The sunlit microoxic niche of the archaeal eukaryotic ancestor comes to light

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

[1]Department of Molecular Biology and Biotechnology, Faculty of Biology and Geology, Babeş-Bolyai University, Cluj-Napoca, Romania.
[2]Institute of Hydrobiology, Department of Aquatic Microbial Ecology, 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, Seestrasse 187, CH-8802 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.

# 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

Summary

The discovery of Asgardaeota archaea, the closest extant relative of eukaryotes to date, has reignited the two-domain of life theory. While it is apparent that Asgardaeota encode multiple eukaryotic-specific proteins, the lack of genomic information and metabolic characterization has precluded inferences about the closest eukaryotic ancestor and the metabolic landscape that laid the grounds for the emergence of the hallmark eukaryotic subcellular architecture. Here, we propose that Heimdallarchaeia (phylum within Asgardaeota) are the closest extant relatives to all eukaryotes and shed light on their facultative aerobic lifestyle, characterized by the capacity to use Sun's energy and aerobic metabolic pathways unique among archaea. Remarkably, the visualization of Heimdallarchaeia organisms revealed a compacted genetic material that is highly unusual for prokaryotes at large. Our results support an evolutionary model in which both protoeukaryote ancestors (archaeal and bacterial) possessed the metabolic repertoire for oxygenic respiration and point towards a late mitochondrial acquisition.

Main

At the dawn of genomics, the eukaryotes were recognized as amalgamated genetic jigsaws that bore components of both bacterial and archaeal descent[1,2]. This genomic chimerism served as a source of speculation and debate over the nature of the protoeukaryote ancestors and inspirited a plethora of hypothetical scenarios for the processes that led to eukaryogenesis[2,3]. Even though, in light of recent research, eukaryotes came into existence through the interplay between an archaeal host[4] and a bacterial endosymbiont[5], the metabolic milieu of the ancestral prokaryotic lineages and their phylogenetic blueprint still remain elusive. Here, we bridge state-of-the-art cultivation-independent genomics, sensitive molecular phylogenetic analyses and genome-scale metabolic reconstructions in order to shed light upon the deep archaeal ancestors of eukaryotes, as well as the metabolic landscape that favored the rise of the 'nucleated cellular architecture'. The genomic catalogue generated during this study enabled us to confidently resolve the backbone of the Asgardaeota superphylum[6] (the closest archaeal relatives of eukaryotes described to date) and to narrow down the eukaryotes' branching point within the tree of life. Collectively, our analyses revealed that contradictory to current opinions[7] the archaeal protoeukaryote ancestor was likely a facultative aerobe that possessed the ability to harness the Sun's energy *via* rhodopsins, and whose mixotrophic metabolism had already acquired unprecedented (within archaea) circuitries for *de novo* aerobic NAD^+ synthesis.

Asgardaeota phylogenomics: Homology-based searches were employed to recover Asgardaeota-related contigs from *de novo* metagenomic assemblies of two deep-sequenced lake sediment samples (with contrasting salinities). By utilizing a hybrid binning strategy and performing manual inspection and data curation, we obtained thirty-five Asgardeota MAGs (metagenome-assembled genomes), spanning three (out of 4) evolutionary lineages within the superphylum: Lokiarchaeia (23), Thorarchaeia (10), and Heimdallarchaeia (2). To the best of our knowledge, by accurately binning 6 026 contigs (total length 55.75 Mbp, average contig length 9 252.5 bp) we generated the largest genomic dataset available to date for this superphylum (in contrast all publicly available MAGs amount to 47.2 Mbp). Due to the challenges associated with reconstructing the evolutionary relationships between archaea and eukayotes[6], in our inferences we used only those MAGs (n= 8; **Supplementary Table S1**) that harbored at least 75% of total phylogenetic markers (See **Supplementary Table S3**). The maximum-likelihood phylogenetic tree, based on concatenation of small (SSU) and large (LSU) ribosomal RNA genes, pictured for the first time a topology in which eukaryotes branched with high-support from within Asgardeota (archaea) (**Supplementary Figure S1a**). Even more remarkably, additionally to recreating a previously described Asgardeota/Eukaryota

branching pattern[6], we provide unprecedented support for a close evolutionary linkage between Heimdallarchaeia (Asgardaeota superphylum) and eukaryotes (SH-aLRT=97.5; UFBoot=100) (**Supplementary Figure S1a**). The genome-focused phylogeny of Asgardaeota revealed a pattern of ancestry, divergence, and descent, in which Heimdallarchaeia comprises the basal branch of the superphylum and Thor-/Odinarchaeia the youngest one (**Figure 1a**). Although dissimilar in branching pattern with the SSU + LSU tree, the phylogenomic one was found to be robust (**Figure 1a**) and to support a topology bought into attention by an earlier study[6]. The SR4-recoded[8] Bayesian tree (maxdiff=0.1) resolved with high support (PP=1) the monophyly of Asgardaeota/Eukaryota, but failed to confidently resolve the internal topology of the superphylum and the branching point of eukaryotes (**Figure 1b**). Noteworthy, in both SR4-recoded Bayesian (**Figure 1b**) and maximum-likelihood phylogenies (**Supplementary Figure S1b**) the eukaryotes caused branch instability for Heimdallarchaeia, which was attracted without statistical support within the superphylum (PP=0.52; SH-aLRT=92.4 and UFBoot=88). To further substantiate the phylogenetic connection between Asgardaeota members and eukaryotes, we screened all the recovered MAGs and the publicly available ones (14) for the presence of potential eukaryotic signature proteins (ESP). Similar to previous reports[6,9,10], the MAGs were found to be highly enriched with ESP (**Figure 1c**), which further reinforced their ancestral linkage to eukaryotes. In addition to the reported ESP[6], we identified a potential subunit of the COPII vesicle coat complex (associated with intracellular vesicle traffic and secretion) in Thorarchaeia and proteins that harbor the N-terminal domain of folliculin - a eukaryote-specific protein which is known to be involved in membrane trafficking in humans[11] (**Figure 1c)** in Lokiarchaea. Furthermore, we retrieved conclusive hits for the ESP-related domains Ezrin/radixin/moesin C-terminal domain and active zone protein ELKS in Lokiarchaeia.

A novel clade of rhodopsins: Recent findings reporting the presence of a novel family of rhodopsins[12] (i.e. heliorhodopsins; abbreviated as HeR) in monoderms[13] encouraged us to perform a dedicated screening in all available Asgardaeota MAGs. The results indicated that one of the Heimdallarchaeia MAGs (i.e Heimdallarchaeota RS678) encoded two heliorhodopsins and what appears to be, as suggested by the presence of a *Exiguobacterium*-like DTK motif[14] and phylogenetic proximity, a type-1 proton-pumping rhodopsin (**Figure 2**). To the best of our knowledge, this is the first report of a proton-pumping rhodopsin in Asgardaeota. Remarkably, we found that the Asgardaeota MAGs recovered during this study encoded rhodopsin sequences similar in membrane orientation to type-1 rhodopsins, and which organized during phylogenetic analysis in a monophyletic clade (SBS=1) placed in-between HeR and type-1 ones (**Figure 2**). Multiple sequence alignments showed: i) homology between transmembrane helices 1, 6 and 7 of these new rhodopsins and the type-1 rhodopsins, while helix 3 was homologous to HeR and ii) presence of additional characteristic HeR motifs (e.g. RWxF motif similar to RWxE of HeR rather than the RYxD motif in most type-1 rhodopsins, and replacement of a proline residue (P91) conserved in type-1 rhodopsins by serine (S91) in both HeR and the new Asgardaeota-found ones) (**Supplementary Figure S2**). Given their phylogenetically intermediate position, as type-1 rhodopsins closest to HeR, and presence of features found in both type-1 and HeR, we denote them as schizorhodopsins (schizo, 'split', plus 'rhodopsin', abbreviated as SzR). The very recent discovery of HeR and their inconclusive functional role[12,13] precludes tentative functional assertions for SzR capacity in Asgardaeota. However, the plethora of rhodopsins that we identified in Heimdallarchaeia (putative type-1 proton pumps, HeR and SzR), together with the SzR found in Lokiarchaeia and Thorarchaeia suggests that, during its evolutionary history, Asgardaeota was present in light-exposed habitats. Moreover, we consider that the primary niche of these rhodopsin-bearing microbes is most likely the top, light exposed sediment layers. Their recovery from deeper strata may be explained by the high deposition rates characteristic for the sampling locations (typically a few cm per year)[15].

Evidences for an aerobic lifestyle in Heimdallarchaeia: The genome-scale metabolic reconstruction placed special emphasis on Heimdallarchaeia, since it was suggested by the above-mentioned phylogenetic analyses to encompass the most probable candidates (to date) for the archaeal protoeukaryote. While the anaerobic lifestyles inferred for Loki-[7] and Thorarchaeia[10] were considered to be accompanied by autotrophy[7] and respectively mixotrophy[10], no consistent metabolic reconstructions exist to date for Heimdallarchaeia. The performed physiological inferences pointed towards mixotrophic lifestyles (for Asgardaeota), simultaneously showing the presence of transporters for the uptake of exogenous organic matter and the metabolic circuitry responsible for its catabolism (see Supplemental Results and Discussion). Remarkably, we found oxygen-dependent metabolic pathways in Heimdallarchaeia, which will be further presented in contrast to the ones harbored by the anaerobic Loki- and Thorarchaeia.

Heimdallarchaeia were inferred to possess components of the aerobic respiration blueprint: a complete tricarboxylic acid cycle (TCA) supported by an electron transport chain (ETC) containing: V/A-type ATPase, succinate dehydrogenase, NADH:quinone oxidoreductase, and the cytochrome c oxidase (**Figure 3**). While in Thor- various components of the TCA were found to be missing, in Lokiarchaeia the complete TCA was associated with: isocitrate dehydrogenases, 2-oxoglutarate-ferredoxin oxidoreductases, and ATP-citrate lyases, pointing towards the presence of a reverse tricarboxylic acid cycle (rTCA). Thus, in contrast with Heimdallarchaeia, which utilizes TCA to fuel their catabolic machinery (**Figure 3**), Lokiarchaeia uses rTCA for autotrophic CO_2 assimilation. While the V/A-type ATPase complex appears to be complete in Loki- and Thorarchaeia, the other components involved in the oxidative phosphorylation processes were not identified.

As nicotinamide adenine dinucleotide (NAD^+) is an essential cofactor in redox biochemistry and energetics[16] (e.g. linking TCA and ETC), we investigated its *de novo* synthesis mechanisms (**Figure 4**). As expected all Asgardaeota phyla were found to harbor the aspartate pathway[17]- a set of metabolic transformations that can occur in both presence or absence of oxygen[18], and which are characteristic for most prokaryotes and the plastid-bearing eukaryotes (obtained through lateral gene transfer from their cyanobacterial endosymbiont)[16]. Surprisingly, Heimdallarchaeia presented in addition to the aspartate pathway the exclusively aerobic kynurenine one[19] (**Figure 4**), which is reported to be present in few bacterial groups and eukaryotes[16]. The phylogenetic reconstruction and evolutionary history inferences showed that this pathway, which is considered to be present in the protoeukaryote ancestor[16], was probably acquired by Heimdallarchaeia through lateral gene transfer from bacteria (**Supplementary Figure S3**). As far as the authors are aware, Heimdallarchaeia are the first archaeal organisms with the aerobic kynurenine pathway. Curiously while Heimdall_LC_3 was found to contain the complete set of genes required for both pathways, Heimdall_LC_2 and Heimdall_RS678 encoded just the genes affiliated with the kynurenine one (**Figure 4**). As the aspartate pathway was reported to function in both oxygen absence (L-aspartate oxidase uses fumarate as electron acceptor)[18] and presence (L-aspartate oxidase uses O_2 as electron acceptor), the existence of the kynurenine pathway in Heimdall_LC_3 appears redundant. By corroborating the presence/absence pattern of the aspartate pathway in Asgardaeota (**Figure 4**) with the reconstructed evolutionary history of Heimdallarchaeia (**Figure 1a**, b; **Supplementary Figure S1a, b**) and blastp similarity searches (for Heimdall_LC_3 L-aspartate oxidase), we inferred that this pathway functioned exclusively under anaerobic conditions. Furthermore, the introgression of kynurenine genes in Heimdallarchaeia appears to be caused by an expansion towards an oxygen-containing niche, which during evolutionary history (from Heimdall_LC_3 to Heimdall_LC_2/Heimdall_RS678) favored the xenologous replacement of the aspartate pathway with the kynurenine one.

Within the anaplerotic metabolism, the reversible transformation of pyruvate into acetyl-CoA and formate can be accomplished by pyruvate formate lyases, which were inferred to be present in all three phyla. Formate produced during this enzymatic process, or by the activity of arylformamidase (kynurenime formamidase) in Loki and Heimdallarchaeia could be further oxidized (by formate

dehydrogenases) and used for quinone/cytochrome pool reduction, or introduced into the one-carbon metabolism and utilized for the synthesis of: purines, glycine, formylmethionine, etc. (**Figure 3**). Uniquely in Heimdallarchaeia we inferred that formate could act as electron donor during aerobic respiration through the actions of the heterotrimeric formate dehydrogenase O. This enzyme facilitates the usage of formate under aerobiosis, and together with nitrate reductase Z (also present solely in Heimdallarchaeia) may participate in a formate to nitrate electron transport pathway that is active when cells are shifted from aerobic to anaerobic conditions[20,21]. The presence of genes encoding pyruvate oxidases (poxL) in Heimdallarchaeia (i.e. LC_2 and LC_3) implies further oxygen usage, as the enzyme employs it in the pyruvate pool decarboxylation process (**Figure 3**).

The comparative genomic analyses also revealed that the three Asgardaeota phyla rely upon glycolysis (i.e. type Embden-Meyerhof-Parnas) to fuel their metabolic machinery. Unexpectedly, three Heimdallarchaeia MAGs (LC_3, AB_125 and AMARA_4) were found to employ non-canonical ADP-dependent kinases that use ADP instead of the typical ATP as the phosphoryl group donor[22] in their glycolytic pathways. Furthermore, they seemed to be bifunctional ADP-dependent glucokinase/phosphofructokinase, which was puzzling since the presence of 6-phosphofructokinases (LC_3 and AB_125) would render their bifunctionality redundant. In order to elucidate the role of the putative bifunctional enzymes, we reconstructed the evolutionary history of the ADP-dependent kinases and inferred that they possess glucokinase activity (based on tree topology and the conserved functional residue E172) (**Supplementary Figure S4**). Additionally, we observed that the deepest branching Heimdallarchaeia (LC_3) harbored the archaeal-type of the enzyme, while the younger ones (AB_125 and AMARA_4) clustered together with the eukaryotic-type (**Supplementary Figure S4**). Although it is easy to assume that cells under low energy conditions (e.g. limiting O_2 availability) could highly benefit from using residual ADP to activate sugar moieties and fuel their glycolysis[23], the metabolic advantage conferred by these ADP-dependent kinases is unclear.

Although pentoses could be recycled via nucleotide degradation in all Asgardaeota phyla, their synthesis differs between Loki-/Heimdallarchaeia that likely utilize the reverse ribulose monophosphate pathway, and Thorarchaeia that employ the xylulose part (of the non-oxidative branch) of the hexose monophosphate one. The identified homologues for ribulose 1,5-bisphosphate carboxylase/oxygenase (RubisCO) genes were found to appertain to the types: III (Loki- and Heimdallarchaeia) and IV (Loki- and Thorarchaeia) (**Supplementary Figure S5**). While RuBisCO is a key enzyme for CO_2 fixation in the Calvin-Benson-Bassham cycle, the absence of phosphoribulokinase renders this metabolic pathway highly improbable. However, we consider that the MAGs encoding type III-like RuBisCO (assigned to Loki- and Heimdallarchaeia) use the nucleotide monophosphate degradation pathway[24,25], thus performing CO_2 fixation by linking nucleoside catabolism to glycolysis/gluconeogenesis. This conclusion is supported by the co-occurrence of genes encoding for: RuBisCO type III, AMP phosphorylases, ribose 1,5-bisphosphate isomerases, and carbonic anhydrases. While carbon monoxide (CO) can be used as carbon and energy source in both aerobic and anaerobic metabolisms[26], the types of enzymes involved in the reaction are dependent upon the available electron acceptor. Thus, while Heimdallarchaeia harbor all three major subunits of the aerobic carbon monoxide dehydrogenases (CODH), Loki- and Thorarchaeia encoded the oxygen-sensitive carbon monoxide dehydrogenase/acetyl-CoA synthase (CODH/ACS). We infer that while Heimdallarchaeia uses CO to obtain energy by shuttling the electrons generated from CO oxidation to oxygen or nitrate, Thor- and Lokiarchaeia may utilize CO as both electron source and intermediary substrate in the ancient Wood–Ljungdahl carbon fixation pathway[27,28] (through CODH/ACS).

Among Asgardaeota, the Heimdallarchaeia were found to possess genes encoding for sulfide:quinone oxidoreductases, enzymes used in sulfide detoxification and energy generation through quinone pool reduction (**Figure 3**). As sulfide binds to cytochrome c oxidase system and

inhibits aerobic respiration[29], the presence of these enzymes in Heimdallarchaeia could point towards a detoxification role. The fact that one Heimdallarchaeia MAGs described in this study (i.e. AMARA_4) had the sulfide:quinone oxidoreductase gene and the other Asgardaeota 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 anaerobic metabolism, nor is it part of a conserved energy generation strategy in Asgardaeota. The superoxide dismutase, catalases, and glutathione peroxidases found in Heimdallarchaeia may act in alleviating the oxidative damaged generated by a facultative aerobic metabolism.

CARD-FISH visualization of Loki- and Heimdallarchaeia. Two phylogenetic probes targeting the 16S rRNA of Loki- and Heimdallarchaeia, respectively (**Supplementary Table S8, Supplementary Figure S7**) were successfully applied to sediment samples of different depth layers. Members of both phyla were rare and seemed to be totally absent below sediment depths of 40 cm. All observed Heimdallarchaeia were similar in cell size (2.0±0.4 µm length x 1.4±0.3 µm width, n=15) and of conspicuous shape with DNA condensed (0.8±0.2 x 0.5±0.1 µm) at the center of the cells (**Figure 5 a-c, Supplementary Figure S8**), what is rather atypical for prokaryotes. In contrast, Lokiarchaeia were very diverse in shape and size and we could distinguish at least three different morphotypes: small-medium sized ovoid cells (2.0±0.5 x 1.3±0.3 µm, n=23, **Figure 5 d-f, Supplementary Figure S9**), large round cells (3.8 x 3.6 µm, **Figure 5 g-i**) with condensed DNA at the center, and large rods/filaments (4.4±1.2 x 1.4±0.5 µm, n=6, **Figure 5 j-l, Supplementary Figure S9**) with filamentous, condensed DNA (2.7±1.4 x 0.4±0.1 µm) that were exclusively present in 30-40 cm sediment depth. This high diversity in morphology in Lokiarchaeia most likely also reflects a higher sampling of the phylogenetic diversity within the phylum.

Discussion

The mainstream theories on the subject of eukaryogenesis[3,30] (which date back to late 20th century), have been recently refuted by improved phylogenetic methods and increased genomic sampling[6,7]. Even after experiencing a new revival[7], the current endosymbiotic theory fails to envision the environmental and metabolic context in which the 'nucleated cellular architecture' emerged. Moreover, it appears to find itself gravitating around a three-decade old theory centered on anaerobic syntrophy[31] (i.e. hydrogen hypothesis). In this study, we show that: i) mixotrophy and harnessing Sun's energy (*via* rhodopsins) are the *modus vivendi* of Asgardaeota, and that ii) aerobic respiration was present in the archaeal protoeukaryote ancestor before the emergence of eukaryogenesis-associated phenomena. The 'aerobic-protoeukaryotes' model surpasses some of the theoretical shortcomings of the 'hydrogen hypothesis' by envisioning an endosymbiotic association in which the primordial function of the bacterial counterpart (i.e. oxidative phosphorylation) would not be detrimental for the existence of the archaeal one (caused by oxygen exposure). We postulate that in the light of these results the previous hypotheses regarding the metabolic capacities of the protoeukaryote ancestors need to be reevaluated.

Methods

Sampling: Sediment sampling was performed on 10 October 2017 at 12:00 in Tekirghiol Lake, Romania, (44°03.19017 N, 28°36.19083 E) and on 11 October 2017 at 15:00 in Amara Lake, Romania, (44°36.30650 N, 27°19.52950 E). Two plunger cores of 0.3 m each were collected from a water depth of 0.8 m in Tekirghiol Lake and 4 m in Amara Lake. Sediment samples were stored in the dark at 4 °C and processed within 24 hours after collection.

DNA extraction and purification: DNA was extracted from approximately 10 g of wet sediment from each mixed core sample using the DNeasy PowerMax Soil Kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. Extracted DNA was further purified by passing it through humic acid removal columns (type IV-HRC) provided in the ZR Soil Microbe DNA MiniPrep kit (Zymo Research,

Irvine, CA, USA). Purified DNA was quality checked and quantified using a ND-1000 NanoDrop spectrophotometer (Thermo Scientific, Waltham, MA, USA). DNA integrity was assessed by agarose gel (1%) electrophoresis and ethidium bromide staining. The samples were denominated as AMARA and TEKIR in accordance with their site of origin. From each sample, 4 µg of pure DNA were vacuum dried in a SpeedVac concentrator (Thermo Scientific, Hilden, Germany) and shipped for library construction and NGS sequencing to Macrogen (Seoul, South Korea).

Sequencing and data preprocessing: Library preparation was performed by the commercial company by using the TruSeq DNA PCR Free Library prep kit (Illumina). Whole-genome shotgun sequencing of the 150 paired-end libraries (350bp insert size) was done using a HiSeq X (Illumina) platform. The amount of total raw sequence data generated for each metagenome was: 64.5Gbp for Amara and 57.6 Gbp for Tekirghiol. Preprocessing of raw Illumina reads was carried out by using a combination of software tools from the BBMap[32] project (https://github.com/BioInfoTools/BBMap/). Briefly, bbduk.sh was used to remove poor quality sequences (qtrim=rl trimq=18), to identify phiX and p-Fosil2 control reads (k=21 ref=vectorfile ordered cardinality), and to remove Illumina adapters (k=21 ref=adapterfile ordered cardinality).

Abundance estimation for Loki- and Heimdallarchaeia: Preprocessed Illumina sets from Amara and Tekirghiol lakes, as well as published[9] set SRX684858 from Loki's castle marine sediment metagenome, were subsampled to 20 million reads by reformat.sh[33]. Each subset was queried for putative RNA sequences by scanning with UBLAST[34] against the non-redundant SILVA SSURef_NR99_132 database[35], that was priorly clustered at 85% sequence identity by UCLUST[34]. Identified putative 16S rRNA sequences (e-value < 1e-5) were screened using SSU-ALIGN[36]. Resulting bona fide 16S rRNA sequences were compared by blastn[37] (e-value <1e-5) against the curated SILVA SSURef_NR99_132 database. Matches with identity ≥ 80% and alignment length ≥90 bp were considered for downstream analyses. Sequences assigned to Loki- and Heimdallarcheia were used to calculate abundances for these taxa in their originating environments **(Figure 5).**

Metagenome assembly and binning: *De novo* assembly of preprocessed paired-end Illumina reads was done by Megahit[38] v.1.1.1 with k-mer list: 39, 49, 69, 89, 109, 129, 149, and with default parameters. Assembled contigs with minimum nucleotide fragment length of 3 kbp were binned by a combination of taxonomy-dependent and -independent methods. Protein coding genes were predicted by MetaProdigal[39]. Taxonomy dependent binning was achieved by first assigning taxonomy labels to the predicted genes by performing screenings with MMseqs2[40] against the NR database. All contigs with a minimum of 30 % genes assigned to Asgardaeota were used for taxonomy-independent binning. Mean base coverage for each contig was computed with bbwrap.sh (default parameters) by mapping to assembled contigs the preprocessed reads from AMARA and TEKIR datasets. Hybrid binning (based on tetranucleotide frequencies and coverage data) was performed using MetaBAT2[41] with default parameters. Bin completeness, contamination and strain heterogeneity were estimated using CheckM[42] with default parameters. Poorly resolved bins (i.e. contamination > 10%, unbinned contigs) were further manually curated by a combination of tetranucleotide frequency PCA graphs and repeated rounds of contamination/completeness assessment by CheckM. Final curated bins with CheckM estimated completeness above 10% and contamination below 10% were denominated as metagenome assembled genomes (MAGs). A total of 35 MAGs were recovered: 23 Lokiarchaeia, 10 Thorarchaeia and 2 Heimdallarchaeia (**Supplementary Table S1**). Unbinned contigs were kept for further analyses (total nucleotide bases/site: 3.46 Mbp Amara and 4.06 Mbp Tekirghiol).

Genome annotation: Publicly available Asgardeota genomes were downloaded from the NCBI Genome section (https://www.ncbi.nlm.nih.gov/genome). Coding sequences were predicted *de novo* with Prokka[43] for all available Asgard MAGs (35 from this study, 14 from NCBI – Accession numbers can be found in **Supplementary Table S2**). BlastKOALA[44] was used to assign KO identifiers (K numbers) to orthologous genes (**Supplementary Table S3**). Inferences of metabolic pathways and

general biological functions were conducted with the online KEGG mapping tools (https://www.genome.jp/kegg/kegg1b.html) using summarized KO numbers assigned to each group. Odinarchaeia was not considered for metabolic reconstruction due to lack of new genomic data. Ribosomal RNA (rRNA)-coding regions (16S, 23S) and transfer RNA (tRNA)-coding regions were predicted with Barrnap (https://github.com/tseemann/barrnap) and tRNAscan-SE[45], respectively. All predicted proteins were queried against NCBI NR, COGs (cluster of orthologous groups) and arCOGs (archaeal cluster of orhtologous groups, 2014)[46]. A local version of InterProScan[47] was used with default settings to annotate protein domains. Potential eukaryote specific proteins (ESPs) were identified based on previously published lists of IPR domains[6] (**Supplementary Table S4**) identified in Asgard archaea. New ESPs were searched based on key words related to eukaryotic specific processes and/or structures. All IPR domains present exclusively in newly recovered Asgardaeota genomes were manually screened by querying accession numbers against the online InterPro database (https://www.ebi.ac.uk/interpro/search/sequence-search), for associations with eukaryotic specific domains. A previously identified[6] ESP - DNA polymerase epsilon, catalytic subunit (IPR029703) - was identified by querying all MAG proteomes with human sequences (**Supplementary Table S4**). Several candidate ESP sequences were further analyzed using jackhmmer[48], Phyre2[49] and Phobius[50].

Phylogenetic trees: A total of 131 taxa were considered for concatenated small subunit (SSU) and larger subunit (LSU) ribosomal RNA phylogenetic analyses, consisting of: 97 archaea (37 Euryarchaeota, 24 Crenarchaeota, 2 Bathyarchaeota, 15 Thaumarchaeota, 3 Aigarchaeota, 2 Korarchaeota, 14 Asgardeota), 21 bacteria and 13 eukaryotes (**Supplementary Table S5**). SSU and LSU sequences were aligned independently by PRANK[51] (parameters: -DNA +F), trimmed using BMGE[52] (–m DNAPAM250:4 –g 0.5) and concatenated. Members of the DPANN group of Archaea (Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota, Nanohaloarchaeota, Woesearchaeota, and Pacearchaeota) were not included due to their known tendency to cause phylogenetic artefacts (detailed previously[6]). Maximum-Likelihood phylogeny for concatenated SSU/LSU gene sequences was inferred using IQ-TREE (-m GTR+I+G4) with ultrafast bootstrapping -bb 1000 and Shimodaira-Hasegawa testing –alrt 1000[53,54].

A total of 93 taxa were considered for concatenated ribosomal protein phylogenomic analyses, consisting of: 85 Archaea (25 Euryarchaeota, 22 Crenarchaeota, 2 Bathyarchaeota, 4 Thaumarchaeota, 1 Aigarchaeota, 3 Korarchaeota, 21 Asgardeota, 7 DPANN) and 8 eukaryotes (**Supplementary Table S5**). Selection criteria for phylogenomic trees of ribosomal proteins conserved between archaea and eukaryotes has been previously described[6]. Amino-acid sequences for the 55 ribosomal proteins were queried and retrieved based on arCOG annotations. Markers not found in the majority of organisms were discarded, obtaining a final set of 48 markers (**Supplementary Table S6**). Additionally, some proteins that were not identified by arCOG scanning were retrieved from the NCBI Protein section (https://www.ncbi.nlm.nih.gov/protein). Sequences were aligned using PRANK (-protein +F), trimmed with BMGE[52] (-m BLOSUM30 -t AA -g 0.2 -b 3), concatenated, and subjected to SR4 amino acid recoding[8]. Maximum-likelihood trees were generated by IQ-TREE (-bb 1000, -alrt 1000) with ultrafast bootstraping[53] and the custom 'C60SR4' model described in a previous study[6]. Bayesian inference phylogenies were constructed using PhyloBayes MPI 1.8[55], using the CAT-Poisson model. Four chains were run in parallel until estimated maxdiff values calculated by bp_comp (-x 5000 10) fell below the recommended 0.3 threshold, indicating convergence between chains.

Multiple sequence alignment of rhodopsins: The three groups of rhodopsins (type-1, schizorhodopsins and heliorhodopsins), were first aligned independently using T_Coffee[56] (http://tcoffee.crg.cat/) in accurate mode, that employs protein structure information, wherever available, or sequence comparisons with homologues in databases to improve accuracy. These alignments were aligned to each other using the profile alignment mode in T_Coffee.

RuBisCO tree reconstruction: The multiple sequence RuBisCO alignment was built upon a core structural alignment of diverse set of sequences, to which additional sequences were added using T_Coffee. A total of 392 sequences were used for the alignment. MUSCLE[57] was used for aligning the sequences (n=146) of the large subunit of RubisCO (types I-III) and RubisCO-like (type IV) (rbcL, K01601) proteins. Sequences not generated in this study were recovered from previous studies[58,59]. For both alignments the maximum likelihood tree was constructed with FastTree2 using a JTT model, a gamma approximation, and 100 bootstrap replicates.

Phylogenetic inference of Heimdallarchaeia glucokinases and kynurenine pathway proteins: ADP-dependent phosphofructokinase/glucokinase protein sequences were identified by their assigned KO number (K00918) in 3 MAGs (AMARA_4, Heimdall_AB_125, Heimdall_LC_3). Retrieved sequences were used along with 49 other sugar kinases published in a previous study[60]. Protein sequences of components of the kynurenine pathway - tryptophan 2,3-dioxygenase (TDO), kynurenine 3-monooxygenase (KMO) and 3-hydroxyanthranilate 3,4-dioxygenase (HAAO) – that were identified only in Heimdallarchaeia MAGs, were used along with sequences of corresponding enzymes from 12 Eukaryotes and 15 Bacteria that were retrieved from NCBI RefSeq (Accession numbers in **Supplementary Table S7**). MAFFT-L-INS-i[61] (default parameters) and PRANK[51] (parameters: -DNA +F) were used for aligning sugar kinase and respectively kynurenine pathway enzyme sequences followed by trimming using BMGE[52] (-m BLOSUM30 -t AA -g 0.5 -b 3). Single protein maximum likelihood trees were constructed with FastTree2[62], using an accurate search strategy (-mlacc 2 –spr 4 –slownni), and 100 bootstrap replicates.

Probe design and CARD-FISH (Catalyzed reporter deposition fluorescence *in situ* hybridization): All assembled 16S sequences classified as Asgardaeota were aligned with the SINA aligner[63], manually optimized in ARB[64] using SILVA database SSURef_NR99_132[35], and a RAxML tree (Randomized Axelerated Maximum Likelihood tree with GTR-GAMMA model, 100 bootstraps[65]) was constructed (**Supplementary Figure S7**). Probe design for Heimdallarchaeia and Lokiarchaeia based on almost full-length sequences of high quality was done with the probe_design and probe_check tools in ARB. Probes were tested *in silico*[66] and in the laboratory with different formamide concentrations in the hybridization buffer until stringent conditions were achieved (**Supplementary Table S8**). Sediment sampling was performed using a custom mud corer on 22 April 2018 at 12:00 in Tekirghiol Lake, Romania, (44°03.19017 N, 28°36.19083 E) and Amara Lake (44°36'N, 27°20'E, 23 April 2018, 14:00). Seven sediment layers (0-70 cm, in 10-cm ranges) were samples in Tekirghiol Lake and the top 10 cm was sampled in Amara Lake. Sediment samples were fixed with formaldehyde, treated with sonication, vortexing and centrifugation to detach cells from sediment particles[67] and aliquots were filtered onto white polycarbonate filters (0.2 µm pore size, Millipore). CARD-FISH was conducted as previously described with fluorescein labelled tyramides[68]. Filters were counterstained with DAPI and inspected by epifluorescence microscopy (Zeiss Imager.M1). Micrographs of CARD-FISH stained cells were recorded with a highly sensitive charge-coupled device (CCD) camera (Vosskühler) and cell sizes were estimated with the software LUCIA (Laboratory Imaging Prague, Czech Republic).

Accession numbers: All sequence data produced during the study is deposited in the Sequence Read Archive (SRA) database of the National Center for Biotechnology Information (NCBI) and can be found linked to the Bioproject PRJNA483005.

References

1. Rivera, M. C., Jain, R., Moore, J. E. & Lake, J. A. Genomic evidence for two functionally distinct gene classes. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 6239–6244 (1998).
2. McInerney, J. O., O'Connell, M. J. & Pisani, D. The hybrid nature of the Eukaryota and a

consilient view of life on Earth. *Nat. Rev. Microbiol.* **12,** 449 (2014).

3. de Duve, C. The origin of eukaryotes: a reappraisal. *Nature reviews. Genetics* **8,** 395–403 (2007).
4. Eme, L., Spang, A., Lombard, J., Stairs, C. W. & Ettema, T. J. G. Archaea and the origin of eukaryotes. *Nat. Rev. Microbiol.* **15,** 711–723 (2017).
5. Martijn, J., Vosseberg, J., Guy, L., Offre, P. & Ettema, T. J. G. Deep mitochondrial origin outside the sampled alphaproteobacteria. *Nature* **557,** 101–105 (2018).
6. Zaremba-Niedzwiedzka, K. *et al.* Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* **541,** 353–358 (2017).
7. Sousa, F. L., Neukirchen, S., Allen, J. F., Lane, N. & Martin, W. F. Lokiarchaeon is hydrogen dependent. *Nat. Microbiol.* **1,** 16034 (2016).
8. Susko, E. & Roger, A. J. On reduced amino acid alphabets for phylogenetic inference. *Mol. Biol. Evol.* **24,** 2139–2150 (2007).
9. Spang, A. *et al.* Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* **521,** 173 (2015).
10. Liu, Y. *et al.* Comparative genomic inference suggests mixotrophic lifestyle for Thorarchaeota. *ISME J.* **12,** 1021–1031 (2018).
11. Dodding, M. P. Folliculin – A tumor suppressor at the intersection of metabolic signaling and membrane traffic. *Small GTPases* **8,** 100–105 (2017).
12. Pushkarev, A. *et al.* A distinct abundant group of microbial rhodopsins discovered using functional metagenomics. *Nature* **558,** 595–599 (2018).
13. Flores-Uribe, J. *et al.* Heliorhodopsins are absent in diderm (Gram-negative) bacteria: Some thoughts and possible implications for activity. *bioRxiv* (2018).
14. Petrovskaya, L. E. *et al.* Predicted bacteriorhodopsin from Exiguobacterium sibiricum is a functional proton pump. *FEBS Lett.* **584,** 4193–4196 (2010).
15. Alexe, M. *Studiul lacurilor sărate din Depresiunea Transilvaniei*. (Presa Universitară Clujeană, 2010).
16. Ternes, C. M. & Schönknecht, G. Gene Transfers Shaped the Evolution of De Novo NAD(+) Biosynthesis in Eukaryotes. *Genome Biol. Evol.* **6,** 2335–2349 (2014).
17. Lin, H., Kwan, A. L. & Dutcher, S. K. Synthesizing and Salvaging NAD+: Lessons Learned from Chlamydomonas reinhardtii. *PLOS Genet.* **6,** e1001105 (2010).
18. Gazzaniga, F., Stebbins, R., Chang, S. Z., McPeek, M. A. & Brenner, C. Microbial NAD Metabolism: Lessons from Comparative Genomics. *Microbiol. Mol. Biol. Rev.* **73,** 529–541 (2009).
19. Kurnasov, O. *et al.* Aerobic tryptophan degradation pathway in bacteria: novel kynurenine formamidase. *FEMS Microbiol. Lett.* **227,** 219–227 (2003).
20. Sawers, G., Heider, J., Zehelein, E. & Bock, A. Expression and operon structure of the sel genes of Escherichia coli and identification of a third selenium-containing formate dehydrogenase isoenzyme. *J. Bacteriol.* **173,** 4983–4993 (1991).
21. Abaibou, H., Pommier, J., Benoit, S., Giordano, G. & Mandrand-Berthelot, M. A. Expression and characterization of the Escherichia coli fdo locus and a possible physiological role for aerobic formate dehydrogenase. *J. Bacteriol.* **177,** 7141–7149 (1995).
22. 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).
23. Dorr, C., Zaparty, M., Tjaden, B., Brinkmann, H. & Siebers, B. The hexokinase of the hyperthermophile Thermoproteus tenax. ATP-dependent hexokinases and ADP-dependent glucokinases, teo alternatives for glucose phosphorylation in Archaea. *J. Biol. Chem.* **278,** 18744–18753 (2003).
24. Aono, R. *et al.* Enzymatic characterization of AMP phosphorylase and ribose-1,5-bisphosphate isomerase functioning in an archaeal AMP metabolic pathway. *J. Bacteriol.* **194,** 6847–6855

(2012).
25. Kono, T. *et al.* A RuBisCO-mediated carbon metabolic pathway in methanogenic archaea. *Nat. Commun.* **8,** 14007 (2017).
26. Techtmann, S. M. *et al.* Evidence for horizontal gene transfer of anaerobic carbon monoxide dehydrogenases. *Front. Microbiol.* **3,** 132 (2012).
27. Martin, W. F. Hydrogen, metals, bifurcating electrons, and proton gradients: The early evolution of biological energy conservation. *FEBS Lett.* **586,** 485–493 (2012).
28. Takami, H. *et al.* A Deeply Branching Thermophilic Bacterium with an Ancient Acetyl-CoA Pathway Dominates a Subsurface Ecosystem. *PLoS One* **7,** e30559 (2012).
29. 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).
30. Martin, W. F., Garg, S. & Zimorski, V. Endosymbiotic theories for eukaryote origin. *Philos. Trans. R. Soc. B Biol. Sci.* **370,** 20140330 (2015).
31. Martin, W. & Muller, M. The hydrogen hypothesis for the first eukaryote. *Nature* **392,** 37–41 (1998).
32. Bushnell, B., Rood, J. & Singer, E. BBMerge - Accurate paired shotgun read merging via overlap. *PLoS One* **12,** e0185056 (2017).
33. Bushnell, B. BBMap short read aligner. *Univ. California, Berkeley, California. URL http//sourceforge. net/projects/bbmap* (2016).
34. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26,** 2460–2461 (2010).
35. 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).
36. Nawrocki, E. P. Structural RNA Homology Search and Alignment using Covariance Models. (Washington University School of Medicine, 2009).
37. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410 (1990).
38. Li, D., Liu, C.-M., Luo, R., Sadakane, K. & Lam, T.-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31,** 1674–1676 (2015).
39. Hyatt, D., LoCascio, P. F., Hauser, L. J. & Uberbacher, E. C. Gene and translation initiation site prediction in metagenomic sequences. *Bioinformatics* **28,** 2223–2230 (2012).
40. Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35,** 1026 (2017).
41. Kang, D. D., Froula, J., Egan, R. & Wang, Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* **3,** e1165 (2015).
42. 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).
43. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30,** 2068–2069 (2014).
44. Kanehisa, M., Sato, Y. & Morishima, K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *J. Mol. Biol.* **428,** 726–731 (2016).
45. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25,** 955–964 (1997).
46. Makarova, K. S., Wolf, Y. I. & Koonin, E. V. Archaeal Clusters of Orthologous Genes (arCOGs): An Update and Application for Analysis of Shared Features between Thermococcales, Methanococcales, and Methanobacteriales. *Life (Basel, Switzerland)* **5,** 818–840 (2015).
47. Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics*

30, 1236–1240 (2014).
48. Finn, R. D. *et al.* HMMER web server: 2015 update. *Nucleic Acids Res.* **43,** W30-8 (2015).
49. 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).
50. Käll, L., Krogh, A. & Sonnhammer, E. L. L. Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. *Nucleic Acids Res.* **35,** W429–W432 (2007).
51. Loytynoja, A. Phylogeny-aware alignment with PRANK. *Methods Mol. Biol.* **1079,** 155–170 (2014).
52. Criscuolo, A. & Gribaldo, S. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol. Biol.* **10,** 210 (2010).
53. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* **35,** 518–522 (2018).
54. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.* **32,** 268–274 (2015).
55. Lartillot, N., Rodrigue, N., Stubbs, D. & Richer, J. PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Syst. Biol.* **62,** 611–615 (2013).
56. Notredame, C., Higgins, D. G. & Heringa, J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302,** 205–217 (2000).
57. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32,** 1792–1797 (2004).
58. 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).
59. 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).
60. 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).
61. Katoh, K. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30,** 3059–3066 (2002).
62. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One* **5,** e9490 (2010).
63. Pruesse, E., Peplies, J. & Glockner, F. O. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* **28,** 1823–1829 (2012).
64. Ludwig, W. *et al.* ARB: a software environment for sequence data. *Nucleic Acids Res.* **32,** 1363–1371 (2004).
65. Stamatakis, A., Ludwig, T. & Meier, H. RAxML-II: a program for sequential, parallel and distributed inference of large phylogenetic trees. *Concurr. Comput. Pract. Exp.* **17,** 1705–1723 (2005).
66. Yilmaz, L. S., Parnerkar, S. & Noguera, D. R. mathFISH, a web tool that uses thermodynamics-based mathematical models for in silico evaluation of oligonucleotide probes for fluorescence in situ hybridization. *Appl. Environ. Microbiol.* **77,** 1118–1122 (2011).
67. Ishii, K., Mussmann, M., MacGregor, B. J. & Amann, R. An improved fluorescence in situ hybridization protocol for the identification of bacteria and archaea in marine sediments. *FEMS Microbiol. Ecol.* **50,** 203–213 (2004).
68. Pernthaler, A., Pernthaler, J. & Amann, R. Fluorescence in situ hybridization and catalyzed reporter deposition for the identification of marine bacteria. *Appl. Environ. Microbiol.* **68,** 3094–3101 (2002).

End notes.

Acknowledgements

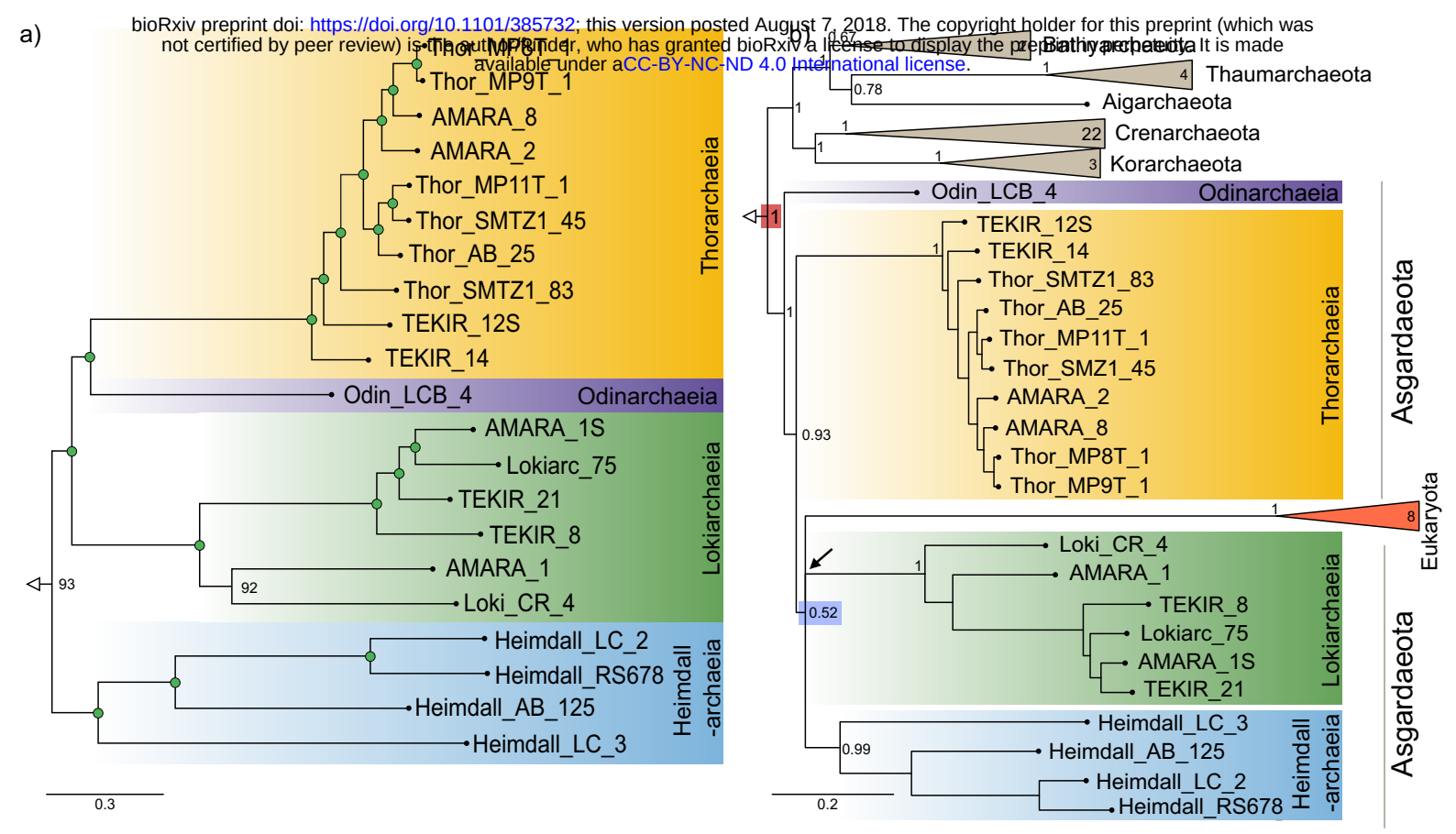
We are thankful to: Z. Keresztes, V. Muntean, M. Alexe, A. Cristea, and A. Baricz for their technical support during sampling and sample preparation. The contribution of E. A. Levei and M. Senila in performing chemical analysis is kindly acknowledged. P-A.B was supported by the research grant PN-III-P4-ID-PCE-2016-0303 (Romanian National Authority for Scientific Research). H.B.L. was supported by the research grants: PN-III-P4-ID-PCE-2016-0303 (Romanian National Authority for Scientific Research) and STAR-UBB Advanced Fellowship-Intern (Babeș-Bolyai University). A-Ş.A. was supported by the research grants: 17-04828S (Grant Agency of the Czech Republic) and MSM200961801 (Academy of Sciences of the Czech Republic). M.M. was supported by the Postdoctoral program PPPLZ (Academy of Sciences of the Czech Republic).R.G. was supported by the research grant 17-04828S (Grant Agency of the Czech Republic).

Contributions

H.L.B. and P-A.B. designed the study. P-A.B., A-Ş.A. and R.G. wrote the manuscript. P-A.B., A-Ş.A., R.G., M.M.S and M.M. analyzed and interpreted the data. O.B., K.I. and H.K. performed rhodopsin data analyses. M.M.S. did CARD-FISH imaging. All authors commented on and approved the manuscript.

Competing interests

The authors declare no competing interests.



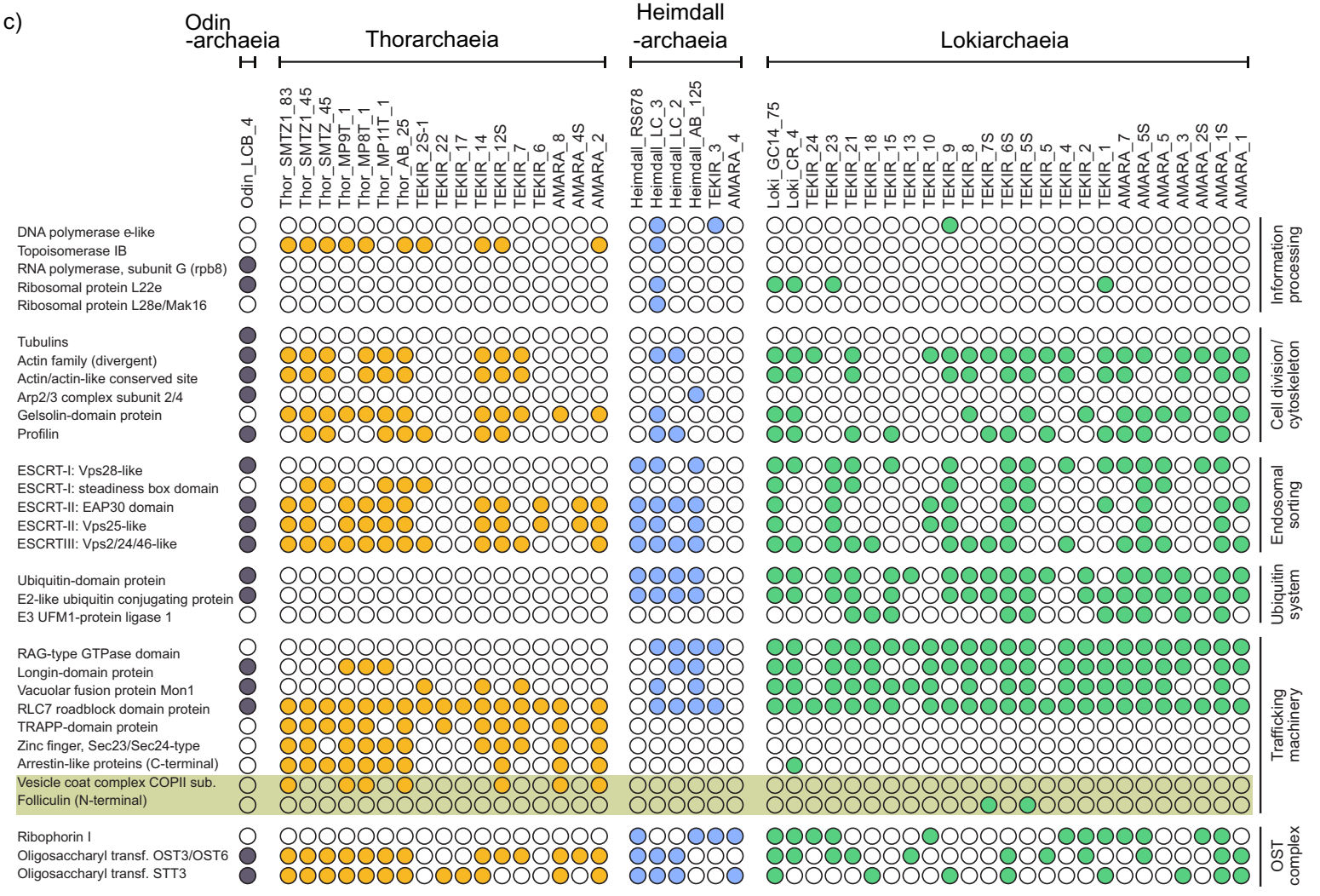


Figure 1. Asgardaeota phylogenomics. a) Maximum-likelihood phylogeny of Asgardaeota superphylum. The green circles highlight UFBoot values higher than 95. b) Asgardaeota phylogeny generated through Bayesian inference. The posterior probability values are shown above the internal nodes. High support for Eukaryota/Asgardaeota monophyly, and the low support for Eukaryota/Heimdallarchaeia association is indicated by red and blue rectangles on the nodes respectively. The black arrow indicates the unresolved position of Lokiarchaeia. Scale bars indicate the number of substitutions per site. Panel (c) provides a census of the eukaryotic signature proteins (ESP) found in the recovered MAGs. The grey box highlights new ESP identified during this study.

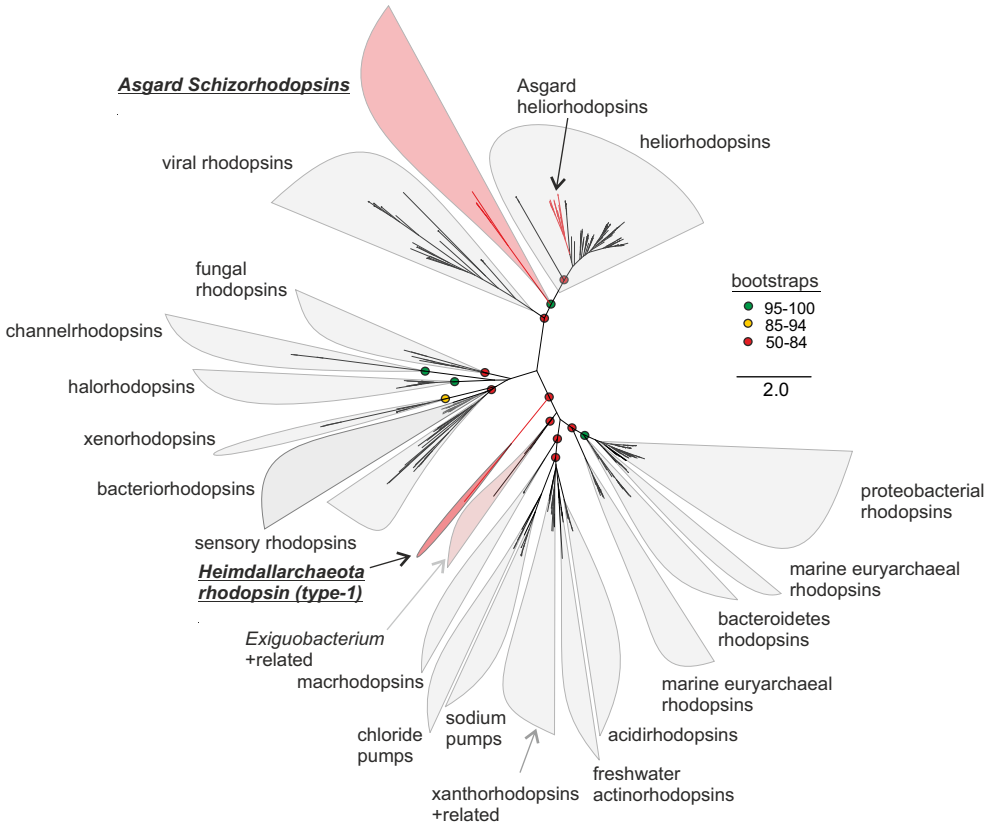


Figure 2. Phylogenetic analysis of rhodopsins. An unrooted maximum-likelihood tree of all Asgardaeota schizorhodopsins identified in this work, heliorhodopsins and representative known type-1 rhodopsins, is shown. The branches colored red are sequences from the Asgardaeota. Bootstrap values on nodes are indicated by colored circles (see color key at the right).

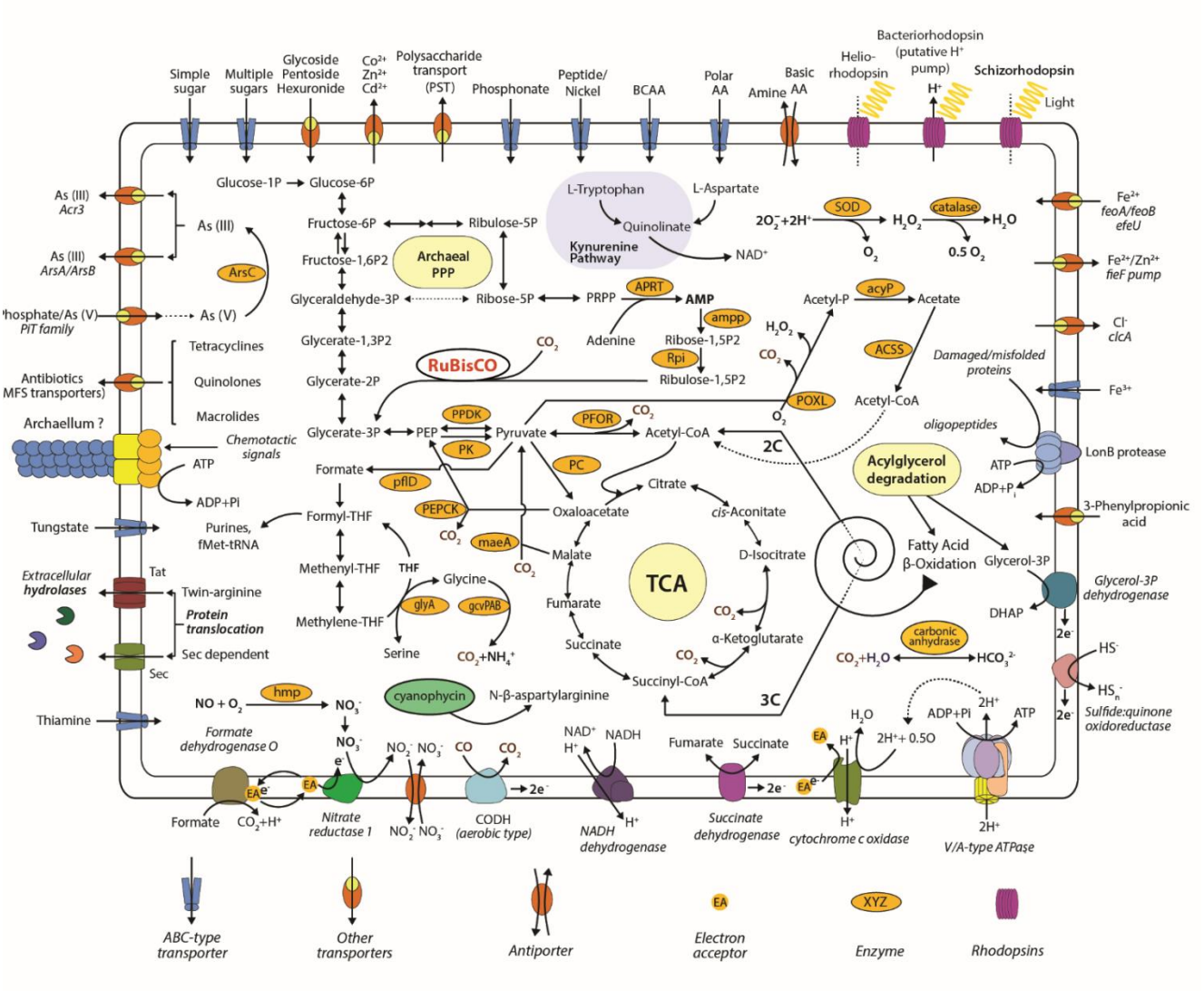


Figure 3. Metabolic reconstruction of Heimdallarchaeia. The text present in the yellow panels depicts names of pathways and metabolic processes. Abbreviations: ArsC - arsenate reductase (glutaredoxin); hmp - nitric oxide dioxygenase; PPDK - pyruvate, phosphate dikinase; PK - pyruvate kinase; PEPCK - phosphoenolpyruvate carboxykinase; maeA - malate dehydrogenase (decarboxylating); glyA - glycine hydroxymethyltransferase; gcvPAB - glycine dehydrogenase; PFOR - pyruvate ferredoxin oxidoreductase; APRT - AMP pyrophosphorylase; ampp - AMP phosphorylase; Rpi - ribose-5-phosphate isomerase; SOD – superoxide dismutase; catalase; acyP - acylphosphatase; poxL - pyruvate oxidase; ACSS - acetyl-CoA synthetase and carbonic anhydrase. RuBisCO - Ribulose-1,5-bisphosphate carboxylase/oxygenase is shown in red.

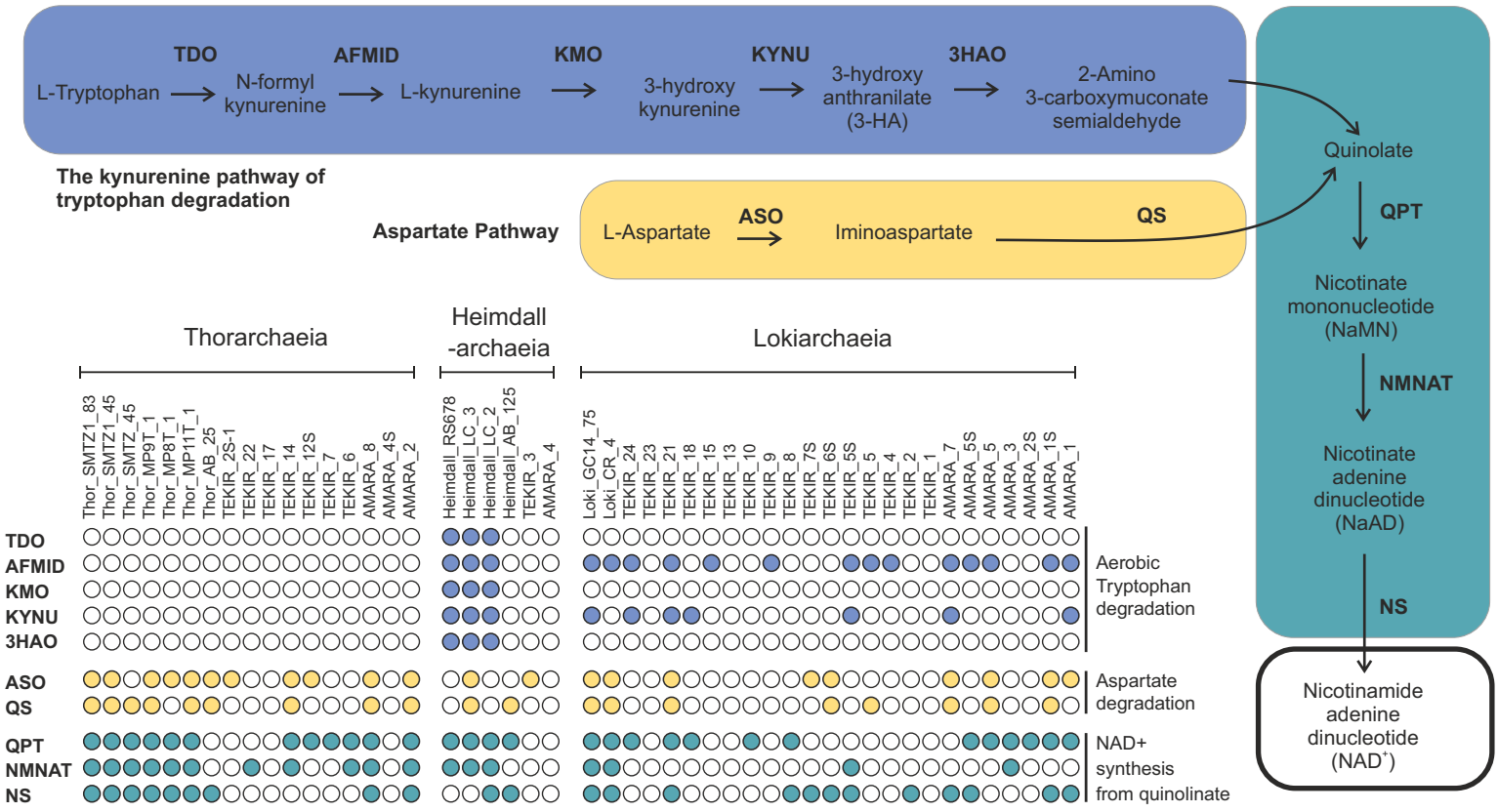


Figure 4. *De novo* NAD$^+$ synthesis pathways. The colored boxes show a schematic representation of the kynurenine and aspartate pathways involved in *de novo* NAD$^+$synthesis. The presence of the enzymes involved in these pathways is indicated for each MAG by using a colored circle. TDO - tryptophan 2,3-dioxygenase; AFMID – arylformamidase; KMO - kynurenine 3-monooxygenase; KYNU – kynureninase; 3HAO - 3-hydroxyanthranilate 3,4-dioxygenase; ASO - L-aspartate oxidase; QS - quinolinate synthase; QPT - nicotinate-nucleotide pyrophosphorylase; NMNAT - nicotinamide-nucleotide adenylyltransferase; NS - NAD+ synthase

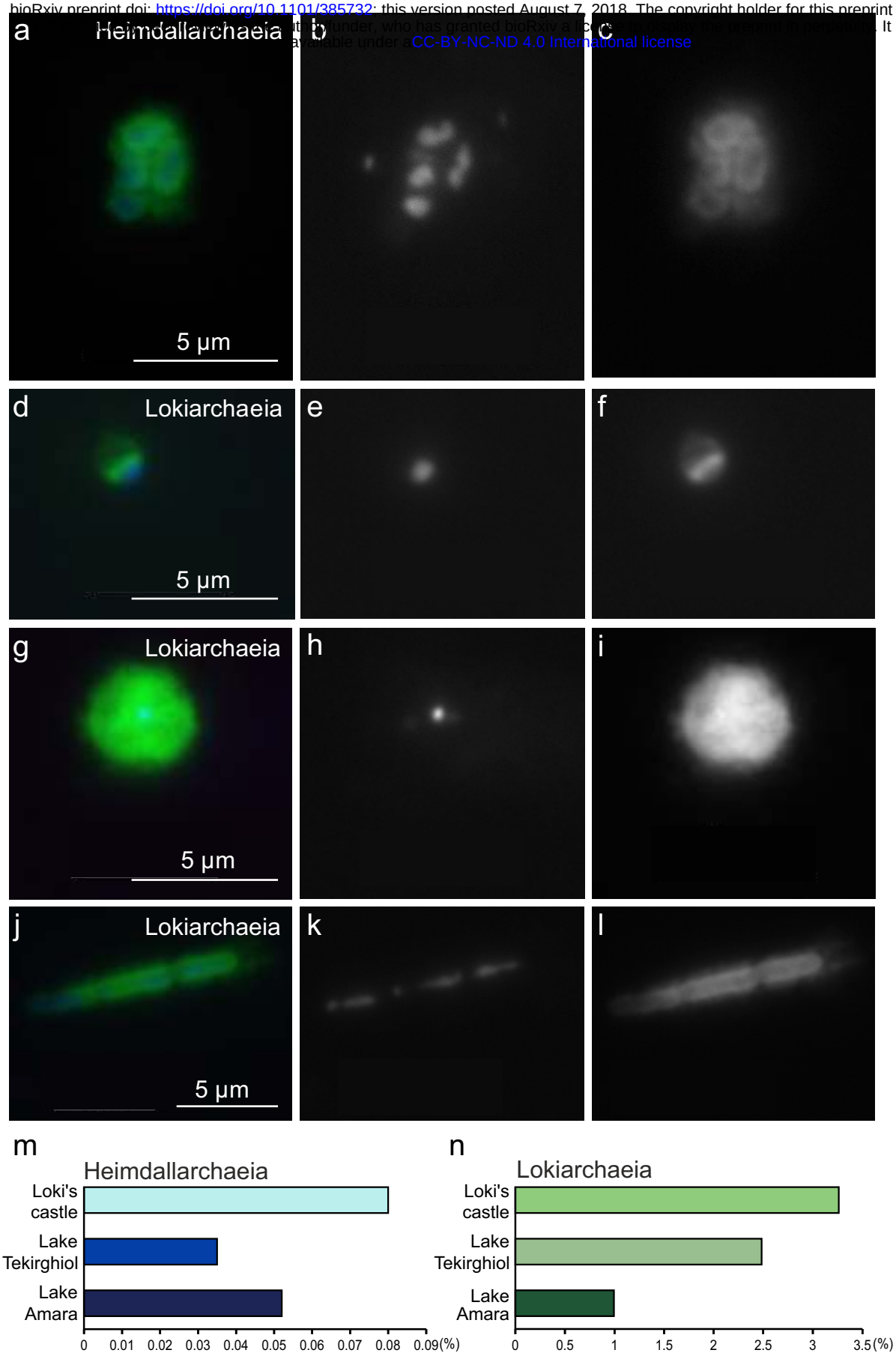


Figure 5. CARD-FISH imaging of Heimdallarchaeia and Lokiarchaeia. The upper panel (a-c) shows several cells of Heimdallarchaeia hybridized with probe heimdall-526, the lower three panels (d-l) diverse morphologies of Lokiarchaeia targeted by probe loki1-1184. The left panels (a, d, g, j) display overlay images of probe signal (green), DAPI staining (blue) and autofluorescence (red), the middle panels (b, e, h, k) DAPI staining of DNA, the right panels (c, f, i, l) CARD-FISH staining of proteins. Abundance estimation of recovered 16S rRNA reads affiliated with Heimdallarchaeia (m) and Lokiarchaeia (n) within metagenomes from Loki's castle, Lake Tekirghiol, and Lake Amara.

Supplementary Information

The sunlit microoxic niche of the archaeal eukaryotic ancestor comes to light

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

[1]Department of Molecular Biology and Biotechnology, Faculty of Biology and Geology, Babeş-Bolyai University, Cluj-Napoca, Romania.
[2]Institute of Hydrobiology, Department of Aquatic Microbial Ecology, 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, Seestrasse 187, CH-8802 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.

# 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 Methods and Discussion

Sampling sites: Amara and Tekirghiol are naturally-formed shallow lakes, located in the South-Eastern Romania, that harbour large deposits of organic-rich sediments (or 'sapropels'). Amara Lake (44°36'N, 27°20'E; 32 m a.s.l.; 1.3 km^1 area; maximum and average depths of 6 m and 2 m respectively) is an oxbow lake with brackish water, originating from an early meander of the Ialomita river (Romanian Plain) supposedly at the end of the Neolithic Black Sea transgression (ca. 3000 BC)[2]. Tekirghiol Lake (44°03'N, 28°36'E; 0.8 m a.s.l.; 11.6 km^2 area; maximum and average depths of 9 m and 3 m respectively) is a saline coastal lake derived from a marine lagoon which was isolated from the Black Sea by a narrow (~200 m wide) sand barrier, most probably during Phanagorian Black Sea regression (ca. 500-700 BC)[1].

Sediment chemical analyses: The leachable major ions were water-extracted using a sediment-to-(milli-Q)water ratio of 1:10 at room temperature. The suspension was centrifuged and the supernatant was filtered through 0.22 µm-pore sized membranes. The obtained filtrate was further analyzed for ion content (**Supplementary Table S9**). Cations (Na^+, K^+, Mg^{2+}) were measured by inductively coupled plasma atomic emission spectrometry (ICP-AES) using Optima 5300DV spectrometer (Perkin Elmer, USA). Chloride (Cl^-) ions were measured by titrimetric method. Sulfate (SO_4^{2-}) was assessed by ion chromatography on ICS-1500 (Dionex, Sunnyvale, CA, USA). The analysis of salt contents of Tekirghiol and Amara sediments indicated that dominant cations and anions were ($g \cdot Kg^{-1}$): 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). All chemical analyses were performed by E.A. Levei and M. Șenilă at INCDO-INOE 2000 - Research Institute for Analytical Instrumentation (Cluj-Napoca, Romania).

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)[3], Baltic Sea AB_25[4] and the White Oak River Estuary SMTZ1_45[4]; forming a genus-level clade (as assessed by amino acid identity values). 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)[5] (**Supplementary Figure S5**) in Thorarchaeia supports the '(more)recent diversification', as the other Asgardaeota (sampled to date) still maintain the ancestral type III (**Supplementary Figure S5**).

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 Asgardaeota superphylum at large: type-1, HeR and schizorhodopsins. In contrast Loki- and Thorarchaeia were found to contain only schizorhodopsins (SzR). Lokiarchaeia was 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 S6**) 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 Asgardaeota in habitats with light exposure possibility (e.g. lake water column, estuarine sediments, mangrove sediments,

microbial mats, etc.)[3,4] fulfils the condition needed for rhodopsin usage, further experimental data is needed to clarify the functions of these newly described proteins.

Metabolism: We observed that while all Asgardaeota 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 Asgardaeota) 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. The newly available 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-[3,6], 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 Asgardaeota, 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 Asgardaeota 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[7].

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 Asgardaeota 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[8].

While a canonical pentose phosphate pathway (PPP) is lacking in all analyzed Asgardaeota, 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*[9]. 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 Asgardaeota 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[3], 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[10].

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[11] - 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[12].

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[13].

Through analyzing new Asgardaeota genomes, we confirm previous findings[3,14] 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[15] 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[16,17]. In spite of the fact that in anaerobic as well as aerobic microbes, CO may be used as both energy and carbon source[18], the ability to utilize CO is conditioned by the presence of enzyme complexes known as carbon monoxide dehydrogenases (CODHs)[19]. 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[20–22]. However we observe that Thor- and Lokiarchaeia encode all components of the bifunctional and oxygen-sensitive[23] 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[24]) 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 through lateral gene transfer from bacteria (**Supplementary Figure S3**). The phylogenetic trees, constructed with key enzymes of the pathway, pointed towards enzyme-specific evolutionary rates (**Supplementary Figure S3**). 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).

In Heimdallarchaeia the following archaellum[24] 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 References

1. Gastescu, P. & Teodorescu, D. C. The lakes of the Romanian Black Sea coast. Maninduced changes, water regime, present state. *Rev. Roum. Gėogr./Rom. Journ. Geogr.* **60,** 27–42 (2016).

2. Fedorov, P. V. Postglacial transgression of the Black Sea. *Int. Geol. Rev.* **14,** 160–164 (1972).

3. Liu, Y. *et al.* Comparative genomic inference suggests mixotrophic lifestyle for Thorarchaeota. *ISME J.* **12,** 1021–1031 (2018).

4. Zaremba-Niedzwiedzka, K. *et al.* Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* **541,** 353–358 (2017).

5. 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).

6. 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).

7. 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).

8. 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).

9. 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).

10. 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).

11. 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).

12. Tittmann, K., Golbik, R., Ghisla, S. & Hübner, G. Mechanism of Elementary Catalytic Steps of Pyruvate Oxidase from Lactobacillus plantarum. *Biochemistry* **39,** 10747–10754 (2000).

13. 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).

14. Sousa, F. L., Neukirchen, S., Allen, J. F., Lane, N. & Martin, W. F. Lokiarchaeon is hydrogen dependent. *Nat. Microbiol.* **1,** 16034 (2016).

15. 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).

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

17. 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).

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

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

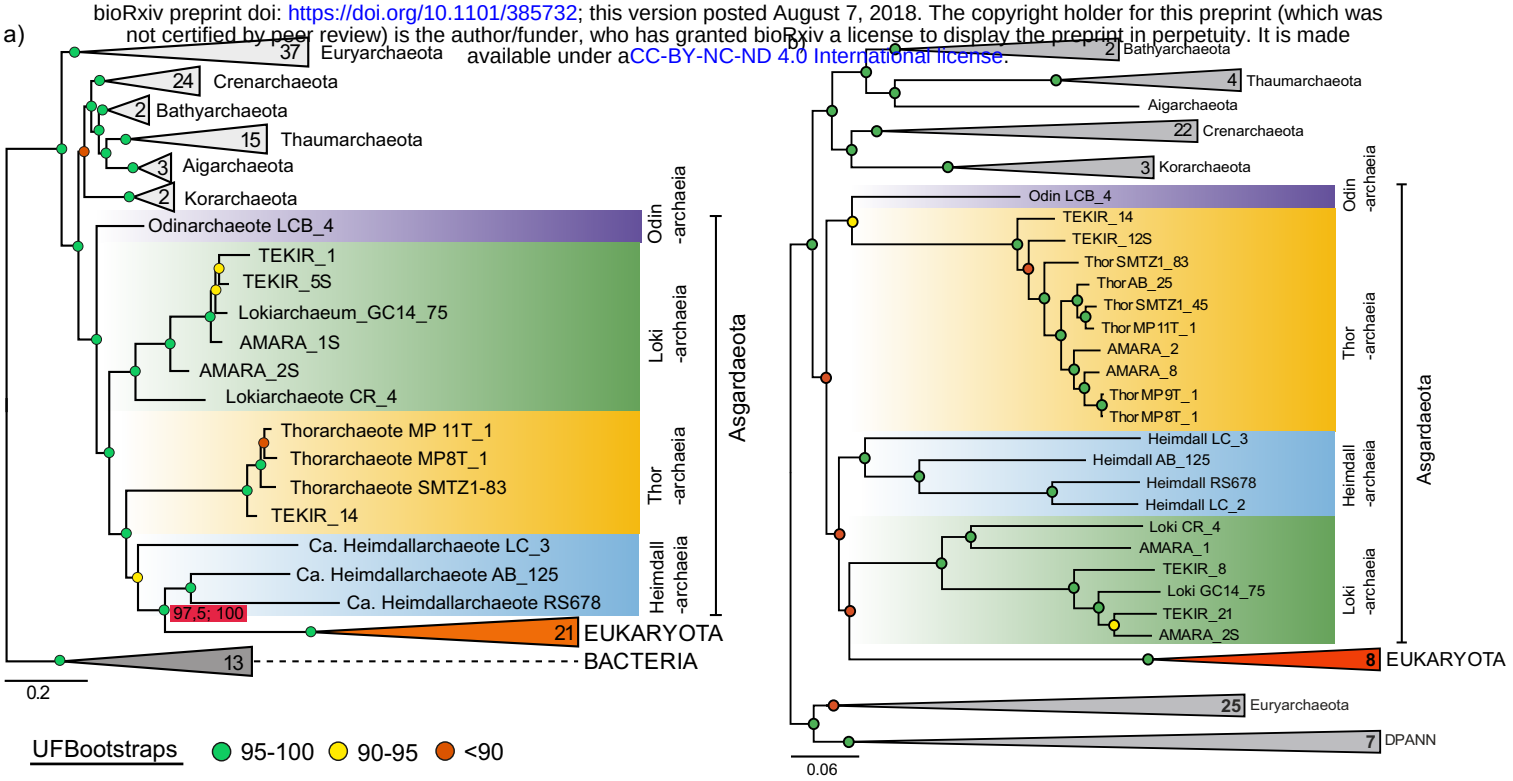
20. King, G. M. Uptake of carbon monoxide and hydrogen at environmentally relevant concentrations by mycobacteria. *Appl. Environ. Microbiol.* **69,** 7266–7272 (2003).

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

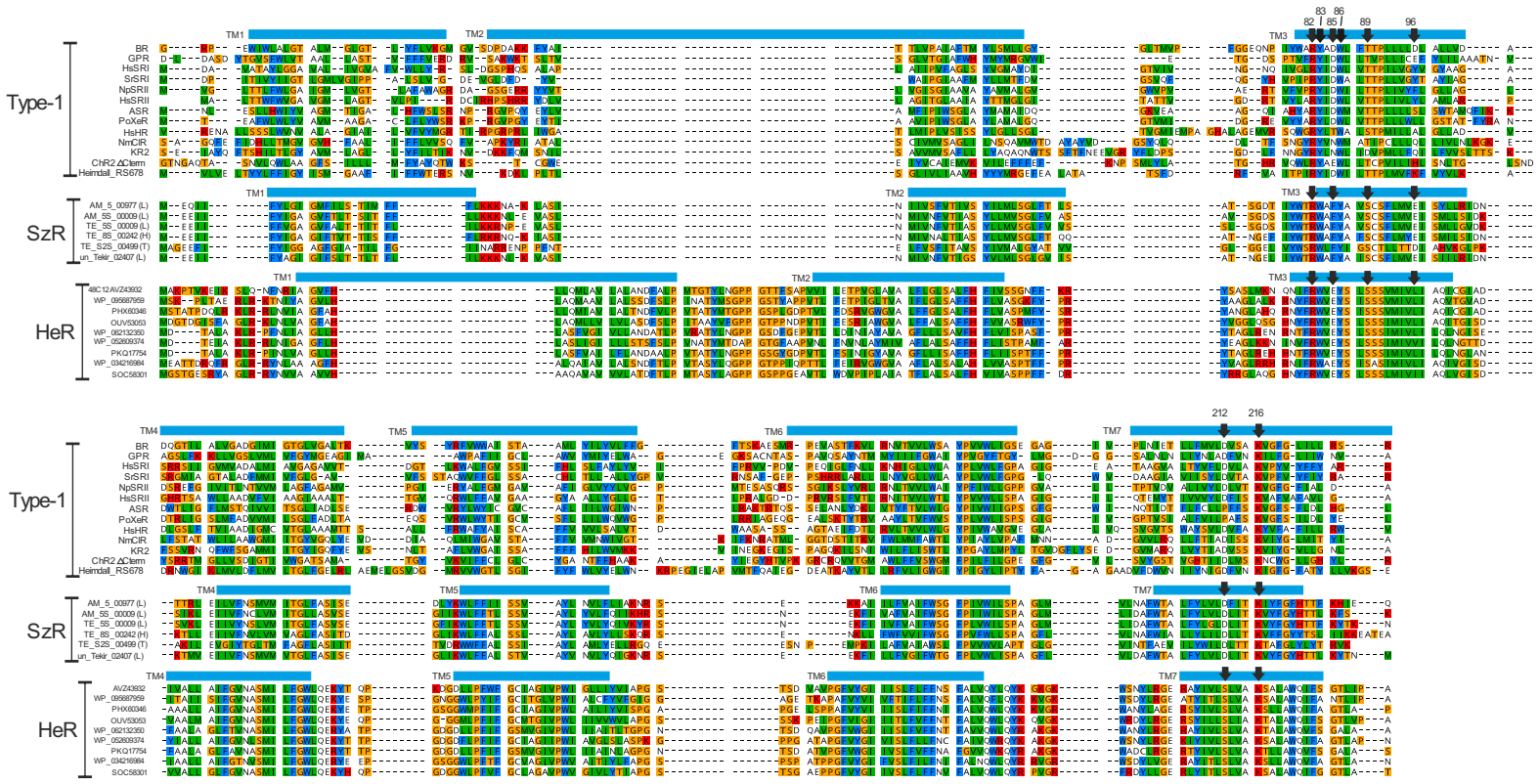
22. King, G. M. & Weber, C. F. Distribution, diversity and ecology of aerobic CO-oxidizing bacteria. *Nat. Rev. Microbiol.* **5,** 107–118 (2007).

23. 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.* (2018).

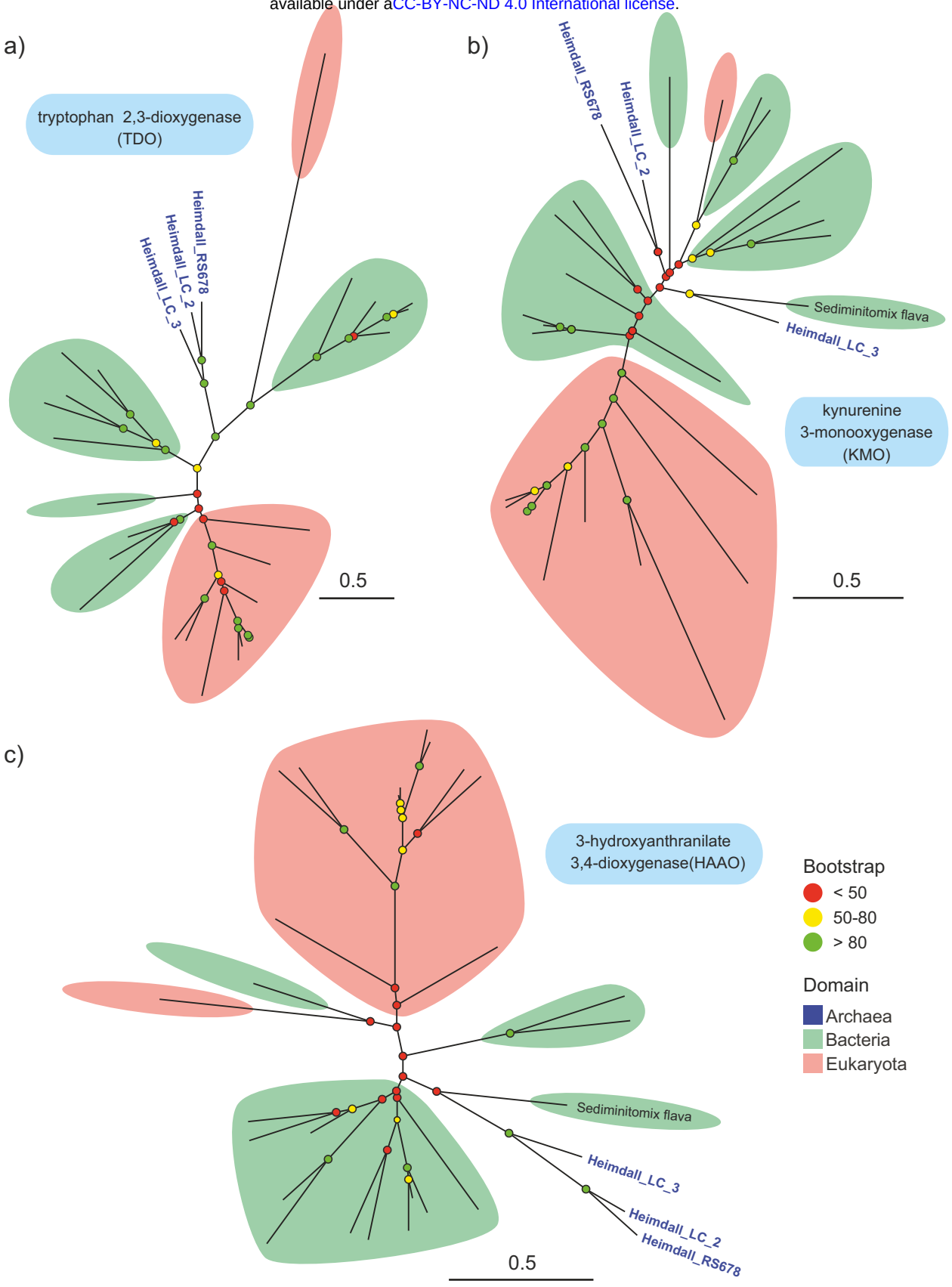
24. Ternes, C. M. & Schönknecht, G. Gene Transfers Shaped the Evolution of De Novo NAD(+) Biosynthesis in Eukaryotes. *Genome Biol. Evol.* **6,** 2335–2349 (2014).



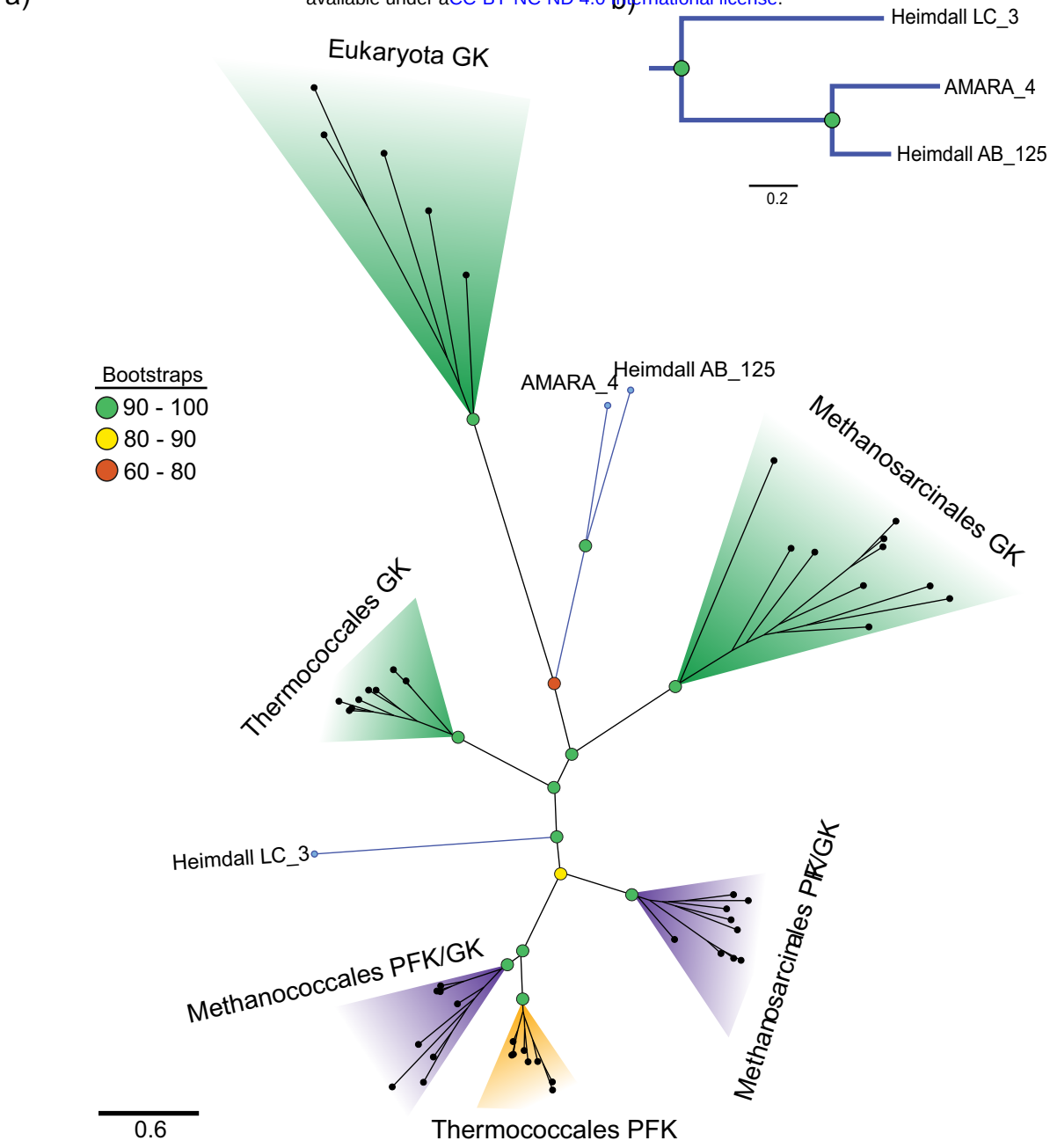
Supplementary Figure S1. a) Maximum-likelihood phylogeny of concatenated SSU/LSU gene sequences spanning the three domains of life. The red rectangle indicates the results of the Shimodaira-Hasegawa test and ultrafast bootstrapping for the Eukaryota/Heimdallarchaeia cluster. b) Maximum-likelihood phylogenomic tree based on 48 ribosomal proteins. Bootstrap 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 S2. 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 - BR: 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-pump rhodopsin from *Nonlabens marinus* (NmClR), sodium- pump rhodopsin from *Krokinobacter eikastus* (KR2), and cation channelrhodopsin2 from *Chlamydomonas reinhardtii* (C-terminal side omitted, ChR2 ΔC-term) and putative proton-pump from Heimdallarchaeota MAG RS678 assembled from Red Sea metagenome. Sequences of Asgardaeota 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.



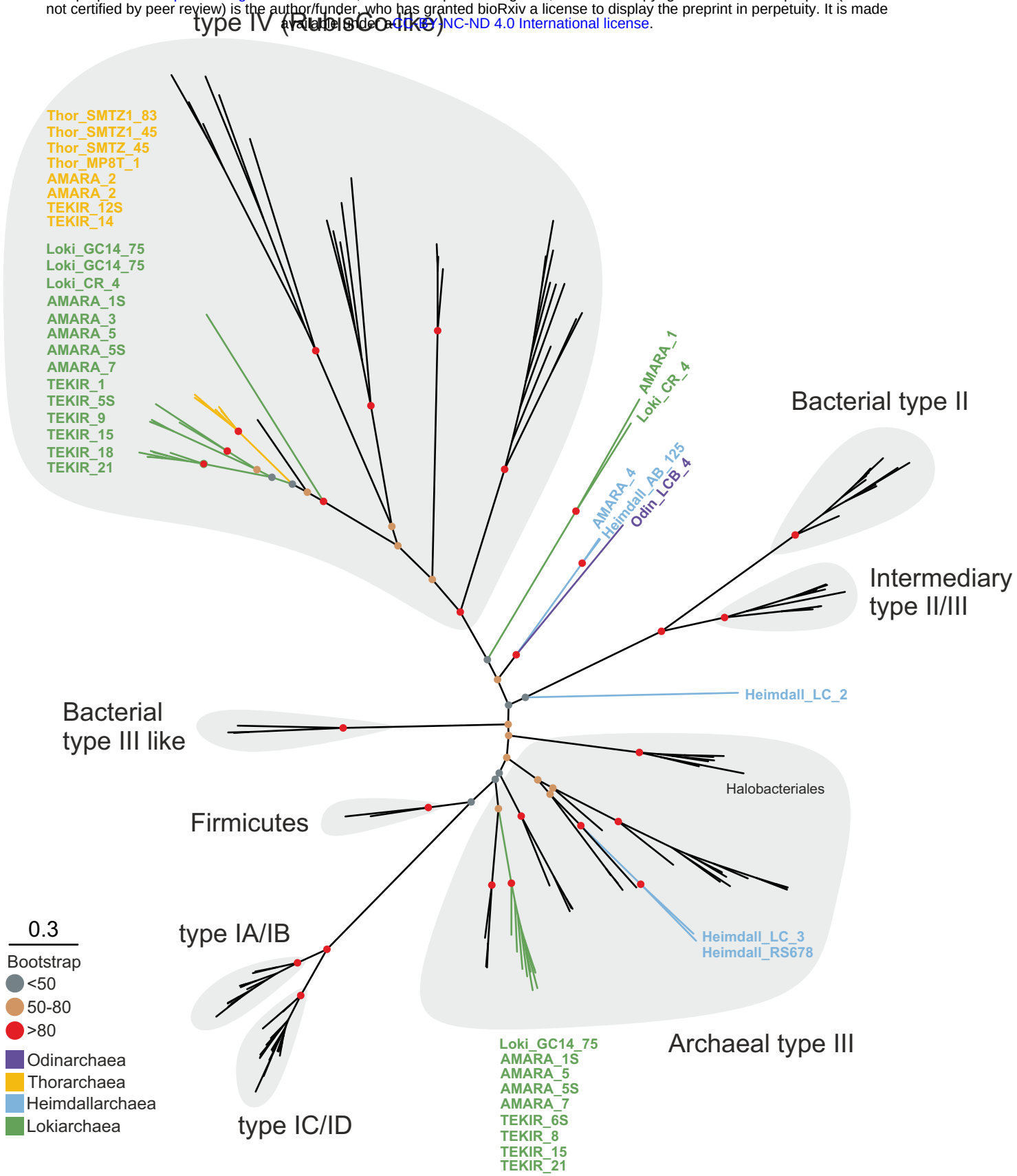
Supplementary Figure S3. Maximum-likelihood phylogenetic trees of key enzymes of the kynurenine pathway: a) tryptophan 2,3-dioxygenase, b) kynurenine 3-monooxygenase and 3-hydroxyanthranilate 3,4-dioxygenase, respectively. Bootstrap values are indicated by colored circles (legend in the lower-right part of the figure). The names of Heimdallarchaeia MAGs are highlighted with blue color.



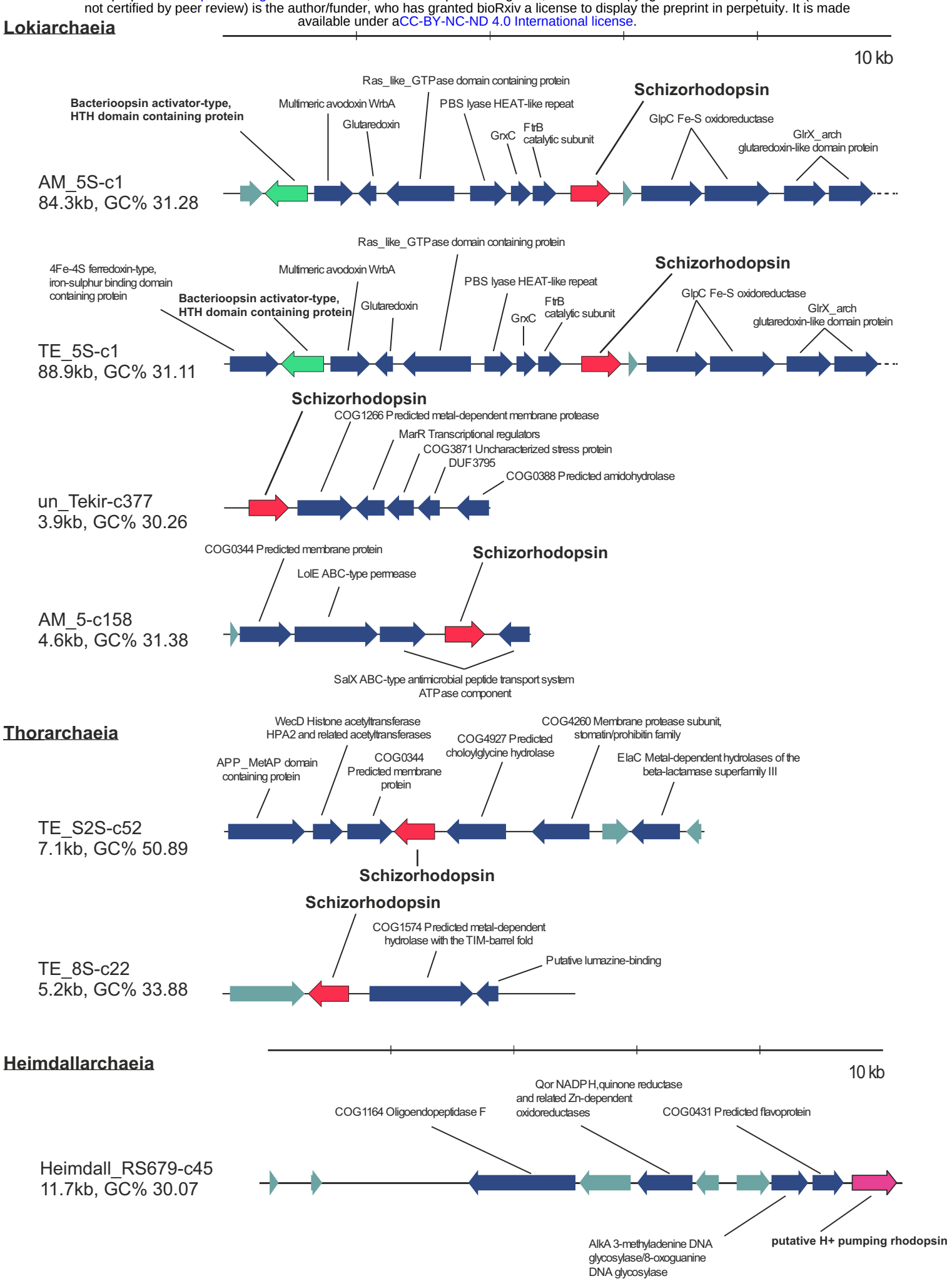
c)

		172										
	Heimdall LC_3	Ser	Ile	Ser	Ser	Gly	Lys	Ala	Lys	Gln	Ile
GK	Thermococcus sibiricus	Ser	Ile	Lys	Leu	Gly	Lys	Ala	Ile	Glu	Trp
GK	Pyrococcus abyssi	Ser	Val	Lys	Arg	Gly	Lys	Ala	Ala	Glu	Leu
GK	Thermococcus gammatolerans	Ser	Ile	Arg	Arg	Gly	Lys	Ala	Ala	Glu	Leu
PFK	Thermococcus onnurineus	Ala	Leu	Lys	Thr	Gly	Lys	Pro	Met	Ala	Val
PFK	Thermococcus barophilus	Ala	Leu	Lys	Thr	Gly	Lys	Pro	Met	Ala	Val
PFK	Thermococcus zilligii	Ala	Leu	Lys	Thr	Gly	Lys	Pro	Met	Ala	Val
PFK/GK	Methanocaldococcus jannaschii	Ser	Ile	Lys	Thr	Gly	Lys	Pro	Ala	Glu	Val
PFK/GK	Methanocaldococcus vulcanius	Ser	Ile	Lys	Thr	Gly	Lys	Pro	Ala	Glu	Val
PFK/GK	Methanococcus vannielii	Ala	Met	Arg	Ser	Gly	Lys	Pro	Val	Glu	Val
PFK/GK	Methanohalobium evestigatum	Ala	Met	Arg	Asp	Gly	Lys	Ala	Ala	Glu	Val
PFK/GK	Methanosarcina acetivorans	Ser	Met	Arg	Glu	Gly	Lys	Ala	Ala	Glu	Val
PFK/GK	Methanolobus psychrophilus	Ser	Met	Arg	Asp	Gly	Lys	Ala	Ala	Glu	Val
PFK/GK	Methanotorris igneus	Ser	Met	Lys	Thr	Gly	Lys	Pro	Ala	Glu	Val

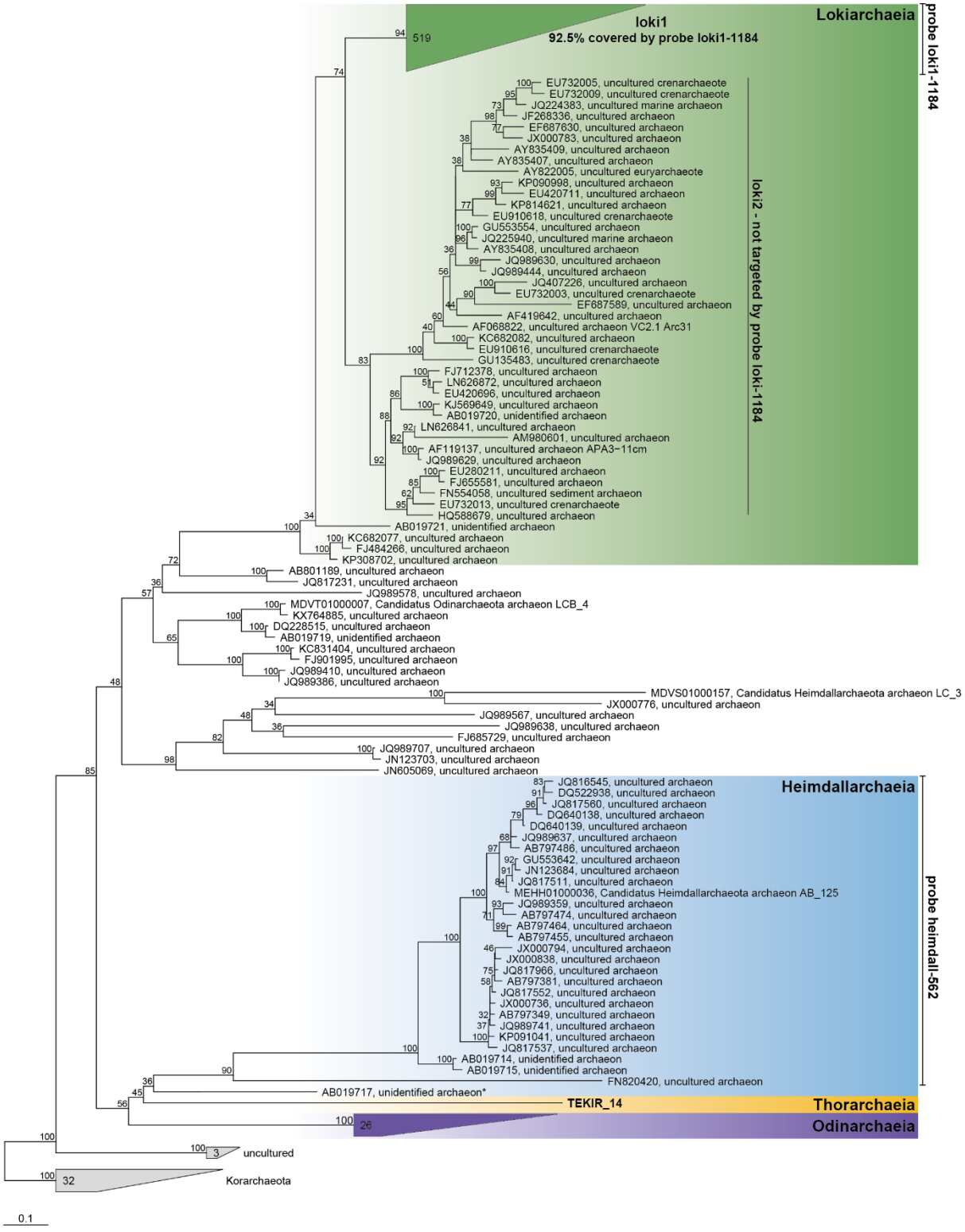
Supplementary Figure S4. a) Maximum-likelihood phylogenetic tree of the the ADP-dependent kinases family in Archaea and Eukaryota. The branches belonging to Heimdallarchaeia are colored in blue. The colored panels highlight clades with: glucokinase activity-green, phosphofructokinase activity -orange, and both gluco- and phosphofructokinase activity-purple. b) Phylogenomic subtree, generated using maximum-likelihood methods, which shows Heimdall_LC_3 MAG as the oldest within the ones that were found to harbor ADP-dependent kinases. The scale bars indicate number of substitutions per site. Bootstrap values are indicated by colored circles (legend in the lower-right part of the figure). c) Sequence alignment of ADP-dependent kinases highlighting the functionally important position 172, as inferred from Castro-Fernandez et al. 2017. The similarity-coloring scheme is based on the BLOSUM62 matrix.



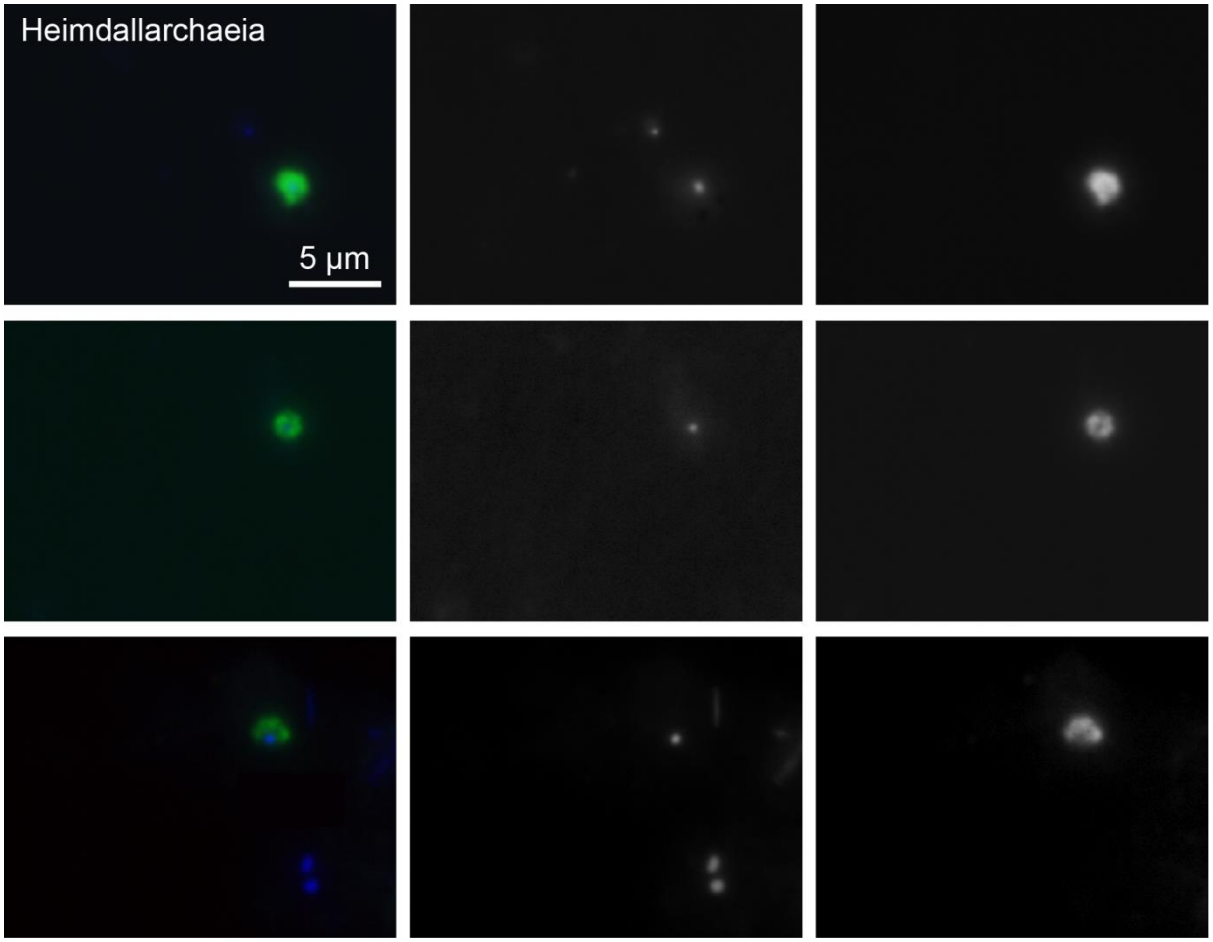
Supplementary Figure S5- Maximum likelihood tree of the large subunit of RubisCo (types I-III) and RubisCo-like (type IV) (rbcL, K01601) protein sequences (n=146) of bacterial and archaeal taxa. Reference sequences were chosen based on previous trees constructed by Tabita et al. 2007 and Wrighton et al. 2016. The branches representing the members of Asgard superphylum are colored based on their affiliated phylum (legends on the bottom left). The scale bar indicates the amino acid substitutions
per site.



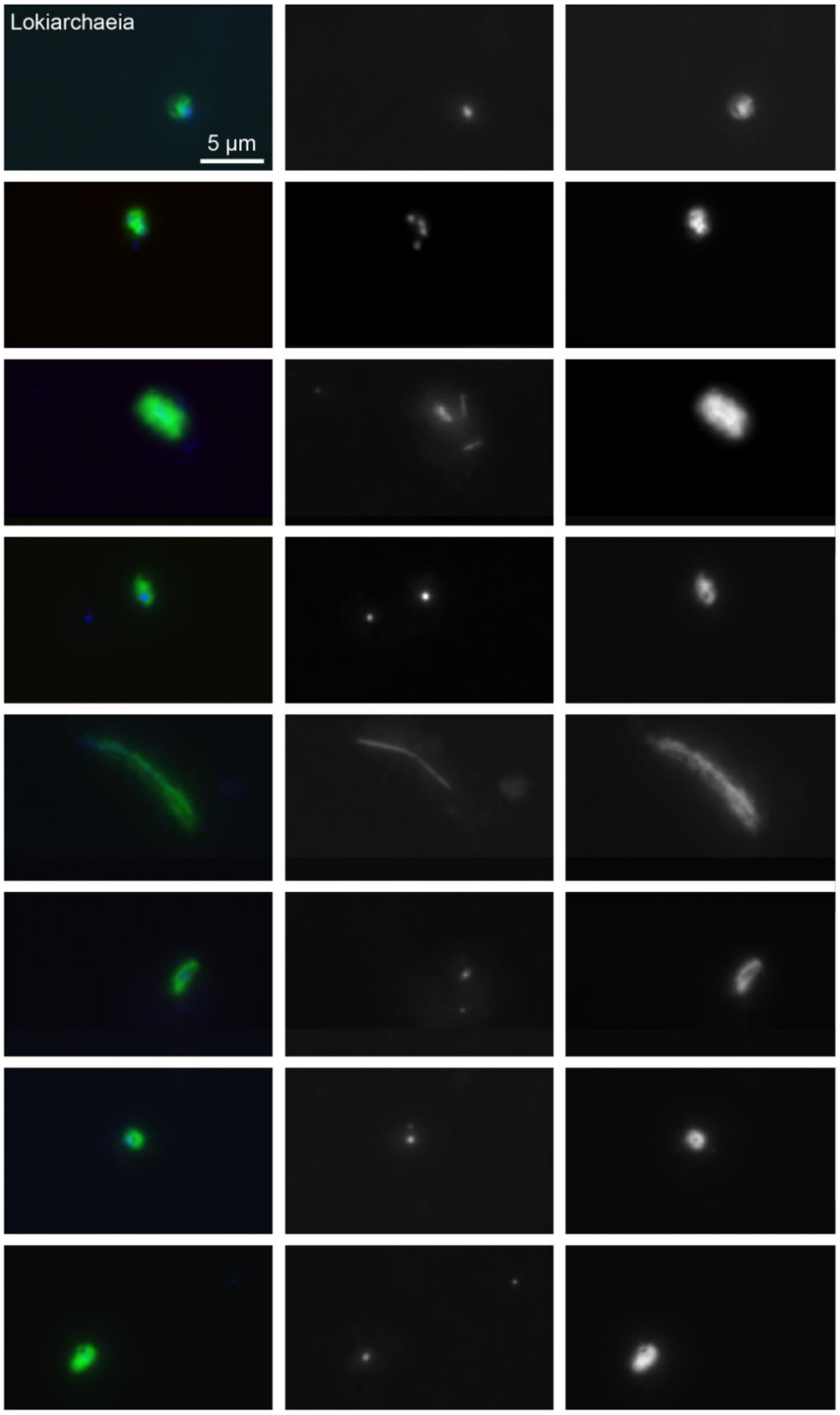
Supplementary Figure S6. Genomic context of rhodopsins in Asgardaeota MAGs. Schizorhodopsins are shown in red, bacteriorhodopsin activator-like protein in green, bacteriorhodopsin in purple and the hypothetical proteins in grey. 10 kb scale bars are shown at the top of each category.



Supplementary Figure S7. RAxML tree (Randomized Axelerated Maximum Likelihood tree with GTR-GAMMA model, 100 bootstraps) of 16S rRNA genes of Asgardaeota with target hits of probes loki1-1184 and heimdall-562. Branches with bootstrap supports <30% were collated to multifurcations. The number of sequences is given for collapsed branches; an asterisk indicates the sequence not targeted by probe heimdall-562.



Supplementary Figure S8. CARD-FISH imaging of Heimdallarchaeia hybridized with probe heimdall-526. The left panels display overlay images of probe signal (green), DAPI staining (blue) and autofluorescence (red), the middle panels DAPI staining of DNA, the right panels CARD-FISH staining of proteins. Individual microphotographs of autofluorescent objects are not displayed because of low intensities and no interference with probe signals. The scale bar (5 µm) top left applies to all microphotographs.



Supplementary Figure S9. CARD-FISH imaging of Lokiarchaeia hybridized with probe loki1-1184. The left panels display overlay images of probe signal (green), DAPI staining (blue) and autofluorescence (red), the middle panels DAPI staining of DNA, the right panels CARD-FISH staining of proteins. Individual microphotographs of autofluorescent objects are not displayed because of low intensities and no interference with probe signals. The scale bar (5 µm) top left applies to all microphotographs.

Supplementary_Table_S1.xlsx

This file contains general statistics for MAGs recovered from the Amara and Tekirghiol Lakes. Estimated genome length (EGS) was determined for MAGs with completeness >70%.

Supplementary Table S2. List of Asgardaeota 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_S3.xlsx

KEGG orthology annotation for Asgardaeota MAGs.

Supplementary Table S4. Accession numbers list for INTERPRO (IPR) domains, COGs (Cluster of Orthologous Groups) and UniProtKB protein sequences used to identify potential Eukaryotic Signature Proteins in Asgardaeota 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	COG03569
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_S5.xlsx

List of taxa used for phylogenetic ribosomal RNA (small subunit SSU and large subunit LSU) as well as ribosomal protein-based inferrences based mainly on a previously published list https://www.nature.com/articles/nature21031. Silva SSU and LSU rRNA IDs are indicated. For MAGs recovered in this study, contig ID and positions are indicated (highlighted).

Supplementary Table S6. List of the 48 ribosomal proteins used for phylogenomic analyses. Total number of taxa (from the 93 included) 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 S7. Protein RefSeq accession numbers for 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 S8. Probes designed for CARD-FISH.

Probe name	Targeted lineage	Targeted MAGs	Lineage coverage (all Lokiarchaeia)	Outgroup hits	Probe sequence (5'-3')	% Formamide	Average length (µm) (±SD)	Average width (µm) (±SD)
loki1-1184	Lokiarchaeia lineage loki1	AMARA_1S, GC14_75	92.5% (85.1%)	none	GACCTGCCTTTGCCCGC	55	2.556 ± 1.19	1.410 ± 0.55
heimdall-526	Heimdallarchaeia	Heimdall_AB_125	96.7%	none	CACTCGCAGAGCTGGTTTTACCG	40	2.006 ± 0.43	1.348 ± 0.34
Competitor name					**Probe sequence (5'-3')**			
loki-1184-C1	competitor 1 for loki-1184				GACCTGCCGTTGCCCGC			
loki-1184-C2	competitor 2 for loki-1184				GACATGCCTTTGCCCGC			
heimdall-526-C1	competitor 1 for heimdall-526				CACTCGRAGAGCTGGTTTTACCG			
heimdall-526-C2	competitor 2 for heimdall-526				CACTCGCAGAGCTGGTATTACCG			
heimdall-526-C3	competitor 3 for heimdall-526				CACTCGCGGAGCTGGTTTTACCG			

Supplementary Table S9. Major water leachable ions and elements from sediments collected in Tekirghiol and Amara Lakes during October 2017.

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 * indicates element total concentration.