Casting light on Asgardarchaeota metabolism in a sunlit microoxic niche

Bulzu, Paul-Adrian ; Andrei, Adrian-Ştefan ; Salcher, Michaela M ; Mehrshad, Maliheh ; Inoue, Keiichi ; Kandori, Hideki ; Beja, Oded ; Ghai, Rohit ; Banciu, Horia L

Abstract: Recent advances in phylogenomic analyses and increased genomic sampling of uncultured prokaryotic lineages have brought compelling evidence in support of the emergence of eukaryotes from within the archaeal domain of life (eocyte hypothesis)1,2. The discovery of Asgardarchaeota and its supposed position at the base of the eukaryotic tree of life3,4 provided cues about the long-awaited identity of the eocytic lineage from which the nucleated cells (Eukaryota) emerged. While it is apparent that Asgardarchaeota encode a plethora of eukaryotic-specific proteins (the highest number identified yet in prokaryotes)5, the lack of genomic information and metabolic characterization has precluded inferences about their lifestyles and the metabolic landscape that favoured the emergence of the protoeukaryote ancestor. Here, we use advanced phylogenetic analyses for inferring the deep ancestry of eukaryotes, and genome-scale metabolic reconstructions for shedding light on the metabolic milieu of Asgardarchaeota. In doing so, we: (1) show that Heimdallarchaeia (the closest eocytic lineage to eukaryotes to date) are likely to have a microoxic niche, based on their genomic potential, with aerobic metabolic pathways that are unique among Archaea (that is, the kynurenine pathway); (2) provide evidence of mixotrophy within Asgardarchaeota; and (3) describe a previously unknown family of rhodopsins encoded within the recovered genomes.

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

Casting light on Asgardarchaeota metabolism in a sunlit microoxic niche

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

[1]Department of Molecular Biology and Biotechnology, Faculty of Biology and Geology, Babeş-Bolyai University, Cluj-Napoca, Romania
[2]Department of Aquatic Microbial Ecology, Institute of Hydrobiology, Biology Centre of the Academy of Sciences of the Czech Republic, České Budějovice, Czech Republic.
[3]Limnological Station, Institute of Plant and Microbial Biology, University of Zurich, Kilchberg, Switzerland.
[4]Faculty of Biology, Technion Israel Institute of Technology, Haifa, Israel.
[5]The Institute for Solid State Physics, The University of Tokyo, Kashiwa, Japan
[6]Department of Life Science and Applied Chemistry, Nagoya Institute of Technology, Nagoya, Japan.
[7]Molecular Biology Center, Institute for Interdisciplinary Research in Bio-Nano-Sciences, Babeş-Bolyai University, Cluj-Napoca, Romania

# These authors contributed equally to this work

*Corresponding author: Rohit Ghai

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

Phone: +420 387 775 881

Fax: +420 385 310 248

E-mail: ghai.rohit@gmail.com

Recent advances in phylogenomic analyses and increased genomic sampling of uncultured prokaryotic lineages have brought compelling evidence in support of the emergence of eukaryotes from within the archaeal domain of life (eocyte hypothesis)[1,2]. The discovery of Asgardarchaeota and its supposed position at the base of the eukaryotic tree of life[3,4] provided cues about the long-awaited identity of the eocytic lineage from which the nucleated cells (Eukaryota) emerged. While it is apparent that Asgardarchaeota encode a plethora of eukaryotic-specific proteins (the highest number identified yet in prokaryotes)[5], the lack of genomic information and metabolic characterization has precluded inferences about their lifestyles and the metabolic landscape that favored the emergence of the protoeukaryote ancestor. Here, we use advanced phylogenetic analyses for inferring the deep ancestry of eukaryotes, and genome-scale metabolic reconstructions for shedding light on the metabolic milieu of Asgardarchaeota. In doing so, we: i) show that Heimdallarchaeia (the closest eocytic lineage to eukaryotes to date) are likely to have a microoxic niche, based on their genomic potential, with aerobic metabolic pathways unique among Archaea (i.e. kynurenine pathway), ii) provide evidence of mixotrophy within Asgardarchaeota and iii) describe a previously unknown family of rhodopsins encoded within the recovered genomes.

At the dawn of genomics, the eukaryotes were recognized as amalgamated genetic jigsaws that bore components of both bacterial and archaeal descent[6]. This genomic chimerism served as a source of speculation and debate over the nature of the protoeukaryote ancestors[3,4,7] and inspirited a plethora of hypothetical scenarios for the processes that led to eukaryogenesis[6,8]. The hypothesis that eukaryotes emerged from within the archaeal radiation was put forward more than three decades ago, when ribosomal structural patterns were shown to support a sister-group relationship between eukaryotes and an extant archaeal lineage[9]. This already diversified archaeal group, which was found to bear more similarity (in ribosomal morphology) with eukaryotes than to Bacteria and the rest of the Archaea was denominated as Eocyta[9]. Subsequently, the eocyte hypothesis[9] was overshadowed by the three-domains tree of life, which depicts Archaea and Eukarya as monophyletic groups that share a common ancestor[10]. More than two decades passed till the usage of increased taxonomic sampling and advanced phylogenetic approaches fueled the revival of the eocyte hypothesis[1,2], and alongside the debate regarding the topology of the tree of life[4,5,7]. Even though, in light of recent research, eukaryotes came into existence through the interplay between an archaeal host[11] and a bacterial endosymbiont[12], the metabolic milieu of the ancestral prokaryotic lineages still remain elusive.

Homology-based searches were employed to recover Asgardarchaeota-related contigs from *de novo* metagenomic assemblies of two deep-sequenced, shallow, brackish lake sediment samples (sediment pore-water salinities: 5.7% in Tekirghiol Lake and 3.9% in Amara Lake). By utilizing a hybrid binning strategy and performing manual inspection and data curation, we obtained eleven high- and medium-quality (> 50% completeness; < 2% contamination) and twenty-four low completeness (<50% completeness; < 3% contamination) MAGs (metagenome-assembled genomes), spanning three (out of four) evolutionary lineages within the superphylum: Lokiarchaeia (23), Thorarchaeia (10), and Heimdallarchaeia (2). The maximum likelihood phylogenetic tree, based on concatenation of small (SSU) and large (LSU) ribosomal RNA genes, pictured a topology in which eukaryotes branched with high-support from within Asgardarchaeota (Archaea) (**Supplementary Figure 1a**). Even more remarkably, in addition to recreating a previously described Asgardarchaeota/Eukaryota branching pattern[5], we provide support for a close evolutionary linkage between Heimdallarchaeia (Asgardarchaeota superphylum) and eukaryotes (SH-aLRT=97.5; UFBoot=100) (**Supplementary Figure 1a**). The genome-focused phylogeny of Asgardarchaeota revealed a pattern of ancestry, divergence, and descent, in which Heimdallarchaeia comprise the basal branch of the superphylum and Thor-/Odinarchaeia the youngest one (**Figure 1a**). Although dissimilar in branching pattern with the SSU + LSU tree (Figure S1a), the phylogenomic one was found to be robust (**Figure 1a**) and to support a topology brought into attention by an earlier study[5]. The SR4-recoded[13] Bayesian tree (maxdiff=0.1) resolved with high support (PP=1) the monophyly of

Asgardarchaeota/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 1b**) the eukaryotes caused branch instability for Heimdallarchaeia, which were attracted without statistical support within the superphylum (PP=0.52; SH-aLRT=92.4 and UFBoot=88). To further substantiate the phylogenetic connections between Asgardarchaeota members and eukaryotes, we screened all recovered MAGs and the publicly available ones (14 in July 2018) for the presence of potential eukaryotic signature proteins (ESP). Similar to previous reports[3,5,14], the MAGs were found to be highly enriched in ESP (**Figure 1c**), which further reinforced their ancestral linkage to eukaryotes. In addition to the reported ESP[5], 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 (known to be involved in membrane trafficking in humans[15]) (**Figure 1c)** in Lokiarchaeia. Furthermore, we retrieved conclusive hits for the ESP-related domains Ezrin/radixin/moesin C-terminal domain and active zone protein ELKS in Lokiarchaeia.

Recent findings reporting the presence of a previously uncharacterized family of rhodopsins[16] (i.e. heliorhodopsins; abbreviated as HeR) in monoderms[17] encouraged us to perform a dedicated screening in all available Asgardarchaeota MAGs. The results indicated that one of the Heimdallarchaeia MAGs (i.e Heimdallarchaeota RS678) encoded two HeR and what appears to be, as suggested by the presence of a *Exiguobacterium*-like DTK motif[18] and phylogenetic proximity, a type-1 proton-pumping rhodopsin (**Figure 2**; **Supplementary Table 4**), suggestive of light sensitivity. Remarkably, we found that the Asgardarchaeota 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 (Standard Bootstrap: SBS=1) placed in-between HeR and type-1 rhodopsins (**Figure 2**). Multiple sequence alignments showed: i) homology between transmembrane helices 1, 6 and 7 of these 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 ones that we identified in Asgardarchaeota) (**Supplementary Figure 2**). 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[16,17] precludes tentative functional assertions for SzR capacity in Asgardarchaeota. 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 their evolutionary history, Asgardarchaeota were 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)[19].

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 ancestor. While the anaerobic lifestyles inferred for Loki-[20] and Thorarchaeia[14] were considered to be accompanied by autotrophy[20] and mixotrophy[14], respectively, no consistent metabolic reconstructions exist for Heimdallarchaeia. The physiology inferred here pointed towards mixotrophic lifestyles (for Asgardarchaeota), simultaneously showing the presence of transporters for the uptake of exogenous organic matter and the metabolic circuitry responsible for its catabolism (see Supplementary Discussion). Noteworthy, 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 to Heimdallarchaeia, which utilize TCA to fuel their catabolic machinery (**Figure 3**), Lokiarchaeia use 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[21] (e.g. linking TCA and ETC), we investigated its *de novo* synthesis mechanisms (**Figure 4**). As expected, all Asgardarchaeota phyla harbored the aspartate pathway - a set of metabolic transformations that can occur both in presence or absence of oxygen[22], and which are characteristic for most prokaryotes and plastid-bearing eukaryotes (obtained through lateral gene transfer from their cyanobacterial endosymbionts)[21]. Surprisingly, in addition to the aspartate pathway, Heimdallarchaeia also encoded the exclusively aerobic kynurenine pathway of NAD^+ biosynthesis[23] (**Figure 4**), which is present in few bacterial groups and eukaryotes[21]. The phylogenetic reconstruction and evolutionary history inferences showed that this pathway, which is considered to be present in the protoeukaryote ancestor[21], was probably acquired by the ancestor of Heimdallarchaeia through lateral gene transfer from bacteria (**Supplementary Figure 3**). As far as the authors are aware, Heimdallarchaeia are the first reported archaeal organisms harboring 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 exclusively the genes affiliated with the kynurenine pathway (**Figure 4**). As the aspartate pathway was reported to function in both oxygen absence (L-aspartate oxidase uses fumarate as electron acceptor)[22] 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 Asgardarchaeota (**Figure 4**) with the reconstructed evolutionary history of Heimdallarchaeia (**Figure 1a**, **b**; **Supplementary Figure 1a, 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 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 aerobic conditions, 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[24]. 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**).

Comparative genomic analyses also revealed that the three Asgardarchaeota phyla rely upon glycolysis (i.e. of Embden-Meyerhof-Parnas type) 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 phosphoryl group donor[25] in their glycolytic pathways. Furthermore, they seemed to be bifunctional ADP-dependent glucokinase/phosphofructokinases, 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 4**). Additionally, we observed that the deepest branching Heimdallarchaeia (LC_3) harbored the archaeal-type enzyme, while the younger ones (AB_125 and AMARA_4) clustered together with the eukaryotic-type (**Supplementary Figure 4**). 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[26], the metabolic advantage conferred by these ADP-dependent kinases is unclear.

Although pentoses could be recycled *via* nucleotide degradation in all Asgardarchaeota 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 pathway. The identified homologues for ribulose 1,5-bisphosphate carboxylase/oxygenase (RuBisCO) genes were found to appertain to types III (Loki- and Heimdallarchaeia) and IV (Loki- and Thorarchaeia) (**Supplementary Figure 5**). 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) utilize the nucleotide monophosphate degradation pathway[27], 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[28], 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 use 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[29] (through CODH/ACS).

The mainstream theories on the subject of eukaryogenesis[8] which date back to late 20th century have been recently challenged by improved phylogenetic methods[2] and increased genomic sampling[5,20]. Even after experiencing a revival[20], the current endosymbiotic theory fails to envision the environmental and metabolic context from which the protoeukaryote ancestor emerged. In order to study the phylogenetic relationships and the metabolic milieu of Asgardarchaeota we leveraged metagenome-derived data. Despite the fact that the performed phylogenomic analyses recovered an eocyte tree topology (in agreement with previous studies[4,5]), they did not reach consensus regarding the branching point of eukaryotes from within the phylum. Thus, while the SSU+LSU tree showed the Eukaryota branching from within Heimdallarchaeia, the ribosomal protein one pointed towards a sister-group relationship. As the available genomic data represents a fraction of the extant Asgardarchaeota diversity, we consider that additional sampling will further improve the resolution of the phylogenomic tree and clarify the phylogenetic relationship between Eukaryota and Heimdallarchaeia. Overall, the performed metabolic reconstructions indicate that Heimdallarchaeia are mixotrophic and have a facultative aerobic metabolism. The presence of oxygen-dependent

pathways in Heimdallarchaeia raises the possibility that the archaeal protoeukaryote ancestor could have also been a facultative aerobe. Thus, based upon phylogenetic and metabolic reconstructions, we propose a hypothesis (i.e. 'aerobic-protoeukaryotes' model) in which both the archaeal and bacterial eukaryotic ancestors have an oxygen-dependent metabolism. This 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 to the existence of the archaeal host (caused by oxygen exposure). This model is in agreement with a recent high-scale time-calibrated phylogenomic tree, which shows that the archaeal-bacterial endosymbiosis that gave birth to the protoeukaryote ancestor took place after the Great Oxidation Event[30]. Even though the present data offers an updated perspective on the lifestyles of Asgardarchaeota, it is based on a fraction of the extant environmental diversity. Thus, further studies will be needed in order to elucidate the lifestyle strategies and evolutionary histories within the Asgardarchaeota. We consider that further environmental exploration will undoubtedly offer access to additional eocytic lineages and improve our apprehension of the evolution of Archaea and Eukaryota.

Methods

Sampling: Amara and Tekirghiol are naturally-formed shallow lakes in South-Eastern Romania that harbor large deposits of organic-rich sediments (or 'sapropels'). Amara Lake (44°36.30650 N, 27°19.52950 E; 32 m a.s.l.; 1.3 km[31] area; maximum and average depths of 6 m and 2 m respectively) is an oxbow lake with brackish water (salinity ca. 1%), originating from an early meander of the Ialomița river (Romanian Plain) supposedly at the end of the Neolithic Black Sea transgression (ca. 3000 BC)[32]. Tekirghiol Lake (44°03.19017 N, 28°36.19083 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 (salinity ca. 6%) derived from a marine lagoon which was isolated from the Black Sea by a narrow (~200 m wide) sand barrier, most probably during the Phanagorian Black Sea regression (ca. 500-700 BC)[31]. Sediment sampling was performed in the same manner in two campaigns, in both lakes. In the first sampling campaign, sediment for exploratory chemical and metagenomics analyses was collected on 10 October 2017 at 12:00 in Tekirghiol Lake and on 11 October 2017 at 15:00 in Amara Lake. The successful recovery of Asgardarchaeota genomes prompted a second sampling campaign for fine chemical profiling on 22 April 2018 at 12:00 in Tekirghiol Lake, on the site of the previous sampling. In Tekirghiol Lake, sampling was performed in the shallow shore area of the lake (approx. 0.8 m depth) using a custom sediment corer (1 m length, 10 cm diameter, sampling area of 78.5 cm^2) with a sharpened bottom rim and a removable plug. Five sediment layers were collected in 10 cm intervals (0 – 50 cm) and deposited in sterile Falcon tubes. In Amara Lake the dense vegetation and increased water depth (>1.5m) in the shore area hampered the usage of the custom sediment corer. Therefore, the sampling was performed using a Petite Ponar dredge (Wildco, Saginaw, MI, USA) handled from an inflatable boat. The grab penetrated the sediment layer to a depth of about 10 cm (sampling area 225 cm^2). Mixed sediment samples obtained from three casts (within a 2 m radius) were collected and deposited in sterile 1 L containers. Sediment samples were stored in the dark at 4 °C and processed within 24 hours after collection.

Sediment chemical analyses: Chemical analyses were performed on both mixed samples (0 – 40 cm Tekirghiol and 0 – 10 cm Amara) taken in 2017 and the vertical profile obtained from Tekirghiol Lake in 2018. 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 PTFE membranes. The obtained filtrate was further analyzed for ion content (Supplementary Figure 10). The water extractable elements Na, K, Ca, Mg, P, Fe, and Mn, 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. Sulphate (SO_4^{2-}) was assessed by ion chromatography on ICS-1500 (Dionex, Sunnyvale, CA, USA).

Dissolved carbon (DC) and dissolved inorganic carbon (DIC) were measured by catalytic combustion and infrared detection of CO_2 using a Multi N/C 2100S Analyser (Analytik Jena, Germany). Dissolved organic carbon (DOC) was obtained by subtracting DIC from DC. Total dissolved nitrogen (DN) as bound nitrogen (including free ammonia, ammonium, nitrite, nitrate, and organic nitrogen) was analyzed by catalytic combustion followed by oxidation of nitrogen monoxide to nitrogen dioxide with ozone and subsequent chemiluminescence detection. Ammonium ions were analysed by spectrophotometry using Lambda 25 UV-VIS spectrophotometer (Perkin Elmer, Beaconsfield, UK) following formation of colored complexes: indophenol blue complex (ammonium), yellow complex formed with sulphosalicylic acid (nitrate), and red colored azo dye formed from diazonium salt in the presence of N-(1-Naphthyl) ethylenediamine and sulphanilamide under acidic conditions (nitrite). The concentration of sulphides was determined by methylene blue method after fixation of samples with 2% (v/v) Zn-acetate. Humidity was estimated by loss-on-ignition (LOI) method following oven-drying of sediments at 105^{o}C for 24 h. The pH and salinity of pore water were measured with a portable HI 9828 multiparameter (Hanna Instruments, Smithfield, RI, USA). All chemical analyses were performed by E.A. Levei, A.M. Incze, and M. Șenilă at INCDO-INOE 2000 - Research Institute for Analytical Instrumentation (Cluj-Napoca, Romania).

DNA extraction and purification: DNA was extracted from approximately 10 g of wet mixed sediment samples (0 – 40 cm Tekirghiol and 0 – 10 cm Amara) collected in 2017 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, Waltham, MA, USA) and shipped for library construction and NGS sequencing to Macrogen (Seoul, South Korea).

Sequencing and data preprocessing: Library preparation was performed by a 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[33] 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[3] set SRX684858 from Loki's castle marine sediment metagenome, were subsampled to 20 million reads by reformat.sh[34]. Each subset was queried for putative RNA sequences by scanning with UBLAST[35] against the non-redundant SILVA SSURef_NR99_132 database[36], that was priorly clustered at 85% sequence identity by UCLUST[35]. Identified putative 16S rRNA sequences (e-value < 1e-5) were screened using SSU-ALIGN[37]. Resulting bona fide 16S rRNA sequences were compared by blastn[38] (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 **(Supplementary Table 1).**

Environmental distribution of Heimdallarchaeia: In order to investigate the existence of a habitat preference we extracted all the available environmental data from the SILVA database[36] (version 132) that was associated with Heimdallarchaeia sequences.

Metagenome assembly and binning: *De novo* assembly of preprocessed paired-end Illumina reads was done by Megahit[39] 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[40]. Taxonomy dependent binning was achieved by first assigning taxonomy labels to the predicted genes by performing screenings with MMseqs2[41] against the NR database. All contigs with a minimum of 30 % genes assigned to Asgardarchaeota were used for taxonomy-independent binning. Mean base coverage for each contig was computed with bbwrap.sh (default parameters) by mapping preprocessed reads from AMARA and TEKIR datasets to the assembled contigs. Hybrid binning (based on tetranucleotide frequencies and coverage data) was performed using MetaBAT2[42] with default parameters. Bin completeness, contamination and strain heterogeneity were estimated using CheckM[43] 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 3% were denominated as metagenome assembled genomes (MAGs). A total of 35 MAGs were recovered: 23 Lokiarchaeia, 10 Thorarchaeia and 2 Heimdallarchaeia (**Supplementary Table 1**). Unbinned contigs were kept for further analyses (total nucleotide bases/site: 3.46 Mbp Amara and 4.06 Mbp Tekirghiol).

Genome annotation: Publicly available Asgardarchaeota genomes were downloaded from the NCBI Genome section (https://www.ncbi.nlm.nih.gov/genome). Coding sequences were predicted *de novo* with Prokka[44] for all available Asgard MAGs (35 from this study, 14 from NCBI – Accession numbers can be found in **Supplementary Table 2**). BlastKOALA[45] was used to assign KO identifiers (K numbers) to orthologous genes (**Supplementary Table 3**). 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 were not considered for metabolic reconstruction due to paucity of genome-level 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[46], respectively. All predicted proteins were queried against NCBI NR, COGs (cluster of orthologous groups) and arCOGs (archaeal cluster of orthologous groups)[47]. A locally installed version of InterProScan[48] was used with default settings to annotate protein domains. Potential eukaryote specific proteins (ESP) were identified based on previously published lists of IPR domains[5] (**Supplementary Table 5**) identified in Asgardarchaeota. Previously unidentified ESP were searched based on key words related to eukaryotic specific processes and/or structures. All IPR domains present exclusively in Asgardarchaeota genomes assembled in this study 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[5] ESP - DNA polymerase epsilon, catalytic subunit (IPR029703) - was identified by querying all MAG proteomes with human sequences (**Supplementary Table 5**). Several candidate ESP sequences were further analyzed using jackhmmer[49], Phyre2[50] and Phobius[51].

Phylogenetic trees: While all the genomic data was used in metabolic reconstructions, the phylogenetic analyses were performed only with MAGs that had >50% completeness. Due to the challenges associated with reconstructing the evolutionary relationships between archaea and eukayotes[5], in our inferences we used only those MAGs (n= 8; **Supplementary Table 1**) that harbored at least 75% of total phylogenetic markers (See **Supplementary Table 7**).

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 Asgardarchaeota), 21 bacteria and 13 eukaryotes (**Supplementary Table 6**). SSU and LSU sequences

were aligned independently by PRANK[52] (parameters: -DNA +F), trimmed using BMGE[53] (–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[5]. Maximum likelihood phylogeny for concatenated SSU+LSU gene sequences was inferred using IQ-TREE (-m GTR+I+G4+F) with ultrafast bootstrapping -bb 1000 and Shimodaira-Hasegawa testing –alrt 1000[54,55].

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 Asgardarchaeota, 7 DPANN) and 8 eukaryotes (**Supplementary Table 6**). Selection criteria for phylogenomic trees of ribosomal proteins conserved between archaea and eukaryotes have been previously described[5]. 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 7**). Additionally, some proteins that were not identified by arCOG scanning were retrieved from NCBI Protein (https://www.ncbi.nlm.nih.gov/protein). Sequences were aligned using PRANK (-protein +F), trimmed with BMGE[53] (-m BLOSUM30 -t AA -g 0.2 -b 3), concatenated, and subjected to SR4 amino acid recoding[13]. Maximum likelihood trees were generated by IQ-TREE (-bb 1000, -alrt 1000) with ultrafast bootstrapping[54] and the custom 'C60SR4' model described in a previous study[5]. Bayesian inference phylogenies were constructed using PhyloBayes MPI 1.8[56], using the CAT-Poisson model. Four chains were run in parallel until estimated maxdiff values calculated by bpcomp (-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[57] (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: MUSCLE[58] 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[59,60]. 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[61]. 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 8**). MAFFT-L-INS-i[62] (default parameters) and PRANK[52] (parameters: -PROTEIN +F) were used for aligning sugar kinase and kynurenine pathway enzyme sequences, respectively, followed by trimming using BMGE[53] (-m BLOSUM30 -t AA -g 0.5 -b 3). Single protein maximum likelihood trees were constructed with FastTree2[63], using an accurate search strategy (-mlacc 2 –spr 4 –slownni), and 100 standard bootstrap replicates.

Data and code availability

Sequence data generated during this study has been deposited in the NCBI Sequence Read Archive (SRA) under study number SRP155597 and linked to BioProject ID PRJNA483005. The Whole Genome

Shotgun project containing genome bins assembled in this study has been deposited at DDBJ/ENA/GenBank under the accessions SDMS00000000-SDOA00000000. The versions described in this paper are version SDMS01000000-SDOA01000000. Derived data supporting the findings presented in this paper are available in the Figshare repository with identifier DOI: 10.6084/m9.figshare.702262, [https://figshare.com/s/eecba719e91d17c60c3f]. All other relevant data supporting the findings of this study are available within the paper and its supplementary information files. No custom code that is central to the conclusions of this study was generated. All programs used in data analyses are listed in detail with their version numbers in the Nature Research Reporting Summary linked to this article.

References

1. Cox, C. J., Foster, P. G., Hirt, R. P., Harris, S. R. & Embley, T. M. The archaebacterial origin of eukaryotes. *Proc. Natl. Acad. Sci.* **105,** 20356 LP-20361 (2008).
2. Williams, T. A., Foster, P. G., Cox, C. J. & Embley, T. M. An archaeal origin of eukaryotes supports only two primary domains of life. *Nature* **504,** 231 (2013).
3. Spang, A. *et al.* Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* **521,** 173 (2015).
4. Spang, A. *et al.* Asgard archaea are the closest prokaryotic relatives of eukaryotes. *PLOS Genet.* **14,** e1007080 (2018).
5. Zaremba-Niedzwiedzka, K. *et al.* Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* **541,** 353–358 (2017).
6. 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).
7. Da Cunha, V., Gaia, M., Nasir, A. & Forterre, P. Asgard archaea do not close the debate about the universal tree of life topology. *PLOS Genet.* **14,** e1007215 (2018).
8. de Duve, C. The origin of eukaryotes: a reappraisal. *Nature reviews. Genetics* **8,** 395–403 (2007).
9. Lake, J. A., Henderson, E., Oakes, M. & Clark, M. W. Eocytes: a new ribosome structure indicates a kingdom with a close relationship to eukaryotes. *Proc. Natl. Acad. Sci.* **81,** 3786–3790 (1984).
10. Woese, C. R., Kandler, O. & Wheelis, M. L. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. U. S. A.* **87,** 4576–4579 (1990).
11. 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).
12. Martijn, J., Vosseberg, J., Guy, L., Offre, P. & Ettema, T. J. G. Deep mitochondrial origin outside the sampled alphaproteobacteria. *Nature* **557,** 101–105 (2018).
13. Susko, E. & Roger, A. J. On reduced amino acid alphabets for phylogenetic inference. *Mol. Biol. Evol.* **24,** 2139–2150 (2007).
14. Liu, Y. *et al.* Comparative genomic inference suggests mixotrophic lifestyle for Thorarchaeota. *ISME J.* **12,** 1021–1031 (2018).
15. Dodding, M. P. Folliculin – A tumor suppressor at the intersection of metabolic signaling and membrane traffic. *Small GTPases* **8,** 100–105 (2017).
16. Pushkarev, A. *et al.* A distinct abundant group of microbial rhodopsins discovered using functional metagenomics. *Nature* **558,** 595–599 (2018).
17. Flores-Uribe, J. *et al.* Heliorhodopsins are absent in diderm (Gram-negative) bacteria: Some thoughts and possible implications for activity. *Environ. Microbiol. Rep.* **0,** (2019).
18. Petrovskaya, L. E. *et al.* Predicted bacteriorhodopsin from Exiguobacterium sibiricum is a functional proton pump. *FEBS Lett.* **584,** 4193–4196 (2010).

19. Alexe, M. *Studiul lacurilor sărate din Depresiunea Transilvaniei*. (Presa Universitară Clujeană, 2010).
20. Sousa, F. L., Neukirchen, S., Allen, J. F., Lane, N. & Martin, W. F. Lokiarchaeon is hydrogen dependent. *Nat. Microbiol.* **1,** 16034 (2016).
21. Ternes, C. M. & Schönknecht, G. Gene Transfers Shaped the Evolution of De Novo NAD(+) Biosynthesis in Eukaryotes. *Genome Biol. Evol.* **6,** 2335–2349 (2014).
22. 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).
23. Kurnasov, O. *et al.* Aerobic tryptophan degradation pathway in bacteria: novel kynurenine formamidase. *FEMS Microbiol. Lett.* **227,** 219–227 (2003).
24. 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).
25. 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).
26. 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).
27. Kono, T. *et al.* A RuBisCO-mediated carbon metabolic pathway in methanogenic archaea. *Nat. Commun.* **8,** 14007 (2017).
28. Techtmann, S. M. *et al.* Evidence for horizontal gene transfer of anaerobic carbon monoxide dehydrogenases. *Front. Microbiol.* **3,** 132 (2012).
29. Martin, W. F. Hydrogen, metals, bifurcating electrons, and proton gradients: The early evolution of biological energy conservation. *FEBS Lett.* **586,** 485–493 (2012).
30. Betts, H. *et al.* Integrated genomic and fossil evidence illuminates life's early evolution and eukaryote origins. *Nat. Ecol. Evol.* (2018).
31. 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).
32. Fedorov, P. V. Postglacial transgression of the Black Sea. *Int. Geol. Rev.* **14,** 160–164 (1972).
33. Bushnell, B., Rood, J. & Singer, E. BBMerge - Accurate paired shotgun read merging via overlap. *PLoS One* **12,** e0185056 (2017).
34. Bushnell, B. BBMap short read aligner. *Univ. California, Berkeley, California. URL http//sourceforge. net/projects/bbmap* (2016).
35. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26,** 2460–2461 (2010).
36. 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).
37. Nawrocki, E. P. Structural RNA Homology Search and Alignment using Covariance Models. (Washington University School of Medicine, 2009).
38. 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).
39. 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).
40. 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).

41. Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35,** 1026 (2017).
42. 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).
43. 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).
44. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30,** 2068–2069 (2014).
45. 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).
46. 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).
47. 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).
48. Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30,** 1236–1240 (2014).
49. Finn, R. D. *et al.* HMMER web server: 2015 update. *Nucleic Acids Res.* **43,** W30-8 (2015).
50. 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).
51. 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).
52. Loytynoja, A. Phylogeny-aware alignment with PRANK. *Methods Mol. Biol.* **1079,** 155–170 (2014).
53. 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).
54. 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).
55. 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).
56. 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).
57. 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).
58. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32,** 1792–1797 (2004).
59. 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).
60. 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).
61. 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).
62. Katoh, K. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30,** 3059–3066 (2002).
63. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2--approximately maximum-likelihood trees

for large alignments. *PLoS One* **5,** e9490 (2010).

Correspondence and requests for materials should be addressed to R.G. (ghai.rohit@gmail.com).

Acknowledgements

We are thankful to Z. Keresztes, V. Muntean, T. Szőke-Nagy, M. Alexe, A. Cristea, and A. Baricz for their technical support during sampling and sample preparation. The contribution of E. A. Levei and M. Șenilă in performing chemical analyses 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.L.B. 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 L200961651 (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. R.G., O.B., K.I. and H.K. performed rhodopsin data analyses. All authors commented on and approved the manuscript.

Competing interests

The authors declare no competing interests.

Figures Legends

Figure 1. Asgardarchaeota phylogenomics. a, Maximum likelihood (LG+C60, general matrix and 60-profile protein models of amino acid substitution) phylogeny of the Asgardarchaeota superphylum. The green circles highlight UFBoot values higher than 95. **b**, Asgardarchaeota phylogeny generated through Bayesian inference (CAT-Poisson, CAT model of amino acid substitution with uniform global exchange rates). The posterior probability values are shown above the internal nodes. High support for Eukaryota/Asgardarchaeota 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. **a**,**b**, Both phylogenies are based on 48 concatenated ribosomal protein markers (Supplementary Table 7) with **a**, 85 archaeal lineages (Supplementary Table 6) selected for ML-inference and **b**, 85 archaeal + 8 eukaryotic lineages used for Bayesian inference (Supplementary Table 6). Panel (**c**) provides a census of the eukaryotic signature proteins (ESP) found in all available MAGs. The grey box highlights ESP identified during this study.

Figure 2. Phylogenetic analysis of rhodopsins. An unrooted maximum likelihood tree of all Asgardarchaeota schizorhodopsins (n=6) identified in this work, heliorhodopsins and representative known type-1 rhodopsins, is shown. The branches colored red are sequences from the Asgardarchaeota. Bootstrap values on nodes are indicated by colored circles (see color key at the right). A total of 392 sequences, spanning known rhodopsin famillies and including schizorhodopsins retrieved in this study, were used for phylogenetic inference. All related data was deposited in Figshare.

Figure 3. Metabolic reconstruction of Heimdallarchaeia. The text in the yellow panels depicts names of pathways and metabolic processes. Abbreviations: ACSS - acetyl-CoA synthetase and carbonic anhydrase; acyP - acylphosphatase; ampp - AMP phosphorylase; APRT – AMP pyrophosphorylase; ArsC - arsenate reductase (glutaredoxin); BCAA - branched-chain amino acid; CODH - carbon monoxide dehydrogenase; gcvPAB – glycine dehydrogenase; glyA – glycine hydroxymethyltransferase; hmp – nitric oxide dioxygenase; maeA - malate dehydrogenase (decarboxylating); PC - pyruvate carboxylase; PEPCK - phosphoenolpyruvate carboxykinase; pflD - formate C-acetyltransferase; PFOR - pyruvate ferredoxin oxidoreductase; PK - pyruvate kinase; poxL – pyruvate oxidase; PPDK - pyruvate, phosphate dikinase; PPP - pentose phosphate pathway; Rpi - ribose-5-phosphate isomerase; RuBisCO - Ribulose-1,5-bisphosphate carboxylase/oxygenase; SOD - superoxide dismutase; TCA - tricarboxylic acid cycle.

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. Abbreviations: 3HAO - 3-hydroxyanthranilate 3,4-dioxygenase; AFMID - arylformamidase; ASO - L-aspartate oxidase; KMO - kynurenine 3-monooxygenase; KYNU - kynureninase; NMNAT - nicotinamide nucleotide adenylyltransferase; NS - NAD^+ synthase; QPT - nicotinate-nucleotide pyrophosphorylase; QS - quinolinate synthase; TDO - tryptophan 2,3-dioxygenase.

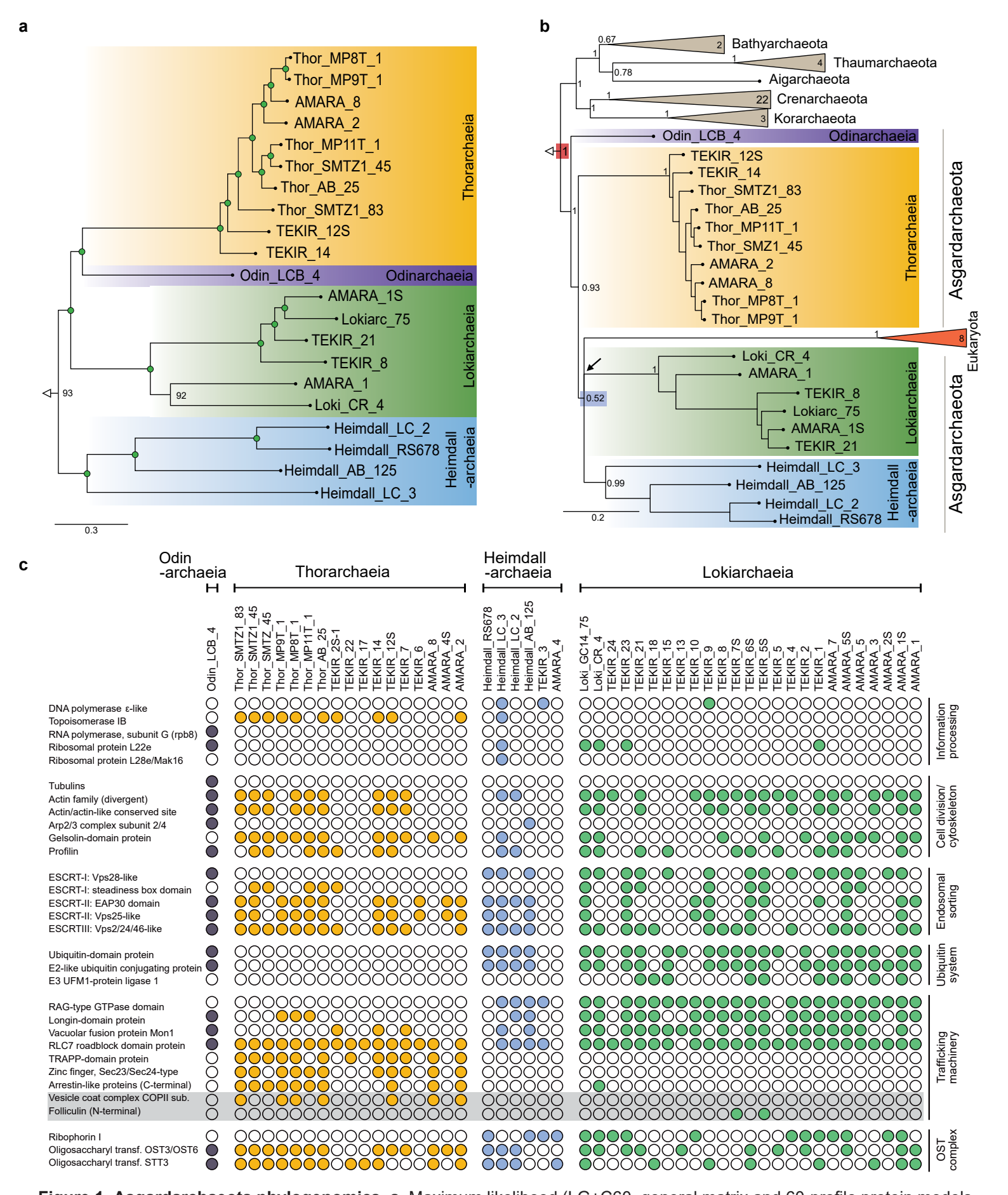

Figure 1. Asgardarchaeota phylogenomics. a, Maximum likelihood (LG+C60, general matrix and 60-profile protein models of amino acid substitution) phylogeny of the Asgardarchaeota superphylum. The green circles highlight UFBoot values higher than 95. **b**, Asgardarchaeota phylogeny generated through Bayesian inference (CAT-Poisson, CAT model of amino acid substitution with uniform global exchange rates). The posterior probability values are shown above the internal nodes. High support for Eukaryota/Asgardarchaeota 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. **a**,**b**, Both phylogenies are based on 48 concatenated ribosomal protein markers (Supplementary Table 7) with **a**, 85 archaeal lineages (Supplementary Table 6) selected for ML-inference and **b**, 85 archaeal + 8 eukaryotic lineages used for Bayesian inference (Supplementary Table 6). Panel (**c**) provides a census of the eukaryotic signature proteins (ESP) found in all available MAGs. The grey box highlights ESP identified during this study.

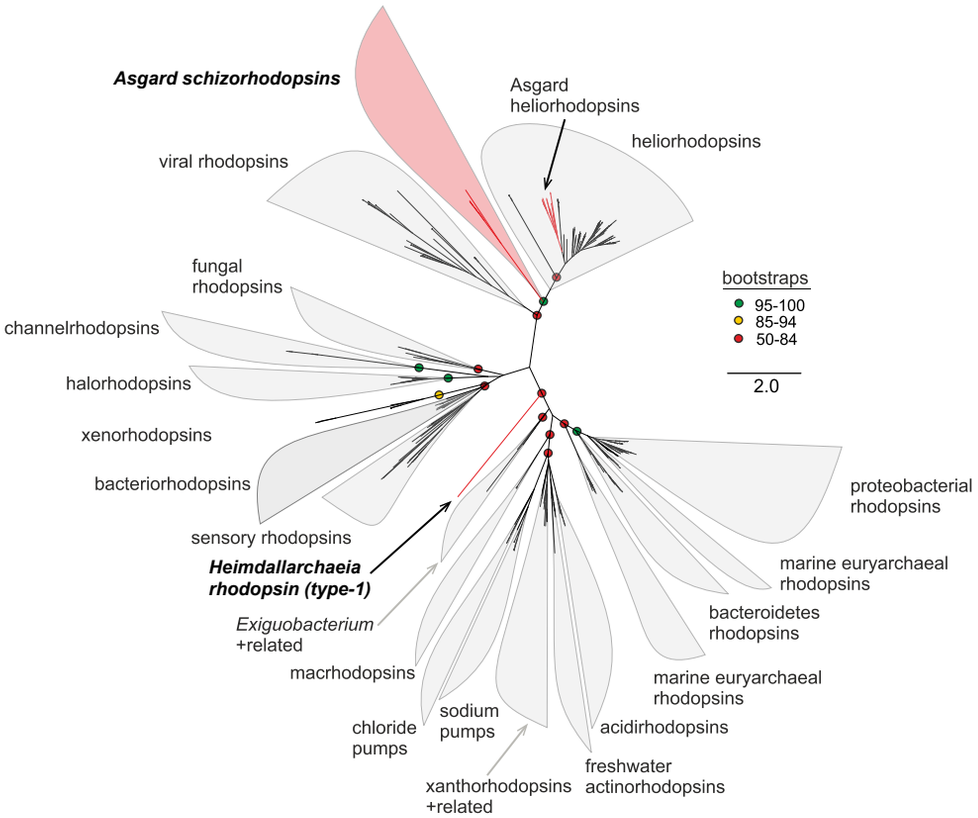

Figure 2. Phylogenetic analysis of rhodopsins. An unrooted maximum likelihood tree of all Asgardarchaeota schizorhodopsins (n=6) identified in this work, heliorhodopsins and representative known type-1 rhodopsins, is shown. The branches colored red are sequences from the Asgardarchaeota. Bootstrap values on nodes are indicated by colored circles (see color key at the right). A total of 392 sequences, spanning known rhodopsin famillies and including schizorhodopsins retrieved in this study, were used for phylogenetic inference. All related data was deposited in Figshare.

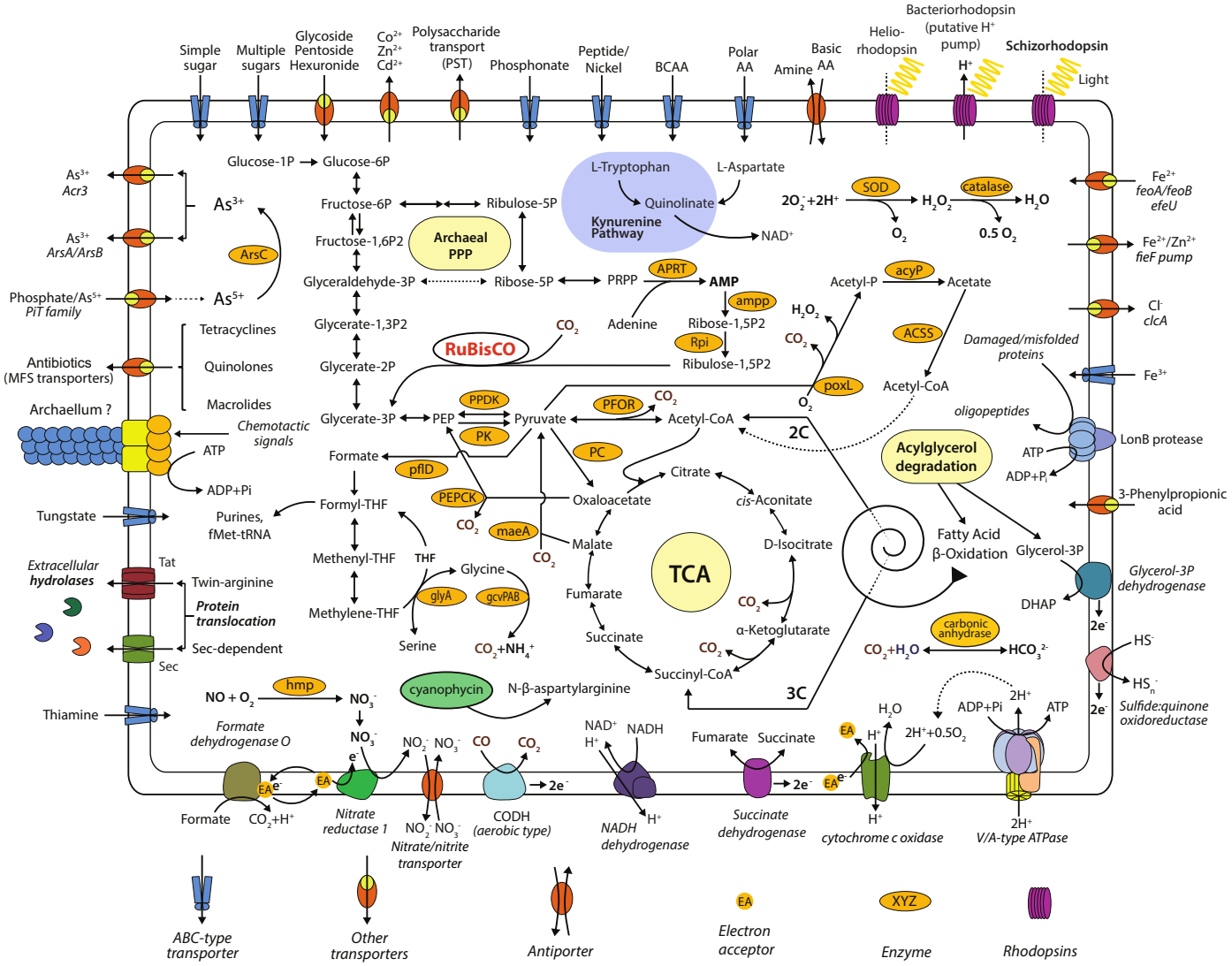

Figure 3. Metabolic reconstruction of Heimdallarchaeia. The text in the yellow panels depicts names of pathways and metabolic processes. Abbreviations: ACSS - acetyl-CoA synthetase and carbonic anhydrase; acyP - acylphosphatase; ampp - AMP phosphorylase; APRT - AMP pyrophosphorylase; ArsC - arsenate reductase (glutaredoxin); BCAA - branched-chain amino acid; CODH - carbon monoxide dehydrogenase; gcvPAB – glycine dehydrogenase; glyA - glycine hydroxymethyltransferase; hmp - nitric oxide dioxygenase; maeA - malate dehydrogenase (decarboxylating); PC - pyruvate carboxylase; PEPCK - phosphoenolpyruvate carboxykinase; pflD - formate C-acetyltransferase; PFOR - pyruvate ferredoxin oxidoreductase; PK - pyruvate kinase; poxL – pyruvate oxidase; PPDK - pyruvate, phosphate dikinase; PPP - pentose phosphate pathway; Rpi - ribose-5-phosphate isomerase; RuBisCO - Ribulose-1,5-bisphosphate carboxylase/oxygenase; SOD - superoxide dismutase; TCA - tricarboxylic acid cycle.

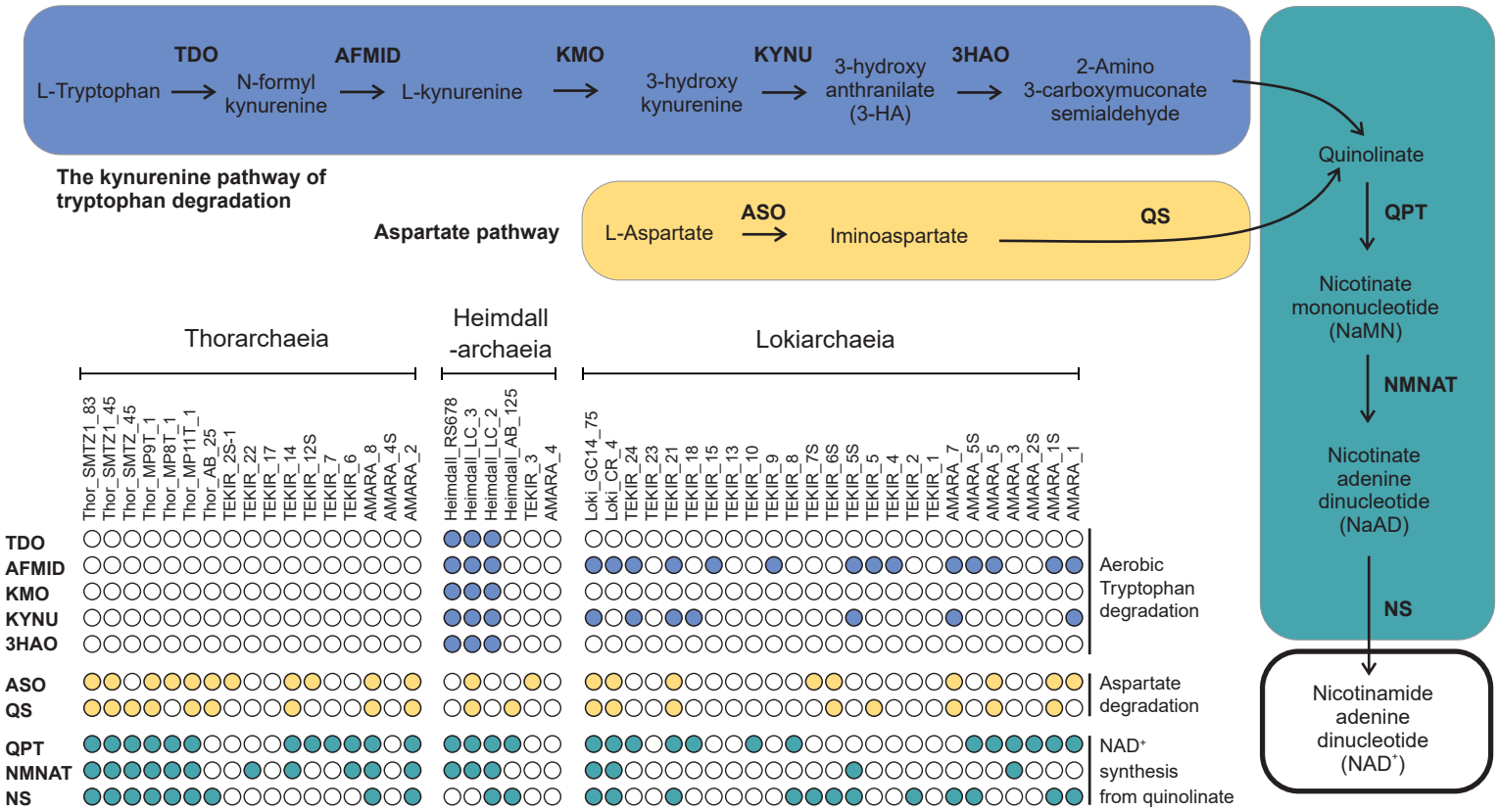

Figure 4. *De novo* NAD⁺ synthesis pathways. The colored boxes show a schematic representation of the kynurenine and aspartate pathways involved in *de novo* NAD⁺ synthesis. The presence of the enzymes involved in these pathways is indicated for each MAG by using a colored circle. Abbreviations: 3HAO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