
Supplementary information

**Oxygen metabolism in descendants of the
archaeal-eukaryotic ancestor**

In the format provided by the
authors and unedited

Oxygen metabolism in descendants of the archaeal-eukaryotic ancestor

Kathryn E. Appler¹, James P. Lingford², Xianzhe Gong^{1,3}, Kassiani Panagiotou⁴, Pedro Leão^{1†}, Marguerite Langwig^{1,5}, Chris Greening^{2,6}, Thijs J. G. Ettema⁴, Valerie De Anda^{1,7†}, and Brett J. Baker^{1,7*}

¹Department of Marine Science, Marine Science Institute, University of Texas at Austin, Port Aransas, TX, USA

²Department of Microbiology, Biomedicine Discovery Institute, Monash University, Clayton, VIC, Australia

³Institute of Marine Science and Technology, Shandong University, Qingdao, Shandong, China

⁴Laboratory of Microbiology, Wageningen University & Research, Wageningen, The Netherlands

⁵Department of Bacteriology and Department of Freshwater and Marine Sciences, University of Wisconsin-Madison, Madison, WI, USA

⁶Securing Antarctica's Environmental Future, Monash University, Clayton, VIC, Australia

⁷Department of Integrative Biology, University of Texas at Austin, Austin, TX, USA

[†]Present Address: Department of Microbiology - RIBES, Radboud University, Nijmegen, The Netherlands

[†]Present Address: Department of Functional and Evolutionary Ecology, Microbial Oceanography Unit, University of Vienna, Djerassiplatz 1, 1030 Vienna, Austria

*Corresponding author: acidophile@gmail.com

Supplementary Information

Supplementary Methods	4
1. Sampling sites	4
2. Asgardarchaeota database curation	4
3. Phylogenetic analyses of the expanded Asgard genomic catalog	5
4. Phylogenetic and structural analyses of Complex I-like hydrogenases	6
a. Sequence homology- Synteny	6
b. Structural modeling	6
i. AlphaFold2 multimer predictions	6
ii. Cofactor modeling	6
iii. NiFe modeling	7
iv. FeS modeling in NuoB	7
v. FeS modeling in NuoI	7
vi. FeS modeling in MbhO	8
vii. FeS modeling in MbhP	8
viii. Electron transfer path	8
ix. Quinone binding site modeling	8
x. NuoEFG multimer predictions	8
Supplementary Discussion	10
1. Asgard archaea and Alphaproteobacterial abundance and co-occurrence	10
2. Asgardarchaeota co-occurrence in the Bohai Sea and Guaymas Basin	11
3. Phylogenomic analyses of the expanded Asgardarchaeota genomic catalog	13
a. Discrepancies within the branching order	13
b. Phylogenetic diversity	14
4. Metabolism of new classes	15
a. Ranarchaeia	15
b. Asgardarchaeia	17
5. Guild metabolism	17
a. Guild 11 and 12	17
i. Hodarchaeales	18
6. Evolutionary and ecological transitions in metabolic traits	20
7. Key metabolic proteins	21
a. Divergent [FeFe]-hydrogenases	21
b. Non-methanogenic methylotrophy	23
c. Oxygen metabolism and ROS tolerance	24
i. Ribulose-1,5-bisphosphate carboxylase/oxygenase family	24
ii. Superoxide dismutase	24

iii. Catalase	25
iv. Peroxidase	25
8. Identified ESPs and eukaryotic characteristics	25
a. <i>recA/RAD51</i> gene family	25
b. <i>Hsp90</i>	26
c. L28e/Mak16 domain	26
d. PF10281	26
9. Phylogenetic and structural analyses of [NiFe]-hydrogenases	27
a. Phylogenetic distribution	27
i. Novel [NiFe]-hydrogenase subgroups	27
ii. Distribution of novel [NiFe]-hydrogenase subgroups	27
b. Structural architecture and putative mechanisms of group 4 [NiFe]-hydrogenase complexes	29
i. Complex I-like vs. MBH-like structural architecture	29
ii. AlphaFold2 model confidence and limitations	30
iii. Complex I-like peripheral arm subunits	30
iv. Complex I-like membrane arm subunits	32
v. MBH-like peripheral arm subunits	32
vi. MBH-like membrane arm subunits	33
vii. Separate NuoEFG-like complexes	34
c. Evolutionary implications	38
10. Tracing the evolution of respiratory complexes	40
a. Complex I-like NADH:ubiquinone oxidoreductase subunits	40
b. Complex II- Succinate dehydrogenase	44
c. QcrABC- Cytochrome c reductase	44
d. Nitrate respiratory complexes	45
e. Complex IV- HCO superfamily	45
f. Complex V-ATP synthase	46
11. Potential impact on eukaryogenesis	46
a. Actin homologs	46
b. ROS signaling	46
c. SCO1	47
12. SeqCode Taxonomic Descriptions	47
Supplementary Figures	53
Supplementary Tables	144
Supplementary References	149

Supplementary Methods

Sampling sites

The Bohai Sea (BH; 38°00.849 N, 119°00.789 E) is a shallow coastal ocean off the coast of China influenced by numerous river inputs, which are the dominant sediment source with an estimated sedimentation rate between 1.5-10 mm/year¹. The distribution and abundance of microbial diversity correspond with organic matter availability and increase in anthropogenic influence close to shore^{2,3}. Guaymas Basin (GB) is a site of active seafloor spreading in the Gulf of California characterized by its hydrothermal vents and hydrocarbon-rich sediments, increasing biodiversity, archaeal abundance, and metabolic partitioning^{4,5}. The heavy organic matter rainfall, 0.23-1 mm/year sedimentation rate, of the GB and the dynamic nature of the young marginal rift basin, provide a continued source of genomic diversity, including 13 of the Asgard lineages, including new classes, Asgardarchaeia and Ranarchaeia^{5,6}.

This study provides the first description of GE3 (GB expedition), including sampling, extraction, sequencing, assembly, binning, etc. This sequencing expedition includes three Alvin dives from which we obtained five cores split into 12 depth horizons (Supplementary Fig. 2). The naming scheme for the corresponding MAGs is Dive_Core(s)_Bin#. The first two portions of the name refer to the Alvin dive and cores to assist in tracking sediment samples across different studies. The horizon numbers increase from shallow to deeper layers of sediment (e.g., H1 shallowest, H2 middle layer, H3 deepest; Supplementary Fig. 2). Only modification to the name is D4493_5_LG, which is the same layer as D4493_5_H3. However, this sediment was sampled, extracted, and sequenced separately. Additional information on the sampling sites (i.e., location, elevation, depth, and temperature) of GE3 is included in Supplementary Table 1.

Asgardarchaeota database curation

To compare metabolism across the Asgardarchaeota phylum, we used the most updated taxonomic delineation⁷. This paper considered GTDB-tk and SeqCode to preserve taxonomic congruence while adhering to updated policies. Recent validation of the superphylum, Asgard archaea, as the phylum, Asgardarchaeota⁸ reclassified all previously proposed phyla as classes, orders, and families. To preserve these previous taxonomic descriptions and add Ranarchaeia, we split the genomes into 17 taxonomic lineages (order-family) shown in Fig. 1a and Supplementary Table 4. One lineage currently has two proposed names (see previous discussion), so we use Freyarchaeia or Frey/Jordarchaeia⁹. Additionally, we provided an environmental perspective gathering location of all previously published MAGs. The elevation of one genome was listed as 5-1000 mbsl, so this MAG was categorized as mesopelagic in our analyses (GCA_002728275.1)¹⁰.

Phylogenetic analyses of the expanded Asgard genomic catalog

Using phylogenomic analyses, we identified several Asgard representatives mislabeled as TACK and Euryarchaeota in public databases with 30 MAGs listed as unclassified or misidentified. To reduce confusion in this rapidly expanding lineage, we noted an additional 25 MAGs with taxonomic incongruence in prior work (Supplementary Table 4). These genomes did not include modifying phyla to lower taxonomic levels or the updated separation of Heimdallarchaeia into multiple lineages^{7,11}.

For the construction of NM47 individual protein phylogenies, IQ-TREE v2.1.3¹² was employed, incorporating model selection from ModelFinder¹³. The best fitting model was selected among the combination of the LG, Q.pfam, WAG models by adding the mixture model C20 with rate heterogeneity (+R4 or +G4). To assess the robustness of the best scoring evolutionary model for each marker a total of 1000 ultrafast bootstraps¹⁴ and 1000 SH-like approximate likelihood ratio test replicates (SH-aLRT)¹⁵ were calculated. The resulting phylogenies were examined to identify sequences indicative of contamination, paralogy or horizontal gene transfer events. After removing such sequences, the remaining sequences were realigned and trimmed as described in the Methods section of the main text and concatenated into a supermatrix containing 1178 sequences. A phylogeny was inferred with IQ-TREE v2.1.3 using this supermatrix. Model selection was performed among the LG+C10, Q.pfam+C10 and WAG+C10 models and rate heterogeneity components (+R4 or +G4). The final species tree was reconstructed using a PMSF approximation of the selected model (WAG+C10+R4) and the previous tree as a guide with 100 nonparametric bootstrap pseudoreplicates (Fig. 1a and Supplementary Table 2). In the un-recoded dataset (Fig. 1a), Heimdallarchaeia appear paraphyletic and attracted toward the outgroup possibly indicating the presence of a long branch attraction artifact due to the long branch introduced by Njordarchaeales^{7,16}.

Given the positioning of Njordarchaeales outside the Heimdallarchaeial clade, we decided to mitigate artifacts related to compositional bias and/or saturation by recoding the initial untreated alignment into four saturation-prone character states (SR4-recoding)¹⁷ using BioKIT v0.0.9¹⁸ and to employ more complex mixture models (60 frequency vectors instead of 10). A phylogenetic tree for the same dataset was then inferred using this recoded alignment with IQ-TREE v2.1.3 under the GTR+G4+C60 model adapted to SR4 as provided by Xavier Grau-Bové (<https://github.com/xgrau/recoded-mixture-models>). A PMSF approximation¹⁹ of the selected model using the resulting tree as a guide was then used to reconstruct a final tree with 100 nonparametric bootstrap pseudoreplicates (Fig. 1b). The recoded analysis conserves the monophyly of Heimdallarchaeia and Lokiarchaeia.

To ensure the recoding did not compromise the phylogenetic signal, we dereplicated our dataset to include species-level representatives of the Asgardarchaeota ingroup, enabling the application of the more computationally intensive C60 mixture model. With this model, we observed increased support for the depicted topologies and

a monophyletic Heimdallarchaeia clade (as was the case for the recoded dataset) (Supplementary Fig. 3). There are a few discrepancies in the branching order between various profile mixture models and recoding but the monophyly of each of the 17 clades remains maximally supported (Fig. 1ab and Supplementary Fig. 3)⁷. Additional high-quality genomes will continue to increase the phylogenomic support, especially for deeper nodes of the Asgard phylogeny.

Phylogenetic and structural analyses of Complex I-like hydrogenases

Sequence homology- Synteny

To narrow down the sequences for AlphaFold2 (AF2) structural modeling, we identified the upstream and downstream annotations of the [NiFe]-hydrogenase large catalytic subunit. Using the functional databases (see methods), several of the NADH:quinone oxidoreductase subunit D-like and [NiFe]-hydrogenases were flanked by other respiratory subunits. Scaffolds containing the consecutive subunits were used for structural analyses with AF2. NADH:quinone oxidoreductase subunits E, F, and G were identified across Asgardarchaeota and in other archaea separate from the other NADH:quinone oxidoreductase subunits. Similarly, scaffolds upstream and downstream from these subunits were used for structural analyses with AF2.

Structural modeling

AlphaFold2 multimer predictions

Gene sequences identified in the maximum step (see “Sequence homology- Synteny” section above) were used as input for AF2 multimer²⁰. See Methods section for parameter details used for running AF2.

Cofactor modeling

We used simple structural comparisons and modeling methods to infer the approximate positions of putative cofactors and ligands in our Complex I-like AF2 models. First, subunits of interest from the top scoring AF2 models (i.e. NuoD, NuoB, NuoI, MbhL, MbhJ, MbhN, MbhO, and MbhP) were queried through the FoldSeek web server^{21,22} (<https://foldseek.com>, 3Di/AA search mode) to find similar structures in the PDB100 database. Second, PDB structures from the FoldSeek search were selected based on low E-value scores, low RMSD, and sequence conservation of key metal coordinating cysteine residues (for a complete list of selected PDBs and their RMSDs with the AF2 models, see Supplementary Table 9). The PDB and AF2 models were structurally superimposed in UCSF ChimeraX v1.8²³ with the matchmaker command (Needleman-Wunsch algorithm settings). The resulting structural superposition of cofactors over the “apo” AF2 models produces an approximate “holo” AF2 model. “Holo” AF2 models free of steric clashes

between the cofactors/ligands and the surrounding protein were judged as biochemically reasonable. Modeling of individual cofactors and ligands are described in the following sections.

NiFe modeling

To check the metal cofactors of the Asgard sequences, we queried FoldSeek against the NuoD (Kari-1) AF2 model returned *E. coli* HycE (PDB: 7Z0T²⁴, large subunit chain E) as the top result (E-value: 3.30e-29). Structural superposition indicates that all AF2 multimer models of the group 4 [NiFe]-hydrogenase membrane-bound complexes ($7 \times$ NuoD/NADH:quinone oxidoreductase subunit D, $2 \times$ MbhL/Membrane-bound hydrogenase) can plausibly coordinate a [NiFe] cofactor (Supplementary Fig. 47b). Some residues in the AF2 models have incorrect side-chain conformations. This inaccuracy may be a limitation of AF2 in predicting side-chain conformations²⁵. The [NiFe] cofactor could not be plausibly modeled into the AF2 multimer model of the archaeal Complex I (Hod-3), indicating this complex is not a group 4 [NiFe]-hydrogenase (Supplementary Fig. 79).

FeS modeling in NuoB

For the group 4 [NiFe]-hydrogenase membrane-bound complexes, we queried FoldSeek with NuoB from the Kari-1 AF2 model, which returned *P. furiosus* MbhJ (PDB: 6CFW²⁶, chain J) as the top result (E-value: 2.44e-18). Structural superposition indicates that all group 4 [NiFe]-hydrogenase membrane-bound complexes AF2 models ($7 \times$ NuoB, $2 \times$ MbhJ) can plausibly coordinate an iron-sulfur (FeS) cluster (Supplementary Fig. 48a). NuoB from the Hod-3 AF2 model was queried separately, which returned *A. thaliana* NuoB (PDB: 8BEE²⁷, chain B) as the top result (E-value: 3.50e-20). Structural superposition indicated that the Hod-3 NuoB can plausibly coordinate one FeS cluster (Supplementary Fig. 79).

FeS modeling in NuoI

For the group 4 [NiFe]-hydrogenase membrane-bound complexes, we queried FoldSeek with NuoI from the Kari-1 AF2 model, which returned *T. elongatus* NDH-I (PDB ID: 6NBY²⁸, chain I) as the top result (E-value: 1.98e-9). Structural superposition indicates that all nine AF2 models ($7 \times$ NuoB, $2 \times$ MbhJ) can plausibly coordinate two FeS clusters (Supplementary Fig. 48b). NuoI from the Hod-3 AF2 model was queried separately, which returned *A. thaliana* NuoI (PDB: 8BEE²⁷, chain I) as the top result (E-value: 9.72e-8). Structural superposition indicated that the Hod-3 NuoI can plausibly coordinate two FeS clusters (Supplementary Fig. 79).

FeS modeling in MbhO

FoldSeek queried against the MbhO (Asgard-1) AF2 model returned *E. coli* HycB (PDB ID: 7Z0T²⁴, chain B) as the third result (E-value: 1.22e-11) and was chosen over the top two results due to this structure also being relevant for the structural modeling of NuoD (see above). Structural superposition indicates that all nine AF2 models (7 × NuoB, 2 × MbhJ) can plausibly coordinate four FeS clusters (Supplementary Fig. 48c).

FeS modeling in MbhP

FoldSeek queried against the MbhP (Asgard-1) AF2 model returned *M. thermolithotrophicus* heterodisulfide reductase subunit H (HdrH; PDB ID: 5ODH²⁹) as the second result (E-value: 4.08e-25) and was chosen over the first result as it contains additional information about the heterodisulfide ligand. Structural superposition indicates that MbhP can plausibly coordinate two non-cubane [4Fe-4S] clusters, which are characteristic of this class of heterodisulfide reductase (Supplementary Fig. 48d)²⁹.

Electron transfer path

Putative electron transfer paths are based on the edge-to-edge distances between the modeled FeS and [NiFe] cofactors. Distances were estimated in UCSF ChimeraX v1.8²³ and all found to be <14 Å distance expected for effective electron transfer on biological timescales (Supplementary Fig. 49)³⁰.

Quinone binding site modeling

To infer if quinone binding could occur in the membrane-bound respiratory complexes, the NuoDB and MbhLJ subunits from our AF2 models were structurally superimposed to *M. smegmatis* Complex I NuoD subunit (PDB ID: 8E9H³¹, chain D) (Supplementary Fig. 51). The *Ms*Complex I structure was chosen for structural comparison and quinone site modeling because it binds menaquinone (MK), which is the most relevant quinone species in the context of archaeal metabolism. The group 4 [NiFe]-hydrogenase membrane-bound AF2 models (subunits NuoDB and MbhLJ) contain residues that sterically occlude the MK quinone headgroup in its equivalent Complex I binding site. Therefore, we could not model MK into the group 4 [NiFe]-hydrogenase membrane-bound models (Supplementary Fig. 79). However, in the Hod-3 AF2 model, only a minor steric clash between the MK quinone headgroup and a nearby residue (Met54, NuoB) occurs. Since this Met54 residue is located in a loop region of NuoB, its spatial position may be conformationally flexible and not inhibit quinone binding (Supplementary Fig. 79).

NuoEFG multimer predictions

AF2 multimer indicated that annotated NuoEFG subunits in the Asgard lineages we tested did not form complexes with the other Nuo subunits (P and Q-modules) present

in the same genome. However, AF2 multimer predictions of NuoEFG subunits in combination with the neighboring proteins often predicted multimers. To investigate the possible novel roles of NuoEFG in Asgard archaea, we selected 26 gene clusters containing NuoEFG from 14 Asgardarchaeota and three archaeal lineages and iteratively ran AF2 multimer between NuoEFG and neighboring genes using AlphaPulldown³². Candidate multimeric complexes identified by AlphaPulldown were manually input into AF2 multimer and iteratively ran with different combinations of subunit stoichiometries. For example, NuoEFG inputs were duplicated and ran as Nuo(EFG)₂. AF2 multimer predictions with the highest predicted template modeling score (pTM) and interface pTM (ipTM) scores and lowest overall predicted aligned error (PAE) scores were selected for structural analyses. When running AF2 multimer on two gene clusters from the coastal guild 11 Hodarchaeales (MAGs GCA_016840025_1_ASM1684002v1 and GCA_020353515_1_ASM2035351v1), we noticed that initial AF2 multimer predictions indicated the presence of complexes larger than 5000 amino acids but could not be computed in full due to AF2's limitations on sequence input size (maximum input of ~4000 amino acids). To circumvent this limitation, we used CombFold³³ to model these two larger NuoEFG complexes (Supplementary Figs. 28-29).

CombFold³³ was run following its three-to-four step pipeline as described in the source code documentation. For the putative NuoEFG complex in MAG GCA_016840025, this involved defining the subunits to include in the prediction, running an all-against-all AF2 multimer prediction for all subunits, then providing these AF2 multimer predictions to the combinatorial assembly algorithm. For the putative NuoEFG complex in MAG GCA_020353515, the combinatorial assembly step produced inconsistent predictions when provided the all-against-all AF2 multimer predictions, possibly due to the high degree of structural symmetry along multiple spatial axes this putative complex exhibits. To circumvent this, we used CombFolds' optional step of predicting the structures of larger groupings of subunits and providing these as inputs for the combinatorial assembly step. These larger groupings of subunits were predicted using the AlphaFold3 (AF3) webserver³³ to accommodate their large input sequence size. The CombFold predicted structures have characteristics of flavin based electron-bifurcating complexes (see discussion).

Supplementary Discussion

Asgard archaea and Alphaproteobacterial abundance and co-occurrence

High-coverage sequencing and manual curation of these MAGs from these three projects greatly expand the Asgard diversity by 46.4% at the time of submission (Fig. 1b and Supplementary Table 3). Asgardarchaeota had an average abundance based on read coverage of 1.10% (9 lineages), 1.16% (13 lineages), and 0.82% (8 lineages) across the samples based on read coverage for the GE2, GE3, and BH, respectively. Hodarchaeales was most abundant based on read coverage in BH at 0.19% in comparison to 0.09% in GE2 and 0.003% in GE3 (Supplementary Table 14).

Our analysis also reconstructed 310 Alphaproteobacteria MAGs (9 GE2, 20 GE3, and 281 BH) identified with GTDB-tk v2.1.1 to look at modern representatives from the group from which mitochondria are believed to have evolved (Supplementary Table 1)³⁴. Abundance of Alphaproteobacteria was 0.11%, 0.09%, and 2.07% based on read coverage for the GE2, GE3, and BH, respectively. Several samples had a relatively high amount of Asgard reads, mapping 10.77% of GE2 MegaMat 12-14 cm deep.

Heimdallarchaeia were found to co-occur with Alphaproteobacteria genomes in three GE3 and three BH sampling sites but not in GE2 (Supplementary Table 10). Yet, we did identify a low abundance of Alphaproteobacteria in GE2 samples containing Gerdarchaeales, Heimdallarchaeaceae, and Hodarchaeales (Supplementary Table 14). Limited sequencing of these lower abundant representatives in a complex community likely impacted the ability to bin GE2 Alphaproteobacteria MAGs from those samples. Within GE3, three cores (D4991_C11, D4993_C5, and D4998_C2223) had occurrences of Alphaproteobacteria with Heimdallarchaeia from 0-6 cm deep in Cathedral Hill core 11 (5 cm 17°C), 0-15 cm deep in Southern mat mound core 5 (10-29°C), and 10-20 cm deep at Marker 14 in Aceto balsamico core 22 and 23 (0 cm 4°C; 25 cm 22-25°C). Only one Gerdarchaeales MAG in core 11 and one Gerdarchaeales and Heimdallarchaeaceae in combined cores 22 and 23, but several Gerdarchaeales (6), Kariarchaeaceae (1), Heimdallarchaeaceae (9), and Hodarchaeales MAGs (1) in core 5.

Within the Bohai Sea MAGs, we identified co-occurrence of Alphaproteobacteria at sites BHB10 (17 mbsl) from 0-2; 24-26; 36-38 cm, M3 (26.5 mbsl) from 28-30; 36-38; 42-44 cm, and M8 (30.5 mbsl) from 42-44; 56-58; 60-62 cm. We identified 2, 21, and 15 Hodarchaeales MAGs, 1, 4, and 3 Kariarchaeaceae MAGs, and 1, 2, and 1 Heimdallarchaeaceae MAGs with Alphaproteobacteria at BHB10, M3, and M8 sites, respectively (Supplementary Table 10). Hodarchaeales had 39 MAGs co-occurring with Alphaproteobacteria with all but one from coastal Bohai Sea sediments. In comparison, Gerdarchaeales, Kariarchaeaceae, and Heimdallarchaeaceae had 8, 9, and 14, respectively. Thereby providing the first evidence of the co-occurrence of Hodarchaeales with Alphaproteobacteria across sites and depths within coastal sediments.

As Alphaproteobacteria includes anaerobic lineages, we also identified the specific orders co-occurring with Heimdallarchaeia, including 24 GTDB-tk orders and several unnamed lineages (o__). These lineages include eight described lineages (Kiloniellales, Micabrivirionales, Rhizobiales, Rhodobacterales, Rhodospirillales, Rickettsiales, Sphingomonadales, and Tistrellales) and several undescribed lineages (GCA-2731375, JAAEOK01, JACZXZ01, JAJFIK01, JAPULJ01, JAUHZY01, M2410, MPNO01, Rs-D84, SMXQ01, UBA11136, UBA1280, UBA6615, UBA7887, UBA828, UBA8366).

Interestingly, Rickettsiales, hotly debated as the potentially closest descendant of the ancestral symbiont, does co-occur with Hodarchaeales at the Bohai Sea site M8, 56-58 cm deep. Kiloniellales, Rhizobiales, UBA6615, and UBA8366 are found in all Bohai Sea sites with Hodarchaeales. JAPULJ01, Rhizobiales, Rhodobacterales, Rs-D84, Sphingomonadales, Tistrellales, and UBA6615 co-occur in the Guaymas Basin sediments. Kiloniellales, Rhodobacterales, and Sphingomonadales are aerobic lineages; whereas Rhizobiales is both aerobic and facultatively anaerobic, and the others have limited metabolic descriptions. More specifically, we classified five genera-level Alphaproteobacteria co-occurring with Hodarchaeales (*Altererythrobacter*, *Amylibacter*, *Boseongicola*, *Methyloceanibacter*, and *Neomegalonema*). Previous analysis has suggested all of these genera type species are aerobic with only *Altererythrobacter* split between aerobic and facultatively anaerobic³⁵.

Asgardarchaeota co-occurrence in the Bohai Sea and Guaymas Basin

We also report the co-occurrence of Asgardarchaeota genomes with all binned microbial lineages from the Bohai Sea and Guaymas Basin. For each site, we identified the number of Asgardarchaeota MAGs at the order level (GTDB v2.4.0) to assess co-occurrence patterns across all binned genomes (see Supplementary Table 10). We also ran pairwise co-occurrence analysis with the R package (cooccur)³⁶. These results were filtered for probability of co-occurrence (>0.5 prob_cooccur), and we provided additional details on potential positive co-occurrences of lineages with both Hodarchaeales and Kariarchaeaceae (<0.05 p_gt).

At the Bohai Sea site, Hodarchaeales MAGs had a probability of co-occurrence greater than 0.5 with Acidobacteriota, Chloroflexota, Desulfobacterota_E, Desulfobacterota, Gemmatimonadota, Myxococcota_A, Nitrospirota, Planctomycetota, and Pseudomonadota phyla. Within Pseudomonadota, there are representatives from Alphaproteobacteria (orders Rhizobiales and UBA8366) and Gammaproteobacteria classes (orders GCA-272945, HK1, and Woeseiales). At the same site, Lokiarchaeia and Thorarchaeia co-occur with both archaea and bacteria. Thorarchaeia co-occurs with the same bacterial phyla and an additional six phyla (Bacteroidota, BMS3Abin14, Calditrichota, Desulfobacterota_B, Myxococcota, and Krumholzibacteriota). Lokiarchaeales co-occurs all the same phyla as Thorarchaeia with additional Desulfobacterota_B and Myxococcota_A. Thorarchaeia also co-occurs with

Alphaproteobacterial orders Rhizobiales, UBA8366, Kiloniellales and Gammaproteobacterial orders GCA-272945, HK1, and Woeseiales, Burkholderiales, GCA-001735895, Pseudomonadales, and UBA9214. We also note archaeal (Aenigmataarchaeota, Hydrothermarchaeota, Nanoarchaeota, and Thermoplasmatota) and bacterial (Acidobacteriota, Armatimonadota, Caldritrichota, CG2-30-53-67, Chloroflexota, Dependitiae, Desulfobacterota, Gemmatimonadota, JADFOP01, JAUYMQ01, Marinisomatota, Myxococcota, Nitrospirota, Omnitrophota, Patescibacteria, Planctomycetota, Pseudomonadota, Spirochaetota, and Zixibacteria) lineages that have potential positive co-occurrence ($p_{gt} < 0.05$) with both Hodarchaeales and Kariarchaeaceae and bacterial phyla. This more accurately represents co-occurrence at this shallow coastal site.

In analyzing the Guaymas Basin samples we included both sequencing expeditions. In these samples, only three Asgardarchaeota lineages had a probability of co-occurrence greater than 0.5. Heimdallarchaeaceae co-occurred with bacterial phyla Acidobacteriota, Bipolaricaulota, and Planctomycetota. Gerdarchaeales co-occurred with those lineages and Chloroflexota and Desulfobacterota, as well as the archaeal phylum Thermoproteota. Lokiarchaeales co-occurred with all the same lineages as Heimdallarchaeaceae plus Bacteroidota, Caldiseriicota, and Zixibacteria. Hodarchaeales had potential positive co-occurrence ($p_{gt} < 0.05$) with archaeal (Nanoarchaeota and Thermoplasmatota, and Lokiarchaeales) and bacterial phyla (Abyssobacteria, Acidobacteriota, Bacillota, CG02, Chlamydiota, Chloroflexota, Desulfobacterota, JAGLTZ01, Myxococcota, Omnitrophota, Patescibacteria, Pseudomonadota, and Zixibacteria). However, we did not identify any lineages with positive co-occurrence ($p_{gt} < 0.05$) for Kariarchaeaceae.

Further, we used Sandpiper v0.3.0³⁷ to determine abundance estimates in water columns versus sediment estimates for Heimdallarchaeia lineages. We manually checked the 1687 SRA accessions to obtain the correct sampling location since sampling was often listed as metagenomic data or marine system without specifying the actual source. We found higher mean relative abundance for Kariarchaeaceae within the water column and higher mean relative abundance of Hodarchaeales in sediments and soils (Supplementary Table 10). However, we note the marine system is dynamic, with the water column (e.g., dead zones) and sediments (e.g., subject to bioturbation and physical reworking) having variable oxygen conditions.

Therefore, we scanned all the available metadata for the SRAs with Heimdallarchaeia sequences based on our Sandpiper results. Then, we searched for any information on oxygen via keywords, including oxygen, diss_oxy, oxy_sat, anaerobe, aerobe, microaerophile, and oxygen_requirement. Based on this, we found 4 samples containing Hodarchaeales with aerobic conditions (listed as aerobic, oxygen_mass > 6.5, DO > 2.0 mg/L, DO > 60.98 $\mu\text{mol/kg}$, DO > 1.4 mL/L, DO > 62.5 $\mu\text{mol/L}$, Oxygen saturation > 90 pct), 2 with microaerobic conditions (listed as microaerophilic,

oxygen_mass $0.8 \leq x < 6.5$ mg/L, DO $0.3 \leq x < 2.0$ mg/L, DO $9.15 \leq x < 60.98$ $\mu\text{mol/kg}$, DO $0.21 \leq x < 1.4$ mg/L), and 26 samples with anaerobic conditions (listed as anaerobic, oxygen_mass < 0.8 mg/L, DO < 0.3 mg/L, DO < 9.15 $\mu\text{mol/kg}$, DO < 0.21 mg/L). In cases where data points differed, we took the author's provided relationship to oxygen. Unfortunately, oxygen concentrations were only gathered for sediment or microbial mat samples. No oxygen concentration was reported for the 44 water samples containing Hodarchaeales.

We found 114 samples containing Kariarchaeaceae with aerobic conditions, 8 with microaerobic conditions, 10 samples with anaerobic conditions, and 5 in oxygen minimum zone (no reported oxygen concentration). Thus, supporting metabolic flexibility of Hodarchaeales as a facultative anaerobe or at the very least containing the capacity to tolerate oxygen. However, this additional data further supports that Kariarchaeaceae likely have an aerobic lifestyle. We acknowledge results from both these added analyses are reliant on correct information being submitted to NCBI (Supplementary Table 10).

Phylogenomic analyses of the expanded Asgardarchaeota genomic catalog

Discrepancies within the branching order

We noticed discrepancies in the branching order of the lineages among the resulting phylogenies between non-recoded and recoded alignments, full and subset datasets, and within this expanded catalog in comparison to Eme et al., 2023⁷. Such discrepancies may stem from several factors. The taxon sampling in this study differs significantly, including three additional high-ranking lineages. However, the increased size of the dataset makes it impractical to apply more sophisticated evolutionary models in order to more robustly infer deeper nodes. Despite the computational burden, we were able to test the entire dataset using a site-heterogeneous model with 10 compositional categories (C10) (Fig. 1a; see methods). The resulting phylogeny provides an overview of the entire expanded genomic catalog. To address potential sources of error stemming from compositional bias and/or saturation, we first applied four-character state recoding (SR4) to the full dataset. We then dereplicated the dataset, which allowed us to apply an even more complex site-heterogeneous model (C60).

The observed incongruence between the two species trees of the expanded dataset (un-recoded; Fig. 1a & recoded; Fig. 1b) likely results from the application of an inadequate evolutionary model (C10) that fails to properly accommodate artifacts stemming from compositional bias or mutational saturation. Such artifacts are mitigated when applying recoding but indeed the bootstrap support of inner nodes remains low possibly because of decreased phylogenetic signal. Yet, the well-defined classes, including Lokiarchaeia (Lokiarchaeales and Helarchaeales) and the Heimdallarchaeia orders remain monophyletic.

We dereplicated the dataset to include species-level representatives of the Asgard ingroup. In the species tree inferred from the dereplicated dataset, deeper nodes show increased bootstrap support when applying a more complex mixture model (C60). Such models are better suited to model the heterogeneity of the amino acid substitution process across sites compared to site-homogeneous models^{38,39}. This is now computationally feasible due to the reduced size of the dataset, which still captures the representative diversity of the Asgard clades. The application of more sophisticated models and/or treatments consistently recovers Heimdallarchaeia as monophyletic (Fig. 1b and Supplementary Fig. 3). Although the selected mixture model (C60) may not fully address compositional heterogeneity, Bayesian inference, as demonstrated previously⁷, could be more effective. However, Bayesian inference is impractical for this expanded dataset due to its excessive size.

Based on both the recoded (Fig. 1b) and dereplicated phylogenies (Supplementary Fig. 3), we support the previously suggested monophyly of Heimdallarchaeia and Lokiarchaeia^{7,16}. These clades have been recovered with more robust Bayesian inference methods and now with this expanded dataset. Deep branching lineages, such as, Atabeyarchaeia, Freyarchaeia, Hermodarchaeia, Odinararchaeia, and Ranarchaeia remain consistent as class-level lineages, although their exact placement within the Asgardarchaeota phylum remains unresolved between analyses. Yet, we do recover similar topology for at least 2 of the 3 presented phylogenies for each of the 17 defined Asgardarchaeota lineages. For example, Atabeyarchaeia, Freyarchaeia, Odinararchaeia, and Ranarchaeia group together in both the full un-recoded and dereplicated datasets. The more complex and realistic model of evolution used in Supplementary Fig. 3 likely mitigated the impact of predicted thermophiles within this group⁷. Hermodarchaeia similarly clustered with Thorarchaeia in two phylogenies, which supports its previously predicted position. Future analyses will continue to refine the positions of these deep-branching lineages⁷.

Phylogenetic diversity

We assessed phylogenetic diversity (PD) to quantify the contribution of the genomes added by this study. This analysis calculated PD using the phylogeny shown in Fig. 1a, both across all Asgardarchaeota and specifically within the Hodarchaeales and Kariarchaeaceae lineages.

For the analysis in Fig. 1a, we pruned the phylogeny to produce two trees. The first tree retained all the Asgardarchaeota MAGs, removing the outgroup. The second tree kept only the publicly available Asgardarchaeota genomes. PD was calculated as the sum of all branch lengths in each tree. We determined PD gain by subtracting the PD of the public-only tree from the PD of the full Asgardarchaeota tree. For percentage of PD gain, we divided the PD gain by the PD of the public tree and multiplied by 100. This

analysis was also applied to a pruned phylogeny of the Hodarchaeales and Kariarchaeaceae lineages.

The genomes added in this study resulted in a 27.46% increase in overall Asgardarchaeota PD, a 44.43% increase within Hodarchaeales, and a 20.30% within Kariarchaeaceae. In addition to identifying a new Asgardarchaeota class, Ranarchaeia, and providing the first comprehensive metabolic description of Asgardarchaeia, these results further support the evolutionary distinctiveness of the genomes added by this study, particularly a substantial expansion in Hodarchaeales diversity.

Metabolism of new classes

Ranarchaeia

Here we provide the first description of the novel lineage Candidatus *Ranarchaeia* and type species *Ranarchaeum aquilius* represented by MAG, D4993_C5_H3_Bin_486^{TS}. We have identified seven genome representatives supported across several phylogenetic marker sets. Through our sampling and binning, we identified four MAGs from the Guaymas Basin (GE3) and two MAGs from the Bohai Sea (BH). In searching, through public datasets we also identified another MAG, clustering with this group from cold seep sediment⁴⁰. *Ranarchaeia* has been extracted from sediment 17-2306 mbsl and reported at least 0-44 cm deep. We also predicted ranges for temperature (32.10-54.69°C), salinity (1.27-6.74% w/v NaCl), and pH (5.48-9.08), using amino acid modeling (Supplementary Table 4)⁴¹.

We distinguished the characteristics of the lineage between the shallow coastal and bathypelagic zones of the ocean. Based on our annotations, we predict *Ranarchaeia* are acetogenic mixotrophs (Supplementary Fig. 68 and Supplementary Table 8). *Ranarchaeia* code for genes predicted to have four extracellular peptidases subfamilies (S12; serine, T03; threonine, M28E and M28F; metallo) for the import of amino acids complemented by external glycoside hydrolase CAZymes (Carbohydrate-Active enZymes), GH5 and potentially GH172 predicted to have multiple cellular locations to break down polypeptides and complex carbohydrates (Supplementary Table 8)^{42,43}. Interestingly, we were not able to annotate any coastal transporters (likely due to limited MAGs) but found deep-sea transport of tungstate, nucleosides, phosphate, and cobalt/nickel (missing CbiN). Genomes from both locations can breakdown glucose (C6) to glyceraldehyde-3-phosphate (GAP) with coastal genomes encoding phosphorylating glyceraldehyde 3-phosphate dehydrogenase to produce fructose 1,6-bisphosphate (F 1,6BP). All *Ranarchaeia* lack the capacity to convert F 1,6BP to 1,3-bisphosphoglycerate (1,3BPG) and instead have less common ability among Asgardarchaeota to use the semi-phosphorylative Entner-Doudoroff Pathway (sp-ED), producing GAP and pyruvate from 2-keto-3-deoxy-6-phosphogluconate (KDPG). Based on recent analysis it is expected this pathway arose later in eukaryotic evolution⁴⁴. Therefore, *Ranarchaeia* could be a missing

link from LECA reconstructions or obtained this pathway more recently. Either provides an interesting view of the evolution of carbon metabolism in these archaea. The encoded RuMP (Ribulose Monophosphate Pathway) and NO-PPP (Non-Oxidative Pentose Phosphate Pathway) provide alternative pathways from glucose metabolism, producing and consuming glycolytic intermediates (F6P and GAP). Similar to other Asgardarchaeota, Ranarchaeia has a partial TCA cycle (pTCA) and likely reversibly generates acetate via acetate---CoA ligase (AcdAB) from pyruvate, resulting in energy conservation via substrate-level phosphorylation⁷⁻⁹.

Ability to use C1 compounds via WLP is thus far limited to the bathypelagic representatives, which could be due to niche specialization or limited sampling of coastal genomes. The deep-sea MAGs encode the H₄MPT methyl and carbonyl branches of the Wood-Ljungdahl Pathway (WLP), converting CO₂ to Acetyl-CoA. Unlike Asgardarchaeia, Ranarchaeia encode acetyl-CoA synthase (CdhC) and CODH/ACS complex subunit gamma/delta (CdhDE) to produce acetyl-CoA from carbon monoxide and CH₃-H₄MPT, connecting the carbonyl and H₄MPT branches, respectively, to glycolysis and acetogenesis⁸. The coastal genomes only encode NADH methylenetetrahydrofolate reductase (MetF), methylenetetrahydromethanopterin dehydrogenase (Mtd), and 5,10-methylenetetrahydromethanopterin reductase (Mer), but no complete branches of the WLP. Interestingly, genomes from both locations (6 MAGs) encode a RuBisCO (ribulose biphosphate carboxylase oxygenase)-like long-chain without an identifiable short chain fixing CO₂ to 3-phosphoglycerate (3PG) in the AMP Salvage Pathway⁴⁵. Thus, we predict this novel class across environments could use both external organic (glucose) and inorganic (CO₂) carbon sources to produce acetate.

Ranarchaeia expands our understanding of anaerobic methylotrophy in Asgardarchaeota and the potential for the breakdown of quaternary amines. Across the seven MAGs, we identified several methyltransferases (MttB (5), MtmB (4), and MgsC (2)), corrinoid protein (MtbC1; COG5012 (3)), and a subunit (MtrH; COG1962 (6)) from the methyl-coenzyme M reductase (MCR) complex. Similar to Asgardarchaeia, Atabayarchaeia, and Frey/Jordarchaeia, this group lacks a complete MCR complex and pyrrolysyl-tRNA synthetase^{8,9}. Therefore, we predict the encoded *mttB* potentially transfers methyl groups from glycine betaine or L-carnitine (quaternary amines) to tetrahydromethanopterin (H₄MPT) in the WLP producing acetyl-CoA and acetate⁴⁶⁻⁴⁸ (Supplementary Fig. 68).

Ranarchaeia encodes known energy-converting cytoplasmic [NiFe]-hydrogenases (Groups 3b and 3c), NADH:ubiquinone oxidoreductase subunits, and a V/A-type ATPase. Hydrogenases likely assist in establishing redox equilibrium and the flanking NADH:ubiquinone oxidoreductase subunits likely help establish an ion gradient used by the ATPase for energy conservation (Supplementary Table 8). The suite of hydrogenases may enable lithoautotrophic growth under certain conditions.

Asgardarchaeia

The metabolism of this class has been limited to one MAG⁸. So here, we expand this description, including ten genomes from the Gulf of California in two hydrothermally-associated basins: 7 Guaymas Basin and 3 Pescadero Basin MAGs. This lineage may be endemic to this ecosystem. These genomes were extracted from sediment 0-33 cm deep from 2,000-3,675.5 mbsl. Our analysis supports this organism as an acetogenic chemoheterotroph. *Asgardarchaeia* code for genes predicted to be extracellular peptidases in the cysteine, metallo, asparagine, and serine proteolytic families for the import of amino acids. Our analysis expands the previous description identifying an external CAZyme (glycoside hydrolase family 15), providing potential of the deep-sea MAG (D4993_C5_LG_Bin_231) to break down larger complex carbohydrates^{42,43}. The added MAGs enabled us to identify additional transporters, including nucleoside, cobalt/nickel (missing CbiN) in addition to previously identified tungstate/molybdate and phosphate. *Asgardarchaeia* breaks down C5 and C6 (ribose and glucose) molecules to form pyruvate via EMP with additional substrates produced through RuMP but lack transketolase for NO-PPP. Like the previous description of this group, these genomes encode a partial TCA cycle (pTCA)⁷⁻⁹. We also add potential for fatty acid anabolism and catabolism through beta oxidation. This group can use acetate via acetyl-CoA synthetase and reversibly generate acetate with acetate---CoA ligase (AcdAB) from pyruvate. Acetate production results in energy conservation via substrate-level phosphorylation.

These genomes use C1 organic compounds, including formaldehyde and formate via RuMP (Hps-Phi and HxlA) and Wood-Ljungdahl Pathway H₄MPT methyl branch (WLP; anaerobic carbon-monoxide dehydrogenase). Additional MAGs did not suggest these microbes encode any full branches of the WLP. However, we were able to better characterize the [NiFe]-hydrogenase groups, which now include 3b, 4a, and 4d along with several novel groups (3k, 4k, 4l, 4q), see discussion of novel [NiFe]-hydrogenases for more details. We support the previous prediction that this organism relies on external organic carbon sources for assimilation and hydrogenases may enable lithoautotrophic growth⁸.

Guild metabolism

Guild 11 and 12

Here we provide a more detailed description of the metabolism provided for Hodarchaeales (guild 11 and 12). Notably, we found that the Hodarchaeales in guild 11 encode specific enzymes like Delta3-Delta2-enoyl-CoA isomerase (ECI1, DCI), acyl-CoA:acyl-CoA alkyltransferase (OleA), olefin beta-lactone synthetase (OleC), and a gliding motility-associated binding protein (GldA). ECI is involved in auxiliary fatty acid metabolism, lipid modification, and energy production and was previously described only in bacteria and eukaryotic organelles (peroxisomes, mitochondria, glyoxysomes)⁴⁹.

Hodarchaeales in guild 11 have the entire OleABCD head-to-head hydrocarbon biosynthesis pathway, which has only been found in bacteria as a mechanism for fatty acid metabolism in lipid-rich environments⁵⁰. We also discovered increased ribosomal-associated proteins within guild 11 (Mtr4) and in guilds 11 and 12 (L28e and Bacteroidetes VLRf1). bVLRf1 in bacteria and a few identified divergent archaea is predicted to provide additional protection from transcription and post-transcriptional processing in translation^{7,51}.

Hodarchaeales

Hodarchaeales is an order of Heimdallarchaeia, the Asgardarchaeota lineage currently most closely related to eukaryotes, but all previous analysis has been limited to 23 Hodarchaeales MAGs, but our analysis adds 49 additional genomes⁷. We split the MAGs into four groups by source location (coastal and bathypelagic) and metabolic guilds (11 and 12) to provide an in-depth metabolic description (Supplementary Fig. 36). Discussion regarding impact of the archaeal host has focused on information processing and cytoskeletal development with less emphasis on metabolism. Our analysis uncovered several unique traits conserved between environment and guild, which we use to support previous ancestral reconstructions and recent description of the origins of central carbon metabolism (CCM) in the LECA⁴⁴.

We predict Hodarchaeales are heterotrophs with ability to break down complex carbohydrates with various CAZymes (Supplementary Table 8), as well as simple sugars (e.g., glucose and ribose)⁴². All groups encoded the ability to breakdown glucose except guild 11 MAGs from the deep-sea, instead entering the EMP glycolysis pathway with G6P or F6P. The other genomes encoded a variety of enzymes to degrade glucose (UDPglucose 6-dehydrogenase (UGDH), and ADP-dependent phosphofructokinase/glucokinase (PfkC). With the EMP, only guild 11 coastal genomes encode glucokinase (Glk), non-archaeal-specific glucose-6-phosphate isomerase (Pgi), fructose-1,6-bisphosphatase I, and fructose-bisphosphate aldolase / 2-amino-3,7-dideoxy-D-threo-hept-6-ulose synthase for the first four steps. Guild 11 from all environments also encode 6-phosphofructokinase 1 (PfkA), fructose-bisphosphate aldolase, class II (FBA), 2,3-bisphosphoglycerate-dependent phosphoglycerate mutase (PGAM), pyruvate water dikinase (Pps/PpsA), putative phosphotransacetylase (K15024), and acetate kinase (AckA). Only fructose 1,6-bisphosphate aldolase/phosphatase (K01622) from bathypelagic found to be guild 12-specific. Straying away from the Heimdallarchaeia ancestor reconstructions, modern Hodarchaeales have not lost gluconeogenesis. The eukaryotic glycolysis enzymes predicted to originate in Asgardarchaeota include ADP-dependent glucokinase (ADPGK/PfkC), 2,3-bisphosphoglycerate-independent phosphoglycerate mutase (APGM), 2,3-bisphosphoglycerate-independent phosphoglycerate mutase (GPMI), and enolase (ENO). We found ADPGK in all but deep-sea guild 11, APGM and GPMI in all but coastal guild 12, ENO⁴⁴. Acetate is reversible

generated from Acetyl-CoA (ADP-forming acetate---CoA ligase; AcdAB) and utilized via acetyl-CoA synthetase. AcdAB is another enzyme predicted to arise from the Asgard archaeal host. Guild 11 also encodes a putative reversible two-step process (phosphate acetyltransferase and putative phosphotransacetylase).

The coastal genomes encode a complete TCA cycle with bathypelagic guild 11 missing malate dehydrogenase (Mdh). Only bathypelagic guild 12 encodes the glyoxylate shunt, bypassing decarboxylation (release of CO₂) in the TCA cycle. This pathway conserves carbon intermediates useful in CO₂ limited environments, which may indicate competition in the deep-sea consortium⁵². Predicted eukaryotic TCA components from Asgardarchaeota include Mdh absent in all bathypelagic Hodarchaeales, supporting LAsECA from shallow environments. The other two steps are malate to fumarate (fumarate hydratase) found in all four subsets and potentially acetyl-CoA and oxaloacetate to citrate (ATP-citrate lyase). ATP-citrate lyase subunit alpha is absent in all Hodarchaeales groups, but only guild 11 encodes the beta subunit. If this did stem from the archaeal host, it may have been lost post-eukaryogenesis.

Reconstruction of RuMP, NO-PPP, O-PPP (Oxidative Pentose Phosphate Pathway), and WLP showed both guild and environmentally specific pathways. Only guild 11 encodes the multi-step RuMP pathway, converting ribose (C5) or Ru5P to F6P in coastal and bathypelagic MAGs, respectively. The reversible NO-PPP is found only in coastal guild 11 with unidirectional O-PPP only in guild 12⁵³. The O-PPP enables production of Ru5P, reducing two NADP⁺ to two NADPH, providing additional reducing power to these microbes. Suggesting increased energy conservation in guild 11. Then, genomes with NO-PPP can reversibly rearrange carbon compounds and break down pentose sugars (e.g., ribose). Although lacking the capacity to complete the NO-PPP, ribose 5-phosphate isomerases (RPIA) is found across environment and guild and a predicted host-derived portion of eukaryotic metabolism⁴⁴. Despite the predicted loss of the WLP in the Heimdallarchaeia ancestor, guild 11 encodes both carbonyl and H₄MPT methyl branches. Most of the enzymes in these pathways are found in <20% of the Hodarchaeales groups, so this could have been more recently acquired. These enzymes likely enable these microbes to potentially use inorganic C1 compounds (CO₂) to produce Acetyl-CoA.

We also traced non-CCM pathways, imports/exports, and coenzyme biosynthesis. We identified portions of the shikimate pathway in coastal guild 11 and bathypelagic guild 12, which is an important step in production of chorismate, which is converted into menaquinone (MK), a vital component of electron transport chains. We only found the AMP salvage pathway in coastal guild 11, but RuBisCO in all Hodarchaeales subsets. Beta oxidation for fatty acid catabolism and generation of Acetyl-CoA and Glycine Cleavage System (GCS) for one carbon and amino acid metabolism was found across guild and environment.

Transporters and coenzyme biosynthesis varied between groups, which corresponds to varying environmental conditions and necessary metabolic substrates. Like *Ranarchaeia* and *Asgardarchaeia*, phosphate transport was found only in deep-sea MAGs (guild 12). This could suggest phosphate is a limiting resource for these archaea in shallower marine systems but is more available for transport in the deep-sea⁵⁴. Tungstate import is found in coastal guild 11 and bathypelagic guild 12. In guild 11, we identified multidrug efflux with iron(III), nickel, and potential lipoprotein import found only in coastal MAGs. The influx of iron(III) may be related to its role in electron transport due to the increased complexity of that guild's ETC. Nickel may be connected as an enzymatic cofactor in hydrogenases and as this was the only group to synthesize coenzyme M. Coastal guild 11 has the capacity to synthesize coenzyme F, an electron carrier. Both coenzymes are used in methanogenesis and one-carbon metabolism. Vitamin B12 (vB12), necessary for coenzyme synthesis, is absent in all coastal *Hodarchaeales* genomes. Indicating a potential metabolic shift from vB12 prototrophy to auxotrophy between deep-sea and coastal environments. This adaptation could be influenced by nutrient availability or microbial interactions. vB12 has been suggested as linked in other host-symbiont metabolic interactions⁵⁵⁻⁵⁷. All guilds had subunits to synthesize coenzyme A, import branched chain amino acids and nucleoside and potential to export heme (missing *CcmD*). Heme is essential for cytochromes and aerobic respiration can increase intracellular oxidative stress, interacting with iron to produce ROS. Therefore, heme export is important to reduce the toxic impact of ROS and has been shown to increase interaction between exporting hosts and scavenging microbes⁵⁸. This could be another potential link between *Hodarchaeales* and other microbes.

We further identify proteins involved in chemotaxis, including CheWARY in coastal guild 11. A few components are also found in bathypelagic guild 11 (CheWRY) and guild 12 (CheW). These were identified with a methyl-accepting chemotaxis protein (K03406). The methyl-accepting chemotaxis proteins were also identified in the HADEG (Hydrocarbon Aerobic Degradation Enzymes and Genes) database results for potential involvement in alkane metabolism in chemotaxis complex^{59,60} (Supplementary Table 8). Further, through our analysis we also identified potential Lux proteins (PF00296, PF05893, and PF00196) known for their involvement in quorum signaling.

As discussed in the main text, *Kariarchaeaceae* representatives also occur in guild 12. Metabolic reconstruction indicates that bathypelagic and shallow coastal *Kariarchaeaceae* MAGs share broadly similar metabolic potential within guild 12, but only the shallow coastal MAGs encode a complete beta-oxidation pathway and the capacity to use ribose as a carbon source through the NO-PPP.

Evolutionary and ecological transitions in metabolic traits

To better compare the metabolic capabilities of *Asgard* archaea with other organisms, we then performed a comparative genomic analysis on 959 single-cell

eukaryotes, 438 Alphaproteobacteria MAGs, and the genomes from each of the 17 Asgardarchaeota taxonomic groups to identify orthologous groups (OGs, from eggNOG database) of genes (Supplementary Table 6). Asgardarchaeota and Alphaproteobacteria share 1,661 and 2,502 OGs with eukaryotes, respectively (Supplementary Fig. 7 and Supplementary Table 6). Our analysis did not include all prokaryotes, so we cannot exclude other donors, but it delineates eukaryotic OGs in each Asgardarchaeota lineage that are absent in Alphaproteobacteria (Supplementary Fig. 6 and Supplementary Table 6). Our comparative genomic approach identified the most functionally distinct orthologs in Lokiarchaeia, Thorarchaeia, and Heimdallarchaeia (Supplementary Fig. 6 and see methods). A large number of uniquely conserved genes are present in Lokiarchaeia (689) and Thorarchaeia (210), which suggests their metabolic diversity as the clades with the largest number of MAGs from various environments (Supplementary Figs. 6-10). The five Heimdallarchaeia lineages, Hodarchaeales (145 OGs), Kariarchaeaceae (98 OGs), Gerdarchaeales (52 OGs), Heimdallarchaeaceae (49 OGs), and Njordarchaeales (20 OGs), had the next largest number of lineage-specific orthologs, which are involved in a wide variety of cellular processes, including energy, amino acid metabolism, DNA maintenance, membrane biogenesis, post-translational modification, and signal transduction (Supplementary Table 6). This genome versatility is also reflected in the average predicted genome size (see methods) with Lokiarchaeia (guild 1-3), Hodarchaeales (guild 11 & 12), and Kariarchaeaceae (guild 12) guilds being greater than 4.0 Mb with a statistically significant (one-sided t-test, p value= 2.37×10^{-16}) increase in genome size from bathypelagic to coastal environments (Supplementary Figs. 11-14, and Supplementary Table 7)⁶¹. These data suggest that Heimdallarchaeia are more metabolically versatile than other Asgardarchaeota.

Key metabolic proteins

Divergent [FeFe]-hydrogenases

The production and use of hydrogen in the progenitors of the eukaryotic cell has been linked as a potential shared metabolic intermediate⁶²⁻⁶⁴. H₂ is a pivotal electron sink for central metabolic processes, an energy source in microbial communities, and an invaluable resource for biotechnology applications⁶⁵. Despite the diversity of [NiFe]-hydrogenases identified in archaea, the discovery of archaeal [FeFe]-hydrogenases has been limited to Woesarchaeota⁶⁶. Recently an in-depth analysis identified [FeFe]-hydrogenases in nine archaeal phyla and fused [NiFe] and [FeFe]-hydrogenases in an additional ten archaeal orders⁶⁷. Identification [FeFe]-hydrogenases with verified production of H₂ would provide support to eukaryogenesis models, suggesting the archaeal host provided hydrogen to their bacterial partners.

We searched for [FeFe]-hydrogenases within all the Asgardarchaeota MAGs (Supplementary Fig. 67). We identified Asgardarchaeota hydrogenases in Groups A, B,

and a divergent megaclade of Asgard-only hydrogenases temporarily called G. There were two MAGs from cold seep sediment (Lokiarchaeales and Thorarchaeia) and one MAG from hot springs (Kariarchaeaceae), encoding A3 clade hydrogenases. This clade assists in internal redox balance reversibly bifurcating electrons from hydrogen, reducing or oxidizing NAD⁺ and ferredoxin. Within Group A, there is another divergent clade of Asgard-specific hydrogenases preliminarily labeled here as A5 that will require additional structural and domain organization analyses to confirm as a subgroup. This divergent clade has hits from Guaymas Basin sediment, including one Asgardarchaeia and two Njordarchaeales MAGs with 14 total Lokiarchaeales hydrogenases split between hydrothermal vent sediment (10), cold seep sediment (2), a biogas reactor (1), and lake sediment (1). Interestingly, we identified additional Group A [FeFe]-hydrogenases, flanking some of the NuoEFG subunits.

We identified 39 Lokiarchaeales Group B [FeFe]-hydrogenases. As Group B remains functionally undefined, this added diversity could help identify the role of these enzymes across environments, including hydrothermal vent rock/sediment (24), biogas reactor (1), cold seep sediment (8), freshwater spring sediment (2), coastal microbial mat/sediment (2), and permafrost subsurface sediment (1). Despite the geographic range of these Lokiarchaeales, their hydrogenases cluster separately from all other lineages except a single *Holophaga foetida* (Acidobacteria WP_026939128.1) hydrogenase⁶⁸. This protein, however, is now suppressed from the NCBI record and removed from the genomes, potentially due to contamination from Lokiarchaeales.

Within the Asgard-specific clade, we identified 360 Asgardarchaeota hydrogenases with one bacterial hydrogenase (WP_015709814) that has also been recently suppressed. This clade included genomes from Heimdallarchaeia (Gerdarchaeales, Heimdallarchaeaceae, Hodarchaeales, and Kariarchaeaceae), Helarchaeales, Hermodarchaeia, Lokiarchaeales, Sifarchaeia, and Thorarchaeia. The novel clade contains 15 [FeFe]-hydrogenases from Hodarchaeales, the closest Asgardarchaeota lineage to eukaryotes. These hydrogenases are found in 10 coastal and two hydrothermal vent sediment MAGs. Further, three Hodarchaeales MAGs have two copies of these potential [FeFe]-hydrogenases. Additional analysis is necessary to determine the function of these enzymes, especially the Heimdallarchaeia [FeFe]-hydrogenases. This study identifies the first [FeFe]-hydrogenases from Hodarchaeales. Like the other [FeFe]-clades, there is an extensive environmental range from the permafrost to hot spring, from groundwater and inland water, from epipelagic to bathypelagic zones of the ocean, and from both sediment and the water column. Our analysis increases the number of archaeal lineages with [FeFe]-hydrogenases, identifying these enzymes within 6 classes and 11 named Asgardarchaeota lineages, providing potential targets for long-term green energy applications and evidence of the importance of hydrogenases in the modern-day descendants of the archaeal host. Yet there is no support at this time that these novel Hodarchaeales [FeFe]-hydrogenases produce H₂.

Using the HydDB, the proteins from the Asgard-specific clade classify as both [NiFe] and [FeFe] clades. The similarity of the catalytic subunit to various histidine kinases and response regulators, as well as clustering with group C2 indicates a potential role as H₂-sensors in regulatory cascades. Many of the enzymes are fused to a PAS domain similar to those [FeFe]-hydrogenases in Group C. The recently accepted manuscript describing novel archaeal [FeFe]-hydrogenase clades is not yet available, but based on our analysis it appears the domain structure and the branching of this Asgard-specific clade does not resemble the newly identified Groups D-F^{67,69}. Instead, tentative Group G is a sister-clade to [FeFe] Clade C2. The identity of these Asgard hydrogenases will need to be verified upon the publication of Greening et al., alignments. These Hodarchaeales [FeFe]-hydrogenases may provide yet another link to eukaryogenesis with several classifying as both [NiFe]-Group 4 with homology to Complex I subunit (NuoD) and [FeFe] Clade C2 with potential ability to sense hydrogen.

Non-methanogenic methylotrophy

Recently several archaeal lineages have been proposed to be non-methanogenic methylotrophs, including two TACK lineages (Brockarchaeia⁷⁰ and Culexarchaeia⁷¹) and four Asgardarchaeota lineages (Asgardarchaeia⁸, Atabeyarchaeia, Frey/Jordarchaeia⁹, and Sifarchaeia^{72,73}). This previously overlooked metabolism has been implicated to have roles in production and consumption of substrates necessary for methane production, potentially modulating methanogenesis⁹.

To determine how widespread this metabolism is in Asgardarchaeota, we determined the presence of methanogenic and methylotrophy-associated proteins. Phylogenomic analyses of trimethylamine methyltransferase subunit, MttB (COG5598), found sequences in all 17 described Asgardarchaeota lineages (Supplementary Fig. 69). Besides previously described clades, we identified several corrinoid proteins: MtaA in coastal Lokiarchaeales (guild 1) and Sifarchaeia (guild 7) and MtaC in coastal Wukongarchaeia (guild 7). We identified two subunits (MtrA and MtrH) of the methyl-coenzyme M reductase (MCR) complex. MtrAH were found in guild 7, Thorarchaeia guild 4 and 6, and deep-sea Lokiarchaeales guild 1. We identified these subunits across inland (Atabeyarchaeia, Frey/Jordarchaeia, Sifarchaeia, and Thorarchaeia), coastal (Sifarchaeia and Thorarchaeia), mesopelagic (Thorarchaeia), bathypelagic (Frey/Jordarchaeia, Sifarchaeia, Thorarchaeia and Lokiarchaeales), hot spring (Frey/Jordarchaeia and Thorarchaeia), and permafrost (Thorarchaeia) environments. We also found MtrA in Hodarchaeales guilds 11 (coastal) and 12 (bathypelagic) and guild 8 Njordarchaeales (bathypelagic). MtrH was found in coastal and bathypelagic Baldrarchaeia (guild 7), bathypelagic, coastal, and inland Helarchaeales (guild 2), coastal Hermodarchaeia (guild 5), bathypelagic Lokiarchaeales (guild 2), coastal Lokiarchaeales (guild 1 and 2), mesopelagic Sifarchaeia (guild 7), and coastal Wukongarchaeia (guild 7).

Helarchaeales also contain the previously described divergent *mcr*-like gene cluster for short-chain alkane metabolism⁷⁴.

It is likely the encoded MttB transfers methyl groups from quaternary amines (glycine betaine or L-carnitine) to substrates in the WLP⁴⁶. This pathway has been previously proposed for MttB-encoding genomes in the absence of a complete MCR and pyrrolysyl-tRNA synthetase (IPR023877). Yet we did identify seven epi-bathypelagic and inland Sifarchaeia (guild 7) MAGs encoding pyrrolysyl-tRNA synthetase, which might change use of the amines pool in those genomes. Our analysis indicates potential for this metabolism across Asgardarchaeota lineages, guilds, and environments. Thus, increasing our understanding of non-methanogenic methylotrophic metabolism in archaea (adding seven Asgardarchaeota lineages and the first Heimdallarchaeia) and expands this methanogenesis-modulating metabolism to new environments (e.g., permafrost).

Oxygen metabolism and ROS tolerance

Ribulose-1,5-bisphosphate carboxylase/oxygenase family: Our phylogenomic analysis supports previous identification of archaeal type III and type IV Ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) in Asgardarchaeota^{7,63,75} (Supplementary Fig. 70). We also identified RuBisCO sequences from 10 Lokiarchaeales MAGs branching with type II references, so potentially they encode a different type of the family. RuBisCO was found in all 17 Asgard archaea lineages across guilds and environments, expanding the known presence of this enzyme. However, as previously suggested, archaeal type III and type IV are not predicted to be involved in carbon fixation via the Calvin-Benson Bassham (CBB) cycle. Instead, these types are involved in assimilating salvaged nucleosides into central carbon metabolism. The AMP salvage pathway converts adenosine monophosphate (AMP) to Ribose 1,5-bisphosphate (R5P) and then to Ribulose 1,5-bisphosphate (Ru5P). In this process, RuBisCO can convert Ru5P to 3PG, connecting the AMP salvage pathway to EMP (Supplementary Fig. 36 and 68 step 130).

Superoxide dismutase: Superoxide dismutase (SOD) provides protection from oxidative damage, including byproducts of aerobic respiration. This enzyme converts superoxide to hydrogen peroxide and oxygen through oxidation and reduction of metal ions. We identified SOD in 58 Asgardarchaeota genomes from 8 taxonomic groups (Atabeyarchaeia, Frey/Jordarchaeia, Hodarchaeales, Kariarchaeaceae, Sifarchaeia, Lokiarchaeales, Odinararchaeia, and Thorarchaeia) (Supplementary Fig. 62 and Supplementary Table 8). These enzymes spanned three different families with different catalytic metals, including Ni, CuZn, and FeMn (sometimes described separately as Fe and Mn)⁷⁶. Atabeyarchaeia, Frey/Jordarchaeia, Hodarchaeales, Odinararchaeia, Sifarchaeia, and Thorarchaeia also encode SODs from the most ancestral family, FeMnSOD. This

family has been previously described in lineages including Sulfolobales, Atabeyarchaeia, and Frey/Jordarchaeia, supporting a more widespread distribution in archaea of this enzyme^{9,77}. Of these groups only Hodarchaeales, Kariarchaeaceae, Lokiarchaeales, and Thorarchaeia also encode CuZnSOD predicted to have evolved later than the other forms and previously unidentified in archaea⁷⁸. Only guild 11 and 12 (Hodarchaeales and Kariarchaeaceae) encode NiSOD. Previous work has connected the evolution of NiSOD to the oxygenation of the world's oceans (GOE) as Fe levels decreased, which is necessary to form FeMnSOD⁷⁸. Further connecting Hodarchaeales and Kariarchaeaceae to increasing tolerance of the toxic byproducts of aerobic respiration and evolution of enzymes during the GOE.

Catalase: Catalases also provide protection against ROS, breaking down hydrogen peroxide to water and oxygen. We identified 116 Asgardarchaeota MAGs encoding ancestral clade 1 and clade 3 catalases (Supplementary Fig. 63 and Supplementary Table 8). The ancestral clade 1 catalases were previously described for Atabeyarchaeia and Frey/Jordarchaeia⁹. However, we also identified clade 3 catalases in Heimdallarchaeaceae, Hermodarchaeia, Helarchaeales, Hodarchaeales, Kariarchaeaceae, Lokiarchaeales, Odinararchaeia, and Thorarchaeia. Clade 3 catalases are known to bind heme b and NADH found across all domains^{79,80}.

Peroxidase: Our analysis also identified three types of peroxidases, including non-heme peroxidase, glutathione peroxidase, and catalase-peroxidase. Non-heme (chloroperoxidase) peroxidases were found in Baldrarchaeia, Frey/Jordarchaeia, Helarchaeales, Sifarchaeia, and Thorarchaeia all from guild 6 or 7. These haloperoxidases incorporate halogen atoms into organic substrates, using hydrogen peroxide as oxidizing agent, releasing water. Other families include glutathione peroxidase found in Hodarchaeales guild 12 and one Thorarchaeia guild 6 MAG. Catalase-peroxidases were identified in Frey/Jordarchaeia, Hermodarchaeia, Hodarchaeales, Helarchaeales, Kariarchaeaceae, Lokiarchaeales, and Odinararchaeia. This increases the suite of enzymes to mitigate the toxic impacts of aerobic and photosynthetic metabolic byproducts.

Identified ESPs and eukaryotic characteristics

recA/RAD51 gene family: In preparing the genomic catalog, several duplications of *recA* were flagged as contamination within the genomes by CheckM⁸¹. After manual inspection, it was identified that most Asgardarchaeota genomes contain several copies of *recA*, a recombinase functioning in recombination, DNA repair, and genome stability (Supplementary Table 8). It is a bacterial single copy marker gene in many quality control analyses for prokaryotes^{81,82}. Archaea have been previously described to have 2-3 types (RadA, RadB, RadAB) with eukaryotes having several copies predicted to be from

endosymbiotic gene transfer and several duplication events^{82,83}. Yet, we annotated more than 8 copies in several Asgardarchaeota genomes, including MAGs from Hodarchaeales and Kariarchaeaceae. This required more extensive manual cleaning of the MAGs during the quality control stages to ensure these sequences were not contaminated despite higher-than-expected redundancy scores. Archaeal and eukaryotic types are predicted to be a product of bacterial *recA* duplications, but the identification of numerous types in Asgardarchaeota provide an interesting link to genome stability and maintenance given increasing genome size in the groups closest to eukaryotes^{82,83}.

Hsp90: Heat shock protein 90 is a molecular chaperone that preserves structural integrity of proteins primarily in eukaryotes and bacteria. Through our comparative genomic analysis, we identified this protein (PF00183) in Lokiarchaeales, Gerdarchaeales, Hodarchaeales, Kariarchaeaceae, and Thorarchaeia. The alignment and tree files are found in the data on Figshare.

L28e/MAK16 domain: This ribosomal protein has been previously described in Eme et al., 2023⁷ as a eukaryotic signature protein. We were able to determine in the expanded dataset that this protein is only found in Hodarchaeales. It is found across guilds and environments with 27 total genomes. The alignment and tree files are found in the data on Figshare.

PF10281: This putative nuclear envelope organization protein was found in 43 Lokiarchaeales genomes from guilds 1 and 2 (excluding guild 3), spanning environments and sampling expeditions. Taxonomic distribution appears to be limited to viruses, eukaryotes, bacteria, and Lokiarchaeales. In fungi, this PFAM is a repeat that is used to identify the protein Ish1, which is linked to stress-response in the nuclear envelope^{84,85}. These proteins also share sequence similarity with other smaller regulating cytoskeletal, cytoplasmic, and nuclear proteins, including gelsolin repeat and SAP domain.

Additional eukaryotic-like and metabolic proteins identified in the comparative genomic analysis that support our finding of increasing energy conservation and cellular complexity include Staygreen protein (PF12638), Transmembrane 231 (PF10149), Cytochrome bd-I (PF01654), DNA photolyase (PF00875), Mtr4 (PF13234), and DNA photolyase (PF01654). Supplementary Table 12 provides an overview of previously described ESPs in this expanded database.

Phylogenetic and structural analyses of [NiFe]-hydrogenases

Phylogenetic distribution

Novel [NiFe]-hydrogenase subgroups

In exploring links between Complex I and [NiFe]-hydrogenase Group 4, we aimed to identify and classify Asgardarchaeota [NiFe]-hydrogenases within this expanded dataset. We filtered the predicted Asgard [NiFe]-hydrogenases for the CxxC+200 residues+CxxC motif, resulting in 1143 potential Asgard archaea hydrogenases. These sequences were further filtered with several rounds of alignments and phylogenies, resulting in 1077 Asgard [NiFe]-hydrogenases from Groups 1, 3, and 4. To determine their evolutionary relationship with Complex I, we added the Asgardarchaeota sequences to 1931 previously classified hydrogenases⁶⁵. Our analysis identified several novel [NiFe]-hydrogenase subgroups from Group 1 (1m-n), 3 (3e-l), and 4 (4j-r) more than doubling the number of subgroups in Group 3 and 4 (Supplementary Fig. 30). There were a few megaclades (>90 MAGs), including 3f (142 MAGs), 3k (99 MAGs), 4p (95 MAGs), 4q (128 MAGs). We conservatively kept megaclade 3k as one group, despite higher complex mixture models indicating the potential to split into several smaller novel groups. Additional work is necessary to describe the functional diversity of the novel enzymes and their metal cofactors.

There are two sequences (one from each Ranarchaeia and Njordarchaeales) that have been labeled 3l because the hydrogenases branch around the new 3l clade and do not currently cluster with high bootstrap support within any of the previously described hydrogenases. Additional sequences will hopefully resolve these more divergent hydrogenases. These two sequences may represent another novel subclade that lacks the additional sequences to provide phylogenetic resolution. Further, as mentioned in several previous publications, [NiFe]-hydrogenase Group 4g references cluster in multiple positions as more genetic diversity is added to the phylogeny and likely represent multiple subgroups⁸⁶.

Distribution of novel [NiFe]-hydrogenase subgroups

To further describe these novel subgroups, we compared their phylogenetic, environmental distribution, and metabolic guild distribution within the Asgardarchaeota. Understanding the taxonomic distribution was necessary to detail the evolution of these enzymes and potential impact on eukaryogenesis. All 17 Asgard lineages encoded novel [NiFe]-hydrogenase subgroups, but some subgroups were only identified in a single lineage (1n;Wukongarchaeia, 3i;Hermodarchaeia, 4j;Hodarchaeales, 4k;Asgardarchaeia, 4m;Kariarchaeaceae, 4o;Ranarchaeia, and 4r; Thorarchaeia). Interestingly, 4j appears to be the most ancestral group 4 [NiFe]-hydrogenase subgroup

with only Hodarchaeales sequences (Supplementary Figs. 30-31). However, 4p and 4q have increased numbers of MAGs (95 and 128) and lineages (6 and 11).

Previous work⁸⁶ has illustrated that these [NiFe]-hydrogenase groups have conserved catalytic directions. In the Asgardarchaeota genomes, we identified Group 1 (respiratory and hydrogen-oxidizing), 3 (cofactor-coupled and bidirectional), and 4 (energy-converting and typically hydrogen-evolving). Within our phylogenetic analysis (Supplementary Fig. 30), we see clear division between the four main hydrogenase groups (Group 2 not identified in Asgardarchaeota). We identified hydrogen-oxidizing Group 1 in Baldrarchaeia, Hermodarchaeia, Kariarchaeaceae, Njordarchaeales, Sifarchaeia, and Wukongarchaeia. Specifically, Kariarchaeaceae encodes subclade 1d, which has previously been characterized as oxygen-tolerant and used for electron input in aerobic respiration and oxygen-tolerant anaerobic respiration⁸⁷. Group 3 hydrogenases were identified in all Asgardarchaeota lineages, while Group 4 hydrogenases were identified in every Asgardarchaeota lineage except Lokiarchaeales.

However, it should be noted some hydrogenases vary in directionality depending on the physiology of their host. A few are even physiologically reversible and will switch directionality depending on factors such as cellular redox state. Group 3 and 4 [NiFe]-hydrogenases, especially, vary in their directionality. These enzymes often act in a hydrogen-oxidizing direction in autotrophic organisms, for example in methanogens, but instead primarily produce hydrogen in heterotrophic organisms including the cultured Asgard archaea. Overall, it is unlikely that the Asgard archaeal lineages encoding both Group 3 and 4 [NiFe]-hydrogenases require a hydrogen-producing partner but more likely are hydrogen producers. We do not deny hydrogen could be provided by a partner, but that it is not a necessity for these lineages.

Based on the metabolic guilds, we identified greater diversity among guild 7, which includes 8 taxonomic lineages (Asgard, Atabey, Baldr, Frey/Jord, Odin, Ran, Sif, and Wukong) in comparison to least diversity among guild 1 (Loki), guild 3 (Loki), and guild 4 (Thor) (Supplementary Fig. 32). For the group closest to Eukaryotes, Hodarchaeales, we can explore guild 11 (Hod) and guild 12 (Hod and Kari). Both guild 11 and 12 encode groups 3c, 3e, 3g, 3j, 4j, and 4p [NiFe]-hydrogenases, but guild 11 also encodes groups 3b, 3d, 3h and guild 12 encodes groups 1d, 4m, and 4q.

We also looked at the environmental distribution of all the Asgard [NiFe]-hydrogenase subgroups. A few were only identified in hot springs (1d) or coastal (1a, 1n, 3d, 3h, 3i) and bathypelagic zones (4a, 4k, and 4m) with several (3a, 4b, 4d, 4e, 4g, 4o) only identified in the Guaymas Basin sediment (Supplementary Fig. 33). This highlights the enzymatic diversity, especially with the Guaymas Basin sediments. Based on our analysis, it appears MAGs from <200 mbsl had more diversity of Group 3 and those >1000 mbsl had more diversity of Group 4. Group 3b [NiFe]-hydrogenases, which uses NADPH to produce hydrogen gas was typically the most common hydrogenase among the environments except biogas reactor (3f), groundwater well (3f), hot springs (4q), oilfield

(equal 3b, 4l, 4q), and permafrost (3f). Group 3f was typically the second most common among the [NiFe]-hydrogenase subclades.

In looking at the NM47 phylogeny, we can begin to map predicted gains of the various subgroups. At the base of the Asgardarchaeota group (Atabeyarchaeia and Frey/Jordarchaeia), we predict the presence of groups 3b, 3f, and 4q [NiFe]-hydrogenases. We can identify differences in the distribution of [NiFe]-hydrogenases among the groups closest to eukaryotes, Gerdarchaeales (3b, 3e, 3g, 3k, 4b, and 4q), Heimdallarchaeaceae (3b, 3c, 3k, 4p), Hodarchaeales (3b, 3c, 3d, 3e, 3g, 3h, 3j, 4j, 4p, 4q), Kariarchaeaceae (1d, 3c, 3e, 3g, 3j, 4m, 4p), and Njordarchaeales (1m, 3a, 3b, 3k, 3l, 4q). Within the Heimdallarchaeia, Hodarchaeales and Kariarchaeaceae display the greatest diversity of [NiFe]-hydrogenases (Supplementary Fig. 34). Of all these lineages, we identified based on synteny the subgroups most similar to Complex I-like complexes were groups 4m, 4p, and 4q. Within Heimdallarchaeia, group 4m was found in Kariarchaeaceae, guild 12, bathypelagic regions; group 4p was found in Hodarchaeales (guild 11), Kariarchaeaceae (guild 12), and Heimdallarchaeaceae (guild 9) from coastal to bathypelagic regions and in the permafrost; group 4q found in Hodarchaeales, Gerdarchaeales, and Njordarchaeales, guild 8, 10, and 12 across numerous environments (Supplementary Fig. 35). Of those, the largest number of Hodarchaeales MAGs and scaffolds is found in group 4p [NiFe]-hydrogenases from coastal regions.

Structural architecture and putative mechanisms of group 4 [NiFe]-hydrogenase complexes

Complex I-like vs. MBH-like structural architecture

Complex I-like structures can be distinguished from MBH-like structures by the position of the antiporter-like subunits relative to the peripheral arm subunits. For example, when viewing the membrane arm subunits “face-on” so that all antiporter subunits and the helix-HL are visible, the peripheral arm is packed against the right-hand side of the membrane arm in Complex I. Whereas in MBH, the peripheral arm is rotated 180° and packed against the left-hand side of the membrane arm^{26,88}. Of the ten *nuo* respiratory complexes we selected for AF2 modeling, eight have a Complex I-like structural architecture (from Kariarchaeaceae, Gerdarchaeales, Njordarchaeales, Hodarchaeales) and two have a MBH-like structural architecture (from Asgardarchaeia) (Supplementary Fig. 46). We named the eight Complex I-like complexes to track lineage “Kari-1” (GCA_015523565), “Kari-2” (GCA_024280435), “Njord-1” (M288), “Njord-2” (D4998_C1112_H3_Bin_254), “Hod-1” (M3_38_Bin_455), “Hod-2” (GCA_024276395), “Hod-3” (M3_30_Bin_130), and “Gerd” (GCA_016839405). Similarly, the two MBH-like complexes are named “Asgard-1” (AB3033_2), and “Asgard-2” (GCA_021162905).

AlphaFold2 model confidence and limitations

We judged the accuracy of AF2 models by considering their confidence scores as well as their structural similarity with experimentally characterized Complex I-like and MBH-like structures. The average predicted local distance difference test (pLDDT) scores were high for all structures (>80). AF2 appeared to have the most inaccuracy (pLDDT 50 - 70) with predicting the “tips” of the transmembrane helices in the membrane arms (Supplementary Fig. 45). One possibility is that these regions are less evolutionarily conserved, thus reducing the multiple-sequence alignment depth which is important for AF2 accuracy⁸⁹. There are a few regions in the peripheral arm subunits that have pLDDT scores <50, which we believe to be intrinsically disordered regions^{90,91} (Supplementary Fig. 45). The pTM and ipTM scores for each model range from moderate to high in their accuracy (>0.7 and >0.6 respectively). The PAE scores indicate that residues within the peripheral arms align confidently with each other, as do residues within the membrane arms. However, most of our AF2 models suffer moderate PAE scores in the alignment between the membrane arm subunits and the peripheral arm subunits (Supplementary Fig. 45). This could indicate the presence of conformational flexibility between the position of the peripheral arm relative to the membrane arm^{92,93}. Indeed, single-particle cryo-EM analysis of Complex I has observed the peripheral arm shifting in position relative to the membrane arm under different turnover states^{94,95}.

A complete analysis of the membrane-arm subunit functions is not possible with our AF2 models. Inferring the proton transfer pathways in membrane-arm subunits requires information about the placement of water molecules^{94,96–98}, and this information is absent from AF2 predictions. An additional limitation is that some of our AF2 models are likely missing subunits as a result of incomplete input information. For instance, the NuoL subunits from “Gerd” and “Hod-2” have large predicted aligned errors relative to the rest of the predicted multimer (Supplementary Fig. 45a), which may indicate that additional subunits interacting with it were not captured during our annotation step. The predicted aligned error between NuoA and the rest of the multimer in “Hod-1” are also large and could indicate that this gene was misannotated as NuoA and is not truly a part of the complex (Supplementary Fig. 45a). Finally, the AF2 multimer models from Asgardarchaea are missing a complete set of Na⁺/H⁺ antiporter subunits as expected from comparison with MBH (Supplementary Fig. 45b). These missing subunits may be found with improvements to our gene annotation steps. Our AF2 multimer models are primarily starting points for hypothesis generation and require future experimental validation.

Complex I-like peripheral arm subunits

Complex I-like complexes found in Heimdallarchaea have a peripheral arm composed of NuoBCDI subunits but are missing NuoEFG subunits. NuoEFG forms the NADH-oxidizing module (N-module) in typical Complex I and provides electrons to the

NuoBCDI subunits, which forms the quinone reduction module (Q-module). *nuoEFG* genes were not observed in the putative *nuo* gene clusters in Heimdallarchaea, and *nuoEFG* genes that were observed elsewhere in the genome failed to form a multimer with the other *nuo* genes when provided as input to AF3 (Supplementary Fig. 78). Therefore, we suspect that the electrons required by the complex are provided by ferredoxin (Fd). This mechanism of electron input into the complex would be similar to that described in other respiratory complexes such as the archaeal Complex I⁹⁹, MBH²⁶, and photosynthetic Complex I (PSI)¹⁰⁰. Indeed, the “top” surface of NuoI has a predicted electropositive charge, which could allow an electronegatively charged reduced Fd to transiently bind and transfer electrons. A similar Fd docking surface has been observed in PSI’s NdhI subunit, which is homologous to NuoI¹⁰⁰. We modeled FeS clusters into the NuoI and NuoB subunits to better understand the electron transfer path in these complexes (Supplementary Fig. 45a-b). Similar to Complex I, electrons could be transferred through a series of FeS [4Fe-4S] clusters within NuoI and NuoB to a reduction site at NuoD (Supplementary Fig. 47a).

Structural modeling suggests that seven of the eight Complex I-like complexes are active [NiFe]-hydrogenases, since the predicted structures of NuoD subunits can coordinate a [NiFe] cofactor and harbor key residues important for hydrogenase activity (Supplementary Figs. 47 and 80). This differs from NuoD found in Complex I, which lacks hydrogenase activity, and instead provides a site for quinone binding and reduction. Furthermore, quinone reductase activity does not appear to be possible in these Complex I-like complexes since they lack a quinone tunnel (Supplementary Figs. 50-51) and lack an appropriate quinone binding site (Supplementary Fig. 52). Therefore, we expect that these complexes oxidize Fd_{red} and reduce protons to H₂ (not reduce quinone). To our knowledge, these are the first reported membrane-bound [NiFe]-hydrogenase complexes that have a Complex I-like structural architecture. This is notable since the NuoD subunit of Complex I is known to be evolutionarily related to the [NiFe]-hydrogenase large subunit, and further research into these Complex I-like [NiFe]-hydrogenases may provide greater insight into the overall structural evolution of Complex I^{101,102}.

One exception in our Complex I-like models was the “Hod-3” model, which is not predicted to be a [NiFe]-hydrogenase. The NuoD subunit of “Hod-3” harbors only two of the necessary four cysteines required to coordinate with a [NiFe] cofactor, and therefore it would not retain [NiFe]-hydrogenase activity (Supplementary Fig. 79). Analysis of the “Hod-3” structure indicates that it has the structural hallmarks of Complex I’s quinone binding activity, including having a quinone tunnel in NuoH, a large lipophilic cavity within the peripheral arm subunits, and a quinone binding site at NuoD (Supplementary Figs. 79-80). Importantly, the putative quinone binding site in NuoD conserves the tyrosine (Tyr66) and histidine (His17) residues required for transferring protons to quinone during the Complex I catalytic cycle¹⁰³ (equivalent to residues Tyr144 and His95 in *Y. lipolytica* NDUFS2). Therefore, we predict that the “Hod-

3" complex is a Fd-dependent Complex I that oxidizes reduced Fd and reduces quinone. Fd-dependent Complex I's have been previously reported in archaea^{99,104–106}, however, their structures have yet to be experimentally resolved.

Complex I-like membrane arm subunits

Unlike Complex I, the membrane arms of all Complex I-like group 4 [NiFe]-hydrogenase models notably lack any NuoM subunit (Supplementary Fig. 46). Given that each antiporter-like subunit in Complex I is expected to pump one proton across the lipid membrane⁹⁶ some of these [NiFe]-hydrogenase complexes may therefore pump three protons instead. However, in the case of MBH and the energy converting hydrogenase (Ech) complexes, the Gibbs free energy released from the oxidation of Fd_{red} and reduction of protons to H₂ is only enough to power the translocation of one to two protons across the membrane¹⁰⁷. Resolving this discrepancy would require an in-depth investigation of the Complex I-like [NiFe]-hydrogenase antiporter subunits, which is beyond the scope of this work (see 'AlphaFold2 model confidence and limitations' section). In the "Gerd" and "Hod-2" complexes, NuoN is also absent, which suggests that these respiratory complexes may pump only one proton per catalytic cycle similar to Ech. In certain Kariarchaeaceae and Njordarchaeales species, NuoL has undergone a gene fission event resulting in the split of the NuoL subunit into two parts, where the transmembrane helices (TMs) 5-14 form one NuoL and TMs 1-4 form a smaller subunit, which we have named NuoL'. AF2 predicts these two variant NuoL subunits form a heterodimer, which may indicate they still form a functional antiporter-like subunit. Another interesting feature not seen in Complex I is that the "Njord-2" complex is predicted to contain an additional small subunit adjacent to the NuoKJA subunits. This subunit is structurally similar to the MbhE subunit of MBH²⁶. RMSD between the 28 pruned atom pairs of both subunits is 0.889 Å (across all 77 pairs is 6.707 Å) with a shared sequence identity of 21.4%. NuoKJA and MbhA-G subunits are implicated in forming the "composite proton path" in the membrane-arm⁸⁸, therefore the additional MbhE-like subunit in Njordarchaeales complex might contribute to the functions of NuoKJA.

The membrane arm of the "Hod-3" complex retains all subunits found in Complex I (NuoLMNJKA), which suggests this complex can translocate four protons per catalytic cycle. Furthermore, the Gibbs free energy released from the oxidation of Fd_{red} and reduction of menaquinone, which is the predicted redox reaction in "Hod-3", is also comparable to that of the oxidation of NADH and reduction of ubiquinone in Complex I⁹⁹. Therefore, we predict that this "Hod-3" complex couples the oxidation of Fd_{red} and reduction of menaquinone to the pumping of four protons.

MBH-like peripheral arm subunits

The Asgardarchaeia MBH-like complex, "Asgard-2", has a peripheral arm composed of the subunits MbhJLKN. The overall subunit composition of "Asgard-2" is highly similar

to a previously described structure of MBH²⁶ and has a putative [NiFe]-hydrogenase active site in MbhL (Supplementary Figs. 46-47). The peripheral arm of “Asgard-1” is also composed of MbhJLKN and harbors a putative [NiFe]-hydrogenase active site in MbhL (Supplementary Fig. 47). However, “Asgard-1” is unique in that it has two additional subunits which we’ve named MbhO and MbhP (Supplementary Fig. 46). MbhO and MbhP are predicted to form a heterodimer which binds to two extended regions on MbhK and MbhN. MbhO is a predicted ferredoxin domain with four putative [4Fe-4S] iron-sulfur clusters (Supplementary Fig. 48c). MbhP is a predicted heterodisulfide reductase domain and is structurally similar to *M. thermolithotrophicus* HdrH (PDB ID: 5ODH)²⁹ (Supplementary Fig. 48d). Heterodisulfide reductases typically reduce a heterodisulfide coenzyme (CoM-S-S-CoB) to coenzyme M (CoM-SH) and coenzyme B (HS-CoB)¹⁰⁸. We propose the reverse reaction occurs in the “Asgard-1” complex, where CoM-SH + HS-CoB is oxidized to CoM-S-S-CoB and is coupled to the reduction of protons to H₂ at the [NiFe]-hydrogenase catalytic site in MbhL. The cofactors we modeled into the “Asgard-1” peripheral arm subunits indicate that a series of iron-sulfur clusters could bridge the heterodisulfide reductase domain of MbhP and the [NiFe]-hydrogenase active site in MbhL for efficient electron transfer (Supplementary Figs. 48-49b). Heterodisulfide oxidation activity has not been previously described in MBH-like respiratory complexes⁸⁸, therefore “Asgard-1” may constitute a novel class of MBH-like complexes.

MBH-like membrane arm subunits

The structural architecture of the two Asgardarchaea complexes we modeled are consistent with an MBH-like architecture (Supplementary Fig. 46). Similarities with MBH include the presence of Na⁺/H⁺ antiporter subunits, the packing of MbhM against MbhH in the membrane arm (which are homologous to NuoH and NuoN respectively), and the presence of a helix HL-like structure²⁶. There are some notable structural differences though. First, the helix HL structure from *P. furiosus* MBH is a small subunit (MbhI), whereas the Asgardarchaea helix HL is fused to the larger antiporter subunit, MbhH. Second, the MbhG and MbhE subunits are also fused in the “Asgard-2” complex. Third, the largest structural difference is the inclusion of an additional antiporter-like subunit in “Asgard-1”. This subunit is homologous to an antiporter-like subunit from *P. furiosus* sulfur reductase complex (MBS) termed MbsH¹⁰⁹. Given the similarity between these subunits, we termed this additional membrane subunit MbhH’. The predicted “Asgard-1” MBH-like complex is to our knowledge the first example of a MBH with a membrane arm composition that resembles MBS. Given its similarities with the MBS membrane arm, we expect “Asgard-1” might also be able to translocate two protons and a sodium ion across the membrane⁸⁸.

Separate NuoEFG-like complexes

A prototypical Complex I includes the NADH oxidation module composed of subunits NuoEFG. We observed gene clusters encoding subunits for Complex I-like complexes (i.e., *nuoLNJKABDCI*) were missing *nuoEFG* genes. Given that gene neighborhood organization tends to conserve co-interacting proteins, this raised the question whether NuoEFG subunits in Asgardarchaeota are forming separate complexes from the Complex I-like complexes, and if so, what metabolic roles these NuoEFG subunits might possibly be playing instead. First, we attempted to predict if NuoEFG and other Complex I-like subunits from the same MAG form a larger multimeric complex, despite their lack of genetic colocalization. Of our eight MAGs we modeled Complex I-like complexes from (see Supplementary Fig. 46), we annotated *nuoEFG* genes in two genomes (Hodarchaeales MAG: M3_30_Bin_130 and Gerdarchaeales MAG: GCA_016839405). We input these two sets of *nuoEFG* genes and the genes for their respective genome co-localized Complex I-like complexes ("Hod-3" and "Gerd") for structural prediction with AF3. No protein-protein interaction was predicted between these NuoEFG subunits and the Complex I-like complexes (Supplementary Fig. 78), suggesting they are each standalone complexes.

We then asked if NuoEFGs were instead forming protein-protein interactions with other proteins from its surrounding gene neighborhood. We selected 26 gene neighborhoods containing *nuoEFG*, covering a diverse phylogenetic distribution, and ran a pipeline of AlphaPulldown and AF2 multimer to predict the multimeric structures of NuoEFG-like complexes. Of the 26 gene neighborhoods containing *nuoEFG* we selected for AF2 multimer, 21 of them are predicted to form larger multimeric complexes, either as heteromers with additional proteins or as homomeric complexes (i.e. forming a dimer of trimers in a Nuo(EFG)₂ complex), while five are predicted to only form a NuoEFG 3mer. We used FoldSeek-multimer and structural comparisons with known complexes in the PDB to help group 24 of the NuoEFG-like complexes into five broad categories. These complexes share homology and structural similarity with the flavin-based electron-bifurcating (FBEB) family of enzymes known as the bifurcating ABC-containing family (BfuABC)¹¹⁰. Our nomenclature choices in describing these complexes tries to follow the conventions set by Feng et al. 2024¹¹⁰, where BfuA, BfuB, and BfuC are homologous to NuoG, NuoF, and NuoE respectively. We have named them as (i) Formate dehydrogenase (Fdh)-BfuABC-like complexes, (ii) NADH-dependent reduced ferredoxin:NADP⁺ oxidoreductase (Nfn)-BfuABC-like complexes (also known as the *Sporomusa ovata* transhydrogenase StnABC complex), (iii) Nfn-BfuABCLS-like complexes, (iv) [NiFe]-BfuABCLS-like complexes, (v) and NuoEFG-like complexes with small NuoG subunits (Supplementary Figs. 76-77). Two of the NuoEFG-like complexes were harder to assign to a group due to their unique structural configurations. We refer to these particular complexes as the Hodarchaeales electron bifurcating megacomplexes, and they were modelled in CombFold due to their large structures (Supplementary Figs.

28-29). The following paragraphs describe the predicted functions and putative functions of these different complexes.

Fdh-BfuABC-like complexes are a dimer of trimers, having a similar structural architecture to Fdh and contain the catalytic formate dehydrogenase subunit¹¹¹ (Supplementary Fig. 76a). Hence, we expect their function is to reduce CO₂ to formate (HCOO⁻). There are some noteworthy structural differences between our Fdh-BfuABC-like complexes and the Fdh structure from *Rhodobacter capsulatus*. One difference is the lack of a FdsD-like subunit in our models. FdsD is expected to help lock the bis-MGD cofactor inside the FdsA subunit¹¹¹. Another major difference is the structure of the FdsB subunits strongly resemble that of BfuB in [FeFe]-BfuABC (also known as HydABC), with a C-terminal Fd-domain harboring two [4Fe-4S] iron-cluster sites. In [FeFe]-BfuABC, this C-terminal Fd-domain forms a bridge for electron transfer between HydC and soluble Fd as part of the flavin-based electron bifurcation (FBEB) mechanism¹¹². This C-terminal Fd-domain is absent in *Rhodobacter capsulatus* FDH, and hence RcFDH is incapable of FBEB. Therefore, the presence of this domain in 5 out of the 6 FDH-like models here indicates that these are putative electron-bifurcating formate dehydrogenases, where Fd_{red} and NADH is oxidized to reduce CO₂ to formate (HCOO⁻)^{113,114}.

Nfn-BfuABC-like complexes are putative transhydrogenases that couple the reversible oxidation of NADH and Fd_{red} to the reduction of NADP⁺ via a FBEB mechanism (Supplementary Fig. 76b). They are highly structurally similar to the *Sporomusa ovata* transhydrogenase Nfn-BfuABC complex (which has also been named the StnABC or NfnABC complex)¹¹⁵⁻¹¹⁷. Like the SoNfn-BfuABC, we also expect these complexes to form a massive complex of a tetramer of trimers, though we chose not to pursue the computational structural modeling of this putative larger complex. Nfn-BfuABC complexes share structural and functional similarity with NfnLS complexes¹¹⁵⁻¹¹⁷, and are both capable of FBEB. Notably, the N-terminal region of the Nfn-BfuA subunit (also named StnC) is structurally similar to the NfnL subunit. Despite this structural similarity, Nfn-BfuA is not capable of FBEB, whereas NfnL is. Rather, Nfn-BfuABC's site of FBEB takes place in the Nfn-BfuBC subunits¹¹⁵⁻¹¹⁷.

Nfn-BfuABCLS-like complexes are, to our knowledge, previously undescribed variation on the Nfn-BfuABC complex. The Nfn-BfuABCLS-like complexes form either tetramers or pentamers (we found some subunits were fused in two of our models), and they share structural similarity to the aforementioned Nfn-BfuABC (StnABC) complex (Supplementary Fig. 76c). However, there are some key structural features that distinguish the Nfn-BfuABCLS-like complexes from Nfn-BfuABC. First, Nfn-BfuABCLS-like complexes contain an NfnS-like subunit (or in one case, a fusion domain of NfnS with BfuA), whereas Nfn-BfuABC contains only the NfnL domain. Second, there are major structural differences between the BfuA subunits of Nfn-BfuABCLS-like complexes when compared to Nfn-BfuABC complexes. As previously discussed, the Nfn-BfuA (StnC) subunit is effectively a fusion of Nfn-BfuA and NfnL, where the N-terminal region of

Nfn-BfuA has structural similarity to NfnL and is the site of NADP(H) binding¹¹⁵. However, the Nfn-BfuABCLS-like complexes have a smaller BfuA subunit and are missing the NfnL domain. Moreover, in one of our AF2 multimer models, the BfuA subunit is fused with a NfnS domain, not NfnL. Third, the orientation of the NfnL subunit/fusion-domain relative to the rest of the complex is different between the Nfn-BfuABCLS-like and Nfn-BfuABC complexes. Fourth, based on our analysis of the AF2 models, we hypothesize that the Nfn-BfuABCLS-like complexes harbour two FBEB sites, as opposed to just one as is the case for both Nfn-BfuABC and NfnLS. We base this on the structural similarity between the *Thermotoga maritima* NfnLS (PDB ID: 5JCA¹¹⁵) and the NfnLS subunits/fusion-domains of Nfn-BfuABCLS-like complexes. For instance, *Tm*NfnL R201 hydrogen-bonds with the N5 nitrogen of flavin adenine dinucleotide (FAD), which prevents N5 from becoming protonated, thereby allowing a reduced FAD to form a low-potential (i.e. high-energy) anionic semiquinone and thus become capable of FBEB. However, the equivalent site in *So*Nfn-BfuA (PDB ID: 8OH5¹¹⁵) is R239, which is not within hydrogen-bonding distance of the FAD N5 nitrogen, resulting in no FBEB taking place in the Nfn-BfuA subunit¹¹⁵. The homologous arginine residues in our Nfn-BfuABCLS-like AF2 models do, however, appear to be within hydrogen-bonding distance (2.7 - 3.3 Å) of the FAD N5 nitrogen, which we modelled in position by structural superposition with *Tm*NfnLS. This will require experimental validation though, and we are currently unsure if Nfn-BfuABCLS's unique structural configuration implies it also has unique functional implications. The presence both NfnLS subunits together in a single BfuABC complex is surprising since both Nfn-BfuABC (StnABC) and NfnLS perform the same transhydrogenase reaction: the reversible oxidation of NADH and Fd_{red} coupled to the reduction of NADP⁺, catalysed via a flavin-based electron bifurcation mechanism. Nfn-BfuABCLS presumably performs the same reaction, albeit with an additional redundant electron-bifurcating step as a result of its structure containing two putative FBEB sites. Therefore, Nfn-BfuABCLS may represent a case of apparently useless or redundant features being maintained throughout evolution¹¹⁸.

[NiFe]-BfuABCLS-like complexes are pentamers composed of BfuABC subunits bound to the small and large subunits of a [NiFe]-hydrogenase (Supplementary Fig. 76d). These complexes are similar to the *Acetomicrobium mobile* [NiFe]-BfuABCLS complex¹¹⁹ (also named the [NiFe]-HydABCLS complex), though have a different structural arrangement of subunits, have a smaller BfuA subunit, and are not predicted to form a dimer of pentamers like *Am*[NiFe]-BfuABCLS. Furthermore, their [NiFe]-hydrogenase small subunit (SSU) is unusual in that it's larger than typical [NiFe]-hydrogenase SSUs (like in *Am*[NiFe]-BfuABCLS). The [NiFe]-hydrogenase SSUs of [NiFe]-BfuABCLS-like complexes are a fusion of F₄₂₀-reducing hydrogenase subunits, FrhG and FrhB. Therefore, these new [NiFe]-BfuABCLS-like complexes may represent a previously undescribed structural variation of the [NiFe]-BfuABCLS class. That said, their expected function is the same: catalyzing the oxidation of H₂ to reduce Fd_{ox} and NAD⁺ via a FBEB mechanism.

NuoEFG-like complexes with small NuoG subunits are mostly trimers, except for one complex which forms a tetramer containing a glutamate synthase subunit (Supplementary Fig. 76e). These complexes are structurally similar to other BfuABC enzymes, except they are not predicted to form multimers larger than a 3mer, and their NuoG (BfuA) subunits are small and supposedly non-functional. We are unable to assign any function to these complexes seeing as their NuoG subunits lack the functionality of homologs like [FeFe]-BfuA, Fdh, or Nfn-BfuA. We chose to name these complexes as NuoEFG-like and avoid the BfuABC-like nomenclature due to this lack of apparent electron-bifurcating function. It is possible these small NuoEFG-like complexes could form larger complexes with additional subunits that are not co-localized in the same gene neighborhood, but that were missed by our AF2 multimer workflow.

Additional AF2 multimer and CombFold modeling indicated that NuoEFG subunits found in two different Hodarchaeales MAGs may form part of very large electron bifurcating complexes (Supplementary Figs. 28-29). One of these complexes found in MAG GCA_016840025, is composed of NuoEFG subunits forming a dimer of trimers, and attached to this central NuoEFG core are four additional enzymatic modules. These additional modules are two [NiFe]-hydrogenases (MvhADG) and two electron transferring flavoproteins complexed with a heterodisulfide reductase domain (EtfAB-Hdr), which together make a complex of Nuo(EFG)₂-Mvh(ADG)₂-Etf(AB)₂-Hdr₂ (Supplementary Figs. 28a-29a). A few putative electron bifurcating reactions may be occurring in this predicted complex. First, NADH may be oxidized and coupled to the reduction of CoM-S-S-CoB (a high potential electron acceptor) and Fd_{ox} (a low potential electron acceptor). This reaction would involve the Nuo and Hdr subunits. Second, H₂ may be oxidized and coupled to the reduction of NAD⁺ (high potential) and Fd_{ox} (low potential). This reaction is similar to that of [FeFe]-BfuABC¹¹² and would involve the Mvh and Nuo subunits. Third, the EtfAB-Hdr module may also be capable of oxidizing NADH to reduce both CoM-S-S-CoM and Fd_{ox}. An interesting possibility is that the EtfAB-Hdr module is capable of electron confurcation similar to Ldh-EtfAB¹²⁰, making the NuoEFG complex a multipurpose metabolic hub for both electron bifurcation and confurcation. In this case, CoM-SH + HS-CoB and Fd_{red} are both oxidized to reduce NAD⁺. A second large complex, found in MAG GCA_020353515, is predicted to be made of NuoEFG subunits forming a tetramer of trimers and bind to four additional [NiFe]-hydrogenases (FrhAG), forming a Nuo(EFG)₄-Frh(AG)₄ complex (Supplementary Figs. 28b-29b). We expect that this complex couples the oxidation of H₂ to the reduction of NAD⁺ and Fd_{ox} in an electron bifurcation reaction similar to that observed in [FeFe]-BfuABC¹¹². Seeing as some [NiFe]-hydrogenases can exhibit oxygen tolerance unlike the [FeFe]-hydrogenases present in [FeFe]-BfuABC (which are typically oxygen intolerant and irreversibly damaged in the presence of oxygen)¹²¹, the presence of [NiFe]-hydrogenases in both Nuo(EFG)₂-Mvh(ADG)₂-Etf(AB)₂-Hdr₂ and Nuo(EFG)₄-Frh(AG)₄ may suggest an adaptation toward oxygen tolerance in the Hodarchaeales metabolism.

The structural models presented here demonstrate an impressive amount of structural diversity contained within the NuoEFG-like family of proteins. They also highlight the role of modularity in protein evolution, given how different redox modules can be combined into larger complexes with unique predicted structural architectures. However, knowledge of these putative novel large electron bifurcating complexes is limited to structural modeling methods and will require experimental characterization.

Evolutionary implications

Complex I (NADH:ubiquinone oxidoreductase) is evolutionarily related to group 4 [NiFe]-hydrogenases^{26,88,101,102,122–124}. The ancestor of Complex I is hypothesized to have been a membrane-bound [NiFe]-hydrogenase complex that was structurally and functionally similar to extant membrane-bound hydrogenase (MBH)⁹⁹. The evolution of Complex I from an ancestral MBH-like complex would have required major structural changes, but the exact steps and the structures of ancestral intermediates have yet to be resolved. The identification of Heimdallarchaeal group 4 [NiFe]-hydrogenase membrane-bound complexes with Complex I-like structures may provide deeper insight into the evolutionary history of Complex I and potential bioenergetic advantages.

Complex I can be divided into has three structural modules with specific functions: the N-module to oxidize the electron donor molecule NADH, Q-module to reduce an electron accepting quinone, and P-module to pump protons across a lipid membrane¹²⁵. Each module originated from ancestral standalone complexes. As such, it is predicted Complex I evolved via the assembly of these distinct “prebuilt” modules. How each module may have evolved and its relationship with group 4 [NiFe]-hydrogenases Complex I-like structures is discussed below.

In the evolution of Complex I, the acquisition of an N-module provides a significant shift between lineages that use NADH as the electron donor as opposed to reduced ferredoxin (Fd_{red}). This transition is expected to have occurred late in the evolution history of Complex I¹⁰⁴. The N-module likely originates from an ancestral NAD⁺-reducing hydrogenase complex. Group 4 [NiFe]-hydrogenase Complex I-like complexes notably lack an N-module (i.e., NuoEFG subunits)^{126–129}. Despite the absence of NuoEFG subunits in these Complex I-like structures (Supplementary Fig. 78), we identified these subunits forming separate putatively electron-bifurcating complexes (Supplementary Figs. 28-29 and 76-77). Our finding suggests that the common ancestor of archaea and eukaryotes may have had all the components necessary to acquire an N-module for its ancestral Complex I. This could also provide new insight into how evolution repurposes modular subunits for different electron transferring complexes. Interestingly, NuoEFG is also missing in aerobic TACK archaea Nitrososphaerales (Thaumarchaeota) and Angelarchaeales^{130–132}. Instead, the aerobic heterotrophic marine Thaumarchaeota are predicted to encode a 2M respiratory system, which lacks NuoEFG and has a duplication of NuoM predicted to translocate an additional electron¹²⁵.

The Q-module (i.e. NuoBDCI subunits) is the site of quinone reduction in Complex I. NuoB and NuoD share a common ancestor with the group 4 [NiFe]-hydrogenases small and large subunits respectively¹²⁷. Indeed, all four NuoBDCI subunits share ancestry with the MbhJLKN subunits in MBH^{88,99}. These MBH subunits constitute the hydrogen reduction module (H-module)¹²⁴. As previously mentioned, it is likely that the ancestor of Complex I was a membrane bound hydrogenase similar to MBH, and over the course of evolution lost hydrogenase activity but gained activity to bind and reduce quinone, meaning that an ancestral H-module evolved into the present day Q-module⁹⁹. The Q-module also has the function of energetically coupling the free energy released from quinone reduction to the P-module, powering proton translocation activity¹⁰³. Remarkably, this energetic coupling activity is also seen in the H-module of extant MBH, despite the H-module and P-module having a different connection from that seen in Complex I (i.e., the junction between the membrane arms and peripheral arms is not conserved between MBH and Complex I)^{26,88}. The structure of Complex I-like group 4 [NiFe]-hydrogenases suggests that the energetic coupling mechanism between a H-module and the P-module is not limited only to the structural configuration seen in MBH but can take place in Complex I-like structural configuration too. A similar structural configuration between H- and P-modules may have taken place in the common ancestors of Complex I and Complex I-like group 4 [NiFe]-hydrogenases.

Modeling of the Asgardarchaeota NuoBI and homologs MbhJN, MbhO, and MbhP predicted plausible coordination of variable FeS clusters. NuoB/MbhJ, NuoI/MbhN, and MbhO are predicted to coordinate 1, 2, and 4 [4Fe-4S] clusters, respectively (Supplementary Fig. 48a-c). MbhP is expected to use non-cubane [4Fe-4S] clusters characteristic of heterodisulfide reductases (Supplementary Fig. 48d, 4b). This transition also changes the predicted FeS [4Fe-4S] stoichiometry from non-cubane (present in one subunit of MBH) to cubane (Complex I). Cubane FeS clusters with occupied coordination sites are more tolerant to reactive oxygen species (ROS), providing protection from the byproducts of aerobic respiration¹³³. MbhP was found only in the deeper-rooted anaerobic Asgardarchaeia Mbh-like complexes but absent in predicted aerobic lineages.

One obviously striking feature of the Asgardarchaeota group 4 [NiFe]-hydrogenase Complex I-like complexes is their short membrane arm (P-module). While the P-module of prototypical Complex I is composed of NuoLMN antiporter subunits, NuoM is absent from group 4 [NiFe]-hydrogenase Complex I-like complexes. As previously discussed, modifications of the NuoLMN impact the number of translocated protons. Yet, the structural preservation of NuoL, including the C-terminal amphipathic helix, indicates the ability to modify the position of peripheral subunits to pump protons. The previously described 2M respiratory system described in TACK contains a second full-length NuoM¹²⁵. Instead, here, we identify the absence of any NuoM in Kariarchaeaceae and Hodarchaeales and the absence of both NuoMN in Njordarchaeales and Gerdarchaeales predicted structural complexes. Instead of pumping four protons as

predicted for 2M, it would be reasonable to suggest these complexes pump one or two protons per reaction cycle. The helix-HL of NuoL – which spans the length of the membrane-arm and forms contacts with NuoM and NuoN¹³⁴ – is shorter in our Complex I-like models. Conversely, the helix-HL is longer in some bacterial species which contain two NuoM subunits instead of the one, and hence have a longer membrane-arm¹²⁵. Other researchers have suggested that the evolution of Complex I membrane-arm length is contingent on potentiating mutations in helix-HL¹²⁵. The observations here of shorter helix-HL lengths corresponding to shorter membrane arms may provide additional support for this hypothesis.

We previously discussed the loss and duplication of NuoM in different Asgardarchaeota groups further supporting the unique evolution trajectory of the membrane arms of these complexes. This complex in Thaumarchaeota is distinct from sister-lineages of ammonium-oxidizing archaea and in combination with a III-a RuBisCO suggests these organisms are aerobic. Several Asgarchaeota lineages, including Hodarchaeales and Kariarchaeaceae also encode type III RuBisCO. In the absence of NuoEFG it has been proposed these organisms use pyrroloquinoline quinone (PQQ)-dependent dehydrogenases to provide the reducing equivalents for the ETC more similar to deeper branching Asgardarchaeota¹³². Sifarchaeia and Ranarchaeia encode other mechanisms to obtain electrons, flavin-utilizing monooxygenases, various dehydrogenases, Fd/NAD(P) binding oxidoreductases. The evidence of several Nuo subunits in aerobic TACK suggest the acquisition of these complexes is deeply rooted within archaea (Fig. 5).

Bioenergetic equations for Complex I described in Chadwick et al., 2017¹²⁵ implicate several factors beyond the number of subunits in determining the number of protons crossing the membrane, including a direct relationship with extracellular pH (increase pH; increase number of pumped protons). Based on amino acid frequency modeling, Hodarchaeales has one of the highest optimum pH averages (7.93) and predicted pH maximum of 9.45, potentially increasing the number of pumped protons per reaction (Supplementary Table 4). These complexes assist in the production of a proton motive force capitalized by the Complex V ATPase for energy conservation (Supplementary Table 8).

Tracing the evolution of respiratory complexes

Complex I-like NADH:ubiquinone oxidoreductase subunits

To map the presence of NADH:ubiquinone oxidoreductase (Nuo) subunits A-N across the Asgardarchaeota phylogeny. NuoD was excluded from this analysis because it is homologous to the large subunit in [NiFe]-hydrogenases we are reviewing separately. In analyzing both sequence and structural homology, we recognized several of the Asgard Nuo subunits were divergent and missed by standard annotation pipelines. An Asgard-

specific hmm profile was made for each subunit from identified annotations and those verified with structural analyses. We conducted a hmmsearch through 664 bacteria (501 Gracilicutes (276 Alphaproteobacteria); 112 Terrabacteria; 46 CPR; 5 Fusobacteria), 1,005 archaea (26 TACK; 30 DPANN; 869 Asgard archaea; 2 Methanopyrales; 4 Methanococcales; 4 Methanobacteriales; 4 Hydrothermarchaea; 22 Diaphorarchaea; 26 Methanotecta; 10 Acherontia; 8 Stygia), and 15,980 mitochondrial genomes. In cases with a high number of similar sequences, we chose a subset of the mitochondrial genomes to avoid oversaturation by the mitochondrial sequences (see Figshare data). We conducted multiple rounds of alignment for each subunit to determine the best cutoffs for each subunit (Supplementary Table 15) and manual removal of false hits. Final filtered sequences were aligned and trimmed with MAFFT v7.490 and trimAl v1.4.rev15, respectively. Individual protein phylogenies produced with IQ-TREE and used to compare these subunits to bacteria and mitochondrial representatives (Supplementary Figs. 15-27). This provided an overview of the presence of these subunits across lineages. This approach did remove a few positive hits in ETC subunits (i.e., SdhB in cultured Lokiarchaeales).

We traced the presence of these subunits across the NM47 phylogeny. Atabayarchaeia and Frey/Jordarchaeia lack NuoEFG with only a few instances of subunit NuoG homologs. There is also patchy distribution of NuoJK between environments, typically absent in deep sea and hot spring genomes. NuoEFG was identified in most Ranarchaeia, but only one Odinararchaeia with no NuoK in this next branching point. Similarly, there is a patchy distribution of NuoEFG in Asgardarchaeia, Thorarchaeia, and Sifarchaeia, but more instances of NuoK. Lokiarchaeia orders (Helarchaeales and Lokiarchaeales) are missing several Nuo subunits, primarily identified to have homologs to NuoI and NuoEFG. Wukongarchaeia also appears to be lacking several Nuo subunits, however this lineage has only three genomes, which limits the analysis. Hermodarchaeia also typically lacks NuoEFG and NuoK. Moving from Njordarchaeales to Gerdarchaeales, Kariarchaeaceae, and Heimdallarchaeaceae we see increasing presence of NuoEFG and NuoK. We identified several Hodarchaeales with homologs for NuoA-N. In observing these transitions, we notice the patchy distribution of NuoJK, which may be linked to the absence of homologs in MBH complexes. Further, we start to identify increasing instances of NuoEFG. Since our analysis included bacterial, archaeal, and mitochondrial sequences we could also trace these subunits across other lineages. Interestingly, we did identify NuoEFG in other archaeal lineages, including Methanobacteriales, Thermococcales, Persephonarchaeia, Theionarchaeia, Bathyarchaeota, Izemarchaeia, Methanophagales, and Nanohaloarchaeota.

Our analyses support the eukaryotic electron transport chain subunits likely did not arise from the archaeal ancestor. Given the unlikelihood of an Asgard archaeal origin, we report the lineages clustering with the mitochondrial subunits in our phylogenetic analyses. Complex I is modular, so we considered each protein separately and within the

three functional modules (P, N, and Q) ¹²⁵. Briefly, we found that most subunits clustered with Alphaproteobacterial orders: Kiloniellales, Rhizobiales, and UBA8366.

The P-module of the ETC Complex I is membrane bound and includes NuoA, NuoH, NuoJ, NuoK, NuoL, NuoM and NuoN subunits¹²⁵. Phylogenetic analysis revealed that mitochondrial NuoA subunits clustered with Alphaproteobacterial lineages, specifically orders Rickettsiales (families Rickettsiaceae and Anaplasmataceae) and UBA1280 (family JALTLI01), with the UBA1280 genome recovered from the Bohai Sea. Similarly, mitochondrial NuoH subunits also grouped with Alphaproteobacteria, including orders: Rhodospirillales (family UXAT02), Rhodospirillales_A (families CALINB01 and JABHGV01), Tistrellales (family Tistrellaceae), and UBA8366 (family JADFVY01). The Rhodospirillales_A are from the Guaymas Basin and the Bohai Sea, while all the UBA8366 clustering with Nuo subunits are from the Bohai Sea. NuoJ mitochondrial proteins clustered with members of Pseudomonadota classes: Alphaproteobacteria, Gammaproteobacteria, and Zetaproteobacteria. Within Alphaproteobacteria, most of the sequences were from Kiloniellales, Rhizobiales, Rhodobacterales, and several UBA lineages (UBA11136, UBA1280, UBA6615, UBA7887, UBA8366). Zetaproteobacterial NuoJ proteins were all from order Mariprofundales (family Mariprofundaceae), while the Gammaproteobacteria included Acidiferrobacterales, Acidithiobacillales, and Burkholderiales orders.

We identified unique clustering in the antiporter subunits of the P-module (NuoL, NuoM, and NuoN). Unlike the other mitochondrial proteins NuoL is clustered with both archaeal and bacterial sequences. The archaeal sequences included representatives from Euryarchaeota (Halobacteriota) and TACK (Thermoproteota) lineages: Archaeoglobi, Syntropharchaeia, and Sulfolobales. The bacterial NuoL were distributed across multiple phyla, including Aquificota, Hydrothermota, Cyanobacteriota, Desulfobacterota_D (class UBA1144), Gemmatimonadota, Nitrospinota, Nitrospirota_A, Nitrospirota, Planctomycetota, and Pseudomonadota (class Gammaproteobacteria). NuoM mitochondrial subunits grouped with Desulfobacterota, Nitrospinota, Nitrospirota, Nitrospirota_A, Pseudomonadota, Thermodesulfobiota, and Verrucomicrobiota. The Pseudomonadota lineages included Alphaproteobacteria (orders JAAEOK01, Kiloniellales, Rickettsiales, Sphingomonadales, and UBA1280) and Zetaproteobacteria (order Mariprofundales) classes. NuoN mitochondrial subunits were clustered with Chrysiogenota, Cyanobacteriota, and Pseudomonadota. Pseudomonadota had several orders, but the majority were from Kiloniellales, Rhizobiales, and UBA8366.

Overall, subunits within the P-module are primarily grouped with Alphaproteobacteria, particularly the orders Kiloniellales, Rhizobiales, UBA8366. Other Alphaproteobacterial orders appeared across multiple subunits, including JAAEOK01, Rhodospirillales, Rhodospirillales_A, Rickettsiales, Sphingomonadales, and UBA1280. The mitochondrial NuoE, NuoF, and NuoK subunits did not meet the strict cut-offs

and/or did not align well with the bacterial and archaeal sequences, producing long branches. Therefore, they are excluded from this discussion.

The N-module involved in NADH oxidation is discussed in detail in *Separate NuoEFG-like complexes* and *Evolutionary implications* sections. Briefly, the mitochondrial NuoG subunits cluster with several Alphaproteobacterial orders. The greatest number of hits were from JAAEOK01, Kiloniellales, Rhizobiales, Rhodobacterales, SMXQ01, Shingomonadales, UBA6615, and UBA8366. The sequences grouping with mitochondrial proteins included several UBA lineages and metabolically undescribed Alphaproteobacterial groups.

Lastly, the Q-module contains four subunits NuoB, NuoC, NuoD, and NuoI. The evolutionary implication of the [NiFe]-hydrogenases and NuoD homology is discussed at length in both main and supplementary texts. NuoB and NuoI mitochondrial sequences group with representative sequences from Cyanobacteriota and Pseudomonadota phyla. Mitochondrial NuoB grouped with Pseudomonadota classes Alphaproteobacteria, Gammaproteobacteria, and Zetaproteobacteria. Whereas NuoC and NuoI both grouped with Alphaproteobacteria. Similarly to previous subunits the majority of the NuoB hits are from Kiloniellales, Rhizobiales, Rhodobacterales, and UBA8366. NuoI mitochondrial subunits grouped with JAAEOK01, Kiloniellales, Rickettsiales, Tistrellales, and UBA1280. NuoC mitochondrial subunits also grouped with Kiloniellales and UBA1280 along with Ferrovibrionales, GCA-2731375, Rhodospirillales_A, and SMXQ01. Across the three analyzed subunits in the Q-module, we reported grouping primarily with Cyanobacteriota and Alphaproteobacterial lineages.

Interestingly, a subset of plant mitochondrial subunits (NuoN, NuoB, and NuoI) group with Cyanobacteria away from the other eukaryotic sequences. The cyanobacterial classes include Cyanobacteriia (Oxyphotobacteria) and Vampiropvibrionia. Within Cyanobacteriia, we see representatives from the orders Cyanobacteriales, Gloeobacterales, PCC-6307, Synechococcales, and Thermosynechococcales. Within Vampiropvibrionia (Melainabacteria), we see two orders 2-02-FULL-35-15 and Vampiropvibrionales. All the mitochondrial subunits grouping with cyanobacterial sequences are from plants: *Zingiber leptorrhizum*, *Allardia glabra*, *Melicope pteleifolia*, *Viola labradorica*, *Osmanthus didymopetalus*, *Psychotria viridis*, and *Dracocephalum moldavica*. Vampiropvibrionia are non-photosynthetic cyanobacteria and have been found to be primarily host-associated¹³⁵. If this grouping is robust, it suggests these mitochondrial subunits are either a result of early endosymbiotic cyanobacteria transfer (given wide phylogenomic diversity) to early plant ancestors or more likely horizontally transferred through host-associated cyanobacteria. HGT is more plausible currently, but not the focus of this study.

Alphaproteobacteria were the only class that grouped consistently with mitochondrial sequences across all the analyzed Complex I subunits. More specifically,

Ferrovibrionales, GCA-2731375, JAAEOK01, Kiloniellales, Rhodospirillales, Rhodospirillales_A, Rickettsiales, SMXQ01, Sphingomonadales, UBA1280, and UBA8366 representatives grouped with at least 5 of the analyzed mitochondrial subunits. We identified only a few Asgardarchaeota single sequences clustering with the eukaryotic electron transport chain subunits. Closer inspection indicates these single hits are either contamination or recent transfer events. This supports previous claims of the electron chain subunit's origin in bacteria, more specifically Alphaproteobacteria or their sister lineages.

We leveraged the Bohai Sea and Guaymas Basin Alphaproteobacterial genomes from the same sequencing project as the Asgardarchaeota released in this study. This allowed us to rule out recent horizontal gene transfers that could falsely resemble evolution of mitochondrial sequences. Instead, we observed these Alphaproteobacterial genomes released here greatly increase the diversity of undescribed clades and provide most of the hits clustering with mitochondrial subunits. Further highlighting the evolutionary insight provided by these genomes within these sampling sites.

Complex II- Succinate dehydrogenase

Succinate dehydrogenase components (Complex II; Sdh) are expected to have evolved from the anaerobic fumarate respiratory complex¹³⁶. Fumarate reductase (Frd) converts fumarate to succinate, which is reversed in Sdh. Typical Sdh includes four subunits: SdhA flavoprotein with a binding site for the oxidation of succinate, SdhB contains the electron transferring FeS clusters, SdhC is membrane-bound, transferring electrons and binding quinone, and SdhD anchors the complex¹³⁷. Within the Asgardarchaeota lineages, we identify homologs for SdhA in all 17 lineages with SdhB in 15 (missing from Asgardarchaeia and Odinararchaeia) likely due to limited number of MAGs from these groups and the stricter cut-offs in this analysis. However, we only find SdhCD homologs in Heimdallarchaeia lineages Hodarchaeales, Kariarchaeaceae, and Njordarchaeales. This increasing addition of subunits has also been described in other archaea, including Sulfolobales¹³⁷. As these are the membrane-anchoring subunits it is possible these provide localization necessary to facilitate electron transfers.

QcrABC- cytochrome c reductase

We identified the QcrABC in five Hodarchaeales guild 11 coastal genomes and potentially one Hermodarchaeia guild 5 coastal MAG. To more thoroughly analyze these subunits, we only included sequences that had three flanking proteins for the hmm search. Asgardarchaeota sequences following QcrA were very similar, so we used several annotation pipelines to separate QcrB and C subunits. The addition of these subunits in Hodarchaeales increases the proton gradient for the generation of ATP. Further biochemical analysis is necessary to confirm these putative complexes.

Nitrate respiratory complexes

In analyzing nitrate reductase complex subunits, we found that guild 12 deep sea Hodarchaeales and deep sea and hot spring Kariarchaeaceae encode NarGHIJ. However, NarGHJ were absent in all coastal genomes from these groups. Further, we also identified NarI two guild 7 MAGs (Baldrarchaeia and Ranarchaeia), as well as guild 11 coastal Hodarchaeales and guild 9 Heimdallarchaeaceae (Supplementary Figs. 53-56). NarI is a membrane-bound subunit involved in transferring electrons from quinones¹³⁸. Yet its role in the absence of the other complex subunits (NarGHJ) in the coastal genomes suggest an environmental metabolic shift in the evolution of Hodarchaeales.

Considering both the presence and absence of Complex III and NarI we identify two potential evolutionary scenarios to explain the distribution: 1) Hodarchaeales' ancestor encoded Complex III, which was later lost in a portion of the genomes. 2) Hodarchaeales' ancestor encoded a Complex III-NarI complex similar to the previously proposed hypothesis⁷ and has more recently gain additional Complex III subunits. In both cases, the complex provides the capacity to shuttle electrons.

Complex IV- HCO superfamily

Cytochrome bc subunits involved in respiration (bc_L) and photosynthesis (bc_Lf) contribute to the generation of an electrochemical gradient across the organelle internal membranes. Presence of cytochrome bc in archaea likely occurred after the last universal common ancestor (LUCA) through LGT from bacteria and potentially co-occurs with fatty acid metabolism and the availability to harness a variety of electron acceptors, including oxygen (copper-containing cytochromes and terminal oxidases)¹³⁹.

To further distinguish the likelihood of oxygen as a terminal electron acceptor, we ran phylogenomic analyses on Asgardarchaeota cytochrome c oxidases within the heme-copper oxidase (HCO) superfamily. This superfamily is typically split into NO reductases (NOR), nitric oxide dismutase families (NOD), and oxygen-reducing families (Type A-I)¹⁴⁰. We identified three oxygen-reducing families A, K (B9), and G (B, B6) (Supplementary Fig. 72). Type A included 41 Hodarchaeales, 22 Kariarchaeaceae, 1 Gerdarchaeales, and 1 Lokiarchaeales MAGs. This family is well-known for presence in mitochondria and bacteria, which reduce oxygen to water to increase the proton gradient¹⁴⁰. All except one Kariarchaeaceae and one Lokiarchaeales sequence cluster together with previously classified type A sequence (Ga0213864_100007002) near the base separating Type A and Type H (B2). Distinguishing characteristics of Type A include two proton channels (K and D), oxidizing cytochrome c and in some cases quinol. Type A has also been identified in non-Asgard archaeal lineages, including Thermoproteia, Poseidoniia, Thermoplasmata, Halobacteria, Iainarchaeia, Bathyarchaeia, and Nitrososphaeria¹⁴⁰.

Additionally, we identified one K-type Kariarchaeaceae and one G-type Hodarchaeales HCO¹⁴⁰. The rooting of the superfamily has been a major hurdle, but the

most robust analyses suggest NO reduction is likely a derived function¹⁴¹. Type K and G have a single proton channel, oxidizing cytochrome c. Previous studies identified Type K in archaeal lineages, Heimdallarchaeia and Thermoproteia; whereas, Type G primarily reported in bacteria has also recently been reported in archaeal clade Thermoproteia¹⁴⁰.

Complex V- ATP synthase

Across Asgardarchaeota, we identify V/A-type ATPase subunits. However, guild 2 coastal Lokiarchaeales and Hodarchaeales MAGs had hits to a few F-type ATPase subunits and 75% of Hermodarchaeia guild 5 coastal genomes encoding V-type subunit (ATPeV0C; K02155). This supports previous descriptions of Complex V in Asgardarchaeota^{7,63}.

Potential impact on eukaryogenesis

Actin homologs

Heimdallarchaeia contain actin homologs like those in Lokiarchaeales, which in Rodrigues-Oliveira et al., 2023, produce protrusions (Supplementary Table 12). Yet in a cultured bacterium (*Candidatus Uab amorphum*¹⁴²), actin does not produce this morphology, necessitating image verification. Protrusions may increase surface-to-volume ratio to access terminal electron acceptors and initial compartmentalization for concentrated proton gradients (Supplementary Table 12)^{62,143}. Portions of Complexes I, III, and IV are typically present in aerobic eukaryotes mitochondrial DNA (mtDNA). When Complex IV is encoded in the nucleus, subunit CoxI mistargets the endoplasmic reticulum (ER) instead of the mitochondrial inner membrane^{144,145}. Conservation of mtDNA enables confined redox regulation while reducing the challenges of transferring hydrophobic subunits outside the mitochondria (i.e., CoxI). The Inside-Out Model predicted Asgard blebs (reminiscent of now observed protrusions) gave rise to the ER, which could explain this mistargeting^{146–148}. Regardless of protrusion production, oxygen tolerance mediated by ROS defenses would enable close-knit spatial interactions with potential aerobic or facultatively anaerobic partners in the microbial consortium.

ROS signaling

Under aerobic conditions, Complex I and III provide the largest generation of free electrons, enabling production of ROS¹⁴⁹. Previous analysis suggests catalase and peroxidase develop before the GOE in response to rising oxygen levels¹⁵⁰. The gain of Complex III increases ROS production, which is converted to H₂O₂ by superoxide dismutase (SOD) and exported. Despite its toxic impact, ROS is also considered an ancestral signaling response, documented in single-cell algae and arising before multicellular eukaryotes¹⁵⁰. The increased export of H₂O₂ may have instigated a ROS wave to surrounding cells, cascading to allow quicker community response to

environmental stress (O₂) and tightening community structure with cell-to-cell contact. Further enabling redox signaling and potentially CoxD to harness the partner's cell cycle (proliferation and apoptosis) and ATP production, respectively¹⁵¹. We also detected myoglobin (Mb) and protoglobin (Pgbs), which may have provided the ability to sense and bind oxygen at low levels. Impact of oxygen stress could have been mitigated by protoglobin binding molecular oxygen, aiding diffusion to the endosymbiont and control of ROS. It would also reduce the impact of oxygen stress from the partner's respiration (fatty acid oxidation), killing the host¹⁵². Today H₂O₂ and control of ROS provides pivotal intercellular signaling between host and organelles¹⁵³.

SCO1

Sco proteins have been linked to various functions, including COX regulation and assembly, protection against oxidative stress, and redox regulation through Cu-binding signaling molecules or Cu homeostasis¹⁵⁴. SCO1 (Sco1) has been implicated in copper transfer (ion loading) to cytochrome c oxidase, likely involved in the posttranslational steps of the complex's formation^{155–157}. It has been proposed that regulation of the ETC and endosymbiont ROS production arose after eukaryogenesis¹⁵⁸. However, here we show that SCO1 was co-localized with CoxABC in Heimdallarchaeia, suggesting coordination of this pathway could have originated before eukaryogenesis. Separate analysis of pathogenic *Neisseria gonorrhoeae* indicate Sco is related to its prevention of oxidative damage rather than in the formation of Cu-containing proteins¹⁵⁹. Therefore, it may also provide additional protection against byproducts of aerobic metabolism.

SeqCode Taxonomic Descriptions

We provide high-quality genomes as type material to update the Asgardarchaeota phylum nomenclature. These taxonomic names will be submitted to the SeqCode Registry¹⁶⁰. These new classifications are supported by AAI (Supplementary Table 13) and phylogenomic analyses (Fig. 1 and Supplementary Fig. 3).

Candidatus Heimdallarchaeum bathyobscurus. *Candidatus* Heimdallarchaeum bathyobscurus (bathyobscurus. N.L. neut. adj. bathy, deep and obscurus, mysterious). A deep-sea taxon from sediment sampling, predicted to optimally prefer neutrophilic, oligohaline, and thermophilic conditions. This uncultured lineage is represented by the type species “D4993_C5_LG_Bin_739_clean”, recovered from Guaymas Basin sediments, and defined as a high-quality metagenome-assembled genome with an estimated completeness of 93.46% and 4.67% contamination, 3.14 Mb genome size and 3.45 Mb predicted genome size.

Candidatus Heimdallarchaeum abyssi. *Candidatus* Heimdallarchaeum abyssi (abyssi. L.

neut. adj. abyssi, deep). A deep-sea taxon from sediment sampling, predicted to optimally prefer neutrophilic, mesohaline, and mesophilic conditions. This uncultured lineage is represented by the type species “Meg22_1214_Bin_408”, recovered from Guaymas Basin sediments, and defined as a high-quality metagenome-assembled genome with an estimated completeness of 90.65% and 4.21% contamination, 3.43 Mb genome size and 3.76 Mb predicted genome size.

Candidatus Freyarchaeum bathylpelagicus. *Candidatus* Freyarchaeum bathylpelagicus (bathypelagicus. N.L. neut. adj. bathy, deep and pelagicus, water). A deep-sea taxon from sediment sampling, predicted to optimally prefer neutrophilic, mesohaline, and hyperthermophilic conditions. This uncultured lineage is represented by the type species “D4998_C1112_H3_Bin_55_clean”, recovered from Guaymas Basin sediments, and defined as a high-quality metagenome-assembled genome with an estimated completeness of 90.76% and 3.12% contamination, 2.76 Mb genome size and 3.02 Mb predicted genome size.

Candidatus Prometheoarchaeum guaymasis. *Candidatus* Prometheoarchaeum guaymasis (guaymasis. N.L. neut. adj. of Guaymas Basin, Gulf of California). A deep-sea taxon from sediment sampling, predicted to optimally prefer neutrophilic, mesohaline, and mesophilic temperature conditions. This uncultured lineage is represented by the type species “D4993_C5_LG_Bin_279”, recovered from Guaymas Basin sediments, and defined as a high-quality metagenome-assembled genome with an estimated completeness of 91.67% and 2.8% contamination, 3.75 Mb genome size, and 4.06 Mb predicted genome size.

Candidatus Helarchaeum bathyoculti. *Candidatus* Helarchaeum bathyoculti (bathyoculti. N.L. neut. adj. bathy, deep and occulti, hidden from sight). A deep-sea taxon from sediment sampling, predicted to optimally prefer neutrophilic, oligohaline, and thermophilic conditions. This uncultured lineage is represented by the type species “Meg22_1416_Bin_67”, recovered from Guaymas Basin sediments, and defined as a high-quality metagenome-assembled genome with an estimated completeness of 91.28% and 3.74% contamination, 4.89 Mb genome size, and 5.31 Mb predicted genome size.

Candidatus Helarchaeum guaymasis. *Candidatus* Helarchaeum guaymasis (guaymasis. N.L. neut. adj. of Guaymas Basin, Gulf of California). A deep-sea taxon from sediment sampling, predicted to optimally prefer neutrophilic, mesohaline, and thermophilic conditions. This uncultured lineage is represented by the type species “Meg22_1214_Bin_169”, recovered from Guaymas Basin sediments, and defined as a high-quality metagenome-assembled genome with an estimated completeness of 95.48% and 4.21% contamination, 4.76 Mb genome size, and 4.97 Mb predicted genome size.

Candidatus Helarchaeum guayarcanus. *Candidatus Helarchaeum guayarcanus* (guayarcanus. N.L. neut. adj. of Guaymas Basin, Gulf of California). A deep-sea taxon from sediment sampling, predicted to optimally prefer neutrophilic, mesohaline, and thermophilic conditions. This uncultured lineage is represented by the type species “Meg22_1012_Bin_314”, recovered from Guaymas Basin sediments, and defined as a high-quality metagenome-assembled genome with an estimated completeness of 95.95% and 3.63% contamination, 5.56 Mb genome size, and 5.78 Mb predicted genome size.

Candidatus Helarchaeum bathyrypticus. *Candidatus Helarchaeum bathyrypticus* (bathyrypticus. N.L. neut. adj. bathy, deep and crypticus, hidden). A deep-sea taxon from sediment sampling, predicted to optimally prefer neutrophilic, oligohaline, and thermophilic conditions. This uncultured lineage is represented by the type species “Meg22_1416_Bin_72”, recovered from Guaymas Basin sediments, and defined as a high-quality metagenome-assembled genome with an estimated completeness of 92.06% and 3.27% contamination, 4.90 Mb genome size, and 5.28 Mb predicted genome size.

Candidatus Helarchaeum abyssi. *Candidatus Helarchaeum abyssi* (abyssi. L. neut. adj. abyssi, deep). A deep-sea taxon from sediment sampling, predicted to optimally prefer neutrophilic, mesohaline, and thermophilic conditions. This uncultured lineage is represented by the type species “D4994_C39_H1_Bin_674”, recovered from Guaymas Basin sediments, and defined as a high-quality metagenome-assembled genome with an estimated completeness of 92.68% and 3.27% contamination, 3.99 Mb genome size, and 4.28 Mb predicted genome size.

Candidatus Helarchaeum guayelus. *Candidatus Helarchaeum guayelus* (guayelus. N.L. neut. adj. guay, isolated from the Guaymas Basin and elusus, elusive). A deep-sea taxon from sediment sampling, predicted to optimally prefer neutrophilic, mesohaline, and thermophilic conditions. This uncultured lineage is represented by the type species “Meg22_1214_Bin_20”, recovered from Guaymas Basin sediments, and defined as a high-quality metagenome-assembled genome with an estimated completeness of 93.39% and 3.74% contamination, 3.80 Mb genome size, and 4.04 Mb predicted genome size.

Candidatus Thorarchaeum guaymasis. *Candidatus Thorarchaeum guaymasis* (guaymasis. N.L. neut. adj. of Guaymas Basin, Gulf of California). A deep-sea taxon from sediment sampling, predicted to optimally prefer neutrophilic, moderately halophilic, and thermophilic conditions. This uncultured lineage is represented by the type species “D4993_C5_H2_Bin_79”, recovered from Guaymas Basin sediments, and defined as a high-quality metagenome-assembled genome with an estimated completeness of 93.93% and 3.74% contamination, 2.88 Mb genome size, and 3.06 Mb predicted genome size.

Candidatus Thorarchaeum bohaisis. *Candidatus* Thorarchaeum bohaisis (bohaisis. N.L. neut. adj. of the Bohai Sea). A coastal taxon from sediment sampling, predicted to optimally prefer neutrophilic, moderately halophilic, and mesophilic temperature conditions. This uncultured lineage is represented by the type species “M3_30_Bin_136”, recovered from Bohai Sea sediments, and defined as a high-quality metagenome-assembled genome with an estimated completeness of 91.12% and 1.87% contamination, 3.71 Mb genome size, and 4.04 Mb predicted genome size.

Candidatus Thorarchaeum abyssi. *Candidatus* Thorarchaeum abyssi (abyssi. L. neut. adj. abyssi, deep). A deep-sea taxon from sediment sampling, predicted to optimally prefer neutrophilic, moderately halophilic, and mesophilic temperature conditions. This uncultured lineage is represented by the type species “Meg22_810_Bin_118”, recovered from Guaymas Basin sediments, and defined as a high-quality metagenome-assembled genome with an estimated completeness of 93.46% and 4.67% contamination, 3.79 Mb genome size, and 4.03 Mb predicted genome size.

Candidatus Thorarchaeum epipelagicus. *Candidatus* Thorarchaeum epipelagicus (epipelagicus. L. neut. adj. epipelagic, shallow). A coastal taxon from sediment sampling, predicted to optimally prefer mesophilic, moderately halophilic, and mesophilic temperature conditions. This uncultured lineage is represented by the type species “M3_44_Bin_221”, recovered from Bohai Sea sediments, and defined as a high-quality metagenome-assembled genome with an estimated completeness of 90.19% and 4.67% contamination, 4.14 Mb genome size, and 4.54 Mb predicted genome size.

Candidatus Gerdarchaeum guaymasis. *Candidatus* Gerdarchaeum guaymasis (guaymasis. N.L. neut. adj. of Guaymas Basin, Gulf of California). A deep-sea taxon from sediment sampling, predicted to optimally prefer neutrophilic, mesohaline, and mesophilic temperature conditions. This uncultured lineage is represented by the type species “Meg22_1214_Bin_23”, recovered from Guaymas Basin sediments, and defined as a high-quality metagenome-assembled genome with an estimated completeness of 92.37% and 4.74% contamination, 3.87 Mb genome size, and 4.17 Mb predicted genome size.

Candidatus Idunarchaeum gen. nov. (Idunarchaeum. Old Norse.fem. n. Idun, Norse goddess of the apples and youth, N.L. neut. n. archaeum, unicellular microorganism which lack an organized nucleus; N.L. neut. n. Idunarchaeum, referring to the type genus Idunarchaeum). Type species: *Candidatus* Idunarchaeum guaymasis within the Lokiarchaeaceae family.

Candidatus Idunarchaeum guaymasis. *Candidatus* Idunarchaeum guaymasis (guaymasis. N.L. neut. adj. of Guaymas Basin, Gulf of California). A primarily deep-sea taxon from

sediment sampling, predicted to optimally prefer neutrophilic, mesohaline, and thermophilic conditions. This uncultured lineage is represented by the type species “Meg22_1214_Bin_189”, recovered from Guaymas Basin sediments, and defined as a high-quality metagenome-assembled genome with an estimated completeness of 92.52% and 2.34% contamination, 5.41 Mb genome size, and 5.81 Mb predicted genome size.

Candidatus Lokiarchaeum abyssi. *Candidatus* Lokiarchaeum abyssi (abyssi. L. neut. adj. abyssi, deep). A deep-sea taxon from sediment sampling, predicted to optimally prefer neutrophilic, oligohaline, and thermophilic temperature conditions. This uncultured lineage is represented by the type species “Meg22_1012_Bin_159”, recovered from Guaymas Basin sediments, and defined as a high-quality metagenome-assembled genome with an estimated completeness of 91.3% and 3.74% contamination, 5.34 Mb genome size, and 5.80 Mb predicted genome size.

Candidatus Skadiarchaeaceae N.L. neut. n. (Skadiarchaeaceae, referring to the type genus Skaidarchaeum; -aceae, ending to denote a family; N.L. fem. pl. N. Skadiarchaeaceae, the Skadiarchaeum family). Type species: *Candidatus* Skadiarchaeum guaymasis within the Freyarchaeales order.

Candidatus Skadiarchaeum gen. nov. (Skadiarchaeum. Old Norse.fem. n. Skadi, Norse goddess of winter, N.L. neut. n. archaeum, unicellular microorganism which lack an organized nucleus; N.L. neut. n. Skadiarchaeum, referring to the type genus Skadiarchaeum). Type species: *Candidatus* Skadiarchaeum guaymasis within the Freyarchaeales order.

Candidatus Skadiarchaeum guaymasis. *Candidatus* Skadiarchaeum guaymasis (guaymasis. N.L. neut. adj. of Guaymas Basin, Gulf of California). A deep-sea taxon from sediment sampling, predicted to optimally prefer neutrophilic, oligohaline, and thermophilic conditions. This uncultured lineage is represented by the type species “D4998_C1112_H3_Bin_32_clean”, recovered from Guaymas Basin sediments, and defined as a high-quality metagenome-assembled genome with an estimated completeness of 95.79% and 4.67% contamination, 2.37 Mb genome size, and 2.47 Mb predicted genome size.

Candidatus Sigynarchaeaceae N.L. neut. n. (Sigynarchaeaceae, referring to the type genus Skaidarchaeum; -aceae, ending to denote a family; N.L. fem. pl. N. Sigynarchaeaceae, the Sigynarchaeum family). Type species: *Candidatus* Sigynarchaeum guaymasis within the Lokiarchaeales order.

Candidatus Sigynarchaeum gen. nov. (Sigynarchaeum. Old Norse.fem. n. Sigyn, Norse

goddess connected to compassion and mercy, N.L. neut. n. archaeum, unicellular microorganism which lack an organized nucleus; N.L. neut. n. Sigynarchaeum, referring to the type genus Sigynarchaeum). Type species: Candidatus Sigynarchaeum guaymasis within the Lokiarchaeales order.

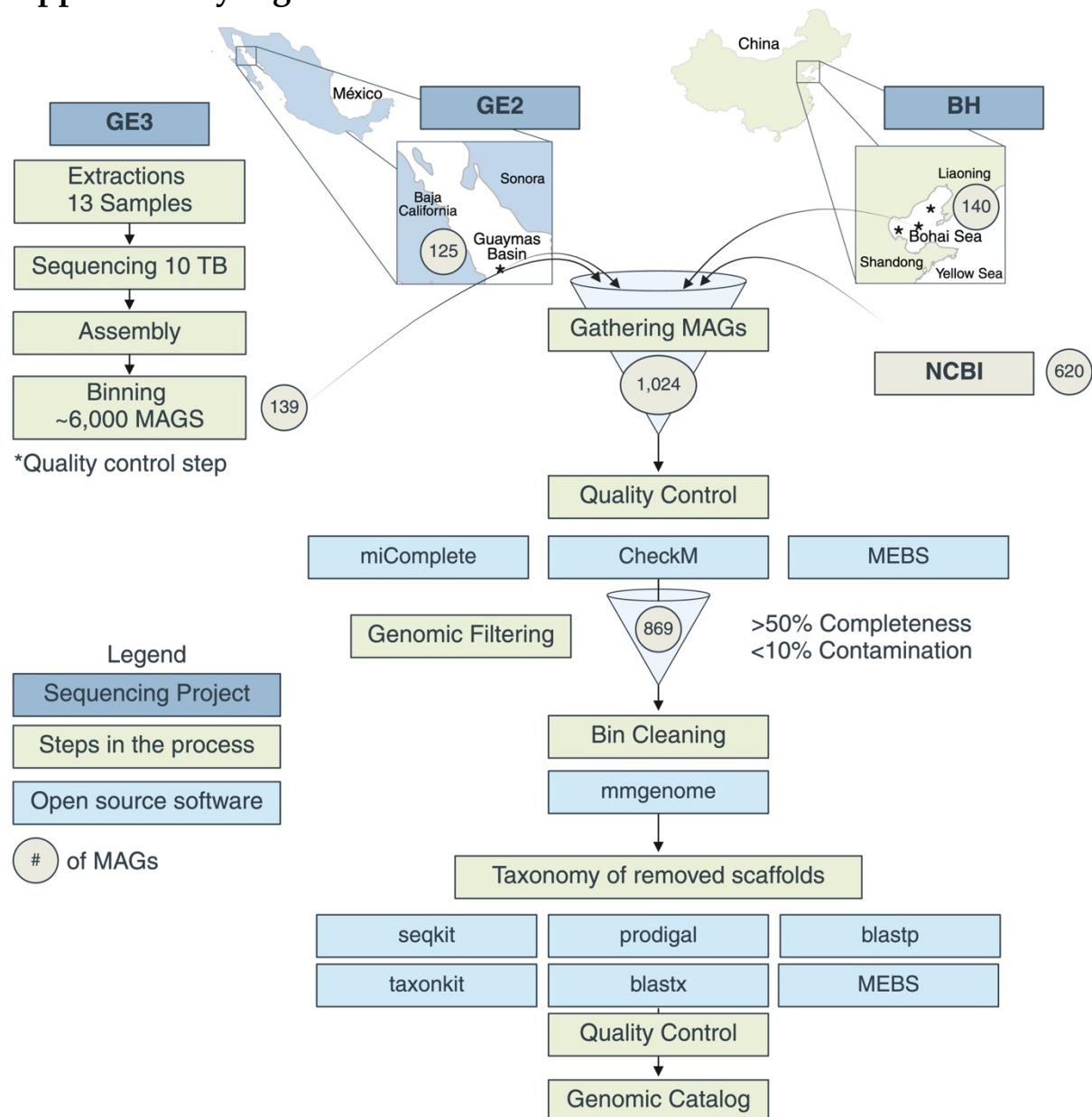
Candidatus Sigynarchaeum guaymasis. Candidatus Sigynarchaeum guaymasis (guaymasis. N.L. neut. adj. of Guaymas Basin, Gulf of California). A deep-sea taxon from sediment sampling, predicted to optimally prefer neutrophilic, oligohaline, and mesophilic temperature conditions. This uncultured lineage is represented by the type species “D4994_C39_H1_Bin_319”, recovered from Guaymas Basin sediments, and defined as a high-quality metagenome-assembled genome with an estimated completeness of 92.52% and 3.74% contamination, 4.89 Mb genome size and 5.25 Mb predicted genome size.

Candidatus Friggarchaeaceae N.L. neut. n. (Friggarchaeaceae, referring to the type genus Friggarchaeum; -aceae, ending to denote a family; N.L. fem. pl. N. Friggarchaeaceae, the Friggarchaeum family). Type species: Candidatus Friggarchaeum guaymasis within the Odinararchaeales order.

Candidatus Friggarchaeum gen. nov. (Friggarchaeum. Old Norse.fem. n. Frigg, Norse goddess connected to wisdom and marital bonds, N.L. neut. n. archaeum, unicellular microorganism which lack an organized nucleus; N.L. neut. n. Friggarchaeum, referring to the type genus Friggarchaeum). Type species: Candidatus Friggarchaeum guaymasis within the Odinararchaeales order.

Candidatus Friggarchaeum guaymasis. Candidatus Friggarchaeum guaymasis (guaymasis. N.L. neut. adj. of Guaymas Basin, Gulf of California). A deep-sea taxon from sediment sampling, predicted to optimally prefer neutrophilic, mesohaline, and thermophilic conditions. This uncultured lineage is represented by the type species “Meg22_1214_Bin_101”, recovered from Guaymas Basin sediments, and defined as a high-quality metagenome-assembled genome with an estimated completeness of 96.26% and 4.21% contamination, 1.58 Mb genome size, and 1.64 Mb predicted genome size.

Supplementary Figures

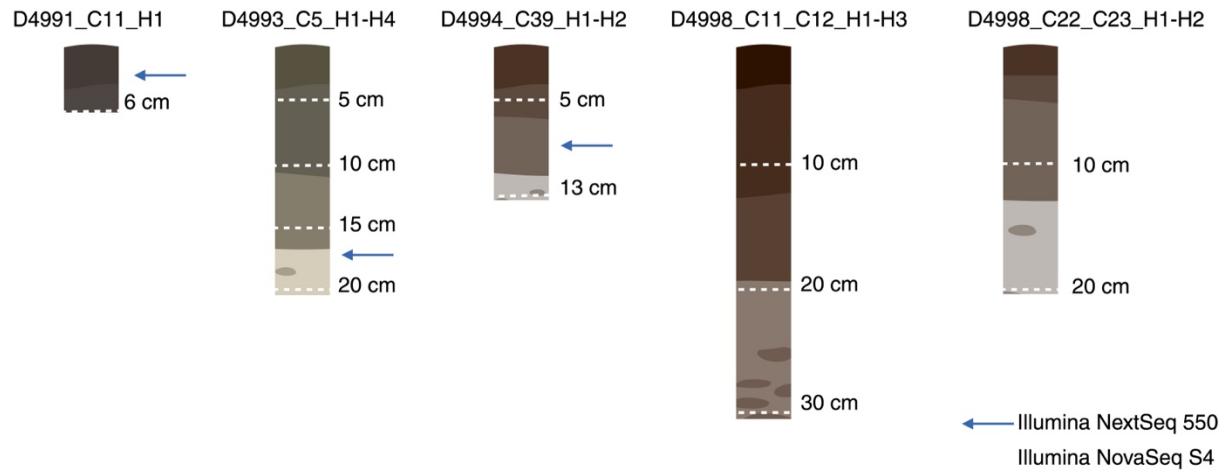


Supplementary Fig. 1 | **Schematic highlighting the steps in curation and expansion of Asgard archaea genomic catalog.** Metagenome-assembled genomes (MAGs) described in this study from the Guaymas Basin (GE2 & GE3) and Bohai Sea (BH) with the published genomes totaled 1,024 MAGs. Quality control and genomic filtering with the listed open-source software resulted in a set of 869 medium to high-quality MAGs for further analysis. Legend provides information on steps in the process and the open-source software for each step. Created in BioRender. Appler, K. (2025) <https://BioRender.com/ahcn7mk>.

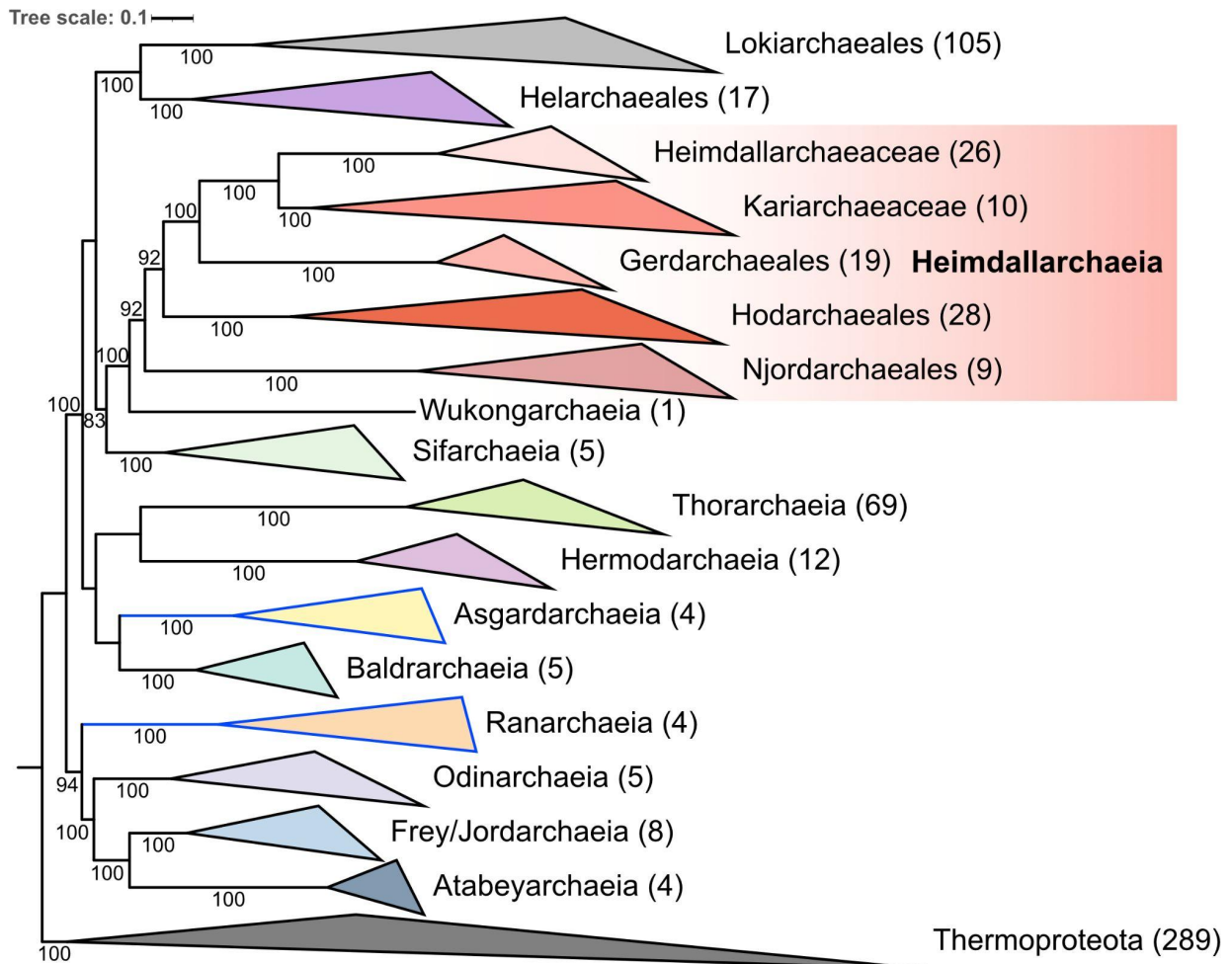
GE3

First described in this study

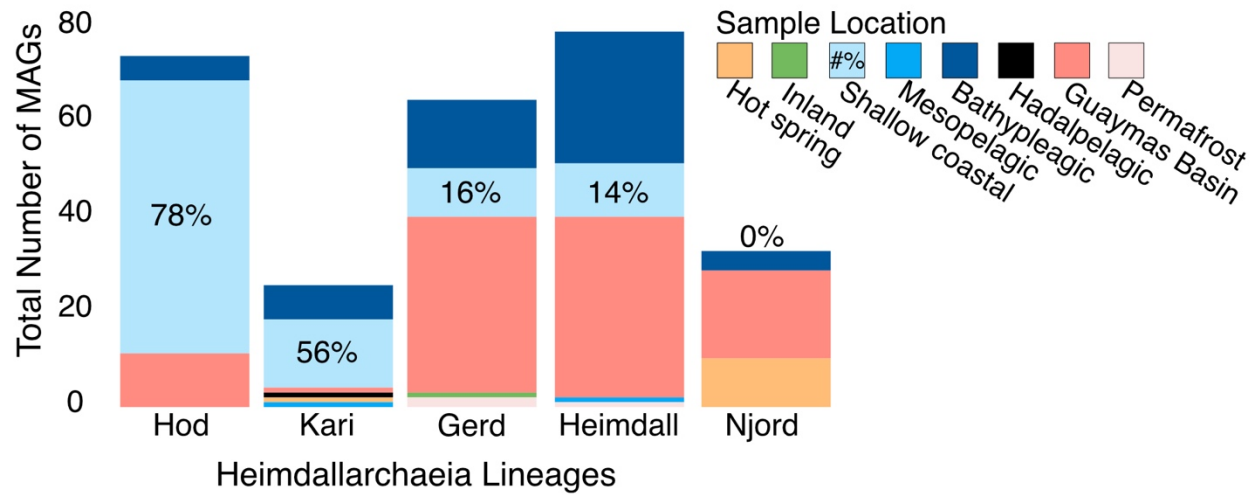
Dive_Cores_Horizons



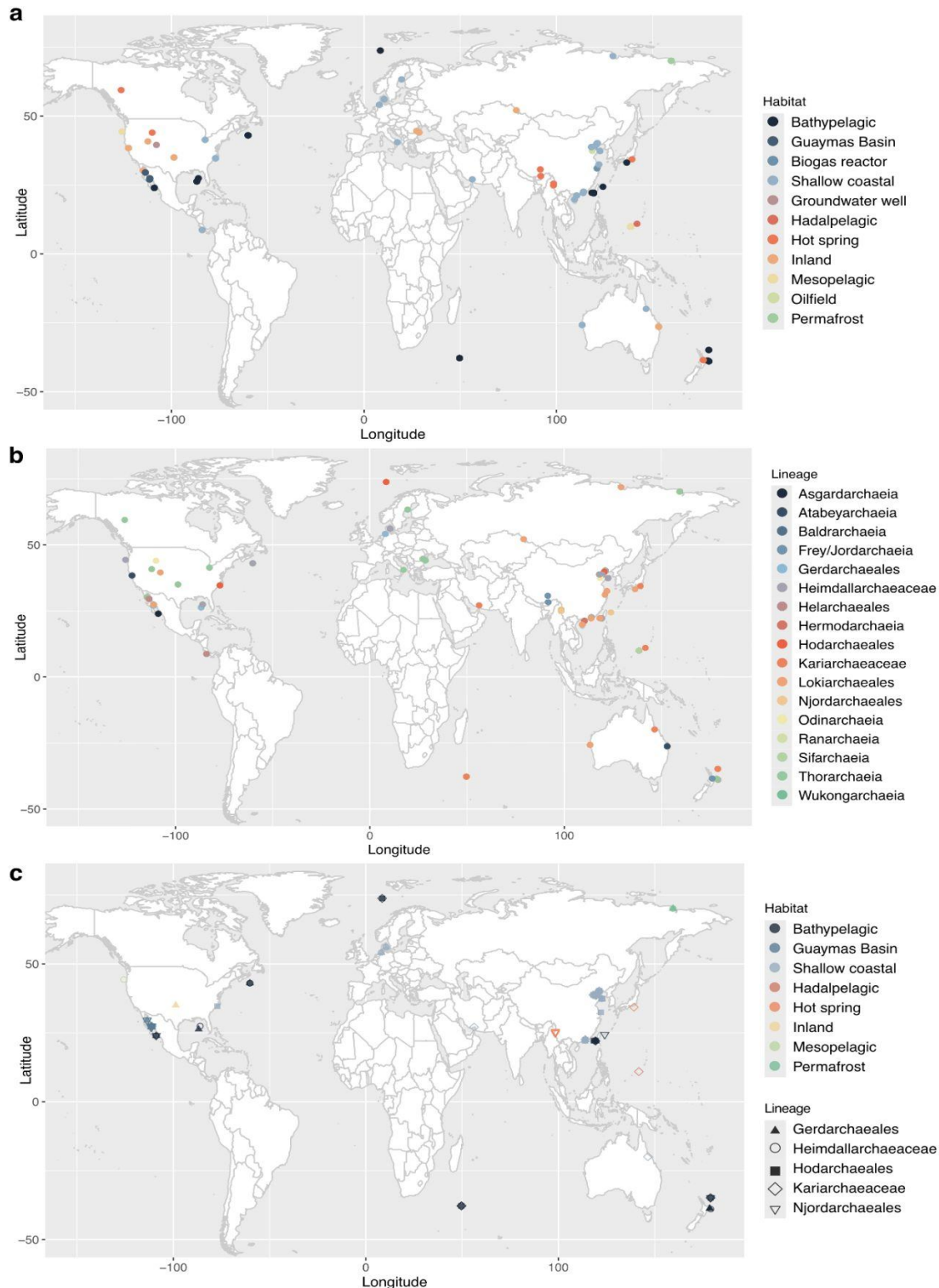
Supplementary Fig. 2 | **Schematic Guaymas Basin Sediment cores from GE3.** Displays the five cores and 12 depth horizons in sequencing project GE3, first described in this study. The blue arrow highlights the three sediment horizons that were sequenced with Illumina NextSeq 550 with all other depths using Illumina NovaSeq S4. Created in BioRender. Appler, K. (2025) <https://BioRender.com/ho68x03>.



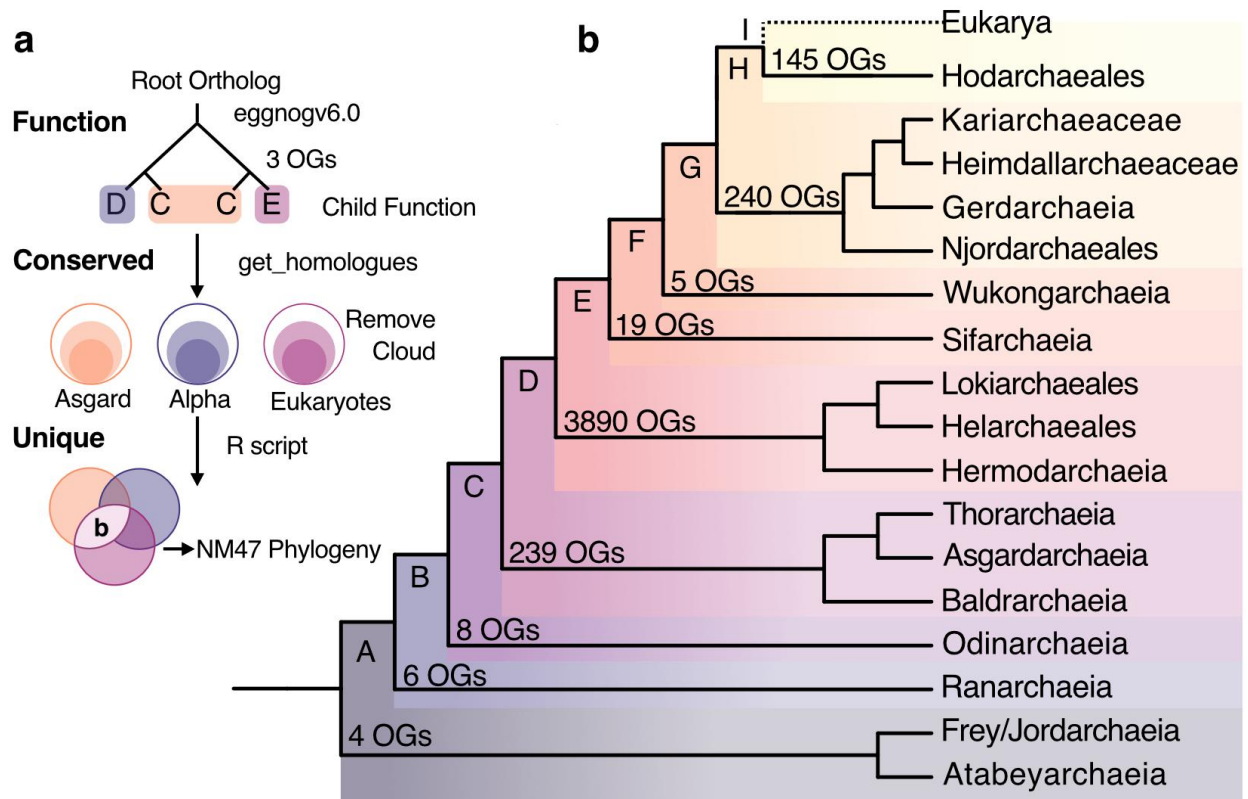
Supplementary Fig. 3 | **Phylogenetic tree of the dereplicated Asgardarchaeota database.** We dereplicated the genomic catalog shown in (Fig. 1ab) to include species-level Asgard representatives (95% similarity with fastANI). The final tree included 331 Asgardarchaeota and 289 genus-level Thermoproteota genomes (excluding Korarchaeia) and is based on a supermatrix of 15,994 amino acid positions. Sequences were aligned with mafft-linsi v7.453 and trimmed with trimAl v1.4.rev22 (-gt 0.5). The model was selected between LG+C60, Q.pfam+C60 and WAG+C60 models and rate heterogeneity components (+R4 or +G4) by IQ-TREE v2.1.3 ModelFinder. The final tree was reconstructed using a PMSF approximation (model: WAG+C60+R4) with 100 nonparametric bootstrap pseudoreplicates (see main text for detailed methods). The number of MAGs in each lineage is listed behind the collapsed branches. Blue branches highlight the two previously metabolically undescribed lineages (Asgardarchaeia and Ranarchaeia). The previously described Heimdallarchaeia orders and families are highlighted with red.



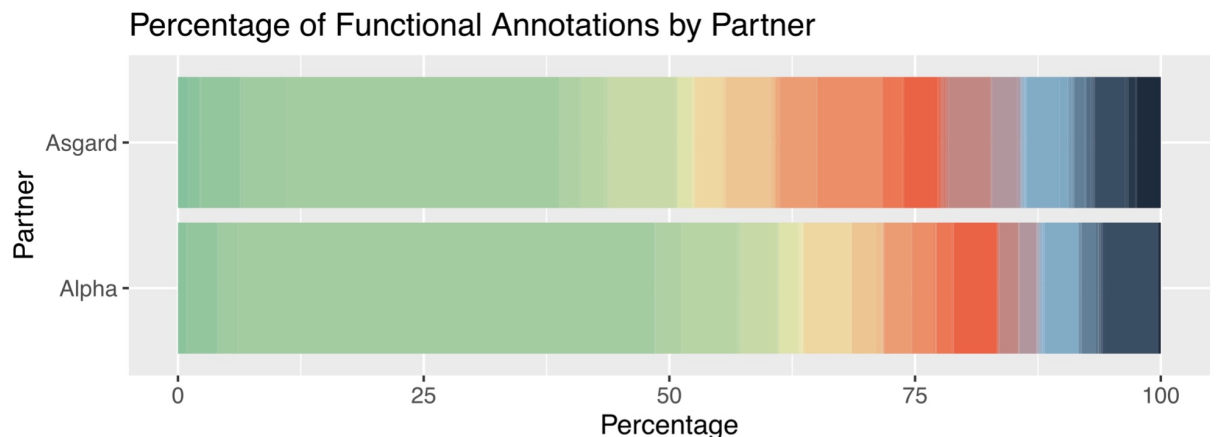
Supplementary Fig. 4 | **Environmental distribution for the five Heimdallarchaeial lineages** (Hodarchaeales, Kariarchaeaceae, Njordarchaeales, Heimdallarchaeaceae, Gerdarchaeales) determined from collected metadata detailed in Supplementary Table 4. The percentage of coastal zone (<200 m) MAGs is listed.



Supplementary Fig. 5 | **Global view of all the isolated Asgard archaea representatives used in this analysis.** Of the 869 Asgardarchaeota MAGs used in this study, we were able to identify coordinates for 817 genomes. **a.** Distribution of these Asgardarchaeota MAGs color coded by habitat **b.** and taxonomy. **c.** The 254 Heimdallarchaeia MAGs with known coordinates are color coded by the habitat from which the genome was extracted and shape for the five known lineages.



Supplementary Fig. 6 | **Conserved functional OGs unique to Asgardarchaeota lineages and single-cell eukaryotes across evolutionary branching points.** a. Schematic showing the workflow to determine conserved functional OGs unique to different taxonomic groups. b. Maximum likelihood phylogeny from Fig. 1b based on 47 concatenated non-ribosomal proteins depicting the phylogenetic relationships between all Asgard MAGs (see methods). The branch lengths have been modified and superimposed with Eukarya to map the number of OGs uniquely shared between the Asgardarchaeota lineages at each branching point with the single-cell eukaryote database (Supplementary Table 6). A dashed lined connects Eukarya to Asgardarchaeota as previously found in Eme et al., 2023. See Fig. 1b for bootstrap values and branch lengths. The subsequent figures (Supplementary Fig. 7-10), show percentage of OGs by partner, lineage, and taxonomic branching point (A-I listed), along with the number of OGs by lineage.

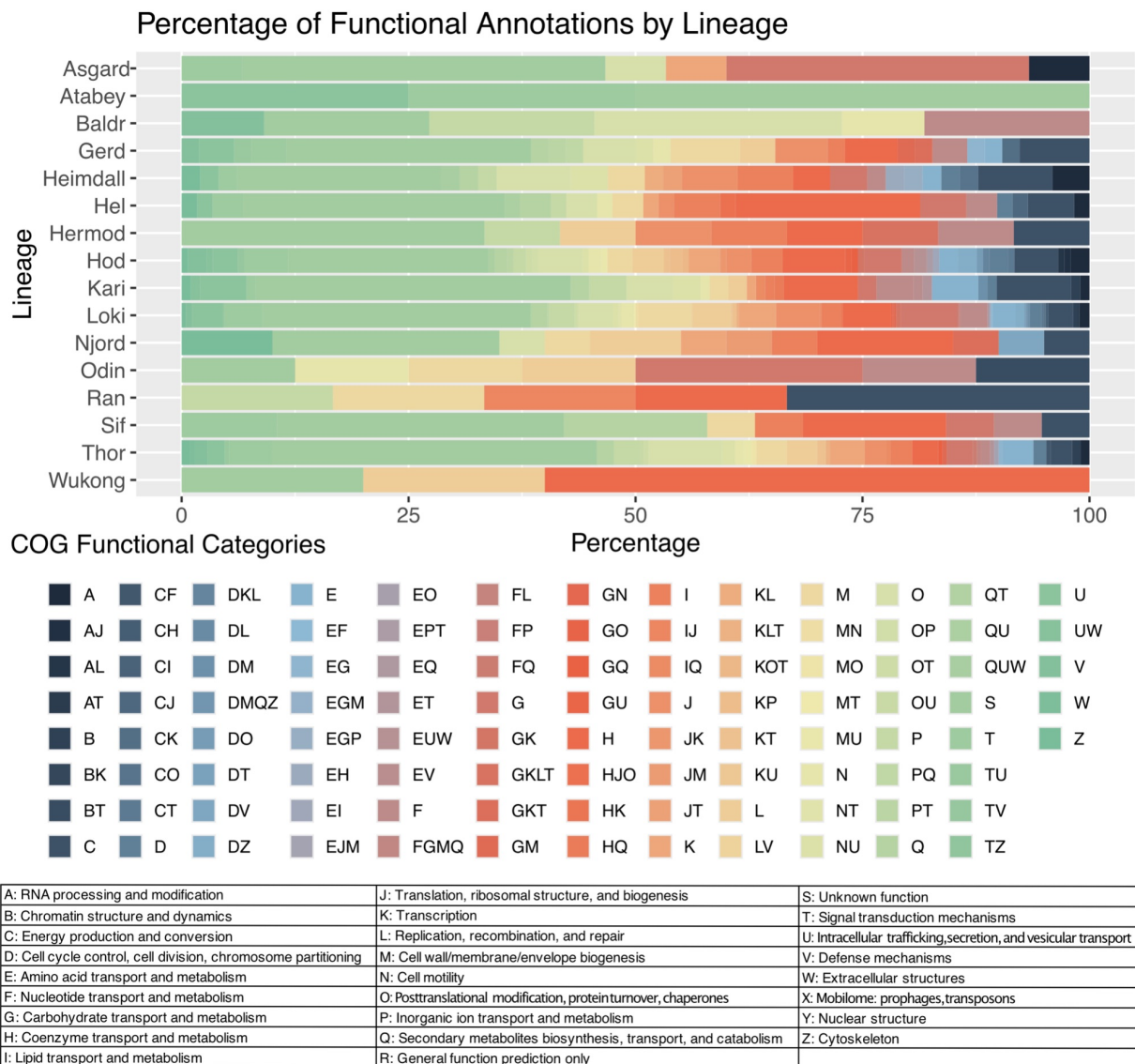


COG Functional Categories

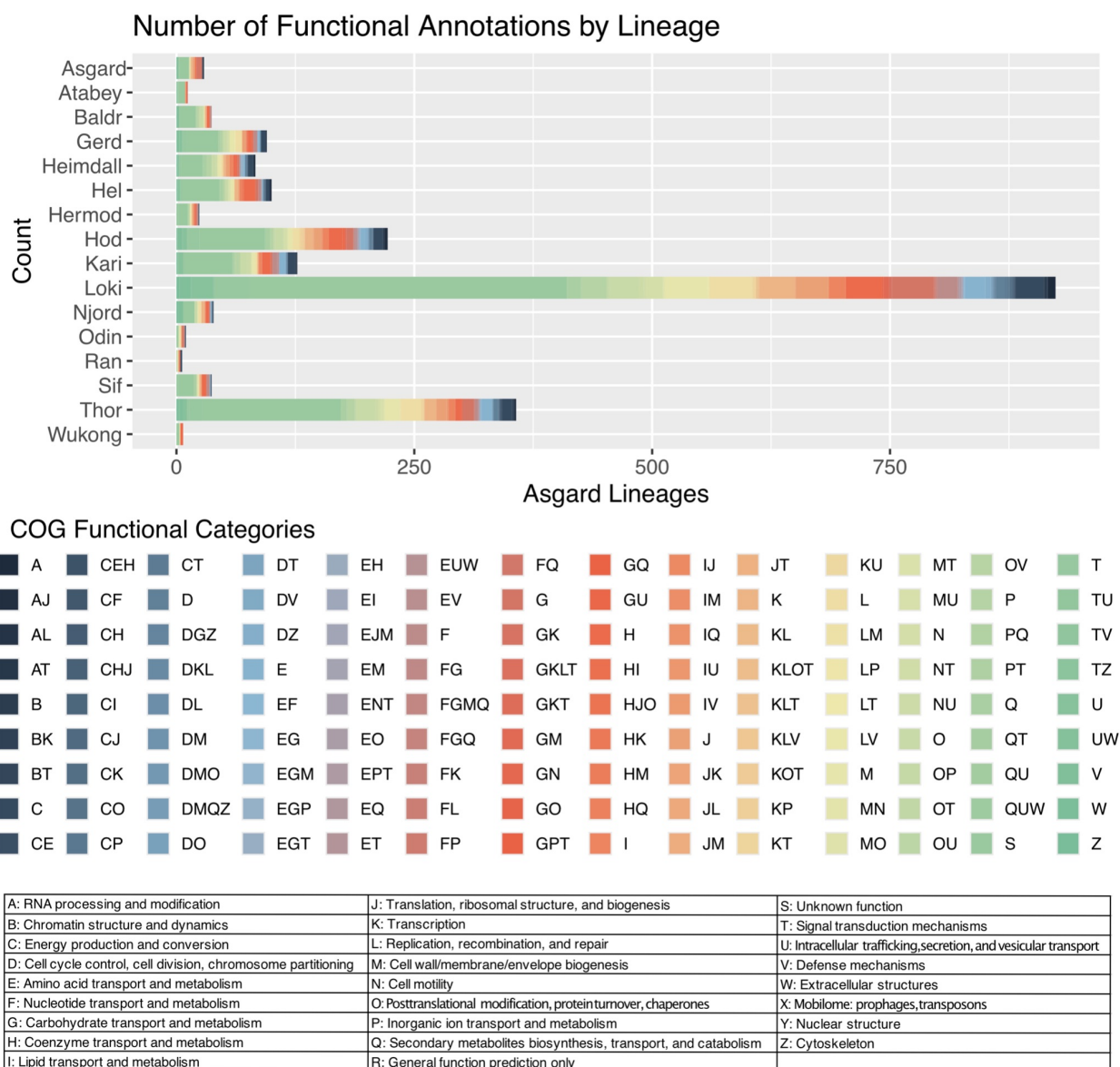
A	CEH	D	E	EO	FP	GOQ	HP	JT	KU	MQ	NU	QUW
ADK	CG	DEGMUZ	EF	EP	FQ	GOT	HQ	K	L	MCU	O	S
AJ	CH	DKL	EG	EPT	G	GOU	I	KL	LO	MT	OP	T
AL	CI	DL	EGIPQ	EQ	GH	GPU	IJ	KLO	LP	MU	OQ	TU
B	CJ	DM	EGM	ET	GK	GQ	IK	KLT	LT	MV	OT	TV
BD	CK	DMQZ	EGP	EUW	GKLT	GT	IM	KLTU	LV	N	OU	TZ
BDL	CNT	DN	EH	EV	GKT	GU	IQ	KMT	M	NO	P	U
BDLTU	CO	DO	EI	F	GL	GUW	IT	KNU	MN	NOU	PQ	UW
BK	COT	DT	EJ	FG	GM	H	IU	KOT	MNO	NPT	PT	UZ
BT	CP	DTZ	EK	FGMQ	GMOU	HI	J	KP	MO	NPTU	Q	V
C	CQ	DV	EM	FH	GN	HJK	JL	KQ	MOT	NQ	QT	W
CE	CT	DZ	ENT	FL	GO	HK	JM	KT	MP	NT	QU	Z

A: RNA processing and modification	J: Translation, ribosomal structure, and biogenesis	S: Unknown function
B: Chromatin structure and dynamics	K: Transcription	T: Signal transduction mechanisms
C: Energy production and conversion	L: Replication, recombination, and repair	U: Intracellular trafficking, secretion, and vesicular transport
D: Cell cycle control, cell division, chromosome partitioning	M: Cell wall/membrane/envelope biogenesis	V: Defense mechanisms
E: Amino acid transport and metabolism	N: Cell motility	W: Extracellular structures
F: Nucleotide transport and metabolism	O: Posttranslational modification, protein turnover, chaperones	X: Mobilome: prophages, transposons
G: Carbohydrate transport and metabolism	P: Inorganic ion transport and metabolism	Y: Nuclear structure
H: Coenzyme transport and metabolism	Q: Secondary metabolites biosynthesis, transport, and catabolism	Z: Cytoskeleton
I: Lipid transport and metabolism	R: General function prediction only	

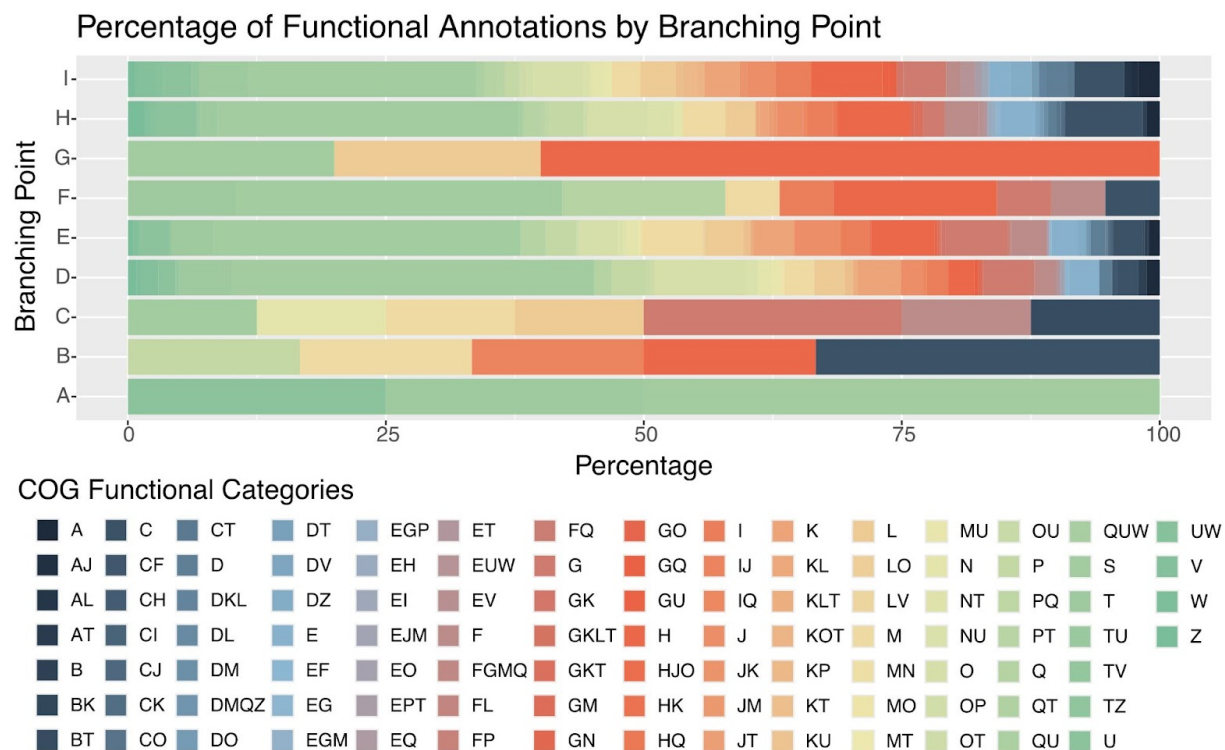
Supplementary Fig. 7 | **Intersection of functional genes between each descendant of the predicted partners in the origin of the eukaryotic cell.** Display the percentage of each COG functional category at the intersection between descendants of the chimeric eukaryotic cell, excluding the other partner. Detailed description of the annotations and genomes used in this analysis are shown in Supplementary Table 6. The table lists the letter acronyms for all COG functional categories. Abbreviations include Asgard (Asgardarchaeota) and Alpha (Alphaproteobacteria).



Supplementary Fig. 8 | **Intersection of functional genes between each named Asgardarchaeota lineage (class-family) within single cell eukaryotes.** Frey/Jordarchaeia is not shown because it did not have any unique OGs. The bar chart displays the percentage of each COG functional category, excluding all other Asgardarchaeota lineages. Detailed description of the annotations and genomes used in this analysis are shown in Supplementary Table 6. The table lists the letter acronyms for all COG functional categories. Abbreviations for Asgardarchaeota remove the taxonomic suffix.

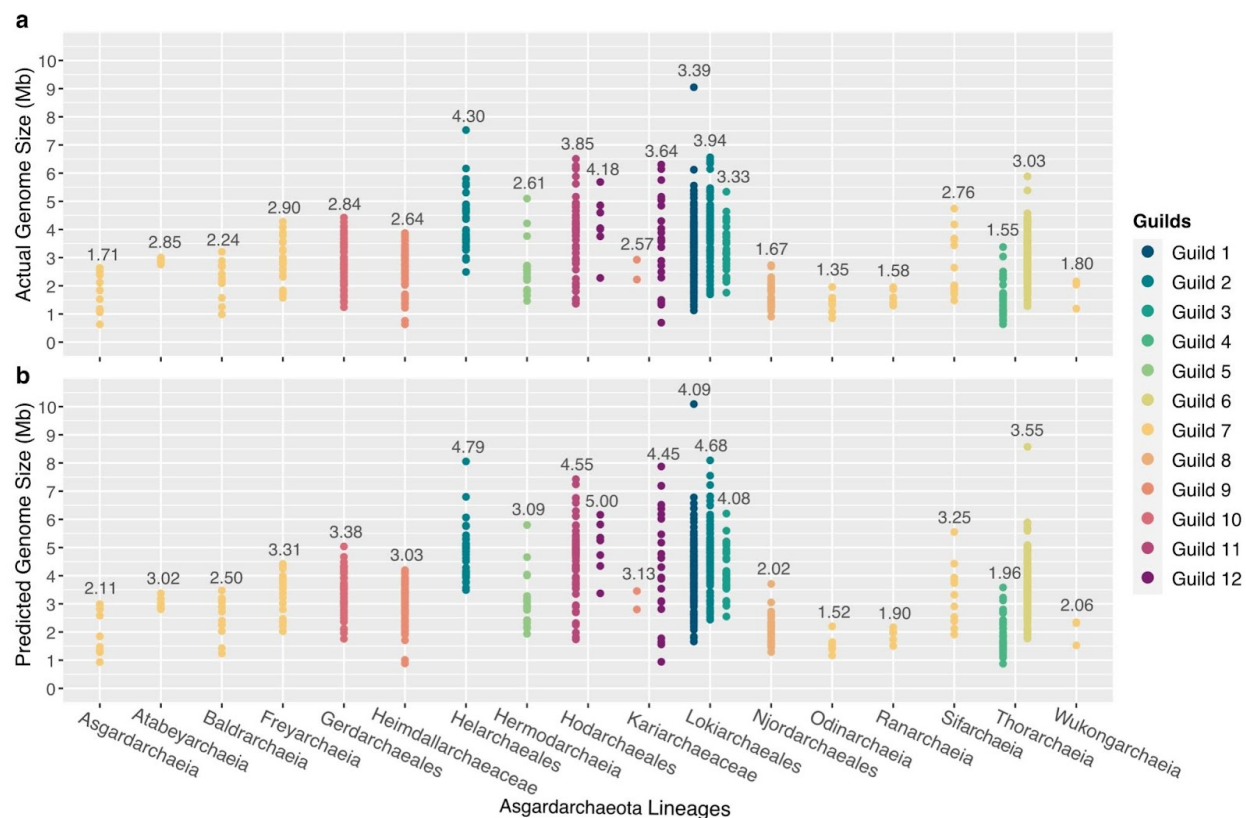


Supplementary Fig. 9 | **Unique functional genes for each Asgardarchaeota lineage (class-family).** Frey/Jordarchaeia is not shown because it did not have any unique OGs. The bar chart displays the percentage of each COG functional category, excluding all other Asgardarchaeota lineages. Detailed description of the annotations and genomes used in this analysis are shown in Supplementary Table 6. The table lists the letter acronymss for all COG functional categories. Abbreviations for Asgardarchaeota remove the taxonomic suffix.

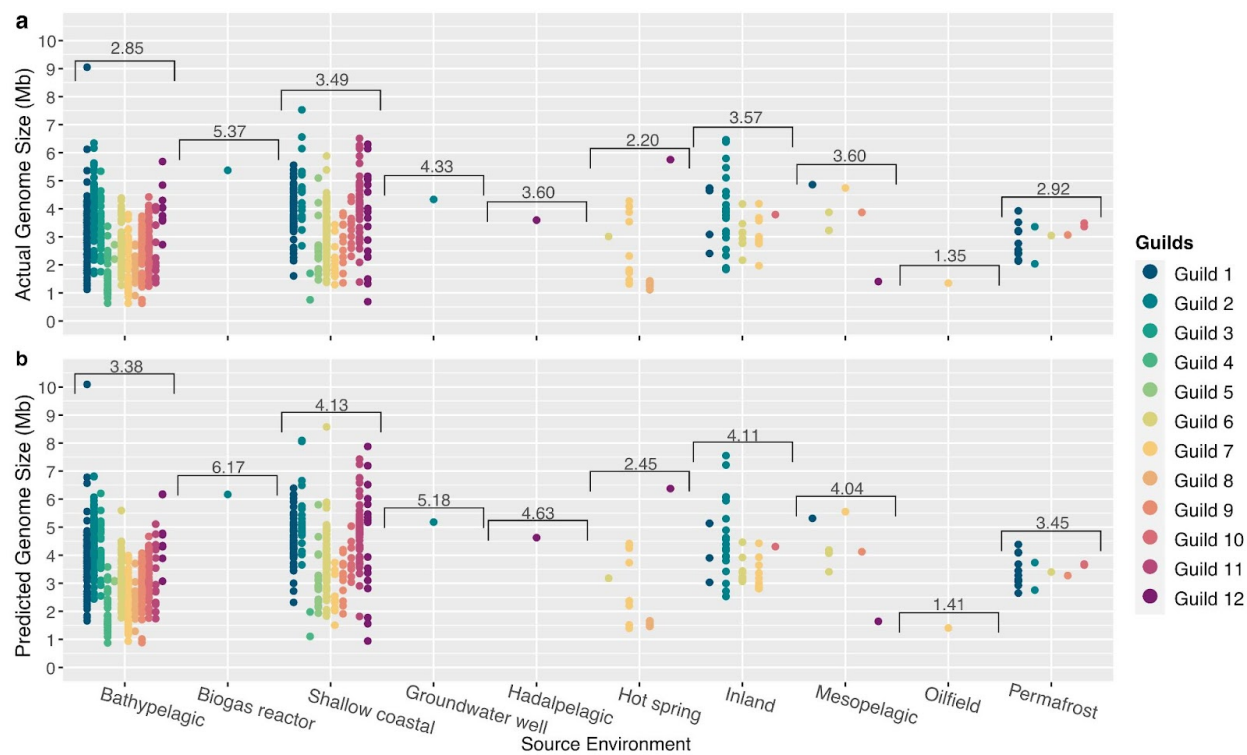


A: RNA processing and modification	J: Translation, ribosomal structure, and biogenesis	S: Unknown function
B: Chromatin structure and dynamics	K: Transcription	T: Signal transduction mechanisms
C: Energy production and conversion	L: Replication, recombination, and repair	U: Intracellular trafficking, secretion, and vesicular transport
D: Cell cycle control, cell division, chromosome partitioning	M: Cell wall/membrane/envelope biogenesis	V: Defense mechanisms
E: Amino acid transport and metabolism	N: Cell motility	W: Extracellular structures
F: Nucleotide transport and metabolism	O: Posttranslational modification, protein turnover, chaperones	X: Mobilome: prophages, transposons
G: Carbohydrate transport and metabolism	P: Inorganic ion transport and metabolism	Y: Nuclear structure
H: Coenzyme transport and metabolism	Q: Secondary metabolites biosynthesis, transport, and catabolism	Z: Cytoskeleton
I: Lipid transport and metabolism	R: General function prediction only	

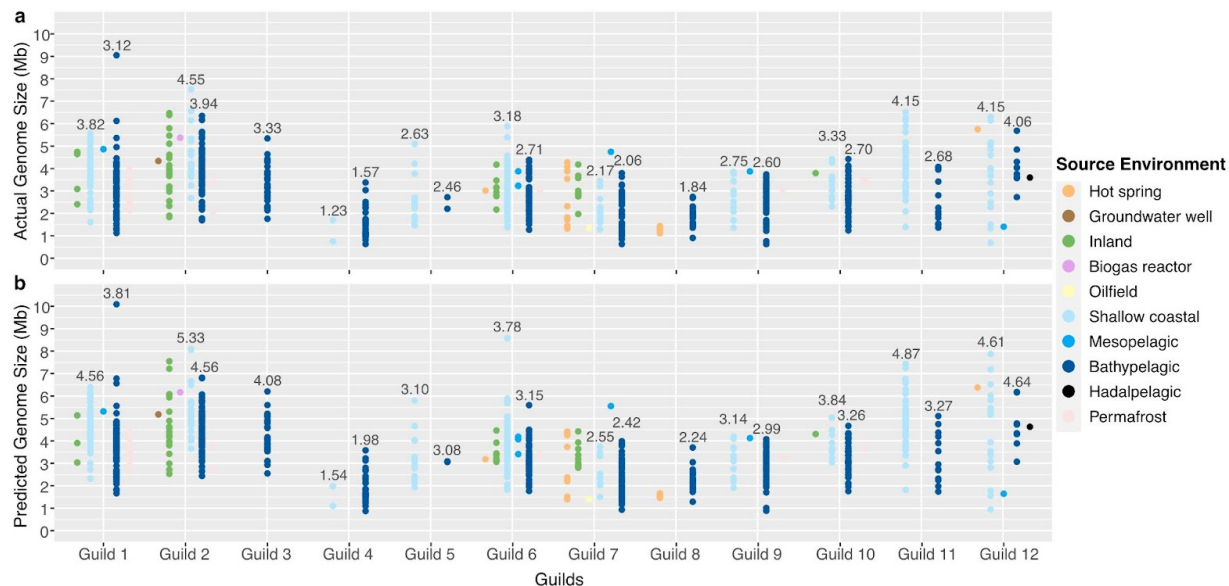
Supplementary Fig. 10 | **Functional comparative genomic NM47 phylogeny branching point.** To trace gain of shared eukaryotic OGs by Asgardarchaeota lineage we determined the percentage of COG functional categories (color) for each branching point in the NM47 phylogeny (Fig. 1b). Delineating of the branching points are labeled in Supplementary Fig. 6 (i.e., A Frey/Jordarchaeia and Atabeyarchaeia). Detailed description of the annotations and genomes used in this analysis are shown in Supplementary Table 6. The table lists the letter acronyms for all COG functional categories.



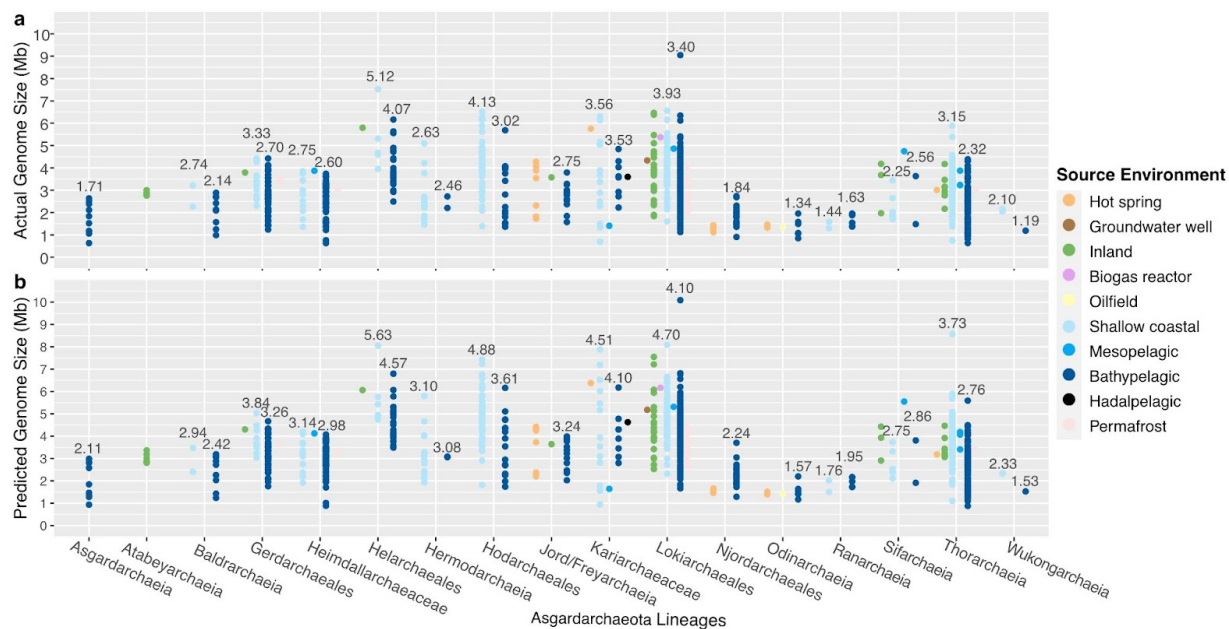
Supplementary Fig. 11 | **Actual and predicted genome size (Mb) of the Asgard genomic catalog separated by taxonomy and metabolic guild.** **a.** Actual and predicted genome size (Mb) of the Asgardarchaeota genomic catalog separated by taxonomy and metabolic guilds. Notice members of the Heimdallarchaeia (guilds 11 and 12, Hodarchaeales and Kariarchaeaceae) and Lokiarchaeia (guilds 1 and 2, Lokiarchaeales and Helarchaeales). Listed above each subset of MAGs by lineage and guild (color-coded) is the average **a.** actual genome size (Mb) and **b.** predicted genome size (Mb) corrected for CheckM v1.2.1 completeness and contamination (see methods). Data for these figures are found in Supplementary Table 4.



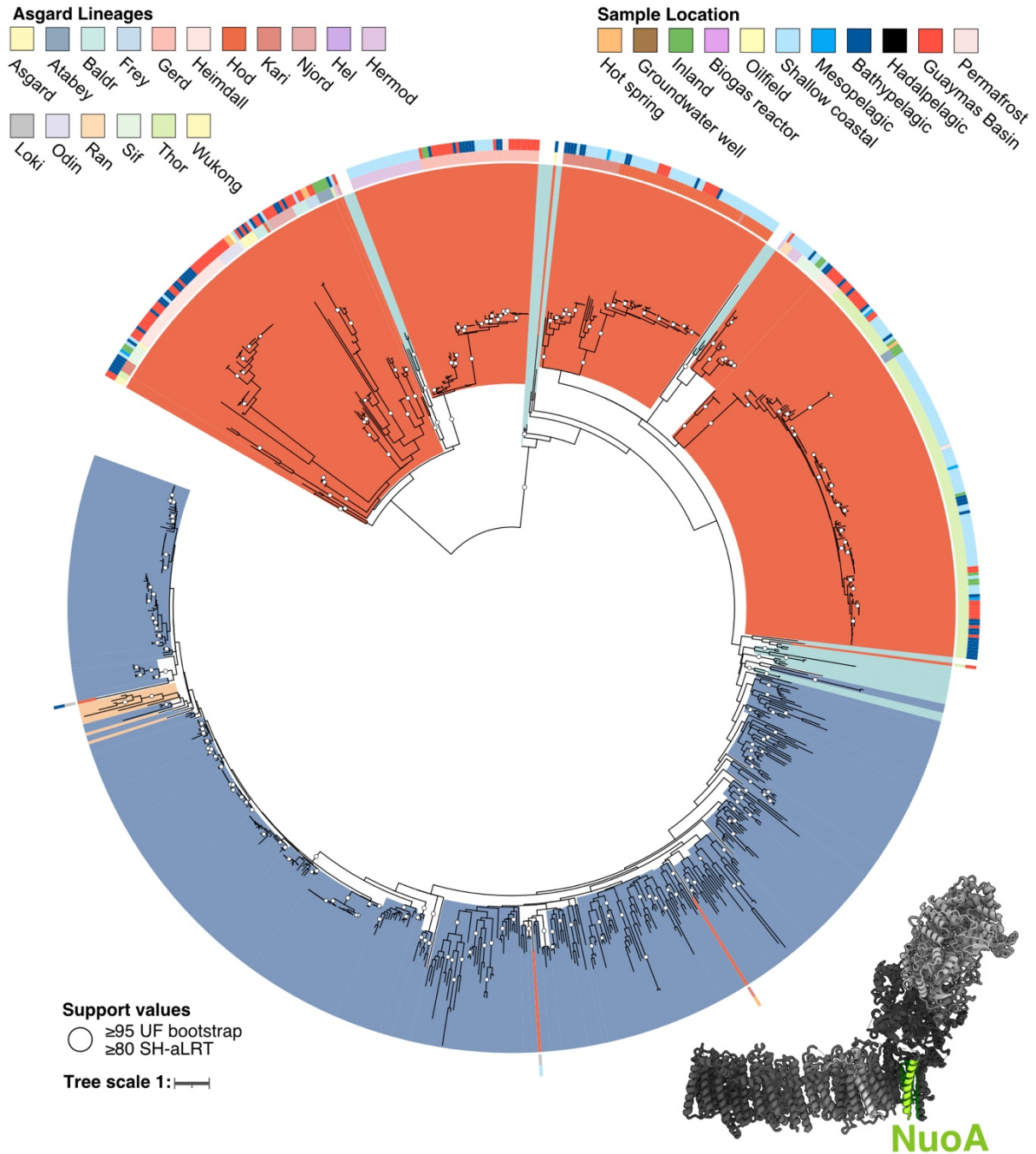
Supplementary Fig. 12 | **Actual and predicted genome size (Mb) of the Asgard genomic catalog separated by source environment and metabolic guild.** Listed above each subset of MAGs is the average **a.** actual genome size (Mb) and **b.** predicted genome size (Mb) corrected for CheckM v1.2.1 completeness and contamination (redundancy single copy markers). Data for these figures are found in Supplementary Table 4. Guaymas Basin was included within the Bathypelagic environment due to depth (~2000 mbsl).



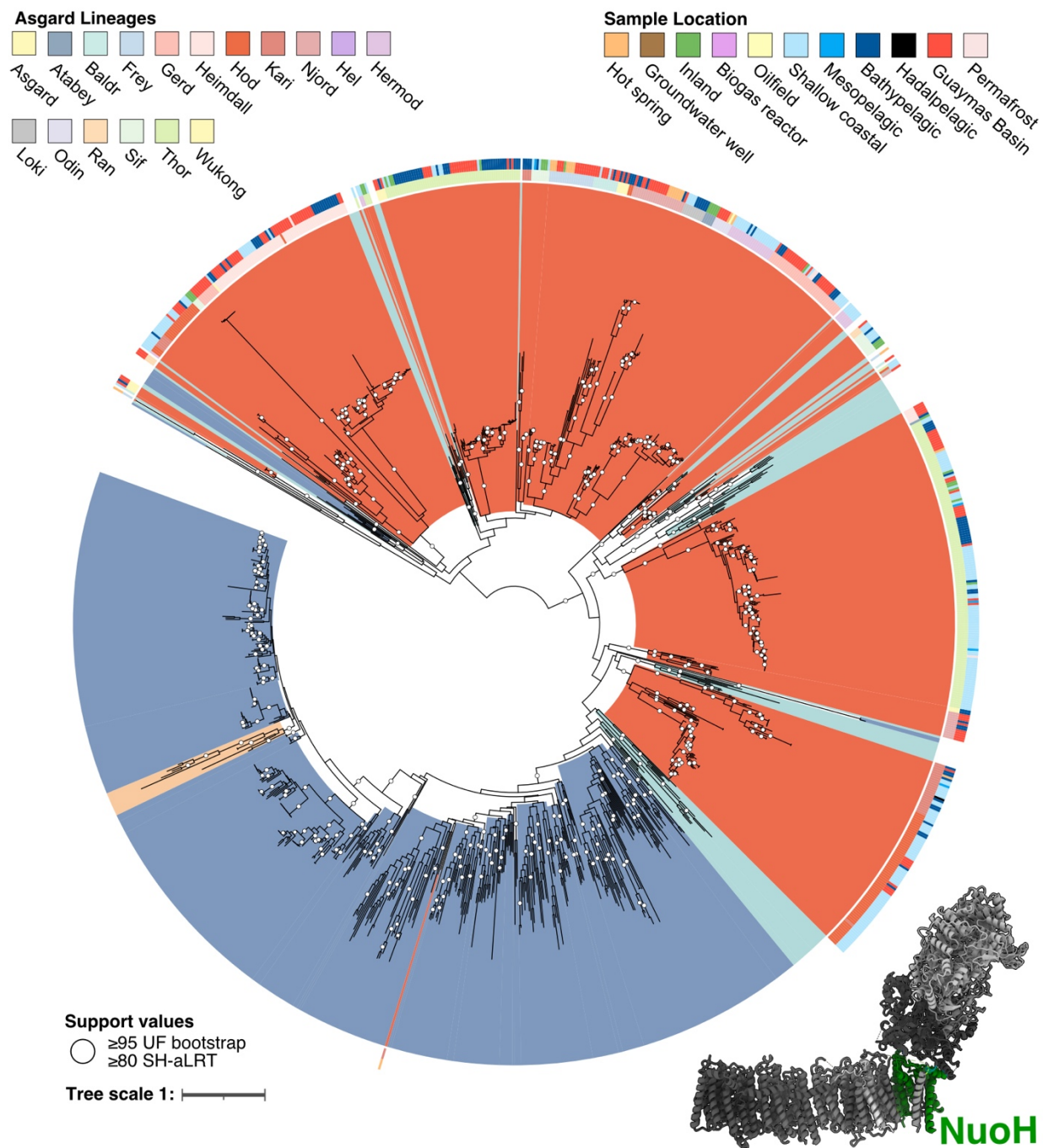
Supplementary Fig. 13 | **Actual and predicted genome size (Mb) of the Asgard genomic catalog separated by metabolic guild and source environment.** Listed above each subset of MAGs by guild is the average **a.** actual genome size (Mb) and **b.** predicted genome size (Mb) corrected for CheckM v1.2.1 completeness and contamination (redundancy single copy markers) for coastal (light blue) and bathypelagic (dark blue). Data for these figures are found in Supplementary Table 4. Guaymas Basin was included within the Bathypelagic environment due to depth (~2000 mbsl).



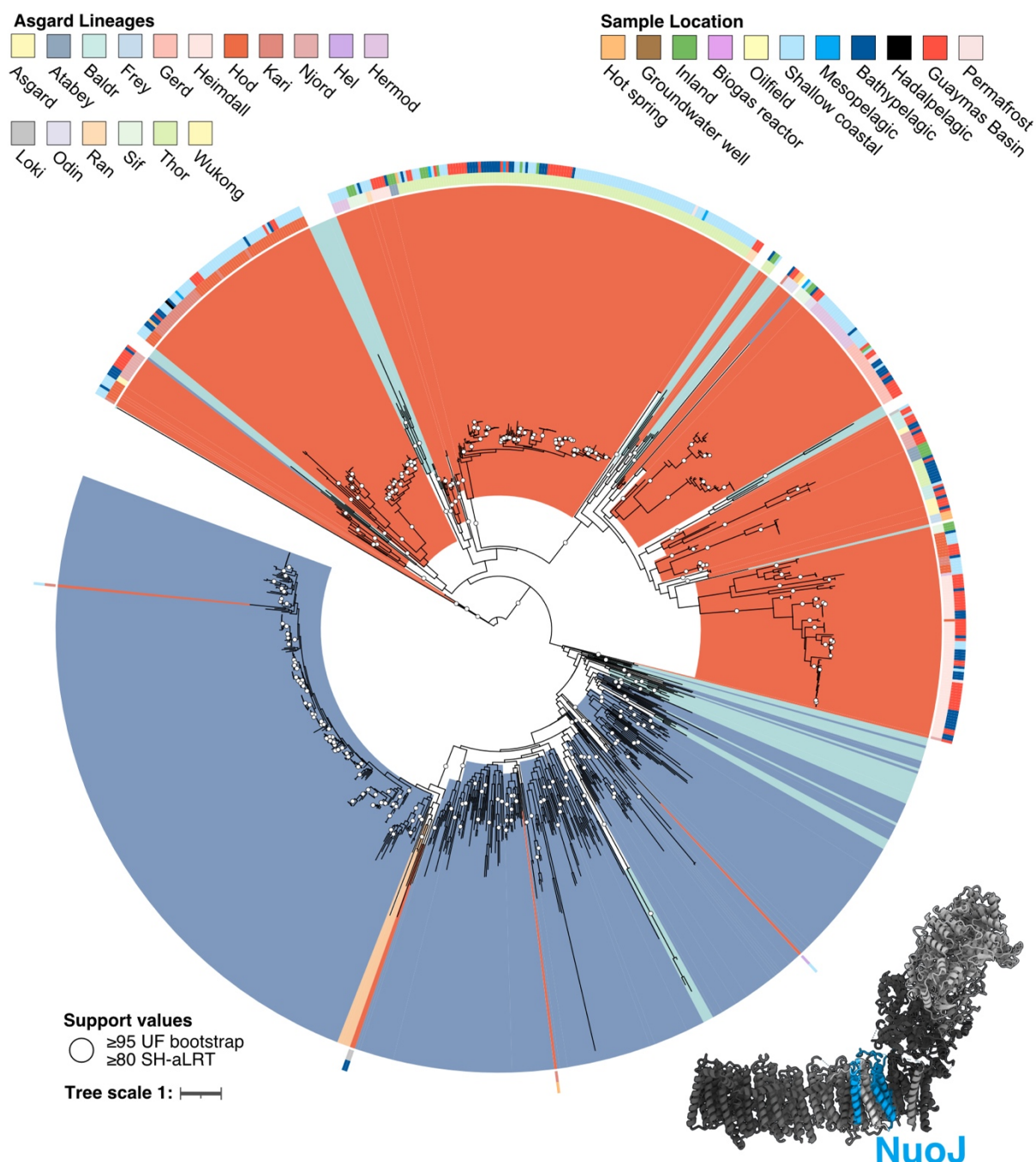
Supplementary Fig. 14 | **Actual and predicted genome size (Mb) of the Asgard genomic catalog separated by taxonomy and source environment.** Listed above each subset of MAGs by taxonomy is the average **a.** actual genome size (Mb) and **b.** predicted genome size (Mb) corrected for CheckM v1.2.1 completeness and contamination (redundancy single copy markers) for coastal (light blue) and bathypelagic (dark blue). Data for these figures are found in Supplementary Table 4. Guaymas Basin was included within the Bathypelagic environment due to depth (~2000 mbsl).



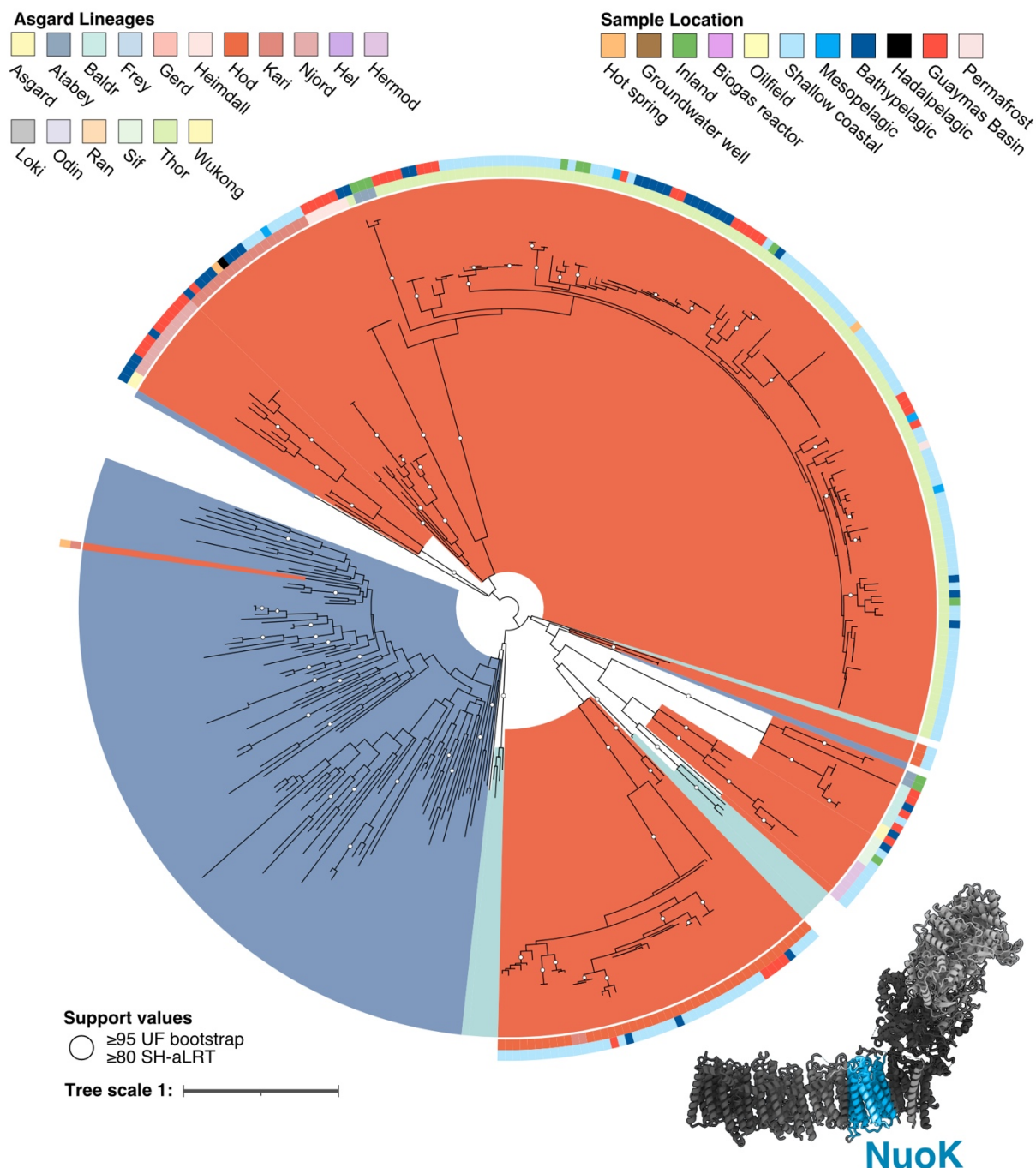
Supplementary Fig. 15 | **P-module subunit NADH:quinone oxidoreductase subunit A (NuoA) phylogeny.** Maximum-likelihood phylogeny was inferred from a supermatrix of 913 sequences and 119 amino acid positions (from 735 initial positions), displaying the diversity of NuoA subunits across Asgard archaea (red), other archaeal lineages (teal), bacteria (dark blue), and mitochondria (orange). The tree was generated using IQ-TREE v2.0.7 (1000 ultrafast bootstrap replicates and calculation of 1000 SH-aLRT) with WAG+C50+R8 best-fit model chosen according to Bayesian Information Criterion (BIC). The two external rings show the Asgard lineage and source location, respectively. Support values are shown when ultrafast (UF) bootstrap support is greater than or equal to 95 and SH-aLRT support is greater than or equal to 80. The phylogeny was visualized with iTOL and midpoint rooted. The bottom right-hand corner shows the NuoA subunit in the *E. coli* reference PDB (7P62).



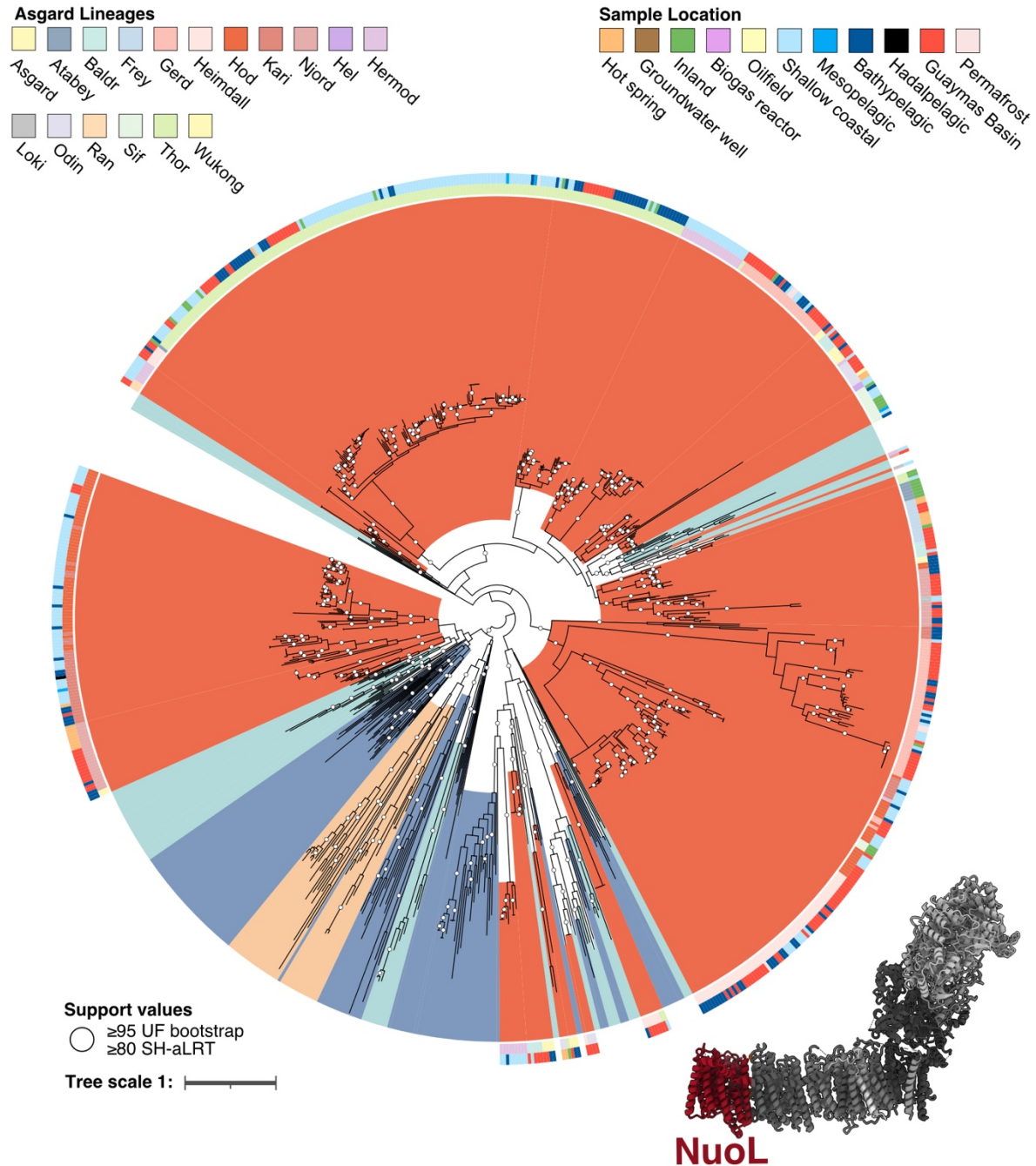
Supplementary Fig. 16 | **P-module subunit NADH:quinone oxidoreductase subunit H (NuoH) phylogeny.** Maximum-likelihood phylogeny was inferred from a supermatrix of 1230 sequences and 323 amino acid positions (from 1808 initial positions), displaying the diversity of NuoH subunits across Asgard archaea (red), other archaeal lineages (teal), bacteria (dark blue), and mitochondria (orange). The tree was generated using IQ-TREE v2.0.7 (1000 ultrafast bootstrap replicates and calculation of 1000 SH-aLRT) with WAG+C50+R8 best-fit model chosen according to Bayesian Information Criterion (BIC). The two external rings show the Asgard lineage and source location, respectively. Support values are shown when ultrafast (UF) bootstrap support is greater than or equal to 95 and SH-aLRT support is greater than or equal to 80. The phylogeny was visualized with iTOL and midpoint rooted. The bottom right-hand corner shows the NuoH subunit in the *E. coli* reference PDB (7P62).



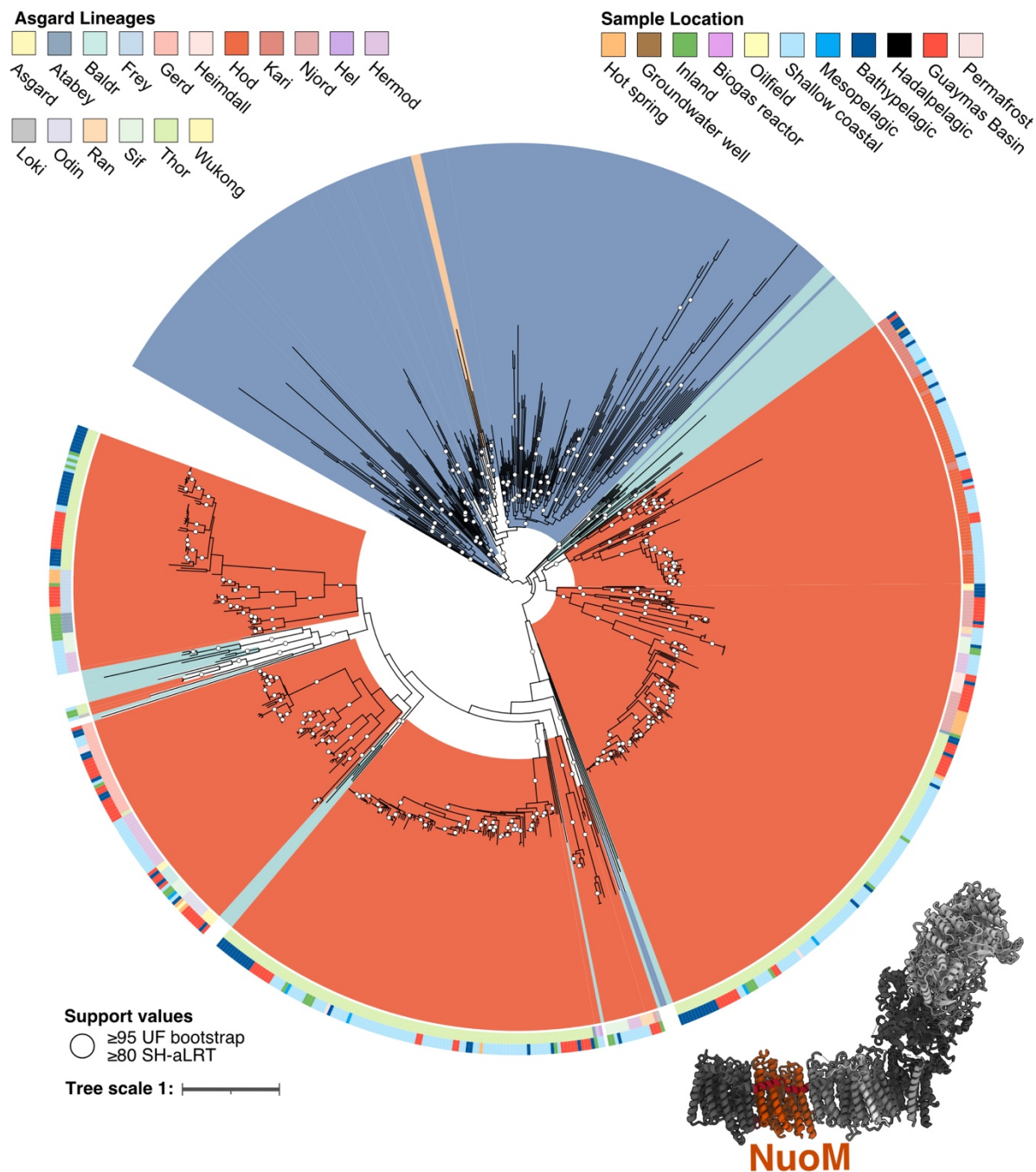
Supplementary Fig. 17 | **P-module subunit NADH:quinone oxidoreductase subunit J (NuoJ) phylogeny.** Maximum-likelihood phylogeny was inferred from a supermatrix of 1017 sequences and 166 amino acid positions (from 1153 initial positions), displaying the diversity of NuoJ subunits across Asgard archaea (red), other archaeal lineages (teal), bacteria (dark blue), and mitochondria (orange). The tree was generated using IQ-TREE v2.0.7 (1000 ultrafast bootstrap replicates and calculation of 1000 SH-aLRT) with LG+C50+R8 best-fit model chosen according to Bayesian Information Criterion (BIC). The two external rings show the Asgard lineage and source location, respectively. Support values are shown when ultrafast (UF) bootstrap support is greater than or equal to 95 and SH-aLRT support is greater than or equal to 80. The phylogeny was visualized with iTOL and midpoint rooted. The bottom right-hand corner shows the NuoJ subunit in the *E. coli* reference PDB (7P62).



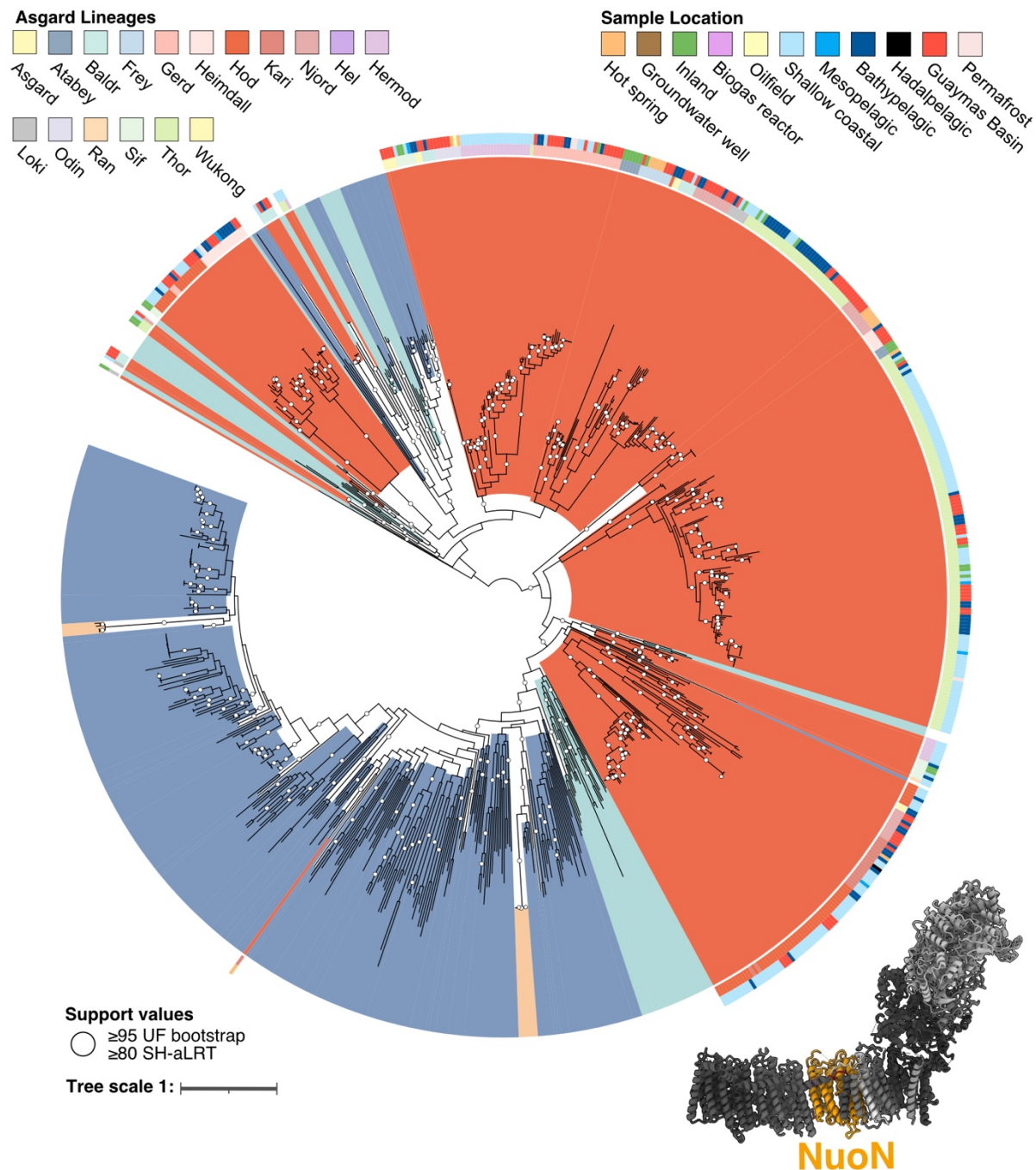
Supplementary Fig. 18 | **P-module subunit NADH:quinone oxidoreductase subunit K (NuoK) phylogeny.** Maximum-likelihood phylogeny was inferred from a supermatrix of 360 sequences and 102 amino acid positions (from 196 initial positions), displaying the diversity of NuoK subunits across Asgard archaea (red), other archaeal lineages (teal), and bacteria (dark blue). The tree was generated using IQ-TREE v2.0.7 (1000 ultrafast bootstrap replicates and calculation of 1000 SH-aLRT) with LG+C50+R6 best-fit model chosen according to Bayesian Information Criterion (BIC). The two external rings show the Asgard lineage and source location, respectively. Support values are shown when ultrafast (UF) bootstrap support is greater than or equal to 95 and SH-aLRT support is greater than or equal to 80. The phylogeny was visualized with iTOL and midpoint rooted. The bottom right-hand corner shows the NuoK subunit in the *E. coli* reference PDB (7P62).



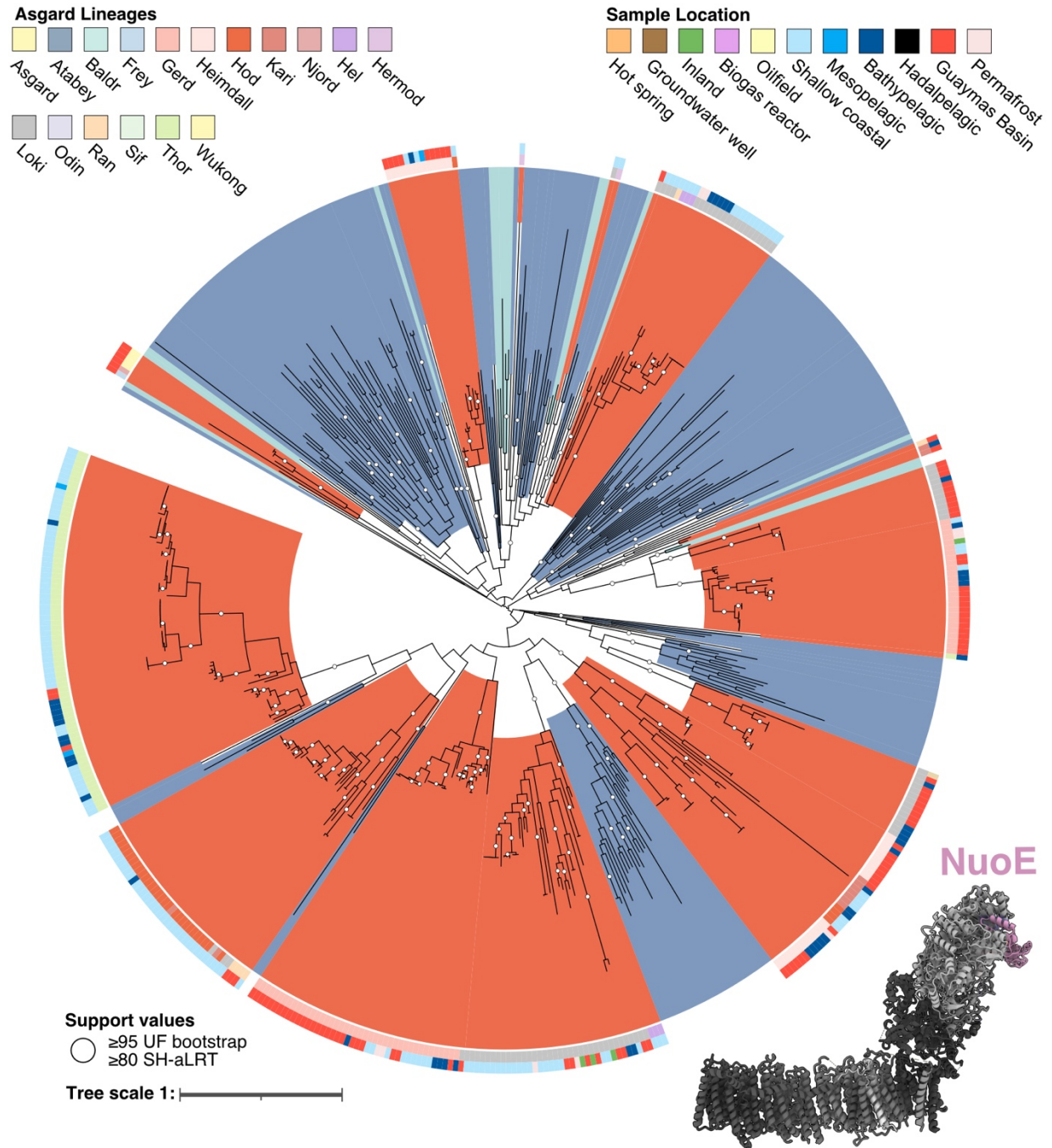
Supplementary Fig. 19 | **P-module subunit NADH:quinone oxidoreductase subunit L (NuoL) phylogeny.** Maximum-likelihood phylogeny was inferred from a supermatrix of 862 sequences and 631 amino acid positions (from 3769 initial positions), displaying the diversity of NuoL subunits across Asgard archaea (red), other archaeal lineages (teal), bacteria (dark blue), and mitochondria (orange). The tree was generated using IQ-TREE v2.0.7 (1000 ultrafast bootstrap replicates and calculation of 1000 SH-aLRT) with WAG+C50+R8 best-fit model chosen according to Bayesian Information Criterion (BIC). The two external rings show the Asgard lineage and source location, respectively. Support values are shown when ultrafast (UF) bootstrap support is greater than or equal to 95 and SH-aLRT support is greater than or equal to 80. The phylogeny was visualized with iTOL and midpoint rooted. The bottom right-hand corner shows the NuoL subunit in the *E. coli* reference PDB (7P62).



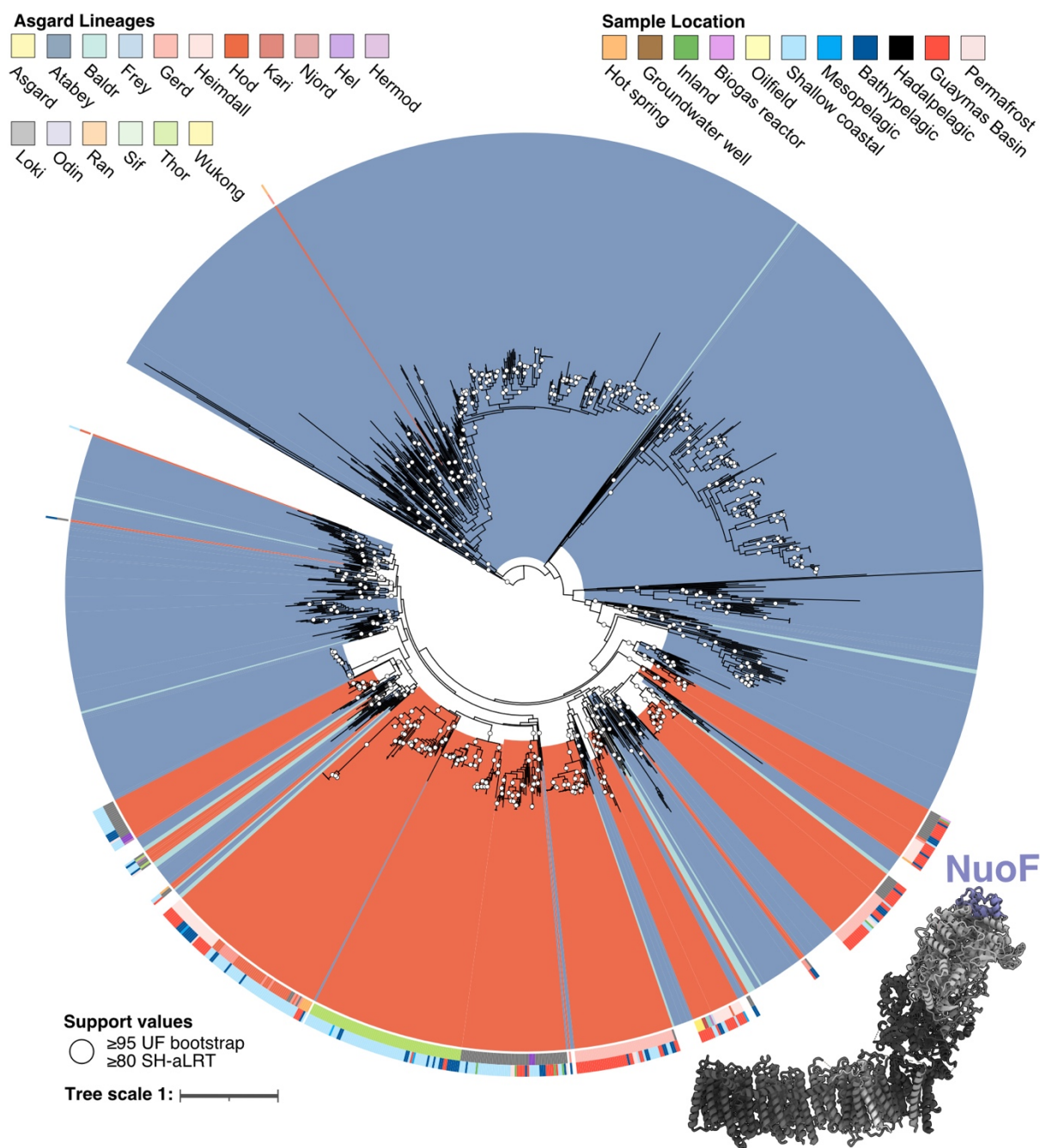
Supplementary Fig. 20 | **P-module subunit NADH:quinone oxidoreductase subunit M (NuoM) phylogeny.** Maximum-likelihood phylogeny was inferred from a supermatrix of 833 sequences and 488 amino acid positions (from 3133 initial positions), displaying the diversity of NuoM subunits across Asgard archaea (red), other archaeal lineages (teal), bacteria (dark blue), and mitochondria (orange). The tree was generated using IQ-TREE v2.0.7 (1000 ultrafast bootstrap replicates and calculation of 1000 SH-aLRT) with WAG+C50+R8 best-fit model chosen according to Bayesian Information Criterion (BIC). The two external rings show the Asgard lineage and source location, respectively. Support values are shown when ultrafast (UF) bootstrap support is greater than or equal to 95 and SH-aLRT support is greater than or equal to 80. The phylogeny was visualized with iTOL and midpoint rooted. The bottom right-hand corner shows the NuoM subunit in the *E. coli* reference PDB (7P62).



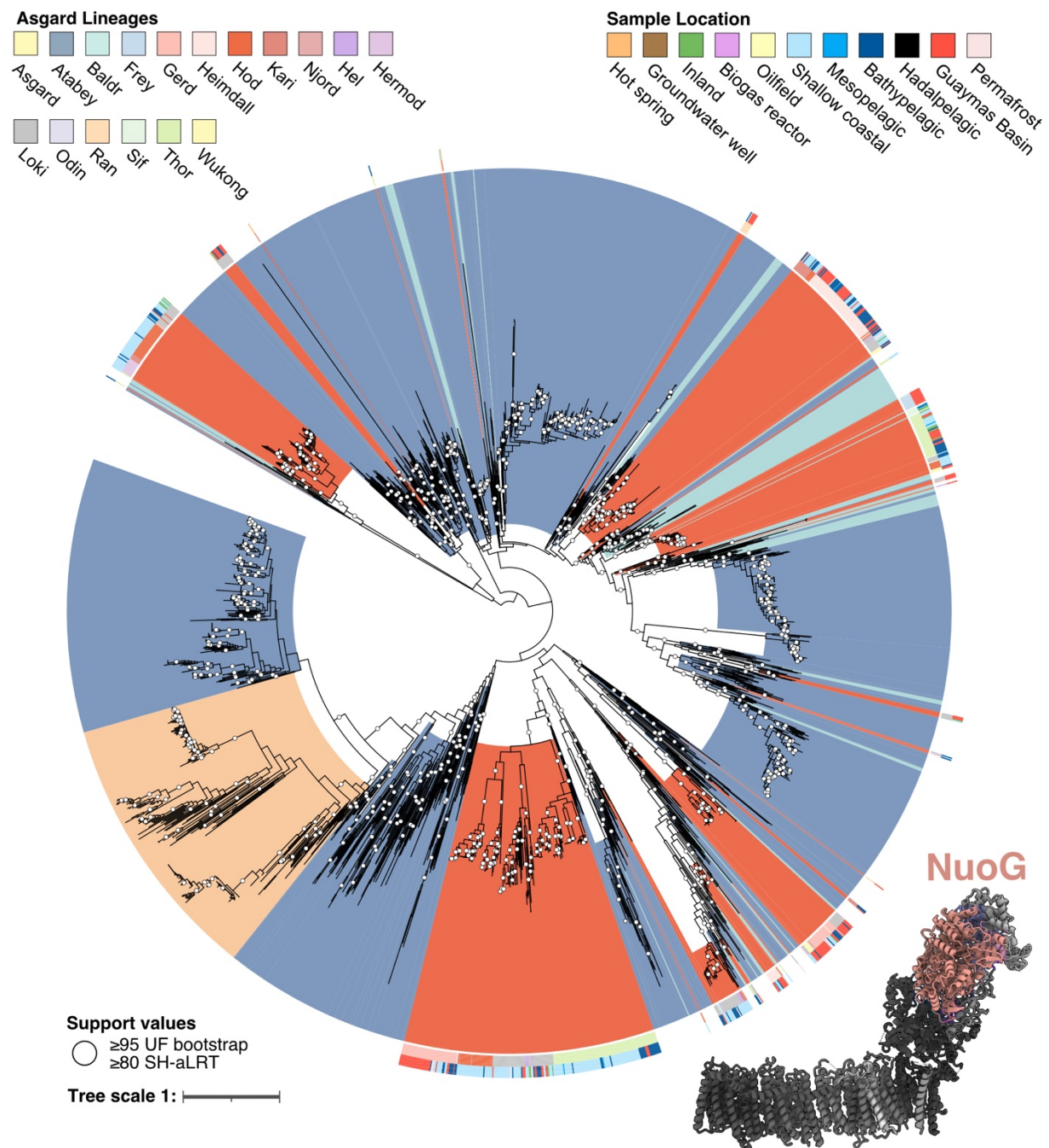
Supplementary Fig. 21 | **P-module subunit NADH:quinone oxidoreductase subunit N (NuoN) phylogeny.** Maximum-likelihood phylogeny was inferred from a supermatrix of 840 sequences and 480 amino acid positions (from 3211 initial positions), displaying the diversity of NuoN subunits across Asgard archaea (red), other archaeal lineages (teal), bacteria (dark blue), and mitochondria (orange). The tree was generated using IQ-TREE v2.0.7 (1000 ultrafast bootstrap replicates and calculation of 1000 SH-aLRT) with WAG+C50+R8 best-fit model chosen according to Bayesian Information Criterion (BIC). The two external rings show the Asgard lineage and source location, respectively. Support values are shown when ultrafast (UF) bootstrap support is greater than or equal to 95 and SH-aLRT support is greater than or equal to 80. The phylogeny was visualized with iTOL and midpoint rooted. The bottom right-hand corner shows the NuoN subunit in the *E. coli* reference PDB (7P62).



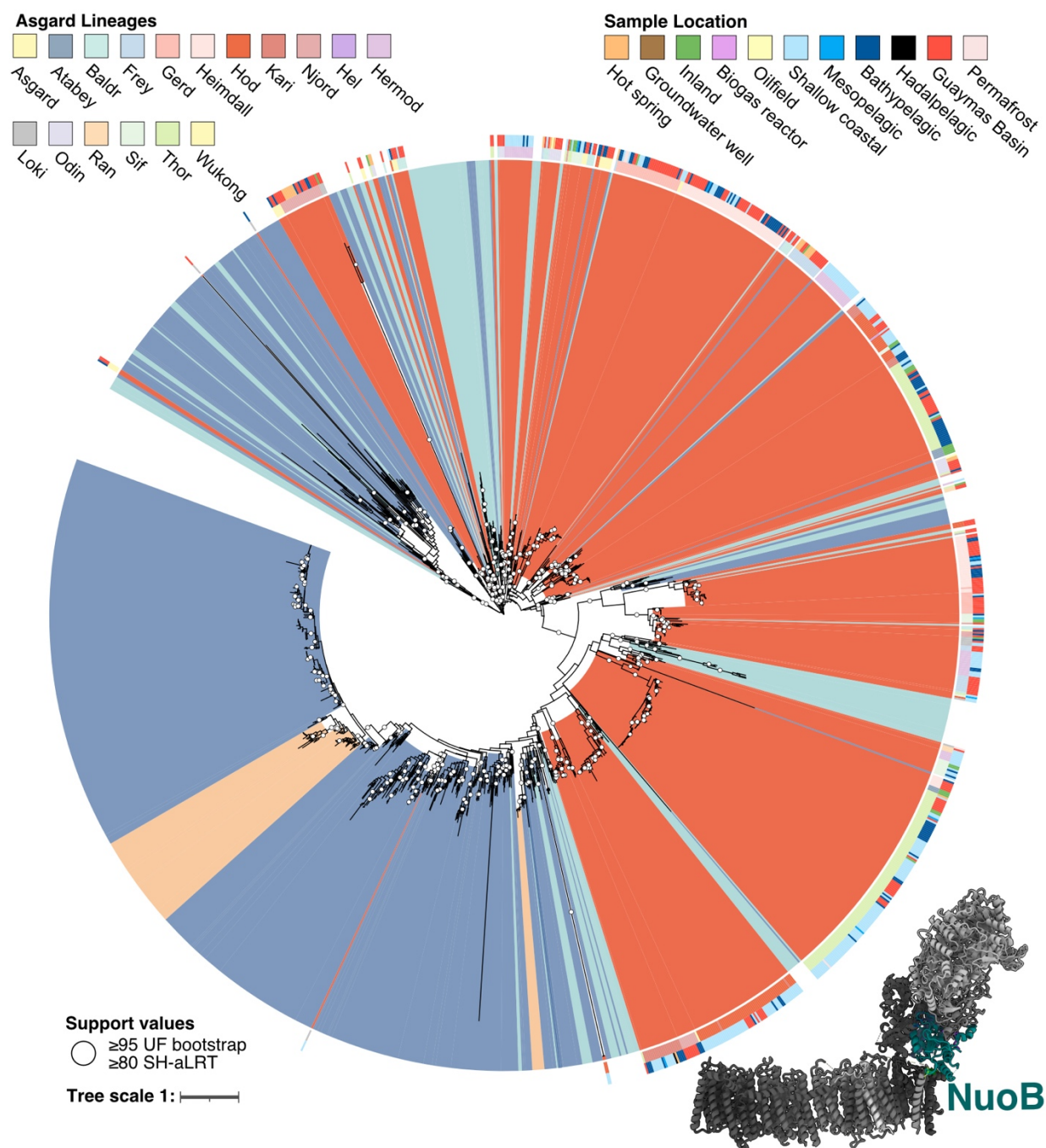
Supplementary Fig. 22 | **N-module subunit NADH:quinone oxidoreductase subunit E (NuoE) phylogeny.** Maximum likelihood phylogeny was inferred from a supermatrix of 530 sequences and 161 amino acid positions (from 1788 initial positions), displaying the diversity of NuoE subunits across Asgard archaea (red), other archaeal lineages (teal), and bacteria (dark blue). The tree was generated using IQ-TREE v2.0.7 (1000 ultrafast bootstrap replicates and calculation of 1000 SH-aLRT) with WAG+C60+R8 best-fit model chosen according to Bayesian Information Criterion (BIC). The two external rings show the Asgard lineage and source location, respectively. Support values are shown when ultrafast (UF) bootstrap support is greater than or equal to 95 and SH-aLRT support is greater than or equal to 80. The phylogeny was visualized with iTOL and midpoint rooted. The bottom right-hand corner shows the NuoE subunit in the *E. coli* reference PDB (7P62).



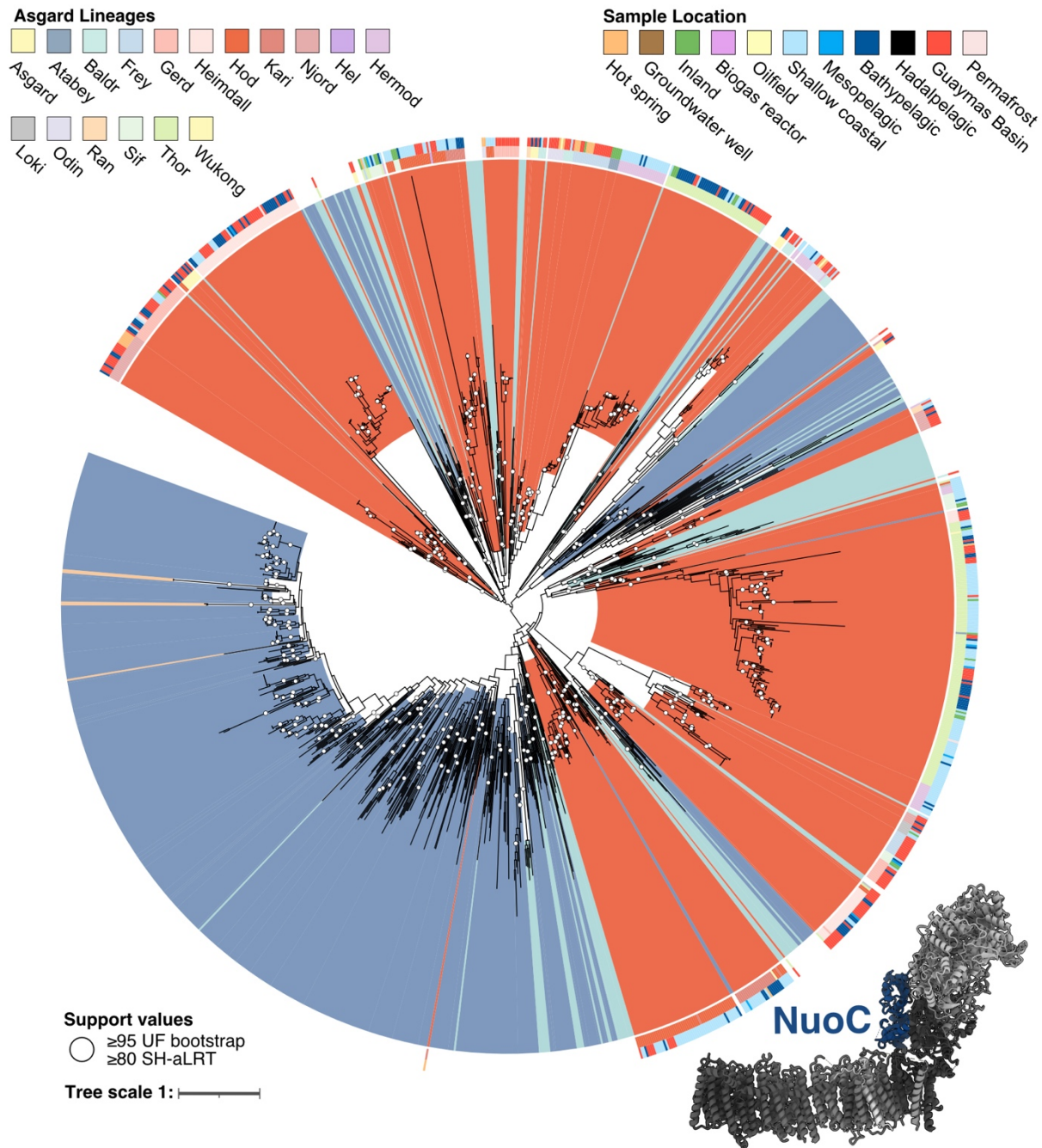
Supplementary Fig. 23 | **N-module subunit NADH:quinone oxidoreductase subunit F (NuoF) phylogeny.** Maximum-likelihood phylogeny was inferred from a supermatrix of 1420 sequences and 514 amino acid positions (from 3334 initial positions), displaying the diversity of NuoF subunits across Asgard archaea (red), other archaeal lineages (teal), and bacteria (dark blue). No mitochondrial genomes passed the threshold for this subunit. The tree was generated using IQ-TREE v2.0.7 (1000 ultrafast bootstrap replicates and calculation of 1000 SH-aLRT) with WAG+C50+R8 best-fit model chosen according to Bayesian Information Criterion (BIC). The two external rings show the Asgard lineage and source location, respectively. Support values are shown when ultrafast (UF) bootstrap support is greater than or equal to 95 and SH-aLRT support is greater than or equal to 80. The phylogeny was visualized with iTOL and midpoint rooted. The bottom right-hand corner shows the NuoF subunit in the *E. coli* reference PDB (7P62).



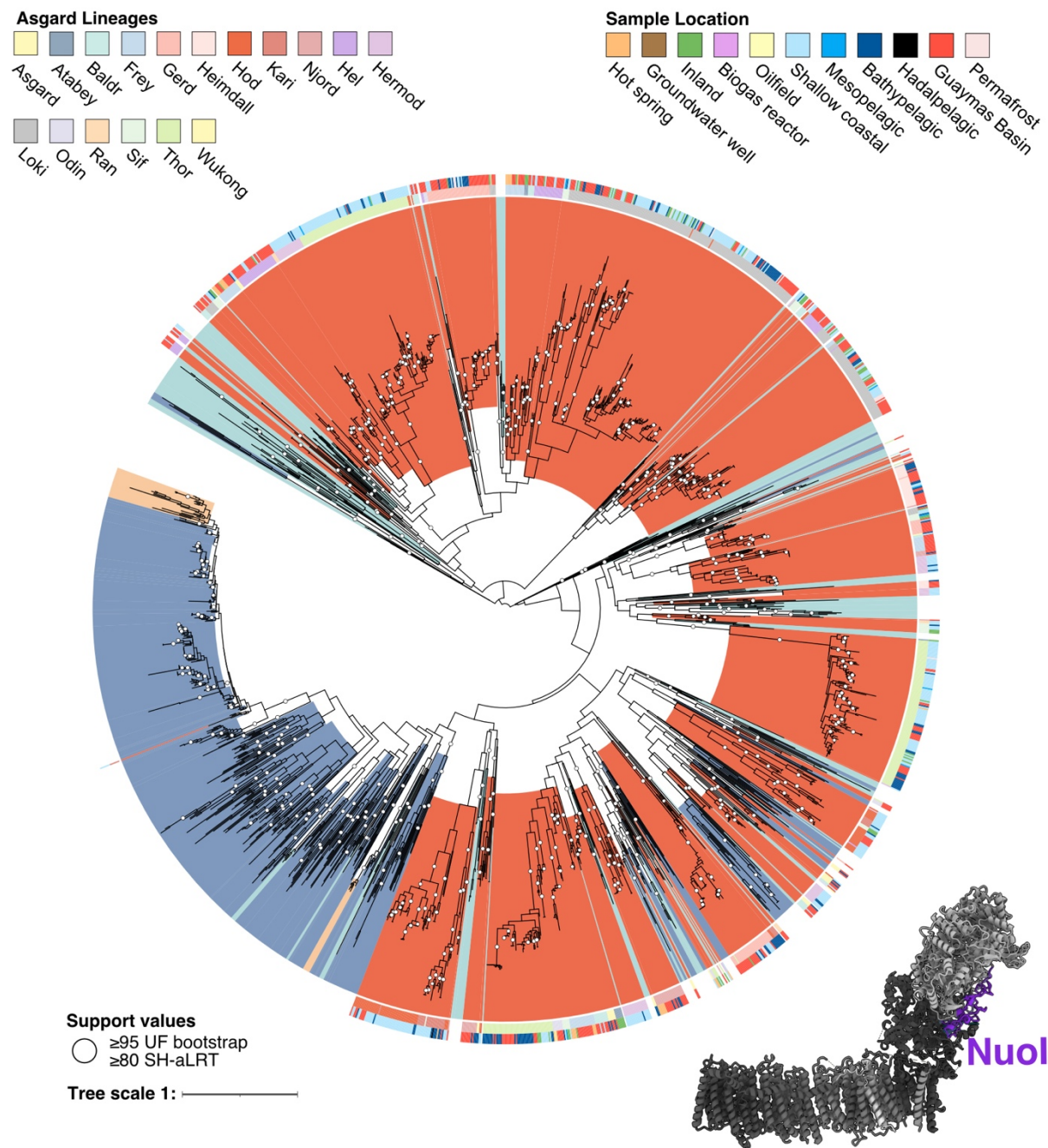
Supplementary Fig. 24 | **N-module subunit NADH:quinone oxidoreductase subunit G (NuoG) phylogeny.** Maximum-likelihood phylogeny was inferred from a supermatrix of 2530 sequences and 648 amino acid positions (from 8625 initial positions), displaying the diversity of NuoG subunits across Asgard archaea (red), other archaeal lineages (teal), bacteria (dark blue), and mitochondria (orange). The tree was generated using IQ-TREE v2.0.7 (1000 ultrafast bootstrap replicates and calculation of 1000 SH-aLRT) with WAG+C50+R8 best-fit model chosen according to Bayesian Information Criterion (BIC). The two external rings show the Asgard lineage and source location, respectively. Support values are shown when ultrafast (UF) bootstrap support is greater than or equal to 95 and SH-aLRT support is greater than or equal to 80. The phylogeny was visualized with iTOL and midpoint rooted. The bottom right-hand corner shows the NuoG subunit in the *E. coli* reference PDB (7P62).



Supplementary Fig. 25 | **Q-module subunit NADH:quinone oxidoreductase subunit B (NuoB) phylogeny.** Maximum-likelihood phylogeny was inferred from a supermatrix of 1600 sequences and 153 amino acid positions (from 2547 initial positions), displaying the diversity of NuoB subunits across Asgard archaea (red), other archaeal lineages (teal), bacteria (dark blue), and mitochondria (orange). The tree was generated using IQ-TREE v2.0.7 (1000 ultrafast bootstrap replicates and calculation of 1000 SH-aLRT) with Q.pfam+C60+R8 best-fit model chosen according to Bayesian Information Criterion (BIC). The two external rings show the Asgard lineage and source location, respectively. Support values are shown when ultrafast (UF) bootstrap support is greater than or equal to 95 and SH-aLRT support is greater than or equal to 80. The phylogeny was visualized with iTOL and midpoint rooted. The bottom right-hand corner shows the NuoB subunit in the *E. coli* reference PDB (7P62).



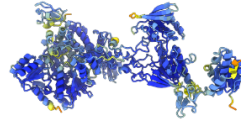
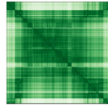
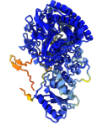
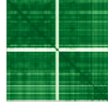
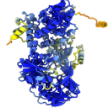
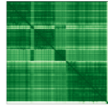
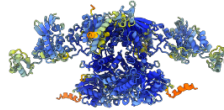
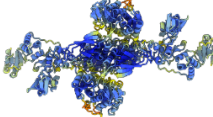
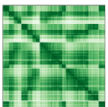
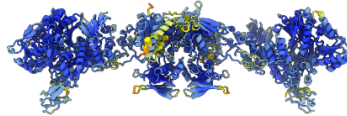
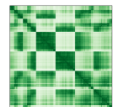
Supplementary Fig. 26 | **Q-module subunit NADH:quinone oxidoreductase subunit C (NuoC) phylogeny.** Maximum-likelihood phylogeny was inferred from a supermatrix of 1435 sequences and 164 amino acid positions (from 1886 initial positions), displaying the diversity of NuoC subunits across Asgard archaea (red), other archaeal lineages (teal), bacteria (dark blue), and mitochondria (orange). The tree was generated using IQ-TREE v2.0.7 (1000 ultrafast bootstrap replicates and calculation of 1000 SH-aLRT) with WAG+C60+R8 best-fit model chosen according to Bayesian Information Criterion (BIC). The two external rings show the Asgard lineage and source location, respectively. Support values are shown when ultrafast (UF) bootstrap support is greater than or equal to 95 and SH-aLRT support is greater than or equal to 80. The phylogeny was visualized with iTOL and midpoint rooted. The bottom right-hand corner shows the NuoC subunit in the *E. coli* reference PDB (7P62).



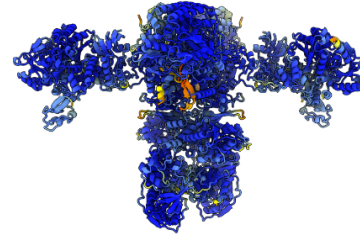
Supplementary Fig. 27 | **Q-module subunit NADH:quinone oxidoreductase subunit I (NuoI) phylogeny.** Maximum-likelihood phylogeny was inferred from a supermatrix of 2060 sequences and 137 amino acid positions (from 5287 initial positions), displaying the diversity of NuoI subunits across Asgard archaea (red), other archaeal lineages (teal), bacteria (dark blue), and mitochondria (orange). The tree was generated using IQ-TREE v2.0.7 (1000 ultrafast bootstrap replicates and calculation of 1000 SH-aLRT) with WAG+C60+R8 best-fit model chosen according to Bayesian Information Criterion (BIC). The two external rings show the Asgard lineage and source location, respectively. Support values are shown when ultrafast (UF) bootstrap support is greater than or equal to 95 and SH-aLRT support is greater than or equal to 80. The phylogeny was visualized with iTOL and midpoint rooted. The bottom right-hand corner shows the NuoI subunit in the *E. coli* reference PDB (7P62).

a

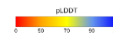
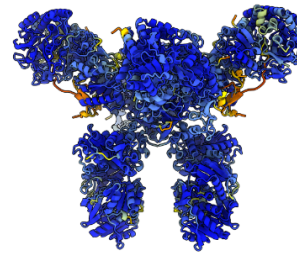
Hodarchaeales (MAG: GCA_016840025)

NuoEFGipTM: 0.852
pTM: 0.836**MvhADG**ipTM: 0.915
pTM: 0.911**EtfAB-Hdr**ipTM: 0.840
pTM: 0.867**Nuo(EG)₂-Hdr₂**ipTM: 0.872
pTM: 0.879**Nuo(EG)₂-MvhD₂**ipTM: 0.852
pTM: 0.836**Nuo(EFG)₂**ipTM: 0.738
pTM: 0.751**Nuo(EFG)₂-Mvh(ADG)₂-Etf(AB)₂-Hdr₂**

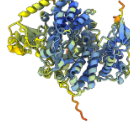
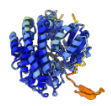
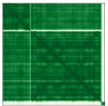
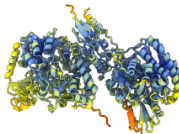
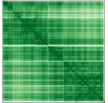
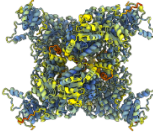
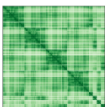
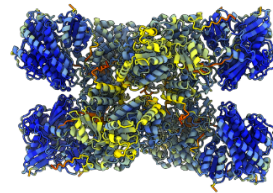
CombFold confidence score: 90.5

CombFold
→

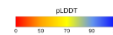
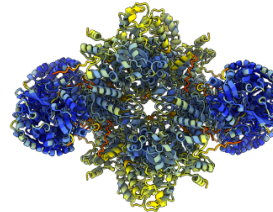
90°

**b**

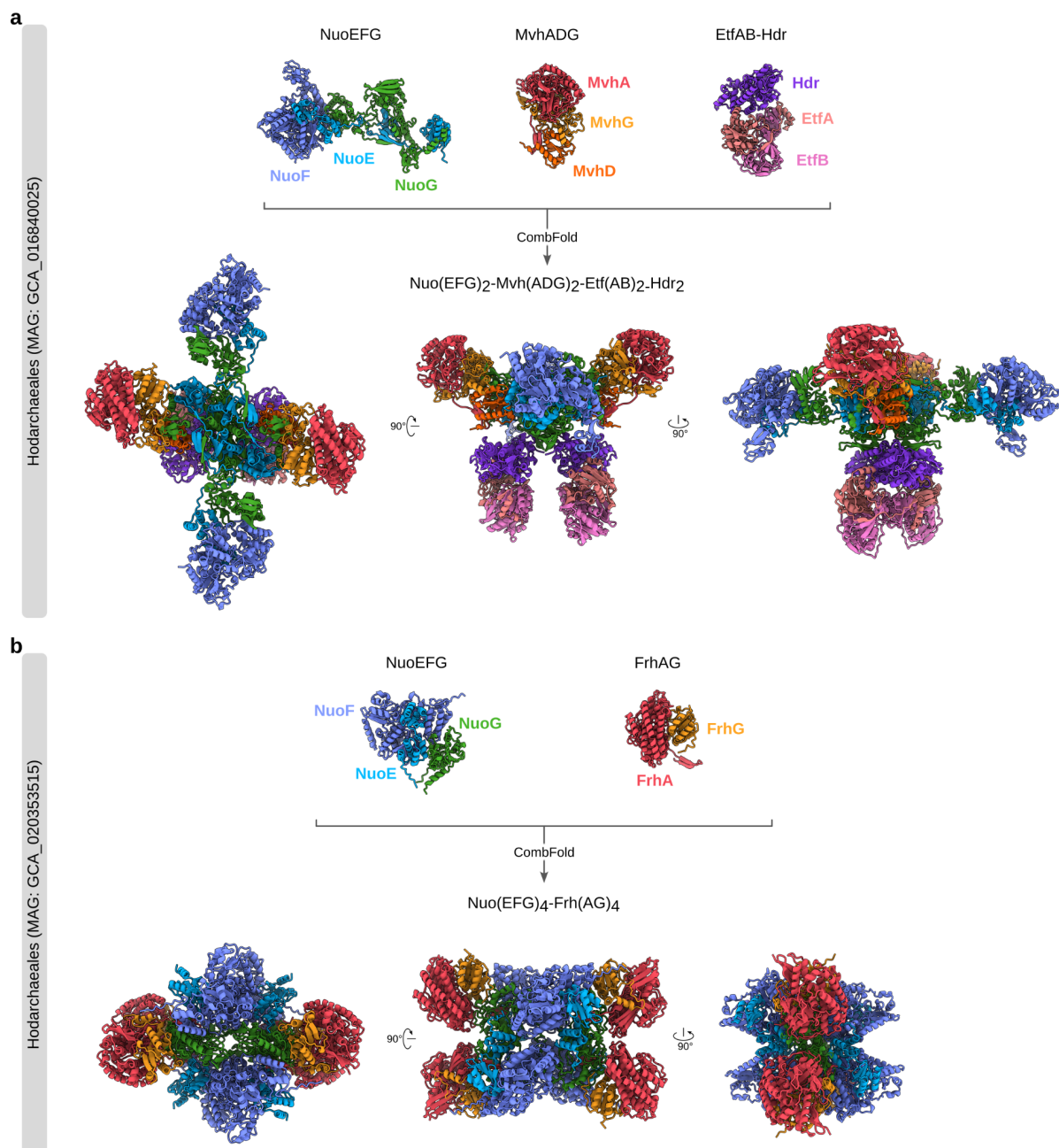
Hodarchaeales (MAG: GCA_020353515)

NuoEFGipTM: 0.85
pTM: 0.86**FrhAG**ipTM: 0.88
pTM: 0.91**NuoEFG-FrhAG**ipTM: 0.84
pTM: 0.86**Nuo(EFG)₄-FrhG₄**ipTM: 0.78
pTM: 0.80**Nuo(EFG)₄-Frh(AG)₄**
CombFold confidence score: 86.7CombFold
→

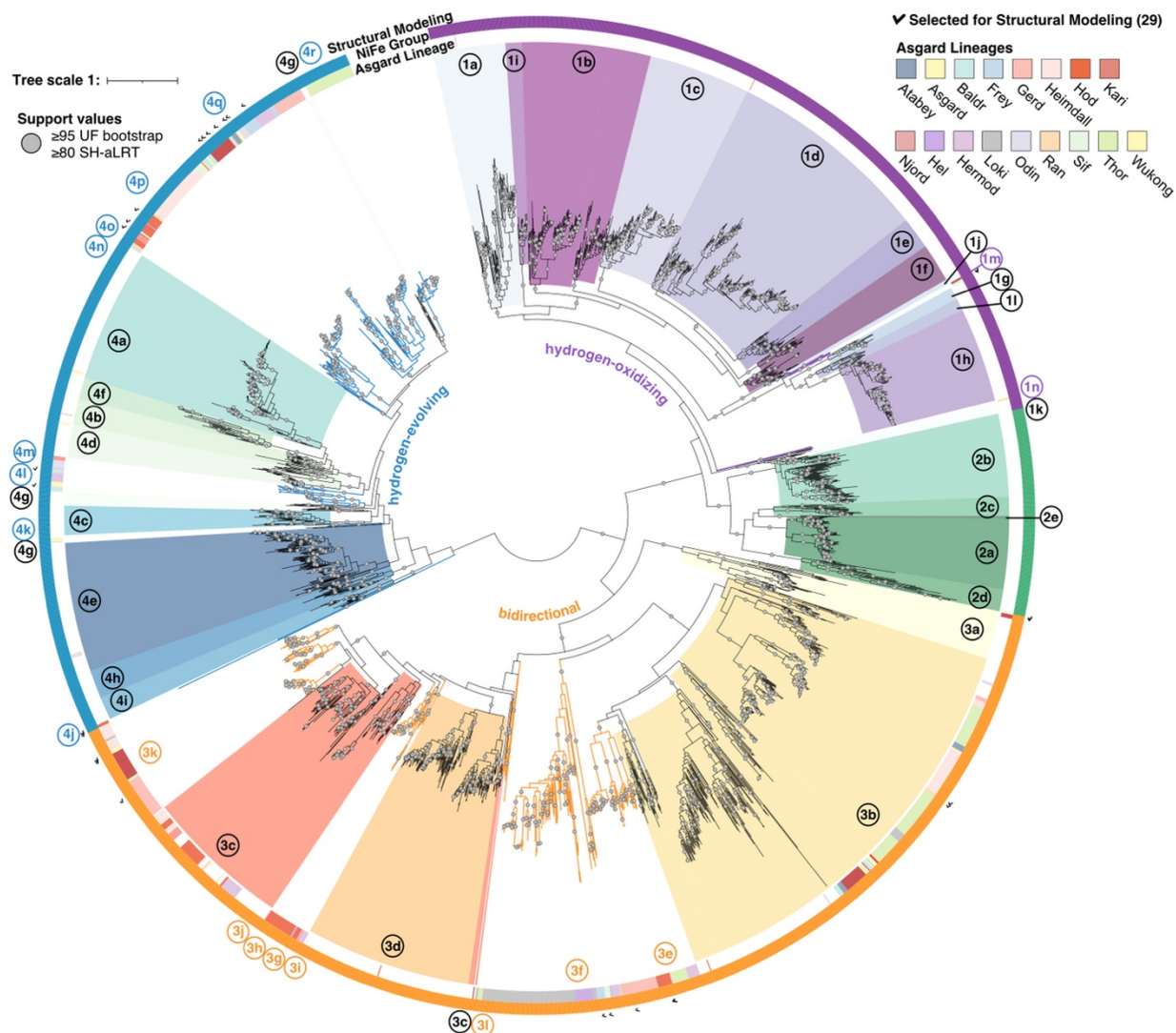
90°



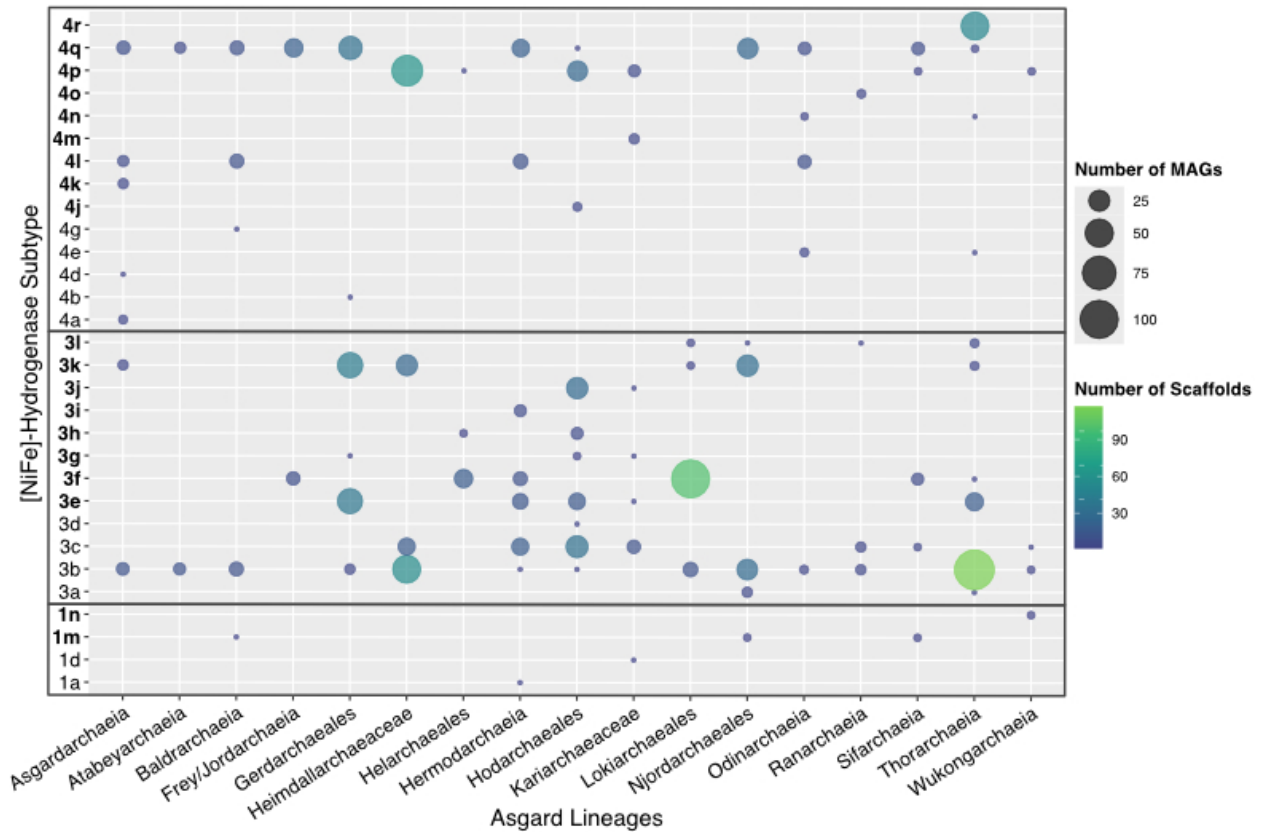
Supplementary Fig. 28 | **AlphaFold and CombFold model confidence scores of putative electron bifurcating complexes.** Due to the large size of each complex (>5000 amino acids), smaller portions of the complex were modelled with AF2 or AF3 (depicted left with PAE plots and pLDDT colored models) and assembled into a full model with CombFold (depicted right with pLDDT colored models) **a.** The putative Nuo(EFG)₂-Mvh(ADG)₂-Etf(AB)₂-Hdr₂ electron bifurcating complex found in Hodarchaeales MAG GCA_016840025. Left: smaller multimeric portions of the larger complex modeled with AF2 multimer. Right: total CombFold assembled complex shown from two different views. **b.** The putative Nuo(EFG)₄-Frh(AG)₄ electron bifurcating complex found in Hodarchaeales MAG GCA_020353515. Left: smaller multimeric portions of the larger complex modeled with AF3. Right: total CombFold assembled complex shown from two different views. Protein models are colored by predicted local distance difference test (pLDDT) scores. Subunit identities within the Predicted aligned error (PAE) plots are denoted by colored boxes. PAE plots generated by PAE Viewer webserver¹⁶¹. ipTM: interface predicted template modeling score. pTM: predicted template modeling score. Frh: F₄₂₀-reducing hydrogenase (a [NiFe]-hydrogenase). Mvh: Methyl viologen-reducing hydrogenase (a [NiFe]-hydrogenase). Etf: electron transferring flavoprotein. Hdr: heterodisulfide reductase.



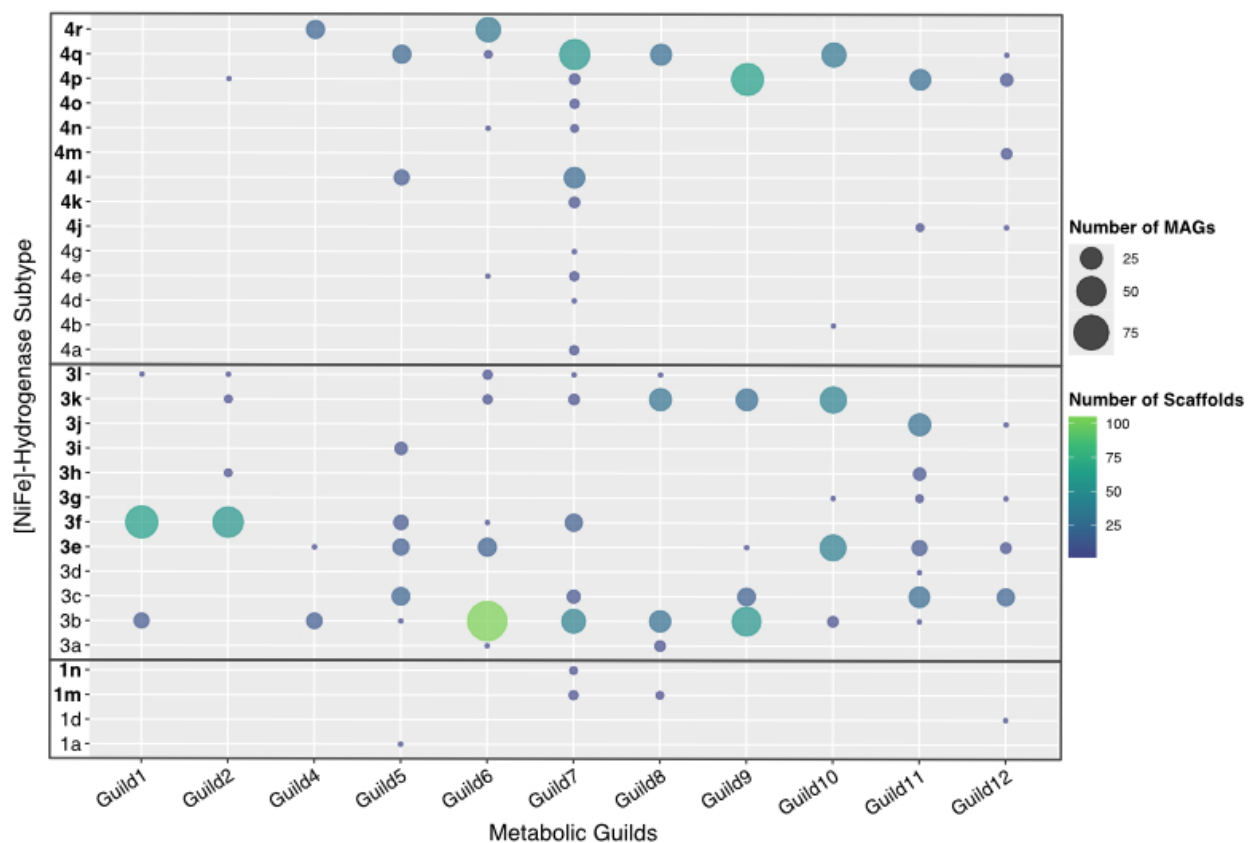
Supplementary Fig. 29 | **CombFold models of putative electron bifurcating complexes.** **a.** The putative Nuo(EFG)₂-Mvh(ADG)₂-Etf(AB)₂-Hdr₂ electron bifurcating complex found in Hodarchaeales MAG GCA_016840025. Top: individual AF2 models of the NuoEFG, MvhADG, and EtfAB-Hdr modules forming part of the total complex. Bottom: the CombFold assembled model shown from three different views. Subunits are colored according to the labelling shown in the top panel. **b.** The putative Nuo(EFG)₄-Frh(AG)₄ electron bifurcating complex found in Hodarchaeales MAG GCA_020353515. Top: individual AF2 models of the NuoEFG and FrhAG modules forming part of the total complex. Bottom: the CombFold assembled model shown from three different views. Subunits are colored according to the labelling shown in the top panel. Frh: F₄₂₀-reducing hydrogenase (a [NiFe]-hydrogenase). Mvh: Methyl viologen-reducing hydrogenase (a [NiFe]-hydrogenase). Etf: electron transferring flavoprotein. Hdr: heterodisulfide reductase.



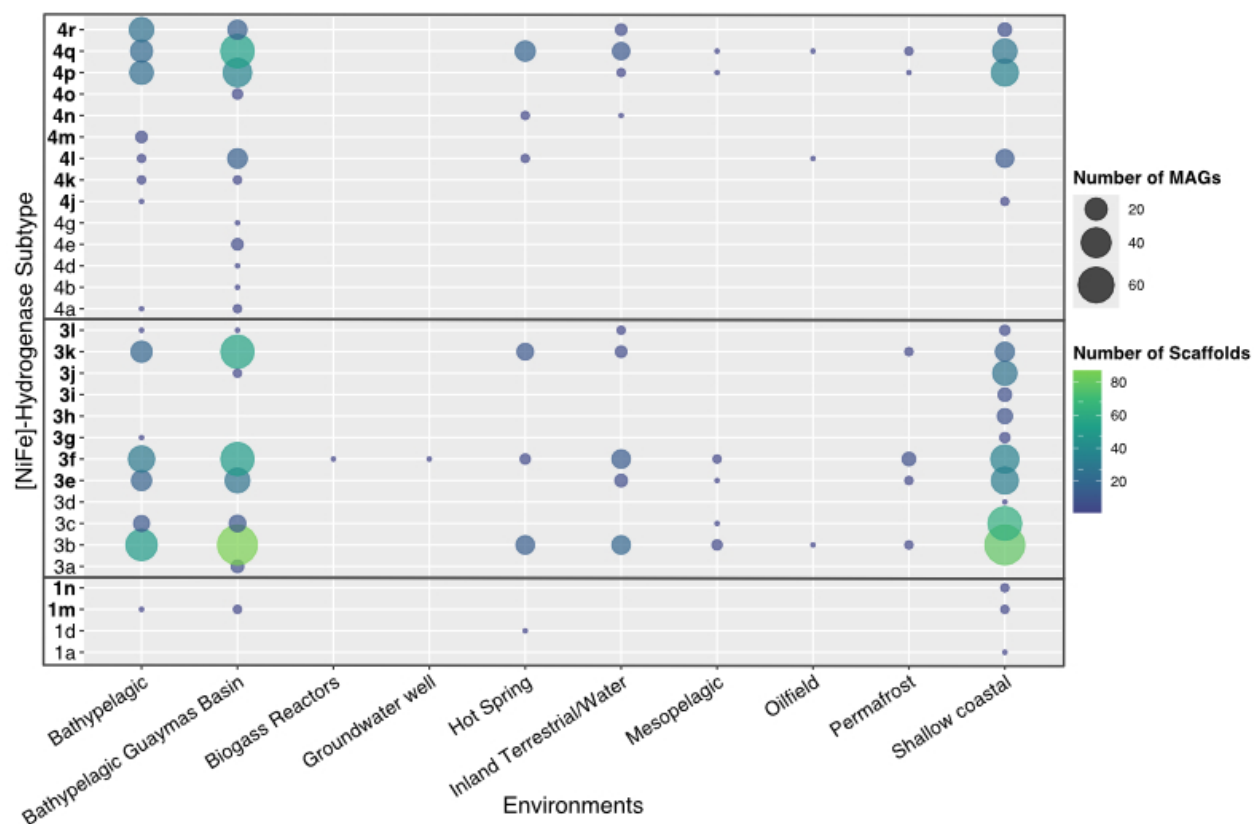
Supplementary Fig. 30 | **Expanded [NiFe]-hydrogenase diversity in Asgardarchaeota.** Maximum likelihood phylogeny was inferred from a supermatrix of 3007 sequences and 440 amino acid positions (from 2216 initial positions), displaying the diversity of Asgard archaea hydrogenases denoted by lineage in the innermost ring, spanning [NiFe]-Groups 1 (purple; respiratory and hydrogen-oxidizing), 2 (green), 3 (orange; cofactor-coupled bidirectional), and 4 (blue; energy-converting and typically hydrogen-evolving) identified by the external ring. Previously identified subgroups are labeled in black. The novel hydrogenase groups identified and named in this study are distinguished by branch and label color-coded according to the group. The black check marks highlight the 29 large catalytic subunits chosen for monomer modeling (Supplementary Figs. 41-44). The tree was generated using IQ-TREE v2.0.7 (1000 ultrafast bootstrap replicates and calculation of 1000 SH-aLRT) with Q.pfam+C50+R8 best-fit model chosen according to Bayesian Information Criterion (BIC). Support values are shown when ultrafast (UF) bootstrap support is greater than or equal to 95 and SH-aLRT support is greater than or equal to 80. The phylogeny was visualized with iTOL and rooted between the known hydrogenase clades.



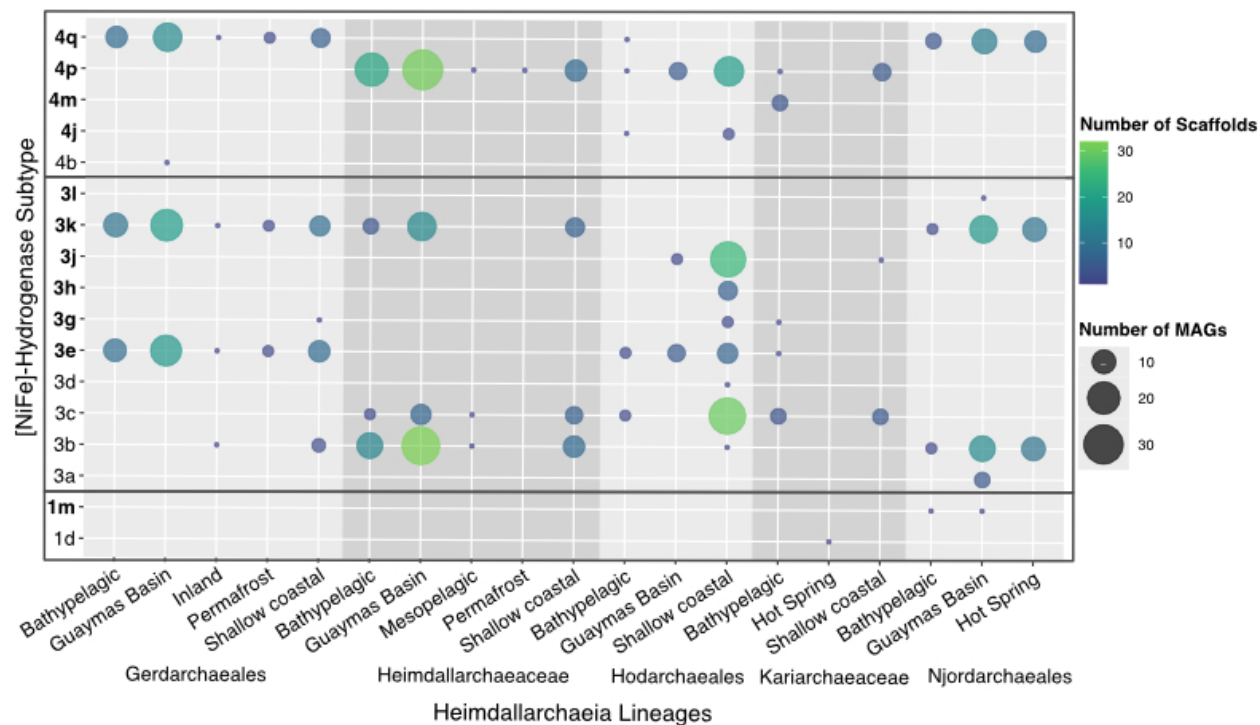
Supplementary Fig. 31 | **[NiFe]-Hydrogenase Enzyme Subtypes by Asgard archaea lineage**. The number of MAGs is shown with increasing circle size and the number of scaffolds for each lineage, increasing from blue to green. The novel lineages are shown in bold.



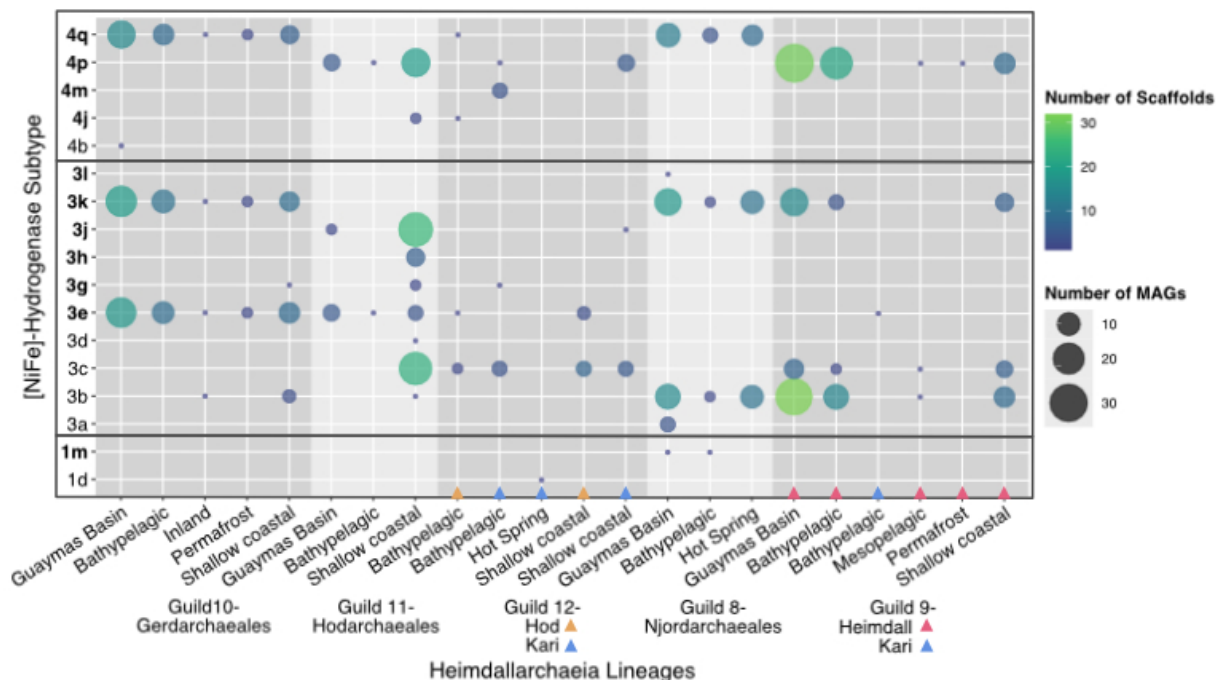
Supplementary Fig. 32 | **[NiFe]-Hydrogenase Enzyme Subtypes by Asgard archaea metabolic guild.** The number of MAGs is shown with increasing circle size and the number of scaffolds for each guild, increasing from blue to green. The novel lineages are shown in bold.



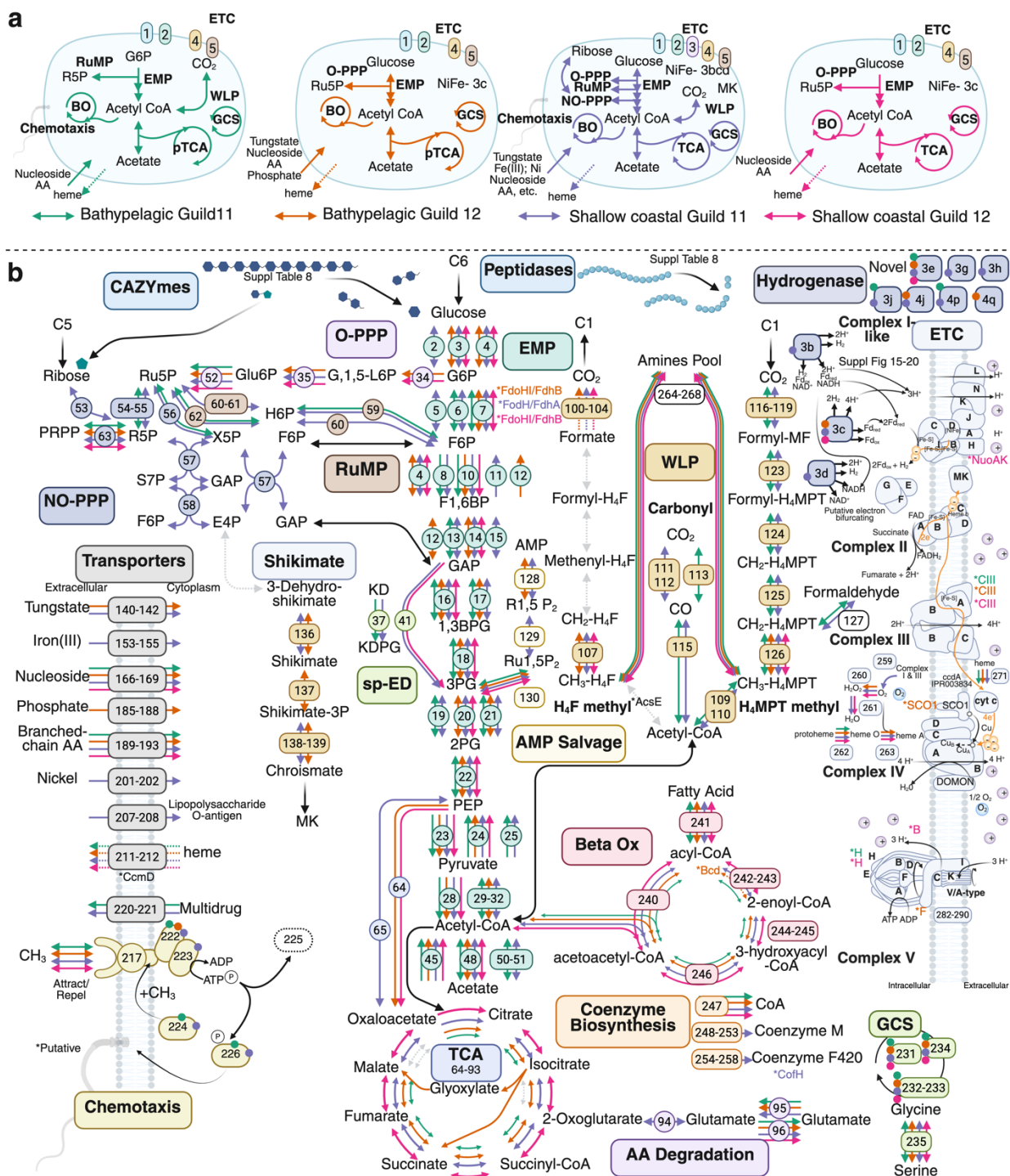
Supplementary Fig. 33 | **[NiFe]-Hydrogenase Enzyme Subtypes by Asgard archaea MAG extraction environments.** The number of MAGs is shown with increasing circle size and the number of scaffolds for each environment, increasing from blue to green. The novel lineages are shown in bold.



Supplementary Fig. 34 | [NiFe]-Hydrogenase Enzyme Subtypes for the five Heimdallarchaeia clades by environment. The number of MAGs is shown with increasing circle size and the number of scaffolds for each guild, increasing from blue to green. The background color distinguishes the clades. The novel lineages are shown in bold.

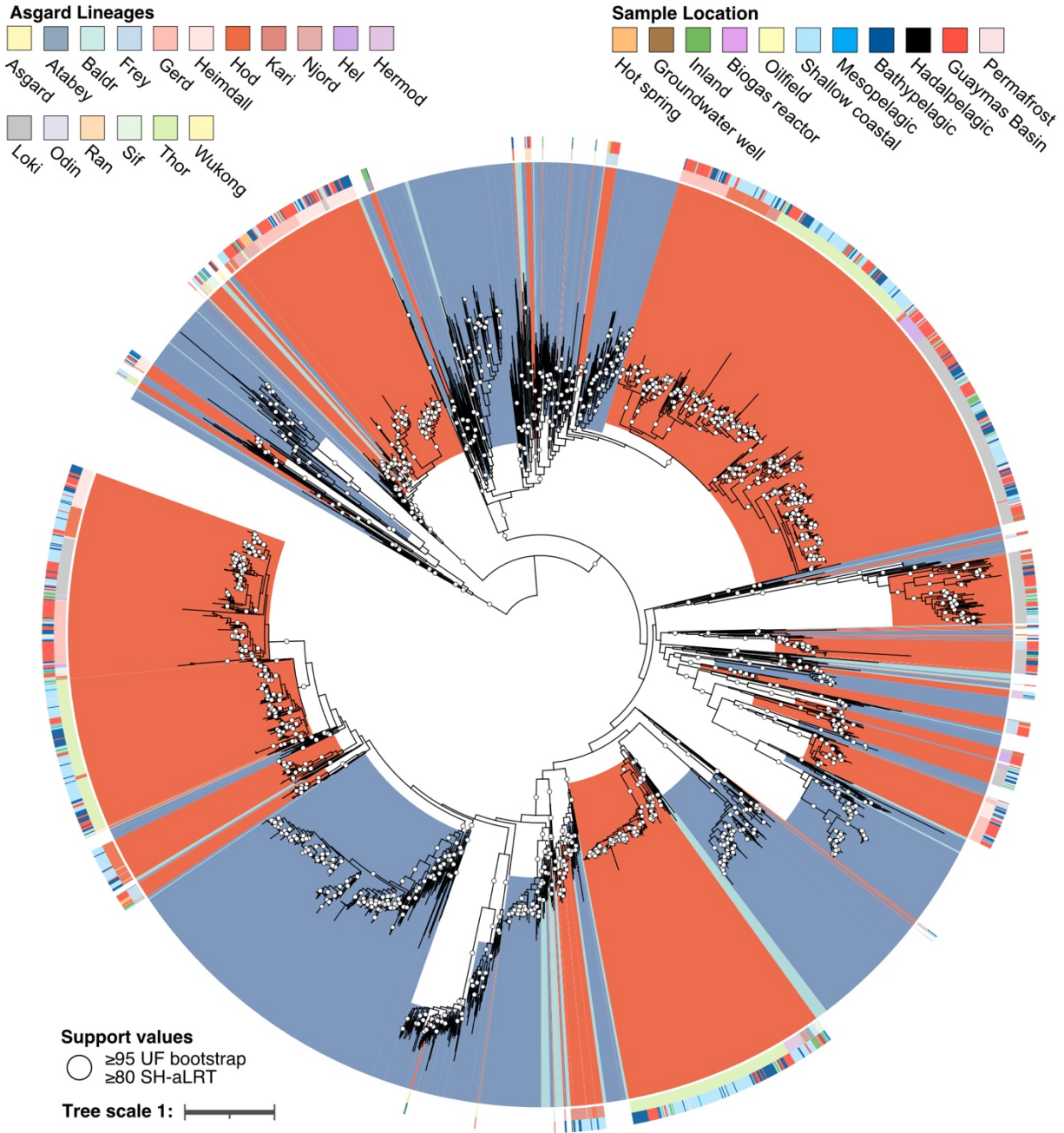


Supplementary Fig. 35 | [NiFe]-Hydrogenase Enzyme Subtypes for the five Heimdallarchaeia clades by **environment and guild**. The number of MAGs is shown with increasing circle size and the number of scaffolds for each subset, increasing from blue to green. The background color distinguishes the guild and the colored triangles the clades (Hodarchaeales-orange; Kariarchaeaceae- blue; Heimdallarchaeaceae-pink). The novel lineages are shown in bold.

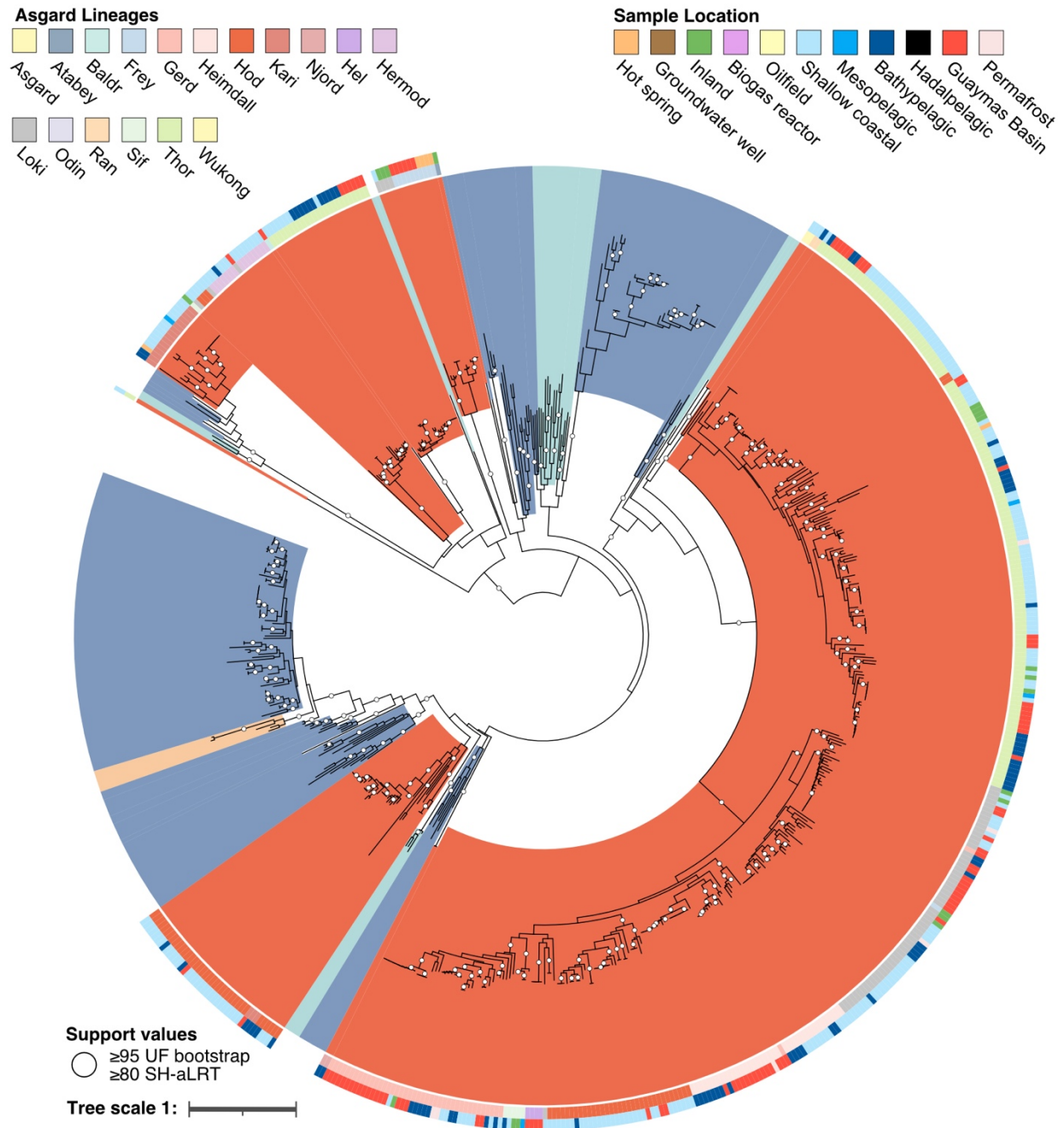


Supplementary Fig. 36 | **Metabolic reconstruction of Hodarchaeales across metabolic guilds and environments.** **a.** Overview of each subset of guild and environment with Hodarchaeales in guild 11 (green and purple) and guild 12 (orange and pink) from bathypelagic (green and orange) and coastal (purple and pink) regions of the ocean. These are distinguished with colored arrows and dots. **b.** Detailed view of the metabolic pathways. Not all predicted reactions are shown in the figure due to space, including the novel [FeFe]-hydrogenases. Missing subunits are listed in the same color as the arrows with an *. All numbers listed correspond to the numbers in Supplementary Table 8, which provides metabolic overviews for all

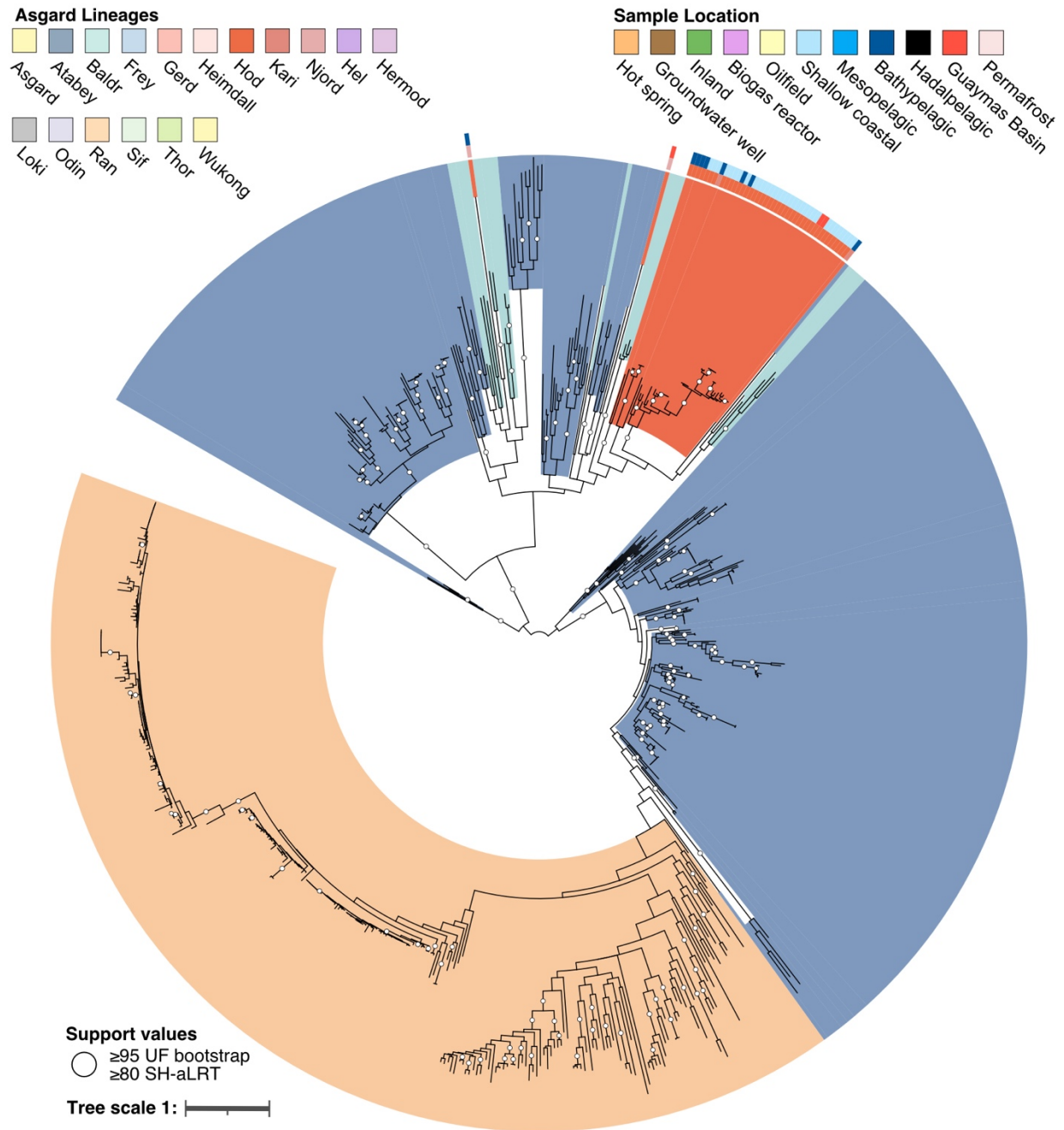
Asgardarchaeota lineages. Abbreviations included RuMP (Ribulose Monophosphate Pathway), BO (Beta Oxidation), AA (Amino Acids), TCA /pTCA (Tricarboxylic Acid Cycle and partial version), WLP (Wood-Ljungdahl Pathway), GCS (Glycine Cleavage System), ETC (Electron Transport Chain), NO/O-PPP (Non-Oxidative/Oxidative Pentose Phosphate Pathway), EMP (Embden-Meyerhof-Parnas Pathway), MK (menaquinone), sp-ED (semi-phosphorylative Entner–Doudoroff), C# (compound with # carbons), and CAZymes (Carbohydrate-Active enZymes). Created in BioRender. Appler, K. (2025) <https://BioRender.com/ldf2y58>.



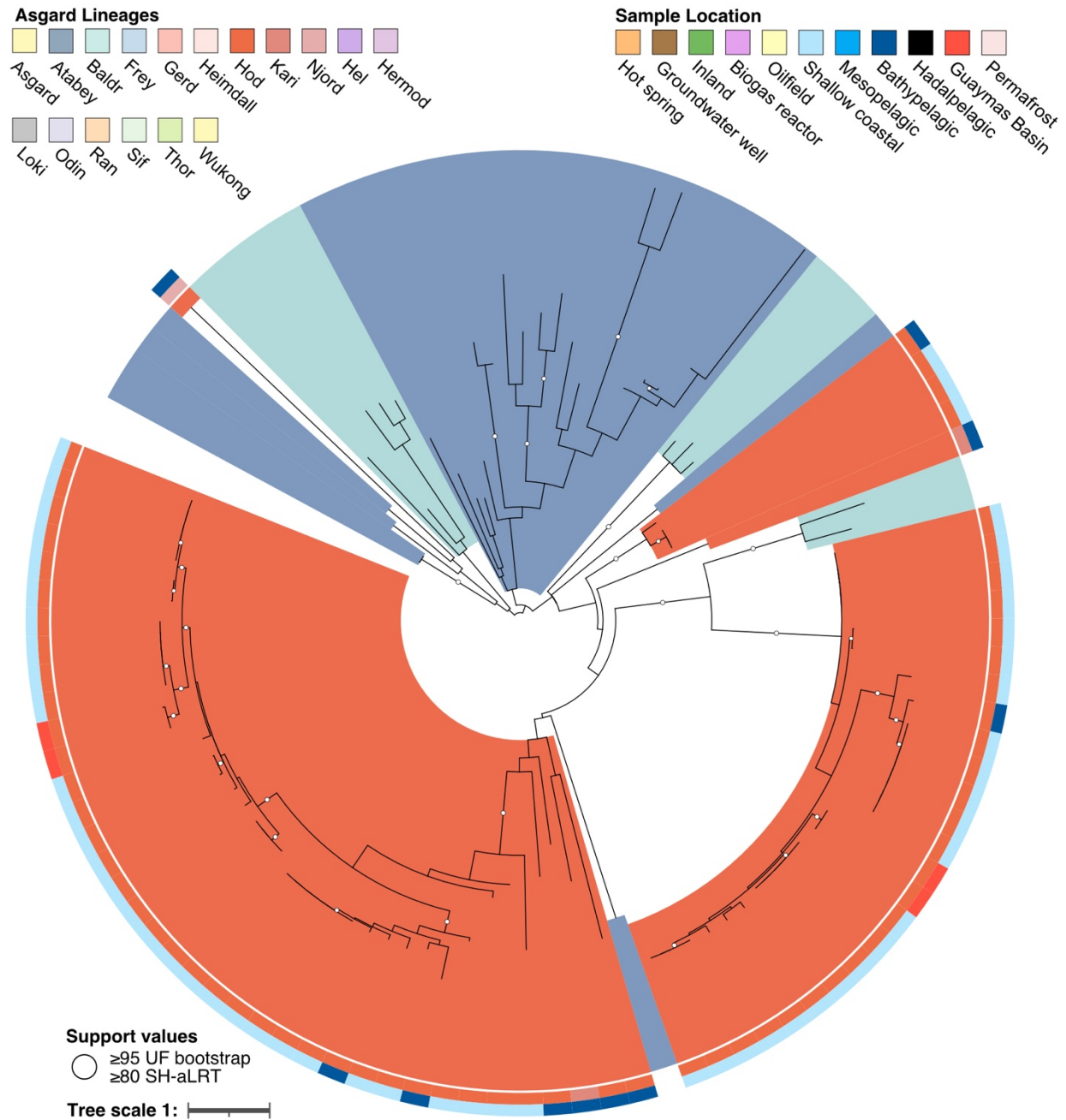
Supplementary Fig. 37 | **Succinate Dehydrogenase Subunit A (SdhA) phylogeny.** Maximum likelihood phylogeny was inferred from a supermatrix of 3437 sequences and 537 amino acid positions (from 5403 initial positions), displaying the diversity of SdhA subunits across Asgard archaea (red), other archaeal lineages (teal), and bacteria (dark blue). The tree was generated using IQ-TREE v2.0.7 (1000 ultrafast bootstrap replicates and calculation of 1000 SH-aLRT) with WAG+C60+R8 best-fit model chosen according to Bayesian Information Criterion (BIC). The two external rings show the Asgard lineage and source location, respectively. Support values are shown when ultrafast (UF) bootstrap support is greater than or equal to 95 and SH-aLRT support is greater than or equal to 80. The phylogeny was visualized with iTOL and midpoint rooted.



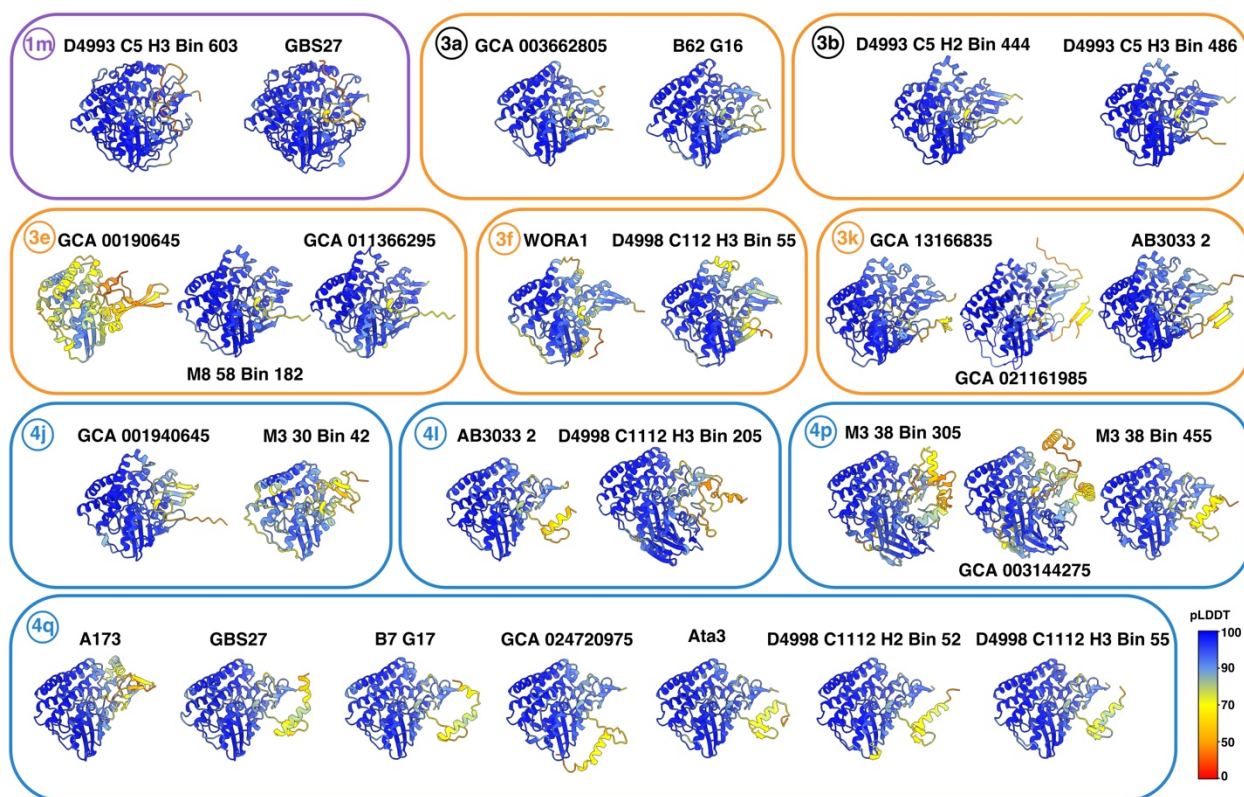
Supplementary Fig. 38 | **Succinate Dehydrogenase Subunit B (SdhB) phylogeny.** Maximum-likelihood phylogeny was inferred from a supermatrix of 661 sequences and 243 amino acid positions (from 1325 initial positions), displaying the diversity of SdhB subunits across Asgard archaea (red), other archaeal lineages (teal), bacteria (dark blue), and mitochondria (orange). The tree was generated using IQ-TREE v2.0.7 (1000 ultrafast bootstrap replicates and calculation of 1000 SH-aLRT) with WAG+C50+R8 best-fit model chosen according to Bayesian Information Criterion (BIC). The two external rings show the Asgard lineage and source location, respectively. Support values are shown when ultrafast (UF) bootstrap support is greater than or equal to 95 and SH-aLRT support is greater than or equal to 80. The phylogeny was visualized with iTOL and midpoint rooted.



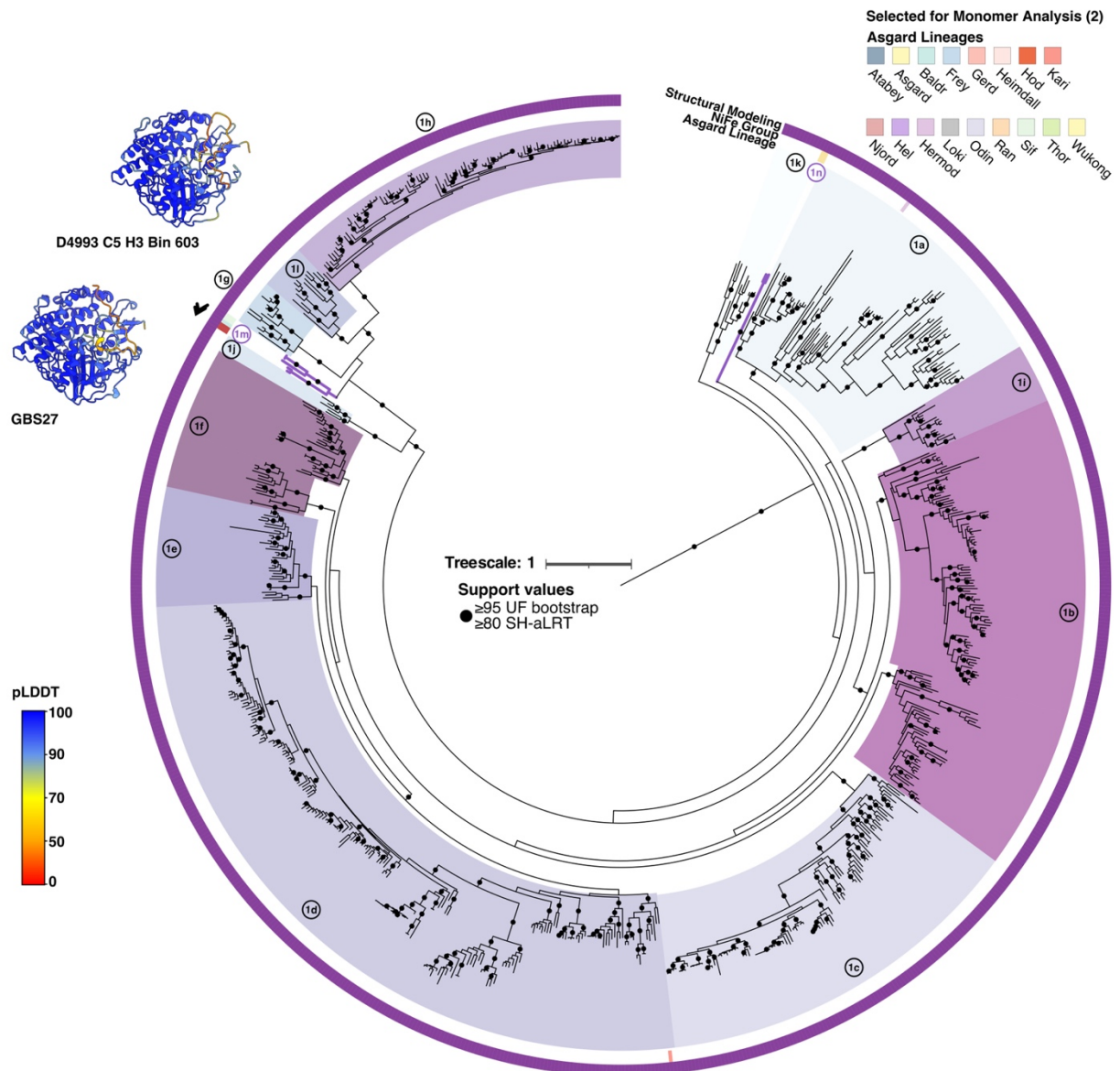
Supplementary Fig. 39 | **Succinate Dehydrogenase Subunit C (SdhC) phylogeny.** Maximum-likelihood phylogeny was inferred from a supermatrix of 705 sequences and 128 amino acid positions (from 824 initial positions), displaying the diversity of SdhC subunits across Asgard archaea (red), other archaeal lineages (teal), bacteria (dark blue), and mitochondria (orange). The tree was generated using IQ-TREE v2.0.7 (1000 ultrafast bootstrap replicates and calculation of 1000 SH-aLRT) with WAG+C50+R8 best-fit model chosen according to Bayesian Information Criterion (BIC). The two external rings show the Asgard lineage and source location, respectively. Support values are shown when ultrafast (UF) bootstrap support is greater than or equal to 95 and SH-aLRT support is greater than or equal to 80. The phylogeny was visualized with iTOL and midpoint rooted.



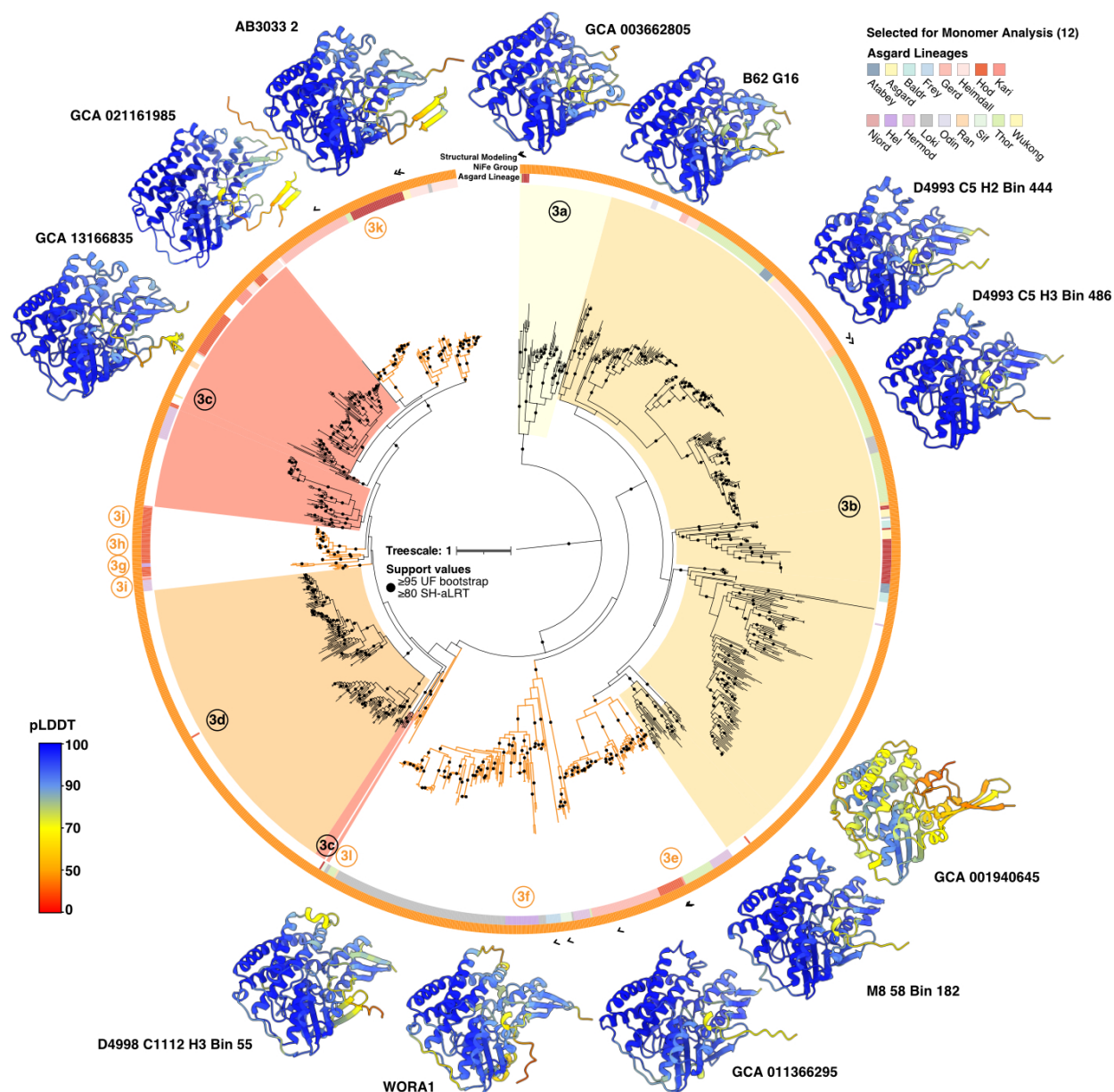
Supplementary Fig. 40 | **Succinate Dehydrogenase Subunit D (SdhD) phylogeny.** Maximum-likelihood phylogeny was inferred from a supermatrix of 105 sequences and 135 amino acid positions (from 481 initial positions), displaying the diversity of SdhD subunits across Asgard archaea (red), other archaeal lineages (teal), bacteria (dark blue), and mitochondria (orange). The tree was generated using IQ-TREE v2.0.7 (1000 ultrafast bootstrap replicates and calculation of 1000 SH-aLRT) with WAG+C50+G4 best-fit model chosen according to Bayesian Information Criterion (BIC). The two external rings show the Asgard lineage and source location, respectively. Support values are shown when ultrafast (UF) bootstrap support is greater than or equal to 95 and SH-aLRT support is greater than or equal to 80. The phylogeny was visualized with iTOL and midpoint rooted.

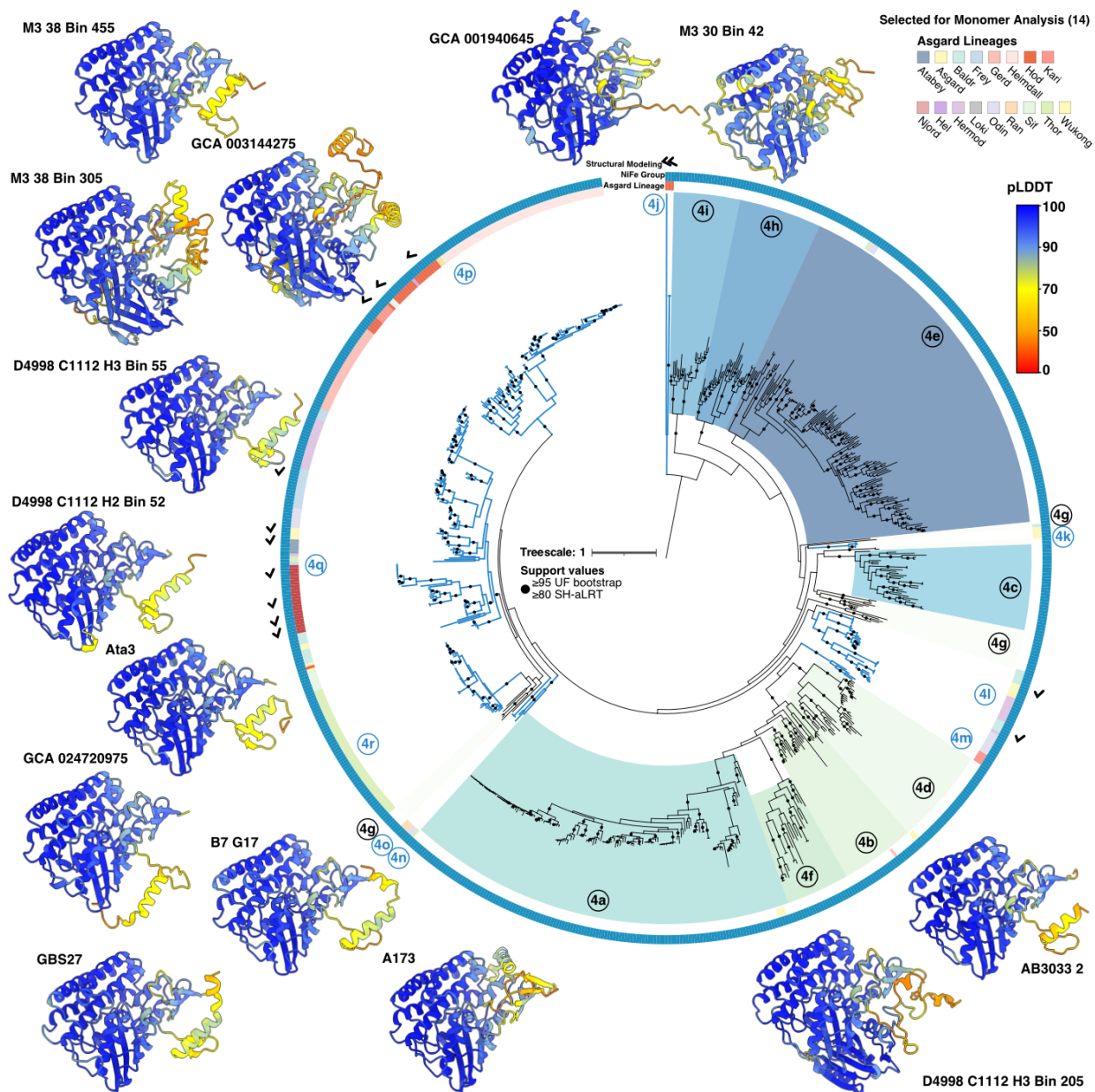


Supplementary Fig. 41 | **AlphaFold2 monomer models of [NiFe]-hydrogenase large subunits.** Proteins are displayed by identified hydrogenase groups and subgroups with genome names. Individual monomers colored by the predicted local distance difference test (pLDDT) scores. Only top ranked AF2 models are shown.

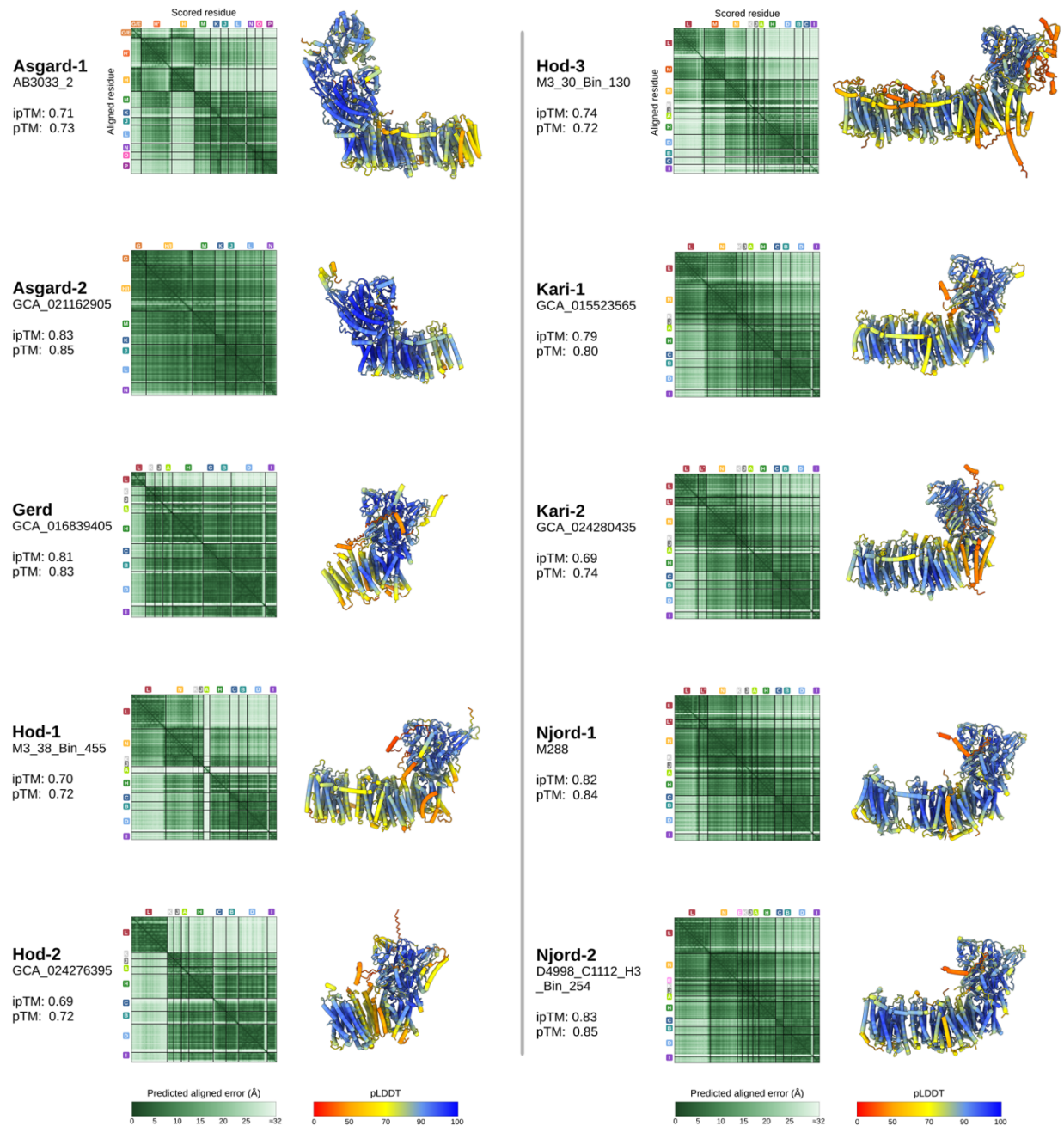


Supplementary Fig. 42 | **AlphaFold2 monomer models of [NiFe]-hydrogenase large subunits Group 1.** Maximum likelihood phylogeny was inferred from a supermatrix of 3007 sequences and 440 amino acid positions (from 2216 initial positions), displaying the diversity of Asgard archaea hydrogenases collapsed to highlight [NiFe]-Groups 1 (purple) identified by the external ring. The tree was generated using IQ-TREE v2.0.7 (1000 ultrafast bootstrap replicates and calculation of 1000 SH-aLRT) with Q.pfam+C50+R8 best-fit model chosen according to Bayesian Information Criterion (BIC). Previously identified subgroups are labeled in black. The novel hydrogenase groups identified and named in this study are distinguished by branch and label color-coded according to the group. The black check marks highlight the 2 large catalytic subunits chosen for monomer modeling for subgroup 1m. Support values are shown when ultrafast (UF) bootstrap support is greater than or equal to 95 and SH-aLRT support is greater than or equal to 80. Individual monomers colored by the predicted local distance difference test (pLDDT) scores. Only top ranked AF2 models are shown. The phylogeny was visualized with iTOL and midpoint rooted between the known hydrogenase clades and pruned to show each group, separately.



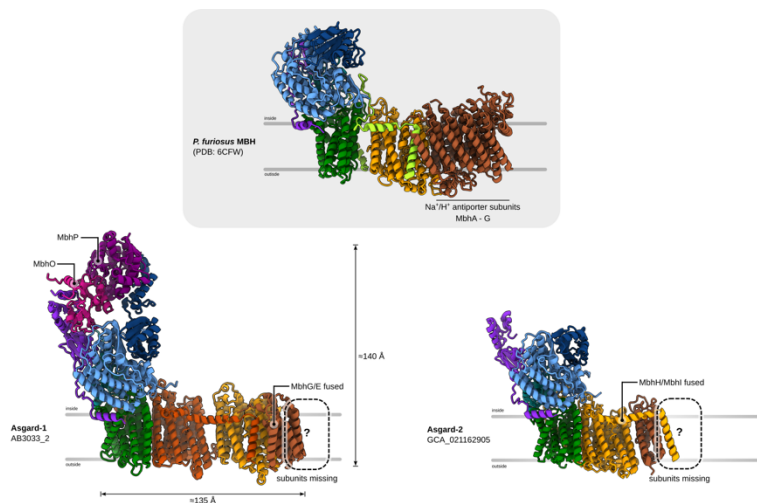


Supplementary Fig. 44 | **AlphaFold2 monomer models of [NiFe]-hydrogenase large subunits Group 4.** Maximum likelihood phylogeny was inferred from a supermatrix of 3007 sequences and 440 amino acid positions (from 2216 initial positions), displaying the diversity of Asgard archaea hydrogenases collapsed to highlight [NiFe]-Groups 4 (blue) identified by the external ring. The tree was generated using IQ-TREE v2.0.7 (1000 ultrafast bootstrap replicates and calculation of 1000 SH-aLRT) with Q.pfam+C50+R8 best-fit model chosen according to Bayesian Information Criterion (BIC). Previously identified subgroups are labeled in black. The novel hydrogenase groups identified and named in this study are distinguished by branch and label color-coded according to the group. The black check marks highlight the 14 large catalytic subunits chosen for monomer modeling for subgroups 4j, 4l, 4p, and 4q. Support values are shown when ultrafast (UF) bootstrap support is greater than or equal to 95 and SH-aLRT support is greater than or equal to 80. Individual monomers colored by the predicted local distance difference test (pLDDT) scores. Only top ranked AF2 models are shown. The phylogeny was visualized with iTOL and midpoint rooted between the known hydrogenase clades and pruned to show each group, separately.

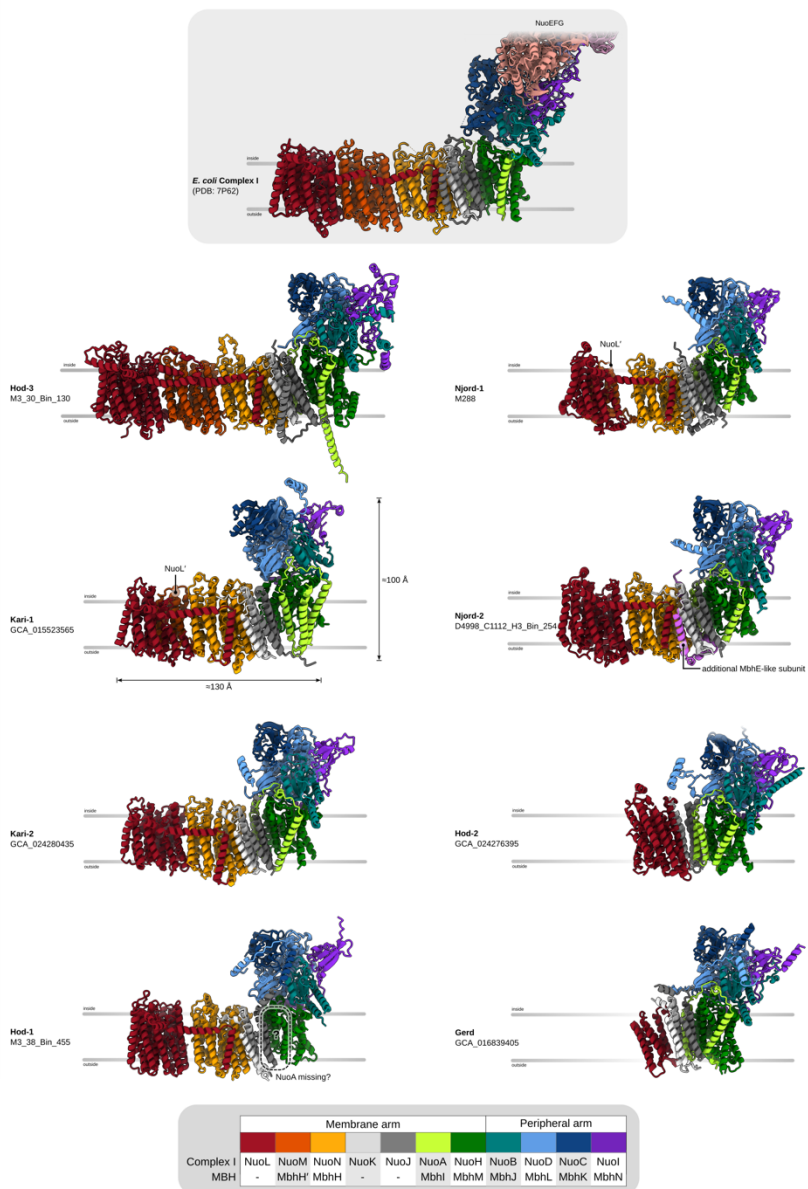


Supplementary Fig. 45 | **AlphaFold2 multimer confidence scores of Complex I-like and MBH-like models.** Top-ranked multimers of *Heimdallarchaea* and *Asgardarchaea* group 4 [NiFe]-hydrogenase membrane-bound complexes. Protein models are colored by predicted local distance difference test (pLDDT) scores. Subunit identities within the Predicted aligned error (PAE) plots are denoted by colored boxes and follow the *nuo* naming convention for *Heimdallarchaea* and the *mbh* naming convention for *Asgardarchaea* (see Supplementary Fig. 46). PAE plots generated by PAE Viewer webserver¹⁶¹. ipTM: interface predicted template modeling score. pTM: predicted template modeling score. Acronyms for the complexes, removing the taxonomic suffix and adding numbering Kari-1 (GCA_015523565), Kari-2 (GCA_024280435), Njord-1 (M288), Njord-2 (D4998_C1112_H3_Bin_254), Hod-1 (M3_38_Bin_455), Hod-2 (GCA_024276395), Hod-3 (M3_30_Bin_130), Gerd (GCA_016839405), Asgard-1 (AB3033_2), and Asgard-2 (GCA_021162905).

MBH-like architecture



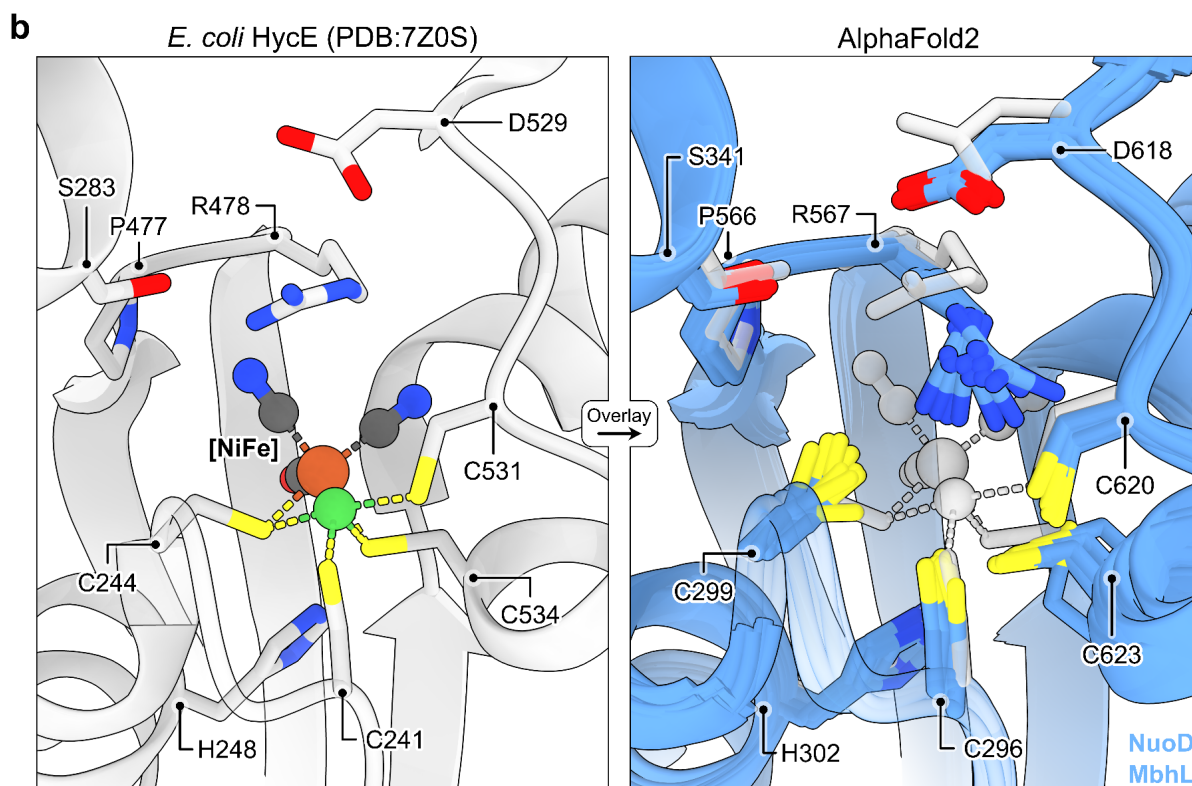
Complex I-like architecture



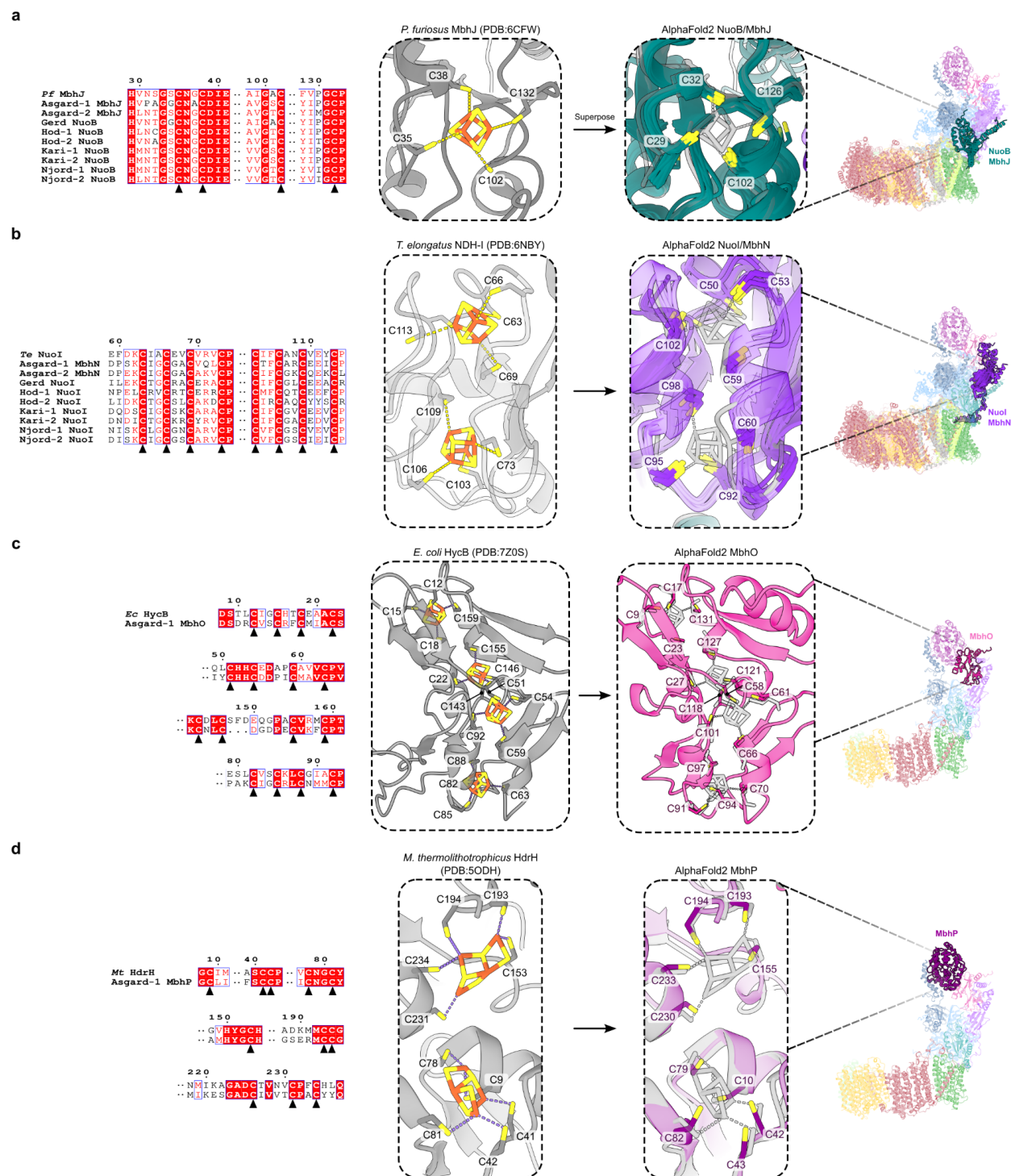
Supplementary Fig. 46 | **Overall structural architecture and shared subunit homology in Complex I-like and MBH-like complexes.** AF2 multimer models are divided into structures that share a Complex I-like or an MBH-like structural architecture. For MBH-like complexes, the *P. furiosus* membrane-bound hydrogenase, MBH (PDB ID: 6CFW)²⁶, is shown for comparison. The order and orientation of the membrane-arm subunits relative to the peripheral-arm subunits is characteristic of MBH and distinguishes those structures from Complex I-like complexes^{26,88}. For Complex I-like complexes, the *E. coli* Complex I (PDB ID: 7P62)⁹⁴ is shown for comparison. The order and orientation of membrane-arm subunits relative to the peripheral-arm subunits in these AF2 models is characteristic of Complex I. In other words, when all homologous antiporter subunits are viewed in the same register, the peripheral-arm subunits in MBH-like complexes will appear rotated at ~180° relative to Complex I-like peripheral-arm subunits. We have highlighted areas on our AF2 models where expected subunits are absent. Missing subunits may be a result of our annotation step failing to capture all relevant sequences for AF2 input. A table showing subunit names and their corresponding coloring in protein models is shown as a color key. Homologous subunits share the same coloring. The colors of different subunits are consistent across this manuscript unless stated otherwise. Acronyms for the complexes, removing the taxonomic suffix and adding numbering Kari-1 (GCA_015523565), Kari-2 (GCA_024280435), Njord-1 (M288), Njord-2 (D4998_C1112_H3_Bin_254), Hod-1 (M3_38_Bin_455), Hod-2 (GCA_024276395), Hod-3 (M3_30_Bin_130), Gerd (GCA_016839405), Asgard-1 (AB3033_2), and Asgard-2 (GCA_021162905).

a

		Metal coordination				Canopy residues				Metal coordination		
	<i>E. coli</i> NuoCD	IEYLGGCVNE			...	NSH	...	ATKG	...	IDFVMSDV		
	<i>E. coli</i> HycE	V CGI CGFAHS	...	H SH	...	A PRG	...	L DP CYSCT				
AlphaFold2	Hod-1 NuoD	L CGI CSVAHQ	...	Q SH	...	A PRG	...	I DP CIGCC				
	Kari-1 NuoD	I CGI CSFAHN	...	H SH	...	A PRG	...	N DP CFSCT				
	Kari-2 NuoD	I CGI CSFSHQ	...	H SH	...	A PRG	...	C DP CFSCT				
	Gerd NuoD	I CGI CSHIHS	...	H SH	...	A PRG	...	L DP CFSCE				
	Njord-1 NuoD	V CGI CSSSHQ	...	H SH	...	A PRG	...	I DP CMSCM				
	Njord-2 NuoD	V CGI CSSAHQ	...	H SH	...	A PRG	...	I DP CMSCM				
	Hod-2 NuoD	T CGI CSGVHQ	...	H SH	...	A PRG	...	I DP CMSCM				
	Asgard-2 MbhL	V CAI CTQAHS	...	E SH	...	A PRG	...	L DP CIACA				
Asgard-1 MbhL	V CGI CSGVHT	...	H SH	...	A PRG	...	V DP CFSCT					
		295	300	340		565		620				

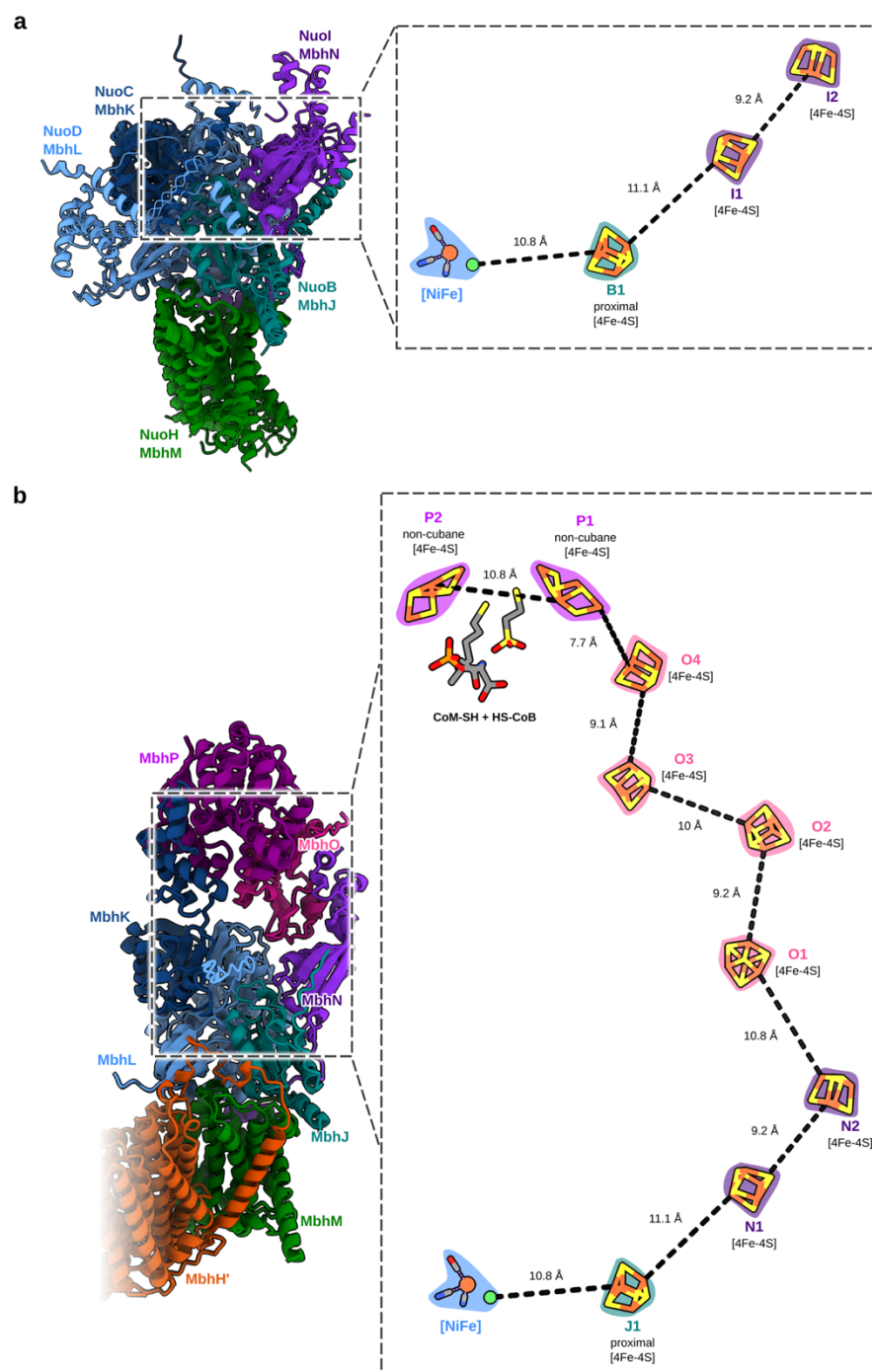


Supplementary Fig. 47 | **Structure of putative [NiFe]-hydrogenase active site in group 4 [NiFe]-hydrogenase membrane-bound complexes.** **a.** Multiple sequence alignment of [NiFe]-hydrogenase large subunits. Residues that are necessary for coordinating the [NiFe] cofactor are highlighted and set in bold font. *E. coli* NuoCD, which lacks hydrogenase activity, also lacks [NiFe] coordinating residues. Conversely, *E. coli* HycE is hydrogenase active and conserved [NiFe] coordinating residues²⁴. **b.** The [NiFe] site in *E. coli* HycE (UnitProt ID: P16431; PDB ID: 7Z0S)²⁴ compared to AF2 models of the same site. Superposition of the AF2 models (blue) with *EcHycE* (gray) illustrates that the putative active sites in the AF2 models share structural similarity with the active hydrogenase *EcHycE*. Numbering of residues in AF2 are based on alignment with *E. coli* NuoCD (UniProt ID: P33599). Multiple sequence alignment made with Clustal Omega¹⁶².

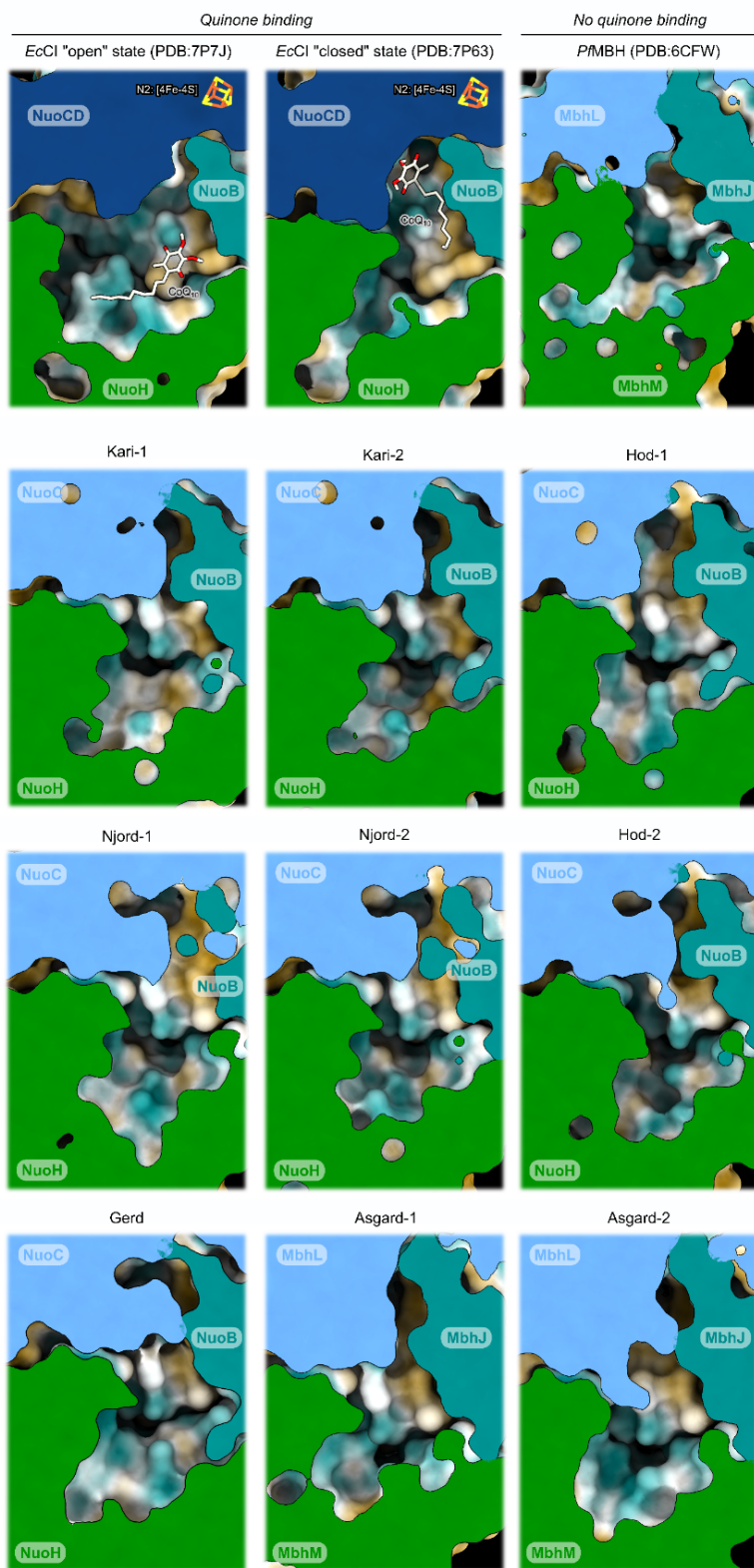
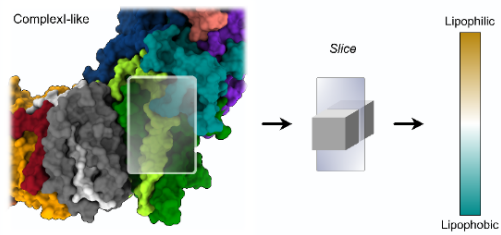


Supplementary Fig. 48 | **Structure of putative iron-sulfur clusters sites in the peripheral arm subunits of group 4 [NiFe]-hydrogenase membrane-bound complexes.** First column: multiple sequence alignment of subunits, with conserved metal-coordinating cysteines labeled with black arrows. The residue numbering (on top) is conserved across the alignments. Second column: close-up view of iron-sulfur binding sites in PDB models selected for comparative structural modeling. Third column: superposition of PDB models (grey) with AF2 models of group 4 [NiFe]-hydrogenase complex subunits. Fourth column: location of subunit relative to the larger complex. **a.** *P. furiosus* MbhJ (UniProt ID: Q8U0Z8; PDB ID: 6CFW)²⁶

structurally aligned with the [NiFe]-hydrogenase small subunit AF2 models, NuoB (7 ×) and MbhJ (2 ×). **b.** *T. elongatus* NDH-I (UniProt ID: Q8DL31; PDB ID: 6NBY)²⁸ structurally aligned with the ferredoxin subunit AF2 models, NuoI (7 ×) and MbhN (2 ×). **c.** *E. coli* HycB (UniProt ID: P0AAK1; PDB ID: 7Z0S)²⁴ structurally aligned with the AF2 model of the large ferredoxin subunit (MbhO) from Asgardarchaeia MBH “Asgard-1”. **d.** The *M. thermolithotrophicus* HdrH (UniProt ID: A0A2D0TCB9; PDB ID: 5ODH)²⁹ structurally aligned with the putative heterodisulfide reductase (MbhP) from “Asgard-1”. Sequence alignments were made with Clustal Omega¹⁶² and visualized with ESPript 3¹⁶³.



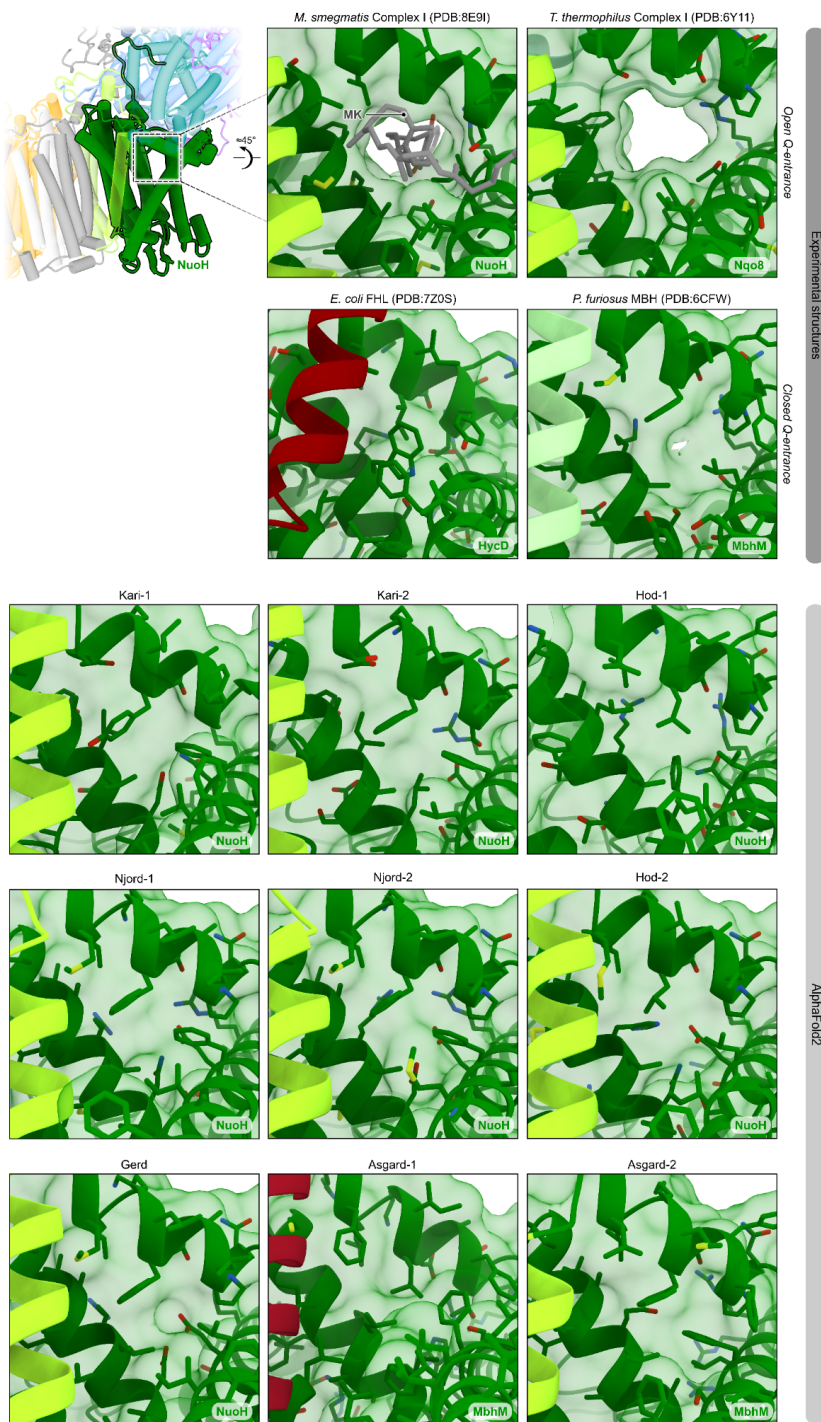
Supplementary Fig. 49 | **Putative electron transfer paths in group 4 [NiFe]-hydrogenase membrane-bound complexes.** **a.** The putative electron transfer path for the 7 X Complex I-like and the MBH-like complex, but not the Asgardarchaea MBH-like complex, “Asgard-1”. Approximate placement of putative FeS clusters and the [NiFe] cofactor is based on structural alignment as described previously in Supplementary Figs. 30-31. Shortest distances between each FeS cluster and the [NiFe] cofactor are marked with black dotted lines. Edge-to-edge distances are labeled in ångström (Å). Membrane-arm subunits not shown in the AF2 multimer superposition. **b.** The putative electron transfer path for the Asgardarchaea MBH-like complex. As before, approximate placement of putative FeS clusters and the [NiFe] cofactor is based on structural alignment as described previously in Supplementary Figs. 47-48.



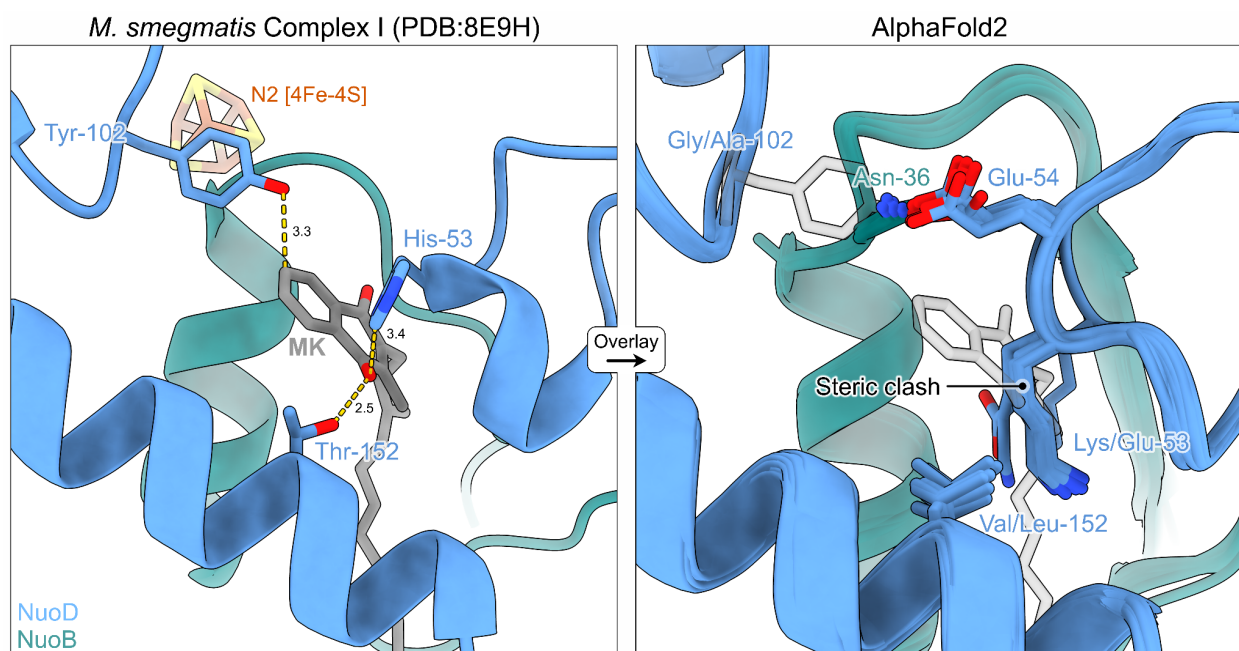
Experimental structures

AlphaFold2

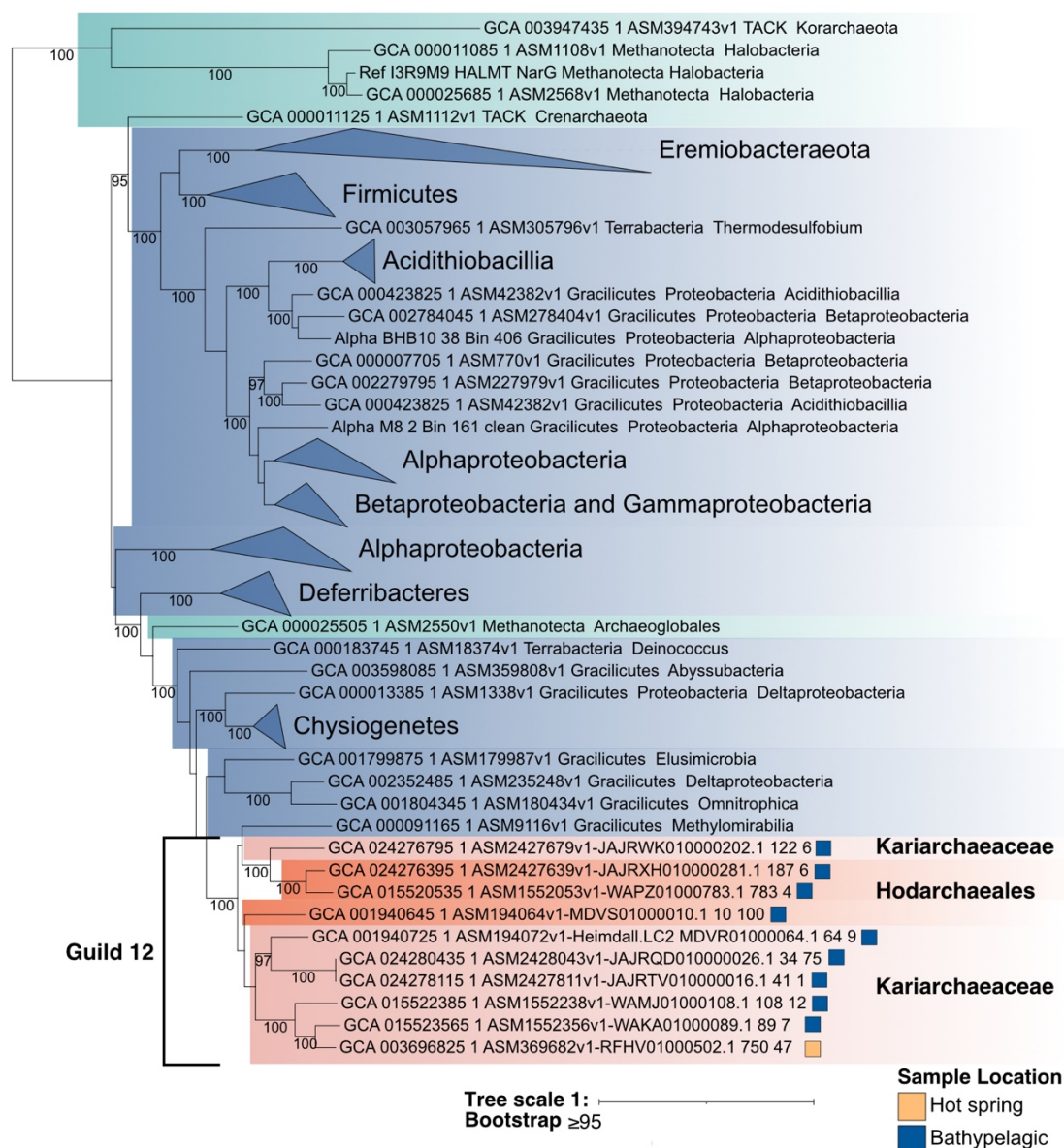
Supplementary Fig. 50 | **Peripheral arm cavity in group 4 [NiFe]-hydrogenase membrane-bound complexes.** Sliced view of the peripheral arm reveals potential quinone binding cavities. For comparison, the experimentally determined quinone binding cavity of *E. coli* Complex I in two different turnover states is shown (PDB IDs: 7P7J and 7P63)⁹⁴, as well as the non-quinone binding cavities from *E. coli* FHL (PDB ID: 7Z0S)²⁴ and *Pyrococcus furiosus* MBH (PDB ID: 6CFW)²⁶. Cavity lipophilicity was calculated in UCSF ChimeraX v1.8²³. CoQ10: Coenzyme Q10. Acronyms for the complexes, removing the taxonomic suffix and adding numbering Kari-1 (GCA_015523565), Kari-2 (GCA_024280435), Njord-1 (M288), Njord-2 (D4998_C1112_H3_Bin_254), Hod-1 (M3_38_Bin_455), Hod-2 (GCA_024276395), Gerd (GCA_016839405), Asgard-1 (AB3033_2), and Asgard-2 (GCA_021162905).



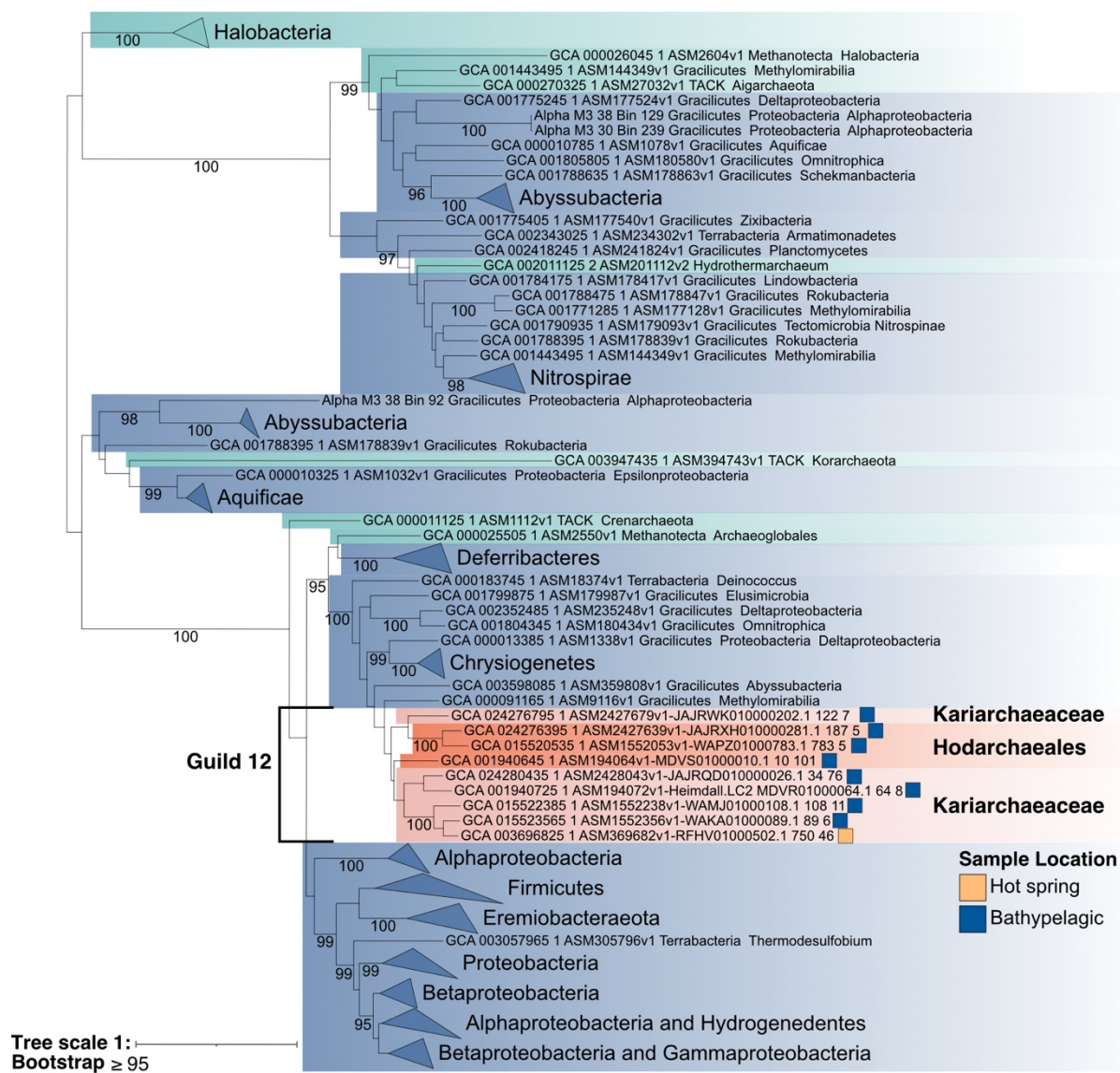
Supplementary Fig. 51 | **Lack of quinone tunnel in group 4 [NiFe]-hydrogenase membrane-bound complexes.** AF2 models compared to experimental structures of quinone binding Complex I (PDB ID's: 8E9I, 6Y11)³¹ and non-quinone binding Complex I-like proteins (PDB ID's: 7Z0S, 6CFW)²⁶. A tunnel opening is seen between three helices of Nuoh in classic Complex I, whereas the same site is occluded by bulky residues in non-quinone binding Complex I-like proteins. AF2 models lack this quinone tunnel. MK: menaquinone. Acronyms for the complexes, removing the taxonomic suffix and adding numbering Kari-1 (GCA_015523565), Kari-2 (GCA_024280435), Njord-1 (M288), Njord-2 (D4998_C1112_H3_Bin_254), Hod-1 (M3_38_Bin_455), Hod-2 (GCA_024276395), Gerd (GCA_016839405), Asgard-1 (AB3033_2), and Asgard-2 (GCA_021162905).



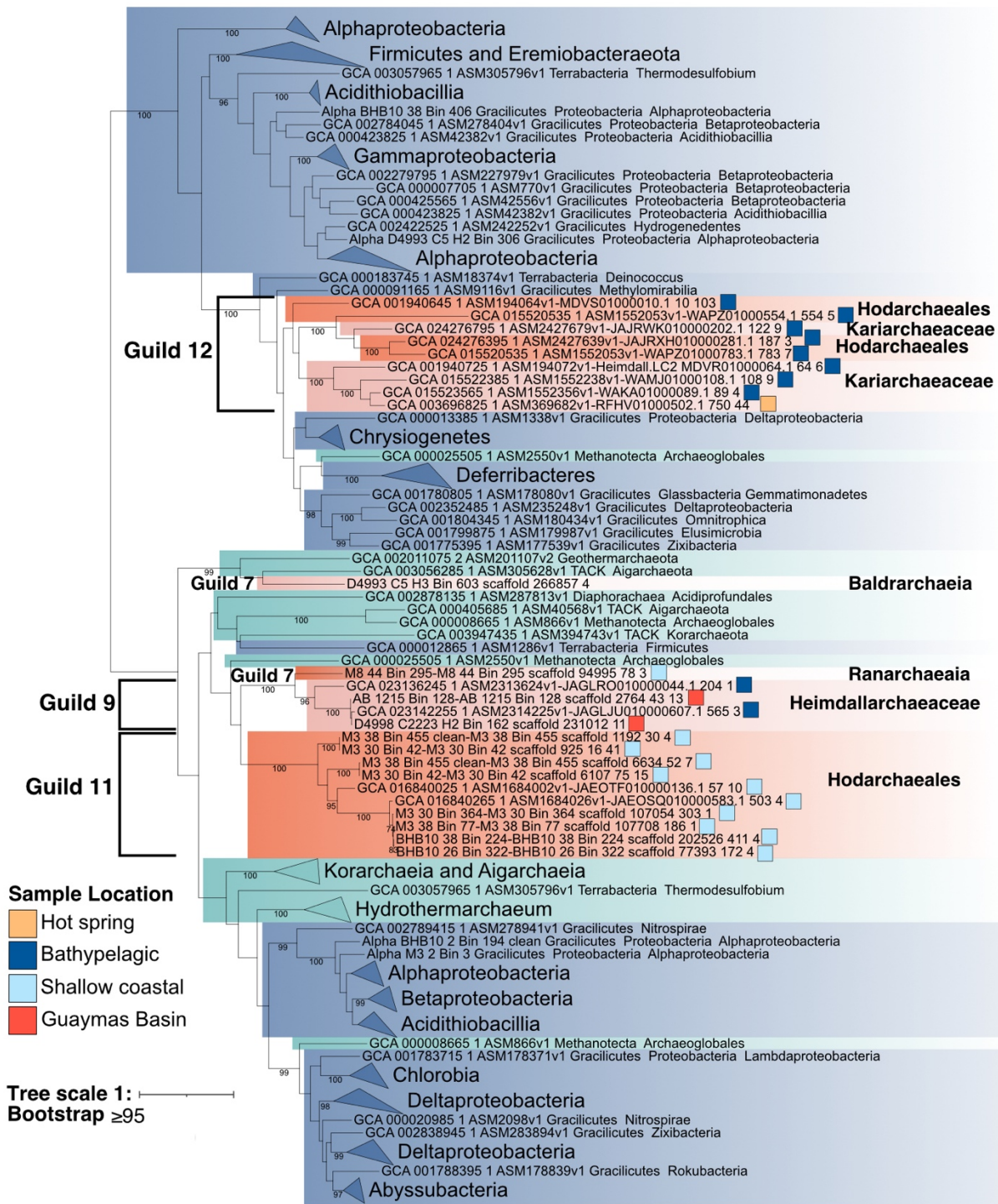
Supplementary Fig. 52 | **Lack of quinone binding site in group 4 [NiFe]-hydrogenase membrane-bound complexes.** Structural superposition of AF2 models with the MsComplex I quinone binding site illustrates that quinone binding is unlikely. *Mycobacterium smegmatis* Complex I (PDB ID: 8E9H)³¹ binds to menaquinone, and this binding is likely made through hydrogen bonds with a threonine and histidine in NuoD. A tyrosine is also implicated in transferring electrons from the N2 FeS cluster to reduce menaquinone⁹⁶. The equivalent binding site in our AF2 models contain substitutions that suggests they do not bind quinone. Notably, the key tyrosine is substituted with a glycine or alanine, and the histidine is substituted with a bulkier lysine or glutamate which occupies the quinone space. Numbering of AF2 models based on alignment with *Mycobacterium smegmatis* NuoD and NuoB (UniProt ID: A0QU33 and A0QU35 respectively). Contacts are depicted with yellow dashed lines and numbers correspond to atomic distance in ångström. MK: menaquinone.



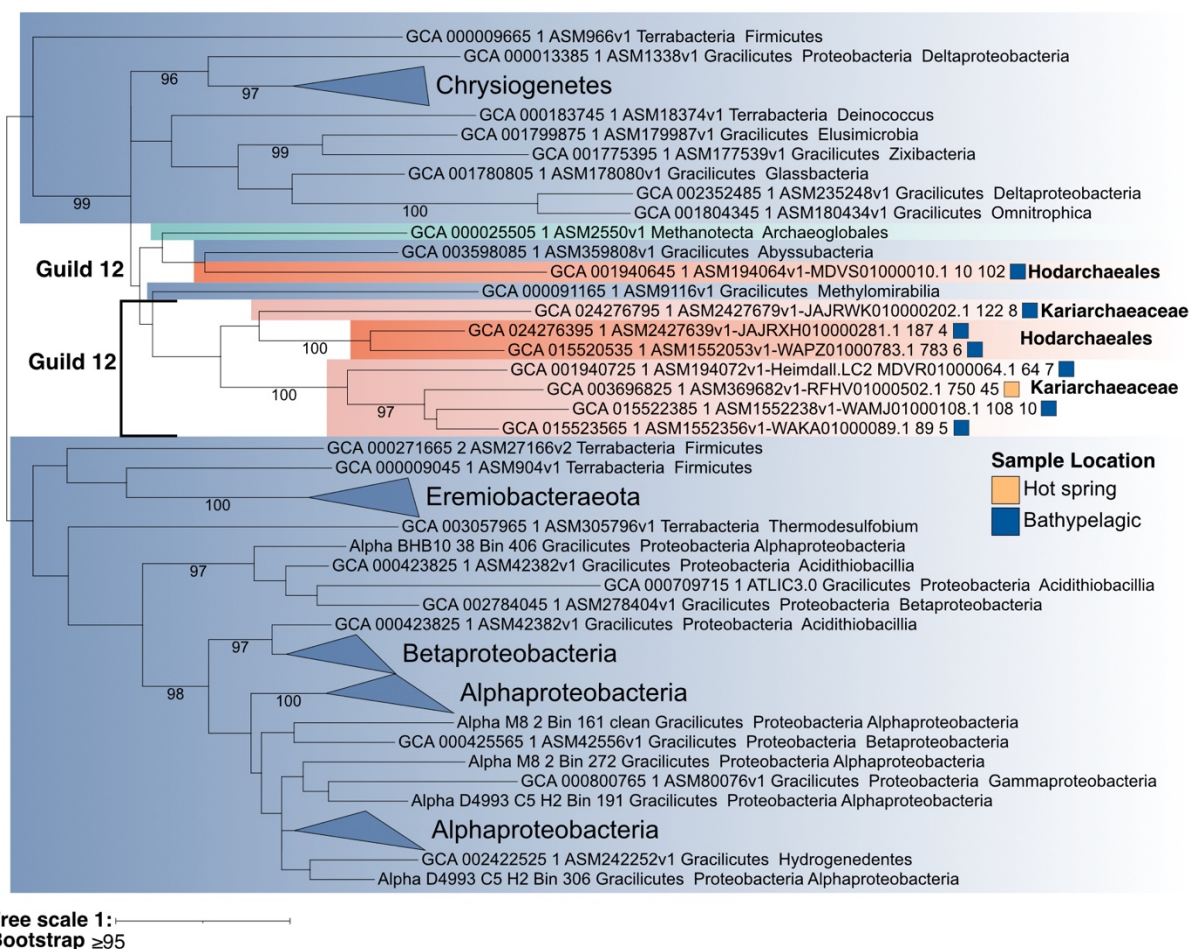
Supplementary Fig. 53 | **Nitrate Reductase Complex subunit G (NarG) phylogeny, catalyzing reduction of nitrate to nitrite.** Maximum-likelihood phylogeny was inferred from a supermatrix of 146 sequences and 1237 amino acid positions (from 2113 initial positions), displaying the diversity of NarG subunits across Asgard archaea (red), other archaeal lineages (teal), and bacteria (dark blue). The tree was generated using IQ-TREE v2.0.7 with LG+R7 best-fit model chosen according to Bayesian Information Criterion (BIC) with 1000 ultrafast bootstrap replicates (shown ≥ 95). The two external rings show the Asgard lineage and source location, respectively. The phylogeny was visualized with iTOL and midpoint rooted.



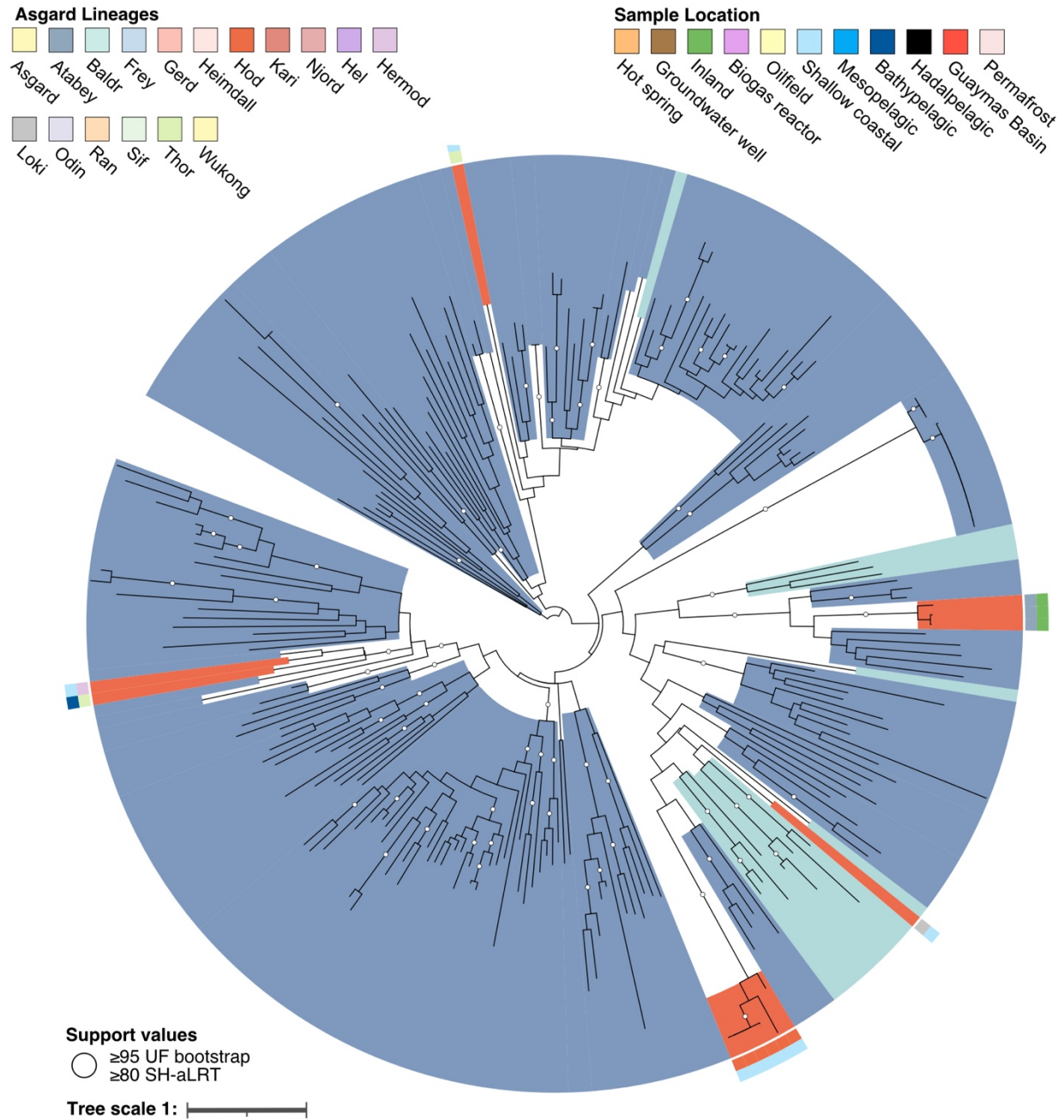
Supplementary Fig. 54 | **Nitrate Reductase Complex subunits H (NarH) and Y (NarY) phylogeny, catalyzing reduction of nitrate to nitrite.** Maximum-likelihood phylogeny was inferred from a supermatrix of 162 sequences and 487 amino acid positions (from 2024 initial positions), displaying the diversity of NarH and NarY subunits across Asgard archaea (red), other archaeal lineages (teal), and bacteria (dark blue). The tree was generated using IQ-TREE v2.0.7 with LG+R7 best-fit model chosen according to Bayesian Information Criterion (BIC) with 1000 ultrafast bootstrap replicates (shown ≥ 95). The two external rings show the Asgard lineage and source location, respectively. The phylogeny was visualized with iTOL and midpoint rooted.



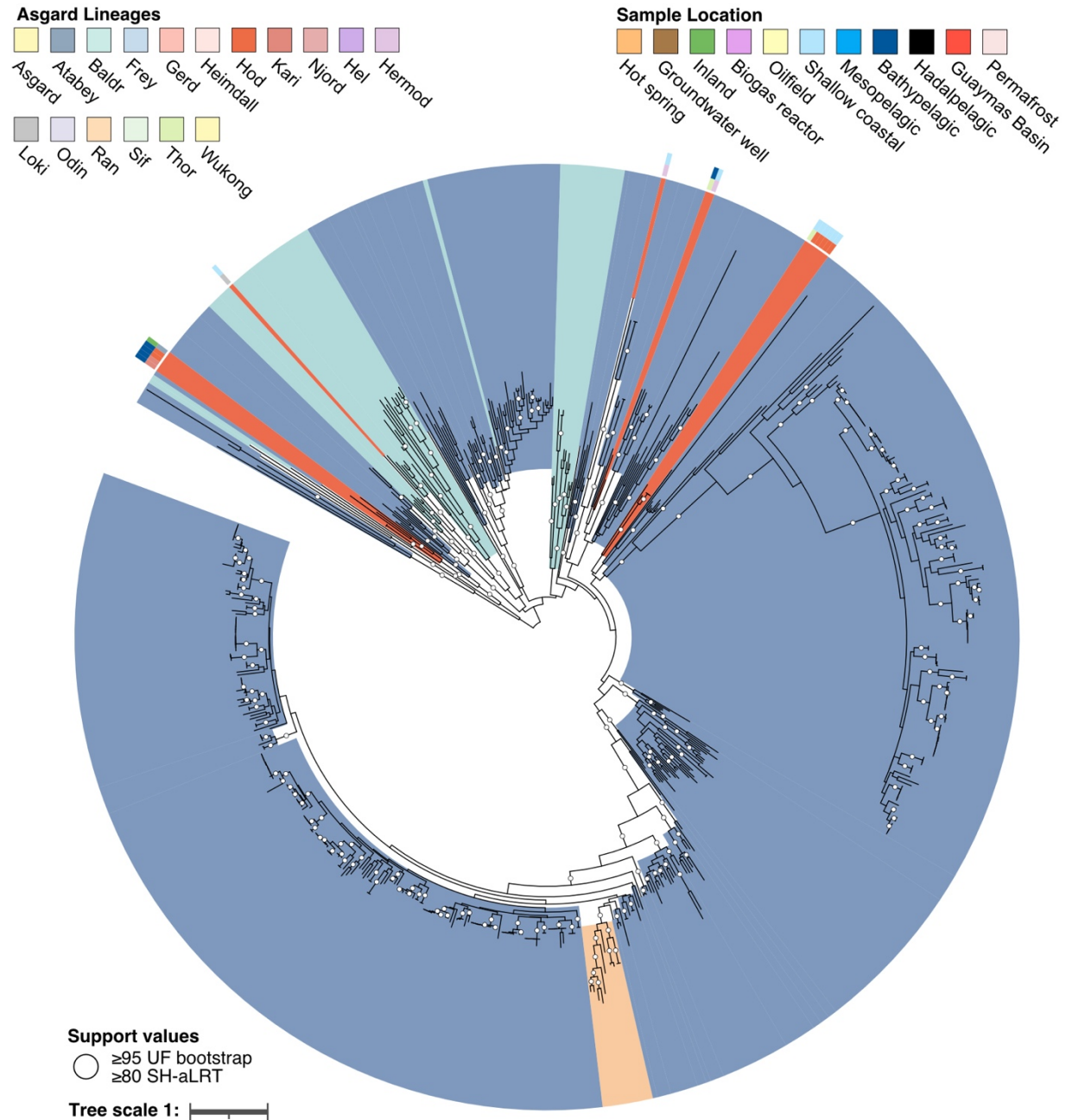
Supplementary Fig. 55 | **Nitrate Reductase Complex subunit I (NarI) phylogeny, catalyzing reduction of nitrate to nitrite.** Maximum likelihood phylogeny was inferred from a supermatrix of 196 sequences and 225 amino acid positions (from 1190 initial positions), displaying the diversity of NarI subunits across Asgard archaea (red), other archaeal lineages (teal), and bacteria (dark blue). The tree was generated using IQ-TREE v2.0.7 with LG+F+R5 best-fit model chosen according to Bayesian Information Criterion (BIC) with 1000 ultrafast bootstrap replicates (shown ≥ 95). The two external rings show the Asgard lineage and source location, respectively. The phylogeny was visualized with iTOL and midpoint rooted.



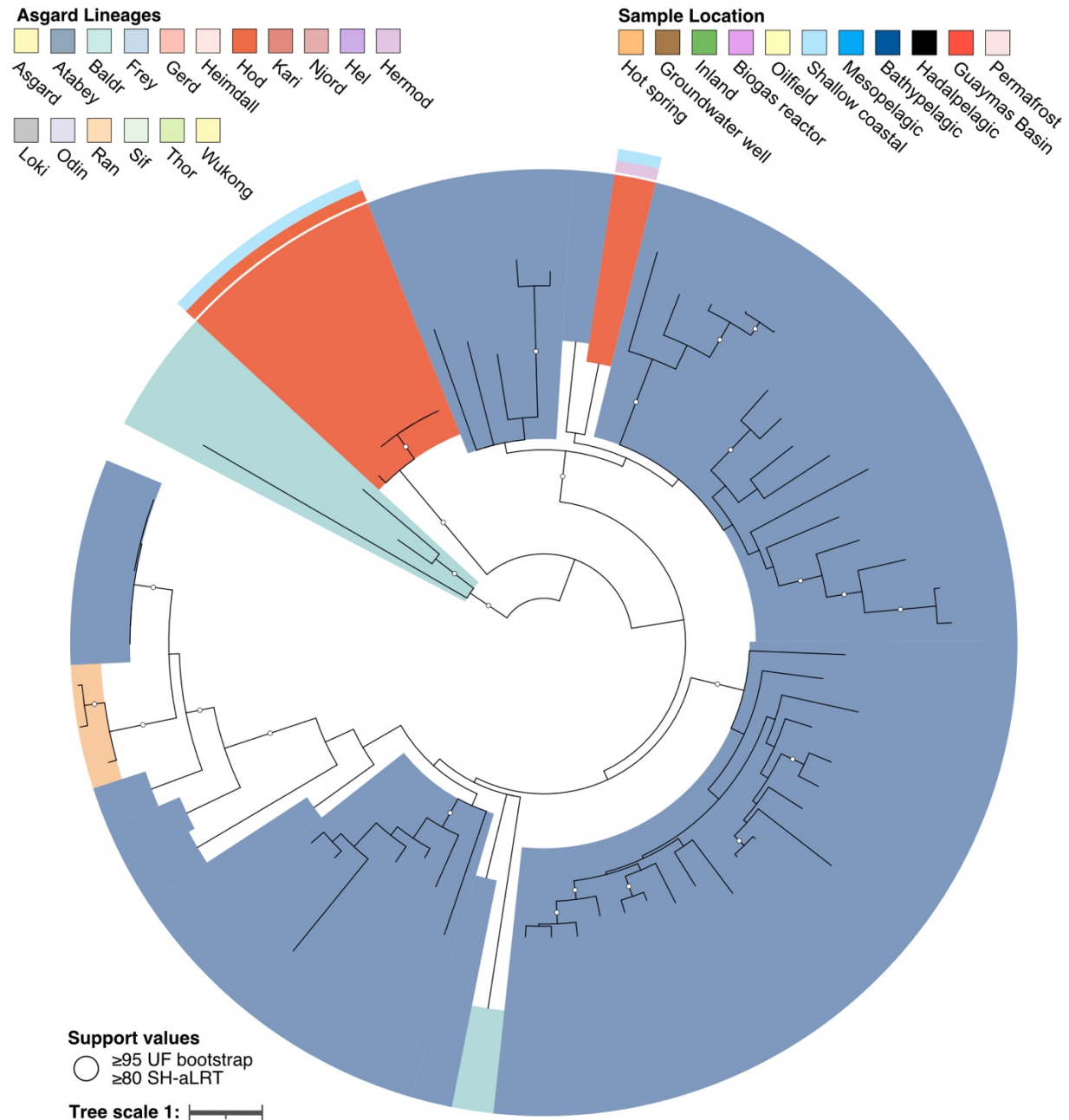
Supplementary Fig. 56 | **Nitrate Reductase Complex subunit J (NarJ) phylogeny, catalyzing reduction of nitrate to nitrite.** Maximum likelihood phylogeny was inferred from a supermatrix of 82 sequences and 224 amino acid positions (from 480 initial positions), displaying the diversity of NarJ subunits across Asgard archaea (red), other archaeal lineages (teal), and bacteria (dark blue). The tree was generated using IQ-TREE v2.0.7 with LG+R5 best-fit model chosen according to Bayesian Information Criterion (BIC) with 1000 ultrafast bootstrap replicates (shown ≥ 95). The two external rings show the Asgard lineage and source location, respectively. The phylogeny was visualized with iTOL and midpoint rooted.



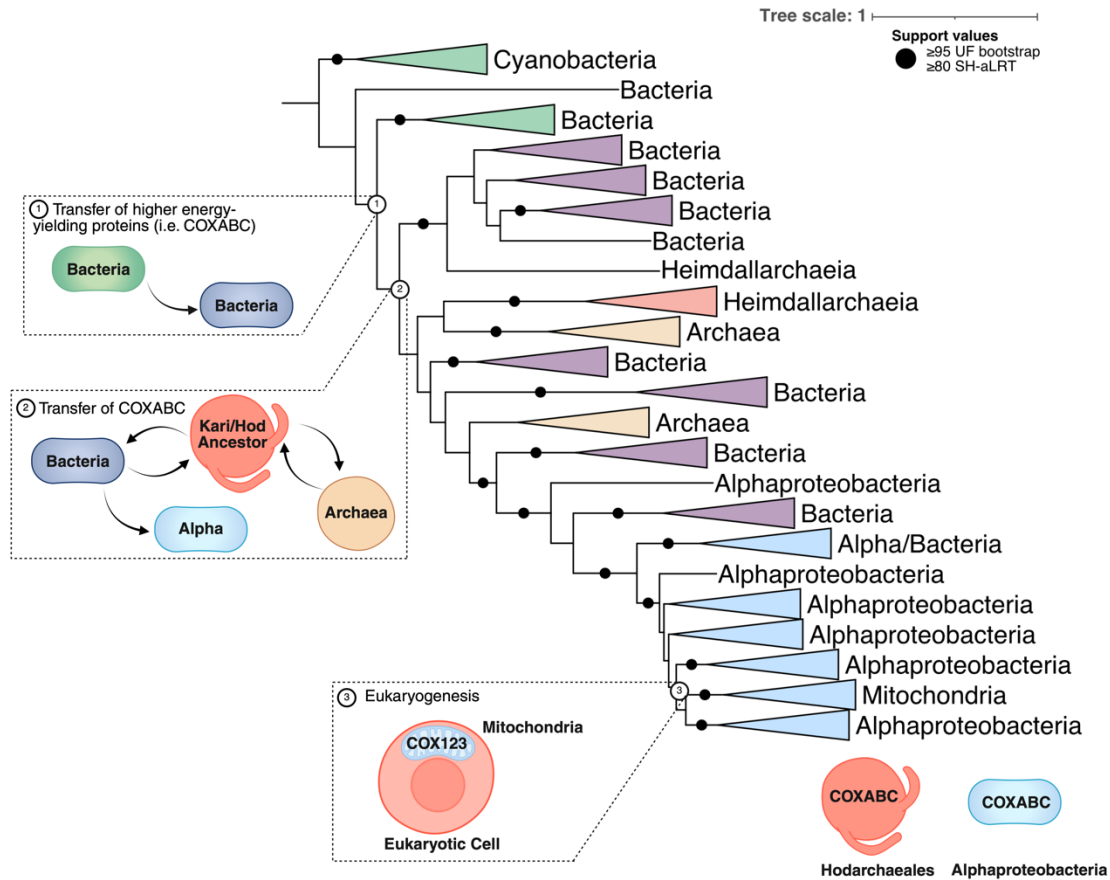
Supplementary Fig. 57 | **Ubiquinol-Cytochrome c Reductase Subunit A (QcrA) phylogeny.** Maximum likelihood phylogeny was inferred from a supermatrix of 241 sequences and 148 amino acid positions (from 1543 initial positions), displaying the diversity of QcrA subunits across Asgard archaea (red), other archaeal lineages (teal), and bacteria (dark blue). The tree was generated using IQ-TREE v2.0.7 (1000 ultrafast bootstrap replicates and calculation of 1000 SH-aLRT) with WAG+C50+R8 best-fit model chosen according to Bayesian Information Criterion (BIC). The two external rings show the Asgard lineage and source location, respectively. Support values are shown when ultrafast (UF) bootstrap support is greater than or equal to 95 and SH-aLRT support is greater than or equal to 80. The phylogeny was visualized with iTOL and midpoint rooted. The most deeply rooted bacterial clades also are annotated as NirD, which explain the longer branches. These cluster away from the Asgardarchaeota sequences.



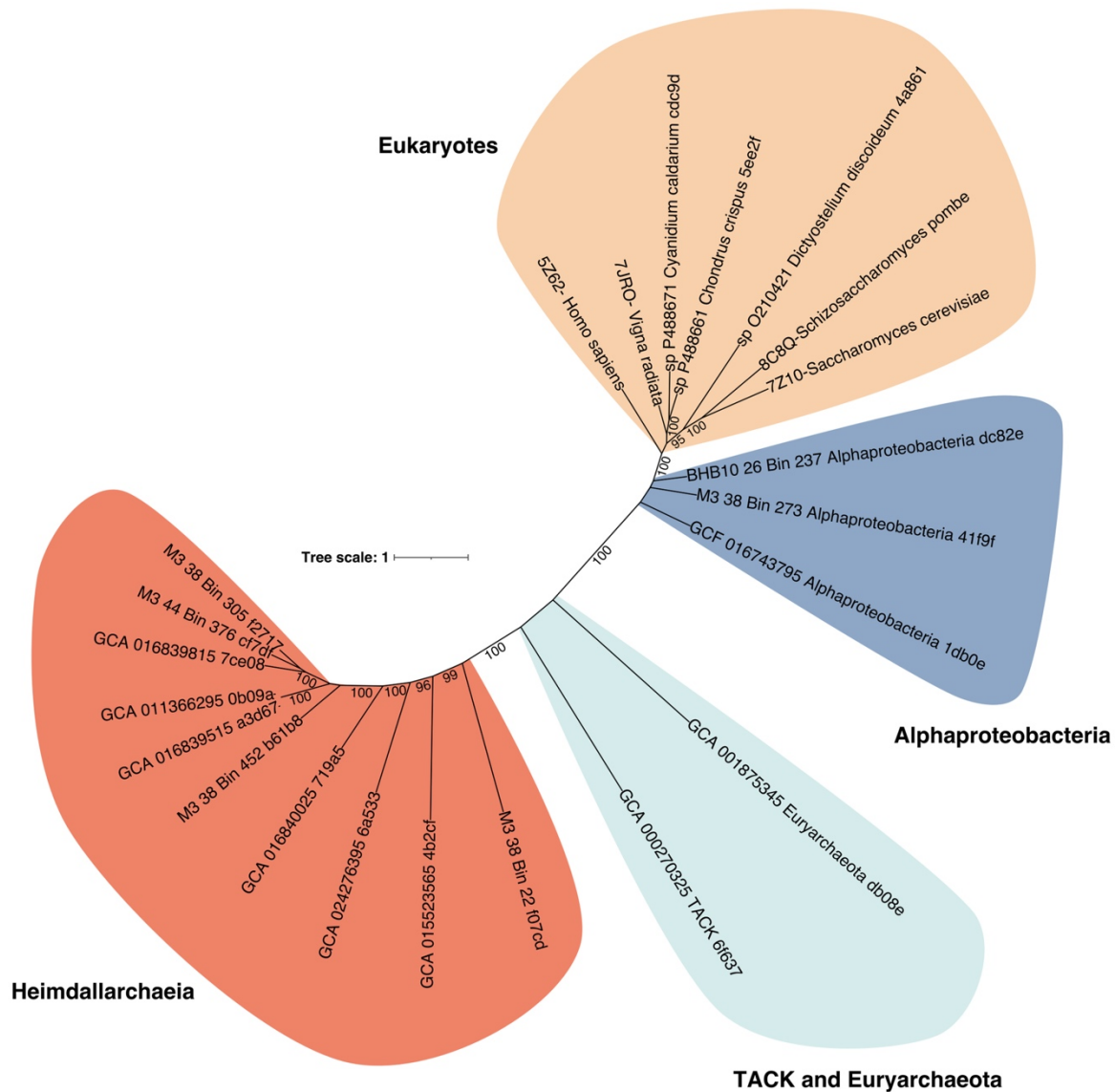
Supplementary Fig. 58 | **Ubiquinol-Cytochrome c Reductase Subunit B (QcrB) phylogeny.** Maximum likelihood phylogeny was inferred from a supermatrix of 623 sequences and 380 amino acid positions (from 2719 initial positions), displaying the diversity of QcrB subunits across Asgard archaea (red), other archaeal lineages (teal), bacteria (dark blue), and mitochondria (orange). The tree was generated using IQ-TREE v2.0.7 (1000 ultrafast bootstrap replicates and calculation of 1000 SH-aLRT) with WAG+C50+R8 best-fit model chosen according to Bayesian Information Criterion (BIC). The two external rings show the Asgard lineage and source location, respectively. Support values are shown when ultrafast (UF) bootstrap support is greater than or equal to 95 and SH-aLRT support is greater than or equal to 80. The phylogeny was visualized with iTOL and midpoint rooted.



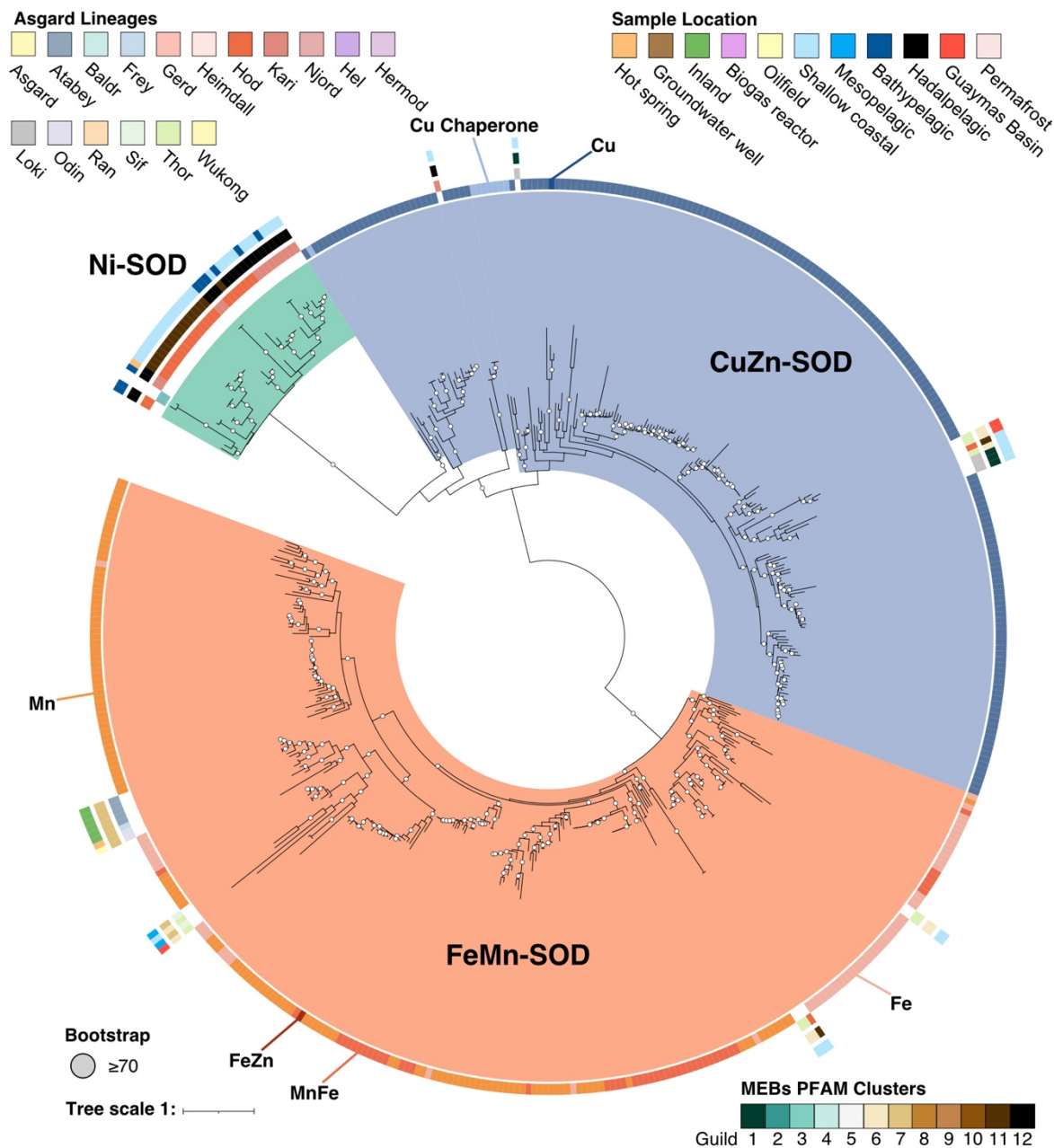
Supplementary Fig. 59 | **Ubiquinol-Cytochrome c Reductase Subunit (QcrC) phylogeny.** Maximum likelihood phylogeny was inferred from a supermatrix of 70 sequences and 359 amino acid positions (from 979 initial positions), displaying the diversity of QcrC subunits across Asgard archaea (red), other archaeal lineages (teal), bacteria (dark blue), and mitochondria (orange). The tree was generated using IQ-TREE v2.0.7 (1000 ultrafast bootstrap replicates and calculation of 1000 SH-aLRT) with WAG+C50+R6 best-fit model chosen according to Bayesian Information Criterion (BIC) with ultrafast bootstrap with 1000 replicates. The two external rings show the Asgard lineage and source location, respectively. Support values are shown when ultrafast (UF) bootstrap support is greater than or equal to 95 and SH-aLRT support is greater than or equal to 80. The phylogeny was visualized with iTOL and midpoint rooted.



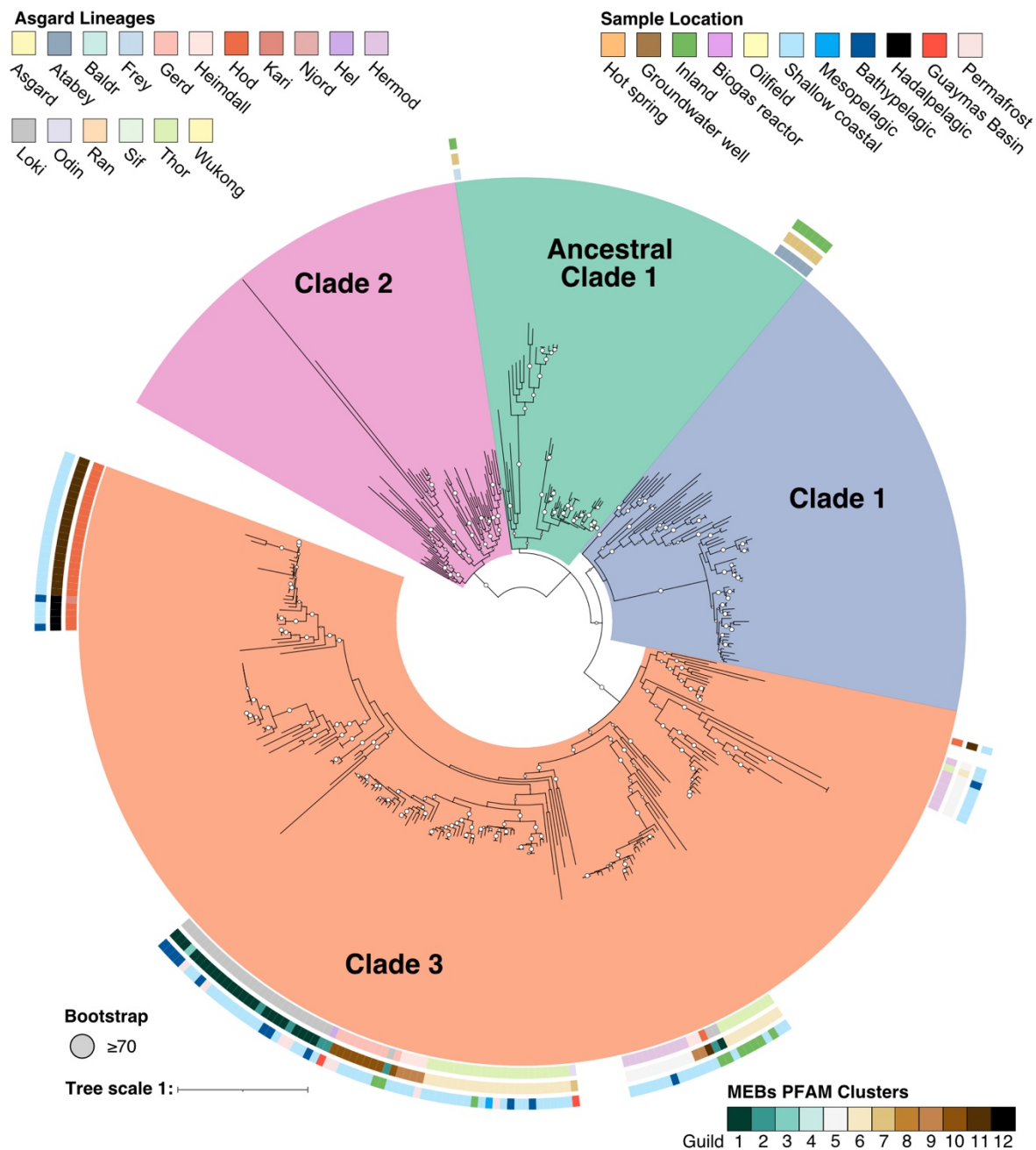
Supplementary Fig. 60 | **Tracing transfers of cytochrome c oxidase ABC (COXABC) in the scheme of the Great Oxidation Event.** Displaying potential transfers of COXABC between the three domains of life mapped onto the phylogeny from Fig. 3b (see methods for details). The phylogeny is rooted with Cyanobacteria and shows potential lateral transfer between bacterial groups (green and purple) (1). Next, the phylogenomic analysis indicates various potential transfers between Heimdallarchaeia groups (red), bacteria (purple), and archaea (orange) (2). However, mitochondrial subunits cluster with Alphaproteobacteria, supporting the eukaryotic cell COX123 (homologs of COXABC) came from the mitochondrial ancestor (3). This is one potential timeline of events and future ancestral reconstructions will further elucidate this transition. Support values are shown when ultrafast (UF) bootstrap support is greater than or equal to 95 and SH-aLRT support is greater than or equal to 80. The phylogeny was visualized with iTOL and final figure created in BioRender. Appler, K. (2025) <https://BioRender.com/36a3dnh>.



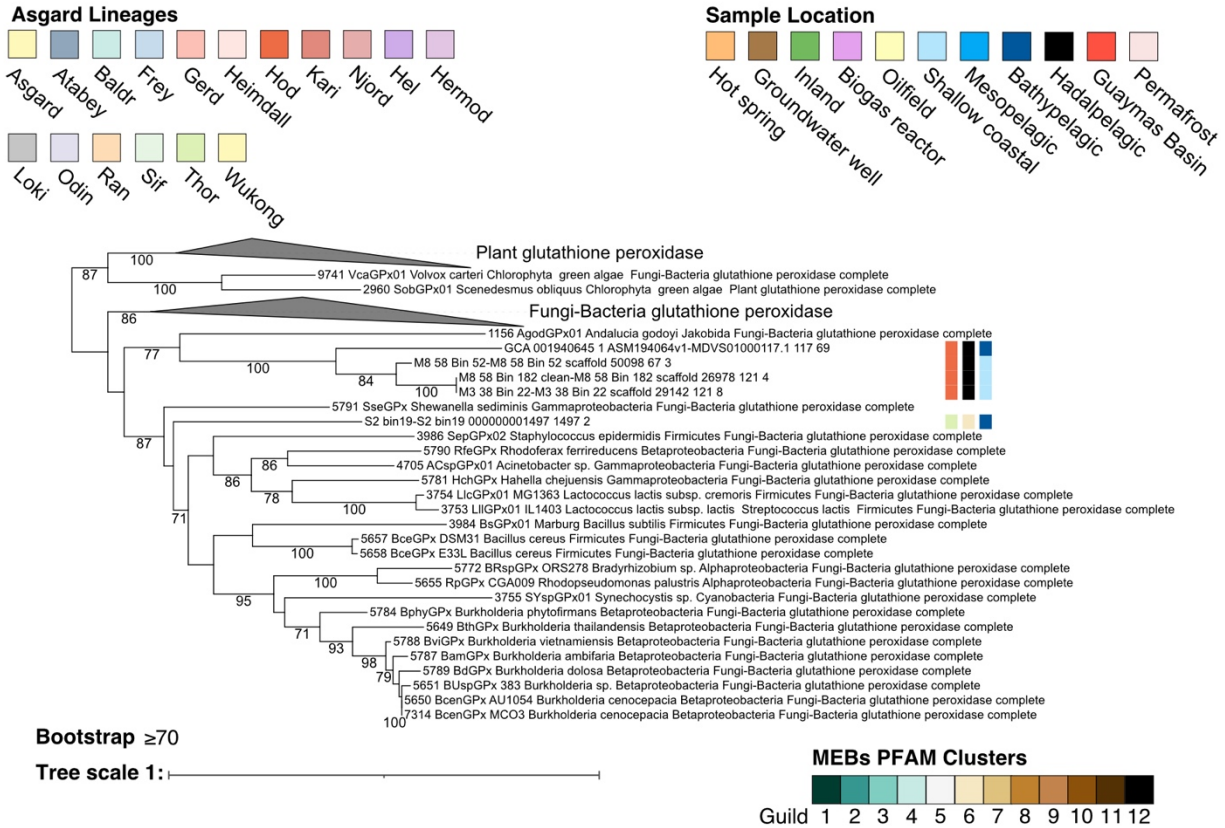
Supplementary Fig. 61 | **Cytochrome c oxidase structural phylogeny** of selected proteins from Heimdallarchaeia (red), TACK and Euryarchaeota (light blue), Bacteria (dark blue), and Eukaryote mitochondria (yellow), supporting the sequence homology-based phylogeny (Fig. 3b). The tree was inferred from a supermatrix of 22 sequences and 3658 positions. The protein references are listed for each genome and correspond to the files uploaded on Figshare. The tree was inferred using IQ-TREE v2.1.4-beta with LG+F+R5 best-fit model chosen according to Bayesian Information Criterion (BIC) with 1000 ultrafast bootstrap replicates (shown ≥ 95). The phylogeny was visualized with iTOL.



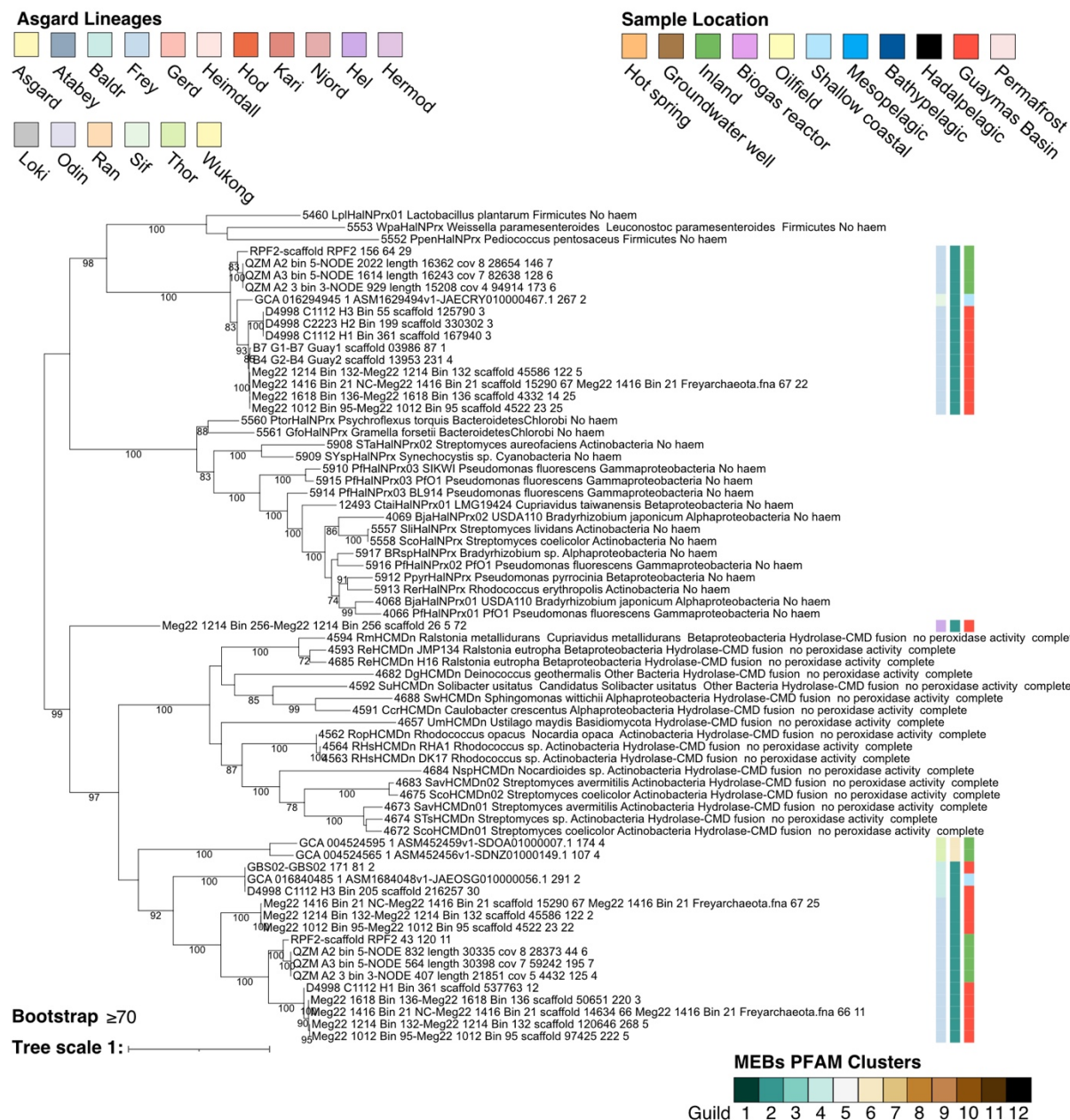
Supplementary Fig. 62 | **Maximum likelihood phylogeny of Copper/Zinc (CuZn), Nickel (Ni), and Iron/Manganese (FeMn) Asgardarchaeota superoxide dismutase (SOD).** We identified 63 Asgardarchaeota SOD sequences aligned with 192 Cu/Zn (PF00080; blue), 2 Ni (PF09055; teal), and 236 FeMn (IPR001189; red/orange reviewed references). A few references also had variations on the major SOD families (e.g., Mn, FeZn, MnFe, Fe, Cu, and Cu Chaperon) and are shown in the innermost ring. The external three rings display the Asgardarchaeota lineage, guild, and sour environment (in-out). The sequences were aligned and trimmed with MAFFT v7.490 and trimAl v1.4.rev15. The tree was generated with IQ-TREE v2.0.7 (-bb 1000 -bnni), model WAG+R6 was chosen according to Bayesian Information Criterion (BIC). The tree was inferred from a supermatrix of 493 sequences and 189 amino acid positions (from 992 initial positions). Ultrafast bootstrap values greater than or equal to 70 are shown but should only be relied on when ≥ 95 . The phylogeny was visualized with iTOL and rooted (between Ni-SOD/CuZn-SOD and FeMn-SOD).



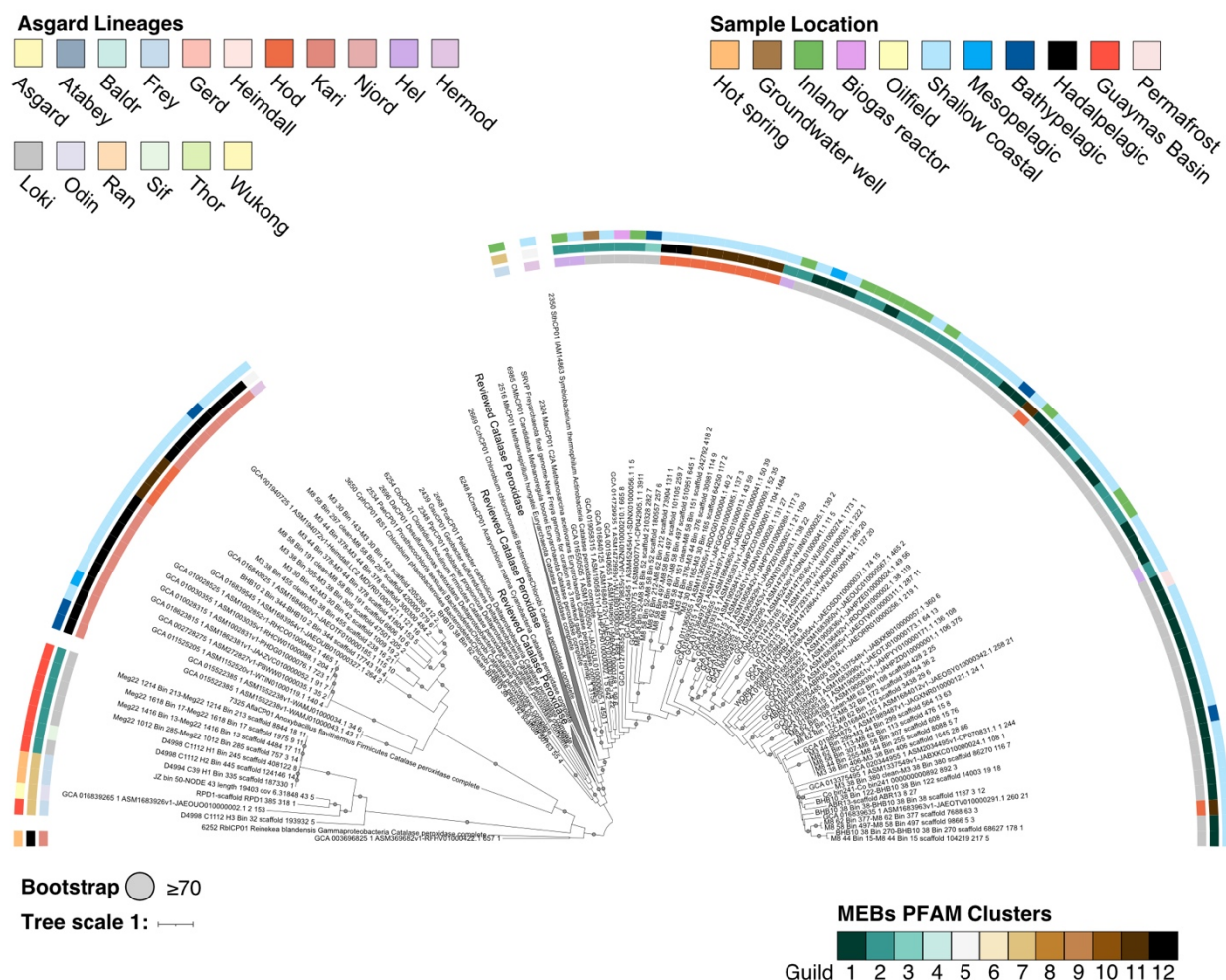
Supplementary Fig. 63 | **Distribution of catalases groups across the Asgardarchaeota phylum.** We identified 129 Asgardarchaeota catalase sequences (KatE) aligned with 278 references. The references included reviewed sequences from IPR024711 (clades 1 teal/purple and 3 orange) and IPR024712 (clade 2 pink) and subset (via MMSeq2¹⁶⁴) of Yuan et al., 2021⁸⁰ references⁹. The sequences were aligned and trimmed with MAFFT v7.490 and trimAl v1.4.rev15. The maximum-likelihood phylogeny was generated with IQ-TREE v2.0.7 (-bb 1000 -bnni), model LG+R8 was chosen according to Bayesian Information Criterion (BIC). The tree was inferred from a supermatrix of 407 sequences and 485 amino acid positions (from 1455 initial positions). The rings display the Asgardarchaeota lineage and source environment. Ultrafast bootstrap values greater than or equal to 70 are shown but should only be relied on when ≥ 95 . The phylogeny was visualized with iTOL and rooted, separating Clade 2 from the other known groups.



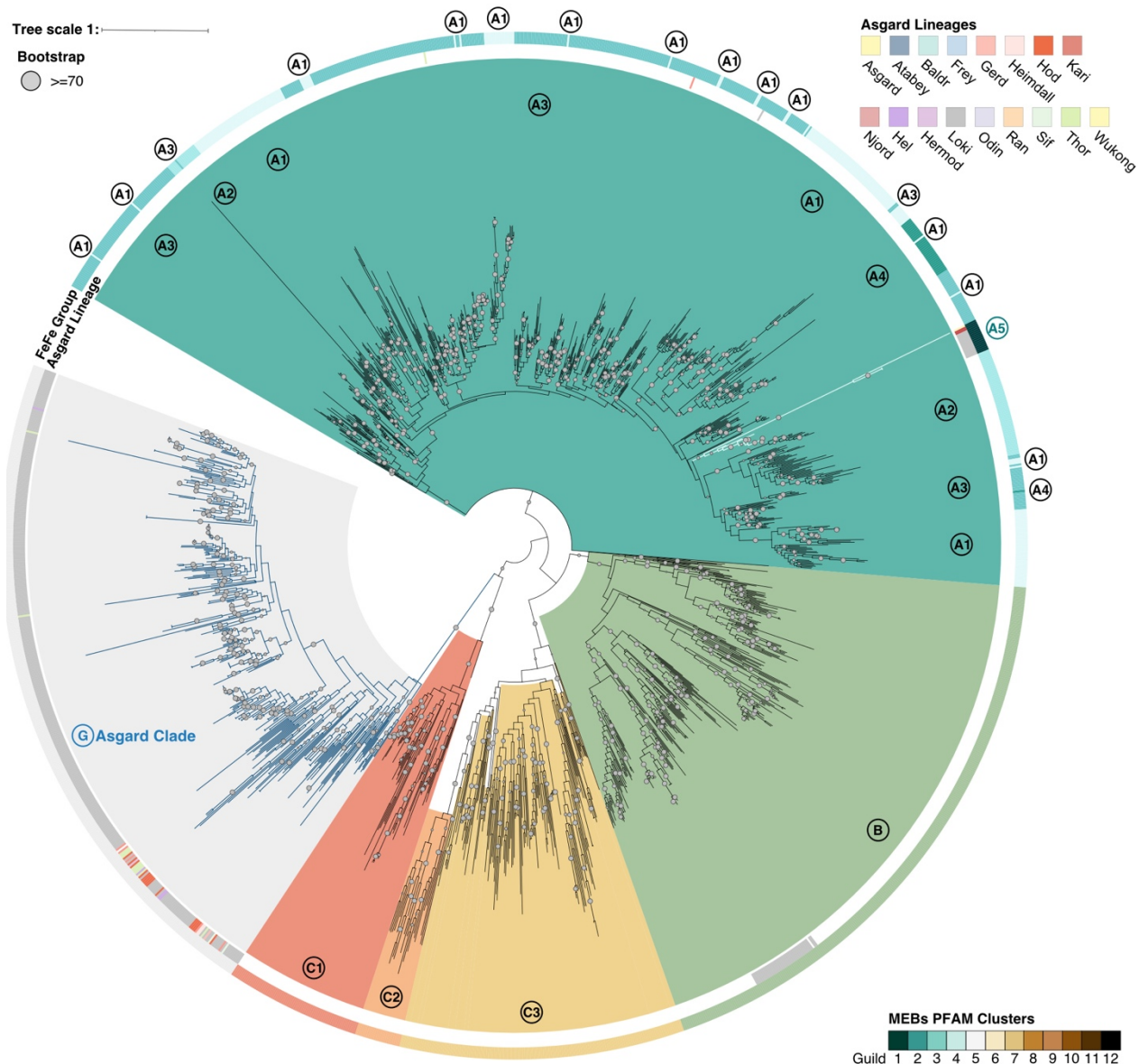
Supplementary Fig. 64 | **Maximum-likelihood phylogeny of glutathione peroxidases in Asgardarchaeota.** We identified 5 Asgardarchaeota (Hodarchaeales guild 12 and Thorarchaeia guild 6) peroxidases annotated as glutathione peroxidase (K00432). A few representatives were blasted against the PeroxiBase/RedoxiBase¹⁶⁵ in July 2024 to produce a non-redundant set of references (169 glutathione peroxidases). Sequences were aligned with MAFFT v7.490 and trimmed with trimAl v1.4.rev15, separately. IQ-TREE v2.0.7 (-bb 1000 -bnni) was used to produce the maximum-likelihood phylogeny, using models LG+I+G4 chosen according to Bayesian Information Criterion (BIC). The tree was inferred from a supermatrix of 174 sequences and 167 amino acid positions (from 392 initial positions). The rings display the Asgardarchaeota lineage, guild, and sour environment (in-out). Ultrafast bootstrap values greater than or equal to 70 are shown but should only be relied on when ≥ 95 . The phylogeny was visualized with iTOL and midpoint rooted.



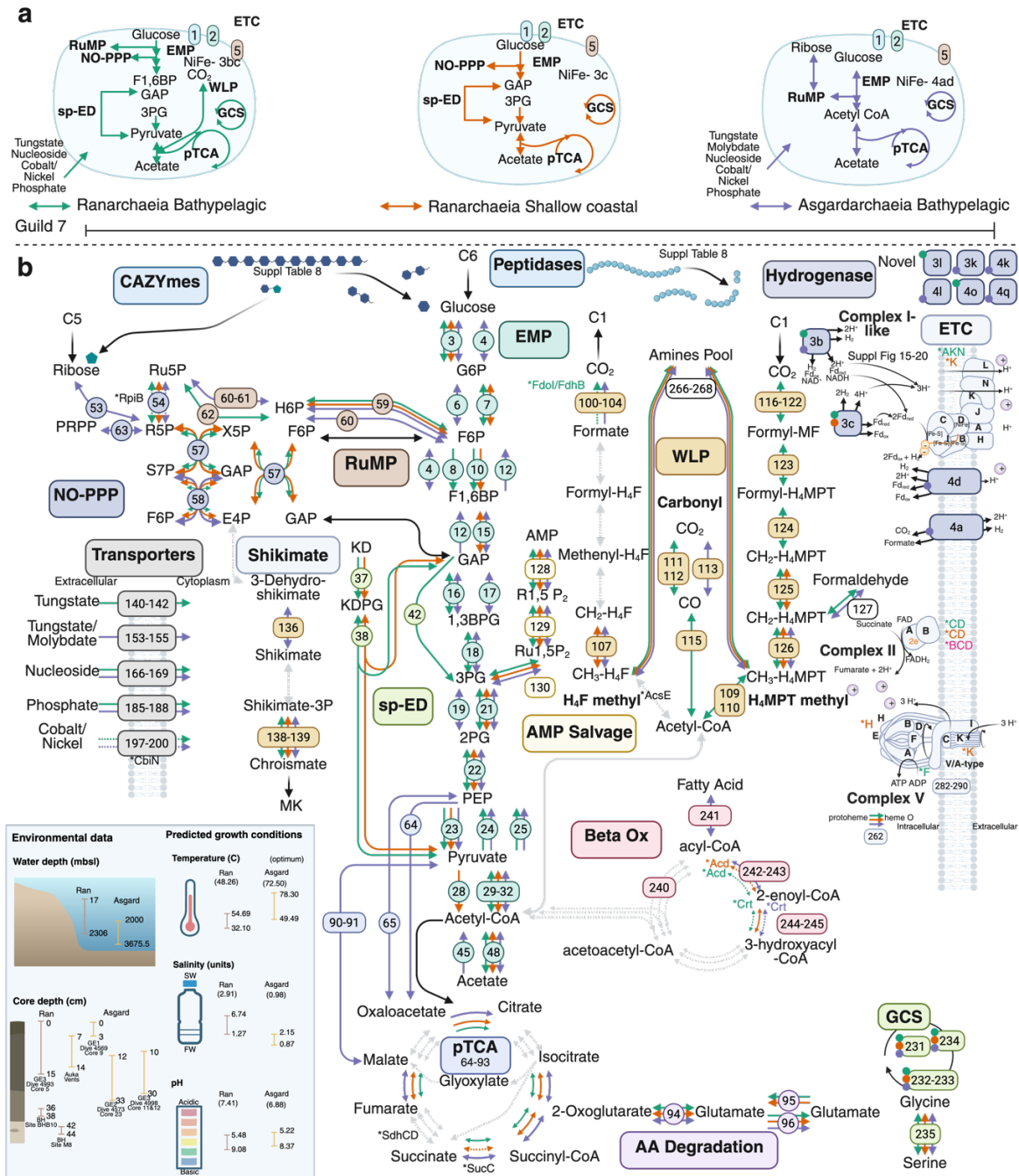
Supplementary Fig. 65 | **Maximum-likelihood phylogeny of non-heme peroxidases in Asgardarchaeota.** We identified 32 Asgardarchaeota peroxidases annotated as non-heme peroxidases (K00432) or more no haem no metal haloperoxidase (PS52061). Sequences included metabolic guilds 2 (Helarchaeales), 6 (Thorarchaeia), but primarily 7 (Frey/Jordarchaeia, Baldarchaeia, and Sifarchaeia). A few representatives were blasted against the PeroxiBase/RedoxiBase¹⁶⁵ in July 2024 to produce a non-redundant set of references (37 non-heme peroxidases). Sequences were aligned with MAFFT v7.490 and trimmed with trimAl v1.4.rev15, separately. IQ-TREE v2.0.7 (-bb 1000 -bnni) was used to produce the maximum-likelihood phylogeny, using model LG+R4 chosen according to Bayesian Information Criterion (BIC). The tree was inferred from a supermatrix of 69 sequences and 263 amino acid positions (from 772 initial positions). The rings display the Asgardarchaeota lineage, guild, and sour environment (in-out). Ultrafast bootstrap values greater than or equal to 70 are shown but should only be relied on when ≥ 95 . The phylogeny was visualized with iTOL and midpoint rooted.



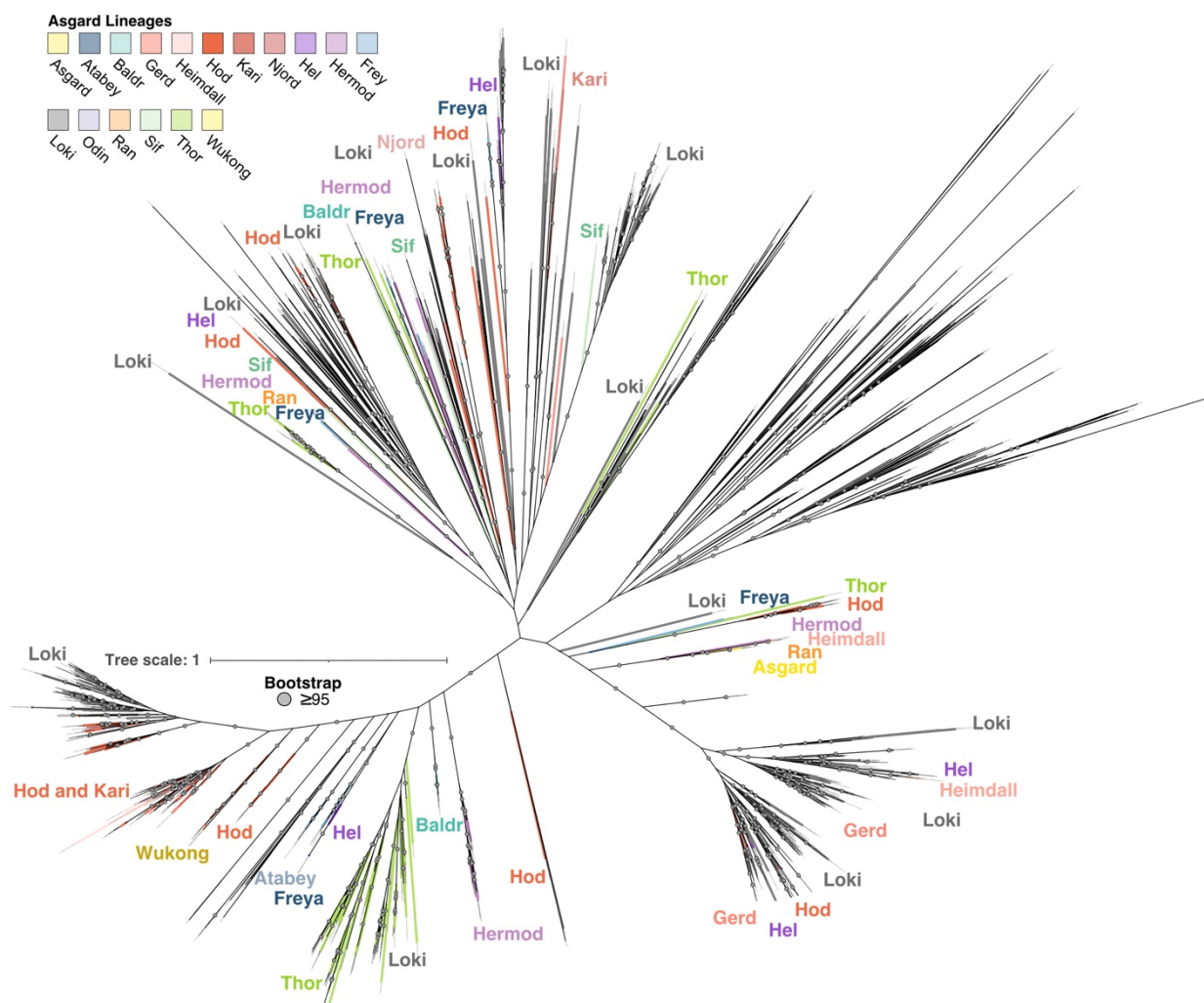
Supplementary Fig. 66 | **Maximum-likelihood phylogeny of catalase-peroxidases in Asgardarchaeota.** We identified 100 Asgardarchaeota peroxidases annotated as catalase-peroxidases (K03782). A few representatives were blasted against the PeroxiBase/RedoxiBase¹⁶⁵ in July 2024 to produce a non-redundant set of references (113 catalase peroxidases). Sequences were aligned with MAFFT v7.490 and trimmed with trimAl v1.4.rev15, separately. IQ-TREE v2.0.7 (-bb 1000 -bnni) was used to produce the maximum-likelihood phylogeny, using model LG+R7 chosen according to Bayesian Information Criterion (BIC). The tree was inferred from a supermatrix of 213 sequences and 731 amino acid positions (from 1119 initial positions). The rings display the Asgardarchaeota lineage, guild, and sour environment (in-out). Bootstraps greater than or equal to 70 are shown but should only be relied on when ≥ 95 . The phylogeny was visualized with iTOL and midpoint rooted.



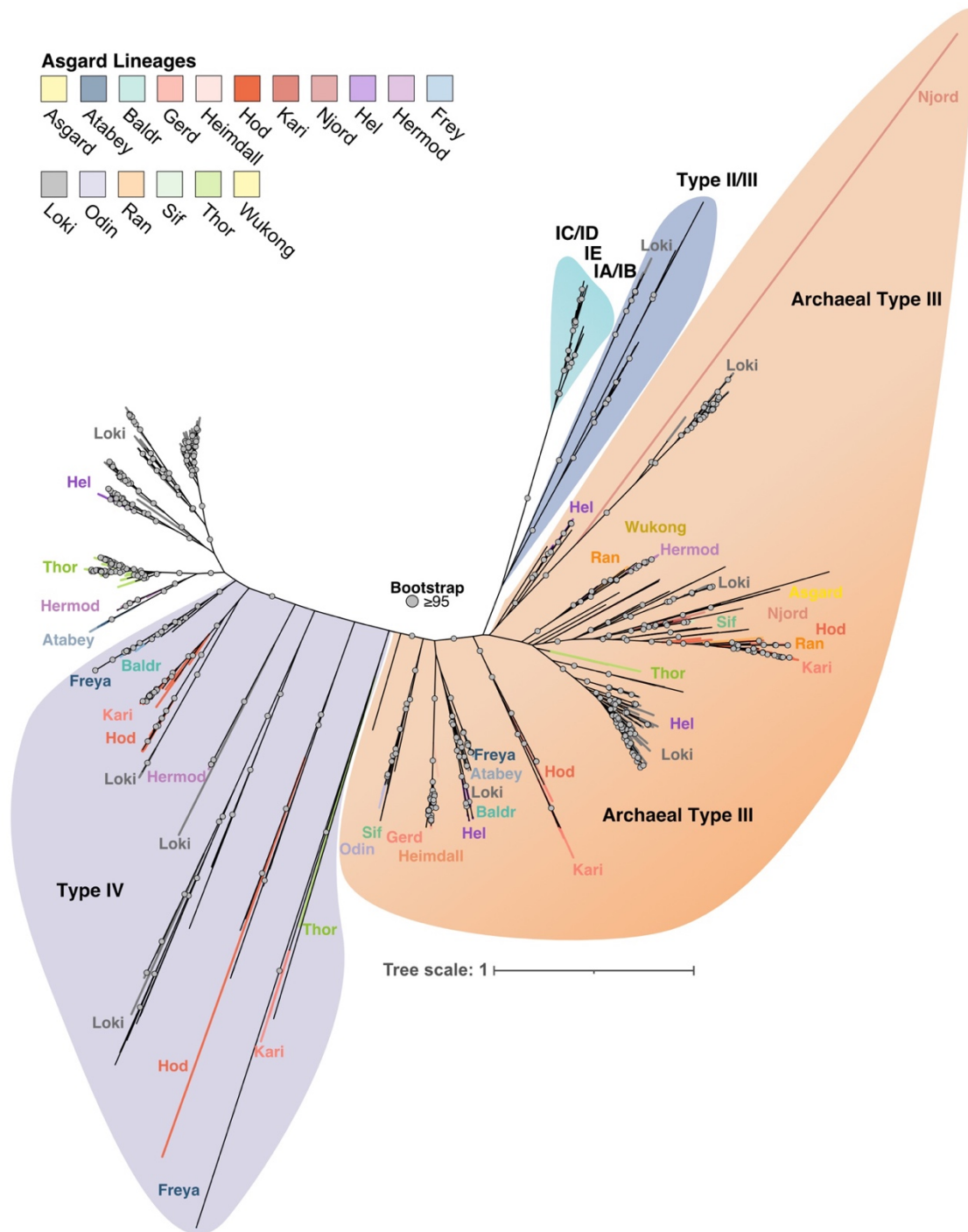
Supplementary Fig. 67 | **Expanded [FeFe]-hydrogenase diversity in Asgardarchaeota.** Maximum likelihood phylogeny of displaying the diversity of Asgard archaea hydrogenases denoted by lineage in the innermost ring, spanning [FeFe]-Groups A (teal), B (green), C3 (yellow), C2 (orange), C1 (red), and preliminarily G (gray) identified by the external ring. Previously identified subgroups are labeled in black. The potential novel hydrogenase groups identified and preliminarily named in this study, including A5 and G. Groups A5 and G are distinguished by white and dark blue branches, respectively. The phylogeny was generated using MAFFT v7.3.10, trimAl v1.4.rev15 (-gt 0.5), and IQ-TREE v1.6.1 with LG+R10 best-fit model chosen according to Bayesian Information Criterion (BIC) with 1000 ultrafast bootstrap replicates (shown ≥ 70). The tree was inferred from a supermatrix of 1639 sequences and 206 amino acid positions (from 5431 initial positions). Ultrafast bootstrap support should only be relied on when ≥ 95 . The phylogeny was midpoint rooted and visualized with iTOL.



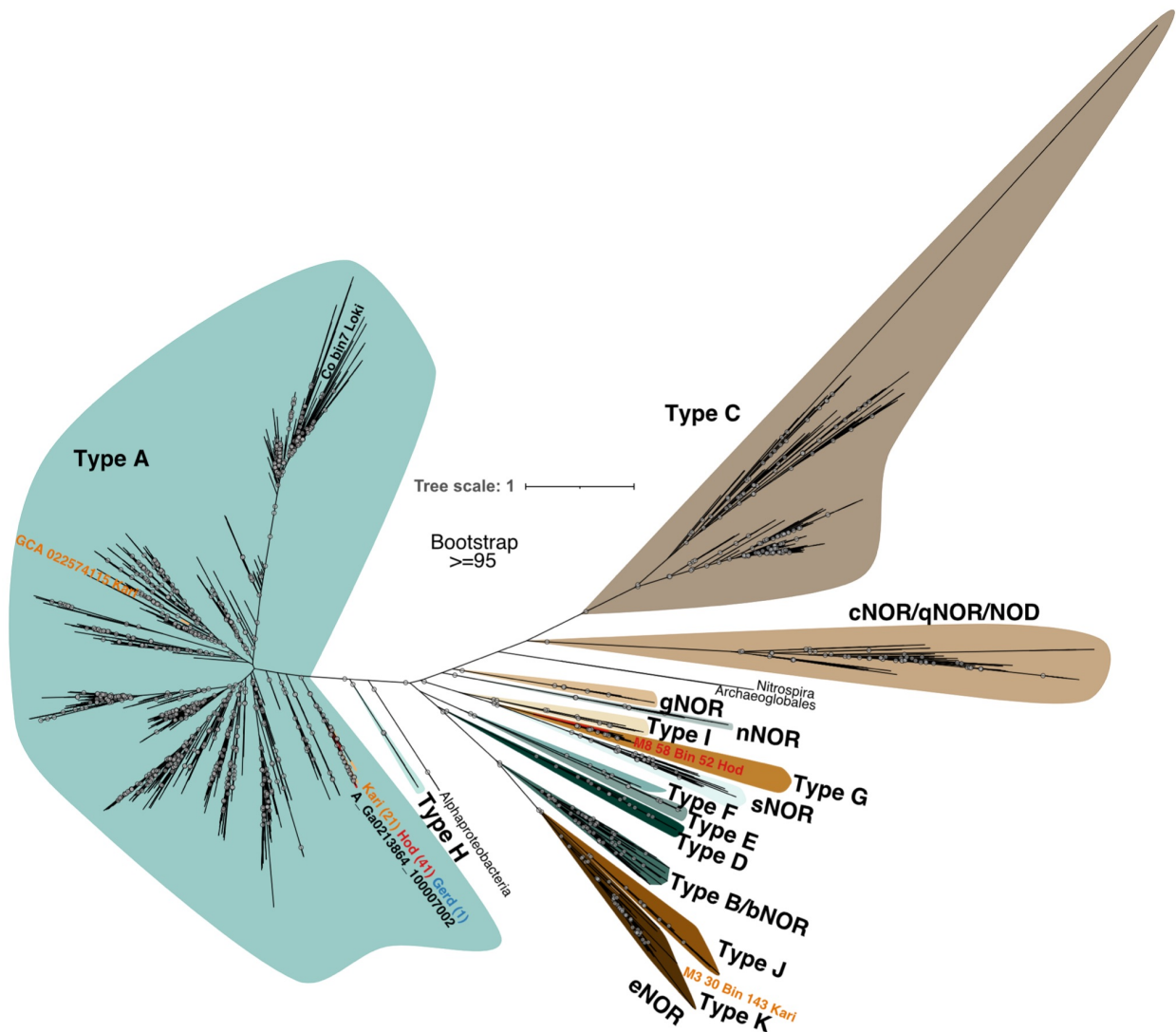
Supplementary Fig. 68 | **Metabolic reconstruction of Ranarchaea and Asgardarchaea across environments.** **a.** Overview of taxonomy and environmental subsets for novel lineage Ranarchaea from coastal (orange) and bathypelagic (green) and Asgardarchaea (previously only one MAG)⁸ from bathypelagic (purple) within guild 7. These are distinguished with colored arrows and dots. **b.** All numbers listed correspond to the numbers in Supplementary Table 8, which provides a metabolic overview for all Asgardarchaeota lineages. Missing subunits are listed in the same color as the arrows with an *. The inset shows the elevation and core depth for MAG representatives and predicted growth condition (temperature, salinity, pH). Abbreviations included RuMP (Ribulose Monophosphate Pathway), BO (Beta Oxidation), AA (Amino Acids), TCA /pTCA (Tricarboxylic Acid Cycle and partial version), WLP (Wood-Ljungdahl Pathway), GCS (Glycine Cleavage System), ETC (Electron Transport Chain), NO-PPP (Non-Oxidative Pentose Phosphate Pathway), EMP (Embden-Meyerhof-Parnas Pathway), sp-ED (semi-phosphorylative Entner-Doudoroff), C# (compound with # carbons), and CAZymes (Carbohydrate-Active enZymes). Asterisks denote pathways that are missing specific subunits to help compare across the environments. Created in BioRender. Appler, K. (2025) <https://BioRender.com/fs8a1vy>.



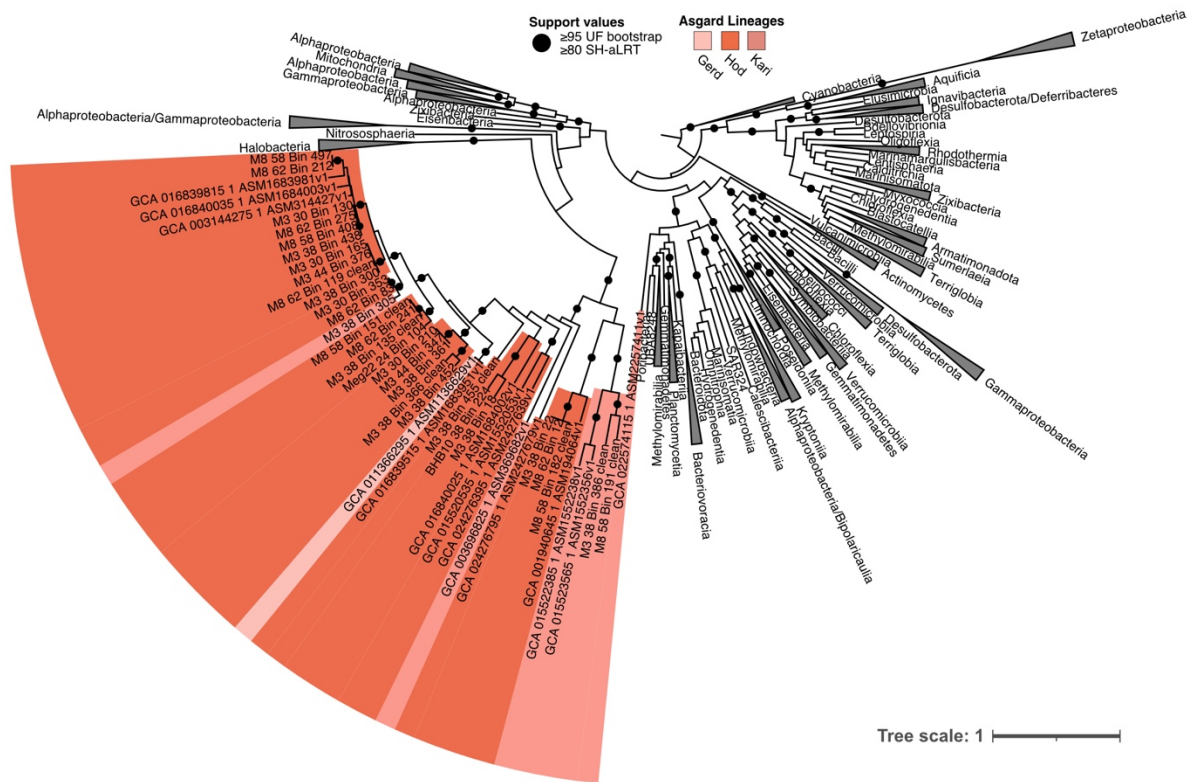
Supplementary Fig. 69 | **Trimethylamine methyltransferases in the expanded Asgardarchaeota genome catalog.** Maximum-likelihood phylogeny of Non-Pyl trimethylamine methyltransferase subunit (MttB). We identified 1,152 MttB Asgardarchaeota sequences, which were aligned with a subset of 294 references¹⁶⁶ based on protein identity. The branches are color-coded by Asgardarchaeota lineage. The sequences were aligned and trimmed with MAFFT v7.490 and trimAl v1.4.rev15. The tree was generated with IQ-TREE v2.0.7 (-bb 1000 -bnni), model LG+F+R10 was chosen according to Bayesian Information Criterion (BIC). The tree was inferred from a supermatrix of 1446 sequences and 476 amino acid positions (from 1461 initial positions). Ultrafast bootstrap values greater than or equal to 95 are shown.



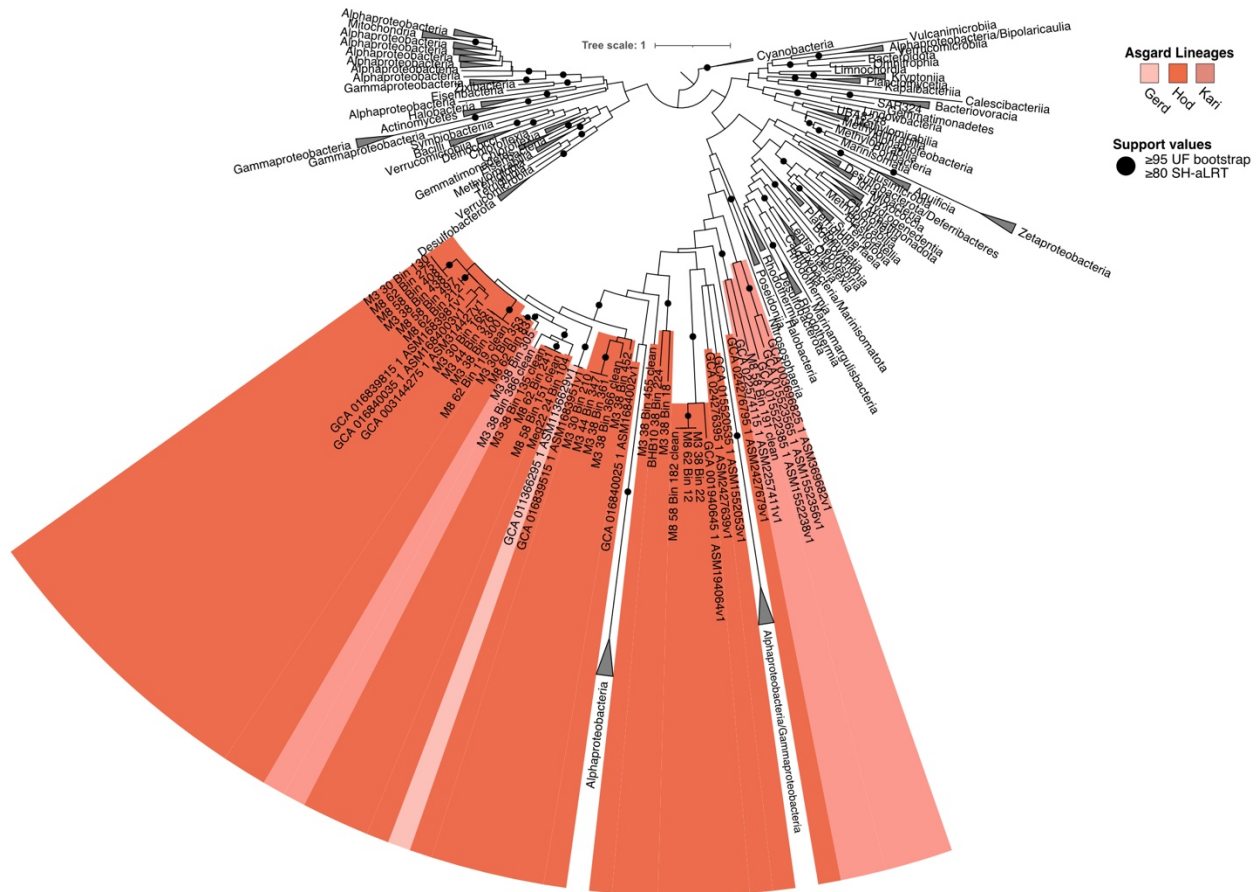
Supplementary Fig. 70 | **Maximum likelihood phylogeny of ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) family enzymes.** We identified 1,113 RuBisCO Asgardarchaeota sequences from Type II/III (blue), Archaeal Type III (orange), and Type IV (purple), which were aligned with 133 references⁶³. The branches are color-coded by Asgardarchaeota lineage. The sequences were aligned and trimmed with MAFFT v7.490 and trimAl v1.4.rev15. The tree was generated with IQ-TREE v2.0.7 (-bb 1000 -bnni), model LG+R10 was chosen according to Bayesian Information Criterion (BIC). The tree was inferred from a supermatrix of 1246 sequences and 417 amino acid positions (from 1105 initial positions). Ultrafast bootstrap values greater than or equal to 95 are shown.



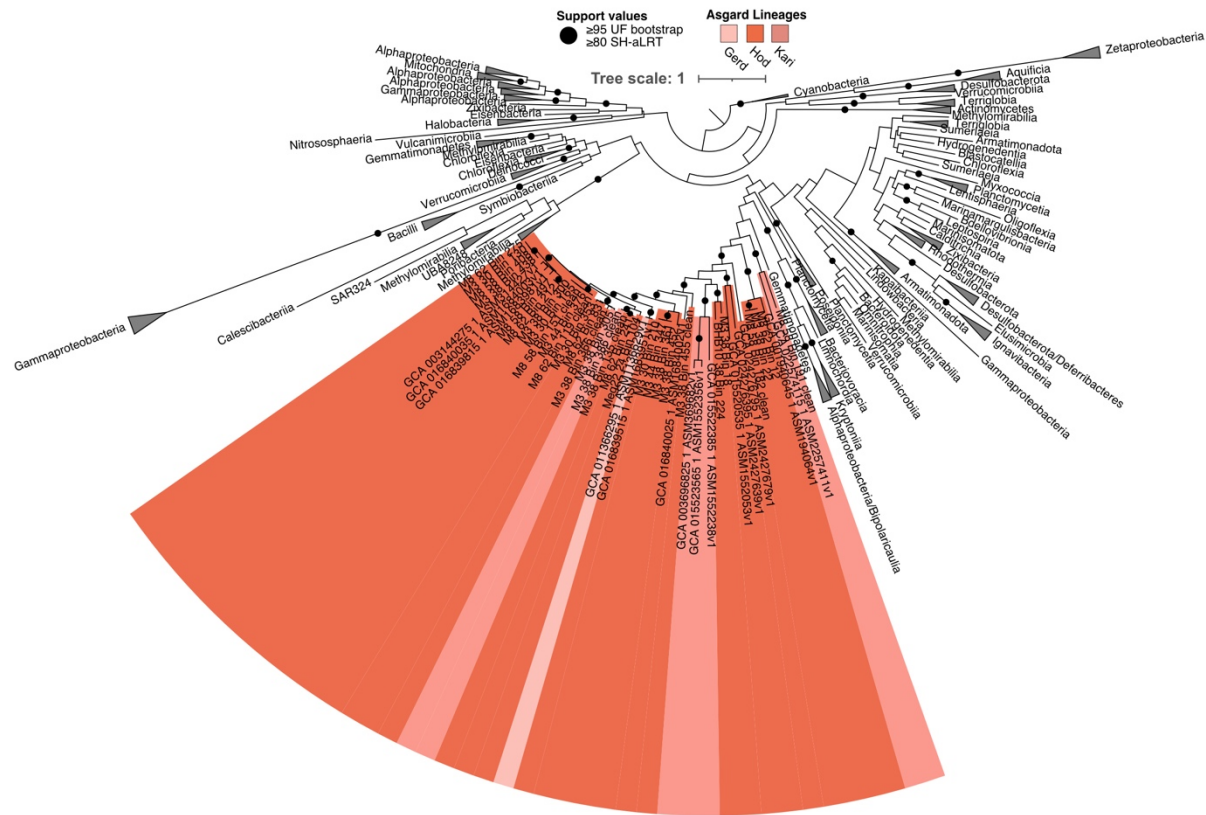
Supplementary Fig. 71 | **Maximum likelihood tree, displaying the Heme-Copper Oxidase (HCO) Superfamily, including oxygen and nitrate reducing families.** Within this phylogeny, we have 67 Asgardarchaeota sequences (colored branches), 23 Kariarchaeaceae (orange), 41 Hodarchaeales (red), 1 Gerdarchaeales (light blue), and 1 Lokiarchaeales (dark blue) MAGs. The phylogeny included 3,176 sequences with 269 amino acid positions (from 11518 initial positions) from all classified types of HCO (subset from https://github.com/ranjani-m/HCO/tree/main/Sequence_database/HCO-family-sequences_unique) and aligned with MAFFT v7.490, trimmed with BMGE v.2.0, and inferred with model LG+F+G with IQ-TREE v2.0.7 (-bb 1000 -bnni). Ultrafast bootstrap values greater than or equal 95 are shown. The HCO family is split into NO reductases and nitric oxide dismutase families (NOR and NOD), including cNOR, qNOR, gNOR, sNOR, eNOR and oxygen-reducing A-I-Type oxidases. There are several naming schemes for the HCO families. To avoid confusion here we provide the name in phylogeny and alternative names see Murali et al., 2022 for more detailed description of the discrepancies. Sequences listed in phylogeny with multiple classification names: A (A1, A2), D (B, B3), E (B, B4, SoxB), F (B, B5), G (B, B6), H (B2), I (B7), J (B8), K (B9), sNOR (B), eNOR (B), bNOR (B), and gNOR (B)¹⁴⁰.



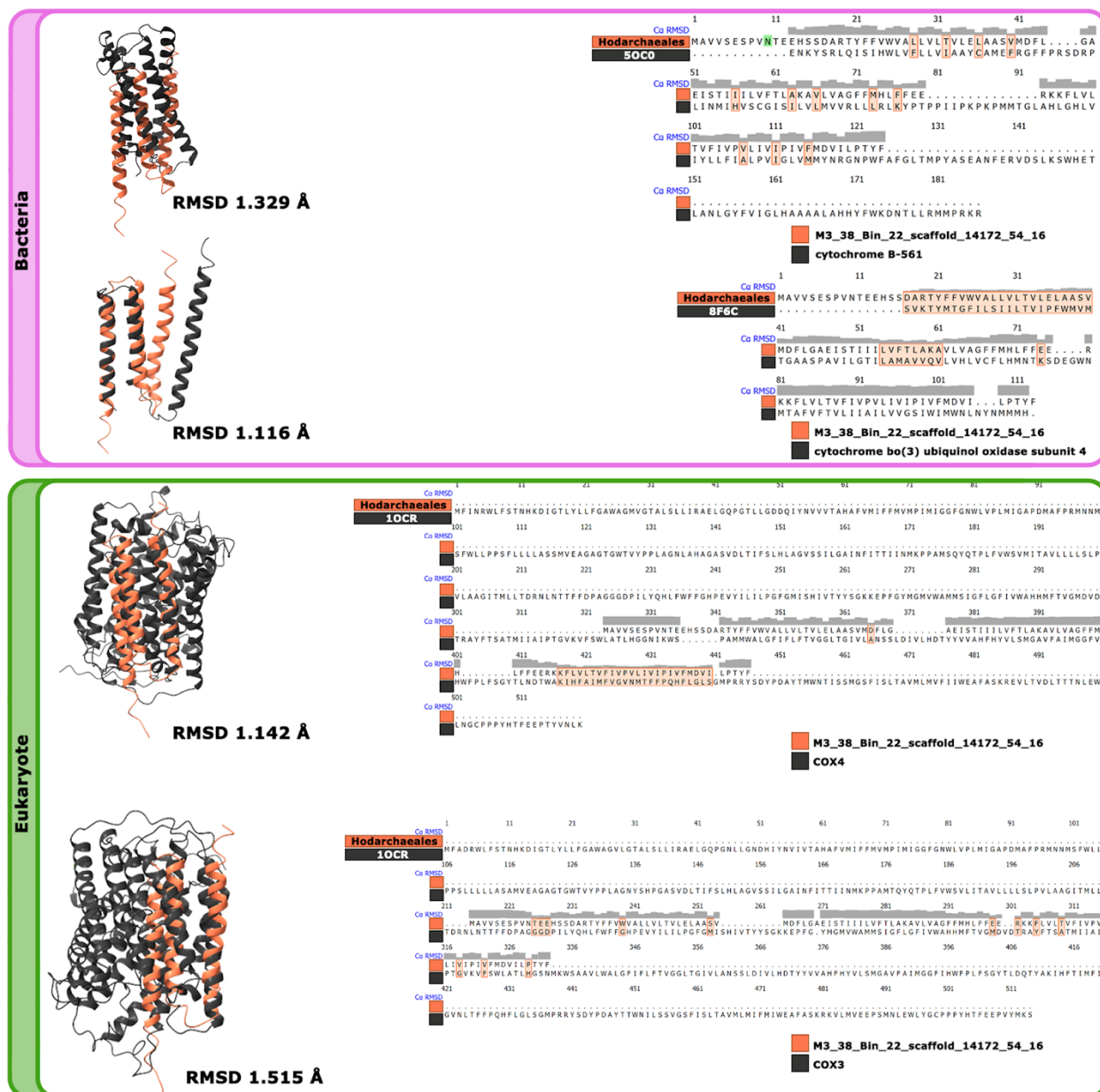
Supplementary Fig. 72 | **Cytochrome c oxidase subunit A (CoxA) phylogeny.** Maximum-likelihood phylogeny was inferred from a supermatrix of 407 sequences and 466 amino acid positions (from 1849 initial positions), displaying the diversity of CoxA subunits across Asgard archaea (varying shades of red for Heimdallarchaeia lineages), other archaeal lineages, bacteria, and mitochondria. The tree was generated using IQ-TREE v 2.0.7 (1000 ultrafast bootstrap replicates and calculation of 1000 SH-aLRT) with WAG+C50+R8 best-fit model chosen according to Bayesian Information Criterion (BIC). Support values are shown when ultrafast (UF) bootstrap support is greater than or equal to 95 and SH-aLRT support is greater than or equal to 80.



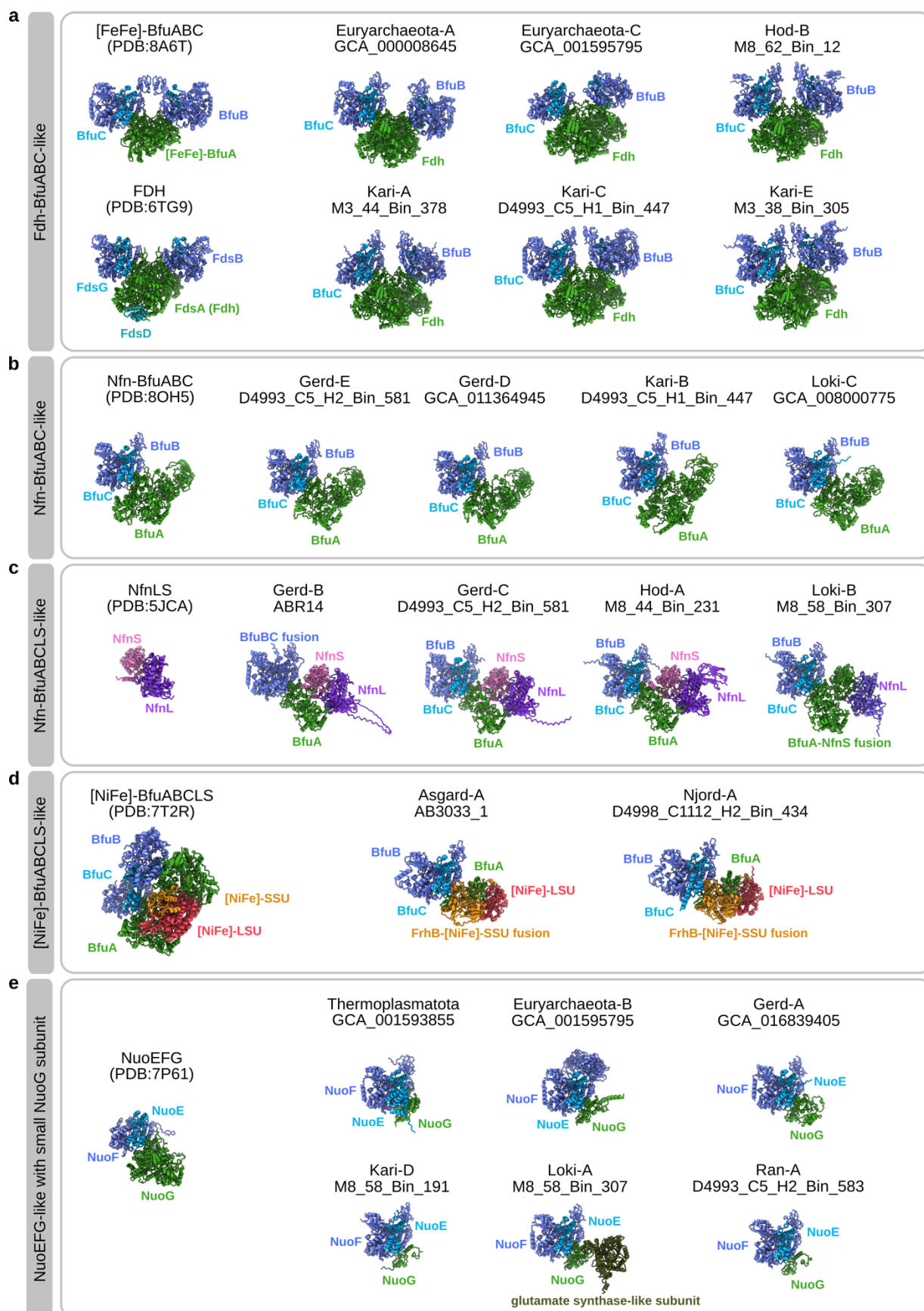
Supplementary Fig. 73 | **Cytochrome c oxidase subunit B (CoxB) phylogeny.** Maximum-likelihood phylogeny was inferred from a supermatrix of 407 sequences and 214 amino acid positions (from 1664 initial positions), displaying the diversity of CoxB subunits across Asgard archaea (varying shades of red for Heimdallarchaeia lineages), other archaeal lineages, bacteria, and mitochondria. The tree was generated using IQ-TREE v 2.0.7 (1000 ultrafast bootstrap replicates and calculation of 1000 SH-aLRT) with WAG+C60+R8 best-fit model chosen according to Bayesian Information Criterion (BIC). Support values are shown when ultrafast (UF) bootstrap support is greater than or equal to 95 and SH-aLRT support is greater than or equal to 80.



Supplementary Fig. 74 | **Cytochrome c oxidase subunit C (CoxC) phylogeny.** Maximum-likelihood phylogeny was inferred from a supermatrix of 407 sequences and 226 amino acid positions (from 1476 initial positions), displaying the diversity of CoxB subunits across Asgard archaea (varying shades of red for Heimdallarchaea lineages), other archaeal lineages, bacteria, and mitochondria. The tree was generated using IQ-TREE v 2.0.7 (1000 ultrafast bootstrap replicates and calculation of 1000 SH-aLRT) with WAG+C50+R8 best-fit model chosen according to Bayesian Information Criterion (BIC). Support values are shown when ultrafast (UF) bootstrap support is greater than or equal to 95 and SH-aLRT support is greater than or equal to 80.

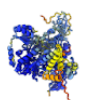


Supplementary Fig. 75 | **Asgardarchaeota cytochrome c oxidase subunit D (CoxD) structural homology with bacterial and eukaryotic monomers.** We used AlphaFold2 to predict Hodarchaeales CoxD from the MAG, M3_38_Bin_22 (shown in orange) and compared to bacterial and eukaryotic verified structures, boxed in pink and green, respectively. The root mean square deviation (RMSD) of the bacterial oxidase (PDB ID: 5O0C¹⁶⁷) and cytochrome bo₃ ubiquinol oxidase from *E. coli* (PDB ID: 8F6C¹⁶⁸) aligned with a Hodarchaeales CoxD is 1.329 and 1.116 Å, respectively. The eukaryotic Complex I (PDB ID: 1OCR¹⁶⁹) subunits 3 (Cox3) and 4 (Cox4) aligned with the Hodarchaeales protein have a RMSD of 1.142 and 1.515 Å, respectively. Lower RMSD indicates higher structural similarity. Alignment of the Hodarchaeales CoxD and cytochrome bo₃ ubiquinol oxidase. In bacteria, cytochrome bo₃ ubiquinol oxidase is a subunit of Complex IV in the electron transport chain, transferring electrons to reduce oxygen to water. To the right of each overlaid verified and predicted structures, we provide the structural alignment, highlighting the aligned portions.

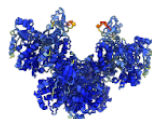


Supplementary Fig. 76 | **Overall structural architecture and shared subunit homology in NuoEFG-like complexes.** AF2 multimer predicted complexes containing NuoEFG subunits are divided into five broad groups based on structural similarity with known electron-bifurcating ABC (BfuABC) complexes in the PDB. The groups are the **a.** Formate dehydrogenase (Fdh)-BfuABC-like complexes, **b.** NADH-dependent reduced ferredoxin:NADP⁺ oxidoreductase (Nfn)-BfuABC-like complexes, **c.** Nfn-BfuABCLS-like complexes, **d.** [NiFe]-BfuABCLS-like complexes, and **e.** NuoEFG-like complexes with a small NuoG subunit. We compare the Asgardarchaeota and other archaeal predicted multimers to the following verified structures: *Thermoanaerobacter kivui* HydABC, also referred to as [FeFe]-BfuABC (PDB ID: 8A6T¹¹²), *Rhodobacter capsulatus* FDH (PDB ID: 6TG9¹¹¹), *Sporomusa ovata* StnABC, also referred to as Nfn-BfuABC¹¹⁰, (PDB ID: 8OH5¹¹⁵), *Pyrococcus furiosus* NfnLS (PDB ID: 5JCA¹⁷⁰), *Acetomicrobium mobile* [NiFe]-HydABCLS, also referred to as [NiFe]-BfuABCLS (PDB ID: 7T2R¹¹⁹), and *E. coli* NuoEFG (PDB ID: 7P61⁹⁴). Homologous subunits share the same coloring. Acronyms for the complexes, removing the taxonomic suffix and adding letter with the name below for reference throughout the manuscript (i.e., Kari-A is from M3_44_Bin_378, a Kariarchaeaceae MAG).

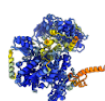
Thermoplasmatota
GCA_000008645
ipTM: 0.903
pTM: 0.894



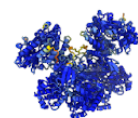
Euryarchaeota-A
GCA_000008645
ipTM: 0.901
pTM: 0.911



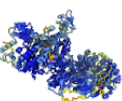
Euryarchaeota-B
GCA_000008645
ipTM: 0.849
pTM: 0.871



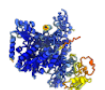
Euryarchaeota-C
GCA_000008645
ipTM: 0.916
pTM: 0.924



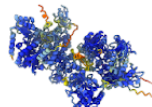
Asgard-A
AB3033_1
ipTM: 0.882
pTM: 0.892



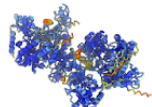
Gerd-A
GCA_016839405
ipTM: 0.897
pTM: 0.906



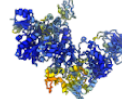
Gerd-B
ABR14
ipTM: 0.831
pTM: 0.858



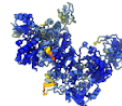
Gerd-C
D4993_C5_H2_Bin_581
ipTM: 0.858
pTM: 0.872



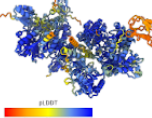
Gerd-D
GCA_011364945
ipTM: 0.911
pTM: 0.895



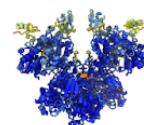
Gerd-E
D4993_C5_H2_Bin_581
ipTM: 0.903
pTM: 0.913



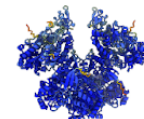
Hod-A
M8_44_Bin_231
ipTM: 0.842
pTM: 0.856



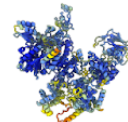
Hod-B
M8_62_Bin_12
ipTM: 0.903
pTM: 0.914



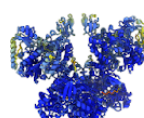
Kari-A
M3_44_Bin_378
ipTM: 0.910
pTM: 0.905



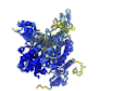
Kari-B
D4993_C5_H1_Bin_447
ipTM: 0.900
pTM: 0.874



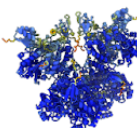
Kari-C
D4993_C5_H1_Bin_447
ipTM: 0.899
pTM: 0.911



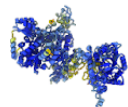
Kari-D
M8_58_Bin_191
ipTM: 0.883
pTM: 0.902



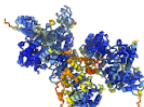
Kari-E
M3_38_Bin_305
ipTM: 0.905
pTM: 0.914



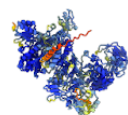
Loki-A
M8_58_Bin_307
ipTM: 0.820
pTM: 0.849



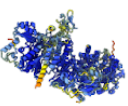
Loki-B
M8_58_Bin_307
ipTM: 0.809
pTM: 0.819



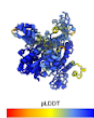
Loki-C
GCA_008000775
ipTM: 0.897
pTM: 0.855



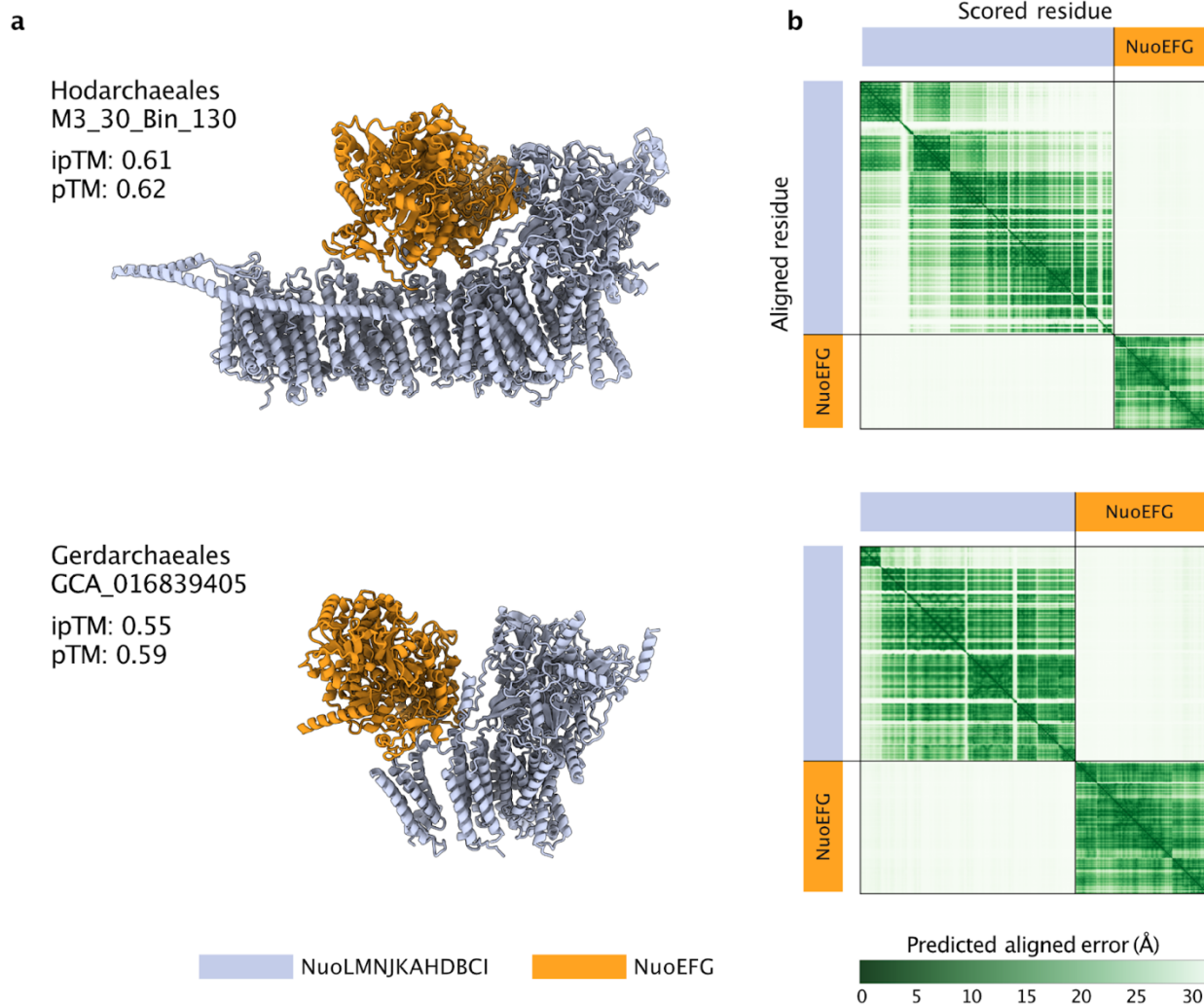
Njord-A
D4998_C1112_H2_Bin_434
ipTM: 0.906
pTM: 0.914



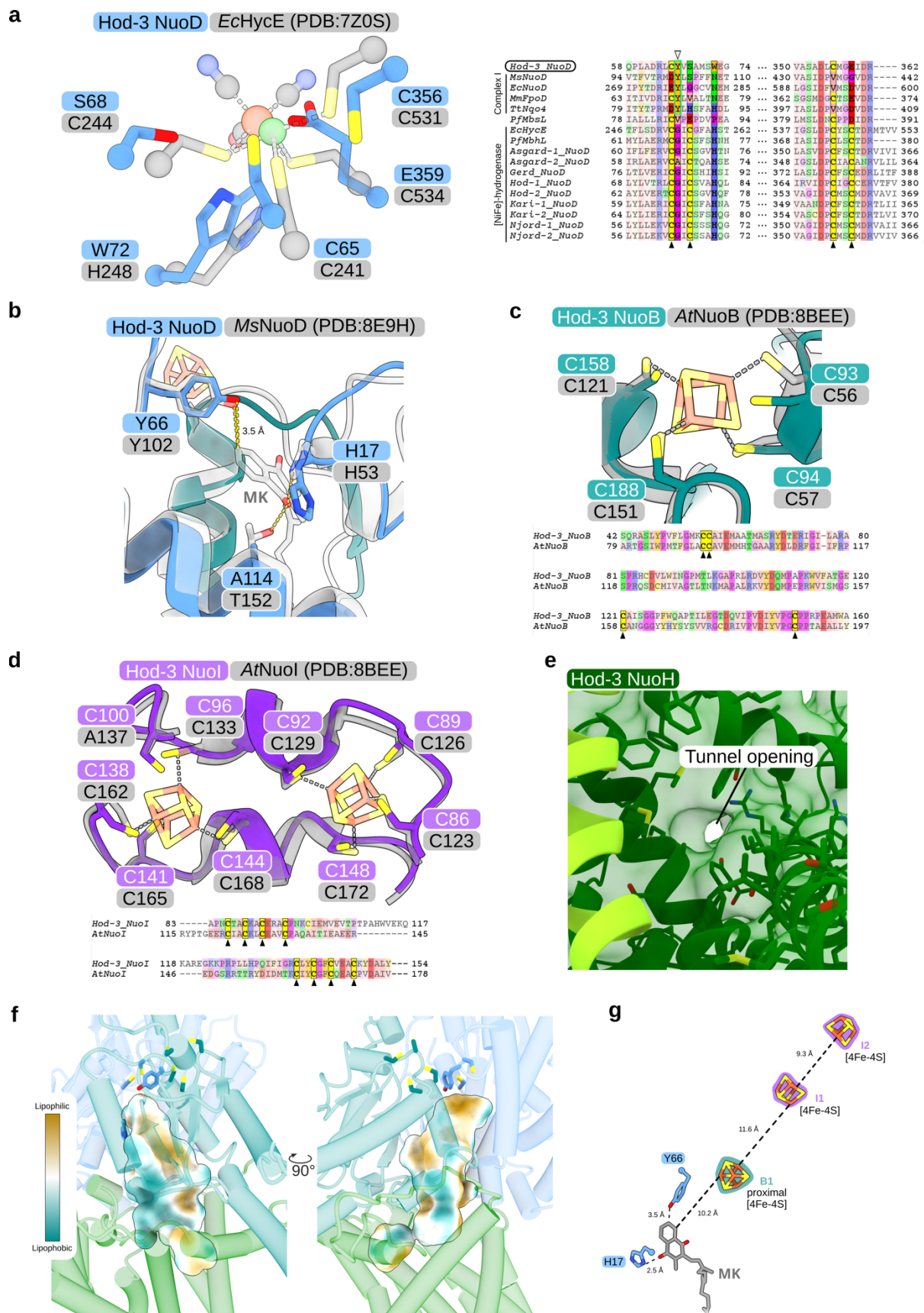
Ran-A
D4993_C5_H2_Bin_583
ipTM: 0.875
pTM: 0.909



Supplementary Fig. 77 | **AlphaFold2 multimer confidence scores of NuoEFG-like complexes.** Top-ranked multimers of the 18 Asgardarchaeota and 4 archaeal NuoEFG-like complexes. These complexes are split into families (see Supplementary Fig. 76). Protein models are colored by predicted local distance difference test (pLDDT) scores. Subunit identities within the Predicted aligned error (PAE) plots are denoted by colored boxes. PAE plots generated by PAE Viewer webserver¹⁶¹. Acronyms for the complexes, removing the taxonomic suffix and adding letter with the name below for reference throughout the manuscript (i.e., Asgard-A is from AB3033_1, an Asgardarchaeia MAG). Other abbreviations include: pLDDT (predicted local distance difference test), ipTM (interface predicted template modeling score), and pTM (predicted template modeling score).



Supplementary Fig. 78 | **No predicted protein-protein interactions between Complex I-like complexes and NuoEFG subunits.** **a.** Two Asgardarchaeota Complex I-like AlphaFold 3 (AF3) models with the NuoEFG subunits in yellow. Genome taxonomy and MAG name are shown next to each predicted structure. AF3 places the NuoEFG subunits at a large distance away from the NuoDBCI subunits and predicts low ipTM (interface predicted template modeling score) and pTM (predicted template modeling score) scores, suggesting no protein-protein interaction occurs between NuoEFG and other Nuo subunits. **b.** Predicted Aligned Error (PAE) plots displaying the predicted alignment errors between the residues of the NuoEFG subunits (labeled in yellow on the axes) and the rest of the Nuo subunits (labeled in grey). High PAE values (lighter green) between the NuoEFG and Complex I-like complexes suggest poor structural alignment, predicting that there is no protein-protein interaction between these two groups of proteins within the same genome. PAE plots generated by PAE Viewer webserver¹⁶¹.



Supplementary Fig. 79 | **AlphaFold2 multimer model of Hodarchaeales Fd-dependent Complex I.** The Complex I-like “Hod-3” (M3_30_Bin_130) model is not predicted to be a [NiFe]-hydrogenase, so the structural analysis of this complex is presented separately here. **a.** Left panel: structural superposition of the Hod-3 NuoD subunit with the [NiFe] coordination site of *E. coli* HycE (PDB ID: 7Z0S²⁴) indicates that [NiFe] cofactor binding is not plausible in Hod-3 NuoD. Notably, the Hod-3 NuoD lacks two of the four cysteine residues necessary for coordinating the [NiFe] cofactor. **a.** Right panel: multiple sequence alignment (MSA) of Hod-3 NuoD with other NuoD sequences from Complex I's or membrane-bound [NiFe]-hydrogenase complexes. Four cysteines are conserved in the membrane-bound [NiFe]-hydrogenase complexes (positions marked with black arrows), whereas Complex I's conserve a key tyrosine residue important for proton transfer to quinone (Tyr66 position in Hod-3; positions marked with both a gray arrow and green box). **b.** Structural superposition of Hod-3 NuoD with *M. smegmatis* NuoD (PDB ID: 8E9H³¹) indicates that menaquinone could plausibly be modeled into the Hod-3 complex. Hod-3 NuoD conserves key tyrosine (Tyr66) and histidine (His17) residues for transferring protons to reduced quinone¹⁰³, further indicating that the Hod-3 complex likely binds to and reduces quinone. **c.** Structural modeling of a FeS cluster into Hod-3 NuoB via structural superposition with *A. thaliana* NuoB (PDB ID: 8BEE²⁷). Sequence alignment between the two subunits show that the four cysteines required to coordinate a FeS cluster are conserved (positions marked with black arrows). **d.** Structural modeling of two FeS clusters into Hod-3 NuoI via structural superposition with *A. thaliana* NuoI (PDB ID: 8BEE²⁷). Sequence alignment between the two subunits show that all eight cysteines required to coordinate the two FeS clusters are conserved (positions marked with black arrows). **e.** Surface representation of Hod-3 NuoH reveals an opening that may be a site for quinone entry into the complex from the lipid membrane. **f.** Surface representation of a large cavity embedded within the Hod-3 complex peripheral arm (subunits NuoHBD). Such a cavity in the Hod-3 complex may be able to accommodate a quinone. The key tyrosine, histidine, and cysteine residues from panels **b** and **c** are also shown. Cavity surface was calculated with a probe radius of 1.8 Å and colored by lipophilicity with UCSF ChimeraX v1.8¹⁷¹. **g.** Overall model of the electron transfer path in Hod-3. Electrons donated by Fd_{red} to the “I2” FeS cluster in NuoI are transferred down to a quinone, which gets reduced. Edge-to-edge distances between adjacent FeS clusters and quinone are labeled in ångström (Å) and are all within electron transfer distance³⁰. Atomic distances calculated in UCSF ChimeraX. Sequence alignments were made with Clustal Omega¹⁶² and visualized in JalView v2.11¹⁷². Abbreviations: MK, menaquinone.

10 20 30 40 50 60 70 80 90

EcHycE 185 ELGDKKNVVPIDPLHVTSEPGHFRLLFVDCENIIDADYRLFYVHRGMEKLAETRMGYN-EVTFLSDRVQCI-CGFAHSTAYTTSVE---NAMGI--QV 276

PfMbhL 1 MKKVEYVWYKIPFPIHPIHGLLEPEKFIITLDGERIVNVVDVLKGYNLRGVQWIGMRR-NYV-QIMYLAERMCI-CFSFHNTYVRAVE---EMAGI--EV 91

Hod-1_NuoD 8 PIPETSDYAISVSGPOHPTHEPEVRFIFQVQVETQVDVLDRIQGNHRIKAFENR-TWL-KNLYLVTRLQCI-CSVAHQALAYVHAARKCM-IIDQ--SV 100

Kari-1_NuoD 1 --MNEQETSIVSGPOHPTHEPEVRFIFQVQVETQVDVLDRIQGNHRIKAFENR-DYN-KTLYLAERICI-CFAMNANPARAVE---FVADV--KV 89

Kari-2_NuoD 4 LSEKIQGKEINISGPOHPTHEPEVRFIFQVQVETQVDVLDRIQGNHRIKAFENR-NSD-QNLYLIERICI-CFSHMGNYAKGVF---EAGI--EA 94

Gerd_NuoD 16 AACMPRTMIDPLHGPALKEPFTFLLEVEGIVVATRIQGVHRIKLAETK-TYS-ANLTLVERICI-CSHHISLANARFETALAESEF--EV 109

Njrd-1_NuoD 1 ----MSVKIPIDGPIHPIHGLLEPEKFIITLDGERIVNVVDVLKGYNLRGVQWIGMRR-NYV-QIMYLAERMCI-CFSFHNTYVRAVE---EMAGI--EV 91

Njrd-2_NuoD 1 ----MSVKLPIDGPIHPIHGLLEPEKFIITLDGERIVNVVDVLKGYNLRGVQWIGMRR-TWF-QNLYLLEKVCIC-CSSAHGCCYAQGA---KIEK--EF 86

Hod-2_NuoD 2 EKSSSTKFGLPVGPPIHPIHGLLEPEKFIITLDGERIVNVVDVLKGYNLRGVQWIGMRR-TYL-QTAYLVERTICI-CSGVHQTYYVSCVE---QLGDI--KI 92

Asgard-1_NuoD 1 -MSKNHSFPIVGPPIHPIHGLLEPEKFIITLDGERIVNVVDVLKGYNLRGVQWIGMRR-TYW-RDIFLFEVRCIC-CSGVHNTFPCQVTE---ELVGI--EI 90

Asgard-2_NuoD 1 ---MPRETIIPFPGYHYALFEPVFMKLSVLGKIVDVLGHAHRIEYLMKK-RFE-DCIRLAERVCAI-CTOAHSECFVMGIF---RIAGI--EP 88

PfMbsL 1 YPIDKDTYELFFGPOHPTHEPEVRFIFQVQVETQVDVLDRIQGNHRIKAFENR-PWY-TNIALLLRICVP-PDVPEBAIYSMAVD---EIGW--EV 108

HsNuoCD 184 AADSDTFMFLNLGPHNPATGVLHVEITLDGERIVNVVDVLKGYNLRGVQWIGMRR-TYRQIMPYPDRWYFPGGILNEWAYARTAS---DLADI--DV 276

EcNuoD 209 GTENEDFMFLNLGPHNPATGVLHVEITLDGERIVNVVDVLKGYNLRGVQWIGMRR-SWH-SYIPYTDRIKYL-GGVNEMPVVLAVE---KLGI--TV 299

Hod-3_NuoD 1 ---MSEQMYIHMGPOHNPATGVLHVEITLDGERIVNVVDVLKGYNLRGVQWIGMRR-PYD-TFQPLADRLCYV-SAMSEGTYYMAIE---KLLGM--EI 88

AtNuoD 4 RKRQIKNTLNFPGPOHNPATGVLHVEITLDGERIVNVVDVLKGYNLRGVQWIGMRR-TYL-QALPYFDRLLCYV-SAMSEGTYYMAIE---KLLNC--EV 94

YLNDFUS2 76 ANTRIRAFITNFPGPOHNPATGVLHVEITLDGERIVNVVDVLKGYNLRGVQWIGMRR-TYM-QALPYFDRLLCYV-SAMSEGTYYMAIE---KLLNV--EV 166

MmFpoD 3 EMLESNEMINLHGPPOHNPATGVLHVEITLDGERIVNVVDVLKGYNLRGVQWIGMRR-TYL-QGITIVDRICVL-MALNBECCYVGCVE---KLLDI--EP 93

TtNqo4 19 KELRETVMTLHGPPOHNPATGVLHVEITLDGERIVNVVDVLKGYNLRGVQWIGMRR-TYL-QNITYTPRMDL-HSFADLAYALAVE---KLLGA--VV 109

MsNuoD 34 AAQAGERIVVMHGPPOHNPATGVLHVEITLDGERIVNVVDVLKGYNLRGVQWIGMRR-NWT-QGVTVTRMDL-HSFADLAYALAVE---KLLGITDDI 126

a b a b

100 110 120 130 140 150 160 170 180 190

EcHycE 277 PERAEYIRAILVERLHLLNLGLACHFG--FDSGFMQFVRVETSMKMAHILTGARKTYGLNLIGCIRDLK---DDMIQITROLAQMRREVEQ 370

PfMbhL 2 PERAEYIRVIVGLERHLLNLGLGVGHGIG--YDVLVHLTWLAERVMVDLAVSGNVRNYSMTYGGVRRDICE---KQKRLILDMIKYREVLQP 185

Hod-1_NuoD 101 PERAEYIRVIVGLERHLLNLGLGVGHGIG--YDVLVHLTWLAERVMVDLAVSGNVRNYSMTYGGVRRDICE---KQKRLILDMIKYREVLQP 185

Kari-1_NuoD 90 PKRAELFVLVLAERHLLNLGLGVGHGIG--FDTFYLLVNDRTIMDLLETITGNVRNYSMTYGGVRRDICE---EIKKIMRDKIEFKPRLEK 183

Gerd_NuoD 95 PRANYIRVIVLAERHLLNLGLGVGHGIG--FDTFYLLVNDRTIMDLLETITGNVRNYSMTYGGVRRDICE---MMIKQLFSLPLILQRIK 188

Njrd-1_NuoD 110 PKRAQYIRVIVLAERHLLNLGLGVGHGIG--FDTFYLLVNDRTIMDLLETITGNVRNYSMTYGGVRRDICE---AQMDIMKYMDLTERVRR 203

Njrd-2_NuoD 87 PEKALYIRVIVLAERHLLNLGLGVGHGIG--FDTFYLLVNDRTIMDLLETITGNVRNYSMTYGGVRRDICE---RMLSKALDVANFIAERAF 180

Hod-2_NuoD 93 PRARHYIRVIVLAERHLLNLGLGVGHGIG--FDTFYLLVNDRTIMDLLETITGNVRNYSMTYGGVRRDICE---EIDIRKALDEVESEIR 186

Asgard-1_NuoD 91 PERAKYIRVIVLAERHLLNLGLGVGHGIG--FDTFYLLVNDRTIMDLLETITGNVRNYSMTYGGVRRDICE---ALARGINKLDSLEKALK 184

Asgard-2_NuoD 89 PKRAKYIRVIVLAERHLLNLGLGVGHGIG--FDTFYLLVNDRTIMDLLETITGNVRNYSMTYGGVRRDICE---KQVSRVLSKLEKLEESIR 182

PfMbsL 109 PERAQYIRVIVLAERHLLNLGLGVGHGIG--FDTFYLLVNDRTIMDLLETITGNVRNYSMTYGGVRRDICE---DKWLKQVDRVTEYLKDKRL 203

HsNuoCD 277 PEYAQILRMGGELCRICMMLAGTFCDDIGDPAIFPMYAMRDRVNVQNLLEDLTQGMMLNLRGLGGVVMVPEPRAEFFESVDFDLDELPERLE 375

EcNuoD 300 PRDYNVIRVILSLRTRHLLNLGLGVGHGIG--AMVRFPAFTDQKYLVDLAEITGFMPMFAIRIGVADLPR---GWRLLREFLDHMKPRLAS 393

Hod-3_NuoD 89 PKRAQYIRVIVLAERHLLNLGLGVGHGIG--FDTFYLLVNDRTIMDLLETITGNVRNYSMTYGGVRRDICE---GFIDARMLETYYLDIRVAE 182

AtNuoD 95 PERAKYIRVIVLAERHLLNLGLGVGHGIG--FDTFYLLVNDRTIMDLLETITGNVRNYSMTYGGVRRDICE---GLCRDIDSPQFASRIE 188

YLNDFUS2 167 PERAKYIRVIVLAERHLLNLGLGVGHGIG--FDTFYLLVNDRTIMDLLETITGNVRNYSMTYGGVRRDICE---GLLDDIYMWATQGDRLD 260

MmFpoD 94 PERHYIRVIVLAERHLLNLGLGVGHGIG--FDTFYLLVNDRTIMDLLETITGNVRNYSMTYGGVRRDICE---GFKERTIPVNLKLEKRLD 187

TtNqo4 110 PERAEYIRVIVLAERHLLNLGLGVGHGIG--FDTFYLLVNDRTIMDLLETITGNVRNYSMTYGGVRRDICE---EFVPELKKLEVLPHRIDE 203

MsNuoD 127 PERASVIRVIMMLRNLRLHLLNLGLGVGHGIG--AMSAMFYGRERREILRVFESITGLMMHAYIRPGGLAADLP---DAITQVRLVELPKRLK 220

c

200 210 220 230 240 250 260 270 280 290

EcHycE 371 LVDVLSTPMEORTVIGRLDPEIARDFSGNCPMVRASGHARVTRADHPFVGYGLL--PMEVH-----SEQG-CVIRSLKVRIN--VYALNMDY 459

PfMbhL 186 IDYVLDHSTIEARLRDVAVPKKAIEMGAVGPTARGSGIKESRWSGLGVYDPL--GKIPVTPEDVTGEKAR-GVYDRMAVRIGLWMSLDLLEH 281

Hod-1_NuoD 195 HRAYMKERSFPIYSGKAILSKDAKAYCAVGPTRVARGSGIKESRWSGLGVYDPL--GKIPVTPEDVTGEKAR-GVYDRMAVRIGLWMSLDLLEH 281

Kari-1_NuoD 184 IRDVLVNSKTIARTNGVSTISKDLARKTCVGPTRVARGSGIKESRWSGLGVYDPL--GKIPVTPEDVTGEKAR-GVYDRMAVRIGLWMSLDLLEH 281

Kari-2_NuoD 189 IKKAIYNDKTIARNGVSTISKDLARKTCVGPTRVARGSGIKESRWSGLGVYDPL--GKIPVTPEDVTGEKAR-GVYDRMAVRIGLWMSLDLLEH 281

Gerd_NuoD 204 YDIALKEDTIRLRTQIGILLPSKAKELCAVGPTRVARGSGIKESRWSGLGVYDPL--GKIPVTPEDVTGEKAR-GVYDRMAVRIGLWMSLDLLEH 281

Njrd-1_NuoD 181 VERVULRRTILDRCGIGLLSKVDARILRTVARGSGIKESRWSGLGVYDPL--GKIPVTPEDVTGEKAR-GVYDRMAVRIGLWMSLDLLEH 281

Njrd-2_NuoD 175 YDIALKEDTIRLRTQIGILLPSKAKELCAVGPTRVARGSGIKESRWSGLGVYDPL--GKIPVTPEDVTGEKAR-GVYDRMAVRIGLWMSLDLLEH 281

Hod-2_NuoD 187 YKTLILQDLSSKRTENVGFTPKRTAMBLGAVGPTRVARGSGIKESRWSGLGVYDPL--GKIPVTPEDVTGEKAR-GVYDRMAVRIGLWMSLDLLEH 281

Asgard-1_NuoD 185 LQELSTDPILFARVKEIILDKNTAKKYCVGPTRVARGSGIKESRWSGLGVYDPL--GKIPVTPEDVTGEKAR-GVYDRMAVRIGLWMSLDLLEH 281

Asgard-2_NuoD 183 PREIFERDKTIRMAKKEVYVSKQDALKLVGPVARGSGIKESRWSGLGVYDPL--GKIPVTPEDVTGEKAR-GVYDRMAVRIGLWMSLDLLEH 281

PfMbsL 204 FDNVLFENYITKRLGEGIVMDKKFALLEGVTCGNLRATGVAYTVRKSDEYLYPEL--DFEIP-----VLKE-GDALARVLRVRLLEODYQYIE 292

HsNuoCD 376 YSDILSTNEIFQARTVNTGLDYETAKDYGVTGPVARGSGIKESRWSGLGVYDPL--GKIPVTPEDVTGEKAR-GVYDRMAVRIGLWMSLDLLEH 281

EcNuoD 394 YEKALQNTILKSGSGVAAAGAKALEWGTGAGLRATGIDFVTRKARPYSGYENF--DFEIP-----VGGGVSCYTRVLMKVESLRQSLRIE 483

Hod-3_NuoD 183 YSAMLDESKVFMRLSKVNLISKENALNLGATGPVLRASGTQGRKVEPYLYDPL--EFEVP-----VGTG-GDALARVLRVRLLEODYQYIE 292

AtNuoD 189 LDEMLTGNRIWKRLVDITGTVAQAKDWGFSVCMRLGSGVCMVLRRAAPDYVDPL--DFDVP-----VGTG-GDALARVLRVRLLEODYQYIE 292

YLNDFUS2 261 IEELLTDNRILKRLVNIITGTVAQAKNLGLSGMPLRGSGIPFTRKNAFYDAYDKV--DFDVP-----VGMN-GCYDRYLIRMAFQOSLRIRIE 377

MmFpoD 188 YEEMFYSITRYERTIGIGVLTADKASLGSVGPVLRATGVPFTRKNEPYLYYRDL--DFKVC-----TETA-GDALARVLRVRLLEODYQYIE 292

TtNqo4 204 YALFESPIFYENARGVPEPVAIDLGLGGLSRLASGVNYVTRKAPYSGYETY--TFDVP-----LGER-GVDFRMLVIRLMRESYKIIQ 329

MsNuoD 221 LEDLLNENYTKARTGVGVYDLDLTCMALGIGPILRSTGLPHLRLKAQPCYGYENY--EFDVI-----TDDR-CYSGRYIRVRLMHESVKVIRIE 309

c

300 310 320 330 340 350 360 370 380 390

EcHycE 460 GLDNLPG-GPIWVEG-----FTYIPHRFALGFALFNGDDIHWSMT--CDNQKLYVWRCRA 513

PfMbhL 282 ALDQMPG-EKIKTFPKDN-----ILVAKLLLDGEGIGRYGAGELVHYVGRQKGRDGPVWKKPRF 344

Hod-1_NuoD 284 ALDAMPA-GDILPWRN-----VPRRPETSEGIQVRAFRGEDIHYIRS--NGTKDPDRHKIRAP 340

Kari-1_NuoD 273 LDILIPS-GKIPQDV-----RGRFEGKAVSRISAPRGENFHYVEF--RKSKNPYRLKVRAP 325

Kari-2_NuoD 278 ALNMLNP-GPILSKI-----RGKFNGATISRLAPRGEDHLYIDF--RNDKIPYRLKVRAP 330

Gerd_NuoD 295 MENLPK-GPIFAK-----PRKVEAQILSRITAPRGELIHYLWS--NDTNKPYRLKVRAP 326

Njrd-1_NuoD 270 IFVNMKPNKPIAAKITH-----FTKPPENAISSVAPRGELFYIYS--NGSLKPERVKIRTP 326

Njrd-2_NuoD 270 IFVNMKPNKPIAAKITH-----FTKPPENAISSVAPRGELFYIYS--NGSLKPERVKIRTP 326

Hod-2_NuoD 276 LDIDPE-GPIETKE-----KRPFDWVSFVSAPRGELIYSIRG--NGTDHPERVKIRTP 329

Asgard-1_NuoD 274 ALKDLPE-GPISVEI-----EFPFEGIAISRCAAPRGELFYIKS--NGTNIPERVKIRTP 326

Asgard-2_NuoD 272 ALENLPK-GDITKEY-----EKIPAGEVTSRVAPRGELLYIKT--TGSITPERVKIRTP 324

PfMbsL 293 LDDMGPPSGPYKQDPKLNLP-----RKVPPEGAFAHVRAKQDFGAYVVS--DGGHKPYRVHIRGP 354

HsNuoCD 465 CVDLLE-GWPEDEBDIQANVPRTL-----KPEAGTEYTRTVGAGRGELGVYIRA--DGTEKPGRFKIRSP 526

EcNuoD 484 CNLNMPP-EGPFKADHPITPPP-----KERTLQHIETLITHLQVSGVPMVPAESFCMIAKNGINSYLYTS--DGSMTSYRTRIRTP 564

Hod-3_NuoD 272 ALDNLPL-ABEGINR-----APRIIPKGEVYFRDPRNGEAGYIYIG--NEDKIPYRVKIRTP 326

AtNuoD 278 CNLNMPP-SOMKADKDRKCPPS-----RCRMKLSMESLIHHFLYTGFSVPASSTYTAFAKRGEGVLYS--NGSNRPYRVKIRAP 358

YLNDFUS2 350 CNDMP-AGAVKVEDFKINSPP-----RNLMKDEMEALIHHPFLYTKGVSPPGTYTALFAKRGEMGVYVS--DGSEPRYRVKIRAP 358

MmFpoD 277 CDDMIP-NGPIPEGTP-----YGRKTPVMRVPAVEFHRVDRGEMGVYVS--DGSDRPYRVKIRAP 338

TtNqo4 293 ALERLE-PGPVDRPNPQITPPP-----RHLLETSEMAVYHFKHYTGEFHPKGEVYVPTSARGELGYIYS--DGSMPYRVKIRAP 373

MsNuoD 310 CLARKL-PGPVMSIDKLLAWPADLKLGPDLGNSPEHIAKIMGRSMEGLIHHFKLVTEGIRVPPGVYVAVSEARGELGHVMS--DGGTRPYRVKIRAP 406

400 410 420 430

EcHycE 514 TYANW-PTLRYMLRGNTVSLAPLIGSLDPCYSCTDR 549

PfMbhL 345 TFPNL-FTIAKALEGNELALVVAIASIDPLSCTDR 380

Hod-1_NuoD 341 TFPNL-PSLVHRLKGVQVAPPVIRVIPCIGCCER 376

Kari-1_NuoD 326 TYANV-VSSVEALVGHIAIPVIVAANDPCFSCCTDR 361

Kari-2_NuoD 331 TYANL-VSSMNVILVGHIAIPVIVASCPFCFSCCTDR 366

Gerd_NuoD 349 TLANI-LSVAEMLKGYIAISTISLASLPCFSCCTDR 384

Njrd-1_NuoD 327 TLANI-VSVIHMVKGHYIAIPPIVAGILFCMCMCMTDR 362

Njrd-2_NuoD 327 TLANI-ASVVIHMVKGHYIAIPPIVAGILFCMCMCMTDR 362

Hod-2_NuoD 330 TLANW-QSIAEMLRGNTVSLAPLIGSLDPCYSCTDR 365

Asgard-1_NuoD 327 SYMNI-PSVKPMLIDETLAVPIILASVPCFSCCTDR 362

Asgard-2_NuoD 325 TILNL-IAAKPILVGGTIAAPVSIISLPCICACANR 360

PfMbsL 315 SIAGHVRVLEQLVIGARLAVPAILMSLNCNPPHIDR 391

HsNuoCD 527 CFSNL-SALKEMSLEYIPDLVATLGLSLVILGVDR 562

EcNuoD 565 SYAHL-QQIPAAIRGSLVSLIVLYLSIFVMSVDR 600

Hod-3_NuoD 327 AYSNL-CLTSHLITGKVAIPPIVASANDLCMCMCMTDR 362

AtNuoD 359 GFAHS-QGLDFMSKHHMLAVVTIGTQIVMFGVDR 394

YLNDFUS2 431 GFAHL-GAFDHIARGHFLDPAVAIIGTQIVMFGVDR 466

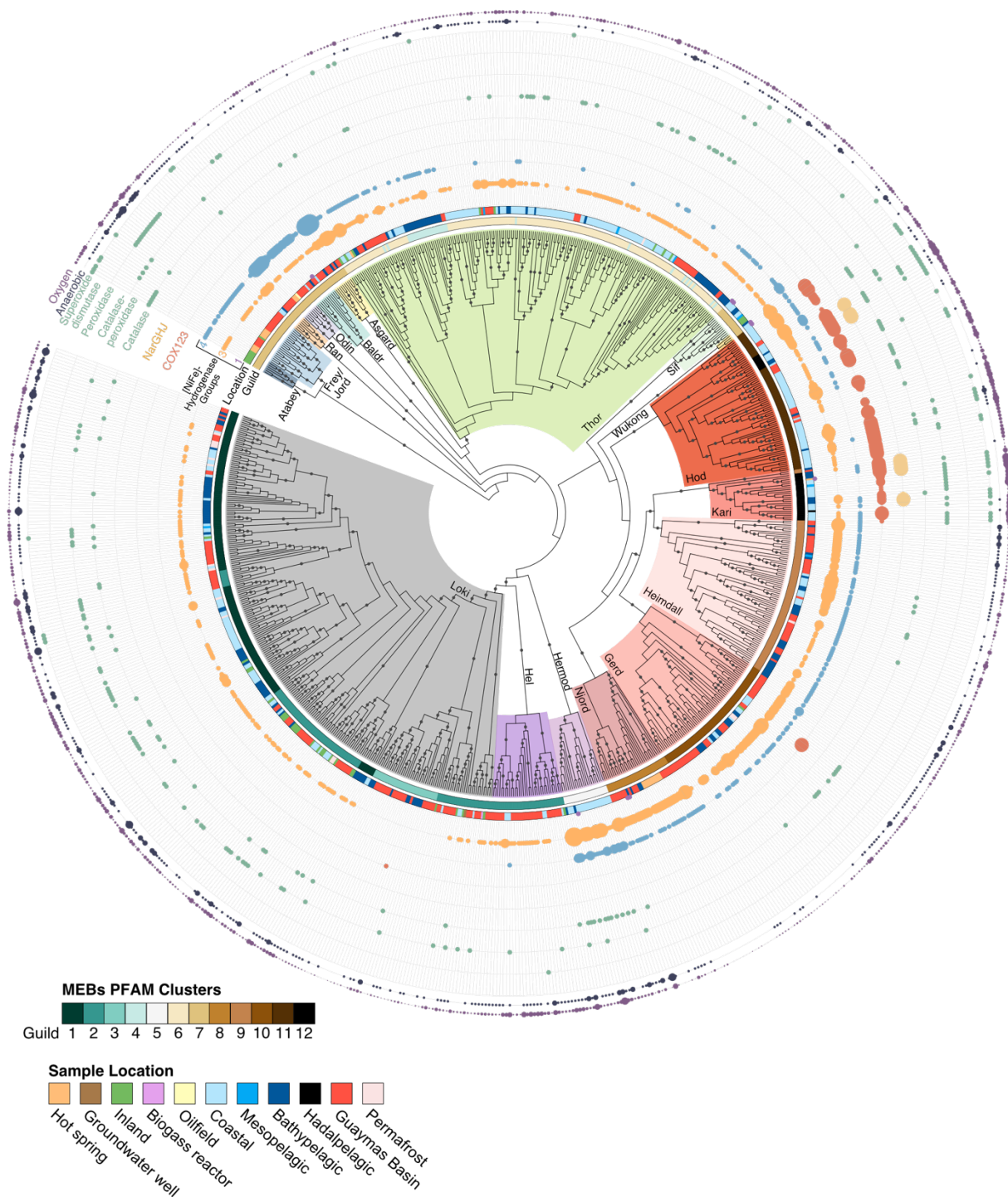
MmFpoD 339 YPTL-QALPFLITGTVAIMVSIQSGMCTSCVDR 374

TtNqo4 374 SFVNL-QSLPYACKGEQVPMVAIASLPMVMGVDR 409

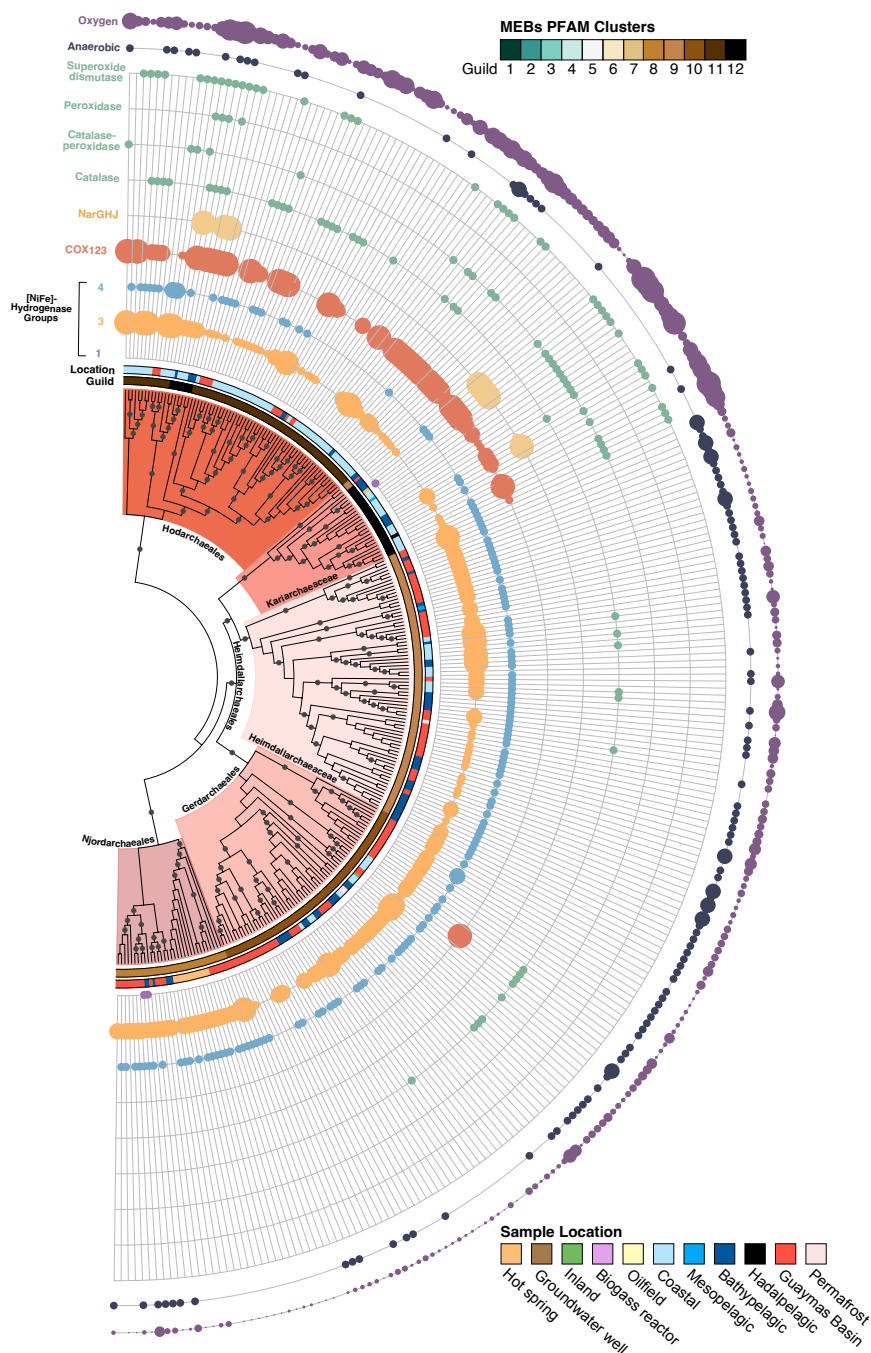
MsNuoD 407 SFTNL-QAVAATCEGGMVAIAIAVASLPMVMGVDR 442

c b

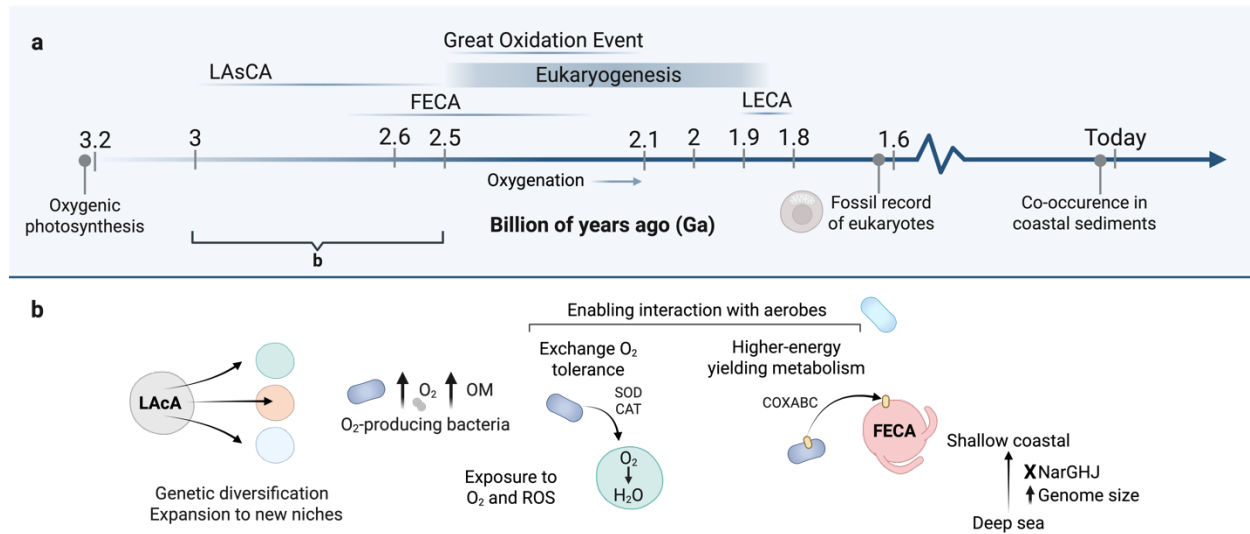
Supplementary Fig. 80 | **Multiple sequence alignment of NuoD homologs.** NuoD subunits from AlphaFold models described in this study aligned against selected homologs with Clustal Omega¹⁶² and visualized in Jalview v2.11¹⁷². Selected homologs cover all three domains of life. Positions of interest are marked with letters. **a.** Positions of both the histidine and tyrosine residues conserved in Complex I and necessary for the protonation of quinone (Tyr144 and His95 in *Y. lipolytica* NDUFS2¹⁰³). Not conserved in [NiFe]-hydrogenase complexes. **b.** Positions of four cysteines and one histidine residue conserved in [NiFe]-hydrogenases and necessary for [NiFe] metal coordination. Not all [NiFe] coordinating residues are conserved in Complex I, MBS, or FPO complexes. **c.** Positions of “canopy residues” (serine, proline, arginine, aspartate) that are important for [NiFe]-hydrogenase function and conserved in [NiFe]-hydrogenase complexes. Information for sequence names from top to bottom: *EcHycE*, *Escherichia coli* HycE (membrane-bound [NiFe]-hydrogenase; UniProt ID: P16431); *PfMbhL*, *Pyrococcus furiosus* MbhL (membrane-bound [NiFe]-hydrogenase; UniProt ID: Q8U0Z6); Hod-1 (M3_38_Bin_455); Kari-1 (GCA_015523565); Kari-2 (GCA_024280435); Gerd (GCA_016839405); Njord-1 (M288); Njord-2 (D4998_C1112_H3_Bin_254); Hod-2 (GCA_024276395); Asgard-1 (AB3033_2); Asgard-2 (GCA_021162905); *PfMbsL*, *Pyrococcus furiosus* MbsL (membrane-bound sulfane reductase; UniProt ID: I6V297); *HsNuoCD*, *Halobacterium salinarum* NuoCD (archaeal Fd-dependent Complex I; UniProt ID: Q9HRL9); *EcNuoD*, *Escherichia coli* NuoD (prokaryotic Complex I; UniProt ID: P33599); Hod-3 (M3_30_Bin_130); *AtNuoD*, *Arabidopsis thaliana* NuoD (eukaryotic Complex I; UniProt ID: P93306); YINDUFS2, *Yarrowia lipolytica* NDUF2 (eukaryotic Complex I; UniProt ID: F2Z626); *MmFpoD*, *Methanosarcina mazei* FpoD (archaeal F₄₂₀:phenazine oxidoreductase; UniProt ID: F1SVE4); *TtNqo4*, *Thermus thermophilus* Nqo4 (prokaryotic Complex I; UniProt ID: Q56220); *MsNuoD*, *Mycobacterium smegmatis* NuoD (prokaryotic Complex I; UniProt ID: A0QU33).



Supplementary Fig. 81 | **Updated Asgardarchaeota phylogeny with identified proteins.** Extensive mapping on the phylogeny described in the Fig. 1b (see main methods and figure for more details). The concentric rings (in-out) highlight the metabolic guilds based on Pfam clustering, sampling locations, [NiFe]-hydrogenase groups (1, purple; 3, yellow; 4, blue), number of COX123 proteins (red), NarGHJ (yellow), and ROS tolerance (green). The outermost rings are the number of identified oxidases enriched in anaerobes (EC: 1.6.3.4, 1.3.3.13, 1.14.13.163, 1.14.11.6, 1.13.12.7, 1.13.12.5, 1.14.13.208; dark purple) and number of oxygen-utilizing enzymes (light purple).



Supplementary Fig. 82 | **Updated Heimdallarchaeia phylogeny with identified proteins.** Extensive mapping on the phylogeny described in the Fig. 1b (see main methods and figure for more details) and subset Heimdallarchaeia from Fig. 81. The concentric rings (in-out) highlight the metabolic guilds based on Pfam clustering, sampling locations, [NiFe]-hydrogenase groups (1, purple; 3, yellow; 4, blue), number of COX123 proteins (red), NarGHJ (yellow), and ROS tolerance (green). The outermost rings are the number of identified oxidases enriched in anaerobes (1.6.3.4, 1.3.3.13, 1.14.13.163, 1.14.11.6, 1.13.12.7, 1.13.12.5, 1.14.13.208; dark purple) and total number of oxygen-utilizing enzymes (light purple).



Supplementary Fig. 83 | **Extended model details.** **a.** Timeline shown in Fig. 5a. **b.** Our data suggest genetic diversification in descendants of the LAsCA (LasCA: grey; descendant Asgardarchaeota: green, orange, blue) allowed their occupation of a variety of new niches. Oxygen-producing bacteria enabled oxygenation and the increase of organic matter (OM). Exposure to oxygen and the byproducts of aerobic respiration (ROS) before the Great Oxidation Event triggered the exchange of oxygen tolerance (e.g., SOD: superoxide dismutase and CAT: catalases) and increasing interactions with the aerobic mitochondrial ancestor (blue rod). The potential gain of ETC complexes (e.g., III and IV: yellow) prior to the first eukaryotic common ancestor (FECA; red) would have increased energy yield and reactive oxygen species (ROS) production. In Heimdallarchaeia, we identify the loss of nitrate reduction (NarGHJ) and increase in genome size in the transition from deep-sea to coastal genomes. Additional abbreviations include Ga (billions of years ago) and LECA (last eukaryotic common ancestor). Created in BioRender. Appler, K. (2025) <https://BioRender.com/5ct659q>.

Supplementary Tables

Supplementary Table 1. Description of the 404 Asgardarchaeota and 310 Alphaproteobacteria MAGs added by this study from the three separate sequencing efforts BH, GE2, and GE3. The table includes metadata for each MAG, including collection date, depth, elevation, and NCBI submission details (accession number, BioSample, BioProject, organism name, and isolate name). A description of the GE3 samples is further detailed in the GE3 tab.

Supplementary Table 2. List of the 47 non-ribosomal markers (NM47) selected from a set of 200 markers previously identified as core archaeal proteins for species phylogenies (Fig. 1ab and Supplementary Fig. 3). Description of the 869 Asgard archaea, including those used for the dereplicated analyses (Asgard_TaxonSet) and 309 Euryarchaeota and TACK outgroup (Outgroup_TaxonSet) MAGs. CheckM v1.2.1 completeness and contamination are reported along with genome quality score (GQS) for all MAGs (GQS = completeness (%) - 5 x contamination (%), as previously described). See methods for more detailed information.

Supplementary Table 3. Statistics by taxonomic lineage of the expanded genomic catalog, including the 404 Asgardarchaeota genomes added by this study (Fig. 1b; Added Diversity) and actual genome size calculated with SeqKit v2.3.0 and corrected for completeness with CheckM v1.2.1 (Genome Size).

Supplementary Table 4. Expansion of genomic diversity of the 869 Asgard archaea MAGs used for this study. To describe unique traits, we used several dimensionality reduction approaches, including lineage based on phylogenomic analysis (NM47), metabolic guilds based on unsupervised clustering of 19,179 protein families (Pfam v34) with Jaccard distance and Ward variance minimization with distance threshold of 0.35 (MEBs v2), and environmental source location. To compare with the sampling location, we include microbial growth conditions predicted using amino acid frequency with GenomeSPOT. Values for oxygen tolerance (probability), pH minimum, pH maximum, salinity minimum (% w/v NaCl), salinity maximum (% w/v NaCl), temperature maximum (°C), and temperature minimum (°C). Prediction of aerobicity with the XGBoost model is shown: 0 = anaerobe | 1 = aerobe, with associated probability. The MetaData tab includes all values and errors for all 869 Asgard MAGs. The Lineage tab includes values for each growth condition and the predicted conditions for growth. We used the average minimum to maximum range of each lineage to classify the pH (acidophilic <4, neutrophilic 4-9, and alkaliphilic >9), salinity (oligohaline <0.5, mesohaline 0.5-3, and moderate halophile 3-15 % w/v NaCl), and temperate predicted

conditions (psychrophilic <15, mesophilic 15-45, thermophilic 45-80, and hyperthermophilic 80-100°C). Additional genome statistics include GTDB-tk v2.1.1 taxonomic classification, quality metrics, and actual and predicted MAG Size (Mb) from SeqKit v2.3.0. See methods for a more detailed description.

Supplementary Table 5. Proposed taxonomic descriptions for Asgardarchaeota MAGs added by this study were submitted to the SeqCode Registry. Genome representatives passing quality control for type species are described in the MAGs sheet.

Supplementary Table 6. Comparative genomic analysis across Asgardarchaeota, Alphaproteobacteria, and single-cell eukaryotes. This analysis was split into three primary goals: 1) to identify functional annotations unique to Asgardarchaeota for each database by taxonomy (_Taxonomy), environment (_Environment), and guild (_Guild), 2) to determine distinct traits of Hodarchaeales between environment, taxonomy, and guild (Hod_), and 3) to identify potential partner functional contributions (Alphaproteobacteria and Asgardarchaeota) to the eukaryotic cell (see methods). The resulting profiles for parsing the 869 Asgardarchaeota genomes are shown for each annotation pipeline, separated by database (PFAM, KEGG, InterPro), using cutoffs of presence in 100%, 75%, 25%, 10%, and 5% of the focal group and absence in all others. Outputs from all three functional databases are also shown for Hodarchaeales, specifically distinguishing between environments and guilds. Genomes used for the tertiary goal are listed in the MAGs tab. Using these datasets, we determined potential gains at each branching point (A-I) in the NM47 phylogeny that are shared with single-cell eukaryotes (Functional_BranchingPoint; Supplementary Fig. 6). The next tab (Functional_Partner) shows annotations identified in either Alphaproteobacteria or Asgardarchaeota that are shared with single-cell eukaryotes. The final tab (Functional_Lineage) reports counts and raw annotations shared between each Asgardarchaeota lineage and single-cell eukaryotes, but not Alphaproteobacteria. See methods for more detailed descriptions.

Supplementary Table 7. Statistical analysis of the genome size and environmental distribution across 869 Asgard MAGs. The Environment sheet displays results from a one-sided t-test testing the null hypothesis that there is no difference in the actual and predicted genome size between genomes from depths >1000 m and < 1000 m. Additional sheets provide the ANOVA analysis testing the null hypotheses that the actual and predicted genome size are the same across different environments (Environment Actual Genome Size and Environment Predicted Genome Size) and taxonomic groups (Taxonomy Actual Genome Size and Taxonomy Predicted Genome Size).

Supplementary Table 8. Asgardarchaeota metabolic annotations. Key metabolic

annotations used for metabolic descriptions in the Supplementary cell diagrams, providing guild-, lineage-, and environmental-specific metabolic traits, are provided in the Asgard sheet. Presence/absence tables used for comparative genomic analyses and metabolic reconstruction from publicly available KEGG, InterPro, and MEROPS databases are also included. Results from dbCAN, HADEG, and MEBs annotation databases are provided in additional tabs. PSORTb was used to predict the location of peptidases and CAZymes (Carbohydrate-Active enZymes) for Hodarchaeales, Kariarchaeaceae, Ranarchaeia, and Asgardarchaeia. The final five sheets list reactive oxygen species (ROS) enzyme types, [NiFe]- and [FeFe]-hydrogenase clades and families, potential non-Asgard [NiFe]-hydrogenases grouping with novel clades or families, and RecA scaffolds (see methods for version information and detailed descriptions).

Supplementary Table 9. Information for the structural models of the group 4 [NiFe]-hydrogenase membrane-bound complexes, Fd-dependent Complex I complex, and electron bifurcating complexes, including models shown in Fig. 3a and Supplementary Figs. 41-44, and 45-52. The table includes sheets for the names and MAG identifiers of each model (Model Names), Foldseek scores of the group 4 [NiFe]-hydrogenase membrane-bound complexes and Fd-dependent Complex I complex (MbC Foldseek), Foldseek-multimer scores of the group 4 [NiFe]-hydrogenase membrane-bound complexes and Fd-dependent Complex I complex (MbC Foldseek-multimer), and Foldseek-multimer scores for the putative electron bifurcating complexes shown in Supplementary Figs. 28-29 and 76-77 (EbC Foldseek-multimer). The RMSD scores between AlphaFold2 models and experimental structures used in the cofactor modeling are provided in the next sheet (Cofactor Model RMSD). We used the abbreviations MbC for membrane-bound complexes and EbC for electron-bifurcating complexes. The final sheet (PDB IDs) lists the visualized Protein Data Bank structures.

Supplementary Table 10. Co-occurrence of Asgardarchaeota lineages and all the binned MAGs from the Bohai Sea (BH) and the Guaymas Basin (GE2 and GE3) is shown in the This_Study sheet. Coverage and relative abundance results from Sandpiper (v0.3.0), which searched the NCBI SRA public metagenome datasets prior to 2021 for OTUs linked with GTDB R220 Asgardarchaeota lineages and c__Alphaproteobacteria (138 orders), are provided in the Public_Coverage and Public_RelativeAbundance sheets, respectively. Additional information on SRA sampling sites and oxygen content for Hodarchaeales and Kariarchaeaceae is provided in the SRA_Site and SRA_Oxygen sheets, respectively. All site and oxygen metadata are based on author-submitted information and available associated literature.

Supplementary Table 11. Data used to produce the structural models and structural phylogeny of cytochrome c oxidase, including model(s) from each domain for Fig. 3b

(Models Fig 3b) and the structural phylogeny (Structural Phylogeny Fig 3b). The TMscore and RMSD used to produce the structural phylogeny are in the next two sheets (Pairwise TMscore Fig 3b and Pairwise RMSD Fig 3b). We also provide Supplementary Fig. 75 models (Model Suppl Fig 75) and FoldSeek results (FoldSeek Suppl Fig 75). A list of the visualized Protein Data Bank structures is also provided (PDB IDs).

Supplementary Table 12. Identification of previously described eukaryotic signature proteins (ESPs) in Eme et al., 2023 across the expanded dataset based on annotations.

Supplementary Table 13. Amino Acid Identity (AAI) between 869 Asgard archaea genomes calculated with CompareM v0.1.2 with aai_wf flag (<https://github.com/donovan-h-parks/CompareM>).

Supplementary Table 14. Abundance for Asgard archaea and Alphaproteobacteria MAGs across each of the three sequencing projects (BH, GE2, GE3), calculated with CoverM version 0.7.0 for the Guaymas Basin and version 0.6.1 for the Bohai Sea, using the relative abundance method.

Supplementary Table 15. Cutoffs for Asgard-specific HMM (hidden Markov model) profiles used to search for electron transport chain subunits across domains.

Subunit	E-value Cut-off
NuoA	1e-10
NuoB	1e-10
NuoC	1e-10
NuoE	1e-50
NuoF	1e-50
NuoG	1e-10
NuoH	1e-10
NuoI	1e-15
NuoJ	1e-3
NuoK	1e-10

NuoL	1e-20
NuoM	1e-50
NuoN	1e-109
SdhA	1e-10
SdhB	1e-70
SdhC	1e-2
SdhD	1e-10
QcrA	1e-10
QcrB	1e-10
QcrC	1e-10
NarH	1e-40
NarI	1e-20
NarJ	1e-10
NarG	1e-90

Supplementary References

1. Qiao, S. *et al.* Sediment accumulation and budget in the Bohai Sea, Yellow Sea and East China Sea. *Mar. Geol.* **390**, 270–281 (2017).
2. Lu, M. *et al.* Nutrients and heavy metals mediate the distribution of microbial community in the marine sediments of the Bohai Sea, China. *Environ. Pollut.* **255**, 113069 (2019).
3. Hu, B. *et al.* Heavy metal in surface sediments of the Liaodong Bay, Bohai Sea: distribution, contamination, and sources. *Environ. Monit. Assess.* **185**, 5071–5083 (2013).
4. Dombrowski, N., Teske, A. P. & Baker, B. J. Expansive microbial metabolic versatility and biodiversity in dynamic Guaymas Basin hydrothermal sediments. *Nat. Commun.* **9**, 4999 (2018).
5. Teske, A. *et al.* The Guaymas Basin Hiking Guide to Hydrothermal Mounds, Chimneys, and Microbial Mats: Complex Seafloor Expressions of Subsurface Hydrothermal Circulation. *Front. Microbiol.* **7**, 75 (2016).
6. Teske, A. *et al.* Characteristics and Evolution of sill-driven off-axis hydrothermalism in Guaymas Basin - the Ringvent site. *Sci. Rep.* **9**, 13847 (2019).
7. Eme, L. *et al.* Inference and reconstruction of the heimdallarchaeial ancestry of eukaryotes. *Nature* **618**, 992–999 (2023).
8. Tamarit, D. *et al.* Description of *Asgardarchaeum abyssi* gen. nov. spec. nov., a

- novel species within the class Asgardarchaeia and phylum Asgardarchaeota in accordance with the SeqCode. *Syst. Appl. Microbiol.* **47**, 126525 (2024).
9. Valentin-Alvarado, L. E. *et al.* Asgard archaea modulate potential methanogenesis substrates in wetland soil. *Nat. Commun.* **15**, 6384 (2024).
 10. Tully, B. J., Graham, E. D. & Heidelberg, J. F. The reconstruction of 2,631 draft metagenome-assembled genomes from the global oceans. *Sci Data* **5**, 170203 (2018).
 11. Liu, Y. *et al.* Expanded diversity of Asgard archaea and their relationships with eukaryotes. *Nature* **593**, 553–557 (2021).
 12. Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
 13. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).
 14. 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 (2017).
 15. Guindon, S. *et al.* New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
 16. Huang, W.-C. *et al.* Phylogenomic analyses reveal that Panguiararchaeum is a clade of genome-reduced Asgard archaea within the Njordarchaeia. *Mol. Biol. Evol.* (2025)

doi:10.1093/molbev/msaf201.

17. Susko, E. & Roger, A. J. On reduced amino acid alphabets for phylogenetic inference. *Mol. Biol. Evol.* **24**, 2139–2150 (2007).
18. Steenwyk, J. L. *et al.* BioKIT: a versatile toolkit for processing and analyzing diverse types of sequence data. *Genetics* **221**, (2022).
19. Wang, H.-C., Minh, B. Q., Susko, E. & Roger, A. J. Modeling Site Heterogeneity with Posterior Mean Site Frequency Profiles Accelerates Accurate Phylogenomic Estimation. *Syst. Biol.* **67**, 216–235 (2018).
20. Evans, R. *et al.* Protein complex prediction with AlphaFold-Multimer. *bioRxiv* 2021.10.04.463034 (2022) doi:10.1101/2021.10.04.463034.
21. van Kempen, M. *et al.* Fast and accurate protein structure search with Foldseek. *Nat. Biotechnol.* **42**, 243–246 (2024).
22. Kim, W. *et al.* Rapid and sensitive protein complex alignment with Foldseek-Multimer. *Nat. Methods* **22**, 469–472 (2025).
23. Pettersen, E. F. *et al.* UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).
24. Steinhilper, R., Höff, G., Heider, J. & Murphy, B. J. Structure of the membrane-bound formate hydrogenlyase complex from *Escherichia coli*. *Nat. Commun.* **13**, 5395 (2022).
25. Terwilliger, T. C. *et al.* AlphaFold predictions are valuable hypotheses and

- accelerate but do not replace experimental structure determination. *Nat. Methods* 1–7 (2023).
26. Yu, H. Structure of an ancient respiratory system. *Cell* **173**, (2018).
 27. Klusch, N. *et al.* Cryo-EM structure of the respiratory I + III₂ supercomplex from *Arabidopsis thaliana* at 2 Å resolution. *Nat. Plants* **9**, 142–156 (2023).
 28. Laughlin, T. G., Bayne, A. N., Trempe, J.-F., Savage, D. F. & Davies, K. M. Structure of the complex I-like molecule NDH of oxygenic photosynthesis. *Nature* **566**, 411–414 (2019).
 29. Wagner, T., Koch, J., Ermler, U. & Shima, S. Methanogenic heterodisulfide reductase (HdrABC-MvhAGD) uses two noncubane [4Fe-4S] clusters for reduction. *Science* **357**, 699–703 (2017).
 30. Page, C. C., Moser, C. C., Chen, X. & Dutton, P. L. Natural engineering principles of electron tunnelling in biological oxidation-reduction. *Nature* **402**, (1999).
 31. Liang, Y. *et al.* Structure of mycobacterial respiratory complex I. *Proc. Natl. Acad. Sci. U. S. A.* **120**, e2214949120 (2023).
 32. Yu, D., Chojnowski, G., Rosenthal, M. & Kosinski, J. AlphaPulldown-a python package for protein-protein interaction screens using AlphaFold-Multimer. *Bioinformatics* **39**, (2023).
 33. Shor, B. & Schneidman-Duhovny, D. CombFold: predicting structures of large protein assemblies using a combinatorial assembly algorithm and AlphaFold2. *Nat.*

Methods **21**, 477–487 (2024).

34. Martijn, J., Vosseberg, J., Guy, L., Offre, P. & Ettema, T. J. G. Deep mitochondrial origin outside the sampled alphaproteobacteria. *Nature* **557**, 101–105 (2018).
35. Hördt, A. *et al.* Analysis of 1,000+ Type-Strain Genomes Substantially Improves Taxonomic Classification of Alphaproteobacteria. *Front. Microbiol.* **11**, 468 (2020).
36. Griffith, D. M., Veech, J. A. & Marsh, C. J. cooccur: Probabilistic Species Co-Occurrence Analysis in R. *J. Stat. Softw.* **69**, (2016).
37. Woodcroft, B. J. *et al.* Comprehensive taxonomic identification of microbial species in metagenomic data using SingleM and Sandpiper. *Nat. Biotechnol.* (2025) doi:10.1038/s41587-025-02738-1.
38. Lartillot, N. & Philippe, H. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol. Biol. Evol.* **21**, 1095–1109 (2004).
39. Quang, L. S., Gascuel, O. & Lartillot, N. Empirical profile mixture models for phylogenetic reconstruction. *Bioinformatics* **24**, 2317–2323 (2008).
40. Dong, X. *et al.* Thermogenic hydrocarbon biodegradation by diverse depth-stratified microbial populations at a Scotian Basin cold seep. *Nat. Commun.* **11**, 5825 (2020).
41. Barnum, T. P. *et al.* Predicting microbial growth conditions from amino acid composition. *bioRxiv* 2024.03.22.586313 (2024) doi:10.1101/2024.03.22.586313.
42. Huang, L. *et al.* dbCAN-seq: a database of carbohydrate-active enzyme (CAZyme)

- sequence and annotation. *Nucleic Acids Res.* **46**, D516–D521 (2018).
43. Rawlings, N. D., Barrett, A. J. & Bateman, A. MEROPS: the peptidase database. *Nucleic Acids Res.* **38**, D227-33 (2010).
44. Santana-Molina, C., Williams, T. A., Snel, B. & Spang, A. Chimeric Origins and Dynamic Evolution of Central Carbon Metabolism in Eukaryotes. *bioRxiv* 2024.05.29.596406 (2024) doi:10.1101/2024.05.29.596406.
45. Tabita, F. R., Satagopan, S., Hanson, T. E., Kreel, N. E. & Scott, S. S. Distinct form I, II, III, and IV Rubisco proteins from the three kingdoms of life provide clues about Rubisco evolution and structure/function relationships. *J. Exp. Bot.* **59**, 1515–1524 (2008).
46. Ticak, T., Kountz, D. J., Girosky, K. E., Krzycki, J. A. & Ferguson, D. J., Jr. A nonpyrrolysine member of the widely distributed trimethylamine methyltransferase family is a glycine betaine methyltransferase. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E4668-76 (2014).
47. Picking, J. W., Behrman, E. J., Zhang, L. & Krzycki, J. A. MtpB, a member of the MttB superfamily from the human intestinal acetogen *Eubacterium limosum*, catalyzes proline betaine demethylation. *J. Biol. Chem.* **294**, 13697–13707 (2019).
48. Ellenbogen, J. B., Jiang, R., Kountz, D. J., Zhang, L. & Krzycki, J. A. The MttB superfamily member MtyB from the human gut symbiont *Eubacterium limosum* is a cobalamin-dependent γ -butyrobetaine methyltransferase. *J. Biol. Chem.* **297**,

101327 (2021).

49. Mursula, A. M., van Aalten, D. M., Hiltunen, J. K. & Wierenga, R. K. The crystal structure of delta(3)-delta(2)-enoyl-CoA isomerase. *J. Mol. Biol.* **309**, 845–853 (2001).
50. Rude, M. A. *et al.* Terminal olefin (1-alkene) biosynthesis by a novel p450 fatty acid decarboxylase from *Jeotgalicoccus* species. *Appl. Environ. Microbiol.* **77**, 1718–1727 (2011).
51. Verma, R. *et al.* Vms1 and ANKZF1 peptidyl-tRNA hydrolases release nascent chains from stalled ribosomes. *Nature* **557**, 446–451 (2018).
52. Kornberg, H. L. & Krebs, H. A. Synthesis of cell constituents from C2-units by a modified tricarboxylic acid cycle. *Nature* **179**, 988–991 (1957).
53. Bräsen, 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).
54. Tsandev, I., Reed, D. C. & Slomp, C. P. Phosphorus diagenesis in deep-sea sediments: Sensitivity to water column conditions and global scale implications. *Chem. Geol.* **330–331**, 127–139 (2012).
55. Jerlström-Hultqvist, J. *et al.* A unique symbiosome in an anaerobic single-celled eukaryote. *Nat. Commun.* **15**, 9726 (2024).
56. Croft, M. T., Lawrence, A. D., Raux-Deery, E., Warren, M. J. & Smith, A. G. Algae acquire vitamin B12 through a symbiotic relationship with bacteria. *Nature* **438**, 90–

93 (2005).

57. Grant, M. A. A., Kazamia, E., Cicuta, P. & Smith, A. G. Direct exchange of vitamin B12 is demonstrated by modelling the growth dynamics of algal-bacterial cocultures. *ISME J.* **8**, 1418–1427 (2014).
58. Donegan, R. K., Moore, C. M., Hanna, D. A. & Reddi, A. R. Handling heme: The mechanisms underlying the movement of heme within and between cells. *Free Radic. Biol. Med.* **133**, 88–100 (2019).
59. Wang, W. & Shao, Z. The long-chain alkane metabolism network of *Alcanivorax dieselolei*. *Nat. Commun.* **5**, 5755 (2014).
60. Rojas-Vargas, J., Castelán-Sánchez, H. G. & Pardo-López, L. HADEG: A curated hydrocarbon aerobic degradation enzymes and genes database. *Comput. Biol. Chem.* **107**, 107966 (2023).
61. Goyal, A. Metabolic adaptations underlying genome flexibility in prokaryotes. *PLoS Genet.* **14**, e1007763 (2018).
62. Imachi, H. *et al.* Isolation of an archaeon at the prokaryote-eukaryote interface. *Nature* **577**, 519–525 (2020).
63. Spang, A. *et al.* Proposal of the reverse flow model for the origin of the eukaryotic cell based on comparative analyses of Asgard archaeal metabolism. *Nat Microbiol* **4**, 1138–1148 (2019).
64. López-García, P. & Moreira, D. The Syntrophy hypothesis for the origin of

- eukaryotes revisited. *Nat Microbiol* **5**, 655–667 (2020).
65. Greening, C. *et al.* Genomic and metagenomic surveys of hydrogenase distribution indicate H₂ is a widely utilised energy source for microbial growth and survival. *ISME J.* **10**, 761–777 (2016).
66. [No title]. <https://academic.oup.com/ismecommun/article/7461079>.
67. Greening, C. *et al.* Unique Minimal and Hybrid Hydrogenases are Active in Anaerobic Archaea. (2023) doi:10.2139/ssrn.4520792.
68. Calusinska, M., Happe, T., Joris, B. & Wilmotte, A. The surprising diversity of clostridial hydrogenases: a comparative genomic perspective. *Microbiology* **156**, 1575–1588 (2010).
69. Land, H. *et al.* Characterization of a putative sensory [FeFe]-hydrogenase provides new insight into the role of the active site architecture. *Chem. Sci.* **11**, 12789–12801 (2020).
70. De Anda, V. *et al.* Brockarchaeota, a novel archaeal phylum with unique and versatile carbon cycling pathways. *Nat. Commun.* **12**, 2404 (2021).
71. Kohtz, A. J., Jay, Z. J., Lynes, M. M., Krukenberg, V. & Hatzenpichler, R. Culexarchaeia, a novel archaeal class of anaerobic generalists inhabiting geothermal environments. *ISME Commun* **2**, 86 (2022).
72. Farag Ibrahim F., Zhao Rui & Biddle Jennifer F. “Sifarchaeota,” a Novel Asgard Phylum from Costa Rican Sediment Capable of Polysaccharide Degradation and

- Anaerobic Methylotrophy. *Appl. Environ. Microbiol.* **87**, e02584-20 (2021).
73. [No title]. <https://academic.oup.com/ismecommun/article/7462909>.
74. Seitz, K. W. *et al.* Asgard archaea capable of anaerobic hydrocarbon cycling. *Nat. Commun.* **10**, 1822 (2019).
75. MacLeod, F., Kindler, G. S., Wong, H. L., Chen, R. & Burns, B. P. Asgard archaea: Diversity, function, and evolutionary implications in a range of microbiomes. *AIMS Microbiol* **5**, 48–61 (2019).
76. Priya, B. *et al.* Comparative analysis of cyanobacterial superoxide dismutases to discriminate canonical forms. *BMC Genomics* **8**, 435 (2007).
77. Ursby, T., Adinolfi, B. S., Al-Karadaghi, S., De Vendittis, E. & Bocchini, V. Iron superoxide dismutase from the archaeon *Sulfolobus solfataricus*: analysis of structure and thermostability. *J. Mol. Biol.* **286**, 189–205 (1999).
78. Sutherland, K. M., Ward, L. M., Colombero, C.-R. & Johnston, D. T. Inter-domain horizontal gene transfer of nickel-binding superoxide dismutase. *Geobiology* **19**, 450–459 (2021).
79. Zamocky, M., Furtmüller, P. G. & Obinger, C. Evolution of catalases from bacteria to humans. *Antioxid. Redox Signal.* **10**, 1527–1548 (2008).
80. Yuan, F. *et al.* The Richness and Diversity of Catalases in Bacteria. *Front. Microbiol.* **12**, 645477 (2021).
81. 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).
82. Chintapalli, S. V. *et al.* Reevaluation of the evolutionary events within recA/RAD51 phylogeny. *BMC Genomics* **14**, 240 (2013).
83. Lin, Z., Kong, H., Nei, M. & Ma, H. Origins and evolution of the recA/RAD51 gene family: evidence for ancient gene duplication and endosymbiotic gene transfer. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 10328–10333 (2006).
84. Dey, G. *et al.* Closed mitosis requires local disassembly of the nuclear envelope. *Nature* **585**, 119–123 (2020).
85. Asakawa, H., Hirano, Y., Shindo, T., Haraguchi, T. & Hiraoka, Y. Fission yeast Ish1 and Les1 interact with each other in the lumen of the nuclear envelope. *Genes Cells* **27**, 643–656 (2022).
86. Søndergaard, D., Pedersen, C. N. S. & Greening, C. HydDB: A web tool for hydrogenase classification and analysis. *Sci. Rep.* **6**, 34212 (2016).
87. Jespersen, M. *et al.* Diverse lineages and adaptations of oxygen-adapted hydrogenases. *Trends Biochem. Sci.* **50**, 596–609 (2025).
88. Yu, H., Schut, G. J., Haja, D. K., Adams, M. W. W. & Li, H. Evolution of complex I-like respiratory complexes. *J. Biol. Chem.* **296**, 100740 (2021).
89. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

90. Necci, M., Piovesan, D., CAID Predictors, DisProt Curators & Tosatto, S. C. E. Critical assessment of protein intrinsic disorder prediction. *Nat. Methods* **18**, 472–481 (2021).
91. Ruff, K. M. & Pappu, R. V. AlphaFold and implications for intrinsically disordered proteins. *J. Mol. Biol.* **433**, 167208 (2021).
92. Guo, H.-B. *et al.* AlphaFold2 models indicate that protein sequence determines both structure and dynamics. *Sci. Rep.* **12**, 10696 (2022).
93. McCafferty, C. L., Pennington, E. L., Papoulas, O., Taylor, D. W. & Marcotte, E. M. Does AlphaFold2 model proteins' intracellular conformations? An experimental test using cross-linking mass spectrometry of endogenous ciliary proteins. *Commun. Biol.* **6**, 421 (2023).
94. Kravchuk, V. *et al.* A universal coupling mechanism of respiratory complex I. *Nature* **609**, 808–814 (2022).
95. Grba, D. N., Wright, J. J., Yin, Z., Fisher, W. & Hirst, J. Molecular mechanism of the ischemia-induced regulatory switch in mammalian complex I. *Science* **384**, 1247–1253 (2024).
96. Kaila, V. R. I. Long-range proton-coupled electron transfer in biological energy conversion: towards mechanistic understanding of respiratory complex I. *J. R. Soc. Interface* **15**, (2018).
97. Grba, D. N., Chung, I., Bridges, H. R., Agip, A.-N. A. & Hirst, J. Investigation of

- hydrated channels and proton pathways in a high-resolution cryo-EM structure of mammalian complex I. *Sci. Adv.* **9**, (2023).
98. Grba, D. N. & Hirst, J. Mitochondrial complex I structure reveals ordered water molecules for catalysis and proton translocation. *Nat. Struct. Mol. Biol.* **27**, (2020).
99. Schut, G. J. *et al.* The role of geochemistry and energetics in the evolution of modern respiratory complexes from a proton-reducing ancestor. *Biochim. Biophys. Acta* **1857**, 958–970 (2016).
100. Schuller, J. M. *et al.* Structural adaptations of photosynthetic complex I enable ferredoxin-dependent electron transfer. *Science* **363**, 257–260 (2019).
101. Albracht, S. P. Intimate relationships of the large and the small subunits of all nickel hydrogenases with two nuclear-encoded subunits of mitochondrial NADH: ubiquinone oxidoreductase. *Biochim. Biophys. Acta* **1144**, 221–224 (1993).
102. Pilkington, S. J., Skehel, J. M., Gennis, R. B. & Walker, J. E. Relationship between mitochondrial NADH-ubiquinone reductase and a bacterial NAD-reducing hydrogenase. *Biochemistry* **30**, 2166–2175 (1991).
103. Kim, H., Saura, P., Pöverlein, M. C., Gamiz-Hernandez, A. P. & Kaila, V. R. I. Quinone catalysis modulates proton transfer reactions in the membrane domain of respiratory Complex I. *J. Am. Chem. Soc.* **145**, 17075–17086 (2023).
104. Moparthy, V. K. & Hägerhäll, C. The evolution of respiratory chain complex I from a smaller last common ancestor consisting of 11 protein subunits. *J. Mol. Evol.* **72**, 484–

- 497 (2011).
105. Gonzalez, O. *et al.* Systems analysis of bioenergetics and growth of the extreme halophile *Halobacterium salinarum*. *PLoS Comput. Biol.* **5**, e1000332 (2009).
106. She, Q. *et al.* The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 7835–7840 (2001).
107. Downing, B. E. & Nayak, D. D. Innovations in the electron transport chain fuel archaeal methane metabolism. *Trends Biochem. Sci.* (2025)
doi:10.1016/j.tibs.2025.02.004.
108. Hedderich, R., Berkessel, A. & Thauer, R. K. Purification and properties of heterodisulfide reductase from *Methanobacterium thermoautotrophicum* (strain Marburg). *Eur. J. Biochem.* **193**, 255–261 (1990).
109. Yu, H. Structure of the respiratory MBS complex reveals iron-sulfur cluster catalyzed sulfane sulfur reduction in ancient life. *Nat. Commun.* **11**, (2020).
110. Feng, X., Schut, G. J., Adams, M. W. W. & Li, H. Structures and electron transport paths in the four families of flavin-based electron bifurcation enzymes. *Subcell. Biochem.* **104**, 383–408 (2024).
111. Radon, C. *et al.* Cryo-EM structures reveal intricate Fe-S cluster arrangement and charging in *Rhodobacter capsulatus* formate dehydrogenase. *Nat. Commun.* **11**, 1912 (2020).
112. Katsyv, A. *et al.* Molecular Basis of the Electron Bifurcation Mechanism in the

- [FeFe]-Hydrogenase Complex HydABC. *J. Am. Chem. Soc.* **145**, 5696–5709 (2023).
113. Buckel, W. & Thauer, R. K. Flavin-based electron bifurcation, A new mechanism of biological energy coupling. *Chem. Rev.* **118**, 3862–3886 (2018).
114. Wang, S., Huang, H., Kahnt, J. & Thauer, R. K. *Clostridium acidurici* electron-bifurcating formate dehydrogenase. *Appl. Environ. Microbiol.* **79**, 6176–6179 (2013).
115. Kumar, A. *et al.* Molecular architecture and electron transfer pathway of the Stn family transhydrogenase. *Nat. Commun.* **14**, 5484 (2023).
116. Li, H., Schut, G. J., Feng, X., Adams, M. W. W. & Li, H. Cryo-EM reveals a composite flavobicluster electron bifurcation site in the Bfu family member NfnABC. *Commun. Biol.* **8**, 239 (2025).
117. Xiao, X. *et al.* Cryo-EM structures define the electron bifurcating flavobicluster and ferredoxin binding site in an Archaeal nfn-Bfu transhydrogenase. *J. Biol. Chem.* 108410 (2025).
118. Sendker, F. L. *et al.* Frequent transitions in self-assembly across the evolution of a central metabolic enzyme. *Nat. Commun.* **15**, 10515 (2024).
119. Feng, X., Schut, G. J., Haja, D. K., Adams, M. W. W. & Li, H. Structure and electron transfer pathways of an electron-bifurcating NiFe-hydrogenase. *Sci. Adv.* **8**, eabm7546 (2022).
120. Kayastha, K. *et al.* Structure-based electron-confurcation mechanism of the Ldh-EtfAB complex. *Elife* **11**, (2022).

121. Fritsch, J., Lenz, O. & Friedrich, B. Structure, function and biosynthesis of O₂-tolerant hydrogenases. *Nat. Rev. Microbiol.* **11**, 106–114 (2013).
122. Batista, A. P., Marreiros, B. C. & Pereira, M. M. The antiporter-like subunit constituent of the universal adaptor of complex I, group 4 membrane-bound [NiFe]-hydrogenases and related complexes. *Biol. Chem.* **394**, 659–666 (2013).
123. Hedderich, R. Energy-converting [NiFe] hydrogenases from archaea and extremophiles: ancestors of complex I. *J. Bioenerg. Biomembr.* **36**, 65–75 (2004).
124. Brandt, U. Adaptations of an ancient modular machine. *Science (New York, N.Y.)* vol. 363 230–231 (2019).
125. Chadwick, G. L., Hemp, J., Fischer, W. W. & Orphan, V. J. Convergent evolution of unusual complex I homologs with increased proton pumping capacity: energetic and ecological implications. *ISME J.* **12**, 2668–2680 (2018).
126. Efremov, R. G. & Sazanov, L. A. The coupling mechanism of respiratory complex I - a structural and evolutionary perspective. *Biochim. Biophys. Acta* **1817**, 1785–1795 (2012).
127. Friedrich, T. & Weiss, H. Modular evolution of the respiratory NADH:ubiquinone oxidoreductase and the origin of its modules. *J. Theor. Biol.* **187**, 529–540 (1997).
128. Marreiros, B. C., Batista, A. P., Duarte, A. M. S. & Pereira, M. M. A missing link between complex I and group 4 membrane-bound [NiFe] hydrogenases. *Biochim. Biophys. Acta* **1827**, 198–209 (2013).

129. Shomura, Y. *et al.* Structural basis of the redox switches in the NAD⁺-reducing soluble [NiFe]-hydrogenase. *Science* **357**, 928–932 (2017).
130. Diamond, S. *et al.* Soils and sediments host Thermoplasmata archaea encoding novel copper membrane monooxygenases (CuMMOs). *ISME J.* **16**, 1348–1362 (2022).
131. Spang, A. *et al.* The genome of the ammonia-oxidizing Candidatus Nitrososphaera gargensis: insights into metabolic versatility and environmental adaptations. *Environ. Microbiol.* **14**, 3122–3145 (2012).
132. Aylward, F. O. & Santoro, A. E. Heterotrophic Thaumarchaea with Small Genomes Are Widespread in the Dark Ocean. *mSystems* **5**, (2020).
133. Bruska, M. K., Stiebritz, M. T. & Reiher, M. Analysis of differences in oxygen sensitivity of Fe-S clusters. *Dalton Trans.* **42**, 8729–8735 (2013).
134. Efremov, R. G. & Sazanov, L. A. Structure of the membrane domain of respiratory complex I. *Nature* **476**, 414–420 (2011).
135. Grettenberger, C., Gold, D. A. & Brown, C. T. Distribution of early-branching Cyanobacteria and the potential habitats that gave rise to the earliest oxygenic phototrophs. *mSphere* **10**, e0101324 (2025).
136. Glass, J. B., Elbon, C. E. & Williams, L. D. Something old, something new, something borrowed, something blue: the anaerobic microbial ancestry of aerobic respiration. *Trends Microbiol.* **31**, 135–141 (2023).

137. Jardim-Messeder, D. *et al.* Fumarate reductase superfamily: A diverse group of enzymes whose evolution is correlated to the establishment of different metabolic pathways. *Mitochondrion* **34**, 56–66 (2017).
138. Kraft, B., Strous, M. & Tegetmeyer, H. E. Microbial nitrate respiration--genes, enzymes and environmental distribution. *J. Biotechnol.* **155**, 104–117 (2011).
139. Dibrova, D. V., Shalaeva, D. N., Galperin, M. Y. & Mulkidjanian, A. Y. Emergence of cytochrome bc complexes in the context of photosynthesis. *Physiol. Plant.* **161**, 150–170 (2017).
140. Murali, R., Hemp, J. & Gennis, R. B. Evolution of quinol oxidation within the heme-copper oxidoreductase superfamily. *Biochim. Biophys. Acta Bioenerg.* **1863**, 148907 (2022).
141. Gribaldo, S., Talla, E. & Brochier-Armanet, C. Evolution of the haem copper oxidases superfamily: a rooting tale. *Trends Biochem. Sci.* **34**, 375–381 (2009).
142. Shiratori, T., Suzuki, S., Kakizawa, Y. & Ishida, K.-I. Phagocytosis-like cell engulfment by a planctomycete bacterium. *Nat. Commun.* **10**, 5529 (2019).
143. Rodrigues-Oliveira, T. *et al.* Actin cytoskeleton and complex cell architecture in an Asgard archaeon. *Nature* **613**, (2023).
144. Roger, A. J., Muñoz-Gómez, S. A. & Kamikawa, R. The origin and diversification of mitochondria. *Curr. Biol.* **27**, R1177–R1192 (2017).
145. Butenko, A., Lukeš, J., Speijer, D. & Wideman, J. G. Mitochondrial genomes

- revisited: why do different lineages retain different genes? *BMC Biol.* **22**, 15 (2024).
146. Baum, D. A. & Baum, B. An inside-out origin for the eukaryotic cell. *BMC Biol.* **12**, 76 (2014).
147. Björkholm, P., Ernst, A. M., Hagström, E. & Andersson, S. G. E. Why mitochondria need a genome revisited. *FEBS Lett.* **591**, 65–75 (2017).
148. Baum Buzz & Spang Anja. On the origin of the nucleus: a hypothesis. *Microbiol. Mol. Biol. Rev.* **87**, e00186-21 (2023).
149. Chen, Q., Vazquez, E. J., Moghaddas, S., Hoppel, C. L. & Lesnefsky, E. J. Production of reactive oxygen species by mitochondria: central role of complex III. *J. Biol. Chem.* **278**, 36027–36031 (2003).
150. Fichman, Y., Rowland, L., Oliver, M. J. & Mittler, R. ROS are evolutionary conserved cell-to-cell stress signals. *Proc. Natl. Acad. Sci. U. S. A.* **120**, e2305496120 (2023).
151. Speijer, D., Hammond, M. & Lukeš, J. Comparing Early Eukaryotic Integration of Mitochondria and Chloroplasts in the Light of Internal ROS Challenges: Timing is of the Essence. *MBio* **11**, (2020).
152. Speijer, D. Oxygen radicals shaping evolution: why fatty acid catabolism leads to peroxisomes while neurons do without it: FADH₂/NADH flux ratios determining mitochondrial radical formation were crucial for the eukaryotic invention of peroxisomes and catabolic tissue differentiation. *Bioessays* **33**, 88–94 (2011).

153. Shadel, G. S. & Horvath, T. L. Mitochondrial ROS signaling in organismal homeostasis. *Cell* **163**, 560–569 (2015).
154. Banci, L., Bertini, I., Cavallaro, G. & Ciofi-Baffoni, S. Seeking the determinants of the elusive functions of Sco proteins. *FEBS J.* **278**, 2244–2262 (2011).
155. Balatri, E., Banci, L., Bertini, I., Cantini, F. & Ciofi-Baffoni, S. Solution structure of Sco1: a thioredoxin-like protein Involved in cytochrome c oxidase assembly. *Structure* **11**, 1431–1443 (2003).
156. Tan, S. H. *et al.* A Sco protein among the hypothetical proteins of *Bacillus lehensis* G1: Its 3D macromolecular structure and association with Cytochrome C Oxidase. *BMC Struct. Biol.* **14**, 11 (2014).
157. Banci, L., Bertini, I., Cavallaro, G. & Rosato, A. The functions of Sco proteins from genome-based analysis. *J. Proteome Res.* **6**, 1568–1579 (2007).
158. Pierron, D. *et al.* Cytochrome c oxidase: evolution of control via nuclear subunit addition. *Biochim. Biophys. Acta* **1817**, 590–597 (2012).
159. Seib, K. L., Jennings, M. P. & McEwan, A. G. A Sco homologue plays a role in defence against oxidative stress in pathogenic *Neisseria*. *FEBS Lett.* **546**, 411–415 (2003).
160. Hedlund, B. P. *et al.* SeqCode: a nomenclatural code for prokaryotes described from sequence data. *Nat Microbiol* **7**, 1702–1708 (2022).
161. Elfmann, C. & Stülke, J. PAE viewer: a webserver for the interactive visualization of

the predicted aligned error for multimer structure predictions and crosslinks.

Nucleic Acids Res. **51**, W404–W410 (2023).

162. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).

163. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320-4 (2014).

164. Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35**, 1026–1028 (2017).

165. Koua, D. *et al.* PeroxiBase: a database with new tools for peroxidase family classification. *Nucleic Acids Res.* **37**, D261-6 (2009).

166. Bojanova, D. P. *et al.* Well-hidden methanogenesis in deep, organic-rich sediments of Guaymas Basin. *ISME J.* **17**, 1828–1838 (2023).

167. Lundgren, C. A. K. *et al.* Scavenging of superoxide by a membrane-bound superoxide oxidase. *Nat. Chem. Biol.* **14**, 788–793 (2018).

168. Guo, Y. *et al.* Monomer and dimer structures of cytochrome bo3 ubiquinol oxidase from *Escherichia coli*. *Protein Sci.* **32**, e4616 (2023).

169. Yoshikawa, S. *et al.* Redox-coupled crystal structural changes in bovine heart cytochrome c oxidase. *Science* **280**, 1723–1729 (1998).

170. Lubner, C. E. *et al.* Mechanistic insights into energy conservation by flavin-based electron bifurcation. *Nat. Chem. Biol.* **13**, 655–659 (2017).

171. Meng, E. C. *et al.* UCSF ChimeraX: Tools for structure building and analysis. *Protein Sci.* **32**, e4792 (2023).
172. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).