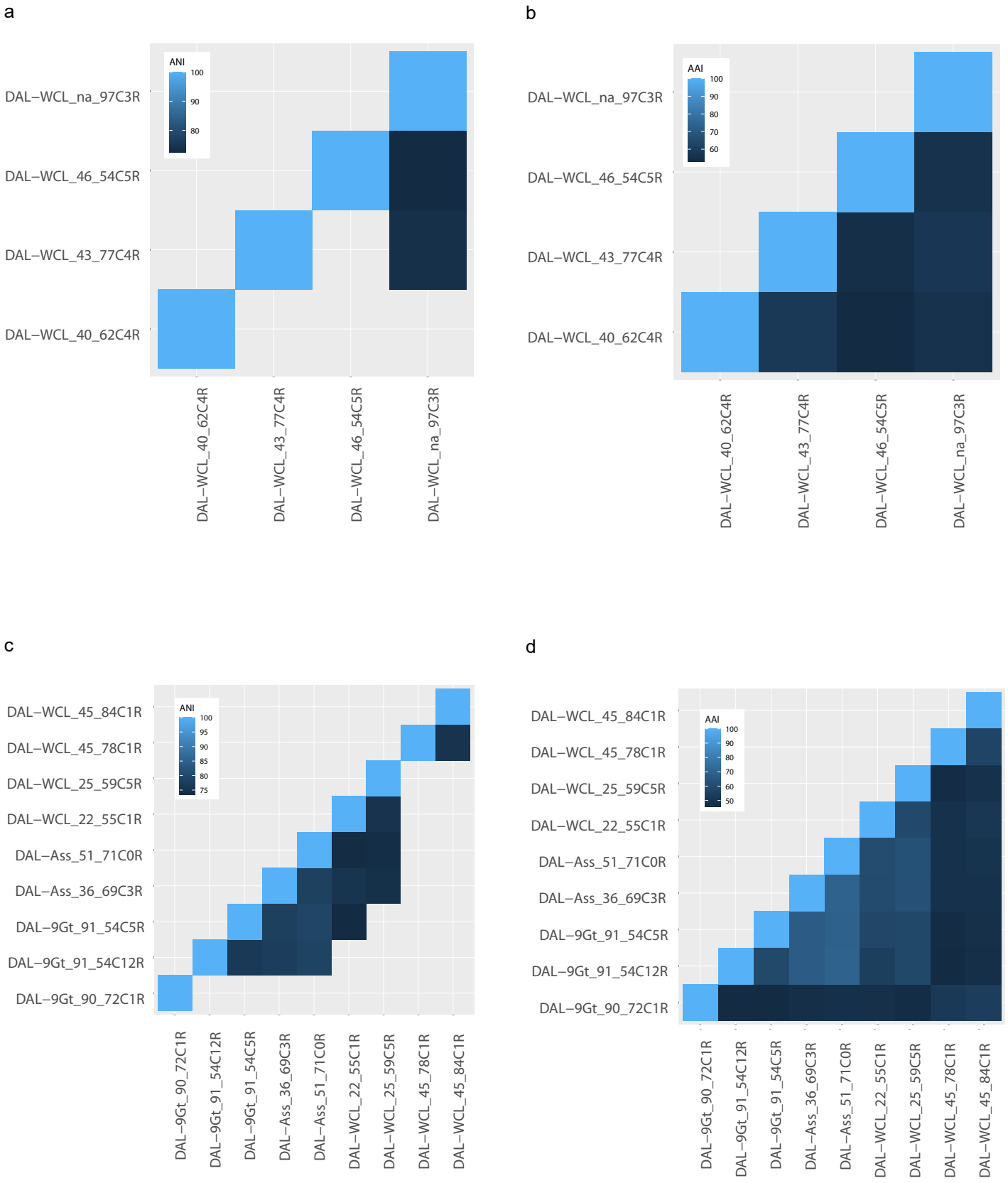
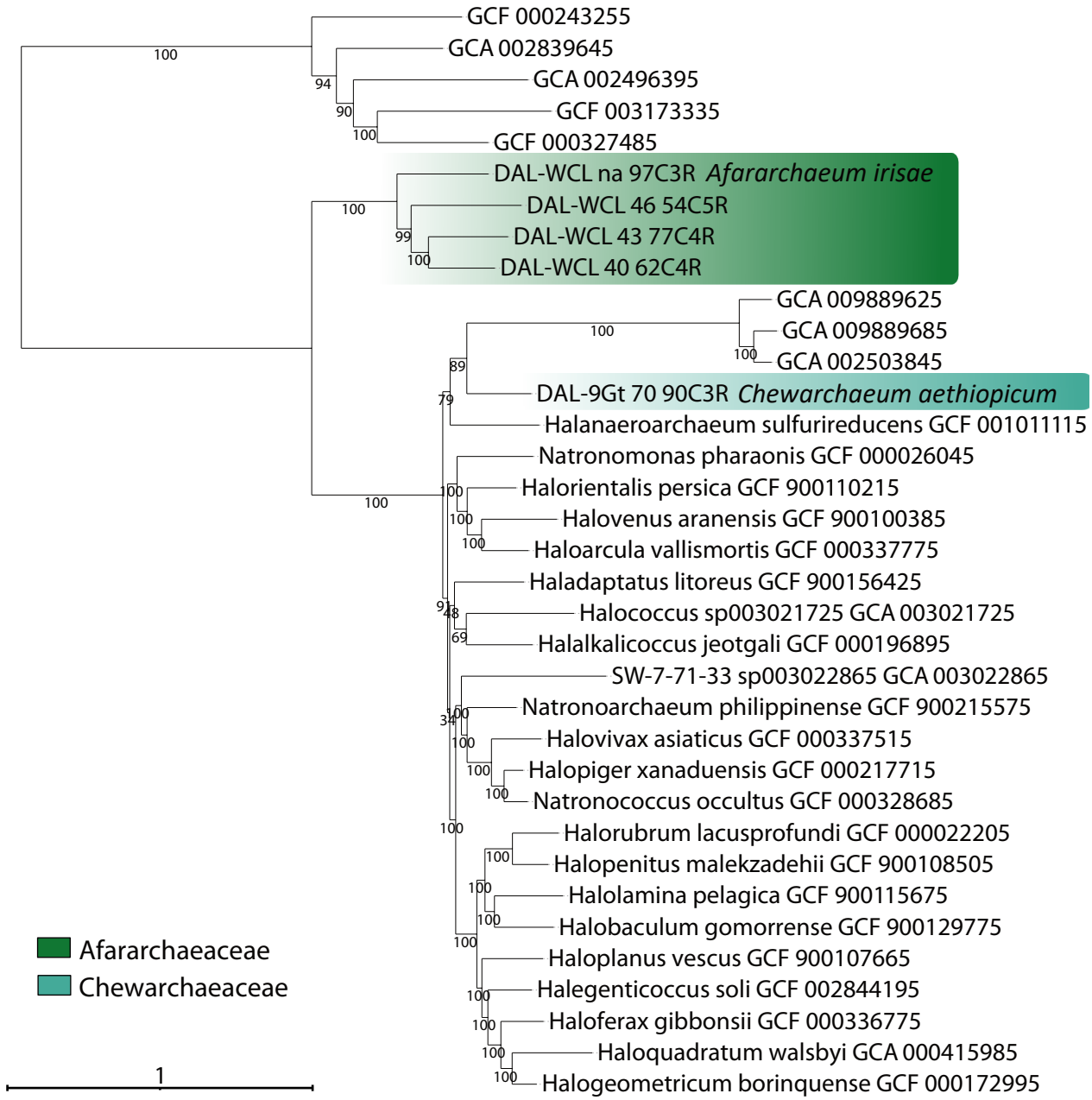
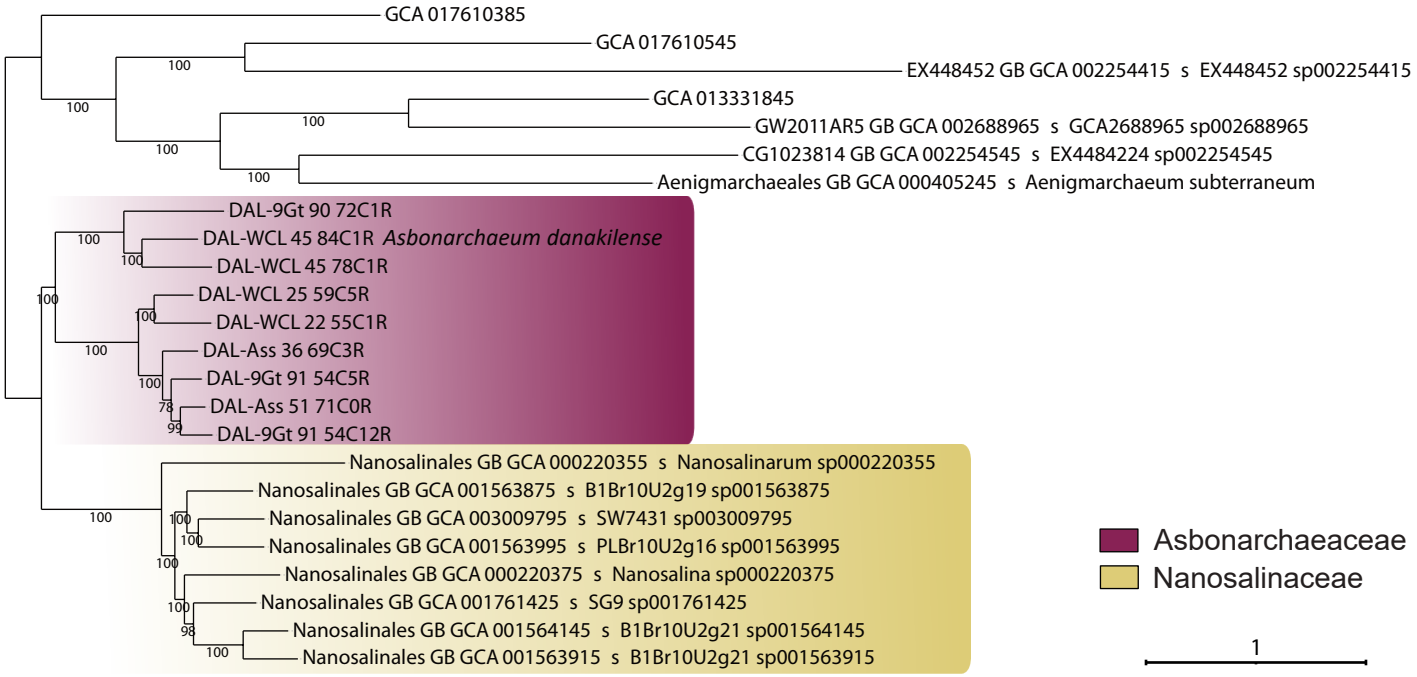


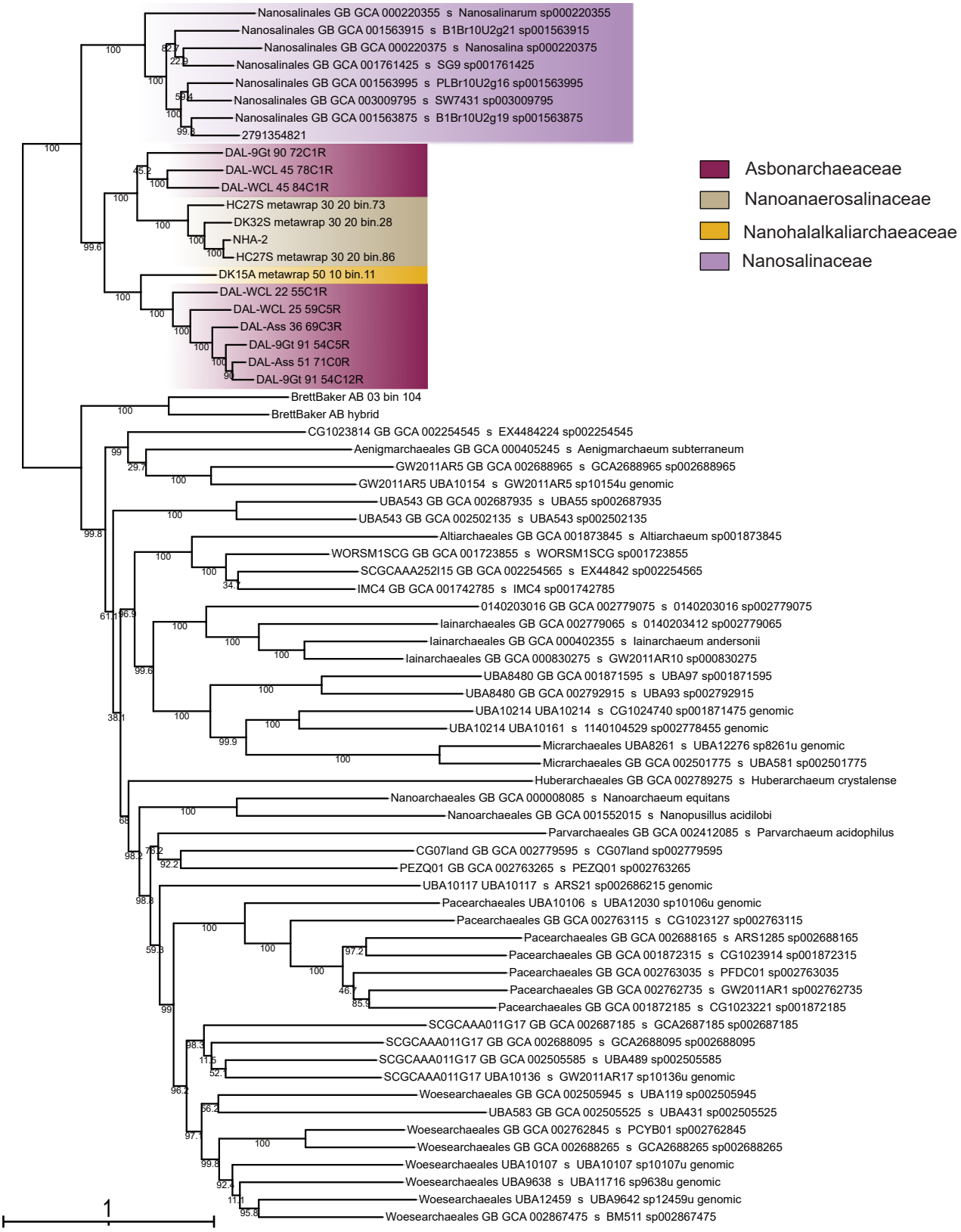
Supplementary Figure 1 | Average nucleotide identities (ANI) and average amino acid identities (AAI) among the Afararchaeaceae and Asbonarchaeaceae MAGs. (a,c) Pairwise ANI for the four Afararchaeaceae and nine Asbonarchaeaceae MAGs. (b,d) Pairwise AAI for the four Afararchaeaceae and nine Asbonarchaeaceae MAGs. ANI are only reported for values above 70% identty.



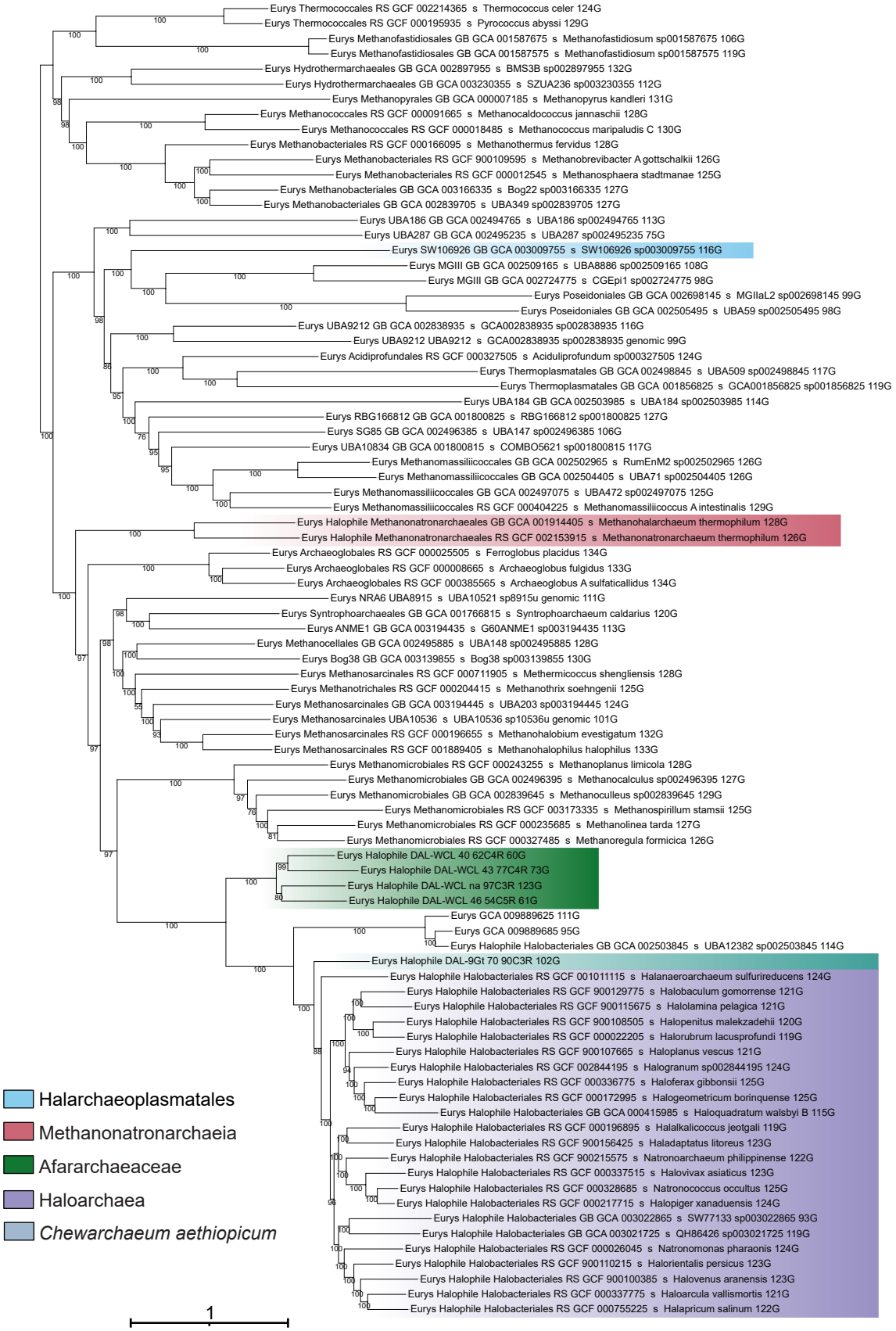
Supplementary Figure 2 | Maximum likelihood phylogeny of 35 archaeal taxa based on the concatenated alignment of 122 single-copy proteins obtained from the Genome Taxonomy Database (GTDB). The ML tree was constructed using the LG+C60+ F+Γ4 model of evolution. Branch support was assessed using 1,000 ultrafast bootstraps via IQ-TREE implementation. The four new afararchaeal MAGs are highlighted in green, and the new deep-branching Haloarchaea MAG '*Chewarchaeum aethiopicum*' is highlighted in teal. The scale bar represents the estimated number of substitutions per site. Each tip contains a GTDB identification label and the species name when available.



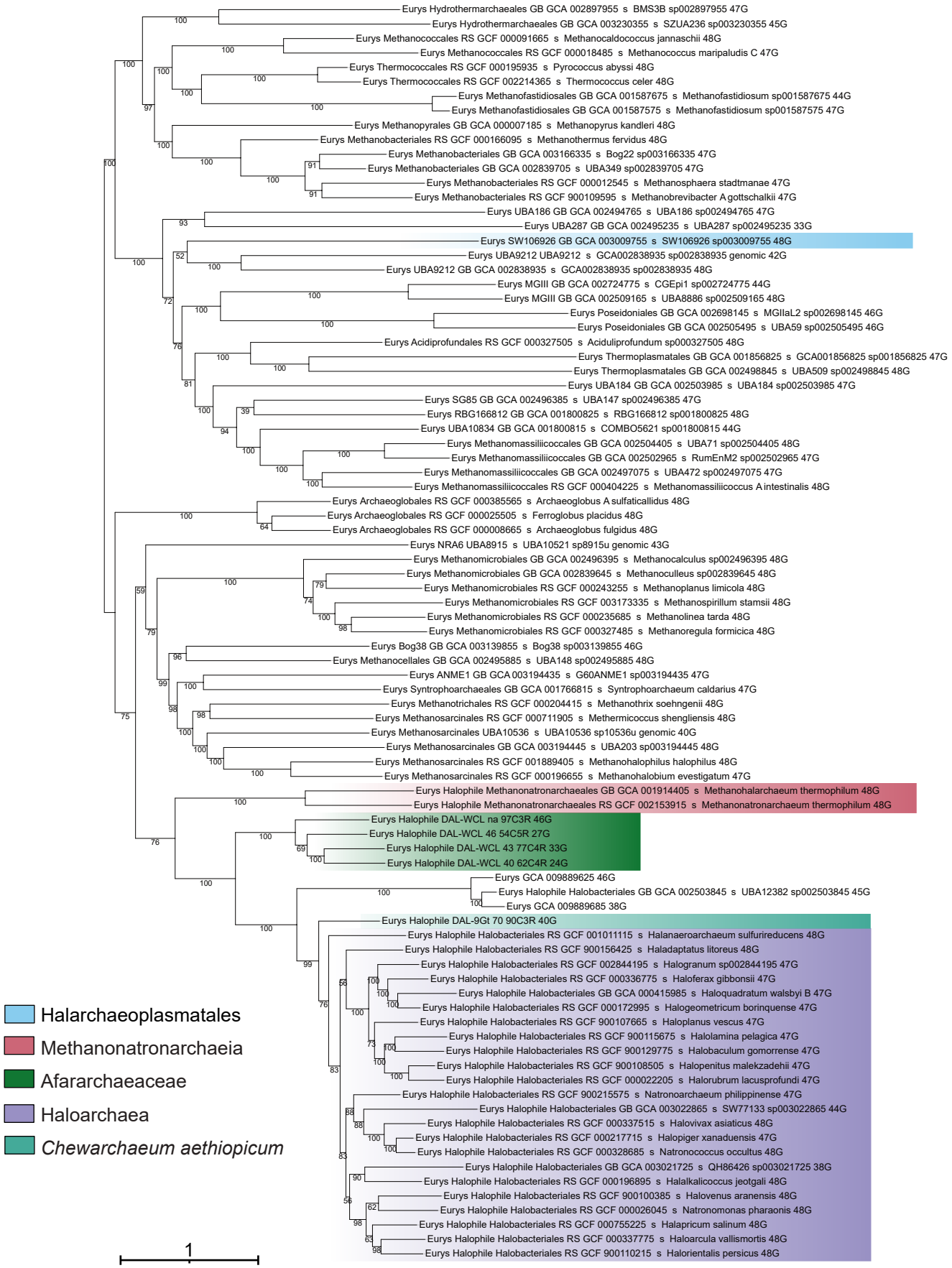
Supplementary Figure 3 | Maximum likelihood phylogeny of 24 DPANN archaea based on the concatenated alignment of 99 single-copy proteins obtained from the Genome Taxonomy Database (GTDB). The ML tree was constructed using the LG+C60+F+Γ4 model of evolution. Branch support was assessed using 1,000 ultrafast bootstraps via IQ-TREE implementation. The scale bar represents the estimated number of substitutions per site. Each tip contains a GTDB identification label, the GTDB order, and the species name when available.



Supplementary Figure 4 | Maximum likelihood phylogeny of 70 DPANN archaea based on the concatenated alignment of 24 large subunit ribosomal proteins. The ML tree was constructed using the LG+C20+Γ4 model of evolution. Branch support was assessed using 1,000 ultrafast bootstraps via IQ-TREE implementation. The scale bar represents the estimated number of substitutions per site. Each tip contains a GTDB identification label, the GTDB order, and the species name when available.

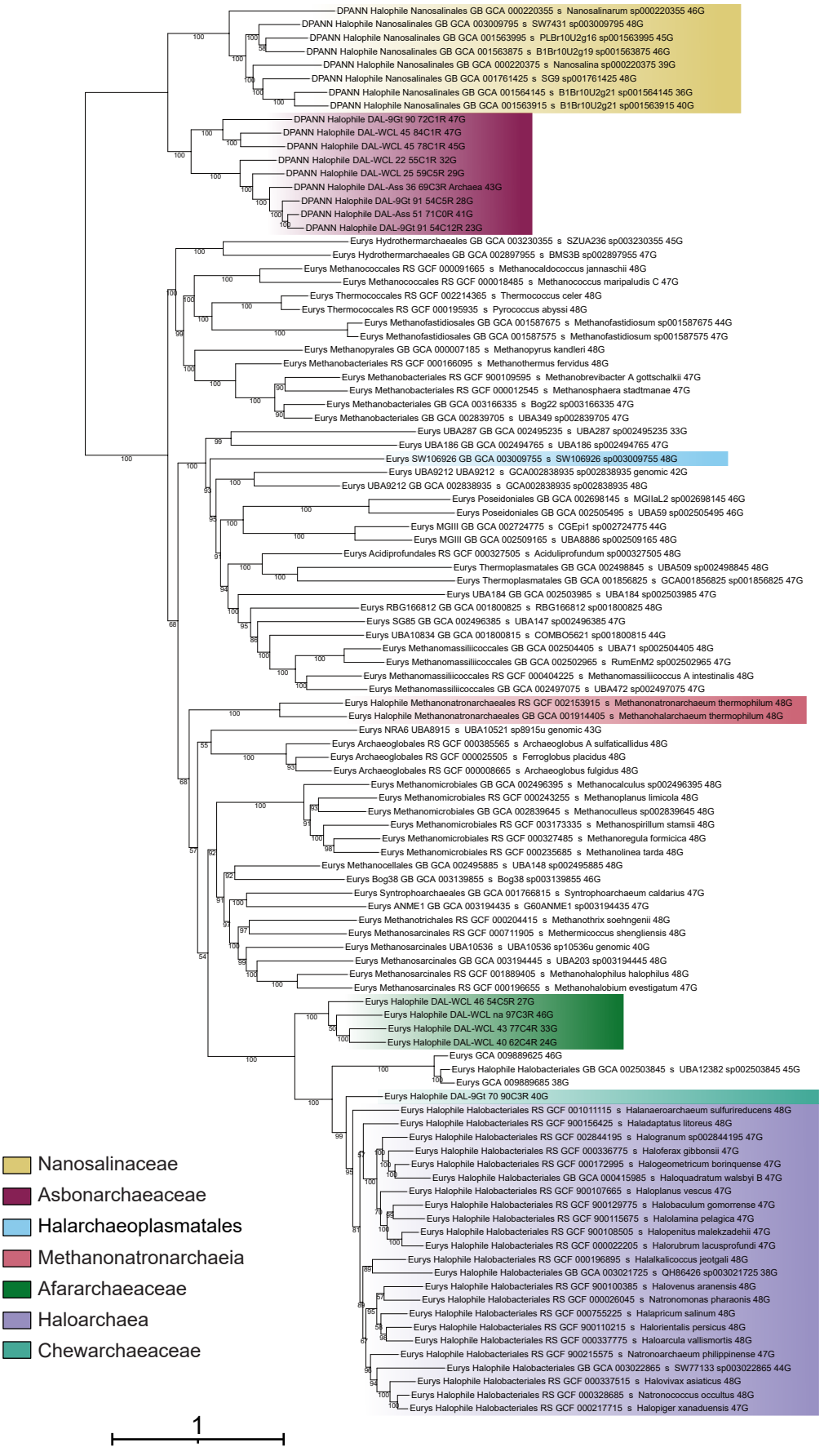


Supplementary Figure 5 | Maximum likelihood phylogeny of 87 archaeal taxa based on the concatenated alignment of 136 new marker (NM) protein dataset. The ML tree was constructed using the LG+C60+F+Γ4 model of evolution. Branch support was assessed using 1,000 ultrafast bootstraps via IQ-TREE implementation. The scale bar represents the estimated number of substitutions per site. Each tip label provides information about the archaeal supergroup, taxonomic order based on GTDB, GTDB accession number, species identification, and the number of markers identified for each taxon out of the 136 NM proteins.

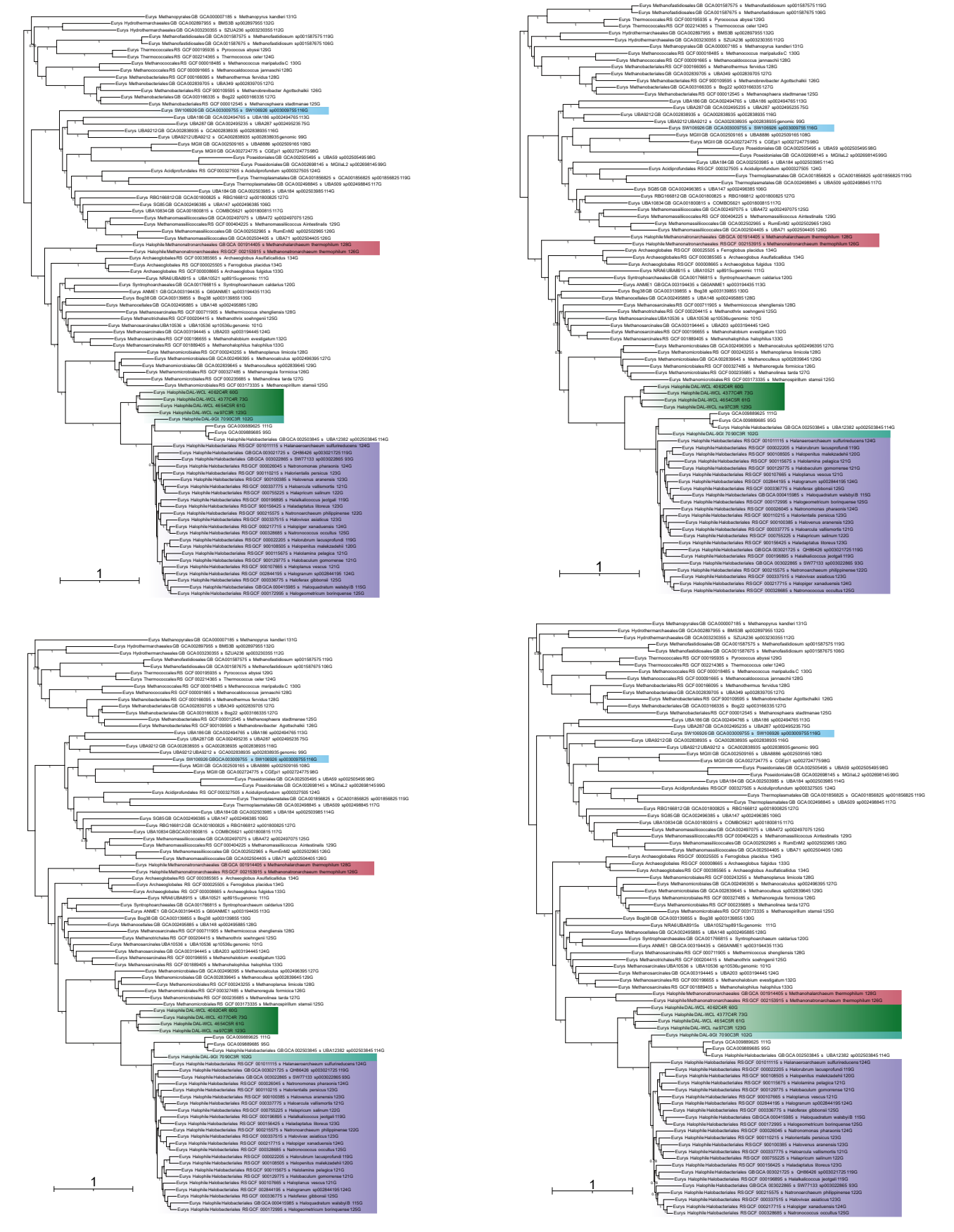


Supplementary Figure 6 | Maximum likelihood phylogeny of 87 archaeal taxa based on the concatenated alignment of 48 ribosomal protein (RP) dataset. The ML tree was constructed using the LG+C60+F+Γ4 model of evolution. Branch support was assessed using 1000 ultrafast bootstraps via IQ-TREE implementation. The scale bar represents the estimated number of substitutions per site. Each tip label provides information about the archaeal super-group, taxonomic order based on GTDB, GTDB accession number, species identification, and the number of markers identified for each taxon out of the 48 RP proteins.

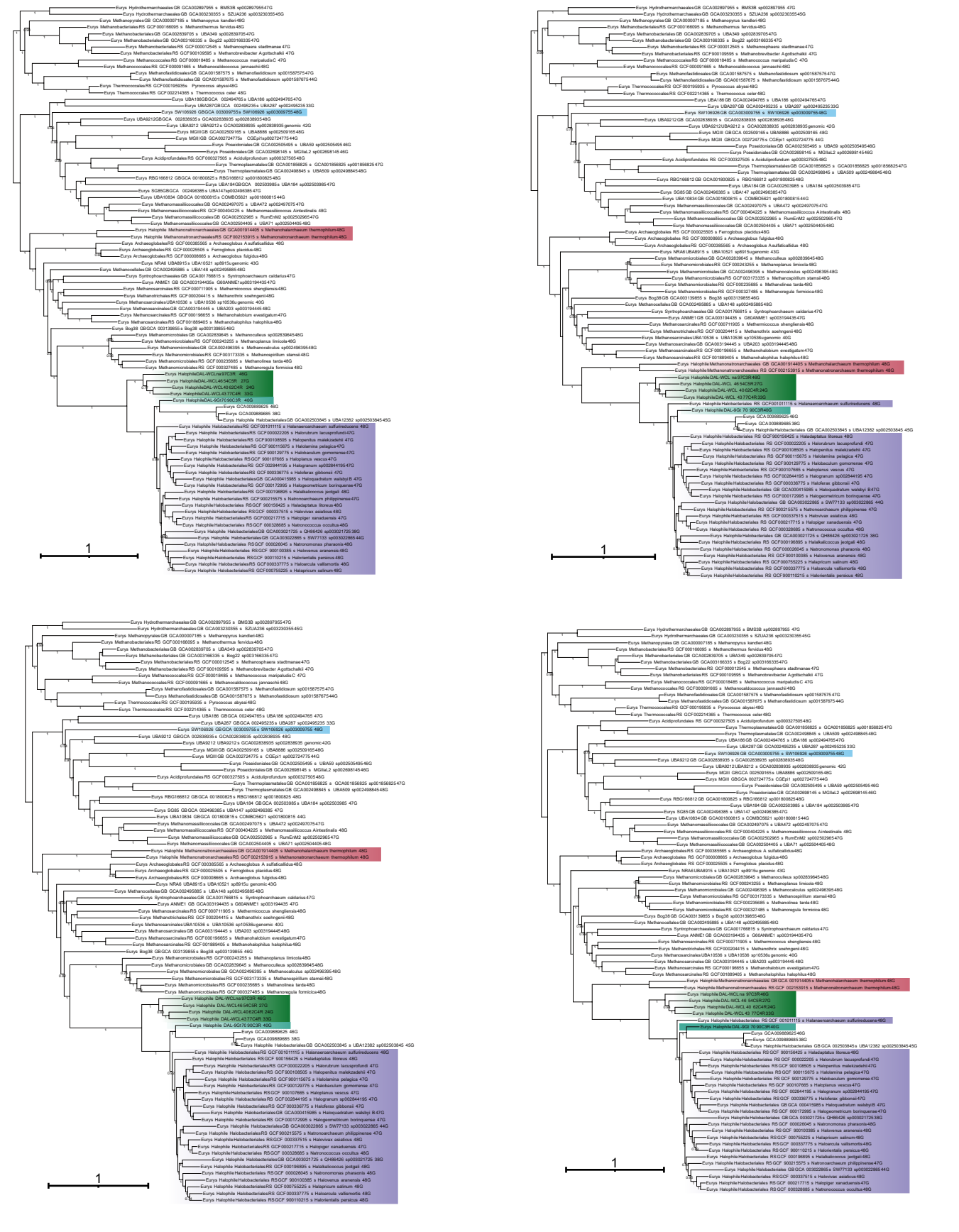
Supplementary Figure 7 | Maximum likelihood phylogeny of 104 archaeal taxa based on the concatenated alignment of 136 new marker (NM) protein dataset. The ML tree was constructed using the LG+C60+F+Γ4 model of evolution. Branch support was assessed using 1000 ultrafast bootstraps via IQ-TREE implementation. The scale bar represents the estimated number of substitutions per site. Each tip label provides information about the archaeal supergroup, taxonomic order based on GTDB, GTDB accession number, species identification, and the number of markers identified for each taxon out of the 136 NM proteins.



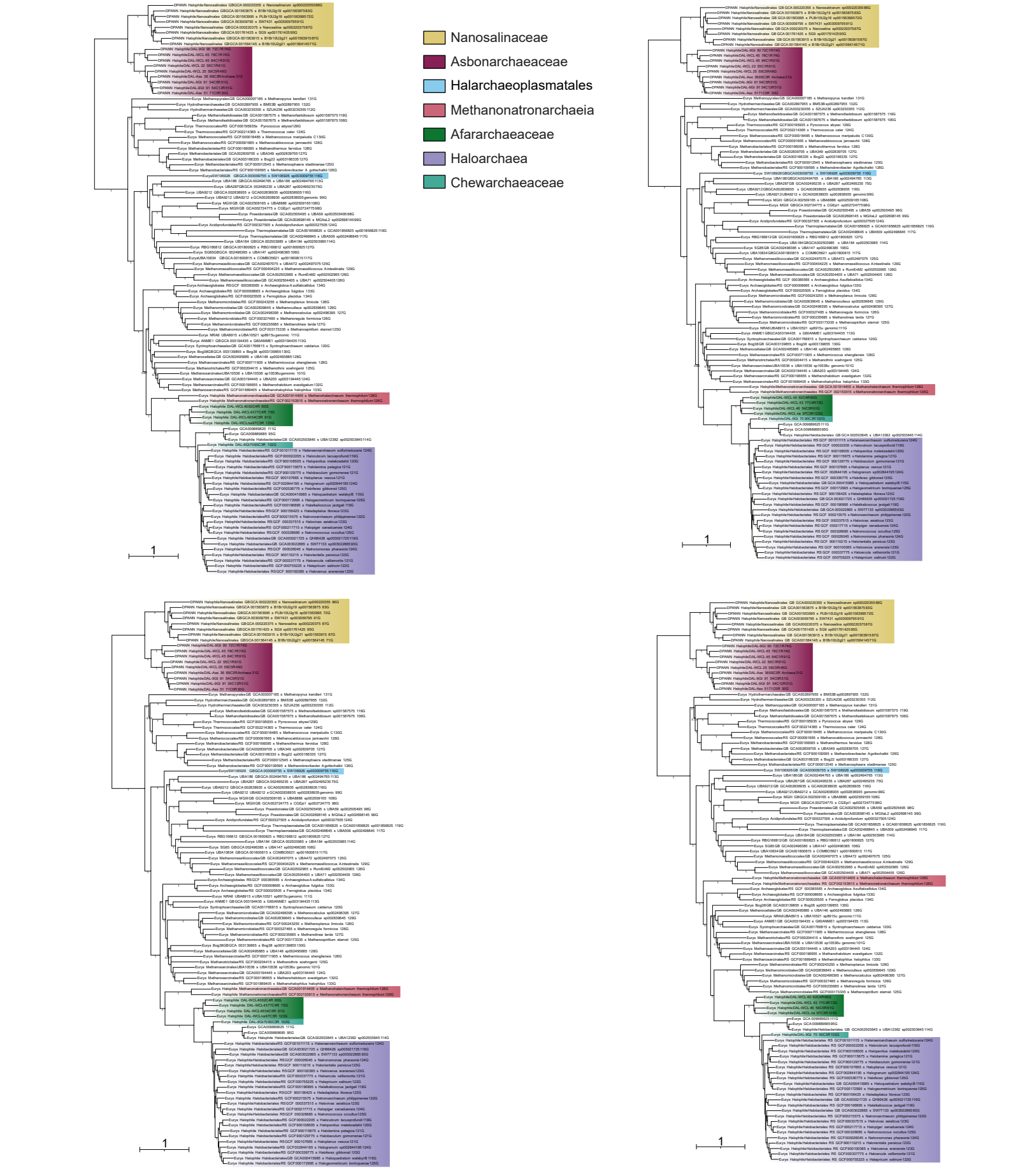
Supplementary Figure 8 | Maximum likelihood phylogeny of 104 archaeal taxa based on the concatenated alignment of 48 ribosomal protein (RP) dataset. The ML tree was constructed using the LG+C60+F+Γ4 model of evolution. Branch support was assessed using 1000 ultrafast bootstraps via IQ- TREE implementation. The scale bar represents the estimated number of substitutions per site. Each tip label provides information about the archaeal super-group, taxonomic order based on GTDB, GTDB accession number, species identification, and the number of markers identified for each taxon out of the 48 RP proteins.



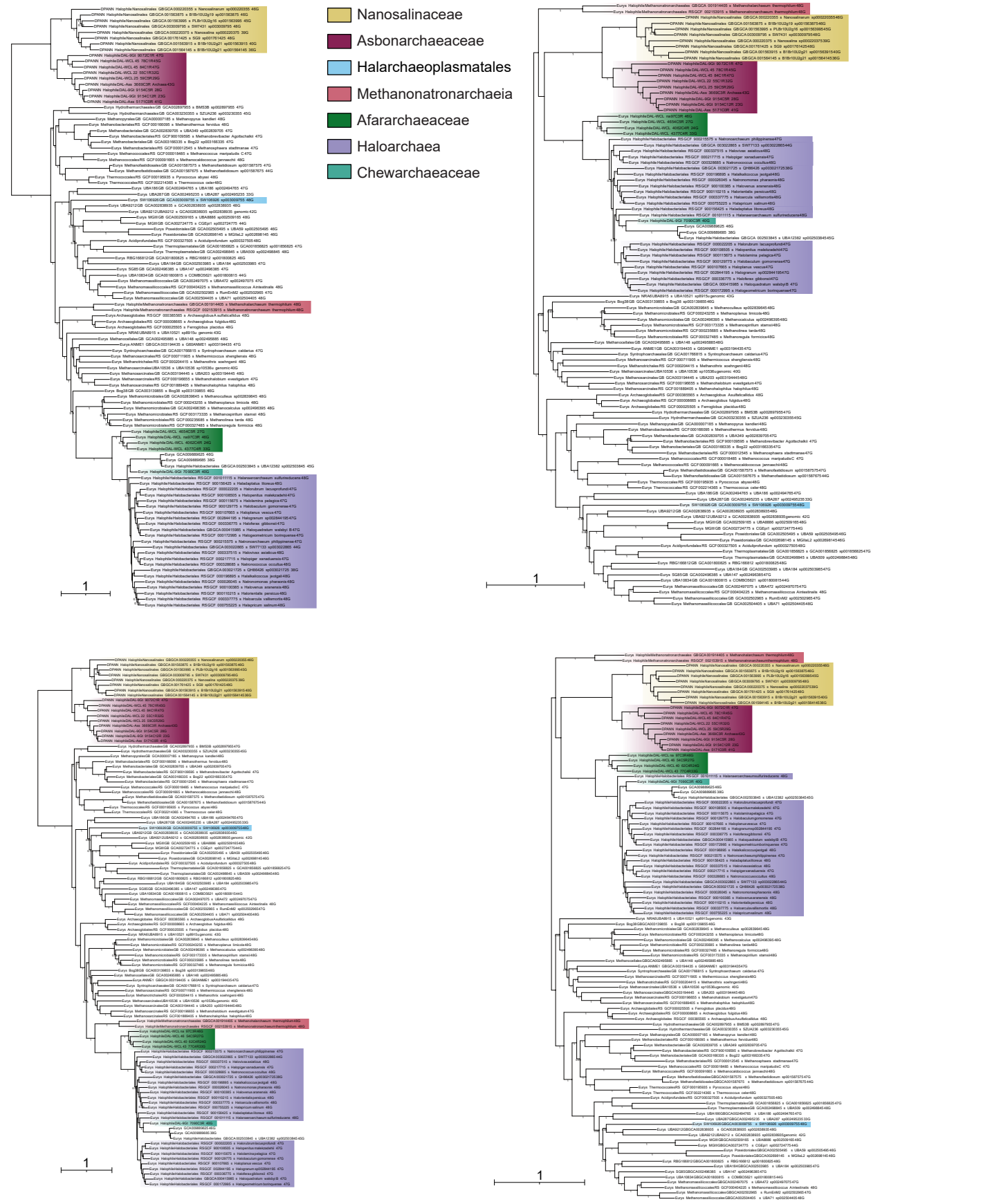
Supplementary Figure 9 | Consensus tree for each of the four MCMC chains for the 87-NM dataset. Four independent Markov chain Monte Carlo (MCMC) chains were inferred using PhyloBayes (CAT+GTR, 15,000 generations with a burn-in of 3,000). Support at branches corresponds to posterior probabilities estimated post-burnin. The scale bar represents the estimated number of substitutions per site. Different halophilic clades are visually represented with distinct colors: halarchaeoaplasmatales (cyan), methanonatronarchaea (rose), afararchaea (green), chewarchaea (teal), and haloarchaea (violet). Each tip label provides information about the archaeal supergroup, taxonomic order based on GTDB, GTDB accession number, species identification, and the number of markers identified for each taxon out of the 136 NM proteins.



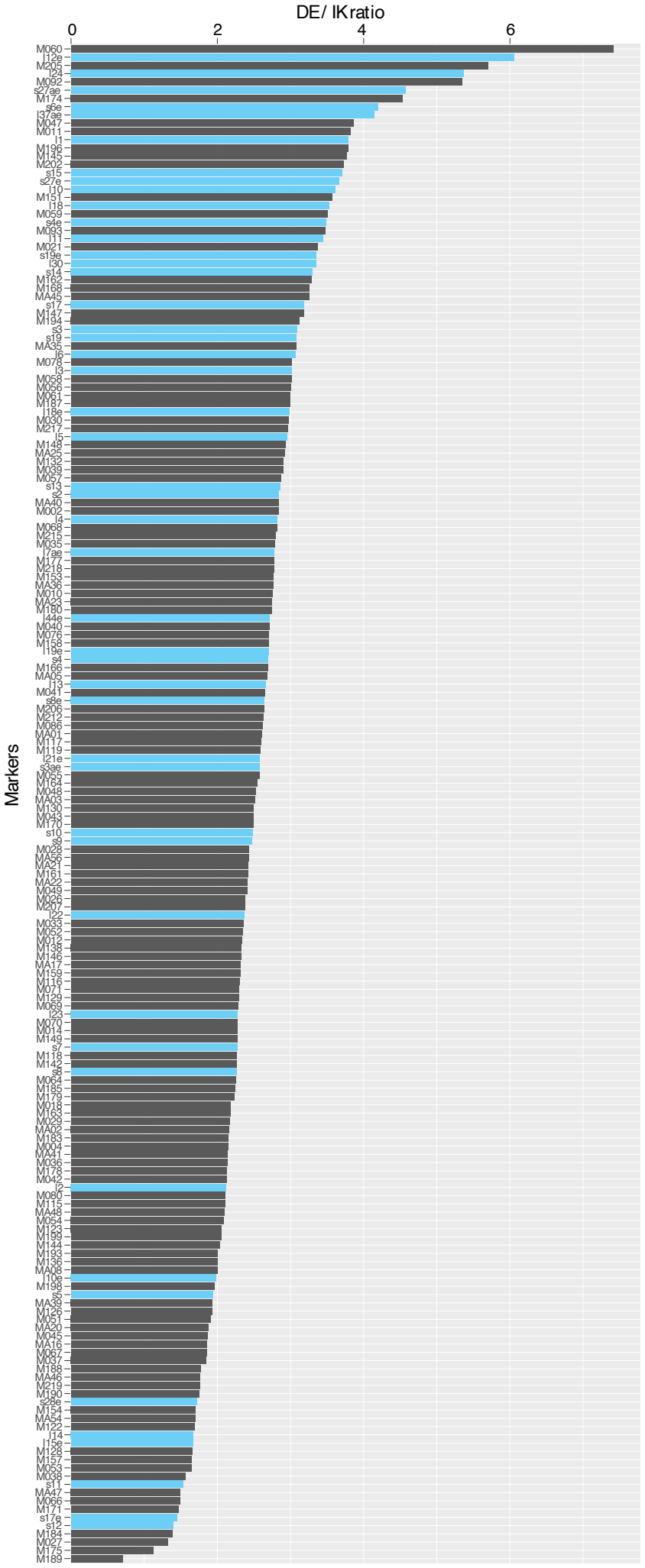
Supplementary Figure 10 | Consensus tree for each of the four MCMC chains for the 87-RP dataset. Four independent Markov chain Monte Carlo (MCMC) chains were inferred using PhyloBayes (CAT+GTR, 15,000 generations with a burn-in of 3,000). Support at branches corresponds to posterior probabilities estimated post-burnin. The scale bar represents the estimated number of substitutions per site. Different halophilic clades are visually represented with distinct colors: halarchaeoaplasmatales (cyan), methanonatronarchaea (rose), afararchaea (green), chewarchaea (teal), and haloarchaea (violet). Each tip label provides information about the archaeal super group, taxonomic order based on GTDB, GTDB accession number, species identification, and the number of markers identified for each taxon out of the 48 RP proteins.



Supplementary Figure 11 | Consensus tree for each of the four MCMC chains for the 104-NM dataset. Four independent Markov chain Monte Carlo (MCMC) chains were inferred using PhyloBayes (CAT+GTR, 15,000 generations with a burn-in of 3,000). Support at branches corresponds to posterior probabilities estimated post-burnin. The scale bar represents the estimated number of substitutions per site. Different halophilic clades are visually represented with distinct colors. Each tip label provides information about the archaeal supergroup, taxonomic order based on GTDB, GTDB accession number, species identification, and the number of markers identified for each taxon out of the 136 NM proteins.

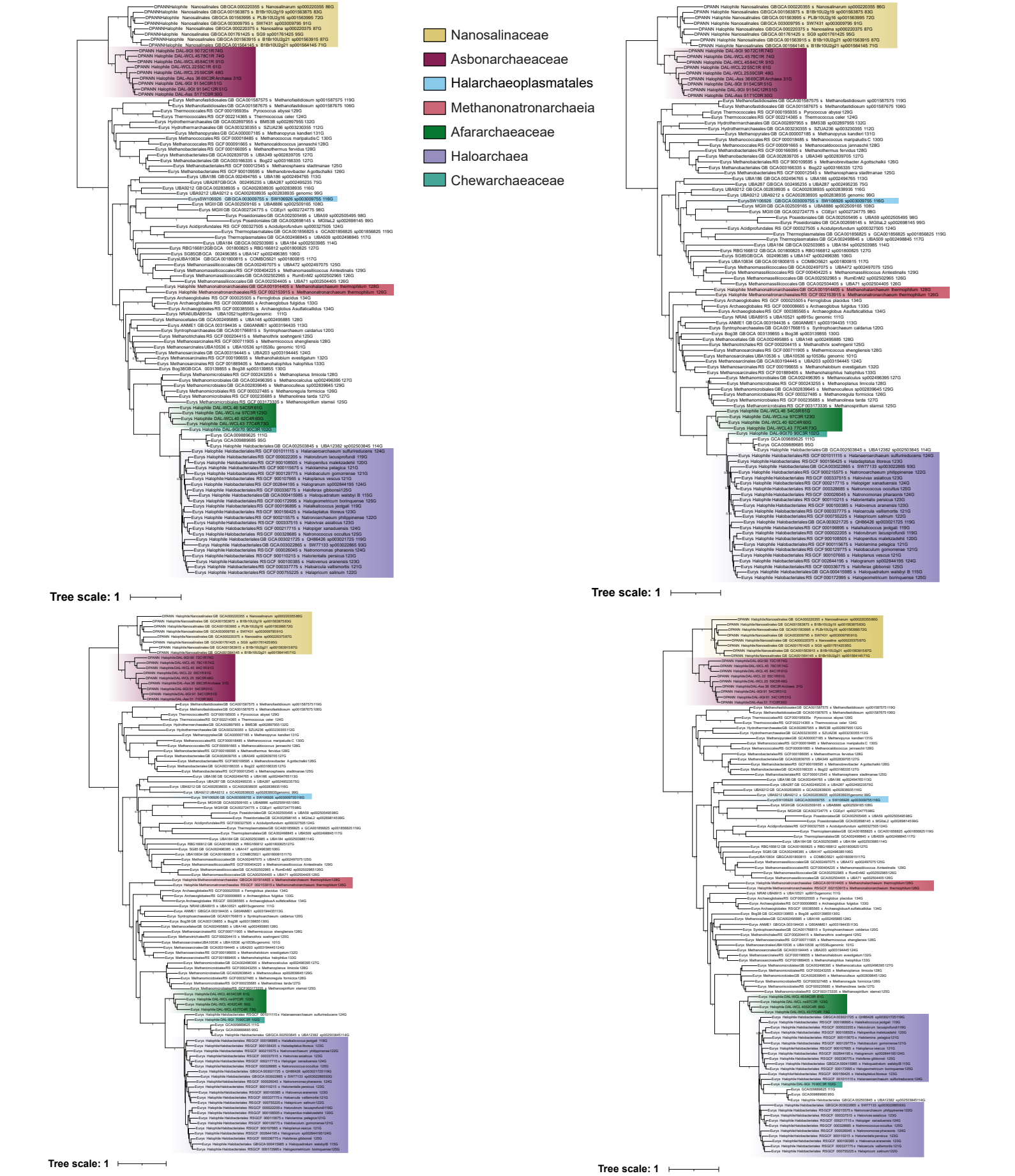


Supplementary Figure 12 | Consensus tree for each of the four MCMC chains for the 104-RP dataset. Four independent Markov chain Monte Carlo (MCMC) chains were inferred using PhyloBayes (CAT+GTR, 15,000 generations with a burn-in of 3,000). Support at branches corresponds to posterior probabilities estimated post-burnin. The scale bar represents the estimated number of substitutions per site. Different halophilic clades are visually represented with distinct colors. Each tip label provides information about the archaeal super-group, taxonomic order based on GTDB, GTDB accession number, species identification, and the number of markers identified for each taxon out of the 48 RP proteins.

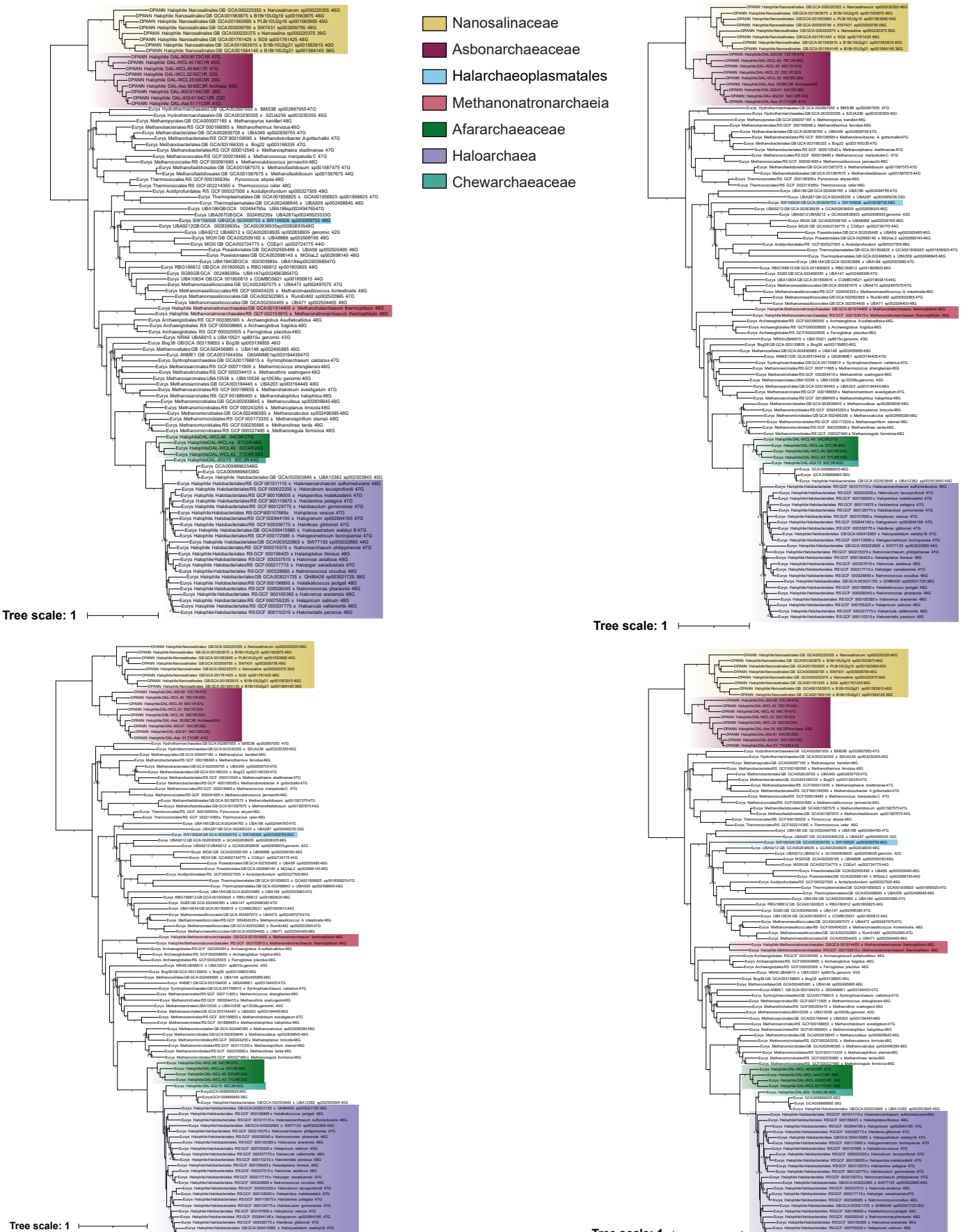


Supplementary Figure 13 | RP (blue) and NM (grey) markers ranked from the most to least biased based on the calculated DE/IK ratio of halophiles versus the DE/IK ratio of non-halophiles.

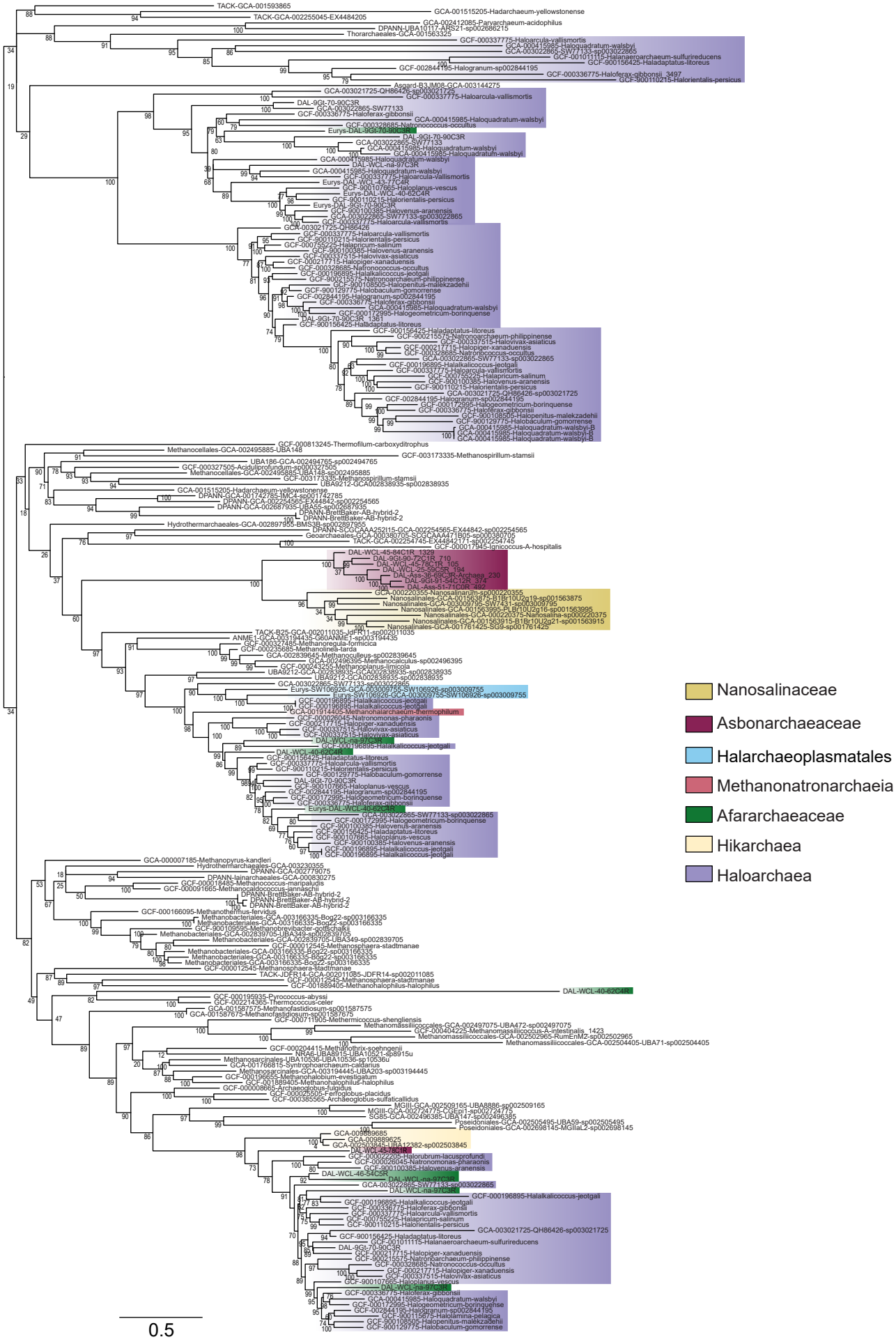
Supplementary Figure 14 | Maximum likelihood phylogeny based on concatenating the 18 most biased ribosomal proteins. The ML tree was constructed using the LG+C60+F+Γ4 model of evolution. Branch support was assessed using 1,000 ultrafast bootstraps via IQ-TREE implementation. Different halophilic clades are visually represented with distinct colors. The scale bar represents the estimated number of substitutions per site. Each tip label provides information about the archaeal super-group, taxonomic order based on GTDB, GTDB accession number, species identification, and the number of markers identified for each taxon out of the 18 RP proteins.



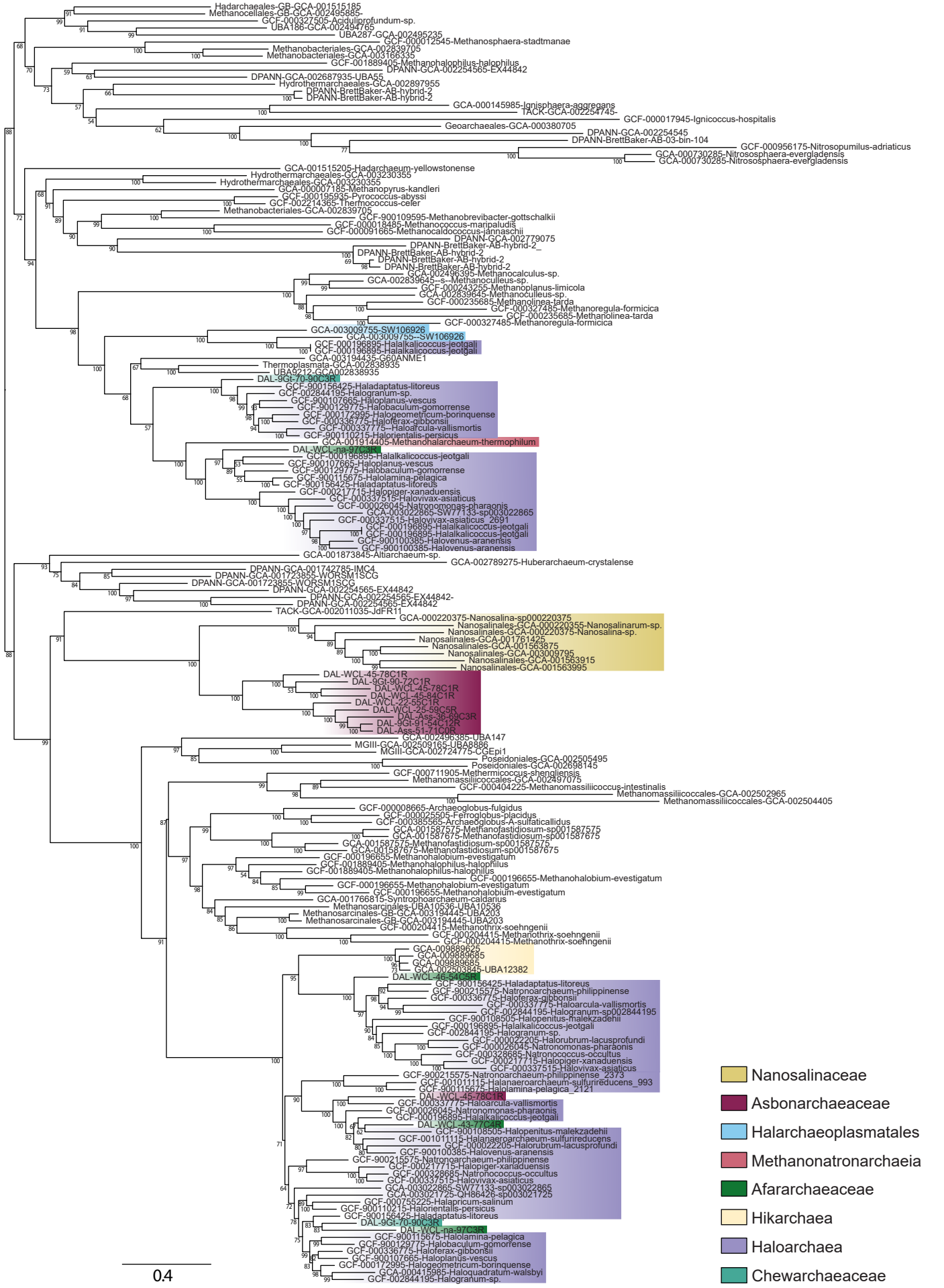
Supplementary Figure 15 | Consensus tree for each of the four MCMC chains for the 104-RP dataset with 20% of most biased sites removed. Four independent Markov chain Monte Carlo (MCMC) chains were inferred using PhyloBayes (CAT+GTR, 15,000 generations with a burn-in of 3,000). Support at branches corresponds to posterior probabilities estimated post-burnin. The scale bar represents the estimated number of substitutions per site. Different halophilic clades are visually represented with distinct colors. Each tip label provides information about the archaeal super group, taxonomic order based on GTDB, GTDB accession number, species identification, and the number of markers identified for each taxon out of the 48 RP proteins.



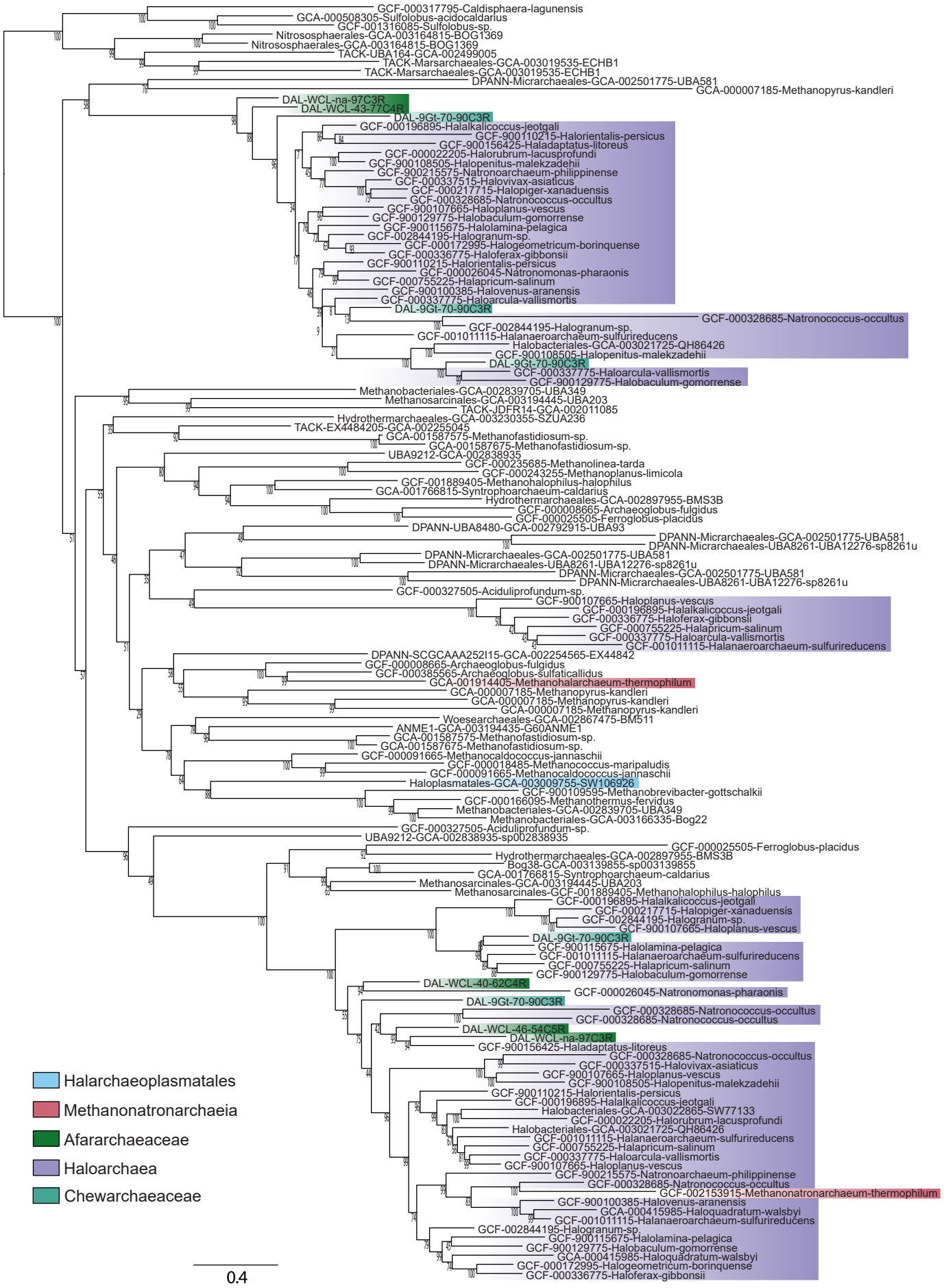
Supplementary Figure 16 | Consensus tree for each of the four MCMC chains for the 104-NM dataset with 20% of most biased sites removed. Four independent Markov chain Monte Carlo (MCMC) chains were inferred using PhyloBayes (CAT+GTR, 15,000 generations with a burn-in of 3,000 generations). Support at branches corresponds to posterior probabilities estimated post-burnin. The scale bar represents the estimated number of substitutions per site. Different halophilic clades are visually represented with distinct colors. Each tip label provides information about the archaeal supergroup, taxonomic order based on GTDB, GTDB accession number, species identification, and the number of markers identified for each taxon out of the 136 NM proteins.



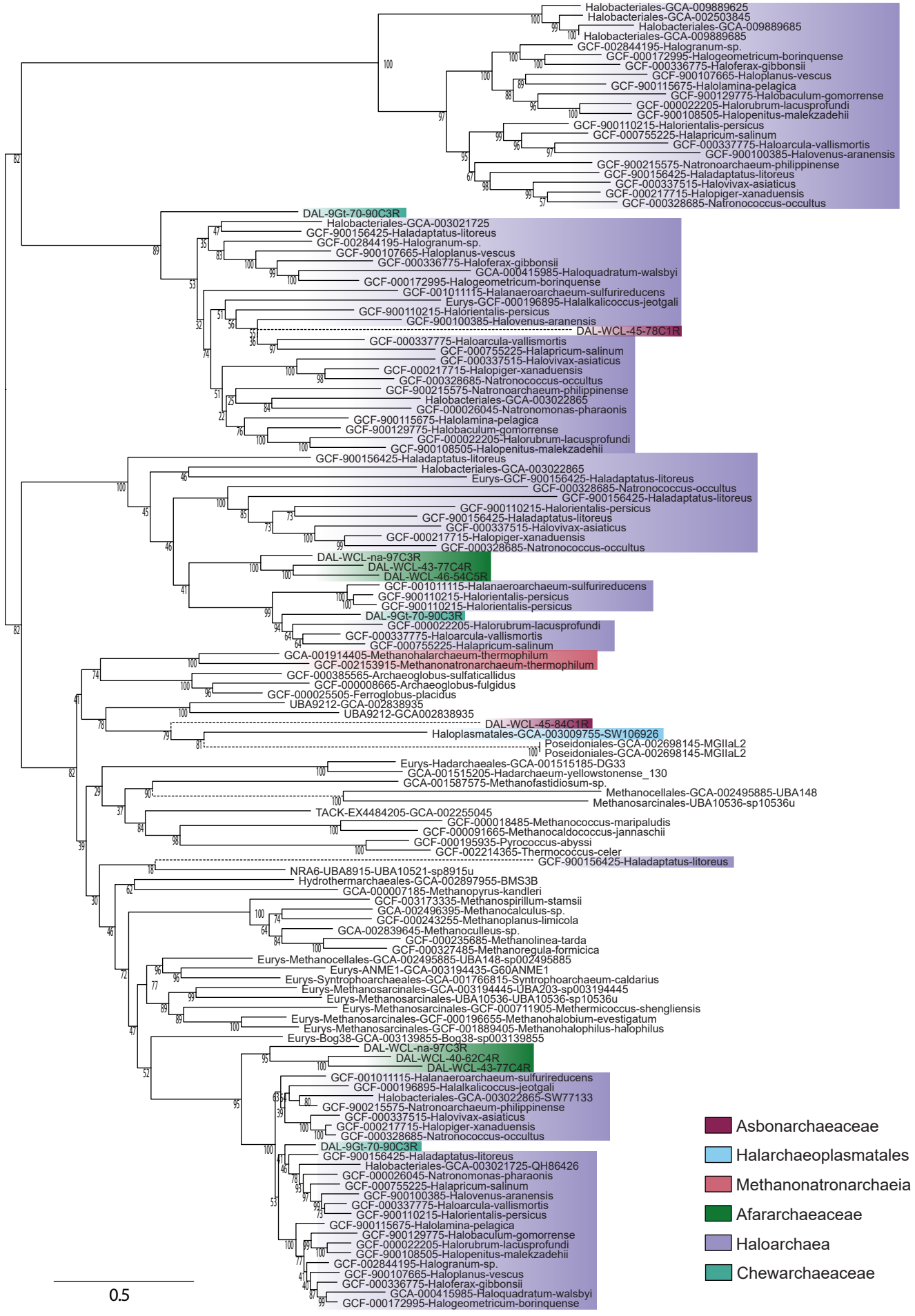
Supplementary Fig. 17 | Maximum likelihood phylogenetic tree of TrkA K+ transporter. The tree was constructed with the LG+C60+F+Γ4 model of sequence evolution. Branch support was assessed using 1,000 ultrafast bootstraps. The scale bar represents the number of substitutions per position.



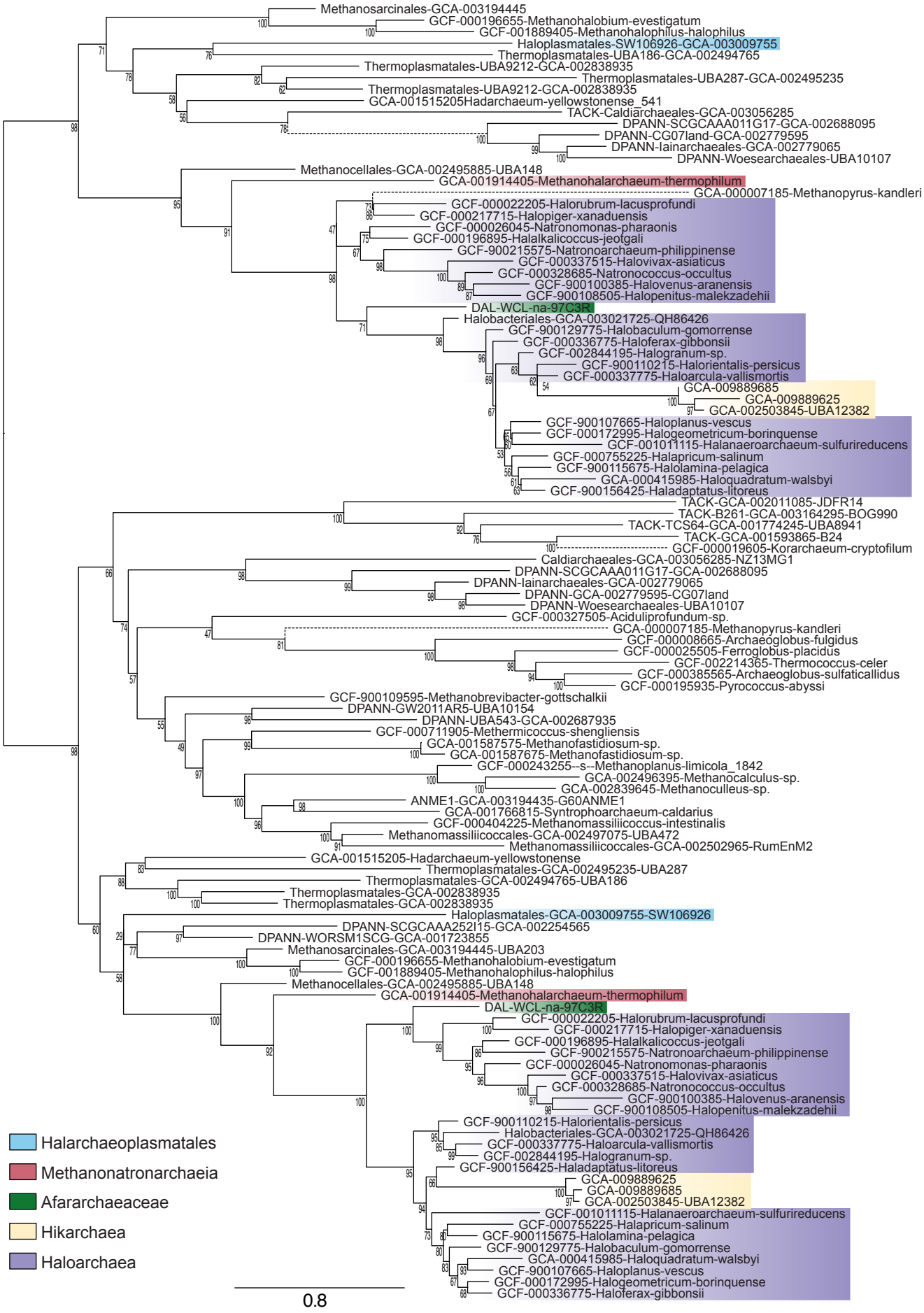
Supplementary Fig. 18 | Maximum likelihood phylogenetic tree of TrkH K+ transporter. The tree was constructed with the LG+C60+F+Γ4 model of sequence evolution. Branch support was assessed using 1,000 ultrafast bootstraps. The scale bar represents the number of substitutions per position.



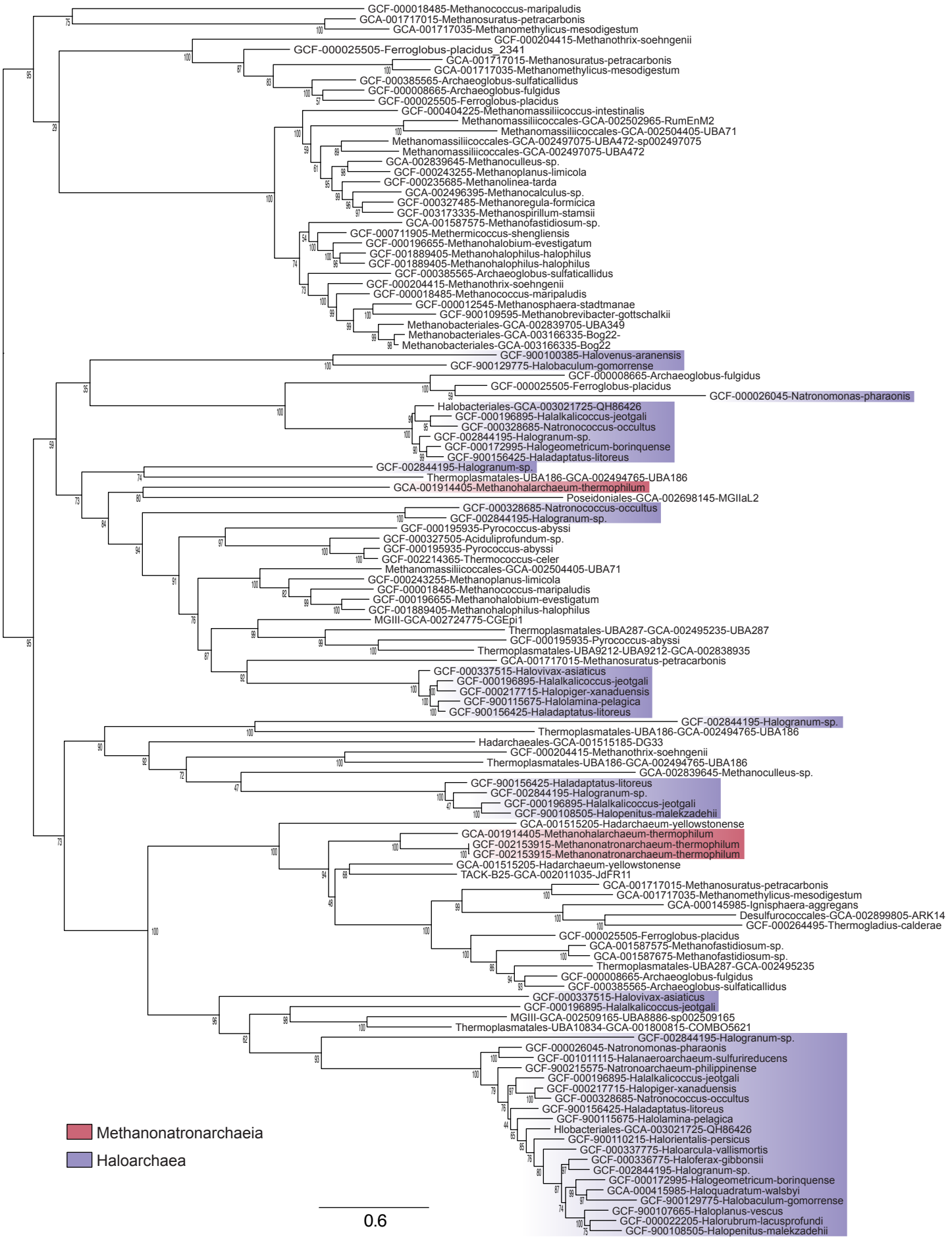
Supplementary Fig. 19 | Maximum likelihood phylogenetic tree of TrkA-N K+ transporter. The tree was constructed with the LG+C60+F+Γ4 model of sequence evolution. Branch support was assessed using 1,000 ultrafast bootstraps. The scale bar represents the number of substitutions per position.



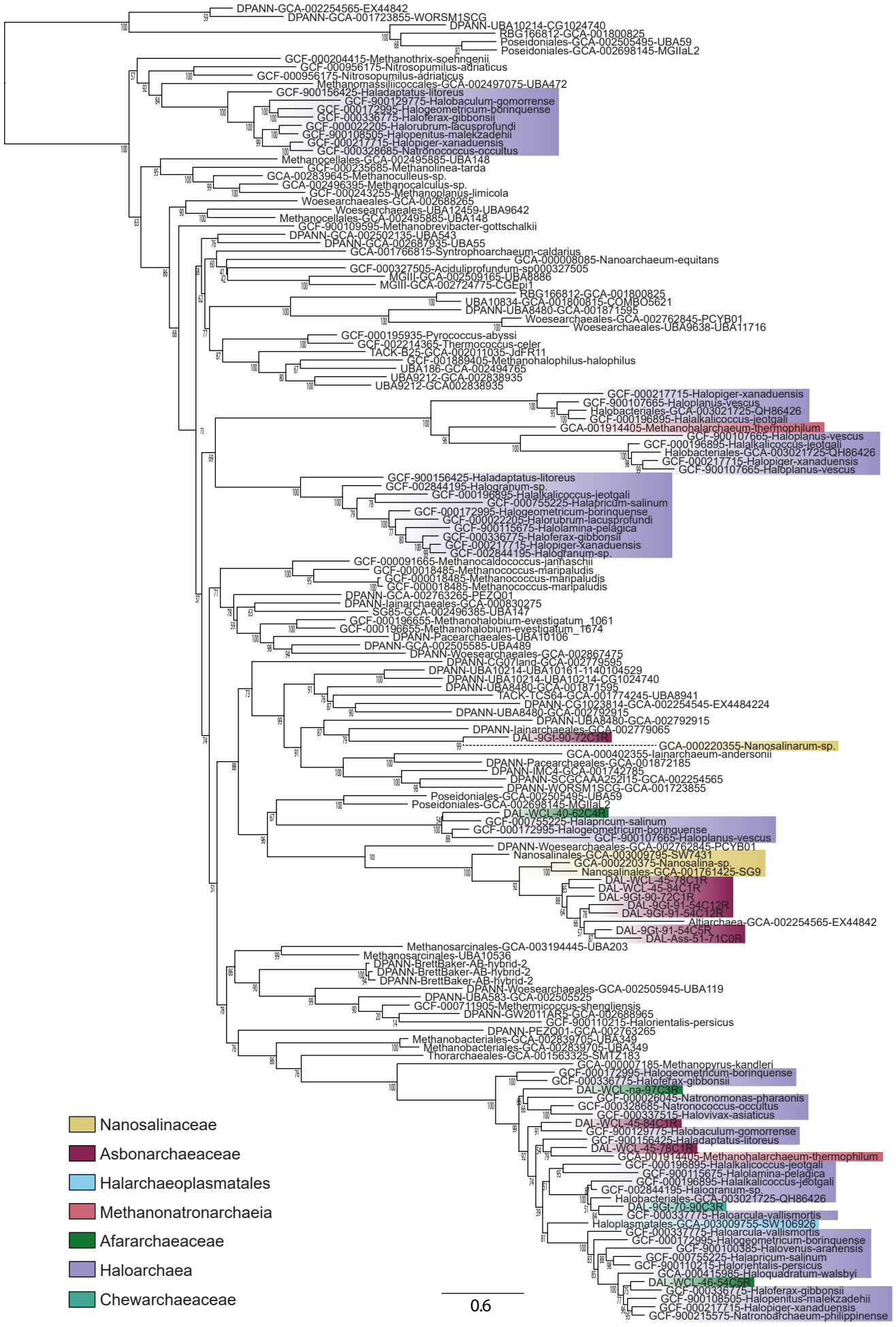
Supplementary Fig. 20 | Maximum likelihood phylogenetic tree of Kef K+ transporters. The tree was constructed with the LG+C60+F+Γ4 model of sequence evolution. Branch support was assessed using 1,000 ultrafast bootstraps. The scale bar represents the number of substitutions per position.



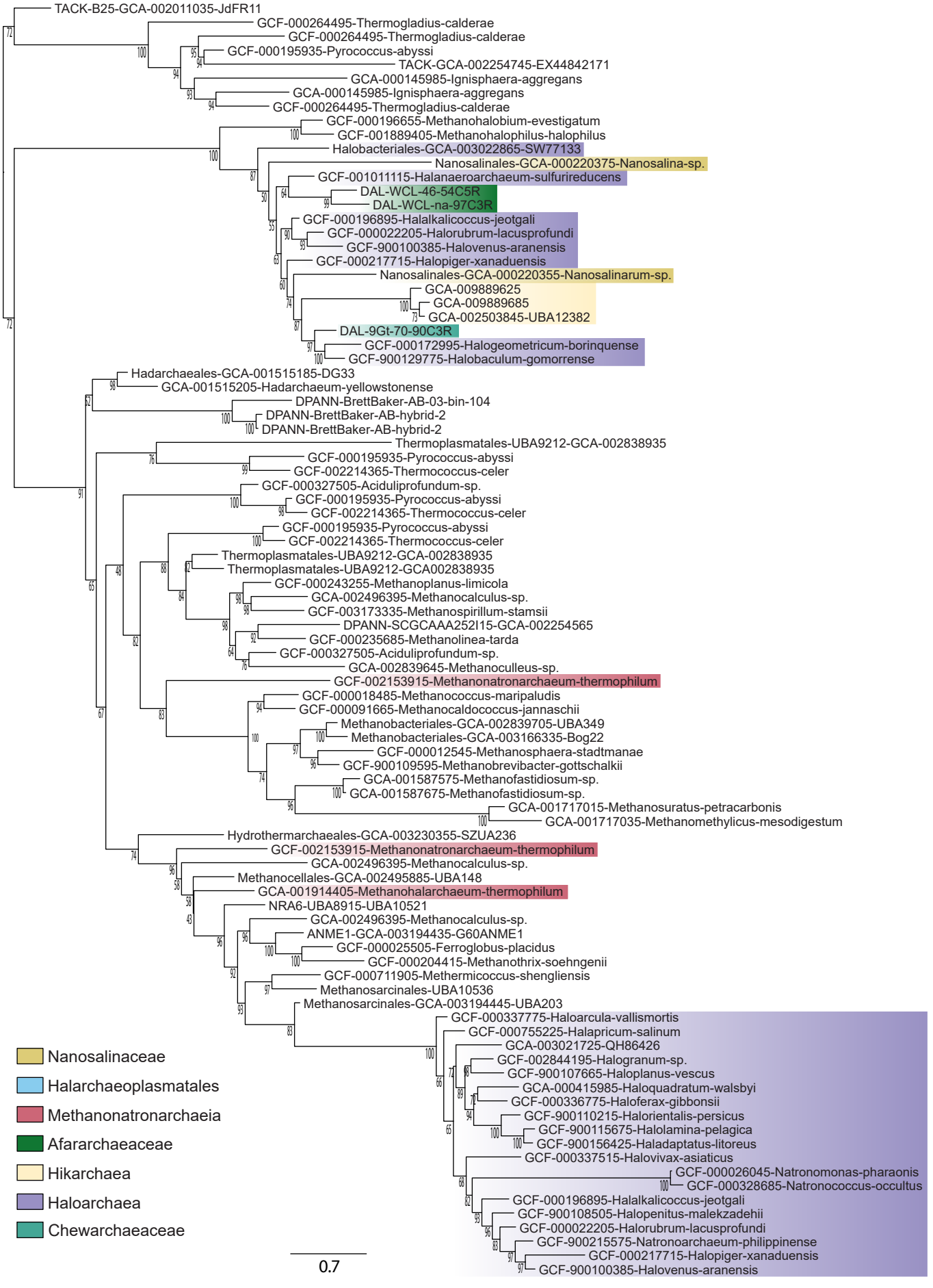
Supplementary Fig. 21 | Maximum likelihood phylogenetic tree of Mg2+ transporters. The tree was constructed with the LG+C60+F+Γ4 model of sequence evolution. Branch support was assessed using 1,000 ultrafast bootstraps. The scale bar represents the number of substitutions per position.



Supplementary Fig. 22 | Maximum likelihood phylogenetic tree of SSF Na+/solute symporters. The tree was constructed with the LG+C60+F+Γ4 model of sequence evolution. Branch support was assessed using 1,000 ultrafast bootstraps. The scale bar represents the number of substitutions per position.

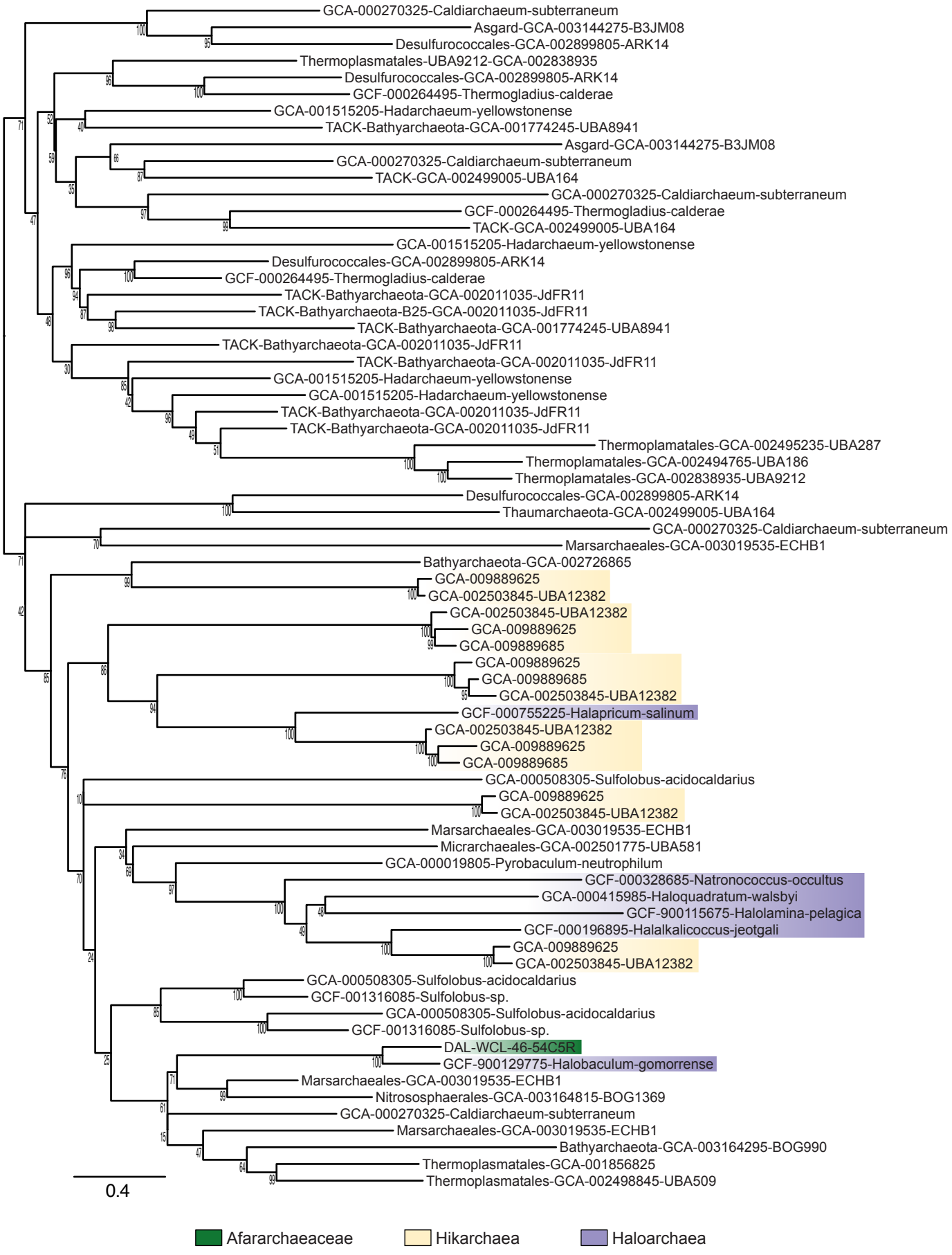


Supplementary Fig. 23 | Maximum likelihood phylogenetic tree of Ca+/Na+ antiporters. The tree was constructed with the LG+C60+F+Γ4 model of sequence evolution. Branch support was assessed using 1,000 ultrafast bootstraps. The scale bar represents the number of substitutions per position.

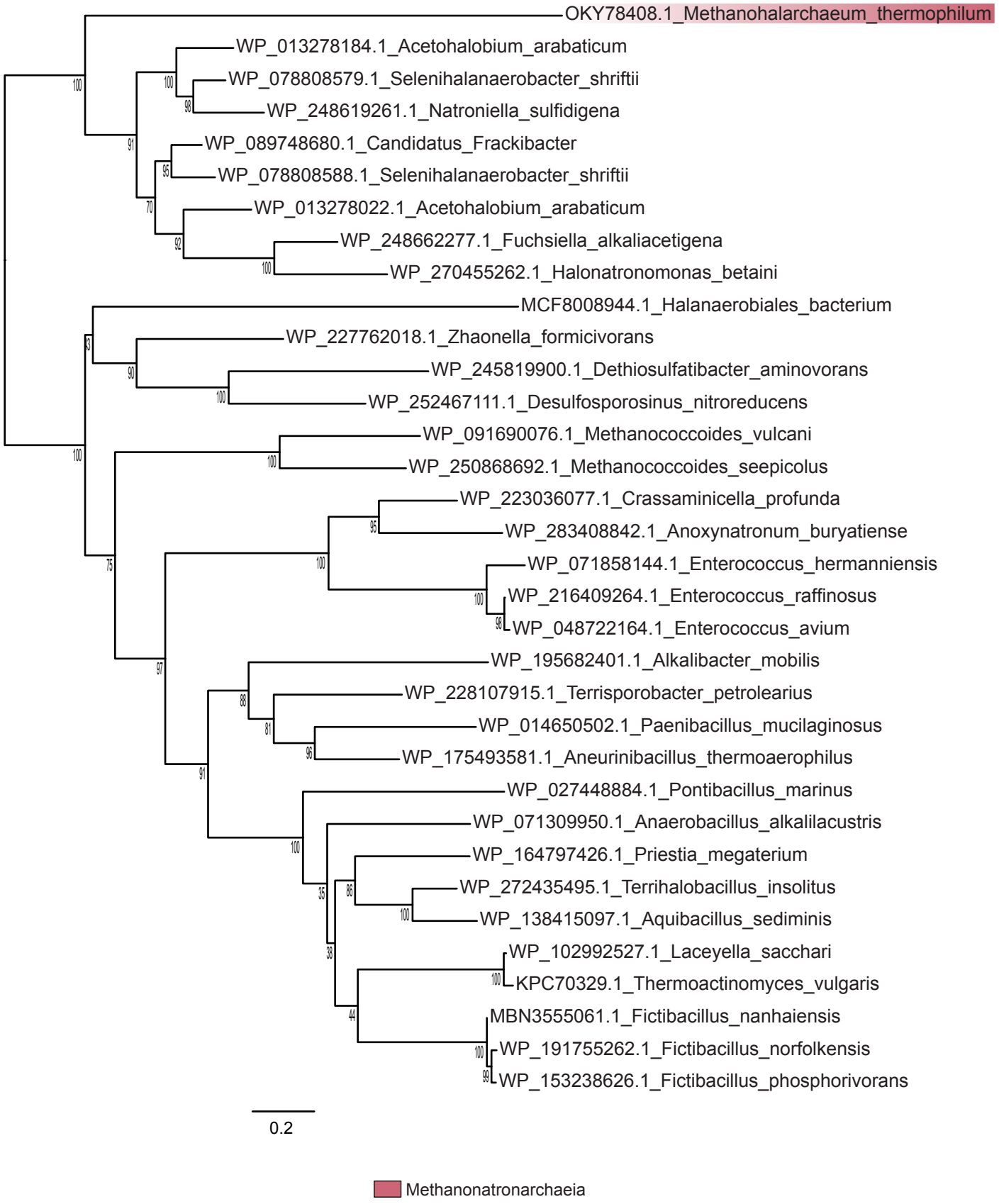


Supplementary Fig. 24 | Maximum likelihood phylogenetic tree of Na+/H+ antiporters. The tree was constructed with the LG+C60+F+Γ4 model of sequence evolution. Branch support was assessed using 1,000 ultrafast bootstraps. The scale bar represents the number of substitutions per position.

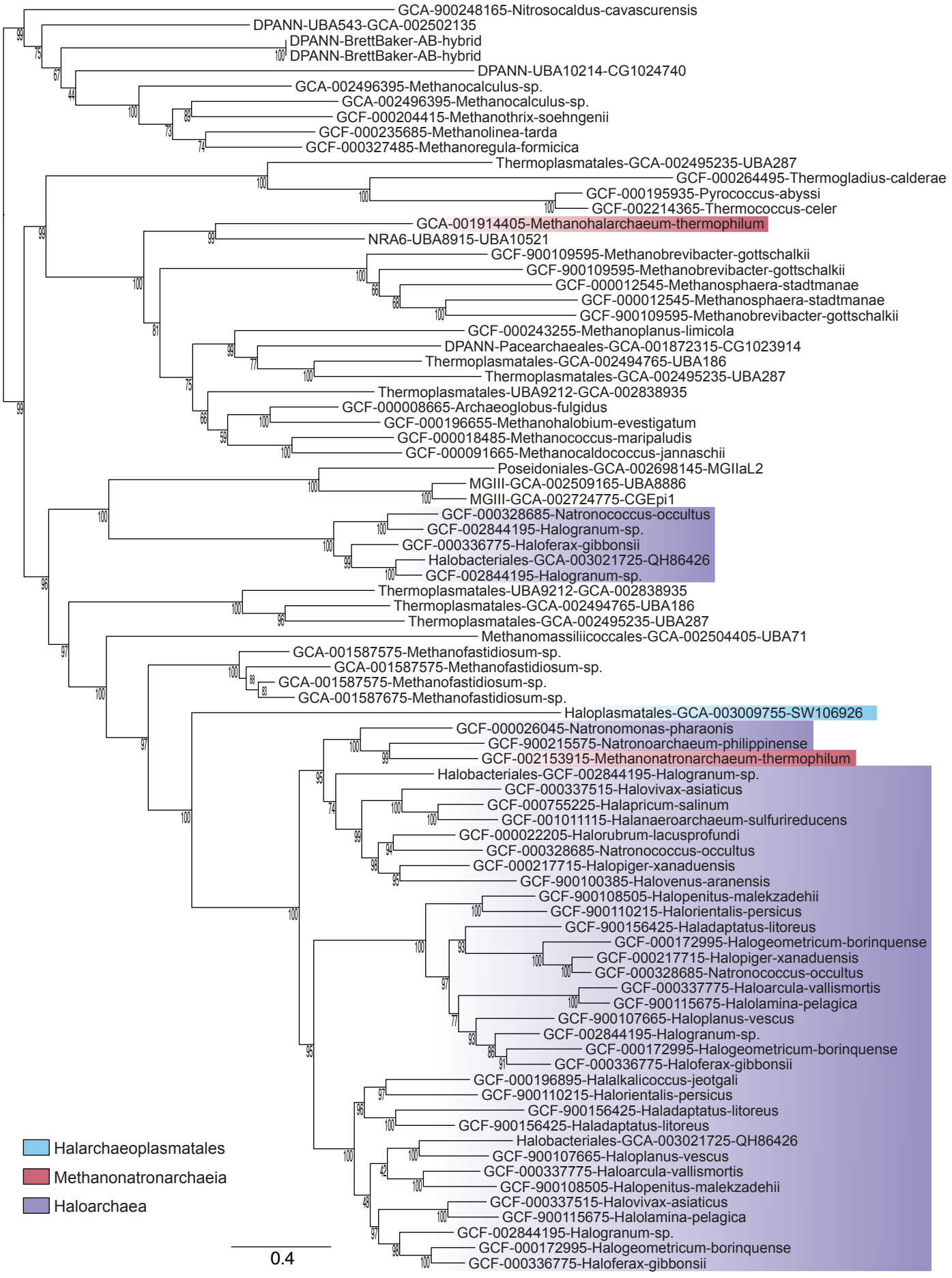
Supplementary Fig. 25 | Maximum likelihood phylogenetic tree of the chaperone GrpE. The tree was constructed with the LG+C60+F+Γ4 model of sequence evolution. Branch support was assessed using 1,000 ultrafast bootstraps. The scale bar represents the number of substitutions per position.



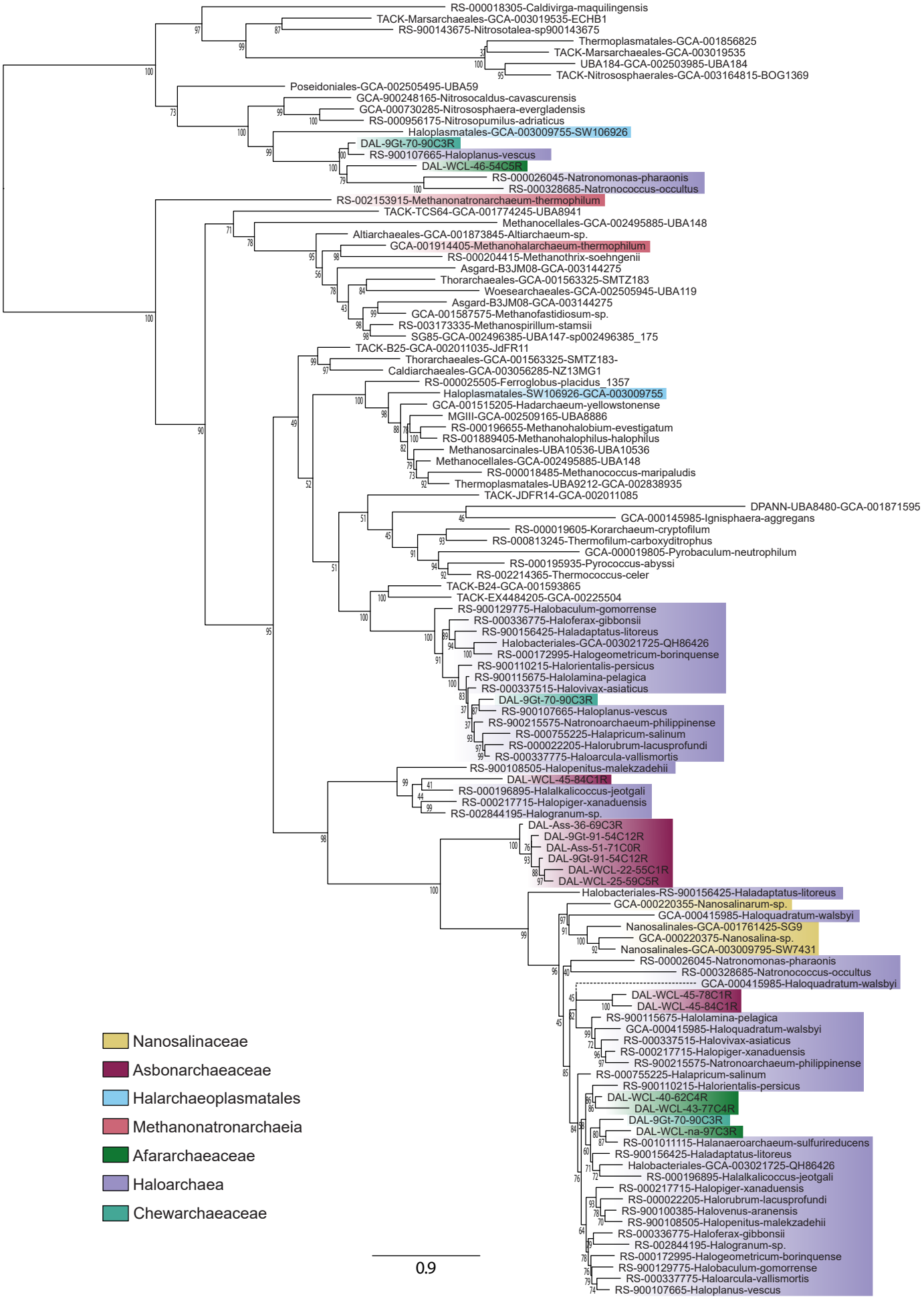
Supplementary Fig. 26 | Maximum likelihood phylogenetic tree of the aerobic-type carbon monoxide dehydrogenase. The tree was constructed with the LG+C60+F+Γ4 model of sequence evolution. Branch support was assessed using 1,000 ultrafast bootstraps. The scale bar represents the number of substitutions per position.



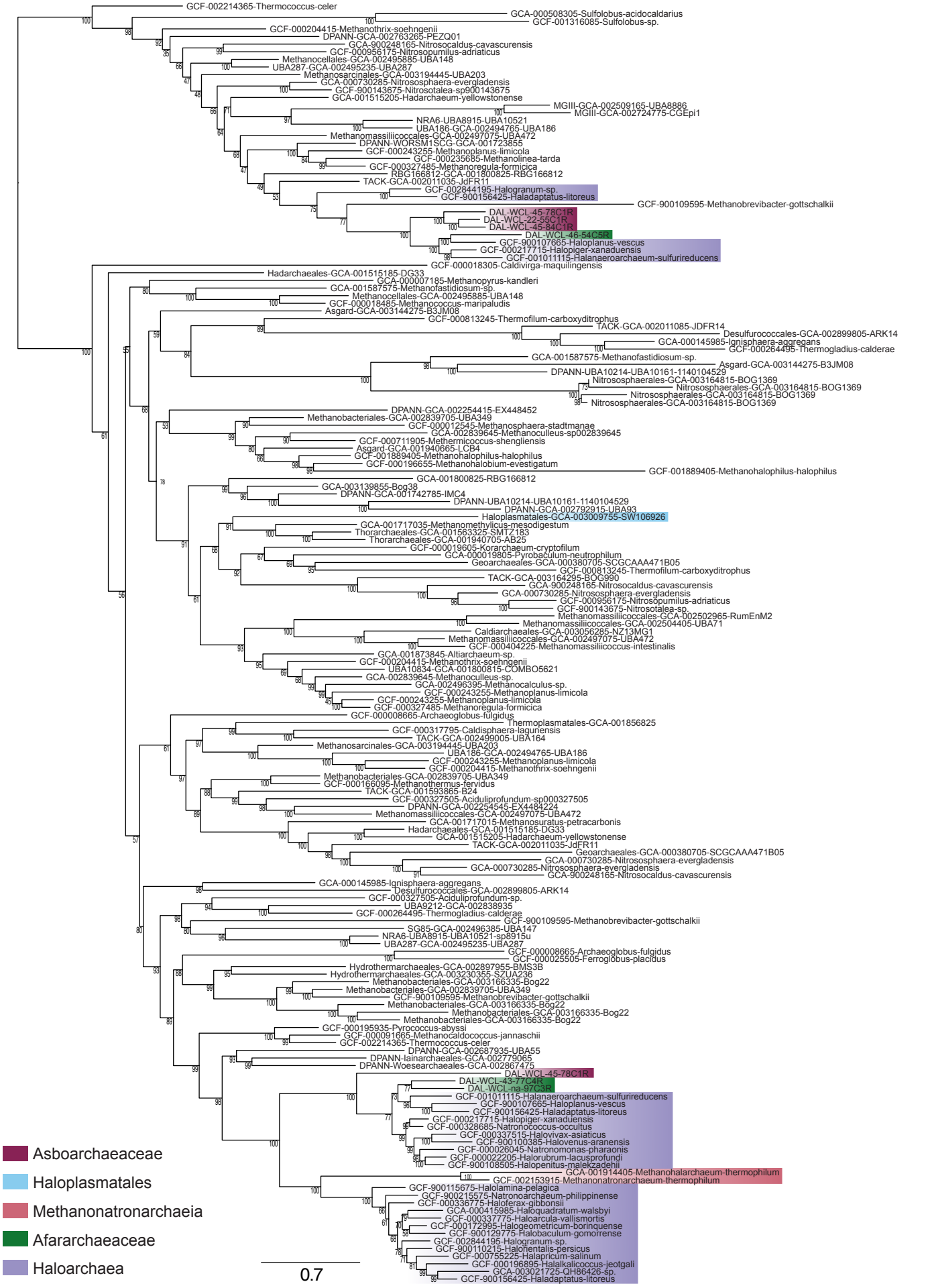
Supplementary Fig. 27 | Maximum likelihood phylogenetic tree of BCCT transporter. The tree was constructed with the LG+C60+F+Γ4 model of sequence evolution. Branch support was assessed using 1,000 ultrafast bootstraps. The scale bar represents the number of substitutions per position.



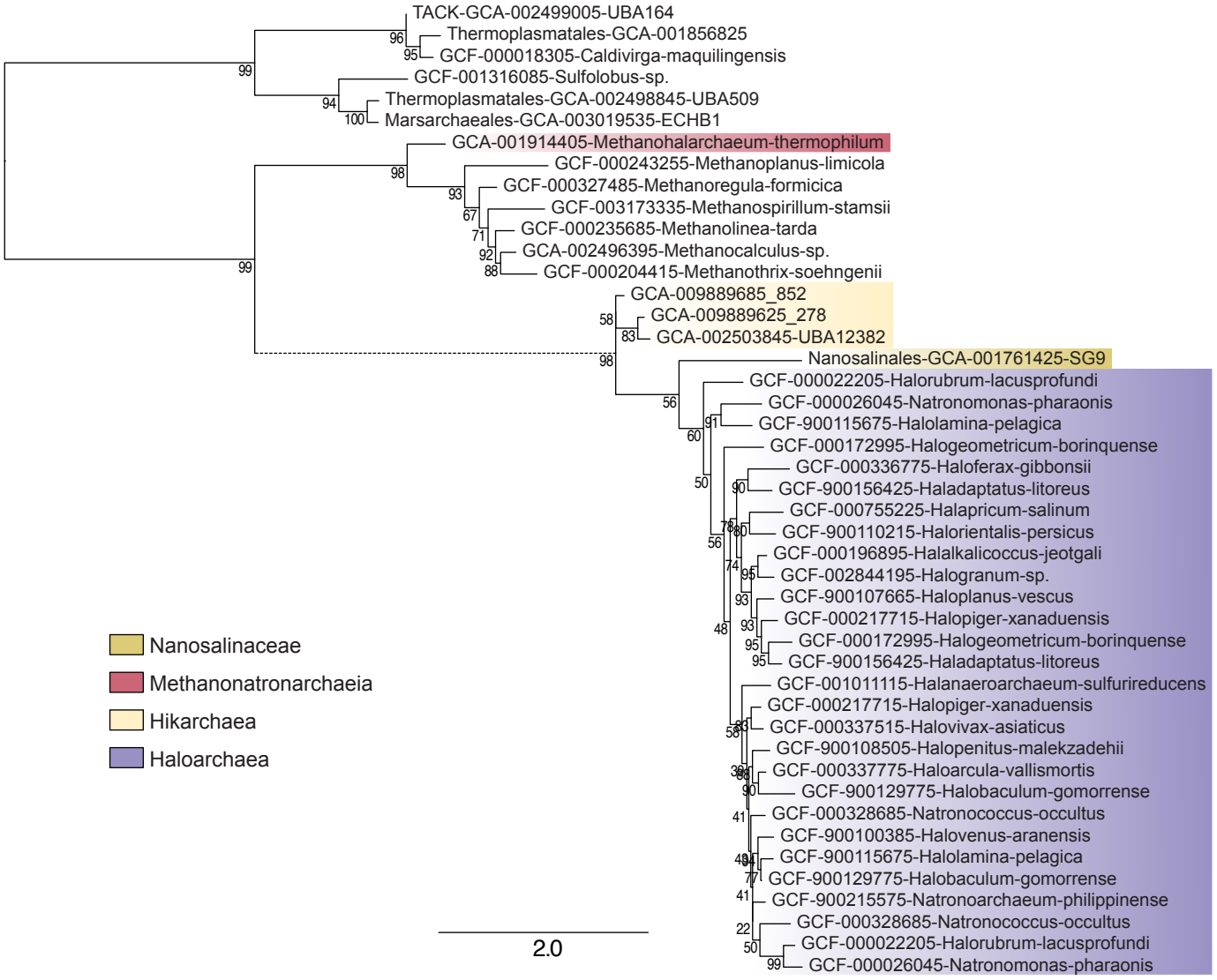
Supplementary Fig. 28 | Maximum likelihood phylogenetic tree of SNF-family Na+-dependent transporters. The tree was constructed with the LG+C60+F+Γ4 model of sequence evolution. Branch support was assessed using 1,000 ultrafast bootstraps. The scale bar represents the number of substitutions per position.



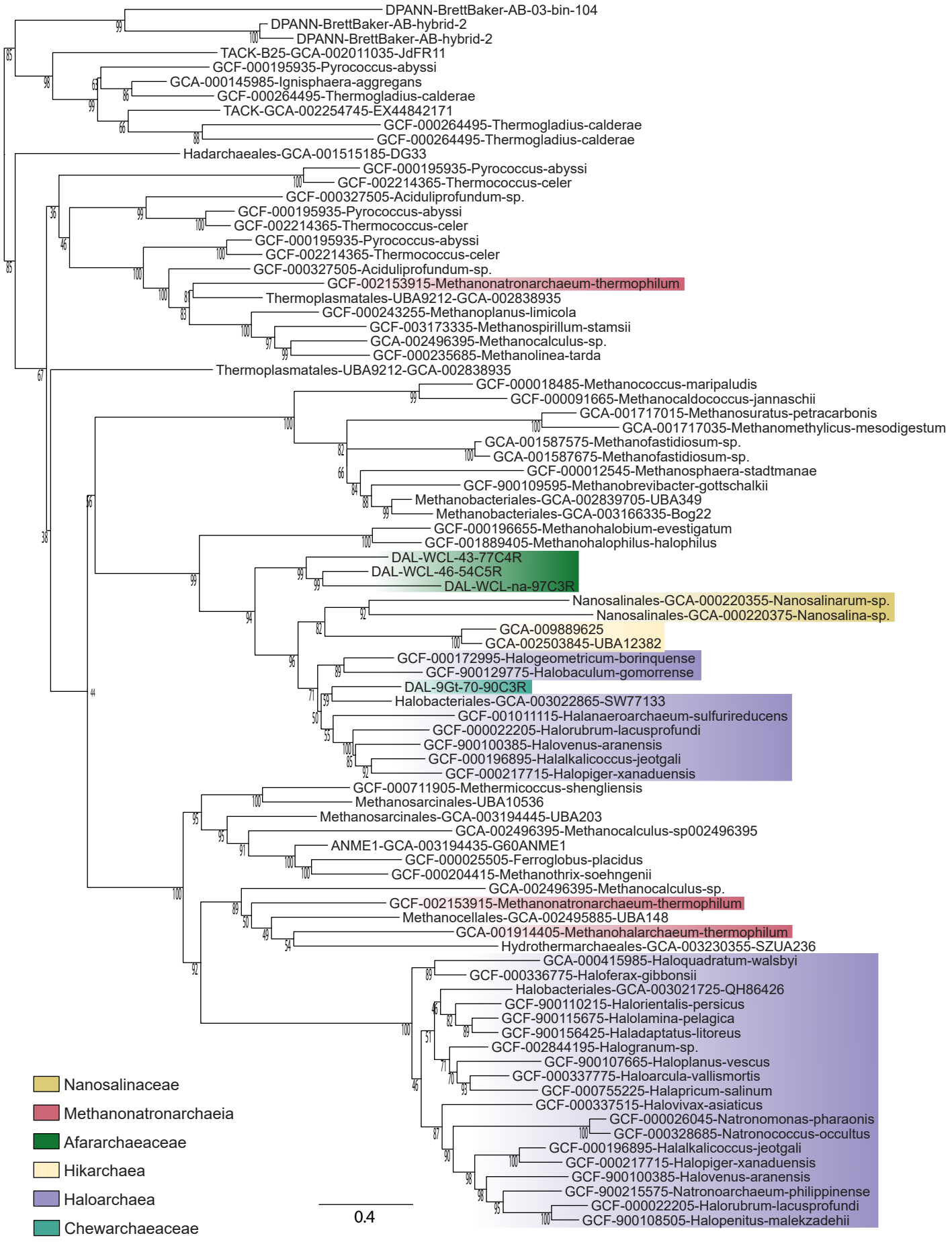
Supplementary Fig. 29 | Maximum likelihood phylogenetic tree of ZupT metal transporters. The tree was constructed with the LG+C60+F+Γ4 model of sequence evolution. Branch support was assessed using 1,000 ultrafast bootstraps. The scale bar represents the number of substitutions per position.



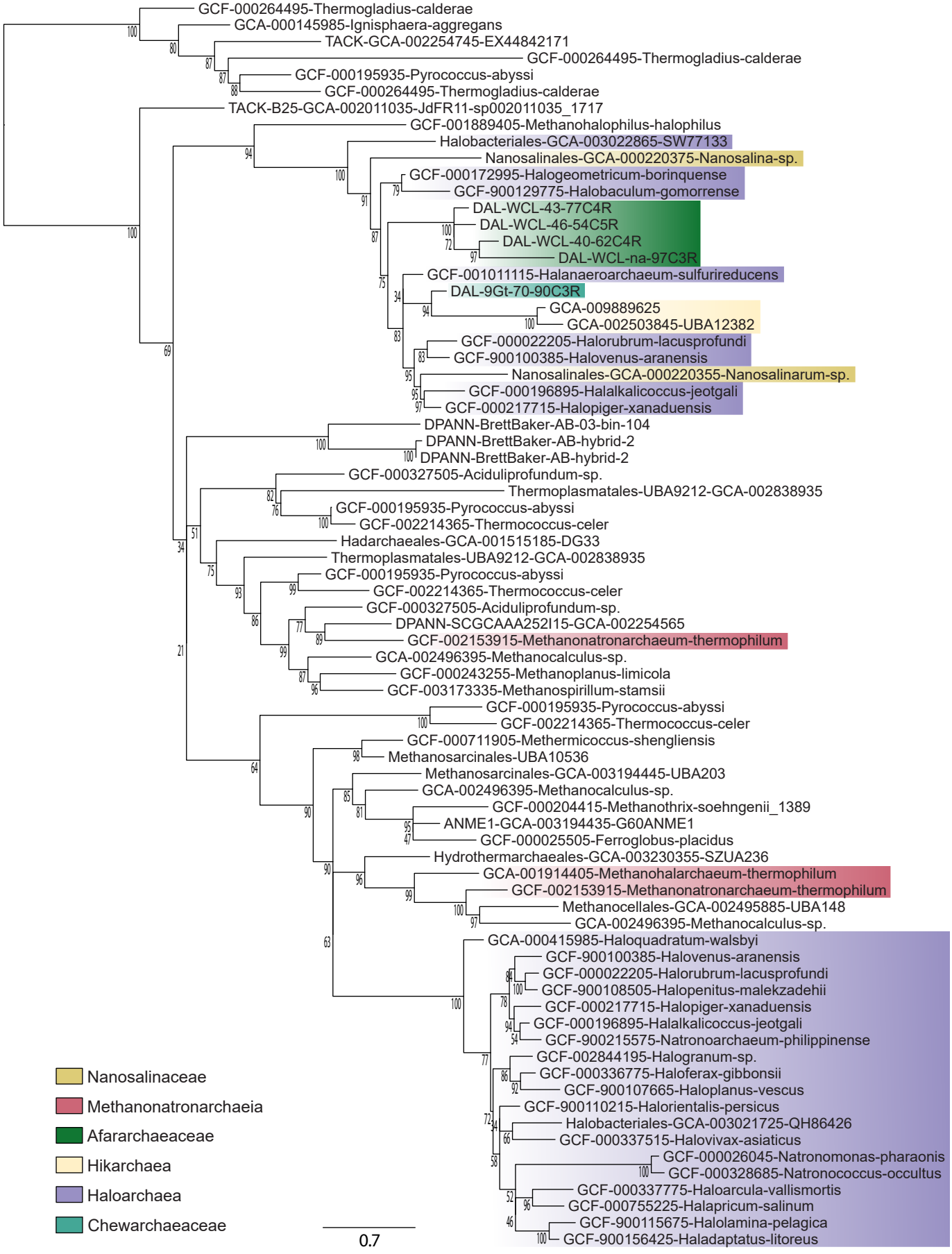
Supplementary Fig. 30 | Maximum likelihood phylogenetic tree of FieF metal transporters. The tree was constructed with the LG+C60+F+Γ4 model of sequence evolution. Branch support was assessed using 1,000 ultrafast bootstraps. The scale bar represents the number of substitutions per position.



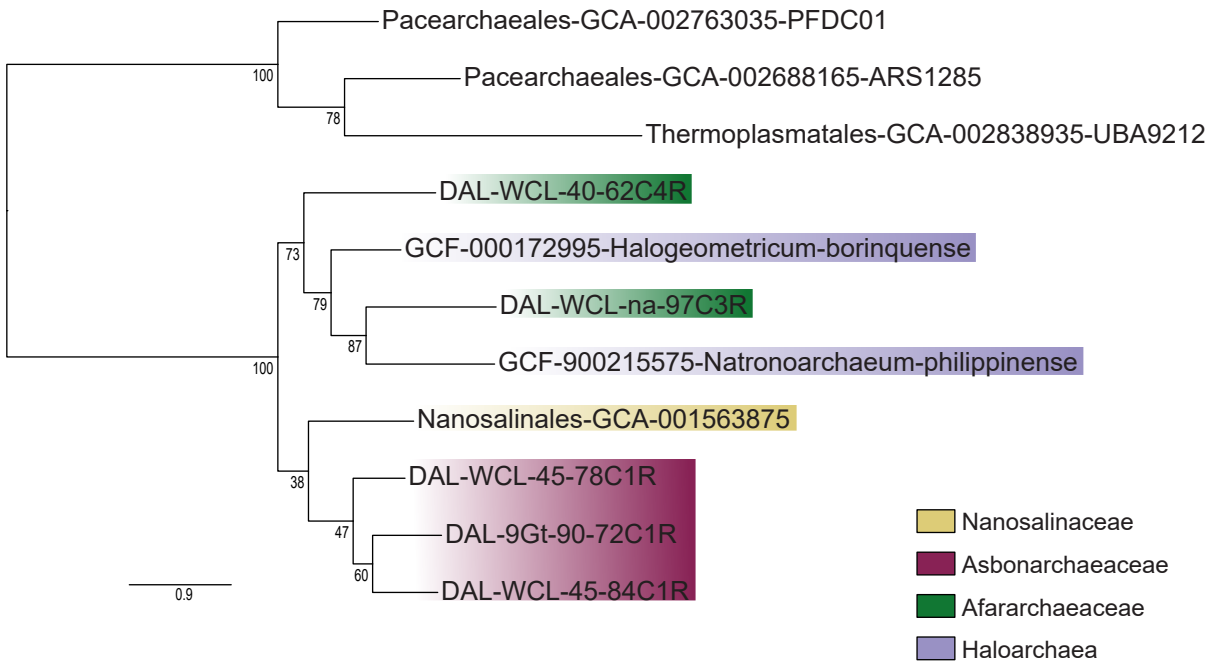
Supplementary Fig. 31 | Maximum likelihood phylogenetic tree of sulfur transporters. The tree was constructed with the LG+C60+F+Γ4 model of sequence evolution. Branch support was assessed using 1,000 ultrafast bootstraps. The scale bar represents the number of substitutions per position.



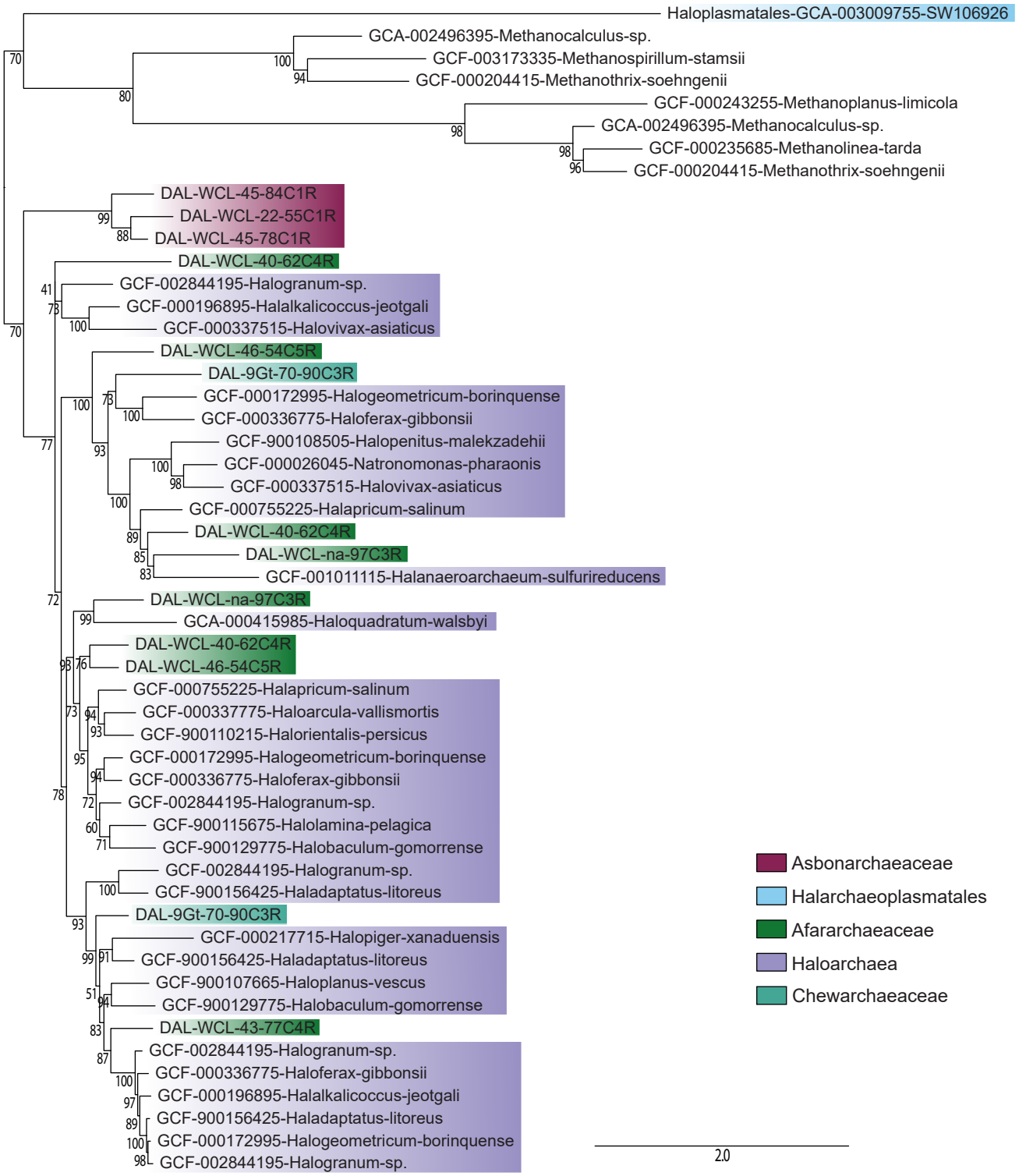
Supplementary Fig. 32 | Maximum likelihood phylogenetic tree of Na+/H+ antiporters. The tree was constructed with the LG+C60+F+Γ4 model of sequence evolution. Branch support was assessed using 1,000 ultrafast bootstraps. The scale bar represents the number of substitutions per position.



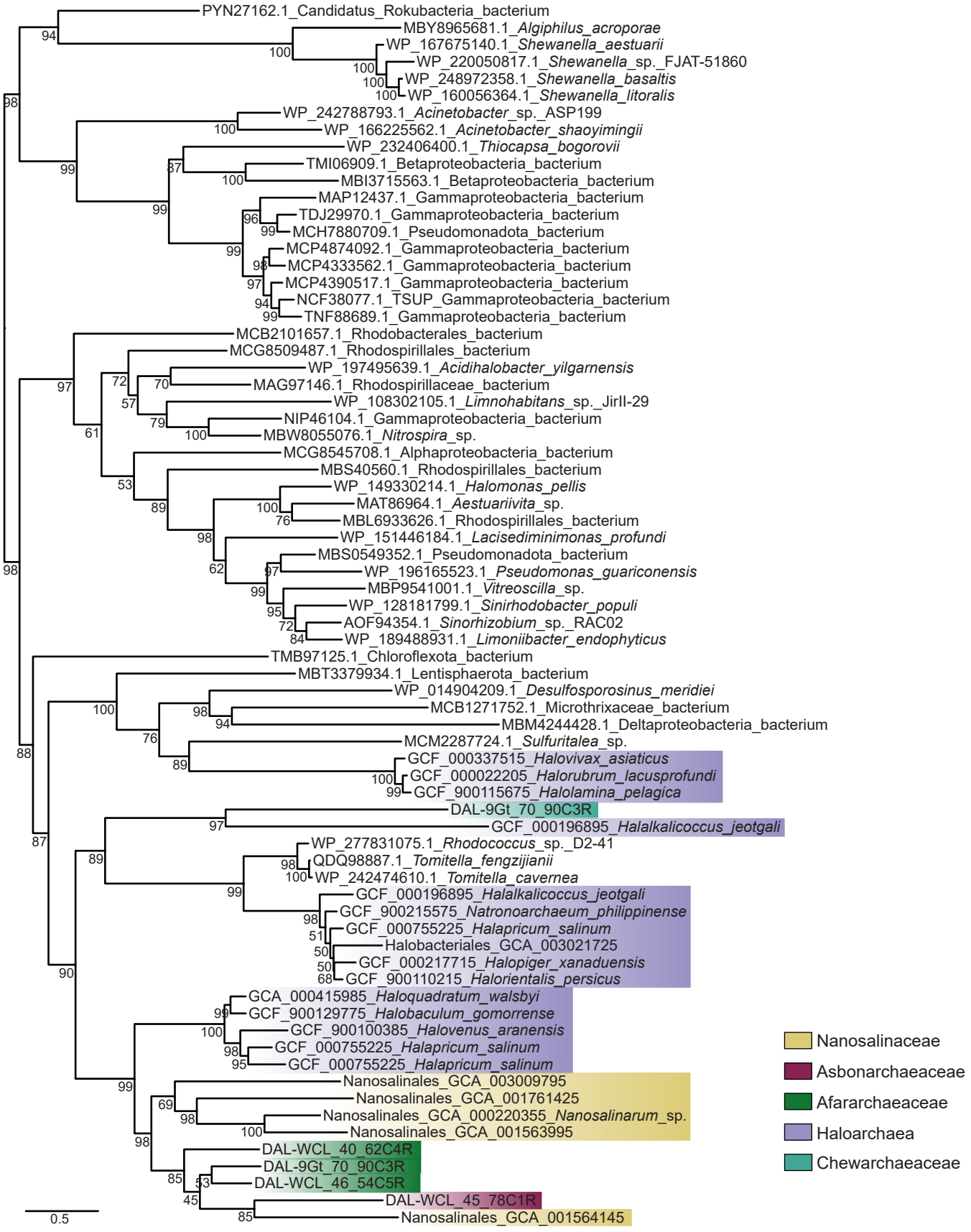
Supplementary Fig. 33 | Maximum likelihood phylogenetic tree of Na+/H+ antiporters. The tree was constructed with the LG+C60+F+Γ4 model of sequence evolution. Branch support was assessed using 1,000 ultrafast bootstraps. The scale bar represents the number of substitutions per position.



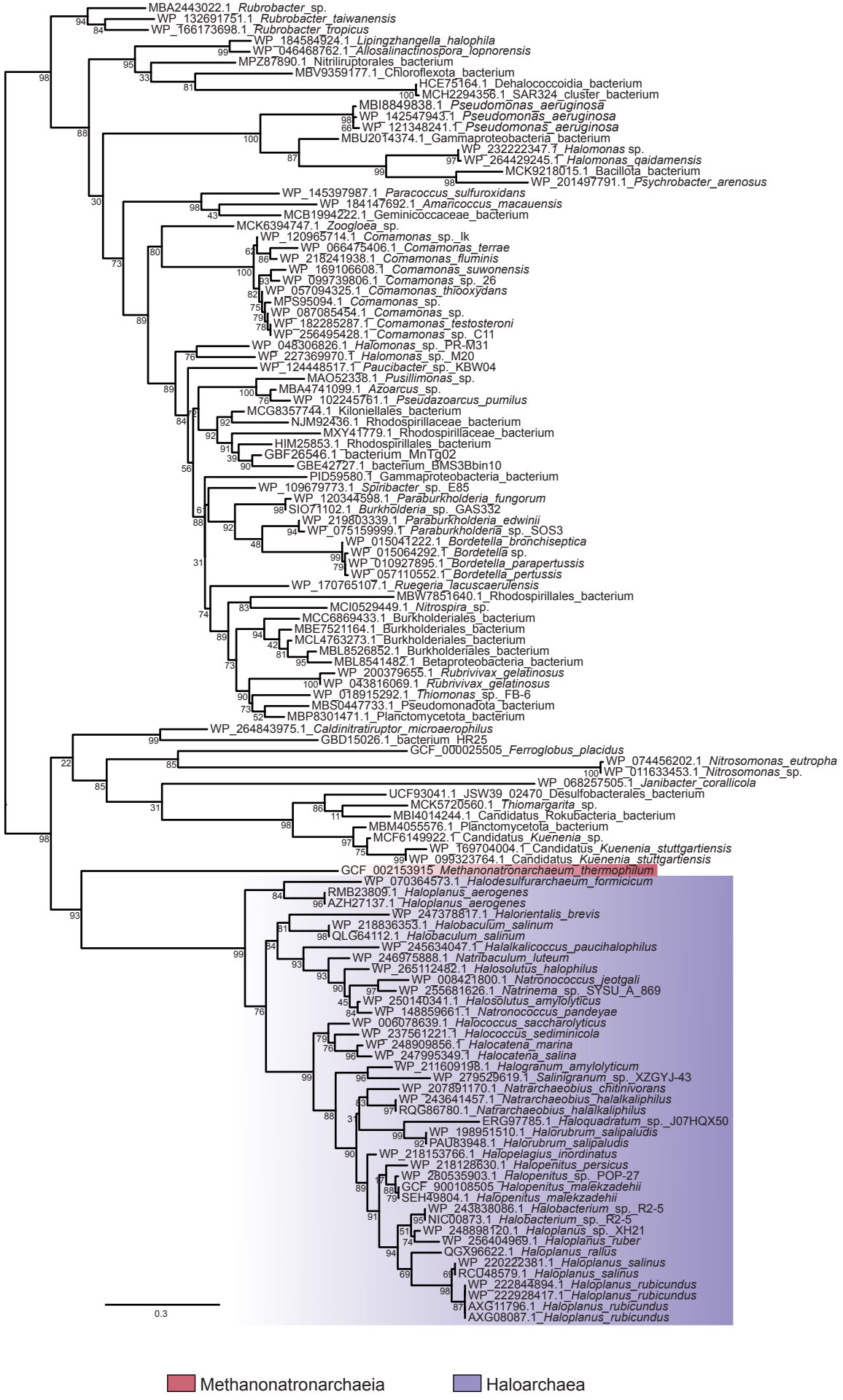
Supplementary Fig. 34 | Maximum likelihood phylogenetic tree of Na+/phosphate symporters. The tree was constructed with the LG+C60+F+Γ4 model of sequence evolution. Branch support was assessed using 1,000 ultrafast bootstraps. The scale bar represents the number of substitutions per position.



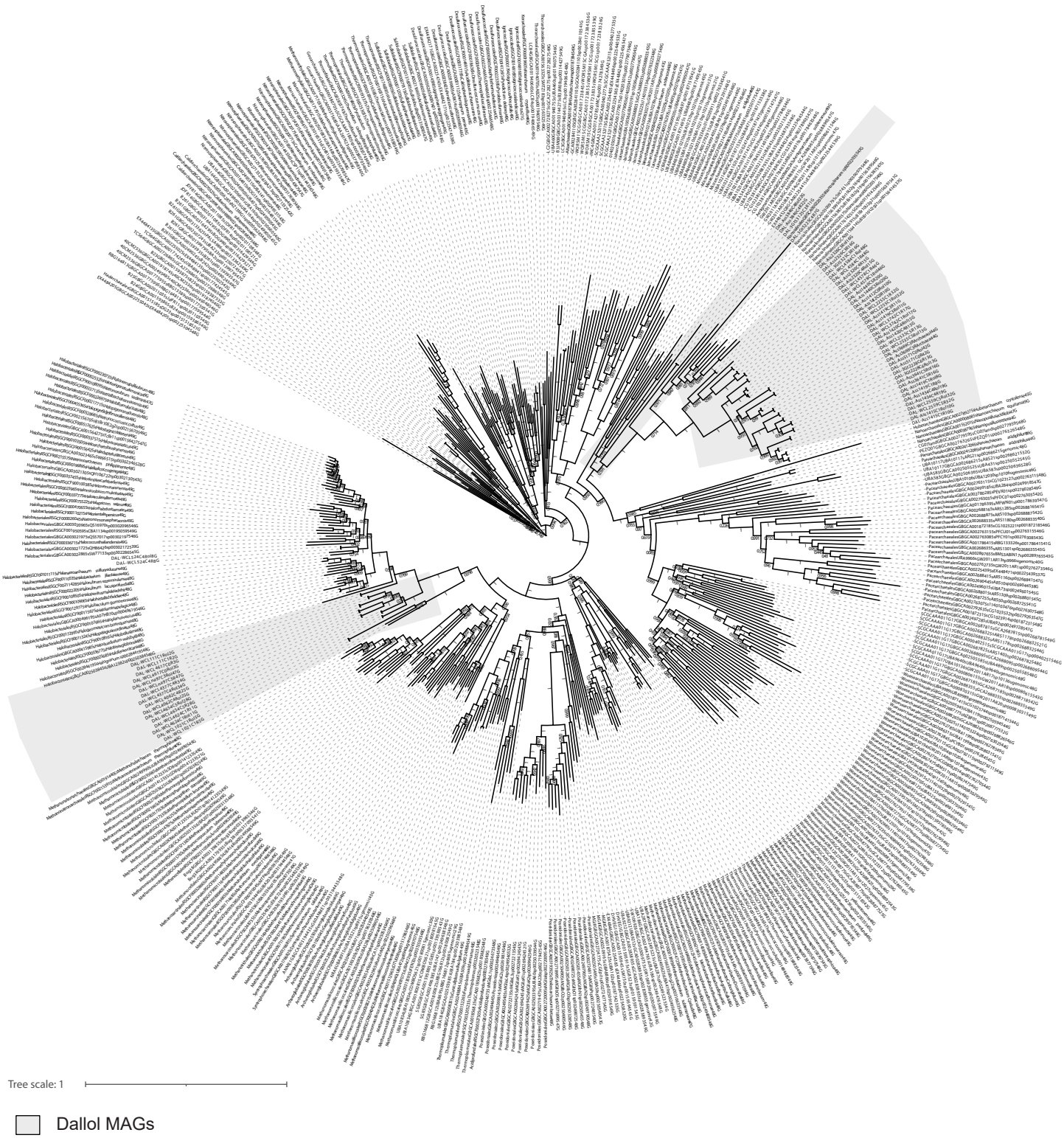
Supplementary Fig. 35 | Maximum likelihood phylogenetic tree of transporters of di- and tricarboxylate Krebs cycle intermediates. The tree was constructed with the LG+C60+F+Γ4 model of sequence evolution. Branch support was assessed using 1,000 ultrafast bootstraps. The scale bar represents the number of substitutions per position.



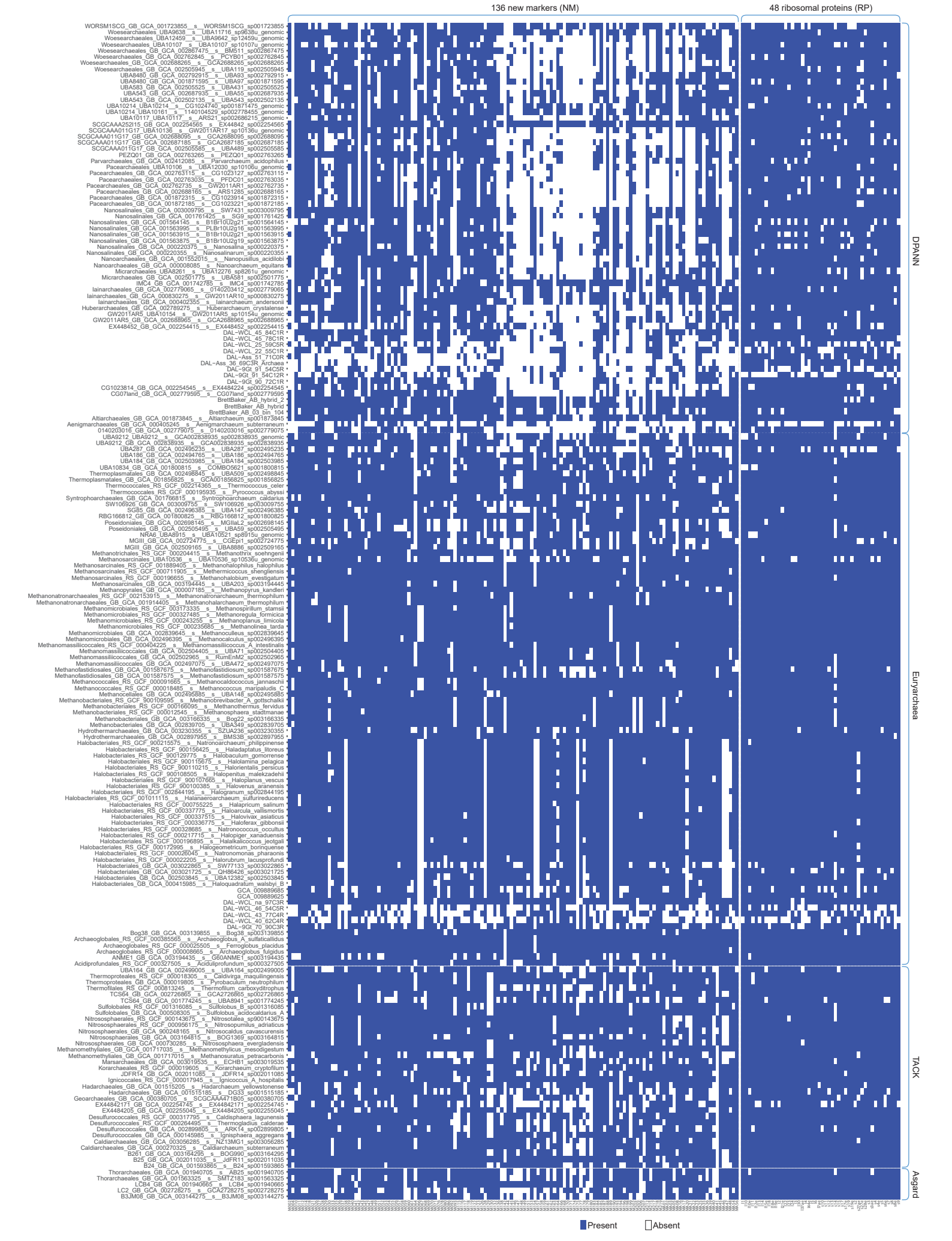
Supplementary Fig. 36 | Maximum likelihood phylogenetic tree of the AmiS/UreI urea transporter. The tree was constructed with the LG+C60+F+Γ4 model of sequence evolution. Branch support was assessed using 1,000 ultrafast bootstraps. The scale bar represents the number of substitutions per position.



Supplementary Fig. 37 | Maximum likelihood phylogenetic tree of TauE/SafE sulfite exporter. The tree was constructed with the LG+C60+F+Γ4 model of sequence evolution. Branch support was assessed using 1,000 ultrafast bootstraps. The scale bar represents the number of substitutions per position.



Supplementary Fig. 38 | Maximum likelihood phylogeny of 427 archaeal taxa based on the concatenated alignment of 49 ribosomal proteins. he ML tree was constructed using the LG+C20+F+Γ4 model of evolution. Branch support was assessed using 1,000 ultrafast bootstraps via IQ-TREE implementation. The scale bar represents the estimated number of substitutions per site. Each tip label provides information about the archaeal supergroup, taxonomic order based on GTDB, GTDB accession number, species identification, and the number of markers identified for each taxon out of the 49 ribosomal proteins**.**



Supplementary Fig. 39 | Distribution of NM and RP markers in the 192 archaeal taxa used in this study.