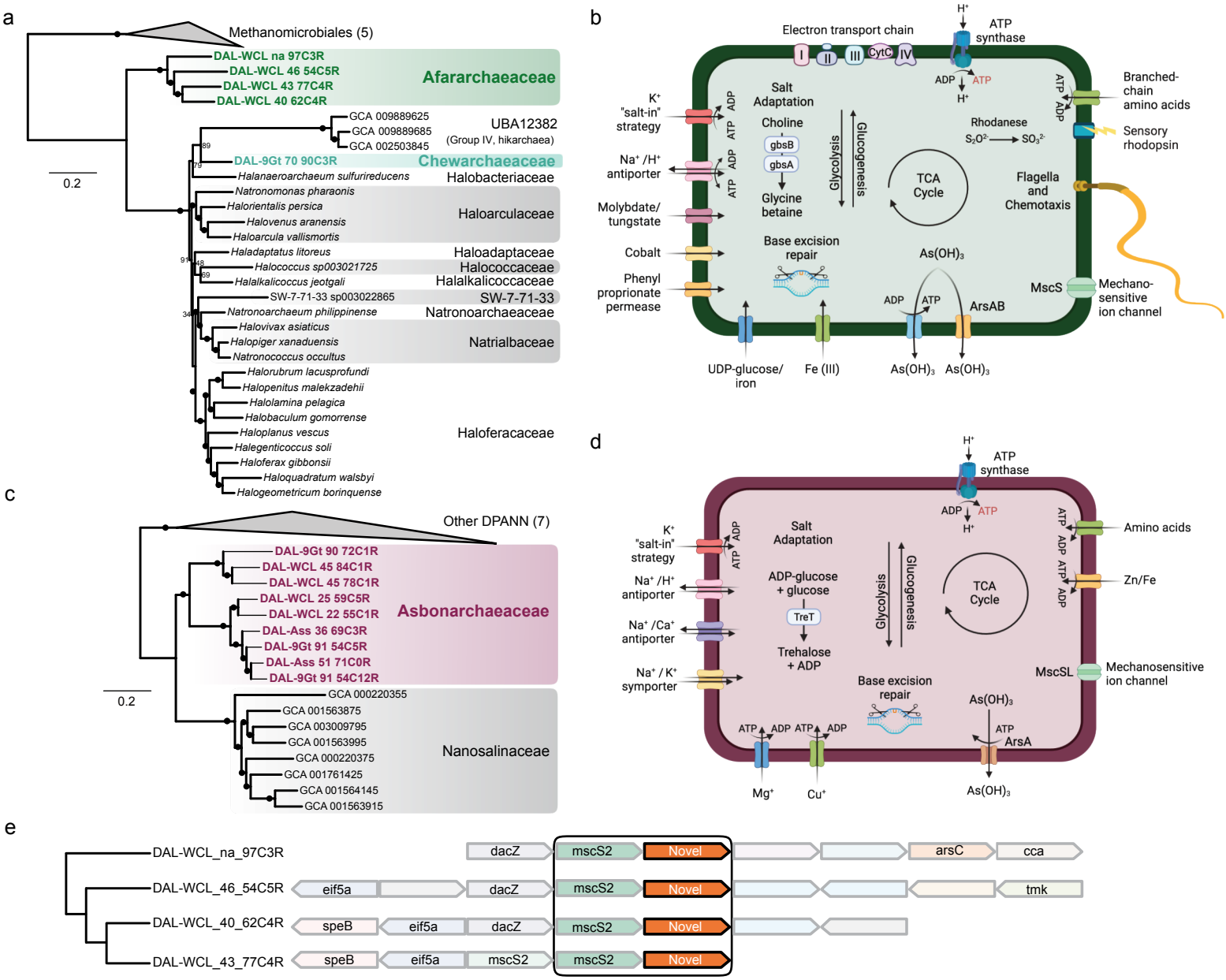


**Fig. 1 | Phylogenetic position and metabolic potential of the new families Afararchaeaceae and Asbonarchaeaceae. (a)** Maximum likelihood phylogenetic tree of 35 euryarchaea, including the four new Afararchaeaceae MAGs (highlighted in green), based on the concatenation of 122 single-copy proteins obtained from the Genome Taxonomy Database (GTDB). The tree was inferred via IQ-TREE with the LG+C60+F+Γ4 model of sequence evolution. The statistical support for branches, with filled circles representing values equal to or larger than 99% support, corresponds to 1,000 ultra-fast bootstrap replicates. The scale bar indicates the expected average number of substitutions per site. All taxonomic ranks shown are based on the GTDB r207 family-level classification. See Supplementary Fig. 2 for the uncollapsed tree. **(b)** Non-exhaustive metabolic scheme based on the predicted gene content of the most complete afararchaeal MAG (DAL-WCL_na_97C3R). A detailed table of the predicted gene content can be found in Supplementary Data 3. **(c)** Maximum likelihood phylogenetic tree of 24 DPANN archaea, including the nine new Asbonarchaeaceae MAGs (highlighted in wine), based on the concatenation of 99 single-copy proteins obtained from GTDB. The tree was inferred by IQ-TREE with the LG+C60+F+Γ4 model of sequence evolution. The statistical support for branches corresponds to 1,000 ultra-fast bootstrap replicates. The scale bar indicates the expected average number of substitutions per site. All taxonomic ranks are based on the GTDB r207 family-level classification. See Supplementary Fig. 3 for the uncollapsed tree. **(d)** Non-exhaustive metabolic scheme based on the predicted gene content of the most complete asbonarcharchaeal MAG (DAL-WCL_45_84C1R). A detailed table of the predicted gene content can be found in Supplementary Data 4. **(e)** Gene maps showing a novel gene family (orange) linked to a conserved mechanosensitive ion channel (mscS2) in the afararchaeal MAGs. Gene abbreviations are as follows: agmatinase (speB), eukaryotic initiation factor 5A (eif5a), di-adenylate cyclase (dacZ), arsenate reductase (arsC), tRNA nucleotidyltransferase (cca), thymidylate kinase (tmk).

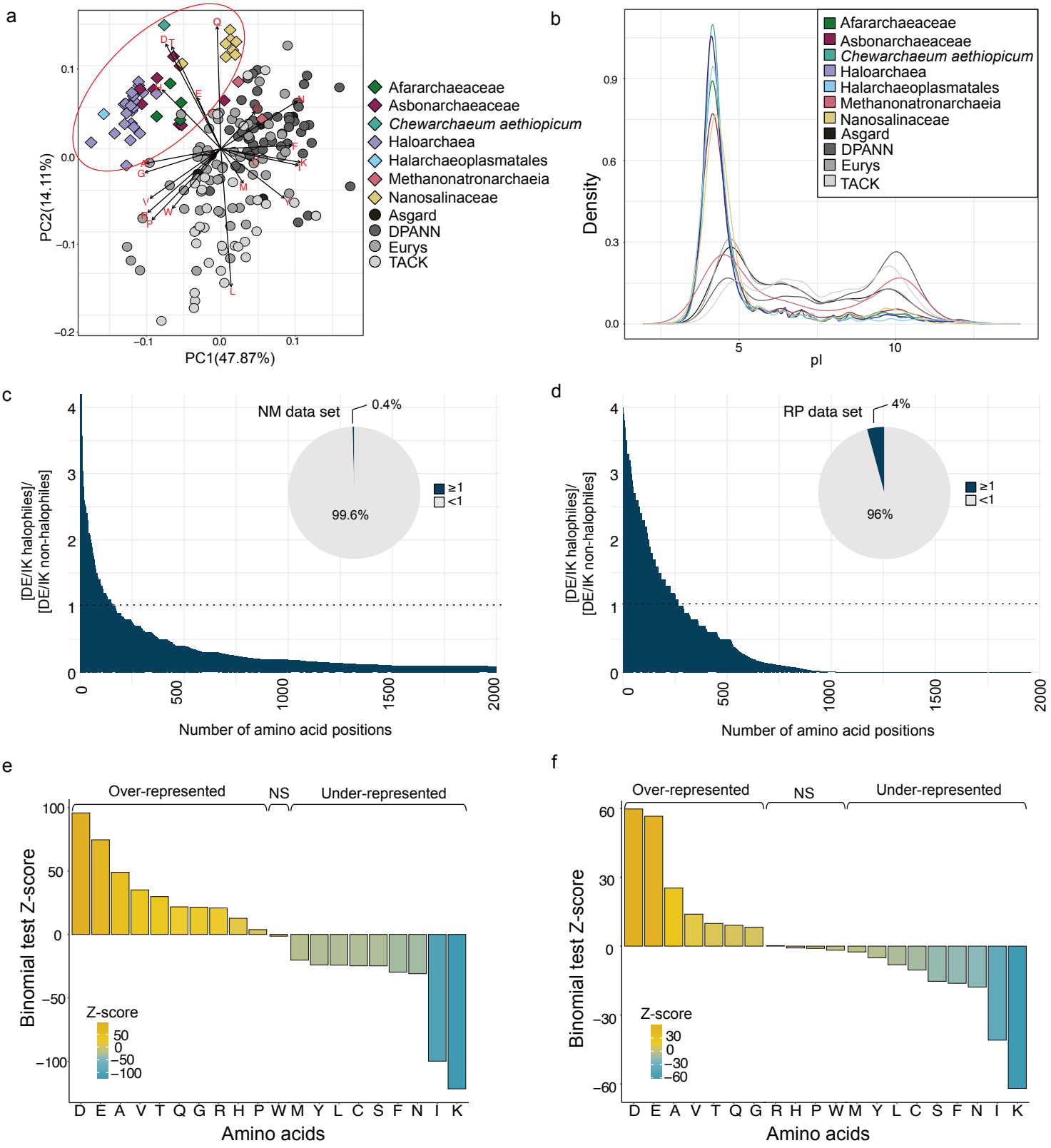


**Fig. 2 | Protein amino acid compositional biases in extremely halophilic archaeal lineages. (a)** PCA plot of 192 archaeal proteomes based on amino acid frequencies. The red ellipse indicates the clustering of all extreme halophiles (colored diamonds), including the newly identified families Afararchaeaceae (green color) and Asbonarchaeaceae (wine color). **(b)** Isoelectric point (pI) distribution of 192 archaeal proteomes. Non-halophilic archaea (grey lines) display a bimodal distribution of pI values, while extreme halophiles (colored lines) exhibit a single spike at pI ~4, indicating a highly acidic proteome. **(c,d)** D+E/I+K site-by-site bias (defined as the ratio [D+E/I+K for halophiles]/[D+E/I+K for non-halophiles]) for the 2,000 most biased sites of the **(c)** NM dataset (39,385 amino acid positions) and **(d)** RP dataset (6,792 amino acid positions). Inset pie charts depict the proportion of amino acids with a ratio greater than or equal to 1 (dark blue) versus less than 1 (grey). **(e,f)** Binomial tests for the **(e)** NM and **(f)** RP datasets compare the proportions of all 20 amino acids between extreme and non-halophiles. Z-scores were calculated relative to extreme halophiles, with |Z| > 1.96 indicating significant enrichment of a given amino acid in extreme halophile sequences ("Over-represented"), |Z| < -1.96 indicating significant depletion of a given amino acid in extreme halophile sequences ("Under-represented"), and some amino acids showing no significant bias ("NS").

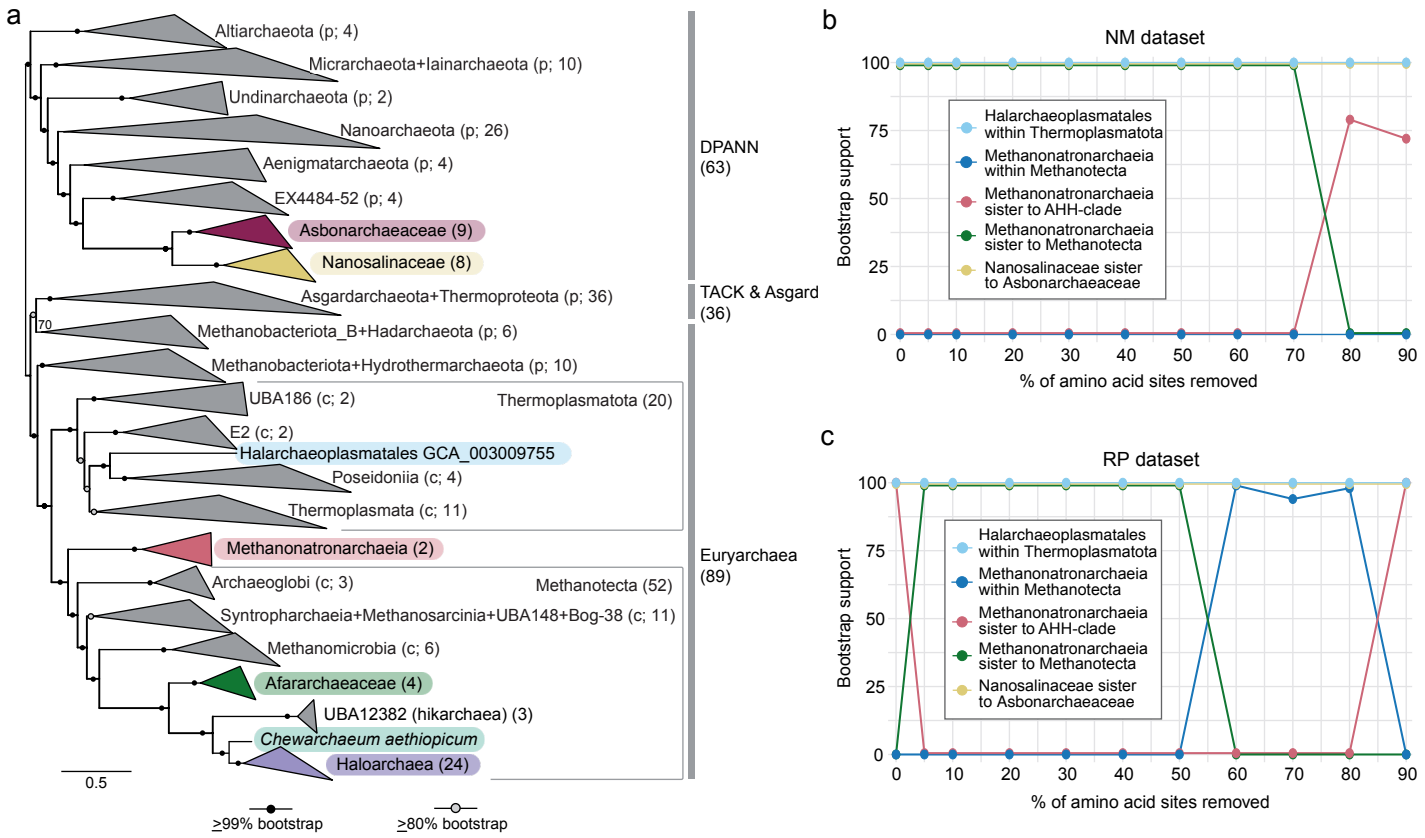


**Fig. 3 | Maximum likelihood phylogeny of archaea, including the new groups Afararchaeaceae and Asbonarchaeaceae. (a)** Phylogenetic tree based on the concatenation of 136 conserved markers (NM dataset) across 192 taxa (39,385 sites) via IQ-TREE under the LG+C60+F+Γ4 model of evolution. Statistical support indicated on the branches corresponds to 1,000 ultra-fast bootstrap replicates. The scale bar indicates the number of substitutions per site. Colors indicate the currently known groups of extremely halophilic archaea. The size of collapsed clades is indicated in parentheses; see Extended Data Fig. 3 for the uncollapsed tree. **(b,c)** Impact of the progressive removal (in steps of 10%) of the most compositionally biased sites from the **(b)** 192-NM (39,385 amino acid positions) and **(c)** 192-RP (6,792 amino acid positions) datasets. Lines show the statistical support values for the position of each of the halophilic clades of interest. These support values were estimated using the ultrafast bootstrap approximation from the ML tree reconstruction (LG+C60+F+Γ4 model) for each site-removal step.

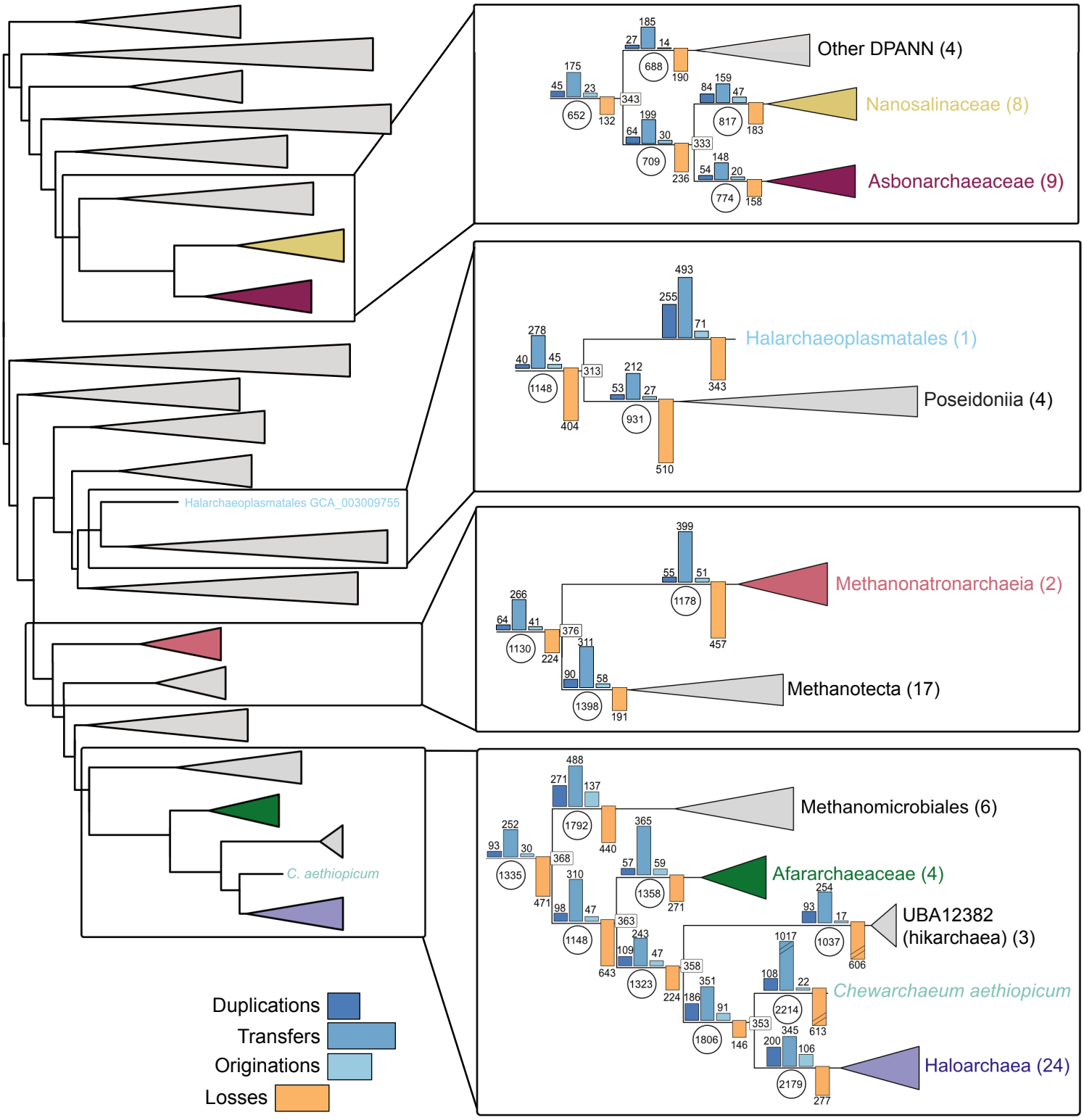


**Fig. 4 | Schematic representation of the tree reconciliation analysis based on the NM species tree.** The full archaeal tree is shown on the left; boxes on the right highlight the details for the four main groups of halophilic archaea: Nanosalinaceae+Asbonarchaeaceae, Halarchaeoplasmatales, Methanonatronarchaeia, and Afararchaeaceae+Haloarchaea. The bar plots on the branches represent the number of gene duplications, transfers, originations, and losses, and the circles indicate the number of predicted ancestral gene copy numbers. The number of taxa in each collapsed clade is indicated by the number in parentheses next to the clade name. The complete version of this tree with the events for all archaeal nodes can be found in Extended Data Fig. 8.