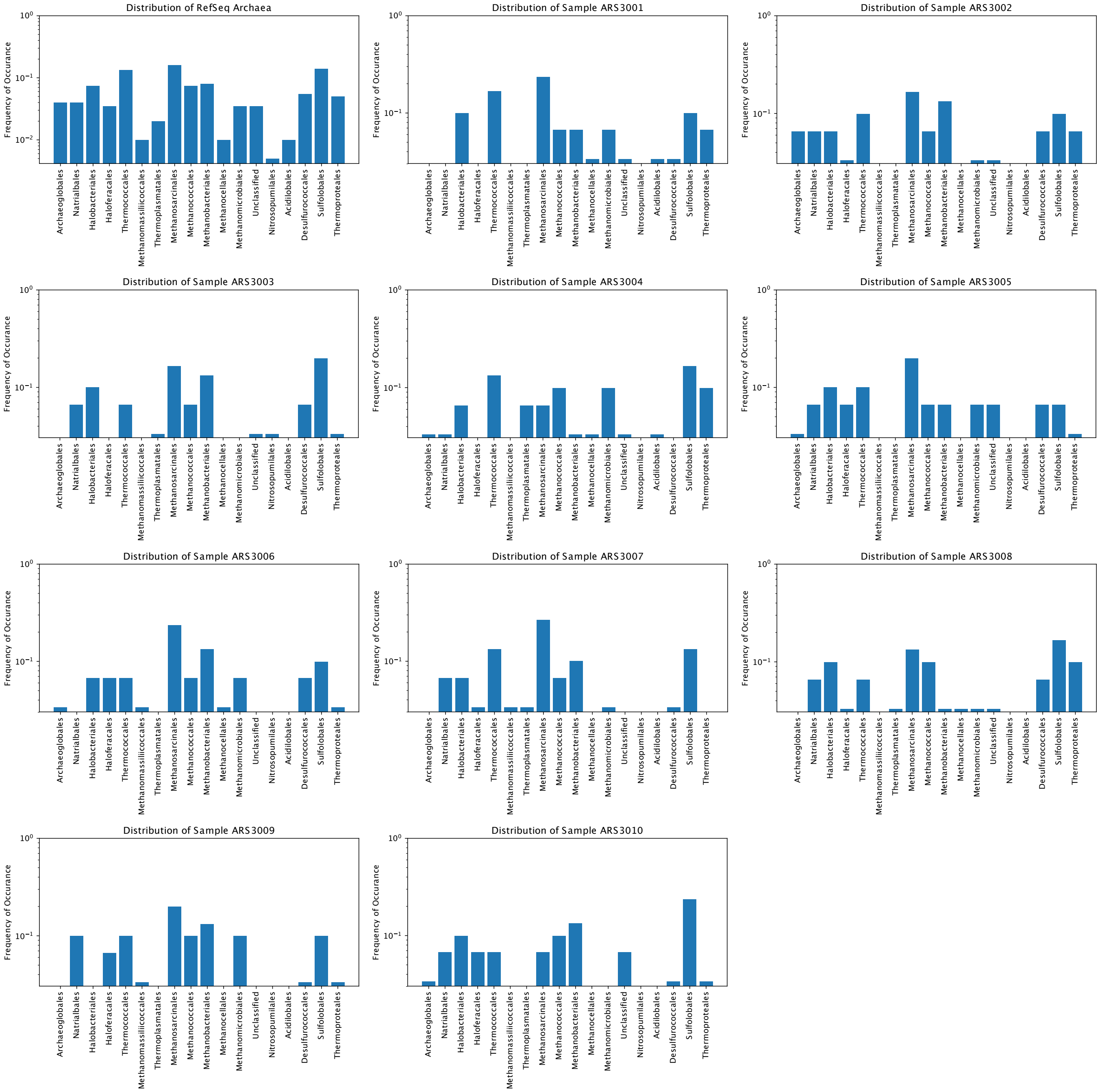
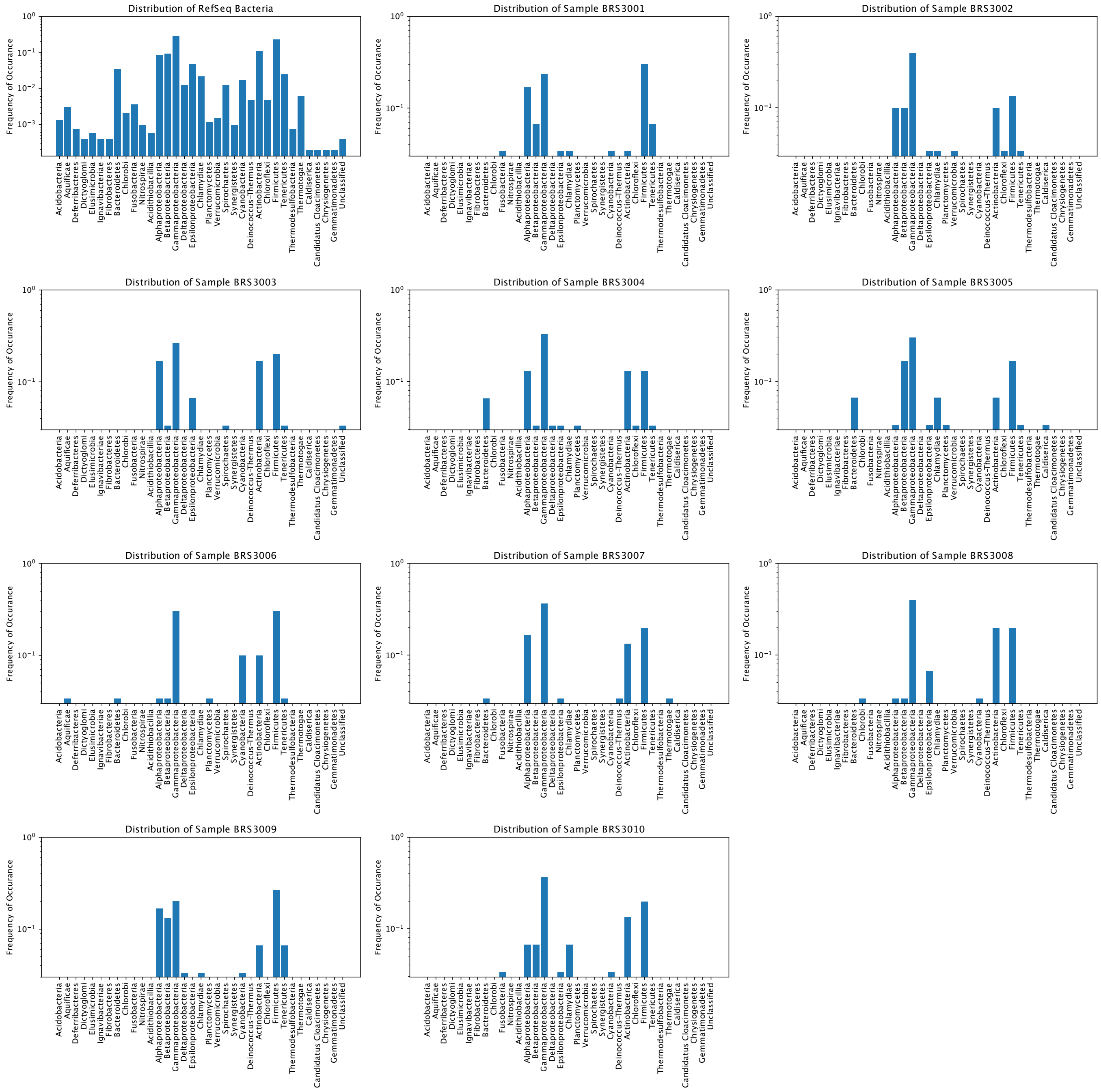


Supplementary Fig. S1 | Distribution of archaeal genomes across taxonomic groups. Each sample used in the study is represented by a panel, which indicates the taxonomic distribution of the organisms within that sample from across 212 completely sequenced archaeal RefSeq genomes. For each sample we chose 30 genomes such that they approximate the diversity across all 212 genomes.



Distribution of RefSeq Bacteria

Distribution of Sample BRS3001

Distribution of Sample BRS3002

Distribution of Sample BRS3003

Distribution of Sample BRS3004

Distribution of Sample BRS3005

Distribution of Sample BRS3006

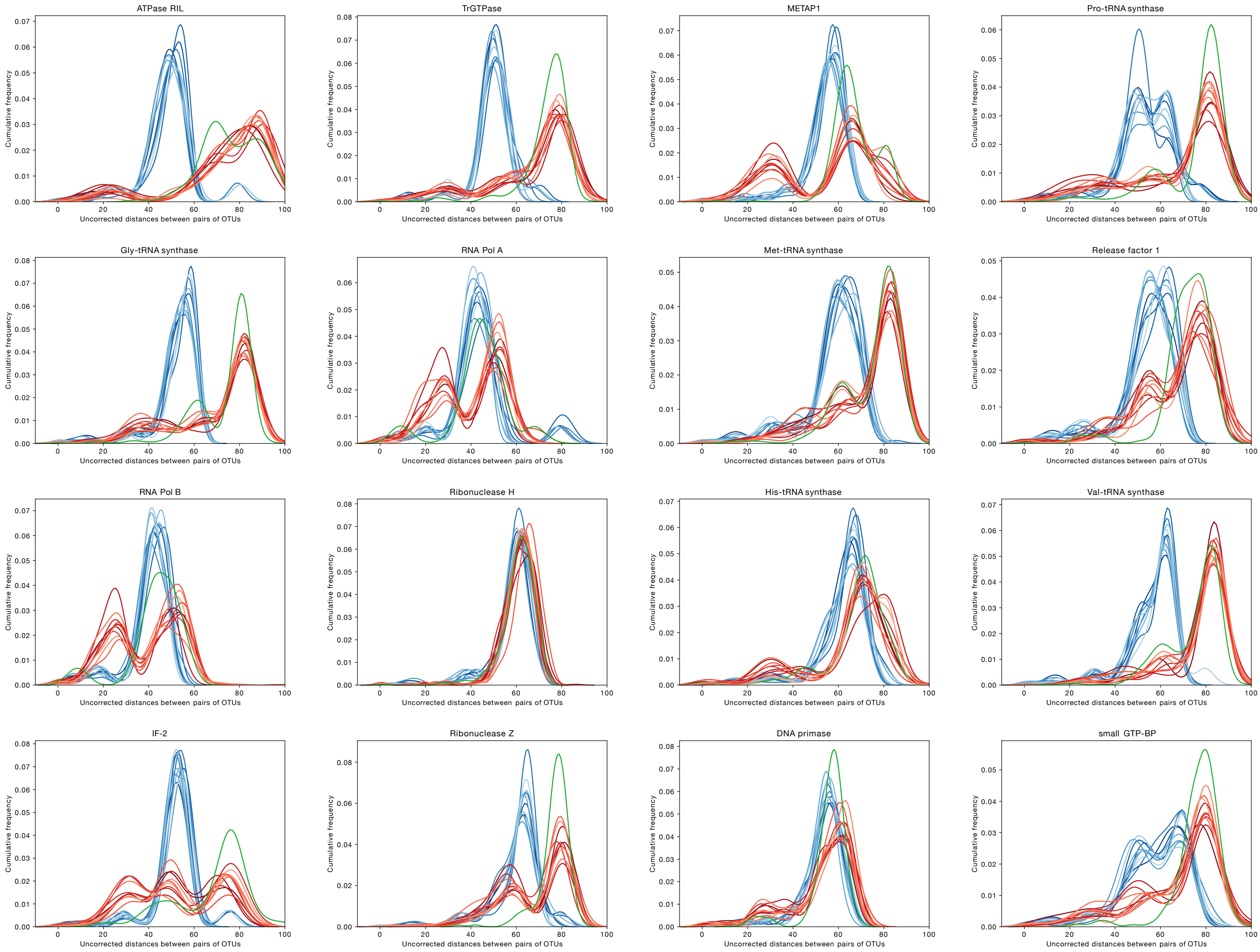
Distribution of Sample BRS3007

Distribution of Sample BRS3008

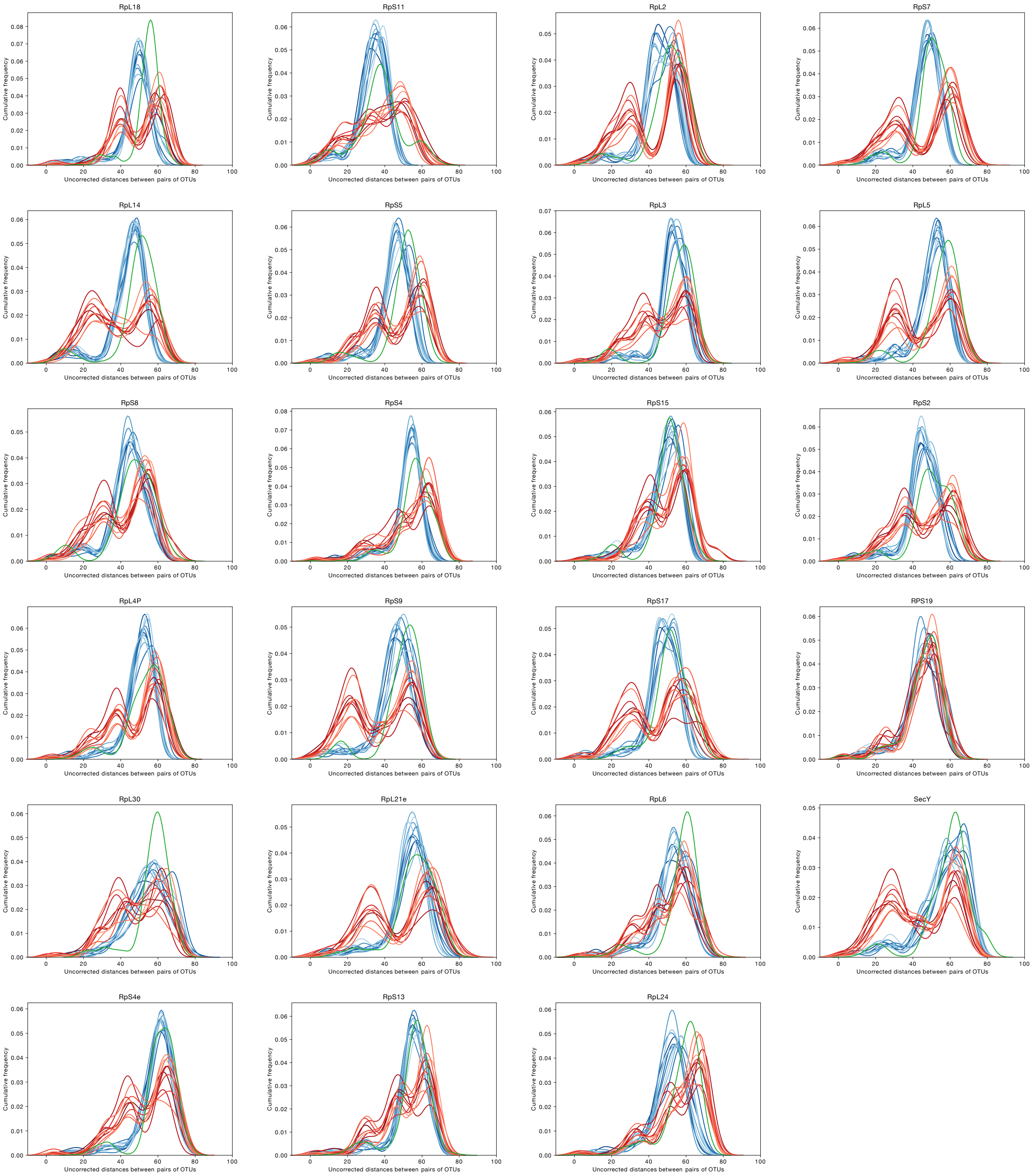
Distribution of Sample BRS3009

Distribution of Sample BRS3010

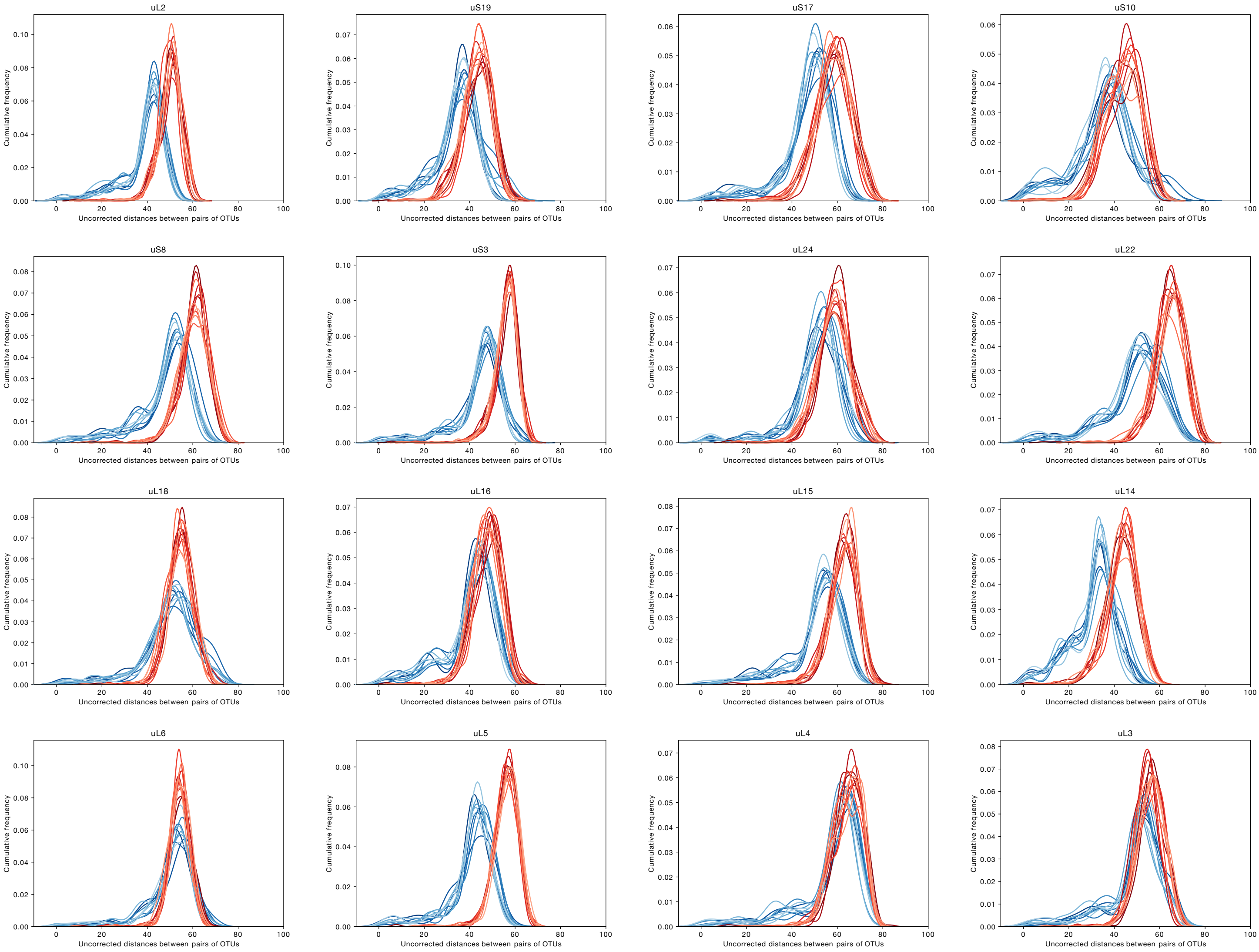
Supplementary Fig. S2 | Distribution of bacterial genomes across taxonomic groups. Each sample used in the study is represented by a panel which indicates the taxonomic distribution of the organisms within that sample from across 5443 completely sequenced bacterial RefSeq genomes. For each sample we chose 30 genomes such that they approximate the diversity and the representation across all 5443 genomes.



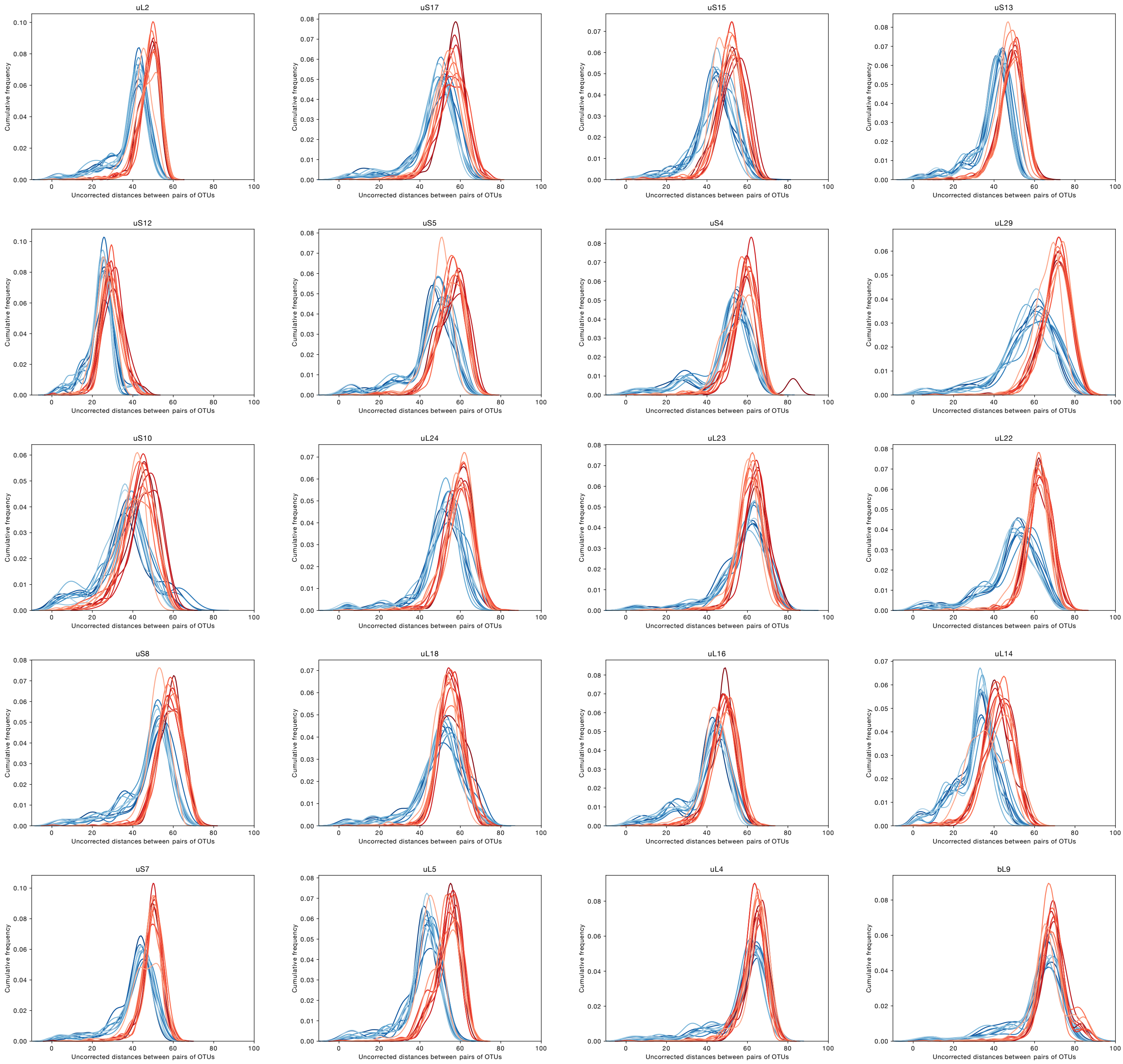
Supplementary Fig. S3 | Uncorrected p-distances for 16 archaeal non-ribosomal proteins: To examine the possibility that topological differences in trees could stem from the phylogenetic depth (sequence divergence) within the individual protein trees, we plotted uncorrected p-distances for each of the 16 universal non-ribosomal proteins from 10 archaeal RefSeq samples (blue), 10 non-asgard archaeal MAGs (red) and one asgard archaeal MAG sample (green), samples consisting of 30 organisms each.



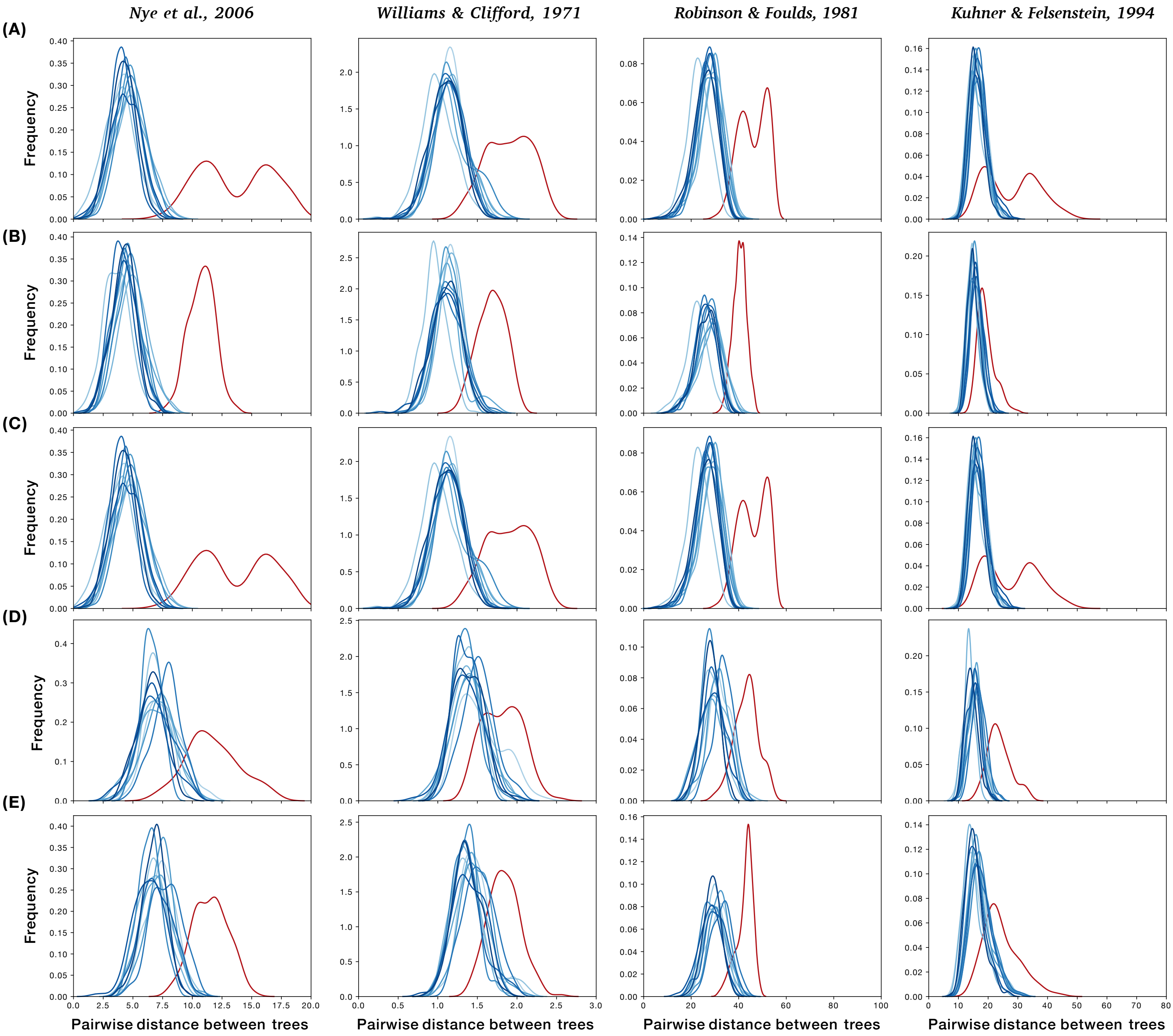
Supplementary Fig. S4 | Uncorrected p-distances for 23 archaeal ribosomal proteins: Same as in Supplemental Fig. S3, but for 23 ribosomal proteins from 10 archaeal RefSeq samples (blue), 10 non-asgard archaeal MAGs (red) and one asgard archaeal MAG sample (green), samples consisting of 30 organisms each. Note that p-distances for non-ribosomal proteins (Supplemental Fig. S3) of MAGs are often shifted towards higher divergence, which is a property of the MAGs themselves, while p-distances between 23 r-proteins are more uniform for the same samples, though often bimodal for non-asgard MAGs, but not bimodal for asgard MAGs. Note that the distribution pf p-distances for r-ptoeins from asgard MAGs closely follows the distribution for archaeal orgDNA.



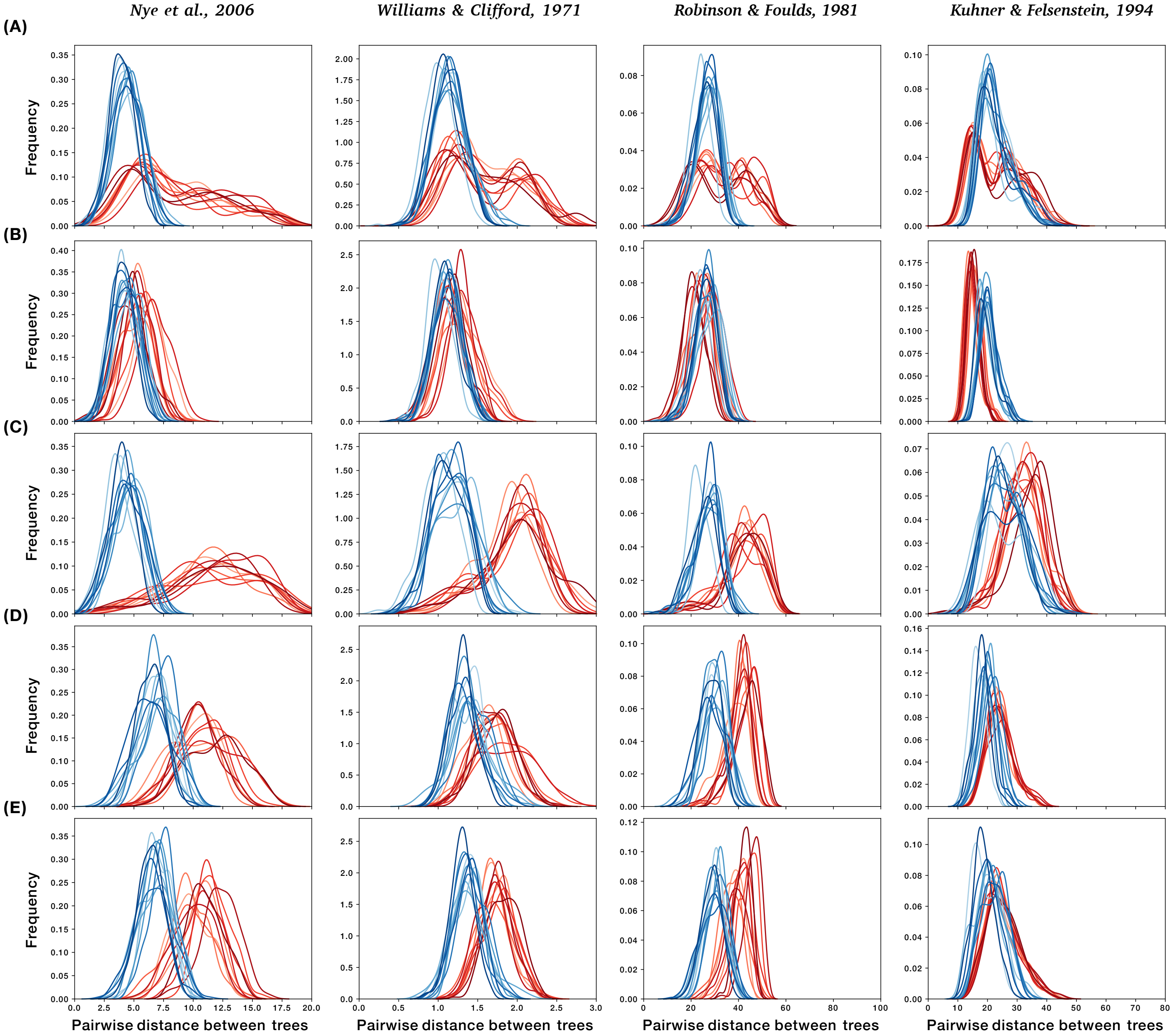
Supplementary Fig. S5 | Uncorrected p-distances for 16 ribosomal proteins from bacterial RefSeq genomes and CPR MAGs: The uncorrected p-distances for each of the 16 universal ribosomal proteins from 10 bacterial RefSeq samples (blue) and 11 CPR MAG samples (red) each consisting of 30 organisms each are plotted.



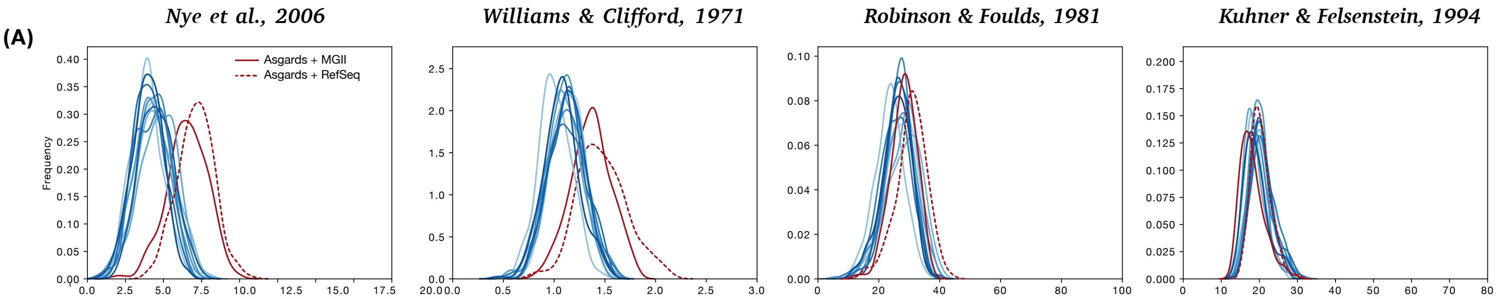
Supplementary Fig. S6 | Uncorrected p-distances for 20 ribosomal proteins from bacteria: The uncorrected p-distances for each of the 20 universal ribosomal proteins from 10 bacterial RefSeq samples (blue) and 10 bacterial non-CPR MAG samples (red) each consisting of 30 organisms each are plotted.



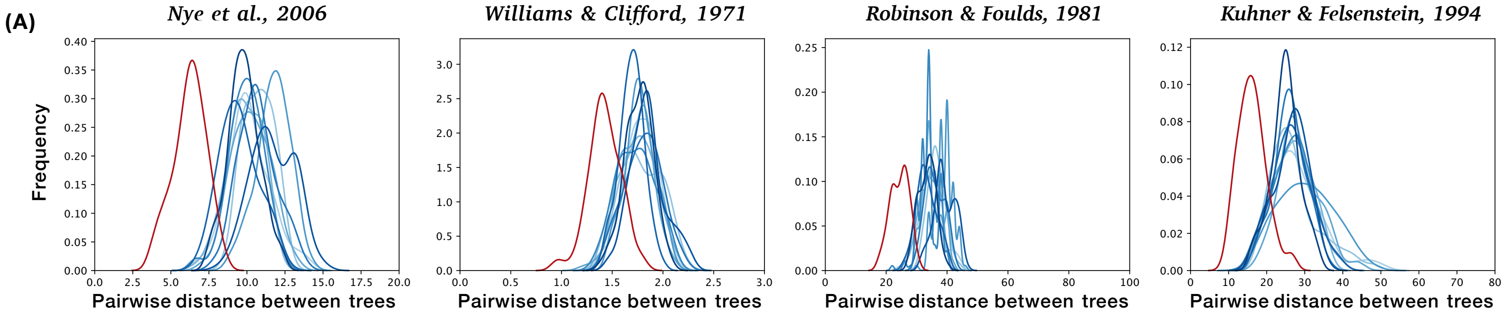
Supplementary Fig. S7 | Distribution of pairwise tree-distances between MAGs and RefSeq after site exlusion: In order to test whether the topological inconsistency that we observe for MAG samples in Fig. 2 was due to phylogenetic effects stemming from highly variable sites, we trimmed sites from the alignment using BMGE[36]. In each case, alignments for a matched universal set of proteins from MAGs (shown in red) and 10 random samples of RefSeq (shown in shades of blue) were subjected to site exclusion, tree construction and comparison. Relative to Fig. 2, site-exclusion[36] does not increase or detract from the topological consistency of trees within a set. **(A)** Trees for 39 universal proteins (site excluded alignments) from 30 Asgard archaeal MAGs are compared with those from 10 samples of 30 archaeal RefSeq genomes. **(B)** Trees for 23 ribosomal proteins (site excluded alignments) from Asgard archaeal MAGs are compared with those from archaeal RefSeq genomes. **(C)** Trees for 16 non-ribosomal proteins (site excluded alignments) from archaeal MAGs are compared with those from archaeal RefSeq genomes. **(D)** Trees for 16 ribosomal proteins (site excluded alignments) from CPR MAGs are compared with those from bacterial RefSeq genomes. **(E)** Trees for 20 ribosomal proteins from bacterial MAGs are compared with those from 10 samples of 30 RefSeq genomes. Note that the comparison of (r-protein or other) tree similarity within and between sets of genomes reported here is distinct from the inspection of individual genomes in which the position of one new set of (r-protein) sequences in a genome would be added to a reference system of closed genomes to be tested for statistical inconsistency in branching behaviour.



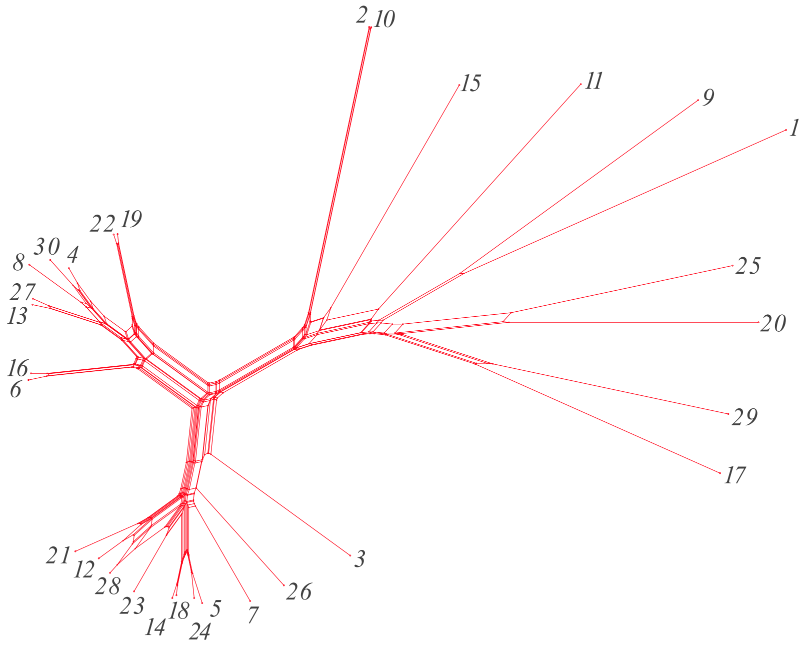
Supplementary Fig. S8 | Distribution of pairwise tree-distances between RefSeq and metagenomes: The pairwise comparisons of tree distances computed using four different metrics (see Methods) are shown. In each case, a matched set of proteins present in 10 MAG samples and 10 samples from RefSeq are taken to plot comparable distributions. The MAG sample is always shown in red. **(A)** Trees for 39 universal proteins from 10 samples of 30 non-asgard archaeal MAGs each are compared with trees for the 39 homologues from 10 samples of 30 archaeal RefSeq genomes **(B)** Trees for 23 ribosomal proteins from 10 samples of 30 non-asgard archaeal MAGs each are compared with those from 10 samples of 30 archaeal RefSeq genomes. **(C)** Trees for 16 non-ribosomal proteins from 30 archaeal MAGs are compared with those from 10 samples of 30 archaeal RefSeq genomes. **(D)** Trees for 16 ribosomal proteins from 30 candidate phyla radiation CPR MAGs from ref. 4 are compared with those from 10 samples of 30 bacterial RefSeq genomes. **(E)** Trees for 20 ribosomal proteins from non-CPR 30 bacterial MAGs are compared with those from 10 samples of 30 RefSeq genomes. In all panels, blue curves represent the 10 independent reference samples while the red curve represents the MAGs. Individual p-values for each comparison are given in Table S4.



Supplementary Figure S9 | Distribution of pairwise tree-distances between MAGs and RefSeq: The pairwise comparisons of tree distances for 23 universal ribosomal proteins computed using four different metrics (see Methods) are shown. In each case, a matched set of proteins present in MAGs and 10 random samples from RefSeq are taken to plot comparable distributions. 16 Asgard MAG samples combined with 14 non-Asgard archaeal MAGs is indicated with a red line. Dotted lines represent the tree comparison metrics for a sample of 16 Asgard MAGS combined with 14 RefSeq genomes. Given that the RefSeq PCGs and marine surface archaeal MAGs has minimal incongruence between trees (Figure 2, Supplementary Figure S7) the differences in the tree comparison metrics in this case are determined only by the Asgard MAGs.

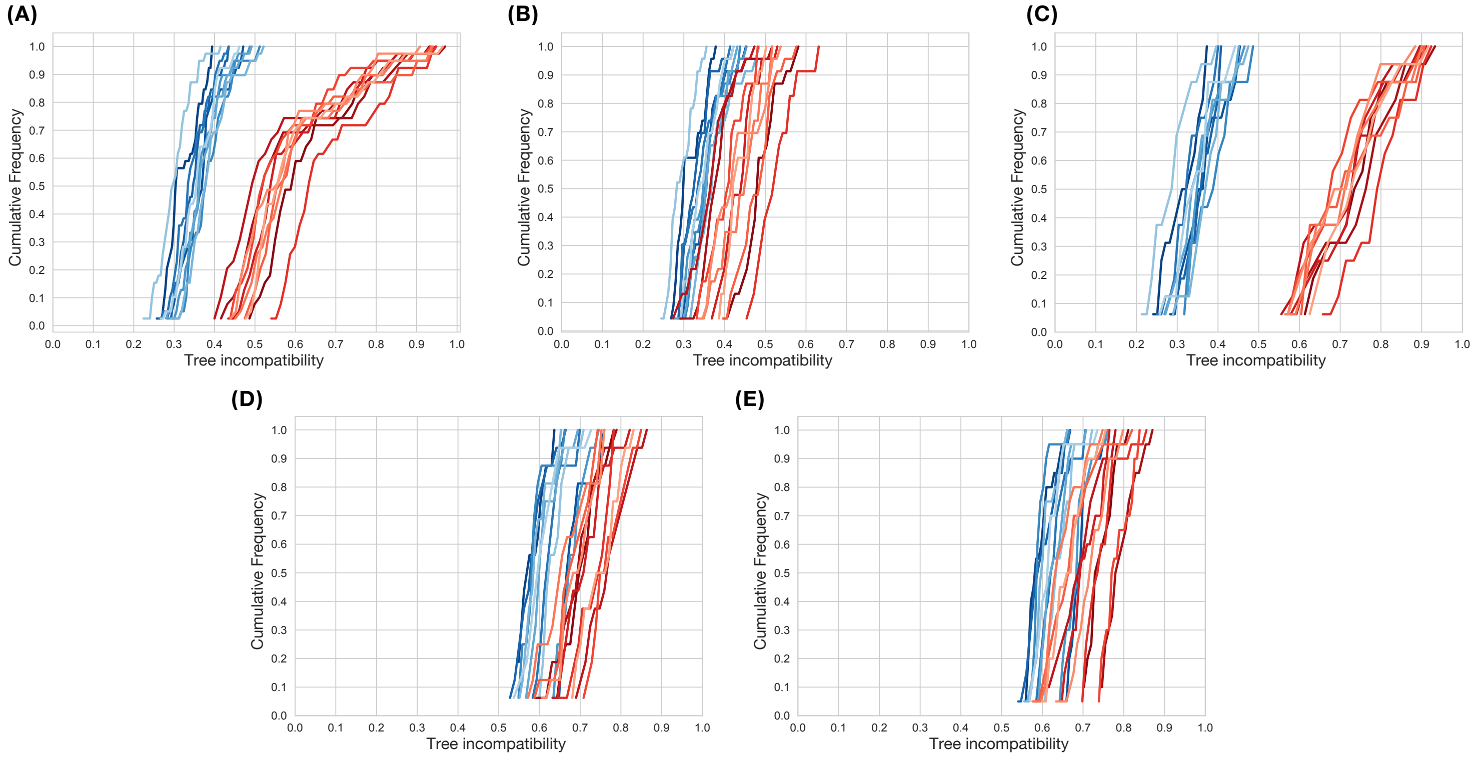


Supplementary Fig. S10 | Distribution of pairwise tree-distances between RefSeq and closed genomes. The pairwise comparisons of tree distances computed using four different metrics (see Methods) are shown. In each case, a matched set of proteins present in the MAG samples and 10 samples from RefSeq are taken to plot comparable distributions. The MAG sample is shown in red. **(A)** Trees for 16 universal ribosomal proteins from 30 closed CPR genomes are compared with trees for the 16 homologues from 10 samples with 30 archaeal RefSeq genomes each (see Table S1 for description of samples). Individual p-values for each comparison are given in Table S9.

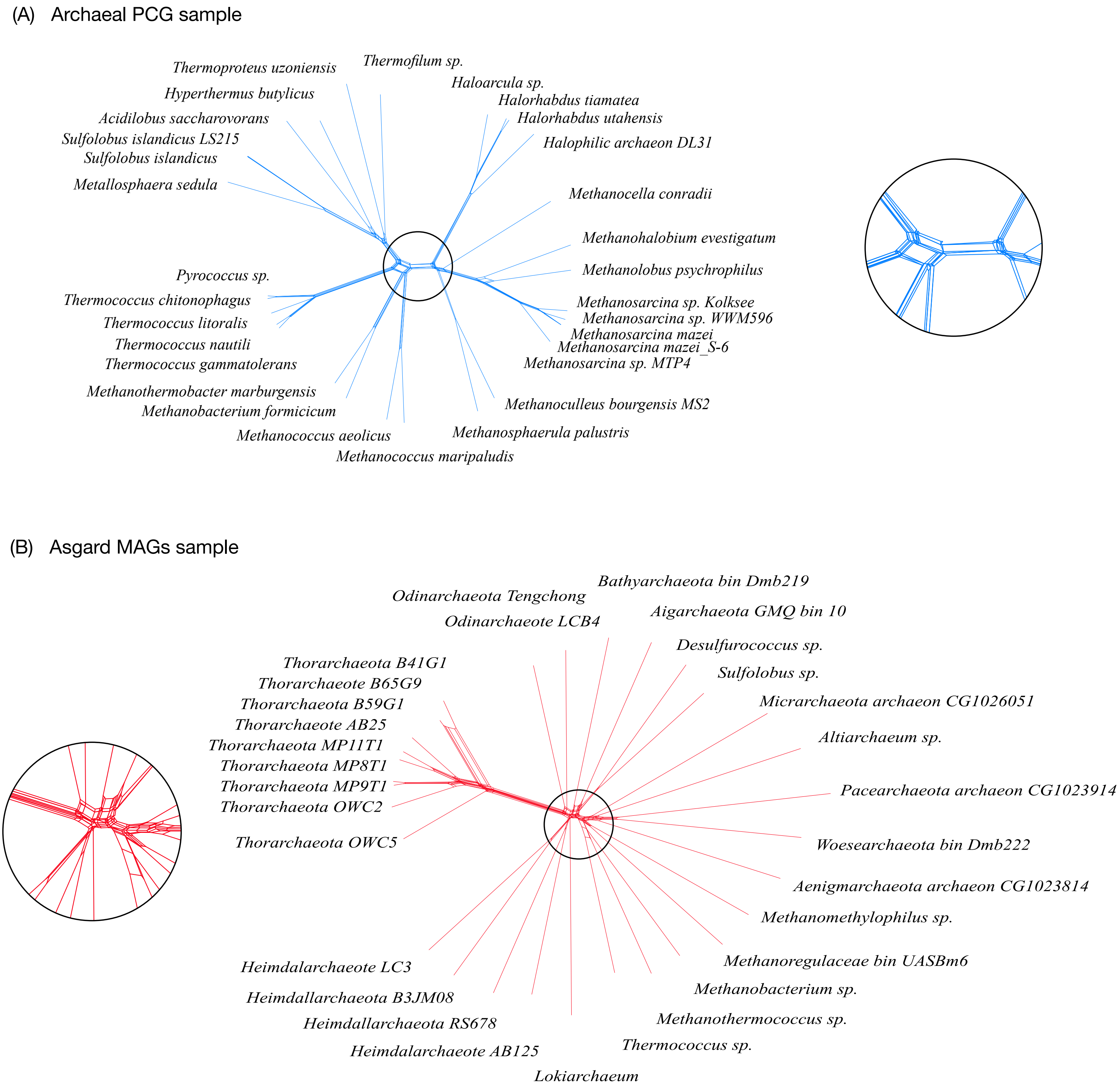


Supplementary Fig. S11 | Neighbor-Nets reconstructed from concatenated alignments of 23 ribosomal proteins for non-Asgard archaeal MAGs: A Neighbor net drawn from a concatenated alignment of 23 ribosomal proteins from archaeal MAGs that do not include Asgards results in a network with a tree-like structure contrary to a Neighbor net for Asgard MAGS (Figure 3b).

1:Sediment microbial community from Chocolate Pots hot springs Yellowstone National Park Wyoming USA Combined Assembly of Gp0156111 Gp0156114 Gp0156117.115; 2:ERR868455_ERR868455.25; 3:ERR599118_ERR599118.63; 4:ERR1726572_ERR1726572.17; 5:SOUTH_ATLANTIC_OCEANS_ERR599165.36; 6:ERR598996_ERR598996.71; 7:ERR598981_ERR598981.9; 8:SOUTH_ATLANTIC_OCEANS_ERR599298.3; 9:Sediment microbial community from Chocolate Pots hot springs Yellowstone National Park Wyoming USA Combined Assembly of Gp0156111 Gp0156114 Gp0156117.57; 10:ERR594345_ERR594345.102; 11:Methanosarcina mazei 1HT25; 12:ERR594338_ERR594338.96; 13:SRR6877511_SRR6877511.51; 14:ERR594321_ERR594321.62; 15:Sediment microbial community from Chocolate Pots hot springs Yellowstone National Park Wyoming USA Combined Assembly of Gp0156111 Gp0156114 Gp0156117.37; 16:ERR594347_ERR594347.23; 17:ERR594345_ERR594345.11; 18:ERR599056_ERR599056.12; 19:SRR6877513_SRR6877513.1; 20:ERR594345_ERR594345.48; 21:ERR599166_ERR599166.64; 22:ERR594347_ERR594347.27; 23:ERR599006_ERR599006.13; 24:ERR598969_ERR598969.10; 25:ERR599062_ERR599062.99; 26:ERR598942_ERR598942.155; 27:ERR315856_ERR315856.59; 28:ERR599007_ERR599007.77; 29:ERR594346_ERR594346.47; 30:SOUTH_ATLANTIC_OCEANS_ERR594335.88



Supplementary Fig. S12 | Tree compatibility scores for samples of tree reconstructed from PCGs and MAGs. Cumulative distribution of tree incompatibility scores within sets of gene trees. In each case every curve represents a set of 30 organisms where the RefSeq samples are shown in shades of blue, the MAG samples are shown in shades of red **(A)** Trees for 39 universal proteins sampled from 10 archaeal RefSeq genomes versus 10 non-Asgard archaeal MAGs. **(B)** Trees for a subset of 23 ribosomal proteins sampled from 10 archaeal RefSeq genomes versus 10 non-Asgard archaeal MAGs. Note that trees for ribosomal proteins from non-Asgard archaeal MAGs (here) are more topologically similar to each other than trees for ribosomal proteins from Asgard MAGs (main text, Fig. 4b). **(C)** Trees for the set of 16 non-ribosomal proteins sampled from 10 archaeal RefSeq genomes versus 10 non-Asgard archaeal MAGs. **(D)** Trees for 16 ribosomal proteins sampled from 10 bacterial RefSeq genomes versus 10 CPR MAGs. **(E)** Trees for 20 ribosomal proteins sampled from 10 bacterial RefSeq genomes versus 10 non-CPR bacterial MAGs.



Supplementary Fig. S13 | Neighbor-Nets reconstructed from concatenated reversed alignments of 23 ribosomal proteins for Archaeal PCGs and asgard archaeal MAGs. (A) The Neighbor-Net of a concatenated reversed alignment of 23 ribosomal proteins in the archaeal PCGs sample shows very little conflict throughout, resulting in a tree-like network. (B) A Neighbor-Net drawn from a concatenated reversed alignment of the same 23 ribosomal proteins from asgard archaeal MAGs results in a network with a star-like structure.