

## Comparative metagenomics of bathypelagic plankton and bottom sediment from the Sea of Marmara

Achim Quaiser, Yvan Zivanovic, David Moreira and Purificación López-García

**Figure S1.** CTD profile of the water column sampled at the Central Basin, Marmara Sea. 0.2-5  $\mu\text{m}$  plankton fractions from 500, 1000 and 1250 m depth were used for SSU-rRNA-based diversity analyses and DNA from the sample of 1000 m depth was pyrosequenced.

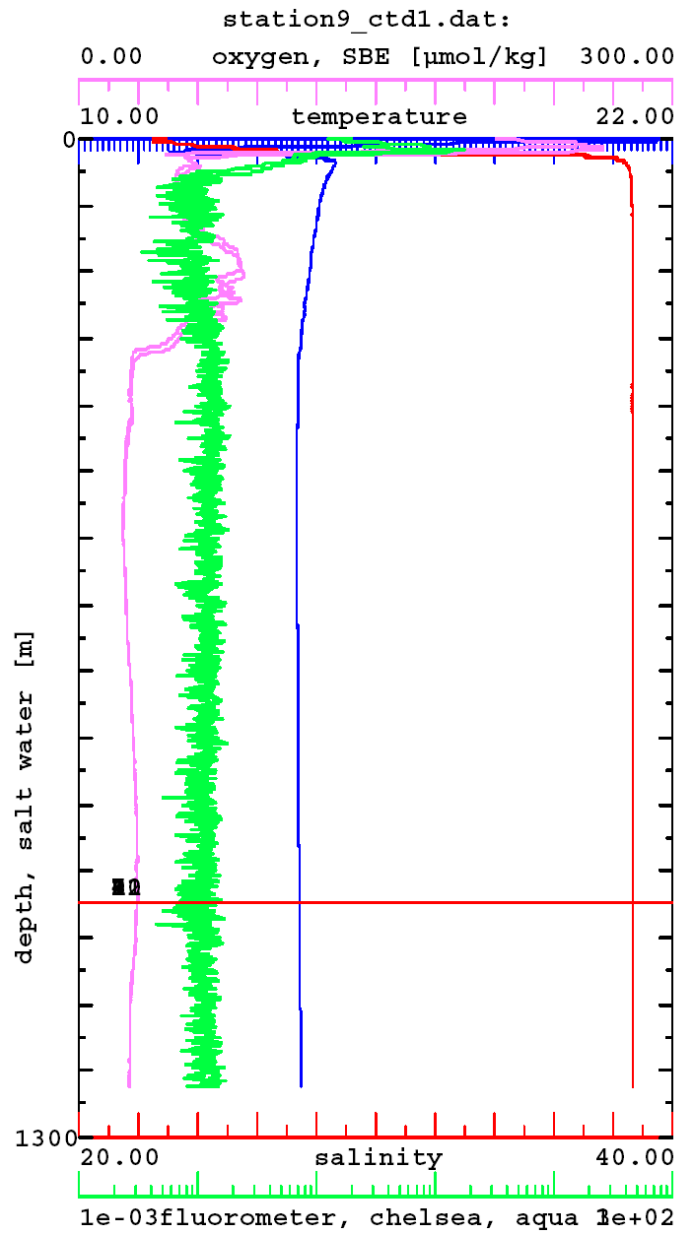
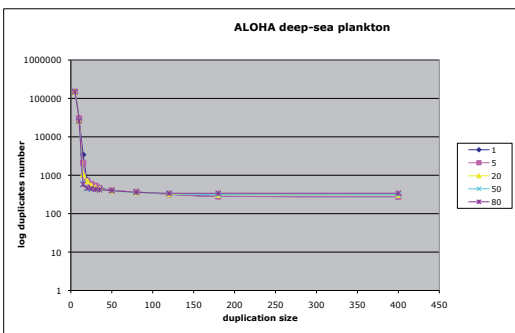
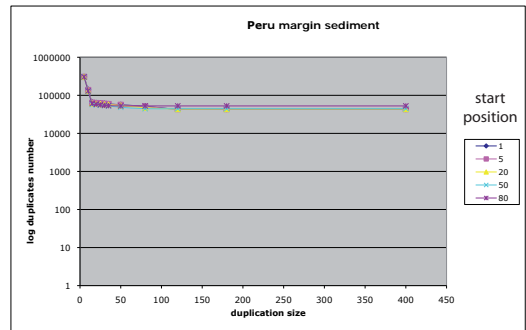
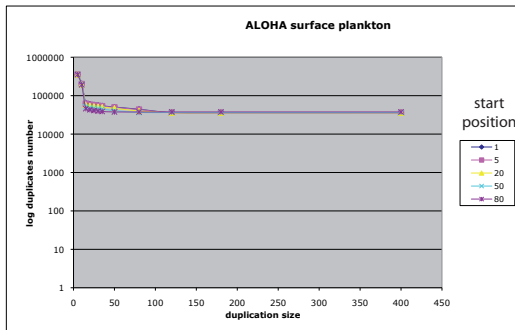
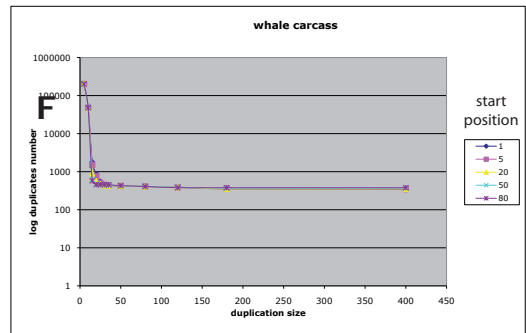
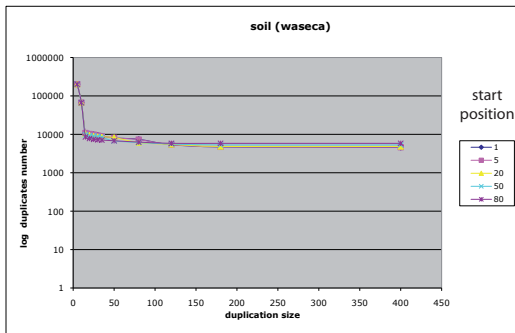
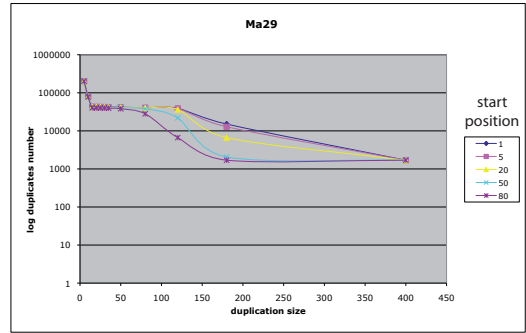
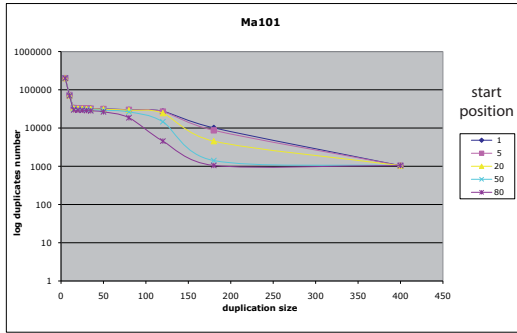
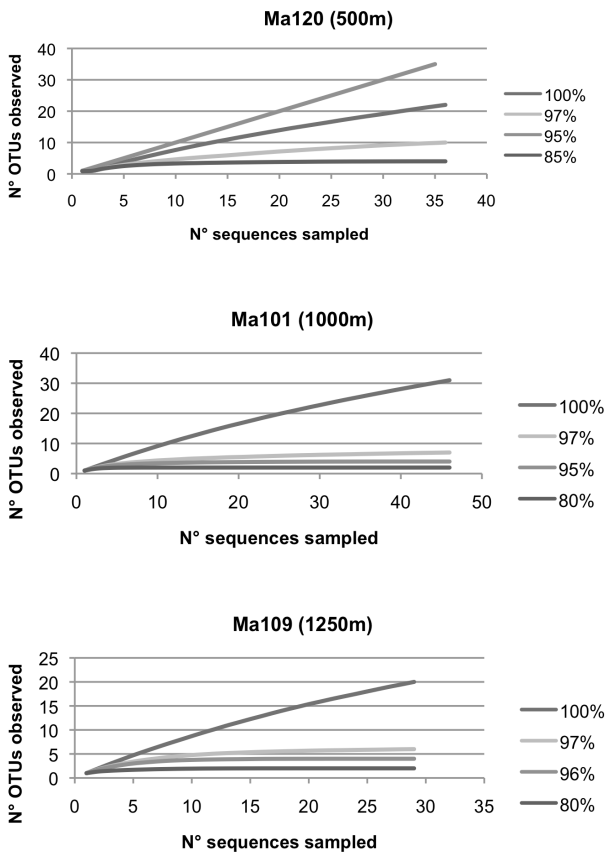


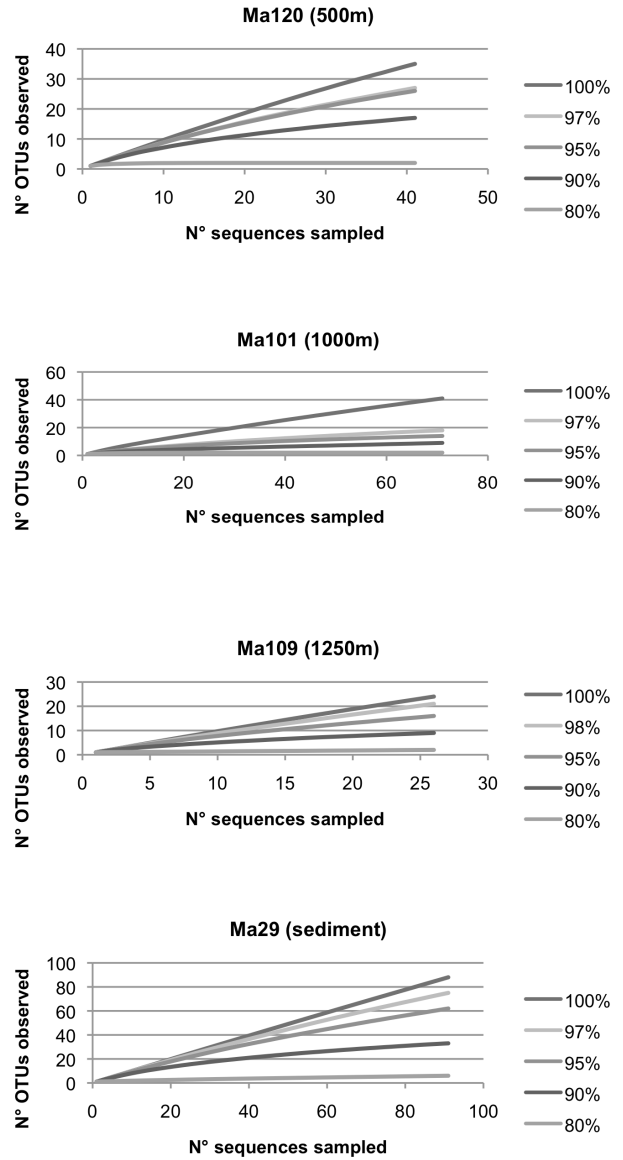
Figure S2. Number of 454 duplicated sequences as function of duplication size and start position



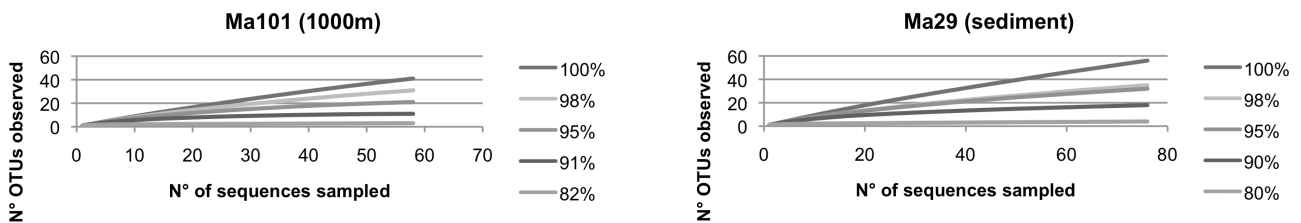
## Archaea



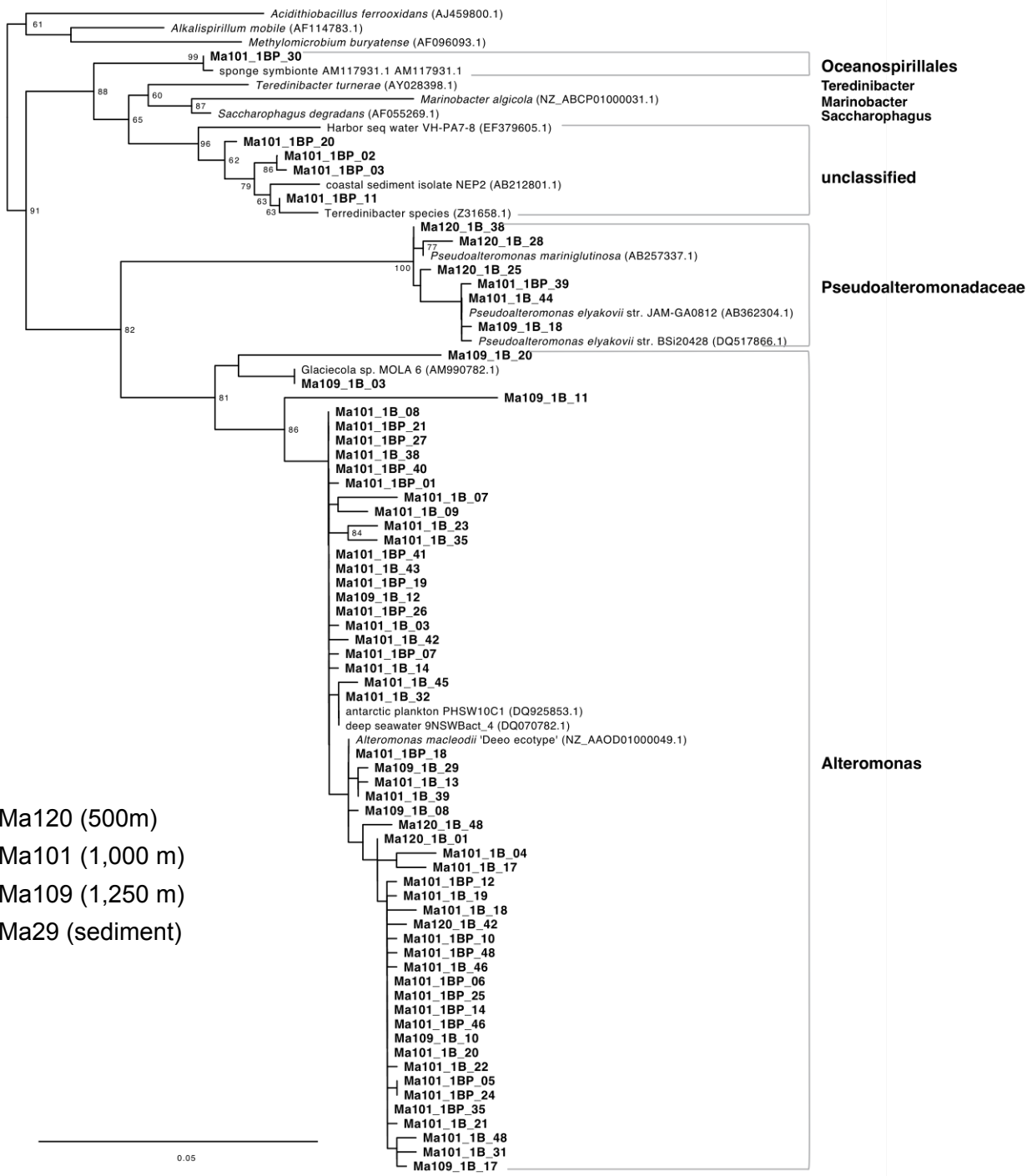
## Bacteria



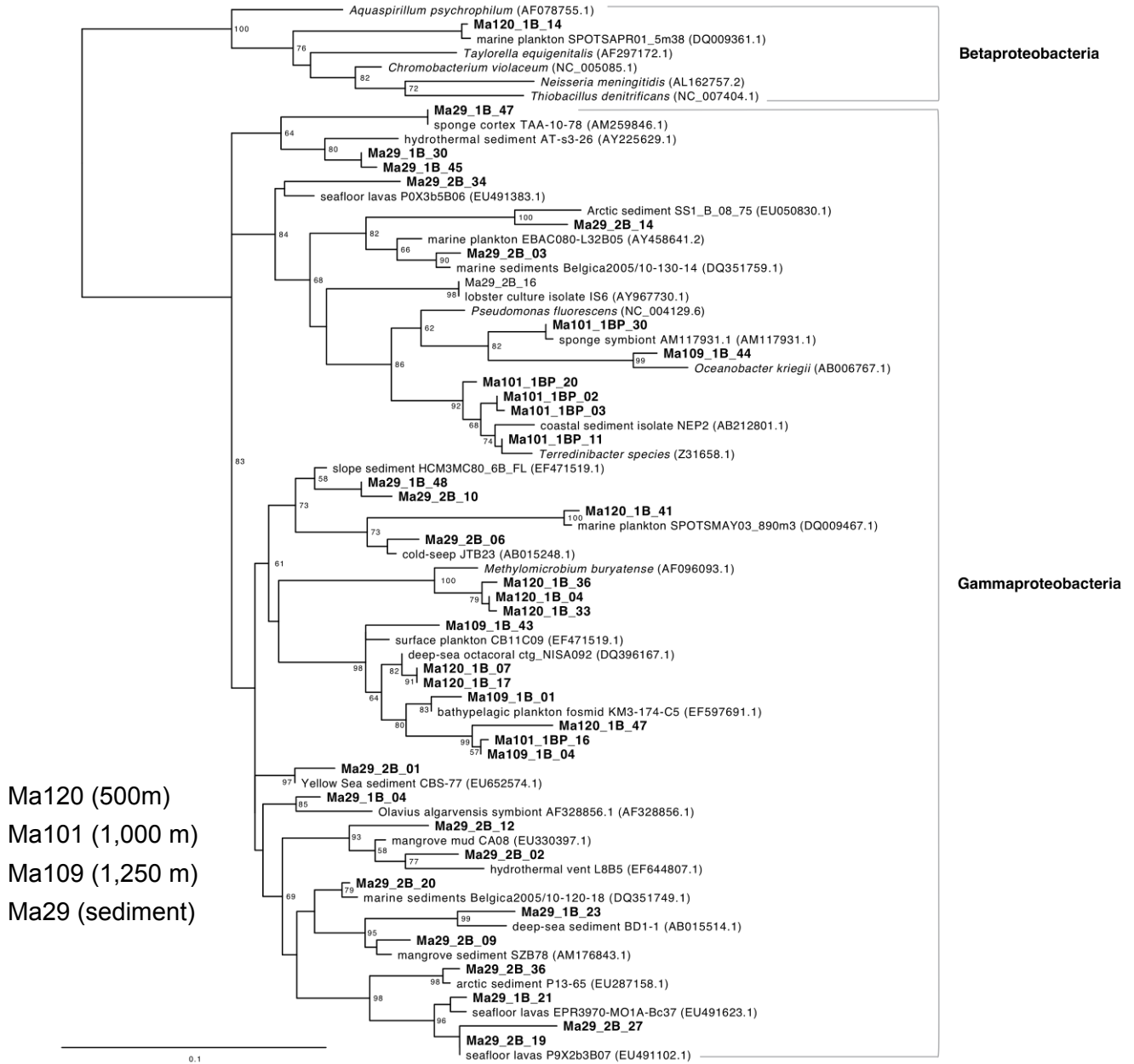
## Eucarya



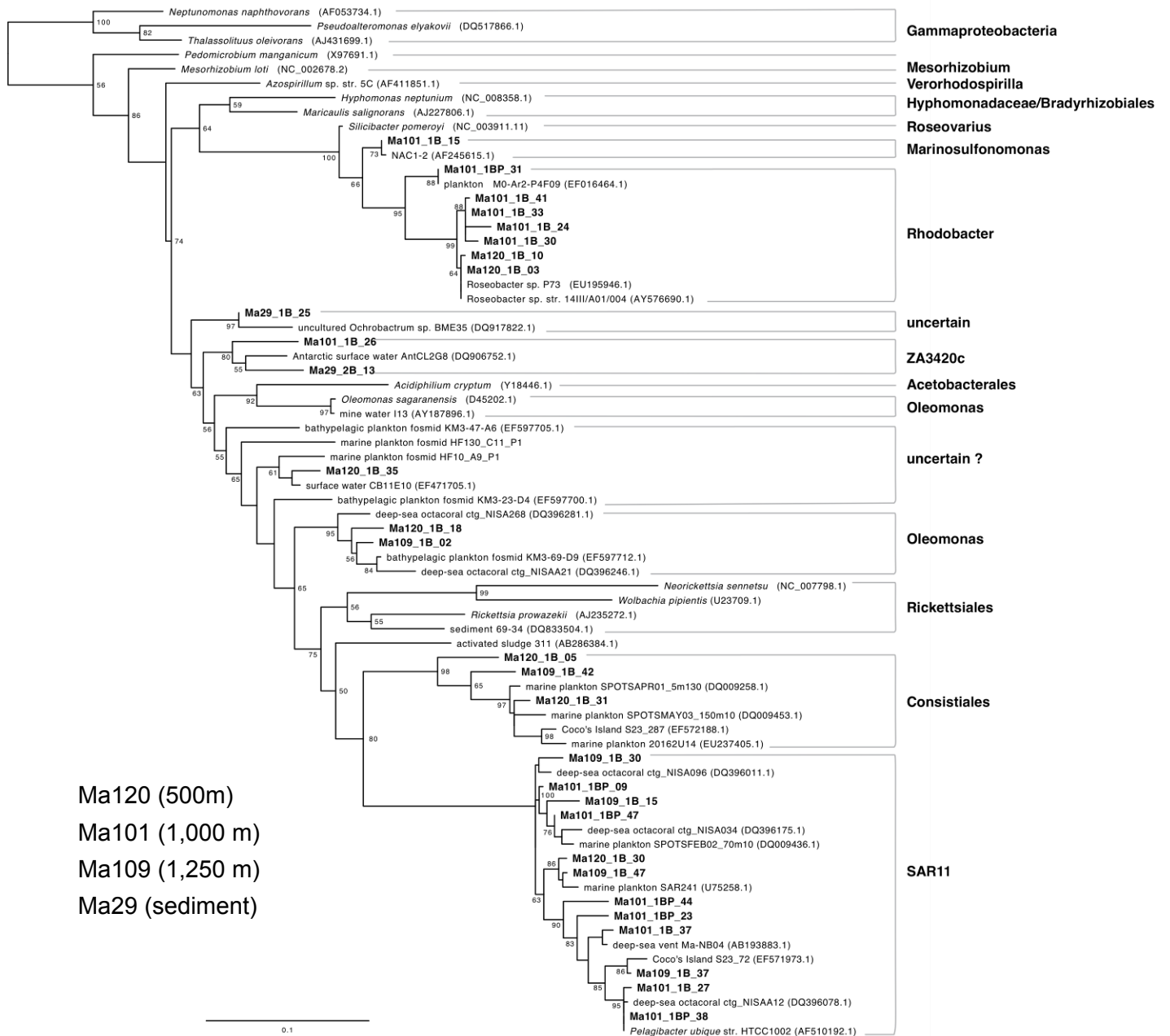
**Figure S3.** Accumulation curves of SSU rRNA gene sequences in the different environmental libraries from meso- and bathypelagic plankton and bottom sediment of the Marmara Sea.



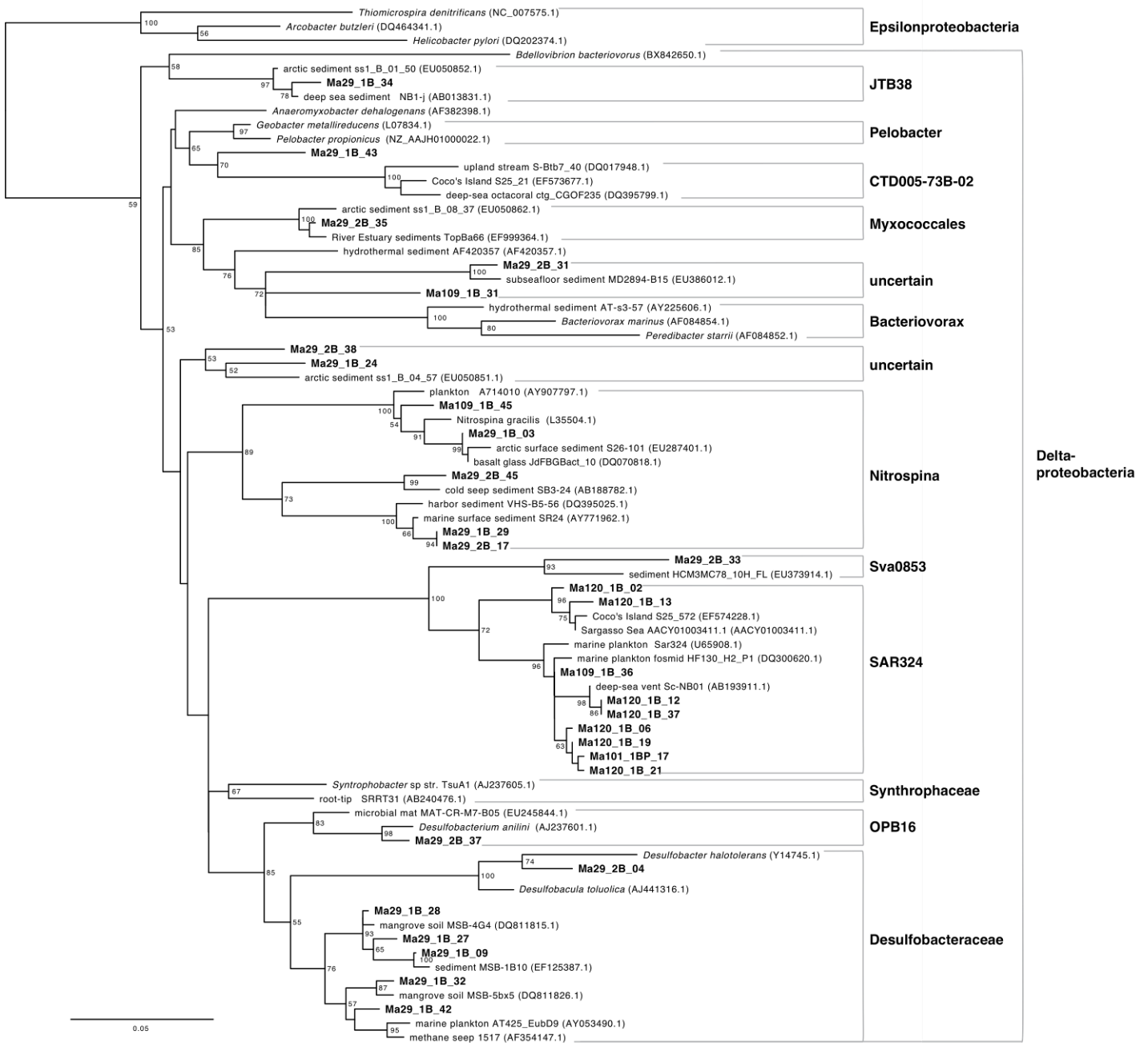
**Figure S4.** Maximum likelihood phylogenetic tree showing the position of the SSU rDNA sequences belonging to the Alteromonadales (Gammaproteobacteria) identified in Marmara samples. The tree was constructed using 677 positions and several other gammaproteobacterial sequences were used as outgroup. Bootstrap values higher than 50% are shown at nodes.



**Figure S5.** Maximum likelihood phylogenetic tree showing the position of the SSU rDNA sequences belonging to the Beta- and Gammaproteobacteria (other than the Alteromonadales) identified in Marmara samples. The tree was constructed using 677 positions. Bootstrap values higher than 50% are shown at nodes.

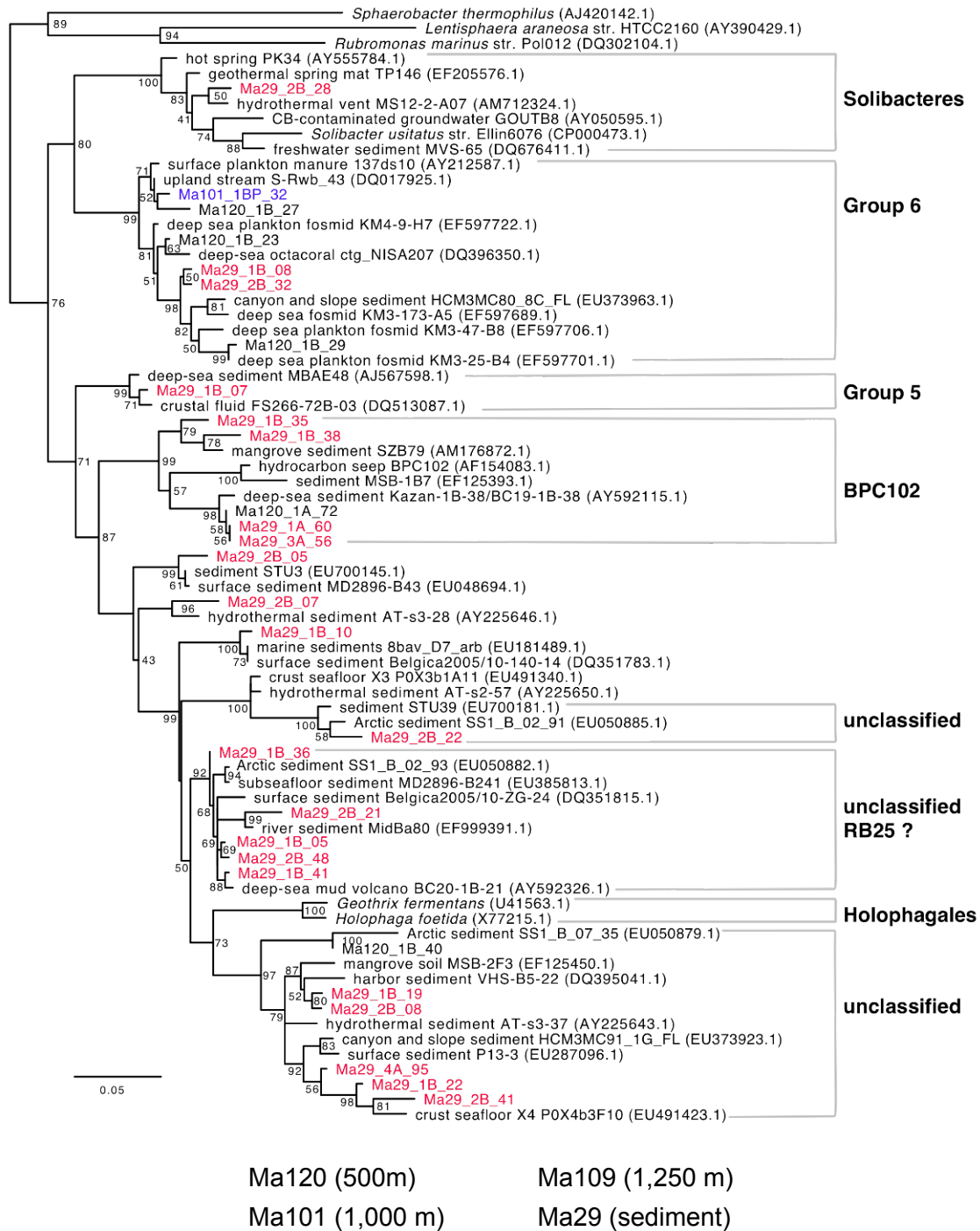


**Figure S6.** Maximum likelihood phylogenetic tree showing the position of the SSU rDNA sequences belonging to the Alphaproteobacteria identified in Marmara samples. The tree was constructed using 677 positions. Bootstrap values higher than 50% are shown at nodes.



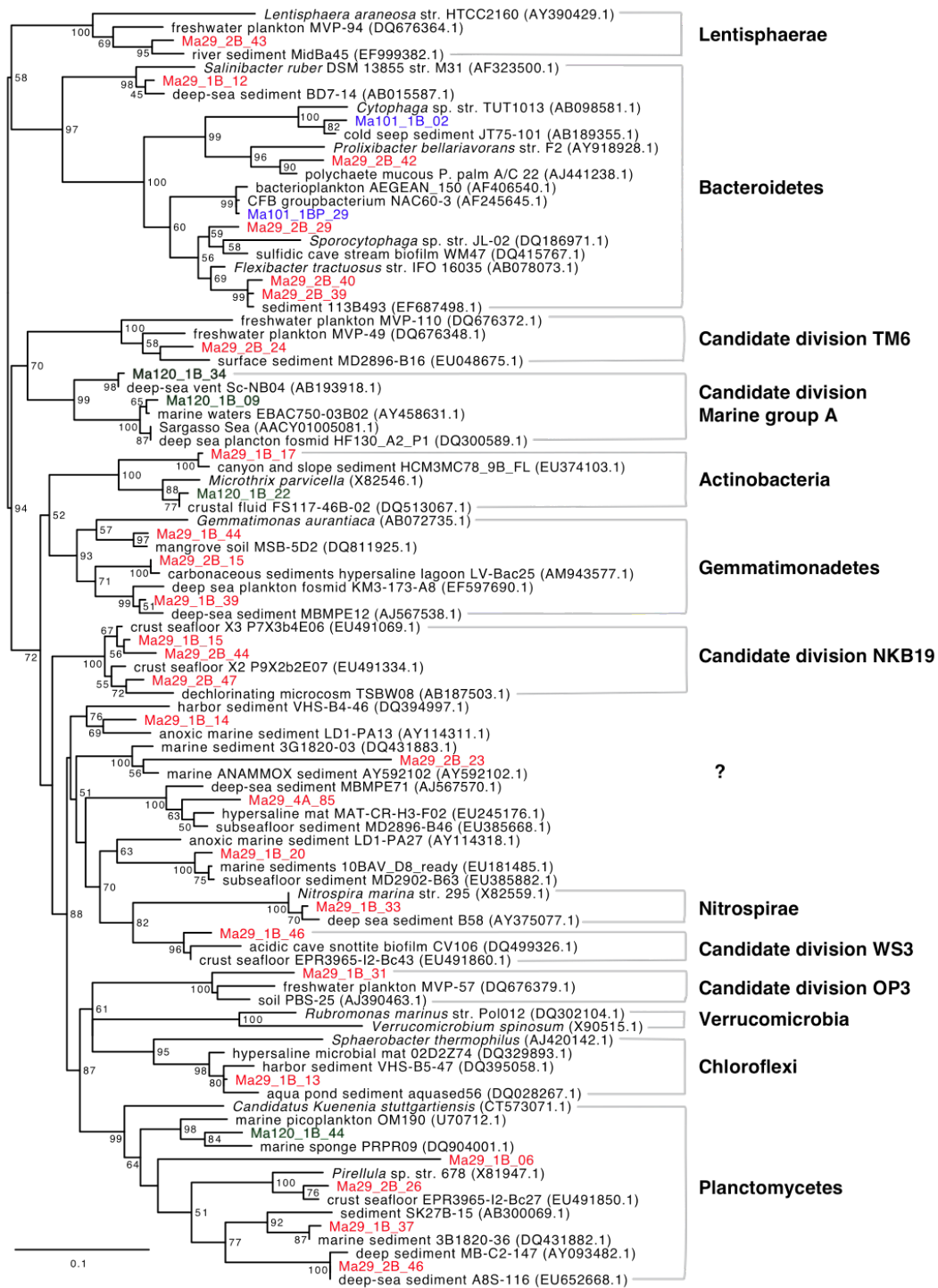
Ma120 (500m)                      Ma109 (1,250 m)  
 Ma101 (1,000 m)                  Ma29 (sediment)

**Figure S7.** Maximum likelihood phylogenetic tree showing the position of the SSU rDNA sequences belonging to the Deltaproteobacteria identified in Marmara samples. The tree was constructed using 677 positions. Bootstrap values higher than 50% are shown at nodes.



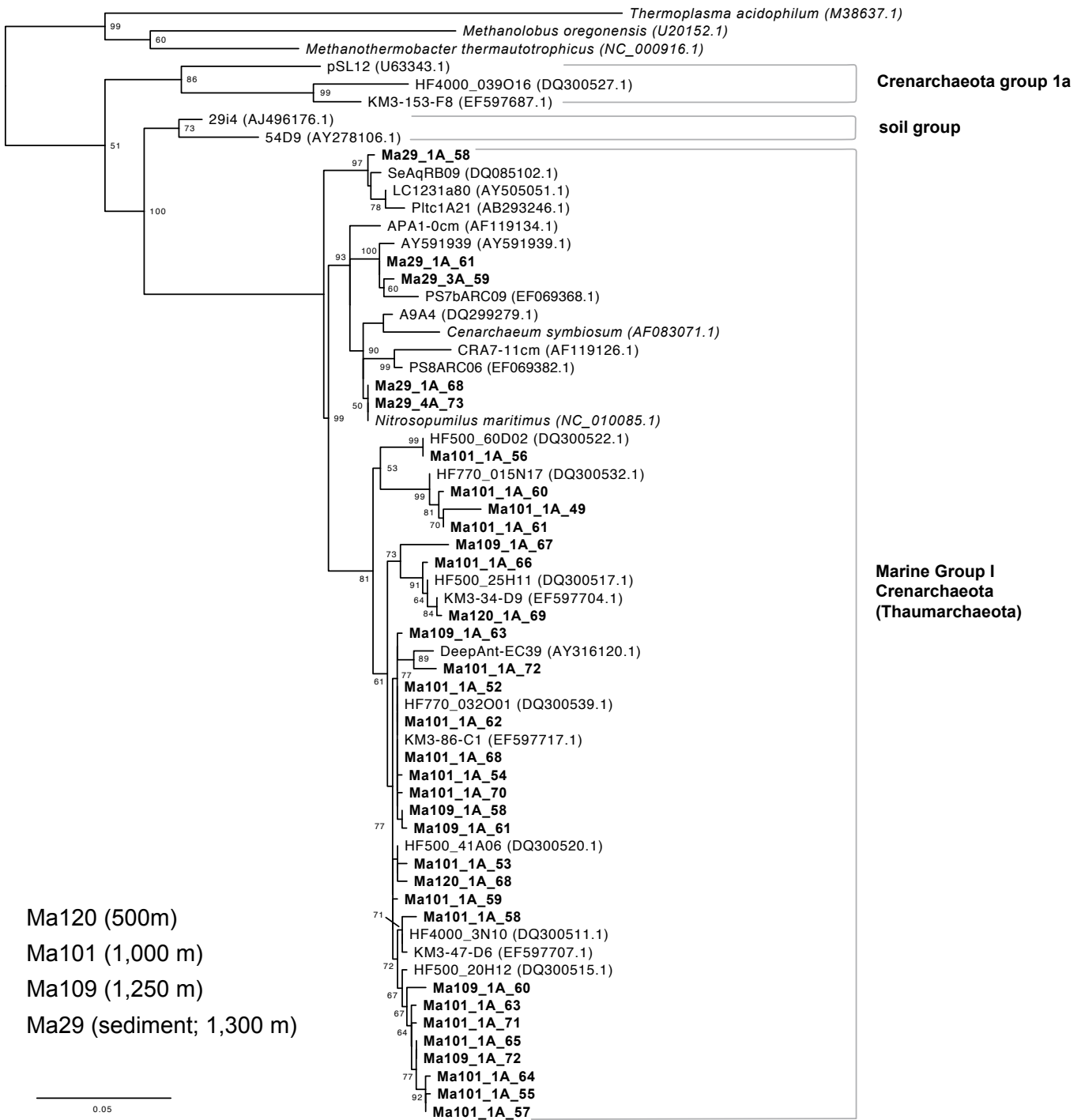
**Figure S8.** Maximum likelihood phylogenetic tree showing the position of the SSU rDNA sequences belonging to the Acidobacteria identified in Marmara samples. The tree was constructed using 636 positions. Bootstrap values higher than 50% are shown at nodes.



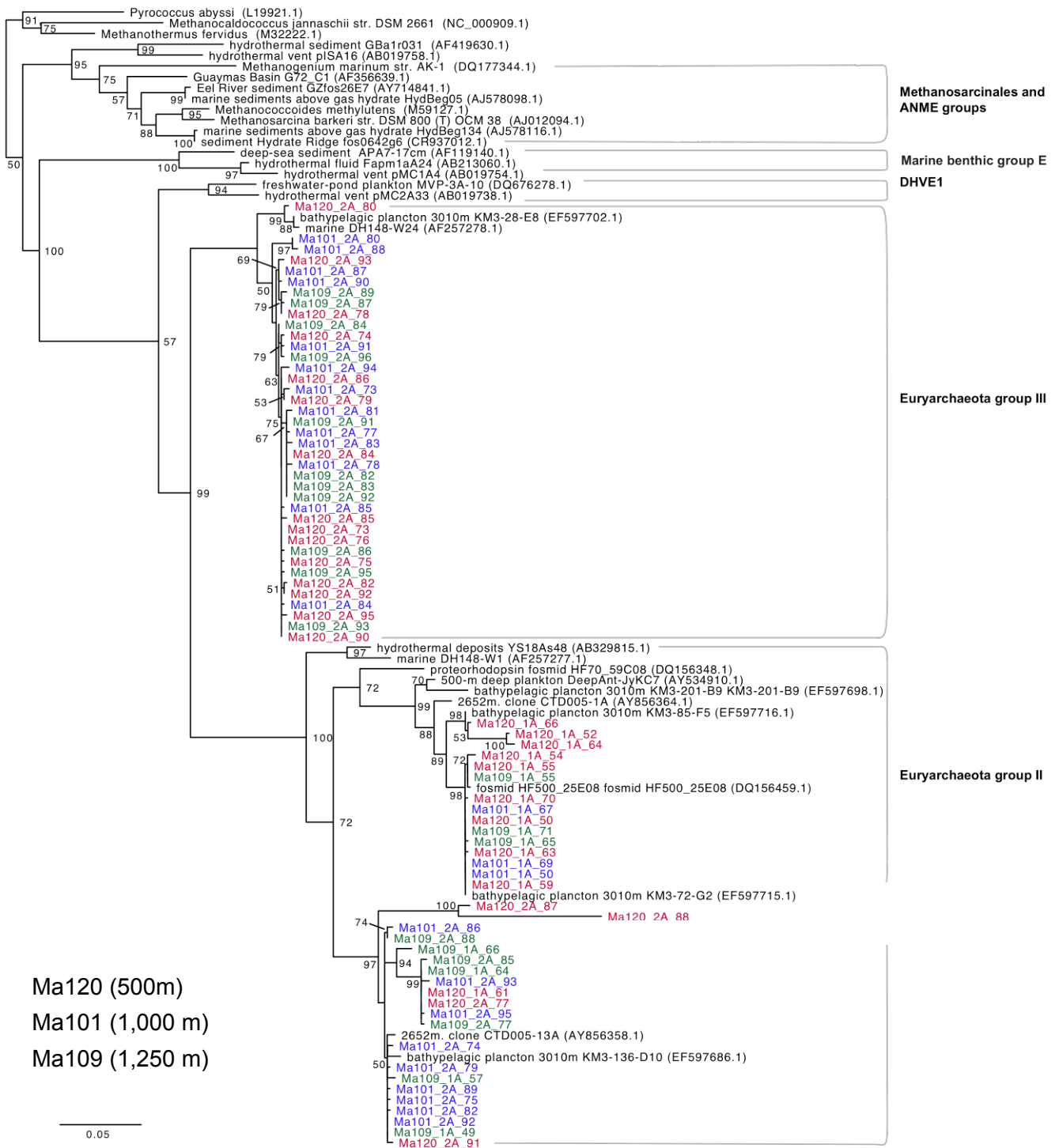


Ma120 (500m)                      Ma109 (1,250 m)  
 Ma101 (1,000 m)                Ma29 (sediment)

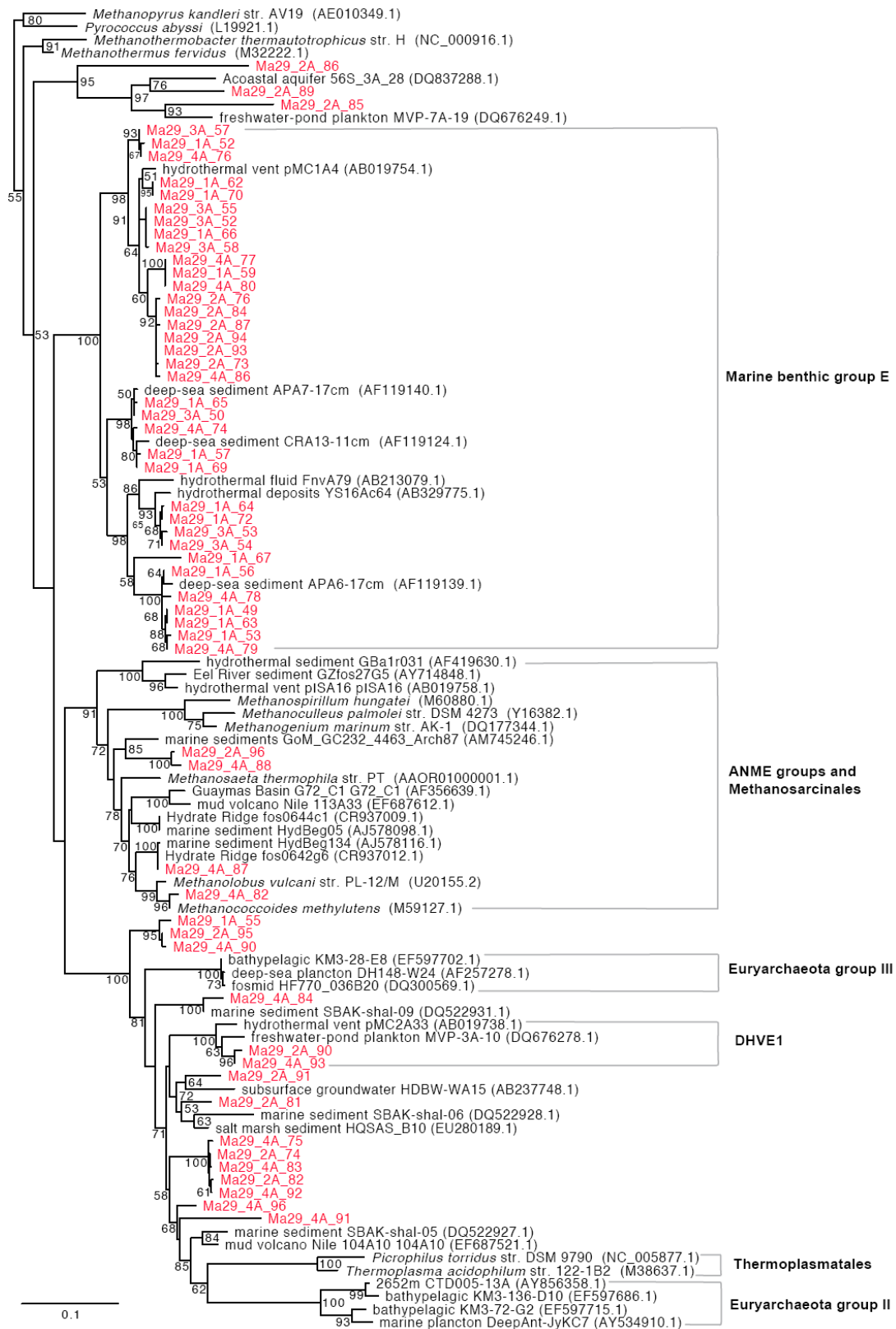
**Figure S9.** Maximum likelihood phylogenetic tree showing the position of the SSU rDNA sequences belonging to the different bacterial taxa other than the Proteobacteria and the Acidobacteria identified in Marmara samples. The tree was constructed using 677 positions. Bootstrap values higher than 50% are shown at nodes.



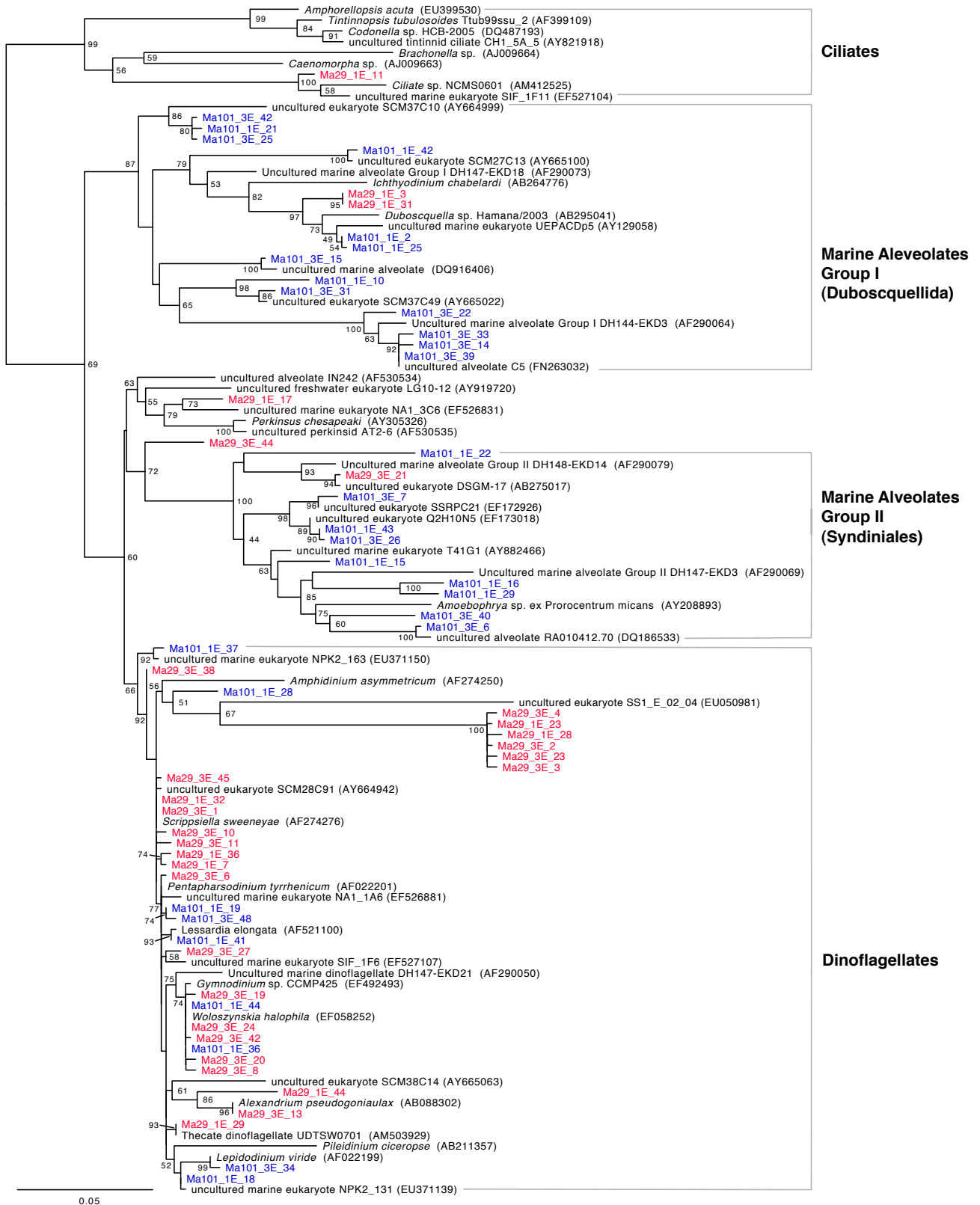
**Figure S10.** Maximum likelihood phylogenetic tree showing the position of the SSU rDNA sequences belonging to the Crenarchaeota / Thaumarchaeota (Group I Crenarchaeota) identified in Marmara deep-sea plankton and sediment. The tree was constructed using 652 positions. Bootstrap values higher than 50% are shown at nodes.



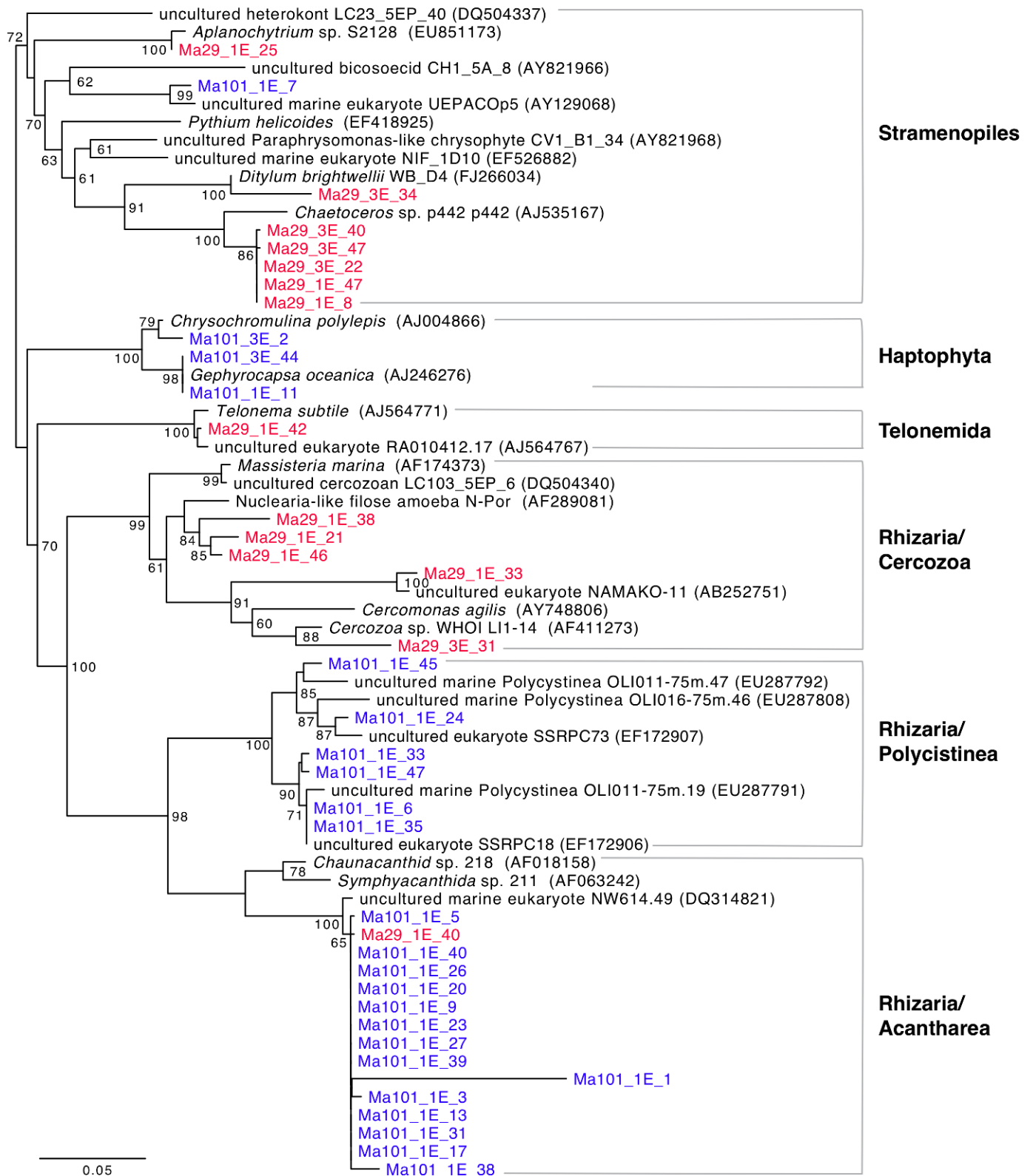
**Figure S11.** Maximum likelihood phylogenetic tree showing the position of the SSU rDNA sequences belonging to Euryarchaeota identified in Marmara deep-sea plankton. The tree was constructed using 699 positions. Bootstrap values higher than 50% are shown at nodes.



**Figure S12.** Maximum likelihood phylogenetic tree showing the position of the SSU rDNA sequences belonging to Euryarchaeota identified in Marmara deep-sea sediment (Ma29). The tree was constructed using 708 positions. Bootstrap values are shown at nodes.

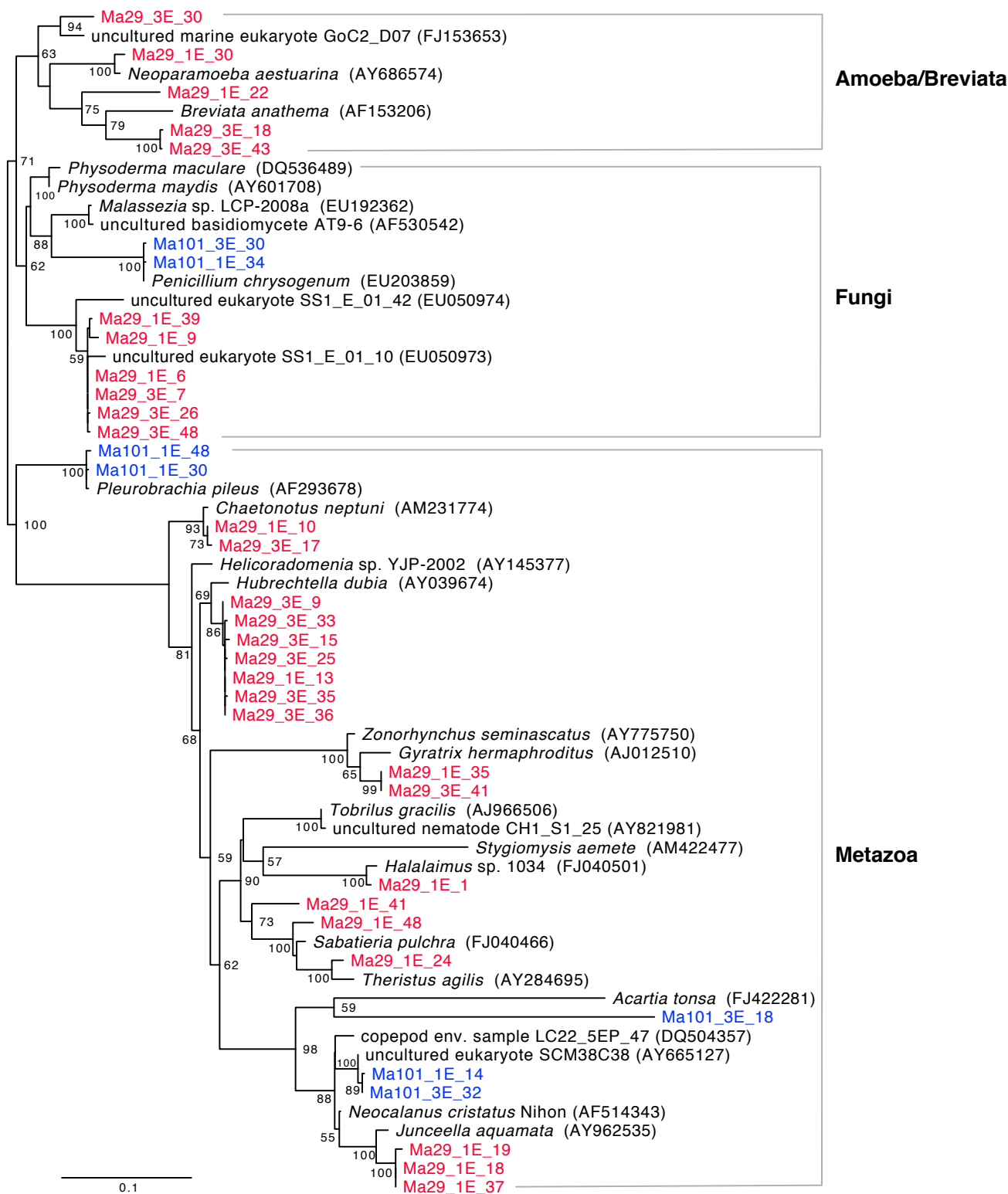


**Figure S13.** Maximum likelihood phylogenetic tree showing the position of the eukaryotic SSU rDNA sequences affiliated to the Alveolates in Marmara samples. The tree was constructed using 783 positions. Bootstrap values higher than 50% are shown at nodes. Ma101, 1,000 m deep plankton; Ma29, sediment.

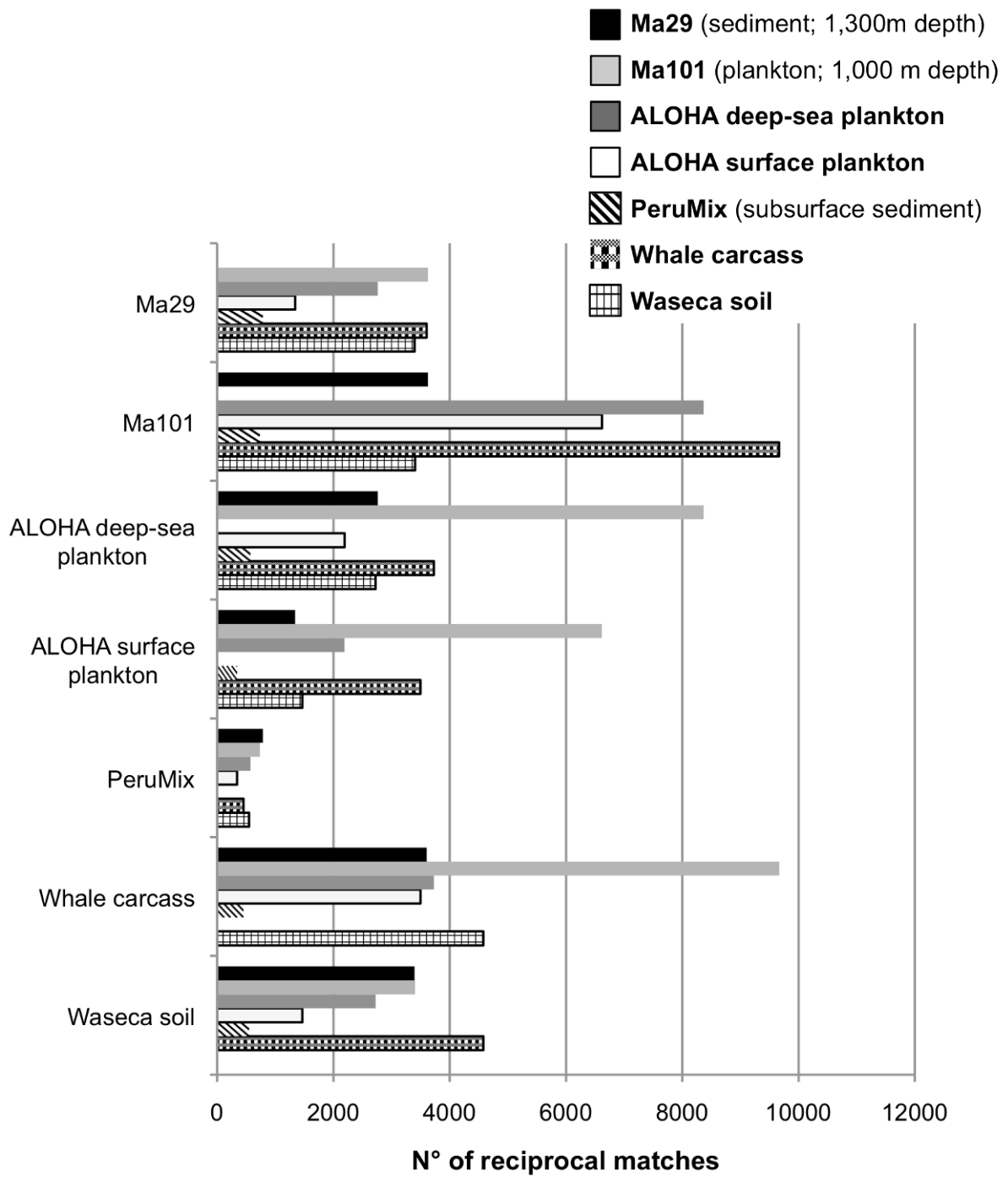


**Figure S14.** Maximum likelihood phylogenetic tree showing the position of the eukaryotic SSU rDNA sequences affiliated to the Telonemida, Rhizaria, Haptophyta and Stramenopiles in Marmara samples. The tree was constructed using 801 positions. Bootstrap values higher than 50% are shown at nodes. Ma101, 1,000 m deep plankton; Ma29, sediment.



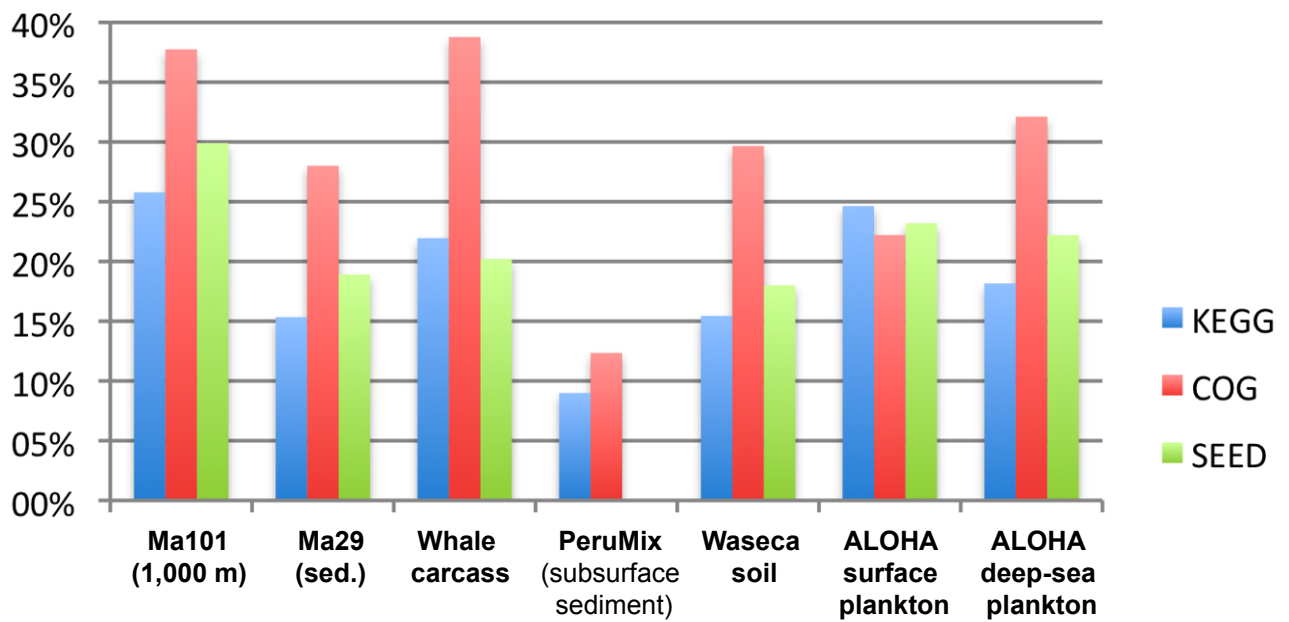


**Figure S15.** Maximum likelihood phylogenetic tree showing the position of the eukaryotic SSU rDNA sequences belonging to the Opisthokonta and Amebozoa in Marmara samples. The tree was constructed using 766 positions. Bootstrap values higher than 50% are shown at nodes. Ma101, 1,000 m deep plankton; Ma29, sediment.

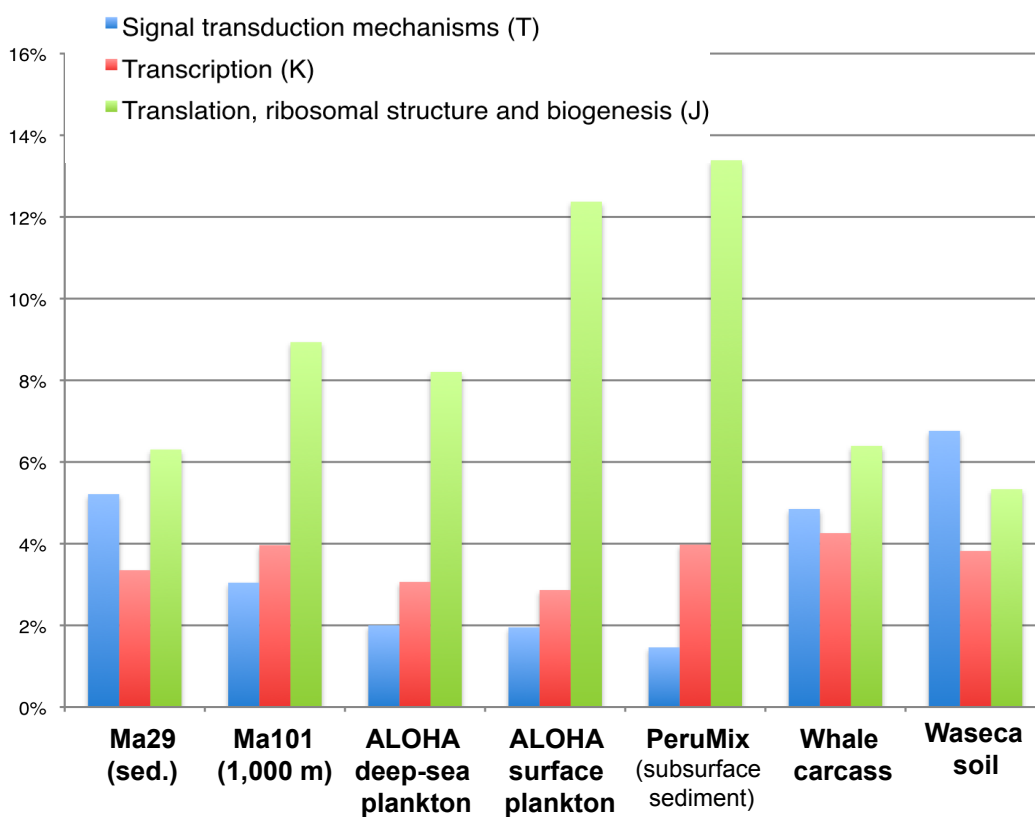


**Figure S16.** Number of shared matches in selected metagenomes as identified by Mummer.

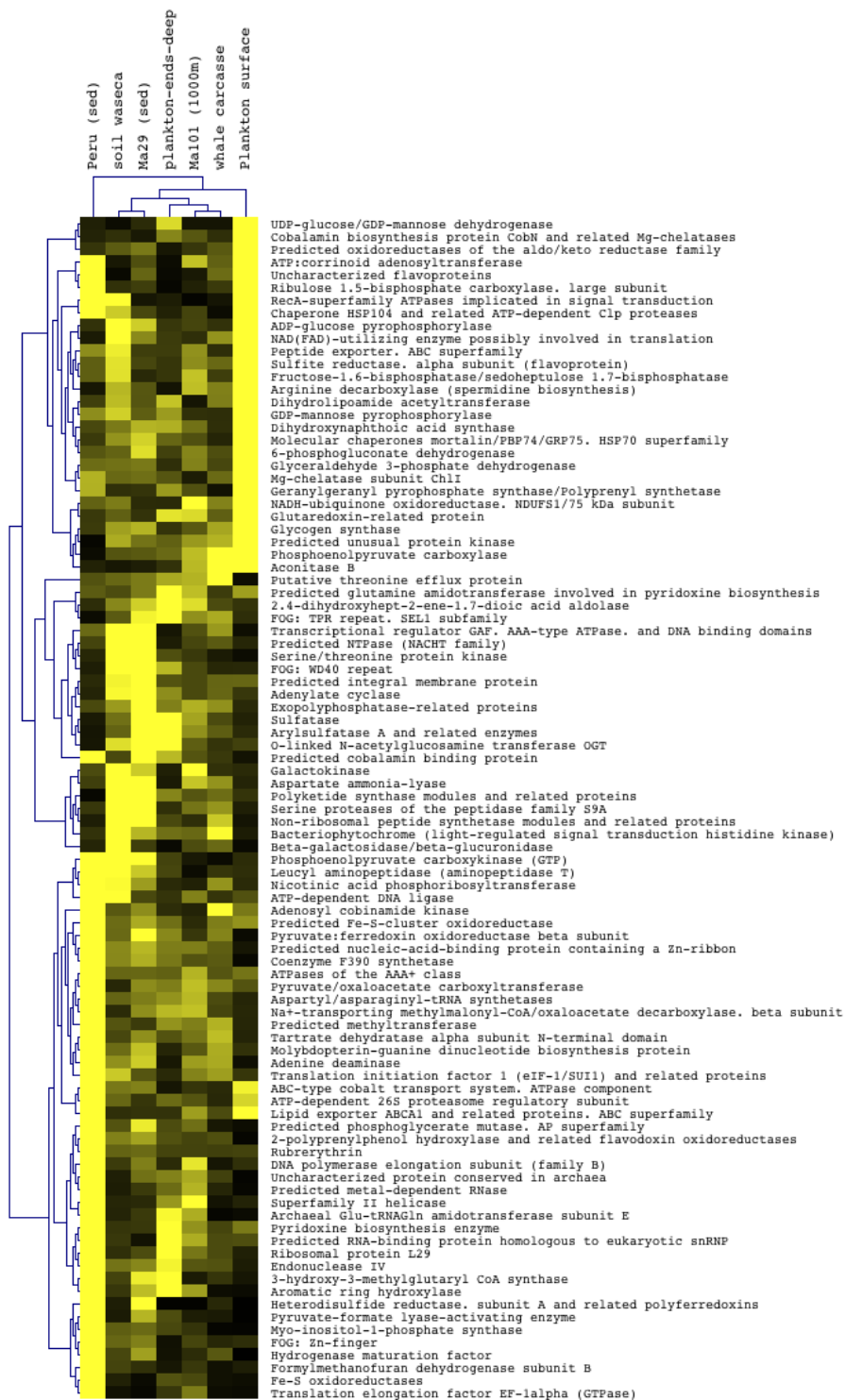




**Figure S17.** Relative proportions of metagenome matches to COG, KEGG and SEED databases normalised with respect to the number of sequences using a cut-off score >40 bits (COG, KEGG) and 1e-05 (SEED).



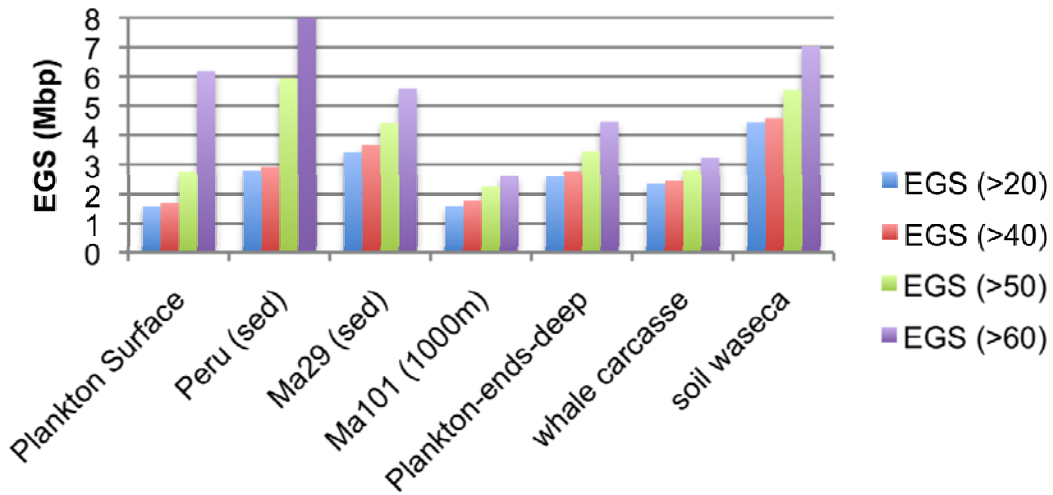
**Figure S18.** Normalised relative proportions of metagenome matches to typical housekeeping genes versus genes involved in signal transduction.



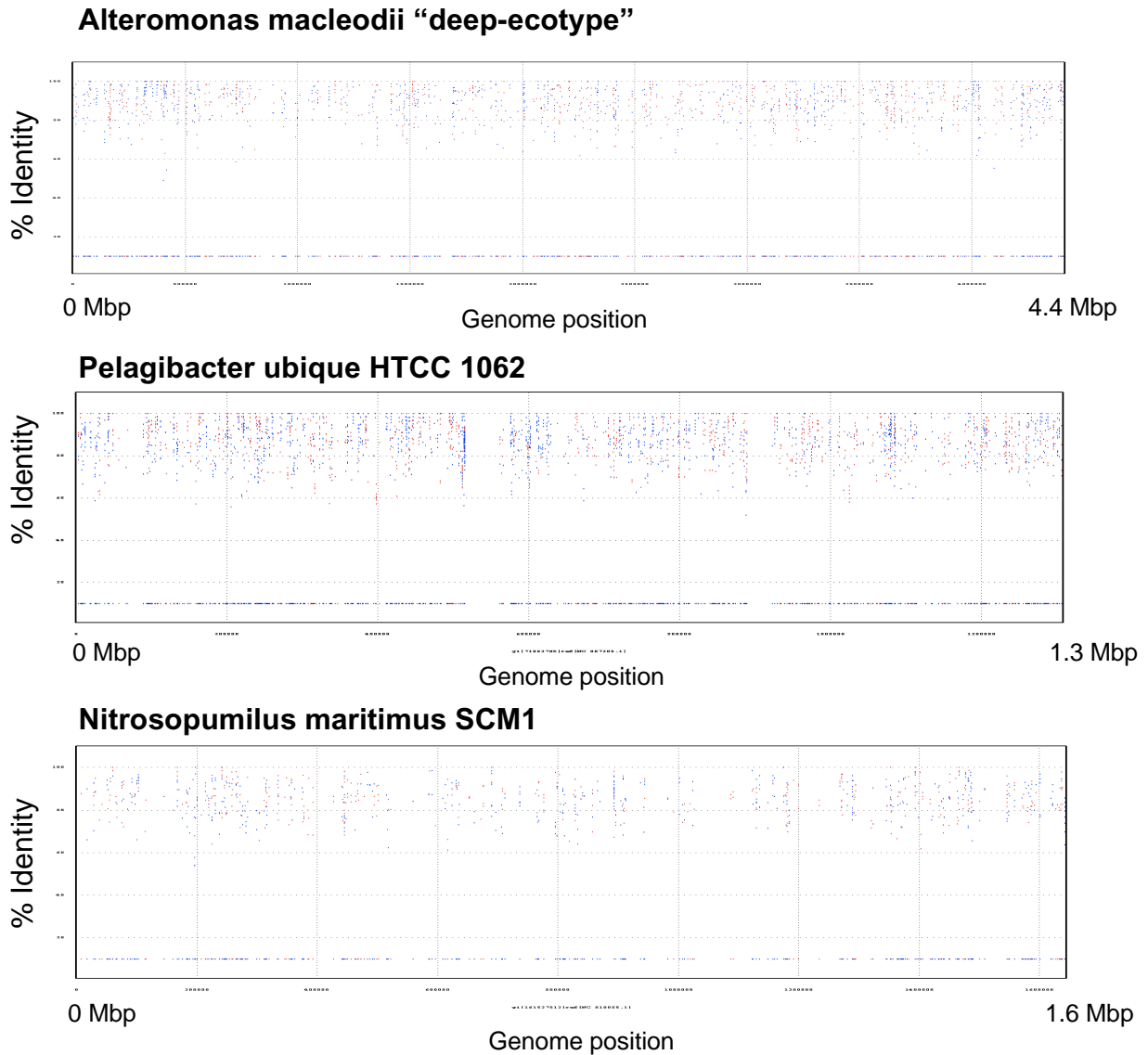
**Figure S19.** Cluster analysis of several metagenomes based on matches to different COG categories normalised to the respective number of sequence fragments. The analysis was done using the software MeV 4.4. Only functional categories having at least one match in all seven metagenomes were retained (1265 COGs). The figure show the 94 most represented functional categories.



**Figure S20.** Cluster analysis of several metagenomes based on matches to different KEGG categories normalised to the respective number of sequence fragments. The analysis was done using the software MeV 4.4. Only functional categories having at least one match in all seven metagenomes or at least 20 matches in one metagenome were retained (148 KEGGSs).



**Figure S21.** Effective genome size (EGS) calculated with different cut-offs using the method by Raes et al. (2007).



**Figure S22.** Recruitments plots (Mummer) of selected abundant genomes in the deep-plankton Ma101 sample. The star represents all the matches on one line.