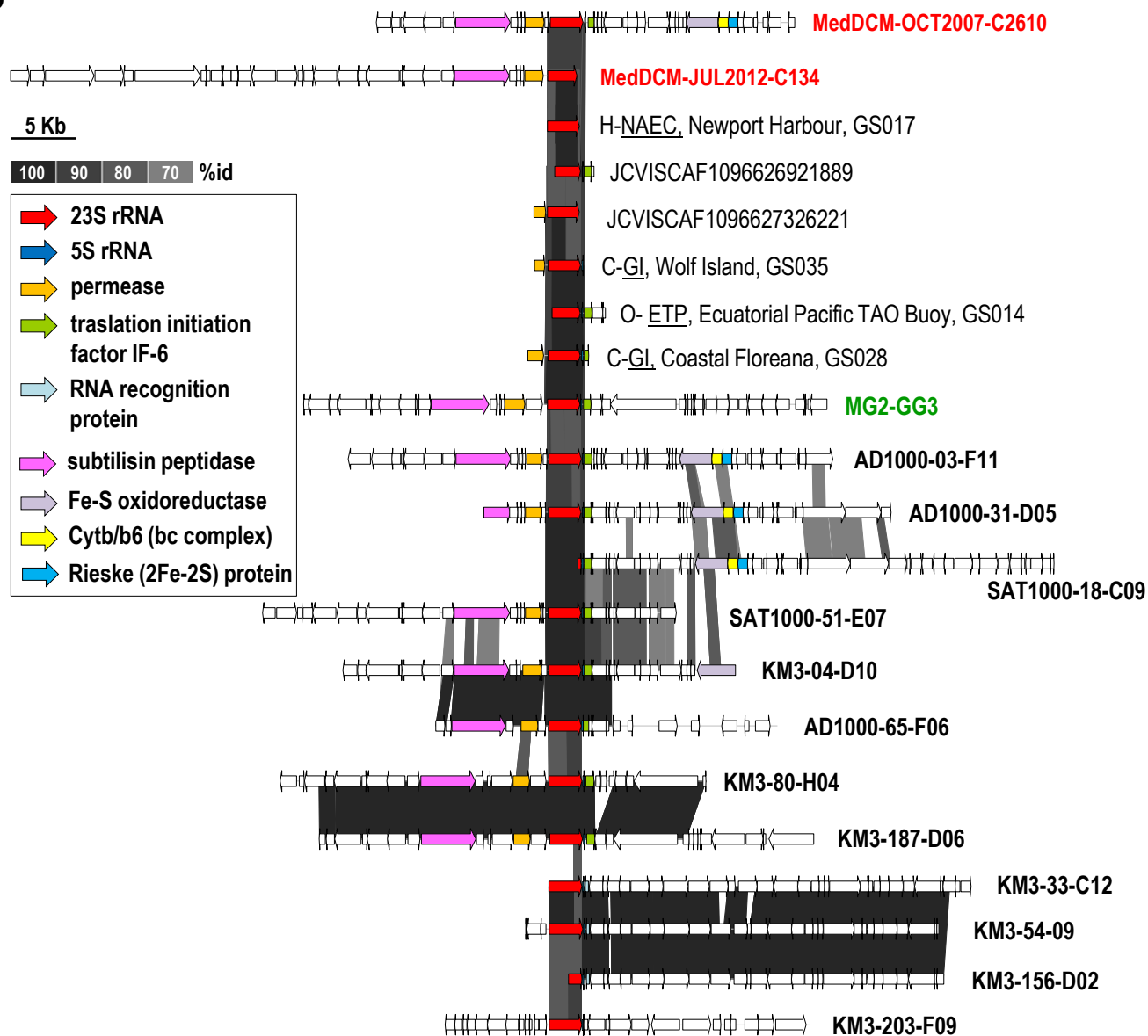


b

Supplementary Figure 1. Comparison of marine low GC thalassoarchaeal contigs containing rRNA genes (bold and red) to scaffolds from the Global Ocean Sampling (GOS) dataset and other MGII described in literature (MG2-GG3 in bold and green) (using BLASTN). The oceanic habitat (C-Coastal, CRA-Coral Reef Atoll, O-Open Ocean, E-Estuary), sampling locations (NAEC-North American East Coast, GI-Galapagos Islands, ETP-Eastern Tropical Pacific, PA-Polynesia Archipelagos, SS-Sargasso Sea, CS-Caribbean Sea) and the GOS dataset identifier are shown next to each GOS scaffold. All ribosomal RNA genes are highlighted in red color and sequence identity amongst the contigs is shown in shades of grey (see color scale). **(a)** Contigs containing 16S rRNA. **(b)** Contigs containing 23S rRNA.