

SM3. Methodology details for the functional prediction database

The functional prediction used bacterial and archaeal taxonomy as entries of an edited database and the assignment to potential functions was performed using the FAPROTAX tool (Louca et al., 2016). The database FAPROTAX_1.2.4 (Louca et al., 2021), available at <http://www.zoology.ubc.ca/louca/FAPROTAX>, was amended with the addition of 2,051 taxonomic entries (from phylum to species), and 22 new ecological roles based on cultured from IJSEM database (Barberan, 2016), LPSN database (Parte, 2018); function and gene searches on NCBI (<https://www.ncbi.nlm.nih.gov/search/>); SILVA database (Yilmaz et al., 2014); and genome and metagenome-assembled genome references. Functions were also revised to allocate interconnected ones (taxa associated with multiple functions, and levels of functions). Our edited database is available at <https://doi.org/10.5281/zenodo.7226298>. The following 22 new functions were added: iron oxidation, hydrogen sulfide production, dimethylsulfoniopropionate demethylation, methasulfonate oxidation, methanotrophy by aerobic ammonia oxidation, CO oxidation, aerobic acetogenesis, syntrophy with hydrogenotrophs, methanogenesis with phosphonates by cyanobacteria, methanogenesis by disproportionation of methyl groups, methoxydotrophic methanogenesis, anaerobic methanotrophy, anaerobic methanotrophy syntrophy, methanotrophy facultative methylotrophy, anaerobic hydrocarbon degradation, EPS marine snow formation, marine oil snow formation, methanogenic degradation of crude oil alkanes, alkane disproportionation, potassium solubilization, phosphate solubilization. Six habitats were added as an ecological characteristic which comprises commonly found taxa or microbial bioindicators, as indicated in the habitat name: habitat detection carbonate crusts active seeps, habitat detection carbonate crusts, habitat detection active seeps, habitat bioindicator seeps, habitat detection sediment nodules, habitat detection marine bottom water.

In addition, new taxonomy entries were added to the 21 functions previously present (Louca et al., 2021): iron respiration, aerobic anoxygenic phototrophy, sulfate respiration, aerobic ammonia oxidation, nitrate reduction, reductive acetogenesis, methanogenesis by CO₂ reduction with H₂, methanogenesis using formate, acetoclastic methanogenesis, methnogenesis by reduction of methyl compounds with H₂, hydrogenotrophic methanogenesis, methanogenesis, methanotrophy, methylotrophy, aromatic hydrocarbon degradation, aromatic compound degradation, oil bioremediation, aliphatic non-methane hydrocarbon degradation, non-methane hydrocarbon degradation.