
Molecular hydrogen in seawater supports growth of diverse marine bacteria

In the format provided by the
authors and unedited

Supplementary information

Table S1 (xlsx). Trace gas oxidation rates and power calculations for Port Phillip Bay and Munida Transect samples.

Table S2 (xlsx). Sequencing statistics of the 15 newly sequenced metagenomes.

Table S3 (xlsx). Abundance of metabolic marker genes in the short read metagenomic and metatranscriptome data. For the newly sequenced metagenomes, this includes the calculated average copies per organism for each marker gene, average copies per organism for each hydrogenase subgroup, and a list of short-read hits and their corresponding match in the database. For the *Tara* Oceans dataset, this includes the calculated average copies per organism, transcript levels, and expression ratios (RNA:DNA ratio) for each marker gene and hydrogenase subgroup.

Table S4 (xlsx). Marker genes identified in the assembled contigs in the newly sequenced metagenomes. This lists the metabolic marker genes annotated across all contigs over 2000 bp (including both unbinned contigs and those that were assigned to a bin).

Table S5 (xlsx). Summary and annotation of metagenome-assembled-genomes (MAGs). For the newly sequenced metagenomes, this includes taxonomic information for each MAG, relative abundance of each MAG per sample calculated by CoverM for the newly sequenced genomes, a summary of metabolic marker genes identified in each MAG, and a full list of metabolic marker genes identified in the MAGs with

alignment information and protein sequences. For the *Tara* Oceans metagenomes, this includes taxonomic information and metabolic summary for each MAG.

Table S6 (xlsx). Spearman's correlation analyses based on metadata from the *Tara* Oceans metagenomes and metatranscriptomes. Raw and false discovery rate adjusted two-sided p-values are provided, and were calculated using the asymptotic *t* approximation (R function `cor.test()`, `method = "spearman"`, `exact = FALSE`).

Figure S1. Map of the sites sampled in this study.

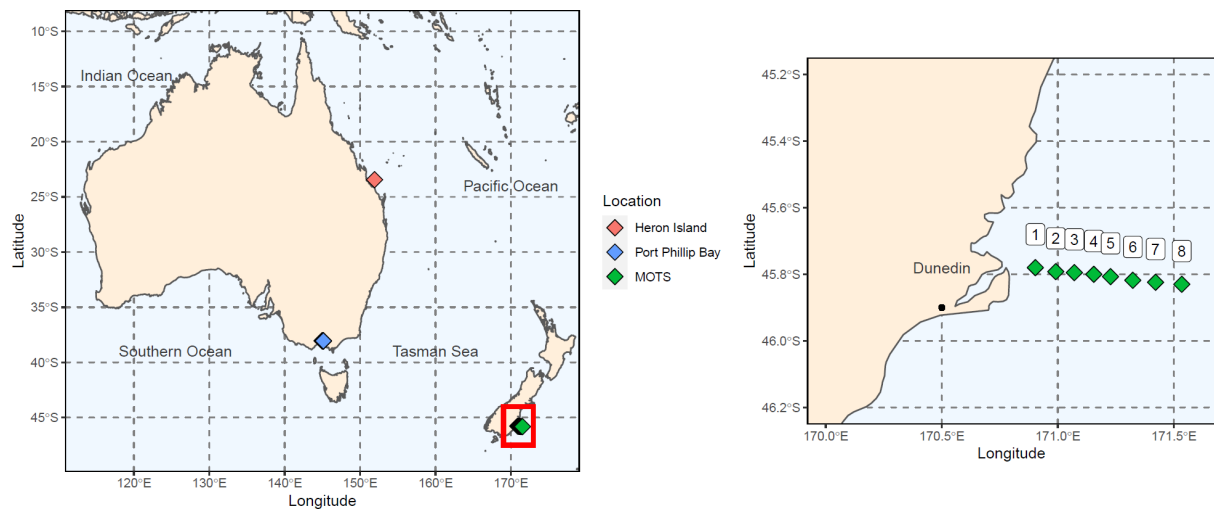


Figure S2. Eight stations sampled in the Munida Observation Time Series. The transect extends 65 km in an Eastern direction from Taieri Head, Otago. It spans neritic waters (NW; black; two stations), a transitional subtropical frontal zone (STW; yellow; two stations), and subantarctic waters (SAW; blue; four stations). Temperature and salinity data reflect the transitions between the three zones.

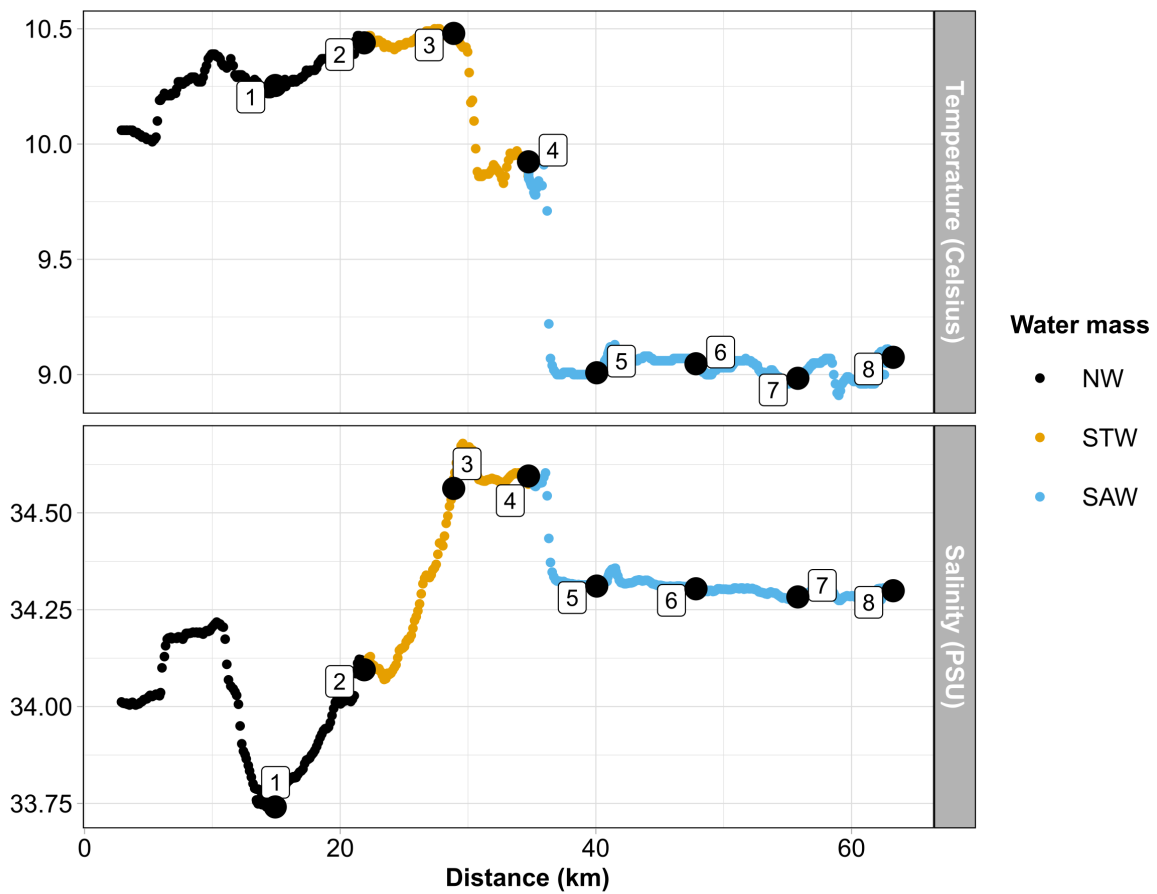


Figure S3. Trace gas oxidation by marine microbial communities in Heron Island. Samples were spiked with 10 ppm CO and H₂ and incubated in the dark. Relative concentrations were calculated by dividing concentrations at later times with the measured concentration at the start of each incubation.

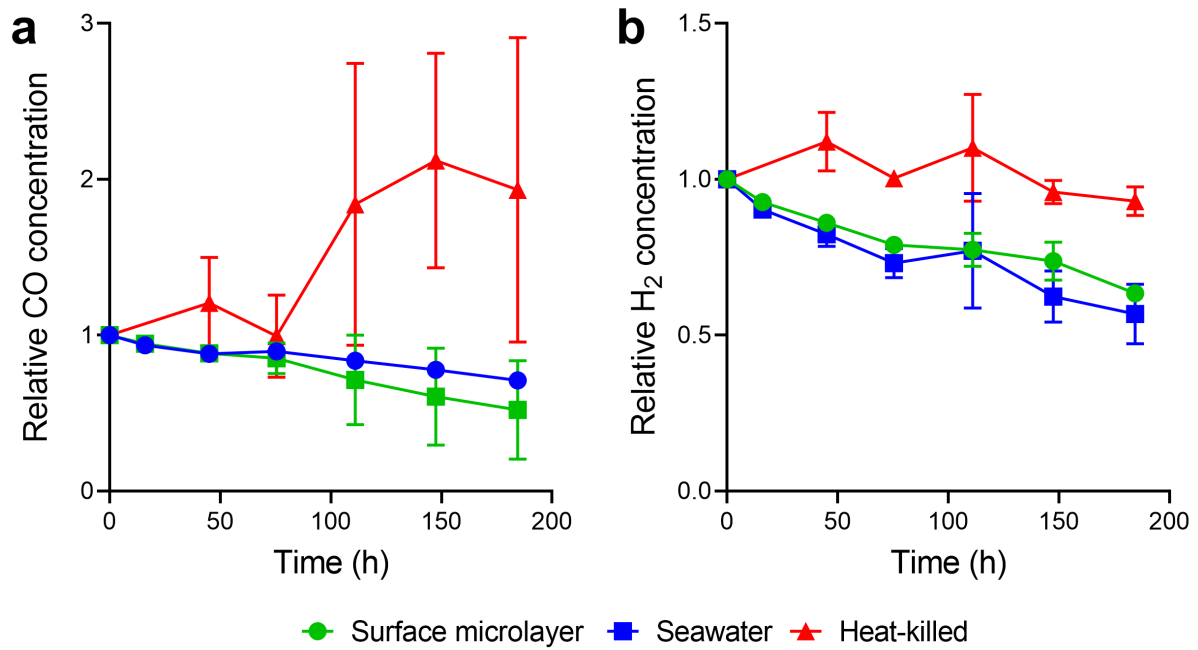


Figure S4. Community composition of marine microbial communities based on metagenomes. The top 10 bacterial and archaeal orders with highest mean relative abundance across all samples are shown. Relative abundances are based on the assembly and classification of the 16S rRNA gene by PhyloFlash. PPB = Port Phillip Bay, SML = Surface microlayer, STW = Subtropical waters, SAW = Subantarctic waters.

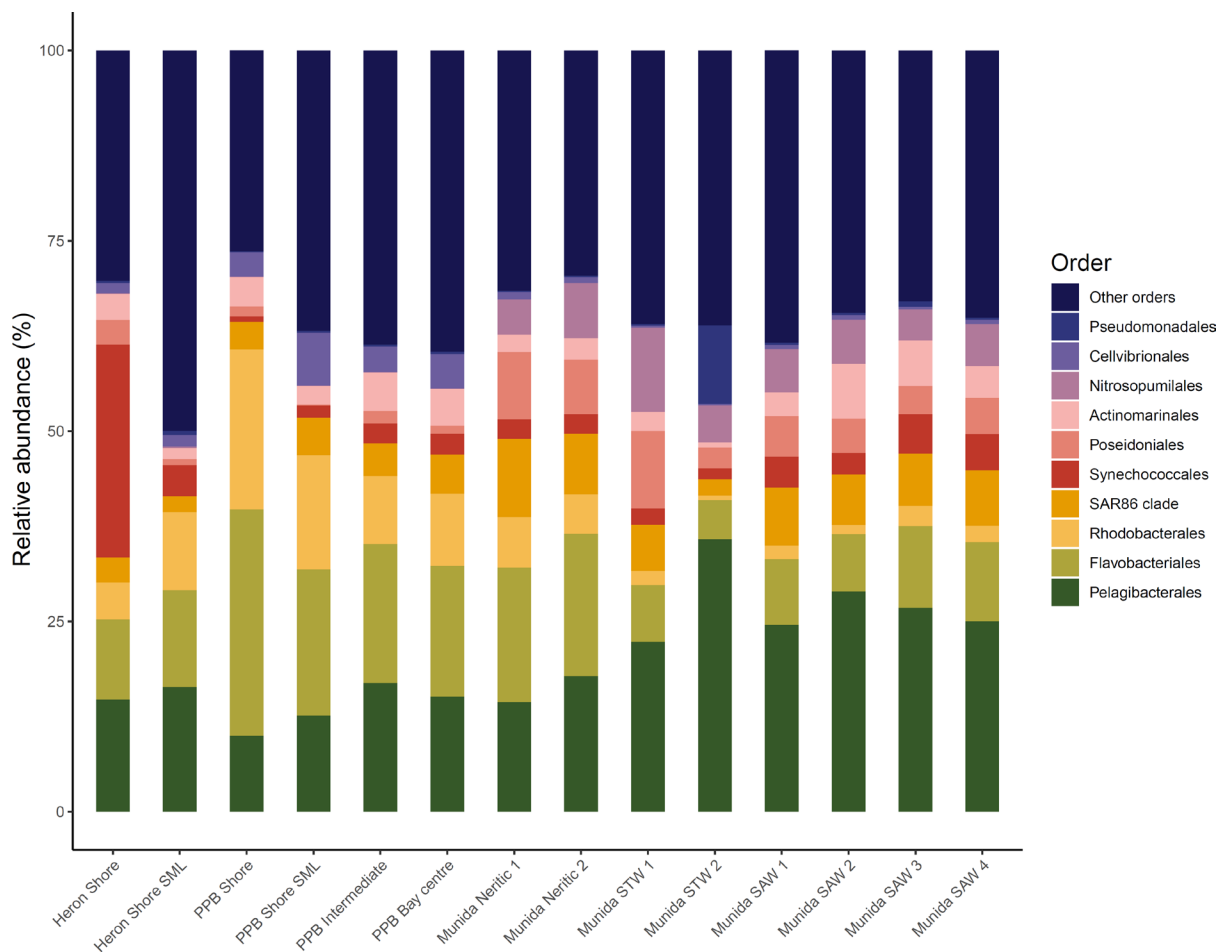


Figure S5. Simple linear correlations between trace gas oxidiser abundance and trace gas oxidation rates. H₂ and CO oxidizer abundance are based on the proportion of bacteria encoding H₂-uptake hydrogenases and form I CO dehydrogenases in the metagenomes. H₂ and CO oxidation rates are calculated based on bulk oxidation rates at measured *in situ* concentrations as per **Table S1**. Pearson's R² values show goodness of fit and *p* values confirm each slope significantly deviates from zero.

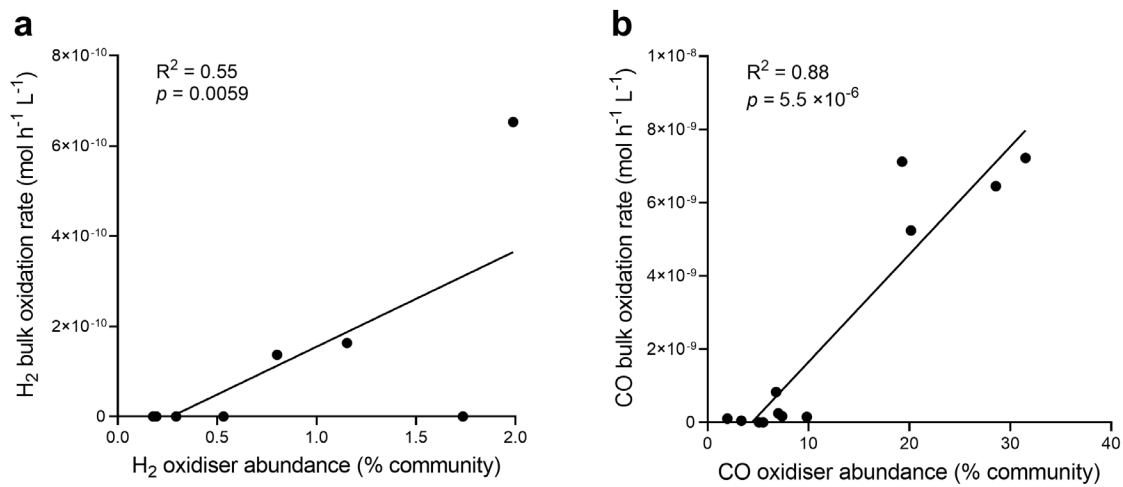


Figure S6. Distribution of metabolic genes in marine bacteria from the three study sites. Bubble plot showing metabolic potential of the 110 metagenome-assembled genomes (MAGs) constructed from the three study sites. MAGs are summarised at phylum level, with the size of the circle corresponding to the number of genomes in that order with a given gene, and the colour reflecting the percentage of genome completeness. Marker genes are omitted that were not detected in any MAG.

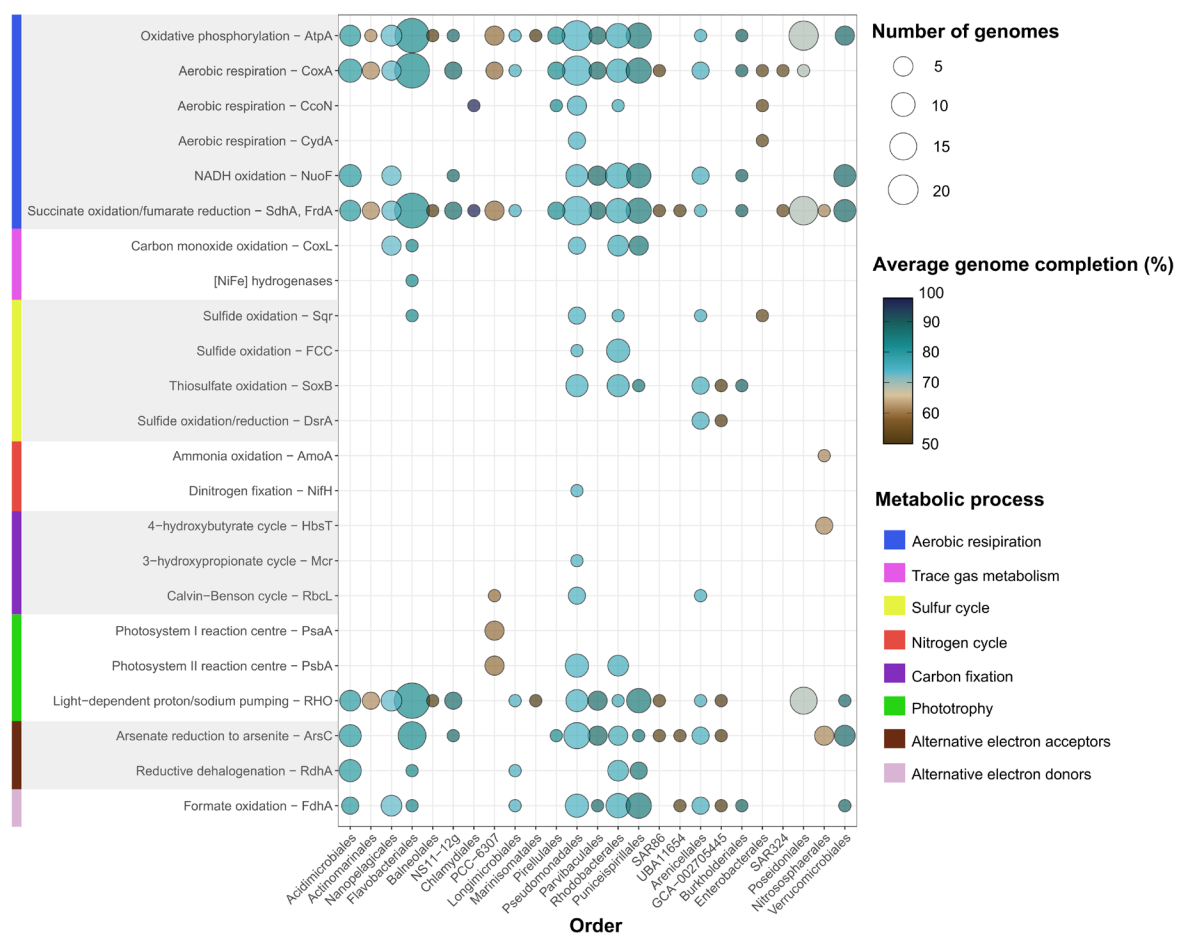


Figure S7. Linear maximum-likelihood phylogenetic tree of marine group 1 and 2 [NiFe]-hydrogenases. Hydrogenase catalytic subunit sequences retrieved from the new MAGs (coloured green) and *Tara* MAGs (coloured green) are shown alongside representative reference sequences (coloured black). Evolutionary history was inferred by using the JTT matrix-based model, the tree was bootstrapped using 50 replicates, and the tree is rooted using the outgroup group 4a [NiFe]-hydrogenase sequence. The tree includes hydrogenase subgroups implicated in aerobic respiration (group 1d, 1f, 1l, 2a), anaerobic respiration (group 1a, 1b, 1c, 1e), and H₂ sensing (group 2b and 2c).

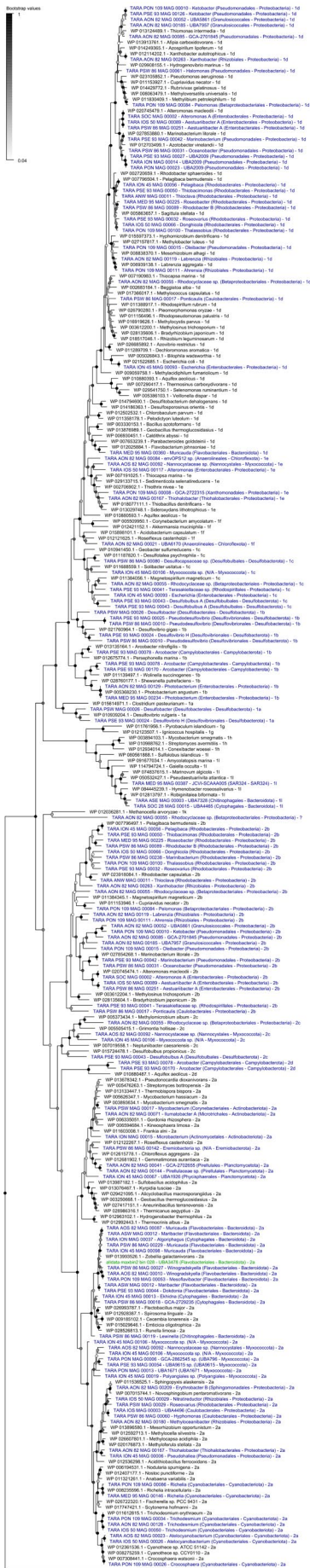
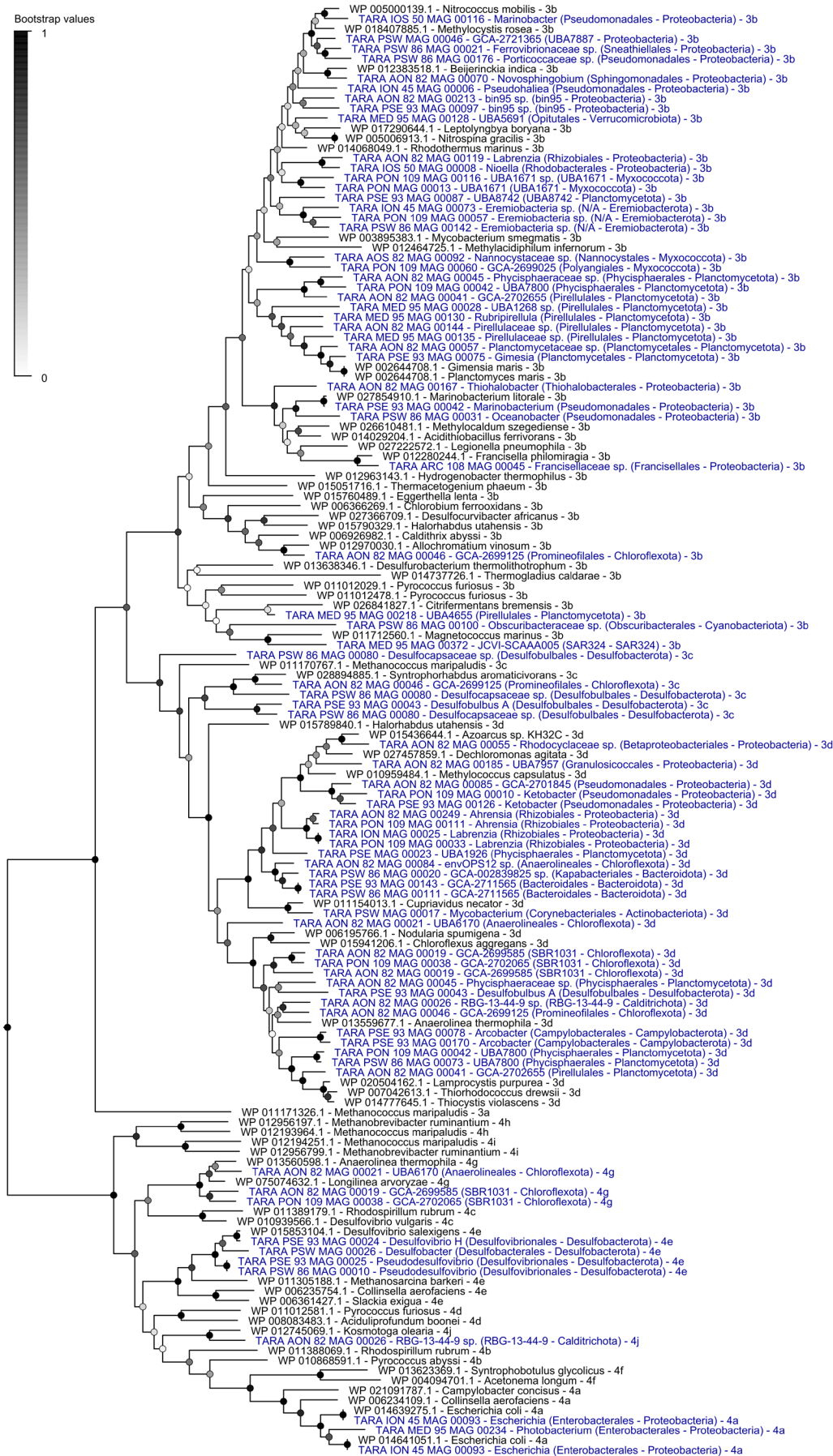


Figure S8. Linear maximum-likelihood phylogenetic tree of marine group 3 and 4 [NiFe]-hydrogenases. Hydrogenase catalytic subunit sequences retrieved from the new MAGs (coloured green) and *Tara* MAGs (coloured green) are shown alongside representative reference sequences (coloured black). Evolutionary history was inferred by using the JTT matrix-based model, the tree was bootstrapped using 50 replicates, and the tree is midpoint-rooted. The tree includes hydrogenase subgroups implicated in fermentation (group 3b, 4a), carbon fixation (group 3d), and energy conversion (group 4e).



0.3

Figure S9. Linear maximum-likelihood phylogenetic tree of marine carbon monoxide dehydrogenases. Catalytic subunit sequences retrieved from the new MAGs (coloured green) and *Tara* MAGs (coloured green) are shown alongside representative reference sequences (coloured black). Evolutionary history was inferred by using the JTT matrix-based model, the tree was bootstrapped using 50 replicates, and the tree is rooted using outgroups including the form II CO dehydrogenase.

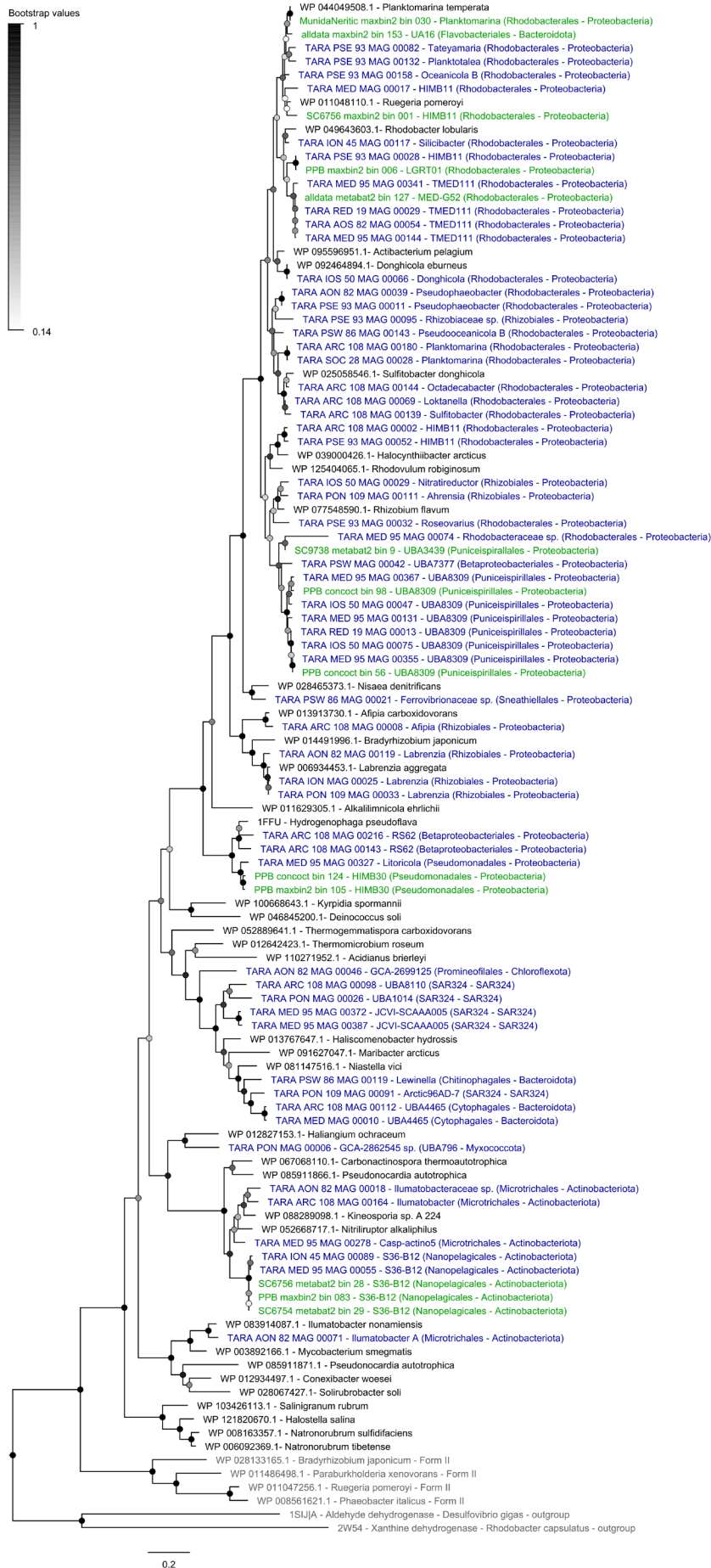


Figure S10. Correlation matrix of environmental variables based on *Tara* Oceans metadata. Pearson correlations were performed on the original *Tara* Oceans metadata without imputation of missing values: each pairwise correlation was calculated using all complete pairs of observations for that variable.

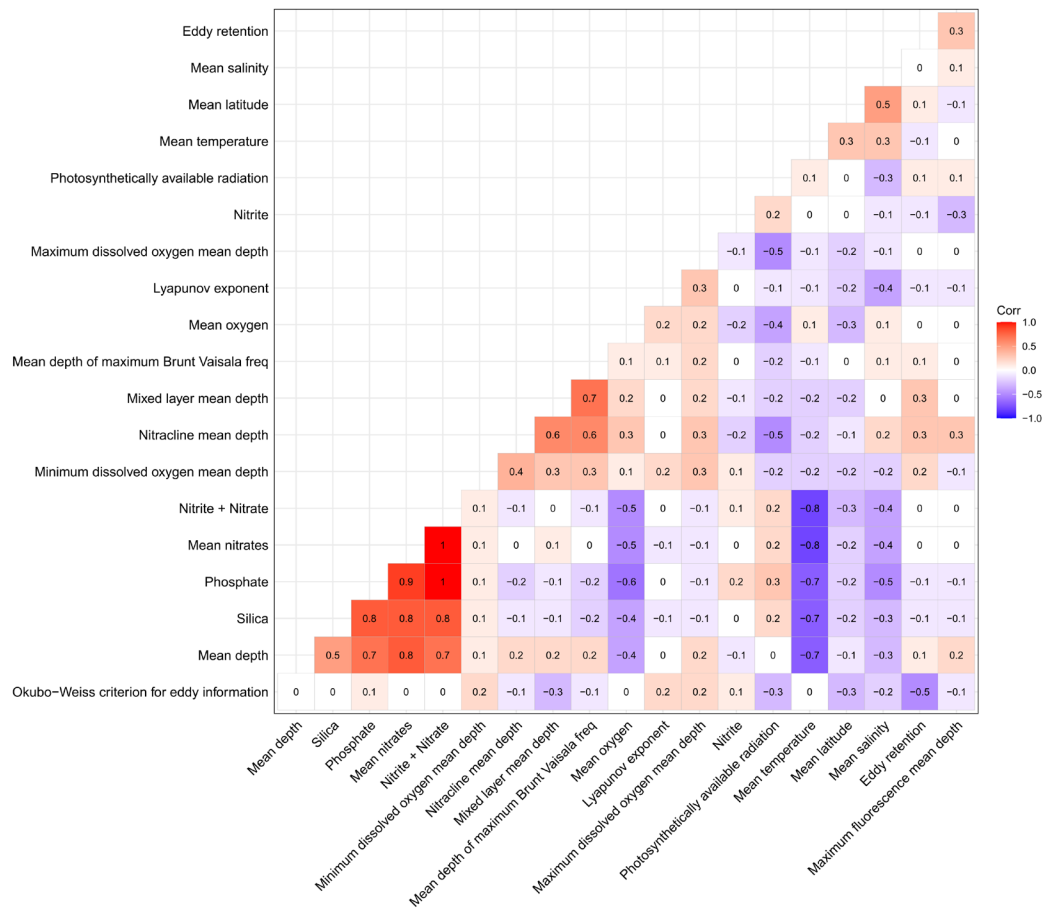


Figure S11. Random forest analysis of strongest environmental predictors of metabolic gene abundance. Results are shown based on the mean gene copies per organism of energy-converting rhodopsins, CO dehydrogenases, and various [NiFe]-hydrogenases based on the *Tara* Oceans metagenomes and accompanying metadata.

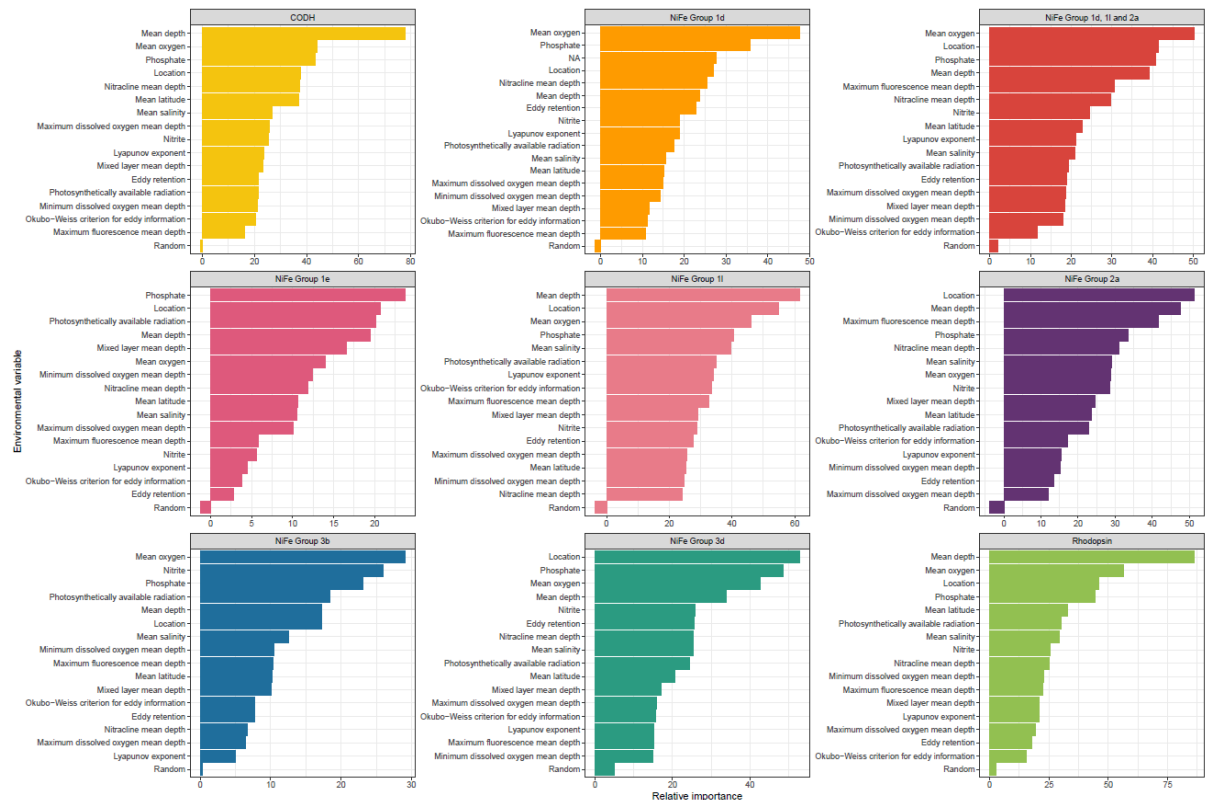


Figure S12. Random forest analysis of strongest environmental predictors of metabolic gene expression. Results are shown based on the reads per kilobase million (RPKM; log-transformed) of energy-converting rhodopsins, CO dehydrogenases, and various [NiFe]-hydrogenases based on the *Tara* Oceans metatranscriptomes and accompanying metadata.

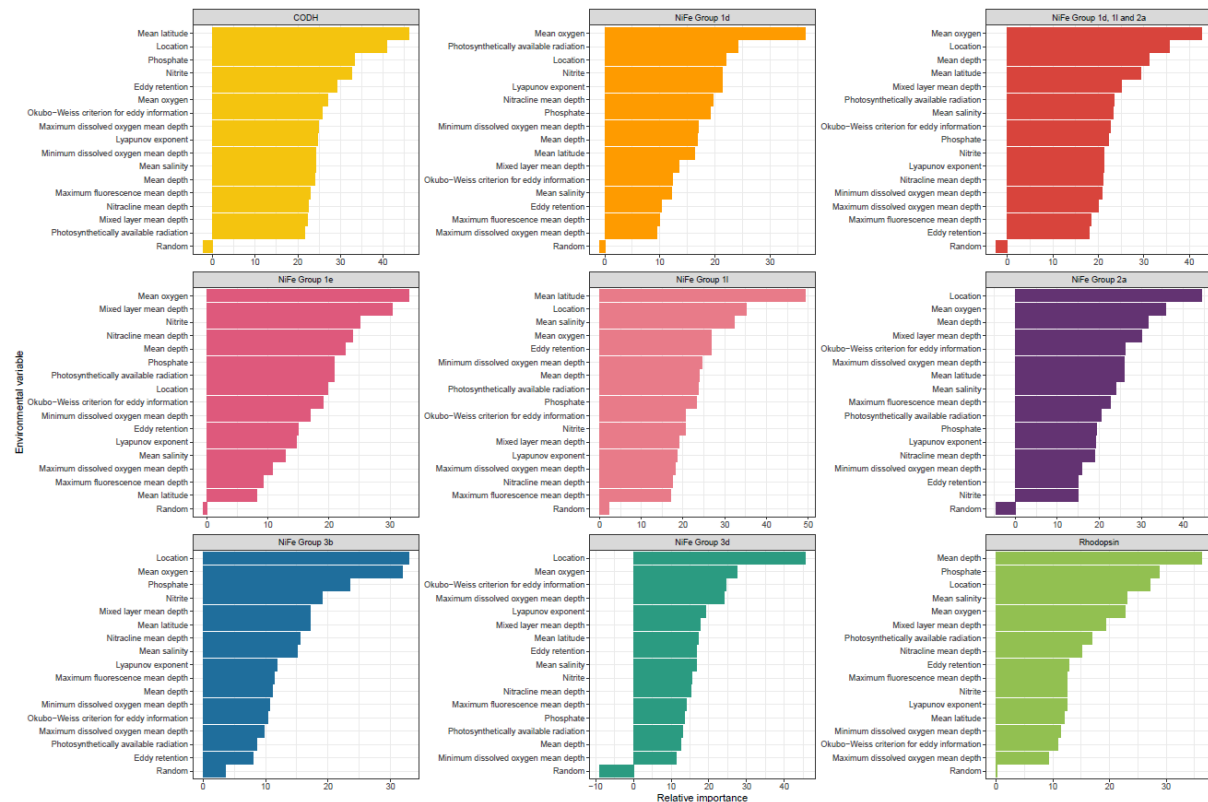


Figure S13. Relationship between latitude and trace gas oxidation capacity in the *Tara* Oceans dataset. Samples are coloured by surface (blue), deep chlorophyll maximum (red) or mesopelagic (green) depth. Gene abundance is represented on the top row, and gene expression on the bottom row.

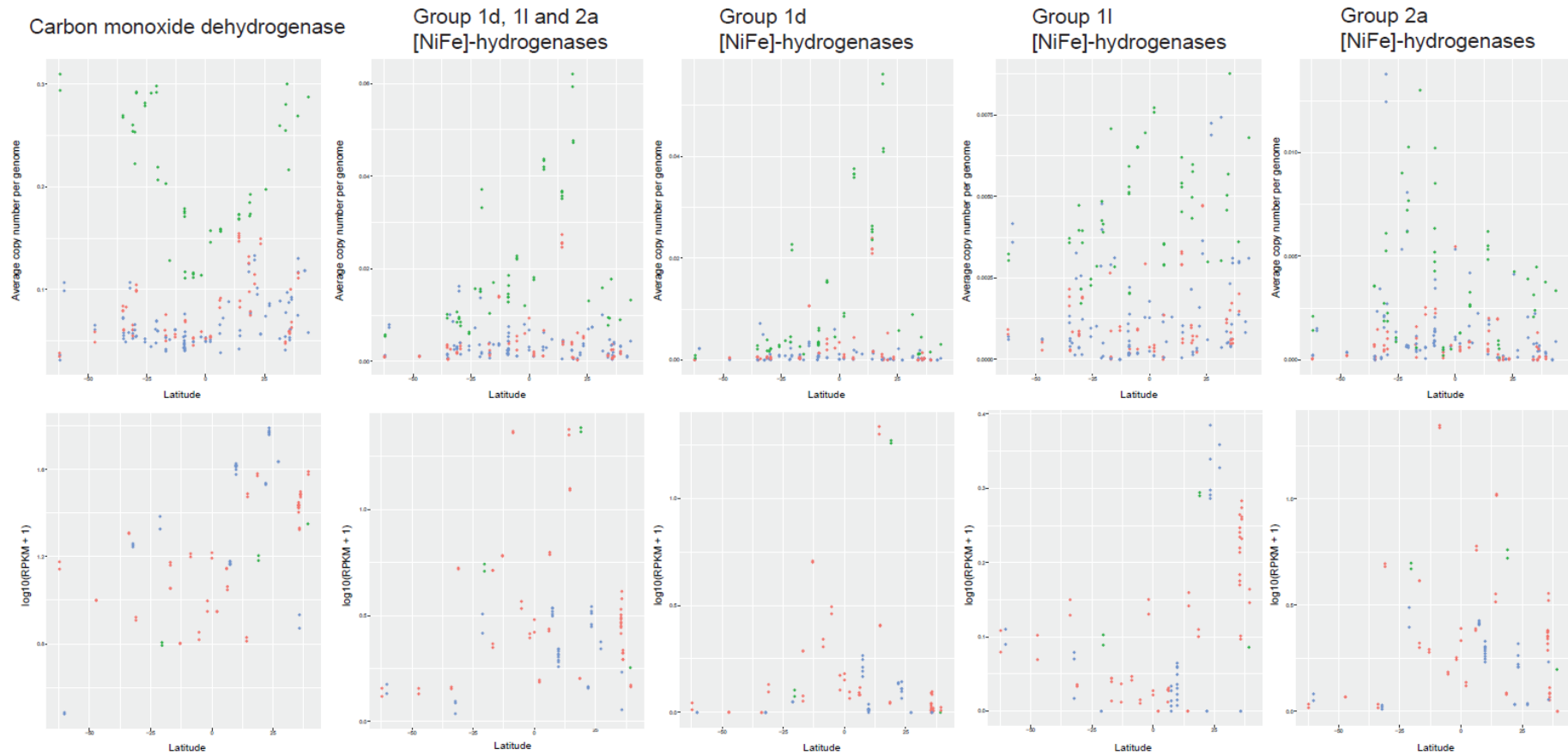


Figure S14. Relationship between oxygen concentration and trace gas oxidation capacity in the *Tara* Oceans dataset.

Samples are coloured by surface (blue), deep chlorophyll maximum (red) or mesopelagic (green) depth. Gene abundance is represented in the top row, and gene expression on the bottom row.

