**Supplementary Information for**

**“Microbial iron and carbon metabolism as revealed by taxonomy-specific functional diversity in the Southern Ocean”**

Ying Sun1\*†, Pavla Debeljak1,2\*, Ingrid Obernosterer1

1 CNRS, Sorbonne Université, Laboratoire d’Océanographie Microbienne, LOMIC, F-66650 Banyuls/mer, France. 2 University of Vienna, Department of Functional and Evolutionary Ecology, A-1090 Vienna, Austria.

\* These authors contributed equally to this study.

† Correspondence to ying.sun@obs-banyuls.fr.

**This file includes:**

Supplementary Methods

Supplementary Results

Supplementary Figure legends

Supplementary Data

Supplementary Code availability

Supplementary References

**Other supplementary material for this manuscript includes the following:**

Supplementary Fig. 1 to 14

Supplementary Table 1 to 6.

**Supplementary Methods**

***Environmental conditions and sample collection***

The Marine Ecosystem Biodiversity and Dynamics of Carbon around Kerguelen (MOBYDICK) cruise took place in the Indian Sector of the Southern Ocean during the austral late summer period from Feb 18th to Mar 29th in 2018. Three stations in contrasting oceanic regions were chosen for our study, including one located in the naturally iron-fertilized waters above the central Kerguelen Plateau (M2) and two off-plateau stations within the High Nutrient Low Chlorophyll (HNLC) waters (M3 and M4; **Supplementary Fig. 1A** and **Supplementary Table 1**). The timing of the sample collection corresponded to about two months after the peak of the summer phytoplankton blooms (**Supplementary Fig. 1B**). In central Kerguelen Plateau waters maximum seasonal Chlorophyll *a* (Chl *a*) concentrations, based on satellite images, were about 1.5 µg L-1, and they were substantially lower (0.3 to 0.5 µg L-1) in HNLC waters. During the MOBYDICK cruise in late austral summer, differences in the Chl *a* concentrations among sites were less pronounced. As a consequence of the sustained seasonal phytoplankton activity, concentrations of dissolved organic carbon (DOC) were higher in Kerguelen Plateau waters as compared to those in HNLC waters and during previous investigations in the study region in spring [[1](#_ENREF_1)] (**Supplementary Fig. 1BC**). Concentrations of dissolved iron were shown not to differ in the on- and off-plateau surface waters in summer (0.09 nM) [[2](#_ENREF_2)], due to the rapid utilization of iron supplied by natural fertilization [[3](#_ENREF_3)]. The seasonal phytoplankton activity led to a build-up of heterotrophic prokaryotic abundance, associated to higher heterotrophic prokaryotic production, in fertilized waters as compared to the HNLC sites. Station M2 was visited three times at an 8-day interval, and station M3 and M4 were visited twice at a two-week interval. Samples were collected for all visits from surface waters (10 m) with 12 L Niskin bottles mounted on a rosette equipped with a conductivity, temperature, depth sensor (Seabird SBE-911 plus CTD unit).

***Microbial community DNA extraction and metagenome library preparation***

Total genomic DNA was extracted from the Sterivex filter units using the AllPrep DNA/RNA kit (Qiagen, Hiden, Germany) with the following modifications: Filter units were thawed and closed with a sterile pipette tip at the outflow. Lysis buffer was added (40 mM EDTA, 50 mM Tris, 0.75 M sucrose) and three freeze-and-thaw cycles were performed using dry ice in ethanol and a water bath at 65 °C. Lysozyme solution (0.2 mg mL-1 final concentration) was added, and filter units were placed on a rotary mixer at 37 °C for 45 min. Proteinase K (0.2 mg mL-1 final concentration) and SDS (1% final concentration) were added and filter units were incubated at 55 °C with gentle agitation every 10 min for 1 h. To each filter unit, 1 550 µl RLT plus buffer was added and inverted to mix. The lysate was recovered by using a sterile 5 ml syringe and loaded in three additions onto the DNA columns by centrifuging at 10 000x g for 30 sec. DNA purification was performed following manufacturer’s guidelines. The concentration of double-stranded DNA was quantified by PicoGreen fluorescence assay (Life Technologies). DNA quality was checked on an Agilent 2100 Bioanalyzer/Agilent Nano DNA chip (Agilent, Santa Clara, CA, USA). Triplicate DNA extracts were pooled in equimolar amounts providing 1 pooled DNA extract per visit and station. The DNA extracts from the repeated visits (3 at M2 and 2 at each M3 and M4) were then pooled for each station to achieve 1 µg in 30 µL Tris for sequencing purposes.

***Microbial community RNA extraction and metatranscriptome library preparation***

Triplicate samples (each 10 L) for prokaryotic RNA extraction were collected from the Niskin bottles and immediately filtered. The filtration procedure did not exceed 10 min and 10 ml of RNA-later was added to the filter in a Greiner tube prior to storage at -80 °C. RNA was extracted using the NucleoSpin® RNA Midi kit (Macherey-Nagel, Düren, Germany). Filters stored in RNA-later were thawed, cut in half, removed from the RNA-later solution, re-frozen in liquid nitrogen and shattered using a mortar. The second half of the filter was kept and extracted when sufficient RNA could not be obtained in the first round of extraction, resulting in different values for normalization per volume (L) used (**Supplementary Table 2**). The obtained ‘powder-like’ filter-pieces were added to the denaturing lysis buffer supplied by the NucleoSpin® RNA Midi kit (Macherey-Nagel, Düren, Germany) and vortexed for 2 min. Two internal standard RNA molecules of 1 006 nucleotides (nt) were synthesized and added to each sample with known copy numbers (**Supplementary Table 2** and **Supplementary Data**), to enable absolute quantification of transcripts in the downstream analysis [[4](#_ENREF_4), [5](#_ENREF_5)]. RNA was treated with two rounds of DNA digestion and purified using the RNA Clean & Concentrator™-5 kit (Zymo Research, OZyme, France). Ribosomal RNA (rRNA) was removed with the RiboZero rRNA stranded RNA protocol.

***Metagenomic read quality control, assembly, and mapping***

Metagenomic reads was evaluated using FastQC (v0.11.7) [[6](#_ENREF_6)] and processed with Trimmomatic (v0.39) [[7](#_ENREF_7)]. An amount of 327.8, 336.3, and 285.0 million high-quality reads were retained for station M2, M3 and M4, respectively (**Supplementary Table 2**). Sequencing coverage and complexity of microbiomes in each metagenomic dataset was assessed using Nonpareil (v3.303) [[8](#_ENREF_8)] with the k-mer based approach (**Supplementary Fig. 2A**). Combined with another 337.3 M high-quality reads from station M2 sampled in a previous cruise in early spring, de novo co-assembly was performed using MEGAHIT (v1.0.4) [[9](#_ENREF_9)] with default settings. CD-HIT-EST (v4.7; -c 0.99 -aS 1 -g 1) [[10](#_ENREF_10)] was used to measure the redundancy of the assembled contigs. Metagenomic reads were aligned back to the contigs using Bowtie2 (v2.3.5) [[11](#_ENREF_11)], with duplicates removed by the markdup function of SAMtools (v1.9) [[12](#_ENREF_12)]. On average, 30.32% of reads in each metagenomic dataset could be recruited back to the assembled contigs.

***Southern Ocean assemblies vs. NCBI nt and nr databases***

We queried all the contigs against the NCBI nt database using BLASTN (v2.7.1) [[13](#_ENREF_13)] with an e-value threshold of 1e-10. The bitscore and percentage of identity of the best alignment for each contig were collected. For those contigs which could not find a match in the nt database, zeroes were assigned. The statistics of the two features are shown in (**Supplementary Fig. 2BC**).

We also searched for homologs of our predicted genes, at both the nucleotide and amino acid level, against the NCBI nt database using BLASTN (v2.7.1) [[13](#_ENREF_13)] and nr database using Diamond (v0.9.24; BLASTP mode) with an e-value threshold of 1e-10. The bitscore and percentage of identity of the best alignment for each gene were collected. For those genes which could not find a match in the nt (or nr) database, zeroes were assigned. The statistics of the two features are shown in (**Supplementary Fig. 2D-G**).

***Metagenome Southern Ocean vs. TARA***

A total of 3 588‬ draft genomes from the TARA Ocean Global Expedition Project, including 957 from Delmont, et al. [[14](#_ENREF_14)] and 2 631 from Tully, et al. [[15](#_ENREF_15)], were downloaded. Besides, another 41 newly released Arctic metagenomic datasets with 11 709 809 contigs were also collected [[16](#_ENREF_16)]. We leveraged FastANI (v1.3) [[17](#_ENREF_17)] for estimating pairwise ANI values between our 133 MAGs and the TARA draft genomes (**Supplementary Fig. 4A**). Read recruitment from our Southern Ocean samples to the TARA assemblies, including both the draft genomes and the un-binned Arctic contigs, were implemented using BBMap (v38.22) [[18](#_ENREF_18)] to verify the differences observed through ANI. The percentage of reads mapped to each assembly dataset, with a minimum percentage of identity at 90%, were summarized, and further normalized by the size of each dataset to a “per Mbp” unit (**Supplementary Fig. 4BCE**). Sequence similarities between our 949 228 non-redundant contigs and the TARA Arctic contigs were assessed using BLASTN (v2.7.1) [[13](#_ENREF_13)]. Matches were identified with a minimum percentage of identity at 90%, a minimum coverage of the shorter sequence at 90% and an e-value threshold at 1e-3 (**Supplementary Fig. 4D**). Interestingly, in contrary to the low sequence similarity shared among contigs, the TARA Arctic datasets showed higher abundance through read recruitment than our MOBYDICK assemblies (adjusted P-value < 0.05 by Wilcoxon rank sum test; **Supplementary Fig. 4E**). To further assess the similarity between our metagenomes and the TARA Arctic ones, we downloaded the 41 TARA Arctic metagenomes and implemented BBMap (v38.22) [[18](#_ENREF_18)] to map TARA Arctic reads to both the TARA Arctic contigs and our MOBYDICK contigs. The read recruitment statistics are summarized and further normalized by the size of each dataset to a “per Mbp” unit, as shown in **Supplementary Fig. 4F**. TARA Arctic metagenomes and contigs are classified into two groups based on sampling depth. In addition to the comparisons at the DNA level, we employed OrthoFinder (v2.2.3) [[19](#_ENREF_19)] to identify orthologous gene clusters among our Southern Ocean MAGs and TARA draft genomes (**Supplementary Fig. 4GH**). Results and discussion could be found in **Supplementary Results**.

***Reconstruction of prokaryotic16S rRNA markers from metagenomes***

Reconstruction of prokaryotic16S rRNA markers from metagenomes is a complementary tool for evaluating microbial diversity and abundance. We implemented 16S rRNA-based community characterization using the MATAM (v1.6.0) [[20](#_ENREF_20)] and phyloFlash (v3.4) [[21](#_ENREF_21)] assemblers with the SILVA SSU rRNA database (v138.1) [[22](#_ENREF_22)], resulting 6 groups of SSU assemblies (3 samples × 2 assemblers). The V4-V5 region of each 16S rRNA assembly was extracted by SSU-ALIGN (v0.1.1) [[23](#_ENREF_23)]. Non-redundant representative V4-V5 sequences were selected across samples using CD-HIT-EST (v4.7; -c 1 -aS 1 -g 1) [[10](#_ENREF_10)], among which 1 460 were at least 200 bp in length and verified with a prokaryotic origin by the assignTaxonomy function of DADA2 (v1.16.0) with its precompiled SILVA training set (v138.1) [[24](#_ENREF_24)].

The identified archaeal and bacterial 16S rRNA V4-V5 representative sequences were aligned by SSU-ALIGN (v0.1.1) [[23](#_ENREF_23)] according to their secondary-structure covariance models. We built a phylogenetic tree from the sequence alignment using FastTree (v2.1.10; -slow -nt -gtr -gamma -bionj) [[25](#_ENREF_25)], with the taxonomic assignment of the sequences as constraints at the order level (-constraints). Metagenomic reads were mapped back to the 16S rRNA V4-V5 representative sequences using Bowtie2 (v2.3.5) [[11](#_ENREF_11)]. A pseudo-OTU table was generated by jgi\_summarize\_bam\_contig\_depths [[26](#_ENREF_26)], taking the depth of coverage as an approximation to amplicon counts. Visualization and statistics were performed in R (v3.6.1) using the phyloseq package (v1.28.0) [[27](#_ENREF_27)]: Closely-related taxa were clustered through phylogenetic agglomeration by tip\_glom (**Supplementary Fig. 5A**); The pseudo-OTU table was transformed for equal sampling depth by rarefy\_even\_depth; Relative abundance was calculated with the aid of transform\_sample\_counts (**Supplementary Fig. 5B**); Alpha diversity was assessed according to the Shannon index and visualized via plot\_richness (**Supplementary Fig. 5C**); To describe the differences across sites, we applied the double principal coordinate analysis (DPCoA) with regard to both phylogenetic structure and relative abundance, and visualized the results using plot\_ordination (**Supplementary Fig. 5D**). Environmental vectors (M2, M3 and M4) were added according to the coordinates of the sampling sites reported in the DPCoA result. The first principal component, which accounts for more than 99% of the total variance, is associated positively with M2 and negatively with M3 and M4. Given that the three sampling sites are representatives of the on- (M2) and off-plateau (M3 and M4) environments, the first principal component could primarily measure the major differences that distinguish the contrasting oceanic regions, such as the availability of iron and organic carbon.

Clusters of taxonomic groups are distributed along the first axis. The SAR11 group and Bacteroidota are well separated from each other and other taxonomic groups. A Monte-Carlo permutation test (1 000 replicates) applied to the DPCoA result was used to evaluate whether the observed compositional differences among microbial communities are higher than expected in a random distribution, by implementing the randtest.dpcoa function in the ade4 package (nrep=1000 and alter=“greater”) [[28](#_ENREF_28)]. The simulated p-value approximates 0.001, assuring phylogenetic clustering as observed in **Supplementary Fig. 5D** given that phylogeny was used to build the distance matrix for our DPCoA analysis.

We also estimated the root square of Rao’s dissimilarity coefficient between samples by the disc function in the ade4 package, using the OTU table and the phylogenetic distance (generated by the cophenetic.phylo function in the ape package) involved in the DPCoA analysis. This resulted a Rao’s dissimilarity value of 0.1624 between M2 and M3, 0.1807 between M2 and M4, and 0.0259 between M3 and M4, in consistent with the correlations between the environmental vectors as shown in **Supplementary Fig. 5D**. To evaluate the statistical significance of categorical explanatory variables, such as the on- and off-plateau environments, we performed another permutation test based on constrained DPCoA (cDPCoA) [[29](#_ENREF_29)]. Our cDPCoA analysis extended the DPCoA result with a partitioning factor of sampling sites (on- and off-plateu), using the bca (between-class) function in the ade4 package. The Monte-Carlo permutation test (1 000 replicates) on the cDPCoA result, using the aforementioned randtest.dpcoa function, resulted in a p-value of 0.3446553 (> 0.05). Considering that we only have three metagenomes, this insignificant outcome is not surprising when the analysis focused on the effect of environmental types. We undertook simulations with pseudo replicates for each sampling sites to examine how the number of samples biased the significance test. We considered two cases, including 2 and 3 replicates per site. Assuming that biological replicates from the same sampling site should consist of similar microbial communities, the pseudo replicate(s) was generated by replacing the original measurement (*x*) with a random number selected using the rnorm function with mean=*x* and sd=0.25*x*. The same cDPCoA and Monte-Carlo permutation test was performed on each simulated OTU table. A total of 100 simulations were carried out in the case of 2 replicates per site, and the mean value of the resulted p-values is 0.06286713 (±0.01416026 standard deviation; mean > 0.05). When 3 replicates per site were considered, the p-value approximates 0.01042957±0.004679961 (mean < 0.05).

***Assembly-free metagenome taxonomic profiling***

One limitation of working with metagenomes is the considerable amount of reads that could not be assembled into contigs (**Supplementary Table 2**). Such a limitation might be explained by the prevalence of genomes with low genome abundance and the deficiency in de Bruijn assembly algorithms [[30](#_ENREF_30)]. Therefore, we also performed taxonomic profiling solely based on unassembled metagenomic reads. We implemented three types of taxonomic classifiers with distinct pros and cons [[31](#_ENREF_31)], including DNA-to-DNA (Centrifuge v1.0.4 [[32](#_ENREF_32)] with its precompiled nt database and Kraken2 v2.0.7-beta [[33](#_ENREF_33)] with its standard database), DNA-to-protein (Kaiju v1.7.0 [[34](#_ENREF_34)] with its precompiled nr databases), and phylogenetic-marker-based (mOTUs2 v2.5.1 [[35](#_ENREF_35)]) methods. The mOTUs2 intrinsically incorporates the correction for copy number and genome length by using universal single-copy phylogenetic marker genes [[35](#_ENREF_35)] (**Supplementary Fig. 5E**). However, to derive accurate relative abundance, the other three tools require subsequent normalization by the elusive total sequence size of individual taxon in the reference databases. To avoid this, we divided out the effects of reference sizes, as well as other multiplicative systematic biases, by considering the fold changes between taxon ratios across sampling sites [[36](#_ENREF_36)]. The fold change is defined as $\frac{O\_{i}(s)}{O\_{j}(s)}/\frac{O\_{i}(t)}{O\_{j}(t)}$, where *O* represents the observed number of reads from sample *s* (or *t*) that were assigned to individual taxonomic group *i* (or *j*) [[36](#_ENREF_36)]. We used *Candidatus Pelagibacter* as the denominator (taxon *j*) given its prevalence and high abundance across sampling sites. The taxon ratios measured the “relative abundance” of a taxonomic group to *Candidatus Pelagibacter*, instead of to the whole community. The fold change of taxon ratios, by crossing out potential multiplicative systematic biases [[36](#_ENREF_36)], was used as an indicator of the relative fitness of a taxon as compared to the ubiquitous SAR11 population in contrasting environments (M3 vs. M2, or M4 vs. M2; **Supplementary Fig. 5F**). Regarding possible detection biases against rare taxa, here, we summarized read counts at the genus levels and only considered genera whose read counts was no less than 1‰ of *Candidatus Pelagibacter*’s.

***Metatranscriptomic read quality control and internal standards quantification***

An initial round of read processing was provided by the company using Trimmomatic (v0.39) [[7](#_ENREF_7)]. The standard Illumina adapters and low-quality bases were removed with the following parameters “2:30:10 SLIDINGWINDOW:4:5”. Further, we performed another round of quality control and refinement with Trim Galore (v0.5.0) using a minimum Phred score of 20 and a length threshold of 50 bp. To remove rRNA, tRNA and internal standard sequences, a two-step clean-up procedure was performed, starting with SortMeRNA (v2.1b) [[37](#_ENREF_37)] and followed by a BLASTN (v2.7.1) [[13](#_ENREF_13)] search against a custom database consisting of 545 336 reference rRNA and tRNA sequences from diverse taxa along with the 2 internal standard sequences [[38](#_ENREF_38)]. Finally, an average of 22.38 M high-quality paired-end protein-coding reads were retained from each replicate, and the per-liter calculations were performed based on the recovery rate of internal standards as previously described [[38-40](#_ENREF_38)] (**Supplementary Table 2**).

***Identification of PUL-like structures and fucose utilization loci***

In this study, we adopted a more general term CAZyme gene clusters (CGCs) [[41](#_ENREF_41)] to identify PUL-like gene clusters in a broader sense, requiring at least one CAZyme accompanied with transcription factors (TFs) and/or transporters (TCs) (**Supplementary** **Fig. 13A-D**). To identify fucose utilization loci among Verrucomicrobiae MAGs, the nine *Lentimonas* genomes provided in Sichert et al. [[42](#_ENREF_42)] were downloaded from NCBI GenBank database. Orthologous proteins between the MAGs and the *Lentimonas* references were determined using OrthoFinder (v2.2.3) [[19](#_ENREF_19)] (**Supplementary** **Fig. 13E**).

***Comparative analysis of Pseudomonadales HTCC2089 pangenomes***

The sequences of 19 Pseudomonadales HTCC2089 draft genomes were downloaded from NCBI GenBank database, including 15 from UBA4421 genus and 4 from UBA9926 (**Supplementary** **Table 6**). Pangenome is defined on the basis of cluster of orthologous genes identified by OrthoFinder (v2.2.3) [[19](#_ENREF_19)], and comparative analysis between genus UBA4421 and UBA9926 was carried out from three aspects, including the core, accessory and singleton genes (**Supplementary Table 6**). A gene cluster enhancing the use of light as an energy resource was identified exclusively in the UBA4421 genus (**Supplementary Fig. 14A**). A gene encoding bacteriorhodopsin was also identified in one of the UBA9926 genome (GCA\_002728935), but genes around it are not related to light harvesting (**Supplementary Fig. 14A**). Genes encoding bacteriorhodopsin were identified and their protein sequences were retrieved. The top 5 homologous bacteriorhodopsin sequences were selected from the MicRhoDE database [[43](#_ENREF_43)] via BLASTP search. Another two archaeal rhodopsin sequences were obtained from MicRhoDE and used as outgroups for phylogeny analysis. All bacteriorhodopsin proteins were aligned by MAFFT (v7.313) [[44](#_ENREF_44)] and processed by trimAl (v1.4; -automated1) [[45](#_ENREF_45)]. Phylogenetic reconstruction was performed using IQ-Tree (v1.6.8; -m MFP -bb 1000 -bnni) [[46](#_ENREF_46)]. The bacteriorhodopsins from distinct genera are well separated on the phylogeny tree with high support values (**Supplementary Fig. 14B**).

**Supplementary Results**

***Comparison between MOBYDICK metagenome assemblies and the TARA ocean dataset***

The TARA Ocean Global Expedition Project provides an extensive survey on the global surface ocean microbiomes [[47](#_ENREF_47)], providing 3 588‬ draft genomes [[14](#_ENREF_14), [15](#_ENREF_15)] and another 41 newly released Arctic metagenomic datasets with 11 709 809 contigs [[16](#_ENREF_16)]. However, to date, only 3 TARA prokaryote-enriched metagenomes from the surface Southern Ocean are publicly available, and their sampling sites were far away from ours. The aforementioned comparisons between our Southern Ocean assemblies and the NCBI databases highlight the uniqueness of the Southern Ocean microbiomes (**Supplementary Fig. 2B-G**) and further request the comparisons between our Southern Ocean MAGs and the TARA assemblies. Our analyses were performed from three main aspects, including the pairwise average nucleotide identity (ANI) assessment, the Southern Ocean metagenomic read recruitment and the identification of orthologous protein families between the two datasets (see **Supplementary** **Methods**). Only 13 of the Southern Ocean MAGs conform to > 95% intra-species ANI values with TARA counterparts, despite the high similarity shared among the TARA MAGs (**Supplementary Fig. 4A** and **Supplementary Table 3**) [[17](#_ENREF_17)]. Metagenomic read recruitment analysis further confirmed that, except for the TARA Southern Ocean samples (SOC), the TARA MAGs recruited a significantly lower amount of our metagenomic reads as compared to our Southern Ocean MAGs (adjusted P-value < 0.05 by Wilcoxon rank sum test; **Supplementary Fig. 4BC**), assuring that the novelty of our Southern Ocean assemblies is not derived from biases introduced during metagenome assembly and binning. In addition, only 4.18% (±1.33%; S.D.) of our Southern Ocean contigs could find a match with high similarity in the TARA Arctic datasets (**Supplementary Fig. 4D**). Interestingly, in contrary to the low sequence similarity shared among contigs, the TARA Arctic datasets showed higher abundance through read recruitment than our Southern Ocean assemblies (adjusted P-value < 0.05 by Wilcoxon rank sum test; **Supplementary Fig. 4E**) and the vice-versa read recruitment analysis confirmed the same pattern (**Supplementary Fig. 4F**), suggesting that Arctic and Southern Ocean metagenomes may have some taxonomic and functional similarity.

***Genera with anomalously deviated fold changes between taxon ratios across sampling sites***

Particularly, *Polaribacter* presented an intense increase in abundance in the on-plateau M2 site as compared to the HNLC waters (**Supplementary Fig. 5F**). *Polaribacter* are heterotrophic bacteria belonging to the family Flavobacteriaceae. They are widely distributed in marine habitats and rely heavily on the phytoplankton-derived dissolved organic matters [[48-51](#_ENREF_48)]. It can develop rapidly under favorable conditions, and was reported to contribute to a large share of biomass production during the spring phytoplankton blooms surrounding Kerguelen Island [[52](#_ENREF_52)]. Another interesting Flavobacteriaceae genus is *Formosa* (**Supplementary Fig. 5F**; named as Hel1-33-131 under the GTDB taxonomy system [[53](#_ENREF_53)] in **Supplementary Table 3**), which accounted for up to 24% of all bacteria during diatom-dominated spring blooms off the Helgoland Island in North Sea [[54](#_ENREF_54)]. In general, marine Flavobacteriaceae are active degraders of biopolymers and was observed to be abundant in nutrient-rich habitats [[49](#_ENREF_49), [55](#_ENREF_55), [56](#_ENREF_56)]. They are key players in Southern Ocean carbon cycling. Therefore, it is important to understand the underlying enzymatic mechanisms and adaptation that drive the specialization of these highly competitive bacteria in the on-plateau M2 site. To the contrary, the Gammaproteobacterial genus *Psychrobacter* were of greater success in HNCL waters, characterized with a log fold change of taxon ratios below 0 (**Supplementary Fig. 5F**). We have a Southern Ocean MAG (MAG\_130) from this genus, which showed an overall high abundance in all three sampling sites and were featured with an average coverage of depth per million reads ~12 times higher in M3 and M4 than M2 (**Fig. 1** and **Supplementary Table 3**). *Psychrobacter* is known to produce siderophores in iron limiting conditions and able to increase growth rate under iron fertilization, reassuring its success in both environments [[57](#_ENREF_57)].

***Genomic differences between MAG\_103 and MAG\_62***

We identified a putative methanogenesis gene cluster in MAG\_62, which was conserved in all but one UBA9926 genomes and absent from the UBA4421 genus (**Supplementary Table 6**). It consisted of one trimethylamine methyltransferase (*mttB*; K14083), a corrinoid methyltransferase, a protein of unknown function, a methyltetrahydrofolate-homocysteine methyltransferase (MTR; K00548). and a ferredoxin. Methylotrophic prokaryotes, which consume methylated compounds such as trimethylamine and dimethyl sulfate, were reported to be common in marine and hypersaline, sulfate-rich sediments [[58](#_ENREF_58)]. Just adjacent to putative methanogenesis loci, there was another cluster consisting of four sarcosine oxidase-encoding genes (*soxADG*), which could only be found in one UBA4421 genomes. Another sarcosine oxidase subunit, *soxB*, was also identified in all UBA9926 genomes and 2 UBA4421 gnomes but absent from MAG\_103. Sulfur oxidation mediated by Sox proteins have also been widely identified as an important process of energy metabolism in sediments and deep-sea vent ecosystems [[59](#_ENREF_59), [60](#_ENREF_60)]. The co-localization and conservation of the *sox* and *mtt* genes in MAG\_62 suggested a deeper water origin of this species, which is coherent with the higher abundances of MAG\_62 in spring characterized by a deep mixed layer (**Supplementary Table 3**).

We could not detect fundamental advantages in substrate utilization that would explain the flourishing of MAG\_62 in the spring surface waters. Other possible explanations could be related to RecBCD DNA repair enzymes and the biosynthesis of antibiotics. MAG\_62 had a gene cluster encoding the three exodeoxyribonuclease V gamma subunits in a row (*recB*, *recC* and *recD*; K03582, K03583 and K03581), which was shared by all UBA9926 genomes but absent from the UBA4421 genus. RecBCD complex is responsible for the repair of DNA double-strand breaks by homologous recombination [[61](#_ENREF_61)], and was considered to play an effective role in handle antibiotic-induced oxidative DNA damage [[62](#_ENREF_62)]. Additionally, RecD was also reported to play a critical role for deep-sea bacteria to grow under high pressure [[63](#_ENREF_63)]. Prokaryotes have evolved various strategies, such as motility, antibiotic synthesis and antibiotic resistance, to increase their probability to survive and increase their competence for larger shared of resources [[64](#_ENREF_64)]. Both MAG\_103 and MAG\_62 were capable of producing multiple antibiotics, but their approach to protect themselves from antibiotic-induced damages were not the same (**Supplementary Table 6**). Penicillin-binding proteins (PBPs) are targets of β-lactam antibiotics. Prokaryotes can confer antibiotic resistance by modifying their PBP structures to either reduce the binding capability of antibiotics or directly degrade the antibiotics [[65](#_ENREF_65)]. Both the UBA9926 and UBA4421 genera could produce penicillin-binding protein 1A (*mrcA*; K05366) and penicillin-binding protein 2 (*mrdA*; K05515) which were considered to participate in beta-Lactam resistance. Besides, MAG\_62 could produce a putative multi antimicrobial extrusion protein (PF01554.18) which was reported to mediate resistance to multiple antimicrobial agents [[66](#_ENREF_66)]. As mentioned above, the RecBCD complex also plays are role in antibiotic resistance, therefore, MAG\_62 might be better equipped to deal with antibiotic stress.

**Supplementary Figure Legends**

**Supplementary Fig. 1** Three sampling sites (M2, M3 and M4) of the MOBYDICK cruise. **A**, Station M2 is located above the central Kerguelen Plateau, where Chlorophyll *a* (Chl *a*) concentrations are higher than at the off-plateau stations (M3 and M4). The colour code in (**A**) represents the monthly (March) climatological mean of Chl *a* concentration (mg m-3) from 2003-2018. Data was collected from NASA’s moderate-resolution imaging spectroradiometer (MODIS-Aqua) dataset (<https://oceancolor.gsfc.nasa.gov/data/aqua/>). The location of the Antarctic Polar Front (APF) is also highlighted in (**A**), showing that during the MOBYDICK cruise M2 and M4 were located south of the polar front in Antarctic waters whereas M3 was situated in the polar front zone. The monthly APF data was collected from Pauthenet et al., (2018) [[67](#_ENREF_67)]. The yearly mean latitude of APF at a resolution of 0.1 longitude degree (the black curve in **A**) was determined by calculating the mean value of all 12-month measurements within a 0.2 degree sliding window, which moves 0.1 longitude degree at one time. The standard error of the mean (SEM) was also calculated and is shown as the grey shading around the black curve in (**A**). The seasonal mean and SEM, covering austral summer and autumn, were determined in the same way, except that only the data from December to May was involved (the blue curve with light blue shading in **A**). The calculation scheme is exemplified in the inset panel of (**A**), showing how the data at 60 °E were assessed: The blue dots represent data for austral summer and autumn, while the orange ones are for austral winter and spring; The mean and SEM values are calculated based on data points falling within the sliding widow which is shaded in grey. **B**, The variation of Chl *a* concentration (mg m-3) of the three sampling sites (M2 in orange, M3 in blue and M4 in green) from October 2017 to March 2018. The background shade in orange displays previous observations from 1998 to 2017, with Mean+S.D. as the upper and Mean-S.D. as the lower limits of the shaded region. Data was collected from Copernicus Marine Service (<http://marine.copernicus.eu/>). **C**, Depth profiles of temperature (°C), salinity and dissolved organic carbon (DOC, µM) as determined during the MOBYDICK cruise. The measurements from 0 to 200 m are illustrated. The colour scheme for data points from different sampling sites and distinct visits is shown at the bottom.

**Supplementary Fig. 2** **A**, Sequencing coverage and complexity of microbial communities in our samples determined by Nonpareil [[8](#_ENREF_8)]. Nonpareil examines the redundancy of the reads in metagenomic datasets to assess the average coverage and predict the amount of sequences required to achieve full coverage [[8](#_ENREF_8)]. Colours indicate different datasets used for the estimation, and “R1/R2” represents the forward/reverse reads. Solid lines show the projection curves of the estimated coverage per sequencing effort. The empty circles indicate the actual size of each metagenomic dataset (x-axis) and the corresponding fraction of microbiome covered by DNA sequencing (y-axis). The horizontal dashed lines in red indicate the 95% and 100% coverages. The arrows at the bottom indicate the sequence diversity (Nonpareil *Nd*). Curves on the right display higher sequence diversity than those on the left. **B** and **C**, The 949 228 assembled contigs were queried against the NCBI nt database. The statistics of the bitscore (**B**) and percentage of identity (**C**) of the best alignment for each contig are represented by bar plots. **D** to **G**, The similarity between our 3 003 586 protein-coding genes and the NCBI nt (or nr) database were also assessed. The corresponding statistics of the best hit for each gene are summarized here. **D** and **F**, the x axis represents different ranges of the bitscores, and the y axis shows the percentage of genes falling into each category. Similarly, **E** and **G**, the x axis represents the percentage of identical match, and the y axis illustrates the percentage of genes in each group.

**Supplementary Fig. 3** Maximum-likelihood phylogenetic tree of the 133 MAGs (**A**) and their corresponding sequence-discrete populations revealed by metagenomic read recruitment (**B** to **G**). **A**, Bootstrap values are shown in percentages at internal nodes. Leaves (MAGs) are coloured according to their taxonomic assignment at the class level. **B**, Fraction of reads recruited to a MAG (y axis) at each percentage of nucleotide identity level (x axis) by using the metagenome reads from Station M2. Bars are coloured in proportion to their heights. MAGs are arranged according to their position on the phylogenetic tree. **C** to **D**, Statistics of recruited metagenome reads from Station M3 and M4. **E** to **G**, Statistics of recruited metatranscriptome reads from Station M2, M3 and M4.

**Supplementary Fig. 4** Comparisons between MOBYDICK Southern Ocean MAGs and the TARA assemblies, including both draft genomes and un-binned contigs, suggesting high genetic diversity at the DNA level but low functional novelty with regards to proteins. **A**, Pairwise ANI values between our MAGs and the TARA draft genomes were calculated by using fastANI [[17](#_ENREF_17)]. If one of our MAGs has multiple matches in TARA, only the highest ANI value was used for the statistics. MAGs without a TARA counterpart (ANI < 75%) were not considered. The violin plot describes the distribution and the density trace of the 95 (out of 133) ANI values. The 95 spotted dots are grouped and coloured according to three ranges: “red” represents for > 95% intra-species ANI values; “salmon” are ANI values ranging between 83 and 95%; and “black” shows < 83% (but ≥ 75%) inter-species ANI values [[17](#_ENREF_17)]. (**B**, **C** and **E**) shows the statistics of the read recruitment analyses against all the metagenome assemblies, including 949 228 contigs and 133 MAGs from our study, as well as 11 709 809 contigs and 3 588‬ draft genomes from the TARA Ocean Project [[14-16](#_ENREF_14)]. Data were summarized according to their sampling locations as specified in the original papers [[14-16](#_ENREF_14)], and the list of acronyms is provided at the end of this paragraph. The boxplots illustrate the distribution of the fraction of reads in each sample (M2, M3 or M4) that could be mapped to the assemblies by BBMap (v38.22) [[18](#_ENREF_18)] with a minimum percentage of identity at 90% and further normalized by the size of each dataset to a “per Mbp” unit. Non-parametric pairwise Wilcoxon rank sum tests with Holm adjustment demonstrated that the non-Southern Ocean TARA draft genomes recruit a significantly lower amount of reads from each sample (M2, M3 or M4), as compared to our 133 MAGs (p.adj < 0.05; **B** and **C**). In contrary to **B** and **C**, the Arctic contigs showed an overall higher rate of read recruitment than our assemblies (p.adj < 0.05; **E**). **D**, Our 949 228 contigs were queried against the 41 TARA Arctic datasets using BLASTN (v2.7.1) [[13](#_ENREF_13)]. The violin plot describes the distribution and the density trace of the percentage of our contigs which have an Arctic match with high sequence similarity. In **F**, Statistics of TARA Arctic metagenome read recruitment to the TARA Arctic contigs and our MOBYDICK contigs. Similar to **E**, Arctic contigs recruited significantly more reads per Mbp than ours. In **G** and **H**, Prodigal (v2.6.3) was used to recover protein sequences from the TARA assemblies under meta mode (-p meta) [[68](#_ENREF_68)]. Protein sequences were pooled according to their sampling locations as mentioned before, resulting one combined site-specific proteome for our study (named as “SO MAGs”), 12 for Delmont, et al. [[14](#_ENREF_14)] and 10 for Tully, et al. [[15](#_ENREF_15)]. Orthologous groups shared among these 23 site-specific proteomes were identified using OrthoFinder (v2.2.3) [[19](#_ENREF_19)]. **G**, A total of 519 468 orthologous groups were identified, among which 84 192 were shared by all. Only 71 orthologous groups were unique to our MAGs (Singletons were not included). **H**, Barplots illustrate the percentage of proteins in each site-specific proteome that are involved in orthologous groups with at least one member from our MAGs. The TARA SOC MAGs have the highest percentage of shared orthologs, whereas the TARA ASW ones have the lowest percentage. Due to computational limitation, the TARA Arctic datasets were not included in the orthology analysis. Acronyms: Delmont, et al. has 12 locations, including ANE (Atlantic northeast), ANW(Atlantic northwest), ASE (Atlantic southeast), ASW (Atlantic southwest), ION (Indian Ocean north), IOS (Indian Ocean south), MED (Mediterranean), PON (Pacific Ocean north), PSE (Pacific Ocean southeast), PSW (Pacific Ocean southwest), RED (Red Sea) and SOC (Southern Ocean) [[14](#_ENREF_14)]. And, Tully, et al. consists of 10 sites, including ARS (Arabian Sea), CPC (Chile-Peru coastal), EAC (east Africa coastal), IN (Indian Ocean), MED (Mediterranean), NAT (north Atlantic), NP (north Pacific), RS (Red Sea), SAT (south Atlantic) and SP (south Pacific) [[15](#_ENREF_15)].

**Supplementary Fig. 5** Taxonomic profiling confirmed that natural iron fertilization and DOC availability altered the diversity and abundance of the prokaryotic communities across the sampling sites. **A** to **D**, Taxonomic profiling through 16S rRNA assemblies (as defined under the SILVA v138.1 taxonomy). **A**, The outmost circle shows the phylogenetic tree constructed using FastTree [[25](#_ENREF_25)] based on the MATAM/phyloFlash-assembled prokaryotic 16S rRNA sequences [[20](#_ENREF_20)] (**Supplementary Methods**). Taxonomic assignment of each 16S rRNA is labelled on the tree tip. The taxonomic colour scheme for the labels is shown in the middle. The red numbers on the internal nodes (≥ 0.5) indicate the local support values provided by the Shimodaira-Hasegawa (SH) test [[25](#_ENREF_25)]. **B**, shows the stacked bar charts of the relative abundance at the class level of the prokaryotic communities from different sampling sites. **C**, The Shannon index of each sampling site suggests that M3 and M4 share similar and higher species diversity and evenness as compared to M2. **D**, Double principal coordinate analysis (DPCoA), taking both phylogenetic and abundance data into account, reveals a strong clustering of different taxonomic groups by sampling regions (off- vs. on-plateau). **E** and **F**, Taxonomic profiling based on the direct assignment of metagenomic reads to phylogenetic categories (as defined under the NCBI taxonomy). **E**, The stacked bar charts of the relative abundance at the class level from different sampling sites estimated by mOTUs2 [[35](#_ENREF_35)]. **F**, Statistics of the fold changes between taxon ratios across sampling sites. The taxon ratios measured the “relative abundance” of a taxonomic group to that of *Candidatus Pelagibacter*, instead of to the whole community. The fold change of taxon ratios, by crossing out potential multiplicative systematic biases [[36](#_ENREF_36)], was used as an indicator of the relative fitness of a taxon as compared to the ubiquitous SAR11 population in contrasting environments (M2 vs. M3, or M2 vs. M4). The median value of the log fold change varied between 1.16 and 1.34, indicating approximately one-fold increasing in the relative fitness of the studied genera in comparison with *Candidatus Pelagibacter* in the on-plateau region. Genera with anomalously deviated fold changes could be quintessential representatives of specific environment, such as *Polaribacter*, *Formosa* and *Psychrobacter* (**Supplementary Results**). The blue and green colors are used to represent different comparison pairs. The shapes are used to distinguish tools. The size of each symbol is in proportion with the total number of reads assigned to the taxonomic group by all tools.

**Supplementary Fig. 6** Taxonomic composition and abundance of KEGG Orthology groups (KOs) in the SO metagenomes (**A** to **D**) and metatranscriptomes (**E** to **J**). In **A** to **D**, the relative abundance of a taxonomic category *i* given a KO*j* was measured: **A**, by the proportion of genes, among all the protein-coding genes belonging to KO*j*, assigned to taxonomy *i*, as defined by the division of $\frac{f\_{Counts}(gene \in Tax\_{i} ∩ KO\_{j})}{f\_{Counts}(gene \in KO\_{j})}$; **B to D**, based on gene abundance, in the form of FPKM (fragments per kilobase of transcript per million mapped reads) values, as defined by the division of $\frac{\sum\_{gene \in Tax\_{i} ∩ KO\_{j}}^{}f\_{FPKM}(gene)}{\sum\_{gene \in KO\_{j}}^{}f\_{FPKM}(gene)}$ in each metagenome. In **E** to **G**, the absolute abundance of each KO group in each metagenome was estimated using the normalized transcripts per liter according to the internal standards protocol [[40](#_ENREF_40)], as defined by the sum of $\sum\_{gene \in KO\_{j}}^{}f\_{Trascripts L^{-1}}(gene)$. The values of the total transcripts L-1 assigned to each KO are shown in the middle of each cell. The colour code, from blue via yellow to red, represents the total number of transcripts L-1 assigned to each KO group from low to hight. In **H** to **J**, the relative abundance of a taxonomic category *i* given a KO*j* was evaluated by the division of $\frac{\sum\_{gene \in Tax\_{i} ∩ KO\_{j}}^{}f\_{Trascripts L^{-1}}(gene)}{\sum\_{gene \in KO\_{j}}^{}f\_{Trascripts L^{-1}}(gene)}$ in each metatranscriptome. The mean of the three replicates was used to represent their corresponding metatranscriptomic sample. The KO groups in all panels are ordered decreasingly according to the value of transcripts L-1 in M2 (as shown in **E**). Only functional groups consisting of at least 10 genes, out of 3 003 587 protein-coding genes predicted from the metagenome assemblies, were shown.

**Supplementary Fig. 7** Abundance of functional groups in each sample and the relative contribution of different taxonomic categories to each functional group were evaluated based on the carbohydrate-active enzymes (CAZyme) [[69](#_ENREF_69)]. The calculation and the order of panels within each section are the same as shown in **Supplementary Fig. 6**.

**Supplementary Fig. 8** Statistics of functional diversity (**A**) and shifts in taxonomic composition within functional groups (**B**) across sampling sites, based on the community-level metagenomic gene abundance represented by FPKM (fragments per kilobase of transcript per million mapped reads) values (**B**-**D** in **Supplementary Fig. 6**-**7**). In **B**, the relative contribution (%) of a specific taxonomic category (e.g., Gammaproteobacteria) to a functional group (e.g., ferrous iron transporter FeoA) in each sampling station was calculated (Materials and methods). Shifts in the relative contribution across stations were estimated using the ratio of the relative contribution in M2 to that in M3 (or M4) and visualized by violin plots. A ratio value less than 1 indicates that the taxonomic category accounts for a larger share of the genes of a functional group in the off-plateau HNLC waters, and vice versa. Multiple databases were considered, including CAZy, FeGenie, KEGG, Pfam, Superfamily and TCDB. Five dominant taxonomic groups in gene pool and transcript inventories across all sampling sites were shown (**Supplementary Fig. 6**-**7**). Colour code is the same as **Supplementary Fig. 6-7**. Only functional groups consisting of at least 50 genes, out of the 3 003 587 protein-coding genes predicted from the metagenome assemblies, were used in the calculation.

**Supplementary Fig. 9** **A**, An exemplary illustration of the influence of genome abundance on the assessment of gene expression levels, showing that fluctuations in transcript abundance could be a result of shifting genome copies rather than changes in expression levels. **B**, The pipeline to generate the metagenome-normalized metatranscriptomic pseudo-read count table. All the analyses were performed under R environment (v3.6.1). Initially, two matrices of read counts of genes were generated using featureCounts (v2.0.0) [[70](#_ENREF_70)]. Each matrix contains 3 003 586 rows, representing the total number of protein-coding genes predicted in our 949,228 assembled contigs (≥ 1 kbp). The metagenomic matrix consists of 3 columns representing 3 sampling sites M2, M3 and M4. The metatranscriptomic matrix has 9 columns corresponding to 3 biological replicates × 3 sampling sites. Secondly, genes with low read counts were filtered out. Only genes whose FPM (fragments per million mapped fragments) value is no less than 1 in at least 1 sample, which must be satisfied in both metagenomic and metatranscriptomic data, were remained. Variance stabilizing transformation (VST) was carried out on both matrices. Given that $Abundance\_{transcript}\~Abundance\_{gene}×Expression Level (EL\_{gene})$, we used the ratio between transcript abundance and gene abundance as approximations of gene expression levels ($EL\_{gene}\~\frac{Abundance\_{transcript}}{Abundance\_{gene}}$). Here, $Abundance\_{transcripts}$ represents the metatranscriptomic read counts of a gene in a sample, and $Abundance\_{gene}$ equals to the metagenomic read counts of a gene in a sample. Since the transformed data was on log2 scale, the formula could also be written as $EL\_{gene}\~2^{VST\left(Abundance\_{transcript}\right)-VST(Abundance\_{gene})}$. Then, in each sample (each column), $EL\_{gene}$ values are normalized (divided) by the corresponding sample (column) maximum. Finally, we converted the normalized ratios to pseudo counts by multiplying with 106 and rounding to the nearest integer. **C**, Heatmap of the sample-to-sample distance, and **D**, Principal component analysis (PCA) based on the normalized read counts after variance stabilizing transformation by DESeq2.

**Supplementary Fig. 10** Counts of significantly differentially expressed genes (SDEGs) in CAZymes [[69](#_ENREF_69)] across contrasting oceanic regions (on- and off-plateau). Each row represents one functional group. For CAZymes, enzyme families are further clustered based on their functions, including the breakdown, biosynthesis and modification of carbohydrates and glycoconjugate (e.g., glycoside hydrolases (GHs) hydrolyse the glycosidic bonds between carbohydrates). The three vertical panels show: (1st) The comparison of SDEGs obtained with vs. without the metagenome-based normalization. The black segment of each bar represents the number of SDEGs identified by both methods; the grey segment displays the SDEGs only found based on the metagenome-normalized metatranscriptomic pseudo counts (with normalization); the white segment shows the SDEGs only reported without the normalization procedure. (2nd) The SDEGs based on the metagenome-normalized metatranscriptomic pseudo counts. (3rd) The log2-based fold change values of SDEGs involved in the 2nd panel. In the bi-direction bar plots, the bars pointing to the left indicate the number of genes that are significantly higher expressed in the on-plateau iron-fertilized M2 site, as compared to the off-plateau HNLC M3 and M4 sites. To the contrary, the bars pointing to the right represent genes that are significantly higher expressed in the off-plateau HNLC waters. The colour scheme of taxonomy is shown on top.

**Supplementary Fig. 11** Counts of significantly differentially expressed genes (SDEGs) in KEGG Orthology groups (KOs) [[71](#_ENREF_71)] across contrasting oceanic regions (on- and off-plateau). We built a list of KOs relevant to iron uptake and metabolism, the tricarboxylic acid (TCA) cycle, the Embden Meyerhof-Parnas (EMP) pathway, proposed C1 and methylated compound oxidation pathways, and etc (**Supplementary Table 4**). The vertical panel arrangement and colour schemes are the same as shown in **Supplementary Fig. 10**. Several KOs that are involved in multiple pathways are displayed multiple times.

**Supplementary Fig. 12** The distribution of significantly differentially expressed genes in the 133 MAGs among diverse functional categories related to iron uptake and carbon metabolism. From left to right, the panels represent the phylogenetic tree (the same as shown in Fig. 1), the KOs involved in the EMP glycolysis/gluconeogenesis, C1 and methylated compound oxidation pathways, oxidative phosphorylation, nitrogen metabolism, transcription factors, ABC transporters, sulfur metabolism, Sec (secretion) system, type II secretion system, type IV secretion system and the two-component system. The symbol and colour schemes are the same as shown in **Fig. 5**.

**Supplementary Fig. 13** Polysaccharide utilization loci (PULs) identified in MAG\_78 (**A**), 134 (**B**), 3 (**C**) and 73 (**D**), as well as fucose utilization loci found in MAG\_58, 98, 6 and 17 (**E**). Arrows represent genes and direction of arrows illustrates their strand locations. The colour scheme is: “orange” represents genes that are significantly higher expressed in the on-plateau iron-fertilized water (M2 site); “yellow” shows genes that are higher, but not significantly, expressed in the on-plateau site; “blue” indicates genes that are significantly higher expressed in the off-plateau HNLC waters (M3 and M4 sites); “cyan” are genes that are higher, but not significantly, expressed in the HNLC waters. Genes encoding the fucose metabolic pathway of *Lentimonas* sp. CC4 [[42](#_ENREF_42)] is used as the reference to arrange their counterparts identified in MAGs from the *Verrucomicrobiae* class (under the GTDB taxonomy system [[53](#_ENREF_53)]). Light blue shades between genes are used to indicate orthologous groups and facilitate visualization, because the incompleteness of MAGs leads to the fragmented distribution of genes among contigs.

**Supplementary Fig. S14** (A) A gene cluster potentially engaged in using light as a complementary energy resource was identified exclusively in the UBA4421 genus. A gene encoding bacteriorhodopsin was also detected in one of the UBA9926 genome (GCA\_002728935), but genes adjacent to it do not form similar light-harvesting gene clusters as found in the UBA4421 genus. Light blue shades between genes are used to indicate orthologous genes. The colour scheme is: “orange” represents genes that are significantly higher expressed in the on-plateau iron-fertilized water (M2 site); “yellow” shows genes that are higher, but not significantly, expressed in the on-plateau site; “white” are genes from reference genomes. (B) Bacteriorhodopsins from different genera are well separated on the phylogeny tree with high support values, indicating distinct origins of the rhodopsin genes. The emergence of the accessory functions of light harvesting might be a result of adaptive evolution. (C) Sequence alignment of transmembrane peptide segment of bacteriorhodopsin sequence of MAG\_103 and its top 20 matches in the MicRhoDE database [[43](#_ENREF_43)], highlighting variation in residues involved in spectral tuning and ion pumping.

**Supplementary Data**

Sequences of the two internal standards add to RNA libraries:

>MTST51006NT

GGGUUCGGUGGUCUAUACUACUACCUAAGUUGGAUGUACUGGUGGAAGUGCUACCAACACAGUAAUGCUGGUAUAGGUAG

GCAACACUACGACUUCAGGAAGAGUCUAACGAAUGUAUGCAUAAUACUAAUGCCUUACAUGUGGAAGCACCCUAUAACGG

ACAGGAUGAGGCACAGGCACAUGUGCAGGCAAGCUUUCAAGUGGAUGUGCGGUGAAAAUAAGUUCUGGCUAGUAAGGGCU

AUGGAAAAUCAACCUGACGAAAGGAUACUAGCUCAAAUGACGAUAACGGACAGUGACUGGCAACCUGAAGAAUGGUACAA

GAAGAGGCACGACCCUGGUGAAAAUGACGUAAUAAGGUGCGUAUACAUAGGUAUAGAAAAUCUAGUAAAUGCUACGUGCC

CUGACAUGGAAGACUACUACGCUAUGACGGGUAAUAAGCCUCUACUAGAACUAAAUAGUAUAGGUCCUUGCACGCAAUGC

ACGGUACACAAGCUAGAAGGUGUACACUGCAUAUGGUGGAUAGUAAGGAGGGACCACUUCCCUGUACCUAUAAUACAAAU

AGUAGACGUAUUCAAUCUAUACAAUUUCGCUAGUGGUACGGUACUAUGCAUACAACACGCUGCUCACCCUUGGGGUGACU

GGAUGUUCGACGUACAAUACGAAAGUUGCAGGAUGUACAGGUGGUGGAUGACGAGGAAUGACUGGAGUGGUCCUAAUAAG

UGGAGUGGUGCUCACAGUAUAUGCCAACCUCACUGCUGCAGGAGUGACGACAGUAGGGUAAGUAAGAGGAUGGCUACGGU

AACGAAGGAAGUAGUAGAAAUGAGUCACAUGGACCUAAAGAGGAGUGCUUACUGCAAUAGGACGCAACUAGAAGAAUACG

ACGCUUUCUACACGAGGUGGAAGUUCGUACCUUGGAUGUACCCUGCUCCUCUACCUUGCGAAGUACAAGACUUCGUAACG

AGGAGGACGUUCGACUACCCUGACCCUACGGCUUGCGGUCUAGUUU

>MTST61006NT

GGGAAUCUAAUAACGGAACCUAUAUGGUACGUAAGUAUGAGUAGUGGUUUCGGUGUAAAUUUCAGGGACCAACAAGGUGC

UAUAGACGUAAGGCUAGCUUGCCUACAAAUAGACCCUAAGGUAUGCCACCAAAGGGGUAGUAAUAGUUACUGGUGGAAUU

GGAGUUGGCUAGACAUGAAUACGUUCUACCCUGCUUACUUCUACGUAAUACAAAGUUGGGAAAGGUGGACGGACUGGAAG

UACAAUUACUACCCUGGUGCUGUAAUGCCUAGGGGUGAACCUAAGUGGACGCCUUGCAGGCAAUGCAUGCCUAGUCACAU

AAAUGGUAGGACGAAUAGGAUGCAAAAGGCUCUAGGUUUCCAAAGGAGGGUAUACUGCUACCCUUUCAUAGAAGCUGAAA

UGAGGGAAUGCAUGGAAAGUGACCCUCACCAAGACAAUGUAAAGUGGAUAAUAUUCCUAAAGGUAAAUGUAAUAUACUUC

AUGUGGGACAUAUACGUAAGGGCUGAAUGCAAUCUAACGCAACCUUUCAUGUUCUGGAUGAAGCAAGGUUUCAGGUUCCU

AAGUUUCACGCACGAAUUCAAUCACGACACGCUACACUGGUACCUACACCAACUAGGUAGUGCUCCUAAUGUAGCUGGUG

GUUGCAGGACGAAGAGUUGCCAAACGGACUACAAUUUCAAUCACCCUGACGUAAAUUACCCUAAUCAAUGCGACUUCUGC

CCUUUCAUGCACUGGCUAUGCAAUGAAGCUAUAAAGAAGGACAAUAGUCUAACGAUAAAUAUAAGGUGCGAACAAGAAUG

CAAUUGGGGUGGUGACGACCACGUAGACGCUGCUUGCAUACCUCAAAUGUGCUACAGGAUGGCUAUAGUAAAUACGAGGA

GGAGUCACCUAUACGAAUACAGUGUAUGCAAUAGGGAAAGGUACACGCUAGAAAUGGUACCUACGGCUCAACCUCACAUA

AGGAUAGCUCACAUAAAUUGGGUAUUCGACUUCAGGGUACACGUUU

**Supplementary Code Availability**

All software used in this study are listed below:

*BBTools*: <https://jgi.doe.gov/data-and-tools/bbtools>; *Biostrings*: <https://bioconductor.org/packages/release/bioc/html/Biostrings.html>; *BLAST*: ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST; *Bowtie2*: <https://sourceforge.net/projects/bowtie-bio/files/bowtie2>; *Centrifuge*: <https://ccb.jhu.edu/software/centrifuge>; *CD-HIT*: <https://github.com/weizhongli/cdhit>; *CheckM*: <https://ecogenomics.github.io/CheckM>; *CONCOCT*: <https://github.com/BinPro/CONCOCT>; *Circlize*: <https://cran.r-project.org/web/packages/circlize>; *dbCAN2*: <http://bcb.unl.edu/dbCAN2/index.php>; *DADA2*: <https://github.com/benjjneb/dada2>; *DESeq2*: <https://bioconductor.org/packages/release/bioc/html/DESeq2.html>; *DIAMOND*: <http://www.diamondsearch.org/index.php>; *eggNOG-mapper*: <https://github.com/eggnogdb/eggnog-mapper>; *FastANI*: <https://github.com/ParBLiSS/FastANI>; *FastQC*: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>; *FastTree*: <http://www.microbesonline.org/fasttree>; *FeatureCounts*: <http://subread.sourceforge.net>; *FeGenie*: <https://github.com/Arkadiy-Garber/FeGenie>; genoPlotR: <http://genoplotr.r-forge.r-project.org>; *GhostKOALA*: <https://www.kegg.jp/ghostkoala>; *GTDB-Tk*: <https://github.com/Ecogenomics/GTDBTk>; *HMMER*: <http://hmmer.org>; *IQ-Tree*: <http://www.iqtree.org>; *Kaiju*: <http://kaiju.binf.ku.dk>; *KofamKOALA*: <https://www.genome.jp/tools/kofamkoala>; *Kraken2*: <https://ccb.jhu.edu/software/kraken2>; *MAFFT*: <https://mafft.cbrc.jp/alignment/software>; *MATAM*: <https://github.com/bonsai-team/matam>; *MaxBin2*: <https://sourceforge.net/projects/maxbin2>; *MEGAHIT*: <https://github.com/voutcn/megahit>; *MetaBAT2*: <https://bitbucket.org/berkeleylab/metabat>; *MetaWRAP*: <https://github.com/bxlab/metaWRAP>; *mOTUs2*: <https://motu-tool.org>; *mvoutlier*: <https://cran.r-project.org/web/packages/mvoutlier>; *Nonpareil*: <https://github.com/lmrodriguezr/nonpareil>; *OrthoFinder*: <https://github.com/davidemms/OrthoFinder>; *phyloFlash* <https://github.com/HRGV/phyloFlash>; *Phyloseq*: <https://bioconductor.org/packages/release/bioc/html/phyloseq.html>; *Prodigal*: <https://github.com/hyattpd/Prodigal>; *R*: <https://www.r-project.org>; *SAMTools*: <http://www.htslib.org>; *SortMeRNA*: <https://bioinfo.lifl.fr/RNA/sortmerna>; *Superfamily*: <http://supfam.org/SUPERFAMILY>; *SSU-ALIGN*: <http://eddylab.org/software/ssu-align>; *trimAl*: <http://trimal.cgenomics.org>; *Trimmomatic*: <http://www.usadellab.org/cms/?page=trimmomatic>; *Trim Galore*: <https://www.bioinformatics.babraham.ac.uk/projects/trim_galore>.

**Supplementary References**

1. Tremblay L, Caparros J, Leblanc K, Obernosterer I. Origin and fate of particulate and dissolved organic matter in a naturally iron-fertilized region of the Southern Ocean. Biogeosciences. 2015;12(2):607-21.

2. Blain S, Sarthou G, Laan P. Distribution of dissolved iron during the natural iron-fertilization experiment KEOPS (Kerguelen Plateau, Southern Ocean). Deep Sea Res Part II Top Stud Oceanogr. 2008;55(5-7):594-605.

3. Blain S, Queguiner B, Armand L, Belviso S, Bombled B, Bopp L, et al. Effect of natural iron fertilization on carbon sequestration in the Southern Ocean. Nature. 2007;446(7139):1070-4.

4. Moran MA, Satinsky B, Gifford SM, Luo H, Rivers A, Chan LK, et al. Sizing up metatranscriptomics. ISME J. 2013;7(2):237-43.

5. Satinsky BM, Gifford SM, Crump BC, Moran MA. Use of internal standards for quantitative metatranscriptome and metagenome analysis. Methods Enzymol. 2013;531:237-50.

6. Andrew S. FastQC: a quality control tool for high throughput sequence data. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>. 2010.

7. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114-20.

8. Rodriguez RL, Gunturu S, Tiedje JM, Cole JR, Konstantinidis KT. Nonpareil 3: fast estimation of metagenomic coverage and sequence diversity. mSystems. 2018;3(3).

9. Li D, Luo R, Liu CM, Leung CM, Ting HF, Sadakane K, et al. MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. Methods. 2016;102:3-11.

10. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 2006;22(13):1658-9.

11. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357-9.

12. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25(16):2078-9.

13. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403-10.

14. Delmont TO, Quince C, Shaiber A, Esen OC, Lee ST, Rappe MS, et al. Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes. Nat Microbiol. 2018;3(7):804-13.

15. Tully BJ, Graham ED, Heidelberg JF. The reconstruction of 2,631 draft metagenome-assembled genomes from the global oceans. Sci Data. 2018;5:170203.

16. Ibarbalz FM, Henry N, Brandao MC, Martini S, Busseni G, Byrne H, et al. Global trends in marine plankton diversity across kingdoms of life. Cell. 2019;179(5):1084-97 e21.

17. Jain C, Rodriguez RL, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun. 2018;9(1):5114.

18. Bendall ML, Stevens SL, Chan LK, Malfatti S, Schwientek P, Tremblay J, et al. Genome-wide selective sweeps and gene-specific sweeps in natural bacterial populations. ISME J. 2016;10(7):1589-601.

19. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019;20(1):238.

20. Pericard P, Dufresne Y, Couderc L, Blanquart S, Touzet H. MATAM: reconstruction of phylogenetic marker genes from short sequencing reads in metagenomes. Bioinformatics. 2018;34(4):585-91.

21. Gruber-Vodicka HR, Seah BKB, Pruesse E. phyloFlash: rapid small-subunit rRNA profiling and targeted assembly from metagenomes. mSystems. 2020;5(5).

22. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013;41(Database issue):D590-6.

23. Nawrocki E. Structural RNA homology search and alignment using covariance models 2009.

24. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13(7):581-3.

25. Price MN, Dehal PS, Arkin AP. FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS One. 2010;5(3):e9490.

26. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ. 2019;7:e7359.

27. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One. 2013;8(4):e61217.

28. Thioulouse J, Dray S, Dufour A-B, Siberchicot A, Jombart T, Pavoine S. Analysing patterns of biodiversity. Multivariate Analysis of Ecological Data with ade4. New York, NY: Springer New York; 2018. p. 281-94.

29. Dray S, Pavoine S, Aguirre de Carcer D. Considering external information to improve the phylogenetic comparison of microbial communities: a new approach based on constrained Double Principal Coordinates Analysis (cDPCoA). Mol Ecol Resour. 2015;15(2):242-9.

30. Steinegger M, Mirdita M, Soding J. Protein-level assembly increases protein sequence recovery from metagenomic samples manyfold. Nat Methods. 2019;16(7):603-6.

31. Ye SH, Siddle KJ, Park DJ, Sabeti PC. Benchmarking metagenomics tools for taxonomic classification. Cell. 2019;178(4):779-94.

32. Kim D, Song L, Breitwieser FP, Salzberg SL. Centrifuge: rapid and sensitive classification of metagenomic sequences. Genome Res. 2016;26(12):1721-9.

33. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome Biol. 2019;20(1):257.

34. Menzel P, Ng KL, Krogh A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nat Commun. 2016;7:11257.

35. Milanese A, Mende DR, Paoli L, Salazar G, Ruscheweyh HJ, Cuenca M, et al. Microbial abundance, activity and population genomic profiling with mOTUs2. Nat Commun. 2019;10(1):1014.

36. McLaren MR, Willis AD, Callahan BJ. Consistent and correctable bias in metagenomic sequencing experiments. Elife. 2019;8.

37. Kopylova E, Noe L, Touzet H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics. 2012;28(24):3211-7.

38. Durham BP, Dearth SP, Sharma S, Amin SA, Smith CB, Campagna SR, et al. Recognition cascade and metabolite transfer in a marine bacteria-phytoplankton model system. Environ Microbiol. 2017;19(9):3500-13.

39. Satinsky BM, Crump BC, Smith CB, Sharma S, Zielinski BL, Doherty M, et al. Microspatial gene expression patterns in the Amazon River Plume. Proc Natl Acad Sci U S A. 2014;111(30):11085-90.

40. Satinsky BM, Gifford SM, Crump BC, Moran MA. Use of internal standards for quantitative metatranscriptome and metagenome analysis. Methods in enzymology. 531: Elsevier; 2013. p. 237-50.

41. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, et al. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 2018;46(W1):W95-W101.

42. Sichert A, Corzett CH, Schechter MS, Unfried F, Markert S, Becher D, et al. Verrucomicrobia use hundreds of enzymes to digest the algal polysaccharide fucoidan. Nat Microbiol. 2020:1-14.

43. Boeuf D, Audic S, Brillet-Gueguen L, Caron C, Jeanthon C. MicRhoDE: a curated database for the analysis of microbial rhodopsin diversity and evolution. Database. 2015;2015.

44. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772-80.

45. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25(15):1972-3.

46. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32(1):268-74.

47. Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, et al. Ocean plankton. Structure and function of the global ocean microbiome. Science. 2015;348(6237):1261359.

48. Fernandez-Gomez B, Richter M, Schuler M, Pinhassi J, Acinas SG, Gonzalez JM, et al. Ecology of marine Bacteroidetes: a comparative genomics approach. ISME J. 2013;7(5):1026-37.

49. Williams TJ, Wilkins D, Long E, Evans F, DeMaere MZ, Raftery MJ, et al. The role of planktonic Flavobacteria in processing algal organic matter in coastal East Antarctica revealed using metagenomics and metaproteomics. Environ Microbiol. 2013;15(5):1302-17.

50. Luria CM, Amaral-Zettler LA, Ducklow HW, Rich JJ. Seasonal succession of free-living bacterial communities in coastal waters of the Western Antarctic Peninsula. Front Microbiol. 2016;7:1731.

51. Dadaglio L, Dinasquet J, Obernosterer I, Joux F. Differential responses of bacteria to diatom-derived dissolved organic matter in the Arctic Ocean. Aquat Microb Ecol. 2018;82(1):59-72.

52. Obernosterer I, Catala P, Lebaron P, West NJ. Distinct bacterial groups contribute to carbon cycling during a naturally iron fertilized phytoplankton bloom in the Southern Ocean. Limnol Oceanogr. 2011;56(6):2391-401.

53. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil PA, et al. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat Biotechnol. 2018;36(10):996-1004.

54. Teeling H, Fuchs BM, Bennke CM, Kruger K, Chafee M, Kappelmann L, et al. Recurring patterns in bacterioplankton dynamics during coastal spring algae blooms. Elife. 2016;5:e11888.

55. Mann AJ, Hahnke RL, Huang S, Werner J, Xing P, Barbeyron T, et al. The genome of the alga-associated marine flavobacterium Formosa agariphila KMM 3901T reveals a broad potential for degradation of algal polysaccharides. Appl Environ Microbiol. 2013;79(21):6813-22.

56. Xing P, Hahnke RL, Unfried F, Markert S, Huang S, Barbeyron T, et al. Niches of two polysaccharide-degrading Polaribacter isolates from the North Sea during a spring diatom bloom. ISME J. 2015;9(6):1410-22.

57. Sinha AK, Parli Venkateswaran B, Tripathy SC, Sarkar A, Prabhakaran S. Effects of growth conditions on siderophore producing bacteria and siderophore production from Indian Ocean sector of Southern Ocean. J Basic Microbiol. 2019;59(4):412-24.

58. Lyu Z, Shao N, Akinyemi T, Whitman WB. Methanogenesis. Curr Biol. 2018;28(13):R727-R32.

59. Meyer B, Imhoff JF, Kuever J. Molecular analysis of the distribution and phylogeny of the soxB gene among sulfur-oxidizing bacteria - evolution of the Sox sulfur oxidation enzyme system. Environ Microbiol. 2007;9(12):2957-77.

60. Jørgensen BB, Nelson DC, Amend JP, Edwards KJ, Lyons TW. Sulfide oxidation in marine sediments: Geochemistry meets microbiology. Sulfur Biogeochemistry - Past and Present. 379: Geological Society of America; 2004. p. 0.

61. Cromie GA. Phylogenetic ubiquity and shuffling of the bacterial RecBCD and AddAB recombination complexes. J Bacteriol. 2009;191(16):5076-84.

62. Stohl EA, Seifert HS. Neisseria gonorrhoeae DNA recombination and repair enzymes protect against oxidative damage caused by hydrogen peroxide. J Bacteriol. 2006;188(21):7645-51.

63. Bidle KA, Bartlett DH. RecD function is required for high-pressure growth of a deep-sea bacterium. J Bacteriol. 1999;181(8):2330-7.

64. Hibbing ME, Fuqua C, Parsek MR, Peterson SB. Bacterial competition: surviving and thriving in the microbial jungle. Nat Rev Microbiol. 2010;8(1):15-25.

65. Sun S, Selmer M, Andersson DI. Resistance to beta-lactam antibiotics conferred by point mutations in penicillin-binding proteins PBP3, PBP4 and PBP6 in Salmonella enterica. PLoS One. 2014;9(5):e97202.

66. Morita Y, Kodama K, Shiota S, Mine T, Kataoka A, Mizushima T, et al. NorM, a putative multidrug efflux protein, of Vibrio parahaemolyticus and its homolog in Escherichia coli. Antimicrob Agents Chemother. 1998;42(7):1778-82.

67. Pauthenet É, Roquet F, Madec G, Guinet C, Hindell M, McMahon C, et al. Seasonal meandering of the Polar Front upstream of the Kerguelen Plateau. Geophys Res Lett. 2018;45(18):9774-81.

68. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 2010;11:119.

69. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Res. 2009;37(Database issue):D233-8.

70. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30(7):923-30.

71. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28(1):27-30.