

1 **Supplementary Information for**

2 **“Microbial iron and carbon metabolism as revealed by taxonomy-specific functional**

3 **diversity in the Southern Ocean”**

4
5 Ying Sun^{1*†}, Pavla Debeljak^{1,2*}, Ingrid Obernosterer¹

6 1 CNRS, Sorbonne Université, Laboratoire d’Océanographie Microbienne, LOMIC, F-66650 Banyuls/mer, France.

7 2 University of Vienna, Department of Functional and Evolutionary Ecology, A-1090 Vienna, Austria.

8 * These authors contributed equally to this study.

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

10
11
12
13
14 **This file includes:**

15 Supplementary Methods

16 Supplementary Results

17 Supplementary Figure legends

18 Supplementary Data

19 Supplementary Code availability

20 Supplementary References

21
22 **Other supplementary material for this manuscript includes the following:**

23 Supplementary Fig. 1 to 14

24 Supplementary Table 1 to 6.

25 **Supplementary Methods**

26 *Environmental conditions and sample collection*

27 The Marine Ecosystem Biodiversity and Dynamics of Carbon around Kerguelen (MOBYDICK)
28 cruise took place in the Indian Sector of the Southern Ocean during the austral late summer
29 period from Feb 18th to Mar 29th in 2018. Three stations in contrasting oceanic regions were
30 chosen for our study, including one located in the naturally iron-fertilized waters above the
31 central Kerguelen Plateau (M2) and two off-plateau stations within the High Nutrient Low
32 Chlorophyll (HNLC) waters (M3 and M4; **Supplementary Fig. 1A** and **Supplementary Table**
33 **1**). The timing of the sample collection corresponded to about two months after the peak of the
34 summer phytoplankton blooms (**Supplementary Fig. 1B**). In central Kerguelen Plateau waters
35 maximum seasonal Chlorophyll *a* (Chl *a*) concentrations, based on satellite images, were about
36 $1.5 \mu\text{g L}^{-1}$, and they were substantially lower (0.3 to $0.5 \mu\text{g L}^{-1}$) in HNLC waters. During the
37 MOBYDICK cruise in late austral summer, differences in the Chl *a* concentrations among sites
38 were less pronounced. As a consequence of the sustained seasonal phytoplankton activity,
39 concentrations of dissolved organic carbon (DOC) were higher in Kerguelen Plateau waters as
40 compared to those in HNLC waters and during previous investigations in the study region in
41 spring [1] (**Supplementary Fig. 1BC**). Concentrations of dissolved iron were shown not to
42 differ in the on- and off-plateau surface waters in summer (0.09 nM) [2], due to the rapid
43 utilization of iron supplied by natural fertilization [3]. The seasonal phytoplankton activity led
44 to a build-up of heterotrophic prokaryotic abundance, associated to higher heterotrophic
45 prokaryotic production, in fertilized waters as compared to the HNLC sites. Station M2 was
46 visited three times at an 8-day interval, and station M3 and M4 were visited twice at a
47 two-week interval. Samples were collected for all visits from surface waters (10 m) with 12 L
48 Niskin bottles mounted on a rosette equipped with a conductivity, temperature, depth sensor
49 (Seabird SBE-911 plus CTD unit).

50 *Microbial community DNA extraction and metagenome library preparation*

51 Total genomic DNA was extracted from the Sterivex filter units using the AllPrep DNA/RNA
52 kit (Qiagen, Hiden, Germany) with the following modifications: Filter units were thawed and
53 closed with a sterile pipette tip at the outflow. Lysis buffer was added (40 mM EDTA, 50 mM
54 Tris, 0.75 M sucrose) and three freeze-and-thaw cycles were performed using dry ice in ethanol
55 and a water bath at $65 \text{ }^\circ\text{C}$. Lysozyme solution (0.2 mg mL^{-1} final concentration) was added, and
56 filter units were placed on a rotary mixer at $37 \text{ }^\circ\text{C}$ for 45 min. Proteinase K (0.2 mg mL^{-1} final

57 concentration) and SDS (1% final concentration) were added and filter units were incubated at
58 55 °C with gentle agitation every 10 min for 1 h. To each filter unit, 1 550 µl RLT plus buffer
59 was added and inverted to mix. The lysate was recovered by using a sterile 5 ml syringe and
60 loaded in three additions onto the DNA columns by centrifuging at 10 000x g for 30 sec. DNA
61 purification was performed following manufacturer's guidelines. The concentration of
62 double-stranded DNA was quantified by PicoGreen fluorescence assay (Life Technologies).
63 DNA quality was checked on an Agilent 2100 Bioanalyzer/Agilent Nano DNA chip (Agilent,
64 Santa Clara, CA, USA). Triplicate DNA extracts were pooled in equimolar amounts providing
65 1 pooled DNA extract per visit and station. The DNA extracts from the repeated visits (3 at M2
66 and 2 at each M3 and M4) were then pooled for each station to achieve 1 µg in 30 µL Tris for
67 sequencing purposes.

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

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

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

86 Metagenomic reads was evaluated using FastQC (v0.11.7) [6] and processed with Trimmomatic
87 (v0.39) [7]. An amount of 327.8, 336.3, and 285.0 million high-quality reads were retained for
88 station M2, M3 and M4, respectively (**Supplementary Table 2**). Sequencing coverage and
89 complexity of microbiomes in each metagenomic dataset was assessed using Nonpareil (v3.303)

90 [8] with the k-mer based approach (**Supplementary Fig. 2A**). Combined with another 337.3 M
91 high-quality reads from station M2 sampled in a previous cruise in early spring, de novo
92 co-assembly was performed using MEGAHIT (v1.0.4) [9] with default settings. CD-HIT-EST
93 (v4.7; -c 0.99 -aS 1 -g 1) [10] was used to measure the redundancy of the assembled contigs.
94 Metagenomic reads were aligned back to the contigs using Bowtie2 (v2.3.5) [11], with
95 duplicates removed by the markdup function of SAMtools (v1.9) [12]. On average, 30.32% of
96 reads in each metagenomic dataset could be recruited back to the assembled contigs.

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

98 We queried all the contigs against the NCBI nt database using BLASTN (v2.7.1) [13] with an
99 e-value threshold of 1e-10. The bitscore and percentage of identity of the best alignment for
100 each contig were collected. For those contigs which could not find a match in the nt database,
101 zeroes were assigned. The statistics of the two features are shown in (**Supplementary Fig.**
102 **2BC**).

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

109 *Metagenome Southern Ocean vs. TARA*

110 A total of 3 588 draft genomes from the TARA Ocean Global Expedition Project, including 957
111 from Delmont, et al. [14] and 2 631 from Tully, et al. [15], were downloaded. Besides, another
112 41 newly released Arctic metagenomic datasets with 11 709 809 contigs were also collected
113 [16]. We leveraged FastANI (v1.3) [17] for estimating pairwise ANI values between our 133
114 MAGs and the TARA draft genomes (**Supplementary Fig. 4A**). Read recruitment from our
115 Southern Ocean samples to the TARA assemblies, including both the draft genomes and the
116 un-binned Arctic contigs, were implemented using BMap (v38.22) [18] to verify the
117 differences observed through ANI. The percentage of reads mapped to each assembly dataset,
118 with a minimum percentage of identity at 90%, were summarized, and further normalized by
119 the size of each dataset to a “per Mbp” unit (**Supplementary Fig. 4BCE**). Sequence similarities
120 between our 949 228 non-redundant contigs and the TARA Arctic contigs were assessed using
121 BLASTN (v2.7.1) [13]. Matches were identified with a minimum percentage of identity at 90%,
122 a minimum coverage of the shorter sequence at 90% and an e-value threshold at 1e-3

123 (**Supplementary Fig. 4D**). Interestingly, in contrary to the low sequence similarity shared
124 among contigs, the TARA Arctic datasets showed higher abundance through read recruitment
125 than our MOBYDICK assemblies (adjusted P-value < 0.05 by Wilcoxon rank sum test;
126 **Supplementary Fig. 4E**). To further assess the similarity between our metagenomes and the
127 TARA Arctic ones, we downloaded the 41 TARA Arctic metagenomes and implemented
128 BMap (v38.22) [18] to map TARA Arctic reads to both the TARA Arctic contigs and our
129 MOBYDICK contigs. The read recruitment statistics are summarized and further normalized by
130 the size of each dataset to a “per Mbp” unit, as shown in **Supplementary Fig. 4F**. TARA Arctic
131 metagenomes and contigs are classified into two groups based on sampling depth. In addition to
132 the comparisons at the DNA level, we employed OrthoFinder (v2.2.3) [19] to identify
133 orthologous gene clusters among our Southern Ocean MAGs and TARA draft genomes
134 (**Supplementary Fig. 4GH**). Results and discussion could be found in **Supplementary**
135 **Results**.

136 *Reconstruction of prokaryotic 16S rRNA markers from metagenomes*

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

146 The identified archaeal and bacterial 16S rRNA V4-V5 representative sequences were
147 aligned by SSU-ALIGN (v0.1.1) [23] according to their secondary-structure covariance models.
148 We built a phylogenetic tree from the sequence alignment using FastTree (v2.1.10; -slow -nt
149 -gtr -gamma -bionj) [25], with the taxonomic assignment of the sequences as constraints at the
150 order level (-constraints). Metagenomic reads were mapped back to the 16S rRNA V4-V5
151 representative sequences using Bowtie2 (v2.3.5) [11]. A pseudo-OTU table was generated by
152 jgi_summarize_bam_contig_depths [26], taking the depth of coverage as an approximation to
153 amplicon counts. Visualization and statistics were performed in R (v3.6.1) using the phyloseq
154 package (v1.28.0) [27]: Closely-related taxa were clustered through phylogenetic
155 agglomeration by tip_glom (**Supplementary Fig. 5A**); The pseudo-OTU table was transformed
156 for equal sampling depth by rarefy_even_depth; Relative abundance was calculated with the aid

157 of transform_sample_counts (**Supplementary Fig. 5B**); Alpha diversity was assessed according
158 to the Shannon index and visualized via plot_richness (**Supplementary Fig. 5C**); To describe
159 the differences across sites, we applied the double principal coordinate analysis (DPCoA) with
160 regard to both phylogenetic structure and relative abundance, and visualized the results using
161 plot_ordination (**Supplementary Fig. 5D**). Environmental vectors (M2, M3 and M4) were
162 added according to the coordinates of the sampling sites reported in the DPCoA result. The first
163 principal component, which accounts for more than 99% of the total variance, is associated
164 positively with M2 and negatively with M3 and M4. Given that the three sampling sites are
165 representatives of the on- (M2) and off-plateau (M3 and M4) environments, the first principal
166 component could primarily measure the major differences that distinguish the contrasting
167 oceanic regions, such as the availability of iron and organic carbon.

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

176 We also estimated the root square of Rao's dissimilarity coefficient between samples by the
177 disc function in the ade4 package, using the OTU table and the phylogenetic distance
178 (generated by the cophenetic.phylo function in the ape package) involved in the DPCoA
179 analysis. This resulted a Rao's dissimilarity value of 0.1624 between M2 and M3, 0.1807
180 between M2 and M4, and 0.0259 between M3 and M4, in consistent with the correlations
181 between the environmental vectors as shown in **Supplementary Fig. 5D**. To evaluate the
182 statistical significance of categorical explanatory variables, such as the on- and off-plateau
183 environments, we performed another permutation test based on constrained DPCoA (cDPCoA)
184 [29]. Our cDPCoA analysis extended the DPCoA result with a partitioning factor of sampling
185 sites (on- and off-plateau), using the bca (between-class) function in the ade4 package. The
186 Monte-Carlo permutation test (1 000 replicates) on the cDPCoA result, using the
187 aforementioned randtest.dpcoa function, resulted in a p-value of 0.3446553 (> 0.05).
188 Considering that we only have three metagenomes, this insignificant outcome is not surprising
189 when the analysis focused on the effect of environmental types. We undertook simulations with
190 pseudo replicates for each sampling sites to examine how the number of samples biased the

191 significance test. We considered two cases, including 2 and 3 replicates per site. Assuming that
192 biological replicates from the same sampling site should consist of similar microbial
193 communities, the pseudo replicate(s) was generated by replacing the original measurement (x)
194 with a random number selected using the `rnorm` function with $\text{mean}=x$ and $\text{sd}=0.25x$. The same
195 cDPCoA and Monte-Carlo permutation test was performed on each simulated OTU table. A
196 total of 100 simulations were carried out in the case of 2 replicates per site, and the mean value
197 of the resulted p-values is 0.06286713 (± 0.01416026 standard deviation; $\text{mean} > 0.05$). When 3
198 replicates per site were considered, the p-value approximates $0.01042957 \pm 0.004679961$ (mean
199 < 0.05).

200 *Assembly-free metagenome taxonomic profiling*

201 One limitation of working with metagenomes is the considerable amount of reads that could not
202 be assembled into contigs (**Supplementary Table 2**). Such a limitation might be explained by
203 the prevalence of genomes with low genome abundance and the deficiency in de Bruijn
204 assembly algorithms [30]. Therefore, we also performed taxonomic profiling solely based on
205 unassembled metagenomic reads. We implemented three types of taxonomic classifiers with
206 distinct pros and cons [31], including DNA-to-DNA (Centrifuge v1.0.4 [32] with its
207 precompiled nt database and Kraken2 v2.0.7-beta [33] with its standard database),
208 DNA-to-protein (Kaiju v1.7.0 [34] with its precompiled nr databases), and
209 phylogenetic-marker-based (mOTUs2 v2.5.1 [35]) methods. The mOTUs2 intrinsically
210 incorporates the correction for copy number and genome length by using universal single-copy
211 phylogenetic marker genes [35] (**Supplementary Fig. 5E**). However, to derive accurate relative
212 abundance, the other three tools require subsequent normalization by the elusive total sequence
213 size of individual taxon in the reference databases. To avoid this, we divided out the effects of
214 reference sizes, as well as other multiplicative systematic biases, by considering the fold
215 changes between taxon ratios across sampling sites [36]. The fold change is defined as
216 $\frac{O_i(s)}{O_j(s)} / \frac{O_i(t)}{O_j(t)}$, where O represents the observed number of reads from sample s (or t) that were
217 assigned to individual taxonomic group i (or j) [36]. We used *Candidatus Pelagibacter* as the
218 denominator (taxon j) given its prevalence and high abundance across sampling sites. The taxon
219 ratios measured the “relative abundance” of a taxonomic group to *Candidatus Pelagibacter*,
220 instead of to the whole community. The fold change of taxon ratios, by crossing out potential
221 multiplicative systematic biases [36], was used as an indicator of the relative fitness of a taxon
222 as compared to the ubiquitous SAR11 population in contrasting environments (M3 vs. M2, or
223 M4 vs. M2; **Supplementary Fig. 5F**). Regarding possible detection biases against rare taxa,

224 here, we summarized read counts at the genus levels and only considered genera whose read
225 counts was no less than 1‰ of *Candidatus Pelagibacter*'s.

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

227 An initial round of read processing was provided by the company using Trimmomatic (v0.39)
228 [7]. The standard Illumina adapters and low-quality bases were removed with the following
229 parameters “2:30:10 SLIDINGWINDOW:4:5”. Further, we performed another round of quality
230 control and refinement with Trim Galore (v0.5.0) using a minimum Phred score of 20 and a
231 length threshold of 50 bp. To remove rRNA, tRNA and internal standard sequences, a two-step
232 clean-up procedure was performed, starting with SortMeRNA (v2.1b) [37] and followed by a
233 BLASTN (v2.7.1) [13] search against a custom database consisting of 545 336 reference rRNA
234 and tRNA sequences from diverse taxa along with the 2 internal standard sequences [38].
235 Finally, an average of 22.38 M high-quality paired-end protein-coding reads were retained from
236 each replicate, and the per-liter calculations were performed based on the recovery rate of
237 internal standards as previously described [38-40] (**Supplementary Table 2**).

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

239 In this study, we adopted a more general term CAZyme gene clusters (CGCs) [41] to identify
240 PUL-like gene clusters in a broader sense, requiring at least one CAZyme accompanied with
241 transcription factors (TFs) and/or transporters (TCs) (**Supplementary Fig. 13A-D**). To identify
242 fucose utilization loci among Verrucomicrobiae MAGs, the nine *Lentimonas* genomes provided
243 in Sichert et al. [42] were downloaded from NCBI GenBank database. Orthologous proteins
244 between the MAGs and the *Lentimonas* references were determined using OrthoFinder (v2.2.3)
245 [19] (**Supplementary Fig. 13E**).

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

247 The sequences of 19 Pseudomonadales HTCC2089 draft genomes were downloaded from
248 NCBI GenBank database, including 15 from UBA4421 genus and 4 from UBA9926
249 (**Supplementary Table 6**). Pangenome is defined on the basis of cluster of orthologous genes
250 identified by OrthoFinder (v2.2.3) [19], and comparative analysis between genus UBA4421 and
251 UBA9926 was carried out from three aspects, including the core, accessory and singleton genes
252 (**Supplementary Table 6**). A gene cluster enhancing the use of light as an energy resource was
253 identified exclusively in the UBA4421 genus (**Supplementary Fig. 14A**). A gene encoding
254 bacteriorhodopsin was also identified in one of the UBA9926 genome (GCA_002728935), but
255 genes around it are not related to light harvesting (**Supplementary Fig. 14A**). Genes encoding

256 bacteriorhodopsin were identified and their protein sequences were retrieved. The top 5
257 homologous bacteriorhodopsin sequences were selected from the MicRhoDE database [43] via
258 BLASTP search. Another two archaeal rhodopsin sequences were obtained from MicRhoDE
259 and used as outgroups for phylogeny analysis. All bacteriorhodopsin proteins were aligned by
260 MAFFT (v7.313) [44] and processed by trimAl (v1.4; -automated1) [45]. Phylogenetic
261 reconstruction was performed using IQ-Tree (v1.6.8; -m MFP -bb 1000 -bnni) [46]. The
262 bacteriorhodopsins from distinct genera are well separated on the phylogeny tree with high
263 support values (**Supplementary Fig. 14B**).

264 **Supplementary Results**

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

266 The TARA Ocean Global Expedition Project provides an extensive survey on the global surface
267 ocean microbiomes [47], providing 3 588 draft genomes [14, 15] and another 41 newly released
268 Arctic metagenomic datasets with 11 709 809 contigs [16]. However, to date, only 3 TARA
269 prokaryote-enriched metagenomes from the surface Southern Ocean are publicly available, and
270 their sampling sites were far away from ours. The aforementioned comparisons between our
271 Southern Ocean assemblies and the NCBI databases highlight the uniqueness of the Southern
272 Ocean microbiomes (**Supplementary Fig. 2B-G**) and further request the comparisons between
273 our Southern Ocean MAGs and the TARA assemblies. Our analyses were performed from three
274 main aspects, including the pairwise average nucleotide identity (ANI) assessment, the
275 Southern Ocean metagenomic read recruitment and the identification of orthologous protein
276 families between the two datasets (see **Supplementary Methods**). Only 13 of the Southern
277 Ocean MAGs conform to > 95% intra-species ANI values with TARA counterparts, despite the
278 high similarity shared among the TARA MAGs (**Supplementary Fig. 4A** and **Supplementary**
279 **Table 3**) [17]. Metagenomic read recruitment analysis further confirmed that, except for the
280 TARA Southern Ocean samples (SOC), the TARA MAGs recruited a significantly lower
281 amount of our metagenomic reads as compared to our Southern Ocean MAGs (adjusted P-value
282 < 0.05 by Wilcoxon rank sum test; **Supplementary Fig. 4BC**), assuring that the novelty of our
283 Southern Ocean assemblies is not derived from biases introduced during metagenome assembly
284 and binning. In addition, only 4.18% ($\pm 1.33\%$; S.D.) of our Southern Ocean contigs could find
285 a match with high similarity in the TARA Arctic datasets (**Supplementary Fig. 4D**).

286 Interestingly, in contrary to the low sequence similarity shared among contigs, the TARA Arctic
287 datasets showed higher abundance through read recruitment than our Southern Ocean
288 assemblies (adjusted P-value < 0.05 by Wilcoxon rank sum test; **Supplementary Fig. 4E**) and
289 the vice-versa read recruitment analysis confirmed the same pattern (**Supplementary Fig. 4F**),
290 suggesting that Arctic and Southern Ocean metagenomes may have some taxonomic and
291 functional similarity.

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

293 Particularly, *Polaribacter* presented an intense increase in abundance in the on-plateau M2 site
294 as compared to the HNLC waters (**Supplementary Fig. 5F**). *Polaribacter* are heterotrophic
295 bacteria belonging to the family Flavobacteriaceae. They are widely distributed in marine

296 habitats and rely heavily on the phytoplankton-derived dissolved organic matters [48-51]. It can
297 develop rapidly under favorable conditions, and was reported to contribute to a large share of
298 biomass production during the spring phytoplankton blooms surrounding Kerguelen Island [52].
299 Another interesting Flavobacteriaceae genus is *Formosa* (**Supplementary Fig. 5F**; named as
300 Hel1-33-131 under the GTDB taxonomy system [53] in **Supplementary Table 3**), which
301 accounted for up to 24% of all bacteria during diatom-dominated spring blooms off the
302 Helgoland Island in North Sea [54]. In general, marine Flavobacteriaceae are active degraders
303 of biopolymers and was observed to be abundant in nutrient-rich habitats [49, 55, 56]. They are
304 key players in Southern Ocean carbon cycling. Therefore, it is important to understand the
305 underlying enzymatic mechanisms and adaptation that drive the specialization of these highly
306 competitive bacteria in the on-plateau M2 site. To the contrary, the Gammaproteobacterial
307 genus *Psychrobacter* were of greater success in HNCL waters, characterized with a log fold
308 change of taxon ratios below 0 (**Supplementary Fig. 5F**). We have a Southern Ocean MAG
309 (MAG_130) from this genus, which showed an overall high abundance in all three sampling
310 sites and were featured with an average coverage of depth per million reads ~12 times higher in
311 M3 and M4 than M2 (**Fig. 1** and **Supplementary Table 3**). *Psychrobacter* is known to produce
312 siderophores in iron limiting conditions and able to increase growth rate under iron fertilization,
313 reassuring its success in both environments [57].

314 ***Genomic differences between MAG_103 and MAG_62***

315 We identified a putative methanogenesis gene cluster in MAG_62, which was conserved in all
316 but one UBA9926 genomes and absent from the UBA4421 genus (**Supplementary Table 6**). It
317 consisted of one trimethylamine methyltransferase (*mttB*; K14083), a corrinoid
318 methyltransferase, a protein of unknown function, a methyltetrahydrofolate-homocysteine
319 methyltransferase (MTR; K00548). and a ferredoxin. Methylophilic prokaryotes, which
320 consume methylated compounds such as trimethylamine and dimethyl sulfate, were reported to
321 be common in marine and hypersaline, sulfate-rich sediments [58]. Just adjacent to putative
322 methanogenesis loci, there was another cluster consisting of four sarcosine oxidase-encoding
323 genes (*soxADG*), which could only be found in one UBA4421 genomes. Another sarcosine
324 oxidase subunit, *soxB*, was also identified in all UBA9926 genomes and 2 UBA4421 gnomes
325 but absent from MAG_103. Sulfur oxidation mediated by Sox proteins have also been widely
326 identified as an important process of energy metabolism in sediments and deep-sea vent
327 ecosystems [59, 60]. The co-localization and conservation of the *sox* and *mtt* genes in MAG_62
328 suggested a deeper water origin of this species, which is coherent with the higher abundances of
329 MAG_62 in spring characterized by a deep mixed layer (**Supplementary Table 3**).

330 We could not detect fundamental advantages in substrate utilization that would explain
331 the flourishing of MAG_62 in the spring surface waters. Other possible explanations could be
332 related to RecBCD DNA repair enzymes and the biosynthesis of antibiotics. MAG_62 had a
333 gene cluster encoding the three exodeoxyribonuclease V gamma subunits in a row (*recB*, *recC*
334 and *recD*; K03582, K03583 and K03581), which was shared by all UBA9926 genomes but
335 absent from the UBA4421 genus. RecBCD complex is responsible for the repair of DNA
336 double-strand breaks by homologous recombination [61], and was considered to play an
337 effective role in handle antibiotic-induced oxidative DNA damage [62]. Additionally, RecD was
338 also reported to play a critical role for deep-sea bacteria to grow under high pressure [63].
339 Prokaryotes have evolved various strategies, such as motility, antibiotic synthesis and antibiotic
340 resistance, to increase their probability to survive and increase their competence for larger
341 shared of resources [64]. Both MAG_103 and MAG_62 were capable of producing multiple
342 antibiotics, but their approach to protect themselves from antibiotic-induced damages were not
343 the same (**Supplementary Table 6**). Penicillin-binding proteins (PBPs) are targets of β -lactam
344 antibiotics. Prokaryotes can confer antibiotic resistance by modifying their PBP structures to
345 either reduce the binding capability of antibiotics or directly degrade the antibiotics [65]. Both
346 the UBA9926 and UBA4421 genera could produce penicillin-binding protein 1A (*mrcA*;
347 K05366) and penicillin-binding protein 2 (*mrdA*; K05515) which were considered to participate
348 in beta-Lactam resistance. Besides, MAG_62 could produce a putative multi antimicrobial
349 extrusion protein (PF01554.18) which was reported to mediate resistance to multiple
350 antimicrobial agents [66]. As mentioned above, the RecBCD complex also plays are role in
351 antibiotic resistance, therefore, MAG_62 might be better equipped to deal with antibiotic stress.

352 **Supplementary Figure Legends**

353 **Supplementary Fig. 1** Three sampling sites (M2, M3 and M4) of the MOBYDICK cruise.
354 **A**, Station M2 is located above the central Kerguelen Plateau, where Chlorophyll *a* (Chl *a*)
355 concentrations are higher than at the off-plateau stations (M3 and M4). The colour code in (**A**)
356 represents the monthly (March) climatological mean of Chl *a* concentration (mg m^{-3}) from
357 2003-2018. Data was collected from NASA's moderate-resolution imaging spectroradiometer
358 (MODIS-Aqua) dataset (<https://oceancolor.gsfc.nasa.gov/data/aqua/>). The location of the
359 Antarctic Polar Front (APF) is also highlighted in (**A**), showing that during the MOBYDICK
360 cruise M2 and M4 were located south of the polar front in Antarctic waters whereas M3 was
361 situated in the polar front zone. The monthly APF data was collected from Pauthenet et al.,
362 (2018) [67]. The yearly mean latitude of APF at a resolution of 0.1 longitude degree (the black
363 curve in **A**) was determined by calculating the mean value of all 12-month measurements
364 within a 0.2 degree sliding window, which moves 0.1 longitude degree at one time. The
365 standard error of the mean (SEM) was also calculated and is shown as the grey shading around
366 the black curve in (**A**). The seasonal mean and SEM, covering austral summer and autumn,
367 were determined in the same way, except that only the data from December to May was
368 involved (the blue curve with light blue shading in **A**). The calculation scheme is exemplified in
369 the inset panel of (**A**), showing how the data at 60 °E were assessed: The blue dots represent
370 data for austral summer and autumn, while the orange ones are for austral winter and spring;
371 The mean and SEM values are calculated based on data points falling within the sliding widow
372 which is shaded in grey. **B**, The variation of Chl *a* concentration (mg m^{-3}) of the three sampling
373 sites (M2 in orange, M3 in blue and M4 in green) from October 2017 to March 2018. The
374 background shade in orange displays previous observations from 1998 to 2017, with Mean+S.D.
375 as the upper and Mean-S.D. as the lower limits of the shaded region. Data was collected from
376 Copernicus Marine Service (<http://marine.copernicus.eu/>). **C**, Depth profiles of temperature
377 (°C), salinity and dissolved organic carbon (DOC, μM) as determined during the MOBYDICK
378 cruise. The measurements from 0 to 200 m are illustrated. The colour scheme for data points
379 from different sampling sites and distinct visits is shown at the bottom.

380

381 **Supplementary Fig. 2** **A**, Sequencing coverage and complexity of microbial communities
382 in our samples determined by Nonpareil [8]. Nonpareil examines the redundancy of the reads in
383 metagenomic datasets to assess the average coverage and predict the amount of sequences
384 required to achieve full coverage [8]. Colours indicate different datasets used for the estimation,
385 and "R1/R2" represents the forward/reverse reads. Solid lines show the projection curves of the

386 estimated coverage per sequencing effort. The empty circles indicate the actual size of each
387 metagenomic dataset (x-axis) and the corresponding fraction of microbiome covered by DNA
388 sequencing (y-axis). The horizontal dashed lines in red indicate the 95% and 100% coverages.
389 The arrows at the bottom indicate the sequence diversity (Nonpareil N_d). Curves on the right
390 display higher sequence diversity than those on the left. **B** and **C**, The 949 228 assembled
391 contigs were queried against the NCBI nt database. The statistics of the bitscore (**B**) and
392 percentage of identity (**C**) of the best alignment for each contig are represented by bar plots. **D**
393 to **G**, The similarity between our 3 003 586 protein-coding genes and the NCBI nt (or nr)
394 database were also assessed. The corresponding statistics of the best hit for each gene are
395 summarized here. **D** and **F**, the x axis represents different ranges of the bitscores, and the y axis
396 shows the percentage of genes falling into each category. Similarly, **E** and **G**, the x axis
397 represents the percentage of identical match, and the y axis illustrates the percentage of genes in
398 each group.

399

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

409

410 **Supplementary Fig. 4** Comparisons between MOBYDICK Southern Ocean MAGs and
411 the TARA assemblies, including both draft genomes and un-binned contigs, suggesting high
412 genetic diversity at the DNA level but low functional novelty with regards to proteins. **A**,
413 Pairwise ANI values between our MAGs and the TARA draft genomes were calculated by using
414 fastANI [17]. If one of our MAGs has multiple matches in TARA, only the highest ANI value
415 was used for the statistics. MAGs without a TARA counterpart (ANI < 75%) were not
416 considered. The violin plot describes the distribution and the density trace of the 95 (out of 133)
417 ANI values. The 95 spotted dots are grouped and coloured according to three ranges: “red”
418 represents for > 95% intra-species ANI values; “salmon” are ANI values ranging between 83
419 and 95%; and “black” shows < 83% (but \geq 75%) inter-species ANI values [17]. (**B**, **C** and **E**)

420 shows the statistics of the read recruitment analyses against all the metagenome assemblies,
421 including 949 228 contigs and 133 MAGs from our study, as well as 11 709 809 contigs and 3
422 588 draft genomes from the TARA Ocean Project [14-16]. Data were summarized according to
423 their sampling locations as specified in the original papers [14-16], and the list of acronyms is
424 provided at the end of this paragraph. The boxplots illustrate the distribution of the fraction of
425 reads in each sample (M2, M3 or M4) that could be mapped to the assemblies by BMAP
426 (v38.22) [18] with a minimum percentage of identity at 90% and further normalized by the size
427 of each dataset to a “per Mbp” unit. Non-parametric pairwise Wilcoxon rank sum tests with
428 Holm adjustment demonstrated that the non-Southern Ocean TARA draft genomes recruit a
429 significantly lower amount of reads from each sample (M2, M3 or M4), as compared to our 133
430 MAGs ($p_{\text{adj}} < 0.05$; **B** and **C**). In contrary to **B** and **C**, the Arctic contigs showed an overall
431 higher rate of read recruitment than our assemblies ($p_{\text{adj}} < 0.05$; **E**). **D**, Our 949 228 contigs
432 were queried against the 41 TARA Arctic datasets using BLASTN (v2.7.1) [13]. The violin plot
433 describes the distribution and the density trace of the percentage of our contigs which have an
434 Arctic match with high sequence similarity. In **F**, Statistics of TARA Arctic metagenome read
435 recruitment to the TARA Arctic contigs and our MOBYDICK contigs. Similar to **E**, Arctic
436 contigs recruited significantly more reads per Mbp than ours. In **G** and **H**, Prodigal (v2.6.3) was
437 used to recover protein sequences from the TARA assemblies under meta mode (-p meta) [68].
438 Protein sequences were pooled according to their sampling locations as mentioned before,
439 resulting one combined site-specific proteome for our study (named as “SO MAGs”), 12 for
440 Delmont, et al. [14] and 10 for Tully, et al. [15]. Orthologous groups shared among these 23
441 site-specific proteomes were identified using OrthoFinder (v2.2.3) [19]. **G**, A total of 519 468
442 orthologous groups were identified, among which 84 192 were shared by all. Only 71
443 orthologous groups were unique to our MAGs (Singletons were not included). **H**, Barplots
444 illustrate the percentage of proteins in each site-specific proteome that are involved in
445 orthologous groups with at least one member from our MAGs. The TARA SOC MAGs have the
446 highest percentage of shared orthologs, whereas the TARA ASW ones have the lowest
447 percentage. Due to computational limitation, the TARA Arctic datasets were not included in the
448 orthology analysis. Acronyms: Delmont, et al. has 12 locations, including ANE (Atlantic
449 northeast), ANW (Atlantic northwest), ASE (Atlantic southeast), ASW (Atlantic southwest),
450 ION (Indian Ocean north), IOS (Indian Ocean south), MED (Mediterranean), PON (Pacific
451 Ocean north), PSE (Pacific Ocean southeast), PSW (Pacific Ocean southwest), RED (Red Sea)
452 and SOC (Southern Ocean) [14]. And, Tully, et al. consists of 10 sites, including ARS (Arabian
453 Sea), CPC (Chile-Peru coastal), EAC (east Africa coastal), IN (Indian Ocean), MED

454 (Mediterranean), NAT (north Atlantic), NP (north Pacific), RS (Red Sea), SAT (south Atlantic)
455 and SP (south Pacific) [15].

456

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

486

487 **Supplementary Fig. 6** Taxonomic composition and abundance of KEGG Orthology
488 groups (KOs) in the SO metagenomes (**A to D**) and metatranscriptomes (**E to J**). In **A to D**, the
489 relative abundance of a taxonomic category i given a KO_j was measured: **A**, by the proportion
490 of genes, among all the protein-coding genes belonging to KO_j , assigned to taxonomy i , as
491 defined by the division of $\frac{f_{counts}(gene \in Tax_i \cap KO_j)}{f_{counts}(gene \in KO_j)}$; **B to D**, based on gene abundance, in the
492 form of FPKM (fragments per kilobase of transcript per million mapped reads) values, as
493 defined by the division of $\frac{\sum_{gene \in Tax_i \cap KO_j} f_{FPKM}(gene)}{\sum_{gene \in KO_j} f_{FPKM}(gene)}$ in each metagenome. In **E to G**, the
494 absolute abundance of each KO group in each metagenome was estimated using the normalized
495 transcripts per liter according to the internal standards protocol [40], as defined by the sum of
496 $\sum_{gene \in KO_j} f_{Transcripts L^{-1}}(gene)$. The values of the total transcripts L^{-1} assigned to each KO
497 are shown in the middle of each cell. The colour code, from blue via yellow to red, represents
498 the total number of transcripts L^{-1} assigned to each KO group from low to high. In **H to J**, the
499 relative abundance of a taxonomic category i given a KO_j was evaluated by the division of
500 $\frac{\sum_{gene \in Tax_i \cap KO_j} f_{Transcripts L^{-1}}(gene)}{\sum_{gene \in KO_j} f_{Transcripts L^{-1}}(gene)}$ in each metatranscriptome. The mean of the three replicates
501 was used to represent their corresponding metatranscriptomic sample. The KO groups in all
502 panels are ordered decreasingly according to the value of transcripts L^{-1} in M2 (as shown in **E**).
503 Only functional groups consisting of at least 10 genes, out of 3 003 587 protein-coding genes
504 predicted from the metagenome assemblies, were shown.

505
506 **Supplementary Fig. 7** Abundance of functional groups in each sample and the relative
507 contribution of different taxonomic categories to each functional group were evaluated based on
508 the carbohydrate-active enzymes (CAZyme) [69]. The calculation and the order of panels
509 within each section are the same as shown in **Supplementary Fig. 6**.

510
511 **Supplementary Fig. 8** Statistics of functional diversity (**A**) and shifts in taxonomic
512 composition within functional groups (**B**) across sampling sites, based on the community-level
513 metagenomic gene abundance represented by FPKM (fragments per kilobase of transcript per
514 million mapped reads) values (**B-D** in **Supplementary Fig. 6-7**). In **B**, the relative contribution
515 (%) of a specific taxonomic category (e.g., Gammaproteobacteria) to a functional group (e.g.,
516 ferrous iron transporter FeoA) in each sampling station was calculated (Materials and methods).
517 Shifts in the relative contribution across stations were estimated using the ratio of the relative
518 contribution in M2 to that in M3 (or M4) and visualized by violin plots. A ratio value less than 1

519 indicates that the taxonomic category accounts for a larger share of the genes of a functional
520 group in the off-plateau HNLC waters, and vice versa. Multiple databases were considered,
521 including CAZy, FeGenie, KEGG, Pfam, Superfamily and TCDB. Five dominant taxonomic
522 groups in gene pool and transcript inventories across all sampling sites were shown
523 (**Supplementary Fig. 6-7**). Colour code is the same as **Supplementary Fig. 6-7**. Only
524 functional groups consisting of at least 50 genes, out of the 3 003 587 protein-coding genes
525 predicted from the metagenome assemblies, were used in the calculation.

526

527 **Supplementary Fig. 9** A, An exemplary illustration of the influence of genome abundance
528 on the assessment of gene expression levels, showing that fluctuations in transcript abundance
529 could be a result of shifting genome copies rather than changes in expression levels. B, The
530 pipeline to generate the metagenome-normalized metatranscriptomic pseudo-read count table.
531 All the analyses were performed under R environment (v3.6.1). Initially, two matrices of read
532 counts of genes were generated using featureCounts (v2.0.0) [70]. Each matrix contains 3 003
533 586 rows, representing the total number of protein-coding genes predicted in our 949,228
534 assembled contigs (≥ 1 kbp). The metagenomic matrix consists of 3 columns representing 3
535 sampling sites M2, M3 and M4. The metatranscriptomic matrix has 9 columns corresponding to
536 3 biological replicates \times 3 sampling sites. Secondly, genes with low read counts were filtered
537 out. Only genes whose FPM (fragments per million mapped fragments) value is no less than 1
538 in at least 1 sample, which must be satisfied in both metagenomic and metatranscriptomic data,
539 were remained. Variance stabilizing transformation (VST) was carried out on both matrices.
540 Given that $Abundance_{transcript} \sim Abundance_{gene} \times Expression\ Level\ (EL_{gene})$, we used
541 the ratio between transcript abundance and gene abundance as approximations of gene
542 expression levels ($EL_{gene} \sim \frac{Abundance_{transcript}}{Abundance_{gene}}$). Here, $Abundance_{transcripts}$ represents the
543 metatranscriptomic read counts of a gene in a sample, and $Abundance_{gene}$ equals to the
544 metagenomic read counts of a gene in a sample. Since the transformed data was on \log_2 scale,
545 the formula could also be written as $EL_{gene} \sim 2^{VST(Abundance_{transcript}) - VST(Abundance_{gene})}$. Then,
546 in each sample (each column), EL_{gene} values are normalized (divided) by the corresponding
547 sample (column) maximum. Finally, we converted the normalized ratios to pseudo counts by
548 multiplying with 10^6 and rounding to the nearest integer. C, Heatmap of the sample-to-sample
549 distance, and D, Principal component analysis (PCA) based on the normalized read counts after
550 variance stabilizing transformation by DESeq2.

551

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

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

577
578 **Supplementary Fig. 12** The distribution of significantly differentially expressed genes in
579 the 133 MAGs among diverse functional categories related to iron uptake and carbon
580 metabolism. From left to right, the panels represent the phylogenetic tree (the same as shown in
581 Fig. 1), the KOs involved in the EMP glycolysis/gluconeogenesis, C1 and methylated
582 compound oxidation pathways, oxidative phosphorylation, nitrogen metabolism, transcription
583 factors, ABC transporters, sulfur metabolism, Sec (secretion) system, type II secretion system,
584 type IV secretion system and the two-component system. The symbol and colour schemes are
585 the same as shown in **Fig. 5**.

586

587 **Supplementary Fig. 13** Polysaccharide utilization loci (PULs) identified in MAG_78 (A),
588 134 (B), 3 (C) and 73 (D), as well as fucose utilization loci found in MAG_58, 98, 6 and 17 (E).
589 Arrows represent genes and direction of arrows illustrates their strand locations. The colour
590 scheme is: “orange” represents genes that are significantly higher expressed in the on-plateau
591 iron-fertilized water (M2 site); “yellow” shows genes that are higher, but not significantly,
592 expressed in the on-plateau site; “blue” indicates genes that are significantly higher expressed
593 in the off-plateau HNLC waters (M3 and M4 sites); “cyan” are genes that are higher, but not
594 significantly, expressed in the HNLC waters. Genes encoding the fucose metabolic pathway of
595 *Lentimonas* sp. CC4 [42] is used as the reference to arrange their counterparts identified in
596 MAGs from the *Verrucomicrobiae* class (under the GTDB taxonomy system [53]). Light blue
597 shades between genes are used to indicate orthologous groups and facilitate visualization,
598 because the incompleteness of MAGs leads to the fragmented distribution of genes among
599 contigs.

600

601 **Supplementary Fig. S14** (A) A gene cluster potentially engaged in using light as a
602 complementary energy resource was identified exclusively in the UBA4421 genus. A gene
603 encoding bacteriorhodopsin was also detected in one of the UBA9926 genome
604 (GCA_002728935), but genes adjacent to it do not form similar light-harvesting gene clusters
605 as found in the UBA4421 genus. Light blue shades between genes are used to indicate
606 orthologous genes. The colour scheme is: “orange” represents genes that are significantly
607 higher expressed in the on-plateau iron-fertilized water (M2 site); “yellow” shows genes that
608 are higher, but not significantly, expressed in the on-plateau site; “white” are genes from
609 reference genomes. (B) Bacteriorhodopsins from different genera are well separated on the
610 phylogeny tree with high support values, indicating distinct origins of the rhodopsin genes. The
611 emergence of the accessory functions of light harvesting might be a result of adaptive evolution.
612 (C) Sequence alignment of transmembrane peptide segment of bacteriorhodopsin sequence of
613 MAG_103 and its top 20 matches in the MicRhoDE database [43], highlighting variation in
614 residues involved in spectral tuning and ion pumping.

615 **Supplementary Data**

616 **Sequences of the two internal standards add to RNA libraries:**

617 >MTST51006NT
618 GGGUUCGGUGGUCUAUACUACUACCUAAGUUGGAUGUACUGGUGGAAGUGCUACCAACACAGUAAUGCUGGUUAGGUAG
619 GCAACACUACGACUUCAGGAAGAGUCUAAACGAAUGUAUGCAUAAUACUAAUGCCUUAACAUGUGGAAGCACCCUUAACGG
620 ACAGGAUGAGGCACAGGCACAUGUGCAGGCAAGCUUUAAGUGGAUGUGCGGUGAAAAUAAGUUCUGGCUAGUAAGGGCU
621 AUGGAAAAUCAACCUAGCAGAAAGGAUACUAGCUCAAAUGACGAUAAACGGACAGUGACUGGCAACCUGAAGAAUGGUACAA
622 GAAGAGGCACGACCCUGGUGAAAAUGACGUAAUAAGGUGCGUAUACAUAGGUUAAGAAAAUCUAGUAAAUGCUACGUGCC
623 CUGACAUGGAAGACUACUACGCUAUGACGGUAAUAAGCCUCUACUAGAACUAAAUAGUAUAGGUCCUUGCACGCAAUGC
624 ACGGUACACAAGCUAGAAGGUGUACACUGCAU AUGGUGGAUAGUAAGGAGGGACCACUCCUGUACCUAAUACAAA
625 AGUAGACGUAAUCAAUCUUAACAUAUUCGCUAGUGGUACGGUACU AUGCAUACAACACGUCUCACCCUUGGGGUGACU
626 GGAUGUUCGACGUACAAUACGAAAGUUGCAGGAUGUACAGGUGGUGGAUGACGAGGAAUGACUGGAGUGGUCAAUAAG
627 UGGAGUGGUCACAGUAU AUGCCAACCUACUGCUGCAGGAGUGACGACAGUAGGGUAAGUAAGAGGAUGGCUACGGU
628 AACGAAGGAAGUAGUAGAAAUGAGUCACAUGGACCUAAAAGAGGAGUGCUUACUGCAAUAGGACGCAACUAGAAGAAUACG
629 ACGCUUUCUACACGAGGUGAAGUUCGUACCUUGGAUGUACCCUGCUCUACCUUGCGAAGUACAAGACUUCGUAACG
630 AGGAGGACGUUCGACUACCCUGACCCUACGGCUUGCGGUCUAGUUU
631 >MTST61006NT
632 GGGAAUCUAAUAACGAACCUAUAUGGUACGU AAGUAUGAGUAGUGGUUUCGGUGUAAAUUCAGGGACCAACAAGGUGC
633 UAUAGACGUAAAGGUAGCUUGCCUACAAAUAGACCCU AAGGU AUGCCACCAAAGGGUAGUAAUAGUUACUGGUGAAUU
634 GGAGUUGGCUAGACAUGAAUACGUUCUACCCUGCUUACUUCUACGUAAUACAAAGUUGGGAAAGGUGGACGGACUGGAAG
635 UACAAUUCUACCCUGGUGCUGUAAUGCCUAGGGGUGAACCUAAGUGGACGCCUUGCAGGCAAUGCAUGCCUAGUCACAU
636 AAAUGGUAGGACGAAUAGGAUGCAAAGGCUCUAGGUUUCCAAAGGAGGGUUAACUGCUACCCUUAUAGAAGCUGAAA
637 UGAGGGAAUGCAUGGAAAGUGACCCUACCAAGACAAUGUAAAGUGGAUAAUUAUCCUAAAGGUAAAUGUAAUUAUUCU
638 AUGUGGGACAUAUACGUAAAGGGCUGAAUGCAAUCUACGCAACCUUUAUGUUCUGGAUGAAGCAAGGUUUCAGGUUCCU
639 AAGUUUCACGCACGAAUCAAUCACGACACGCUACACUGGUACCUACACCAACUAGGUAGUGCUCUAAUGUAGCUGGUG
640 GUUGCAGGACGAAGAGUUGCCAAACGGACUACAAUUCAAUCACCCUGACGUAAAUAUACCUAAUCAAUGCGACUUCUGC
641 CCUUAUGCACUGGCUAUGCAAUGAAGCUAUAAGAAGGACAAUAGUCUACGAUAAAUAUAAAGGUGCGAACAAGAAUG
642 CAAUUGGGUGGUGACGACCAGUAGACGCUUGCAUACCUAAAUGUGCUACAGGAUGGCUAUAAGUAAAACGAGGA
643 GGAGUCACCUAUAACGAAUACAGUGUAUGCAAUAGGGAAAGGUACACGCUAGAAAUGGUACCUACGGCUAACCUCACAU
644 AGGAUAGCUCACAUAAAUUGGGUAUUCGACUUCAGGGUACACGUUU

645 **Supplementary Code Availability**

646 All software used in this study are listed below:

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

671 **Supplementary References**

- 672 1. Tremblay L, Caparros J, Leblanc K, Obernosterer I. Origin and fate of particulate and dissolved
673 organic matter in a naturally iron-fertilized region of the Southern Ocean. *Biogeosciences*.
674 2015;12(2):607-21.
- 675 2. Blain S, Sarthou G, Laan P. Distribution of dissolved iron during the natural iron-fertilization
676 experiment KEOPS (Kerguelen Plateau, Southern Ocean). *Deep Sea Res Part II Top Stud Oceanogr*.
677 2008;55(5-7):594-605.
- 678 3. Blain S, Queguiner B, Armand L, Belviso S, Bombléd B, Bopp L, et al. Effect of natural iron
679 fertilization on carbon sequestration in the Southern Ocean. *Nature*. 2007;446(7139):1070-4.
- 680 4. Moran MA, Satinsky B, Gifford SM, Luo H, Rivers A, Chan LK, et al. Sizing up metatranscriptomics.
681 *ISME J*. 2013;7(2):237-43.
- 682 5. Satinsky BM, Gifford SM, Crump BC, Moran MA. Use of internal standards for quantitative
683 metatranscriptome and metagenome analysis. *Methods Enzymol*. 2013;531:237-50.
- 684 6. Andrew S. FastQC: a quality control tool for high throughput sequence data. Available online at:
685 <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>. 2010.
- 686 7. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data.
687 *Bioinformatics*. 2014;30(15):2114-20.
- 688 8. Rodriguez RL, Gunturu S, Tiedje JM, Cole JR, Konstantinidis KT. Nonpareil 3: fast estimation of
689 metagenomic coverage and sequence diversity. *mSystems*. 2018;3(3).
- 690 9. Li D, Luo R, Liu CM, Leung CM, Ting HF, Sadakane K, et al. MEGAHIT v1.0: A fast and scalable
691 metagenome assembler driven by advanced methodologies and community practices. *Methods*.
692 2016;102:3-11.
- 693 10. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or
694 nucleotide sequences. *Bioinformatics*. 2006;22(13):1658-9.
- 695 11. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*.
696 2012;9(4):357-9.
- 697 12. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map
698 format and SAMtools. *Bioinformatics*. 2009;25(16):2078-9.
- 699 13. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*.
700 1990;215(3):403-10.
- 701 14. Delmont TO, Quince C, Shaiber A, Esen OC, Lee ST, Rappe MS, et al. Nitrogen-fixing populations
702 of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes. *Nat Microbiol*.
703 2018;3(7):804-13.
- 704 15. Tully BJ, Graham ED, Heidelberg JF. The reconstruction of 2,631 draft metagenome-assembled
705 genomes from the global oceans. *Sci Data*. 2018;5:170203.
- 706 16. Ibarbalz FM, Henry N, Brandao MC, Martini S, Busseni G, Byrne H, et al. Global trends in marine
707 plankton diversity across kingdoms of life. *Cell*. 2019;179(5):1084-97 e21.
- 708 17. Jain C, Rodriguez RL, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of
709 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun*. 2018;9(1):5114.

- 710 18. Bendall ML, Stevens SL, Chan LK, Malfatti S, Schwientek P, Tremblay J, et al. Genome-wide
711 selective sweeps and gene-specific sweeps in natural bacterial populations. *ISME J.*
712 2016;10(7):1589-601.
- 713 19. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics.
714 *Genome Biol.* 2019;20(1):238.
- 715 20. Pericard P, Dufresne Y, Couderc L, Blanquart S, Touzet H. MATAM: reconstruction of phylogenetic
716 marker genes from short sequencing reads in metagenomes. *Bioinformatics.* 2018;34(4):585-91.
- 717 21. Gruber-Vodicka HR, Seah BKB, Pruesse E. phyloFlash: rapid small-subunit rRNA profiling and
718 targeted assembly from metagenomes. *mSystems.* 2020;5(5).
- 719 22. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene
720 database project: improved data processing and web-based tools. *Nucleic Acids Res.*
721 2013;41(Database issue):D590-6.
- 722 23. Nawrocki E. Structural RNA homology search and alignment using covariance models 2009.
- 723 24. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: High-resolution
724 sample inference from Illumina amplicon data. *Nat Methods.* 2016;13(7):581-3.
- 725 25. Price MN, Dehal PS, Arkin AP. FastTree 2--approximately maximum-likelihood trees for large
726 alignments. *PLoS One.* 2010;5(3):e9490.
- 727 26. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, et al. MetaBAT 2: an adaptive binning algorithm
728 for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ.* 2019;7:e7359.
- 729 27. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics
730 of microbiome census data. *PLoS One.* 2013;8(4):e61217.
- 731 28. Thioulouse J, Dray S, Dufour A-B, Siberchicot A, Jombart T, Pavoine S. Analysing patterns of
732 biodiversity. *Multivariate Analysis of Ecological Data with ade4.* New York, NY: Springer New
733 York; 2018. p. 281-94.
- 734 29. Dray S, Pavoine S, Aguirre de Carcer D. Considering external information to improve the
735 phylogenetic comparison of microbial communities: a new approach based on constrained Double
736 Principal Coordinates Analysis (cDPCoA). *Mol Ecol Resour.* 2015;15(2):242-9.
- 737 30. Steinegger M, Mirdita M, Soding J. Protein-level assembly increases protein sequence recovery from
738 metagenomic samples manyfold. *Nat Methods.* 2019;16(7):603-6.
- 739 31. Ye SH, Siddle KJ, Park DJ, Sabeti PC. Benchmarking metagenomics tools for taxonomic
740 classification. *Cell.* 2019;178(4):779-94.
- 741 32. Kim D, Song L, Breitwieser FP, Salzberg SL. Centrifuge: rapid and sensitive classification of
742 metagenomic sequences. *Genome Res.* 2016;26(12):1721-9.
- 743 33. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol.*
744 2019;20(1):257.
- 745 34. Menzel P, Ng KL, Krogh A. Fast and sensitive taxonomic classification for metagenomics with Kaiju.
746 *Nat Commun.* 2016;7:11257.
- 747 35. Milanese A, Mende DR, Paoli L, Salazar G, Ruscheweyh HJ, Cuenca M, et al. Microbial abundance,
748 activity and population genomic profiling with mOTUs2. *Nat Commun.* 2019;10(1):1014.
- 749 36. McLaren MR, Willis AD, Callahan BJ. Consistent and correctable bias in metagenomic sequencing
750 experiments. *Elife.* 2019;8.

- 751 37.Kopylova E, Noe L, Touzet H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in
752 metatranscriptomic data. *Bioinformatics*. 2012;28(24):3211-7.
- 753 38.Durham BP, Dearth SP, Sharma S, Amin SA, Smith CB, Campagna SR, et al. Recognition cascade
754 and metabolite transfer in a marine bacteria-phytoplankton model system. *Environ Microbiol*.
755 2017;19(9):3500-13.
- 756 39.Satinsky BM, Crump BC, Smith CB, Sharma S, Zielinski BL, Doherty M, et al. Microspatial gene
757 expression patterns in the Amazon River Plume. *Proc Natl Acad Sci U S A*. 2014;111(30):11085-90.
- 758 40.Satinsky BM, Gifford SM, Crump BC, Moran MA. Use of internal standards for quantitative
759 metatranscriptome and metagenome analysis. *Methods in enzymology*. 531: Elsevier; 2013. p.
760 237-50.
- 761 41.Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, et al. dbCAN2: a meta server for automated
762 carbohydrate-active enzyme annotation. *Nucleic Acids Res*. 2018;46(W1):W95-W101.
- 763 42.Sichert A, Corzett CH, Schechter MS, Unfried F, Markert S, Becher D, et al. Verrucomicrobia use
764 hundreds of enzymes to digest the algal polysaccharide fucoidan. *Nat Microbiol*. 2020:1-14.
- 765 43.Boeuf D, Audic S, Brillet-Gueguen L, Caron C, Jeanthon C. MicRhoDE: a curated database for the
766 analysis of microbial rhodopsin diversity and evolution. *Database*. 2015;2015.
- 767 44.Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in
768 performance and usability. *Mol Biol Evol*. 2013;30(4):772-80.
- 769 45.Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. trimAl: a tool for automated alignment
770 trimming in large-scale phylogenetic analyses. *Bioinformatics*. 2009;25(15):1972-3.
- 771 46.Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic
772 algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 2015;32(1):268-74.
- 773 47.Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, et al. Ocean plankton.
774 Structure and function of the global ocean microbiome. *Science*. 2015;348(6237):1261359.
- 775 48.Fernandez-Gomez B, Richter M, Schuler M, Pinhassi J, Acinas SG, Gonzalez JM, et al. Ecology of
776 marine Bacteroidetes: a comparative genomics approach. *ISME J*. 2013;7(5):1026-37.
- 777 49.Williams TJ, Wilkins D, Long E, Evans F, DeMaere MZ, Raftery MJ, et al. The role of planktonic
778 Flavobacteria in processing algal organic matter in coastal East Antarctica revealed using
779 metagenomics and metaproteomics. *Environ Microbiol*. 2013;15(5):1302-17.
- 780 50.Luria CM, Amaral-Zettler LA, Ducklow HW, Rich JJ. Seasonal succession of free-living bacterial
781 communities in coastal waters of the Western Antarctic Peninsula. *Front Microbiol*. 2016;7:1731.
- 782 51.Dadaglio L, Dinasquet J, Obernosterer I, Joux F. Differential responses of bacteria to diatom-derived
783 dissolved organic matter in the Arctic Ocean. *Aquat Microb Ecol*. 2018;82(1):59-72.
- 784 52.Obernosterer I, Catala P, Lebaron P, West NJ. Distinct bacterial groups contribute to carbon cycling
785 during a naturally iron fertilized phytoplankton bloom in the Southern Ocean. *Limnol Oceanogr*.
786 2011;56(6):2391-401.
- 787 53.Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil PA, et al. A standardized
788 bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol*.
789 2018;36(10):996-1004.
- 790 54.Teeling H, Fuchs BM, Bennke CM, Kruger K, Chafee M, Kappelmann L, et al. Recurring patterns in
791 bacterioplankton dynamics during coastal spring algae blooms. *Elife*. 2016;5:e11888.

792 55. Mann AJ, Hahnke RL, Huang S, Werner J, Xing P, Barbeyron T, et al. The genome of the
793 alga-associated marine flavobacterium *Formosa agariphila* KMM 3901T reveals a broad potential for
794 degradation of algal polysaccharides. *Appl Environ Microbiol.* 2013;79(21):6813-22.
795 56. Xing P, Hahnke RL, Unfried F, Markert S, Huang S, Barbeyron T, et al. Niches of two
796 polysaccharide-degrading *Polaribacter* isolates from the North Sea during a spring diatom bloom.
797 *ISME J.* 2015;9(6):1410-22.
798 57. Sinha AK, Parli Venkateswaran B, Tripathy SC, Sarkar A, Prabhakaran S. Effects of growth
799 conditions on siderophore producing bacteria and siderophore production from Indian Ocean sector
800 of Southern Ocean. *J Basic Microbiol.* 2019;59(4):412-24.
801 58. Lyu Z, Shao N, Akinyemi T, Whitman WB. Methanogenesis. *Curr Biol.* 2018;28(13):R727-R32.
802 59. Meyer B, Imhoff JF, Kuever J. Molecular analysis of the distribution and phylogeny of the *soxB* gene
803 among sulfur-oxidizing bacteria - evolution of the Sox sulfur oxidation enzyme system. *Environ*
804 *Microbiol.* 2007;9(12):2957-77.
805 60. Jørgensen BB, Nelson DC, Amend JP, Edwards KJ, Lyons TW. Sulfide oxidation in marine sediments:
806 Geochemistry meets microbiology. *Sulfur Biogeochemistry - Past and Present.* 379: Geological
807 Society of America; 2004. p. 0.
808 61. Cromie GA. Phylogenetic ubiquity and shuffling of the bacterial RecBCD and AddAB recombination
809 complexes. *J Bacteriol.* 2009;191(16):5076-84.
810 62. Stohl EA, Seifert HS. *Neisseria gonorrhoeae* DNA recombination and repair enzymes protect against
811 oxidative damage caused by hydrogen peroxide. *J Bacteriol.* 2006;188(21):7645-51.
812 63. Bidle KA, Bartlett DH. RecD function is required for high-pressure growth of a deep-sea bacterium.
813 *J Bacteriol.* 1999;181(8):2330-7.
814 64. Hibbing ME, Fuqua C, Parsek MR, Peterson SB. Bacterial competition: surviving and thriving in the
815 microbial jungle. *Nat Rev Microbiol.* 2010;8(1):15-25.
816 65. Sun S, Selmer M, Andersson DI. Resistance to beta-lactam antibiotics conferred by point mutations
817 in penicillin-binding proteins PBP3, PBP4 and PBP6 in *Salmonella enterica*. *PLoS One.*
818 2014;9(5):e97202.
819 66. Morita Y, Kodama K, Shiota S, Mine T, Kataoka A, Mizushima T, et al. NorM, a putative multidrug
820 efflux protein, of *Vibrio parahaemolyticus* and its homolog in *Escherichia coli*. *Antimicrob Agents*
821 *Chemother.* 1998;42(7):1778-82.
822 67. Pauthenet É, Roquet F, Madec G, Guinet C, Hindell M, McMahon C, et al. Seasonal meandering of
823 the Polar Front upstream of the Kerguelen Plateau. *Geophys Res Lett.* 2018;45(18):9774-81.
824 68. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene
825 recognition and translation initiation site identification. *BMC Bioinformatics.* 2010;11:119.
826 69. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The
827 Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic*
828 *Acids Res.* 2009;37(Database issue):D233-8.
829 70. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning
830 sequence reads to genomic features. *Bioinformatics.* 2014;30(7):923-30.
831 71. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.*
832 2000;28(1):27-30.

