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

Ying Sun, Pavla Debeljak, Ingrid Obernosterer

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

Running title: Microbial iron and carbon strategies in the Southern Ocean

Ying Sun[1*†], Pavla Debeljak[1,2*], Ingrid Obernosterer[1]

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.

† Corresponding author: Ying Sun, Microbial Oceanography Laboratory (LOMIC), CNRS-Sorbonne University, 1 Avenue Pierre Fabre, 66650 Banyuls sur mer, France. Tel: (+33) 04 68 88 73 53. Email: ying.sun@obs-banyuls.fr.

Abstract

Marine microbes are major drivers of all elemental cycles. The processing of organic carbon by heterotrophic prokaryotes is tightly coupled to the availability of the trace element iron in large regions of the Southern Ocean. However, the functional diversity in iron and carbon metabolism within diverse communities remains a major unresolved issue. Using novel Southern Ocean meta-omics resources including 133 metagenome-assembled genomes (MAGs), we show a mosaic of taxonomy-specific ecological strategies in naturally iron-fertilized and high nutrient low chlorophyll (HNLC) waters. Taxonomic profiling revealed apparent community shifts across contrasting nutrient regimes. Community-level and genome-resolved metatranscriptomics evidenced a moderate association between taxonomic affiliations and iron and carbon-related functional roles. Diverse ecological strategies emerged when considering the central metabolic pathways of individual MAGs. Closely related lineages appear to adapt to distinct ecological niches, based on their distribution and gene regulation patterns. Our in-depth observations emphasize the complex interplay between the genetic repertoire of individual taxa and their environment and how this shapes prokaryotic responses to iron and organic carbon availability in the Southern Ocean.

Introduction

Heterotrophic prokaryotes play a key role in the cycling of elements in the ocean [1-3]. Efforts to decipher prokaryotes-environment interactions, facilitated by meta-omics technologies [4, 5], have revealed highly complex dynamics in marine microbial community structure and function [6-18]. Recent studies from global ocean expeditions provided insights into microbial diversity coupled with significant changes in gene repertoire and expression along temperature gradients and depths [13, 14, 18]. However, there are still fundamental gaps in understanding how individual prokaryotic taxa regulate their metabolic capacities in response to changes of other environmental factors, such as nutrient availability, and how their adaptation strategies may influence energy flows and nutrient cycling in their ecological niches.

The Southern Ocean remains one of the least explored ocean regions. The perennially cold waters present the largest high-nutrient, low-chlorophyll (HNLC) region of the global ocean, where iron is the primary limiting factor of biological productivity [19]. Heterotrophic prokaryotes experience a double constraint due to low concentrations of bioavailable iron and dissolved organic carbon (DOC) [20, 21]. The availability of these nutrients affects prokaryotic heterotrophic activities, particularly growth and respiration [22, 23], and as a consequence their functions in microbial food webs. Further, iron is present in various chemical forms in the ocean [24] and a multitude of substrates constitute the pool of organic matter [25], challenging the exploration of the functional roles of diverse prokaryotic taxa in accessing these essential resources.

Southern Ocean islands are a source of iron to the surrounding seawater, leading to localized spring phytoplankton blooms [26]. Kerguelen Island, located in the Indian Sector of the Southern Ocean, sustains the largest iron-fertilized region [27-30]. The annually occurring

diatom-dominated phytoplankton blooms east of the island have important consequences on microbial communities. Field studies and onboard experiments identified diatom-derived DOC as a key driver for prokaryotic diversity, activity, and seasonal community succession [21, 31-34]. The naturally-iron fertilized region off Kerguelen Island within HNLC waters provides a natural laboratory to examine phylogenetic and functional diversity of microbial lineages that mediate iron and carbon cycling.

In this study, we provide a comprehensive survey of the structure, genetic repertoire and expression pattern of the free-living (< 0.8 µm size fraction) prokaryotic community in contrasting Southern Ocean productivity regions. We sampled three stations during the Marine Ecosystem Biodiversity and Dynamics of Carbon around Kerguelen (MOBYDICK) cruise in late austral summer (18th February to 30th March 2018), including one located in the naturally iron-fertilized waters and two off-plateau ones within HNLC waters. Metagenomic assembly and curation recovered a novel Southern Ocean meta-omics resource with 3 million protein-coding genes and characterized 133 metagenome-assembled genomes (MAGs) complementary to existing oceanic databases. Our main objective was to explore the distribution of prokaryotic functions related to iron and carbon metabolism in contrasting nutrient regimes and their links to taxonomy. We addressed this objective on the community and taxon-specific level by considering both the functional potential and the gene expression patterns.

Materials and methods

Sample collection, metagenome and metatranscriptome sequencing

Surface seawater (10 m) was collected at three stations in contrasting oceanic regions during the MOBYDICK cruise (**Supplementary Fig. 1A**). Station M2 was located in the naturally iron-fertilized waters above the central Kerguelen Plateau and stations M3 and M4 were located in off-plateau HNLC waters. The timing of the cruise covered the demise of the summer phytoplankton blooms (**Supplementary Fig. 1B**), as reflected in enhanced concentrations of dissolved organic carbon, prokaryotic abundance and heterotrophic production in on-plateau surface waters as compared to HNLC waters (**Supplementary Fig. 1C, Supplementary Table 1** and **Supplementary Methods**) [35]. Station M2 was visited three times at an 8-day interval, and station M3 and M4 were visited twice at a two-week interval. For metagenomes, triplicate 6 L seawater samples, collected by Niskin bottles, were each filtered through 0.8 µm Polycarbonate filters (PC, Nuclepore). The cells in the < 0.8 µm fraction were concentrated in 0.2 µm Sterivex filter units (Millipore). Total genomic DNA was extracted from the Sterivex filter units using the AllPrep DNA/RNA kit (Qiagen, Hiden, Germany) with modifications (**Supplementary Methods**). The DNA was extracted from the triplicate seawater samples collected during each of the repeated visits per station. 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. Three metagenomic libraries (one per station) were prepared using the Illumina Nano library preparation kit. For metatranscriptomes, 10 L seawater samples were immediately pre-filtered through 0.8 µm PC filters (Nucelpore) and the cells in the < 0.8 µm fraction concentrated on 0.22 µm Express Plus Polyethersulfate (PES) filters (Millipore). RNA

was extracted from the samples collected during the first visit at each site using the NucleoSpin® RNA Midi kit (Macherey-Nagel, Düren, Germany). Two internal standard RNA molecules were synthesized and added to each sample with known copy numbers. Technical details are provided in **Supplementary Methods**. Nine metatranscriptomic libraries (3 triplicates × 3 stations) were prepared using the Illumina TruSeq Stranded mRNA Library Prep kit. Paired-end sequencing (2 × 150 bp) was performed on Illumina HiSeq 4000 platform at Fasteris SA, Inc. (Switzerland).

Metagenome assembly and binning

Quality control (QC) passed reads from each sample were co-assembled using MEGAHIT (v1.0.4) [36] with default settings, resulting 949,228 contigs with a minimum length of 1,000 bp (**Supplementary Methods**). Sequencing and assembly statistics are summarized in **Supplementary Fig. 2** and **Supplementary Table 2**. MetaWRAP (v1.1.3) [37] was used to assign contigs (≥ 2,500 bp) into metagenome-assembled genomes (MAGs) with the aid of three binning tools, including CONCOCT (v1.0.0) [38], MaxBin (v2.2.5; -markerset 40 -prob_threshold 0.5) [39] and MetaBAT (v2.12.1) [40]. Further refinement was implemented, based on read coverage and GC content of each contig in a MAG, by multivariate outlier detection using the aq.plot function in the mvoutlier package (v2.0.9) from R (v3.6.1). MAGs after contig outlier removal were reassessed using CheckM (v1.1.2) [41] for completeness and redundancy. Sequence-discrete populations closely related to the MAGs were identified as previously described using BBMap (v38.22) [42] (**Supplementary Fig. 3**). Comparisons between our Southern Ocean assemblies and existing databases, including the NCBI and the TARA Ocean Global Expedition Project, were performed to evaluate the novelty of our data (**Supplementary Fig. 2** and **4**). Technical details are thoroughly described in **Supplementary Methods**.

Metagenome functional profiling

A total of 3,003,586 protein-coding genes were identified from the 949,228 contigs by Prodigal (v2.6.3) [43] under meta mode (-p meta). Functional annotation was carried out against eggnog [44] using eggNOG-mapper (v1.0.3) [45], Pfam [46] using HMMER (v3.2.1) [47], KEGG [48] using GhostKOALA (v2.2) [49] and KofamKOALA (v1.0.0) [50], TCDB [51] using BLASTP (v2.7.1) [52], CAZy [53] using dbCAN2 (v2.0.1) [54], and MEROPS [55] using BLASTP (v2.7.1) [52]. Iron-related genes were further examined by FeGenie [56], and Fe-containing domains were characterized using Superfamily (v1.75) [57].

Metagenome taxonomic profiling

Taxonomy classification of the 133 MAGs was determined using the classify_wf function of the GTDB-Tk toolkit [58] based on the Genome Taxonomy Database (v0.3.0) (**Supplementary Table 3**). For phylogeny inference, 218 single-copy orthologous gene families shared by at least 20 (out of 133) MAGs were identified by OrthoFinder (v2.2.3) [59], aligned with MAFFT (v7.313) [60] and filtered by trimAl (v1.4) [61]. Maximum Likelihood (ML) phylogenetic reconstruction was performed based on the concatenation of the proteins using IQ-Tree (v1.6.8; -m TESTMERGE -bb 1000 -bnni) [62] (**Fig. 1**). Metagenome-assembled genes which were not included in the MAGs were subjected to taxonomic classification using Kaiju (v1.7.0) [63] with its precompiled nr databases. We also quantified taxonomic diversity and relative abundance in each sample by using SSU reconstruction and assembly-free taxonomic classifiers (**Supplementary Fig. 5**). Technical details are thoroughly described in **Supplementary Methods**.

Metatranscriptome transcript abundance and gene expression profiling

Read counts of each gene were generated using featureCounts (v2.0.0) [64] with the BAM files produced by mapping QC-passed meta-genomic/transcriptomic reads to the 949,228 annotated contigs using Bowtie2 (v2.3.5) [65] (**Supplementary Methods**). Besides common shared options including "-Q 1 --primary -p -B -P -C", different settings were used for metagenomic ("-O --fracOverlap 0.25 --ignoreDup -s 0") and metatranscriptomic ("-s 2") reads. Based on internal standard recoveries (**Supplementary Methods**), we estimated the quantitative inventories of transcripts per liter of each gene, and enumerated transcripts mediating key iron uptake and carbon metabolism pathways (**Fig. 2**, **Supplementary Fig. 6-8**, and **Supplementary Table 4**). Functional diversity was measured by Shannon index based on the abundance matrix of functional groups in each sample using the "diversity" function of the vegan package in R (v3.6.1). The abundance of each functional group in a sample (M2, M3 or M4) was defined as the sum of all transcripts of genes assigned with the corresponding function, $f_{Abundance}(Func_i) = \sum_{gene \in Func_i} f_{Transcripts\ L^{-1}}(gene)$. The taxonomic composition of a functional group was assessed by the ratio of $\frac{\sum_{gene \in Func_i \cap Tax_j} f_{Transcripts\ L^{-1}}(gene)}{\sum_{gene \in Func_i} f_{Transcripts\ L^{-1}}(gene)}$.

Further, differential gene expression analyses were performed in two ways, including on the original metatranscriptomic read counts and on the metagenome-normalized metatranscriptomic profile. The normalization was performed to minimize the influence of genome abundance on the assessment of gene expression levels, given that fluctuations in transcript abundance could be a result of shifting genome copies rather than changes in expression levels (**Supplementary Fig. 9A**) [14]. We normalized the metatranscriptomic profile by relative gene abundance through the division of variance-stabilizing transformed count tables, and then converted the ratios to integer pseudo-counts in a range of 0 to 10^6 (**Supplementary Fig. 9B**). DESeq2 (v1.24.0) was applied to identify significantly differentially expressed genes

(SDEGs) across contrasting oceanic regions (on- vs. off-plateau) at a false discovery rate (FDR) threshold of 0.1 [66]. Considering that during the MOBYDICK cruise the two HNLC sites (M3 and M4) were located at distinct water bodies separated by the Antarctic Polar Front (**Supplementary Fig. 1A**), we included in our design formula, besides the factor of iron concentration gradients, a term representing the influence of the Antarctic Polar Front. Genes, identified as significantly differentially expressed, were further summarized according to their functional categories (**Fig. 3, Supplementary Fig. 10-11** and **Supplementary Table 4-5**).

Results and Discussion

A novel Southern Ocean meta-omics resource

An average of 316.4 million (M) pairs of high-quality metagenomic reads were obtained from each station, achieving approximately 95% average coverage of the sampled communities (**Supplementary Fig. 2A** and **Supplementary Table 2**) [67]. Combined with another pre-sequenced metagenome from station M2 in early spring, a total of 1,286.5 M pairs of reads were assembled into 949,228 non-redundant contigs (≥ 1,000 bp), on which 3,003,586 protein-coding genes were identified. Most contigs and genes in the assembled metagenome had low similarity to sequences in the NCBI nt database (**Supplementary Fig. 2B-E** and **Supplementary Methods**), underlying the novelty of our data. Although almost half of the predicted proteins displayed homology with high similarity to sequences in the NCBI nr database (bitscore ≥ 200 and E-value < 1e-10) [68], the amount of near identical matches (percentage of identity ≥ 90%) does not exceed 15.32% and another 717,088 (23.87%) proteins have no significant homologs found in the nr database (**Supplementary Fig. 2FG**).

A total of 133 Southern Ocean MAGs were recovered, among which 116 have a completeness ≥ 50% and a redundancy < 5% (**Fig. 1** and **Supplementary Table 3**). The Southern Ocean MAGs represent a wide range of taxonomic groups, including 4 archaeal and 129 bacterial genomes. The classes of Alphaproteobacteria (n=34), Gammaproteobacteria (n=35) and Bacteroidia (n=39) dominated the bacterial Southern Ocean MAGs, while other members belonging to Planctomycetota (n=5), Myxococcota (n=3), Verrucomicrobiota (n=4) and Actinobacteriota (n=3) were also present. Metagenomic read recruitment revealed variable abundance for some taxonomic groups across sampling sites, including the Pelagibacterales order (also known as the SAR11 clade) and Flavobacteriaceae family (**Supplementary Fig. 3**).

Only 13 of our Southern Ocean MAGs conform to > 95% intra-species ANI values with counterparts from the TARA Ocean Global Expedition (**Supplementary Fig. 4A** and **Supplementary Table 3**) [69]. Metagenomic read recruitment analysis further confirmed that the novelty of our Southern Ocean assemblies is not derived from biases introduced during metagenome assembly and binning (**Supplementary Fig. 4B-F** and **Supplementary Information**). The protein-level comparison indicated more functional similarity than diversity, given that around 90.51% of the proteins in the Southern Ocean MAGs were assigned with orthologs in the TARA assemblies (**Supplementary Fig. 4GH**).

Taxonomic profiling of prokaryotic communities

Regarding the considerable amount of metagenomic reads that could not be assembled into MAGs (**Supplementary Table 2**), we carried out additional taxonomic profiling analyses using both 16S rRNA reconstruction-based and assembly-free methods (**Supplementary Methods**), in order to obtain a less biased estimate of the microbial community composition in our samples. Overall, the dominance of Alphaproteobacteria, Gammaproteobacteria and Bacteroidia, as well as the variations in diversity and abundance of individual taxa across samples (**Supplementary Fig. S5ABE**), agreed with the observations in the 133 Southern Ocean MAGs. The SAR11 clade was ubiquitous and abundant across all sampling sites, and its 16S rRNA assemblies displayed high phylogenetic diversity (**Supplementary Fig. 5A**). Species diversity, measured by Shannon index, was higher for the microbial communities in the off-plateau HNLC M3 and M4 sites than those in the on-plateau M2 site (**Supplementary Fig. S5C**). Microbial community variability among stations was explored with double principal coordinate analysis (DPCoA) followed by Monte Carlo permutation tests, incorporating not only information on abundance patterns but also phylogenetic structures (**Supplementary Methods**). Ordination of communities by DPCoA

revealed a significant clustering of taxonomic groups along the first principal component correlated with contrasting nutrient regimes (p-value ≈ 0.001; **Supplementary Fig. S5D** and **Supplementary Methods**). However, the statistical significance of categorical explanatory variable (on- vs. off-plateau waters) could not be appropriately assessed due to the small number of representative samples per environment (**Supplementary Methods**).

Community functional potential and gene expression patterns

To assess the functional potential and gene expression patterns at the community level, we examined key functions in carbon and iron metabolism and the contribution of prokaryotic taxa to the respective functional groups across the metagenomes and metatranscriptomes obtained from the different sites (**Supplementary Fig. 6-7** and **Supplementary Table 4**). The overall functional potential, based on gene presence and absence in metagenomes, was similar across sampling sites at the community level (**A-D** in **Supplementary Fig. 6-7**). Key metabolic genes were universally present in all samples, including those involved in iron uptake and carbohydrate-active enzymes (CAZymes). We further examined the community-level transcript abundance based on the normalized per-liter transcripts estimated following the internal standards protocol (**E-J** in **Supplementary Fig. 6-7**) [70]. We did not observe an overall enrichment of functional groups related to iron and carbon metabolism in either the on- or off-plateau prokaryotic communities, but the patterns were quite patchy. For example, the siderophore transporters belonging to ExbD (K03559), ExbB (K03561), TonB (K03832), TonB-dependent outer membrane receptors (K16087 and K02014), ferrous iron transporter FeoA (K04758) and FeoB (K04759), as well as two transcriptional regulators Fur (K09823) and TroR (K03709), had higher abundance in the iron-fertilized waters. To the contrary, the vitamin B12 transporter (K16092), putative hemin transport protein HmuS (K07225), heme iron utilization

protein HugZ (K07226), vacuolar iron transporter VIT (K22736), ferredoxin/flavodoxin switch relate HemG (K00230), ferric transporters FbpA (K02012), a siderophore transporter (K02016), and another Fur transcriptional regulator Irr (K09826) was more abundant in the off-plateau waters.

To better explain this mosaic pattern, we explored the possible link between taxonomy and function. We measured the functional diversity using the Shannon index based on the abundance matrix of functional groups across samples. In contrary to the species diversity (**Supplementary Fig. S5C**), the functional diversity of the on-plateau M2 site was no less than the average of the off-plateau M3 and M4 sites (**Fig. 2A** and **Supplementary Fig. 8A**). That is, the evidently lower species diversity in the on-plateau waters was decoupled from the community functional structure. The taxonomic compositions within functional groups across study sites provided a complementary perspective (**Supplementary Fig. 6-7**). Overall, the SAR11 clade contributed slightly higher to the gene pool and transcript inventories in the off-plateau waters, whilst Flavobacteriales made up a larger share in the iron-fertilized on-plateau zone (**Fig. 2B** and **Supplementary Fig. 8B**). The Roseobacterales and Gammaproteobacteria adopted more flexible ecological strategies, as their contributions to the functional pool were similar in different waters. This pattern is consistent with the clear separation in community taxonomic composition across divergent environmental conditions (**Supplementary Fig. 5D**) and suggests the aforementioned mosaic transcript abundance as a result of environmental nutrient availability and microbial life strategies. That is to say, when the variation in environmental conditions leads to the selection for specific metabolic functions (e.g., DOC degradation, access to iron), the taxonomic variation within functional groups would be a result

of both the importance of the specific function and the phylogenetic distribution of those functions [71].

We further recovered the gene expression profiles by normalizing the metatranscriptomic transcript abundance using the metagenomic gene abundance (**Supplementary Fig. 9**). The SDEGs obtained with and without the metagenome-based normalization were partially overlapped (**Supplementary Fig. 10-11**), confirming that prokaryotic community transcripts vary as a function of shifts in both community composition and gene expression levels [14]. We classified SDEGs according to their functional groups and taxonomic affiliations, and confirmed that gene expression patterns were not fully determined by nutrient regimes, but more taxonomy-resolved with microenvironmental considerations (**Fig. 3** and **Supplementary Fig. 10-11**).

The Flavobacteriales group and Gammaproteobacteria constituted the majority of the SDEGs belonging to the glycoside hydrolysis (GH) and glycosyltransferase (GT) families, which primarily have higher expression levels in the on-plateau iron-fertilized waters. Among them, the most corresponding GH families included GH16 and GH17 responsible for the decomposition of glucans and galactans, GH92 for the degradation of mannoses, and the β-1,3-D-glucan phosphorylases GH149 with inconclusive roles [72]. Gammaproteobacteria and Flavobacteriales were also enriched in the SDEGs of the GH3 family, which facilitates the utilization of glucose, arabinose and xylose. These GH3 SDEGs were higher expressed in either the on- or off-plateau region. This intra-taxonomy difference in expression patterns across different nutrient regimes suggests a mixture of copio- and oligotrophic life strategists applied within these two taxonomic groups [73]. On the contrary, the SDEGs of the GH23 and GH73 families were mainly from the SAR11 clade. GH23 (lytic transglycosylases) and GH73 (β-N-acetylglucosaminidases) are both involved in peptidoglycan degradation, an essential macromolecule of the bacterial outer cell

wall. The higher expression of these families in the off-plateau HLNC waters could indicate the use of peptidoglycan as carbon source or the accelerated growth of SAR11.

The SDEGs involved in Fe-uptake and Fe-related pathways exhibited similar patterns (**Fig. 3** and **Supplementary Fig. 11**). SAR11 constituted a great proportion of the SDEGs coding for two iron-related transcriptional regulators (IscR and Irr; **Fig. 3**). IscR monitors Fe-S cluster homeostasis and is responsible for the autorepression of genes involved in Fe-S cluster biogenesis, such as the *sufBCD* operon [74]. Under oxidative stress and iron starvation, IscR is in its apoform and relieves its repression of the *suf* operon [75, 76]. Irr, a global regulator of iron homeostasis, functions as a sensor of the cellular heme biosynthesis and accumulates under iron limitation to control target genes [77, 78]. It is reported to be conserved in the SAR11 subgroup Ia and maintained by selection due to fitness advantage [79, 80]. The induction of the glyoxylate shunt (GS) is an efficient strategy for heterotrophic prokaryotes to maintain growth and respiration rates under iron stress [23, 81]. We examined three key enzymes related to the GS, including isocitrate lyase (K01637; ICL encoded by *aceA*) and malate synthase (K01638; MS encoded by *aceB*) within the GS pathway, as well as isocitrate dehydrogenase that catalyses the oxidative decarboxylation of isocitrate (K00031; IDH encoded by *icd*) (**Supplementary Fig. 11**). The upregulation of *aceA* and *aceB* indicates the elevation of the GS under stress conditions, whereas the variant expression of *icd* may provide a clue to the competition between IDH and ICL for the substrate isocitrate. The SAR11 clade accounts for a large share of the *aceA* and *aceB* SDEGs which have higher expression levels in off-plateau waters. The community-level abundance of GS-related transcripts generally agree with the observations of SDEGs (**Supplementary Fig. 6**). For instance, the amount of SAR11 *aceA* transcripts L^{-1} were more than doubled in the off-plateau samples. The enrichment of both *aceA*/*aceB* and *icd* suggests that

the SAR11 GS system functions supplementary to the classic TCA cycle in response to iron and/or carbon limitation. Gammaproteobacteria and Flavobacteriales dominated the SDEGs encoding the TonB-ExbB-ExbD complex for siderophore uptake. Related functional groups were generally enriched with SDEGs in both the iron-fertilized and the HNLC waters at the community level, however associated to different taxa.

Taxon-specific ecological roles

To resolve ecological roles of prokaryotic taxa across contrasting oceanic waters, we proceeded our data mining effort at a finer resolution with the 133 MAGs (**Fig. 1**). Initially, we performed a systematic survey for metabolically active prokaryotes through transcript abundances. Ribosomal proteins (RP) are critical for protein synthesis and levels of RP transcripts have been proposed as an indicator for prokaryotic growth rates [82-85]. We surveyed 93 prokaryotic RP KEGG Orthology groups (KOs) through all our assemblies (**Supplementary Table 4**). Generally, taxa with higher grow rates (more RP transcripts) had also high cell metabolism (more total transcripts) in the off-plateau HNCL waters ($R^2 > 0.9$), whereas in the on-plateau zone several MAGs showed an all-vs.-RP ratio depart from the fitted line ($R^2 < 0.5$; **Fig. 4**). This provided us with interesting insights. First, environmental properties in the iron-fertilized on-plateau zone lead to a decoupling between cell metabolism and growth of several MAGs. Secondly, individual species from closely related taxonomic groups revealed diverse ecological strategies. For example, MAG_91, although it forms a monophyletic clade with MAG_126 on the phylogeny tree (**Fig. 1**), were better adapted to the HNLC environment at site M4 as compared to the on-plateau zone (**Fig. 4**).

The diversity in ecological strategies became more evident when we examined central metabolic pathways gene by gene (**Fig. 5**, **Supplementary Fig. 12** and **Supplementary Table 4-**

5). While the SAR11 SDEGs tuned their expression in a relatively consistent manner among individual MAGs, we observed diverse expression patterns of SDEGs belonging to Flavobacteriales and Gammaproteobacteria. Generally, the Flavobacteriales MAGs constituted more genes that were significantly higher expressed in the iron-fertilized on-plateau zone, whereas a limited number of them displayed the opposite pattern by downregulating the expression of genes responsible for iron uptake and carbon metabolism in the same water.

Particularly, with respect to polysaccharide metabolism, two MAGs (MAG_51 and 99), belonging to the same Flavobacteriales UA16 genus but sharing an inter-species ANI value of 79.05% [69], exhibited contrasting expression patterns. Moreover, four MAGs (MAG_78, 3, 134 and 73) from the Flavobacteriales 1G12 family but distinct genera, which formed a monophyletic clade on the phylogenomic tree, also reflected niche divergence. For instance, polysaccharide utilization loci (PULs) are operon-like gene structures that encode co-regulated proteins that specialize in polysaccharide detection, uptake and hydrolysis [86]. PULs are prevalent in the Bacteroidota phylum, and typically feature with SusD-like substrate-binding proteins, TonB-dependent receptors (TBDR) and various CAZymes. We identified several PUL-like loci in four Flavobacteriales UA16 MAGs (**Supplementary Fig. 13A-D** and **Supplementary Methods**). All the UA16 MAGs consisted of a "GH149 + GH30 + GH16 (and/or GH17) + transporters" PUL structure, indicating a general utilization of glucans and galactans, but MAG_73 contained a unique "GH92 + GH78 (+ CBM67 domain) + transporter" locus. MAG_73 was ubiquitous in all sampling sites but more abundant in the off-plateau waters (**Fig. 1** and **Supplementary Table 3**) and exhibited an opposite expression pattern to the other three (**Fig. 5**). The GH92 family exo-α-mannosidases function on α-linked mannose residues in an exo-acting manner and therefore are responsible for the depolymerization of α-linked mannans [87]. Algal mannans are widely

distributed in marine ecosystems, and α-mannans were identified in red seaweed [88, 89] and the diatom *Phaeodactylum tricornutum* [90]. Recent studies demonstrated that marine bacteria, especially Bacteroidota, can degrade mannans [91-93]. Although whether the GH92 PUL facilitates the growth of MAG_73 across all sampling sites requires further study due to the incompleteness of MAGs. The utilization of mannans elucidates the metabolic potential of MAG_73 and suggests that it could occupy distinct niches as compared to MAG_78, 3 and 134. Fucose is another bioavailable monosacharide common in marine waters and released by diatoms [94]. We identified candidate gene clusters specific for fucose utilization among Verrucomicrobiae MAGs, showing orthology to the recently discovered functional loci in *Lentimonas* sp. CC4 (**Supplementary Fig. 13E**) [95].

We also observed MAGs with similar expression patterns in late summer above the plateau, but contrasting abundances in samples from early spring at the same site (**Supplementary Table 3**) providing subtle clues to seasonal adaptation. This was specifically the case for the metabolically active on-plateau MAG_103 and MAG_62 belonging to the Pseudomonadales HTCC2089 family but different genera (**Fig. 5**). To explore the potential genetic reasons behind this observation, we constructed pan-genomes by using both our MAGs and their closely related reference genomes. The incorporation of reference genomes into our analysis is to compensate the incompleteness of our MAGs. For the comparison between the MAG_103 and MAG_62, a total of 19 Pseudomonadales HTCC2089 draft genomes were retrieved from NCBI GenBank database based on the phylogenetic information provided by GTDB [58], including 15 from UBA4421 genus and 4 from UBA9926 (**Supplementary Methods**). The two MAGs shared most of their polysaccharide degradation and proteolysis genes and perceived to be competitors for similar resources (**Supplementary Table 6**). We

identified one singleton chitinase (GH18) unique to MAG_103 but missing from all other Pseudomonadales HTCC2089 draft genomes, and three GHs (GH17, GH149 and GH158) conserved in MAG_62 but absent from the UBA4421 genus. However, as discussed above, GH149 is with indecisive function, and GH17 and GH158 share similar substrate specificities with GH16, which is common in HTCC2089. The number of peptidases-encoding genes was also comparable in MAG_103 (n=102) and MAG_62 (n=104).

We made an unexpected observation that could explain the different abundance patterns of these MAGs in early spring and late summer. We identified a gene cluster in MAG_103 related to light-induced energy acquisition, which was conserved in the UBA4421 genus but absent from the UBA9926 genomes (**Supplementary Fig. 14AB** and **Supplementary Table 6**). This gene cluster consisted of 6 genes, encoding a bacteriorhodopsin (PF01036.18), a synthase (PF00348.17), a phytoene desaturase (*crtI*; K10027), a 15-cis-phytoene synthase (*crtB*; K02291), a lycopene beta-cyclase (*crtL1*; K06443), and a beta-carotene 15,15'-dioxygenase (*blh*; K21817). Bacteriorhodopsin could facilitate MAG_103 with the capability to use light as a supplemental energy source [96, 97]. The MAG_103 bacteriorhodopsin sequence contained a blue light absorbing glutamine “Q” and a proton pumping motif “DTE” (**Supplementary Fig. 14C**) [98]. The *crtI*, *crtB* and *crtL1* genes are involved in the internal retinal biosynthesis system and responsible for beta-carotene biosynthesis. Further, the beta-carotene dioxygenase encoded by *blh* cleaves the beta-carotene to produce all-trans retinal, which could be used by MAG_103 as its photoreactive chromophore [99]. The activity of beta-carotene dioxygenase was reported to be iron dependent [100], and the expression of the *blh* gene was slightly upregulated, but not significantly, in the iron-fertilized region. All the other four genes involved in light harvest were significantly higher expressed in the on-plateau region as compared to the off-plateau waters.

The extra energy supplied from light might be the reason that MAG_103 became more competitive in stratified summer surface waters.

MAG_103 and MAG_62 differed in their potential to use and resist to antibiotics (**Supplementary Results** and **Supplementary Table 6**). While competing with other prokaryotes in the late summer surface waters when bulk abundances reached 1.18×10^9 cells L^{-1} (**Supplementary Table 1**), the production of antibiotics might greatly facilitate MAG_103's dominance over other species. MAG_103 genes involved in antibiotic production were higher expressed in the on-plateau region. The expression level of MAG_103 genes encoding the general secretory pathway proteins (*gspC*, *gspD*, *gspE*, *gspF*, *gspG* and *gspL*), as well as the *sec* translocase system (*secA*, *secB* and *secD*), were significantly upregulated in the iron-fertilized on-plateau water. The type II secretion (T2S) pathway, coupled with Sec translocon, is regarded as the main protein secretion pathway of bacteria, which is capable of transporting a wide range of substrates, including proteases, lipases, phosphatases, carbohydrates-degrading enzymes and toxins [101]. The translocation of antibiotics produced by MAG_103 through the outer membrane might be mainly facilitated by the T2S. Within the two-component system (TCS), MAG_103 genes encoded a phosphate regulon sensor histidine kinase PhoR, a phosphate regulon response regulator OmpR, an osmolarity sensor histidine kinase EnvZ, and an invasion response regulator UvrY were all significantly upregulated, possibly due to stress specific responses. Further, MAG_103 genes related to iron uptake, encoding a ferric transporter (*fbpA*), a ferredoxin (*fdx*), the TonB-ExbB-ExbD system, a bacterioferritin (*bfr*), a vitamin B12 transporter (*btuB*), a bacteriorhodopsin, and the Fe-S cluster assembly proteins (*nfuA* and *sufBCD*), were also significantly higher expressed in the iron-fertilized water. The significant increase in expression levels of TCA cycle enzymes, CAZymes, more than half of the peptidases

[55], and the aerobic carbon-monoxide dehydrogenase subunits (*coxS* and *coxL*) indicated an enhanced carbon flux between phytoplankton and MAG_103 represented Pseudomonadales population during the bloom decline. Although these accessory genomic features and their corresponding expression patterns are not direct evidence related to iron and carbon metabolism, the enhanced competence may facilitate the survival and growth of the microbes, whose abundances influence their roles in nutrient cycles.

Deciphering the many unknowns regarding the ecological roles of marine prokaryotes inhabiting the Southern Ocean, undoubtedly a region of key importance in ongoing global warming, remains profoundly challenging. Here we provide a comphrehensive investigation of prokaryotic functional activities from the community level to individual taxa, targeting in situ responses linked to iron and carbon cycling. Despite remarkable shifts in community composition across contrasting nutrient regimes, we observed conservation of functional diversity through functional redundancy among community members inhabiting each ecosystem. The distinct gene expression patterns of individual taxa illustrate the link between the genetic repertoire of prokaryotic taxa and their diverse responses to the multitude of environmental factors. Our observations of a mosaic of taxonomy-specific ecological strategies in the cycling of iron and organic carbon provides insights how the habitat shapes microbial diversity in the ocean.

Acknowledgements

We thank B. Quéguiner, the PI of the MOBYDICK project, for providing us the opportunity to participate to this cruise, the captain and crew of the R/V Marion Dufresne for their enthusiasm and support aboard during the MOBYDICK–THEMISTO cruise (https://doi.org/10.17600/18000403). This work was supported by the French oceanographic fleet ("Flotte océanographique française"), the French ANR ("Agence Nationale de la Recherche", AAPG 2017 program, MOBYDICK Project number: ANR-17-CE01-0013), the French Research program of INSU-CNRS-LEFE/CYBER ("Les enveloppes fluides et l'environnement" - "Cycles biogéochimiques,environnement et ressources") and the Austrian FWF grant under the number P28781-B21. The authors thank the Roscoff Analyses and Bioinformatics for Marine Sciences Platform (ABiMS; http://abims.sb-roscoff.fr/) and the French Institute of Bioinformatics (IFB; https://www.france-bioinformatique.fr) for providing computational facilities and technical supports. We thank S. Blain for providing the satellite-based chlorophyll a data for the period 1998-2017. We also thank M. A. Moran and M. Landa for providing the internal standards for further *in vitro* transcription. Two reviewers provided detailed and insightful comments that helped to improve previous versions of our manuscript.

Author Contributions

I.O. conceived the project and designed the experiments. P.D. and I.O. participated to the cruise. P.D. collected the samples and carried out the nucleic acid extraction. P.D. performed metagenome read processing and contig assembly. Y.S. carried out bioinformatics analysis. Y.S. and I.O. wrote the manuscript, and P.D. provided input to the results and commented on the manuscript.

Competing Interests

The authors declare no competing interests.

Data Availability

The data sets generated and analysed during the current study are available in the European Nucleotide Archive (ENA) repository at https://www.ebi.ac.uk/ena under the project ID PRJEB37465 (metagenome) and PRJEB37466 (metatranscriptome). The metagenome reads are under the accession number ERR4234198- 4234200. The metatranscriptome reads are under the accession number ERR4234183-4234191. The 949,228 contigs are under the accession number ERZ1694383. The 133 MAGs are under the accession number ERZ1694384-1694516.

References

1. Azam F, Malfatti F. Microbial structuring of marine ecosystems. Nat Rev Microbiol. 2007;5(10):782-91.
2. Lechtenfeld OJ, Hertkorn N, Shen Y, Witt M, Benner R. Marine sequestration of carbon in bacterial metabolites. Nat Commun. 2015;6:6711.
3. Buchan A, LeCleir GR, Gulvik CA, Gonzalez JM. Master recyclers: features and functions of bacteria associated with phytoplankton blooms. Nat Rev Microbiol. 2014;12(10):686-98.
4. DeLong E (ed). Microbial Metagenomics, Metatranscriptomics, and Metaproteomics, 1st edn. San Diego, CA, USA: Academic Press; 2013.
5. White III RA, Callister SJ, Moore RJ, Baker ES, Jansson JK. The past, present and future of microbiome analyses. Nat Protoc. 2016;11(11):2049.
6. Hanson CA, Fuhrman JA, Horner-Devine MC, Martiny JB. Beyond biogeographic patterns: processes shaping the microbial landscape. Nat Rev Microbiol. 2012;10(7):497-506.
7. Gilbert JA, Steele JA, Caporaso JG, Steinbruck L, Reeder J, Temperton B, et al. Defining seasonal marine microbial community dynamics. ISME J. 2012;6(2):298-308.
8. Needham DM, Fuhrman JA. Pronounced daily succession of phytoplankton, archaea and bacteria following a spring bloom. Nat Microbiol. 2016;1:16005.
9. Lindh MV, Sjostedt J, Andersson AF, Baltar F, Hugerth LW, Lundin D, et al. Disentangling seasonal bacterioplankton population dynamics by high-frequency sampling. Environ Microbiol. 2015;17(7):2459-76.
10. Fuhrman JA, Cram JA, Needham DM. Marine microbial community dynamics and their ecological interpretation. Nat Rev Microbiol. 2015;13(3):133-46.

11. Ruiz-González C, Logares R, Sebastián M, Mestre M, Rodríguez-Martínez R, Galí M, et al. Higher contribution of globally rare bacterial taxa reflects environmental transitions across the surface ocean. Mol Ecol. 2019;28(8):1930-45.

12. 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.

13. 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.

14. Salazar G, Paoli L, Alberti A, Huerta-Cepas J, Ruscheweyh HJ, Cuenca M, et al. Gene expression changes and community turnover differentially shape the global ocean metatranscriptome. Cell. 2019;179(5):1068-83 e21.

15. Boeuf D, Edwards BR, Eppley JM, Hu SK, Poff KE, Romano AE, et al. Biological composition and microbial dynamics of sinking particulate organic matter at abyssal depths in the oligotrophic open ocean. Proc Natl Acad Sci U S A. 2019;116(24):11824-32.

16. Ghiglione JF, Galand PE, Pommier T, Pedros-Alio C, Maas EW, Bakker K, et al. Pole-to-pole biogeography of surface and deep marine bacterial communities. Proc Natl Acad Sci U S A. 2012;109(43):17633-8.

17. Frias-Lopez J, Shi Y, Tyson GW, Coleman ML, Schuster SC, Chisholm SW, et al. Microbial community gene expression in ocean surface waters. Proc Natl Acad Sci U S A. 2008;105(10):3805-10.

18. Acinas SG, Sánchez P, Salazar G, Cornejo-Castillo FM, Sebastián M, Logares R, et al. Metabolic architecture of the deep ocean microbiome. bioRxiv. 2019:635680. https://doi.org/10.1101/635680.

19. Martin JH, Gordon RM, Fitzwater SE. Iron in Antarctic waters. Nature. 1990;345(6271):156-8.

20. Church MJ, Hutchins DA, Ducklow HW. Limitation of bacterial growth by dissolved organic matter and iron in the Southern ocean. Appl Environ Microbiol. 2000;66(2):455-66.

21. Obernosterer I, Fourquez M, Blain S. Fe and C co-limitation of heterotrophic bacteria in the naturally fertilized region off the Kerguelen Islands. Biogeosciences. 2015;12(6):1983-92.

22. Fourquez M, Obernosterer I, Blain S. A method for the use of the radiotracer 55Fe for microautoradiography and CARD-FISH of natural bacterial communities. FEMS Microbiol Lett. 2012;337(2):132-9.

23. Koedooder C, Gueneugues A, Van Geersdaële R, Vergé V, Bouget F-Y, Labreuche Y, et al. The role of the glyoxylate shunt in the acclimation to iron limitation in marine heterotrophic bacteria. Front Mar Sci. 2018;5:435.

24. Blain S, Tagliabue A (eds). Iron Cycle in Oceans, 1st edn. London, UK: ISTE Ltd and John Wiley & Sons, Inc.; 2016.

25. Dittmar T, Arnosti C. An inseparable liaison: marine microbes and nonliving organic matter. In: Gasol JM, Kirchman DL, editors. Microbial Ecology of the Oceans, 3rd edn. Hoboken NJ, USA: John Wiley and Sons, Inc.; 2018, pp 189–229.

26. 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.

27. Lasbleiz M, Leblanc K, Armand LK, Christaki U, Georges C, Obernosterer I, et al. Composition of diatom communities and their contribution to plankton biomass in the naturally ironfertilized region of Kerguelen in the Southern Ocean. FEMS Microbiol Ecol. 2016;92:fiw171.

28. 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.

29. Blain S, Capparos J, Guéneuguès A, Obernosterer I, Oriol L. Distributions and stoichiometry of dissolved nitrogen and phosphorus in the iron-fertilized region near Kerguelen (Southern Ocean). Biogeosciences. 2015;12(2):623-35.

30. d'Ovidio F, Della Penna A, Trull TW, Nencioli F, Pujol M-I, Rio M-H, et al. The biogeochemical structuring role of horizontal stirring: Lagrangian perspectives on iron delivery downstream of the Kerguelen Plateau. Biogeosciences. 2015;12(19):5567-81.

31. Landa M, Blain S, Christaki U, Monchy S, Obernosterer I. Shifts in bacterial community composition associated with increased carbon cycling in a mosaic of phytoplankton blooms. ISME J. 2016;10(1):39-50.

32. Landa M, Blain S, Harmand J, Monchy S, Rapaport A, Obernosterer I. Major changes in the composition of a Southern Ocean bacterial community in response to diatom-derived dissolved organic matter. FEMS Microbiol Ecol. 2018;94(8).

33. Fourquez M, Beier S, Jongmans E, Hunter R, Obernosterer I. Uptake of Leucine, chitin, and iron by prokaryotic groups during spring phytoplankton blooms induced by natural iron fertilization off Kerguelen Island (Southern Ocean). Front Mar Sci. 2016;3:256.

34. Debeljak P, Toulza E, Beier S, Blain S, Obernosterer I. Microbial iron metabolism as revealed by gene expression profiles in contrasted Southern Ocean regimes. Environ Microbiol. 2019;21(7):2360-74.

35. Christaki U, Gueneugues A, Liu Y, Blain S, Catala P, Colombet J, et al. Seasonal microbial food web dynamics in contrasting Southern Ocean productivity regimes. Limnol Oceanogr. 2021;66(1):108-22.

36. 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.

37. Uritskiy GV, DiRuggiero J, Taylor J. MetaWRAP-a flexible pipeline for genome-resolved metagenomic data analysis. Microbiome. 2018;6(1):158.

38. Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, et al. Binning metagenomic contigs by coverage and composition. Nat Methods. 2014;11(11):1144-6.

39. Wu YW, Simmons BA, Singer SW. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics. 2016;32(4):605-7.

40. 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.

41. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015;25(7):1043-55.

42. 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.

43. 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.

44. Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, et al. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res. 2016;44(D1):D286-93.

45. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, et al. Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper. Mol Biol Evol. 2017;34(8):2115-22.

46. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, et al. The Pfam protein families database in 2019. Nucleic Acids Res. 2019;47(D1):D427-D32.

47. Eddy SR. Accelerated profile HMM searches. PLoS Comput Biol. 2011;7(10):e1002195.

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

49. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J Mol Biol. 2016;428(4):726-31.

50. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, et al. KofamKOALA: KEGG ortholog assignment based on profile HMM and adaptive score threshold. Bioinformatics. 2019;36:2251–2.

51. Saier MH, Jr., Reddy VS, Tsu BV, Ahmed MS, Li C, Moreno-Hagelsieb G. The transporter classification database (TCDB): recent advances. Nucleic Acids Res. 2016;44(D1):D372-9.

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

53. 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.

54. 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.

55. Rawlings ND, Barrett AJ, Thomas PD, Huang X, Bateman A, Finn RD. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. Nucleic Acids Res. 2018;46(D1):D624-D32.

56. Garber AI, Nealson KH, Okamoto A, McAllister SM, Chan CS, Barco RA, et al. FeGenie: a comprehensive tool for the identification of iron genes and iron gene neighborhoods in genome and metagenome assemblies. Front Microbiol. 2020;11:37.

57. Dupont CL, Yang S, Palenik B, Bourne PE. Modern proteomes contain putative imprints of ancient shifts in trace metal geochemistry. Proc Natl Acad Sci U S A. 2006;103(47):17822-7.

58. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics. 2019;36:1925–7.

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

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

61. 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.

62. 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.

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

64. 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.

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

66. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.

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

68. Pearson WR. An introduction to sequence similarity ("homology") searching. Curr Protoc Bioinformatics. 2013;Chapter 3:Unit3 1.

69. 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.

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

71. Louca S, Polz MF, Mazel F, Albright MBN, Huber JA, O'Connor MI, et al. Function and functional redundancy in microbial systems. Nat Ecol Evol. 2018;2(6):936-43.

72. Kuhaudomlarp S, Patron NJ, Henrissat B, Rejzek M, Saalbach G, Field RA. Identification of Euglena gracilis beta-1,3-glucan phosphorylase and establishment of a new glycoside hydrolase (GH) family GH149. J Biol Chem. 2018;293(8):2865-76.

73. Ho A, Di Lonardo DP, Bodelier PL. Revisiting life strategy concepts in environmental microbial ecology. FEMS Microbiol Ecol. 2017;93(3).

74. Rodionov DA, Gelfand MS, Todd JD, Curson AR, Johnston AW. Computational reconstruction of iron- and manganese-responsive transcriptional networks in alpha-proteobacteria. PLoS Comput Biol. 2006;2(12):e163.

75. Rincon-Enriquez G, Crete P, Barras F, Py B. Biogenesis of Fe/S proteins and pathogenicity: IscR plays a key role in allowing Erwinia chrysanthemi to adapt to hostile conditions. Mol Microbiol. 2008;67(6):1257-73.

76. Py B, Barras F. Building Fe–S proteins: bacterial strategies. Nat Rev Microbiol. 2010;8(6):436-46.

77. Zappa S, Bauer CE. Iron homeostasis in the Rhodobacter genus. Adv Bot Res. 2013;66:289–326.

78. Jaggavarapu S, O'Brian MR. Differential control of Bradyrhizobium japonicum iron stimulon genes through variable affinity of the iron response regulator (Irr) for target gene promoters and selective loss of activator function. Mol Microbiol. 2014;92(3):609-24.

79. Grote J, Thrash JC, Huggett MJ, Landry ZC, Carini P, Giovannoni SJ, et al. Streamlining and core genome conservation among highly divergent members of the SAR11 clade. mBio. 2012;3(5):e00252-12.

80. McAdams HH, Srinivasan B, Arkin AP. The evolution of genetic regulatory systems in bacteria. Nat Rev Genet. 2004;5(3):169-78.

81. Fourquez M, Devez A, Schaumann A, Guéneuguès A, Jouenne T, Obernosterer I, et al. Effects of iron limitation on growth and carbon metabolism in oceanic and coastal heterotrophic bacteria. Limnol Oceanogr. 2014;59(2):349-60.

82. Wilson DN, Nierhaus KH. The weird and wonderful world of bacterial ribosome regulation. Crit Rev Biochem Mol Biol. 2007;42(3):187-219.

83. Wei Y, Lee JM, Richmond C, Blattner FR, Rafalski JA, LaRossa RA. High-density microarray-mediated gene expression profiling of Escherichia coli. J Bacteriol. 2001;183(2):545-56.

84. Hendrickson EL, Liu Y, Rosas-Sandoval G, Porat I, Soll D, Whitman WB, et al. Global responses of Methanococcus maripaludis to specific nutrient limitations and growth rate. J Bacteriol. 2008;190(6):2198-205.

85. Gifford SM, Sharma S, Booth M, Moran MA. Expression patterns reveal niche diversification in a marine microbial assemblage. ISME J. 2013;7(2):281-98.

86. Sonnenburg ED, Zheng H, Joglekar P, Higginbottom SK, Firbank SJ, Bolam DN, et al. Specificity of polysaccharide use in intestinal bacteroides species determines diet-induced microbiota alterations. Cell. 2010;141(7):1241-52.

87. Gregg KJ, Zandberg WF, Hehemann JH, Whitworth GE, Deng L, Vocadlo DJ, et al. Analysis of a new family of widely distributed metal-independent alpha-mannosidases provides unique insight into the processing of N-linked glycans. J Biol Chem. 2011;286(17):15586-96.

88. Matulewicz M, Cerezo A. Water-soluble sulfated polysaccharides from the red seaweed Chaetangium fastigiatum. Analysis of the system and the structures of the α-D-(1→3)-linked mannans. Carbohydr Polym. 1987;7(2):121-32.

89. Kolender AA, Pujol CA, Damonte EB, Matulewicz MC, Cerezo AS. The system of sulfated α-(1→3)-linked D-mannans from the red seaweed Nothogenia fastigiata: structures, antiherpetic and anticoagulant properties. Carbohydr Res. 1997;304(1):53-60.

90. Le Costaouëc T, Unamunzaga C, Mantecon L, Helbert WJAr. New structural insights into the cell-wall polysaccharide of the diatom Phaeodactylum tricornutum. Algal Res. 2017;26:172-9.

91. Teeling H, Fuchs BM, Becher D, Klockow C, Gardebrecht A, Bennke CM, et al. Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. Science. 2012;336(6081):608-11.

92. 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.

93. Chen J, Robb CS, Unfried F, Kappelmann L, Markert S, Song T, et al. Alpha-and beta-mannan utilization by marine Bacteroidetes. Environ Microbiol. 2018;20(11):4127-40.

94. Biersmith A, Benner R. Carbohydrates in phytoplankton and freshly produced dissolved organic matter. Mar Chem. 1998;63(1):131-44.

95. 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; 5:1026–39.

96. Beja O, Aravind L, Koonin EV, Suzuki MT, Hadd A, Nguyen LP, et al. Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. Science. 2000;289(5486):1902-6.

97. DeLong EF, Beja O. The light-driven proton pump proteorhodopsin enhances bacterial survival during tough times. PLoS Biol. 2010;8(4):e1000359.

98. Olson DK, Yoshizawa S, Boeuf D, Iwasaki W, DeLong EF. Proteorhodopsin variability and distribution in the North Pacific Subtropical Gyre. ISME J. 2018;12(4):1047-60.

99. Kim SY, Waschuk SA, Brown LS, Jung KH. Screening and characterization of proteorhodopsin color-tuning mutations in Escherichia coli with endogenous retinal synthesis. Biochim Biophys Acta. 2008;1777(6):504-13.

100. von Lintig J, Vogt K. Filling the gap in vitamin A research. Molecular identification of an enzyme cleaving beta-carotene to retinal. J Biol Chem. 2000;275(16):11915-20.

101. Korotkov KV, Sandkvist M, Hol WG. The type II secretion system: biogenesis, molecular architecture and mechanism. Nat Rev Microbiol. 2012;10(5):336-51.

Figure Legends

Fig. 1 Genomic features of the 133 Southern Ocean (SO) metagenome-assembled genomes (MAGs) visualized using the circlize package (v0.4.9) in R (v3.6.1). The outmost circle shows the phylogenetic tree derived from the concatenation of 218 single-copy orthologous genes shared by the MAGs. The tip labels and branches are coloured according to their taxonomic affiliations determined by GTDB-Tk [58]. The 50 MAGs whose genes were excluded from the DGE analysis due to low read counts in either metagenomes or metatranscriptomes are marked with asterisks. The second circle ("Cov.") is a heatmap displaying the average coverage of depth (per million reads) of each MAG in each sample. The third to seventh circles are heatmaps showing the number of total transcripts ("Trans L^{-1}"), transcripts from genes encoding ribosomal proteins ("Ribo."), genes involved in Fe-related metabolic activities ("Fe") and TCA cycle ("TCA"), as well as genes encoding carbohydrate-active enzymes ("CAZy") in one liter of sampled seawaters (L^{-1}). The value of transcripts L^{-1} of each MAG was further normalized by the length of the MAG (Mbp). The colour schemes are given at the left bottom. The 8th circle illustrates the number of significantly differentially expressed genes at contrasting oceanic regions (on-plateau iron-fertilized vs. off-plateau HNLC waters). The orange bars represent the number of genes that are significantly higher expressed in the on-plateau M2 site, as compared to the off-plateau M3 and M4 sites. The blue bars summarize genes that are significantly higher expressed in the off-plateau HNLC waters.

Fig. 2 Community functional diversity and taxonomic composition within functional groups. Shannon index based on the abundance of functional groups (A) and shifts in taxonomic composition within functional groups (B) across sampling sites were calculated based on the

community-level transcript abundance represented by the normalized per-liter transcripts estimated following the internal standards protocol [70] (**E-J** 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 transcripts (L^{-1}) 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. Colour code is the same as **Supplementary Fig. 6-7**. Only functional groups consisting of at least 50 genes, out of the 3,003,586 protein-coding genes predicted from the metagenome assemblies, were used in the calculation.

Fig. 3 Statistics of significantly differentially expressed genes (SDEGs) involved in glycoside hydrolysis and key iron metabolic pathways. Panels from top to bottom represent glycoside hydrolase (GH), iron uptake regulators (Reg.), ferrous uptake (Fe2+), ferric uptake (Fe3+), siderophore biosynthesis and uptake (Sid. Syn./Upt.), heme uptake (Heme), iron storage (Sto.) and Ferredoxin/Flavodoxin switch (F/F). The full list of KEGG Orthology groups (KOs) related to iron metabolism examined in this study could be found in **Supplementary Table 4**. Each row represents one functional group. The two vertical panels show statistics of the SDEGs based on the metagenome-normalized metatranscriptomic pseudo counts and the corresponding log2-based fold changes. 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.

Fig. 4 Ratios of ribosomal-protein versus all transcripts (L^{-1} Mbp^{-1}) from 133 MAGs. Lines were calculated from Model II linear regression analyses. The corresponding formulas and R-squared measures are shown at the bottom.

Fig. 5 The distribution of significantly differentially expressed genes (SDEGs) in the MAGs among diverse functional categories related to iron uptake and carbon metabolism. Only 47 MAGs with SDEGs are shown here. From left to right, the panels represent the phylogenetic tree (the same as shown in **Fig. 1**), the iron-related KEGG Orthology groups (KOs), the KOs involved in the tricarboxylic acid (TCA) cycle, and carbohydrate-active enzymes (CAZymes). Each square block describes the statistics of a protein family in a MAG. An empty square suggests that no genes in the MAG (y axis) are classified into the corresponding functional group (x axis). A circle in the square block indicates the identification of homologs to a protein family in the MAG, with its size proportional to the number of genes assigned to that family. The square blocks are coloured according to the differential expression patterns of its gene(s). As illustrated in **Fig. 1**, genes, which are significantly higher expressed in the iron-fertilized site M2 as compared to the HNLC M3 and M4 sites, are highlighted in orange; vice versa, in blue. Given that genes belonging to the same functional group might not be synchronized in their expression patterns, the transparency of each square block shows the percentage of genes that are

significantly differentially expressed. We have not detected protein families whose genes were significantly shifting their expression levels in opposite directions (e.g., parts of the genes in the same family significantly upregulate their expression levels whilst others significantly downregulate theirs). The KO “K00240” (marked with an asterisk) is shown twice, because it is a Fe-S protein family and also involved in the TCA cycle. Among the pathways involved in carbon metabolism, KOs shared by multiple pathways are only shown once. All the information illustrated in this graph is summarized based on the differential expression analysis performed on the metagenome-normalized metatranscriptomic profile (see **Methods**).

Fig. 1

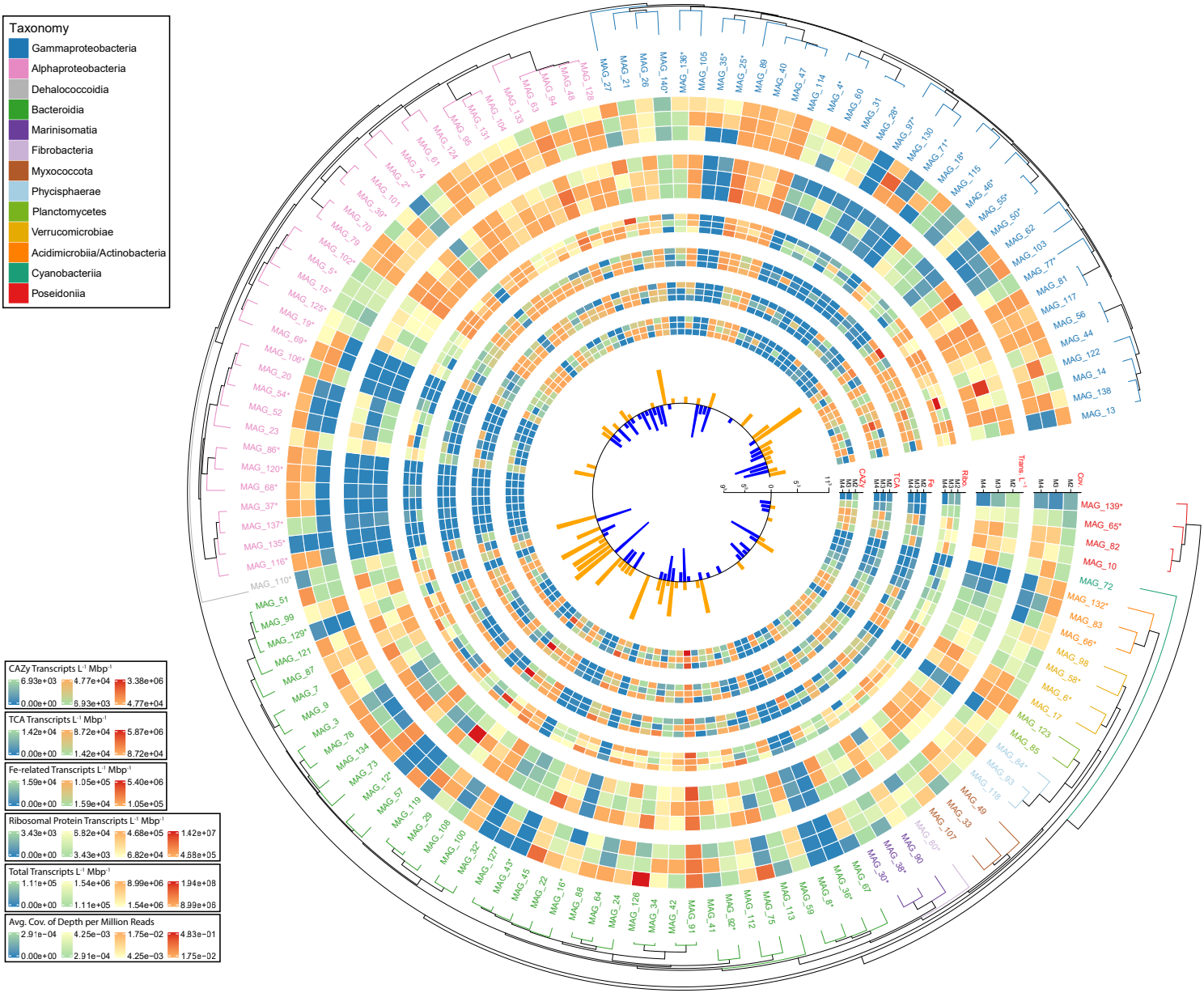

Taxonomy
Gammaproteobacteria
Alphaproteobacteria
Dehalococcoidia
Bacteroidia
Marinisomatia
Fibrobacteria
Myxococcota
Phycisphaerae
Planctomycetes
Verrucomicrobiae
Acidimicrobiia/Actinobacteria
Cyanobacteriia
Poseidoniia
CAZy Transcripts L-1 Mbp-1
6.93e+03 4.77e+04 3.38e+06
0.00e+00 6.93e+03 4.77e+04
TCA Transcripts L-1 Mbp-1
1.42e+04 8.72e+04 5.87e+06
0.00e+00 1.42e+04 8.72e+04
Fe-related Transcripts L-1 Mbp-1
1.59e+04 1.05e+05 5.40e+06
0.00e+00 1.59e+04 1.05e+05
Ribosomal Protein Transcripts L-1 Mbp-1
3.43e+03 6.82e+04 4.68e+05 1.42e+07
0.00e+00 3.43e+03 6.82e+04 4.68e+05
Total Transcripts L-1 Mbp-1
1.11e+05 1.54e+06 8.99e+06 1.94e+08
0.00e+00 1.11e+05 1.54e+06 8.99e+06
Avg. Cov. of Depth per Million Reads
2.91e−04 4.25e−03 1.75e−02 4.83e−01
0.00e+00 2.91e−04 4.25e−03 1.75e−02
CAZy
TCA
Fe
Ribo
Trans. L-1
Cov.

Fig. 2

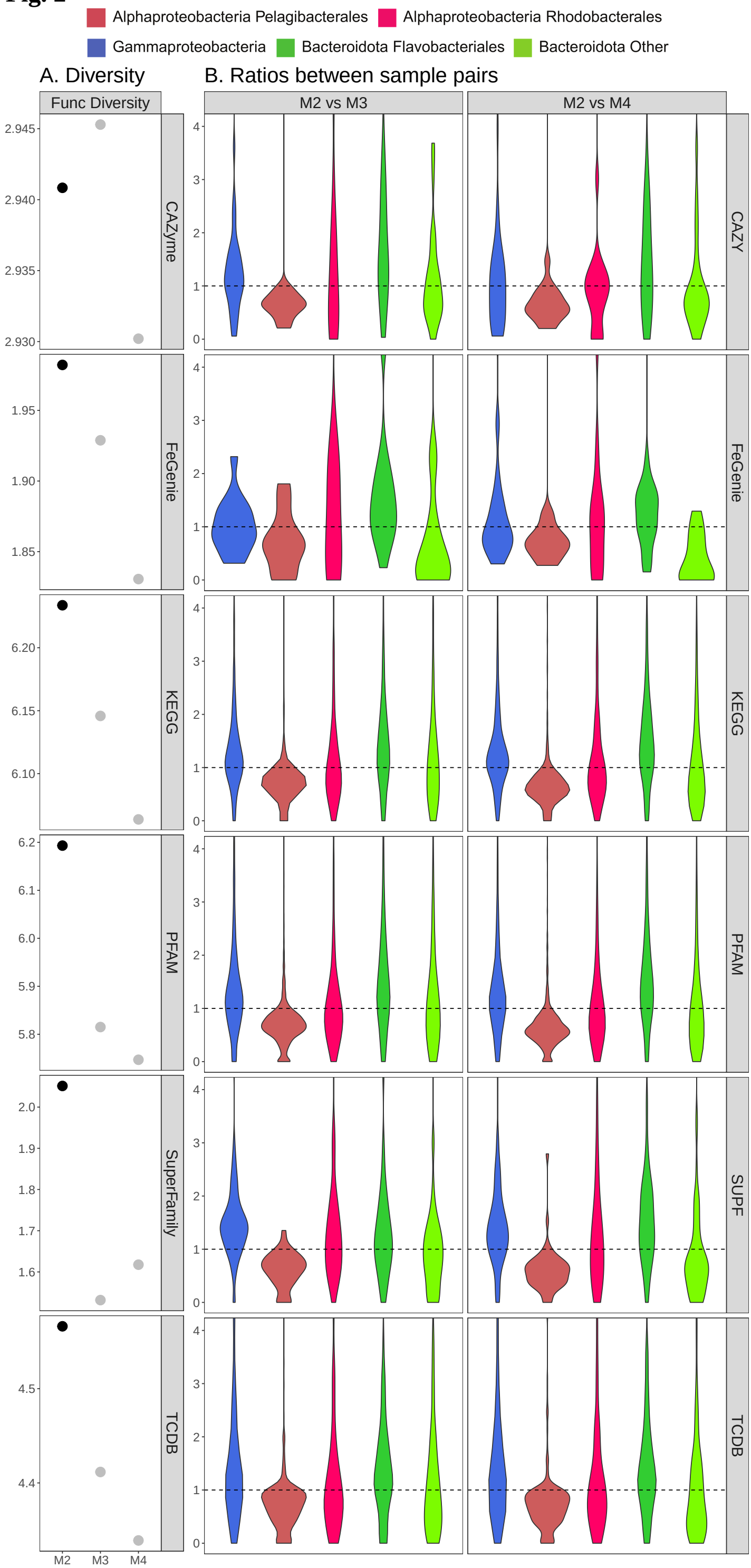

Alphaproteobacteria Pelagibacterales
Alphaproteobacteria Rhodobacterales
Gammaproteobacteria
Bacteroidota Flavobacteriales
Bacteroidota Other
A. Diversity
B. Ratios between sample pairs
Func Diversity
M2 vs M3
M2 vs M4
CAZyme
CAZY
FeGenie
KEGG
PFAM
SuperFamily
SUPF
TCDB
M2
M3
M4

Fig. 3

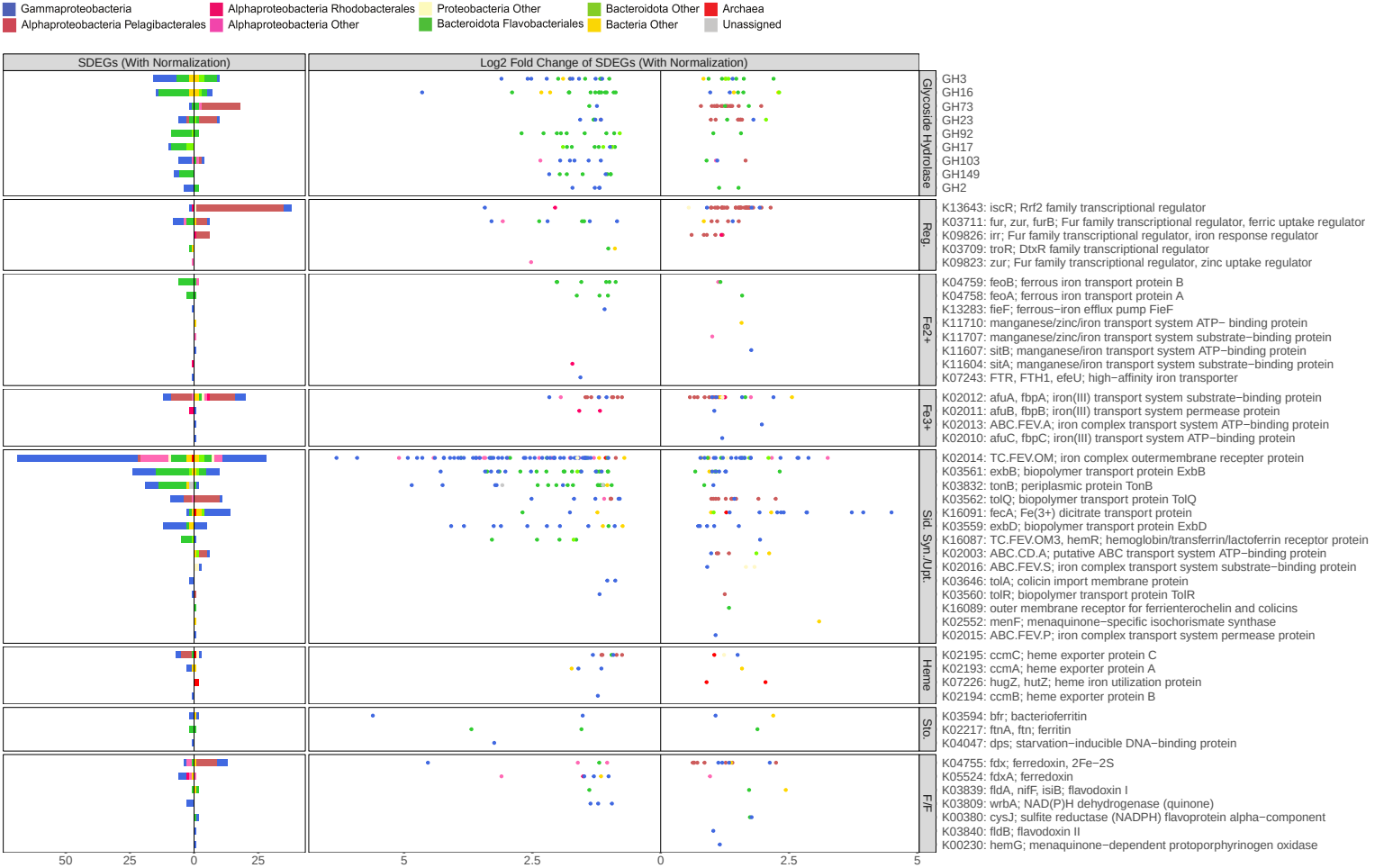

Gammaproteobacteria
Alphaproteobacteria Pelagibacterales
Alphaproteobacteria Rhodobacterales
Alphaproteobacteria Other
Proteobacteria Other
Bacteroidota Flavobacteriales
Bacteroidota Other
Bacteria Other
Archaea
Unassigned
SDEGs (With Normalization)
Log2 Fold Change of SDEGs (With Normalization)
Glycoside Hydrolase
GH3
GH16
GH73
GH23
GH92
GH17
GH103
GH149
GH2
Reg.
K13643: iscR; Rrf2 family transcriptional regulator
K03711: fur, zur, furB; Fur family transcriptional regulator, ferric uptake regulator
K09826: irr; Fur family transcriptional regulator, iron response regulator
K03709: troR; DtxR family transcriptional regulator
K09823: zur; Fur family transcriptional regulator, zinc uptake regulator
Fe2+
K04759: feoB; ferrous iron transport protein B
K04758: feoA; ferrous iron transport protein A
K13283: fieF; ferrous−iron efflux pump FieF
K11710: manganese/zinc/iron transport system ATP− binding protein
K11707: manganese/zinc/iron transport system substrate−binding protein
K11607: sitB; manganese/iron transport system ATP−binding protein
K11604: sitA; manganese/iron transport system substrate−binding protein
K07243: FTR, FTH1, efeU; high−affinity iron transporter
Fe3+
K02012: afuA, fbpA; iron(III) transport system substrate−binding protein
K02011: afuB, fbpB; iron(III) transport system permease protein
K02013: ABC.FEV.A; iron complex transport system ATP−binding protein
K02010: afuC, fbpC; iron(III) transport system ATP−binding protein
Std. Syn./Upt.
K02014: TC.FEV.OM; iron complex outermembrane recepter protein
K03561: exbB; biopolymer transport protein ExbB
K03832: tonB; periplasmic protein TonB
K03562: tolQ; biopolymer transport protein TolQ
K16091: fecA; Fe(3+) dicitrate transport protein
K03559: exbD; biopolymer transport protein ExbD
K16087: TC.FEV.OM3, hemR; hemoglobin/transferrin/lactoferrin receptor protein
K02003: ABC.CD.A; putative ABC transport system ATP−binding protein
K02016: ABC.FEV.S; iron complex transport system substrate−binding protein
K03646: tolA; colicin import membrane protein
K03560: tolR; biopolymer transport protein TolR
K16089: outer membrane receptor for ferrienterochelin and colicins
K02552: menF; menaquinone−specific isochorismate synthase
K02015: ABC.FEV.P; iron complex transport system permease protein
Heme
K02195: ccmC; heme exporter protein C
K02193: ccmA; heme exporter protein A
K07226: hugZ, hutZ; heme iron utilization protein
K02194: ccmB; heme exporter protein B
Sto.
K03594: bfr; bacterioferritin
K02217: ftnA, ftn; ferritin
K04047: dps; starvation−inducible DNA−binding protein
F/F
K04755: fdx; ferredoxin, 2Fe−2S
K05524: fdxA; ferredoxin
K03839: fldA, nifF, isiB; flavodoxin I
K03809: wrbA; NAD(P)H dehydrogenase (quinone)
K00380: cysJ; sulfite reductase (NADPH) flavoprotein alpha−component
K03840: fldB; flavodoxin II
K00230: hemG; menaquinone−dependent protoporphyrinogen oxidase
50
25
0
25
5
2.5
0
2.5
5

Fig. 4

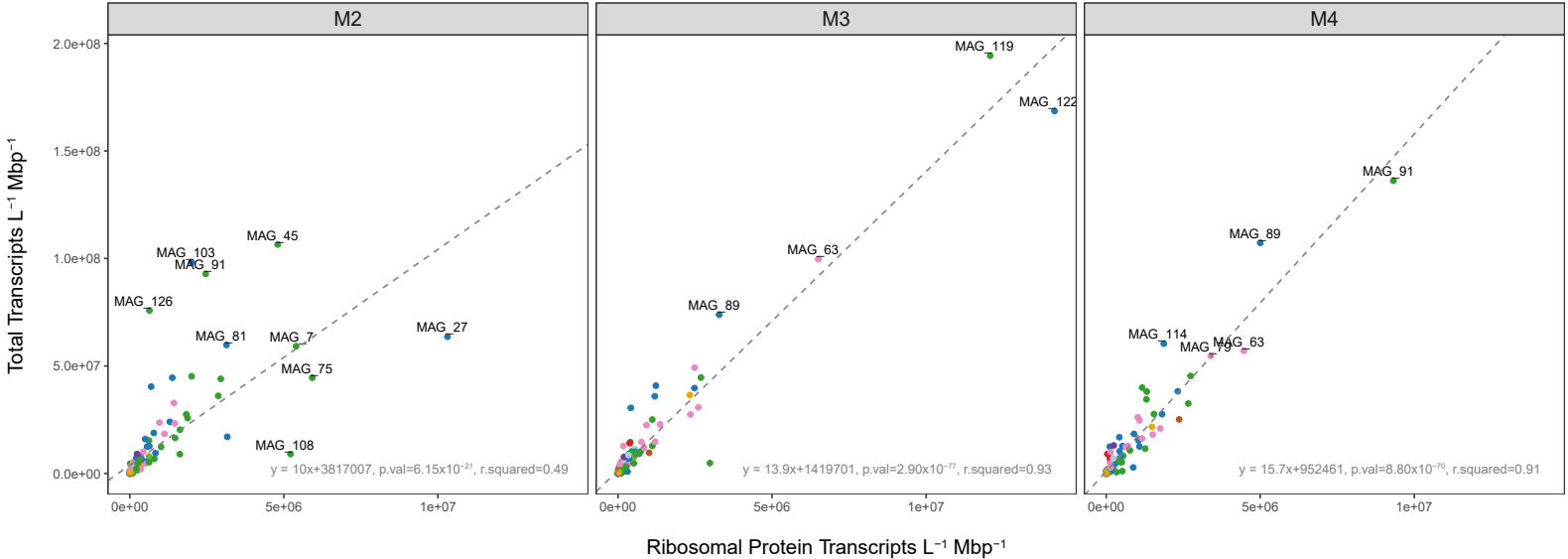
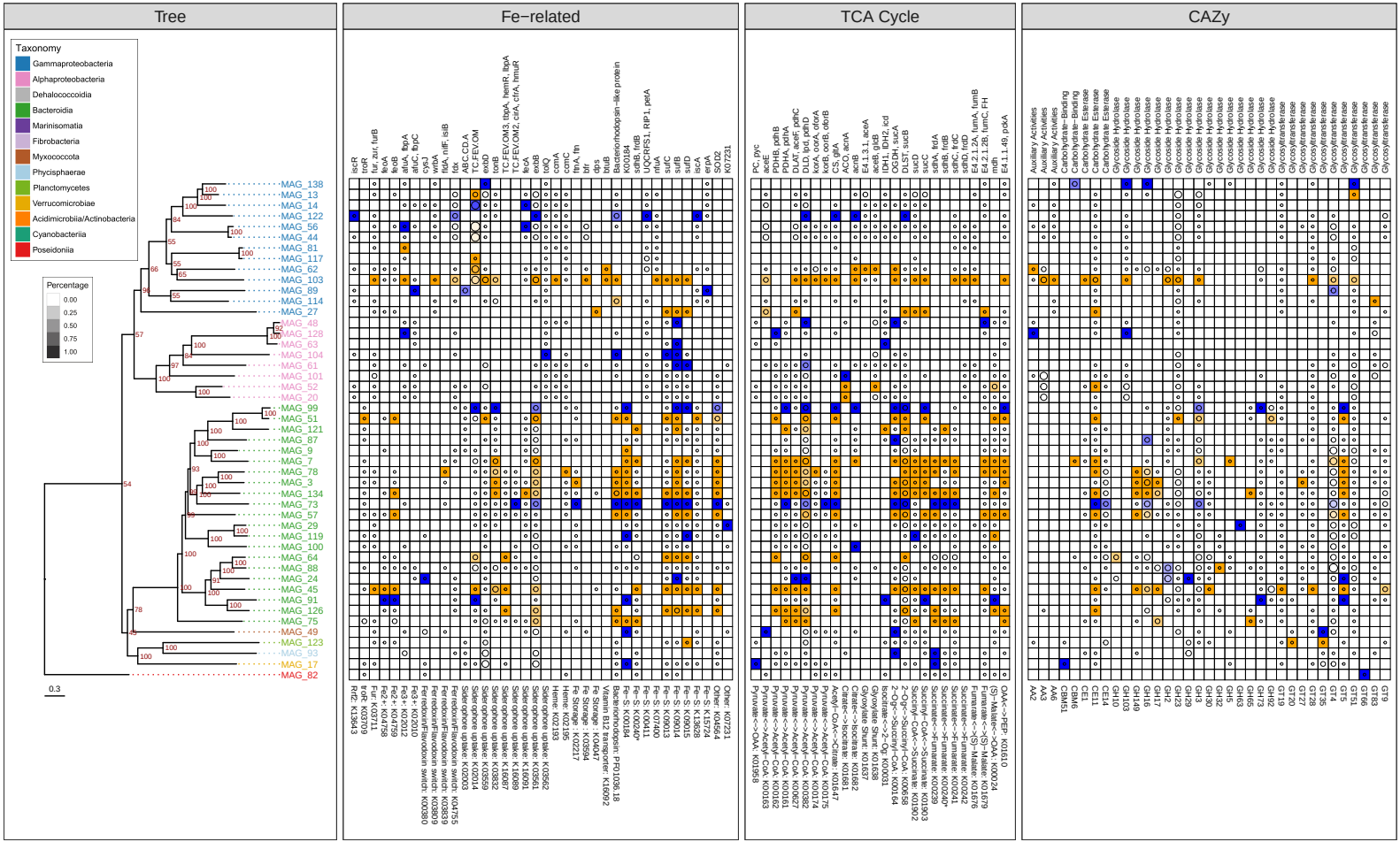

M2
M3
M4
Total Transcripts L⁻¹ Mbp⁻¹
Ribosomal Protein Transcripts L⁻¹ Mbp⁻¹
2.0e+08
1.5e+08
1.0e+08
5.0e+07
0.0e+00
0e+00
5e+06
1e+07
MAG_45
MAG_103
MAG_91
MAG_126
MAG_81
MAG_7
MAG_27
MAG_75
MAG_108
y = 10x+3817007, p.val=6.15x10⁻²¹, r.squared=0.49
MAG_119
MAG_122
MAG_63
MAG_89
y = 13.9x+1419701, p.val=2.90x10⁻⁷⁷, r.squared=0.93
MAG_91
MAG_89
MAG_114
MAG_79
MAG_63
y = 15.7x+952461, p.val=8.80x10⁻⁷⁰, r.squared=0.91

Fig. 5


Tree
Taxonomy
Gammaproteobacteria
Alphaproteobacteria
Dehaloccocoidia
Bacteroidia
Marinisomatia
Fibrobacteria
Myxococcota
Phycisphaerae
Planctomycetes
Verrucomicrobiae
Acidimicrobiia/Actinobacteria
Cyanobacteriia
Poseidoniia
Percentage
0.00
0.25
0.50
0.75
1.00
0.3
Fe−related
TCA Cycle
CAZy