

ARTICLE



<https://doi.org/10.1038/s41467-021-22409-4>

OPEN

# Metabolic pathways inferred from a bacterial marker gene illuminate ecological changes across South Pacific frontal boundaries

Eric J. Raes <sup>1,2✉</sup>, Kristen Karsh<sup>1</sup>, Swan L. S. Sow <sup>1,3,4</sup>, Martin Ostrowski <sup>5</sup>, Mark V. Brown<sup>6</sup>, Jodie van de Kamp <sup>1</sup>, Rita M. Franco-Santos <sup>3</sup>, Levente Bodrossy <sup>1</sup> & Anya M. Waite<sup>2</sup>

Global oceanographic monitoring initiatives originally measured abiotic essential ocean variables but are currently incorporating biological and metagenomic sampling programs. There is, however, a large knowledge gap on how to infer bacterial functions, the information sought by biogeochemists, ecologists, and modelers, from the bacterial taxonomic information (produced by bacterial marker gene surveys). Here, we provide a correlative understanding of how a bacterial marker gene (16S rRNA) can be used to infer latitudinal trends for metabolic pathways in global monitoring campaigns. From a transect spanning 7000 km in the South Pacific Ocean we infer ten metabolic pathways from 16S rRNA gene sequences and 11 corresponding metagenome samples, which relate to metabolic processes of primary productivity, temperature-regulated thermodynamic effects, coping strategies for nutrient limitation, energy metabolism, and organic matter degradation. This study demonstrates that low-cost, high-throughput bacterial marker gene data, can be used to infer shifts in the metabolic strategies at the community scale.

<sup>1</sup>CSIRO Oceans and Atmosphere, Hobart, TAS, Australia. <sup>2</sup>Ocean Frontier Institute and Department of Oceanography, Dalhousie University, Halifax, NS, Canada. <sup>3</sup>Institute for Marine and Antarctic Studies, University of Tasmania, Hobart, TAS, Australia. <sup>4</sup>NIOZ Royal Netherlands Institute for Sea Research, Department of Marine Microbiology and Biogeochemistry, Den Burg, The Netherlands. <sup>5</sup>Climate Change Cluster, University of Technology Sydney, Sydney, NSW, Australia. <sup>6</sup>School of Environmental and Life Sciences, The University of Newcastle, Callaghan, NSW, Australia. ✉email: [eric.raes@dal.ca](mailto:eric.raes@dal.ca)

The oceans cover 71% of our planet, and the microbial organisms that inhabit them catalyze important ecosystem services (such as  $O_2$  production, C sequestration, and elemental cycling) which sustain life on Earth<sup>1</sup>. Because microbes execute key roles in numerous biogeochemical pathways, it is important to understand how the species and functional composition of these communities respond to environmental changes. On a geological timescale, a 13-million-year-long nanoplankton abundance time series analysis suggested that ecological functions are more important to community resilience and biochemical functions than species richness<sup>2</sup>. Mapping microbial biogeography in relation to abiotic and biotic parameters therefore merits intensive investigation<sup>3</sup>. A better understanding of the ocean genome<sup>4</sup> will allow society to better preserve and utilize the vast genetic diversity in marine ecosystems. Global oceanographic initiatives such as the GO-SHIP<sup>5</sup> and the GEOTRACES<sup>6</sup> programs originally surveyed only abiotic essential ocean variables such as temperature, salinity, and dissolved inorganic trace metals. These initiatives have recently started to include biological essential ocean variables, such as marker gene sequencing and metagenomics, in their sampling programs, potentially enabling scientists to fulfill critical knowledge gaps on how microbial diversity relates to the microbial community metabolic potential. The combination of these two data sets with measurements on microbial processes, such as primary productivity, will provide insights into how local environmental conditions modulate the relationship between functional microbial diversity and productivity across frontal zones and within ocean provinces.

The highly conserved 16S rRNA gene is commonly sequenced for prokaryotic identification and microbial community profiling; an analysis that has been employed to study many biomes around the world<sup>3,7–9</sup>. 16S rRNA gene sequencing, however, does not provide direct information on the metabolic capacity of the microbial communities studied; this information can be obtained from shotgun metagenomics and genome sequencing. Because metagenome assembly, binning, and taxonomic profiling is a complex process and requires a higher level of computational resources than 16S rRNA gene analyses<sup>10</sup>; and because the amount of spatial and temporal 16S rRNA gene data readily available vastly surpasses that of shotgun data<sup>11,12</sup>, evolutionary modelers have often inferred the potential metabolic profiles of microbial communities from sequence data of marker genes such as the 16S rRNA gene<sup>13</sup>. Although this is an indirect method to estimate microbial metabolic functions, it has been shown that 16S rRNA gene data analysis with the open source software Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt2<sup>13,14</sup>) results in predictions of metabolic microbial profiles that strongly agree (i.e., high Spearman correlations) with results from shotgun metagenomics<sup>14</sup>. The best predictions generated so far were for the human microbiome, followed by those for the ocean biome<sup>14,15</sup>. It should be mentioned, however, that high Spearman correlations should be carefully interpreted, as metabolic profiles are highly conserved in bacteria<sup>16</sup>.

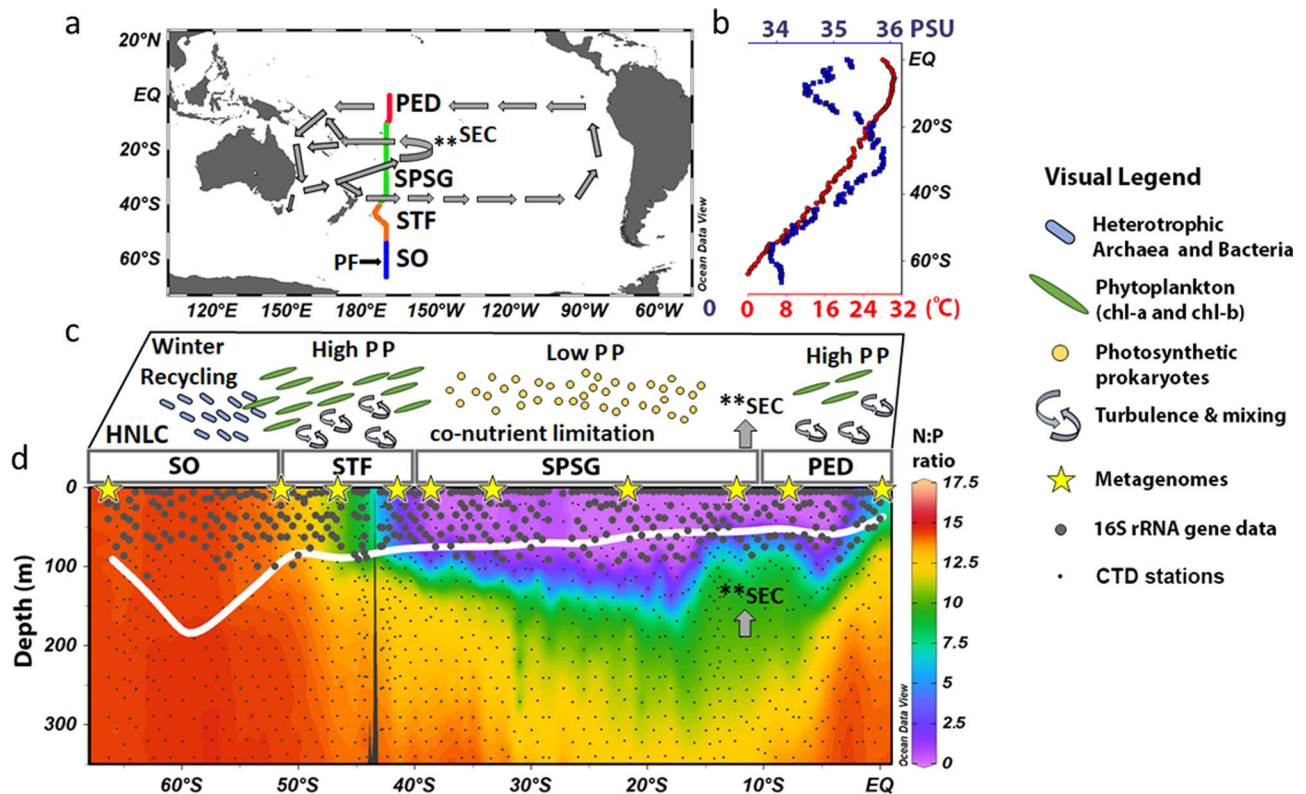
Our aim in this study is to test whether 16S rRNA gene-based metabolic reconstructions generated by the software PICRUSt2 can predict broad-scale latitudinal patterns in microbial metabolic capacity which agree with our current mechanistic understanding of functional microbial biogeography, both within and across ecological provinces in the South Pacific Ocean<sup>17,18</sup>. More specifically, by contrasting biomass estimators (concentrations of various photosynthetic pigments and of particulate organic carbon and nitrogen (PON)), primary productivity, and N assimilation measurements with existing quantitative and qualitative data from oceanographic literature, we test the validity of the following hypotheses:

- H1: Primary productivity will be positively correlated with pathways associated with  $CO_2$ -fixation. Frontal zones, which stimulate primary productivity<sup>19</sup>, should display a higher relative abundance of pathways associated with  $CO_2$ -fixation and energy metabolism than less productive regions.
- H2: Cell metabolism pathways are positively affected by thermodynamics<sup>20</sup>—as temperature increases, more kinetic energy (adenosine triphosphate (ATP)) is required to maintain the cellular machinery and fuel metabolic processes. Therefore, it can be expected that an increase in temperatures will lead to an increase in the relative abundance of cell biosystems machinery (cell structure and cell wall biosynthesis pathways).
- H3: Pathways which reflect microbial strategies for coping with trace metal and macro-nutrient limitations will show latitudinal trends corresponding to element-specific abundances (i.e., high relative abundances of cofactor and secondary metabolite biosynthesis pathways due to iron limitation in the Southern Ocean (SO)<sup>21</sup> and to co-nutrient stress in the oligotrophic gyre<sup>22</sup>).
- H4: The high availability of nutrients and seasonally defined production of organic matter in the SO and in the Subtropical Frontal Zone (STF) will result in higher relative abundances of pathways associated with energy production, such as lipid and carbohydrate biosyntheses, in these ocean provinces<sup>23,24</sup>.
- H5: Higher rates of bacterial degradation of particulate and dissolved organic material in the SO<sup>25,26</sup> in comparison to the other zones should result in the former having a greater diversity (and relative abundance) of degradation pathways, including the presence of more complex compound degradation pathways.

In this study we compare these hypotheses and infer metabolic pathways with measured physico-biochemical parameters, 11 corresponding metagenomes and show evidence supporting that microbial functional diversity follows trends within and between oceanographic provinces as expected from existing quantitative and qualitative oceanographic literature. Such analyses may provide insight into the drivers of ecological changes and, overall, into the effects of biodiversity on marine ecosystem functioning.

## Results and discussion

**Hydrographic conditions.** This study was conducted during late autumn and early winter in 2016 along the decadal repeated P15S GO-SHIP transect, which runs from the Antarctic ice edge to the equator along the 170° W meridional in the South Pacific Ocean (Fig. 1a). Sea surface temperatures along the transect gradually increased from  $-1.5^\circ C$  at 66° S to  $30.4^\circ C$  at 5.5° S, and then decreased slightly, due to the equatorial upwelling, to  $28.1^\circ C$  at the equator (Fig. 1a, b). Surface salinity was lowest in the SO (33.80–34.30), increasing north of the Polar Front (60° S) to a maximum of 35.87 at 30° S, then decreasing north of 30° S to 34.50 at 10° S, and then increasing to 35.25 at the equator (Fig. 1a, b). Dissolved  $NO_3^-$ , Si, and  $PO_4^{3-}$  concentrations covaried above the mixed layer depth (MLD), and were closely linked to the major oceanographic provinces<sup>17</sup>.  $NO_3^-$  concentrations above the MLD were  $>16\ \mu mol\ l^{-1}$  in the SO; between 1 and  $16\ \mu mol\ l^{-1}$  in the STF;  $\leq 0.05\ \mu mol\ l^{-1}$  in the South Pacific Subtropical Gyre Province (SPSG); and  $\leq 2\ \mu mol\ l^{-1}$  in the Pacific Equatorial Divergence Province (PED) (Supplementary Fig. 1). Differences in  $NO_3^-:PO_4^{3-}$  ratios above the MLD showed differences that illustrated the four distinct oceanographic provinces, averaging ( $\pm$ standard deviation (SD))  $14.24 \pm 0.25$  in the SO,  $9.76 \pm 2.3$  in the STF,  $0.43 \pm 0.87$  in the oligotrophic SPSG, and  $5.6 \pm 1.4$  in the PED (Fig. 1c).



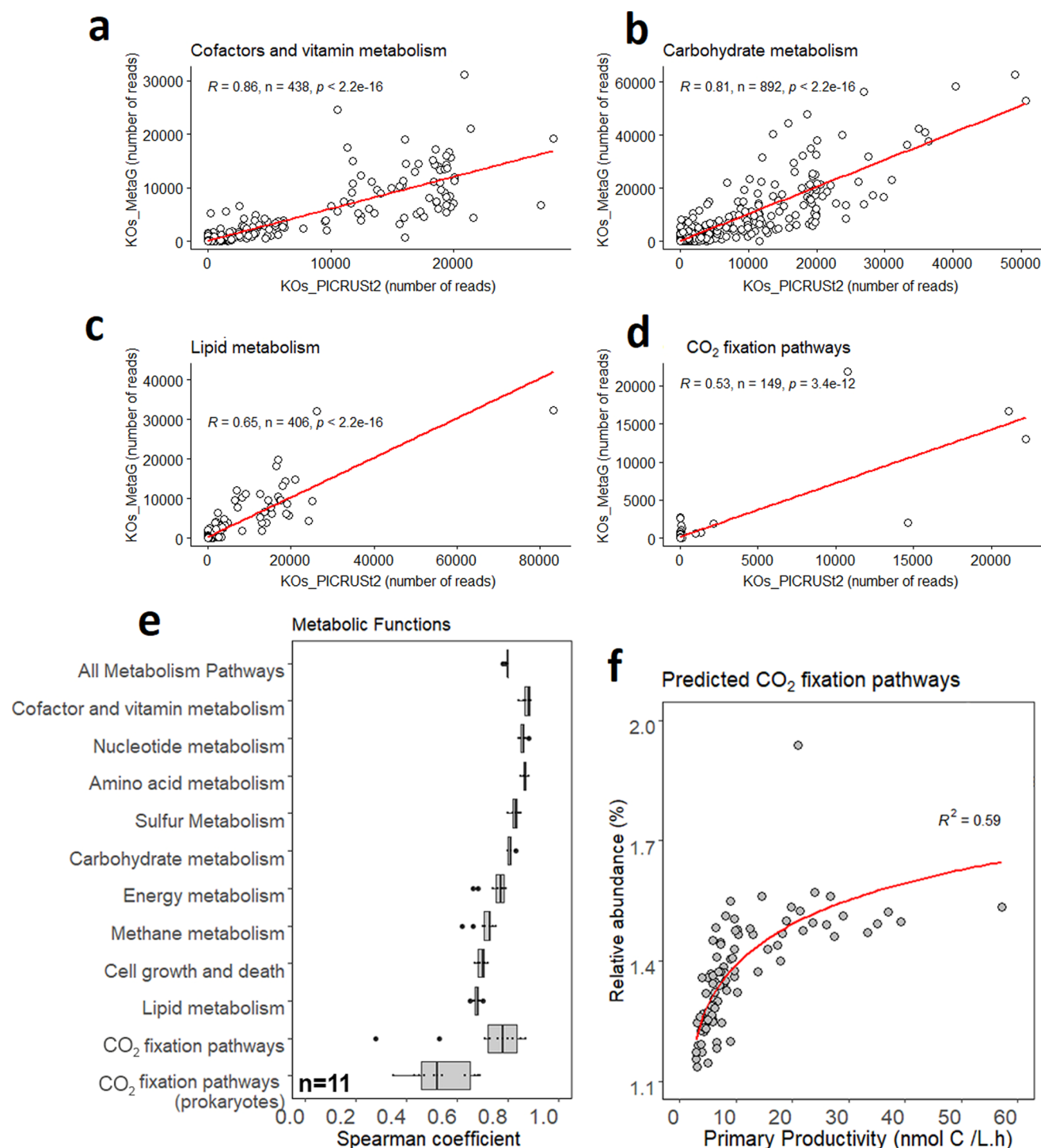
**Fig. 1** Oceanographic transect with abiotic water column data and conceptual mechanistic understanding of the microbial biogeography. **a** The GO-SHIP P15S transect along the 170° W meridian in the South Pacific Ocean. Gray arrows indicate the South Equatorial Current (SEC) within the upper ocean, and include a westward returning branch between 20° and 10° S (denoted by \*\*SEC<sup>69,70</sup>). The oceanographic provinces<sup>54</sup> are color coded: the Southern Ocean (SO; blue), the Subtropical Frontal Zone (STF; orange), the South Pacific Subtropical Gyre Province (SPSG; green), the Pacific Equatorial Divergence Province (PED; red), the Polar Front (PF) is in addition highlighted by a black arrow. **b** Profiles for SST (red; bottom x-axis) and salinity (blue; top x-axis). **c** Conceptual mechanistic understanding of relative changes in the functional prokaryotic and microbial-eukaryotic biogeography. Blue rods represent heterotrophic archaea and bacteria, which recycle in winter the organic matter produced in the summer and autumn months in the high nutrient low chlorophyll (HNLC) region of the SO. High primary productivity (PP) driven by phytoplankton rich in chlorophyll (green discs) is expected in the STF due to turbulence and mixing (curved arrows). The oligotrophic SPSG is characterized by low PP and nutrient co-limitation, as well as by higher abundances of photosynthetic prokaryotes (yellow circles). The westward returning branch of the SEC is indicated with the gray arrow in **c, d** and can be the source of the increase in N:P ratios at 100 m depth. Equatorial upwelling and mixing in the PED results in an increase of the N:P ratio at the surface and, thus, in increased PP and chlorophyll concentrations. **d** Latitudinal plot of N:P ratios from the surface to 350 m depth. The thick white line represents the MLD; 16S rRNA sampling stations are shown in large gray circles; shotgun metagenome samples are represented by yellow stars; and CTD stations appear as small gray dots.

### PICRUSt2 predictions, Shotgun metagenomes, and PP data.

Ninety-two percentage of the 16S rRNA gene sequencing data were used to infer Kyoto Encyclopedia of Genes and Genomes (KEGG) Orthologs (KO) using PICRUSt2. On average, 30% of the shotgun metagenome sequences (MGS) could be assigned to a KO (Supplementary Table 1), and  $1.4 \pm 0.9\%$  (average  $\pm$  SD) of the total reads were eukaryotic. A comparative analysis between the KO predictions from PICRUSt2 and the KOs profiled from corresponding MGS was performed in a similar way as presented by Douglas et al.<sup>14</sup> to validate the PICRUSt2 predicted metabolic pathways. Spearman correlation coefficients (shown as R in Fig. 2a–d) were calculated between the KO abundances predicted by PICRUSt2 and the KO abundances profiled from MGS for all individual samples and for each distinct metabolic function. The Spearman correlation coefficients, which represent the similarity in rank ordering of KO abundances between the predicted and observed data, are summarized for the 11 samples and for each metabolic function in Fig. 2e (pathway correlations for all samples are shown in the Supplementary methods). Overall, the correlation coefficients showed strong associations between the ranks of the predicted KO and the MGS KO abundances, with the strongest correlation coefficients recorded for cofactor and

vitamin biosynthesis pathways. Although the correlation coefficients for CO<sub>2</sub>-fixation pathways were the weakest, these PICRUSt2 predicted pathways were positively correlated with PP (Fig. 2f). Similar to the work of Agrawal et al.<sup>15</sup>, who validated their PICRUSt2 predictions with qPCR data, we herein show that PICRUSt2 predictions related to CO<sub>2</sub>-fixation pathways can be verified (independently from MGS data) with PP rates. We do note that the saturation of the relative abundance of predicted CO<sub>2</sub>-fixation pathways (at  $>20 \text{ nmol C l}^{-1} \text{ h}^{-1}$ ) requires further investigation when extrapolating marker gene predictions to rates.

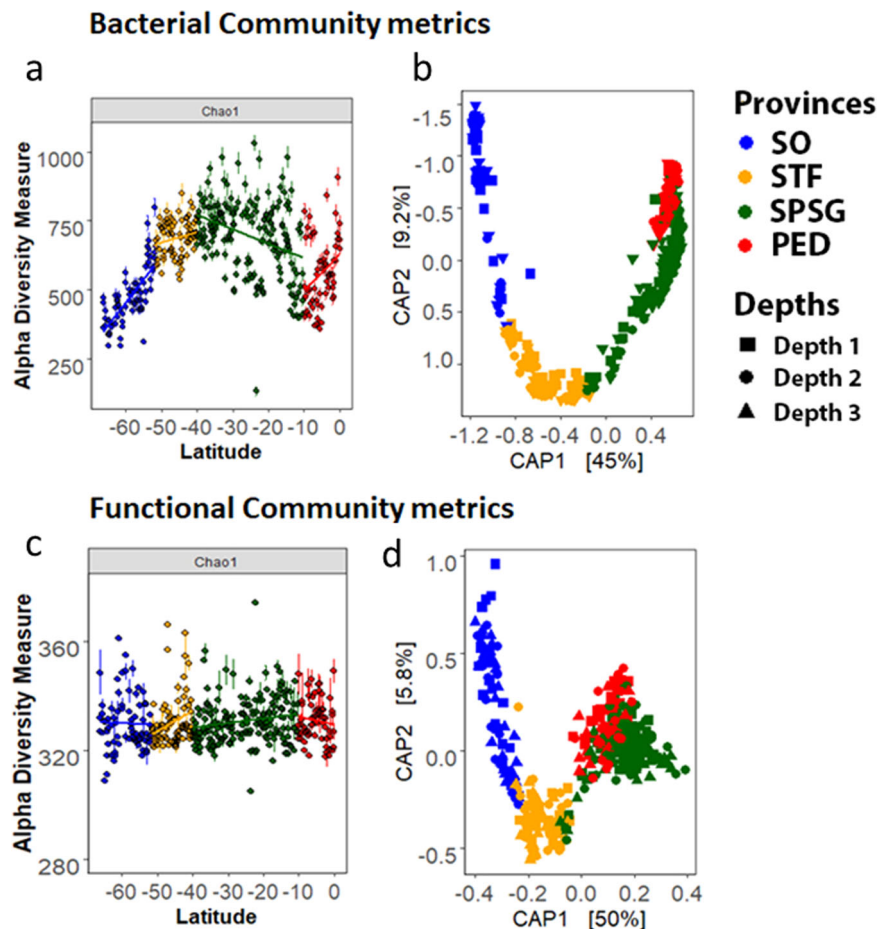
**Bacterial and metabolic community diversities.** Bacterial alpha diversity (Chao index) increased from the SO to the northern edge of the STF (between 66 and 40° S). Diversity decreased within the SPSG to then increase again north of 10° S (Fig. 3a). Canonical analysis of principal coordinates (CAP) plots and ANOSIM results of the sequence data revealed clear and significant differences in the bacterial communities (beta diversity) between all four oceanographic provinces (Fig. 3b and Supplementary Table 2). PICRUSt2 analyses of 387 DNA samples



**Fig. 2** Comparative analyses between the KEGG orthologs (KO) predictions from PICRUSt2 and the KOs profiled from corresponding shotgun metagenomes (MGS). Two-sided Spearman correlation coefficients (indicated as  $R$  on plots **a–d** for the station at 66 °S) between the number of assigned KO reads are shown for the following pathways: **a** cofactor and vitamin metabolism; **b** carbohydrate metabolism; **c** lipid metabolism; and **d** CO<sub>2</sub>-fixation pathways. **e** Summary of the Spearman correlation coefficients between the predicted KO and MGS KO abundances for the 11 samples and for each metabolic function. Boxplots in **e** show median values (center line); the upper and lower quartiles are shown by box hinges; the whiskers present 1.5× interquartile range and the points outside the whiskers are plotted individually. Predicted PICRUSt2 pathways for CO<sub>2</sub>-fixation and primary productivity correlation with an exponential order (red curve; **f**). On average of  $1.4 \pm 0.9\%$  ( $\pm$ SD) of the metagenome reads were Eukaryotic.

resulted in the inference of 400 MetaCyc pathways (the inferred functional diversity). No significant differences were observed in the alpha diversity of MetaCyc pathways diversities between the four oceanographic provinces (two-sided Wilcoxon test  $p > 0.05$ ;

Fig. 3c). CAP plots and ANOSIM results for the MetaCyc pathways, however, revealed significant differences in the composition of the metabolic potential between the four oceanographic provinces (Fig. 3d and Supplementary Table 2).



**Fig. 3 Bacterial community (OTUs clustered at 97% similarity; upper panels) and functional community (MetaCyc pathways) metrics.** Alpha diversity shown as Chao diversity (**a, c** with standard errors as defined for the Chao1 model for estimating richness), and beta diversity shown as CAP plots (**b, d**). Linear models were fitted through the data points on **a, c** for each of the four oceanographic provinces, which appear in the figure color coded. Data for CAP plots were rarefied and square-root transformed, and also show color-coded oceanographic provinces. The three different sampling depths are indicated by symbols: squares for depth 1 (13–36.7 m), circles for depth 2 (18.6–85.5 m), and triangles for depth 3 (39.9–185 m). ANOSIM outputs are shown in Supplementary Table 2.

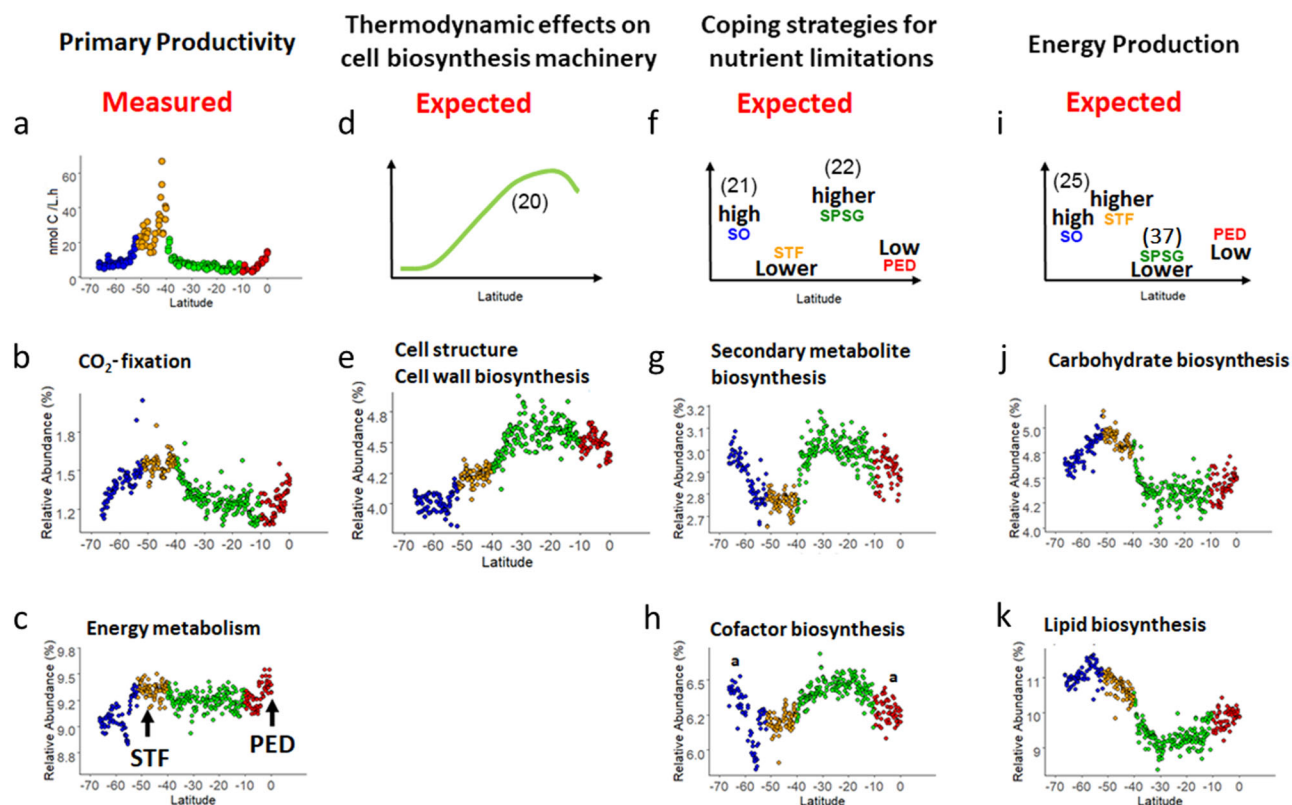
**Metabolic community composition.** The 400 inferred MetaCyc pathways could be collapsed into 41 secondary superclasses (Supplementary Data 1). Across the transect, the sum of the relative abundance of ten of these superclasses accounted approximately for 75–80% of the total relative abundance. These metabolic functions covered (in descending order of relative abundance) pathways of (1) amino acid and (2) nucleotide biosynthesis; (3) energy metabolism; (4) lipid and (5) carbohydrates biosynthesis; (6) cell structure and cell wall biosynthesis; (7) cofactor and (8) secondary metabolite biosynthesis; (9) vitamin biosynthesis; and (10) fermentation. CO<sub>2</sub>-fixation pathways accounted for <2% of the total relative abundance but are singled out to illustrate patterns in the data, as are a few degradation pathways discussed below. Trends in the relative abundance of these metabolic functions (see Fig. 4 for functions 3–8 and Supplementary Fig. 2 for the other functions) are described in more detail in the following paragraphs in relation to our hypotheses.

**Primary productivity shapes ecological functions of the bacterial community (H1).** Our first hypothesis, based upon the fact that PP is stimulated by increases in nutrient concentrations in frontal zones<sup>19</sup>, proposed that CO<sub>2</sub>-fixation pathways in frontal systems should be positively related to autotrophic production. Raes et al.<sup>18</sup> showed that, along the P15S GO-SHIP transect, the SO and the SPSG were areas of low PP, whereas the STF and the

PED had relatively high PP (Figs. 1b and 4a). The authors also observed an important trend along this transect: a switch from net autotrophy (i.e., high CO<sub>2</sub>-fixation) in the STF to heterotrophy (i.e., high nitrification rates and degradation of organic matter) when light availability is reduced during the winter months in the SO<sup>18</sup> (more information under the subheading H3 and H4—Energy production and Degradation).

Pathways associated with CO<sub>2</sub>-fixation increased from 66° S to the northern edge of the STF at 40° S (35%) and from 10° S to the equator (by 30%), and decreased (by 36%) northwards within the SPSG (all oceanographic provinces were significantly different from one another; two-sided Wilcoxon tests,  $p < 0.05$ ; Fig. 4b). Energy metabolism and nucleotide biosynthesis pathways were positively related to the frontal zones, with both the STF and the equatorial upwelling showing significantly higher relative abundances than the SO and SPSG (two-sided Wilcoxon tests,  $p < 0.05$ ; Fig. 4c and Supplementary Figs. 2B and 3).

Metabolic pathways associated with CO<sub>2</sub>-fixation, energy metabolism, and nucleotide biosynthesis, thus, showed similar latitudinal trends, which in turn were well aligned with basin-wide variations in PP (Fig. 4a–c). These trends were also in agreement with the distinction between productive (the STF and the PED) and less productive (the SO and the SPSG) oceanographic provinces (Fig. 1b). In our model predictions, the concentrations of nanoplankton and 19'-hexanoyloxyfucoxanthin (a diagnostic



**Fig. 4** Latitudinal trends for metabolic inferences from a 16S rRNA marker gene survey. Columns illustrate (from left to right) hypotheses and their associated secondary superclass (functional) pathways: H1 (PP), H2 (thermodynamic effects on cell biosystems machinery), H3 (coping strategies for nutrient limitation), and H4 (energy production). Graphs in the upper row illustrate measured data for PP (**a**) and expected values (according to the literature) for the processes relating to H2 (**d**<sup>20</sup>), H3 (**f**<sup>21</sup> and <sup>22</sup>), and H4 (**i**<sup>25</sup> and <sup>37</sup>). Graphs in the middle and lower rows show relative abundances of functional pathways predicted from 16S rRNA gene sequencing: CO<sub>2</sub>-fixation (**b**), energy metabolism (**c**), cell structure and cell wall biosynthesis (**e**), secondary metabolite biosynthesis (**g**), cofactor biosynthesis (**h**), carbohydrate biosynthesis (**j**), and lipid biosynthesis pathways (**k**). The oceanographic provinces are color coded: the Southern Ocean (SO; blue), the Subtropical Frontal Zone (STF; orange), the South Pacific Subtropical Gyre Province (SPSG; green) and the Pacific Equatorial Divergence Province (PED; red). Significant differences were observed between provinces unless otherwise indicated by **a**.

pigment for Prymnesiophytes) alone explained 52% of the latitudinal changes in PP, whereas total pigments and the concentrations of chl *b* and of nanoplankton explained ~50% of the latitudinal trends for CO<sub>2</sub>-fixation (Figs. 1b and 4a–c). Shifts in salinity and temperature characterize frontal zones, and both parameters were the main explanatory variables in predicting energy metabolism pathways. Besides supporting H1, these results also support previous findings that PP (autotrophic energy production) is a main driver for archaeal and bacterial richness across frontal boundaries in the South Pacific Ocean<sup>17</sup>.

**Thermodynamic effects on cell biosynthesis machinery (H2).** Our second hypothesis was formulated upon the concept that cell metabolism pathways are positively (though not solely) affected by thermodynamics<sup>20</sup>. As temperature increases, so does the energetic demand for an organism to maintain cellular machinery and metabolic processes (basal metabolism). In our study, metabolic pathways associated with cell biosystems machinery showed trends similar to those that would be expected with an increase in temperature. The relative abundances of cell structure and cell wall biosynthesis pathways were significantly different between all oceanographic provinces (two-sided Wilcoxon test,  $p < 0.05$ ). Temperature-wise, these abundances showed a quasi-linear increase of 20% from ~3 °C (55° S) in the SO to ~25 °C (22° S; Fig. 4e), though they decreased significantly thereafter (north of 22° S; two-sided Wilcoxon test,  $p < 0.05$ ).

Though thermodynamic regulation by increasing basal metabolic rates comes with an energetic cost, the resulting increase in cellular machinery will ultimately enable the organism a more active lifestyle and faster responses to environmental challenges<sup>20</sup>. It should also be noted that basal metabolism involves ATP-requiring pathways that are essential for cell survival, with a significant proportion of these pathways being protein synthesis, which maintains potential energy gradients across membranes<sup>20</sup>. Furthermore, increasing temperatures have also been positively linked to a higher production of cell wall proteins<sup>27</sup> and to a change in the composition of cell structural membranes<sup>28</sup>. The fact that temperature has a regulating effect on the cell biosystems machinery suggests that cellular physiological adaptations, including qualitative and quantitative variations in the cell wall proteins<sup>27,28</sup>, would result in a higher relative abundance of (inferred) cell structure and cell wall biosynthesis pathways. As expected, we recorded an increase in relative abundances of cell structure and cell wall biosynthesis pathways as temperatures increased northward along the transect.

Although temperatures predicted 56% of the variability in our regression models for these pathways, we should also note that the concentrations of NH<sub>4</sub><sup>+</sup> and of photosynthetic prokaryotes (picoplankton) together predicted 22% of those models (Fig. 4e and Supplementary Data 2). The steepest increase in relative abundances of cell structure and cell wall biosynthesis pathways was observed in the SPSG. This tropical oceanic province is characterized by relatively higher NH<sub>4</sub><sup>+</sup> uptake rates (in

comparison to other provinces), high picoplankton concentrations, and low concentrations of organic matter with relatively high  $\delta^{15}\text{N}$ —indications of a food web that is dominated by high turnover of organic material<sup>18</sup>. The turnover rate of organic material and its incorporation into new bacterial biomass has been shown to be regulated by bacterial particle colonization<sup>29</sup>, which itself has also been positively correlated to the presence of cell wall proteins<sup>30</sup>. We therefore suggest another possible (additional) explanation for the increase in cell structure and cell wall biosynthesis pathways: in a system with low concentrations of organic matter, the ability to display multiple particle adhesion strategies would be an added advantage in the competition for particle colonization.

**Coping strategies for nutrient limitations (H3).** Our third hypothesis postulated that the inferred metabolic predictions would result in latitudinal trends corresponding to element-specific abundances and, thus, reflect microbial strategies for coping with trace metal and macro-nutrient limitations. As expected, we observed bimodal latitudinal trends for the biosynthesis of secondary metabolites and cofactors, which are related to trace metal and co-nutrient limitations. The relative abundance of these pathways significantly decreased from 66° S toward the STF, increased north of the STF (with highest values in the SPSG), and then decreased again toward the equator (two-sided Wilcoxon tests,  $p < 0.05$ ; Fig. 4g, h and Supplementary Fig. 2A–C), such that they were lower in the more productive regions (STF and PED) than in the least productive regions. Amino acid and vitamin biosynthesis pathways showed the same significant trends as described for secondary metabolite and cofactor biosynthesis pathways (two-sided Wilcoxon test,  $p < 0.05$ ; Supplementary Fig. 2A, C).

The observed trends in the above-mentioned pathways agree with our conceptual understanding of these oceanic provinces, and with the available literature on the topic<sup>21,22</sup>. These results suggest that 16S rRNA data can potentially be used to track changes in how the microbial community copes with (essential) micro- and macro-nutrient limitation. Transport membrane proteins, much like secondary metabolites and cofactors, also play an important role in the transport of vital compounds, and the relative abundance of genes regulating their expression could also be higher in the nutrient-limited provinces. This is, however, still to be verified, as the functional output from PICRUST2 does not resolve specific transporter pathways.

**Seasonally defined energy production (H4).** Our fourth hypothesis suggested that pathways associated with energy production, such as lipid and carbohydrate biosynthesis, would show higher relative abundances in the SO and in the STF because of the seasonally defined production of organic matter in these provinces. Indeed, we observed higher relative abundances of carbohydrates and lipid biosynthesis pathways in the SO and STF in comparison to the other provinces. The relative abundance of carbohydrate biosynthesis pathways increased significantly (by 12%) from 66° S toward the southern edge of the STF (52° S; Fig. 4j), and the average value in the SPSG was significantly lower (15%) than that recorded at the northern edge of the STF (40° S). North of 10° S their relative abundance increased significantly once again (by 10%; all trends were significant; two-sided Wilcoxon tests,  $p < 0.05$ ; Fig. 4j). Similar trends were observed for changes in the relative abundance of lipid biosynthesis pathways (Fig. 4k), with values being highest in the SO, then declining by 22% from ~52° S until ~30° S, and subsequently increasing (by 10%) toward the equator (two-sided Wilcoxon tests,  $p < 0.05$ ; Fig. 4k).

Lipids and carbohydrates are structurally essential molecules and important energy sources<sup>31</sup>; and several prokaryotes are known to accumulate large amounts of lipid reserves, as these are advantageous for survival/against starvation and confer a strong evolutionary advantage<sup>32,33</sup>. The type (structural or storage) of lipid synthesized by microorganisms will depend on stress imposed on cells, the growth phase, and on environmental conditions such as nutrient availability in relation to (abundant) C sources, with storage lipids usually being accumulated under nutrient-limiting conditions<sup>32,34</sup>. The strong differences in light availability between the winter and summer in the SO and STF profoundly impact lipid trophodynamics<sup>35</sup>. During early winter the bacterial community will consume the remainder of the autumn production, likely using it to fuel (structural and/or storage) lipid synthesis. Our study was conducted during early winter in the SO and the STF, which could explain the high relative abundance of lipid biosynthesis pathways in the bacterial community at these latitudes (Fig. 3k, l).

The relative abundances likely also reflect bacterial adaptation to low temperatures (changes in phospholipid fatty acid composition to maintain membrane fluidity<sup>23</sup>). The low temperatures in the SO, which result in slow cell growth and division, coupled with a high availability of C, N, and  $\text{PO}_4^{3-}$ , likely enable bacteria in this region to synthesize both structural and storage lipids. Any decreasing trends in the relative abundance of lipid biosynthesis pathways seen in the STF (in relation to the SO) would likely be due to increasing metabolic activities (due to increasing temperatures) and, thus, decreasing synthesis of lipids for storage; and to a decreasing availability of  $\text{PO}_4^{3-}$  for phospholipid biosynthesis. North of the STF in the tropical region the seasons are not as distinguishable, and the conditions are oligotrophic (Fig. 1b, c). Although laboratory studies have shown that prokaryotes will readily synthesize storage lipids under N-limiting conditions, this is only true when a C source is abundant. The low availability of organic C, regardless of nutrient input, in the SPSG thus likely explains the rapidly declining relative abundance of lipid and carbohydrate biosynthesis pathways observed in the region (Fig. 4j, k).

The bioavailability of  $\text{PO}_4^{3-}$  might also explain the declining trend in the relative abundance of lipid biosynthesis pathways<sup>36</sup>. It has been shown that, in  $\text{PO}_4^{3-}$ -deficient environments (such as the SPSG; Fig. 1c), heterotrophic bacteria and photosynthetic prokaryotes (picocyanobacteria) are able to engage in lipid remodeling (substituting phospholipids with non-phosphorus lipids, such as sulfolipids or glycolipids<sup>37</sup>), a strategy which increases their survival at an evolutionary scale in oligotrophic areas of the ocean<sup>37,38</sup>. As this lipid remodeling is expressed at a community level, the shift in trends of the metabolic pathways might give insight regarding how bacterial communities cope with  $\text{PO}_4^{3-}$ -limitation in the South Pacific Ocean. Slightly higher relative abundances in the PED than in the SPSG would reflect an increase in  $\text{PO}_4^{3-}$  availability in this region (phosphorus-associated pathways were also observed predominantly in the PED).

**Degradation pathways (H5).** Our last hypothesis suggested that we would detect more degradation-type pathways during the onset of winter in the SO, given that it has been previously shown that the SO is a region of high nutrient recycling rates and breakdown of organic matter in winter (e.g., measurements of high nitrification rates<sup>18</sup>). Although the secondary superclasses constituted ~75–80% of the total abundance of pathways, other superclasses were also of importance when distinguishing (significant differences in) the main bacterial processes occurring within and between the ocean provinces. The results from our

indicator analysis (see Supplementary methods) revealed that degradation pathways (proxies for heterotrophy; Supplementary Fig. 4B–H) were characteristic of the SO, even though they occurred in low relative abundances. More specifically, we observed a greater variety of degradation pathways and the presence of more complex compound degradation pathways, such as aromatic compound and amino acid degradation (which decreased significantly between the SO and SPSG by 153% and 107%, respectively; two-sided Wilcoxon tests,  $p < 0.05$ ), in the SO. Other degradation pathways included carbohydrate, sugars and acids, and fatty acid and lipid degradation (all decreased significantly from the SO to the SPSG by 54%, 133%, and 50%, respectively; two-sided Wilcoxon tests,  $p < 0.05$ ; Supplementary Fig. 4A–F). Our results also support those of Manganello and Malfatti<sup>26</sup>, who concluded that bacteria and archaea are the most important producers of organic particles via organic degradation when light availability is reduced at higher latitudes.

**Other important pathways.** Fermentation pathways were observed along the entire transect (~2% relative abundance). On average, the highest relative abundances were recorded in the SO, followed by a steady decline (by 26%) from the SO to the southern edge of the PED. All oceanographic provinces were significantly different from each other, except the SPSG and the PED, for which differences in the relative abundances of fermentation pathways were not significant (two-sided Wilcoxon tests,  $p < 0.05$ ; Supplementary Fig. 2D). The anaerobic degradation of organic material (including fermentation) contributes significantly to the degradation processes in marine sediments<sup>39</sup>. Because fermentation is favored under anoxic environments, studies targeting the potential of this process in the (mostly oxygenated) photic zone are absent to the best of our knowledge. Nevertheless anaerobic N-cycling processes such as denitrification and anammox have been shown to occur in anoxic and suboxic marine aggregates in oxygenated waters of the photic zone<sup>40,41</sup>. These microhabitats offer niches for a diverse range of metabolic pathways<sup>42</sup>, and the anoxic zones within marine snow particles could potentially harbor fermentative bacteria. We should note that the presence of fermentation pathways could be an artifact due to the presence of inactive sulfate-reducing bacteria and methanogenic archaea, which are capable of fermenting under favorable environmental condition<sup>39</sup>.

Sulfur metabolism (which significantly decreased by 18% between the SO and the SPSG; Supplementary Fig. 4A) pathways were also found to be characteristic of the SO by the indicator analysis. The SO is a known hotspot for sulfur cycling processes, in particular the production of dimethylsulfoniopropionate (DMSP) (as shown by the high presence of *Phaeocystis* sp. in Sow and Trull<sup>43</sup>) and of the climate cooling dimethylsulfide gas (DMS<sup>44,45</sup>). Members of the SAR11 and *Planktomarina* genera are also known DMSP degraders and were found to dominate the SO in this data set<sup>17</sup>, explaining the higher relative abundance of sulfur-metabolizing genes in this province (Supplementary Fig. 4A). The relative abundance of sulfur-associated pathways declined significantly north of the STF, but pathways were still detectable up to the equator (Supplementary Fig. 4A). The 16S rRNA-derived predictions were, therefore, in agreement with Landa et al.<sup>46</sup>, whose metagenomic analyses from the Tara Ocean data set showed that a large range of marine bacteria are able to use dissolved organic sulfur metabolites, and that the latter play an important part in the global pelagic ocean.

The indicator analyses also highlighted the importance of denitrification pathways in the SO, with relative abundances decreasing significantly by 165% thereafter to the SPSG (two-sided Wilcoxon tests,  $p < 0.05$ ; Supplementary Fig. 4G).

Methanogenesis-associated pathways were characteristic of the STF (Supplementary Fig. 5), whereas phosphorous compounds-associated pathways were characteristic of the SPSG and of the PED, respectively (Supplementary Fig. 6).

**Regression modeling.** From the 22 biotic and abiotic parameters used to model latitudinal trends in metabolic pathways and PP (jointly referred to herein as pathways), only five were identified as the main predictors for  $\geq 4$  pathways (Table 1 and Supplementary Data 2). Temperature was the major predictor of energy metabolism; lipid, carbohydrate, cell structure and cell wall, and vitamin biosynthesis; fermentation; and CO<sub>2</sub>-fixation pathways (12–56%), and one of the main predicting parameters for amino acid and nucleotide biosynthesis pathways (11–15%). Total pigment concentration was the main predictor of cofactor biosynthesis pathways (43%), but also helped predict trends in amino acid, secondary metabolite, and vitamin biosynthesis and CO<sub>2</sub>-fixation pathways (11–20%). Chl b and nanoplankton concentrations and salinity each helped predict four pathways, being the main predictors for secondary metabolite biosynthesis pathways (29%), PP (26%), and nucleotide biosynthesis pathways (20%), respectively. Chl b concentrations also helped predict variations in amino acid biosynthesis, energy metabolism, and CO<sub>2</sub>-fixation pathways (8–13%); whereas nanoplankton concentrations helped predict nucleotide and secondary metabolite biosynthesis and CO<sub>2</sub>-fixation pathways (7–15%) and salinity helped predict energy metabolism and lipid and cofactor biosynthesis pathways (5–12%).  $\delta^{15}\text{N}$  and the concentrations of dissolved inorganic nutrients ( $\text{NH}_4^+$ ,  $\text{NO}_2^-$ ,  $\text{NO}_3^-$ , and Si) and of picoplankton were one of the main predicting parameters for only two pathways.

**Considerations on the use of 16S rRNA gene sequencing for inferences on microbial functional ecology.** We acknowledge that 16S rRNA metabarcoding is a broad-brush approach with a number of limitations for drawing conclusions about changes in functional ecology. Douglas et al.<sup>14</sup> and Langille et al.<sup>13</sup> clearly noted two main criticisms of functional estimates based on 16S rRNA amplicon-based hidden-state predictions. The first is that the predictions are obviously biased toward the available reference genomes (which was empirically quantified by Langille et al.<sup>13</sup>). This limitation will be partially addressed in the near future as the number of metagenome-assembled genomes, and sequenced genomes in general, continues to increase. The second criticism, also confirmed through permutation analyses by Douglas et al.<sup>14</sup>, is that the 16S rRNA-based predictions do not provide the necessary resolution to detect biogeographic patterns of bacterial ecotypes of interest<sup>47</sup>.

We should note two examples from our results that clearly illustrate these limitations. First, the N<sub>2</sub>-fixation and nitrification metabolic pathways, which have been shown to be important in the South Pacific Ocean<sup>18</sup>, were not present in the PICRUSt2 MetaCyc outputs. This is likely because N<sub>2</sub>-fixation is not well resolved by 16S rRNA gene sequencing<sup>48</sup> and because bacteria involved in nitrification made up only 1% of the bacterial biomass in the samples (see Supplementary Fig. 9 in Raes et al.<sup>18</sup>). Underestimating the occurrence of these pathways that contribute to inputs of new (N<sub>2</sub>-fixation) and regenerated (nitrification) N could ultimately lead to global oceanic models underestimating PP<sup>49,50</sup>. This limitation can, however, be addressed with the use of additional amplicon-based analyses. For example, N<sub>2</sub>-fixation functional gene (*nifH*) sequencing data have been coupled with direct rate measurements to reveal biogeographic patterns of the diazotrophic community<sup>18</sup>. Secondly, our analyses do not provide the necessary resolution to

**Table 1 Contributions of the 22 biotic and abiotic predictors to boosted regression tree models relating latitudinal trends to the (relative) abundance of the ten secondary superclasses + CO<sub>2</sub>-fixation pathways and PP at Depth 1 (n = 75) along the GO-SHIP P15S transect.**

| Pathway/process                            | Parameter 1           | Parameter 2                           | Parameter 3                           | Parameter 4                          | Parameter 5                          | cv           |
|--------------------------------------------|-----------------------|---------------------------------------|---------------------------------------|--------------------------------------|--------------------------------------|--------------|
| PP                                         | [Nanoplankton] (26%)  | [HexFuco] (26%)                       | [Pras] (16%)                          | [POC] (6%)                           | [NH <sub>4</sub> <sup>+</sup> ] (4%) | 0.91 ± 0.02  |
| CO <sub>2</sub> -fixation                  | Temperature (25%)     | [Total pigment] (20%)                 | [Nanoplankton] (15%)                  | [chl b] (3%)                         | [O <sub>2</sub> ] (7%)               | 0.842 ± 0.04 |
| Energy metabolism                          | Temperature (12%)     | Salinity (12%)                        | [NH <sub>4</sub> <sup>+</sup> ] (10%) | [chl b] (9%)                         | [Microplankton] (8%)                 | 0.757 ± 0.08 |
| Nucleotide biosynthesis                    | Salinity (20%)        | δ <sup>13</sup> C (16%)               | Temperature (15%)                     | [Nanoplankton] (7%)                  | [Silicate] (6%)                      | 0.725 ± 0.13 |
| Cell structure and cell wall biosynthesis. | Temperature (56%)     | [NH <sub>4</sub> <sup>+</sup> ] (18%) | [Zeaxanthin] (4%)                     | δ <sup>13</sup> C (3%)               | [Picoplankton] (3%)                  | 0.955 ± 0.01 |
| Amino acid biosynthesis                    | [Picoplankton] (34%)  | Temperature (11%)                     | [Total pigment] (11%)                 | chl b (8%)                           | δ <sup>15</sup> N (5%)               | 0.796 ± 0.05 |
| Secondary metabolite biosynthesis          | chl b (29%)           | [Total pigment] (20%)                 | [Nanoplankton] (8%)                   | δ <sup>15</sup> N (7%)               | C:N ratios (5%)                      | 0.822 ± 0.05 |
| Cofactor biosynthesis                      | [Total pigment] (43%) | Salinity (8%)                         | [PON] (7%)                            | [Silicate] (7%)                      | δ <sup>15</sup> N (4%)               | 0.837 ± 0.03 |
| Vitamin biosynthesis                       | Temperature (30%)     | [NO <sub>3</sub> <sup>-</sup> ] (17%) | [Total pigment] (13%)                 | [NO <sub>2</sub> <sup>-</sup> ] (9%) | [Fuco] (6%)                          | 0.946 ± 0.02 |
| Lipid biosynthesis                         | Temperature (45%)     | [NO <sub>3</sub> <sup>-</sup> ] (9%)  | [NO <sub>2</sub> <sup>-</sup> ] (6%)  | Salinity (5%)                        | [DV:Chl-a] (5%)                      | 0.956 ± 0.01 |
| Carbohydrate biosynthesis                  | Temperature (56%)     | [NH <sub>4</sub> <sup>+</sup> ] (18%) | [Zeaxanthin] (4%)                     | δ <sup>13</sup> C (3%)               | Salinity (3%)                        | 0.945 ± 0.01 |
| Fermentation                               | Temperature (28%)     | [Picoplankton] (23%)                  | [NO <sub>2</sub> <sup>-</sup> ] (11%) | [Fuco] (10%)                         | δ <sup>13</sup> C (4%)               | 0.928 ± 0.02 |

The relative contribution of each parameter is shown in brackets (%). The cumulative contribution of all listed parameters within a pathway/PP amounts to ~60–80% of the model predictability, except for the energy metabolism pathway, which shows a cumulative predictability of 51%. Square brackets are used to represent concentrations. A full breakdown of all predictor variables can be found in Supplementary Data 2. cv coefficient of variation.

detect biogeographic patterns of ecotypes of interest. While ecotypes such as the Pelagibacter SAR11, the cyanobacteria Prochlorococcus, and the prymnesiophyte *Phaeocystis* sp., among others, appear functionally redundant in a broad, amplicon-based functional analysis, the fine-scale metabolic variations that have evolved among these ecotypes may have important bearing on the temporal and spatial structure of the community and productivity of the ecosystem<sup>43,51–53</sup>. Such limitations have been addressed with focused taxonomic analysis of, for example, amplicon sequence variant (ASV) data<sup>43</sup>.

**A tool for large-scale functional ecology.** We set out to test five hypotheses under the assumption that 16S rRNA gene sequences can offer significant insight into the functional diversity of bacterial communities in oceanographic studies. Our results demonstrated that the observed latitudinal trends in metabolic pathways generated by the PICRUST2 software were consistent with measured physico-chemical parameters such as temperature, nutrient bioavailability, diagnostic pigments (e.g., fucoxanthin, prasinoxanthin, chl b, and zeaxanthin), and the isotopic fractionation of PON, among others. In addition, our observations aligned with our measurements of biogeochemical rates, with quantitative and qualitative predictions from the available literature, and with our overall mechanistic understanding of functional microbial biogeography in the South Pacific Ocean. A comparative analysis between the KO predictions from PICRUST2 and the KOs profiled from the corresponding MGS provided support to the inferred metabolic pathways and, thus, to the proposed hypotheses. Our results exemplify the potential for low-cost, high-throughput mapping of (functional) biodiversity and ecosystem change in global monitoring campaigns such as GO-SHIP and (bio)GEOTRACES. Community-level metabolic information directly speaks to the state of, and changes in, ecosystems, while also complementing the information provided by abiotic variables, which are more routinely used to monitor the state of the oceans. The ability to query metabolic pathways in existing and future 16S rRNA gene data sets on a global scale establishes the opportunity to test hypotheses regarding how biodiversity influences functional diversity, and how these are related to energy production in the ocean. Deriving metabolic profiles from 16S rRNA gene data sets obtained by oceanic sampling programs on a global scale may, thus, provide a better understanding of the components of a resilient marine ecosystem and of how that resilience is tested through existing and emerging environmental stressors.

Methods

**Study area and water sampling.** This oceanographic study was conducted in the South Pacific Ocean onboard the R.V. Investigator from 23 April to 29 June 2016 along the longitudinal P15S GO-SHIP line at 170° W (Fig. 1a). The P15S GO-SHIP line is a transect that runs from the ice edge (~66° S) to the equator (0° S; Fig. 1; <http://www.go-ship.org/>). The results presented herein are a continuation of the work from Raes et al.<sup>17,18</sup> along the P15S GO-SHIP transect. For clarity, we briefly reiterate some of the methodology applied to our study, but for an in-depth explanation on the physical and biochemical data validation and the presented C and N rate measurements we refer the reader to the aforementioned papers and the Supplementary methods.

For the purposes of this study, the P15S GO-SHIP transect was divided into four oceanographic provinces or Longhurst provinces<sup>54</sup>. From south to north the transect covered: (1) the SO between 66° and 52° S; (2) the STF between 52° and 40° S; (3) the SPSG between 40° and 10° S; and (4) the PED, between 10° and 0° S. Physical, biogeochemical and metadata were collected from 36 depth horizons at 140 stations (approximately every half a latitudinal degree). Full depth profiles for temperature, salinity, and dissolved oxygen were conducted using a Seabird (SBE25 plus) conductivity–temperature–depth profiler with a SBE43 O<sub>2</sub> sensor mounted on a 36 Niskin bottle rosette sampler.

**DNA sequencing and bioinformatics.** Samples for DNA analyses were collected from 12 litres Niskin bottles at three depths within the mixed layer. Depth horizons

ranged between 1.3 and 36.7 m ( $6.6 \pm 4.1$  m (average  $\pm$  SD)) for Depth 1; 18.6 and 85.8 m ( $35.3 \pm 13.1$  m) for Depth 2; and 39.9 and 185 m ( $70.7 \pm 18.0$  m) for Depth 3 (Fig. 1c). A peristaltic pump was used to filter two litres of seawater through a 0.22- $\mu$ m pore size Sterivex<sup>TM</sup> filter (catalog no. SVGPL10RC; Millipore, Germany). Samples were stored at  $-80^{\circ}\text{C}$  until DNA extraction. DNA was isolated from the filters following a modified organic (phenol:chloroform:isoamyl-alcohol-based) extraction protocol<sup>55</sup> of the PowerWater<sup>®</sup> Sterivex<sup>TM</sup> DNA isolation kit (Mo Bio Laboratories-QIAGEN, Germany). Bacterial diversity was investigated via tag sequencing targeting the V1–V3 region of the 16S rRNA gene with the bacterial forward 27F (AGAGTTTGATCTMTGGCTCAG) and reverse 519R (GWAT-TACCGCGGCKGCTG) primer sets (Lane et al.<sup>56</sup> and Lane<sup>57</sup>) using the Illumina MiSeq<sup>TM</sup> platform. 16S rRNA amplicons were generated using 300-bp paired-end sequencing at the Ramaciotti Centre for Genomics (University of New South Wales, Sydney). ASV tables were prepared after Bissett et al.<sup>58</sup> and as outlined in the Supplementary methods. To reduce run time, and because PICRUSt2 estimates the genome from the nearest ancestor, we clustered the ASV's at the 97% similarity threshold (which were generated using the USEARCH cluster\_fast function with -id 0.97; please refer to the Supplementary methods for the full workflow).

**Shotgun metagenomics.** Illumina Novaseq shotgun sequencing was performed for 11 of the samples collected for DNA analyses in order to complement the 16S rRNA gene data. Genomic DNA shearing and library preparations were performed at the Ramaciotti Centre for Genomics (UNSW, Sydney, Australia; see Supplementary methods for details). The sequencing depth for the 11 samples ranged from 60 to 109 million reads (with an average of  $8.5 \pm 17$  million reads; see Supplementary Table 1). Illumina adapters were removed from the R1 and R2 reads using the Trimmomatic software (with the TruSeq3-PE-2.fa:2:30:10:2:keepBothReads setting; version: 0.38; Bolger et al.<sup>59</sup>). The R1 and R2 reads from each sample were then merged using bbmerge (version 38.37; Bushnell et al.<sup>60</sup>). The unmerged sequences were quality controlled (QCed) using the Trimmomatic software (version: 0.38; Bolger et al.<sup>59</sup>) with the following settings: leading:3, trailing:3, slidingwindow:10:15, minlen:50 bp. After QC the merged and unmerged sequences were concatenated into one file per sample. Samples were not rarefied. Functional assignments were done on the concatenated files which contained both QCed merged and unmerged sequences with the sqm\_reads analysis mode using the SqueezeMeta software version 1.2.0 with default settings<sup>61</sup>. Functions were assigned using Diamond Blastx alignments<sup>62</sup> of the reads against the Clusters of Orthologous Groups of proteins and KEGG databases<sup>63–65</sup> using the lowest common ancestor (Huson et al.<sup>66</sup>) and fun3 methods. The script make\_databases.pl was used to download and format the latest versions of the databases (database creation on Tuesday May 5, 2020). The script combine-sqm-tables.py was then used to generate and combine tabular outputs from all 11 samples. All scripts used in this study are available on the SqueezeMeta GitHub page <https://github.com/jtamames/SqueezeMeta>.

**Functional composition.** The software PICRUSt2 (version 2.3.0b)<sup>13,14</sup> was used with default settings to infer approximate functional potential of the microbial communities sampled across the 7000 km transect in the South Pacific Ocean. The average Nearest Sequenced Taxon Index (NSTI) score, based on 387 samples (covering the three depths), was  $0.134 \pm 0.031$  ( $\pm$ SD). Approximately 1% of OTUs (51 out of 4189) were above the maximum NSTI cut-off score of values  $>2$ , and were removed. These removed OTUs represented 2.9% of the relative abundance of the bacterial community, and included two OTUs which showed 94 and 99% match with *Bathycoccus prasinos* mitochondrial DNA and represented 2.8% of the bacterial community relative abundance. It should be noted that chloroplasts and mitochondrial sequences (7% of the data) were removed from the data set prior to the PICRUSt2 analyses. The relative abundance of each OTU (including its sequence at the 97% similarity threshold) with NSTI values  $>2$  is shown in Supplementary Data 3. Two pathways with  $<10$  reads were removed from the data set, namely PWY-6948/sitosterol degradation to androstenedione and PWY-6713/L-rhamnose degradation II. The final predicted metagenome pathway abundance data were converted to relative abundances per sample by rarefying to the lowest abundance per sample as suggested by Douglas et al.<sup>14</sup>.

**Statistical analyses.** Statistical analyses and data visualizations were performed with R version 3.6.1<sup>67</sup> and the PRIMER v7 software. For a list of all analyses performed and detailed information and citations regarding the software versions of the R packages used in analyses, please refer to the Supplementary methods. The workflow code for the analyses reported in our study are available on GitHub<sup>68</sup>.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

All physical, biogeochemical, and meta- data are available at the CLIVAR and Carbon Hydrographic Data Office (CCHDO; <https://cchdo.ucsd.edu/>; GO-SHIP transect P155 expocode: 096U20160426). Genomic data are available at <https://www.ncbi.nlm.nih.gov/bioproject/385736>. Primary productivity and nitrogen assimilation data from this study are available at <https://doi.pangaea.de/10.1594/PANGAEA.884052> and <https://doi.pangaea.de/10.1594/PANGAEA.885169>. Pigment data are available at <https://doi.pangaea.de/10.1594/PANGAEA.884052>. All the code, key data files, and workflows to reproduce the analyses and figures are available on GitHub<sup>68</sup>.

## Code availability

Scripts, key data files, and workflows to reproduce the analyses and plots are available from GitHub<sup>68</sup>.

Received: 27 August 2020; Accepted: 9 March 2021;

Published online: 13 April 2021

## References

- Falkowski, P. G. Evolution of the nitrogen cycle and its influence on the biological sequestration of CO<sub>2</sub> in the ocean. *Nature* **387**, 272 (1997).
- Alvarez, S. A. et al. Diversity decoupled from ecosystem function and resilience during mass extinction recovery. *Nature* **574**, 242–245 (2019).
- Sunagawa, S. et al. Structure and function of the global ocean microbiome. *Science* **348**, 1261359 (2015).
- Blasiak, R. et al. The ocean genome: conservation and the fair, equitable and sustainable use of marine genetic resources. Blue Paper (2020).
- Boss, E. et al. Recommendations for plankton measurements on the GO-SHIP program with relevance to other sea-going expeditions. SCOR Working Group 154 GO-SHIP Report (2020).
- Billar, S. J. et al. Marine microbial metagenomes sampled across space and time. *Sci. Data* **5**, 180176 (2018).
- Méthé, B. A. et al. A framework for human microbiome research. *Nature* **486**, 215–221 (2012).
- Thompson, L. R. et al. A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* **551**, 457–463 (2017).
- Clarke, L. J. et al. Mainstreaming microbes across biomes. *BioScience* **70**, 589–596 (2020).
- Sczyrba, A. et al. Critical assessment of metagenome interpretation—a benchmark of metagenomics software. *Nat. Methods* **14**, 1063–1071 (2017).
- Karl, D. M. & Church, M. J. Microbial oceanography and the Hawaii Ocean Time-series programme. *Nat. Rev. Microbiol.* **12**, 699–713 (2014).
- Brown, M. V. et al. Systematic, continental scale temporal monitoring of marine pelagic microbiota by the Australian Marine Microbial Biodiversity Initiative. *Sci. Data* **5**, 180130 (2018).
- Langille, M. G. et al. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat. Biotechnol.* **31**, 814–821 (2013).
- Douglas, G. M. et al. PICRUSt2 for prediction of metagenome functions. *Nat. Biotechnol.* **38**, 685–688 (2020).
- Agrawal, S. et al. Determining uncertainties in PICRUSt analysis—an easy approach for autotrophic nitrogen removal. *Biochemical Eng. J.* **152**, 107328 (2019).
- Sun, S., Jones, R. B. & Fodor, A. A. Inference-based accuracy of metagenome prediction tools varies across sample types and functional categories. *Microbiome* **8**, 1–9 (2020).
- Raes, E. J. et al. Oceanographic boundaries constrain microbial diversity gradients in the South Pacific Ocean. *Proc. Natl Acad. Sci. USA* **115**, E8266–E8275 (2018).
- Raes, E. J. et al. N<sub>2</sub> Fixation and new insights into nitrification from the ice-edge to the equator in the South Pacific Ocean. *Front. Mar. Sci.* **7**, 1–20 (2020).
- Floodgate, G. D., Fogg, G. E., Jones, D. A., Lochte, K. & Turley, C. M. Microbiological and zooplankton activity at a front in Liverpool Bay. *Nature* **290**, 133–136 (1981).
- Clarke, A. & Fraser, K. Why does metabolism scale with temperature? *Funct. Ecol.* **18**, 243–251 (2004).
- Boyd, P. W. et al. A mesoscale phytoplankton bloom in the polar Southern Ocean stimulated by iron fertilization. *Nature* **407**, 695–702 (2000).
- Browning, T. J. et al. Nutrient co-limitation at the boundary of an oceanic gyre. *Nature* **551**, 242–246 (2017).
- de Mendoza, D. & Cronan, J. E. Jr Thermal regulation of membrane lipid fluidity in bacteria. *Trends Biochem. Sci.* **8**, 49–52 (1983).
- Parrish, C. C. & Wangersky, P. J. Growth and lipid class composition of the marine diatom, *Chaetoceros gracilis*, in laboratory and mass culture turbidostats. *J. Plankton Res.* **12**, 1011–1021 (1990).
- Rivkin, R. B. & Legendre, L. Biogenic carbon cycling in the upper ocean: effects of microbial respiration. *Science* **291**, 2398–2400 (2001).
- Manganelli, M. et al. Major role of microbes in carbon fluxes during austral winter in the southern drake passage. *PLoS ONE* **4**, e6941 (2009).
- Mattarelli, P., Biavati, B., Pesenti, M. & Crociani, F. Effect of growth temperature on the biosynthesis of cell wall proteins from *Bifidobacterium globosum*. *Res. Microbiol.* **150**, 117–127 (1999).

28. Schouten, S., Hopmans, E. C., Schefuß, E. & Sinninghe Damsté, J. S. Distributional variations in marine crenarchaeotal membrane lipids: a new tool for reconstructing ancient sea water temperatures? *Earth Planet Sci. Lett.* **204**, 265–274 (2002).
29. Enke, T. N., Leventhal, G. E., Metzger, M., Saavedra, J. T. & Cordero, O. X. Microscale ecology regulates particulate organic matter turnover in model marine microbial communities. *Nat. Commun.* **9**, 2743 (2018).
30. Busscher, H. J. et al. Measurement of the surface free energy of bacterial cell surfaces and its relevance for adhesion. *Appl. Environ. Microbiol.* **48**, 980–983 (1984).
31. Parrish, C. C. Lipids in marine ecosystems. *ISRN Oceanogr.* **2013**, 1–16 (2013).
32. Wältermann, M. & Steinbüchel, A. Neutral lipid bodies in prokaryotes: recent insights into structure, formation, and relationship to eukaryotic lipid depots. *J. Bacteriol.* **187**, 3607–3619 (2005).
33. Kalscheuer, R. et al. Analysis of storage lipid accumulation in *Alcanivorax borkumensis*: evidence for alternative triacylglycerol biosynthesis routes in bacteria. *J. Bacteriol.* **189**, 918–928 (2007).
34. Alvarez, H., Kalscheuer, R. & Steinbüchel, A. Accumulation and mobilization of storage lipids by *Rhodococcus opacus* PD630 and *Rhodococcus ruber* NCIMB 40126. *Appl. Microbiol. Biotechnol.* **54**, 218–223 (2000).
35. Phleger, C. F., Nichols, P. D. & Virtue, P. Lipids and trophodynamics of Antarctic zooplankton. *Comp. Biochem. Physiol. Part B: Biochem. Mol. Biol.* **120**, 311–323 (1998).
36. DyhrMaN, S. T., Ammerman, J. W. & Van Mooy, B. A. Microbes and the marine phosphorus cycle. *Oceanography* **20**, 110–116 (2007).
37. Van Mooy, B. A., Rocap, G., Fredricks, H. F., Evans, C. T. & Devol, A. H. Sulfolipids dramatically decrease phosphorus demand by picocyanobacteria in oligotrophic marine environments. *Proc. Natl Acad. Sci. USA* **103**, 8607–8612 (2006).
38. Sebastián, M. et al. Lipid remodelling is a widespread strategy in marine heterotrophic bacteria upon phosphorus deficiency. *ISME J.* **10**, 968–978 (2016).
39. Kirchman, D. L., Hanson, T. E., Cottrell, M. T. & Hamdan, L. J. Metagenomic analysis of organic matter degradation in methane-rich Arctic Ocean sediments. *Limnol. Oceanogr.* **59**, 548–559 (2014).
40. Woebken, D., Fuchs, B. M., Kuypers, M. M. & Amann, R. Potential interactions of particle-associated anammox bacteria with bacterial and archaeal partners in the Namibian upwelling system. *Appl. Environ. Microbiol.* **73**, 4648–4657 (2007).
41. Ploug, H. Small-scale oxygen fluxes and remineralization in sinking aggregates. *Limnol. Oceanogr.* **46**, 1624–1631 (2001).
42. Rogge, A. et al. Hard and soft plastic resin embedding for single-cell element uptake investigations of marine-snow-associated microorganisms using nano-scale secondary ion mass spectrometry. *Limnol. Oceanogr. Methods* **16**, 484–503 (2018).
43. Sow, S. L. S., Thomas, W. T. & Bodrossy, L. Oceanographic fronts shape phaeocystis assemblages: a high-resolution 18S rRNA gene survey from the ice-edge to the equator of the South Pacific. *Front. Microbiol.* **11**, 1847 (2020).
44. Berresheim, H. Biogenic sulfur emissions from the Subantarctic and Antarctic Oceans. *J. Geophys. Res. Atmos.* **92**, 13245–13262 (1987).
45. Sheehan, C. E. & Petrou, K. Dimethylated sulfur production in batch cultures of Southern Ocean phytoplankton. *Biogeochemistry* **147**, 53–69 (2020).
46. Landa, M. et al. Sulfur metabolites that facilitate oceanic phytoplankton–bacteria carbon flux. *ISME J.* **13**, 2536–2550 (2019).
47. Brown, M. V., Ostrowski, M., Grzymski, J. J. & Lauro, F. M. A trait based perspective on the biogeography of common and abundant marine bacterioplankton clades. *Mar. Genomics* **15**, 17–28 (2014).
48. Gaby, J. C. & Buckley, D. H. A comprehensive aligned *nifH* gene database: a multipurpose tool for studies of nitrogen-fixing bacteria. *Database (Oxford)* **2014**, bau001 (2014).
49. Dugdale, R. & Goering, J. Uptake of new and regenerated forms of nitrogen in primary productivity 1. *Limnol. Oceanogr.* **12**, 196–206 (1967).
50. Eppley, R. W. & Peterson, B. J. Particulate organic matter flux and planktonic new production in the deep ocean. *Nature* **282**, 677–680 (1979).
51. Brown, M. V. et al. Global biogeography of SAR11 marine bacteria. *Mol. Syst. Biol.* **8**, 595 (2012).
52. Martiny, A. C., Coleman, M. L. & Chisholm, S. W. Phosphate acquisition genes in *Prochlorococcus* ecotypes: evidence for genome-wide adaptation. *Proc. Natl Acad. Sci. USA* **103**, 12552–12557 (2006).
53. Braakman, R., Follows, M. J. & Chisholm, S. W. Metabolic evolution and the self-organization of ecosystems. *Proc. Natl Acad. Sci. USA* **114**, E3091–E3100 (2017).
54. Longhurst, A. R. *Ecological Geography of the Sea* (Elsevier, 2010).
55. Appleyard, S. A., Abell, G. & Watson, R. *Tackling Microbial-related Issues in Cultured Shellfish via Integrated Molecular and Water Chemistry Approaches* (CSIRO Marine and Atmospheric Research, 2013).
56. Lane, D. J. et al. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proc. Natl Acad. Sci. USA* **82**, 6955–6959 (1985).
57. Lane, D. 16S/23S rRNA sequencing. In *Nucleic Acid Techniques in Bacterial Systematics* (eds Stackebrandt, E. & Goodfellow, M.) (Wiley, 1991).
58. Bissett, A. et al. Introducing BASE: the Biomes of Australian Soil Environments soil microbial diversity database. *GigaScience* **5**, 21 (2016).
59. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
60. Bushnell, B., Rood, J. & Singer, E. BBMerge—accurate paired shotgun read merging via overlap. *PLoS ONE* **12**, e0185056 (2017).
61. Tamames, J. & Puente-Sanchez, F. SqueezeMeta, a highly portable, fully automatic metagenomic analysis pipeline. *Front. Microbiol.* **9**, 3349 (2019).
62. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2014).
63. Clark, K., Karsch-Mizrachi, L., Lipman, D. J., Ostell, J. & Sayers, E. W. GenBank. *Nucleic Acids Res.* **44**, D67–D72 (2016).
64. Tatusov, R. L. et al. The COG database: an updated version includes eukaryotes. *BMC Bioinforma.* **4**, 1–14 (2003).
65. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
66. Huson, D. H., Auch, A. F., Qi, J. & Schuster, S. C. MEGAN analysis of metagenomic data. *Genome Res.* **17**, 377–386 (2007).
67. R Core Team. *R: A Language and Environment for Statistical Computing* (R Core Team, 2013).
68. Raes, E. J. et al. Metabolic pathways inferred from a bacterial marker gene illuminate ecological changes across South Pacific frontal boundaries. (Version v1.0.0). Zenodo. <https://doi.org/10.5281/zenodo.4567694> (2021).
69. Talley, L. D., Pickard, G. L., Emery, W. J. & Swift, J. H. Chapter 10—Pacific Ocean. In *Descriptive Physical Oceanography* 6th edn (eds L. D., Talley, G. L., Pickard, W. J., Emery & J. H. Swift) (Academic Press, 2011).
70. Ganachaud, A. et al. The Southwest Pacific Ocean circulation and climate experiment (SPICE). *J. Geophys. Res. Oceans* **119**, 7660–7686 (2014).

## Acknowledgements

We would like to thank the captain, officers and crew of the R.V. Investigator during cruise IN2016\_V03 for their technical assistance while at sea; and Dr. Susan Wijffels and Dr. Bernadette Sloyan for the opportunity to piggyback on the P155 GO-SHIP transect voyage number IN2016\_V03. We would also like to thank Dr. Bernhard Tschitschko, Dr. Nicole Gail Hellesey, and Gabriela Paniagua Cabarrus for their help at sea. We would like to acknowledge the contribution of the Marine Microbes consortium in the generation of data used in this publication. The Marine Microbes project was supported by funding from Bioplatforms Australia and the Integrated Marine Observing System (IMOS) through the Australian Government National Collaborative Research Infrastructure Strategy (NCRIS) in partnership with the Australian research community. This work was supported by CSIRO, the Australian Climate Change Science Program, and the Marine National Facility. A.M.W. was supported by grants from the Alfred Wegener Institute, the University of Western Australia, and the Ocean Frontier Institute. This work was supported by Australian Research Council awards DP150102326 to M.B. and M.O. Work at CSIRO was supported by CSIRO Office of Community Engagement Science Leader Fellowship R-04202 (to L.B.) and by CSIRO Oceans and Atmosphere Environmental Genomics Grant R-02412. We sincerely thank Dr. Tom Trull, Dr. Sharon Hook, and Dr. Richard Matear for their valuable comments on an earlier draft.

## Author contributions

E.J.R. oversaw data collection and analyzed data. K.K., S.L.S.S., L.B., R.M.F.-S., and A.M.W. contributed to data analyses and data interpretation. E.J.R., L.B., J.v.d.K., M.O., and M.B. contributed to the study design. All authors contributed to the writing of the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-22409-4>.

**Correspondence** and requests for materials should be addressed to E.J.R.

**Peer review information** *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021