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

Eric Raes (eric.raes@csiro.au)
CSIRO
Kristen Karsh
CSIRO
Swan Sow
Royal Netherlands Institute for Sea Research
Martin Ostrowski
University of Technology Sydney https://orcid.org/0000-0002-4357-3023
Mark Brown
University of Newcastle Australia
Jodie van de Kamp
CSIRO https://orcid.org/0000-0003-2167-0938
Rita Franco-Santos
University of Tasmania
Levente Bodrossy
CSIRO Marine and Atmospheric Research https://orcid.org/0000-0001-6940-452X
Anya Waite
Ocean Frontier Institute

Article

Keywords: Latitudinal Trends, GO-SHIP, Metabolic Reconstruction, Marker genes, South Pacific Ocean, Bacteria, Functional diversity, MetaCyc pathways

Posted Date: September 4th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-67387/v1

Version of Record: A version of this preprint was published at Nature Communications on April 13th, 2021. See the published version at https://doi.org/10.1038/s41467-021-22409-4.

Abstract

Global oceanographic monitoring initiatives started by measuring abiotic essential ocean variables but are currently incorporating biological and metagenomic sampling. There is, however, a large gap between the taxonomic information produced by bacterial genomic analyses and information on bacterial functions, which is sought by biogeochemists, ecologists, and modellers. Here, we provide a mechanistic understanding of how a bacterial marker gene (16S rRNA) can be used to derive latitudinal trends for core metabolic pathways and, ultimately, be used for mapping ecosystem function change in global monitoring campaigns. From a transect spanning 7000 km in the South Pacific Ocean we identified ten metabolic pathways, which were related to ecological processes of primary productivity, temperature-regulated growth, coping strategies for nutrient limitation, energy metabolism, and degradation. We compared and contrasted these metabolic pathways with measured physico-biochemical parameters within and between oceanographic provinces, and found that functional diversity is as affected by oceanographic boundaries as is taxonomic composition. This study demonstrates that bacterial marker gene data, sampled and analysed with low costs and high throughput, can be used to infer on metabolic changes at the community scale. Such analyses may provide insight into the drivers of ecological changes and, overall, into the effects of biodiversity on marine ecosystem functioning.

Introduction

“No water, no life. No blue, no green” (Dr. Sylvia Earle). The oceans cover 71% of our planet, and the microbial organisms inhabiting the oceans catalyse important ecosystem services (such as O_2 production, C sequestration and elemental cycling) which sustain life on Earth [1]. Because microbial communities execute key roles in numerous biogeochemical pathways, it is important to understand how these communities respond to environmental changes. Mapping microbial biogeography in relation to abiotic and biotic parameters therefore merits intensive investigation [2]. A better understanding of the ‘ocean genome’ [3] will also allow society to better preserve and utilise the vast genetic diversity in marine ecosystems. Global oceanographic initiatives such as the GO-SHIP [4] and the bioGEOTRACES[5] programs, which started out by surveying abiotic essential ocean variables (EOVs) such as temperature, salinity and dissolved inorganic trace metals, have started to include biological EOVs, such as metagenomics, in their global sampling program. These biotic EOVs can fill critical knowledge gaps relating to how ecological drivers shape microbial community structure and influence their functional diversity in the open ocean.

Recent metagenomics and metatranscriptomics work from the Tara Oceans Project suggested that, in polar regions, the effects of ocean warming on bacterial communities will be stronger on a community (composition) level than on a gene (regulatory) level [6]. On a basin scale, Raes, Bodrossy [7] showed that permanent oceanographic frontal systems can have a significant influence on both the alpha- and beta-diversity of prokaryotes and microbial eukaryotes. In addition, their data supported a strong positive link between productivity and microbial diversity. However, overall, we still lack insight on how local environmental conditions modulate the correlation between functional microbial diversity and productivity, both across frontal zones and within ocean provinces.

The highly conserved 16S ribosomal RNA gene (16S rRNA) is commonly sequenced for prokaryotic identification and microbial community profiling; an analysis that has been employed to study several biomes around the world [2, 8, 9, 10]. 16S rRNA sequencing, however, does not provide direct information on the metabolic functions of the microbial communities, data which can only be obtained (and with greater resolution regarding diversity) via shotgun metagenomics [11]. However, because the costs of shotgun metagenomics are currently ~5 times those of 16S rRNA sequencing, and because the amount of spatial and temporal 16S rRNA data available in the literature vastly surpasses that of shotgun data[12, 13], evolutionary modellers have often inferred about the potential functional profiles of microbial communities from marker genes such as 16S rRNA sequence data [14]. Although this is an indirect method, it has been

shown that the analysis of 16S rRNA data with the open source software Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt[14] and the improved version PICRUSt2 [15]) results in predictions of functional microbial profiles that are in strong agreement (i.e., high Spearman correlations) with results from shotgun metagenomics. The best predictions presented so far relate to communities in the human microbiome, followed by those in the ocean biome. Sun, Jones [16], however, provided evidence that correlation coefficients such as Spearman are an unreliable measure to test the performance of prediction tools such as PICRUSt2.

Our aim was to test six hypotheses to determine if 16S rRNA sequences hold significant information regarding microbial ecological functions. More specifically, by using biomass (i.e., the concentrations of various photosynthetic pigments and of particulate organic carbon (C) and nitrogen (N)), primary productivity, and N assimilation measurements we set out to test the validity of the following six hypotheses:

H1: Metabolic reconstructions based on marker gene surveys, using evolutionary prediction programs such as PICRUSt2, can predict broad-scale latitudinal patterns in microbial functional processes which agree with our current mechanistic understanding of functional microbial biogeography, both within and across ecological provinces in the South Pacific Ocean (such as laid out by Raes, Bodrossy [7] and Raes, van de Kamp [17]);

H2: Frontal zones, which stimulate primary productivity [18], should display a higher relative abundance of pathways associated with CO_2-fixation and energy metabolism;

H3: The number of pathways related to cell structure and cell wall biosynthesis pathways increase with increasing temperatures (in accord with the expectation that higher temperatures increase growth rates; see Eppley [19], Thomas, Kremer [20] and Hoppe, Gocke [21]);

H4: Latitudinal trends will be identified for pathways which reflect microbial strategies in coping with trace metal and macro nutrient limitations (i.e., high relative abundances of co-factor and secondary metabolite biosynthesis pathways due to iron limitation in the Southern Ocean [22] and co-nutrient stress in the oligotrophic gyre Browning, Achterberg [23]);

H5: Pathways associated with energy storage (such as lipid and carbohydrate synthesis) should be most active in the Southern Ocean and in the productive subtropical frontal zone (due to high temporal variability and seasonal change in these environments [24]);

H6: Degradation-type pathways should occur in higher relative abundances in the Southern Ocean due to higher rates of bacterial degradation of particulate and dissolved organic material (Rivkin and Legendre [25]).

Results

Hydrographic conditions

This oceanographic study was conducted in late Austral autumn and early winter along the decadal repeated P15S GO-SHIP transect, which ran 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 ^{o}C at 66 °S to 30.4 ^{o}C at 5.5 °S, and then decreased slightly, due to the equatorial upwelling, to 28.1 ^{o}C at the equator (Fig. 1 A-B). Surface salinity was lowest in the Southern Ocean (33.8-34.3), increasing north of the Polar Front (60 °S) to a maximum of 35.87 at 30 °S. Salinity decreased north of 30 °S to 34.5 at 10 °S and then increased to 35.25 at the equator (Fig. 1 A-B). NO_3^-, Si and PO_4^{3-} concentrations covaried above the mixed layer depth (MLD), and were closely linked to the major oceanographic features (Raes, Bodrossy [7]). NO_3^- concentrations above the MLD were > 16 µmol L^{-1} in the Southern Ocean; between 1

and 16 μmol L^{-1} in the STF; ≤ 0.05 μmol L^{-1} in the SPSG; and up to 2 μmol L^{-1} in the PED (Supplementary Fig. 1). NO_3^-:PO_4^{3-} ratios above the MLD showed differences that emphasized the four distinct oceanographic provinces, averaging 14.24 ± 0.25 in the Southern Ocean, 9.76 ± 2.3 in the STF, 0.431 ± 0.874 in the oligotrophic SPSG, and 5.6 ± 1.4 in the PED (Fig. 1 C).

Bacterial and functional 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. 2 A). 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. 2 B; and Supplementary Table 1. PICRUSt2 analyses of 387 DNA samples resulted in the inference of 399 MetaCyc pathways derived from 387 samples. No significant differences were not observed in the alpha-diversity of MetaCyc pathways diversities between the four oceanographic provinces (Wilcoxon test $p>0.05$; and Fig. 2 C). CAP plots and ANOSIM results analyses for the MetaCyc pathways data revealed significant differences in the functional bacterial pathway community compositions between the four oceanographic
provinces (Fig. 2 D; and Supplementary Table 1).

Functional community composition

The 399 inferred MetaCyc pathways (supplementary Table 2) could be collapsed into 41 secondary superclasses. Across the transect, the sum of the relative abundance of ten of these superclasses accounted for up to 80% of the total relative abundance, and they are hereon referred to as the core ecosystem functions. They covered (in descending order of relative importance) 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. Core functions (see Fig. 3 for functions 1-9 and Supplementary Fig. 2 for function 10) are described in more detail in the following paragraphs in relation to PP (H2), temperature-regulated growth (H3), coping strategies for micro and macronutrient limitation (H4), energy storage (H5), and degradation (H6).

Pathways associated with energy metabolism, nucleotide biosynthesis and CO_2-fixation (the latter contributed less than 2% to the total relative abundance) showed trends similar to those of PP and which are in agreement with the distinction between productive (the STF and the PED) and less productive (the Southern Ocean and the SPSG) oceanographic provinces, as suggested by our H2 (Figs 1B and 3A-D). Pathways associated with CO_2-fixation increased from 66 ^oS to the northern edge of the STF at 40 ^oS (by 34%) and from 10 ^oS to the equator (by 30%), and decreased (by 27%) within the SPSG (Fig. 3 B and Supplementary Fig. 3 B).The highest relative abundances of energy metabolism pathways were found at the transition between the SO and the STF. No significant difference were found between the SPSG and PED (Fig. 3 C and Supplementary Fig. 3 C; Wilcoxon tests, $p>0.05$). The relative abundances of nucleotide biosynthesis pathways showed an increase of 11% from 65 ^oS to 52 ^oS (Fig. 3 D). A decrease was then noted northwards before reversing to an increase by 6.7% from 10 ^oS to the equator (Fig. 3 D), a trend which followed PP rates (Fig. 3 A). Although small differences were noted in the mean relative abundances of nucleotide biosynthesis pathways, they were significantly different between all oceanographic provinces (Wilcoxon tests, $p<0.05$; Supplementary Fig. 3 D).

The relative abundance of cell structure and cell wall biosynthesis pathways were significantly different between all oceanographic provinces (Wilcoxon tests, $p<0.05$; Supplementary Fig. 3 F). They showed an increase from 66 ^oS to approximately 20 ^oS (11% between the SO and the SPSG; Fig. 3 F); north of 20 ^oS, the relative abundances of cell structure and cell wall biosynthesis pathways showed a declining trend (Fig. 3 F and Supplementary Fig. 3 F). The

relative abundances of pathways associated with amino-acid biosynthesis were not significantly different between the SO and the STF (Wilcoxon test, p>0.05; 17.9% ±0.003%; Fig. 3 G and Supplementary Fig. 3 G). Amino-acid biosynthesis pathways increased by 6.4% north of the STF, with the highest relative contribution in the SPSG, and decreased north of 10 ^{o}S (Fig. 3 G and Supplementary Fig. 3 G).

The relative abundances of secondary metabolite biosynthesis, cofactor biosynthesis, and vitamin biosynthesis pathways showed similar bimodal latitudinal trends, in which relative abundances decrease from 66 ^{o}S towards the STF, increase north of the STF (with highest relative abundances in the SPSG), and then decrease again towards the equator (Figs. 3 H-K). The relative abundance of secondary metabolite biosynthesis pathways declined by 10% from 66 ^{o}S towards the STF; increased by an average of 8% within the SPSG (compared to the STF), and significantly decreased from 10 ^{o}S towards the equator (Wilcoxon tests, p<0.05; Fig. 3 I and Supplementary Fig. 3 I). Cofactor biosynthesis pathways decreased by 4.5% from 66 ^{o}S to the STF and increased by 7% between the STF and the SPSG (Fig. 3 J, Supplementary Fig. 3 J). Overall, significant trends were observed in cofactor biosynthesis pathways, characterized by relatively higher abundances south of the Polar Front (PF) in the SO, low abundance in the STF, higher abundances in the SPSG and low abundances in the PED compared to the SPSG (Wilcoxon tests, p<0.05; Supplementary Fig. 3 J). The relative abundance of vitamin biosynthesis pathways was significantly different between all provinces (Wilcoxon tests, p<0.05; Supplementary Fig. 3K). These pathways decreased by 15.5% from 66 ^{o}S to 50 ^{o}S (south of the PF), increased by 30% until 30 ^{o}S and then decreased again towards the equator (Fig. 3 K).

The relative abundances of carbohydrates biosynthesis pathways increased from 66 ^{o}S towards the southern edge of the STF (52 ^{o}S), but no significant differences were found between these provinces (Wilcoxon test, p>0.05; Fig. 3 M and supplementary Fig. 3 M). From the northern edge of the STF (40 ^{o}S), carbohydrates biosynthesis pathways declined by 7.2% compared to the average in the SPSG (Fig. 3 M and Supplementary Fig. 3 M). No significant differences were found between the SPSG and the PED (Wilcoxon test, p>0.05 and Supplementary Fig. 3 M). Similar trends were observed for changes in the relative abundances of lipid biosynthesis pathways (Fig. 3 N). Again, the highest contribution of lipid biosynthesis pathways were found in the Southern Ocean and the STF, and no significant differences were found between these two ocean provinces (Wilcoxon test, p>0.05 and Supplementary Fig. 3 N). North of the STF, the relative abundance of lipid biosynthesis pathways declined until ~30 ^{o}N (13.7% compared to the average in the SPSG), increasing afterwards towards the PED (Fig. 3 N and Supplementary Fig. 3 N). Fermentation pathways were found across the whole transect. Highest relative abundances were noted in the SO; the relative abundances declined steadily (30%) from the SO to southern edge of the PED. No significant difference were found between the SPSG and the PED (Supplementary Fig. 2).

Although the above mentioned secondary superclasses were considered as core ecosystem functions, other superclasses were also of importance when distinguishing (significant differences in) main ecological processes within and between each ocean province. The SO was mainly characterized by sulphur-metabolism pathways (which significantly decreased by 12% between the SO and the SPSG; Supplementary Fig. 4 A), and degradation pathways (a proxy for heterotrophy (H6); Supplementary Fig. 4 B-H). Various degradation pathways were found in the SO, among which aromatic compound degradation, carbohydrate degradation, amino acid degradation, denitrification (significant decreases of 135, 56, 88, and 146% between the SO and SPSG, respectively), sugars and acids degradation, sugars and polysaccharides degradation, and alcohol degradation (significant decreases of 150, 29, and 5% between the SO and the STF, respectively) (Supplementary Fig. 4). Pathways characteristic of the STF were those of energy metabolism and reactions associated with methanogenesis (supplementary Fig. 5). Cofactor-biosynthesis pathways (Fig. 3 M) were characteristic of the SPSG, while pathways associated with phosphorus compounds were found observed predominantly in the PED (Wilcoxon test, p>0.05 and supplementary Fig. 6).

Predictive modelling

The predictive capability of Boosted Regression Tree (BRT) modelling was used to estimate which biotic and abiotic parameters influenced metabolic pathways and PP at Depth 1 (n=75). Latitudinal changes in CO_2-fixation and energy pathways, including PP, were strongly predicted by concentrations in nanoplankton and chl-b (Table 1). These two parameters contributed up to 47, 50, and 58%, respectively, of the models predicting CO_2-fixation pathways, energy pathways, and PP (Table 1 and supplementary Table 3). Temperature, NH_4^+, and zeaxanthin concentrations contributed 53, 20, and 6%, respectively, to the models predicting latitudinal trends in cell structure and cell wall biosynthesis pathways (Table 1 and supplementary Table 3). Amino acid biosynthesis pathways were largely predicted by the abundance of picoplankton and temperature (Table 1). Secondary metabolite biosynthesis pathways were mainly predicted by total pigment concentrations, chl-b, nanoplankton and picoplankton abundances, and C:N ratios (Table 1). Co-factor biosynthesis pathways were largely predicted by total pigment concentrations, chl-b, and salinity (Table 1). Vitamin biosynthesis pathways were mainly predicted by total pigment and NO_3^- concentrations and temperature (Table 1). Changes in lipid biosynthesis pathways could largely be predicted by the total pigment and PO_4^{3-} concentrations, divinyl-chl-*a* concentrations, and changes in salinity and temperature (Table 1). Carbohydrate biosynthesis pathways were primarily (up to 67%) predicted by changes in temperature (Table 1). Fermentation pathways were mainly predicted by temperature, dissolved oxygen concentrations and micro- and picoplankton concentrations (Table 1 and Supplementary Table 3).

Table 1: Shows the relative influence of 22 predictor variables in explaining variation in the latitudinal trends for the core ecosystem functions, including PP and CO_2-Fixation pathways at Depth 1 (*n* = 75) along the p15S GO-SHIP transect determined by BRT. The relative contribution of each parameter is shown in brackets (%). Summing the contribution for each parameter for each pathway results in a contribution >60% to the model predictability.

Pathway	Parameter 1	Parameter 2	Parameter 3	Parameter 4	Parameter 5	cv
PP	[chl-b] (37%)	[Nanoplankton] (20%)	[Total pigment] (9%)			0.917±0.02
CO_2-fixation	[Nanoplankton] (25%)	[chl-b] (23%)	Temperature (16%)			0.878±0.03
Energy-Metabolism	[chl-b] (30%)	[Nanoplankton] (20%)	[Microplankton] (6%)	[NH_4^+] (5%)		0.752±0.04
Nucleotide Biosynthesis	Salinity (20%)	Temperature (20%)	δ13C (16%)	[Silicate] (6%)		0.892±0.019
Cell-Structure Biosynthesis	Temperature (53%)	[NH_4^+] (20%)	[Zeaxanthin] (6%)			0.955±0.01
Amino acid biosynthesis	[Picoplankton] (38%)	Temperature (32%)				0.896±0.03
Secondary Metabolite Biosynthesis	[Total pigment] (20%)	chl-b (13%)	[Nanoplankton] (11%)	[Picoplankton] (10%)	C:N ratios (7%)	0.857±0.03
Co-factor Biosynthesis	[Total pigment] (39%)	chl-b (8%)	Salinity (7%)	[Silicate] (6%)		0.808±0.05
Vitamin Biosynthesis	[Total pigment] (26%)	[NO_3^-] (24%)	Temperature (14%)			0.948±0.01
Lipid Biosynthesis	[Total pigment] (39%)	[PO_4^{3-}] (10%)	[divinyl-chl-*a*] (9%)	Salinity (7%)		0.926±0.01
Carbohydrate Biosynthesis	Temperature (67%)	[Total pigment] (7%)				0.946±0.01
Fermentation	Temperature (36%)	[Oxygen] (10%)	[Microplankton] (10%)	[Picoplankton] (7%)		0.944±0.01

Discussion

This study set out to test six hypotheses under the assumption that 16S rRNA sequences can offer significant insight into microbial ecological functions in oceanographic studies. The metabolic reconstructions showed strong latitudinal gradients in the microbial metabolic pathways which coincide geographically with the four oceanographic provinces and are in agreement with the current mechanistic understanding of microbial biogeography in the South Pacific Ocean [7, 17] (confirming H1, Fig. 1B). More specifically, the trends in microbial functions mirrored the latitudinal variation in physico-chemical and biological measurements (which included temperature; nutrient bioavailability; diagnostic pigments, such as fucoxanthin, peridinin, chl-b, and zeaxanthin; and the isotopic fractionation of particulate organic nitrogen). This shows that our 16S rRNA sequencing data can complement previously published datasets on biomass and rate measurement for microbial communities in the South Pacific Ocean and, thus, improve our understanding of ecological changes observed across this basin. We were also able to find evidence which supported H2-H6, as described in more detail in the following sections.

Primary productivity shapes ecological functions of the bacterial community (H2)

Our second hypothesis, based upon the fact that primary productivity (PP) is stimulated by increases in nutrient concentrations at frontal zones, proposed that the types of microbial pathways (and overall functional community structure) in frontal systems should be related to biomass production. Raes, van de Kamp [17] showed that, within the GO-

SHIP P15S transect, the SO and the SPSG were areas of low PP, whereas the STF and the PED had relatively high PP (Fig. 1 B, 3 A). The authors also observed an important trend along this transect: a switch from net autotrophy (i.e. high C-fixation) in the STF to heterotrophy (i.e. high nitrification and degradation of organic matter) in the SO [17]. In the oligotrophic SPSG we observed high NH_4^+ assimilation rates, an increase in the δ^{15}N-PON, and a higher abundance of picoplankton, which suggests an active microbial loop in this region. All pathways associated with CO_2-fixation, energy metabolism, and nucleotide biosynthesis showed similar latitudinal trends, which aligned with variations in PP (Fig. 3 A, B, C). In our model predictions, concentrations of nanoplankton and chl-b alone explained 58% of the latitudinal changes in PP, but also explained ~50% of the latitudinal trends for CO_2-fixation and energy pathways (Fig. 1 B and 3, A, B and C). Besides supporting H2, these results also support previous findings that PP (energy production) is a main driver for archaeal and bacterial richness across frontal boundaries in the South Pacific Ocean [7].

Temperature-regulated growth and nutrient limitation (H3 and H4)

Our third hypothesis was formulated upon the concept that micro-organisms follow a second order growth curve with temperature [19, 20]. We recorded an increase in the percentage of the bacterial community which displayed cell structure and cell wall biosynthesis pathways as temperature -a proxy for latitude- increased (Fig. 4 E). However, we also found that NH_4^+ and the concentration of photosynthetic prokaryotes, together with temperature, explained ~80% of the variance in cell structure and cell wall biosynthesis pathways in our models (Supplementary Table 3). This is in agreement with previously published data for the tropical (i.e., high temperatures), oligotrophic SPSG, which was characterized by relatively higher NH_4^+ uptake rates and picoplankton concentrations as well as organic matter with higher δ^{15}N [17]; indications of a food web that is dominated by high turnover of organic material (greater cell wall biosynthesis). It seems, thus, that the presence and relative abundance of cell growth-related pathways is not only directly related to temperature, but also to picoplankton abundance.

Our fourth hypothesis postulated that the inferred metabolic predictions would result in latitudinal trends which reflect microbial strategies in coping with trace metal and macro nutrient limitations. As expected, we observed bimodal latitudinal trends for the biosynthesis of secondary metabolites, co-factors (which were identified as key pathways), and vitamins, which we related to trace metal and co-nutrient limitations. In the more productive regions (STF and PED), expression of the above-mentioned pathways was lower relative to the least productive regions. Iron supply to the SO is limited, causing the region to have low chlorophyll production despite its high nutrient concentrations[22]. Similarly, inorganic macronutrient concentrations in the oligotrophic SPSG are at or below the detection limit in surface waters [23]. The above-mentioned biosynthesis pathways thus represent strategies used by the microbial community to cope with (essential) micro and macro nutrient limitation.

Energy storage and degradation (H5 and H6)

Our fifth hypothesis suggested that pathways associated with energy storage, such as lipid and carbohydrate synthesis, would be most active in the SO and in the STF. Lipids and carbohydrates are structurally essential molecules and important energy sources [29], such that, when nutrients are abundant, microorganisms allocate C to lipid biosynthesis [30]. The SO and STF are highly seasonal environments where the strong differences in light availability between the winter and summer profoundly impact lipid trophodynamics, such that energy stored during the light season becomes critical for survival in the darker months [24, 31]. Our study was conducted during the beginning of winter in the SO and the STF, when the remainder of the biomass produced during the light season is consumed and, most likely, allocated to energy storage in the form of lipids. This could explain the high relative abundance of lipid biosynthesis pathways in the bacterial community at the higher latitude (the SO and STF; Fig. 3 K, L). North of the STF in the tropical region the seasons are not as distinguishable, and the conditions are oligotrophic (Fig. 1 B and C). This means that organic matter

is likely to be rapidly recycled and taken up for cell growth (e.g. high NH_4^+ uptake in the SPSG see Raes, van de Kamp [17]), which may explain the rapidly declining lipid and carbohydrate biosynthesis pathways observed north of the STF (Fig. 3 K, L). Another possible explanation for the declining trend in lipid biosynthesis pathways might be related to the bioavailability of PO_4^{3-} [32]. It has been shown that, in PO_4^{3-}-deficient environments (such as the SPSG; Fig. 1C), heterotrophic bacteria and photosynthetic prokaryotes (picocyanobacterial) are able to engage in lipid remodelling (substituting phospholipids with alternative, non-phosphorus lipids, such as sulfolipids or glycolipids[28]), a strategy which increases their survival at an evolutionary scale in oligotrophic areas of the ocean [28, 33]. As this lipid remodelling is expressed at a community level, the shift in trends of the metabolic pathways might give insight regarding how bacterial communities cope with PO_4^{3-}-limitation in the South Pacific Ocean.

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 [17]). Relative to the productive STF we measured lower inorganic C assimilation rates in the Southern Ocean, during the onset of the darker winter months (Fig. 1 B). Our result also supports those of Manganelli, Malfatti [34], who concluded that bacteria and archaea are the most important producers of organic particles via heterotrophic production (organic degradation) when light availability is reduced at higher latitudes during the winter. Overall, these results independently confirm that C-based degradation pathways are indeed key functions in the Southern Ocean during winter.

Other Important pathways

The SO is also a hotspot for sulphur cycling, in particular the production of dimethylsulfoniopropionate (DMSP; as shown by the high presence of *Phaeocystis* in Raes, Bodrossy [7] and Sow, Trull [35])) and the climate cooling dimethylsulfide gas (DMS; Berresheim [36] and Sheehan and Petrou [37]). Recent work by Landa, Burns [38], who surveyed 1.4 million bacterial genome equivalents from the Tara Ocean's dataset found that 1 in 5 of those genomes have the capacity to use DMSP. Members of the SAR11 and Planktomarina genera are known DMSP degraders ; they were dominant bacteria in our data set in the SO [7], explaining the higher relative abundance and presence of sulfur metabolising genes in the SO (Supplementary Fig. 4 A). Sulphur pathways declined significantly north of the STF but were still detectable up to the equator (Supplementary Fig. 4 A). These findings are in agreement with the study from Landa, Burns [38], which claims that a large range of marine bacteria have the capacity to use dissolved organic sulfur metabolites, and that the sulphur metabolites play an important part in the global pelagic ocean.

Fermentation pathways were included in the ten core ecosystem functions, which showed the highest relative abundances in the SO but were present across the entire transect (Supplementary Fig. 2). The anaerobic degradation of organic material (including fermentation) contributes significantly to the degradation processes in (marine) sediments [39]. Because fermentation is favoured under anoxic environments, studies targeting the potential of this process in the photic zone are absent to the best of our knowledge. 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 [40, 41, 42]. These microhabitats offer niches for a diverse range of metabolic pathways [43], and the anoxic zones within marine snow particles could potentially harbour fermentative bacteria. We do note that the indication of fermentation pathways, however, could also be an artefact due to the presence of inactive sulphate-reducing bacteria and methanogenic archaea, which are capable of fermenting under favourable environmental conditions [39].

Considerations on the use of 16S rRNA 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 metabolic activity. Douglas, Maffei [15] clearly noted two main criticisms on functional estimates based on 16S rRNA amplicon-based hidden-state predictions. The first is that the predictions are obviously biased towards the available reference genomes, a limitation which will be partially addressed as the number of metagenome-assembled genomes (MAGs), and sequenced genomes in general, continues to increase. The second criticism is that the 16S rRNA-based predictions do not provide the necessary resolution to detect biogeographic pattern of ecotypes of interest, such as shown by Brown, Lauro [44] for the pelagibacter SAR11.

We should note two examples from our results that clearly illustrate these limitations of amplicon-based functional estimates. First, the metabolic pathways N_2 fixation and nitrification, which have been shown to be important in the South Pacific Ocean[17], were not statistically important, and therefore not present, in the PICRUSt2 MetaCyc outputs. This is likely because N_2 fixation is not well resolved by 16S rRNA tag sequencing (e.g., Gaby and Buckley [45]) and because bacteria involved in nitrification made up only 1% of the bacterial biomass (see supplementary Fig. 9 in Raes, van de Kamp [17]). Neglecting these metabolic pathways that contribute to new (N_2 fixation) and regenerated (nitrification) inputs of N will contribute to high uncertainties when estimating the f-ratio [46, 47], and thus, consequently lead to an underestimation when modelling global oceanic primary productivity. Secondly, our analyses do not provide the necessary resolution to detect biogeographic patterns of ecotypes of interest. The pelagibacter SAR 11, the cyanobacteria *Prochlorococcus* and the prymnesiophyte *Phaeocystis,* to name some examples, show ecological and evolutionary distinct ecotypes that are vertically and horizontally partitioned throughout the water column [35, 44, 48, 49]. While these ecotypes 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.

The limitations noted above can, however, be addressed with additional existing tools in molecular ecology. For example, using this dataset, it was possible to combine amplicon-based analysis of the functional gene associated with N_2 fixation (*nifH*) and direct rate measurements to reveal biogeographic patterns in the capacity and occurrence of these metabolic pathways[17]. Similarly for this dataset, focused taxonomic analysis of the 18S rRNA data revealed fine-scale ecotype distributions of the haptophyte *Phaeocystis* sp. [35]

As with any endeavour in science, the tool a researcher chooses to use depends on the question being asked. Within the toolbox available to molecular ecologists, amplicon-based functional derivation 1) uses low cost, high throughput sample collection and standard analyses and 2) yields the information regarding ecosystem function that is required to construct biogeochemical and ecological models. Amplicon-based functional estimates yield functional information at the same breadth (i.e. basin-scale) and depth (i.e. a wide range of metabolic functions) as global ocean monitoring campaigns such as GO-SHIP and bioGEOTRACES. Mapping ecosystem function at this scale provides the opportunity to investigate the relationship between biodiversity and functional diversity, whether diversity drives productivity or the other way around, and whether biochemical resource limitation sets the ultimate control on productivity and ecosystem resilience. On a geological timescale, an analysis of 13-million-year-long nanoplankton abundance time series suggests that ecological functions, rather than species richness, are more important to community resilience and biochemical functions [50]. Deriving functional profiles from 16S rRNA datasets obtained by oceanic sampling programs on a global scale may 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.

Conclusion

Our study demonstrated that the metabolic reconstructions based on the marker gene 16S rRNA resulted in latitudinal trends in microbial functions which were consistent with physico-chemical parameters, results of laboratory experiments, and our overall mechanistic understanding of functional microbial biogeography in the South Pacific Ocean. The strong latitudinal trends in the derived metabolic pathways suggest that not only taxonomic composition, but also functional diversity, is affected by oceanographic boundaries. The validation of the derived ecosystem functions from 16S rRNA tag sequencing for the South Pacific Ocean showcases the potential for mapping ecosystem function in global monitoring campaigns such as GO-SHIP and bioGEOTRACES. Community-level functional information directly speaks to the state of, and changes in, ecosystems, while also complementing the information provided by physico-chemical parameters, which are more routinely used to monitor the state of the oceans. The ability to query metabolic pathways in existing and future 16S rRNA datasets on a global scale establishes the ability to test hypotheses regarding how biodiversity influences functional diversity, and how these are related to energy production in the ocean.

Declarations

Competing interests

The authors declare no competing interests.

Data availability

Genomic data are available at https://www.ncbi.nlm.nih.gov/bioproject/385736 under the accession no. PRJNA385736. All physical, biogeochemical, and meta- data are available at the CLIVAR and Carbon Hydrographic Data Office (CCHDO; https://cchdo.ucsd.edu/; GO-SHIP transect P15S expocode: 096U20160426).

Author contributions: EJR analysed data. KK, SLSS, LB, RMFS and AMW contributed to data analyses and data interpretation. EJR, LB, JvdK, MO and MB contributed to the study design. All authors contributed to the writing of the manuscript.

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 P15 GO-SHIP transect voyage number IN2016_V03. We would also like to thank Bernhard Tschitschko, Nicole Gail Hellessey, 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 Alfred Wegener Institute and the University of Western Australia. This work was supported by Australian Research Council awards DP150102326 to M.V.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 in an earlier draft.

Material And Methods

Study region and water sampling

Our oceanographic study was conducted in the south Pacific Ocean onboard the *R.V. Investigator* from 23 April - 29 June 2016 along the longitudinal P15S GO-SHIP line at 170^{o} W (Fig. 1A). The P15S GO-SHIP line is a transect that runs from the ice edge (~66^{o} S) to the equator (0^{o} S; Fig. 1;http://www.go-ship.org/). In this study we build further upon the work along the P15S GO-SHIP transect from Raes, Bodrossy [7] and Raes, van de Kamp [17]). For clarity we briefly reiterate some of the material and methods, but for an in-depth explanation on the physical and bio-chemical data validation and the presented C and N rate measurements we refer the reader to the aforementioned papers and the supplementary material. The oceanographic transect could be divided into four oceanographic provinces or Longhurst provinces [51]. From South to North the transect covered: 1) the Southern Ocean between 66° - 52° S; 2) the Subtropical Frontal zone (STF) between 52° - 40° S; 3) the South Pacific Subtropical Gyre (SPSG) between 40° - 10° S, and 4) the Pacific Equatorial Divergence Province (PED), between 10° - 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 (DO) were conducted using a Seabird (SBE25 plus) conductivity-temperature-depth (CTD) profiler with a SBE43 O_2 sensor mounted on a 36 Niskin bottle rosette sampler.

DNA sequencing and bioinformatics

Samples for DNA analyses were collected from 12 L Niskin bottles at three depths in the upper Ocean. Samples for Depth 1 ranged between 1.3 and 36.7 m (with an of average 6.6±4.1 m (±standard deviation (sd)). Depth 2 had a depth range between 18.6 and 85.8 m (average 35.3±13.1 m), and samples from Depth 3 ranged between 39.9 and 185 m (average 70.7±18.0 m; Fig. 1 C). A peristaltic pump was used to filter two litres of seawater through a 0.22-µm pore size Sterivex^{TM} filter (catalogue no. SVGPL10RC; Millipore, Germany). Samples were stored at -80 ^{o}C until DNA extraction. DNA was isolated from the filters following a modified organic (phenol:chloroform:isoamyl-alcohol-based) extraction protocol [52] of the PowerWater® Sterivex^{TM} 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 and reverse 519R primer sets (Lane, Pace [53] and Lane [54]) using the Illumina MiSeq^{Tm} platform. 16S rRNA amplicons were generated using 300-bp paired-end sequencing at the Ramaciotti Centre for Genomics (University of New South Wales, Sydney). Amplicon sequence variant (ASV) tables were prepared after Bissett, Fitzgerald [55] and as outlined in the supplementary material. 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; see supplementary material for the full workflow).

Functional composition

PICRUSt2 (see Langille, Zaneveld [14], Douglas, Beiko [56] and Douglas, Maffei [15] was used to infer an approximate functional potential of the microbial communities across the 7000 km transect in the South Pacific Ocean. We ran the PICRUSt2 pipeline (version picrust/2.3.0b) with default settings, and sequences with NSTI scores >2 were removed. The average NSTI score, based on 387 samples (covering three depths), was 0.145±0.031 (±standard deviation). 140 of the 4411 OTUs (a total of 0.3%) were above the maximum NSTI cut-off of 2 and were removed. The 0.3% of OTUs which were removed contributed to 3.5% of the relative abundance of the bacterial community. One OTU, which showed a 99% match (440bp/441bp) with *Bathycoccus prasinos* mitochondrial DNA, contributed to 2.5% of bacterial community relative abundance. The relative abundance of each OTU (including its sequence at the 97% similarity threshold) with NSTI values > 2 is shown in supplementary Table 4. Pathways with less than 10 reads were removed, which resulted in the removal of two pathways (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, Maffei [15].

Statistical Analysis

The Phyloseq package (v.1.28.0; McMurdie and Holmes [57]) was used to analyse, visualize and plot the microbiome and metadata along the P15S GO-SHIP transect. Statistical tests were conducted using the Vegan package version 2.5-6 [58] in R version 3.6.1 [59] and the PRIMER v7 software [60]. The plyr package (v.1.8.4) was used to calculate means and summarise the data [61]. The ggpubr package [62] and Wilcoxon tests [63] were used to test for differences between ocean provinces and the Bonferoni method was used to control for multiple tests. The Ocean Data View (ODV; v.5.0.0.) software package was used to visualize the physical and bio-chemical data. Analysis of similarities (ANOSIM) was used to test whether we could identify statistical differences between the oceanographic provinces based on the sequence data and the MetaCyc pathway data. ANOSIM tests were performed on Bray-Curtis dissimilarity matrices generated from square-root-transformed data. The 'indicspecies' package (ver. 1.7.8; De Cáceres [64]) with the 'multipatt' function and 9999 permutations was used to identify indicator pathways in the four oceanographic provinces. Pearson's phi coefficients of association [65] were calculated to determine significant indicator pathways in the four oceanographic provinces. The phi coefficient was corrected using the function="r.g." to accommodate for the fact some oceanographic provinces had more stations than others [66]. Twenty-two biotic and abiotic parameters were used to predict trends in the various metabolic pathways including primary productivity rates using boosted regression trees (BRT; Elith, Leathwick [67]). BRTs were implemented in the R v.3.6.1 software environment using the gbm.step function, a gaussian error structure, 10 fold cross-validation, and the following settings learning.rate=0.001, tree.complexit =10, bag.fraction=0.5 [67]. All models resulted in > 2050 trees.

References

1. Falkowski PG. Evolution of the nitrogen cycle and its influence on the biological sequestration of CO2 in the ocean. *Nature* **387**, 272 (1997).
2. Sunagawa S, *et al.* Structure and function of the global ocean microbiome. *Science* **348**, 1261359 (2015).
3. Blasiak R, *et al.* The Ocean Genome: Conservation and the Fair, Equitable and Sustainable Use of Marine Genetic Resources.). Blue Paper (2020).
4. 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).
5. Biller SJ, *et al.* Marine microbial metagenomes sampled across space and time. *Scientific Data* **5**, 180176 (2018).
6. Salazar G, *et al.* Gene expression changes and community turnover differentially shape the global ocean metatranscriptome. *Cell* **179**, 1068–1083. e1021 (2019).
7. Raes EJ, *et al.* Oceanographic boundaries constrain microbial diversity gradients in the South Pacific Ocean. *Proceedings of the National Academy of Sciences* **115**, E8266-E8275 (2018).
8. Methé BA, *et al.* A framework for human microbiome research. *Nature* **486**, 215–221 (2012).
9. Thompson LR, *et al.* A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* **551**, 457 (2017).
10. Clarke LJ, *et al.* Mainstreaming Microbes across Biomes. *BioScience* **70**, 589–596 (2020).
11. Brumfield KD, Huq A, Colwell RR, Olds JL, Leddy MB. Microbial resolution of whole genome shotgun and 16S amplicon metagenomic sequencing using publicly available NEON data. *PLOS ONE* **15**, e0228899 (2020).
12. Karl DM, Church MJ. Microbial oceanography and the Hawaii Ocean Time-series programme. *Nature Reviews Microbiology* **12**, 699–713 (2014).
13. Brown MV, *et al.* Systematic, continental scale temporal monitoring of marine pelagic microbiota by the Australian Marine Microbial Biodiversity Initiative. *Scientific data* **5**, 180130 (2018).
14. Langille MG, *et al.* Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature biotechnology* **31**, 814 (2013).

15. Douglas GM, *et al.* PICRUSt2 for prediction of metagenome functions. *Nat Biotechnol* **38**, 669–688 (2020).
16. Sun S, Jones RB, Fodor AA. Inference-based accuracy of metagenome prediction tools varies across sample types and functional categories. *Microbiome* **8**, 1–9 (2020).
17. Raes EJ, *et al.* N2 Fixation and New Insights Into Nitrification From the Ice-Edge to the Equator in the South Pacific Ocean. *Frontiers in Marine Science* **7**, (2020).
18. Floodgate GD, Fogg GE, Jones DA, Lochte K, Turley CM. Microbiological and zooplankton activity at a front in Liverpool Bay. *Nature* **290**, 133–136 (1981).
19. Eppley RW. Temperature and phytoplankton growth in the sea. *Fish bull* **70**, 1063–1085 (1972).
20. Thomas MK, Kremer CT, Klausmeier CA, Litchman E. A global pattern of thermal adaptation in marine phytoplankton. *Science* **338**, 1085–1088 (2012).
21. Hoppe H-G, Gocke K, Koppe R, Begler C. Bacterial growth and primary production along a north–south transect of the Atlantic Ocean. *Nature* **416**, 168–171 (2002).
22. Boyd PW, *et al.* A mesoscale phytoplankton bloom in the polar Southern Ocean stimulated by iron fertilization. *Nature* **407**, 695–702 (2000).
23. Browning TJ, *et al.* Nutrient co-limitation at the boundary of an oceanic gyre. *Nature* **551**, 242–246 (2017).
24. Hagen W, Auel H. Seasonal adaptations and the role of lipids in oceanic zooplankton. *Zoology* **104**, 313–326 (2001).
25. Rivkin RB, Legendre L. Biogenic carbon cycling in the upper ocean: effects of microbial respiration. *Science* **291**, 2398–2400 (2001).
26. Talley LD, Pickard GL, Emery WJ, Swift JH. Chapter 10 - Pacific Ocean. In: *Descriptive Physical Oceanography (Sixth Edition)* (eds Talley LD, Pickard GL, Emery WJ, Swift JH). Academic Press (2011).
27. Ganachaud A, *et al.* The Southwest Pacific Ocean circulation and climate experiment (SPICE). *Journal of Geophysical Research: Oceans* **119**, 7660–7686 (2014).
28. Van Mooy BA, Rocap G, Fredricks HF, Evans CT, Devol AH. Sulfolipids dramatically decrease phosphorus demand by picocyanobacteria in oligotrophic marine environments. *Proceedings of the National Academy of Sciences* **103**, 8607–8612 (2006).
29. Parrish CC. Lipids in marine ecosystems. *ISRN Oceanography* 2013, (2013).
30. Kattner G, Gercken G, Hammer K. Development of lipids during a spring plankton bloom in the northern North Sea: II. Dissolved lipids and fatty acids. *Marine Chemistry* **14**, 163–173 (1983).
31. Phleger CF, Nichols PD, Virtue P. Lipids and trophodynamics of Antarctic zooplankton. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* **120**, 311–323 (1998).
32. DyhrMaN ST, Ammerman JW, Van Mooy BA. Microbes and the marine phosphorus cycle. *Oceanography* **20**, 110–116 (2007).
33. Sebastián M, *et al.* Lipid remodelling is a widespread strategy in marine heterotrophic bacteria upon phosphorus deficiency. *The ISME Journal* **10**, 968–978 (2016).
34. Manganelli M, Malfatti F, Samo TJ, Mitchell BG, Wang H, Azam F. Major Role of Microbes in Carbon Fluxes during Austral Winter in the Southern Drake Passage. *PLOS ONE* **4**, e6941 (2009).
35. Sow SLS, Trull 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. *Frontiers in Microbiology*, (in press).
36. Berresheim H. Biogenic sulfur emissions from the Subantarctic and Antarctic Oceans. *Journal of Geophysical Research: Atmospheres* **92**, 13245–13262 (1987).
37. Sheehan CE, Petrou K. Dimethylated sulfur production in batch cultures of Southern Ocean phytoplankton. *Biogeochemistry* **147**, 53–69 (2020).

38. Landa M, *et al.* Sulfur metabolites that facilitate oceanic phytoplankton–bacteria carbon flux. *The ISME journal* **13**, 2536–2550 (2019).
39. Kirchman DL, Hanson TE, Cottrell MT, Hamdan LJ. Metagenomic analysis of organic matter degradation in methane-rich Arctic Ocean sediments. *Limnology and oceanography* **59**, 548–559 (2014).
40. Woebken D, Fuchs BM, Kuypers MM, Amann R. Potential interactions of particle-associated anammox bacteria with bacterial and archaeal partners in the Namibian upwelling system. *Applied and Environmental Microbiology* **73**, 4648–4657 (2007).
41. Ploug H. Small-scale oxygen fluxes and remineralization in sinking aggregates. *Limnology and Oceanography* **46**, 1624–1631 (2001).
42. Klawonn I, Bonaglia S, Brüchert V, Ploug H. Aerobic and anaerobic nitrogen transformation processes in N 2-fixing cyanobacterial aggregates. *The ISME journal* **9**, 1456–1466 (2015).
43. 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. *Limnology and Oceanography: Methods* **16**, 484–503 (2018).
44. Brown MV, *et al.* Global biogeography of SAR11 marine bacteria. *Molecular systems biology* **8**, 595 (2012).
45. Gaby JC, Buckley DH. A comprehensive aligned nifH gene database: a multipurpose tool for studies of nitrogen-fixing bacteria. *Database (Oxford)* 2014, bau001-bau001 (2014).
46. Dugdale R, Goering J. Uptake of new and regenerated forms of nitrogen in primary productivity 1. *Limnology and oceanography* **12**, 196–206 (1967).
47. Eppley RW, Peterson BJ. Particulate organic matter flux and planktonic new production in the deep ocean. *Nature* **282**, 677 (1979).
48. Martiny AC, Coleman ML, Chisholm SW. Phosphate acquisition genes in Prochlorococcus ecotypes: evidence for genome-wide adaptation. *Proceedings of the National Academy of Sciences* **103**, 12552–12557 (2006).
49. Braakman R, Follows MJ, Chisholm SW. Metabolic evolution and the self-organization of ecosystems. *Proceedings of the National Academy of Sciences* **114**, E3091-E3100 (2017).
50. Alvarez SA, Gibbs SJ, Bown PR, Kim H, Sheward RM, Ridgwell A. Diversity decoupled from ecosystem function and resilience during mass extinction recovery. *Nature* **574**, 242–245 (2019).
51. Longhurst AR. *Ecological geography of the sea*. Elsevier (2010).
52. Appleyard SA, Abell G, Watson R. *Tackling microbial related issues in cultured shellfish via integrated molecular and water chemistry approaches*. CSIRO Marine and Atmospheric Research (2013).
53. Lane DJ, Pace B, Olsen GJ, Stahl DA, Sogin ML, Pace NR. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proceedings of the National Academy of Sciences* **82**, 6955–6959 (1985).
54. Lane D. 16S/23S rRNA sequencing. Nucleic acid techniques in bacterial systematics (Stackebrandt E & Goodfellow M, eds).). Wiley, New York (1991).
55. Bissett A, *et al.* Introducing BASE: the Biomes of Australian Soil Environments soil microbial diversity database. *GigaScience* **5**, 21 (2016).
56. Douglas GM, Beiko RG, Langille MG. Predicting the functional potential of the microbiome from marker genes using PICRUSt. In: *Microbiome Analysis*). Springer (2018).
57. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PloS one* **8**, (2013).
58. Oksanen J, *et al.* The vegan package. *Community ecology package* **10**, 631–637 (2007).
59. Team RC. R: A language and environment for statistical computing. (2013).

60. Clarke K, Gorley R. Primer. *PRIMER-e, Plymouth*, (2006).
61. Wickham H. The split-apply-combine strategy for data analysis. *Journal of Statistical Software* **40**, 1–29 (2011).
62. Kassambara A. ggpubr:"ggplot2" based publication ready plots. *R package version 01* **6**, (2017).
63. Wilcoxon F, Katti S, Wilcox RA. Critical values and probability levels for the Wilcoxon rank sum test and the Wilcoxon signed rank test. *Selected tables in mathematical statistics* **1**, 171–259 (1970).
64. De Cáceres M. How to use the indicspecies package (ver. 1.7. 1). *R Proj* **29**, (2013).
65. Chytrý M, Tichý L, Holt J, Botta-Dukát Z. Determination of diagnostic species with statistical fidelity measures. *Journal of Vegetation science* **13**, 79–90 (2002).
66. Tichy L, Chytry M. Statistical determination of diagnostic species for site groups of unequal size. *Journal of Vegetation Science* **17**, 809–818 (2006).
67. Elith J, Leathwick JR, Hastie T. A working guide to boosted regression trees. *Journal of Animal Ecology* **77**, 802–813 (2008).

Figures

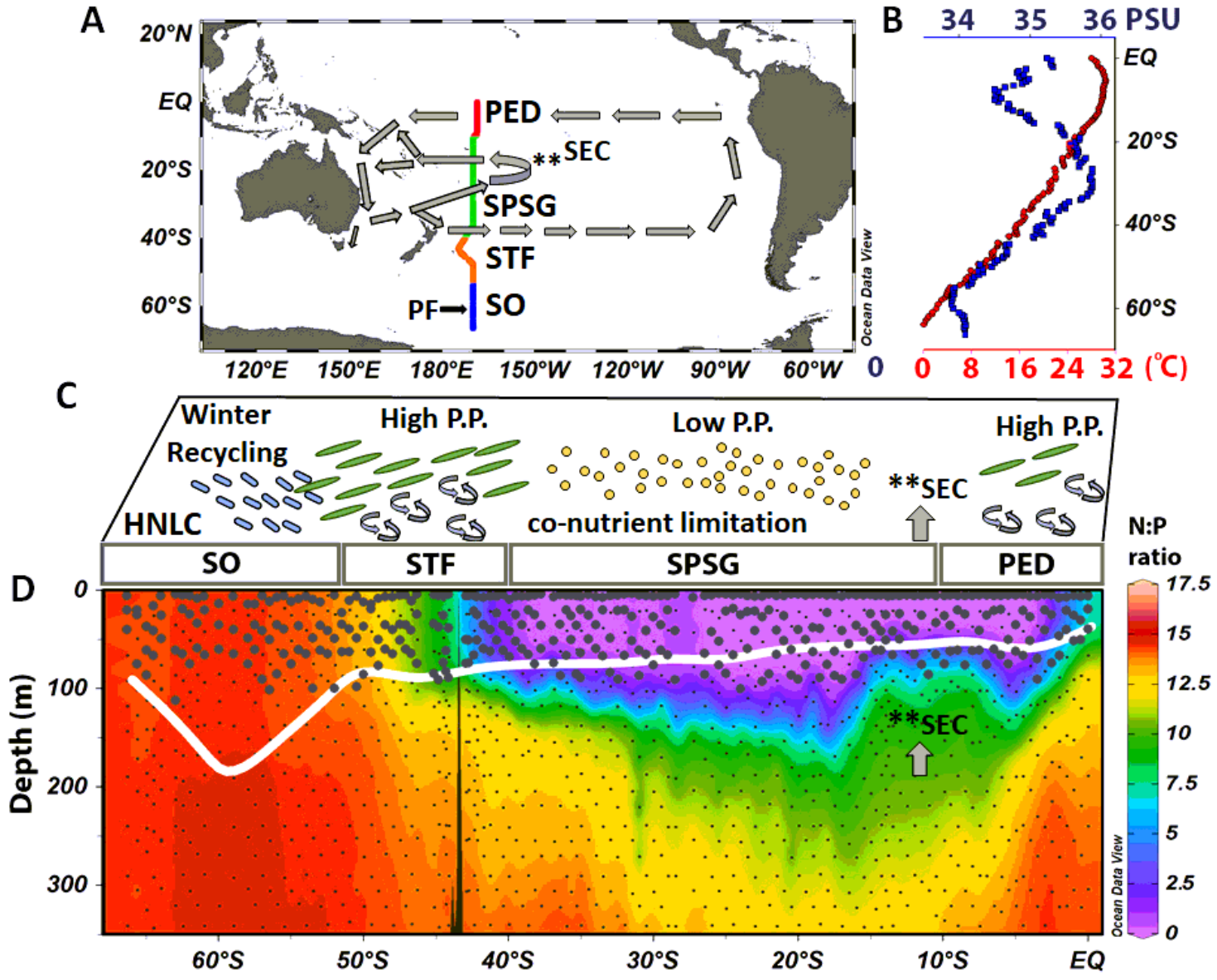

Figure 1

(A) The GO-SHIP P15S transect along the 170o W meridian in the South Pacific Ocean. Grey arrows indicate the direction of the wind-driven South Equatorial Current (SEC) within the upper Ocean, and include a westward returning branch between 20 and 10 °S (denoted by **SEC; based on Talley, Pickard 26 and Ganachaud, Cravatte 27 under non La Niña or non El Niño events). The four oceanographic provinces described by Longhurst (2010) are colour coded: the Southern Ocean (SO; blue sector; 66° S - 52°), the Subtropical Frontal zone (STF; orange sector; 52° S - 40° S), the South Pacific Subtropical Gyre Province (SPSG; green sector; 40° S - 10° S), and the Pacific Equatorial Divergence Province (PED; red sector; 10° S - 0° S). The Polar Front (PF) in the Southern Ocean is indicated by the black arrow at 60° S. (B) Profiles for sea surface temperature (red circles) and salinity (blue squares). (C) Conceptual mechanistic understanding of the functional prokaryotic and microbial-eukaryotic biogeography within and across ecological provinces along the transect in the South Pacific Ocean. Temperature above the mixed layer depth (MLD) is shown as a red gradient. 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) is expected in the STF due to turbulence and mixing (indicated by curved arrows); phytoplankton containing high concentrations of chlorophyll-a (chl-a) and chlorophyll-b (chl-b) are indicated by green discs. The oligotrophic SPSG is characterized by low PP and nutrient co-limitation, as well as by higher abundances of photosynthetic prokaryotes (represented as closed yellow circles). The westward returning branch of the SEC (**SEC) is indicated with the grey arrow, and can be characterized by an increase in the N:P ratio at 100 m depth. Equatorial upwelling and mixing (also shown as curved arrows) result in an increase of the N:P ratio at the surface and, thus, in increased PP rates and in chl-a and chl-b concentrations. (C) Latitudinal plot of N:P ratios from the surface to 350 m depth along the transect. The thick white line represents MLD, and sampling stations are shown in large grey circles (DNA sampling) and in small grey dots (CTD sampling). The location of the boundaries between the four oceanographic provinces in (C) coincides with the acronyms in the bottom part of (D).

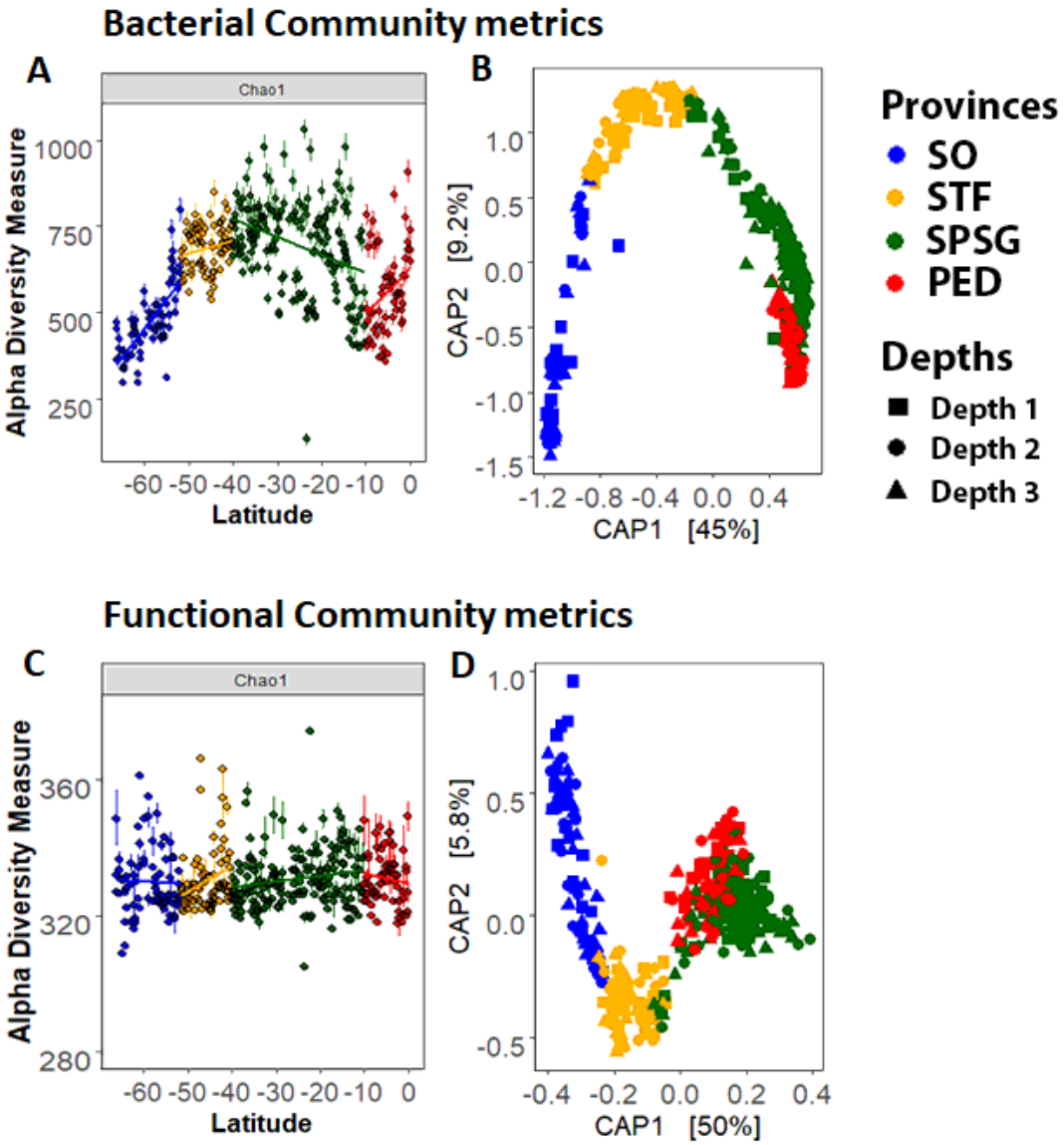

Figure 2

Bacterial community (OTUs clustered at 97% similarity; upper panels) and functional community (MetaCyc pathways) metrics: Alpha diversity shown as Chao diversity (A, C) and beta-diversity shown as CAP plots (B, D). Linear models were fitted through the data points on (A) and (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 (1.3 – 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.

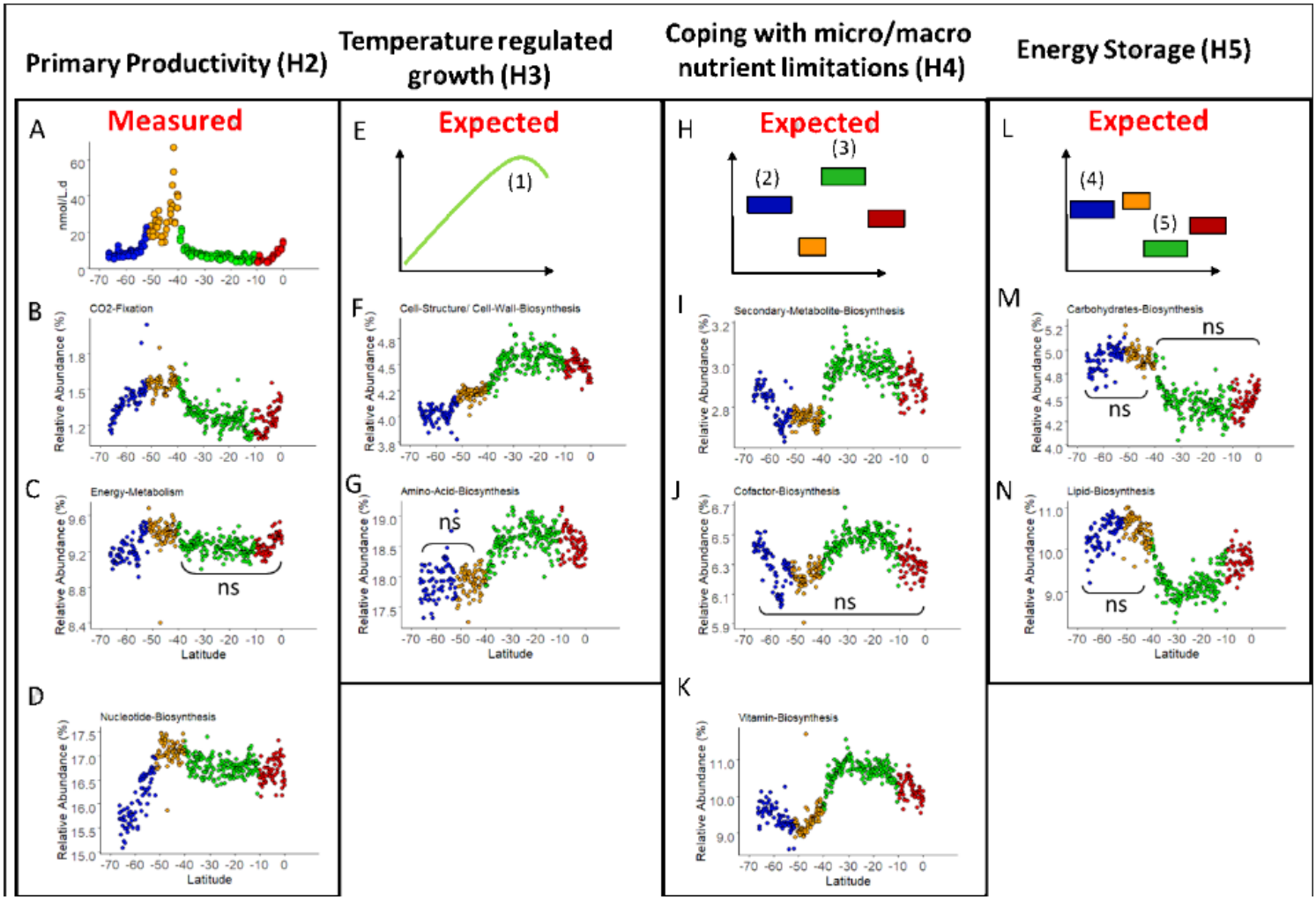

Figure 3

Latitudinal trends for the core ecosystem functions (except for fermentation) and for CO2-fixation pathway. Each column shows pathways associated with a process approached in our hypotheses: PP (H2, left column), temperature-regulated growth (H3, second column from the left), coping strategies for micro and macro nutrient limitation (H4, second column from the right), and energy storage (H5, right column). Graphs in the upper row present measured data for PP (A) and expected values for the other processes (D, G, and J, respectively) according to (1) Eppley 18, (2) Boyd, Watson 21, (3) Browning, Achterberg 22, (4) Rivkin and Legendre 24, (5) Van Mooy, Rocap 28. Graphs in the middle and lower rows present relative abundance data obtained from 16S rRNA tag sequencing for the ten core ecosystem functions. Significant differences were observed between provinces unless otherwise indicated by 'ns'.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Raesetal.2020SupplementaryMaterialFunctions.docx
- SupplementaryTable2MetaCycPathwaysGOSHIPP15S.csv
- SupplementaryTable3BRTmodeloutputs.csv
- SupplementaryTable4OTU97sequenceswithNSTIvaluesgreaterthan2rar20163Picrust2.3.0b.csv