On site analysis of bacterial communities of the ultra-oligotrophic South Pacific Gyre

Greta Reintjes[1#], Halina E. Tegetmeyer[1, 2, 3], Miriam Bürgisser[1], Sandi Orlić [4,5], Ivo Tews [6], Mikhail Zubkov[7], Daniela Voß[8], Oliver Zielinski[8], Christian Quast[1], Frank Oliver Glöckner[1, 9], Rudolf Amann[1], Timothy G. Ferdelman[1], Bernhard M. Fuchs[1]

Max Planck Institute for Marine Microbiology, Bremen, Germany(1); Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany(2); Center for Biotechnology – CeBiTec, Bielefeld University, Bielefeld, Germany(3); Institut Ruđer Bošković, Zagreb, Croatia(4); Center of excellence for Science and Technology – integration of Mediterranean region – STIM, Zagreb, Croatia(5); Biological Sciences, Institute for Life Sciences (IfLS), University Southampton, Southampton, United Kingdom(6); National Oceanography Centre, Southampton, United Kingdom(7); Institute for Chemistry and Biology of the Marine Environment, University Oldenburg, Germany(8); Jacobs University, Bremen, Germany(9).

Running title: Bacterial communities of the South Pacific Gyre

Address correspondence to Greta Reintjes, Max Planck Institute for Marine Microbiology, Celsiusstr. 1, 28359, Bremen, Germany. Tel: +49 421 2028 935, Fax: +49 421 2028 790, email: greintje@mpi-bremen.de

Keywords: Microbial community structure, AEGEAN-169, *Prochlorococcus*, South Pacific Subtropical Gyre, Biogeography.

Abstract

The South Pacific Gyre (SPG) covers 10% of the ocean's surface and is often regarded as a marine biological desert. To gain an on-site overview of the remote, ultra-oligotrophic microbial community of the SPG we developed a novel on-board analysis pipeline, which combines next-generation sequencing with fluorescence *in situ* hybridisation and automated cell enumeration. We tested the pipeline during the SO245 "UltraPac" cruise from Chile to New Zealand and found that the overall microbial community of the SPG was highly similar to that of other oceanic gyres. The SPG was dominated by 20 major bacterial clades, including SAR11, SAR116, AEGEAN-169 marine group, SAR86, *Prochlorococcus*, SAR324, SAR406, and SAR202. Most of the bacterial clades showed a strong vertical (20 m – 5000 m), but only a weak longitudinal (80°W – 160°W), distribution pattern. Surprisingly, in the central gyre *Prochlorococcus*, the dominant photosynthetic organism, had only low cellular abundances in the upper waters (20 - 80 m) and were more frequent around the 1% irradiance zone (100 - 150 m). Instead, the surface waters of the central gyre were dominated by SAR11, SAR86, and SAR116 clades known to harbour light-driven proton pumps. The alphaproteobacterial AEGEAN-169 marine-group was particularly abundant in the surface waters of the central gyre indicating a potentially interesting adaptation to ultraoligotrophic waters and high solar irradiance. In the future, the newly developed community analysis pipeline will allow for on-site insights into a microbial community within 35 hours of sampling, which will permit more targeted sampling efforts and hypothesis-driven research.

Importance

The South Pacific Gyre is due to its vast size and remoteness one of the least studied oceanic regions on earth. However, both remote sensing and *in situ* measurements indicated that the activity of its microbial community contributes significantly to global biogeochemical cycles. Presented here is an unparalleled investigation of the microbial community of the SPG

from 20 - 5000 m depth covering a geographic distance of ~7000 km. This insight was achieved through the development of a novel on-board analysis pipeline, which combines next-generation sequencing with fluorescence *in situ* hybridisation and automated cell enumeration. The pipeline is well comparable to on-shore systems based on the Illumina platforms and yields microbial community data in less than 35 hours after sampling. Going forward the ability to gain on-site knowledge of a remote microbial community will permit hypothesis-driven research, through the generation of novel scientific questions and subsequent additional targeted sampling efforts.

Introduction

Oligotrophic gyres are vast ocean biomes which represent 60% of the oceans and 40% of the Earth's surface. The largest of these gyres is the South Pacific Gyre (SPG) which has a total area of 37 million km^2 and represents ~10% of the oceans total area (328 $x10^6$ km^2, (1)). The SPG is a unique, ultra-oligotrophic habitat, which has some of the clearest waters ever reported, near undetectable surface nitrate concentrations and the lowest sea surface chlorophyll-*a* concentrations (0.023 nmole l^{-1}) (2, 3). Although often defined as ultra-oligotrophic and a "biological desert", estimates of the SPG's contribution to global biogeochemical cycles show that it plays a significant role in global carbon and nitrogen cycling (4-7). This contribution is calculated primarily from remote sensing data obtained via satellites because, due to its large size, the SPG has received only limited direct scientific attention (8). The two major SPG expeditions (BIOSCOPE and Ocean Drilling Program (IODP) Expedition 329) have shown that while the waters are ultra-oligotrophic, there is still a considerable amount of microbial activity, specifically carbon and nutrient cycling (4-6).

Although microorganisms appear to be major players in the SPG our understanding of their abundance and distribution patterns is very limited. The few available studies have begun to highlight the abundance of specific clades (4, 9, 10) or the community composition at individual depths (6, 11), but due to the diverse array of methodologies applied in these

studies, it is difficult to obtain a comprehensive picture. This shortcoming hinders our ability to draw conclusions about the potential metabolic contributions of individual microorganisms to biogeochemical cycles in the SPG.

The primary reason for the lack of microbial data of the SPG is due to its vast size and remoteness; few scientific expeditions have traversed it due to high expedition costs. Moreover, in-depth analyses of microbial communities of remote sampling sites, such as the SPG, are hindered by on-board methodological limitations. Specifically, unlike direct measurements (temperature, salinity), samples for microbial ecology cannot readily be analysed on-site and need to be preserved for later analysis in a lab. Realistically the analysis of these samples occurs only weeks to months after sampling. Furthermore, in-depth follow-up studies must wait for future sampling campaigns, which may take years due to site remoteness and limitations in project funding. The consequence of this discrepancy between sampling and obtaining results are costly “shot in the dark” sampling efforts and prevents targeted on-site experimentation.

The goal of our study was two-fold. Firstly, we wanted to address the lack of on-site microbial diversity and abundance analyses by developing a mobile, high-throughput sequencing and data analysis pipeline and combining this with fluorescence in situ hybridisation (FISH) and on-board automated cell counting (12). The pipeline should have the capacity to quickly and inexpensively give a comprehensive insight into a microbial community on-site. Secondly, we tested and operated the newly developed pipeline on board the RV Sonne during the SO-245 “UltraPac” cruise, in the SPG from the Chilean upwelling waters (-84°E), crossing the centre of the oligotrophic gyre, to the coast of New Zealand (-159°E, 7000 km).

99 **Results**

100 **On-board Next Generation Sequencing (NGS)**

101 During the SO-245 cruise, we investigated the bacterial diversity, composition and
102 abundance of the SPG using a newly designed field-based analysis pipeline, which functions
103 even under challenging conditions, such as ship-board pitch and roll movements (e.g. at
104 Station 12 max wave height: 3.1 m, max heave 4.7°, max pitch 5.2°, max roll angle 8.5°).

105 A total of 147 samples were taken from multiple depths at 11 stations, during the SO-
106 245 cruise, to validate the pipeline's capacity for on-site diversity profiling (Supplementary
107 Table S1a). The samples were examined directly on board the RV Sonne to test the individual
108 steps of the pipeline and to ensure a high number of high-quality reads could be obtained
109 within the shortest possible time. We advanced previous efforts of on-board sequencing by
110 optimising each step from DNA extraction to data processing (Supplementary Table S2) and
111 by addressing the previously encountered issues of unexpected equipment failure due to
112 transportation and computational limitations (13).

113 The pipeline had a minimum sampling requirement of 10^7 cells l^{-1} and DNA could be
114 extracted from all 147 samples with an average concentration of 4.2 ng μl^{-1}. The DNA
115 concentration was proportional to the TCC, with lower DNA concentrations obtained from
116 deeper waters (3000 – 5000 m; 2.1 x 10^4 cells ml^{-1}; 0.5 ng μl^{-1}) and the highest DNA
117 concentrations acquired from above the DCM (75 - 100 m; 6 x 10^5 cells ml^{-1}; 7.5 ng μl^{-1},
118 Table 1).

119 Sequencing was performed on an Ion Torrent PGM platform, which was selected due
120 to its physical robustness, compact dimensions, and because it has previously been
121 highlighted as a suitable platform for on-board sequencing (13). In total, 1100 Mbp were
122 sequenced on board the RV Sonne, which equated to 3.9 million reads with a median read
123 length of 290 bases. On-board sequencing was tested using multiple chip types (Ion V2 314,

316, 318) and raw data processing methods (default and stringent, Table 1). There was no difference in the read quality between chip types, but more stringent quality trimming decreased in the total number of bases (by 30%) and the total number of reads (47%). Additionally, a more stringent processing method resulted in quicker processing times and increased the mean read length from 278 bp to 368 bp (Table 1).

A major issue encountered in previous attempts at producing a remote sequencing pipeline were field-based computational issues. We specifically addressed this by developing a novel offline version of the SILVAngs pipeline (14) on a dedicated mobile server (Supplementary Table S2). Before the cruise, the server was pre-installed with all necessary software and tested in controlled settings (see Methods). For validation, two mock community data sets were analysed on both the SILVAngs online web service and the newly developed offline version of SILVAngs using the SILVA 16S rRNA database (SSU REF 123) as a reference (14). For the offline server the alignment of the rRNA in the SSUref123 database was shortened to match with the amplified 16S region to increase the classification speed (see Methods). Cluster analysis showed that samples analysed by the normal and modified SILVAngs versions yielded highly similar results (Supplementary Figure S1). Mantel tests showed no significant difference between the community compositions of the two systems ($R = 0.996$, $P = 0.001$ based on 1000 permutations).

All 147 sequencing samples were processed on the SILVAngs offline server on board the RV Sonne, which equated to a total of 3.2×10^7 reads. For all stations, a minimum of 3,500 reads per sample were obtained. The median read abundance for main stations was ~24,000 and for intermediate stations ~8,800 reads (Table 1). A higher sequencing depth was obtained for the main stations to test the SILVAngs offline server ability to classify rare bacterial populations.

Overall the field-based sequencing and data analysis pipeline yielded equivalent results, at similar costs, to previous lab-based investigations. Results could be obtained within

30 – 34 hours of sampling (Supplementary Table S3). The cost of DNA extraction, PCR, size selection and sequencing using our field-based pipeline was approximately 450 euro for a single run yielding 60-100 Mb (400 bp reads, not including machine or personnel costs).

On-board Microbial Abundance Profiling

Although 16S rRNA tag sequencing provides an in-depth insight into the composition of a microbial community, it is semi-quantitative and, consequently, does not provide a comprehensive interpretation of a microbial community (15). Knowing on-site if a target organism is present or absent and, even more, its absolute cellular abundance and vertical distribution improves sampling and experimental efforts particularly for cultivation, metagenomics or single cell analyses (16, 17). We, therefore, combined our newly developed on-board sequencing pipeline with the high-throughput image acquisition and cell enumeration system described by C. M. Bennke et al. (12). We applied the cell enumeration pipeline, in parallel to the sequencing pipeline, during the SO-245 cruise to the total and relative microbial abundance in 257 samples from 15 stations at various depths (Supplementary Table 1b).

Absolute abundances of particular bacterial clades were determined using specific FISH probes, which were selected based on the prior acquired sequencing results (10). By combining the two methods, the specificity and coverage of each FISH probe could be tested before FISH; preventing unnecessary, labour intensive FISH procedures. One limitation of the combined approach is that on-board FISH analysis can only be done using previously described probes, which are selected based on prior studies. For the “unknown” clades, such as AEGEAN-169 in this study, there were no available probes and new specific probes needed to be designed. The counts for such “unknown” clades cannot be performed directly on board.

Physicochemical properties of the SPG

The SO-245 "Ultra-Pac" cruise crossed through the oligotrophic "eye" of the SPG (Figure 1a & b). The most pronounced changes in physicochemical conditions occurred in the top 500 m of the SPG (Supplementary Figure S2a, b, c). The central gyre region (stations 4 – 9; 100°W – 120°W) had characteristically high surface water temperatures between 20 - 25°C and there was virtually no chlorophyll fluorescence measurable in the surface waters down to 70 m. At station 6 (110°W), marking the very centre of the gyre, temperature peaked at 24.9°C at the surface and was 19.9°C at 200 m depth (Figure 1b). There, the deep chlorophyll maximum (DCM) descended down to a maximum depth of 190 – 200 m with 0.5 µg l^{-1} fluorescence. Along the transect, chlorophyll fluorescence was highest in the surface waters at station 14 (160°W; 1.9 µg l^{-1}), indicating increase in primary productivity towards New Zealand (Supplementary Figure S2e & f). The depth of the euphotic layer, representing the depth where downward photosynthetic available radiation (PAR, as defined from 400 - 700 nm) irradiance is reduced to 1% of its surface value, varied between 162 m in the SPG (stations 4 - 9), 110 m for station 1, and 69 m for station 14 (Figure 1c). The DCM depths are below the euphotic layer for all SPG stations, but within the 1% irradiance layer if only blue light (430 – 490 nm) is considered (down to 210 m for the SPG, data not shown). Below 500 m the physicochemical parameters stayed relatively consistent across the SPG, except in the oxygen profile which showed the extent of the well-documented oxygen minimum zone (OMZ) within the water column (Supplementary Figure S2, (18, 19)).

Niche Partitioning in the Bacterial Community of the SPG

The bacterial community composition of the SPG was highly similar across a geographic distance of ~7000 km (Figure 2a, Table 2a), but showed a significant change with depth, which could be directly correlated to the change in light availability (Figure 2b & c and Table 2b & c). Correspondingly, the total cellular abundance decreases with the decrease in

available light (Figure 1c). It was higher in the surface waters (top 200 m; 1 x 10^6 to 2 x 10^5 cell ml^{-1}) and decreased to 7 x 10^4 cell ml^{-1} by depth of 500 m; below which it stayed relatively constant (Figure 1c, Supplementary Figure S3 a & b). The highest total cell counts (TCC) of the SPG were found just above the 1% irradiance zone at 90°W (1.1 x 10^6 cells ml^{-1}) and at 40 m depth at 139°W (9.2 x 10^5 cells ml^{-1}, Figure 1c). In the centre of the gyre (100°W – 120°W) there were 3.9 x 10^5 cells ml^{-1} in the surface waters and this increased to ~5 x 10^5 cells ml^{-1} at 100 m depth (Figure 1c).

There were 20 dominant bacterial clades within the SPG with a relative read abundance of >0.5% in at least two stations (Figure 3b). These clades showed a distinct distribution with depth; having a higher read (determined by sequencing) and cellular abundance (determined for 8 clades by FISH) either in the euphotic zone or below the euphotic zone (Figure 3, Table 3 and Supplementary Figure S3). In the euphotic zone (0 - 150 m) members of the SAR86, SAR11 surface group 4 and 1, SAR116, *Rickettsiales* S25 593, *Ascidiaceihabitans*, *Prochlorococcus*, *Rhodobacteraceae* and AEGEAN-169 marine group had high relative read abundances (Figure 3b). The AEGEAN-169 marine group had an abundance of 3 - 6% (1.6 x 10^4 cells ml^{-1}, determined by FISH) throughout the surface water (top 100 m), with a particularly high relative abundance in the top 20 m of the centre of the gyre (Figure 3a). Contrastingly, the SAR86 group was more abundant (3 – 5%, 1.7 x 10^4 cells ml^{-1}, determined by FISH) in the surface waters (top 100 m) outside of the central gyre (Figure 3a). SAR11 was enumerated using a clade specific FISH probe and therefore exhibited a relative abundance of 10 - 50% (average 2 x 10^5 cells ml^{-1}) throughout the upper water column, which decreased slightly with depth and towards the eastern end of the transect (Figure 3a, Supplementary Figure S3).

The most significant changes in bacterial composition occurred in the 1% irradiance zone (Figure 3b), where there was a decrease in abundance in the euphotic clades and an increase in the mesopelagic clades. Additionally, the phototrophic bacteria, *Prochlorococcus*

and *Synechococcus* exhibited a distinct distribution profile around the 1% irradiance zone. *Prochlorococcus* was present in high abundance (5 – 30%, 7.9 x 104 cell ml^{-1}, determined by FISH) throughout the top 250 m (Figure 3a) and remained high in abundance within the 1% irradiance zone, but decreased in abundance just below it (150 – 200 m, Figure 3a). Comparatively, *Synechococcus* was low in read abundance (not counted by FISH) in the surface waters and the 1% irradiance zone and increased only below the peak of *Prochlorococcus* at 150 – 250 m depth (Figure 3b).

Below the euphotic and 1% irradiance layer (aphotic: 150 – 5125 m) the well-known mesopelagic bacterial clades: SAR324, SAR406, SAR202, *Sulfitobacter* and Sva0996 marine group, increased in relative and absolute abundance (Figure 3b & c, (27-29)). In addition to the bacterial clades with high read abundances, there was also a large rare bacterial community throughout the SPG (relative read abundance < 0.1%). About 550 clades had a low relative abundance (<0.5%) and were detectable at only a few sites (3 or less). Whereas 120 clades had a low abundance (<0.5%) but were ubiquitously present. These ubiquitous but rare clades were predominantly from the *Verrucomicrobia* (*Puniceicoccaceae*), *Planctomycetes*, *Deltaproteobacteria* and the *Bacteroidetes* (*Flavobacteriaceae*) (Supplementary Figure S5).

Discussion

Oligotrophic gyres cover vast areas of the Earth's surface and contribute, due to microbial carbon and nitrogen cycling, significantly to global biogeochemical cycles (4, 5, 7). However, our current understanding of the abundance and distribution patterns of the microbial community of the largest of these gyres, the SPG, is still limited due to both infrequent sampling and lack of on-site community analysis. Therefore, during the SO245 cruise, we developed an on-board microbial community analysis pipeline which enabled the on-site sequencing of 147 samples and enumeration of 275 samples by FISH. The outcome of

our method developments is a readily applicable system for an efficient, cost-effective, field-based, comprehensive microbial community analysis.

Picoplankton community of the SPG

Using our newly established pipeline, we discovered that the microbial community of the SPG showed a pronounced vertical distribution pattern. The community changed significantly in composition with depth, which was directly correlated to the availability of light (Figure 2 and Table 2). Such a noticeable vertical distribution has also been observed in other oceanic gyres (North Pacific-, South Atlantic- and Northern Atlantic Gyres, (25, 30-33)) and was linked to the significant changes in the physicochemical conditions related to depth: changes in temperature, nutrient concentrations, availability of light and the availability of labile organic matter (34-36).

The euphotic surface waters of the central gyre were extremely limited in inorganic macronutrients and especially in nitrogen salts (3, 7). The low nutrient availability restricts growth to specialist oligotrophic organisms, which was reflected by the low cellular abundance in the surface mixed layer (4×10^5 cells ml^{-1}, Figure 1c). Dominant clades were *Prochlorococcus*, SAR11, SAR116, SAR86 and the AEGEAN-169 marine group (Figure 3b) all of which, except for the AEGEAN-169 marine group, are well documented to be optimised for an oligotrophic lifestyle (37-40). Cultured and genome-sequenced representatives of these clades are also reported to have streamlined genomes and specialised resource acquisition abilities (41-43).

Additionally, *Prochlorococcus*, SAR11, SAR86 and SAR116 are equipped with the genetic potential for photosynthesis or phototrophy via proteorhodopsins (37, 44, 45). Although our current knowledge of the genetic potential of the AEGEAN-169 marine group is limited (33, 46-48); their high cellular abundance, of up to 3×10^4 cells ml^{-1} in the surface waters of the central gyre indicates a specialised oligotrophic lifestyle. Previous studies have

highlighted some of the potential factors affecting the AEGEAN-169 marine group distribution patterns (47), but contrastingly found them to have a high relative abundance in deeper waters (500 m, 33). A possible explanation for these dissimilarities is the presence of multiple ecological species of the AEGEAN-169 marine group (47). Future metagenomic studies of these organisms are required to examine the importance of this abundant clade in the most oligotrophic surface waters in the SPG.

Prochlorococcus, the dominant primary producer in oligotrophic ocean regions (49, 50), was also the most abundant autotrophic organism in the surface waters of the SPG. However, in comparison to studies in the Atlantic Gyres, its absolute abundance in surface waters was low (Figure 3a, (25, 51, 52)). Interestingly, the abundance of *Prochlorococcus* increased with depth and peaked between 100 – 150 m in and around the 1% irradiance zone. The low abundance of *Prochlorococcus* in the surface waters of the SPG could be an indication that low nutrients, high solar irradiance or a combination of both inhibits its growth (9, 53, 54).

The measured chlorophyll fluorescence in the surface waters of the SPG was below the detection limit in our study (Figure 1c), although previous studies measured up to 0.017 µg l^{-1} (55). Chlorophyll fluorescence peaked deep in the water column around 200 m and could be measured down to nearly 300 m depth (Supplementary Figure S2f). Light availability as indicated by the 1% irradiance layer was maximal in the central waters of the SPG, reaching down to 162 m. The deep penetration of blue light in the water column of the SPG, down to 210 m, indicated that the light conditions were suitable for photosynthetic activity even at these depths. Similar chlorophyll measurements taken in the North and South Atlantic Gyres show comparable fluorescence profiles, although the depth of the DCM in the Atlantic is considerably higher in the water column (120 - 165 m) than in the SPG and surface waters are not entirely depleted in chlorophyll (25, 56).

In the mesopelagic zone of the SPG, where light became limiting (57), there was a distinct change in the microbial community dominated by SAR11 surface clade 1, SAR86 and *Prochlorococcus* to SAR324, SAR406 and SAR202. Although there are currently no cultured representatives of these three bacterial groups, metagenomic analyses have revealed some insight into their possible metabolic capabilities. SAR202 and SAR324 have been associated with carbon, and sulphur oxidation (20, 21, 58, 59) and are likely chemolithoautotrophs ubiquitous in the dark oceans. In particular, SAR324 has also hypothesised to degrade the lipid chains of chlorophyll *a*, which may explain its increased abundance below the DCM (60). Interestingly, in the mesopelagic zone, a novel and so-far undescribed group called SVA0996 of the *Actinobacteria* was found highly represented in the 16S rRNA tag reads. Because of its abundance, this group could be of interest in future studies.

We designed and optimised an on-board sequencing and data analysis pipeline that enabled us to obtain on-site microbial community diversity results of the SPG within 34 hours of sampling. In surface waters, the community was dominated by a few key oligotrophic organisms, which are adapted to extreme physicochemical conditions. The ability to obtain "direct" insights into the microbial diversity, even at extremely remote oligotrophic sampling sites, enables the close examination of novel discovered microbial clades, such as the AEGEAN-169 marine group in the surface waters or the SVA0996 group (this study) in the deeper water layers. Additionally, and most importantly, it allows microbial ecologists to perform a more targeted sampling, thereby furthering our understanding of the diversity and metabolic capabilities of key microorganisms.

Materials and Methods

Sampling

Seawater samples were collected aboard the RV Sonne during the „UltraPac" cruise (SO-245) from Antofagasta, Chile (17.12.2015) to Wellington, New Zealand (28.01.2016).

Water samples were taken from a total of 15 stations (Supplementary Table S1) using a Seabird sbe911+ CTD (Seabird Scientific, WA, USA), attached to a SBE32 Carousel Water Sampler containing 24x12-liter bottles. Two types of stations were sampled: main stations and intermediate stations. On main stations, the CTD was cast through the entire water column to 50 - 100 m above the seafloor and samples were taken at various depths throughout the water column (61). Generally, 4 to 5 CTDs were cast to reduce the time between sampling at depth and processing of the samples. Intermediate stations consisted of a single CTD cast down to 500 m and samples were taken from variable depths (61). For diversity analysis, a total of 1 L of seawater was sampled. The water was directly filtered onto a 47 mm polycarbonate filter (0.2 µm pore size) using a bottle top Nalgene filter holder (Thermo Fisher, MA, USA) and a vacuum pump. After filtration samples were immediately used for DNA extraction.

Physicochemical Data

Physicochemical characteristics were examined using a CTD (Sea-Bird Electronics Inc. SBE 911plus probe). The system was equipped with: double temperature (SBE 3) and conductivity probes (SBE 4), a pressure sensor (Digiquartz), an oxygen sensor (SBE 43), an altimeter (Bentos) and a chlorophyll fluorometer combined with a turbidity sensor (WET Labs ECO-AFL/FL). The sensors were pre-calibrated by the manufacturers. The data were recorded with the Seasave V7.23.1 software and processed using SeaBird SBE Data Processing software. Data were despiked, and also visually checked. The ship position was derived from the shipboard GPS-system linked to the CTD data. The time zone is given in UTC. Salinity was quality checked by reference samples [n=30], measured with an Optimare Precision Salinometer (OPS S/N 004) 5 month after the cruise. All CTD data was obtained from and is available on Pangaea (www.pangaea.de, (57)). The physicochemical data was visualised using the ODV4 software (www.odv.awi.de).

Underwater light field was measured utilizing a HyperPro II Profiler (Satlantic Inc., Canada) according to the procedures described in (62). For these measurements, the profiler was lowered into the water at least 30 m behind the vessel to avoid ship shadowing when free falling. A downward irradiance reference sensor was mounted at an elevated, non-shaded location. Profiler data processing and calculation of photosynthetic available radiation were performed with ProSoft 7.7.16 (Satlantic Inc., Canada).

DNA Extraction, Polymerase Chain Reaction (PCR), Size Selection and Quantification

Each step from DNA extraction to data processing was selected and optimised to achieve a high level of high quality reads in the shortest possible processing time. The advantages and disadvantages of each processing step are highlighted in Supplementary Table S2. The final optimal protocol for the SPG study is described below.

DNA extractions were done using the MoBio Power Water DNA Extraction Kit (MoBio Laboratories, Inc., CA, USA) as recommended by the manufacturer. PCR was carried out using the Platinum PCR SuperMix High Fidelity polymerase kit (Thermo Fisher), using the primers S-D-Bact-0341-b-S-17 and S-D-Bact-0785-a-A-21 targeting the V3-V4 variable region of the 16S rRNA, evaluated by (63). Both primers were fusion primers with additional adaptor and barcode sequences at the 5' end to allow sequencing and separation of samples in down-stream analyses. The reverse primers contained the Ion tr-P1 adaptor at the 5' end of the primer and the forward primers contained both the Ion A adaptor and one of 40 IonXpress barcodes (Ion Xpress 1 - 40) as well as the key sequence (GAT) before the primer. Reverse fusion primer sequence: (5'- CCTCTCTATGGGCAGTCGGTGAT GACTACHVGGGTA TCTAATCC-3'). Forward fusion primer sequence: (5'- CCATCTCATCCCTGCGTGTCTCCGACTCAG XXXXXXXXXX GAT CCTACGGGNGGCWGCAG-3') [XXXXXXXXXX = barcode sequence 1...40]. After

amplification the PCR amplicons were size selected using Agencourt AMPure XP (BeckmanCoulter, Krefeld, Germany).

A successful sequencing reaction requires precise quantities of the template library to ensure a clonal amplification on individual Ion Sphere Particles (ISPs) (see below). A fragment analyser (AATI) was used to determine the quality and quantity of the extracted DNA, the size selected PCR products and the final sequencing pools. Genomic DNA was analysed using the DNF – 488 high sensitivity genomic DNA analysis kit (AATI, size range from 50 bp to 40,000 bp). All template libraries and final sequencing pools were analysed using the DNF - 472 standard sensitivity NGS kit sizing DNA (AATI, size range from 25 bp – 5,000 bp and up to a minimum of 0.1 ng μl^{-1}) as recommended by the manufacturer. The fragment was adapted to ship movements by adding magnets to the individual sample trays, thereby preventing the accidental dropping of a sampling tray caused by ship pitches, during plate movement or at the “on hold” position inside the tray drawers. The internal plate lift was mechanically stabilised for ship movement and vibration by the installation of an additional guide rail on the upper side connected via rubber mounts. Additionally, a specialised stand with transport handles and attachments was applied for easy manual transport and to allow for secure attachment to a surface (Supplementary Figure S6 a-e).

Ion Torrent Sequencing and Raw Sequence Processing

The Ion Torrent PGM was adapted for onboard use by securing it to a 2 cm thick polyethylene base plate and the internal hard drives were replaced by SSDs. The base was equipped with handles that could be used for manual transportation of the sequencer and to fix it to the surface (Supplementary Figure S6 f-h). A similar base was fastened to the Ion OneTouch2 Instrument (Thermo Fisher) and Ion OneTouch ES instrument (Thermo Fisher). The Torrent Server (Thermo Fisher) was also adapted to withstand ship-board vibration and

transport by placing it in a custom-made metal frame using rubber mounts (Supplementary Figure S6 f-h)

Sequencing was carried out as recommended by the manufacturer using an Ion Torrent PGM sequencer (Thermo Fisher). Emulsion PCR and enrichment of template-positive ion sphere particles (ISP) was done using the Ion PGM Hi-Q OT2 Kit (Thermo Fisher) on the Ion OneTouch 2 Instrument (Thermo Fisher) and Ion OneTouch ES instrument (Thermo Fisher) following the Ion Torrent user manual. Subsequently, the library fragments (attached to the ISP) were sequenced using the Ion PGM Hi-Q Sequencing Kit (Thermo Fisher) following the user manual on an Ion PGM system (Thermo Fisher). Sequencing was done on Ion 314, 316 and 318 chip Kit v2 (Thermo Fisher) with a total of 1200 flows per sequencing run. The chips vary in their capacity (number of sensors) and therefore total output, run time and processing time. Specifically, the Ion 314 chip has 1.2 M sensors, a total output of up to 100 Mb and a run time of 2 - 4 h. The Ion 316 chip has 6.1 M sensors, an output of up to 1 Gb and runs for 3 - 5 hours. The Ion 318 chip has 11 M sensors, a total output of up to 2 Gb and runs from 4 - 7 hours.

The Torrent Suite software, which converts the raw signals (raw pH values) into incorporation measurements and ultimately into basecalls for each read, was used for initial quality trimming. The standard Torrent Suite settings and more stringent settings were applied. The standard settings and stringent setting were defined in the basecaller arguments of the Torrent Suite Software. Standard: BaseCaller --barcode-filter 0.01 --barcode-filter-minreads 20 --barcode-mode 1 --barcode-cutoff 3 --trim-qual-cutoff 10 --trim-qual-window-size 20 --trim-min-read-len 100. Stringent: Basecaller --barcode-mode 1 --barcode-cutoff 0 --trim-qual-cutoff 15 --trim-qual-window-size 10 --trim-min-read-len 250. Finally the reads were exported as .sff files using the file exporter plugin in the Torrent suite software. The .sff files were split into individual sample FASTA files using mothur version 1.35.1 (64)

(sffinfo()) and analysed using the offline SILVAngs Pipeline called "Lab on a ship" (see below).

All sequence data was deposited in the European Nucleotide Archive (ENA, (65)) using the data brokerage service of the German Federation for Biological Data (GFBio, (66)), in compliance with the MIxS standard (67). The INSDC accession number for the data is: PRJEB39460 and is available under https://www.ncbi.nlm.nih.gov/bioproject/PRJEB39460.

Offline SILVAngs Pipeline "Lab on a ship"

The computer cluster "lab on a ship" was developed to facilitate offline 16S rRNA sequence classification using the SILVAngs pipeline. Previously this was only available using the online platform (14). The benefit of having an offline version is the potential to use it on board a research vessel. To ensure a quick classification system an efficient computing cluster was obtained for the offline analysis and consisted of 4x Intel Xeon E5-4607 6-core, 2.6 GHz, 256 GB RAM mounted on a supermicro X9QR7-TF+ main board (Supermicro, CA, USA), 3 x 480 GB SATA/600 hard disks (Samsung) for fast data read/write processes and 5 x 2 TB SATA3 server RAID hard disks, Ultrastar (HGST, USA) for data storage. The server was installed in a portable 19'' standard rack and placed in the ships' server room.

The server was pre-installed with an offline copy of the SILVAngs pipeline including BLAST (68), ARB software package (69), as well as the SINA aligner (v1.2.11) (70). The complete SILVAngs pipeline can be run using a single command line argument. Additionally, the mothur software (version 1.35.1) (64) and R-Studio with all required packages (71) were installed to offer users further analysis and graphing options. Version 123 of the SILVA (SSU) dataset was used as the classification reference by both the offline and online pipelines. The standard SILVAngs settings for alignment (min. alignment identity 50%, min. alignment score 40, min. base pair score 30%), quality trimming (min. sequence quality 30%, min. length 250, max. ambiguities 2%, max. homopolymers 2%), clustering (CD-Hit Version 4.6,

min. OTU identity 98%) and classification (BLAST version SINA v1.2.10-pre (revision 24275M), similarity 86%) were applied. To increase the speed of the alignment stage, a custom alignment SEED was used by the offline version of the pipeline that uses an alignment trimmed to match the sequence region of the SSU gene. The off-line server cluster enables the classification of 40 sample with an average of 18,000 reads per sample within 3h. To test the "Lab on a ship" server and ensure that similar community composition results are obtained using different quality trimming methods a mock community analysis was done. The mock community samples consisted of a defined number of quality trimmed reads (10,000) obtained from ten Ion Torrent sequenced marine sample from the Atlantic Ocean. The community classification output from the two servers was then compared using cluster analysis.

Statistical Analysis

The interpretation and visualisation of the microbial diversity data was done using normalised genus abundance to site matrices in the R software with the packages Vegan (community ecology package (71)) and Rioja (Analysis of Quaternary Science Data (72)). Normalisation was done using the decostand (method = "total") function of the Vegan software package. For beta diversity analysis and related hypothesis testing, Bray-Curtis dissimilarity matrices of the normalised read abundances of all samples were constructed. Differences in the community structure between sampling sites were analysed by comparing all samples by analysis of similarity (ANOSIM) and visualised in non-metric multi-dimensional scaling plots (NMDS). To test for significant changes in the community composition by longitude, depth and irradiance ANOSIM analyses were performed and visualised using NMDS plots. Subsequently, permutation multivariant analysis of variance (PERMANOVA) with pairwise analyses were performed to identify the amount of variance associated with individual factors.

473 **Total Cell Counts and FISH**

474 DAPI-staining and CARD-FISH were carried out as described in (12, 73). DAPI and
475 FISH stained cells were visualised and counted automatically using a fully automated image
476 acquisition and cell enumeration system (12). FISH probes sequences are listed in Table 3,
477 with their corresponding competitors and helper oligonucleotides, their specificity and
478 formamide concentrations in the hybridization buffer. For this study, a new probe specific for
479 the AEGEAN-169 clade was designed and tested, based on the latest SILVA 16S rRNA
480 database (refnr 128). Total cellular abundances were also determined by flow cytometry
481 (FACSort, Becton Dickinson) as described in Zubkov and Tarran (74).

482 **Acknowledgements**

483 We thank Captain Mallon and the crew of the RV Sonne for assistance at sea. We also
484 thank Andreas Ellrott, Jörg Wulf, Rohan Henkel, Gabriele Klockgether and Gaute Lavik for
485 technical assistance in the lab and at sea. The authors thank Anke Meyerdierks for helpful
486 discussions. This project was funded by the Max Planck Society and the UltraPac Expedition
487 (SO-245) was funded by the Federal Ministry of Education and Research of Germany (Grant
488 03G0245A).

489

490 **Figure Legends**

491 Figure 1. a) Map showing 12 sampling sites in the South Pacific Gyre indicated by black dots.
492 b) Contour plots of temperature (°C) data derived from CTD measurements at 12 stations
493 during the SO245 cruise with depth from 0 - 500 m. c) Contour plot of total cellular
494 abundance enumerated by DAPI staining (cell ml^{-1}) with depth from 0 - 500 m. Dashed white
495 line represents the euphotic layer (m). Also shown is the chlorophyll fluorescence indicated
496 by dark grey (0.5 µg l^{-1}) and light grey (0.25 µg l^{-1}) line. The stations are indicated on the axis
497 below the plots. All data is publicly available from Pangaea:

https://doi.org/10.1594/PANGAEA.870806 (75). All figures were created using the Ocean Data Viewer software (76).

Figure 2. NMDS plots showing Bray-Curtis dissimilarity in community composition a) across longitude (°E), b) by depth (m) and c) by irradiance zone. Each dot represents an individual sample and the communities are coloured coded according to the legends.

Figure 3. Niche partitioning of the bacterial community of the SPG. a) Contour plots of absolute cellular abundance of SAR86, SAR11, AEGEAN-169 and *Prochlorococcus* (cell ml^{-1}) enumerated by FISH with depth from 0 - 500 m. b) Bubble plot showing the relative read abundance (determined by tag sequencing) and depth distribution of the 20 dominant bacterial clades (relative read abundance > 5%) in the SPG. All samples were first sorted by depth and then irradiance zone before plotting. The samples of the 1% irradiance zone are plotted by station. The euphotic, 1% irradiance and aphotic zones are represented by blue shading. c) Contour plots of relative cellular abundance of SAR324, SAR406, and SAR202 (% of total cell counts determined by FISH) with depth from 0 - 5000 m.

References

1. Longhurst A, Sathyendranath S, Platt T, Caverhill C. 1995. An estimate of global primary production in the ocean from satellite radiometer data. Journal of Plankton Research 17:1245-1271.

2. Morel A, Gentili B, Claustre H, Babin M, Bricaud A, Ras J, Tièche F. 2007. Optical properties of the “clearest” natural waters. Limnology and Oceanography 52:217-229.

3. Raimbault P, Garcia N, Cerutti F. 2008. Distribution of inorganic and organic nutrients in the South Pacific Ocean − evidence for long-term accumulation of organic matter in nitrogen-depleted waters. Biogeosciences 5:281-298.

4. Van Wambeke F, Obernosterer I, Moutin T, Duhamel S, Ulloa O, Claustre H. 2008. Heterotrophic bacterial production in the eastern South Pacific: longitudinal trends and coupling with primary production. Biogeosciences 5:157-169.

5. Halm H, Lam P, Ferdelman TG, Lavik G, Dittmar T, LaRoche J, D'Hondt S, Kuypers MM. 2012. Heterotrophic organisms dominate nitrogen fixation in the South Pacific Gyre. ISME J 6:1238-49.

6. Walsh EA, Smith DC, Sogin ML, D'Hondt S. 2015. Bacterial and archaeal biogeography of the deep chlorophyll maximum in the South Pacific Gyre. Aquatic Microbial Ecology 75:1-13.

7. Letscher RT, Knapp AN, James AK, Carlson CA, Santoro AE, Hansell DA. 2015. Microbial community composition and nitrogen availability influence DOC remineralization in the South Pacific Gyre. Marine Chemistry 177:325-334.

8. Claustre H, Huot Y, Obernosterer I, Gentili B, Tailliez D, Lewis M. 2008. Gross community production and metabolic balance in the South Pacific Gyre, using a non intrusive bio-optical method. Biogeosciences 5:463-474.

9. Lami R, Cottrell MT, Ras J, Ulloa O, Obernosterer I, Claustre H, Kirchman DL, Lebaron P. 2007. High abundances of aerobic anoxygenic photosynthetic bacteria in the South Pacific Ocean. Appl Environ Microbiol 73:4198-205.

10. West NJ, Lepere C, Manes CL, Catala P, Scanlan DJ, Lebaron P. 2016. Distinct Spatial Patterns of SAR11, SAR86, and Actinobacteria Diversity along a Transect in the Ultra-oligotrophic South Pacific Ocean. Front Microbiol 7:234.

11. Yin Q, Fu B, Li B, Shi X, Inagaki F, Zhang XH. 2013. Spatial variations in microbial community composition in surface seawater from the ultra-oligotrophic center to rim of the South Pacific Gyre. PLoS One 8:e55148.

12. Bennke CM, Reintjes G, Schattenhofer M, Ellrott A, Wulf J, Zeder M, Fuchs BM. 2016. Modification of a High-Throughput Automatic Microbial Cell Enumeration System for Shipboard Analyses. Appl Environ Microbiol 82:3289-3296.

13. Lim YW, Cuevas DA, Silva GG, Aguinaldo K, Dinsdale EA, Haas AF, Hatay M, Sanchez SE, Wegley-Kelly L, Dutilh BE, Harkins TT, Lee CC, Tom W, Sandin SA, Smith JE, Zgliczynski B, Vermeij MJ, Rohwer F, Edwards RA. 2014. Sequencing at sea: challenges and experiences in Ion Torrent PGM sequencing during the 2013 Southern Line Islands Research Expedition. PeerJ 2:e520.

14. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glockner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 41:D590-6.

15. Props R, Kerckhof FM, Rubbens P, De Vrieze J, Hernandez Sanabria E, Waegeman W, Monsieurs P, Hammes F, Boon N. 2017. Absolute quantification of microbial taxon abundances. ISME J 11:584-587.

16. DeLong EF. 2009. The microbial ocean from genomes to biomes. Nature 459:200-6.

17. Iverson V, Morris RM, Frazar CD, Berthiaume CT, Morales RL, Armbrust EV. 2012. Untangling genomes from metagenomes: revealing an uncultured class of marine Euryarchaeota. Science 335:587-90.

18. Lam P, Lavik G, Jensen MM, van de Vossenberg J, Schmid M, Woebken D, Gutierrez D, Amann R, Jetten MS, Kuypers MM. 2009. Revising the nitrogen cycle in the Peruvian oxygen minimum zone. Proc Natl Acad Sci U S A 106:4752-7.

19. Pinti DL. 2014. Oxygen-Minimum Zone, p 1-2. *In* Amils R, Gargaud M, Cernicharo Quintanilla J, Cleaves HJ, Irvine WM, Pinti D, Viso M (ed), Encyclopedia of

Astrobiology doi:10.1007/978-3-642-27833-4_5164-2. Springer Berlin Heidelberg, Berlin, Heidelberg.

20. Rappé MS, Connon SA, Vergin KL, Giovannoni SJ. 2002. Cultivation of the ubiquitous SAR 11 marine bacterioplankton clade. Nature 418:630-633.

21. Morris RM, Rappé MS, Connon SA, Vergin KL, Siebold WA, Carlson CA, Giovannoni SJ. 2002. SAR 11 clade dominates ocean surface bacterioplankton communities. Nature 420:806-810.

22. Gómez-Pereira PR, Schüler M, Fuchs BM, Bennke C, Teeling H, Waldmann J, Richter M, Barbe V, Bataille E, Glöckner FO. 2012. Genomic content of uncultured *Bacteroidetes* from contrasting oceanic provinces in the North Atlantic Ocean. Environmental Microbiology 14:52-66.

23. West NJ, Schönhuber WA, Fuller NJ, Amann RI, Rippka R, Post AF, Scanlan DJ. 2001. Closely related Prochlorococcus genotypes show remarkably different depth distributions in two oceanic regions as revealed by in situ hybridization using 16S rRNA-targeted oligonucleotides. Microbiology 147:1731-1744.

24. Zubkov MV, Fuchs BM, Burkill PH, Amann R. 2001. Comparison of cellular and biomass specific activities of dominant bacterioplankton groups in stratified waters of the Celtic Sea. Applied and environmental microbiology 67:5210-5218.

25. Schattenhofer M, Fuchs BM, Amann R, Zubkov MV, Tarran GA, Pernthaler J. 2009. Latitudinal distribution of prokaryotic picoplankton populations in the Atlantic Ocean. Environ Microbiol 11:2078-93.

26. Fuchs BM, Woebken D, Zubkov MV, Burkill P, Amann R. 2005. Molecular identification of picoplankton populationsin contrasting waters of the Arabian Sea. Aquatic Microbial Ecology 39:145-157.

595 27. Morris RM, Rappe MS, Urbach E, Connon SA, Giovannoni SJ. 2004. Prevalence of
596 the Chloroflexi-related SAR202 bacterioplankton cluster throughout the mesopelagic
597 zone and deep ocean. Appl Environ Microbiol 70:2836-42.
598 28. Sheik CS, Jain S, Dick GJ. 2014. Metabolic flexibility of enigmatic SAR324 revealed
599 through metagenomics and metatranscriptomics. Environ Microbiol 16:304-17.
600 29. Milici M, Vital M, Tomasch J, Badewien TH, Giebel H-A, Plumeier I, Wang H,
601 Pieper DH, Wagner-Döbler I, Simon M. 2017. Diversity and community composition
602 of particle-associated and free-living bacteria in mesopelagic and bathypelagic
603 Southern Ocean water masses: Evidence of dispersal limitation in the Bransfield
604 Strait. Limnology and Oceanography 62:1080-1095.
605 30. Friedline CJ, Franklin RB, McCallister SL, Rivera MC. 2012. Bacterial assemblages
606 of the eastern Atlantic Ocean reveal both vertical and latitudinal biogeographic
607 signatures. Biogeosciences 9:2177-2193.
608 31. DeLong EF, Preston CM, Mincer T, Rich V, Hallam SJ, Frigaard NU, Martinez A,
609 Sullivan MB, Edwards R, Brito BR, Chisholm SW, Karl DM. 2006. Community
610 genomics among stratified microbial assemblages in the ocean's interior. Science
611 311:496-503.
612 32. Agogue H, Lamy D, Neal PR, Sogin ML, Herndl GJ. 2011. Water mass-specificity of
613 bacterial communities in the North Atlantic revealed by massively parallel
614 sequencing. Mol Ecol 20:258-74.
615 33. Cram JA, Chow CE, Sachdeva R, Needham DM, Parada AE, Steele JA, Fuhrman JA.
616 2015. Seasonal and interannual variability of the marine bacterioplankton community
617 throughout the water column over ten years. ISME J 9:563-80.
618 34. Pakulski JD, Benner R. 1994. Abundance and distribution of carbohydrates in the
619 ocean. Limnology and Oceanography 39:930-940.

620 35. Osterholz H, Niggemann J, Giebel HA, Simon M, Dittmar T. 2015. Inefficient
621 microbial production of refractory dissolved organic matter in the ocean. Nat Commun
622 6:7422.
623 36. Moran MA, Kujawinski EB, Stubbins A, Fatland R, Aluwihare LI, Buchan A, Crump
624 BC, Dorrestein PC, Dyhrman ST, Hess NJ, Howe B, Longnecker K, Medeiros PM,
625 Niggemann J, Obernosterer I, Repeta DJ, Waldbauer JR. 2016. Deciphering ocean
626 carbon in a changing world. Proc Natl Acad Sci U S A 113:3143-51.
627 37. Giovannoni SJ, Bibbs L, Cho JC, Stapels MD, Desiderio R, Vergin KL, Rappe MS,
628 Laney S, Wilhelm LJ, Tripp HJ, Mathur EJ, Barofsky DF. 2005. Proteorhodopsin in
629 the ubiquitous marine bacterium SAR11. Nature 438:82-5.
630 38. Molloy S. 2012. Marine microbiology: SAR86: streamlined for success. Nat Rev
631 Microbiol 10:82.
632 39. Swan BK, Tupper B, Sczyrba A, Lauro FM, Martinez-Garcia M, Gonzalez JM, Luo
633 H, Wright JJ, Landry ZC, Hanson NW, Thompson BP, Poulton NJ, Schwientek P,
634 Acinas SG, Giovannoni SJ, Moran MA, Hallam SJ, Cavicchioli R, Woyke T,
635 Stepanauskas R. 2013. Prevalent genome streamlining and latitudinal divergence of
636 planktonic bacteria in the surface ocean. Proc Natl Acad Sci U S A 110:11463-8.
637 40. Brown MV, Ostrowski M, Grzymski JJ, Lauro FM. 2014. A trait based perspective on
638 the biogeography of common and abundant marine bacterioplankton clades. Mar
639 Genomics 15:17-28.
640 41. Dupont CL, Rusch DB, Yooseph S, Lombardo MJ, Richter RA, Valas R, Novotny M,
641 Yee-Greenbaum J, Selengut JD, Haft DH, Halpern AL, Lasken RS, Nealson K,
642 Friedman R, Venter JC. 2012. Genomic insights to SAR86, an abundant and
643 uncultivated marine bacterial lineage. ISME J 6:1186-99.
644 42. Tripp HJ. 2013. The unique metabolism of SAR11 aquatic bacteria. J Microbiol
645 51:147-53.

646 43. Luo H, Moran MA. 2013. Assembly-free metagenomic analysis reveals new
647 metabolic capabilities in surface ocean bacterioplankton. Environ Microbiol Rep
648 5:686-96.
649 44. Beja O, Aravind L, Koonin EV, Suzuki MT, Hadd A, Nguyen LP, Jovanovich SB,
650 Gates CM, Feldman RA, Spudich JL, Spudich EN, DeLong EF. 2000. Bacterial
651 rhodopsin: evidence for a new type of phototrophy in the sea. Science 289:1902-6.
652 45. Oh HM, Kwon KK, Kang I, Kang SG, Lee JH, Kim SJ, Cho JC. 2010. Complete
653 genome sequence of "Candidatus Puniceispirillum marinum" IMCC1322, a
654 representative of the SAR116 clade in the Alphaproteobacteria. J Bacteriol 192:3240-
655 1.
656 46. Yang C, Li Y, Zhou B, Zhou Y, Zheng W, Tian Y, Van Nostrand JD, Wu L, He Z,
657 Zhou J, Zheng T. 2015. Illumina sequencing-based analysis of free-living bacterial
658 community dynamics during an Akashiwo sanguine bloom in Xiamen sea, China. Sci
659 Rep 5:8476.
660 47. Cram JA, Xia LC, Needham DM, Sachdeva R, Sun F, Fuhrman JA. 2015. Cross-depth
661 analysis of marine bacterial networks suggests downward propagation of temporal
662 changes. ISME J 9:2573-86.
663 48. Bougouffa S, Yang JK, Lee OO, Wang Y, Batang Z, Al-Suwailem A, Qian PY. 2013.
664 Distinctive microbial community structure in highly stratified deep-sea brine water
665 columns. Appl Environ Microbiol 79:3425-37.
666 49. Flombaum P, Gallegos JL, Gordillo RA, Rincon J, Zabala LL, Jiao N, Karl DM, Li
667 WK, Lomas MW, Veneziano D, Vera CS, Vrugt JA, Martiny AC. 2013. Present and
668 future global distributions of the marine Cyanobacteria Prochlorococcus and
669 Synechococcus. Proc Natl Acad Sci U S A 110:9824-9.
670 50. Partensky F, Hess WR, Vaulot D. 1999. Prochlorococcus, a marine photosynthetic
671 prokaryote of global significance. Microbiol Mol Biol Rev 63:106-27.

672 51. Zwirglmaier K, Jardillier L, Ostrowski M, Mazard S, Garczarek L, Vaulot D, Not F,
673 Massana R, Ulloa., Scanlan DJ. 2008. Global phylogeography of marine
674 Synechococcus and Prochlorococcus reveals a distinct partitioning of lineages among
675 oceanic biomes. Environmental Microbiology 10:147–161.
676 52. Gomez-Pereira PR, Hartmann M, Grob C, Tarran GA, Martin AP, Fuchs BM, Scanlan
677 DJ, Zubkov MV. 2013. Comparable light stimulation of organic nutrient uptake by
678 SAR11 and Prochlorococcus in the North Atlantic subtropical gyre. ISME J 7:603-14.
679 53. Partensky F, Hoepffner N, Li W, Ulloa O, Vaulot D. 1993. Photoacclimation of
680 Prochlorococcus sp. (Prochlorophyta) Strains Isolated from the North Atlantic and the
681 Mediterranean Sea. Plant Physiol 101:285-296.
682 54. Moore LR, Rocap G, Chisholm SW. 1998. Physiology and molecular phylogeny of
683 coexisting Prochlorococcus ecotypes. Nature 393:464-7.
684 55. Ras J, Claustre H, Uitz J. 2008. Spatial variability of phytoplankton pigment
685 distributions in the Subtropical South Pacific Ocean: comparison between in situ and
686 predicted data. Biogeosciences 5:353-369.
687 56. Robinson C, Poulton AJ, Holligan PM, Baker AR, Forster G, Gist N, Jickells TD,
688 Malin G, Upstill-Goddard R, Williams RG. 2006. The Atlantic Meridional Transect
689 (AMT) programme: a contextual view 1995–2005. Deep Sea Research Part II: Topical
690 Studies in Oceanography 53:1485-1515.
691 57. Zielinski O, Henkel R, Vofl D, Ferdelman TG. 2017. Physical oceanography during
692 SONNE cruise SO245 (UltraPac) doi:10.1594/PANGAEA.870806. PANGAEA.
693 58. Biers EJ, Sun S, Howard EC. 2009. Prokaryotic genomes and diversity in surface
694 ocean waters: interrogating the global ocean sampling metagenome. Appl Environ
695 Microbiol 75:2221-9.
696 59. Swan BK, Martinez-Garcia M, Preston CM, Sczyrba A, Woyke T, Lamy D,
697 Reinthaler T, Poulton NJ, Masland ED, Gomez ML, Sieracki ME, DeLong EF, Herndl

GJ, Stepanauskas R. 2011. Potential for chemolithoautotrophy among ubiquitous bacteria lineages in the dark ocean. Science 333:1296-300.

60. Chitsaz H, Yee-Greenbaum JL, Tesler G, Lombardo MJ, Dupont CL, Badger JH, Novotny M, Rusch DB, Fraser LJ, Gormley NA, Schulz-Trieglaff O, Smith GP, Evers DJ, Pevzner PA, Lasken RS. 2011. Efficient de novo assembly of single-cell bacterial genomes from short-read data sets. Nat Biotechnol 29:915-21.

61. Reintjes G, Tegetmeyer H, Wulf J, Bürgisser M, Fuchs BM. 2017. On-board sequencing of the microbial community of the South Pacific Gyre. doi:https://doi.pangaea.de/10.1594/PANGAEA.882015 PANGAEA.

62. Holinde L, Zielinski O. 2016. Bio-optical characterization and light availability parameterization in Uummannaq Fjord and Vaigat–Disko Bay (West Greenland). Ocean Science 12:117-128.

63. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, Glockner FO. 2013. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Res 41:e1.

64. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 75:7537-41.

65. Toribio AL, Alako B, Amid C, Cerdeno-Tarraga A, Clarke L, Cleland I, Fairley S, Gibson R, Goodgame N, Ten Hoopen P, Jayathilaka S, Kay S, Leinonen R, Liu X, Martinez-Villacorta J, Pakseresht N, Rajan J, Reddy K, Rosello M, Silvester N, Smirnov D, Vaughan D, Zalunin V, Cochrane G. 2017. European Nucleotide Archive in 2016. Nucleic Acids Res 45:D32-D36.

723 66. Diepenbroek M, Glöckner F, Grobe P, Güntsch A, Huber R, König-Ries B,
724 Kostadinov I, Nieschulze J, Seeger B, Tolksdorf R, Triebel D. 2014. Towards an
725 Integrated Biodiversity and Ecological Research Data Management and Archiving
726 Platform: The German Federation for the Curation of Biological Data (GFBio), vol
727 232. Bonn: Köllen Verlag, Lecture Notes in Informatics (LNI) – Proceedings.
728 67. Yilmaz P, Kottmann R, Field D, Knight R, Cole JR, Amaral-Zettler L, Gilbert JA,
729 Karsch-Mizrachi I, Johnston A, Cochrane G. 2011. Minimum information about a
730 marker gene sequence (MIMARKS) and minimum information about any (x)
731 sequence (MIxS) specifications. Nature biotechnology 29:415-420.
732 68. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment
733 search tool. J Mol Biol 215:403-10.
734 69. Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar, Buchner A, Lai
735 T, Steppi S, Jobb G, Forster W, Brettske I, Gerber S, Ginhart AW, Gross O, Grumann
736 S, Hermann S, Jost R, Konig A, Liss T, Lussmann R, May M, Nonhoff B, Reichel B,
737 Strehlow R, Stamatakis A, Stuckmann N, Vilbig A, Lenke M, Ludwig T, Bode A,
738 Schleifer KH. 2004. ARB: a software environment for sequence data. Nucleic Acids
739 Res 32:1363-71.
740 70. Pruesse E, Peplies J, Glockner FO. 2012. SINA: accurate high-throughput multiple
741 sequence alignment of ribosomal RNA genes. Bioinformatics 28:1823-9.
742 71. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson
743 GL, Solymos P, Henry M, Stevens H, Wagner H. 2013. vegan: Community Ecology
744 Package. R package v2.0-10. http://CRAN.R-project.org/package=vegan.
745 72. Juggins S. 2016. Rioja: Analysis of Quaternary Science Data, v0.9-9. Cran.
746 http://www.staff.ncl.ac.uk/stephen.juggins/.

747 73. Thiele S, Fuchs BM, Amann RI. 2011. Identification of microorganisms using the
748 ribosomal RNA approach and fluorescence *in situ* hybridization. Treatise on Water
749 Science 3:171-189.
750 74. Zubkov MV, Tarran GA. 2008. High bacterivory by the smallest phytoplankton in the
751 North Atlantic Ocean. Nature 455:224-6.
752 75. Zielinski O, Henkel R, Voß D, Ferdelman TG. 2017. Physical oceanography during
753 SONNE cruise SO245 (UltraPac). Institute for Chemistry and Biology of the Marine
754 Environment, Carl von Ossietzky University, Oldenburg, Germany. PANGAEA
755 doi:10.1594/pangaea.870806.
756 76. Schlitzer R. 2017. Ocean Data View. http://odv.awi.de.

Table 1. On-board sequencing pipeline results. a) Averaged DNA concentration (µg ml^{-1}) with standard deviations and average total cell counts (TCC, DAPI stained cell ml^{-1}) over different depths ranges of the SPG. b) Torrent Suite analysis of raw sequencing data using three different sequencing chip types (Ion V2 314, 316, 318) and two analysis methods (default and stringent). Loading refers to the percentage of wells of each sequencing chip which are filled with ISPs. c) Sequencing read abundances for all stations and within different station types (main and intermediate).

On-board Sequencing Results

a) DNA Extraction and Total Cell Counts

Depth (m)	*DNA conc. (ng μl^{-1})*	*Standard deviation*	*TCC (cells ml^{-1})*	*Number of samples*
20	4.8	1.95	5.25×10^5	8
40 - 60	5.8	2.84	5.93×10^5	11
75 - 100	7.4	2.85	5.95×10^5	12
125 -150	7.7	3.61	4.57×10^5	16
160 - 175	6.9	4.71	3.41×10^5	11
200 - 250	4.9	3.11	1.91×10^5	22
300 - 500	2.2	1.3	1.17×10^5	21
750 - 1500	0.9	0.72	4.14×10^4	10
2000 - 3000	0.4	0.15	2.22×10^5	7
3500 - 5000	0.5	0.21	2.14×10^4	7
Average	4.2		2.90×10^5	13

b) Sequencing Performances

Settings	*Chip type (Ion v2)*	*Number of runs*	*Total bases (Mbp)*	*Total number of reads (Q20)*	*Loading (%)*	*Mean length (bp)*	*Median (bp)*
Default	Ion 314	5	83	3.06×10^5	72	265	258
Stringent			56	1.48×10^5		379	439
Difference			- 33%	- 52%		114	181
Default	Ion 316	1	465	1.49×10^6	52	312	364
Stringent			322	8.47×10^5		380	433
Difference			- 31%	- 43%		68	69
Default	Ion 318	2	551	2.12×10^6	50	259	248
Stringent			404	1.17×10^6		345	362
Difference			- 27%	- 45%		86	114

c) Sequencing statistics

	All stations	*Main*	*Intermediate*
Average	18630	23247	10274
Median	13051	24035	8809
Mode	9321	8531	9321
Minimum	3454	5433	3454
Maximum	61928	61928	34095

Table 2 | Permutational multivariant analysis of variance (PERMANOVA), analysis of similarly (ANOSIM) and Mantel tests of bacterial community composition based on Bray-Curtis dissimilarities of relative read abundance. Factors using for grouping observations were a) station (geographic distance (Mantel test)), b) irradiance and c) depth (m).

	PERMANOVA				*ANOSIM*	*Mantel Test*
a) Analysis by station (geographic distance)						
Source of Variance	**d.f.**	**SS**	***pseudo* F**	**R^2**	**R**	**R**
Station (longitude)	1	0.323	02.11	0.018	0.012	0.03
Residuals	116	17.92		0.15		
Total	118	18.24		1		
b) Analysis by depth (m)						
Depth	1	3.32	26.05	0.18*	0.69*	
Residuals	117	14.92		0.13		
Total	118	18.24		1		
c) Analysis by irradiance zone						
Irradiance	2	10.31	75.42	0.56*	0.76*	
Residuals	116	7.93		0.068		
Total	118	18.24		1		

** denotes significance of ($p<0.001$). PERMANOVA p-values were obtained using sums of squares and 999 permutations. d.f.: degrees of freedom, SS: sum of squares. ANOSIM performed with 999 permutations. Mantel test was equated using a distance matrix of the with geographic distance between stations.*

Table 3. List of specific oligonucleotide probes for fluorescence *in situ* hybridization (FISH) applied in this study. For the detection of members of the SAR11-clade the following probes were mixed according to (22): SAR11-152R, SAR11-441R(ori), SAR11-441Rmod, SAR11-487mod, SAR11-542R, SAR11-732R, and helper SAR11-487-h3; For the detection of members of the AEGEAN169 clade the following probes were mixed: AEGEAN169-395, AEGEAN169-744, and their helper AEGEAN169-395-h1, AEGEAN169-395-h2, AEGEAN169-744-h1, AEGEAN169-744-h2, and the competitor AEGEAN169-395-comp; for SAR86 the probe SAR86-1245 was mixed with the helper SAR86-1245-h3 and SAR86-1245-h5; All probes were mixed in equimolar concentrations; abbreviation FA: formamide.

probe	target organisms	sequence (5' → 3')	FA [%]	reference
SAR11-152R	SAR11-clade	ATTAGCACAAGTTTCCYCGTGT	25	(20)
SAR11-441R(ori)	SAR11-clade	TACAGTCATTTTCTTCCCCGAC	25	(20)
SAR11-441Rmod	SAR11-clade	TACCGTCATTTTCTTCCCCGAC	25	(20) modified
SAR11-487mod	SAR11-clade	CGGACCTTCTTATTCGGG		(21) modified
SAR11-487-h3	helper to SAR11-487mod	CGGCTGCTGGCACGAAGTTAGC		(22)
SAR11-542R	SAR11-clade	TCCGAACTACGCTAGGTC	25	(20)
SAR11-732R	SAR11-clade	GTCAGTAATGATCCAGAAAGY TG	25	(20)
PRO405	*Prochlorococcus*	AGAGGCCTTCGTCCCTCA	15	(23)
AEGEAN169-395	AEGEAN169-clade	GTCACTCACGCTGCATTG	20	this study
AEGEAN169-395-comp	competitor to AEGEAN169-395	GTCACTCACGCGGCATTG	20	this study
AEGEAN169-395-h1	helper to AEGEAN169-395	CTGGATCAGGGTTTCCCC	20	this study
AEGEAN169-395-h2	helper to AEGEAN169-395	TACTTCCCTAAGGCCTTC	20	this study
AEGEAN169-744	AEGEAN169-clade	ATCTCAGCGTCAAAAATGG	20	this study
AEGEAN169-744-h1	helper to AEGEAN169-744	CCTAGTTAGTCGCCTTCG	20	this study
AEGEAN169-744-h2	helper to AEGEAN169-744	TGCTACCCACGCTTTCGT	20	this study
SAR86-1245	SAR86 clade	TTAGCGTCCGTCTGTAT	35	(24)

SAR86-1245-h3	helper to SAR86	GGATTRGCACCACCTCGCGGC	35	(24)
SAR86-1245-h5	helper to SAR86	CCATTGTAGCACGTGTGTAGC	35	(24)
SAR202-312R	SAR202 clade	TGTCTCAGTCCCCCTCTG	40	(27)
SAR324-1412	SAR324 clade	GCCCCTGTCAACTCCCAT	35	(25)
SAR406-97	SAR406 clade	CACCCGTTCGCCAGTTTA	40	(26)