

Functional traits of particle-associated microbial assemblages in the low-latitude Pacific Ocean

Hideto Takami (✉ takamih@jamstec.go.jp)

Japan Agency for Marine-Earth Science and Technology <https://orcid.org/0000-0003-3433-7018>

Yuya Tada

Kaiyo Kenkyu Kaihatsu Kiko Yokohama Kenkyujo

Takashi Okubo

Kaiyo Kenkyu Kaihatsu Kiko Yokohama Kenkyujo

Wataru Arai

Kaiyo Kenkyu Kaihatsu Kiko Yokohama Kenkyujo

Yoshitoshi Ogura

Kyushu Daigaku Igakubu Daigakuin Igakukei Gakufu Daigakuin Igaku Kenkyuin

Tetsuya Hayashi

Kyushu Daigaku Igakubu Daigakuin Igakukei Gakufu Daigakuin Igaku Kenkyuin

Shotaro Suzuki

Tsukuba Daigaku - Taiki Kaio Kenkyujo

Koji Hamasaki

Tokyo Daigaku Taiki Kaiyo Kenkyujo

Research

Keywords: particle-associated, free-living, metagenomics, functionome, Pacific Ocean, Genomaple

Posted Date: January 22nd, 2020

DOI: <https://doi.org/10.21203/rs.2.21628/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background Organic particles are hotspots for microbial activity and serve as sites of organic matter mineralisation in the water column of marine systems. In nutrient-limited surface water, degradation of organic matter and nutrient regeneration by marine microbes is crucial. Although free-living (FL) bacteria vastly outnumber those on particles, particle-associated (PA) bacteria can reach locally higher concentrations. Accordingly, to achieve a better understanding of marine microbial ecosystems, it is important to elucidate the differences in not only microbial community structures, but also functional traits, between PA and FL environmental sample fractions. In a previous study, we demonstrated that the Genomape (formerly MAPLE) system could successfully differentiate the functional potentials and diversity of contributors to each function in four metagenomic datasets generated by the Global Ocean Sampling expedition. Hence, we also used this system to highlight functional traits in PA microbial assemblages. Results The PA and FL fractions could be distinguished from one another by their taxonomic compositions, inferred from ribosomal proteins and relative abundance of module functions. Module functions that were more abundant among PA assemblages than FL assemblages were shared between both subtropical gyres, and their taxonomic compositions were similar. Bacterial transport systems associated with adhesive molecules used for forming microbial assemblages through particulate organic matter were more abundant in the PA fractions. Bacterial regulatory system elements for C 4 -dicarboxylate transport and B-vitamin biosynthesis were also abundant among PA assemblages, suggesting mutual relationships between bacteria and algae involved in exchange of nutrient sources. On the other hand, module functions related to amino acid biosynthesis and bacterial transport systems for inorganic nitrogen, phosphorus, and urea were significantly more abundant in the PA assemblages of more oligotrophic North and South Pacific subtropical gyres than eastern equatorial Pacific regions. Conclusions Comprehensive functional metagenomic analyses based on functional abundance revealed some notable functional traits in PA assemblages related to cell adhesion and nutrient acquisition, enabling the microbes to survive in subtropical regions that are more oligotrophic than the equatorial regions.

Background

Although bacteria can grow rapidly when appropriate nutrients are available, some still grow slowly under optimum conditions. Such microbes are usually present in oligotrophic environments, and are rarely present in copiotrophic environments [1]. Subtropical and tropical ocean surface water is nutritionally poor due to the low productivity of such ecosystems, which are characterised by strong thermal stratification. In particular, the southern Pacific is extremely oligotrophic, with nutrient concentrations that are undetectable by standard methods [2]. Phosphate concentrations of the North Pacific subtropical gyre (NPSG) are lower than those of the South Pacific subtropical gyre (SPSG) [3,4]. By contrast, inorganic nitrogen and phosphorus concentrations in the equatorial (EQ) Pacific regions are relatively high as a result of upwelling [5]. Nevertheless, primary productivity is insufficient to exhaust macronutrients owing

to the limited supply of iron received as dust from Eurasia, which is a high-nutrient and low-chlorophyll area [6,7].

Organic particles are hotspots for microbial activity, and microbes play critical roles in the degradation and remineralisation of both sinking [8,9] and suspended particulate organic matter (POM) [10,11]. However, the physiological differences between particle-associated (PA) and free-living (FL) assemblages remain poorly understood [12,13]. The primary and most important process of POM formation is photoautotrophic fixation of CO₂ into biomass in pelagic ecosystems, usually by phytoplankton. Particles associated with microbes constitute a significant fraction of sinking particles and thus contribute to the biological carbon pump [14–16]. In nutrient-limited surface water, degradation of POM and nutrient regeneration by marine microbes is crucial and more directly linked to phytoplankton primary production than in nutrient-replete ocean areas. Thus, phytoplankton development affects the community structure and dynamics of marine bacteria in a way that depends on changes in nutrient availability, algal exudates, and biological surfaces [17,18]. Although FL bacteria vastly outnumber those on particles, PA bacteria can reach locally higher concentrations [19]. Accordingly, it is important to elucidate the differences not only in microbial community structures but also in functional traits between PA and FL environmental sample fractions to reach a better understanding of marine microbial ecosystems. Although most bacterial community structure studies do not distinguish between PA and FL assemblages, several studies have compared microbial diversity between PA and FL fractions at individual marine sites [17, 20–23].

Metagenomic approaches have been increasingly recognised as a starting point for understanding microbial functions and community structures [24,25]. However, many metagenomic microbial community analyses simply infer bacterial community structure based on potentially biased polymerase chain reaction (PCR) amplicons of 16S rRNA sequences, despite the fact that gene copy number can vary even within the same species [26]. On the other hand, we have confirmed the effectiveness of using ribosomal proteins, instead of problematic 16S rRNA sequences, to identify organisms and conduct community structure analysis [27]. An unrealised major goal of metagenomics is the inference of both species diversity and potential functionomes, defined as comprehensive sets of biological functions arising through the combinations of coding genes underlying such functions, e.g., carbon fixation, nitrogen fixation, and amino acid metabolism [28,29]. Indeed, only a fraction of genes corresponding to specific metabolic pathways and transporters have been employed in functional characterisation of habitats. To resolve this problem, we developed software to evaluate metabolic and physiological potential, Genomable (formerly MAPLE) [27–29], which enables estimation of comprehensive and detailed functional potential based on completion ratios of every functional module defined by the Kyoto Encyclopedia of Genes and Genome (KEGG) [30] and the abundance of complete modules. Indeed, given that Genomable has successfully differentiated the functional potentials and diversity of contributors to each function in four metagenomic datasets generated by the Global Ocean Sampling (GOS) expedition [27], it should be capable of revealing differences in functional traits between PA and FL microbial assemblages.

Metatranscriptomic analyses of microbial assemblages from NPSG surface water have previously revealed community-wide metabolic activities and taxon-specific expression patterns [31–35]. However, it remains unclear how and to what extent PA and FL functional potentials and diversities of microbial consortia differ among the Pacific subtropical gyres and equatorial regions. Accordingly, a functional metagenomic approach has become indispensable for detecting differences in functional potentials and their host organisms, both between PA and FL assemblages and among NPSG, SPSG, and EQ assemblages within the respective PA and FL fractions.

In this study, our analyses, based on the functional modules responsible for pathways and molecular complexes using Genomale [27,28], have enabled a novel, comprehensive, and detailed functional characterisation of the PA and FL fractions. Here, we provide comparative analyses of functional abundance and a reanalysis of prokaryotic community structure based on ribosomal proteins instead of problematic 16S rRNA sequences. We place special emphasis on functional traits of the PA fraction in the subtropical regions and the biodiversity contributing to functional abundance of the highlighted modules. We also focus on the differences in the functional traits of the PA fraction between subtropical and equatorial regions.

Results And Discussion

Comparison of community structure between PA and FL fractions.

We performed prokaryotic community structure analyses of metagenomes obtained in the NPSG, SPSG, and EQ regions (Additional File 1: Fig. S1) based on ribosomal proteins (Additional File 2: Table S1). The relative abundance of bacteria in the PA fractions of each region exceeded 96%, and there was no significant difference in the ratio of bacteria and archaea between the NPSG and EQ regions (Fig. 1a). However, the relative abundance of archaea in the SPSG regions was extremely low (0.3–0.8%) in comparison with other regions. The FL fraction exhibited a similar tendency to the PA fraction, except that archaeal abundance in the EQ region was about 2-fold higher (2.6–3.6%) than in the NPSG region.

Among bacterial community members, α -proteobacteria were predominant across all samples, and their relative abundance was lower in the PA fraction (23.9–45.8%) than in the FL fraction (36.1–57.3%) (Fig. 1b and Additional File 3: Table S2). Cyanobacteria were the second most prominent members in both NPSG fractions except for the PA fraction from site NPSG1; in contrast to SPSG, where Cyanobacteria was only second most abundant for SPSG1 in the FL fraction. Thus, remarkable differences were observed in relative abundance between the NPSG and SPSG sites even though both are oligotrophic and subtropical. The relative abundance of Bacteroidetes in PA assemblages was more than twice as high as in the FL fraction in almost all regions, and their abundance in the NPSG was about half of that in other regions (Fig. 1b). We previously performed community structure analyses of metagenomes based on ribosomal proteins (GS030 and GS031) [27], obtained near the Galápagos Islands by the GOS Expedition [36]. The community structure of these GOS sites, which are near the EQ1 site, was mainly composed of α -proteobacteria, γ -proteobacteria, and Bacteroidetes, but the relative abundance of Cyanobacteria was

very low (< 2%), unlike the FL fraction of site EQ1 (14%). In the GOS expedition, an 0.8- μm filter was used for prefiltration, unlike the 3.0- μm filter used in this study, and the smaller pores may have more efficiently trapped the longer, narrower Cyanobacteria cells. Non-metric multidimensional scaling (nMDS) analysis of the compositional patterns of the prokaryotic ribosomal proteins revealed that the community structure in the FL and PA fractions differed, and this difference was supported by Bray–Curtis dissimilarity (Fig. 2a). In addition, the difference in prokaryotic community structure among the subtropical and equatorial regions in each fraction (PA and FL) was statistically significantly similar to that observed in the former analysis (Fig. 3a).

In most metagenomic analyses, amplicon sequences of 16S rRNA genes are used to compare microbial community structures among different environments [37]. Recently, this PCR-based amplicon approach has generally targeted the V4 region because different 16S rRNA gene regions yield varying degrees of accuracy in taxonomic assignment [38]. However, because prokaryotic species exhibit 16S rRNA copy number variation, it is impossible to determine the copy numbers of individual uncultivable microbes present in natural environments. Thus, because taxonomic compositions based on amplicon sequences are strongly influenced by copy number and PCR bias, this approach is suboptimal for community structure analysis. We used three types of primer sets containing the V3–V4 region to confirm the variability among results, and we found that the patterns of bacterial composition in the PA and FL fractions of the NPSG, SPSG, and EQ regions varied significantly depending on the primer sets [2]. The composition of the four main bacterial taxa (i.e., α -proteobacteria, Cyanobacteria, γ -proteobacteria, and Bacteroidetes) in both fractions of each sampling site differed by factors of 1.2–6, and for Verrucomicrobia, the difference was several tens of fold because these species were barely detected by the V1–V2 primer set (Table 1). Copy numbers of the *rrn* operon in α -proteobacteria, Cyanobacteria, γ -proteobacteria, and Bacteroidetes range from 1 to 11 (mean, 2.7), 1 to 6 (mean, 2.6), 1 to 21 (mean, 6.2), and 1 to 13 (mean, 3.4), respectively [26]. A comparison of the relative abundances of each bacterial taxon based on ribosomal proteins with amplicon-based relative abundances revealed that while some relative abundance estimates were similar, there was no consistency among the results of amplicon primer sets. Indeed, although δ -proteobacteria were specifically detected by all types of amplicons in the PA fraction of NPSG2, they were hardly detected based on ribosomal proteins in either fraction from any region. Furthermore, the relative abundance of Bacteroidetes was clearly overestimated in the PA fractions of the EQ2 and SPSG2 sites.

Localisation pattern and biodiversity of archaea

Among archaeal community members, Marine group II Euryarchaeota (MGII) [39,40] predominated among archaeal ribosomal proteins in both PA and FL fractions of all regions, with SPSG samples exhibiting especially high relative abundances exceeding 95%, and about 80% of MGII sequences consisted of MGIIA (Fig. 4a). In NPSG samples, marine group III Euryarchaeota (MGIII) accounted for about 30% and 20% in the PA and FL fractions, respectively, with MGIIIB predominant in both fractions, in

contrast to the SPSG samples. MGIII was also detected in EQ samples, but at a lower relative abundance than in the NPSG samples (16.6% and 6.8% in the PA and FL fractions, respectively). Similar to the SPSG samples, MGIIA was predominant in EQ samples, but the region's archaeal community structure was intermediate between those of the NPSG and SPSG regions. Although some uncultivated marine Euryarchaeota can adhere to particulate organic matter [41], no significant differences in archaeal community structure were observed between the PA and FL fractions in this study. Most archaeal ribosomal proteins assigned by Genomape in each metagenome matched uncultured draft genomes by top score in homology searches. Some of those ribosomal proteins exhibited more than 90% identity with Euryarchaeota archaeon TMED252 and TMED215 proteins within MGIIA and MGIIIB, respectively, in all regions [34], and with MGIII Euryarchaeota CG-Epi1 and CG-Epi3 [42] in the NPSG and EQ regions (Fig. 4b). However, many ribosomal proteins did not match other draft genomes, with identities reaching only 80%, indicating greater marine euryarchaeal diversity than expected, as reported previously for the NPSG region [41].

Functional discrimination based on relative module abundance.

Using Genomape, we analysed a total of 18 metagenomic sequence datasets from the PA and FL fractions of each sampling site (Additional file 4: Table S3). The following results were recorded: module completion ratio (MCR) by individual taxonomic ranks (ITRs; see Methods), number of ITRs with completed modules (i.e., diversity of ITR completed modules), and sum of relative module abundance of each ITR-completed module (Additional file 5: Tables S4). We used the relative module abundance commonly completed in all metagenomic samples to compare the functional traits among samples and also examined the composition of each ITR responsible for completion of each module. First, we performed statistical analysis based on the pattern of relative module abundance of each sample to confirm that the FL and PA assemblages, as well as their microbial community structures, were functionally distinct. The nMDS analysis of Bray–Curtis dissimilarity showed that the prokaryotic functional potential of the FL fraction significantly differed from that of the PA fraction (Fig. 2b). Next, we examined the significance of the differences among sampling sites for the same fraction (i.e., FL or PA) in the same manner. In the FL fraction, while NPSG, SPSG, and EQ regions were significantly differentiated (Fig. 3b), the SPSG1 site, with a prokaryotic compositional pattern similar to that of the EQ3 site (Fig. 1b), deviated from the other SPSG sites. Additionally, in the PA fraction, each region was significantly differentiated according to a pattern similar to that of the FL fraction, as supported by ANOSIM and PERMANOVA. These results indicated that the PA and FL fractions had predominant functions, and that their microbial assemblages were independent of each other. In addition, we expected that PA and FL would have respective functional characteristics that reflect the environmental factors at the sampling location. Therefore, in the next step, we performed comparative analyses to highlight the predominant physiological functions characterising the PA fractions, as well as the subtropical regions.

Common functional traits of PA fraction in NPSG and SPSG.

Particle association is likely a prominent strategy for utilisation organic matter by heterotrophic bacteria and archaea, especially those in oligotrophic marine environments. Accordingly, we focused on prokaryotic functions that predominate in the PA fractions, and combined prokaryotic sequences from three sampling sites in the same region as representative datasets from the NPSG, EQ, and SPSG regions to highlight differences in the prokaryotic potential functionomes of each region (Additional File 4: Table S3). We focused on modules with relative abundances > 2-fold higher in the PA fraction than in the FL fraction, based on the Genomape results (Additional File 6: Table S5), to determine common PA functional traits in both oligotrophic regions. Consequently, 14 modules were commonly highlighted in NPSG and SPSG regions, three of which were the modules responsible for biosynthesis of B vitamins such as pantothenate (B₅), pyridoxal (B₆), and thiamine (B₁) (Fig. 5). Most eukaryotic phytoplankton require exogenous B vitamins, and are thus B-vitamin auxotrophs. For example, a species within the Dinoflagellates, *Lingulodinium polyendrum*, obtains vitamin B₁ from a member of clade *Roseobacter* within α -proteobacteria, followed by the γ -proteobacterium *Marinobacter flavimaris* [43]. Photosynthetic picoeukaryotic algae that contribute to primary production in oligotrophic marine environments also require B vitamins, and the heterotrophic bacterial partner *Dinoroseobacter shibae* within α -proteobacteria frequently associates with marine algae within Mamiellales, providing vitamins B₁ and B₁₂ to its host [44]. Vitamin B₁ plays a pivotal role in intermediary carbon metabolism and is a cofactor for a number of enzymes involved in primary carbohydrate and branched-chain amino acid metabolism. Indeed, the major contributors to the relative abundance of the vitamin B₁ biosynthesis module (M00127) are the α - and γ -proteobacterial species containing *Dinoroseobacter* and *Marinobacter*, respectively. In addition, other types of Dinoflagellate and Mamiellales were detected by eukaryotic community structure analysis based on ribosomal proteins (Additional File 7: Fig. S2).

Another notable feature of the PA fractions was the fairly high relative module abundance of two-component regulatory systems related to transport of C₄-dicarboxylates (fumarate, succinate, and malate) [45] in both the NPSG and SPSG fractions. Contributors to module abundance consisted of three major proteobacterial classes, namely α -, γ -, and β -proteobacteria (Fig. 5). Although the process of physiological adaptation of diatoms under nitrogen starvation is poorly understood, proteomic research on the marine diatom *Thalassiosira pseudonana* has revealed that many of the TCA cycle proteins that yield C₄-dicarboxylate, such as succinate dehydrogenase and fumarate dehydratase, are upregulated by 1.6- to 2.5-fold under nitrogen starvation [46]. Diatoms within Thalassiosirales were detected in all PA fractions, at 1–6% relative abundance, and were most abundant in the EQ region, where upwelling occurs (Additional File 7: Fig. S2). Heterotrophic bacterial assemblages adhering to diatomaceous surfaces may exploit C₄-dicarboxylate released from the diatoms as a nutrient source. Additionally, expression of genes encoding portions of the C₄-dicarboxylate transport system and binding proteins has been observed in *Pelagibacter* within α -proteobacteria in the NPSG region, suggesting the importance of C₄-dicarboxylate as a carbon and energy source for these bacteria [33]. Although such a transcriptomic approach has not yet been applied to SPSG samples, similar results can be expected. Our findings further emphasise the

importance of C₄-dicarboxylate utilisation, especially for proteobacteria adhered to POM in oligotrophic marine environments.

The relative abundance of the module responsible for adhesin protein transport system (APTS) in the PA fraction was more than 3-fold greater than in the FL fraction (Fig. 5). POM is initially living biomass of primary producers that, through various processes and transfer steps in the food web, eventually becomes detrital POM, which often dominates total POM. Actually, eukaryotic ribosomal proteins mainly corresponding to various taxa of alga were significantly overrepresented in trapped biomass on the 3.0- μ m pre-filter (Additional File 7: Fig. S2). During the first step of bacterial interaction with POM, including living biomass, bacterial cells develop a series of surface proteinaceous adhesins that promote specific or nonspecific adhesion under various environmental conditions and achieve irreversible attachment [47]. Although APTS occurs in many kinds of bacterial species, major contributors to this system, such as *Dinoroseobacter* and *Planctomarina* within α -proteobacteria, were shared between the NPSG and SPSSG regions. In addition, the relative abundance of the capsular polysaccharide and lipopolysaccharide (LPS) ABC transport system was higher in the PA fraction. Complex carbohydrates in the form of LPS, polysaccharide, capsules, and lipooligosaccharides are commonly found on the cell surface [48]. Capsules allow organisms to adhere to various surfaces, as well as to other bacteria, facilitating the colonisation of various ecological niches and formation of biofilms [49]. The capsule provides the bacterium with mechanisms to avoid non-specific host defences, including complement-mediated bacteriolysis and opsonophagocytosis [50]. Because LPS is also implicated in surface adhesion, bacteriophage sensitivity, and interactions with predators, these aggregative and protective molecules may greatly contribute to the formation and maintenance of PA bacterial assemblages.

Thus, PA bacteria can be more advantageous to gain nutrient and energy sources through particle-mediated interactions between microbial communities than FL bacteria, but they also become susceptible to stress caused by various factors, especially antibiotics, produced by secondary metabolism. As one manifestation of this concept, the relative module abundance of the two-component regulatory system related to the envelope stress response (ESR) and cationic antimicrobial peptide resistance (CAPR) was remarkably high. ESR is an inducible bacterial response that senses and mediates adaptation to insults to the outer compartment of the cell, and has been implicated in the ability to resist or endure various types of antibiotics [51]. Antimicrobial peptides are part of the innate immune response found among all classes of life. Antimicrobial peptides can kill Gram-negative and Gram-positive bacteria, as well as enveloped viruses and fungi [52]. Bacteria use various resistance strategies to avoid being killed by antimicrobial peptides [53], but only γ -proteobacterial species contributed to the relative abundance of the CAPR and ESR modules.

Significant functional traits in subtropical regions.

As mentioned previously, surface seawater of subtropical latitudes is generally more oligotrophic than that of tropical latitudes, but the concentrations of inorganic salts such as nitrate, phosphate, and iron complex differ between the subtropical and tropical regions, and also the north and south subtropical regions. Hence, we calculated the relative abundance ratio of each module between the subtropical (NPSG and SPSG) and EQ regions (i.e., NPSG/EQ and SPSG/EQ) in the PA fraction. Fifteen modules for which the NPSG/EQ and SPSG/EQ values exceeded a 2-fold difference were identified (Fig. 6 and Additional File 6: Table S5). Especially, the relative abundance of the module responsible for the Entner–Doudoroff (ED) pathway was more than 3-fold higher in the PA fraction than in the EQ regions (Fig. 5). Aerobic and facultative anaerobes are more likely to have the ED pathway despite its low energy yield, as these microbes have other non-glycolytic pathways for creating ATP, such as oxidative phosphorylation. Thus, it is thought that the ED pathway is especially favoured among species thriving in the oligotrophic environments due to the lower amounts of protein required to generate ATP. Another notable feature of the subtropical oligotrophic regions was the high relative abundance of modules related to biosynthesis of amino acids such as serine, phenylalanine, tyrosine, and histidine; moreover, there was no significant difference in the contributors to these modules between NPSG and SPSG, with the exception of serine biosynthesis (Fig. 6). In addition to amino acid biosynthesis, the relative abundance of the module responsible for urea transport system was more than 4-fold higher than in the EQ regions. The concentration of ammonium in the subtropical regions corresponding to the sampling sites in this study is very low, less than 0.15 $\mu\text{mol/l}$ [54]. In the NPSG regions, more than 60% of the contributors to the abundance of urea transport system were cyanobacterial species similar to *Prochlorococcus* sp. MIT 604 with a complete set of urease genes. In contrast to the NPSG, SPSG, in which the proportion of cyanobacteria in the total community is small (Fig. 1b), contributed only 15% to the urea transport system. However, most of the cyanobacteria species were *Prochlorococcus* sp. MIT 604, as in the case of NPSG. As well as ammonium, nitrate and nitrite (N+N) concentrations in the NPSG and SPSG regions were much lower than in the EQ region (Additional File 1: Fig. 1). Corresponding to this condition, the relative abundance ratio of the two transport systems related to nitrate and nitrite in the subtropical regions, especially the nitrate/nitrite transport system (M00438), was 8–10 times higher than that of the EQ (Fig. 6). Interestingly, there was a significant difference between the NPSG and SPSG in the composition of contributors to this function. Thus, these results reflect the differences in concentrations of carbon and nitrogen sources between the subtropical and tropical regions; indeed, the concentration of chlorophyll a in the EQ region is 3–7 times higher than in the subtropical regions (Additional File: Fig. S1). In addition, cell number by DAPI count was 3–4 times higher, suggesting larger amounts of organic matter than in the NPSG and SPSG regions. Therefore, PA assemblages in the subtropical regions are presumably adapted to oligotrophic environments due to higher capacity for amino acid synthesis and uptake of urea and N+N.

The other highlighted module included molybdate transport systems. Nitrogen fixation activity is not observed in the equatorial region, which is rich in soluble reactive phosphorus (SRP), whereas NPSG, which has been depleted of SRP, has significantly higher N_2 -fixation activity [2]. In fact, the relative abundance ratio of the two-component regulatory system for phosphate starvation response was higher

in both the NPSG and SPSG regions, where the concentration of SRP is much lower than in the EQ. Nitrogen fixation contributes up to 84% of new production in the upper waters of subtropical gyre, where the diazotroph community includes γ -proteobacteria and Cyanobacteria [55]. Considering that nitrogenase, the core enzyme of nitrogen fixation, is Mo-dependent [56], higher Mo-transport potential is presumably connected to the significant nitrogen fixation activity in the NPSG [55], as the major contributors to the relative abundance of the Mo transport system are the γ -proteobacteria. On the other hand, the two-component regulatory system related to nitrogen fixation also had a high relative abundance ratio, although its association with nitrogen fixation in the NPSG was not clear, as α -proteobacteria were the main contributor to this function.

As mentioned above, the PA microbial assemblage has higher functional potential for adhesion and self-protection in oligotrophic surface water. In fact, considering the extremely high relative abundance ratios of modules related to the adhesion protein transport system, this seems to indicate that the subtropical region has a higher capacity than the EQ region to form microbial assemblages capable of surviving in an oligotrophic environment. On the other hand, other types of drug resistance system (i.e., multidrug efflux pump and aminoglycoside resistance) were highlighted with significant high relative abundance ratio unlike former comparison. Aminoglycoside antibiotics exhibit bactericidal activity against mainly Gram-negative aerobes and some anaerobic bacilli [57]. Streptomycin is a first-in-class aminoglycoside antibiotic produced by *Streptomyces griseus*; indeed, various actinobacterial species containing *S. griseus* were detected in PA assemblages by the community structure analysis based on the ribosomal proteins, although their population was small (Fig. 1b and Additional File: Table S2).

Conclusions

Through comprehensive functional metagenomic analyses based on functional abundance, we highlighted the functional traits of PA fractions associated with formation and maintenance of microbial assemblage in comparison with the FL fractions and locations. We could also reveal the functions of PA assemblages that were required for survival in oligotrophic ocean surface water. Accordingly, this approach can enable more precise interpretations of biological information harboured in massive metagenomic datasets and lead to a better understanding of marine ecosystems. Our study also reinforces the importance of considering PA and FL prokaryotic assemblages as independent microbial communities. In addition, we reconfirmed that community structure analysis using problematic 16S rRNA amplicon sequences can often be misleading in studies of this kind.

Methods

Sample collection and treatment and DNA extraction.

As previously reported, during two excursions of the R/V Hakuho-Maru across the NPSG, SPSG, and EQ regions of the Pacific Ocean (Additional File 1: Fig. S1), 18 seawater samples were collected from the surface at nine locations [2]. To separate the PA and FL assemblages, 100 L of seawater were serially

filtered with 3.0- μm pore Nucleopore polycarbonate membrane filters (Whatman, Florham Park, NJ, USA) and 0.22- μm pore Steripak-GP cartridge filter units (Millipore, Billerica, MA, USA). DNA was extracted using a previously reported method with some modifications [31]. Lysozyme (1 mg/mL as a final concentration) was added to 20 mL of sucrose lysis buffer (50 mM Tris HCl, pH 8.0, 40 mM EDTA, 0.75 M sucrose) and transferred to the Steripak-GP cartridge filter in which the cell masses were trapped. After a 30-min incubation at 37°C, the mixture including Proteinase K (0.1 mg/ml) and SDS (1%) was divided into 2.0-mL tubes with 3.0- μm membrane filter cartridges, and each sealed tube was incubated at 65°C for 90 min. This treatment was repeated to increase lysis efficiency. After concentration of cell lysate with Amicon Ultra 15 (Millipore), DNA was purified according to a standard method.

Sequencing and raw data processing.

Paired-end libraries with 480-bp inserts were prepared using the Nextera XT DNA Library Prep Kit (Illumina, San Diego, CA, USA) and sequenced on the Illumina MiSeq platform, generating 2–5 million reads with 250-bp read lengths. PEAR (v0.9.6) was used to merge these paired-end reads after sequences with low-quality scores were removed using FASTX-Toolkit and duplicates were eliminated using PRINSEQ [58]. Gene prediction from the merged sequences was performed using MetaGene Annotator [59]. These sequences were divided into partial and complete genes, and all sequences, except for partial sequences shorter than 50 amino acids (a.a.) in length, were randomly selected to compile multi-FASTA files for Genomagle analysis. The number of sequences obtained from each sampling site is listed in Table S3 (Additional File 4).

Functional annotation and evaluation of metabolic potential.

Each multi-FASTA file of a.a. sequences (Additional File 4: Table S3) was submitted to Genomagle 2.3.2 using GHOSTX as a homology search engine [28]. The single-directional best-hit method [27] was used to assign a KEGG Orthology (KO) ID to the query gene using the KEGG Automatic Annotation Server [60]. The MCR of each functional module was calculated as described previously [27,29]. To normalise KO abundance, the total number of query sequences assigned to each KO was divided by the average length of each KO group [27]. To enable comparisons among different sites, module abundance was normalised against the abundance of all ribosomal proteins in each sample (yielding module abundance per ribosomal protein) [27]. Taxonomic assignment of contributors to complete modules was performed based on ITRs, such as phylum and class (i.e. secondary taxonomic rank defined by KEGG).

Comparative analysis of module abundance.

PA and FL datasets were created using three combined sequences from the same region, as shown in Table S3 (Additional file 4). These six sets of sequence data were submitted to Genomape to calculate the relative abundance of each functional module per ribosomal protein appearing in each fraction (FL and PA) from the NPSG, SPSG, and EQ regions. The abundance ratio was calculated between the PA and FL fractions in the same region, and also between regions in the same fraction, based on the Genomape results (Additional File 6: Table S5).

Community structure analysis.

The proportional representation of bacteria, archaea, and eukaryotes in the metagenomes was calculated based on the mapping pattern of virtual module M90000 for prokaryotic ribosomes (Additional File 8: Fig. S3) [27]. Because the archaeal ribosome contains six more proteins than the bacterial ribosome (58 vs. 52), we normalised the total number of archaeal ribosomal proteins to the number of bacterial ribosomal proteins by multiplying the archaeal ribosomal protein count data by 52/58. We summed the number of bacterial ribosomal proteins and normalised archaeal ribosomal proteins, and then used this sum as the denominator in subsequent calculations. We calculated the proportion of prokaryotes at the ITR level similarly. For a more detailed analysis of archaeal communities, we searched all sequences assigned to archaeal ribosomal proteins by Genomape using the NCBI non-redundant (nr) protein database.

Phylogenetic analysis.

To determine the phylogenetic position of the top hit euryarchaeotic species based on homology searches, we retrieved ribosomal proteins from each complete or draft genome registered in the NCBI database. Subsequently, 10 commonly conserved ribosomal proteins were selected from among the draft genomes. These concatenated protein sequences were aligned with the corresponding sequence of *Halobacterium salinarum*, used as an outgroup. The LG+G model in MEGA 6.06 was used to construct a phylogenetic tree using the maximum likelihood method [61].

Statistics.

A dataset based on the mapping patterns of module M90000 was prepared for comparative prokaryotic community structure analysis, as shown in Table S1 (Additional File 2). Another dataset was also prepared for comparative functional analysis based on module abundance patterns, as shown in Table S4 (Additional File 5). Only abundance data for modules commonly completed among all metagenomic samples were used for this analysis. Bray–Curtis distance [62] was used to create the distance matrix. The differences in community structure and module abundance patterns, both between the PA and FL fractions and among locations, were analysed using non-metric multidimensional scaling (nMDS). The

differences between categories (i.e., different treatments) were tested using ANOSIM [63] and PERMANOVA [64], with 9999 permutations performed for both analyses using the VEGAN package in R (ver. 3.3.3) [65].

Declarations

Acknowledgements

We thank Professor K. Kurokawa and Professor H. Mori for their assistance with processing and registering the data. Thanks are also due to the captain and crew of the R/V Hakuho-Maru for assistance with sample collection.

Author Contributions

H.T. and K.H. conceived and designed the analysis; K.H., S.S., Y.O., and T.H. performed the experiments; H.T., T.O., Y.T., and W.A. analysed the data; H.T. wrote the manuscript. All authors read and approved the final version of the manuscript.

Funding

This work was supported by KAKENHI Grants-in-Aid for Scientific Research (to H.T., 15KT0039, 17H00793, and 18K19846); by KAKENHI Grants-in-Aid for Scientific Research (to K.H., 24121004 and 241210030), as part of the NEOPS New Ocean Paradigm on Its Biogeochemistry, Ecosystem and Sustainable Use/Project on Innovative Areas; by a grant from the Cross-Ministerial Strategic Innovation Promotion Program (to H.T.); and by MEXT KAKENHI grant No. 221S0002 (to K.H.).

Availability of data and materials

All sequence data used in this study have been deposited in DDBJ under BioSample accession numbers SSUB004934 and SAMD00042636–00042654 for E1PA, E1FL, E2AP, E2FL, E3AP, E3FL, N1AP, N1FL, N2AP, N2FL, N3AP, N3FL, S1AP, S1FL, S1FL_add, S2AP, S2FL, S3AP, and S3FL. All data supporting the findings of this study are available within the article and its additional files. Amino acid sequences used for Genomape analysis are available at <https://zenodo.org/deposit/3233522>.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare no competing interests.

Author details

¹Microbial Genome Research Group, Yokohama Institute, Japan Agency for Marine-Earth Science & Technology, Yokohama, Japan, ²Department of Bacteriology, Faculty of Medical Sciences, Kyushu University, Fukuoka, Japan, ³Atmosphere and Ocean Research Institute, The University of Tokyo, Kashiwa, Japan

References

1. Koch AL. Oligotrophs versus copiotrophs. 2001;23:657–61.
2. Suzuki S, Kaneko R, Kodama T, Hashihama F, Suwa S, Tanita I, Furuya K, Hamasaki K. Comparison of community structures between particle associated and free-living prokaryotes in tropical and subtropical Pacific Ocean surface waters. *J Oceanogr.* 2017;73:383–
3. Moutin T, Karl DM, Duhamel S, Rimmelin P, Raimbault P, Van Mooy BAS, Claustre H. Phosphate availability and the ultimate control of new nitrogen input by nitrogen fixation in the tropical Pacific Ocean. *Biogeosciences.* 2008;5:95–
4. Hashihama F, Furuya K, Kitajima S, Takeda S, Takemura T, Kanda J. Macro-scale exhaustion of surface phosphate by dinitrogen fixation in the western North Pacific. *Geophys Res Lett.* 2009;36(3):L03610.
5. Murry JW, Johnson E, Garside CA. U.S. JGOFS process study in the equatorial Pacific EqPac: Introduction. *Deep-sea Res Part II* 1995;42(2-3):275–
6. Miller BC, Wheeler Habitat determinants of primary production in the sea. In: *Biological Oceanography*, 2nd edition, Wiley-Blackwell, Chichester, West Sussex; 2012. p. 49–72. ISBN 978-1-444-33301-5.
7. Pitchford JW, Brindley J. Iron limitation, grazing pressure and oceanic high nutrient-low chlorophyll HNLC; regions. *J Plankton Res.* 1999;21(3):525–47.
8. Cho BC, Azam F. Major role of bacteria in biogeochemical fluxes in the ocean's interior. *Nature.* 1988;332:441–3.
9. Steinberg DK, Van Mooy BAS, Buesseler KO, Boyd PW, Kobari T, Karl DM. et al. Bacterial vs. zooplankton control of sinking particle flux in the ocean's twilight zone. *Limnol Oceanogr.* 2008;53(4):1327–38.
10. Karl DM, Knauer GA, Martin JH. Downward flux of particulate organic matter in the ocean—a particle decomposition paradox. *Nature.* 1988;332:438–41.
11. Giering SL, et al. Reconciliation of the carbon budget in the ocean's twilight zone. *Nature,* 2014;507:480–3.
12. Herndl GJ, Reinthaler T. Microbial control of the dark end of the biological pump. *Nat Geosci.* 2013;6:718–24.

13. Ganesh S, Parris DJ, DeLong EF, Stewart FJ. Metagenomic analysis of size-fractionated picoplankton in a marine oxygen minimum zone. *ISME J.* 2014;8:187–211.
14. Simon M, Grossart HP, Schweitzer B, Ploug H. Microbial ecology of organic aggregates in aquatic ecosystems. *Aquat Microb Ecol.* 2002;28:175–211.
15. Alldredge A, Silver M. Characteristics, dynamics and significance of marine snow. *Prog Oceanogr.* 1988;20(1):41–82.
16. Volkman J, Tanoue E. Chemical and biological studies of particulate organic matter in the ocean. *J Oceanogr.* 2002;58:265–79.
17. Rösler S, Grossart HP. Contrasting dynamics in activity and community composition of free-living and particle-associated bacteria in spring. *Aquat Microb Ecol.* 2012;66(2):169–81.
18. Tada Y, Suzuki K. Changes in the community structure of free-living heterotrophic bacteria in the open tropical Pacific Ocean in response to microalgal lysate-derived dissolved organic matter. *FEMS Microbiol Ecol.* 2016;92(7):fiw099.
19. Caron DA, Davis PG, Madin LP, Sieburth JM. Heterotrophic bacteria and bacterivorous protozoa in oceanic macroaggregates. *Science.* 1982;218(4574):795–7.
20. Bižić-Ionescu M, Zeder M, Ionescu D, Orlić S, Fuchs BM, Grossart HP, Amann R. Comparison of bacterial communities on limnic versus coastal marine particles reveals profound differences in colonization. *Environ Microbiol.* 2015;17(10):3500–14.
21. Rieck A, Herlemann DP, Jürgens K, Grossart HP. Particle-associated differ from free-living bacteria in surface waters of the Baltic Sea. *Front Microbiol.* 2015;6:1297.
22. DeLong ED., Franks DG, Alldredge, AL. Phylogenetic diversity of aggregate attached vs. free-living marine bacterial assemblages. *Limnol Oceanogr.* 1993;38(5):924–34.
23. Whalen KE, Becker JW, Schrecengost AM, Gao Y, Giannetti N, Harvey E. Bacterial alkylquinolone signalling contributes to structuring microbial communities in the ocean. *Microbiome.* 2019;7:93.
24. Kennedy J, Marchesi JR, Dobson ADW. Marine metagenomics: strategies for the discovery of novel enzymes with biotechnological applications from marine environments. *Microbial Cell Factories.* 2008;7:27.
25. Bengtsson-Palme J, Rosenblad MA, Molin M, Blomberg A. Metagenomics reveals that detoxification systems are underrepresented in marine bacterial communities. *BMC Genomics.* 2014;15:749.
26. Stoddard SF, Smith BJ, Hein R, Roller BR, Schmidt TM. *rnrDB*: improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. *Nucleic Acids Res.* 2015;43:D593–8.
27. Takami H, Taniguchi T, Arai W, Takemoto K, Moriya Y, Goto S. An automated system for evaluation of the potential functionome: MAPLE version 2.1.0, *DNA Res.* 2016;23(5):467–75.
28. Arai W, Taniguchi T, Goto S, Moriya Y, Uehara H, Takemoto K, Ogata H, Takami H. MAPLE 2.3.0: an improved system for evaluating the functionomes of genomes and metagenomes. *Biosci Biotech Biochem.* 2018;82(9):1515–7.

29. Takami H, Taniguchi T, Moriya Y, Kuwahara T, Kanehisa M, Goto S. Evaluation method for the potential functionome harbored in the genome and metagenome. *BMC Genomics*. 2012;13:699.
30. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T. et al. KEGG for linking genomes to life and the environment. *Nucleic Acids Res*. 2008;36:D480–4.
31. Frias-Lopez, J. et al. Microbial community gene expression in ocean surface waters. *Proc Natl Acad Sci USA*. 2008;105(10):3805–10.
32. Poretsky RS, Hewson I, Sun S, Allen AE, Zehr JP, Moran MA. Comparative day/night metatranscriptomic analysis of microbial communities in the North Pacific subtropical gyre. *Environ Microbiol*. 2009;11(6):1358–75.
33. Shi Y, Tyson GW, Eppley JM, DeLong EF. Integrated metatranscriptomic and metagenomic analyses of stratified microbial assemblages in the open ocean. *ISME J*. 2011;5:999–1013.
34. Aylwarda FO, Eppleya JM, Smithb JM, Chavezb FP, Scholinb CA, DeLonga EF. Microbial community transcriptional networks are conserved in three domains at ocean basin scales. *Proc Natl Acad Sci USA*. 2015;112(17):5443–8.
35. Robidart JC, Magasin JD, Shilova IN, Turk-Kubo KA, Wilson ST, Karl DM, Scholin CA, Zehr JP. Effects of nutrient enrichment on surface microbial community gene expression in the oligotrophic North Pacific Subtropical Gyre. *ISME J*. 2019;13:374–87.
36. Yutin N, Suzuki MT, Teeling H, Weber M, Venter JC, Rusch DB, Béjà O. Assessing diversity and biogeography of aerobic anoxygenic phototrophic bacteria in surface waters of the Atlantic and Pacific Oceans using the Global Ocean Sampling expedition metagenomes. *Environ Microbiol*. 2007;9(6):1464–75.
37. Tringe SG, Rubin E M. Metagenomics: DNA sequencing of environmental samples. *Nat Rev Genet*. 2005;6(11):805–14.
38. Liu Z, Lozupone C, Hamady M, Bushman FD, Knight R. Short pyrosequencing reads suffice for accurate microbial community analysis. *Nucleic Acids Res*. 2007;35:e120.
39. Iverson V, Morris RM, Frazar CD, Berthiaume CT, Morales RL, Armbrust EV. Untangling genomes from metagenomes: Revealing an uncultured class of marine Euryarchaeota. *Science*. 2012;335(6068):587–90.
40. Tully BJ, Sachdeva R, Graham, ED, Heidelberg JF. 290 metagenome-assembled genomes from the Mediterranean Sea: a resource for marine microbiology. *PeerJ*. 2017;5:e3558.
41. Orsi WD, Smith JM, Wilcox HM, Swalwell JE, Carini P, Worden AZ, Santoro AE. Ecophysiology of uncultivated marine euryarchaeal is linked to particulate organic matter. *ISME J*. 2015;9:1747–63.
42. Haro-Moreno JM, Rodriguez-Valera F, López-García P, Moreira D, Martin-Cuadrado AB. New insights into marine group III Euryarchaeota, from dark to light. *ISME J*. 2017;11:1102–17.
43. Cruz-López R, Maske H. The vitamin B1 and B12 required by the marine Dinoflagellate *Lingulodinium polyedrum* can be provided by its associated bacterial community in culture. *Front Microbiol*. 2016;7:560.

44. Cooper MB, Kazamia E, Helliwell KE, Kudahl UJ, Sayer A, Wheeler, GL, Smith AG. Cross-exchange of B-vitamins underpins a mutualistic interaction between *Osterococcus tauri* and *Dinoroseobacter shibae*. ISME J. 2019;13:334-345.
45. Janausch IG, Zients E, Tran QH, Kroger A, Unden G. C₄-dicarbolylate carriers and sensors in bacteria. Biochim Biophys Acta. 2002;1553(1-2):39–
46. Hockin NL, Mock T, Mulholland F, Kopriva S, Malin G. The response of diatom central carbon metabolism to nitrogen starvation is different from that of green alga and higher plants. Plant Physiol. 2012;158:299–
47. Klemm P, Schembri MA. Bacterial adhesins: function and structure. Int J Med Microbiol. 2000;290(1):27–
48. Silver RP, Prior K, Nsahlai C, Wright LF. ABC transporters and the export of capsular polysaccharides from Gram-negative bacteria. Res Microbiol. 2001;152:357–364.
49. Roberts IS. The biochemistry and genetics of capsular polysaccharide production in bacteria. Annu Rev Microbiol. 1996;50:285–315.
50. Moxon ER, Kroll JS. The role of bacterial polysaccharide capsules as virulence factors. Curr Top Microbiol Immunol. 1990; 150:65–85.
51. Guest RL, Raivio TL. Role of Gram-negative envelope stress response in the presence of antimicrobial agents. Trends Microbiol. 2016;24(5):377–390.
52. Robert EW, Diamond G. The role of cationic antimicrobial peptides in innate host defences. Trends Microbiol. 2000;8(9):402–410.
53. Reddy KV, Yedery RD, Aranha C. Antimicrobial peptides: premises and promises. Int J Antimicrob Agents. 2004;24(6):536–47.
54. Yasunaka S. et al. Basin-scale distribution of NH₄⁺ and NO₂⁻ in the Pacific Ocean. J Oceanogr. 2018; 74:1–11.
55. Shinozaki T, Bombar D, Reimann L, Hashihama F, Takeda S, Yamaguchi T, Ehama M, Hamasaki K, Furuya K. Basin-scale variability of active diazotrophs and nitrogen fixation in the North Pacific, from the tropics to the subarctic Bering Sea. Global Biogeochem Cycles. 2017;31(6):996–1009.
56. Scott DJ, May HD, Newton WE, Brigle KE, Dean DR. Role of the nitrogenase MoFe protein α -subunit in FeMo-cofactor binding and catalysis. Nature 1990;343:188–190.
57. Davies J, Wright GD. Bacterial resistance to aminoglycoside antibiotics. Trends Microbiol. 1997;5(6):234–240.
58. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. Bioinformatics. 2011;27(6):863–864.
59. Noguchi H, Taniguchi T, Itoh T. MetaGeneAnnotator: Detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes. DNA Res. 2008;15(6):387–96.

60. Moriya Y, Itoh M, Okuda S, Yoshizawa A, Kaneshisa, M. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* 2007;35:W182–5.
61. Tamura K, Stecher G, Peterson D, Filipski A, Kumar, S. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol.* 2013;30(12):2725–9.
62. Bray JR, Curtis JT. An ordination of the upland forest communities of Southern Wisconsin. *Ecol Monogr.* 1957;27:326–49.
63. Clarke KR. Nonparametric multivariate analyses of changes in community structure. *Aust J Ecol.* 1993;18:117–43.
64. McArdle BH, Anderson MJ. Fitting multivariate models to community data: a comment on distance-based redundancy analysis. *Ecology.* 2001;82(1):290–7.
65. R Development Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2014

Additional File Legends

Additional File 1: Fig. S1. Sampling locations and environmental variables of surface water.

Additional File 2: Table S1. Taxonomic composition of prokaryotes in each metagenomic sample.

Additional File 3: Table S2. Taxonomic composition of Bacteria in each metagenomic sample.

Additional File 4: Table S3. Number of sequence reads used for Genomapple analysis.

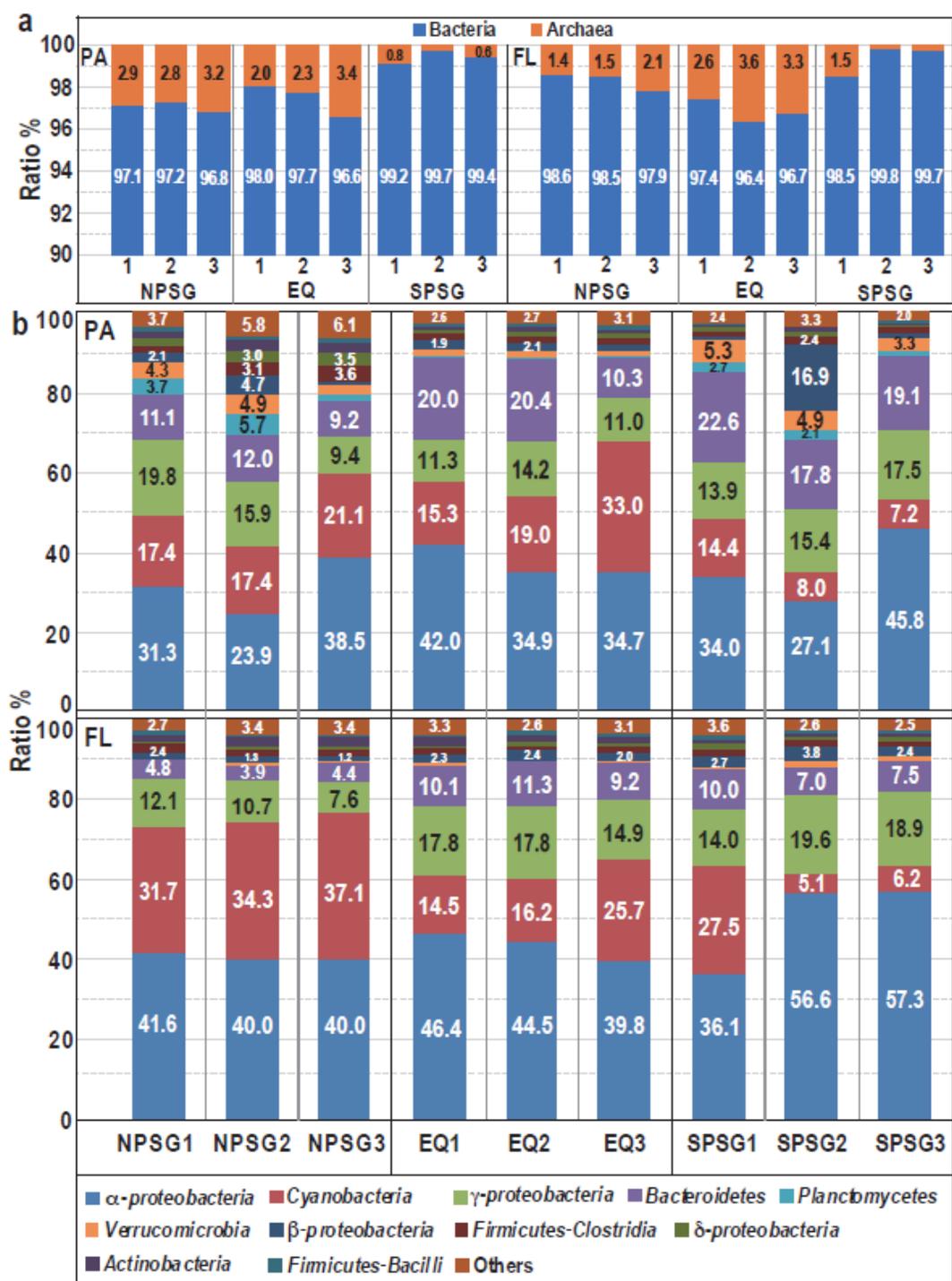
Additional File 5: Table S4. Comparison of Genomapple results for metagenomic sequence of the PA and FL fractions from each sampling site from the NPSG, EQ, and SPSG regions.

Additional File 6: Table S5. Comparison of Genomapple results based on integrated metagenomic sequences of three sites in each region.

Additional File 7: Fig. S2. Composition of eukaryotes based on ribosomal proteins of the NPSG, EQ, and SPSG regions.

Additional File 8: Fig. S3. Virtual module for the prokaryotic ribosome (M90000).

Figures



Takami et al. Fig. 1

Figure 1

Prokaryotic community structure analysis based on ribosomal proteins. a Comparison of bacterial and archaeal ratio between FL and PA fractions from each sampling site. b Comparison of bacterial composition between FL and PA fractions from each site. Relative abundances of each taxon were calculated based on mapping patterns for M90000 modules (Additional File 8: Fig. S3), which are virtual common ribosome modules for prokaryotes.

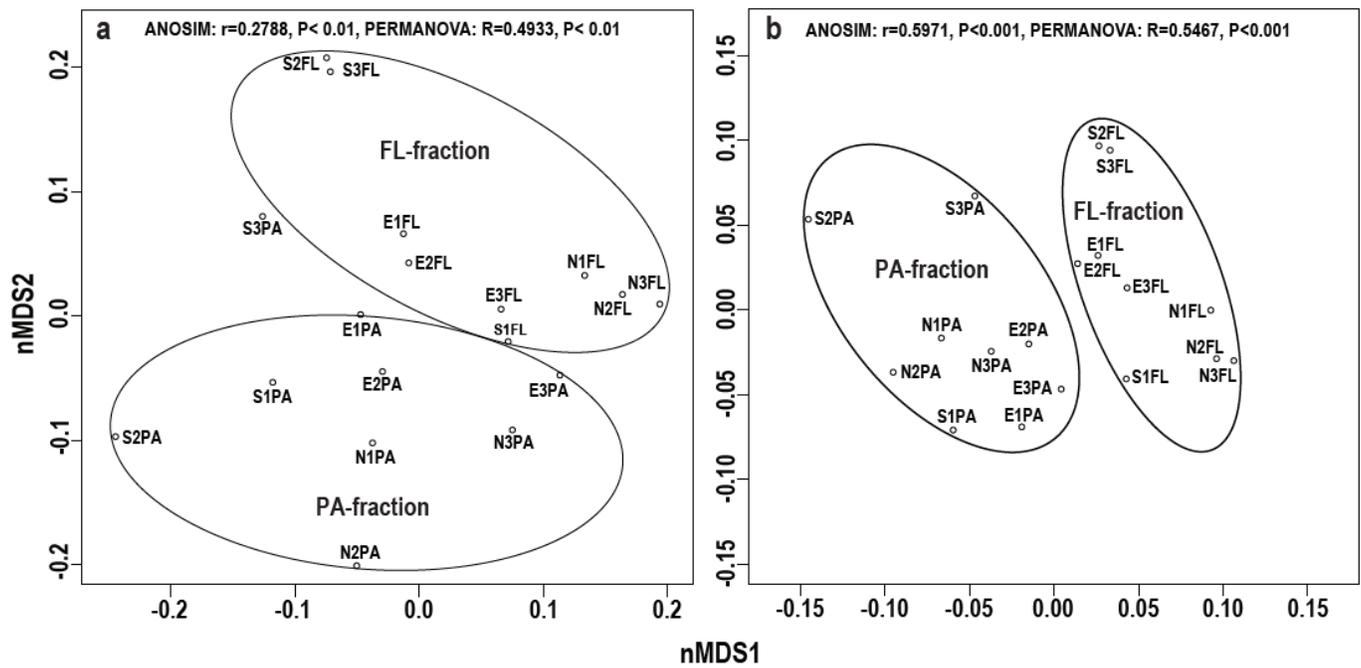


Figure 2

Non-metric multidimensional scaling (nMDS) analysis. a Prokaryotic taxonomic composition. b Relative abundance of modules completed by prokaryotes. a Prokaryotic taxonomic composition. b Abundance of complete modules. PA, particle-associated fraction; FL, free-living fraction. The Bray–Curtis distance was used to create the distance matrix. Data were analysed using nMDS. Differences between categories (i.e. different treatments) were tested using analysis of similarity (ANOSIM) and permutational multivariate analysis of variance (PERMANOVA) with 9999 permutations used for both analyses.

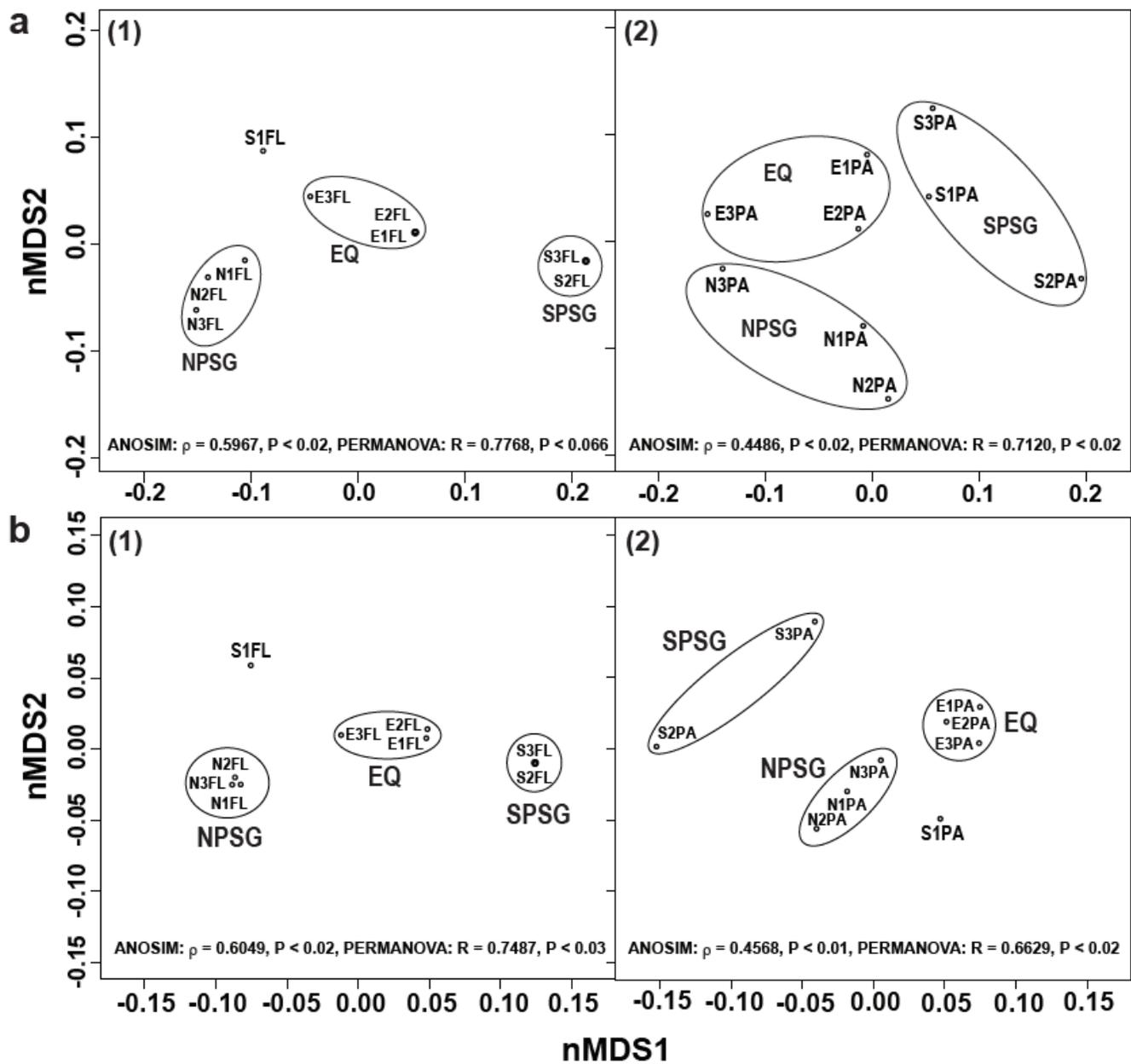


Figure 3

Non-metric multidimensional scaling (nMDS) analysis of complete modules in each fraction from each sampling site using prokaryotic taxonomical composition and abundance. a Prokaryotic taxonomic composition. b Abundance of complete modules. (1) particle-associated fraction (PA); (2) free-living fraction (FL). Statistical analysis was performed as in Fig. 2.

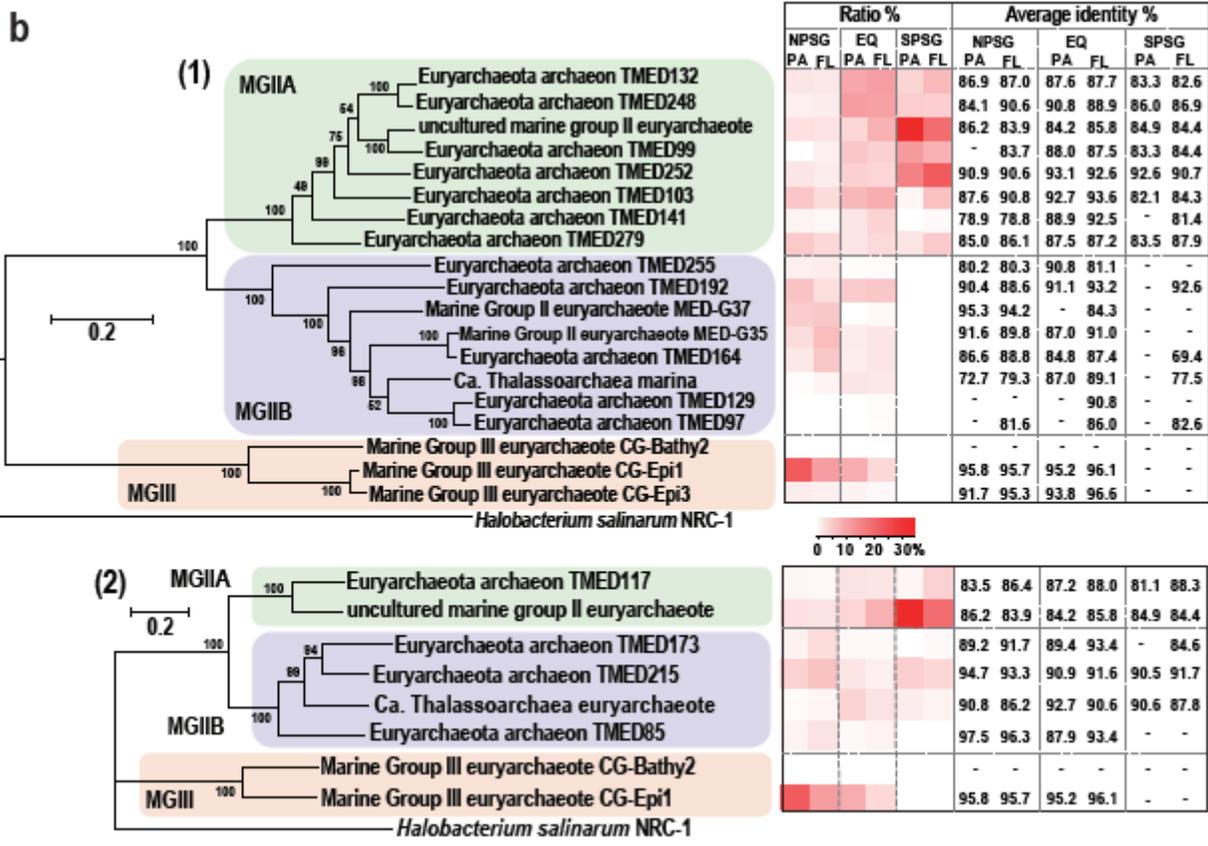
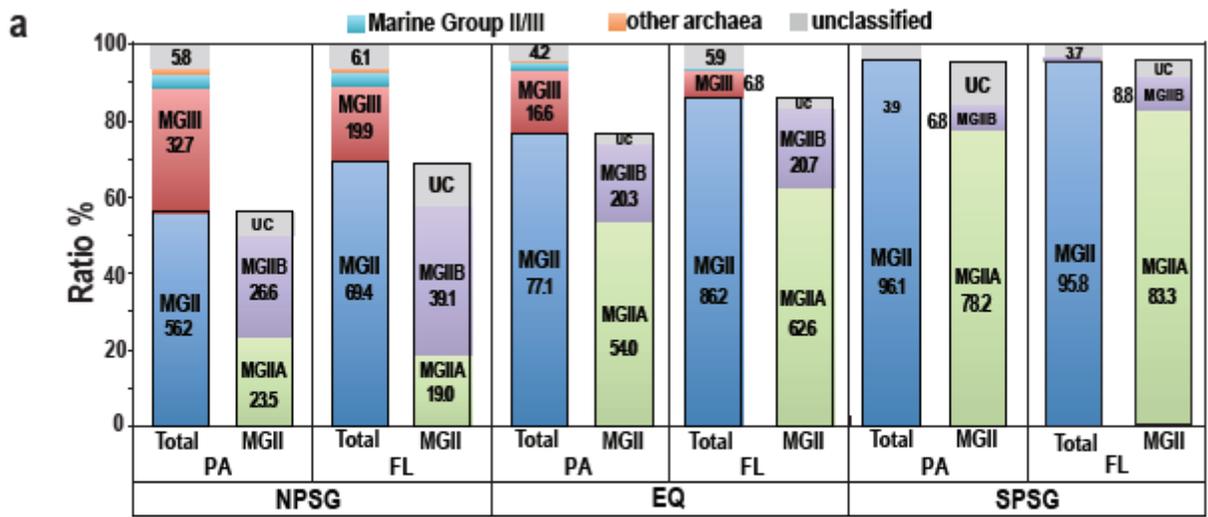


Figure 4

Classification of archaea based on ribosomal proteins. a Taxonomic composition based on archaeal ribosomal proteins assigned by Genomaple. The colour-coded portions of each bar show the percentage of ribosomal proteins assigned to each archaeal group. b Distribution of archaeal members within each metagenomic sample. The red bar shows the percentage of ribosomal proteins assigned to each archaeal member whose draft genome sequence has been registered in NCBI. 'Average identity' indicates the percentage of ribosomal proteins appearing in each metagenome relative to those of each archaeal

member within MGII and MGIII, respectively. The following two sets of concatenated ribosomal proteins were used for this study: (1) rpIN (K02874), rpIB (K02886), rpIV (K02890), rpIC (K02906), RPL4 (K02930), rpIE (K02931), rpsQ (K02961), rpsS (K02965), rpsC (K02982), and RPS4 (K02987); (2) rpIA (K02863), rpIJ (K02864), RPL31 (K02910), rpsK (K02948), rpsQ (K02956), RPS17 (K02962), RPS19 (K02966), RPS3A (K02984), rpsD (K02986), and RPS6 (K02991). A maximum likelihood tree was constructed using the LG+G model implemented in MEGA 6.06 [61].

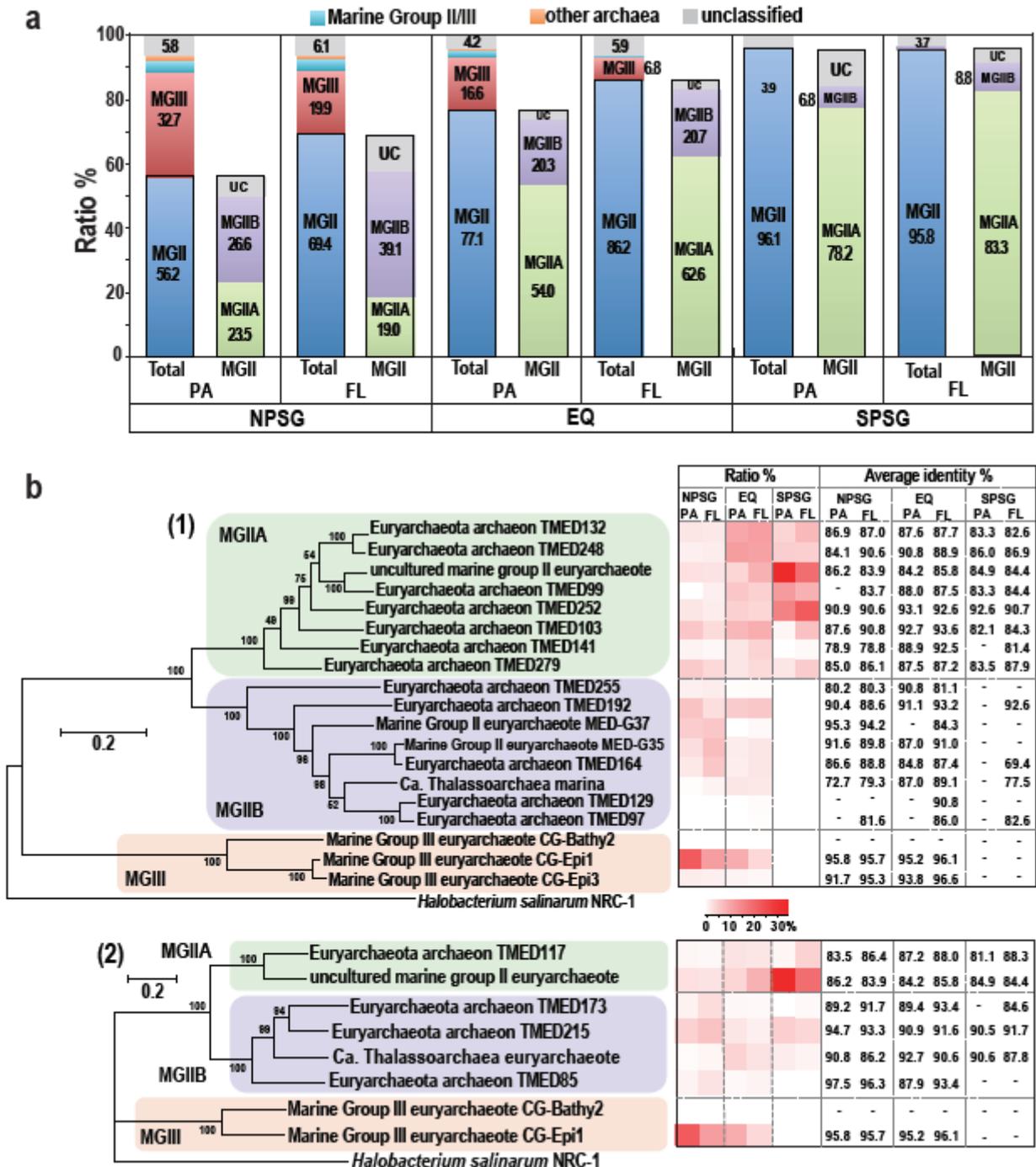


Figure 5

Abundant functions in PA fraction common to subtropical regions, and their biological diversity. PA/FL is the ratio of abundance in the PA fraction vs. the FL fraction. Abundance ratios were calculated based on the relative abundance of the modules commonly completed in both regions, and only modules with 2-fold or higher abundance ratios are highlighted in this figure. Blue text indicates rare modules that were completed by less than 30% of 2367 prokaryotic species registered in Genomape. Histograms indicate the composition of ITRs contributing to the module function in the NPSG and SPSG regions. (1) Effective ATP generation; (2) Nutrient source exchange between bacteria and alga; (3) Adhesion of cells via POM; (4) Population stress and cellular self-protection; (5) Nitrogen fixation; (6) Uptake of minerals and trace organic matter.



Takami et al. Fig. 6

Figure 6

Functions that were more abundant in the PA fraction in subtropical regions than in equatorial regions, and their biodiversity. Only modules with abundance ratios exceeding 2-fold are shown in this figure. NPSG/EQ, abundance ratio of NPSG to EQ; SPSG/EQ, abundance ratio of SPSG to EQ. Module ID, coloured in blue, indicates rare modules, as shown in Fig. 5. Histograms indicate the composition of ITRs contributing to the module function in the NPSG and SPSG regions. (1) Effective ATP generation; (2) Nutrient source exchange between bacteria and alga; (3) Adhesion of cells via POM; (4) Population stress

and cellular self-protection; (5) Nitrogen fixation; (6) Uptake of minerals and trace organic matter; (7) Amino acid biosynthesis.

Supplementary Files

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

- [FigureS3.pdf](#)
- [FigureS2.pdf](#)
- [TableS2.pdf](#)
- [TableS3.pdf](#)
- [TableS1.pdf](#)
- [TableS5.xlsx](#)
- [FigureS1.pdf](#)
- [TableS4.xlsx](#)
- [Table1.pdf](#)