

Original citation:

Richa, Kumari, Balestra, Cecilia, Piredda, Roberta, Benes, Vladimir, Borra, Marco, Passarelli, Augusto, Margiotta, Francesca, Saggiomo, Maria, Biffali, Elio, Sanges, Remo, Scanlan, David J. and Casotti, Raffaella. (2017) Distribution, community composition and potential metabolic activity of bacterioplankton in an urbanized Mediterranean Sea coastal zone. *Applied and Environmental Microbiology*, 83 (17). e00494-17.

Permanent WRAP URL:

<http://wrap.warwick.ac.uk/90095>

Copyright and reuse:

The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions. Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Publisher's statement:

A note on versions:

The version presented here may differ from the published version or, version of record, if you wish to cite this item you are advised to consult the publisher's version. Please see the 'permanent WRAP URL' above for details on accessing the published version and note that access may require a subscription.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk

Distribution, community composition and potential metabolic activity of bacterioplankton in an urbanized Mediterranean Sea coastal zone

Kumari Richa*, Cecilia Balestra*, Roberta Piredda*, Vladimir Benes[°], Marco Borra*, Augusto Passarelli*,
Francesca Margiotta*, Maria Saggiomo*, Elio Biffali*, Remo Sanges*, David J Scanlan⁺ and Raffaella
Casotti*[#]

* Stazione Zoologica A. Dohrn, Villa Comunale I80121 Naples, Italy

⁺ School of Life Sciences, University of Warwick, UK

[°] Scientific Core Facilities & Technologies, GeneCore, European Molecular Biology Laboratory, Heidelberg,
Germany

[#] Corresponding author

Running title: Bacterioplankton in an urbanized coastal zone

Key words: Bacterioplankton, 16S rDNA and rRNA, potential metabolic activity, Illumina sequencing, Gulf
of Naples

Abstract

Bacterioplankton are fundamental components of marine ecosystems and also influence the entire biosphere by contributing to the global biogeochemical cycles of key elements. Yet, there is a significant gap in knowledge about their diversity, specific activity and environmental factors that shape their community composition and function. Here, the distribution and diversity of surface bacterioplankton along the coastline of the Gulf of Naples (GON, Italy) was investigated using flow cytometry coupled with high-throughput sequencing of 16S rRNA gene (rDNA). Heterotrophic bacteria numerically dominated the bacterioplankton, and comprised mainly *Alphaproteobacteria*, *Gammaproteobacteria* and *Bacteroidetes*. Distinct communities occupied river-influenced, coastal and offshore sites, as indicated by Bray-Curtis dissimilarity, distance metric (UniFrac), LefSe and multivariate analyses. Heterogeneity in diversity and community composition was mainly due to salinity and changes in environmental conditions across sites as defined by nutrient and chlorophyll *a* concentrations. Bacterioplankton communities were composed of a few dominant taxa and a large proportion (92%) of rare taxa (here defined as OTUs accounting for <0.1% of total sequence abundances), the majority of which were unique to each site. The relationship between 16S rRNA and 16S rDNA, i.e. between potential metabolic activity and abundance, was positive for the whole community. However, analysis of individual Operational Taxonomical Units (OTU) revealed high 16S rRNA:rDNA ratios for most (71.6±16.7%) of the rare taxa, suggesting that these low abundance organisms were potentially active and hence might be playing an important role in ecosystem diversity and functioning in the GON.

Importance

The study of bacterioplankton in coastal zones is of critical importance considering that these areas are highly productive and also anthropogenically impacted. Their richness and evenness as well as their potential activity are very important to assess ecosystem health and functioning. Here, we investigated bacterial distribution, community composition and potential metabolic activity in the GON which is an ideal test due to its heterogeneous environment characterized by a complex hydrodynamics and terrestrial inputs of varying quantities and quality. Our study demonstrates that bacterioplankton communities in this region are highly

45 diverse and strongly regulated by a combination of different environmental factors leading to their
46 heterogeneous distribution, with the rare taxa contributing to a major proportion of diversity and shifts in
47 community composition and potentially holding a key role in ecosystem functioning.

48

Introduction

Microbes dominate the abundance, diversity and activity of marine ecosystems, being key components of marine food webs and playing vital roles in major biogeochemical cycles, climate regulation and the remineralization of organic matter (1, 2). Bacteria (and Archaea) numerically dominate the microbial fraction, including both autotrophic primary producers that carry out photosynthesis and heterotrophic organisms that recycle the dissolved organic carbon and nutrients, processing nearly one half of global marine primary production (3). Bacterioplankton communities exhibit high phylogenetic and physiological diversity (4) with a remarkable capacity to transform and adapt to the environment around them (5). As a consequence, knowledge of microbial community composition and diversity patterns is critical to determining the health and functioning of marine ecosystems (1).

High-throughput sequencing of ribosomal genes facilitated an increasing recognition of the vast diversity of marine microbes (6). Bacterioplankton diversity and distribution patterns have been extensively studied both at local (7, 8) and global scales (9, 10, 11) using this technique, yet some oceanic regions, particularly in highly populated coastal areas, still remain underexplored (12). Determining bacterioplankton metabolic activity is crucial for estimating their potential contribution to ecosystem processes. Previous studies have examined bulk activity in terms of biomass production and growth rates, for instance using bacterial production assays (13, 14), or by measuring incorporated labelled precursors (15, 16). However, much less is known of the activity of specific taxa within complex bacterioplankton communities, an aspect that can be analyzed using 16S rRNA:rDNA ratios as a proxy of potential growth rate relative to abundance (17, 18). Although this approach is not free from bias, for instance due to more than one copy of rRNA gene per cell (see 17, 19 and references therein), it can readily provide taxon-specific growth potential information, thus acting as a proxy for their contribution to ecosystem processes, as well as facilitating the identification of the key drivers of such processes (19). Previously, it was presumed that under steady-state conditions and in the absence of top-down regulation, most active bacteria are the ones that are most abundant, with higher growth rates leading to higher biomass (17). Consequently, the rare bacteria were generally considered to

represent a “seed bank” encompassing slow-growing or dormant individuals ready to respond only when environmental conditions became favorable again (20). However, a few recent studies showed that rare Operational Taxonomical Units (OTUs) can have higher levels of 16S rRNA (ribosomes) per rDNA (cell number) and that they may be disproportionately active relative to their abundances (21, 22). These findings have encouraged studies on distribution patterns and significance of the rare biosphere in ecosystem functioning and prompted further analyses at finer taxonomic resolution.

The Gulf of Naples (GON) is a semi-enclosed deep embayment opening into the southern Tyrrhenian Sea, in the mid-western Mediterranean basin, with an average depth of 170 m and a surface area of 870 km². Situated besides the urbanized and densely populated city of Naples and its surroundings, the GON is subjected to severe anthropogenic pressures such as land and industrial run-offs, improperly treated sewage discharge and maritime trafficking (23). In its southern part, it receives several inputs from the Sarno river (one of the most polluted rivers in Europe), particularly following heavy rain events or from uncontrolled urban discharges. The GON is characterized by complex hydrodynamics with a variable boundary between coastal and offshore waters, whose position depends upon the general circulation of the Tyrrhenian Sea, which intrudes eddies into the GON, mainly in the fall and winter (24). The extension of the boundary between mesotrophic coastal and oligotrophic offshore waters strongly depends on local physical geography and bottom topography (25), but also on seasonal variability (26, 27). At times when northerly currents dominate, freshwater inputs from the Volturno and Garigliano rivers can be traced into the GON from the northern Gulf of Gaeta (28). The high diversity and dynamic nature of the GON make it an ideal site to study microbial community structure and dynamics in response to environmental factors. Extensive investigations in this region have surveyed planktonic eukaryotes (23, 29), but bacterial communities have received much less attention except for cyanobacteria (30, 31) and pathogens (e.g. 32, 33). However, there has been a recent focus on using picophytoplankton to track physical processes such as eddies, vertical mixing or upwelling events (e.g. 30, 31).

The aim of this study was to characterize GON bacterioplankton communities in terms of abundance and distribution, and to relate them to the general environmental conditions present along the coastline of the GON. To facilitate this we used flow cytometry and high-throughput sequencing of 16S rDNA amplicons to obtain high resolution taxonomic identification. At three sites, the correlation between bacterioplankton abundance and activity was also investigated, the latter *via* comparison of 16S rRNA:rDNA ratios.

Materials and Methods

Sample collection, chlorophyll and nutrient analysis

Sixteen surface seawater samples (Fig. 1) were collected from 10th to 29th April 2013 on board the R/V Minerva Uno and the R/V Vettoria using a rosette sampler equipped with 10 L Niskin bottles and a CTD probe (SBE 911 plus, SeaBird Electronics, USA). At all stations seawater samples for chlorophyll *a*, inorganic nutrient (NO₂, NO₃, NH₄, PO₄ and SiO₄) analyses, flow cytometry counts, 16S rDNA and 16S rRNA library preparation were collected from the same Niskin bottle. Duplicate samples for flow cytometry (1 ml each) were placed in 1.5 ml cryovials, fixed with a mix of 0.05% (v/v) glutaraldehyde and 1% (w/v) paraformaldehyde (final concentrations) for 10 minutes in the dark, frozen in liquid nitrogen and stored at -80°C until analysis as in [31](#). For DNA collection, 2 l of each sample was pre-filtered through a 10 µm pore-size mesh net to remove large eukaryotes and debris and successively filtered through 47 mm Glass Fiber/A filters (Whatman, UK), whose nominal pore size (1.67 µm) is larger than most free-living marine bacteria ([34](#)). Cells were then collected onto a 0.22 µm pore-size Sterivex filter cartridge (Millipore, USA) using a peristaltic pump at 24 rpm. Sterivex filters were then sealed, frozen in liquid nitrogen and stored at -80°C until DNA extraction. For RNA collection, the procedure was exactly the same as for the DNA extraction, except for the last filter, which was a 0.22 µm pore-size 47 mm Durapore filter (Millipore, USA), using a vacuum pump at low vacuum (2.9 psi). After filtration, each filter was stored in a cryovial, frozen in liquid nitrogen and stored at -80 °C until RNA extraction.

Samples for the measurement of total chlorophyll *a* were filtered onto 25 mm Whatman GFF filters, frozen and stored in liquid nitrogen until analysis. Concentrations were determined using a Shimadzu (Japan) RF-5301PC spectrofluorometer as previously described (35). Discrete samples for inorganic nutrient concentrations (NO₃, NO₂, NH₄, PO₄, SiO₄) were frozen and stored at -20°C until analysis. Concentrations were determined using a FlowSys Autoanalyzer (Systea SpA, Italy) (36). Detection limits were 0.01 µM for nitrates, nitrites and phosphates, 0.05 µM for ammonia and 0.1 µM for silicates.

Flow cytometry

Bacterioplankton cell concentrations were estimated using a Becton Dickinson FACScalibur (Becton Dickinson, USA), following standard procedures (37). For assessment of heterotrophic bacteria frozen samples were thawed and stained with SYBR Green I (Molecular Probes Inc., USA). Cell abundances were extracted using CellQuest software (Becton Dickinson, USA).

Nucleic acid extraction and sequencing

DNA extraction was performed using the DNeasy Blood and Tissue Kit (Qiagen, USA) according to the manufacturer's instructions, with the only minor modification being overnight incubation at 56°C following addition of the Qiagen lysis buffer and Proteinase K to the Sterivex filters. RNA extraction was performed using the PowerWater[®] RNA Isolation kit (Mo Bio laboratories, USA) according to the manufacturer's protocol. DNA contamination was checked by PCR of 16S rRNA gene using standard primers and conditions (38). The minimum amount (20 ng) of extracted RNA was reverse transcribed with random primers using the iScript[™] cDNA Synthesis kit (Biorad, USA), following the manufacturer's instructions. DNA and cDNA samples were shipped to the GeneCore Genomics Core Facility at the European Molecular Biology Laboratory, Heidelberg (Germany), for paired-end multiplexed sequencing using the Illumina MiSeq platform. The PCR protocol used for amplifying DNA and cDNA was standardized to use minimum amount of template. Primers used for library preparation covered V4 and V5 hypervariable regions (515F-926R) of the 16S rRNA gene, as in reference 39, with only one base difference (Table 1).

Nucleotide sequence accession numbers

Sequences obtained in this study have been deposited in the European Nucleotide Archive (ENA) with accession numbers from ERS1162689 to ERS1162707 (study accession number- PRJEB14040).

Sequence analyses

Low quality raw reads with a contaminating adapter sequence or with a Phred value less than 25 were excluded using Trimmomatic (40). The resulting sequences were further processed and analysed using the Mothur software package (version 1.33.1, 41), following the MiSeq standard operating procedure (SOP) (http://www.mothur.org/wiki/MiSeq_SOP, 42). Reads were joined into contigs using the sequence and quality score data from the Fastq files. Primers were trimmed and sequences were further screened according to the following parameters: minimum length=370, maximum length=376, maximum number of Ns=0, maximum homopolymers=8. Unique sequences, i.e. randomly screened sequences from groups of identical sequences present in two or more copies, were screened out in order to speed up the process and to avoid computational memory problems. The resulting sequences were aligned against the SILVA reference database (version 102) and pre-clustered allowing a maximum of three differences between sequences. Chimeric sequences were detected with the UCHIME algorithm (43) and removed. After this filtration, taxonomic assignment was realized using the SILVA taxonomy string with the k-nearest-neighbour algorithm at a bootstrap confidence score of 100 and a minimum similarity threshold of 80%. Across all 16S rDNA libraries, a total of 8,811,540 raw reads were obtained, from which 1,216,213 high quality sequences with an average read length of 370 to 376 were used for taxonomic assignment. Only a very small number (10 sequences) of Archaeal sequences assigned to *Euryarchaeota* and *Crenarchaeota* appeared in the whole dataset, and these were not included in further analysis. Sequences flagged as chloroplast, mitochondria or eukaryotes were also excluded. The remaining 1,210,304 sequences were clustered *de novo* using the average neighbor algorithm at a distance cut-off of 97% similarity resulting in 2681 OTUs. OTUs identified as singletons were excluded from the dataset to limit the inflation caused by spurious OTUs (PCR artefacts) and consensus taxonomy for each of the remaining 1235 OTUs was obtained. The full lists of OTUs for the 16S rDNA and the 16S rRNA libraries are provided as supplementary files “dataset 1 and 2”.

Diversity estimates and statistical analyses

The 16S rDNA libraries prepared using samples collected from all the 16 stations produced 26,177 to 151,291 reads, 487 to 1362 unique sequences, and 281 to 715 OTUs after quality filtration (details for each station are in Table S1). Rarefaction curves for the observed OTUs, Chao I and Shannon Diversity Index were generated using the “rarefaction.single” command built in Mothur (Fig. S2) and compared as in [44](#) to evaluate the sampling effort and the alpha diversity of the different communities. Good’s coverage was >99% for all the samples, indicating that the majority of the species present were sufficiently sampled. Good’s coverage, alpha diversity indices (Shannon-Wiener ([45](#), [46](#)) and invsimpson ([47](#)), richness estimators (ChaoI ([48](#)) and ACE (Abundance-based richness estimation, [49](#)), were computed based on multiple times randomly resampled to make the sample size equal to the sample with the lowest number of sequences (26177, MC station). The Berger-Parker dominance index ([50](#)) was also calculated but on non-normalized data. Beta diversity measures were performed on normalized sample sizes in R using the Vegan package. Sample clustering using the Bray-Curtis dissimilarity index was carried out using the functions “vegdist” and “hclust” in R software. Statistic comparison of Bray-Curtis dissimilarity between community clusters and sub-clusters was performed using pairwise analysis of similarity (ANOSIM) using the software Primer 6 ([51](#)). The main operationally defined taxa contributing to > 90% similarity within Bray-Curtis clusters of bacterial communities were identified using similarity percentage (SIMPER) analysis ([52](#)) in Primer 6 ([51](#)). Based on SIMPER analysis, we were able to identify what taxa are driving the differences observed in the composition between sites in the GON, and then grouped all the OTUs with the same affiliation at the Order/Family level.

Relationships between bacterial communities and environmental parameters were explored using Canonical Correspondence Analysis (CCA) using the Vegan package in R software, using envfit parameter. Data were log (x+1) transformed prior to any statistical analysis. Distribution patterns of all the abundant taxa in each station were visualized by generating heatmaps in the MeV program (<http://www.tm4.org/mev.html>).

A Linear Discriminant Analysis (LDA) Effect Size pipeline (LEfSe, [53](#)), available at <http://huttenhower.sph.harvard.edu/galaxy/>, was used to determine significant differentially distributed OTUs. A threshold of 2.0 was chosen as threshold for the logarithmic LDA scores.

A sequence from each OTU was randomly selected using “get.oturep” command in Mothur and used to align in ClustalW incorporated as an accessory application in the BioEdit software package (version 7.2.5) ([54](#)). Maximum-likelihood phylogenetic distances were calculated using the aligned representative sequences in FastTree ([55](#)). The corresponding tree was visualized in MEGA (version 6) ([56](#)) and was used for UniFrac analysis ([57](#)), in order to compare the phylogenetic distance between microbial communities as a function of different sites. In order to assess which stations in the tree had similar bacterial assemblages, a hierarchical clustering with abundance weight was performed, based on the distance matrix generated by computing pairwise UniFrac distance. Robustness of the clustering was determined by weighted Jackknife analysis with random resampling of the sequences and clustering with 1000 permutations while calculating support for each node.

Potential metabolic activity

The ratio of 16S rRNA:rDNA was used to assess potential specific activity relative to abundance of each operationally defined taxon at the three stations marked in Fig. 1 (VE3, PO1 and TA1), from where we could obtain good quality RNA. All samples were extracted, amplified and analyzed using the same procedure to avoid any bias. Singletons were removed and data were natural log(x+1) transformed prior to statistical testing. All tests were performed on normalized datasets with equal number of sequences for both 16S rDNA and 16S rRNA libraries of each sample. Since assumptions of parametric regressions were not fulfilled even after normalization of the data, nonparametric Spearman and Kendall’s tau tests were performed in R software. Heatmaps were generated to visualize differences in 16S rDNA and 16S rRNA abundances of the same OTU between sites and of different OTUs within each site. Differences in 16S rRNA:rDNA ratios between OTUs were also assessed within and between sites. Sequences were weighted to bacterial abundances.

Results and Discussion

Temperature, salinity, inorganic nutrients and chlorophyll a

To gain comprehensive information about bacterial diversity and community composition in surface waters of the GON, a total of sixteen stations spanning both coastal and offshore sites were sampled (Fig. 1). The values of all measured environmental parameters are reported in Table 2. Temperature and salinity values were quite homogeneous across the sampled region, ranging from 15.75°C (SA6) to 16.56 °C (TA2) and from 36.13 PSU (MC) to 37.52 PSU (PO3), respectively, indicating active mixing in the area. As a consequence, stations located in the Sarno river catchment (SA1 to SA5) did not show significantly lower salinity values, as expected from freshwater discharge, though relatively higher nutrient concentrations (especially N and P) detected at these stations (especially SA1) were interpreted as indicators of river influence (Table 2). The highest salinity (37.40 PSU and 37.52 PSU) and lowest chlorophyll *a* (0.93 and 0.62 µg L⁻¹) values were recorded at stations PO3 and VE3, respectively, indicative of offshore oligotrophy.

Bacterial abundance

Abundances of phototrophic and heterotrophic bacteria enumerated by flow cytometry are reported in Table 2 and Fig. S1. Heterotrophs dominated numerically with average values of 1.15 ± 0.57 (SD) $\times 10^6$ cells ml⁻¹, while autotrophs were dominated by *Synechococcus* (average values 1.11 ± 0.45 (SD) $\times 10^4$ cells ml⁻¹), at all stations. *Prochlorococcus* was detected at only 5 stations with concentrations between 1.11×10^3 and 2.52×10^3 cells ml⁻¹. No clear patterns in *Synechococcus* abundance were observed, while heterotrophic bacteria were more abundant at the MC station and near the Sarno river and less abundant at the offshore stations VE3 and PO3.

Alpha and beta diversity

The rarefaction curves of the observed OTUs (Fig. S2a) approached saturation only for stations SA1, TA1, TA2, SA2 and SA5, indicating insufficient sampling depth for the others. Although Chao I curves (Fig. S2b) were steeper, they did not saturate for each library. However, the Shannon rarefaction curves (Fig. S2c),

which consider both richness and evenness, showed saturation and also a stable pattern in all samples, suggesting that sampling was sufficient to accurately describe the trends in alpha-diversity. Furthermore, Good's coverage was >99% for all the samples (Table 3), indicating that, although possibly more rare taxa could be recovered by a deeper sampling, the rarefied sequencing depth we used was satisfactorily representing alpha diversity at the sampling sites.

The Chao I diversity index was higher at most stations near the mouth of the Sarno river (Table 3) and was positively correlated with ammonium (Pearson's $r=0.66$, $p<0.05$). This pattern has been previously observed along coast to offshore transects, for instance in the Southern Adriatic Sea (58), Blanes Bay in the northwestern Mediterranean Sea (59), the South China Sea (60) and Moreton Bay in Australia (61). The higher richness in coastal areas may be due to terrestrial inputs, the availability of organic and inorganic matter at high concentration and to continuous mixing of local and external communities, resulting numerous favourable niches for bacterioplankton communities. The higher rates of primary production in coastal areas provide a larger fraction of organic matter in the form of detritus, most of which is degraded by heterotrophic bacteria prior to entering higher trophic levels, supporting bacterial communities with higher richness in these areas (62).

The Bray-Curtis dissimilarity dendrogram clustered the samples into four different groups (Fig. 2a). Group I grouped the offshore PO3 and VE3 stations, confirming the similarity inferred from salinity and chlorophyll *a* data. Group II included stations TG1, TG2 and PO1, located in front of the heavily urbanized areas of the neapolitan province. Group III included stations TA1, TA2, SA5 and SA8, north and south of the zone of influence of the Sarno river, whilst Group IV included stations SA7, SA1, SA2, SA3, SA4 and SA6, which are closest to the Sarno river. Groups III and IV sub-clustered stations TA1 and TA2 (IIIa) and SA5 and SA8 (IIIb). Group IV sub-clustered stations SA1 and SA2 (IVa) and SA4 and SA6 (IVb). The MC station was not included in any of the clusters. Bray-Curtis clustering of samples was confirmed to be significant using ANOSIM R statistic (Table S2), except for the lack of difference between group II and III ($R=0.75$, $p>0.05$), hence these two groups were considered as one cluster from now on. The OTUs responsible for Bray-Curtis

similarity within groups were identified using similarity percentage (SIMPER) analysis (Table S3). Bacterial communities in Group I, II+III and IV were 88.7%, 88.6% and 88.8% similar respectively, mainly due to similar proportions of the same OTUs (Table S2). The MC station showed relatively high proportions of some OTUs (including some rare ones) as compared to all other stations, such as those attributed to *Nereida* spp., *Flavobacteraceae*, *Gammaproteobacteria*, *Oceanospirillaceae*, the Roseobacter clade, *Rickettsiales*, *Flavobacterium* and *Cryomorphaceae*, which are probably responsible for its outgrouping in the Bray-Curtis dendrogram.

Bacterial community composition

A total of 1235 OTUs, after excluding singletons, were used to assess bacterial community composition; 99% of these could be classified at the phylum level. These were affiliated to 17 major phyla and 3 candidate divisions (Fig. 3a). Overall, the surface bacterioplankton at all stations was dominated by *Proteobacteria* (~69%) followed by *Bacterioidetes* (~27%), as in other studies from the Mediterranean Sea (63). *Proteobacteria* were also the most diverse bacterial group, comprising the highest number of retrieved OTUs (47.2% of the whole dataset). Of these, the most frequently occurring OTUs belonged to *Gammaproteobacteria* and *Alphaproteobacteria* (35% and 32% of *Proteobacteria*, respectively), mainly represented by *Oceanospirillales* and *Alteromonadales* the former (18% and 14% of the *Gammaproteobacteria*, respectively) and *Rickettsiales* the latter (19% and 12% of the *Alphaproteobacteria*, respectively).

Similar to studies conducted in surface waters at the global scale (9, 10), the dominant components of bacterial communities of the GON were present at all stations but in different proportions (Fig. 3b). Based on SIMPER analysis and following OTU grouping, *Rhodobacterales* resulted to be the highest contributing clade at coastal sites (Group II+III stations), while SAR11 and SAR86 were the most abundant offshore (Group I stations). This is consistent with previous observations in the Southern Adriatic Sea (58) and the north-western Mediterranean Sea (64) about SAR11 distribution. Members of SAR11 and SAR86 are well known for their ability to grow at low substrate concentrations, due to their streamlined genomes (65, 66) and light-

harvesting proteorhodopsins for ATP production (66, 67) likely conferring them a selective advantage in oligotrophic waters (68). In contrast, the dominance of members of the *Rhodobacterales* (mainly of the genus *Nereida* and the *Roseobacter* clade) at coastal stations is likely due to the higher nutrient concentrations and chlorophyll *a* levels at these sites, given their known ability to thrive in more eutrophic environments with high primary productivity (69) and to use phytoplankton-derived dissolved organic carbon, especially originating from diatoms (e.g. 70), which abundantly bloom in the GON (71). At river-influenced stations (Group IV) *Alteromonadales* (mainly *Glaciecola* and SAR92) dominated, as related to their preference for high nutrient concentrations, a feature in good agreement with their copiotrophic nature (72), the exception being the SAR92 clade which was more abundant at offshore stations, as already reported (73).

Cyanobacteria represented ~1% of the total bacteria, with *Synechococcus* and *Prochlorochoccus* present at all stations, but increasing their contribution at both coastal and offshore (*Synechococcus*) or offshore stations alone (*Prochlorochoccus*), only partially matching flow cytometry data (Table 2, Fig. S1). This shows that sequencing was robust enough to retrieve *Prochlorochoccus* even in samples where it was below detection limits of flow cytometry. Contrary to other studies (e.g. 58, 74, 75), *Cyanobacteria* did not dominated in the GON neither showed a clear distribution pattern, which may support their high genetic diversity with physiologically diverse clades in terms of light and nutrient adaptation (76, 77, 78). Thus, a more precise mapping of the microdiversity among closely related cyanobacterial lineages is critical to understand their distribution pattern in this area.

In order to identify significant differential distribution of bacterial groups at the OTU level in relation to environmental parameters, LefSe analysis was applied. For this test, we considered stations within Group I (offshore), II +III together (coastal), and IV (river-influenced), as the ones identified by the Bray-Curtis dendrogram (Fig. 2a), and confirmed by ANOSIM (Table S2). A total of 87 OTUs, including both “rare” and “abundant” (<0.1% and >0.1% of total sequences, respectively), were found to be differentially abundant in the three areas, with 14 OTUs specialized for river-influenced, 18 for coastal and 55 for offshore stations (Fig. 5a). However, many OTUs, expected to be specialized at offshore stations (e.g. SAR11, SAR92) were not

highlighted by this test, due to the fact that they showed similar numerical distribution at both coastal and river-influenced stations, and LefSe takes into account only the OTUs that are differentially abundant between the groups considered. Therefore, we performed a second LefSe test considering river-influenced and coastal stations together (Group II+III+IV), named “coastal+river-influenced”, and compared this merged group with the offshore group (Group I). This second LefSe test showed that a total of 113 OTUs, including both “rare” and “abundant”, were differentially abundant, with 82 specialized for offshore and 31 for coastal+river-influenced (Fig. 5b). It is important to note that the offshore area consisted in a higher number of significant, differentially abundant OTUs than the coastal areas, contributed mainly by rare OTUs, suggesting a niche-adaptation strategy adopted by these low-abundant and highly diverse taxa (34). The analysis at the OTU level is necessary to understand bacterial distribution patterns with the highest resolution. Our results highlight that variability of bacterial community composition across different sites in the GON is driven not only by abundant but also by rare species, whose spatial dynamics needs further thorough description and understanding.

OTU frequency

Forty-eight OTUs (95.8% of the total sequences) were “abundant” (here defined as OTUs accounting for >0.1% of the total sequences) at either all or $\geq 50\%$ of total stations. 48 OTUs (2.6% of the total sequences) were observed to be abundant at <50% of all stations, whilst 1139 OTUs (1.6% of the total sequences) were considered “rare” (here defined as OTUs accounting for <0.1% of total sequences). These numbers suggest that a tremendous bacterial diversity in the GON is accounted for by the so-called “rare biosphere” *sensu* (79), which is in agreement with other studies (58, 59, 80). Among the rare OTUs, *Arcobacter*, *Chlamydiae*, *Fermicutes* (mainly *Ruminococcaceae*, *Veillonellaceae*, *Lachnospiraceae*, *Staphyococcaceae*), *Fusobacteria*, *Lentisphaeraceae* (*Lentisphaera* and *Victivallis*), *Spirochaetes*, *Synergisestes* and *Vibrio* spp. were observed ($\leq 0.01\%$ of total sequences) and were attributed to sewage-associated-bacteria or pathogens (e.g. 81, 82, 83), indicating the dispersal of faecal bacteria from the anthropogenically affected coastal area and suggesting that they can survive in marine waters and therefore represent a reservoir of pathogenicity potentially harmful for

human health.. Similarly, the presence of *Rhodocyclales*, which are known to degrade several pollutants (84) is suggestive of pollutant biodegradation occurring in the GON, as related to the many anthropogenic activities along the GON coastline.

Bacterial communities and environmental factors

In order to identify the specific environmental factors that might explain bacterial community composition in the GON, Canonical Correspondence Analysis (CCA) was performed using OTU abundances and the measured environmental parameters (Fig. 4). The first two axes explained 78% and 18% of the cumulative variance, respectively. Monte Carlo permutations (999 permutations) indicated that only salinity ($P=0.004$), chlorophyll *a* ($P=0.001$) and ammonium concentrations ($P=0.012$) were significant determinants of bacterial community structure in terms of abundance of OTUs, based on the envfit parameter. Ammonium was a determinant at the Sarno river-influenced stations (Fig. 2a Group IV), which was dominated by *Alteromonadales*. Likewise, bacterial community structure at the MC station, which is largely explained by chlorophyll *a*, is dominated by members of the *Rhodobacterales*. Salinity was the dominant variable at the oligotrophic sites VE3 and PO3 that were mainly dominated by SAR11. The effect of other environmental parameters such as temperature, nitrate, nitrite, silicate and phosphate were probably masked by the complex physical features of the GON at the time of sampling, including mixing. These results indicate that, similar to other marine areas (58, 85), it is a combination of several environmental factors influencing the local distribution of bacterial populations in the GON. Indeed, the unexplained variation in the CCA is likely due to unmeasured environmental variables and processes, the latter including top-down controls like grazing and viral lysis that are also known to shape bacterial community composition.

In order to further compare the phylogenetic diversity of the bacterial communities between stations, UniFrac distance metrics were calculated based on a Maximum-Likelihood tree constructed using one representative sequence from each OTU (Fig. 2b). Weighted hierarchical clustering showed that community composition was more similar at those sites located in close vicinity to each other or with similar features.

Although similar clustering was also mirrored by Bray Curtis dendrogram (Fig. 2a) and CCA (Fig. 4), UniFrac analysis revealed that environmental heterogeneity not only regulates the spatial distribution of abundance of different taxonomic groups at the local scale in the GON, but may also have resulted in distinct phylogenetic diversity (see Fig. 2b).

Bacterial potential metabolic activity

The potential activity of each operationally defined taxon and the correlation between potential activity (16S rRNA frequency) and abundance (16S rDNA frequency), was assessed at the three stations where good quality RNA could be obtained: VE3, PO1, and TA1. These stations are also representative of the coastal (TA1 and PO1) and offshore areas (VE3). 16S rRNA:rDNA ratios may have been affected by the extraction of DNA and RNA from different lysates, but libraries were normalized by equal number of sequences before performing calculations or statistical tests, in order to avoid any bias (as also indicated in the Material and Methods section).

Overall, rRNA and rDNA libraries were positively correlated at all the three sites (Fig. 6) (Spearman's and Kendall's $P < 0.0001$, Kendalls' τ : VE3=0.22, PO1=0.52, TA1= 0.53). The average 16S rRNA:rDNA ratio for rare (<0.1% of total sequence abundances) and abundant bacteria was 2.3 (0.2 - 8.47, n=209) and 0.71 (0.09 - 1.51, n=79) respectively, at station VE3; 1.63 (0.27 - 6.67, n=103) and 0.95 (0.18 - 2.2, n=57) respectively, at station PO1; and 1.2 (0.22 - 4.90, n=87) and 0.8 (0.19 - 1.92, n=57) respectively, at station TA1, with the ratio for rare bacteria being significantly higher than the ratio for abundant bacteria ($P < 0.05$, t-test). At all three stations, most (79.80% at VE3, 54.40% at PO1 and 70.20% at TA1) of the abundant taxa had a ratio <1, whereas most (84.60% at VE3; 77.66% at PO1 and 52.84% at TA1) of the rare taxa had a ratio ≥ 1 . For instance, an abundant SAR11 OTU (ID-10), showed ratios <1 at all the three sites (Fig. 7), similar to what observed in other studies for SAR11 (22). Likewise, OTUs attributed to *Rhodobacteraceae* and *Flavobacteriaceae* showed a 16S rRNA:rDNA ratio <1. However, cultivated members of these groups have been previously reported to possess more than one ribosomal operon (22), which may explain this ratio. As for the rare OTUs, *Blastopirellula* (*Planctomycetaceae*) showed a very high 16S rRNA:rDNA ratio (8.5) at

the VE3 site. The possibility of a sequencing artifact for this high ratio was excluded since the same behavior was observed for other *Blastopirellula* OTUs identified at the same station. These results suggest that, although potential activity and abundance showed a significant positive correlation at the whole community level, at the individual taxa level this may not hold true and therefore, abundance may not reflect the degree of their potential activity, as also observed in previous studies in other geographical areas (17, 18). Whilst the role of the abundant phylotypes in ecosystem functioning is well recognized, contribution of the rare ones remained largely underestimated until recently, due to the previous assumption of them being merely the slow-growing and dormant members of the community (20). In general, the abundant bacterial phylotypes altogether contribute a higher proportion of biomass production or activity, but not always on a per-cell basis, while rare bacteria can have 16S rRNA:rDNA ratios >1, therefore exhibiting higher potential activity than the abundant ones, as reported by us and others (17, 18, 22). Unfortunately, due to our small sample size, correlations between 16S rRNA and rDNA for individual OTUs could not be statistically tested.

OTUs were ranked on the basis of their abundance in each library. Differences in the relative rank of the same OTU among different sites, and between rDNA and rRNA libraries of the same site, were observed. For instance, the majority of the top 50 most abundant OTUs were shared by all six libraries, but the relative abundance of these OTUs varied between sites. In addition, variations in sequence abundance of the same OTU between rDNA and rRNA libraries resulted in substantial variation in the 16S rRNA:rDNA ratio among OTUs, both within and between stations (Fig. 7). These differences can be attributed to different biotic and abiotic environmental factors (as in 18, 86) or to differences in evenness of diversity between rDNA and rRNA at different stations (as also found by 17). Both hypotheses are possible in our case, since the average Shannon index of the rDNA and rRNA libraries were 3.05 ± 0.24 and 2.44 ± 0.80 respectively, and values of the measured environmental factors were also different. Differences in the 16S rRNA:rDNA ratio can also be due to life strategies adopted by some bacteria, for instance, the increase of ribosome concentration as they enter dormancy to achieve higher protein synthesis potential to be used after recovery to vegetative stage in response to favorable cues (87).

Conclusions

The observed strong spatial structuring of bacterial communities in the GON is tightly regulated by environmental factors and local hydrological features, suggesting a niche-based community assembly, in accordance with the famous dictum ‘everything is everywhere, but the environment selects’ (88). Further evaluation of bacterial association and co-occurrence network is needed to gain additional insights on the potential relationships within these communities as well as to reveal ecological processes such as top-down and bottom-up regulation and community succession, as suggested by other researchers (eg. 89). Regular monitoring of bacterioplankton community structure and activity, both at the seasonal and annual scale will facilitate a better understanding and predictions of their responses to the changing environment. In particular, in our data salinity appears to play a relevant role, and in this context the GON may represent an interesting test site to predict the responses of the microbial components to the changing climate-induced salinity increase in the oceans (90). Our study also highlights the relevance of the “rare” phylotypes in the heterogeneous bacterial diversity and their potential role in ecosystem functioning, suggesting that they harbour a persistent functional pool of ecological potential rather than acting merely as a “seed bank” (91). Therefore, assessment of their spatio-temporal dynamics and mechanisms controlling their population size is crucial.

Acknowledgements

We thank Jed A Fuhrman for his advice on the 16S rDNA primers used in this study, Patrick Schloss and his team for their suggestions regarding “Mothur”, Adriana Zingone for advice, Francesco Musacchia for Trimmomatic analysis, the Captains and crew of the RV “Minerva Uno” and “Vettoria”, Fabio Conversano, Gianluca Zazo, Violante Stefanino and Marco Cannavacciuolo for valuable help during sampling and for chlorophyll analyses. We also thank two anonymous referees who provided useful comments to improve the manuscript. This study was supported by the RITMARE Flag Project from the Italian Ministry of the

University and Research (MIUR). K.R. was supported by a Fellowship from Stazione Zoologica ‘Anton Dohrn’ in the framework of the SZN-Open University PhD Program. K.R. prepared the DNA, analyzed the FCM samples, performed the bioinformatics analyses, elaborated all data and wrote the first draft of the manuscript. C.B. collected the samples and ran part of the flow cytometry samples, R.P. helped with the statistical analyses, V.B. performed the sequencing, E.B. and M.B. performed a previous library preparation, F.M. coordinated the nutrient and chlorophyll analyses, A.P. coordinated the sampling and performed nutrient analyses, M.S. participated to sampling, R.S. supported and supervised K.R. on bioinformatics, D.S. supervised K.R. work and reviewed the manuscript, R.C. conceived the study, coordinated the different contributions, supervised K.R. and elaborated, prepared and reviewed the manuscript together with K.R.

References

1. **Fuhrman, JA.** 2009. Microbial community structure and its functional implications. *Nature* **459**: 193-199.
2. **Amaral-Zettler L, Artigas LF, Baross J, Bharathi L, Boetius A, Chandramohan D, Herndl G, Kogure K, Neal P, Pedrós-Alió C, Ramette A, Schouten S, Stal L, Thessen A, Leeuw JD, Sogin M.** 2010. A global census of marine microbes, p 223-245. *In* McIntyre AD (ed), *Life in the world’s oceans*, Blackwell Publishing Ltd, Oxford.
3. **Benner R, Herndl G.** 2011. Bacterially derived dissolved organic matter in the microbial carbon pump, P 46-48. *In* Jiao N, Azam F, Sanders (ed), *Microbial carbon pump in the ocean*, edited by Jiao, N., Azam, F, Sanders, Science/AAAS, Washington, DC.
4. **Pace NR.** 1997. A molecular view of microbial diversity and the biosphere. *Science* **276**:734-740.
5. **Follows MJ, Dutkiewicz S, Grant S, Chisholm SW.** 2007. Emergent biogeography of microbial communities in a model ocean. *Science* **315**:1843-1846.
6. **DeLong EF, Karl DM.** 2005. Genomic perspectives in microbial oceanography. *Nature* **439**: 1014.

7. **Herlemann DP, Labrenz M, Jürgens K., Bertilsson S, Waniek JJ, Andersson AF.** 2011. Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *ISME J.* **5**:1571-1579.
8. **Zheng X, Dai X, Huang L.** 2016. Spatial variations of prokaryotic communities in surface water from India Ocean to Chinese marginal seas and their underlining environmental determinants. *Front. Mar. Sci.* **3**:17-26.
9. **Pommier T, Canbäck B, Riemann L, Boström K, Simu ., Lundberg P, Tunlid A, and Hagström Å.** 2007. Global patterns of diversity and community structure in marine bacterioplankton. *Mol. Ecol.* **16**: 867-880.
10. **Zinger L, Amaral-Zettler LA, Fuhrman JA, Horner-Devine MC, Huse SM, Welch DBM, Martiny JB, Sogin M, Boetius A, and Ramette A.** 2011. Global patterns of bacterial beta-diversity in seafloor and seawater ecosystems. *PloS One* **6**: e24570.
11. **Armbrust EV, Palumbi SR.** 2015. Uncovering hidden worlds of ocean biodiversity. *Science* **348**:865–867.
12. **Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K., Salazar G, Djahanschiri B, Zeller G, Mende DR, Alberti A.** 2015. Structure and function of the global ocean microbiome. *Science* **348**:10.1126/science.1261359.
13. **Ducklow H.** 2000. Bacterial production and biomass in the oceans, p 85-120. *In* Kirchman DL (ed), *Microbial Ecology of the Oceans*, John Wiley and Sons, New York.
14. **Fouilland E, Mostajir B.** 2010. Revisited phytoplanktonic carbon dependency of heterotrophic bacteria in freshwaters, transitional, coastal and oceanic waters. *FEMS Microbiol. Ecol.* **73**:419-429.
15. **Fuhrman J, Azam F.** 1982. Thymidine incorporation as a measure of heterotrophic bacterioplankton production in marine surface waters: evaluation and field results. *Mar. Biol.* **66**:109-120.
16. **Kirchman D, K'nees E, Hodson R.** 1985. Leucine incorporation and its potential as a measure of protein synthesis by bacteria in natural aquatic systems. *Appl. Environ. Microbiol.* **49**:599-607.

17. **Hunt DE, Lin Y, Church MJ, Karl DM, Tringe SG, Izzo LK, Johnson ZI.** 2013. Relationship between abundance and specific activity of bacterioplankton in open ocean surface waters. *Appl. Environ. Microbiol.* **79**:177-184.
18. **Campbell BJ, Kirchman DL.** 2013. Bacterial diversity, community structure and potential growth rates along an estuarine salinity gradient. *ISME J.* **7**:210-220.
19. **Blazewicz SJ, Barnard RL, Daly RA, Firestone MK.** 2013. Evaluating rRNA as an indicator of microbial activity in environmental communities: limitations and uses. *ISME J.* **7**:2061-2068.
20. **Lennon JT, Jones SE.** 2011. Microbial seed banks: the ecological and evolutionary implications of dormancy. *Nature Rev. Microbiol.* **9**:119-130.
21. **Jones SE, Lennon JT.** 2010. Dormancy contributes to the maintenance of microbial diversity. *Proc. Natl. Acad. Sci. U.S.A.* **107**: 5881- 5886.
22. **Campbell BJ, Yu L, Heidelberg JF, Kirchman DL.** 2011. Activity of abundant and rare bacteria in a coastal ocean. *Proc. Natl. Acad. Sci. U.S.A.* **108**:12776-12781.
23. **Zingone A, Dubroca L, Iudicone D, Margiotta F, Corato F, d'Alcalà MR, Saggiomo V, Sarno D.** 2010. Coastal phytoplankton do not rest in winter. *Estuaries Coasts* **33**:342-361.
24. **Cianelli D, Uttieri M, Buonocore B, Falco P, Zambardino G, Zambianchi E.** 2011. Dynamics of a very special Mediterranean coastal area: the Gulf of Naples, p 129-150. *In* Williams GS (ed), *Mediterranean Ecosystems: Dynamics, Management and Conservation*, Nova Science Publishers.
25. **Grieco L, Tremblay LB, Zambianchi E.** 2005. A hybrid approach to transport processes in the Gulf of Naples: an application to phytoplankton and zooplankton population dynamics. *Cont. Shelf Res.* **25**: 711-728.
26. **Carrada GC, Fresi E, Marino D, Modigh M, Ribera d'Alcalà M.** 1981. Structural analysis of winter phytoplankton in the Gulf of Naples. *J. Plankton Res.* **3**:291-314.

27. **Marino D, Modigh M, Zingone A.** 1984. General features of phytoplankton communities and primary production in the Gulf of Naples and adjacent waters, p 89-100. *In* Holm-Hansen O, Bolis L, Gilles L (ed), *Marine Phytoplankton and Productivity*, Springer, Berlin.
28. **Carrada G, Hopkins T, Bonaduce G, Ianora A, Marino D, Modigh M, Ribera d'Alcalà M, Scotto CB.** 1980. Variability in the hydrographic and biological features of the Gulf of Naples. *Mar. Ecol.* **1**:105-120.
29. **Mazzocchi MG, Dubroca L, García-Comas C, Di Capua I, Ribera d'Alcalà M.** 2012. Stability and resilience in coastal copepod assemblages: the case of the Mediterranean long-term ecological research at Station MC (LTER-MC). *Prog. Oceanogr.* **97**:135-151.
30. **Modigh M, Saggiomo V, Ribera d'Alcala M.** 1996. Conservative features of picoplankton in a Mediterranean eutrophic area, the Bay of Naples, *J. Plankton Res.* **18**:87-95.
31. **Casotti R, Brunet C, Aronne B, Ribera d'Alcala M.** 2000. Mesoscale features of phytoplankton and planktonic bacteria in a coastal area as induced by external water masses. *Mar. Ecol. Prog. Ser.* **195**:15-27.
32. **Izzo G, Tosti E, Volterra L.** 1983. Faecal contamination of marine sediments in a stretch of the Gulf of Naples. *Water Air Soil Pollut.* **20**:191-198.
33. **Volterra L, Tosti E, Vero A, Izzo G.** 1985. Microbiological pollution of marine sediments in the southern stretch of the Gulf of Naples. *Water Air Soil Pollut.* **26**:175-184.
34. **Alonso-Sáez L, Díaz-Pérez L, Morán XA.** 2015. The hidden seasonality of the rare biosphere in coastal marine bacterioplankton. *Environ. Microbiol.* **17**:3766-80.
35. **Holm-Hansen O, Lorenzen CJ, Holmes RW, Strickland JD.** 1965. Fluorometric determination of chlorophyll. *J. Cons. Int. Explor. Mer.* **30**:3-15.
36. **Hansen HP, K Grasshoff.** 1983. Automated chemical analysis. *In* K Grasshoff, M Ehrhardt, K Kremling (ed), *Methods of Seawater Analysis*, Verlag Chemie, Weinheim.

37. **Balestra C, Alonso-Sáez L, Gasol JM, Casotti R.** 2011. Group-specific effects on coastal bacterioplankton of polyunsaturated aldehydes produced by diatoms. *Aquat. Microb. Ecol.* **63**:123-31.
38. **Lane DJ.** 1991. 16S/23S rRNA sequencing, p 115–147. *In*: Stackebrandt E, Goodfellow M (eds). *Nucleic Acid Techniques in Bacterial Systematics*, John Wiley & Sons, New York.
39. **Parada A, Needham DM, Fuhrman JA.** 2015. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time-series and global field samples. *Environ. Microbiol.* <http://dx.doi.org/10.1111/1462-2920.13023>.
40. **Bolger AM, Lohse M, Usadel B.** 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 10.1093/bioinformatics/btu170.
41. **Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ.** 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75**:7537-7541.
42. **Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss P D.** 2013. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl. Environ. Microbiol.* **79**:5112-5120.
43. **Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R.** 2011. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**:2194-2200.
44. **Lundin D, Severin I, Logue JB, Östman Ö, Andersson AF, Lindström ES.** 2012. Which sequencing depth is sufficient to describe patterns in bacterial α - and β - diversity?. *Environ. Microbiol. Rep.* **4**:367-72.
45. **Shannon CE, Weaver W.** 1949. *The mathematical theory of information*. University of Illinois Press, Urbana, IL.
46. **Krebs CJ.** 1989. *Ecological Methodology*. Harper & Row, New York.
47. **Simpson EH.** 1949. Measurement of diversity. *Nature*, **163**:688.

48. **Chao A.** 1984. Non-parametric estimation of the number of classes in a population. *Scand. J. Stat.* **11**:265–270.
49. **Chao A, SM Lee.** 1992. Estimating the number of classes via sample coverage. *JASA.* **87**:210–217.
50. **Magurran M.** 1988. *Ecological Diversity and its Measurement.* Cambridge University Press, Cambridge, UK.
51. **Clarke KR, Gorley RN.** 2006. *PRIMER v6: user manual/tutorial*, p 190. Plymouth: PRIMER-E.
52. **Clarke KR.** 1993. Non-parametric multivariate analyses of changes in community structure. *Aust. J. Ecol.* **18**:117-143.
53. **Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, Huttenhower C.** 2011. Metagenomic biomarker discovery and explanation. *Genome Biol.* **12**:R60.
54. **Hall TA.** 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **41**:95-98.
55. **Price MN, Dehal PS, Arkin AP.** 2010. FastTree 2—approximately maximum-likelihood trees for large alignments. *PloS One* **5**:e9490.
56. **Tamura K, Stecher G, Peterson D, Filipski A, Kumar S.** 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **30**:2725-2729.
57. **Lozupone C, Knight R.** 2005. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* **71**:8228-8235.
58. **Quero GM, Luna GM.** 2014. Diversity of rare and abundant bacteria in surface waters of the Southern Adriatic Sea. *Mar. Genomics* **17**:9-15.
59. **Pommier T, Neal PR, Gasol JM, Coll M, Acinas SG, Pedrós-Alió C.** 2010. Spatial patterns of bacterial richness and evenness in the NW Mediterranean Sea explored by pyrosequencing of the 16S rRNA. *Aquat. Microb. Ecol.* **61**:221-33.
60. **Du J, Xiao K, Li L, Ding X, Liu H, Lu Y, Zhou S.** 2013. Temporal and spatial diversity of bacterial communities in coastal waters of the South China Sea. *PloS One.* **8**:e66968.

61. **Hewson I, Fuhrman JA.** 2004. Richness and diversity of bacterioplankton species along an estuarine gradient in Moreton Bay, Australia. *Appl. Environ. Microbiol.* **70**: 3425-3433.
62. **Mou X, Sun S, Edwards RA, Hodson RE, Moran MA.** 2008. Bacterial carbon processing by generalist species in the coastal ocean. *Nature* **451**:708-11.
63. **Luna GM.** 2015. Diversity of marine microbes in a changing Mediterranean Sea. *Rend. Lincei* **26**: 49-58.
64. **Salter I, Galand PE, Fagervold SK, Lebaron P, Obernosterer I, Oliver MJ, Suzuki MT, Tricoire C.** 2015. Seasonal dynamics of active SAR11 ecotypes in the oligotrophic Northwest Mediterranean Sea. *ISME J.* **9**:347-360.
65. **Grote J, Thrash JC, Huggett MJ, Landry ZC, Carini P, Giovannoni SJ, Rappé MS.** 2012. Streamlining and core genome conservation among highly divergent members of the SAR11 clade. *MBio* **3**:e00252-12.
66. **Dupont CL, Rusch DB, Yooseph S, Lombardo MJ, Richter RA, Valas R, Novotny M, Yee-Greenbaum J, Selengut JD, Haft DH, Halpern AL.** 2012. Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage. *ISME J.* **6**:1186-99.
67. **Giovannoni SJ, Bibbs L, Cho JC, Stapels MD, Desiderio R, Vergin KL, Rappé MS, Laney S, Wilhelm LJ, Tripp HJ.** 2005. Proteorhodopsin in the ubiquitous marine bacterium SAR11. *Nature* **438**:82-85.
68. **Gómez-Pereira PR, Hartmann M, Grob C, Tarran GA, Martin AP, Fuchs BM, Scanlan DJ, Zubkov MV.** 2013. Comparable light stimulation of organic nutrient uptake by SAR11 and *Prochlorococcus* in the North Atlantic subtropical gyre. *ISME J.* **7**:603-614.
69. **Brinkhoff T, Giebel HA, Simon M.** 2008. Diversity, ecology, and genomics of the *Roseobacter* clade: a short overview. *Arch. Microbiol.* **189**: 531-539.
70. **Taylor JD, Cottingham SD, Billinge J, Cunliffe M.** 2014. Seasonal microbial community dynamics correlate with phytoplankton-derived polysaccharides in surface coastal waters. *ISME J.* **8**: 245-248.

71. **Ribera d'Alcalà M, Conversano F, Corato F, Licandro P, Mangoni O, Marino D, Mazzocchi M, Modigh M, Montresor M, Nardella M.** 2004. Seasonal patterns in plankton communities in a pluriannual time series at a coastal Mediterranean site (Gulf of Naples): an attempt to discern recurrences and trends. *Sci. Mar.* **68**: 65-83.
72. **Liu M, Dong Y, Zhao Y, Zhang G, Zhang W, Xiao T.** 2011. Structures of bacterial communities on the surface of *Ulva prolifera* and in seawaters in an *Ulva* blooming region in Jiaozhou Bay, China. *World J. Microb. Biot.* **27**:1703-1712.
73. **Cho, JC, Giovannoni SJ.** 2004. Cultivation and growth characteristics of a diverse group of oligotrophic marine *Gammaproteobacteria*. *Appl. Environ. Microbiol.* **70**: 432-440.
74. **Crespo BG, Pommier T, Fernández-Gómez B, and Pedrós-Alió C.** 2013. Taxonomic composition of the particle-attached and free-living bacterial assemblages in the Northwest Mediterranean Sea analyzed by pyrosequencing of the 16S rRNA. *Microbiologyopen* **2**:541-552.
75. **Li J, Li N, Li F, Zou T, Yu S, Wang Y, Qin S, Wang G.** 2014. Spatial diversity of bacterioplankton communities in surface water of northern South China Sea. *PloS One* **9**:e113014.
76. **Zwirgmaier K, Jardillier L, Ostrowski M, Mazard S, Garczarek L, Vaultot D, Not F, Massana R, Ulloa O, Scanlan DJ.** 2008. Global phylogeography of marine *Synechococcus* and *Prochlorococcus* reveals a distinct partitioning of lineages among oceanic biomes. *Environ. Microbiol.* **10**:147–161.
77. **Scanlan DJ, Ostrowski M, Mazard S, Dufresne A, Garczarek L, Hess WR, Post AF, Hagemann M, Paulsen I, Partensky F.** 2009. Ecological genomics of marine picocyanobacteria. *Microbiol. Mol. Biol. Rev.* **73**:249-99.
78. **Kent AG, Dupont CL, Yooseph S, Martiny AC.** 2016. Global biogeography of *Prochlorococcus* genome diversity in the surface ocean. *ISME J.* doi:10.1038/ismej.2015.265.

79. **Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR, Arrieta JM, Herndl GJ.** 2006. Microbial diversity in the deep sea and the underexplored “rare biosphere”. *Proc. Natl. Acad. Sci. U.S.A.* **103**:12115-12120.
80. **Galand PE, Casamayor EO, Kirchman DL, Lovejoy C.** 2009. Ecology of the rare microbial biosphere of the Arctic Ocean. *Proc. Natl. Acad. Sci. U.S.A.* **106**:22427-22432.
81. **Vandamme P, Vancanneyt M, Pot B, Mels L, Hoste B, Dewettinck D, Vlaes L, Van Den Borre C, Higgins R, Hommez J.** 1992. Polyphasic taxonomic study of the emended genus *Arcobacter* with *Arcobacter butzleri* *comb. nov.* and *Arcobacter skirrowii* *sp. nov.*, an aerotolerant bacterium isolated from veterinary specimens. *Int. J. Syst. Bacteriol.* **42**:344-356.
82. **Campbell AM, Fleisher J, Sinigalliano C, White JR, Lopez JV.** 2015. Dynamics of marine bacterial community diversity of the coastal waters of the reefs, inlets, and wastewater outfalls of southeast Florida. *Microbiologyopen*: [dx.doi.org/10.1002/mbo3.245](https://doi.org/10.1002/mbo3.245).
83. **Luna GM, Quero GM, Perini L.** 2016. Next generation sequencing reveals distinct fecal pollution signatures in aquatic sediments across gradients of anthropogenic influence. *Adv. in Oceanogr. Limnol.* **25**:157.
84. **Loy A, Schulz C, Lückner S, Schöpfer-Wendels A, Stoecker K, Baranyi C, Lehner A, Wagner M.** 2005. 16S rRNA gene-based oligonucleotide microarray for environmental monitoring of the betaproteobacterial order “*Rhodocyclales*”. *Appl. Environ. Microbiol.* **71**:1373-1386.
85. **Liu J, Fu B, Yang H, Zhao M, He B, Zhang XH.** 2015. Phylogenetic shifts of bacterioplankton community composition along the Pearl Estuary: the potential impact of hypoxia and nutrients. *Front. Microb.* **6**:10.3389/fmicb.2015.00064.
86. **Aanderud ZT, Vert JC, Lennon JT, Magnusson TW, Breakwell DP, Harker AR.** 2016. Bacterial dormancy is more prevalent in freshwater than hypersaline lakes. *Front. Microbiol.* **7**: [doi: 10.3389/fmicb.2016.00853](https://doi.org/10.3389/fmicb.2016.00853).

87. **Sukenik A, Kaplan-Levy RN, Welch JM, Post AF.** Massive multiplication of genome and ribosomes in dormant cells (akinetes) of *Aphanizomenon ovalisporum* (Cyanobacteria). 2012. ISME J. **6**:670-679.
88. **Baas-Becking, L. GM.** 1934. Geobiologie of Inleiding Tot de Milieukunde. In *Diligentia Wetensch*, serie 18/19, edited by Stokkun, V., and Zoon, W. P. The Hague, The Netherlands.
89. **Vergin KL, Jhirad N, Dodge J, Carlson CA, Giovannoni SJ.** 2017. Marine bacterioplankton consortia follow deterministic, non-neutral community assembly rules. *Aquat. Microb. Ecol.* **79**:165-75.
90. **Curry R, Dickson B, Yashayaev I.** 2003. A change in the freshwater balance of the Atlantic Ocean over the past four decades. *Nature* **426**:826-9.
91. **Lynch MD, Neufeld JD.** 2015. Ecology and exploration of the rare biosphere. *Nature Rev. Microbiol.* **13**:217-29.

689 Table 1. Forward and reverse primers used to amplify the V4-V5 hypervariable region of the 16S rRNA gene
690 (from Parada et al. 2016).

515F (5'-GTGCCAGCMGCCGCGGTAA-3')

906R (5'-CCGYCAATTYMTTTRAGTTT-3')

691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711

712 Table 2. Bacterial abundance, geographic coordinates and environmental parameters (temperature, salinity, chlorophyll *a* (Chl *a*), nitrate (NO₃),
713 nitrite (NO₂), ammonium (NH₄), phosphate (PO₄) and silicate (SiO₄) concentration) at each sampling station in the GON (see Fig. 1).

Station ID	Coordinates	Temp (° C)	Salinity (PSU)	Chl <i>a</i> (µg l ⁻¹)	NO ₃ (µM)	NO ₂ (µM)	NH ₄ (µM)	PO ₄ (µM)	SiO ₄ (µM)	<i>Prochlorococcus</i> (cells ml ⁻¹)	<i>Synechococcus</i> (cells ml ⁻¹)	Heterotrophic bacteria (cells ml ⁻¹)
SA1	40°43'58.08"N 14°27'48.96"E	15.87	36.99	1.15	33.67	1.30	8.22	0.82	46.50	0	1.33 x 10 ⁴	2.50 x 10 ⁶
SA2	40°43'19.92" N 14°27'52.92"E	16.00	37.08	3.45	2.53	0.25	4.33	0.17	5.22	0	6.16 x 10 ³	1.14 x 10 ⁶
SA3	40°43'28.92"N 14°27'41.04"E	15.99	36.98	3.11	4.23	0.31	3.86	0.30	8.00	0	8.9 x 10 ³	5.65 x 10 ⁵
SA4	40°44'8.88"N 14°27'21.96"E	15.95	37.36	2.64	14.51	0.59	5.14	0.24	21.90	1.21 x 10 ³	1.22 x 10 ⁴	1.33 x 10 ⁶
SA5	40°43'36.84"N 14°27'12.96"E	16.13	37.06	3.10	11.34	0.49	2.55	0.17	17.10	2.52 x 10 ³	1.46 x 10 ⁴	1.64 x 10 ⁶
SA6	40°42'51.12"N 14°28'12.72"E	15.75	37.15	2.49	4.30	0.20	1.36	0.12	5.65	0	1.56 x 10 ⁴	3.74 x 10 ⁴
SA7	40°44'5.64"N 14°26'39.48"E	16.20	36.88	2.59	5.52	0.27	1.85	0.15	8.67	0	1.42 x 10 ⁴	1.80 x 10 ⁶
SA8	40°42'33.84"N 14°27'34.92"E	16.36	36.92	2.15	1.23	0.12	0.67	0.07	2.11	0	1.3 x 10 ⁴	1.60 x 10 ⁶
VE3	40°40'59.88" N 14°24'28.80"E	16.13	37.4	0.62	2.12	0.12	0.80	0.07	1.93	1.80 x 10 ³	1.71 x 10 ⁴	1.44 x 10 ⁶
TA1	40°45'5.40" N 14°23'54.60"E	16.35	36.55	2.12	0.05	0.06	0.45	0.05	0.34	0	1.32 x 10 ⁴	9.95 x 10 ⁵
TA2	40°44'39.48"N 14°22'33.60"E	16.56	36.87	1.80	0.00	0.03	0.64	0.01	0.14	1.32 x 10 ³	1.6 x 10 ⁴	1.44 x 10 ⁶
PO1	40°48'18.00"N 14°19'12"E	16.23	36.61	3.09	2.43	0.17	0.87	0.06	2.99	0	7.6 x 10 ³	9.77 x 10 ⁵
TG2	40°46'28.20"N 14°20'12.12"E	16.29	37.08	2.82	3.99	0.22	0.95	0.08	5.07	0	8.38 x 10 ³	8.16 x 10 ⁵
TG1	40°46'44.40"N 14°21'4.32"E	16.21	36.72	3.60	0.15	0.05	0.63	0.04	0.18	0	2.05 x 10 ³	8.17 x 10 ⁵
PO3	40°47'27.60"N 14°16'33.60"E	15.96	37.52	0.93	0.04	0.05	0.58	0.02	0.30	0	2.31 x 10 ³	7.65 x 10 ⁵
MC	40°48'42.84"N 14°15'00"E	16.26	36.13	5.17	5.50	0.47	4.72	0.25	6.12	2.24 x 10 ³	1.4 x 10 ⁴	5.47 x 10 ⁵

Table 3. Average (\pm SD) Good's coverage and diversity indices at GON stations. ACE* is the richness estimation of the non-subsampled 16S rDNA libraries. Different number of asterisks indicates significant differences between groups of stations (as in Fig. 2a) as obtained using ANOSIM R statistic (Fig. S2). Note: Standard deviations are generated by Mothur after multiple resampling of the OTU abundance values in order to normalize the data in order to have equal number of sequences in each library.

Station ID	Good's coverage (%)	ACE*	ACE	Chao	Berger-Parker	Shannon	Invsimpson
PO3*	99.7 \pm 0.02	399	354 \pm 42	331 \pm 27	0.24	3.25 \pm 0.008	11.98 \pm 0.11
VE3*	99.7 \pm 0.02	459	350 \pm 39	347 \pm 29	0.31	3.05 \pm 0.010	7.97 \pm 0.08
TG1**	99.6 \pm 0.02	542	455 \pm 69	373 \pm 36	0.14	3.23 \pm 0.007	14.64 \pm 0.10
PO1**	99.6 \pm 0.02	507	479 \pm 68	392 \pm 40	0.13	3.17 \pm 0.007	14.43 \pm 0.10
TG2**	99.6 \pm 0.02	551	461 \pm 56	401 \pm 31	0.13	3.23 \pm 0.005	15.00 \pm 0.08
TA1**	99.8 \pm 0.02	614	333 \pm 54	277 \pm 36	0.13 \pm	2.66 \pm 0.008	9.15 \pm 0.05
TA2**	99.8 \pm 0.02	730	340 \pm 60	279 \pm 33	0.14	2.78 \pm 0.008	10.01 \pm 0.06
SA5**	99.5 \pm 0.03	714	594 \pm 74	482 \pm 48	0.16	3.26 \pm 0.008	15.12 \pm 0.11
SA8**	99.6 \pm 0.03	523	520 \pm 55	428 \pm 37	0.13	3.36 \pm 0.007	16.66 \pm 0.11
SA7***	99.6 \pm 0.03	724	609 \pm 80	438 \pm 49	0.32	2.73 \pm 0.008	7.29 \pm 0.06
SA1***	99.5 \pm 0.03	753	727 \pm 108	498 \pm 62	0.25	2.85 \pm 0.009	8.93 \pm 0.07
SA2***	99.3 \pm 0.04	935	962 \pm 111	688 \pm 64	0.26	2.96 \pm 0.009	8.97 \pm 0.08
SA3***	99.4 \pm 0.02	782	788 \pm 74	561 \pm 42	0.21	3.18 \pm 0.006	11.95 \pm 0.07
SA4***	99.5 \pm 0.03	829	706 \pm 74	542 \pm 50	0.16	3.26 \pm 0.007	13.85 \pm 0.10
SA6***	99.7 \pm 0.02	378	343 \pm 42	293 \pm 25	0.19	3.00 \pm 0.005	11.18 \pm 0.06
MC	99.6	418	418	400	0.22	3.17	11.74

719 Figure Legends

720 Fig. 1. Map of the Gulf of Naples highlighting the location of the stations sampled. Samples for flow cytometry
721 and 16S rDNA sequencing were collected at all stations, whereas samples for 16S rRNA sequencing were
722 obtained only from stations PO1, TA1 and VE3 (indicated with red symbols). Red square indicates Sarno
723 river-influenced stations.

724 Fig. 2. Station clustering based on a) Bray Curtis dissimilarity and b) the distance matrix generated by calculating
725 pairwise UniFrac metrics. The scale bar in b) shows the distance between clusters in UniFrac units; if two
726 or more environments have similar lineages, they have a distance of 0. The significance of the cluster
727 nodes was determined using the jackknife analysis. Jackknife significance values are: >99.9%=1, 90-
728 99%=2, 70-90%=3; 50-70%=4, <50%=5 and higher values indicate a higher adaptation of communities to
729 the existing environmental conditions.

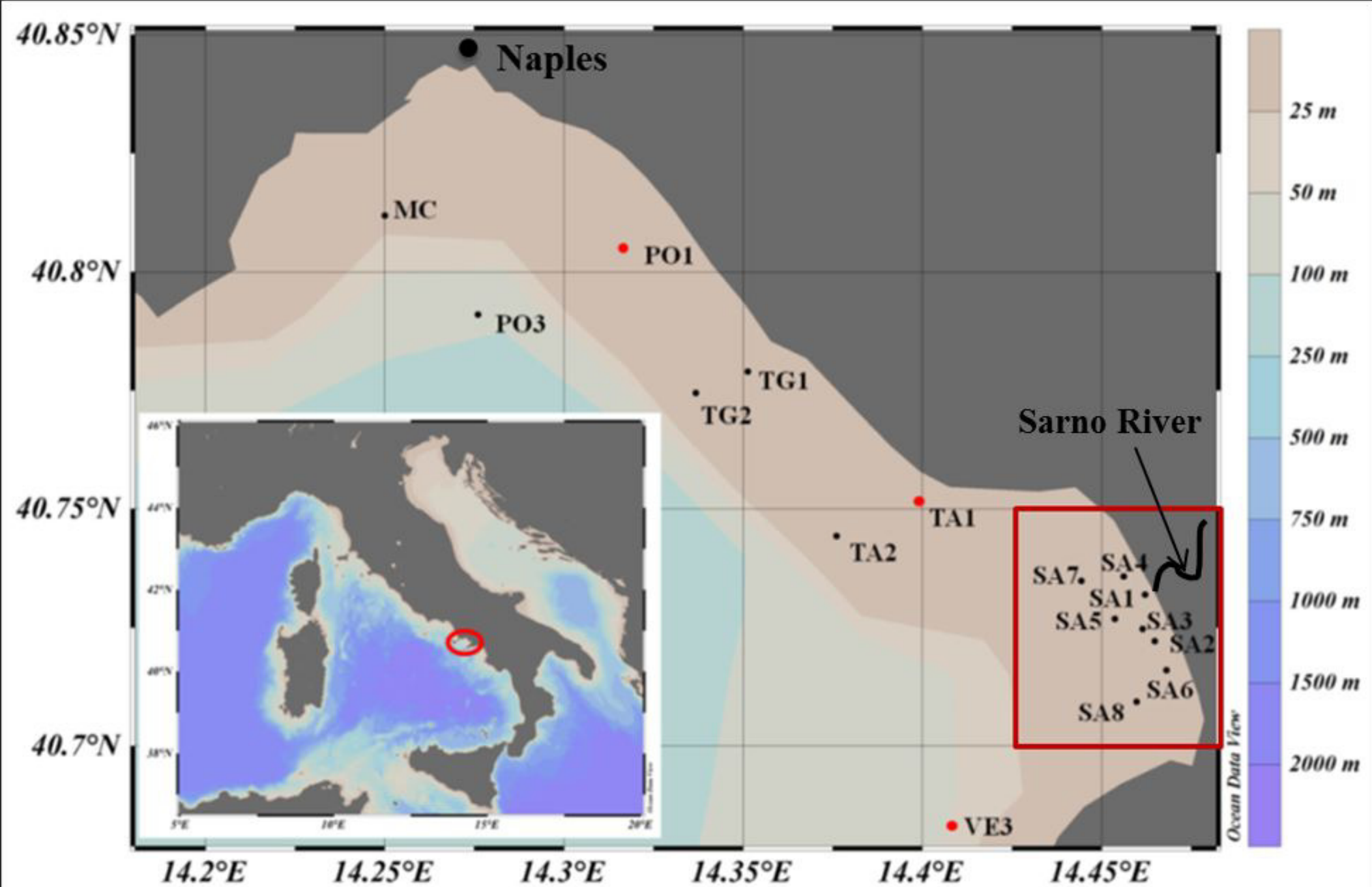
730 Fig. 3. a) Bacterial community composition at the phylum level at each GON station sampled. The phylum
731 *Proteobacteria* is split into several classes (*Alpha*-, *Beta*-, *Delta*-, *Gamma*-, and *Epsilon*-). Bacterial OTUs
732 that could not be classified were labelled as “other bacteria” b) The relative abundance of the top 12
733 abundant phylotypes in the GON at a lower taxonomic level (Order/Family level).

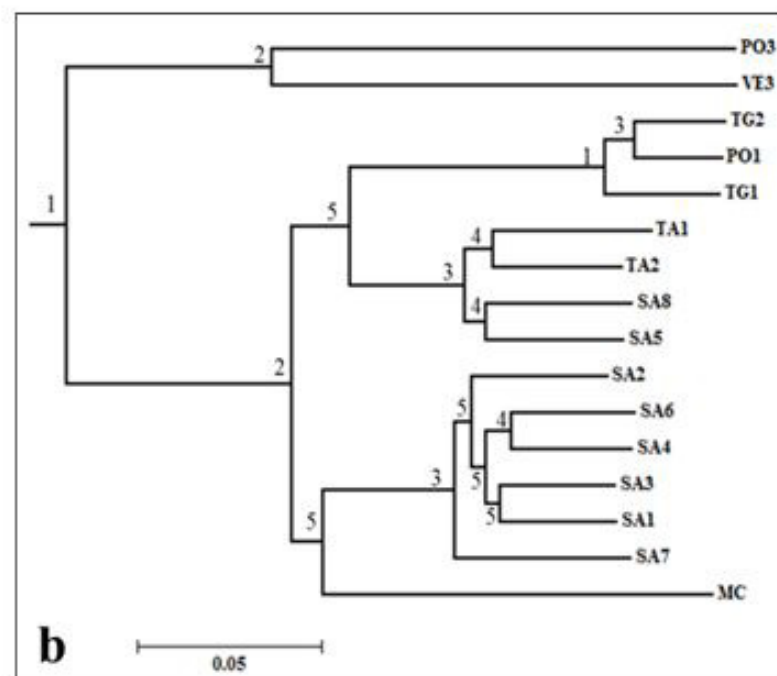
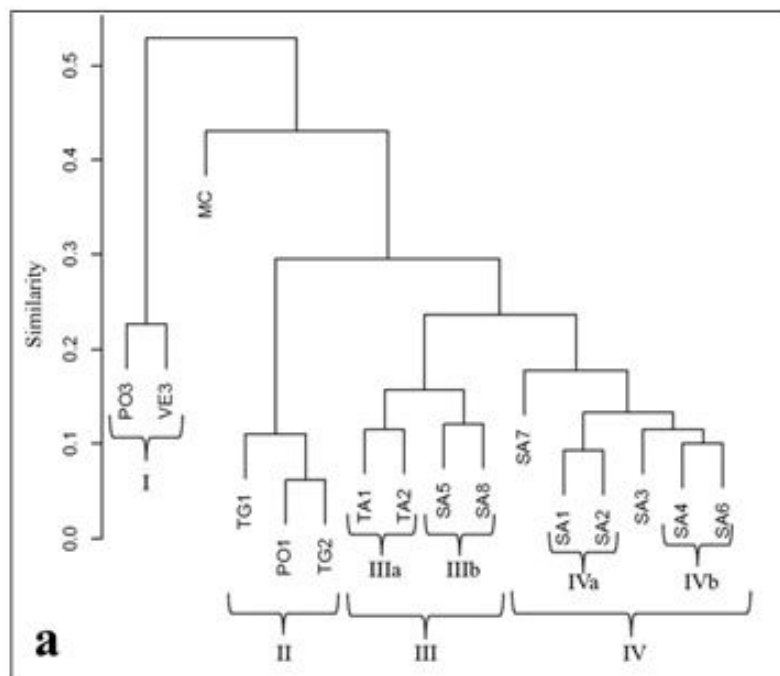
734 Fig. 4. CCA ordination plot depicting the relationship between environmental parameters and bacterial
735 community structure as represented by 16S rRNA gene sequence data. The first two axes explained 78%
736 and 18% of the cumulative variance, respectively.

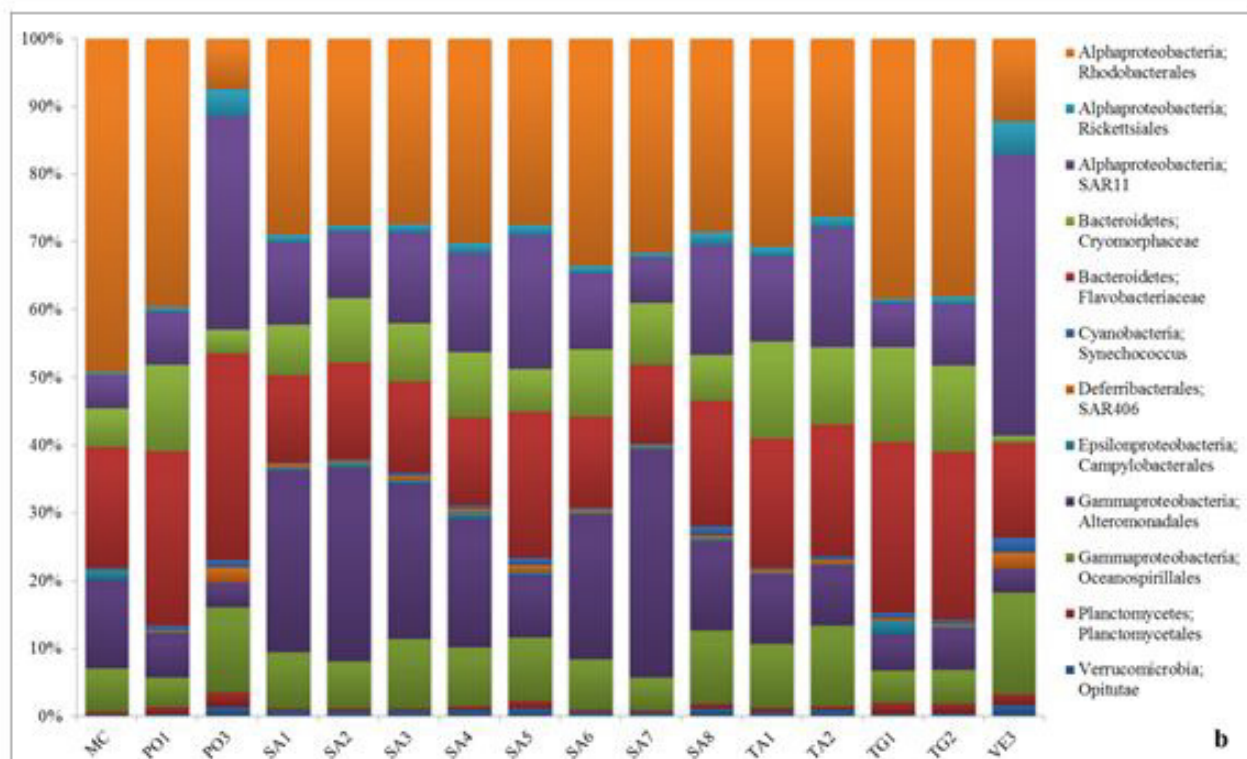
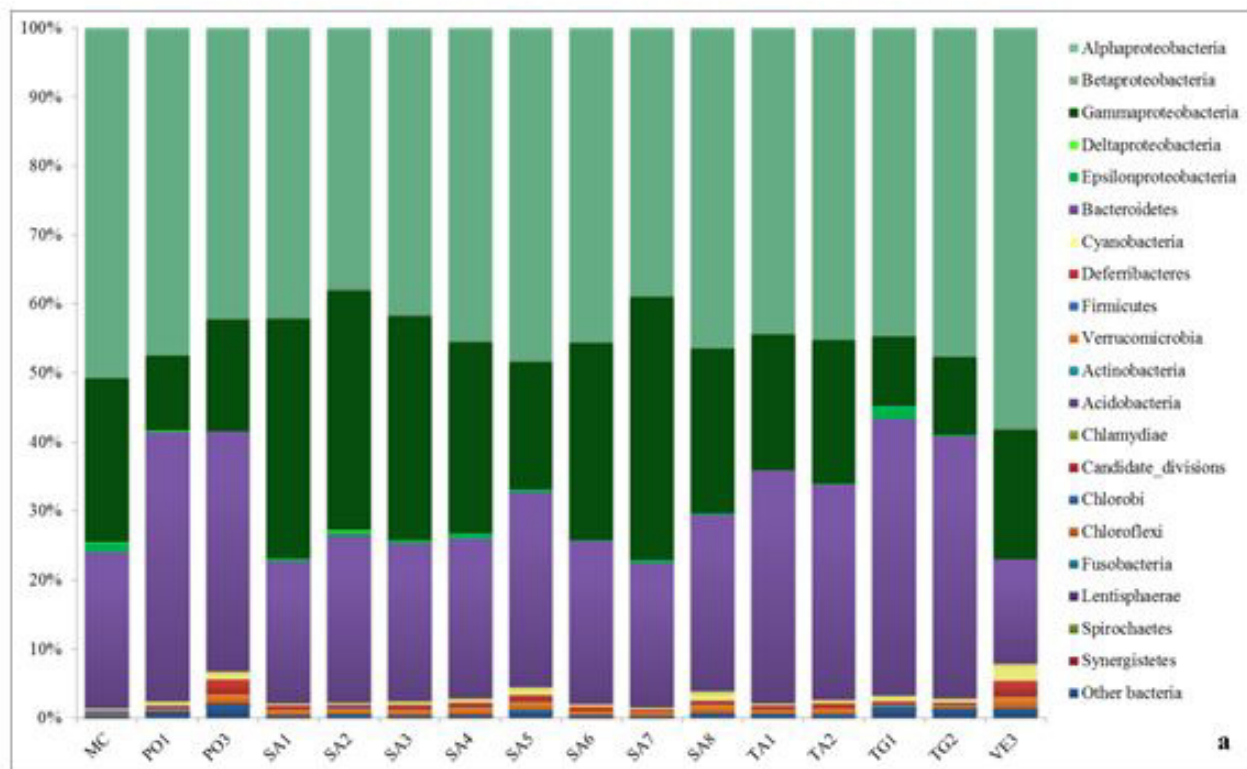
737 Fig. 5. LefSe analysis, indicating significantly, differential distribution of taxa in the different groups of stations
738 a) as identified by Bray-Curtis dendrogram (Fig. 2a) followed by ANOSIM R statistic, and b) after
739 grouping coastal+river-influenced stations. Red asterisks indicate “abundant” OTUs that were present in
740 all the libraries, while blue asterisks indicate OTUs that shifted between “abundant” and “rare” among
741 stations. OTUs without asterisk are “rare” (<0.1% of total sequences of the libraries).

Fig. 6. Correlation between 16S rRNA and 16S rDNA frequencies at stations VE3, PO1 and TA1 with slopes for each station (slope for the total data is $y=0.6191x + 0.7343$ $r^2= 0.3845$). Individual data points correspond to paired $\log(rRNA+1)$ and $\log(rDNA+1)$ for each individual OTU. Data were log transformed in order to eliminate bias. Therefore, correlations are limited to OTUs where both rRNA and rDNA sequences were present.

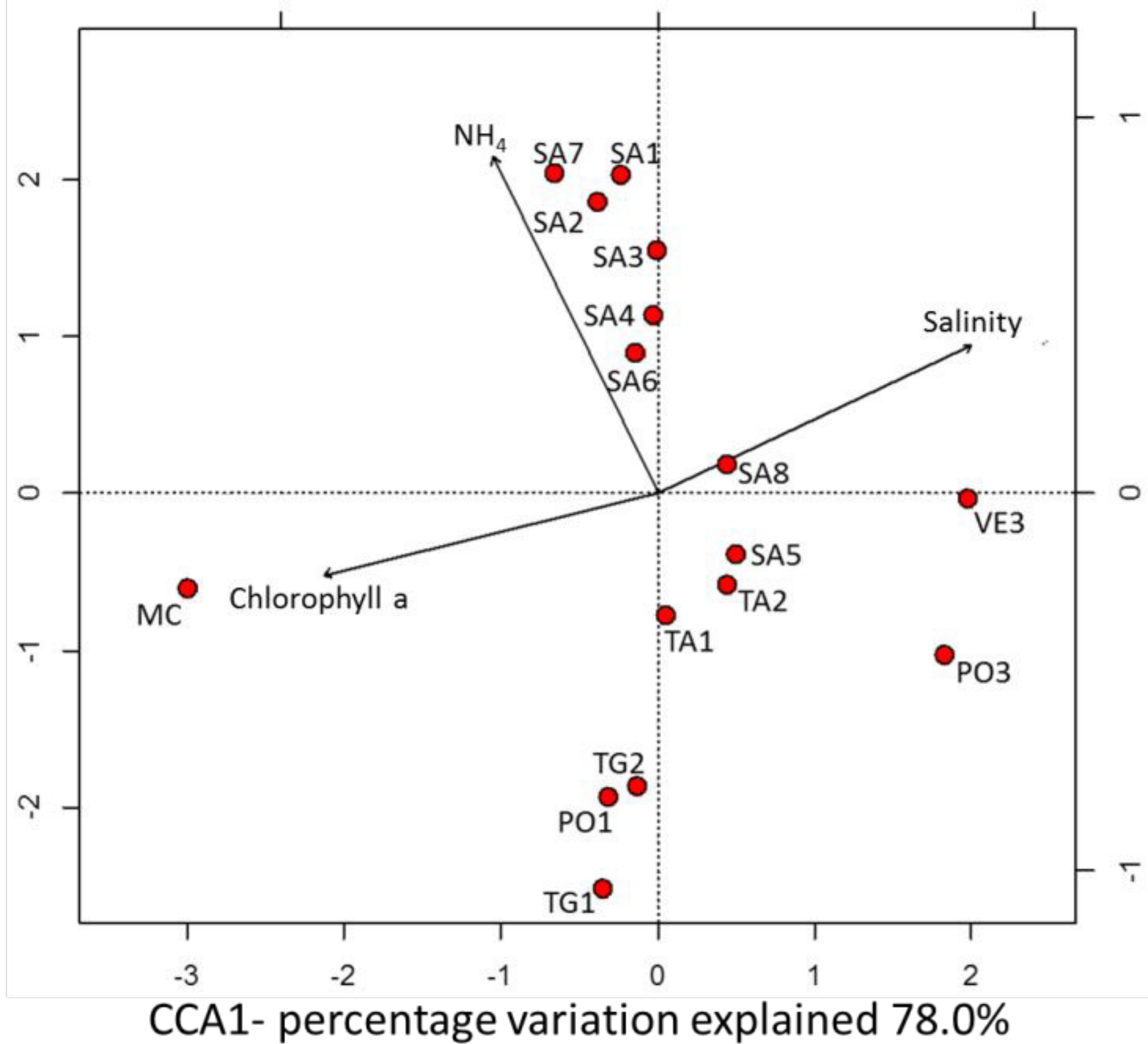
Fig. 7. Comparison of the relative abundance and potential activity of different phylogenetic groups within and between different sites in the GON. A) maximum-likelihood phylogenetic tree of the 50 most abundant OTUs present in all the equally subsampled 16S rRNA and 16S rDNA libraries; *Methanococcoides burtonii*, an archaeal species was included as an out-group. B) Heatmap showing the log abundance ($\log(\text{observations}+1)$) for each 16S rRNA and 16S rDNA OTU present in the tree. C) Specific potential activity relative to abundance of each OTU present at the VE3, PO1 and TA1 stations calculated using the rRNA:rDNA ratio.





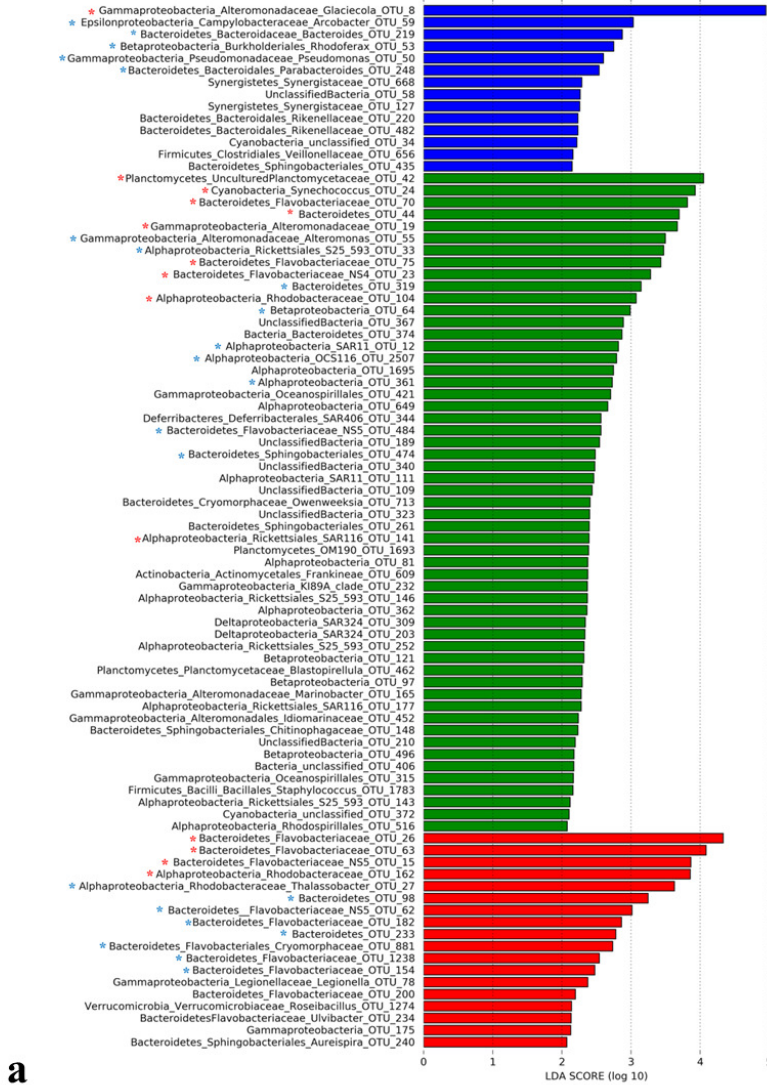


CCA2- percentage variation explained 18.0%



CCA1- percentage variation explained 78.0%

■ Coastal ■ Offshore ■ River-influenced



■ Coastal + River-influenced ■ Offshore

