# Identification of key bacteria for ecological dynamics in a coastal marine observatory

Erick Mateus-Barros[1]*, Emiliano Pereira[1], Juan Zanetti[1], Luciana Griffero[1], Rudolf Amann[2], Cecilia Alonso[1]

**Running Title:** Temporal dynamics in key marine microbiome

**Affiliation:** [1]Grupo de Ecología Microbiana de Sistemas Acuáticos, Centro Universitario Regional del Este, Universidad de la República. Uruguay; [2]Molecular Ecology Department, Max Planck Institute for Marine Microbiology, Bremen Germany

***Corresponding Author:** Erick Mateus-Barros, e-mail address: erickmbarros@gmail.com

## Abstract

Bacteria are essential for ecosystem maintenance, drive biogeochemical cycles, and influence responses to climate change. Additionally, many are involved in blooms and toxin production, affecting environmental quality. Despite their crucial role, some of their characteristics are only now being understood with the development of molecular techniques independent of cultivation. Thus, the recent effort to unify classical and microbial ecology studies has led to a better understanding of their dynamics. Here, we investigate the temporal processes that influence bacterial abundance and persistence, as well as their role in promoting or inhibiting the presence of certain organisms in the core of a metacommunity. Using molecular approaches such as amplicon and metagenomic analyses, we retrieved information on key bacteria (those with high abundance and/or persistence) from the South Atlantic Microbial Observatory (SAMO) to assess their interactions with environmental and biological factors. We identified a significant relationship between diversity changes, temperature, and nutrient availability, in a recurrent pattern that clusters winter-spring and summer-autumn communities, suggesting a connection with the seasonal changes in marine currents characteristic of this region. Among the key organisms, SAR11 and Flavobacteriaceae stand out. Flavobacteriaceae is known for its adaptive capacity and large genomes, while SAR11 has a reduced genome, making it dependent on compounds produced by other organisms. We identified an annual fluctuation in its abundance linked to Synechococcus blooms. The significant co-occurrence of ASVs from both groups reinforces the evidence of a biological interaction that sustains SAR11 through the exchange of these compounds. Through this study, we contribute to clarifying the factors that locally influence the geographic patterns of marine bacteria and identifying the pathways that promote their dynamics and functions in the ecosystem.

**Keywords:** marine bacterioplankton, temporal dynamics, amplicon, CAZimes, biological interactions

## Introduction

Bacteria are essential for diverse ecological functions and the structure of communities. Being the most abundant organisms in any ecosystem they are key for ecosystems maintenance (Chiriac, Haber et al. 2022), driving the major biogeochemical cycles (Myklestad 2000), and influencing the responses of ecosystems to climate change (Cavicchioli, Ripple et al. 2019). In addition, many of them are related to phenomena such as blooms and toxin production (Huisman and Hulot 2005), which affect environmental quality. Despite their crucial environmental roles, some of their characteristics could be fully understood only after the advent of culture-independent molecular techniques (Rappé and Giovannoni 2003) , and key aspects of their ecology remain still to be addressed. For example, recent studies on microbial biogeography have shown clearly identifiable patterns of dispersion and isolation for these organisms (Martiny, Bohannan et al. 2006, Lindh, Sjostedt et al. 2017, Soininen, Heino et al. 2018, Mateus-Barros, de Melo et al. 2021), which is contrary to what was initially supposed based on their small size and large populations.

In this context, there is a running effort to unify studies in classical and microbial ecology (Horner-Devine and Bohannan 2006) to boost the development of both disciplines (Acinas, Klepac-Ceraj et al. 2004, Barberán, Casamayor et al. 2014). On one hand, this interaction allows us to apply mathematical models and ecological theories to microbial ecology (e.g. Lindh, Sjostedt et al. 2017, Soininen, Heino et al. 2018); on the other hand, microbial datasets represent an ideal sample number to test the limits of ecological theories (Barberán, Casamayor et al. 2014).

The core-satellite hypothesis (Hanski 1982) is a conceptual framework of classical ecology that emerges as an attractive way to explore the macroecological structure of microorganisms (Lindh, Sjostedt et al. 2017, Jeong, Choi et al. 2020, Mateus-Barros, de Melo et al. 2021, Escalas, Paula et al. 2022). This hypothesis states that the frequency of occupancy in a metapopulation should be bimodal. This spatial pattern of species distribution indicates that, when the sampling effort is ideal, most organisms found should be present in one or few sites, while others will be distributed throughout the entire landscape (Gleason 1929). This distribution dichotomy evidences the existence of two distinct groups of organisms regarding occupancy: on the one hand, there are the satellites which are those restricted to one or few sites, while on the other, the most persistent organisms are called core (Hanski 1982) . This phenomenon seems to be mainly related to abundance (Papp and Izsák 1997) since these organisms are generally those that manage to (re)colonize new sites in a dispersal wave that gradually increases the area covered by a species (Brown 1984).

Utilizing the core-satellite framework it is possible to recover some interesting information about biodiversity since its pattern depends on the dispersion capacity inherent to each organism that should impact this structural feature. For example, some aquatic bacteria can be regionally persistent even when they are not abundant (Lindh, Sjostedt et al. 2017). Additionally, while the satellite encompasses the major

part of regional diversity (Mateus-Barros, de Melo et al. 2021), the core is composed by organisms that are potentially key to ecosystem functions (Fonte, Latombe et al. 2021). The core-satellite approach is also a valuable way to retrieve information concerning the metacommunity structure (i.e. a set of local communities linked by a flux of individuals). Also, this approach can recover details on topics such as colonization and extinction dynamics, adaptiveness and dispersal capacity, niche partition, and competition (Gaston, Blackburn et al. 2000, Mehranvar and Jackson 2001, Lindh, Sjostedt et al. 2017).

The bimodal distribution of species is probably directly related to the abundance variation through space and time (Brown 1984). Locally, environmental factors like pH and temperature (Lindström, Kamst-Van Agterveld et al. 2005, Sunagawa, Coelho et al. 2015), and large scale processes like shifts in marine currents domains or seasonal conditions (Messer, Ostrowski et al. 2020) can determine which groups will likely be more abundant and, consequently, have increased chances to disperse a highest number of individuals and dominate spatially (Gaston, Blackburn et al. 2000), having consequences at the ecosystem scale. For example, a long-term study on the dynamics of a phytoplanktonic community in a tropical lake showed that the change in environmental conditions changed dominance features, and allowed a non-seasonal change of ecological state, from filamentous cyanobacteria to green algae (Rugema, Darchambeau et al. 2019).

Biological interactions can also be a relevant feature to be considered. A microbial community is connected in an intricate interaction web (i.e. microbial loop Azam, Fenchel et al. 1983) by compounds incorporation made available by exudation, excretion, and cell lysis by other organisms (Kawasaki and Benner 2006, Alonso, Musat et al. 2012, Sarmento and Gasol 2012). The most abundant bacterial clades living in the marine and freshwater surface layers are specialized to optimize their genomes. They maintain genes necessary to key functions (Escalas, Paula et al. 2022) and to use the resources released by other bacteria and picophytoplankton, explaining the highly interdependent behavior of many microbial ecosystems (Giovannoni, Cameron Thrash et al. 2014, Chiriac, Haber et al. 2022).

The degradation of algal-derived Dissolved Organic Matter (DOM) is a crucial process of the marine Carbon cycle. Algal-derived labile DOM is rapidly consumed by the heterotrophic bacterioplankton leading to CO2 production or alternatively transformed into compounds that escape degradation for a long period, contributing to carbon sequestration (Jiao, Herndl et al. 2010). Multiyear systematic evaluations of spring blooms revealed a pattern of recurrent taxonomic successions within the bacterioplankton, paralleled by changes in the metagenomic repertoire of membrane transporters and carbohydrate-active enzymes (Teeling, Fuchs et al. 2016). Such evaluations rely on intensively sampled long-term time series which have been rarely performed outside the temperate regions and are particularly lacking in the Southern Hemisphere.

Here, our aim was to investigate the local temporal processes that may contribute to the abundance and persistence of bacteria, and their roles in guarantying or preventing certain organisms from being included in the core of a metacommunity. More specifically, we intend to (1) correlate the temporal variation on the bacterial

presence and abundance with environmental factors and the variation in others biological components, (2) identify potential roles of these more abundant organisms, by correlating with environmental variables, other biological components and genes related to biogeochemical cycles, in special for the Carbon, and (3) to understand the temporal factors in local microbial communities that impact their spatial features.

## Methods

*Study Sites and Sampling Design*

In this study we used a dataset collected monthly in the South Atlantic Microbial Observatory (SAMO) located in the marine zone of the “Laguna de Rocha” protected area (34° 42’ S, 54° 15’ W) to test the impact of a series of local environmental and biological variables on the persistence and abundance variation of bacteria aiming to understand what makes them have a core or satellite distribution along time.

The SAMO, launched in 2014 from the Ocean Sampling Day international initiative (Kopf, Bicak et al. 2015), is located at the Uruguay–Buenos Aires shelf ecoregion. This marine region is categorized by having a complex hydrodynamic system, as it is impacted by the Río de la Plata outflow and the confluence of Brazil and Malvinas currents (Burone, Franco-Fraguas et al. 2021). This creates an intricate balance of temperature, salinity and nutrients uptake that varies through the year depending of the water mass that is majoritarely influencing the region (Burone, Franco-Fraguas et al. 2021). While the warmer and salty Brazil current mainly impacts the region at the Winter-Spring period, the Malvinas current turns the ocean colder in the summer-Autumn period (Burone, Franco-Fraguas et al. 2021).

The SAMO is part of the Latin American Aquatic Microbial Observatory Network (AMOLat), which born as a scientific consortium for the creation and maintenance of long-term microbial observatories all around the continent, by the sharing of knowledge, experiences and opportunities (Fermani, Martyniuk et al. 2024).

*Sampling procedures, physicochemical and molecular analyses*

The sampling of sub-surface marine water has been performed monthly since 2018 to nowadays. In the field, a multi-parameter probe is used to recover environmental information (salinity, temperature, pH, turbidity, dissolved oxygen, suspended solids, and fluorescence). Further physicochemical characterization includes: i) the quantification of nutrients (total phosphorus, total nitrogen, phosphate, ammonium, nitrite, nitrate, reactive silica) and chlorophyll a, following standard methods (Apha 1999), and ii) the spectral characterization of the chromophoric and fluorescent dissolved organic matter (cDOM and fDOM) (Amaral, Graeber et al. 2016).

For the molecular analyses, departing from the microbial biomass recovered after water samples filtration, the eDNA is extracted following standard methods (Alonso et al 2010). The prokaryotic community diversity is determined by amplifying a variable region of the 16S rRNA gene with the primers 515F-Y (5’-GTG YCA GCM

GCC GCG GTA A- 3') and 916R (5'-CCG YCA ATT YM TTT RAG TTT-3') (Parada, Needham et al. 2015), while the small planktonic eukaryotic components are determined by amplifying a variable region of the 18S rRNA gene using the primers TAReukFWD1 (5'-CCAGCASCYGCGGTAATTCC-3') and TAReukREV3 (ACTTTCGTTCTTGATYRA) (Stoeck, Bass et al. 2010).

The 16S and 18S amplified material, as well as the extracted eDNA without prior amplification are sequenced using Illumina platforms, to obtain the prokaryotic, eukaryotic, and metagenomic sequences of the whole bacterioplankton community, respectively.

*Bioinformatics*

The amplicon sequence data was processed with the DADA2 pipeline (Callahan, McMurdie et al. 2016), integrating the following tasks: removing primers if present, trimming low-quality sequence regions, dereplicating, identifying amplicon sequence variants (ASVs), merging paired-end reads, and removing chimeras. Prior to the analyses, the ASVs tables were cleaned to remove sequences non-identified as bacteria and eukaryotes in the 16S and 18S datasets, respectively. After, the tables were rarefied by the sample with least sequences number, and the ASVs counting five or less sequences were removed.

To pre-process the metagenomic data, the methodology described below is applied: 1) adapter sequences are removed; 2) paired-end reads are merged; 3) merged and unmerged reads are quality trimmed and filtered. It is worth noting that this project will be based on a gene-centric approach, and thus, the construction of contigs and scaffolds will not be necessary. However, we will explore the possibility of generating open reading frame assemblies at the protein level (Steinegger, Mirdita et al. 2019). The preprocessed metagenomic data is analyzed utilizing a custom-developed pipeline (https://github.com/pereiramemo/Mg-Traits (Pereira-Flores, Barberan et al. 2021), which performs the computation of several functional aggregated traits at the metagenome level, ranging from GC variance to average genome size, and includes the functional annotation based on the Pfam database (Mistry, Chuguransky et al. 2021) and the carbohydrates-active enzymes database (Drula, Garron et al. 2022), among others. The results of the metagenomic sequence data analysis are represented as Functional Abundance Matrices (FAMs). In addition, a matrix summarizing the metagenomic trait values computed for each metagenomic sample is generated. All this data is used as an input for the analyses described below.

*Statistical analyses*

As stated before, the occupancy capacity of an organism seems to be related to their abundance features in each location (Papp and Izsák 1997), as those more abundant tends to more easily (re)colonize other sites in a landscape (Brown 1984). That being said, bacteria are some of organisms that have greater dispersal capacity in nature, which means that some groups can reach longer distances while maintaining

an lower mean relative abundance (Lindh, Sjostedt et al. 2017). In this sense, to better understand the bacterioplanktonic temporal dynamics observed in SAMO, we carried out statistical measurements that consider the entire community and sub-sampled the community into those ASVs considered as key, by observing two main characteristics: Their mean relative abundance and persistency. Each ASV' abundance was classified from the approach proposed by Pedrós-Alió (2012), which has defined as abundant those organisms who reaches at least 1% of local relative abundance. The persistence was defined as the capacity of being present in more than 90% of all samples, in an approach that is similar from those proposed by Hanski (1982) to classify the organisms which better occupy a landscape. Considering it, we could reach 3 main classes: those ASVs presenting an mean relative abundance greater than 1%, but persistence less than 90% samples were classified as "Abundant"; complementarily, the ASVs who have less than 1% mean relative abundances, but reached 90% of samples persistence were called "Persistent". The "Core" of bacterioplankton was those who have mean relative abundances greater than 1% and persistence greater than 90%. Finally, we also classified the ASVs that do not reached both Abundance and Persistence cutoff as "non-Key".

To determine if there is a bimodal distribution on the persistence of microorganisms, we applied the Mitchell-Olds & Shaw test of quadratic extremes (Mitchell-Olds and Shaw 1987). A significant bimodal distribution means that there is a group of ASVs always present in SAMO independently of the conditions, while other group is seemed in very specific situations (Gleason 1929, Hanski 1982).

To determine the relationship between the temporal variations of bacteria abundances and the environmental factors we applied the non-Metric Multidimensional Scaling (nMDS) and distance-based Redundancy Analysis (dbRDA). To understand how the community dissimilarity fluctuated through the years, we applied a Time-decay Relationship (TdR) analysis. To measure potential (in)direct biological relationships between the bacteria classified as key in this community and whit Eukaryotes, we used paired co-occurrences measurements. For this purpose, we used the 16S and 18S data, and the contigs found in the metagenomes that contains information for the CAZimes family synthesis. All statistical analyses described above were performed using the R software (R Core Team 2019) and various functions provided by the vegan (Oksanen, Blanchet et al. 2016) and Tidyverse (Wickham, Averick et al. 2019) packages. The co-occurrence analysis was performed with the co-occur package (Griffith, Veech et al. 2016).

## Results

After rarefaction and rarest ASVs cutoff, 2 586 ASVs were obtained from 326'540 reads (mean of 9'330 reads per sample), were we could identify a total of 30 phyla. These ASVs were classified were it was pertinent considering Order, Family or Genus into 31 groups of interest. The most abundant groups of interest encompassed the SAR11 (17.75% of total reads), followed by Flavobacteriaceae (14.77%), Rhodobacterales (8.79%), other Gammaproteobacteria (6.05%), Cryomorphaceae (5.85%), and Synechococcus (5.32%) (Fig. 1).

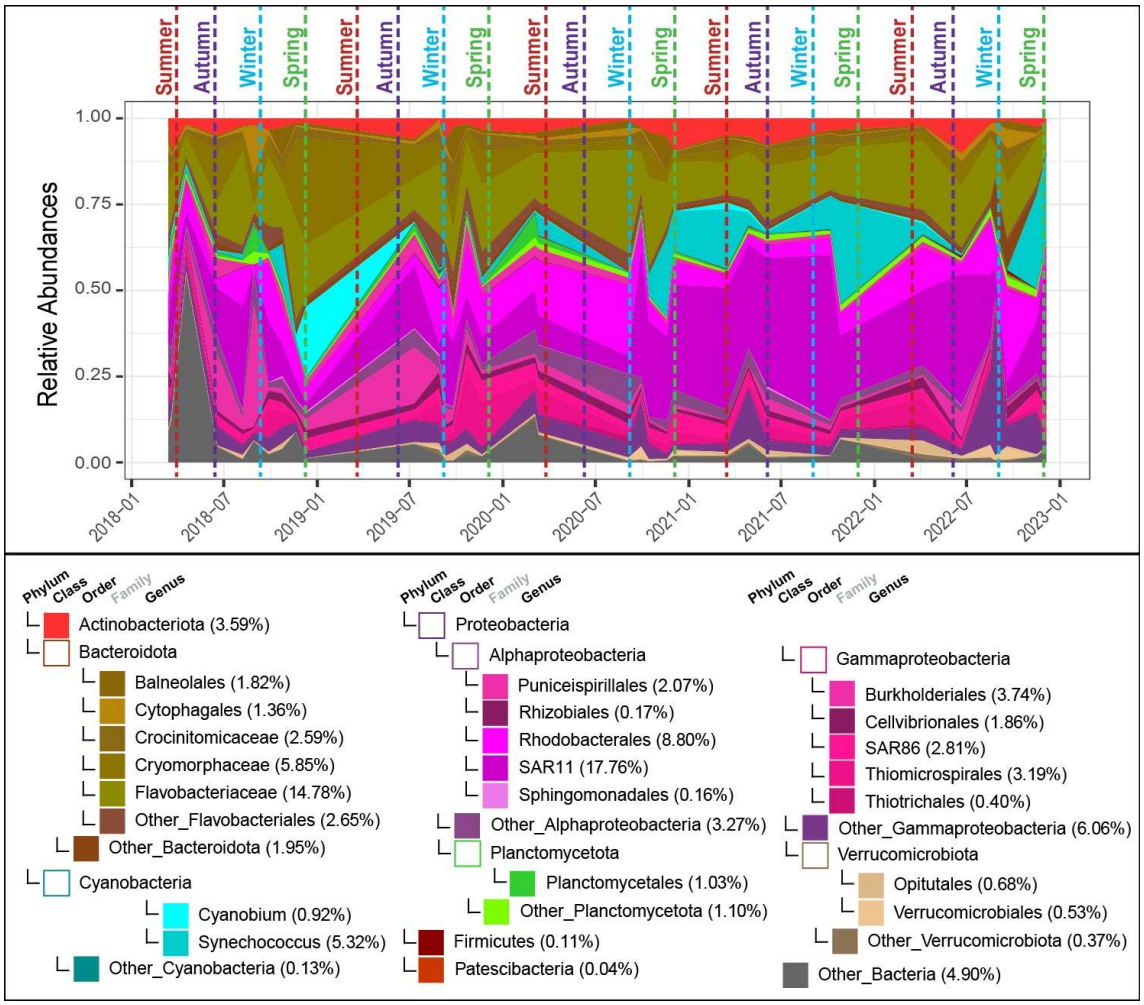


**Figure 1.** Temporal variation in the relative abundance of bacterioplankton. Each ASV was assigned to a phylogenetic phylum. Groups with known relevance to local dynamics were classified at the Class, Order, or Genus level, depending on the level of taxonomic resolution achieved for each phylum. Parenthetical values represent the total relative abundance of each group

Although the bimodality test did not show a significant result (Fig. S1), temporal variation was observed in the ASVs that make up the key groups —i.e., Abundant, Persistent, and Core— throughout the semesters. Although this variation does not present high dissimilarity values, it is enough to explain the absence of an observable bimodal pattern (Fig. S2). Additionally, we observed a significant correlation between community variation and certain environmental and biological components. The nMDS analysis (Fig. 2) shows that large-scale variations in the community appear to be related to seasonal patterns that divide the samples into two main groups by season: Summer-Autumn and Winter-Spring, in a recurrent arrangement that does not vary considerably with interannual changes, but is also reflected in the observed diversity variations between seasons (Fig. S3). We also found a trophic state effect on this observed temporal variation. The beta-dispersion analysis (Fig. S4) indicates that this variation is not significantly related to variation within the groups. By subdividing the community into key bacteria (Core, Abundant, and Persistent) and non-key bacteria, a

similar result can be recovered in the Abundant and non-Key groups, while in the Core and Persistent groups, a variation of lower intensity is observed (Fig. S5).

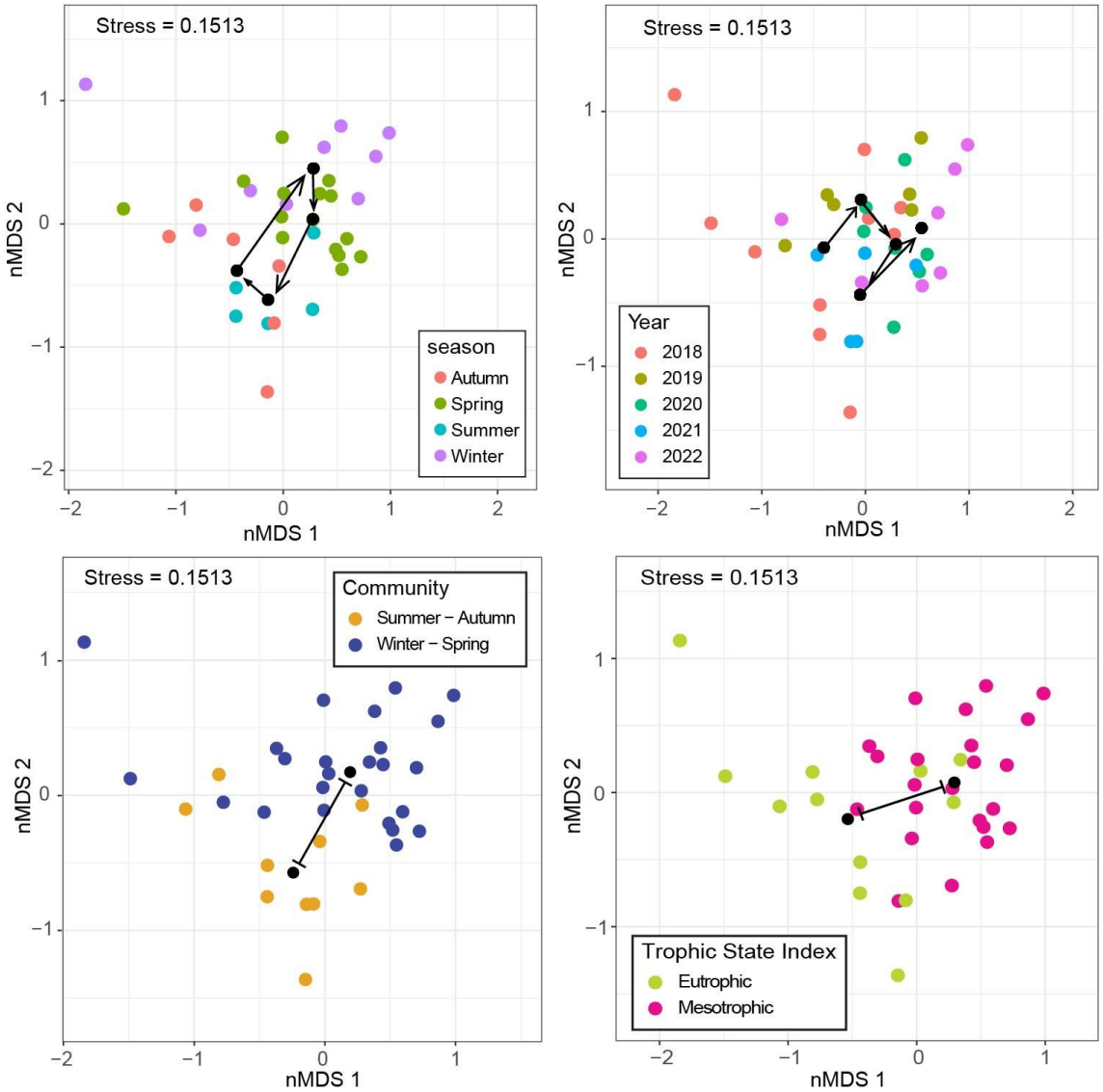


**Figure 2.** nMDS (non-metric Multi-Dimensional Scaling) plots depicting bacterioplankton community dissimilarity in response to large-scale environmental changes across seasons (top-left box), years (top-right box), the Brazil-Malvinas Confluence (bottom-left box), and trophic state (bottom-right box)

Complementarily, the dbRDA (Fig. 3) shows the relationship between this temporal community variation and more refined environmental factors fluctuations. This community seems to be mainly related to salinity, and by nutrients uptake. The type of nutrient varies, encompassing from nitrogen compounds (Total and Dissolved Nitrogen), going through trophic indexes related to the availability of Phosphorus (TSI – Phosphorus) e sintetização de carbono (TSI – Secchi and TSI – Chla), and reaching to protein derived compounds (Component 4). In addition, there being a negative relationship to Dissolved Oxygen availability. The second dimension indicates a less intense link between the microbial diversity and Temperature, and to some of the

above listed nutrient compounds (Total Nitrogen and TSI indexes) and Dissolved Oxygen.

Also, subsampling the key and non-key groups, we can observe that, besides salinity and nutritive compounds remained de main drivers of variation, each one turned to a slightly different direction (Fig. S6). While the Salinity and Nutrients remained the main drivers of non-Key diversity, pH joined Temperature as secondary drivers of their variation. The Abundant group are more strongly related to Salinity and Temperature in one direction while also strongly related to Nutrients into other direction. In turn, Persistent bacteria seems form three groups, one more closely related to Salinity, other to Dissolved Oxygen, and the last to Nutrients and Turbidity. Finally, the Core group exhibits a strong relationship with salinity, temperature, and nutrients, notably, components 3, 4, and 5 also stand out here, which were characterized as a protein fraction, indicating a potentially more labile biological origin

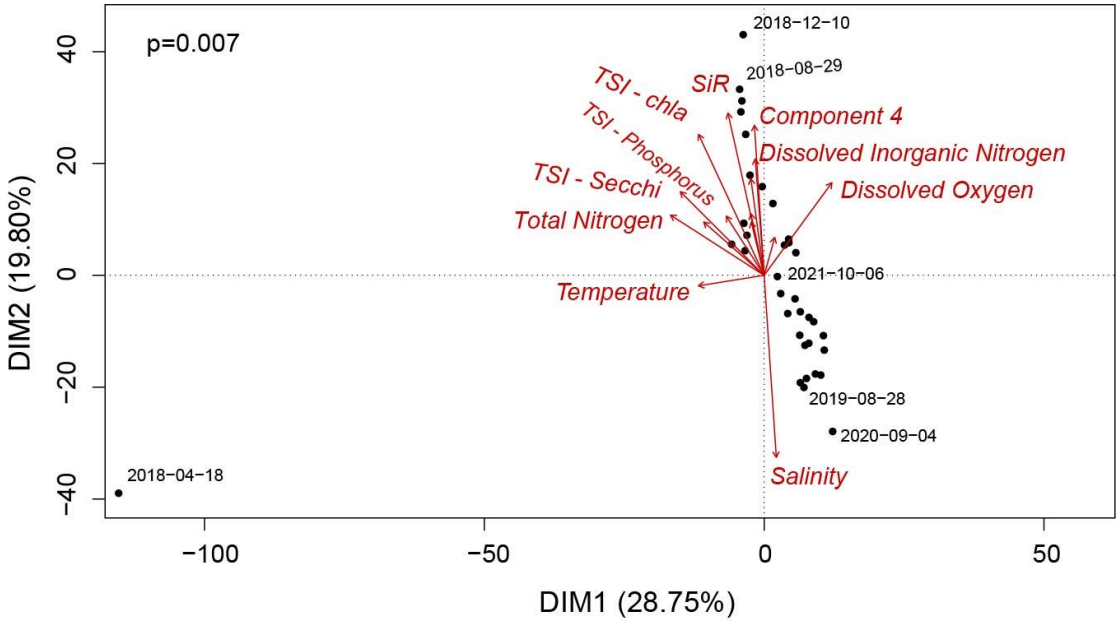


**Figure 3.** dbRDA (distance-based ReDundancy Analysis) plots displaying the maximum similarity between sites, considering the relationship between them and the environmental factors that best explain the observed bacterioplankton dissimilarity

The TdR shows the variation in dissimilarity between samples as the temporal distance between each compared sample increases (Fig. 4). This analysis demonstrates an expected gradual increase in observed dissimilarity, displaying a characteristic wave-like pattern with a one year periodicity. This suggests that samples collected in the same season exhibit a certain degree of similarity, likely associated with the large-scale inter-annual variation in environmental conditions, a pattern also observed in Figure 2. This trend, observed in the community as a whole, is more prominent in the subset of Abundant and non-Key groups, while the variation is less evident among Core bacteria and absent in the Persistent ones. These two groups also

exhibit the lowest rates of change when comparing samples collected at highest temporal distances.

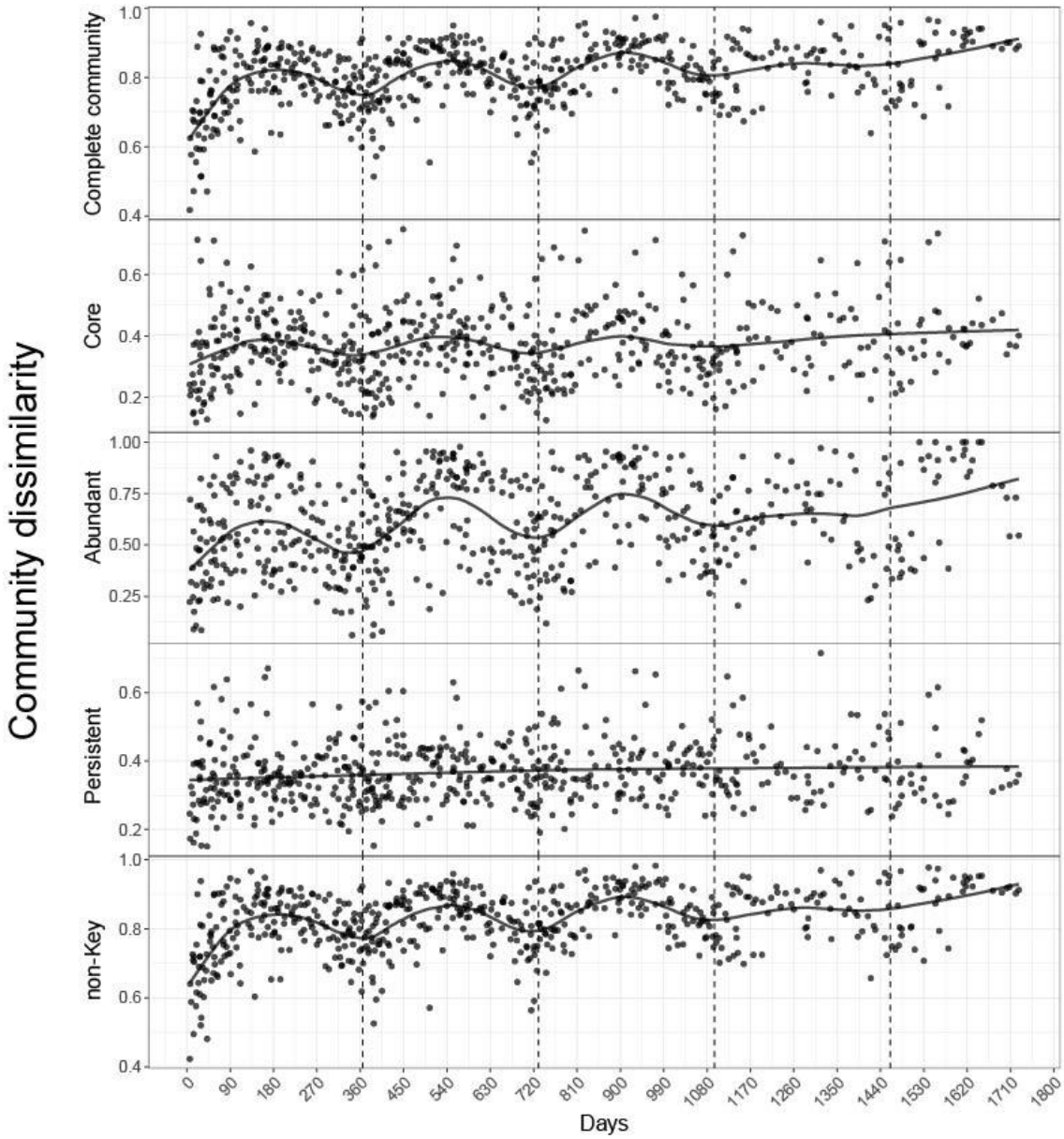


**Figure 4.** TdR (Time-decay Relationship) plots illustrating temporal variation in sample dissimilarity for the whole community (first box) and, in order, the key groups (Core, Abundant, and Persistent) and non-key groups

The community exhibited higher change rates associated with the transition from Winter-Spring toward Summer-Autumn (Fig. 5). This pattern is identical when considering bacteria classified as non-Key, and is similar for the Abundant bacteria. However, the pattern is reversed for the Core and Persistent groups. In this case, the highest change rates were recorded during the transition from Summer-Autumn to Winter-Spring.

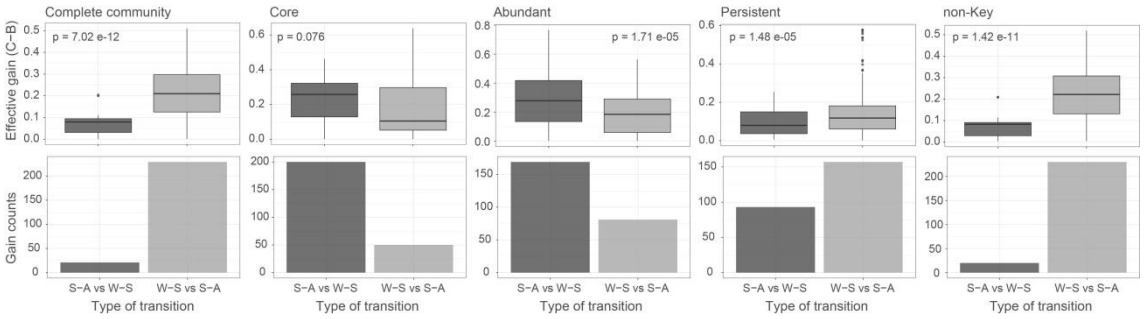


**Figure 5.** Gains in dissimilarity from the winter-spring transition to summer-autumn, considering, from left to right, the whole community, key groups (Core, Abundant, Persistent), and non-key groups

We also analyzed changes in picoeukaryote abundance to identify patterns that might explain how interactions with prokaryotic organisms might be affected. We observed fluctuations in the abundance of ASVs retrieved from 18S (Fig. S7) that do not seem to correlate with those observed in prokaryotes. Moreover, these organisms appear to follow a distinct persistence pattern that, on one hand, separates the spring samples from the others and, on the other, the summer ones (Fig. S8), which likely reflects periodic blooms of Metazoa and Dinoflagellata, respectively, but is not significantly related to environmental fluctuations (Fig. S3).

Finally, to understand the paired relationship between the bacteria and eukaryotic ASVS and CAZimes occurrence, we applied a co-occurrence analysis. The bacterial groups that exhibited the highest number of interactions with both 16S, 18S, and CAZymes were SAR11, Flavobacteriaceae, and Thiomicrospirales (Tab. 1; Tab. S1). SAR11 and Flavobacteriaceae also stand out as those representing the major number of nodes in this co-occurrence network (Fig. 6), which includes various representatives such as Thiomicrospirales, Synechococcus, Balneolales, among others. The interaction network also reveals an isolated network that encompassed just SAR11 and Synechococcus representatives.

**Table 1.** Top 5 bacteria among key groups (Core, Abundant, and Persistent) with the most significant co-occurrences with other bacteria, CAZymes and picoeukaryotes. For a complete list, see Table S1

| Class | Group 1 | Number of co-ocurrences | Group 2 |
|---|---|---|---|
| Prokaryota VS Prokaryota | SAR11 | 14 | SAR11 (6); Synechococcus (5)$^{+}$; Burkholderiales (1); Cellvibrionales (1); Other Alphaproteobacteria (1) |
| | Flavobacteriaceae | 12 | Flavobacteriaceae (2); SAR86 (2); Synechococcus (2)$^{+}$; Burkholderiales (1); Cryomorphaceae (1); Other Flavobacteriales (1); Thiomicrospirales (1)$^{+}$ |
| | Synechococcus | $2^{+7}$ | Balneolales (1); Cellvibrionales (1) |

| | | | |
|---|---|---|---|
| | Thiomicrospirales | $4^{+1}$ | Cryomorphaceae (1); Puniceispirillales (1); SAR86 (1); Other Flavobacteriales (1) |
| Prokaryota VS CAZimes | SAR11 | 67 | AA (3); CBM (2); CE (2); GH (30); GT (14);<br>PL (16) |
| | Flavobacteriaceae | 35 | AA (2); CBM (2); GH (31) |
| | Burkholderiales | 17 | GH (12); GT (3); PL (2) |
| | Other_Bacteria | 16 | GH (9); GT (1); PL (6) |
| | Thiomicrospirales | 13 | AA (1); GH (7); GT (3); PL (2) |
| Prokaryota VS Eukaryota | Thiomicrospirales | 204 | Protalveolata (61); Ochrophyta (38); Dinoflagellata (24); Ciliophora (22); Metazoa (22); Other_Eukaryota (19); Chlorophyta (8); Cryptophyceae (8); Cercozoa (2) |
| | Other_Bacteria | 142 | Protalveolata (54); Other Eukaryota (22); Ciliophora (20); Ochrophyta (16); Dinoflagellata (14); Metazoa 10; Chlorophyta (3); Cryptophyceae (3) |
| | SAR11 | 140 | Protalveolata (60); Other_Eukaryota (18); Ochrophyta (17); Ciliophora (16); Dinoflagellata (16); Metazoa (7); Chlorophyta (4); Cercozoa (2); |
| | Flavobacteriaceae | 128 | Protalveolata (42); Dinoflagellata (19); Other_Eukaryota (17); Ciliophora (16); Ochrophyta (16); Metazoa (9); Cryptophyceae (7); Chlorophyta (2) |
| | Burkholderiales | 100 | Protalveolata (47); Ciliophora (16); Metazoa (15); Ochrophyta (9); Dinoflagellata (7); Chlorophyta (3); Other Eukaryota (3) |

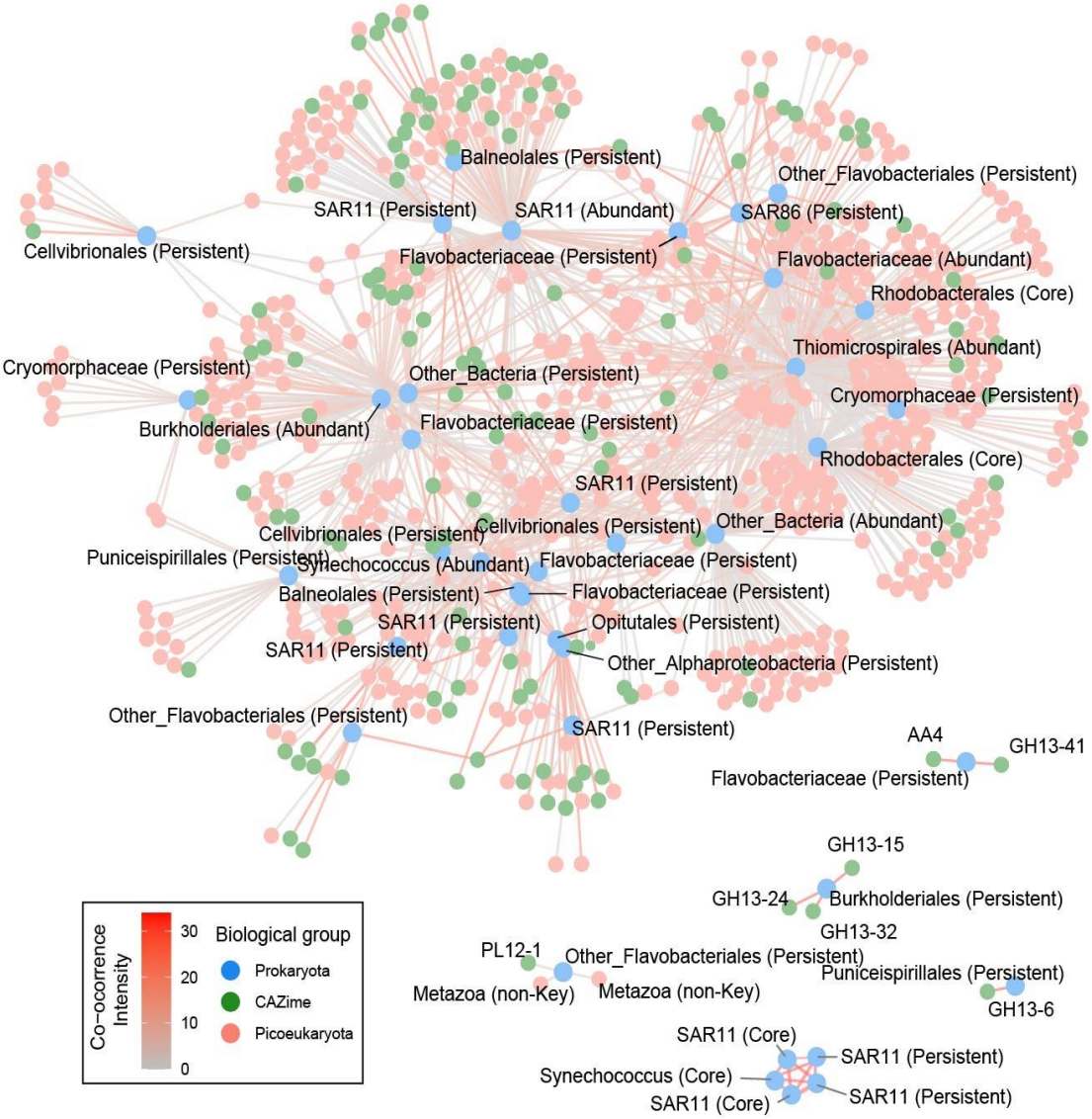


**Figure 6.** Co-occurrence network illustrating the relationship between bacteria from the key groups (Core, Abundant, and Persistent) among themselves and with CAZymes and picoeukaryotes. For a complete image including the correlations between CAZymes and picoeukaryotes, see Figure S9

## Discussion

Bacteria are pivotal organisms for understanding ecological processes occurring globally (Chiriac, Haber et al. 2022). While their relationships with environmental factors have been extensively identified (Lindström, Kamst-Van Agterveld et al. 2005, Sunagawa, Coelho et al. 2015), their associations with other factors, such as geography (Zorz, Willis et al. 2019, Mateus-Barros, de Melo et al. 2021) and potential biological interactions (Giovannoni 2017, Kim, Kang et al. 2019), have become increasingly clearer following the advent of culture-independent sequencing technologies (Rappé and Giovannoni 2003). These advancements have enabled the identification of previously little known metabolic process chains, suggesting the necessity of certain interactions that were not previously anticipated (e.g. Giovannoni 2017). Additionally, the application of well-established mathematical

models from classical ecology provides a robust theoretical framework for better comprehending these processes in microorganisms (Horner-Devine and Bohannan 2006). Here, we employ a straightforward ecological model—the core-satellite species hypothesis (Hanski 1982)—to identify patterns of persistence over a relatively extensive temporal interval and to pinpoint groups of potential interest for studies of marine bacterioplankton. We also aim to determine which environmental and biological factors potentially drive this community. Through this approach, we identified a clear and recurrent temporal variation that groups characteristic patterns into Winter-Spring and Summer-Autumn, indicating a relationship with the seasonal changes in marine currents, which is characteristic of this region (Burone, Franco-Fraguas et al. 2021). We also observed a possible biological interaction between ASVs identified as SAR11 and Synechococcus.

The theory presented by Hanski (1982) posits that, in an ideal sampling, it would be possible to observe a bimodal distribution in the occupancy frequency of any given community, with a large group of organisms being found in all sites. This pattern would be related to the observable abundances for each species within the community, as the most abundant organisms are also those capable of dispersing more efficiently across the landscape (Brown 1984), in a phenomenon called as rescue effect (Gotelli 1991). In this case, in this local study of large-scale temporal spam, we sought to understand whether this spatial pattern was also reflected in temporal processes, in a pattern that would reflect the ability of the more resilient bacteria to persist over time and, thus, be those most likely to eventually dominate the (meta)community (Leibold, Holyoak et al. 2004). However, this bimodal pattern cannot be observed in the persistence distribution over multiple years. A study conducted in the Baltic Sea (Lindh, Sjostedt et al. 2017) demonstrated the existence of a bimodal pattern in the marine bacterioplankton occupation frequency, also related to the high abundance of organisms found at most sites. Although a bimodal pattern in frequency persistency was not directly measured at these sites, the recurrent spatial pattern and repeated finding of common representatives throughout the study, suggest that this could likely be observed. In a more recent example, the measurement of temporal persistence in a tropical freshwater environment demonstrated the existence of a bimodal distribution (Mateus-Barros 2023). In these two cases, a relevant geographical feature that may have contributed to it is a more pronounced geographical isolation, which would hinder the action of climate factors on a continental scale while facilitating the action of species sorting (Leibold, Holyoak et al. 2004) by environmental factors, thus guiding these communities. Here, we have a sampling site highly connected to the Atlantic Ocean, which undergoes successive climate changes related to the seasonal influence of the Brazil and Malvinas currents (i.e. Brazil-Malvinas Confluence; Burone, Franco-Fraguas et al. 2021). The Brazil current performing a higher influence during the Winter-Spring period, is notable for raising the water temperatures in the region, also bringing water richer in nutrients from the tropics (Burone, Franco-Fraguas et al. 2021). This factor is favors the recurrence of cyanobacteria blooms that occur every spring in the region. On the other hand, the colder waters brought by the Malvinas current, particularly during the Summer-Autumn period, result in a profound change in local characteristics, which creates challenges for a large portion of the organisms that make up the local bacterioplankton to manage to persist through time.

Although a significant bimodality in bacterial frequency persistency cannot be observed, there is still a certain number of ASVs that could be found in all the samples here analyzed. Therefore, we applied an approach that made it possible to identify organisms that may play key roles in local processes. We used the proposed cutoffs to identify the “Abundant” (i.e. relative abundance greater than 1% of the total; Pedrós-Alió 2012) and “Persistent” ASVs (i.e. persistence greater than 90%, similar to the proposed occupancy frequency; Hanski 1982). We also identified a “Core” of bacteria (abundance greater than 1% and persistence greater than 90%) and referred to the remaining community as "non-Key" bacteria. Although each of the groups exhibited their own variation, a recurring factor across all groups is the relationship between this variation and local temperature fluctuations and nutrient availability from various sources. This pattern reinforces the link with large-scale changes driven by the Brazil-Malvinas currents confluence, and aligns with other studies in the global ocean indicating that temperature is a predominant factor guiding microbial diversity across the globe (Sunagawa, Coelho et al. 2015, Zorz, Willis et al. 2019).

Among the organisms most commonly found among the key taxa, SAR11 and Flavobacteriaceae stand out. Notably, Flavobacteriaceae are well known for their adaptive capacity, being found in different environments (Chiriac, Haber et al. 2022). These organisms possess a complex genome capable of decoding proteins that degrade a wide range of molecules, including CAZymes (Chiriac, Haber et al. 2022). Considering the nutrient-limited conditions of the oceanic environment, this ability to degrade carbon compounds of varying complexity may explain their consistent presence over time. In contrast, SAR11 are known for having a reduced genome (i.e., streamlined organisms). A growing effort in the genomic mapping of streamlined organisms has led to significant discoveries regarding these fascinating microbes. Bacteria currently identified only by their genetic codes, such as hgcI in freshwater environments (Kim, Kang et al. 2019) and SAR11 in marine environments (Giovannoni 2017), have had their metabolic pathways mapped. This has allowed the identification of molecules essential for their survival but that they cannot synthesize themselves due to the absence a set of essential genes for producing the required proteins. These findings enabled their cultivation in laboratory by the synthesis and provision of these molecules (Giovannoni 2017, Kim, Kang et al. 2019). Beyond being a major step toward species-level identification of these groups, this discovery also raises an important ecological question. The presence of these organisms in a given environment implies that these molecules are naturally synthesized—either through the physical breakdown of more complex molecules or via metabolic processing and bioavailability facilitated by other organisms. Moreover, it becomes evident that the availability of these resources may act as a limiting factor for biogeographical processes at both medium and large scales. In the specific case of SAR11, genomic mapping has identified a deficiency in the production of 4-amino-5-hydroxymethyl-2-methylpyrimidine, which can be found in isolates from marine cyanobacterial (Giovannoni 2017). This alone serves as strong evidence of SAR11’s dependency on other organisms. Additionally, here we identified a recurring fluctuation in their abundance over the years, with an increasing abundance shortly after the annual Synechococcus bloom events and then gradually declining until the following year’s bloom event. Co-occurrence analyses further support this idea, showing that distinct ASVs identified as belonging to these two groups repeatedly co-occur, even forming a

cluster in the co-occurrence network composed exclusively of ASVs from these two taxa.

## Conclusions

By using a simple tool for classifying ASVs among distinct biological components, we had a simpler way to identify groups of interest and factors that guide the variation of diversity as a whole. Here, we were able to demonstrate that the impact of periodical changes in current dynamics seems to be the most relevant factor to explain temporal variations in this community, by influencing the local availability of nutrients and local temperature fluctuations, factors widely recognized as guiding bacterioplankton diversity in the global ocean (Sunagawa, Coelho et al. 2015). Still, we managed to highlight co-occurrence patterns between some biological components, in a relationship that appears to be mediated by the sharing of organic compounds (Giovannoni 2017). Overall, this approach shows an interesting potential by separating distinct biological components into groups that, at least here, exhibit ecological coherence, facilitating the identification of processes occurring within this community. Thus, it proves to be an interesting tool for identifying factors guiding their dynamics, both spatially, as previously demonstrated, and temporally.

## Acknowledgements

This project was funded by the ANII-MPI grant (ANII_MPI_ID_2017_1_1007663) awarded to C.A. and R.A., the full-time dedication program (DT) of the Universidad de la República, and PEDECIBA support to C.A. The National Agency for Research and Innovation (ANII) is also acknowledged for supporting E.M.B. as a postdoctoral fellow during the execution of this study (PD_NAC_2023_1_177028). We further acknowledge the support of the Scientific Research Commission of the Universidad de la República Uruguay (CSIC) through the Grupos I+D 2022 program, which contributed to the completion of this work.

Special thanks are extended to Danilo Calliari, Maite Colina, Carolina Crisci, Carolina Lescano, Ana Martínez, Mariana Meerhoff, Andrés Pérez, Lorena Rodríguez, Nicolás Silvera, and the personnel from La Paloma Naval Prefecture for their dedication to maintaining the SAMO observatory time series, which is part of the microbial observatories initiative within the µSudAqua network.

## References

Acinas, S. G., V. Klepac-Ceraj, D. E. Hunt, C. Pharino, I. Ceraj, D. L. Distel and M. F. Polz (2004). "Fine-scale phylogenetic architecture of a complex bacterial community." Nature **430**(6999): 551-554.

Alonso, C., N. Musat, B. Adam, M. Kuypers and R. Amann (2012). "HISH–SIMS analysis of bacterial uptake of algal-derived carbon in the Río de la Plata estuary." Systematic and applied microbiology **35**(8): 541-548.

Amaral, V., D. Graeber, D. Calliari and C. Alonso (2016). "Strong linkages between DOM optical properties and main clades of aquatic bacteria." Limnology and Oceanography: n/a-n/a.

Azam, F., T. Fenchel, J. G. Field, J. S. Gray, L.-A. Meyer-Reil and F. Thingstad (1983). "The ecological role of water-column microbes in the sea." Marine ecology progress series: 257-263.

Barberán, A., E. O. Casamayor and N. Fierer (2014). "The microbial contribution to macroecology." Front Microbiol **5**: 203.

Brown, J. H. (1984). "On the relationship between abundance and distribution of species." The american naturalist **124**(2): 255-279.

Burone, L., P. Franco-Fraguas, A. Carranza, D. Calliari, M. Michaelovitch de Mahiques, M. Gómez, Y. Marin, O. Gutiérrez and L. Ortega (2021). "Physical Drivers and Dominant Oceanographic Processes on the Uruguayan Margin (Southwestern Atlantic): A Review and a Conceptual Model." Journal of Marine Science and Engineering **9**(3): 304.

Callahan, B. J., P. J. McMurdie, M. J. Rosen, A. W. Han, A. J. Johnson and S. P. Holmes (2016). "DADA2: High-resolution sample inference from Illumina amplicon data." Nat Methods **13**(7): 581-583.

Cavicchioli, R., W. J. Ripple, K. N. Timmis, F. Azam, L. R. Bakken, M. Baylis, M. J. Behrenfeld, A. Boetius, P. W. Boyd, A. T. Classen, T. W. Crowther, R. Danovaro, C. M. Foreman, J. Huisman, D. A. Hutchins, J. K. Jansson, D. M. Karl, B. Koskella, D. B. Mark Welch, J. B. H. Martiny, M. A. Moran, V. J. Orphan, D. S. Reay, J. V. Remais, V. I. Rich, B. K. Singh, L. Y. Stein, F. J. Stewart, M. B. Sullivan, M. J. H. van Oppen, S. C. Weaver, E. A. Webb and N. S. Webster (2019). "Scientists' warning to humanity: microorganisms and climate change." Nat Rev Microbiol **17**(9): 569-586.

Chiriac, M. C., M. Haber and M. M. Salcher (2022). "Adaptive genetic traits in pelagic freshwater microbes." Environ Microbiol.

Drula, E., M.-L. Garron, S. Dogan, V. Lombard, B. Henrissat and N. Terrapon (2022). "The carbohydrate-active enzyme database: functions and literature." Nucleic acids research **50**(D1): D571-D577.

Escalas, A., F. S. Paula, F. Guilhaumon, M. Yuan, Y. Yang, L. Wu, F. Liu, J. Feng, Y. Zhang and J. Zhou (2022). "Macroecological distributions of gene variants highlight the functional organization of soil microbial systems." The ISME Journal **16**(3): 726-737.

Fermani, P., N. Martyniuk, M. Saraceno, M. Gerea, C. Sabio, M. R. Schiaffino, M. L. Sánchez, S. A. Dohle, C. Alonso and M. A. Barrena (2024). "A Latin American network of microbial observatories for monitoring aquatic ecosystems." Ecología Austral **34**(3): 633-643.

Fonte, L. F. M. d., G. Latombe, M. Gordo, M. Menin, A. P. de Almeida, C. Hui and S. Lötters (2021). "Amphibian diversity in the Amazonian floating meadows: a Hanski core-satellite species system." Ecography **44**(9): 1325-1340.

Gaston, K. J., T. M. Blackburn, J. J. Greenwood, R. D. Gregory, R. M. Quinn and J. H. Lawton (2000). "Abundance–occupancy relationships." Journal of Applied Ecology **37**(s1): 39-59.

Giovannoni, S. J. (2017). "SAR11 Bacteria: The Most Abundant Plankton in the Oceans." Ann Rev Mar Sci **9**: 231-255.

Giovannoni, S. J., J. Cameron Thrash and B. Temperton (2014). "Implications of streamlining theory for microbial ecology." The ISME journal **8**(8): 1553-1565.

Gleason, H. (1929). "The significance of Raunkiaer's law of frequency." Ecology **10**(4): 406-408.

Gotelli, N. J. (1991). "Metapopulation models: the rescue effect, the propagule rain, and the core-satellite hypothesis." The American Naturalist **138**(3): 768-776.

Griffith, D. M., J. A. Veech and C. J. Marsh (2016). "Cooccur: probabilistic species co-occurrence analysis in R." Journal of Statistical Software **69**: 1-17.

Hanski, I. (1982). "Dynamics of Regional Distribution: The Core and Satellite Species Hypothesis." Oikos **38**(2): 210-221.

Horner-Devine, M. C. and B. J. Bohannan (2006). "Unifying ecology to include all creatures great and small." Trends in ecology & evolution **21**(9): 473.

Huisman, J. and F. D. Hulot (2005). Population dynamics of harmful cyanobacteria. Harmful cyanobacteria, Springer**:** 143-176.

Jeong, S. Y., J. Y. Choi and T. G. Kim (2020). "Coordinated Metacommunity Assembly and Spatial Distribution of Multiple Microbial Kingdoms within a Lake." Microb Ecol **79**: 801-814.

Jiao, N., G. J. Herndl, D. A. Hansell, R. Benner, G. Kattner, S. W. Wilhelm, D. L. Kirchman, M. G. Weinbauer, T. Luo, F. Chen and F. Azam (2010). "Microbial production of recalcitrant dissolved organic matter: long-term carbon storage in the global ocean." Nat Rev Microbiol **8**(8): 593-599.

Kawasaki, N. and R. Benner (2006). "Bacterial release of dissolved organic matter during cell growth and decline: molecular origin and composition." Limnology and Oceanography **51**(5): 2170-2180.

Kim, S., I. Kang, J.-H. Seo and J.-C. Cho (2019). "Culturing the ubiquitous freshwater actinobacterial acI lineage by supplying a biochemical 'helper'catalase." The ISME journal **13**(9): 2252-2263.

Kopf, A., M. Bicak, R. Kottmann, J. Schnetzer, I. Kostadinov, K. Lehmann, A. Fernandez-Guerra, C. Jeanthon, E. Rahav and M. Ullrich (2015). "The ocean sampling day consortium." Gigascience **4**: 1-5.

Leibold, M. A., M. Holyoak, N. Mouquet, P. Amarasekare, J. M. Chase, M. F. Hoopes, R. D. Holt, J. B. Shurin, R. Law, D. Tilman, M. Loreau and A. Gonzalez (2004). "The metacommunity concept: a framework for multi-scale community ecology." Ecology Letters **7**(7): 601-613.

Lindh, M. V., J. Sjostedt, B. Ekstam, M. Casini, D. Lundin, L. W. Hugerth, Y. O. Hu, A. F. Andersson, A. Andersson, C. Legrand and J. Pinhassi (2017). "Metapopulation theory identifies biogeographical patterns among core and satellite marine bacteria scaling from tens to thousands of kilometers." Environ Microbiol **19**(3): 1222-1236.

Lindström, E. S., M. P. Kamst-Van Agterveld and G. Zwart (2005). "Distribution of typical freshwater bacterial groups is associated with pH, temperature, and lake water retention time." Applied and environmental microbiology **71**(12): 8201-8206.

Martiny, J. B. H., B. J. M. Bohannan, J. H. Brown, R. K. Colwell, J. A. Fuhrman, J. L. Green, M. C. Horner-Devine, M. Kane, J. A. Krumins, C. R. Kuske, P. J. Morin, S. Naeem, L. Ovreas, A.-L. Reysenbach, V. H. Smith and J. T. Staley (2006). "Microbial biogeography: putting microorganisms on the map." Nat Rev Micro **4**(2): 102-112.

Mateus-Barros, E. (2023). "Padrões e processos que estruturam (meta) comunidades de bacterioplâncton, da escala local à intercontinental."

Mateus-Barros, E., M. L. de Melo, I. L. Bagatini, A. Caliman and H. Sarmento (2021). "Local and Geographic Factors Shape the Occupancy-Frequency Distribution of Freshwater Bacteria." Microbial Ecology.

Mehranvar, L. and D. A. Jackson (2001). "History and taxonomy: their roles in the core-satellite hypothesis." Oecologia **127**(1): 131-142.

Messer, L. F., M. Ostrowski, M. A. Doblin, K. Petrou, M. E. Baird, T. Ingleton, A. Bissett, J. Van de Kamp, T. Nelson, I. Paulsen, L. Bodrossy, J. A. Fuhrman, J. R. Seymour and M. V. Brown (2020). "Microbial tropicalization driven by a strengthening western ocean boundary current." Glob Chang Biol **26**(10): 5613-5629.

Mistry, J., S. Chuguransky, L. Williams, M. Qureshi, G. A. Salazar, E. L. Sonnhammer, S. C. Tosatto, L. Paladin, S. Raj and L. J. Richardson (2021). "Pfam: The protein families database in 2021." Nucleic acids research **49**(D1): D412-D419.

Mitchell-Olds, T. and R. G. Shaw (1987). "REGRESSION ANALYSIS OF NATURAL SELECTION: STATISTICAL INFERENCE AND BIOLOGICAL INTERPRETATION." Evolution **41**(6): 1149-1161.

Myklestad, S. M. (2000). Dissolved organic carbon from phytoplankton. Marine chemistry, Springer**:** 111-148.

Oksanen, J., F. G. Blanchet, M. Friendly, R. Kindt, P. Legendre, D. McGlinn, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, E. Szoecs and H. Wagner (2016). "Vegan: community ecology package."

Papp, L. and J. Izsák (1997). "Bimodality in occurrence classes: a direct consequence of lognormal or logarithmic series distribution of abundances: a numerical experimentation." Oikos **79**(1): 191-194.

Parada, A., D. M. Needham and J. A. Fuhrman (2015). "Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time-series and global field samples." Environ Microbiol.

Pedrós-Alió, C. (2012). "The rare bacterial biosphere." Ann Rev Mar Sci **4**: 449-466.

Pereira-Flores, E., A. Barberan, F. O. Glöckner and A. Fernandez-Guerra (2021). "Mg-Traits pipeline: advancing functional trait-based approaches in metagenomics." ARPHA Conference Abstracts **4**.

R Core Team (2019). R: A Language and Environment for Statistical Computing. Vienna, Austria, R Foundation for Statistical Computing.

Rappé, M. S. and S. J. Giovannoni (2003). "The uncultured microbial majority." Annual review of microbiology **57**(1): 369-394.

Rugema, E., F. Darchambeau, H. Sarmento, M. Stoyneva-Gärtner, M. Leitao, W. Thiery, A. Latli and J. P. Descy (2019). "Long-term change of phytoplankton in Lake Kivu: The rise of the greens." Freshwater Biology.

Sarmento, H. and J. M. Gasol (2012). "Use of phytoplankton-derived dissolved organic carbon by different types of bacterioplankton." Environ Microbiol **14**(9): 2348-2360.

Soininen, J., J. Heino and J. Wang (2018). "A meta-analysis of nestedness and turnover components of beta diversity across organisms and ecosystems." Global Ecology and Biogeography **27**(1): 96-109.

Steinegger, M., M. Mirdita and J. Söding (2019). "Protein-level assembly increases protein sequence recovery from metagenomic samples manyfold." Nature methods **16**(7): 603-606.

Stoeck, T., D. Bass, M. Nebel, R. Christen, M. D. Jones, H. W. BREINER and T. A. Richards (2010). "Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water." Molecular ecology **19**: 21-31.

Sunagawa, S., L. P. Coelho, S. Chaffron, J. R. Kultima, K. Labadie, G. Salazar, B. Djahanschiri, G. Zeller, D. R. Mende, A. Alberti, F. M. Cornejo-Castillo, P. I. Costea, C. Cruaud, F. d'Ovidio, S. Engelen, I. Ferrera, J. M. Gasol, L. Guidi, F. Hildebrand, F. Kokoszka, C. Lepoivre, G. Lima-Mendez, J. Poulain, B. T. Poulos, M. Royo-Llonch, H. Sarmento, S. Vieira-Silva, C. Dimier, M. Picheral, S. Searson, S. Kandels-Lewis, T. O. coordinators, C. Bowler, C. de Vargas, G. Gorsky, N. Grimsley, P. Hingamp, D. Iudicone, O. Jaillon, F. Not, H. Ogata, S. Pesant, S. Speich, L. Stemmann, M. B. Sullivan, J. Weissenbach, P. Wincker, E. Karsenti, J. Raes, S. G. Acinas and P. Bork (2015). "Structure and function of the global ocean microbiome." Science **348**(6237).

Teeling, H., B. M. Fuchs, C. M. Bennke, K. Kruger, M. Chafee, L. Kappelmann, G. Reintjes, J. Waldmann, C. Quast, F. O. Glockner, J. Lucas, A. Wichels, G. Gerdts, K. H. Wiltshire and R. I. Amann (2016). "Recurring patterns in bacterioplankton dynamics during coastal spring algae blooms." Elife **5**: e11888.

Wickham, H., M. Averick, J. Bryan, W. Chang, L. D. A. McGowan, R. François, G. Grolemund, A. Hayes, L. Henry and J. Hester (2019). "Welcome to the Tidyverse." Journal of open source software **4**(43): 1686.

Zorz, J., C. Willis, A. M. Comeau, M. G. I. Langille, C. L. Johnson, W. K. W. Li and J. LaRoche (2019). "Drivers of Regional Bacterial Community Structure and Diversity in the Northwest Atlantic Ocean." Front Microbiol **10**: 281.