

RESEARCH ARTICLE

Open Access

Reducing the arbitrary: fuzzy detection of microbial ecotones and ecosystems – focus on the pelagic environment



Antoine Bagnaro^{1*}, Federico Baltar^{1,2,3}, Gretchen Brownstein⁴, William G. Lee⁴, Sergio E. Morales⁵, Daniel W. Pritchard^{1,6} and Christopher D. Hepburn¹

Abstract

Background: One of the central objectives of microbial ecology is to study the distribution of microbial communities and their association with their environments. Biogeographical studies have partitioned the oceans into provinces and regions, but the identification of their boundaries remains challenging, hindering our ability to study transition zones (i.e. ecotones) and microbial ecosystem heterogeneity. Fuzzy clustering is a promising method to do so, as it creates overlapping sets of clusters. The outputs of these analyses thus appear both structured (into clusters) and gradual (due to the overlaps), which aligns with the inherent continuity of the pelagic environment, and solves the issue of defining ecosystem boundaries.

Results: We show the suitability of applying fuzzy clustering to address the patchiness of microbial ecosystems, integrating environmental (Sea Surface Temperature, Salinity) and bacterioplankton data (Operational Taxonomic Units (OTUs) based on 16S rRNA gene) collected during six cruises over 1.5 years from the subtropical frontal zone off New Zealand. The technique was able to precisely identify ecological heterogeneity, distinguishing both the patches and the transitions between them. In particular we show that the subtropical front is a distinct, albeit transient, microbial ecosystem. Each water mass harboured a specific microbial community, and the characteristics of their ecotones matched the characteristics of the environmental transitions, highlighting that environmental mixing lead to community mixing. Further explorations into the OTU community compositions revealed that, although only a small proportion of the OTUs explained community variance, their associations with given water mass were consistent through time.

Conclusion: We demonstrate recurrent associations between microbial communities and dynamic oceanic features. Fuzzy clusters can be applied to any ecosystem (terrestrial, human, marine, etc) to solve uncertainties regarding the position of microbial ecological boundaries and to refine the relation between the distribution of microorganisms and their environment.

* Correspondence: antoine.bagnaro@postgrad.otago.ac.nz

¹Department of Marine Sciences, University of Otago, Dunedin, New Zealand
Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Nature is not uniform but is comprised of patches [1]. Identifying the boundaries between ecosystems and their characteristics remains a fundamental challenge for biogeographers and ecologists ([2, 3] for a review). Information on the spatial distribution of communities and ecosystems is key to understand community dynamics [4], and local and global diversity patterns [5] as well as guiding conservation strategies and ecosystem risk assessment protocols [6]. Initial qualitative assessment of boundaries within or between ecosystems can be done to direct sampling effort [7], but this approach is restricted to environments where delineations are clear (e.g. forest patches, [8], seagrass meadows, [9], agricultural land, [10]). Many ecosystems challenge such *a priori* designs by presenting no obvious patchiness [11, 12], gradualness [13, 14], by being dynamic [15, 16], because the kind of organisms under consideration are too small for visual assessment [17, 18], or a combination of all or some of the above. In these cases, the selection of threshold values, or the definition of arbitrary boundaries, is often required to classify the different components of the ecosystems. The definition of thresholds may be arguable [19, 20] and these values can vary as much as the system they define [21, 22]. Together, these uncertainties reduce the potential comparability between studies [23] and opportunities for broader investigations across ecosystems.

The marine pelagic environment and its microbial communities offer an important example of this problem. On one hand, advances in satellite imagery [24] and recent sampling cruises [25, 26] reveal a patchy and dynamic ocean, with associated difficulties in the detection and delineation of water masses and ecosystems [11, 12]. On the other hand, the constant growth of sequencing techniques produces increasingly complex microbial data sets that challenge ecological and statistical interpretation [27].

Despite the challenges, microbial communities demand attention since they fulfil key functions in the oceans, including as nutrient and carbon recyclers [28]. Approximately one-half of the carbon fixed by marine autotrophs is directly processed by bacteria [28, 29]. Bacteria also facilitate the regeneration of nitrogen and phosphorus [30] and the release of iron [31], processes crucial to primary productivity [32]. Identifying spatial patterns in marine bacterial assemblages have the potential to reveal heterogeneity in oceanic productivity or biogeochemical patterns [33]. Recent studies did high-light associations between bacterioplankton communities and oceanographic features in the pelagic oceans [17, 18, 34–38] and in coastal waters along estuarine to marine transitions [39, 40]. Consistently identifying these associations requires methods to place the boundaries of

ecosystems and to characterise the transitions between them (see [41, 42]). Transition zones between ecosystems, or ecotones, can play important ecological roles. They have been advocated as harbouring higher levels of productivity or diversity ([43] for a list of ecotone properties and further definition of ecotones) and to regulate the links between neighbouring systems [44, 45].

To examine ecosystems and their transitions in the pelagic environment, we used fuzzy clustering [46, 47], a technique that creates overlapping sets of clusters. A graphical explanation of the type of output generated by fuzzy clustering algorithms, in comparison to the more common k-means clustering, is given in Fig. 1. In fuzzy set theory, observations (usually sample sites in ecology) are assigned membership values in each fuzzy cluster in the range (0, 1), which expresses the degree to which the observation meets the definition of each cluster centroid [48]. In other words, the fuzzy cluster centroids correspond to the archetypical composition of the different communities [2], and each site is given membership values for each of the cluster, depending on how well the site composition reflects the centroid composition. Once plotted against the spatial dimensions of the gradient, the membership grades for a given cluster usually present a plateau surrounded by two declining edges, corresponding to a community core surrounded by two ecotones. As a consequence, the outputs of the analyses appear both structured (into clusters) and gradual (due to the overlaps), which aligns with the inherent continuity of natural systems, and particularly of the pelagic environment. Although not new, fuzzy logic approaches have been bypassed in ecological studies, where other types of gradient analyses prevail [10, 49]. However, rare applications have produced useful results [2, 50, 51]. The efficiency of fuzzy sets to describe ecological data has already been proven, when compared to ordinations such as canonical correspondence analysis (CCA) and distance-based redundancy analysis (DB-RDA) [48].

Here, fuzzy clusters were applied for the first time in the pelagic environment. We used the Munida Microbial Observatory Time Series (MOTS), a pelagic transect (for details see: Munida transect, [52, 53] for chemistry [17, 38, 54]); to explore the spatial associations of microbial communities (from OTUs based on 16S rRNA gene) and the surrounding oceanic structures (Sea Surface Temperature and Salinity) with a focus on ecotones.

Materials and methods

Study Site & Sampling

The sampling was done along the Munida Microbial Observatory Time Series (MOTS, Fig. 2a). MOTS is a coastal transect crossing the subtropical front off the East Coast of the South Island, New Zealand. This 65 km long transect crosses multiple major oceanic

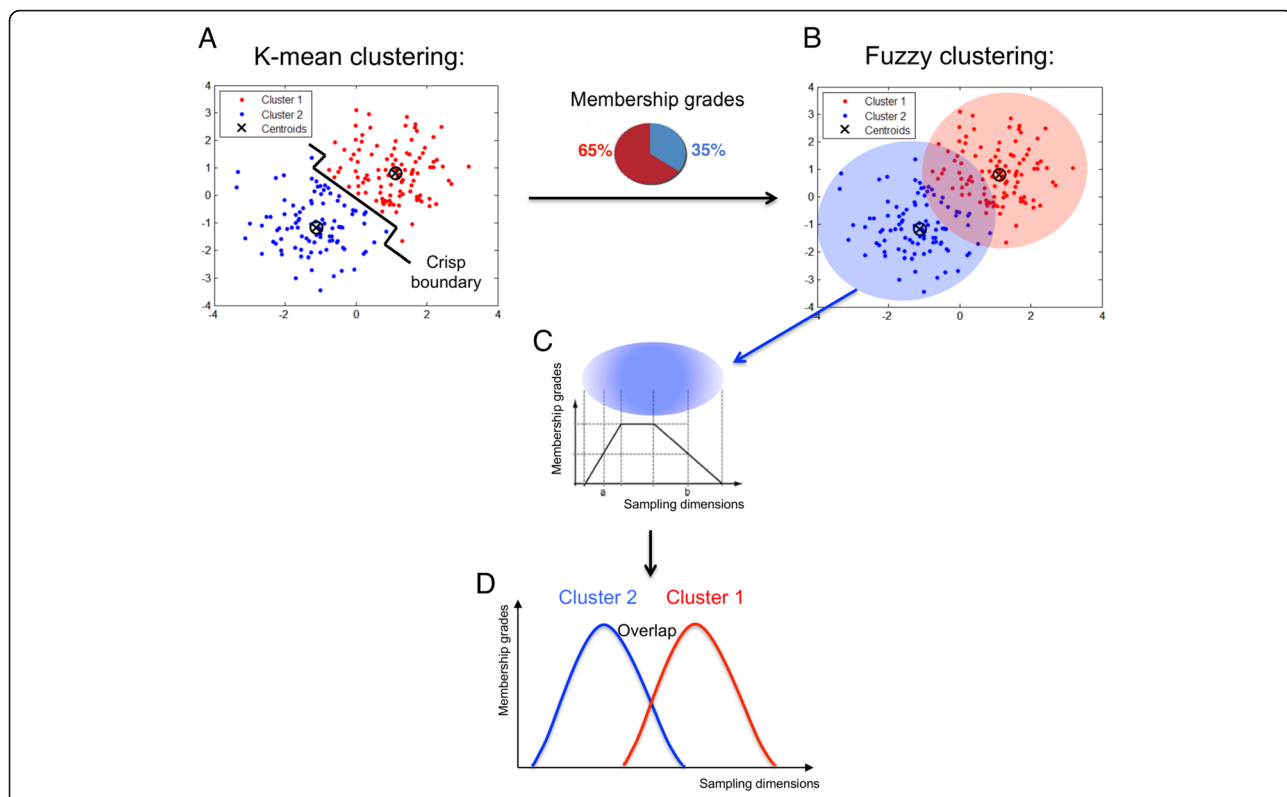


Fig. 1 **a** Typical outputs of K-means clustering (2 clusters). The separation between the two clusters is crisp, as observations can only belong to one cluster, regardless of their actual distance to the cluster centroids. **b** Typical outputs of fuzzy clustering (same 2 clusters). Each observation now belongs to both clusters, according to their degree of similarity with each cluster centroid. This translates in their membership grades (e.g. 65% in cluster 1 and 35% in cluster 2). There is no strict boundary between the two clusters, as they now overlap. **c** Schematic representation of the membership grade profile of a single cluster over the spatial dimensions of the sampling. **d** Schematic representation of the overlap between the two clusters membership grade profiles over the spatial dimensions of the sampling

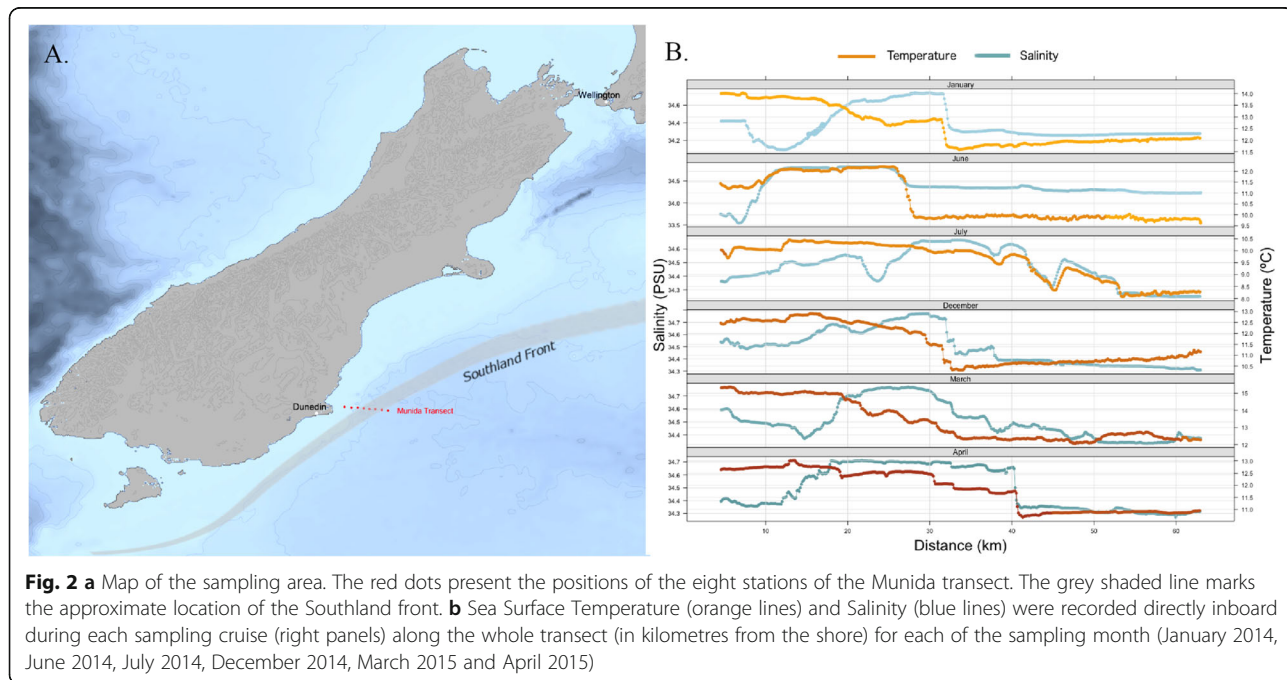


Fig. 2 **a** Map of the sampling area. The red dots present the positions of the eight stations of the Munida transect. The grey shaded line marks the approximate location of the Southland front. **b** Sea Surface Temperature (orange lines) and Salinity (blue lines) were recorded directly inboard during each sampling cruise (right panels) along the whole transect (in kilometres from the shore) for each of the sampling month (January 2014, June 2014, July 2014, December 2014, March 2015 and April 2015)

features: neritic waters (NW), subtropical waters (STW), the subtropical front (FRONT), and sub-Antarctic waters (SAW). For this reason, it has been continuously investigated for almost two decades. The position and the width of these water masses are dynamic throughout the year and the subtropical front in particular has been extensively surveyed [22, 55]. It displays seasonal variations, being farther offshore during winter and closer inshore during summer (around 43 km and 27 km respectively, estimated from [22]). It is also narrowest during winter (15.07 km) and broadest during spring (23.88 km), with an average width of 8.36 km [56] or 18 km [22].

Sampling took place between January 2014 and April 2015 on the RV *Polaris* (see Fig. 2b for the 6 sampling months: January 2014, June 2014, July 2014, December 2014, March 2015 and April 2015). Continuous sea surface temperature (SST) and sea surface salinity (SSS) data were measured using a Sea Bird SBE45 thermosalinograph and associated SBE38 remote temperature sensor, with position information appended from the vessel GPS system.

Microbial communities were determined at each of the 8 stations of the MOTS from surface samples (2 m below the surface) and collected in 5 l acid-rinsed plastic bottles and then filtered (0.5–0.8 l) through 0.22 µm polycarbonate filters (Millipore). Duplicate samples for DNA extractions were taken together with samples for chlorophyll-a (Chl-a) analysis from each station [17, 38].

DNA was extracted separately from each filter using a PowerSoil® DNA Isolation Kit (MoBio, Carlsbad, CA, USA) and the manufacturer's protocol. Bead beating was performed using a Geno/Grinder for 2 × 15 s and the final elution was done using 50 µL of solution C6 (sterile elution buffer, 10 mM Tris). DNA concentration was measured using a Nanodrop Spectrophotometer from Thermo Fisher. Amplicon libraries were generated using the Earth Microbiome Project barcoded primer set (V4 Primer 806R) and conditions [57]. All samples (independent replicates) were run on an Illumina Miseq run.

All amplicon data was processed using QIIME 1.9.1 and quality filtered using default parameters [57] with all fragments kept being 151 bp. Samples with less than 10,000 sequences were removed. Sequences were grouped into operational taxonomic units (OTUs) at 97% similarity. Open-reference OTU picking was carried out using the SILVA 119 release reference library and UCLUST [58]. Taxonomic assignments were performed using BLAST and the SILVA reference database. Multiple rarefactions (10) were performed to a depth of 10,000 sequences per sample and the files were merged to create an averaged OTU table. All data in the merged biom file were rounded prior to downstream analysis using the phyloseq package [59] in R (R Core Team 2019, [60]). Relative abundance (%) was calculated as the number of reads matching a particular OTU relative to the total

number of reads. The data were then averaged for each of the sampling station (8 stations per month over 6 months) according to their positions along the transect (expressed in kilometres from the coastline) in order to build the microbial community matrices used for the statistical analysis.

Statistical analysis

Fuzzy clusters (Fuzzy C-Means (FCM)) were constructed independently based on either the continuous temperature and salinity data (for water mass identification) or the microbial OTU data (for microbial communities identification) (fanny function, cluster package, [61], complemented by the vegclust function of the vegclust package, [47], to extract the fuzzy clusters centroid composition). The final number of clusters was chosen through an iterative process, with regard to the normalized partition coefficient and the normalized partition entropy [47]. Due to mathematical limitations, the number of fuzzy clusters cannot exceed $n/2 - 1$, where n is the total number of observations. This limitation meant that it was not possible to compute more than 3 fuzzy clusters out of the 8 sampled stations for the microbial communities. We therefore could not test for the existence of a fourth (or more) cluster in the OTU data.

First and second derivatives were calculated for each cluster from the regression of membership score on distance along the transect. The local maximum of the first derivative was used to pinpoint the location of the transitions between main environmental or community types and to assess the abruptness of these transitions (maximum slopes). The interval between two local maxima or minima was used as a proxy for the width of the environmental or community types. The widths of the transitions were assessed as being the interval between two local maxima of the second derivatives. We extracted these values for each of the sampling months and each of the clusters and compared the values grouped per month, seasons and water masses or communities. The statistical significance of the differences between groups was assessed by non-parametric Kruskal-Wallis tests followed by Dunn's test for multiple pairwise comparisons and Bonferroni p -values adjustments (dunn.test package, [62]).

We extracted the fuzzy cluster centroids compositions (with the vegclust package) for each of the community clusters. These centroids correspond to the prototypical vector of the clusters in the multidimensional space of species abundances and can be used to represent community types ([47] and references therein). Every "species" in the dataset is present in the cluster centroids, and the most representative ones can be kept to describe the taxonomic composition of the clusters [2]. We compared the evolution of the fuzzy clusters at increasing taxonomic resolutions (from phyla to OTUs, Fig. S1.2). Further analyses on community composition were then

conducted at the OTU level, to achieve a finer level of description [63]. Due to the high number of sampled OTUs, these results are presented as networks (qgraph package, [64]), in which the links correspond to the number of shared OTUs (OTU identities) between clusters. In order to avoid arbitrary decisions, a range of different thresholds was used to select the most representative OTUs in each cluster (ranging from 25 to 1% most representative, see supplementary information, Fig. SI.3A-3F). Networks often have several and equally correct graphical representations (in terms of where the nodes and edges should be drawn). We used the spinglass algorithm (repeated over 1000 networks, igraph package, [65]) to determine the actual number of communities in the network.

Results

Water masses and Bacterioplankton communities

The continuous retrieval of temperature and salinity data allowed for a precise recording of the main water

parameters along the Munida transect. The positions of the water masses could be directly approximated from the evolutions of SST and salinity (Fig. 2b) according to several thresholds ([55] for a complete list of values). The environmental fuzzy clusters, based on these two parameters, consistently described these water masses (Fig. 3a and Fig. SI. 1), and mean values for SST and SSS could be extracted from the fuzzy cluster centroids. Neritic waters (NW), characterised by high temperatures and low salinities (13.15 ± 1.25 °C, 34.32 ± 0.28 PSU), were the first to be crossed by the transect and finished between 9 and 19 km offshore depending on the sampling month. Subtropical waters (STW), characterised by higher temperatures and salinities (12.32 ± 0.93 °C, 34.67 ± 0.08 PSU), were next and continued until 27 to 32 km offshore. The subtropical front itself was described as a separate cluster by the FCM (11.30 ± 0.88 °C, 34.42 ± 0.13 PSU), with a width between 11 km and 22 km (for an average of 17 km), and occurred

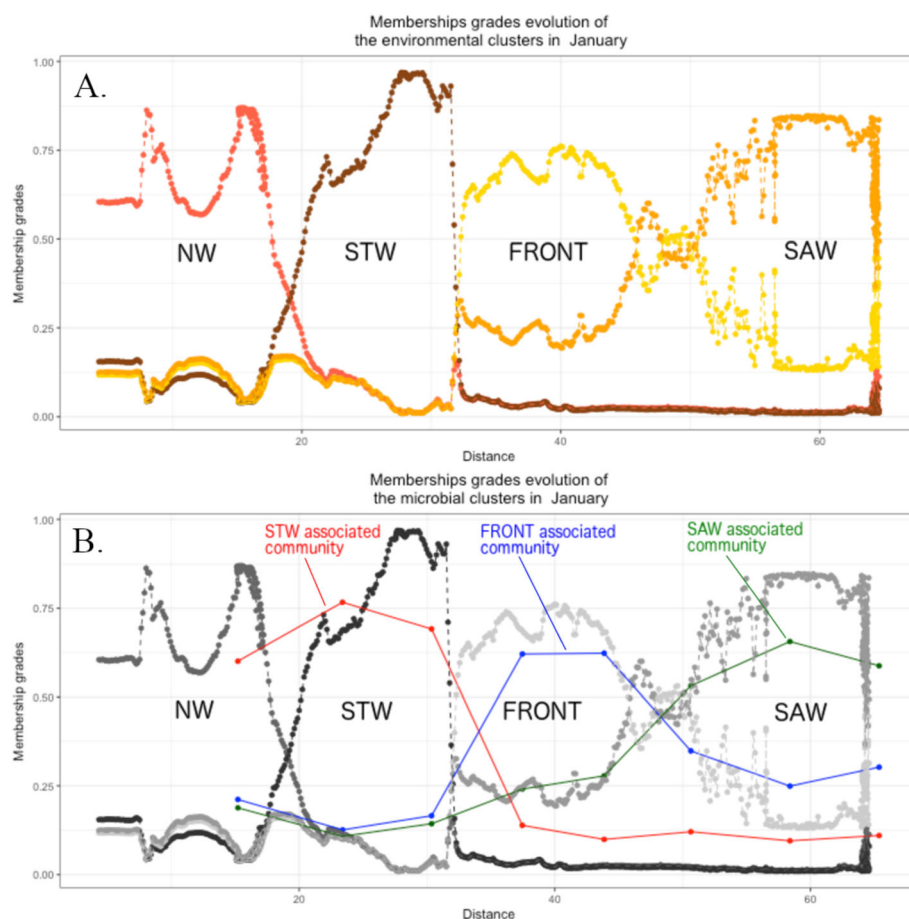


Fig. 3 Example of the evolution of the fuzzy membership grades along the Munida transect (in kilometres from the shore) in January 2014. **a** The clusters based on environmental parameters (Surface temperature and Salinity) describe the different water masses: Neritic waters (red), Sub-Tropical waters (brown), Front (yellow) and Sub-Antarctic waters (orange). **b** Evolution of the microbial community clusters (based on OTUs) along the transect. The environmental clusters are kept in shades of grey for easiness of comparison. The evolution of the fuzzy membership grades for the other sampling months can be found in the SI, Fig. 1A to F

between the end of the STW and the beginning of the sub-Antarctic waters (SAW), which began between 41 and 53 km offshore. The low temperatures and salinity values, characteristic of the SAW ($10.78 \pm 1.32 \text{ }^\circ\text{C}$, $34.31 \pm 0.06 \text{ PSU}$), then continued until the end of the transect. The values given here are intervals, to account for the shape of the FCM and the monthly variability. We could not depict any clear seasonal patterns in the positions of the water masses (see supplementary information, Fig. SI. 1A-1F).

The FCM also described a spatial segregation of bacterioplankton communities along the transect (Fig. 3b and Fig. SI. 1). Due to the number of sampling stations

and to computational constraints, the number of possible clusters was limited to three. Nevertheless, they were spatially organised and their positions coincided with the position of particular water masses. As such, it was possible to associate the bacterioplankton communities with the water masses in which they occurred. The STW and SAW could always be linked to a distinct microbial community, however the NW (present in July, March and April) and the subtropical front (present in January, June and December) inconsistently fluctuated between distinct and shared community clusters with bordering water masses. The relative positions of the fuzzy clusters did not vary dramatically depending on

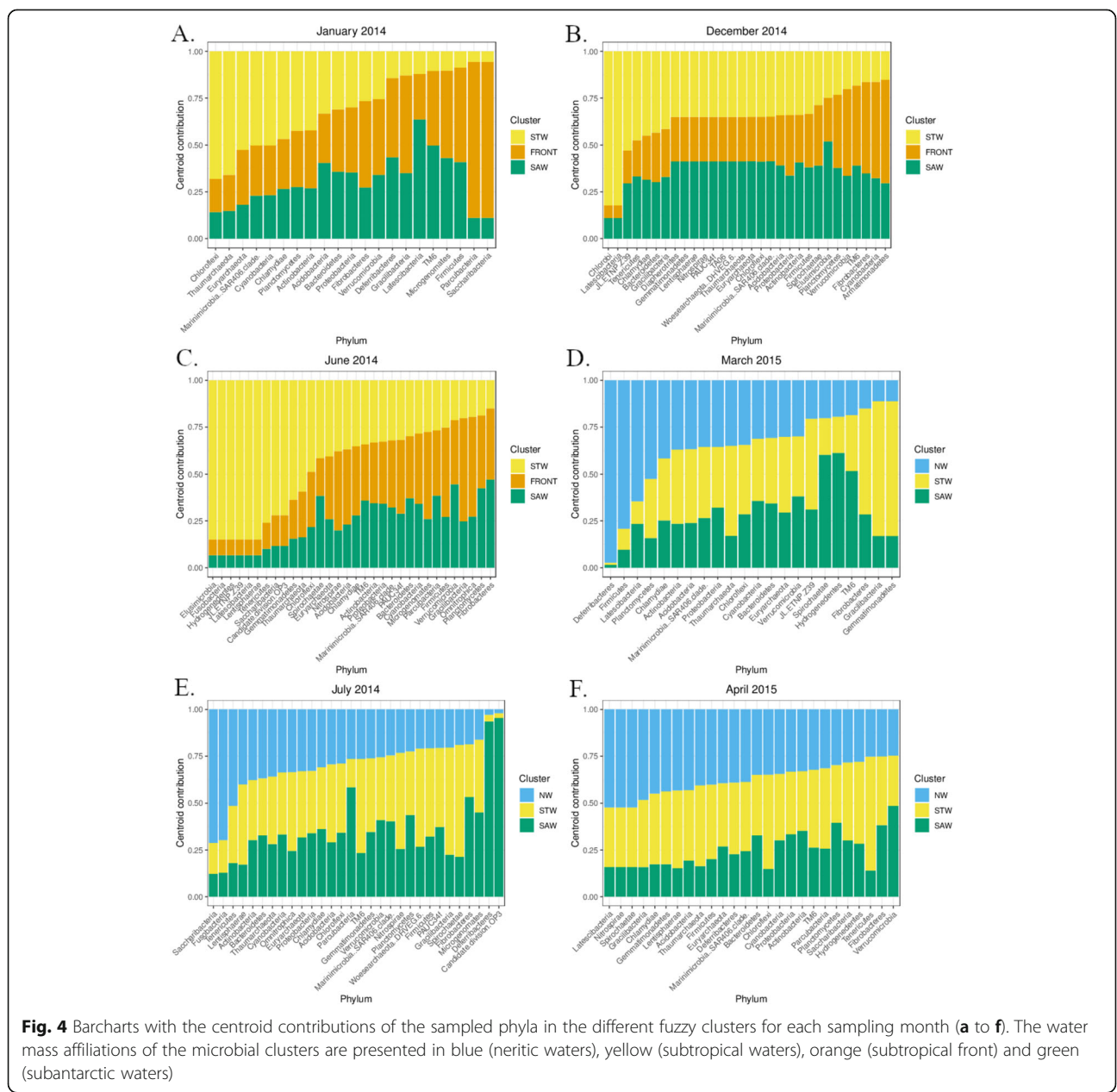


Fig. 4 Barcharts with the centroid contributions of the sampled phyla in the different fuzzy clusters for each sampling month (a to f). The water mass affiliations of the microbial clusters are presented in blue (neritic waters), yellow (subtropical waters), orange (subtropical front) and green (subantarctic waters)

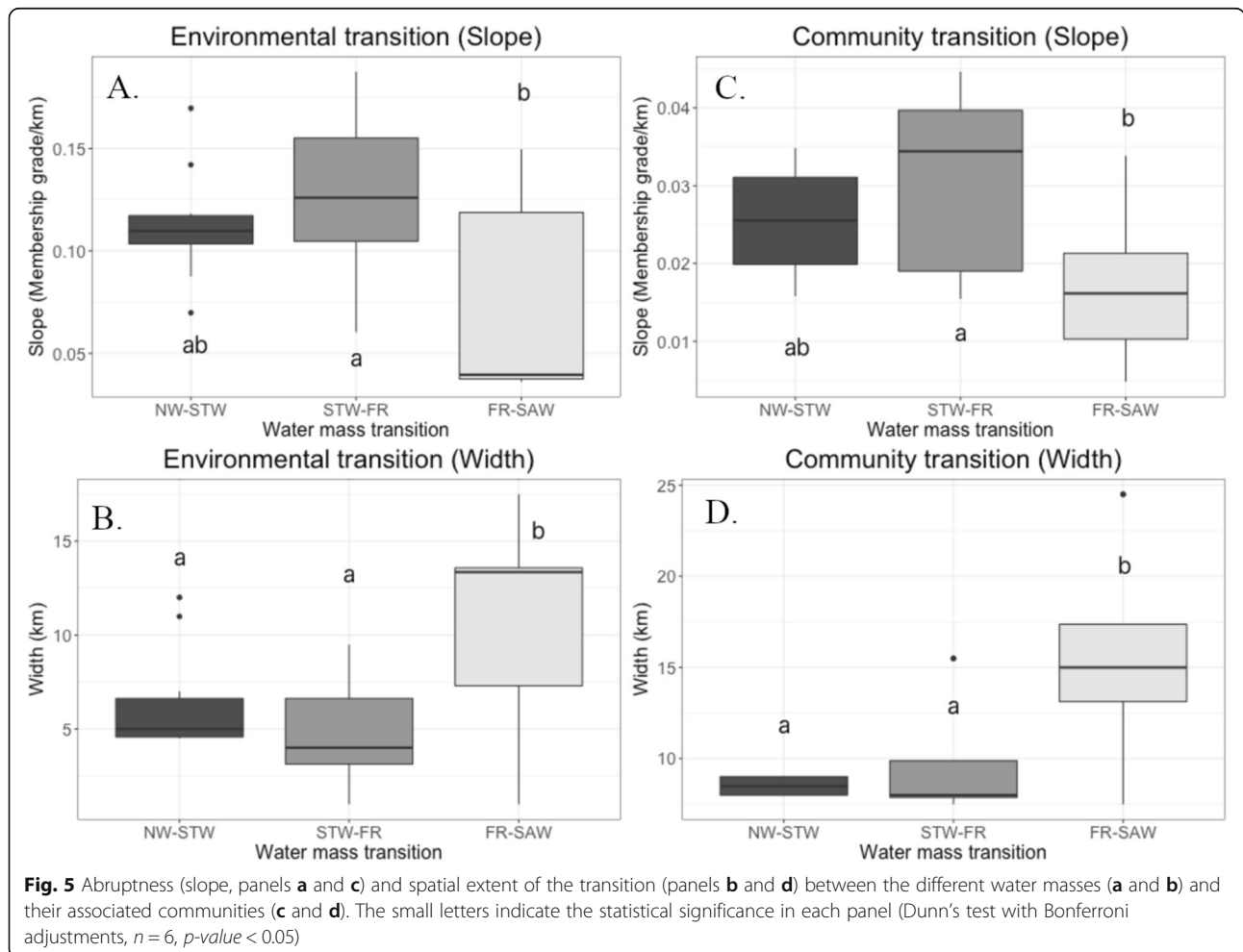
the taxonomic accuracy, but their spatial fit with the environmental clusters seemed to improve with increasing taxonomic resolution (Fig. SI. 2). The phylum composition of the fuzzy cluster centroids varied between months (Fig. 4). Some phyla (e.g. *Euryarchaeota*, *Verrucomicrobia*, *Planctomycetes* and *Bacteroidetes*) were consistently scoring intermediate values in each cluster, indicating ubiquitous distribution. Other phyla were more consistently associated with a single cluster at a time (e.g. *Latescibacteria* and *Tenericutes*, in NW or STW) and their spatial distribution therefore coincided with distinct water masses.

The overlaps and shapes created by the evolution of the FCM membership grades formed overlaps and areas of more or less mixed conditions between the different clusters. This representation offers a straightforward way to visualise the spatial organisation – or patchiness – of the environment from multivariate data. The status of boundary data, i.e. of data located in the outer part of the clusters, can therefore be more easily addressed.

Transition patterns

The characteristic (i.e., abruptness and the extent, here measured as slope and width) of the environmental transition zones along the transect were estimated (Fig. 5a-b). Overall, the transitions between the STW and the subtropical front were steeper and shorter than the transition from the subtropical front to the SAW (respectively $\times 1.69$ steeper and $\times 2.29$ km shorter on average, assessed with Kruskal-Wallis rank sum test, p -value < 0.05 , $n = 12$), whereas the characteristics of the transitions between the NW and the STW did not significantly differ from the transitions between the STW and the subtropical front. All these values were subjected to important variations from 1 month to the other but did not reveal any strong seasonal trends.

The shapes of the transitions between the microbial clusters were congruent with those between the environmental clusters (Fig. 4c-d). The transitions between the communities in the STW and the subtropical front were $\times 1.78$ steeper and $\times 1.59$ shorter than the transitions



between the communities in the subtropical front and the SAW. These differences were statistically significant (Kruskal-Wallis rank sum test, $p\text{-value} < 0.05$, $n = 6$), meaning that the environmental transitions and ecotones between the STW and the subtropical front were sharper and shorter than the ones between the front and the SAW. The transitions between the NW and the STW were not statistically different from the transitions from the STW and the subtropical front.

System dynamic and relations between water masses and microbial communities

As mentioned above, the variations in the positions and width of the microbial clusters matched those of the water mass (i.e. environmental) clusters. A network analysis (Fig. 6a) showed that these spatial associations also corresponded to a higher relatedness of the microbial communities that were representative of a particular water mass throughout the sampling period. As such, the links between the microbial communities that belonged to the same water masses were stronger than the links uniting water masses of different origins.

A visual analysis of the network structure reveals three different groups (i.e., the NW, STW and SAW associated communities), whereas the subtropical front communities are poorly linked with each other but draw a gap between the NW and STW communities on one side and the SAW communities on the other. Nevertheless, the spin-glass algorithm only detected 2 statistical communities in the network (one comprised of the NW and STW communities and the other of the SAW and frontal communities), with the rare occurrence of the subtropical front community in June as a third isolate (>1% of the networks). Although not in a different group, the STW community in December often appeared slightly apart from the other STW communities.

Importantly, the number of OTUs that were present in each cluster centroid only accounted for a small portion of the total number of OTUs. Such proportions are indicative of the degree of uniqueness in the microbial communities, as only a fraction of the OTUs played a role in linking the clusters (number of common OTUs between cluster centroids presented in Fig. 6b). Each month harboured an average of 41.1% ($\pm 4.7\%$) of the total number

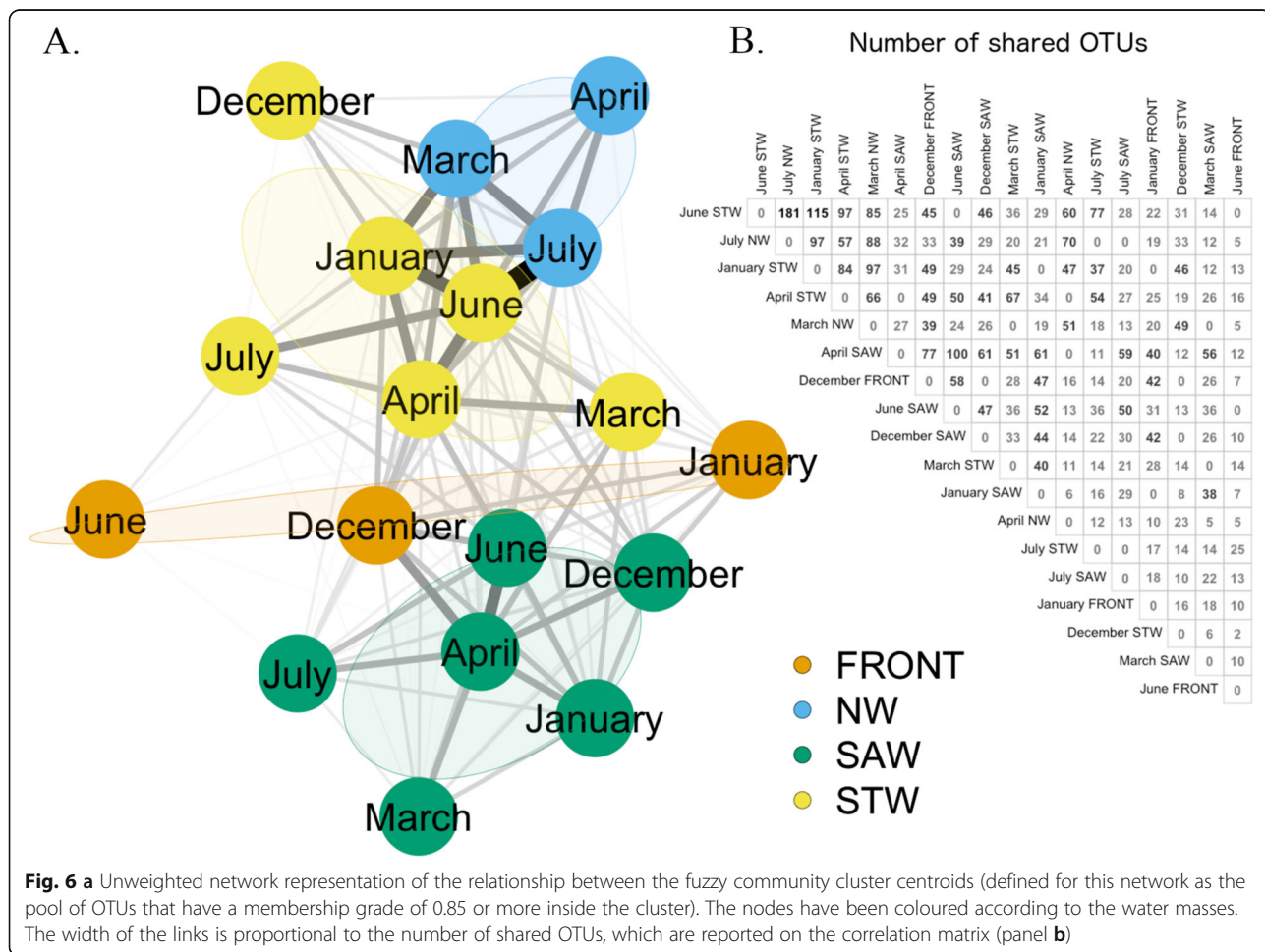


Fig. 6 a Unweighted network representation of the relationship between the fuzzy community cluster centroids (defined for this network as the pool of OTUs that have a membership grade of 0.85 or more inside the cluster). The nodes have been coloured according to the water masses. The width of the links is proportional to the number of shared OTUs, which are reported on the correlation matrix (panel **b**)

of OTUs over the whole sampling period, and shared 51.5% ($\pm 4.5\%$) of these remaining OTUs with other months. In the example network presented here (threshold value of 0.85, Fig. 6a), an average of 16.3% ($\pm 5.3\%$) of the total number of OTUs/cluster centroid were above the threshold and considered as representative of a community. The remaining OTUs were therefore too ubiquitous to be kept in one particular cluster centroid. Among the representative OTUs of each cluster, an average of 6.4% ($\pm 1.0\%$) were shared. The majority of the OTUs were therefore either too ubiquitous to represent a particular cluster, or too specific to be found in more than one cluster.

Discussion

Method considerations, suitability and efficiency

In this study, we used a fuzzy clustering approach for the first time to unravel the patchiness and the associated ecotones of microbial communities in the pelagic environment. We demonstrate that fuzzy clustering provides an efficient tool to unravel the heterogeneity of apparently continuous environments, particularly when the use of multivariate data is required, as it is the case for water masses (temperature and salinity) and microbial communities. There are three main advantages. First, the obtained graphical representation respects the continuity of the original data. It thus avoids arbitrary decisions regarding the assignation of spatial boundaries between clusters [66]. Second, FCM can be applied both on species (here OTU) and environmental data, with similar graphical outputs, which highlight spatial associations and makes comparison more straightforward. Finally, information from the cluster centroids can be extracted. These sets of vectors can be interpreted as biological description of the clusters [2, 47] or as representative values in the case of environmental variables. By keeping the ecosystem description structured in patches, it offers the possibility to bring together studies centred on ecosystem cores and research on ecotones, two approaches that have often been treated separately, despite their relation to understand landscapes as dynamic and networked entities [67, 68]. Recognising that patches can grade into each other aligns with more recent ecological theories, such as network theory [69] and complexity theory [67], that recognise that exchanges, links and continuity between entities are crucial for the description and functioning of ecosystems. Moreover, several quantitative metrics can be extracted from the shape of the fuzzy clusters, such as the extent and strength of the mixing area (here measured as widths and rate of change). Modelling studies suggest that the sharpness of the environmental transition is a major factor in controlling the characteristics of the ecotone,

however, there is still a lack of empirical data to calibrate these models [4]. Using FCM along gradients may provide some of the metrics needed to progress in this field and prove useful for the parameterisation of statistical models that aim to account for fluxes and exchanges between compartments [70].

One drawback of this approach is the determination of the correct number of fuzzy clusters. Some mathematical indices have been developed to support this decision process, but they sometime fail to indicate a clear optimum [47]. The number of clusters is also constrained by the number of observations, which may be limiting at low sampling intensity.

Water masses and microbial communities

The high variability we observed in the positions of the different water masses between the successive cruises describes the study area as a very dynamic system, with a wide range of possible states. We do not report seasonal trends in the position of the front, but the lack of time averaging in this study may have captured ephemeral variations, such as eddies, plumes and meanders, that can hide seasonal trends or associated cycles (as revealed in longer time series, [22, 56]). Therefore, although seasonality affects the average position of the oceanic features in the area, the range of possible locations remains high throughout the year (as reported in [22, 71, 72]). Here, this can be seen in the July data where the temperature and salinity records were disturbed.

FCMs offer an alternative approach in oceanography, and contrast to existing techniques for oceanic jet detection in the oceans (that are mostly based on satellite imagery algorithms, the use of gradient thresholding methods, probability density functions and contour methods, see [72] for a review). These methods make use of additional physical parameters such as Sea Surface Height (SSH) that are further away from the field of biological oceanography. Although FCMs could also be applied on SSH data, the use of SST and SSS facilitate the comparison of the subtropical front with surrounding oceanic features, as salinity and temperature are the most common descriptors to define water masses. These descriptors are also ecologically meaningful [73, 74] and the FCM do not exclude the possibility of including additional ecological parameters to better understand the drivers of species and community distribution.

Patterns of association between plankton and oceanic circulation have been demonstrated at larger scales, mainly through ocean colour imagery (Chlorophyll a, but also with other pigments, proteins and lipids [18, 75–77]). The phylum composition of the different fuzzy cluster centroids highlighted a ubiquitous distribution of some of the phyla along the 60 km of the transect,

among which the *Euryarchaeota* (methanogens), *Verrucomicrobia*, *Planctomycetes*, *Bacteroidetes* (known particle degraders, [78]) and SAR406. Many of these phyla have been linked with anaerobic metabolisms in deeper parts of the oceans [79]. Their presence in surface samples may either indicate a strong mixing of the surface waters with deeper layers, the ability of some of the members of these phyla to live in more oxic conditions, or their association with suboxic microhabitats in particles (see [80, 81]). Other phyla were more consistently associated with a single cluster at a time, e.g. *Latescibacteria* (NW in March and April 2015, STW in June 2014), *Tenericutes* (NW July 2014, STW in June 2014 and April 2015). These particular associations mostly happened in the nearshore sections of the transect and could indicate that the subtropical front and the SAW are poorly differentiated at the phylum level, comparatively to the NW and STW. Information on the basic metabolism and ecological properties of these microorganisms are still scarce [82] and drawing ecological conclusions from these patterns would still require too many unsupported assumptions, particularly as microbial taxonomic diversity may be a poor predictor of microbial functional diversity [83]. The spatial segregation of the fuzzy clusters was only slightly more pronounced at increased taxonomic resolution (Fig. SI. 2), which may suggest that members of the same phylum might share general ecological strategies [84]. Ecological communities are shaped by more than just the sum of individual responses to surrounding environmental conditions. They are usually stabilised by facilitative or mutualistic relationships (sometime centred around a few key species [85]). Genomic studies have suggested that many bacterial phyla, and in particular those with small genome sizes could be obligate syntrophs, depending on larger bacterial hosts (e.g. in the Candidate Phyla Radiation and TM6 [82, 86, 87]). Such microbial associations may introduce reinforcement loops in the composition of microbial community, particularly at taxonomic resolutions that capture more specialised relationships. Investigating the co-occurrence of particular taxa in environmental datasets could help exploring these associations [82].

To account for these microbial associations, the community networks were thus constructed at the highest taxonomic resolution (OTU). The consistency of the links that united the bacterial communities of a particular water mass demonstrate the existence of biogeographic patterns in pelagic microbial communities (see also [88, 89] for bacterial communities), even at local scales over few kilometres. These patterns may be driven by the dispersal limitations induced by water masses circulation or by “ecosystem filtering” (including biotic and abiotic filters). We nevertheless recognise that these drivers unequally affected the microbial communities. Many OTUs were ubiquitous in the study area and therefore unrelated to a

particular cluster centroid, conversely many others were too rare, i.e. occurring in only one cluster, to play a role in the networks. Overall, only a small proportion of the OTUs explained the community variance. An average of 41% of the total number of OTUs sequenced over the whole sampling period were present per month and only 16.3% of that number were representative of a given cluster (the other 83.7% were too ubiquitous to represent a cluster). 6.4% were shared between clusters. These numbers can provide estimates for alpha and beta diversity (respectively defined as the diversity in a single community and the total dissimilarity between two or more communities, [90, 91]). If considered month per month, the high percentage of ubiquitous OTU advocate for low beta diversity between the water masses and a high alpha diversity in each water mass. Other microbial studies on land have found patterns of low beta diversity and high alpha diversity in microbial communities ([92] on soil microbiota), which is consistent with the results of this study. However, we point out that the low percentage of OTUs shared between the successive months suggests much higher beta diversity over time. In the case of microorganisms, their ability to evolve quickly may result in emergent patterns of beta diversity [93] and may provide an explanation for these findings. Inasmuch as we can tell from the limited data of this study, we advocate that time alone could be an important driver of microbial diversity patterns (along with dispersal limitations and environmental filtering) and that its effect on microbial diversity should be further isolated and explored, notably as it might provide a link between the slow-evolving macroorganisms and the fast-evolving microorganism diversity patterns.

Considering the study area, higher numbers of shared OTUs were reported between the microbial communities in the NW (and STW in a lesser extent) compared to the offshore communities (subtropical front and SAW, Fig. SI. 3), thus revealing more conserved communities over time. The strong influence of the Clutha River in the region, with its consistent inputs of freshwater and sediments, is responsible for the existence of a well differentiated NW, and is already known to support higher levels of primary productivity [94]. These terrestrial inputs may have had an additional stabilizing effect on the particle degrading bacterioplankton communities in the surface layers sampled in this study.

This study was indeed limited to the surface waters. Depth influences the water mass characteristics ([95], for New Zealand), and can drive microbial community differentiation, particularly on both side of the euphotic zone [96]. Mixing events however regularly occur and can lead to a homogenisation of the water column down to depths of around 100 m [22]. The conclusions of our study thus mainly apply to these depths.

Transition patterns, oceanic front and ecotone

It is generally accepted that water mass properties change rapidly across the front, yet remain approximately constant along the front [71]. As such, frontal areas have been regarded as transition zones between water masses and create potential ecotones [17, 38]. The FCMs were able to isolate the front as a distinct environmental cluster. The physical characteristics of the transitions between the different water masses and microbial communities along the continental shelf revealed that the passing from the frontal waters to the SAW were more gradual and over longer distances than the other transitions, which is probably the consequence of the sub-Antarctic origin of the subtropical front that makes its composition more similar to the SAW (90% SAW and 10% STW, according to [97]).

Despite this similarity with the SAW, the subtropical front in this study often maintained specific microbial clusters, comprising relatively low numbers of shared OTUs with the surrounding water masses (Fig. SI. 3). The origin of these few shared OTUs appeared to be equally partitioned between the STW and the SAW, but remained marginal in the cluster composition, highlighting the emergence of a separate community. These particular microbial communities were especially evident over the summer months, when stronger gradients in nutrient concentrations and temperature prevail and when phytoplankton blooms may occur [22, 52]. In these conditions, phytoplankton dynamics may drive the emergence of a distinctive microbial community [17, 38, 98, 99]. By contrast, a microbial community associated with the less salty neritic waters replaced it during the winter and autumn months, as the gradients associated with the front weaken [22]. Nevertheless, the appearance of a front-associated microbial cluster in June is at odds with this interpretation and we cannot reject a detrimental effect of the limited number of possible microbial fuzzy clusters.

The subtropical front did act as a clear separator between the STW and the SAW. Its position as an ecotone or as a separate and distinct ecosystem remains debatable. Temporal – and possibly seasonal – patterns may be the main drivers for whether or not the front harbours a separate community, or is an ecotone. Transient ecotones have been reported in the terrestrial literature [100]. This distinction is important, as the exchanges between the STW and the SAW are likely to be impacted by the presence of a distinctive microbial community between them, as well as the capacity of the frontal area to support higher levels of productivity [101] and attract organisms belonging to higher trophic levels [102].

Our study revealed the existence of another ecotone between the coastal NW and the STW, with the persistence of a coastal microbial biome. The occurrence of coastal front associated with freshwater inputs have

already been reported [103] and linked with distinct plankton communities [104]. Additionally, we demonstrate its influence on microbial communities, which has implications for primary productivity in coastal waters [28, 105]. Microbial ecotones may represent gradients of differing biochemical activities, where the frontal micro-organism benefit from the combined access to different sources of nutrients, and thus overcome potential growth limitations. Oceanic fronts are usually presented as areas of enhanced productivity [55, 106] and our ability to pinpoint and characterise the pelagic environment as continuous and comprised of both patches and the transitions between them might prove crucial to understand the linkage between the different trophic levels in the oceans, in turn improving our understanding of global biogeochemical cycles [33, 107].

Conclusion

The use of fuzzy clustering offers a novel way of integrating environmental and biotic entities in continuous representations, thus bringing together both patches and ecotones in ecosystem studies. Here, this technique allowed us to reveal the water mass organisation along the subtropical convergence zone off South New Zealand. In particular, we report the subtropical front as a distinct oceanic structure. Each water mass harboured distinctive bacterioplankton communities with high alpha diversity but low beta diversity although there was an important effect of time on microbial community composition. The characteristics of the ecotones between these communities matched the characteristics of the transitions between the different water masses, indicating that the mixing of the waters also resulted in the mixing of the microbial communities. It is noteworthy, that a pelagic oceanic front was used as a case study, but fuzzy clustering may be applied to a wide range of ecological systems and proves a powerful asset to resolve uncertainties regarding the patchiness along environmental gradients.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s40793-020-00363-w>.

Additional file 1. Fig. SI.1: Variations of the membership grades of the environmental fuzzy clusters (in shades of black) and of the microbial clusters (in red, blue and green) along the Munida transect (in kilometres from the shore) for the different month of the sampling. The microbial communities have been labelled according to the water mass that they occupy. The computations of the two sets of clusters (environmental and microbial) were independent thus their spatial correspondence is incidental. Fig. SI.2: Evolutions of the fuzzy clusters at the different sampling time (A to F) depending on the taxonomic resolution of the microbial data. Taxonomic levels spans from Phylum (lowest resolution, row 1) to OTUs (highest resolution, row 4), with order (row 2) and genus (row 3) as intermediate taxonomic resolutions. The water masses fuzzy clusters are presented in the background and are identical to Fig. SI.1. The relative positions of the transitions between the microbial clusters

are hardly affected by the taxonomic resolution. Their shapes, however, are and the microbial clusters match the environmental cluster best at increased taxonomic resolutions. Fig. S1.3: Network representations of the number of shared OTUs between the microbial communities in association to their water masses (colours) and the sampling month (labels). The thicknesses of the links are proportional to the number of shared OTUs. The exact numbers are reported on the matrix on the right of each network. The coloured ellipse areas correspond to the 95% probability region for the water masses in the networks. Note that the higher the threshold, the weaker the links. For the highest thresholds (0.95 and 0.99), some nodes are not linked to any others and their positions should not be interpreted.

Acknowledgements

We thank the skippers and crew of the RV *Polaris II*. We also wish to thank Dr. Kim Currie (NIWA), for access to the environmental data from the transect, and without whom the MOTS would not exist.

Authors' contributions

A.B., F.B., S.E.M., C.D.H., designed the study. S.E.M. analysed the sequencing data. A.B., G.B., D.W.P., performed the bioinformatic analyses. A.B., F.B., G.B., W.G.L., S.E.M., D.W.P., C.D.H., wrote the manuscript. All authors discussed the results and significantly contributed to critical revisions of the final manuscript. The authors read and approved the final manuscript.

Funding

This research was supported by a University of Otago doctoral scholarship and University of Otago Research Grants.

Availability of data and materials

The raw sequence reads have been deposited to the NCBI Short Read Archive (SRA) under BioProject accession number PRJNA430292 (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA430292/>). The datasets used during the current study are available as supplementary files, as well as the R codes used for the data analysis, according to the description given in the materials & Method section.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Marine Sciences, University of Otago, Dunedin, New Zealand. ²NIWA, University of Otago, Dunedin, New Zealand. ³Department of Functional & Evolutionary Ecology, Center of Functional Ecology, University of Vienna, Vienna, Austria. ⁴Manaaki Whenua, Landcare Research, Dunedin, New Zealand. ⁵Department of Microbiology & Immunology, University of Otago, Dunedin, New Zealand. ⁶Te Ao Tūroa, Te Rūnanga o Ngāi Tahu, Dunedin, New Zealand.

Received: 8 August 2019 Accepted: 18 July 2020

Published online: 13 August 2020

References

- Macdonald DW, Johnson DDP. Patchwork planet: the resource dispersion hypothesis, society, and the ecology of life. *J Zool* 3rd ed. 2015;295:75–107.
- Bandelj V, Solidoro C, Curiel D, Cossarini G, Melaku Canu D, Rismondo A. Fuzziness and heterogeneity of benthic Metacommunities in a complex transitional system. *Thrus S*, editor. *PLoS One*. 2012;7:e52395.
- Ebach MC, Parenti LR. The dichotomy of the modern bioregionalization revival. *J Biogeogr*. 2015;42:1801–8.
- Jiang J, DeAngelis DL, Teh S-Y, Krauss KW, Wang H, Li H, et al. Defining the next generation modeling of coastal ecotone dynamics in response to global change. *Ecol Model* Elsevier B.V. 2016;326:168–76.
- Cabral JS, Valente L, Hartig F. Mechanistic simulation models in macroecology and biogeography: state-of-art and prospects. *Ecography*. 2017;40:267–80.
- Rice J, Gjerde KM, Ardron J, Arico S, Cresswell I, Escobar E, et al. Policy relevance of biogeographic classification for conservation and management of marine biodiversity beyond national jurisdiction, and the GOODS biogeographic classification. *Ocean Coast Manag* Elsevier Ltd. 2011;54:110–22.
- Murray NJ, Keith DA, Bland LM, Ferrari R, Lyons MB, Lucas R, et al. The role of satellite remote sensing in structured ecosystem risk assessments. *Sci Total Environ* Elsevier B.V. 2017;619–620:249–57.
- Olivera I, Malhi Y. Many shades of green: the dynamic tropical forest–savannah transition zones. *Philos Trans Royal Soc B: Biol Sci*. 2016;371: 20150308.
- Barnes R, Hamylton S. On the very edge: faunal and functional responses to the interface between benthic seagrass and unvegetated sand assemblages. *Mar Ecol Prog Ser*. 2016;553:33–48.
- Plexida S, Solomou A, Poirazidis K, Sfougaris A. Factors affecting biodiversity in agrosylvopastoral ecosystems with in the Mediterranean Basin: a systematic review. *J Arid Environ*. 2018;151:125–33.
- Pauthenet E, Roquet F, Madec G, Nerini D. A linear decomposition of the Southern Ocean thermohaline structure. *J Phys Oceanogr*. 2017;47: 29–47.
- Kim YS, Orsi AH. On the variability of Antarctic circumpolar current fronts inferred from 1992–2011 altimetry. *J Phys Oceanogr*. 2014;44:3054–71.
- Minissale P, Sciandrello S. The sabulicolous therophytic associations in Sicily: new insights through the statistical approach, stressing the continuum versus discrete model of plant communities. *Acta Botanica Gallica*. 2015;162: 55–78.
- Whittaker RH. Gradient analysis of vegetation. *Biol Rev*. 1967;49:207–64.
- Yando ES, Osland MJ, Hester MW. Microspatial ecotone dynamics at a shifting range limit: plant–soil variation across salt marsh–mangrove interfaces. *Oecologia*. Springer. Berlin Heidelberg. 2018;187:319–31.
- Bearup D, Blasius B. Ecotone formation induced by the effects of tidal flooding: a conceptual model of the mud flat–coastal wetland ecosystem. *Ecol Complex* Elsevier B.V. 2017;32:217–27.
- Baltar F, Currie K, Stuck E, Roosa S, Morales SE. Oceanic fronts: transition zones for bacterioplankton community composition. *Environ Microbiol Rep*. 2016;8:132–8.
- Hernando-Morales V, Ameneiro J, Teira E. Water mass mixing shapes bacterial biogeography in a highly hydrodynamic region of the Southern Ocean. *Hallsworth JE*, editor. *Environ Microbiol* 2nd ed. 2017;19:1017–29.
- Juanes JA, Puente A, Ramos E. A global approach to hierarchical classification of coastal waters at different spatial scales: the NEA case. *J Marine Biol Assoc U K University of Otago Library*. 2017;97:465–76.
- Makowski C, Finkl CW, Vollmer HM. Geospatially integrated seafloor classification scheme (G-ISCS): a new method for cognitively interpreting benthic biogeomorphological features. *J Coast Res*. 2015;31: 488–504.
- Maxwell SM, Hazen EL, Lewison RL, Dunn DC, Bailey H, Bograd SJ, et al. Dynamic Ocean management: defining and conceptualizing real-time management of the ocean. *Mar Policy* Elsevier. 2015;58:42–50.
- Hopkins J, Shaw AGP, Challenor P. The Southland front, New Zealand: variability and ENSO correlations. *Cont Shelf Res* Elsevier. 2010;30:1535–48.
- Baziuké D, Juščenko N, Šiaulyš A. Switching to numerical scale in marine environmental decision support systems: fuzzy logic approach. *Ocean Coast Manag* Elsevier Ltd. 2014;101:35–41.
- Krug LA, Platt T, Sathyendranath S, Barbosa AB. Ocean surface partitioning strategies using ocean colour remote sensing: a review. *Prog Oceanogr* Elsevier Ltd. 2017;155:41–53.
- Müller O, Wilson B, Paulsen ML, Rumińska A, Armo HR, Bratbak G, et al. Spatiotemporal dynamics of ammonia-oxidizing Thaumarchaeota in distinct Arctic water masses. *Front Microbiol*. 2018;9:286.
- Guerrero-Feijóo E, Nieto-Cid M, Sintés E, Dobal-Amador V, Hernando-Morales V, Álvarez M, et al. Optical properties of dissolved organic matter relate to different depth-specific patterns of archaeal and bacterial community structure in the North Atlantic Ocean. *King G*, editor. *FEMS Microbiol Ecol*. 2017;93:fw224.
- Weiss S, Xu ZZ, Peddada S, Amir A, Bittinger K, Gonzalez A, et al. Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome* Microbiome. 2017;5:1–18.

28. Buchan A, LeClerc GR, Gulvik CA, González JM. Master recyclers: features and functions of bacteria associated with phytoplankton blooms. *Nat Rev Microbiol* Nature Publishing Group. 2014;12:686–98.
29. Ducklow HW, Kirchman DL, Quinby HL, Carlson CA, Dam HG. Stocks and dynamics of bacterioplankton carbon during the spring bloom in the eastern North Atlantic Ocean Pergamon press Ltd, editor. *Deep-Sea Research II*. 1993;40:245–63.
30. Azam F, Malfatti F. Microbial structuring of marine ecosystems. *Nat Rev Microbiol*. 2007;5:782–91.
31. Mioni CE, Pakulski JD, Poorvin L, Baldwin A, Twiss MR, Jeffrey WH, et al. Variability in the in situ bioavailability of Fe to bacterioplankton communities in the eastern subtropical Pacific Ocean. *Aquat Microb Ecol*. 2007;46:239–51.
32. Ramin M, Perhar G, Shimoda Y, Arhonditsis GB. Examination of the effects of nutrient regeneration mechanisms on plankton dynamics using aquatic biogeochemical modeling. *Ecol Model* Elsevier B.V. 2012;240:139–55.
33. Britten GL, Primeau FW. Biome-specific scaling of ocean productivity, temperature, and carbon export efficiency. *Geophys Res Lett*. 2016;43:5210–6.
34. Baltar F, Stegui JAI, Gasol JM, Lekanberri I, Herndl GJ. Mesoscale eddies: hotspots of prokaryotic activity and differential community structure in the ocean. *ISME J* Nature Publishing Group. 2010;4:975–88.
35. Chust G, Irigoien X, Chave J, Harris RP. Latitudinal phytoplankton distribution and the neutral theory of biodiversity. Duarte CM, editor. *Glob Ecol Biogeogr*. 2012;22:531–43.
36. Fondi M, Karkman A, Tamminen MV, Bosi E, Virta M, Fani R, et al. "Every gene is everywhere but the environment selects": global Geolocalization of gene sharing in environmental samples through network analysis. *Genome Biol Evol*. 2016;8:1388–400.
37. Djurhuus A, Boersch-Supan PH, Mikalsen S-O, Rogers AD. Microbe biogeography tracks water masses in a dynamic oceanic frontal system. *R Soc Open Sci*. 2017;4:170033.
38. Morales SE, Meyer M, Currie K, Baltar F. Are oceanic fronts ecotones? Seasonal changes along the subtropical front show fronts as bacterioplankton transition zones but not diversity hotspots. *Environ Microbiol Rep*. 2018;10:184–9.
39. Fortunato CS, Herfort L, Zuber P, Baptista AM, Crump BC. Spatial variability overwhelms seasonal patterns in bacterioplankton communities across a river to ocean gradient. *ISME J* Nature Publishing Group. 2012;6:554–63.
40. Doherty M, Yager PL, Moran MA, Coles VJ, Fortunato CS, Krusche AV, et al. Bacterial biogeography across the Amazon River-ocean continuum. *Front Microbiol*. 2017;8:1–17.
41. Hufkens K, Scheunders P, Ceulemans R. Ecotones in vegetation ecology: methodologies and definitions revisited. *Ecol Res*. 2009;24:977–86.
42. Ferro I, Morrone JJ. Biogeographical transition zones: a search for conceptual synthesis. *Biol J Linn Soc*. 2014;113:1–12.
43. Brownstein G, Döbert T, Dobbie L, Hashim N, Bastow WJ. Functional traits shed new light on the nature of ecotones: a study across a bog-to-forest sequence. *Community Ecol*. 2013;14:31–40.
44. Wiens JA, Crawford CS, Gosz JR. Boundary dynamics: a conceptual framework for studying landscape ecosystems Wiley, editor. *Oikos*. 1985;45:421–7.
45. Valentine JF, Heck KL Jr, Blackmon D, Goecker ME, Christian J, Kroutil RM, et al. Food web interactions along seagrass-coral reef boundaries: effects of piscivore reductions on cross-habitat energy exchange. *Mar Ecol Prog Ser*. 2007;333:37–50.
46. Zadeh LA. Fuzzy sets. *Infect Control*. 1965;8:338–53.
47. De Cáceres M, Font X, Oliva F. The management of vegetation classifications with fuzzy clustering. *J Veg Sci*. 2010;21:1138–51.
48. Roberts DW. Comparison of multidimensional fuzzy set ordination with CCA and DB-RDA. *Ecology*. 2009;90:2622–34.
49. Ejrnæs R. Can we trust gradients extracted by Detrended correspondence analysis? *J Veg Sci*. 2000;11:565–72.
50. Lasram FBR, Hattab T, Halouani G, Romdhane MS, Le Loc'h F. Modeling of Beta diversity in Tunisian waters: predictions using generalized dissimilarity modeling and Bioregionalisation using fuzzy clustering. Bianchi CN, editor. *PLoS One*. 2015;10:e0131728.
51. Olivero J, Marquez AL, Real R. Integrating fuzzy logic and statistics to improve the reliable delimitation of biogeographic regions and transition zones. *Syst Biol*. 2013;62:1–21.
52. Currie KI, Hunter KA. Seasonal variation of surface water CO₂ partial pressure in the Southland current, east of New Zealand. *Mar Freshw Res*. 1999;50:375.
53. Currie KI, Reid MR, Hunter KA. Interannual variability of carbon dioxide drawdown by subantarctic surface water near New Zealand. *Biogeochemistry*. 2011;104:23–34.
54. Baltar F, Stuck E, Morales S, Currie K. Bacterioplankton carbon cycling along the subtropical frontal zone off New Zealand. *Prog Oceanogr* Elsevier Ltd. 2015;135:168–75.
55. Jones KN, Currie KI, McGraw CM, Hunter KA. The effect of coastal processes on phytoplankton biomass and primary production within the near-shore subtropical frontal zone. *Estuar Coast Shelf Sci* Elsevier Ltd. 2013;124:44–55.
56. Shaw AGP, Vennell R. Measurements of an oceanic front using a front-following algorithm for AVHRR SST imagery. *Remote Sens Environ*. 2001;75:47–62.
57. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J* Nature Publishing Group. 2012;6:1621–4.
58. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yara P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res*. 2012;41:D590–6.
59. McMurdie PJ, Holmes S. Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data Watson M, editor. *PLoS One*. 2013;8:e61217.
60. The R Core Team. R: a language and environment for statistical computing [internet]. Vienna: R Foundation for Statistical Computing; 2019. p. 1–3695. Available from: <https://www.r-project.org>.
61. Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K. Cluster: cluster analysis basics and extensions. R package version 2.0.9; 2019. p. 1–82. Available from: <https://svn.r-project.org/R-packages/trunk/cluster>.
62. Dinno A. DunnTest: Dunn's test of multiple comparisons using rank sums. R package version 1.3.5; 2017. p. 1–7. Available from: <https://CRAN.R-project.org/package=dunn.test>.
63. Martiny JBH, Jones SE, Lennon JT, Martiny AC. Microbiomes in light of traits: a phylogenetic perspective. *Science*. 2015;350:aac9323.
64. Epskamp S, Cramer AOJ, Waldorp LJ, Schmittmann VD, Borsboom D. Qgraph: network visualizations of relationships in psychometric data. *J Stat Softw*. 2012;48:1–18.
65. Csárdi G, Nepusz T. The igraph software package for complex network research. *InterJ Complex Systems*. 2006;Complex Systems:1–9.
66. Kim D-W, Lee KH, Lee D. Fuzzy clustering of categorical data using fuzzy centroids. *Pattern Recogn Lett*. 2004;25:1263–71.
67. Cadenasso ML, Pickett STA, Grove JM. Dimensions of ecosystem complexity: heterogeneity, connectivity, and history. *Ecological Complexity*. 2006;3:1–12.
68. Ewers RM, Marsh CJ, Wearn OR. Making statistics biologically relevant in fragmented landscapes. *Trends Ecol Evol* Elsevier Ltd. 2010;25:699–704.
69. Lau MK, Borrett SR, Baiser B, Gotelli NJ, Ellison AM. Ecological network metrics: opportunities for synthesis. *Ecosphere* 2nd ed. 2017;8:e01900.
70. Hou W, Walz U. An integrated approach for landscape contrast analysis with particular consideration of small habitats and ecotones. *NC*. 2016;14:25–39.
71. Chapman CC. New perspectives on frontal variability in the Southern Ocean. *J Phys Oceanogr*. 2017;47:1151–68.
72. Chapman CC. Southern Ocean jets and how to find them: improving and comparing common jet detection methods. *J Geophys Res Oceans*. 2014; 119:4318–39.
73. Winter C, Matthews B, Suttle CA. Effects of environmental variation and spatial distance on bacteria, archaea and viruses in sub-polar and arctic waters. *Nat Publ Group*. 2013;7:1507–18.
74. Belanger CL, Jablonski D, Roy K, Berkea SK, Krug AZ, Valentine JW. Global environmental predictors of benthic marine biogeographic structure. *PNAS*. 2012;109:14046–51.
75. Reygondeau G, Longhurst A, Martinez E, BEAUGRAND G, Antoine D, Maury O. Dynamic biogeochemical provinces in the global ocean. *Global Biogeochem Cycles*. 2013;27:1046–58.
76. Mouw CB, Hardman-Mountford NJ, Alvain S, Bracher A, Brewin RJW, Bricaud A, et al. A Consumer's guide to satellite remote sensing of multiple phytoplankton groups in the Global Ocean. *Front Mar Sci*. 2017;4:1–19.
77. Roy S. Distributions of phytoplankton carbohydrate, protein and lipid in the world oceans from satellite ocean colour. *ISME J* Springer US. 2018;12:1457–72.
78. Bachmann J, Heimbach T, Hassenrück C, Kopprio GA, Iversen MH, Grossart HP, et al. Environmental drivers of free-living vs. particle-attached bacterial community composition in the Mauritania upwelling system. *Front Microbiol*. 2018;9:1–13.
79. Yilmaz P, Yara P, Rapp JZ, Glöckner FO. Expanding the world of marine bacterial and archaeal clades. *Front Microbiol*. 2016;6:788.

80. Vojvoda J, Lamy D, Sintés E, García JAL, Turk V, Herndl GJ. Seasonal variation in marine-snow-associated and ambient-water prokaryotic communities in the northern Adriatic Sea. *Aquat Microb Ecol*. 2014;73:211–24.
81. Kjørboe T, Ploug H, Thygesen UH. Fluid motion and solute distribution around sinking aggregates. I. Small-scale fluxes and heterogeneity of nutrients in the pelagic environment. *Mar Ecol Prog Ser*. 2001;211:1–13.
82. Solden L, Lloyd K, Wrighton K. The bright side of microbial dark matter: lessons learned from the uncultivated majority. *Curr Opin Microbiol* Elsevier Ltd. 2016;31:217–26.
83. Louca S, Jacques SMS, Pires APF, Leal JS, Srivastava DS, Parfrey LW, et al. High taxonomic variability despite stable functional structure across microbial communities. *Nature Ecol Evol* Nature Publishing Group. 2016;1:1–12.
84. Yeh CF, Soininen J, Teittinen A, Wang J. Elevational patterns and hierarchical determinants of biodiversity across microbial taxonomic scales. *Mol Ecol*. 2018;28:86–99.
85. Ellison AM. Foundation species, non-trophic interactions, and the value of being common. *ISCIENCE* Elsevier Inc. 2019;13:254–68.
86. He X, McLean JS, Edlund A, Yooshep S, Hall AP, Liu S-Y, et al. Cultivation of a human-associated TM7 phylotype reveals a reduced genome and epibiotic parasitic lifestyle. *PNAS*. 2015;112:244–9.
87. Gong J, Qing Y, Guo X, Warren A. "Candidatus *Sonnebornia yantaiensis*," a member of candidate division OD1, as intracellular bacteria of the ciliated protist *Paramecium bursaria* (Ciliophora, Oligohymenophorea). *Syst Appl Microbiol* Elsevier GmbH. 2014;37:35–41.
88. Agogue H, Lamy D, Neal PR, Sogin ML, Herndl GJ. Water mass-specificity of bacterial communities in the North Atlantic revealed by massively parallel sequencing. *Mol Ecol*. 2011;20:258–74.
89. Mapelli F, Varela MM, Barbato M, Alvaríño R, Fusi M, Álvarez M, et al. Biogeography of planktonic bacterial communities across the whole Mediterranean Sea. *Ocean Sci*. 2013;9:585–95.
90. Baselga A, Leprieur F. Comparing methods to separate components of beta diversity. *Spencer M, editor. Methods Ecol Evol*. 2015;6:1069–79.
91. Jost L. Partitioning diversity into independent alpha and Beta components. *Ecology*. 2007;88:2427–39.
92. Lentendu G, Mahé F, Bass D, Rueckert S, Stoeck T, Dunthorn M. Consistent patterns of high alpha and low beta diversity in tropical parasitic and free-living protists. *Mol Ecol* 2nd ed. 2018;27:2846–57.
93. Sauterey B, Ward BA, Follows MJ, Bowler C, Claessen D. When everything is not everywhere but species evolve: an alternative method to model adaptive properties of marine ecosystems. *J Plankton Res*. 2015;37:28–47.
94. Haywood GJ. Some effects of river discharges and currents on phytoplankton in the sea off Otago, New Zealand. *N Z J Mar Freshw Res*. 2004;38:103–14.
95. Stevens CL, O'Callaghan JM, Chiswell SM, Hadfield MG. Physical oceanography of New Zealand/Aotearoa shelf seas – a review. *N Z J Mar Freshw Res*. 2019;1–40. <https://doi.org/10.1080/00288330.2019.1588746>.
96. DeLong EF, Preston CM, Mincer T, Rich V, Hallam SJ, Frigaard N-U, et al. Community genomics among stratified microbial assemblages in the Ocean's interior. *Science*. 2006;311:496–503.
97. Sutton PJH. The Southland current: a subantarctic current. *N Z J Mar Freshw Res*. 2003;37:645–52.
98. Kim H, Lee DE, Ducklow HW. Mixing regime-dependent causality between phytoplankton and bacteria in the subtropical North Atlantic Ocean ecosystem. *Mar Ecol Prog Ser*. 2018;600:41–53.
99. Zhou J, Richlen ML, Sehein TR, Kulis DM, Anderson DM, Cai Z. Microbial community structure and associations during a marine dinoflagellate bloom. *Front Microbiol*. 2018;9:1–21.
100. Tölgyesi C, Zalatnai M, Erdős L, Bátorfi Z, Hupp NR, Körmöczy L. Unexpected ecotone dynamics of a sand dune vegetation complex following water table decline. *JPECOL*. 2016;72:rtv032.
101. Chafee M, Fernández-Guerra A, Buttigieg PL, Gerdts G, Eren AM, Teeling H, et al. Recurrent patterns of microdiversity in a temperate coastal marine environment. *ISME J* Nature Publishing Group. 2018;12:237–52.
102. Zamon JE, Phillips EM, Guy TJ. Marine bird aggregations associated with the tidally-driven plume and plume fronts of the Columbia River. *Deep-Sea Res II* Elsevier. 2014;107:85–95.
103. Yeleğçi Ö, Charria G, Capet X, Reverdin G, Sudre J, Yahia H. Spatial and seasonal distributions of frontal activity over the French continental shelf in the Bay of Biscay. *Cont Shelf Res* Elsevier Ltd. 2017;144:65–79.
104. Albaina A, Irigoien X. Fine scale zooplankton distribution in the Bay of Biscay in spring 2004. *J Plankton Res*. 2007;29:851–70.
105. Raveh O, David N, Rilov G, Rahav E. The temporal dynamics of coastal phytoplankton and Bacterioplankton in the eastern Mediterranean Sea. *Dam HG, editor. PLoS One*. 2015;10:e0140690.
106. Kavanaugh MT, Hales B, Saraceno M, Spitz YH, White AE, Letelier RM. Hierarchical and dynamic seascapes: a quantitative framework for scaling pelagic biogeochemistry and ecology. *Prog Oceanogr* Elsevier Ltd. 2014;120:291–304.
107. Hanson BT, Hewson I, Madsen EL. Metaproteomic survey of six aquatic habitats: discovering the identities of microbial populations active in biogeochemical cycling. *Microb Ecol*. 2014;67:520–39.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

