Microbial diversity through an oceanographic lens: refining the concept of ocean provinces through trophic-level analysis and productivity-specific length scales

Cora Hörstmann 1,2*, Pier Luigi Buttigieg 3, Uwe John 1,4 Eric J. Raes 5, Dieter Wolf-Gladrow 1, Astrid Bracher 1,6 and Anya M. Waite 5
1 Alfred Wegener Institute Helmholtz Center for Polar and Marine Science, Bremerhaven, Germany.
2 Department of Life Sciences and Chemistry, Jacobs University gGmbH, Bremen, Germany.
3 Helmholtz Metadata Collaboration, GEOMAR, Kiel, Germany.
4 Helmholtz Institute for Functional Marine Biodiversity, Oldenburg, Germany.
5 Ocean Frontier Institute and Department of Oceanography, Dalhousie University, Halifax, NS, Canada.
6 Institute of Environmental Physics, University of Bremen, Bremen, Germany.

Summary
In the marine realm, microorganisms are responsible for the bulk of primary production, thereby sustaining marine life across all trophic levels. Longhurst provinces have distinct microbial fingerprints; however, little is known about how microbial diversity and primary productivity change at finer spatial scales. Here, we sampled the Atlantic Ocean from south to north (~50°S–50°N), every ~0.5° latitude. We conducted measurements of primary productivity, chlorophyll-a and relative abundance of 16S and 18S rRNA genes, alongside analyses of the physicochemical and hydrographic environment. We analysed the diversity of autotrophs, mixotrophs and heterotrophs, and noted distinct patterns among these guilds across provinces with high and low chlorophyll-a conditions. Eukaryotic autotrophs and prokaryotic heterotrophs showed a shared interprovince diversity pattern, distinct from the diversity pattern shared by mixotrophs, cyanobacteria and eukaryotic heterotrophs. Additionally, we calculated samplewise productivity-specific length scales, the potential horizontal displacement of microbial communities by surface currents to an intrinsic biological rate (here, specific primary productivity). This scale provides key context for our trophically disaggregated diversity analysis that we could relate to underlying oceanographic features. We integrate this element to provide more nuanced insights into the mosaic-like nature of microbial provincialism, linking diversity patterns to oceanographic transport through primary production.

Introduction
The continuous movement of seawater and turnover of microbial biomass and diversity in marine ecosystems form a time-varying mosaic of phylogenetic and functional biodiversity across ocean basins. This complicates efforts to map, understand and monitor key marine ecosystem attributes such as microbial growth, biodiversity and carbon cycling (Stec et al., 2017).

Marine ecosystems comprise a wide array of microbial life: microbial (photo-)autotrophic, pro- and eukaryotes, form the base of the marine food web (Hutchins et al., 2015; Sunagawa et al., 2015), and sustain energy exchange, provision and recycling of resources (Falkowski et al., 2008; Guidi et al., 2016) for higher trophic levels. Heterotrophs and mixotrophs remineralize most of the carbon and nutrients from the primary production via the microbial loop, before these can be exported to the deep sea (Azam et al., 1983; Azam and Malfatti, 2007).

In terrestrial systems, primary producer communities have been used to define major biomes (Woodward, 1987; Woodward et al., 2004), biogeographic realms and ecoregions (Olson et al., 2001) through their physical and functional structuring of ecosystems (Cardinale et al., 2011). In the ocean, Longhurst (2007) used chlorophyll-a (chl a) concentrations as a proxy for
phytoplankton biomass to delimit ocean provinces, alongside water temperature to distinguish water masses. Longhurst provinces are used to define oceanographic biogeographic subdivisions; however, static applications of these provinces typically overlook the dynamic interactions, life histories, endemism and/or vicariance within ecological assemblages, needed to truly map microbial biogeography. Thus, multiple variables including biomass, primary productivity (PP), and diversity need to be considered with carefully structured sampling across space and time (Kollmann et al., 2016). Here, we investigate how integrating multiple physical and biochemical variables – accounting for their horizontal displacement by surface currents – can improve our understanding of microbial provincialization across a high-resolution transect (~0.5° latitude) of the Atlantic Ocean.

Microbial assemblages can act as ‘fingerprints’ for water masses thanks to their high diversity, responsiveness and (typically) fast generation time (hours to days; Martiny et al., 2006), aiding the definition of biogeographic boundaries in pelagic environments (Raes et al., 2011; Fuhrman et al., 2015). However, the dynamic nature of the oceans – with their fronts, currents, eddies, up- and downwellings, and other hydrographic features – has the potential to add additional complexity to the ecological variation between and within such regions (Oliver and Irwin, 2008; Hernando-Morales et al., 2017).

Hydrographic features create structural variability in the ocean, which (through, e.g. modifying nutrient distributions) provides more diverse niche space for microbial communities to colonize. Unless mitigated, this would increase microbial diversity (e.g. Kemp and Mitsch, 1979; Cadotte, 2006) and should favor phytoplankton productivity (Legendre, 1981). While basin-scale horizontal dispersal of organisms by major ocean currents is known to reduce microbial β-diversity (Raes et al., 2011; Richter et al., 2020; Sommeria-Klein et al., 2021), mesoscale and sub-mesoscale horizontal transport of microbial communities, and the impact on their activity, is rarely taken into account, except in a few frameworks that couple ecology and hydrography (e.g. the ‘dual-lens’ approach; Oldham et al., 2013). Indeed, hydrographic dynamics within and between ocean provinces interact with both neutral and selective ecological processes, resulting in communities in different successional states (Zhou et al., 2014) and/or shaped by opportunistic responses (e.g. Duffy and Stachowicz, 2006; Hartmann et al., 2012; Fadeev et al., 2021).

Here, we assess the relative importance of regional water mass characteristics (physicochemical parameters and hydrography) on microbial diversity. Furthermore, we resolve diversity responses along major trophic groups (auto-, mixo-, heterotrophs), an important but understudied perspective in microbial ecology (reviewed in Seibold et al., 2018). We leverage a conceptual framework that asserts that microbial communities are distinct within oceanographic regions, separated by fronts and currents limiting microbial dispersal (Martiny et al., 2006; Milici et al., 2016; Raes et al., 2018). We then assess how ecosystem structure (i.e. hydrography), bottom-up and top-down factors can qualify and advance traditional partitioning of the Atlantic into biogeographic provinces.

Results

Delineation of oceanographic, ecosystemic provinces

Our analyses of remote sensing observations of geostrophic currents, sea surface temperature and chl a combined with in situ measurements of oceanographic features, chl a and microbial diversity informed our determination of ecosystem boundaries shown in Fig. 1. The ecosystem boundaries broadly overlapped with the Longhurst Provinces and we thus maintain the same naming conventions (Fig. S1). One exception was the South Atlantic Gyral (SATL) province, wherein sites were clearly ordinated into separate groups; one with cool waters (COLD; 20.4–23.3°C) in the Argentine Basin, south of the Rio Grande rise and northern, oligotrophic warmer waters (HOT; 25.1–27.5) in the Brazil Basin (Figs S2 and S3). Furthermore, Stations 24 and 92 were outliers with respect to the temperature-salinity plot used for identifying ecosystem boundaries (Fig. S2), and were thus not grouped in an ocean province.

On a broad scale, we could separate ocean provinces into provinces with low and high chl a with significant differences in chl a concentrations (Wilcoxon, \( p = 9.112 \times 10^{-6}, n_1 = 38, n_2 = 39 \)), and significant differences in PP \( ^* \) (Wilcoxon, \( p = 3.686 \times 10^{-5}, n_1 = 38, n_2 = 39 \); Table 1). In our principal component analysis (PCA) (Fig. S4), sites in the Southwest Atlantic Shelves province (FKLD), Brazil Current Coast province (BRAZ), Canary Current Coast province (CNRY), North Atlantic Subtropical Gyre province (NAST-E) and North Atlantic Drift province (NADR) (i.e. high-chl a) provinces were associated with high particulate organic matter and high dissolved inorganic nutrient concentrations. Furthermore, the SATL-COLD, SATL-HOT, Western Tropical Atlantic province (WTTRA) and North Atlantic Tropical Gyre province (NATR) (i.e. low-chl a provinces) provinces were associated with low temperatures and larger distance to coast (>600 km; Fig. S4).

Province PP and chl a concentrations correlate across the transect (Pearson correlation; \( r = 0.48, p = 6.68 \times 10^{-6}, n = 80 \)), but did not always correlate within provinces (e.g. WTTRA: \( r = 0.08, p = 0.8, n = 11 \)). We measured the highest PP in the CNRY province, a peak in chl a concentrations relative to adjacent ocean provinces.
Table 1), and total cell numbers of 1207 ± 737 cells μl⁻¹ (mean ± SD, n = 30; Table S1).

Biomass turnover and transport of pelagic microbial communities

In the following analyses, we used a quantitative comparison of biomass turnover (specific PP; \(P^β\)) and current speed (see Methods, Eqs. 3 and 4) to estimate the distance a microbial community has travelled through passive advection before half of its biomass has been turned over: the productivity-specific length scale (Fig. 2A).

Based on the biological and physical component of the productivity-specific length scale, we observed strong sample variations in both variables [\(P^β\) based on PP per day (d) and chl a, Table 1, and surface current speed] within and between provinces (Fig. 2B). For example, the SATL-HOT province had high variations in \(P^β\) (0.007–0.07 mg C m⁻³ d⁻¹, n = 15) with relatively constant horizontal current speed (1–3 m d⁻¹). In contrast, we measured a large range of horizontal current speed (2–7 m d⁻¹, n = 4) at relatively low \(P^β\) (0.002–0.01 mg C m⁻³ d⁻¹, n = 4) within the BRAZ province resulting in the largest productivity-specific length scales (up to 22 km) of our dataset.

We used our calculations of sample-specific productivity-specific length scales to examine variabilities in beta diversity (β-diversity) between neighbouring sites.
Table 1. Physico-chemical properties, primary productivity, Chl a concentrations and productivity-specific length scale of ocean provinces during the PS13 expedition. Provinces are separated into high-Chl a provinces and low-Chl a provinces.

<table>
<thead>
<tr>
<th>Province</th>
<th>H-CHL</th>
<th>L-CHL</th>
<th>α-div</th>
<th>β-div</th>
<th>Phosphate (µM)</th>
<th>Nitrate (µM)</th>
<th>POC (nmól L⁻¹)</th>
<th>PN (nmól L⁻¹)</th>
<th>n</th>
<th>Range</th>
<th>Mean ± SD</th>
<th>Chl a (mg m⁻² h⁻¹)</th>
<th>PP (µmol C m⁻² h⁻¹)</th>
<th>Range</th>
<th>Mean ± SD</th>
<th>PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRZ</td>
<td>1.70</td>
<td>0.86</td>
<td>1.22</td>
<td>0.72</td>
<td>0.44</td>
<td>0.32</td>
<td>0.18</td>
<td>0.13</td>
<td>4</td>
<td>3.2</td>
<td>0.09 ± 0.07</td>
<td>0.04 ± 0.07</td>
<td>2.2 ± 2.1</td>
<td>108.1</td>
<td>65.77</td>
<td>0.81</td>
</tr>
<tr>
<td>SATL-COLD</td>
<td>29.14</td>
<td>28.33</td>
<td>11.78</td>
<td>10.01</td>
<td>0.06</td>
<td>0.02</td>
<td>0.02</td>
<td>0.04</td>
<td>9</td>
<td>2.1</td>
<td>0.08 ± 0.06</td>
<td>0.07 ± 0.06</td>
<td>1.2 ± 1.1</td>
<td>268.4</td>
<td>142.75</td>
<td>0.14</td>
</tr>
</tbody>
</table>

Alpha diversity (α-diversity) across functional groups and oceanographic provinces

Rarefaction curves for both prokaryotes and eukaryotes saturated in all samples (Fig. S5). Prokaryotic Shannon diversity was highest (260.1) at Station 11 (BRAZ province) and lowest (59) at Station 97 (WTRA province; Table S2). Eukaryotic Shannon diversity was highest (766.8) at Station 11 (BRAZ province) and lowest (11.2) at Station 8 (BRAZ province; Table S2).

We observed that α-diversity varied across trophic groups, provinces and latitude (Fig. 3A). The α-diversity of both eukaryotic and prokaryotic autotrophs and eukaryotic mixotrophs was lower than that of heterotrophs. Relative to other functional groups, heterotrophic prokaryotes had higher α-diversity in the Southernmost and northernmost provinces (FKLD, BRAZ, NADR). Shannon diversity increased towards the equator and was most pronounced for heterotrophic pro- and eukaryotes (Fig. 3A) with the highest α-diversity in the SATL-HOT, SATL-COLD and WTRA provinces (Table S2); the warmest and most oligotrophic provinces sampled. Similarly, the low-chl a provinces had significantly higher mixotrophic α-diversity than the high-chl a provinces (Wilcoxon, p < 2.2e⁻¹⁶, n₁ = 39, n₂ = 68).

In relation to the physical–chemical parameters, α-diversity was positively and linearly correlated with physical variables (temperature and salinity), and negatively correlated with nutrient and biochemical concentrations (NO₃, PO₄, POC, PN, Chl a) across ocean provinces (Fig. 3B). The correlation between temperature and α-diversity was significant for autotrophic eukaryotes, mixotrophs and heterotrophic prokaryotes.
Table S3). POC, PN and chl $a$ were significantly negatively correlated with the $\alpha$-diversity of autotrophic eukaryotes. The $\alpha$-diversity of heterotrophic prokaryotes was negatively correlated with NO$_3$, PO$_4$, POC, PN (Fig. 3B; Table S3). However, within provinces, we found both positive and negative correlations of functional groups with temperature, nutrients and particulate matter (e.g. PN correlated positively with cyanobacterial diversity in the SATL-COLD province: $r = 0.8$, $p = 3.85 \times 10^{-4}$, $n = 15$, Fig. 3B).

PP had no significant correlations with picoplankton cell number or microbial $\alpha$-diversity across any trophic functional group (Table S3). However, in the CNRY province, where productivity rates were significantly higher than in other provinces (Wilcoxon, $p$-value $= 5.75 \times 10^{-4}$, $n_1 = 17$, $n_2 = 77$), diversity of autotrophic eukaryotes and cyanobacteria were negatively correlated with PP ($r = -0.6$, $p = 0.04$, $n = 14$, and $r = -0.6$, $p = 0.01$, $n = 14$ respectively; Fig. 3B).

We noted that $\alpha$-diversity of trophic functional groups are similarly structured in their correlations with environmental (physical + chemical) parameters; however, the magnitudes and significances of the correlations varied between groups (e.g. NAST-E PO$_4$, auto.cyano; $r = 0.83$, $p = 0.01$; PO$_4$, het.prok, $r = 0.39$, $p = 0.27$, Figs S6 and S7), as did the distribution of residuals (Figs S8–S15). Between functional groups, we observed a significant positive correlation between the $\alpha$-diversity of autotrophic eukaryotes and cyanobacteria in the CNRY provinces or between autotrophic eukaryotes and mixotrophs in the CNRY and NADR provinces as well as across the transect (Fig. 3B, S16).
We observed that each province has a distinct correlation pattern across the variables we examined (Fig. 3B). Correlations were more pronounced in the CNRY province (Fig. 3B). We could not identify pairs of parameters with correlations that were consistently repeated across provinces (e.g. temperature, auto.euk SATL-COLD; temperature, auto.euk SATL-HOT; Fig. 3B). Furthermore, we noted that samples were not evenly spread along bivariate plots, especially in physically energetic regions (i.e. large scatter of temperature – salinity profile; Fig. S17).

**Beta diversity (β-diversity) patterns of auto-, mixo- and heterotrophs across provinces**

Sites belonging to high-chl a and low-chl a provinces were well separated along the first axes of our redundancy analysis (RDA) plots for each trophic functional
group (Fig. 4; between 40% and 61% of variance constrained). Sites belonging to high-chl a provinces were associated with higher dissolved inorganic nutrient concentrations, particulate organic matter and chl a concentrations. Sites belonging to low-chl a provinces were associated with higher temperatures, salinity and larger distances to province boundaries and the coast.

We found that sites in low-chl a and high-chl a provinces were differentially ordinated in our RDA biplots. We observed the shortest distances between sites in low-chl a provinces when analysing autotrophic euakaryotic diversity. In contrast, heterotrophic euakaryotic and mixotrophic diversity led to sites in low-chl a provinces being ordinated furthest apart from those in other provinces (Table S4). While autotrophic euakaryotic communities were distinctly separated between the high-chl a provinces, mixotrophic communities had short relative ordination distances or overlapped in these provinces (Fig. 4A and C; Table S4). Across all trophic groups, the samples from the CNRY province showed the largest spread along the second RDA axes and partially overlapped with samples from the NAST-E province (Fig. 4).

Prokaryotic heterotrophic communities were more similar to autotrophic euakaryotes in their pattern of the biplots (procrustes analysis, Table S5), with shorter relative distances in low-chl a than in high-chl a provinces (Fig. 4A and E; Table S4). In contrast, cyanobacteria, heterotrophic and mixotrophic euakaryotes were more similar to each other (procrustes analysis, Table S5); shorter relative distances in high-chl a than in low-chl a provinces (Fig. 4; Table S4).

Analysis of the $\beta$-diversity among communities of heterotrophic pro- and euakaryotes, and mixotrophs originated station 24 between those that are characteristic of the BRAZ province and the southern, cold part of the SATL-COLD province (Fig. 4C–E). Similarly, analysis of $\beta$-diversity of the autotrophic euakaryotes and mixotrophs at Station 92 (located between the SATL-HOT and the WTRA province) ordinated the station between the SATL-HOT and WTRA (Fig. 4A and C).

The SATL province, as defined by Longhurst (2007), clustered into two distinct groups in microbial $\beta$-diversity (Fig. 4) which was also apparent in our biogeochemical inspection and division into a southern dynamic (COLD) part, and a northern, oligotrophic (HOT) part in the Brazil Basin (Fig. 4).

Discussion

Our study revealed that microbial trophic groups showed variable $\alpha$- and $\beta$-diversity patterns in relation to physical and biogeochemical environmental parameters across our south-to-north transect. Below, we discuss how our integrative results can contribute to refining microbial, biogeographic provincialism. Our results provide an estimate of sample-based spatial scales of differences in biodiversity signals, suggesting that observations of ecosystem function and stability need regional high-resolution sampling.

In situ measurements of province characteristics deviate from Longhurst provinces. Broadly, traditional Longhurst province boundaries (Longhurst, 2007) are determined by clustering remote sensing data on chl a and temperature data (e.g. Devred et al., 2007; Hardman-Mountford et al., 2008) or in situ measurements of phytoplankton pigment composition (Barlow et al., 2007; Taylor et al., 2011; Bracher et al., 2020). Our delineation largely overlapped (for about 80% of the stations) with the provinces delineated by Bracher et al., (2020) using hierarchical cluster analysis on phytoplankton group composition data derived from the HPLC marker pigments (Table S6). However, we observed regional divergence. Most pronounced was the separation of the SATL province into SATL-HOT and SATL-COLD; supported by our physical, chemical and microbial observations. This difference could not be detected in phytoplankton pigment-based analyses (see Bracher et al., 2020). Such observations (if confirmed) exemplify how new provincialisation can arise from microbial oceanographic perspectives.

However, we also noted that some of our in situ measurements could lead to spurious distinctions within provinces. For example, we occasionally measured high nutrient and chl a concentrations in provinces known to be oligotrophic (Longhurst, 2007; e.g. up to 0.43 mg m$^{-3}$ chl a in the NATR province). Given our sampling regime, this may simply be the result of sampling a transient patch of high-nutrient/high-chl a water. Our $\beta$-diversity supports this possibility, as NATR sites were largely ordinated with sites from other oligotrophic provinces along the first major axis in each of our RDA triplots (Fig. 4). We note that this shows the value of microbial diversity data in holistically describing oceanographic provinces, especially when sampling dynamic regions.

Ocean currents can drive overlaps in $\beta$-diversity signals: For example, ordinations showed that sites from the NAST-E and CNRY province overlapped (Fig. 4), likely caused by transport and dispersal along the Azores Current which joins the Canary Current (Fedoseev, 1970; Fig. S2a). The complex dynamics in the CNRY province (driven, for example, by a weakening of the Canary Current through coastal upwelling and eddy formations as described in accompanying studies of the same sampling campaign; (Bracher et al., 2020; von Appen et al., 2020) likely contributed to increased microbial $\beta$-diversity, relative to most other provinces. Similarly, samples in physically energetic provinces (e.g. BRAZ province), defined by an increased scatter in their temperature-salinity...
profiles (Figs S2 and S17), were more distributed in their α-diversity suggesting that the habitat structure is more disrupted or patchy in these ecosystems and high-resolution sampling (<50 km) is needed for stable biodiversity assessments as shown in other studies across ocean filaments (Fadeev et al., 2021) and fronts (Mousing et al., 2016).

Productivity-specific length scales and their role in microbial biogeography. The ratio of PP rate and current speed gives us a productivity-specific length over which a biological community is carried during a single biomass turnover. This can be interpreted as a scale relating the potential for observed changes in β-diversity to horizontal transport. This may be seen as a time-sensitive version of earlier relationships described between chl a patchiness and current speed (Powell et al., 1975). Specifically, we estimated the PP rate (normalized $P_P$) at each sampling site through 13C stable isotope experiments and biomass assessment (chl a concentration).

Fig. 4. Redundancy analysis (RDA) of CLR-transformed ASV counts, subsequently partitioned into separate tables for each functional group (A) autotrophic eukaryotes, (B) autotrophic prokaryotes (C) eukaryotic mixotrophs, (D) heterotrophic eukaryotes and (E) heterotrophic prokaryotes. Contextual spatial and environmental data (z-scored) were used as explanatory variables, represented as arrows. Samples cluster according to ocean provinces. Samples separate between low-chl a provinces (SATL-COLD, SATL-HOT, WTRA, NATR) (indicated with light grey star) and high-chl a provinces (FKLD, BRAZ, CNRY, NAST-E, NADR) (indicated with dark grey star) along the first axis (RDA 1) across all functional groups.
estimate of the spatial scale of ecosystem patchiness in epipelagic systems, ultimately controlled by a key biological rate (PP) and horizontal current speed.

Across the Atlantic, we identified three regimes of β-diversity associated with different magnitudes of productivity-specific length scales [high diversity at low productivity-specific length scales (Fig. 2C Box 1), low diversity at low productivity-specific length scales (Fig. 2C Box 2), and increasing diversity with increasing productivity-specific length scales (Fig. 2C Box 3)], affected by different magnitudes of $P^\beta$ and current speed within different provinces (Fig. 2B).

For example, prokaryotic growth in oligotrophic regions with low current speed, such as the SATL-HOT province, has been shown to be controlled by heterotrophic grazing (Teira et al., 2019). In our data, the change in β-diversity within the SATL-HOT province occurred at intermediate productivity-specific length scales with higher $P^\beta$, higher proportional heterotrophic biomass and higher eukaryotic heterotrophic richness relative to other sites (Fig. 2B and C – box 2, Tables S1 and S2). This can suggest that differences in productivity-specific length scales are a result of greater biological top-down control on local biodiversity than differences in horizontal current speed. The β-diversity signals in provinces with high horizontal current speed increased with increasing productivity-specific length scales (Fig. 2C Box 3 SATL_COLD, BRAZ), suggesting a possible shift in mechanisms for β-diversity at higher productivity-specific length scales. The observed increase in β-diversity was likely due to the variable local energetics of the Brazilian Coastal Current (De Souza and Robinson, 2004); reflected in a large range of temperature-salinity signals in our dataset (Fig. S2). These energetics can result in a confluence zone of communities potentially far from their region of origin (Malvinas-Brazil Confluence; Clayton et al., 2013). However, at this point, we cannot differentiate between the physical and biological mechanisms as we also observed different magnitudes of ordination distances within different trophic functional groups. For most of our sites (Fig. 2C, Box 2), the length scale was <5 km and did not correlate with the relatively low β-diversity signals in our RDA analyses (<0.05 ordination distance). In this case, β-diversity can be controlled by one or more of the multiple environmental variables shaping microbial communities such as vertical mixing (Cheng et al., 2020), atmospheric processes (Mayol et al., 2014), macroorganisms (Troussellier et al., 2017) and anthropogenic impact (Nogales et al., 2011).

It is important to note that at this point, we discuss surface dynamics only, and do not include any consideration of vertical mixing, itself clearly an important feature in generating diversity changes (Delong et al., 2006), and likely the source for the great heterogeneity of β-diversity in the CNRY province (Fig. 2C – Box 1). Understanding the interplay of biological and physical controls would require more complete analyses of biological and physical interactions and is thus only a first-order estimate.

Essentially, the productivity-specific length scale can be seen as defining a relevant scale of the mosaic of the biodiversity signatures measured in this study, revealing where higher-resolution sampling (<5 km) would have added more value to our data set and thus where more careful interpretations of our β-diversity results are necessary. It also provides a perspective on the resolution future campaigns should adopt to develop biogeographic understandings in each region.

We are convinced that Lagrangian sampling – at scales informed by sample-based productivity-specific length scales – could enhance our understanding of microbial pelagic biogeography. This approach would help characterize how far microbial communities in a given water packet could disperse within provinces boundaries, giving shape to their internal mosaics of biodiversity. This is analogous to how the Damköhler number relates exposure timescale and processing timescale of a chemical reaction, to express how spatially and temporally extended the impacts of that reaction will be felt (Oldham et al., 2013).

Province-dependent correlation structures between α-diversity and environmental variables. Across our entire transect, correlations between microbial α-diversity and environmental variables followed well-described latitudinal temperature–diversity relationships (e.g. Fuhrman et al., 2008; Ibarbalz et al., 2019). However, we observed more faceted environment–diversity relationships within provinces, especially within physically energetic provinces (Fig. 3B). More measured variables, including temperature, correlated with α-diversity in energetic provinces relative to those with less energetic profiles. This suggests that the temperature–diversity relationship is not controlled by thermal energy alone (see, e.g. Giebel et al., 2009) but is nested in a more faceted microbial response to local water mass characteristics. For example, the high-temperature provinces we sampled were also oligotrophic (low chl a), which would confound any attempt to independently assess temperature–diversity relationships without accounting for the impact of oligotrophic conditions on α-diversity (Fig. 3A; Santi et al., 2019).

Our results showed a nuanced relationship of PP to microbial α-diversity within provinces and do not corroborate previous observations, which detected significant positive correlations (Raes et al., 2018). We observed moderate positive (but not significant at $n = 20$) correlations of PP and prokaryotic heterotrophic α-diversity in low-chl a provinces. This may be due to more
oligotrophic conditions supporting more even saturation of available niches, rather than boom-and-bust dynamics characteristic of eutrophic environments and events (Duffy and Stachowicz, 2006). In contrast, we observed a negative and significant correlation between α-diversity and PP in the highly productive CNRY province, suggesting the rise and succession (via invisibility) of a few, opportunistic phylotypes (Steiner and Leibold, 2004), due to input of limiting nutrients from Saharan dust (von Appen et al., 2020). Together, our observations in the Atlantic and those in the Pacific (Raes et al., 2018) suggest that the relationship between PP and microbial α-diversity is nuanced, and observed signals depend on province-specific characteristics, which drive competition.

Notably, the differences in sampling size (n_min = 8, n_max = 20) of different provinces impact individual correlation and correlation significance. Thus, we excluded the FKLD (n = 4) and NATR (n = 6) province from our correlation analysis. The great variability in significant correlations between environmental variables and microbial diversity showed that refined observations within provinces are needed to confirm the observed individual environment–diversity relationships.

Low and high chl a conditions correspond to contrasting diversity patterns between trophic groups. We observed contrasting β-diversity patterns in microbial functional groups associated with low- and high-chl a conditions, supporting the observations by Irwin et al. (2006) of differences in the importance of environmental predictors for different phytoplankton functional types. Our results show that environmental conditions are important predictors not only on a phylogenetic level (e.g. diatoms vs. dinoflagellates, Irwin et al., 2006) but also within trophic functional groups.

We observed that eukaryotic mixotrophic communities had relatively high α-diversity at each site and high β-diversity between sites in low-chl a provinces (vs. high-chl a provinces), suggesting that mixotrophy is supported under low-nutrient, low chl a conditions (see, e.g. Hartmann et al., 2012). Similarly, the β-diversity of cyanobacteria and eukaryotic heterotrophs was greater in sites under low chl a conditions, suggesting these functional groups are more susceptible to selective forces in these provinces (e.g. increased cyanobacterial diversity under low phosphate conditions; Thompson et al., 2013).

In contrast, we observed that there is greater β-diversity between sites across high-chl a provinces (vs. low-chl a provinces) when examining eukaryotic autotrophic and prokaryotic heterotrophic communities (Fig. 4A and E), which are known to structure one another (Seymour et al., 2017). However, we did not observe any proportional change in the α-diversity of these groups between low- and high-chl a provinces. This suggests that higher productivity is contributing more to species turnover between sites, rather than greater diversity within sites (Vallina et al., 2014). Observing a stable number of Hutchinsonian niches suggests that their occupancy is driven more by stochastic than deterministic processes in these provinces. These processes may include the dilution of slow-growing cells in dynamic systems (Irwin et al., 2006) as well as the prevalence and favouring of r-strategists in the community. However, further work is needed to explore these speculations.

Overall, our results extend previous findings (Legendre, 1981) of coupling between microbial diversity and productivity, showing trophy-specific diversity patterns between low chl a and high chl a conditions. Only more temporally and spatially expanded observations (factoring in functional diversity using metagenomics-/transcriptomics, improved hydrographic descriptions, and physicochemical/nutrient profiles in these regions would allow our observations to be more confidently linked to ecosystem states driven by productivity (Chase and Leibold, 2002) and broader biogeographic descriptions.

Conclusion

Here, we assessed how measures of microbial diversity and activity can better inform ecological partitioning of the ocean, granting new perspectives on functional microbial biogeography across, within and between provinces. We showed that eukaryotic autotrophs and prokaryotic heterotrophs show similar cross-province β-diversity patterns, distinct from those shared by mixotrophs, cyanobacteria and eukaryotic heterotrophs. Our calculations of a productivity-specific length scale are, to our knowledge, the first attempt to quantify – on a per-sample basis – the influence of surface ocean currents on microbial communities coupled to PP measurements. This provides a first-order estimate of how spatially extended a microbial diversity signature may be, and thus the scale of a recognizable patch in a larger biogeographic area. Our findings also show the value of exploring functional communities (i.e. guilds) of microorganisms to more holistically understand community ecology with phylogenetic diversity data. We conclude that highly resolved sampling of these factors along more Lagrangian designs would help microbial ecology coherently map the subdivisions of the ocean’s biogeographic provinces.

Materials and methods

Sample collection

Our sampling was part of the PS113 (ANT-XXXIII/4) campaign onboard RV Polarstern from Punta Arenas, Chile.
matches in nutrient concentrations (Supplementary S1, Figs S1 and S2). We classified samples into provinces by identifying (i) clusters on a temperature-salinity plot (i.e. by water mass) where clusters were constrained by geographic proximity (Fig. S2), and (ii) boundary currents that coincided with province boundaries (i.e. where surface velocity vectors (geostrophic currents) were strong at province boundaries as identified above; Fig. S3a; Copernicus, 2020). We also identified ocean ridges in bathymetry profiles as features potentially structuring the modified provinces (GEBCO 2019). We compared our classification into Longhurst provinces with delineations proposed by Bracher et al. (2020) (Table S6).

We classified provinces as high-chl a provinces if their mean chl a concentrations were above 0.3 μg m⁻³ chl a, or as low-chl a provinces if their mean concentrations were below 0.1 μg m⁻³ chl a. Provinces with mean concentrations between 0.1 and 0.3 μg m⁻³ were treated as ambiguous, and we defaulted to classifications in Longhurst (2007). For details see Supplementary S1.

We calculated the distances to the coast and to province boundaries in qgis 3.14 (QGIS Geographic Information System, 2020) by densifying the vector coastlines and extracting vertices, followed by extracting the distance from each site points to the nearest hub on the coast contours. We used the ‘110 m vector coastline’ (v4.1.0, Natural Earth Data) and the province boundaries (delineated above) in our calculations.

**Biochemical and PP profiling**

At each station, we collected biochemical samples for dissolved inorganic nutrients (silicate, phosphate, nitrate and nitrite) as well as particulate organic matter (POC and PN). For details see Supplementary S2.

We measured PP with 200 μmol NaH¹³CO₃ stable isotope incubations in triplicates over a time period of ~24 h. For details on experimental design see Supplementary S3 and incubation conditions in Table S7. Analysis of ¹³C incorporated into organic matter was carried out on a PDZ Europa ANCA-GSL elemental analyser interfaced to a PDZ Europa 20–20 isotope ratio mass spectrometer (Sercon, Cheshire, UK) by the Isotopic Laboratory at the UC Davis, California campus. PP was calculated as in Eq. 1:

\[
PP = \frac{\text{at.\%}(\text{enriched}) - \text{at.\%}(\text{NA}) \times \text{POC}}{\left(100 \times \text{DIC}(13C) / \text{DIC}(13) + \text{DIC}(12C)\right) - \text{at.\%}(\text{NA}) \times \text{Incubation time}}\]

(1)

with at.\%(enriched) as the atom percent derived from ¹³C-spiked samples; at.\%(NA) is the atom percent of natural abundance derived from particulate organic carbon (POC) samples. Dissolved inorganic carbon (DIC) is assumed as 2000 μmol kg⁻¹ (after Zeebe and Wolf-Gladrow, 2001). Incubation time is the time between sample spike and filtration (~24 h).

We calculated specific PP \(P^B\) based on PP and chl a concentration, as shown in Equation 2.

\[
P^B = \frac{PP}{\text{chl} a} \]

(2)

where PP is expressed in mg C m⁻³ d⁻¹ and chl a concentration in mg chl a in m⁻³, resulting in a \(P^B\) value in mg C m⁻³ d⁻¹ chl a⁻¹.

**Productivity-specific length scale**

To link the biomass turnover of microbial communities to their pelagic advection, we calculate a length scale for biological–physical coupling (productivity-specific length scale) using a quantitative comparison of measured biomass turnover and measured current speed.

\[
\text{productivity} - \text{specific length scale} = \ln(2) \times \text{current speed} \times \lambda^{-1}
\]

(3)

where the productivity-specific length scale is expressed in kilometer. Current speed reflects average horizontal current speed from the VM-ADCP which provided data to Bremerhaven, Germany, from 2018-05-08 to 2018-06-10 (Strass, 2018). We took discrete measurements for biophysical analyses of sea surface water at 193 stations in the Atlantic Ocean from about 11 m depth through the ship’s seawater system (Teflon® tubing with a membrane pump) at an interval of ~0.5° latitude.

**Province delineation after Longhurst**

We defined ecological regions based on the variables suggested by Longhurst (2007), which include gradients in sea surface temperature, salinity, chl a, and checked for provincialism of microbial trophic groups. PP was calculated as in Eq. 1:

© 2021 The Authors. Environmental Microbiology published by Society for Applied Microbiology and John Wiley & Sons Ltd., Environmental Microbiology
for the depth range 20–50 m. The \( \lambda \) is derived as in Eq. 4.

\[
\lambda = \frac{\text{PP}}{\text{standing stock biomass}}
\]  

(4)

PP is expressed in mg C m\(^{-3}\) d\(^{-1}\) and the standing stock biomass corresponds to the chl a concentration (in mg m\(^{-3}\)) times 23 mg chl a (mg C\(^{-1}\)) [for conversion between chl a and carbon concentrations (Geider, 1987); validated against in situ POC concentrations; Table S8].

**Microbial sampling, processing and amplicon sequence analyses**

**Flow cytometry.** At each station, samples for cell counts were taken into 2 ml Eppendorf tubes and incubated in the dark prior to fixation using 0.2% paraformaldehyde. Samples were snap-frozen and stored at −80°C while at sea. Samples were analysed on a BD Accuri™ C6 Plus Flow Cytometer (BD Biosciences-US) according to Gasol and Moran (2015). For details on measurements see Supplementary S4.

**Microbial DNA sampling.** For 16S and 18S rRNA gene sequencing analysis, 4 L of seawater were filtered through 0.2 μm Sterivex\® filters. Filters were purged with air, tightly closed, snap-frozen in liquid nitrogen, and stored at −80°C until further analysis.

**DNA extraction and amplicon sequencing.** DNA was extracted using a DNeasy\® PowerWater\® DNA extraction kit (QIAGEN, Valencia, CA, USA, Catalogue No./ID: 149000) following the manufacturer’s instructions. After DNA extraction, DNA concentration was quantified using a Quantum™ Fluorometer and normalized to 2 ng μl\(^{-1}\).

Amplicons targeting the variable region 4 (V4) of the bacterial 16S rRNA gene (515F - 806R) (Caporaso et al., 2016) and the variable region 4 (V4) of the eukaryotic 18S rRNA gene (TA-Reuk454FWD1 – TAR-eukREV3) (Stoeck et al., 2010) were generated following standard protocols of amplicon library preparation (16S Metagenomic Sequencing Library Preparation, Illumina, Part #15044223 Rev.B) and sequenced using a MiSeq Sequencer (Illumina). 16S and 18S rRNA gene amplicon reads were generated using 300-bp paired-end sequencing using Nextera XT Index Kit v2 Set A-B (Illumina) index primer. Samples were demultiplexed using bcl2fastq (Illumina) with barcode mismatches set to 1.

ASV tables for both 16S and 18S rRNA gene amplicon sequences were constructed using the DADA2 R package, v1.15.1 (Callahan et al., 2016). Samples were processed using the DADA2 pipeline (v1.15) with an additional step where primers were trimmed using cutadapt v1.18 after pre-filtering of FASTQ files (dada2_16S.R, dada2_18S.R in Supplementary_Code/dada2/). Diagnostics of each filtering step are in Table S9 and number of reads plotted for each filtering step in density plots (Fig. S18).

**Taxonomic assignment and functional grouping.** Taxonomic assignment was performed outside DADA2 using SilvaNGS (v1.4) (Quast et al., 2013) pipeline for 16S rRNA gene data with the similarity threshold set to 1. Reads were aligned using SINA v1.2.10 (Pruesse et al., 2012), and classified using BLASTn (v2.2.30) (Camacho et al., 2009) with the Silva database (v132) as a reference database (Supplementary 2). For taxonomic assignment of 18S rRNA gene amplicons, we used the QIIME 2 Plugin ‘feature-classifier’ (v2019.7.0 from package ‘q2-feature-classifier’) in qiime 2 (Bokulich et al., 2018) and the pr2 database (v4.12) (Guillou et al., 2013).

We classified eukaryotic taxa as either autotroph, mixotroph, heterotroph, or unknown if no information of trophic lifestyle was available by performing an unstructured literature search to validate expert-led assignment by UJ (see outcomes in Table S10). Prokaryotes, except Cyanobacteria, were considered to be dominated by a heterotrophic lifestyle (sensu Herndl et al., 2008), see Table S10 for further details.

**Ecological data analyses**

We performed our data analysis using R v4.0.3 (R Core Team, 2020) and RStudio v1.4.1103 (RStudio Team, 2021). Biophysical parameters, concentrations of particulate organic matter, dissolved inorganic nutrients, microbial cell counts and PP measurements were used as environmental metadata for statistical analyses with microbial sequencing data.

Alpha diversity (\( \alpha \)-diversity) was determined by calculating Hill numbers (Chao et al., 2014) of richness, Shannon entropy and Simpson concentration (\( q = 0, q = 1 \) and \( q = 2 \)), using the iNEXT package (v2.0.20) repeating calculations 100 times (Supplementary_Code/alpha_diversity/alpha_diversity.R). Rarefaction curves of all Hill numbers were plotted using fortify() of the ggplot2 package (v3.3.3). Further analyses were performed using Shannon diversity measures of each trophic functional group, as it reflects true diversity (richness + evenness) and is less susceptible to fluctuations in rarer phyotypes.

Pearson correlations between microbial Shannon diversity and environmental parameters (salinity, temperature, PO4, H2SiO4, NO3, dissolved oxygen, distance to province boundary, distance to coast, PN, POC, chl a) were calculated and plotted using the corplot() function of the corplot package (v0.84) and adjusted for multiple testing.
using the Holm–Bonferroni method (Supplementary_Code/correlation_analyses/general_complot.R). Residuals of all correlations were screened for notable departures from Gaussian distributions, in which case these correlations were excluded from our results (Supplementary_Code/correlation_analyses/Residuals_complot.R, Figs S8–S15).

ASV tables and environmental metadata were transformed for comparability and statistical downstream analyses (general_clr_hellinger_transformations.R, z-scoring_subset.R in Supplementary_Code/data_transformations/). Before transformations, we removed all ASVs with ≤3 instances across all samples. Furthermore, before CLR transformation, we performed Bayesian-multiplicative treatments of zeros in the ASV tables using cmultRepl() function of the zComposition package (v1.3.4): this uses sample-wise totals to convert zero counts (which will lead to errors in log-ratios) into near-zero estimates, assuming undersampling rather than absence. To account for compositionality effects in our ASV tables (see Gloor et al., 2017), we performed a CLR transformation for RDA. Prior to permutational MANOVA (PERMANOVA) analyses using the decostand() function in vegan (v2.5.6), Environmental variables were checked for normal distribution (Fig. S19) and z-scored for scale-independent intercomparability.

We examined the distribution of sites among environmental gradients using a PCA on our microbial diversity metadata. We tested differences of ocean provinces using two-sided Wilcoxon rank-sum tests (hereafter: Wilcoxon) (Supplementary_Code/environmental_analyses/PP_plots. R L.81–84).

To examine microbial beta diversity (β-diversity) and its relation to environmental and contextual variables, we performed a set of RDA using the CLR-transformed ASV tables as response matrices and tables of environmental variables as explanatory matrices. We calculated sun azimuth and altitude based on sampling time and location using the suncalc (v.0.5.0) package in R. We performed stepwise model to identify significant environmental variables using ordR2step() function in vegan (Supplementary_Code/multivariate_analysis_ordistep.R). Differences of microbial dissimilarity between ocean provinces were tested with a PERMANOVA (Anderson, 2001) on Aitchinson distance of the CLR-transformed ASV tables using the adonis2() function along with a beta dispersion test to evaluate the homogeneity of dispersion using the betadisper() function in vegan (Supplementary_Code/multivariate_analysis/ RDA_functions.R).

Acknowledgements

We thank Volker Strass and the crew of RV Polarstern for their support during the PS113 expedition. We thank Pauline Thome for her support in sampling during the expedition and Susanne Spahic for her help during the expedition. We thank Simon Dreutter for help with the qGIS environment. We thank Dr. Wilken-Jon von Appen for his help in oceanographic data analyses. We thank Prof. Dr. Matthias Ulrich, Prof. Dr. Frank Oliver Glöckner as well as two anonymous reviewers for their comments on this study. This study was supported by the Helmholtz PÖMAR Graduate School, the POF IV Research Programme topic 6 and subtopic 6.2 of the Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Germany. Research funding was further provided by the Ocean Frontier Institute, through an award from the Canada First Research Excellence Fund.

Data Availability Statement

Chlorophyll a (chl a) data were derived from https://doi.org/10.1594/PANGAEA.913514 and published in Bracher et al. (2020). Dissolved inorganic nutrients, particulate organic matter and primary productivity measurements are publicly archived (https://doi.org/10.1594/PANGAEA.926458, https://doi.org/10.1594/PANGAEA.926460 and https://doi.org/10.1594/PANGAEA.926462 respectively). Sequence data for this study have been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under accession number PRJEB42499 (https://www.ebi.ac.uk/ena/browser/text-search?query=PRJEB42499), using the data brokerage service of the German Federation for Biological Data [GFBio, (Dippenbroek and Glöckner, 2014), in compliance with the Minimal Information about any (X) Sequence (MiXSiS) standard (Yilmaz et al., 2011). The analysis based on satellite data was generated using E.U. Copernicus Marine Service Information and are available at http://marine.copernicus.eu (Copernicus, 2020) using the GLOBAL_ANALYSISFORECAST_PHY_CPL_001_015 data (Lea et al., 2015). The 110 m vector coastline (ESRI shapefile; v4.1.0) is available at https://www.naturaearthdata.com/.

Code Availability

Statistical analysis code is available in the Supplementary_Code supplement of this paper with documentation in an RNotebook (top-level directory) and all code files. Code is also publicly archived on github in the tagged release: https://zenodo.org/badge/latestdoi/329689664 (released with publication).

References


Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher’s web-site:

**Appendix S1**: Supplementary Information.