

RESEARCH

Open Access



Particle-attached bacteria act as gatekeepers in the decomposition of complex phytoplankton polysaccharides

Feng-Qing Wang¹, Daniel Bartosik^{2,3†}, Chandni Sidhu^{1†}, Robin Siebers⁴, De-Chen Lu^{1,5}, Anke Trautwein-Schult⁴, Dörte Becher⁴, Bruno Huettel⁶, Johannes Rick⁷, Inga V. Kirstein⁷, Karen H. Wiltshire⁷, Thomas Schweder^{2,3}, Bernhard M. Fuchs¹, Mia M. Bengtsson^{4*}, Hanno Teeling^{1*} and Rudolf I. Amann^{1*}

Abstract

Background Marine microalgae (phytoplankton) mediate almost half of the worldwide photosynthetic carbon dioxide fixation and therefore play a pivotal role in global carbon cycling, most prominently during massive phytoplankton blooms. Phytoplankton biomass consists of considerable proportions of polysaccharides, substantial parts of which are rapidly remineralized by heterotrophic bacteria. We analyzed the diversity, activity, and functional potential of such polysaccharide-degrading bacteria in different size fractions during a diverse spring phytoplankton bloom at Helgoland Roads (southern North Sea) at high temporal resolution using microscopic, physicochemical, biodiversity, metagenome, and metaproteome analyses.

Results Prominent active 0.2–3 μm free-living clades comprised *Aurantivirga*, “Formosa”, *Cd. Prosilicoccus*, NS4, NS5, *Amylibacter*, *Planktomarina*, SAR11 Ia, SAR92, and SAR86, whereas BD1-7, *Stappiaceae*, *Nitricolaceae*, *Methylophagaceae*, *Sulfitobacter*, NS9, *Polaribacter*, *Lentimonas*, CL500-3, *Algibacter*, and *Glaciecola* dominated 3–10 μm and > 10 μm particles. Particle-attached bacteria were more diverse and exhibited more dynamic adaptive shifts over time in terms of taxonomic composition and repertoires of encoded polysaccharide-targeting enzymes. In total, 305 species-level metagenome-assembled genomes were obtained, including 152 particle-attached bacteria, 100 of which were novel for the sampling site with 76 representing new species. Compared to free-living bacteria, they featured on average larger metagenome-assembled genomes with higher proportions of polysaccharide utilization loci. The latter were predicted to target a broader spectrum of polysaccharide substrates, ranging from readily soluble, simple structured storage polysaccharides (e.g., laminarin, α -glucans) to less soluble, complex structural, or secreted polysaccharides (e.g., xylans, cellulose, pectins). In particular, the potential to target poorly soluble or complex polysaccharides was more widespread among abundant and active particle-attached bacteria.

Conclusions Particle-attached bacteria represented only 1% of all bloom-associated bacteria, yet our data suggest that many abundant active clades played a pivotal gatekeeping role in the solubilization and subsequent degradation

[†]Daniel Bartosik and Chandni Sidhu contributed equally to this work.

*Correspondence:

Mia M. Bengtsson
mia.bengtsson@uni-greifswald.de
Hanno Teeling
hteeling@mpi-bremen.de
Rudolf I. Amann
ramann@mpi-bremen.de

Full list of author information is available at the end of the article



© The Author(s) 2024. **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.

of numerous important classes of algal glycans. The high diversity of polysaccharide niches among the most active particle-attached clades therefore is a determining factor for the proportion of algal polysaccharides that can be rapidly remineralized during generally short-lived phytoplankton bloom events.

Keywords Algal bloom, Algal polysaccharide, Bacterioplankton, *Bacteroidota*, Carbohydrate-active enzyme, Carbon budget, Carbon cycle, Free-living bacteria, Helgoland Roads LTER, Marine microbes, Particle-attached bacteria, Particulate organic matter, Polysaccharide utilization locus

Background

Global photosynthetic net primary production (NPP) amounts to an estimated 104.9 gigatons of carbon per year [1]. Almost half of this is allotted to algae, in particular, to the small unicellular planktonic algae (phytoplankton) that dominate the world's oceans [2]. Diatoms (*Bacillariophyta*) represent the most prominent phytoplankton group, in particular, in polar and upwelling regions, and have been estimated to fix up to 20 gigatons of carbon annually [3]. It has been suggested that the silicate shells (frustules) of diatoms provide a competitive advantage over other phytoplankton by allowing them to save energy for cytoskeleton maintenance [4]. Further abundant and globally distributed phytoplankton taxa include photosynthetic dinoflagellates and haptophytes (*Haptophyta*), such as coccolithophorids. For the haptophyte genus *Phaeocystis*, it has been shown that their small cells with high surface-to-volume ratios can out-compete diatom productivity under certain conditions [5]. *Phaeocystis* often dominates spring and summer blooms after diatoms have peaked in the coastal North Sea [6], where they can account for up to 65% of the annual primary production [7].

Primary production by marine phytoplankton is not constant but culminates during phytoplankton blooms. Such blooms can be massive, yet they are usually short-lived. Bloom termination is often initiated by nutrient depletion and can be amplified by a number of factors, such as self-shading, grazing (e.g., by copepods), and various infections, e.g., by viruses, algicidal bacteria, parasitic peronosporomycetes (oomycetes), dinoflagellates, and marine fungi [8–11]. Also, the coagulation of algae and increased sinking of the formed particles due to reduced buoyancy can play a role [12].

During phytoplankton blooms, copious amounts of algal organic matter are released as dissolved or particulate organic matter (DOM, POM). Most of this is rapidly remineralized by heterotrophic bacteria and zooplankton, but the exact proportions are a matter of debate. It has been estimated that 62% of the daily phytoplankton production is on average consumed by small zooplankton [13] (reviewed in [14]). Zooplankton sloppy feeding and excretion in turn increase the DOM and POM pools available to bacteria [15], and measurements of bacterial

respiration rates have suggested that bacteria remineralize 70–92% of the POM within the mesopelagic zone (–200 to –1000 m) [16]. Only about 1–3% of biological net primary production reaches bathypelagic depths (below –1000 m) [17] via the so-called biological pump, where it can be sequestered for longer periods of time. According to recent estimates, about 10 gigatons of carbon are exported to the deep sea annually, including 1.3 gigatons by the biological pump, 15% of which is phyto-detritus [18].

About 95 to > 99% of the epipelagic marine bacteria typically consist of DOM-decomposing free-living (FL) planktonic bacteria (bacterioplankton) and the remainder of POM-decomposing particle-attached (PA) bacteria [19–21]. However, high particle abundances can elevate proportions of PA bacteria, at times possibly even above those of FL bacteria [19]. FL bacteria have been estimated to mediate 53% of the DOM and PA bacteria 50% of the POM fluxes [16]. Likewise, PA bacteria have been shown to exhibit higher per-cell activities (e.g., [20]) and higher proportions of hydrolytic enzymes [22, 23]. However, currently, we have only a poor understanding of the factors that determine the fractions of the organic matter that are remineralized by FL bacteria, PA bacteria, and the fractions that either feed the pool of recalcitrant DOM or sink out to the sea floor.

Depending on the developmental stage and physiological condition, up to 75% [24] or even more [25] of the dry weight of algae can consist of various polysaccharides, e.g., as intracellular stores of biochemical energy and as cell matrix and cell wall components. Many of these polysaccharides have no counterparts in terrestrial plants, in particular, those that are anionic, e.g., due to sulfation. The dominating polysaccharides in marine macroalgae (seaweeds) are well known, such as laminarins, fucoidans, cellulose, and alginates in brown algae (*Phaeophyta*); cellulose, xylans, and ulvans in green algae (*Chlorophyta*); and agars, carrageenans, and galactans (including porphyran and furcellaran) in red algae (*Rhodophyta*). Less is known about microalgal polysaccharides. Brown macroalgae and other stramenopiles, including diatom and raphidophyte phytoplankters, contain laminarin as a store of photoassimilated biochemical energy [26]. Laminarin, which is also used by haptophyte phytoplankters,

is a water-soluble, structurally simple β -1,3-linked helical homopolymer of glucose with occasional β -1,6-branches that typically consist of 20 to 30 monomers [27]. The dry weight of diatoms can consist of up to 35% of laminarin during exponential growth, and even up to 80% has been reported for the stationary phase [28]. Laminarin is therefore one of the most abundant polysaccharides on Earth [29].

The polysaccharides that are encrusted in the siliceous diatom frustules are more heterogeneous, and little is known about their structures. Studies of *Phaeodactylum tricorutum* have identified a sulfated glucuronomanan as a major cell wall component that might be widespread in diatoms [30]. However, monosaccharides other than mannose and glucuronic acid have been identified in frustules, including fucose, galactose, glucose, xylose, rhamnose, and arabinose [31]. Compositions depend on diatom species and physiological state, which would indicate a huge diversity in corresponding structures. Considering the prevalence of diatoms, these polysaccharides are produced in large quantities and play a non-negligible role in global carbon cycling.

Many microalgae also exude polysaccharide-rich extracellular polymeric substances (EPS). EPS have many functions, e.g., providing a nutritious matrix to attract beneficial bacteria, particularly in the immediate algal phycosphere. EPS also increase cell surface adhesiveness and thereby promote algae aggregation and flocculation [32–34]. Likewise, a portion of the EPS itself can coagulate into more dense transparent extracellular particles (TEP). The amount of EPS that algae produce depends on many factors. It tends to increase when nutrients become limiting, which is commonly interpreted as a mechanism to dispose excess carbon [35] as a substitute for adaptive photosynthesis downregulation. Not much is known about EPS composition, which may vary depending on algal species and physiological conditions, but most EPS seem to contain high proportions of sulfated polysaccharides [36].

Due to the inherent chemical heterogeneity and structural complexity, no bacterium can harbor the genes required to decompose all algal polysaccharides. Instead, bacteria specialize in subsets, which is why the remineralization of algal polysaccharides is a collective endeavor of polysaccharide-degrading bacteria with distinct substrate niches. Genes that code for the polysaccharide degradation machinery in bacterial genomes are often co-located as operons or regulons. Such polysaccharide utilization loci (sg. polysaccharide utilization locus (PUL)) are particularly prominent in the genomes of polysaccharide-degrading *Bacteroidota*, where they typically comprise a *susCD* gene tandem that codes for a SusD-like substrate-binding and for a SusC-like channel

protein of a TonB-dependent transporter (TBDT) [37]. These are accompanied by genes coding for degradative carbohydrate-active enzymes (CAZymes), namely glycoside hydrolases (GHs), carbohydrate esterases (CEs), polysaccharide lyases (PLs), and by accessory genes coding for, e.g., surface glycan-binding proteins (e.g., [38]), sulfatases, ABC transporters, and other associated functions. PUL lengths depend on the target substrate and can vary considerably. While a typical laminarin PUL consists of around 20 genes (e.g., [39]), PUL-rich loci can also encompass close to 100 genes (e.g., [40]).

We have analyzed the microbial response of bacteria to spring phytoplankton blooms in a series of studies at the long-term ecological research (LTER) site “Kabeltonne” off Helgoland Island in the southern North Sea [39, 41–44], in which we focused on the response of FL (0.2–3 μ m) bacteria and their associated polysaccharide niches. Recently, we could exemplarily show that abundant bloom-associated FL bacterioplankton clades preferentially consume water-soluble, low-complexity storage polysaccharides such as laminarin and α -glucans, which therefore exert a strong community structuring effect [39]. Also, other polysaccharides, such as alginate or mannose-containing polysaccharides, play a role, albeit in lower quantities [39]. An unknown proportion of the dissolved polysaccharides originate from POM that has been solubilized by PA bacteria and diffused away before uptake. However, so far, little is known about the involved polysaccharide-degrading PA bacteria and their connection to FL bacteria. A recent comparative study of FL and PA metagenome-assembled genomes (MAGs) from different water depths in the North Pacific Subtropical Gyre has shown that PA bacteria are characterized by higher predicted growth efficiencies and, on average, larger genomes with higher proportions of genes for peptidases, CAZymes, secretion, sensing and motility [23].

In this study, we investigated a diverse spring phytoplankton bloom that took place in 2018 off Helgoland Roads at high temporal resolution (51 sampling dates over a 90-day period). We aimed to disentangle the roles of PA bacteria in comparison with FL bacteria with respect to their potential to degrade phytoplankton-derived polysaccharides. We collected microscopic algal biodiversity and biovolume data, eukaryote 18S rRNA gene amplicon data, and 16S rRNA gene amplicon data of bacterial communities from FL (0.2–3 μ m), PA3 (3–10 μ m), and PA10 (> 10 μ m) filter fractions together with a broad range of physicochemical data. In addition, we performed metagenomics of FL (18 samples), PA3 (16 samples), and PA10 (8 samples) bacterial communities, reconstructed MAGs of abundant key players, and compared their polysaccharide degradation potentials. These data were complemented by metaproteomes from

10 selected time points during the bloom to link bacterial protein to the decomposition of algal glycans.

Results

The 2018 Helgoland spring phytoplankton bloom was diverse and polyphasic

Based on microscopically determined phytoplankton taxa, corresponding biovolume estimates (< 0.1 to $1.71 \text{ mm}^3 \text{ L}^{-1}$, Additional file 1: Table S1), and chlorophyll *a* measurements (~ 2 to 33.8 units, Additional file 1: Table S1), the 2018 Helgoland spring bloom consisted of a pre-bloom phase dominated by lowly abundant diatoms and *Phaeocystis* sp. haptophytes (March 1 to April 9), a diatom-dominated phase (April 10 to May 08) largely overlapping with a notable bloom of *Chattonella* raphidophytes (April 19 to May 11), and a late phase dominated by *Phaeocystis* sp. haptophytes and few *Dinophyceae* (May 09 to May 31) (Fig. 1).

Diatoms comprised various *Chaetoceros* species and *Thalassiosira rotula*. After they went into decline, *Phaeocystis* sp. and *Dinophyceae* numbers increased, with *Phaeocystis* sp. becoming dominant until the first wave of blooming algae ended about 1 week into June. A remarkable correlation was obtained between chlorophyll *a* measurements and estimated biovolumes of

photosynthetic plankters (Additional file 2: Fig. S1A). Additional non-photosynthetic plankters comprised in particular dinoflagellates, e.g., *Noctiluca scintillans*. The latter was detected at the end of May and, despite being low in numbers, dominated the biovolume of unicellular eukaryotic plankters due to large cell sizes (Additional file 1: Table S1, Additional file 2: Fig. S1B).

Analysis of the 15 most abundant 18S rRNA gene amplicon sequence variants (ASVs) largely supported microscopic observations (Additional file 2: Fig. S2). For example, the *Chaetoceros* bloom was detected in the PA10, and the *Phaeocystis* and *Noctiluca* blooms in the PA3 fractions (*Phaeocystis* cells are small, and *Noctiluca* cells are fragile and thus broke during filtration). One inconsistency was that *Chattonella* could not be detected, likely because their particularly fragile, large, wall-less cells disintegrated during filtration. In addition, 18S rRNA ASV data revealed a noteworthy peak of *Cryptothecomonas* nanoflagellates towards the end of the diatom bloom.

We focused on the period from March 1 to May 31. FL bacterial total cell counts (TCC) increased continuously from $0.5 \times 10^9 \text{ L}^{-1}$ on March 1 to a peak of $3.3 \times 10^9 \text{ L}^{-1}$ on May 24 (Fig. 1). This increase was gradual during the pre- and main bloom phases and progressed more

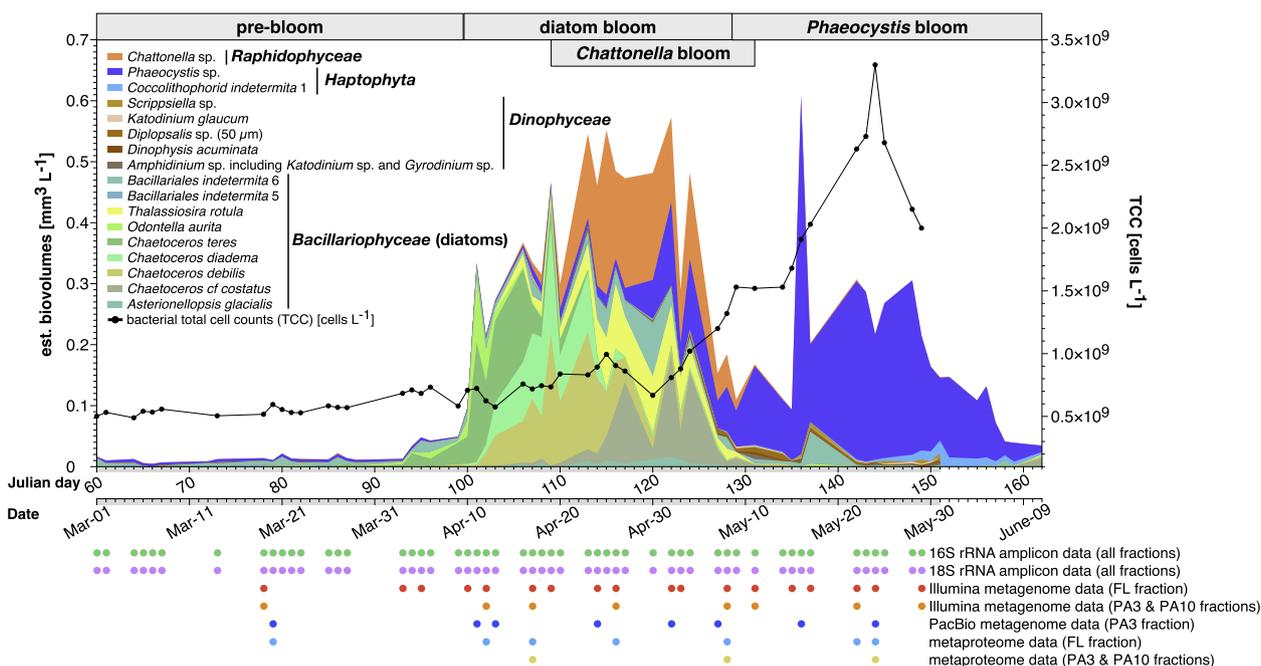


Fig. 1 2018 spring phytoplankton bloom at Helgoland Roads and associated datasets. Estimated biovolumes of abundant phytoplankton taxa (stacked colored areas) as assessed by microscopy and of total cell counts (TCC) of DAPI (4',6-diamidino-2-phenylindole)-stained bacteria (black line). Complete data was sampled until the end of May. Additional data for June is shown to visualize the end of the bloom but does not include TCC and *Dinophyceae* biovolume information. The core sampling period from March 1 until May 31 is indicated by a gray area on the abscissa. Additional data including abundances of non-photosynthetic plankters are provided in Additional file 2: Fig. S1 and Additional file 3. Associated omics sampling dates are indicated by colored circles at the bottom

rapidly after diatoms peaked at the end of April. Likewise, flagellate numbers increased throughout the bloom, ranging from $2.4 \times 10^6 \text{ L}^{-1}$ on March 1 to $1.2 \times 10^7 \text{ L}^{-1}$ on May 16 (Additional file 2: Fig. S3A). Flagellate numbers correlated well with Chl *a* and biovolume estimates, indicating that flagellates not only preyed on bacteria but also on microalgae. The remaining zooplankton was dominated by various copepod species with undulating abundances over time that showed no clear correlation to phytoplankton data, possibly due to vertical migration in and out of the sampled surface water (Additional file 1: Table S2).

An influx of nutrient-rich coastal water triggered the onset of the bloom

Physicochemical data indicated an incursion of nutrient-rich coastal waters at the onset of the diatom bloom, as on April 10 nitrate concentrations spiked to $19.0 \mu\text{M}$, silicate concentrations spiked to $10.7 \mu\text{M}$, and salinity decreased from 33.8 to 32.6 (Additional file 2: Fig. S3B-D). A second influx likely occurred from May 22 to 29 and was accompanied by an increase in silicate concentrations from 1.0 to $4.0 \mu\text{M}$ and a drop of salinity to 31.7. A spike in phosphate concentrations from 0 to $0.7 \mu\text{M}$ was also detected during this period (Additional file 2: Fig. S3E).

Wind directional data (Additional file 1: Table S3) supported these influx events, since northeasterly to easterly winds dominated from April 9 to 14 and from May 23 to 30 (Additional file 2: Fig. S4; see [39] for details). Additional rain and sunshine data are provided in Additional file 1: Table S4.

FL and PA bacterial communities exhibited distinct diversities and compositional shifts over the bloom's progression

16S rRNA gene amplicon sequencing of 153 samples from FL and PA fractions yielded 24,356 unique ASVs (Additional file 1: Table S5). Good's coverage (a measure for the proportion of singletons) indicated that this was adequate to capture basically all of the diversity of the FL communities (avg. coverage, ~ 1.0) and most of the diversities of both PA fractions (avg. coverages, 0.96 and 0.91, respectively) (Additional file 2: Fig. S5A). FL bacterial communities had significantly lower alpha diversity indices (Chao1, Simpson's, Shannon) than PA3 and PA10 communities (ANOVA, $p < 0.001$), whereas the difference between both PA communities was less pronounced (Additional file 2: Fig. S5B-D). Shannon indices exhibited distinct patterns over time (Fig. 2A, B), with a high correlation of PA3 and PA10 samples ($p < 0.0001$, Fig. 2A). FL Shannon indices were highest prior to the pre- and early diatom-bloom, decreased and stayed low during

the main diatom and early *Phaeocystis* bloom phases, and finally increased again towards the bloom's end in the late *Phaeocystis* bloom. No such trend was observed for PA communities (Fig. 2A, B).

NMDS analyses based on weighted UniFrac distances corroborated that PA3 and PA10 communities were more alike and FL communities more distinct. Pre-bloom communities grouped well and were distinct from main and late-bloom communities (Fig. 2C, left to right). While these differences were less pronounced between the two main bloom phases, they were still detectable. The average distances between the pre-bloom and the main bloom stages were also smaller in the FL than in both PA fractions (Fig. 2C), indicating that the bloom caused a more profound community change in both PA fractions.

Distinct bloom phases selected for distinct genera in all size fractions

Both, FL and PA communities showed clear temporal successions of distinct bacterial clades. PA communities, however, were not only more diverse but also dominated by different taxa and exhibited more dynamic compositional shifts (Fig. 3, Additional file 1: Table S5). While FL communities were dominated by *Alphaproteobacteria*, proportions were lower within PA3 and PA10 communities (Additional file 2: Fig. S6A). In contrast, *Gammaproteobacteria* exhibited particularly high relative abundances in PA3 and PA10 communities but less so in FL communities. Likewise, *Verrucomicrobiota* and *Planctomycetota* exhibited higher relative abundances in PA3 and PA10 than in FL communities, whereas *Bacteroidota* were ubiquitous in all samples (Additional file 2: Fig. S6A). *Flavobacteriaceae* accounted for similar percentages in all fractions before May 4 (FL, 7–25%; PA3, 4–19%; PA10, 5–24%, Fig. 3). Afterwards, *Flavobacteriaceae* proportions increased rapidly during May 4 to May 8 and May 15 to May 29 in the FL (up to 30%) but not in both PA fractions (up to 18%) (Fig. 3). *Cryomorphaceae* relative abundances were higher in FL than in PA communities, while it was the opposite for *Saprosiraceae* (Fig. 3).

As reported for FL bacterioplankton sampled at Helgoland Roads in previous years [41, 42], SAR11 clade Ia, *Planktomarina*, and *Amylibacter* accounted for a substantial fraction of ASVs. These three alphaproteobacterial clades had high relative abundances in all FL samples but exhibited lower relative abundances in PA3 and were even rare in PA10 samples (Additional file 2: Fig. S7A). Members of alphaproteobacterial unclassified *Stappiaceae* were thriving in PA3 during the late *Phaeocystis* bloom, while alphaproteobacterial *Sulfitobacter* simultaneously ramped up in PA10 samples.

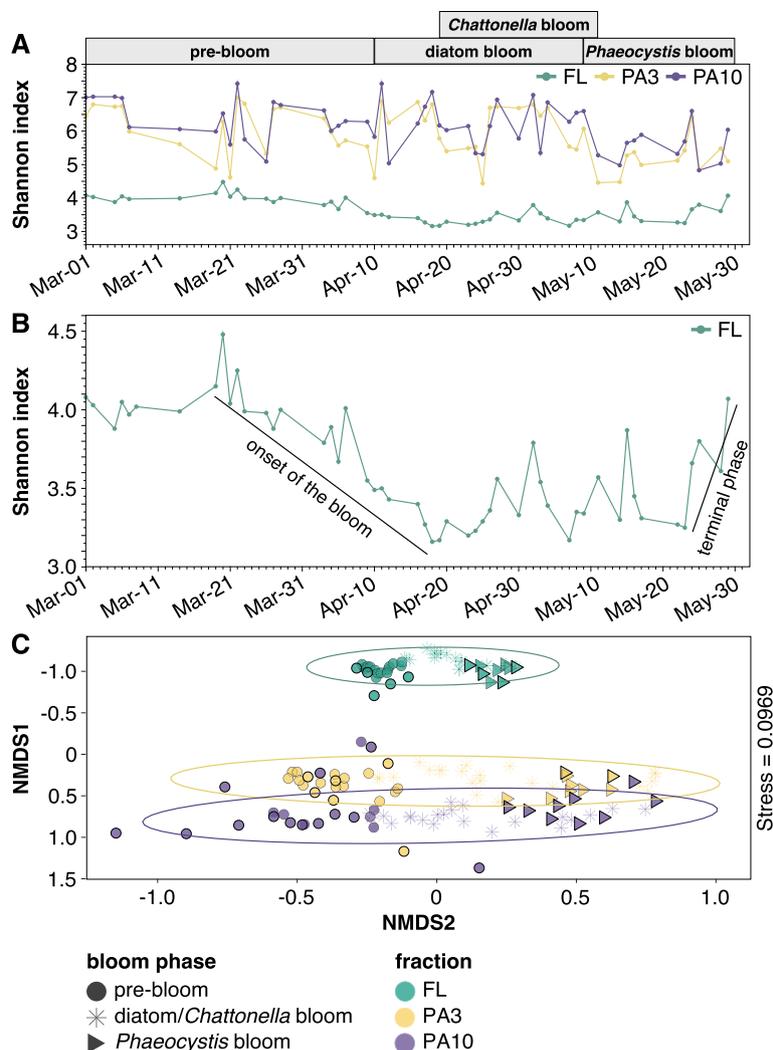


Fig. 2 Diversity as assessed by 16S rRNA gene amplicon data. **A** Shannon index of the sampled bacterial communities over time. Colors correspond to sample fractions (FL, 0.2–3 μm ; PA3, 3–10 μm ; PA10, > 10 μm). **B** Shannon index of the FL fraction over time. **C** Non-metric multidimensional scaling (NMDS) analysis of the bacterial community of all three fractions. Colors correspond to sample fractions and shapes to bloom phases

During the late diatom and *Phaeocystis* bloom phases, the FL bacterial community consisted primarily of *Bacteroidota*, including *Cd. Prosiliococcus* [45], *Aurantivirga*, “Formosa”, and members of the NS3a and NS5 marine groups, *Gammaproteobacteria* including SAR92 and unclassified *Nitrincolaceae*, as well as a distinct group of *Verrucomicrobiota* including *Lentimonas* (Additional file 2: Fig. S7A-B). Members of *Cd. Prosiliococcus*, the NS5 marine group, and *Aurantivirga* were also detected in the PA fractions but with lower relative abundances, some of which were probably due to carryover during fractionating filtration (Additional file 2: Fig. S7A). “Formosa” was present with similar low overall maximum relative abundance in FL and PA fractions (Additional file 2: Fig. S7B). *Aligabacter* was mostly detected in the PA

fractions and increased during the main bloom phases in PA3 communities (Additional file 2: Fig. S7B).

Polaribacter was not as abundant in FL communities as in previous [42] or later [39] years. More *Polaribacter* and unclassified *Saprospiraceae* were detected in both PA than in FL communities. *Maribacter* and *Winogradskyella* [21] were thriving during the late bloom phases but only in PA10 communities (Additional file 2: Fig. S7B).

During the bloom, *Gammaproteobacteria* had higher relative abundances in PA than in FL communities. For instance, in comparison with FL communities, members of the BD1-7 clade and *Colwellia* had higher relative abundances in PA communities during the diatom and *Phaeocystis* bloom phases, while unclassified

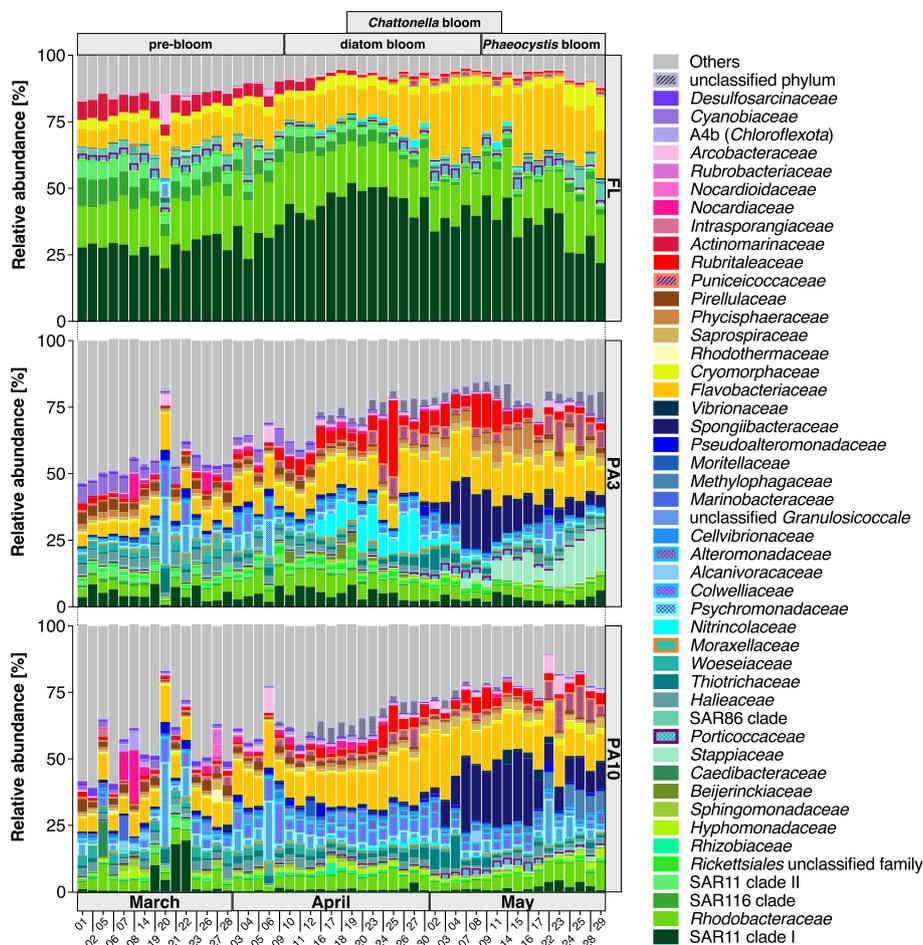


Fig. 3 Compositional differences among fractions as assessed by 16S rRNA gene amplicon data. Relative abundances of families are shown that belong to the topmost five abundant in at least two samples. The remaining families are subsumed as “Others.” Details are provided in Additional file 1: Table S5

Nitrincolaceae had higher relative abundances in PA3 communities (Additional file 2: Fig. S7A). Likewise, unclassified *Methylophagaceae* and *Glaciecola* exhibited higher relative abundances in PA10 communities (Additional file 2: Fig. S7A-B).

Furthermore, *Persicirhabdus* (*Verrucomicrobiota*) were proportionally more abundant in PA communities during the diatom bloom, and members of CL500-3 (*Planctomycetota*), known to be also associated with blooming freshwater algae [46], were proportionally more abundant in PA3 communities during the late diatom and *Phaeocystis* bloom phases. Similar bloom-associated temporal dynamics were also discernible in several groups that were not among the selected topmost genera. For instance, *Arenicella* and “Formosa” members were present only during the late diatom and *Phaeocystis* bloom phases in all fractions (Additional file 2: Fig. S7B). Likewise, members of the SUP05 cluster and NS7 marine group were present during the pre-bloom and early

diatom bloom phases in FL communities, while at the same time, members of the DEV007 clade (*Verrucomicrobiota*) were present in PA3 communities (Additional file 2: Fig. S7B). Further composition dynamics at the ASV level are provided in Additional file 3.

FL and PA community members exhibited distinct CAZyme composition dynamics

We selected 42 samples for metagenome sequencing, namely 18 FL (Illumina), 16 PA3 (Illumina, 8; PacBio, 8), and 8 PA10 (Illumina) samples (Additional file 1: Table S6). The resulting metagenomes amounted to 1.6 Tbp raw sequences. *K*-mer-based metagenome composition analyses corroborated significant differences between fractions (PERMANOVA, $p = 0.003$; Additional file 2: Fig. S8A). Distances between the pre-bloom and the two bloom periods were closer for FL than for PA data, corroborating 16S rRNA gene amplicon-based NMDS analyses (Fig. 2C). Individual assemblies of all

metagenomes yielded 18.2 Gbp with 2.5 kbp minimum length (Additional file 1: Table S6).

We computed and compared CAZyme relative frequencies in assembled Illumina metagenome data over time (eight samples of each fraction, Additional file 1: Table S7). Overall, genes targeting β -1,3-glucan (laminarin) were most frequent, with peaking relative frequencies towards the end of the diatom bloom (Fig. 4A). Respective genes were dominated by *Bacteroidota* and *Gammaproteobacteria* in all fractions, whereas *Verrucomicrobiota* (more prominent in FL fractions) and *Planctomycetota* (more prominent in PA fractions) contributed only little (Additional file 1: Table S7). These data suggested an overall increase of laminarin-consuming bacteria when the diatom bloom collapsed, most notably in PA3 communities. This corroborates recent data from FL bacteria during the 2020 Helgoland spring bloom, where laminarin PULs were the most frequent and highest expressed of all PULs

[39]. In terms of gene compositions, β -glucan PULs comprised the previously described variant-1 [39] coding for GH149, GH17, GH16, GH158 and GH30 enzymes (including variations), variant-2 coding for GH16 or GH17 and GH3 enzymes [39], and a PUL type coding only for GH16 enzymes (Additional file 2: Fig. S9).

Genes targeting α -glucans exhibited the second highest relative frequencies and exhibited no discernible trend (Fig. 4B). For the most part, respective genes were more frequent in PA than in FL communities. More *Gammaproteobacteria* and *Planctomycetota* contained these genes in PA than in FL communities. Four types of α -glucan PULs were present: type I coding for only one or more GH13 enzymes; type II coding for GH13, GH65, and sometimes an additional GH31 enzyme; type III coding for GH13, GH77, and GH57 enzymes; and type IV coding only for GH13 and GH31 enzymes (Additional file 2: Fig. S9). Genes targeting alginate were also

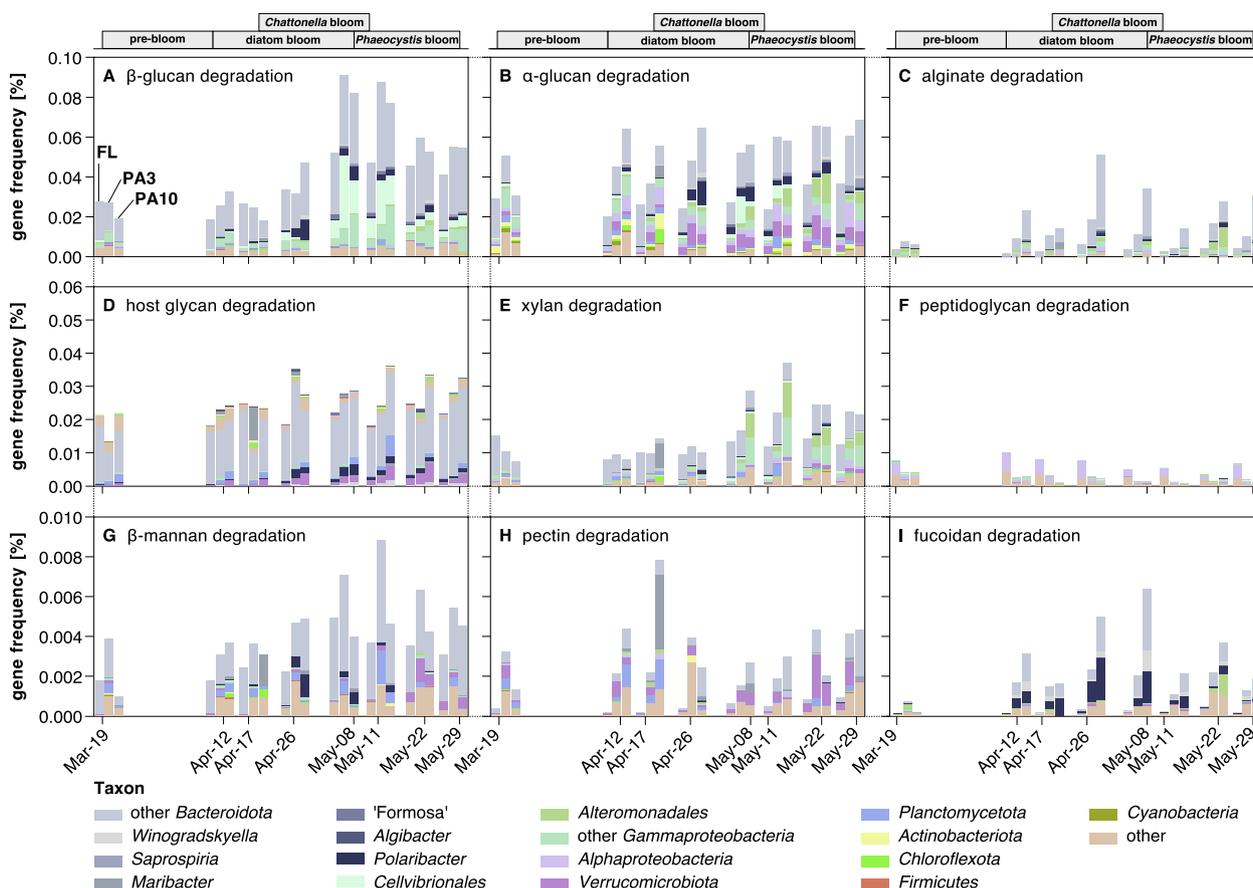


Fig. 4 Frequencies of CAZyme genes attributed to the degradation of specific polysaccharide substrates. CAZyme genes in metagenomes and corresponding substrates were predicted using the dbCAN3-sub database. Gene frequencies were calculated as follows: frequency = $\Sigma(\text{average coverage of target genes}) \times 100 / \Sigma(\text{average coverage of all genes})$. Only dates with data for all fractions were plotted as stacked bar charts (left to right: FL, 0.2–3 μm ; PA3, 3–10 μm ; PA10, > 10 μm). Colors represent dominating taxa. Additional data for α -rhamnosides, chitin, arabinans, α -mannans, cellulose, and sialic acids is depicted in Additional file 2: Fig. S10, and complete data is summarized in Additional file 1: Table S7

frequent, with notable higher proportions in PA communities, particularly in PA10 (Fig. 4C).

Relative frequencies of host glycan degradation genes, e.g., genes targeting eukaryotic N-glycans, showed no trend in FL communities but were for the most part higher in PA communities, where they increased during the diatom and *Phaeocystis* bloom phases (Fig. 4D). Respective genes attributed to unclassified *Bacteroidota*, *Polaribacter*, *Verrucomicrobiota*, and *Planctomycetota*, with the latter preferring PA10 fractions.

Xylan degradation genes were rarer. Their relative frequencies ramped up in PA communities after the diatom bloom abated, in particular, in PA3 with high proportions of *Alteromonadales* and other *Gammaproteobacteria* during the *Phaeocystis* bloom (Fig. 4E). Genes targeting peptidoglycan (murein) were notably more frequent among FL than PA bacteria. Proportions were highest during the pre- and diatom bloom stages, and lower during the late bloom (Fig. 4F).

For α -mannans, there were no consistent differences between FL and PA communities (Additional file 2: Fig. S10D). Frequencies were highest during the early diatom bloom phase and leveled off towards the end of the *Phaeocystis* bloom. In contrast, proportions of genes for β -mannan degradation were often highest among PA3 bacteria, in particular, during the diatom to *Phaeocystis* bloom transition phase (Fig. 4G). A transition in α -mannan degradation from *Flavobacteriales* to other *Bacteroidota* was observed before, during, and after the bloom (Additional file 2: Fig. S10D), whereas bacterial communities harboring β -mannan degradation genes were dominated by *Planctomycetota*, *Verrucomicrobiota*, and *Bacteroidota* during the main and late bloom stages (Fig. 4G).

Proportions of genes targeting pectins (Fig. 4H) were notably higher in both PA fractions and increased during the diatom and late *Phaeocystis* blooms. During the diatom and *Phaeocystis* bloom phases, fucoidan degradation genes had much higher proportions in PA10 than in PA3 or FL communities, which coincided with a notable increase in the proportions of *Polaribacter*, *Maribacter*, and other *Bacteroidota* (Fig. 4I).

Genes for the degradation of sialic acids were more frequent in PA3 communities but ramped up in FL communities after the diatom bloom abated (Additional file 2: Fig. S10F). However, during the *Phaeocystis* bloom, sialic acid degradation potential appeared to have shifted towards PA communities. This shift was characterized by an increase in *Planctomycetota* and unclassified *Bacteroidota*, along with the emergence of *Saprospiria* (Additional file 2: Fig. S10F). Genes for the degradation of α -rhamnosides, chitooligosaccharides, arabinans, and cellulose were notably more frequent in both PA fractions

and increased as the bloom progressed (Additional file 2: Fig. S10A-C, E).

Representative PA and FL community MAGs

From all 42 assembled metagenomes, we reconstructed 1944 initial bins (Illumina, 1721; PacBio, 223). Manual refinement retained 146 MAGs that fulfilled the MIMAG high-quality (HQ) criteria ($> 90\%$ completeness; $< 5\%$ contamination; presence of 23S, 16S, and 5S rRNA genes; and ≥ 18 tRNAs) [47], 964 MAGs (Illumina 895; PacBio, 69) of at least medium quality (MQ) ($\geq 50\%$ completeness, $< 10\%$ contamination), and 399 Illumina MAGs that did adhere to the “near complete” category by Almeida et al. [48] and were thus treated like HQ MAGs in downstream analyses (Additional file 1: Table S8, Additional file 2: Fig. S11). Dereplication of these in total 1509 MAGs at 95% ANI yielded 305 species-level MAGs of 16 known phyla (Additional file 1: Table S8, Additional file 2: Fig. S12), including 139 (45.7%) HQ MAGs. The average number of recovered MAGs per sample increased with decreasing filter pore size in fractionating filtration (PA10, 20; PA3, 26; FL, 57), due to the decreased capture of eukaryotic biomass [49]. The average size of HQ MAGs was larger in PA than in FL communities (Additional file 2: Figs. S13-14). Sizes of abundant MAGs ranged from 0.8 to 8.6 Mbp in PA and from 0.6 to 2.9 Mbp in FL communities (Fig. 5). Details are provided in Additional file 1: Table S9 and Additional file 3.

Based on 16S rRNA gene amplicon data (Additional file 1: Table S5), we selected 40 high-abundance genera (Additional file 1: Table S10), 39 of which were represented by corresponding MAGs. MAGs from FL community samples included previously identified relevant clades at Helgoland Roads, such as *Aurantivirga* [44], *Polaribacter* [50], “Formosa” species Hel1_33_131 [51], *Cd. Prosilicoccus* [45], the NS4 and NS5 marine groups [52], *Amylibacter*, and the SAR11 Ia, SAR92, and SAR86 clades. MAG abundances based on read frequencies were also not unusual compared to earlier years with one notable exception. We detected *Polaribacter* clade 2-b [50] (Additional file 2: Fig. S15), but in agreement with ASV data, its maximum relative abundance of $< 0.1\%$ was much lower than in previous observations, with recorded maxima of 14.7% (2010), 19.8% (2011), and 34.5% (2012) [53].

More *Planctomycetota* (FL, 2; PA3, 9; PA10, 2) and *Chitinophagales* (FL, 2; PA3, 7; PA10, 5) MAGs were obtained from PA3 and PA10 metagenomes, whereas more alphaproteobacterial MAGs were obtained from FL metagenomes (FL, 31; PA3, 22; PA10, 23). Also, the genus diversity of *Flavobacteriaceae* MAGs was higher in PA than in FL communities (FL, 12; PA3, 17; PA10, 14). Based on ASV and MAG abundance data, we categorized

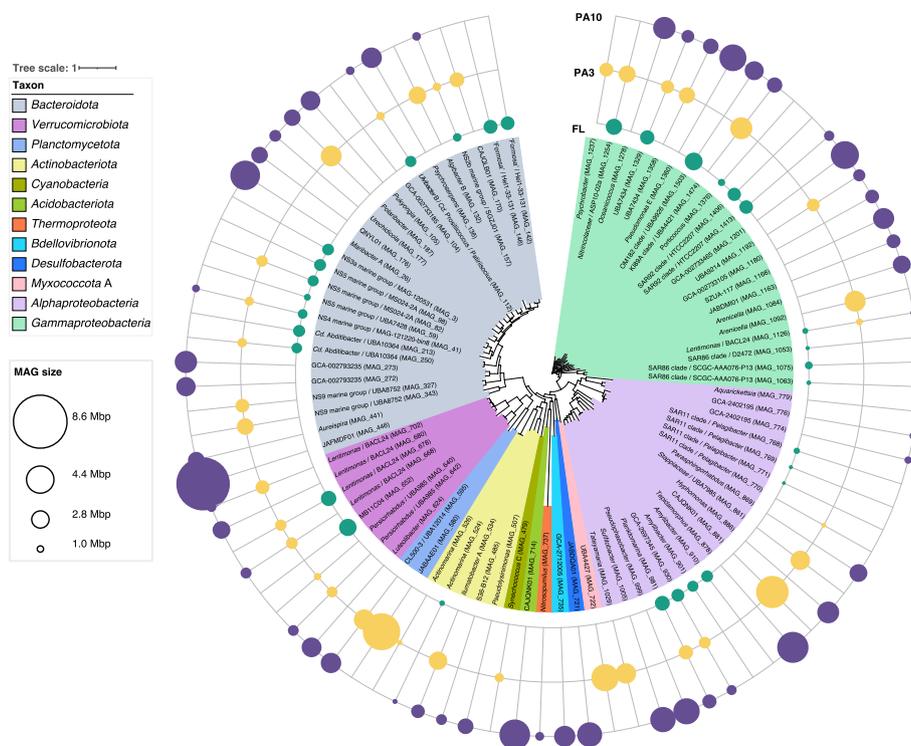


Fig. 5 Sizes of the ten topmost abundant MAGs across each sample ($n = 136$). De-replication was carried out within each fraction, and MAGs belonging to the top ten in terms of abundance (as determined by MAG abundance) were specifically chosen for this analysis. The central maximum-likelihood tree of MAGs was computed in anvio v7.1 based on protein sequences of 38 universal single-copy genes. Circle area sizes correspond to MAG sizes with colors representing the fractions from which the MAGs were retrieved

MAGs into those that were most abundant in either FL or PA communities and those that were abundant in both. In total, 152 MAGs belonged to the abundant PA MAGs (Additional file 1: Table S11), including *Polaribacter*. Further details are provided in Additional file 3.

Novel PA community MAGs

Comparison of all MAGs to those obtained for FL bacteria during the 2010 to 2016 spring blooms [43, 44] identified 100 species-level PA MAGs (HQ, 42; MQ, 58) that were uniquely obtained in 2018 (Additional file 1: Table S8, Additional file 2: Fig. S16). According to GTDB r207_v2, 76 of these MAGs represented novel species (Additional file 1: Table S8).

The 100 PA MAGs comprised a high proportion of over 3.5 Mbp (38/100 as compared to 57/305 for all species-level MAGs) (Additional file 2: Fig. S16C). They were dominated by *Paraglaciecola* (MAG_1218), *Pseudalteromonas* (MAG_1211 and 1212), and UBA12014 (CL500-3, *Planctomycetota*, MAG_591) in PA3 and GCA-002793235 (*Vicingaceae*, MAG_272), GCA-002733465 (*Kangiellaceae*, MAG_1201), *Maribacter* A (MAG_26), *Polaribacter* (MAG_189), and *Pseudolysiniomonas* (MAG_507) in PA10 communities (Additional

file 2: Fig. S12, Additional file 3). We also obtained HQ MAGs of *Acidobacteriota* and *Chloroflexota*, plus seven of ten *Planctomycetota* MAGs and four *Polaribacter* MAGs representing species that we did not identify at Helgoland Roads before.

Few particularly active but distinct MAGs dominated FL and PA communities

Seven sampled FL metaproteomes yielded 43,750 unique proteins (Additional file 1: Table S12). A total of 15,906 of these proteins were assigned to 177 FL MAGs, with up to 16.3% SusC- and SusD-like proteins and various TBBDTs (Additional file 2: Fig. S17, Additional file 3). Such high proportions agree with previous metaproteome studies on bloom-associated bacterial communities [41, 43, 54]. Conversely, only 5018 proteins were obtained from three PA3 and PA10 metaproteome sampling dates, which were dominated by eukaryotic proteins (42.7 to 64.0%). Just 932 of these proteins could be assigned to bacterial MAGs (Additional file 1: Table S12, Additional file 3), which is why we used these data only to pinpoint the most active PA MAGs (Additional file 2: Fig. S18).

Protein abundance data corresponded well with calculated MAG abundances from corresponding Illumina

metagenomes (Additional file 2: Fig. S19). Abundant MAGs with high overall protein expression on all seven FL sampling dates comprised members of the NS4 marine group, *Planktomarina*, and the OM182 and SAR11 clade Ia clades. The highest overall expression was observed in a *Nitrincolaceae* ASP10-02a clade MAG, but only during the diatom bloom phase (Fig. 6). The expression of SusC-like proteins is indicative of oligosaccharide uptake in *Bacteroidota*. MAGs with high SusC-like protein expression during the diatom and *Phaeocystis* bloom phases comprised members of *Cd. Prosilicoccus*, *Aurantivirga*, the NS3a, NS5, and NS4 marine groups, *Cd. Abditibacter*, and the *Cyclobacteriaceae* clade UBA4465 (Fig. 6). Other MAGs expressed SusC-like proteins only during distinct bloom phases, such as during the diatom bloom phase (other members of the NS5 marine group), the pre- and diatom bloom phases (a member of the NS2b marine group), the late diatom and *Phaeocystis* bloom phases (members of “Formosa”), or only the *Phaeocystis* bloom phase (again a member of the NS5 marine group).

Apart from bacteroidotal SusC-like proteins, TBDTs for the uptake of larger substrates, possibly including oligosaccharides, were predominantly expressed by members of various gammaproteobacterial clades. MAGs of the OM182, SAR92, and SAR86 clades exhibited high TBDT expression during all sampling dates, while others showed such expression mostly during the *Phaeocystis* bloom, e.g., MAGs of the SAR86 clade and *Glaciecola*.

In accordance with ASV data, *Polaribacter* were found to be only lowly abundant and hardly expressed

in FL communities but prominent in PA communities (Additional file 2: Fig. S7A). Further data are shown in Additional file 2: Fig. S20. It is noteworthy that *Alphaproteobacteria* in FL communities expressed mostly ABC-type transporters, whereas in PA10 communities, members of the alphaproteobacterial genera *Parasphingopyxis*, *Parasphingorhabdus*, *Maricaulis*, and *Hyphomonas* featured the highest TBDT expressions (Additional file 2: Fig. S19). *Parasphingopyxis* species have been isolated from red macroalgae and *Maricaulis* from dinoflagellate phycospheres [55, 56], while *Parasphingorhabdus* species have been found in mollusk guts [57]. *Maricaulis* and *Hyphomonas* can attach to surfaces via prosthecae and feature complicated life cycles [58, 59]. Further details on active MAGs are provided in Additional file 3.

Active CAZymes, PULs, and PUL-like clusters

Expressed CAZymes in FL community metaproteome data mapped to PULs and PUL-like clusters that were predicted to target host glycans, α-glucans, β-glucans, xyloglucans, fucose, alginate, and chitin (Additional file 1: Table S12). High expression was also observed for α-glucan degradation CAZymes with a peak on April 26, and β-glucan (laminarin) degradation CAZymes, which were particularly expressed during the diatom bloom’s end and the second bloom phase (May 8, 22, and 24) (Additional file 2: Fig. S21). On May 8, after the diatom bloom, also few CAZymes targeting fucose-containing polysaccharides were expressed. Complementary

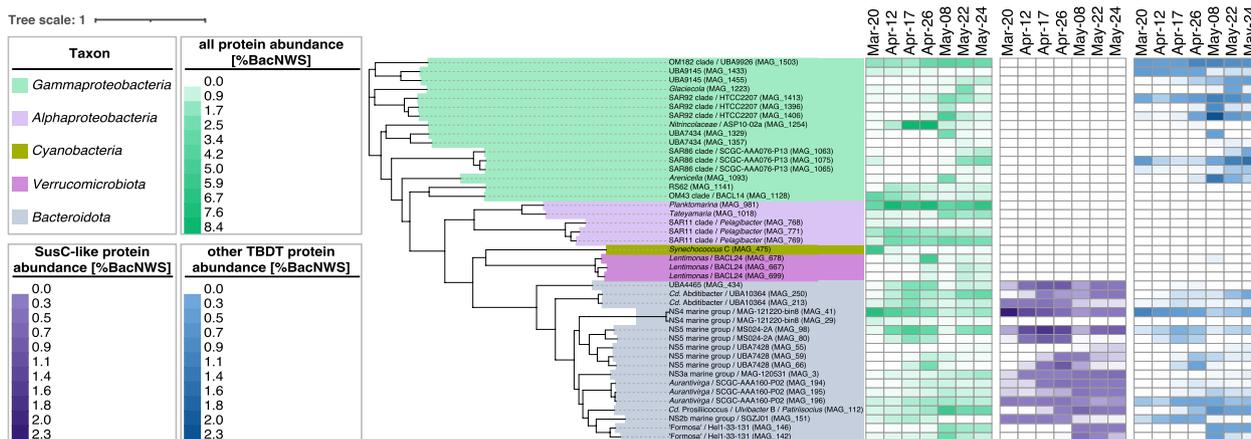


Fig. 6 Topmost expressed bacterial MAGs obtained from sampled FL bacteria. In total, 44 highly expressed bacterial FL MAGs were selected, and a maximum-likelihood tree was computed in anvio v7.1 based on protein sequences of 38 universal single-copy genes. Taxonomic affiliations according to GTDB r207_v2 are provided for each MAG with corresponding Silva r138.1 affiliations in parenthesis. Corresponding phyla (class for *Proteobacteria*) are represented by background colors. Protein expression based on summarized protein abundances as a percentage of bacterial normalized weighted spectra (%BacNWS) are represented by heatmaps: (i) overall expression (shades of green), (ii) expression of SusC-like proteins (shades of purple), and (iii) expression of TonB-dependent transporters (TBDTs) except for SusC-like proteins (shades of blue)

information on the PA metaproteome data is provided in Additional file 2: Fig. S18 and Additional file 3.

MAG analyses highlight distinct polysaccharide degradation potentials in abundant FL and PA community members

Out of the 305 species-level MAGs of all fractions, 244 contained degradative CAZymes including 161 candidate PULs (*susCD* gene tandems plus at least 1 degradative CAZyme), 1056 PUL-like clusters (1 *susC*-, *susD*-like or other TBSDT gene plus at least 1 degradative CAZyme), and 652 CAZyme-rich gene clusters (at least 3 degradative CAZyme genes) (Additional file 1: Table S13, Additional file 2: Fig. S22).

We linked MAGs with 16S rRNA gene amplicon data to leverage the high temporal resolution amplicon data to uncover variations in MAG abundances (Additional file 1: Table S10, Additional file 2: Fig. S23), for which we selected the 71 most abundant MAGs for in-depth PUL analysis. Nine of these harbored 40 or more CAZyme genes, all of which were prevalent in PA communities (Fig. 7). A description of the most prominent MAGs, their links to ASVs and changes over time as well as their

key CAZyme genes and inferred substrates is provided in Additional file 3, whereas a more holistic summary of the main results is provided in the subsequent discussion.

Discussion

The 2018 spring phytoplankton bloom at Helgoland Roads was among the most diverse in terms of phytoplankton species richness that we analyzed since 2009 [42, 43], in particular, compared to that of 2020, where algal biomass was almost entirely dominated by few diatom species during two sharply separated bloom phases [39]. The 2018 spring bloom in contrast was characterized by more complex gradual successions of diatoms, raphidophytes, haptophytes, and—to a lesser extent—photosynthetic dinoflagellates.

In 2018, an influx of nitrate- and silicate-rich freshwater around April 10 was likely instrumental in bolstering the diatom bloom, which resulted in an almost complete consumption of free silicate within a fortnight. A second influx event around May 23 during the late *Phaeocystis* bloom coincided with the emergence of *Noctiluca scintillans*, a heterotrophic giant dinoflagellate (0.2–2 mm diameter) that frequently occurs in Helgoland waters

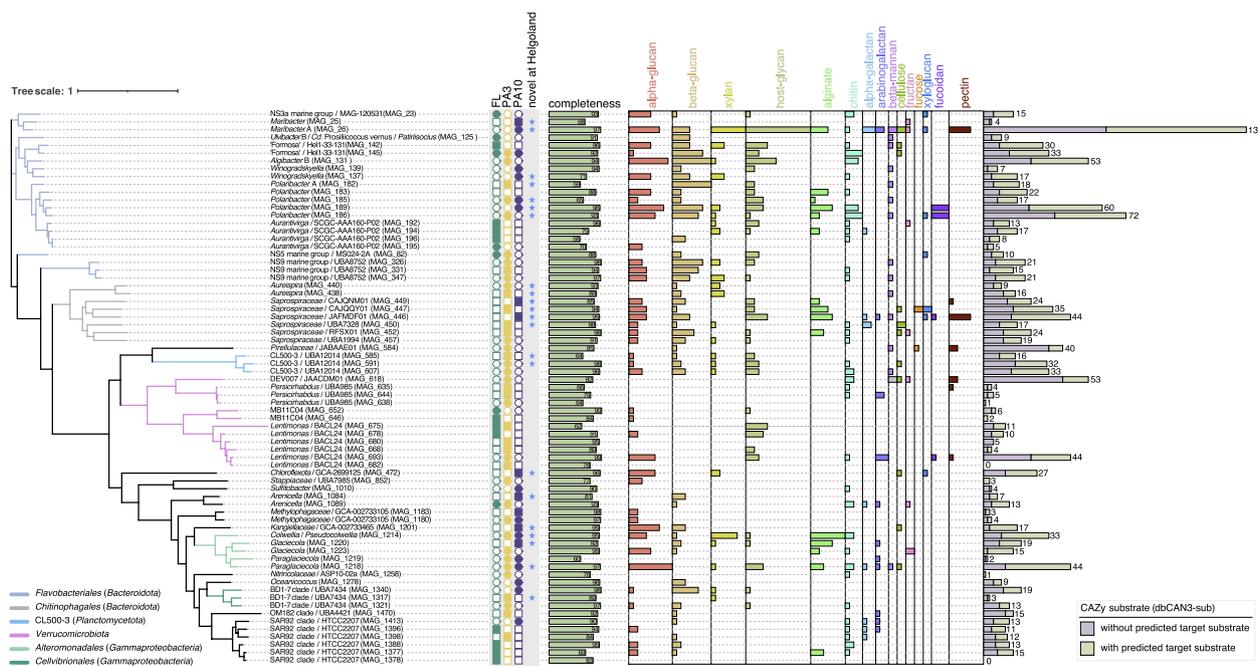


Fig. 7 Predicted substrates of PULs and PUL-like clusters in abundant MAGs. Based on 16S rRNA amplicon and corresponding MAG abundance data, we selected 71 MAGs from bacterial clades that were prominent either during particular bloom phases and/or in particular size fractions. A maximum-likelihood tree (computed in anvio v7.1 based on protein sequences of 38 universal single-copy genes) is shown to the left. PULs and PUL-like clusters in these MAGs were annotated and corresponding target substrate classes were using the dbCAN3-sub database. The numbers of involved degradative CAZyme genes (GH, PL, CE) are represented by bar charts. The stacked bars to the very right correspond to the total number of CAZyme genes per MAG with and without predicted substrate specificities. In addition, the fractions are indicated where the MAGs were most abundant (FL, 0.2–3 μm; PA3, 3–10 μm; PA10, > 10 μm), with circles representing data obtained from corresponding ASVs and squares representing read-based MAG abundances. Further details are provided in Additional file 1: Table S7

from June to August [60]. *N. scintillans* was most probably transported with coastal waters to Helgoland, and since *N. scintillans* prey on *Phaeocystis* [61], likely contributed to the *Phaeocystis* bloom's demise. Likewise, *Cryothecomonas* nanoflagellates, detected after the diatom bloom's peak, prey on diatoms [62], and thus likely contributed to the termination of the diatom bloom.

The bacterioplankton responded to the spring bloom with swift successions of distinct clades, which were more dynamic in the more diverse PA communities. Some bacterial clades correlated with distinct phytoplankton bloom phases, e.g., *Polaribacter*, *Winogradskyella*, and unclassified *Nitrincolaceae* with the diatom bloom, and unclassified *Stappiaceae*, *Sulfitobacter*, and unclassified *Methylophagaceae* with the *Phaeocystis* bloom. Other clades were abundant during both, the late diatom and *Phaeocystis* bloom phases, e.g., *Cd. Prosilicoccus*, "Formosa", *Algibacter*, *Glaciecocola*, and the BD1-7 clade.

The positive selection of bloom-adapted bacterial clades resulted in a decline in the diversity of FL bacteria and a size increase of the most abundant MAGs, notably in *Bacteroidota*, *Gammaproteobacteria*, and, to a lesser extent, *Planctomycetota*, *Verrucomicrobiota*, and *Alphaproteobacteria*. Diversity increased again during the collapse of the *Phaeocystis* bloom with the proliferation of more opportunistic generalists, such as members of the SAR86 clade and *Methylophagaceae* [63]. These clades featured smaller genomes, which was reflected in a decrease in the average size of abundant MAGs during the terminal bloom phase. Contrasting patterns were observed in both PA fractions, where diversities did not decrease during the main bloom phases, while the sizes of the most abundant MAGs decreased during the diatom bloom and increased notably during the late *Phaeocystis* bloom towards the bloom's end. This illustrates that different selective forces shaped FL and PA communities.

PA bacteria harbored more genes to degrade hardy soluble and structurally complex polysaccharides

Both abundant FL and PA bacteria featured high proportions of polysaccharide-degrading bacteria, however, with distinct CAZyme and PUL repertoires. Genes for the degradation of laminarins and α -glucans, both abundant, soluble, and structurally simple storage glucans, were the most prominent among FL bacteria, corroborating previous observations at Helgoland Roads [39]. Surprisingly, such genes were proportionally even more abundant among PA bacteria. Owing to sheer numbers, FL bacteria likely decomposed the bulk of laminarins and α -glucans, but the high proportion of respective genes in PA bacteria indicates that they are far from insignificant in this process. PUL analyses of 71 abundant MAGs from all fractions (Fig. 7) substantiated the salient role

of storage glucans, as 40 contained β -glucan and 43 α -glucan PULs (Additional file 2: Fig. S9), fortifying the view that these glucans become available to PA bacteria that colonize senescent or dead algae.

Like during the 2020 spring bloom [39], α -glucan PULs were also dominated by type I α -glucan PULs in 2018. In addition to the previously described α -glucan PULs types I, II, and IV [39, 44], we also identified an additional type III comprising GH13, GH57, and GH77 (see [64]) genes in *Gammaproteobacteria*. MAG analyses showed an increase in PA *Gammaproteobacteria* with α - and β -glucan utilization genes after the diatom bloom (Additional file 2: Fig. S9). For example, MAG_1223 (*Glaciecocola*) and MAG_1218 (*Paraglaciocola*) contained abundant genes for α -glucan hydrolysis, while MAG_1340 (BD1-7 clade) was rich in β -glucan hydrolytic genes (Fig. 7), suggesting that besides *Bacteroidota* also *Gammaproteobacteria* are significant α - and β -glucan consumers during algae die off phases.

Many marine macroalgae, e.g., *Saccharina* and *Fucus* brown algae, release gel-forming alginate and pectin-like polysaccharides [65]. Besides, alginate biosynthesis genes have been found in bloom-associated marine SAR92 clade *Gammaproteobacteria* [63]. Metatranscriptome analyses have furthermore suggested that alginate is an abundant bacterial substrate during spring blooms at Helgoland Roads [39]. Alginate and pectin degradation gene frequencies were notably more abundant in PA communities, reflecting the low solubilities of both substrates. Alginate gene frequencies in general prevailed over pectin degradation gene frequencies. This corroborates studies on the bacterial colonization of synthetic alginate and pectin particles by Bunse et al., where alginate was the preferred substrate [66]. We found alginate utilization genes predominantly in metagenome sequences attributed to unclassified *Bacteroidetes*, *Polaribacter*, and *Alteromonadales*. This was corroborated by MAG analyses, with alginate PULs present in PA *Saprosiraceae* (*Bacteroidota*), *Polaribacter* (*Bacteroidota*), and *Colwellia* (*Alteromonadales*) (Fig. 7). These in situ data also support the in vitro experiments of Bunse et al., who identified *Colwellia* as among the primary colonizers on synthetic alginate particles [66].

Xylan degradation gene frequencies were largely stable among FL bacteria but increased considerably in PA bacteria during the late diatom and *Chattonella* bloom phases, surpassing FL frequencies more than twofold. This likely reflects an increased availability of structural xylans from disintegrating algae as well as poor xylan solubilities. Studies on the diatom *Thalassiosira weissflogii* have shown that its xylans and mannans are primarily found in POM and only little in DOM [67], consistent with functions as cell wall polysaccharides [30, 68]. We

found xylan degradation genes in PA bacteria (e.g., *Collwellia*) and in FL bacteria (e.g., “Formosa”), but overall gene proportions suggest a higher proportion of xylan degrading bacteria among PA bacteria.

The bacterial cell wall polysaccharide peptidoglycan is also hardly soluble. However, peptidoglycan degradation gene [69] frequencies were substantially higher in FL than in PA bacteria, suggesting that peptidoglycan is rapidly solubilized and recycled. This is consistent with the fact that peptidoglycan is not known to significantly accumulate in POM [70].

Conversely, fucoidan utilization gene frequencies were consistently higher in PA than in FL communities. Apart from the well-known *Verrucomicrobiota* [71], we observed such genes also in *Polaribacter*-affiliating metagenome sequences during the diatom bloom. This was confirmed by the presence of fucoidan-targeting gene clusters in various PA MAGs, including *Polaribacter* MAG_186 and MAG_189, *Saprospiraceae* MAG_446 and MAG_449, *Planctomycetota* MAG_584, and *Lentimonas* MAG_693 (Fig. 7). *Polaribacter* MAG_189 was only abundant at the beginning of the diatom bloom, whereas *Polaribacter* MAG_186 prevailed during the diatom bloom, and the *Planctomycetota*, *Saprospiraceae*, and *Lentimonas* MAGs were most abundant during the *Phaeocystis* bloom. The presence of fucoidan throughout the bloom is plausible, since fucoidan-containing polysaccharides secreted by diatoms [67, 72] are rather persistent to bacterial degradation [73].

Genes for host glycan recognition, binding, and degradation were present in 33 of the 71 studied abundant MAGs, in particular in *Bacteroidota*, *Verrucomicrobiota*, and *Planctomycetota* (Fig. 7), and comprised GH92 (α -mannosidase) as well as GH20 and GH109 (β -1,6-N-acetylglucosaminidase) family genes. Host glycans are branched heteropolysaccharides that decorate eukaryotic host cell surfaces, e.g., mucin O-linked glycans, N-linked glycoproteins, and highly sulfated glycosaminoglycans (GAGs) in the human gut [74, 75]. Microalgae are also decorated with host glycans that are known to play a role in algal symbiont interactions with their hosts (e.g., [76]). In particular, mannose-rich N-glycans have been detected in microalgae [77, 78]. Binding to host glycans allows bacteria to initiate colonization of eukaryote surfaces, but host glycans also constitute an important substrate, not only for human gut bacteria but also for PA bacteria during phytoplankton blooms. For human gut *Bacteroides*, it has been shown that problematic antennary monosaccharides are removed from host glycans before uptake [75]. It is likely that such extracellular pre-digestion also occurs among marine bacteria, and if the selectively removed monosaccharides can diffuse away, this would explain the source of soluble sulfated

methylpentoses (fucose, rhamnose) that constitute a preferential substrate for recurring small-celled FL *Verrucomicrobiota* at Helgoland Roads [79].

The highly adaptable CAZome

CAZyme repertoires can vary considerably even between species of the same genus (e.g., [39]). For instance, *Winogradskyella* HQ MAG_139 harbored a much lower number of PULs and CAZyme-rich gene clusters than MAG_137, even though the latter was of lesser quality (94% vs 71%, Fig. 7). *Lentimonas* represents another illustrative case with three HQ MAGs, of which only MAG_693 contained abundant CAZyme genes (Fig. 7). CAZyme gene and PUL repertoires thus confer information about the adaptation of a given species towards a specific polysaccharide niche rather than its overall phylogenetic position in the tree of life. This implicates that the process of polysaccharide niche adaptation must considerably outpace the evolution of novel species, possibly by frequent lateral gene transfer [80].

CAZyme-rich MAGs often exhibited higher relative abundances during the bloom than closely related ones with fewer CAZymes, particularly in the PA fractions. This trend was evident for MAGs of *Maribacter*, *Winogradskyella*, *Polaribacter*, *Lentimonas*, and the CL500-3 and BD1-7 clades. *Lentimonas* MAG_693 for example harbored CAZyme genes targeting a variety of polysaccharide substrates (Fig. 7). Corresponding ASV data confirmed that this MAG represented a distinctively PA-associated species, consistent with a previous study on *Lentimonas* [81]. In this study, we found additional preferentially FL *Lentimonas* species, highlighting a broader niche spectrum within members of this genus (Fig. 7). Similarly, members of *Polaribacter* usually exhibit high FL abundances during diatom-dominated blooms at Helgoland Roads [50]. *Polaribacter* MAG_183 in this study corresponds to an abundant FL *Polaribacter* (MAG P_MB288) that we observed during the Helgoland spring bloom in 2020 [39]. However, in contrast to 2020, the dominating *Polaribacter* during the 2018 bloom were distinct and exhibited a clear preference for PA communities. In accordance with general trends, the dominant PA *Polaribacter* featured a larger MAG (MAG_189) with a higher number of CAZyme genes.

Noteworthy clades with low CAZyme gene proportions

Members of the BD1-7 clade, unclassified *Stappiaceae*, and *Nitri-colaceae* featured high abundances but not CAZyme-rich MAGs. The gammaproteobacterial BD1-7 clade belongs to the group of Oligotrophic Marine *Gammaproteobacteria* (OMG) [82]. Members of this clade are known to associate with phytoplankton [83] as with other eukaryotes, such as sponges [84], corals [85], brown algae

[86], and squids [87], where they may exert symbiotic functions. Alphaproteobacterial *Stappiaceae* are related to the abundant *Roseobacteraceae*. The genera include bacteriochlorophyll-producing *Roseibium*, members of which have been isolated from red algae [88], corals [89], oysters [90], and dinoflagellates [91], possibly also in a symbiotic function. *Nitricolaceae* (formerly *Oceanospirillaceae*), abundant in PA3 communities, are opportunistic *Gammaproteobacteria* that frequently associate with phytoplankton (e.g., [92]), including *Reinekea* species, which we have observed in high abundance at Helgoland Roads before [42, 93]. Members of the *Nitricolaceae* ASP10-2a clade are known to be diatom-associated [94].

Bloom-associated PA bacteria and global carbon cycling

The bulk of particles during phytoplankton blooms are either formed directly by aggregation of algal necromass or indirectly via the formation and excretion of fecal pellets by grazing of small zooplankton. The latter have been estimated to consume almost two-thirds of the phytoplankton cells on a daily basis [13], which is why fecal pellets constitute high proportions of the POM during phytoplankton blooms. Copepods are abundant zooplankters and have short gut transit times (30 to 90 min [95, 96]), which is why their fecal pellets contain considerable proportions of only partially degraded microalgae [97]. A substantial part of the captured PA communities in our study thus represent primary or secondary fecal pellet colonizers that consume residual pelleted algal polysaccharides. According to recent estimates, phytoplankton-specific loss rates to zooplankton grazing constitute the greatest uncertainty in CMIP6 marine biogeochemical models used to assess bacterial remineralization versus sequestration rates of algal biomass. These uncertainties range in the gigatons of carbon per year [98], which is substantial considering that recent anthropogenic carbon emissions have been estimated at around 10 gigatons per year (IPCC for the year 2018).

Concluding remarks

Marine bacteria that colonize suspended particles inhabit a much more diverse habitat with ampler niche spaces and closer interactions than free-floating bacteria in the water column. However, to obtain quantitative data on bacterial polysaccharide degradation on particles in situ poses a considerable challenge. This applies in particular to reliable biochemical data on polysaccharide turnover rates, precise cell counts of PA bacteria, and even precise PA bacterial diversities owing to high proportions of chloroplast sequences in corresponding 16S rRNA gene amplicon data. Also, to obtain corresponding sufficiently deep metaproteome data remains a challenge. Finally,

our sampling method does neither allow to discriminate different types of particles apart from broad size ranges nor to discriminate between loosely particle-associated and truly particle-attached bacteria—a limitation that we have discussed in detail in a previous study on the diversity, isolation, and cultivation of PA bacteria during the 2018 Helgoland spring bloom [21].

These challenges notwithstanding, we could demonstrate that PA bacterial communities were more diverse and underwent more dynamic changes in response to the 2018 spring phytoplankton bloom at Helgoland Roads than their FL counterparts. PA communities also featured a substantially higher metabolic potential for the degradation of a wide variety of polysaccharides. This was not only evident from assembled metagenome data but also in representative MAGs of abundant and active species. In the aforementioned study [21], we have also shown that PA bacteria represented less than 1% of the total bacterial community during spring 2018 at Helgoland Roads. However, considering that a major proportion of algal necromass passes through the POM pool, these bacteria must act as gatekeepers for the solubilization and subsequent remineralization of significant, yet-to-be-quantified proportions of algal polysaccharides during and after phytoplankton blooms, in spite of being considerably outnumbered by FL bacteria.

Materials and methods

Sampling, physicochemical, and phytoplankton data

Seawater samples were collected during spring 2018 (March 1 to May 29) off the North Sea island Helgoland (German Bight) at the LTER site “Kabeltonne” (54° 11.3' N, 7° 54.0' E, DEIMS.iD: <https://deims.org/1e96f9b-0915-4661-849f-b3a72f5aa9b1>) by fractionating filtration (FL, 0.2–3 µm; PA3, 3–10 µm; PA10, > 10 µm) as described previously [42] (see Additional file 3 for details).

Wind direction data were obtained from the Climate Data Store of the Copernicus Climate Change Service [99]. Other physicochemical data, such as Secchi depth, water temperature, salinity, chlorophyll *a* content, dissolved inorganic nitrogen (NO_2^- , NO_3^- , NH_4^+), silicate, and phosphate as well as microscopic algae and zooplankton counts and taxonomic classifications were obtained as part of the Helgoland Roads LTER time series [100, 101]. Biovolumes of abundant plankters were determined in the framework of the Sylt Roads time series [102] as described elsewhere [103]. These data are summarized in Additional file 1: Table S1. Both the Helgoland and Sylt Roads time series are conducted by the Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research (Bremerhaven, Germany).

16S and 18S rRNA gene amplicon sequencing and analysis

Sequencing of 16S rRNA gene amplicons was performed at the Max Planck Genome Centre Cologne (Germany). DNA from biomass retained on filters was extracted as described elsewhere [21] and amplified using primers 341F and 805R targeting the V3 and V4 regions [104] for the FL samples, and primers 515F and 806R targeting the V4 region [105] for the PA3 and PA10 samples. Sequencing was carried out on an Illumina HiSeq 2500 (Illumina, San Diego, CA, USA) in rapid mode with 2×250 bp paired-end reads.

Sequences were analyzed for single nucleotide-resolved amplicon sequence variants (ASVs) using the DADA2 v1.19.2 package [106] with R v4.0.3 (<http://www.R-project.org>) (Additional file 3). ASVs assigned to chloroplasts, mitochondria, *Eukarya*, *Archaea*, or unclassified sequences were excluded from further analyses (Additional file 2: Fig. S6B). Possible impacts of the different primer sets and the omission of rarefaction are provided in Additional file 3, as well as details on the analysis of the 18S rRNA amplicon data.

Metagenome sequencing and assembly

Metagenomes were sequenced at the Max Planck Genome Centre Cologne, 34 on an Illumina HiSeq 2500 using 2×150 bp chemistry, and eight additional PA3 metagenomes on a PacBio Sequel II (Menlo Park, CA, USA) using one SMRT cell per sample in long-read HiFi mode. The quality of Illumina reads was assessed with FastQC v0.11.9 [107].

Quality-filtered reads from FL metagenomes were assembled individually within SPAdes v3.11.1 [108]. Quality-filtered reads from PA3 and PA10 Illumina metagenomes were assembled individually using MEGAHIT v1.2.9 [109]. Assemblies of PA3 PacBio metagenomes were generated using Flye v2.9.1 [110] (Additional file 3). Assembly quality was assessed with QUAST [111]. Contigs below 2.5 kbp were removed using *anvi-script-reformat-fasta* within *anvi'o* v6.2 [112].

Metagenome-assembled genome (MAG) retrieval and analysis

MAG retrieval was performed as described previously [43] (Additional file 3). MAGs were classified into low-, medium-, and high-quality categories according to the criteria described in Bowers et al. [47] using CheckM v1.1.3 [113]. Only medium- and high-quality MAGs were used in further analyses. Dereplication was done using dRep v3.0.0 [114] with an average nucleotide identity (ANI) > 95%. ANI was calculated using FastANI [115]. MAG abundances were calculated as described previously [116] (Additional file 3). 16S rRNA gene sequences

were extracted from MAGs using *barrnap* v0.9 (<https://github.com/tseemann/barrnap>) and subsequently classified in the Silva Incremental Aligner (SINA) with Silva SSU 138.1 taxonomy [117]. MAG taxonomies were determined by GTDB-Tk v2.1.0 [118, 119] with GTDB release R207_v2. Differences in the denominations of taxa between Silva and GTDB were resolved as described previously [44]. A phylogenomic tree of dereplicated MAGs was constructed using FastTree [120] from within *anvi'o* v7.1 and visualized using interactive Tree of Life (iTol) v6.5.6 [121].

Interrelation of 16S rRNA gene amplicon and MAG data

Blastn was used to search all prevalent ASVs from abundant genera within the 16S rRNA gene amplicon dataset against all MAG-derived 16S rRNA gene sequences. For identical hits with 100% coverage, we assumed that changes in ASV relative abundance reflected changes of the corresponding MAG over time. Since not all MAGs contained 16S rRNA genes, we extended our search to all MAGs that we obtained from the Helgoland metagenome samples from 2010 [44], 2012 [44], 2016 [43, 44], and 2020 [39]. MAGs from 2018 without 16S rRNA gene sequence were considered to match MAGs from other sampling years, if both exhibited an ANI of at least 95%. In addition, we included two matching MAGs from the GTDB database. Details are provided in Additional file 1: Table S10.

Metagenome and MAG annotation

For assembled Illumina metagenomes, protein-coding sequences were predicted using Prodigal [122], Aragorn [123], and *barrnap* (<https://github.com/tseemann/barrnap>) as implemented in Prokka v1.14.6 [124] (default settings). For PacBio data, FragGeneScan v1.31 [125] was used (setting *-w 1*) due to a higher number of frameshifts. Functional MAG annotations were done as described in Additional file 3.

Gene frequency analyses

Eukaryotic and unclassified reads were removed from unassembled Illumina metagenomes according to Kaiju v1.9.0 [126] annotations. Metagenomes were subsequently assembled with MEGAHIT v1.2.9, and frequencies of genes of interest were computed for each metagenome as follows: gene frequency = (sum of average coverage of target gene(s)) \times 100 / (sum of average coverage of all genes) [42]. The average coverages of target genes were determined in SqueezeMeta v1.3.1 [127] using bowtie2 [128] for mapping. CAZymes were predicted as described in Additional file 3.

Prediction of CAZyme-rich gene clusters and PULs

CAZyme-rich gene clusters and PULs were identified in a sliding window approach as described previously [40, 43] with a window length of ten genes. When at least three genes within the window coded for either GHs, PLs, CEs, sulfatase, TBDTs, or SusD-like proteins, we considered this a candidate locus. The resulting CAZyme-rich loci were manually annotated based on a combination of multiple databases (Additional file 3). Putative target substrate classes of PULs, PUL-like, and CAZyme-rich gene clusters in MAGs were predicted using the dbCAN3-sub database [129].

Metaproteome analyses

Metaproteomes were analyzed on seven dates for FL samples (2018/03/20, 2018/04/12, 2018/04/17, 2018/04/26, 2018/05/08, 2018/05/22, 2018/05/24) and on three dates for PA samples (2018/04/17, 2018/05/08, 2018/05/24). Proteins were extracted from filtered biomass and subsequently analyzed as described elsewhere [43, 130, 131] (Additional file 3).

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40168-024-01757-5>.

Additional file 1: Supplementary tables.

Additional file 2: Supplementary figures.

Additional file 3: Supplementary text.

Acknowledgements

We acknowledge the Helgoland sampling team, including Lily Franzmeyer, Mirja Meiners, Sabine Kühn, Peter Rücknagel, Jörg Wulf, Nina Heinzmann, Greta Giljan, Jan Brüwer, and Anneke Heins. We also acknowledge the BAH team, including Eva Maria Brodte, Antje Wichels, and FS Aade and FS Uthörn captains and crews for the help with sampling, analyses, logistics, and provision of lab space. Furthermore, we thank our colleagues from the Max Planck Genome Centre Cologne for their assistance in metagenomics. Further thanks go to Lilly Franzmeyer, Thomas Sura, Daniela Zühlke, Doreen Schultz, Katharina Riedel, and Pierre-Alexander Mücke from the University of Greifswald for the technical assistance in MS-based metaproteomics. Finally, we thank our former colleagues Karen Krüger and T. Ben Francis for the assistance in MAG CAZyme analysis and Luis H. Orellana for the MAG abundance calculations. Fengqing Wang is a member of the International Max Planck Research School of Marine Microbiology (MarMic).

Authors' contributions

F.W., C.S., and D.L.: bioinformatics and data integration and analysis. D.B., R.S., A.T.S., and D.B.: metaproteomics. B.H.: sequencing. J.R.: algal biovolumes. I.V.K. and K.H.W.: Helgoland Roads time series data. B.M.F.: sampling logistics. M.M.B.: 16S and 18S rRNA gene amplicon sampling and analysis. F.W., C.S., D.L., and H.T.: compilation of the manuscript. B.M.F., T.S., M.M.B., H.T., and R.I.A.: study design. All authors read, discussed, and approved the final manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL. This study was funded by the Max Planck Society and supported by the German Research Foundation (DFG) in the framework of the research unit FOR2406 "Proteogenomics of Marine Polysaccharide Utilization (POMPU)" by grants of D.B. (BE

3869/4-3), B.M.F. (FU 627/2-3), T.S. (SCHW 595/10-3, SCHW 595/11-3), M.M.B. (RI 969/9-2), H.T. (TE 813/2-3), and R.I.A. (AM 73/9-3). The Helgoland time series is supported by the Biological Station Helgoland, Alfred Wegener Institute, and Helmholtz Center for Polar and Marine Research (AWI_BAH_o 1).

Availability of data and materials

Metagenome reads, assemblies, and MAGs were deposited in the European Nucleotide Archive (ENA) under project numbers PRJEB38290 and PRJEB67502. 16S rRNA gene amplicon sequences of FL and PA fractions were deposited in ENA under project numbers PRJEB51721 and PRJEB51816, respectively. Mass spectrometry proteome data were deposited at the ProteomeXchange Consortium via the PRIDE partner repository [132]. Original mass spectrometry proteome data of FL bacteria are accessible as project PXD042676 and data for PA bacteria as project PXD046705.

Declarations

Ethics approval and consent to participate

Ethics approval was not required for the study.

Competing interests

The authors declare no competing interests.

Author details

¹Max Planck Institute for Marine Microbiology, Celsiusstraße 1, 28359 Bremen, Germany. ²Institute of Pharmacy, University of Greifswald, Felix-Hausdorff-Straße 3, 17489 Greifswald, Germany. ³Institute of Marine Biotechnology, Walther-Rathenau-Straße 49a, 17489 Greifswald, Germany. ⁴Institute of Microbiology, University of Greifswald, Felix-Hausdorff-Straße 8, 17489 Greifswald, Germany. ⁵Marine College, Shandong University, Weihai 264209, China. ⁶Max Planck Genome Centre Cologne, Carl von Linné-Weg 10, 50829 Cologne, Germany. ⁷Alfred Wegener Institute for Polar and Marine Research, Biologische Anstalt Helgoland, Helgoland 27483, Germany.

Received: 7 November 2023 Accepted: 4 January 2024

Published online: 20 February 2024

References

- Field CB, Behrenfeld MJ, Randerson JT, Falkowski P. Primary production of the biosphere: integrating terrestrial and oceanic components. *Science*. 1998;281(5374):237–40.
- Falkowski PG, Barber RT, Smetacek VV. Biogeochemical controls and feedbacks on ocean primary production. *Science*. 1998;281(5374):200–6.
- Mann DG. The species concept in diatoms: evidence for morphologically distinct, sympatric gamodemes in four epipelagic species. *Plant Syst Evol*. 1989;164:215–37.
- Inomura K, Karlusich JJP, Dutkiewicz S, Deutsch C, Harrison PJ, Bowler C. High growth rate of diatoms explained by reduced carbon requirement and low energy cost of silica deposition. *Microbiol Spectr*. 2023:e03311–22.
- Arrigo KR, Robinson DH, Worthen DL, Dunbar RB, DiTullio GR, VanWoert M, et al. Phytoplankton community structure and the drawdown of nutrients and CO₂ in the southern ocean. *Science*. 1999;283(5400):365–7.
- Lancelot C, Gypens N, Billen G, Garnier J, Roubeix V. Testing an integrated river-ocean mathematical tool for linking marine eutrophication to land use: the *Phaeocystis*-dominated Belgian coastal zone (Southern North Sea) over the past 50 years. *J Mar Syst*. 2007;64:216–28.
- Alderkamp AC, Buma AGJ, van Rijssel M. The carbohydrates of *Phaeocystis* and their degradation in the microbial food web. *Biogeochemistry*. 2007;83:99–118.
- Vincent F, Gralka M, Schleyer G, Schatz D, Cabrera-Brufau M, Kuhlisch C, et al. Viral infection switches the balance between bacterial and

- eukaryotic recyclers of organic matter during coccolithophore blooms. *Nat Commun.* 2023;14(1):510.
9. Zhang ZH, Li DH, Xie RZ, Guo RY, Nair S, Han H, et al. Plastoquinone synthesis inhibition by tetrabromo biphenyldiol as a widespread algalicidal mechanism of marine bacteria. *ISME J.* 2023;17(11):1979–92.
 10. Scholz B, Guillou L, Marano AV, Neuhauser S, Sullivan BK, Karsten U, et al. Zoospore parasites infecting marine diatoms - a black box that needs to be opened. *Fungal Ecol.* 2016;19:59–76.
 11. Garvetto A, Nézan E, Badis Y, Bilién G, Arce P, Bresnan E, et al. Novel widespread marine oomycetes parasitising diatoms, including the toxic genus *Pseudo-nitzschia*: genetic, morphological, and ecological characterisation. *Front Microbiol.* 2018;9:2918.
 12. Fernández-Méndez M, Wenzhöfer F, Peeken I, Sørensen HL, Glud RN, Boëtius A. Composition, buoyancy regulation and fate of ice algal aggregates in the Central Arctic Ocean. *PLoS ONE.* 2014;9(9):e107452.
 13. Schmoker C, Hernández-León S, Calbet A. Microzooplankton grazing in the oceans: impacts, data variability, knowledge gaps and future directions. *J Plankton Res.* 2013;35(4):691–706.
 14. Worden AZ, Follows MJ, Giovannoni SJ, Wilken S, Zimmerman AE, Keeling PJ. Rethinking the marine carbon cycle: factoring in the multifarious lifestyles of microbes. *Science.* 2015;347(6223):1257594.
 15. Vargas CA, Cuevas LA, González HE, Daneri G. Bacterial growth response to copepod grazing in aquatic ecosystems. *J Mar Biol Assoc UK.* 2007;87(3):667–74.
 16. Giering SL, Sanders R, Lampitt RS, Anderson TR, Tamburini C, Boutfir M, et al. Reconciliation of the carbon budget in the ocean's twilight zone. *Nature.* 2014;507(7493):480–3.
 17. De La Rocha CL, Passow U. Factors influencing the sinking of POC and the efficiency of the biological carbon pump. *Deep Sea Research Part II: Topical Studies in Oceanography.* 2007;54(5–7):639–58.
 18. Nowicki M, DeVries T, Siegel DA. Quantifying the carbon export and sequestration pathways of the ocean's biological carbon pump. *Glob Biogeochem Cycles.* 2022;36:e2021GB007083.
 19. Simon M, Grossart HP, Schweitzer B, Ploug H. Microbial ecology of organic aggregates in aquatic ecosystems. *Aquat Microb Ecol.* 2002;28(2):175–211.
 20. Turley CM, Stutt ED. Depth-related cell-specific bacterial leucine incorporation rates on particles and its biogeochemical significance in the Northwest Mediterranean. *Limnol Oceanogr.* 2000;45(2):419–25.
 21. Heins A, Reintjes G, Amann R, Harder J. Particle collection in Imhoff sedimentation cones enriches both motile chemotactic and particle-attached bacteria. *Front Microbiol.* 2021;12:643730.
 22. Azúa I, Unanue M, Ayo B, Artolozaga I, Arrieta JM, Iriberrí J. Influence of organic matter quality in the cleavage of polymers by marine bacterial communities. *J Plankton Res.* 2003;25(12):1451–60.
 23. Leu AO, Eppley JM, Burger A, DeLong EF. Diverse genomic traits differentiate sinking-particle-associated versus free-living microbes throughout the oligotrophic open ocean water column. *mBio.* 2022;13(4):e01569–22.
 24. Patel AK, Vadrade AP, Singhania RR, Michaud P, Pandey A, Chen SJ, et al. Algal polysaccharides: current status and future prospects. *Phytochem Rev.* 2023;22:1167–96.
 25. Myklesstad S. Production of carbohydrates by marine planktonic diatoms. I. comparison of nine different species in culture. *J Exp Mar Biol Ecol.* 1974;15(3):261–74.
 26. Chen J, Yang J, Du H, Aslam M, Wang W, Chen W, et al. Laminarin, a major polysaccharide in stramenopiles. *Mar Drugs.* 2021;19(10):576.
 27. Zvyagintseva TN, Shevchenko NM, Popivnich IB, Isakov VV, Scobun AS, Sundukova EV, et al. A new procedure for the separation of water-soluble polysaccharides from brown seaweeds. *Carbohydr Res.* 1999;322(1–2):32–9.
 28. Myklesstad SM. Production, chemical structure, metabolism, and biological function of the (1→3)-linked, β3-D-glucans in diatoms. *Biol Oceanogr.* 1989;6(3–4):313–26.
 29. Becker S, Tebben J, Coffinet S, Wiltshire K, Iversen MH, Harder T, et al. Laminarin is a major molecule in the marine carbon cycle. *Proc Natl Acad Sci USA.* 2020;117(12):6599–607.
 30. Le Costaouëc T, Unamunzaga C, Mantecon L, Helbert W. New structural insights into the cell-wall polysaccharide of the diatom *Phaeodactylum tricornutum*. *Algal Res.* 2017;26:172–9.
 31. Gügi B, Le Costaouëc T, Burel C, Lerouge P, Helbert W, Bardor M. Diatom-specific oligosaccharide and polysaccharide structures help to unravel biosynthetic capabilities in diatoms. *Mar Drugs.* 2015;13(9):5993–6018.
 32. Mühlenthaler M, Grossart HP, Eigemann F, Voss M. Mini-review: Phytoplankton-derived polysaccharides in the marine environment and their interactions with heterotrophic bacteria. *Environ Microbiol.* 2018;20(8):2671–85.
 33. Shnyukova EI, Zolotariova YK. Ecological role of exopolysaccharides of *Bacillariophyta*: a review. *Int J Algae.* 2017;19(1):5–20.
 34. Thornton DCO. Diatom aggregation in the sea: mechanisms and ecological implications. *Eur J Phycol.* 2002;37(2):149–61.
 35. Babiak W, Krzemińska I. Extracellular polymeric substances (EPS) as microalgal bioproducts: a review of factors affecting EPS synthesis and application in flocculation processes. *Energies.* 2021;14(13):4007.
 36. Martino PD. Extracellular polymeric substances, a key element in understanding biofilm phenotype. *Aims Microbiol.* 2018;4(2):274–88.
 37. Glenwright AJ, Pothula KR, Bhamidimarri SP, Chorev DS, Baslé A, Firbank SJ, et al. Structural basis for nutrient acquisition by dominant members of the human gut microbiota. *Nature.* 2017;541(7637):407–11.
 38. White JBR, Silale A, Feasey M, Heunis T, Zhu Y, Zheng H, et al. Outer membrane utilosomes mediate glycan uptake in gut Bacteroidetes. *Nature.* 2023;618(7965):583–9.
 39. Sidhu C, Kirstein IV, Meunier CL, Rick J, Fofonova V, Wiltshire KH, et al. Dissolved storage glycans shaped the community composition of abundant bacterioplankton clades during a North Sea spring phytoplankton bloom. *Microbiome.* 2023;11(1):77.
 40. Lu D, Wang F, Amann RI, Teeling H, Du JZ. Epiphytic common core bacteria in the microbiomes of co-located green (*Ulva*), brown (*Saccharina*) and red (*Grateloupia*, *Gelidium*) macroalgae. *Microbiome.* 2023;11(1):126.
 41. Teeling H, Fuchs BM, Becher D, Klockow C, Gardebrecht A, Bennis CM, et al. Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. *Science.* 2012;336(6081):608–11.
 42. Teeling H, Fuchs BM, Bennis CM, Krüger K, Chafee M, Kappelmann L, et al. Recurring patterns in bacterioplankton dynamics during coastal spring algae blooms. *eLife.* 2016;5:e11888.
 43. Francis TB, Bartosik D, Sura T, Sichert A, Hehemann JH, Markert S, et al. Changing expression patterns of TonB-dependent transporters suggest shifts in polysaccharide consumption over the course of a spring phytoplankton bloom. *ISME J.* 2021;15(8):2336–50.
 44. Krüger K, Chafee M, Francis TB, del Rio TG, Becher D, Schweder T, et al. In marine *Bacteroidetes* the bulk of glycan degradation during algae blooms is mediated by few clades using a restricted set of genes. *ISME J.* 2019;13(11):2800–16.
 45. Francis TB, Krüger K, Fuchs BM, Teeling H, Amann RI. *Candidatus* *Prosilicoccus vernus*, a spring phytoplankton bloom associated member of the *Flavobacteriaceae*. *Syst Appl Microbiol.* 2019;42(1):41–53.
 46. Okazaki Y, Fujinaga S, Tanaka A, Kohzu A, Oyagi H, Nakano S. Ubiquity and quantitative significance of bacterioplankton lineages inhabiting the oxygenated hypolimnion of deep freshwater lakes. *ISME J.* 2017;11(10):2279–93.
 47. Bowers RM, Kyrpidis NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol.* 2017;35(8):725–31.
 48. Almeida A, Mitchell AL, Boland M, Forster SC, Gloor GB, Tarkowska A, et al. A new genomic blueprint of the human gut microbiota. *Nature.* 2019;568(7753):499–504.
 49. Kavagutti VS, Bulzu PA, Chiriac CM, Salcher MM, Mukherjee I, Shabarova T, et al. High-resolution metagenomic reconstruction of the freshwater spring bloom. *Microbiome.* 2023;11(1):15.
 50. Avci B, Krüger K, Fuchs BM, Teeling H, Amann RI. Polysaccharide niche partitioning of distinct *Polaribacter* clades during North Sea spring algal blooms. *ISME J.* 2020;14(6):1369–83.
 51. Hahnke RL, Bennis CM, Fuchs BM, Mann AJ, Rhiel E, Teeling H, et al. Dilution cultivation of marine heterotrophic bacteria abundant after a spring phytoplankton bloom in the North Sea. *Environ Microbiol.* 2015;17(10):3515–26.

52. Alonso C, Warnecke F, Amann R, Pernthaler J. High local and global diversity of *Flavobacteria* in marine plankton. *Environ Microbiol.* 2007;9(5):1253–66.
53. Chafee M, Fernández-Guerra A, Buttigieg PL, Gerds G, Eren AM, Teeling H, et al. Recurrent patterns of microdiversity in a temperate coastal marine environment. *ISME J.* 2018;12(1):237–52.
54. Williams TJ, Wilkins D, Long E, Evans F, DeMaere MZ, Raftery MJ, et al. The role of planktonic *Flavobacteria* in processing algal organic matter in coastal East Antarctica revealed using metagenomics and metaproteomics. *Environ Microbiol.* 2013;15(5):1302–17.
55. Jeong SE, Kim KH, Baek K, Jeon CO. *Parasphingopyxis algicola* sp. nov., isolated from a marine red alga *Asparagopsis taxiformis* and emended description of the genus *Parasphingopyxis* Uchida et al. 2012. *Int J Syst Evol Microbiol.* 2017;67(10):3877–81.
56. Zhang XL, Qi M, Li QH, Cui ZD, Yang Q. *Maricaulis alexandrii* sp. nov., a novel active biofloculants-bearing and dimorphic prosthecate bacterium isolated from marine phycosphere. *Antonie Van Leeuwenhoek.* 2021;114(8):1195–203.
57. Yoo JH, Han JE, Lee JY, Jeong SW, Jeong YS, Lee JY, et al. *Parasphingogorhabdus cellanae* sp. nov., isolated from the gut of a Korean limpet, *Cellana toreuma*. *Int J Syst Evol Microbiol.* 2022;72(8):005470.
58. Abraham WR, Strömpl C, Meyer H, Lindholm S, Moore ERB, Christ R, et al. Phylogeny and polyphasic taxonomy of *Caulobacter* species. Proposal of *Maricaulis* gen. nov. with *Maricaulis maris* (Poindexter) comb. nov. as the type species, and emended description of the genera *Brevundimonas* and *Caulobacter*. *Int J Syst Evol Microbiol.* 1999;49(3):1053–73.
59. Abraham WR, Rohde M. The family *Hyphomonadaceae*. In: Rosenberg E, EF DL, Lory S, Stackebrandt E, Thompson F, editors. *The prokaryotes: Alphaproteobacteria and Betaproteobacteria*. Berlin, Heidelberg: Springer Berlin Heidelberg. 2014:p. 283–99.
60. Löder MGJ, Kraberg AC, Aberle N, Peters S, Wiltshire KH. Dinoflagellates and ciliates at Helgoland Roads North Sea. *Helgoland Mar Res.* 2012;66:11–23.
61. Weisse T, Tande K, Verity P, Hansen F, Gieskes W. The trophic significance of *Phaeocystis* blooms. *J Mar Syst.* 1994;5(1):67–79.
62. Schnepf E, Kühn SF. Food uptake and fine structure of *Cryptothecomonas longipes* sp. nov., a marine nanoflagellate invertebrate feeding phagotrophically on large diatoms. *Helgoland Mar Res.* 2000;54(1):18–32.
63. Francis TB, Urich T, Mikolasch A, Teeling H, Amann R. North Sea spring bloom-associated *Gammaproteobacteria* fill diverse heterotrophic niches. *Environ Microbiome.* 2021;16:15.
64. Mareček F, Møller MS, Svensson B, Janeček Š. A putative novel starch-binding domain revealed by in silico analysis of the N-terminal domain in bacterial amylomaltses from the family GH77. *3 Biotech.* 2021;11(5):229.
65. Koch H, Dürwald A, Schweder T, Noriega-Ortega B, Vidal-Melgosa S, Hehemann JH, et al. Biphasic cellular adaptations and ecological implications of *Alteromonas macleodii* degrading a mixture of algal polysaccharides. *ISME J.* 2019;13(1):92–103.
66. Bunse C, Koch H, Breider S, Simon M, Wietz M. Sweet spheres: succession and CAZyme expression of marine bacterial communities colonizing a mix of alginate and pectin particles. *Environ Microbiol.* 2021;23(6):3130–48.
67. Huang GY, Vidal-Melgosa S, Sichert A, Becker S, Fang Y, Niggemann J, et al. Secretion of sulfated fucans by diatoms may contribute to marine aggregate formation. *Limnol Oceanogr.* 2021;66(10):3768–82.
68. Hecky RE, Mopper K, Kilham P, Degens ET. The amino acid and sugar composition of diatom cell-walls. *Mar Biol.* 1973;19(4):323–31.
69. Humann J, Lenz LL. Bacterial peptidoglycan-degrading enzymes and their impact on host muropeptide detection. *J Innate Immun.* 2009;1(2):88–97.
70. Kitayama K, Hama T, Yanagi K. Bioreactivity of peptidoglycan in seawater. *Aquat Microb Ecol.* 2007;46:85–93.
71. Sichert A, Corzett CH, Schechter MS, Unfried F, Markert S, Becher D, et al. *Verrucomicrobia* use hundreds of enzymes to digest the algal polysaccharide fucoidan. *Nat Microbiol.* 2020;5(8):1026–39.
72. Vidal-Melgosa S, Sichert A, Francis TB, Bartosik D, Niggemann J, Wichels A, et al. Diatom fucan polysaccharide precipitates carbon during algal blooms. *Nat Commun.* 2021;12(1):1150.
73. Bligh M, Nguyen N, Buck-Wiese H, Vidal-Melgosa S, Hehemann JH. Structures and functions of algal glycans shape their capacity to sequester carbon in the ocean. *Curr Opin Chem Biol.* 2022;71:102204.
74. Brown HA, Koropatkin NM. Host glycan utilization within the *Bacteroidetes* Sus-like paradigm. *Glycobiology.* 2021;31(6):697–706.
75. Briiliūtė J, Urbanowicz PA, Luis AS, Baslé A, Paterson N, Rebello O, et al. Complex N-glycan breakdown by gut *Bacteroides* involves an extensive enzymatic apparatus encoded by multiple co-regulated genetic loci. *Nat Microbiol.* 2019;4(9):1571–81.
76. Tivey TR, Parkinson JE, Mandelare PE, Adpressa DA, Peng W, Dong X, et al. N-linked surface glycan biosynthesis, composition, inhibition, and function in cnidarian-dinoflagellate symbiosis. *Microb Ecol.* 2020;80(1):223–36.
77. Baiet B, Burel C, Saint-Jean B, Louvet R, Menu-Bouaouiche L, Kiefer-Meyer MC, et al. N-glycans of *Phaeodactylum tricornutum* diatom and functional characterization of its N-acetylglucosaminyltransferase I enzyme. *J Biol Chem.* 2011;286(8):6152–64.
78. Mócsai R, Figl R, Troschl C, Strasser R, Svehla E, Windwarder M, et al. N-glycans of the microalga *Chlorella vulgaris* are of the oligomannosidic type but highly methylated. *Sci Rep.* 2019;9(1):331.
79. Orellana LH, Francis TB, Ferraro M, Hehemann JH, Fuchs BM, Amann R. *Verrucomicrobiota* are specialist consumers of sulfated methyl pentoses during diatom blooms. *ISME J.* 2022;16(3):630–41.
80. Hehemann JH, Corec G, Barbeyron T, Helbert W, Czejek M, Michel G. Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. *Nature.* 2010;464(7290):908–12.
81. Ren YH, Luo ZH, Liu Q, Wei B, Wu YH, Shu WS, et al. Insights into community assembly mechanisms, biogeography, and metabolic potential of particle-associated and free-living prokaryotes in tropical oligotrophic surface oceans. *Front Mar Sci.* 2022;9:923295.
82. Cho JC, Giovannoni SJ. Cultivation and growth characteristics of a diverse group of oligotrophic marine *Gammaproteobacteria*. *Appl Environ Microbiol.* 2004;70(1):432–40.
83. Wemheuer B, Güllert S, Billerbeck S, Giebel HA, Voget S, Simon M, et al. Impact of a phytoplankton bloom on the diversity of the active bacterial community in the southern North Sea as revealed by metatranscriptomic approaches. *FEMS Microbiol Ecol.* 2014;87(2):378–89.
84. Holert J, Cardenas E, Bergstrand LH, Zaikova E, Hahn AS, Hallam SJ, et al. Metagenomes reveal global distribution of bacterial steroid catabolism in natural, engineered, and host environments. *mBio.* 2018;9(1):e02345–17.
85. Yu XP, Yu KF, Liao ZH, Chen B, Deng CQ, Yu JY, et al. Seasonal fluctuations in symbiotic bacteria and their role in environmental adaptation of the scleractinian coral *Acropora pruinosa* in high-latitude coral reef area of the South China Sea. *Sci Total Environ.* 2021;792:148438.
86. Paix B, Othmani A, Debroas D, Culioli G, Briand JF. Temporal covariation of epibacterial community and surface metabolome in the Mediterranean seaweed holobiont *Taonia atomaria*. *Environ Microbiol.* 2019;21(9):3346–63.
87. Hu XJ, Su HC, Zhang P, Chen ZZ, Xu Y, Xu WJ, et al. Microbial community characteristics of the intestine and gills of medium-form populations of *Sthenoteuthis oualaniensis* in the South China Sea. *Fishes.* 2022;7(4):191.
88. Suzuki T, Muroga Y, Takahama M, Nishimura Y. *Roseigium denhamense* gen. nov., sp. nov. and *Roseibium hemelinense* sp. nov., aerobic bacteriochlorophyll-containing bacteria isolated from the east and west coasts of Australia. *Int J Syst Evol Microbiol.* 2000;50(6):2151–6.
89. Couceiro JF, Keller-Costa T, Marques M, Kyrpidis NC, Woyke T, Whitman WB, et al. The *Roseibium album* (*Labrenzia alba*) genome possesses multiple symbiosis factors possibly underpinning host-microbe relationships in the marine benthos. *Microbiol Resour Announc.* 2021;10(34):e0032021.
90. Karimi E, Keller-Costa T, Slaby BM, Cox CJ, da Rocha UN, Hentschel U, et al. Genomic blueprints of sponge-prokaryote symbiosis are shared by low abundant and cultivatable *Alphaproteobacteria*. *Sci Rep.* 2019;9(1):1999.
91. Aguirre EG, Carlson HK, Kenkel CD. Complete genome sequence of *Roseibium* sp. strain Sym1, a bacterial associate of *Symbiodinium linucheae*, the microalgal symbiont of the anemone *Aiptasia*. *Microbiol Resour Announc.* 2023;12(3):e0111822.

92. Thiele S, Vader A, Thomson S, Saubrekka K, Petelenz E, Armo HR, et al. The summer bacterial and archaeal community composition of the northern Barents Sea. *Prog Oceanogr*. 2023;103054.
93. Avci B, Hahnke RL, Chafee M, Fischer T, Gruber-Vodicka H, Tegetmeyer HE, et al. Genomic and physiological analyses of *Reinekea forsetii* reveal a versatile opportunistic lifestyle during spring algae blooms. *Environ Microbiol*. 2017;19(3):1209–21.
94. Bertrand EM, McCrow JP, Moustafa A, Zheng H, McQuaid JB, Delmont TO, et al. Phytoplankton–bacterial interactions mediate micronutrient colimitation at the coastal Antarctic sea ice edge. *Proc Natl Acad Sci USA*. 2015;112(32):9938–43.
95. Mayzaud P, Tirelli V, Bernard JM, Roche-Mayzaud O. The influence of food quality on the nutritional acclimation of the copepod *Acartia clausi*. *J Mar Syst*. 1998;15(1–4):483–93.
96. Tirelli V, Mayzaud P. Relationship between functional response and gut transit time in the calanoid copepod *Acartia clausi*: role of food quantity and quality. *J Plankton Res*. 2005;27(6):557–68.
97. Köster M, Sietmann R, Meuche A, Paffenhöfer GA. The ultrastructure of a doliolid and a copepod fecal pellet. *J Plankton Res*. 2011;33(10):1538–49.
98. Rohr T, Richardson AJ, Lenton A, Chamberlain MA, Shadwick EH. Zooplankton grazing is the largest source of uncertainty for marine carbon cycling in CMIP6 models. *Commun Earth Environ*. 2023;4(1):212.
99. Hersbach H, Bell B, Berrisford P, Biavati G, Horányi A, Muñoz SJ, et al. ERA5 hourly data on single levels from 1940 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). <https://doi.org/10.24381/cds.adbb2d47>.
100. Wiltshire KH, Kraberg A, Bartsch I, Boersma M, Franke HD, Freund J, et al. Helgoland Roads, North Sea: 45 years of change. *Estuar Coast*. 2010;33(2):295–310.
101. Kraberg A, Kieb U, Peters S, Wiltshire KH. An updated phytoplankton check-list for the Helgoland Roads time series station with eleven new records of diatoms and dinoflagellates. *Helgoland Mar Res*. 2019;73:9.
102. Armonies W, Asmus H, Buschbaum C, Lackschewitz D, Reise K, Rick J. Microscopic species make the diversity: a checklist of marine flora and fauna around the Island of Sylt in the North Sea. *Helgoland Mar Res*. 2018;72(1):11.
103. Hillebrand H, Dürsel CD, Kirschtel D, Pollinger U, Zohary T. Biovolume calculation for pelagic and benthic microalgae. *J Phycol*. 1999;35(2):403–24.
104. Herlemann DPR, Labrenz M, Jürgens K, Bertilsson S, Waniek JJ, Andersson AF. Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *ISME J*. 2011;5(10):1571–9.
105. Lucas J, Wichels A, Teeling H, Chafee M, Scharfe M, Gerds G. Annual dynamics of North Sea bacterioplankton: seasonal variability superimposes short-term variation. *Fems Microbiol Ecol*. 2015;91(9):fiv099.
106. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat Methods*. 2016;13(7):581–3.
107. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010: Available online at: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
108. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*. 2012;19(5):455–77.
109. Li DH, Luo RB, Liu CM, Leung CM, Ting HF, Sadakane K, et al. MEGAHIT v1.0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods*. 2016;102:3–11.
110. Kolmogorov M, Bickhart DM, Behsas B, Gurevich A, Rayko M, Shin SB, et al. metaFlye: scalable long-read metagenome assembly using repeat graphs. *Nat Methods*. 2020;17(11):1103–10.
111. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics*. 2013;29(8):1072–5.
112. Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, et al. Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ*. 2015;3:e1319.
113. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res*. 2015;25(7):1043–55.
114. Olm MR, Brown CT, Brooks B, Banfield JF. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J*. 2017;11(12):2864–8.
115. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun*. 2018;9(1):5114.
116. Orellana LH, Francis TB, Krüger K, Teeling H, Müller MC, Fuchs BM, et al. Niche differentiation among annually recurrent coastal Marine Group II Euryarchaeota. *ISME J*. 2019;13(12):3024–36.
117. Pruesse E, Peplies J, Glöckner FO. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics*. 2012;28(14):1823–9.
118. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics*. 2020;36(6):1925–7.
119. Parks DH, Chuvochina M, Chaumeil PA, Rinke C, Mussig AJ, Hugenholtz P. A complete domain-to-species taxonomy for *Bacteria* and *Archaea*. *Nat Biotechnol*. 2020;38(9):1079–86.
120. Price MN, Dehal PS, Arkin AP. FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS ONE*. 2010;5(3):e9490.
121. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res*. 2021;49(W1):W293–6.
122. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform*. 2010;11:119.
123. Laslett D, Canback B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res*. 2004;32(1):11–6.
124. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014;30(14):2068–9.
125. Rho M, Tang H, Ye Y. FragGeneScan: predicting genes in short and error-prone reads. *Nucleic Acids Res*. 2010;38(20):e191.
126. Menzel P, Ng KL, Krogh A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat Commun*. 2016;7(1):11257.
127. Tamames J, Puente-Sánchez F. SqueezeMeta, a highly portable, fully automatic metagenomic analysis pipeline. *Front Microbiol*. 2018;9:3349.
128. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9(4):357–9.
129. Zheng J, Ge Q, Yan Y, Zhang X, Huang L, Yin Y. dbCAN3: automated carbohydrate-active enzyme and substrate annotation. *Nucleic Acids Res*. 2023;51(W1):W115–21.
130. Deusch S, Seifert J. Catching the tip of the iceberg - evaluation of sample preparation protocols for metaproteomic studies of the rumen microbiota. *Proteomics*. 2015;15(20):3590–5.
131. Schultz D, Zühlke D, Bernhardt J, Francis TB, Albrecht D, Hirschfeld C, et al. An optimized metaproteomics protocol for a holistic taxonomic and functional characterization of microbial communities from marine particles. *Environ Microbiol Rep*. 2020;12(4):367–76.
132. Perez-Riverol Y, Bai J, Bandla C, García-Seisdedos D, Hewapathirana S, Kamatchinathan S, et al. The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res*. 2022;50(D1):D543–52.

Publisher's Note

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