

## DATASET BRIEF OPEN ACCESS

# Metaproteomic Dataset on Semi-Diurnal Variability of the Bacterioplankton Communities During a Spring Phytoplankton Bloom in the North Sea

Vaikhari Kale<sup>1</sup> | Jürgen Bartel<sup>1</sup> | Daniel Bartosik<sup>2,3</sup> | Philip Berhard Lude<sup>1</sup> | Chandni Sidhu<sup>4</sup> | Hanno Teeling<sup>4</sup> | Rudolf Amann<sup>4</sup> | Thomas Schweder<sup>2,3</sup> | Dörte Becher<sup>1,3</sup> | Anke Trautwein-Schult<sup>1</sup> 

<sup>1</sup>Institute of Microbiology, University of Greifswald, Greifswald, Germany | <sup>2</sup>Institute of Pharmacy, University of Greifswald, Greifswald, Germany | <sup>3</sup>Institute of Marine Biotechnology e.V., Greifswald, Germany | <sup>4</sup>Max Planck Institute for Marine Microbiology, Bremen, Germany

**Correspondence:** Anke Trautwein-Schult ([anke.trautwein-schult@uni-greifswald.de](mailto:anke.trautwein-schult@uni-greifswald.de))

**Received:** 16 October 2024 | **Revised:** 4 June 2025 | **Accepted:** 4 June 2025

**Funding:** This study was funded by the Max Planck Society and supported by the German Research Foundation (DFG) in the framework of the research unit FOR 2406 “Proteogenomics of Marine Polysaccharide Utilization (POMPU)” by grants to D.B. (BE 3869/4-3), T.S. (SCHW 595/10-3, SCHW 595/11-3), R.A. (AM 73/9-3), and H.T. (TE 813/2-3). The Helgoland Time Series of the Alfred Wegener Institute is supported by the Helmholtz Association as an LK-II performance category program.

**Keywords:** *Aurantivirga* | day-night variability | metaproteomics | phytoplankton bloom | *Candidatus Prosiliicoccus*

## ABSTRACT

Phytoplankton blooms create a substrate-rich environment that supports the growth of bacterial planktonic heterotrophs. Previously, we studied the dynamics of such bacterioplankton at a long-term ecological research site near the coast of Helgoland Island (North Sea) once a day. Here, we present a novel dataset (available under the PRIDE-ID: PXD055396) indicating significant differences at the protein level in a semi-diurnal analysis. Using metaproteomics, we studied changes in the free-living (0.2–3 µm) bacterial community that occurred between early (7 am) and late (9 pm) sampling over 3 days. The results highlight the sensitivity, robustness, and reproducibility of mass spectrometry-based metaproteomic analyses to assess changes in the activities of the bacterioplankton communities. Taxonomic analyses revealed significant changes in the abundance of 65 bacterial genera. Particularly, proteins from the flavobacterial genera *Candidatus Prosiliicoccus* and *Aurantivirga* were significantly more abundant in the late samples. This comprehensive dataset highlights semi-diurnal changes in bacterial community composition and metabolic activity during a phytoplankton bloom that would have remained undetected with a once-per-day sampling approach.

Just as other meta-omics technologies, microbial metaproteomics expands the application of proteomics from pure cultures and simple model systems to complex biological systems consisting of entire microbial communities and possibly host cells [1]. However, conducting metaproteomic studies is often challenging due to sample diversity and complexity [2–5]. Over the past decade, advancements in the preparation of complex samples and

innovative methods of data acquisition and analysis have allowed for a more widespread and routine use of metaproteomics. In recent years, numerous metaproteome studies have provided unique insights into the activity of microbial communities in various environments, such as the oceans [6], the human gut [7–9], soil [10, 11], wastewater treatment plants [12, 13], and even extreme habitats [14] like high-CO<sub>2</sub> aquifers [15]. Studies focusing

**Abbreviations:** CAZymes, carbohydrate-active enzymes; LTER, long-term ecological research; MAGs, metagenome-assembled genomes; PCoA, principal coordinate analysis; R<sup>2</sup>, Pearson correlation coefficient.

This is an open access article under the terms of the [Creative Commons Attribution](#) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *Proteomics* published by Wiley-VCH GmbH.

on the microbial nutrient transport in the South Atlantic Ocean [16], biogeochemical processes within the oxygen minimum zones at a seasonally stratified fjord [17], or the dynamics of the carbon flux in the marine ecosystems [18, 19] showed the potential and challenges of ocean metaproteome studies [6].

Marine microbiomes have increasingly garnered attention due to their impact on global biogeochemical cycles. Marine phytoplankton blooms can trigger large-scale bacterioplankton blooms, which, though short-lived, play a crucial role in marine carbon cycling [20–22]. The blooming and subsequent decay of algae releases substantial amounts of organic matter, including large amounts of various algal polysaccharides. Distinct clades of heterotrophic planktonic bacteria benefit from these substrates during bloom events. Bacterial polysaccharide utilization specialists degrade algal polysaccharides via specific carbohydrate-active enzymes (CAZymes) [20, 23, 24]. Due to their usually high substrate specificity, a diverse array of such CAZymes is required for the complete breakdown and utilization of each distinct algal polysaccharide [25–28]. In conjunction with CAZymes, specific types of TonB-dependent transporter proteins allow these bacteria to bind and transport longer oligosaccharides into the periplasm [23, 29, 30].

During phytoplankton blooms, dynamic changes in environmental conditions, nutrient availability, and substrate supply, as well as in interactions among organisms (animals, zooplankton, phytoplankton, bacterioplankton, and viruses), result in highly dynamic marine microbiome composition. Additionally, phototrophic organisms exhibit a specific temporal relationship between transcriptome, proteome, and metabolome, due to the presence of light-dark cycles [31, 32], adding another level of complexity in these dynamic systems. For instance, in a controlled light-dark *in vitro* study of the phototrophic marine cyanobacterium *Prochlorococcus* MED4, significant variabilities were detected in the relationship between mRNA and protein abundances, with more pronounced oscillations in transcript abundances and later peaks in protein abundances [31]. Several *in situ* studies on diel oscillation have been conducted in marine ecosystems with surface water temperatures above 20°C, such as the North Pacific Subtropical Gyre, the Daya Bay in the South Chinese Sea, and the North Atlantic Ocean near New Jersey [33–35], but lacking for temperate marine ecosystems. Some studies confirmed diel oscillations of around half of all *Prochlorococcus* transcripts [36], and 70% of the targeted metabolites [37]. These studies highlight that different effects can be observed on the transcript, protein, and metabolite levels.

To monitor changes in plankton composition, a long-term ecological research (LTER) site was established off the coast of Helgoland Island in the rather cold (about 10°C water surface temperature during springtime) southern North Sea (“Kabeltonne”, 54°11.3'N, 7°54.0'E, DEIMS.iD: <https://deims.org/1e96ef9b-0915-4661-849f-b3a72f5aa9b1>). For this LTER, phytoplankton data has been systematically collected since 1962, and zooplankton data since 1975, together with accompanying physicochemical data [38, 39]. Regular sampling of bacterioplankton at the Helgoland Roads LTER site has been conducted since 2009 and analyzed using a combination of methods, including state-of-the-art metaproteomics [20, 22, 40–44]. These long-term sampling campaigns revealed that the complex bacterial communities that

thrive during and after phytoplankton blooms are dominated by annually recurrent clades [40, 45] with substantial changes in the relative abundances, sometimes within a day or two. It was recently demonstrated that, especially during spring blooms with little top-down (predator) control, the composition of this bacterial community is significantly shaped by the availability of dissolved algal glycans, as well as by bacterial glucans that living bacteria recycle from bacterial necromass [43]. In 2020, we sampled at the Helgoland Roads LTER site between 2nd March and 26th May. Previously, we published data on total bacterial cell counts and cell counts for specific clades, chlorophyll *a*, wind direction data, physicochemical and phytoplankton, glucan and saccharide measurements, as well as metagenomes, metatranscriptomes, and additionally for 15 sampling time points (7 am), metaproteomics data [42–44].

So far, diurnal changes have been out of the scope of the long-term studies. Here, we applied a metaproteomics workflow in a temperate marine ecosystem (Table S1) to investigate short-term changes in the free-living (0.2–3 µm) bacterial community. We used samples collected at semi-diurnal intervals at 7 am (“early”) and 9 pm (“late”) for 3 days (5th, 6th, and 7th May 2020) and compared them with samples taken before (29th April, 7 am) and afterwards (11th May, 7 am). The early samples were taken between 67 (29th April, 7 am) to 91 min (11th May, 7 am) after sunrise, and the late samples were taken 10 min (5th May, 9 pm) to 14 min (7th May, 9 pm) before sunset (Table S1). Here, we present brief first insights into short-term changes in the community compositions and activities of the free-living bacterioplankton communities that were sampled with different time intervals of 10 (between 9 pm and 7 am) and 14 h (between 7 am and 9 pm). Therefore, seawater samples were collected at 1 m depth and sequentially filtered through 142 mm polycarbonate membrane filters (Millipore, Schwalbach, Germany) with decreasing pore sizes (10, 3, and 0.2 µm) as previously described [43] and in the Supporting Information. The proteins of the free-living bacterial community were extracted from one-eighth of a 0.2 µm filter by adding lysis buffer, heating, and sonication treatment in triplicate. After centrifugation (1 min, 4°C, 5000 × g), the proteins in the supernatant were precipitated by adding pre-cooled trichloroacetic acid (20% [v/v]), pelleted via centrifugation (1 h, 4°C, 12,000 × g), washed three times with pre-cooled acetone and dried. The remaining protein pellet was resuspended in 2× Laemmli SDS sample loading buffer (4% SDS, 20% glycerol, 10% 2-mercaptoethanol, 0.002% bromophenol blue in 0.125 M Tris-HCl [pH 6.8]) and incubated for 5 min at 95°C before separation via SDS-PAGE (Criterion TG 4%–20% Precast Midi Gel, BIO-RAD Laboratories, Inc., Hercules, CA, USA). Afterwards, the gel was fixated (40% ethanol [v/v], 10% acetic acid [v/v]), and stained with Coomassie [46], before each gel line was cut into 20 equally sized pieces. The gel pieces were cut into smaller pieces, destained three times with washing buffer (200 mM ammonium bicarbonate in 30% ACN [v/v]), and dehydrated as described by Bonn et al. [47] before the proteins were in-gel reduced and alkylated. After washing, dehydrating, and drying of the gel pieces, the proteins were tryptically digested (2 µg/mL trypsin, Promega, Fitchburg, WI, USA), incubated for 15 h at 37°C and eluted first with 120 µL solvent A (0.1% acetic acid [v/v]) and second with 100 µL 30% ACN (v/v) for 15 min in an ultrasonication bath (SONOREX SUPER RK 102 H, Bandelin, Berlin, Germany). The pooled eluates were desalting via

ZipTip C18 (Merck Millipore, P10 tip size, Burlington, MA, USA) according to the manufacturer's instructions and loaded onto in-house packed capillary columns (20 cm length, 75  $\mu$ m inner diameter, Dr. Maisch GmbH, Ammerbruch, Germany, RepsoSil pur C18 material with pore size 120  $\text{\AA}$  and particle size 1.9  $\mu$ m) via an Easy nLC1000 LC (Thermo Fisher Scientific, Waltham, MA, USA) coupled to a Q Exactive mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA) in data-dependent acquisition mode and separated using a non-linear binary gradient (131 min) from 1% to 99% solvent B (99.9% ACN [v/v], 0.1% acetic acid [v/v]) in solvent A at a constant flow rate of 300 nL/min. The MS1 scan was recorded in the orbitrap with a resolution of 140,000 at 200 m/z and a mass window of 300 to 1650 m/z. The 15 most abundant precursor ions were selected for higher-energy C-Trap dissociation fragmentation with enabled dynamic exclusion.

MS/MS spectra were searched using a two-step searching strategy [48]. First, all MS/MS spectra were searched against a sample-specific, non-redundant metagenome-derived database as described in the Supporting Information using Mascot [49] (Table S2) with the following parameters: fragment ion mass tolerance and parent ion tolerance of 10 ppm, no missed cleavages, variable modification on methionine (oxidation), and fixed modification on cysteine (carbamidomethylation). Search outputs from replicates were combined in Scaffold [50] (using X!Tandem [51] for validation with default settings). Peptide and protein identifications were accepted with a probability higher than 95% and 99%, respectively. The eight subset databases for each time point (including only identified proteins) were combined, and after filtering for redundancy using CD-HIT [52] (clustering threshold of 97% identity), decoy entries were added, and all MS/MS spectra were searched again with the constructed non-redundant subset database (88,300 entries including reverse entries). The Mascot [49] outputs were analyzed in Scaffold [50] as described above. Peptide and protein identifications were accepted if they contained at least two unique peptides, whereas this criterion was not applied to the first searching step. Peptide probabilities from Mascot and X!Tandem were assigned by the Scaffold Local FDR algorithm or Peptide Prophet algorithm [53] with delta-mass correction, respectively. Protein identifications were accepted if they could be established at greater than 99% probability by the Protein Prophet algorithm [54] and contained at least two unique peptides. Proteins that contained similar peptides and could not be differentiated based on the MS/MS analysis alone were grouped to satisfy the principles of parsimony. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (<https://proteomecentral.proteomexchange.org>) via the PRIDE partner repository [55] with the dataset identifier PXD055396. Functional and taxonomic annotation was performed by an in-house pipeline. For this aim, protein sequences were mapped onto 2020 high-quality MAGs [42] using Diamond BLASTP (v2.1.1.155, flags: -evaluate 1E-4 -id 95 -query-cover 70 -subject-cover 70. Sequences that remained unmapped were classified as either "prokaryotic" or "non-prokaryotic" using BASTA (v1.4.1, with the options: -m 1 -i 0 -l 0 -e 0.001 -p 90), along with Diamond BLASTP (options: -evaluate 0.001 -k 100) against the NCBI non-redundant protein database ("NCBI\_nr"; as of 22nd February 2023). For "non-prokaryotic" sequences, the NCBI\_nr results were used for Last Common Ancestor predictions via BASTA, while "prokaryotic" sequences were classified based on Diamond BLASTP results

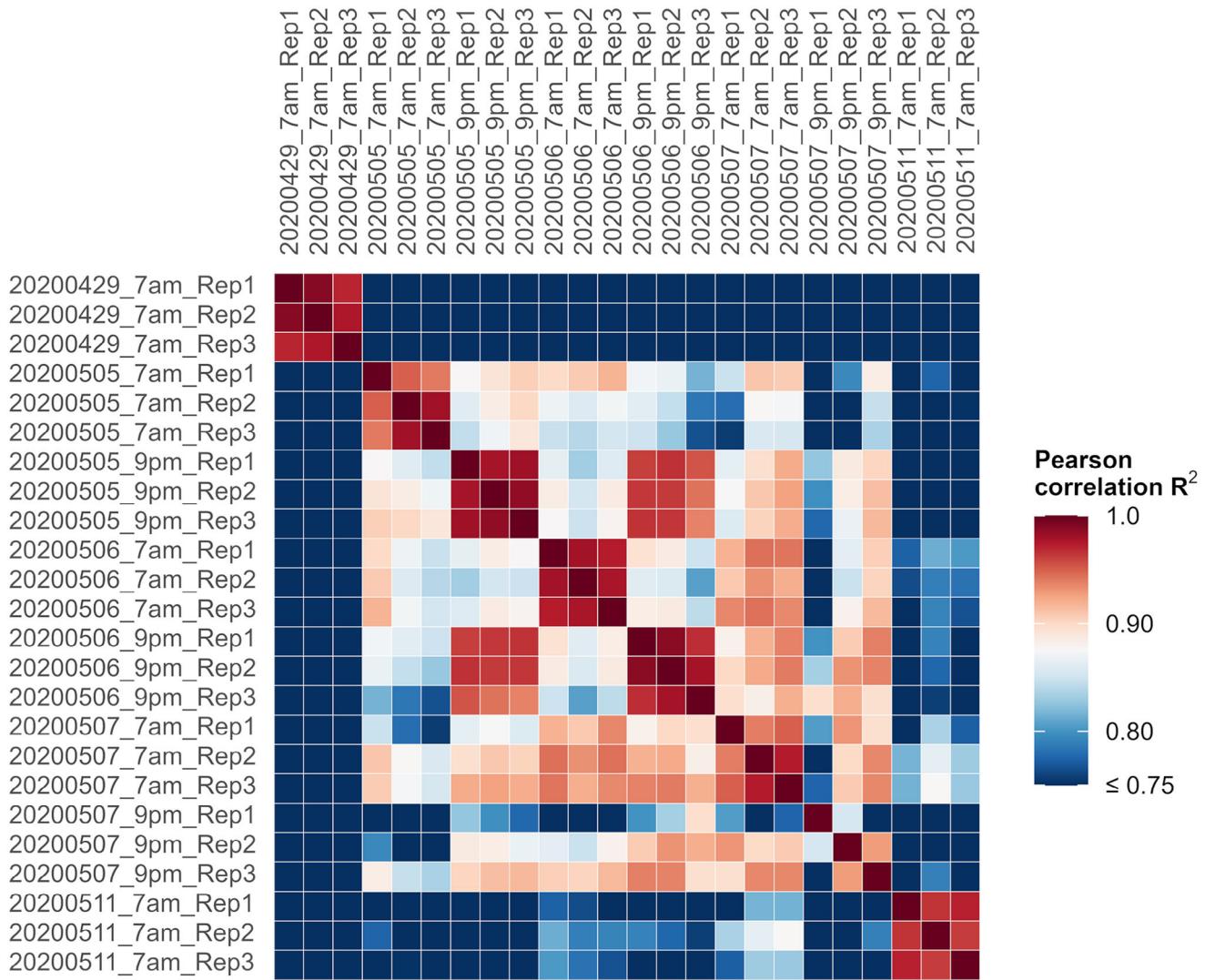
against the Genome Taxonomy Database ("GTDB"; r214.1) [56] with identical thresholds.

All LC-MS/MS measurements generated a total of 20.5 million recorded spectra, of which 5.9 million could be confidently identified. On average, 858,049 ( $\pm$  132,309) spectra were generated, of which 245,663 ( $\pm$  45,788) were identified per sample, resulting in an average identification rate of 28.8% ( $\pm$  5.0%). In classical metaproteomic studies, spectral identification rates typically range from 5% to 30% [5]. A cross-laboratory (Critical Assessment of MetaProteome Investigation—CAMPI) study on stool samples reported identification rates ranging from 15.7% to 41.9% (average 23.2%) [4]. This variation was attributed to differences in protein extraction, sample preparation, and mass detection, followed by the same search strategy between the different labs using the X!Tandem search engine and a publicly available database.

To check the quality and robustness of the data, we calculated the squared Pearson correlation coefficient ( $R^2$ ) of the weighted spectral counts for protein groups as a measure of the linear correlation between two individual technical replicates (Figure 1). For each pairwise combination of the three replicates of each sampling time point, the  $R^2$  values indicated a high reproducibility. All  $R^2$  values between replicates exceeded 0.94. These  $R^2$  values are comparable to stool samples that were analyzed directly after collecting fecal specimens compared to the same sample after being frozen for 2 months [57]. Only samples from the 7th May at 9 pm were an exception with a lower correlation value of 0.73 between replicates 1 and 3.

In total, we identified 26,263 protein groups in the complete dataset (excluding contaminations and decoy hits) ranging from 6961 (7th May 2020, 9 pm, replicate 2) to 13,066 protein groups (6th May 2020, 7 am, replicate 1) per sample (with an average of  $10,759 \pm 1536$ ) (Table S3). More spectra were identified per protein group in the 9 pm samples (considering weighted spectra, which represent the sum of unique peptide spectra per protein group and a fractional contribution of spectra from shared peptides, on average 24.8 identified spectra per protein group, with a 5% to 95% percentile of 1.86 to 88.5) compared to the 7 am samples (on average 21.0 identified spectra per protein group, with a 5% to 95% percentile of 1.81 to 76.7). After filtering for protein groups present in at least two of three replicates in at least one sampling time point, we quantified 19,933 protein groups, with numbers ranging from 7684 (7th May 2020, 9 pm) to 12,248 (6th May 2020, 7 am) protein groups per sampling time point. On average, significantly more protein groups were identified and quantified in the 7 am samples ( $11,426 \pm 1250$  identified and  $11,110 \pm 1032$  quantified) compared to the 9 pm samples ( $9561 \pm 1226$  identified and  $9273 \pm 1145$  quantified, paired Student's *t*-test, *p* value < 0.0002).

At the end of April 2020, the first phase of the spring phytoplankton bloom that was dominated by *Ditylum brightwellii* diatoms ended, and a second bloom was dominated by *Cerataulina pelagica* and to a lesser extent *Chaetoceros* sp. diatoms [42]. At the same time, total bacterial cell counts increased again after a decline of *D. brightwellii* numbers towards the end of the first bloom. The frequency of metagenomic 16S rRNA gene sequences indicated a shift from *Alphaproteobacteria* dominating during the first blooming phase towards *Bacteroidota* taking over at



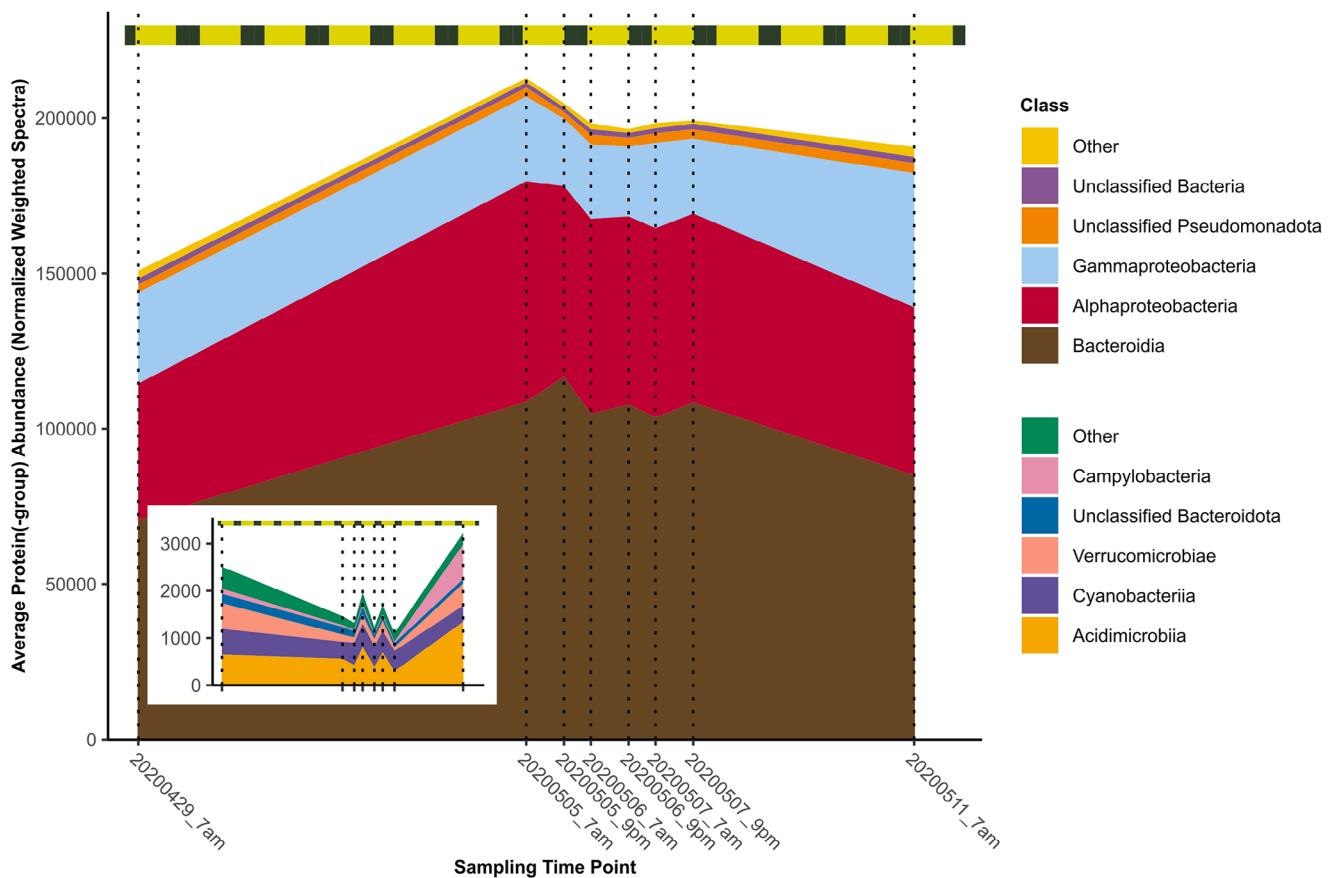
**FIGURE 1** | Heatmap representing the squared Pearson correlation coefficients ( $R^2$ ) of the raw quantitative values (weighted spectra) from different samples taken between the 29th April and the 11th May 2020. The color gradient represents the correlation, with red indicating higher correlation values and blue representing lower correlation values. The x- and y-axes correspond to the different samples.

the beginning of the second phytoplankton bloom [42]. Consequently, a noticeable shift in the metaproteome composition was detected over the full 13-day period between 29th April and 11th May 2020. In total, 4379 protein groups in our dataset were quantified (in at least two of the three replicates) in all samples, representing the largest overlap (22.0%) compared to all possible intersections between the eight sampling time points (Figure S1). The first (29th April) and last (11th May) sampling time points differed from the others, with 2531 (12.7%) and 1426 (7.2%) protein groups being detected exclusively in these two sampling time points, respectively. Furthermore, 1048 (5.3%) protein groups were detected in all samples except on 29th April, confirming the overall shift in bacterioplankton composition in the second bloom phase. The semi-diurnal shifts measured from 5th to 7th May were smaller than the shifts within 4 to 6 days.

To test for short-term dynamics on the level of summed protein group abundances between the different taxonomic groups in our dataset, we initially compared alpha-diversity based on the Shannon index calculated for the genus level between early and

late samples from 5th to 7th May (Figure S2A). This metric describes the taxonomic richness and evenness of the abundance distribution of a microbial community within a sample with greater values indicating a more diverse community. A significantly higher diversity was detected in 7 am samples compared to 9 pm samples (paired Student's *t*-test, *p* value < 0.0002).

In the spring bloom study of 2020, we investigated bacterioplankton metagenomes from 30 time points (including the eight time points we analyzed here) from the same bloom, which allowed us to reconstruct a total of 251 non-redundant metagenome-assembled genomes (MAGs) of free-living bacteria [42]. We mapped the 19,933 protein groups quantified in our dataset back to these MAGs resulting in 11,622 (59%) protein groups mapping to 198 MAGs (79% of all MAGs detected). This indicates that the constructed MAGs represented the most abundant and active taxa and underscores the high quality of the metaproteome dataset. A phylogenetic tree was generated based on 16S rRNA sequences for individual MAGs, previously [42]. Using summed protein group abundances from those mapped protein groups, pair-wise



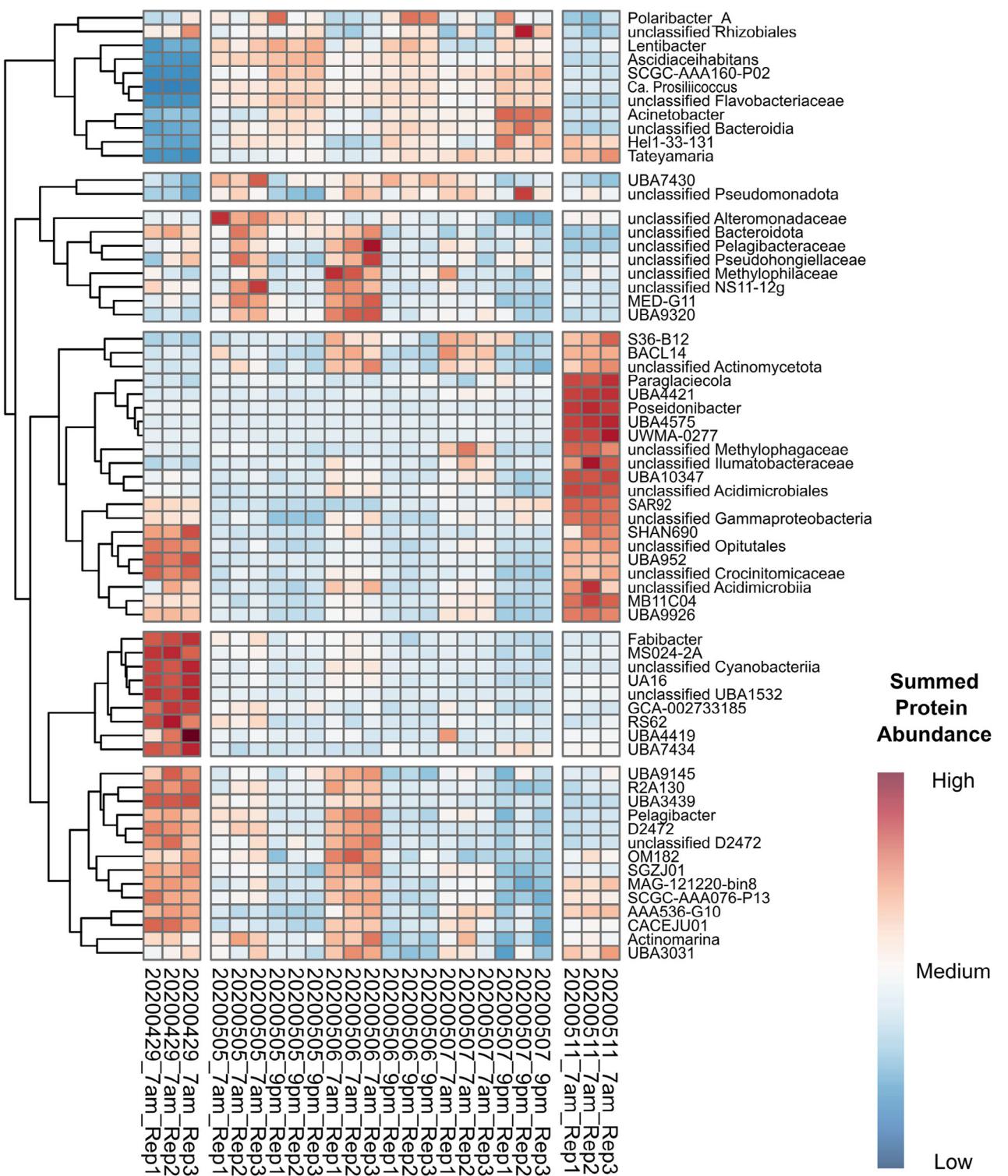
**FIGURE 2** | The composition of marine microbiota between the 29th April and 11th May 2020 at the Helgoland LTER station “Kabeltonne” in the North Sea based on metaproteomics data. The quantified protein groups assigned to the class level were summed. Low-abundant members of taxonomical groups were merged to “Others”. The smaller inset shows additional classes that are included as “Others” in the main plot. In the absence of an assignment to any taxonomical group, respective proteins were assigned as unclassified to the next higher taxonomic level. The vertical dotted lines indicate the different sampling time points. The time between sunrise and sunset is indicated in yellow and the time between sunset and sunrise in black in the horizontal bar at the top of the plot (weather data, Table S1).

weighted Unifrac distances were calculated as a measure of beta diversity. These distances were then used as the input for a principal coordinate analysis (PCoA, Figure S2B). All sample points were separated from each other with days being sorted along the x-axis and sampling time along the y-axis.

After these general trends, we analyzed short- and long-term trends across various taxonomic levels based on summed protein group abundances. The increase in protein group abundances of the bacterial community (Figure S3A) at the beginning of the *C. pelagica*-dominated phytoplankton bloom phase could be mainly attributed to changes in the dominating phyla *Bacteroidota* and *Pseudomonadota* (formerly *Proteobacteria*) (Figure S3B). However, less abundant phyla, such as *Campylobacterota*, also exhibited long-term changes in protein group abundances, whereas the phylum *Actinobacteriota* exhibited short-term fluctuations between the 7 am and 9 pm samples. The dominant classes *Alphaproteobacteria*, *Gammaproteobacteria*, and *Bacteroidia* were subject to strong fluctuations (Figure 2), which mainly involved the orders *Rhodobacterales*, *Pelagibacterales*, *Pseudomonadales*, *Enterobacterales*, and *Flavobacterales* (Figure S3D). *Rhodobacteraceae*, *Porticoccaceae*, *Pelagibacteraceae*, and *Alteromonadaceae* (Figure S3E) were dominating proteobacterial families including the abundant genera *Amylibacter*, *Plank-*

*tomarina* and members of the gammaproteobacterial SAR92 (HTCC2207) clade (Figure S3F). Especially the dominant genera *Cd. Prosiliicoccus* (previously *Ulvibacter*) and *Aurantivirga* (SCGC-AAA160-P02) (Figure S3F) of the family *Flavobacteriaceae* showed higher abundance at the later sampling time points (Figure S3E). A rapid increase of *Bacteroidota* abundance after 29th April has already been shown for corresponding metagenomic 16S rRNA gene frequencies [42]. The indicated daytime-dependent changes in abundances in the metagenomic gene frequencies were confirmed with our metaproteomic data, for the abundant genera *Amylibacter*, *Polaribacter*, and *Aurantivirga*.

We identified significant changes in 65 relevant bacterial genera based on summed protein group abundances between early and late samples taken between 5th May and 7th May. As depicted in Figure 3, the different protein abundances between early and late samples of these genera were clustered depending on their abundance in all eight sampling time points (including 29th April and 11th May). The most interesting cluster consisted of genera with low overall protein group abundances at the beginning of May and higher protein group abundances in the 9 pm samples than the 7 am samples. This group consisted of the previously mentioned *Cd. Prosiliicoccus* and *Aurantivirga*, which were also more abundant at the beginning of the second bloom phase (peak



**FIGURE 3** | Heatmap and hierarchical clustering of significantly changed genera. The metaproteomic dataset from 5th May (7 am) until 7th May (9 pm) was analyzed using the R package ANCOMBC2 using summed protein abundances on the genus level. Significantly changed genera were hierarchically clustered and protein abundances of 29th April and 11th May were included.

on 6th May for *Cd. Prosiliicoccus* and 8th May for *Aurantivirga*) as indicated by the abundances inferred from metagenomic data or microscopic cell counting via catalyzed reporter deposition-fluorescence in situ hybridization [42]. For most other clusters, protein groups were more abundant in the 7 am samples than the 9 pm samples, although these clusters showed different protein

group abundances on the 29th April and 11th May. Members of the *Alteromonadaceae*, *Pelagibacteraceae*, and *Methylophilaceae* families showed the highest protein group abundance for the 5th and 6th May, whereas the cluster containing unclassified *Actinomycetota*, unclassified *Gammaproteobacteria* and *Poseidonibacter* showed highest abundance on 11th May and the cluster

containing *Fabibacter* peaked at the end of the first phytoplankton bloom phase (29th April).

Finally, we were interested in the function of the proteins with significant short-term abundance changes. For this purpose, protein groups with significantly different abundances were identified using the R package DEqMS [58] (Table S3, labeled with “both” in the column DEqMS result). The analysis identified 422 such protein groups, 369 of which were more abundant in the morning samples. Functional annotation was performed using MAG annotations. For sequences that could not be assigned to MAGs, BLAST alignments of the protein sequences were conducted against the Genome Taxonomy Database (GTDB) [56] to find the most similar and related sequence. While 72 of the different abundant proteins were annotated as hypothetical proteins, the second highest number of differentially abundant proteins (47 significantly changed protein groups) were annotated as TonB-dependent receptors, which mediate substrate-specific transport across the outer membrane [59].

Previous studies have shown that bacterioplankton composition is highly dynamic during spring in the North Sea [20, 22, 43, 44, 60–64]. Changes over several days are more pronounced compared to the changes that occur within a day, as is shown in both our analysis and metagenome gene frequency analyses [42]. Nevertheless, our comprehensive metaproteome dataset suggests that significant changes occurred in the abundances of distinct bacterial protein groups at both, the taxonomic and functional levels, between the early and late samples. An even higher temporal resolution would be necessary for a more detailed analysis of diurnal adaptive changes of the bacterial physiology in the complex microbiomes during phytoplankton blooms. We anticipate that this study will pave the way towards such studies, and hope that the presented dataset will be a useful resource for researchers during the development of data evaluation and modeling approaches. Additionally, our preliminary study reveals that the timing of sampling is crucial in long-term studies. In the case of irregular sampling, detected changes could be caused by diurnal changes and the associated changes in environmental conditions like the solar irradiation intensity. Changes at the transcriptome level are more dynamic as shown for dial oscillation in phototrophic bacteria [31]. Yet, as we show here, changes do also occur at the protein level. Owing to contemporary refined metaproteomics, such changes can nowadays be detected even in highly complex, diverse, and dynamic systems *in situ*.

## Acknowledgments

We acknowledge Lilly Franzmeyer for technical assistance in the preparation of the samples.

Open access funding enabled and organized by Projekt DEAL.

## Conflicts of Interest

The authors declare no conflicts of interest.

## Data Availability Statement

The Mass Spectrometry Proteomics Data Have Been Deposited to the ProteomeXchange Consortium (<https://proteomecentral.proteomexchange.org>) via the PRIDE partner repository with the dataset identifier PXD055396.

## References

1. P. Wilmes and P. L. Bond, “Metaproteomics: Studying Functional Gene Expression in Microbial Ecosystems”, *Trends in Microbiology* 14, no. 2 (2006): 92–97, <https://doi.org/10.1016/j.tim.2005.12.006>.
2. J. Armengaud, “Metaproteomics to Understand How Microbiota Function: The Crystal Ball Predicts a Promising Future”, *Environmental Microbiology* 25, no. 1 (2022): 115–125, <https://doi.org/10.1111/1462-2920.16238>.
3. D. J. Nebauer, L. A. Pearson, and B. A. Neilan, “Critical Steps in an Environmental Metaproteomics Workflow”, *Environmental Microbiology* 26, no. 5 (2024): 16637, <https://doi.org/10.1111/1462-2920.16637>.
4. T. Van Den Bossche, B. J. Kunath, K. Schallert, et al., “Critical Assessment of MetaProteome Investigation (CAMPI): A Multi-Laboratory Comparison of Established Workflows”, *Nature Communications* 12, no. 1 (2021): 7305, <https://doi.org/10.1038/s41467-021-27542-8>.
5. R. Heyer, K. Schallert, R. Zoun, B. Becher, G. Saake, and D. Benndorf, “Challenges and Perspectives of Metaproteomic Data Analysis”, *Journal of Biotechnology* 261 (2017): 24–36, <https://doi.org/10.1016/j.jbiotec.2017.06.1201>.
6. M. A. Saito, E. M. Bertrand, M. E. Duffy, et al., “Progress and Challenges in Ocean Metaproteomics and Proposed Best Practices for Data Sharing”, *Journal of Proteome Research* 18, no. 4 (2019): 1461–1476, <https://doi.org/10.1021/acs.jproteome.8b00761>.
7. A. Tanca, A. Palomba, G. Fiorito, M. Abbondio, D. Pagnozzi, and S. Uzzau, “Metaproteomic Portrait of the Healthy human Gut Microbiota”, *NPJ Biofilms and Microbiomes* 10, no. 1 (2024): 54, <https://doi.org/10.1038/s41522-024-00526-4>.
8. Z. Sun, Z. Ning, and D. Figeys, “The Landscape and Perspectives of the Human Gut Metaproteomics”, *Molecular & Cellular Proteomics* 23, no. 5 (2024): 100763, <https://doi.org/10.1016/j.mcpro.2024.100763>.
9. M. Stamboulian, J. Canderan, and Y. Ye, “Metaproteomics as a Tool for Studying the Protein Landscape of Human-Gut Bacterial Species”, *PLoS Computational Biology* 18, no. 3 (2022): 1009397, <https://doi.org/10.1371/journal.pcbi.1009397>.
10. J.-Y. Lee, H. D. Mitchell, M. C. Burnet, et al., “Uncovering Hidden Members and Functions of the Soil Microbiome Using *De Novo* Metaproteomics”, *Journal of Proteome Research* 21, no. 8 (2022): 2023–2035, <https://doi.org/10.1021/acs.jproteome.2c00334>.
11. H. Pan, R. Wattiez, and D. Gillan, “Soil Metaproteomics for Microbial Community Profiling: Methodologies and Challenges”, *Current Microbiology* 81, no. 8 (2024): 257, <https://doi.org/10.1007/s00284-024-03781-y>.
12. H. B. C. Kleikamp, D. Grouzdev, P. Schaasberg, et al., “Metaproteomics, Metagenomics and 16S rRNA Sequencing Provide Different Perspectives on the Aerobic Granular Sludge Microbiome”, *Water Research* 246 (2023): 120700, <https://doi.org/10.1016/j.watres.2023.120700>.
13. A. Elahinik, M. Haarsma, B. Abbas, et al., “Glycerol Conversion by Aerobic Granular Sludge”, *Water Research* 227 (2022): 119340, <https://doi.org/10.1016/j.watres.2022.119340>.
14. K. Maseh, N. Ehsan, S. Mukhtar, S. Mehnaz, and K. A. Malik, “Metaproteomics: An Emerging Tool for the Identification of Proteins from Extreme Environments”, *Environmental Sustainability* 4, no. 1 (2021): 39–50, <https://doi.org/10.1007/s42398-020-00158-2>.
15. C. J. Moore, T. L. V. Bornemann, P. A. Figueroa-Gonzalez, et al., “Time-Series Metaproteogenomics of a High-CO<sub>2</sub> Aquifer Reveals Active Viruses with Fluctuating Abundances and Broad Host Ranges”, *Microlife* 5 (2024): uqae011, <https://doi.org/10.1093/femsml/uqae011>.
16. R. M. Morris, B. L. Nunn, C. Frazer, D. R. Goodlett, Y. S. Ting, and G. Rocap, “Comparative Metaproteomics Reveals Ocean-Scale Shifts in

Microbial Nutrient Utilization and Energy Transduction", *ISME Journal* 4, no. 5 (2010): 673–685, <https://doi.org/10.1038/ismej.2010.4>.

17. A. K. Hawley, H. M. Brewer, A. D. Norbeck, L. Pasa-Tolic, and S. J. Hallam, "Metaproteomics Reveals Differential Modes of Metabolic Coupling Among Ubiquitous Oxygen Minimum Zone Microbes", *Proceedings of the National Academy of Sciences of the United States of America* 111, no. 31 (2014): 11395–11400, <https://doi.org/10.1073/pnas.1322132111>.

18. E. K. Moore, B. L. Nunn, D. R. Goodlett, and H. R. Harvey, "Identifying and Tracking Proteins Through the Marine Water Column: Insights into the Inputs and Preservation Mechanisms of Protein in Sediments", *Geochimica Et Cosmochimica Acta* 83 (2012): 324–359, <https://doi.org/10.1016/j.gca.2012.01.002>.

19. M. C. Bridoux, J. Neibauer, A. E. Ingalls, B. L. Nunn, and R. G. Keil, "Suspended Marine Particulate Proteins in Coastal and Oligotrophic Waters", *Journal of Marine Systems* 143 (2015): 39–48, <https://doi.org/10.1016/j.jmarsys.2014.10.014>.

20. H. Teeling, B. M. Fuchs, D. Becher, et al., "Substrate-Controlled Succession of Marine Bacterioplankton Populations Induced by a Phytoplankton Bloom", *Science* 336, no. 6081 (2012): 608–611, <https://doi.org/10.1126/science.1218344>.

21. P. G. Falkowski, "The Role of Phytoplankton Photosynthesis in Global Biogeochemical Cycles", *Photosynthesis Research* 39, no. 3 (1994): 235–258, <https://doi.org/10.1007/BF00014586>.

22. T. B. Francis, D. Bartosik, T. Sura, et al., "Changing Expression Patterns of TonB-Dependent Transporters Suggest Shifts in Polysaccharide Consumption over the Course of a Spring Phytoplankton Bloom", *ISME Journal* 15 (2021): 2336–2350, <https://doi.org/10.1038/s41396-021-00928-8>.

23. M. Bäumgen, T. Dutschke, and U. T. Bornscheuer, "Marine Polysaccharides: Occurrence, Enzymatic Degradation and Utilization", *Chem-BioChem* 22, no. 13 (2021): 2247–2256, <https://doi.org/10.1002/cbic.202100078>.

24. E. Drula, M.-L. Garron, S. Dogan, V. Lombard, B. Henrissat, and N. Terrapon, "The Carbohydrate-Active Enzyme Database: Functions and Literature", *Nucleic Acids Research* 50, no. D1 (2022): D571–D577, <https://doi.org/10.1093/nar/gkab1045>.

25. J. M. Grondin, K. Tamura, G. Déjean, D. W. Abbott, and H. Brumer, "Polysaccharide Utilization Loci: Fueling Microbial Communities", *Journal of Bacteriology* 199, no. 15 (2017): e00860–00816, <https://doi.org/10.1128/JB.00860-16>.

26. B. Yu, Z. Lu, S. Zhong, and K.-L. Cheong, "Exploring Potential Polysaccharide Utilization Loci Involved in the Degradation of Typical Marine Seaweed Polysaccharides by *Bacteroides thetaiotaomicron*", *Frontiers in Microbiology* 15 (2024): 1332105, <https://doi.org/10.3389/fmicb.2024.1332105>.

27. L. Reisky, A. Préchoux, M.-K. Zühlke, et al., "A Marine Bacterial Enzymatic Cascade Degrades the Algal Polysaccharide Ulvan", *Nature Chemical Biology* 15, no. 8 (2019): 803–812, <https://doi.org/10.1038/s41589-019-0311-9>.

28. T. Dutschke, I. Beidler, D. Bartosik, et al., "Marine Bacteroidetes Enzymatically Digest Xylans from Terrestrial Plants", *Environmental Microbiology* 25, no. 9 (2023): 1713–1727, <https://doi.org/10.1111/1462-2920.16390>.

29. A. Trincone, "Update on Marine Carbohydrate Hydrolyzing Enzymes: Biotechnological Applications", *Molecules* 23, no. 4 (2018): 901, <https://doi.org/10.3390/molecules23040901>.

30. K. Schauer, D. A. Rodionov, and H. De Reuse, "New Substrates for TonB-Dependent Transport: Do We Only See the 'Tip of the Iceberg'?", *Trends in Biochemical Sciences* 33, no. 7 (2008): 330–338, <https://doi.org/10.1016/j.tibs.2008.04.012>.

31. J. R. Waldbauer, S. Rodrigue, M. L. Coleman, and S. W. Chisholm, "Transcriptome and Proteome Dynamics of a Light-Dark Synchronized Bacterial Cell Cycle", *PLoS ONE* 7, no. 8 (2012): 43432, <https://doi.org/10.1371/journal.pone.0043432>.

32. A. Werner, C. D. Broeckling, A. Prasad, and C. A. M. Peebles, "A Comprehensive Time-Course Metabolite Profiling of the Model *Cyanobacterium synechocystis* sp. PCC 6803 under Diurnal Light:Dark Cycles", *Plant Journal* 99, no. 2 (2019): 379–388, <https://doi.org/10.1111/tpj.14320>.

33. F. O. Aylward, J. M. Eppley, J. M. Smith, F. P. Chavez, C. A. Scholin, and E. F. Delong, "Microbial Community Transcriptional Networks Are Conserved in Three Domains at Ocean Basin Scales", *Proceedings of the National Academy of Sciences of the United States of America* 112, no. 17 (2015): 5443–5448, <https://doi.org/10.1073/pnas.1502883112>.

34. S. Chen, M. Z. U. Arifeen, M. Li, et al., "Diel Patterns in the Composition and Activity of Planktonic Microbes in a Subtropical Bay", *Ocean-Land-Atmosphere Research* 3 (2024): Article 0044, <https://doi.org/10.34133/olar.0044>.

35. J. R. Zehnpfennig, C. M. Hansel, S. D. Wankel, et al., "Diel Patterns in Marine Microbial Metatranscriptomes Reflect Differences in Community Metabolic Activity Over Depth on the Continental Shelf of the North Atlantic", *Frontiers in Marine Science* 9, (2022): Article 798877, <https://doi.org/10.3389/fmars.2022.798877>.

36. E. A. Ottesen, C. R. Young, S. M. Gifford, et al., "Multispecies Diel Transcriptional Oscillations in Open Ocean Heterotrophic Bacterial Assemblages", *Science* 345, no. 6193 (2014): 207–212, <https://doi.org/10.1126/science.1252476>.

37. A. K. Boysen, L. T. Carlson, B. P. Durham, et al., "Particulate Metabolites and Transcripts Reflect Diel Oscillations of Microbial Activity in the Surface Ocean", *mSystems* 6, no. 3 (2021): e00896–20, <https://doi.org/10.1128/mSystems.00896-20>.

38. K. H. Wiltshire, A. Kraberg, I. Bartsch, et al., "Helgoland Roads, North Sea: 45 Years of Change", *Estuaries and Coasts* 33, no. 2 (2009): 295–310, <https://doi.org/10.1007/s12237-009-9228-y>.

39. A. Kraberg, U. Kieb, S. Peters, and K. H. Wiltshire, "An Updated Phytoplankton Check-List for the Helgoland Roads Time Series Station with Eleven New Records of Diatoms and Dinoflagellates", *Helgoland Marine Research* 73, no. 1 (2019), <https://doi.org/10.1186/s10152-019-0528-8>.

40. H. Teeling, B. M. Fuchs, C. M. Bennke, et al., "Recurring Patterns in Bacterioplankton Dynamics during Coastal Spring Algae Blooms", *eLife* 5 (2016): 11888, <https://doi.org/10.7554/eLife.11888>.

41. B. Francis, T. Urich, A. Mikolasch, H. Teeling, and R. Amann, "North Sea Spring Bloom-Associated Gammaproteobacteria Fill Diverse Heterotrophic Niches", *Environmental Microbiome* 16, no. 1 (2021): 15, <https://doi.org/10.1186/s40793-021-00385-y>.

42. C. Sidhu, I. V. Kirstein, C. L. Meunier, et al., "Dissolved Storage Glycans Shaped the Community Composition of Abundant Bacterioplankton Clades During a North Sea Spring Phytoplankton Bloom", *Microbiome* 11, no. 1 (2023): 77, <https://doi.org/10.1186/s40168-023-01517-x>.

43. I. Beidler, N. Steinke, T. Schulze, et al., "Alpha-Glucans from Bacterial Necromass Indicate an Intra-Population Loop Within the Marine Carbon Cycle", *Nature Communications* 15, no. 1 (2024): 4048, <https://doi.org/10.1038/s41467-024-48301-5>.

44. M.-K. Zühlke, E. Ficko-Blean, D. Bartosik, et al., "Unveiling the Role of Novel Carbohydrate-Binding Modules in Laminarin Interaction of Multimodular Proteins from Marine Bacteroidota during Phytoplankton Blooms", *Environmental Microbiology* 26, no. 5 (2024): 16624, <https://doi.org/10.1111/1462-2920.16624>.

45. M. Chafee, A. Fernández-Guerra, P. L. Buttigieg, et al., "Recurrent Patterns of Microdiversity in a Temperate Coastal Marine Environment", *ISME Journal* 12, no. 1 (2018): 237–252, <https://doi.org/10.1038/ismej.2017.165>.

46. V. Neuhoff, N. Arold, D. Taube, and W. Ehrhardt, "Improved Staining of Proteins in Polyacrylamide Gels Including Isoelectric Focusing Gels with Clear Background at Nanogram Sensitivity Using Coomassie Brilliant Blue G-250 and R-250", *Electrophoresis* 9, no. 6 (1988): 255–262, <https://doi.org/10.1002/elps.1150090603>.

47. F. Bonn, J. Bartel, K. Büttner, M. Hecker, A. Otto, and D. Becher, “Picking Vanished Proteins from the Void: How to Collect and Ship/Share Extremely Dilute Proteins in a Reproducible and Highly Efficient Manner”, *Analytical Chemistry* 86, no. 15 (2014): 7421–7427, <https://doi.org/10.1021/ac501189j>.

48. P. Jagtap, J. Goslinga, J. A. Kooren, et al., “A Two-Step Database Search Method Improves Sensitivity in Peptide Sequence Matches for Metaproteomics and Proteogenomics Studies”, *Proteomics* 13, no. 8 (2013): 1352–1357, <https://doi.org/10.1002/pmic.201200352>.

49. D. N. Perkins, D. J. C. Pappin, D. M. Creasy, and J. S. Cottrell, “Probability-Based Protein Identification by Searching Sequence Databases Using Mass Spectrometry Data”, *Electrophoresis* 20, no. 18 (1999): 3551–3567, [https://doi.org/10.1002/\(sici\)1522-2683\(19991201\)20:18<3551::aid-elps3551>3.0.co;2-2](https://doi.org/10.1002/(sici)1522-2683(19991201)20:18<3551::aid-elps3551>3.0.co;2-2).

50. B. C. Searle, “Scaffold: A Bioinformatic Tool for Validating MS/MS-Based Proteomic Studies”, *Proteomics* 10, no. 6 (2010): 1265–1269, <https://doi.org/10.1002/pmic.200900437>.

51. R. Craig and R. C. Beavis, “TANDEM: Matching Proteins with Tandem Mass Spectra”, *Bioinformatics* 20, no. 9 (2004): 1466–1467, <https://doi.org/10.1093/bioinformatics/bth092>.

52. W. Li and A. Godzik, “Cd-hit: A Fast Program for Clustering and Comparing Large Sets of Protein or Nucleotide Sequences”, *Bioinformatics* 22, no. 13 (2006): 1658–1659, <https://doi.org/10.1093/bioinformatics/btl158>.

53. A. Keller, A. I. Nesvizhskii, E. Kolker, and R. Aebersold, “Empirical Statistical Model to Estimate the Accuracy of Peptide Identifications Made by MS/MS and Database Search”, *Analytical Chemistry* 74 (2002): 5383–5392, <https://doi.org/10.1021/ac025747h>.

54. A. I. Nesvizhskii, A. Keller, E. Kolker, and R. Aebersold, “A Statistical Model for Identifying Proteins by Tandem Mass Spectrometry”, *Analytical Chemistry* 75 (2003): 4646–4658, <https://doi.org/10.1021/ac0341261>.

55. Y. Perez-Riverol, J. Bai, C. Bandla, et al., “The PRIDE Database Resources in 2022: A Hub for Mass Spectrometry-Based Proteomics Evidences”, *Nucleic Acids Research* 50, no. D1 (2022): D543–D552, <https://doi.org/10.1093/nar/gkab1038>.

56. D. Parks and P. Hugenholtz (2024). Genome Taxonomy Database r2141 Version 192.

57. C. Henry, A. Bassignani, M. Berland, O. Langella, H. Sokol, and C. Juste, “Modern Metaproteomics: A Unique Tool to Characterize the Active Microbiome in Health and Diseases, and Pave the Road Towards New Biomarkers—Example of Crohn’s Disease and Ulcerative Colitis Flare-Ups”, *Cells* 11, no. 8 (2022): 1340, <https://doi.org/10.3390/cells11081340>.

58. Y. Zhu, L. M. Orre, Y. Zhou Tran, et al., “DEqMS: A Method for Accurate Variance Estimation in Differential Protein Expression Analysis”, *Molecular & Cellular Proteomics* 19, no. 6 (2020): 1047–1057, <https://doi.org/10.1074/mcp.TIR119.001646>.

59. M. Fujita, K. Mori, H. Hara, S. Hishiyama, N. Kamimura, and E. Masai, “A TonB-Dependent Receptor Constitutes the Outer Membrane Transport System for a Lignin-Derived Aromatic Compound”, *Communications Biology* 2 (2019): 432, <https://doi.org/10.1038/s42003-019-0676-z>.

60. L. Kappelmann, K. Krüger, J.-H. Hehemann, et al., “Polysaccharide Utilization Loci of North Sea Flavobacteriia as Basis for Using SusC/D-Protein Expression for Predicting Major Phytoplankton Glycans”, *ISME Journal* 13, no. 1 (2019): 76–91, <https://doi.org/10.1038/s41396-018-0242-6>.

61. K. Krüger, M. Chafee, T. Ben Francis, 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 Journal* 13, no. 11 (2019): 2800–2816, <https://doi.org/10.1038/s41396-019-0476-y>.

62. B. Avci, K. Krüger, B. M. Fuchs, H. Teeling, and R. I. Amann, “Polysaccharide Niche Partitioning of Distinct *Polaribacter* Clades During North Sea Spring Algal Blooms”, *ISME Journal* 14, no. 6 (2020): 1369–1383, <https://doi.org/10.1038/s41396-020-0601-y>.

63. G. Reintjes, B. M. Fuchs, M. Scharfe, K. H. Wiltshire, R. Amann, and C. Arnosti, “Short-Term Changes in Polysaccharide Utilization Mechanisms of Marine Bacterioplankton During a Spring Phytoplankton Bloom”, *Environmental Microbiology* 22, no. 5 (2020): 1884–1900, <https://doi.org/10.1111/1462-2920.14971>.

64. F.-Q. Wang, D. Bartosik, C. Sidhu, et al., “Particle-Attached Bacteria Act as Gatekeepers in the Decomposition of Complex Phytoplankton Polysaccharides”, *Microbiome* 12, no. 1 (2024): 32, <https://doi.org/10.1186/s40168-024-01757-5>.

## Supporting Information

Additional supporting information can be found online in the Supporting Information section.

**Supporting Figure 1:** pmic70001-sup-0001-figuresS1-

S3.docx **Supporting Table 1:** pmic70001-sup-0001-

tableS1.xlsx **Supporting Table 2:** pmic70001-sup-0001-

tableS2.xlsx **Supporting Table 2:** pmic70001-sup-0001-tableS3.xlsx