Spatial heterogeneity in carbohydrates and their utilisation by microbial communities in the high North Atlantic

Taylor Priest[1]*, Silvia Vidal-Melgosa[1,2], Jan-Hendrik Hehemann[1,2], Rudolf Amann[1], Bernhard M. Fuchs[1]

[1] Max Planck Institute for Marine Microbiology, Bremen, Germany
[2] MARUM, University of Bremen, Bremen, Germany

***Corresponding author**
Taylor Priest
tpriest@mpi-bremen.de

ABSTRACT

Carbohydrates in the ocean are chemically and structurally diverse, represent a substantial fraction of marine organic matter and are key substrates for enzymes of heterotrophic microbes. Our current understanding on microbial carbohydrate utilisation however, is largely derived from spring phytoplankton blooms in temperate regions. Here, we characterise the carbohydrate fraction of particulate organic matter (POM) and the utilisation of carbohydrates by marine microbes during late summer in high North Atlantic waters, using a combination of analytical techniques with metagenomics and metatranscriptomics. The composition and abundance of monosaccharides and varied detection of specific polysaccharide epitopes revealed high spatial and vertical heterogeneity. Complex polysaccharides, known to accumulate in POM over time, were consistently detected, whilst more labile, simple structures, such as β-1,3-glucan, exhibited patchy distribution, indicating local variations in primary productivity and rapid utilisation by microbes. Metatranscriptomics revealed the presence of active and dynamic bacterial populations, with distinct assemblages dominating carbohydrate utilisation across samples. Transcription of genes encoding carbohydrate-active enzymes revealed substrate-based niche

partitioning among carbohydrate specialists, typically involving complex compounds such as α-mannans and alginate, along with communal substrates, such as laminarin, targeted by multiple populations in each sample. The high spatial heterogeneity observed across samples and depths, indicates that local biological and physical processes continue to shape the carbohydrate pool during late summer in high latitude waters and microbial populations remain responsive to such changes.

INTRODUCTION

Marine carbohydrates are highly diverse, chemically and structurally, and represent a substantial fraction of characterised organic matter [1]. The structural diversity of carbohydrates is owed to the numerous monosaccharide building blocks that can be connected with diverse linkages along with the substitution by a range of other chemical groups [2]. In the oceans, carbohydrates are primarily produced by micro- and macroalgae, wherein they serve structural and storage functions and can constitute between 13 – 90% of biomass [3]. Phytoplankton-derived carbohydrates range from low-molecular weight (LMW) oligosaccharides to complex high-molecular weight (HMW) polysaccharides, with the composition varying across taxa and with life stage and environmental conditions [4, 5]. Through exudation and upon cell death and lysis, a cocktail of LMW and HMW carbohydrates are released from phytoplankton and integrated into the particulate and dissolved organic matter pools (POM and DOM) [6–8]. These compounds subsequently act as valuable substrates for heterotrophic microbes.

Exogenous carbohydrate utilisation is a widespread trait amongst bacterial and archaeal taxa but the mechanisms employed and the degradative capabilities vary [9–11]. Mono-, di- and trisaccharides can be readily taken up into the cell by many species through porins or dedicated transporters, whereas uptake of oligosaccharides requires more specialised systems, such as TonB-dependent transporters (TBDTs). To make use of HMW polysaccharides, microbes must first depolymerize the polysaccharide extracellularly using specialised excreted or outer

membrane-bound glycoside hydrolases (GHs) or polysaccharide lyases (PLs), often followed by uptake of the LMW products. These enzyme classes, together with carbohydrate-binding modules (CBMs) and carbohydrate esterases (CEs), are collectively referred to as carbohydrate-active enzymes (CAZymes). CAZymes are classified into families based on sequence similarity and common ancestry [12, 13], with many such families known to target specific glycosidic linkage types within polysaccharides. Some complex marine polysaccharides are decorated with chemical groups such as sulfates, which require additional genetic machinery for complete degradation, e.g. sulfatases. As such, comparing CAZyme gene profiles in conjunction with sulfatases and transporters can provide valuable insights into the carbohydrate utilisation potential of microbes [10, 14].

Carbohydrate utilisation by microbial populations exhibit spatial and temporal variations. The rate of hydrolysis and the substrate spectrum of extracellular CAZymes has been shown to reduce and narrow with depth [15] and with distance from the coast [16]. In addition, a latitudinal gradient has also been evidenced, with a broader spectrum of CAZyme activities measureable in temperate compared to high latitude waters [17]. Temporal shifts in CAZyme, sulfatase and transporter gene profiles are also evident following spring phytoplankton blooms [18]. These patterns are congruent with microbial population dynamics and shifts in microbial community composition. In particular, community-level patterns are shaped by the presence and composition of specialised carbohydrate degraders, such as members of the *Bacteroidetes. Bacteroidetes* typically harbour large CAZyme repertoires [10, 14] and exhibit distinct successional-like dynamics following spring phytoplankton blooms [18]. These dynamics have been described as evidence of substrate-based niche partitioning and thus driven by the availability of substrate [19, 20]. Such detailed assessments on microbial carbohydrate utilisation have been largely focused on spring phytoplankton blooms in temperate ecosystems. Whether the same microbial clades occupy discrete substrate-based niches at later seasons and if these patterns are also evident in higher latitude waters is yet to be determined.

In this study, we combine analytical techniques to characterise the carbohydrate fraction of particulate organic matter (POM) along with metagenomics and metatranscriptomics to assess microbial carbohydrate utilisation patterns in high latitudes of the North Atlantic during late summer. Through the application of high performance anion-exchange chromatography- and carbohydrate microarray-based approaches, we were able to not only quantify monosaccharides from the particulate fraction but also assess spatial and vertical distribution patterns of specific polysaccharide epitopes in the upper euphotic zone. Concurrently, combining novel analysis pipelines for long PacBio HiFi metagenome reads in concert with short-read metatranscriptomes, we could characterise shifts in the potential and active carbohydrate utilisation patterns of entire microbial communities as well as specific populations.

METHODS

Sample collection

Seawater samples were collected from ten stations located in the eastern Fram Strait and around the Svalbard archipelago in September 2020 during the MSM95 Maria S. Merian research cruise [21]. A map of the sampling locations (Figure 1) was generated using publically available bathymetric data from the International Bathymetric Chart of the Arctic Ocean (IBCAO) [22, 23] and the QGIS v3.14.16-Pi (qgis.org) software. Seawater was collected using a CTD-rosette sampler from surface water (SRF), typically 2 m depth, and the bottom of the surface mixed layer (BML). The BML depth was defined by the beginning of the thermo- and halocline and drop in surface fluorescence values (Figure 1). One location was sampled twice over a two day period, with additional samples collected at 100 and 200 m depth, these are labelled as S1 and S6 (Supplementary Table S1). Of the water collected, 4 L was filtered sequentially through a 3 and 0.2 µm pore-size polycarbonate membrane filter (142 mm diameter) and immediately stored at -80 °C for ‘omics analysis. A second 4 L of seawater was filtered through a 0.7 µm pre-combusted

Whatman Grade GF/F filter (47 mm diameter) and immediately stored at -80 °C for carbohydrate analysis.

Monosaccharide and polysaccharide analysis

The GF/F filters from all samples (ten stations and two depths) were cut into ten equally-sized circular pieces, with a diameter of 10 mm. The monosaccharide composition of and polysaccharide structures present on the filter pieces was analysed as described previously [24] and detailed in Supplementary Methods. First, two of the filter pieces were hydrolysed using acid (1 M HCl for 24 h at 100 °C) and the resulting monosaccharides were analysed using High-Performance Anion Exchange Chromatography with Pulsed Amperometric Detection along with monosaccharide standards. Acidic monosaccharides could not be detected due to a problem with the detector. The remaining filter pieces were subject to polysaccharide extraction using a sequential solvent protocol, with 1) MilliQ water, 2) 0.2 M EDTA (pH 7.5) and 3) 4 M NaOH with 0.1% w/v $NaBH_4$. The identification and semi-quantitative analysis of polysaccharide compounds was performed using a microarray and antibody-based approach. Polysaccharide extracts were printed in quadruplicates onto nitrocellulose membranes (0.45 µm pore-size) using a microarray robot (Sprint, Arrayjet, Roslin, UK). The printed arrays were blocked for 1 h in 1 x PBS with 5% (w/v) non-fat milk powder (MPBS), followed by incubation for 2 h with polysaccharide-specific monoclonal antibodies (Supplementary Table S2). After incubation, the arrays were washed in PBS and incubated for 2 h in anti-rat, anti-mouse or anti-His tag secondary antibodies conjugated to alkaline phosphatase. Arrays were thoroughly washed in PBS and deionized water before being developed in a solution containing 5-bromo-4-chloro-3-indolylphosphate and nitro blue tetrazolium in alkaline phosphatase buffer (100 mM NaCl, 5 mM $MgCl_2$, 100 mM Tris-HCl, pH 9.5). Developed arrays were scanned and the binding of each probe against each spotted sample was quantified using Array-Pro Analyzer 6.3 (Media Cybernetics). Mean antibody binding signal

intensity (from four replicates) was determined for each extract of each sample, with the highest mean signal of the dataset being set to 100 and the remainder normalised accordingly.

Metagenome and metatranscriptome sequencing

Filtered seawater samples of the 0.2 – 3 µm fraction from SRF and BML depths of four different stations (S8, S10, S25, S26) were subject to a dual nucleic acid isolation protocol using the DNA/RNA Mini Prep Plus kit from Zymo Research (Irvine, CA, USA), according to the manufacturer's instructions. The quality of extracted DNA was assessed using capillary electrophoresis with a FEMTOpulse (Agilent), whilst RNA quality was assessed using a PicoChip on a Bioanalyser (Agilent, CA, USA). Ultra-low DNA libraries were prepared from the eight samples without further fragmentation by the protocol "Procedure & Checklist - Preparing HiFi SMRTbell® Libraries from Ultra-Low DNA Input" of PacBio (CA, USA). Libraries were sequenced on 4 x 8M SMRT cells on a Sequel II platform for 30 h with sequencing chemistry 2.0 and binding kit 2.0 (two samples multiplexed per SMRT cell). Four of the samples were additionally selected for metatranscriptome sequencing (S10_surface, S25_Surface, S25_BML, S26_BML). Illumina-compatible libraries were produced from extract RNA using the Universal Prokaryotic RNA-Seq library preparation kit, incl. Prokaryotic AnyDeplete® (Tecan Genomics, CA, USA). Libraries were sequenced on a HiSeq 3000 platform with 2 x 150 bp paired-end read mode.

HiFi read taxonomic classification

A custom pipeline was employed to taxonomically classify metagenomic HiFi reads against a GTDB-based protein database. A Diamond blast (v2.0.14) [25] database was generated from the gene amino acid sequences of all GTDB species-representatives (release 207) after clustering at 99% sequence identity, to remove redundancy. NCBI-style taxdump files (nodes.dmp, names.dmp and accession2taxid) were then generated using scripts from https://github.com/nick-youngblut/.

Open reading frames were predicted on raw HiFi reads using FragGeneScan v1.31 [26]. Gene sequences were aligned to the generated GTDB protein database using Diamond blastp (parameters: --id 50 --top 5 --fast). A secondary filtering step was applied to the output including identity threshold >65%, e-value <1E-10 and query-cover >50%. Using the remaining hits, a single taxonomic classification for each gene was determined using a last common ancestor approach, *lca* command from TaxonKit [27]. To further increase the number of genes taxonomically classified, the last common ancestor algorithm was then applied to all genes within a single HiFi read, resulting in a single taxonomic classification for each HiFi read, and it's containing genes.

HiFi read functional annotation

For functional characterisation, the predicted gene sequences (see above) from HiFi reads were subject to a custom annotation pipeline, modified from Priest *et al.* [28]. In brief, genes were annotated against the Pfam database (release 35.0) using HMMsearch v3.3.2 (parameters: cut_ga), UniProtKB database (05.2022) using Diamond v2.0.14 blastp (parameters: -k 1 --evalue 1e-10 --query-cover 50 --id 40 --sensitive) and the KEGG database (07.2022) using kofam_scan (https://github.com/takaram/kofam_scan ; parameters: -E 0.0001). Additional annotations were obtained by searching more specialised databases using HMMScan (Transporter Classification database; obtained 11.2021, TonB HMM profiles from TIGRFAM, and dbCAN; v10) and Diamond blastp (CAZyDB; release 09242021, SulfAtlas; v1.3, MEROPS; v12.1) using the same settings as described above, except for HMMScan against the dbCAN database (parameters; -E 1E-15).

Single-copy ribosomal protein (SC-RBP) gene analysis

From the gene annotations, 16 single-copy ribosomal protein (SC-RBP) genes [29] were identified and extracted from each metagenome. The average number of SC-RBP was used as a proxy for the number of genomes recovered in each metagenome. A subset of four SC-RBP genes (RBP L3, L4, L6 and S8) were clustered at previously defined gene-specific ANI thresholds [30] and the

average number of clusters across the four was used as a proxy for the number of species captured. The composition of metagenomes and metatranscriptomes was compared based on the species clusters from the RBP L6 gene, selected due to its high recoverability and species delineation accuracy [30]. The taxonomy of each cluster was determined based on a majority vote between the taxonomy of the contained genes, derived from original HiFi read classifications.

Assembly, binning and metagenome-assembled genome recovery

The assembly of Hifi reads was performed using MetaFlye v2.8 [31] (parameters: --meta –pacbio-hifi –hifi-error 0.01 –keep-haplotypes). Coverage information was obtained through mapping HiFi reads to assembled contigs using Minimap2 v2.1 (parameters: -x map-hifi –MD). Contigs were binned using Metabat2 [32]. The resulting bins were subject to manual refinement using the Anvi'o v7 [33] interactive interface to generate metagenome-assembled genomes (MAGs). MAGs were dereplicated at a 99% ANI threshold using dRep v3.2.2 (parameters: --comp 50 --con 5 --sa 0.99 --nc 0.6). The completeness and contamination of representative MAGs was estimated using CheckM v1.1.2 [34]. A two-pronged approached was used for taxonomic classification of MAGs, the classify_wf pipeline of GTDB-tk v2.1.1 [35] (Release 207) and the extraction of 16S rRNA gene sequences using Barrnap v0.9 [36] and classification against the SILVA_SSU_Ref138.1_NR99 database, following the same process described in '*Phylogenetic characterisation of communities'*. Of the species-representative MAGs, 84% contained a complete 16S rRNA gene and thus received dual taxonomies.

MAG relative abundance estimation

The relative abundance of representative MAGs was determined using a similar approach to Orellana et al. [37]. In brief, reads were competitively recruited from each metagenome to the MAG representatives. Mapped reads were converted into depth values using Genomecov (-bga option) from the Bedtools package [38] and the 80% central truncated average of the sequencing

depth (TAD) was determined using the 'BedGraph.tad.rb' script (option range 80) from the enveomics collection [39]. The relative abundance was then determined as the quotient between the TAD value and the number of microbial genomes captured in each metagenome, determined from SC-RBP genes.

MAG functional characterisation

The functional characterisation of MAGs was performed following the same procedure described for the HiFi reads above except for an additional process of polysaccharide utilisation loci (PULs) detection. PULs were defined as genetic loci containing a SusC/SusD gene pair with two or more degradative CAZymes or the presence of at least three degradative CAZymes in close proximity (maximum six genes apart). PULs were manually inspected and visualised at BioRender.com.

Transcription level of genes at the community- and MAG-level

Adapters and low quality reads were removed from the metatranscriptomes using BBDuk of the BBtools program v38.73 (http://bbtools.jgi.doe.gov) (parameters: ktrim=r, k=29, mink=12, hdist=1, tbo=t, tpe=t, qtrim=rl, trimq=20, minlength=100). Although an rRNA depletion step was performed prior to sequencing, it is expected that 5 – 15% of reads would still be related to rRNA. As such, SortMeRNA v2.0 [41] was used to filter out rRNA sequences from the dataset, with the SILVA SSU Ref 138 NR99 database as a reference. The transcription level of genes was determined by read recruitment of transcripts to the predicted gene sequences from the HiFi reads using BBmap (v35; parameters: minid=98 idfilter=98). Mapped read values were converted to transcripts per million (TPM), according to Wagner *et al.* [42]. For MAGs, transcripts were competitively recruited to all MAG genes, with same procedure as above, and the values were converted to TPM using the total number of transcripts recruited to the whole metagenome-predicted genes as the total transcript values. In order to compare the transcription level of MAGs across samples, we determined the average TPM value of the 16 SC-RBPs for each MAG in each sample, and took

the quotient of this and the average TPM value of the 16 SC-RBPs in the whole sample – providing proportional transcription of all genomes recovered.

RESULTS & DISCUSSION

Seawater samples were collected from surface waters (SRF) and the bottom of the surface mixed layer (BML) in the Eastern Fram Strait region to investigate the distribution of carbohydrates and their utilisation by microbial communities (Figure 1). The ten sites were grouped into three categories based on the underlying seafloor topography (above-slope, above-shelf and open-ocean) as this also corresponds to differences in hydrographic conditions. The main water mass in this region is of North Atlantic origin. The West Spitsbergen Current (WSC) transports Atlantic water northward into the Arctic Ocean, with the main branch flowing above the continental slope. At the shelf break, a temperature salinity-front occurs, whilst above the West Spitsbergen shelf,

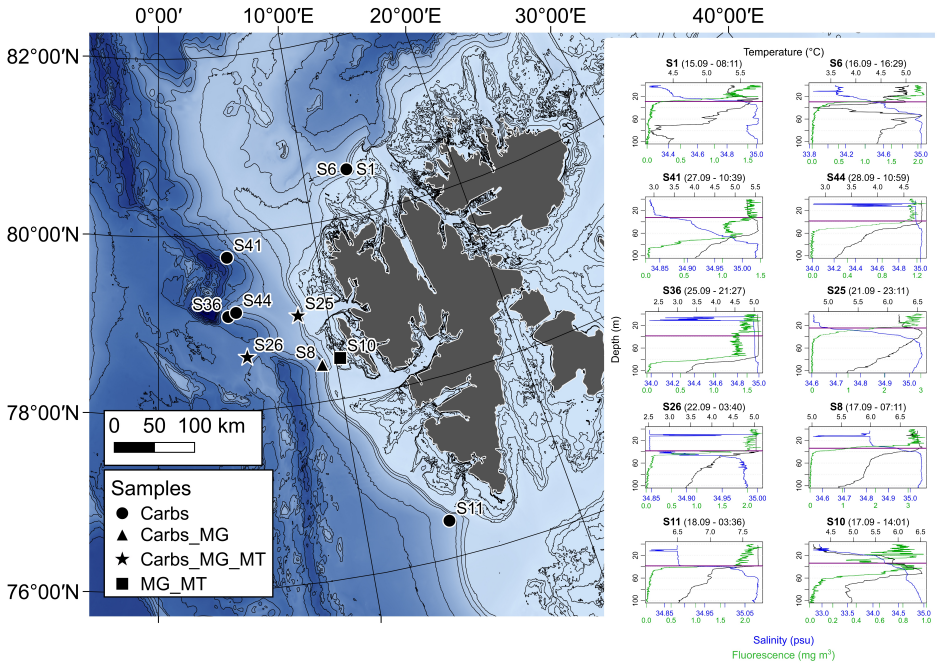

Figure 1. Bathymetric map with sampling locations, types of samples collected and vertical profiles from CTD casts. Map legend: Carbs = samples collected for carbohydrate analysis on particulate organic matter fraction, MG = sampled for metagenomics, MT = sampled for metatranscriptomics. Vertical profiles of temperature, salinity and fluorescence were derived from sensor measurements during CTD casts. The purple horizontal lines on each profile represent the bottom mixed layer (BML) sampling depth.

Atlantic water (AW) converges and mixes with Arctic water and freshwater from land, resulting in intra-annual variability in hydrographic properties [43]. Previous studies have outlined typical thresholds that can be used to distinguish between the main water masses in this region, which include >34.9 psu and >4.1 °C for AW [44]. The temperatures observed in SRF and BML samples were indicative of AW, ranging from 4.1 – 7.7 °C, however, the salinity values of above-shelf (S10) and three above-slope stations (S1, S6 and S8) were below these thresholds (Figure 1 and Supplementary Table S1). These observations suggest an influence of either polar-derived water or freshwater from Spitsbergen in the surface waters of these stations.

Carbohydrate analysis of POM samples

Carbohydrates of POM (>0.7 µm) in SRF and BML samples were analysed with two different approaches. POM carbohydrates were hydrolysed with acid and the resulting monosaccharides were quantified by chromatography revealing monosaccharide depth- and location-related patterns (Figure 2). Total neutral and amino monosaccharide concentrations ranged from 1.4 – 13.8 µg per L of seawater (hereon µg L). Higher values were typically observed in SRF compared to BML samples, average of 8.8 µg L and 5.2 µg L, respectively (Supplementary Table S3). However, the magnitude of change between the two depths was station-dependent, from a negligible difference at station S8 to a threefold decrease from SRF to BML depths at station S6 (Figure 2a). These patterns are in agreement with previous findings that showed decreasing monosaccharide concentrations with depth [45]. In addition, stations categorised as above-slope typically contained higher monosaccharide concentrations than open-ocean stations (Figure 2 and Supplementary Figure S1). During early summer in this region, chlorophyll *a* and dissolved organic compounds have also been shown to reach higher concentrations in surface waters

above the continental slope (~8 °E) [46]. These patterns likely reflect hydrographic processes, such as the frontal zone situated above the shelf break.

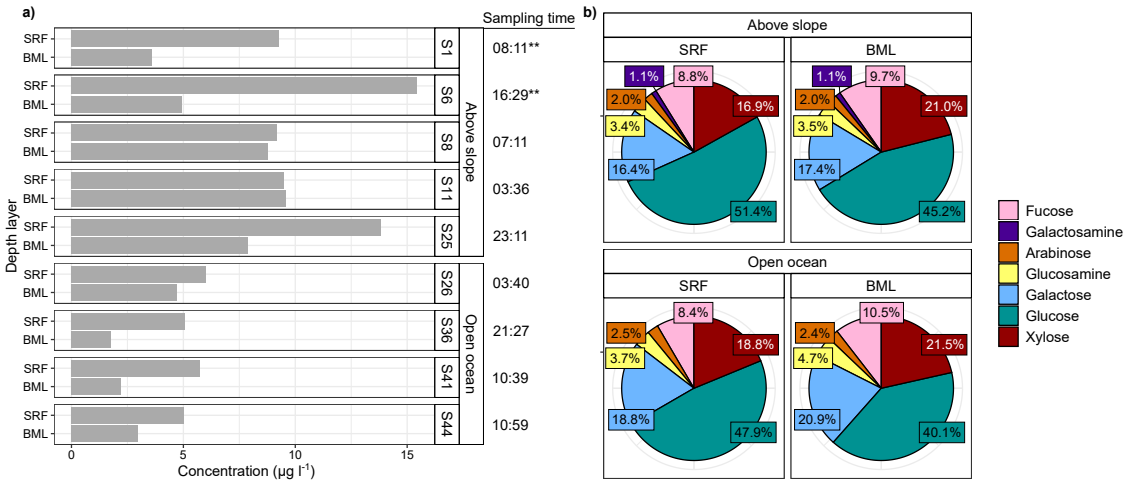

Figure 2. Total concentration and relative abundance of monosaccharides in carbohydrates from POM fraction. a) Total monosaccharide concentrations as ug L of seawater at each station and depth**, b)** Relative abundance of monosaccharides grouped by station location in relation to continental slope. Data show neutral and amino monosaccharides. SRF = Surface water sample, BML = sample from bottom of the surface mixed layer.

The most abundant monosaccharide detected in all samples was glucose. The relative proportion of glucose was higher in SRF than BML depths, with an average of 49% and 43% respectively, and in above-slope compared to open-ocean samples, average of 47% and 44% respectively (Figure 2b). These proportions are in line with previous observations in marine surface waters (31 – 55%) [47–49] but lower than during a phytoplankton bloom, where concentrations can reach >70% [50]. Furthermore, in contrast to findings from the high North Pacific [47], we observed a decrease in glucose proportions with depth and an increase in all other monosaccharides. This pattern was most pronounced for xylose that, on average, increased by 5% in relative proportion from SRF to BML depths (Figure 2b). This suggests a preferential utilisation of glucose-based carbohydrates in surface waters.

Carbohydrates of POM were also analysed using carbohydrate microarrays combined with molecular probes, detecting 17 distinct polysaccharide epitopes (Supplementary Figure S2 and Supplementary Table S4). This provides a semi-quantitative insight into the presence and abundance of polysaccharide epitopes, with changes in antibody binding signal correlating to

epitope abundance [24, 51]. Variations in the detection and abundance of epitopes revealed location and sample-specific patterns. The polysaccharide epitopes observed most frequently included fucose-containing sulfated polysaccharides (FCSPs) and glucuronoxylan in 91% of samples (Figure 3). FCSPs are unique to and important structural components of brown algal cell walls [52] whilst glucuronoxylans are features of terrestrial plant and macroalgae cell walls [53]. However, both of these complex polysaccharides have been previously detected during microalgal blooms [24]. In particular, FCSPs have been evidenced as a main carbohydrate excreted by diatoms in culture [54] and shown to accumulate in POM over a period of weeks following a spring phytoplankton bloom [24] as well as contribute to long-term carbon sequestration in sediments [55]. In contrast to this, β-1,3-glucan is a structurally simple polysaccharide (laminarin-like) that has previously been shown to accumulate only for days to a week in POM during a phytoplankton bloom [24]. This polysaccharide was absent in some open-ocean samples, but present in above-slope sites (Supplementary Figure S2. Due to its lability and the widespread capacity for its degradation in marine bacteria [14], the distribution of β-1,3-glucan could indicate patchiness in recent phytoplankton production, which reflects previous observations from this region [56]. In combination with the changes in monosaccharide concentrations, the carbohydrate fraction of POM could be predominantly characterised as that of a late summer setting but with above-slope samples also showing signs of more recently produced organic matter.

The collection of samples from the same location over a two day period (stations S1 and S6) and at additional depths (100 and 200 m) provided further insights into spatiotemporal changes (Supplementary Figure S4). S1 samples were retrieved at 08:00 whilst S6 samples were collected one day later at 16:30. For the hydrolysed POM carbohydrates, temporal shifts in absolute monosaccharide concentrations were observed at all depths, with S6 samples

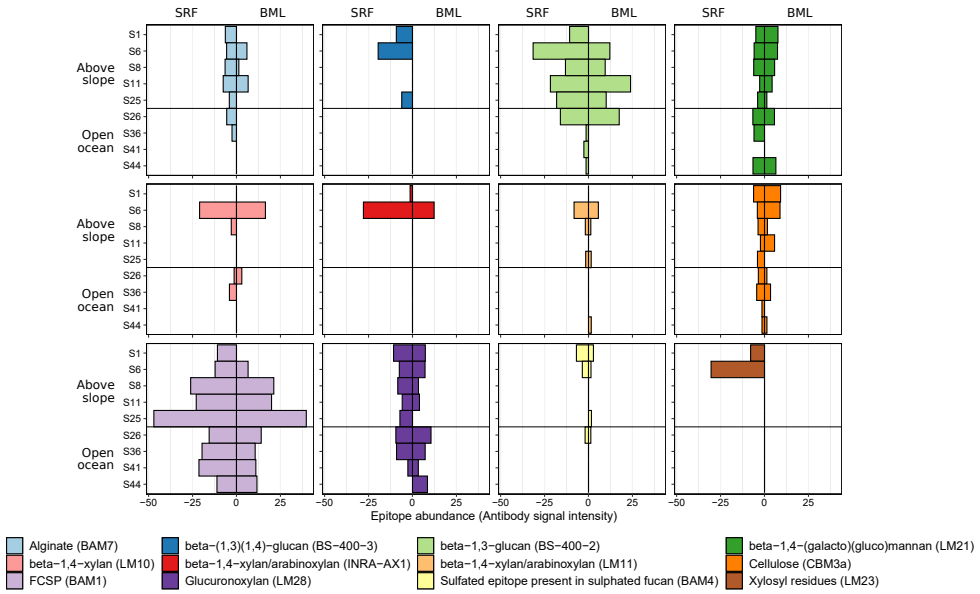

Figure 3. Polysaccharide structures in the particulate organic matter fraction. Plots show the relative abundance (antibody signal intensity) of polysaccharide epitopes detected in POM samples. Signal intensities for each extract against each antibody were quantified and the highest mean signal value (n=4 replicates per extract) in the data set (which corresponded to a standard control) was set to 100 and all other values were normalised accordingly. The epitope abundances shown here represent the summed values from the mean signal of H_2O, EDTA and NaOH extractions. Each antibody has a different avidity, thus the absolute numbers should not be compared between probes to infer different concentrations, but signal from a single antibody determines the relative abundance of the epitope within the sample set. Specific polysaccharide structures that the molecular probes (depicted in parentheses) bind to are shown in the legend. SRF = Surface water sample, BML = bottom of the surface mixed layer. FCSP = Fucose-containing sulphated polysaccharide. Some epitopes that were only detected at low intensity in <3 samples were not included in this plot.

containing 1.1 – 3.6 x higher concentrations than S1 samples. In SRF depths, glucose concentrations were twofold higher at S6 along with the presence of arabinose that was not detected in S1 samples. Clear temporal changes in polysaccharide structures were also observed, particularly with respect to xylans. Most notably, in SRF depths, S6 contained a higher signal intensity of α-1,5-arabinan and β-1,3-glucan as well as β-1,4-xylan epitopes and xylosyl residues. The only epitopes that had a higher signal intensity in S1 samples were alginate, and glucuronoxylan. It is important to note that the alginate targeting antibody used, BAM7, also has cross reactivity with FCSP. The pronounced shift in glucose and β-1,3-glucan between the two sampling time points may reflect diurnal fluctuations in laminarin production by diatoms (Becker et al. 2020), with highest production during the day and a partial consumption at night. Such diel periodicity is also evident at the particulate organic carbon level [57], with accumulations during

the day to a maximum concentration at dusk strong diel periodicity [58]. In addition, the pronounced increase in xylosyl residues and β-1,4-xylan epitopes at S6 could suggest the introduction of, or change in the primary producers between the two sampling time points, potentially resulting from changes in water mass dynamics.

Whole microbial community analysis

In order to assess microbial carbohydrate utilisation patterns, eight PacBio HiFi read metagenomes were generated from SRF and BML depths of two above-slope (S8 and S25), one open-ocean (S26) and one above-shelf (S10) station. From four of the samples, metatranscriptomes were also generated (Supplementary Table S5 and S6). Although samples for monosaccharide and polysaccharide characterisation were not obtained for the above-shelf station S10, the lower salinity values observed in these samples (psu < 33) suggests an influence from polar marine/freshwater and thus we predict it may provide a valuable comparison.

An initial taxonomy-independent comparison was performed with previously published metagenomes from the Fram Strait [28] and Arctic Ocean [59] in order to place the sampled communities into context. Based on sequence composition dissimilarity, our metagenomes were most closely related to those previously generated from WSC, high North Atlantic and Barents Sea samples in June and July and most dissimilar to those from the polar water mass of the western Fram Strait (East Greenland Current) (Supplementary Figure S5). This indicates that the communities captured by the metagenomes are representative of summer communities in a North Atlantic water mass, which is also in agreement with the post spring phytoplankton bloom composition of carbohydrates in POM.

As the focus of this study was on microbial carbohydrate utilisation, we initially classified reads at kingdom-level in order to remove eukaryotic sequences. Across the metagenomes, 16 – 46% of reads were identified as eukaryotic (Supplementary Figure 6a), which is surprising due to the size fractionated filtration used during sampling (see Methods). The analysis conducted

hereon was performed only on the metagenomic fraction classified as prokaryotic in origin, however we extracted and analysed 18S rRNA genes to provide insights into the eukaryotic taxa present in each sample (Supplementary Figure S7). Using the average sequencing depth of single-copy ribosomal protein (SC-RBP) genes, we determined that the number of microbial genomes sequenced ranged from 761 in S8_SRF to 1619 in S10_BML. Furthermore, using a subset of SC-RBPs [30], we determined that the number of species captured ranged from 374 – 554 (Supplementary Figure S6b). Combining these values and normalising them by sequencing depth indicates differences in species evenness across metagenomes, with an almost twofold variation in the number of genomes per species per million reads, 2.0 – 3.7, observed across the metagenomes.

Composition of metagenome and metatranscriptome microbial communities

The composition and structure of microbial communities varied across samples, with no consistent patterns in relation to location or depth. The most prominent families identified across the metagenomes were the *Flavobacteriaceae* (4 – 14%), *Rhodobacteraceae* (3 – 13%), D2472 (SAR86; 5 – 12%) and *Poseidoniaceae* (2 – 12%) except for in S25_BML, which was dominated by *Alteromonadaceae* (25%) (Supplementary Figure S8). The same families were also the largest contributors to community transcription, although in some cases, the proportions were markedly different, such as *Flavobacteriaceae* that represented 14 - 30% relative proportion of transcription. Twenty genera were identified as reaching >2.5% relative abundance and together represented, on average, 42% of the microbial communities (Figure 4). These included *Pseudoalteromonas* (0 – 29%), MGIIa-L1 (<0.1 - 10%), *Amylibacter* (<0.1 – 6%), D2472 (SAR86; 3 – 6%), ASP10-02a (<0.1 – 6%) and HTCC2207 (SAR92; 1 – 6%). However, only some of these genera were amongst those contributing the most to community transcription. The maximum relative proportions observed for genera in the four metatranscriptomes were for *Pseudoalteromonas* (45%), *Vibrio* (29%), *Flavobacterium* (13%), *Amylibacter* (9%), SAR86 (9%) and *Pseudothioglobus* (9%).

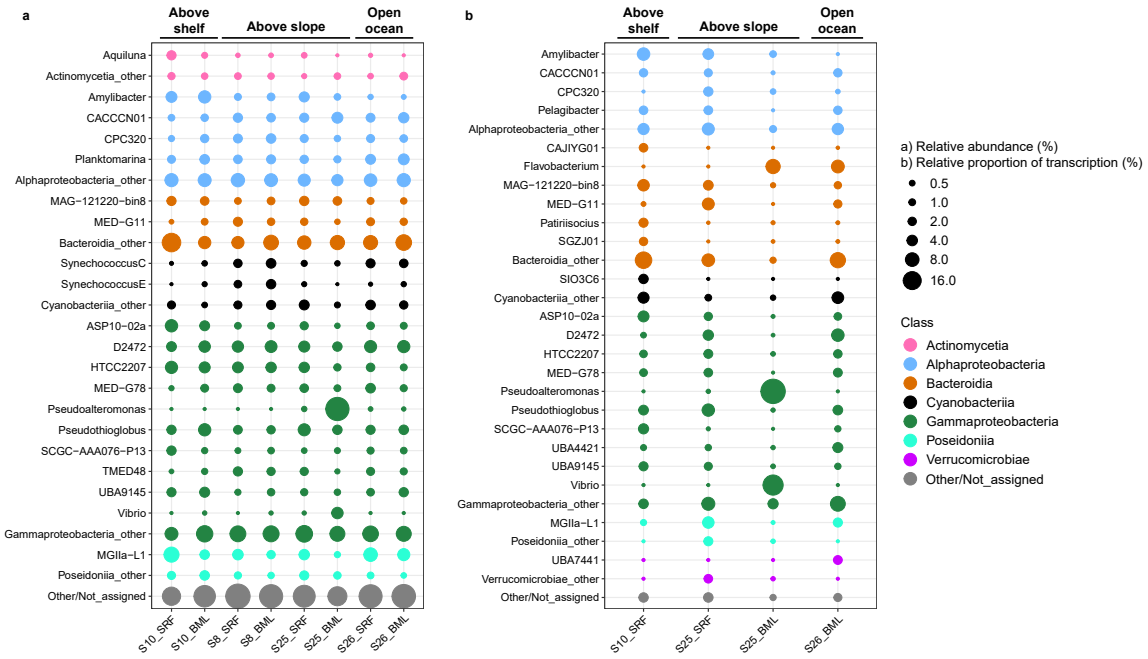

Figure 4. Genus-level composition and transcription of sampled microbial communities based on ribosomal protein L6. Relative abundance is described as the relative proportion to total ribosomal protein L6 sequences identified in the sample. Relative proportion of transcription is determined as the relative proportion to total sample ribosomal protein L6 TPM values. Taxonomy of ribosomal protein L6 sequences was derived from taxonomic classification of the original HiFi read, which was classified using a GTDB-based database.

Discrepancies in the community structures between metagenomes and metatranscriptomes highlight that gene abundance is not always intrinsically linked to transcription. This was particularly evident for MGIIa-L1, which represented a larger proportion of gene abundance than transcription, and genera within the *Bacteroidia*, such as *Flavobacterium* and MAG121220-bin8 (NS4), which showed the opposite trend. Such patterns are not uncommon in marine microbial communities [60] and can be influenced by lifestyle, fitness and differences in metabolism.

Sample S25_BML harboured a distinct community from all others, with a large proportion of *Pseudoaltermonas* and *Vibrio.* These genera have previously been observed under nutrient-rich conditions and are known to be associated with eukaryotic hosts and phytoplankton blooms [61, 62]. Considering the high proportion of eukaryotic reads captured in some of the metagenomes, the observed pattern in S25_BML may represent signals of processes occurring in the larger size fraction. In support of this, the eukaryotic community of this sample (Supplementary Figure S7) contained a large proportion of 18S rRNA genes affiliated with

copepods, whose microbiomes are typically dominated by *Pseudoalteromonas* and *Vibrio* [63, 64].

In general, the microbial community compositions observed here more closely resemble those previously described from summer [65, 66] than from late September in 2018 in the WSC [67]. This likely reflects the inter-annual variability in seasonal transitions, as September is considered the shoulder month between summer and the on-set of winter. However, methodological differences also play a major role. This is particularly evident with respect to the high proportions of MGIIa-L1 in our samples, which has not been observed before in the WSC, and is typical of late summer communities in temperate coastal ecosystems [68]. The previous employment of bacterial-specific 16S rRNA gene based primers [66] and other potential biases in methodologies have led to the Marine Group II Archaea, and Archaea more generally, being overlooked in the WSC, wherein they could represent an important fraction of the microbial community.

Whole community carbohydrate utilisation patterns

Microbial carbohydrate utilisation was assessed through the abundance and transcription of carbohydrate-active enzymes (CAZymes) in relation to the number of microbial genomes captured. In particular, we focused on those involved in degradative processes (glycoside hydrolase, GH; carbohydrate-binding module, CBM; carbohydrate esterase, CE; polysaccharide lyase, PL), hereon collectively referred to as CAZymes.

CAZyme genes represented a minor proportion of total gene content but a higher proportion of gene transcription. The number of CAZyme genes ranged from 7 – 19 per microbial genome (PMG), which corresponded to 0.3% of total sample gene content, on average. In contrast, CAZyme genes represented between 1.5% and 3.0% of total gene transcription in the four metatranscriptomes. Clear differences were also observed with respect to the number of CAZyme genes being transcribed, with 3159 genes in S10_SRF compared to 494 genes in

S25_BML. Employing a dissimilarity-based approach, we observed that the CAZyme compositions of samples from S8 and S26 were grouped together based on station (location) whereas those of S10 and S25 showed no coherent clustering (Supplementary Figure S9).

The CAZyme gene profiles comprised a core backbone of universally abundant and transcribed gene families. The most abundant gene families in all samples were those involved in peptidoglycan synthesis and degradation, which is in agreement with previous observations and reflects the core machinery required for bacterial cell membrane construction and maintenance [18]. These included, CE11, GH23, GH103 and GH73 that together represented, on average, 27% of CAZyme gene abundance, with an average of 0.83 PMG for each, and 35% of CAZyme gene transcription (Figure 5 and Supplementary Tables S7 – S10). Several CAZyme gene families involved in the degradation of algal-derived polysaccharides also represented a high proportion of gene abundance and transcription. The GH16_3 gene family, known to encode enzymes for degradation of diatom laminarin constituted 3.1% of CAZyme gene abundance (0.3 PMG) and 2.7% of CAZyme gene transcription. In addition, gene families known to target sialic acids (GH33; average of 1.4% relative abundance or and 1.7% relative transcription), α-mannan (GH92; average of 1.7% relative abundance and 1.3% relative transcription) and α-fucan (GH29; average of 1.3% relative abundance and 1.2% relative transcription) were also among the most abundant (Figure 5 and Supplementary Tables S7 – S10). The transcription of these, and most, CAZyme gene families was tightly coupled to gene abundance, with strong linear relationships being observed in all samples except S25_BML (Supplementary Figure S10) – a higher abundance resulted in higher transcription. Despite this, there were cases where abundance and transcription were disconnected, with relative proportions of transcription being about twofold higher than abundance (Supplementary Figure S11). The predicted substrates of these gene families are also known to be derived from algae and include the alginate-targeting PL7_5 in S10_SRF and the galacturonan-targeting GH28 in S26_BML. Although the microbial community analysis was based on the free-living fraction, some of these patterns corresponded to the

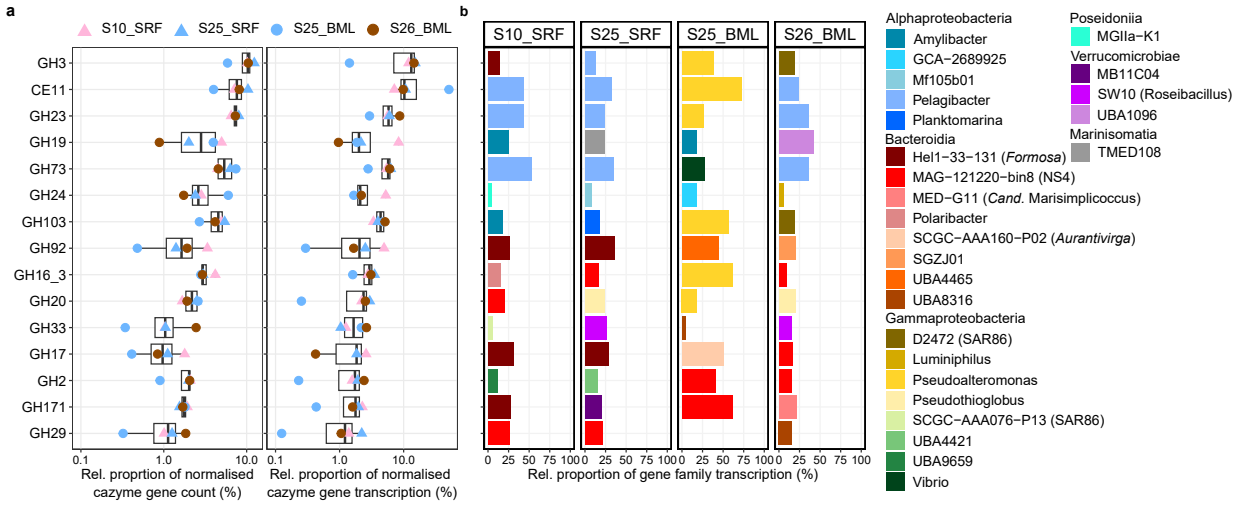

Figure 5. Abundance, transcription and taxonomic information of the CAZyme gene families with the highest proportional transcription across samples. a) The abundance and transcription of carbohydrate-active enzyme (CAZyme) genes was normalised by the number of microbial genomes per sample and then converted to relative proportions. The selected CAZyme gene families visualised are those that reached the highest proportions of transcription across samples. **b)** The genera that contributed the most to transcription of each gene family in each metatranscriptome. Taxonomic classification of CAZyme genes was derived from the classification of the read from which it was derived.

distribution of carbohydrates in POM, such as the widespread presence of FCSP and alginate and the transcription of α-fucan- and alginate-targeting genes. Carbohydrates that are part of the algae cell belong to the POM pool, but they can become part of the DOM pool due to processes such as algal cell lysis, viral infection and grazing (in several of the here studied stations presence of copepods was confirmed by our 18S data). Also, the majority of polysaccharide epitopes detected during spring phytoplankton blooms are present in both DOM and POM pools [24], suggesting that most of the carbohydrates we detected in POM were also present in DOM and thus available to the microbial free-living fraction. The capacity to degrade algal-derived polysaccharides, such as laminarin and α-mannan, is not a widespread trait amongst marine heterotrophs and indicates the presence and activity of carbohydrate-degrading specialists responding to substrate availability.

We observed two distinct patterns when comparing dominant microbial taxa responsible for transcription of CAZyme gene families in each sample. The first pattern was exhibited in, for example, the peptidoglycan recycling gene families (e.g. GH23 and GH73), whereby a single

genus dominated transcription (*Pelagibacter* that accounted for up to 45% of the gene family transcription). This pattern is apparent in gene families that are common in marine microbes, resulting in abundance and transcription patterns determined by the taxa with the highest population size. In contrast, the transcription of other gene families, particularly those targeting algal-derived polysaccharides, was dominated by distinct *Bacteroidia*-affiliated genera across samples, such as *Polaribacter* and NS4 for laminarin, *Formosa* for α-mannan and NS4 and UBA8316 for α-fucan. Several of these genera are well known as carbohydrate-degrading specialists in temperate coastal ecosystems and are among the main microbial responders to spring phytoplankton blooms [18, 69]. In particular, *Polaribacter* and *Formosa* have been identified as annually recurring after phytoplankton blooms in the German Bight, North Sea, reaching > 5% relative abundance, and exhibiting successional-like dynamics indicative of substrate-based niche partitioning [10, 18]. The presence and activity of these genera in our samples indicates that they are also key players in carbohydrate cycling at higher latitudes of the North Atlantic and at later seasonal stages.

To provide higher resolution insights into spatial distribution patterns of carbohydrate utilisation and assess if substrate-based niche partitioning is evident amongst the carbohydrate-degrading specialists, we investigated patterns at the population-level using metagenome-assembled genomes.

Recovery of metagenome-assembled genomes

We recovered and analysed 83 population-representative metagenome-assembled genomes (MAGs), delineated at a 99% ANI threshold (Supplementary Information S1 and Supplementary Table S8). The MAGs captured between 48 – 88% of the microbial metagenomic reads (Supplementary Table S12) and 11 – 37% of the microbial metatranscriptomic reads (Supplementary Table S13) (Figure 6). Similar to the patterns observed in the community-level analysis, the coupling of relative abundance and relative proportion of transcription varied across

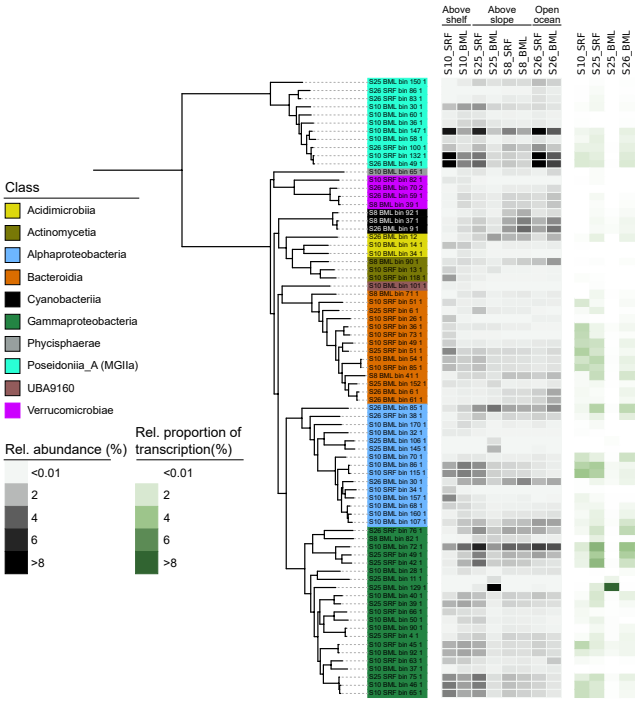

Figure 6. Population-representative metagenome-assembled genome phylogenetic tree with relative abundance of CAZyme genes and relative proportion of transcription. The tree was calculated from a concatenated alignment of 16 single-copy ribosomal protein genes (SC-RBPs), with a threshold of at least 8 genes per MAG for inclusion in the tree. Relative abundance of MAGs was defined as the quotient between the truncated average sequencing depth of the MAG and the average sequencing depth of 16 SC-RBPs in the sample. Relative proportion of transcription was defined by the average TPM value of the MAG-encoded 16 SC-RBPs and the whole sample average TPM values for the 16 SC-RBPs.

MAGs and typically differed between taxonomic groups. For example, MGIIa-L1-affiliated MAGs exhibited high relative abundance (up to 12%) but a low proportion of transcription (up to 1.5%) whilst others showed the opposite trend, such as the *Aurantivirga* MAG with up to 0.7% relative abundance but 2.7% relative proportion of transcription. A recent study describing the ecology of MGII representatives in the German Bight, identified MGIIa-L1 as being indicative of post phytoplankton bloom situations and distinct from the dominant populations during winter that were affiliated with the MGIIb [68]. This observation is consistent with the late summer sampling conducted in this study. One MAG recovered, S25_BML_bin_129, shares 97.1% ANI to the cultured representative of *Pseudoalteromonas primoryensis,* and constituted 22.8% relative abundance of sample S25_BML. This indicates that the dominance of *Pseudoalteromonas* observed in the S25_BML sample reflects a single species.

Population-level carbohydrate utilisation patterns

Carbohydrate utilisation potential was assessed across the MAGs using frequencies of CAZymes, TonB-Dependent Transporters (TBDTs), sulfatases and peptidases. Peptidases were included for comparison as proteins are another key substrate used by heterotrophic microbes. The largest CAZyme gene repertoires were observed in MAGs affiliated with *Verrucomicrobiae* and *Bacteroidia* that, on average, harboured 15 and 14 CAZymes per Mbp, respectively, whilst the lowest was observed in *Poseidoniia*-affiliated MAGs, 1 per Mbp (Figure 7). *Poseidoniia* (MGIIa)-affiliated MAGs typically harboured high peptidase to CAZyme ratios, 3.7:1 on average, indicating a preference for proteinaceous substrates, in agreement with previous reports [68]. In addition to large CAZyme repertoires, *Verrucomicrobiae* MAGs also encoded the largest number of sulfatase genes, with 24 per Mbp, on average. Despite these broad taxa-related patterns, large variations were evident at the MAG level, with the number of CAZymes in *Bacteroidia* representatives ranging from 5 to 25 per Mbp – the largest being identified in an NS4 representative (S10_BML_bin_54_1) (Figure 7). Differences in genetic repertoires for carbohydrate degradation reflect previous descriptions for these taxa, where *Bacteroidia* typically harbour high numbers of CAZymes and *Verrucomicrobiae* are known to specialise in the degradation of more complex, sulfated polysaccharides [37, 70].

Comparing transcription levels of the focal gene groups across MAGs revealed distinct assemblages of populations that dominated carbohydrate utilisation in each sample (Figure 7). In the above shelf sample S10_SRF, CAZyme, sulfatase and TBDT transcription values were dominated by only a few *Bacteroidia* representatives, particularly *Formosa* (25_SRF_bin_51_1) and NS2b (S10_SRF_bin_49_1). In contrast, the above slope sample S25_SRF was characterised by a larger number and diversity of populations of the *Bacteroidia* and *Gammaproteobacteria* that exhibited comparable transcription values. The key contributors to CAZyme transcription in S25_SRF included two distinct populations affiliated with the NS4 Marine Group (S10_SRF_bin_85_1 and S10_BML_bin_54_1) along with one of *Formosa*

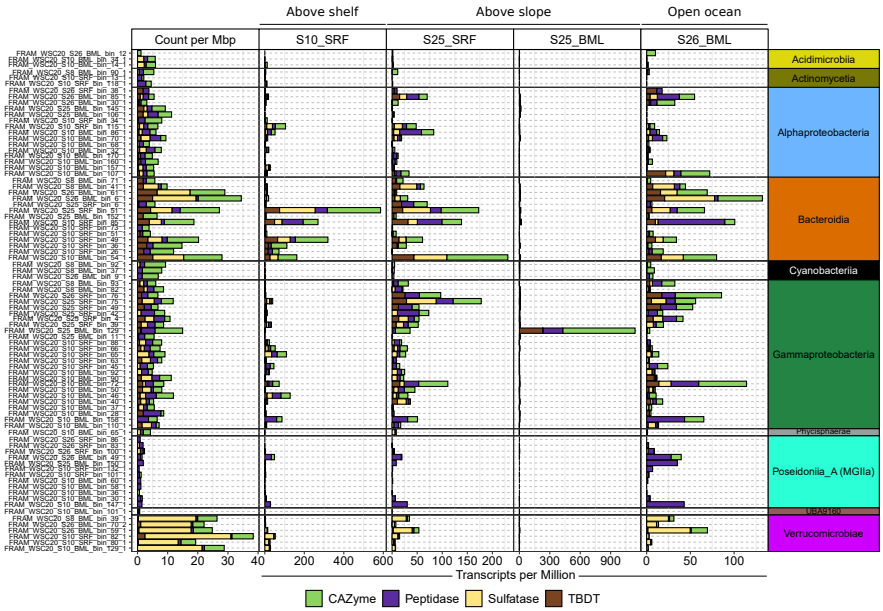

Figure 7. Count and transcription level of carbohydrate utilisation genes for population-representative MAGs. Carbohydrate utilisation genes were clustered into four groups based on function. The count and TPM value of all genes within each group were summed. Gene counts were further normalised by MAG genome size. TBDT = TonB-dependent transporter.

(S25_SRF_bin_51_1), SAR92 (S25_SRF_bin_75_1) and SAR86 (S10_BML_bin_72_1). The highest TBDT transcription levels were also associated with these populations. High sulfatase transcription in S25_SRF was observed in two populations of *Roseibacillus* (S8_BML_bin_39_1 and S26_BML_bin_59_1) that contributed little to transcription in S10_SRF. In sample S26_BML, comparably high CAZyme gene transcription levels were observed across populations assigned to *Bacteroidia, Gammaproteobacteria, Poseidoniia, Alphaproteobacteria* and *Verrucomicrobiae.* In this sample, several distinct populations that contributed little in SRF samples, were amongst the largest contributors to CAZyme transcription, including those assigned to *Cand.* Arcticimaribacter (S26_BML_bin_6_1), *Flavobacteriaceae* (S26_BML_bin_6_1) and *Planktomarina* (S10_BML_bin_107_1). As could be expected, the TPM values for the focal gene groups in S25_BML were strongly dominated by the *Pseudoalteromonas* representative. The MAG analysis revealed highly dynamic population-level patterns in the utilisation of carbohydrates across samples that further illustrates heterogeneity over spatial scales and indicates that the

taxonomic patterns observed at the community-level can be attributed to a varying number of discrete populations.

Substrate-based niche partitioning of populations

Amongst the population representatives that dominated CAZyme gene transcription in each sample, we observed patterns indicative of substrate-based niche partitioning and the utilisation of “communal” as well as “specialist” substrates. We focused only on MAGs that exhibited the highest CAZyme transcription values and further identified within each, those gene families whose transcription was higher than single-copy ribosomal proteins – considered up-transcribed in relation to the genome (Figure 8). From this, several notable patterns emerged. The first was the contribution of multiple MAGs to the transcription of a single gene family in one sample, which provides indications of “communal” substrate usage. For these families, even the combined transcription of the selected MAGs still only captured a small proportion of the community-level gene family transcription. This was particularly true for gene families targeting laminarin (GH16_3, GH17 and GH149) (Figure 8). Laminarin is a major storage compound of diatoms with an estimated annual production of 12 ± 8 gigatons, which can constitute 26 ± 17% of the marine particulate organic carbon pool [71]. The large production and relatively simple structure of laminarin makes it a prime target for heterotrophic bacteria. In agreement with our observations here, the capacity to degrade laminarin was identified as a widespread trait in microbes responding to phytoplankton blooms [10, 14]. In contrast to laminarin, other polysaccharide compounds were only targeted by one of the dominant population representatives. This was particularly evident for the *Bacteroidia*-affiliated populations, which exhibited a higher diversity of transcribed CAZyme gene families that corresponded to a broader diversity of substrates compared to other taxa, which reflects the known carbohydrate-specialist nature of these microbes. Amongst the *Bacteroidia* populations, we observed patterns indicative of substrate-based niche partitioning. Such partitioning was evident either through the exclusive

targeting of a polysaccharide by one population, such as rhamnogalacturonan (GH28 and GH78) by NS5 (S26_BML_bin_6_1), or through changes in the composition of populations making use of the same polysaccharide across samples, which was clear for β-xylan (GH10) and β-mannan (GH130). From this analysis, we could conclude that "communal" substrates were being targeted by multiple populations while "specialist" substrates were being utilised by either single or a unique composition of populations across samples.

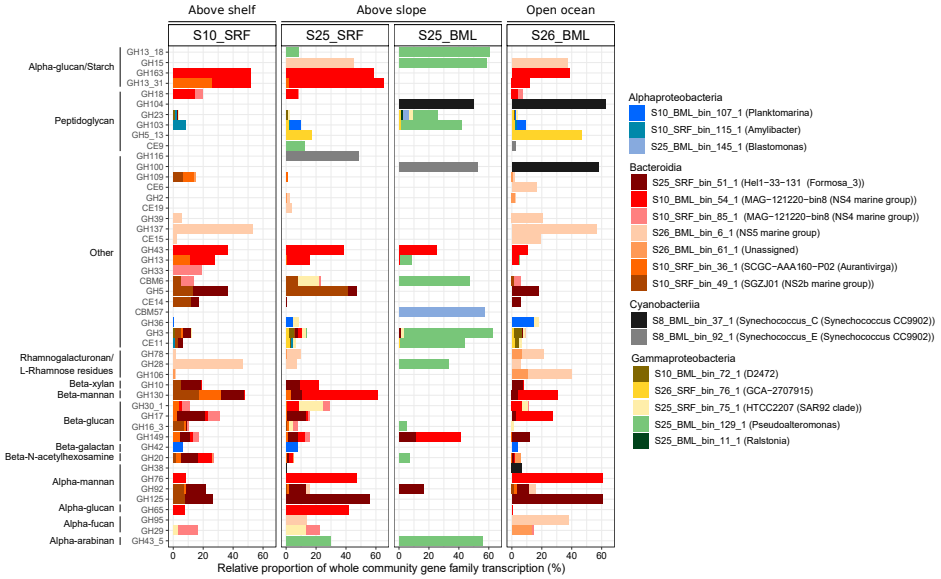

Figure 8. Comparison of CAZyme gene family transcription by selected MAGs across the four metatranscriptome samples. The ten MAGs that exhibited the highest CAZyme transcription values in each sample were selected. Within each of those MAGs, only CAZyme gene families with transcription values higher than ribosomal proteins were retained. To determine the proportion of transcription at the gene family level, we mapped all transcripts identified as CAZymes at the community level to MAG predicted genes. Presented here is the proportion of transcripts of each gene family associated with the selected MAGs – proportion of sample total transcription for that gene family.

Presence and transcription of polysaccharide utilisation loci

We further investigated carbohydrate utilisation by *Bacteroidia* populations through the transcription of their polysaccharide utilisation loci (PULs). PULs are the result of genomic rearrangement through evolution to optimise the transcription of genes required to extracellularly break down and uptake specific polysaccharide molecules. Many of the key carbohydrate degraders of *Bacteroidia* observed in the marine environment, e.g. after spring phytoplankton blooms, typically harbour PULs. Investigating the composition of PULs in genomes has been a

valuable approach previously used to disentangle substrate-based niche partitioning [10, 14, 19]. However, we only identified PULs in seven of the *Bacteroidia* MAGs recovered here (Supplementary Table S14 and S15), an observation that could be influenced by the completeness of the MAGs. Assessing the transcription of these PULs provided little additional insights into polysaccharide utilisation patterns. In cases where a MAG harboured multiple PULs targeting different polysaccharides, such as the *Formosa* representative, they exhibited similar transcription levels across samples. Only in the case of the NS2b representative was a clear variation observed, where the xylan-targeting PUL exhibited a twofold higher transcription than the mannan-targeting PUL in S26_BML. This is an important observation for future studies investigating carbohydrate utilisation by members of *Bacteroidia*, as it indicates that the entire genomic repertoire of CAZymes, sulfatases and TBDTs needs to be considered and not only the information encapsulated within PULs. Furthermore, the lack of transcriptional variability across PULs indicates a low level of regulation and is more indicative of a priming effect, where the detection of one substrate results in transcription of all PULs.

CONCLUSION

During late summer in the high North Atlantic, the carbohydrate fraction of POM and the utilisation of carbohydrates by microbial communities exhibits vertical and lateral spatial heterogeneity. Monosaccharide and polysaccharide analysis of POM showed distinct differences across locations and depths. Typically, higher abundances were observed at locations above the continental slope compared to the open-ocean. Monosaccharide compositions were dominated by glucose, which decreased in proportion with depth, suggesting preferential utilisation in surface waters. The polysaccharide pool of POM was consistently characterised by complex compounds known to accumulate over weeks to months during phytoplankton blooms, such as FCSP, along with a patchy detection and varied abundance of more labile substrates, such as β-1,3-glucan. These observations suggest spatial heterogeneity in recent productivity that reflects previously

described patchiness in phytoplankton dynamics in this region in early summer [46]. Concurrently, metatranscriptomics revealed the presence of active and dynamic populations, with distinct assemblages dominating carbohydrate utilisation in each sample and making use of labile (communal) and more complex (specialist) substrates. Carbohydrate specialists, such as those affiliated with *Formosa, Aurantivirga* and NS4 Marine Group, each exhibited distinct profiles of transcribed CAZyme genes across samples, which indicated substrate-based niche partitioning. These specialists taxonomically resembled, at the genus-level, those known to be primary responders and the key carbohydrate degraders during phytoplankton blooms in temperate ecosystems. In combination, these results highlight that local biological and physical processes shape the carbohydrate pool in the high North Atlantic in late summer and that the key heterotrophic microbial populations, particularly of the *Bacteroidia* and *Gammaproteobacteria*, respond to differences in glycan availability.

DATA AVAILABILITY

The measurements of several abiotic parameters from sensors mounted on the CTD have been published under the PANGAEA accession 943220 [72]. The monosaccharide concentrations have been deposited under the PANGAEA accession 957737. The metagenomic raw reads, assemblies and metagenome-assembled genomes along with the metatranscriptomic raw reads were deposited at European Nucleotide Archive, under the project accession PRJEB58071 (Supplementary Table S16). Detailed information on how to reproduce our results and generate the figures presented in this manuscript is provided at https://github.com/tpriest0/FRAM_STRAIT_WSC20_data_analysis.

AUTHOR CONTRIBUTIONS

TP wrote the manuscript, performed the metagenomic, metatranscriptomic and carbohydrate data analysis. TP and SVM extracted the carbohydrates and SVM subsequently carried out the

microarray analysis. SVM, JHH, BMF and RA contributed to the interpretation of the results and the formulation of the story. BMF, RA and TP planned the work and devised the project. All authors contributed to reviewing and improving the manuscript.

ACKNOWLEDGEMENTS

We would like to thank the captain and crew of the *RV. Maria. S. Merian* for their support throughout all sampling aspects of this project. We thank Alek Bolte for his assistance with HPAEC-PAD and Tina Trautmann for her assistance with the carbohydrate microarray analysis. We thank Bruno Hüttel and the team at the Max Planck Genome Centre in Cologne for their work in generating the metagenomes and metatranscriptomes. We thank A. Murat Eren for his valuable comments and feedback with respect to data analysis.

REFERENCES

1. Benner R, Pakulski JD, Mccarthy M, Hedges JI, Hatcher PG. Bulk chemical characteristics of dissolved organic matter in the ocean. Science. 1992;255:1561–1564.
2. Hofmann J, Hahm HS, Seeberger PH, Pagel K. Identification of carbohydrate anomers using ion mobility–mass spectrometry. Nature. 2015;526:241–244.
3. Myklestad S. Production of carbohydrates by marine planktonic diatoms. I. Comparison of nine different species in culture. J Exp Mar Biol Ecol. 1974;15:261–274.
4. Bellinger B, Abdullahi A, Gretz M, Underwood G. Biofilm polymers: relationship between carbohydrate biopolymers from estuarine mudflats and unialgal cultures of benthic diatoms. Aquat Microb Ecol. 2005;38:169–180.
5. Abdullahi AS, Underwood GJC, Gretz MR. Extracellular matrix assembly in diatoms (Bacillariophyceae). V. Environmental effects on polysaccharide synthesis in the model diatom, Phaeodactylum Tricornutum1. J Phycol. 2006;42:363–378.

6. Wetz MS, Wheeler PA. Release of dissolved organic matter by coastal diatoms. Limnol Oceanogr. 2007;52:798–807.
7. Møller EF. Production of dissolved organic carbon by sloppy feeding in the copepods Acartia tonsa, Centropages typicus, and Temora longicornis. Limnol Oceanogr. 2007;52:79–84.
8. Biddanda B, Benner R. Carbon, nitrogen, and carbohydrate fluxes during the production of particulate and dissolved organic matter by marine phytoplankton. Limnol Oceanogr. 1997;42:506–518.
9. Reintjes G, Arnosti C, Fuchs B, Amann R. Selfish, sharing and scavenging bacteria in the Atlantic Ocean: a biogeographical study of bacterial substrate utilisation. ISME J. 2019;13:1119–1132.
10. Krüger K, Chafee M, Ben Francis T, Glavina del Rio T, 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;13:2800–2816.
11. Arnosti C, Wietz M, Brinkhoff T, Hehemann J-H, Probandt D, Zeugner L, et al. The Biogeochemistry of marine polysaccharides: sources, inventories, and bacterial drivers of the carbohydrate cycle. Annu Rev Mar Sci. 2021;13:81–108.
12. Henrissat B. A classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem J. 1991;280:309–316.
13. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Res. 2009;37:D233–D238.
14. Kappelmann L, Krüger K, Hehemann J-H, Harder J, Markert S, Unfried F, et al. Polysaccharide utilization loci of North Sea Flavobacteriia as basis for using SusC/D-protein expression for predicting major phytoplankton glycans. ISME J. 2019;13:76–91.
15. Hoarfrost A, Arnosti C. Heterotrophic extracellular enzymatic activities in the Atlantic Ocean follow patterns across spatial and depth regimes. Front Mar Sci. 2017;4.

16. D'Ambrosio L, Ziervogel K, MacGregor B, Teske A, Arnosti C. Composition and enzymatic function of particle-associated and free-living bacteria: a coastal/offshore comparison. ISME J. 2014;8:2167–2179.
17. Arnosti C, Steen AD, Ziervogel K, Ghobrial S, Jeffrey WH. Latitudinal gradients in degradation of marine dissolved organic carbon. PLOS ONE. 2011;6:e28900.
18. Teeling H, Fuchs BM, Bennke CM, Krüger K, Chafee M, Kappelmann L, et al. Recurring patterns in bacterioplankton dynamics during coastal spring algae blooms. eLife. 2016;5:e11888.
19. Avcı 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;14:1369–1383.
20. Teeling H, Fuchs BM, Becher D, Klockow C, Gardebrecht A, Bennke CM, et al. Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. Science. 2012;336:608–611.
21. Purser A, Hoge U, Busack M, Hagemann J, Lehmenhecker S, Dauer E, et al. Arctic Seafloor Integrity, Cruise No. MSM95 - (GPF 19-2_05), 09.09.2020 - 07.10.2020, Emden (Germany) - Emden (GERMANY). 2020.
22. GEBCO Compilation Group. GEBCO 2020 Grid. 2020.
23. Jakobsson M, Mayer LA, Bringensparr C, Castro CF, Mohammad R, Johnson P, et al. The international bathymetric chart of the Arctic Ocean version 4.0. Sci Data. 2020;7:176.
24. 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;12:1150.
25. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015;12:59–60.
26. Rho M, Tang H, Ye Y. FragGeneScan: predicting genes in short and error-prone reads. Nucleic Acids Res. 2010;38:e191.

27. Shen W, Ren H. TaxonKit: A practical and efficient NCBI taxonomy toolkit. J Genet Genomics. 2021;48:844–850.
28. Priest T, Orellana LH, Huettel B, Fuchs BM, Amann R. Microbial metagenome-assembled genomes of the Fram Strait from short and long read sequencing platforms. PeerJ. 2021;9:e11721.
29. Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, et al. A new view of the tree of life. Nat Microbiol. 2016;1:1–6.
30. Olm MR, Crits-Christoph A, Diamond S, Lavy A, Matheus Carnevali PB, Banfield JF. Consistent metagenome-derived metrics verify and delineate bacterial species boundaries. mSystems. 2020;5:e00731-19.
31. Kolmogorov M, Bickhart DM, Behsaz B, Gurevich A, Rayko M, Shin SB, et al. metaFlye: scalable long-read metagenome assembly using repeat graphs. Nat Methods. 2020;17:1103–1110.
32. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ. 2019;7:e7359.
33. 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.
34. 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;25:1043–1055.
35. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics. 2020;36:1925–1927.
36. Seeman T. barrnap 0.9: rapid ribosomal RNA prediction.

37. Orellana LH, Francis TB, Ferraro M, Hehemann J-H, Fuchs BM, Amann RI. Verrucomicrobiota are specialist consumers of sulfated methyl pentoses during diatom blooms. ISME J. 2021;1–12.

38. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26:841–842.

39. Rodriguez-R LM, Konstantinidis KT. The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. PeerJ Prepr. 2016;4:e1900v1.

40. Bushnell B. BBTools software package. http://bbtools.jgi.doe.gov. .

41. Kopylova E, Noé L, Touzet H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics. 2012;28:3211–3217.

42. Wagner GP, Kin K, Lynch VJ. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. Theory Biosci. 2012;131:281–285.

43. Nilsen F, Cottier F, Skogseth R, Mattsson S. Fjord–shelf exchanges controlled by ice and brine production: The interannual variation of Atlantic Water in Isfjorden, Svalbard. Cont Shelf Res. 2008;28:1838–1853.

44. Richter ME, von Appen W-J, Wekerle C. Does the East Greenland Current exist in the northern Fram Strait? Ocean Sci. 2018;14:1147–1165.

45. Skoog A, Benner R. Aldoses in various size fractions of marine organic matter: Implications for carbon cycling. Limnol Oceanogr. 1997;42:1803–1813.

46. von Jackowski A, Grosse J, Nöthig E-M, Engel A. Dynamics of organic matter and bacterial activity in the Fram Strait during summer and autumn. Philos Trans R Soc Math Phys Eng Sci. 2020;378:20190366.

47. Tanoue E, Handa N. Monosaccharide composition of marine particles and sediments from the Bering Sea and northern North Pacific. Oceanol Acta. 1987;10:91–99.

48. Liebezeit G, Bölter M. Water-extractable carbohydrates in particulate matter of the Bransfield Strait. Mar Chem. 1991;35:389–398.

49. Compiano A-M, Romano J-C, Garabetian F, Laborde P, de la Giraudièrea I. Monosaccharide composition of particulate hydrolysable sugar fraction in surface microlayers from brackish and marine waters. Mar Chem. 1993;42:237–251.

50. Ittekkot V, Brockmann U, Michaelis W, Degens E. Dissolved Free and Combined Carbohydrates During a Phytoplankton Bloom in the Northern North Sea. Mar Ecol Prog Ser. 1981;4:299–305.

51. Øbro J, Sørensen I, Moller I, Skjøt M, Mikkelsen JD, Willats WGT. High-throughput microarray analysis of pectic polymers by enzymatic epitope deletion. Carbohydr Polym. 2007;70:77–81.

52. Deniaud-Bouët E, Hardouin K, Potin P, Kloareg B, Hervé C. A review about brown algal cell walls and fucose-containing sulfated polysaccharides: Cell wall context, biomedical properties and key research challenges. Carbohydr Polym. 2017;175:395–408.

53. Ebringerová A, Heinze T. Xylan and xylan derivatives – biopolymers with valuable properties, 1. Naturally occurring xylans structures, isolation procedures and properties. Macromol Rapid Commun. 2000;21:542–556.

54. Huang G, 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;66:3768–3782.

55. Vidal-Melgosa S, Lagator M, Sichert A, Priest T, Pätzold J, Hehemann J-H. Not digested: algal glycans move carbon dioxide into the deep-sea. 2022. bioRxiv. , 2022.03.04.483023

56. Trudnowska E, Gluchowska M, Beszczynska-Möller A, Blachowiak-Samolyk K, Kwasniewski S. Plankton patchiness in the Polar Front region of the West Spitsbergen Shelf. Mar Ecol Prog Ser. 2016;560:1–18.

57. Granum E, Kirkvold S, Myklestad SM. Cellular and extracellular production of carbohydrates and amino acids by the marine diatom Skeletonema costatum: diel variations and effects of N depletion. Mar Ecol Prog Ser. 2002;242:83–94.

58. Boysen AK, Carlson LT, Durham BP, Groussman RD, Aylward FO, Ribalet F, et al. Particulate metabolites and transcripts reflect diel oscillations of microbial activity in the surface ocean. mSystems. 2021;6:e00896-20.

59. Salazar G, Paoli L, Alberti A, Huerta-Cepas J, Ruscheweyh H-J, Cuenca M, et al. Gene expression changes and community turnover differentially shape the global ocean metatranscriptome. Cell. 2019;179:1068-1083.e21.

60. Rinta-Kanto JM, Sun S, Sharma S, Kiene RP, Moran MA. Bacterial community transcription patterns during a marine phytoplankton bloom. Environ Microbiol. 2012;14:228–239.

61. Pontiller B, Martínez-García S, Joglar V, Amnebrink D, Pérez-Martínez C, González JM, et al. Rapid bacterioplankton transcription cascades regulate organic matter utilization during phytoplankton bloom progression in a coastal upwelling system. ISME J. 2022;1–13.

62. Holmström C, Kjelleberg S. Marine Pseudoalteromonas species are associated with higher organisms and produce biologically active extracellular agents. FEMS Microbiol Ecol. 1999;30:285–293.

63. Moisander PH, Sexton AD, Daley MC. Stable associations masked by temporal variability in the marine copepod microbiome. PLOS ONE. 2015;10:e0138967.

64. Shoemaker KM, Moisander PH. Microbial diversity associated with copepods in the North Atlantic subtropical gyre. FEMS Microbiol Ecol. 2015;91:fiv064.

65. Wietz M, Bienhold C, Metfies K, Torres-Valdés S, von Appen W-J, Salter I, et al. The polar night shift: seasonal dynamics and drivers of Arctic Ocean microbiomes revealed by autonomous sampling. ISME Commun. 2021;1:1–12.

66. Fadeev E, Salter I, Schourup-Kristensen V, Nöthig E-M, Metfies K, Engel A, et al. Microbial communities in the east and west Fram Strait during sea ice melting season. Front Mar Sci. 2018;5.

67. von Jackowski A, Becker KW, Wietz M, Bienhold C, Zäncker B, Nöthig E-M, et al. Variations of microbial communities and substrate regimes in the eastern Fram Strait between summer and fall. Environ Microbiol. 2022;24:4124–4136.

68. Orellana LH, Ben Francis T, Krüger K, Teeling H, Müller M-C, Fuchs BM, et al. Niche differentiation among annually recurrent coastal Marine Group II Euryarchaeota. ISME J. 2019;13:3024–3036.

69. Francis B, Urich T, Mikolasch A, Teeling H, Amann R. North Sea spring bloom-associated Gammaproteobacteria fill diverse heterotrophic niches. Environ Microbiome. 2021;16:15.

70. 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;5:1026–1039.

71. 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. 2020;117:6599–6607.

72. Priest T, Merten V, Purser A. Physical oceanography during RV MARIA S. MERIAN cruise MSM95. 2022. PANGAEA.