MASTERARBEIT | MASTER'S THESIS

Titel | Title

Characterization of bacterial communities living in marine snow and their proteome

verfasst von | submitted by

Jan Eulering

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien | Vienna, 2025

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 833
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Ecology and Ecosystems
Betreut von | Supervisor:	Univ.-Prof. i.R. Dr. Gerhard Herndl

Acknowledgements

First, I want to say thank my supervisor Prof. Gerhard J. Herndl for guiding me through my thesis project, for improving my scientific writing skills and for always being patient and kind with me.

I am also very grateful for all the help and feedback from the members of the Microbial Oceanography Group with my research project. Special thanks to Maria Papadatou for showing me how to work properly and efficiently in the laboratory and also to Zihao Zhao, who introduced me to the world of bioinformatics, multi-omics techniques and how to work with high-performance computer clusters. As a bachelor's student, I would never have thought that I would be able to do something like this.

I am very thankful for my parents' help in every possible way, even though they were 1000 kilometers away. And finally, I want to thank my girlfriend, who I left behind in Germany. Thank you for your patience and support during these tough times!

Abstract

Marine snow acts as a key component of the biological pump, which are the biogeochemical processes that transfer carbon from surface waters into the ocean's interior, where it is partially remineralized and partially stored over long time periods. These remineralization processes are carried out mostly by a variety of heterotrophic bacteria associated with unique metabolic capabilities and ecological niches. Therefore, it is important to get detailed knowledge about what these key players are metabolizing on marine snow particles and in the ambient water and how these processes influence the fate of marine snow particles, the biological pump and therefore, influence the food availability for a variety of larger marine metazoan species and ultimately, the climate. As this field has become more important during recent years, a variety of new research methods became available in the last decades to determine the reactivity of organic matter in the ocean water column and the role of microbes transforming this organic matter as it sinks through the water column.

To test the reliability of these new methods and to get a deeper understanding of bacterial metabolic pathways and the community composition on marine snow and ambient water ecosystems, a multi-omics approach, consisting of metagenomics and proteomics was chosen in this study. Additionally, we used the relatively understudied method of ^{15}N stable isotope labeling to trace patterns of nitrogen assimilation.

Using laboratory experiment and artificial marine snow, we found that marine snow particles and the ambient water represented different niches, each with a unique microbial community composition and significant differences in metabolic activities, signaling processes and nutrient acquisition processes. All samples were dominated in gene abundance by the diverse phylum of *Pseudomonadota,* followed by *Bacteroidetes* which seemed to prefer a particle-associated lifestyle. In particle associated samples general nutrient acquisition pathways were dominant, while free-living bacteria in ambient water samples had more active genes associated with competition and stress responses.

We found that multi-omics approaches are superior to single-omics methods, because genetic potentials and protein expression patterns do not necessarily correlate. We also found a large fraction of proteins with unknown functions.

Zusammenfassung

Meeresschnee fungiert als Schlüsselkomponente der biologischen Pumpe, das heisst der biogeochemischen Prozesse, die Kohlenstoff aus dem Oberflächenwasser ins Innere des Ozeans transportieren, wo er teilweise remineralisiert und teilweise über lange Zeiträume gespeichert wird. Diese Remineralisierungsprozesse werden größtenteils von heterotrophen Bakterien durchgeführt, die jeweils einzigartige Stoffwechselfähigkeiten und ökologischen Nischen besitzen. Daher ist es wichtig, detaillierte Kenntnisse darüber zu erlangen, was diese Hauptakteure auf Meeresschneepartikeln und im Umgebungswasser metabolisch tun und wie diese Prozesse das Schicksal der Meeresschneepartikel, die biologische Pumpe und in Folge auch die Nahrungsverfügbarkeit für eine Vielzahl größerer mehrzelliger Meeresbewohner und auch das Klima beeinflussen. Da dieser Bereich in den letzten Jahren immer mehr an Bedeutung gewonnen hat, ist in den letzten Jahrzehnten eine Vielzahl neuer Forschungsmethoden verfügbar geworden, die ermöglichen chemische Prozesse und die Rolle von Mikroben, die dieses organische Material umsetzen, während es in der Wassersäule sinkt, zu erforschen.

Um die Zuverlässigkeit dieser neuen Methoden zu testen und ein tieferes Verständnis der bakteriellen Stoffwechselwege und der Zusammensetzung der Lebensgemeinschaften in Meeresschnee- und Umgebungswasser-Ökosystemen zu erlangen, wurde in dieser Studie ein Multi-omics-Ansatz, bestehend aus Metagenomics und Proteomics, gewählt. Zusätzlich verwendeten wir die relativ wenig untersuchte Methode der ^{15}N-Stabilisotopenmarkierung, um die Muster der Stickstoffassimilation zu verfolgen.

Wir fanden mit einem Experiment mit im Labor erzeugten Meeresschnee heraus, dass marine Schneepartikel und das Umgebungswasser unterschiedliche Nischen darstellen, die jeweils eine einzigartige bakterielle Gemeinschaft und signifikante Unterschiede in Bezug auf Stoffwechselaktivitäten, Signalprozesse und Prozesse der Nährstoffgewinnung aufweisen. In allen Proben dominierte das vielfältige Phylum der *Pseudomonadota*, gefolgt von *Bacteroidetes*, die eine partikelassoziierte Lebensweise zu bevorzugen schienen. In Partikel-assoziierten Proben schien der Schwerpunkt auf Nährstoffaufnahmepfaden zu liegen, während freilebende Bakterien in Umgebungswasserproben aktive Gene aufwiesen, die mit Konkurrenz und Stressreaktionen in Verbindung stehen.

Wir haben festgestellt, dass Multi-omics den Single-omics-Methoden überlegen sind, da genetische Potenziale und Proteinexpressionsmuster nicht unbedingt korrelieren. Wir fanden einen großen Anteil von Proteinen mit unbekannter Funktion.

Table of contents

1. Introduction

Detritus from the ocean's surface that sinks to the deep ocean is referred to as marine snow, which is essential for the marine carbon and nutrient cycling (Simon et al., 1990). Furthermore, marine snow is an essential component of the biological pump (de la Rocha, 2006). Marine snow originates from various sources, including dead plankton cells, fecal pellets, mucus webs from zooplankton, and other organic matter. These materials aggregate to form marine snow, which can range in size from at least 500µm to several centimeters (Turner, 2015).

Understanding marine snow cycling is important, particularly in the context of climate change, as it helps in transferring carbon from the atmosphere to the ocean’s interior (Ploug et al., 1999). Microbial communities on marine snow and in the surrounding water are key players in this process (Simon et al., 1990). Climate affects the intensity of the biological pump, which, in turn, influences the food availability for larger organisms and impacts benthic species (Alldredge et al., 1988; Zhao et al., 2024c).

There are a number of studies focusing on the abundance and activity of microbial communities on marine snow (Thiele et al., 2015; Boeuf et al., 2019; Zhao et al., 2024a). Heterotrophic microbes prefer particulate organic matter as their main carbon source rather than dissolved organic matter, pointing to the importance of studying how marine snow is utilized by bacteria (Baltar et al.; Zhao et al., 2024b; Thiele et al., 2015).

Multi-omics approaches can provide valuable insights into the community composition and ecosystem processes, identifying how different taxa utilize marine snow (Haas et al., 2017; Kieft et al., 2021). Combining metagenomics and -proteomics offers a better understanding of microbial activities, as actual protein expression patterns are often decoupled from the genetic potential (Bergauer et al., 2018; Zhao et al., 2024a; Zhao et al., 2024c). This study aims to explore metabolic pathways and ecological niches of bacterial communities linked to marine snow and ambient water while investigating nitrogen assimilation patterns using ^{15}N-labeled nitrogen as a tracer (Fogel et al., 1993; Tinta et al., 2023). This approach has the potential to reveal new insights into ecological processes, but is relatively understudied (Taubert et al., 2013). A variety of bioinformatic tools were used to analyze differences in underlying processes of the individual samples and explore possible niche differentiation of heterotrophic microbes.

2. Materials and Methods

2.1 Biological sample preparation

To produce a sufficient amount of marine snow, incubation of *Tysochrysis intea* was initiated in three bioreactors with 1L volume each. Guillard's medium was used, however, with ^{15}N labeled $NaNO_3$. The bioreactors were connected to a hose which supplied them with filtered air for CO_2 supply. The algae grew under artificial light for 16:8 h light: dark cycle.

The exponential growth rate of *T. intea* ended after about seven days. Then, the whole content of the three bioreactors was transferred to a sterile 5L Schott bottle. The whole 3L volume was autoclaved at 120°C for 30min. Then, *T. intea* was transferred to three 1L bottles to allow formation of marine snow.

After 4 days, the dead cells aggregated into visible particles. After removing 400mL of excess water, water containing the particles was inoculated with a natural microbial community collected from the surface waters of the Adriatic Sea off Rovinj, Croatia. Surface water temperature was 19°C when the sampling took place. This water was stored in the home laboratory at 21°C in the dark for about one month prior to being used as an inoculum. The bottles were then put on a rolling table. Again, after four days the small particles from the beginning formed sphere-shaped marine snow of about 0.5 to 1 cm in diameter. Sampling of marine snow was performed by removing excess water until a sterile 1.5mL Eppendorf tube was filled. The tube was centrifuged three times for a few seconds to form a pellet.

The remaining seawater (0.5 L) from the Schott bottles was filtered first through a 1.2µm pore size polycarbonate filter to collect larger bacteria and small particles (Sample A12). From this 1.2 µm filtrate, 0.5L was subsequently filtered through a 0.2µm pore size filter to collect free-living bacterial cells (Sample A02). The collected samples were stored at -70°C until further processing.

2.2 Protein and DNA extraction

The protein extraction process was done on 6 different days. On the first day two buffers were prepared. Fifty mL Buffer A was made of 77.125mg DTT (1,4-Dithiothreitol), 5ml Sodium Dodecyl-sulfate (SDS) 20% in 45mL autoclaved Milli-Q water. Buffer B (50mL) consisted of 77.125mg DTT, 2g CHAPS, 7.612g Thiourea, 21g Urea, one Protease inhibitor cocktail tablet, 1mL EDTA, 2.5mL Tris–HCl, 2.5mL Triethylammonium bicarbonate (TEAB) buffer, mixed in 44mL autoclaved Milli-Q water.

Samples were collected from the freezer and placed on ice to prevent destruction of proteins by enzymes. Filter samples were cut into pieces and transferred into Falcon tubes. Buffer A was added to fully cover all the filter pieces in the tube, then incubated at 90°C for 10 min, followed by 30 min sonication. The buffers were added in the same way to the pellet sample. The supernatant liquid (fraction A) was transferred into a new 15mL Falcon tube and stored at 4°C overnight. The remaining parts of the samples (fraction B) were then covered with Buffer B and incubated on a shaker at 2g overnight.

On the next day, samples of fraction B were vortexed horizontally for 1:15h, followed by a few seconds of short centrifugation to spin down the filter pieces that were attached to tube walls. After that, the samples of fraction B were sonicated for 30 min, while samples of fraction A were incubated at ~55°C for 10 min. The two fractions A+B were combined by pipetting the liquid fraction A into the tube of B. The combined fractions A+B of each sample were then transferred to a Pierce protein concentrator column and centrifuged at 45°C, 3220x *g* at room temperature. This process was repeated until the volume in the column reached approximately 500µL in each sample. The concentrated samples were then transferred from the column into a sterile 2mL Eppendorf tube and stored at -70°C.

Starting on day 3, to the samples -20°C cold acetone was added (three times the volume of the sample) resulting in 1.5mL volume. Samples were briefly vortexed and then stored overnight.

On the following day, fresh 50mM TEAB buffer was prepared (200µL/sample). Samples were collected from the -20°C freezer and centrifuged for 30min at 4°C, 13000x *g*. Afterwards the supernatant acetone was removed. The tubes were left open for around 5 min to let the remaining acetone evaporate. Then 200µL of TEAB-buffer were added into the sample and resuspended. To these samples we added a 9:1 ratio of ice-cold ethanol (200µL sample – 1800µL ethanol). It was tried to suspend the pellet with the vortex, but the pellets were not fully dissolving at this point. Samples were then stored at -20°C until further processing.

In the following step, samples were collected from the freezer and centrifuged at 4°C for 30 min with 13000x *g*. Supernatant ethanol was then removed with a pipette, leaving the lids open for the remaining ethanol to evaporate. Again, samples were stored at -20°C overnight.

On the next working day, TEAB buffer and fresh 1M alkylation buffer were prepared. TEAB buffer was prepared like in the steps before. For 3mL of the alkylation buffer 0.56g Iodoacetamide (IAA) was dissolved in 3mL of the TEAB buffer. The alkylation buffer is light sensitive; therefore, it was stored in the dark. TEAB buffer (110µL) was added to each sample. The samples were resuspended by vortexing and sonicating for a few seconds. Two µL of reduction buffer was added into each sample. The reduction buffer was ready to use. Samples were then incubated at 60°C for 30 min in the heat block, followed by a 5 min cool down to room temperature. In the dark, 2µL of alkylation buffer was added to each sample, followed by 30 min incubation in the dark. Then, a trypsin solution was prepared using 1mL TEAB buffer per trypsin powder vial (Merck – Sigma-Aldrich) with a 1mL sterile syringe. Then 50 µL of trypsin was added to each sample and then incubated at 37°C, in the heat block for 16 h.

For the last step, three solutions were prepared. TFA-ACN solution, containing 0.5% trifluoroacetic acid (TFA) in 5% acetonitrile (CAN) and 94.5% autoclaved MilliQ-water was prepared. The activation solution contained 50% methanol and 50% autoclaved MilliQ-water. The elution buffer was 70% ACN with 0.1% formic acid in 29.9% MilliQ-water. The first step was performed 16 h after trypsin addition. To stop the reaction, 1.5µL of 100% TFA was added to each sample. After a few min, samples were centrifuged at room temperature for 1 min at 10000x *g*. The supernatant liquid was transferred to a new sterile Eppendorf tube and the pellets kept at -20°C. The sample tubes were put into the SpeedVac concentrator with open lid at room temperature, V-AQ mode for 15 min per run. It took 4 cycles until the pellets were in an acceptable jelly-like condition for the next working step. For the purification of peptides in the next step, the Pierce™ C18 Spin Columns (Thermo Scientific, USA, product number 89873) were used. It is important to note that a TFA-ACN solution was used as substitute for sample buffer, equilibration solution and wash solution in this protocol. After purification, samples were dissolved in 40µL elution buffer per sample. The samples were then put into the SpeedVac concentrator with lid open again, this time at 45°C, V-AQ mode for 20 min. As the samples were not fully dry at this point, the last step was repeated for another 10 min. Samples were then stored at -70°C. Final processed samples were submitted for mass spectrometry analysis at the Department of Functional and Evolutionary Ecology in the University of Vienna using an Orbitrap-MS as described in Zhao et al. (2024b).

DNA extraction was done using the DNeasy® PowerWater® Kit by QIAGEN (Qiagen, Hilden, Germany) according to the manufacturer's instructions.

For an analysis of DNA quality and quantity, samples were tested in a NanoDrop (Thermo Scientific, USA) with the elusion buffer used as blank, as everything was dissolved in this buffer. As a measure of quality, the DNA concentration was measured for each sample with a required conc. of >10ng/µL needed for further analysis. As a quality measurement the quotient of wavelengths (260nm/280nm) was determined with an optimal value between 1 and 2.

2.3 Bioinformatic data preparation and processing

The metagenomic raw data consisted of four files per sample, reads from two lanes with forward and reverse reads. Data preparation was done using a workflow with software packages as described in Zhao et al. (2024a) and depicted in Figure 1. The computational power was provided by the Life Science Compute Cluster (LiSC) of the University of Vienna. First, the metagenomics raw data were assembled with Megahit into contigs, overlapping reads, which are then used to reconstruct genes by mapping them to a reference genome and predicting genes. This process was subsequently done with BBMap and Prodigal software packages. CD-Hit was used to remove redundancies and contaminations by clustering genes with a similarity of 90% into groups. Abundance data in form of normalized read counts were calculated with Burrows-Wheeler-Alignment (BWA) algorithm. The processed data sets were searched against the UniProt peptidase database for identification of peptidases and against the UniProt non-redundant taxonomy database for identification of bacterial phyla. HmmPress was used to search against the dbCAN-database for Cazymes and kofamscan and EggNog Mapper were used for functional annotations of pathways by searching against their respective reference databases. This resulted in genes that were each associated with a KO (KeGG (Kyoto Encyclopedia of Genes and Genomes) Orthology) number, which allowed identification of their involvement in major metabolic pathways by searching these numbers against the KO database.

The data were quantified by reads per kilobase per million (RPKM) method. Hereby, the read count for each gene was normalized by dividing the mapped reads by the gene's length (in kilobases) to account for the differences in length between individual genes.

$$RPKM = \frac{number\ of\ reads\ mapped\ to\ gene\ sequence}{(gene\ length\ \div 1000) \times (total\ number\ of\ mapped\ reads \div 1.000.000)}$$

This corrected for longer genes having higher read counts due to size. Next, the result was divided by the total number of reads (in millions) mapped to the entire sample. This normalized for differences in sequencing depth, ensuring a sample with more reads did not show higher gene expression just due to more sequencing. Before statistical testing, the resulting values were normalized to approach a gaussian normal distribution by using log-transformation, if not stated otherwise. Information on the function of genes/proteins or pathways was derived from the annotations of EggNog-Mapper (evolutionary genealogy of genes: Non-supervised Orthologous Groups), the KEGG Database or by UniProt Blast search.

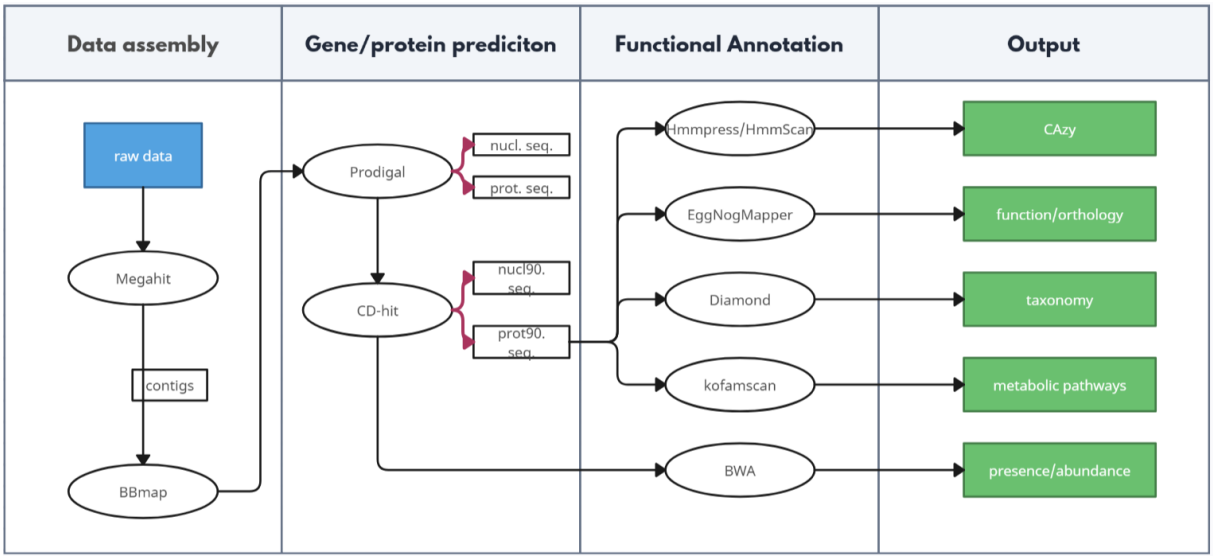

Figure 1: Workflow of metagenomics raw data processing with software packages and algorithms used in each step

The three samples with extracted proteins were processed in the mass-spectrometry facilities of the University of Vienna using an Orbitrap-MS, resulting in one (*.raw)-file per sample. Further processing was done using Proteome Discover (Thermo Scientific, USA). This tool searched the information obtained from the (*.raw)-files against a reference peptide sequence database, in this case the FASTA-databases built in the metagenomic processing steps. Proteins were quantified in relation to the total protein amount and the ^{15}N labeling was determined by software by comparing peptide mass shifts to the expected mass of unlabeled peptides. This quantification method was implemented manually under the assumption that labeled amino acids were heavier than regular ^{14}N containing amino acids. As the mass shift from one atom of ^{14}N to ^{15}N is defined as approximately 0.997u, labeled amino acids could be identified by mass shifts of multiples of 0.997u, depending on the number of nitrogen atoms in the respective amino acid (Naudé, 1929).

3. Results

3.1 Taxonomy

A total of 3.89 x 10^{10} genes of assembled sequencing data were matched with their respective phyla in the marine snow sample (sample P). For the ambient water sample filtered onto 1.2µm (A12) a total of 2.72 x 10^{10} genes were found and for the sample filtered onto 0.2µm filters (sample A02) 1.54 x 10^{10} genes were identified. The results of the taxonomic analysis across three sample types, marine snow (P), ambient water filtered onto 0.2 µm pore size (A02), and ambient water filtered onto 1.2 µm pore size (A12) filters revealed significant differences (Kruskal-Wallis p < 0.00001) in the abundances of bacterial phyla among samples. The Mann-Whitney pairwise tests revealed significant differences of absolute count numbers among individual samples P and A12 with Bonferroni corrected p < 0.00001, and P and A02 with p=0.0005. The difference between the two ambient water samples was not significant with p=0.36. These differences are illustrated in Figure 2, where total abundances of the three most abundant phyla *Pseudomonadota, Bacteroidota* and *Planctomycetota* were visualized.

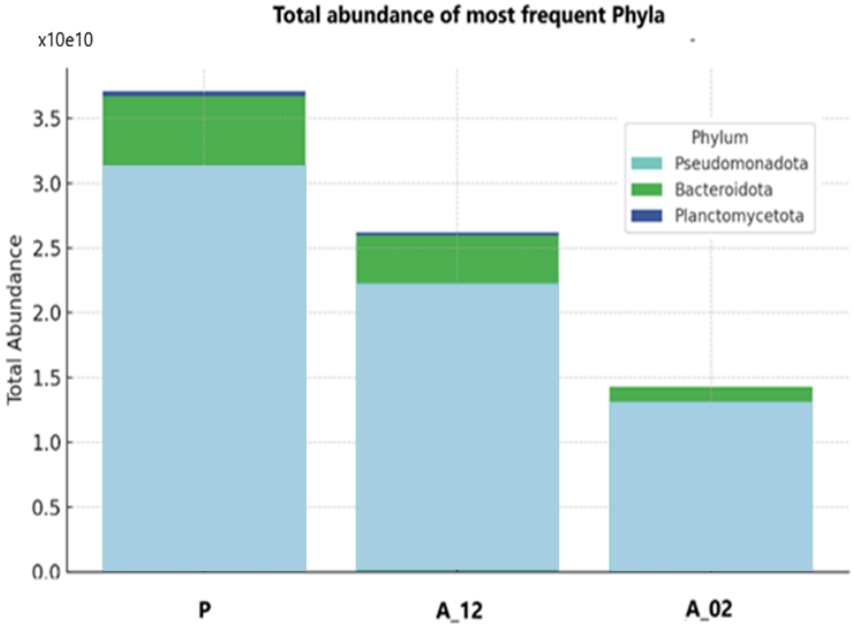

Figure 2: Total abundance comparison of the three most common phyla for the three treatments P, A12 and A02. Pseudomonadota and Bacteroidota are by far the most common phyla in all samples

As shown in Figure 2 and 3, *Pseudomonadota* was by far the most dominant phylum in absolute and relative numbers across all samples, with the particle sample showing the highest abundance (313.96 billion), higher than both ambient water samples, A02 (223.41 billion) and A12 (131.27 billion). The difference between particles and the free-living fraction in A02 represented an increase of 139%, while the difference between P and A12 was approximately 40.5% and A12 and A02 differed by 70.1%. As seen in Figure 3, the dominance of *Pseudomonadota* phylum was also reflected by the high relative gene counts of around 80.6% for P, 82% for A12 and 84,7% for A02.

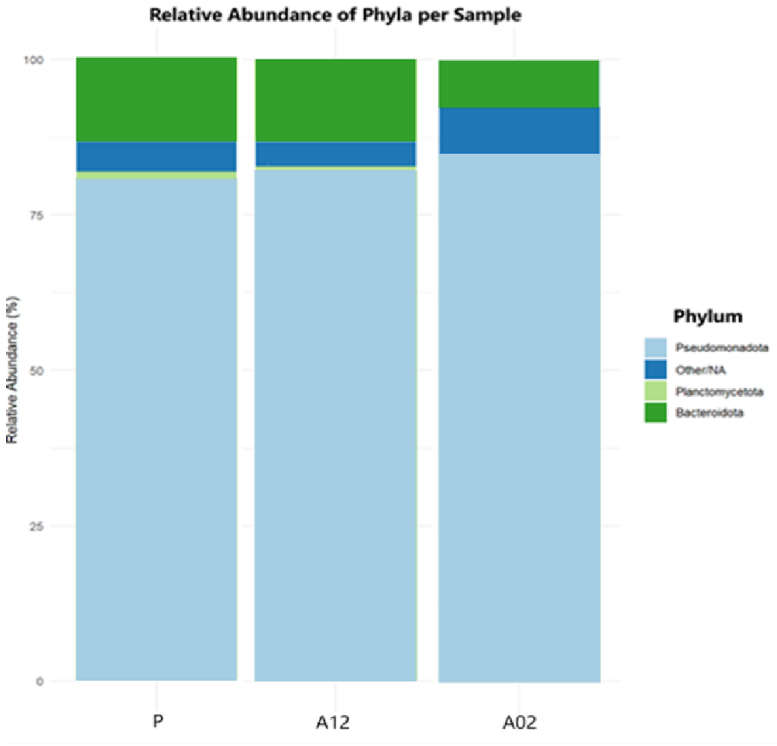

Figure 3: Relative abundances of the three most common and aggregated remaining phyla per sample; dominance of Pseudomonadota and relative difference in Bacteroidota gene counts between particle associated and free-living sample is clearly visible

The absolute abundance of *Bacteroidota genes* was also high but far lower than the *Pseudomonadota* gene count, with P at 52.74 billion, A12 at 36.18 billion, and A02 at 11.38 billion. The increase from the free-living fraction in the ambient water A02 to the particle sample P was 363.6%, while the intermediate A12 showed a 217.8% increase compared to A02 as shown in Figure 2. The relative number of *Bacteroidota* genes was notably decreasing from particles containing samples to the free-living community in A02. The fraction in the intermediate sample was highest with 13.5%, followed by the particle associated sample P with 13.2%. In A02 the relative amount of *Bacteroidetes* genes was only 7.3%.

The *Planctomycetota* gene abundance was highest in P (4.2 billion), with much lower counts in A12 (2.22 billion) and A02 (103.3 million). This phylum displayed a marked reduction from the particle-associated sample to the filtered water samples, with decreases of 97.5% and 47.1% from P to A02 and A12, respectively.

In P, *Thermodesulfobacteriota* gene counts were 147.99 million, compared to 118.75 million in A12 and 82.63 million in A02. The relative comparison showed a 78.7% higher gene abundance

in P compared to A02, with A12 showing an increase of 43.7% increase to the ambient water sample A02. The phylum with relatively low count differences between each sample was *Cyanobacteriota*, with 183.89 million in P, 133.2 million in A12, and 74.09 million in A02. While the abundance was highest in the particle sample, the differences to A02 and A12 were relatively small compared to other phyla. It was also notable that some phyla were present in one fraction but not in the other fractions. In sample P there were 21 phyla, which were exclusively found in this sample. The results of the taxonomic analysis showed significant differences in absolute bacterial phylum gene abundance among each of the three sample fractions, yet the overall pattern of relative dominance and presence appeared to be relatively consistent, as seen in figure 3, where the constant dominance of *Pseudomonadota* genes is visible. The fraction of phyla that occurred in low abundances was almost twice as high in A02 compared to the other samples. When comparing relative count numbers, it is clearly visible that *Bacteroidota* preferred particle-associated lifestyles, while *Pseudomonadota* were not associated with a particular sample, as the relative gene count for this phylum was reduced by almost 50% from sample A12 to A02, whereas the difference between P and A12 was quite small. Generally, bacterial gene occurrence decreased to a high degree with each filtration step.

Diversity indices were calculated using Past Software with the relative frequencies to account for differences in total count numbers. The alpha-diversity within each sample was relatively low, with Shannon indices of $\alpha = 0.68$ for sample P, $\alpha = 0.59$ for A02 and $\alpha = 0.63$ for A12. The total number of taxa present across all samples was $\gamma = 105$. Therefore, Whittaker's beta-diversity was calculated as $\beta = 164.96$.

This high beta-diversity indicated strong differences in taxonomic composition between the samples and suggested high species turnover or variation among the samples.

3.2 Carbohydrate active enzymes (CAZymes)

The gene count numbers of CAZyme genes demonstrated the metabolic diversity of the samples. The particle sample P displayed the highest overall CAZymes gene abundance, while total gene count numbers decreased with each filtration step. The findings generally suggested functional specialization in degrading different kinds of plant matter as shown by glycoside hydrolase and glycosyl transferase families. As fitting the measured values into a normal distribution was not achievable by data transformation, Kruskal-Wallis and Mann-

Whitney pairwise tests with Bonferroni correction were used to test for significant differences. The overall difference among samples was highly significant ($p < 0.00001$), individual differences were also significant with relatively high p-values for the comparison of P and A12 ($p = 0.0282$) and the difference of A02 and A12 ($p = 0.01634$). The difference between P and A02 was highly significant ($p < 0.00001$). A total of 13,955 CAZyme genes were found for P, with decreasing numbers for A12 (10,144) and for A02 (6,884). The total most frequent families and their abundances are shown in Figure 4.

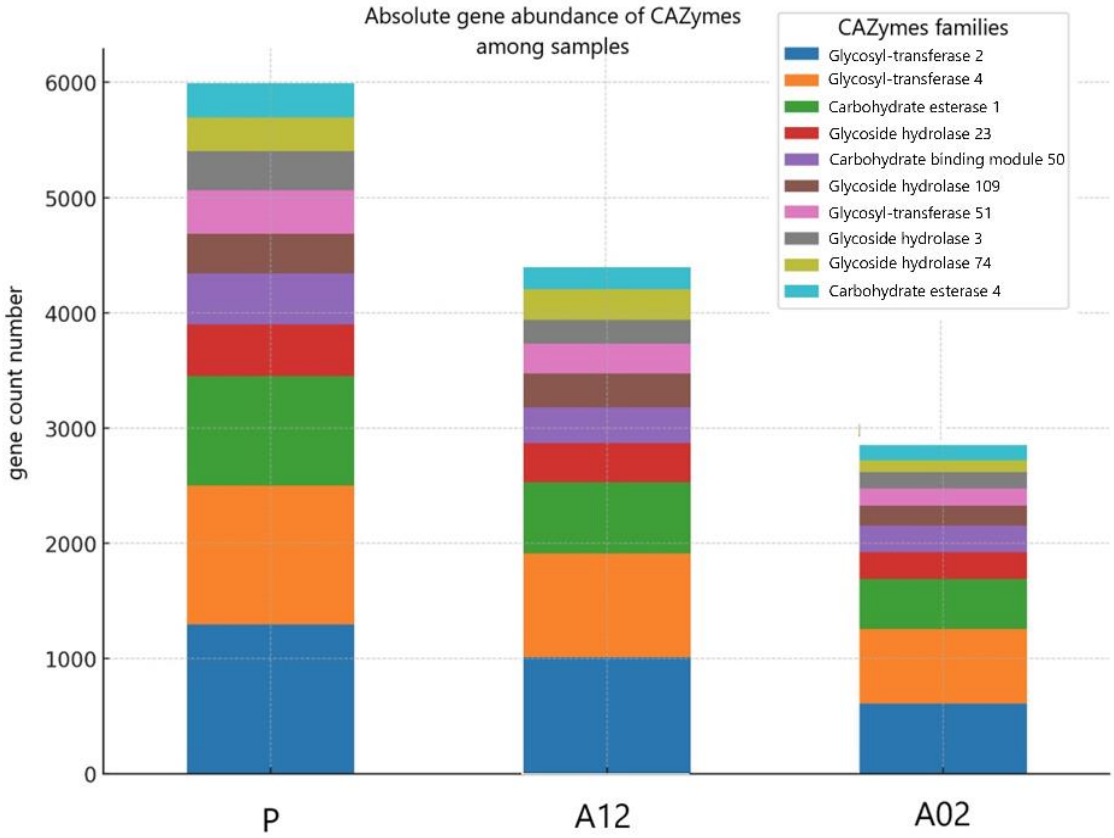

Figure 4: Gene abundance comparison for the most frequent CAZymes genes among the three samples; absolute counts are decreasing with each filtration step, general pattern of dominance is similar

Glycosyltransferase family 2 with 2911 total occurrences and Glycosyltransferase family 4 with 2762 total occurrences were the most abundant CAZyme families, therefore they appeared to be important in all analyzed samples. Both families consistently showed the highest gene counts in each sample. Families like the Carbohydrate Esterase Family 1 (1977 occurrences), Glycoside Hydrolase Family 23 (1025) and Carbohydrate-Binding Module Family 50 (977) were moderately abundant. That indicated active carbohydrate esterification and binding processes in the samples. These families also showed relatively similar distributions across all samples.

The particle fraction showed the highest frequency of CAZyme genes across almost all families, implying more active metabolism involved in carbohydrate degradation compared to the ambient water samples A02 and A12. A12 had a higher count of Glycoside Hydrolase Family 74 (268 occurrences) and Glycoside Hydrolase Family 109 (297 occurrences) than the other two samples. A02 generally showed the lowest frequencies, but the sample had a notable number of Glycosyl transferase 4 (650) and Glycosyl transferase 2 genes (606) occurrences.

The predominance of glycosyl transferase families (e.g., GT2, GT4, GT51) indicated that glycosylation was the most common carbohydrate metabolism process across all samples. Glycoside hydrolase families (GH23, GH74, GH109) were enzymes involved in carbohydrate degradation, which are linked to the breakdown of plant or other microbial polysaccharides, while Carbohydrate esterase families (CE1, CE4) have enzymatic activity related to the modification of polysaccharides.

The most frequent genes for each individual sample did not differ much, as the most abundant genes were the same for each sample, with minimal deviations in order, as depicted in Table 1. When comparing relative numbers, there were not many enzyme families with a deviation greater than 1% across samples.

Family (P)	Frequency (P)	Family (A12)	Frequency (A12)	Family (A02)	Frequency (A02)
GT2_Glycos_transf	1294	GT2_Glycos_t	1011	GT4	650
GT4	1210	GT4	902	GT2_Glycos_transf	606
CE1	946	CE1	617	CE1	434
GH23	451	GH23	339	GH23	235
CBM50	440	CBM50	309	CBM50	228
GT51	379	GH109	297	GH109	172
GH109	344	GH74	268	GT51	150
GH3	338	GT51	256	GH3	141
CE4	299	GH3	206	CE4	133
GH74	293	CE4	190	AA3	114

Table 1: The ten most frequent CAZymes genes for each individual sample are shown; it is clearly visible that there is only variety in numbers, but not in enzyme patterns

It was notable that most CAZymes were involved in carbohydrate degradation, namely glycoside-hydrolase (GHs) and glycoside-transferase families (GTs). Additionally, relative contributions were quite consistent across samples. The only exceptions with >1% variation was GT2 in the particle sample, GH74 in the intermediate A12 and GT4 in the free-living fraction of the ambient water, A02.

3.3 Peptidases

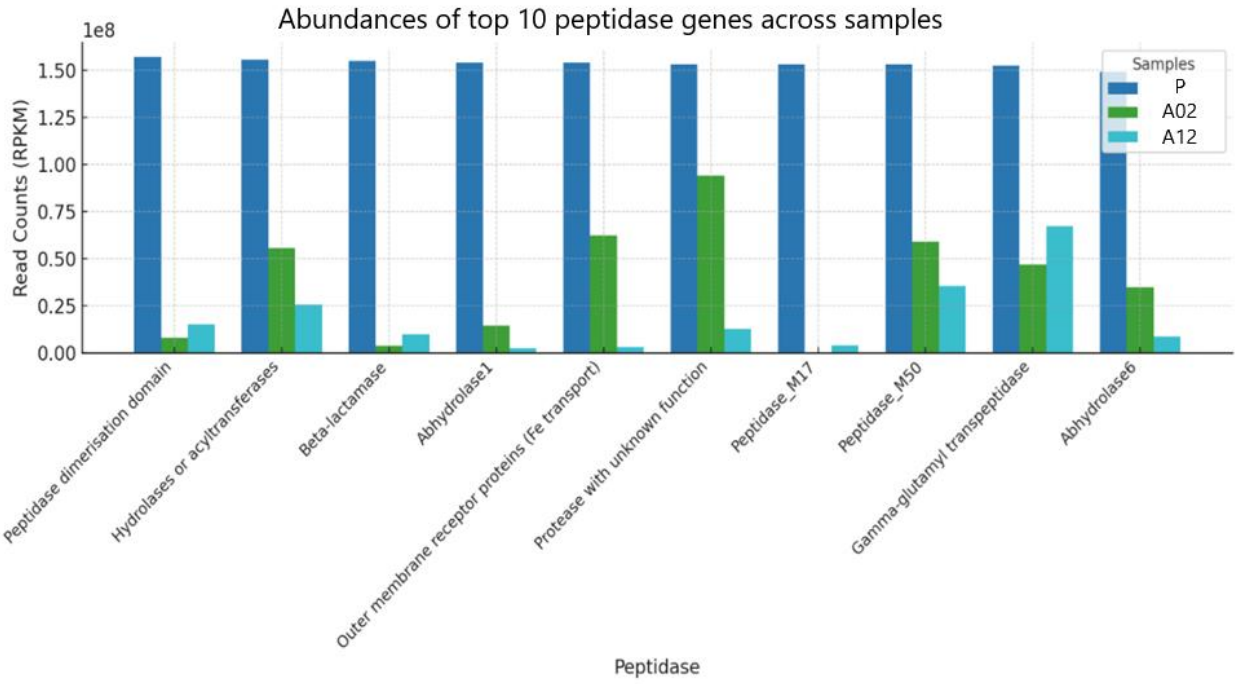

Figure 5: The ten total most frequent peptidase genes and their RPKM in each sample. The bars indicate that the abundance is higher for every gene in P, but there is no clear pattern for A12 and A02

A peptidase analysis was done for the ten most common peptidase genes across the three samples, P, A02, and A12 based on their respective read counts. Kruskal-Wallis ANOVA revealed highly significant differences for all comparisons ($p < 0.00001$).

As shown in Figure 5, the peptidase dimerization domain had the highest read counts in the particle associated sample with 157.018.795 total reads. It was also found in A02 with 7.897.111 reads and A12 (15.131.776 reads). Hydrolases or acyltransferase genes were very abundant in P with a total of 155.373.745 reads and present in a moderate level in A02 with 55.574.708 reads and A12 with 25.557.015 reads. The genes for Beta-lactamase were found in all samples, particularly in P (154.907.055), with lower counts in A02 (3.515.692) and A12 (9.781,887). Abhydrolase1 genes were highly represented in P (154.020.405) and A02 (14.444,318), with lower counts in A12 (2.473.377). Outer membrane receptor proteins for Fe transport were found in the particle associated sample (153.887.415) and in the free-living fraction of the ambient water A02 (62.443.227), and were represented minimally in the intermediate A12 (3.190.625). A protease with an unknown function was present in all samples, highest in A02 (94.113.064) followed by P (152.947.055) and A12 (12.679.559).

These results showed that some peptidases, for example the peptidase dimerization domain and hydrolases or acyltransferases, were found across all samples, while others were found

exclusively in individual samples. This proved the importance of these enzymes in marine snow ecosystems, while the peptidases like peptidase M17, which were highly present in some samples and almost absent in other samples probably indicated functions associated with the individual properties of the respective sample. It was also notable that the general trend in the previous analyses, where count numbers and frequencies decreased in order P > A12 > A02 was not reflected for a notable number of peptidase gene counts, where count numbers for A02 were higher than the numbers for A12 (Figure 5). P still dominated the pattern of gene counts for the top 1% of the most frequent peptidase genes as it is clearly visible for the ten most common peptidases in Figure 5.

3.4 Metabolic pathways

Metabolic pathway analysis was done with a variety of different tools. EggNog-Mapper, KEGG Orthologies and UniProt BLAST searches were done to identify, describe and correctly annotate metabolic pathways. The total RPKM values for individual genes were different among the samples. As count numbers sometimes differed by several orders of magnitude, relative numbers are depicted in Figure 6 for better comparability. Figure 6 shows the most frequent pathways for all samples combined and the relative contribution of each sample. The difference between samples was highly significant ($p < 0.00001$) for all comparisons. Again, Kruskal-Wallis and Mann Whitney pairwise tests with Bonferroni correction were used due to the non-normal distribution of the dataset.

The analysis of EggNog-Mapper annotations showed that the gene numbers for the most commonly found pathways were highly different for each sample. The gene for fructose-1,6-bisphosphatase had a total read count of 262.215,471 across all samples. The highest expression for these genes was found in the free-living fraction of the ambient water A02 (118.339,494) and the lowest in the particle associated sample. (44.105,032). Other highly expressed pathway genes were those involved in isoprenoid synthesis and aldehyde dehydrogenase family proteins. These pathways also showed varied expression patterns. That suggested that metabolic

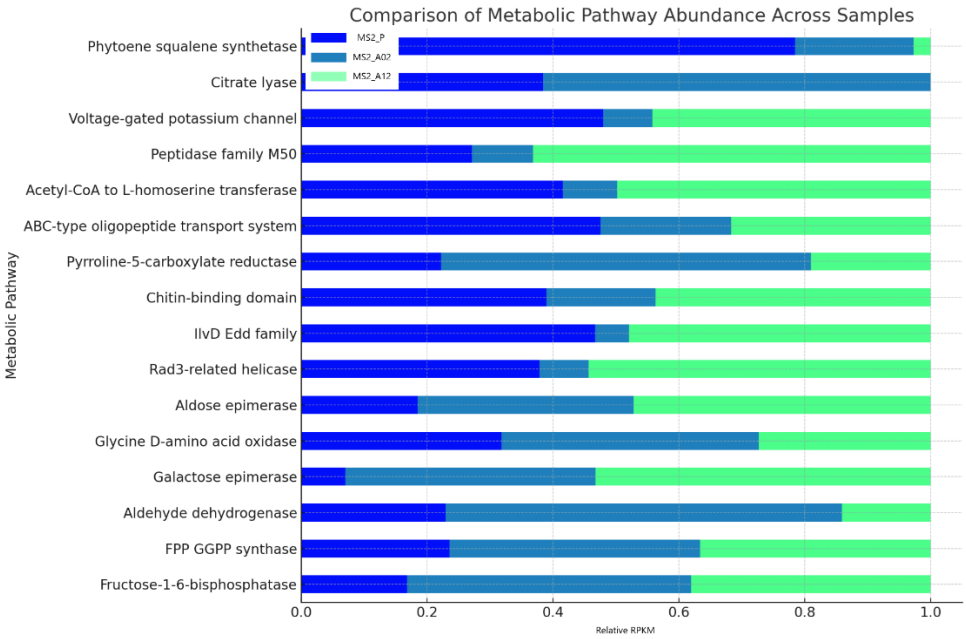

Figure 6: Comparison of contribution of each sample to the total most common identified metabolic pathways with EggNog Mapper. It is clearly visible that different processes dominated each sample

adaptations were present in each sample. Genes associated with efficient transport and defense, or stress mechanisms were found preferably in decreasing order A02 > A12 > P, processes associated with carbohydrate degradation and nutrient acquisition were found in higher numbers in the opposite order.

Comparing the genes with taxonomic annotations revealed that the pathways were found in all bacterial phyla detected in the previous analysis, but were mainly associated with *Pseudomonadota* and *Bacteroidota.* These phyla are known for their key roles in marine snow degradation. This was in accordance with the findings from the taxonomical analysis, which identified the genes of these phyla as the most abundant. *Alphaproteobacteria* expressed genes for enzymes like glycine deaminating oxidases, whereas *Gammaproteobacteria* highly expressed genes for chitin-binding proteins. *Bacteroidetes* showed high numbers of carbohydrate metabolism enzyme genes. It is notable that in absolute numbers the gene counts for the most frequent pathways from EggNog annotations were relatively even across all samples.

A similar pathway analysis was done with the KeGG database. Again, the data showed that marine snow bacteria break down and recycle the POM, including nitrogen compounds, and lipids. These are essential processes for nutrient recycling in marine snow and the biological pump. While general dominating processes showed relatively similar patterns (Figure 8), the analysis of the most frequent genes was very different for each individual sample (Figure 7).

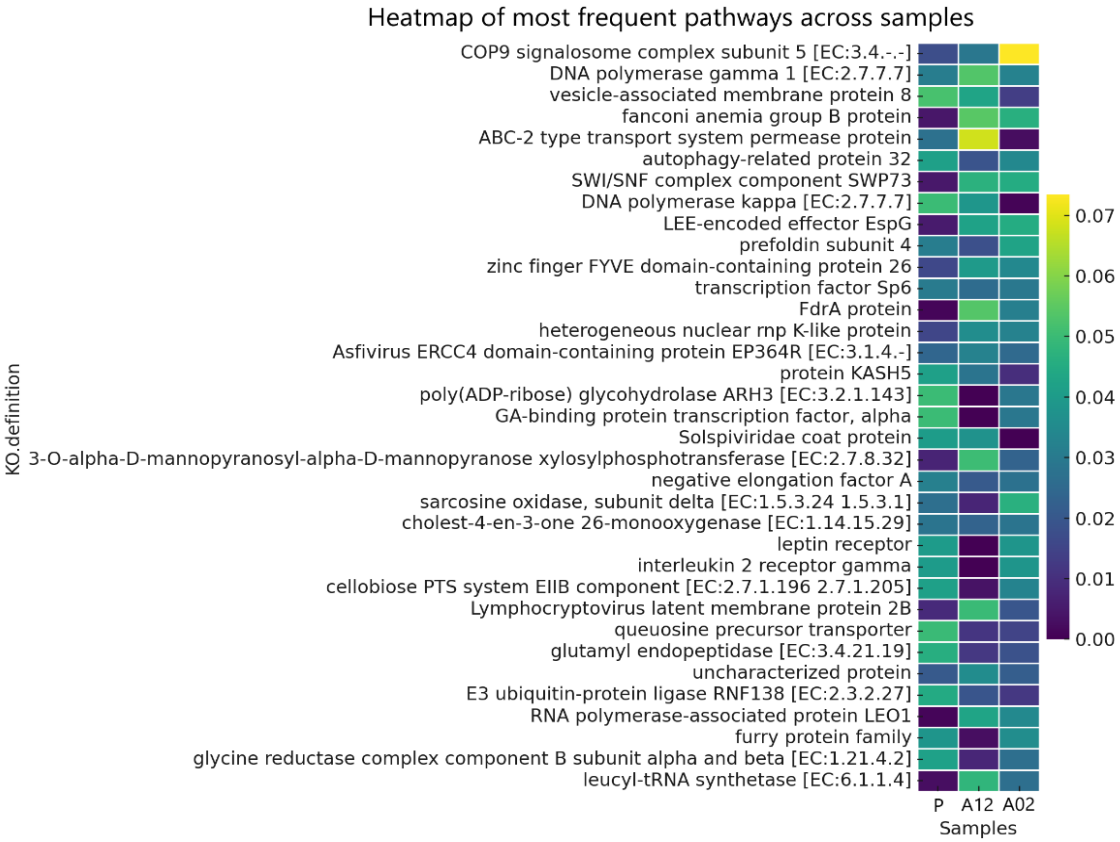

Figure 7: Heatmap of most frequent pathways according to Kegg Orthologs. Again, it was clearly visible that patterns in pathways were very different for each sample. KO Orthologs are more specific and provide insights into detailed cellular processes, while EggNog Annotations provide a more general overview.

The most frequent metabolic pathways in particle sample P were involved in transport processes and genetic information processing. The most frequent pathways (in order) included Vesicle-associated membrane protein 8, a pathway with a role in vesicle trafficking and intracellular transport. DNA polymerase kappa, an enzyme involved in DNA repair, was the second most abundant. Poly-ADP-ribose-Glycohydrolase ARH3, which degrades poly-ADP-ribose, was involved in energy metabolism. GA-binding protein transcription factor alpha was part of active transcriptional regulation processes. The Queuosine precursor transporter is again involved in nucleotide transport and processing.

The metabolic pathways in the intermediate sample A12 were dominated by amino acid metabolism and defense and repair mechanisms. COP9 signalosome complex subunit 5, which is involved in protein homeostasis and signaling pathways, was second most abundant. Sarcosine oxidase, subunit delta, an enzyme involved in amino acid metabolism was the third most abundant. The next most frequent pathway in sample A12 was Fanconi anemia group B protein, which plays a role in DNA repair. SWI/SNF complex component SWP73, a regulator

of chromatin remodeling, is also involved in gene expression regulation. The LEE-encoded effector Esp, typically associated with pathogenicity of the expressing organism and therefore a stress response, was the next most abundant.

In the free-living fraction of the ambient water (A02) the most frequent pathway genes were identified in the following order: ABC-2 type transport system permease protein has a role in nutrient uptake and is therefore part of the energy metabolism pathway, followed by Fanconi anemia group B protein, which was also highly abundant in A12 is involved in DNA repair, meaning it is involved in genetic information processing pathways. FdrA protein is associated with stress responses. Lastly, DNA polymerase gamma 1 is involved in DNA replication processes and again involved in the genetic information processing pathways.

Searching the KO numbers of the individual genes against the KeGG pathway reconstruction database revealed the most present pathways. The most predominant pathways were associated with carbohydrate, amino acid, energy, lipid and nucleotide metabolism, genetic information processing pathways and environmental information processing pathways, as depicted in Figure 8. This graph shows that environmental signaling genes were the most expressed genes in each sample, followed by genetic information processing. It is notable that the relative number of genes associated with lipid metabolism was far higher in the particle associated sample compared to the other two samples. Carbohydrate and amino acid metabolism genes showed a slight increase with each filtration step, while nucleotide metabolism and energy metabolism associated gene counts appeared to be relatively consistent across samples and therefore independent of potential ecological niches.

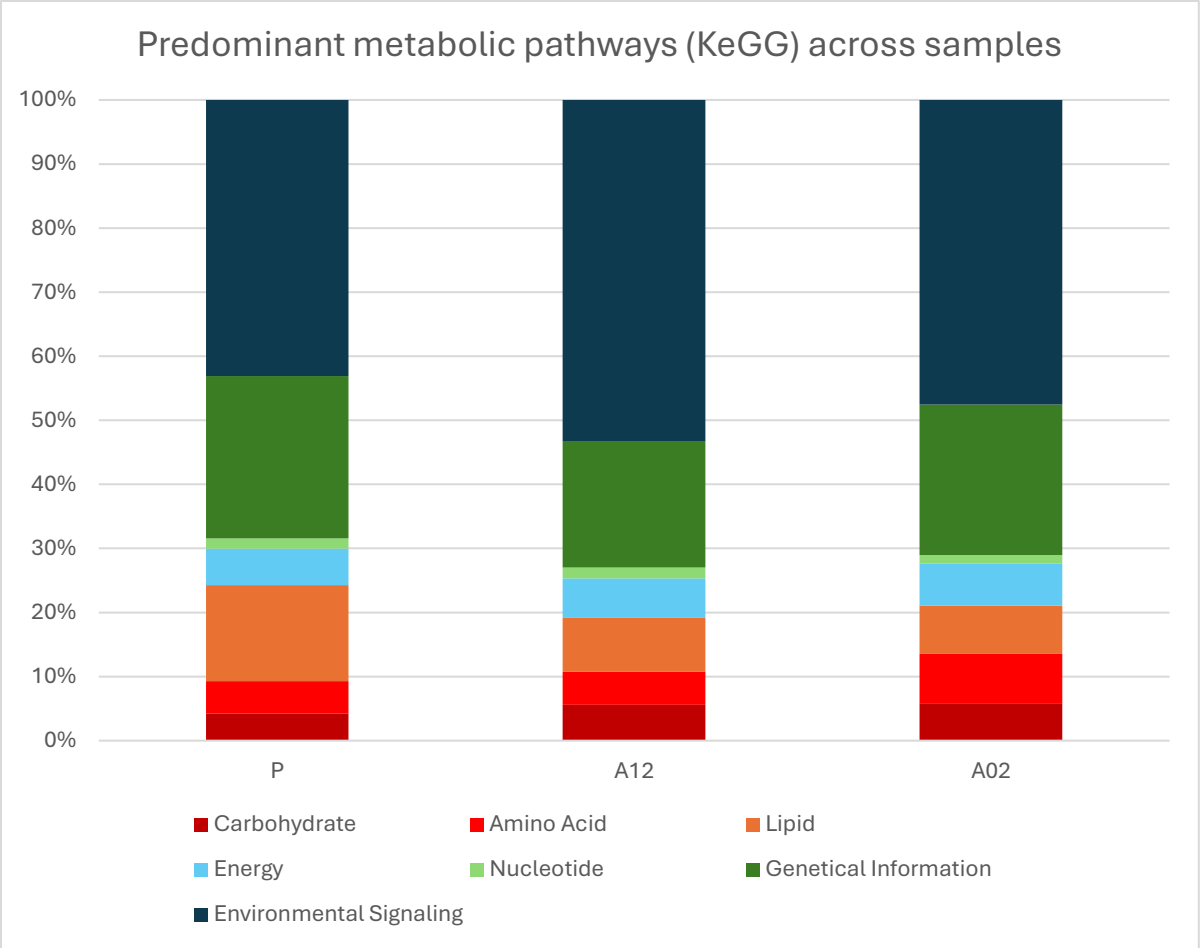

Figure 8: Relative contributions of predominant enzymatic pathways from each sample. Environmental signaling genes are by far the most abundant genes, comprising about half of the found genes. Lipid metabolism was strongly associated with particles, while amino acid and carbohydrate metabolism increased in intermediate and free-living ambient water samples. Energy and nucleotide metabolism genes are consistently common across samples.

3.5 Proteomics

The proteomics analysis was performed with Proteome Discoverer (Thermo Scientific, USA). The results showed that the number of identified proteins was quite low across the samples and this identification rate suggested that the dataset may not represent the full proteome accurately. Important results and values derived from this analysis, for example the total number of peptides and unique peptide counts per protein, again indicated low overall coverage. This implied that there is a need for further optimization of the experimental setup to achieve deeper proteome coverage. The low number of detected proteins was probably caused by the self-implemented method for isotope detection, as Proteome Discoverer Software does not offer an integrated workflow for ^{15}N label detection. Analysis was still

done by comparing the relative frequencies of the detected proteins and the affiliated genes and checking for general patterns between labeled and unlabeled proteins.

The labeled proteins in the particle fraction revealed association with key pathways for nitrogen metabolism (e.g., aminotransferase and alpha-L-glutamate ligase) as well as signaling pathways (adenylate and guanylate cyclase catalytic domain) and post-translational modifications in genetical information processing (SPOUT MTase). These proteins indicated that the bacteria in the particle fraction were incorporating nitrogen from the labeled algae, with a focus on proteins involved in amino acid metabolism, nitrogen assimilation, and cellular communication. The labeled proteins found in A02 were connected to environmental sensing (MCPsignal) and there were also defense proteins (CRISPR-associated protein Cse3) along with nitrogen metabolism associated proteins (glycine degradation). The P-Pant transferase protein was again involved in genetical information processing. The labeled proteins in the intermediate A12 were dominated by a combination of stress responses (AI-2E family transporter and Chaperone SurA) and and genetical information processing proteins (methyl esterification). These proteins clearly showed that the bacteria were not only processing nitrogen like in the particle sample, instead they were also adapting to environmental and cellular stress.

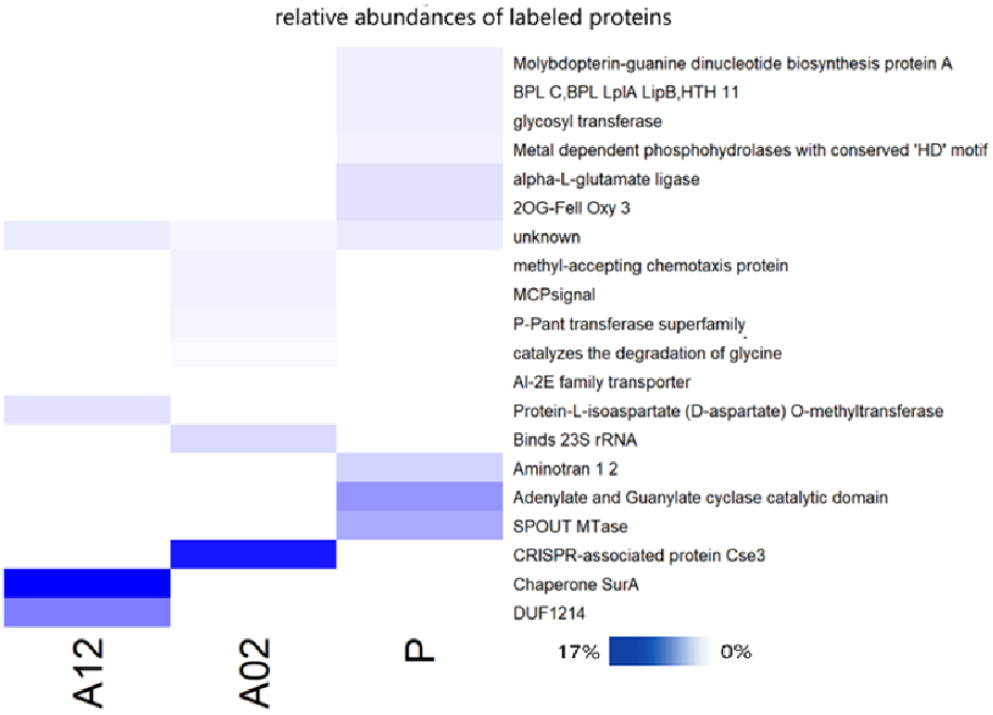

Figure 9: Relative abundances of labeled proteins that contribute to more than 1% of the total labeled proteins. Proteins with unknown function were aggregated. Dark blue color indicates higher concentrations of up to 17% for

protein Cse3 in A02 and Chaperone SurA in A12. DUF1214 follows with 5% for A12, and SPOUT MTase with 3%. Light blue colors indicate between frequency between 1%-3%

Most of the most abundant ^{15}N-labeled proteins were linked to clearly described pathways and functions as shown in Figure 9. In sample A02, many labeled proteins had the annotation "binds 23S rRNA", but no detailed explanation on function could be provided with the used databases. The found proteins are involved in general carbohydrate, lipid and amino acid metabolism.

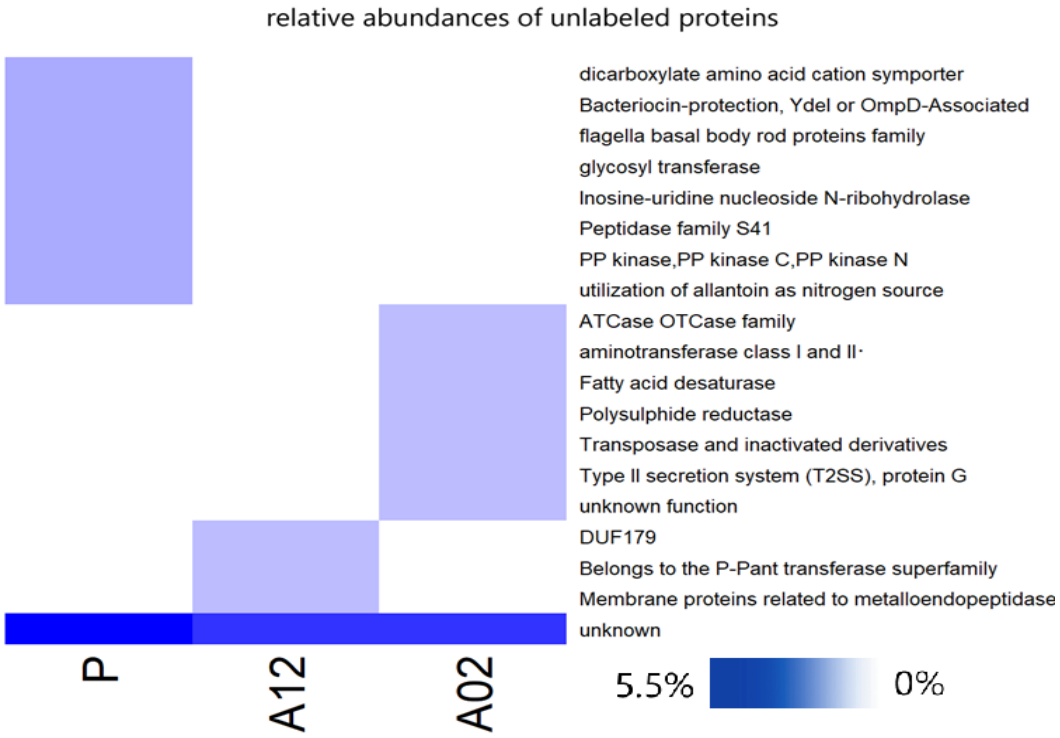

Figure 10: Relative abundances of unlabeled proteins that contribute to more than 1% of the total unlabeled proteins. Proteins of unknown function clearly dominate with up to 5.5% of total proteins. Light blue color indicates frequencies between 1-3%.

The unlabeled could often not be associated with any greater pathway categories or even had unknown functions as seen in Figure 10. Additionally, there were proteins found without clear functional characterization like DUF179 and also many entries with missing annotations, which made functional characterization impossible. This poorly described fraction amounted to a total of 13% of the unlabeled proteins. Searching their sequences against the UniProt BLAST database also did not lead to any results. Most of the characterized unlabeled annotated proteins were Transporters and Secretory Proteins. Several proteins such as dicarboxylate amino acid cation symporter, AI-2E family transporter, and Type II secretion system proteins play roles in energy metabolism and intercellular signaling. Some proteins are also involved in carbohydrate (glycosyl transferase), protein (peptidase S41) and lipid (fatty

acid desaturase) metabolism but they were only represented in individual samples to a small degree.

An ANOVA revealed the significant difference ($p=0.012$) between labeled and unlabeled proteins, after log-transformation of the datasets.

Taxonomic patterns for ^{15}N labels were analyzed by comparing 108 most dominant and well annotated proteins, 54 labeled and 54 unlabeled and their associated bacterial classes. No significant difference could be identified between groups (Kruskal-Wallis $p=0.65$). Almost half of the bacteria of each group belonged to the *Gammaproteobacteria* class (21 and 28 for labeled vs unlabeled proteins), followed by *Alphaproteobacteria* (24 and 14, respectively). The unlabeled proteins mostly belonged to poorly defined functional categories without meaningful annotations, with some exceptions, mainly proteins for transporters and signaling. Labeled proteins were mostly connected to the well-characterized pathways of general nutrient metabolism, genetical information processing, signaling and energy metabolism (Figure 9 & 10).

4. Discussion

4.1 Marine snow particles as microhabitat for bacteria

The observed taxonomic patterns of bacteria living on particles and ambient water reflect the existence of ecological niches and their habitat preference. Even though there is no distinct pattern of phyla across samples, the abundance data indicates preferred ecological niches. The high abundance of *Pseudomonadota* and *Bacteroidota* genes (Figure 2&3) in marine snow aligns with other studies that have documented the preferential association of these phyla with particulate organic matter in marine systems (Azam & Long, 2001; Simon et al., 2002). Such particles provide a concentrated source of nutrients, creating microenvironments for heterotrophic bacteria that can utilize complex organic matter (DeLong et al., 2006). The enrichment of the *Planctomycetota* phylum in the particle sample shows that this phylum has a preference for this substrate type, because these microbes prefer low-oxygen microhabitats within POM aggregates and is metabolically versatile (Wiegand et al., 2018). Furthermore, the

presence of other mostly anaerobic taxa also indicates that living in and on marine snow provides anaerobic microhabitats (Vojvoda et al., 2014). This is underlined by the almost twofold difference in relative gene counts for *Bacteroidetes* in A02, while the diverse and adaptive phyla of *Pseudomonadota* is consistently abundant in all samples. *Bacteroidetes* are known to contribute to a huge degree to the degradation of organic matter and particularly thrive during algal blooms, which underlines their preference for growth attached to marine snow particles (Fernández-Gómez et al., 2013). These algal blooms lead to an increase in POM particles, which is the substrate for *Bacteroidota* and therefore lead to an increase in cell abundance. The diverse *Pseudomonadota* phylum can be associated with free-living, and particle associated lifestyles and is therefore found in similar counts in all samples. The different adaptations are reflected in the metabolic pathways (Giovannoni et al., 2012).
This underlines the uniqueness of marine snow as an ecosystem, as marine bacteria in other habitats are metabolically compositionally different (Kiørboe et al., 2003). Their metabolic capabilities make these bacteria a huge contributor in biogeochemical cycles like the biological pump in marine ecosystems, because they cycle nutrients and decompose the organic matter in marine snow (Kaltenböck & Herndl, 1992; Orsi et al., 2018; Grossart et al., 2007). The findings that there are distinct habitats and niche partitioning on marine snow and ambient waters are also underlined by the high β-diversity.

Generally, nutrients seem to be more available in particular form, as overall gene counts for bacteria decrease with each particle filtration step, and less nutrients provide less support for life.

4.2 Inferences from predominant enzymes and metabolic pathways

The analysis of CAZyme genes revealed that relative numbers were consistent across samples, with one exception per sample. The specific enrichment of certain families is difficult to interpret, as families with varying functions were analyzed, and even if equal numbers of the same family are found, there could be very different processes active in every sample, as revealed by the EggNog annotations.

The bacterial communities in marine snow were also found to have different functional characteristics. The metabolic pathway analysis with EggNog-Mapper and KeGG Orthologies revealed that the major processes are associated with the classes *Flavobacteria*,

Alphaproteobacteria, and *Gammaproteobacteria*, followed by *Firmicutes*, each specialized in different processes. This again suggests niche differentiation (Baquero et al., 2021).

The KeGG analysis of the most frequent pathways revealed that *Alphaproteobacteria*, such as *Sulfitobacter* and *Rhodobacteraceae* were involved in pathways associated in the biosynthesis of isoprenoids (K02523), the synthesis of branched-chain amino acids (K01687), and the transportation systems of peptides (K02035). This is an indicator that they live a generalist, opportunistic lifestyle with a focus nutrient acquisition which is in accordance with the general knowledge about the lifestyles of *Alphaproteobacteria,* which are known as a metabolically diverse taxon (Grossart et al., 2006). Because they are opportunistic generalists, *alphaproteobacteria* use their metabolic adaptability to adjust to changing nutritional conditions. They live mostly in areas of marine snow with accessible dissolved organic materials, as evidenced by their dependence on peptide absorption and lipid metabolism. (Gupta, 2005).
Aldose epimerase (K01785), citrate lyase (K18118), and chitinases (K01183) are the most present in *gammaproteobacteria*. This suggests adaptation to breaking down organic materials with a high molecular weight, such as chitin (Beier, 2013).

Glycerophosphoryl diester phosphodiesterase (K01113) was used to identify *firmicutes*. This enzyme is used as transporter in the breakdown of phospholipids and cell wall synthesis (Larson et al., 1983).
Generally, the enzyme patterns demonstrate niche specialization and functional complementarity. This is statistically proven by the highly significant differences in all comparisons regarding gene frequencies across samples (Figure 6, 7).
This partitioning shows the effectiveness of microbial communities in breaking down and recycling organic material in marine snow. which underlines the importance of microbes in the biological pump and in increasing nutrient availability for the larger ecosystem (Duret et al., 2019). Analyzing the predominant pathways from each sample individually illustrates the functional complementarity between samples, when investigating the individually most frequent genes for each sample and their associated pathways.
The detailed EggNog analysis reveals that the particle sample has a broader focus on general cellular processes like vesicle trafficking, transcription regulation, and nucleotide metabolism, while A12 appears to have more processes involved in protein assembly, amino acid metabolism, DNA repair, so processes involved in active gene regulation and energy utilization. Sample A02 focuses on transport systems and carbohydrate metabolism and adaptations to nutrient uptake.

These findings again suggest functional complementarity between the microbial communities, underscoring their collective ability to adapt to distinct ecological niches and efficiently recycle nutrients. The partitioning of metabolic functions likely contributes to the overall stability and resilience of the microbial ecosystem (Cornell et al., 1992). There are general processes which appear to be quite evenly distributed and expressed among samples, but processes that contribute to acquiring energy or defense mechanisms are differently expressed among samples. If predominant EggNog pathways are explored in depth and compared to functional databases there is again a pattern visible, which shows the nutrient availability in POM, as metabolic pathways here are focused on general degradation pathways (e.g. high number of CAZymes, phytoene synthetase, high number of peptidases) while pathways in A02 are focused on efficient transport (COP9), protein homeostasis (peptidase M50), stress responses (CRISPR Cse3) and general competition mechanisms. Sample A12 tends to be an intermediate level between the particle associated communities and the free-living communities. The strong decrease of peptidase gene frequencies suggests that amino acids are metabolized preferentially by marine heterotrophic bacteria, either because they are the most limited nutrient or because they are more easily degradable, compared to the complex carbohydrates from marine snow (Wünsch et al., 2019). The results of the KO pathway analysis revealed that lipid metabolism genes were the most abundant of all genes involved in macronutrient uptake in the three samples. The count number was notably high in the particle associated sample. As *T. intea* is particularly known for its high lipid contents, it can be assumed that this is the reason for the high number in lipid metabolism associated genes (Huang et al., 2019). Therefore, niche partitioning, community composition and metabolic pathways might be shaped by substrate properties. This means that in situ produced marine snow, especially when originating from only one algal species, might not reflect actual chemical properties of marine snow from the ocean. It is also important to note that EggNog and KeGG pathway analyses highlight very different processes. While KeGG analyses reveal very general pathway patterns, the EggNog analysis is way more detailed and can lead to different conclusions and assumptions.

4.3 Inferences from multi-omics approaches and stable isotope proteomics

The presence of unlabeled proteins can lead to the assumption that specific proteins are synthesized or utilized in response to distinct environmental conditions, resource availability, or ecological roles. This would also support the finding that different taxa occupy niches on

marine snow and ambient water, depending on nutrient availability and competition. In a diverse microbial community, bacteria often partition available resources to optimize their ecological roles (Duret et al., 2019).

The labeled proteins indicate involvement in metabolic pathways actively used through the experiment by incorporating nitrogen from the ^{15}N-labeled algae. These proteins are directly involved in the breakdown, assimilation, and use of the algal nitrogen. The presence of unlabeled proteins suggests that these molecules are either unrelated to the immediate assimilation of algae-derived nitrogen or they could also represent pathways specialized for other niches, connected to alternative nitrogen sources present in the water. As the bacterial community originated from the Adriatic Sea, this could be a source of alternative nitrogen. This unlabeled nitrogen could be obtained from free floating detritus or maybe through recycling of the original bacterial community. The unlabeled proteins might correspond to bacteria and pathways adapted to metabolize alternative nitrogen or carbon sources e.g., dissolved inorganic nitrogen. The findings in this study confirm that there is a variety of proteins in marine bacteria that are uncharacterized (Sanejouand, 2023). More research is needed, to get deeper insights into niche development and to fully understand degradation processes of marine organic matter (Watson et al., 2005; Evans et al., 2007; Lozano et al., 2022).

The unlabeled proteins could also reflect proteins that were not actively synthesized during the experiment. Some of the annotated proteins can be associated with defense or stress mechanisms rather than nutrient acquisition. These proteins are generally synthesized at much lower rates under normal conditions (Perez-Llarena et al., 2016). Such a division indicates that the bacterial community has specialized metabolic strategies to adapt to variable conditions and that the individuals are occupying different niches (Bochdansky et al., 2017). Most labeled proteins are involved in active nitrogen metabolism for example, such as amino acid synthesis and protein production.

A correlation between RNA levels and protein expression was not detected in this experiment. Generally, there is evidence for decoupling of these two variables, as shown in Zhao et al. (2024a).

By analyzing the species information based on the lowest common ancestor algorithm associated with the incorporation patterns of ^{15}N into their proteins, there is more evidence for niche partitioning. The results show that the particle-attached community (P) specializes in degradation and recycling processes of POM sources and the chemical modifications of the recycled molecules with proteins like Adenylate and Guanylate cyclase catalytic Domain,

Aminotransferases and SPOUT Methyltransferases. The most commonly found proteins in this sample are expressed by *Bradyrhizobiaceae*, *Ruegeria*, *Alteromonadaceae* and *Shewanellaceae.* In contrast to the particle sample, the community in the ambient water (A12) relied more on nutrient uptake and maintenance of cellular function rather than direct POM degradation. This is a niche strategy with the focus on scavenging and survival in an environment with less available nutrients. This lifestyle is indicated by proteins like the AI-2E family Transporter, the Chaperone SurA or a protein involved in methyl esterification of amino aicd residues. Most labeled proteins of this sample are associated with *Colwelliaceae*, an unclassified *Gammaproteobacteria* bacterium, *Alphaproteobacteria* and *Hyphomonadaceae*. The patterns in the free-living fraction suggest that the A02 microbial community is adapted to a lifestyle where optimized resource acquisition and modification, as well as defense mechanisms are advantageous. This is reflected, for example, by the CRISPR-associated protein Cse3, the methyl-accepting chemotaxis Proteins or the P-Pant transferase superfamily. The associated bacteria in this sample are *Vibrionales* and *Alteromonadaceae.* These functional differences reflect the existence of individual ecological strategies that minimize direct competition. Another factor that indicates niche partitioning is the difference in bacterial species that express the described proteins in each sample and use most of the labeled nitrogen source for different processes depending on the environmental conditions. By focusing on the labeled proteins, we gain information about metabolic processes that actively happened during the experiment. Thus, we can conclude that there is active niche partition due to the functional complementarity between samples, and because the predominant proteins in each sample were produced by different types of bacteria (Titus et al., 2024, Von Bergen et al., 2013).

The coexistence of labeled and unlabeled proteins in all three treatments implies that there are other ways present to acquire nitrogen, than by degradation of algae. To further explore these patterns, an optimized analysis workflow is needed, especially for identification of nitrogen isotopes in Proteome Discoverer. Even though the quality of the results in this study is not good enough for robust statistical tests, analyses for general patterns are possible and SIP-proteomics with ^{15}N appear to be a valuable method for in-depth analysis of nitrogen assimilation patterns in relation to different variables (Taubert et al., 2013; Jehmlich et al., 2016). Additionally, it can be concluded that there are not many software packages available that offer implemented workflows for the detection of ^{15}N. Other programs like MaxQuant, MSFragger, ProteoWizard and OpenPepXL were considered during the analysis, but none of them has built-in options for

^{15}N detection. Therefore, software with integrated option for ^{15}N label detection is urgently needed, to improve quality of results and time efficiency in data analysis.

4.4 Possible limitations and challenges

The proteomics analysis using Proteome Discoverer found a lower than expected number of identified proteins. This limited protein identification may be caused by several factors, for example insufficient protein abundance, mistakes in sample preparation, or errors during mass spectrometry analysis. Additionally, high filtering thresholds applied in the data analysis with Proteome Discoverer may have led to the low protein identification. This might exclude lower confidence matches, leading to a smaller number of identified peptides. As a result, the dataset does not provide an optimal representation of the proteome, because there is a chance that some proteins remain undetected. This might restrict the ecological insights that can be derived from the analysis. For the future, optimization of sample preparation and especially data analysis workflows could improve proteome coverage and enhance the quality of the results. Trends are still visible and can be discussed, even though small sample sizes may lead to not very robust results. Problems in low significance results for stable isotope analyses are namely an insufficient number of available peptides, as well as the overall quality of the MS data, especially regarding peptide masses and signal intensities (Taubert et al., 2013). For this study the low overall coverage in combination with the self-implemented detection algorithm for Proteome Discoverer might have been the key problem in identifying labeled proteins. Generally, protein-based stable isotope probing is an understudied, but valid method that is capable of producing high quality results in analyzing patterns in proteomes. Used methods that lead to high quality results in association with ^{15}N labeling in mass spectrometry often used software for mass spectra visualization like QualBrowser (Thermo Fisher Scientific, USA) and manual analysis of the data (Seifert et al., 2013; Kieft et al., 2021). This seems to be a promising and effective approach, but requires a lot of time and moderate expertise, especially with large scale datasets. If all of this is taken into account, ^{15}N labeling methods might be even more precise than using carbon isotopes, because nitrogen stays in the sample and carbon is always lost due to outgassing in the form of CO_2, especially when utilized by heterotrophic bacteria, which is already shown in experiments with plants and might also apply to marine bacteria (Abadie et al., 2018).

For the analysis of different enzyme classes like peptidases and CAZymes, it can be difficult to make comprehensive conclusions, as the interpretation of frequency differences of enzymes that belong to the same enzyme family (e.g. glycosidase transferase family 2 & 4) is hardly possible as they are functionally very similar, yet measured amounts differ by orders of magnitudes. Also, equal concentrations of the same families across different samples, might hide the fact that there are completely different enzymes present, as these families are not necessarily substrate specific, as described in databases like cazypedia.org and cazy.org and therefore it is difficult to make coherent conclusions with this information.

There is also some criticism about multi-omics approaches which cannot be neglected. First, during taxonomic analysis most studies only analyze and compare higher taxonomic classes, due to the nature of the LCA-algorithm (lowest common ancestor), which can only distinguish genes to a certain taxonomic level. Therefore, actual biodiversity levels might not be accurately reflected, and especially niche differentiation can be underestimated, due to the high adaptivity of phyla like *Pseudomonadota* (Buchan et al., 2014). *Alphaproteobacteria* and *Gammaproteobacteria* prefer different lifestyles and ecological niches as indicated by large occurrence differences in marine samples, yet they are analyzed as one phylum in most metagenomics studies (Franco et al., 2017).

Even though multi-omics are proven to be a valuable tool in the identification of ecosystem processes, some general technical and biological challenges still apply. The many variables in biological systems may lead to difficulties in comparability of datasets. Genetic potential and actual protein expression patterns can be uncoupled or correlated, and it is difficult to take external factors like RNA degradation and complexity of enzyme families into account, which can lead to possible sampling bias (Subramanian et al., 2020; Zhao et al., 2024a). There can also be technological limitations on several levels as shown in this experiment.

5. References

- Abadie, C., Bathellier, C., & Tcherkez, G. (2018). Carbon allocation to major metabolites in illuminated leaves is not just proportional to photosynthesis when gaseous conditions (CO2 and O2) vary. *New Phytologist*, *218*(1), 94-106.

- Alldredge, A. L., & Silver, M. W. (1988). Characteristics, dynamics and significance of marine snow. *Progress in oceanography*, *20*(1), 41-82.

- Azam, F., & Long, R. A. (2001). Sea snow microcosmos. *Nature*, *414*(6863), 495-498.

- Baltar, F., J. Aristegui, J.M. Gasol, E. Sintes, G.J. Herndl, (2009). Evidence for dependence of prokaryotic metabolism on suspended particulate organic matter in the dark waters of the subtropical North Atlantic. Limnol. Oceanogr., 54: 182-193

- Baquero, F., Coque, T. M., Galán, J. C., & Martinez, J. L. (2021). The origin of niches and species in the bacterial world. *Frontiers in microbiology*, *12*, 657986.

- Bergauer, K., Fernandez-Guerra, A., Garcia, J. A., Sprenger, R. R., Stepanauskas, R., Pachiadaki, M. G., ... & Herndl, G. J. (2018). Organic matter processing by microbial communities throughout the Atlantic water column as revealed by metaproteomics. *Proceedings of the National Academy of Sciences*, *115*(3), E400-E408.

- Bochdansky, A. B., Clouse, M. A., & Herndl, G. J. (2017). Eukaryotic microbes, principally fungi and labyrinthulomycetes, dominate biomass on bathypelagic marine snow. *The ISME Journal*, *11*(2), 362-373.

- Boeuf, D., Edwards, B. R., Eppley, J. M., Hu, S. K., Poff, K. E., Romano, A. E., ... & DeLong, E. F. (2019). Biological composition and microbial dynamics of sinking particulate organic matter at abyssal depths in the oligotrophic open ocean. Proceedings of the National Academy of Sciences, 116(24), 11824-11832.

- Buchan, A., LeCleir, G. R., Gulvik, C. A., & González, J. M. (2014). Master recyclers: features and functions of bacteria associated with phytoplankton blooms. *Nature Reviews Microbiology*, *12*(10), 686-698.

- Cornell, H. V., & Lawton, J. H. (1992). Species interactions, local and regional processes, and limits to the richness of ecological communities: a theoretical perspective. *Journal of animal ecology*, 1-12.

- de La Rocha, C. L., & Passow, U. (2006). The biological pump. *Treatise on geochemistry*, *6*, 83-111.

- DeLong, E. F., Preston, C. M., Mincer, T., Rich, V., Hallam, S. J., Frigaard, N. U., ... & Karl, D. M. (2006). Community genomics among stratified microbial assemblages in the ocean's interior. *Science*, *311*(5760), 496-503.

- DNeasy® PowerWater® Kit protocol by QIAGEN. (2022). Qiagen, Hilden, Germany.

- Duret, M. T., Lampitt, R. S., & Lam, P. (2019). Prokaryotic niche partitioning between suspended and sinking marine particles. Environmental microbiology reports, 11(3), 386-400.

- Evans, F. F., Raftery, M. J., Egan, S., & Kjelleberg, S. (2007). Profiling the secretome of the marine bacterium Pseudoalteromonas tunicata using amine-specific isobaric tagging (iTRAQ). *Journal of proteome research*, *6*(3), 967-975.

- Fernández-Gómez B, Richter M, Schüler M, Pinhassi J, Acinas SG, González JM, Pedrós-Alió C. Ecology of marine Bacteroidetes: a comparative genomics approach. ISME J. 2013 May;7(5):1026-37. doi: 10.1038/ismej.2012.169. Epub 2013 Jan 10. PMID: 23303374; PMCID: PMC3635232.

- Fogel, M. L., & Cifuentes, L. A. (1993). Isotope fractionation during primary production. In *Organic geochemistry: principles and applications* (pp. 73-98). Boston, MA: Springer US.

- Franco, D. C., Signori, C. N., Duarte, R. T., Nakayama, C. R., Campos, L. S., & Pellizari, V. H. (2017). High prevalence of gammaproteobacteria in the sediments of admiralty bay and north bransfield Basin, Northwestern Antarctic Peninsula. *Frontiers in Microbiology*, *8*, 153.

- Giovannoni, S. J., & Vergin, K. L. (2012). Seasonality in ocean microbial communities. Science, 335(6069), 671-676.

- Grossart, H. P., Kiørboe, T., Tang, K. W., Allgaier, M., Yam, E. M., & Ploug, H. (2006). Interactions between marine snow and heterotrophic bacteria: aggregate formation and microbial dynamics. *Aquatic microbial ecology*, *42*(1), 19-26.

- Grossart, H. P., Tang, K. W., Kiørboe, T., & Ploug, H. (2007). Comparison of cell-specific activity between free-living and attached bacteria using isolates and natural assemblages. *FEMS microbiology letters*, *266*(2), 194-200.

- Grossart, H. P., Tang, K. W., Kiørboe, T., & Ploug, H. (2007). Comparison of cell-specific activity between free-living and attached bacteria using isolates and natural assemblages. *FEMS microbiology letters*, *266*(2), 194-200

- Gupta, R. S. (2005). Protein signatures distinctive of alpha proteobacteria and its subgroups and a model for α–proteobacterial evolution. *Critical reviews in microbiology*, *31*(2), 101-135.

- Huang, B., Marchand, J., Thiriet-Rupert, S., Carrier, G., Saint-Jean, B., Lukomska, E., ... & Mimouni, V. (2019). Betaine lipid and neutral lipid production under nitrogen or phosphorus limitation in the marine microalga Tisochrysis lutea (Haptophyta). *Algal Research*, *40*, 101506.

- Janda, J. M., & Abbott, S. L. (2007). 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. Journal of clinical microbiology, 45(9), 2761-2764.

- Jehmlich, N., Vogt, C., Lünsmann, V., Richnow, H. H., & von Bergen, M. (2016). Protein-SIP in environmental studies. *Current opinion in biotechnology*, *41*, 26-33.

- Kaltenböck, E., & Herndl, G. J. (1992). Ecology of amorphous aggregations (marine snow) in the Northern Adriatic Sea. IV. Dissolved nutrients and the autotrophic community associated with marine snow. *Marine Ecology-Progress Series*, *87*, 147-147.

- Kieft, B., Li, Z., Bryson, S., Hettich, R., Pan, C., Mayali, X., & Mueller, R. (2021). Phytoplankton exudates and lysates support distinct microbial consortia with specialized metabolic and ecophysiological traits. *Proceedings of the National Academy of Sciences of the United States of America*, 118.

- Kiørboe, T., Tang, K., Grossart, H. P., & Ploug, H. (2003). Dynamics of microbial communities on marine snow aggregates: colonization, growth, detachment, and grazing mortality of attached bacteria. *Applied and Environmental Microbiology*, *69*(6), 3036-3047.

- Landa, M., Blain, S., Christaki, U., Monchy, S., & Obernosterer, I. (2015). Shifts in bacterial community composition associated with increased carbon cycling in a mosaic of phytoplankton blooms. *The ISME Journal*, 10, 39-50.

- Larson TJ, Ehrmann M, Boos W (1983). "Periplasmic glycerophosphodiester phosphodiesterase of Escherichia coli, a new enzyme of the glp regulon". J. Biol. Chem. 258 (9): 5428–32.

- Lozano, C., Kielbasa, M., Gaillard, J. C., Miotello, G., Pible, O., & Armengaud, J. (2022). Identification and characterization of marine microorganisms by tandem mass spectrometry proteotyping. *Microorganisms*, *10*(4), 719.

- Meyer, K., Ridgwell, A., Ridgwell, A., & Payne, J. (2016). The influence of the biological pump on ocean chemistry: implications for long-term trends in marine redox chemistry, the global carbon cycle, and marine animal ecosystems. *Geobiology*, 14, 207 - 219.

- Naudé, S. M. (1929). An isotope of nitrogen, mass 15. *Physical Review*, *34*(11), 1498.

- Orsi, W. D. (2018). Ecology and evolution of seafloor and subseafloor microbial communities. *Nature Reviews Microbiology*, *16*(11), 671-683.

- Osterholz, H., Singer, G., Wemheuer, B., Daniel, R., Simon, M., Niggemann, J., & Dittmar, T. (2016). Deciphering associations between dissolved organic molecules and bacterial communities in a pelagic marine system. *The ISME Journal*, 10, 1717-1730.

- Pérez-Llarena, F. J., & Bou, G. (2016). Proteomics as a tool for studying bacterial virulence and antimicrobial resistance. *Frontiers in microbiology*, *7*, 410.

- Ploug, H., Grossart, H. P., Azam, F., & Jørgensen, B. B. (1999). Photosynthesis, respiration, and carbon turnover in sinking marine snow from surface waters of Southern California Bight: implications for the carbon cycle in the ocean. *Marine Ecology Progress Series*, *179*, 1-11.

- Raimundo, I., Silva, R., Meunier, L., Valente, S. M., Lago-Lestón, A., Keller-Costa, T., & Costa, R. (2021). Functional metagenomics reveals differential chitin degradation and utilization features across free-living and host-associated marine microbiomes. *Microbiome*, *9*, 1-18.

- Rath, J., Wu, K. Y., Herndl, G. J., & DeLong, E. F. (1998). High phylogenetic diversity in a marine-snow-associated bacterial assemblage. *Aquatic Microbial Ecology*, *14*(3), 261-269.gr

- Sanejouand, Y. H. (2023). On the unknown proteins of eukaryotic proteomes. *Journal of Molecular Evolution*, *91*(4), 492-501.

- Seifert, J., Taubert, M., Jehmlich, N., Schmidt, F., Völker, U., Vogt, C., ... & Von Bergen, M. (2012). Protein-based stable isotope probing (protein-SIP) in functional metaproteomics. *Mass spectrometry reviews*, *31*(6), 683-697.

- Silver, M. W., Coale, S. L., Steinberg, D. K., & Pilskaln, C. H. (1995). Marine Snow: What It Is and How It Affects Ecosystem Functioning. In Linking Species & Ecosystems (pp. 45-51). Boston, MA: Springer US.

- Silver, M., Shanks, A., & Trent, J. (1978). Marine Snow: Microplankton Habitat and Source of Small-Scale Patchiness in Pelagic Populations. *Science*, 201, 371 - 373.

- Simon, C., & Daniel, R. (2010). Metagenomic Analyses: Past and Future Trends. *Applied and Environmental Microbiology*, 77, 1153 - 1161.

- Simon, M., Alldredge, A. L., & Azam, F. (1990). Bacterial carbon dynamics on marine snow. *Marine Ecology Progress Series*, 205-211.

- Simon, M., Grossart, H. P., Schweitzer, B., & Ploug, H. (2002). Microbial ecology of organic aggregates in aquatic ecosystems. *Aquatic microbial ecology*, *28*(2), 175-211.

- Taubert, M., Baumann, S., von Bergen, M. *et al.* Exploring the limits of robust detection of incorporation of ^{13}C by mass spectrometry in protein-based stable isotope probing (protein-SIP). *Anal Bioanal Chem* **401**, 1975–1982 (2011). https://doi.org/10.1007/s00216-011-5289-

- Taubert, M., von Bergen, M., & Seifert, J. (2013). Limitations in detection of 15 N incorporation by mass spectrometry in protein-based stable isotope probing (protein-SIP). *Analytical and bioanalytical chemistry*, *405*, 3989-3996.

- Thiele, S., Fuchs, B. M., Amann, R., & Iversen, M. H. (2015). Colonization in the photic zone and subsequent changes during sinking determine bacterial community composition in marine snow. *Applied and environmental microbiology*, *81*(4), 1463-1471.

- Thomas, F., Bordron, P., Eveillard, D., & Michel, G. (2017). Gene expression analysis of Zobellia galactanivorans during the degradation of algal polysaccharides reveals both substrate-specific and shared transcriptome-wide responses. *Frontiers in Microbiology*, *8*, 1808.

- Thomas, T., Egan, S., Burg, D., Ng, C., Ting, L., & Cavicchioli, R. (2007). Integration of genomics and proteomics into marine microbial ecology. *Marine Ecology Progress Series*, *332*, 291-299.

- Tinta T, Zhao Z, Bayer B, Herndl GJ. Jellyfish detritus supports niche partitioning and metabolic interactions among pelagic marine bacteria. Microbiome. 2023;11(1):156.

- Titus, B. M., Froehlich, C. Y., Vondriska, C., Baker, R., & Caves, E. M. (2024). Stable isotopes disentangle niche partitioning and co-occurrence in a multi-species marine mutualism. *Oikos*, e10553.

- Turner, J. T. (2015). Zooplankton fecal pellets, marine snow, phytodetritus and the ocean's biological pump. *Progress in Oceanography*, *130*, 205-248.

- Vojvoda, J., Lamy, D., Sintes, E., Garcia, J. A., Turk, V., & Herndl, G. J. (2014). Seasonal variation in marine-snow-associated and ambient-water prokaryotic communities in the northern Adriatic Sea. *Aquatic Microbial Ecology*, *73*(3), 211-224.

- Von Bergen, M., Jehmlich, N., Taubert, M., Vogt, C., Bastida, F., Herbst, F. A., ... & Seifert, J. (2013). Insights from quantitative metaproteomics and protein-stable isotope probing into microbial ecology. *The ISME journal*, *7*(10), 1877-1885.

- Wiegand, S., Jogler, M., & Jogler, C. (2018). On the maverick Planctomycetes. *FEMS microbiology reviews*, *42*(6), 739-760.

- Wilmes, P., & Bond, P. (2006). Metaproteomics: studying functional gene expression in microbial ecosystems. *Trends in microbiology*, 14 2, 92-7 .

- Wünsch D, Trautwein K, Scheve S, Hinrichs C, Feenders C, Blasius B, Schomburg D, Rabus R. Amino Acid and Sugar Catabolism in the Marine Bacterium Phaeobacter inhibens DSM 17395 from an Energetic Viewpoint. Appl Environ Microbiol. 2019 Nov 27;85(24):e02095-19.

- Zhao, Z., Amano, C., Reinthaler, T., Baltar, F., Orellana, M. V., & Herndl, G. J. (2024c). Metaproteomic analysis decodes trophic interactions of microorganisms in the dark ocean. *Nature Communications*, *15*(1), 6411.

- Zhao, Z., Amano, C., Reinthaler, T., Orellana, M. V., & Herndl, G. J. (2024b). Substrate uptake patterns shape niche separation in marine prokaryotic microbiome. *Science Advances*, *10*(20), eadn5143.

- Zhao, Z., Baltar, F., & Herndl, G. J. (2024a). Decoupling between the genetic potential and the metabolic regulation and expression in microbial organic matter cleavage across microbiomes. *Microbiology Spectrum*, *12*(5), e03036-23.