Brief Report

An optimized metaproteomics protocol for a holistic taxonomic and functional characterization of microbial communities from marine particles

Doreen Schultz,[1] Daniela Zühlke,[1] Jörg Bernhardt,[1]
Thomas Ben Francis,[2] Dirk Albrecht,[1]
Claudia Hirschfeld,[1] Stephanie Markert[3] and
Katharina Riedel [iD][1]*
[1]*Institute of Microbiology, University of Greifswald, Greifswald, Germany.*
[2]*Max-Planck Institute for Marine Microbiology, Bremen, Germany.*
[3]*Institute of Pharmacy, University of Greifswald, Greifswald, Germany.*

Summary

This study aimed to establish a robust and reliable metaproteomics protocol for an in-depth characterization of marine particle-associated (PA) bacteria. To this end, we compared six well-established protein extraction protocols together with different MS-sample preparation techniques using particles sampled during a North Sea spring algae bloom in 2009. In the final optimized workflow, proteins are extracted using a combination of SDS-containing lysis buffer and cell disruption by bead-beating, separated by SDS-PAGE, in-gel digested and analysed by LC–MS/MS, before MASCOT search against a metagenome-based database and data processing/visualization with the *in-house*-developed bioinformatics tools *Prophane* and *Paver*. As an application example, free-living (FL) and particulate communities sampled in April 2009 were analysed, resulting in an as yet unprecedented number of 9354 and 5034 identified protein groups for FL and PA bacteria, respectively. Our data suggest that FL and PA communities appeared similar in their taxonomic distribution, with notable exceptions: eukaryotic proteins and proteins assigned to *Flavobacteriia*, *Cyanobacteria*, and some proteobacterial genera were found more abundant on particles, whilst overall proteins belonging to *Proteobacteria* were more dominant in the FL fraction. Furthermore, our data points to functional differences including proteins involved in polysaccharide degradation, sugar- and phosphorus uptake, adhesion, motility, and stress response.

Introduction

A 20% of marine bacteria lives attached to algae or marine particles (Azam *et al*., 1983). These marine particles consist of various kinds of organic matter, i.e. dead/dying zoo- or phytoplankton, bacterioplankton, as well as inorganic small particles held together by a sugary matrix consisting of polysaccharide-composed transparent extracellular particles (TEPs) composed of polysaccharides, which are exuded mostly by phytoplankton but also bacteria (Alldredge *et al*., 1993). These particle-associated (PA) microbial communities have adapted to strive and survive in marine environments. Whilst some bacteria are only loosely associated with algae, others colonize algal surfaces (Grossart, 1999), where they form commensalistic or symbiotic communities with their host or even forage on algae (Sohn *et al*., 2004; Amin *et al*., 2012). Marine particles grow while sinking and thus contribute largely to the 'biological pump' by transporting carbon to deeper waters and sediments (Volkman and Tanoue, 2002). These aggregates may reach several centimetres in diameter. They are enzymatically well equipped to metabolize high-molecular weight substrates, thus providing nutrition to the attached community as well as leaving smaller carbon compounds to the surrounding water column community (Simon *et al*., 2002; Grossart, 2010).

About one decade ago, scientists started to link molecular systems biology of microorganisms to ecosystem level processes (e.g. reviewed in the study by Raes and Bork, 2008). Metagenomic studies provide valuable knowledge about diversity and distribution of microorganisms in natural environments. Moreover, metatranscriptomics and metaproteomics approaches were established to investigate, which genes are

*For correspondence. E-mail riedela@uni-greifswald.de; Tel. (+49) 3834 420 5900; Fax 49 (0) 3834 420 5902

expressed at a given time point and which proteins are particularly abundant in complex biological systems. Metaproteomics has meanwhile widely proven its potential to revisit microbial ecology concepts by linking genetic and functional diversity in microbial communities and relating taxonomic and functional diversity to ecosystem stability (Schneider and Riedel, 2010). Numerous studies, describing large-scale proteome analyses of acid-mine drainage (AMD) biofilms (Ram *et al*., 2005), wastewater treatment plants (Wilmes *et al*., 2008), and fresh-water stream biofilms (Hall *et al*., 2012) have demonstrated the power of metaproteomics to unveil molecular mechanisms involved in function, physiology, and evolution of surface-associated aquatic microbial communities. Marine metaproteomics has meanwhile been widely applied (Wang *et al*., 2014; Saito *et al*., 2019), in particular in habitats such as ocean scale shifts (Morris *et al*., 2010), the Atlantic (Bergauer *et al*., 2018) or Antarctic oceans (Williams *et al*., 2012), e.g. to investigate *Roseobacter* clade (Christie-Oleza and Armengaud 2015) and bacterioplankton (e.g. Wöhlbrand *et al*., 2017a) physiology. Teeling *et al*. (2012) studied the bacterioplankton response to a diatom bloom in the North Sea by an integrated meta-omics approach employing metagenomics and metaproteomics and provided strong evidence that distinct free-living (FL) populations of *Bacteroidetes*, *Gammaproteobacteria*, and *Alphaproteobacteria* specialize in a successive decomposition of algal-derived organic matter. As mentioned above, a significant fraction of decaying algal biomass is, however, mineralized by heterotrophic bacteria living on particles, which process a large fraction of the biosynthesized organic matter (Azam, 1998) and are thus greatly contributing to large-scale carbon fluxes (Bauer *et al*., 2006).

So far, the majority of the published studies focused on FL bacterioplankton, thereby leaving the PA bacterial communities largely unexplored. This is mainly due to the high complexity of PA samples, the presence of DNA/protein-binding polysaccharides and process-interfering substances as well as a lack of (meta)genomic information (e.g. Wöhlbrand *et al*., 2017b) although information on marine metagenomes is constantly growing (reviewed by Mineta and Gojobori 2016; Alma'abadi *et al*., 2015). Previous experiments also indicate that a high abundance of eukaryotic proteins contributes to these challenges (Smith *et al*., 2017; Saito *et al*., 2019).

Our goal was therefore to establish a robust and reproducible metaproteomics protocol enabling in-depth analyses of marine particles. For this purpose, we tested different established protocols for their applicability for protein extraction from PA bacteria in order to unravel the PA community's specific contribution to polysaccharide decomposition in marine habitats. We hypothesize that these communities express specific genes to adapt to the sessile life style and to the availability of specific polysaccharides (as observed by Ganesh *et al*., 2014).

Results and discussion

Establishment of a metaproteomics pipeline for PA microbial communities

As stated above, metaproteomic analyses of PA microbial communities are severely hampered by their high complexity, the presence of a large proportion of eukaryotic proteins, the sugary particle-matrix as well as the lack of (meta)genomic information on PA-specific pro- and eukaryotes (Wöhlbrand *et al*., 2017b; Saito *et al*., 2019). Whilst the metaproteomics analyses by Teeling *et al*. (2012) and Kappelmann *et al*. (2019) of FL bacterioplankton (harvested on 0.2 μm filters) sampled during spring blooms from 2009 to 2012 off the German island Helgoland (54°11′03″N, 7°54′00″E) resulted in the identification of several thousand protein groups, the PA microbial communities retained on 3 and 10 μm pore-sized filters were notably more difficult to analyse by the integrated metagenomic/metaproteomic approach employed at that time. For sample preparation, 500 l raw seawater were subjected to a fractionating filtration through 10 μm, 3 μm (PA fractions) and 0.2 μm (FL fractions) pore size filters (for further information about the sampling process please refer to the supporting information of Teeling *et al*. 2012). Everything that is retained on a 3 μm pore size filter is regarded as marine particles in the following.

Protein extraction. Efficient protein extraction is a crucial step for successful metaproteomics analyses of microbial communities. In a first step, we therefore tested five different protein extraction methods that employ different strategies and that were already successfully applied for metaproteome analyses of microbial communities from different environments, i.e. sewage sludge (phenol extraction; Kuhn *et al*., 2011), leaf litter (SDS-TCA; Schneider *et al*., 2012), stream hyporheic biofilms (SDS-acetone; Hall *et al*., 2012), hypersaline microbial mats (bead beating; Moog, 2012), and soil (freezing and thawing; Thompson *et al*., 2008; Chourey *et al*., 2010). In addition, the commercially available TRI-Reagent® (Sigma-Aldrich) for simultaneous isolation of RNA, DNA and proteins was tested (Table 1 and Supporting Information Fig. S1A). Filter samples used for protocol evaluation originated from several sampling events in the 2009 spring bloom sampling campaign (two to four sampling events, Supporting Information Table S1). Total protein amounts extracted from the filters by each of the applied methods were quite variable (Table 1; Supporting Information Table S1 and Fig. S1B). Highest protein yield as determined by the Pierce™ BCA Protein Assay and 1D SDS-PAGE was obtained using the SDS-acetone or bead beating approach (Table 1 and Supporting Information Fig. S1B). In conclusion, SDS-acetone- and bead beating-based protocols turned out to be most efficient for protein extraction from

Table 1. Comparison of the six tested protein extraction protocols.

Protocol Nr.	1 - Phenol	2 - SDS-TCA	3 - TRI-reagent®	4 - Freeze and thaw	5 - SDS-acetone	s6 - Bead beating
References	Kuhn et al. (2011)	Schneider et al. (2012)	Sigma-Aldrich	Chourey et al. (2010) and Thompson et al. (2008)	Hall et al. (2012)	Moog (2012)
Originally used for	Sewage sludge from biomembrane reactors	Leaf litter	Simultaneous extraction of RNA, DNA, and proteins	Soil	Stream hyporheic biofilms	Hypersaline microbial mats
Composition of the protein extraction buffer	0.1 M NaOH	1% (w/v) SDS, 50 mM Tris/HCl, pH 7	TRI-Reagent® (guanidine thiocyanate and phenol monophase solution)	5% (w/v) SDS, 50 mM Tris/HCl, 0.1 mM EDTA, 0.15 M NaCl, 1 mM $MgCl_2$, 50 mM DTT, pH 8.5	1% (w/v) SDS, 50 mM Tris/HCl, pH 6.8	5% (w/v) SDS, 0.05 mM Tris/HCl, 0.1 M DTT, 0.01 M EDTA, 10% (v/v) glycerol, 1.7 mM PMSF, pH 6.8
Cell disruption methodology	Sonication 3 × 30 s (20% power output)*	Sonication 3 × 40 s (20% power output)*	TRI-Reagent®	Two freeze and thaw cycles (liquid nitrogen, rt), 10 min boiling	Sonication 5 × 1 min (20% power output), 15 min boiling, procedure repeated on the pellet	FastPrep® 6.5 m/s, 4 × 30 s
Additional protein purification	phenol extraction (2×)	/	Chloroform extraction, ethanol extraction	/	/	/
Protein precipitation	0.1 M ammonium-acetate in methanol (1:5)**	10% TCA	2-propanol (1:1.5)**	25% TCA	acetone (1:5)**	acetone (1:4)**
Mean total protein amount (μg) 3–10 μm fraction	8.9	22.7	16.1	25.2	38.6	27.3
Mean total protein amount (μg) ≥ 10 μm fraction	8.8	12.8	38.2	24.3	102.1	114.2

* Sonopuls HD2200 Bandelin electronic, Berlin, Germany.
** ratio samle: precipitant.

particles and were therefore used for optimizing the downstream MS sample preparation procedure.

MS sample preparation. Total protein was extracted by the SDS-acetone and bead beating method from filters collected on 28 April 2009 and separated by 1D SDS-PAGE (Supporting Information Fig. S1B). Even though MS sample preparation via GeLC–MS/MS is more time-consuming compared to 1D or 2D-LC approaches, it has proven valuable to purify protein extracts and remove polymeric contaminants (e.g. Lassek *et al.* 2015; Keiblinger and Riedel, 2018) and yields comparable results as LC-based peptide fractionation (Hinzke *et al.*, 2019). To determine whether an increase in the total number of individual gel sub-fractions will lead to more protein identifications, gel lanes (two technical replicates for each protocol) were cut in either 10 or 20 equally-sized fractions, proteins were *in gel* trypsin-digested and the resulting peptides were subjected to LC–MS/MS analysis. Moreover, we tested whether reduction and alkylation of the proteins prior to tryptic digestion increased protein identification rates (Supporting Information Fig. S1B). Searching the acquired spectra in the so far available 0.2 μm 2009 (MIMAS) database (Teeling *et al.*, 2012) revealed that the best results (Supporting Information Fig. S1B and Fig. S2) were obtained by higher fractionation (20 gel pieces) without reduction and alkylation.

Optimizing databases. Metagenomic sequencing, assembly and annotation of FL (0.2 μm pore-sized filters) and PA (3 and 10 μm pore-sized filters) fractions of water samples collected during the Helgoland spring bloom 2009 was performed in parallel to the optimization of the metaproteomics protocol (for details see Supplemental Experimental Procedures). Unfortunately, most probably due to the high amount of eukaryotic DNA, length and number of assembled sequences of the large particulate fraction (10 μm pore-sized filters) was not sufficient for a valid data interpretation. Thus, the metagenomic database used for subsequent database searches was only composed of sequences of FL bacteria (0.2 μm pore-sized filters) and microbial communities present in the medium particulate fraction (3 μm pore-sized filters).

The LC–MS/MS spectra obtained with the bead beating protocol were searched against four different databases to identify the database that results in the highest number of reliably identified protein groups (Supporting Information Fig. S3): (i) the non-redundant NCBI database (NCBInr), (ii) a database with Uniprot sequences from abundant bacteria and diatoms identified by Teeling *et al.* (2012) (PABD), (iii) the database used by Teeling *et al.* (2012) containing proteins based on translated metagenomes of FL bacteria (0.2 μm pore-sized filters from different sampling time points) of the spring bloom 2009 (MIMAS) and (ivV) a database based on the metagenomes of the 0.2 and 3 μm pore-sized filters from samples of the 14 April 2009 (0.2 + 3 μm 2009). Best results were obtained with the 0.2 + 3 μm 2009 database (Supporting Information Fig. S3), which is not surprising as the resolving power of metaproteome analyses relies heavily upon the database used for protein identification (e.g. Schneider and Riedel, 2010; Teeling *et al.*, 2012). It is, moreover, well accepted that metaproteomic data are most

informative in combination with complementary omics approaches, i.e. genomics and transcriptomics (e.g. Banfield *et al.*, 2005; Ram *et al.*, 2005).

Since the bead beating-based protocol resulted in more reproducible protein yields compared to the SDS-acetone extraction protocol (Supporting Information Fig. S1B), was less time-intensive and resulted in the identification of the highest number of unique protein groups, bead beating was used for protein extraction in all subsequent analyses. A possible explanation for the efficiency of the chosen protocol is the effective disintegration of the particulate matrix by EDTA added to the extraction buffer (Passow, 2002).

Application example – comparative metaproteome analyses of FL and PA bacterioplankton

To evaluate whether the optimized protocol is suitable for a comparative metaproteomic analysis of FL and PA microbial communities, the procedure was applied to several fractions of a microbial community sampled in April 2009, i.e. 0.2–3 μm (= FL), 3–10 μm and ≥ 10 μm (= PA) fractions. Five technical replicates of each sample were subjected to the final optimized workflow (Fig. 1) and the resulting MS/MS-data were searched against the matching metagenome-based database (0.2 + 3 μm 2009). Employing our optimized pipeline, we were able to record 360 000–460 000 spectra per technical replicate and about 20 000 spectra per gel fraction, which subsequently led to the identification of 9354 protein groups (19.4% of spectral IDs; 89 240 out of 460 000), 2263 protein groups (10.2% of spectral IDs; 36 720 out of 360 000), and 2771 protein groups (10.7% of spectral IDs; 47 080 out of 440 000) for the 0.2–3 μm (Supporting Information Table S2), 3–10 μm (Table S3), and ≥ 10 μm (Table S4) fractions respectively. This is, at least to our knowledge, the largest number of protein groups ever identified for marine particles. Comparable studies addressing metaproteomic analyses of marine sediments of the Bering Sea (Moore *et al.*, 2012), the coastal North Sea, and the Pacific Ocean (Wöhlbrand *et al.*, 2017b) identified less than 10% of the protein identification numbers resulting from the here presented novel metaproteomic pipeline.

About 1956 of the identified protein groups of the two PA fractions were also identified in the FL fraction and only 276 proteins were exclusively found in the PA fractions (Supporting Information Fig. S4). This suggests that protein expression profiles of planktonic and particulate bacteria vary less than expected. However, this might also be due to the fact that PA bacteria are known to be tychoplanktic and, e.g. as offspring cells searching for a place to settle, may thus only temporarily be part of the planktonic community (Ghiglione *et al.*, 2007; Grossart, 2010; Crespo *et al.*, 2013). Moreover, clogging of filter pores by particles may cause retention of FL bacteria
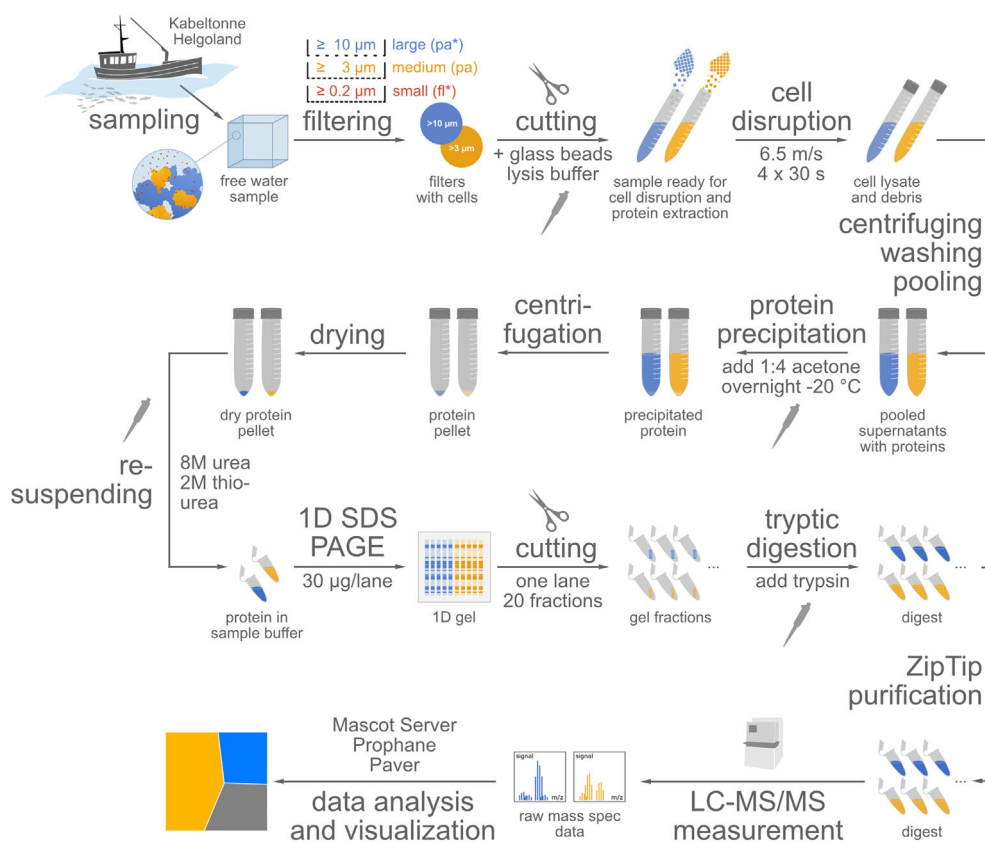
Fig. 1. Final metaproteomics protocol. Protein extraction from filters was conducted using 5% (w/v) SDS-containing lysis buffer, cell disruption by FastPrep-mediated bead beating, separation of proteins by 1D-SDS-PAGE, tryptic in-gel digestion, LC–MS/MS analyses on an Orbitrap Velos™ mass spectrometer, MASCOT database search against the metagenome-based database (0.2 + 3 μm 2009) and data-processing and visualization with the *in-house*-developed bioinformatics tools *Prophane 3.1* and *Paver*.

thus contaminating the PA fractions by planktonic bacteria.

Taxonomic differences between FL and PA bacterioplankton. Besides the similarity of the FL and PA metaproteomic data sets, the phylogenetic assignment of the identified protein groups indicated some notable taxonomic differences between the FL and PA fractions (Fig. 2 and Supporting Information Table S5).

As expected, PA fractions contained considerably more eukaryotic proteins than the FL fraction (Fig. 2A). These proteins comprised 43% (3–10 μm fraction) and 54% (≥ 10 μm fraction) of the protein groups identified from particles but contributed only 11% to the protein groups identified in the planktonic fraction. Many of these proteins were assigned to known phytoplankton taxa (Supporting Information Fig. S5) reflecting the ongoing phytoplankton bloom. Moreover, abundant proteins were also assigned to Oomycetes (water moulds, e.g. Peronosporales, Saprolegniales) and Fungi (e.g. Cryptomycota), indicating that saprotrophic or parasitic eukaryotes (Jones *et al.*, 2011; Nigrelli and Thines, 2013) may contribute to phytoplankton biomass degradation during algal blooms (Supporting Information Fig. S5). Notably, the number of viral protein groups was found to be almost three times higher in the two particulate fractions when compared

to their planktonic counterpart (Fig. 2A). The most abundant bacterial phyla within both, the FL and PA fractions, were *Proteobacteria* (FL 55%; PA 41% and 39%, 3 μm and 10 μm pore-sized filters) and *Bacteroidetes* (FL 40%; PA 48% and 47%, 3 μm and 10 μm pore-sized filters). In the study of Teeling *et al.* (2012), proteins derived from planktonic samples were mostly assigned to *Bacteroidetes*, *Alphaproteobacteria* and *Gammaproteobacteria*. The most prominent clades within the *Bacteroidetes* were *Flavobacteria* and, to a lower extent, *Ulvibacter*, *Formosa* and *Polaribacter*. *Gammaproteobacteria* were dominated by the clades *Reinekea* and *Roseobacter*. These findings are in good accordance with our results obtained for the planktonic fraction. Proteins expressed by *Alphaproteobacteria*, *Betaproteobacteria* and *Gammaproteobacteria* were generally more dominant in the FL bacteria, whilst proteins assigned to *Cyanobacteria* (e.g. *Synechococcus*, *Arthrospira*), *Opitutae*, *Flavobacteriia* (e.g. *Arenitalea*, *Olleya*, *Algibacter*, *Lacinutrix*), and some proteobacterial genera (e.g. *Oceanicoccus*, *Candidatus* Puniceispirillum, *Neptuniibacter*, *Halioglobus*, *Ramlibacter*) were more abundant in the PA fraction (Fig. 2B). This is in good accordance to other studies, which reported *Bacteroidetes* in both, FL and PA, bacterioplankton (DeLong *et al.*, 1993; Eilers *et al.*, 2001; Abell and Bowman, 2005; Alonso *et al.*, 2007). Moreover, *Flavobacteriia* have been found highly abundant during
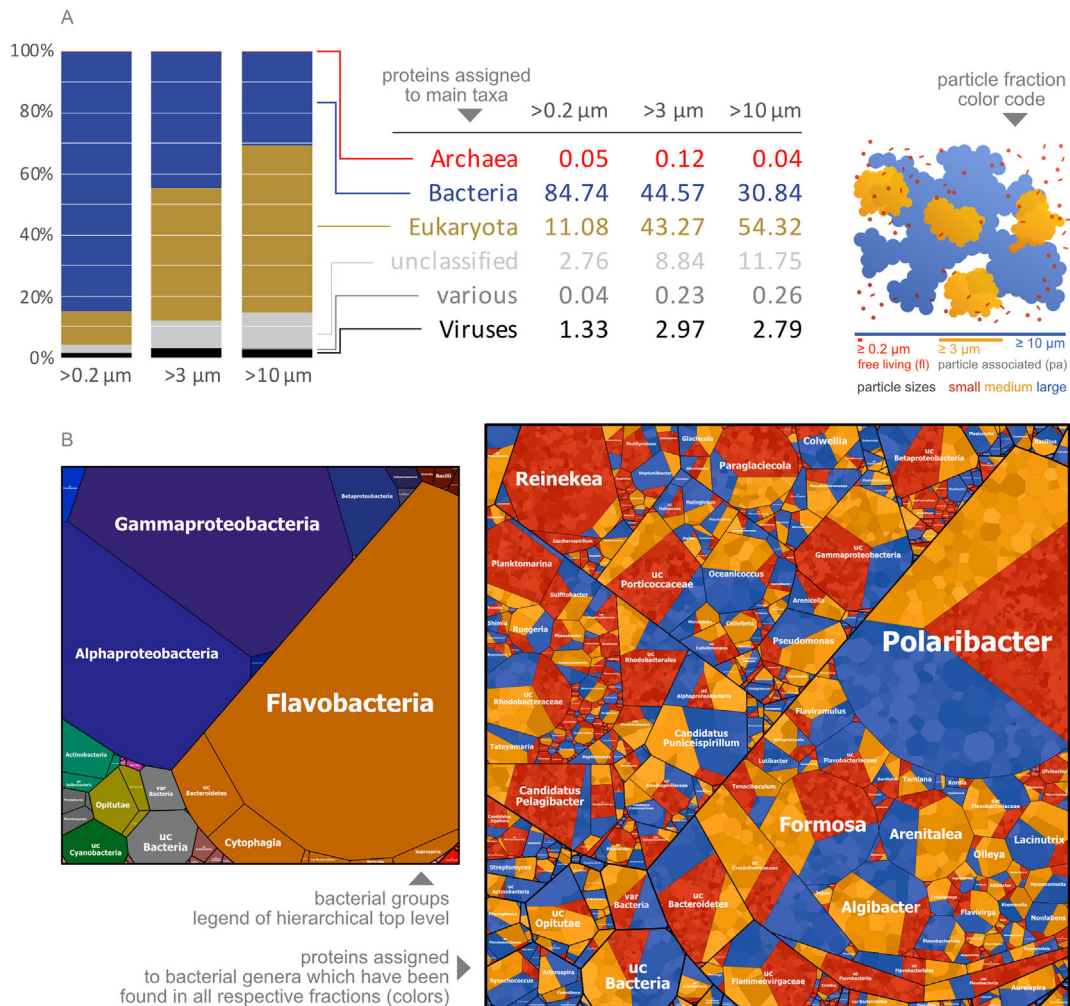
Fig. 2. Taxonomic affiliation of proteins of FL and PA metaproteomes during the spring bloom on 14 April 2009 at 'Kabeltonne' Helgoland. A. Distribution of pro- and eukaryotes in the FL (0.2–3 μm) and PA (3–10 μm, ≥ 10 μm) fractions based on the relative abundance of protein groups assigned to the different phylogenetic groups. B. Voronoi tree maps visualizing the phylogenetic assignment of bacterial protein groups identified in FL (red) and PA (yellow and blue) fractions. Cell size corresponds to the relative abundance of the respective bacterial genus on protein level. Proteins of *Reinekea* for example are most abundant in the FL fraction and are therefore encoded by a large red tree map cell. In the PA fractions they can be detected only in traces, resulting in very small cell sizes (coloured in yellow and blue). *Algibacter* protein abundance, on the other hand, was notably higher in the PA fractions, compared to the FL fraction.

phytoplankton blooms indicating that they play an important role as consumers of algal-derived organic matter (Simon *et al*., 1999; Riemann *et al*., 2000; Pinhassi *et al*., 2004; Grossart *et al*., 2005; Teeling *et al*., 2016; Chafee *et al*., 2018).

Functional differences between FL and PA bacterioplankton. Notably, differences in the protein profiles between FL and PA bacteria seemed more evident on the functional level (Fig. 3). Most importantly, the SusC/D utilization system, specific glycoside hydrolases, i.e. GH family 1, 13, and 16 (including beta-glucosidases, alpha-1,4-amylases, and exo- and endo-1,3-beta-glycanases), glycosyl transferases and TonB-dependent transporters were found with higher overall expression levels in the PA fractions compared to the FL fraction (Fig. 3A). This is in good accordance with the high substrate availability (Caron *et al*., 1982; Grossart *et al*., 2003; Fernández-Gómez *et al*., 2013), especially the presence of

highly abundant microalgae storage polysaccharides, i.e. alpha- and beta-glucans (Kroth *et al*., 2008), in the particles. Sulfatases, capable of cleaving sulfate sugar ester bonds, are contributing to the degradation of specific sulphated algal polysaccharides such as mannans and fucans (Gómez-Pereira *et al*., 2012). This is well supported by our finding that sulfatases are strongly expressed by PA *Flavobacteriia*, especially *Formosa* sp. (Fig. 3A and B).

Moreover, our data indicate that FL and PA seem to employ different strategies for phosphate acquisition, stress response as well as adhesion and motility (for further information, please see the Supporting Information).

Conclusions and outlook

In this study, we present a broadly applicable metaproteomics protocol for successful extraction of proteins
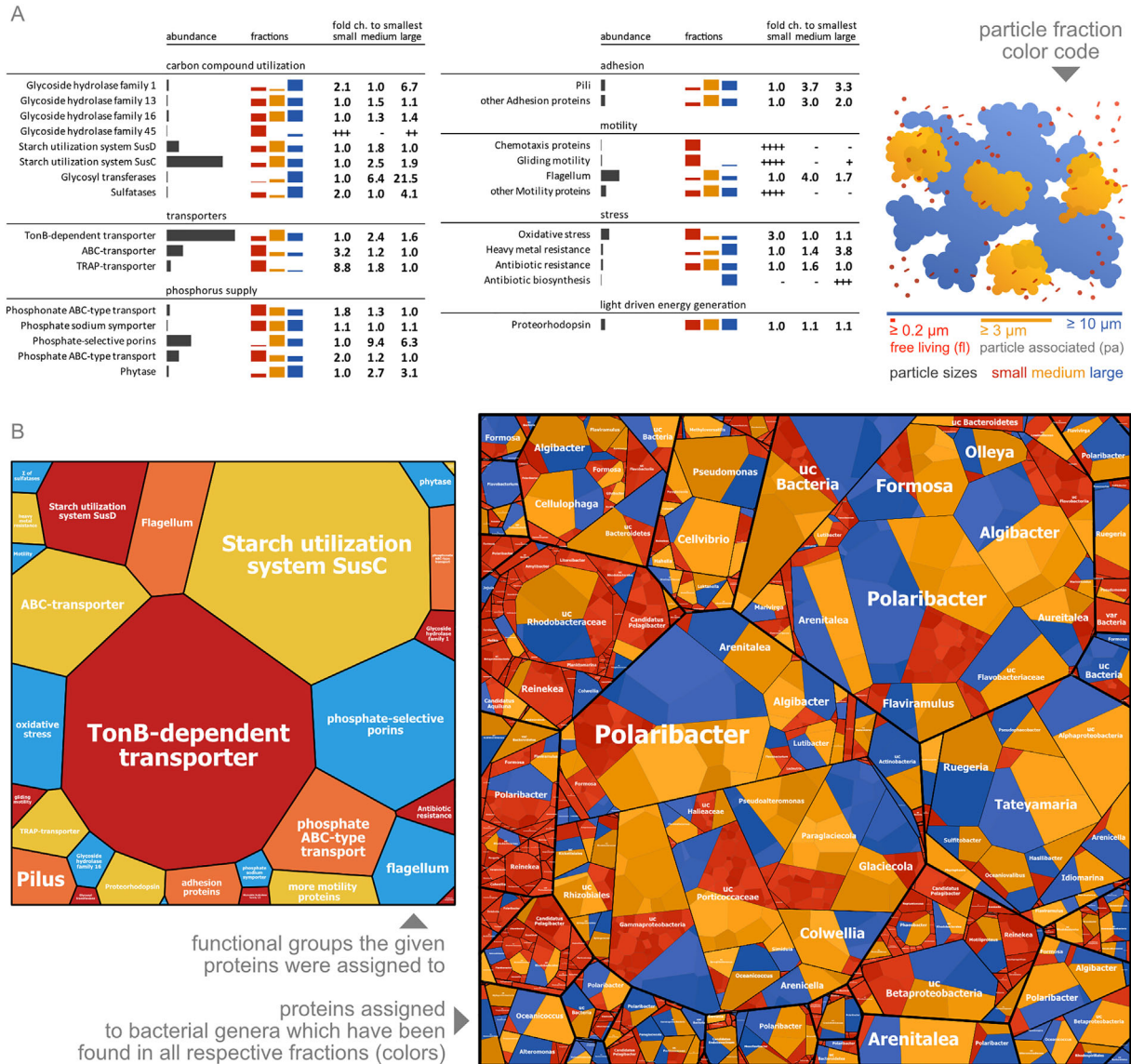
Fig. 3. Functional assignment of proteins in FL and PA metaproteomes during the spring bloom on 14 April 2009 at 'Kabeltonne' Helgoland. A. Total (black bars) and relative (red for FL, yellow and blue for PA) abundance of selected protein groups with assigned functions in the FL (0.2–3 μm = small) and PA (3–10 μm = medium and ≥ 10 μm = large) fractions. B. Voronoi tree maps showing the phylogenetic assignment of selected functional protein groups identified in FL (red) and PA (yellow and blue) fractions. Cell size corresponds to the relative abundance of the respective genus within specific functional categories.

from marine particles. Our comparative metaproteomic analyses of marine microbial communities living either planktonically or attached to particles resulted in an as yet unequalled number of identified protein groups for marine particles and gave first insights into the expression of life style-specific functions of bacteria living on particles.

Although our optimized metaproteomic workflow significantly improved the identification rate of PA proteins, the number of protein identifications from the particles is still considerably lower compared to FL bacterial

communities. We assume that especially the high abundance of eukaryotic proteins poses problems in protein identification due to the complexity and diversity of microbial eukaryote genomes and the presence of introns and repeats in the metagenomic DNA sequence databases, which hinders peptide identification (Saito *et al.*, 2019). Metaproteome coverage of marine particles could be significantly improved by employing customized databases including eukaryotic metatranscriptomic (RNA-based) sequence data (Keeling *et al.*, 2014). This can be achieved by generating metatranscriptomes from the

particular fractions. Alternatively, protein identification could also be substantially improved by extracting already existing metatranscriptomic and metagenomic data from relevant eukaryotic taxa from public databases. Key to the latter approach is reliable information on which eukaryotic organisms make up the particles, which can be attained by 18S rRNA gene amplicon sequencing. Perspectively, we will extend our analyses on eukaryotic taxa and analyse multiple time points during phytoplankton blooms to investigate succession of taxonomical clades and expressed functions of marine particles from pre-bloom to post-bloom conditions.

Acknowledgements

References

Abell, G.C.J., and Bowman, J.P. (2005) Colonization and community dynamics of class Flavobacteria on diatom detritus in experimental mesocosms based on Southern Ocean seawater. *FEMS Microbiol Ecol* **53**: 379–391.

Alma'abadi, A.D., Gojobori, T., and Mineta, K. (2015) Marine metagenome as a resource for novel enzymes. *GPB* **13**: 290–295.

Alldredge, A.L., Passow, U., and Logan, B. (1993) The existence, abundance, and significance of large transparent exopolymer particles in the ocean. *Deep-Sea Res I* **40**: 1131–1140.

Alonso, C., Warnecke, F., Amann, R., and Pernthaler, J. (2007) High local and global diversity of Flavobacteria in marine plankton. *Environ Microbiol* **9**: 1253–1266.

Amin, S.A., Parker, M.S., and Armbrust, E.V. (2012) Interactions between diatoms and bacteria. *MMBR* **76**: 667–684.

Azam, F., Fenchel, T., Field, J.G., Gray, J.S., Meyer-Reil, L. A., and Thingstad, F. (1983) The ecological role of water-column microbes in the sea. *Mar Ecol Prog Ser* **10**: 257–263.

Azam, F. (1998) Microbial control of oceanic carbon flux: the plot thickens. *Science* **280**: 694–696.

Banfield, J.F., Verberkmoes, N.C., Hettich, R.L., and Thelen, M.P. (2005) Proteogenomic approaches for the molecular characterization of natural microbial communities. *OMICS* **9**: 301–333.

Bauer, M., Kube, M., Teeling, H., Richter, M., Lombardot, T., Allers, E., *et al*. (2006) Whole genome analysis of the marine *Bacteroidetes* '*Gramella frosetii*' reveals adaptations to degradation of polymeric organic matter. *Environ Microbiol* **8**: 2201–2213.

Bergauer, K., Fernandez-Guerra, A., Garcia, J.A.L., Sprenger, R.R., Stepanauskas, R., Pachiadaki, M.G., *et al*. (2018) Organic matter processing by microbial communities throughout the Atlantic water column as revealed by metaproteomics. *Proc Natl Acad Sci U S A* **115**: E400–E408.

Caron, D.A., Davis, P.G., Madin, L.P., and Sieburth, J.M. (1982) Heterotrophic bacteria and bacterivorous protozoa in oceanic macroaggregates. *Science* **218**: 795–797.

Chafee, M., Fernàndez-Guerra, A., Buttigieg, P.L., Gerdts, G., Eren, A.M., Teeling, H., and Amann, R.I. (2018) Recurrent patterns of microdiversity in a temperate coastal marine environment. *ISME J* **12**: 237–252.

Chourey, K., Jansson, J., VerBerkmoes, N., Shah, M., Chavarria, K.L., Tom, L.M., *et al*. (2010) Direct cellular lysis/protein extraction protocol for soil metaproteomics. *J Proteome Res* **9**: 6615–6622.

Christie-Oleza, J.A., and Armengaud, J. (2015) Proteomics of the Roseobacter clade, a window to the marine microbiology landscape. *Proteomics* **15**: 3928–3942.

Crespo, B.G., Pommier, T., Fernández-Gómez, B., and Pedrós-Alió, C. (2013) Taxonomic composition of the particle-attached and free-living bacterial assemblages in the Northwest Mediterranean Sea. *Microbiol Open* **2**: 541–552.

DeLong, E.F., Franks, D.G., and Alldredge, A.L. (1993) Phylogenetic diversity of aggregate-attached vs. free-living marine bacterial assemblages. *Limnol Oceanogr* **38**: 924–934.

Eilers, H., Pernthaler, J., Peplies, J., Glöckner, F.O., Gerdts, G., and Amann, R. (2001) Isolation of novel pelagic bacteria from the German bight and their seasonal contribution to surface picoplankton. *AEM* **67**: 5134–5142.

Fernández-Gómez, B., Richter, M., Schüler, M., Pinhassi, J., Acinas, S., González, J.M., *et al*. (2013) Ecology of marine Bacteroidetes. A comparative genomics approach. *ISME J* **7**: 1026–1037.

Ganesh, S., Parris, D.J., DeLong, E.F., and Stewart, F.J. (2014) Metagenomic analysis of size-fractionated picoplankton in a marine oxygen minimum zone. *ISME J* **8**: 187–211.

Ghiglione, J.F., Mevel, G., Pujo-Pay, M., Mousseau, L., Lebaron, P., and Goutx, M. (2007) Diel and seasonal variations in abundance, activity, and community structure of particle-attached and free living bacteria in NW Mediterranean Sea. *Microb Ecol* **54**: 217–231.

Gómez-Pereira, P.R., Schüler, M., Fuchs, B.M., Bennke, C., Teeling, H., Waldmann, J., *et al*. (2012) Genomic content of uncultured Bacteroidetes from contrasting oceanic provinces in the North Atlantic Ocean. *Environ Microbiol* **14**: 52–66.

Grossart, H.P. (1999) Interactions between marine bacteria and axenic diatoms (*Cylindrotheca fusiformis*, *Nitzschia laevis*, and *Thalassiosira weissflogii*) incubated under various conditions in the lab. *Aquat Microb Ecol* **19**: 1–11.

Grossart, H.P., Hietanen, S., and Ploug, H. (2003) Microbial dynamics on diatom aggregates in Oresund, Denmark. *Mar Ecol Prog Ser* **249**: 69–78.

Grossart, H.P., Levold, F., Allgaier, M., Simon, M., and Brinkhoff, T. (2005) Marine diatom species harbour distinct bacterial communities. *Environ Microbiol* **7**: 860–873.

Grossart, H.P. (2010) Ecological consequences of bacterioplankton lifestyles: changes in concepts are needed. *Environ Microbiol Rep* **2**: 706–714.

Hall, E.K., Besemer, K., Kohl, L., Preiler, C., Riedel, K., Schneider, T., *et al*. (2012) Effects of resource chemistry

on the composition and function of stream hyporheic biofilms. *Front Microbiol* **3**: 1–14.

Hinzke, T., Kouris, A., Hughes, R.A., Strous, M., and Kleiner, M. (2019) More is not always better: evaluation of 1D and 2D-LC-MS/MS methods for Metaproteomics. *Front Microbiol* **10**: 238.

Jones, M.D.M., Forn, I., Gadelha, C., Egan, M.J., Bass, D., Massana, R., and Richards, T.A. (2011) Discovery of novel intermediate forms redefines the fungal tree of life. *Nature* **474**: 200–203.

Kappelmann, L., Krüger, K., Hehemann, J.H., Harder, J., Markert, S., Unfried, F., *et al*. (2019) Polysaccharide utilization loci of North Sea *Flavobacteriia* as basis for using SusC/D-protein expression for predicting major phytoplankton glycans. *ISME J* **13**: 76–91.

Keeling, P.J., Burki, F., Wilcox, H.M., Allam, B., Allen, E.E., Amaral-Zettler, L.A., *et al*. (2014) The marine microbial eukaryote transcriptome sequencing project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biol* **12**: e1001889–e1001886.

Keiblinger, K.M., and Riedel, K. (2018) Sample preparation for metaproteome analyses of soil and leaf litter. *Methods Mol Biol* **1841**: 303–318.

Kroth, P.G., Chiovitti, A., Gruber, A., Martin-Jezequel, V., Mock, T., Schnitzler Parker, M., *et al*. (2008) A model for carbohydrate metabolism in the diatom *Phaeodactylum tricornutum* deduced from comparative whole genome analysis. *PLoS One* **3**: e1426.

Kuhn, R., Benndorf, D., Rapp, E., Reichl, U., Palese, L.L., and Pollice, A. (2011) Metaproteome analysis of sewage sludge from membrane bioreactors. *Proteomics* **11**: 2738–2744.

Lassek, C., Burghartz, M., Chaves-Moreno, D., Otto, A., Hentschker, C., Fuchs, S., *et al*. (2015) A metaproteomics approach to elucidate host and pathogen protein expression during catheter-associated urinary tract infections (CAUTIs). *Mol Cell Proteomics* **14**: 989–1008.

Mineta, K., and Gojobori, T. (2016) Databases of the marine metagenomics. *Gene* **576**: 724–728.

Moog, G. (2012) Optimierung der Probenaufbereitung für Metaproteomanalysen mittels Massenspektrometrie. Diploma thesis. University of Greifswald, Institute of Microbiology.

Moore, E.K., Nunn, B.L., Goodlett, D.R., and Harvey, H.R. (2012) Identifying and tracking proteins through the marine water column: insights into the inputs and preservation mechanisms of protein in sediments. *Geochim Cosmochim Acta* **83**: 324–359.

Morris, R.M., Nunn, B.L., Frazar, C., Goodlett, D.R., Ting, Y. S., and Rocap, G. (2010) Comparative metaproteomics reveals ocean-scale shifts in microbial nutrient utilization and energy transduction. *ISME J* **4**: 673–685.

Nigrelli, L., and Thines, M. (2013) Tropical oomycetes in the German bight - climate warming or overlooked diversity? *Fungal Ecol* **6**: 152–160.

Passow, U. (2002) Transparent exopolymer particles (TEP) in aquatic environments. *Prog Oceanogr* **55**: 287–333.

Pinhassi, J., Sala, M.M., Havskum, H., Peters, F., Guadayol, O., Malits, A., *et al*. (2004) Changes in bacterioplankton composition under different phytoplankton regimens. *AEM* **70**: 6753–6766.

Raes, J., and Bork, P. (2008) Molecular eco-systems biology: towards an understanding of community function. *Nat Rev Microbiol* **6**: 693–699.

Ram, R.J., Verberkmoes, N.C., Thelen, M.P., Tyson, G.W., Baker, B.J., Blake, R.C., *et al*. (2005) Community proteomics of a natural microbial biofilm. *Science* **308**: 1915–1920.

Riemann, L., Steward, G.F., and Azam, F. (2000) Dynamics of bacterial community composition and activity during a mesocosm diatom bloom. *AEM* **66**: 578–587.

Saito, M.A., Bertrand, E.M., Duffy, M.E., Gaylord, D.A., Held, N.A., Hervey, W.J., *et al*. (2019) Progress and challenges in ocean metaproteomics and proposed best practices for data sharing. *J Proteome Res* **18**: 1461–1476.

Schneider, T., and Riedel, K. (2010) Environmental proteomics Analysis of structure and function of microbial communities. *Proteomics* **10**: 785–798.

Schneider, T., Keiblinger, K.M., Schmid, E., Sterflinger-Gleixner, K., Ellersdorfer, G., Roschitzki, B., *et al*. (2012) Who is who in litter decomposition? Metaproteomics reveals major microbial players and their biogeochemical functions. *ISME J* **6**: 1749–1762.

Simon, M., Glöckner, F.O., and Amann, R. (1999) Different community structure and temperature optima of heterotrophic picoplankton in various regions of the Southern Ocean. *Aquat Microb Ecol* **18**: 275–284.

Simon, M., Grossart, H.P., Schweitzer, B., and Ploug, H. (2002) Microbial ecology of organic aggregates in aquatic ecosystems. *Aquat Microb Ecol* **28**: 175–211.

Smith, M.W., Herfort, L., Fortunato, C.S., Crump, B.C., and Simon, H.M. (2017) Microbial players and processes involved in phytoplankton bloom utilization in the water column of a fast-flowing, river-dominated estuary. *Microbiol Open* **6**: 4.

Sohn, J.H., Lee, J.H., Yi, H., Chun, J., Bae, T.S., Ahn, T.Y., *et al*. (2004) *Kordia algicida* gen. Nov., sp. nov., an aligicidal bacterium isolated from red tide. *Int J Syst Evol Microbiol* **54**: 675–680.

Teeling, H., Fuchs, B.M., Becher, D., Klockow, C., Gardebrecht, A., Bennke, C.M., *et al*. (2012) Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. *Science* **336**: 608–611.

Teeling, H., Fuchs, B.M., Bennke, C.M., Krüger, K., Chafee, M., Kappelmann, L., *et al*. (2016) Recurring patterns in bacterioplankton dynamics during coastal spring algae blooms. *Elife* **5**: e1188.

Thompson, M.R., Chourey, K., Froelich, J.M., Erickson, B.K., VerBerkmoes, N.C., and Hettich, R.L. (2008) Experimental approach for deep proteome measurements from small-scale microbial biomass samples. *Anal Chem* **80**: 9517–9525.

Volkman, J., and Tanoue, E. (2002) Chemical and biological studies of particulate organic matter in the ocean. *J Oceanogr* **58**: 265–279.

Wang, D.Z., Xie, Z.X., and Zhang, S.F. (2014) Marine metaproteomics: current status and future directions. *J Proteomics* **97**: 27–35.

Williams, T.J., Long, E., Evans, F., Demaere, M.Z., Lauro, F. M., Raftery, M.J., *et al.* (2012) A metaproteomic assessment of winter and summer bacterioplankton from Antarctic peninsula coastal surface waters. *ISME J* **6**: 1883–1900.

Wilmes, P., Wexler, M., and Bond, P.L. (2008) Metaproteomics provides functional insight into activated sludge wastewater treatment. *PLoS One* **3**: e1778.

Wöhlbrand, L., Wemheuer, B., Feenders, C., Ruppersberg, H.S., Hinrichs, C., Blasius, B., *et al.* (2017a) Complementary metaproteomic approaches to assess the bacterioplankton response toward a phytoplankton spring bloom in the southern North Sea. *Front Microbiol* **8**: 442.

Wöhlbrand, L., Feenders, C., Nachbaur, J., Freund, H., Engelen, B., Wilkes, H., *et al.* (2017b) Impact of extraction methods on the detectable protein complement of metaproteomic analyses of marine sediments. *Proteomics* **17**: 22–34.

Supporting Information

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

Figure S1 (A) Sampling strategy and (B) evaluation of protein extraction and MS sample preparation protocols. (A) Water samples collected at 'Kabeltonne' Helgoland during the spring bloom 2009 were sequentially filtered to obtain the 0.2–3 μm (FL) and two PA (3–10 μm = medium, ≥ 10 μm = large) fractions as described in Teeling *et al*. (). Filters were initially cut into three or four pieces, which were subsequently shredded and mixed with the respective extraction buffer. **(B)** Filters (medium particles = yellow; large particles = blue) from different sampling time points (turquoise, green and red) were processed according to the six different protocols described in the Experimental Procedure section. With regard to the extracted protein amount, the bead beating and SDS-acetone approaches obviously outcompeted the four other protocols. However, the SDS-acetone protocol was less reproducible than the bead beating protocol. Considering bead beating and SDS-acetone as best performing protocols, they were employed to test different MS sample preparation approaches, i.e., different number of SDS gel fractions for tryptic digestion together with protein reduction (red.) and alkylation (alk.) prior to tryptic digestion. The subsequent LC–MS/MS analyses revealed best results for the bead beating protocol followed by GeLC–MS/MS from 20 fractions without protein reduction and alkylation as shown in the bottom line of the figure. Bubble sizes for the large (blue) and medium (yellow) particles correspond to the number of identified protein groups (see also Fig. S2).

Figure S2 Protein identifications obtained by different extraction and protein pre-fractionation protocols. For the medium particle size fraction (3–10 μm, yellow), 20 gel fractions after standard treatment, i.e. without protein reduction (red.) and alkylation (alk.), resulted in the highest number of identified protein groups, no matter which protein extraction protocol (SDS-acetone (red) or bead beating (green)) was applied. For the large particle fraction (≥ 10 μm, blue) the general trend was similar. However, the beat beating protocol performed better compared to the SDS-acetone protocol.

Figure S3 Number of identified protein groups obtained with different databases: (I) the non-redundant NCBI database (NCBInr, 136,216,794 entries), **(II)** a database with Uniprot sequences of known abundant bacteria and diatoms identified by the study of Teeling *et al*. () (PABD, 2,638,314 entries), **(III)** a metagenome-based database employed for the FL bacterial fraction within the study of Teeling *et al*. () (MIMAS, 1,579,724 entries) and **(IV)** a database based on translated metagenomes of the FL fraction on the 0.2 μm filters and particles on the 3 μm filters sampled on the 14th of April 2009 (0.2 + 3 μm 2009, 1,463,572 entries).

Figure S4 Venn diagram showing the number of fraction-specific and shared proteins of the planktonic and particulate fractions. The FL (0.2 μm) fraction is depicted in red, the medium PA fraction (0.2–3 μm) in yellow and the large PA fraction (≥ 10 μm) in blue.

Figure S5 Phylogenetic assignment of eukaryotic proteins present in the FL and PA fractions during the spring bloom on 14th of April 2009 at 'Kabeltonne' Helgoland. (A) Distribution of different eukaryotes in the FL (0.2 μm) and PA (0.2–3 μm and ≥ 10 μm) fractions as shown by relative protein abundances assigned to the different eukaryotic phylogenetic groups. (B) Voronoi tree map visualizing the relative abundance of eukaryotic taxa based on the abundance of assigned proteins extracted from the FL (red) and PA (yellow and blue) fractions. Cell size corresponds to the relative abundance of the respective genus. In this preliminary analysis, protein identification is based on metagenomic (DNA-based) information from the filtered fractions, which suffers limitations for eukaryotic protein identification, probably resulting in incomplete functional and taxonomic profiles.

Table S1 Comparison of total protein amounts and number of biological replicates of the six tested protein-extraction protocols

Table S2 Prophane output for proteins extracted from 0.2 μm pore-sized filters

Table S3 Prophane output for proteins extracted from 3 μm pore-sized filters

Table S4 Prophane output for proteins extracted from 10 μm pore-sized filters

Table S5 Distribution of phylogenetic groups among proteins extracted from the 0.2 μm, 3 μm and 10 μm pore-sized filters.

Appendix S1: Supporting Information
Appendix S2: Supporting Information