Title: A robust metaproteomics pipeline for a holistic taxonomic and functional characterization of microbial communities from marine particles

Authors: 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 [1]*

Addresses: [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.

***Corresponding author mailing address:** Institute of Microbiology, Center for Functional Genomics of Microbes (C_FunGene), University of Greifswald, Felix-Hausdorff-Straße 8, DE-17487 Greifswald, Germany. E-mail riedela@uni-greifswald.de; Tel (+49) 3834 420 5900

Running title: Metaproteomic pipeline to analyse marine particles

Originality-Significance Statement

Marine particles consist of organic particulate matter (e.g. phyto- or zooplankton) and particle-associated (PA) microbial communities, which are often embedded in a sugary matrix. A significant fraction of the decaying algal biomass in marine ecosystems is expected to be mineralized by PA heterotrophic communities, which are thus greatly contributing to large-scale carbon fluxes. Whilst numerous studies have investigated the succession of planktonic marine bacteria along phytoplankton blooms, the community structure and functionality of PA bacterial communities remained largely unexplored and knowledge on specific contributions of these microorganisms to carbon cycling is still surprisingly limited. This has been mostly been due to technical problems, i.e. to the difficulty to retrieve genomic DNA and proteins from these polysaccharide-rich entities, their enormous complexity and the high abundance of eukaryotic microorganisms.

Our study presents an innovative, robust, reproducible, and reliable metaproteomics pipeline for marine particles, which will help to address and fill the above-described knowledge gap. Employing the here established workflow enabled us to identify more than 5,000 PA proteins, which is, at least to our knowledge, the largest number of protein groups ever assigned to marine particles. Notably, the novel pipeline has been validated by a first, comparative metaproteome analysis of free-living and PA bacterial communities indicating a significant functional shift enabling surface-associated bacteria to adapt to particle-specific living conditions. In conclusion, our novel metaproteomics pipeline presents a solid and promising methodological groundwork for future culture-independent analyses of seasonal taxonomic and functional successions of PA microbial communities in aquatic habitats.

Summary

This study aimed to establish a robust, reproducible and reliable metaproteomic pipeline 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 this 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 proof of principle, free-living (FL) and particulate communities sampled in April 2009 were analysed, resulting in an as yet unprecedented number of 9,354 and 5,034 identified protein groups for FL and PA bacteria, respectively. Our data revealed 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. In contrast, significant functional differences including proteins involved in polysaccharide degradation, sugar- and phosphorus uptake, adhesion, motility, and stress response were detected.

Introduction

The oceans provide about one half of the global net primary production (NPP) (Field *et al.*, 1998; Falkowski *et al.*, 1998; Azam and Malfatti, 2007), a large part of which is reprocessed by bacterioplankton in the so-called "microbial loop", a pathway during which dissolved organic carbon is returned to higher trophic levels via its incorporation into bacterial biomass (Azam, 1998). Notably, about 20% of the bacterioplankton 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 present one of the most ancient forms of "multicellularity" (Hall-Stoodley *et al.*, 2004) and 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 predate 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 nutrients 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 Raes and Bork, 2008). Metagenomic studies were initiated to provide valuable knowledge about diversity and distribution of microorganisms in natural environments. Moreover, metatranscriptomics and metaproteomics approaches were established to investigate, which genes are 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 (Saito *et al.,* 2019; Wang *et al.,* 2014), in particular in habitats such as ocean scale shifts (Morris *et al*., 2010), the Atlantic (Bergauer *et al*., 2017) 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* specialized 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 (Battin *et al*., 2003; Bauer *et al.*, 2006).

So far, the majority of the published studies focused on FL bacterioplankton, thereby leaving the PA bacterial communities largely unexplored. The particulate fraction of bacterioplankton has proven challenging to comprehensive meta-omics characterization, due to its high complexity, presence of DNA/protein-binding polysaccharides, process-interfering substances and lack of (meta)genomic information on marine particles (e.g. Wöhlbrand *et al*., 2017b), despite the fact that 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, reproducible and reliable pipeline enabling in-depth metaproteomics analyses of marine particles by testing different established protocols for their applicability for protein extraction from PA bacterioplankton in order to unravel the PA community's specific contribution to polysaccharide decomposition in marine habitats. As stated above, heterotrophic microbial communities are supposed to be well equipped to metabolize algal high molecular weight substrates (Simon *et al*., 2002; Grossart, 2010). We thus hypothesize that these communities either differ taxonomically from their FL counterparts (as shown by Bizic-Ionescu *et al*. (2015)) and/or express different genes to adapt to the sessile life style and the availability of specific polysaccharides such as insoluble glycan fibres (as observed by Ganesh *et al*. (2014)). Moreover, we postulated that this adaption will mostly affect the expression of proteins involved in motility, adhesion, stress response as well as CAZymes together with appropriate sugar transporters. In order to test these hypotheses, a metaproteomics pipeline for marine particles was established and tested as a proof of concept on spring bloom samples collected in April 2009, for which the FL bacterial fraction had previously been characterized (Teeling *et al*., 2012).

Results and Discussion

Establishment of a metaproteomics pipeline for PA microbial communities

As stated above, the metaproteomics analysis of PA microbial communities is 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 Teeing *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 at "Kabeltonne" Helgoland resulted in the identification of several thousand protein groups, the PA microbial communities retained on 3 and 10 µm pore-sized filters emerged as difficult to analyse by the integrated metagenomic/metaproteomic approach employed at that time.

We thus aimed to develop a suitable and effective metaproteomics pipeline for in-depth analyses of taxonomic composition and functionality of marine PA microbial communities. To this end, six different well-established protein extraction protocols for various particulate samples were tested using biomass from 3 µm and 10 µm pore-sized filters (assigned as PA fraction) of the MIMAS bio-archive (www.mimas-project.de) covering different stages of the bloom, which had been collected in 2009 during the spring phytoplankton bloom sampling campaign by Teeling *et al*. (2012). Moreover, various MS sample preparation protocols and different protein sequence databases were evaluated.

[**Table 1**]

Protein extraction. Efficient protein extraction is a crucial step for successful metaproteomics analyses of the 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; Chourey *et al*., 2010; Thompson *et al*., 2008). In addition, the commercially available TRI-Reagent® (Sigma Aldrich) for simultaneous isolation of RNA, DNA and proteins was tested (**Table 1**). Filter samples used for protocol evaluation originated from several sampling dates in the 2009 spring bloom sampling campaign (9th of February, 7th of April, 21st of April, and 16th of June 2009). Filters were cut into pieces and subjected to the respective extraction protocol (**Fig. 1A**). Total protein amounts extracted from the filters by each of the applied methods were quite variable. 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**; **Fig. 1B**). In conclusion, SDS-acetone- and bead beating-based protocols turned out to be most efficient for protein extraction from 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 28th of April 2009 and separated by 1D SDS-PAGE (**Fig. 1B**). Even though MS sample preparation via GeLC MS/MS is more time-consuming compared to 1D or 2D-LC approaches, it has been 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 were cut in either 10 or 20 equal-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 (**Fig. 1B**). Searching the acquired spectra in the so far available 0.2 µm 2009 (MIMAS) database (Teeling *et al.*, 2012) revealed that the best results (**Fig. 1B**, **Fig. S1**) were obtained by higher fractionation (20 gel pieces) without reduction and alkylation.

[**Figure 1]**

Optimizing databases. Metagenomic sequencing, assembling 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 experimental procedures). Unfortunately, the coverage and quality of the metagenome sequences of the large particulate fraction (10 µm pore-sized filters) was not sufficient to be correctly assembled, annotated and translated. 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 (**Fig. S2**): (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 (IV) a database based on the metagenomes of the 0.2 and 3 µm pore-sized filters from samples of the 14th of April 2009 (0.2 + 3 µm 2009). Best results were obtained with the 0.2 + 3 µm 2009 database (**Fig. S2)**, 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).

Based on our results, we finally decided to use the bead beating-based protocol for protein extraction since this method resulted in more reproducible protein yields compared to the SDS-acetone extraction protocol (**Fig. 1B**). Furthermore, this method was less time-intensive compared to the SDS-acetone approach and resulted in the identification of the

highest number of unique protein groups, most probably due to the effective disintegration of the particulate matrix by EDTA added to the extraction buffer (Passow, 2002). In summary, the optimized metaproteomic pipeline for marine particles (**Fig. 2**) comprises protein extraction by bead beating, protein fractionation by 1D SDS-PAGE (20 fractions), followed by in-gel tryptic digestion, LC-MS/MS and database search against the matching metagenome database (0.2 + 3 µm 2009).

[**Figure 2]**

Proof of principle – comparative metaproteome analyses of FL and PA bacterioplankton

As proof of principle, the newly established protocol was applied for a comparative metaproteomic analysis of FL and PA microbial communities sampled at the 14th of April 2009 in 0.2 µm - 3 µm (= FL), 3 - 10 µm and ≥ 10 µm (= PA) fractions. Five technical replicates of each sample were subjected to the final optimized workflow 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 460,000, 360,000, and 440,000 spectra, which subsequently led to the identification of 9,354 protein groups (19.4% of spectral IDs), 2,263 protein groups (10.2% of spectral IDs), and 2,771 protein groups (10.7% of spectral IDs) for the 0.2 - 3 µm (**Table S1**), 3 - 10 µm (**Table S2**), and ≥ 10 µm (**Table S3**) 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.

1,956 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 (**Fig. S3**). This suggests that protein expression profiles of planktonic and particulate bacteria vary less than

234 expected. However, this might also be due to the fact that PA bacteria are known to hop on
235 and off particles and, e.g. as offspring cells searching for a place to settle, may thus only
236 temporarily be part of the planktonic community (Ghiglione *et al*., 2007; Grossart, 2010;
237 Crespo *et al*., 2013). Moreover, clogging of filter pores by particles may cause retention of FL
238 bacteria thus contaminating the PA fractions by planktonic bacteria. In addition, the lack of ≥
239 10 µm pore-sized filter metagenomic sequences hampers comprehensive protein
240 identifications in this PA fraction that may result in a virtually lower abundance than
241 expected.

242 **Taxonomic differences between FL and PA bacterioplankton.** Besides the somewhat
243 unexpected similarity of the FL and PA metaproteomic datasets, the phylogenetic
244 assignment of the identified protein groups indicated some notable taxonomic differences
245 between the FL and PA fractions (**Fig. 3, Table S4**).

246 [**Figure 3**]

247 The number of eukaryotic protein groups was significantly higher in the PA fractions
248 (**Fig. 3A**) and comprised 43% and 54% of the protein groups identified for the 3 - 10 µm and
249 ≥ 10 µm fractions, respectively, in contrast to only 11% of the protein groups identified for the
250 FL phytoplankton. Moreover, the number of viral protein groups was found to be almost three
251 times higher in the two particulate fractions when compared to their planktonic counterpart
252 (**Fig. 3A**). The most abundant phyla within both, the FL and PA fractions, were
253 *Proteobacteria* (FL 55%; PA 41% and 39%, 3 µm and 10 µm pore-sized filters) and
254 *Bacteroidetes* (FL 40%; PA 48% and 47%, 3 µm and 10 µm pore-sized filters). Proteins
255 expressed by *Alpha*-, *Beta*- and *Gammaproteobacteria* were generally more dominant in the
256 FL bacteria, whilst proteins assigned to *Cyanobacteria* (e.g. *Synechococcus*, *Arthrospira*),
257 *Opitutae*, *Flavobacteriia* (e.g. *Arenitalea*, *Olleya*, *Algibacter*, *Lacinutrix*), and some
258 proteobacterial genera (e.g. *Oceanicoccus*, *Candidatus* Puniceispirillum, *Neptuniibacter*,
259 *Halioglobus*, *Ramlibacter*) were more abundant in the PA fraction (**Fig. 3B**). This is in good
260 accordance to other studies confirming that *Bacteroidetes* have been identified 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 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 were more evident on the functional level (**Fig. 4**). 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. 4A**). 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 sulphate 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. 4A & B**).

[Figure 4]

FL and PA bacteria seem to employ different phosphate acquisition strategies: whilst in FL bacteria phosphate and phosphonate ABC-type transporters appeared highly expressed, PA bacteria rather seem to employ phytases and phosphate-selective porins. As expected, various proteins involved in stress response were differentially expressed. Interestingly, functions involved in oxidative stress defense appeared to be less abundant in the PA fraction (maybe due to shading, reducing solar irradiation stress in the particles), whilst proteins for heavy metal and antibiotic resistance were strongly expressed in the ≥ 10 µm fraction, which also contained the highest proportion of eukaryotic proteins (**Fig. 4A**). This

might be due to the fact that some algae take up and store heavy metals (Gaudry *et al.,* 2007) and are capable of producing antibiotics (Grossart, 1999). This indicates that close eukaryote-bacterial interactions in particles require such defense strategies of the associated bacteria. As expected, adhesion proteins as well as proteins involved in motility, i.e. flagella and type IV-pili, were more abundant on the particles, emphasizing their importance for biofilm/aggregate formation (O´Toole & Kolter, 1998; Lemon *et al.*, 2007; Houry *et al.,* 2010; Burke *et al.*, 2011). Interestingly, proteorhodopsin, an inner membrane protein involved in light-dependent energy generation, which has been proposed to enable FL bacteria such as *Polaribacter* (Fernández-Gómez *et al*., 2013) and *Pelagibacter* (Giovannoni *et al*., 2005) to survive under low nutrient conditions, was also abundantly identified in PA bacteria such as *Polaribacter*, *Paraglaciecola*, and *Marinosulfomonas* in our analyses.

Eukaryotes are highly abundant and might contribute to polysaccharide degradation on marine particles. Our metaproteomics data demonstrate a high abundance of eukaryotes on the particles (**Fig. 5**). Preliminary analyses indicate that these include numerous microalgal groups, e.g. diatoms, *Pelagophytes*, *Raphidiophytes*, *Cryptophytes*, *Dinoflagellates* and *Haptophytes*, but also fungi and various protozoa (**Table S2 and S3**). This clearly sets particles apart from the FL fraction and highlights the importance of direct eukaryote-bacterial interactions in particles. Previous work on FL bacteria showed that bacterial succession was largely independent of phytoplankton composition, and instead determined by broad substrate availability (Teeling *et al*., 2016). PA bacterial composition is more likely to be directly controlled by algal composition due to the intimate nature of their interactions (Grossart *et al*., 2005), although functional redundancy may be substantial (Burke *et al*., 2011). Moreover, eukaryotes may also contribute to polysaccharide degradation in concert with bacteria. For example, fungal taxa can be abundant in marine particles and have been shown to utilize algal polysaccharides such as laminarin (Bochdansky *et al*., 2016; Cunliffe *et al.,* 2017).

[Figure 5]

315 *Conclusions and outlook*

316 Our comparative metaproteomic analyses of marine microbial communities living either
317 planktonically or attached to particles resulted in an as yet unequalled number of identified
318 protein groups for marine particles. Interestingly, the great overlap between metaproteomes
319 of FL and PA heterotrophic bacterial communities indicates that taxonomic differences
320 between them might be less pronounced than previously thought. This might be due to the
321 fact that (I) FL bacteria can rapidly adapt to the surface-associated life style, as the majority
322 of these bacteria seems to be also present on the particles and proteins important for biofilm-
323 formation, i.e. motility and adhesion proteins, are also expressed when living planktonically,
324 and that (II) FL or PA-specific bacteria are frequently hopping on and off the particles.
325 Notably, there is strong evidence that bacteria, when living on the particles, express life style-
326 specific functions, i.e. special CAZymes, sugar transporters and proteins involved in certain
327 stress responses, which enable them to cope with the unique living conditions on marine
328 particles.

329 Although our optimized metaproteomic workflow significantly improved the identification
330 rate of PA proteins, the number of protein identifications from the particles is still
331 considerably lower compared to FL bacterial communities. We assume that especially the
332 high abundance of eukaryotic proteins poses problems in protein identification due to the
333 complexity and diversity of microbial eukaryote genomes and the presence of introns and
334 repeats in the metagenomic DNA sequence databases, which hinders peptide identification
335 (Saito *et al*., 2019). Metaproteome coverage of marine particles could be significantly
336 improved by employing customized databases including eukaryotic metatranscriptomic
337 (RNA-based) sequence data (Keeling *et al*., 2014). This can be achieved by generating
338 metatranscriptomes from the particular fractions. Alternatively, protein identification could
339 also be substantially improved by extracting already existing metatranscriptomic and
340 metagenomic data from relevant eukaryotic taxa from public databases. Key to the latter
341 approach is reliable information on which eukaryotic organisms make up the particles, which

342 can be attained by 18S rRNA gene amplicon sequencing. Perspectively, we will extend our
343 analyses on eukaryotic taxa and analyse multiple time points during phytoplankton bloom to
344 investigate succession of taxonomical clades and expressed functions of marine particles
345 from pre-bloom to post-bloom conditions.

Experimental Procedures

Bacterial biomass samples. Sampling of bacterioplankton was performed as described previously (Teeling *et al.*, 2012). Briefly, surface water samples were taken during spring 2009 at the station "Kabeltonne" (50° 11.3' N, 7° 54.0' E) between the main island Helgoland and the minor island Düne about 40 km offshore in the southeastern North Sea in the German Bight. Bacterial biomass for protein extraction was sequentially filtered with peristaltic pumps onto 10 µm, 3 µm, and 0.2 µm pore-sized filters (142 mm diameter) to separate PA and FL bacteria. All filters were stored at -80 °C until further analyses.

Testing of protein extraction protocols. To test six different existing protein extraction protocols for their applicability on PA bacteria, filters from several time points containing varying amounts of biomass were chosen. Sample preparation for the metaproteomic analysis included cutting the filters into quarters and subsequently into small pieces (1-2 mm in diameter). Pieces of one quarter filter were transferred into 15 ml falcon tubes and treated according to the respective protocol.

Protocol 1 - Phenol. Filter pieces were incubated in 2.4 ml of a 0.1 M NaOH solution for 10 min at room temperature and were then sonicated three times for 30 s at 20% power output (Sonopuls HD2200 with microtip MS 73; Bandelin electronic, Germany). Subsequently the sample was centrifuged for 15 min at 12,500 x g at 20 °C to separate the supernatant from the filter pieces. The supernatant was transferred into a new tube and protein extraction using phenol was performed according to the protocol published by Kuhn and colleagues (Kuhn *et al.,* 2011).

Protocol 2 - SDS-TCA. Filter pieces were mixed with 5 ml extraction buffer (1% (w/v) SDS, 50 mM Tris/HCl, pH 7.0) and vigorously shaken for 2 min at room temperature. The cell disruption by sonication, boiling and shaking was performed according to the protein extraction protocol published by Schneider and colleagues (Schneider *et al.,* 2012). Subsequently, proteins were precipitated with 10% TCA over night at 4 °C. The precipitated

372 proteins were centrifuged for 20 min at 12,500 x g at 4 °C and the pellet was washed two
373 times in ice-cold acetone.

374 *Protocol 3 - TRI-Reagent®.* The TRI-Reagent® (Sigma-Aldrich, product-number T9424) is
375 used for the simultaneous isolation of RNA, DNA, and proteins. Filter pieces were transferred
376 into 4 ml TRI-Reagent and shaken vigorously for 5 min. Subsequently the proteins were
377 extracted according to the manufacturer's guidelines.

378 *Protocol 4 - Freeze and Thaw.* Protein extraction was carried out according to a combination
379 of the extraction protocols of Chourey *et al.* (2010) and Thompson *et al.* (2008). To this end,
380 filter pieces were mixed with 4 ml lysis buffer (5% SDS, 50 mM Tris/HCl, 0.1 mM EDTA,
381 0.15 M NaCl, 1 mM $MgCl_2$, 50 mM DTT, pH 8.5) and vigorously shaken for 3 min.
382 Subsequently, the samples were boiled for 10 min, followed by two freezing and thawing
383 cycles with liquid nitrogen. After cooling at 4 °C the samples were vigorously shaken for
384 3 min. To remove cell debris, samples were centrifuged for 20 min at 12,500 x g at 4 °C. The
385 proteins in the supernatant were precipitated with 25% TCA over night at 4 °C. Precipitated
386 proteins were centrifuged for 20 min at 12,500 x g at 4 °C and the resulting protein pellet was
387 washed with ice-cold acetone.

388 *Protocol 5 - SDS-Acetone.* Filter pieces were mixed with 5 ml extraction buffer (50 mM Tris,
389 1% (w/v) SDS, pH 7.5) and vortexed vigorously. Proteins were extracted by sonication,
390 boiling and shaking as described by Hall and colleagues (Hall *et al*., 2012). Subsequently,
391 proteins were precipitated with five volumes acetone over night at -20 °C. The precipitated
392 proteins were centrifuged for 20 min at 12,500 x g at 4 °C and the pellet washed two times in
393 ice-cold acetone.

394 *Protocol 6 – bead beating*. Protein extraction was carried out according to the extraction
395 protocol of Moog (2012), which is based on the protocol of Teeling and colleagues (Teeling
396 *et al*., 2012). To this end, filter pieces were covered with 4 ml lysis buffer (0.1 M DTT, 0.01 M
397 EDTA, 10% Glycerol (v/v), 1.7 mM PMSF, 5% SDS (w/v), 0.05 M Tris/HCl, pH 6.8) and 2 ml
398 glass beads (0.1 – 0.11 mm diameter) were added. The cells on the filter pieces were

subsequently disrupted four times for 30 s with 6.5 m/s via bead beating with a Fast Prep™-24 (MP Biomedicals, Germany). To remove cell debris and glass beads, samples were centrifuged for 20 min at 12,500 x g at 4 °C and the supernatant was transferred into new tubes. This washing step was repeated 2 to 4 times until the beads were colourless. The glass beads were washed with 3 ml lysis buffer and vigorously shaken. Proteins enriched in the pooled supernatants were precipitated with 1:4 acetone at -20 °C over night. Precipitated proteins were centrifuged for 20 min at 12,500 x g at 4 °C and the resulting protein pellet washed with ice-cold acetone.

All resulting protein pellets were air-dried and resolved in 8 M urea / 2 M thiourea.

Determination of protein concentrations. . Protein concentrations were determined using the Pierce™ BCA Protein Assay Kit (Thermo Fisher Scientific). Protein extracts were prepared with the Compat-Able™ Protein Assay Preparation Reagent Kit (Thermo Fisher Scientific) according to the manufacturer's guidelines.

SDS-PAGE protein separation. 30 µg protein or 30 µl protein extract was mixed with 4x SDS sample buffer (20% glycerol, 100 mM Tris/HCl, 10% (w/v) SDS, 5% β-mercaptoethanol, 0.8% bromphenol blue, pH 6.8) and loaded on TGX precast 4-20% gels (Biorad, Germany). Samples were separated by electrophoresis at 150 V for 45 min. After fixation (10% acetic acid, 40% ethanol, 30 min) the gels were stained with Brilliant Blue G250 Coomassie and imaged.

Protein digestion and MS-sample preparation. Three different protocols were tested on proteins extracted from 3 µm and 10 µm filters.

Protocol 1 - 10 gel pieces. Protein lanes were cut into 10 equal-sized pieces and washed with a buffer containing 50 mM ammoniumbicarbonate and 30% (v/v) acetonitrile. Prior to tryptic digestion, gel pieces were dried in a vacuum concentrator and re-swollen with 2 ng/µl trypsin solution (sequencing grade trypsin, Promega, USA) followed by overnight digestion at 37 °C. After digestion the gel pieces were covered with water and peptides were eluted from

the gel in an ultrasonic bath. The eluted peptides were desalted with C18 Millipore® ZipTip columns (Millipore) according to the manufacturer's guidelines.

Protocol 2 - 20 gel pieces. The protein lanes were cut into 20 equal-sized pieces and treated as described above.

Protocol 3 - 20 gel pieces with reduction and alkylation. The protein lanes were cut into 20 equal pieces and washed with a buffer containing 100 mM ammoniumbicarbonate (NH_4HCO_3) and 50% (v/v) methanol. Subsequently, proteins were reduced in 50 mM NH_4HCO_3 containing 10 mM DTT for 30 min at 60 °C, followed by alkylation in 50 mM NH_4HCO_3 containing 50 mM iodoacetamide (IAA) for 60 min in the dark at room temperature. Prior to tryptic digestion, the gel pieces were dehydrated using 100% acetonitrile and dried, re-swollen with 2 ng/µl trypsin solution and incubated at 37 °C over night. Peptides were eluted from the gel pieces by a six-step procedure, using acetonitrile, 1% (v/v) acetic acid in water, acetonitrile, 10% (v/v) acetic acid and two times acetonitrile. Peptide-containing supernatants were pooled and completely dried in a vacuum concentrator. Samples were subsequently resolved in buffer A (5% (v/v) acetonitrile, 0.1% (v/v) formic acid) and desalted with C18 Millipore® ZipTip columns (Millipore) according to the manufacturer's guidelines.

Constructions of a protein sequence database from marine particle metagenomes. Environmental DNA was extracted from the 0.2 µm, 3 µm and 10 µm pore-sized filters sampled on the 14th of April 2009 by a modified standard protocol of Zhou *et al.* (1996). In detail one polycarbonate filter was cut into 4 pieces and mixed with 13.5 ml extraction buffer (100 mM Tris-HCl (pH 8.0), 100 mM EDTA (pH 8.0), 100 mM Na-phosphate (pH 8.0), 1.5 M NaCl, 1% CTAB (Hexadecyltrimethylammonium-bromide)). Subsequently 100 µl 10 mg/ml Proteinase K was added and the sample was incubated shaking at 37 °C for 30 min. 1.5 ml 20% SDS was added and the sample was incubated shaking at 65 °C for 2 h. The sample was centrifuged at 6,000 x g for 10 min at room temperature and the supernatants were transferred to fresh tubes. Subsequently, an equal volume of chloroform/isoamylalcohol was added and the sample was mixed carefully by shaking and was centrifuged at 10,000 x g for

10 min at room temperature. Afterwards the aqueous upper phase was transferred into a new tube and the DNA was precipitated by addition of 0.6 volumes isopropanol. The sample was moderately shaken over night at 4°C. After centrifugation at 50,000 x g for 20 min at room temperature, the pellet was washed with 10 ml 80% (v/v) ethanol and dried. The pellet was resuspended in 200 µl TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 7) and stored at -20 °C until sequencing.

DNA was sequenced at the Max Planck Sequencing Centre (Cologne, Germany), using the Illumina HiSeq 2500 platform and 2 x 250 bp chemistry. Sequences were then trimmed using bbduk v35.14 (http://bbtools.jgi.doe.gov) with the following parameters: ktrim = r k = 28 mink = 12 hdist = 1 tbo = t tpe = t qtrim = rl trimq = 20 minlength = 100. Read quality for each sample was then confirmed using FastQC v0.11.2 (Andrews, 2010). Trimmed and filtered reads from the three metagenomic datasets were then assembled individually. The 0.2 µm pore-sized filter sample was assembled with metaSPAdes v3.10.1 (Nurk *et al.*, 2017) with kmers of length 21, 33, 55, 77, 99, and 127, and error correction mode switched on. Assembly of the larger size fraction was done with MEGAHIT v1.1.3 (Li et al., 2016) with kmers 21, 33, 55, 77, 99, 127, 155, 183, and 211. Assembled contigs longer than 1500 base pairs were kept for gene predictions. Genes were predicted and annotated using Prokka v1.11 (Seeman, 2014), which implements prodigal v2.6.3 (Hyatt *et al.*, 2010) for ORF prediction.

Raw read sequences and assembled contig sequences have been deposited in the European nucleotide archive (ENA) under the project accession number PRJEB2888.

LC-MS/MS data acquisition and data analysis. Peptides were separated by reversed-phase chromatography on an Easy-nLC 1000 (Thermo Scientific) with self-packed C18 analytical columns (100 µm × 20 cm) and coupled to a LTQ Orbitrap Velos mass spectrometer(Thermo Scientific) using a non-linear binary gradient of 80 minutes from 5 % solvent A (0.1 % (v/v) acetic acid) to 99 % solvent B (0.1 % acetic acid (v/v), 99.9 % acetonitrile (v/v)) and a flow rate of 300 nl/min. Survey scans at a resolution of 30,000 were

recorded in the Orbitrap analyser (m/z 300 - 1700) and the 20 most intense precursor ions were selected for CID fragmentation in the LTQ. Dynamic exclusion of precursor ions was enabled; single-charged ions and ions with unknown charge state were excluded from fragmentation. Internal lock mass calibration was enabled (lock mass 445.120025).

The mass spectrometry raw data were converted into mgf files using MSConvert (64-bit, Proteowizard 3) and subsequently subjected to database searching via Mascot (Matrix Science; version 2.6.0). Four different protein sequence databases were used for peptide to spectrum matching: I) the non-redundant NCBI database (NCBI nr - NCBIprot_20171030 database (136,216,794 entries)), II) a database containing protein sequences of abundant bacteria and diatoms (PABD) based on the study of Teeling *et al*. (2012), and retrieved from Uniprot KB (Uniprot_DoS_complete_20170829 database (2,638,314 entries)), III) a database containing protein sequences of the free-living fraction from Teeling *et al*. (2012) (0.2 µm 2009 (MIMAS) - MIMAS_forward_reverse_all_contaminants database (1,579,724 entries)), and IV) a database based on translated metagenomes from 0.2 µm and 3 µm filters (see below for details) (0.2 + 3 µm 2009 - 02_plus_3_POMPU_nr97_fw_cont_20181015 database (1,463,571 entries). Mascot was searched with a fragment ion mass tolerance of 0.80 Da and a parent ion tolerance of 10.0 ppm. Oxidation of methionine was specified as a variable modification, trypsin was set as digestion enzyme and a maximum of two missed cleavages was allowed.

Scaffold (version Scaffold_4.8.7, Proteome Software Inc.) was used to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 95.0% probability by the Peptide Prophet algorithm (Keller *et al.,* 2002) with Scaffold delta-mass correction. Protein identifications were accepted if they could be established at greater than 99.0% probability and contained at least one identified peptide. Protein probabilities were assigned by the Protein Prophet algorithm (Nesvizhskii *et al*., 2003). Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony. Peptides that

were only found in one of the replicates were excluded from the following data analysis. Mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository (Perez-Riverol *et al*., 2019) with the data set identifier PXD12699 (reviewer account details: username reviewer9795@ebi.ac.uk; password 5CKUi0AF).

For further data analysis, the software *ProPHAnE* (Proteomics result Pruning and Homology group Annotation Engine; version 3.1.1) (Schneider *et al.*, 2011) was used. For the taxonomical classification of the identified protein groups the NCBI NR database (version 2018-08-02; e-value 0.01, query cover 0.9, max-target-seqs 1) and the diamond blastp algorithm (version 0.8.22) were used. For functional classification of the identified protein groups the eggmap database (version 4.5.1, downloaded at 2018-07-31) and the algorithm e-mapper were used.

A list of common contaminants was added to all translated ORF sequences found by metagenome analysis of the 0.2 µm and 3 µm filters from the sampling date 14th of April 2009. Redundant sequences were eliminated (97% redundancy, elimination of shorter sequence) using CD-HIT (www.cdhit.org), a program for clustering and comparing protein or nucleotide sequences, resulting in the database 02_plus_3_POMPU_nr97_fw_cont_20181015 (1,463,571 entries).

Acknowledgements.

We thank the DFG for financial support (RI 969/9-1) in the frame of the Research Unit "POMPU" (FOR 2406). We are grateful to Sabine Kühn for technical assistance and to Thomas Schweder for helpful discussions.

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.

Andrews, S. (2010) FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

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.

Azam, F., and Malfatti, F. (2007) Microbial structuring of marine ecosystems. *Nature Reviews* **5:** 782-791.

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.

Battin, T.J., Kaplan, L.A., Denis Newbold, J., and Hansen, C.M. (2003) Contributions of microbial biofilms to ecosystem processes in stream mesocosms. *Nature* **426:** 439-442.

557 Bauer, M., Kube, M., Teeling, H., Richter, M., Lombardot, T., Allers, E., *et al*. (2006) Whole
558 genome analysis of the marine *Bacteroidetes 'Gramella frosetii'* reveals adaptations to
559 degradation of polymeric organic matter. *Environ Microbiol* **8:** 2201-2213.
560 Bergauer, K., Fernandez-Guerra, A., Garcia, J.A.L., Sprenger, R.R., Stepanauskas, R.,
561 Pachiadaki, M.G., *et al*. (2018) Organic matter processing by microbial communities
562 throughout the Atlantic water column as revealed by metaproteomics. *Proc Natl Acad Sci*
563 *U S A* **115:** E400-E408.
564 Bižić-Ionescu, M., Zeder, M., Ionescu, D., Orlić, S., Fuchs, B.M., Grossart, H.P., *et al*. (2015)
565 Comparison of bacterial communities on limnic versus coastal marine particles reveals
566 profound differences in colonization. *Environ Microbiol* **17:** 3500-3514.
567 Bochdansky, A.B., Clouse, M.A., and Herndl, G.J. (2016). Eukaryotic microbes, principally
568 fungi and labyrinthulomycetes, dominate biomass on bathypelagic marine snow. *ISME J*
569 **11:** 362-373.
570 Burke, C., Steinberg, P., Kjelleberg, S., and Thomas, T. (2011) Bacterial community
571 assembly based on functional genes rather than species. *PNAS* **108:** 14288-14293.
572 Caron, D.A., Davis, P.G., Madin, L.P., and Sieburth, J.M. (1982) Heterotrophic bacteria and
573 bacterivorous protozoa in oceanic macroaggregates. *Science* **218:** 795-797.
574 Chafee, M., Fernàndez-Guerra, A., Buttigieg, P.L., Gerdts, G., Eren, A.M., Teeling, H., *et al.*
575 (2018) Recurrent patterns of microdiversity in a temperate coastal marine environment.
576 *ISME J* **12:** 237-252.
577 Chourey, K., Jansson, J., VerBerkmoes, N., Shah, M., Chavarria, K.L., Tom, L.M., *et al.*
578 (2010) Direct cellular lysis/protein extraction protocol for soil metaproteomics. *J Proteome*
579 *Res* **9:** 6615-6622.
580 Christie-Oleza, J.A., and Armengaud, J. (2015) Proteomics of the Roseobacter clade, a
581 window to the marine microbiology landscape. *Proteomics* **15:** 3928-3942.
582 Crespo, B.G., Pommier, T., Fernández-Gómez, B., and Pedrós-Alió, C. (2013) Taxonomic
583 composition of the particle-attached and free-living bacterial assemblages in the
584 Northwest Mediterranean Sea. *Microbiology Open* **2:** 541-552.

585 Cunliffe, M., Hollingsworth, A., Bain, C., Sharma, V., and Taylor, J.D. (2017) Algal
586 polysaccharide utilisation by saprotrophic planktonic marine fungi. *Fungal Ecol* **30:** 135-
587 138.
588 DeLong, E.F., Franks, D.G., and Alldredge, A.L. (1993) Phylogenetic diversity of aggregate-
589 attached vs. free-living marine bacterial assemblages. *Limnol Oceanogr* **38:** 924-934.
590 Eilers, H., Pernthaler, J., Peplies, J., Glöckner, F.O., Gerdts, G., and Amann, R. (2001)
591 Isolation of novel pelagic bacteria from the German Bight and their seasonal contribution
592 to surface picoplankton. *AEM* **67:** 5134-5142.
593 Falkowski, P.G., Barber, R.T., and Smetacek, V. (1998) Biogeochemical controls and
594 feedbacks on ocean primary production. *Science* **281:** 200-206.
595 Fernández-Gómez, B., Richter, M., Schüler, M., Pinhassi, J., Acinas, S., González, J.M. *et*
596 *al.* (2013) Ecology of marine Bacteroidetes. A comparative genomics approach. *ISME J* **7:**
597 1026-1037.
598 Field, C.B., Behrenfeld, M.J., Randerson, J.T., and Falkowski, P. (1998) Primary production
599 of the biosphere: integrating terrestrial and oceanic components. *Science* **281:** 237-240.
600 Ganesh, S., Parris, D.J., DeLong, E.F., and Stewart, F.J. (2014) Metagenomic analysis of
601 size-fractionated picoplankton in a marine oxygen minimum zone. *ISME J* **8:** 187-211.
602 Gaudry, A., Zeroual, F., Gaie-Levrel, F., Moskura, M., Boujrhal, F.Z., Cherkaoui, E. *et al*.
603 (2007) Heavy metals pollution of the atlantic marine environment by the Moroccan
604 phosphate industry, as observed through their bioaccumulation in *Ulva lactuca*. *Water Air*
605 *Soil Pollut* **178:** 267-285.
606 Ghiglione, J.F., Mevel, G., Pujo-Pay, M., Mousseau, L., Lebaron, P., and Goutx, M. (2007)
607 Diel and seasonal variations in abundance, activity, and community structure of particle-
608 attached and free living bacteria in NW Mediterranean Sea. *Microb Ecol* **54:** 217-231.
609 Giovannoni, S.J., Tripp, H.J., Givan, S., Podar, M., Vergin, K.L., Baptista, D., *et al*. (2005)
610 Genome Streamlining in a Cosmopolitan Oceanic Bacterium. *Science* **309:** 1242-1245.

611 Gómez-Pereira, P.R., Schüler, M., Fuchs, B.M., Bennke, C., Teeling, H., Waldmann, J., *et al*.
612 (2012) Genomic content of uncultured Bacteroidetes from contrasting oceanic provinces
613 in the North Atlantic Ocean. *Environ Microbiol* **14:** 52-66.
614 Grossart, H.P. (1999) Interactions between marine bacteria and axenic diatoms
615 (*Cylindrotheca fusiformis*, *Nitzschia laevis*, and *Thalassiosira weissflogii*) incubated under
616 various conditions in the lab. *Aquat. Microb Ecol* **19:** 1-11.
617 Grossart, H.P., Hietanen, S., and Ploug, H. (2003) Microbial dynamics on diatom aggregates
618 in Oresund, Denmark. *Mar Ecol Prog Ser* **249:** 69-78.
619 Grossart, H.P., Levold, F., Allgaier, M., Simon, M., and Brinkhoff, T. (2005) Marine diatom
620 species harbour distinct bacterial communities. *Environ Microbiol* **7:** 860-873.
621 Grossart, H.P. (2010) Ecological consequences of bacterioplankton lifestyles: changes in
622 concepts are needed. *Environ Microbiol Rep* **2:** 706-714.
623 Hall, E.K., Besemer, K., Kohl, L., Preiler, C., Riedel, K., Schneider, T., *et al*. (2012) Effects of
624 resource chemistry on the composition and function of stream hyporheic Biofilms. *Front*
625 *Microbiol* **3:** 1-14.
626 Hinzke, T., Kouris, A., Hughes, R.A., Strous, M., and Kleiner, M. (2019) More Is Not Always
627 Better: Evaluation of 1D and 2D-LC-MS/MS Methods for Metaproteomics. *Front Microbiol*
628 **10:** 238.
629 Houry, A., Briandet, R., Aymerich, S., and Gohar, M. (2010) Involvement of motility and
630 flagella in *Bacillus cereus* biofilm formation. *Microbiology* **156:** 1009-1018.
631 Hyatt, D., Chen, G.L., LoCascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010)
632 Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC*
633 *Bioinformatics* **11:** 119.
634 Kappelmann, L., Krüger, K., Hehemann, J.H., Harder, J., Markert, S., Unfried, F., *et al*.
635 (2019) Polysaccharide utilization loci of North Sea *Flavobacteriia* as basis for using
636 SusC/D-protein expression for predicting major phytoplankton glycans. *ISME J* **13:** 76-91
637 Keeling, P.J., Burki, F., Wilcox, H.M., Allam, B., Allen, E.E., Amaral-Zettler, L.A., *et al.* (2014)
638 The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP):

639 Illuminating the functional diversity of eukaryotic life in the oceans through transcriptome
640 sequencing. *PLoS Biol* **12:** e1001889-6.
641 Keiblinger, K.M., and Riedel, K. (2018) Sample Preparation for Metaproteome Analyses of
642 Soil and Leaf Litter. *Methods Mol Biol* **1841:** 303-318.
643 Keller, A., Nesvizhskii, A.I., Kolker, E., and Aebersold, R. (2002) Empirical statistical model to
644 estimate the accuracy of peptide identifications made by MS/MS and database search.
645 *Anal Chem* **74:** 5383-5392.
646 Kroth, P.G., Chiovitti, A., Gruber, A., Martin-Jezequel, V., Mock, T., Schnitzler Parker, M., *et*
647 *al*. (2008). A model for carbohydrate metabolism in the diatom *Phaeodactylum tricornutum*
648 deduced from comparative whole genome analysis. *PLoS One* **3:** e1426.
649 Kuhn, R., Benndorf, D., Rapp, E., Reichl, U., Palese, L.L., and Pollice, A. (2011)
650 Metaproteome analysis of sewage sludge from membrane bioreactors. *Proteomics* **11:**
651 2738-2744.
652 Lassek, C., Burghartz, M., Chaves-Moreno, D., Otto, A., Hentschker, C., Fuchs, S., *et al*.
653 (2015) A metaproteomics approach to elucidate host and pathogen protein expression
654 during catheter-associated urinary tract infections (CAUTIs). *Mol Cell Proteomics* **14:** 989-
655 1008.
656 Lemon, K.P., Higgins, D.E., and Kolter, R. (2007) Flagellar motility is critical for *Listeria*
657 *monocytogenes* biofilm formation. *J Bacteriol* **189:** 1009-1018.
658 Li, D., Luo, R., Liu, C.M., Leung, C.M., Ting, H.F., Sadakane, K., *et al*. (2016) MEGAHIT
659 v1.0: A Fast and Scalable Metagenome Assembler driven by Advanced Methodologies
660 and Community Practices. *Methods* **102:** 3-11.
661 Mineta, K., and Gojobori, T. (2016). Databases of the marine metagenomics. *Gene* **576:** 724-
662 728.
663 Moog, G. (2012) Optimierung der Probenaufbereitung für Metaproteomanalysen mittels
664 Massenspektrometrie. Diploma thesis at the University of Greifswald, Institute of
665 Microbiology.

Moore, E.K., Nunn, B.L., Goodlett, D.R., 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.

Nesvizhskii, A.I., Keller, A., Kolker, E., and Aebersold, R. (2003) A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem* **75:** 4646-4658.

Nurk, S., Meleshko, D., Korobeynikov, A., and Pevzner, P.A. (2017) metaSPAdes: a new versatile metagenomic assembler. *Genome Res* **27:** 824-834.

O'Toole, G.A., and Kolter, R. (1998) Flagellar and twitching motility are necessary for *Pseudomonas aeruginosa* biofilm development. *Mol Microbiol* **30:** 295-304.

Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D.J., *et al*. (2019) The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res* **47:** D442-D450.

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.

693 Schneider, T., and Riedel, K. (2010) Environmental proteomics. Analysis of structure and
694 function of microbial communities. *Proteomics* **10:** 785-798.
695 Schneider, T., Schmid, E., de Castro Junior, J.V., Cardinale, M., Eberl, L., Grube, M., *et al.*
696 (2011) Structure and function of the symbiosis partners of the lung lichen (*Lobaria*
697 *pulmonaria* L. Hoffm.) analyzed by metaproteomics. *Proteomics* **11:** 2752-2756.
698 Schneider, T., Keiblinger, K.M., Schmid, E., Sterflinger-Gleixner, K., Ellersdorfer, G.,
699 Roschitzki, B., *et al*. (2012) Who is who in litter decomposition? Metaproteomics reveals
700 major microbial players and their biogeochemical functions. *ISME J* **6:** 1749-1762.
701 Seemann, T. (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30:** 2068-
702 2069.
703 Simon, M., Grossart, H.P., Schweitzer, B., and Ploug, H. (2002) Microbial ecology of organic
704 aggregates in aquatic ecosystems. *Aquat Microb Ecol* **28:** 175-211.
705 Simon, M., Glöckner, F.O., and Amann, R. (1999) Different community structure and
706 temperature optima of heterotrophic picoplankton in various regions of the Southern
707 Ocean. *Aquat Microb Ecol* **18:** 275-284.
708 Sohn, J.H., Lee, J.H., Yi, H., Chun, J., Bae, T.S., Ahn, T.Y., *et al.* (2004) *Kordia algicida* gen.
709 nov., sp. nov., an aligicidal bacterium isolated from red tide. *Int J Syst Evol Microbiol* **54:**
710 675-80.
711 Teeling, H., Fuchs, B.M., Becher, D., Klockow, C., Gardebrecht, A., Bennke, C.M., *et al.*
712 (2012) Substrate-controlled succession of marine bacterioplankton populations induced by
713 a phytoplankton bloom. *Science* **336:** 608-611.
714 Teeling, H., Fuchs, B.M., Bennke, C.M., Krüger, K., Chafee, M., Kappelmann, L., *et al.*
715 (2016) Recurring patterns in bacterioplankton dynamics during coastal spring algae
716 blooms. *Elife* **5:** e1188.
717 Thompson, M.R., Chourey, K., Froelich, J.M., Erickson, B.K., VerBerkmoes, N.C., and
718 Hettich, R.L. (2008) Experimental approach for deep proteome measurements from small-
719 scale microbial biomass samples. *Anal Chem* **80:** 9517-9525.

720 Volkman, J., and Tanoue, E. (2002) Chemical and biological studies of particulate organic
721 matter in the ocean. *J Oceanogr* **58:** 265-279.
722 Wang, D.Z., Xie, Z.X., and Zhang, S.F. (2014). Marine metaproteomics: current status and
723 future directions. *J Proteomics* **97:** 27-35.
724 Williams, T.J., Long, E., Evans, F., Demaere, M.Z., Lauro, F.M., Raftery, M.J., *et al*. (2012) A
725 metaproteomic assessment of winter and summer bacterioplankton from Antarctic
726 Peninsula coastal surface waters. *ISME J* **6:** 1883-1900.
727 Wilmes, P., Wexler, M., and Bond, P.L. (2008) Metaproteomics provides functional insight
728 into activated sludge wastewater treatment. *PLoS One* **3:** e1778.
729 Wöhlbrand, L., Wemheuer, B., Feenders, C., Ruppersberg, H.S., Hinrichs, C., Blasius, B., *et*
730 *al.* (2017a) Complementary metaproteomic approaches to assess the bacterioplankton
731 response toward a phytoplankton spring bloom in the Southern North Sea. *Front Microbiol*
732 **8:** 442.
733 Wöhlbrand, L., Feenders, C., Nachbaur, J., Freund, H., Engelen, B., Wilkes, H., *et al.*
734 (2017b) Impact of extraction methods on the detectable protein complement of
735 metaproteomic analyses of marine sediments. *Proteomics* **17:** 22.
736 Zhou, J., Bruns, M.A., and Tiedje, J.M. (1995) DNA Recovery from Soils of Diverse
737 Composition. *AEM* **62:** 316-322.

738 **Figure legends**

739 **Figure 1: (A) Sampling strategy and (B) evaluation of protein extraction and MS**
740 **sample preparation protocols. (A)** Water samples collected at "Kabeltonne" Helgoland
741 during the spring bloom 2009 were sequentially filtered to obtain the 0.2 - 3 µm (FL) and two
742 PA (3 - 10 µm = medium, ≥ 10 µm = large) fractions as described in Teeling *et al*. (2012).
743 Filters were initially cut into three or four pieces, which were subsequently shredded and
744 mixed with the respective extraction buffer. **(B)** Filters (medium particles = yellow; large
745 particles = blue) from different sampling time points (turquoise, green and red) were
746 processed according to the six different protocols describe in the experimental procedure
747 section. With regard to the extracted protein amount the bead beating and SDS-acetone
748 approaches obviously outcompeted the four other protocols. However, the SDS-acetone
749 protocol was less reproducible than the bead beating protocol. Considering bead beating and
750 SDS-acetone as best performing protocols, they were employed to test different MS sample
751 preparation approaches, i.e., different number of SDS gel fractions for tryptic digestion
752 together with protein reduction (red.) and alkylation (alk.) prior to tryptic digestion. The
753 subsequent LC-MS/MS analyses revealed best results for the bead beating protocol followed
754 by GeLC-MS/MS from 20 fractions without protein reduction and alkylation as shown in the
755 bottom line of the figure. Bubble sizes for the large (blue) and medium (yellow) particles
756 correspond to the number of identified protein groups (see also **Fig. S1**).

757

758 **Figure 2: Final metaproteomics pipeline.** Protein extraction from filters was conducted
759 using 5% (w/v) SDS containing lysis buffer, cell disruption by FastPrep-mediated bead
760 beating, separation of proteins by 1D-SDS-PAGE, tryptic in-gel digestion, LC-MS/MS
761 analyses on an Orbitrap Velos^{TM} mass spectrometer, MASCOT database search against the
762 metagenome-based database (0.2 + 3 µm 2009) and data-processing and visualization with
763 the *in-house*-developed bioinformatics tools *Prophane 3.1* and *Paver*.

764

765 **Figure 3: Taxonomic affiliation of proteins of FL and PA metaproteomes during the**
766 **spring bloom on 14th of April 2009 at "Kabeltonne" Helgoland. (A)** Distribution of pro-
767 and eukaryotes in the FL (0.2 - 3 µm) and PA (3 - 10 µm, ≥ 10 µm) fractions based on the
768 relative abundance of protein groups assigned to the different phylogenetic groups. **(B)**
769 Voronoi treemaps visualizing the phylogenetic assignment of bacterial protein groups
770 identified in FL (red) and PA (yellow and blue) fractions. Cell size corresponds to the relative
771 abundance of the respective bacterial genus on protein level. Proteins of *Reinekea* for
772 example are most abundant in the FL fraction and are therefore encoded by a large red
773 treemap cell. In the PA fractions they can be detected only in traces resulting in very small
774 cell sizes (coloured in yellow and blue). *Algibacter* protein abundance, on the other hand,
775 was notably higher in the PA fractions, compared to the FL fraction.
776
777 **Figure 4: Functional assignment of proteins in FL and PA metaproteomes during the**
778 **spring bloom on 14th of April 2009 at "Kabeltonne" Helgoland. (A)** Total abundance of
779 selected protein groups with assigned functions in the FL (0.2 – 3 µm = small) and PA (3 - 10
780 µm = medium and ≥ 10 µm = large) fractions. **(B)** Voronoi treemaps showing the
781 phylogenetic assignment of selected functional protein groups identified in FL (red) and PA
782 (yellow and blue) fractions. Cell size corresponds to the relative abundance of the respective
783 genus within specific functional categories.
784
785 **Figure 5: Phylogenetic assignment of eukaryotic proteins present in the FL and PA**
786 **fractions during the spring bloom on 14th of April 2009 at "Kabeltonne" Helgoland. (A)**
787 Distribution of different eukaryotes in the FL (0.2 µm) and PA (0.2 - 3 µm and ≥ 10 µm)
788 fractions as shown by the relative protein abundances assigned to the different eukaryotic
789 phylogenetic groups. **(B)** Voronoi treemap visualizing the relative abundance of eukaryotic
790 taxa based on the abundance of assigned proteins extracted from the FL (red) and PA
791 (yellow and blue) fractions. Cell size corresponds to the relative abundance of the respective
792 genus. In this preliminary analysis, protein identification is based on metagenomic (DNA-

793 based) information from the filtered fractions, which suffers limitations for eukaryotic protein
794 identification, probably resulting in incomplete functional and taxonomic profiles.
795
796

Supplemental Material

Figure S1. 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 S2: 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.* (2012) (PABD, 2,638,314 entries), **(III)** a metagenome-based database employed for the FL bacterial fraction within the study of Teeling *et al*. (2012) (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 S3: Venn diagram of overlapping and fraction-specific protein sets.

Table S1: Prophane output for proteins extracted from 0.2 µm pore-sized filters

Table S2: Prophane output for proteins extracted from 3 µm pore-sized filters

Table S3: Prophane output for proteins extracted from 10 µm pore-sized filters

Table S4: Distribution of phylogenetic groups within proteins extracted from the 0.2 µm, 3 µm and 10 µm pore-sized filters.

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**	**6 - Bead beating**
Reference	Kuhn *et al.* (2011)	Schneider *et al.* (2012)	Sigma Aldrich	Chourey *et al.* (2010) 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 $MgCl_2$, 50mM 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 x 30 s (20% power output)	sonication 3 x 40 s (20% power output)	TRI-Reagent®	2 freeze and thaw cycles (liquid nitrogen, rt), 10 min boiling	sonication 5 x 1 min (20% power output), 15 min boiling, procedure repeated on the pellet	FastPrep® 6.5 m/s, 4 x 30 s
Additional protein purification	phenol extraction (2x)	/	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**

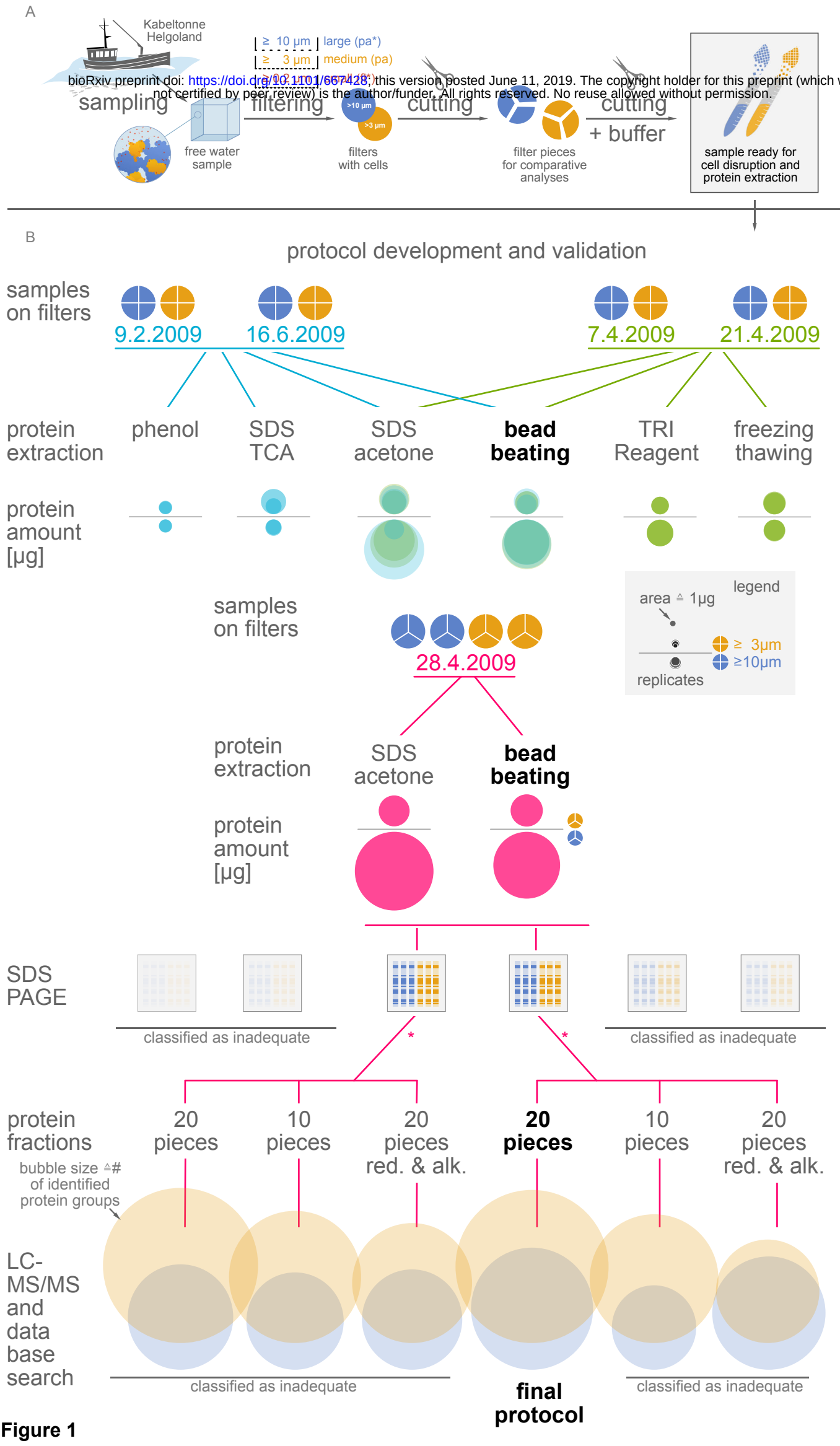


Figure 1

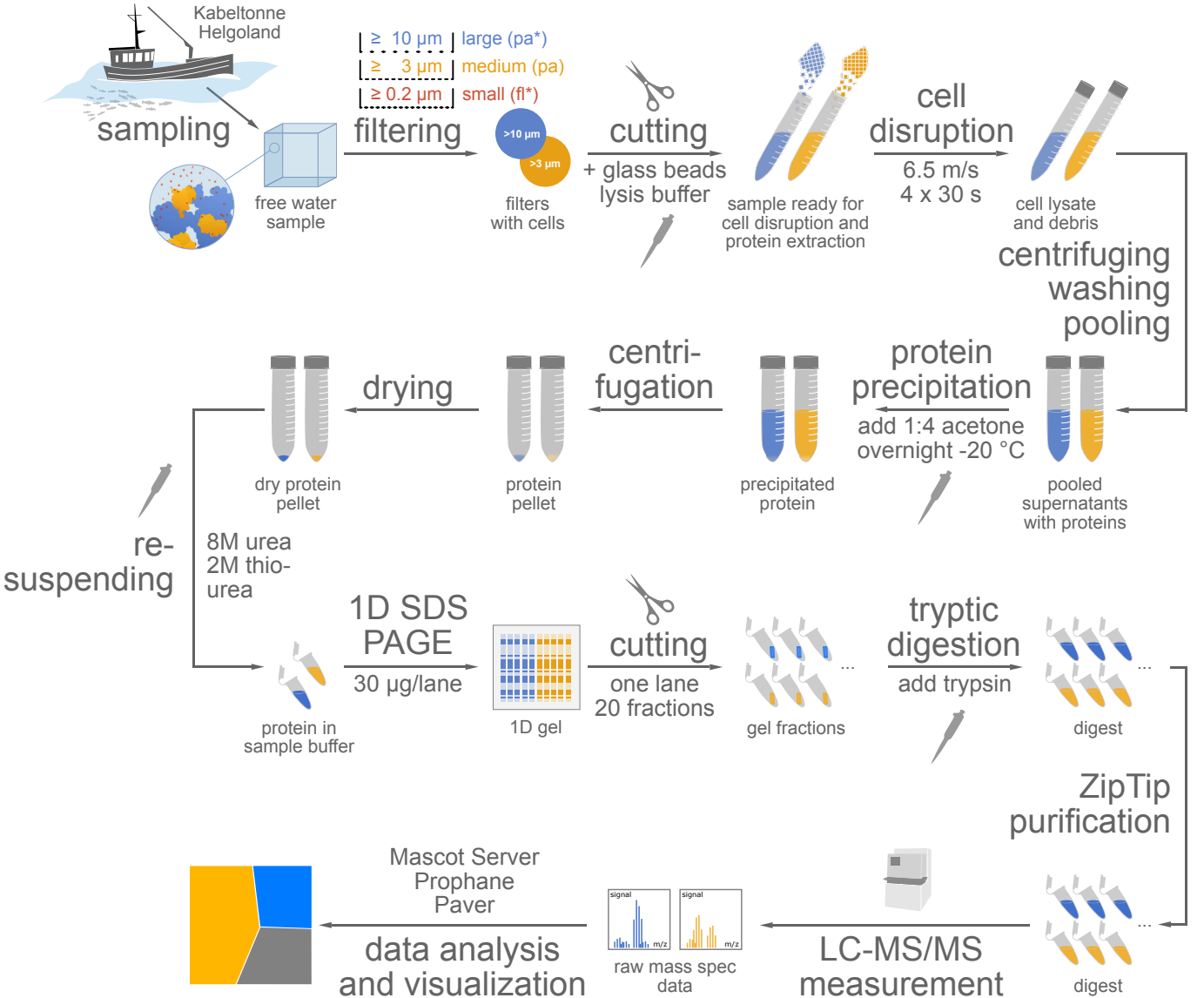

Kabeltonne Helgoland
sampling
free water sample
≥ 10 µm large (pa*)
≥ 3 µm medium (pa)
≥ 0.2 µm small (fl*)
filtering
>10 µm
>3 µm
filters with cells
cutting
+ glass beads lysis buffer
sample ready for cell disruption and protein extraction
cell disruption
6.5 m/s
4 x 30 s
cell lysate and debris
centrifuging
washing
pooling
protein precipitation
add 1:4 acetone
overnight -20 °C
pooled supernatants with proteins
precipitated protein
centri-fugation
protein pellet
drying
dry protein pellet
re-suspending
8M urea
2M thio-urea
protein in sample buffer
1D SDS PAGE
30 µg/lane
1D gel
cutting
one lane
20 fractions
gel fractions
tryptic digestion
add trypsin
digest
ZipTip purification
digest
LC-MS/MS measurement
raw mass spec data
signal
m/z
Mascot Server
Prophane
Paver
data analysis and visualization

Figure 2

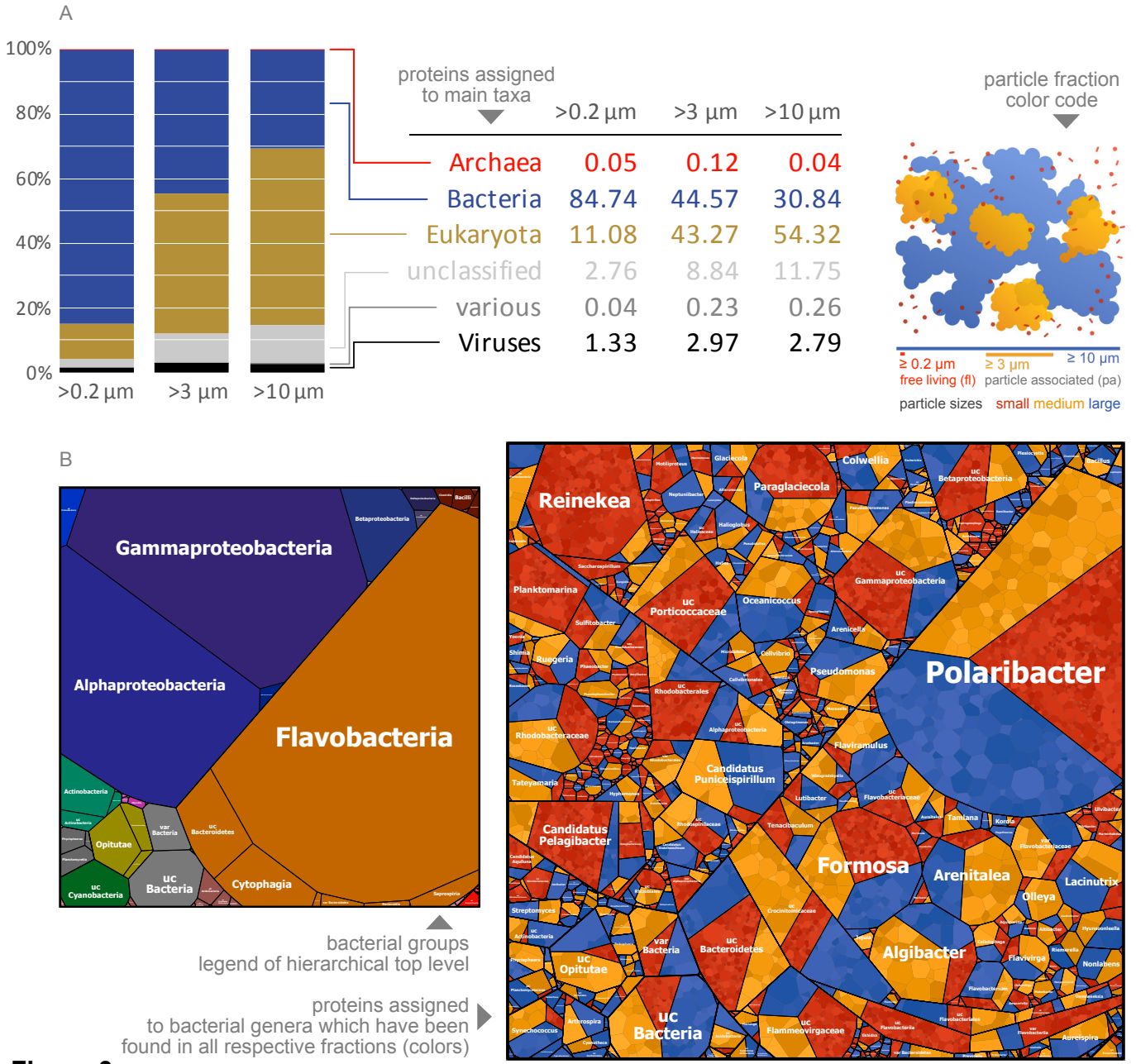

A
100%
80%
60%
40%
20%
0%
>0.2 µm
>3 µm
>10 µm
proteins assigned to main taxa
>0.2 µm
>3 µm
>10 µm
Archaea 0.05 0.12 0.04
Bacteria 84.74 44.57 30.84
Eukaryota 11.08 43.27 54.32
unclassified 2.76 8.84 11.75
various 0.04 0.23 0.26
Viruses 1.33 2.97 2.79
particle fraction color code
≥ 0.2 µm free living (fl)
≥ 3 µm particle associated (pa)
≥ 10 µm
particle sizes small medium large
B
Gammaproteobacteria
Alphaproteobacteria
Flavobacteria
Betaproteobacteria
Bacilli
Actinobacteria
Opitutae
var Bacteria
uc Bacteroidetes
uc Cyanobacteria
uc Bacteria
Cytophagia
Saprospiria
bacterial groups
legend of hierarchical top level
proteins assigned
to bacterial genera which have been
found in all respective fractions (colors)
Reinekea
Glaciecola
Colwellia
Paraglaciecola
uc Betaproteobacteria
Planktomarina
uc Porticoccaceae
Oceanicoccus
uc Gammaproteobacteria
Sulfitobacter
Arenicella
Ruegeria
Pseudomonas
Polaribacter
uc Rhodobacterales
uc Rhodobacteraceae
Flaviramulus
Candidatus Puniceispirillum
Tateyamaria
Lutibacter
Candidatus Pelagibacter
Tenacibaculum
Formosa
Arenitalea
Lacinutrix
Olleya
Algibacter
Streptomyces
uc Opitutae
var Bacteria
uc Bacteroidetes
uc Bacteria
uc Flammeovirgaceae
Synechococcus
Aureispira

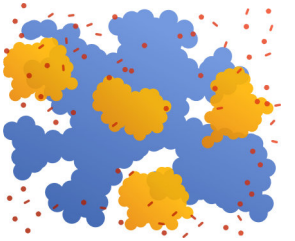
Figure 3

A

	abundance	fractions	fold ch. to smallest small	medium	large
carbon compound utilization					
Glycoside hydrolase family 1			2.1	1.0	6.7
Glycoside hydrolase family 13			1.0	1.5	1.1
Glycoside hydrolase family 16			1.0	1.3	1.4
Glycoside hydrolase family 45			+++	-	++
Starch utilization system SusD			1.0	1.8	1.0
Starch utilization system SusC			1.0	2.5	1.9
Glycosyl transferases			1.0	6.4	21.5
Sulfatases			2.0	1.0	4.1
transporters					
TonB-dependent transporter			1.0	2.4	1.6
ABC-transporter			3.2	1.2	1.0
TRAP-transporter			8.8	1.8	1.0
phosphorus supply					
Phosphonate ABC-type transport			1.8	1.3	1.0
Phosphate sodium symporter			1.1	1.0	1.1
Phosphate-selective porins			1.0	9.4	6.3
Phosphate ABC-type transport			2.0	1.2	1.0
Phytase			1.0	2.7	3.1

	abundance	fractions	fold ch. to smallest small	medium	large
adhesion					
Pili			1.0	3.7	3.3
other Adhesion proteins			1.0	3.0	2.0
motility					
Chemotaxis proteins			++++	-	-
Gliding motility			++++	-	+
Flagellum			1.0	4.0	1.7
other Motility proteins			++++	-	-
stress					
Oxidative stress			3.0	1.0	1.1
Heavy metal resistance			1.0	1.4	3.8
Antibiotic resistance			1.0	1.6	1.0
Antibiotic biosynthesis			-	-	+++
light driven energy generation					
Proteorhodopsin			1.0	1.1	1.1

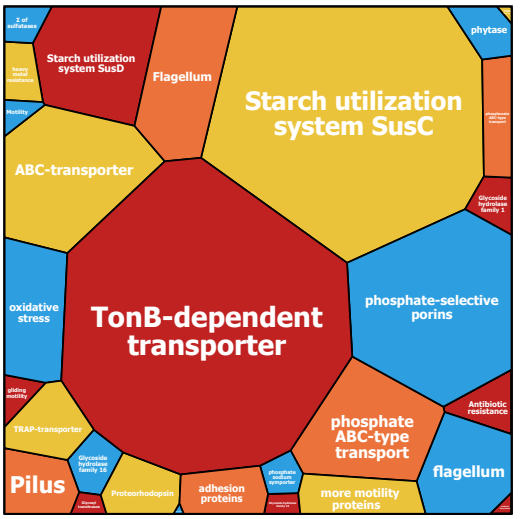
particle fraction color code



≥ 0.2 µm free living (fl) ≥ 3 µm particle associated (pa) ≥ 10 µm

particle sizes small medium large

B



functional groups the given proteins were assigned to

proteins assigned to bacterial genera which have been found in all respective fractions (colors)

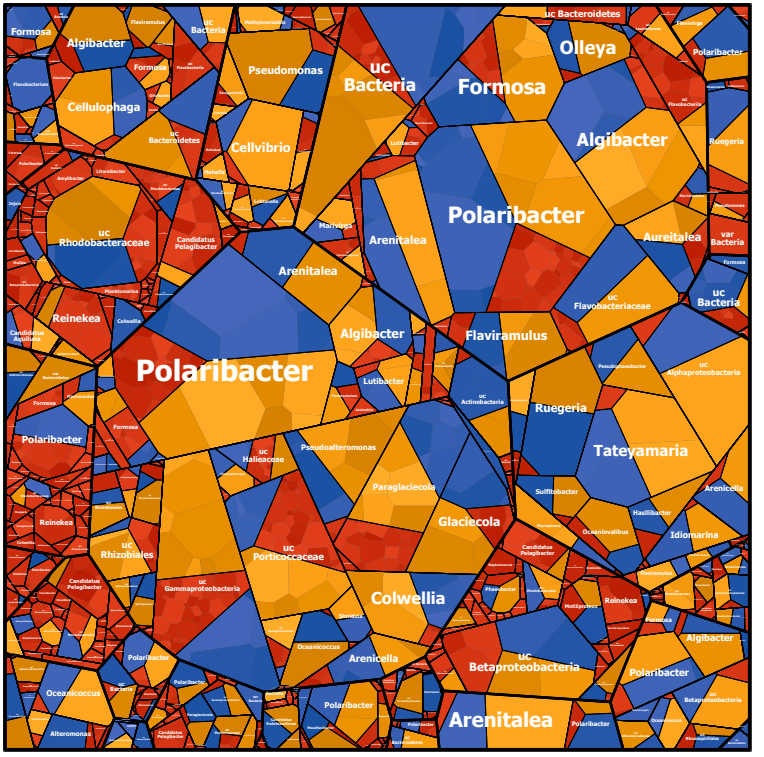


Figure 4

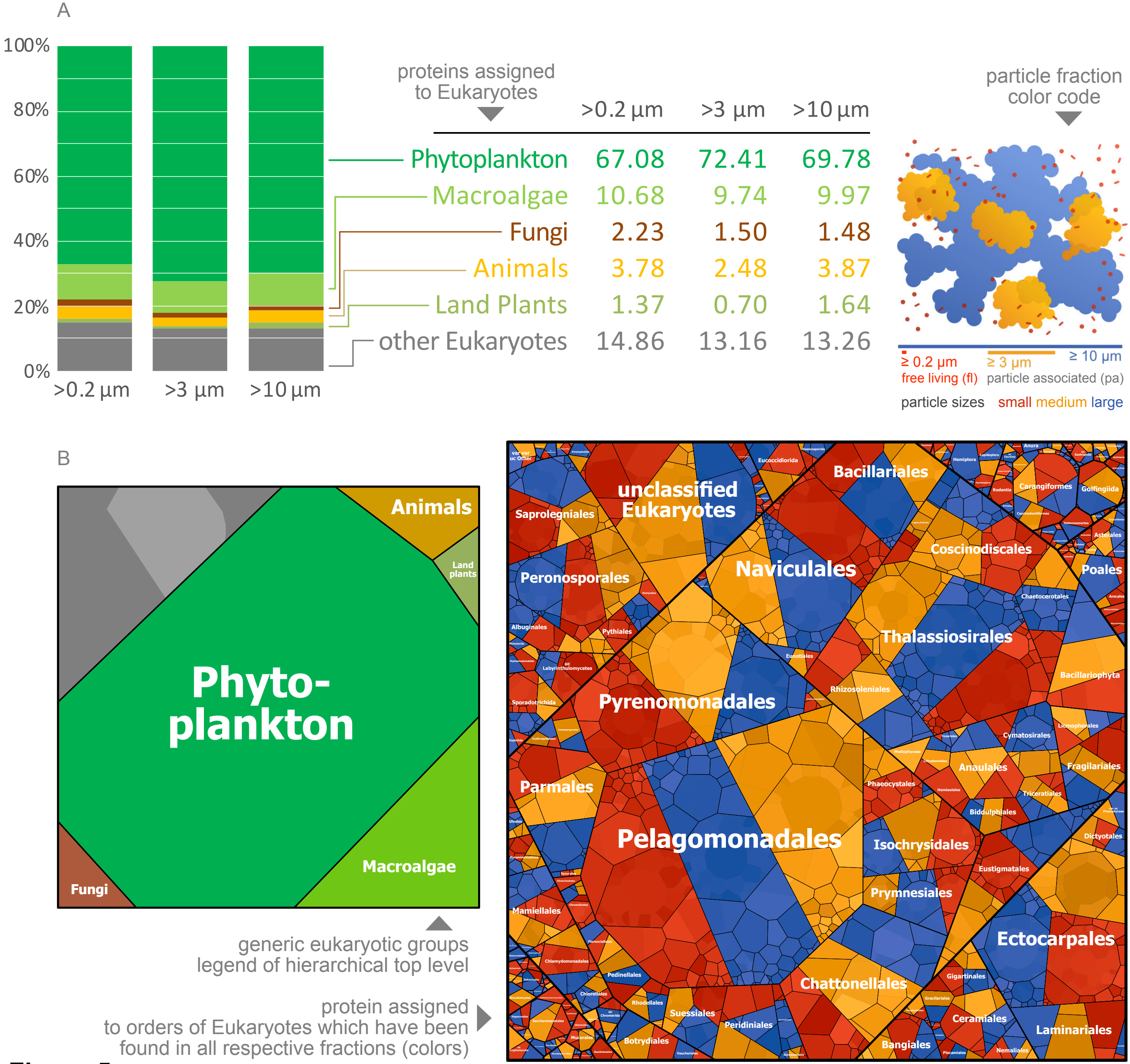

A
100%
80%
60%
40%
20%
0%
>0.2 µm
>3 µm
>10 µm
proteins assigned to Eukaryotes
>0.2 µm >3 µm >10 µm
Phytoplankton 67.08 72.41 69.78
Macroalgae 10.68 9.74 9.97
Fungi 2.23 1.50 1.48
Animals 3.78 2.48 3.87
Land Plants 1.37 0.70 1.64
other Eukaryotes 14.86 13.16 13.26
particle fraction color code
≥ 0.2 µm
free living (fl)
≥ 3 µm
≥ 10 µm
particle associated (pa)
particle sizes
small medium large
B
Animals
Land plants
Phyto-plankton
Macroalgae
Fungi
generic eukaryotic groups
legend of hierarchical top level
protein assigned
to orders of Eukaryotes which have been
found in all respective fractions (colors)
unclassified Eukaryotes
Bacillariales
Saprolegniales
Naviculales
Coscinodiscales
Peronosporales
Poales
Thalassiosirales
Pyrenomonadales
Bacillariophyta
Rhizosoleniales
Parmales
Anaulales
Fragilariales
Pelagomonadales
Isochrysidales
Prymnesiales
Mamiellales
Ectocarpales
Chattonellales
Ceramiales
Laminariales
Bangiales
Suessiales
Peridiniales
Botrydiales
Albuginales
Pythiales
Carangiformes
Golfingiida
Cymatosirales
Licmophorales
Eustigmatales
Dictyotales
Biddulphiales
Triceratiales
Phaeocystales
Gigartinales
Rhodellales
Pedinellales
Chaetocerotales
Sporadotrichida
Eunotiales

Figure 5