

# SCIENTIFIC REPORTS



OPEN

## MS analysis of a dilution series of bacteria: phytoplankton to improve detection of low abundance bacterial peptides

Received: 22 November 2017

Accepted: 6 June 2018

Published online: 18 June 2018

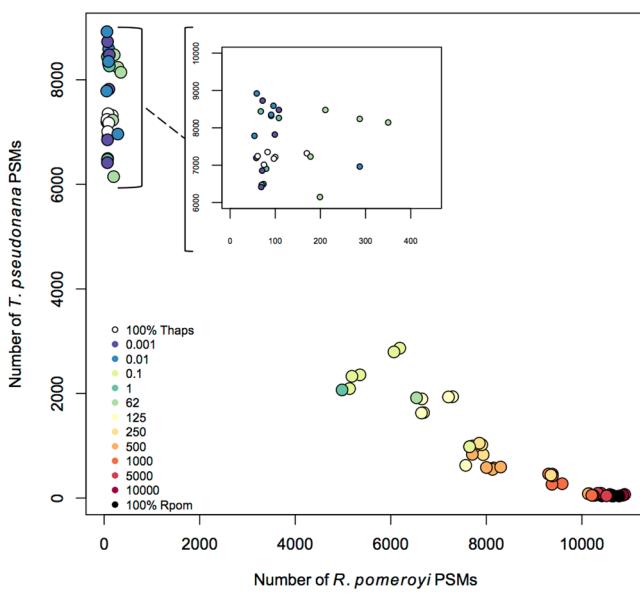
Emma Timmins-Schiffman<sup>1</sup>, Molly P. Mikan<sup>2</sup>, Ying Sonia Ting<sup>1,3</sup>, H. Rodger Harvey<sup>2</sup> & Brook L. Nunn<sup>1</sup>

Assigning links between microbial activity and biogeochemical cycles in the ocean is a primary objective for ecologists and oceanographers. Bacteria represent a small ecosystem component by mass, but act as the nexus for both nutrient transformation and organic matter recycling. There are limited methods to explore the full suite of active bacterial proteins largely responsible for degradation. Mass spectrometry (MS)-based proteomics now has the potential to document bacterial physiology within these complex systems. Global proteome profiling using MS, known as data dependent acquisition (DDA), is limited by the stochastic nature of ion selection, decreasing the detection of low abundance peptides. The suitability of MS-based proteomics methods in revealing bacterial signatures outnumbered by phytoplankton proteins was explored using a dilution series of pure bacteria (*Ruegeria pomeroyi*) and diatoms (*Thalassiosira pseudonana*). Two common acquisition strategies were utilized: DDA and selected reaction monitoring (SRM). SRM improved detection of bacterial peptides at low bacterial cellular abundance that were undetectable with DDA from a wide range of physiological processes (e.g. amino acid synthesis, lipid metabolism, and transport). We demonstrate the benefits and drawbacks of two different proteomic approaches for investigating species-specific physiological processes across relative abundances of bacteria that vary by orders of magnitude.

Microbes and their physiological processes are the foundation of many ecosystems as they regulate the flow of carbon and nutrients between different reservoirs. Although oceanic primary producers fuel the ecosystem with carbon and nitrogen in the form of dissolved and particulate organic matter (POM), heterotrophic bacteria act as the primary catalysts for remineralizing nutrients back into the ecosystem. This diversity of microbes possesses a bounty of enzymes, allowing them to utilize different energy and carbon sources from a wide range of substrates. To better understand how bacteria function as a collective community, impact their local environments, and control the long-term fate of carbon burial, a variety of 'omic techniques are now being applied to marine samples (i.e., genomics, transcriptomics, proteomics, metabolomics), providing an unprecedented view of microbial activities across ecosystems<sup>1</sup>. Although there is an expanded body of information on the potential microbes involved in the transformation of organic matter, our current understanding of how *in situ* heterotrophic bacterial communities actively transform, mobilize, and remineralize carbon and other essential elements is only recently being realized.

Metaproteomics is a new approach that has the potential to unravel bacterial-biogeochemical relationships. Due to the tight cellular regulation of protein synthesis and internal degradation, protein abundances reflect the metabolic status and response of a single cell or community of organisms at the time of harvest<sup>2</sup>. When correlated to time specific biogeochemical data, these measurements, set within the context of environmental stimulus, indicate how cells modify their metabolism to acclimate or adapt to the changing chemistry<sup>3</sup>. Although not a direct indication of enzymatic activity, a proteome analysis infers metabolic status and can provide details of protein abundance for processes of interest within an organism or community<sup>4-7</sup>. More direct measurements of protein expression and activity, such as post-translation modifications (PTMs) and enzyme activity, would provide even

<sup>1</sup>University of Washington, Department of Genome Sciences, Seattle, WA, 98195, USA. <sup>2</sup>Old Dominion University, Department of Ocean, Earth, and Atmospheric Sciences, Norfolk, VA, 23529, USA. <sup>3</sup>Neon Therapeutics, Boston, MA, 02139, USA. Correspondence and requests for materials should be addressed to B.L.N. (email: [brookh@uw.edu](mailto:brookh@uw.edu))



**Figure 1.** Peptide spectral matches (PSMs) from data dependent acquisition (DDA) of bacterial dilutions in phytoplankton. Points are colored by the cellular ratio of Rpom:1 Thaps cell, indicated in the legends. Number of Thaps peptide spectral matches (PSMs) per injection is plotted against number of Rpom PSMs. The insert shows a detailed view of the upper left-hand corner of the graph.

more accurate interpretations of protein activity and function<sup>8,9</sup>, but methodologies to scale these up to a community level are not developed as they are for full proteome abundance surveys and require significantly more biomass. The detection of protein abundance as a proxy for cellular activity continues to be utilized within the basic science, medical, and environmental fields as it provides an unbiased, rapid survey of the relative presence of thousands of proteins across metabolic pathways.

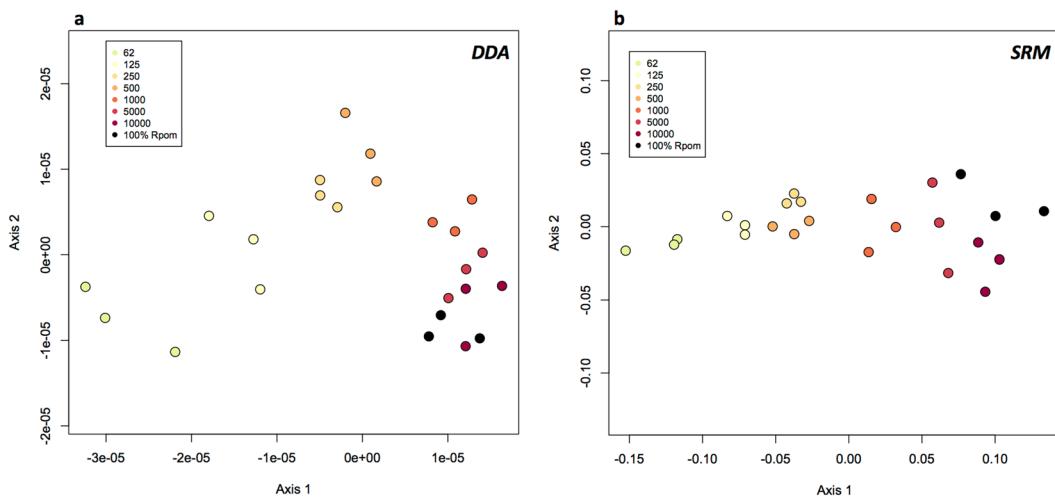
Historically, discovery-based proteomic techniques have been used to survey the microbial community proteome after fractionation by size<sup>4–7</sup>. The standard MS method for these discovery-style proteomics studies uses data dependent acquisition (DDA; e.g. shotgun or bottom-up proteomics), where peptide ions are selected for fragmentation and tandem mass spectrometry (MS2) based on their initial ion intensity. Resulting from the selection of the 10–20 most intense ions for analysis, this method can fail to detect low abundance proteins<sup>10,11</sup>. Examples across environments demonstrate how detection of relatively low abundance bacterial proteins (due to dynamism of expression level or organism abundance) can be obscured by co-occurring biological material, even though bacteria are present and biologically active<sup>12–15</sup>. Without a method to fully characterize bacterial metabolic processes across oceanic conditions, we cannot fully understand their roles in biogeochemical cycles, which inhibits full and accurate parameterization in predictive models.

A wide range of approaches that facilitate low abundance peptide detection have been investigated, with some techniques applicable to specific sample types (sample preparation<sup>16,17</sup>), while others are more broadly applicable (data acquisition<sup>7,10,18–21</sup>). Here we identify key proteins and peptides of interest from bacteria across a range of abundances with DDA and then use selected reaction monitoring (SRM) to improve the detection range to better include lower abundance peptides. Once peptide targets are identified, SRM assays can then detect and measure a suite of peptides with high sensitivity across a broad dynamic range<sup>22,23</sup> (e.g., down to attomolar per milligram of total protein). Specifically, this method allows the user to rapidly analyze many bacterial peptides of interest and thus quantify selected microbial metabolic processes<sup>11,24</sup>. This work builds upon DDA characterization of DOM<sup>25–27</sup> and SRM analyses of a small number of peptides from bacterial cultures and communities<sup>11,24,28–30</sup> to apply SRM to environments with highly variable abundances of different taxonomic groups.

To investigate detection limits of active microbial processes in eukaryote-dominated systems, a model system comprised of a well-known diatom species, *Thalassiosira pseudonana* (Thaps), and the heterotrophic marine bacterium, *Ruegeria pomeroyi* (Rpom), was used to mimic a broad range of bacteria:phytoplankton ratios. A simplified mixture allowed us to demonstrate and define possible variability in bacterial peptide signals among eukaryotic biomass. By tracking this variability, we were able to identify and select bacterial peptides that were below the DDA detection limit through the dilution series and test the ability of SRM to expand the range of detectability in a mixture of proteomes. Because this is a taxonomically complex system, we also emphasize the impact of identical peptide sequences and taxonomically indistinct peptides sequences on the process of SRM assay development.

## Results

**Proteomic differences across the *Ruegeria pomeroyi*: *Thalassiosira pseudonana* gradient.** Across all bacterial dilutions (*R. pomeroyi*: *T. pseudonana* 1:1000, 1:100, 1:10, 1:1, 62:1, 125:1, 250:1, 500:1, 1000:1, 5000:1, 10000:1), 3923 proteins were identified using data dependent acquisition (DDA), with 1967 attributed to *Thalassiosira pseudonana* (Thaps) and 1956 attributed to *Ruegeria pomeroyi* (Rpom). The number



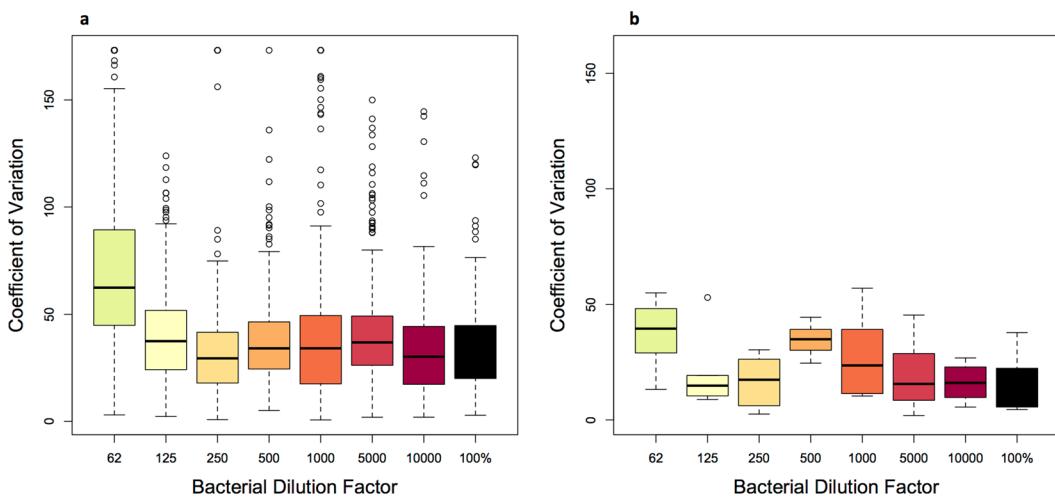
**Figure 2.** Nonmetric multidimensional scaling plots (NMDS) of proteomic data (normalized spectral abundance factor) from Rpom:Thaps dilutions for DDA (a) and SRM (b) analyses. Points are colored by cellular ratio Rpom:1 Thaps cell.

of Rpom peptide spectral matches (PSMs) increased linearly with the cellular ratio of Rpom cells greater than 62:1 Rpom:Thaps (Fig. 1). Below the 62:1 cellular ratio, most bacterial peptides are undetectable within a DDA experiment. Using non-metric multidimensional scaling (NMDS), there is a suite of 36 Rpom proteins within the DDA data matrix that drives a trend along axis 1, the axis that drives the separation of samples along the cellular dilution gradient ( $R = 0.8684$ ,  $p = 0.001$ ; Fig. 2a). These proteins are involved in metal and nucleotide binding. Peptides from these 36 proteins drive this trend in the NMDS because they are under-sampled during MS2 selection, possibly due to low initial relative abundances, HPLC retention time, peptide size, amino acid composition, charge state, and/or hydrophobicity index<sup>31</sup>. Without the collection of a MS2 spectrum, the peptide/protein would remain undetected in a DDA experiment, thus biasing a full proteome analysis and the processes reflected by their expression. This is a recognized limitation in proteomics, however it can be overcome if the researcher begins the experiment with knowledge of what processes, or proteins, they are interested in monitoring.

**Identical peptide sequences in SRM assay development.** Peptide sequences can be conserved across taxonomic groups, ranging from species to kingdoms. Typical MS-based bottom-up proteomics (i.e., peptide identification leads to protein inference) followed by database searching identifies an amino acid sequence from an observed tandem mass spectrum. In a taxonomically complex sample, however, a single peptide sequence might correspond to 10s to 1000s of species<sup>29</sup>. When searching for the specific signal of a biological or a taxonomic group, it is critical to examine sequence identity in a mixed community sample to eliminate non-discriminatory peptides shared across taxa<sup>29</sup>. The Rpom proteome has 400,708 putative tryptic peptides and the Thaps proteome has 1,887,118 putative tryptic peptides. If all putative peptides were to be detected in an MS experiment, there could be an overlap of 9,328 peptides due to sequence similarity between the two proteomes. However, not all peptides predicted from the genome exist in the digested protein lysates, nor can they all be detected. Across all dilutions 31 peptides with identical sequences were detected in collected tandem mass spectra (Supplementary Table S1). These taxonomically ambiguous peptides are predominantly present in proteins that have basal, highly conserved functions (e.g. ATP synthase, elongation factors, succinyl-CoA ligase).

**Determination of peptide taxonomic specificity.** When designing SRM assays for a wider range of environmental contexts, the full taxonomic identity of each peptide must be characterized to determine its taxonomic specificity. UniPept<sup>32</sup> is a web application that allows users to compare provided lists of peptides to the large non-redundant UniProtKB peptide database to reveal taxonomic specificity based on known peptide sequences. For example, UniPept classifies peptide TVINWAQNAEIFR (from the protein Q5LPJ5, cobalt chelatase) to be identifying of the Rhodobacteraceae family. Most of the Rpom peptides in the SRM assay ( $n = 28$ ) were found to be specific to the species *Rugeria pomeroyi*, 1 peptide was found across the *Rugeria* genus, 21 were specific to Rhodobacteraceae (at the family level), 3 are indicative of Alphaproteobacteria (class), 2 were found in all Proteobacteria (phylum), and 6 are found across all Bacteria (kingdom). For the purpose of designing a peptide based SRM assay for metaproteomic samples, a more complete and sample-specific metagenomic-predicted protein database should be investigated for taxonomic specificity since UniPept does not include recent or site-specific protein entries<sup>33</sup>.

**Bacterial biomarker development using targeted proteomics.** Targeted proteomics assays were developed from 64 peptides of interest chosen to expand the range of detectability of the bacterial peptides when they were undetectable among the dominating eukaryotic peptides (see Methods section). From the 64 peptides of interest, 305 Rpom peptide transitions (peptide fragments) were detected (Supplementary Table S2).



**Figure 3.** Integrated peak area coefficient of variation values across: (a) all peptide transitions ( $n = 305$ ) for each bacterial dilution and (b) 10 transitions with the lowest sum of CV across all samples, representing the most reliable transitions (Supplementary Table S5). These 10 peptide transitions with the lowest CV correspond to 7 peptides from 7 proteins involved in diverse functions (e.g., purine metabolism, fatty acid biosynthesis, transcription). X-axis values correspond to the bacterial dilutions, indicated by cellular ratio Rpom:1 Thaps cell. The boxes represent the upper and lower quartiles of the data distribution; the horizontal black line represents median value; “whiskers” extend to the greatest and least values, excluding outliers; open circles represent outliers ( $\pm 1.5$  times the upper or lower quartiles).

Two hundred seventy-five peptide transitions (90%) were detected across all dilutions in at least one replicate. Additionally, NMDS and ANOSIM revealed a statistically significant difference in peptide biomarker abundance along the bacterial dilution gradient ( $R = 0.778$ ,  $p = 0.001$ ; Fig. 2b).

Based on Pearson’s correlation coefficients between SRM peptide transition abundance and bacterial dilution factor, 199 (out of 305 total) transitions had a significant correlation with relative bacterial cell abundance (Pearson’s  $r \geq 0.755$ ; Supplementary Fig. S1). These 199 peptide transitions could be detected and used both as biomarkers of bacterial function and relative abundance within a mixed sample in this particular environmental context. The peptide transitions were chosen as representative biomarkers originating from proteins across cellular functions including translation, amino acid and lipid biosynthesis, membrane transport, signaling, vitamin B biosynthesis, and ATP metabolism.

An examination of coefficients of variation (CVs) for the Skyline-derived integrated peak areas for all the transitions finds that peptide and peptide transition selection is critically important and can significantly affect data reliability. A low CV is desirable because it indicates a consistent peptide transition signal, in this case across the integrated peak areas of biological replicates of the same bacterial dilution. In the lowest bacterial abundance sample, CVs were  $>100\%$  for 19% ( $n = 55$ ) of the chosen peptide transitions. Samples with higher relative bacterial abundances yielded a trend of improved CV across all peptide transitions considered (Fig. 3). In dilutions of 250:1 (bacteria:diatom) or greater, 1–5.6% of the transitions were reported with high CVs ( $>100\%$ ). Quantification of bacteria-specific peptides was more reliable at greater relative cellular abundances.

## Discussion

The range of bacteria:phytoplankton ratios examined here clearly demonstrate that SRM detects bacterial peptides across a wide range of abundances and physiological processes. We demonstrate how mass spectrometry-based proteomics has the flexibility and selectivity to detect known peptides of individual bacterial species with a lower relative protein abundance living amongst larger volume species with higher relative abundances of proteins (phytoplankton), but this method could also be adapted to other mixed communities. We selected informative biomarker peptides discovered during the DDA analysis to characterize bacterial processes across a wide range of phytoplankton concentrations. These biomarker peptides were then analyzed using SRM, thereby building on existing environmental applications for this pipeline<sup>24,28,29</sup>. In many environments, researchers desire the ability to select and detect specific markers for the presence of an organism within a mixed community. Although SRM is a proven technique in open ocean samples enriched for microbial communities using filtration, environments with low microbial to particle/detrital biomass ratios (e.g., during phytoplankton blooms) offer extra challenges. MS-based methods have the resolution, selectivity, and sensitivity to be an ideal analytical tool for these questions with the added benefit of sequence specificity that can reveal information on biological function and even taxonomy<sup>32</sup>.

The DDA-to-SRM pipeline demonstrated here allows users to first develop testable hypotheses (DDA) on a small number of samples and then rapidly apply them to multiple, complex, mixed community environmental samples. Similar pipelines have been previously applied to explore selected peptides to test hypotheses on specific organism-driven processes and proteins, such as iron preservation<sup>28</sup>, iron limitation<sup>30</sup>, and nutrient stress<sup>24</sup>. The

peptides of these targeted proteins were detected at abundances as low as 0.0003–20 fmol/ $\mu$ g protein<sup>28,30</sup>. We expanded upon this foundational work by targeting peptides across a broad range of physiological processes that were technically difficult to detect due to low abundance. These advances show that SRM can be leveraged to interrogate a suite of important physiological processes in species present at low cell counts when peptides specific to the species and process of interest are known. Although experimentally simplified, results demonstrate that SRM can detect peptides in samples with low concentrations of a bacterium that co-occurs with phytoplankton, suggesting that the method could work on unfractionated samples collected during a phytoplankton bloom<sup>34–36</sup>. With this proof of concept, investigators can apply SRM assays to mixed eukaryotic-prokaryotic communities. Assurance of detection and quantification of a signature peptide within a complex sample is increased if the peptide is observed in either DDA or DIA discovery experiments prior to SRM assay development. The workflow can be economized by pooling samples in the DDA or DIA analysis to limit the number of LC-MS/MS analyses and experimentally verifying the presence of high-responding peptides<sup>37</sup>. Although assay development is more streamlined if peptides are first characterized in a preliminary, discovery-based MS step, there will be cases where peptides from a candidate protein of interest are not observed in the discovery phase. These candidate proteins could be selected through genomic predictions, literature discovery, or data mining. Investigators can then utilize one of the many proteotypic calculators that model peptide sequence physiochemical properties to predict high-responding peptide targets for SRM assay development<sup>31</sup>. Additionally, as the available spectral library databases are growing, it is possible to find previous MS experiments that provide charge state, transition lists, and retention times<sup>38,39</sup>. Development of environmental-specific data repositories for this kind of work are currently under construction and will rapidly excel the field<sup>40</sup>. Further work can expand this foundation to more complex, multi-taxa communities.

Taxonomic characterization of peptides in an SRM assay could facilitate our understanding of functional redundancy and specificity across taxonomic groups within a complex bacterial community. We informally demonstrated this potential by analyzing our SRM peptides with UniPept to characterize taxonomic specificity of each peptide. The largest group of peptides are specific to our species of interest, *R. pomeroyi*, but if we were to apply our assay to a mixed environmental community we would also glean information from more general taxonomic groups, from the genus to kingdom levels. These results underline the importance of understanding the taxonomic specificity of peptides when selected for an assay of a mixed community, but also the flexibility of this type of analysis. Carefully selected peptides could reveal the extent of functional redundancy and specificity within a natural community, which would significantly augment our understanding of ecosystem function.

In a standard DDA experiment, only the most relatively abundant peptides are selected for full MS2 analysis<sup>41</sup>, however the advantage is that no prior knowledge of protein sequence or detectability by the mass spectrometer are necessary. The detection limitation of DDA is empirically demonstrated here when the peptide signals from bacterial proteins in a mixed prokaryotic-eukaryotic dilution series were reduced when bacterial abundance was below the 62:1 cellular ratio in the DDA dataset (Fig. 1). This ratio (62:1) and below represent some of the typical oceanic concentrations of bacterial cells. Isolation and selectivity of bacteria-specific peptides using SRM significantly improved the detection of targeted bacterial peptides in all experiments from the dilution series. Similarly, targeted analysis detected B12 synthesis peptides that were below the detection of DDA in a Ross Sea bacterial community<sup>11</sup>. A technical limitation of DDA tandem mass spectrometry is the stochastic nature of how ions are selected for tandem mass analysis (i.e., MS2 spectral generation). Because DDA is an analysis that is strictly tethered to the peptide elution profile, co-eluting peptides will interfere with MS2 collection and subsequent identification. As a result, peptides from abundant proteins in the bacteria fraction may decrease in relative intensity as eukaryotic peptides fill the MS1 survey scan. This would lead to fewer, or zero, MS2 collected on those bacterial peptides, ultimately leading to a potential bias in biological data due to protein detectability. In the demonstrated application, DDA or DIA could be applied to a lab culture or sample of a mixed community to detect a broad range of peptides in an unbiased manner to provide guidance in selecting appropriate SRM targets. Subsequently, an SRM peptide assay could be applied across environmental scenarios to capture the dynamism of organism presence and protein expression.

The loss of peptide signal from low abundance proteins in complex mixtures is an analytical challenge that researchers have been trying to resolve for years<sup>16–18,42,43</sup>. Peptides from a candidate protein may be under-sampled in a mass spectrometry experiment as a result of native protein expression, relative abundance of the organism in a mixture, HPLC retention time, or inherent physicochemical properties such as sequence length, amino acid composition, and hydrophobicity. Valuable information regarding bacterial contributions to ecosystem processes across a dynamic range of expression (whether that be the human gut or open ocean) can be lost due to technical limits of detection. Low abundance peptides in DDA experiments can be obscured by peptides from the host of a microbiome<sup>12</sup> or by the relatively greater abundance of other species within a community<sup>13</sup>. Similarly, in both planktonic and benthic samples of mixed detritus consisting of phytoplankton and bacteria in the Bering Sea, only a few prokaryotic proteins were identified, even though high bacterial counts were observed and bacteria were actively degrading dead phytoplankton<sup>14,15,44</sup>. In these studies, only 2–7 bacteria-specific proteins were detected using DDA, making it difficult to assess bacterial metabolic processes and mechanistic controls, and to quantify the extent of their contribution to overall ecosystem processes. This low detection of bacterial peptides translates to interpretations of bacterial community metabolism based primarily on highly abundant proteins. SRM approaches detect a diverse suite of proteins across abundance ranges that vary by orders of magnitude, including very low intensity peptides such as those that were measured across all samples in this experiment (6 peptides from 2 proteins).

Low quantities of proteins were consistently detected across most dilutions using a targeted approach, however coefficients of variation were high for many of the peptide transitions. As noted by others, this suggests that analytes must be screened for their utility in absolute quantification<sup>45</sup>. It can be difficult to achieve reliable quantification of peptides that occur at relatively low abundance in complex mixtures using MS methods due

#Rpom cells: # Thaps cells	100% Thaps	62	125	250	500	1000	5000	10000	100% Rpom
Rpom cells	0	$2.08 \times 10^9$							
Thaps cells	$8.33 \times 10^6$	$3.33 \times 10^7$	$1.67 \times 10^7$	$8.33 \times 10^6$	$4.17 \times 10^6$	$2.08 \times 10^6$	$4.17 \times 10^5$	$2.08 \times 10^5$	0
Protein ratio		4.9	9.8	19.7	39.3	78.8	393.0	788.0	

**Table 1.** Cell counts of the bacteria *R. pomeroyi* (Rpom) and diatom *T. pseudonana* (Thaps) mixtures produced by serial dilution and the calculated ratio of estimated protein contributed from each source (Rpom g protein/Thaps g protein). The Rpom cellular protein content is estimated from<sup>46</sup> and Thaps cellular protein content is estimated from<sup>84</sup>.

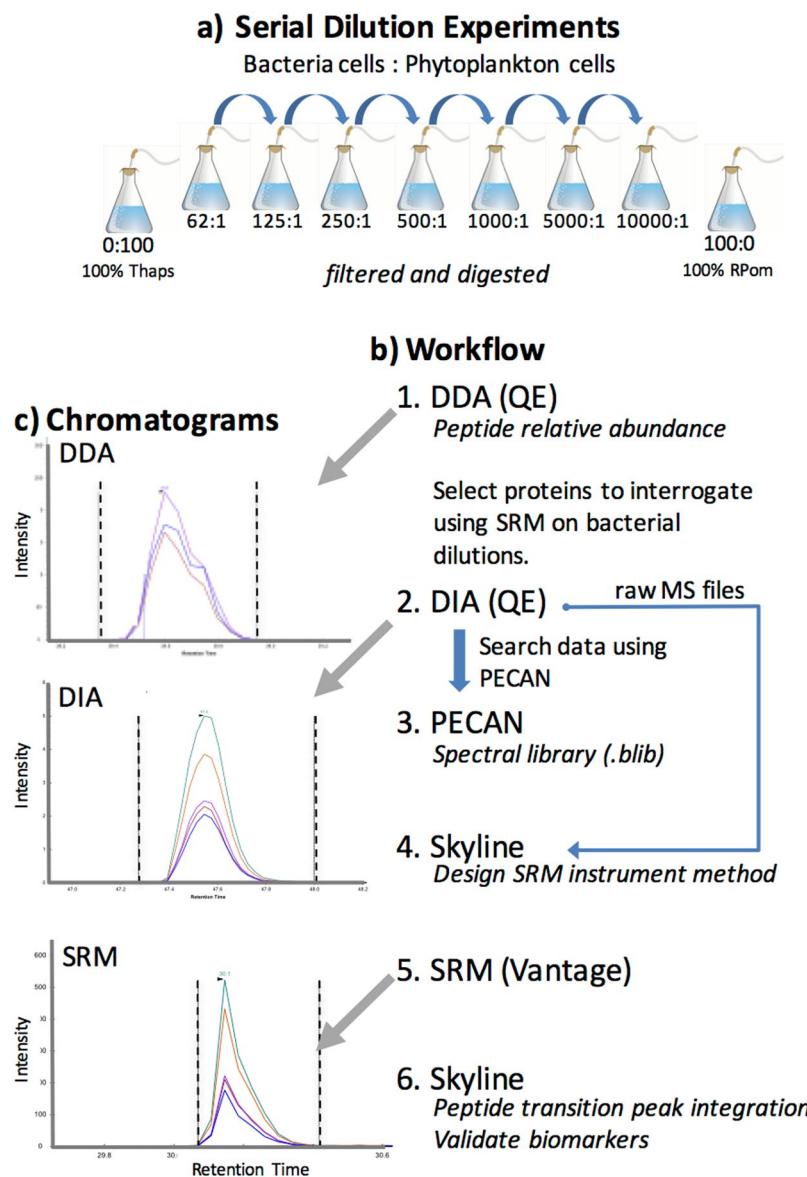
to co-eluting, interfering peptides or misidentified and misquantified peptide transitions<sup>46</sup>. There is no correlation between high CV and peptide transition retention time, suggesting that retention time does not affect peptide transition stability and that there is little interference from co-eluting peptides in target peptide detection. Additionally, peptide transitions yielding the highest CVs were observed in the sample with the lowest bacterial abundance, suggesting that peptide transition quantification accuracy decreases with the target species relative abundance in a sample. In previous work, CVs for SRM assays on low abundance peptides in single-species protein lysates ranged from 5.4–16.8%<sup>23</sup>. Peptide inter-run variability is important to consider in assay development because a target with high inter-run variability would lead to low confidence peptide quantifications. Decreasing the number of peptides and transitions per MS injection would improve quantifications for low abundance peptides by increasing the dwell time of the triple quadrupole for each transition<sup>22</sup>.

Despite these promising results, SRM is not the panacea to low abundance peptide and protein detection. SRM assays designed for complex environmental communities where species have little available sequence data, can require investment in preliminary detection of peptides using DDA or DIA<sup>47</sup>. Even this preliminary step can be biased, with the several alternative methods now available that allow for the development of SRM assays without the requirement of MS-based peptide detection<sup>48–51</sup>. To detect low abundance peptides using only DDA, sample fractionation should be considered to reduce peptide complexity in MS experiments despite their varied recoveries. Chemical fractionation to remove interfering matrices, depletion of high abundance proteins, multiple filtrations steps, protein size fractionations (including gel separations), and selective antibody removal are all examples that would decrease sample complexity in MS experiments, allowing access to lower abundance proteins for DDA experiments. Additionally, not all techniques are applicable to every system; for example, selective antibody removal<sup>17</sup> is common in some model systems but would be challenging to apply to complex, uncharacterized systems when interfering, high abundance protein sequences are unknown. Gas phase fractionation within the mass spectrometer with DDA experiments has been shown to dramatically improve the depth of proteome discovery<sup>10,52</sup>. DIA has been reported to improve peptide discovery by 94%<sup>10</sup>, detects many more peptides than DDA, and does not require sample fractionation or enrichment; however, current bioinformatic pipelines for complex communities are less established than for DDA or SRM. DIA does not limit the mass spectrometer to collecting MS2 only on ions with intense precursor signals, as in DDA, and thus dramatically increases the dynamic range of the MS and increases overall proteome coverage by detecting peptides that occur at lower abundances<sup>10,52</sup>. However, due to the multiple injections per sample required to cover the full range of masses, DIA can require significantly more MS time and starting material than SRM, which may not be feasible with environmental samples<sup>37</sup>. Additionally, once an SRM assay is developed, it is imperative to ensure the specificity of the peptide transitions monitored, especially since multiple peptides can co-elute resulting in different fragments within an MS2 selection window. A peptide's presence in a sample can be verified with total confidence through the inclusion of a synthesized stable isotope labeled peptide of interest. Specific research goals and limitations of experimental design and samples need to be considered when choosing DDA, DIA, or SRM for characterizing a proteome.

In the oceanic ecosystem, many species coexist and compete as they metabolize, degrade and recycle organic material. Standard MS techniques (DDA) can capture proteomic profiles of the most abundant proteins within a system, but since individual microbe protein contributions vary, the desired signal may not be detected. Through applications of SRM to samples containing some realistic cellular ratios of bacteria and phytoplankton, and with peptides previously characterized on a mass spectrometer, we assessed the ability of targeted proteomics to detect selected metabolic processes of an organism present at low cell counts. In DDA mode, the ability of the mass spectrometer to detect bacterial peptides declined with a reduction in relative bacterial abundance; yet targeted SRM analysis reliably detected the metabolic signals of our desired bacterial species of interest across the full dilution series. Although SRM can detect and quantify peptides down to the attomolar level, knowledge of the peptide sequence detectability is required for assay development. This can be obtained with proteotypic peptide calculators that predict detectability based on physio-chemical properties<sup>31,37,53–55</sup>, or experimental determination using DDA or DIA, or the mining of previously published spectral libraries<sup>56</sup>. The reliability of these assays can be determined via a first-round of SRM analysis, with subsequent rounds dedicated to refining the set of peptides included in the assay. These assays could be used to probe microbial metabolic processes across a range of environments to better understand the ecosystem-level transfer of essential nutrients<sup>57</sup>.

## Methods

**Dilution series.** The marine diatom *Thalassiosira pseudonana* (Thaps, CCMP1335) was grown in f/2 media<sup>58,59</sup> with autoclaved and filtered artificial seawater (salinity 30) at ambient room temperature (18–22 °C) under a 13:11 hour light:dark schedule. Diatom growth was monitored by absorbance measurements at 550 nm (Spectronic Educator, Flinn Scientific, Batavia, IL). Cell counts and cellular health were checked throughout the



**Figure 4.** Illustration of experimental setup and workflow for mass spectrometry data acquisition and analysis. (a) Serial dilutions were completed using bacterial cells (RPom) as the diluent (see text). Dilution was based on cell counts to achieve cellular ratios of Rpom (*R. pomeroyi*) to Thaps (*T. pseudonana*). Each serial dilution was then lysed and proteins were digested prior to MS experiments. (b) MS experimental workflow: 1. Data dependent acquisition (DDA) was performed on the Q-Exactive-HF (QE) to assess the limit of detection for a standard, discovery-driven proteomics experiment. 2. Data independent acquisition (DIA) was also completed on the QE to create spectral libraries for selected reaction monitoring (SRM) method development. 3 & 4. These spectral libraries were analyzed with PECAN and Skyline was used to select optimal transitions and to design an instrument method for SRM analyses. 5. SRM was completed on the TSQ Vantage for 309 bacterial peptide transitions. 6. Peptide transition detection and quantification was performed in Skyline. (c) The chromatograms of peptide IPSAVGYQPTLATDMGAMQER (from protein Q5LNP1) were collected using the 3 different MS approaches (DDA, DIA, and SRM) on bacterial dilution 5000:1. Black vertical lines indicate peak integration boundaries, and colored peaks represent the different transitions (i.e. peptide fragments) collected.

growth cycle with a hemocytometer on an Olympus Optical epifluorescence microscope. The culture was harvested during exponential growth.

The marine heterotrophic bacterium *Ruegeria pomeroyi* (Rpom, NCMA B3) was reconstituted in autoclaved and filtered 0.5 YTSS media<sup>60</sup> and slowly transitioned into a low carbon (as 0.625 mM glucose) medium over multiple generations. Cultures were grown under axenic conditions at room temperature and bacterial growth tracked by absorbance measurements at 600 nm. The culture was harvested during early stationary phase for experimental mixtures.

To mimic a wide range of oceanic POM samples, a dilution series involving mixtures of Rpom and Thaps was created using different cellular ratios of bacteria:phytoplankton based on previous publications of bacteria and

phytoplankton counts in (1) mesocosm experiments<sup>61,62</sup>, (2) before, during, and after phytoplankton blooms<sup>34–36</sup>, and (3) as a function of depth<sup>63</sup>. In nature, these ratios can vary by an order of magnitude<sup>34,35,63,64</sup> and we extended our dilution ratios to better define the upper and lower limits of mass spectrometry detection within these complex mixtures. Aliquots of cultures at concentrations of  $10^8$  cell ml<sup>-1</sup> (Rpom) and  $10^5$  cell ml<sup>-1</sup> (Thaps) were mixed to yield desired Rpom:Thaps ratios (outlined in Table 1) in triplicate. Samples were subsequently filtered onto 47 mm, 0.2  $\mu$ m Nucleopore polycarbonate filters (Whatman, Maidstone, UK) to simulate simultaneous *in situ* ocean collections of these mixtures onto a 0.2  $\mu$ m filter. Once filtered, cells were killed with a 5 ml rinse of cold 10% TCA before the filtered samples were frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$ . After protein digestion, additional dilutions were created from these original samples based on calculated cell counts to yield Rpom:Thaps ratios of 1:1000, 1:100, 1:10, and 1:1 for the DDA analysis.

**Protein extraction.** Proteins were extracted from filters by shaking the filters suspended in 500  $\mu\text{l}$  of 6 M urea in a bead beater with no beads (repeat 3 times:1 min shaking; ice 5 minutes). After removing the filters from the liquid, cells were lysed using a sonicating probe<sup>3</sup>. This method removed an average of 42% (range: 9–94%) of total proteins from the filter, determined by amino acid quantification (Supplementary Methods and Supplementary Table S3).

Protein concentrations were measured using the BCA assay (Pierce, Thermo Fisher Scientific), following the manufacturer's protocol. All samples were analyzed in triplicate and concentrations were averaged for a final protein concentration. Digestions of 100  $\mu\text{g}$  of protein were completed following<sup>3</sup>.

**LC-MS/MS and protein inference: DDA.** Liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS) was completed on a Q-Exactive-HF (QE: Thermo Fisher Scientific) in technical duplicate analyses for each sample using data dependent acquisition (DDA) on the top 20 precursor ions (Fig. 4). The analytical column was 20 cm long and packed in house (3  $\mu\text{m}$  C18; Dr. Maisch) with a 3 cm long trap (3  $\mu\text{m}$  C12; Dr. Maisch). Peptides were eluted using a 5–35% ACN gradient over 60 minute at 300 nl/min flow rate. MS1 ions were collected in the scan range of 400–1400  $m/z$ . Automatic gain control threshold was set at  $1 \times 10^6$  for MS1 and  $5 \times 10^4$  for MS2 and dynamic exclusion of 30 s was used for MS2. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository<sup>39</sup> with the dataset identifier PXD004799 (<http://www.proteomexchange.org/>).

**Database search parameters.** Completed proteomes for *Thalassiosira pseudonana* and *Ruegeria pomeroyi* were downloaded from Uniprot (7/2013; [www.uniprot.org](http://www.uniprot.org)). These databases were concatenated with 50 common contaminants, yielding a protein database of 17,395 proteins. To assign spectra to peptide sequences, correlative database searches were completed using Comet v. 2015.01 rev. 2<sup>65,66</sup>. Comet parameters included: Trypsin enzyme specificity, semi-digested, allowed 1 missed cleavage, 50 ppm mass tolerance, cysteine modification of 57 Da (resulting from the iodoacetamide) and modifications on methionine of 15.999 Da (oxidation). Minimum protein and peptide thresholds were set at  $P > 0.9$  on Protein and Peptide Prophet<sup>67</sup>. Protein inferences from the whole-cell lysates were accepted by ProteinProphet if the thresholds noted above were passed, two or more peptides were identified, and at least one terminus was tryptic<sup>68–70</sup>. Normalized spectral abundance factor (NSAF) was calculated<sup>71</sup> for all inferred proteins<sup>72</sup>.

**Proteomic differences across Rpom:Thaps gradient.** Non-metric multidimensional scaling (NMDS) in the vegan package<sup>73</sup> in R v. 3.2.3<sup>74</sup> was applied to assess tightness of technical replicates (Supplementary Fig. S3), excluding the 100% Thaps sample. Technical replicate analyses of individual samples showed consistent proteome characterization so spectral counts were averaged across technical replicates to calculate final NSAF and for NMDS and ANOSIM analyses in the vegan package in R.

The Rpom proteins that had significant loadings ( $p = 0.00099$  and loading  $> 0.99$ ) along the NMDS axis that differentiates the samples based on ratio of Rpom:Thaps (Supplementary Table S4) were analyzed for enrichment of specific biological processes in DAVID v. 6.7<sup>75,76</sup> using the Rpom proteome as the background protein list. These are proteins that are increasingly difficult to detect with DDA methods when bacteria are at relatively low abundance.

**In silico analysis of peptide sequence identity and taxonomic specificity.** The Thaps and Rpom complete proteomes were digested *in silico* using the Protein Digestion Simulator v. 2.2.5350.26597 from PNBL (omics.pnl.gov) to determine if there would be peptide sequence homology between organisms. The following settings were used: Minimum fragment mass = 400, maximum fragment mass = 6000, minimum residue count = 5, max missed cleavages = 3, hydrophobicity mode = Hopp and Woods. This created two files, one containing putative Thaps tryptic peptides and the other containing putative Rpom tryptic peptides.

Rpom peptides selected for SRM analysis (see below) were compared to all known bacterial peptide sequences in UniPept<sup>32,77</sup>, which searches peptide sequences against the entire UniProt database, on February 13, 2017. This analysis gives the taxonomic specificity of each peptide, i.e. if a peptide is species-specific or found across bacterial taxa at a higher taxonomic level.

**LC-MS/MS: DIA, targeted proteomic assay development, and SRM.** Targeted proteomics assays were developed and tested to determine at what point in the dilution steps the bacterial peptides were undetectable among the dominating eukaryotic peptides. Based on the DDA analysis of the bacterial dilution series, peptides in the following categories were selected for targeted assays: (1) peptides present across biological replicates and dilutions, (2) unique peptides identified only in low Rpom:Thaps dilution (i.e., phosphate-specific transport

system (Q5LS18) and ABC transporter, ATP-binding (Q5LLS4)), (3) peptides that drive the differences observed in the NMDS (see Methods, Proteomic differences across Rpom:Thaps gradient). These categories yielded an assay of 64 peptides derived from 24 proteins (Supplementary Table S2). Predetermined peptides with identical sequences between the Rpom and Thaps proteomes were not present in the list of peptides of interest for targeted proteomics.

The three technical replicates from the DDA experiment for the dilution of 5000 Rpom: 1 Thaps cell were pooled in equal quantities to create two new technical replicates for data independent acquisition (DIA) on the QE (Thermo). Each sample included a spiked-in internal quality control peptide standard (375 fmol Peptide Retention Calibration Mix; Pierce, hereafter referred to as “QC”). Sample injections for all DIA experiments included 1  $\mu\text{g}$  protein plus the internal standard in a 2  $\mu\text{l}$  injection. DIA experiments were completed using a 27 cm analytical column with a 3 cm pre-column (3  $\mu\text{m}$  C18; Dr. Maisch). Technical replicates were collected in 4  $m/z$  isolation width windows in 125  $m/z$  ranges (400–525, 525–650, 650–775, 775–900)<sup>10</sup>. For each method, a gradient of 5–80% ACN over 90 minutes was applied for peptide spectra acquisition. Raw data can be accessed via ProteomeXchange (<http://www.proteomexchange.org/>) under identifier PXD004758.

To generate spectral libraries for targeted method development, Peptide Centric Analysis was completed with the software program PECAN<sup>78</sup>. Input files included the list of peptides generated for targeted proteomics, as described above, and the mzML files generated from the raw DIA files using MSConvert<sup>79</sup>. PECAN correlates a list of peptide sequences of interest with the acquired DIA spectra to locate the peptide-specific spectra within the acquired DIA dataset.

The PECAN.blib output file was then imported into Skyline daily v. 3.5.1.9706<sup>80</sup> for targeted method development. The targeted method development workflow, including screenshots, can be found in Supplementary Methods. Peptide transitions are defined as the reproducible fragments of peptides that are produced during the MS2 scan in a mass spectrometer<sup>81</sup>. Peptide transitions were selected if peak morphology was uniform and consistent across the MS2 scans for both technical replicates. Peptides were selected for targeted analysis if they had >3 high quality transitions and >3 peptides per protein. Only 4 transitions per peptide were selected for targeted analysis and no more than 3 peptides per protein were selected. The final list consisted of 334 transitions (based on manual protein selection) and this transition list was divided among two method files for the final SRM analyses (Supplementary Table S2). The Skyline document used to make the SRM assay is freely available at Panorama: <https://panoramaweb.org/labkey/oceanbact.url>.

Selected reaction monitoring (SRM), was completed on a Thermo Vantage for all bacterial dilution samples in Table 1. Samples were prepared as described above for DIA (1  $\mu\text{g}$  of protein per 3  $\mu\text{l}$  injection), and each sample was analyzed individually on the Thermo Vantage. New C18 trap (2 cm) and C18 analytical columns (27.5 cm) were used and each sample was analyzed in two MS experiments to cover the entire peptide transition list (n = 334). Raw data can be accessed in the PeptideAtlas (<http://www.peptideatlas.org/PASS/PASS00917>) under accession PASS00917.

Acquired SRM data were analyzed in Skyline (<https://panoramaweb.org/labkey/oceanbact.url>). Peptide transition MS2 peaks were quantified using peak area integration across all samples. Peak presence was determined based on consistency of retention time (verified by spiked in QC peptides) and peak morphology. Relative retention times for QC and bacterial peptides were correlated between DIA and SRM experiments with an  $R^2 > 0.99$  (Supplementary Fig. S2).

All peptide transition peak intensities were exported from Skyline for analysis. QC transitions were assessed for consistency across runs by calculating the coefficients of variation (CVs) of transition peak area across injections in the raster package<sup>82</sup> in R v. 3.2.3<sup>74</sup>. The eight QC transitions with the lowest CV (<40) were used for inter-run normalization. Peak intensities for 305 bacterial transitions were normalized by dividing by the averaged intensities for 8 QC transitions within a given run. Normalized peak intensities were analyzed using NMDS and ANOSIM, as described above for DDA. Pearson’s r and the critical r value were calculated in R v. 3.2.3<sup>74</sup> for the correlation between peptide transition peak intensity and bacterial dilution factor. A heatmap of average peptide transition peak intensities for each dilution above the cut-off of Pearson’s critical r was constructed in pheatmap<sup>83</sup> in R, with rows (transitions) and columns (dilutions) clustered using Euclidean distance and the average clustering method. Proteins were annotated with Gene Ontology terms using the UniProt Retrieve/ID mapping tool (uniprot.org).

**Data availability.** The datasets generated during and/or analyzed during the current study are available in the repositories ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD004799 for DDA data and PXD004758 for DIA data; Peptide Atlas for SRM data (<http://www.peptideatlas.org/PASS/PASS00917>) under accession PASS00917; and Panorama for Skyline documents (<https://panoramaweb.org/labkey/oceanbact.url>). Other data generated or analysed during this study are including in this published article (and its Supplementary Information files).

## References

1. Sunagawa, S. *et al.* Structure and function of the global ocean microbiome. *Science* **22**, 1–10 (2015).
2. Vogel, C. & Marcotte, E. M. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nature Reviews Genetics* **13**, 227–232 (2012).
3. Nunn, B. L., Slattery, K., Cameron, K. A., Timmins-Schiffman, E. & Junge, K. Proteomics of *Colwellia psychrerythraea* at subzero temperatures - a life with limited movement, flexible membranes and vital DNA repair. *Environmental Microbiology* **111**, 9009–9014 (2015).
4. Georges, A. A., El-Swais, H., Craig, S. E., Li, W. K. W. & Walsh, D. A. Metaproteomic analysis of a winter to spring succession in coastal northwest Atlantic Ocean microbial plankton. *The ISME Journal* **8**, 1–13 (2014).
5. Hanson, B. T., Hewson, I. & Madsen, E. L. Metaproteomic survey of six aquatic habitats: Discovering the identities of microbial populations active in biogeochemical cycling. *Microbial Ecology* **67**, 520–539 (2014).

6. Hawley, A. K., Brewer, H. M., Norbeck, A. D., Pasa-Tolc, L. & Hallam, S. J. Metaproteomics reveals differential modes of metabolic coupling among ubiquitous oxygen minimum zone microbes. *PNAS* **111**, 11395–11400 (2014).
7. Morris, R. M. *et al.* Comparative metaproteomics reveals ocean-scale shifts in microbial nutrient utilizations and energy transduction. *The ISME Journal* **4**, 673–685 (2010).
8. Carpy, A. *et al.* Absolute Proteome and Phosphoproteome Dynamics during the Cell Cycle of *Schizosaccharomyces pombe*. *Molecular & Cellular Proteomics* **13**, 1925–1936 (2014).
9. Jessani, N., Liu, Y., Hupfrey, M. & Cravatt, B. F. Enzyme activity profiles of the secreted and membrane proteome that depict cancer cell invasiveness. *PNAS* **99**, 10335–10340 (2002).
10. Panchaud, A. *et al.* Precursor Acquisition independent from ion count: How to dive deeper into the proteomics ocean. *Analytical Chemistry* **81**, 6481–6488 (2009).
11. Bertrand, E. M., Saito, M. A., Jeon, Y. J. & Neilan, B. A. Vitamin B12 biosynthesis gene diversity in the Ross Sea: the identification of a new group of putative polar B12 biosynthesizers. *Environmental Microbiology* **13**, 1285–1298 (2011).
12. Verberkmoes, N. C. *et al.* Shotgun metaproteomics of the human distal gut microbiota. *The ISME Journal* **3**, 179–189 (2009).
13. Ram, R. J. *et al.* Community proteomics of a natural microbial biofilm. *Science* **308**, 1915–1920 (2005).
14. Moore, E. K., Nunn, B. L., Goodlett, D. R. & Harvey, H. R. Identifying and tracking proteins through the marine water column: Insights into the inputs and preservation mechanisms of protein in sediments. *Geochimica et Cosmochimica Acta* **83**, 324–359 (2012).
15. Moore, E. K., Harvey, H. R., Faux, J. F., Goodlett, D. R. & Nunn, B. L. Protein recycling in Bering Sea algal incubations. *Marine Ecology Progress Series* **515**, 45–59 (2014).
16. Pedersen, S. K. *et al.* Unseen proteome: Mining below the tip of the iceberg to find low abundance and membrane proteins. *Journal of Proteome Research* **2**, 303–311 (2003).
17. Ackermann, B. L. & Berna, M. J. Coupling immunoaffinity techniques with MS for quantitative analysis of low-abundance protein biomarkers. *Expert Review of Proteomics* **4**, 175–186 (2014).
18. Washburn, M. P., Wolters, D. & Yates, J. R. III. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnology* **19**, 242–247 (2001).
19. Geromanos, S. J. *et al.* The detection, correlation, and comparison of peptide precursor and product ions from data independent LC-MS with data dependant LC-MS//MS. *Proteomics* **9**, 1683–1695 (2009).
20. Mattes, T. E. *et al.* Sulfur oxidizers dominate carbon fixation at a biogeochemical hot spot in the dark ocean. *The ISME Journal* **7**, 2349–2360 (2013).
21. Nunn, B. L. *et al.* Diatom Proteomics Reveals Unique Acclimation Strategies to Mitigate Fe Limitation. *Plos One* **8**, e75653 (2013).
22. Picotti, P. & Aebersold, R. Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nature Methods* **9**, 555–566 (2012).
23. Glukhova, V. A., Tomazela, D. M., Findlay, G. D., Monnat, R. J. Jr. & MacCoss, M. J. Rapid assessment of RNAi-mediated protein depletion by selected reaction monitoring mass spectrometry. *Journal of Proteome Research* **12**, 3246–3254 (2013).
24. Saito, M. A. *et al.* Multiple nutrient stresses at intersecting Pacific Ocean biomes detected by protein biomarkers. *Science* **345**, 1173–1177 (2014).
25. Wang, D.-Z., Dong, H.-P., Xie, Z.-X., Dai, M.-H. & Hong, H.-S. Metaproteomic characterization of dissolved organic matter in the water column of the South China Sea. *Limnology and Oceanography* **56**, 1641–1652 (2011).
26. Dong, H.-P., Wang, D.-Z., Xie, Z.-X., Dai, M.-H. & Hong, H.-S. Metaproteomic characterization of high molecular weight dissolved organic matter in the surface seawaters in the South China Sea. *Geochimica et Cosmochimica Acta* **109**, 51–61 (2013).
27. Yoshida, M., Yamamoto, K. & Suzuki, S. Metaproteomic characterization of dissolved organic matter in coastal seawater. *Journal of Oceanography* **70**, 105–113 (2014).
28. Saito, M. A. *et al.* Iron conservation by reduction of metalloenzyme inventories in the marine diazotroph *Crocaphaera watsonii*. *PNAS* **108**, 2184–2189 (2011).
29. Saito, M. A. *et al.* Needles in the blue sea: Sub-species specificity in targeted protein biomarker analyses within the vast oceanic microbial metaproteome. *Proteomics* **15**, 3521–3531 (2015).
30. Mackey, K. R. M. *et al.* Divergent responses of Atlantic coastal and oceanic *Synechococcus* to iron limitation. *PNAS* **112**, 9944–9949 (2015).
31. Fusaro, V. A., Mani, D. R., Mesirov, J. P. & Carr, S. A. Prediction of high-responding peptides for targeted protein assays by mass spectrometry. *Nature biotechnology* **27**(2), 190 (2009).
32. Mesure, B. *et al.* The UniPept metaproteomics analysis pipeline. *Proteomics* **15**, 1437–1442 (2015).
33. Timmins-Schiffman, E. *et al.* Critical decisions in metaproteomics: Achieving high confidence protein annotations in a sea of unknowns. *ISME Journal* **11**, 309–314 (2016).
34. Bunse, C. *et al.* Spatio-Temporal Interdependence of Bacteria and Phytoplankton during a Baltic Sea Spring Bloom. *Front Microbiol* **7**, 517–526 (2016).
35. Suzuki, K. *et al.* Population dynamics of phytoplankton, heterotrophic bacteria, and viruses during the spring bloom in the western subarctic Pacific. *Deep-Sea Research Part I-Oceanographic Research Papers* **58**, 575–589 (2011).
36. Teeling, H. *et al.* Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. *Science* **336**, 608–611 (2012).
37. Searle, B. C., Egertson, J. D., Bollinger, J. G., Stergachis, A. B. & MacCoss, M. J. Using data independent acquisition (DIA) to model high-responding peptides for targeted proteomics experiments. *Molecular & Cellular Proteomics* **14**, 2331–2340 (2015).
38. Vizcaíno, J. A. *et al.* The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013. *Nucleic acids research* **41**, D1063–D1069 (2012).
39. Vizcaíno, J. A. *et al.* Update of the PRIDE database and related tools. *Nucleic Acids Research* **44**, D447–D456 (2016).
40. Ake, H. *et al.* BCO-DMO – a domain-specific repository for oceanographic data from around the world, OD34B-2759. Presented at the 2018 Ocean Sciences Meeting, Portland, OR, 12–16 February (2018).
41. Domon, B. & Aebersold, R. Mass spectrometry and protein analysis. *Science* **312**, 212–217 (2006).
42. Tang, H.-Y. *et al.* A novel four-dimensional strategy combining protein and peptide separation methods enables detection of low-abundance proteins in human plasma and serum proteomes. *Proteomics* **5**, 3329–3342 (2005).
43. Moore, E. K., Nunn, B. L., Faux, J. F., Goodlett, D. R. & Harvey, H. R. Evaluation of electrophoretic protein extraction and database driven protein identification from marine sediments. *Limnology and Oceanography Methods* **10**, 353–366 (2012).
44. Nunn, B. L. *et al.* The path to preservation: Using proteomics to decipher the fate of diatom proteins during microbial degradation. *Limnology and Oceanography* **55**, 1790–1804 (2010).
45. Sherman, J., McKay, M. J., Ashman, K. & Molloy, M. P. How specific is my SRM?: The issue of precursor and product ion redundancy. *Proteomics* **9**, 1120–1123 (2009).
46. Chang, C.-Y. *et al.* Protein significance analysis in selected reaction monitoring (SRM) measurements. *Molecular and Cellular Proteomics* **11**, M111.014662 (2011).
47. Plumel, M. J. *et al.* Mass spectrometry-based sequencing and SRM-based quantitation of two novel vitellogenin isoforms in the leatherback sea turtle (*Dermochelys coriacea*). *Journal of Proteome Research* **12**, 4122–4135 (2013).
48. Abelin, J. G. *et al.* Mass spectrometry profiling of HLA-associated peptidomes in mono-allelic cells enables more accurate epitope prediction. *Immunity* **46**(2), 315–326 (2017).

49. Bianco, L. & Bessant, C. Free computational resources for designing selected reaction monitoring transitions. *Proteomics* **10**, 1106–1126 (2010).
50. Pino, L. K. *et al.* The Skyline ecosystem: Informatics for quantitative mass spectrometry proteomics. *Mass spectrometry reviews* (2017).
51. Li, H. *et al.* Current trends in quantitative proteomics—an update. *Journal of Mass Spectrometry* **52**, 319–341 (2017).
52. Panchaud, A., Jung, S., Shaffer, S. A., Aitchison, J. D. & Goodlett, D. R. Faster, quantitative, and accurate precursor acquisition independent from ion count. *Analytical Chemistry* **83**, 2250–2257 (2011).
53. Monneuse, J.-M. *et al.* Towards the profiling of the *Arabidopsis thaliana* plasma membrane transportome by targeted proteomics. *Proteomics* **11**, 1789–1797 (2011).
54. Cox, D. M. *et al.* Multiple reaction monitoring as a method for identifying protein posttranslational modifications. *Journal of biomolecular techniques* **16**, 83–90 (2005).
55. Anderson, L. & Hunter, C. L. Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. *Molecular & Cellular Proteomics* **5**, 573–588 (2006).
56. Sherwood, C. A. *et al.* MaRiMba: a software application for spectral library-based MRM transition list assembly. *Journal of proteome research* **8**, 4396–4405 (2009).
57. Chan, L.-K. *et al.* Transcriptional changes underlying elemental stoichiometry shifts in a marine heterotrophic bacterium. *Frontiers in Microbiology* **3**, 1–24 (2012).
58. Guillard, R. R. L. & Ryther, J. H. Studies of marine planktonic diatoms. I. *Cyclotella nana* Hustedt and *Detonula confervacea* Cleve. *Canadian Journal of Microbiology* **8**, 229–239 (1962).
59. Guillard, R. R. L. Culture of phytoplankton for feeding marine invertebrates. pp 26–60 In Smith, W. L. & Chanley, M. H. (Eds) *Culture of Marine Invertebrate Animals*. Plenum Press, New York, USA (1975).
60. Gonzalez, J. M., Whitman, W. B., Hodson, R. E. & Moran, M. A. Identifying numerically abundant culturable bacteria from complex communities: an example from a lignin enrichment culture. *Applied and Environmental Microbiology* **62**, 4433–4440 (1996).
61. Smith, D. C., Steward, G. F., Long, R. A. & Azam, F. Bacterial Mediation of Carbon Fluxes during a Diatom Bloom in a Mesocosm. *Deep-Sea Research Part II-Topical Studies in Oceanography* **42**, 75–97 (1995).
62. Reimann, L., Steward, G. F. & Azam, F. Dynamics of bacterial community composition and activity during a mesocosm diatom bloom. *Applied and Environmental Microbiology* **66**, 578–587 (2000).
63. Li, W. K. W., Dickie, P. M., Harrison, W. G. & Irwin, B. D. Biomass and production of bacteria and phytoplankton during the spring bloom in the western North Atlantic Ocean. *Deep Sea Research Part II: Topical Studies in Oceanography* **40**, 307–327 (1993).
64. Ducklow, H. W., Kirchman, D. L., Quniby, H. L., Carlson, C. A. & Dam, H. G. Stocks and dynamics of bacterioplankton carbon during the spring bloom in the eastern North Atlantic Ocean. *Deep-Sea Research II* **40**, 245–263 (1993).
65. Eng, J. K., Jahan, T. A. & Hoopmann, M. R. Comet: an open source tandem mass spectrometry sequence database search tool. *Proteomics* **13**, 22–24 (2012).
66. Eng, J. K. *et al.* A deeper look into Comet - Implementation and features. *Journal of the American Society for Mass Spectrometry* **26**, 1865–1874 (2015).
67. Nesvizhskii, A. I., Keller, A., Kolker, E. & Aebersold, R. A statistical model for identifying proteins by tandem mass spectrometry. *Analytical Chemistry* **75**, 4646–4658 (2003).
68. Keller, A. *et al.* Experimental protein mixture for validating tandem mass spectral analysis. *OMICS: A Journal of Integrative Biology* **6**, 207–212 (2002).
69. Pedrioli, P. G. A. Trans-Proteomic Pipeline: A pipeline for proteomic analysis. *Proteome Bioinformatics* **604**, 213–238 (2010).
70. Deutsch, E. W. *et al.* Trans-Proteomic Pipeline, a standardized data processing pipeline for large-scale reproducible proteomics informatics. *Proteomics Clinical Applications* **9**, 745–754 (2015).
71. Florens, L. *et al.* Analyzing chromatin remodeling complexes using shotgun proteomics and normalized spectral abundance factors. *Methods* **4**, 303–311 (2006).
72. Timmins-Schiffman, E. *et al.* Shotgun proteomics reveals physiological response to ocean acidification in *Crassostrea gigas*. *BMC Genomics* **15**, 951–969 (2014).
73. Oksanen, J. *et al.* Vegan: Community Ecology Package. R package version 2.3–5, <https://CRAN.R-project.org/package=vegan> (2016).
74. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, <https://www.R-project.org> (2015).
75. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. *Nature Protocols* **4**, 44–57 (2009).
76. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research* **37**, 1–13 (2009).
77. Mesuere, B. *et al.* UniPept: Tryptic Peptide-Based Biodiversity Analysis of Metaproteome Samples. *Journal of Proteome Research* **11**, 5773–5780 (2012).
78. Ting, Y. S. *et al.* PECAN: library-free peptide detection for data-independent acquisition tandem mass spectrometry data. *Nature Methods* **14**, 903–908 (2017).
79. Chambers, M. C. *et al.* A cross-platform toolkit for mass spectrometry and proteomics. *Nature Biotechnology* **30**, 918–920 (2012).
80. MacLean, B. *et al.* Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26**, 966–968 (2010).
81. Wolf-Yadlin, A., Hautaniemi, S., Lauffenburger, D. A. & White, F. M. Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks. *PNAS* **104**, 5860–5865 (2007).
82. Hijmans, R. J. Raster: Geographic data analysis and modeling. R package version 2.5–2, <https://CRAN.R-project.org/package=raster> (2015).
83. Kolde, R. Pheatmap: Pretty Heatmaps. R package version 1.0.8, <https://CRAN.R-project.org/package=pheatmap> (2015).
84. Montagnes, D. J. S., Berge, J. A., Harrison, P. J. & Taylor, F. J. R. Estimating carbon, nitrogen, protein, and chlorophyll a from volume in marine phytoplankton. *Limnology and Oceanography* **39**, 1044–1060 (1994).

## Acknowledgements

This work was funded by grants from the National Science Foundation (NSF-OCE 1233014 to B.L.N. and for E.T.S. and M.P.M. and NSF-OCE 1233589 to H.R.H.) as well as a Training Grant from the National Institutes of Health for E.T.S. (T32 HG00035). This work was also supported in part by the University of Washington's Proteomics Resource (UWPR95794). We would like to thank Jarrett Egertson and James Bollinger for their indispensable help in designing SRM assays for this work; Lindsay Pino and Jarrett Egertson for advice on the analysis of the SRM data; Genn Merrihew and Rich Johnson for assistance with DIA and SRM method development; Jimmy Eng and Priska von Haller for help with data acquisition and Comet searches; Nick Shulman for periodic mentoring; Mike MacCoss for guidance and use of lab space; and T.A.N. and I.J.E. for their ongoing inspiration. Mary Ann Moran (UGA) kindly provided the original pure culture of *R. pomeroyi*.

## Author Contributions

H.R.H. and B.L.N. conceived of the experimental plan and contributed to data analysis, interpretation, and manuscript writing. E.T.S. designed the mass spectrometry experiments, was responsible for data analysis and interpretation, and was the primary manuscript author. M.P.M. grew the bacteria and phytoplankton cultures, extracted proteins, and contributed to writing the manuscript. Y.S.T. aided with proteomics data analysis and interpretation.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-27650-4>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018