Linking amino acid sequences to sources and fates of organic matter in aquatic systems

Megan Duffy

April 21, 2020

A description of proposed Ph.D. dissertation research

Supervisory committee

Rick Keil (chair)
Jeff Richey
Gabrielle Rocap
Anitra Ingalls
Sarah Stroup (GSR)

Contents

Overview

Proteins enact life's intent. These macromolecules, directed by genes and informed by the environment, are the engines that power the cells of all biological entities on Earth. From viruses and bacteria to humans and blue whales, proteins serve a vast range of metabolic, transport, communication, and structural purposes. Along with geological and chemical drivers, protein-driven biological processes modulate planetary cycles of carbon, oxygen, and nitrogen.

By unlocking the information stored in peptide and protein sequences, we can learn the biological origins and functions of cells within a community. For organic geochemists, there is useful information here as well: proteins make up a large proportion of organic carbon and nitrogen in aquatic systems. Thus, the cycling and degradation dynamic of proteins is of great importance when thinking about global organic carbon preservation and sequestration.

My proposed Ph.D. research addresses the cycling of proteins and protein-derived organic matter in both laboratory settings and environmental systems. Chapter 1, my M.S. project, describes the usefulness of integrating a different kind of peptide sequencing, *de novo* sequencing, into the traditional environmental proteomics workflow in order to access degraded and unanticipated sequences. This is described in "Protein cycling in the eastern tropical North Pacific oxygen deficient zone: a *de novo*-assisted peptidomic approach", [Duffy et al., *in review*]. The three subsequent projects utilize this technique to ask questions about proteins and peptide cycling in complex environmental systems.

Chapter 2 uses the *de novo*-assisted peptidomic technique to follow the peptides of a diatom through a simulated bloom and subsequent degradation by a natural microbial community. Chapter 3 is a peptidomic-based comparison of organic matter across seasons, stations (on and offshore) and sinking class of marine organic matter in an ODZ. Chapter 4 moves out of the purely marine realm and bridges the span between terrestrial and marine systems in the Amazon River-Atlantic continuum, probing how organic matter travels from land to sea, and how microbes alter its reactivity and character.

1 Proposed work

1.1 Chapter 1: De novo-assisted peptidomics helps in the study of marine carbon flux and protein degradation

Note: The work comprising Chapter 1 is detailed in a submitted manuscript sent to all committee members, and as such will not be fully introduced here.

Developments in high resolution mass spectrometry and computing have facilitated metaproteomic investigations of proteins in suspended Dong et al., 2010; Bridoux et al., 2015; Bergauer et al., 2017 and sinking Moore et al., 2012 organic matter. These powerful metaproteomic tools could help resolve questions about protein and peptide degradation in the ocean, but they are currently limited in that they are designed to identify intact proteins from living cells.

Current metaproteomic workflows adhere to a bottom-up proteomic approach that relies on peptide-spectrum matching through database searching Saito et al., 2019, with databases derived from complementary metagenomic or metatranscriptomic analyses and/or publicly available sequence data. This approach has limitations for protein in environmental organic matter, a complex mixture of living, dead, and degraded material. Deviations from a protein's DNA blueprint, post-translational modifications (PTM), occur as a result of that protein's intended biochemical function and also through degradative processes Kim, Weiss, and Levine, 2014 that are poorly understood. Indeed, it is typical that fewer than half of all tandem mass spectra acquired in shotgun proteomics experiments are successfully matched to a peptide from a database Chick et al., 2015.

De novo sequencing is a database-independent tool that uses the same tandem mass spectral information as a peptide-spectrum matching analysis. Rather than comparing peptides to a protein database, first principles-based algorithms are used to establish sequences directly. Several styles of *de novo* sequencing algorithms exist, but all use respective mass differences between fragment ion peaks to piece back together the original peptide ion amino acid sequence. *De novo* sequencing is generally less accurate than database searching, often due to insufficient mass accuracy Muth et al., 2015 or uneven fragmentation patterns in tandem mass spectrometry Lu and Chen, 2004. While the original method of protein sequencing, *de novo* is now mainly used

in antibody peptide sequencing (i.e., only when a database-derived sequence is unavailable). But analogous to antibody peptides, degraded peptides are inherent unknowns in the environment. They are difficult to predict or make a database entry for, but surely are present and highly relevant to organic matter cycling.

1.1.1 Benchmark: *Prochlorococcus marinus* culture

As a benchmark study, I compared the sequencing performance of a well-established *de novo* algorithm against traditional database searching on a high resolution mass spectrometry (HRMS) dataset from a *Prochlorococcus marinus* culture. *P. marinus*, a ubiquitous free-living marine cyanobacterium found throughout the world's oceans Chisholm et al., 1988, was an advantageous marine microbe for this assessment given its small and relatively well-characterized proteomes Paul et al., 2010.

I performed *de novo* peptide sequencing with a commercially available algorithm (Peaks Ma et al., 2003) and database searching, aligning and comparing the two resulting groups of peptides by several criteria: coverage of the known *Prochlororoccus* proteome, degree and types of mass modification, length (number of amino acids), and taxonomic specificity (the degree to which a certain peptide is identifying of and unique to one organism's protein sequence or many across different taxa.)

The *de novo* algorithm did find peptides unidentified through database searching that boosted the overall protein identifications by 6%. *De novo* peptides were, on average, almost 3 residues shorter than the average of the database-identified peptides [Figure 1]. Peptide length is an important metric in an metaproteomic context because with increased amino acid space, peptide sequences have more potential to be specific to a particular parent protein.

To establish the specificity of *de novo*-only sequences (*de novo* sequences that did not align with database-ID'd peptides), I interrogated their sequences against the UniProt Knowledgebase (UniProtKB) protein database using Unipept, a tool designed for tryptic peptide sorting and identification. Unipept uses a lowest common ancestor (LCA) approach to match peptides to as low a resolved taxonomic identification level as is possible given the input sequence. I found that many *de novo*-only peptides are taxonomically specific, with 17% specific to at least the phylum *Cyanobacteria*, and 15% specific to the species level [Figure 1]. In contrast, of the database-
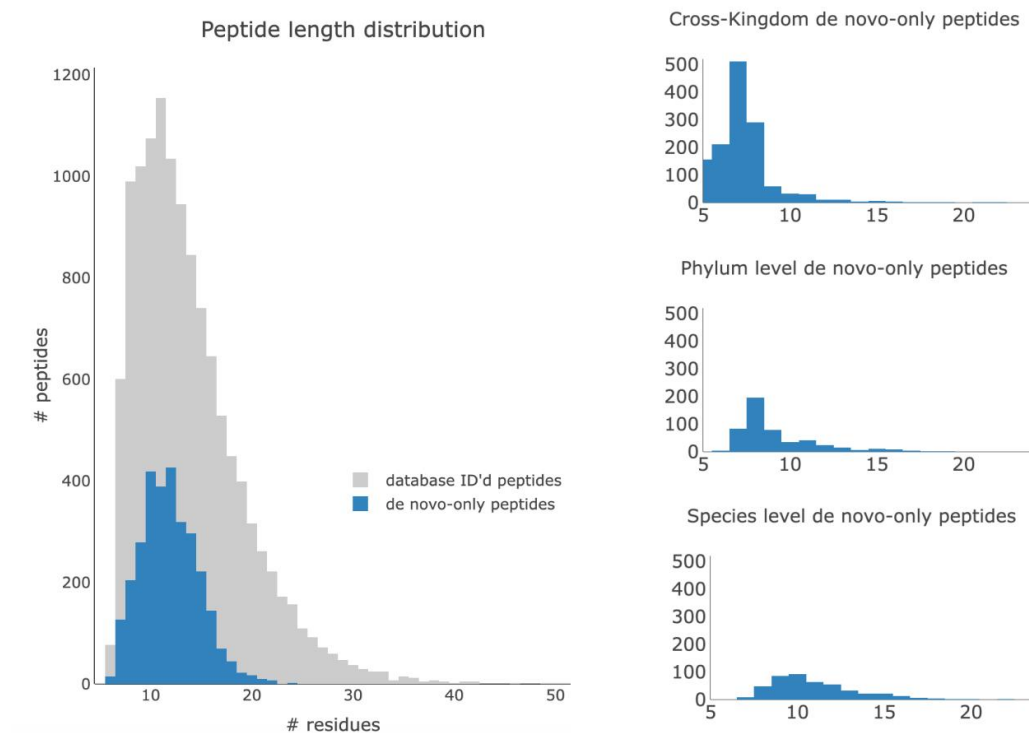
Figure 1: Histograms showing a) overlaid distributions of sequenced peptide lengths (number of amino acid residues) for database search-identified (grey) and *de novo*-only peptides (blue) of cultured *Prochlorococcus marinus* MED4 protein LC-MS/MS dataset. Of 12,160 database peptides, the mean length was 13.8 residues; of 3,019 *de novo* peptides, the mean length was 10.9 amino acids. Distributions of *de novo*-only peptide lengths for individual taxonomic rankings are shown for b) cross-kingdom, c) phylum, and d) species.

identified peptides; 51% of peptides matched to the species level. Overall, this comparison showed that the *de novo*-only peptides sequenced from the *Prochlorococcus* culture data are not as specific to their organismal source to those sequenced by traditional database searching, but still have potential to identify individual organisms in a mixed community. It is important to note that the *de novo*-only non-specific peptides, while not taxon-identifying, nevertheless represent valuable information, especially if one is less concerned about organismal source and wishes to evaluate sequence similarity with respect to preservation and degradation.

I searched for 30 different mass modifications, and the seven most commonly observed in this dataset were carbamidomethylation (on cysteine), oxidation (on methionine), acetylation, formylation, deamidation (on asparagine and glutamine), dehydration and phosphorylation. Since carbamidomethylation is performed during experimental peptide preparation and set during both database and *de novo* searches, I expect no difference in levels of this modification been the peptide pools. Indeed, no significant difference in carbamidomethylation is observed between database and *de novo*-only peptides. However, in the case of every other modification I searched for, the *de novo*-only peptides are more modified - by up to 10% more in the case of acetylation and to lesser extents for oxidation, formylation, deamidation, dehydration, and phosphorylation.

This benchmark culture study shows that while the results of *de novo* searches are not as comprehensive in coverage as the database approach, *de novo* sequencing does complement the database search by a) identifying peptides that would not otherwise have been discovered, b) identifying peptides that have been post-translationally modified, and c) providing output of non-specific peptides for further evaluation.

1.1.2 Preliminary environmental application: the eastern tropical North Pacific

The exciting application of *de novo* peptide sequencing is to undercharacterized systems, exemplified by the challenges of marine environments. In this preliminary study as part of my M.S. work, I evaluated six POM samples from the eastern tropical North Pacific (ETNP) oxygen deficient zone (ODZ) collected in January, 2017 onboard the *R/V Sikuliaq*. Suspended particles were collected with large volume *in situ* pumps on 0.3 μm membranes and sinking particles were collected with free-drifting, unpoisoned sediment traps.

I found that *de novo*-identified peptides included matches to proteins from unanticipated taxa, including many from the fungal subkingdom Dikarya. I also sequenced peptides from the autotrophic phylum *Cyanobacteria* in particles at 1000 m depth, indicating transfer of autotrophic C and N from these surface microbes. Some of these peptides were missed by the database-driven approach, likely because they are in the process of being degraded as they sink to the interior of the ocean – indeed, *de novo*-identified peptides at depth contain more mass modifications relative to those at epipelagic base, including deamidation and oxidation. Deamidation has been hypothesized as a source of ammonium supporting anammox in this region Van Mooy, Richard G Keil, and Allan H Devol, 2002, suggesting that the *de novo* tool provides a molecular-level view into the processes fueling chemoautotrophy.

Expected outcomes

A manuscript introducing *de novo*-assisted approach and containing both the benchmark *Prochlorococcus* peptide analysis and ETNP POM study was submitted to *Limnology and Oceanography* and is currently under review. The *de novo*-assisted peptide sequencing approach is used in Chapters 2, 3, and 4.

1.2 Chapter 2: Tracking protein breakdown in a controlled algae degradation

De novo-assisted peptide sequencing was used to investigate changes in peptide quality over the course of a semi-controlled degradation of a diatom culture in by a natural microbial assemblage. Proteins were extracted and trypsin-digested from a bloom-and-bust simulation sampled at multiple time-points. This builds upon both laboratory Nunn et al., 2010 and environmental investigations Moore et al., 2012 of protein breakdown. The former study, in which Nunn et al. tracked diatom proteins through a similar degradation experiment, highlights both how far metaproteomic instrumental and algorithmic tools have advanced in a short time, and the strong suit of *de novo* sequencing: they identified 340 *Thalassiosira pseudonana* peptides on day 0 and 63 on day 10. I am now able to sequence and identify hundreds of algal peptides in each timepoint, notably in the latter stages of the degradation (day 0 = 576; day 12 = 249).

The Nunn et al. experiment resulted in 4 identifiable diatom peptides after a 23-day degradation, three with or adjacent to transmembrane domains and the forth contained in an organelle. The theory that membrane (or organelle) associated proteins may be preferentially preserved Wolfe et al., 2006 is becoming increasing testable with new tools that can search thousands of peptides for their associated proteins' annotations like GO terms. I will perform such evaluations, and also search for mass modifications: recent work by Abdulla et al. shows an accumulation in anoxic sediment pore water of what seem to be deaminated peptides (shown through Fourier-transform ion cyclotron resonance mass spectrometry derived formulae in DOM, not proteomic sequencing, Abdulla, Burdige, and Komada, 2018). I will also calculate relative abundances of individual amino acids at each timepoint over the course the degradation - this will mimic the expensive body of early diagenesis research that looks at total hydrolyzable amino acids (THAA), and which reveals empirical trends in amino acid composition (for instance, enrichment in glycine, serine, and theonine in degraded OM) and has been used to generate a Degradation Index (DI) Dauwe and Middelburg, 1998. This body of protein degradation work drives my primary question:

Question 2.1 How does the phytoplankton peptide pool change over the course of degradation?

Informed by the literature and results of preliminary ETNP POM peptidomics, I hypothesize that peptides will become, as the degradation experiment progresses:

a) shorter in length

b) more non-tryptic in character (lower tryptic:non-tryptic ratio)

c) more modified, particularly with more oxidation and deamidation

d) closer to DI index relative amino acid ratios

e) increasingly from membrane-associated proteins

I will address this question using LC-HRMS with *de novo*-assisted peptide sequencing and peptide-based tools (UniPept, MetaGOmics) to explore annotations on function and subcellular localization.

1.2.1 Experimental design

The following was performed by collaborators at the University of Maine: a culture of marine diatom *Thalassiosira weissflogii* brought to approximately 106 cells/mL, concentrated and rendered non-viable by freezing at -80 C, then homogenized. This frozen concentrate was thawed, resuspended to 2 g/L (dry weight) in 1 μm filtered, UV-sterilized surface seawater collected from the Gulf of Maine (GoM). Algal cells in suspension were confirmed intact by microscopy. Unsterilized, 1 μm filtered filtered GoM seawater was used to induce bacterial decomposition of algal material, with 1 mL added to 1 5L of seawater. The algal suspension was covered in black plastic and left undisturbed in the dark at 19.5 C, monitored daily and sampled for chemical analyses at days 0, 2, 5, and 12. Algal cell counts were estimated by chlorophyll autofluorescence and the abundance of bacterial cells indicated by SYBRGreen staining.

1.2.2 Peptide evolution over degradation of *T. weissflogii culture*

To make a protein search database, I extracted *T. weissflogii* transcript sequences from the Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP) Keeling et al., 2014 and GoM metagenomic sequences from the Global Ocean Sampling (GOS) data product Yooseph et al., 2007.

De novo-assisted database searching reveals changes in peptide composition over degradation by a natural microbial assemblage. Both tryptic and non-tryptic peptides sequenced from late-stage degradation material (day 5, day 12) are relatively enriched in glycine, serine, and threonine and are depleted in phenylalanine, glutamic acid, tyrosine, and leucine. While similar observations of relative amino acid abundance patterns have been made for organic matter degradation on a bulk level by looking at THAA Dauwe, Middelburg, et al., 1999, to my knowledge **this is a first observation that the enzyme-hydrolyzable peptide pool influences this pattern.**

I also tracked tryptic and non-tryptic peptides that were sequenced from each timepoint through the degradation, monitoring which were membrane-associated and modified. I will use spectral counts to better quantify and expand on these results.

1.2.3 Evolution of peptide mass modifications

Peptide mass modifications changed throughout the degradation, with peptides generally containing more of a selection of the 30 modifications I allowed in both the *de novo* algorithm and database searches. The relative number of peptides with deamidated residues increased, as well as for oxidation, formylation, and dehydration. I will use spectral count normalization to further quantify the contributions of modified peptides at each timepoint.

Progress

All lab work is complete for this project. Data analysis is in a final stage and I expect to have a manuscript ready during Spring Quarter 2020 to go out for comments and edits from collaborators in Larry Mayer's group at the University of Maine.

Expected outcomes

The manuscript, to be submitted to *Limnology and Oceanography*, builds upon a large body of work that asks "What happens to the proteins of marine primary producers when they die?" Because *de novo* sequencing is not reliant on a protein database, it is better able to identify peptides with mass modifications, which is particularly useful in later stages of the time series degradation sequence. The chemical and phylogenetic specificity contained in peptide sequences enables the tracking of protein degradation along a presumed early diagenetic sequence, opening the door for more insights into the origin, evolution and fate of proteinaceous materials under varying ocean conditions.

1.3 Chapter 3: High-resolution marine flux measurements, *in situ* respiration rate determinations, and metaproteomic surveys in the eastern tropical North Pacific oxygen deficient zone

1.3.1 Study site: the eastern tropical North Pacific

Oxygen deficient zones (ODZs) naturally occur where aerobic respiration of organic matter (OM) combines with water column stabilization to form a

persistent, low-oxygen layer at mid-depths. ODZs make up less than 1% by volume of the world ocean, yet account for 30-50% of the oceanic nitrogen loss as N_2 DeVries et al., 2013, driving nitrogen limitation of primary productivity over vast regions of the ocean. The size of ODZs is sensitive to climate change and variability: a 1% reduction of the ocean's O_2 content is predicted to double the size of world OMZs Deutsch et al., 2011. Critically, climate models predict an approximate 5% decrease to the ocean's O_2 reservoir within this century Bopp et al., 2013.

The eastern tropical North Pacific (ETNP) is home to the Earth's largest marine oxygen deficient zone, accounting for approximately 41% of global marine anoxic waters Paulmier and Ruiz-Pino, 2009. Numerous studies have documented an 'enhanced' flux through ODZs, implying that a high fraction of the surface-derived particulate organic carbon sinks to the deep ocean Allan H. Devol and Hartnett, 2001; Van Mooy, Richard G Keil, and Allan H Devol, 2002; Richard G. Keil, Neibauer, et al., 2016. However, there remains much uncertainty about the mechanism(s) explaining these observations. In the ODZ, shifts in zooplankton behavior Richard G. Keil, Neibauer, et al., 2016 and *in situ* production from anammox Ganesh et al., 2018 or deep photoautotrophy by *Prochlorococcus* populations within the anoxic secondary chlorophyll maximum Fuchsman et al., 2019 likely have roles in an enhanced flux. These POM fluxes are important in controlling N_2 loss from denitrification and anammox Fuchsman et al., 2019 and likely the makeup of POM controls the relative contribution of those processes Babbin et al., 2014. Since diffusion of electron acceptors like O_2 and NO_3 can be limiting in the interior of large particles Ploug, Iversen, and Fischer, 2008; Ploug, 2001; D. Bianchi et al., 2018, they might become niches where denitrification and other processes (sulfate reduction, iron reduction) may occur, even if those processes are unfavored in the surrounding water. Because proteinaceous matter comprises the single largest identifiable component of the sinking flux in the tropical eastern Pacific Wakeham et al., 1997, evaluating its processing using the *de novo*-assisted approach may provide insights into POM dynamics and carbon and nitrogen flow in within this ODZ.

My goal is to combine peptide-level molecular determinations with flux and metabolic rate measurements to learn about the microbial actors, their fuel, and (quantitatively) the biogeochemical implications of their lifestyles. This is summarized by four primary questions:
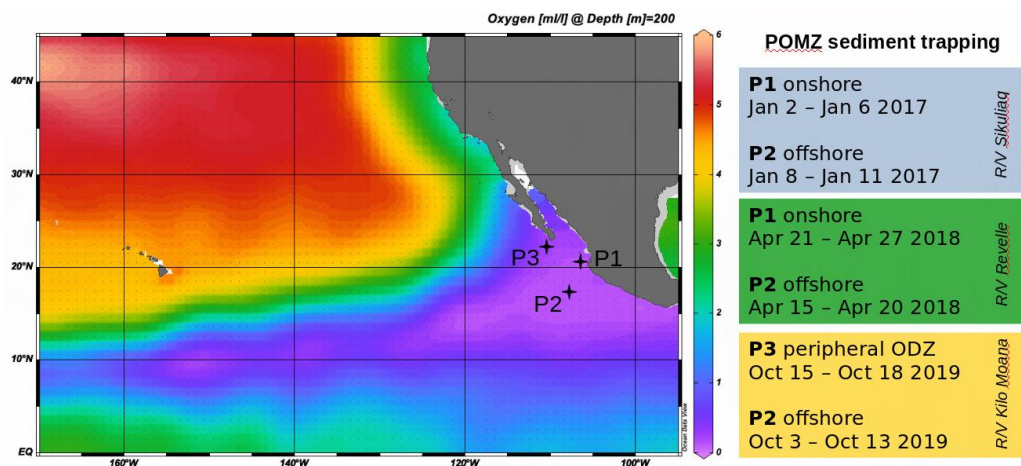
Figure 2: The eastern tropical North Pacific oxygen deficient zone (ETNP ODZ), shown here as O_2 concentration at 200 m from World Ocean Atlas data, 1955-2013. Time series sediment trapping stations (P1, P2, and P3) for the three POMZ expeditions are shown with expedition timings.

Question 3.1 How variable are OM flux profiles across seasons between nearshore and offhsore oligotrophic stations?

Question 3.2 Do rates of nitrogen loss (heterotrophic denitrification and anammox) increase with higher OM fluxes?

Question 3.3 Is there a shift in the microbial OM processing when these profiles look different?

Question 3.4 Are there patterns in protein degradation across sinking and suspended POM through the ODZ, and between near and offshore stations?

Three research cruises to the ETNP have been mounted as part of an NSF Dimensions in Biodiversity grant awarded to PIs Gabrielle Rocap, Allan Devol, Rick Keil, and Curtis Deutsch. A large component of that field work has been week-long time series occupations of a nearshore station (P1) and offshore oligotrophic station (P2) where the Keil group has deployed free-drifting sediment trap-incubator systems designed to both collect sinking particles and also to perform stable isotope labeled incubations with particle-concentrated (and control) chambers at *in situ* conditions [Keil et al., *in*
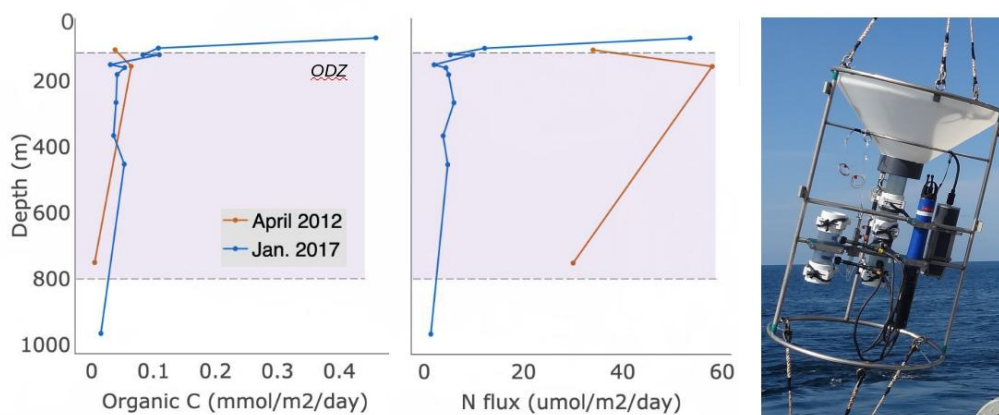
Figure 3: Flux profiles of organic carbon and nitrogen at offshore oligotrophic station P2 in the ETNP ODZ. The ODZ, where O_2 dropped below the limit of detection with a Seabird SBE 43 Dissolved Oxygen Sensor, is shown in purple shading. *Right*: a generation 5 sediment trap-incubator system.

prep]. A third station (P3) was visited in October 2019 to include a deep ODZ with a minor or absent secondary chlorophyll maximum. At all stations we also collected OM using large volume *in situ* McLane pumps onto glass fiber filters (stacked 2.7 and 0.3 μm). Figure 2 shows the locations and timings of ETNP station sediment trap sampling.

1.3.2 Flux of suspended and sinking POM from 2012-2019

Flux attenuation curves are anticipated to vary between cruise years and nearshore and offshore oligotrophic stations. All three cruises' flux samples are at UW in storage, and the 2017 suite of suspended and sinking particle samples has been processed for elemental analyses of particulate inorganic and organic carbon and nitrogen quantities and ^{13}C and ^{15}N isotopic compositions, carried out at the UW IsoLab and UC Davis Stable Isotope Facility on both raw and acidified samples. Comparing with an earlier cruise to stations P2, the organic C and N fluxes are markedly different between April 2012 and January 2017 (Figure 3). To answer **Question 3.1**, I will:

1. Lead preparation of 2019 samples stable C and N isotope measurements at the UW IsoLab or UC Davis facility (2018 samples prepared but not

run).

2. Determine attenuation coefficients or P1, P2, and P3 during all three cruises as in Keil et al., 2016 Richard G. Keil, Neibauer, et al., 2016.

1.3.3 N$_2$ loss rate calculations from sediment trap-incubators

My role in this field work and subsequent sample processing and data analysis is primarily the OM flux and metaproteomic work. However, key to contextualizing the OM flux and microbial processing of OM informed by metaproteomics is determining the N$_2$ loss rates due to particles. Clara Fuchsman and Allan Devol are the leads on this aspect of the project, and my contributions have been determining organic carbon and nitrogen in each incubation and amassing the metadata needed to make these calculations.

Tracers (in this case ^{15}NO$_2$) were injected into the particle-concentrated and control bottles onboard the sediment traps and incubated for 4-32 hours. Once the traps are back onboard, we subsampled both chambers for ^{30}N$_2$ and ^{29}N$_2$ measurement by isotope ratio mass spectrometry (IRMS). Currently we have calculated N$_2$ loss rates for the 2017 cruise (1-5 nM N day^{-1}), which are within range of previous shipboard incubation rates at Station P2 (5-15 nM N day^{-1}) Babbin et al., 2014. We have prepared and are waiting only on organic carbon numbers to normalize rates from the 2018 cruises; we have not yet run samples from the 2019 cruise. To finish this project I will:

1. Run 2019 incubation N$_2$ samples when normal operations resume at UW in the Stable Isotope Lab on the Delta V IRMS.

2. Use the ^{29}N$_2$ and ^{30}N$_2$-based calculation developed by Al Devol and Bonnie Chang to determine N$_2$ loss rate over experiment.

3. Normalize rates to organic carbon determined by measurements at UW IsoLab or UC Davis facilities.

1.3.4 Metapeptidomic and metaproteomic signatures of POM through the ODZ

My M.S. work focused on a subset of ETNP POM from 2017 as a first look using the *de novo*-assisted peptide sequencing approach. For this project I'll expand on that work and technique to address **Question 3.3** and **Question**

15

3.4. **I hypothesize that in the ODZ and perhaps in the oxycline, suspended and sinking particles lead to functional diversity by creating biochemical niches for microbes that facilitate N loss and other processes**. Metaproteomics samples (36 total) from 2017 and 2018 cruises have already been run on a Thermo Q-Exactive Plus at the UW Proteomics Resource Center. Samples come from above, within, and below the ODZ from both P1 and P2 stations. For a functional analysis of OM processing (**Question 3.3**), I will:

1. Perform *de novo*-assisted metaproteomics with Peaks using an assembled metagenome-derived database constructed by Clara Fuchsman, Gabrielle Rocap, and Cedar McKay.

2. Analyze database search results using a peptide-based Gene Ontology (GO) enrichment analysis to explore community function of particles. I'll use MetaGOmics Riffle et al., 2017, which determines the abundance of GO functional categories by combining all GO terms associated with a peptide based on that peptide's matching metagenome database protein(s). The relative abundance for each GO term is then calculated by the combined spectral counts of associated peptides. Finally, to determine the relative contribution of each taxon to a GO term, all peptides are assigned a lowest common ancestor (LCA) of each top BLAST hit for the metagenome database proteins containing a peptide. An example of this analysis, also investigating community OM processing, can be found in Mikan et al., 2020 Mikan et al., 2020.

To address **Question 3.4**, I'll continue to probe both *de novo* and database-identified peptides to:

1. Determine quality and quantity (semi-quantitatively by spectral counts) of peptide mass modifications.

2. Track the export of protein made by primary producers. In my preliminary look at 6 ETNP POM samples I found degraded Cyanobacterial peptides by LCA peptide analysis. This is peptide evidence of bacterial export that has been explored by other means (lipid signatures, Close et al., 2013, and network analysis, Richardson and Jackson, 2007) but not in an ODZ. Figure 4 shows the relative numbers of peptides sequenced and identified as exclusively belonging to the phylum *Cyanobacteria* in suspended and sinking particles from Station P2.
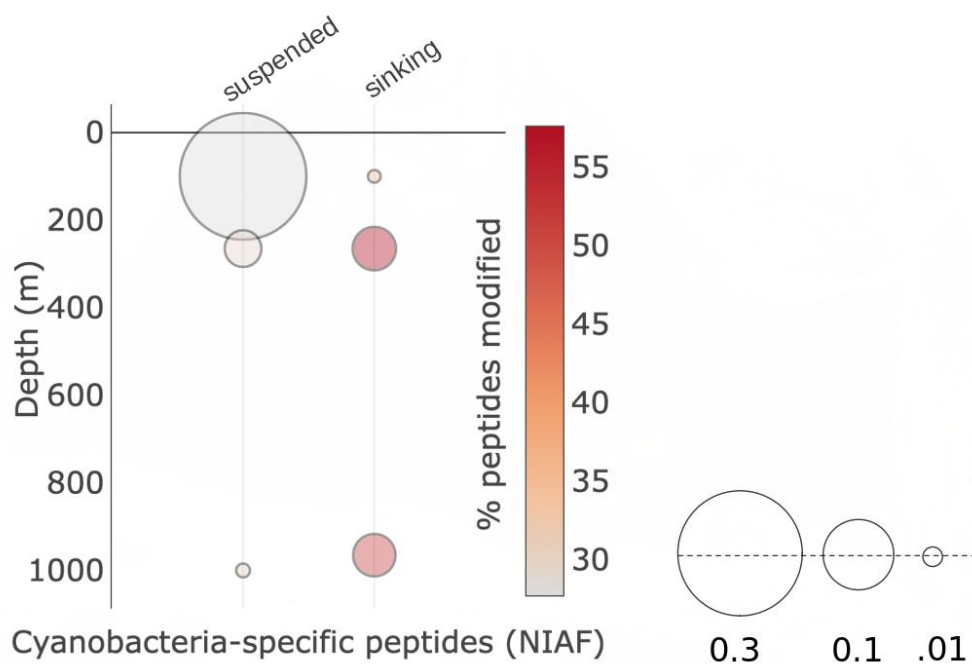
Figure 4: Cyanobacteria-specific peptides sequenced from suspended and sinking POM from offshore Station P2, ETNP January 2017 with sampling depth. Bubble sizes represent peptide abundances scaled by the normalized ion abundance factor (NIAF) of the MS1 parent ion spectral counts. Color scale represents the percentage of peptides in each samples with mass modifications.

Progress

- 2017 and 2018 total mass flux samples analyzed (**Q3.1**)

- 2017 (OC normalized) and 2018 (raw) N_2 loss rates calculated (**Q3.2**)

- 2017 and (most) 2018 metaproteomics samples analyzed (**Q3.3, 3.4**)

Expected outcomes

Recent work has deconvoluted the boundaries of different water masses in the ETNP Evans et al., 2020 and modeled injections of oxygen into the ODZ Margolskee et al., 2019. Questions of seasonality in this ODZ have given way to questions about smaller time scale variability driven by eddies, countercurrents, and subsurface zonal jets. The impacts of this variability has yet to be linked to the organic carbon cycling of the ETNP. The work of this large field endeavor provides an extremely valuable carbon-based dataset that covers time and space in ways that can inform future modeling efforts to predict future changes of the ocean's largest ODZ.

This continues to be highly collaborative project. I am leading the creation of a single flux, OM processing, and nitrogen loss rate manuscript with exact authorship to be determined.

1.4 Chapter 4: Organic matter exchange in the Lower Amazon River

1.4.1 The Lower Amazon River-to-Ocean Continuum

The classically conceived role of rivers is that they simply export OM to the oceans and that once there, long-term preservation of terrestrially-derived OM occurs largely along continental margins. However, in recent decades rivers have come to be seen not simply as transitory pipes, but as themselves transformers and regulators that adjust the carbon cycle of not only their watersheds but also of the marine receiving waters. This paradigm shift is a result of the discovery that rivers and other inland waters outgas immense quantities of CO_2 and CH_4 to the atmosphere Butman and Raymond, 2011; Richey et al., 2002.

Estimating these riverine carbon fluxes is logistically difficult, and depending on their actual magnitude, the global terrestrial CO_2 sink may

prove to be smaller than currently estimated because rivers may mobilize and remineralize a significant component of the OM pool that is considered sequestered in soils. This poses two primary questions – **what are the geographic distributions and magnitude of aquatic CO_2 outgassing, and what are the dynamics that drive this outgassing?**

The high levels of CO_2 supersaturation that drive gas evasion from large rivers are thought to be largely produced through *in situ* respiration by heterotrophic microbes. These communities utilize OM that is fixed on land and then flushed into rivers Mayorga et al., 2005; N. D. Ward, Richard G. Keil, et al., 2013; N. D. Ward, T. S. Bianchi, et al., 2016. However, there is little knowledge of what terrestrial OM compounds actually fuel this respiration or the range of their turnover rates. Outstanding unknowns to achieving this goal are: evaluating the types of OM that are degraded in the river, their sources, the microorganisms and consortia that degrade the fixed carbon, and the metabolic pathways (aerobic, anaerobic) that lead to the large outgassing of CO_2 and CH_4. A large team of scientists from Brazil and the U.S. led by Dr. Jeff Richey is funded to address these issues in the lower Amazon, by far the world's largest river system by discharge, using a combination of organic geochemical and molecular biological tools.

My role in this collaboration is 1) providing a proteomic window into the microbial community processing of OM as it changes along the continuum and with hydrologic conditions, and 2) characterizing the proteinaceous component of this OM along the same gradients. The primary research questions I will address are:

Question 4.1 What is the microbial community composition and functioning of the lower river-to-ocean transition - what are the OM degradation pathways fueling the observed CO_2 exchange? Do these shift with different hydrologic regimes (high vs low water)?

Question 4.2 What is the turnover of terrestrial OM (proteins, lignins, black carbon) in the river-to-ocean continuum across the different flow regimes of the year?

I will address **Question 4.1** using a combination of *de novo*-proteomics, underway hydrologic and gas flux measurements, and metatransciptomic analyses from previous Doherty et al., 2017; Satinsky, Zielinski, et al., 2014 and current collaborators. I'll address **Question 4.2** using the aforementioned techniques and datasets in addition to shipboard incubations and

Figure 5: A) Lower Amazon sampling stations, April 2019 on the *B/M Mirage* (C) and B) stations for metagenomic and metatranscriptomic inventories of the lower Amazon river Satinsky, Fortunato, et al., 2015 in May 2011 and the Amazon plume in June, 2010 Satinsky, Zielinski, et al., 2014. Figure adapted from Satinsky et al., 2015 Satinsky, Fortunato, et al., 2015.

collaborators' geochemical measurements of lignin, black carbon, and DOM characterization by FTICR-MS.

1.4.2 Metaproteomic insights across the Continuum

I've participated in one lower Amazon expedition (April 2019) and had additional samples collected by Jeff Richey in November 2019. Currently I have large suite of metaproteomics samples (approximately 200 with replicates) from 4 stations of the lower Amazon reach (Figure 5), spanning high- and low- water regimes of the hydrograph, as well as across tidal cycles. Figure 5 shows the 4 stations sampled in April 2019 (at rising water) on the *B/M Mirage*. Beside proteomics sampling, all four stations were sampled for DOC, DIC, alkalinity, CDOM, nutrients, and major ions. We made continuous underway measurements of pH, dissolved O_2, temperature, conductivity, Chl-a, turbidity, and FDOM using YSI EXO^2 sondes. Likewise, the concentration and stable isotopic composition of CO_2 /CH_4 were continuously measured.

Originally, a future expedition on a boat capable of reaching further into geographic mouth and near plume was planned for Spring 2020; I hope to

participate on this expedition when rescheduled, though those samples' integration into my thesis would depend on timing. This would connect sampling from earlier Richey-led Amazon projects upriver from Macapá and genomic surveys north into the Atlantic's Amazon plume (Figure 5B). To address **Question 4.1**, I will:

1. Perform *de novo*-assisted metaproteomics with Peaks using an assembled metagenome-derived database constructed from published data Satinsky, Zielinski, et al., 2014; Doherty et al., 2017; Satinsky, Fortunato, et al., 2015; Ghai et al., 2011.

2. Analyze database search results with MetaGOmics Riffle et al., 2017, focusing on OM processing.

3. Compare metaproteomics results with underway CTD and gas measurements, ADCP, nutrients, C + N, and DOM (Waterhackweek project, August 2020).

I hypothesize that community structural and functional shifts will co-occur with changes in DOM composition. In particular, I expect that the abundance of proteins related to specific functions (e.g. lignin degradation, primary production, and nutrient cycling) will be closely related to changes in DOM molecular formulae throughout tidal cycles, incubations, and along the study domain. I've been accepted as a participant at the Consortium of Universities for the Advancement of Hydrologic Science Waterhackweek workshop currently rescheduled for August 2020, where I proposed bringing these metaproteomic, geochemical and hydrologic datasets with the goal of synthesizing them into a searchable dataproduct. I anticipate microbial community shifts to be structured by hydrographic and tidal conditions, as has been found in Arctic rivers to the point of using genes as an accurate predictor river discharge ('genohydrography', Good, URycki, and B. C. Crump, 2018). Several other studies have established links between river flow rates and community composition Byron C. Crump and Hobbie, 2005, including far out into the Amazon plume Doherty et al., 2017.

1.4.3 Organic matter degradation incubations

While my goal is to enlarge the geographical lens of research in this section of the river-to-ocean domain and make to connections to pools of DOM and outgassing, I'm also interested in timescales of bacterial OM transformations and the resulting reactivity of processed material. Ward et al. N. Ward et al., 2018 developed rotating incubation systems that better simulate river flow and particle suspension. These have been important in deciphering the source of highly reactive materials, showing that lignins, often thought to be fairly recalcitrant Hedges, Clark, and Come, 1988; Gough et al., 1993; Opsahl and Benner, 1997, are in fact remineralized in the lower Amazon. Conversely, though proteins are thought to be very labile, the mineral-laden river water may slow degradation rates due to mineral surface-protein interactions Richard G. Keil, Montluçon, et al., 1994; Mayer, 1994. For this reason, I conducted incubations using the same rotating chambers paired with biochemical oxygen demand (BOD) O_2 respiration measurements in April 2019 at 4 stations (Figure 5) with the goal of determining protein turnover rates and contribution to heterotrophic respiration (**Question 4.2**). Water (4 L) was collected from 50% river depth and incubated for 24-hr in dark, rotating chambers. I filtered 300 mL onto 2.7 and 0.3 μm glass fiber membranes in triplicate at the beginning and end of each incubation, and retained 1 L filtrate samples in triplicate for geochemical analyses frozen in amber polypropylene bottles. I plan to repeat these incubations further into the river mouth on the next field expedition, if possible.

Progress

1. Extracted protein from the April 2019 filtered water and incubation samples in duplicate in the lab at UW in preparation for LC-HRMS when the UW Proteomics Resource Center (and campus) is open again. Protein extraction concentrations as determined by a Lowry assay show sufficient protein for analysis.

2. Constructed a search database from previous upriver and plume expeditions' metagenomic and metatranscriptomic publications Satinsky, Zielinski, et al., 2014; Doherty et al., 2017; Satinsky, Fortunato, et al., 2015; Ghai et al., 2011(see Figure 5B).

Expected outcomes

This large project will likely produce several publications that will address a key knowledge gap about large rivers' carbon processing variability. My Chapter 4 publication will focus on OM processing across the study domain through a metaproteomic and geochemical viewpoint, looking at least at protein degradation and sources through using *de novo*-assisted peptide sequencing and collaborators' dissolved black carbon, lignin, and DOM characterizations.

2 COVID-19 Considerations

As of writing, UW Oceanography operations are proceeding remotely, with only essential work happening on campus. I have remote access to a computer to run Peaks (*de novo*-assisted sequencing), and have access to all 2017 and most 2018 ETNP proteomic data, stable C and N isotope data, and metadata. All Amazon proteomics samples (April 2019 and November 2019) are in Seattle, with about 1/2 of April 2019 samples extracted and digested (including from all incubations) but not yet run at the UW Proteomics Resource Center. While somewhat unsatisfying, I believe that I can complete all aspects of my thesis, save some ETNP 2019 proteomics samples and further Amazon proteomics work. Rick and I have discussed the unlikely event that I cannot enter the lab again within the next year and we believe that I have enough laboratory work already completed to have a robust and discovery-filled thesis.

3 Timeline of research

Combined M.S./Ph.D. from September 2015 - December 2021 (6.5 years):

	Winter	Spring	Summer	Fall
2019-2020	Co-instruct Ocean 295	General Exam	Outreach: Ocean Intern? (3-weeks)	Waterhackweek (August 31-Sept 6th)
Year 5	Presentations: Ocean Sciences Meeting	Dept. seminar (June 5)		
	Lab work: Chp 4 Amazon proteomics	Proposal writing: Amazon metatranscriptomics at DOE EMS Lab	Lab work: Chp 4 Amazon proteomics	Data analysis: Chp 4 proteomics and data synthesis
		Writing: Chp 2 Algae rot		
			Writing set up: Chp 3 flux; Lab work: rates(?)	Data analysis: Chp 3 flux + proteomics
	Funding: NSF GRFP			Funding: TROCAS II RA
2020-2021			Gordon Research Seminar in Environmental Sciences: Water	Defend dissertation
Year 6	Writing: Chp 4 Amazon proteomics	Possible Amazon River mouth expedition?		
		Final data analysis and writing: Chp 3 rates + proteomics		
	Funding: TROCAS II RA, potential DOE funding			

4 References

Dong, Hong-Po, Da-Zhi Wang, Minhan Dai, and Hua-Sheng Hong (July 1, 2010). "Characterization of particulate organic matter in the water column of the South China Sea using a shotgun proteomic approach". In: *Limnology and Oceanography* 55.4, pp. 1565–1578. ISSN: 1939-5590. DOI: 10.4319/lo.2010.55.4.1565.

Bridoux, Maxime, Jaqui Neibauer, Anitra Ingalls, Brook Nunn, and Richard Keil (Feb. 28, 2015). "Suspended marine particulate proteins in coastal and oligotrophic waters". In: *Journal of Marine Systems* 143. DOI: 10.1016/j.jmarsys.2014.10.014.

Bergauer, Kristin, Antonio Fernandez-Guerra, Juan A. L. Garcia, Richard R. Sprenger, Ramunas Stepanauskas, Maria G. Pachiadaki, Ole N. Jensen, and Gerhard J. Herndl (Dec. 14, 2017). "Organic matter processing by microbial communities throughout the Atlantic water column as revealed by metaproteomics". In: *Proceedings of the National Academy of Sciences*, p. 201708779. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1708779115.

Moore, Eli K., Brook L. Nunn, David R. Goodlett, and H. Rodger Harvey (Apr. 15, 2012). "Identifying and tracking proteins through the marine water column: insights into the inputs and preservation mechanisms of protein in sediments". In: *Geochimica et cosmochimica acta* 83, pp. 324–359. ISSN: 0016-7037. DOI: 10.1016/j.gca.2012.01.002.

Saito, Mak A. et al. (Apr. 5, 2019). "Progress and Challenges in Ocean Metaproteomics and Proposed Best Practices for Data Sharing". In: *Journal of Proteome Research* 18.4, pp. 1461–1476. ISSN: 1535-3893. DOI: 10.1021/acs.jproteome.8b00761.

Kim, Geumsoo, Stephen J. Weiss, and Rodney L. Levine (Feb. 2014). "Methionine Oxidation and Reduction in Proteins". In: *Biochimica et biophysica acta* 1840.2. ISSN: 0006-3002. DOI: 10.1016/j.bbagen.2013.04.038.

Chick, Joel M., Deepak Kolippakkam, David P. Nusinow, Bo Zhai, Ramin Rad, Edward L. Huttlin, and Steven P. Gygi (July 2015). "A mass-tolerant database search identifies a large proportion of unassigned spectra in shotgun proteomics as modified peptides". In: *Nature Biotechnology* 33.7, pp. 743–749. ISSN: 1546-1696. DOI: 10.1038/nbt.3267.

Muth, Thilo, Carolin A. Kolmeder, Jarkko Salojärvi, Salla Keskitalo, Markku Varjosalo, Froukje J. Verdam, Sander S. Rensen, Udo Reichl, Willem M. de Vos, Erdmann Rapp, and Lennart Martens (Oct. 2015). "Navigat-

ing through metaproteomics data: a logbook of database searching". In: *Proteomics* 15.20, pp. 3439–3453. ISSN: 1615-9861. DOI: `10.1002/pmic.201400560`.

Lu, Bingwen and Ting Chen (Mar. 1, 2004). "Algorithms for de novo peptide sequencing using tandem mass spectrometry". In: *Drug Discovery Today: BIOSILICO* 2.2, pp. 85–90. ISSN: 1741-8364. DOI: `10.1016/S1741-8364(04)02387-X`.

Chisholm, Sallie W., Robert J. Olson, Erik R. Zettler, Ralf Goericke, John B. Waterbury, and Nicholas A. Welschmeyer (July 1988). "A novel free-living prochlorophyte abundant in the oceanic euphotic zone". In: *Nature* 334.6180, pp. 340–343. ISSN: 1476-4687. DOI: `10.1038/334340a0`.

Paul, Sandip, Anirban Dutta, Sumit K Bag, Sabyasachi Das, and Chitra Dutta (Feb. 10, 2010). "Distinct, ecotype-specific genome and proteome signatures in the marine cyanobacteria Prochlorococcus". In: *BMC Genomics* 11, p. 103. ISSN: 1471-2164. DOI: `10.1186/1471-2164-11-103`.

Ma, Bin, Kaizhong Zhang, Christopher Hendrie, Chengzhi Liang, Ming Li, Amanda Doherty-Kirby, and Gilles Lajoie (2003). "PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry". In: *Rapid communications in mass spectrometry: RCM* 17.20, pp. 2337–2342. ISSN: 0951-4198. DOI: `10.1002/rcm.1196`.

Van Mooy, Benjamin A. S, Richard G Keil, and Allan H Devol (Feb. 1, 2002). "Impact of suboxia on sinking particulate organic carbon: Enhanced carbon flux and preferential degradation of amino acids via denitrification". In: *Geochimica et Cosmochimica Acta* 66.3, pp. 457–465. ISSN: 0016-7037. DOI: `10.1016/S0016-7037(01)00787-6`.

Nunn, Brook L., Ying S. Ting, Lars Malmström, Yihsuan S. Tsai, Angela Squier, David R. Goodlett, and H. Rodger Harvey (2010). "The path to preservation: Using proteomics to decipher the fate of diatom proteins during microbial degradation". In: *Limnology and Oceanography* 55.4, pp. 1790–1804. ISSN: 1939-5590. DOI: `10.4319/lo.2010.55.4.1790`.

Wolfe, Alexander, Mark Edlund, Arthur Sweet, and Steven Creighton (June 26, 2006). "A First Account of Organelle Preservation in Eocene Nonmarine Diatoms: Observations and Paleobiological Implications". In: *PALAIOS* 21, pp. 298–304. DOI: `10.2110/palo.2005.p05-14e`.

Abdulla, Hussain A., David J. Burdige, and Tomoko Komada (Feb. 15, 2018). "Accumulation of deaminated peptides in anoxic sediments of Santa Barbara Basin". In: *Geochimica et Cosmochimica Acta* 223, pp. 245–258. ISSN: 0016-7037. DOI: `10.1016/j.gca.2017.11.021`.

Dauwe, Birgit and Jack J. Middelburg (1998). "Amino acids and hexosamines as indicators of organic matter degradation state in North Sea sediments". In: *Limnology and Oceanography* 43.5, pp. 782–798. ISSN: 1939-5590. DOI: `10.4319/lo.1998.43.5.0782`.

Keeling, Patrick J. et al. (June 24, 2014). "The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): Illuminating the Functional Diversity of Eukaryotic Life in the Oceans through Transcriptome Sequencing". In: *PLOS Biology* 12.6, e1001889. ISSN: 1545-7885. DOI: `10.1371/journal.pbio.1001889`.

Yooseph, Shibu et al. (Mar. 13, 2007). "The Sorcerer II Global Ocean Sampling Expedition: Expanding the Universe of Protein Families". In: *PLOS Biology* 5.3, e16. ISSN: 1545-7885. DOI: `10.1371/journal.pbio.0050016`.

Dauwe, Birgit, Jack J. Middelburg, Peter M. J. Herman, and Carlo H. R. Heip (Nov. 1, 1999). "Linking diagenetic alteration of amino acids and bulk organic matter reactivity". In: *Limnology and Oceanography* 44.7, pp. 1809–1814. ISSN: 1939-5590. DOI: `10.4319/lo.1999.44.7.1809`.

DeVries, T., C. Deutsch, P. A. Rafter, and F. Primeau (Apr. 15, 2013). "Marine denitrification rates determined from a global 3-D inverse model". In: *Biogeosciences* 10.4, pp. 2481–2496. ISSN: 1726-4189. DOI: `10.5194/bg-10-2481-2013`.

Deutsch, Curtis, Holger Brix, Taka Ito, Hartmut Frenzel, and LuAnne Thompson (July 15, 2011). "Climate-Forced Variability of Ocean Hypoxia". In: *Science* 333.6040, pp. 336–339. ISSN: 0036-8075, 1095-9203. DOI: `10.1126/science.1202422`.

Bopp, L., L. Resplandy, J. C. Orr, S. C. Doney, J. P. Dunne, M. Gehlen, P. Halloran, C. Heinze, T. Ilyina, R. Séférian, J. Tjiputra, and M. Vichi (Oct. 2, 2013). "Multiple stressors of ocean ecosystems in the 21st century: projections with CMIP5 models". In: *Biogeosciences* 10.10, pp. 6225–6245. ISSN: 1726-4189. DOI: `10.5194/bg-10-6225-2013`.

Paulmier, A. and D. Ruiz-Pino (Mar. 2009). "Oxygen minimum zones (OMZs) in the modern ocean". In: *Progress in Oceanography* 80.3, pp. 113–128. ISSN: 00796611. DOI: `10.1016/j.pocean.2008.08.001`.

Devol, Allan H. and Hilairy E. Hartnett (Nov. 1, 2001). "Role of the oxygen-deficient zone in transfer of organic carbon to the deep ocean". In: *Limnology and Oceanography* 46.7, pp. 1684–1690. ISSN: 1939-5590. DOI: `10.4319/lo.2001.46.7.1684`.

Keil, Richard G., Jacquelyn A. Neibauer, Christina Biladeau, Kelsey van der Elst, and Allan H. Devol (Apr. 8, 2016). "A multiproxy approach to

understanding the "enhanced" flux of organic matter through the oxygen-deficient waters of the Arabian Sea". In: *Biogeosciences* 13.7, pp. 2077–2092. ISSN: 1726-4170. DOI: `https://doi.org/10.5194/bg-13-2077-2016`.

Ganesh, Sangita et al. (Nov. 2018). "Single cell genomic and transcriptomic evidence for the use of alternative nitrogen substrates by anammox bacteria". In: *The ISME Journal* 12.11, pp. 2706–2722. ISSN: 1751-7370. DOI: `10.1038/s41396-018-0223-9`.

Fuchsman, Clara A., Hilary I. Palevsky, Brittany Widner, Megan Duffy, Michael C. G. Carlson, Jacquelyn A. Neibauer, Margaret R. Mulholland, Richard G. Keil, Allan H. Devol, and Gabrielle Rocap (June 27, 2019). "Cyanobacteria and cyanophage contributions to carbon and nitrogen cycling in an oligotrophic oxygen-deficient zone". In: *The ISME Journal*, p. 1. ISSN: 1751-7370. DOI: `10.1038/s41396-019-0452-6`.

Babbin, A. R., R. G. Keil, A. H. Devol, and B. B. Ward (Apr. 25, 2014). "Organic Matter Stoichiometry, Flux, and Oxygen Control Nitrogen Loss in the Ocean". In: *Science* 344.6182, pp. 406–408. ISSN: 0036-8075, 1095-9203. DOI: `10.1126/science.1248364`.

Ploug, Helle, Morten Iversen, and Gerhard Fischer (Sept. 1, 2008). "Ballast, sinking velocity, and apparent diffusivity within marine snow and zooplankton fecal pellets: Implications for substrate turnover by attached bacteria". In: *Limnology and oceanography* 53, pp. 1878–1886. DOI: `10.4319/lo.2008.53.5.1878`.

Ploug, Helle (2001). "Small-scale oxygen fluxes and remineralization in sinking aggregates". In: *Limnology and Oceanography* 46.7, pp. 1624–1631. ISSN: 1939-5590. DOI: `10.4319/lo.2001.46.7.1624`.

Bianchi, Daniele, Thomas S. Weber, Rainer Kiko, and Curtis Deutsch (Apr. 2018). "Global niche of marine anaerobic metabolisms expanded by particle microenvironments". In: *Nature Geoscience* 11.4, pp. 263–268. ISSN: 1752-0908. DOI: `10.1038/s41561-018-0081-0`.

Wakeham, Stuart G., Cindy Lee, John I. Hedges, Peter J. Hernes, and Michael J. Peterson (Dec. 1, 1997). "Molecular indicators of diagenetic status in marine organic matter". In: *Geochimica et Cosmochimica Acta* 61.24, pp. 5363–5369. ISSN: 0016-7037. DOI: `10.1016/S0016-7037(97)00312-8`.

Riffle, Michael, Damon H. May, Emma Timmins-Schiffman, Molly P. Mikan, Daniel Jaschob, William Stafford Noble, and Brook L. Nunn (Dec. 27, 2017). "MetaGOmics: A Web-Based Tool for Peptide-Centric Functional

and Taxonomic Analysis of Metaproteomics Data". In: *Proteomes* 6.1. ISSN: 2227-7382. DOI: 10.3390/proteomes6010002.

Mikan, Molly P., H. Rodger Harvey, Emma Timmins-Schiffman, Michael Riffle, Damon H. May, Ian Salter, William S. Noble, and Brook L. Nunn (Jan. 2020). "Metaproteomics reveal that rapid perturbations in organic matter prioritize functional restructuring over taxonomy in western Arctic Ocean microbiomes". In: *The ISME Journal* 14.1, pp. 39–52. ISSN: 1751-7370. DOI: 10.1038/s41396-019-0503-z.

Close, Hilary G., Sunita R. Shah, Anitra E. Ingalls, Aaron F. Diefendorf, Eoin L. Brodie, Roberta L. Hansman, Katherine H. Freeman, Lihini I. Aluwihare, and Ann Pearson (July 30, 2013). "Export of submicron particulate organic matter to mesopelagic depth in an oligotrophic gyre". In: *Proceedings of the National Academy of Sciences* 110.31, pp. 12565–12570. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1217514110.

Richardson, Tammi L. and George A. Jackson (Feb. 9, 2007). "Small Phytoplankton and Carbon Export from the Surface Ocean". In: *Science* 315.5813, pp. 838–840. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1133471.

Evans, Zachary C., Elisabeth Boles, Jarek V. Kwiecinski, Susan Mullen, Martin Wolf, Allan H. Devol, Rintaro Moriyasu, SungHyun Nam, Andrew R. Babbin, and James W. Moffett (2020). "The role of water masses in shaping the distribution of redox active compounds in the Eastern Tropical North Pacific oxygen deficient zone and influencing low oxygen concentrations in the eastern Pacific Ocean". In: *Limnology and Oceanography* n/a (n/a). ISSN: 1939-5590. DOI: 10.1002/lno.11412.

Margolskee, A., H. Frenzel, S. Emerson, and C. Deutsch (2019). "Ventilation Pathways for the North Pacific Oxygen Deficient Zone". In: *Global Biogeochemical Cycles* 33.7, pp. 875–890. ISSN: 1944-9224. DOI: 10.1029/2018GB006149.

Butman, David and Peter A. Raymond (Dec. 2011). "Significant efflux of carbon dioxide from streams and rivers in the United States". In: *Nature Geoscience* 4.12, pp. 839–842. ISSN: 1752-0908. DOI: 10.1038/ngeo1294.

Richey, Jeffrey E., John M. Melack, Anthony K. Aufdenkampe, Victoria M. Ballester, and Laura L. Hess (Apr. 2002). "Outgassing from Amazonian rivers and wetlands as a large tropical source of atmospheric CO 2". In: *Nature* 416.6881, pp. 617–620. ISSN: 1476-4687. DOI: 10.1038/416617a.

Mayorga, Emilio, Anthony K. Aufdenkampe, Caroline A. Masiello, Alex V. Krusche, John I. Hedges, Paul D. Quay, Jeffrey E. Richey, and Thomas A.

Brown (July 2005). "Young organic matter as a source of carbon dioxide outgassing from Amazonian rivers". In: *Nature* 436.7050, pp. 538–541. ISSN: 1476-4687. DOI: `10.1038/nature03880`.

Ward, Nicholas D., Richard G. Keil, Patricia M. Medeiros, Daimio C. Brito, Alan C. Cunha, Thorsten Dittmar, Patricia L. Yager, Alex V. Krusche, and Jeffrey E. Richey (July 2013). "Degradation of terrestrially derived macromolecules in the Amazon River". In: *Nature Geoscience* 6.7, pp. 530–533. ISSN: 1752-0908. DOI: `10.1038/ngeo1817`.

Ward, Nicholas D., Thomas S. Bianchi, Henrique O. Sawakuchi, William Gagne-Maynard, Alan C. Cunha, Daimio C. Brito, Vania Neu, Aline de Matos Valerio, Rodrigo da Silva, Alex V. Krusche, Jeffrey E. Richey, and Richard G. Keil (2016). "The reactivity of plant-derived organic matter and the potential importance of priming effects along the lower Amazon River". In: *Journal of Geophysical Research: Biogeosciences* 121.6, pp. 1522–1539. ISSN: 2169-8961. DOI: `10.1002/2016JG003342`.

Doherty, Mary, Patricia L. Yager, Mary Ann Moran, Victoria J. Coles, Caroline S. Fortunato, Alex V. Krusche, Patricia M. Medeiros, Jérôme P. Payet, Jeffrey E. Richey, Brandon M. Satinsky, Henrique O. Sawakuchi, Nicholas D. Ward, and Byron C. Crump (2017). "Bacterial Biogeography across the Amazon River-Ocean Continuum". In: *Frontiers in Microbiology* 8, p. 882. ISSN: 1664-302X. DOI: `10.3389/fmicb.2017.00882`.

Satinsky, Brandon M., Brian L. Zielinski, Mary Doherty, Christa B. Smith, Shalabh Sharma, John H. Paul, Byron C. Crump, and Mary Ann Moran (May 15, 2014). "The Amazon continuum dataset: quantitative metagenomic and metatranscriptomic inventories of the Amazon River plume, June 2010". In: *Microbiome* 2.1, p. 17. ISSN: 2049-2618. DOI: `10.1186/2049-2618-2-17`.

Satinsky, Brandon M., Caroline S. Fortunato, Mary Doherty, Christa B. Smith, Shalabh Sharma, Nicholas D. Ward, Alex V. Krusche, Patricia L. Yager, Jeffrey E. Richey, Mary Ann Moran, and Byron C. Crump (Sept. 10, 2015). "Metagenomic and metatranscriptomic inventories of the lower Amazon River, May 2011". In: *Microbiome* 3, p. 39. ISSN: 2049-2618. DOI: `10.1186/s40168-015-0099-0`.

Ghai, Rohit, Francisco Rodriguez-Valera, Katherine D. McMahon, Danyelle Toyama, Raquel Rinke, Tereza Cristina Souza de Oliveira, José Wagner Garcia, Fernando Pellon de Miranda, and Flavio Henrique-Silva (2011). "Metagenomics of the water column in the pristine upper course of the

Amazon river". In: *PloS One* 6.8, e23785. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0023785.

Good, S. P., D. R. URycki, and B. C. Crump (2018). "Predicting Hydrologic Function With Aquatic Gene Fragments". In: *Water Resources Research* 54.3, pp. 2424–2435. ISSN: 1944-7973. DOI: 10.1002/2017WR021974.

Crump, Byron C. and John E. Hobbie (2005). "Synchrony and seasonality in bacterioplankton communities of two temperate rivers". In: *Limnology and Oceanography* 50.6, pp. 1718–1729. ISSN: 1939-5590. DOI: 10.4319/lo.2005.50.6.1718.

Ward, Nicholas, Elise Morrison, Yina Liu, Albert Rivas-Ubach, Todd Osborne, Andrew Ogram, and Thomas Bianchi (Dec. 3, 2018). "Marine microbial community responses related to wetland carbon mobilization in the coastal zone: Coastal microbial dynamics". In: *Limnology and Oceanography Letters.* DOI: 10.1002/lol2.10101.

Hedges, John I., Wayne A. Clark, and Gregory L. Come (1988). "Fluxes and reactivities of organic matter in a coastal marine bay". In: *Limnology and Oceanography* 33.5, pp. 1137–1152. ISSN: 1939-5590. DOI: 10.4319/lo.1988.33.5.1137.

Gough, Mark A., R. Fauzi, C. Mantoura, and Martin Preston (Mar. 1, 1993). "Terrestrial plant biopolymers in marine sediments". In: *Geochimica et Cosmochimica Acta* 57.5, pp. 945–964. ISSN: 0016-7037. DOI: 10.1016/0016-7037(93)90032-R.

Opsahl, Stephen and Ronald Benner (Apr. 1997). "Distribution and cycling of terrigenous dissolved organic matter in the ocean". In: *Nature* 386.6624, pp. 480–482. ISSN: 1476-4687. DOI: 10.1038/386480a0.

Keil, Richard G., Daniel B. Montluçon, Fredrick G. Prahl, and John I. Hedges (Aug. 1994). "Sorptive preservation of labile organic matter in marine sediments". In: *Nature* 370.6490, pp. 549–552. ISSN: 1476-4687. DOI: 10.1038/370549a0.

Mayer, Lawrence M. (June 1, 1994). "Relationships between mineral surfaces and organic carbon concentrations in soils and sediments". In: *Chemical Geology* 114.3, pp. 347–363. ISSN: 0009-2541. DOI: 10.1016/0009-2541(94)90063-9.