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: directed by genes and informed by the environment. These macromolecules 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. Life, in turn, along with geological and chemical drivers, modulates the planetary cycles of carbon, oxygen, and nitrogen: at the foundation of photosynthetic fixation of atmospheric CO_2 into organic carbon is a suite of proteins, enzymes shuttling electrons, transforming energy from the sun into stored fuel.

Through unlocking the information stored in peptide and protein sequences, we learn the biological origins and functions of cells within a community. For organic geochemists, there is useful information here as well: their biological function aside, 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. In applying the recent developed tools of systems biologists, we can begin to disentangle the origins and organic matter, as well as the metabolic means of its breakdown.

My proposed Ph.D. research addresses the cycling of proteins and protein-derived organic matter in both laboratory settings and environmental systems. Chapter 1 of my Ph.D., 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 following 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

Developments in high resolution mass spectrometry and computing have facilitated metaproteomic investigations of proteins in suspended [1, 2, 3] and sinking [4] 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 proteins from living biomass, often reliant on genomic or transcriptomic information derived from the same (or very similar) sample.

Current metaproteomic workflows adhere to a bottom-up proteomic approach, relying on peptide-spectrum matching through database searching [5]. In this regime, proteins from the extracted environmental sample are proteolytically digested with an enzyme (often trypsin) that site-specifically cleaves proteins at known amino acid residues to generate a complex mixture of peptides. These resulting peptides are then separated by liquid chromatography and analyzed by high resolution mass spectrometry (LC-HRMS). Mass spectra from these observed peptides (precursor mass spectra, or MS) along with their fragmentation product spectra (MS^2) are then compared with theoretical spectra generated from an *in silico* digestion of protein sequences within a database, a process referred to as peptide-spectrum matching. Typically, these protein sequence databases are derived either from complementary metagenomic or metatranscriptomic analyses, a compilation of annotated reference proteomes of taxa thought likely to be represented in the sample, or a combination of these two. This approach has limitations given a) the taxonomically diverse and uncharacterized nature of environmental samples, and b), the fact that proteins often deviate from gene product predictions due to alternative splicing, mutations, polymorphisms, or post-translational modifications (PTM) as a result of that protein's intended biochemical function or through degradative processes [6]. 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 [7].

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 the *de novo* respective mass differences between fragment ion peaks to piece back together the original MS peptide ion amino acid sequence. Historically 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). This is because *de novo* sequencing is generally less accurate than database searching, often because of insufficient mass accuracy [8] or uneven fragmentation patterns in tandem mass spectrometry [9]. But analogous to antibodies 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

Due to its database independence, *de novo* sequencing has exciting potentials for complementing the traditional database-driven approach to environmental peptidomic and proteomic studies. The number of studies using any kind of *de novo* peptide sequencing in environmental samples is small [10], but will likely grow with as mass spectrometer resolution and scanning power increase. As such, it is critical to evaluate the accuracy of this tool in a well-characterized system. As a benchmark study, I compared the sequencing performance of a well-established *de novo* algorithm against a traditional database search on a *Prochlorococcus marinus* culture protein dataset. *P. marinus*, a ubiquitous free-living marine cyanobacterium found throughout the world's oceans [11], was an advantageous marine microbe for this assessment given its small and relatively well-characterized proteomes [12].

In this benchmark study of the *P. marinus* HRMS spectral dataset, I performed *de novo* peptide sequencing with a commercially available algorithm, Peaks [13], and database searching, aligning and comparing the two resulting groups of peptides on 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.)

In terms of proteome coverage, *de novo* sequencing did as well as ex-
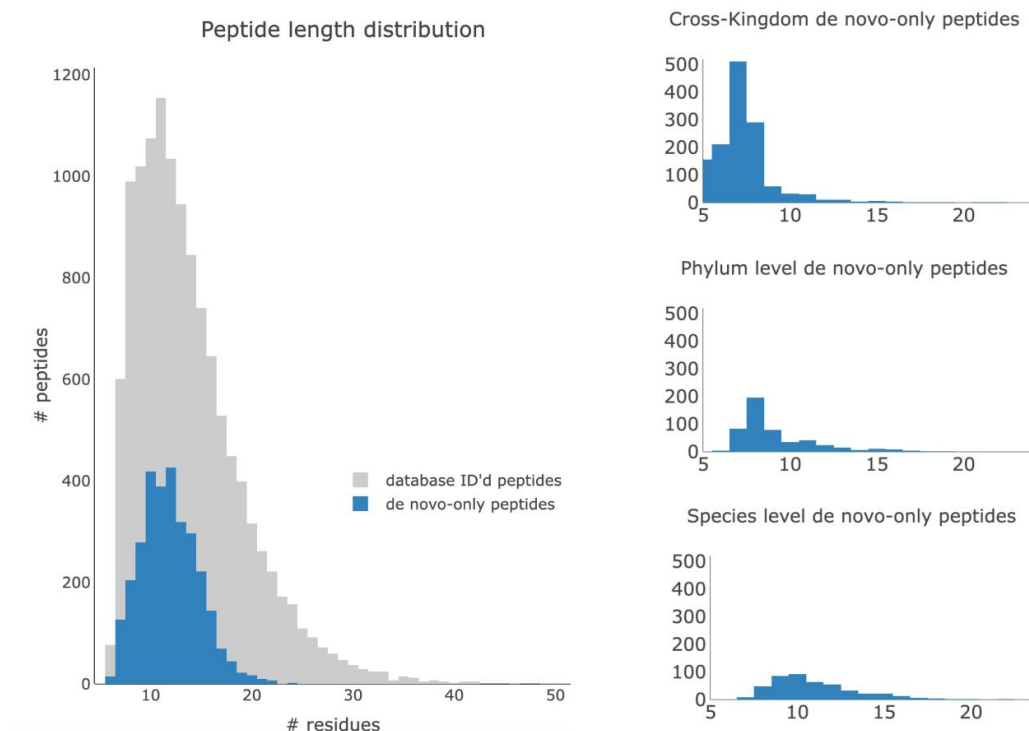
Figure 1: Histograms showing a) overlaid distributions of sequenced peptide lengths (number of amino acid residues) for database search-identified (grey) and *de novo*-only peptides (blue) of cultured *Prochlorococcus marinus* MED4 protein LC-MS/MS dataset. Of 12,160 database peptides, the mean length 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.

pected: 28% identification rate of the total predicted proteins, compared with 67% for traditional database searching. However, the *de novo* algorithm did find peptides unidentified through database searching that boosted the overall protein identification rate by 6%. The *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 and peptidomic context because with increased amino acid space, peptide sequences have more potential to be specific to a particular parent protein. Given the limitations of *de novo* sequencing, we expect the lengths of confident *de novo* peptides to be shorter than those identified by database searching.

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-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, we 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 we searched for, the

7

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.

Thus, the 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 wouldn't 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 uncharacterized systems, exemplified by the challenges to marine metaproteomic and peptidomic study. 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*. POM samples were collected with large volume *in situ* pumps that filtered between 300 and 1000 L of seawater onto 0.3 μm membranes, which is presumed to include both free-living microbes as well as small suspended detrital organic matter. These samples are hence called suspended particles. Sinking particles were collected with free-drifting, unpoisoned sediment net traps with deployments lasting between 24 and 96 hours.

I applied the *de novo*-assisted approach to this small ETNP POM sample set, finding 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 [14], suggesting that the *de novo* tool provides a molecular-level view into the processes fueling chemoautotrophy.

Progress

Both the benchmark *Prochlorococcus* peptide analysis and ETNP POM study, was submitted to *Limnology and Oceanography* in January, 2020 and is currently under review. The *de novo*-assisted peptide sequencing approach is used in Chapter 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 natural seawater. Proteins were extracted and trypsin-digested from a bloom-and-bust simulation sampled at multiple timepoints. This builds upon both laboratory [15] and environmental investigations [4] of protein breakdown. The latter example, from 2012, in which Moore and colleagues tracked diatom proteins from surface to early sediments in the Bering Sea, highlights how far metaproteomic instrumental and algorithmic tools have advanced in a short time: they identified 4 *Thalassiosira pseudonana* proteins in the surface and just 1 into early sediments through traditional database searching. We are now able to sequence and identify hundreds of algal proteins in each timepoint in this degradation dataset.

   Question 2.1 How does the peptide pool of the phytoplankton change through degradation?
   Recent work by Abdulla and colleagues showed an accumulation in anoxic sediment pore water of what seem to be deaminated peptides (shown through FTIRMS, and not proteomic sequencing, [16])I hypothesized that sequenced peptides would become:

   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

1.2.1   Experimental design

A culture of marine diatom *Thalassiosira weissflogii* brought to approximately 106 cells/mL, concentrated, 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 used to induce bacterial decomposition of algal material: 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) [17] and GoM metagenomic sequences from the Global Ocean Sampling (GOS) data product [18].

Application of *de novo*-assisted database searching reveals changes in peptide composition over degradation by a natural microbial community. Both tryptic and non-tryptic peptides sequenced from late-stage degradation material are relatively enriched amounts of glycine, serine, and threonine and depleted 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 total hydrolyzable amino acids [19], this is the first observation that the enzyme-hydrolyzable peptide pool influences this pattern. Because *de novo* sequencing is not reliant on a protein database, we are also able to identify proteins or protein fragments from organisms unexpected in the sample, particularly useful 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.

1.2.3 Evolution of peptide mass modifications

Comparison of total modification distribution to DI papers. Serine abundance. Future directions of this kind of work?

Progress

1.3 Chapter 3: High-resolution marine flux measurements and *in situ* respiration rate determinations, and metaproteomic surveys of particulate matter 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 [20], 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 [21]. Critically, climate models predict an approximate 5% decrease to the ocean's O_2 reservoir within this century [22].

The exciting application of *de novo* peptide sequencing is to uncharacterized systems, perfectly exemplified by the challenges to marine metaproteomic and peptidomic study. The eastern tropical North Pacific (ETNP) is home to the Earth's largest marine oxygen deficient zone (ODZ), accounting for approximately 41% of global marine anoxic waters [23]. 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 [24, 14, 25]. However, there remains much uncertainty about the mechanism(s) explaining these observations. In the ODZ, shifts in zooplankton behavior [25] and *in situ* production from anammox [26] 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 [27] and likely the makeup of POM controls the relative
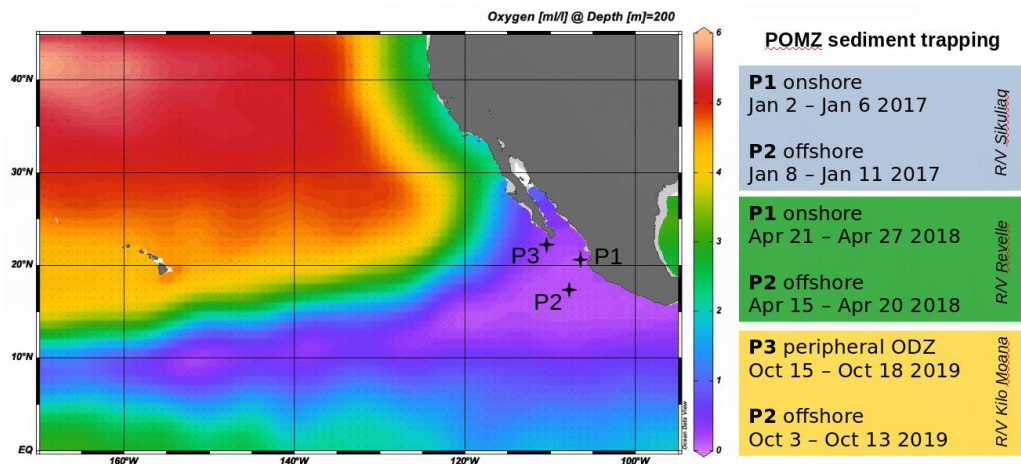
Figure 2: The eastern tropical North Pacific oxygen deficient zone (ETNP ODZ), shown here as O_2 concentration at 200 m from World Ocean Atlas data, 1955-2013. Time series sediment trapping stations (P1, P2, and P3) for the three POMZ expeditions are shown with expedition timings.

contribution of those processes [28]. Since proteinaceous matter comprises the single largest identifiable component of the sinking flux in the tropical eastern Pacific [29], evaluating its processing using the de novo-assisted approach might provide insights 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 can be described in three primary questions:

Question 3.1 How variable are OM flux profiles across seasons and from on to offshore? Is there a shift in microbial community structure and OM processing when these profiles look different?

Question 3.2 Are there patterns in peptide and protein degradation in sinking and suspended POM, with depth and between on- and offshore stations?

Question 3.3 How variable are OM flux profiles across seasons and from

12

on to offshore? Is there a shift in microbial community structure and OM processing with when these profiles look different?

1.3.2 Field Work and Preliminary Observations

Describe 3 cruises here. Describe sediment traps, briefly, and McLane pumps. Figure 2 the locations and timings of ETNP station sediment trap sampling.

1.3.3 Flux of suspended and sinking POM from 2017-2019

This highly collaborative work addresses **Question 3.1**. The Keil group has deployed sediment trap-incubator systems and large volume *in situ* pumps (McLane) on three recent research cruises to the ETNP between January, 2017 and October, 2019 (see Figure 2).

Fluxes from 2017-2018 here. Make an R version from talk Excel fig.

1.3.4 Metapeptidomic and metaproteomic signatures of POM through the ODZ

Table of run proteomics samples and extracted proteomics samples.

For a functional analysis of OM processing, I'll use the peptide-based tool Unipept [30] to identify peptides indicative of certain metabolic processes based on gene ontologies, or GO terms. Unipept uses peptide spectral counts to generate semi-quantitative BLAST-like sequence alignment against a manually annotated and reviewed section of the UniProtKB database.

Figure 3 shows the relative numbers of peptides sequenced and identified as exclusively belonging to the phylum *Cyanobacteria* in suspended and sinking particles from offshore Station P2 in the ETNP, January 2017. Color scale indicates the percentage of peptides in each sample that have mass modifications as described in Section 1.2.

1.3.5 Metabolic rate calculations from *in situ* incubation-sediment traps

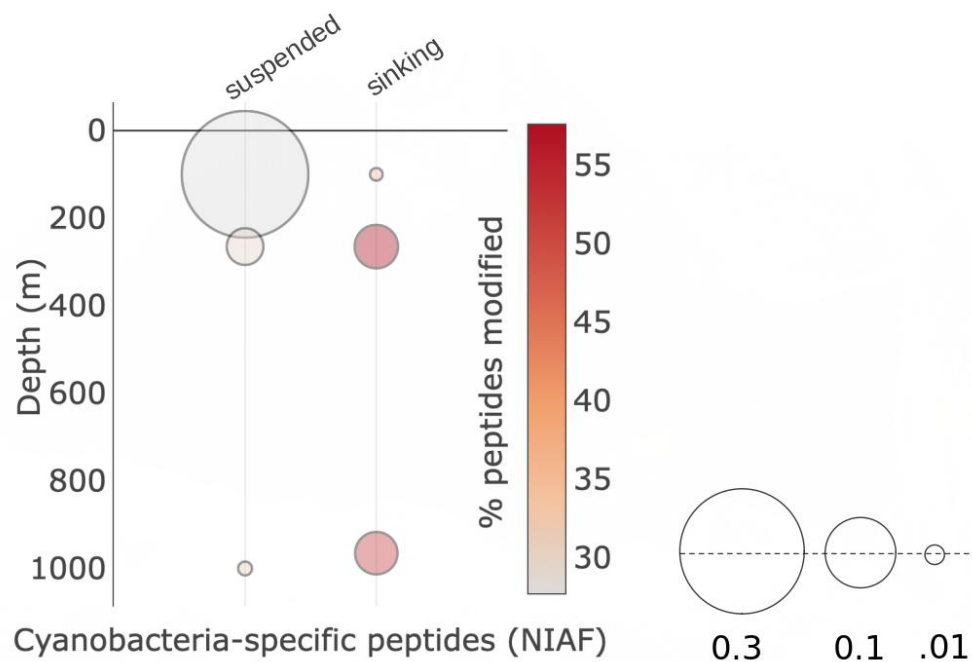2017 and 2018 rates here.

Figure 3: 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.

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

1.4.1 The Lower Amazon River-to-Ocean Continuum

Water is a major vector that moves biogeochemical constituents through Earth's reservoirs and across the continuum of the atmosphere, biota, soils, inland waters, oceans, and sediments [31]. Biological and physical mechanisms like weathering, production, respiration, as well as human land use serve to transform organic carbon-containing molecules along their journey through these systems. The classically conceived role of rivers in this transport 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 [32, 33].

Globally, it's estimated that inland waters process, transport, and bury 2.7 Pg C y^{-1} [34, 35]. As this figure nearly equals current estimates of the terrestrial sink for anthropogenic carbon (2.8 Pg C y^{-1}) [36], carbon cycling in rivers then relocates and/or mitigates terrestrial sequestration of human forcing. But 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 as a substrate [37, 38, 39]. However, there is little knowledge of what terrestrial OM compounds actually fuel this respiration or the range of their turnover rates. Work done in marine systems and

small streams has propagated an assumption that lignin and cellulose, the two major classes of macromolecles in terrestrial OM, are generally refractory in all aquatic systems [40, 41, 42]. However, recent work in the lower Amazon basin have shown that lignins do in fact contribute to heterotrophic respiration in the main stem of the river [38]. In particular, rotating incubation systems that better simulate the flow conditions of the river have been important in deciphering the source of highly reactive materials [43]. This is critical to understanding the remineralization potential of specific OM types and to elucidating the underlying mechanisms controlling their fate.

Such developments have made significant headway towards achieving the ultimate goal: calculating of the overall carbon balance of a basin and making predictions of future change. 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 world's largest river system, the Amazon, 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 hydologic conditions, and 2) characterizing the proteinaceous component of this OM along the same gradients. The primary research question I will address are:

Question 4.1 What is the microbial community composition and functioning of the lower river-to-ocean transition, and 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 How do different components of terrestrial OM (proteins, lignins, black carbon), cycling in the river-to-ocean continuum across the different flow regimes of the year?

As detailed below, I will address **Question 4.1** using a combination of *de novo*-proteomics, underway hydrologic and gas flux measurements, and metatransciptomic analyses from previous [44, 45] and current collaborators. I'll address **Question 4.2** using the aforementioned datasets and collaborators' geochemical measurements of lignin, amino acids, and POM/DOM characterization.

16

Figure 4: Lower Amazon sampling stations, April 2019.

1.4.2 Field Work and Preliminary Observations

I've participated in one lower Amazon expedition (April, 2019) and had additional samples collected in November, 2019. Currently I have large suite of metaproteomics samples (approximately 200 with replicates) from 4 stations of the lower Amazon reach (Figure 4), spanning high- and low- water regimes of the hydrograph, as well as across tidal cycles. I plan to participate in, or get samples from, a future expedition on a boat capable of reaching further into the plume (timing TBA, given current events). This will connect sampling from earlier Richey-led Amazon projects, TROCAS I (upriver from Macapá) and ROCA (into the Atlantic plume).

Map of TROCAS I and ROCA stations

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. Protein abundance will rapidly respond to changing conditions, revealing microbial stress signals and processes that

17

are too dynamic/rapid to capture through traditional genomics. 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', [46]). Several other studies have established links between river flow rates and community composition [47], including out in the Amazon plume [44]. There is precedent for anticipating that two scenarios: first, where a microbial community shifts in composition with changing OM. For example, in a study of an Arctic fjord, Paulsen and colleagues identified specific taxa that were associated with degradation of fluorescent DOM (FDOM) [48]. In a second scenario, the functioning of the community the the main response to OM, as shown in recent metaproteomic work by Mikan and colleagues with similar aims in to this project in the Chuchki Sea. They showed a shift in functional response to OM shifts in shipboard incubations simulating a phytoplankton bloom, with some slight change in the heterotrophic community makeup [49].

Figure 4 shows the 4 proteomics/DOM/incubation stations sampled in April, 2019 at rising water on the *B/M Mirage*. All four stations were also sampled for full cross channel ADCP current velocity profiles over tidal cycles, total suspended sediments, nutrients, total C and N, and gas exchange.

1.4.3 Degradation tracking incubation experiment

While my goal is to enlarge the geographical lens of research in this section of the river and make connection to the DOM pools, I'm also interested in timescales of bacterial DOM transformations and the resulting reactivity of processed material. For this reason, I conducted incubations in river condition-mimicking rotating chambers in triplicate [43].

Progress

1. Extracted protein from the April 2019 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 metatransciptomic publications [45, 44, 50,

51].

2 Timeline of research

Combined M.S./Ph.D. from September 2015 - September 2021 (6 years):

Winter 2020

- Co-instruct Ocean 295: Chemistry of Marine Organic Carbon

- Presentations: Ocean Sciences Meeting

- Lab work: Chp 4 Amazon proteomics

- Writing: Chp 2 Algae rot

- Funding: NSF GRFP

Spring 2020

- Data analysis: Chp 3 ETNP metapeptidomics

- Lab work: Chp 4 Amazon proteomics

- Funding: NSF GRFP

- Apply for DOE student grant

Summer 2020

- Outreach: Ocean Intern (3-4 weeks)

- Waterhackweek (August 31-Sept 6th)

- Writing: Chp 3

- Funding: NSF GRFP

Autumn 2020

- Funding: NSF TROCAS II

Winter 2021

- Funding: NSF TROCAS II

Spring 2021

- Funding: NSF TROCAS II

Summer 2021

- Funding: NSF TROCAS II

Autumn 2021

- Funding: NSF TROCAS II

3 References

[1] Hong-Po Dong, Da-Zhi Wang, Minhan Dai, and Hua-Sheng Hong. "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 (July 1, 2010), pp. 1565–1578. ISSN: 1939-5590. DOI: 10.4319/lo.2010.55.4.1565.

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

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

[4] Eli K. Moore, Brook L. Nunn, David R. Goodlett, and H. Rodger Harvey. "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 (Apr. 15, 2012), pp. 324–359. ISSN: 0016-7037. DOI: 10.1016/j.gca.2012.01.002.

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

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

[7] Joel M. Chick, Deepak Kolippakkam, David P. Nusinow, Bo Zhai, Ramin Rad, Edward L. Huttlin, and Steven P. Gygi. "A mass-tolerant database search identifies a large proportion of unassigned spectra in shotgun proteomics as modified peptides". In: *Nature Biotechnology*

33.7 (July 2015), pp. 743–749. ISSN: 1546-1696. DOI: 10.1038/nbt.3267.

[8]   Thilo Muth, 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. "Navigating through metaproteomics data: a logbook of database searching". In: *Proteomics* 15.20 (Oct. 2015), pp. 3439–3453. ISSN: 1615-9861. DOI: 10.1002/pmic.201400560.

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

[10]   Thilo Muth and Bernhard Y. Renard. "Evaluating de novo sequencing in proteomics: already an accurate alternative to database-driven peptide identification?" In: *Briefings in Bioinformatics* (). DOI: 10.1093/bib/bbx033.

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

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

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

[14]   Benjamin A. S Van Mooy, Richard G Keil, and Allan H Devol. "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 (Feb. 1, 2002), pp. 457–465. ISSN: 0016-7037. DOI: 10.1016/S0016-7037(01)00787-6.

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

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

[17] Patrick J. Keeling et al. "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 (June 24, 2014), e1001889. ISSN: 1545-7885. DOI: `10.1371/journal.pbio.1001889`.

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

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

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

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

[22] L. Bopp, 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. "Multiple stressors of ocean ecosystems in the 21st century: projec-

tions with CMIP5 models". In: *Biogeosciences* 10.10 (Oct. 2, 2013), pp. 6225–6245. ISSN: 1726-4189. DOI: `10.5194/bg-10-6225-2013`.

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

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

[25] Richard G. Keil, Jacquelyn A. Neibauer, Christina Biladeau, Kelsey van der Elst, and Allan H. Devol. "A multiproxy approach to understanding the "enhanced" flux of organic matter through the oxygen-deficient waters of the Arabian Sea". In: *Biogeosciences* 13.7 (Apr. 8, 2016), pp. 2077–2092. ISSN: 1726-4170. DOI: `https://doi.org/10.5194/bg-13-2077-2016`.

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

[27] Clara A. Fuchsman, 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. "Cyanobacteria and cyanophage contributions to carbon and nitrogen cycling in an oligotrophic oxygen-deficient zone". In: *The ISME Journal* (June 27, 2019), p. 1. ISSN: 1751-7370. DOI: `10.1038/s41396-019-0452-6`.

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

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

[30] The UniProt Consortium. "UniProt: the universal protein knowledge-base". In: *Nucleic Acids Research* 46.5 (Mar. 16, 2018), pp. 2699–2699. ISSN: 0305-1048. DOI: `10.1093/nar/gky092`.

[31] J. J. Cole. *Plumbing the Global Carbon Cycle: Integrating Inland Waters into the Terrestrial Carbon Budget*. 2007. URL: `https://link-springer-com.offcampus.lib.washington.edu/article/10.1007/s10021-006-9013-8` (visited on 04/02/2020).

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

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

[34] Tom J. Battin, Sebastiaan Luyssaert, Louis A. Kaplan, Anthony K. Aufdenkampe, Andreas Richter, and Lars J. Tranvik. "The boundless carbon cycle". In: *Nature Geoscience* 2.9 (Sept. 2009), pp. 598–600. ISSN: 1752-0908. DOI: `10.1038/ngeo618`.

[35] Lars J. Tranvik et al. "Lakes and reservoirs as regulators of carbon cycling and climate". In: *Limnology and Oceanography* 54.6 (2009), pp. 2298–2314. ISSN: 1939-5590. DOI: `10.4319/lo.2009.54.6_part_2.2298`.

[36] Josep G. Canadell, Corinne Le Quéré, Michael R. Raupach, Christopher B. Field, Erik T. Buitenhuis, Philippe Ciais, Thomas J. Conway, Nathan P. Gillett, R. A. Houghton, and Gregg Marland. "Contributions to accelerating atmospheric CO2 growth from economic activity, carbon intensity, and efficiency of natural sinks". In: *Proceedings of the National Academy of Sciences of the United States of America* 104.47 (Nov. 20, 2007), pp. 18866–18870. ISSN: 0027-8424. DOI: `10.1073/pnas.0702737104`.

[37] Emilio Mayorga, Anthony K. Aufdenkampe, Caroline A. Masiello, Alex V. Krusche, John I. Hedges, Paul D. Quay, Jeffrey E. Richey, and Thomas A. Brown. "Young organic matter as a source of carbon dioxide

outgassing from Amazonian rivers". In: *Nature* 436.7050 (July 2005), pp. 538–541. ISSN: 1476-4687. DOI: `10.1038/nature03880`.

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

[39] Nicholas D. Ward, 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. "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 (2016), pp. 1522–1539. ISSN: 2169-8961. DOI: `10.1002/2016JG003342`.

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

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

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

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

[44] Mary Doherty, 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. "Bacterial Biogeography across the Amazon River-Ocean Continuum". In: *Frontiers in Micro-*

biology 8 (2017), p. 882. ISSN: 1664-302X. DOI: `10.3389/fmicb.2017.00882`.

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

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

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

[48]   Maria Lund Paulsen, Oliver Müller, Aud Larsen, Eva Friis Møller, Mathias Middelboe, Mikael K. Sejr, and Colin Stedmon. "Biological transformation of Arctic dissolved organic matter in a NE Greenland fjord". In: *Limnology and Oceanography* 64.3 (2019), pp. 1014–1033. ISSN: 1939-5590. DOI: `10.1002/lno.11091`.

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

[50]   Brandon M. Satinsky, 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. "Metagenomic and metatranscriptomic inventories of the lower Amazon River, May 2011". In: *Microbiome* 3 (Sept. 10, 2015), p. 39. ISSN: 2049-2618. DOI: `10.1186/s40168-015-0099-0`.

[51]   Rohit Ghai, Francisco Rodriguez-Valera, Katherine D. McMahon, Danyelle
        Toyama, Raquel Rinke, Tereza Cristina Souza de Oliveira, José Wag-
        ner Garcia, Fernando Pellon de Miranda, and Flavio Henrique-Silva.
        "Metagenomics of the water column in the pristine upper course of the
        Amazon river". In: *PloS One* 6.8 (2011), e23785. ISSN: 1932-6203. DOI:
        `10.1371/journal.pone.0023785`.