From microbes to microbiomes: bacterial proteomes reveal strategies for overcoming environmental challenges

Miranda Cristen Mudge

A dissertation submitted

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

University of Washington

2024

Reading Committee:

Michael J. MacCoss, Chair

Brook L. Nunn

Jodi Young

Program Authorized to Offer Degree:

Molecular and Cellular Biology

University of Washington

Abstract

From microbes to microbiomes: bacterial proteomes reveal strategies for overcoming environmental challenges

Miranda Cristen Mudge

Chair of the Supervisory Committee:

Michael J. MacCoss

Genome Sciences

Bacteria exist in every environment studied on Earth due to their wide-ranging metabolic capacities. In recent years, many studies have highlighted the significant contributions of bacteria as they interact with and respond to both other organisms and their environment through chemical signals. Because bacteria quickly respond to small physical and/or chemical environmental changes by adjusting their proteome, mass spectrometry-based proteomic techniques have gained momentum as a means to study the metabolic changes of mixed communities to environmental perturbations. In the following chapters, I utilize three different mass spectrometry techniques to identify the unique ways that microbes overcome environmental challenges in three unique projects. The complexity of the mass spectrometry data acquisition strategy utilized increases with chapter number, moving from the identification of the most abundant proteins in a single-organism culture to the eventual discovery of peptides from a complex ocean microbiome that can be isolated and targeted with mass spectrometry.

In Chapter 1, I outline the many decisions that should be considered when conducting a complex proteomic or metaproteomic study. In Chapter 2, I describe adaptations of a model psychrophilic bacterium to subzero, hypersaline conditions to analyze the long-term effects of temperature, salinity, and nutrients on viability and survival and identify potential biomarkers for life detection through enrichment analyses of short amino acid chains. Chapter 3 transitions to a more complex biological system and quantitative mass spectrometry techniques. A high-resolution time series analysis of the bacterial metaproteome preceding a harmful algal bloom was utilized to isolate peptide biomarkers through quantitative discovery and targeted validation mass spectrometry analyses. In Chapter 4, the same microbiome was further characterized to elucidate metabolic patterns in the context of multiple algal blooms. Metabolic scheduling, the idea that bacteria prioritize distinct pathways at specific times of day, was explored to reveal functions that reflect global changes in the metabolic strategy of the microbiome between blooms. Finally, in Chapter 5 I reflect on the challenges still present in microbiome research, specifically in the context of metaproteomics, and discuss future directions for the research.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

This section is dedicated to all of the people who helped make this work possible. First, to Brook Nunn, thank you for being my mentor. You've given me the freedom during my PhD to pursue the projects that excite me. I feel more confident and competent as a scientist after working with you and will always remember your positive attitude and encouragement. You lead a lab that everyone wants to be a part of, and I feel very fortunate that I was able to be your student. To Mike MacCoss, thank you for letting me join your lab. I have learned so much from working alongside so many fantastic people. To my committee, thank you for your continued support, excitement, and encouragement. To Ben Kerr, Jodi Young, Bill Noble, and Ajai Dandekar, you've challenged me to think outside the box and your investment in my growth as a interdisciplinary scientist has allowed me to reach my goal of obtaining a Ph.D.

This work would not have been possible without massive efforts by individuals both in the lab and across many different universities. To Emma Timmins-Schiffman, thank you for being a steady guide throughout my time in the lab. You helped me to find the thread when all of my problems seemed like a permanently tangled mess more times than I can count, and I greatly appreciate the balance you bring to the lab. To Mike Riffle, you've been a continued source of guidance and aid throughout my projects, and I could not have completed this work without your help. To the Kubanek lab, specifically Julia Kubanek and Gabi Chelbi, thank you for partnering with me in this work. I've always felt like an integral part of the team because of you. To everyone in the Genome Sciences program who helped with fieldwork and data analysis, thank you for your excitement about the projects. To my previous mentors, Robert Delong and Nima Mosammaparast, I would not have pursued an advanced degree without your encouragement. A special acknowledgment is deserved by every individual who helped make the fieldwork possible. This project is built on your efforts. Thank you.

Finally, I'd like to thank my family. To my mom Melissa Bollinger, you've helped me believe through my whole life that I am smart and capable of doing anything I put my mind to. Thank you for helping to shape me into who I am today. To my dad Mike Welker, who has always been my family, thank you for always offering to write me a note to escape any troubles I was facing. I am better able to face challenges head on knowing that I will always have your support. To my brother Dylan Welker, who has taught me to slow down and enjoy what is important in life, you've shown me how to be adaptable to life's challenges while remaining confident in myself. Thank you to Kyle Mears for supporting me in all the ways a good partner should and for being exactly what I need even when I have no idea what I'm doing. All of my favorite memories since moving across the country have been made with you, and you've made Seattle feel like my home. To the Mears family, including Kevin, Linda, Kelly, Aaron, Eloise, Declan, Nicole, and Ben, thank you for welcoming me as family. It was a difficult transition to move away from the place I've lived most of my life, but I am always happy knowing that I now have family in Washington. And finally, thank you to my dog Brody for sticking with me through all of the late nights, long walks, and trips across the country. Throughout my PhD, one thing that remained assured is that I would always be able to return home to the happiest welcome I could hope for from you.

DEDICATION

To everyone who motivated me to keep going.

Chapter 1: INTRODUCTION

1.1 MICROBIOLOGY MODERN PERSPECTIVE

Bacteria are ubiquitous on Earth and, as such, have a large influence on both human health and the environment (O'Malley 2008). Studies of the microbiome have exponentially risen with the advent of technologies such as high-throughput DNA sequencing, increasing our ability to detect and quantify the diversity of community members and dramatically altering our understanding of the pivotal role bacteria play in major health outcomes and disease (Moran 2015; Compant *et al.,* 2019; Trevathan-Tackett *et al.,* 2019; Ogunrinola *et al.,* 2020). For example, research on the gut microbiome has profoundly transformed the field of gastroenterology (Nichols and Davenport 2021). Many of these same microbes are also found throughout terrestrial, marine, and extreme environments on Earth (Rocca *et al.,* 2019). Many specific niches are characterized by the presence of keystone taxa, or bacteria with unique metabolisms that allow them to adaptively occupy a given location. Additionally, the wide diversity of organisms on Earth shows that the microbiome can change in dynamic response to perturbations through cascades of altered taxonomic composition and active functionality, with effects observed on both human and environmental outcomes (Rocca *et al.,* 2019). The research herein provides an in-depth look at the importance of bacteria in two understudied environments that are anticipated to be adversely influenced by global climate change. Using mass spectrometry-based proteomics to interrogate the proteome in response to simulated and natural perturbations, I explore the changing functions of the bacterium *Colwellia psychrerythraea* and an ocean microbiome. Specifically, the chapters detail the temporal changes that occur in response to challenging environmental perturbations and how these changes reveal potential biomarkers for various applications, with strong implications for translational environmental microbiology.

1.2 PROTEOMICS IN MICROBIOLOGY

Proteomics is the study of the complete set of proteins within a biological system. In contrast to other ‘omics approaches, proteomics reflects the active functions of cells at the time of harvest. Proteomic techniques, like mass spectrometry (MS), are capable of capturing functionality through the measurement of peptides (Wu and MacCoss 2002), the component pieces of proteins that are tightly regulated and rapidly modified in response to minute environmental perturbations (Maier *et al.,* 2011). Recent advances in MS techniques and methods have enabled accurate quantification of peptides in a variety of sample types across a wide dynamic range (Brunner *et al.,* 2022; Zubarev 2013). While proteomics studies analyze a single organism, metaproteomics is the collection and analysis of a mixed community of organisms. Most bacteria cannot be cultured (Zamkovaya *et al.,* 2021) and even those that can, often don’t live alone in the environment (Moran 2015, Fierer 2017). Taking this into consideration, metaproteomic studies are especially useful for observing the response of a mixed community to environmental perturbations as these analyses capture the true dynamic response of cells as opposed to their predicted functional capacity (Morris *et al.,* 2010; Kleiner *et al.,* 2017; Mikan *et al.,* 2020; Shrestha *et al.,* 2021; Zhao *et al.,* 2023).

The building blocks of proteins, amino acids, have been isolated from asteroids and space dust (Potiszil *et al.,* 2023), hinting at the potential for their use by other lifeforms on exoplanets. The sequence of those amino acids yield specific peptides that can retain both taxonomic and functional information which are directly measured in traditional bottom-up mass spectrometry experiments (Gillet, Leitner, and Aebersold 2016); (M. Miller and M. Smith 2023). The confident identification of those peptides in different samples allows for direct conclusions in proteomic comparative studies and avoids inherent pitfalls of protein inference, such as mis-identification and over-interpretation (Rosenberger *et al.,* 2017; Plubell *et al.,* 2021). Proteins, in their role as cellular biological machines, reflect active functionality at the time of cell capture and reveal regulatory mechanisms and adaptations to environmental changes as organisms can change their proteomes rapidly. Microbiome analyses uncover community-scale interactions that allow certain bacteria to dominate while others are limited, while also highlighting the prevalence of functional redundancy and conservation of metabolic processes across taxonomies. Additionally, enrichments of specific bacteria in some studies may indicate the competitive advantages of keystone taxa (Venter, Smith, and Payne 2011). The study herein demonstrates the importance of selecting the appropriate analytical scale, showcasing detailed analyses at multiple levels and their applications across various fields.

1.2.1 MASS SPECTROMETRY-BASED PROTEOMIC TECHNIQUES

Recent advancements in mass spectrometry are increasing the applicability and relevance of proteomic studies. While microbiome studies are currently dominated by metagenomics through the application of high-throughput sequencing technologies (Zhang and Ning 2015), the field is expanding to harness the power of metaproteomics through applied mass spectrometry. To expand on considerations for experimental design, the specific mass spectrometry technique or method applied in a study will make a considerable impact on the biological conclusions one can make. In this work, only bottom-up proteomics, in which proteins are broken into their component peptides prior to MS analysis, will be considered (M. Miller and M. Smith 2023). In bottom-up MS proteomics, the targets being directly measured are peptides; for this reason, all interpretations will be made at the peptide level to avoid the problems that come with protein inference (Rosenberger *et al.,* 2017), that are further exacerbated when samples contain mixed communities (see discussion in Mikan *et al.,* 2020).

Data-dependent acquisition (DDA) mass spectrometry is a robust technique that captures the most abundant peptides in a sample in a stochastic manner (Bateman *et al.,* 2014). Though not directly quantitative, this approach has broad applicability and has been widely adopted in the proteomics field. More recently, data-independent acquisition (DIA) mass spectrometry was developed, which is a quantitative method that systematically samples all ions within a specified mass-to-charge window (Pino *et al.,* 2020). DIA is a powerful tool but requires targeted extraction of the signal ions resulting from fragmentation and downstream interpretation relies on a database (Lee *et al.,* 2016generally the metagenome of the sample itself or a highly related sample or organism. Database selection has a major impact on biological conclusions from metaproteomic MS analyses (Timmins-Schiffman *et al.,* 2017) and could be seen as a weakness of proteomics due to the added costs incurred, however recent advances in both AI and de novo sequencing are playing a part in making up the gap to advance proteomic capabilities (R. S. Johnson *et al.,* 2020). While both DDA and DIA are discovery-based techniques, targeted analyses are also possible. Parallel reaction monitoring is a targeted method that allows for highly sensitive and accurate quantification of a subset of peptides (Remes, Yip, and MacCoss 2020). All of these methods have advantages and disadvantages (Hu, Noble, and Wolf-Yadlin 2016). DDA is affordable and produces more manageable data, but the stochastic nature in which ions are selected for fragmentation limits researchers to a qualitative or comparative quantitation analysis. DIA allows for quantification of all peptides across a wide *m/z* range regardless of the precursor ion intensities. That said, coeluting peptides can create difficulties in downstream interpretation (Rosenberger *et al.,* 2017), especially when samples contain mixed or complicated proteomes as is the case with microbiome samples (Saito *et al.,* 2019). Quantitative accuracy is improved with targeted analyses (Doerr 2012), but only a few targets can be measured at a time and they must be known prior to conducting the analysis (van Bentum and Selbach 2021). Due to the inherent complexities in downstream analyses, experimental design when using mass spectrometry-based proteomics is a critical decision point. While this is true in any protein study, it becomes even more important when trying to understand how a mixed microbial community changes through time.

1.3 COMPLEXITY OF MICROBIOME DATA ANALYSIS

Despite MS-based proteomics generating an abundance of peptide spectral data, final interpretation of active metabolic functions from big data remains challenging. In the following chapters, a variety of approaches are utilized to extract meaningful and statistically significant trends without bias across various data structures and biological questions. This work repeatedly emphasizes the power of dimensionality reduction techniques, such as non-metric multidimensional scaling (NMDS)(Mudge *et al.,* 2021) and principal component analysis (PCA)(Boyd *et al.,* 2016), which help isolate drivers of variation between samples under various conditions or across timepoints. Choosing the most relevant statistical method is also crucial for accurate downstream data interpretation. For example, in Chapter 2 log fold change analysis distinguishes significant differences in protein abundance between experimental conditions, revealing relevant metabolic modifications to environmental perturbations. In Chapters 3 and 4 more advanced time series statistics were applied to find significant trends in environmental metaproteomic data for the first time. This allowed for the non-biased isolation of changepoints in underlying peptide abundance states for use in revealing potential biomarkers. As the peptide and protein amino acid sequences are informative and specific to both taxonomic and functional annotations, I was able to reveal who was doing what through time. Taxonomic composition is often used in microbiome studies to analyze changes in community composition (Sunagawa *et al.,* 2015; Shin *et al.,* 2018); this approach, however, does not elucidate what the microbes present are actually doing. Functionality, as categorized through gene ontology (GO) terms, can be more difficult to isolate and interpret due to the innate complexity of GO term structure; however, understanding the active metabolic responses to conditions is crucial to scientific advancement. In this work, a tool called MetaGOmics (Riffle *et al.,* 2017) is utilized and improved to assign annotations of both taxa and function across a microbiome, facilitating the analysis of overarching trends. Chapters 3 and 4 aim to draw statistically robust biological conclusions by using both taxonomy and function to address a variety of hypothesis-driven and environmentally relevant questions.

Though microbiomes often exhibit a high level of diversity, common metabolic processes are shared across many bacteria, making them relevant to a variety of studies regardless of the biological question being addressed. For example, carbon, nitrogen, and fatty acid metabolism are all critical for life as we know it (Morris *et al.,* 2010). Recent literature has uncovered evidence that these processes can occur with regularity even in bacteria, with metabolic cycles ranging from seasonal to daily (Lindh *et al.,* 2015; Arandia-Gorostidi *et al.,* 2022). This regularity suggests the presence of metabolic scheduling, with distinct pathways prioritized at specific times of day. This general idea of metabolic scheduling is relatively new and fundamentally changes our view of how organisms as simple as bacteria can operate within complex communities, and perhaps even how and why certain mixtures maintain the balance they do. Choosing a specific time of day to sample could have drastic implications for the biological pathways expressed at the time of collection and should be further examined. Additionally, most microorganisms have specific adaptations down to the strain level that can offer them a competitive advantage against neighboring bacteria in the midst of environmental change (Scheuerl *et al.,* 2020; Lieberman 2022). Teasing apart the influence of metabolic scheduling and competitive strategies in the context of complex biological processes is difficult but could provide diagnostic or forecasting capabilities in a range of fields.

1.4 ORGANIZATION OF THIS DISSERTATION

While the experiments presented here are diverse, there are overarching themes worth noting. One major goal of the work was the development of biomarkers to aid in the translation from bench to real world applications. The development of peptide-based biomarkers is gaining momentum in medical research, with a range of potential molecular or antigen rapid tests in development (Pandey, Malviya, and Chottova Dvorakova 2021). Another theme of this research was the characterization of adaptation strategies in the midst of challenging environments, both in the context of a single organism and a diverse microbiome. Finally, the backbone of this research was the use of mass spectrometry-based proteomics techniques to isolate functional strategies of bacteria when responding to change through a peptide-centric lens.

In Chapter 2, the model psychrophilic bacterium *Colwellia psychrerythraea* 34H (Cp34H) was used to analyze the long-term effects of temperature, salinity, and nutrients on viability and survival. Specific adaptation strategies were isolated by examining the proteome after incubations of four months across two temperatures and two nutrient conditions. Potential biomarkers for life in subzero environments were additionally isolated by searching for enriched combinations of short (2-5) amino acid chains, small enough for life detection strategies proposed on future missions. In Chapter 3, a metaproteomic analyses of a complete time series that captured an ocean microbiome preceding a harmful algal bloom of the non-toxic phytoplankton *Chaetoceros socialis* was performed. Pre-bloom dynamics were examined to isolate peptide biomarkers through the development of a novel, statistically robust pipeline for discovery. After the identification of peptide biomarkers that predict the onset of bloom formation, the peptides were tested in a targeted mass spectrometry assay on another in situ time-series to determine their validity in predicting blooms. In Chapter 4, peptides from the same pre-bloom time series in Chapter 3 were further characterized in the context of the entire 22 day time series collected during the 2021 field season (Nunn 2024) in which both pre-bloom and post-bloom dynamics of Chaetoceros influenced the metabolic scheduling of the surrounding microbiome. GO terms were examined to identify unique patterns in metabolism across various dominant bacterial classes of the microbiome; additionally, GO terms were examined independent of taxonomy to reveal functions that reflect global changes in the metabolic strategy of the microbiome between blooms. Through these chapters a variety of mass spectrometry techniques were used and optimized for analyzing low-biomass microbial communities, including data-dependent acquisition, data-independent acquisition, and parallel reaction monitoring, to better understand how microbiomes respond to their environment and how they may in turn also influence the environment in tangible, applicable, and relevant ways.

Chapter 2: SUBZERO, SALINE INCUBATIONS OF COLWELLIA PSYCHRERYTHRAEA REVEAL STRATEGIES AND BIOMARKERS FOR SUSTAINED LIFE IN EXTREME ICY ENVIRONMENTS

Chapter 2 is adapted with minimal modification from:

Mudge, Miranda C., Brook L. Nunn, Erin Firth, Marcela Ewert, Kianna Hales, William E. Fondrie, William S. Noble, Jonathan Toner, Bonnie Light, and Karen A. Junge. 2021. "Subzero, Saline Incubations of Colwellia Psychrerythraea Reveal Strategies and Biomarkers for Sustained Life in Extreme Icy Environments." *Environmental Microbiology* 23 (7): 3840–66. https://doi.org/10.1111/1462-2920.15485.

M. C. Mudge contribution: Proteomic mass spectrometry, including sample preparations and all proteomic data analyses, and biomarker identifications.

2.1 ABSTRACT

Colwellia psychrerythraea is a marine psychrophilic bacterium known for its remarkable ability to maintain activity during long-term exposure to extreme subzero temperatures and correspondingly high salinities in sea ice. These microorganisms must have adaptations to both high salinity and low temperature to survive, be metabolically active, or grow in the ice. Here, we report on an experimental design that allowed us to monitor culturability, cell abundance, activity, and proteomic signatures of *Colwellia psychrerythraea* strain 34H (*Cp*34H) in subzero brines and supercooled sea water through long-term incubations under eight conditions with varying subzero temperatures, salinities, and nutrient additions. Shotgun proteomics found novel metabolic strategies used to maintain culturability in response to each independent experimental variable, particularly in pathways regulating carbon, nitrogen, and fatty acid metabolism. Statistical analysis of abundances of proteins uniquely identified in isolated conditions provide metabolism-specific

protein biosignatures indicative of growth or survival in either increased salinity, decreased temperature, or nutrient limitation. Additionally, to aid in the search for extant life on other icy worlds, analysis of detected short peptides in -10°C incubations after four months identified over 500 potential biosignatures that could indicate the presence of terrestrial-like cold-active or halophilic metabolisms on other icy worlds.

2.2 INTRODUCTION

As seawater freezes, organisms are trapped within brine channels and inclusions. This habitable, liquid fraction of the sea ice environment is spatially heterogeneous and variably interconnected (Golden *et al*., 2007; Ewert and Deming, 2014), with decreasing temperatures resulting in reduced brine volume and increased brine salinity (Cox and Weeks, 1983). This environment is dynamic and heterogeneous over spatial and temporal scales, including many factors such as bacterial abundances and diversities, extracellular polymeric substances (EPS) (Collins *et al*., 2008), and dissolved organic carbon (DOC) (Underwood *et al.*, 2010). In the Arctic, conditions can be simultaneously dynamic and extreme for considerable periods of time; data recorded within sea ice near Utqiaġvik, Alaska show temperatures remaining below -10°C and calculated brine salinities remaining above 144 ppt (14.4% salt content) for greater than 100 consecutive days (Ewert and Deming, 2014). Organisms living in sea ice thus may be considered polyextremophiles, experiencing both subzero temperatures (psychrotolerance or psychrophily; Junge *et al*., 2019) and high salinities (halotolerance or halophily; Oren, 2006), as well as highly variable nutrient availability.

Polyextremophiles have a particular relevance to astrobiology, and given that most planets and moons with the potential to harbor life have surface temperatures well below freezing with putative high salt content liquids (Hand *et al.*, 2009; Martin and McMinn, 2018), it is reasonable to examine polyextremophilic psychrophiles. Of distinct interest in the search for signs of extraterrestrial life is the Europan chaos terrain (Hand *et al*., 2017, Europa Lander Report). Organisms sampled from this environment would be expected to have passed through habitable subsurface liquid into uninhabitable surface ice, thus it is important to understand the survivability of organisms and biosignatures when transferred to conditions beyond their known growth limits for considerable periods of time, specifically conditions that are colder, more saline, and potentially nutrient limited. Additionally, liquid water, ranging in size from droplets to predicted subsurface oceans, has been detected on various icy worlds throughout the Solar System due to the high concentration of salts within (Carr *et al*., 1998; Hendrix *et al.,* 2018; Lauro *et al*., 2020), making brines on such worlds a suitable target for the search for biomarkers of life. Martin et al. (2020) modeled one such example, Titan's subsurface ocean, to report potential protein compactness, flexibility, and backbone dihedral distributions by comparing proteins found on Earth to simulations in Titan waters, adding weight to our hypothesis described herein that proteins can exist in extreme hypersaline conditions.

The theoretical minimum growth temperatures for psychrophilic organisms (such as those predicted for Antarctic isolates in Bowman *et al*., 1998) will be limited by the equilibrium salinity of the brines, unless the organism is inhabiting supercooled solutions. On Earth, supercooled water exists in clouds, in the surface waters of open leads or ice islands in the Arctic ocean (Untersteiner and Sommerfeld, 1964; Katsaros, 1973), or brines underneath Antarctic ice sealed lakes (Murray *et al.,* 2012); evidence of bacterial activity has been observed in supercooled cloud droplets and Antarctic brines, suggesting such liquids provide suitable habitats for life (Sattler, 2001; Murray *et al*., 2012; Creamean, 2019). Adaptations that limit sea ice freezing, favor supercooling, or allow the organism to locate itself in conditions of favorable salinity will be advantageous to microorganisms in sea ice. Survival at subzero temperatures in equilibrium brines in comparison to supercooled seawater at the same temperature has not been examined to date. It is furthermore unclear whether low-temperature or high-salinity adaptations, or what combination of both, grant an organism the ability to grow under the corresponding equilibrium brine salinity conditions in subzero environments. Thus, in order to characterize the responses of organisms to polyextremophilic environments, we used multivariable laboratory experiments with both supercooled seawater and equilibrium brines that allowed us to investigate the individual and interactive effects of salinity and temperature and at the same time explore microbial activities in supercooled seawater.

Here we specifically examined Colwellia psychrerythraea 34H (Cp34H), a marine psychrophilic gammaproteobacterium from subzero Arctic marine sediments (Huston *et al*., 2000) that is also found in sea ice (Methe *et al*., 2005; Boetius *et al*., 2015). *Cp*34H is considered the model marine psychrophile for cold oceans (Czajkaa *et al.*, 2018); it has been shown to exhibit protein synthesis at very low subzero temperatures in artificial sea ice (Junge *et al.*, 2006), and to possess many adaptations typical of psychrophilic sea ice bacteria (Methe *et al*., 2005; Ewert and Deming, 2011; Collins and Deming, 2013; Nunn *et al*., 2015; Firth *et al.,* 2016; Showalter and Deming, 2017). The current temperature growth range of *Cp*34H in nutrient rich marine media is -12 to 19°C (Wells and Deming, 2006a,b), corresponding to an equilibrium brine salinity maximum of approximately 181 ppt (Cox and Weeks, 1986; Eicken and Salganek, 2009), well outside the observed salinity growth range of *Cp*34H at its optimal growth temperature of 8°C in nutrient rich marine medium (15 to 70 ppt; Huston, 2003). The equilibrium brine conditions studied here were chosen based on natural conditions encountered by organisms within sea ice based on known eutectic point values. Below -5°C, conductive fluid flow in sea ice ceases to exist (Golden *et al*., 2007) and brine pockets containing a high salt content are isolated. By comparing results from cells in-brine with cells in-supercooled seawater, we can specifically elucidate the impact that salts alone have on microbial activities at a given temperature and without ice formation. This controlled analysis removes the factor of the ice itself, as examined in previous in-ice experimental studies (e.g., Junge *et al*., 2006; Nunn *et al*., 2015) or field work (e.g., Junge *et al*., 2004), where it is impossible to distinguish between effects of temperature and salinity - since in the natural habitat of sea ice both are always intricately linked. Although past work has characterized the effects of temperature and time on psychrophilic activities and proteomes in ice (Nunn *et al*., 2015), to our knowledge the functional responses unique to temperature, salinity, and nutrients in *Cp*34H have yet to be characterized.

In this study, *Cp*34H was incubated for four months at two subzero temperatures within its growth range, in supercooled seawater or in equilibrium brine, with or without nutrient additions. By tracking bacterial cell counts, metabolic activity, culturability, and proteomic responses in tandem across the eight conditions, we investigated proteome shifts to determine potential protein indicators for identifying present or extant life in each independent condition, whether on Earth or other icy worlds. Individual and interactive effects of temperature, salinity, and nutrient availability were disentangled, providing the molecular details of cellular culturability maintenance after extended periods of time. We furthermore analyzed the detected peptides from Cp34H held at -10°C for four months for short polypeptide enrichments to determine if specific amino acids were enriched relative to the entire predicted proteome, representing abundant temperature- or salinity-dependent biomarkers with high detectability and survivability that are compatible with life on Earth found in environments analogous to those proposed off-Earth. As proteins and amino acids survive in the environment for extended periods of time and have been detected in extraterrestrial bodies (e.g., McKay *et al*., 1996; refs within Cobb and Pudritz, 2014; McGeoch *et al.*, 2020), we propose and provide a list of short three or four amino acid long polypeptides (referred to as 3- and 4-mers) as candidate biomarkers for future exobiology missions — especially relevant given that most worlds in the solar system are at temperatures well below the freezing point of water with high salinity and limited water activity (Carr *et al*., 1998; Khurana *et al*., 1998; Hand *et al*., 2009; Hendrix *et al*., 2018; Parro *et al*., 2018).

2.3 METHODS

Supercooled sea water and equilibrium brines We prepared three sterile saline solutions (artificial seawater (ASW) at 6°C, equilibrium brine at −5°C and −10°C). The ionic composition of the solutions was chosen to reflect the freezing-equilibrium concentrations of seawater at −1°, −5° and −10°C resulting in salinities of 35, 84, and 140 ppt with respective salt content of 3.5, 8.4, and 14% (Table 2.1, Marion *et al.,* 1999). Recipes for saline solutions were calculated to achieve the desired molal concentrations for each temperature (Table 2.2). Solutions were prepared with 1000 g of milli-Q water previously amended with 1.3 g/L of the buffer TAPSO (3-[N- tris(hydroxymethyl) methylamino]-2-hydroxypropane-sulfonic acid) and titrated to a pH of 7.6. Individual marine salts listed in Table 2.2 were added to the buffered milli-Q water at room temperature. Care was taken to prevent mineral precipitation at room temperature by mixing $CaCl_2$ separately in half of the water volume and slowly titrating the $CaCl_2$ solution into a solution containing the remaining salts. After preparation, saline solutions were filter-sterilized using 0.2 µm cap-filters. Equilibrium ionic concentrations in Table 2.1 were converted to salt concentrations (%) following Marion and Kargel (2008). Water activity for brines was calculated based on the equilibrium constant of ice, a function of temperature (Spencer *et al.,* 1990).

Table 2.1 Molal concentrations of sea salt ions used to prepare equilibrium solutions.

Equilibrium temperature	Na	K	Ca	Mg	Cl	SO_4	% Salt	Ionic strength (mol/kg)	Water activity
−1°C	0.486	0.011	0.010	0.055	0.567	0.029	3.5	0.7	0.990
−5°C	1.228	0.027	0.024	0.1383	1.432	0.074	8.4	1.8	0.953
−10°C	2.172	0.051	0.046	0.263	2.720	0.060	14.0	3.2	0.907

Concentrations were calculated assuming the equilibrium freezing pathway of seawater (Marion *et al.,* 1999). Salt concentration (%) was calculated according to Marion & Kargel (2008). Water activity for Brine was calculated based on the equilibrium constant of ice, a function of temperature (Spencer *et al.,* 1990). Note that the equilibrium solution at −1°C corresponds to artificial seawater (ASW).

Table 2.2. Recipe used in the preparation of experimental saline solutions.

Solution	**KCl**	**$MgCl_2$**	**NaCl**	**$MgSO_4$**	**$CaCl_2$**
ASW	0.8	2.4	28.4	3.5	1.1
-5°C Brine	2.0	6.1	71.8	8.9	2.7
-10°C Brine	3.8	19.3	126.9	7.3	5.1

Salts (g) added to 1000 g of water.

Bacterial strain and cellular preparations Cell preparations were obtained from cultures of *Colwellia psychrerythraea* strain 34H (ATCC No. BAA-681; GenBank Accession No. AF396670). *Cp*34H was cultured from frozen glycerol stocks, and then grown to early stationary phase in half-strength Marine Broth 2216 (SIGMA, Junge *et al.,* 2006)), rocking, at 0°C to a final cell density of 1.04×10^9 cells mL^{-1} (OD_{600} = 0.44).

The culture was split into three equal parts and centrifuged under sterile conditions at 2,800 x g for 20 min at 4°C (Nunn *et al.,* 2015). After discarding the supernatant, cells were resuspended in either one of the three pre-chilled sterile saline solutions (see above), centrifuged again at 2,800 x g for 10 min at 4°C and resuspended in their corresponding pre-chilled saline solution for a final OD_{600} of 0.26 (ASW), 0.29 (−5°C brine) and 0.28 (−10°C brine). Triplicate 100 µL aliquots of cell resuspension were fixed in 2% formaldehyde and stored at 4°C for the determination of bacterial abundance through epifluorescence microscopy (see Culturability and cell abundance (cell counts)); another set of triplicate 20 µL aliquots was used to determine culturability through the most probable number (MPN) procedure as described below. The resulting initial cell concentrations were 3.12×10^8, 3.39×10^8 and 2.75×10^8 cells mL^{-1} for ASW, −5°C and −10°C brine, respectively. The initial MPN of culturable cells were 2.15×10^8, 3.75×10^8, and 2.15×10^8 cells mL^{-1} for ASW, −5°C and −10°C brine, respectively.

Cells resuspended in saline solutions were either used without further amendments (-nutr superscript) or amended with glucose (1 g/L), yeast extract (14 mg/L) and leucine (4 mg/L) (+nutr superscript). Nutrient concentrations used here are known to support growth in *Cp*34H (Marx *et al.*, 2009; Czajka *et al.,* 2018). Cell resuspensions intended for activity measurements were amended with glucose and yeast extract only, as the radioactive marker already contained leucine. Parallel growth curves of *Cp*34H in ASW with and without nutrients at -1°C for 16 days were also performed to confirm that this minimal amount of nutrients can support growth and culturability under optimal temperature and salinity conditions.

Incubations Prior to incubation at the two temperatures, cell resuspensions were distributed into1.5 mL microcentrifuge tubes (Fisherbrand™ Premium) to be used for measurements of metabolic activity (500 µL cell resuspensions for protein synthesis through [3H]-leucine incorporation), measurements of cell abundance and culturability (600 µL cell resuspension, shared) and proteomics (600 µL cell resuspension). To determine protein synthesis activity, triplicate samples of both live incubations and killed control cells were retrieved at 2, 12, 24 hours and one month. Metabolic activity incubations were completed at one month since previous studies had shown that subzero activity rates based on increases of leucine incorporations over time until maximum incorporation amounts are reached can usually be calculated during initial 24 hour incubation periods, thus allowing for the determination of activity rates for the different treatments (Junge *et al.*, 2006, Nunn *et al.*, 2015). Samples for cell counts and cell culturability were retrieved after initial set-up and after one (n = 3) and four months (n = 4). Samples for proteomic analysis were harvested after four months (n = 4).

Microcentrifuge tubes were placed in plastic racks, wrapped in bubble wrap casings to reduce temperature fluctuations, and transferred to −5°C and −10°C incubators. Fluctuations in incubator temperature were monitored prior to the start of the experiment, and the minimum temperature registered in a few days period was either −5°C or −10°C. Based on measured cooling rates, samples reached their target incubation temperature within one hour.

To generate a reference proteomic dataset (control) from which all experimental proteomic datasets were compared, *Cp*34H was grown at -1°C and harvested at mid-exponential growth. Specifically, three cultures of *Cp*34H were grown in 20 mL of ½ x MB2216 (BD Difco) at -1°C, stirring, to a final cell density of $4.82 \times 10^8 \pm 3.18 \times 10^7$ cells mL^{-1}. Initial inocula were derived from the same frozen glycerol stock used for all other cell resuspensions. From each flask, one 1 mL aliquot was obtained and killed with an immediate addition (i.e., within seconds) of 200 µL of pre-chilled 50% trichloroacetic acid (TCA). Any possible effect of opening the tube (and allowing more oxygen to enter for a few seconds) is considered negligible. Another 1 mL aliquot was fixed with 2% formaldehyde and stored at 4°C for analysis of microbial abundance. TCA-killed samples were centrifuged at 5,000 x g for 20 min at 4°C, then washed with cold, sterile ASW and centrifuged again at 18,000 x g for 15 min at 4°C. Pellets were stored at -70°C until proteomic sample preparation (as described below).

Culturability and cell abundance (cell counts) At each time point, each microcentrifuge tube was subsampled for culturability, cell abundance and proteomic analysis. Culturability was determined by aliquoting 20 µL into triplicate wells of a pre-chilled 96-well microplate containing 180 µL of ½ x MB 2216. Ten serial 1:10 dilutions were performed on each triplicate sample. Microplates were incubated at 4°C until wells in the microplate showed visible turbidity. The number of wells displaying growth was converted into three-digit codes used to calculate the most probable number of culturable cells (MPN) according to standard tables from the FDA Bacteriological Analytical Manual (Blodget, 2011). To determine cell abundances, 100 µL aliquots were fixed in 2% formaldehyde and kept at 4°C until epifluorescence microscopy enumeration. Cells were stained with the DNA-fluorescent stain DAPI (4′,6′-diamidino-2-phenylindole 2HC) and microscopic counts were performed in triplicate as in Junge *et al.* (2004).

Metabolic activity: [3H]-leucine incorporation Bacterial metabolic activity was measured using [3H]-leucine incorporation assays previously developed for the analysis of bacterial protein synthesis activity of *Cp*34H in saline ice (Junge *et al.,* 2006; Nunn *et al.*, 2015). Briefly, incubations for the metabolic activity assays contained 100 µL of the radioactive tracer ([3H]-leucine (10 Ci mL^{-1} ; diluted 1:100 from stock of [3H]-leucine [ICN Biomedicals, Catalog No. 20036E; 40–60 Ci mL^{-1} in sterile 2:98 ethanol–water mixture])). For controls, 100 µL of 50% chilled trichloroacetic acid (TCA) were added to each microtube before tracer addition (as in Junge *et al.,* 2006 and Nunn *et al*., 2015). Triplicate controls were completed at each temperature and time point. Methods described by Junge and colleagues (2006, 2015) were primarily followed, with the exception of one modification where two ethanol washes were performed during the processing stage.

Proteomic sample preparations After four months of incubation, a 440 µL aliquot from each 600 µL cell suspension sample was subsampled, killed with 88 µL of cold 50% TCA, and cell debris and proteins were pelleted at 18,000 x g for 15 min at 4°C. After discarding supernatant, pellets were frozen at -70°C until proteomic sample preparation. All TCA-precipitated cell pellets for proteomic analyses were mechanically lysed (Branson 250 Sonifier; 20kHz, 10 x 10 s on ice) and then proteins were solubilized, quantified, and digested following the detailed methods outlined in Nunn *et al.* (Nunn *et al.*, 2015). Briefly, 100 µg aliquots of protein were reduced (tris(2-carboxyethyl)phosphine (TCEP), alkylated (iodoacetamide) and then digested with Promega modified trypsin (4 hr 37°C). Peptides were desalted using C18 miniprep centrifugal spin columns following manufacturer's instructions (Nest group MicroSpin columns) and cleaned up peptides were evaporated to dryness and resuspended in 2% acetonitrile (ACN), 0.1% formic acid with final concentration of 0.5 µg protein μL^{-1}.

Proteomic mass spectrometry The mass spectrometry analysis was performed on a QExactive Hybrid Quadrupole Orbitrap with an inline EASY-nLC 1200 (ThermoFisher Scientific). Reverse-phase chromatography was achieved using a Manufactured PicoTip fused silica capillary column (40 cm long, 75 µm i.d.) packed in-house with C18 particles (Dr. Maisch ReproSil-Pur; C18-Aq, 120 Å, 3 µm) fitted with an in-house packed kasil-frit 3 cm long, 100 µm i.d. precolumn (Dr. Maisch ReproSil-Pur; C18-Aq, 120 Å, 3 µm). To maintain a stable back-pressure on the liquid chromatography column, the 40 cm analytical column was held at 40°C. Peptides were eluted using an acidified (formic acid, 0.1% v/v) water-acetonitrile gradient (2–35% acetonitrile in 120 min). The top 20 most intense ions were selected for MS2 acquisition from precursor ion scans of 400–1200 *m/z*. Centroid full MS resolution data was collected at 70,000 with AGC target of $1x10^6$ and centroid MS2 data was collected at resolution of 35,000 with AGC target of $5x10^4$. Dynamic exclusion was set to 10 seconds and +2, +3, +4 ions were selected for MS2 using data dependent acquisition mode (DDA). Sample analyses on the MS were randomized to reduce batch effects and quality control (QC) peptide mixtures were analyzed every fifth injection to monitor chromatography and MS sensitivity. Skyline was used to determine that QC standards did not deviate >10% through all analyses (MacLean *et al.*, 2010). The mass spectrometry data are deposited to the Proteome Xchange Consortium via the PRIDE partner repository with the dataset identifier PXD022428.

Protein data interpretation Peptide identifications from mass spectrometry data were completed using Comet 2018.01 rev. 2 (Eng *et al.*, 2013; 2015) and the Uniprot *Colwellia psychrerythraea* protein database (downloaded 4/2/2019) concatenated with 50 common contaminants and the quality control peptides. Comet parameters included: reverse concatenated sequence database search, trypsin enzyme specificity, cysteine modification of 57 Da and modifications on methionine of 15.999 Da (oxidation). PeptideProphet and ProteinProphet were used to validate peptide spectral matches (PSMs) and determine thresholds for a false discovery rate < 0.01 (proteins accepted if ≥ 2 peptides identified, and p > 0.95 on ProteinProphet and p > 0.99 on PeptideProphet (Nesvizhskii *et al.*, 2003). Abacus (in the QPROT software suite) was used to generate normalized spectral abundance factors (NSAF), the number of unique peptides identified for each protein across all experiments, and peptide spectral match (PSM) counts for all confidently identified proteins (Fermin *et al.*, 2011; Choi *et al.*, 2015). NSAF corrects for the effect of protein length on final spectral counts, as larger proteins inherently contribute more peptides. NSAF is calculated as the number of spectral counts for a protein (SpecCnt), divided by the protein length (L) (i.e., the number of amino acids), divided by the sum of SpecCnt/L for all proteins identified in a mass spectrometry experiment. NSAF values were further analyzed with QPROT to determine pairwise log2 fold changes between treatments and were plotted using ggplot2 to generate volcano plots (Choi *et al.*, 2015; Wickham, 2016). All data presented to be significantly changing passed $\log_2$ fold-change ≥ |0.5| and Z-statistic score ≥ |2.0| (equivalent to p-value < 0.05; two tailed) (Choi *et al.*, 2015). Nonparametric multidimensional scaling analysis (NMDS) was completed on NSAF values for each bioreplicate with ≥ 2 unique peptides identified per treatment using the package “vegan” with R studio version 1.2.1335 (Oksanen *et al*., 2013; Rstudio Team, 2020). Hierarchical clustering analysis was completed using the package “pheatmap” on peptides meeting more stringent significance thresholds of $\log_2$ fold-change ≥ |2| and Z-statistic score ≥ |2.0|. Upset plots and venn diagrams were completed on datasets where ≥ 2 unique peptides were detected across all experiments, and ≥ 2 spectral counts per bioreplicate (Chen, 2018; Conway and Gehlenborg, 2019). A defined list of 21 functional annotation terms for categorizing proteins was selected using EggNOG-mapper version 5.0, a database for defining proteins using orthologous relationships and gene evolutionary histories (Huerta-Cepas *et al.*, 2019). All identified proteins were also annotated with Kyoto Encyclopedia of Genes and Genomes (KEGG) using KOALA (Kanehisa *et al.*, 2015), assigning K numbers to identified metabolic enzymes. Proteins identified to be significantly changing were mapped onto preexisting metabolic maps using KEGG Mapper - Search&Color Pathway tool (https://www.genome.jp/kegg/tool/map_pathway2.html).

Enrichment analysis of salinity and temperature-specific short polypeptide

Enrichment analyses of the amino acid sequences retained after four months at -10°C were completed to determine if there were short peptide sequences (3-4 amino acids) that were significantly enriched relative to the short peptide sequences from all of the proteins detected across these experiments. Peptide-spectrum matches (PSMs) were post-processed using Percolator (Käll *et al.*, 2007) to detect peptides in each run at a false discovery rate (FDR) threshold of 1%. NSAF values from all MS experiments on the four treatment conditions were analyzed using two-factor analysis of variance (ANOVA) with 1% FDR control via the Benjamini-Hochberg (BH) procedure to detect proteins associated with changes in temperature and salinity. This analysis provided sets of proteins associated with temperature (-5°C vs. -10°C), salinity (ASW vs. equilibrium brines), and the interaction of the two terms (all conditions vs. -10°C brine treatments). From the proteins identified in the ANOVA, $\log_2$ fold change was computed followed by a one-tailed t-test (with BH FDR control at 1%) to determine significantly changing proteins between different treatments ($\log_2$ fold change ≥ 0.5). Proteins identified to be significantly different in response to high salinity, -10°C temperatures, or the combined effects in the -10°C brines were then digested *in silico* to 3-4 amino acid long k-mers, and statistically evaluated relative to the background lists of 3-mers and 4-mers detected across all experiments to identify condition-specific enriched short polypeptides. A chi-squared test was performed on the k-mer counts in each condition, with BH multiple testing correction, providing a q-value and $\log_2$ fold change value for each k-mer and allowing identification of statistically different frequencies of the k-mers observed in the different treatments. A second chi-squared test was performed to determine which, if any, amino acid compositions of the identified 3-mers and 4-mers were significantly increased or decreased relative to the amino acid distribution of all detected peptide k-mers. Heatmaps were generated to visualize the distribution of amino acids in the 3-mers and 4-mers, irrespective of the location in the sequence, where rows representing amino acid distributions were clustered using Euclidean distances. Principal component analysis was completed using Python using scikit-learn package v.0.21.3, and all code and scripts are provided in the Github repository github.com/noble-lab/2020_cp34h.

2.4 RESULTS

2.4.1 CULTURABILITY, ABUNDANCE AND ACTIVITY ANALYZED AFTER ONE MONTH INCUBATION

Incubations of *Cp*34H (Fig. 2.1) in artificial seawater (ASW; Table 2.1) with nutrient media at optimal salinity and warmer temperature (-1°C) demonstrated that *Cp*34H can grow in this media as indicated by the increase in culturability (quantified as the number of culturable cells determined through MPN technique; see experimental procedures below) over time up to a maximum of 1.45×10^{8} cells mL^{-1} (Fig. 2.2). When incubated at −5°C in supercooled ASW, *Cp*34H maintained a high number of culturable cells (~1×10^{8} cells mL^{-1}) over four months in experiments where nutrients were present (Fig. 2.2a). Exposure to lower temperature (*i.e.,* -10°C, Fig. 2.2b, d) resulted in a complete elimination of culturable cells; higher salinity (*i.e.,* 8.4% [for the nutrient-depleted treatment only], Fig. 2.2c) resulted in a sharp decrease in the number of culturable cells (down to an average of 5.38×10^{1} +/- 1.08×10^{2} cells mL^{-1}). With the exception of most of the -10°C treatments, culturability in brine samples increased in the presence of nutrients. Cells remained culturable after one month of incubation in -10°C brine treatments, with total cell abundances as high or higher than in -5°C incubations (Fig. 2.3). Total cell abundance in -5°C brine treatments after four months showed a significant decrease ($p < 0.01$; Fig. 2.3).

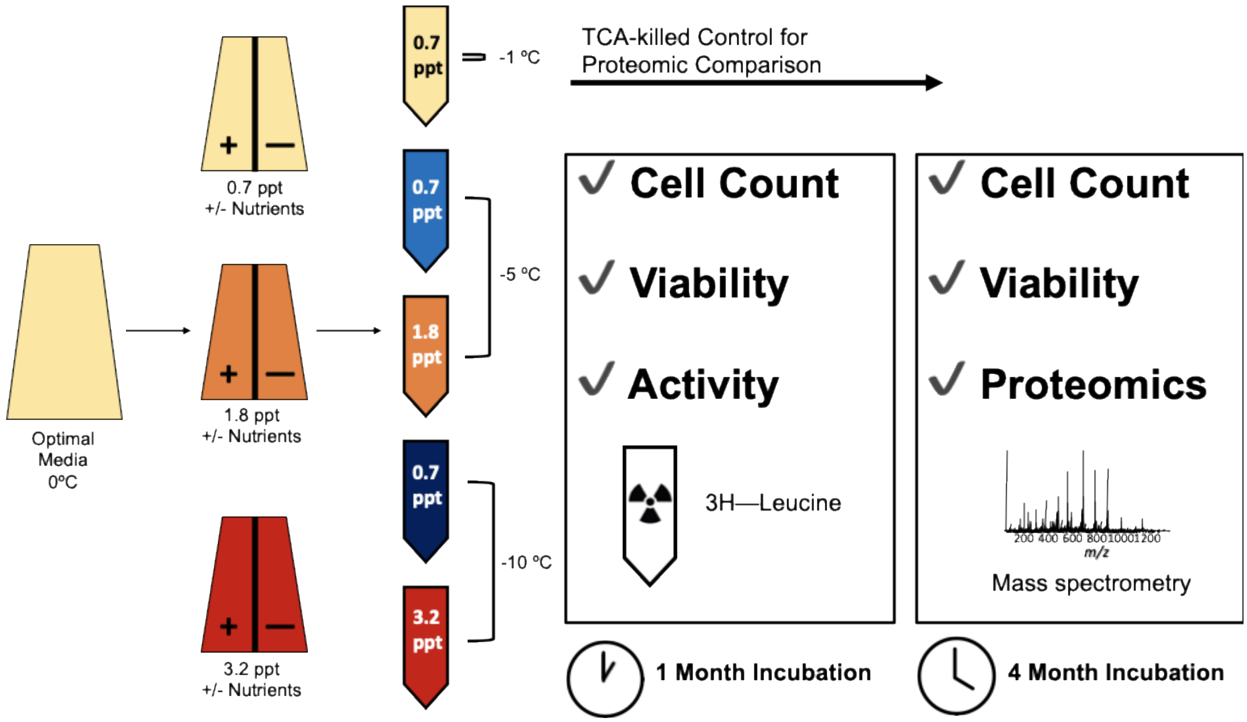


Figure 2.1. Experimental design for long-term incubations of *Cp*34H in extreme temperatures and salinities. Cells were grown in optimal media at 0°C until desired concentration was reached. Cells were then aliquoted into three separate media with salinity concentrations calculated in Table 2.1-2. Media for each salinity was divided into 2 flasks, with or without additional nutrients. Cells were then aliquoted in triplicate to eppendorf tubes and incubated at -5°C or -10°C for one and four months. Samples were harvested and cell count and culturability measures were performed. Additionally, samples incubated one month were measured for activity using [3H]-leu and samples incubated for four months were analyzed via mass spectroscopy.

[3H]-leucine incorporation into *Cp*34H was tracked for the first month of the incubation and used to approximate the rate of oxygen depletion within the cultures, not accounting for possible oxygen flux into or out of the sealed 1.5 mL eppendorf tubes as the microcentrifuge tubes were likely not gas tight (Table 2.3). Generally, *Cp*34H incorporated [3H]-leucine into newly made proteins at both temperatures, usually peaking at 1 hr (Table 2.3). Incorporation rates ranged from 3.63×10^{-7} (-10°C, brine^{-nutr}) to 2.37×10^{-2} (-5°C, ASW^{+nutr}) g bacterial carbon produced per g

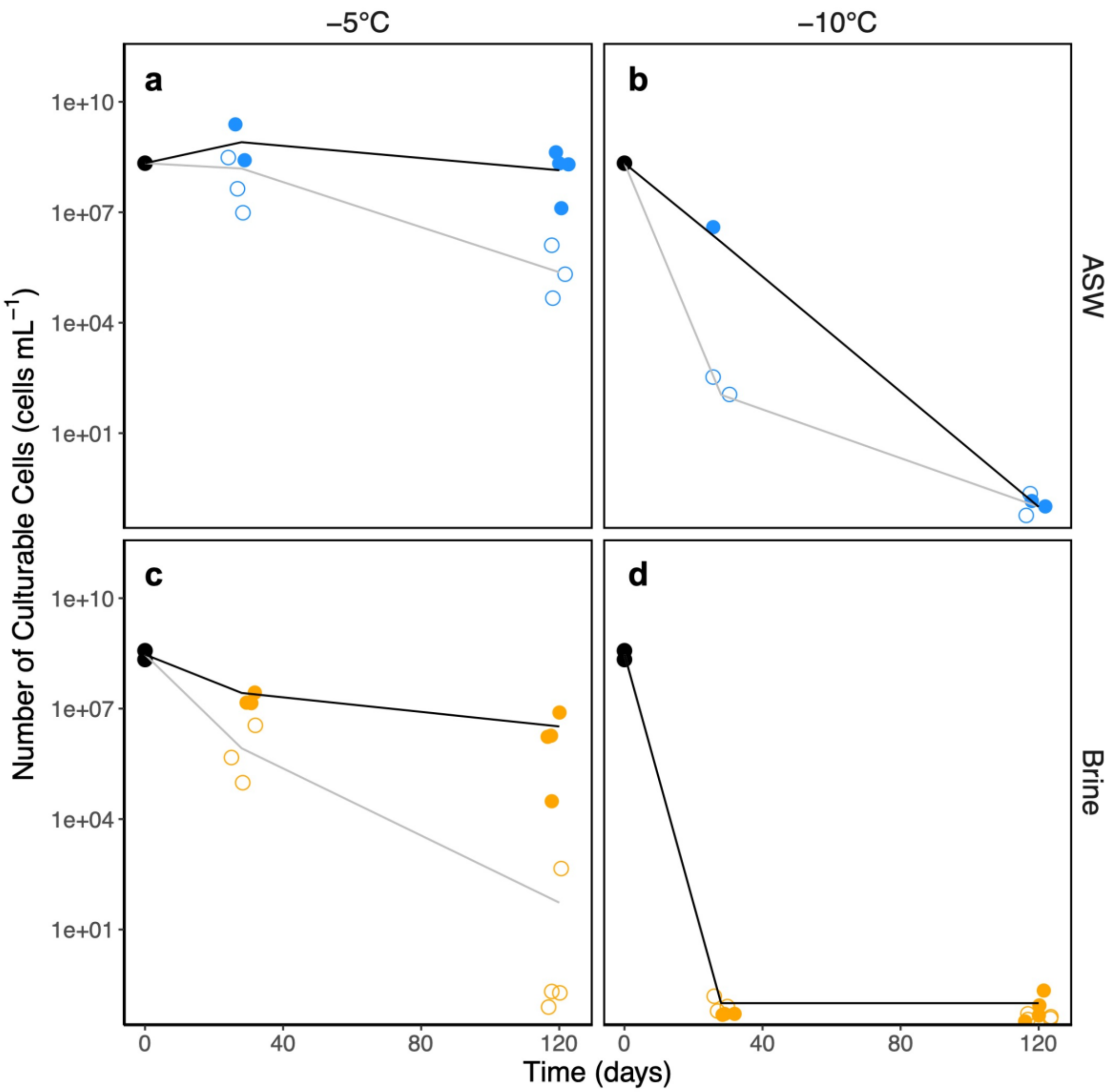


Figure 2.2. Most probable number (MPN) of culturable cells (logarithmic scale) after incubation for one to four months at −5°C or −10°C. Open symbols indicate cells incubated in saline solution without nutrients (either ASW: blue, −5°C brine or −10°C brine: orange); filled symbols indicate cells incubated in saline solutions amended with nutrients. Black filled circle indicates initial starting condition cell cultures. Lines connect average values at each time point ($^{+nuts}$: black, $^{-nuts}$: gray). Note that the salinity of brines is different for each temperature (see Table 2.1).

bacteria carbon per hour, with incorporation rates in ASW two to five orders of magnitude higher than those from corresponding brine treatments (Table 2.3). At one month, brine^{+nutr} at -5°C doubled incorporation of [3H]-leucine; cells incubated in brine at -10°C showed no measurable levels of activity, and all non-brine treatments showed no major change in [3H]-leucine incorporation (Fig. 2.4).

Table 2.3 Rates and Metabolic States of Cp34H Across Experimental Treatments.

Temp. (°C)	**Condition**	**Time** (hour)	**Max Rate*** ($gC(gC^{-1}h^{-1})$)	**Oxygen status at 4 months†**	**Metabolic State****
-5	ASW^{-nutr}	2	$3.64x10^{-4}$	Depleted	Maintenance
-5	ASW^{+nutr}	1	$2.37x10^{-2}$	Depleted	Growth
-5	$brine^{-nutr}$	1	$7.51x10^{-6}$	Limited	Maintenance
-5	$brine^{+nutr}$	1	$4.71x10^{-4}$	Depleted	Maintenance
-10	ASW^{-nutr}	1	$6.51x10^{-4}$	Depleted	Maintenance
-10	ASW^{+nutr}	1	$1.59x10^{-2}$	Depleted	Growth
-10	$brine^{-nutr}$	1	$3.63x10^{-7}$	Limited	Survival
-10	$brine^{+nutr}$	24	$1.47x10^{-4}$	Unknown	Maintenance

Maximum [3H]-leucine incorporation rates were calculated from [3H]-leucine uptake and scaled to bacterial abundance according to the methods and values used in Simon and Azam (1989), with an isotope dilution of 126 for nutrient samples and assuming 65 fg C bacterium^{-1} following Junge et al. (2006). Metabolic states were derived from the findings of Price and Sowers (2004), assuming a value of 1E-3 for µg for both temperatures listed. †Oxygen depletion threshold calculated as the number of hours to use all oxygen in 600 µL solution at 20°C.

2.4.2 PROTEOMES ANALYZED AFTER FOUR MONTH INCUBATION

Using mass spectrometry, we identified a total of 2362 proteins (FDR < 0.01) from a combined analysis of all eight modified temperature, salinity and nutrient treated cell states (47 total MS experiments). Non-metric multidimensional scaling analysis (NMDS) of the relative abundance (normalized spectral abundance factors: NSAF) of all proteins identified in these experimental treatments revealed that temperature, salinity, and nutrient additions drive clustering of MS experimental results on bioreplicates into distinct groups (Fig. 2.5). Importantly, sub-populations with unique protein profiles emerge after four months compared to the Cp34H -1°C controls (Fig. 2.5, black), despite very low levels of metabolic activities and culturabilities in some of the incubations (Table 2.3). The supercooled ASW samples separate along NMDS component 2, with clusters of -5°C separated by the presence/absence of nutrients. $Brine^{+nutr}$ samples exhibit the greatest separation from the other samples along NMDS component 1 (Fig. 2.5). Samples incubated at -10°C cluster together in a subpopulation independent of salinity (Fig. 2.5, red). Immediately adjacent to this cluster are samples from the -5°C $brine^{-nutr}$ condition (Fig. 2.5).

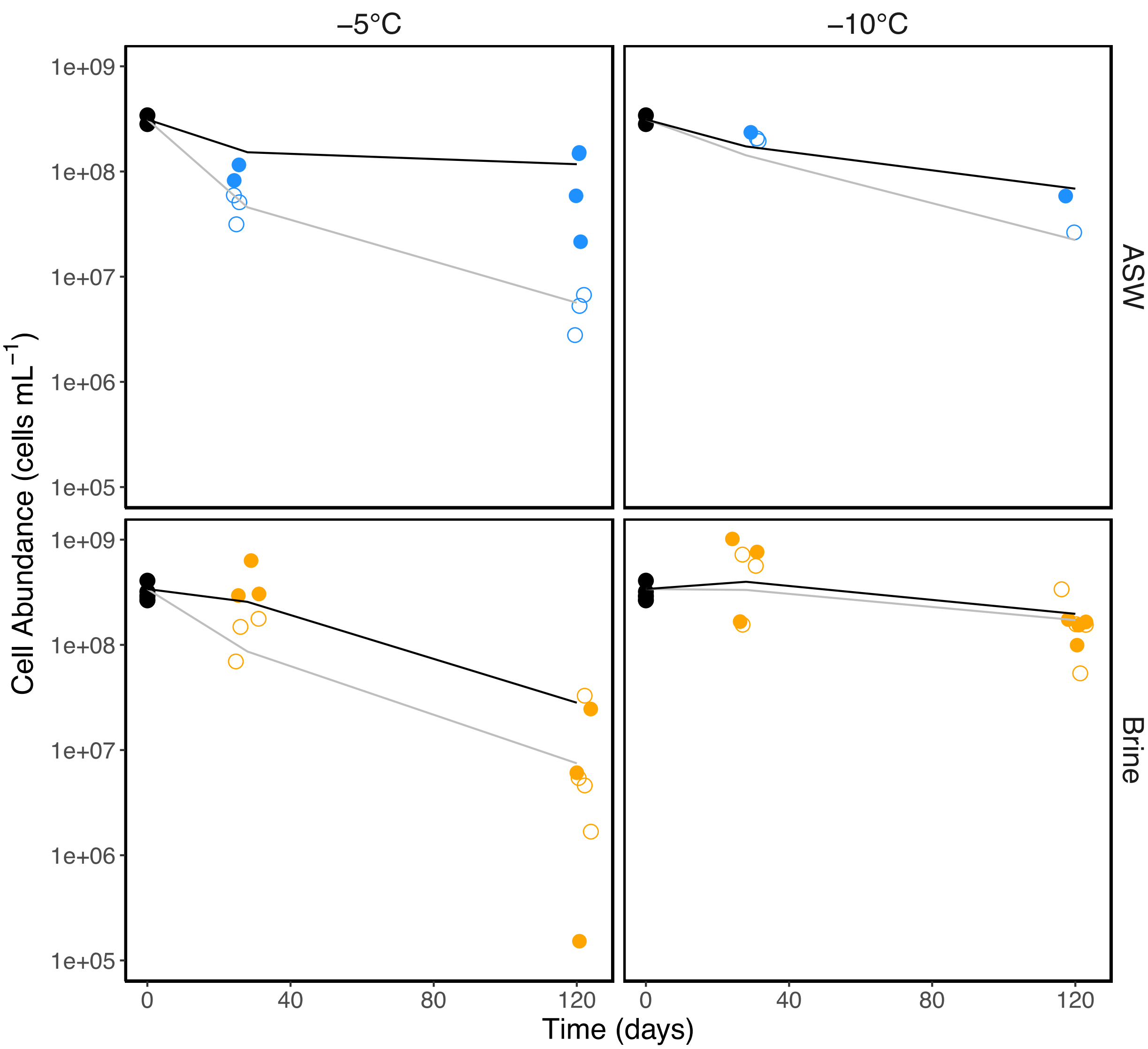


Figure 2.3. Cell abundance (logarithmic scale) after incubation for one and four months at −5°C or −10°C. Cell count was performed on cells harvested after incubation for one or four months. Open symbols indicate cells incubated in saline solution without nutrients (either ASW: blue, −5°C brine or −10°C brine: orange); filled symbols indicate cells incubated in saline solutions amended with nutrients. Lines connect average values at each time point (+nuts: black, -nuts: gray). Note that the salinity of equilibrium brines is different for each temperature (see Table 2.1).

2.4.3 GROWING AND MAINTENANCE ACTIVITY SIGNATURES OF CULTURABLE CELL POPULATIONS AT -5°C

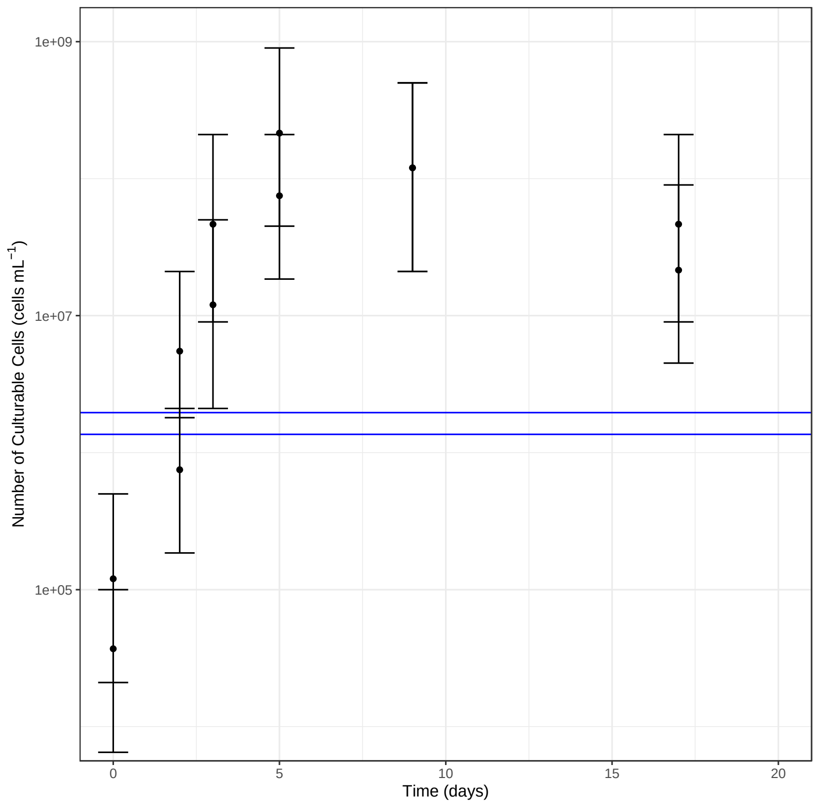


Figure 2.4. [3H]-leucine incorporation during early incubation time points. Metabolic activity (3H-leucine incorporation) after incubation first 24 h and 1 month at −5 °C or −10 °C. Open symbols indicate cells incubated in saline solution without nutrients (either ASW: blue, −5°C brine or −10°C brine: orange); filled symbols indicate cells incubated in saline solutions amended with nutrients. Lines connect average values at each time point ($^{+nuts}$: black, $^{-nuts}$: gray). Dashed line corresponds to the median of controls. Lines connect average values at each time point.

Analysis of the set of proteins shared between conditions identified 874 proteins common among all groups. A total of 68 proteins found within the control were not observed in any of the other treatments at four months (Fig. 2.6b, black), whereas 16 proteins were shared among all treatments at four months (Fig. 2.6b, red), indicative of a time- and/or temperature-dependent response. The second highest number of unique proteins shared were present in the control, ASW^{+nutr}, ASW^{-nutr}, and brine^{-nuts}. Brine^{-nuts}, the condition experiencing the most external stressors, had 45 proteins shared with the control and 12 proteins unique to the condition, the highest abundance of proteins unique to any individual experimental group (Fig. 2.6b).

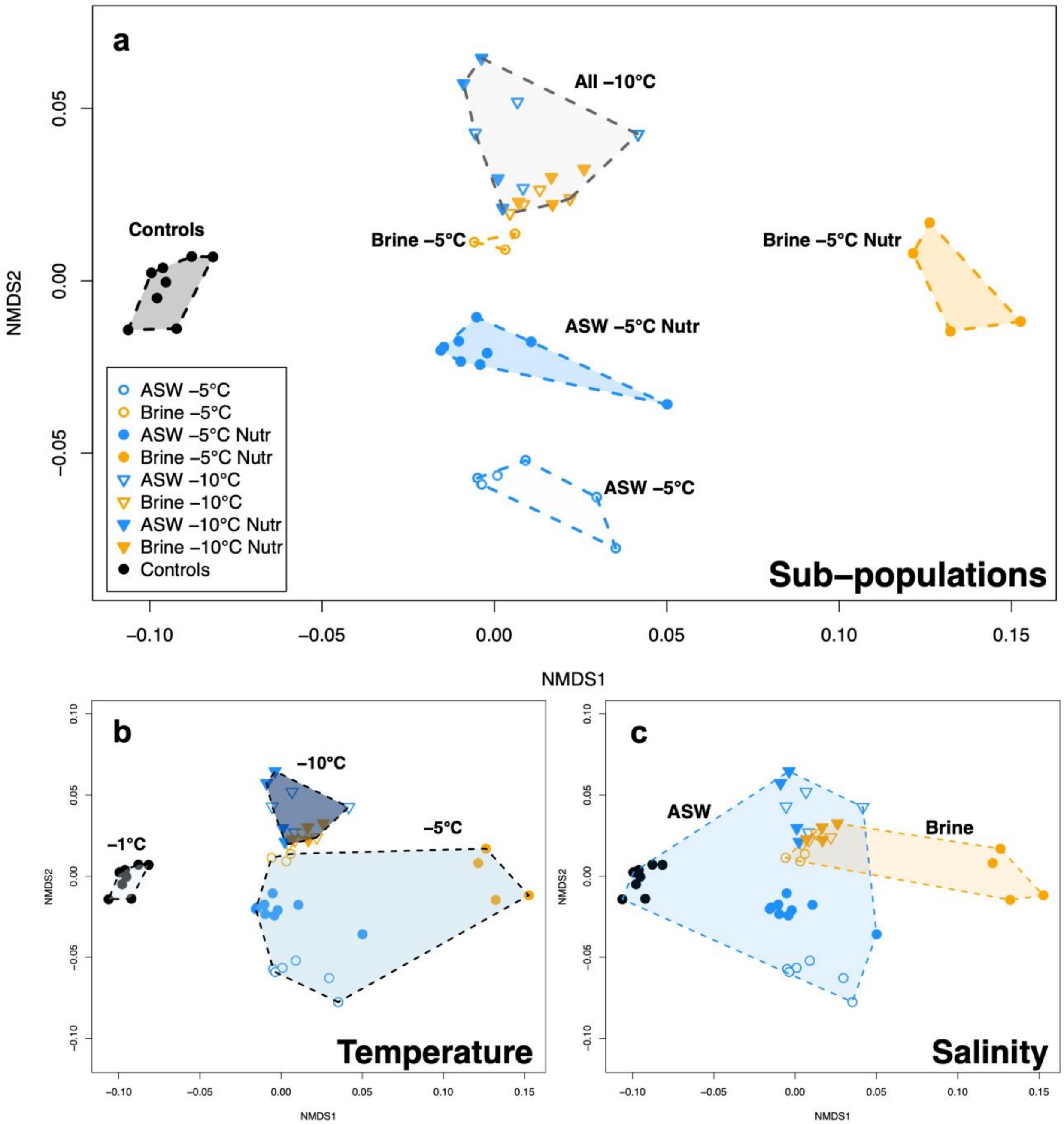


Figure 2.5. Non-metric multidimensional scaling analysis (NMDS) of the NSAF values for all proteins confidently identified in each bioreplicate from each experimental condition harvested at four months. Sample clustering and culturability driven by condition. Open symbols indicate cells incubated in saline solution without nutrients (either ASW: blue, −5°C brine or −10°C brine: orange); filled symbols indicate cells incubated in saline solutions amended with nutrients. Black filled circles represent -1°C starting cell culture conditions (controls). Clustering reveals subgroups based on temperature (dotted line, −5°C: green, −10°C: red), salinity (dashed line, ASW: blue, brine: orange), or salinity-nutrient nexus at −5°C (solid line, color respective to salinity).

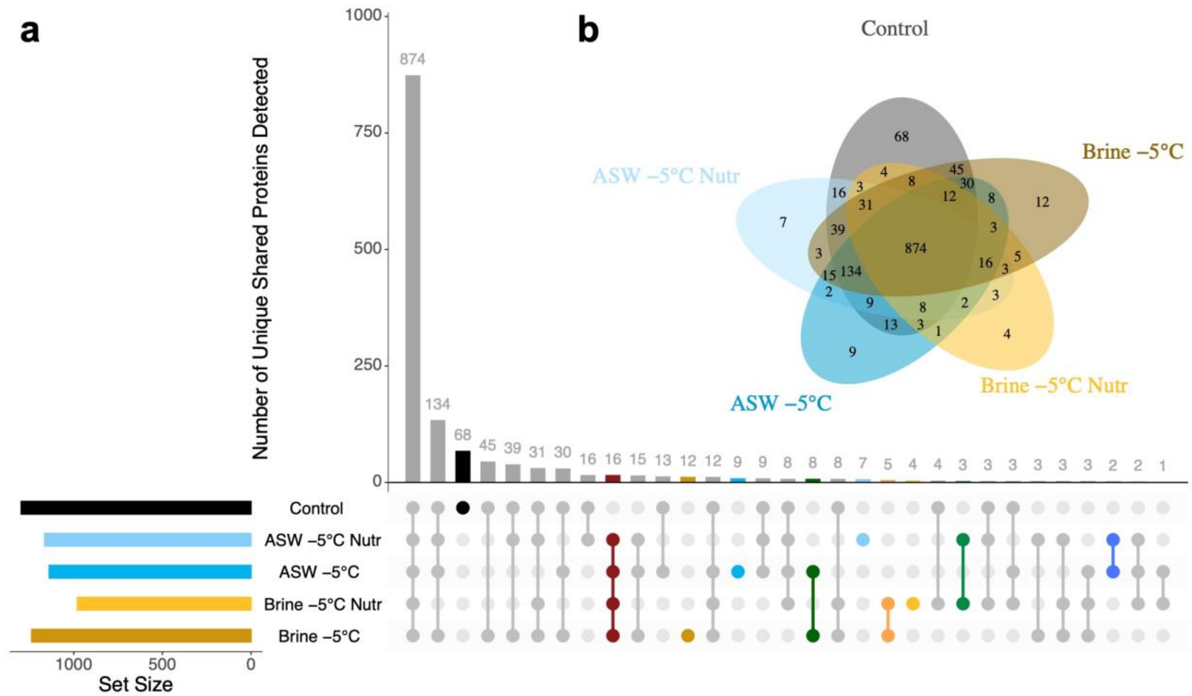


Figure 2.6. Visual analysis of quantitative proteomic data identifying unique and shared proteins across experimental conditions. a) Upset plot revealing the number of proteins uniquely shared between the conditions indicated on the parameter grid. The horizontal barchart on the left represents the number of proteins present in each condition (Set Size). The parameter grid indicates which conditions share unique detected proteins as presented in the vertical barchart. Parameter grid colors indicate intersections of proteins unique to the control (black), all experimental conditions (red), ASW-specific conditions (blues), brine-specific conditions (yellow and oranges), and unique to nutrient additions (greens). b) Venn diagram revealing the overlap and uniqueness of proteins identified across all samples.

The four treatment conditions yielded proteomes distinct from the control (Fig. 2.7, cluster 1, 4-5, 7-9). Vertical clustering of protein abundance (NSAF) in the five different treatment conditions revealed that nutrient status controlled the resulting proteome to a greater extent than salinity (i.e., brine vs. ASW), though both ASW samples were more similar to the control than brine samples with the same nutrient status (Fig. 2.7). ASW^{+nutr} had the second greatest number of proteins observed, which included a variety of proteins involved in nitrogen regulation and iron transport (Fig. 2.7, cluster 3; GlnK, GlnG, NrtA, NirB, and NasA, and 4 iron complex outer membrane transport proteins: Fe-Omp). Clusters 4 and 11 are primarily driven by the brine^{+nutr} treatment proteome, with multiple general stress response proteins differentially abundant in cluster 11, including Hfq and CPS_3912 (Fig. 2.7). Two proteins involved in the generation of exopolymers are represented in cluster 5 (Fig. 2.7, cluster 5; ExbB, Wza). Cluster 9 consisted of multiple enzymes involved in metabolism and synthesis of secondary metabolites (e.g., AceK, HepA, TypB, ACSL) and in nucleotide metabolism (PpnN, WecB). Brine^{-nutr} contained one protein with a distinct increase in abundance compared to all other conditions: CPS_3809 is a type IV pili methyl-accepting chemotaxis transducer N-terminal domain-containing protein (Fig. 2.7, cluster 10).

Around 300 proteins in both ASW treatments and around 200 proteins associated with both brine treatments were significantly more abundant relative to controls, respectively (Fig. 2.8; Table 2.4). The majority of the proteins that increased in abundance were attributed to energy production, amino transport and metabolism processes. Compared to the two brine states, ASW had a greater number of proteins involved in transcription and cell wall biogenesis (Fig. 2.9). ASW^{+nutr} had twice as many enzymes increase in abundance (17 vs. 7) involved in carbohydrate transport and metabolism, while ASW^{-nutr} had twice the number of proteins involved in lipid metabolism significantly increase (13 vs. 6, Fig. 2.9). Brine^{+nutr} was observed to have over four times the number of significant proteins decrease in abundance compared to all other conditions (489 total, 193 unique, Table 2.4), representing the functional categories amino acid transport and metabolism (69), energy production and conversion (52), translation, ribosomal structure, and biogenesis (35), and function unknown (85) (Fig. 2.9).

2.4.4 SHORT POLYPEPTIDES INDICATIVE OF HIGH SALINITY OR EXTREME LOW TEMPERATURE

After four months, samples incubated at -10°C maintained similar cell counts to those incubated at -5°C despite a complete loss of culturability (Fig. 2.3). We investigated if specific 3-

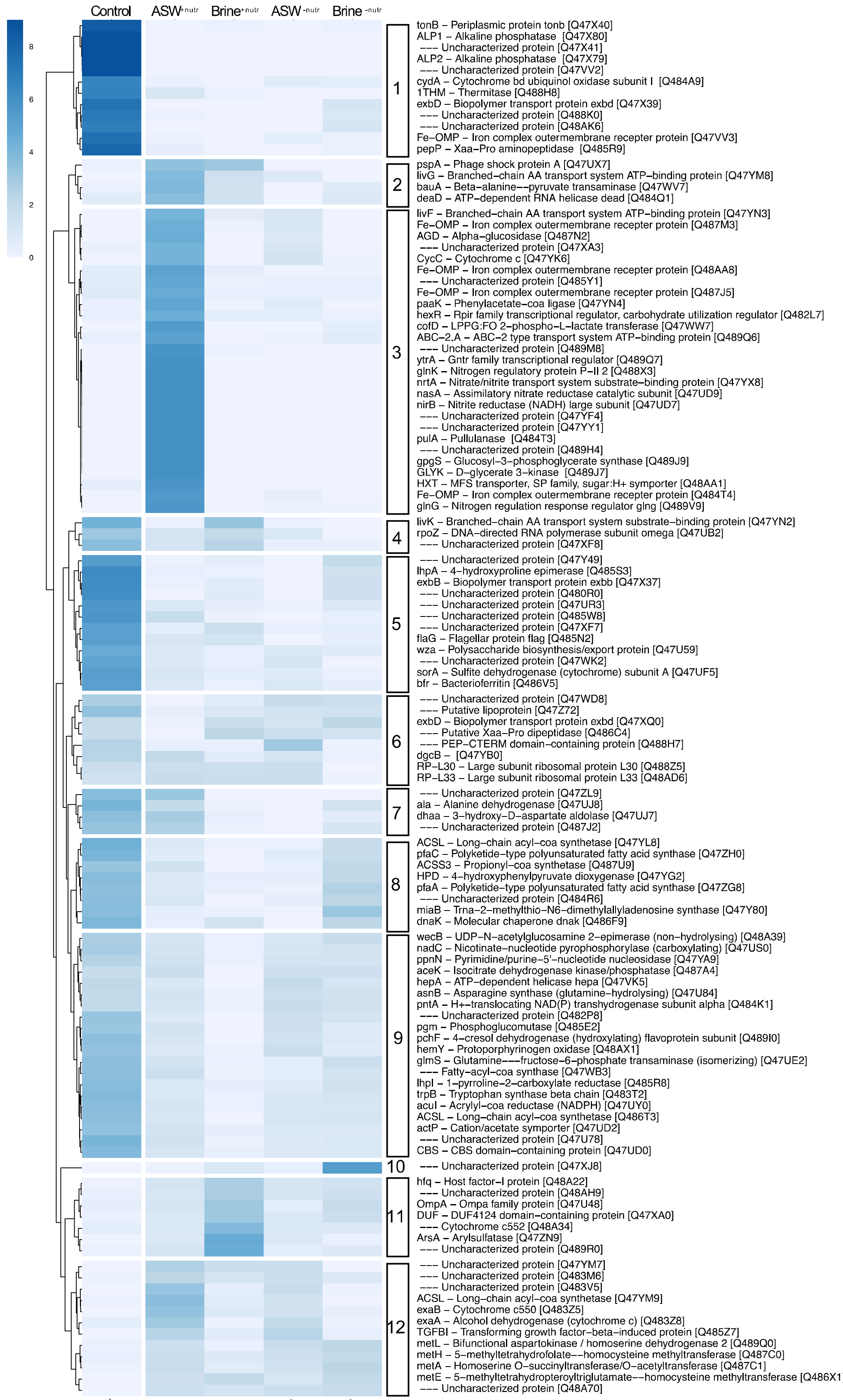

Control
ASW+nutr
Brine+nutr
ASW-nutr
Brine-nutr
8
6
4
2
0
1
2
3
4
5
6
7
8
9
10
11
12
tonB – Periplasmic protein tonb [Q47X40]
ALP1 – Alkaline phosphatase [Q47X80]
––– Uncharacterized protein [Q47X41]
ALP2 – Alkaline phosphatase [Q47X79]
––– Uncharacterized protein [Q47VV2]
cydA – Cytochrome bd ubiquinol oxidase subunit I [Q484A9]
1THM – Thermitase [Q488H8]
exbD – Biopolymer transport protein exbd [Q47X39]
––– Uncharacterized protein [Q488K0]
––– Uncharacterized protein [Q48AK6]
Fe–OMP – Iron complex outermembrane recepter protein [Q47VV3]
pepP – Xaa–Pro aminopeptidase [Q485R9]
pspA – Phage shock protein A [Q47UX7]
livG – Branched–chain AA transport system ATP–binding protein [Q47YM8]
bauA – Beta–alanine––pyruvate transaminase [Q47WV7]
deaD – ATP–dependent RNA helicase dead [Q484Q1]
livF – Branched–chain AA transport system ATP–binding protein [Q47YN3]
Fe–OMP – Iron complex outermembrane recepter protein [Q487M3]
AGD – Alpha–glucosidase [Q487N2]
––– Uncharacterized protein [Q47XA3]
CycC – Cytochrome c [Q47YK6]
Fe–OMP – Iron complex outermembrane recepter protein [Q48AA8]
––– Uncharacterized protein [Q485Y1]
Fe–OMP – Iron complex outermembrane recepter protein [Q487J5]
paaK – Phenylacetate–coa ligase [Q47YN4]
hexR – Rpir family transcriptional regulator, carbohydrate utilization regulator [Q482L7]
cofD – LPPG:FO 2–phospho–L–lactate transferase [Q47WW7]
ABC–2.A – ABC–2 type transport system ATP–binding protein [Q489Q6]
––– Uncharacterized protein [Q489M8]
ytrA – Gntr family transcriptional regulator [Q489Q7]
glnK – Nitrogen regulatory protein P–II 2 [Q488X3]
nrtA – Nitrate/nitrite transport system substrate–binding protein [Q47YX8]
nasA – Assimilatory nitrate reductase catalytic subunit [Q47UD9]
nirB – Nitrite reductase (NADH) large subunit [Q47UD7]
––– Uncharacterized protein [Q47YF4]
––– Uncharacterized protein [Q47YY1]
pulA – Pullulanase [Q484T3]
––– Uncharacterized protein [Q489H4]
gpgS – Glucosyl–3–phosphoglycerate synthase [Q489J9]
GLYK – D–glycerate 3–kinase [Q489J7]
HXT – MFS transporter, SP family, sugar:H+ symporter [Q48AA1]
Fe–OMP – Iron complex outermembrane recepter protein [Q484T4]
glnG – Nitrogen regulation response regulator glng [Q489V9]
livK – Branched–chain AA transport system substrate–binding protein [Q47YN2]
rpoZ – DNA–directed RNA polymerase subunit omega [Q47UB2]
––– Uncharacterized protein [Q47XF8]
––– Uncharacterized protein [Q47Y49]
lhpA – 4–hydroxyproline epimerase [Q485S3]
exbB – Biopolymer transport protein exbb [Q47X37]
––– Uncharacterized protein [Q480R0]
––– Uncharacterized protein [Q47UR3]
––– Uncharacterized protein [Q485W8]
––– Uncharacterized protein [Q47XF7]
flaG – Flagellar protein flag [Q485N2]
wza – Polysaccharide biosynthesis/export protein [Q47U59]
––– Uncharacterized protein [Q47WK2]
sorA – Sulfite dehydrogenase (cytochrome) subunit A [Q47UF5]
bfr – Bacterioferritin [Q486V5]
––– Uncharacterized protein [Q47WD8]
––– Putative lipoprotein [Q47Z72]
exbD – Biopolymer transport protein exbd [Q47XQ0]
––– Putative Xaa–Pro dipeptidase [Q486C4]
––– PEP–CTERM domain–containing protein [Q488H7]
dgcB – [Q47YB0]
RP–L30 – Large subunit ribosomal protein L30 [Q488Z5]
RP–L33 – Large subunit ribosomal protein L33 [Q48AD6]
––– Uncharacterized protein [Q47ZL9]
ala – Alanine dehydrogenase [Q47UJ8]
dhaa – 3–hydroxy–D–aspartate aldolase [Q47UJ7]
––– Uncharacterized protein [Q487J2]
ACSL – Long–chain acyl–coa synthetase [Q47YL8]
pfaC – Polyketide–type polyunsaturated fatty acid synthase [Q47ZH0]
ACSS3 – Propionyl–coa synthetase [Q487U9]
HPD – 4–hydroxyphenylpyruvate dioxygenase [Q47YG2]
pfaA – Polyketide–type polyunsaturated fatty acid synthase [Q47ZG8]
––– Uncharacterized protein [Q484R6]
miaB – Trna–2–methylthio–N6–dimethylallyladenosine synthase [Q47Y80]
dnaK – Molecular chaperone dnak [Q486F9]
wecB – UDP–N–acetylglucosamine 2–epimerase (non–hydrolysing) [Q48A39]
nadC – Nicotinate–nucleotide pyrophosphorylase (carboxylating) [Q47US0]
ppnN – Pyrimidine/purine–5'–nucleotide nucleosidase [Q47YA9]
aceK – Isocitrate dehydrogenase kinase/phosphatase [Q487A4]
hepA – ATP–dependent helicase hepa [Q47VK5]
asnB – Asparagine synthase (glutamine–hydrolysing) [Q47U84]
pntA – H+–translocating NAD(P) transhydrogenase subunit alpha [Q484K1]
––– Uncharacterized protein [Q482P8]
pgm – Phosphoglucomutase [Q485E2]
pchF – 4–cresol dehydrogenase (hydroxylating) flavoprotein subunit [Q489I0]
hemY – Protoporphyrinogen oxidase [Q48AX1]
glmS – Glutamine–––fructose–6–phosphate transaminase (isomerizing) [Q47UE2]
––– Fatty–acyl–coa synthase [Q47WB3]
lhpI – 1–pyrroline–2–carboxylate reductase [Q485R8]
trpB – Tryptophan synthase beta chain [Q483T2]
acuI – Acrylyl–coa reductase (NADPH) [Q47UY0]
ACSL – Long–chain acyl–coa synthetase [Q486T3]
actP – Cation/acetate symporter [Q47UD2]
––– Uncharacterized protein [Q47U78]
CBS – CBS domain–containing protein [Q47UD0]
––– Uncharacterized protein [Q47XJ8]
hfq – Host factor–I protein [Q48A22]
––– Uncharacterized protein [Q48AH9]
OmpA – Ompa family protein [Q47U48]
DUF – DUF4124 domain–containing protein [Q47XA0]
––– Cytochrome c552 [Q48A34]
ArsA – Arylsulfatase [Q47ZN9]
––– Uncharacterized protein [Q489R0]
––– Uncharacterized protein [Q47YM7]
––– Uncharacterized protein [Q483M6]
––– Uncharacterized protein [Q483V5]
ACSL – Long–chain acyl–coa synthetase [Q47YM9]
exaB – Cytochrome c550 [Q483Z5]
exaA – Alcohol dehydrogenase (cytochrome c) [Q483Z8]
TGFBI – Transforming growth factor–beta–induced protein [Q485Z7]
metL – Bifunctional aspartokinase / homoserine dehydrogenase 2 [Q489Q0]
metH – 5–methyltetrahydrofolate––homocysteine methyltransferase [Q487C0]
metA – Homoserine O–succinyltransferase/O–acetyltransferase [Q487C1]
metE – 5–methyltetrahydropteroyltriglutamate––homocysteine methyltransferase [Q486X1]
––– Uncharacterized protein [Q48A70]

Figure 2.7. Vertically and horizontally clustered heatmap view of all proteins revealed to be increased in abundance in one of the four treatment conditions at −5°C (ASW^{+nuts}, brine^{+nuts}, ASW^{-nuts}, brine^{-nuts}) relative to the control condition by $\log_2$ fold $\geq$ 2 ($p < 0.05$). Vertical dendrogram indicates hierarchical clustering of the different samples, while horizontal dendrogram clustered protein NSAF values normalized by row (12 clusters, 118 proteins).

4 amino acid-long polypeptide combinations ("k-mers") indicative of the individual stressors of extreme low temperature (-10°C), high salinity (brine), or the co-occurrence of these stressors accumulate in extremophile cells. ANOVA analysis of the semi-quantitative proteomic data from the -5°C and -10°C treatments identified unique sets of proteins specifically associated with extreme low temperature (885 proteins), high salinity (586 proteins), and the interaction of the two stressors (1802 proteins) at a 1% FDR threshold.

In general, there were fewer temperature-specific k-mers (79) than brine-specific k-mers (180; Fig. 2.10A). Proteins identified to be significantly increased in the -10°C brine treatments contained the highest number of unique enriched k-mers (367; Fig. 2.10a). Temperature-specific 3-mers in -10°C conditions were significantly enriched in glutamic acid (E) and valine (V), while aspartic acid (D) and threonine (T) were depleted relative to all other detected protein k-mers. The distributions of amino acids were not significantly different in the temperature-specific 4-mers. Salinity-specific 3-mers in brine conditions were significantly enriched in alanine (A) and depleted in aspartic acid (Q) relative to all other detected protein k-mers (Fig. 2.10b). The distributions of amino acids in brine-specific 4-mers also had significantly higher frequencies of alanine (A), but also higher abundances of the polar amino acid glutamine (Q). These brine-specific 4-mers were likewise depleted in aspartic acid (D), and the two non-polar amino acids phenylalanine (F) and tryptophan (W) (Fig. 2.10b; Appendix A.7).

One k-mer, *Lys-Asp-Ala* (KDA), was identified to be enriched in each of the three isolated analyses of extreme low temperatures (-10°C), high salinity (brine), and a combination of both (-10°C brine) responses (Fig. 2.10a). The four k-mers that were identified to be enriched in the brine and -10°C treatments separately, but not in the co-occurring stressor treatment, were

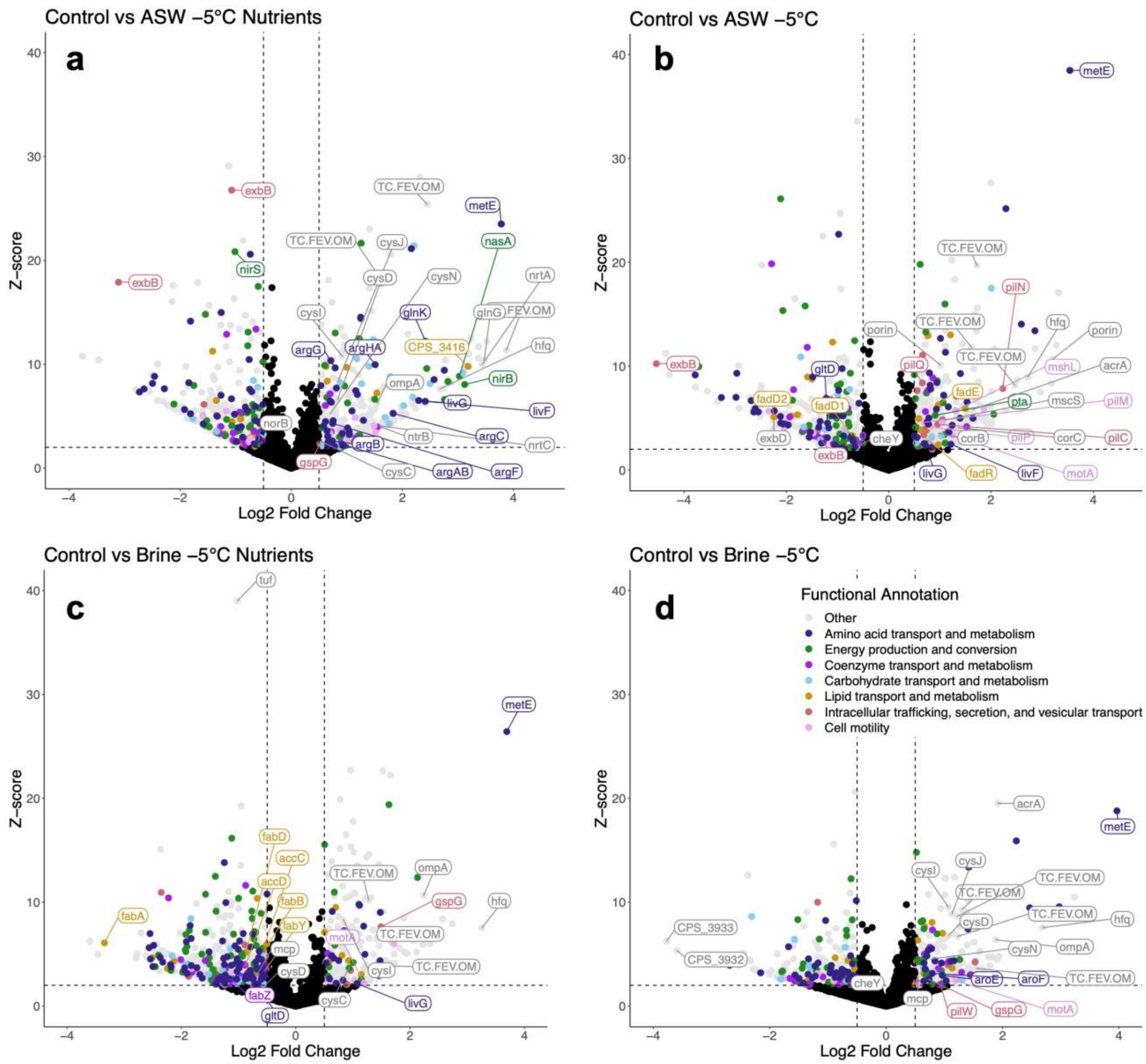


Figure 2.8. Differential abundances (log2 fold change: LFC) of all proteins detected in four experimental -5°C conditions compared to control cells and their respective significance z-score calculated. Average protein abundances were determined from spectral counts from the control samples and compared to the average spectral counts for the experimental conditions a) ASW^{+nuts}, b) ASW^{-nuts}, c) brine^{+nuts}, d) brine^{-nuts}. Proteins meeting a significance threshold of LFC ≥ |0.5| and a z-score ≥ 2 are indicated in color where each color represents a different EggNOG term. Functional annotations are: dark blue: amino acid transport and metabolism, green: energy production and conversion, magenta: coenzyme transport and metabolism, light blue: carbohydrate transport and metabolism, orange: lipid transport and metabolism, dark pink: intracellular trafficking secretion and vesicular transport, light pink: cell motility. Proteins were labeled with a protein name if the LFC ≥ |2| or z-score ≥ 15.

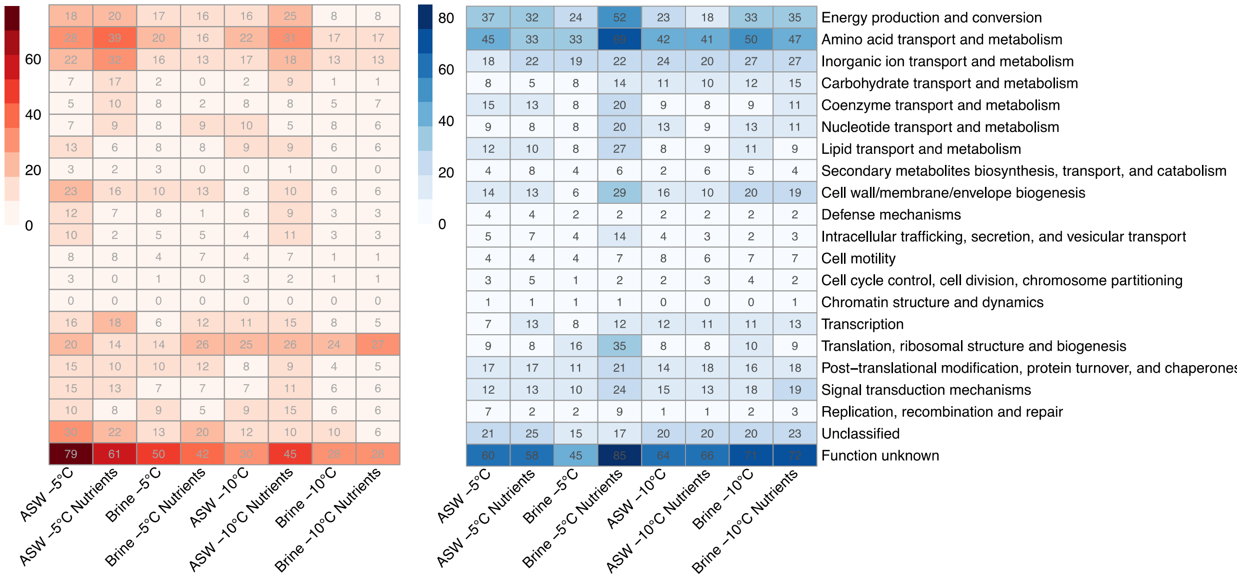


Figure 2.9. Proteins quantified and assigned to EggNOG terms. Proteins identified via mass spectrometry to undergo significant changes in abundance compared to optimal conditions (increases shown in red, decreases shown in blue) were mapped to functional orthology enrichment terms using EggNOG. A count was determined (shown in individual boxes) for each experimental condition to identify differences between temperature, salinity, and nutrient responses.

Table 2.4: Summary of unique proteins decreased in abundance at -5°C compared to controls held at -1°C.

Experimental Condition	Total Number of Proteins Decreased (compared to -1°C)	Total Number of Proteins Unique to Each Condition	Percentage of Proteins Identified*
ASW **+nuts**	301	37	54.1
ASW **-nuts**	312	41	61
Common		33	60.6
Brine **+nuts**	**489**	**193**	81.9
Brine **-nuts**	237	23	73.9
Common		27	81.5

All experimental conditions contain samples incubated at -5°C, where Common indicates proteins shared between salinities (ASW or Brine). Protein abundance was compared to control to calculate log fold change (LFC), where LFC <-0.5 indicates significant decrease for calculating total number of proteins per condition. Number of unique proteins shows venn diagram results of proteins uniquely decreased in a specific condition or shared salinity.*Calculated as number of proteins identified by EggNog (Gene Ontology analysis) out of total number proteins significantly changed, indicates percentage of proteins with known classification.

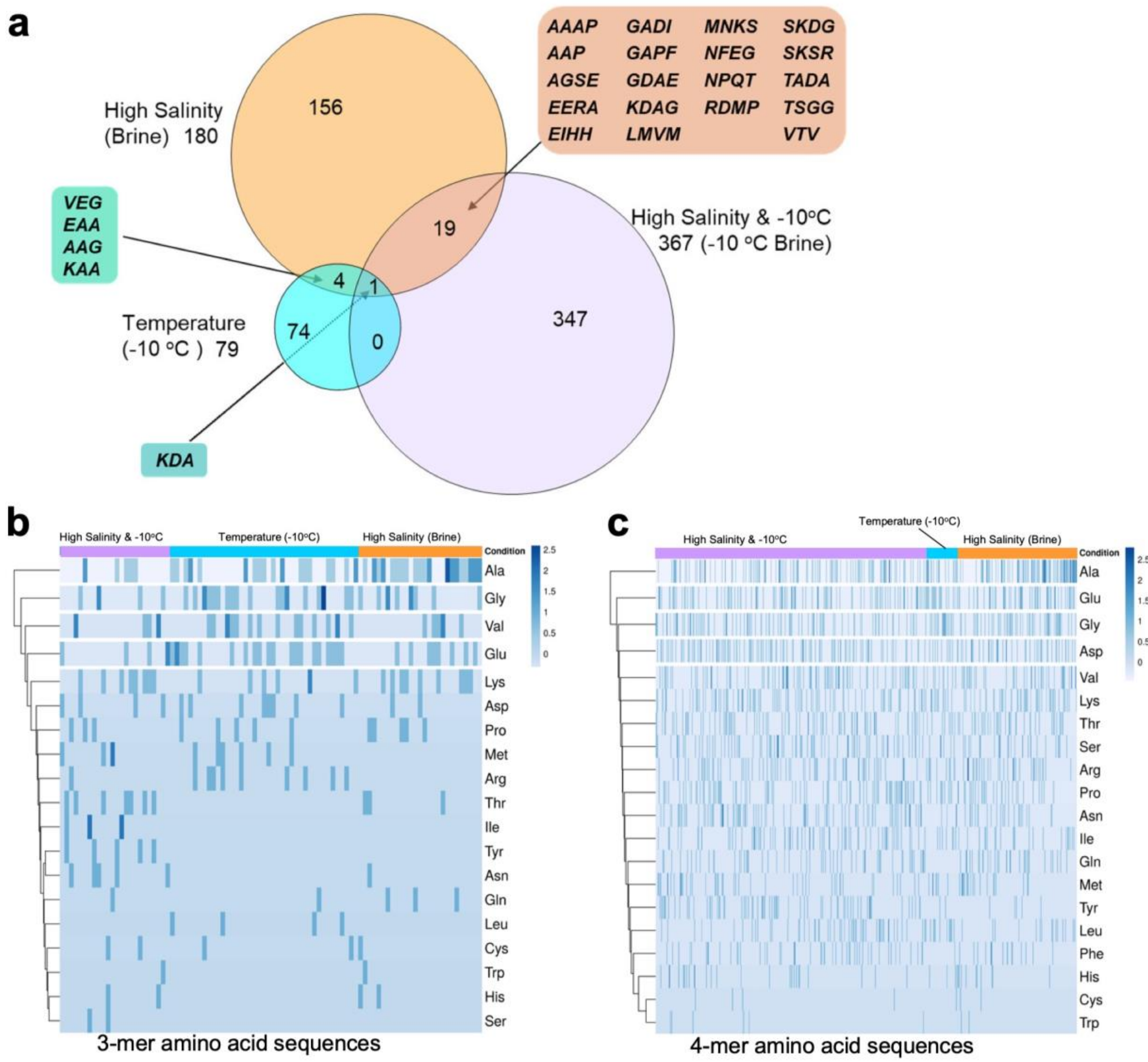


Figure 2.10. Enriched k-mers attributed to extreme environmental stressor conditions. a) Distribution of k-mers (3-4 amino acid long polypeptides) identified to be enriched in proteins in significantly higher abundances in high salinity (i.e., brine; orange), extreme low temperature (i.e., -10°C; blue), and the combined analysis of high salinity and low temperature (i.e., -10°C brine; purple). Subsets include lists of 3-mers and 4-mers identified to be enriched in more than one condition. b) Heatmap of the distribution of amino acids found in 3-mers (irrespective of position). c) Heatmap of the distribution of amino acids found in 4-mers (irrespective of position). For heatmaps, amino acid counts in each k-mer are represented by rows with no scaling applied, and rows are clustered using Euclidean distance and average linkage. Columns represent treatments and are not clustered.

dominated by the presence of alanine*: Val-Glu-Gly, Glu-Ala-Ala, Ala-Ala-Gly, and Lys-Ala-Ala* (VEG, EAA, AAG, KAA; Fig. 2.10a). Analysis of shared k-mers between the high salinity and the combined high salinity and low temperature treatment revealed 19 k-mers. Of the 534 4-mers identified to be enriched in the proteins inferred from our peptide data, 96% (512) of them were actually detected by mass spectrometry in our condition-specific samples.

2.5 DISCUSSION

2.5.1 CULTURABILITY AND ACTIVITY ANALYSIS

As is the case for sea ice (when bacterial cells get included into recently formed ice immediately experiencing a change in salinity and temperature in their new habitat), cells of *Cp*34H were immediately transitioned from control conditions into their respective media destined for long-term incubations without any prior acclimatization steps. As shown previously for in-ice incubations (Junge *et al.* 2006; Nunn *et al.,* 2015), *Cp*34H is capable of withstanding this rapid transition into subzero solutions, evidenced by positive measures of incorporation of [3H]-leucine (i.e., protein synthesis) at rates that indicate that cells in the majority of treatments were in metabolic states of either growth or maintenance by 24 hours (metabolism-terms according to established rates in Price and Sowers, 2004; Table 2.3; Fig. 2.4). The only exception was -10°C brine^{-nutr}, with an activity rate representative of a physiological survival state (Table 2.3). Here, we characterize the response of *Cp*34H after one month incubation as being in one of four physiological activity states: *i)* growing and culturable cells (high levels of activity and culturability; -5°C, ASW), ii) active cells (low levels of activity and moderate culturability; -5°C, brine), *iii)* viable but not culturable, or VBNC, cells (moderate levels of activity and very low to no culturability; -

10°C, ASW^{-nutr}), and *iv)* surviving cells (very low to no levels of activity and not culturable but persisting as evidenced by maintained total counts; -10°C, brine).

Throughout this paper we chose to use the term culturability as an assessment of cell state as it most accurately reflects what is being measured in our samples. In the literature, culturability has also been used synonymously with viability, alluding to the fact that cells must be viable in order to be amenable to culturing (Pinto *et al*., 2015). For many bacterial species, as was the case in our study, cells can still be viable (as indicated by positive measures of [3H]-leucine based protein synthesis) but not culturable (as indicated by no recoverability in culture). This cell state, referred to in the literature as the VBNC (Viable But Not Culturable) cell state, has been described as a survival strategy for cells coping with environmental stress without entering dormancy (Pinto *et al*., 2013; Ayrapetyan and Oliver, 2016). Evidence of the VBNC state in our -10°C ASW^{-nutr} samples at one month warrants further study, especially in view of the fact that a better understanding of this adaptive strategy may help inform future astrobiological searches, particularly given the relevance of maintaining a measurable level of activity that ensures long-term survival in salty and super cold environments such as might be found on other icy worlds.

Our finding of continued protein synthesis activity at -10°C in supercooled ASW, but not in high salinity brine, suggests salinity stress as the most detrimental factor under the conditions studied here. This is congruent with the known growth salt content range of *Cp*34H ([1.5 - 7.0%, in marine medium], Huston, 2003), and the fact that cells were not culturable after one month of incubation in -10°C brine. After four months at -10°C, cells in all treatments were unculturable (Fig. 2.2), suggesting that both the lower temperature and the higher salinity affected culturability to a greater degree after prolonged incubation. Organisms frozen in sea ice at −10°C are expected to be partitioned in the brine fraction of the ice (Junge *et al*., 2019), and thus experience conditions equivalent to those of our −10°C equilibrium brine treatments (Fig. 2.2d). Our data suggest that *Cp*34H would not be active under the more extreme temperature and salt content conditions within one month of incorporation into sea ice. However, natural wintertime sea ice samples are known to contain an abundant supply of potentially cryo- and halo-protective exopolymers thought to be important for microbial survival (Krembs *et al*., 2002, 2008, 2011) and in-ice activity (Junge *et al*., 2006), and it is possible that *Cp*34H remains active when embedded within such substances in the ice (Junge *et al*., 2006; Marx *et al*., 2009). Further studies that include polymer amendments could shed light on the likelihood of active growth in high salinity brines of *Cp*34H with significance for sea ice environmental conditions.

2.5.2 THE DYNAMIC PROTEOMIC RESPONSE OF Cp34H IN SUBZERO TEMPERATURES

The predicted proteome for *Cp*34H consists of 4,910 proteins (Methe *et al*., 2005), of which we identified 2,362 proteins (48% of the predicted proteome). This expands on previous studies by Nunn *et al.* who identified 1,763 proteins (36%; Nunn *et al.*, 2015). Technical and biological replicates of proteomic signatures from each condition form distinct subpopulation clusters in an NMDS analysis, providing confidence that the four month-long treatments yield distinct protein responses based on temperature, salinity, and nutrients in *Cp*34H (Fig. 2.5). Because samples at -5°C retained culturability (Fig. 2.2a, c), these proteins were analyzed to provide information about how active cells mitigated decreased temperature, increased salinity, and nutrient limitation; samples at -10°C, which lost culturability but contained persisting cells (Fig. 2.2b, d; 2.3), were analyzed for the enrichment of polypeptides that could serve as biomarkers for previously active life on icy worlds.

2.5.3 CONSTITUTIVE PROTEOMIC RESPONSE OF Cp34H TO FOUR MONTH EXPOSURES OF -5°C

The constitutive proteomic response from *Cp*34H suggested a metabolic shift toward an anaerobic-like state through the increased expression of proteins involved in the Entner-Doudoroff pathway (Czajka *et al.*, 2018) and sulfate reduction (Parro *et al.*, 2018) (Fig. 2.8a-d). The alteration of these central core metabolisms in four month-long incubations support calculations and previous suggestions that the cellular chemical reactions in these pathways occur more spontaneously (i.e., more negative value calculated for Gibbs free energy) than traditional C, N, and S acquisition (Flamholz *et al*., 2013; Noor *et al*., 2014; Czajka *et al*., 2018). Observed increases in amino acid transport across all conditions emphasizes the rapid response of these cells to the available extracellular substrates (*i.e.,* free leucine) and intracellular recycling of amino acids as a primary nitrogen source. In particular, cysteine and methionine metabolism, a metabolic source of sulfur and nitrogen, were observed to increase in all conditions. Additionally, proteins involved in cellular motility, including flagellar, pilus, and chemotaxis proteins, were abundant across all conditions, supporting previous examinations of *Cp*34H in response to subzero temperatures (Nunn *et al*., 2015; Showalter and Deming, 2018) and presenting additional motility-specific biomarkers for extant life as desired by NASA's Ladder of Life Detection (Neveu *et al*., 2018).

2.5.4 RESPONSES TO TEMPERATURE, SALINITY, AND NUTRIENTS

Our analysis of the proteins that significantly diverge in critical metabolic pathways of each condition at -5°C in comparison to the control identified unique mitigation strategies in each individual treatment. These condition-specific analyses revealed that the availability of nutrients triggered the use of new pathways of carbon and nitrogen assimilation and the additional stress of low temperatures or high salinity initiated intracellular recycling to strengthen cell walls. In addition to analyzing each individual condition, a broader analysis of environmental variables was performed to isolate effects of salinity, nutrients, and temperature on *Cp*34H.

ASW with Nutrients. Cells incubated in ASW^{+nutr} at -5°C experienced conditions closest to the -1°C control, allowing for an isolated examination of long-term temperature- and reduced available oxygen- dependent effects. The reported growth range for *Cp*34H extends from 19°C to -12°C, with an optimal growth temperature of 8°C (Huston, 2003; Wells and Deming, 2006a,b; Junge *et al*., 2003). Within the context of this range, we hypothesize that *Cp*34H experiences minor temperature stress, not enough to permanently damage cells, but enough to activate metabolic pathway networks within the cell that lead to an optimized maintenance of functionality. At -5°C, cells were able to maintain levels of culturability equivalent to that experienced by cells in optimal conditions (Fig. 2.2a). Activity as measured by the maximum incorporation rate of [3H]-leucine after 1 hour was also observed at rates (2.37×10^{-2} g C (g C)$^{-1}$ h^{-1}) consistent with either growing or metabolically active cells (Price and Sowers, 2004; Junge *et al*., 2006) (Table 2.3).

The predominant mitigation strategy of -5°C ASW^{+nutr} to withstand long-term incubation was the increased utilization of amino acid metabolism (Fig. 2.8a; 2.9), facilitated primarily by mechanisms of assimilatory nitrate reduction described herein, that allowed *Cp*34H to preserve activation of necessary pathways for maintaining cell activity over prolonged time frames (Fig. 2.7). An increased production of amino acids in ASW^{+nutr} requires a readily available nitrogen source. Though it is difficult to ascertain if cells in ASW^{+nutr} were nitrogen limited by month four, the increase in abundance of nitrogen regulatory proteins, including nitrate ABC transporter (NrtA), nitrate reductase (NasA), nitrite reductase NADPH (NirB), nitrogen regulatory protein PII (GlnK), and DNA-binding transcriptional regulator NtrC (GlnG) (Fig. 2.7, cluster 3; Appendix A.1), suggests *Cp*34H modulates its need for nitrogen through the directed acquisition of extracellular nitrate, a predominant and readily available form of nitrogen shown to drive biogeography and trophic status in Arctic surface waters and sea ice (Ardyna and Arrigo, 2020; Clark *et al*., 2020;

Henley *et al*., 2020). Notable nitrate transporters included NrtA, an ABC transporter responsible for high-affinity nitrate acquisition at low concentrations (Nagore *et al*., 2006; Akhtar *et al*., 2015), and GlnG, a component of the NtrB/NtrC system that responds to nitrogen limitation and plays an important role in the regulation of transcription (Pahel *et al*., 1982). *Cp*34H's preferential expression of high-affinity transporters in this analysis suggested that the initial nutrient concentration supplied did not sustain the metabolic needs of the cells over the entire four month incubation, possibly due to ASW^{+nutr} at -5°C possessing the highest rates of activity during early time point measurements (Table 2.3). Other notable nitrogen metabolism proteins increased in ASW^{+nutr}, such as nitrate reductase (NasA) and nitrite reductase NADPH (NirB), suggested that *Cp*34H may utilize assimilatory nitrate reduction pathways for producing ammonia (Fig. 2.7, cluster 3; Appendix A.1). For example, *Mycobacterium tuberculosis* incubated under hypoxia conditions was able to utilize nitrite as a sole nitrogen source due to NirB, an iron-sulfur containing protein that can reduce nitrite directly to ammonia (Jackson *et al*., 1981; Wang and Gunsalus, 2000; Chen and Wang, 2015), allowing critical nitrogen regulation when cells entered dormancy states (Akhtar *et al*., 2013). This assembly of proteins emphasized the benefits of metabolic diversification in nitrogen regulation in the dual presence of low temperature and potential reduced oxygen stress.

In close relation to nitrogen metabolism, ASW^{+nuts} uniquely revealed increased abundances of enzymes involved in arginine biosynthesis, while all other conditions decreased the abundance of these enzymes (Appendix A.2). Previous studies have suggested psychrophiles reduce arginine content in proteins because of increased hydrogen bonding and reduced protein flexibility (Maayer *et al*., 2014), while other studies have revealed that marine psychrophiles grown with readily available nutrients, such as this, use chemotaxis and active transport to acquire and assimilate arginine (Geesey and Morisa, 1979). Arginine can be utilized in the biosynthesis of glutamate and evidence here supports the hypothesis that glutamate is being synthesized from the abundant branched-chain amino acids (Wu and Morris, 1998) (Fig. 2.8a, putative branched chain amino acid transporters: CPS_3416, CPS_3412/LivF, CPS_3417/LivG). Significant increases in GlnK, a component of the nitrogen regulatory protein PII, also suggested that cells tightly regulate acetyl CoA and arginine metabolism (Watzer *et al*., 2019). Overall, these preferential mechanisms of nitrogen regulation through nitrate reduction proteins and arginine biosynthesis pathways allow ASW^{+nuts} to maintain a wide array of metabolisms that sustain its high culturability over four month incubations at a decreased temperature.

ASW without Nutrients. In the Polar oceans, nutrient levels vary widely, have changed considerably with the climate change pronounced there, and show low and high nutrient conditions occurring at different locations or at different times of the year (Ardyna and Arrigo, 2020; Henley *et al*., 2020). In cold, nutrient-depleted waters, organisms require multiple strategies for nutrient acquisition and cold-response maintenance for long in-ice residence times. In our incubations, the one order of magnitude reduction in culturability of cells observed in ASW^{-nutr} compared to cells in ASW^{+nuts} indicated that the presence of nutrients plays a critical role in maintaining higher levels of culturability under subzero conditions (Fig. 2.5), confirming the widely accepted hypothesis that psychrophilic growth is enhanced with increased nutrient availability (Moyer *et al*., 2017). Primary drivers of metabolism in ASW^{-nutr} included modifications to the membrane through the regulation of two-component systems, increases in membrane receptors, and alterations to fatty acid metabolism (Fig. 2.8b, 2.9; Appendix A.4).

The major membrane proteins substantially increased were TonB receptors, with 12 enzymes (*i.e.,* TonB, ExbB, ExbD, and Fe-OMP iron complex outer membrane receptors) discovered to have a significant $\log_2$ fold change in ASW^{-nutr} (Fig. 2.8B). TonB receptors are active transporters that facilitate uptake of a wide range of molecules and nutrients (i.e., vitamin B12, siderophores, polysaccharides, aromatic compounds (Pawelek *et al.,* 2006; Shultis *et al.,* 2006; Bolam and Berg, 2018; Fujita *et al.,* 2019)). TonB is part of a membrane complex that includes ExbB and ExbD proteins; in our data, these proteins were observed to decrease significantly in all conditions compared to the control cells (Fig. 2.7). As extracellular polysaccharides (EPS) play a key role in stabilizing extracellular environments for *Cp*34H and other psychrophiles (Deming and Young, 2017; Bowman, 2017), it was interesting that despite TonB receptors being significantly increased, ExbB and ExbD are only increased in control conditions (Fig. 2.7). This may be a time-dependent expression, suggesting that EPS is already present in the cell matrix and does not need to be remade through the four month incubation. Notably, several other membrane proteins were observed to significantly increase in ASW^{-nutr} (12 efflux pumps: RND, Acr family A/B/D/E/F, magnesium CorC and cobalt; 9 type IV pilus biogenesis proteins: Pil-N/O/C/P/M/Q, MshL; 2 Omp porins; 4 mechanosensitive ion channel proteins; Fig. 2.8B). These membrane modifications may serve to aid cells in their acquisition of extracellular nutrient sources through control of membrane permeability or nutrient depletion-directed motility (Ferenci 2005; Castillo-Keller *et al*., 2006; Ni *et al*., 2016).

Additionally, the existence of proteins from both biosynthesis and degradation pathways in fatty acid metabolism suggested that long-chain fatty acids were constructed or deconstructed as needed, confirming the critical importance of adapting membrane fluidity for maintaining activity at low temperatures (Bowman 2017). Specifically, lipopolysaccharide assembly protein LptD was increased, along with AccB, an acetyl-CoA carboxylase that leads to the production of malonyl-CoA (Fig. 2.8b). Malonyl-CoA signals bacteria to increase their production of fatty acids and phospholipid biosynthesis by acting as a potent negative regulator of fatty acid catabolism (Fujita *et al.,* 2007). Additionally, increases in fatty acid degradation proteins FadB and FadE were offset by the presence of FadR, a regulator for fatty acid metabolism (Fujita *et al.,* 2007; Zhang *et al.,* 2012). *Cp*34H encodes multiple copies of various fatty acid metabolism regulatory proteins in its genome (Wan *et al.,* 2016), emphasizing the need for high-precision regulation for the utilization of fatty acids in diverse environments. The increase in fatty acid biosynthesis enzymes, even in the absence of external nutrients, showed that 1) *Cp*34H has multiple methods for coping with nutrient limitation, 2) cells are adept in balancing multiple external stressors, and 3) the combination of the more moderate temperature and nutrient stresses tested here are not enough to engage the prioritization of fatty acid breakdown, thus showing that cells are prepared to handle more extreme conditions than those in ASW^{-nutr} incubations.

Brine with Nutrients. Shelf Arctic ocean waters are fed by nutrient-rich sources (Torres-Valdés *et al*., 2016), and brine inclusions within land-fast sea ice are often respectively nutrient replete, offering a variety of extracellular substances for organisms residing in such inclusions to utilize for prolonged survival. Some of the metabolites captured in brine inclusions contain classes of compatible solutes that can serve as osmoprotectants against the increased salinity imposed by the brine as the ice ages (Welsh, 2000; Firth *et al*., 2016). The slight decrease in culturability in brine^{+nutr} (Fig. 2.2c) reflected an expansive salinity and growth response that decreased both the diversity and abundance of proteins detected (Fig. 2.8c; Table 2.4), yielding the most distinct matrix of proteins tested in this study (Fig. 2.5). Of the four conditions at -5°C, brine^{+nutr} had the lowest number of unique proteins (Fig. 2.6, yellow) and the highest number of proteins significantly decreased in abundance compared to the control, independent of functional category (Fig. 2.7c; Table 2.4). Large decreases in ribosomal subunits were observed that were not apparent in other conditions (Appendix A.5), suggesting that part of the overall decrease in protein abundance is due to the lack of available cellular machinery for engaging in translation. The general decrease observed in both protein count and abundance detected compared to other conditions suggests that brine^{+nutr} cells increased protein turnover in order to prioritize a subset of enzymes to survive the increased salinity exposure (Fig. 2.7).

Potential protein biomarkers of salinity stress observed in this study are Hfq, OmpA, and type IV pilus proteins (Fig. 2.8c, d). The most notable protein contributing to these changes is Hfq, a stress-activated RNA-binding chaperone protein that directs a profound number of biological processes in bacteria including carbon metabolism (Vogel and Luisi, 2011), fatty acid biosynthesis (Bianco *et al.*, 2019; Huber *et al.,* 2020), chemotaxis and motility (Soutourina and Bertin, 2003), and membrane proteins (Deng *et al.,* 2016). Hfq had the second highest increase in log fold change (Fig. 2.8c) in brine^{+nuts}; while Hfq was detected as increased at high levels in all experiment conditions (Fig. 2.8a-d), the combination of brine and nutrients extends the influence of this RNA chaperone to lead the induction of widespread protein decreases. Because of its increased prevalence with the addition of multiple stressors, Hfq could serve as a potential biomarker for cells in extreme environments. The Hfq gene is required for the regulation of OmpA, an outer membrane protein that is osmoregulated, involved in maintaining membrane integrity (Udekwu *et al.,* 2005; Choi and Lee*,* 2019), and was increased in abundance in brine^{+nuts}. Evidence of increased OmpA in halophilic bacteria supports the use of this protein as an indicator for a prolonged salinity stress response (Yun *et al.,* 2018). Additionally, OmpA mRNA is notably stable, which could increase the likelihood of translation under extreme stress (Nilsson *et al.,* 1984; Vytvytska *et al.,* 1998, 2000). The increase in type IV pilus protein MshD, 1 of 4 proteins uniquely detected in brine^{+nutr} (Fig. 2.6), reveals a brine^{+nuts} - driven response and may represent a biomarker for psychrophiles in higher nutrient brines. Additionally, a significant increase in branched-chain amino acid ABC transporters was observed both in brine^{+nutr} and ASW^{+nutr} conditions (LivG, LivF, LivK, Fig. 2.7, clusters 2, 3, 4; Appendix A.6), confirming the preferential utilization of leucine, which was provided in our nutrient additions, and thus validating our use of leucine as a protein synthesis tracker for activity assessment (Table 2.3).

Proteins with the greatest decrease in abundance in brine^{+nutr} were involved in fatty acid metabolism (Fig. 2.8c, Table 2.4). Enzymes involved included acetyl-CoA carboxylase carboxyl transferase (AccC/A/D, Fig. 2.8c) and acyl-carrier proteins (FabG/V/D/B/A/Z/F/Y, Fig. 2.8c). Brine^{+nutr} was the only condition to induce widespread decreases in fatty acid biosynthesis, though no notable enzymes involved in degradation were detected as significant (Appendix A.7). This indicated that the entire process of fatty acid metabolism has been included in the observed protein decrease experienced. It is possible to infer that cells avoid expending energy on the degradation of fatty acids due to the presence of other utilizable nutrients in their environment that they can acquire through the increased use of type IV pilus proteins. This study reveals that cells experiencing increased salinity in the presence of a readily available nutrient source undergo widespread decreases in metabolism while maintaining select stress proteins that are increased to levels greater than in any other condition.

Brine without Nutrients. Bacteria in these subzero, hypersaline environments possess adaptations of great relevance to astrobiology, as the presence of liquid water on icy worlds facing constant extreme cold temperatures is dependent upon an over-saturation of salts (Zolotov, 2007; Dickson *et al.,* 2013; Vance *et al.,* 2019). As such, proteins increased in these incubations could inform further investigation into valuable biomarkers for life in extreme conditions. In response to varied stressors, brine^{-nutr} cells exhibited the highest number of unique proteins in any experimental group (Fig. 2.8b, c), including a putative gluconeogenesis factor protein for regulating cell shape (CPS_2836), NAD kinase (NadK) which synthesizes NADP and acts as an essential reducing agent (Spaans *et al.,* 2015), an ubiquinone kinase (UbiB) that functions under anaerobic conditions (Meganathan *et al.,* 2001), and membrane-bound proteins such as an RND efflux- and ATP- transporters that can regulate the ion concentrations within cells (Tokunaga *et al.,* 2004; Li and Nikaido, 2009). This diverse set of unique proteins suggests that brine^{-nutr} cells must initiate a suite of metabolic processes to survive cold, high salinity conditions with limited nutrients for extended periods of time. The varied metabolic response of -5°C brine^{-nutr} was only partially effective at maintaining culturability after four months (Fig. 2.2c), with an NMDS analysis revealing that these samples clustered nearest cells incubated at -10°C which completely lost recoverability (Fig. 2.2b, D; 2.5a). Conversely, the comparable condition of ASW^{-nutr}, with equivalent temperature and nutrients, exhibited only a moderate decrease in culturability (Fig. 2.2a), suggesting that proteins specific to brine^{-nutr} were responsible for conferring increased protection against salinity stress. An increase in enzymes involved in aromatic amino acid metabolism, including activation of the shikimate pathway (AroE, CPS_3954, TrpCF, TrpE, HisC, TyrA, Fig. 2.8d), have all been previously noted as defense mechanisms against increased salinity (Falb *et al.,* 2008; Lee *et al.,* 2016).

When cells experienced temperature, salinity, and nutrient stress simultaneously (*i.e.,* brine^{-nutr} samples), we detected an increase in the ATP-dependent peptidase FtsH, an important regulator of growth, cell division, and extracytoplasmic membrane stress that is also required for the synthesis of multiple components of polysaccharides (Ito and Akiyama, 2005)(Fig. 2.8d). FtsH is a membrane protein responsible for degrading thermodynamically unstable proteins and maintaining cellular and membrane lipopolysaccharides (LPS)(Shimohata *et al.*, 2002; Ito and Akiyama, 2005). These membrane-specific changes, combined with increases in the FTS negative regulator protein YccA, associated type II secretion system proteins (GspC, GspL, Fig. 2.8d), and other fatty acid regulation proteins (LptB, DjlA, PaaK, FadH, Fig. 2.8d), suggest that brine^{-nutr} mitigates the matrix of environmental stressors by actively maintaining a highly regulated, LPS-enriched membrane. Fortified (and specialized) membranes can buffer the cells against the subzero saline environment while simultaneously managing any internal damage to cellular processes via DNA damage repair (SbcC, RadA) and purine salvage pathway proteins (Ade), two processes that are likely correlated with the high salt content of the brine incubation (Fig. 2.8d). The nutrient-limited state also increases chemotaxis, motility, and amidohydrolase proteins common to both nutrient-limited conditions (see *General Nutrient-Limitation Response* below). Additionally, enoyl-CoA hydratase (*i.e.,* PaaK) isolated from the haloarchaea *Haloferax mediterranei* was essential for the mobilization of carbon stored in polyhydroxyalkanoate (PHA) through beta-oxidation pathways also present in *Cp*34H, suggesting a possible recycling of fatty acids for use as a carbon source in brine^{-nutr} (Methe *et al.,* 2005*;* Liu *et al.*, 2016).

General Salinity-Dependent Response. In our samples, protein pathways prioritized in response to increased salinity include general amino acid transport and recycling, an intracellular metabolism that would also provide the cell with needed carbon and nitrogen (Fig. 2.9). OmpA was also prioritized and has been characterized in *E. coli* for its significant role in maintaining membrane integrity. The combined correlation of Hfq and OmpA to cellular envelope stress in high salt conditions suggests a distinct salinity response by *Cp*34H (Fig. 2.8C-D, Table 2.) (Choi and Lee, 2019). Additionally, psychrophiles and halophiles have contradictory strategies for the production of branched-chain fatty acids (Ventosa *et al.*, 1998; Chattopadhyay, 2006)(LivG, LivK, LivF, Fig. 2.7), therefore *Cp*34H cells exposed to both low temperatures and high salinities must combine strategies, as was observed here. Due to the increased expression of fatty acid biosynthesis proteins (*i.e.,* LptD, LptB, AccB, PaaK, Fig. 2.8), we also propose that fatty acids be explored for use as potential biomarkers for high salinity candidates. For example, while psychrophiles are documented to increase branched-chain fatty acids (Chattopadhyay 2006), halophiles decreased branch-chains in favor of cyclopropane fatty acids and unsaturated fatty acids (Ventosa *et al.*, 1998). Because the presence of branched-chain C15 monounsaturated fatty acid is so unusual amongst halophilic bacteria, Skerratt *et al*. (1991) proposed it as a taxonomic biomarker for specific strains isolated from Antarctic lakes.

General Nutrient-Limitation Response. Entrapment in sea ice brine inclusions would create a closed system, naturally creating a nutrient limited environment. Both ASW^{-nutr} and brine^{-nutr} cells displayed increased abundance of proteins involved in chemotaxis-directed motility (MotA, Mcp, Fig. 2.8b, d), consistent with previous microscopic evidence that cells actively follow nutrient gradients (Junge *et al.*, 2003; Showalter and Deming, 2018). Nutrient-limited treatments also exhibited increases in pilus proteins responsible for *Cp*34H nutrient acquisition (PilP, PilC). Type IV pilus proteins are present in a diverse array of bacteria (Craig *et al.*, 2019), aiding primarily in the acquisition of nutrients (Gold *et al.*, 2015), suggesting that once detection and abundance thresholds are quantified, they may serve as good biomarkers of generalized nutrient limitation in field work. Additionally, type IV pili possess a twitching motility as well as the ability to sense and adhere to solid surfaces, creating the potential for ice-affine surface adaptations or adhesion to inorganic solutes present within brine channels (Collins, 2009; Baker *et al.*, 2017; Craigz *et al.*, 2019). An increased abundance of amidohydrolase family proteins (CPS_3630, 3629, 0854), which have been associated with increased growth of organisms in nutrient-limited conditions (Cezairliyana and Ausubela, 2017), suggests these hydrolytic enzymes could serve as a general biomarker for low nutrient bioavailability.

General Temperature-Dependent Response after four months at -5°C.

The general response to a decrease in temperature over four months relative to the control culture is primarily characterized by significant changes in carbon metabolism, emphasizing its central role in cell survival as temperature decreases (Czajka *et al.*, 2018) (Appendix A.8-11). Proteins in the glycine cleavage system were increased in *Cp*34H, which allowed cells to utilize products, e.g., glycine betaine, as compatible solutes or as bioavailable sources of carbon and nitrogen (Andreesen, 1994; Methe *et al.*, 2005; Gunde-Cimerman *et al.*, 2018). Enzymes involved in the detoxification of formaldehyde were identified (i.e., S-formylglutathione hydrolase (FrmB), S-(hydroxymethyl)glutathione dehydrogenase (FrmA), alanine-glyoxylate transaminase (AGXT)), highlighting the importance of methane and other small 1-C substrates as a primary carbon source utilized through the folate biosynthesis pathway (i.e., aminomethyltransferase (GcvT), 5-methyltetrahydrofolate-homocysteine methyltransferase (MetH), glycine hydroxymethyltransferase (GlyA)) at the four month time point. In nutrient-replete bacterial incubations, it has previously been noted that these metabolisms are initiated after macromolecular carbon sources are depleted (Mikan *et al.*, 2020). That said, all conditions possess enzymes for the breakdown of glucose through the Entner-Doudoroff pathway (glucokinase (GLK), glyceraldehyde 3-phosphate dehydrogenase (GAPDH); Appendix A.8-11), indicating either that glucose remained in the system, or these enzymes are retained by *Cp*34H and express slow turnover rates. This same network of carbon metabolism enzymes produce formate, methanol, NADH and hydrogen, all byproducts from precursors previously recognized/modeled to increase in ice veins (Price 2000). Similarities in carbon metabolism could be due to the utilization of choline as a substrate for energy production and biosynthesis pathways (Collins and Deming, 2013; Czajka *et al.*, 2018). Choline, a B vitamin and preferentially utilized osmolyte by *Cp*34H in sea ice brines (Firth *et al.*, 2016), would be expected to be present at adequate levels in conditions receiving nutrients due to the inclusion of yeast extract, but it appears that even cells lacking nutrients rely on this pathway for metabolism when other stressors are present.

MetE, a vitamin B12-independent methionine synthase (Mordukhova and Pan, 2013), had the highest $\log_2$ fold change of any protein detected in all conditions (Fig. 2.8a-d), with the exception of ASW^{+nuts}. Additionally, all conditions except ASW^{+nutr} also exhibited a significant increase in nitrous oxide reductase (NosZ, Fig. 2.8b, c), suggesting a buildup of nitrous oxide; this byproduct is a known inhibitor of vitamin B12 metabolism that renders the vitamin inactive (Pema *et al*., 1998; Stockton *et al*., 2017), likely generating cellular B12 deficiencies. Vitamin B12, also known as cobalamin, is essential for a variety of metabolic pathways, such as 1-carbon metabolism and methionine metabolism. B12 has also been thoroughly studied for its thermodynamic stability as a biomolecule that can only be synthesized by microorganisms, yielding broad implications for origin of life studies (Eschenmoser, 2011). Despite its high demand by cells, B12 is often sparsely found in marine environments such as the ocean, where it is largely depleted in surface waters and only measurable in femtomolar to low picomolar concentrations (Menzel and Spaeth, 1962; Sañudo-Wilhelmy *et al*., 2012; Ellis *et al.*, 2016). Supporting our conclusion that these cells were challenged by inhibition or depletion of vitamin B12, vitamin B transporters increased in abundance in all -5°C conditions (BtuB; Fig. 2.7, 5a-d). The lack of intracellular vitamin B12, whether through a cellular inability to acquire or through inactivation via nitrous oxide, appears to cause a number of consequences in the cell.

Sulfate reduction was activated in all -5°C conditions compared to the control, demonstrating a cold-adapted sulfate metabolic response in *Cp*34H. Sulfate reduction has been previously noted to be more active in psychrophilic marine bacteria compared to warmer-adapted marine bacteria (Knoblauch *et al*., 1999) and general carbon turnover rates in Arctic marine sediments are comparable to temperate climates due to their activity (Sageman *et al*., 1998). This, however, is the first analysis to reveal that sulfate reduction enzymes in a psychrophilic bacterium increased in response to lower subzero temperatures (*i.e.,* -1°C to -5°C) regardless of nutrient input (CysC, CysJ, CysI, CysH, CysN, CysD, Fig. 2.8a-d). *Cp*34H has been previously noted to have an enhanced genomic potential for sulfur metabolism, possessing both the ability to oxidize intracellular sulfur and acquire sulfur extracellularly through sulfatase production (Methe *et al*., 2005). Further, sulfur metabolism could be driven by the increase in cysteine and methionine metabolism experienced by cells in all conditions as a general response to a decrease in temperature. Complete activation of the methionine biosynthesis pathway was observed in all conditions (MetA, MetB, MetC, MetE, MetH, MetK, MetL, Appendix A.8-11). Substitutions of various amino acids for methionine, previously observed to be increased in psychrophilic proteins, could stabilize the metal-binding domains of protein structures under low temperature conditions (Harvilla *et al*., 2014; Maayer *et al*., 2014). Because the -5°C treatments were incubated for four months, increases in sulfate reduction may be a result of oxygen limitation. Calculations based on the growth rates tracked with [3H]-leucine suggest that the oxygen is reduced by the end of the first week in the ASW^{+nutr} incubation (Table 2.3), further supporting the presence of anaerobic metabolic pathways being captured by the proteome at month four. These findings suggest that when overwintering, psychrophilic sulfate reducing enzymes can increase in abundance as temperatures drop, potentially influencing global sulfur cycles (Karr *et al*., 2005). Additionally, heterotrophic bacteria in anaerobic conditions of evolving glacial lakes also exhibit active sulfate reduction, suggesting these metabolic pathways may have increased relevance in other planetary lakes, such as on Titan (Parro *et al*., 2018).

2.5.5 SURVIVABLE AND DETECTABLE POLYPEPTIDE BIOMARKERS

These are the first incubations of *Cp*34H in supercooled ASW and equilibrium brine where activity, culturability, and proteomics were simultaneously monitored. Based on previous work that indicates activity at -10°C in ice samples for extended periods of time (Junge *et al*., 2006; Nunn *et al*., 2015), visual evidence that our incubations maintained cell numbers at -10°C in all treatments, and results presented here that detected protein profiles at -10°C significantly differed from controls support the hypothesis that cells in extreme conditions on an icy world could retain specific polypeptide signatures for life at these conditions. This study represents the first statistical evaluation of detected bacterial peptide sequences to determine if short peptide motifs (3-4 amino acid long) were specifically synthesized and/or retained as a result of the high salinity or low -10°C temperature. Evidence of the enrichment of short peptides in icy-world-analog incubations here on Earth provide environment-specific context for future discoveries of detectable biomarkers for life (extant or extinct) on other icy worlds. Currently, no multi-amino acid-based biomarkers for life exist, but here we provide a list of potential candidate polypeptides that could be targeted using high mass accuracy lander-based or in-orbit mass spectrometers (for review see Arevalo *et al*., 2020). Candidate biomarkers must be detectable by mass spectrometry and include polypeptide sequences of 2-8 amino acids in length; for statistical reasons, our analysis focuses on polypeptides of 3-4 amino acids in length.

The most extreme experimental condition, -10°C brine, representative of the response of terrestrial organisms to highly saline ice, especially under low-nutrient circumstances, is of distinct astrobiological relevance. Our analysis identified over 534 k-mers that were found to be significantly increased in *Cp*34H (Fig. 2.10). These peptides are specifically enriched in the amino acids Ala, Val, Glu, and Gln. Over 80 amino acids have been detected in carbonaceous meteorites (Pizzarello and Holmes, 2009; Pizzarello *et al*., 2006; Georgiou, 2018), and the most dominant alpha-amino acids discovered in meteorites follows a distribution of Gly>Ala>Glu>Asp>Ser

(Davila and McKay, 2014), notably containing two of the dominant amino acids detected to be enriched in k-mers (Ala and Glu) at the more extreme conditions studied here. Alanine is the second smallest of the 20 alpha-amino acids and is used in synthetic peptide design as a small spacer amino acid, providing non-charged flexibility (Walters and DeGrado, 2006 and citations within). The increased abundance of alanine in both 3- and 4-mers in both the temperature- and brine-specific enrichment analyses also suggests that alanine provides a flexible spacer to connect functional or acidic residues with higher binding affinities or catalytic propensities (Bartlett *et al*., 2002). Valine, another frequently encountered amino acid in meteorites (Cobb and Pudritz, 2014), is noted to have a short hydrophilic side chain, providing valine-rich peptides with greater flexibility and activity at low temperatures.

Georgiou (2018) proposed that different combinations of amino acids generate specific catalytic units that may be enriched in life on earth, thereby increasing the probability of being discovered in primitive extraterrestrial life forms. Examinations of catalytic sites within enzymes suggest that two of the amino acids revealed to be enriched, retained, and detected in *Cp*34H k-mers have key chemical groups that can increase catalytic potential (Gln: amide, Glu: carboxylate; Bartlett *et al*., 2002). Previous analyses of >190 catalytic enzymes' three-dimensional structures and functional sites identified six common catalytic triads (*Thr-His-His, Ser-His-Asp, Ser-His-Glu, Asp-Thr-Lys, Thr-Lys-Asp, Lys-Glu-Lys*; Gutteridge and Thornton, 2005). None of these 3-mers were identified in our analysis of enriched k-mers with the exception of *Ser-His-Asp* (SHD), a catalytic site of hydrolase serine proteases, identified in one temperature-specific 4-mer *Ser-His-Asp-Arg* (SHDR). The k-mer identified to be enriched in the temperature-, salinity-, and brine-10°C- specific responses, *Lys-Asp-Ala* (KDA), is present in 189 of the predicted proteins from *Cp*34H. Although the sequence is short, it may enhance the functionality of these proteins in subzero temperatures as it has an acidic residue with a basic amide functional group (K), a carboxylate functional group (D), and a small non-polar spacer amino acid (A). In colder temperatures, where protein structures can become rigid and inflexible, the small amino acid spacer would provide a wider opening for other molecules to bind with adjacent catalytic amino acids.

Although individual amino acids have been detected in meteorites, and numerous models and predictions based on physical properties of known enzymes have been made to theorize on novel life detection strategies, this report is the first to take detectable polypeptide biosignatures from psychrophiles incubated in extreme environmental conditions with the goal of identifying common, short, repeating patterns that might be found elsewhere in the Solar System. These peptides could be included in immunoassay-based life detector chips as proposed by Parro *et al*., 2018, or directly detected and verified using tandem mass spectrometry methods (*e.g.,* Arevalo *et al*., 2020). Since peptides fragment in a predictable manner, generating a forward and backward conformational sequence fingerprint with b- and y-ions, high mass-accuracy mass spectrometry methods have a distinct advantage for detecting the proposed list of k-mers. Furthermore, for off-planet missions to detect life, mass spectrometers have flight heritage as they have been tested on missions to Mars dating back to the 1970s Viking mission (Anderson *et al*., 1972). Mass spectrometers can provide sensitive and quantitative analyses of the chemical composition of a range of planetary matrices, decipher isotopic, elemental, and molecular abundances in an unbiased manner, and, with more recent advancements, allow for and direct complex experiments in the gas phase using tandem mass spectrometers, making them capable payload instruments for helping confidently detect biomarkers of life (*e.g.,* Arevalo *et al*., 2019). For example, using tandem mass spectrometry, mission control will be able to design experiments that first accumulate predetermined ions (*i.e.,* specific mass to charge ratios), while ejecting all other ions. This increases the signal to noise ratio, improves sample-specific detection limits, and yields higher resolution fragmentation spectra (*i.e.,* MS2) for downstream interpretation. Mission-ready high mass-accuracy Orbitraps are already under development, showing great promise for detecting 3-4 amino acid long polypeptides (Arevalo *et al*., 2018). With known polypeptide targets, such as we provide here, pre-programmed parallel reaction monitoring mass spectrometry (PRM-

MS) methods can be designed to target, isolate, and fragment desired peptide masses of interest in the gas phase. The sensitivity of these MS based-methods have demonstrated detection of attomolar levels of many peptides within complex matrices (Gillette and Carr, 2013; Glukhova *et al*., 2013). Additionally, mass spectrometers would allow for on-the-fly organic or inorganic target selection from ground control so that these instruments can discover unexpected signals in a targeted-assay-like manner as new information is gained while in orbit.

The list of k-mers provided here reveals important trends in amino acid sequences and should be further refined as they are explored in other organisms. Although peptides have been subjectively placed low on the "evidence for life" scale, directly correlating specific features that are enriched in terrigenous systems similar to those anticipated off-Earth may allow us to link heritable traits found as a molecular signal to specific environmental responses (Neveu *et al*., 2018). Direct correlations of k-mers featured here to high salinity or low temperature would strengthen evidence for the existence of life elsewhere if discovered off-planet. In addition to performing additional tests on these k-mers in other organisms, these peptides can be used in other low biomass, subzero or saline environments to determine their universality as a life detection biomarker. Isotopically-labeled synthetic duplicates should be made as laboratory standards, and limits of quantification and detection should be established with mass spectrometry methods in a variety of matrices, including rock with potential relevance to the joint NASA and ESA future Mars rock and soil sample return campaign.

2.6 CONCLUDING REMARKS

Conditions in the most extreme cold environments on Earth provide a useful analog to study other icy worlds, where survival would require organisms to maintain adaptations to extreme low temperatures and high salinities. Long-term survival in a matrix of extreme conditions can offer valuable clues to the metabolic strategies employed by microbial life-forms enduring similar combinations of stresses as the experimental variables explored in this study. Additionally, as specific proteins were uniquely identified and/or significantly increased or decreased in direct correlation to isolated variables, peptides from those proteins could be selected as biomarkers for extant life here and elsewhere.

Understanding the specific mechanisms by which extremophiles adapt to their surroundings is important for the progress of environmental, technological, and astrobiological advancements. Here we report how environmental variables of low temperature, high salinity, and absence of nutrients independently affect functional proteomic responses in the context of culturability and metabolic activity of *Cp*34H, particularly highlighting the impact of carbon, nitrogen, and fatty acid metabolism. Finally, we propose the use of polypeptide biomarkers in advancing astrobiological searches for life on other icy worlds, providing a suitable list of candidate biomarkers to investigate in future studies.

Chapter 3: HARMFUL ALGAL BLOOMS PRECEDED BY A PREDICTABLE AND QUANTIFIABLE SHIFT IN THE OCEAN MICROBIOME

Chapter 3 is adapted with minimal modification from:

Mudge, M.C., Riffle, M., Chebli, G., Plubell, D., Rynearson, T., Noble, W.S., Timmins-Schiffman, E., Kubanek, J., Nunn, B.L. (2024) Harmful algal blooms preceded by a predictable and quantifiable shift in the oceanic microbiome. Nat Commun. *In review.*

M. C. Mudge contribution: Study design, sample collection, DNA extractions, sample processing, data analysis and interpretation, and manuscript drafting and editing.

3.1 ABSTRACT

Harmful algal blooms (HABs) have become a worldwide environmental and human health problem, stressing the urgent need for a reliable forecasting tool. Dynamic interactions between algae, including harmful algae, and bacteria play a large role regulating water chemistry. Free-living bacteria quickly respond to small physical and/or chemical environmental changes by adjusting their proteome. We hypothesize that this response is detectable at the peptide level and occurs before rapid phytoplankton growth characteristic of harmful bloom events. To characterize the microbiome's physiological changes preceding bloom onset, we collected and analyzed a high-resolution metaproteomic time series of a free-living microbiome in a coastal ecosystem. We developed an agnostic, comprehensive, and statistically robust framework to query the microbial metaproteome for bacterial peptides that undergo a non-reverting state shift in abundance ~24 hours prior to HAB onset. Twelve candidate HAB biomarkers were confirmed to be detectable, quantifiable, and correlated across two pre-bloom periods. This study proposes and verifies that

the changing molecular signals within the ocean microbiome can be used as diagnostic biomarkers for predicting HAB initiation, potentially mitigating detrimental algal bloom outcomes in the future.

3.2 INTRODUCTION

Harmful algal blooms (HABs) are sudden accumulations of photosynthetic phytoplankton that cause adverse effects to aquatic ecosystems (“What Is a Harmful Algal Bloom?”, National Oceanic and Atmospheric Administration). These blooms pose serious threats to both human health and the economy (Berdalet *et al.,* 2015) with negative consequences on fishing, tourism, and real estate industries as well as mass mortalities of coastal marine mammals, seabirds, and fishes. Harmful effects caused by bloom-forming species are primarily attributed to the production of highly toxic compounds such as neurotoxins (Anderson *et al.,* 2021); that said, there are many non-toxic HAB species that achieve remarkably high cell densities that can block sunlight to deeper waters, alter local water chemistry (*e.g.,* pH, nutrients)(Brooks *et al.,* 2016), generate hypoxic or anoxic zones (Turley *et al.,* 2022), and/or obstruct fish gills, all potentially leading to a swift rise in mortalities (Rolton *et al.,* 2022). A variety of human-made and environmental perturbations are capable of initiating these blooms (Glibert and Burkholder 2018); however, in the majority of bloom events, the specific underlying cause for HAB initiation is unknown. In general, two fundamental goals of HAB research are (1) to determine what factors lead to the dominance of HAB-forming species and (2) to predict bloom onset, enabling warning and mitigation strategies to combat the increasing frequency and intensity of HAB events globally (Barton *et al.,* 2016).

Interactions between phytoplankton and bacteria have been shown to play a paramount role in shaping the dynamics of nutrient availability, biogeochemical cycles, and microbial diversity and function in marine ecosystems (Seymour *et al.,* 2017, Mayali 2018; Muratore *et al.,* 2022). In the specific context of HABs, tight associations have been observed within the microbiome as phytoplankton advance through bloom cycles, with bacteria playing a deterministic role in the outcome of HAB progression and algal byproducts (Dong *et. al.,* 2014; Coyne, Wang, and Johnson 2022; Kuhlisch *et al.,* 2024). Additionally, certain consistently abundant taxa have been noted as commonly associated with different phytoplankton species (Hattenrath-Lehmann and Gobler 2017; Deng *et al.,* 2022; Fortin *et al.,* 2022), lending precedent to the idea that community structure and function are driven by specific associations (Grossart *et al.,* 2005). The influence these organisms have on each other's biological potential creates a recognized opportunity (Gajardo *et al.,* 2023) to utilize rapid bacterial active responses to individual and interacting chemical, physical, and biological factors as an in situ predictor for phytoplankton dynamics.

Proteomic techniques are capable of capturing this active functionality of cells through the measurement of peptides, the component pieces of proteins that are tightly regulated and rapidly modified in response to minute environmental perturbations (Maier *et al.,* 2011; Calado 2022). Recent advances in mass spectrometry (MS) techniques and methods have enabled highly accurate quantification of peptides in a variety of sample types across a wide dynamic range (Brunner *et al.,* 2022). Additionally, as the development of peptide-based biomarkers for diagnosing human health (Pandey *et al.,* 2021) is gaining momentum in medical research, a range of potential molecular or antigen rapid tests could be deployed via citizen science (Agarwal *et al.,* 2022) or autonomous vehicle monitoring for environmental applications (Wang *et al.,* 2024).

The high-resolution time series presented here captured the biological and chemical characteristics of a dynamic coastal ecosystem during two pre-bloom periods of the HAB-forming diatom *Chaetoceros socialis*. To interrogate the microbiome for biomarkers capable of predicting a HAB >24 hours in advance, a metaproteomic analysis was performed to quantify bacterial peptides collected every four hours. A multistep framework was developed for selecting and statistically validating potential biomarker peptides using a hidden Markov model (HMM), a statistical approach that can identify when the underlying state of the data undergoes a non-reverting shifts in time series data. Our investigation revealed 314 potential biomarkers for predicting bloom onset. Targeted validation of peptides from a second pre-bloom period with a different chemical profile confirmed a subset of 12 peptides as quantifiable, non-reverting biomarkers. These findings provide evidence towards the use of the bacterial microbiome to generate biomarkers for predicting future HABs.

3.3 METHODS

Sample collection and water column analysis Sample collection methods are detailed in Nunn *et. al.*, 2024. Briefly, field sampling was conducted in East Sound, WA, USA from May 27 - June 18, 2021, by pumping water from 100 m offshore at a constant depth of 2 m and flow rate of 5 L min^{-1}. Whole water was filtered through a 200 µm nylon mesh followed by three subsequent filters at 100 µm, 10 µm, 1 µm before final collection onto a 0.22 µm polyethersulfone (PES) 47 mm filter. Final metaproteomic samples were stored long term in -80°C on 0.22 µm polyethersulfone (PES) filters. Additional samples collected included nutrients, flow cytometry, prokaryotic DNA/proteomics, eukaryotic DNA/RNA/proteomics, and dissolved metabolites. Water chemistry was also monitored at the collection site with a YSI Exo1 Sonde probe capturing chlorophyll *a* concentration, salinity, temperature, dissolved oxygen, and pH every 10 minutes.

Metagenome assembly and database generation DNA extraction was performed according to methods detailed in Nunn *et. al.*, 2024 on samples from all time points collected. All 132 samples were combined to generate a pooled sample with equal amounts of DNA (50 ng) from each sample for whole genome sequencing (WGS) using 150 bp paired-end Illumina sequencing (NextSeq 500 Mid Output v2 Kit - 300 cycle) through the Northwest Genomics Center (NWGC) at the University of Washington. DNA library preparation, quality controls, and WGS were performed by the NWGC. The resulting FASTQs were assembled into a metagenome using MEGAHIT (Li et al. 2015). Samtools (Danecek *et al.,* 2021) 1.14 was used with Python (Van Rossum and Drake 2009) 3.7.7 to filter only mapped reads, and Prodigal (Hyatt *et al.,* 2010) 2.6.3 was used to convert contigs into a FASTA file consisting of 3,457,818 proteins (PRIDE identifier PXD052873).

Proteomic sample processing Samples collected for microbiome proteomic analysis (collected on 0.22 µm PES) were manually agitated within a 3x4 cm Ziplock baggie with 100 µl S-Trap solubilization buffer (5% SDS, 50 mM TEAB, 2 mM $MgCl_2$, 1X HALT protease and phosphatase inhibitors in nanopure water) to ensure equal spreading of the buffer across the filter surface. The cell lysate was collected via pipette in a 2 mL microfuge tube. This was followed by additional rinse of the filter with 100 µl nanopure water followed by manual agitation and immediate transfer of solution to pool fractions in the 2 mL tube. Two additional rinses of the filter with 100 µl nanopure water without manual agitation were performed and the solution added to the pool. The resulting 400 µl whole cell solution was then sonicated (Branson 250 Sonifier; 20 kHz, 30 × 10 s on ice). Samples were then evaporated in a SpeedVac (ThermoFisher Scientific SRF110) at room temperature to reduce the volume to 100 µL (5% SDS). Protein concentrations for each sample were quantified in triplicate using a Bicinchoninic Acid protein assay kit (Pierce Thermo Scientific) using a microplate reader. Enolase protein standard (0.16 µL 100 ng $µl^{-1}$ enolase per 1 µg protein) was added to each sample to track digestion efficiency. Each sample (20 µg protein) was treated with benzonase to degrade DNA (0.5 µL 250 unit $µl^{-1}$ 10 minutes at 95°C), and proteins were reduced with 20 mM dithiothreitol for 10 minutes at 60°C followed by a four minute cool down to room temperature, alkylated with 40 mM iodoacetamide for 30 minutes in the dark, acidified to pH <2 (1.2% aqueous phosphoric acid), and then processed on an S-trap micro column according to manufacturer's recommendations (Protifi S-Trap micro protocol). Proteins were digested with Promega modified trypsin (2 µg for 1:10 ratio, 1 hour 37°C). Resulting peptides were evaporated to dryness and resuspended in 2% acetonitrile (ACN), 0.1% formic acid with final concentration of 0.5 μg protein μl^{-1}. Prior to analysis on the mass spectrometer, each sample received 50 fmol ug^{-1} protein of Pierce Retention Time Calibration (PRTC) as an external standard to monitor MS performance across experiments.

Proteomic mass spectrometry overview All mass spectrometry analyses were performed on an Orbitrap Lumos Fusion Tribrid instrument with an inline EASY-nLC 1200 (ThermoFisher Scientific). Reverse phase chromatography was achieved using a manufactured PicoTip fused silica capillary column (75 μm i.d., 40 cm long) packed in-house with C18 beads (Dr. Maisch ReproSil-Pur; C18-Aq, 120 Å, 3 μm) and maintained at 50°C. The analytical column was fitted with a 4 cm long, 150 μm i.d. in-house packed kasil-frit precolumn (Dr. Maisch ReproSil-Pur; C18-Aq, 120 Å, 3 μm). Peptides were eluted using an acidified (formic acid, 0.1% v/v) water-acetonitrile gradient (2-45% acetonitrile). Sample analyses on the MS were randomized to reduce batch effects and quality control (QC) peptide mixtures were analyzed every sixth injection to monitor chromatography and MS sensitivity. Each sample was analyzed with data-dependent acquisition (DDA) and data-independent acquisition (DIA) in random order. Skyline was used to evaluate the consistency of the resulting peak areas and retention times of QC standards through all analyses (MacLean *et al.,* 2010). The DDA MS data are deposited to the Proteome Xchange Consortium via the PRIDE partner repository with the dataset identifier PXD052873. The DIA and PRM MS data are deposited to Panorama Public (Sharma *et al.,* 2014; 2018): (https://panoramaweb.org/Orcas-Island-Time-Series.url).

DDA analyses Samples from June 12, 2021 (17:00) to June 18, 2021 (13:00) were analyzed on the mass spectrometer operating in DDA mode. The top 20 most intense ions were selected for MS2 acquisition from precursor ion scans of 400–1200 *m/z*. Centroid full MS resolution data was collected at 70,000 with AGC target of 1 $\times 10^6$ and centroid MS2 data was collected at resolution of 35,000 with AGC target of 5×10^4. Dynamic exclusion was set to 10 seconds and +2, +3, +4 ions were selected for MS2. Comet 2022.01 (Eng, Jahan, and Hoopmann 2013; Eng *et al.,* 2015) was used to search the DDA files; parameters included: reverse concatenated sequence database search, 10 ppm precursor mass tolerance, semi-tryptic specificity with 3 allowed missed cleavages, cysteine modification of 57 Da (resulting from the iodoacetamide) and modifications on methionine of 15.999 Da (oxidation). Percolator (version 3.02)(Käll *et al.,* 2007; The *et al.,* 2016) was used to select peptides with 1% FDR. An open source Nextflow workflow (https://github.com/mriffle/nf-filter-fasta) was used to reduce the metagenomic FASTA from 3,457,818 proteins into a smaller DDA-refined database representing all possible proteins from the peptides detected in the samples for proteomic interpretation. These peptides were mapped back to the full metagenomic FASTA to compile a database of all proteins with at least 1 peptide identified by the workflow. The resulting DDA-refined database contained 36,429 protein sequences.

DIA analyses Samples from June 12, 2021 (17:00) to June 18, 2021 (13:00) were analyzed on the mass spectrometer using DIA mode. For each DIA MS experiment, 4 µg digested peptides with 7 fmol PRTC were analyzed. Centroid full MS resolution data was collected at 120,000 and MS2 data was collected at a resolution of 15,000 with an AGC target of 4×10^5 across isolation windows of 8 *m/z*. Prosit (Gessulat *et al.,* 2019) was used to process the DDA-refined database using an HCD intensity and ion retention time prediction model according to methods from Searle *et. al.,* 2020(Searle et al. 2020). Six gas phase fractionations (GPFs) spanning 100 *m/z* for each analysis were completed with narrow isolation window (4 *m/z*) to generate a chromatogram library using Encyclopedia (Searle *et al.,* 2018) v. 1.4.01 following Pino *et. al.*, 2020. The 36 individual time series MS experiments were searched against the resulting chromatogram library using Encyclopedia v. 1.12.31. Encyclopedia-exported results were visualized using Skyline-daily version 22.2.0.527, including quality assurance that every peptide contained at least three quantifiable transitions and external standards (PRTC peptides) had a coefficient of variation below 20%. Quantitative values for each peptide were generated using Skyline which calculated the area under the curve (AUC) for each peptide after total ion current (TIC) normalization using Encyclopedia-inferred peak boundaries.

Taxonomic and Functional Annotations. MetaGOmics (Riffle *et al.,* 2017) software (version 0.2.0; https://www.yeastrc.org/metagomics) identifies the lowest common ancestor on each peptide's taxonomic and functional Gene Ontology (GO) assignment using a provided site-specific metagenomic-generated protein sequence database. Here, we used the DDA-refined database as our background proteome and provided a complete list of all peptides identified at each timepoint with their respective peptide quantities (DIA-MS). The MetaGOmics reports were used to annotate peptides to their respective GO terms and taxonomic classes.

3.3.1 Framework for selecting peptides

Phase 1A: Data dimensionality Reduction- PCA. Quantitative peptide AUC values were first median normalized and log transformed prior to performing the PCA. PC1 was driven by missingness of data across the 36 timepoints and was not included in further exploratory analysis. PC2 and PC3 were separated by day of collection, identifying a critical time point of change. A Mann-Whitney test was used to test for significance between the first and second half of the timeseries. The top 10% of peptides driving variation across PC2 were selected as candidate peptides.

Phase 1B: Taxonomic trends in summed peptide abundances. Peptide abundances resulting from each class as annotated by MetaGOmics were summed at each timepoint and plotted through time. A Spearman's correlation coefficient rho (ρ) was then calculated (Table 3.2; Fig. 3.9) comparing the summed quantitative peptide abundance values for each class and ranked time. All peptides from the bacterial classes with a correlation coefficient $\rho >= |0.5|$ were selected for further development and testing as biomarkers (*i.e.*, Flavobacteriia, Gammaproteobacteria, Verrucomicrobiae, and Opitutae). Alphaproteobacteria was additionally added to the analysis as it represented the most abundant taxonomic class.

Phase 2: HMM State-shift time series analysis. A two-state hidden Markov model (HMM) was then fit to the quantitative values associated with each candidate peptide to identify those that undergo a non-reverting state shift >24 hours prior to the BSD (R package depmixS4 (Visser and Speekenbrink 2010) (Fig. 3.12-13). The time series data was log transformed prior to model fitting. The model was specified to have a Gaussian distribution and transition probabilities were assumed to be stationary. Model parameters were estimated using an Expectation-Maximization (EM) algorithm, with a random seed set to ensure reproducibility. The resulting probability matrix was used to determine the latent state of the observations, and the determined state outcomes were plotted through time. Of the 314 peptides determined to undergo a non-reverting state shift, 87 were selected for the next phase of biomarker development.

Phase 3: Matrix-Matched Calibration Curve (MMCC). To determine if the candidate peptides are quantitative by mass spectrometry, a matrix-matched calibration curve (MMCC) was created (Pino, Searle, *et al.,* 2020). MMCC analysis provides the user with the ability to determine a peptide's detectability within a complex matrix and whether the peptide yields a linear response when analyzed with targeted parallel reaction monitoring (PRM) methods in a dilution series. A pooled sample consisting of 2 µg aliquots of digested peptides from 36 samples from our time series as a representative water microbiome. This microbiome sample was serially diluted with a matched matrix of peptides resulting from the marine bacterium *Colwellia psychrerythraea* 34H (Methé *et al.,* 2005). The resulting dilution series consisted of six microbiome-to-matrix ratios that spanned the expected dynamic range of targeted peptides: 0%, 0.3%, 3%, 50%, 70%, 100% microbiome.

PRM analyses on MMCC The targeted peptide list was generated using the Skyline analysis of the DIA data. Peptides were trapped and separated on a near identical chromatography setup detailed earlier with the exception of the analytical column length (30 cm). Mass spectrometry experiments were completed with 2 µg digested peptides with 150 fmol PRTC. Peptides were eluted using the same gradient as DDA/DIA experiments followed by a 30 minute wash with 80% ACN and 10 minute re-equilibration with 3% ACN. Targeted ions were selected for MS2 acquisition from precursor ion scans of 400–1000 *m/z* with an isolation window of 1.6 *m/z,* flow rate of 300 nL/min, and resolution of 30,000. Centroid full MS and MS2 data was collected with an AGC target of 4×10^5. Data was analyzed with Skyline (MacLean *et al.,* 2010) software to identify peptide peaks and chromatographic time boundaries. Skyline-calculated peptide quantification from AUC were exported and analyzed with R following *Pino et al. 2020.* A total of 27 peptides were identified to yield linear responses via the MMCC with resulting limit of detection (LOD) >1e6.

Biomarker Peptide Validation To validate the final 27 candidate peptides as predictive HAB biomarkers, PRM MS analyses were performed on the set of samples collected from May 28, 2021 (01:00) to June 3, 2021 (01:00) using the same target peptide list. MS experiments were completed on 2 µg peptides with 150 fmol PRTC external standard. Skyline (version 23.0.9.187) was used to visualize the peptide peak boundaries, which were visualized and manually adjusted based on the parameters established by the MMCC methods. The resulting AUC report was exported from Skyline and evaluated in R to analyze the temporal trends for each peptide. A two state HMM was fit to each peptide to select for those that undergo a non-reverting state shift, resulting in the final panel of 12 successful biomarker peptides. A Spearman correlation analysis was performed on the 12 peptides to compare peptide abundances across the discovery and validation data sets and assess for consistency of pre-bloom temporal trends. Taxonomic and functional annotations were assigned to each of the 12 peptides by first mapping each peptide back to its respective parent proteins within the DDA-refined database. Subsequently, a standard protein search was performed against the publicly available NR database using the Basic Local Alignment Search Tool (BLAST; 2/6/2024).

3.4 RESULTS

3.4.1 Field-sampling captures two pre-bloom initiation periods with divergent nutrient profiles

A high-resolution time series of a coastal bacterial microbiome was acquired from May 27 - June 18, 2021 in East Sound, WA, USA, capturing samples every 4 hours across one full HAB cycle and the pre-bloom dynamics of a second HAB (Fig.3.1; 3.2). On-site microscopy determined that the eukaryotic fraction of the ecosystem was dominated by *Chaetoceros socialis* (SoundToxins Phytoplankton Monitoring Network). Two bloom initiations of *C. socialis* were identified from the detailed chlorophyll fluorescence trace (Fig. 3.1b, c). Calculated bloom start dates (BSD)(Brody, Lozier, and Dunne 2013) were May 31 at 23:00 and June 18 at 11:00 (Fig. 3.3). Cell counts of nanoeukaryotes and picoeukaryotes confirmed the May 31st BSD (Fig. 3.1d), while Sound Toxin microscopy and satellite images confirmed the presence of a HAB on June 19 [SoundToxins Phytoplankton Monitoring Network](Fig. 3.4). The nutrient conditions over the two pre-bloom initiation periods revealed divergent chemistries (Fig. 3.1h, i). Specifically, the pre-bloom period in May experienced a 20-30 mM increase in silicate and a 15 mM increase in nitrate

~60 hours in advance of the BSD (Fig. 3.1h), coinciding with a 4°C fluctuation in temperature (Fig. 3.1j). In contrast, in the pre-bloom conditions prior to the June BSD concentrations of all nutrients measured fluctuated < 5 mM, revealing a more stable chemical ecosystem (Fig. 3.1i).

As a result of these divergent nutrient profiles in the pre-bloom conditions, the two time series were considered as separate in situ pre-bloom experiments (discovery and validation) allowing the opportunity to characterize each ecosystem independently (Fig. 3.1a). June 12 - 18 was selected as the discovery period because the environmental parameters monitored were more stable (Fig. 3.1f). May 28 - June 2 was selected as the validation period to test candidate biomarkers (Fig. 3.1a).

3.4.2 Discovery phase: identifying peptide biomarkers through quantitative metaproteomics

A new metaproteomic pipeline was developed to 1) confidently identify peptides from this complex microbiome, 2) generate quantitative peptide data, and 3) annotate peptides using a lowest common ancestor approach.

Briefly, data dependent acquisition (DDA) MS data from the June 12 - 18 sample set (36 samples) were searched against the metagenome-generated database (DB) of 3,457,818 proteins (PRIDE identifier PXD052873). DDA analysis yielded 36,429 protein detections from the 9,765 peptides confidently detected, which were used to generate the smaller DDA-refined database for quantitative data independent acquisition (DIA) analysis and interpretation (Fig. 3.5)(Everett, Bierl, and Master 2010; Bern and Kil 2011). The resulting DDA-refined database represents 1% of the original metagenome. Next, the DIA analysis of the discovery sample set detected 4,645 unique peptides. A conservative approach was taken to only include peptides that yielded a quantitative value for > 50% of the 36 time points, yielding a final set of 4,468 peptides for the next phase of our biomarker screen (Fig. 3.6).

Phase 1A: Data dimensionality reduction (PCA) For Phase 1A, a principal component analysis (PCA) was performed on the time series to identify how the quantified peptides varied by time and day of collection. A PCA of all quantified peptides from the DIA analysis of the discovery period revealed that variation across PC1 (proportion of variance = 12.67%) was driven by absence of peptide signals across the time series (Fig. 3.7). Visualization of PC2 (proportion of variance = 9.17%) vs PC3 (proportion of variance = 6.67%) revealed two clusters of data separating along PC2 (Fig. 3.8a). Analysis of the date and time of collection represented by each datapoint revealed that the separation was driven by the date of collection (between June 15-16; Fig. 3.8a), not the time of day of collection as expected. A Mann-Whitney test of PC2 loadings between June 12 - 15 and June 16 - 18 was significant ($p = 1.916e-09$). A total of 447 peptides, representing the top 10% of peptides driving variation across PC2, were selected for Phase 2 analysis.

Phase 1B: Taxonomic trend analysis In Phase 1B, 2,015 of the total 4,468 peptides were annotated with MetaGOmics (Riffle *et al.,* 2017), revealing representation of 29 total bacterial classes in the quantitative peptide dataset (Fig. 3.9). Nearly half of those annotated peptides were from Alphaproteobacteria (956 peptides). The next five most abundant classes were Flavobacteria (431 peptides), Gammaproteobacteria (339 peptides), Verrucomicrobiae (91 peptides), Bacilli (75 peptides), and Opitutae (39 peptides)(Table 3.1). Seven of the 29 taxa showed significant correlation between summed quantitative peptide abundance values and time over the pre-bloom period (Spearman's rho cut-off = $|0.5|$; Fig. 3.8b; Table 3.2). Verrucomicrobiae, Opitutae, Acidobacteriia, and Zetaproteobacteria exhibited a strong positive trend over time ($\rho \geq 0.7$, Fig. 3.8b), with a distinct inflection in abundance at the same June 15-16 changepoint identified from the PCA (Fig. 3.8a, b). The peptide trend from the most abundant class, Alphaproteobacteria, exhibited a weak positive correlation ($\rho = 0.35$, Fig. 3.8b), while peptide trends from Gammaproteobacteria, Flavobacteria, and Epsilonproteobacteria exhibited a

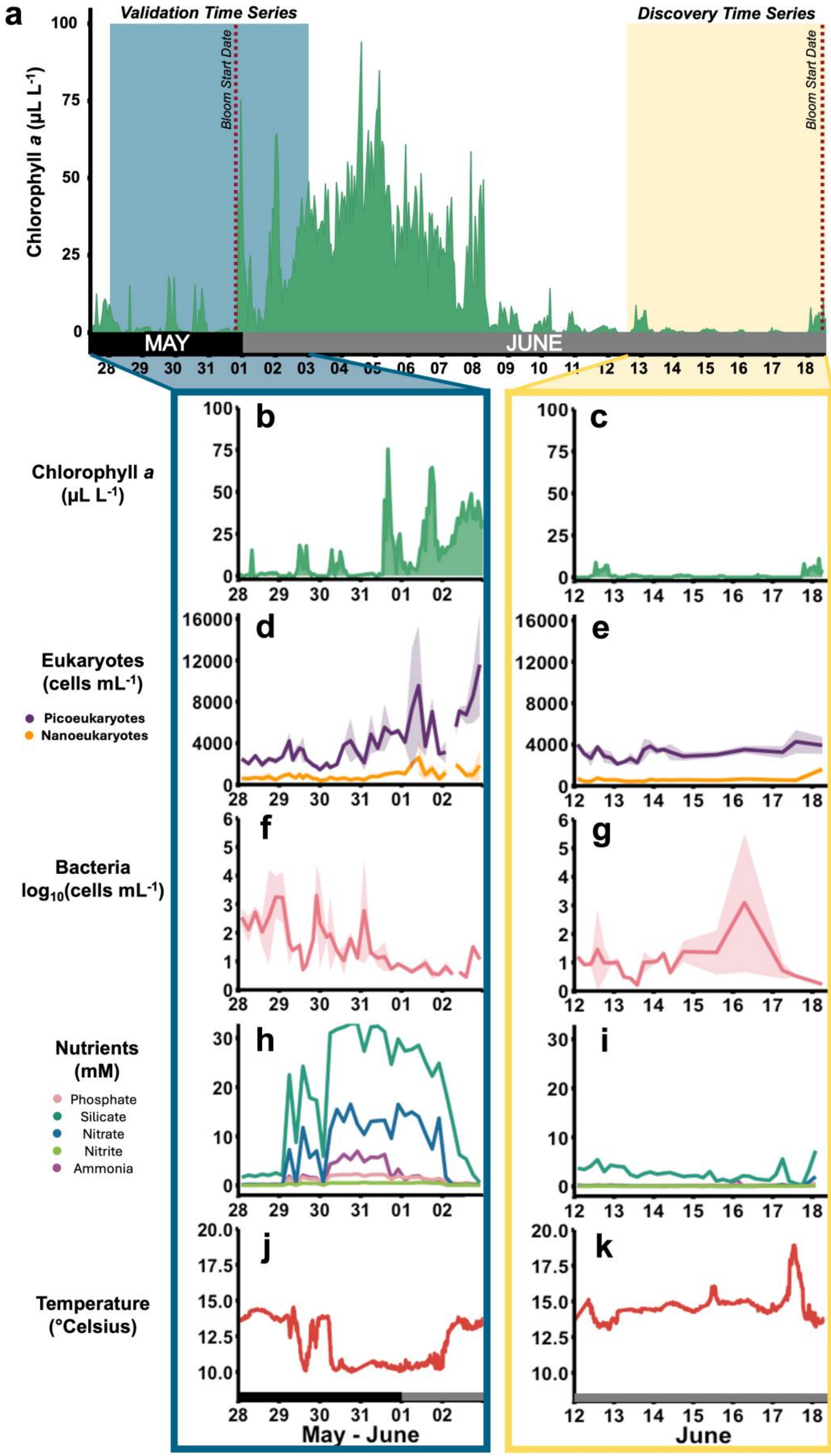

a
Validation Time Series
Discovery Time Series
Bloom Start Date
Bloom Start Date
Chlorophyll a (µL L-1)
100
75
50
25
0
MAY
JUNE
28 29 30 31 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18
b
c
Chlorophyll a (µL L-1)
d
e
Eukaryotes (cells mL-1)
Picoeukaryotes
Nanoeukaryotes
f
g
Bacteria log10(cells mL-1)
h
i
Nutrients (mM)
Phosphate
Silicate
Nitrate
Nitrite
Ammonia
j
k
Temperature (°Celsius)
May - June
June

Figure 3.1. Fieldwork captures pre-bloom dynamics of two *Chaetoceros socialis* blooms. a) Relative phytoplankton levels were captured by chlorophyll *a* measured every 10 min via YSI Exo1 Sonde probe. Time periods selected for discovery (yellow) and validation (blue) are highlighted and bloom start date (BSD) indicated with a dotted red line. b,c) The chlorophyll *a* trace for the discovery and validation datasets are expanded for alignment. d,e) Flow cytometry of nanoeukaryotes and picoeukaryotes, as well as bacterial f,g) cell counts were determined using cell size. h,i) Nutrient concentrations were reported from total nutrient analysis. Temperature j,k) of the water column was captured with a YSI probe.

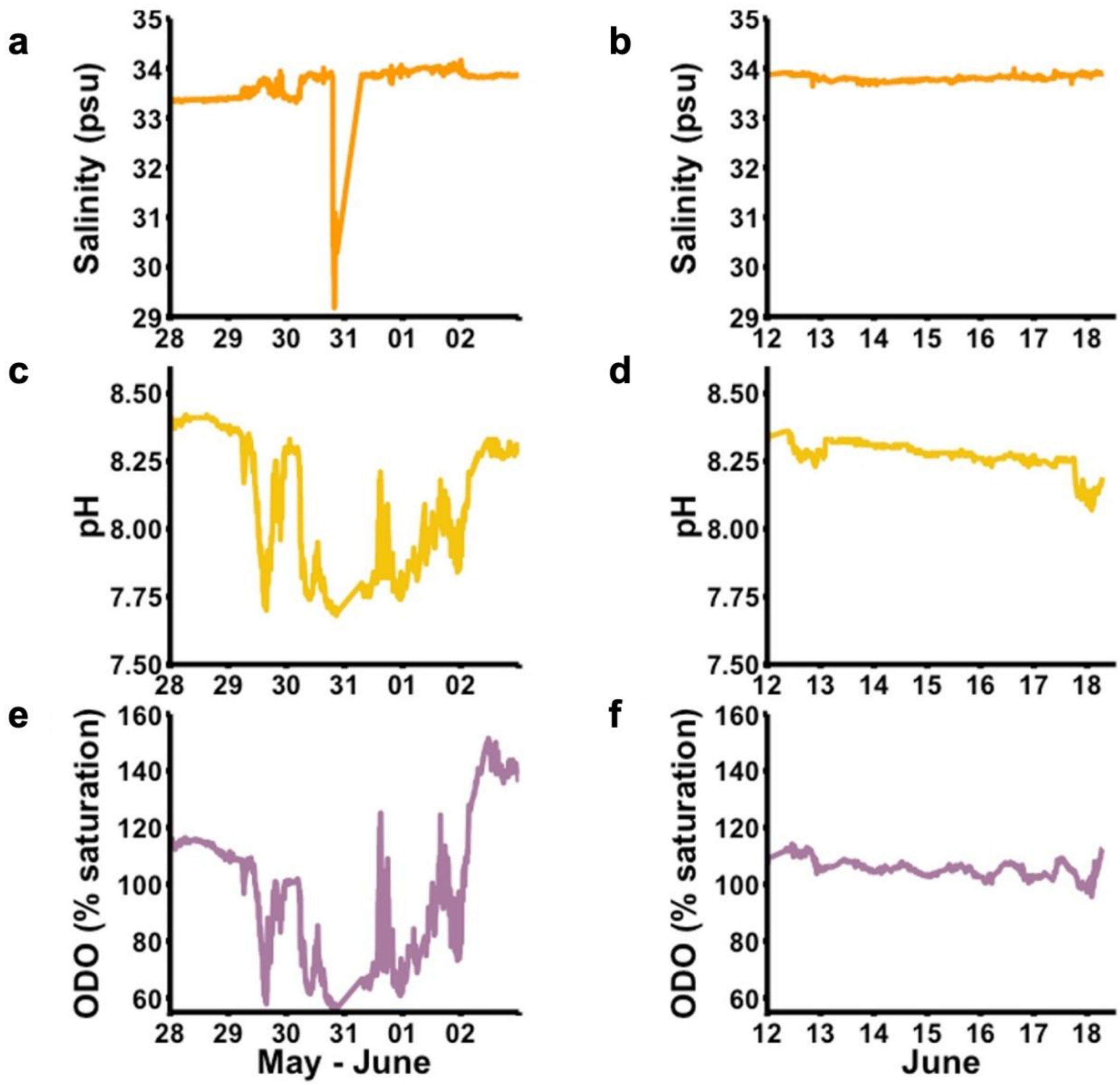


Figure 3.2. Water chemistry measurements during discovery and validation pre-bloom periods. YSI Exo1 Sonde probe used to capture levels of a,b) salinity, c,d) pH, and e,f) dissolved oxygen at sample collection site for the validation and discovery data sets, respectively.

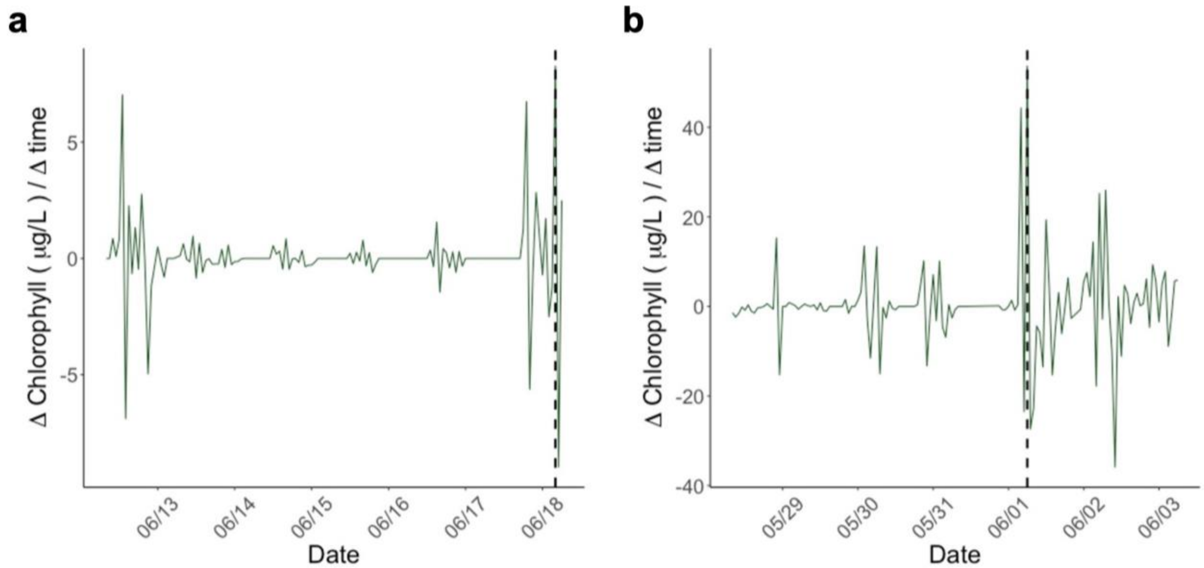


Figure 3.3. Bloom start date (BSD) determined using chlorophyll *a* trace. The rate of change of chlorophyll *a* was calculated to determine the bloom start date for the a) discovery and b) validation data sets. BSD, noted with a dotted line, was annotated as the point of highest rate of change in the time period analyzed.

negative correlation with time (ρ ≤ -0.5, Fig. 3.8b). All peptides from classes with a ρ ≥ |0.5| were selected for the next phase of biomarker discovery. Additionally, peptides from the most abundant class, Alphaproteobacteria, were also selected to proceed through biomarker development. In total, our taxa-driven discovery approach yielded 1858 peptides for consideration as biomarkers in Phase 2.

Phase 2: State shift analysis (HMM) As our goal was to identify biomarkers to predict HAB initiation, a hidden Markov model (HMM) approach was applied to reveal peptides with a non-reverting shift in underlying state processes. A two-state HMM was fit to each of the 2,015 Phase 1 candidate peptides' temporal abundances (Fig. 3.6, yellow). A total of 314 peptides were confirmed to undergo a non-reverting state shift >24 hours prior to the bloom start date (Fig. 3.8c). This set included 88 peptides that were identified from both Phase 1 analyses, as well as 181 that originated from only the Phase 1A PCA and 221 from only the Phase 1B correlation analysis (Fig. 3.6).

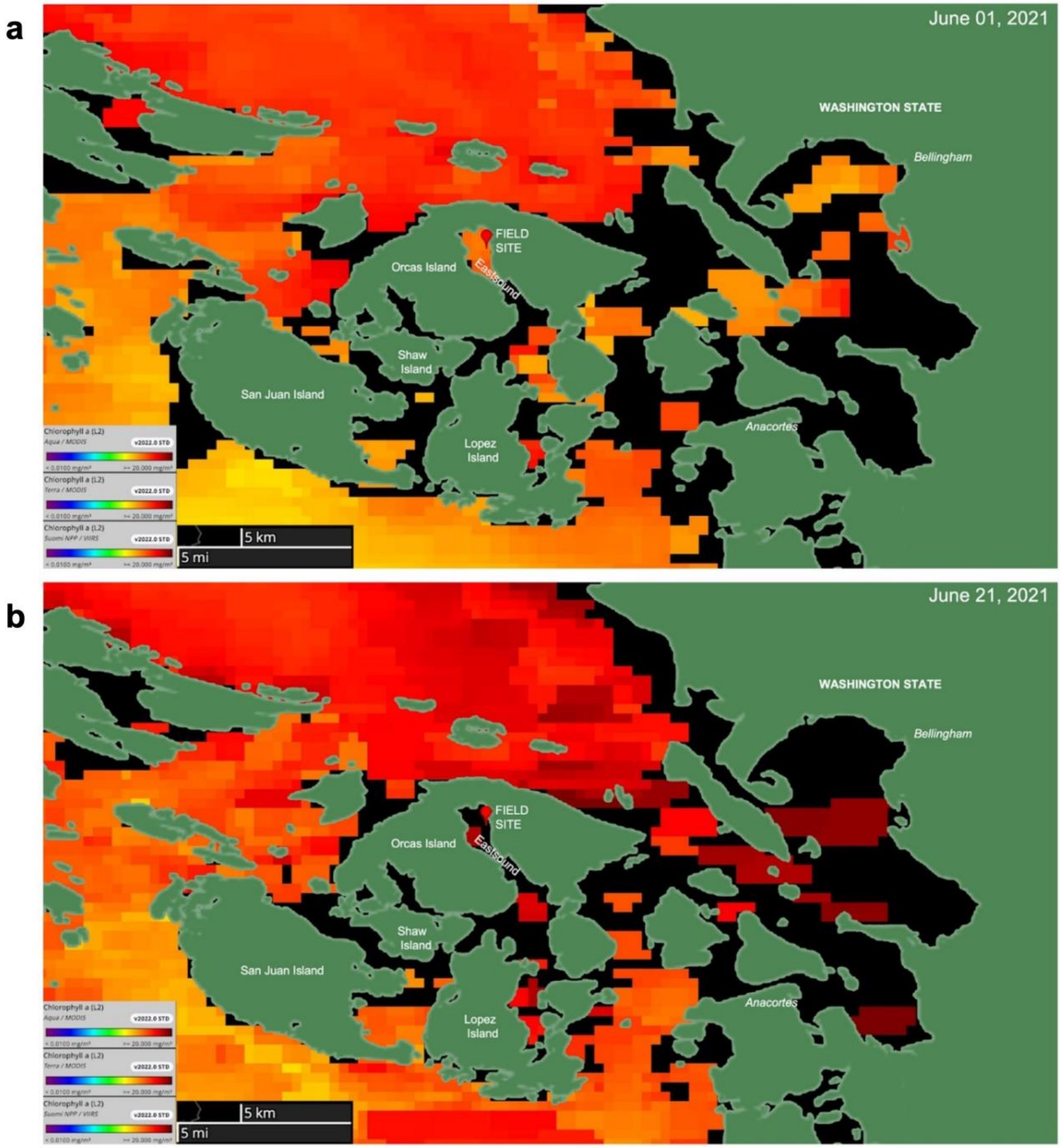


Figure 3.4. Satellite images of chlorophyll *a* concentration confirming HAB presence. MODIS satellite images of near-surface chlorophyll *a* concentration (mg m^{-3}) in waters surrounding East Sound, WA, USA during a) June 1 and b) June 21, 2021. The MODIS Chlorophyll a product is available from both the Terra and Aqua satellites. The sensor and imagery resolution is 1 km, and the temporal resolution is daily. References: MODISA_L2_OC doi:10.5067/AQUA/MODIS/L2/OC/2018

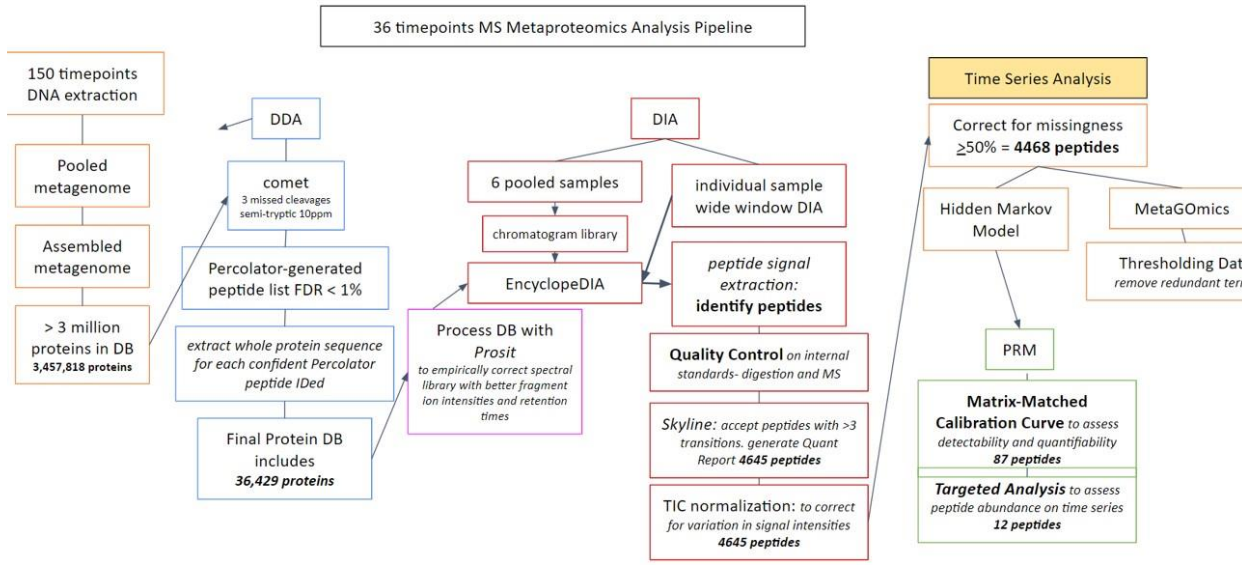


Figure 3.5. Schematic of data analysis pipeline for MS proteomics. A pipeline was created to identify, quantify, and quality assure candidate peptide biomarkers across the experimental analysis.

Phase 3: Matrix-Matched Calibration Curves (MMCC) From the 314 peptides that passed Phase 2, 87 peptides were randomly selected to test for a quantitative response in a dilution series using targeted mass spectrometry (Pino, Searle, *et al.,* 2020). Because metaproteomic measurements using MS techniques are prone to confounding matrix effects, we constructed a matrix-matched calibration curve to ensure that candidate peptides could be accurately and reliably quantified prior to further analysis (Fig. 3.6, blue). The MMCC consisted of a dilution series of this study's microbiome peptides against a matched-matrix of *Colwellia psychrerythraea* 34H peptides to determine if candidate biomarkers exhibited a linear response during targeted MS experiments and establish a limit of detection (LOD) for each peptide. The MMCC method followed by parallel reaction monitoring (PRM) allowed evaluation of the LOD for the 87 peptides down to a detectable threshold of 1e3, and the dilution series demonstrated a linear trend above the LOD (Fig. 3.10). Due to sample limitation, a random subset of 27 peptides that passed the MMCC standards were selected to progress through to the validation phase.

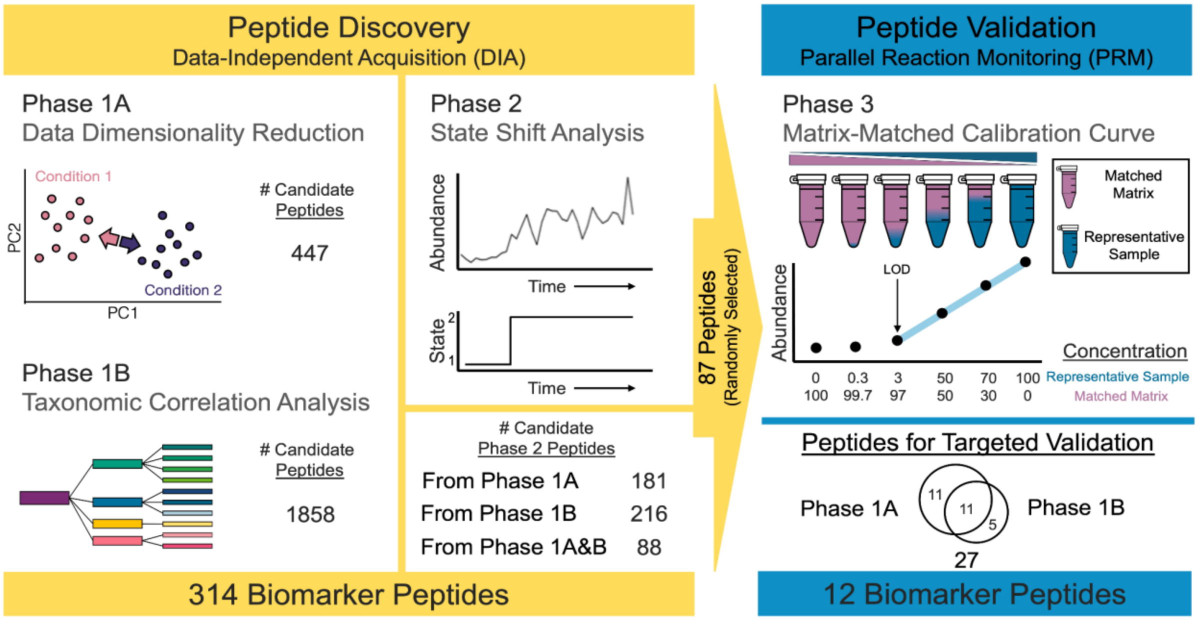


Figure 3.6. Framework for determining candidate peptides. Schematic of the biomarker identification pipeline for systematic peptide discovery (yellow-framed box) and validation (blue-framed box). For peptide discovery, data-independent acquisition was used to quantify peptides from the June 12 -18 samples. Peptides contributing to pre-bloom microbiome proteome variation were identified in Phase 1A through dimensionality reduction (PC loadings indicated by green and purple arrow illustrations in PCA plot) and Phase 1B through correlation analysis of taxonomic trends over time which identified peptides from taxonomic classes that had a time-specific correlation ($\rho \geq |0.5|$). The candidate peptides identified in Phase 1 progressed to Phase 2 state shift analysis using hidden Markov models. In the discovery phase a total of 314 peptides were identified as potential biomarker candidates for downstream testing. A subset of 87 of the candidate peptides were randomly selected and to determine if they yielded quantitative linear responses in Phase 3 using a matrix-matched calibration curve. A final randomly selected subset of 27 peptides that passed Phase 3 was evaluated using targeted analysis to evaluate consistency in temporal abundance from May 28 - June 2 and validate the biomarkers as predictors of HAB initiation.

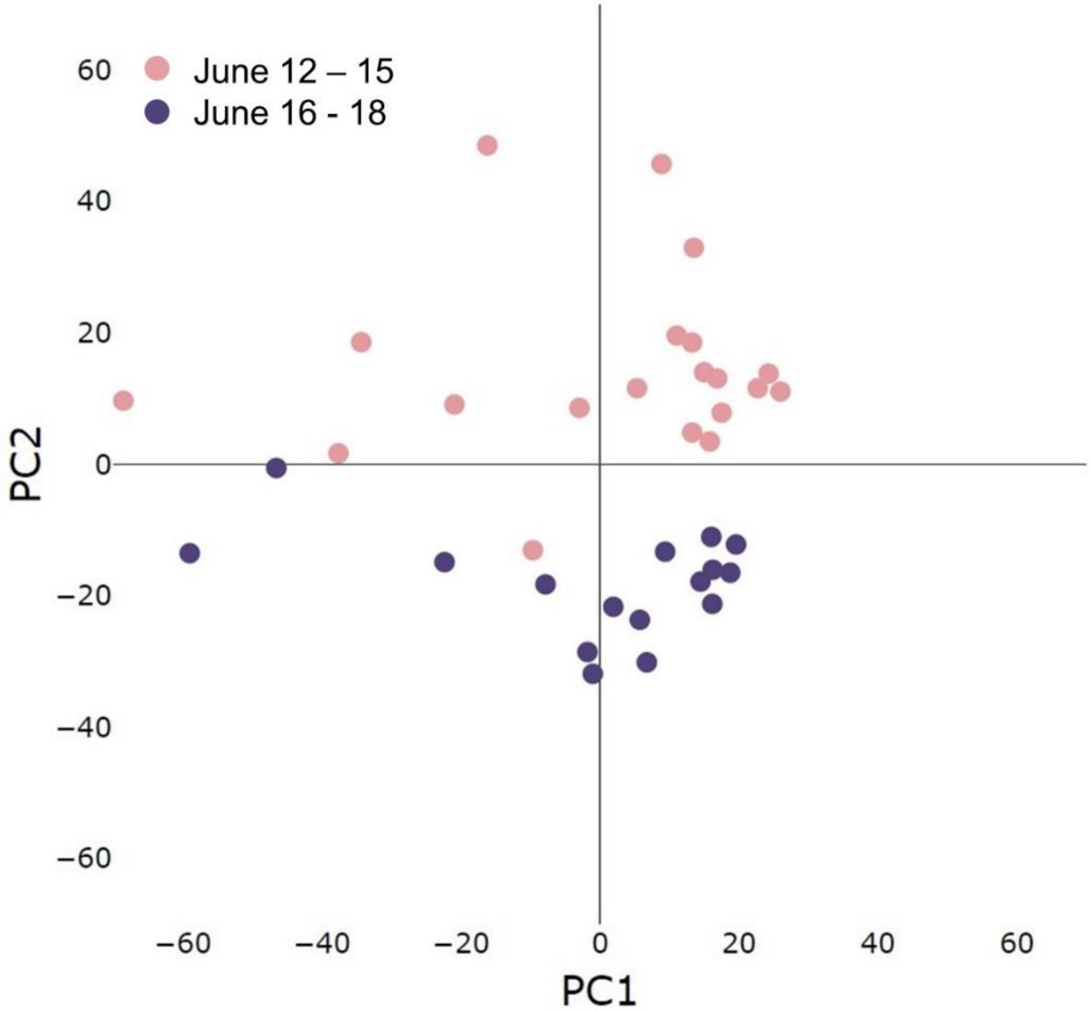


Figure 3.7. Principal component analysis of the bacterial microbiome driven by random missingness. PCA shows that variation across PC1 is driven by random missingness of quantitative peptide abundance across the time series.

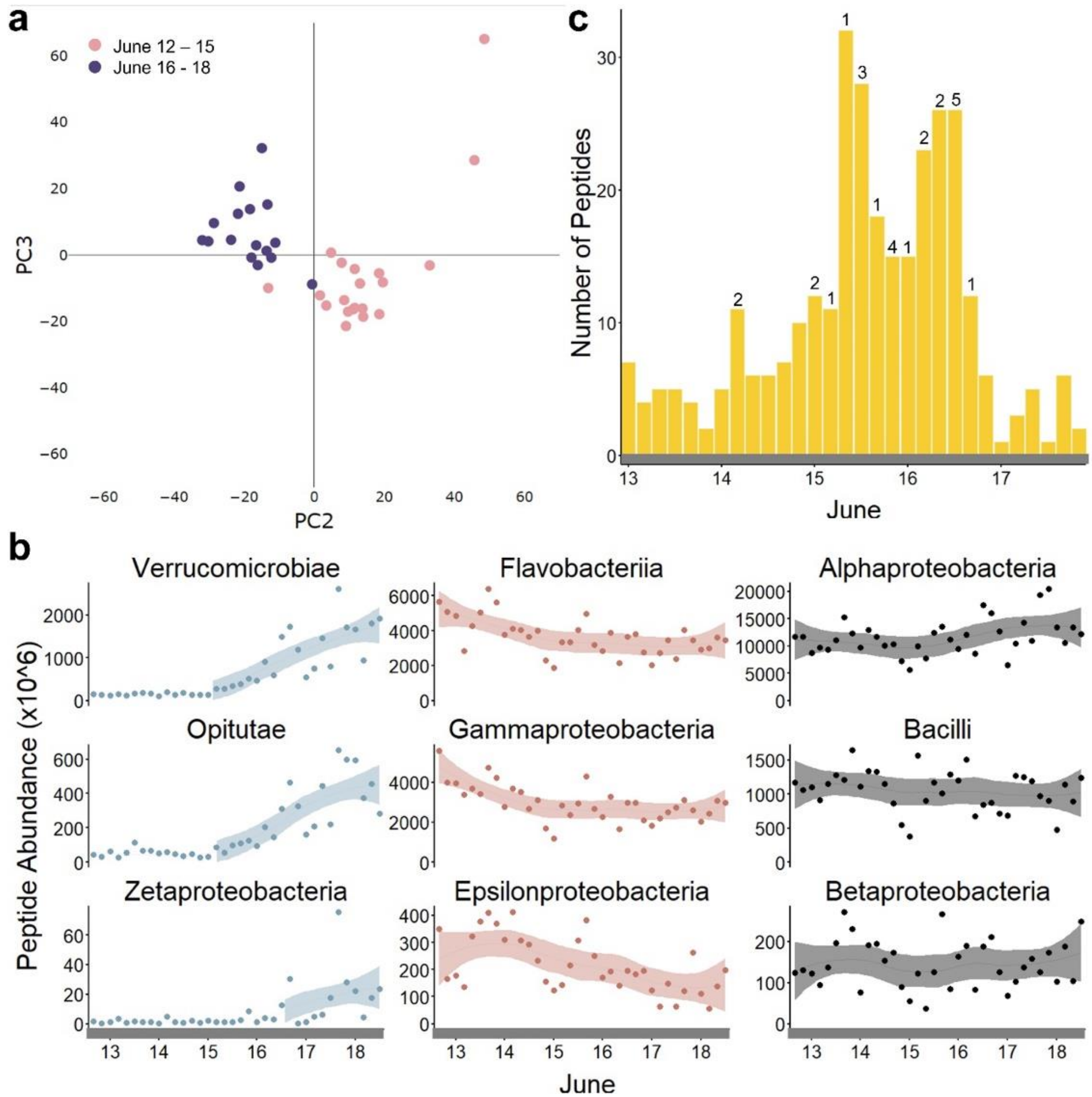


Figure 3.8. Quantified peptides characterized across the microbiome reveal trends in global structure and specific taxa. a) PCA of all quantified bacterial peptide abundances from DIA analyses, colored to illustrate the first (June 12-15, purple) and second (June 16-18, pink) half of the time series, with PC2/3 reported. b) Time series of the total peptide abundances of bacterial classes, with shaded regions representing loess smoothing. Spearman's correlation coefficient rho (ρ) was calculated using each class's summed peptide abundance against ranked time values (positive (≥0.5, blue), negative (≤-0.5, red) and no (black) correlation). c) Distribution of the HMM state shift timing for all non-reverting peptides when average peptide abundance before and after the state shift are compared. Numbers above the bars indicate how many peptides from that time were selected as final biomarkers for validation.

Table 3.1. Spearman's correlation coefficient for class-level summed peptide abundance over time.

Peptide	Correlation coefficient	Significance
Verrucomicrobiae	0.900	2.20E-16
Opitutae	0.849	3.60E-08
Acidobacteriia	0.703	5.09E-06
Zetaproteobacteria	0.642	3.69E-05
Gammaproteobacteria	-0.562	4.56E-04
Epsilonproteobacteria	-0.554	5.55E-04
Flavobacteriia	-0.545	7.19E-04
Bacteroidia	0.458	5.40E-03
Deltaproteobacteria	0.447	6.76E-03
Balneolia	-0.432	2.27E-02
Phycisphaerae	0.421	1.12E-02
Sphingobacteriia	-0.415	1.25E-02
Gloeobacteria	-0.408	1.40E-02
Clostridia	-0.403	1.54E-02
Chlamydiia	-0.363	3.01E-02
Alphaproteobacteria	0.348	3.83E-02
Chitinophagia	-0.298	8.25E-02
Oligoflexia	0.280	1.09E-01
Actinobacteria	0.279	1.10E-01
Planctomycetia	0.262	1.23E-01
Bacilli	-0.228	1.81E-01
Saprospiria	-0.222	1.92E-01
Acidimicrobiia	0.192	2.84E-01
Spirochaetia	0.164	3.54E-01
Tissierellia	0.162	3.99E-01
Lentisphaeria	0.114	5.08E-01
Cytophagia	0.076	6.60E-01
Thermotogae	0.024	8.90E-01
Betaproteobacteria	0.003	9.86E-01

Calculated correlation coefficient (Spearman's rho) and significance (p-value) between summed peptide abundances for each class from the discovery time series.

Table 3.2. Peptide selection proportions by framework phase.

Method	Group	Phase 1	Phase 2
Data dimensionality reduction	Total Number Peptides	447	181
Taxonomic correlation analysis	Total Number Peptides	1858	221
	Alphaproteobacteria	956	61
	Gammproteobacteria	339	34
	Flavobacteria	431	63
	Opitutate	39	21
	Verrucomicrobiae	91	40
	Acidobacteriia	1	1
	Zetaproteobacteria	1	1

The number of peptides evaluated within each biomarker framework phase in the discovery phase are reported. Phase 2 represents the number of peptides from classes that underwent a non-reverting state shift prior to the BSD.

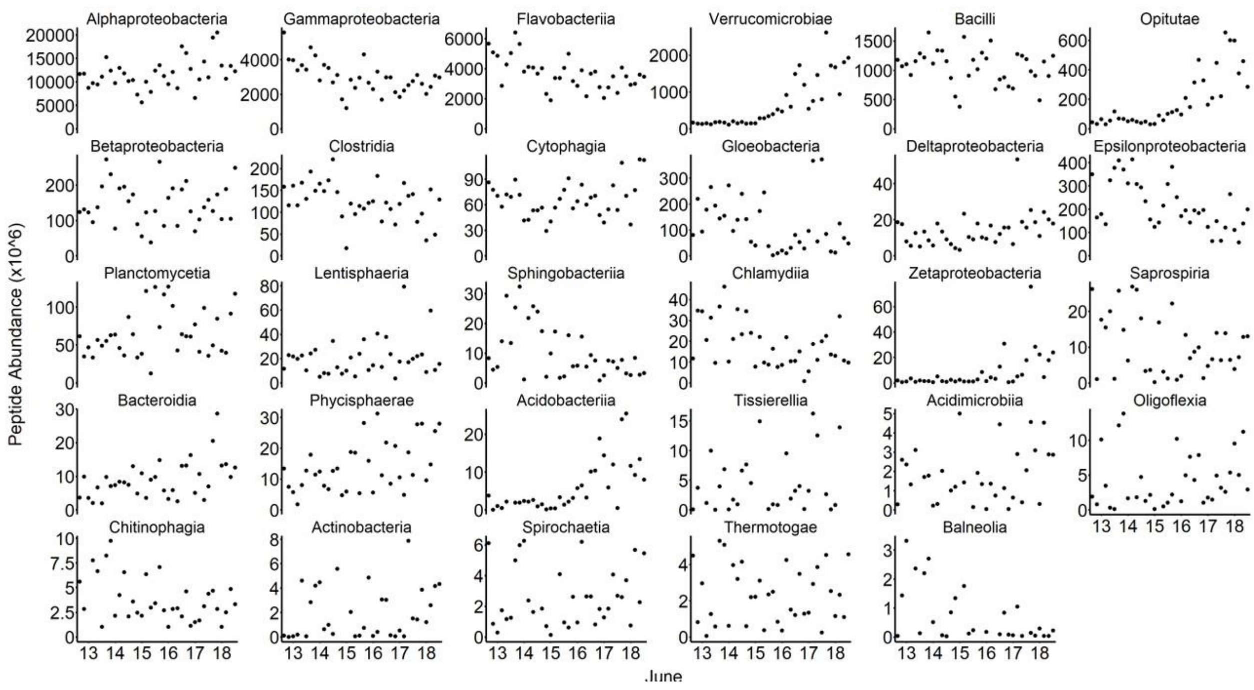


Figure 3.9. Correlation analysis of identified bacterial classes over time shows trends in peptide abundance prior to BSD. Points represent total summed abundance of all peptides annotated per given bacterial class across the discovery dataset.

3.4.3 Validation of biomarkers as predictors of HABs

In order to determine whether our final subset of 27 biomarkers could predict a HAB, targeted MS analyses were performed on the May 28 - June 2 pre-bloom sample set (*i.e.*, the validation period, Fig 3.1c). The goal of the validation experiment was to evaluate whether the 27 biomarkers identified in the discovery bloom data set also exhibited a non-reverting state-shift prior to the June 18 BSD and whether the resulting temporal patterns were consistent across multiple blooms. All peptides identified as potential pre-bloom biomarkers were above the LOD in the validation time series. A two-state HMM was fit to each peptide, revealing a total of 12 peptides that exhibited a non-reverting state shift prior to the BSD (Fig. 3.11a,b; Fig. 3.12-13). The state shift for all peptides occurred at least 24 hours prior to the observed BSD for both blooms (Fig. 3.1c; Fig. 3.11b; Fig. 3.12-13; Table 3.3). Additionally, a Spearman correlation analysis of the discovery and validation blooms showed a strong positive correlation for eight of the quantified peptides when peptide abundance was aligned to the BSD (correlation value ≥ 0.6 (*e.g.*, Fig. 3.11a,b; Fig. 3.14). A detailed annotation of the peptides with Basic Local Alignment Search Tool (BLAST) protein analysis revealed that, of the final 12 validated biomarkers, 8 peptides were from Gammaproteobacteria, 2 peptides were from Flavobacteriia, and 2 peptides were from Verrucomicrobiae. Functionally, 6 of the biomarker peptides were TonB-dependent receptors, 2 peptides were ribosomal, and others are indicative of an active response to the environment (Table 3.4).

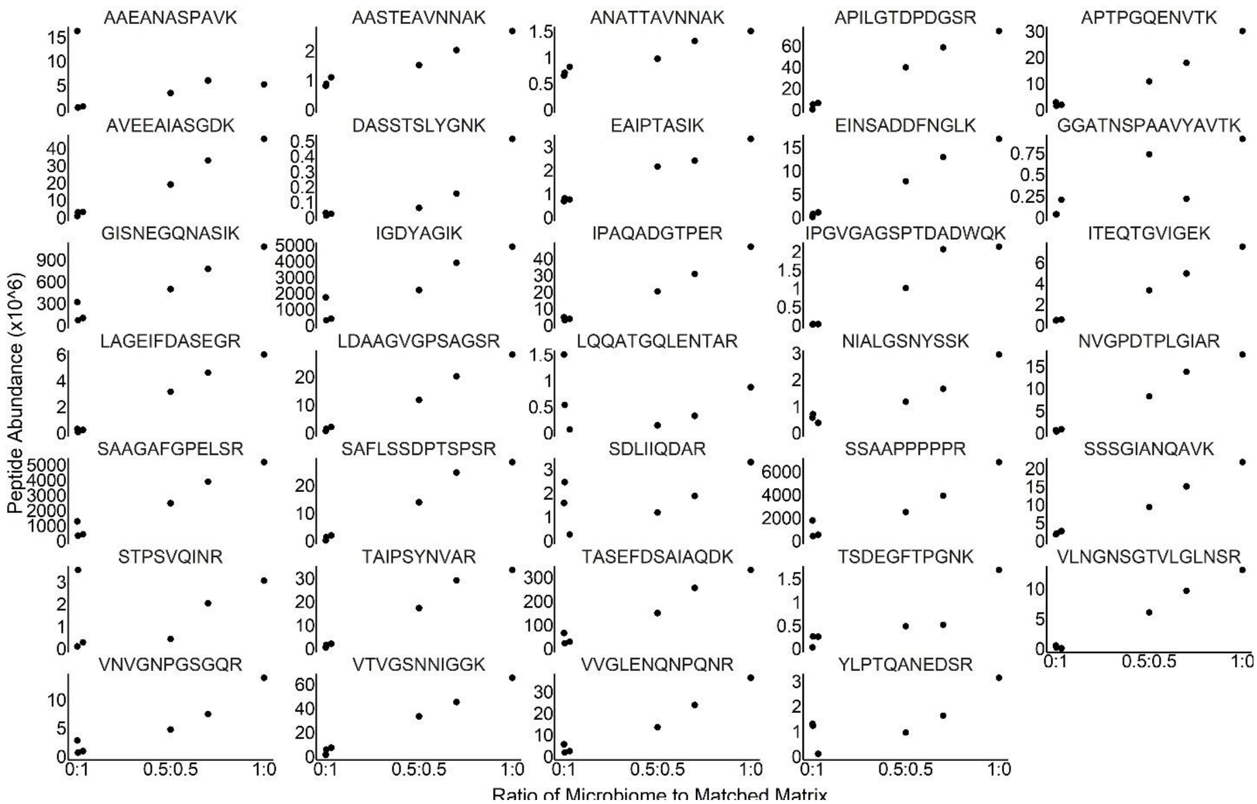


Figure 3.10. Quantitative linearity of biomarker peptides assessed using a matrix-matched calibration curve (MMCC). Peptide abundance of all biomarker peptides in the validation analysis, including controls and standards, for each dilution series of representative HAB sample in a matched-matrix of *Colwellia psychrerythraea* 34H.

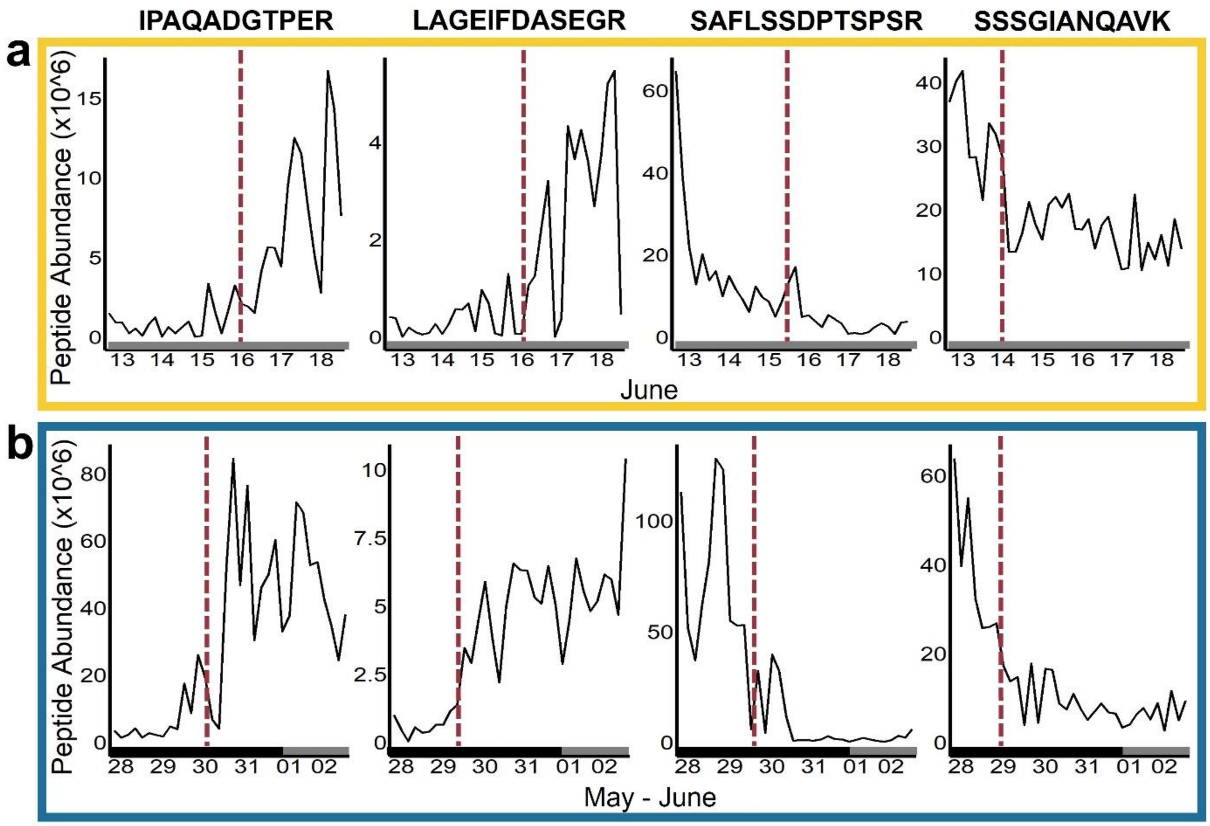


Figure 3.11. Peptides serve as reproducible predictive biomarkers of bloom onset across discovery and validation blooms. A representative subset of 4 of the 12 validated biomarkers' peptide abundances from a the discovery time series (yellow: June 12-18) and b the validation time series (blue: May 28 - June 02). State shift timing predicted by hidden Markov model indicated by dotted line.

Table 3.3. Spearman's correlation coefficient rho and p-values for biomarker peptides.

Peptide	Correlation coefficient	Significance
VVGLENQNPQNR	0.9021314	1.72E-09
APTPGQENVTK	0.8747826	2.59E-06
IPAQADGTPER	0.8286957	2.10E-06
TAIPSYNVAR	0.7955957	1.26E-05
LAGEIFDASEGR	0.7678261	1.90E-05
SAFLSSDPTSPSR	0.7653838	1.32E-05
SSSGIANQAVK	0.6617391	5.86E-04
VLNGNSGTVLGLNSR	0.6108167	1.52E-03
STPSVQINR	0.4792346	1.78E-02
LQQATGQLENTAR	0.3730435	7.34E-02
AASTEAVNNAK	-0.3782609	6.92E-02
AAEANASPAVK	-0.8008696	4.46E-06

Calculated correlation coefficient (Spearman's rho) and significance (p-value) between peptide abundances from the discovery and validation time series by aligning times by BSD.

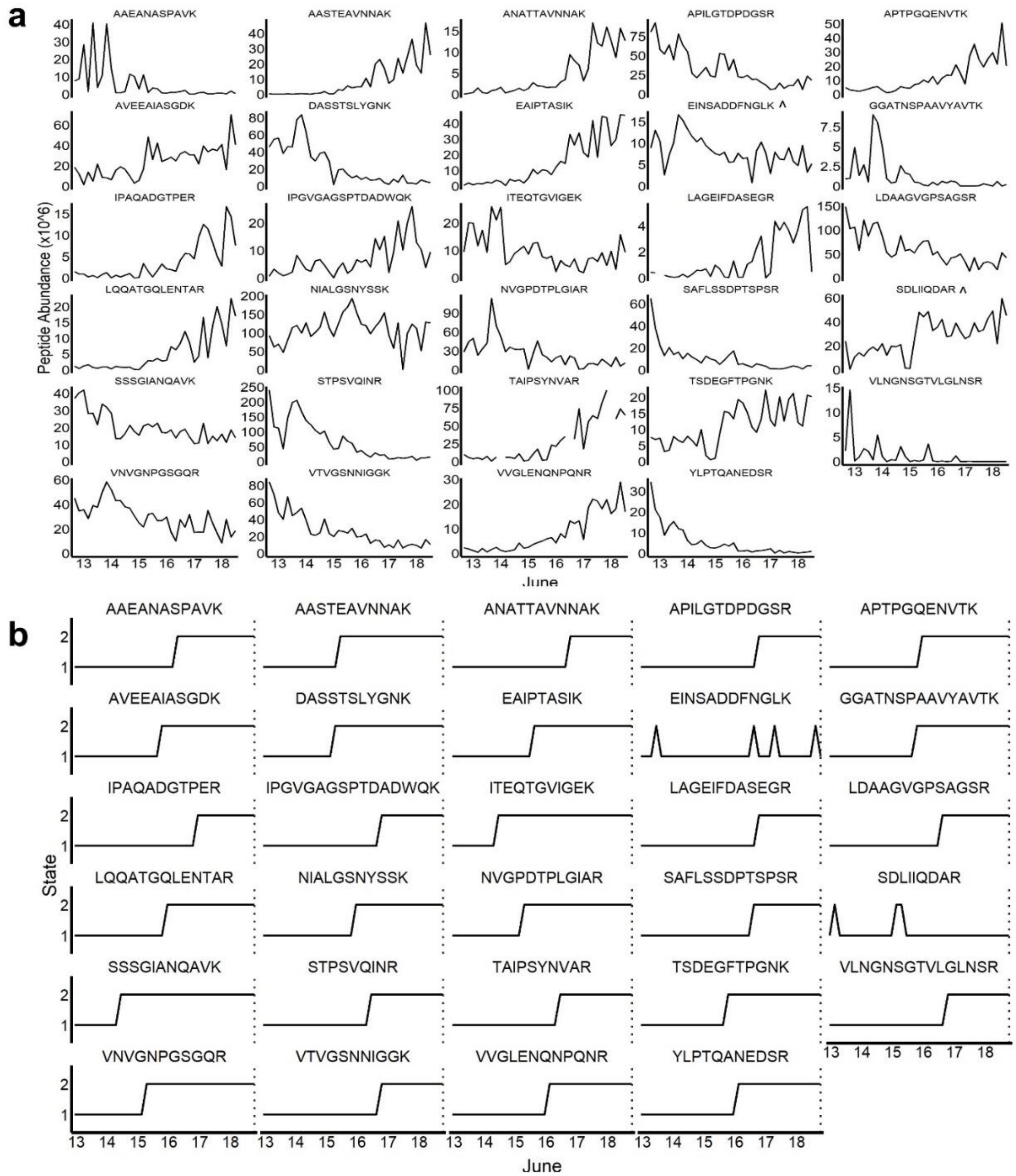


Figure 3.12. Predicted state of peptides from DIA analysis on discovery data set. a Peptide abundances and b predicted state over time (HMM) for the panel of peptides chosen to proceed to Phase 3. Carrot (^) indicates peptides that do not pass Phase 2; these peptides were included in the validation experiment as negative controls. Dotted line indicates bloom start date (BSD).

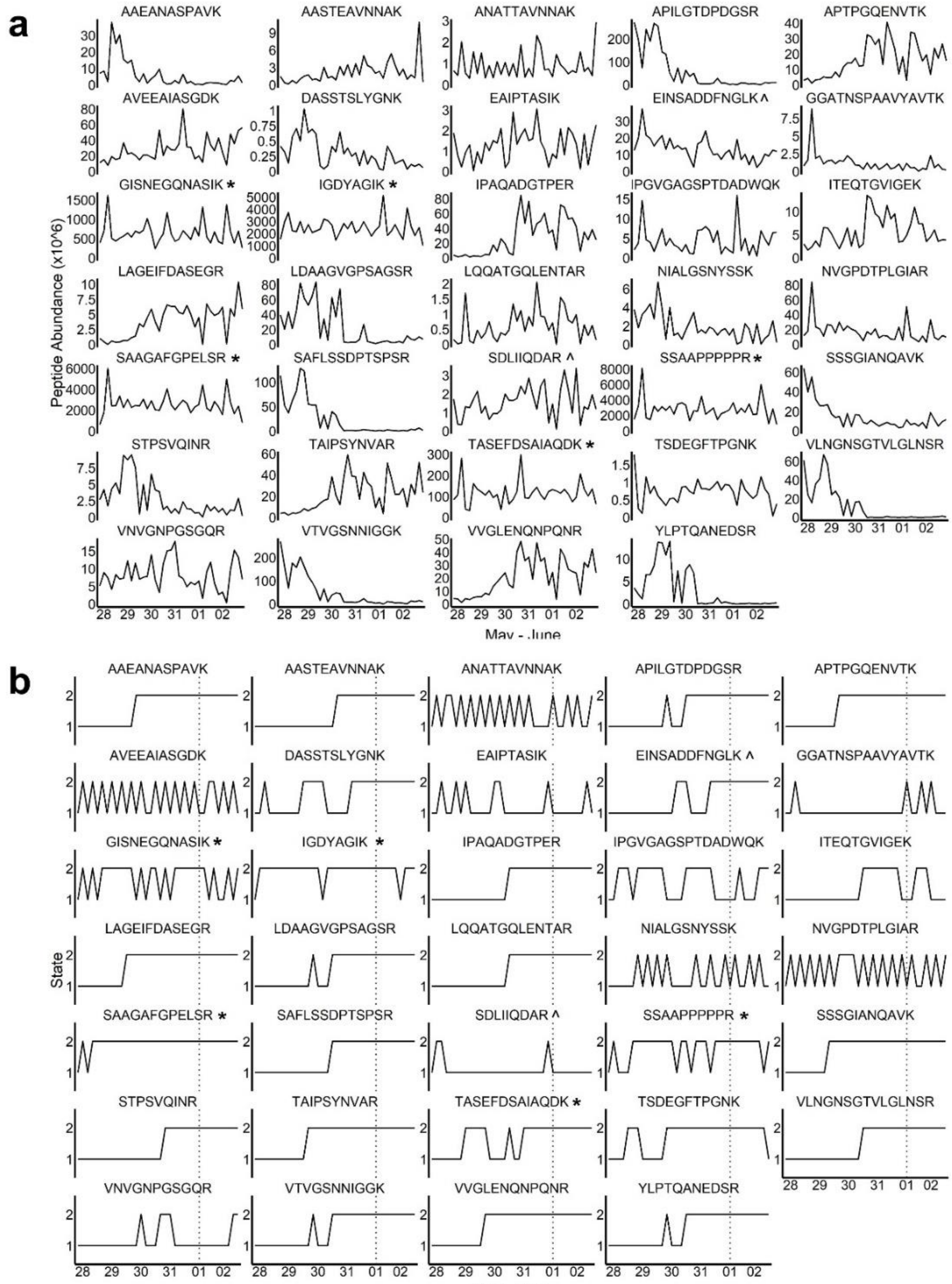


Figure 3.13. Predicted state of peptides from PRM analysis on validation data set. a) Peptide abundances and b) predicted state over time (HMM) for the panel of peptides chosen to proceed to Phase 3. Asterisk (*) indicates external standards; carrot (^) indicates controls. Dotted line indicates bloom start date (BSD).

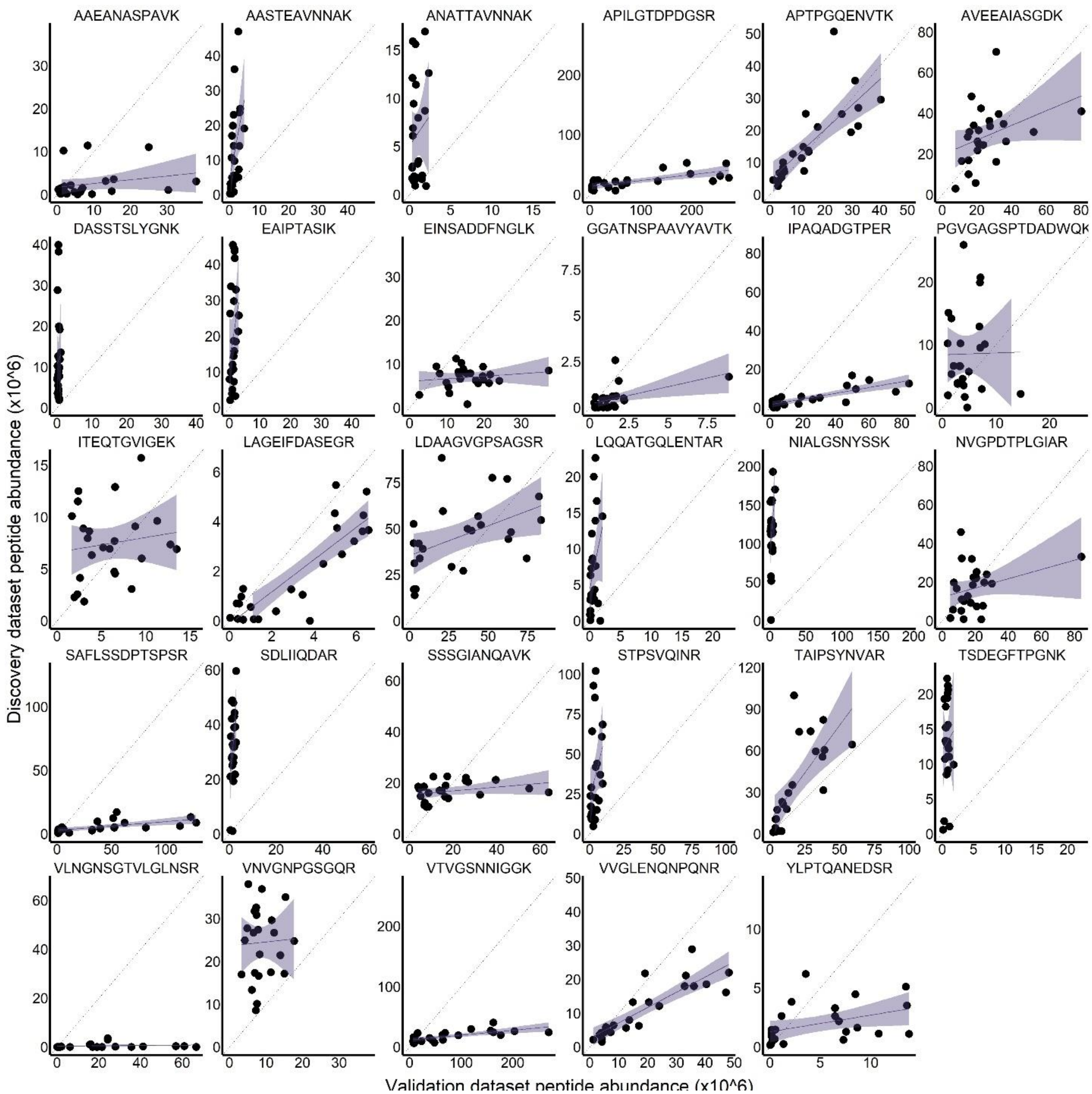


Figure 3.14. Correlation of biomarker peptides between discovery and validation pre-bloom periods. Alignment of peptides determined by setting T_1 as the time point immediately prior to the BSD, with T_{T+1} set as each subsequent time point moving backwards in time from the BSD. Peptide abundances were only compared for time points where both discovery and validation phase possessed a quantitative value (24 samples).

Table 3.4. Biomarker peptide annotations and details.

Taxonomic Class	Peptide	Protein Function	Peptide Abundance Trend	Discovery Pre-bloom State Switch *(hrs before BSD)*	Validation Pre-bloom State Switch *(hrs before BSD)*
Gammaproteobacteria	APTPGQENVTK*	TonB-dependent receptor	+	66	54
	VVGLENQNPQNR*		+	62	54
	IPAQADGTPER	TonB-dependent receptor	+	42	34
	TAIPSYNVAR	TonB-dependent receptor	+	54	54
	LAGEIFDASEGR	30S ribosomal protein S7	+	46	58
	AAEANASPAVK	Bacteriorhodopsin-like protein	-	58	50
	SAFLSSDPTSPSR^	Family 16 glycosylhydrolase (GH16)	-	50	34
	VLNGNSGTVLGLNSR^		-	46	34
Flavobacteriia	SSSGIANQAVK	TonB-dependent receptor	-	102	66
	STPSVQINR	TonB-dependent receptor plug domain-containing protein	-	54	26
Verrucomicrobiae	AASTEAVNNAK	Prepilin-type N-terminal cleavage/methylation domain-containing protein	+	78	30
	LQQATGQLENTAR	50S ribosomal protein L29	+	66	34

*,^ *Peptides map back to the same protein in metagenomic-generated protein sequence database.*

Successful candidate peptides are detailed with taxonomic class, protein of origin function, temporal trend of the quantified peptide abundance (increase as +, decrease as -), and timing of HMM state switch indicated as number of hours prior to BSD in the discovery and validation phytoplankton bloom events.

3.5 DISCUSSION

In this study, we utilized high-temporal resolution sampling of a bacterial microbiome to reveal that certain bacterial peptides undergo explicit and reproducible state shifts ≥24 hours prior to the BSD of two HABs occurring in divergent chemical ecosystems. These peptides originated from bacterial community members whose population densities also reflected changes in total peptide abundance approximately 24 hours prior to HAB onset. These findings confirm our hypothesis that HABs and their associated microbiomes exhibit ecological coupling of metabolic activity in response to the biological or environmental factors that drive HAB initiation, allowing for the development of HAB predictive tools in the form of peptides.

As the long-term goal of this research is to identify candidate peptide biomarkers that can be easily isolated and identified in future applications, sample collection targeted free-living marine bacteria. Targeting free-living bacteria increases the likelihood that any relevant biomarkers would be applicable for predictive use in other bloom scenarios (in comparison to particle-associated bacteria that would be specific to their respective bloom-forming phytoplankton), as well as avoids the practical complications of extracting bacterial signals from abundant phytoplankton biomass (Timmins-Schiffman *et al.,* 2018; Ahern *et al.,* 2021). Additionally, probe-based hourly chlorophyll *a* abundance was used to calculate the BSD (Brody, Lozier, and Dunne 2013), streamlining the assessment of HAB onset independent of laborious, and sometimes unreliable, methods used to quantify cell counts.

To further optimize data analyses, we developed a statistically robust metaproteomic pipeline to increase the accuracy of peptide identifications through an agnostic scaling of the MS search space. Since the DDA MS analysis of site- and time-matched samples provided peptide fragmentation spectra and their corresponding high mass accuracy precursor mass-to-charge ratios, traditional peptide-spectrum match search algorithms were applied to identify all detectable peptide candidates corresponding to the immense metagenomic protein sequence database. This information was then utilized to reduce the metagenomic protein sequence database by 99%. By doing the database reduction using DDA data that is distinct from the DIA data used in the discovery Phases, this approach avoids potential biases incurred by double-dipping methods widespread among other metaproteomic pipelines (Everett, Bierl, and Master 2010). This novel approach allowed for confident detection and quantification of DIA peptide signals in an unbiased, statistically robust manner, while accurately controlling the false discovery rate (FDR)(Fig. 3.5).

Due to the discovery-based nature of this analysis and the innate complexity of the marine microbiome, a combined approach of empirical and taxonomy-specific analyses for biomarker identification was advantageous, as neither individual analysis captured all peptides exhibiting biomarker potential and other ecosystems may exhibit different dynamics than those captured here. That said, many of the biomarker candidates were identified in both peptide selection methods in Phase 1 (88 peptides, 28%; Table 3.1; Fig. 3.6). While Phase 1B (taxonomic correlation analysis) yielded over three times as many candidate peptides as Phase 1A (data dimensionality reduction), each method discovered a similar number of non-reverting peptides in Phase 2 (Phase 1A = 181; Phase 1B = 221). Limited sample quantities for downstream analyses restricted our validation screen to only a subset of peptides (87 of the 314; 28%) for targeted analysis; however, we posit that all 314 biomarkers identified in the discovery analysis have strong potential for development as predictive signals for HAB onset.

Although all samples were collected from the same coastal site, the two blooms began ~17 days apart (Fig. 3.1a) with significantly different chemical landscapes represented by the varied nutrient profiles characterizing the water column prior to each bloom (Fig. 3.1h,i). This pronounced difference between the conditions present during sample collection was seen as a benefit to the study as any bloom biomarker selected for real-world application would need to be consistent across diverse HAB systems and resilient to any external influences from potential perturbations. Despite the contrasting nutrient profiles through time, consistent microbial peptide signals were observed preceding both HABs (Fig. 3.11a,b; Fig. 3.15). This observation underscores the microbiome's potential for revealing key changes in the complex chemical matrix of the ecosystem through highly specific biological responses independent of environmental conditions. Additionally, the timing of the HMM state shifts identified for all the peptides compared with the timing of the BSD for the two blooms revealed that a change in the collective functions of the bacterial microbiome occurred within 24-48 hours prior to bloom onset (Fig. 3.11a,b; Table 3.4). This is the first study to statistically identify and isolate this critical time frame in pre-bloom dynamics for future research and funding mechanisms to focus on.

The class-level bacterial structure of the microbiome in the pre-bloom discovery phase mirrors previously documented marine microbiomes surrounding phytoplankton bloom events (Lindh *et al.,* 2015; Patin *et al.,* 2020; Ferrer-González *et al.,* 2021). The classes dominating our final biomarker screen included commonly occurring Gammaproteobacteria

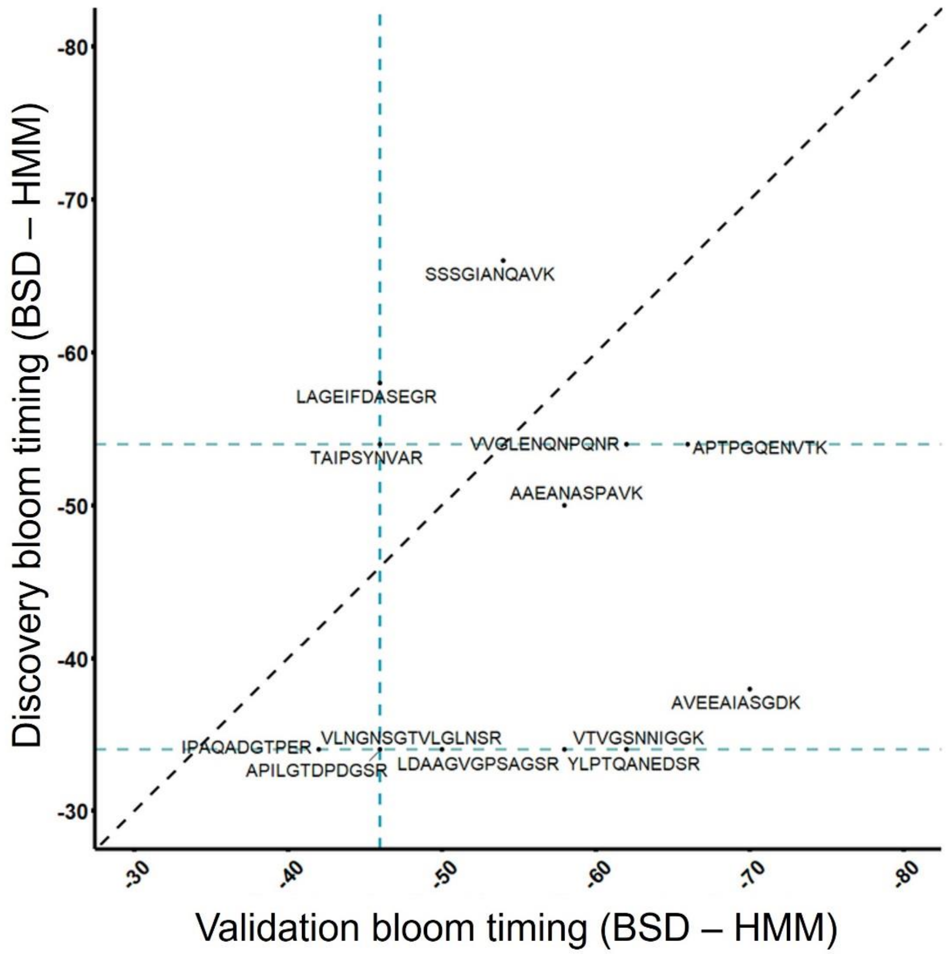


Figure 3.15. Timing of state shift compared to BSD. For each peptide identified as a HAB biomarker, the coordinates for the plot were calculated by subtracting the hour of sample collection immediately following the HMM state shift from the BSD. The amount of time that passed between the HMM state shift and the BSD were compared between the discovery and validation blooms.

Francis *et al.,* 2021), Flavobacteriia (Williams et al. 2013; Teeling *et al.,* 2016), and Verrucomicrobiae (Sichert *et al.,* 2020; L. H. Orellana *et al.,* 2022) (Table 3.4). The shared taxonomic profiles across blooms increase the likelihood that these peptides will be generalizable to other marine phytoplankton blooms and suggest the presence of keystone taxa that precede HAB formation. These classes all yielded detectable, quantifiable, and consistent biomarker peptides across both blooms (Fig. 3.11). A protein BLAST (Altschul *et al.,* 1990) analysis of each of the 12 validated peptides revealed that the active functionality represented by our biomarker panel was driven by general cellular growth and directed responses to chemical conditions of the pre-bloom ecosystem. Two of the biomarker peptides from Gammaproteobacteria uniquely map to the protein family 16 glycosylhydrolase (GH16), a class of carbohydrate-active enzymes (CAZymes) involved in the degradation of polysaccharides, the dominant molecules in transparent exopolymer particles (TEP) released by bloom forming algae (Viborg *et al.,* 2019; Vidal-Melgosa *et al.,* 2021). Another Gammaproteobacteria peptide that decreased as time approached the BSD was the light-driven proton pump bacteriorhodopsin-like protein (Béjà *et al.,* 2000). A reduction in active light-derived energy processes, despite consistent light exposure may indicate a switch to heterotrophic feeding strategies (Guo, Zhang, and Lin 2014; Bar-Shalom *et al.,* 2023). These Gammaproteobacteria peptides may also reflect an active change in metabolism to employ general strategies for combating resource limitation and potential cell death, both conditions suggested by the overall downward trend of total peptides affiliated with the Gammaproteobacteria class (Fig. 3.8b). From Verrucomicrobiae, biomarker peptides mapped to the proteins 50S ribosomal protein L29 and a putative major type IV pilin protein (Prepilin-type N-terminal cleavage/methylation domain containing protein). We would expect an upregulation of protein translation-associated pathways, such as ribosomal proteins, to support the pre-bloom growth observed in Verrucomicrobiae (Scott *et al.,* 2010). Type IV pili are functionally diverse, playing a role in processes ranging from motility and biofilm formation to DNA uptake, protein secretion, and adherence(Bünger *et al.,* 2020; Ellison, Whitfield, and Brun 2022), potentially allowing them to serve as a useful wide-ranging biomarker due to the prevalence of Verrucomicrobiae in marine ecosystems (L. H. Orellana *et al.,* 2022). All biomarkers exhibited a statistically significant, non-reverting state shift in abundance (Fig. 3.12-13), allowing for the possibility of further development as a rapid assessment tool that can differentiate normal and pre-bloom microbiome states using a relative change in abundance threshold (Pandey, Malviya, and Chottova Dvorakova 2021).

While a diverse screen is advantageous for capturing the heterogeneity of potential HAB systems, having multiple biomarkers that respond differently to the same underlying process adds valuable redundancy to a detection system. If one biomarker fails or provides ambiguous results, others may provide diagnostic information, enhancing the overall reliability of the detection system (Ou *et al.,* 2021). In our final peptide panel, 6 of the 12 biomarker peptides were identified as TonB-dependent receptors (TBDRs)(Table 3.4); TBDRs are components of a critical energy-dependent transport system for acquiring scarce vital resources (Silale and Berg 2023). TonB systems are ubiquitous and highly expressed in gram negative bacteria (Morris *et al.,* 2010; Tang *et al.,* 2012; Dong *et. al.,* 2014), and individual TBDRs exhibit high selectivity in substrate uptake, with targeted compounds including (but not limited to) iron (Pawelek *et al.,* 2006), vitamin B_{12} (Shultis *et al.,* 2006), carbohydrates (Blanvillain *et al.,* 2007), lignans (Fujita *et al.,* 2019), and nickel chelates (Schauer *et al.,* 2007). While the high diversity of TBDRs makes it difficult to elucidate their specific functional role in this ecosystem, the contrasting patterns between the TBDRs in bacterial classes detected here (*i.e.*, increasing abundance in Gammaproteobacteria vs. decreasing abundance in Flavobacteria) suggests that distinct chemical signals in the water column may be preferentially beneficial to specific organisms within the microbiome (Russo et al. 2016; Teeling et al., 2016; Pollet, Martin, and Koropatkin 2021). Previous studies have demonstrated that both Gammaproteobacteria and Flavobacteriia utilize TBDRs in response to algal blooms, specifically to consume complex organic molecules (*i.e.*, polysaccharides) released throughout bloom progression by algae as they increase their metabolic productivity (Williams *et al.,* 2013; Dong *et al.,* 2014; Francis *et al.,* 2021). To date, this is the first in situ evidence that TBDR proteins may be active *prior* to bloom formation. Because TonB systems are a metabolically expensive transport method, the cosmopolitan frontline response seen here suggests TBDRs equate to an essential and highly sensitive indicator of changes in water chemistry precluding HAB initiation, posing an ideal target for the development of a forecasting tool.

Techniques for bloom monitoring and forecasting that utilize water chemistry (Marrone *et al.,* 2024), fluid dynamics (Zheng *et al.,* 2021), satellites (Lin *et al.,* 2021), machine learning (Lee *et al.,* 2023), and even the microbiome (Gajardo *et al.,* 2023) have been proposed; however, to date, no reliable predictor of HABs exists. This research highlights both the potential of the microbiome for generating intracellular predictive signals as well as the potential for dissolved metabolomics analyses to elucidate extracellular biochemical signals in the water that may be driving bloom formation. Additionally, further study is needed in both oceanic and coastal marine ecosystems to validate if these findings are translatable to blooms of other planktonic types, including cyanobacteria and toxin-producing blooms. While capturing a bloom system prior to onset remains difficult, isolating this critical time frame in future studies could advance our understanding of what triggers a bloom and how community structure is impacted by environmental factors prompting bloom initiation.

Chapter 4: THE INFLUENCE OF ALGAL BLOOM CYCLES ON METABOLIC SCHEDULING WITHIN THE BACTERIAL MICROBIOME

4.1 INTRODUCTION

The ocean's microbial diversity is immense (Sunagawa *et al.,* 2020). Various studies attempting to quantify bacteria in the ocean estimate cell count ranges spanning several orders of magnitude (Whitman, Coleman, and Wiebe 1998; Kallmeyer *et al.,* 2012). While many studies have attempted to characterize the depth of microbial diversity in marine environments (Brown *et al.,* 2009; Salazar *et al.,* 2016), reports of community composition are consistently dominated by only a few bacteria (Agogué *et al.,* 2011; Allen *et al.,* 2012; Sunagawa *et al.,* 2015; Walker *et al.,* 2022) - namely members of the Proteobacteria and Bacteroidetes phyla. The prevalence of these bacteria suggests an ecological or evolutionary competitive edge that allows them to be suited to resource and nutrient limitations common in the ocean (Giovannoni and Stingl 2005).

Recent research posits that circadian rhythms may confer an advantage in survival (Johnson *et al.,* 2017; Leach, Carrier, and Reitzel 2019). Studies of cyanobacteria have paved the way for greater understanding of circadian rhythmicity (Golden 2020; Tanaka, Shimakawa, and Nakanishi 2020) across various levels of life including prokaryotes, eukaryotes, and archaea. For example, it is known that organisms capable of photosynthesis engage in this metabolic process primarily during the day to optimally utilize sunlight (Eberhard, Finazzi, and Wollman 2008). In this chapter, I propose the idea that many metabolic processes selectively occur at a specific time, introducing a metabolic schedule for the organism that encompasses not just one specific function but rather a continuous cascade of molecular functions that occur on a regulated schedule. When organisms are able to regulate their functionality they may be better able to anticipate environmental perturbations and respond with a competitive advantage for gaining resources (Johnson *et al.,* 2017). The discovery of regular scheduling for functional activities in the microbiome highlights the need for microbiome studies to progress beyond genetic sequencing to further characterize active metabolisms in order to better understand how organisms are responding to changes in their environment.

While blooms can be environmental disasters with a variety of harmful consequences on the environment and surrounding life, they also provide a large natural reservoir for carbon sequestration (Falkowski *et al.,* 2000), yield a source of food for surrounding life (Huisman *et al.,* 2018), and play a role in nutrient cycling in the ocean (Howarth and Marino 2006). The sudden release of nutrients such as dissolved organic matter (DOM) during bloom decline creates an opportunity for rapid microbial community succession (Yang *et al.,* 2015; Zhou *et al.,* 2020; Chen *et al.,* 2021). For example, metatranscriptomics studies of the bacteria-phytoplankton relationships surrounding a bloom show that closely associated bacteria will increase expression of a variety to transporters to take advantage of increased DOM (Rinta-Kanto *et al.,* 2012). Various classes of bacteria like Verrucomicrobiae have been documented as specialized consumers of complex polysaccharides released during bloom progression (Orellana *et al.,* 2022). Members of the class Gammaproteobacteria have been shown to increase early in bloom cycles as rapid responders to specific phytoplankton activity (Tada *et al.* 2012). Flavobacteriia are also rapid responders that dominate microbial communities surrounding peak bloom activity (Williams *et al.*, 2013). Additionally, when bacteria, such as Flavobacteriia, metabolize algal byproducts they, in turn, release DOM preferentially utilized by other bacterial groups, including Gammaproteobacteria and Alphaproteobacteria (Williams *et al.,* 2013). In this way, the rapid response of one bacteria evolved to have a competitive edge for utilizing released nutrients promotes dynamic patterns of bacterial succession throughout the bloom cycle (Buchan *et al.,* 2014). Additionally, microbiomes associated with bloom events exhibit a surprising consistency throughout the bloom cycle (Sapp *et al.,* 2007; Teeling *et al.*, 2016).

The consistency of taxa yields precedent to the idea that these bacterial members of marine populations should also be factored into models of phytoplankton bloom development and progression. Algal bloom cycles are characterized by a sequence of initiation, exponential growth, and demise. The field season detailed in Chapter 3 captured a unique ecosystem during a period of transition in which the number of bacteria peaked between two *Chaetoceros* blooms. Mass spectrometry was used to evaluate the changes in the functional profile of the microbiome. While the dataset immediately precedes a *Chaetoceros* bloom that began on June 18, 2021, it is also important to explore the context of this dynamic microbiome as a result of prior bloom demise. To investigate temporal trends in the context of a changing environment, a two state hidden Markov model was used to identify functions in the microbiome through an analysis of gene ontology (GO) terms. Immediately preceding the second *Chaetoceros* bloom, a rapid onset of metabolic activity was observed in classes of the phylum Verrucomicrobia, specifically the classes Verrucomicrobiae and Opitutae, which underwent a significant increase in all functional terms in the second half of the analyzed time series. Alternatively, the classes Gammaproteobacteria and Flavobacteriia exhibited a non-reverting state shift characterized by an overall decrease in functional output. Given the prominence of these classes in the context of HAB literature, the decrease in peptide abundance is likely attributable to their decline following a depletion of the influx of organic matter produced during the bust cycle of the previous bloom. Additionally, the classes Planctomycetia and Gloeobacteria also exhibited a non-reverting state shift in their respective biological processes though the relevance of these lower abundance bacteria is less clear as they are underrepresented in marine literature. The examination of these changing classes of bacteria and their metabolic cascades in the context of multiple bloom events provides insight into the dynamic nature of microbiome succession surrounding bloom events and can enhance microbiome dynamic models and ecosystem process forecasting.

4.2 METHODS

Methods related to sample collection and data analysis are detailed in Chapter 3.3 or Nunn *et. al.,* 2024.

Taxonomic and functional in summed peptide abundances. MetaGOmics (Riffle *et al.,* 2017) software (version 0.2.0; https://www.yeastrc.org/metagomics) identifies the lowest common ancestor on each peptide's taxonomic and functional Gene Ontology (GO) assignment using a provided site-specific metagenomic-generated protein sequence database. Here, we used the DDA-refined database as our background proteome and provided a complete list of all peptides identified at each timepoint with their respective peptide quantities (DIA-MS). The MetaGOmics reports were used to annotate peptides to their respective GO terms and taxonomic classes. Peptide abundances resulting from each class as annotated by MetaGOmics were summed at each timepoint and plotted through time.

HMM State-shift time series analysis. A two-state hidden Markov model (HMM) was then fit to the quantitative values associated with each class or GO term to identify those that undergo a non-reverting state shift >24 hours prior to the BSD (R package depmixS4). The time series data was log transformed prior to model fitting. The model was specified to have a Gaussian distribution and transition probabilities were assumed to be stationary. Model parameters were estimated using an Expectation-Maximization (EM) algorithm, with a random seed set to ensure reproducibility. The resulting probability matrix was used to determine the latent state of the observations, and the determined state outcomes were plotted through time.

4.3 RESULTS

4.3.1 *Analysis of microbiome between two blooms*

A high-resolution data set of the bacterial metaproteome captured two independent harmful algal bloom initiations of the phytoplankton *Chaetoceros* (Fig. 4.1a; Fig. 3.1a). The first bloom was closely followed by an increase in bacterial cell counts indicative of a bacterial bloom that peaked following the first phytoplankton bloom demise (Fig. 4.1a). This bacterial bloom decreased in cell count number prior to the time frame analyzed by metaproteomics; however, there was an increase in bacterial cell count on June 16, two days preceding the second phytoplankton bloom (Fig. 4.1b).

Using data-independent acquisition methods, a mass spectrometry analysis of the six days immediately preceding the second bloom identified 4468 peptides across a time series of 36 total time points collected every four hours. Of the 4468 peptides in the dataset, 4370 peptides (97.8%) were annotated with taxonomic and functional terms by MetaGOmics. Class level taxonomy annotations were assigned to 2015 peptides (45%) and then used to look for community level trends while GO term grouping was utilized to analyze trends in the timing of functional changes across the microbiome. A total of 29 classes and 803 GO terms were identified.

4.3.2 *A unified cellular perspective: GO term analysis of the total microbiome*

To assess general trends across the microbiome, a complete Gene Ontology Directed Acyclic Graph (GO-DAG) for the bacterial microbiome was assembled using annotations from all GO terms identified by MetaGOmics. Each tree consisted of all possible GO terms with the root node representing the GO type (i.e., biological process), leaf nodes representing the highest specificity GO terms, and parent nodes as any node that generates a child or leaf node (Fig. 4.2).

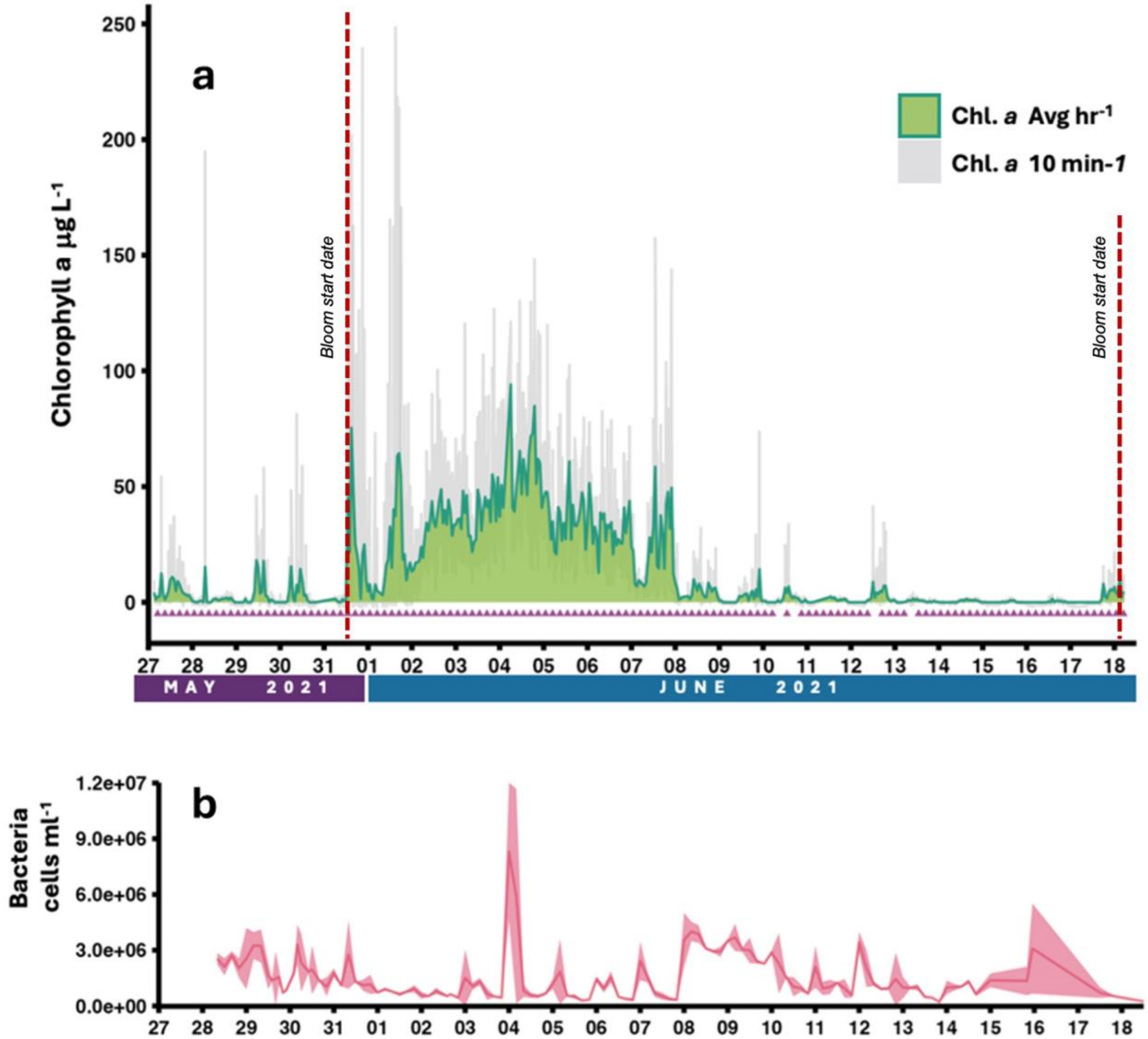


Figure 4.1. Bacterial abundance changes surrounding a phytoplankton bloom-bust cycle. a) Chlorophyll *a* levels across the 22 day field season show the full bloom cycle of one *Chaetoceros* bloom and the pre-bloom period of a second *Chaetoceros* bloom. b) Bacterial cell counts from the field season show an increase in cell count during and immediately following the first bloom as well as prior to the second bloom (adapted from Nunn *et. al.,* 2024).

All analyses here were completed by constructing the GO-DAG generated using the root node “biological process.” This GO type was selected as the most relevant for assessing changes in bacterial function.

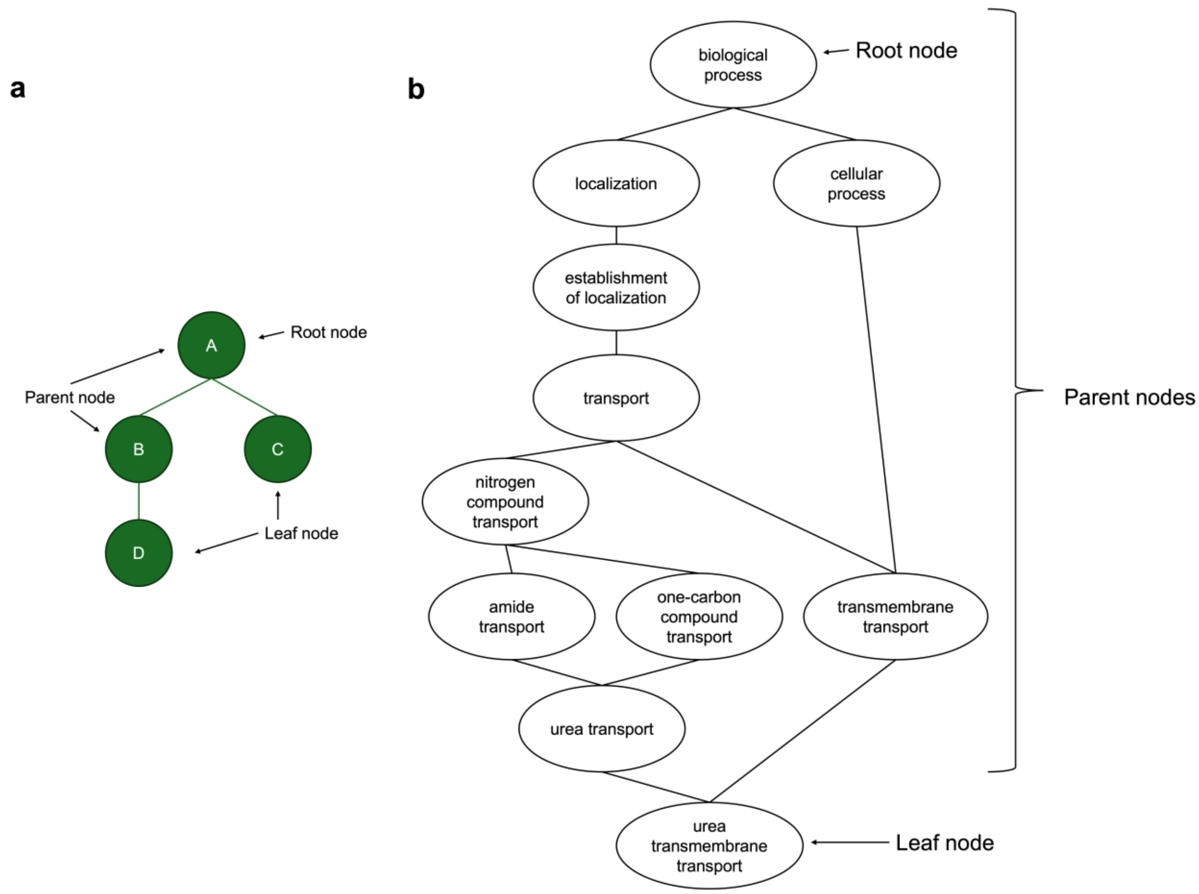


Figure 4.2. Diagram of Directed Acyclic Graph (GO-DAG). a) A full GO-DAG will contain root nodes that form the base of the tree and have no parent node (A), parent nodes that are connected to lower GO terms (A,B), and leaf nodes that form the terminal end of the tree (C, D). b) Example of a GO-DAG for urea transmembrane transport.

All GO terms annotated as “no rank” were selected for a representative analysis to identify trends in abundance patterns across the microbiome independent of taxonomy. Of the 803 GO terms within the data set, 277 were not annotated at the class level. A hidden Markov model determined that 13 unique time series representing 31 GO terms were found to undergo a non-reverting state shift prior to HAB initiation (Fig. 4.3). Of the 31 non-reverting GO terms, 12 were not annotated to have a lowest common ancestor (LCA) class level, suggesting these peptides are conserved across phyla or kingdoms. These terms included aerobic respiration, protein processing, glycine catabolic process, nitrogen cycle metabolic process, serine family amino acid catabolic process, sodium ion transport, and sulfur compound catabolic process (Fig. 4.3).

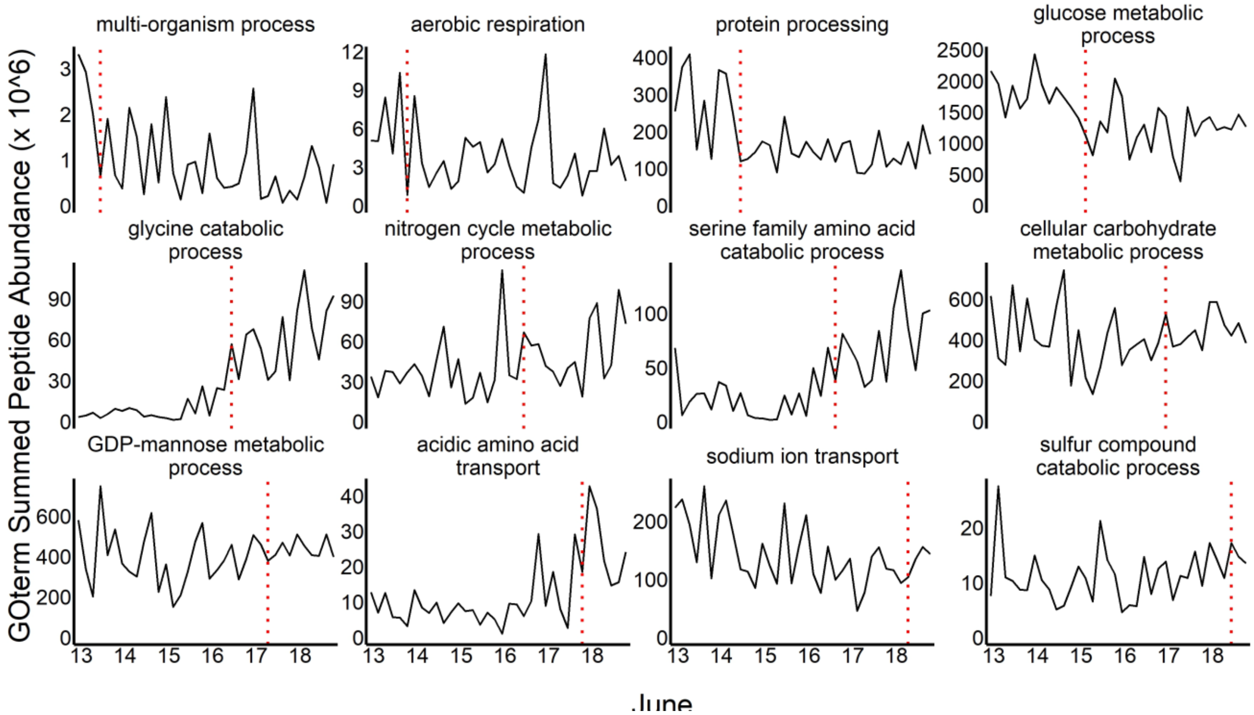


Figure 4.3. Analysis of non-reverting state shifts across the whole microbiome reveal unique GO terms. A two-state hidden Markov model was fit to all GO terms identified within the microbiome time series to identify those terms that underwent a non-reverting state shift prior to HAB initiation. The time series for each non-reverting GO term is ordered here by timing of the underlying state shift.

4.3.3 Pairwise correlations between microbiome functions

To reduce the innate complexity of the GO term analysis, a filtering algorithm, referred to henceforth as MetaGOmics-thresholding, was applied prior to assessing for pairwise correlations to extract a subset of independent GO terms by trimming the DAGS. The settings within the MetaGOmics-thresholder are a peptide q-value of 0.01 and a distinct peptide count of 3. A threshold of zero was utilized to simplify downstream interpretations, thereby reducing the total set of GO terms from 803 down to 184 starting terms. A Spearman correlation analysis was then performed on this subset using summed peptide abundances for each GO term identified by MetaGOmics to identify related functions. A total of 353 strong associations were discovered (correlation ≥|0.8|) between pairwise comparisons of GO terms, with 281 positive correlations and

72 negative correlations. The same analysis conducted at the peptide level yields 4,417 strong positive correlations and 585 strong negative correlations.

4.3.4 Highly conserved GO terms show trends in functionality pre- and post-bloom

Using biological process as a parent node, all child nodes immediately connected to biological process were isolated for downstream analyses, resulting in 10 highly conserved processes represented across 19 of the 29 total classes. These nodes included both parent and leaf nodes (Fig. 4.4). Trends in these 10 highly conserved functional processes were evaluated by plotting the time series of summed peptide abundances for each term (Fig. 4.4a,b), normalizing to each time point (Fig. 4.4c), and normalizing to the GO term itself (Fig. 4.4d). In the context of their specific peptide abundances, the classes Alphaproteobacteria, Flavobacteriia, and Gammaproteobacteria dominated with the highest abundance of all terms analyzed (Fig. 4.4a). Terms that were identified across many bacterial classes exhibited a greater stability in their abundance (Fig. 4.4a-c) and ratios (Fig. 4.4d) through time. For example, the terms “metabolic process” and “cellular process” showed spikes in abundance throughout the time series but these spikes were shared across all of the represented classes (Fig. 4.4c). In other cases, certain classes like Verrucomicrobiae (Fig. 4.4, green) and Opitutae (Fig. 4.4, pink) began to increase on June 16 across all GO terms.

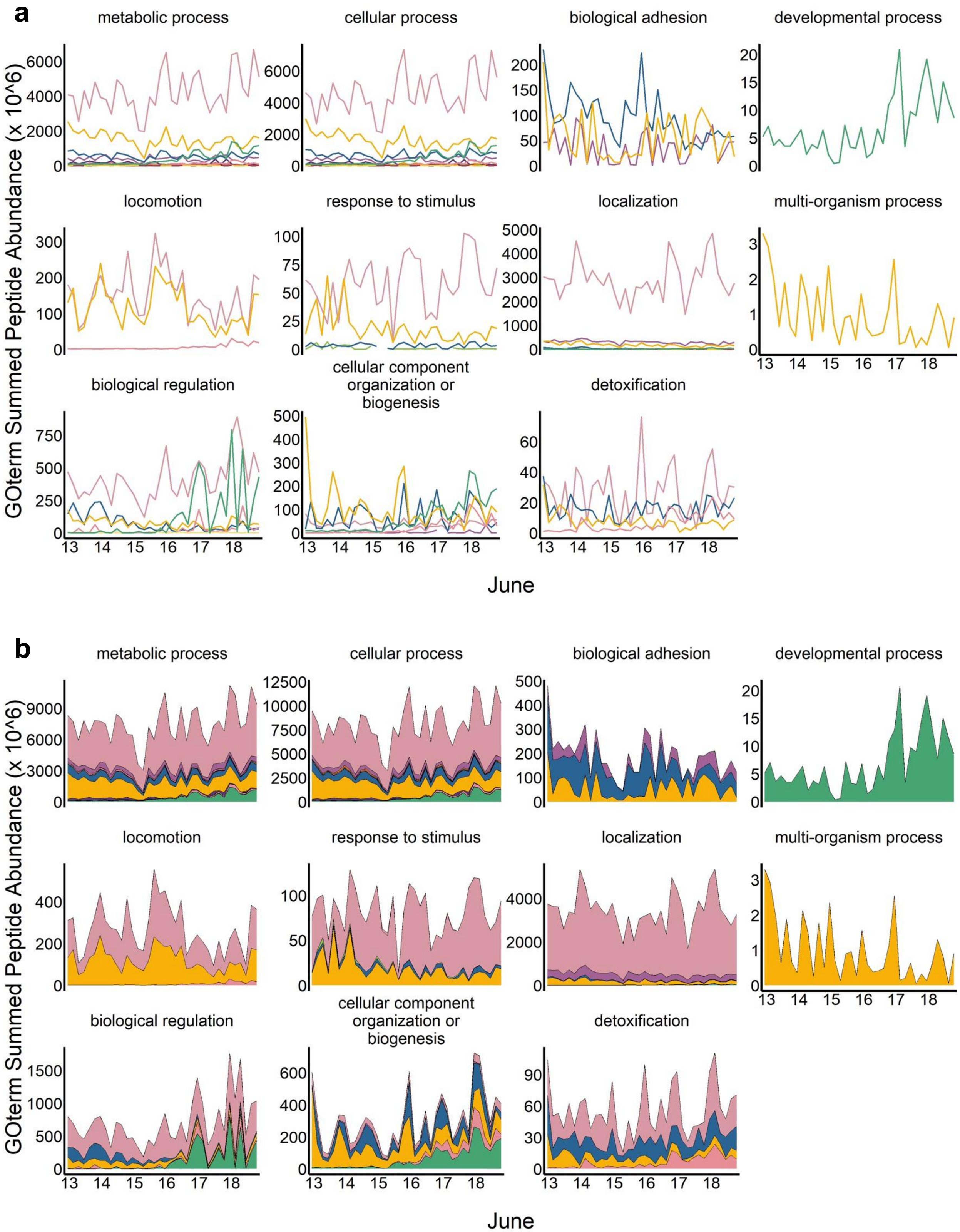
The total abundance of the GO terms directly connected to the root node “biological process” are mostly consistent between classes, though certain classes exhibit unique trends prior to the second *Chaetoceros* bloom. Alphaproteobacteria is the most abundant class represented by total number of peptides, GO terms they are associated with, and relative abundances of both compared to other classes. The next most abundant classes as represented by LCA-assigned peptide abundances are Flavobacteriia, Gammaproteobacteria, Bacilli, Verrucomicrobiae, and Opitutae (Table 4.1; Fig. 3.9). While the functions presented here are chosen to represent highly conserved processes based on their direct connection to the root node "biological process", the number of classes represented by each term varies. The most highly represented GO terms are metabolic process (18 classes) and cellular process (16 classes). On the other hand, "developmental process" is represented by only one class. Terms with higher abundance also experience higher variation in abundance across the time series, as reflected by the individual peptides represented.

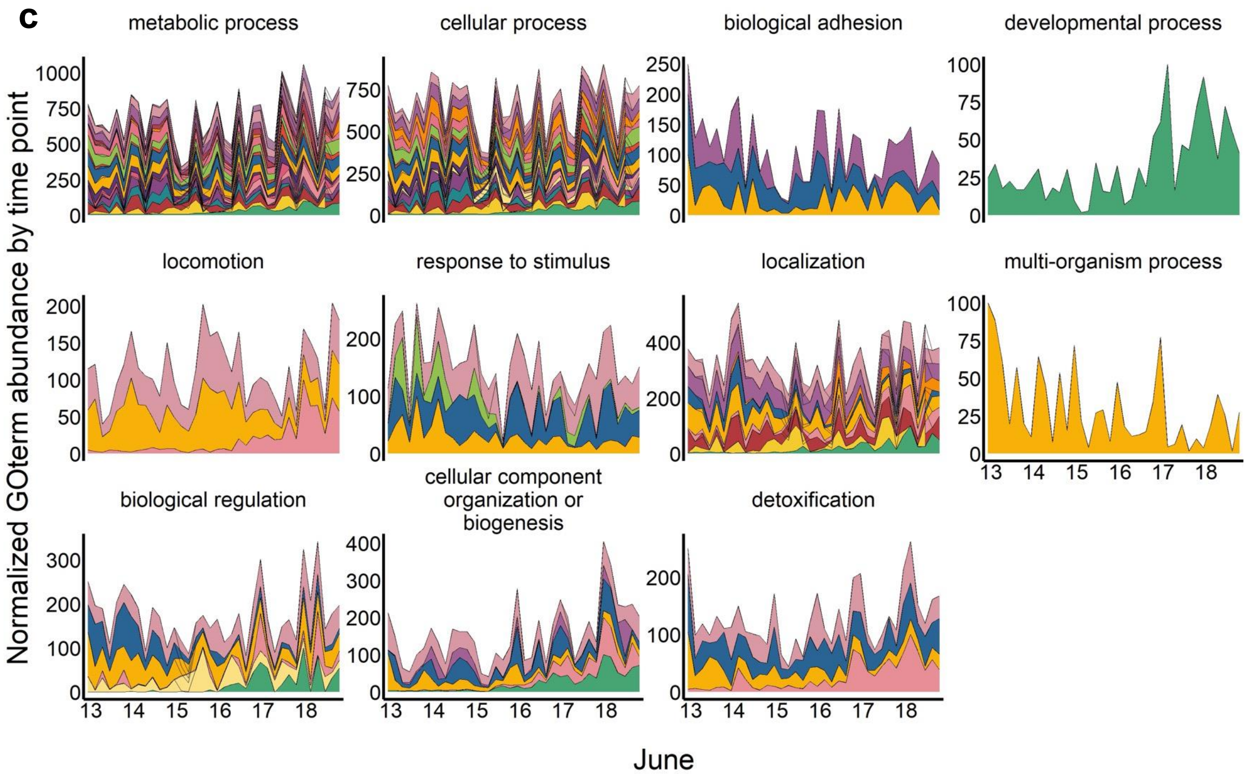
A general trend of increased abundance across the classes Verrucomicrobiae and Opitutae also emerges across high level GO terms (e.g., "biological regulation", "cellular component organization or biogenesis", "developmental process")(Fig. 4.4). Changes in the overall distribution of GO terms across the classes is also evident when GO terms are normalized across the entire time series (Fig. 4d).

Table 4.1. Summary of bacterial classes that undergo a non-reverting state shift within the data set.

Class	HMM state shift	Number GO terms	Number HMM GO terms	Number HMM time points	Total number peptides	General trend
Flavobacteriia	18.21 (06/14)	277	38	7	431	Down
Gammaproteobacteria	18.21 (06/14)	360	79	11	339	Down
Gloeobacteria	18.21 (06/14)	18	18	1	4	Down
Planctomycetia	19.01 (06/15)	10	1	1	6	
Verrucomicrobiae	19.01 (06/15)	203	145	8	91	Up
Opitutae	20.05 (06/16)	125	94	9	39	Up

A detailed summary of each bacterial class that was determined to undergo a non-reverting state shift when a two-state HMM was fit to the total summed peptide abundance, including the timing of the state shift, the total number of GO terms, the number of non-reverting GO terms, the number of time points at which at least one GO term underwent a state shift, the total number of peptides, and the general abundance trend of non-reverting GO terms.


a
metabolic process
cellular process
biological adhesion
developmental process
locomotion
response to stimulus
localization
multi-organism process
biological regulation
cellular component organization or biogenesis
detoxification
GOterm Summed Peptide Abundance (x 10^6)
June
b
metabolic process
cellular process
biological adhesion
developmental process
locomotion
response to stimulus
localization
multi-organism process
biological regulation
cellular component organization or biogenesis
detoxification
GOterm Summed Peptide Abundance (x 10^6)
June


c
metabolic process
cellular process
biological adhesion
developmental process
locomotion
response to stimulus
localization
multi-organism process
biological regulation
cellular component organization or biogenesis
detoxification
Normalized GOterm abundance by time point
June

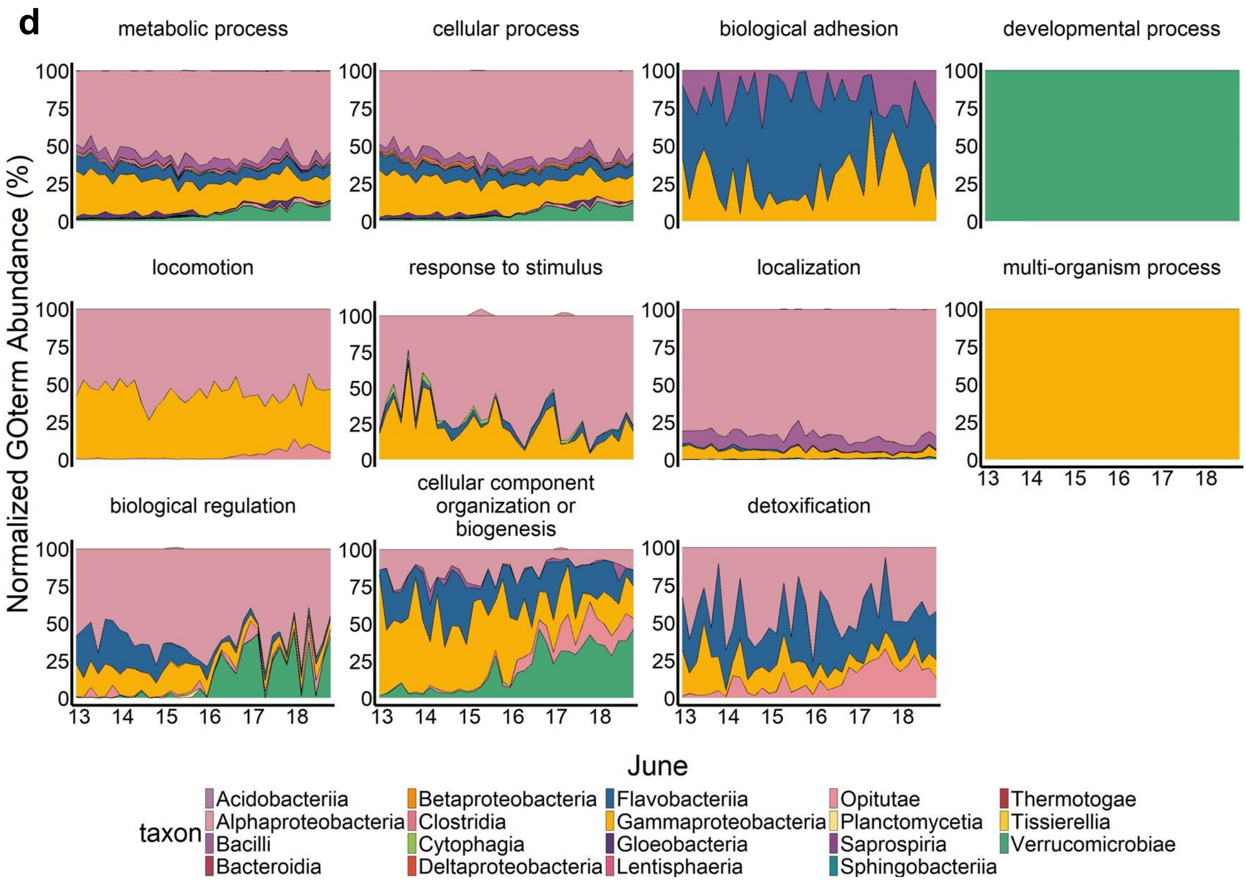

d
metabolic process
cellular process
biological adhesion
developmental process
locomotion
response to stimulus
localization
multi-organism process
biological regulation
cellular component organization or biogenesis
detoxification
Normalized GOterm Abundance (%)
June
taxon
Acidobacteriia
Alphaproteobacteria
Bacilli
Bacteroidia
Betaproteobacteria
Clostridia
Cytophagia
Deltaproteobacteria
Flavobacteriia
Gammaproteobacteria
Gloeobacteria
Lentisphaeria
Opitutae
Planctomycetia
Saprospiria
Sphingobacteriia
Thermotogae
Tissierellia
Verrucomicrobiae

Figure 4.4. Analysis of highly conserved gene ontology (GO) terms across all classes. Class level summed peptide abundances for each GO term were analyzed across the microbiome time series for the GO terms identified as direct child nodes of biological process and visualized as a) line charts and b) stacked area charts. GO terms were also c) normalized by each individual time point and d) by each GO term.

4.3.5 Functional cascade driving non-reverting bacterial classes

To evaluate the composition of the microbiome and its functional output in the context of the surrounding HABS, a hidden Markov model was fit to each class of bacteria to evaluate the summed peptide abundance time series and identify classes that underwent a non-reverting state shift. Of the 29 total classes identified in the microbiome, six were found to undergo a non-reverting state shift: Flavobacteriia, Gammaproteobacteria, Gloeobacteria, Planctomycetia, Verrucomicrobiae, and Opitutae (Fig. 4.5). To further examine the metabolic output of these classes, all GO terms from each non-reverting class were then assessed by HMM.

4.3.6 Flavobacteriia

Flavobacteriia was the most abundant class of bacteria to undergo a non-reverting state shift when total summed peptide abundance was evaluated (Table 4.1). With 431 annotated peptides, Flavobacteriia are the second most abundant class of bacteria in the sampled microbiome. A total of 277 GO terms were annotated to Flavobacteriia, and 38 of those GO terms underwent a non-reverting state shift prior to bloom formation with the first transition June 14th at hour 9:00 and the final transition June 15th at hour 21:00 (Fig. 4.6a). Flavobacteriia functional terms decreased overall, with the exception of protein remodeling and cellular amino acid metabolic process, which increased in abundance (Fig. 4.6b). The GO terms “protein localization” and “translational elongation” underwent a state shift of decreased peptide abundance prior to the total class level shift. “Protein refolding” and “cellular amino acid metabolic process” were the

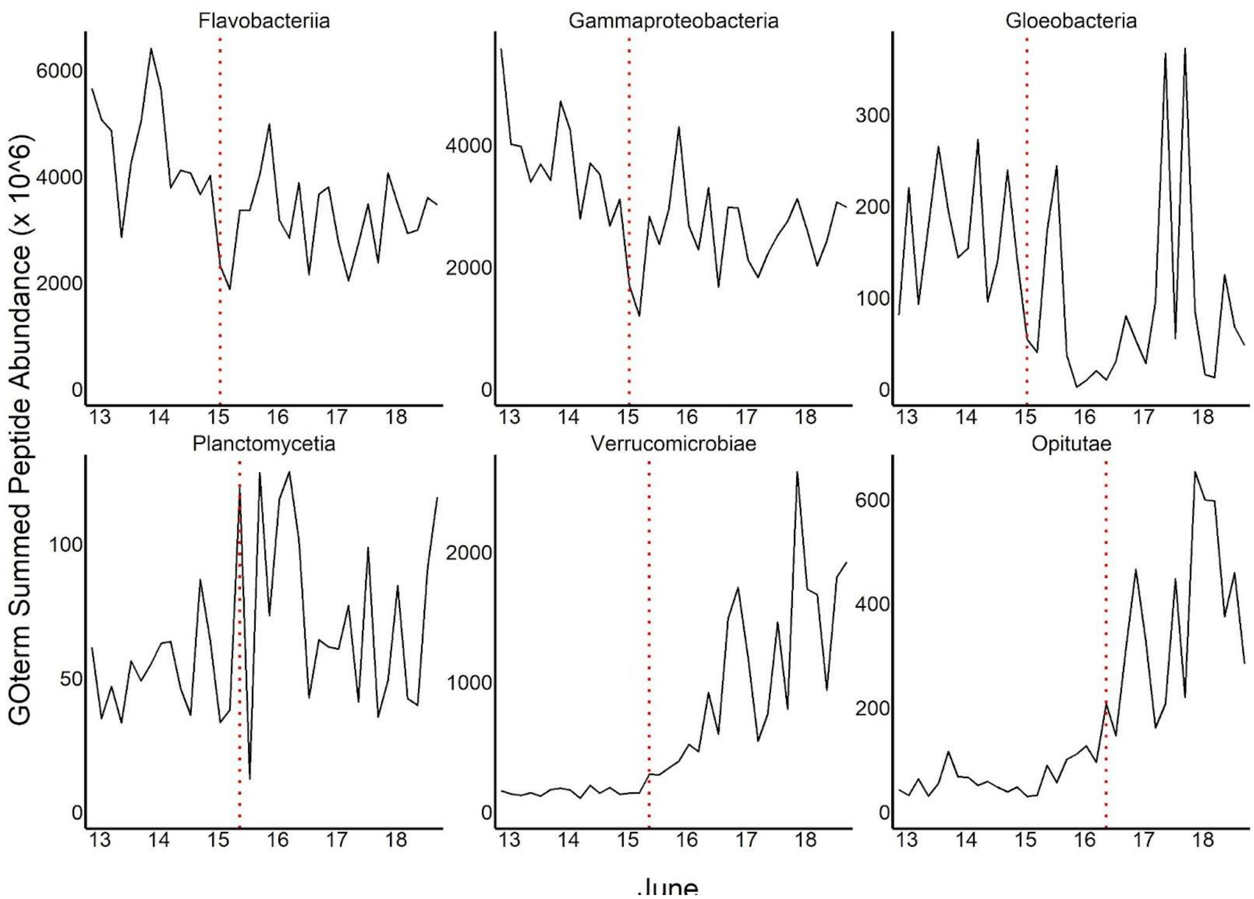


Figure 4.5. Analysis of non-reverting state shifts across individual bacterial classes reveal trends in taxonomy preceding HAB. A two-state hidden Markov model was fit to the summed peptide abundances from each individual bacterial class to identify those terms that underwent a non-reverting state shift prior to HAB initiation. The time series for each non-reverting class is ordered here by timing of the underlying state shift.

only terms to show an increase in abundance on June 15 at hours 09:00 and 13:00, respectively. This change coincided with the observed increase in functional terms in the class Verrucomicrobiae (Fig. 4.8).

4.3.7 Gammaproteobacteria

Gammaproteobacteria was the second most abundant class of bacteria to undergo a non-reverting state shift when total summed peptide abundance was evaluated (Table 4.1). The overall state shift for the class occurred on June 14 at hour 21:00 (Fig. 4.7a), reflecting a decrease

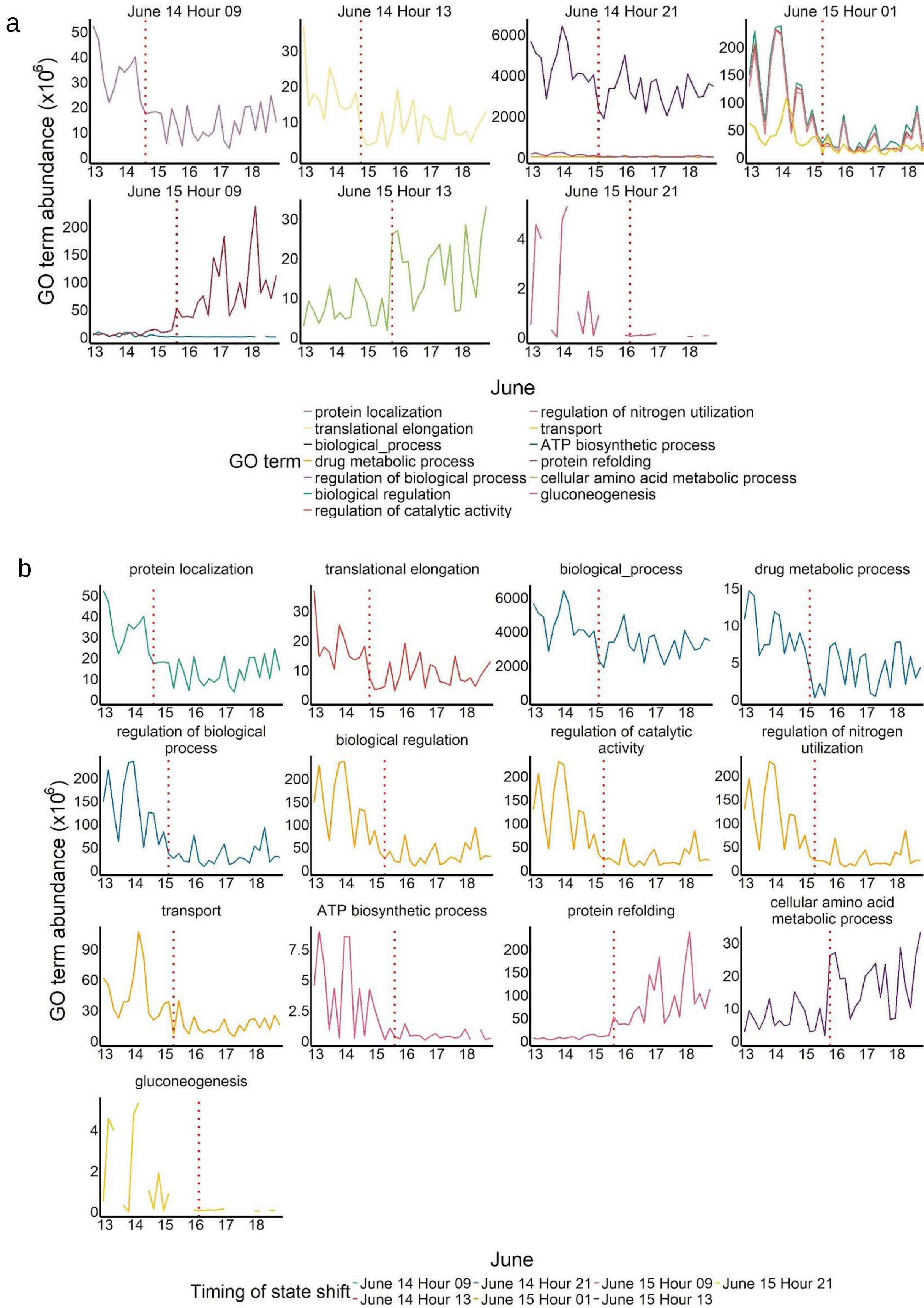
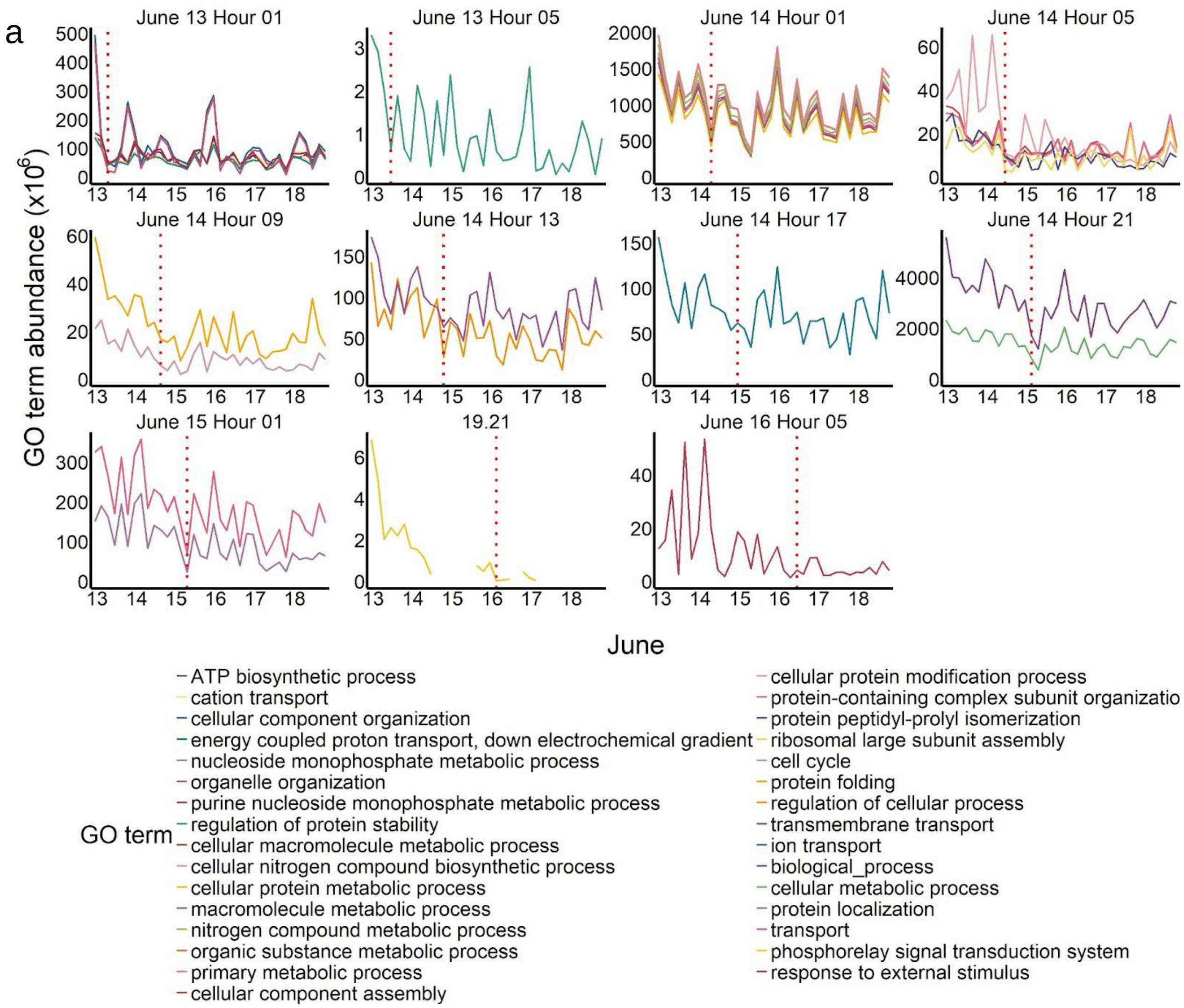

a
June 14 Hour 09
June 14 Hour 13
June 14 Hour 21
June 15 Hour 01
June 15 Hour 09
June 15 Hour 13
June 15 Hour 21
GO term abundance (x10^6)
June
GO term
protein localization
translational elongation
biological_process
drug metabolic process
regulation of biological process
biological regulation
regulation of catalytic activity
regulation of nitrogen utilization
transport
ATP biosynthetic process
protein refolding
cellular amino acid metabolic process
gluconeogenesis
b
protein localization
translational elongation
biological_process
drug metabolic process
regulation of biological process
biological regulation
regulation of catalytic activity
regulation of nitrogen utilization
transport
ATP biosynthetic process
protein refolding
cellular amino acid metabolic process
gluconeogenesis
GO term abundance (x10^6)
June
Timing of state shift
June 14 Hour 09
June 14 Hour 13
June 14 Hour 21
June 15 Hour 01
June 15 Hour 09
June 15 Hour 13
June 15 Hour 21

Figure 4.6. Analysis of non-reverting state shifts across the Flavobacteriia by time point and GO term. All non-reverting GO terms from the class Flavobacteriia, plotted by a) timing of the underlying state shift and b) GO term. GO terms represent summed peptide abundances, with data filtered to include only peptides represented by >50% of the time points.

in peptide abundance matched by all non-reverting GO terms within the class. A total of 360 individual GO terms were annotated to Gammaproteobacteria's 339 total peptides (Table 4.1). Because the two-state hidden Markov model was prone to error during a large multivariate analysis, representative GO terms for all unique GO term time series were selected prior to model fitting to reduce error when evaluating the specific timing of the underlying state shift. A hidden Markov model was fit to the 118 unique time series identified, revealing 79 non-reverting functions represented by 32 unique time series including the highest tiered category, biological process (Fig. 4.7b; Table 4.1). The overall cascade began with a decrease in energy production ("ATP biosynthesis process", "cation transport", etc.), followed by nitrogen related processes, cellular protein modifications, "cell cycle", and finally transport related processes. The final non-reverting function of "response to external stimulus" was observed June 16 at hour 05:00.

4.3.8 Verrucomicrobiae

Verrucomicrobiae had a total of 91 annotated peptides associated with 203 GO terms (Table 4.1); 145 these GO terms underwent a non-reverting state shift prior to the June 18th BSD, the most non-reverting GO terms of any class observed (Fig. 4.8a). These 145 GO terms resulted from 39 unique peptide-abundance time-trends, with state shifts occurring between June 15th at hour 05:00 and June 16th at hour 13:00 (Fig. 4.8a,b). The non-reverting state shift for Verrucomicrobiae as a class based on total summed peptide abundance was identified as occurring on June 15th at hour 05:00, making this the only class to undergo non-reverting state shift in total peptide abundance at the start of its metabolic cascade. This burst of activity was



a
June 13 Hour 01
June 13 Hour 05
June 14 Hour 01
June 14 Hour 05
June 14 Hour 09
June 14 Hour 13
June 14 Hour 17
June 14 Hour 21
June 15 Hour 01
19.21
June 16 Hour 05
GO term abundance (x10^6)
June
GO term
ATP biosynthetic process
cation transport
cellular component organization
energy coupled proton transport, down electrochemical gradient
nucleoside monophosphate metabolic process
organelle organization
purine nucleoside monophosphate metabolic process
regulation of protein stability
cellular macromolecule metabolic process
cellular nitrogen compound biosynthetic process
cellular protein metabolic process
macromolecule metabolic process
nitrogen compound metabolic process
organic substance metabolic process
primary metabolic process
cellular component assembly
cellular protein modification process
protein-containing complex subunit organizatio
protein peptidyl-prolyl isomerization
ribosomal large subunit assembly
cell cycle
protein folding
regulation of cellular process
transmembrane transport
ion transport
biological_process
cellular metabolic process
protein localization
transport
phosphorelay signal transduction system
response to external stimulus

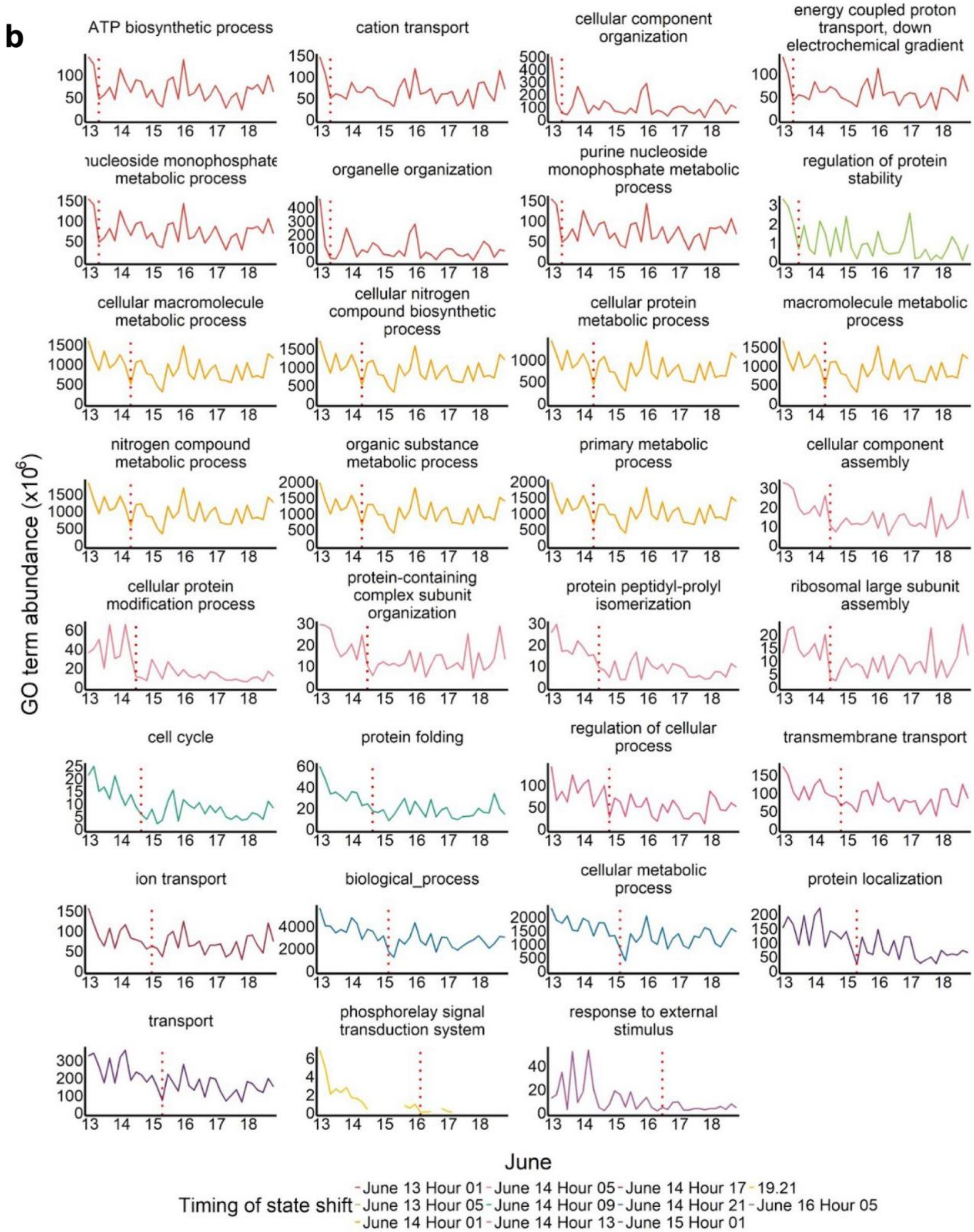
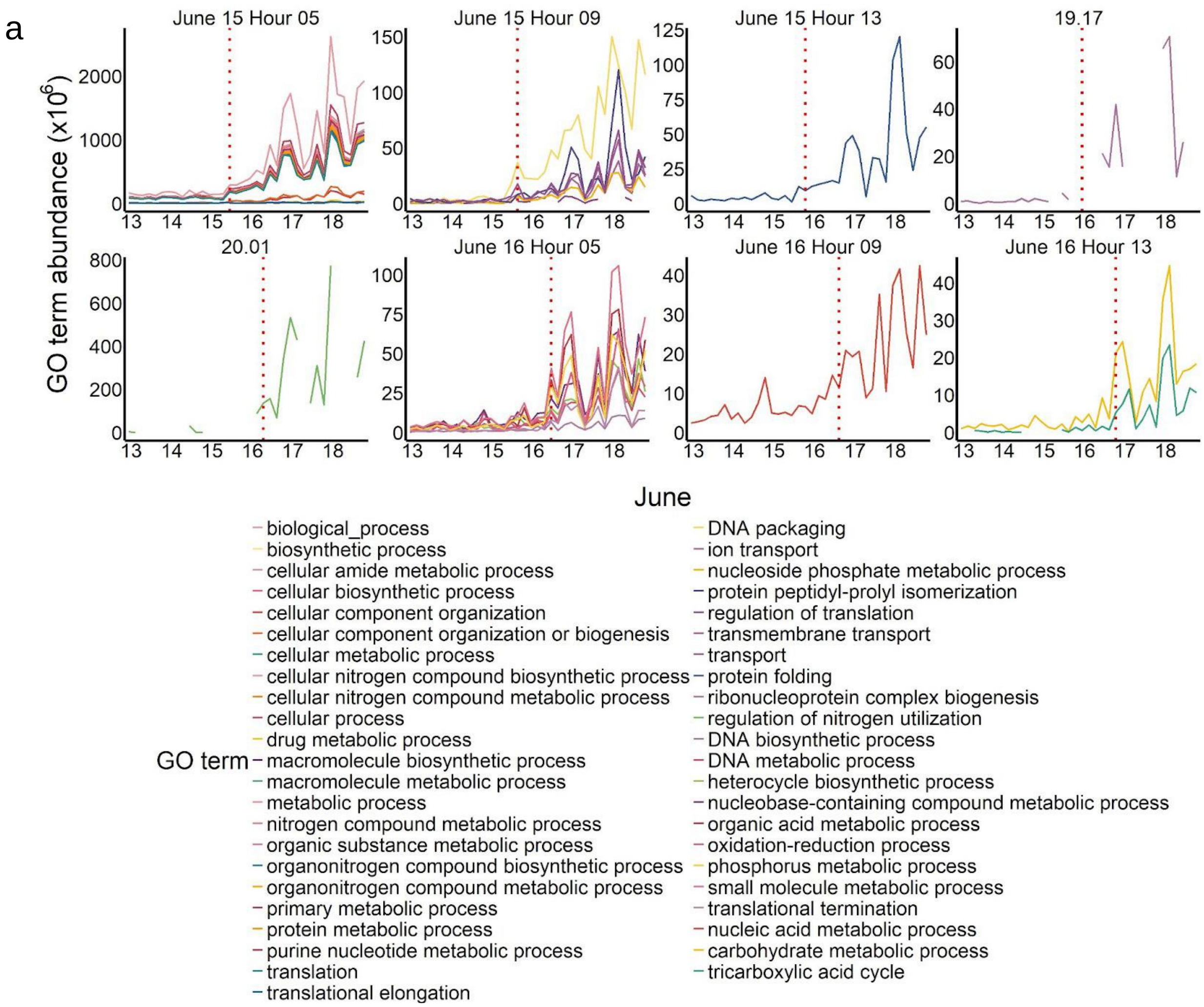
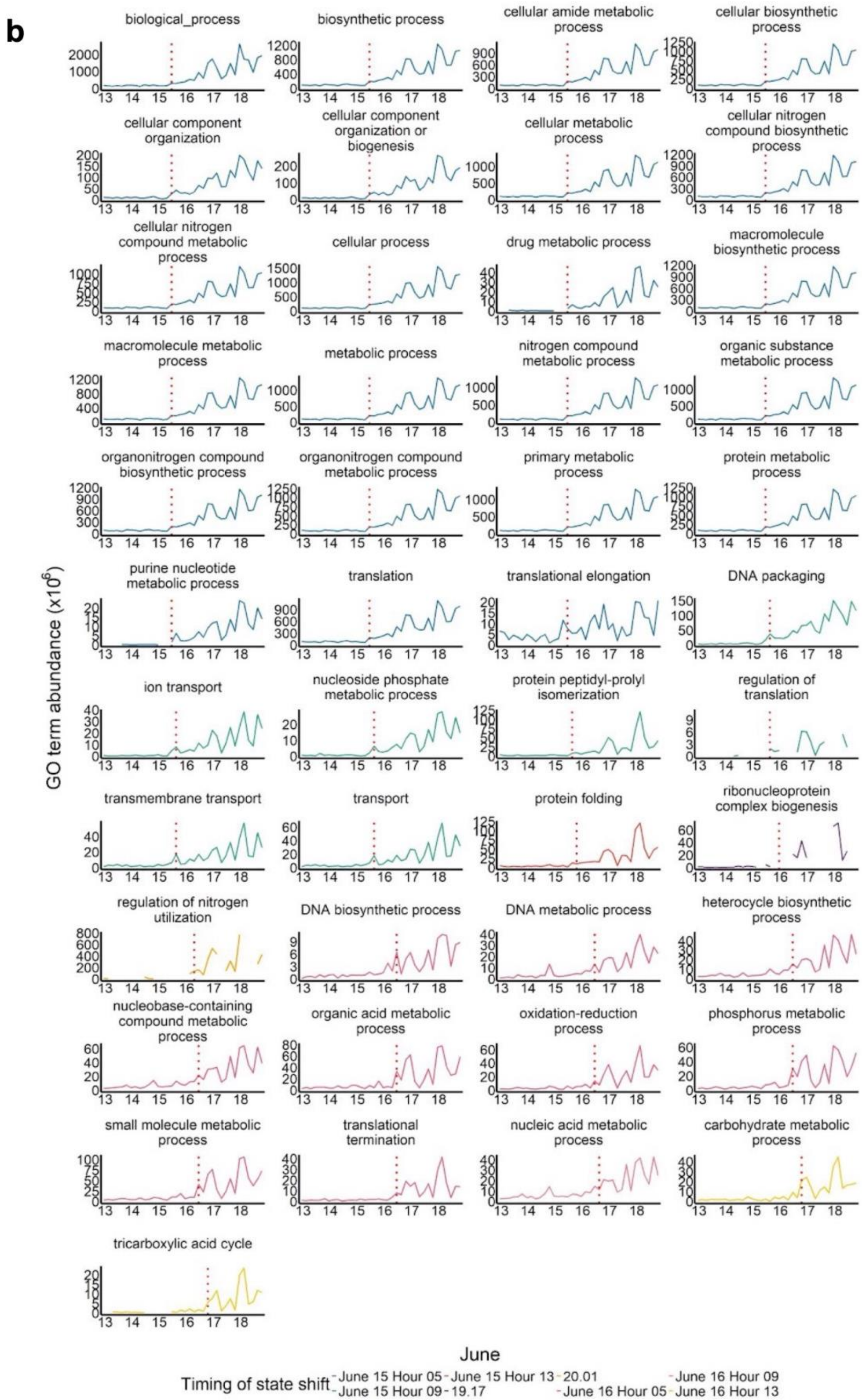


Figure 4.7. Analysis of non-reverting state shifts across the Gammaproteobacteria by time point and GO term. All non-reverting GO terms from the class Gammaproteobacteria, plotted by a) timing of the underlying state shift and b) GO term.


a
June 15 Hour 05
June 15 Hour 09
June 15 Hour 13
19.17
20.01
June 16 Hour 05
June 16 Hour 09
June 16 Hour 13
GO term abundance (x10^6)
June
GO term
biological_process
biosynthetic process
cellular amide metabolic process
cellular biosynthetic process
cellular component organization
cellular component organization or biogenesis
cellular metabolic process
cellular nitrogen compound biosynthetic process
cellular nitrogen compound metabolic process
cellular process
drug metabolic process
macromolecule biosynthetic process
macromolecule metabolic process
metabolic process
nitrogen compound metabolic process
organic substance metabolic process
organonitrogen compound biosynthetic process
organonitrogen compound metabolic process
primary metabolic process
protein metabolic process
purine nucleotide metabolic process
translation
translational elongation
DNA packaging
ion transport
nucleoside phosphate metabolic process
protein peptidyl-prolyl isomerization
regulation of translation
transmembrane transport
transport
protein folding
ribonucleoprotein complex biogenesis
regulation of nitrogen utilization
DNA biosynthetic process
DNA metabolic process
heterocycle biosynthetic process
nucleobase-containing compound metabolic process
organic acid metabolic process
oxidation-reduction process
phosphorus metabolic process
small molecule metabolic process
translational termination
nucleic acid metabolic process
carbohydrate metabolic process
tricarboxylic acid cycle


b
biological_process
biosynthetic process
cellular amide metabolic process
cellular biosynthetic process
cellular component organization
cellular component organization or biogenesis
cellular metabolic process
cellular nitrogen compound biosynthetic process
cellular nitrogen compound metabolic process
cellular process
drug metabolic process
macromolecule biosynthetic process
macromolecule metabolic process
metabolic process
nitrogen compound metabolic process
organic substance metabolic process
organonitrogen compound biosynthetic process
organonitrogen compound metabolic process
primary metabolic process
protein metabolic process
purine nucleotide metabolic process
translation
translational elongation
DNA packaging
ion transport
nucleoside phosphate metabolic process
protein peptidyl-prolyl isomerization
regulation of translation
transmembrane transport
transport
protein folding
ribonucleoprotein complex biogenesis
regulation of nitrogen utilization
DNA biosynthetic process
DNA metabolic process
heterocycle biosynthetic process
nucleobase-containing compound metabolic process
organic acid metabolic process
oxidation-reduction process
phosphorus metabolic process
small molecule metabolic process
translational termination
nucleic acid metabolic process
carbohydrate metabolic process
tricarboxylic acid cycle
GO term abundance (x10^6)
June
Timing of state shift
June 15 Hour 05
June 15 Hour 09
June 15 Hour 13
19.17
20.01
June 16 Hour 05
June 16 Hour 09
June 16 Hour 13

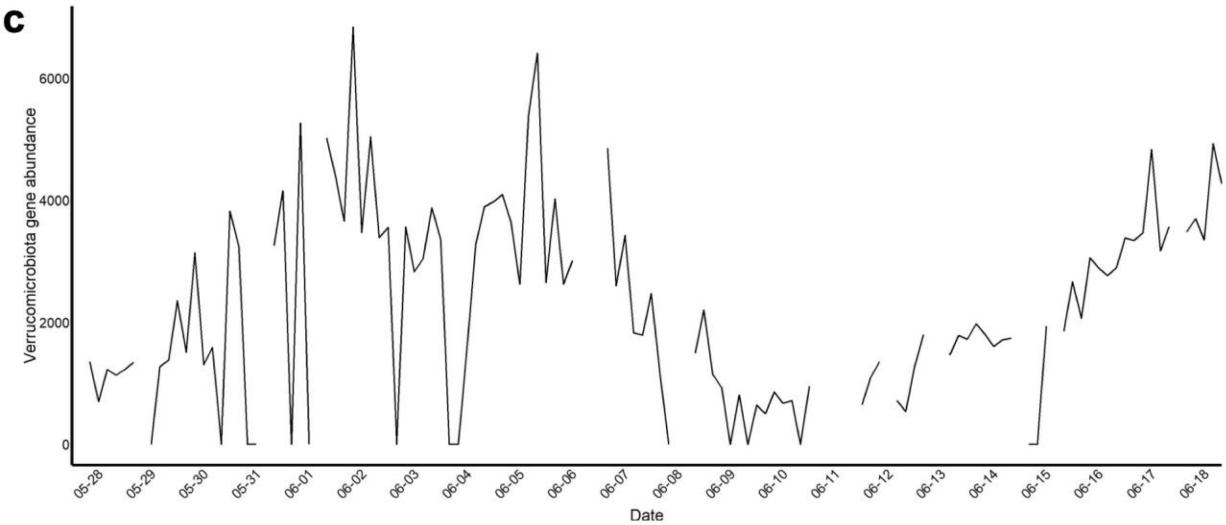


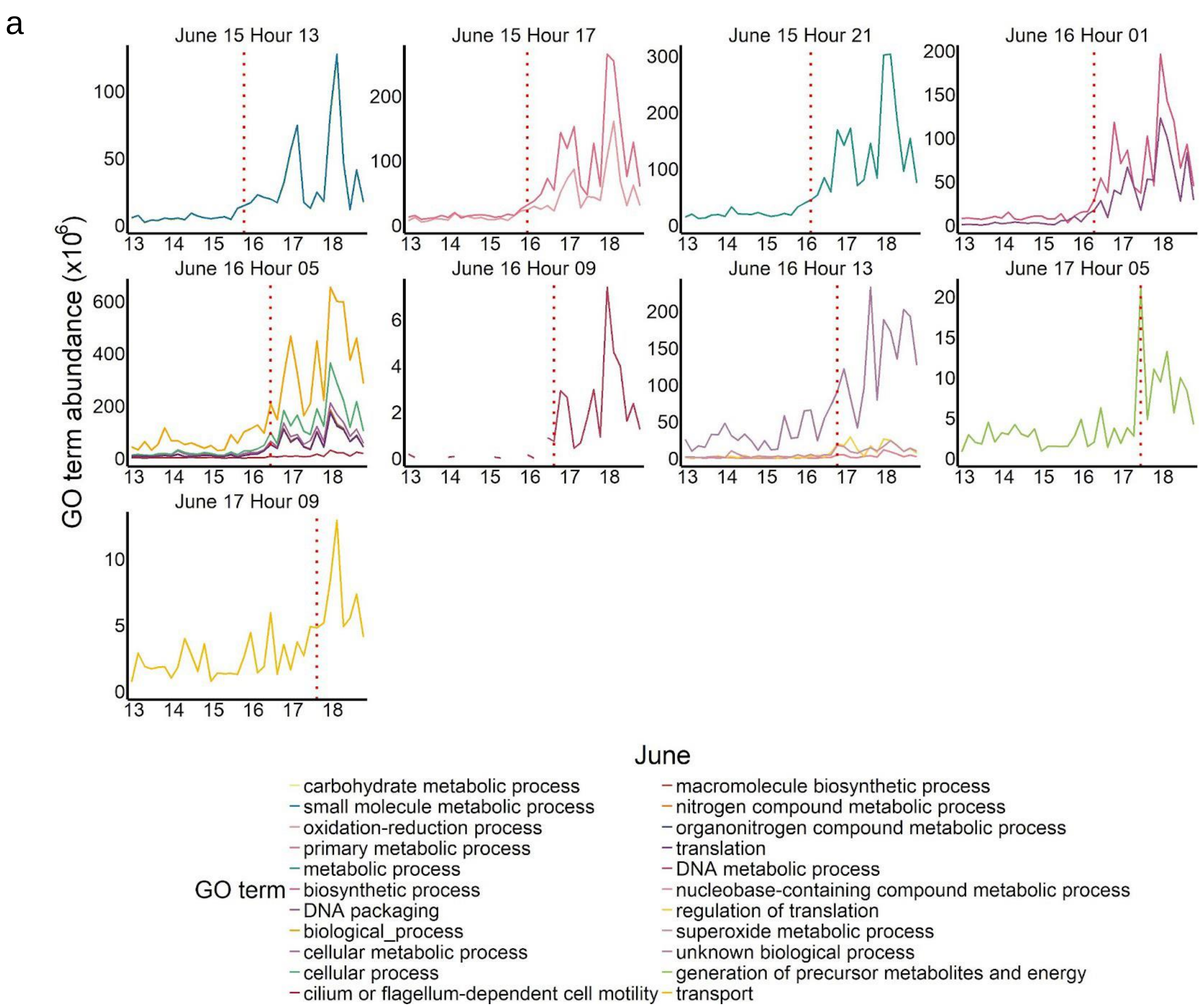
Figure 4.8. Analysis of non-reverting state shifts across the Verrucomicrobiae by time point and GO term. All non-reverting GO terms from the class Verrucomicrobiae, plotted by a) timing of the underlying state shift and b) GO term. c) Gene counts for Verrucomicrobiae across the field season.

also reflected in the overall abundance of GO terms observed, with the highest GO term abundances occurring on June 15 at hour 05:00 while subsequent timepoints had progressively smaller increases in abundance. All non-reverting time series were characterized by a notable increase in peptide abundance immediately following the changepoint. Compared to other classes, Verrucomicrobiae showed a unique increase in the functional term developmental process (Fig. 4.8b), also annotated as cell morphogenesis. The overall increase in peptide abundance was also reflected in the gene counts observed by the class preceding the second bloom, which began to increase on June 8th and continued through the end of the sample collection on June 18th (Fig. 4.8c).

4.3.9 Opitutae

Opitutae had a total of 39 annotated peptides associated with 125 GO terms (Table 4.1), 94 of which underwent a non-reverting state shift prior to the June 18th bloom start date (BSD;

see Chapter 3 for reference). These 94 GO terms mapped back to 25 unique time series, with state shifts occurring between June 15th at hour 13:00 and June 17th at hour 09:00 (Fig. 4.9a). The non-reverting state shift for Opitutae as a class based on total summed peptide abundance occurred on June 16th at hour 05:00 (Fig. 4.9a), making this class the last to undergo a detectable change in state prior to the bloom when evaluated using a two state model. All non-reverting time series were characterized by a notable increase in peptide abundance immediately following the changepoint, similar to the metabolic profile of the class Verrucomicrobiae. Compared to other classes, Opitutae showed a unique increase in functional terms related to cell motility (Fig. 4.9b).



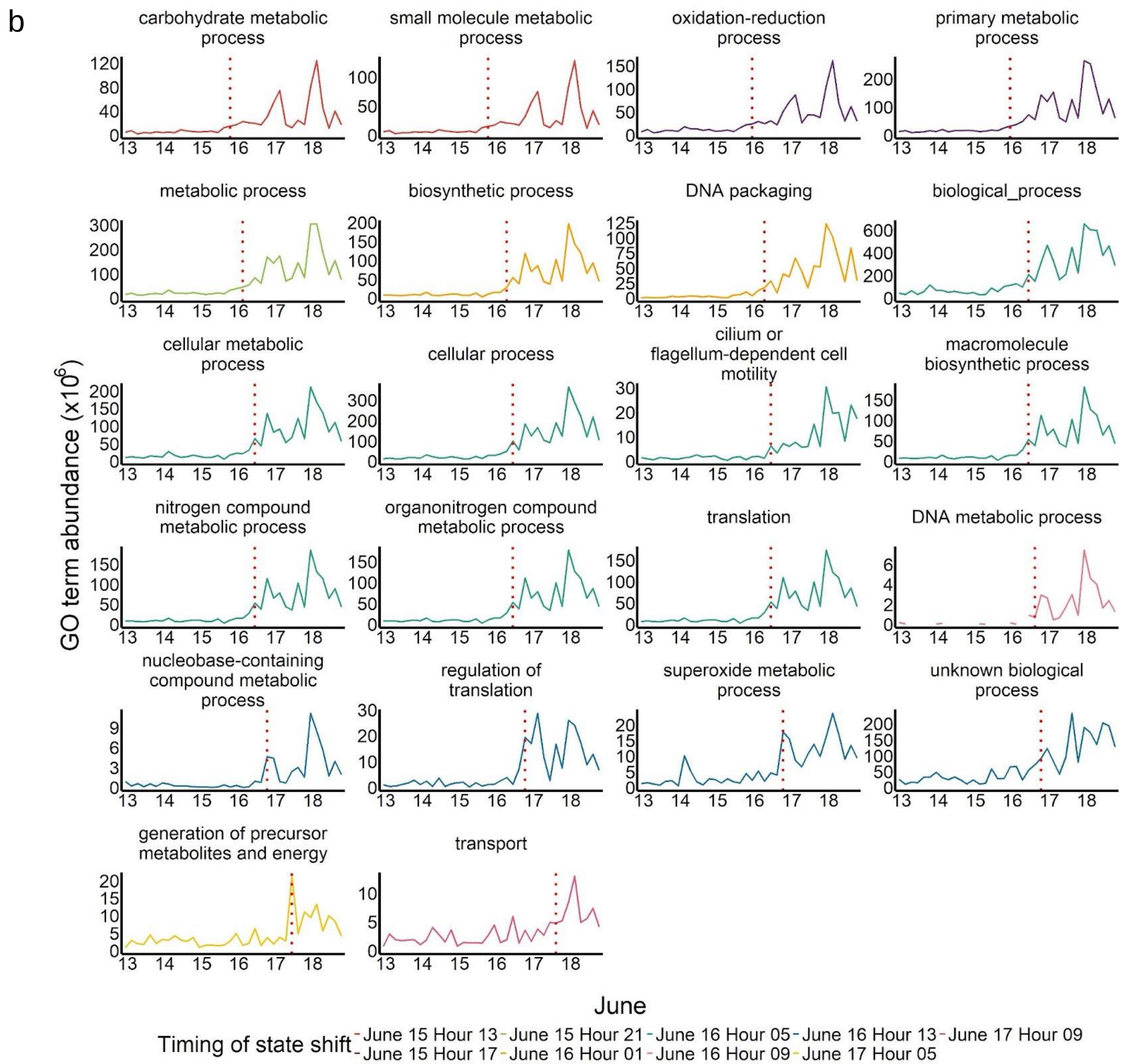


Figure 4.9. Analysis of non-reverting state shifts across the Opitutae by time point and GO term. All non-reverting GO terms from the class Opitutae, plotted by a) timing of the underlying state shift and b) GO term.

4.4 DISCUSSION

Previous studies have evaluated the rapid succession of clades of bacteria closely associated with phytoplankton bloom progression and demise, showing that the emergence of consistent bacterial members is driven by available substrates (Teeling *et al.,* 2016). Alternatively, comparatively few studies have evaluated the dynamics preceding blooms that can trigger bloom onset and even fewer have tracked the distributions of metabolisms among those clades. Here I utilize peptide abundances as an equivalent proxy for the presence of various bacteria and their active functions (Waldbauer *et al.,* 2012; Mikan *et al.,* 2020) to isolate interactions between bacteria as they react and preemptively respond to the conditions surrounding multiple phytoplankton bloom events. A close examination of changing functional output through a metaproteomic analysis of the bacterial microbiome highlights key classes that change surrounding blooms and the metabolic changes they undergo in the time between blooms. This study reveals interesting trends in bacteria influenced by pre- and post-bloom dynamics and also attempts to tackle the complexity of a functional analysis resulting from the interconnectedness of gene ontology terminology.

4.4.1 Multiple bloom cycles influence the bacterial microbiome

To understand the metabolic patterns captured by the mass spectrometry analysis here, it is important to recognize that the field season captured two consecutive *Chaetoceros socialis* blooms. As such, the dynamic functions described herein detail both transitions made prior to a *Chaetoceros* bloom as well as signals from the end of the previous bloom. In the context of known literature, many of the classes represented in the marine ecosystem are consistently prevalent in marine microbiomes as well as recognized degraders of algal bloom byproducts (Brown *et al.,* 2009; Teeling *et al.,* 2016). The first *Chaetoceros* bloom captured in the 2021 field season began on May 31, peaked on June 4, and maintained chlorophyll *a* levels above 25 ug L^{-1} through June 8 (Fig. 4.1a). Immediately preceding the drop to a near zero level of chlorophyll *a* on June 8 at 09:00, there was a spike in bacterial cell count beginning June 7 at 17:00. The dominant classes of bacteria based on gene counts at the time were Alphaproteobacteria, Gammaprotebacteria, and Flavobacteriia (Nunn *et. al.,* 2024).

A hidden Markov model was chosen for this analysis due to the nature of the study, specifically that the dynamics of the system were observed to shift from a stable, low nutrient state rapidly into a HAB. Because of this, a two-state HMM was utilized to take advantage of the high-resolution dynamics to evaluate the precise timing of the shift in microbial functional output preceding the bloom and isolate functions that undergo a significant change. This dynamic change in metabolic profile of the microbiome, both as a whole and at the class level (Fig. 4.5-4.9), revealed a cascade of functions preceding bloom initiation. While the HMM was a valuable model for isolating changepoint-driven state switches, we recognize that further analyses employing other types of models, as well as additional variables like environmental covariates, are necessary to validate the significance of the relationship between the classes. For example, in Chapter 3, a two-state HMM was utilized to identify individual peptides that could predict harmful algal blooms. In Chapter 4, a two state HMM was applied to GO terms by summing related peptide abundances to identify changes in function both across the microbiome and within individual classes. In this way, the two-state model identified both preemptive increases in function preceding the second bloom as well as decreases in function that reflect the demise of the first bloom. Further analyses could include a dynamic factor analysis to identify the covariates driving underlying trends.

4.4.2 Verrucomicrobia phylum increases metabolic activity but differentiates in class-related carbon metabolism strategies

Of the classes that fit a non-reverting two-state HMM, the classes Verrucomicrobiae and Opitutae exhibit a distinct onset of metabolic activity prior to the bloom indicative of a cascade of functional activation immediately prior to the second HAB (Fig. 4.8a-b; Fig. 4.9).

Verrucomicrobiae's increased metabolic activity was coupled with growth as demonstrated by increased cell counts preceding the HAB (Fig. 4.8c). Based on gene counts, the increase in Verrucomicrobiae levels began prior to the period of time analyzed by MS, but the activation of metabolic function didn't begin until later, highlighting the importance of selecting analysis techniques that reflect the active state of the system at the time of capture. Most studies of the microbiome are limited to taxonomic composition, but a study consisting only of metagenomics in this case would not be able to confidently identify the timing of the metabolic cascade that began prior to the bloom. This is particularly important in bloom studies where HAB onset is rapid.

Opitutae activated a cascade of molecular functions starting June 15 beginning with carbon metabolism, followed by chromosome condensation and DNA packaging, then translation and protein processing coupled with an initiation of cell motility related functions, high level cellular process regulation, and finally an increase in electron transport chain and ion transport. Interspersed are increases in functions related to cellular detoxification and the removal of superoxide radicals. This cascade suggests that Opitutae's increase in activity was related to the acquisition of carbon, specifically related to glucose, hexose, monosaccharide, and small molecule metabolic processes. The cascade of processes related to packaging DNA, protein metabolic processes, and translation are likely occurring simultaneously as cells respond to their environment. Multiple terms related to cell motility and locomotion are increased in Opitutae beginning June 16 at hour 05:00, indicating Opitutae may be poised to actively scavenge for resources that other classes receive passively, conferring an advantage prior to bloom formation. As protein production continues to increase, cells activate pathways related to energy acquisition like electron transport chain and ion mobility related proteins. Overall, the increase in Opitutae's peptide abundance follows a general progression of metabolic activity as cells grow in response to resource acquisition.

Verrucomicrobiae also experiences an increase in metabolic activity following a longer progression of cell growth validated by gene counts for the class. Unlike other classes, Verrucomicrobiae's increase in activity correlates with the general class-level peptide abundance as evidenced by the timing of the state shift biological process in tandem with 58 other GO terms including ATP, nitrogen, cellular protein, organic substance, peptide, and ribonucleoside metabolic and biosynthesis processes, as well as general translation on June 15 at hour 05:00. Four hours later, at 09:00, in contrast to Opitutae, Verrucomicrobiae simultaneously increased processes related to chromosome condensation, localization, protein modification, cellular regulation processes, and ion transport (Fig. 4.8b). As time progressed, additional GO terms like ribosome biogenesis, DNA biosynthesis, small molecule metabolic process, and TCA cycle also increased. Within 24 hours, Verrucomicrobiae exhibited an increase in 137 GO terms.

4.4.3 Notable characteristics of the phylum Verrucomicrobia

The phylum Verrucomicrobia was first documented in 1933 and first recognized as a unique phylum in 1997 (Hedlund, Gosink, and Staley 1997). Verrucomicrobia is highly prevalent, near ubiquitous in marine environments (Freitas et al. 2012), and contains a large taxonomic diversity across associated genomes and their sizes. Respective classes are commonly referred to as subdivisions: Verrucomicrobiae (1), Spartobacteria (2), and Opitutae (3,4). The International Census of Marine Microbes (IcoMM) analyzed all major ocean basins across 506 samples and found that Verrucomicrobia was present in 98% of the samples and was on average the sixth most common phylum in the water column (Freitas *et al.,* 2012), making these bacteria of great relevance to studies of marine ecosystems, specifically marine microbiomes. Genomes associated with Opitutae are generally observed to be smaller and cosmopolitan compared to the class Verrucomicrobiae which has the largest genome in the phylum at 2.7 million base pairs.

4.4.4 *Differences between Verrucomicrobiae and Opitutae suggest specialized competitive strategies between classes*

Verrucomicrobiae has been previously associated with phytoplankton blooms, as they possess the capacity for degrading algal byproducts like complex lipopolysaccharides (Orellana *et al.,* 2021). In particular, Verrucomicrobiae is capable of sequestering organic matter and complex polysaccharides released during blooms through the expression of enzymes like glycoside hydrolase, sulfatases, and carbohydrate esterases (Sichert *et al.,* 2020; Orellana *et al.,* 2021). The increase in “carbohydrate derivative biosynthetic process” and “carbohydrate metabolic process” in Verrucomicrobiae preceding the second bloom suggests these processes could also play a role in the bacteria’s response to anticipating oncoming blooms in addition to post-bloom activity. Interestingly, despite the close genetic relationship between the two classes, Opitutae is not associated with phytoplankton blooms. This may be because Opitutae is less documented or prevalent in marine microbiomes.

Disregarding the complex interrelatedness of GO terms and the timing of their state shifts, Verrucomicrobiae and Opitutae shared 75 GO terms; Verrucomicrobiae had more unique GO terms (70 GO terms) identified using a LCA approach, but also more GO terms overall compared to Opitutae (19 GO terms). Terms shared between the two classes from the Verrucomicrobia phylum seemed related to Opitutae’s cell motility and detoxification proteins, while Verrucomicrobiae had many more terms related to purine ribonucleoside processes and ATP synthesis (Fig. 4.8b; Fig. 4.9b).

Additionally, the two Verrucomicrobia classes were uniquely characterized by their specific carbon metabolic processes (Fig. 4.8b; Fig. 4.9b). Opitutae’s metabolism began with an increase in glucose, hexose, and monosaccharide metabolism. These GO terms are related to activation of glycolysis and the pentose phosphate pathway which were activated early in the time series functional cascade. Conversely, Verrucomicrobiae increased citrate and malate metabolism, processes related to the TCA and glyoxylate cycle (Dalcin Martins *et al.,* 2021). Even though these classes are highly genetically related and exhibit similar trends in their metabolic activity and timing, the specific pathways activated differ between the two, emphasizing the importance of looking at active functionality rather than predicted functionality when evaluating biological systems. This may have implications for the bacteria's relationship to phytoplankton blooms, as Verrucomicrobiae are recognized degraders of complex polysaccharides released as bloom byproducts (Cardman *et al.,* 2014; Sichert *et al.,* 2020; Orellana *et al.,* 2022) while Opitutae has no documented association with algal blooms to our knowledge.

4.4.5 Classes that decrease in abundance reflect response driven by loss of productivity after bloom demise

While classes from the Verrucomicrobia phylum experience a pronounced increase in abundance of all related GO terms, the classes Flavobacteriia and Gammaproteobacteria experience a non-reverting state shift in "biological process" corresponding to an overall decrease in the majority of all GO terms identified, respectively. Apart from Alphaproteobacteria (956 peptides), the classes Flavobacteriia (431 peptides) and Gammaproteobacteria (339 peptides) are the second and third most abundant bacteria represented in the microbiome identified using quantitative proteomics (Table 4.1). A smaller proportion of GO terms in both decreasing classes (13.7% Flavobacteriia, 21.9% Gammaproteobacteria) were found to undergo a non-reverting state shift when compared to prominent classes where biological process was increased (71.4% Verrucomicrobiae, 78.4% Opitutae). One possibility is that a general decline in metabolic activity is slower than an increase in metabolic activity. For example, classes that are prominent are more capable of slowing their protein turnover, whereas cells increasing in metabolic activity need to experience a burst to maintain growth. It is also possible that less optimal conditions lead to a general decline in the abundance of Flavobacteriia. Also of note is the timing of decreasing GO terms compared to increasing GO terms. The majority of the GO terms in Gammaproteobacteria and Flavobacteriia undergo a state shift prior to those in Verrucomicrobiae and Opitutae. Additionally, the GO terms in Flavobacteriia that do increase, specifically protein refolding, cellular amino acid metabolic process, and ATP biosynthesis, do so towards the end of the time series around the same time that processes in Verrucomicrobiae and Opitutae begin to increase. This suggests that the trends in metabolic activity across the classes are related to different stimuli, namely the different blooms. Based on these distinct trends, metabolic activity reflected by the changing GO terms in relevant classes will be further analyzed in the context of the bloom which seems to exert the greatest influence.

Both Gammaproteobacteria and Flavobacteriia are commonly associated with algal bloom progression (Morris *et al.,* 2010). The association between Gammaproteobacteria and algal blooms has been well documented, and these bacteria play a large role in marine carbon and nitrogen cycling (Liu and Liu 2020). Gammaproteobacteria are often found in carbon-rich marine environments due to their diversity and abundance of specialized transporters; they have also been found to correlate with the rise of phytoplankton blooms through increased cell abundance (Francis *et al.,* 2021; Vidal-Melgosa *et al.,* 2021). In this system, Gammaproteobacteria-specific peptides were highly abundant with 339 total peptides representing 360 GO terms. While all of the non-reverting GO terms identified in Gammaproteobacteria were found to decrease in abundance, a smaller proportion of terms exhibited a non-reverting state shift compared to other classes like Verrucomicrobiae or Opitutae. Flavobacteriia are another class of bacteria shown to dominate microbial communities across a variety of ecosystems (Williams *et al.,* 2013; Russo *et al.,* 2016). While most of the functional terms in Flavobacteriia decrease, protein refolding and cellular amino acid metabolic process increase, suggesting a coordinated response where the class is responding to some environmental cue that the next bloom is about to begin. This change is also consistent with the timing of the pre-bloom responders Verrucomicrobiae and Opitutae, while all of the functional terms that decrease switch states nearer to the first bloom (Fig. 4.6).

For the dataset analyzed here, these classes appear to reflect trends detailed in the literature in that they seem to have peaked prior to the beginning of the analysis. Figures 4.6 and 4.7 show that both classes are down-regulating their metabolic processes, suggesting that the abundance of biological matter produced in the wake of the first bloom event has been utilized. Nunn *et. al.*, 2024 shows that the first bloom coincided with an influx of nutrients into the water column which could have also influenced the bacterial bloom that peaked prior to the dataset analyzed here. Therefore, a likely explanation for this result is that Gammaproteobacteria and Flavobacteriia are responding to the decrease in available DOM (from the prior date June 1 through the end of the bloom on June 7) and decrease of available nutrients by downregulating most of their metabolic functions.

4.4.6 Potential clues in lower abundance classes

In addition to the classes detailed above, Gloeobacteria and Planctomycetia also underwent a non-reverting state shift in “biological process”, indicating a switch in underlying metabolism. Despite both of these classes having few specific peptides identified as LCA terms, other contributing peptides might be ubiquitous across a range of classes and thus grouped in the overarching category of unannotated peptides. Gloeobacteria only contained one annotated peptide related to terms including oxidation-reduction process, photosynthesis, and organonitrogen compound metabolic process. Gloeobacteria is a class of cyanobacteria with early phylogenetic divergence that has been used in studies to elucidate the early origins of photosynthesis (Saw, Cardona, and Montejano 2021; Grettenberger, 2021), suggesting potentially relevant implications for metabolic scheduling despite a lack of association with phytoplankton blooms. Planctomycetia mapped to two unique peptides, one related to cell redox and homeostasis, and the other only annotated as biological process. Though this class is a known member of many marine and terrestrial environments (Dedysh and Ivanova 2019), it is not commonly associated with algal bloom events directly. While these classes do undergo a detectable state shift, more studies are needed to determine their relevance in the context of microbial community dynamics and their importance in the context of harmful algal blooms.

4.4.7 Annotation tools play a critical role in biological conclusions

Many tools exist for annotating proteomic identifications (Sajulga *et al.,* 2020). We chose to use MetaGOmics (Riffle *et al.,* 2017), but there is the possibility that peptides annotations have been added or updated in the literature since the time of performing the analysis. Additionally, there is the possibility of error; more specifically, it is possible that some of the peptides annotated to bacteria are also found within the eukaryotic fraction of the microbiome. One example includes the presence of GO terms related to photosynthesis, specifically within the classes Opitutae and Gloeobacteria. A literature search shows that Opitutae class does not have members that appear to be capable of photosynthesis. The peptides related to photosynthesis GO terms may actually originate from cyanobacteria, which were noted as present in the microbiome through a flow cytometry analysis (Nunn *et. al.,* 2024). Such ambiguity should be considered in further studies that seek to expand upon the results provided here.

4.4.8 Complexity of GO terms as functional descriptors

While our metaproteomics analysis yielded valuable insight into the functional profile of the microbiome captured along a high-resolution temporal gradient, it is important to recognize that GO term functional analyses are more innately complex than taxonomic analyses of the same system. While both annotation types are hierarchical, GO DAGs display highly interrelated connections between terms. Additionally, a single peptide may map back to more than 30 unique functional terms, making it difficult to draw conclusions about the overall biological system. This complexity was inherent to the analysis performed here. For example, seven GO terms with unique peptide abundance time series annotated as Gammaproteobacteria were shown to undergo a non-reverting state shift on June 14 at hour 01:00 (Fig. 4.7), however the abundance trends of these terms within the class are only identifiable by using time series statistics. This also makes correlation analyses that utilize GO terms as input very difficult, as repeated GO terms that seem highly correlated may stem from a single peptide. This caveat is unavoidable given the prevalence of GO terms as descriptors of gene products but must be taken into account when weighing the significance of functional associations.

In conclusion, analyzing trends in function across the time series revealed interesting functions that may help in future studies of algal blooms. Alphaproteobacteria dominated the microbiome in both number of peptides and overall peptide abundance, remaining stable throughout the entire metaproteomic analysis from June 12 - 18. The next most abundant classes, Flavobacteriia and Gammaproteobacteria, appear to be primarily influenced by the first *Chaetoceros* bloom. These classes underwent a non-reverting state shift on June 14 at hour 21:00 characterized by an overall decrease in functionality likely driven by the bloom demise immediately preceding the analysis. The phylum Verrucomicrobia, specifically the classes Verrucomicrobiae and Opitutae, showed a significant increase in peptide abundance prior to the second *Chaetoceros* bloom. Verrucomicrobiae underwent a non-reverting state shift on June 15 at hour 05:00, with increased peptide abundance across 39 GO terms within a 24 hour period. On June 15 at hour 05:00, Opitutae underwent a non-reverting state shift with 22 GO terms increasing over the following 48 hours. All significant GO terms across Verrucomicrobiae and Opitutae increased in abundance, but the functions prioritized by the two classes differed. While both classes increased carbon acquisition, Verrucomicrobiae activated the TCA and glyoxylate cycle while Opitutae increased metabolic pathways related to glycolysis and the pentose phosphate pathway. Opitutae also uniquely increased cell motility processes. By identifying the timing of these non-reverting biological processes across abundant and relevant bacterial classes, the work here provides a unique characterization of the microbiome's metabolic schedule and how a complex community responds to and influences the dynamics of multiple phytoplankton bloom cycles.

Chapter 5: CLOSING REMARKS

Metaproteomics studies are gaining recognition for their ability to isolate functional changes in a biological system, particularly in microbiome research. The research presented here builds on prior studies by highlighting strategic adaptations to challenging environmental conditions across multiple unique systems. The chapters herein highlight how organisms are able to regulate their functionality to anticipate environmental perturbations and respond with rapid and strategic tactics to survive and even gain a competitive advantage in the face of difficult circumstances. When taken together, this work showcases the applicability of metaproteomics to environmental studies and also highlights a variety of considerations to address in future research of any biological system.

5.1 THE COMPLEXITY OF ANNOTATIONS AND PROTEOMIC DATA ANALYSIS

Biological analyses such as the ones conducted here are directly dependent upon the quality of the data collected. Confident peptide identifications from mass spectra are essential to make biological conclusions, but the value of these peptides is directly linked to the used to generate the data (Timmins-Schiffman *et. al.,* 2017). The ideal database for generating confident identifications is a site- and time-specific metagenome, however this can be difficult for smaller labs with limited resources. Fortunately, advances in high-throughput DNA sequencing are easing the barrier of entry for these types of studies. There is also much potential in the future of de novo sequencing for constructing reliable databases. Researchers pursuing proteomic studies of novel systems should consider the potential benefits of rapidly developing technologies for constructing databases for their own experiments. Conversely, as technologies advance, it will become easier to generate thousands of peptide identifications from any given analysis. A careful balance must be established between obtaining a large number of potentially relevant peptide identifications and narrowing searches to only relevant peptides of interest. Ideally, a proteomic analysis will be performed by searching against the largest possible library of only changing relevant targets.

Data analysis in metaproteomics studies remains difficult and is not a streamlined, automated pipeline yet. Even when confident annotations are established, the inherent complexity of GO terms as detailed in Chapter 4 remains a roadblock when identifying the relevant level at which to draw biological conclusions. Trends in higher level processes like “metabolic process” or “locomotion” can reveal global changes in the microbiome with relevant implications for metabolic scheduling as detailed in Chapter 4. On the other hand, analyzing specific terms revealed by child nodes on a GO-DAG can lead to the identification of unique responses or adaptations as specific taxa change their functionality to gain a competitive edge. Balancing the weight of each level of analysis is crucial for both identifying novel changes occurring as microbiomes respond to environmental perturbations and better characterizing biological systems.

5.2 METAPROTEOMICS IN CONTEXT

While metaproteomic analyses are powerful due to their ability to reveal the active functionality of cells in their natural community setting, the results of any study must be considered in the greater context of the system being analyzed. For example, Chapter 3 highlights the potential of the microbiome for generating intracellular predictive signals. The addition of a dissolved metabolomics analysis to complement this study could elucidate extracellular biochemical signals in the water that may be driving bloom formation. This could be further complemented by a panel testing for the presence of relevant toxins or satellite mapping of chlorophyll levels across the entire time series. Furthermore, there is also potential to map increases in chlorophyll measured via fluorescence signal against the metatranscriptome of the phytoplankton present in the system. All of these variables combined provide a comprehensive view of the changing dynamics that precede the initiation of the harmful algal bloom captured in the 2021 field work. While the focus of the study in Chapter 3 was limited to predictive biomarkers in the form of bacterial peptides, compiling every possible variable analyzed could aid future researchers in pinpointing all contributing factors to HAB initiation.

Though metaproteomics studies are gaining prevalence, the current state of the field of microbiome research is centered around metagenomic studies. Advances in high-throughput DNA sequencing have allowed for exponential progress in reporting on complex systems, especially as costs continue to drop. A literature search of microbiome articles reveals that many researchers report on metrics like community composition and diversity. While these studies capture the innate complexity of a mixed system, they fail to capture what the bacteria or other organisms present are actually doing. Even transcriptomic studies only capture functional potential of the system, these studies are prone to issues due to the high regulation of transcription within cells. Additionally, pipelines for evaluating transcriptomic data tend to select for highly differentially expressed genes that can yield low active protein function and mRNA has been proven as an unreliable indicator for protein activity (Evan *et. al.,* 2015). Metaproteomics analyses are highly valuable as they capture the active functional profile of the cell, yielding promise as an avenue for discovering how communities actually respond to and influence their environment.

When considering the studies presented here in the context of my PhD, I realized that despite the large differences between the studies conducted, the systems analyzed, and the techniques utilized, many of the same proteins shared functions across the studies. One prevalent example was TonB transporters, a major protein pathway that facilitates uptake of a wide range of molecules and nutrients (Pawelek *et al.,* 2006; Shultis *et al.,* 2006; Bolam and Berg, 2018; Fujita *et al.,* 2019). Despite major differences in the bacteria present between studies, certain proteins are preferentially utilized across phylogenies. A logical extension of this observation is that the functionality of any microbiome could be condensed to a system of proteins representative not only of various taxa but of the entire biological system as one unit. In this line of thinking, certain proteins would be expressed at a given time of day, regardless of community structure, due to the functional redundancy present across various taxa. Metabolic scheduling, as explored in detail in Chapter 4, proposes wide ranging implication regarding the timing of sampling in a variety of environmental studies. If sample collection for a study of a marine microbiome is limited to daily, or even monthly, sampling, then the time of day in which the sample is collected could drastically impact the biological conclusions that result from the study. For example, collecting samples once per day, but at different times of day across a time series, could lead to significant changes between samples that result not from a biological response in the system but simply from discrepancies in the experimental design. For this reason, future research in the field should take care to consider the idea of metabolic scheduling across a range of microbiome studies.

5.3 FUTURE DIRECTIONS

5.3.1 The future of life detection research

Biomarkers revealed in Chapter 2 from the analysis on Cp34H revealed metabolic responses to challenges that stretch the adaptive potential even for extremophiles, subsequently revealing relevant biomarkers for life in cold, saline environments. The enrichment analysis detailed in Chapter 2 allowed for isolation of short amino acid chains (k-mers) that could be used to detect present or extant life on other worlds during upcoming space missions. One ongoing future direction is currently being conducted by the Arevalo lab at University of Maryland (Ni *et al.,* 2023). A subset of the k-mers are being synthesized and tested with instrumentation to isolate detectable k-mers in line with NASA's for future life detection missions. The idea of peptides as biomarkers for life open a new avenue for research that could revolutionize the field of astrobiology.

5.3.2 The future of HAB research

Harmful algal bloom research is currently limited to monitoring known sites of bloom activity for the presence of toxin production or biomass accumulation, leading to a period of delay between any bloom start date and confirmed bloom presence that prevents a proper timeline of warning the public against potential hazardous exposure. Additionally, toxin analyses are limited to testing for specific known toxins and could potentially exclude relevant screening for hazards present. Advances in machine learning and technologies like satellite imaging are creating a potential avenue for forecasting that could help advance the field of HAB research. While highly applicable due to the prevalence of publicly available networks of satellite data, the role of the microbiome in influencing bloom activity should not be understated. Additionally, many methods of machine learning rely on the curation of complete datasets that comprehensively capture all variables in the system that may influence bloom development, including biological and environmental factors. The dataset presented here in both Chapters 3 and 4 and Nunn et. al. 2024 is unprecedented in scope and will aid both in the development of future directions presented within the Chapters as well as in potential use for other researchers to test their own modeling approaches against a high-resolution time series preceding bloom development. There are still many potential avenues of scientific investigation that can be pursued with the data presented here.

5.4 CONCLUSION

The field of metaproteomics research has the potential to expand studies ranging from human health applications to the environment to reveal relevant adaptations and strategies for overcoming environmental challenges. In this dissertation, I utilized and helped optimize a wide range of mass spectrometry techniques to analyze and quantify changes in both individual bacteria and complex microbiomes over time as they respond to environmental perturbations. This research was also used to identify and develop potential biomarkers for a wide range of applications. The novel ideas and robust applications presented here are at the forefront of environmental proteomic research and have the potential to expand and exceed the boundaries of microbiome research as we know it.

BIBLIOGRAPHY

Agarwal, Dilip Kumar, Vikas Nandwana, Stephen E. Henrich, Vara Prasad V.N. Josyula, C. Shad Thaxton, Chao Qi, Lacy M. Simons, et al. 2022. “Highly Sensitive and Ultra-Rapid Antigen-Based Detection of SARS-CoV-2 Using Nanomechanical Sensor Platform.” *Biosensors & Bioelectronics* 195 (January):113647. https://doi.org/10.1016/j.bios.2021.113647.

Agogué, Hélène, Dominique Lamy, Phillip R. Neal, Mitchell L. Sogin, and Gerhard J. Herndl. 2011. “Water Mass-Specificity of Bacterial Communities in the North Atlantic Revealed by Massively Parallel Sequencing.” *Molecular Ecology* 20 (2): 258–74. https://doi.org/10.1111/j.1365-294X.2010.04932.x.

Ahern, Olivia M., Kerry A. Whittaker, Tiffany C. Williams, Dana E. Hunt, and Tatiana A. Rynearson. 2021. “Host Genotype Structures the Microbiome of a Globally Dispersed Marine Phytoplankton.” *Proceedings of the National Academy of Sciences of the United States of America* 118 (48): e2105207118. https://doi.org/10.1073/pnas.2105207118.

Akhtar, Shamim, Arshad Khan, Charles D. Sohaskey, Chinnaswamy Jagannath, and Dhiman Sarkar. 2013. “Nitrite Reductase NirBD Is Induced and Plays an Important Role during In Vitro Dormancy of Mycobacterium Tuberculosis.” *Journal of Bacteriology* 195 (20): 4592–99. https://doi.org/10.1128/JB.00698-13.

Akhtar, Naureen, Eugenia Karabika, James R. Kinghorn, Anthony D.M. Glass, Shiela E. Unkles, and Duncan A. Rouch. 2015. “High-Affinity Nitrate/Nitrite Transporters NrtA and NrtB of Aspergillus Nidulans Exhibit High Specificity and Different Inhibitor Sensitivity.” *Microbiology* 161 (Pt 7): 1435–46. https://doi.org/10.1099/mic.0.000088.

Allen, Lisa Zeigler, Eric E Allen, Jonathan H Badger, John P McCrow, Ian T Paulsen, Liam D H Elbourne, Mathangi Thiagarajan, et al. 2012. “Influence of Nutrients and Currents on the Genomic Composition of Microbes across an Upwelling Mosaic.” *The ISME Journal* 6 (7): 1403–14. https://doi.org/10.1038/ismej.2011.201.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. “Basic Local Alignment Search Tool.” *Journal of Molecular Biology* 215 (3): 403–10. https://doi.org/10.1016/S0022-2836(05)80360-2.

Anderson, Donald M., Elizabeth Fensin, Christopher J. Gobler, Alicia E. Hoeglund, Katherine A. Hubbard, David M. Kulis, Jan H. Landsberg, et al. 2021. “Marine Harmful Algal Blooms (HABs) in the United States: History, Current Status and Future Trends.” *Harmful Algae*, Global Harmful Algal Bloom Status Reporting, 102 (February):101975. https://doi.org/10.1016/j.hal.2021.101975.

Anderson, Duwayne M., K. Biemann, Leslie E. Orgel, John Oro, Timothy W. Owen, Garson P. Shulman, Priestley Toulmin Iii, and H. C. Urey. 1972. "Mass Spectrometric Analysis of Organic Compounds, Water and Volatile Constituents in the Atmosphere and Surface of Mars: The Viking Mars Lander." *Icarus* 16 (1): 111–38. https://doi.org/10.1016/0019-1035(72)90140-6.

Andreesen, Jan R. 1994. "Glycine Metabolism in Anaerobes." *Antonie van Leeuwenhoek* 66 (1): 223–37. https://doi.org/10.1007/BF00871641.

Arandia-Gorostidi, Nestor, Anders K. Krabberød, Ramiro Logares, Ina Maria Deutschmann, Renate Scharek, Xosé Anxelu G. Morán, Felipe González, and Laura Alonso-Sáez. 2022. "Novel Interactions Between Phytoplankton and Bacteria Shape Microbial Seasonal Dynamics in Coastal Ocean Waters." *Frontiers in Marine Science* 9 (July). https://doi.org/10.3389/fmars.2022.901201.

Arandia-Gorostidi, Nestor, Anders K. Krabberød, Ramiro Logares, Ina Maria Deutschmann, Renate Scharek, Xosé Anxelu G. Morán, Felipe González, and Laura Alonso-Sáez. 2022. "Novel Interactions Between Phytoplankton and Bacteria Shape Microbial Seasonal Dynamics in Coastal Ocean Waters." *Frontiers in Marine Science* 9 (July). https://doi.org/10.3389/fmars.2022.901201.

Ardyna, Mathieu, and Kevin Robert Arrigo. 2020. "Phytoplankton Dynamics in a Changing Arctic Ocean." *Nature Climate Change* 10 (10): 892–903. https://doi.org/10.1038/s41558-020-0905-y.

Arevalo, Ricardo, Laura Selliez, Christelle Briois, Nathalie Carrasco, Laurent Thirkell, Barnabé Cherville, Fabrice Colin, et al. 2018. "An Orbitrap-Based Laser Desorption/Ablation Mass Spectrometer Designed for Spaceflight." *Rapid Communications in Mass Spectrometry* 32 (21): 1875–86. https://doi.org/10.1002/rcm.8244.

Arevalo Jr, R., Southard, A.E., Grubisic, A., Danell, R., Gundersen, C., Minasola, N.A., Yu, A.W., Fahey, M., Getty, S., Ni, Z. and Briois, C. 2019. "CORALS: Characterization of Ocean Realms and Life Signatures." *AGU Fall Meeting Abstracts 2019*: P34C-04.

Arevalo, Ricardo, Ziqin Ni, and Ryan M. Danell. 2020. "Mass Spectrometry and Planetary Exploration: A Brief Review and Future Projection." *Journal of Mass Spectrometry* 55 (1): e4454. https://doi.org/10.1002/jms.4454.

Ayrapetyan, Mesrop, and James D Oliver. 2016. "The Viable but Non-Culturable State and Its Relevance in Food Safety." *Current Opinion in Food Science*, Food microbiology • Functional foods and nutrition, 8 (April):127–33. https://doi.org/10.1016/j.cofs.2016.04.010.

Bar-Shalom, Rinat, Andrey Rozenberg, Matan Lahyani, Babak Hassanzadeh, Gobardhan Sahoo, Markus Haber, Ilia Burgsdorf, et al. 2023. “Rhodopsin-Mediated Nutrient Uptake by Cultivated Photoheterotrophic Verrucomicrobiota.” *The ISME Journal* 17 (7): 1063–73. https://doi.org/10.1038/s41396-023-01412-1.

Bartlett, Gail J., Craig T. Porter, Neera Borkakoti, and Janet M. Thornton. 2002. “Analysis of Catalytic Residues in Enzyme Active Sites.” *Journal of Molecular Biology* 324 (1): 105–21. https://doi.org/10.1016/S0022-2836(02)01036-7.

Barton, Andrew D., Andrew J. Irwin, Zoe V. Finkel, and Charles A. Stock. 2016. “Anthropogenic Climate Change Drives Shift and Shuffle in North Atlantic Phytoplankton Communities.” *Proceedings of the National Academy of Sciences of the United States of America* 113 (11): 2964–69. https://doi.org/10.1073/pnas.1519080113.

Bateman, Nicholas W., Scott P. Goulding, Nicholas J. Shulman, Avinash K. Gadok, Karen K. Szumlinski, Michael J. MacCoss, and Christine C. Wu. 2014. “Maximizing Peptide Identification Events in Proteomic Workflows Using Data-Dependent Acquisition (DDA).” *Molecular & Cellular Proteomics: MCP* 13 (1): 329–38. https://doi.org/10.1074/mcp.M112.026500.

Béjà, Oded, L. Aravind, Eugene V. Koonin, Marcelino T. Suzuki, Andrew Hadd, Linh P. Nguyen, Stevan B. Jovanovich, et al. 2000. “Bacterial Rhodopsin: Evidence for a New Type of Phototrophy in the Sea.” *Science* 289 (5486): 1902–6. https://doi.org/10.1126/science.289.5486.1902.

Bentum, Mirjam van, and Matthias Selbach. 2021. “An Introduction to Advanced Targeted Acquisition Methods.” *Molecular & Cellular Proteomics : MCP* 20 (October):100165. https://doi.org/10.1016/j.mcpro.2021.100165.

Berdalet, Elisa, Lora E. Fleming, Richard Gowen, Keith Davidson, Philipp Hess, Lorraine C. Backer, Stephanie K. Moore, Porter Hoagland, and Henrik Enevoldsen. 2015. “Marine Harmful Algal Blooms, Human Health and Wellbeing: Challenges and Opportunities in the 21st Century.” *Journal of the Marine Biological Association of the United Kingdom. Marine Biological Association of the United Kingdom* 2015. https://doi.org/10.1017/S0025315415001733.

Bern, Marshall, and Yong J. Kil. 2011. “Comment on ‘Unbiased Statistical Analysis for Multi-Stage Proteomic Search Strategies.’” *Journal of Proteome Research* 10 (4): 2123–27. https://doi.org/10.1021/pr101143m.

Bianco, Colleen M., Kathrin S. Fröhlich, and Carin K. Vanderpool. 2019. “Bacterial Cyclopropane Fatty Acid Synthase mRNA Is Targeted by Activating and Repressing Small RNAs.” *Journal of Bacteriology* 201 (19). https://doi.org/10.1128/JB.00461-19.

Blanvillain, Servane, Damien Meyer, Alice Boulanger, Martine Lautier, Catherine Guynet, Nicolas Denancé, Jacques Vasse, Emmanuelle Lauber, and Matthieu Arlat. 2007. "Plant Carbohydrate Scavenging through TonB-Dependent Receptors: A Feature Shared by Phytopathogenic and Aquatic Bacteria." *PLoS ONE* 2 (2): e224. https://doi.org/10.1371/journal.pone.0000224.

Blodget, R. 2011. "Bacteriological Analytical Manual, Dataset 2, Most Probable Number from Serial Dilutions." FDA.

Boetius, Antje, Alexandre M. Anesio, Jody W. Deming, Jill A. Mikucki, and Josephine Z. Rapp. 2015. "Microbial Ecology of the Cryosphere: Sea Ice and Glacial Habitats." *Nature Reviews Microbiology* 13 (11): 677–90. https://doi.org/10.1038/nrmicro3522.

Bolam, David N, and Bert van den Berg. 2018. "TonB-Dependent Transport by the Gut Microbiota: Novel Aspects of an Old Problem." *Current Opinion in Structural Biology*, Engineering and design: New applications • Membranes, 51 (August):35–43. https://doi.org/10.1016/j.sbi.2018.03.001.

Bowman, John P. 2017. "Genomics of Psychrophilic Bacteria and Archaea." In *Psychrophiles: From Biodiversity to Biotechnology*, edited by Rosa Margesin, 345–87. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-57057-0_15.

BOWMAN, JOHN P., JOHN J. GOSINK, SHAREE A. McCAMMON, TOM E. LEWIS, DAVID S. NICHOLS, PETER D. NICHOLS, JENNY H. SKERRATT, JIM T. STALEY, and TOM A. McMEEKIN. 1998. "Colwellia Demingiae Sp. Nov., Colwellia Hornerae Sp. Nov., Colwellia Rossensis Sp. Nov. and Colwellia Psychrotropica Sp. Nov.: Psychrophilic Antarctic Species with the Ability to Synthesize Docosahexaenoic Acid (22:Ω63)." *International Journal of Systematic and Evolutionary Microbiology,* 48 (4): 1171–80. https://doi.org/10.1099/00207713-48-4-1171.

Boyd, P. W., P. W. Dillingham, C. M. McGraw, E. A. Armstrong, C. E. Cornwall, Y.-y Feng, C. L. Hurd, et al. 2016. "Physiological Responses of a Southern Ocean Diatom to Complex Future Ocean Conditions." *Nature Climate Change* 6 (2): 207–13. https://doi.org/10.1038/nclimate2811.

Brody, Sarah R., M. Susan Lozier, and John P. Dunne. 2013. "A Comparison of Methods to Determine Phytoplankton Bloom Initiation." *Journal of Geophysical Research: Oceans* 118 (5): 2345–57. https://doi.org/10.1002/jgrc.20167.

Brooks, Bryan W., James M. Lazorchak, Meredith D.A. Howard, Mari-Vaughn V. Johnson, Steve L. Morton, Dawn A.K. Perkins, Euan D. Reavie, Geoffrey I. Scott, Stephanie A. Smith, and Jeffery A. Steevens. 2016. "Are Harmful Algal Blooms Becoming the Greatest Inland Water

Quality Threat to Public Health and Aquatic Ecosystems?" *Environmental Toxicology and Chemistry* 35 (1): 6–13. https://doi.org/10.1002/etc.3220.

Brown, Mark V., Gayle K. Philip, John A. Bunge, Matthew C. Smith, Andrew Bissett, Federico M. Lauro, Jed A. Fuhrman, and Stuart P. Donachie. 2009. "Microbial Community Structure in the North Pacific Ocean." *The ISME Journal* 3 (12): 1374–86. https://doi.org/10.1038/ismej.2009.86.

Brunner, Andreas-David, Marvin Thielert, Catherine Vasilopoulou, Constantin Ammar, Fabian Coscia, Andreas Mund, Ole B. Hoerning, et al. 2022. "Ultra-High Sensitivity Mass Spectrometry Quantifies Single-Cell Proteome Changes upon Perturbation." *Molecular Systems Biology* 18 (3): e10798. https://doi.org/10.15252/msb.202110798.

Buchan, Alison, Gary R. LeCleir, Christopher A. Gulvik, and José M. González. 2014. "Master Recyclers: Features and Functions of Bacteria Associated with Phytoplankton Blooms." *Nature Reviews Microbiology* 12 (10): 686–98. https://doi.org/10.1038/nrmicro3326.

Bünger, Wiebke, Xun Jiang, Jana Müller, Thomas Hurek, and Barbara Reinhold-Hurek. 2020. "Novel Cultivated Endophytic Verrucomicrobia Reveal Deep-Rooting Traits of Bacteria to Associate with Plants." *Scientific Reports* 10 (1): 8692. https://doi.org/10.1038/s41598-020-65277-6.

Calado, Cecília R. C. 2022. "Antigenic and Conserved Peptides from Diverse Helicobacter Pylori Antigens." *Biotechnology Letters* 44 (3): 535–45. https://doi.org/10.1007/s10529-022-03238-x.

Cardman, Z., C. Arnosti, A. Durbin, K. Ziervogel, C. Cox, A. D. Steen, and A. Teske. 2014. "Verrucomicrobia Are Candidates for Polysaccharide-Degrading Bacterioplankton in an Arctic Fjord of Svalbard." *Applied and Environmental Microbiology* 80 (12): 3749–56. https://doi.org/10.1128/AEM.00899-14.

Carr, Michael H., Michael J. S. Belton, Clark R. Chapman, Merton E. Davies, Paul Geissler, Richard Greenberg, Alfred S. McEwen, et al. 1998. "Evidence for a Subsurface Ocean on Europa." *Nature* 391 (6665): 363–65. https://doi.org/10.1038/34857.

Castillo-Keller, Maria, Phu Vuong, and Rajeev Misra. 2006. "Novel Mechanism of Escherichia Coli Porin Regulation." *Journal of Bacteriology* 188 (2): 576–86. https://doi.org/10.1128/JB.188.2.576-586.2006.

Cezairliyan, Brent, and Frederick M. Ausubel. 2017. "Investment in Secreted Enzymes during Nutrient-Limited Growth Is Utility Dependent." *Proceedings of the National Academy of Sciences of the United States of America* 114 (37): E7796–7802. https://doi.org/10.1073/pnas.1708580114.

Chattopadhyay, M. K. 2006. “Mechanism of Bacterial Adaptation to Low Temperature.” *Journal of Biosciences* 31 (1): 157–65. https://doi.org/10.1007/BF02705244.

Chen, Huirong, Junjun JIang, Fajun Jiang, Shuangfei Li, and Zhangli Hu. 2021. “Temporal Variability of Free-Living Microbial Culturability and Community Composition after an Akashiwo Sanguinea Bloom in Shenzhen, China.” *Ecotoxicology* 30 (5): 975–85. https://doi.org/10.1007/s10646-021-02407-4.

Chen, H. 2018. “VennDiagram: Generate High-Resolution Venn and Euler Plots.” R package version 1.6.20. URL https://CRAN.R-project.org/package=VennDiagram.

Chen, Ying, and Fengping Wang. 2015. “Insights on Nitrate Respiration by Shewanella.” *Frontiers in Marine Science* 1:80. https://doi.org/10.3389/fmars.2014.00080.

Choi, Hyungwon, Sinae Kim, Damian Fermin, Chih-Chiang Tsou, and Alexey I. Nesvizhskii. 2015. “QPROT: Statistical Method for Testing Differential Expression Using Protein-Level Intensity Data in Label-Free Quantitative Proteomics.” *Journal of Proteomics* 129 (November):121–26. https://doi.org/10.1016/j.jprot.2015.07.036.

Choi, Umji, and Chang-Ro Lee. 2019. “Distinct Roles of Outer Membrane Porins in Antibiotic Resistance and Membrane Integrity in Escherichia Coli.” *Frontiers in Microbiology* 10. https://doi.org/10.3389/fmicb.2019.00953.

Clark, S. C., J. Granger, A. Mastorakis, A. Aguilar-Islas, and M. G. Hastings. 2020. “An Investigation Into the Origin of Nitrate in Arctic Sea Ice.” *Global Biogeochemical Cycles* 34 (2): e2019GB006279. https://doi.org/10.1029/2019GB006279.

Cobb, Alyssa K., and Ralph E. Pudritz. 2014. “Nature’s Starships. I. Observed Abundances and Relative Frequencies of Amino Acids in Meteorites.” *The Astrophysical Journal* 783 (2): 140. https://doi.org/10.1088/0004-637X/783/2/140.

Collins, R. Eric, Shelly D. Carpenter, and Jody W. Deming. 2008. “Spatial Heterogeneity and Temporal Dynamics of Particles, Bacteria, and pEPS in Arctic Winter Sea Ice.” *Journal of Marine Systems*, Sea ice and life in a river-influenced arctic shelf ecosystem, 74 (3): 902–17. https://doi.org/10.1016/j.jmarsys.2007.09.005.

Collins, R. Eric, and Jody W. Deming. 2013. “An Inter-Order Horizontal Gene Transfer Event Enables the Catabolism of Compatible Solutes by Colwellia Psychrerythraea 34H.” *Extremophiles* 17 (4): 601–10. https://doi.org/10.1007/s00792-013-0543-7.

Collins, Roy E. 2009. “Microbial Evolution In Sea Ice: Communities To Genes.” Thesis. https://digital.lib.washington.edu:443/researchworks/handle/1773/15544.

Compant, Stéphane, Abdul Samad, Hanna Faist, and Angela Sessitsch. 2019. “A Review on the Plant Microbiome: Ecology, Functions, and Emerging Trends in Microbial Application.”

Journal of Advanced Research, Special Issue on Plant Microbiome, 19 (September):29–37. https://doi.org/10.1016/j.jare.2019.03.004.

Conway, Jake, and Nils Gehlenborg. 2019. "UpSetR: A More Scalable Alternative to Venn and Euler Diagrams for Visualizing Intersecting Sets." https://cran.r-project.org/web/packages/UpSetR/index.html.

Cox, G. F. N., and W. F. Weeks. 1983. "Equations for Determining the Gas and Brine Volumes in Sea-Ice Samples | Journal of Glaciology | Cambridge Core." 29 (102): 306–16. 1986. "Changes in the Salinity and Porosity of Sea-Ice Samples During Shipping and Storage." *Journal of Glaciology* 32 (112): 371–75. https://doi.org/10.3189/S0022143000012065.

Coyne, Kathryn J., Yanfei Wang, and Gretchen Johnson. 2022. "Algicidal Bacteria: A Review of Current Knowledge and Applications to Control Harmful Algal Blooms." *Frontiers in Microbiology* 13 (April). https://doi.org/10.3389/fmicb.2022.871177.

Craig, Lisa, Katrina T. Forest, and Berenike Maier. 2019. "Type IV Pili: Dynamics, Biophysics and Functional Consequences." *Nature Reviews Microbiology* 17 (7): 429–40. https://doi.org/10.1038/s41579-019-0195-4.

Creamean, J. M., J. N. Cross, R. Pickart, L. McRaven, P. Lin, A. Pacini, R. Hanlon, et al. 2019. "Ice Nucleating Particles Carried From Below a Phytoplankton Bloom to the Arctic Atmosphere." *Geophysical Research Letters* 46 (14): 8572–81. https://doi.org/10.1029/2019GL083039.

Czajka, Jeffrey J., Mary H. Abernathy, Veronica T. Benites, Edward E. K. Baidoo, Jody W. Deming, and Yinjie J. Tang. 2018. "Model Metabolic Strategy for Heterotrophic Bacteria in the Cold Ocean Based on *Colwellia Psychrerythraea* 34H." *Proceedings of the National Academy of Sciences* 115 (49): 12507–12. https://doi.org/10.1073/pnas.1807804115.

Davila, Alfonso F., and Christopher P. McKay. 2014. "Chance and Necessity in Biochemistry: Implications for the Search for Extraterrestrial Biomarkers in Earth-like Environments." *Astrobiology* 14 (6): 534–40. https://doi.org/10.1089/ast.2014.1150.

De Maayer, Pieter, Dominique Anderson, Craig Cary, and Don A. Cowan. 2014. "Some like It Cold: Understanding the Survival Strategies of Psychrophiles." *EMBO Reports* 15 (5): 508–17. https://doi.org/10.1002/embr.201338170.

Dalcin Martins, Paula, Anniek de Jong, Wytze K. Lenstra, Niels A. G. M. van Helmond, Caroline P. Slomp, Mike S. M. Jetten, Cornelia U. Welte, and Olivia Rasigraf. 2021. "Enrichment of Novel Verrucomicrobia, Bacteroidetes, and Krumholzibacteria in an Oxygen-Limited Methane- and Iron-Fed Bioreactor Inoculated with Bothnian Sea Sediments." *MicrobiologyOpen* 10 (1): e1175. https://doi.org/10.1002/mbo3.1175.

Danecek, Petr, James K. Bonfield, Jennifer Liddle, John Marshall, Valeriu Ohan, Martin O. Pollard, Andrew Whitwham, et al. 2021. “Twelve Years of SAMtools and BCFtools.” *GigaScience* 10 (2): giab008. https://doi.org/10.1093/gigascience/giab008.

Dedysh, Svetlana N, and Anastasia A Ivanova. 2019. “Planctomycetes in Boreal and Subarctic Wetlands: Diversity Patterns and Potential Ecological Functions.” *FEMS Microbiology Ecology* 95 (2): fiy227. https://doi.org/10.1093/femsec/fiy227.

Deming, Jody W., and Jodi N. Young. 2017. “The Role of Exopolysaccharides in Microbial Adaptation to Cold Habitats.” In *Psychrophiles: From Biodiversity to Biotechnology*, edited by Rosa Margesin, 259–84. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-57057-0_12.

Denby, Katie J., Jeffrey Iwig, Claudine Bisson, Jodie Westwood, Matthew D. Rolfe, Svetlana E. Sedelnikova, Khadine Higgins, et al. 2016. “The Mechanism of a Formaldehyde-Sensing Transcriptional Regulator.” *Scientific Reports* 6 (December):38879. https://doi.org/10.1038/srep38879.

Deng, Yiqin, Chang Chen, Zhe Zhao, Jingjing Zhao, Annick Jacq, Xiaochun Huang, and Yiying Yang. 2016. “The RNA Chaperone Hfq Is Involved in Colony Morphology, Nutrient Utilization and Oxidative and Envelope Stress Response in Vibrio Alginolyticus.” *PLoS ONE* 11 (9). https://doi.org/10.1371/journal.pone.0163689.

Deng, Yunyan, Kui Wang, Zhangxi Hu, Qiang Hu, and Ying Zhong Tang. 2022. “Identification and Implications of a Core Bacterial Microbiome in 19 Clonal Cultures Laboratory-Reared for Months to Years of the Cosmopolitan Dinoflagellate Karlodinium Veneficum.” *Frontiers in Microbiology* 13 (August):967610. https://doi.org/10.3389/fmicb.2022.967610.

Dickson, James L., James W. Head, Joseph S. Levy, and David R. Marchant. 2013. “Don Juan Pond, Antarctica: Near-Surface CaCl 2 -Brine Feeding Earth’s Most Saline Lake and Implications for Mars.” *Scientific Reports* 3 (1): 1166. https://doi.org/10.1038/srep01166.

Doerr, Allison. 2012. “Targeting with PRM.” *Nature Methods* 9 (10): 950–950. https://doi.org/10.1038/nmeth.2193.

Dong, Hong-Po, Yi-Guo Hong, Songhui Lu, and Lu-Yuan Xie. 2014. “Metaproteomics Reveals the Major Microbial Players and Their Biogeochemical Functions in a Productive Coastal System in the Northern South China Sea.” *Environmental Microbiology Reports* 6 (6): 683–95. https://doi.org/10.1111/1758-2229.12188.

Eberhard, Stephan, Giovanni Finazzi, and Francis-André Wollman. 2008. “The Dynamics of Photosynthesis.” *Annual Review of Genetics* 42 (Volume 42, 2008): 463–515. https://doi.org/10.1146/annurev.genet.42.110807.091452.

Eicken, Hajo, and Maya Salganek. 2009. *Field Techniques for Sea-Ice Research*. University of Alaska Press.

Ellis, Kelsey A., Natalie R. Cohen, Carly Moreno, and Adrian Marchetti. 2017. “Cobalamin-Independent Methionine Synthase Distribution and Influence on Vitamin B12 Growth Requirements in Marine Diatoms.” *Protist* 168 (1): 32–47. https://doi.org/10.1016/j.protis.2016.10.007.

Ellison, Courtney K, Gregory B Whitfield, and Yves V Brun. 2022. “Type IV Pili: Dynamic Bacterial Nanomachines.” *FEMS Microbiology Reviews* 46 (2): fuab053. https://doi.org/10.1093/femsre/fuab053.

Eng, Jimmy K., Tahmina A. Jahan, and Michael R. Hoopmann. 2013. “Comet: An Open-Source MS/MS Sequence Database Search Tool.” *Proteomics* 13 (1): 22–24. https://doi.org/10.1002/pmic.201200439.

Eng, Jimmy K., Michael R. Hoopmann, Tahmina A. Jahan, Jarrett D. Egertson, William S. Noble, and Michael J. MacCoss. 2015. “A Deeper Look into Comet – Implementation and Features.” *Journal of the American Society for Mass Spectrometry* 26 (11): 1865–74. https://doi.org/10.1007/s13361-015-1179-x.

Eschenmoser, Albert. 2011. “Etiology of Potentially Primordial Biomolecular Structures: From Vitamin B12 to the Nucleic Acids and an Inquiry into the Chemistry of Life’s Origin: A Retrospective.” *Angewandte Chemie International Edition* 50 (52): 12412–72. https://doi.org/10.1002/anie.201103672.

Everett, Logan J., Charlene Bierl, and Stephen R. Master. 2010. “Unbiased Statistical Analysis for Multi-Stage Proteomic Search Strategies.” *Journal of Proteome Research* 9 (2): 700–707. https://doi.org/10.1021/pr900256v.

Ewert, M., S. D. Carpenter, J. Colangelo-Lillis, and J. W. Deming. 2013. “Bacterial and Extracellular Polysaccharide Content of Brine-Wetted Snow over Arctic Winter First-Year Sea Ice.” *Journal of Geophysical Research: Oceans* 118 (2): 726–35. https://doi.org/10.1002/jgrc.20055.

Ewert, Marcela, and Jody W. Deming. 2011. “Selective Retention in Saline Ice of Extracellular Polysaccharides Produced by the Cold-Adapted Marine Bacterium *Colwellia Psychrerythraea* Strain 34H.” *Annals of Glaciology* 52 (57): 111–17. https://doi.org/10.3189/172756411795931868.

Ewert, Marcela, and Jody W. Deming. 2014. “Bacterial Responses to Fluctuations and Extremes in Temperature and Brine Salinity at the Surface of Arctic Winter Sea Ice.” *FEMS Microbiology Ecology* 89 (2): 476–89. https://doi.org/10.1111/1574-6941.12363.

Falb, Michaela, Kerstin Müller, Lisa Königsmaier, Tanja Oberwinkler, Patrick Horn, Susanne von Gronau, Orland Gonzalez, Friedhelm Pfeiffer, Erich Bornberg-Bauer, and Dieter Oesterhelt. 2008. "Metabolism of Halophilic Archaea." *Extremophiles* 12 (2): 177–96. https://doi.org/10.1007/s00792-008-0138-x.

Falkowski, P., R. J. Scholes, E. Boyle, J. Canadell, D. Canfield, J. Elser, N. Gruber, et al. 2000. "The Global Carbon Cycle: A Test of Our Knowledge of Earth as a System." *Science (New York, N.Y.)* 290 (5490): 291–96. https://doi.org/10.1126/science.290.5490.291.

Ferenci, Thomas. 2005. "Maintaining a Healthy SPANC Balance through Regulatory and Mutational Adaptation." *Molecular Microbiology* 57 (1): 1–8. https://doi.org/10.1111/j.1365-2958.2005.04649.x.

Fermin, Damian, Venkatesha Basrur, Anastasia K. Yocum, and Alexey I. Nesvizhskii. 2011. "Abacus: A Computational Tool for Extracting and Pre-Processing Spectral Count Data for Label-Free Quantitative Proteomic Analysis." *Proteomics* 11 (7): 1340–45. https://doi.org/10.1002/pmic.201000650.

Ferrer-González, Frank Xavier, Brittany Widner, Nicole R Holderman, John Glushka, Arthur S Edison, Elizabeth B Kujawinski, and Mary Ann Moran. 2021. "Resource Partitioning of Phytoplankton Metabolites That Support Bacterial Heterotrophy." *The ISME Journal* 15 (3): 762–73. https://doi.org/10.1038/s41396-020-00811-y.

Fierer, Noah. 2017. "Embracing the Unknown: Disentangling the Complexities of the Soil Microbiome." *Nature Reviews Microbiology* 15 (10): 579–90. https://doi.org/10.1038/nrmicro.2017.87.

Firth, E., S.D. Carpenter, H.L. Sørensen, R.E. Collins, and J.W. Deming. 2016. "Bacterial Use of Choline to Tolerate Salinity Shifts in Sea-Ice Brines." *Elementa: Science of the Anthropocene* 4 (August):000120. https://doi.org/10.12952/journal.elementa.000120.

Flamholz, Avi, Elad Noor, Arren Bar-Even, Wolfram Liebermeister, and Ron Milo. 2013. "Glycolytic Strategy as a Tradeoff between Energy Yield and Protein Cost." *Proceedings of the National Academy of Sciences of the United States of America* 110 (April). https://doi.org/10.1073/pnas.1215283110.

Fortin, Samantha G., Bongkeun Song, Iris C. Anderson, and Kimberly S. Reece. 2022. "Blooms of the Harmful Algae *Margalefidinium Polykrikoides* and *Alexandrium Monilatum* Alter the York River Estuary Microbiome." *Harmful Algae* 114 (May):102216. https://doi.org/10.1016/j.hal.2022.102216.

Francis, Ben, Tim Urich, Annett Mikolasch, Hanno Teeling, and Rudolf Amann. 2021. "North Sea Spring Bloom-Associated Gammaproteobacteria Fill Diverse Heterotrophic Niches."

Environmental Microbiome 16 (1): 15. https://doi.org/10.1186/s40793-021-00385-y.

Francis, T Ben, Daniel Bartosik, Thomas Sura, Andreas Sichert, Jan-Hendrik Hehemann, Stephanie Markert, Thomas Schweder, et al. 2021. “Changing Expression Patterns of TonB-Dependent Transporters Suggest Shifts in Polysaccharide Consumption over the Course of a Spring Phytoplankton Bloom.” *The ISME Journal* 15 (8): 2336–50. https://doi.org/10.1038/s41396-021-00928-8.

Freitas, Sara, Stephen Hatosy, Jed A Fuhrman, Susan M Huse, David B Mark Welch, Mitchell L Sogin, and Adam C Martiny. 2012. “Global Distribution and Diversity of Marine Verrucomicrobia.” *The ISME Journal* 6 (8): 1499–1505. https://doi.org/10.1038/ismej.2012.3.

Fujita, Masaya, Kosuke Mori, Hirofumi Hara, Shojiro Hishiyama, Naofumi Kamimura, and Eiji Masai. 2019. “A TonB-Dependent Receptor Constitutes the Outer Membrane Transport System for a Lignin-Derived Aromatic Compound.” *Communications Biology* 2 (1): 1–10. https://doi.org/10.1038/s42003-019-0676-z.

Fujita, Yasutaro, Hiroshi Matsuoka, and Kazutake Hirooka. 2007. “Regulation of Fatty Acid Metabolism in Bacteria.” *Molecular Microbiology* 66 (4): 829–39. https://doi.org/10.1111/j.1365-2958.2007.05947.x.

Gajardo, Gonzalo, Jesús Morón-López, Karen Vergara, Shoko Ueki, Leonardo Guzmán, Oscar Espinoza-González, Alondra Sandoval, et al. 2023. “The Holobiome of Marine Harmful Algal Blooms (HABs): A Novel Ecosystem-Based Approach for Implementing Predictive Capabilities and Managing Decisions.” *Environmental Science & Policy* 143 (May):44–54. https://doi.org/10.1016/j.envsci.2023.02.012.

Geesey, Gill G., and Richard Y. Morita. 1979. “Capture of Arginine at Low Concentrations by a Marine Psychrophilic Bacterium.” *Applied and Environmental Microbiology* 38 (6): 1092–97.

Georgiou, Christos D. 2018. “Functional Properties of Amino Acid Side Chains as Biomarkers of Extraterrestrial Life.” *Astrobiology* 18 (11): 1479–96. https://doi.org/10.1089/ast.2018.1868.

Gessulat, Siegfried, Tobias Schmidt, Daniel Paul Zolg, Patroklos Samaras, Karsten Schnatbaum, Johannes Zerweck, Tobias Knaute, et al. 2019. “Prosit: Proteome-Wide Prediction of Peptide Tandem Mass Spectra by Deep Learning.” *Nature Methods* 16 (6): 509–18. https://doi.org/10.1038/s41592-019-0426-7.

Gillet, Ludovic C., Alexander Leitner, and Ruedi Aebersold. 2016. “Mass Spectrometry Applied to Bottom-Up Proteomics: Entering the High-Throughput Era for Hypothesis Testing.” *Annual Review of Analytical Chemistry* 9 (Volume 9, 2016): 449–72. https://doi.org/10.1146/annurev-anchem-071015-041535.

Gillette, Michael A., and Steven A. Carr. 2013. “Quantitative Analysis of Peptides and Proteins in

Biomedicine by Targeted Mass Spectrometry." *Nature Methods* 10 (1): 28–34. https://doi.org/10.1038/nmeth.2309.

Giovannoni, Stephen J., and Ulrich Stingl. 2005. "Molecular Diversity and Ecology of Microbial Plankton." *Nature* 437 (7057): 343–48. https://doi.org/10.1038/nature04158.

Glibert, Patricia M., and JoAnn M. Burkholder. 2018. "Causes of Harmful Algal Blooms." In *Harmful Algal Blooms*, 1–38. John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118994672.ch1.

Glukhova, Veronika A., Daniela M. Tomazela, Geoffrey D. Findlay, Raymond J. Monnat, and Michael J. MacCoss. 2013. "Rapid Assessment of RNAi-Mediated Protein Depletion by Selected Reaction Monitoring Mass Spectrometry." *Journal of Proteome Research* 12 (7): 3246–54. https://doi.org/10.1021/pr400067k.

Gold, Vicki AM, Ralf Salzer, Beate Averhoff, and Werner Kühlbrandt. 2015. "Structure of a Type IV Pilus Machinery in the Open and Closed State." Edited by Stephen C Harrison. *eLife* 4 (May):e07380. https://doi.org/10.7554/eLife.07380.

Golden, K. M., H. Eicken, A. L. Heaton, J. Miner, D. J. Pringle, and J. Zhu. 2007. "Thermal Evolution of Permeability and Microstructure in Sea Ice." *Geophysical Research Letters* 34 (16). https://doi.org/10.1029/2007GL030447.

Golden, Susan S. 2020. "Principles of Rhythmicity Emerging from Cyanobacteria." *The European Journal of Neuroscience* 51 (1): 13–18. https://doi.org/10.1111/ejn.14434.

Grettenberger, Christen L. 2021, "Novel Gloeobacterales Spp. from Diverse Environments across the Globe." *mSphere* 6 (4): e00061-21. https://doi.org/10.1128/mSphere.00061-21.

Grossart, Hans-Peter, Florian Levold, Martin Allgaier, Meinhard Simon, and Thorsten Brinkhoff. 2005. "Marine Diatom Species Harbour Distinct Bacterial Communities." *Environmental Microbiology* 7 (6): 860–73. https://doi.org/10.1111/j.1462-2920.2005.00759.x.

Gunde-Cimerman, Nina, Ana Plemenitaš, and Aharon Oren. 2018. "Strategies of Adaptation of Microorganisms of the Three Domains of Life to High Salt Concentrations." *FEMS Microbiology Reviews* 42 (3): 353–75. https://doi.org/10.1093/femsre/fuy009.

Guo, Zhiling, Huan Zhang, and Senjie Lin. 2014. "Light-Promoted Rhodopsin Expression and Starvation Survival in the Marine Dinoflagellate Oxyrrhis Marina." *PLoS ONE* 9 (12). https://doi.org/10.1371/journal.pone.0114941.

Gutteridge, Alex, and Janet M. Thornton. 2005. "Understanding Nature's Catalytic Toolkit." *Trends in Biochemical Sciences* 30 (11): 622–29. https://doi.org/10.1016/j.tibs.2005.09.006.

Hamamoto, Tetsuo, Nobuhisa Takata, Toshiaki Kudo, and Koki Horikoshi. 1994. "Effect of Temperature and Growth Phase on Fatty Acid Composition of the Psychrophilic Vibrio Sp.

Strain No. 5710." *FEMS Microbiology Letters* 119 (1–2): 77–81. https://doi.org/10.1111/j.1574-6968.1994.tb06870.x.

Hand, K.P., C.F. Chyba, J.C. Priscu, R.W. Carlson, and K.H. Nealson. 2009. "Astrobiology and the Potential for Life on Europa." *Europa*, 589–629.

Hand, K.P, Murray, A.E., Garvin, J.B., Brinckerhoff, W.B., Christner, B.C, Edgett, K.S, *et al.* 2017. "Report of the Europa Lander Science Definition Team."

Harvilla, Paul B., Holly N. Wolcott, and John S. Magyar. 2014. "The Structure of Ferricytochrome C552 from the Psychrophilic Marine Bacterium Colwellia Psychrerythraea 34H." *Metallomics : Integrated Biometal Science* 6 (6): 1126–30. https://doi.org/10.1039/c4mt00045e.

Hattenrath-Lehmann, Theresa K., and Christopher J. Gobler. 2017. "Identification of Unique Microbiomes Associated with Harmful Algal Blooms Caused by Alexandrium Fundyense and Dinophysis Acuminata." *Harmful Algae* 68 (September):17–30. https://doi.org/10.1016/j.hal.2017.07.003.

Hedlund, B. P., J. J. Gosink, and J. T. Staley. 1997. "Verrucomicrobia Div. Nov., a New Division of the Bacteria Containing Three New Species of Prosthecobacter." *Antonie Van Leeuwenhoek* 72 (1): 29–38. https://doi.org/10.1023/a:1000348616863.

Hendrix, Amanda R., Terry A. Hurford, Laura M. Barge, Michael T. Bland, Jeff S. Bowman, William Brinckerhoff, Bonnie J. Buratti, et al. 2019. "The NASA Roadmap to Ocean Worlds." *Astrobiology* 19 (1): 1–27. https://doi.org/10.1089/ast.2018.1955.

Henley, Sian F., Marie Porter, Laura Hobbs, Judith Braun, Robin Guillaume-Castel, Emily J. Venables, Estelle Dumont, and Finlo Cottier. 2020. "Nitrate Supply and Uptake in the Atlantic Arctic Sea Ice Zone: Seasonal Cycle, Mechanisms and Drivers." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 378 (2181): 20190361. https://doi.org/10.1098/rsta.2019.0361.

Howarth, Robert W., and Roxanne Marino. 2006. "Nitrogen as the Limiting Nutrient for Eutrophication in Coastal Marine Ecosystems: Evolving Views over Three Decades." *Limnology and Oceanography* 51 (1part2): 364–76. https://doi.org/10.4319/lo.2006.51.1_part_2.0364.

Hu, Alex, William S. Noble, and Alejandro Wolf-Yadlin. 2016. "Technical Advances in Proteomics: New Developments in Data-Independent Acquisition." *F1000Research* 5 (March):F1000 Faculty Rev-419. https://doi.org/10.12688/f1000research.7042.1.

Huber, Michaela, Kathrin S. Fröhlich, Jessica Radmer, and Kai Papenfort. 2020. "Switching Fatty Acid Metabolism by an RNA-Controlled Feed Forward Loop." *Proceedings of the National*

Academy of Sciences 117 (14): 8044–54. https://doi.org/10.1073/pnas.1920753117.

Huerta-Cepas, Jaime, Damian Szklarczyk, Davide Heller, Ana Hernández-Plaza, Sofia K. Forslund, Helen Cook, Daniel R. Mende, et al. 2019. “eggNOG 5.0: A Hierarchical, Functionally and Phylogenetically Annotated Orthology Resource Based on 5090 Organisms and 2502 Viruses.” *Nucleic Acids Research* 47 (D1): D309–14. https://doi.org/10.1093/nar/gky1085.

Huisman, Jef, Geoffrey A. Codd, Hans W. Paerl, Bas W. Ibelings, Jolanda M. H. Verspagen, and Petra M. Visser. 2018. “Cyanobacterial Blooms.” *Nature Reviews Microbiology* 16 (8): 471–83. https://doi.org/10.1038/s41579-018-0040-1.

Huston, A. L., B. B. Krieger-Brockett, and J. W. Deming. 2000. “Remarkably Low Temperature Optima for Extracellular Enzyme Activity from Arctic Bacteria and Sea Ice.” *Environmental Microbiology* 2 (4): 383–88. https://doi.org/10.1046/j.1462-2920.2000.00118.x.

Huston, Adrienne Louisa. 2003. “Bacterial Adaptation to the Cold: In Situ Activities of Extracellular Enzymes in the North Water Polynya and Characterization of a Cold-Active Aminopeptidase from Colwellia Psychrerythraea Strain 34H.” https://doi.org/10.1016/b978-0-12-382219-2.00095-8.

Hyatt, Doug, Gwo-Liang Chen, Philip F. LoCascio, Miriam L. Land, Frank W. Larimer, and Loren J. Hauser. 2010. “Prodigal: Prokaryotic Gene Recognition and Translation Initiation Site Identification.” *BMC Bioinformatics* 11 (1): 119. https://doi.org/10.1186/1471-2105-11-119.

Ito, Koreaki, and Yoshinori Akiyama. 2005. “Cellular Functions, Mechanism of Action, and Regulation of FtsH Protease.” *Annual Review of Microbiology* 59:211–31. https://doi.org/10.1146/annurev.micro.59.030804.121316.

Jackson, R. H., A. Cornish-Bowden, and J. A. Cole. 1981. “Prosthetic Groups of the NADH-Dependent Nitrite Reductase from Escherichia Coli K12.” *The Biochemical Journal* 193 (3): 861–67. https://doi.org/10.1042/bj1930861.

Johnson, Carl Hirschie, Chi Zhao, Yao Xu, and Tetsuya Mori. 2017. “Timing the Day: What Makes Bacterial Clocks Tick?” *Nature Reviews Microbiology* 15 (4): 232–42. https://doi.org/10.1038/nrmicro.2016.196.

Johnson, Richard S., Brian C. Searle, Brook L. Nunn, Jason M. Gilmore, Molly Phillips, Chris T. Amemiya, Michelle Heck, and Michael J. MacCoss. 2020. “Assessing Protein Sequence Database Suitability Using De Novo Sequencing *.” *Molecular & Cellular Proteomics* 19 (1): 198–208. https://doi.org/10.1074/mcp.TIR119.001752.

Junge, Karen, Karen Cameron, and Brook Nunn. 2019. “Chapter 12 - Diversity of Psychrophilic Bacteria in Sea and Glacier Ice Environments—Insights Through Genomics, Metagenomics, and Proteomics Approaches.” In *Microbial Diversity in the Genomic Era*, edited by Surajit Das and Hirak Ranjan Dash, 197–216. Academic Press. https://doi.org/10.1016/B978-0-12-814849-5.00012-5.

Junge, Karen, Hajo Eicken, and Jody Deming. 2004. “Bacterial Activity at -2 to -20 C in Arctic Wintertime Sea Ice.” *Applied and Environmental Microbiology - AEM* 70 (January):550–57. https://doi.org/10.1128/AEM.70.1.550-557.2004.

Junge, Karen, Hajo Eicken, and Jody W. Deming. 2003. “Motility of Colwellia Psychrerythraea Strain 34H at Subzero Temperatures.” *Applied and Environmental Microbiology* 69 (7): 4282–84. https://doi.org/10.1128/AEM.69.7.4282-4284.2003.

Junge, Karen, Hajo Eicken, Brian D. Swanson, and Jody W. Deming. 2006. “Bacterial Incorporation of Leucine into Protein down to -20 Degrees C with Evidence for Potential Activity in Sub-Eutectic Saline Ice Formations.” *Cryobiology* 52 (3): 417–29. https://doi.org/10.1016/j.cryobiol.2006.03.002.

Käll, Lukas, Jesse D. Canterbury, Jason Weston, William Stafford Noble, and Michael J. MacCoss. 2007. “Semi-Supervised Learning for Peptide Identification from Shotgun Proteomics Datasets.” *Nature Methods* 4 (11): 923–25. https://doi.org/10.1038/nmeth1113.

Kallmeyer, Jens, Robert Pockalny, Rishi Ram Adhikari, David C. Smith, and Steven D’Hondt. 2012. “Global Distribution of Microbial Abundance and Biomass in Subseafloor Sediment.” *Proceedings of the National Academy of Sciences* 109 (40): 16213–16. https://doi.org/10.1073/pnas.1203849109.

Kanehisa, Minoru, Yoko Sato, and Kanae Morishima. 2016. “BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences.” *Journal of Molecular Biology* 428 (4): 726–31. https://doi.org/10.1016/j.jmb.2015.11.006.

Karr, Elizabeth A., W. Matthew Sattley, Melissa R. Rice, Deborah O. Jung, Michael T. Madigan, and Laurie A. Achenbach. 2005. “Diversity and Distribution of Sulfate-Reducing Bacteria in Permanently Frozen Lake Fryxell, McMurdo Dry Valleys, Antarctica.” *Applied and Environmental Microbiology* 71 (10): 6353–59. https://doi.org/10.1128/AEM.71.10.6353-6359.2005.

Katsaros, Kristina. 1973. “Supercooling at the Surface of an Arctic Lead.” *Journal of Physical Oceanography - J PHYS OCEANOGR* 3 (October):482–86. https://doi.org/10.1175/1520-0485(1973)003<0482:SATSOA>2.0.CO;2.

Khurana, K. K., M. G. Kivelson, D. J. Stevenson, G. Schubert, C. T. Russell, R. J. Walker, and C. Polanskey. 1998. “Induced Magnetic Fields as Evidence for Subsurface Oceans in Europa and Callisto.” *Nature* 395 (6704): 777–80. https://doi.org/10.1038/27394.

Kleiner, Manuel, Erin Thorson, Christine E. Sharp, Xiaoli Dong, Dan Liu, Carmen Li, and Marc Strous. 2017. "Assessing Species Biomass Contributions in Microbial Communities via Metaproteomics." *Nature Communications* 8 (1): 1558. https://doi.org/10.1038/s41467-017-01544-x.

Knoblauch, Christian, Bo Barker Jørgensen, and Jens Harder. 1999. "Community Size and Metabolic Rates of Psychrophilic Sulfate-Reducing Bacteria in Arctic Marine Sediments." *Applied and Environmental Microbiology* 65 (9): 4230–33. https://doi.org/10.1128/AEM.65.9.4230-4233.1999.

Knoblauch, Christian, Kerstin Sahm, and Bo B. Jørgensen. 1999. "Psychrophilic Sulfate-Reducing Bacteria Isolated from Permanently Cold Arctic Marine Sediments: Description of Desulfofrigus Oceanense Gen. Nov., Sp. Nov., Desulfofrigus Fragile Sp. Nov., Desulfofaba Gelida Gen. Nov., Sp. Nov., Desulfotalea Psychrophila Gen. Nov., Sp. Nov. and Desulfotalea Arctica Sp. Nov." *International Journal of Systematic and Evolutionary Microbiology,* 49 (4): 1631–43. https://doi.org/10.1099/00207713-49-4-1631.

Kolde, Raivo. 2019. "Pheatmap: Pretty Heatmaps." https://cran.r-project.org/web/packages/pheatmap/index.html.

Krembs, C, H Eicken, K Junge, and J. W Deming. 2002. "High Concentrations of Exopolymeric Substances in Arctic Winter Sea Ice: Implications for the Polar Ocean Carbon Cycle and Cryoprotection of Diatoms." *Deep Sea Research Part I: Oceanographic Research Papers* 49 (12): 2163–81. https://doi.org/10.1016/S0967-0637(02)00122-X.

Krembs, Christopher, and Jody W. Deming. 2008. "The Role of Exopolymers in Microbial Adaptation to Sea Ice." In *Psychrophiles: From Biodiversity to Biotechnology*, edited by Rosa Margesin, Franz Schinner, Jean-Claude Marx, and Charles Gerday, 247–64. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-540-74335-4_15.

Krembs, Christopher, Hajo Eicken, and Jody W. Deming. 2011. "Exopolymer Alteration of Physical Properties of Sea Ice and Implications for Ice Habitability and Biogeochemistry in a Warmer Arctic." *Proceedings of the National Academy of Sciences* 108 (9): 3653–58. https://doi.org/10.1073/pnas.1100701108.

Kuan, Yi-Jehng, Hui-Chun Huang, Steven B. Charnley, Andrew Markwick, Oliver Botta, Pascale Ehrenfreund, Zbigniew Kisiel, and Harold M. Butner. 2003. "Interstellar Organic Molecules: From Clouds to Solar Nebula" 14. http://adsabs.harvard.edu/abs/2003IAUJD..14E..44K.

Kuhlisch, Constanze, Adva Shemi, Noa Barak-Gavish, Daniella Schatz, and Assaf Vardi. 2024. "Algal Blooms in the Ocean: Hot Spots for Chemically Mediated Microbial Interactions." *Nature Reviews Microbiology* 22 (3): 138–54. https://doi.org/10.1038/s41579-023-00975-2.

Kyte, J., and R. F. Doolittle. 1982. “A Simple Method for Displaying the Hydropathic Character of a Protein.” *Journal of Molecular Biology* 157 (1): 105–32. https://doi.org/10.1016/0022-2836(82)90515-0.

Lauro, Sebastian Emanuel, Elena Pettinelli, Graziella Caprarelli, Luca Guallini, Angelo Pio Rossi, Elisabetta Mattei, Barbara Cosciotti, et al. 2020. “Multiple Subglacial Water Bodies below the South Pole of Mars Unveiled by New MARSIS Data.” *Nature Astronomy*, September, 1–8. https://doi.org/10.1038/s41550-020-1200-6.

Leach, Whitney B., Tyler J. Carrier, and Adam M. Reitzel. 2019. “Diel Patterning in the Bacterial Community Associated with the Sea Anemone Nematostella Vectensis.” *Ecology and Evolution* 9 (17): 9935–47. https://doi.org/10.1002/ece3.5534.

Lee, Donghyun, Mingyu Kim, Beomhui Lee, Sangwon Chae, Sungjun Kwon, and Sungwon Kang. 2022. “Integrated Explainable Deep Learning Prediction of Harmful Algal Blooms.” *Technological Forecasting and Social Change* 185 (December):122046. https://doi.org/10.1016/j.techfore.2022.122046.

Lee, Elliot M., Sujatha Srinivasan, Samuel O. Purvine, Tina L. Fiedler, Owen P. Leiser, Sean C. Proll, Samuel S. Minot, Brooke L. Deatherage Kaiser, and David N. Fredricks. 2023. “Optimizing Metaproteomics Database Construction: Lessons from a Study of the Vaginal Microbiome.” *mSystems* 8 (4): e00678-22. https://doi.org/10.1128/msystems.00678-22.

Lee, Seung Jae, Eun-Mi Jeong, Ah Young Ki, Kyung-Seo Oh, Joseph Kwon, Jae-Hyuk Jeong, and Nam-Jin Chung. 2016. “Oxidative Defense Metabolites Induced by Salinity Stress in Roots of Salicornia Herbacea.” *Journal of Plant Physiology* 206 (November):133–42. https://doi.org/10.1016/j.jplph.2016.08.015.

Lewis, Kim, Slava Epstein, Anthony D’Onofrio, and Losee L. Ling. 2010. “Uncultured Microorganisms as a Source of Secondary Metabolites.” *The Journal of Antibiotics* 63 (8): 468–76. https://doi.org/10.1038/ja.2010.87.

Li, Dinghua, Chi-Man Liu, Ruibang Luo, Kunihiko Sadakane, and Tak-Wah Lam. 2015. “MEGAHIT: An Ultra-Fast Single-Node Solution for Large and Complex Metagenomics Assembly via Succinct de Bruijn Graph.” *Bioinformatics* 31 (10): 1674–76. https://doi.org/10.1093/bioinformatics/btv033.

Li, Xian-Zhi, and Hiroshi Nikaido. 2009. “Efflux-Mediated Drug Resistance in Bacteria: An Update.” *Drugs* 69 (12): 1555–1623. https://doi.org/10.2165/11317030-000000000-00000.

Lieberman, Tami D. 2022. “Detecting Bacterial Adaptation within Individual Microbiomes.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 377 (1861): 20210243. https://doi.org/10.1098/rstb.2021.0243.

Lin, Jing, Yingying Meng, Yanjing Shi, and Xuezheng Lin. 2019. “Complete Genome Sequences of Colwellia Sp. Arc7-635, a Denitrifying Bacterium Isolated from Arctic Seawater.” *Current Microbiology* 76 (9): 1061–65. https://doi.org/10.1007/s00284-019-01714-8.

Lin, Junfang, Peter I. Miller, Bror F. Jönsson, and Michael Bedington. 2021. “Early Warning of Harmful Algal Bloom Risk Using Satellite Ocean Color and Lagrangian Particle Trajectories.” *Frontiers in Marine Science* 8 (November). https://doi.org/10.3389/fmars.2021.736262.

Lindh, Markus V., Johanna Sjöstedt, Anders F. Andersson, Federico Baltar, Luisa W. Hugerth, Daniel Lundin, Saraladevi Muthusamy, Catherine Legrand, and Jarone Pinhassi. 2015. “Disentangling Seasonal Bacterioplankton Population Dynamics by High-Frequency Sampling.” *Environmental Microbiology* 17 (7): 2459–76. https://doi.org/10.1111/1462-2920.12720.

Liu, Guiming, Shuangfeng Cai, Jing Hou, Dahe Zhao, Jing Han, Jian Zhou, and Hua Xiang. 2016. “Enoyl-CoA Hydratase Mediates Polyhydroxyalkanoate Mobilization in Haloferax Mediterranei.” *Scientific Reports* 6 (1): 24015. https://doi.org/10.1038/srep24015.

Liu, Shuting, and Zhanfei Liu. 2020. “Distinct Capabilities of Different Gammaproteobacterial Strains on Utilizing Small Peptides in Seawater.” *Scientific Reports* 10 (January):464. https://doi.org/10.1038/s41598-019-57189-x.

MacLean, Brendan, Daniela M. Tomazela, Nicholas Shulman, Matthew Chambers, Gregory L. Finney, Barbara Frewen, Randall Kern, David L. Tabb, Daniel C. Liebler, and Michael J. MacCoss. 2010. “Skyline: An Open Source Document Editor for Creating and Analyzing Targeted Proteomics Experiments.” *Bioinformatics (Oxford, England)* 26 (7): 966–68. https://doi.org/10.1093/bioinformatics/btq054.

Maier, Tobias, Alexander Schmidt, Marc Güell, Sebastian Kühner, Anne-Claude Gavin, Ruedi Aebersold, and Luis Serrano. 2011. “Quantification of mRNA and Protein and Integration with Protein Turnover in a Bacterium.” *Molecular Systems Biology* 7 (July):511. https://doi.org/10.1038/msb.2011.38.

Marion, Giles M., and Jeffrey S. Kargel. 2008. *Cold Aqueous Planetary Geochemistry with FREZCHEM: From Modeling to the Search for Life at the Limits*. Advances in Astrobiology and Biogeophysics. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-75679-8.

Marion, G.M, R.E Farren, and A.J Komrowski. 1999. “Alternative Pathways for Seawater Freezing.” *Cold Regions Science and Technology* 29 (3): 259–66. https://doi.org/10.1016/S0165-232X(99)00033-6.

Marrone, Babetta L., Shounak Banerjee, Anjana Talapatra, C. Raul Gonzalez-Esquer, and

Ghanshyam Pilania. 2024. "Toward a Predictive Understanding of Cyanobacterial Harmful Algal Blooms through AI Integration of Physical, Chemical, and Biological Data." *ACS ES&T Water* 4 (3): 844–58. https://doi.org/10.1021/acsestwater.3c00369.

Martin, Kyle P., Shannon M. MacKenzie, Jason W. Barnes, and F. Marty Ytreberg. 2019. "Protein Stability in Titan's Subsurface Water Ocean." *Astrobiology* 20 (2): 190–98. https://doi.org/10.1089/ast.2018.1972.

Marx, Joseph G., Shelly D. Carpenter, and Jody W. Deming. 2009. "Production of Cryoprotectant Extracellular Polysaccharide Substances (EPS) by the Marine Psychrophilic Bacterium Colwellia Psychrerythraea Strain 34H under Extreme Conditions." *Canadian Journal of Microbiology* 55 (1): 63–72. https://doi.org/10.1139/W08-130.

Mayali, Xavier. 2018. "Editorial: Metabolic Interactions Between Bacteria and Phytoplankton." *Frontiers in Microbiology* 9 (April):727. https://doi.org/10.3389/fmicb.2018.00727.

McGeoch, Malcolm W., Sergei Dikler, and Julie E. M. McGeoch. 2020. "Hemolithin: A Meteoritic Protein Containing Iron and Lithium." *arXiv:2002.11688 [Astro-Ph, Physics:Physics]*, February. http://arxiv.org/abs/2002.11688.

McKay, David S., Everett K. Gibson, Kathie L. Thomas-Keprta, Hojatollah Vali, Christopher S. Romanek, Simon J. Clemett, Xavier D. F. Chillier, Claude R. Maechling, and Richard N. Zare. 1996. "Search for Past Life on Mars: Possible Relic Biogenic Activity in Martian Meteorite ALH84001." *Science* 273 (5277): 924–30. https://doi.org/10.1126/science.273.5277.924.

Meganathan, R. 2001. "Ubiquinone Biosynthesis in Microorganisms." *FEMS Microbiology Letters* 203 (2): 131–39. https://doi.org/10.1111/j.1574-6968.2001.tb10831.x.

Menzel, David W., and Jane P. Spaeth. 1962. "Occurrence of Vitamin B12 in the Sargasso Sea1." *Limnology and Oceanography* 7 (2): 151–54. https://doi.org/10.4319/lo.1962.7.2.0151.

Methé, Barbara A., Karen E. Nelson, Jody W. Deming, Bahram Momen, Eugene Melamud, Xijun Zhang, John Moult, et al. 2005. "The Psychrophilic Lifestyle as Revealed by the Genome Sequence of Colwellia Psychrerythraea 34H through Genomic and Proteomic Analyses." *Proceedings of the National Academy of Sciences* 102 (31): 10913–18. https://doi.org/10.1073/pnas.0504766102.

Mikan, Molly P., H. Rodger Harvey, Emma Timmins-Schiffman, Michael Riffle, Damon H. May, Ian Salter, William S. Noble, and Brook L. Nunn. 2020. "Metaproteomics Reveal That Rapid Perturbations in Organic Matter Prioritize Functional Restructuring over Taxonomy in Western Arctic Ocean Microbiomes." *The ISME Journal* 14 (1): 39–52. https://doi.org/10.1038/s41396-019-0503-z.

Miller, Rachel M., and Lloyd M. Smith. 2023. "Overview and Considerations in Bottom-up

Proteomics." *Analyst* 148 (3): 475–86. https://doi.org/10.1039/D2AN01246D.

Moran, Mary Ann. 2015. "The Global Ocean Microbiome." *Science* 350 (6266): aac8455. https://doi.org/10.1126/science.aac8455.

Mordukhova, Elena A., and Jae-Gu Pan. 2013. "Evolved Cobalamin-Independent Methionine Synthase (MetE) Improves the Acetate and Thermal Tolerance of Escherichia Coli." *Applied and Environmental Microbiology* 79 (24): 7905–15. https://doi.org/10.1128/AEM.01952-13.

Morris, Robert M, Brook L Nunn, Christian Frazar, David R Goodlett, Ying S Ting, and Gabrielle Rocap. 2010. "Comparative Metaproteomics Reveals Ocean-Scale Shifts in Microbial Nutrient Utilization and Energy Transduction." *The ISME Journal* 4 (5): 673–85. https://doi.org/10.1038/ismej.2010.4.

Moyer, Craig L., R. Eric Collins, and Richard Y. Morita. 2017. "Psychrophiles and Psychrotrophs." In *Reference Module in Life Sciences*, B9780128096338022822. Elsevier. https://doi.org/10.1016/B978-0-12-809633-8.02282-2.

Mudge, Miranda C., Brook L. Nunn, Erin Firth, Marcela Ewert, Kianna Hales, William E. Fondrie, William S. Noble, Jonathan Toner, Bonnie Light, and Karen A. Junge. 2021. "Subzero, Saline Incubations of Colwellia Psychrerythraea Reveal Strategies and Biomarkers for Sustained Life in Extreme Icy Environments." *Environmental Microbiology* 23 (7): 3840–66. https://doi.org/10.1111/1462-2920.15485.

Muratore, Daniel, Angela K. Boysen, Matthew J. Harke, Kevin W. Becker, John R. Casey, Sacha N. Coesel, Daniel R. Mende, et al. 2022. "Complex Marine Microbial Communities Partition Metabolism of Scarce Resources over the Diel Cycle." *Nature Ecology & Evolution* 6 (2): 218–29. https://doi.org/10.1038/s41559-021-01606-w.

Murray, Alison E., Fabien Kenig, Christian H. Fritsen, Christopher P. McKay, Kaelin M. Cawley, Ross Edwards, Emanuele Kuhn, et al. 2012. "Microbial Life at −13 °C in the Brine of an Ice-Sealed Antarctic Lake." *Proceedings of the National Academy of Sciences* 109 (50): 20626–31. https://doi.org/10.1073/pnas.1208607109.

Nagore, Daniel, Begoña Sanz, Javier Soria, Marta Llarena, María J. Llama, Juan J. Calvete, and Juan L. Serra. 2006. "The Nitrate/Nitrite ABC Transporter of Phormidium Laminosum: Phosphorylation State of NrtA Is Not Involved in Its Substrate Binding Activity." *Biochimica et Biophysica Acta (BBA) - General Subjects* 1760 (2): 172–81. https://doi.org/10.1016/j.bbagen.2005.12.011.

Nesvizhskii, Alexey I., Andrew Keller, Eugene Kolker, and Ruedi Aebersold. 2003. "A Statistical Model for Identifying Proteins by Tandem Mass Spectrometry." *Analytical Chemistry* 75 (17): 4646–58. https://doi.org/10.1021/ac0341261.

Neveu, Marc, Lindsay E. Hays, Mary A. Voytek, Michael H. New, and Mitchell D. Schulte. 2018. "The Ladder of Life Detection." *Astrobiology* 18 (11): 1375–1402. https://doi.org/10.1089/ast.2017.1773.

Ni, Lei, Shuai Yang, Rongrong Zhang, Zhenyu Jin, Hao Chen, Jacinta C. Conrad, and Fan Jin. 2016. "Bacteria Differently Deploy Type-IV Pili on Surfaces to Adapt to Nutrient Availability." *Npj Biofilms and Microbiomes* 2 (1): 1–10. https://doi.org/10.1038/npjbiofilms.2015.29.

Ni, Ziqin, Ricardo Arevalo, Anais Bardyn, Lori Willhite, Soumya Ray, Adrian Southard, Ryan Danell, *et al.* 2023. "Detection of Short Peptides as Putative Biosignatures of Psychrophiles via Laser Desorption Mass Spectrometry." *Astrobiology* 23 (6): 657–69. https://doi.org/10.1089/ast.2022.0138.

Nichols, Robert G., and Emily R. Davenport. 2021. "The Relationship between the Gut Microbiome and Host Gene Expression: A Review." *Human Genetics* 140 (5): 747–60. https://doi.org/10.1007/s00439-020-02237-0.

Nilsson, G., J. G. Belasco, S. N. Cohen, and A. von Gabain. 1984. "Growth-Rate Dependent Regulation of mRNA Stability in Escherichia Coli." *Nature* 312 (5989): 75–77. https://doi.org/10.1038/312075a0.

Noor, Elad, Arren Bar-Even, Avi Flamholz, Ed Reznik, Wolfram Liebermeister, and Ron Milo. 2014. "Pathway Thermodynamics Highlights Kinetic Obstacles in Central Metabolism." *PLOS Computational Biology* 10 (2): e1003483. https://doi.org/10.1371/journal.pcbi.1003483.

Nunn, Brook L. 2024. "Investigating the Interactions of a Phytoplankton Community and Its Microbiome on a 4-Hour Timescale to Reveal Emerging and Predictive Properties across an Algal Bloom and Bust Cycle." http://dx.doi.org/10.46936/10.25585/60008633.

Nunn, Brook L., Krystal V. Slattery, Karen A. Cameron, Emma Timmins-Schiffman, and Karen Junge. 2015. "Proteomics of Colwellia Psychrerythraea at Subzero Temperatures – a Life with Limited Movement, Flexible Membranes and Vital DNA Repair." *Environmental Microbiology* 17 (7): 2319–35. https://doi.org/10.1111/1462-2920.12691.

Nutrition, Center for Food Safety and Applied. 2020. "BAM Appendix 2: Most Probable Number from Serial Dilutions." *FDA*, October. https://www.fda.gov/food/laboratory-methods-food/bam-appendix-2-most-probable-number-serial-dilutions.

Ogunrinola, Grace A., John O. Oyewale, Oyewumi O. Oshamika, and Grace I. Olasehinde. 2020. "The Human Microbiome and Its Impacts on Health." *International Journal of Microbiology* 2020 (1): 8045646. https://doi.org/10.1155/2020/8045646.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H. and Wagner, H., 2013. "Community ecology package." *R* *package version*, *2*(0).

Oksanen, J., Guillaume Blanchet, F., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., *et al*. 2019. “vegan: Community Ecology Package.” R package version 2.5-6. URL https://CRAN.R-project.org/package=vegan.

O’Malley, Maureen A. 2008. “‘*Everything Is Everywhere*: But *the Environment Selects*’: Ubiquitous Distribution and Ecological Determinism in Microbial Biogeography.” *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 39 (3): 314–25. https://doi.org/10.1016/j.shpsc.2008.06.005.

Orellana, Luis, T. Francis, Marcela Ferraro, Jan-Hendrik Hehemann, Bernhard Fuchs, and Rudolf Amann. 2021. “Verrucomicrobiota Are Specialist Consumers of Sulfated Methyl Pentoses during Diatom Blooms.” *The ISME Journal* 16 (September). https://doi.org/10.1038/s41396-021-01105-7.

Orellana, Luis H., T. Ben Francis, Marcela Ferraro, Jan-Hendrik Hehemann, Bernhard M. Fuchs, and Rudolf I. Amann. 2022. “Verrucomicrobiota Are Specialist Consumers of Sulfated Methyl Pentoses during Diatom Blooms.” *The ISME Journal* 16 (3): 630–41. https://doi.org/10.1038/s41396-021-01105-7.

Oren, Aharon. 2006. *Halophilic Microorganisms and Their Environments*. Springer Science & Business Media.

Ou, Fang-Shu, Stefan Michiels, Yu Shyr, Alex A. Adjei, and Ann L. Oberg. 2021. “Biomarker Discovery and Validation: Statistical Considerations.” *Journal of Thoracic Oncology : Official Publication of the International Association for the Study of Lung Cancer* 16 (4): 537–45. https://doi.org/10.1016/j.jtho.2021.01.1616.

Pahel, G., D. M. Rothstein, and B. Magasanik. 1982. “Complex glnA-glnL-glnG Operon of Escherichia Coli.” *Journal of Bacteriology* 150 (1): 202–13.

Pandey, Shashank, Gaurav Malviya, and Magdalena Chottova Dvorakova. 2021. “Role of Peptides in Diagnostics.” *International Journal of Molecular Sciences* 22 (16): 8828. https://doi.org/10.3390/ijms22168828.

Parro, Víctor, Yolanda Blanco, Fernando Puente-Sánchez, Luis A. Rivas, Mercedes Moreno-Paz, Alex Echeverría, Guillermo Chong-Díaz, Cecilia Demergasso, and Nathalie A. Cabrol. 2016. “Biomarkers and Metabolic Patterns in the Sediments of Evolving Glacial Lakes as a Proxy for Planetary Lake Exploration.” *Astrobiology* 18 (5): 586–606. https://doi.org/10.1089/ast.2015.1342.

Patin, Nastassia V., Emily Brown, Gabriella Chebli, Claire Garfield, Julia Kubanek, and Frank J. Stewart. 2020. “Microbial and Chemical Dynamics of a Toxic Dinoflagellate Bloom.” *PeerJ* 8

(July):e9493. https://doi.org/10.7717/peerj.9493.

Pema, P. J., H. A. Horak, and R. H. Wyatt. 1998. “Myelopathy Caused by Nitrous Oxide Toxicity.” *American Journal of Neuroradiology* 19 (5): 894–96.

Pawelek, Peter D., Nathalie Croteau, Christopher Ng-Thow-Hing, Cezar M. Khursigara, Natalia Moiseeva, Marc Allaire, and James W. Coulton. 2006. “Structure of TonB in Complex with FhuA, E. Coli Outer Membrane Receptor.” *Science* 312 (5778): 1399–1402. https://doi.org/10.1126/science.1128057.
Pema, P. J., H. A. Horak, and R. H. Wyatt. 1998. “Myelopathy Caused by Nitrous Oxide Toxicity.” *American Journal of Neuroradiology* 19 (5): 894–96.
Pino, Lindsay K., Brian C. Searle, Han-Yin Yang, Andrew N. Hoofnagle, William S. Noble, and Michael J. MacCoss. 2020. “Matrix-Matched Calibration Curves for Assessing Analytical Figures of Merit in Quantitative Proteomics.” *Journal of Proteome Research* 19 (3): 1147–53. https://doi.org/10.1021/acs.jproteome.9b00666.

Pino, Lindsay K., Seth C. Just, Michael J. MacCoss, and Brian C. Searle. 2020. “Acquiring and Analyzing Data Independent Acquisition Proteomics Experiments without Spectrum Libraries.” *Molecular & Cellular Proteomics* 19 (7): 1088–1103. https://doi.org/10.1074/mcp.P119.001913.

Pinto, Daniela, Mário A. Santos, and Lélia Chambel. 2015. “Thirty Years of Viable but Nonculturable State Research: Unsolved Molecular Mechanisms.” *Critical Reviews in Microbiology* 41 (1): 61–76. https://doi.org/10.3109/1040841X.2013.794127.

Pizzarello, S, G W Cooper, and G J Flynn. 2006 “The Nature and Distribution of the Organic Material in Carbonaceous Chondrites and Interplanetary Dust Particles,” 27.

Pizzarello, Sandra. 2006. “The Chemistry of Life’s Origin: A Carbonaceous Meteorite Perspective.” *Accounts of Chemical Research* 39 (4): 231–37. https://doi.org/10.1021/ar050049f.

Pizzarello, Sandra, and William Holmes. 2009. “Nitrogen-Containing Compounds in Two CR2 Meteorites: 15N Composition, Molecular Distribution and Precursor Molecules.” *Geochimica et Cosmochimica Acta* 73 (7): 2150–62. https://doi.org/10.1016/j.gca.2009.01.022.

Plubell, Deanna L., Lukas Käll, Bobbie-Jo Webb-Robertson, Lisa Bramer, Ashley Ives, Neil L. Kelleher, Lloyd M. Smith, Thomas J. Montine, Christine C. Wu, and Michael J. MacCoss. 2021. “Can We Put Humpty Dumpty Back Together Again? What Does Protein Quantification Mean in Bottom-up Proteomics?” *bioRxiv*, January, 2021.01.25.428175. https://doi.org/10.1101/2021.01.25.428175.

Pollet, Rebecca M., Lauryn M. Martin, and Nicole M. Koropatkin. 2021. “TonB-Dependent

Transporters in the Bacteroidetes: Unique Domain Structures and Potential Functions.” *Molecular Microbiology* 115 (3): 490–501. https://doi.org/10.1111/mmi.14683.

Potiszil, Christian, Tsutomu Ota, Masahiro Yamanaka, Chie Sakaguchi, Katsura Kobayashi, Ryoji Tanaka, Tak Kunihiro, et al. 2023. “Insights into the Formation and Evolution of Extraterrestrial Amino Acids from the Asteroid Ryugu.” *Nature Communications* 14 (1): 1482. https://doi.org/10.1038/s41467-023-37107-6.

Price, P. Buford. 2000. “A Habitat for Psychrophiles in Deep Antarctic Ice.” *Proceedings of the National Academy of Sciences* 97 (3): 1247–51. https://doi.org/10.1073/pnas.97.3.1247.

Price, P. Buford, and Todd Sowers. 2004. “Temperature Dependence of Metabolic Rates for Microbial Growth, Maintenance, and Survival.” *Proceedings of the National Academy of Sciences* 101 (13): 4631–36. https://doi.org/10.1073/pnas.0400522101.

Rabus, R., A. Ruepp, T. Frickey, T. Rattei, B. Fartmann, M. Stark, M. Bauer, et al. 2004. “The Genome of Desulfotalea Psychrophila, a Sulfate-Reducing Bacterium from Permanently Cold Arctic Sediments.” *Environmental Microbiology* 6 (9): 887–902. https://doi.org/10.1111/j.1462-2920.2004.00665.x.

Remes, Philip M., Ping Yip, and Michael J. MacCoss. 2020. “Highly Multiplex Targeted Proteomics Enabled by Real-Time Chromatographic Alignment.” *Analytical Chemistry* 92 (17): 11809–17. https://doi.org/10.1021/acs.analchem.0c02075.

Riffle, Michael, Damon H. May, Emma Timmins-Schiffman, Molly P. Mikan, Daniel Jaschob, William Stafford Noble, and Brook L. Nunn. 2017. “MetaGOmics: A Web-Based Tool for Peptide-Centric Functional and Taxonomic Analysis of Metaproteomics Data.” *Proteomes* 6 (1): 2. https://doi.org/10.3390/proteomes6010002.

Rinta-Kanto, Johanna M., Shulei Sun, Shalabh Sharma, Ronald P. Kiene, and Mary Ann Moran. 2012. “Bacterial Community Transcription Patterns during a Marine Phytoplankton Bloom.” *Environmental Microbiology* 14 (1): 228–39. https://doi.org/10.1111/j.1462-2920.2011.02602.x.

Rocca, Jennifer D., Marie Simonin, Joanna R. Blaszczak, Jessica G. Ernakovich, Sean M. Gibbons, Firas S. Midani, and Alex D. Washburne. 2019. “The Microbiome Stress Project: Toward a Global Meta-Analysis of Environmental Stressors and Their Effects on Microbial Communities.” *Frontiers in Microbiology* 9 (January). https://doi.org/10.3389/fmicb.2018.03272.

Rolton, Anne, Lesley Rhodes, Kate S. Hutson, Laura Biessy, Tony Bui, Lincoln MacKenzie, Jane E. Symonds, and Kirsty F. Smith. 2022. “Effects of Harmful Algal Blooms on Fish and Shellfish Species: A Case Study of New Zealand in a Changing Environment.” *Toxins* 14 (5):

341. https://doi.org/10.3390/toxins14050341.

Rosenberger, George, Isabell Bludau, Uwe Schmitt, Moritz Heusel, Christie L. Hunter, Yansheng Liu, Michael J. MacCoss, et al. 2017. “Statistical Control of Peptide and Protein Error Rates in Large-Scale Targeted Data-Independent Acquisition Analyses.” *Nature Methods* 14 (9): 921–27. https://doi.org/10.1038/nmeth.4398.

Rotert, Kenneth R., Anthony P. Toste, and John G. Steiert. 1993. “Membrane Fatty Acid Analysis of Antarctic Bacteria.” *FEMS Microbiology Letters* 114 (3): 253–57. https://doi.org/10.1111/j.1574-6968.1993.tb06582.x.

RStudio Team. 2020. “RStudio: Integrated Development for R.” RStudio, PBC, Boston, MA. URL http://www.rstudio.com/.

Russo, David A., Narciso Couto, Andrew P. Beckerman, and Jagroop Pandhal. 2016. “A Metaproteomic Analysis of the Response of a Freshwater Microbial Community under Nutrient Enrichment.” *Frontiers in Microbiology* 7 (August). https://doi.org/10.3389/fmicb.2016.01172.

Sageman, Bradley B., Michael H. Gardner, John M. Armentrout, and Adam E. Murphy. 1998. “Stratigraphic Hierarchy of Organic Carbon-Rich Siltstones in Deep-Water Facies, Brushy Canyon Formation (Guadalupian), Delaware Basin, West Texas.” *Geology* 26 (5): 451–54. https://doi.org/10.1130/0091-7613(1998)026<0451:SHOOCR>2.3.CO;2.

Saito, Mak A., Erin M. Bertrand, Megan E. Duffy, David A. Gaylord, Noelle A. Held, William Judson Hervey, Robert L. Hettich, et al. 2019. “Progress and Challenges in Ocean Metaproteomics and Proposed Best Practices for Data Sharing.” *Journal of Proteome Research* 18 (4): 1461–76. https://doi.org/10.1021/acs.jproteome.8b00761.

Sajulga, Ray, Caleb Easterly, Michael Riffle, Bart Mesuere, Thilo Muth, Subina Mehta, Praveen Kumar, et al. 2020. “Survey of Metaproteomics Software Tools for Functional Microbiome Analysis.” *PLOS ONE* 15 (11): e0241503. https://doi.org/10.1371/journal.pone.0241503.

Salazar, Guillem, Francisco M. Cornejo-Castillo, Verónica Benítez-Barrios, Eugenio Fraile-Nuez, X. Antón Álvarez-Salgado, Carlos M. Duarte, Josep M. Gasol, and Silvia G. Acinas. 2016. “Global Diversity and Biogeography of Deep-Sea Pelagic Prokaryotes.” *The ISME Journal* 10 (3): 596–608. https://doi.org/10.1038/ismej.2015.137.

Sanudo-Wilhelmy, S. A., L. S. Cutter, R. Durazo, E. A. Smail, L. Gomez-Consarnau, E. A. Webb, M. G. Prokopenko, W. M. Berelson, and D. M. Karl. 2012. “Multiple B-Vitamin Depletion in Large Areas of the Coastal Ocean.” *Proceedings of the National Academy of Sciences* 109 (35): 14041–45. https://doi.org/10.1073/pnas.1208755109.

Sapp, Melanie, Antje Wichels, Karen H. Wiltshire, and Gunnar Gerdts. 2007. “Bacterial

Community Dynamics during the Winter-Spring Transition in the North Sea." *FEMS Microbiology Ecology* 59 (3): 622–37. https://doi.org/10.1111/j.1574-6941.2006.00238.x.

Sattler, Birgit, Hans Puxbaum, and Roland Psenner. 2001. "Bacterial Growth in Supercooled Cloud Droplets." *Geophysical Research Letters* 28 (2): 239–42. https://doi.org/10.1029/2000GL011684.

Saw, Jimmy H, Tanai Cardona, and Gustavo Montejano. 2021. "Complete Genome Sequencing of a Novel Gloeobacter Species from a Waterfall Cave in Mexico." *Genome Biology and Evolution* 13 (12): evab264. https://doi.org/10.1093/gbe/evab264.

Schauer, Kristine, Barbara Gouget, Marie Carrière, Agnès Labigne, and Hilde De Reuse. 2007. "Novel Nickel Transport Mechanism across the Bacterial Outer Membrane Energized by the TonB/ExbB/ExbD Machinery." *Molecular Microbiology* 63 (4): 1054–68. https://doi.org/10.1111/j.1365-2958.2006.05578.x.

Scheuerl, Thomas, Meirion Hopkins, Reuben W. Nowell, Damian W. Rivett, Timothy G. Barraclough, and Thomas Bell. 2020. "Bacterial Adaptation Is Constrained in Complex Communities." *Nature Communications* 11 (1): 754. https://doi.org/10.1038/s41467-020-14570-z.

Scott, Matthew, Carl W. Gunderson, Eduard M. Mateescu, Zhongge Zhang, and Terence Hwa. 2010. "Interdependence of Cell Growth and Gene Expression: Origins and Consequences." *Science (New York, N.Y.)* 330 (6007): 1099–1102. https://doi.org/10.1126/science.1192588.

Searle, Brian C., Lindsay K. Pino, Jarrett D. Egertson, Ying S. Ting, Robert T. Lawrence, Brendan X. MacLean, Judit Villén, and Michael J. MacCoss. 2018. "Chromatogram Libraries Improve Peptide Detection and Quantification by Data Independent Acquisition Mass Spectrometry." *Nature Communications* 9 (1): 5128. https://doi.org/10.1038/s41467-018-07454-w.

Searle, Brian C., Kristian E. Swearingen, Christopher A. Barnes, Tobias Schmidt, Siegfried Gessulat, Bernhard Küster, and Mathias Wilhelm. 2020. "Generating High Quality Libraries for DIA MS with Empirically Corrected Peptide Predictions." *Nature Communications* 11 (1): 1548. https://doi.org/10.1038/s41467-020-15346-1.

Seymour, Justin R., Shady A. Amin, Jean-Baptiste Raina, and Roman Stocker. 2017. "Zooming in on the Phycosphere: The Ecological Interface for Phytoplankton–Bacteria Relationships." *Nature Microbiology* 2 (7): 1–12. https://doi.org/10.1038/nmicrobiol.2017.65.

Sharma, Vagisha, Josh Eckels, Greg K. Taylor, Nicholas J. Shulman, Andrew B. Stergachis, Shannon A. Joyner, Ping Yan, et al. 2014. "Panorama: A Targeted Proteomics Knowledge Base." *Journal of Proteome Research* 13 (9): 4205–10. https://doi.org/10.1021/pr5006636.

Sharma, Vagisha, Josh Eckels, Birgit Schilling, Christina Ludwig, Jacob D. Jaffe, Michael J.

MacCoss, and Brendan MacLean. 2018. “Panorama Public: A Public Repository for Quantitative Data Sets Processed in Skyline.” *Molecular & Cellular Proteomics : MCP* 17 (6): 1239–44. https://doi.org/10.1074/mcp.RA117.000543.

Shimohata, Nobuyuki, Shinobu Chiba, Naoya Saikawa, Koreaki Ito, and Yoshinori Akiyama. 2002. “The Cpx Stress Response System of Escherichia Coli Senses Plasma Membrane Proteins and Controls HtpX, a Membrane Protease with a Cytosolic Active Site.” *Genes to Cells: Devoted to Molecular & Cellular Mechanisms* 7 (7): 653–62. https://doi.org/10.1046/j.1365-2443.2002.00554.x.

Shin, HyeonSeok, Eunju Lee, Jongoh Shin, So-Ra Ko, Hyung-Seok Oh, Chi-Yong Ahn, Hee-Mock Oh, Byung-Kwan Cho, and Suhyung Cho. 2018. “Elucidation of the Bacterial Communities Associated with the Harmful Microalgae Alexandrium Tamarense and Cochlodinium Polykrikoides Using Nanopore Sequencing.” *Scientific Reports* 8 (1): 5323. https://doi.org/10.1038/s41598-018-23634-6.

Showalter, G. M., and J. W. Deming. 2018. “Low-temperature Chemotaxis, Halotaxis and Chemohalotaxis by the Psychrophilic Marine Bacterium *Colwellia Psychrerythraea* 34H.” *Environmental Microbiology Reports* 10 (1): 92–101. https://doi.org/10.1111/1758-2229.12610.

Shrestha, Him K., Manasa R. Appidi, Manuel I. Villalobos Solis, Jia Wang, Dana L. Carper, Leah Burdick, Dale A. Pelletier, Mitchel J. Doktycz, Robert L. Hettich, and Paul E. Abraham. 2021. “Metaproteomics Reveals Insights into Microbial Structure, Interactions, and Dynamic Regulation in Defined Communities as They Respond to Environmental Disturbance.” *BMC Microbiology* 21 (1): 308. https://doi.org/10.1186/s12866-021-02370-4.

Shultis, David D., Michael D. Purdy, Christian N. Banchs, and Michael C. Wiener. 2006. “Outer Membrane Active Transport: Structure of the BtuB:TonB Complex.” *Science* 312 (5778): 1396–99. https://doi.org/10.1126/science.1127694.

Sichert, Andreas, Christopher H. Corzett, Matthew S. Schechter, Frank Unfried, Stephanie Markert, Dörte Becher, Antonio Fernandez-Guerra, et al. 2020. “Verrucomicrobia Use Hundreds of Enzymes to Digest the Algal Polysaccharide Fucoidan.” *Nature Microbiology* 5 (8): 1026–39. https://doi.org/10.1038/s41564-020-0720-2.

Silale, Augustinas, and Bert van den Berg. 2023. “TonB-Dependent Transport Across the Bacterial Outer Membrane.” *Annual Review of Microbiology* 77 (Volume 77, 2023): 67–88. https://doi.org/10.1146/annurev-micro-032421-111116.

Simon, Meinhard, and Farooq Azam. 1989. “Protein Content and Protein Synthesis Rates of Planktonic Marine Bacteria.” *Marine Ecology Progress Series* 51 (3): 201–13.

Skerratt, J. H., P. D. Nichols, C. A. Mancuso, S. R. James, S. J. Dobson, T. A. McMeekin, and H. Burton. 1991. "The Phospholipid Ester-Linked Fatty Acid Composition of Members of the Family Halomonadaceae and Genus Flavobacterium: A Chemotaxonomic Guide." *Systematic and Applied Microbiology* 14 (1): 8–13. https://doi.org/10.1016/S0723-2020(11)80355-3.

Soutourina, Olga A., and Philippe N. Bertin. 2003. "Regulation Cascade of Flagellar Expression in Gram-Negative Bacteria." *FEMS Microbiology Reviews* 27 (4): 505–23. https://doi.org/10.1016/S0168-6445(03)00064-0.

Spaans, Sebastiaan K., Ruud A. Weusthuis, John van der Oost, and Servé W. M. Kengen. 2015. "NADPH-Generating Systems in Bacteria and Archaea." *Frontiers in Microbiology* 6 (July). https://doi.org/10.3389/fmicb.2015.00742.

Spencer, Ronald J., Nancy Møller, and John H. Weare. 1990. "The Prediction of Mineral Solubilities in Natural Waters: A Chemical Equilibrium Model for the Na□K□Ca□Mg□Cl□SO4□H2O System at Temperatures below 25°C." *Geochimica et Cosmochimica Acta* 54 (3): 575–90. https://doi.org/10.1016/0016-7037(90)90354-N.

Stockton, Lindsey, Cameron Simonsen, and Susan Seago. 2017. "Nitrous Oxide–Induced Vitamin B12 Deficiency." *Proceedings (Baylor University. Medical Center)* 30 (2): 171–72.

Sunagawa, Shinichi, Luis Pedro Coelho, Samuel Chaffron, Jens Roat Kultima, Karine Labadie, Guillem Salazar, Bardya Djahanschiri, et al. 2015. "Structure and Function of the Global Ocean Microbiome." *Science* 348 (6237): 1261359. https://doi.org/10.1126/science.1261359.

Sunagawa, Shinichi, Silvia G. Acinas, Peer Bork, Chris Bowler, Damien Eveillard, Gabriel Gorsky, Lionel Guidi, et al. 2020. "Tara Oceans: Towards Global Ocean Ecosystems Biology." *Nature Reviews Microbiology* 18 (8): 428–45. https://doi.org/10.1038/s41579-020-0364-5.

Tada, Yuya, Akito Taniguchi, Yuki Sato-Takabe, and Koji Hamasaki. 2012. "Growth and Succession Patterns of Major Phylogenetic Groups of Marine Bacteria during a Mesocosm Diatom Bloom." *Journal of Oceanography* 68 (4): 509–19. https://doi.org/10.1007/s10872-012-0114-z.

Tanaka, Kenya, Ginga Shimakawa, and Shuji Nakanishi. 2020. "Time-of-Day-Dependent Responses of Cyanobacterial Cellular Viability against Oxidative Stress." *Scientific Reports* 10 (1): 20029. https://doi.org/10.1038/s41598-020-77141-8.

Tang, Kai, Nianzhi Jiao, Keshao Liu, Yao Zhang, and Shuhui Li. 2012. "Distribution and Functions of TonB-Dependent Transporters in Marine Bacteria and Environments: Implications for Dissolved Organic Matter Utilization." *PLOS ONE* 7 (7): e41204.

https://doi.org/10.1371/journal.pone.0041204.

Teeling, Hanno, Bernhard M Fuchs, Christin M Bennke, Karen Krüger, Meghan Chafee, Lennart Kappelmann, Greta Reintjes, et al. 2016. "Recurring Patterns in Bacterioplankton Dynamics during Coastal Spring Algae Blooms." *eLife* 5:e11888. https://doi.org/10.7554/eLife.11888.

The, Matthew, Michael J. MacCoss, William S. Noble, and Lukas Käll. 2016. "Fast and Accurate Protein False Discovery Rates on Large-Scale Proteomics Data Sets with Percolator 3.0." *Journal of the American Society for Mass Spectrometry* 27 (11): 1719–27. https://doi.org/10.1007/s13361-016-1460-7.

Timmins-Schiffman, Emma, Damon H. May, Molly Mikan, Michael Riffle, Chris Frazar, H. R. Harvey, William S. Noble, and Brook L. Nunn. 2017. "Critical Decisions in Metaproteomics: Achieving High Confidence Protein Annotations in a Sea of Unknowns." *The ISME Journal* 11 (2): 309–14. https://doi.org/10.1038/ismej.2016.132.

Timmins-Schiffman, Emma, Molly P. Mikan, Ying Sonia Ting, H. Rodger Harvey, and Brook L. Nunn. 2018. "MS Analysis of a Dilution Series of Bacteria:Phytoplankton to Improve Detection of Low Abundance Bacterial Peptides." *Scientific Reports* 8 (1): 9276. https://doi.org/10.1038/s41598-018-27650-4.

Trevathan-Tackett, Stacey M., Craig D. H. Sherman, Megan J. Huggett, Alexandra H. Campbell, Bonnie Laverock, Valentina Hurtado-McCormick, Justin R. Seymour, et al. 2019. "A Horizon Scan of Priorities for Coastal Marine Microbiome Research." *Nature Ecology & Evolution* 3 (11): 1509–20. https://doi.org/10.1038/s41559-019-0999-7.

Tokunaga, Hiroko, Kenjiro Mitsuo, Sachiyo Ichinose, Akira Omori, Antonio Ventosa, Taiji Nakae, and Masao Tokunaga. 2004. "Salt-Inducible Multidrug Efflux Pump Protein in the Moderately Halophilic Bacterium Chromohalobacter Sp." *Applied and Environmental Microbiology* 70 (8): 4424–31. https://doi.org/10.1128/AEM.70.8.4424-4431.2004.

Torres-Valdés, Sinhué, Takamasa Tsubouchi, Emily Davey, Igor Yashayaev, and Sheldon Bacon. 2016. "Relevance of Dissolved Organic Nutrients for the Arctic Ocean Nutrient Budget." *Geophysical Research Letters* 43 (12): 6418–26. https://doi.org/10.1002/2016GL069245.

Turley, Brendan D, Mandy Karnauskas, Matthew D Campbell, David S Hanisko, and Christopher R Kelble. 2022. "Relationships between Blooms of *Karenia Brevis* and Hypoxia across the West Florida Shelf." *Harmful Algae* 114 (May):102223. https://doi.org/10.1016/j.hal.2022.102223.

Vance, Steven D., Laura M. Barge, Silvana S.S. Cardoso, and Julyan H.E. Cartwright. 2019. "Self-Assembling Ice Membranes on Europa: Brinicle Properties, Field Examples, and Possible

Energetic Systems in Icy Ocean Worlds." *Astrobiology* 19 (5): 685–95. https://doi.org/10.1089/ast.2018.1826.

Van Rossum, Guido, and Fred L. Drake. 2009. *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace.

Venter, Eli, Richard D. Smith, and Samuel H. Payne. 2011. "Proteogenomic Analysis of Bacteria and Archaea: A 46 Organism Case Study." *PLoS ONE* 6 (11): e27587. https://doi.org/10.1371/journal.pone.0027587.

Ventosa, Antonio, Joaquin Nieto, and Aharon Oren. 1998. "Biology of Moderately Halophilic Aerobic Bacteria." *Microbiology and Molecular Biology Reviews* 62 (June):504–44. https://doi.org/10.1128/MMBR.62.2.504-544.1998.

Viborg, Alexander Holm, Nicolas Terrapon, Vincent Lombard, Gurvan Michel, Mirjam Czjzek, Bernard Henrissat, and Harry Brumer. 2019. "A Subfamily Roadmap of the Evolutionarily Diverse Glycoside Hydrolase Family 16 (GH16)." *The Journal of Biological Chemistry* 294 (44): 15973–86. https://doi.org/10.1074/jbc.RA119.010619.

Vidal-Melgosa, Silvia, Andreas Sichert, T. Ben Francis, Daniel Bartosik, Jutta Niggemann, Antje Wichels, William G. T. Willats, et al. 2021. "Diatom Fucan Polysaccharide Precipitates Carbon during Algal Blooms." *Nature Communications* 12 (February):1150. https://doi.org/10.1038/s41467-021-21009-6.

Visser, Ingmar, and Maarten Speekenbrink. 2010. "depmixS4: An R Package for Hidden Markov Models." *Journal of Statistical Software* 36 (August):1–21. https://doi.org/10.18637/jss.v036.i07.

Vogel, Jörg, and Ben F. Luisi. 2011. "Hfq and Its Constellation of RNA." *Nature Reviews. Microbiology* 9 (8): 578–89. https://doi.org/10.1038/nrmicro2615.

Vytvytska, O., J. S. Jakobsen, G. Balcunaite, J. S. Andersen, M. Baccarini, and A. von Gabain. 1998. "Host Factor I, Hfq, Binds to Escherichia Coli ompA mRNA in a Growth Rate-Dependent Fashion and Regulates Its Stability." *Proceedings of the National Academy of Sciences of the United States of America* 95 (24): 14118–23. https://doi.org/10.1073/pnas.95.24.14118.

Vytvytska, Oresta, Isabella Moll, Vladimir R. Kaberdin, Alexander von Gabain, and Udo Bläsi. 2000. "Hfq (HF1) Stimulates ompA mRNA Decay by Interfering with Ribosome Binding." *Genes & Development* 14 (9): 1109–18.

Waldbauer, Jacob R., Sébastien Rodrigue, Maureen L. Coleman, and Sallie W. Chisholm. 2012. "Transcriptome and Proteome Dynamics of a Light-Dark Synchronized Bacterial Cell Cycle." *PLOS ONE* 7 (8): e43432. https://doi.org/10.1371/journal.pone.0043432.

Walker, Jordan R., Alaina C. Woods, Mary K. Pierce, Jamie L. Steichen, Antonietta Quigg, Karl Kaiser, and Jessica M. Labonté. 2022. "Functionally Diverse Microbial Communities Show Resilience in Response to a Record-Breaking Rain Event." *ISME Communications* 2 (1): 1–11. https://doi.org/10.1038/s43705-022-00162-z.

Walters, R. F. S., and W. F. DeGrado. 2006. "Helix-Packing Motifs in Membrane Proteins." *Proceedings of the National Academy of Sciences of the United States of America* 103 (37): 13658–63. https://doi.org/10.1073/pnas.0605878103.

Wan, Xia, Yun-Feng Peng, Xue-Rong Zhou, Yang-Min Gong, Feng-Hong Huang, and Gabriel Moncalián. 2016. "Effect of Cerulenin on Fatty Acid Composition and Gene Expression Pattern of DHA-Producing Strain Colwellia Psychrerythraea Strain 34H." *Microbial Cell Factories* 15 (1): 30. https://doi.org/10.1186/s12934-016-0431-9.

Wang, H., and R. P. Gunsalus. 2000. "The nrfA and nirB Nitrite Reductase Operons in Escherichia Coli Are Expressed Differently in Response to Nitrate than to Nitrite." *Journal of Bacteriology* 182 (20): 5813–22. https://doi.org/10.1128/jb.182.20.5813-5822.2000.

Wang, Ying, Jiawang Chen, Jin Guo, Zhenwu Yu, Yuan Lin, and Yuhong Wang. 2024. "Advances and Development in Sampling Techniques for Marine Water Resources: A Comprehensive Review." *Frontiers in Marine Science* 11 (April). https://doi.org/10.3389/fmars.2024.1365019.

Watzer, Björn, Philipp Spät, Niels Neumann, Moritz Koch, Roman Sobotka, Boris Macek, Oliver Hennrich, and Karl Forchhammer. 2019. "The Signal Transduction Protein PII Controls Ammonium, Nitrate and Urea Uptake in Cyanobacteria." *Frontiers in Microbiology* 10. https://doi.org/10.3389/fmicb.2019.01428.

Wells, Le, and Jw Deming. 2006. "Characterization of a Cold-Active Bacteriophage on Two Psychrophilic Marine Hosts." *Aquatic Microbial Ecology* 45 (October):15–29. https://doi.org/10.3354/ame045015.

Wells, Llyd E., and Jody W. Deming. 2006. "Modelled and Measured Dynamics of Viruses in Arctic Winter Sea-Ice Brines." *Environmental Microbiology* 8 (6): 1115–21. https://doi.org/10.1111/j.1462-2920.2006.00984.x.

Welsh, David T. 2000. "Ecological Significance of Compatible Solute Accumulation by Micro-Organisms: From Single Cells to Global Climate." *FEMS Microbiology Reviews* 24 (3): 263–90. https://doi.org/10.1111/j.1574-6976.2000.tb00542.x.

"What Is a Harmful Algal Bloom? | National Oceanic and Atmospheric Administration." Accessed May 3, 2024. https://www.noaa.gov/what-is-harmful-algal-bloom.

Whitman, William B., David C. Coleman, and William J. Wiebe. 1998. "Prokaryotes: The Unseen Majority." *Proceedings of the National Academy of Sciences of the United States of America*

95 (12): 6578–83.

Wickham, Hadley. 2009. *Ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer. https://doi.org/10.1007/978-0-387-98141-3.

Williams, Timothy J., David Wilkins, Emilie Long, Flavia Evans, Mathew Z. DeMaere, Mark J. Raftery, and Ricardo Cavicchioli. 2013. "The Role of Planktonic Flavobacteria in Processing Algal Organic Matter in Coastal East Antarctica Revealed Using Metagenomics and Metaproteomics." *Environmental Microbiology* 15 (5): 1302–17. https://doi.org/10.1111/1462-2920.12017.

Wu, Christine C., and Michael J. MacCoss. 2002. "Shotgun Proteomics: Tools for the Analysis of Complex Biological Systems." *Current Opinion in Molecular Therapeutics* 4 (3): 242–50.

Wu, Guoyao, and Sidney M. Morris. 1998. "Arginine Metabolism: Nitric Oxide and Beyond." *Biochemical Journal* 336 (1): 1–17. https://doi.org/10.1042/bj3360001.

Yang, Caiyun, Yi Li, Benjamin Zhou, Yanyan Zhou, Wei Zheng, Yun Tian, Joy D. Van Nostrand, et al. 2015. "Illumina Sequencing-Based Analysis of Free-Living Bacterial Community Dynamics during an Akashiwo Sanguine Bloom in Xiamen Sea, China." *Scientific Reports* 5 (1): 8476. https://doi.org/10.1038/srep08476.

Yun, Sung Ho, Edmond Changkyun Park, Sang-Yeop Lee, Hayoung Lee, Chi-Won Choi, Yoon-Sun Yi, Hyun-Joo Ro, et al. 2018. "Antibiotic Treatment Modulates Protein Components of Cytotoxic Outer Membrane Vesicles of Multidrug-Resistant Clinical Strain, Acinetobacter Baumannii DU202." *Clinical Proteomics* 15 (1): 28. https://doi.org/10.1186/s12014-018-9204-2.

Zamkovaya, Tatyana, Jamie S Foster, Valérie de Crécy-Lagard, and Ana Conesa. 2021. "A Network Approach to Elucidate and Prioritize Microbial Dark Matter in Microbial Communities." *The ISME Journal* 15 (1): 228–44. https://doi.org/10.1038/s41396-020-00777-x.

Zhang, Fuzhong, Mario Ouellet, Tanveer S. Batth, Paul D. Adams, Christopher J. Petzold, Aindrila Mukhopadhyay, and Jay D. Keasling. 2012. "Enhancing Fatty Acid Production by the Expression of the Regulatory Transcription Factor FadR." *Metabolic Engineering* 14 (6): 653–60. https://doi.org/10.1016/j.ymben.2012.08.009.

Zhang, Houjin, and Kang Ning. 2015. "The Tara Oceans Project: New Opportunities and Greater Challenges Ahead." *Genomics, Proteomics & Bioinformatics* 13 (5): 275–77. https://doi.org/10.1016/j.gpb.2015.08.003.

Zhao, Jinzhi, Yi Yang, Hua Xu, Jianxujie Zheng, Chengpin Shen, Tian Chen, Tao Wang, et al. 2023. "Data-Independent Acquisition Boosts Quantitative Metaproteomics for Deep

Characterization of Gut Microbiota.” *Npj Biofilms and Microbiomes* 9 (1): 1–14. https://doi.org/10.1038/s41522-023-00373-9.

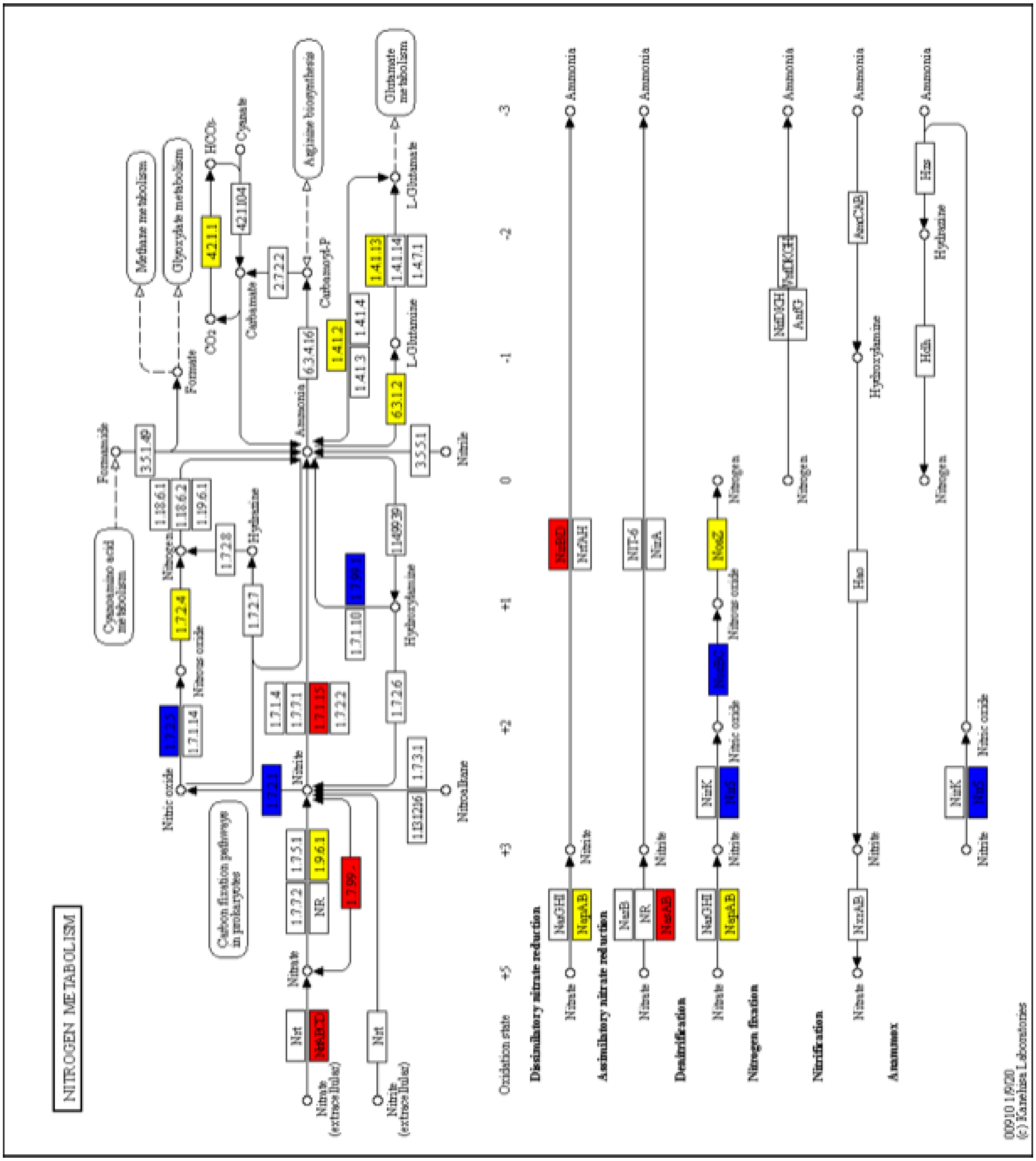
Zheng, Lei, Huipeng Wang, Cao Liu, Shurong Zhang, Aizhong Ding, En Xie, Jian Li, and Shengrui Wang. 2021. “Prediction of Harmful Algal Blooms in Large Water Bodies Using the Combined EFDC and LSTM Models.” *Journal of Environmental Management* 295 (October):113060. https://doi.org/10.1016/j.jenvman.2021.113060.

Zhou, Jin, Yong-min Lao, Jun-ting Song, Hui Jin, Jian-ming Zhu, and Zhong-hua Cai. 2020. “Temporal Heterogeneity of Microbial Communities and Metabolic Activities during a Natural Algal Bloom.” *Water Research* 183 (September):116020. https://doi.org/10.1016/j.watres.2020.116020.

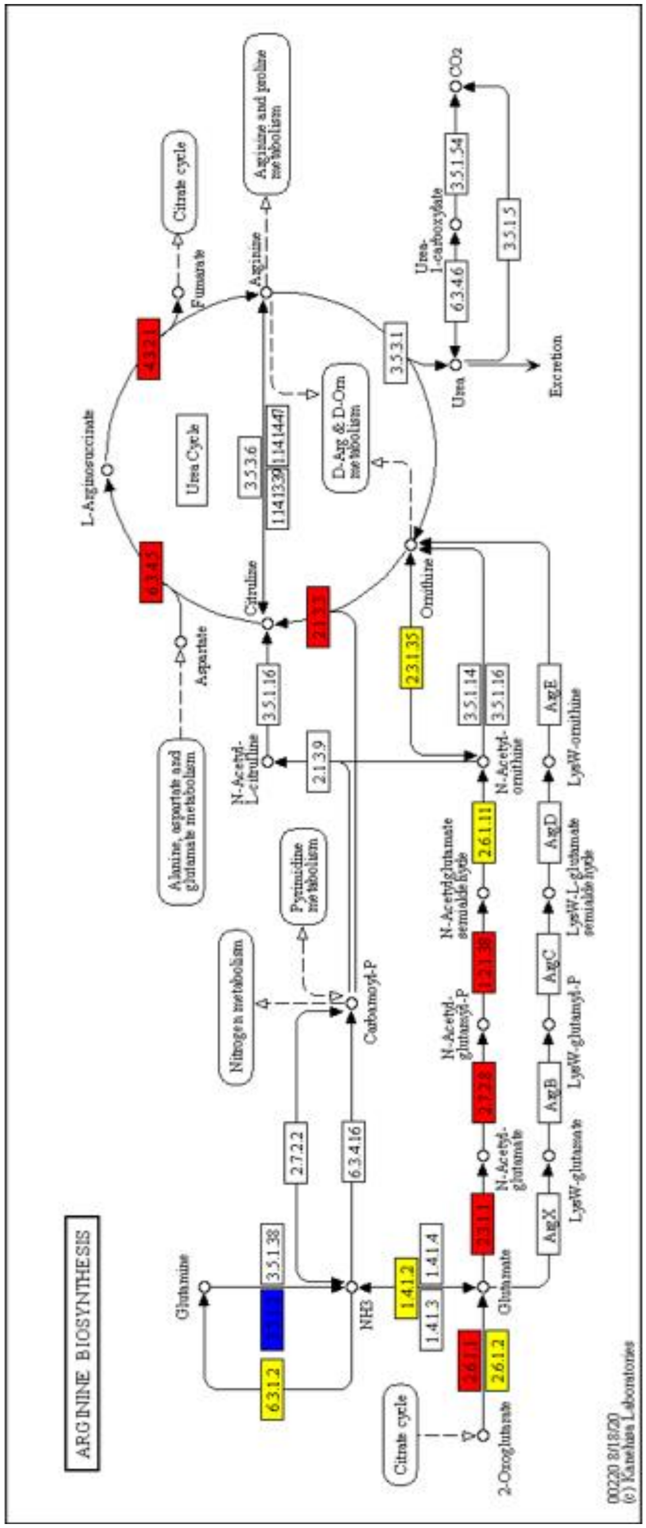
Zolotov, Mikhail Y. 2007. “An Oceanic Composition on Early and Today’s Enceladus.” *Geophysical Research Letters* 34 (23). https://doi.org/10.1029/2007GL031234.

Zubarev, Roman A. 2013. “The Challenge of the Proteome Dynamic Range and Its Implications for In-Depth Proteomics.” *Proteomics* 13 (5): 723–26. https://doi.org/10.1002/pmic.201200451.

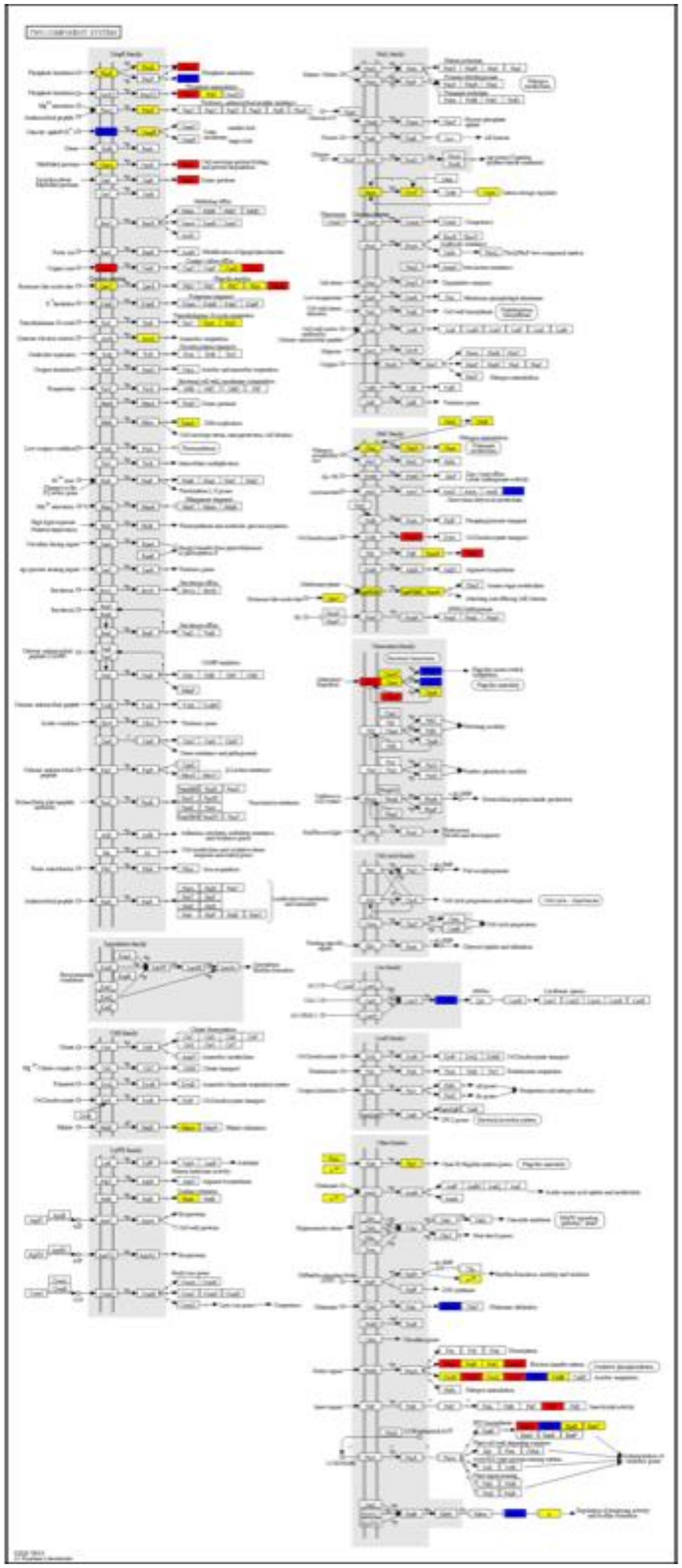
Appendix A: KEGG PATHWAY MAPS OF ENRICHED PATHWAYS IN Cp34H FOUR MONTH INCUBATIONS



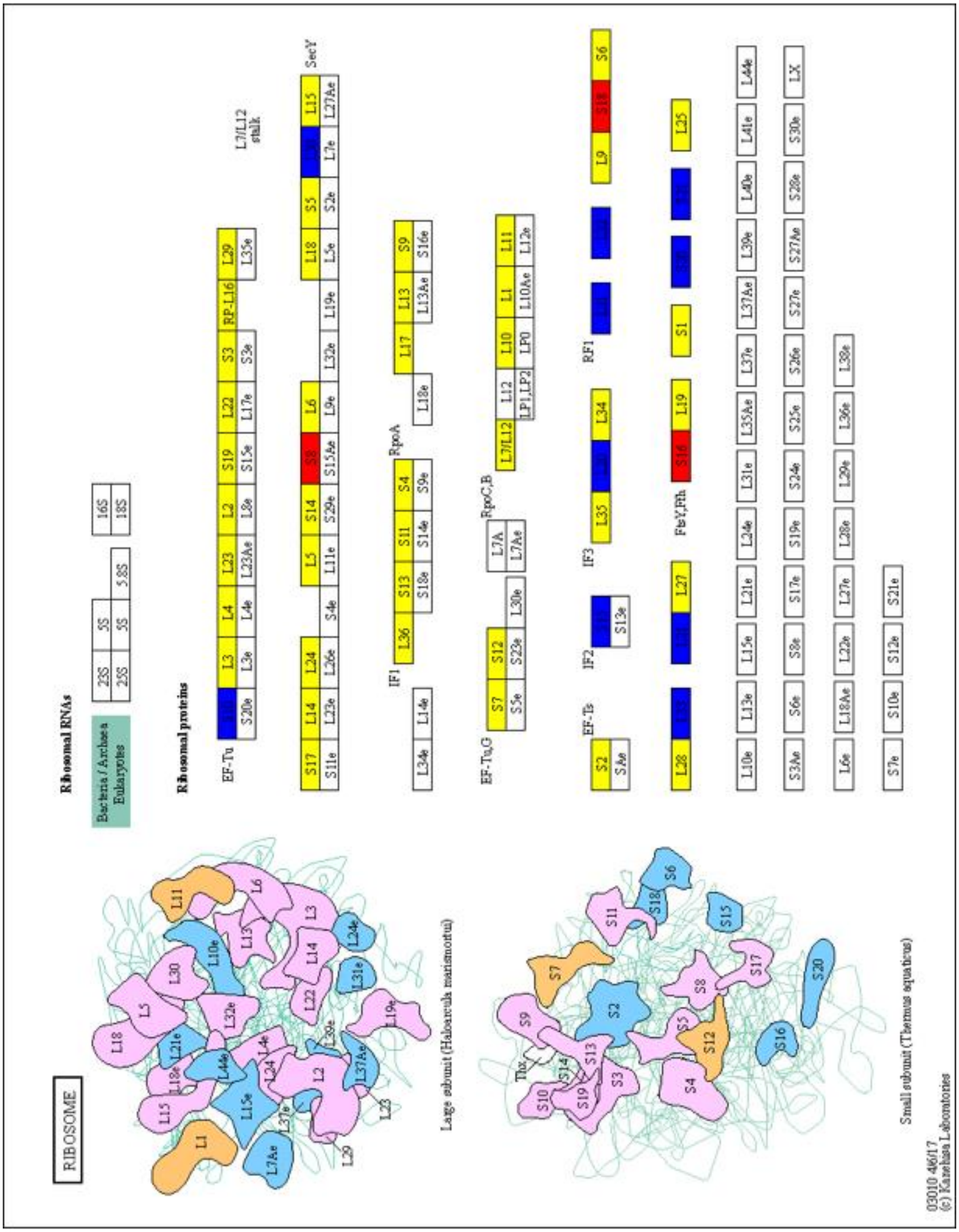
A.1: Nitrogen metabolism in ASW+nuts at -5°C. KEGG pathway map of nitrogen metabolism with proteins detected in ASW^{+nuts} at -5°C as compared to control cells (red: LFC ≥ 0.5, blue: LFC ≤- 0.5, yellow: LFC between 0.5 and -0.5; all: Z-score ≥ 2).



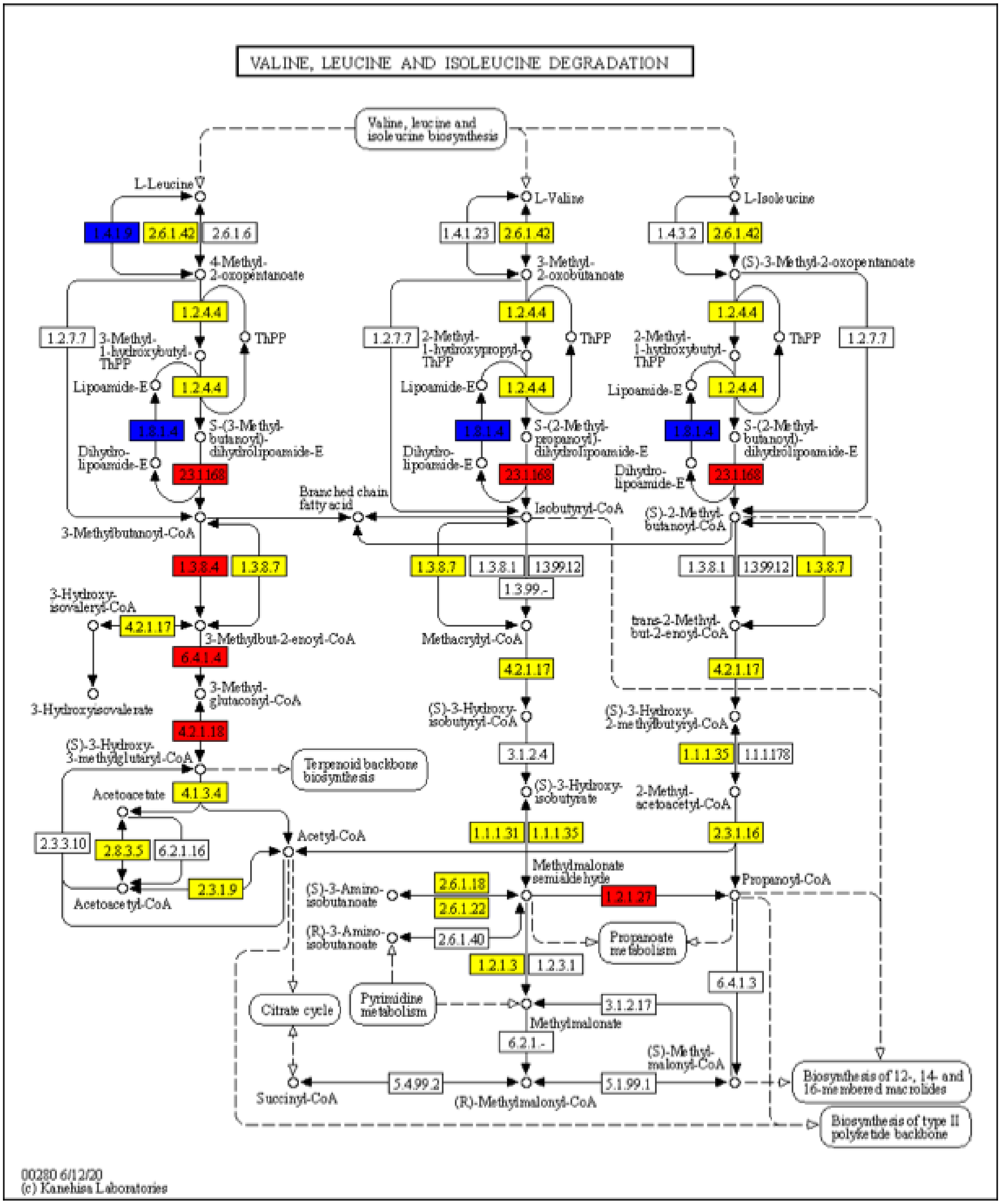
A.2: Arginine biosynthesis in ASW+nuts at -5°C. KEGG pathway map of arginine biosynthesis with proteins detected in ASW^{+nuts} at -5°C as compared to control cells (red: LFC ≥ 0.5, blue: LFC ≤ -0.5, yellow: LFC between 0.5 and -0.5; all: Z-score ≥ 2).



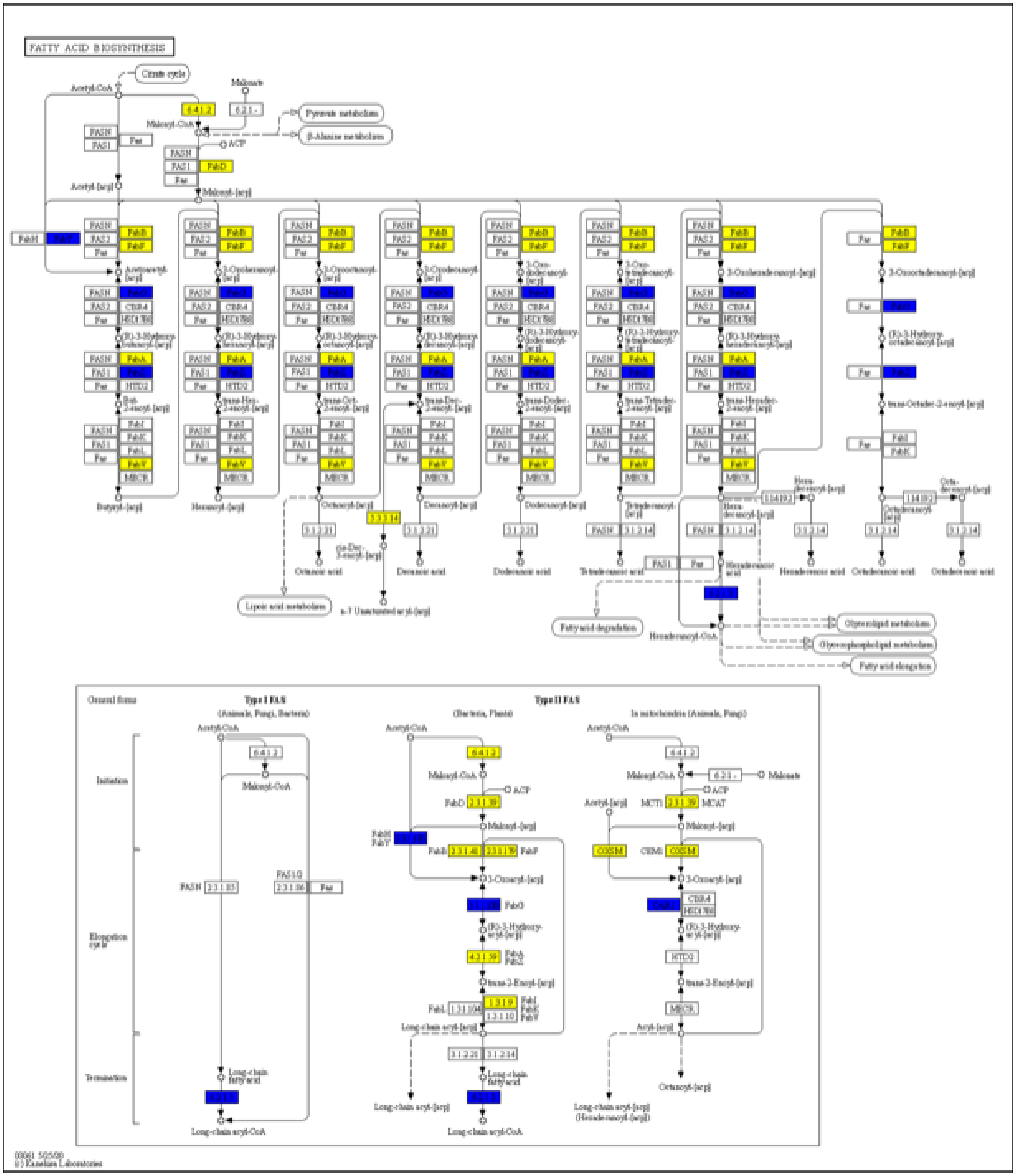
A.3: Two-component systems in ASW-nuts at -5°C. KEGG pathway map of two-component systems with proteins detected in ASW^{-nuts} at -5°C as compared to control cells (red: LFC ≥ 0.5, blue: LFC ≤ -0.5, yellow: LFC between 0.5 and -0.5; all: Z-score ≥ 2).



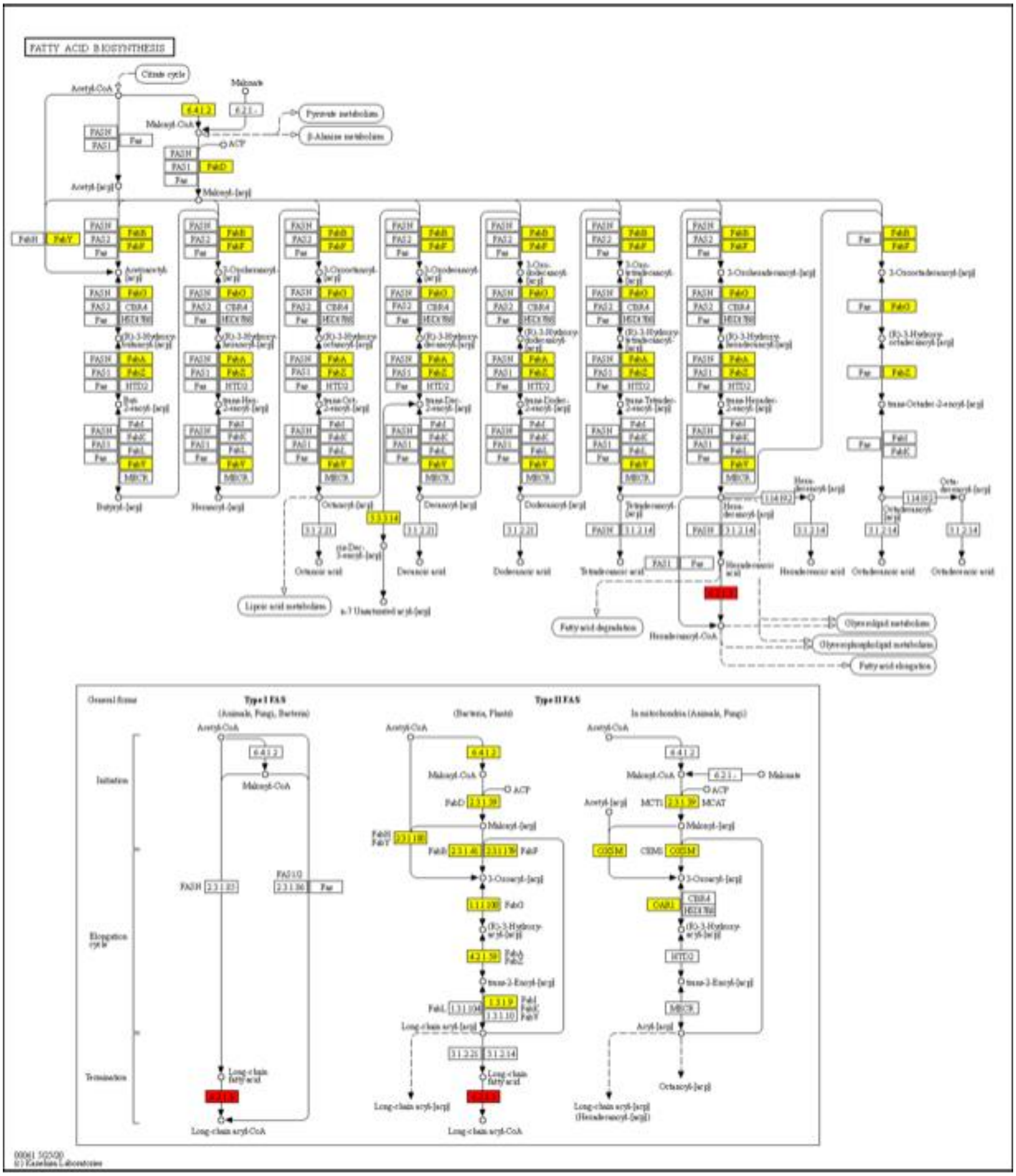
A.4: Ribosomes in brine+nuts at -5°C. KEGG pathway map of individual ribosomal components with proteins detected in brine[+nuts] at -5°C as compared to control cells (red: LFC ≥ 0.5, blue: LFC ≤ -0.5, yellow: LFC between 0.5 and -0.5; all: Z-score ≥ 2).



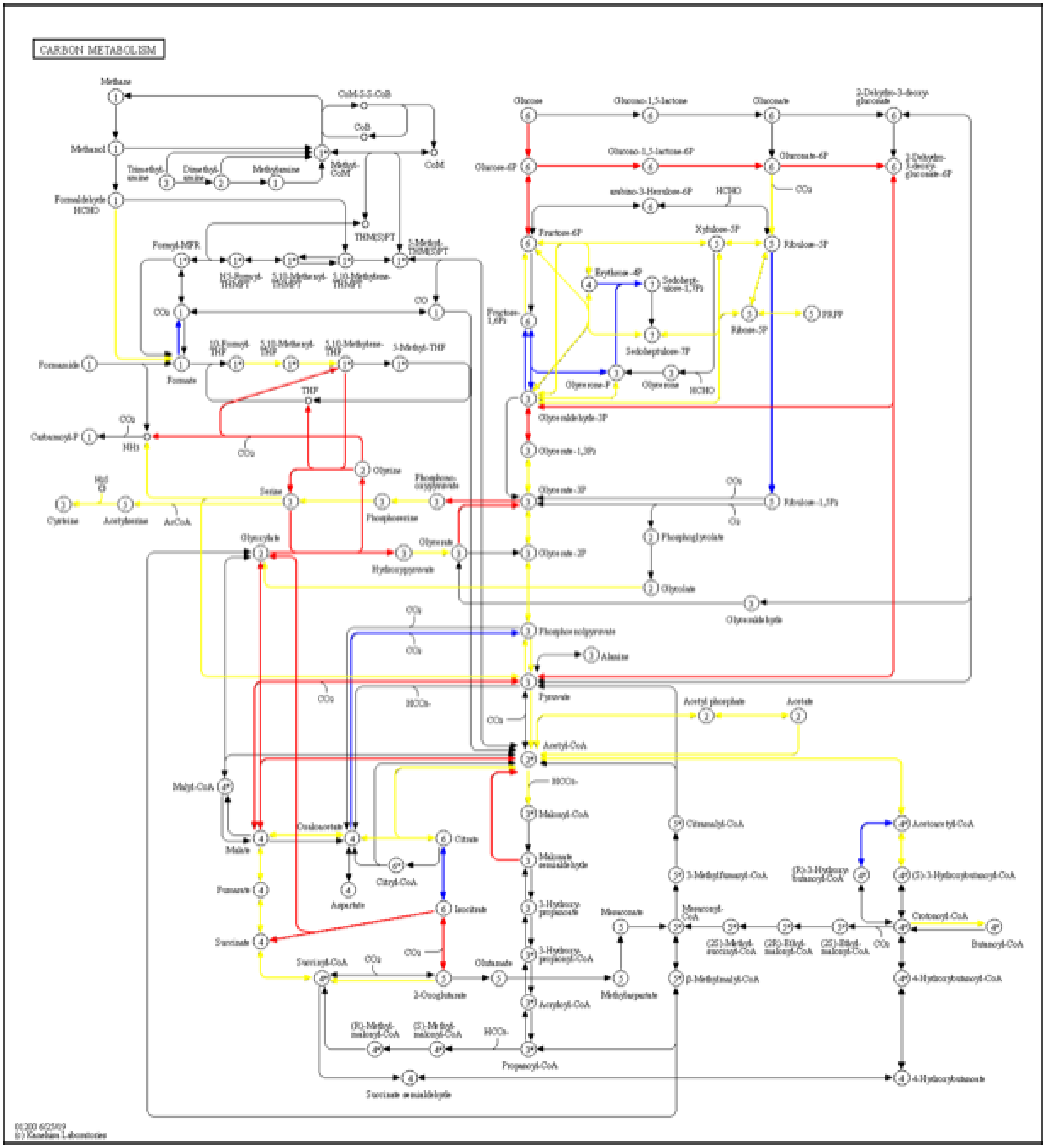
A.5: Valine, leucine and isoleucine degradation in brine+nuts at -5°C. KEGG pathway map of branched-chain amino acid degradation with proteins detected in brine[+nuts] at -5°C as compared to control cells (red: LFC ≥ 0.5, blue: LFC ≤ -0.5, yellow: LFC between 0.5 and -0.5; all: Z-score ≥ 2).



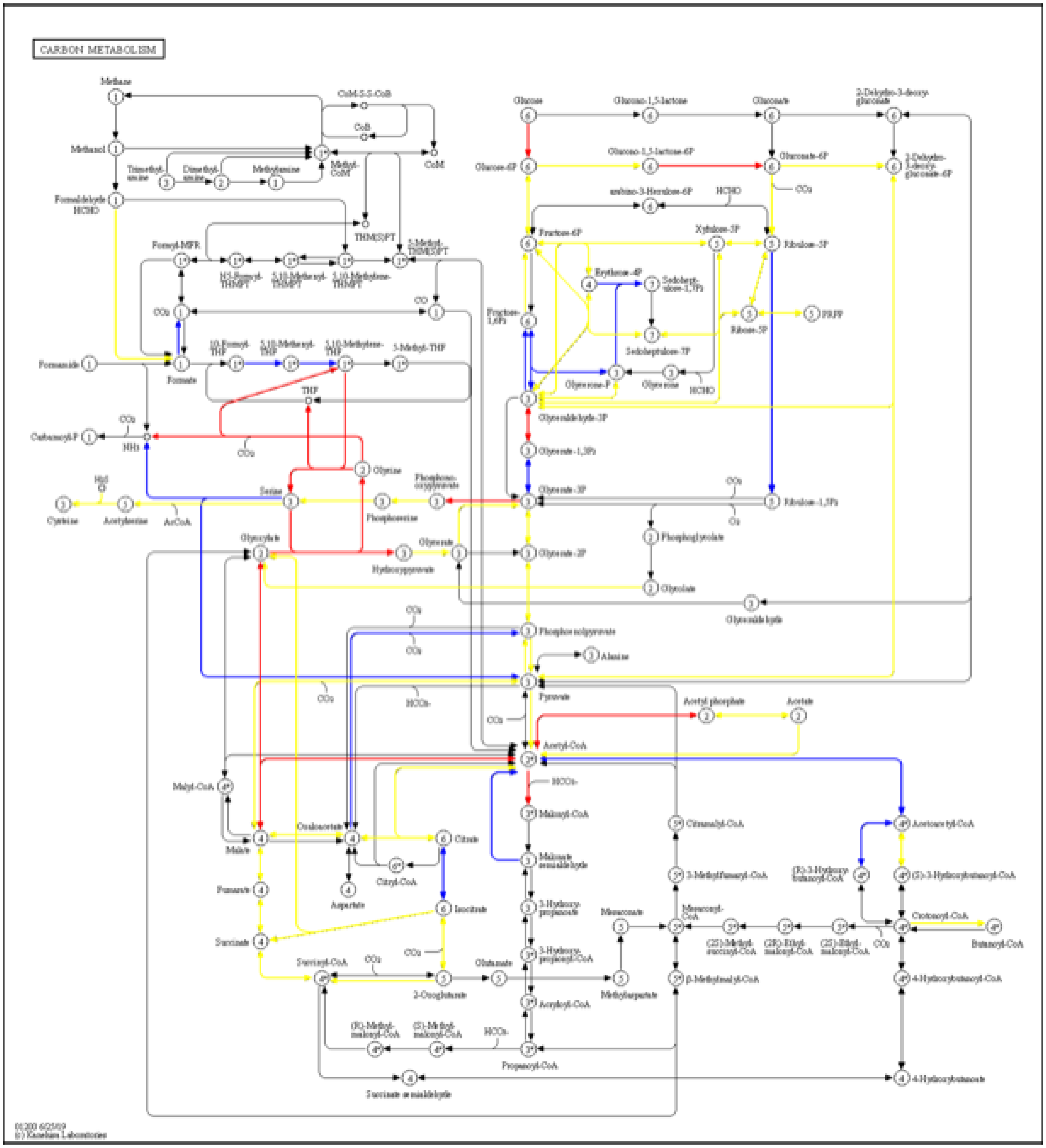
A.6: Fatty acid biosynthesis in brine+nuts at -5°C. KEGG pathway map of fatty acid biosynthesis with proteins detected in brine+nuts at -5°C as compared to control cells (red: LFC ≥ 0.5, blue: LFC ≤ -0.5, yellow: LFC between 0.5 and -0.5; all: Z-score ≥ 2).



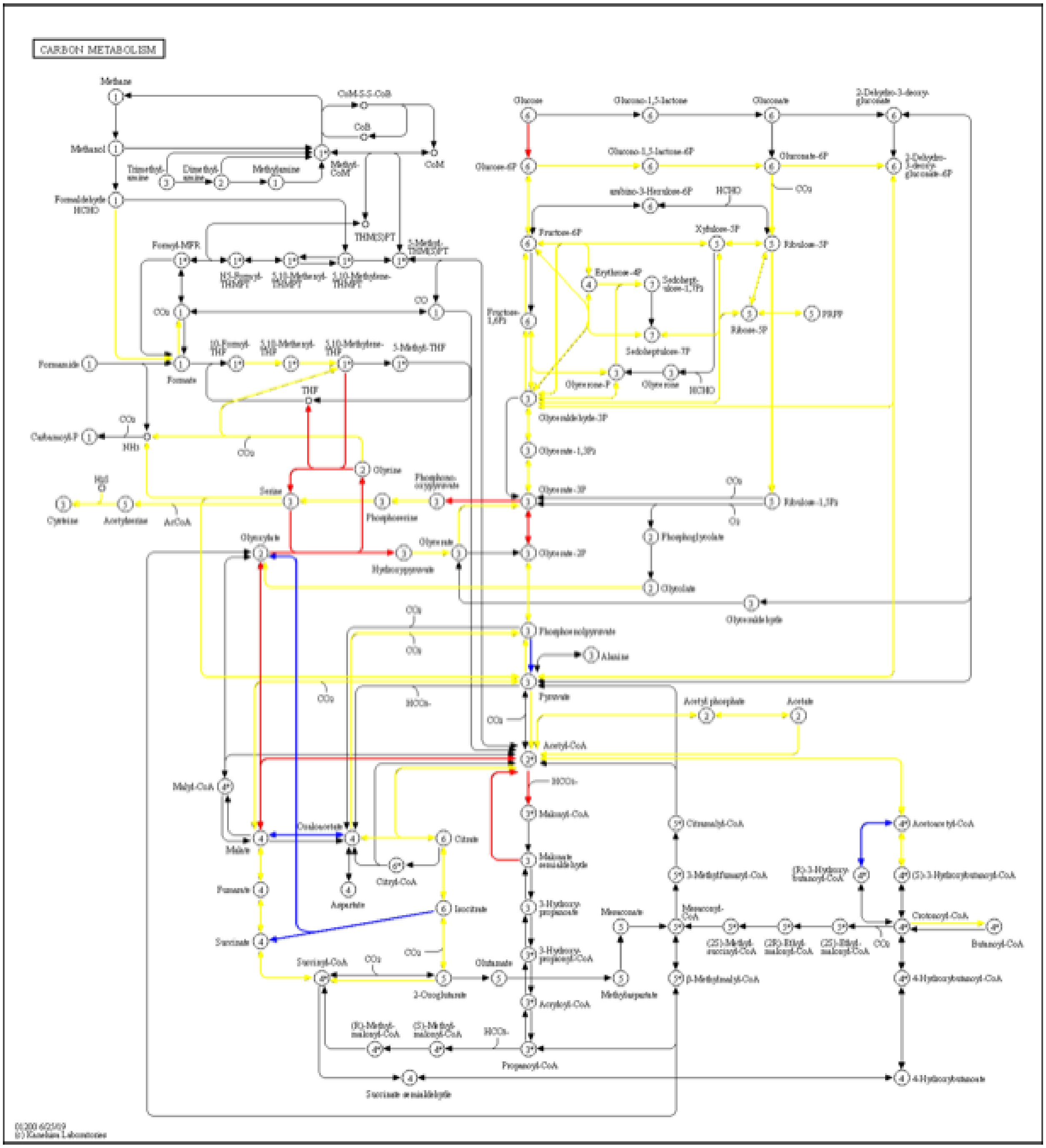
A.7: Phenylalanine, tyrosine and tryptophan biosynthesis in brine-nuts at -5°C. KEGG pathway map of phenylalanine, tyrosine and tryptophan biosynthesis with proteins detected in brine-nuts at -5°C as compared to control cells (red: LFC ≥ 0.5, blue: LFC ≤ -0.5, yellow: LFC between 0.5 and -0.5; all: Z-score ≥ 2).



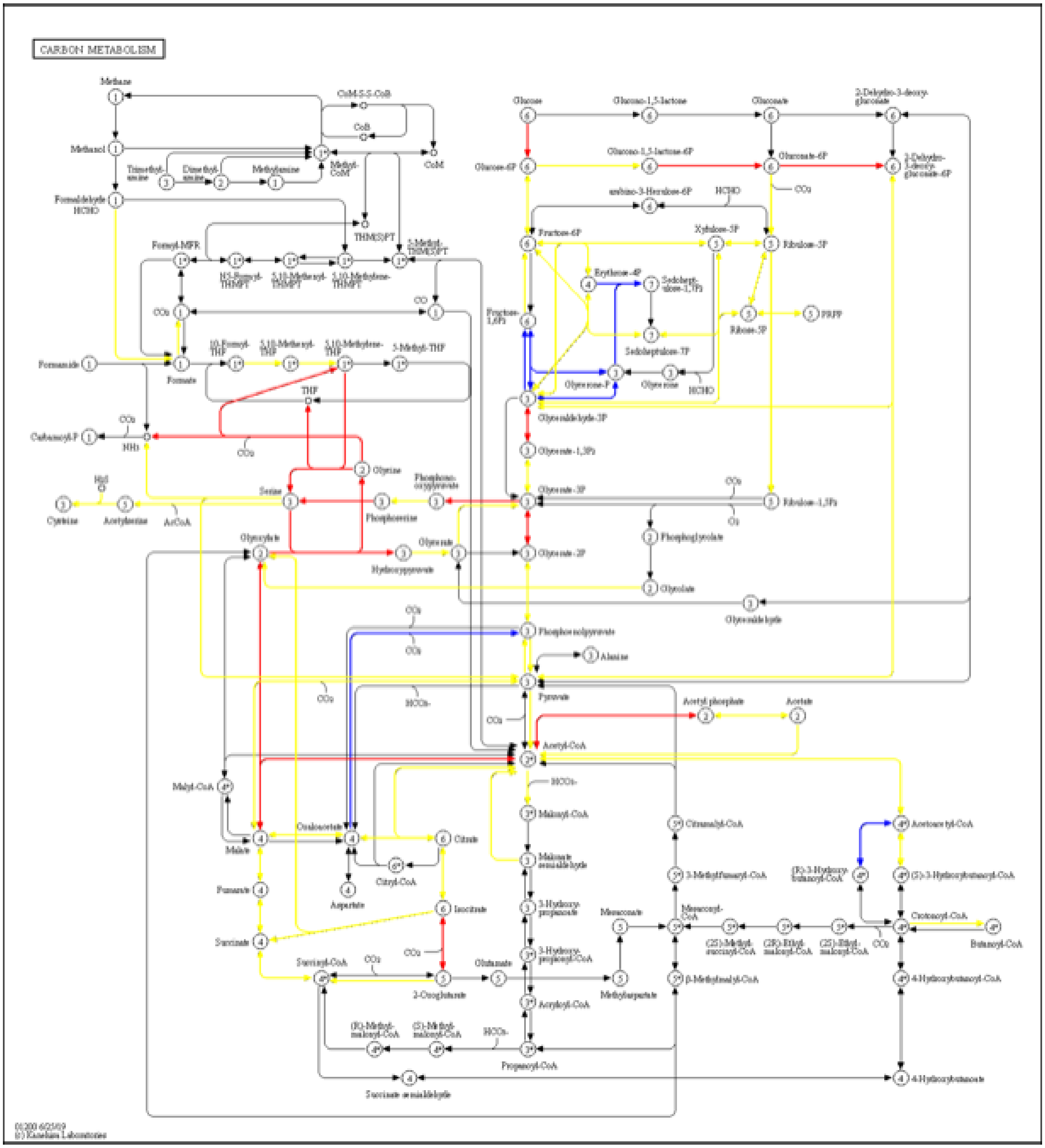
A.8: Carbon metabolism in ASW+nuts at -5°C. KEGG pathway map of carbon metabolism with proteins detected in ASW^{+nuts} at -5°C as compared to control cells (red: LFC ≥ 0.5, blue: LFC ≤ -0.5, yellow: LFC between 0.5 and -0.5; all: Z-score ≥ 2).



A.9: Carbon metabolism in ASW-nuts at -5°C. KEGG pathway map of carbon metabolism with proteins detected in ASW$^{\text{-nuts}}$ at -5°C as compared to control cells (red: LFC ≥ 0.5, blue: LFC ≤ -0.5, yellow: LFC between 0.5 and -0.5; all: Z-score ≥ 2).



A.10: Carbon metabolism in brine+nuts at -5°C. KEGG pathway map of carbon metabolism with proteins detected in brine^{+nuts} at -5°C as compared to control cells (red: LFC ≥ 0.5, blue: LFC ≤ -0.5, yellow: LFC between 0.5 and -0.5; all: Z-score ≥ 2).



A.11: Carbon metabolism in brine-nuts at -5°C. KEGG pathway map of carbon metabolism with proteins detected in brine-nuts at -5°C as compared to control cells (red: LFC ≥ 0.5, blue: LFC ≤ -0.5, yellow: LFC between 0.5 and -0.5; all: Z-score ≥ 2).