Insights from Automated and Untargeted Marine Microbial Metabolomics

William Kumler

2025-06-03

# Acknowledgements

* Lab
  + Anitra
  + Laura
  + Josh
  + Susan
  + Angie/Katherine/Frank
  + Regina
  + Natalie/Iris/Claudia/Raisha/Bryn
  + Leland/Lindsey/Everetta/Raafay/Natalie/Anna/Amy/Alec
* Broader Ocean/UW
  + Committee
  + Armbrust lab
  + Cohort
  + Bowling team
  + OCEAN front office
  + OSB cleaning staff
  + eScience (Dave and Bryna esp.)
  + Fisheries friends (SEAS+Helena/Zoe)
* Research things
  + Funding (ARCS/SCOPE)
  + SCOPE folks
    - Dave K/Sonya/Ed/Angel/Matt/Ben/Daniel
    - SCOPE Ops
    - Captain and crew of the KM
    - PERISCOPE team
  + R packages xcms/data.table/tidyverse/knitr/plotly/shiny/vegan/broom/arrow/duckdb/xml2/XML/Rcolorbrewer/viridis/ggh4x/ggtext/patchwork/pbapply/future/furrr
    - RStudio/Posit, R Markdown
  + Sisweb/Chemcalc/Metlin/HMDB/GNPS/Metabolights/Metabolomics Workbench
* Friends and family
  + Not all my time was spent on boats, in the lab, or the office
    - Stone Gardens/Edgeworks/Seattle Public Libraries
    - Land acknowledgement
  + Past mentors
    - Bethanie, Jim, Mimi, Pete
  + Kaylyn/Hannah/Shawnee/Wave/Kate/Sarra/Rachel Liu/Rachel Lai/Tasha/Sylvia/Colby/Allison/Este/Cassie?/Carly?/Tori?
  + Mom/Dad/Ben/Maggie/Gmas
* And you, dear reader?

# Chapter 1: Introduction

## Marine microbes

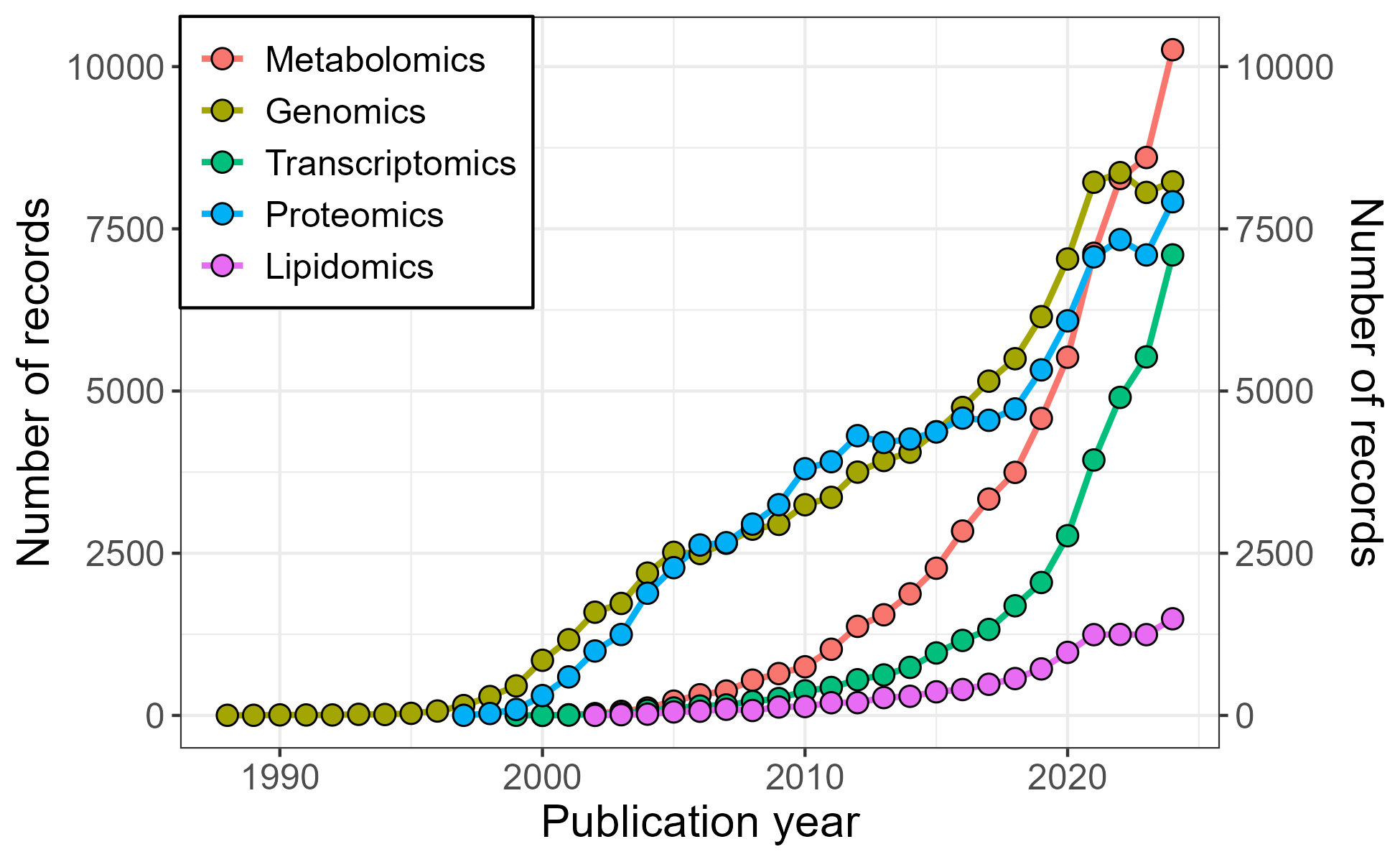
Marine carbon fixation happens at an incredible rate. In the blink of an eye (~100 milliseconds), the ocean converts a blue whale’s mass of atmospheric carbon into biomass and has performed this continuously for at least the last two billion years (Falkowski 1994; Ligrone 2019). Most of this is performed by single-celled organisms too small to see with the naked eye known as microbes (Falkowski 1994; Falkowski, Fenchel, and Delong 2008). The process by which they transform air and nutrients into food is the base of the marine food web and regulates Earth’s climate, with many fates available to the fixed carbon. A large fraction of this particulate matter will be transformed back into CO2 via respiration within the surface ocean, either by the phytoplankton themselves or the rest of the food chain. A smaller fraction makes it out of the euphotic zone via the biological pump and is sequestered for hundreds to thousands of years, while an even smaller fraction survives to the seafloor and can be sequestered for millenia in marine sediments (Iversen 2023; Siegel et al. 2023).

The pathway a particular atom of carbon travels is determined by the structure of the molecule it composes and the environment in which it’s found. Highly labile compounds such as sugars and amino acids can be converted almost instantaneously back into CO2, while ultra-refractory compounds can persist for thousands of years (Moran, Ferrer-González, et al. 2022). Our understanding of the marine environment’s biogeochemistry and community composition has vastly expanded in the past few decades thanks to the establishment of long-term ecological time series and advances in genetic tools, while our characterization of organic carbon lags far behind (Moran, Kujawinski, et al. 2022; Longnecker, Kido Soule, and Kujawinski 2024). Determining the molecular composition of marine carbon and its fluxes through the environment is therefore paramount in improving our ability to accurately model the microbial marine ecosystem (Jones et al. 2024).

## Metabolites and metabolomics

Metabolites are defined simply as the products of cellular metabolism, but this uncomplicated definition belies the dizzying complexity of microbial processes. While technically all biologically produced molecules could fall within this category, the conventional usage refers to the small (<1000 Dalton) organic molecules that act as currencies within the cell while excluding macromolecules such as proteins and lipids. Metabolites are often the reaction intermediates and building blocks of larger molecules but have several important roles of their own, including nutrient and energy storage (Van Mooy et al. 2009; Becker et al. 2018; Mojzeš et al. 2020), antioxidation (Narainsamy et al. 2016), osmotic balance (Yancey et al. 1982; Yancey 2005), buoyancy (Yancey 2005; Boyd and Gradmann 2002), and cell signaling (both beneficial and antagonistic interactions) Thukral, Allen, and Petras (2023). There are likely hundreds of thousands of individual molecules composing the metabolome in the environment, making their comprehensive analysis challenging (Schrimpe-Rutledge et al. 2016).

Nonetheless, metabolomics attempts to do so. The study of “all” small molecules in the cell is a rapidly growing field with over 10,000 publications in 2024 and recently eclipsed all other “omics” fields of study according to a topic search in Web of Science (Figure 1.1, Patti, Yanes, and Siuzdak (2012), Edwards (2023)). These publications span a massive swath of disciplines, with contributions from medicine, polymer chemistry, astronomy, and oceanography. This interdisciplinary nature has resulted in the construction of expansive databases linking organisms’ genetic potential to their realized state (Bauermeister et al. 2022; Kanehisa 2000; Karp et al. 2019).



*Figure 1.1: Number of publications indexed by Web of Science yearly since 1988 across different ’omics disciplines. Data were generated by searching the term in the legend as a topic and tabulated as a bar chart using the WOS Analyze Results option for Publication Years. All data rows were exported to CSV and plotted here using R’s ggplot2 library.*

Quantifying all small molecules in the cell is challenging for many reasons. First, metabolites span a wide range of chemical properties that cannot all be extracted simultaneously or separated on the same type of chromatography (Kido Soule et al. 2015; Cajka and Fiehn 2016; Gika et al. 2019). Second, their wide range of roles in the cell mean that annotating signals is more difficult than proteomics or lipidomics because their building blocks are not shared (Schrimpe-Rutledge et al. 2016). Third, the diversity and novelty of many compounds makes pure standards often unavailable, let alone isotopically-labeled versions necessary for the construction of the gold-standard multipoint internal calibration curve (Patti, Yanes, and Siuzdak 2012; Cajka and Fiehn 2016).

The problems listed above are exacerbated in marine microbial metabolomics. Primarily this is due to their incredibly low concentrations in both the particulate and dissolved phases, with typical values in the picomolar to nanomolar range (Heal et al. 2021; Sacks et al. 2022; Moran, Kujawinski, et al. 2022; Longnecker, Kido Soule, and Kujawinski 2024). An additional problem is the way the salty matrix of seawater behaves similarly to many metabolites during chemical analysis but numerically dominates their abundance by 105 to 1010 molecules per liter. (Boysen et al. 2018; Longnecker, Kido Soule, and Kujawinski 2024). In contrast to other metabolomics specialties where the organism of interest is well studied and genetically documented, environmental metabolomics struggles with a lack of genetic representation and less than 5% of the genetic diversity in the ocean has been captured by reference genomes (DeLong 2005; Salazar and Sunagawa 2017). Certainly fewer than 5% of the organisms in the ocean have been cultured in the lab and their metabolites documented, though work to improve this is underway (Heal et al. 2021; Durham et al. 2022; Kujawinski et al. 2023). Finally, the general inaccessibility of the open ocean results in chronic undersampling and significantly reduced sample sizes relative to land-based metabolomics, resulting in low-power analyses that are only able to detect the largest signals (Karl and Church 2017).

Despite these challenges, marine microbial metabolomics shows significant promise for characterizing the composition of seawater and the organisms that live within it. Metabolites have been used to describe the latitudinal variation in marine particles (Heal et al. 2021; Johnson et al. 2023, 2020), the response of the microbial community to nutrient and vitamin availability (Sañudo-Wilhelmy et al. 2014; Heal et al. 2017; Bertrand et al. 2015; Wilson et al. 2019; Dawson et al. 2020), and the response of phytoplankton to changes in temperature and salinity (Dawson et al. 2023) as well as their response over the diel cycle (Muratore et al. 2022; Boysen et al. 2021). Additionally, recent work on metabolites dissolved in seawater has begun to unlock the vast diversity of organic carbon and nitrogen in the ocean (Sacks et al. 2022; Widner et al. 2021; Johnson, Kido Soule, and Kujawinski 2017). All of these efforts have implications for the way the smallest molecules in the ocean affect its ability to cycle energy and matter through the globe.

## Automated and untargeted liquid-chromatography mass spectrometry

Mass spectrometry (MS) is the dominant analytical platform in metabolomics (Cajka and Fiehn 2016; Gika et al. 2019). Commonly, this technique is paired with chromatographic separation to allow isomers to be quantified independently and to provide additional information about the chemicals’ nature. The disadvantage of this pairing is that the signal must then be integrated in retention time to provide an accurate reconstruction of the original quantity. With noisy signals such as those produced by hydrophilic interaction columns (HILIC, Buszewski and Noga (2012)) and compounds near the limit of detection, this becomes a challenge. The conventional solution is manual integration, in which a mass-spectrometrist manually reviews the extracted chromatograms and determines the start and end of chromatographic peak for integration, often via graphical user interface (GUI). However, this method is time consuming (scaling with the number of compounds and the number of samples) and cannot be guaranteed to be reproducible. This has led to the use of software for automatic peak detection and integration.

Automatic peakpicking and annotation software has been developed in parallel for the better part of two decades by both open-source and commercial endeavours (Smith et al. 2006; Tautenhahn, Böttcher, and Neumann 2008; Heuckeroth et al. 2024; Schmid et al. 2023; Tsugawa et al. 2015; Rafiei and Sleno 2015; Coble and Fraga 2014; Hohrenk et al. 2020). The focus of these tools is typically on untargeted metabolomics (including proteomics and lipidomics), which uses a data-driven approach to compound detection rather than approaching the dataset with a list of anticipated compounds (Gika et al. 2019). This approach is particularly useful for marine microbial metabolomics, where many compounds are yet to be discovered and the additional features detected produce more powerful statistics to compensate for small sample sizes. The untargeted method also comes with significant drawbacks, with imperfect integrations by the peakpickers, multiple signals due to adducts and isotopes, and low-confidence annotations still requiring extensive manual review (Myers et al. 2017a). While untargeted analysis is traditionally associated with hypothesis generation because of its compound discovery capability (Giera, Yanes, and Siuzdak 2022; Thukral, Allen, and Petras 2023), it is perfectly qualified for testing of a well-formed hypothesis as well.

I highlight here the distinction between untargeted MS and automated MS because they are orthogonal philosophies often conflated. It is entirely possible (and often desired!) to have an automatic targeted workflow where specific compounds of interest are quantified with the speed and reproducibility of an algorithm without expanding the analysis to unknowns. Similarly, it is possible to perform untargeted metabolomics with traditionally targeted tools such as Proteowizard’s Skyline (Adams et al. 2020) or even Microsoft Excel as long as the data is used to drive discovery instead of a priori knowledge about the expected compounds. For example, one could imagine an Automated Data Analysis Pipeline (ADAP) type algorithm (Myers et al. 2017b) that recursively extracts the largest intensities in a file and nearby *m/z* values for manual integration. These “alternate” MS methods (targeted automation and manual untargeted) are underutilized simply because the tools for their use have not yet been implemented or documented sufficiently.

## Overview of projects

This thesis presents a body of work spanning data science and oceanography. In the first chapter, I discuss how MS data can be enormously simplified by converting it into a “tidy” format in the sense of Wickham (2014). This allows for the rapid exploration and reproducible analysis that I use in the rest of the thesis. Chapter 3 logically extends this framework into proper database systems which mitigates Chapter 2’s major problems with memory usage. I additionally compare multiple database systems with emphasis on modern column-oriented and online analytical processing methods that show particular promise. The particular strength of these methods is their ability to look at data *across* files rather than within a single one, something that I heavily leverage in later chapters.

Chapter 4 demonstrates the utility of allowing for rapid raw data access by showing how novel peak metrics calculated from the raw data can significantly reduce the rate of false positives in existing peakpicking software. This “cleaned” data set then shows interesting differences between marine microbial samples taken from different depths that were not apparent in the original. In the appendix, I also illustrate how raw data itself can be treated as a multidimensional array with the largest “signals” being those of high-quality peaks, allowing dimensionality reduction techniques to group MS features for rapid quality annotation.

Chapters 5 and 6 are applications of the above philosophy to oceanographic data collected from the North Pacific Subtropical Gyre (NPSG) near Station ALOHA. The NPSG is the largest biome on the planet and, like most of the surface ocean, is limited by the bioavailability of nitrogen despite large standing stocks of dinitrogen gas and DON in addition to the constant upwelling of nitrate from the deep (Moore et al. 2013; Karl and Church 2017). Since nitrogen limits the amount of carbon fixation and export possible, understanding the forms and fluxes of nitrogen-containing molecules and the organisms they compose directly affects our ability to predict marine carbon cycling. As the majority of the nitrogen flux is through small, polar molecules (Moran et al. 2016; Moran, Kujawinski, et al. 2022), metabolomics is particularly well suited to describing and quantifying these elemental cycles.

Chapter 5 documents an exploratory metabolomics dataset collected in the NPSG across two sets of mesoscale eddy features of opposing polarity. In many ways this chapter felt like a return to the Challenger era of observational oceanography which required the use of complex ecological statistics to unravel the impacts of sea level anomaly on the ocean’s metabolome and reported several compounds for the first time in the open ocean. Chapter 6, in contrast, was a deeply-nested experimental framework using short-term incubations with isotopically-labeled nitrogen substrates to test specific hypotheses about microbial nutrient acquisition and use. In both cases, the rapid and intuitive exploration of select chromatograms as well as access to the raw data was key for constructing a confident and coherent narrative of the microbial role in ocean biogeochemistry.

# Chapter 2: Tidy Data Neatly Resolves Mass-Spectrometry’s Ragged Arrays

# Chapter 3: Databases Are an Effective and Efficient Method for Storage and Access of Mass-Spectrometry Data

# Chapter 4: Picky with Peakpicking: Assessing Chromatographic Peak Quality with Simple Metrics in Metabolomics

# Chapter 5: Metabolites Reflect Variability Introduced by Mesoscale Eddies in the North Pacific Subtropical Gyre

# Chapter 6: The Form of Nitrogen Determines its Fate in the North Pacific Subtropical Gyre

# Chapter 7: Conclusions

Note that I’m unhappy I didn’t get to do more MS/MS stuff, partially due to the tools not being very good - diagnostic fragments, wildly varying ways/formats to query, Metlin going private, unclear how to create consensus spectra from multiple scans or match these to knowns, limited database availability, lack of MS/MS stuff in my own data - would love to do this next.

# Bibliography

Adams, Kendra J., Brian Pratt, Neelanjan Bose, Laura G. Dubois, Lisa St. John-Williams, Kevin M. Perrott, Karina Ky, et al. 2020. “Skyline for Small Molecules: A Unifying Software Package for Quantitative Metabolomics.” *Journal of Proteome Research* 19 (4): 1447–58. <https://doi.org/10.1021/acs.jproteome.9b00640>.

Bauermeister, Anelize, Helena Mannochio-Russo, Letícia V. Costa-Lotufo, Alan K. Jarmusch, and Pieter C. Dorrestein. 2022. “Mass Spectrometry-Based Metabolomics in Microbiome Investigations.” *Nature Reviews Microbiology* 20 (3): 143–60. <https://doi.org/10.1038/s41579-021-00621-9>.

Becker, Kevin W., James R. Collins, Bryndan P. Durham, Ryan D. Groussman, Angelicque E. White, Helen F. Fredricks, Justin E. Ossolinski, et al. 2018. “Daily Changes in Phytoplankton Lipidomes Reveal Mechanisms of Energy Storage in the Open Ocean.” *Nature Communications* 9 (1): 5179. <https://doi.org/10.1038/s41467-018-07346-z>.

Bertrand, Erin M., John P. McCrow, Ahmed Moustafa, Hong Zheng, Jeffrey B. McQuaid, Tom O. Delmont, Anton F. Post, et al. 2015. “Phytoplankton–Bacterial Interactions Mediate Micronutrient Colimitation at the Coastal Antarctic Sea Ice Edge.” *Proceedings of the National Academy of Sciences* 112 (32): 9938–43. <https://doi.org/10.1073/pnas.1501615112>.

Boyd, C., and D. Gradmann. 2002. “Impact of Osmolytes on Buoyancy of Marine Phytoplankton.” *Marine Biology* 141 (4): 605–18. <https://doi.org/10.1007/s00227-002-0872-z>.

Boysen, Angela K., Laura T. Carlson, Bryndan P. Durham, Ryan D. Groussman, Frank O. Aylward, François Ribalet, Katherine R. Heal, et al. 2021. “Particulate Metabolites and Transcripts Reflect Diel Oscillations of Microbial Activity in the Surface Ocean.” Edited by Jeff Bowman. *mSystems* 6 (3): e00896–20. <https://doi.org/10.1128/mSystems.00896-20>.

Boysen, Angela K., Katherine R. Heal, Laura T. Carlson, and Anitra E. Ingalls. 2018. “Best-Matched Internal Standard Normalization in Liquid Chromatography–Mass Spectrometry Metabolomics Applied to Environmental Samples.” *Analytical Chemistry* 90 (2): 1363–69. <https://doi.org/10.1021/acs.analchem.7b04400>.

Buszewski, Bogusław, and Sylwia Noga. 2012. “Hydrophilic Interaction Liquid Chromatography (HILIC)—a Powerful Separation Technique.” *Analytical and Bioanalytical Chemistry* 402 (1): 231–47. <https://doi.org/10.1007/s00216-011-5308-5>.

Cajka, Tomas, and Oliver Fiehn. 2016. “Toward Merging Untargeted and Targeted Methods in Mass Spectrometry-Based Metabolomics and Lipidomics.” *Analytical Chemistry* 88 (1): 524–45. <https://doi.org/10.1021/acs.analchem.5b04491>.

Coble, Jamie B., and Carlos G. Fraga. 2014. “Comparative Evaluation of Preprocessing Freeware on Chromatography/Mass Spectrometry Data for Signature Discovery.” *Journal of Chromatography A* 1358 (September): 155–64. <https://doi.org/10.1016/j.chroma.2014.06.100>.

Dawson, H M, E Connors, N G Erazo, J S Sacks, V Mierzejewski, S M Rundell, L T Carlson, et al. 2023. “Microbial Metabolomic Responses to Changes in Temperature and Salinity Along the Western Antarctic Peninsula.” *The ISME Journal* 17 (11): 2035–46. <https://doi.org/10.1038/s41396-023-01475-0>.

Dawson, H M, K R Heal, A Torstensson, L T Carlson, A E Ingalls, and J N Young. 2020. “Large Diversity in Nitrogen- and Sulfur-Containing Compatible Solute Profiles in Polar and Temperate Diatoms.” *Integrative and Comparative Biology* 60 (6): 1401–13. <https://doi.org/10.1093/icb/icaa133>.

DeLong, Edward F. 2005. “Microbial Community Genomics in the Ocean.” *Nature Reviews Microbiology* 3 (6): 459–69. <https://doi.org/10.1038/nrmicro1158>.

Durham, Bryndan P., Angela K. Boysen, Katherine R. Heal, Laura T. Carlson, Rachel Boccamazzo, Chloe R. Deodato, Wei Qin, Rose Ann Cattolico, E. Virginia Armbrust, and Anitra E. Ingalls. 2022. “Chemotaxonomic Patterns in Intracellular Metabolites of Marine Microbial Plankton.” *Frontiers in Marine Science* 9 (September): 864796. <https://doi.org/10.3389/fmars.2022.864796>.

Edwards, Bethanie R. 2023. “Lipid Biogeochemistry and Modern Lipidomic Techniques.” *Annual Review of Marine Science* 15 (1): 485–508. <https://doi.org/10.1146/annurev-marine-040422-094104>.

Falkowski, Paul G. 1994. “The Role of Phytoplankton Photosynthesis in Global Biogeochemical Cycles.” *Photosynthesis Research* 39 (3): 235–58. <https://doi.org/10.1007/BF00014586>.

Falkowski, Paul G., Tom Fenchel, and Edward F. Delong. 2008. “The Microbial Engines That Drive Earth’s Biogeochemical Cycles.” *Science* 320 (5879): 1034–39. <https://doi.org/10.1126/science.1153213>.

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>.

Giera, Martin, Oscar Yanes, and Gary Siuzdak. 2022. “Metabolite Discovery: Biochemistry’s Scientific Driver.” *Cell Metabolism* 34 (1): 21–34. <https://doi.org/10.1016/j.cmet.2021.11.005>.

Gika, Helen, Christina Virgiliou, Georgios Theodoridis, Robert S. Plumb, and Ian D. Wilson. 2019. “Untargeted LC/MS-based Metabolic Phenotyping (Metabonomics/Metabolomics): The State of the Art.” *Journal of Chromatography B* 1117 (June): 136–47. <https://doi.org/10.1016/j.jchromb.2019.04.009>.

Heal, Katherine R., Bryndan P. Durham, Angela K. Boysen, Laura T. Carlson, Wei Qin, François Ribalet, Angelicque E. White, Randelle M. Bundy, E. Virginia Armbrust, and Anitra E. Ingalls. 2021. “Marine Community Metabolomes Carry Fingerprints of Phytoplankton Community Composition.” Edited by Manuel Liebeke. *mSystems* 6 (3): e01334–20. <https://doi.org/10.1128/mSystems.01334-20>.

Heal, Katherine R., Wei Qin, Francois Ribalet, Anthony D. Bertagnolli, Willow Coyote-Maestas, Laura R. Hmelo, James W. Moffett, et al. 2017. “Two Distinct Pools of B12 Analogs Reveal Community Interdependencies in the Ocean.” *Proceedings of the National Academy of Sciences* 114 (2): 364–69. <https://doi.org/10.1073/pnas.1608462114>.

Heuckeroth, Steffen, Tito Damiani, Aleksandr Smirnov, Olena Mokshyna, Corinna Brungs, Ansgar Korf, Joshua David Smith, et al. 2024. “Reproducible Mass Spectrometry Data Processing and Compound Annotation in MZmine 3.” *Nature Protocols* 19 (9): 2597–2641. <https://doi.org/10.1038/s41596-024-00996-y>.

Hohrenk, Lotta L., Fabian Itzel, Nicolai Baetz, Jochen Tuerk, Maryam Vosough, and Torsten C. Schmidt. 2020. “Comparison of Software Tools for Liquid Chromatography–High-Resolution Mass Spectrometry Data Processing in Nontarget Screening of Environmental Samples.” *Analytical Chemistry* 92 (2): 1898–1907. <https://doi.org/10.1021/acs.analchem.9b04095>.

Iversen, Morten H. 2023. “Carbon Export in the Ocean: A Biologist’s Perspective.” *Annual Review of Marine Science* 15 (1): 357–81. <https://doi.org/10.1146/annurev-marine-032122-035153>.

Johnson, Winifred M., Melissa C. Kido Soule, and Elizabeth B. Kujawinski. 2017. “Extraction Efficiency and Quantification of Dissolved Metabolites in Targeted Marine Metabolomics.” *Limnology and Oceanography: Methods* 15 (4): 417–28. <https://doi.org/10.1002/lom3.10181>.

Johnson, Winifred M., Melissa C. Kido Soule, Krista Longnecker, Maya P. Bhatia, Steven J. Hallam, Michael W. Lomas, and Elizabeth B. Kujawinski. 2023. “Particulate and Dissolved Metabolite Distributions Along a Latitudinal Transect of the Western Atlantic Ocean.” *Limnology and Oceanography* 68 (2): 377–93. <https://doi.org/10.1002/lno.12275>.

Johnson, Winifred M., Krista Longnecker, Melissa C. Kido Soule, William A. Arnold, Maya P. Bhatia, Steven J. Hallam, Benjamin A. S. Van Mooy, and Elizabeth B. Kujawinski. 2020. “Metabolite Composition of Sinking Particles Differs from Surface Suspended Particles Across a Latitudinal Transect in the South Atlantic.” *Limnology and Oceanography* 65 (1): 111–27. <https://doi.org/10.1002/lno.11255>.

Jones, Catriona L. C., Judith Camps-Castella, Mike Smykala, Morgan S. Sobol, and Keisuke Inomura. 2024. “Seeing Through the Gray Box: An Integrated Approach to Physiological Modeling of Phytoplankton Stoichiometry.” *Frontiers in Ecology and Evolution* 12 (December): 1505025. <https://doi.org/10.3389/fevo.2024.1505025>.

Kanehisa, M. 2000. “KEGG: Kyoto Encyclopedia of Genes and Genomes.” *Nucleic Acids Research* 28 (1): 27–30. <https://doi.org/10.1093/nar/28.1.27>.

Karl, David M., and Matthew J. Church. 2017. “Ecosystem Structure and Dynamics in the North Pacific Subtropical Gyre: New Views of an Old Ocean.” *Ecosystems* 20 (3): 433–57. <https://doi.org/10.1007/s10021-017-0117-0>.

Karp, Peter D, Richard Billington, Ron Caspi, Carol A Fulcher, Mario Latendresse, Anamika Kothari, Ingrid M Keseler, et al. 2019. “The BioCyc Collection of Microbial Genomes and Metabolic Pathways.” *Briefings in Bioinformatics* 20 (4): 1085–93. <https://doi.org/10.1093/bib/bbx085>.

Kido Soule, Melissa C., Krista Longnecker, Winifred M. Johnson, and Elizabeth B. Kujawinski. 2015. “Environmental Metabolomics: Analytical Strategies.” *Marine Chemistry* 177 (December): 374–87. <https://doi.org/10.1016/j.marchem.2015.06.029>.

Kujawinski, Elizabeth B., Rogier Braakman, Krista Longnecker, Jamie W. Becker, Sallie W. Chisholm, Keven Dooley, Melissa C. Kido Soule, Gretchen J. Swarr, and Kathryn Halloran. 2023. “Metabolite Diversity Among Representatives of Divergent *Prochlorococcus* Ecotypes.” Edited by Jack A. Gilbert. *mSystems* 8 (5): e01261–22. <https://doi.org/10.1128/msystems.01261-22>.

Ligrone, Roberto. 2019. “The Great Oxygenation Event.” In *Biological Innovations That Built the World*, 129–54. Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-030-16057-9_5>.

Longnecker, Krista, Melissa C. Kido Soule, and Elizabeth B. Kujawinski. 2024. “A Witches’ Brew: Dissolved Metabolites in Seawater.” In *Biogeochemistry of Marine Dissolved Organic Matter*, 69–99. Elsevier. <https://doi.org/10.1016/B978-0-443-13858-4.00007-1>.

Mojzeš, Peter, Lu Gao, Tatiana Ismagulova, Jana Pilátová, Šárka Moudříková, Olga Gorelova, Alexei Solovchenko, Ladislav Nedbal, and Anya Salih. 2020. “Guanine, a High-Capacity and Rapid-Turnover Nitrogen Reserve in Microalgal Cells.” *Proceedings of the National Academy of Sciences* 117 (51): 32722–30. <https://doi.org/10.1073/pnas.2005460117>.

Moore, C. M., M. M. Mills, K. R. Arrigo, I. Berman-Frank, L. Bopp, P. W. Boyd, E. D. Galbraith, et al. 2013. “Processes and Patterns of Oceanic Nutrient Limitation.” *Nature Geoscience* 6 (9): 701–10. <https://doi.org/10.1038/ngeo1765>.

Moran, Mary Ann, Frank X. Ferrer-González, He Fu, Brent Nowinski, Malin Olofsson, McKenzie A. Powers, Jeremy E. Schreier, William F. Schroer, Christa B. Smith, and Mario Uchimiya. 2022. “The Ocean’s Labile DOC Supply Chain.” *Limnology and Oceanography* 67 (5): 1007–21. <https://doi.org/10.1002/lno.12053>.

Moran, Mary Ann, Elizabeth B. Kujawinski, William F. Schroer, Shady A. Amin, Nicholas R. Bates, Erin M. Bertrand, Rogier Braakman, et al. 2022. “Microbial Metabolites in the Marine Carbon Cycle.” *Nature Microbiology* 7 (4): 508–23. <https://doi.org/10.1038/s41564-022-01090-3>.

Moran, Mary Ann, Elizabeth B. Kujawinski, Aron Stubbins, Rob Fatland, Lihini I. Aluwihare, Alison Buchan, Byron C. Crump, et al. 2016. “Deciphering Ocean Carbon in a Changing World.” *Proceedings of the National Academy of Sciences* 113 (12): 3143–51. <https://doi.org/10.1073/pnas.1514645113>.

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>.

Myers, Owen D., Susan J. Sumner, Shuzhao Li, Stephen Barnes, and Xiuxia Du. 2017a. “Detailed Investigation and Comparison of the XCMS and MZmine 2 Chromatogram Construction and Chromatographic Peak Detection Methods for Preprocessing Mass Spectrometry Metabolomics Data.” *Analytical Chemistry* 89 (17): 8689–95. <https://doi.org/10.1021/acs.analchem.7b01069>.

———. 2017b. “One Step Forward for Reducing False Positive and False Negative Compound Identifications from Mass Spectrometry Metabolomics Data: New Algorithms for Constructing Extracted Ion Chromatograms and Detecting Chromatographic Peaks.” *Analytical Chemistry* 89 (17): 8696–8703. <https://doi.org/10.1021/acs.analchem.7b00947>.

Narainsamy, Kinsley, Sandrine Farci, Emilie Braun, Christophe Junot, Corinne Cassier-Chauvat, and Franck Chauvat. 2016. “Oxidative-Stress Detoxification and Signalling in Cyanobacteria: The Crucial Glutathione Synthesis Pathway Supports the Production of Ergothioneine and Ophthalmate.” *Molecular Microbiology* 100 (1): 15–24. <https://doi.org/10.1111/mmi.13296>.

Patti, Gary J., Oscar Yanes, and Gary Siuzdak. 2012. “Metabolomics: The Apogee of the Omics Trilogy.” *Nature Reviews Molecular Cell Biology* 13 (4): 263–69. <https://doi.org/10.1038/nrm3314>.

Rafiei, Atefeh, and Lekha Sleno. 2015. “Comparison of Peak-Picking Workflows for Untargeted Liquid Chromatography/High-Resolution Mass Spectrometry Metabolomics Data Analysis.” *Rapid Communications in Mass Spectrometry* 29 (1): 119–27. <https://doi.org/10.1002/rcm.7094>.

Sacks, Joshua S., Katherine R. Heal, Angela K. Boysen, Laura T. Carlson, and Anitra E. Ingalls. 2022. “Quantification of Dissolved Metabolites in Environmental Samples Through Cation-Exchange Solid-Phase Extraction Paired with Liquid Chromatography–Mass Spectrometry.” *Limnology and Oceanography: Methods* 20 (11): 683–700. <https://doi.org/10.1002/lom3.10513>.

Salazar, Guillem, and Shinichi Sunagawa. 2017. “Marine Microbial Diversity.” *Current Biology* 27 (11): R489–94. <https://doi.org/10.1016/j.cub.2017.01.017>.

Sañudo-Wilhelmy, Sergio A., Laura Gómez-Consarnau, Christopher Suffridge, and Eric A. Webb. 2014. “The Role of B Vitamins in Marine Biogeochemistry.” *Annual Review of Marine Science* 6 (1): 339–67. <https://doi.org/10.1146/annurev-marine-120710-100912>.

Schmid, Robin, Steffen Heuckeroth, Ansgar Korf, Aleksandr Smirnov, Owen Myers, Thomas S. Dyrlund, Roman Bushuiev, et al. 2023. “Integrative Analysis of Multimodal Mass Spectrometry Data in MZmine 3.” *Nature Biotechnology* 41 (4): 447–49. <https://doi.org/10.1038/s41587-023-01690-2>.

Schrimpe-Rutledge, Alexandra C., Simona G. Codreanu, Stacy D. Sherrod, and John A. McLean. 2016. “Untargeted Metabolomics Strategies—Challenges and Emerging Directions.” *Journal of the American Society for Mass Spectrometry* 27 (12): 1897–1905. <https://doi.org/10.1007/s13361-016-1469-y>.

Siegel, David A., Timothy DeVries, Ivona Cetinić, and Kelsey M. Bisson. 2023. “Quantifying the Ocean’s Biological Pump and Its Carbon Cycle Impacts on Global Scales.” *Annual Review of Marine Science* 15 (1): 329–56. <https://doi.org/10.1146/annurev-marine-040722-115226>.

Smith, Colin A., Elizabeth J. Want, Grace O’Maille, Ruben Abagyan, and Gary Siuzdak. 2006. “XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification.” *Analytical Chemistry* 78 (3): 779–87. <https://doi.org/10.1021/ac051437y>.

Tautenhahn, Ralf, Christoph Böttcher, and Steffen Neumann. 2008. “Highly Sensitive Feature Detection for High Resolution LC/MS.” *BMC Bioinformatics* 9 (1): 504. <https://doi.org/10.1186/1471-2105-9-504>.

Thukral, Monica, Andrew E. Allen, and Daniel Petras. 2023. “Progress and Challenges in Exploring Aquatic Microbial Communities Using Non-Targeted Metabolomics.” *The ISME Journal* 17 (12): 2147–59. <https://doi.org/10.1038/s41396-023-01532-8>.

Tsugawa, Hiroshi, Tomas Cajka, Tobias Kind, Yan Ma, Brendan Higgins, Kazutaka Ikeda, Mitsuhiro Kanazawa, Jean VanderGheynst, Oliver Fiehn, and Masanori Arita. 2015. “MS-DIAL: Data-Independent MS/MS Deconvolution for Comprehensive Metabolome Analysis.” *Nature Methods* 12 (6): 523–26. <https://doi.org/10.1038/nmeth.3393>.

Van Mooy, Benjamin A. S., Helen F. Fredricks, Byron E. Pedler, Sonya T. Dyhrman, David M. Karl, Michal Koblížek, Michael W. Lomas, et al. 2009. “Phytoplankton in the Ocean Use Non-Phosphorus Lipids in Response to Phosphorus Scarcity.” *Nature* 458 (7234): 69–72. <https://doi.org/10.1038/nature07659>.

Vardi, Assaf, Fabio Formiggini, Raffaella Casotti, Alessandra De Martino, François Ribalet, Antonio Miralto, and Chris Bowler. 2006. “A Stress Surveillance System Based on Calcium and Nitric Oxide in Marine Diatoms.” Edited by Jeffrey Dangl. *PLoS Biology* 4 (3): e60. <https://doi.org/10.1371/journal.pbio.0040060>.

Wickham, Hadley. 2014. “Tidy Data.” *Journal of Statistical Software* 59 (10). <https://doi.org/10.18637/jss.v059.i10>.

Widner, Brittany, Melissa C. Kido Soule, Frank Xavier Ferrer-González, Mary Ann Moran, and Elizabeth B. Kujawinski. 2021. “Quantification of Amine- and Alcohol-Containing Metabolites in Saline Samples Using Pre-extraction Benzoyl Chloride Derivatization and Ultrahigh Performance Liquid Chromatography Tandem Mass Spectrometry (UHPLC MS/MS).” *Analytical Chemistry* 93 (11): 4809–17. <https://doi.org/10.1021/acs.analchem.0c03769>.

Wilson, Samuel T., Nicholas J. Hawco, E. Virginia Armbrust, Benedetto Barone, Karin M. Björkman, Angela K. Boysen, Macarena Burgos, et al. 2019. “Kīlauea Lava Fuels Phytoplankton Bloom in the North Pacific Ocean.” *Science* 365 (6457): 1040–44. <https://doi.org/10.1126/science.aax4767>.

Yancey, Paul H. 2005. “Organic Osmolytes as Compatible, Metabolic and Counteracting Cytoprotectants in High Osmolarity and Other Stresses.” *Journal of Experimental Biology* 208 (15): 2819–30. <https://doi.org/10.1242/jeb.01730>.

Yancey, Paul H., Mary E. Clark, Steven C. Hand, R. David Bowlus, and George N. Somero. 1982. “Living with Water Stress: Evolution of Osmolyte Systems.” *Science* 217 (4566): 1214–22. <https://doi.org/10.1126/science.7112124>.

# Appendix 1: Speedy Quality Assurance via Lasso Labeling for Untargeted Mass-Spectrometry Data