Molecular mechanisms of microbial interactions in model systems and natural environments

Zinka Bartolek

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington
2025

Reading committee:
E. Virginia Armbrust, Chair
Robert Morris
Amy Willis

Program Authorized to Offer Degree:
Oceanography

University of Washington

Abstract

Molecular mechanisms of microbial interactions in model systems and natural environments

Zinka Bartolek

Chair of the Supervisory Committee:
E. Virginia Armbrust
Oceanography

Marine microbes are dominant life forms in the ocean and comprise dynamic populations of viruses, bacteria, archaea and single-celled eukaryotes. Interactions between marine microbes impact global biogeochemical cycles and form the base of marine food webs. Interactions range from symbiotic to parasitic, and are mediated through nutrient exchanges and chemical signals. In this dissertation, I focus on characterizing molecular mechanisms responsible for interactions within two groups of microorganisms, phytoplankton and bacteria. In Chapter 1, I studied how extracellular metabolites of an antagonistic bacterium *Croceibacter atlanticus* impact physiological and transcriptional changes in the model diatom *Thalassiosira pseudonana* through lab co-culture experiments. In response to extracellular bacterial metabolites, *T. pseudonana* showed inhibition of cell division, with transcriptional changes in cell cycle regulation, amino acid production and cell wall stability, suggesting that bacterial metabolites may modulate diatom metabolism in ways that support bacterial growth. In Chapter 2, I developed a new statistical method to detect microbial interactions and their transcriptional signatures from environmental metatranscriptome data collected in the North Pacific. Most consistent transcriptional interactions occurred in cross-kingdom pairs especially between diatoms and Proteobacteria, ranging from common functional mechanisms like cross-feeding and defense to species-specific mechanisms occurring through secondary metabolite exchanges, revealing molecular mechanisms of interactions directly in natural communities. In Chapter 3, I studied metabolic capabilities and interactions of 52 novel heterotrophic bacteria isolates capable of utilizing phytoplankton-derived organic matter from the Equatorial Pacific. I generated and analyzed closed genomes of novel strains, species and genus, and proposed potential metabolic exchanges between bacteria including cross-feeding of vitamins and growth factors that may influence interactions with phytoplankton hosts. In Chapter 4, I designed a course-based undergraduate research experience (CURE) that teaches students laboratory techniques and data analysis skills in genomics through exploration of the metabolism of marine bacteria. Novel Equatorial Pacific isolates and their genomes were used by students in their independent research projects to determine capacity of bacteria to grow on select phytoplankton metabolites and connect the growth phenotype to the genomic capability of the bacteria. This dissertation provides insights into the molecular mechanisms controlling interactions in marine microbial communities, from controlled lab studies to natural environments, and introduces novel methodologies to help advance our understanding of microbial interactions.

Table of Contents

Introduction

Marine microbes are dominant life forms in the ocean accounting for the majority of oceanic biodiversity and biomass. They exist in diverse communities of viruses, bacteria, archaea and phytoplankton and engage in complex interactions with each other. The collective function of these microbial communities arises from intra- and inter-species interactions, resulting in emergent properties that govern global-scale processes. A central framework in understanding these dynamics is the microbial loop, where phytoplankton fix nearly 50% of atmospheric CO_2 into dissolved organic matter (DOM), which is consumed by heterotrophic bacteria and converted to biomass that travels up the food chain, can be exported to the deep ocean or respired and returned to the atmosphere (Azam et al., 2007; Field et al., 1998; Falkowski et al., 2008). By regulating the balance between recycling and export of carbon, nitrogen, and other essential nutrients, microbial interactions play an important role in global biogeochemical cycles. In addition, microorganisms form the base of the marine food web, and interactions between them determine the availability of organic matter to higher trophic levels.

Marine microbes can interact with each other either through direct contact or through exchanges of organic matter and metabolites (Seymour et al., 2017). The essential first step to microbial interaction is encounter, either one cell encountering the other, or encounter between a cell and chemical resources released by a different organism (Słomka et al., 2023). The phycosphere, a diffusive boundary layer surrounding individual phytoplankton cells is an environment where the majority of cell-cell or cell-chemical resource interactions are hypothesized to occur (Amin et al., 2012; Seymour et al., 2017). The microenvironment of the phycosphere is distinct from surrounding seawater since mass transport is limited to diffusion and not affected by turbulence (Bell & Mitchel 1972; Seymour et al., 2017). Chemical conditions including pH differ from surrounding seawater (Liu et al., 2022), nutrient concentrations can exceed background levels by orders of magnitude (Azam et al., 2007), and concentrations of phytoplankton and bacteria released compounds and secondary metabolites are enriched (Meyer et al., 2017).

Microbial interactions in the phycosphere range from mutualism to parasitism and are mediated through exchanges of nutrients and metabolites. The most common bacteria found in the phycosphere of marine phytoplankton belong to Alphaproteobacteria, Gammaproteobacteria and Bacteroidetes, which are copiotrophic bacteria adapted to nutrient rich environments and specialized in degrading complex organic matter (Seymour et al., 2017). These bacteria can sense chemical gradients and chemotax towards nutrient hotpots such as the phycosphere, where they interact with the phytoplankton through direct contact, transient proximity to the cell or through exchanges of nutrients and metabolites (Raina et al., 2022). Mutualistic exchanges between phytoplankton and bacteria include exchanges of bacteria-produced vitamins such as cobalamin (B12) (Zoccarato et al., 2022; Giordano et al., 2024), regulation of growth by growth promoting phytohormones like indole-3-acetic acid (IAA) (Khalil et al., 2024; Durham et al., 2017), and exchanges of scarce nutrients including siderophore-bound iron and amino acids (Amin et al., 2009; Gómez-Consarnau et al., 2018; Croft et al., 2005). On the other hand, algicidal bacteria can disrupt the progression of cell cycle in phytoplankton, inhibit phytoplankton growth and cause lysis through excretion of various compounds ranging from small metabolites to large enzymes (Stock et al., 2020; Guillard 1975). In turn, phytoplankton

can also regulate the composition and activity of their bacterial microbiomes through excretion of several metabolites, resulting in complex interaction dynamics (Stock et al., 2020).

Bacteria engage in complex interactions with each other within the phycosphere, and these bacteria-bacteria interactions can have community level impacts. Phytoplankton associated bacteria occupy distinct niches, where different bacterial taxonomies are specialized in degrading particular types of phytoplankton derived dissolved organic matter (DOM) (Ferrer-González et al., 2021; Mayali et al., 2012). While bacterial use of DOM can be predicted for single bacteria, communities of bacteria have unpredictable patterns of DOM utilization, suggesting emergent properties of bacterial communities resulting from interactions between bacterial taxa (Fuessel et al., 2025). Bacteria interact with each other through exchanges of nutrients, including competition, cross-feeding and syntropy (Seth & Taga, 2014). Similarly, bacteria engage in complex signaling strategies such as quorum sensing and use various defensive strategies through the production of antibiotics and toxins (Coolahan & Whalen, 2025; Rolland et al., 2016). Some of these bacterial communication signals can indirectly impact host phytoplankton physiology, pointing towards a prevalence of higher order interactions (Amin et al., 2015; Durham et al., 2017; Fei et al., 2020; Shibl et al., 2020).

Overall, interactions between marine microorganisms are complex. Interactions are based on attachment or exchange of metabolites and nutrients and can be highly specific to the interacting species or strains. Some interactions involve exchanges of specific metabolites suggesting long term co-evolution (King et al., 2022; Ahern et al., 2021), while others occur through opportunistic uptake of DOM and defense mechanisms (Sarmento & Gasol, 2012; Sarmento et al., 2016). The nature of interactions is dynamic and can change from mutualistic to pathogenic depending on the growth stage of interacting organisms, nutrient availability and other environmental conditions (Daniels et al., 2023; Harvey et al., 2023). Finally, microorganisms interact in complex communities where multiple exchanges might be happening at the same time under changing environmental conditions.

The overarching research question of this dissertation is: What are the molecular mechanisms driving interactions in marine phytoplankton and bacterial communities? I address this question at different scales, starting from controlled lab co-culture experiments of model organisms in Chapter 1, to characterization of new bacterial isolates and predictions of bacterial metabolic exchanges in Chapter 3, and to developing new statistical methods that enable detection of interactions directly from field metatranscriptome data in Chapter 2. In Chapter 4, I develop teaching modules for a Course Undergraduate Research Experience that teaches students genomics approaches to studying microbial interactions. In each chapter of my dissertation, I attempt to identify the underlying mechanisms that control interactions between marine microorganisms. Understanding the molecular basis of these interactions is essential to predicting the future of primary production in the ocean, especially in a rapidly changing climate.

References

Ahern, O. M., Whittaker, K. A., Williams, T. C., Hunt, D. E., & Rynearson, T. A. (2021). Host genotype structures the microbiome of a globally dispersed marine phytoplankton. *Proceedings of the National Academy of Sciences*, *118*(48), e2105207118. https://doi.org/10.1073/pnas.2105207118

Amin, S. A., Green, D. H., Hart, M. C., Küpper, F. C., Sunda, W. G., & Carrano, C. J. (2009). Photolysis of iron–siderophore chelates promotes bacterial–algal mutualism. *Proceedings of the National Academy of Sciences*, *106*(40), 17071–17076. https://doi.org/10.1073/pnas.0905512106

Amin, S. A., Hmelo, L. R., Van Tol, H. M., Durham, B. P., Carlson, L. T., Heal, K. R., Morales, R. L., Berthiaume, C. T., Parker, M. S., Djunaedi, B., Ingalls, A. E., Parsek, M. R., Moran, M. A., & Armbrust, E. V. (2015). Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. *Nature*, *522*(7554), 98–101. https://doi.org/10.1038/nature14488

Amin, S. A., Parker, M. S., & Armbrust, E. V. (2012). Interactions between Diatoms and Bacteria. *Microbiology and Molecular Biology Reviews*, *76*(3), 667–684. https://doi.org/10.1128/MMBR.00007-12

Azam, F., & Malfatti, F. (2007). Microbial structuring of marine ecosystems. *Nature Reviews Microbiology*, *5*(10), 782–791. https://doi.org/10.1038/nrmicro1747

Bell W, Mitchell R. Chemotactic and growth response of marine bacteria to algal extracellular products. Biol. Bull. 1972;143(2):265-77.

Coolahan, M., & Whalen, K. E. (2025). A review of quorum-sensing and its role in mediating interkingdom interactions in the ocean. *Communications Biology*, *8*(1), 179. https://doi.org/10.1038/s42003-025-07608-9

Croft, M. T., Lawrence, A. D., Raux-Deery, E., Warren, M. J., & Smith, A. G. (2005). Algae acquire vitamin B12 through a symbiotic relationship with bacteria. *Nature*, *438*(7064), 90–93. https://doi.org/10.1038/nature04056

Daniels, M., van Vliet, S. & Ackermann, M. Changes in interactions over ecological time scales influence single-cell growth dynamics in a metabolically coupled marine microbial community. *ISME J* **17**, 406–416 (2023). https://doi.org/10.1038/s41396-022-01312-w

Durham, B. P., Dearth, S. P., Sharma, S., Amin, S. A., Smith, C. B., Campagna, S. R., Armbrust, E. V., & Moran, M. A. (2017). Recognition cascade and metabolite transfer in a marine bacteria-phytoplankton model system. *Environmental Microbiology*, *19*(9), 3500–3513. https://doi.org/10.1111/1462-2920.13834

Falkowski, P. G., Fenchel, T., & Delong, E. F. (2008). The Microbial Engines That Drive Earth's Biogeochemical Cycles. *Science*, *320*(5879), 1034–1039. https://doi.org/10.1126/science.1153213

Fei, C., Ochsenkühn, M. A., Shibl, A. A., Isaac, A., Wang, C., & Amin, S. A. (2020). Quorum sensing regulates 'swim-or-stick' lifestyle in the phycosphere. *Environmental Microbiology*, *22*(11), 4761–4778. https://doi.org/10.1111/1462-2920.15228

Ferrer-González, F. X., Widner, B., Holderman, N. R., Glushka, J., Edison, A. S., Kujawinski, E. B., & Moran, M. A. (2021). Resource partitioning of phytoplankton metabolites that support bacterial heterotrophy. *The ISME Journal*, *15*(3), 762–773. https://doi.org/10.1038/s41396-020-00811-y

Field CB, Behrenfeld MJ, Randerson JT, Falkowski P. Primary production of the biosphere: integrating terrestrial and oceanic components. Science. 1998;281(5374):237.

Fuessel, J., Miller, S., Hecker, C., Trigodet, F., Dlugosch, L., Brinkhoff, T., Dittmar, T., Eren, A. M., Lennartz, S., & Bunse, C. (2025). *Bacterial interactions shape the molecular composition of dissolved organic matter*. In Review. https://doi.org/10.21203/rs.3.rs-7321332/v1

Giordano, N., Gaudin, M., Trottier, C., Delage, E., Nef, C., Bowler, C., & Chaffron, S. (2024). Genome-scale community modelling reveals conserved metabolic cross-feedings in epipelagic bacterioplankton communities. *Nature Communications*, *15*(1), 2721. https://doi.org/10.1038/s41467-024-46374-w

Gómez-Consarnau, L., Sachdeva, R., Gifford, S. M., Cutter, L. S., Fuhrman, J. A., Sañudo-Wilhelmy, S. A., & Moran, M. A. (2018). Mosaic patterns of B-vitamin synthesis and utilization in a natural marine microbial community. *Environmental Microbiology*, *20*(8), 2809–2823. https://doi.org/10.1111/1462-2920.14133

Guillard RRL. Culture of Phytoplankton for Feeding Marine Invertebrates. In: Smith WL, Chanley MH, editors. Culture of marine invertebrate animals: proceedings — 1st conference on culture of marine invertebrate animals greenport. Boston, MA: Springer US; 1975. p. 29-60.

Harvey, E., Yang, H., Castiblanco, E., Coolahan, M., Dallmeyer-Drennen, G., Fukuda, N., Greene, E., Gonsalves, M., Smith, S., & Whalen, K. (2023). Quorum sensing signal disrupts viral infection dynamics in the coccolithophore Emiliania huxleyi. *Aquatic Microbial Ecology*, *89*, 75–86. https://doi.org/10.3354/ame01998

Khalil, A., Bramucci, A. R., Focardi, A., Le Reun, N., Willams, N. L. R., Kuzhiumparambil, U., Raina, J.-B., & Seymour, J. R. (2024). Widespread production of plant growth-promoting hormones among marine bacteria and their impacts on the growth of a marine diatom. *Microbiome*, *12*(1), 205. https://doi.org/10.1186/s40168-024-01899-6

King, K., Bramucci, A. R., Labbate, M., Raina, J.-B., & Seymour, J. R. (2022). Heterogeneous Growth Enhancement of Vibrio cholerae in the Presence of Different Phytoplankton Species. *Applied and Environmental Microbiology*, *88*(17), e01158-22. https://doi.org/10.1128/aem.01158-22

Liu, F., Gledhill, M., Tan, Q.-G., Zhu, K., Zhang, Q., Salaün, P., Tagliabue, A., Zhang, Y., Weiss, D., Achterberg, E. P., & Korchev, Y. (2022). Phycosphere pH of unicellular nano- and

micro- phytoplankton cells and consequences for iron speciation. *The ISME Journal*, *16*(10), 2329–2336. https://doi.org/10.1038/s41396-022-01280-1

Mayali, X., Weber, P. K., Brodie, E. L., Mabery, S., Hoeprich, P. D., & Pett-Ridge, J. (2012). High-throughput isotopic analysis of RNA microarrays to quantify microbial resource use. *The ISME Journal*, *6*(6), 1210–1221. https://doi.org/10.1038/ismej.2011.175

Meyer N, Bigalke A, Kaulfuß A, Pohnert G. Strategies and ecological roles of algicidal bacteria. FEMS Microbiol. Rev. 2017;41(6):880-99.

Raina, J.-B., Lambert, B. S., Parks, D. H., Rinke, C., Siboni, N., Bramucci, A., Ostrowski, M., Signal, B., Lutz, A., Mendis, H., Rubino, F., Fernandez, V. I., Stocker, R., Hugenholtz, P., Tyson, G. W., & Seymour, J. R. (2022). Chemotaxis shapes the microscale organization of the ocean's microbiome. *Nature*, *605*(7908), 132–138. https://doi.org/10.1038/s41586-022-04614-3

Rolland, J. L., Stien, D., Sanchez-Ferandin, S., & Lami, R. (2016). Quorum Sensing and Quorum Quenching in the Phycosphere of Phytoplankton: A Case of Chemical Interactions in Ecology. *Journal of Chemical Ecology*, *42*(12), 1201–1211. https://doi.org/10.1007/s10886-016-0791-y

Sarmento, H., & Gasol, J. M. (2012). Use of phytoplankton-derived dissolved organic carbon by different types of bacterioplankton. *Environmental Microbiology*, *14*(9), 2348–2360. https://doi.org/10.1111/j.1462-2920.2012.02787.x

Sarmento, H., Morana, C., & Gasol, J. M. (2016). Bacterioplankton niche partitioning in the use of phytoplankton-derived dissolved organic carbon: Quantity is more important than quality. *The ISME Journal*, *10*(11), 2582–2592. https://doi.org/10.1038/ismej.2016.66

Seymour, J. R., Amin, S. A., Raina, J.-B., & Stocker, R. (2017). Zooming in on the phycosphere: The ecological interface for phytoplankton–bacteria relationships. *Nature Microbiology*, *2*(7), 17065. https://doi.org/10.1038/nmicrobiol.2017.65

Seth, E. C., & Taga, M. E. (2014). Nutrient cross-feeding in the microbial world. *Frontiers in Microbiology*, *5*. https://doi.org/10.3389/fmicb.2014.00350

Shibl, A. A., Isaac, A., Ochsenkühn, M. A., Cárdenas, A., Fei, C., Behringer, G., Arnoux, M., Drou, N., Santos, M. P., Gunsalus, K. C., Voolstra, C. R., & Amin, S. A. (2020). Diatom modulation of select bacteria through use of two unique secondary metabolites. *Proceedings of the National Academy of Sciences*, *117*(44), 27445–27455. https://doi.org/10.1073/pnas.2012088117

Słomka, J., Alcolombri, U., Carrara, F., Foffi, R., Peaudecerf, F. J., Zbinden, M., & Stocker, R. (2023). Encounter rates prime interactions between microorganisms. *Interface focus*, *13*(2), 20220059. https://doi.org/10.1098/rsfs.2022.0059

Stock, F., Bilcke, G., De Decker, S., Osuna-Cruz, C. M., Van Den Berge, K., Vancaester, E., De Veylder, L., Vandepoele, K., Mangelinckx, S., & Vyverman, W. (2020). Distinctive Growth and Transcriptional Changes of the Diatom Seminavis robusta in Response to Quorum Sensing

Related Compounds. *Frontiers in Microbiology*, *11*, 1240. https://doi.org/10.3389/fmicb.2020.01240

Zoccarato, L., Sher, D., Miki, T., Segrè, D., & Grossart, H.-P. (2022). A comparative whole-genome approach identifies bacterial traits for marine microbial interactions. *Communications Biology*, *5*(1), 276. https://doi.org/10.1038/s42003-022-03184-4

Chapter 1: Flavobacterial exudates disrupt cell cycle progression and metabolism of the diatom Thalassiosira pseudonana

Zinka Bartolek[1], Shiri Graff van Creveld[1], Sacha Coesel[1], Kelsy R. Cain[1], Megan Schatz[1], Rhonda Morales[1], E. Virginia Armbrust[1]

[1]School of Oceanography, University of Washington, Seattle, WA 98195

Previously published as:

Bartolek, Z., Van Creveld, S. G., Coesel, S., Cain, K. R., Schatz, M., Morales, R., & Virginia Armbrust, E. (2022). Flavobacterial exudates disrupt cell cycle progression and metabolism of the diatom *Thalassiosira pseudonana*. *The ISME Journal*, *16*(12), 2741–2751. https://doi.org/10.1038/s41396-022-01313-9

Abstract

Phytoplankton and bacteria form the base of marine ecosystems and their interactions drive global biogeochemical cycles. The effects of bacteria and bacteria-produced compounds on diatoms range from synergistic to pathogenic and can affect the physiology and transcriptional patterns of the interacting diatom. Here, we investigate physiological and transcriptional changes in the marine diatom *Thalassiosira pseudonana* induced by extracellular metabolites of a known antagonistic bacterium *Croceibacter atlanticus*. Mono-cultures of *C. atlanticus* released compounds that inhibited diatom cell division and elicited a distinctive morphology of enlarged cells with increased chloroplast content and enlarged nuclei, similar to what was previously observed when the diatom was co-cultured with live bacteria. The extracellular *C. atlanticus* metabolites induced transcriptional changes in diatom pathways that include recognition and signaling pathways, cell cycle regulation, carbohydrate and amino acid production, as well as cell wall stability. Phenotypic analysis showed a disruption in the diatom cell cycle progression and an increase in both intra- and extracellular carbohydrates in diatom cultures after bacterial exudate treatment. The transcriptional changes and corresponding phenotypes suggest that extracellular bacterial metabolites, produced independently of direct bacterial-diatom interaction, may modulate diatom metabolism in ways that support bacterial growth.

Introduction

Diatoms are photosynthetic eukaryotes responsible for up to 20% of annual photosynthesis on Earth [1, 2]. They serve as the base of many marine food webs and can form dense blooms that reach relatively high cell abundances. Diatoms have co-existed and co-evolved with bacteria for millions of years, ultimately developing close interactions that facilitate the exchange of nutrients and metabolites [3]. These interactions are hypothesized to occur within the phycosphere, a diffusive boundary layer that surrounds individual phytoplankton cells. Here, mass transport is limited to diffusion and is not impacted by turbulence [4, 5]. Within this microenvironment, nutrient and fixed carbon concentrations can exceed background levels by orders of magnitude, and bacterial growth can be stimulated by extracellular metabolites of the diatom [6]. The phycosphere also contains enriched concentrations of signaling compounds and secondary metabolites that modify the behavior of the host diatom and associated bacteria [7].

Diatoms and bacteria display pathogenic and mutualistic interactions, some of them obligatory [3, 5, 7]. When co-cultured with diatoms, bacteria can influence diatom growth rate, metabolism, and cell cycle progression [8,9,10,11,12]. Fundamentally, the interactions between diatoms and bacteria are based on various metabolic exchanges. In mutualistic interactions, bacteria provide micronutrients necessary for diatom growth, such as vitamin B_{12} or bioavailable iron in the form of siderophores, while the diatom supplies the bacteria with readily accessible carbon and sulfur sources necessary for bacterial growth [5, 11, 13]. Interactions between diatoms and bacteria are often species-, and even strain-specific, where the interacting bacteria are adapted to a relatively narrow range of compounds secreted by the diatom, resulting in distinct bacterial communities that are consistent across strains and time [14,15,16,17]. Conversely, diatoms can shape the composition of their bacterial communities through excreted metabolites to select for bacterial communities that are beneficial to the diatom and to prevent the establishment of potentially harmful bacteria [18,19,20]. Additionally, phytoplankton exudates can influence the production of the bacterial exo-metabolome [21].

Algicidal bacteria generally belong to the phyla of *Bacteroidetes*, *Gammaproteobacteria*, and *Alphaproteobacteria*, and interact with diatoms through direct contact, close proximity to the diatom cell, or through production of excreted algicides [7]. Bacterial attachment to other phytoplankton such as dinoflagellates and coccolithophorids has also been described [22, 23]. Algicide production by bacteria often depends on the high nutrient availability found in the phycosphere, and their structure and mode of action can range from small metabolites with specific cellular targets to large enzymes catalyzing the change of numerous processes in the cell [7]. For example, bacterial quorum sensing compounds can have antagonistic effects on diatom growth and transcriptional changes, with long chain N-acyl homoserine lactones inhibiting growth of *Seminavis robusta* and inducing changes in lipid metabolism and cell cycle progression [24, 25]. However, the activity of algicidal compounds produced by bacteria in a contact-independent manner is understudied and not well understood.

Using a bacterial consortium previously isolated from the diatom *Pseudo-nitzschia multiseries*, Amin et al. demonstrated the effect of various single strain isolates on diatom physiology through co-culture experiments with different diatom species [11]. The isolated flavobacterium *Croceibacter atlanticus* (SA60) exhibited an inhibitory effect on several different diatoms including the model *Thalassiosira pseudonana* [10]. In co-culture, live *C. atlanticus* attached directly to the surface of diatom cells and inhibited *T. pseudonana* cell division and induced enlarged cells with multiple plastids [10]. Here, we asked whether a similar

diatom morphology could be elicited by *C. atlanticus* bacterial exudates in bacterial-free culturing conditions. We investigated the effect of extracellular bacterial metabolites produced by mono-cultures of *C. atlanticus* on the growth, morphology, and transcriptional profiles of the diatom *T. pseudonana*. Resulting changes in cell growth and transcriptional profiles of diatoms exposed to *C. atlanticus* extracellular metabolites suggest a disruption of cell cycle-related processes and an enhancement of carbohydrate and amino acid metabolism, independent of bacterial attachment. Phenotyping of treated diatom cultures confirmed cell cycle disruption and increased carbohydrate concentrations in the media of treated diatom cells. This indicates that metabolites produced and released by *C. atlanticus* bacteria that were not previously exposed to diatoms can modulate the transcriptional profile, physiology, and phycosphere composition of diatoms to potentially enhance bacterial growth.

Methods

C. atlanticus filtrate and H_2O_2 treatment of T. pseudonana cells

Glycerol stocks of *C. atlanticus* SA60 NCMA B36 (National Center for Marine Algae and Microbiota, NCMA, East Botany, ME, USA) isolated from *Pseudo-nitzschia multiseries* CLNN-17 [11] were streaked on marine agar plates (marine broth (MB) with 1.5% w/v agar) and grown at 30 °C for 3-4 days. Single colonies were grown overnight in MB with shaking at 30 °C until an OD_{600} (NanoDrop One Microvolume UV-Vis Spectrophotometer OD_{600}) of ~0.3 was reached. Cells were inoculated into synthetic seawater medium Aquil (National Center for Marine Algae and Microbiota (NCMA)) with 5% MB (v/v), grown with shaking at 30 °C until the exponentially growing cells reached $OD_{600} = 0.3$ (4×10^8 cells/ml), filtered twice through a 0.22 µm glass fiber filter; and the resulting filtrate was collected and checked microscopically for bacterial contamination by staining with SYBR Green (1:10000) for 20 min. Additionally, 500 µl of filtrate was inoculated into 10 ml of MB, incubated at room temperature for 5 days and inspected visually for bacterial contaminants. This sterile filtrate was stored in the dark at 4 °C prior to use.

To measure the effect of bacterial-free *C. atlanticus* filtrate on *T. pseudonana* cell growth, the filtrate was added at different concentrations (1.6, 3.3, 13, 33, 66; and 100% v/v filtrate to L1 + Si medium) to triplicate treatment flasks of 5000 cells/ml *T. pseudonana* CCMP 1335 (obtained from NCMA) that had been maintained as a semicontinuous batch culture in 30 ml L1 + Si medium (NCMA) [26] at 20 °C in a 16:8 h light:dark cycle at ~ 160 µmol photons m^{-2} sec^{-1}. Control *T. pseudonana* cultures were treated with corresponding volumes of filtered Aquil with 5% v/v MB (the bacterial growth media). A more comprehensive time course experiment was conducted with control cultures and 33% v/v of bacterial-free *C. atlanticus* filtrate in a final volume of 2 L. At 7 h into the light period of each day of the 8-day experiment, in vivo chlorophyll *a* fluorescence (Turner Designs 10-AU fluorometer) was measured, and samples were collected to determine cell abundance (Guava easyCyte Plus Flow Cytometry System) and photochemical yield of photosystem II (Fv/Fm) after a 15 min dark acclimation (Phyto-PAM fluorometer). Samples for microscopy (2 ml) and for cell cycle analysis (2 ml) were taken daily, incubated for 20 min with 2% v/v glutaraldehyde, flash frozen in liquid nitrogen, and stored at −80 °C prior to analysis. Samples (200 ml) for RNA extraction were collected from each replicate flask at three timepoints: 24 h, 72 h, and 120 h after the start of the experiment by filtering diatom cells onto 0.2 µm polycarbonate membrane filters, flash freezing in liquid nitrogen, and storing the filters at −80 °C prior to analysis. In a separate experiment

using the same treatment and control conditions in a final volume of 30 ml, triplicate samples were collected for diatom cell cycle analysis every 2 h for 12 h, fixed with glutaraldehyde, flash frozen in liquid nitrogen and stored at −80 °C.

Exponentially growing *T. pseudonana* cells (f/2+Si media at 20 °C in a 16:8 h light:dark cycle at ~ 160 µmol photons m^{-2} sec^{-1}) were treated in triplicate with 0 µM or 200 µM of hydrogen peroxide. Two hours after treatment, 500 ml of cultures were harvested for RNA extraction as described above.

Cell morphology measurements and imaging

T. pseudonana cells were imaged via epifluorescence microscopy (470/40 nm blue excitation 515 nm long-pass emission filter) and brightfield imaging. To obtain measurements of nuclear area and chloroplast area, 1 ml of glutaraldehyde-preserved diatom cells was concentrated onto black 0.2 µm pore size polycarbonate filters, the DNA of cells collected on the filter was stained using a SYBR Green I stain, and the filter subsequently attached to microscope slides, as described in van Tol et al., 2017 [10]. Fluorescent images at 60x magnification were taken (Nikon eclipse 80i microscope) and processed using ImageJ [27] software. Nucleus and chloroplast areas were obtained by splitting the 60x fluorescent images into three channels (RGB), and the wand tool was used to select nuclei (green channel) and chlorophyll fluorescence (red channel). The Kolmogorov-Smirnov test was used to test for significant differences ($p < 0.05$) in distributions of control compared to treated cells at each time point. To obtain measurements of cell area, 1 ml of glutaraldehyde-preserved diatoms cells was stained with SYBR Green (1:10000), incubated for 20 min at room temperature, spun down at 10000 rpm for 1 min, and the supernatant discarded for a final volume of 10 µl. Resulting concentrated diatom cells were mounted onto transparent slides, and fluorescent and brightfield images at 20x magnification were taken (Leica DMi8 microscope). Cell area and maximum and minimum cell diameter measurements were obtained from 20x brightfield images using the ImageJ [27] wand tool and straight-line selection tool. Analysis was restricted to in-focus cells entirely within the camera frame.

Flow cytometry

Samples for flow cytometry were stained with SYBR Green I (1:10000), incubated for 20 min, and run on a BD Influx flow cytometer (BD, Franklin Lakes, NJ, USA) (488 nm laser; emission 530 nm, 40 nm band pass) on a linear scale with at least 10,000 *T. pseudonana* cells analyzed per sample [10]. *T. pseudonana* cells were gated using the FCSplankton package in R (https://github.com/fribalet/FCSplankton). SYBR Green I-stained cells were manually gated on forward scatter and 530/40 nm emission using the set_gating_params() function. Forward scatter distributions, as a proxy for cell size, were normalized to 1 µm beads (Molecular Probes, Eugene, OR, USA) that were added to each sample as an internal standard. Contrary to control cultures, the G1, S, and G2 populations of the filtrate-treated cultures did not follow a Gaussian distribution, precluding quantitative cell cycle decomposition of these samples. To compare the distributions of the SYBR-stained control and filtrate-treated cells, the distributions were aligned to the mode of the G1 peak. Cell cycle stage boundaries were drawn based on the signal of the control cultures, allowing for qualitative comparison of the G1, S, and G2 population distributions between control and filtrate treated cultures.

Carbohydrate content analysis

Carbohydrate content was determined via the phenol-sulfuric acid assay as described in DuBois et al. 1956 [28]. *T. pseudonana* cells (initial concentration 1×10^5cells/ml) were transferred in triplicate into either control media (Aquil + 5% MB) or undiluted bacterial-free *C. atlanticus* filtrate (30 ml total volume) and 2 ml samples were collected for total (dissolved + particulate) and dissolved carbohydrate content at 24 h and 48 h. Samples for dissolved carbohydrates were filtered through a 0.4 µm filter. Both the resulting filtrate and samples for total carbohydrates were stored at −80 °C. To control for carbohydrates introduced by the bacterial filtrate itself, samples of bacterial-free *C. atlanticus* filtrate were also measured. For each replicate, 250 µl sample was combined with 250 µl of 5% phenol in DI water and 1250 µl of concentrated sulfuric acid, gently mixed, and incubated for 20 min at room temperature. Absorbance at 485 nm was measured and compared to a dilution series of glucose standards. To obtain the final carbohydrate content of treated cells, the carbohydrate content of the *C. atlanticus* filtrate was subtracted from measurements of total and dissolved fractions of treated cells. Particulate carbohydrate content was calculated as the difference between total and dissolved fractions. The students *t*-test was used to test for significance ($p < 0.05$) between control and treated cells at each time point.

T. pseudonana transcriptome sequencing and analysis

The *T. pseudonana* transcriptome processing pipeline of control and filtrated-treated cultures (Sup. 1.1) included RNA extraction, sequencing, and quantification (blue boxes); identification of differentially transcribed genes and genes that were co-transcribed (yellow boxes); and gene enrichment analyses (green boxes) to identify subsets of genes for further analyses. The code applied for this analysis is shared on GitHub page (https://github.com/armbrustlab/Thaps_Catl) and all data are deposited on Zenodo (10.5281/zenodo.6672614).

Total RNA was extracted from 0.2 µm polycarbonate membrane filters with the Zymo Direct-zol RNA MiniPrep Plus kit. Poly-A-selection, library preparation and sequencing were performed at the Northwest Genomics Center (University of Washington) with NextSeq (Illumina). Sequence reads were trimmed using Trimmomatic 0.39 [29], run in paired-end mode with cut adaptor and other Illumina-specific sequences (ILLUMINACLIP) set to TruSeq3-PE.fa:2:30:10:1:true, Leading and Trailing thresholds of 25, a sliding window trimming approach (SLIDINGWINDOW) of 4:15, an average quality level (AVGQUAL) of 20, and a minimum length (MINLEN) of 60. Reads were mapped to the genome of *T. pseudonana* (Thaps3 FilteredModels2) using Hisat2-2.1.0. We calculated the number of *T. pseudonana* reads that aligned to the gene models from resulting SAM alignment files with aligned sequences used in subsequent analyses and normalized using transcripts per million (TPM).

The edgeR pipeline [30] and Weighted Gene Co-expression Network Analysis (WCGNA) [31] were used in parallel to identify the transcriptional changes in response to treatment with the bacterial filtrate (yellow boxes, Sup. 1.1). EdgeR was used to detect differential expression of $\log_2$ transformed transcript levels based on pairwise comparisons between the control samples and the samples treated with bacterial-free *C. atlanticus* filtrate for each sampled time point (24 h, 72 h, and 120 h). A generalized linear model (GLM) quasi-likelihood F-test (QLTF) was used to test for significant differential expression ($p < 0.01$ and false discovery rate (FDR) < 0.05). Significantly differentially expressed genes with $|\log FC| > 1$ were used for downstream analysis.

Signed WCGNA was used to identify modules of co-expressed genes using both control and treatment transcriptomes with log_2 transformed transcripts per million (TPM) as input abundance data; those genes with transcript abundance below the TPM median in greater than 25% of samples were removed from analysis (Sup. 1.2A; Sup. 1.1). Two outlier samples (t24_cont_3 and t72_cont_72) were removed from further network analysis as recommended [31], based on a branch height > 60 in the sample clustering dendrogram of Euclidean distance between samples (Sup. 1.2B). To identify modules of groups of genes with similar expression patterns, a network of gene co-expression was constructed by making a weighted adjacency matrix that was then transformed into a topological overlay matrix (TOM) to estimate connectivity in the network. Average linkage hierarchical clustering with a soft-threshold power=18 (Sup. 1.2C) was used to construct the clustering tree structure of the TOM (Sup. 1.3A) [31]. Modules were assigned based on a dynamic tree cut method (minimum number of genes per module=30; deep split=1), and similar modules were merged (cut height threshold=0.25) (Sup. 1.3). Each module was correlated to different parameters measured during the experiment to identify modules with strongest correlation to treatment with the bacterial filtrate. Modules with a Pearson's correlation >0.6 to *C. atlanticus* filtrate treatment were selected based on module eigengenes (MEs) and module significance >0.5 (average gene significance for each module, where gene significance is the log10 transformed *p* value from the linear regression between gene expression and *C. atlanticus* filtrate treatment) (Sup. 1.4).

Functional enrichment analysis (green boxes, Sup. 1.1) was performed with topGO [32] using a custom gene ontology (GO) annotation file (Joint Genome Institute) and a classic Fisher test ($p < 0.01$). Functional enrichment analysis was performed on each WGCNA module, as well as on the differentially expressed genes at each of the three timepoints identified by edgeR (Table S1). Enrichment analysis of GO Biological Process (BP) terms were visualized as a network using the igraph R package [33]. Node and edge files were derived from GO term similarity analysis using NaviGO [34], with a Lin's similarity score cutoff of 0.7.

Two sets of 'hub-genes' were considered for this work: 1) genes that were identified as those with gene significance and module membership > 0.8 (Sup. 1.5) and differential expression at one or more time points (edgeR, $p < 0.01$) (indicated in red section of Euler diagram, Sup. 1.1; Sup. 1.2) those differentially expressed genes ($p < 0.01$ in the edgeR analysis) present in pathways that were enriched in filtrate-correlated modules (indicated in orange section of Euler diagram, Sup. 1.1). In total 792 genes (out of a total of 11869 transcribed genes) were identified for further analysis.

Data availability

All sequencing data is publicly available on NCBI under the accession number GSE197934. Code used for data processing and statistical analysis is available at GitHub (https://github.com/armbrustlab/Thaps_Catl). All supplementary data tables and secondary data is available on Zenodo (10.5281/zenodo.6672614). The *C. atlanticus* SA60 strain used in these experiments is available at Bigelow National Center for Marine Algae and Microbiota under the identifier NCMA B36.

Results

Physiological responses of T. pseudonana to C. atlanticus filtrate treatment

To determine whether physical interaction is required for *C. atlanticus* antagonism against diatoms [10], we collected bacteria-free filtrate from mono-cultures of *C. atlanticus* cells grown without previous exposure to the diatoms. Different proportions of the bacteria-free filtrate (1.6, 3.3, 13, 33, 66, and 100 % v/v filtrate to diatom growth medium) were added to exponentially growing axenic cultures of *T. pseudonana* and compared to control cultures. Treatment with *C. atlanticus* filtrate affected the relative in vivo chlorophyll *a* fluorescence levels of *T. pseudonana* cultures in a dose-dependent manner, with a half maximal inhibitory concentration of 33% v/v (Sup. 1.6), indicating that part of the morphology in the diatom was induced by bacterial extracellular compounds. The increased relative in vivo chlorophyll a fluorescence at low concentrations of filtrate (<13% v/v) could reflect either an increase in in vivo fluorescence per cell, or an increase in the total number of cells; high concentrations (66, 100%) of *C. atlanticus* filtrate were deleterious to the diatom within 2 days.

A time course of bacterial exudate impacts on diatom cell abundance, physiology, and transcriptional modulation was examined in more detail by adding 33% v/v bacterial filtrate to axenic cultures of *T. pseudonana*. The maximum cell abundance (Fig. 1.1A) and chlorophyll *a* fluorescence (Sup 1.7A) were reduced in the filtrate-treated cultures, along with a slight decrease in specific growth rate for filtrate-treated cultures (0.77 day^{-1}), as compared to control cultures (0.81 day^{-1}; Sup. 1.7B). The photosynthetic efficiency (F_v/F_m) at the initial time point (24 h) was significantly lower in control samples than in filtrate-treated samples (Sup. 1.7C). A subsequent experiment indicated that dilution of *T. pseudonana* control cultures from 1×10^6 cells/ml to 5000 cells/ml led to a comparable reduction in Fv/Fm (Sup. 1.7D). The filtrate-treated cells did not display a similar transient response to dilution.

Microscopy-based cell-size measurements suggested that a subset of *T. pseudonana* cells increased in size over time after exposure to *C. atlanticus* filtrate, with visible changes in cellular morphology (Fig. 1.1B, C, D). The larger cells contained larger chloroplasts (Fig. 1.1D, Sup. 1.8A) and an increased average nuclear area (Fig. 1.1D, Sup. 1.8B). Flow cytometry-derived forward and side scatter distributions differed between control and filtrate-treated cultures, a difference that became apparent after 12 h of exposure to the bacterial-free filtrate (Sup. 1.9). Forward scatter is a proxy for cell size and side scatter reflects cell shape and internal cellular structures [35]. Increases in both proxies in treated cells is consistent with the elongated phenotype observed with microscopy (Fig. 1.1B, C; Sup. 1.10). These morphological changes are similar to that observed when *T. pseudonana* was co-cultured with live *C. atlanticus* cells [10], indicating that bacterial exo-metabolites are at least partially responsible for the observed changes.

Transcriptional responses of T. pseudonana to C. atlanticus filtrate treatment

Triplicate samples for transcriptome analysis were collected at three time points: prior to a difference in cell abundances between treatments (24 h), when the cultures were in early exponential growth phase yet cell abundances between treatments differed (72 h), and when both the control and treated cultures were in mid-late exponential growth phase (120 h, marked by arrows in Fig. 1.1A). The resulting transcriptomes were analyzed with two complementary methods: differential gene expression and Weighted Gene Co-expression Network Analysis (WCGNA) [31] (Sup. 1.1). The number of significant differentially expressed (DE) genes ($p < 0.01$, FDR < 0.05 and $|\log_2$ fold change $| > 1$) in treatment compared to control samples increased over time with 2050, 3039, and 3303 DE genes at 24 h, 72 h, and 120 h, respectively

(Sup. 1.11A). A Mean Squared Deviation analysis of the top 500 most differentially expressed genes showed a separation of samples based on treatment and, to a lesser extent, sampling time, with replicate samples grouping together (Sup 1.11B). Differentially expressed genes were enriched in ribosome biogenesis, RNA translation, amino acid metabolism, response to stress, RNA modification, and transport at all three timepoints (Fisher's exact test; Sup. Table 1.1).

We used WCGNA to cluster *T. pseudonana* genes into modules of genes with comparable transcription profiles. The resulting 12 modules were correlated (Pearson) with measured experimental parameters and phenotypic changes observed in the diatom with or without bacterial filtrate treatment (Sup. Table 1.2). Most modules were correlated with either i) bacterial filtrate treatment-dependent traits (percent of enlarged cells, maximum cell diameter, and minimum cell diameter; modules 9, 5 and 21), or ii) time-dependent traits (chlorophyll *a* fluorescence, cell count, and sampling time; modules 11, 1, 35 and 34) (Sup. 1.4). Modules correlated with bacterial filtrate-treatment (9, 5, 21) were enriched in cell cycle progression, carbohydrate metabolism and biosynthesis, amino acid metabolism, transcriptional regulation and signaling genes (Fig. 1.2; Sup. Table 1.1). In contrast, modules correlated with time-dependent traits were enriched ($p < 0.01$) in transport, localization, RNA processing, ribosomal biogenesis, catabolic processes, and transcriptional regulation (GO Biological Process terms; Sup. Table 1.1). These functional features were also identified by differential expression analysis, highlighting the importance of separating time-dependent and treatment-dependent traits.

Further analysis of genes that were both differentially expressed and correlated to the *C. atlanticus* filtrate treatment (Sup. 1.1) highlighted five cellular pathways in which expression patterns were significantly altered by *C. atlanticus* filtrate treatment: cell cycle progression; transport; amino acid metabolism; chitin and carbohydrate metabolism; and recognition, signaling and regulation (Fig. 1.3). The *C. atlanticus* filtrate induced changes in the abundance of transcripts associated with numerous components of cell cycle regulation: kinase complexes, including regulatory cyclin subunits known to control progression through G1/S, G2/M and mitosis stages of the cell cycle; cyclin dependent kinases (CDKs) that act as catalytic subunits (Fig. 1.3, Sup. Table 1.3); putative casein kinases that inhibit CDK activity; [36] and the mitotic spindle checkpoint protein MAD-2 (Sup. Table 1.3). In addition to regulatory genes, several hub genes (see methods) encoded motor proteins involved in chromosome segregation, kinetochore binding and spindle formation, including a chromokinesin, centrosome-associated actins, and condensin-like proteins (Fig. 1.3, Sup. Table 1.3). Transcripts encoding the plastid division protein FtsZ are elevated at 72 h and 120 h in the treated cultures as compared to control cultures, coinciding with the increased chloroplast content in treated diatom cells at these timepoints (Fig. 1.3, Sup. 1.8A).

Filtrate treatment resulted in upregulation of transcription of genes involved in carbon metabolism such as the TCA cycle, glycolysis and glucogenesis, and chrysolaminarin-based storage mechanisms (Fig. 1.3, Sup. Table 1.3). We also observed an increase in transcripts associated with amino acid biosynthesis pathways in treated cells, with changes associated with biosynthesis pathways of cysteine; aspartate and alanine; valine, leucine and isoleucine; asparagine, threonine, methionine and lysine; and ornithine and arginine (Fig. 1.3, Sup. Table 1.3). Silicic acid transporter-encoding transcript abundance (Sit1 and Sit3) was also upregulated (Fig. 1.3, Sup. Table 1.3). In contrast, the transcript abundance decreased for twenty genes that encode chitinases and chitin synthases and the enzymes required for biosynthesis of cell-wall-associated chitin polymers (Sup. Table 1.3). The treatment also resulted in downregulation of

genes that encode proteins localized to the girdle band region of the cell wall that regulate the stability of the cell wall during cell division, which suggests a destabilization and increased permeability of the cell wall of cells treated with *C. atlanticus* filtrate (Fig. 1.3, Sup. Table 1.3).

To understand whether the observed transcriptional responses were specific to the bacterial exudate or represented a general cellular response to stress, we evaluated the stress response of *T. pseudonana* to H_2O_2-induced oxidative stress after 2 h (Fig. 1.3). A large number (1569) of genes were differentially expressed under both H_2O_2-induced oxidative stress and *C. atlanticus* filtrate treatment and included genes involved in cell cycle regulation, signaling, and amino acid metabolism (Fig. 1.3, Sup. 1.12, Sup. Table 1.1). Genes enriched in processes related to stress response, amino acid biosynthesis, photosynthesis, and general metabolism were differentially expressed in oxidative stress or in both oxidative stress and *C. atlanticus* filtrate treatments. Conversely, genes differentially expressed specifically in response to *C. atlanticus* filtrate treatment were enriched in cell cycle control, including DNA replication and microtubule-based movement (Sup. Table 1.1, Fig. 1.3). Although oxidative stress and ROS signaling can play an important role in microbial interactions [37,38,39], we did not find transcriptional evidence that *C. atlanticus* filtrate induced an oxidative stress response in *T. pseudonana*. Together these results indicate that observed changes in transcription of genes involved in cell cycle control are due to the filtrate treatment rather than a general response to stress.

Phenotypic assessment of transcriptome-identified pathways

The *C. atlanticus* filtrate induced changes in the abundance of transcripts associated with cell cycle regulation. We confirmed disruption of cell cycle progression in treated *T. pseudonana* cells using flow cytometry-based cell cycle analysis (Fig. 1.4, Sup. 1.13). Under control conditions, the proportion of control cells in S phase decreased and the proportion of cells in G1 phase increased between 6–12 h after the start of the experiment, indicating active cell division (Fig. 1.4). In contrast, in treated cultures the proportion of S phase versus G1 phase remained constant during this same interval, suggesting an almost immediate disruption of cell cycle progression in the treated cultures. The difference in DNA distributions between control and treated cultures became more apparent over time with a gradual increase in the proportion of treated cells in S and G2 phases after 96 h. During late exponential and stationary phases of growth (120–168 h), a large proportion of cells in the treated cultures were arrested in S phase and a subset of treated cells displayed DNA content greater than G2, reflecting unregulated DNA replication and the presence of larger nuclei, in accordance with the differential expression patterns of cell-cycle related genes (Figs. 1.3, 1.4).

Treatment with *C. atlanticus* filtrate also resulted in changes in transcript levels associated with amino acid and carbohydrate metabolism as well as cell wall components (Fig. 1.3). These transcriptional profiles suggest an upregulation of these biosynthesis pathways that, with a coinciding disruption of the cell wall stability, could potentially result in enhanced release of organic compounds. To test whether diatoms exposed to *C. atlanticus* filtrate produce and exude higher levels of complex carbohydrates, we subjected cultures to 24 h and 48 h of *C. atlanticus* filtrate and analyzed the extracellular, total, and particulate carbohydrate content of these treated diatoms relative to control cultures. Diatoms were treated with undiluted filtrate to ensure a rapid and uniform response by treated cells (Sup. 1.14). We observed an increase in extracellular carbohydrates after both 24 h and 48 h of treatment in total carbohydrate content,

indicative of increased carbohydrate production (Fig. 1.5). The increase in the dissolved phase suggests an increased release of carbohydrates by the diatom into the media (Fig. 1.5), in accordance with the transcriptomic results.

Discission

Interactions between diatoms and bacteria are mediated through a series of metabolic exchanges that can influence the behavior and metabolism of both organisms. We focused here on a previously characterized antagonistic interaction between the model diatom *T. pseudonana* and the bacterium *C. atlanticus* [10]. In co-culture, live bacteria attach to the surface of diatom cells, inhibit cell division, and modify the morphology of the diatom [10]. Here, we developed an experimental system that relies on bacterial exudates rather than the presence of live bacteria. We found that the *T. pseudonana* morphology of increased cell size with larger chloroplast area and nuclei arises during exposure to bacterial exudates without the presence of live bacteria. In addition, the observed increase in production of carbohydrate-based compounds and inferred increase in amino acid production and cell wall permeability suggested that released bacterial metabolites may modify the metabolism of the diatom to better support bacterial growth. Resulting changes in the diatom's morphology, especially the increased cell size, could have far reaching ecological consequences, including increased sinking rates and grazing susceptibility, and affect light absorption patterns and export of carbon from the surface ocean [40,41,42].

A schematic based on our transcriptome and phenotypic analyses summarizes how extracellular metabolites of *C. atlanticus* impact *T. pseudonana* physiology (Fig. 1.6). Transcriptional changes in diatom cell cycle progression, silicic acid uptake, and chitin metabolism corresponded to the observed morphology. The combination of flow-cytometry-based cell cycle analysis with the larger nuclei observed in microscopy and the differentially expressed genes in the treated cells indicate that released *C. atlanticus* metabolites inhibit *T. pseudonana* cell cycle progression. Changes in cell cycle progression were evidenced by disruption of genes encoding numerous regulatory and motor proteins involved in division (Fig. 1.6), as well as a gradual increase in the proportion of S and G2 phase cells (Fig. 1.4). The atypical increase in S phase cells with *C. atlanticus* filtrate treatment (Fig. 1.4) could indicate replication stress, which can lead to uneven chromosome segregation and aneuploidy in eukaryotic cells [43]. We observed transcriptional changes in genes responsible for spindle formation and chromosome segregation (Fig. 1.6); analogous genes in the centrosome of other eukaryotic cells were associated with mitotic delay and genomic instability [44]. Similarly, observed transcriptional changes in kinetochore function (Fig. 1.6) have been linked to development of aneuploidy, as kinetochores are responsible for binding microtubules, coordinating chromosome movements, and activating spindle checkpoints [45]. Additionally, the downregulation of centrosome-associated actins involved in spindle separation and contractile ring formation during cytokinesis suggests that the filtrate inhibits cytokinesis of *T. pseudonana* cells [46], leading to enlarged diatom cells that continue to replicate their DNA but are unable to divide (Fig. 1.1, Fig. 1.6). Previous work shows that bacterial quorum sensing molecules cause cell cycle arrest in coccolithophores, while at the same time protecting affected cells from viral mortality, highlighting the importance of bacterial signals in multitrophic interactions [47]. The multitrophic effect of cell cycle changes in *T. pseudonana* warrants further investigation and could have implications for viral or grazing susceptibility. Cell volume in *Thalassiosira* is constrained by G1 DNA content [48]; as long as cell metabolism is active in

aneuploid or polyploid cells unable to divide, an increase in size is expected. Although bacterial algicides can induce cell cycle arrest and programmed cell death [49], live diatom cells displaying the morphology induced by the *C. atlanticus* filtrate continued to photosynthesize (Sup. 1.7A) and produce organic material bioavailable to bacteria (Fig. 1.5).

Enlargement of the silica cell wall is a necessary accompaniment for cell expansion; we observed an upregulation of genes encoding silicic acid transporters suggesting that *T. pseudonana* increased silicic acid uptake (Fig. 1.3) [50]. Similarly, transcriptional patterns of genes encoding girdle band localized proteins involved in chitin synthesis and cell wall stability suggested a destabilization of the silica cell wall, which could also allow for the observed cell enlargement (Fig. 1.6). The predominant downregulation of genes required for chitin synthesis suggested a decoupling of silicon precipitation and chitin utilization, potentially leading to an underproduction of extracellular chitin fibers that slow the sinking rate of diatoms. Both filtrate-treated *T. pseudonana* cells and *T. pseudonana* co-cultured with live *C. atlanticus* sank to the bottom of the tube [10].

Bacterially produced compounds may manipulate the metabolism of the diatom to support bacterial growth. This can be achieved by specific enhancement of biosynthesis of compounds beneficial for the bacteria and/or enhanced export of metabolites from the diatom to the phycosphere. Repression of cell division may also indirectly support bacterial growth and allow for an extended leakage of metabolites as cells continue to photosynthesize and fix carbon without dividing. Such mechanisms have been observed in plants; polyploidy is common in cells near sustained nutrient exchange sites between the host and biotroph, where enlarged host cells can support the growth of their biotroph by producing excess carbon [51]. We found that treated *T. pseudonana* cells released twice as many dissolved carbohydrates compared to untreated cells (Fig. 1.5), along with an upregulation of genes involved in chrysolaminarin synthesis, glucogenesis, and glycolysis (Figs. 1.3, 1.6). *C. atlanticus* has the genetic potential to utilize these substrates as a nutrient source using the numerous carbohydrate-active enzymes (CAZymes) encoded in its genome [10]. We also observed an upregulation of multiple genes required for the biosynthesis of alanine, valine, leucine, isoleucine, and arginine, amino acids in *T. pseudonana* for which *C. atlanticus* is auxotrophic [10, 52]. Previous studies showed that the bacterium is able to use external amino acids as a sole nitrogen source [10, 52]. In plant rhizospheres, where the host provides a stable availability of select metabolites, bacterial auxotrophy for specific amino acids can provide a selective fitness advantage by allowing them to conserve the energy required to synthesize these essential compounds [53]. The postulated destabilization of the silica cell wall and chitin structures could lead to an increased permeability of the cell wall, potentially exacerbating the leakage of metabolites from *T. pseudonana* that can support bacterial growth. Therefore, by producing compounds that increase diatom leakiness and production of carbohydrates and essential amino acids, *C. atlanticus* may create a beneficial nutrient rich micro-environment. Leakage-facilitated metabolic exchanges among phytoplankton were described in the Black Queen Hypothesis, where one microorganism relies on a neighbor to produce essential components and thereby conserves its own metabolic potential [54, 55].

Diatoms can select and modulate the composition of their microbiome through the excretion of metabolites [18,19,20]. The antagonistic bacterium described here may be maintained within a diatom microbiome due to the heterogeneity of the diatom's response at the individual cell level. In our experiments in which *T. pseudonana* cultures are exposed to 33% bacterial filtrate, both flow cytometry and microscopy observations indicate that some of the cells exhibit aberrant morphology while other cells appear healthy. This morphological

heterogeneity can be the result of different metabolic states of the cells, difference in the exact amount of bacterial metabolites each cell absorbed, or a random 'bet hedging' strategy [56,57,58,59]. Additionally, antagonistic bacteria can be maintained with the diatom through regulation by other bacteria [60,61,62]. *C. atlanticus* was originally isolated from a consortium of bacteria from the diatom *P. multiseries*, and co-culture of live bacteria induced growth inhibition on a wide range of phylogenetically diverse axenic diatoms [10]. Our bacteria-free system mimicked the overall physiological response of *T. pseudonana* co-cultured with live *C. atlanticus*. *T. pseudonana* response to the bacterial filtrate is dose-dependent, with high concentrations being toxic and low concentrations stimulating fluorescence. The increased relative in vivo chlorophyll a fluorescence at low concentrations of filtrate (< 13% v/v) suggests either an increase in in vivo fluorescence per cell or an increase in the final concentration of cells, perhaps due to hormesis, an apparent increase of viability likely due to enhanced reparative processes due to stress [63]. In contrast to the direct association with live bacteria, we did not observe significant induction of multiple nuclei within individual cells. These differences between the diatom response to live bacterial co-culture and bacteria-free exudates can be a result of several factors, for example, bacterial attachment or close proximity to the diatom can increase exudate concentrations in the diatom phycosphere. Additionally, a yet unexplored diatom-bacteria metabolic crosstalk may alter the pathogenicity of the bacteria and its exudates.

The interactions between phytoplankton and bacteria are dynamic and can change from beneficial to algicidal based on chemical cues released by the host [64]. Although active metabolites released by *C. atlanticus* were produced in the absence of previous exposure to the diatom, the bacteria were grown on a rich media, which may have simulated a nutrient-rich phycosphere environment. Nonetheless, our study indicates that diatom-specific cues are not required for the production of algicidal compounds by *C. atlanticus*. Future work warrants further investigation of how the presence of phytoplankton and phytoplankton-produced compounds impacts the production of algicides by *C. atlanticus*.

Diverse bacterially produced compounds ranging from small molecules to enzymes have algicidal impacts on phytoplankton. For example, bacterial quorum-sensing molecules can inhibit diatom growth and metabolism [24, 25, 47]. Similarly, a flavobacterium produces L-amino acid oxidases, exuding hydrogen peroxide as a byproduct of amino acid oxidation, resulting in an indirect algicidal effect on green algae [64]. The active components of released *C. atlanticus* metabolites are currently unknown, however, the *C. atlanticus* genome encodes cyclomodulins such as gamma-glutamyl transpeptidases that are known to affect the cell cycle of various eukaryotic cells [65, 66]. This study opens the way for further investigations into the isolation and characteristics of bioactive components produced by a ubiquitous marine bacterium. Flavobacteria are known to play important roles in organic matter degradation during phytoplankton blooms [67] and elucidating the effects of bacterially produced metabolites on diatoms has important implications for understanding bloom dynamics and nutrient flux in the ocean.

Figures

Figure 1.1

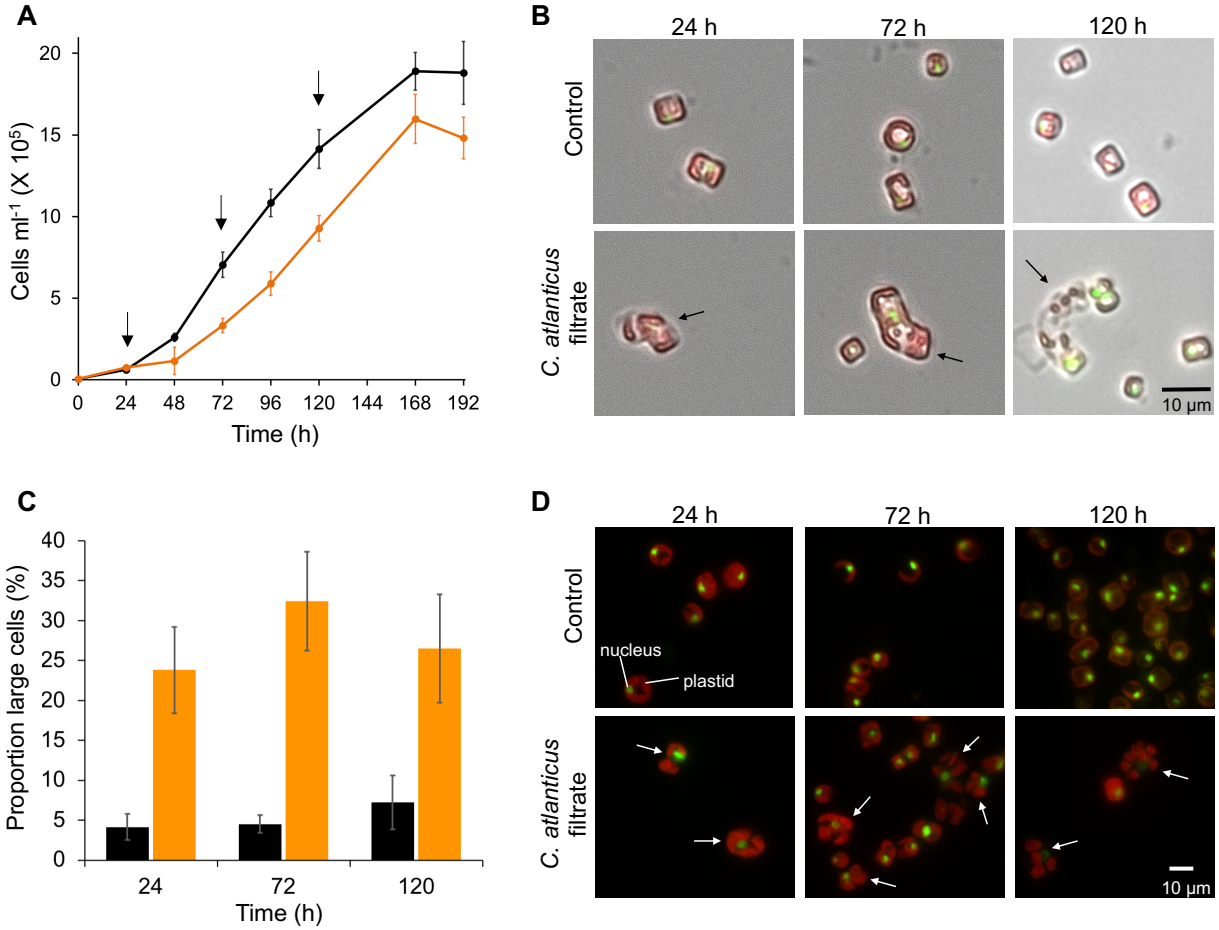


Fig. 1.1. Effect of *C. atlanticus* filtrate on growth and morphology of *T. pseudonana*. (A) Growth of *T. pseudonana* cultured with 33% v/v *C. atlanticus* filtrate (orange) or with 33% v/v control bacterial media (black). Black arrows indicate sampling times for transcriptome analysis. Cell counts at 0 h were not directly measured but calculated from dilution of cultures used to initialize the experiment. **(B)** Representative brightfield images of *T. pseudonana* morphology grown under control conditions (top row) or in *C. atlanticus* filtrate-treated (bottom row) conditions over the course of the experiment. The three time points corresponding to transcriptome sampling time points are shown. Brightfield images show cell size and shape, overlapped with fluorescence images of SYBR Green stained DNA in diatom nuclei (green fluorescence) and chlorophyll fluorescence (red fluorescence). Black arrows indicate treated cells with abnormal morphology. **(C)** Percentage of control (black) and *C. atlanticus* filtrate treated (orange) *T. pseudonana* cells with a cell area 2 standard deviations larger than the mean cell area of the control population at each time point. Cell area measurements were obtained from brightfield microscopy images. **(D)** Representative epifluorescence images of *T. pseudonana* morphology grown under control conditions (top row) or in *C. atlanticus* filtrate-treated (bottom row) conditions over the course of the experiment. Green fluorescence represents SYBR Green stained DNA in the diatom nucleus, and red fluorescence represents chlorophyll fluorescence in plastids. White arrows indicate treated cells with abnormal morphology.

Figure 1.2

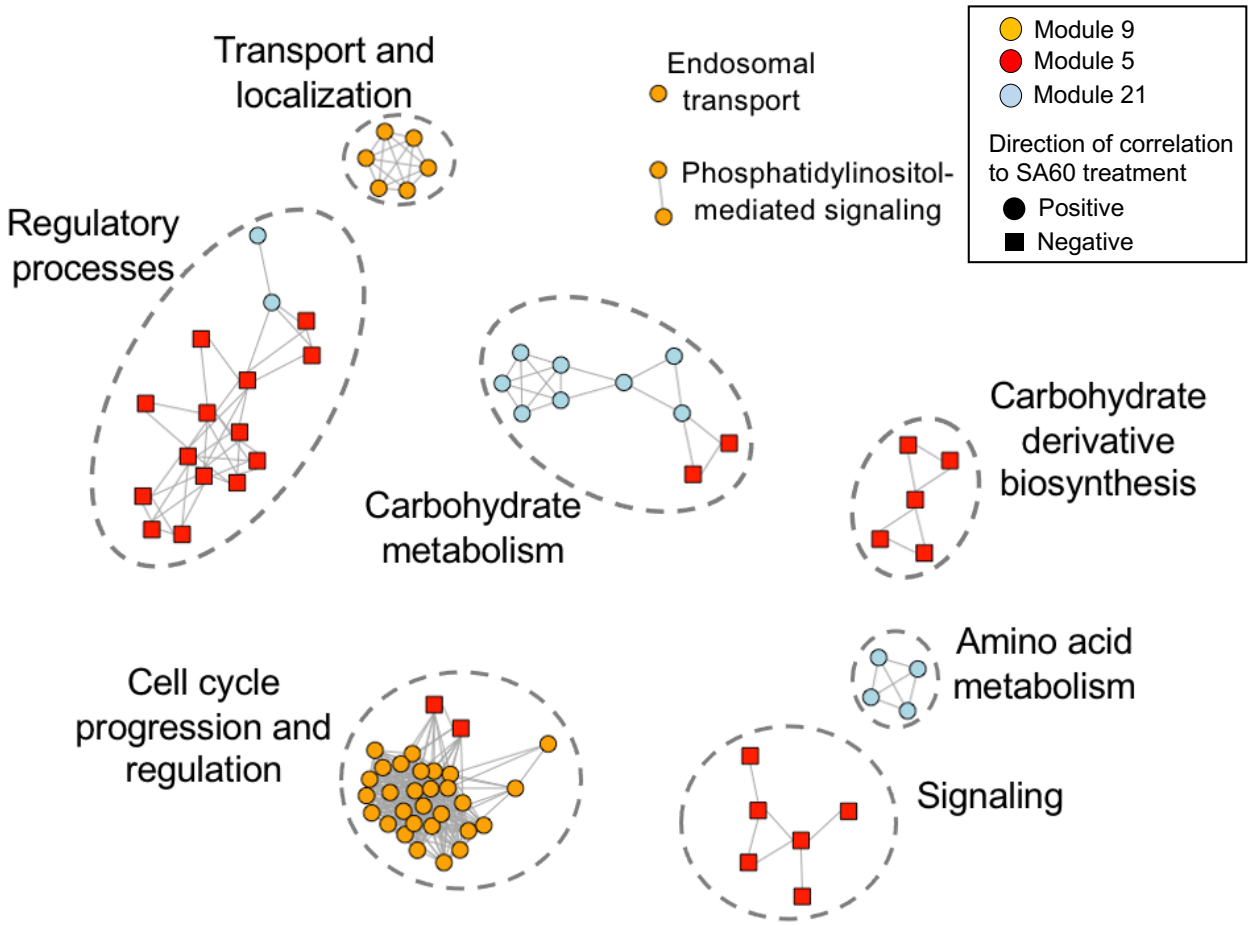


Fig. 1.2. Functional enrichment of genes belonging to *C. atlanticus* filtrate treatment related modules. Network of enriched GO terms in modules 9, 5, and 21 based on a Lin's similarity score cutoff of 0.7. Enrichment analysis was done on each module separately. Each node represents an enriched GO term while clusters represent GO terms of similar function. Enriched terms are colored by the module they belong to and shaped based on directionality of correlation to *C. atlanticus* filtrate treatment. The length of edges connecting the nodes is arbitrary.

Figure 1.3

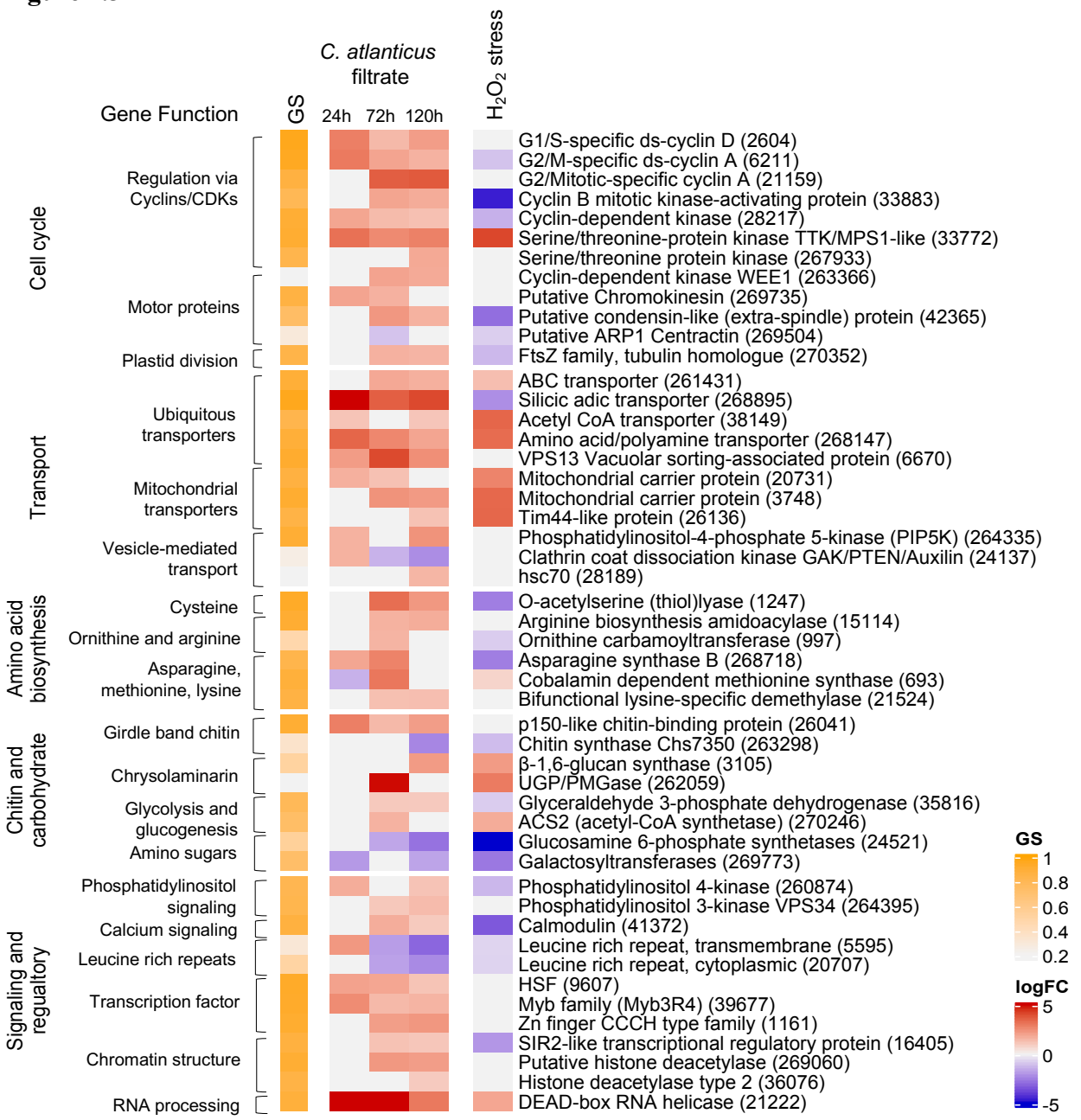


Fig. 1.3. Transcriptional patterns of selected *T. pseudonana* genes in response to *C. atlanticus* filtrate treatment and exposure to oxidative stress. Representative genes with functions in cell cycle progression, transport, amino acid biosynthesis, chitin and carbohydrate metabolism, and signaling and regulatory processes are shown. Columns from left to right: gene significance (GS) for *C. atlanticus* filtrate, differential expression (logFC) of genes in response to *C. atlanticus* filtrate treatment at 24, 72, and 120 h after exposure as compared to control cultures, and differential expression of *T. pseudonana* after 2 h of H_2O_2-induced oxidative stress compared to control conditions (see Methods). Labels on the right indicate encoded protein, with gene ID in parentheses.

Figure 1.4

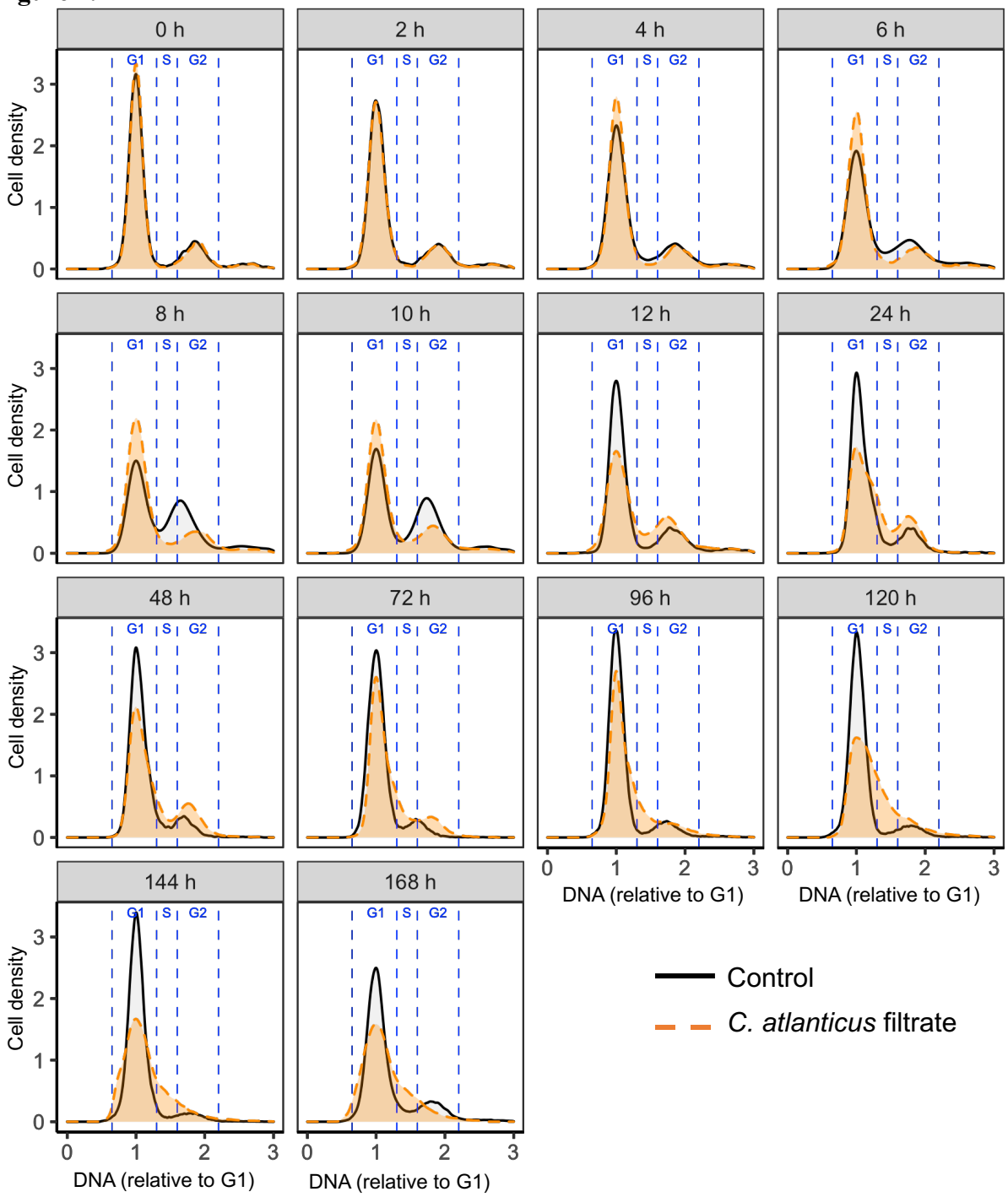


Fig. 1.4. Cell cycle analysis via flow cytometry in control (black) and *C. atlanticus* filtrate treated *T. pseudonana* cells (orange). DNA distributions of representative cytograms are shown (one representative replicate out of three is shown here), with distributions normalized to 1 µm beads and aligned based on the mode of G1 peaks. Each plot represents samples taken at indicated times after initial treatment with the filtrate. Blue dashed lines indicate boundaries of different cell cycle stages (G1, S, G2) and were drawn based on the signal of the control cultures.

Figure 1.5

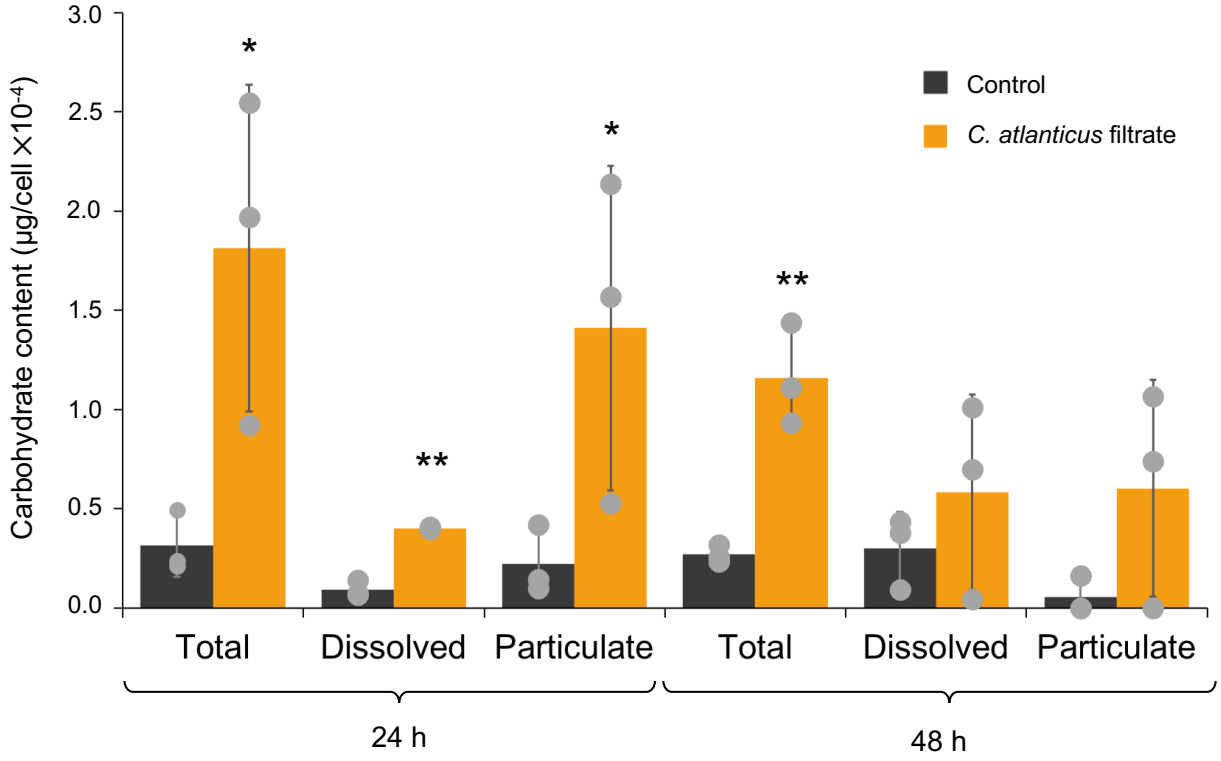


Fig. 1.5. Carbohydrate content in control (black) and *C. atlanticus* filtrate treated (orange) diatom cultures. Gray dots represent individual experimental replicates, stars indicate significant difference between control and filtrate treated cultures at each time point by students t-test (*$p < 0.05$, **$p < 0.01$). Total, dissolved, and particulate carbohydrate measurements were derived from bulk volume measurements normalized to cell concentration to get a carbohydrate content per cell. Particulate carbohydrate content was calculated as the difference between total and dissolved fractions. In filtrate treated samples, carbohydrate content of pure bacterial filtrate was subtracted from sample measurements to estimate carbohydrates produced by diatoms.

Figure 1.6

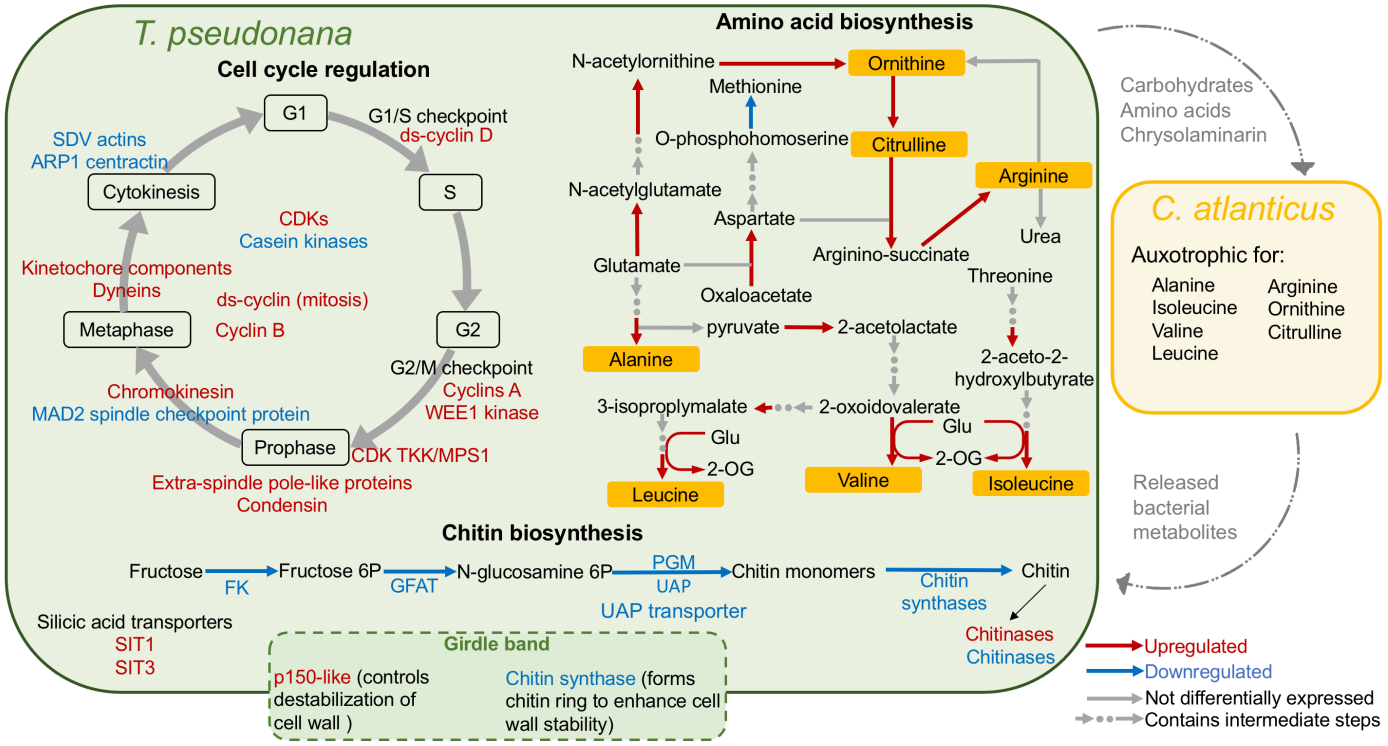
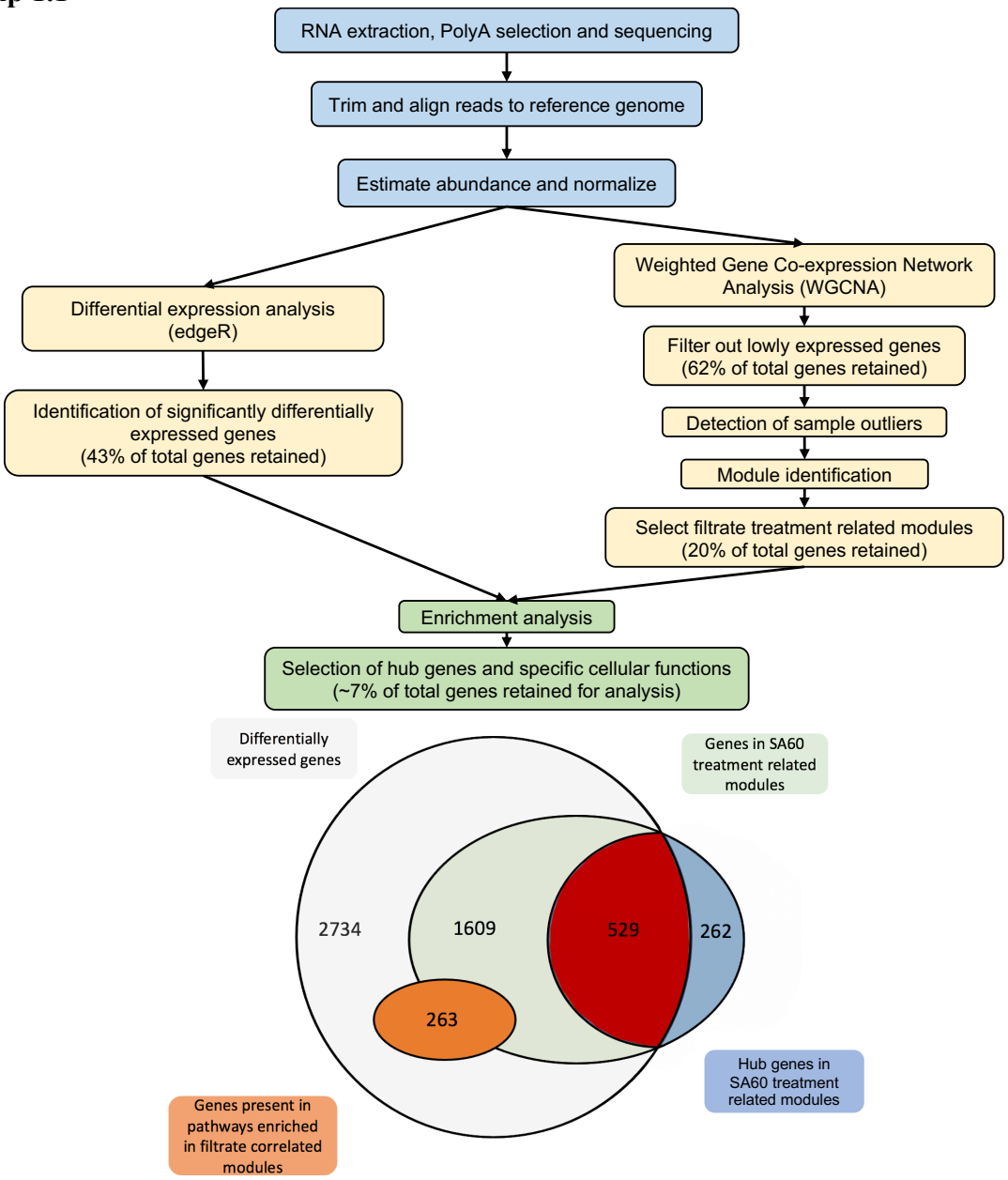


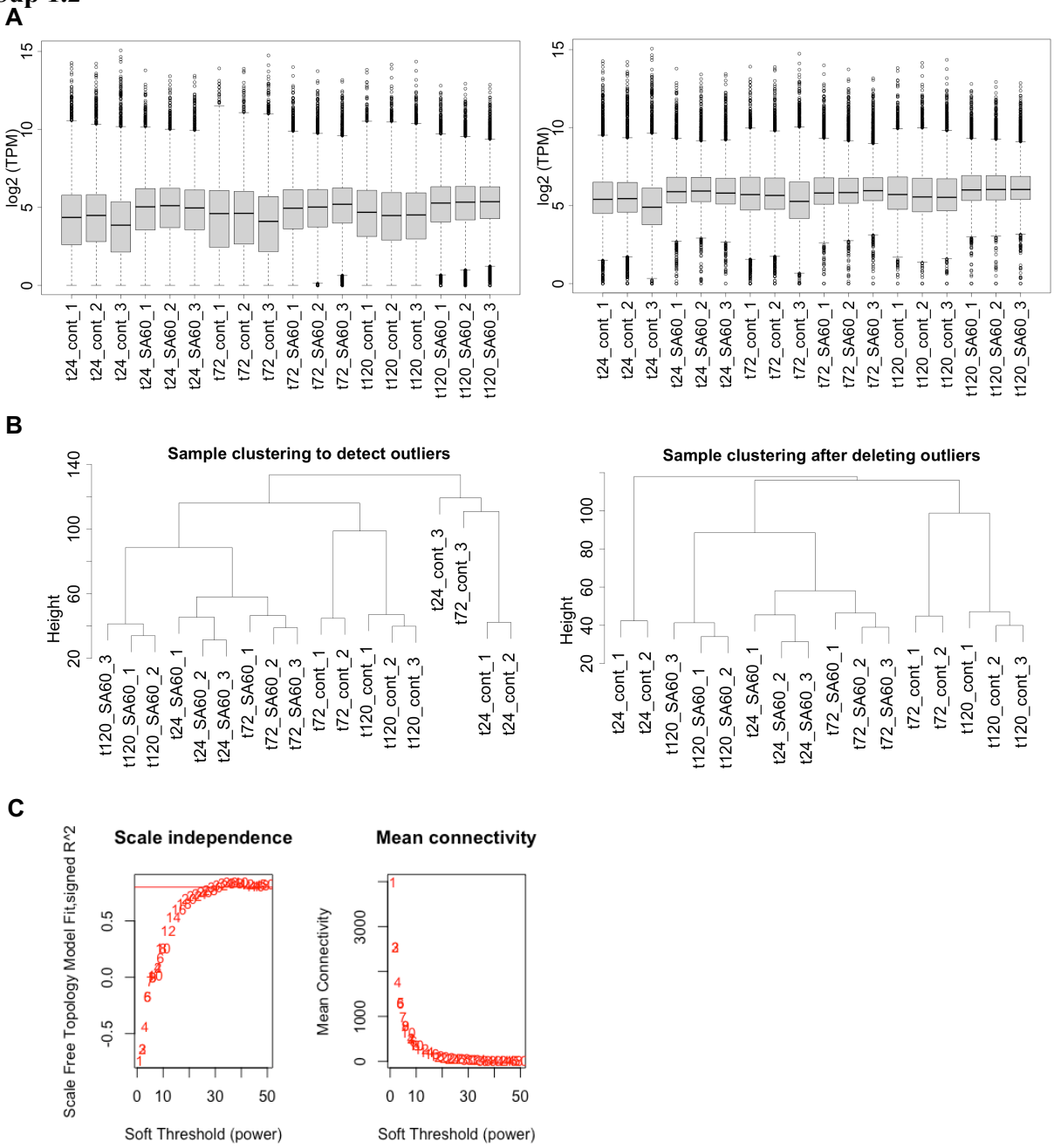
Fig. 1.6. Summary of *T. pseudonana* transcriptional changes in response to *C. atlanticus* filtrate. The green box represents *T. pseudonana*, the yellow box represents *C. atlanticus*. Upregulated (red) and downregulated (blue) *T. pseudonana* genes involved in cell cycle progression, amino acid metabolism, and chitin biosynthesis are shown. Colored letters and arrows represent up or downregulated genes, grey arrows represent genes that are not differentially expressed. Arrows with dots indicate intermediate steps not shown in the pathway. *C. atlanticus* auxotrophic amino acid residues are derived from van Tol et al., 2017[10], and potential interacting metabolites are indicated with the curved grey arrows

Supplemental Figures and Tables
Sup 1.1



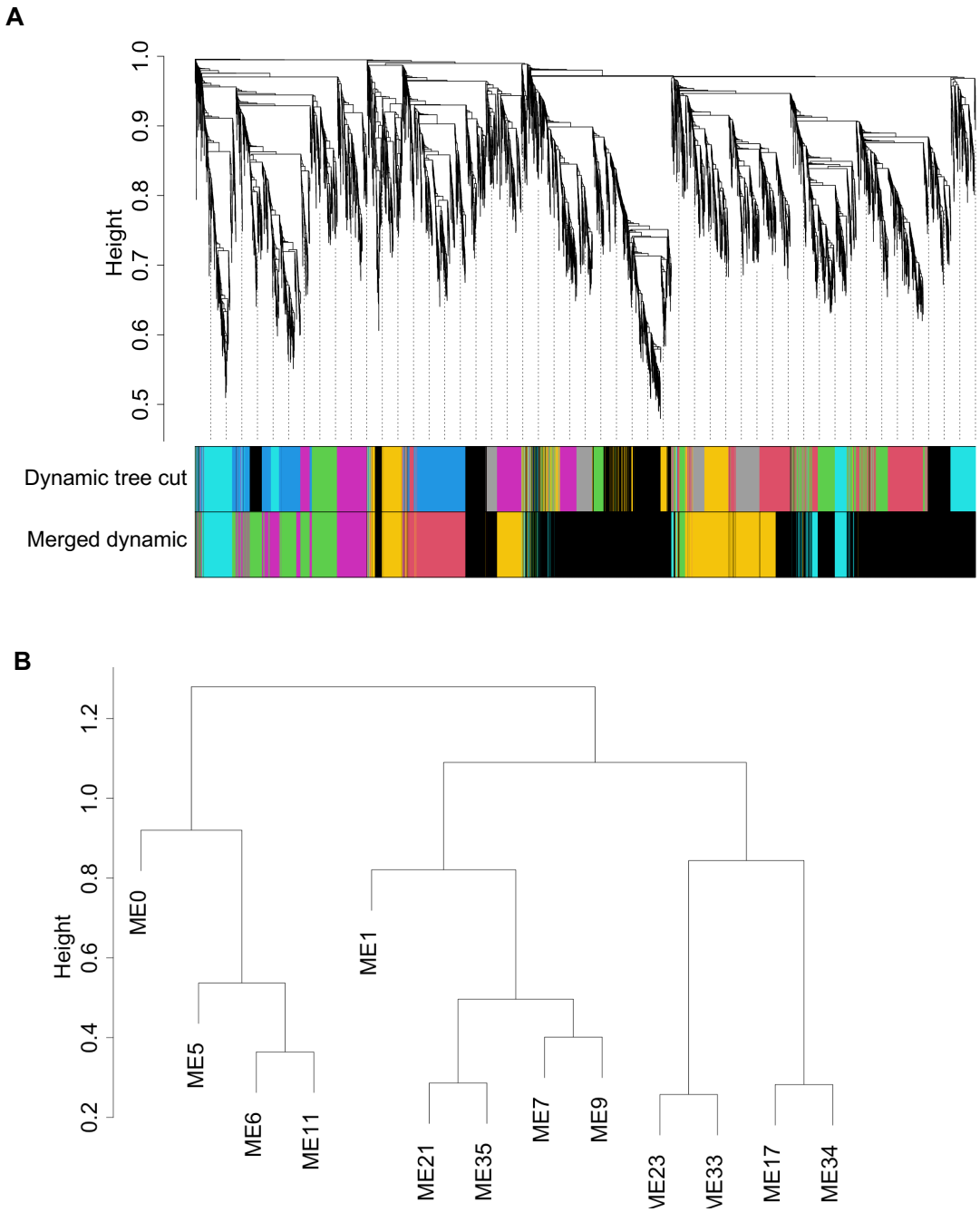
Sup. 1.1. Flowchart of statistical approach used for analyzing transcriptome data. Sample pre-processing (blue boxes), statistical analysis using WGNA clustering and differential expression analysis (yellow boxes), and selection of interesting genes (green boxes) is shown. Euler diagram illustrates the selection of genes considered for detailed analysis. An intersection of genes that were significantly differentially expressed in the *C. atlanticus* filtrate treated samples compared to the control in one or more of the time points sampled and hub genes identified from the four WCGNA modules were considered (red), along with a subset of genes present in pathways that were enriched in filtrate correlated WGCNA modules (orange). In total 792 genes were considered in detail for further analysis, indicated by the red and orange sections.

Sup 1.2



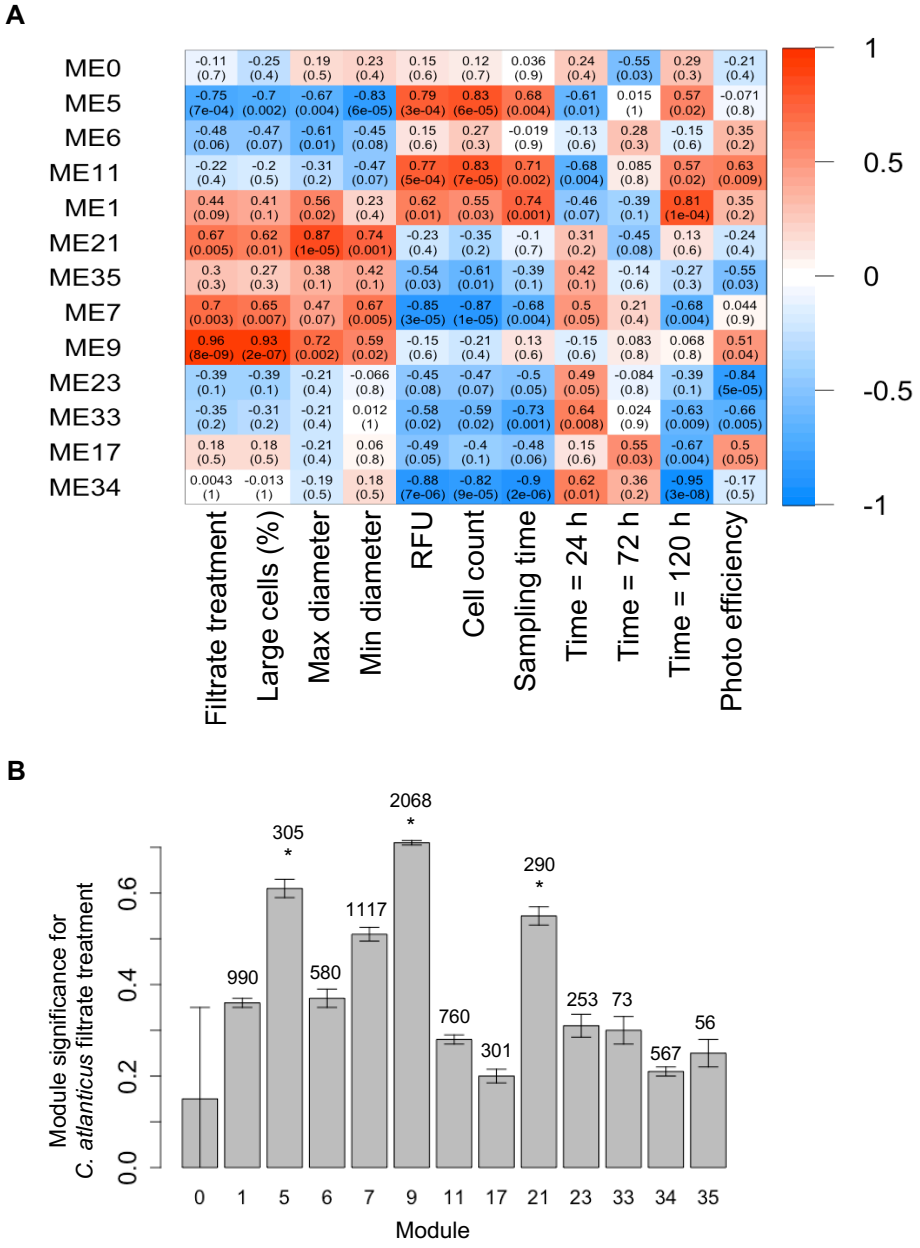
Sup. 1.2. WGCNA sample pre-processing and module construction metrics. (A) Bar plots of log2 TPM distributions in each of the collected samples before pre-processing (left) and after removing genes with expression levels below the median TPM in greater than 25% of the samples (right). Median TPM was calculated across all genes and samples. (B) Sample clustering dendrograms before (left) and after (right) removal of two outlier samples. (C) Scale-free topology and mean connectivity parameters used in network construction. A soft threshold power of 18 was used for network construction. The red line in the left panel represents the 0.8 cutoff value of the scale-free topology.

Sup 1.3



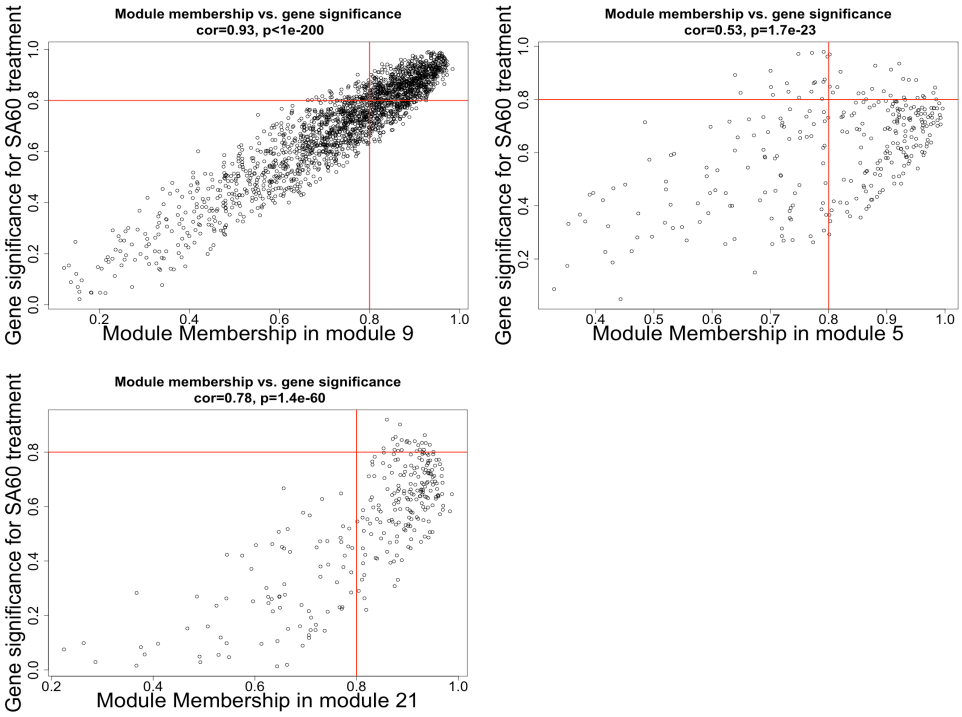
Sup. 1.3. WGCNA network and module construction. (A) Dendrogram of all genes clustered via hierarchical clustering according to a dissimilarity measure (1-TOM), and the corresponding module colors. The top row of module colors indicates module assignments using a dynamic tree cut method, while the bottom row indicates module assignments after similar modules have been merged based on a module dissimilarity threshold height of 0.25. **(B)** One-dimensional clustering of merged modules according to module eigengenes, which correspond to the principal component of each module. The ME0 module captures genes that do not show strong clustering with genes in any other module.

Sup 1.4



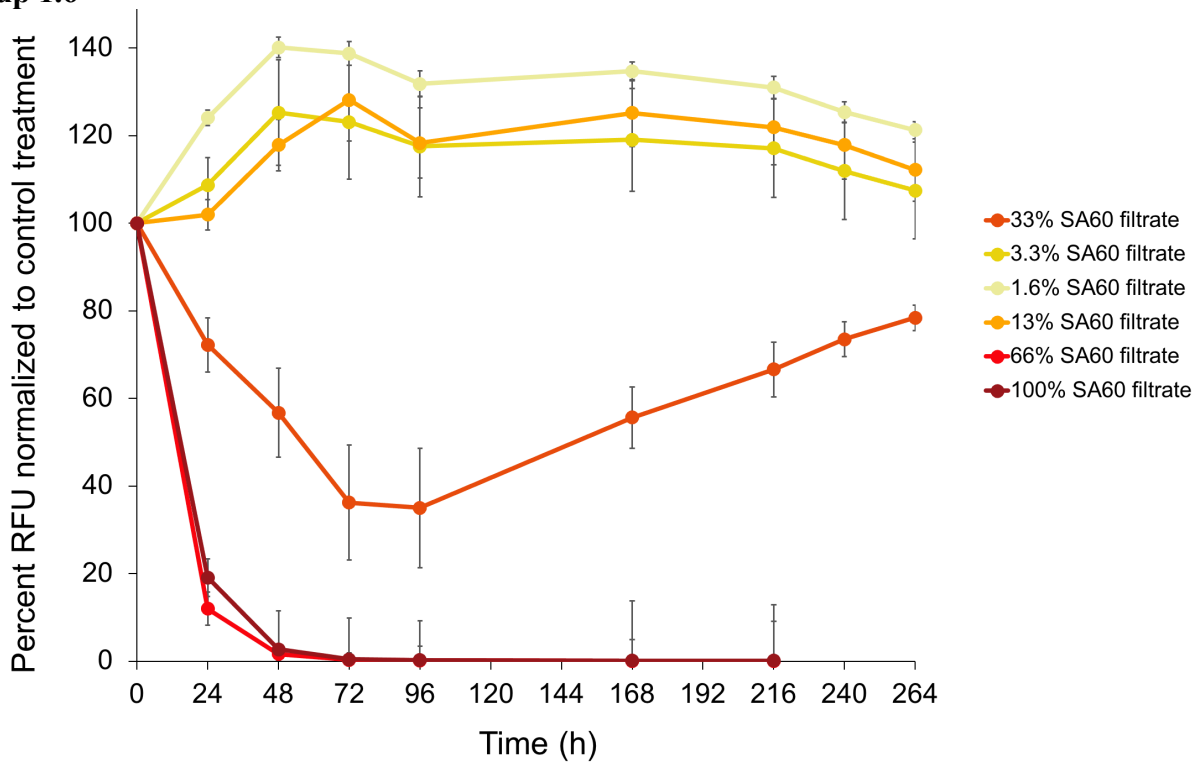
Sup. 1.4. Correlation of WGCNA generated gene modules to *C. atlanticus* filtrate treatment and other experimental traits. (A) Heatmap of the correlation (Pearson) between the experimental traits (columns) and WGCNA-derived module eigengenes (MEs) (rows). Within each cell, upper number is the correlations of the corresponding MEs and traits; *p* values from the linear mixed-effects model in parentheses. Red indicates positive correlations, blue indicates negative correlations, according to the color legend. RFU indicates relative fluorescence units corresponding to Chlorophyll *a* content. Cell size parameters are from microscopy analysis. **(B)** Mean gene significance of each module to treatment with *C. atlanticus* filtrate. Error bars represent standard deviation across all genes in a module. Stars indicate modules with greatest significance to treatment; numbers indicate number of genes in each module.

Sup 1.5



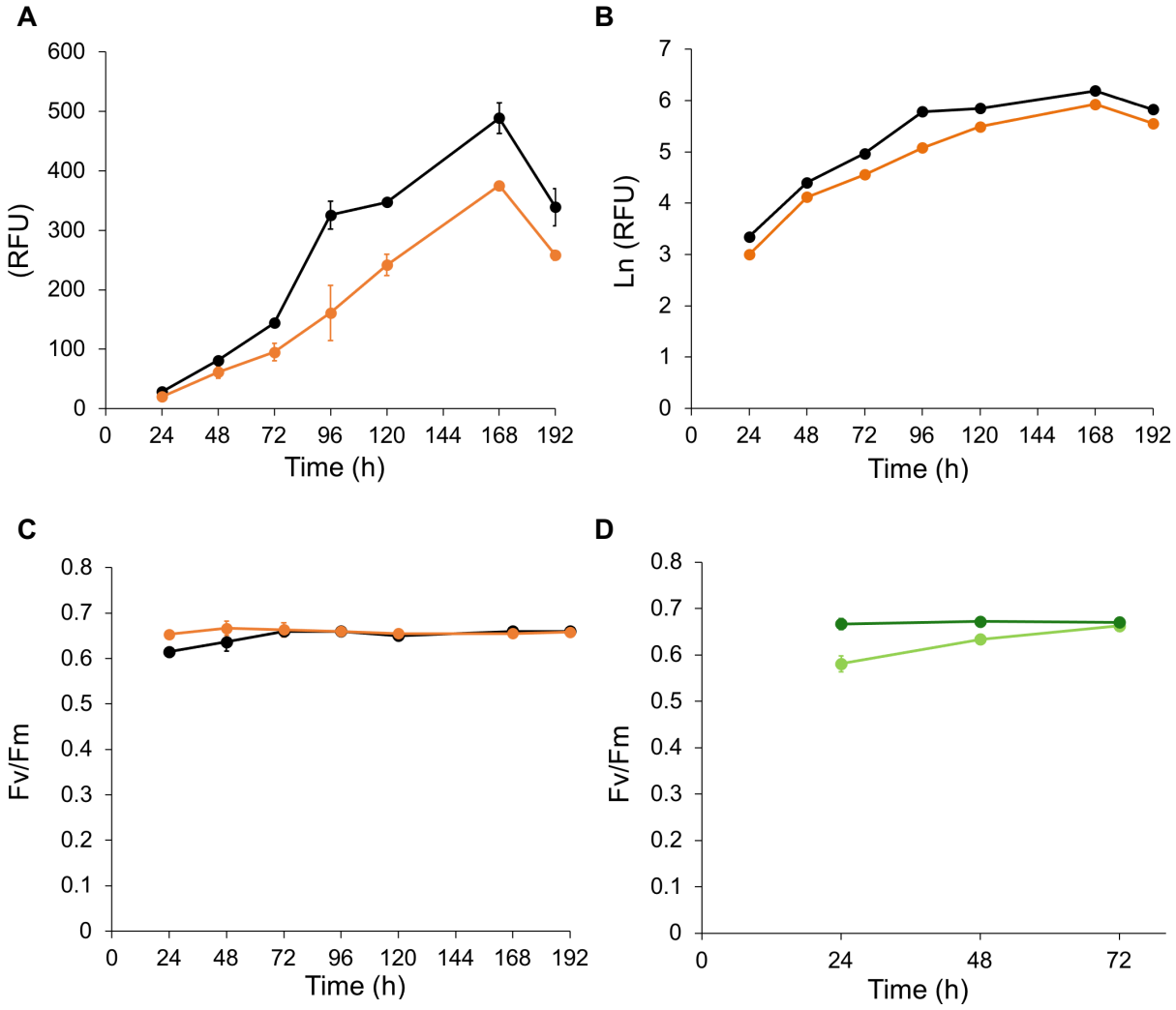
Sup. 1.5. Selection of hub genes that are differentially expressed and correlated to the *C. atlanticus* filtrate treatment. Correlation between gene significance for *C. atlanticus* treatment traits and module membership in module 9, 5 and 21. A hub gene is defined as a gene in one of these modules with both a gene significance > 0.8 and a module membership > 0.8 (red lines).

Sup 1.6



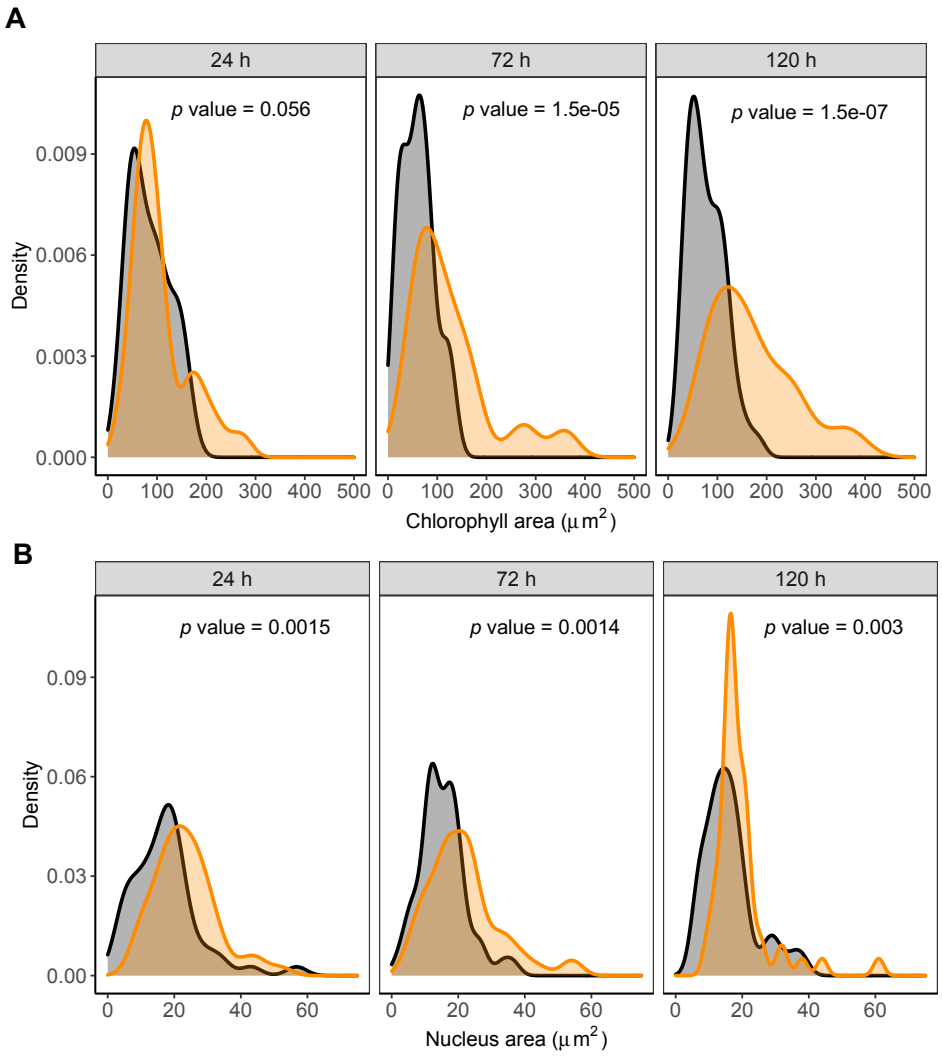
Sup. 1.6. Dose-dependent relative fluorescence ratios of *T. pseudonana* in response to *C. atlanticus* filtrates. Relative chlorophyll fluorescence units (RFU) in response to different concentrations of *C. atlanticus* filtrate. Fluorescence under varying concentrations of *C. atlanticus* filtrate is normalized to fluorescence under control conditions. The mean of three replicate cultures is shown, error bars represent standard deviations.

Sup 1.7



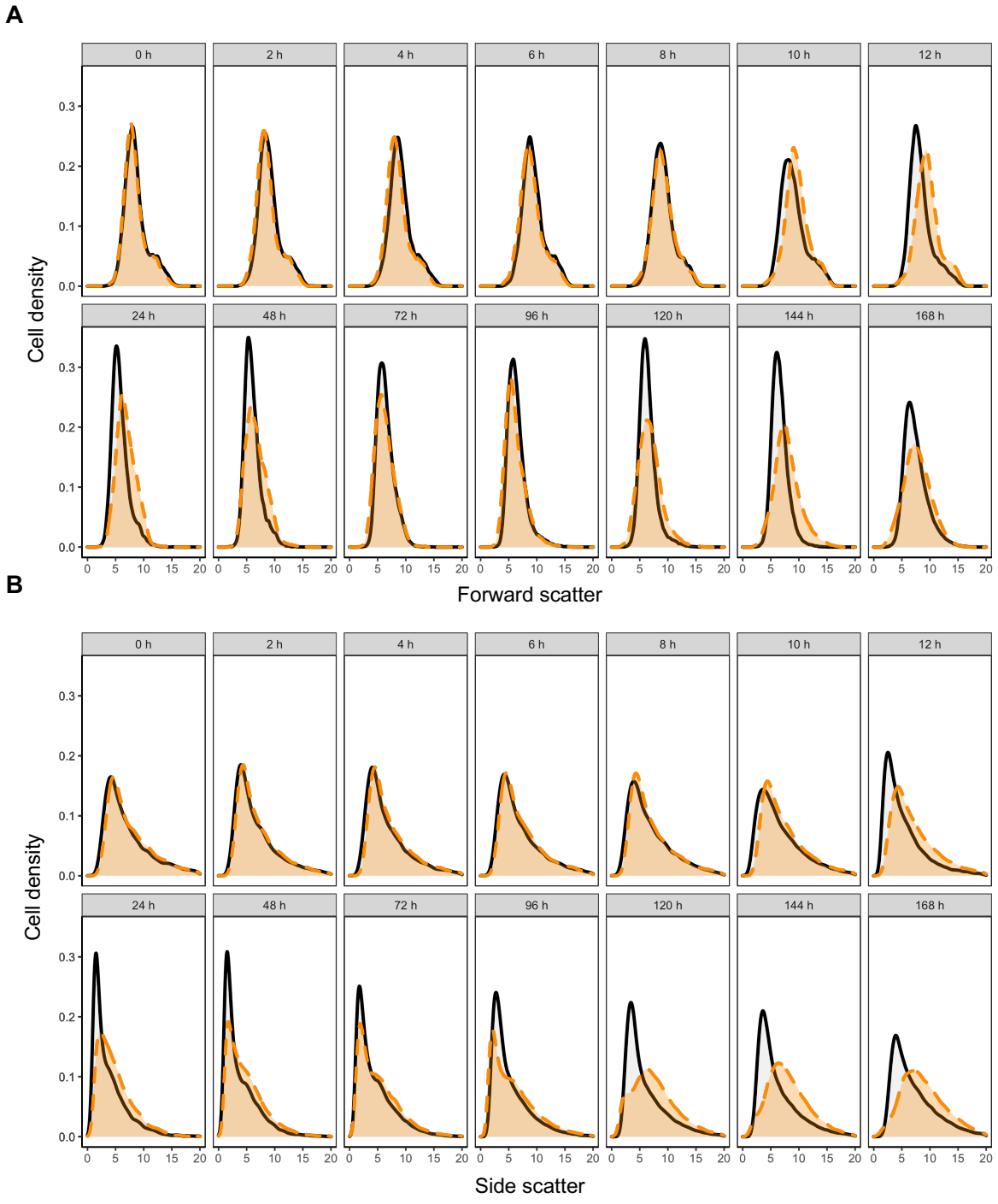
Sup. 1.7. Additional measurements of control (black) and filtrate treated (orange) diatom cultures. (A) Relative chlorophyll *a* fluorescence of *T. pseudonana* in response to 33% v/v *C. atlanticus* filtrate measured in relative fluorescence units (RFU). **(B)** RFU plotted on a logarithmic scale, with specific growth rates calculated between 24 h and 96 h of diatom growth. **(C)** Photosynthetic efficiency (Fv/Fm) of control samples and samples treated with 33% v/v *C. atlanticus* filtrate over the growth curve. (**D**) Fv/Fm of only control cultures measured during an independent experiment where control cultures were diluted to 5000 cells/ml (light green) compared to 5x104 cells/ml (dark green). Error bars represent the standard deviation of triplicate cultures in all panels.

Sup 1.8



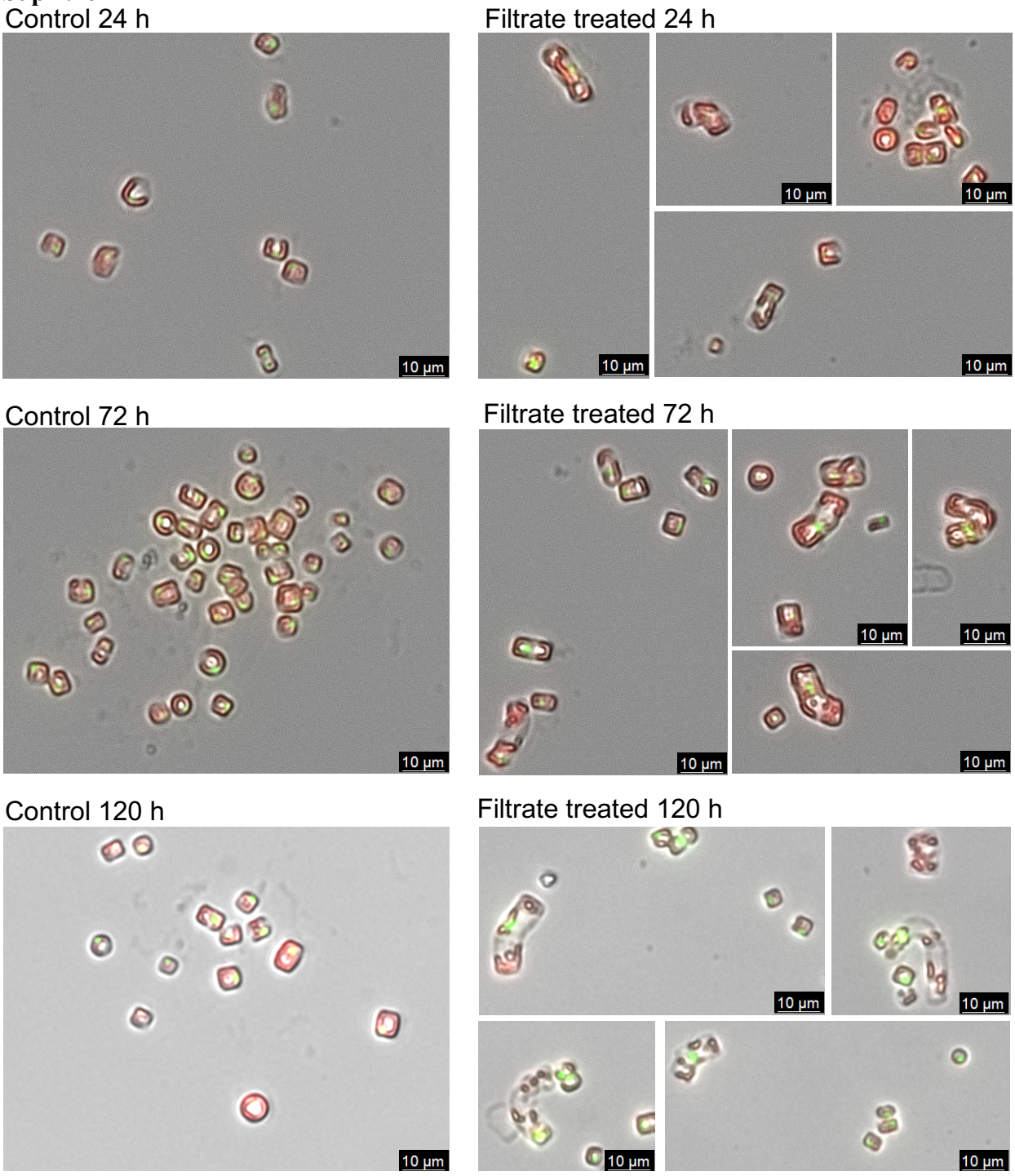
Sup. 1.8. Microscopy measurements of control (black) and *C. atlanticus* filtrate treated (orange) *T. pseudonana* cells. (A, B) Distributions of chlorophyll area (A) and nucleus area (B) per cell in control and *C. atlanticus* filtrate treated cells across three time points. Kolmogorov-Smirnov test was used to test for significance ($p<0.05$) between control and *C. atlanticus* filtrate treated populations at each time point.

Sup 1.9



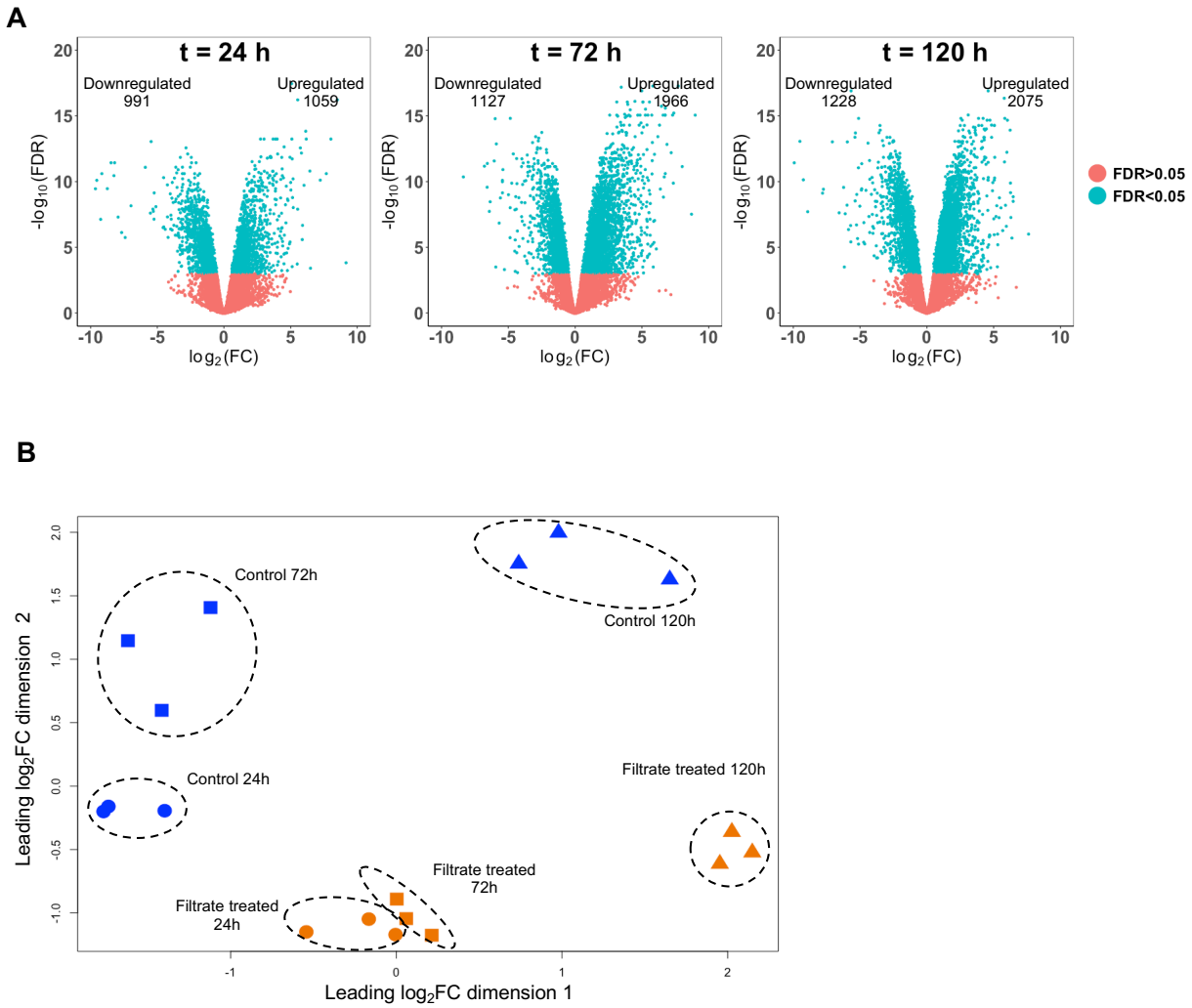
Sup. 1.9. Cell size distributions from flow cytometry data of control (black) and *C. atlanticus* filtrate treated (orange) populations. (A) Distributions of forward scatter as a proxy for cell size. **(B)** Distributions of side scatter can indicate changes in cell shape.

Sup 1.10



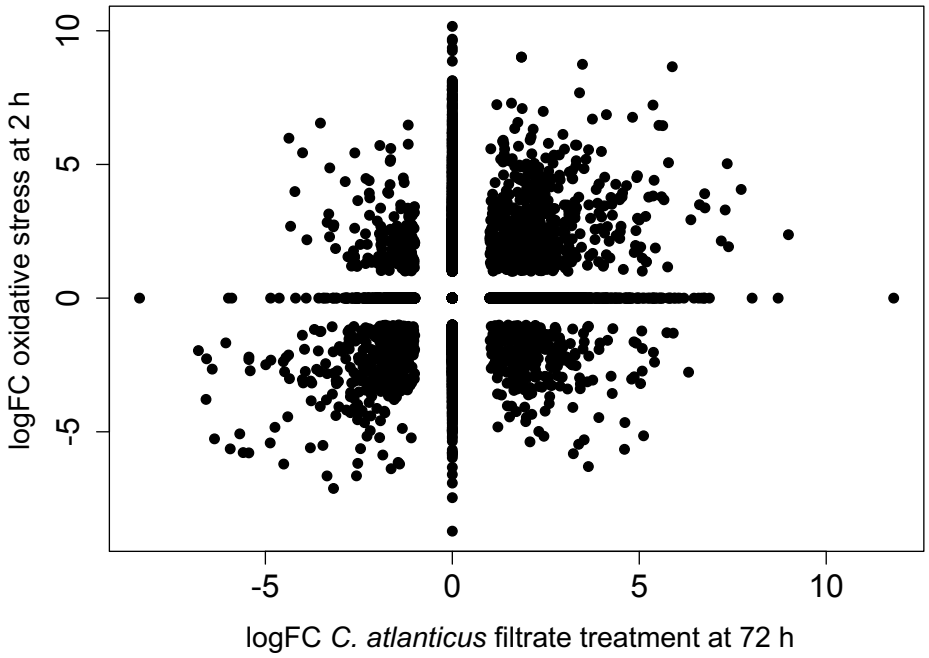
Sup. 1.10: Additional microscopy images of control (left column) and *C. atlanticus* filtrate treated cells (right columns) at three time points. Brightfield images show cell size and shape, overlapped with fluorescence images of SYBR Green stained DNA in diatom nuclei (green fluorescence) and chlorophyll fluorescence (red fluorescence).

Sup 1.11



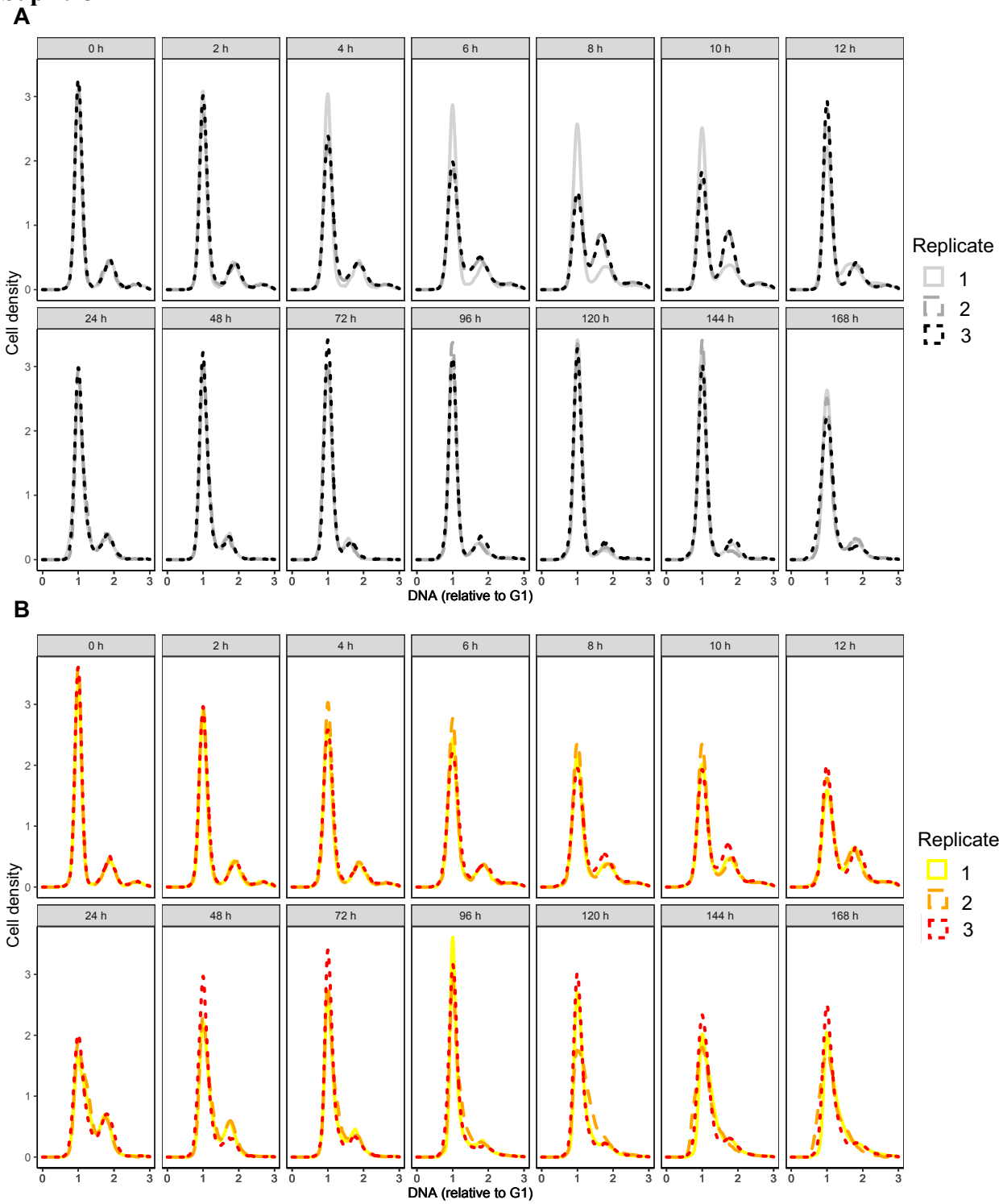
Sup. 1.11. Differential expression analysis of transcriptome data (A) Volcano plots of RNA-seq data for the three sampled time points. Log2 fold changes are plotted against log10 False discover rate (FDR) for *C. atlanticus* filtrate treated samples when compared to the control at all three time points. Significantly differentially expressed genes based on a 5% FDR level are highlighted in blue. The number of significantly up- and down-regulated genes is signified on the plots, where significant differential expression is defined as genes that have $p < 0.01$, FDR < 0.05 and |log2 fold change| > 1 in the treatment compared to control samples. **(B)** The Mean Square Deviation (MSD) of the 500 most differentially expressed genes. Different symbols represent samples taken at t = 24 h (circles), t = 72 h (squares) or and t = 120h (triangles) for either *C. atlanticus* filtrate treatment (orange) or control (blue). Distances between samples represent the log2 fold change between samples, based on the top 500 most discriminative genes.

Sup 1.12



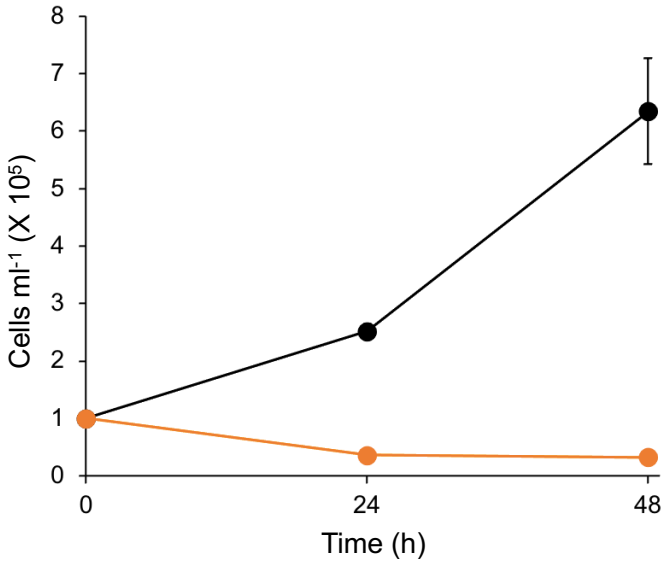
Sup. 1.12. Comparison of *T. pseudonana* transcriptional profiles under *C. atlanticus* filtrate treatment and H2O2-induced oxidative stress. Differential expression of *C. atlanticus* filtrate treated after cells after 72 h compared to differential expression 2 h after addition of 200 μM H2O2-induced oxidative stress.

Sup 1.13



Sup. 1.13: Cell cycle analysis via flow cytometry in control (A) and *C. atlanticus* filtrate treated *T. pseudonana* cells (B). DNA distributions of all cytograms are shown, with distributions normalized to 1 µm beads and aligned based on G1 peaks. For each treatment, all three replicates are plotted to show consistency.

Sup. 1.14



Sup. 1.14: Cell counts of the carbohydrate content phenol-sulfuric acid assay experiment of control (black) and *C. atlanticus* filtrate treated (orange) diatom cells. *T. pseudonana* cells were treated with undiluted bacterial filtrate. At 24 h, cells are still alive, but after 48 h, the undiluted bacterial filtrate treatment becomes lethal.

Sup. Table 1.1. Enriched Gene Ontology (GO) terms in WCGNA modules and differentially expressed genes at each of the three time points. Enrichment was done on GO Biological Process (BP), Molecular Function (MF), and Cellular Component (CC) terms, as well as on KEGG pathways on genes belonging to every module and differentially expressed genes at 24 h, 72 h and 120 h after treatment with *C. atlanticus* filtrate. Only significantly enriched terms with Fisher's test p-value < 0.01 are shown.

Sup. Table 1.2. Gene significance and module membership for genes in all WGCNA modules. Gene significance (GS), p-value for gene significance (p.GS), module membership (MM) and p-value for module membership (p.MM) for each gene in relation to each measured experimental trait and module.

Sup. Table 1.3. Expression patterns of *T. pseudonana* in response to *C. atlanticus* filtrate. Differential expression for three time points and gene significance for *C. atlanticus* filtrate treatment for hub genes with functional annotations, as well as for additional genes involved in relevant pathways and functions that were differentially expressed but were not identified as hub genes.

References

1. Falkowski PG. The role of phytoplankton photosynthesis in global biogeochemical cycles. Photosynth. Res. 1994;39(3):235-58.
2. Field CB, Behrenfeld MJ, Randerson JT, Falkowski P. Primary production of the biosphere: integrating terrestrial and oceanic components. Science. 1998;281(5374):237.
3. Amin SA, Parker MS, Armbrust EV. Interactions between diatoms and bacteria. Microbiol. Mol. Biol. Rev. 2012;76(3):667-84.
4. Bell W, Mitchell R. Chemotactic and growth response of marine bacteria to algal extracellular products. Biol. Bull. 1972;143(2):265-77.
5. Seymour JR, Amin SA, Raina J-B, Stocker R. Zooming in on the phycosphere: the ecological interface for phytoplankton–bacteria relationships. Nat. Microbiol. 2017;2:17065.
6. Azam F, Malfatti F. Microbial structuring of marine ecosystems. Nat. Rev. Microbiol. 2007;5(10):782-91.
7. Meyer N, Bigalke A, Kaulfuß A, Pohnert G. Strategies and ecological roles of algicidal bacteria. FEMS Microbiol. Rev. 2017;41(6):880-99.
8. Windler M, Bova D, Kryvenda A, Straile D, Gruber A, Kroth PG. Influence of bacteria on cell size development and morphology of cultivated diatoms. Phycol. Res. 2014;62(4):269-81.
9. Buhmann MT, Schulze B, Forderer A, Schleheck D, Kroth PG. Bacteria may induce the secretion of mucin-like proteins by the diatom *Phaeodactylum tricornutum*. J. Phycol. 2016;52(3):463-74.
10. van Tol HM, Amin SA, Armbrust EV. Ubiquitous marine bacterium inhibits diatom cell division. ISME J. 2017;11(1):31-42.
11. Amin SA, Hmelo LR, van Tol HM, Durham BP, Carlson LT, Heal KR, et al. Interaction and signaling between a cosmopolitan phytoplankton and associated bacteria. Nature. 2015;522(7554):98-101.
12. Durham BP, Sharma S, Luo H, Smith CB, Amin SA, Bender SJ, et al. Cryptic carbon and sulfur cycling between surface ocean plankton. PNAS. 2015;112(2):453-7.
13. Durham BP, Dearth SP, Sharma S, Amin SA, Smith CB, Campagna SR, et al. Recognition cascade and metabolite transfer in a marine bacteria-phytoplankton model system. Environ. Microbiol. 2017;19(9):3500-13.
14. Grossart H-P, Levold F, Allgaier M, Simon M, Brinkhoff T. Marine diatom species harbour distinct bacterial communities. Environ. Microbiol. 2005;7(6):860-73.
15. Crenn K, Duffieux D, Jeanthon C. Bacterial epibiotic communities of ubiquitous and abundant marine diatoms are distinct in short- and long-term associations. Front Microbiol. 2018;9:2879.
16. Behringer G, Ochsenkühn MA, Fei C, Fanning J, Koester JA, Amin SA. Bacterial communities of diatoms display strong conservation across strains and time. Front Microbiol. 2018;9:659.
17. Schäfer H, Abbas B, Witte H, Muyzer G. Genetic diversity of 'satellite' bacteria present in cultures of marine diatoms. FEMS Microbiol. Ecol. 2002;42(1):25-35.
18. Shibl AA, Isaac A, Ochsenkühn MA, Cárdenas A, Fei C, Behringer G, et al. Diatom modulation of select bacteria through use of two unique secondary metabolites. PNAS. 2020;117(44):27445-55.

19. Fu H, Uchimiya M, Gore J, Moran MA. Ecological drivers of bacterial community assembly in synthetic phycospheres. PNAS. 2020:201917265.
20. Stock W, Blommaert L, De Troch M, Mangelinckx S, Willems A, Vyverman W, et al. Host specificity in diatom-bacteria interactions alleviates antagonistic effects. FEMS Microbiol. Ecol. 2019;95(11).
21. Segev E, Wyche TP, Kim KH, Petersen J, Ellebrandt C, Vlamakis H, et al. Dynamic metabolic exchange governs a marine algal-bacterial interaction. Elife. 2016;5:e17473.
22. Wagner-Döbler I, Ballhausen B, Berger M, Brinkhoff T, Buchholz I, Bunk B, et al. The complete genome sequence of the algal symbiont *Dinoroseobacter shibae*: a hitchhiker's guide to life in the sea. ISME J. 2010;4(1):61-77.
23. Frank O, Michael V, Päuker O, Boedeker C, Jogler C, Rohde M, et al. Plasmid curing and the loss of grip – the 65-kb replicon of *Phaeobacter inhibens* DSM 17395 is required for biofilm formation, motility and the colonization of marine algae. Syst. Appl. Microbiol. 2015;38(2):120-7.
24. Paul C, Pohnert G. Interactions of the algicidal bacterium *Kordia algicida* with diatoms: regulated protease excretion for specific algal lysis. PLoS One. 2011;6(6):e21032.
25. Stock F, Bilcke G, De Decker S, Osuna-Cruz CM, Van den Berge K, Vancaester E, et al. distinctive growth and transcriptional changes of the diatom *Seminavis robusta* in response to quorum sensing related compounds. Front. Microbiol. 2020;11:1240.
26. Guillard RRL. Culture of Phytoplankton for Feeding Marine Invertebrates. In: Smith WL, Chanley MH, editors. Culture of marine invertebrate animals: proceedings — 1st conference on culture of marine invertebrate animals greenport. Boston, MA: Springer US; 1975. p. 29-60.
27. Rasband WS. (2016). ImageJ, U.S. National Institutes of Health, Bethesda, MD, USA. Available at: http://imagej. nih.gov/ij/, 1997–2015.
28. DuBois M, Gilles KA, Hamilton JK, Rebers PA, Smith F. Colorimetric method for determination of sugars and related substances. Anal. Chem. 1956;28(3):350-6.
29. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. BOINFP. 2014;30(15):2114-20.
30. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. BOINFP. 2010;26(1):139-40.
31. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinform. 2008;9(1):559.
32. Alexa A, Rahnenfuhrer J (2021). topGO: Enrichment analysis for gene ontology. R package version 2.46.0.
33. Csardi G, Nepusz T (2006). "The igraph software package for complex network research." InterJournal, Complex Systems, 1695. https://igraph.org.
34. Wei Q, Khan IK, Ding Z, Yerneni S, Kihara D. NaviGO: interactive tool for visualization and functional similarity and coherence analysis with gene ontology. BMC Bioinform. 2017;18(1):177.
35. Shapiro HM. (2003). Physical parameters and their uses. In: Shapiro HM (ed). Practical Flow Cytometry. John Wiley & Sons, Inc.: New York, NY, USA, pp. 273–285.
36. Clercq AD, Inzé D. Cyclin-dependent kinase inhibitors in yeast, animals, and plants: a functional comparison. Crit. Rev. Biochem. Mol. Biol. 2006;41(5):293-313.
37. Zinser ER. The microbial contribution to reactive oxygen species dynamics in marine ecosystems. Environ. Microbiol. Rep. 2018;10(4):412-27.

38. Whalen KE, Kirby C, Nicholson RM, O'Reilly M, Moore BS, Harvey EL. The chemical cue tetrabromopyrrole induces rapid cellular stress and mortality in phytoplankton. Sci. Rep. 2018;8(1):15498-.
39. Sheyn U, Rosenwasser S, Ben-Dor S, Porat Z, Vardi A. Modulation of host ROS metabolism is essential for viral infection of a bloom-forming coccolithophore in the ocean. ISME J. 2016;10(7):1742-54.
40. Finkel ZV, Irwin AJ, Schofield O. Resource limitation alters the ¾ size scaling of metabolic rates in phytoplankton. Mar. Ecol. Prog. Ser. 2004;273:269-80.
41. De Troch M, Chepurnov V, Gheerardyn H, Vanreusel A, Ólafsson E. Is diatom size selection by harpacticoid copepods related to grazer body size? J. Exp. Mar. Biol. Ecol. 2006;332(1):1-11.
42. Finkel ZV. Light absorption and size scaling of light-limited metabolism in marine diatoms. Limnol. Oceanogr. 2001;46(1):86-94.
43. Wilhelm T, Said M, Naim V. DNA replication stress and chromosomal instability: dangerous liaisons. Genes. 2020;11(6):642.
44. Gelot C, Magdalou I, Lopez BS. Replication stress in mammalian cells and its consequences for mitosis. Genes. 2015;6(2).
45. Vogt E, Kirsch-Volders M, Parry J, Eichenlaub-Ritter U. Spindle formation, chromosome segregation and the spindle checkpoint in mammalian oocytes and susceptibility to meiotic error. Mutat. Res. - Genet. Toxicol. Environ. Mutagen. 2008;651(1):14-29.
46. Van de Meene AML, Pickett-Heaps JD. Valve morphogenesis in the centric diatom *Rhizosolenia setigera* (*Bacillariophyceae, Centrales*) and its taxonomic implications. Eur. J. Phycol. 2004;39(1):93-104.
47. Pollara SB, Becker JW, Nunn BL, Boiteau R, Repeta D, Mudge MC, et al. Bacterial quorum-sensing signal arrests phytoplankton cell division and impacts virus-induced mortality. mSphere. 2021;6(3):e00009-21.
48. Von Dassow P, Petersen TW, Chepurnov VA, Virginia Armbrust E. Inter- and intraspecific relationships between nuclear DNA content and cell size in selected members of the centric diatom genus *Thalassiosira (Bacillariophyceae)*. J. Phycol. 2008;44(2):335-49.
49. Pokrzywinski KL, Tilney CL, Warner ME, Coyne KJ. Cell cycle arrest and biochemical changes accompanying cell death in harmful dinoflagellates following exposure to bacterial algicide IRI-160AA. Sci. Rep. 2017;7(1):45102.
50. Durkin CA, Mock T, Armbrust EV. Chitin in diatoms and its association with the cell wall. Eukaryot. Cell. 2009;8(7):1038.
51. Wildermuth MC. Modulation of host nuclear ploidy: a common plant biotroph mechanism. Curr. Opin. Plant Biol. 2010;13(4):449-58.
52. Cho J-C, Giovannoni SJ. Croceibacter atlanticus gen. nov., sp. nov., A Novel Marine Bacterium in the Family *Flavobacteriaceae*. Syst. Appl. Microbiol. 2003;26(1):76-83.
53. Trivedi P, Leach JE, Tringe SG, Sa T, Singh BK. Plant–microbiome interactions: from community assembly to plant health. Nat. Rev. Microbiol. 2020;18(11):607-21.
54. Morris JJ, Lenski RE, Zinser ER. The black queen hypothesis: evolution of dependencies through adaptive gene loss. MBio. 2012;3(2):e00036-12.
55. Ndhlovu A, Durand PM, Ramsey G. Programmed cell death as a black queen in microbial communities. Mol. Ecol. 2021;30(5):1110-9.

56. Schreiber F, Littmann S, Lavik G, Escrig S, Meibom A, Kuypers MMM, et al. Phenotypic heterogeneity driven by nutrient limitation promotes growth in fluctuating environments. Nat. Microbiol. 2016;1(6):16055.
57. Sengupta A, Carrara F, Stocker R. Phytoplankton can actively diversify their migration strategy in response to turbulent cues. Nature. 2017;543(7646):555-8.
58. Levy SF, Ziv N, Siegal ML. Bet hedging in yeast by heterogeneous, age-correlated expression of a stress protectant. PLoS Biol. 2012;10(5):e1001325.
59. Ackermann M. A functional perspective on phenotypic heterogeneity in microorganisms. Nat. Rev. Microbiol. 2015;13(8):497-508.
60. Blair PM, Land ML, Piatek MJ, Jacobson DA, Lu T-YS, Doktycz MJ, et al. Exploration of the biosynthetic potential of the populus microbiome. mSystems. 2018;3(5):e00045-18.
61. Helfrich EJN, Vogel CM, Ueoka R, Schäfer M, Ryffel F, Müller DB, et al. Bipartite interactions, antibiotic production and biosynthetic potential of the *Arabidopsis* leaf microbiome. Nat. Microbiol. 2018;3(8):909-19.
62. Long RA, Qureshi A, Faulkner DJ, Azam F. 2-n-Pentyl-4-quinolinol produced by a marine *Alteromonas* sp. and its potential ecological and biogeochemical roles. AEM. 2003;69(1):568-76.
63. Calabrese EJ. Hormesis: from mainstream to therapy. Cell Commun. Signal. 2014;8(4), 289–291.
64. Chen WM, Sheu FS, Sheu SY. Novel l-amino acid oxidase with algicidal activity against toxic cyanobacterium *Microcystis aeruginosa* synthesized by a bacterium *Aquimarina* sp. Enzyme Microb. Technol. 2011;49(4):372-9.
65. El-Aouar Filho RA, Nicolas A, De Paula Castro TL, Deplanche M, De Carvalho Azevedo VA, Goossens PL, et al. Heterogeneous family of cyclomodulins: smart weapons that allow bacteria to hijack the eukaryotic cell cycle and promote infections. Front. Cell. Infect. Microbiol. 2017;7.
66. Ricci V, Giannouli M, Romano M, Zarrilli R. *Helicobacter pylori* gamma-glutamyl transpeptidase and its pathogenic role. WJG. 2014;20(3):630-8.
67. Teeling H, Fuchs Bernhard M, Becher D, Klockow C, Gardebrecht A, Bennke Christin M, et al. Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. Science. 2012;336(6081):

Chapter 2: Functional patterns of microbial interaction in the North Pacific revealed through statistical modeling of environmental metatranscriptomes

Zinka Bartolek[1], Amy Willis[2], Sarah Teichman[2], E. Virginia Armbrust[1]

[1]School of Oceanography, University of Washington, Seattle, WA, 98195, USA
[2]Department of Biostatistics, University of Washington, Seattle, WA, 98195, USA

Abstract

Marine microorganisms form dynamic networks of interactions that structure food webs and drive global biogeochemical cycles. These interactions range from opportunism, to sustained symbioses to competition, predation and defense. Despite advances in lab experiments, co-occurrence analyses, environmental omics, and computational modeling approaches, the scope of molecular processes involved in microbial interactions remains difficult to observe directly in the environment. Here, we developed Enviromtx, a statistical approach that detects microbial interactions and their transcriptional signatures directly from environmental metatranscriptomes, without *a-priori* assumption of interacting organisms or mechanisms of interaction. We identified the most consistent transcriptional signatures of interactions across the North Pacific Ocean, and show that a small but significant portion of a species transcriptome is involved in an interaction response. The most common interactions occurred in cross-kingdom pairs, especially between diatoms and bacterial phyla such as Pseudomonadota and Bacteroidota, mediated by nutrient exchange, defense, secondary metabolites, and motility. We identified general functions of interactions occurring among many species including cross-feeding and defense, as well as species-specific responses occurring through secondary metabolite exchanges and metabolic regulation. We propose new candidate interactions between specific species-species pairs and highlight specific functions mediating the interactions. By detecting functional signatures of interactions from field metatranscriptomes, Enviromtx reveals broad patterns of species interactions in the North Pacific and generates hypotheses for targeted laboratory or modeling studies.

Introduction

Microorganisms in the ocean engage in complex, dynamic networks of interactions with global-scale consequences as these organisms form the base of marine food webs and influence nutrient cycling in the ocean. Central to these interactions are nutritional exchanges between phytoplankton and bacteria, where phytoplankton produce organic matter through photosynthesis that heterotrophic bacteria consume and remineralize. However, complex interactions occur between phytoplankton and bacteria as well as within bacterial and phytoplankton communities, and include opportunism, competition, predation and symbiosis (Falkowski et al., 2008; Azam & Malfatti, 2007). In the dilute ocean environment, phytoplankton and bacteria exist in close association to each other within the phycosphere, a diffusive boundary layer surrounding each phytoplankton cell, where nutrient concentrations can exceed background levels by orders of magnitude (Seymour et al., 2017), and chemical conditions including pH differ from surrounding seawater (Liu et al., 2022). Bacteria can sense these chemical gradients through chemotaxis, enabling them to remain transiently within the phycosphere or to attach directly to the phytoplankton cell (Raina et al., 2022). Within bacterial communities, bacteria engage in competition for nutrients, syntropy and cross-feeding (Seth & Taga, 2014). Some eukaryotes obtain nutrients through phagocytosis, and mixotrophs can adjust their metabolic strategy towards heterotrophy or phototrophy depending on nutrient and prey availability (Stoecker et al., 2017; Lambert et al., 2022). Numerous microorganisms engage in competition and defense through production of antimicrobial compounds and toxins (Agrawal et al., 2017), as well as communication, signaling and regulation of microbiome composition via quorum sensing (Coolahan & Whalen, 2025; Rolland et al., 2016; Pollara et al., 2021; Stock et al., 2020). Similarly, species engage in exchanges of scarce resources, for example siderophore-mediated iron exchanges and synthesis of amino acids and vitamins by some species are taken up by others that are auxotrophic for the compound (Amin et al., 2009; Gómez-Consarnau et al., 2018; Croft et al., 2005). Some interactions involve specific exchanges of specific metabolites suggesting long term co-evolution (King et al., 2022; Ahern et al., 2021), while others occur through opportunistic uptake of available organic matter and broad range defense mechanisms (Sarmento & Gasol, 2012; Sarmento et al., 2016).

Marine microbial interactions have been extensively studied through co-culture experiments with model organisms, computational models, and *in-situ* community level omics observations. For example, lab co-culture experiments have identified molecular mechanism of interactions between specific phytoplankton-bacteria pairs and found that bacteria can both stimulate and inhibit phytoplankton growth and metabolism through a series of metabolic exchanges (Durham et al., 2017; Amin et al., 2015; Van Tol et al., 2017; Bartolek et al., 2022; Pollara et al., 2021; Stock et al., 2020). Similarly, constraint based metabolic models of individual marine species such as the diatoms *Thalassiosira pseudonana* (Van Tol et al., 2017) and *Phaeodactylum tricornutum* (Levering et al., 2016), as well as various marine bacteria (Fondi & Di Patti, 2019) have identified potential mechanisms of metabolic interchanges between species pairs. Potential microbial interactions within natural communities have been inferred using environmental surveys focused on co-occurrence of specific taxonomic groups. Co-occurrence analysis and microbiome characterization have shown that specific bacterial taxa such as *Bacteroidetes*, *Gammaproteobacteria*, and *Alphaproteobacteria* tend to associate with phytoplankton (Amin et al., 2012; Ramanan et al., 2016; Sterling et al., 2023). Community level metabolic models based on genomic potential have been used to predict the fluxes of nutrients

within bacterial communities both in the human gut microbiome and in marine settings (Magnúsdóttir et al., 2017; Levine et al., 2017), and are moving towards incorporating the effects of cross-feeding and metabolite exchanges between marine phototrophs and heterotrophs into biogeochemical models (Lücken et al., 2025). Similarly, outcomes of pairwise interactions between bacteria can be predicted using machine learning, although, machine learning models require prior knowledge of phenotype (healthy vs. diseased) or interaction outcome (co-existence vs. co-exclusion) classification for training, which is difficult to obtain in an ocean setting (DiMucci et al., 2021). Observing direct expression of specific interaction related gene markers and molecules in the environment is possible through environmental metagenomics, metatranscriptomics and metabolomics (Durham et al., 2022; Park et al., 2025). In the human gut microbiome, transcriptional interactions between pairs of bacteria were identified on a gene-to-gene differential expression level based on presence and absence patterns of individual species across patients (Plichta et al., 2016). These approaches have identified different aspects of microbial interactions and yet, linking specific molecular functions to species interactions in complex natural environments remains a challenge.

Here, we describe a novel statistical method Enviromtx that predicts transcriptional mechanisms of interaction in natural communities based on metatranscriptome data. Enviromtx identifies transcriptional features of interaction between species pairs by testing the expression patterns of all sequenced functions in relation to the two species co-abundance patterns along an environmental gradient. The approach does not rely on isolation and culturing of representative model organisms or on sequenced genomes and metabolic models of select species. Similarly, we approach the problem with no *a priori* expectations of interacting organisms or interaction-associated functional signals. We apply Enviromtx to metatranscriptome data collected on three transect cruises in the North Pacific to study which taxonomic pairs are interacting in the environment, and through what molecular functions. The North Pacific is particularly suitable to our analysis as it crosses numerous chemical, physical and biological gradients, resulting in variation of species abundances across the region (Follett et al., 2021; Dutkiewicz et al., 2024). The North Pacific consists of the subtropical gyre, a transition zone (NPTZ) defined by a salinity front at the southern most border and the transition zone chlorophyll front (TZCF) further north where surface chlorophyll *a* concentrations exceed 0.2 mg/m^3 (Polovina et al., 2001), and the subpolar gyre in the north. This environmental gradient is characterized by changes in nutrient regimes (Juranek et al., 2020), shifts in phytoplankton community structure (Church et al., 2008), and changing functional roles of different microorganisms (Gradoville et al., 2020; Lambert et al., 2022). Comparing results of interaction signatures from three separate latitudinal cruises along 158W allowed us to pull out the most robust interactions in the North Pacific. Detecting functional signatures of interactions directly from field metatranscriptome data makes it possible to survey a broad range of interacting species, identify high level patterns of interactions in the ocean, and propose additional targets for detailed lab or metabolic modeling experiments.

Methods

G1-3 metatranscriptome data collection and processing:

Metatranscriptome Sample Collection

Metatranscriptome samples were collected on three cruises in the North Pacific: Gradients 1 (KOK1606; April 20 - May 4, 2016), Gradients 2 (MGL1704; May 26 - June 4, 2017), and Gradients 3 (KM1906; April 10 - April 29, 2019). These cruises follow a latitudinal transect from ~20°- 40°N along 158°W crossing numerous environmental Gradients. Seawater was collected either from a CTD rosette or the ship's underway system as previously described in Groussman et al., 2024. Briefly, approximately 5-10 L of seawater was subsampled into 10 L acid-washed polycarbonate carboys, pre-filtered through a 100 µm (Gradients 2 and 3) or 200 µm (Gradients 1) mesh, and triplicate samples were sequentially filtered through 3 µm and 0.2 µm polycarbonate filters to generate both a 3-100 or 3-200 µm size fraction and a 0.2-3 µm size fraction. Filters were flash frozen in liquid nitrogen and stored at -80°C before processing.

Metatranscriptome Extraction and sequencing

Metatranscriptomes were generated as previously described (Groussman et al., 2024; Durham et al., 2019; Park et al., 2023). In short, total RNA was extracted using the ToTALLY RNA kit (Invitrogen AM1910) for Gradients 1, and the Direct-zol MiniPrep Plus kit (Zymo Research R2720) for Gradients 2 and Gradients 3. During extraction, a set of 14 synthetic spike-in mRNA standards was added to the samples as previously described (Groussman et al., 2024; Satinsky et al., 2013) to enable estimates of transcript abundance across samples. The extracted RNA was quantified with a Qubit fluorometer 1.0 (ThermoFisher) and quality controlled using a Bioanalyzer 2100 (Agilent). For each extracted filter (0.2 µm and 3 µm for each sample replicate), the RNA was split into two halves to generate either total community metatranscriptomes or eukaryotic community metatranscriptomes. Total community mRNA was generated by rRNA depletion using either Illumina RiboZero (Yeast+Bacteria) kit for Gradients 1 and Gradients 2 samples or ThermoFisher RiboMinus (Yeast+Bacteria; K15503, K15504) kit for Gradients 3 samples. Eukaryotic community mRNA was generated by polyA selection (conducted at the Northwest Genomic Center, University of Washington, Seattle). Resulting mRNA samples were sequenced on two separate plates for each cruise. Gradients 1 polyA samples were sequenced on an Illumina NextSeq500 platform using four high output flow cells with 300 cycles. All other samples were sequenced on the Illumina NovaSeq6000 platform using S2 flow cell with 300 cycles. Library prep and sequencing was done at the Northwest Genomics Center (University of Washington, Seattle).

Metatranscriptome assembly, normalization and annotation

Gradients 1, 2, and 3 polyA selected eukaryotic metatranscriptome sequences and assembled contigs have previously been published and the processing described in Groussmann et al., 2024. The Gradients 1, 2 and 3 rRNA depleted total community metatranscriptome processing is described here (Sup. 2.1). Raw Illumina reads from rRNA depleted samples were quality controlled with Trimmomatic v0.36 (MAXINFO:135:0.5, LEADING:3, TRAILING:3, MINLEN:60, and AVGQUAL:20) (Bolger et al., 2014). Short reads were assembled into longer contigs using Trinity de novo assembler v2.3.2 (--normalize_reads --min_kmer_cov 2 --min_contig_length 300) (Grabherr et al., 2011). All assemblies were performed on short reads from combined replicates. Assembled contigs were translated into amino acid space in six frames with transeq (Rice et al., 2000) vEMBOSS:6.6.0.059 using Standard Genetic Code, and the longest coding frame was retained for future analysis. The best frame translated assemblies were clustered at 99% amino acid identity with MMseqs2 (Steinegger & Söding, 2018). To quantify assembled contig abundances and obtain raw transcript counts, unassembled short reads were

mapped to the assembled contigs using kallisto v0.46.1 (Bray et al., 2016). To estimate transcript concentrations in transcripts per liter, the raw transcript counts for each assembled contig were multiplied with the normalization factors based on the recovery of 4 synthetic internal mRNA sequences (Satinsky et al., 2013; Durham et al., 2019; Park et al., 2023). Standard counts were recovered using the Bowtie 2 aligner v.2.5.2 using the standard pipeline for paired-end reads (Groussman et al., 2024). Taxonomic annotation of the assembled contigs was done with DIAMOND alignment to a combined reference database containing MarFeRReT v1.1 (Groussman et al., 2023) and MARMICRODB v.1.0 (Becker et al., 2019) reference databases. For annotation with MARMICRODB, custom taxIDs from MARMICRODB were converted to NCBI taxIDs, taking care that *Prochlorococcus* and *Synecoccus* taxID's matched, as the taxIDs were not published at the time MARMICRODB was released. Functional annotation of the assembled contigs was done using HMMER 3.3 (Eddy, 2011) against the Kyoto Encyclopedia of Genes and Genomes (KEGG) KOfam Hidden Markov Models (HMMs) database (Kanehisa & Goto, n.d.) with KOfam predictions with a threshold score > 30 retained for each contig, and against Pfam 35.0 (Mistry et al., 2021) HMMs where the 'trusted cutoff' assigned by Pfam to each hmm profile was used as a minimum score threshold. A summary of the total number of assembled contigs and taxonomic and functional annotation efficiency for all rRNA depleted total community metatranscriptomes from Gradients 1, 2 and 3 is available in Sup. Table 2.1.

Enviromtx statistical model:

Overview:

We designed a new statistical method, Environmtx, to detect functional signatures of interactions between microorganisms directly in the environment by determining whether the relative abundance of a given species impacts the gene expression patterns of another species. Enviromtx models pairwise transcriptional interactions between microorganisms using environmental metatranscriptome data. It estimates the percentage difference in the expression-per-unit-abundance of one species (transcript expression of a function in the "responder" species) for changes in the abundance-per-unit-abundance of the other species (abundance of the "companion" species relative to the abundance of the responder). The 'responder/companion-label' is arbitrary and is tested in both directions of a pair. The strength of the interaction is assessed, and a hypothesis test for a null interaction is performed. Each functional feature (KOfam) in the responder species can be tested in response to one companion at a time. The method is robust to unequal detection of different species and functional features between species across the dataset, and can account for correlated observational designs (e.g. replicates and longitudinal sampling).

Model input data:

To detect pairwise transcriptional interactions in the Gradients 1, 2 and 3 North Pacific metatranscriptome data, the Enviromtx statistical model was applied to each Gradients cruise separately, then aggregated (Sup. 2.1). The transcript abundance data for each taxonomy and KOfam along the cruise transect were obtained and aggregated as follows: Contigs with an taxonomic annotation at kingdom level of "Bacteria", "Archaea", and "Viruses" were retrieved from the ribosomal-depleted metatranscriptomes, and contigs with a taxonomic annotation at kingdom level of "Eukaryota" were retrieved from the polyA selected metatranscriptomes. Size fractions were combined per replicate per dataset, resulting in 0.2 - 200 μm size fractionation for

Gradients 1 and 0.2 - 100 µm size fractionation for Gradients 2 and 3. A sample replicate was disregarded if it lacked either a size fraction, or a total community or eukaryotic community component. This resulted in a total of 21 samples including replicates for Gradients 1, 29 for Gradients 2, and 22 for Gradients 3 cruises, spanning the entire transect of each cruise with comparable spatial resolution between cruises (Sup. Table 2.2; Fig. 2.1). Contigs taxonomically annotated at the species level were grouped into species bins. Contigs without species-level annotation were not analyzed further.

Within each species pair, the Enviromtx model was applied to all responder KOfams individually. The responder (X_{ij}) and companion (X_{ij*}) transcript abundance (transcripts/L) was calculated by summing the estimated transcript concentrations per species bin, including both functionally annotated and un-annotated counts. The KOfam abundance (Y_{ijk}) was calculated by summing transcript concentrations of all contigs per responder taxonomic bin annotated with the same KOfam per sample in transcripts/L.

Enviromtx Model Structure

For every species pair detected in environmental metatranscriptome data, we define one member of the pair as the responder and one member of the pair as the companion and test whether the responder shows a transcriptional response, based on KOfam expression, to the abundance of the companion species. For a fixed function *k* in the responder species, responder species abundance *j* and companion species abundance *j**, we estimate the multiplicative increase in the average expression-per-unit-abundance of function *k* when the abundance of species *j** per unit abundance of species *j* changes by 1% (e.g.). The model can also account for abiotic covariates (e.g., temperature, nutrient availability, etc.) to compare across environments with the same abiotic characteristics but different responder-companion abundance ratios. The methodology is implemented in the R package enviromtx available at https://github.com/statdivlab/enviromtx.

The Enviromtx model is as follows. Let Y_{ijk} denote the observed transcript abundance of function *k* in responder species *j* (responder species) in sample *i*, and X_{ij} denote the observed abundance of responder species *j* in sample *i*, and X_{ij*} denote the observed abundance of companion species *j** (companion species) in sample *i.* Here, observed abundance can be on the coverage, count, or relative scale. We then model that a 1% increase in X_{ij*} relative to X_{ij} is associated with a 1.01^{β_1}- fold increase in the average value of Y_{ijk} relative to X_{ij}. In the absence of additional adjustment variables, the model is

$$\mathbb{E}\left[\frac{Y_{ijk}}{X_{ij}}\right] = \gamma_0 \times \left(\frac{X_{ij*}}{X_{ij}}\right)^{\gamma_1} \quad (1)$$

while the covariate-adjusted generalization is

$$\mathbb{E}\left[\frac{Y_{ijk}}{X_{ij}}\right] = \beta_0 \times \left(\frac{X_{ij*}}{X_{ij}}\right)^{\beta_1} \times e^{\beta_2 W_2 + \cdots + \beta_p W_p} \quad (2)$$

where W_2, …, W_p are abiotic covariates. In the covariate-adjusted model, a 1% increase in X_{ij*} relative to X_{ij} is associated with a 1.01^{γ_1}- fold increase in the average value of Y_{ijk} relative to X_{ij} when comparing environments with the same values of the abiotic factors W_2, …, W_p.

Estimation and testing

We estimate the model parameters β_0, β_1, β_2… β_p using Poisson regression with log link function (using a logarithm to relate predictors to the outcome), and accounting for correlations between repeated measurements by using a user-specified correlation structure. The advantages of this estimation approach are numerous: it can accommodate a flexible variety of experimental designs (with respect to both covariate adjustment and non-independence of sampling); it naturally handles sparsity in observed transcript abundances (without pseudocounts); and estimates of β_1 are asymptotically unbiased under weak assumptions (Poisson variance is not necessary).

We tested the hypothesis that the average expression of function k per unit abundance of species j does not vary with fold-differences in the relative abundance of species j* to species j, equivalently, that $\beta_1 = 0$. Evidence against H_0: $\beta_1 = 0$ implies that the average expression-per-unit-abundance differs across different abundance-per-unit-abundance when comparing environments matched on adjustment covariates. To ensure valid error rate control (i.e., correct p-values) with practically feasible sample sizes, sparse Y_{ijk}, and without distributional assumptions on Y_{ijk}, we use model-robust score tests. We developed the raoBust R package for this application: https://github.com/statdivlab/raoBust.

Model assumptions and optimality

In addition to [2], the Enviromtx model makes three assumptions, and if these assumptions hold, β_1 estimates are guaranteed to approach the true difference in expression-per-unit-abundance as the sample size becomes large ("consistent"). The first assumption is that the observed species abundances are, on average, proportional to the product of true species abundances (e.g., in cell or DNA copies per ml), an unknown sample-specific parameter, and an unknown species-specific parameter (McLaren et al., 2019). The second assumption is similar: observed transcript abundances are, on average, proportional to the product of true transcript abundances (e.g., in transcripts per L), an unknown sample-specific parameter, and an unknown transcript-specific detectability parameter. Finally, the ratio of the sample-specific parameters across the transcript and species abundance data is constant across samples.This final assumption is plausible when transcript and species abundances are determined from the same metatranscriptome data. In contrast, it would not apply if species abundances were determined using metagenomics data, for example, while transcript abundances were determined using metatranscriptome data. While estimates of β_1 are consistent under the three stated assumptions, β_0 absorbs KOfam and species detectability parameters, and therefore has no ecological interpretation and should not be interpreted.

The first assumption is valid when responder and companion abundances are determined from the same samples as the sample intensity cancels in the numerator and denominator of [2]. This assumption may be violated when responders and companions are detected from the two different sequencing approaches: the total community metatranscriptomes, and the eukaryotic community (polyA selected) metatranscriptomes. In these cases, the sample-specific intensities will differ across the two selection samples, because these samples were sequenced separately. This limitation is unavoidable in our methodology, and we therefore exercise additional caution interpreting these cases.

In some metatranscriptome samples the responder and companion abundances (X_{ij} and X_{i*}) are zero, because of two possibilities that we cannot distinguish: either the species is truly not present in the sample, or the species is not detected in a given sample within the limits of sequencing depth. To address these missing species abundances, we implement a default

pseudocount option (which can be user-adjusted), using pseudocount values as the minimum abundance per species across samples. Note that pseudocounts are used only for the responder and companion species abundances. They are not required for the transcript abundances of individual functions, which can be sparse across samples.

Applying enviromtx to North Pacific Metatranscriptomes:

Applying enviromtx to Gradients 1, 2 and 3 datasets

Enviromtx was applied to each cruise separately (Sup. 2.1) to test pairwise interactions between all possible pairs of species level taxonomic bins with sufficient functional annotation, here defined as more than 500 annotated KOfams per Eukaryotic species, and more than 300 KOfams per Bacteria, Viral or Archaea species. Within each pair, we considered both options regarding which species was the responder and which was the companion. We adjusted for environmental conditions that varied across the transects by first calculating Pearson's correlation between every pair of measured environmental parameters across the three cruises. We then selected one physical measurement of the transect (Temperature), one macronutrient measurement (Nitrate + Nitrite) and one micronutrient measurement (Manganese) with the least correlation to each other (Pearson's coefficient < 0.5). Each cruise was conditioned with the same set of environmental parameters. Within each cruise, the model was fitted using a cluster correlation structure, with technical replicates in the same cluster, and unknown within-cluster correlation. A subset of function-responder-companion combinations were computationally unstable and did not converge (G1: ~14% of runs; G2: ~6% of runs; G3: ~10% of runs) (Sup. 2.2). We refit these models with a known within-cluster correlation (correlation within technical replicates), using the average of the within-cluster correlations estimated from the stable function-responder-companion combinations (correlation:alpha for G1: -0.014, G2: 0.052, G3: 0.091). If estimation still did not converge, the model was refitted using an independent correlation structure (Sup. 2.2). In these cases, β1 estimates can still be interpreted, but p-values may be anti-conservative (smaller than evidenced).

Enviromtx internal validation

As an initial internal validation of the model, we compared the magnitudes of β_1 for each combination of KOfam, responder and companion in all pairs of cruises for both the entire set of Enviromtx KOfams (with significant and insignificant p-values) and for only those that were significant (Enviromtx p-value < 0.1). KOfams that appear in both cruises were retained and those present in only one cruise were disregarded. We used a one-sided two-sample test for proportions (conf.level = 0.95) to test the null hypothesis that the model produces random guesses regarding the coincidence of the sign of β_1 for a given combination of KOfam, responder and companion.

Combining enviromtx results across three Gradients cruises

Enviromtx was applied to Gradients 1, 2 and 3 data separately and the results were combined to detect the most robust transcriptomic response signals that persisted across years and times of year (Sup. 2.1). Combined β_1 values across three cruises for each responder/companion/function instance were calculated as the average of β_1 from Enviromtx outputs from Gradients 1, 2 and 3. P-values testing the same hypothesis (for each responder-companion-function) were combined into a single p-value by constructing a chi-squared test as

follows. The corresponding quantile of a chi-squared (df = 1) distribution for each p-value was determined, and these quantiles were summed. As each cruise was independent of the others, this sum is chi-squared distributed (df = number of cruises) under the null hypothesis of no change in expression-per-unit-abundance across patterns of abundance-per-unit-abundance, and an overall p-value can therefore be found. There are cases where a responder-companion-function instance is missing a β_1 and p-value prediction in one or more cruises. This results from cases where a particular KOfam in the responder species is either not detected or not expressed across all cruises. In these cases, the combined β_1 and p-values were calculated based on cruises with available results only.

Analysis of Enviromtx outputs:

Selecting highly interacting companions per responder

Highly interacting companions per responder were defined as those companions that elicited more significant KOfams per responder, specifically companions with an outlier number of KOfams, defined as greater than 75% percentile (Q3) + 1.5 Interquartile Range (IQR) KOfams across all companions (Sup. 2.10B). To ensure biological interpretability of the highly interacting pairs, we employed two additional cutoffs: highly interacting companions per responder had above the median number of significant KOfams per pair across the entire dataset (above 7 KOfams per pair), and more than the median number of enriched KEGG pathways per pair across the entire dataset (see KEGG enrichment analysis below) (Sup. 2.10C).

Co-occurrence Pearson calculations

Co-occurrence of species transcriptional abundances across cruises in the North Pacific was calculated based on estimated total transcripts/L per each species taxonomic bin across a combined dataset for Gradients 1, Gradients 2 and Gradients 3 cruises. Pearson's correlation coefficient was calculated for all possible pairwise combinations of species.

KEGG enrichment analysis

KEGG pathways and their associated KOfams were obtained from the KEGG database release 101.0 using the R package KEGGREST (Tenenbaum, 2016). Pathway enrichment analysis was performed at two levels: for each responder/companion pair separately, and for functions that have a consistent interaction signal across pairs in bacteria and eukaryotic responders (see methods on Wilcoxon rank sum test below). To test for KEGG pathway enrichment in each responder/companion pair, a contingency table was constructed for every KEGG pathway in each responder/companion pair, and tested for enrichment of all KEGG pathways in KOfams that had a significant Enviromtx p-value and were therefore involved in mediating the interaction between the two species. We also tested for KEGG pathway enrichment in bacterial and eukaryotic functions consistently involved in interactions across pairs, testing for enrichment in the unique set of functions detected through the Wilcoxson rank sum test for bacteria and eukaryotes separately that also had a significant Enviromtx p-value in at least one pair. The Fisher's Exact Test was used to test for enrichment of KEGG pathways, followed by a Benjamin-Hochberg multiple comparison correction with false discovery rate < 5%.

Wilcoxon rank sum test

We identified KOfams that are most commonly involved in an interaction response across responder-companion pairs for bacterial and eukaryotic responders separately using a Wilcoxson rank sum test. For each KOfam, we evaluate the Enviromtx p-value distribution across all responder-companion pairs the KOfam appears in. We compared the distribution of Enviromtx p-values for each KOfam to the distribution Enviromtx p-values across all KOfams. Multiple comparison correction was done post-hoc using the Bonferroni method (Bland & Altman, 1995). KOfams that showed a significant Wilcoxon rank sum p-value (< 0.01) and that showed a significant Enviromtx p-value (<0.1) in at least one responder/companion pair were considered.

General vs. specific functional response

We calculated the percentage of total companions that elicited a significant interaction response (Enviromtx p-value < 0.1) per KOfam for each responder species across the entire dataset. For each responder, a KOfam was considered to be part of a "specific response" or a "general response" category if the percentage of companions that elicit a significant interaction signal was below (specific response) or above (general response) the median percentage of companions that elicit a significant interaction signal across all KOfams. To evaluate broad patterns of "specific response" or "general response" KOfams across multiple responders, we combined the results across all responders in bacteria and eukaryotes separately. We categorized the KOfams into standard KEGG pathways, KEGG Brite hierarchies and custom made categories that highlight previously identified interaction-specific processes based on literature review (Sup. Table 2.3).

Results

Transcriptional abundance of Bacteria and Eukaryotes across the North Pacific transition zone

Three cruises - Gradients 1 (Spring 2016), 2 (Summer 2017) and 3 (Spring 2019) - each transited from the low nutrient, low chlorophyll North Pacific Subtropical Gyre (NPSG), across the North Pacific Transition Zone (NPTZ) as defined by a salinity front isohaline (34.82) (Fig. 2.1, dashed blue line) and a seasonally migrating chlorophyll front (0.15 mg m^{-3}) (Fig. 2.1, dashed green line), northward towards the iron-limited, high nutrient low chlorophyll (HNLC) North Pacific Subarctic Gyre (Fig. 2.1) (Juranek et al., 2020; Polovina et al., 2001; Ayers & Lozier, 2010). Samples for metatranscriptomes were collected at different stations along the three cruise transects (Fig. 2.1) and were processed for both eukaryotic community (polyA selected) and total community (ribosomal RNA depleted) data. Transcriptional abundance was estimated by aggregating total transcripts/L per taxonomic bin in a given sample.

Abundance patterns of the unicellular eukaryotic and bacterial portions of the community shifted along the three transects. The eukaryotes were dominated by Haptophytes and Dinoflagellates, after correcting Dinoflagellate abundances for unusually large transcript to biomass ratios (Coesel et al., 2025) (Fig. 2.1). The transition zone chlorophyll front was characterized by a higher relative abundance of Pelagophytes (Fig. 2.1). North of the chlorophyll front, the eukaryotic community shifted towards a larger proportion of Chlorophytes and Bacillariophytes, corresponding to higher nutrient and chlorophyll concentrations and a shift towards larger cell sizes (Juranek et al., 2020) (Fig. 2.1). General patterns of transcript

abundances of major bacterial groups mirror abundances derived from 16S amplicon data collected on the same cruises (Key, n.d.) (Sup. 2.3). Bacterial taxa commonly associated with particles or phytoplankton, including Bacteroidota and Alpha- and Gamma-proteobacteria (Amin et al., 2012; Seymour et al., 2017) increased in relative transcript abundance north of the chlorophyll front, coinciding with shifts in eukaryotic community composition (Fig. 2.1; Sup. 2.4; Sup. 2.5). Enterobacterales were dominant in Gradients 1, and absent from Gradients 2 or Gradients 3 cruises. These bacteria were likely introduced as contaminants during sampling or processing on Gradients 1 as they are generally prevalent as human pathogens or in coastal nutrient rich environments (Janda & Abbott, 2021), and were not present in high abundance in 16S amplicon data sampled and processed independently (Sup. 2.3). Across the three transects, Cyanobacteriota, especially *Prochlorococcus* decreased in the north (Fig. 2.1; Sup. 2.5) (Dutkiewicz et al., 2024).

Detecting pairwise transcriptional interactions - Enviromtx statistical model

We developed a statistical model called Enviromtx to detect potential species interactions based on transcriptional patterns derived from the metatranscriptome data (Methods). We used Enviromtx to identify functional features (KOfams) that reflect interactions between species pairs in the environment without a-priori knowledge of which organisms interact with each other or through which processes. Each species pair was analyzed individually with Enviromtx by assigning one species in the pair as the "responder", and the other as the "companion" (Fig. 2.2A). Initial assignment of "responder" and "companion" in a pair was arbitrary, and the assignment was subsequently reversed such that all species were tested as both a responder and a companion. The Enviromtx model estimates β_1, which reflects either a positive or negative correlation of expression of each KOfam function per species pair relative to the abundance of responder and companion species in a pair (Fig. 2.2A). To control for the effect of environmental parameters on gene expression patterns, we adjusted for three environmental parameters with the least correlation to each other based on Pearson's correlation coefficient. This resulted in temperature as a physical measure of the transition zone, nitrate (plus nitrite) concentrations as a macronutrient measurement, and dissolved manganese concentrations as a micronutrient measurement (Pearson correlation coefficient: temperature with $NO_3 + NO_2$= -0.72; temperature with Mn = 0.79; $NO_3 + NO_2$ with Mn = -0.51) (Sup. 2.6).

A total of 119 eukaryotic and 86 bacterial environmental species bins had sufficient (>500 for eukaryotes and >300 for bacteria) KOfam annotations for analysis (Methods; Sup. 2.7; Sup. Table 2.4). The unicellular eukaryotic species bins belonged to 22 phyla dominated by Dinoflagellata (39 out of 119), Haptophyta (16 out of 119) and Bacillariophyta (12 out of 119), and the 86 bacteria species bins belonged to 8 phyla dominated by Pseudomonadota (60 out of 86) and Bacteroidota (11 out of 86) (Fig. 2.2B; Sup. 2.7; Sup. Table 2.4). Interactions between all 205 possible species pairs were tested, resulting in a total of 41,820 species pairs (Fig. 2.2B).

As an internal validation of the model, we tested whether for each species pair, the β_1 for a given KOfam had the same sign (positive or negative) across the three cruises (Methods). The direction of correlation of β_1 was similar between all three cruises for all tested KOfams, regardless of whether the β_1 p-values were significant or not (Sup. 2.8). When only the β_1 values for KOfams with significant p-values were compared, the directions of β_1 were more similar than random between Gradients 1 and Gradients 2 (53% similar, p-value = 3.86^{-5}), and between Gradients 1 and Gradients 3 (15% similar, p-value < 2.2^{-16}). In contrast, the directions of β_1 were

not significantly similar between Gradients 2 and Gradients 3 (52% similar, p-value = 0.3), likely because Gradients 2 and Gradients 3 overlapped in relatively few significant KOfams (Fig. 2.2C; Sup. 2.8). These results indicate consistent biological patterns of species interactions, despite the fact that the three cruises were conducted in different years and at different times of year (Fig. 2.1).

To identify the most consistently observed functional interactions in the North Pacific, we applied the Enviromtx model on each of the Gradients 1, 2 and 3 cruises separately, and combined post-hoc the p-values and β_1 for each KOfam in a species pair across the three cruises (Methods). This resulted in a combined p-value and β_1 for every KOfam in all species pairs (Sup. 2.9A). This approach identified consistent signals where the same responder-companion-function was detected across multiple cruises. Subsequent analyses were based on these combined p-values and β_1's. We tested 74,390,822 responder-companion-function instances, where values of β_1 ranged from ~-10 to ~10, with a median β_1 of zero (Sup. 2.9B). Of all tested KOfams across all pairs and cruises, 792,117 (approximately 1%) were significant (combined p-value < 0.1) (Sup. 2.9A). Per pair, the median percent of KOfams with a significant p-value was ~0.5% of the total number of annotated KOfams in the responder species, with ~10% as highest percent of significant KOfams (Sup. 2.9C). This suggests that across taxonomically diverse responder species, a relatively small fraction of the transcriptome is regulated in response to interactions with companion species.

Highly interacting taxonomic groups in the North Pacific

We identified a subset of highly interacting species pairs (Sup. 2.10A) as those pairs where a given responder species had a greater than typical number (Q3 + 1.5 IQR) of KOfam functions significantly involved in an interaction (Methods; Sup. 2.10A,B,C). We identified 547 highly interacting species pairs, consisting of 112 distinct responder species and 138 distinct companion species (Sup. 2.10A; Sup. Table 2.5). The median percent of significant KOfams per highly interacting pair was 1.3% (Sup. 2.10D), significantly greater than the ~0.5% observed across the entire dataset (Sup. 2.9C). Further, we found that the abundances (estimated total transcripts/L) of the identified highly interacting pairs had a significantly stronger co-occurrence based on Pearson's correlation coefficient than typical pairs (Sup. 2.11), suggesting that highly interacting species pairs generally co-occur more commonly in the environment.

The species level results were grouped at higher taxonomies to determine whether there were general patterns of interactions across kingdom or phylum levels. First we asked whether bacterial responders highly interacted with more bacterial companions or with more eukaryotic companions and vice versa. Across all highly interacting pairs, bacterial responders interacted with a total of 116 bacterial companions and 201 eukaryotic companions, while eukaryotic responders interacted with a total of 139 bacterial companions and 91 eukaryotic companions. For each individual responder species, bacterial responders had a significantly larger proportion of highly interacting eukaryotic companions than bacterial companions, while eukaryotic responders had a significantly larger proportion of highly interacting bacterial companions (Sup. 2.10E; Sup. 2.12A). This pattern is consistent with the hypothesis that cross-kingdom interactions between phytoplankton and bacteria in the surface ocean are dominant in driving nutrient cycling, where bacteria consume and remineralize the organic matter produced by phytoplankton through photosynthesis (Fuhrman, 2009; Thingstad & Lignell, 1997).

We next asked whether the highly interacting species were members of particular phyla. A high proportion of bacterial species within the Pseudomonadota (both Gammaproteobacteria and Alphaproteobacteria) and the Bacteroidota displayed significant interactions with eukaryotic species within the Bacillariophyta (Fig. 2.3A). Interactions of Bacillariophyta with Bacteria involved several groups, where Gammaproteobacteria, specifically Oceanospirillales, Weoseiales and SAR86 cluster, responded to a wide range of Bacillariophyta species, including *Thalassiothrix antarctica, Fragilariopsis kerguelensis*, *Coscinodiscus wailesii*, *Pseudo-nitzschia heimii*, *Chaetoceros dichaeta* and *Thalassiosira bioculata*. Bacillariophyte response to these Gammaproteobacteria companions was dominated by two species, *Chaetoceros dichaeta* and *Thalassiosira sp*. Interactions with Bacteroidota were dominated by a response of Flavobacteria to *Chaetoceros dichaeta* and *Thalassiosira bioculata*, while interactions with Alpaproteobacteria were dominated by a response of Rhodobacterales and Pelagibacteriales, including SAR116 cluster, to the same two diatom species (Sup. 2.12B; Sup. Table 2.5). Most of the bacteria interacting with Bacillariophytes were copiotrophic, aligning with previous reports that copiotrophic bacteria associate with diatom hosts and engage in exchanges of organic matter and other molecules, although oligotrophic SAR116 also showed interactions with diatoms (Cirri & Pohnert, 2019; Amin et al., 2015; Durham et al., 2015; Van Tol et al., 2017; Bartolek et al., 2022; Barak-Gavish et al., 2018). Additional interactions were between Gammaproteobacteria and eukaryotic phyla such as Foraminifera, Chrysophyceae, Pelagophyceae and Bolidophyceae, as well as between Alphaproteobacteria and Dictyochophyceae (Fig. 2.3A). Within bacterial phyla, most interactions occurred between Bacteroides and Alpha and Gammaproteobacteria (Fig. 2.3A). In addition to these general patterns, there were numerous phyla such as Neomarimicrobiota, Dinoflagellata, Chlorophyta, Betaproteobacteria, Haptophyta and Verrucomicrobia that interacted with a range of other taxonomies through highly specific interactions at the species level (Fig. 2.3A; Sup. 2.12B).

To detect broad patterns of functional features of interactions within the highly interacting species pairs, we conducted enrichment analysis of responder KOfams with a significant Enviromtx p-value in every species-species pair (Sup. 2.13; Sup. Table 2.6). We grouped the highly interacting pairs at the phylum level to identify the most commonly enriched pathways (Fig. 2.3B). When Bacillariophyta responded to Gammaproteobacteria, the most commonly enriched pathways included those related to secondary metabolite and natural product synthesis and processing, such as monoterpenoid biosynthesis, carbapenem biosynthesis, puromycin biosynthesis and penicillin biosynthesis as well as cytochrome P450 processes including drug metabolism. Enrichment of these pathways in Bacillariophyta by Gammaproteobacteria is consistent with known roles of secondary metabolites in cross-kingdom interactions (Shibl et al., 2020). Bacillariophytes also responded to Gammaproteobacteria with pathways involved in amino acid biosynthesis and metabolism; fatty acid- and steroid-related pathways; and synthesis of glycan and glycosaminoglycan, which are related to cell wall structure, formation of the extracellular matrix and potentially signaling and interactions with surrounding cells (Fig. 2.3B). Gammaproteobacteria responders had different enriched pathways depending on the companion phyla. For example, glycerophospholipid metabolism was enriched only in response to eukaryotic Bacillariophytes and Pelagophytes, potentially signifying increased nutrient utilization or surface colonization whereas sphingolipid metabolism, furfural degradation, and mannose-type and other O-glycan biosynthesis pathways were enriched only in response to Bacteroidota (Fig. 2.3B). Both of these specific responses suggest potential surface interactions or colonization, metabolic opportunism, and competition. Gammaproteobacteria also

showed a more general response to various eukaryotic and bacterial companions, including ones related to antimicrobial production and defense such as vancomycin, tetracycline, nonribosomal peptides pathways and atrazine degradation; iron and oxidative stress response such as siderophore biosynthesis, ferroptosis and pentose phosphate pathways; motility and attachment such as chemotaxis and biofilm formation; and core metabolism pathways, including glycolysis, glucogenesis and butanoate metabolism (Fig. 2.3B). When the protist group Dictyochophyceae responded to Alphaproteobacteria, the enriched pathways included ones related to membrane function, nutrient transport, photosynthesis and redox balance, and secondary metabolism (Fig. 2.3B). These enriched pathways are consistent with mixotrophic behavior whereby photosynthetic cells carry out phagocytosis in response to the presence of bacteria. Both Alphaproteobacteria and Bacteroidota responded to Bacillariophyta. However, Alphaproteobacteria responded with enrichment in biofilm formation, ferroptosis and various pathways involved in steroid, polyketide, and alkaloid synthesis, whereas Bacteroidota responded with enrichment of amino acid metabolism, glucosinolate and antibiotic synthesis (Fig. 2.3B). Overall, two general patterns emerged from these analyses. First, as expected, responder taxonomy constrains the types of functions with which organisms can respond to each other. Second, and more intriguingly, was that some pathways were enriched across multiple species-species pairs, while others were enriched in only a few pairs, suggesting that some functional features are more commonly involved in interactions, while others are specific to a subset of interacting pairs.

Example interacting pairs at species level

In addition to broad patterns within highly interacting pairs, Enviromtx can be used to identify candidate interactions between specific species pairs and the potential functions involved. Because each species is evaluated as both a responder and a companion, the interactions can be classified as either unidirectional, where only one member of the pair responds, or reciprocal, where both members of the pair respond. Few species-species interactions appeared to be reciprocal; of the 547 highly interacting pairs, only 8 exhibited reciprocal interactions, in which both the responder and the companion demonstrated a highly interacting response (Sup. 2.14). This suggests a predominance of either commensal or opportunistic interactions within the North Pacific microbial community. To illustrate the power of Enviromtx, we highlight two examples of candidate interactions between species bins with distinct functional signatures: a reciprocal interaction between *Pseudoalteromonas sp*. and *Chaetoceros dichtea*; and a unidirectional response of the Alphaproteobacteria OCS116 cluster bacteria to the eukaryote *Stramenopiles sp* based on iron metabolism and siderophore-related KOfams (Fig. 2.4).

Pseudoalteromonas sp. and *Chaetoceros dichaeta* interacted reciprocally, with *Chaetoceros* inducing broader transcriptional changes in *Pseudoalteromonas* (~5% of KOfams) than vice versa (~1%) (Sup. 2.14). The *Pseudoalteromonas* response included shifts in antibiotics and toxins, bacterial toxins, and pyruvate metabolism (Sup. 2.15A). At the KOfam level, multiple enzymes in pyruvate metabolism showed negative β1 values, suggesting reduced internal generation and turnover of pyruvate, while a positive β_1 for LDHD oxidizing D-lactate to pyruvate may reflect selective regeneration of pyruvate from external lactate (Fig. 2.4A). KOfams associated with bacterial chemotaxis including cheV and fliS had a negative β_1, indicating reduced swimming behavior (Fig. 2.4A). In contrast, KOfams for biosynthesis of

secondary metabolites, including violacein (*vioC*) and carbapenem pathways (*thnE*) had a positive β_1 (Fig. 2.4A). On the other hand, the *Chaetoceros dichaeta* response to *Pseudoalteromonas sp*. was dominated by cell cycle, signaling, and transport functions (Sup. 2.15B). Positive β1 values for cell cycle genes (PPP2R1, APC1) and for secretion-related genes (DENND2, SEC62) suggest enhanced cell cycle progression and secretion pathways (Fig. 2.4A). In summary, *Pseudoalteromonas* reduced motility and shifted towards defense and reduced pyruvate consumption, possibly using other carbon sources exuded by the diatom, while *Chaetoceros* increased transcription of cell cycle and secretion pathways, possibly increasing its growth rate and providing the bacteria with organic matter.

The Alphaproteobacteria *OCS116 cluster bacteria* responded to the eukaryote *Stramenopiles sp.* mainly through biosynthesis of secondary metabolites, including siderophore group peptides (Sup. 2.15C). Two key KOfams involved in siderophore biosynthesis have a positive β_1: entE activates a precursor used in synthesis of enterobactin and vibriobactin, and mtbE is a peptide synthase required for mycobactin synthase (Fig. 2.4B). Positive β_1 values of CoA ligase KOfams suggest enhanced biosynthesis of lipids or bioactive compounds (Fig. 2.4B). This unidirectional interaction reveals that Alphaproteobacteria OCS116 responded to *Stramenopiles sp*. by increasing siderophore and secondary metabolite synthesis. These examples illustrate the power of our method to identify potential candidate interactions between specific species pairs in nature and the molecular mechanisms through which they take place.

General vs. specific functional signatures of interactions

In addition to highly interacting pairs and novel species-species interactions, we identified general functional patterns of interactions across the entire North Pacific. A Wilcoxon rank sum test (see Methods) was used to identify KOfams and functional categories with significant interaction signals across all responder-companion pairs for bacterial and eukaryotic responders separately. For the bacterial responders, 1729 KOfams were significantly ranked as consistently involved in an interaction. For the eukaryotic responders, 1166 KOfams were ranked as consistently involved in an interaction. Across a wide range of phyla, the bacterial KOfams were enriched for KEGG pathways including two-component systems, flagellar assembly, and ABC transporters (Sup. 2.16A). These pathways belong to the functional categories of transport, amino acid synthesis and metabolism as well as biosynthesis of cofactors (Sup. 2.16B). The eukaryotic KOfams were enriched for valine, leucine and isoleucine degradation; riboflavin metabolism; and aminoacyl-rRNA biogenesis (Sup. 2.17A). These pathways belong to the functional categories of membrane trafficking, chromosome and associated proteins, transporters and ubiquitin systems (Sup. 2.17B).

We next evaluated how the functional basis of interactions is distributed across all species pairs. Specifically, we asked if certain functions were induced in a given responder by multiple companions (general functions), and if other functions in the same responder were induced by a select set of companions (specific functions) (see Methods). Across all responder species, most interaction-induced functions were specific to a few companions, with a smaller set of functions that were always induced in response to multiple companions (Sup. 2.18A,B). To determine which functional features dominated specific interaction responses compared to general interaction responses across responders, we assigned every responder KOfam into either a “general response” or a “specific response” group, and combined all “general response” and “specific response” KOfams across bacterial and eukaryotic responders (see Methods). We

defined a general response as one elicited by a greater than median number of companions and a specific response as one elicited by lower than median number of companions. For both bacterial and eukaryotic responders, one set of KOfams were always specific (2759 KOfams for bacteria and 2162 KOfams for eukaryotes) and a different set were always general (1141 for bacteria and 2064 for eukaryotes) across responders. A third set of KOfams were either general or specific depending on the responder species (1641 for bacteria and 5972 for eukaryotes) (Sup. 2.18C). As KOfams in both the specific and general response groups can belong to the same functional category, we identified those functional categories with the greatest proportional difference in KOfams that belonged to a specific versus general response group (Fig. 2.5; Sup. Table 2.3). The general response of bacteria was dominated by KOfams belonging to transporters, biosynthesis of amino acids and biosynthesis of various toxins, phytohormones and antibiotics suggesting a prevalence of both cross-feeding as well as competition and modulation of interacting organisms through chemical cues. The specific response of bacteria to select companions included quorum sensing, biofilm formation, metabolism of various B vitamins, siderophore related functions, and various metabolic and regulatory pathways (Fig. 2.5). The general response of eukaryotes was dominated by various B vitamin pathways, transcription factors, 2-component systems, and transporters suggesting a prevalence of keystone metabolites in the eukaryotic response (Durham et al., 2025), as well as sensing and cellular remodeling. The specific response of eukaryotes to select companions included ribosomal processes, endocytosis, cell cycle related pathways, plant hormone signaling and biosynthesis, and various nucleic acid processes (Fig. 2.5), suggesting potential growth, cell cycle changes and export in response to specific companions. Together, these results suggest that microbial interactions are characterized by both general functional responses ubiquitous across the microbial community, and specific functional responses that might enable co-evolution and species-specific interdependencies.

Discussion

Although accumulating evidence highlights the importance of marine microbial networks to ecosystem function, identifying which organisms are involved in the interactions and through which mechanisms remains challenging due to the complexity of these networks. Here, we developed a statistical model Enviromtx for use with metatranscriptome data to identify the basis of potential interactions. The underlying premise of Enviromtx is that the abundance of two species relative to each other dictates their ability to interact, a signal that is detectable through changes in transcriptional patterns of the two species. Species abundance changes across environmental gradients, which creates natural experiments for detecting interaction signals (Fig 2.1; Guidi et al., 2016; Teeling et al., 2012; Buchan et al., 2014). However, changing environmental conditions can also cause transcriptional responses independent of interactions. Enviromtx therefore also incorporates the ability to control for differences in environmental parameters. We applied Enviromtx to 72 metatranscriptomes collected on three Gradients cruises that span nearly 20 degrees of latitude in the North Pacific Ocean. We analyzed close to 41,000 environmental species pairs across the three cruises and identified robust interactions that persist across environmental regimes. The direction of interactions (up- or downregulation of gene expression) were more consistent across cruises than expected by chance, although only 52–59% similar, likely due to variability in sampling years, seasons and local conditions. The Gradients 2 cruise showed the least similarity with the other two cruises, consistent with previously observed

anomalies in community composition and nutrient limitation relative to Gradients 1 and 3 (Hawco et al., 2025).

At higher taxonomic levels, cross-kingdom interactions dominated, bacterial responders frequently interacted with eukaryotic companions and vice versa, consistent with established ideas of phytoplankton-bacteria coupling in the surface ocean (Azam & Malfatti, 2007; Seymour et al., 2017; Buchan et al., 2014; Levine et al., 2017). Detected interactions were characterized by phyla known to co-occur, with diatoms consistently interacting with Alphaproteobacteria, Gammaproteobacteria, and Bacteroidota (Yang et al., 2023; Amin et al., 2012; Ramanan et al., 2016; Sterling et al., 2023). A prevalence of highly interacting species occurred between Gammaproteobacteria and environmental diatoms most closely related to *Chaetoceros dichaeta* and *Thalassiosira sp.* In lab studies the Gammaproteobacteria *Oceanospiralles* induced amino acid production in *Thallasiosira pseudonana* (Paul et al., 2013), and the diatom *Coscinodiscus radiatus* growth was inhibited by numerous Gammaproteobacteria species (Di Costanzo et al., 2023). A prevalence of toxins, antibiotics and defenses in both Bacillariophytes and Gammaproteobacteria interacting with each other suggests predominance of antagonistic relationships in the North Pacific (Coyne et al., 2022). Within bacteria, recurring Bacteroidota - Proteobacteria interactions through complex sugars and phytoplankton produced compounds including cell wall structures reflect both syntropy and competition for liable DOM in particle-associated niches (Ferrer-González et al., 2021). Overall, the functional mechanisms involved in phyla-level interactions reflect the importance of nutritional exchanges as previously demonstrated through lab cultures and modeling studies (Christie-Oleza et al., 2017; Weissberg et al., 2025, Dutkiewicz et al., 2024), however, a large portion of the observed interaction signal is also related to defense and signaling mechanisms, suggesting a more complex interaction network in the North Pacific than purely based on exchanges of organic matter and limiting resources (Amin et al., 2012; Coolahan & Whalen, 2025; King et al., 2022; Ahern et al., 2021; W. Stock et al., 2019). The functional diversity of interactions at the species level is highlighted with two potential new interactions, including a nutritional exchange between copiotrophic *Pseudoalteromonas* and the diatom *Chaetoceros dichaeta*, and a highly specific siderophore mediated interaction between the Alphaproteobacteria OCS116 cluster and a *Stramenopiles* sp.

Across interacting species, we observed both general responses (activated by many companions) and specific responses (activated by few). In bacterial responders, specific interaction responses involved quorum sensing, biofilm formation, B vitamin metabolism, siderophores, and other metabolic and regulatory pathways. Long term phytoplankton associated bacteria were enriched in similar pathways compared to transient phycosphere bacteria (Focardi et al., 2025), suggesting these mechanisms mediate interactions with eukaryotic hosts and mutualistic bacterial partners. For example, quorum sensing regulates both phytoplankton bacteria interactions and bacterial communication (Coolahan & Whalen, 2025). In contrast, defense mechanisms such as toxins and antibiotics were activated in response to many companions, reflecting a generalized antagonistic strategy. This is consistent with previous findings that bacteria use such compounds for intra- and interspecies competition, responding rapidly to diverse biotic cues without companion-specific recognition (Hibbing et al., 2010). In lab studies, numerous interacting pairs were shown to interact through antagonism (Meyer et al., 2017), our results suggest that co-evolution may not be necessary for such interactions to occur. In eukaryotes, specific responses included cell cycle pathways, plant hormone signaling, and nucleic acid processes, while general responses included B vitamin pathways, transcription factors, two-component systems, and transporters. Surprisingly, cell cycle regulation emerged as

a specific rather than general response, contrasting with lab studies showing widespread cell cycle transcript changes during phytoplankton bacteria interactions (Bartolek et al., 2022; Pollara et al., 2021; Paul & Pohnert, 2011).Vitamin related pathways were specific in bacterial responders but part of a general response in eukaryotes. B vitamins mediate phytoplankton bacteria interactions, with many bacteria synthesizing vitamins essential for phytoplankton growth (Croft et al., 2005; Bertrand & Allen, 2012). Eukaryotes may respond broadly to any vitamin source due to its critical role in growth, while bacteria may regulate biosynthesis selectively, investing energy only when interacting with partners that provide reciprocal benefits.

Species interactions in the North Pacific were generally characterized by focused molecular responses rather than transcriptome-wide remodeling, with only a small fraction of the responder transcriptome involved (~0.5% median, 1.3% in highly interacting pairs). This contrasts with lab studies, where 14–25% of genes can be differentially expressed (Shibl et al., 2020; Durham et al., 2017). Both lab studies and our approach assume interactions occur through direct physical contact or exuded molecules (Seymour et al., 2017). However, while lab studies are often done in high-density conditions, our approach implicitly assumes interactions in the environment are dose dependent: if a companion is rare, exuded molecules may be too dilute or physical contact unlikely. In addition to differing assumptions in experimental design, constitutive expression due to long term co-occurrence, multiple simultaneous interactions and enhanced energetic constraints could likely explain the lower magnitude of responses observed in nature. Most species level interactions in highly interacting pairs were unidirectional, with few reciprocal responses, suggesting opportunistic rather than co-evolved associations, allowing taxa to flexibly exploit co-abundant partners for nutrients and cofactors without requiring highly specialized mutualistic capabilities (Morris et al., 2012; Lima-Mendez et al., 2015). One limitation of our method is that we cannot exclude the possibility that some of the unidirectional interactions that we observe are reciprocal, and we are unable to observe their signatures because of limitations of sequencing depth and annotation efficiency. However, as we do not find them in the highly interacting pairs, even if the reciprocal interaction is occurring in the environment, it is through a smaller remodeling of the transcription profile.

The novel statistical approach described here allows us to predict transcriptional mechanisms of interactions in the environment from field metatranscriptome data. We were able to detect both the strength of each function in all possible pairwise interactions as well as assess its significance. Our method can be applied to any metatranscriptomic dataset that contains multiple observations with varying species abundances. Similarly, our method is robust to missing species or functions across samples, and accounts for correlated samples such as due to technical replicates. In our study, we apply Enviromtx on assembled metatranscriptomes at the species level, annotated with KO functions. Alternatively, it is possible to apply this method at coarser taxonomic resolution provided that the subgroups are equally well-detected, as well as annotate at the gene level resolution by mapping metatranscriptomes to reference genomes. In its current formulation, Enviromtx detects most robust interaction features, but is not designed to study how interactions change across environmental conditions. However, there are numerous lab studies that demonstrate interactions between organisms can change based on external conditions, for example, interactions between phytoplankton and bacteria can shift from beneficial to pathogenic based on varying levels of nutrient availability, growth stage and temperature (Mayers et al., 2016; Seyedsayamdost et al., 2011; Wiener et al., n.d.; Lin et al., 2024). In the future, the model could be generalized to allow changes in transcript expression to

vary with abundances in an environmental-dependent manner. Our approach is based on the assumption that interactions occur between two species at a time, however, there are cases where 3-way species interactions have been described, or where higher order interactions are occurring (Harvey et al., 2023; Bairey et al., 2016). In complex natural environments, we expect this to be the case, and future modifications to our methods could incorporate more complex interaction networks. We view our approach as a hypothesis generating tool and a powerful way to observe general functional patterns of interactions in the environment, as well as propose specific species-species pairs that might be interesting candidates for more targeted lab studies.

Figures

Figure 2.1

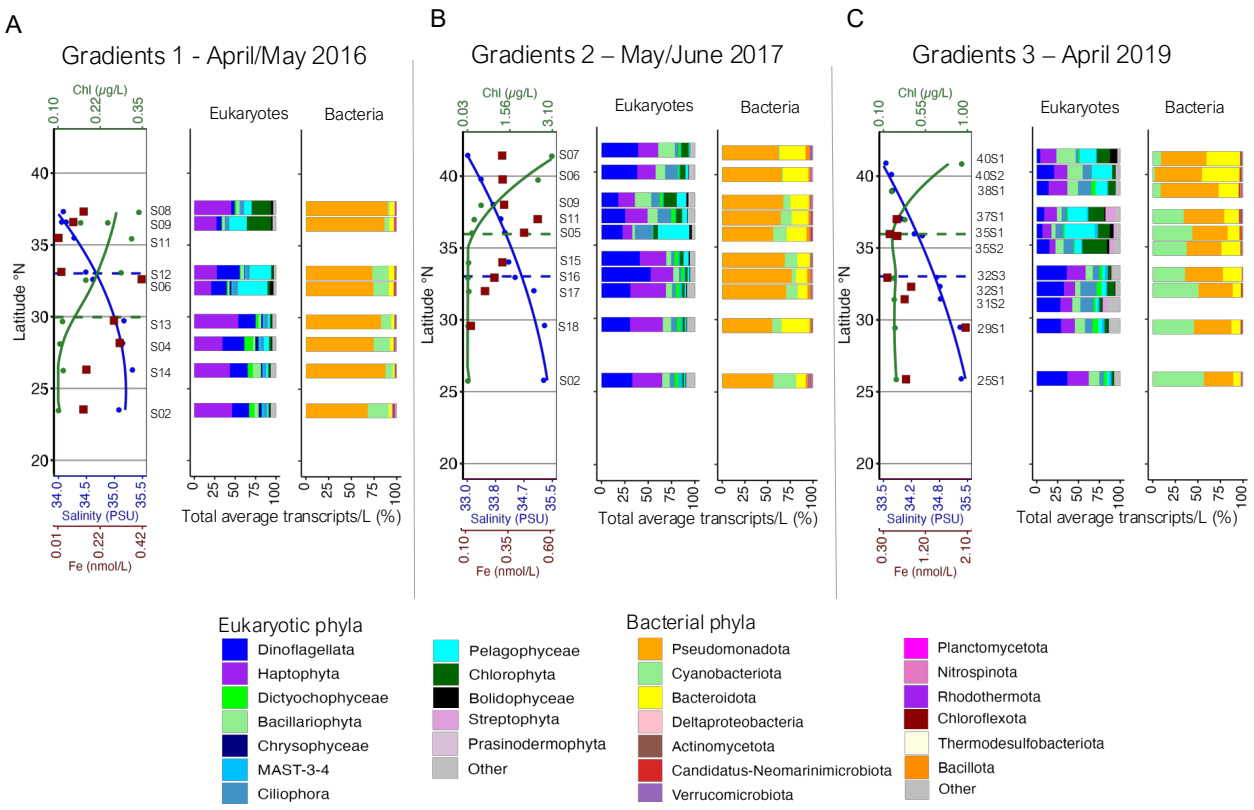


Fig. 2.1. Overview of Gradients 1, 2 and 3 metatranscriptomes. (A-C) Environmental conditions and metatranscriptome sampling sites (left-most plots), relative composition of dominant eukaryotic (middle plots) and bacterial (right-most plots) phyla for three cruises: Gradients 1 in 2016 (A), Gradients 2 in 2017 (B) and Gradients 3 in 2019 (C). Latitudinal distribution of salinity (filled blue circles and blue smoothed trend line), chlorophyll (chl, filled green circles and green smoothed trend line) and dissolved iron (Fe, brown filled squares, taken from underway measurements) concentrations (note change in scales between cruises) along the transects; green dashed line indicates the chlorophyll front (0.15 mg m^{-3}), and blue dashed line indicates the salinity front isohaline (34.82), station numbers (on right) indicate metatranscriptome sampling stations (Juranek et al., 2020; Pinedo-González et al., 2020; Park et al., 2023). Relative proportions of the top 10 eukaryotic phyla are derived from the polyA selected fraction; relative proportions of the top 10 bacterial phyla are derived from the rRNA-depleted fraction of community RNA samples. Relative dinoflagellate abundances were divided by a factor of 6.4 to account for their disproportionate transcript to biomass ratios (Coesel et al., 2025). Other eukaryotic and bacterial phyla comprise the sum of all organisms that are not in the top 10 most abundant groups.

Figure 2.2

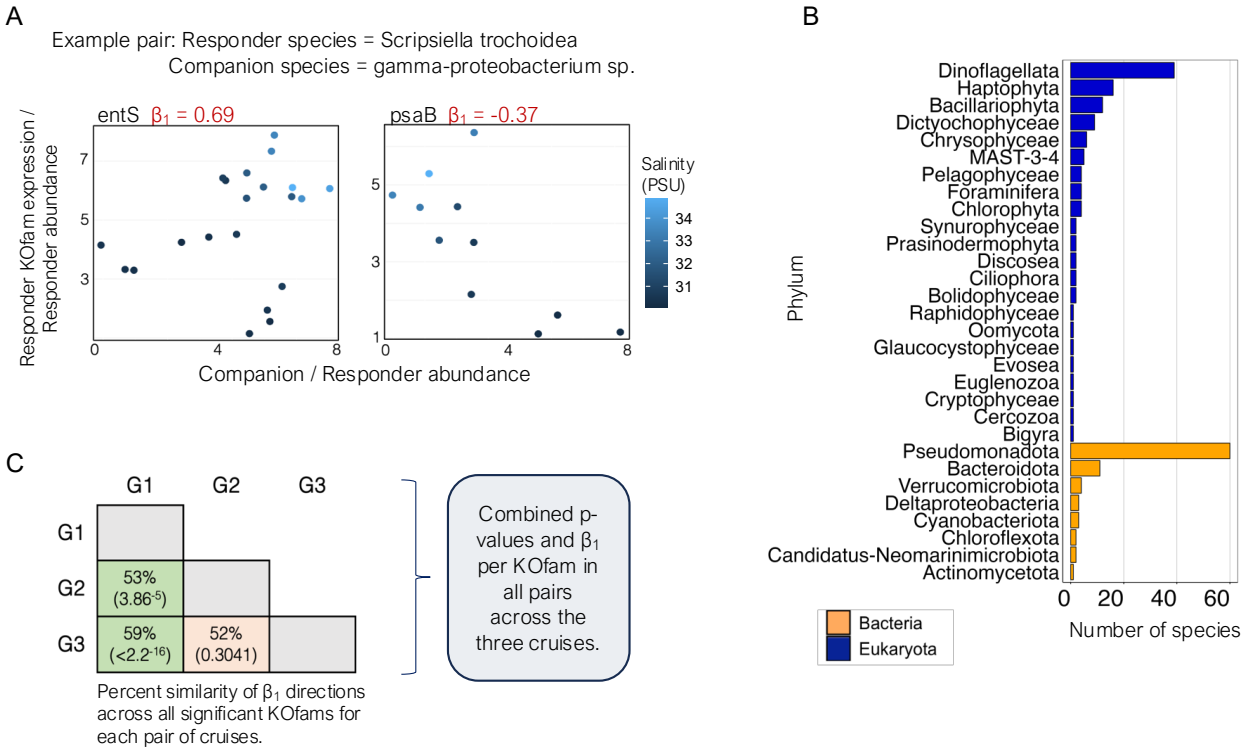


Fig 2.2. Enviromtx applied on cruises in the North Pacific. (A) Example output of the Enviromtx model, which tests for transcriptional interactions between species pairs. Expression of each responder KOfam (transcripts per KOfam) is analyzed individually for each responder/companion species pair in a marginal model of species co-abundances while conditioning for environmental factors. β_1 indicates the correlation across samples of KOfam expression for each responder/companion pair. Examples of β_1 calculated for expression of two KOfams (entS: enterobactin transporter KO8225; psaB: photosystem I P700 chlorophyll *a* apoprotein A2 K02690) for the responder *Scripsiella trochoidea* and the companion gamma-proteobacterium sp. The positive β_1 for entS indicates increased transcript abundance for this KOfam in *S. trochoidea* in response to an increased relative abundance of the gamma-proteobacterium sp.; the negative $\beta_1$for apo A2 indicates decreased transcript abundance for this KOfam in *S. trochoidea* in response to an increased relative abundance of the gamma-proteobacterium sp. Each dot represents a distinct metatranscriptome sample, colored by the salinity of the sampling site representing progression along the environmental gradient. **(B)** Phylum-level distribution of the 205 species level taxonomic bins with sufficient functional annotations across Gradients 1, and 3 (Bacteria – orange, Eukaryota – blue), resulting in tests of 41,820 total pairwise species interactions. **(C)** Percent similarity and p-value in parentheses (significant: green shading; not significant: orange shading) of proportion of positive vs negative β_1 for each cruise pair across all significant KOfams for all responder/companion species pairs.

Figure 2.3

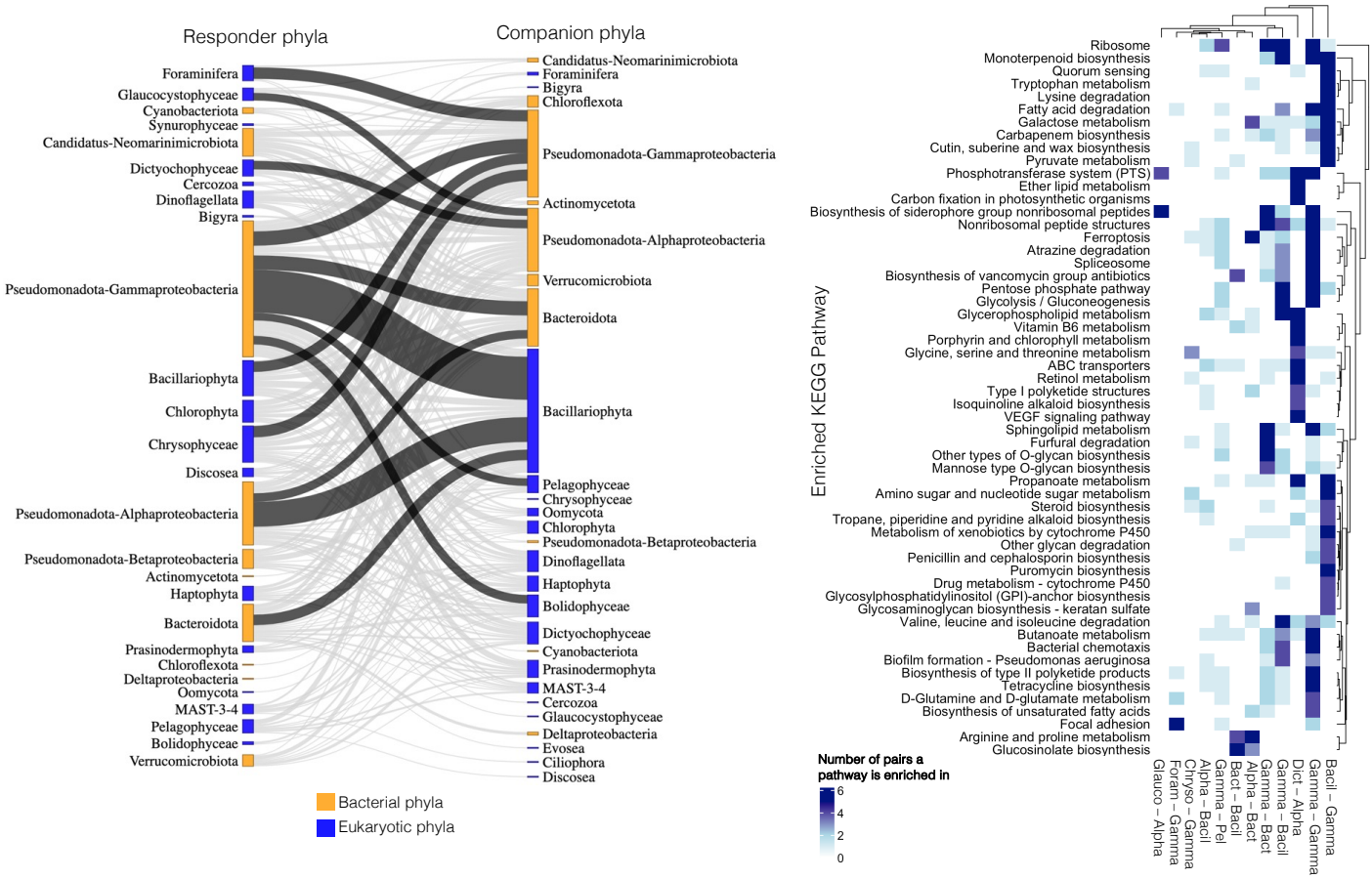


Fig. 2.3. Patterns of highly interacting phyla and their enriched KEGG pathways in the North Pacific. (A) Highly interacting species pairs grouped at the phylum level. The phyla on the left correspond to responder species and the phyla on the right correspond to companion species; colored bars indicate corresponding kingdom of the responder or companion phyla (Bacteria – yellow, Eukaryotes – blue). Colored bar height scales with the total number of highly interacting responder or companion species within a phylum. Connecting lines between the responder and companion phyla indicate highly interacting species pairs. Width of each connecting line scales with number of highly interacting responder/companion species pairs within a phylum. Black connecting lines indicate phyla with more than 7 highly interacting species; grey connecting lines indicate phyla with 1-6 highly interacting species. **(B)** Enriched KEGG pathways in the highly interacting species pairs, grouped at phylum level. Heatmap indicates number of species pairs a pathway is enriched in and is clustered by pathway and responder-companion phyla group. The x-axis labels indicate a responder – companion phyla group: Bacillariophyta responding to Gammaproteobacteria (Bacil– Gamma), Gammaproteobacteria responding to Gammaproteobacteria (Gamma – Gamma), Dictyochophyceae responding to Alphaproteobacteria (Dict – Alpha), Gammaproteobacteria responding to Bacillariophyta (Gamma – Bacil), Gammaproteobacteria responding to Bacteroidota (Gamma – Bact), Alphaproteobacteria responding to Bacteroidota (Alpha – Bact),

Bacteroidota responding to Bacillariophyta (Bact – Bacil), Gammaproteobacteria responding to Pelagophyceae (Gamma – Pel), Alphaproteobacteria responding to Bacillariophyta (Alpha – Bacil), Chrysophyceae responding to Gammaproteobacteria (Chryso – Gamma), Foraminifera responding to Gammaproteobacteria (Foram – Gamma), and Glaucocystophyceae responding to Alphaproteobacteria (Glauco – Alpha).

Figure 2.4

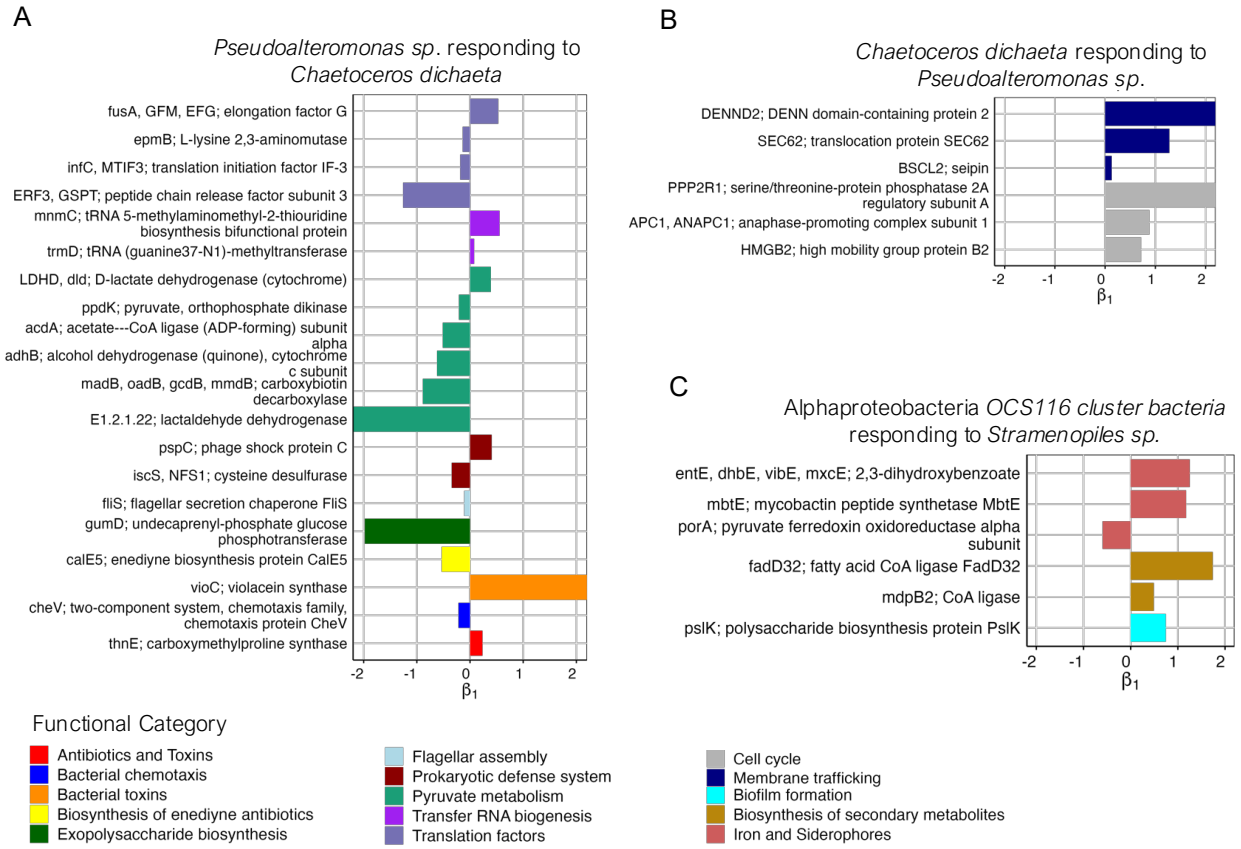


Fig. 2.4. Example reciprocal (A, B) or unidirectional (C, D) species level interactions with involved KOfams indicated by significant β_1 values. (A) *Pseudomonas sp. 520P1 No. 412* response to the eukaryote *Chaetoceros dichaeta* **(B)** *Chaetoceros dichaeta* response to *Pseudomonas sp. 520P1 No. 412,* **(C)** Alphaproteobacteria *OCS116 bacterium AG 390 A19* response to *Stramenopiles sp. TOSAG4-2.* Bar plots colored by functional category indicate sign and magnitude of β_1 values. Positive β_1 indicates increased transcript abundance for responder KOfam as relative abundance of companion increases, and negative β_1 indicates decreased transcript abundance for responder KOfam as relative abundance of companion increases.

Figure 2.5

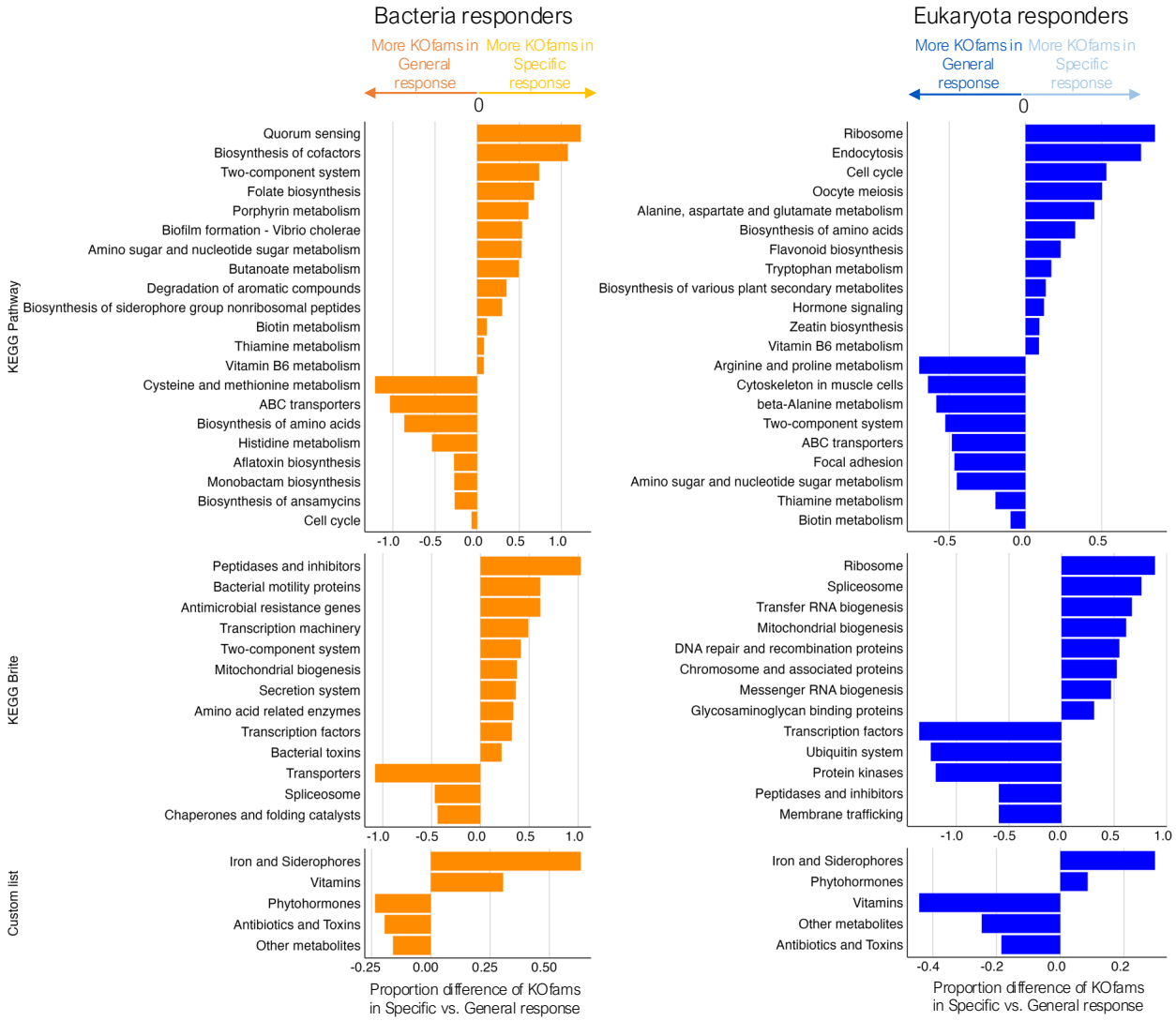
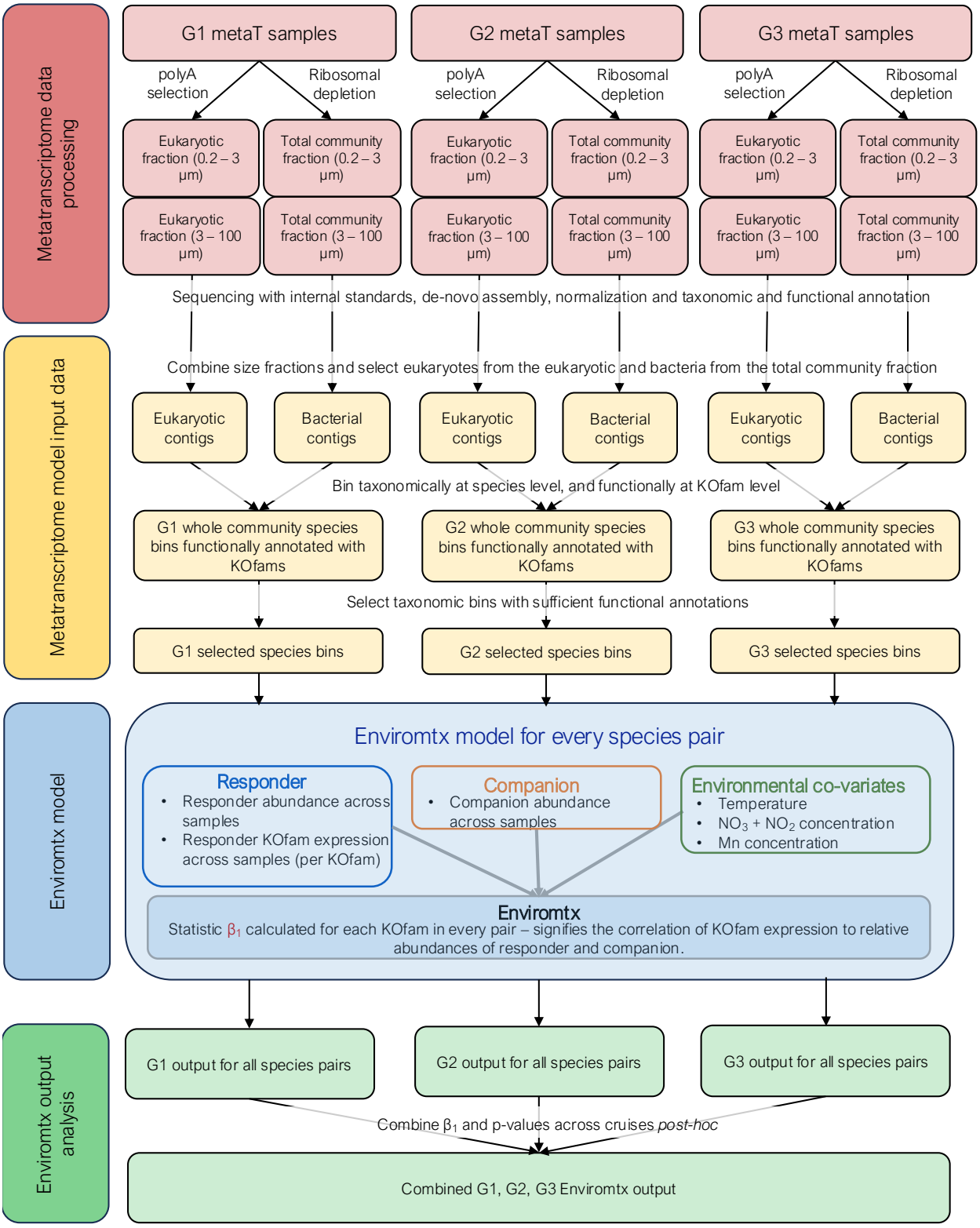


Fig. 2.5. Function features of interactions. A general response is defined as elicited by a greater than average number of companions; a specific response is defined as elicited by a lower than average number of companions. Proportion of KOfams within a pathway in bacteria (orange) or eukaryotes (blue) that are general (negative values) or specific (positive values); a value of zero indicates KOfams within a pathway are equally distributed between the two categories. Functional categories are grouped according to KEGG pathways, KEGG brite hierarchies, or a custom list.

Supplemental Figures and Tables
Sup 2.1



Sup 2.1. Flowchart shows four main steps taken in the analysis, processing metatranscriptome data from the North Pacific (red), formatting the data into a suitable format for analysis with

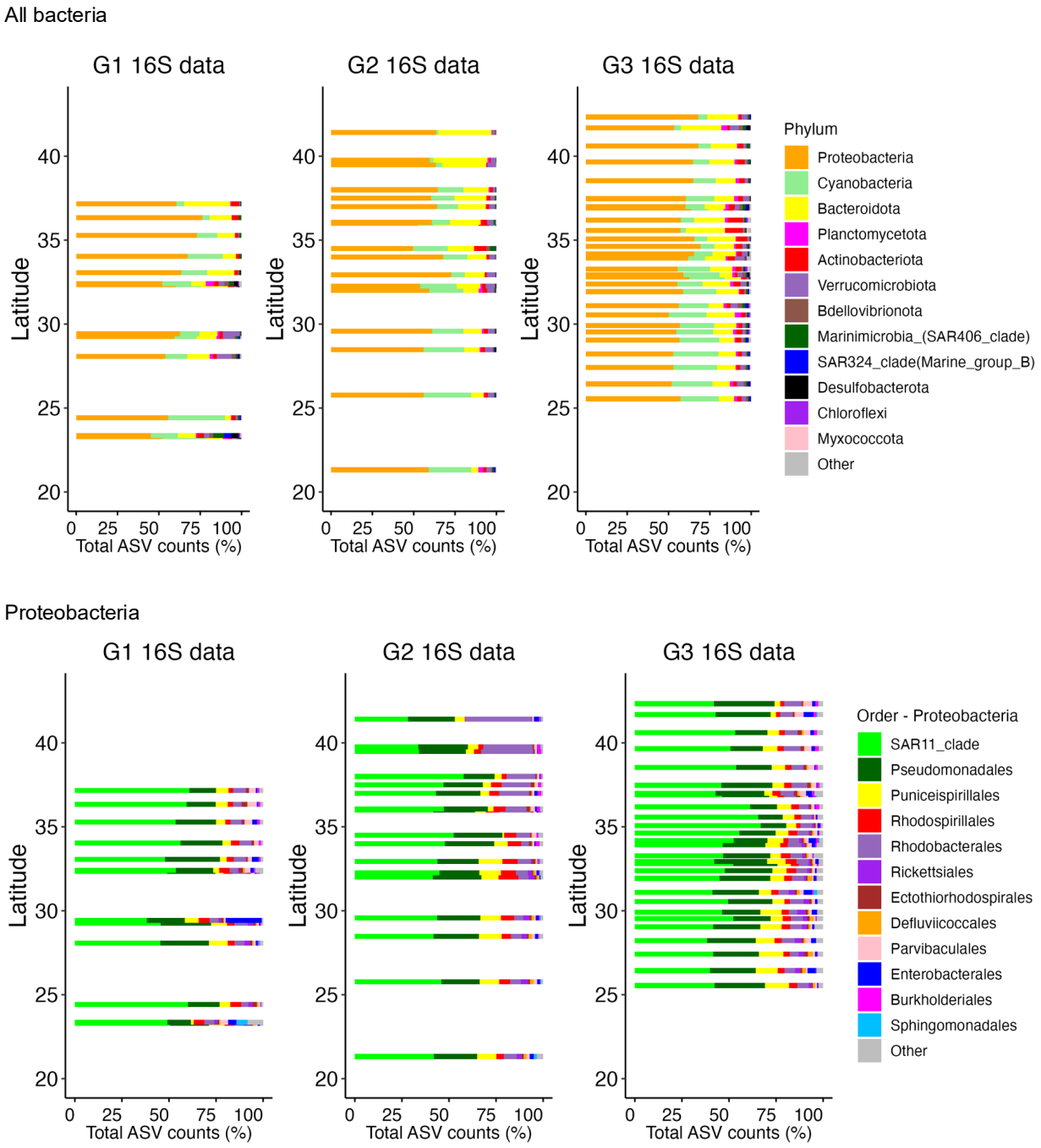
Enviromtx (yellow), applying Enviromtx on each pair separately for every cruise individually (green), and processing the resulting output for further analysis (green).

Sup 2.2

	Total number of run instances	Number that completed in gee (step1)	Average correaltion.Alpha of completed runs in gee	Number that completed with glm (step 2)
G1	34943181	30028760 (85.93597%)	-0.01424921	4914421 (14.06403%)
G2	35497583	33480358 (94.31729%)	0.05203436	2017225 (5.682711%)
G3	48396426	43656493 (90.20603%)	0.09103436	4739933 (9.793973%)
Total	118,837,190			

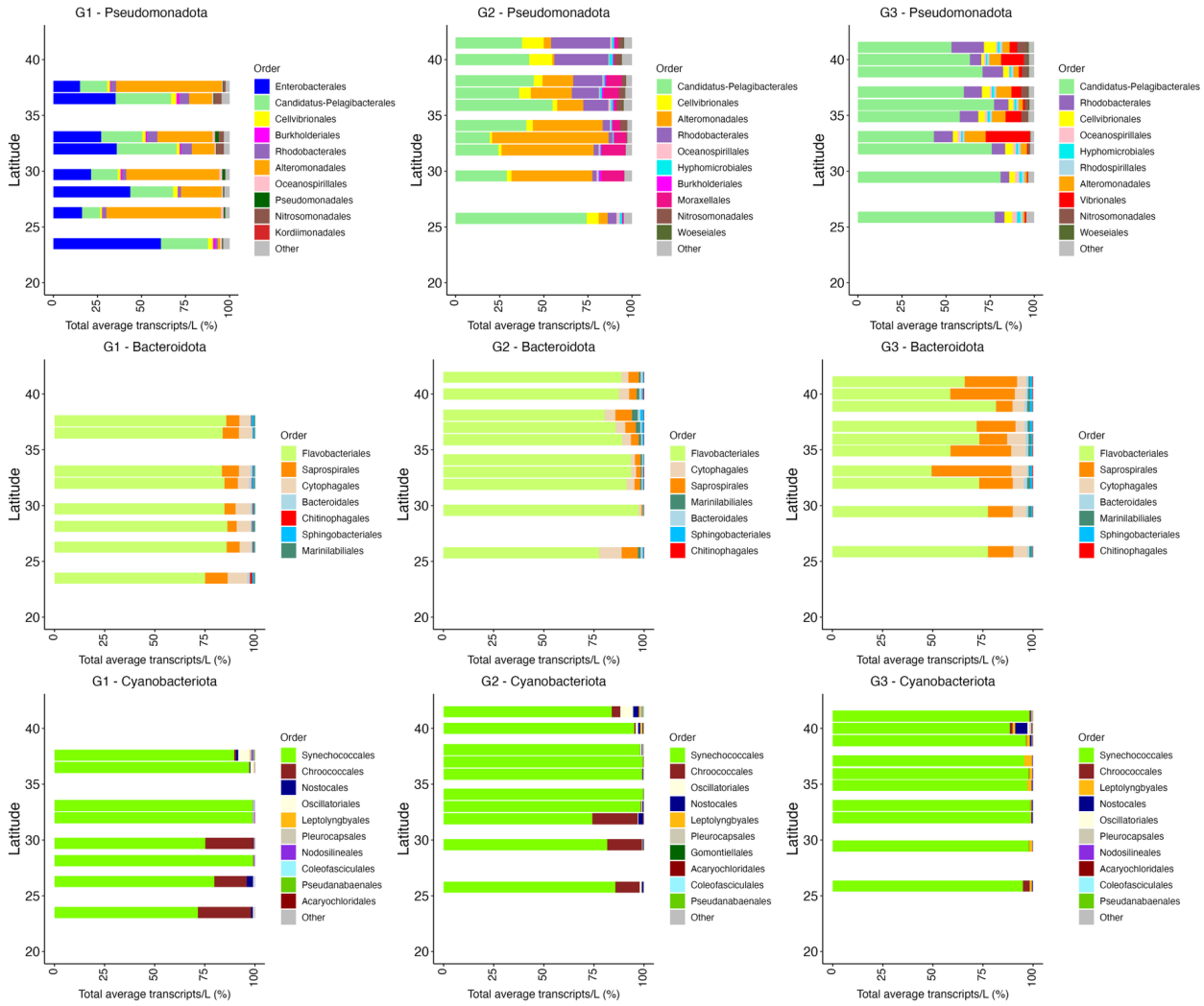
Sup. 2.2. Summary of Enviromtx instances (responder/companion/function) completed in two steps of model fitting. The first step was fitted with gee, and the second step was fitted with glm using the corresponding correlation.Alpha for each cruise.

Sup 2.3



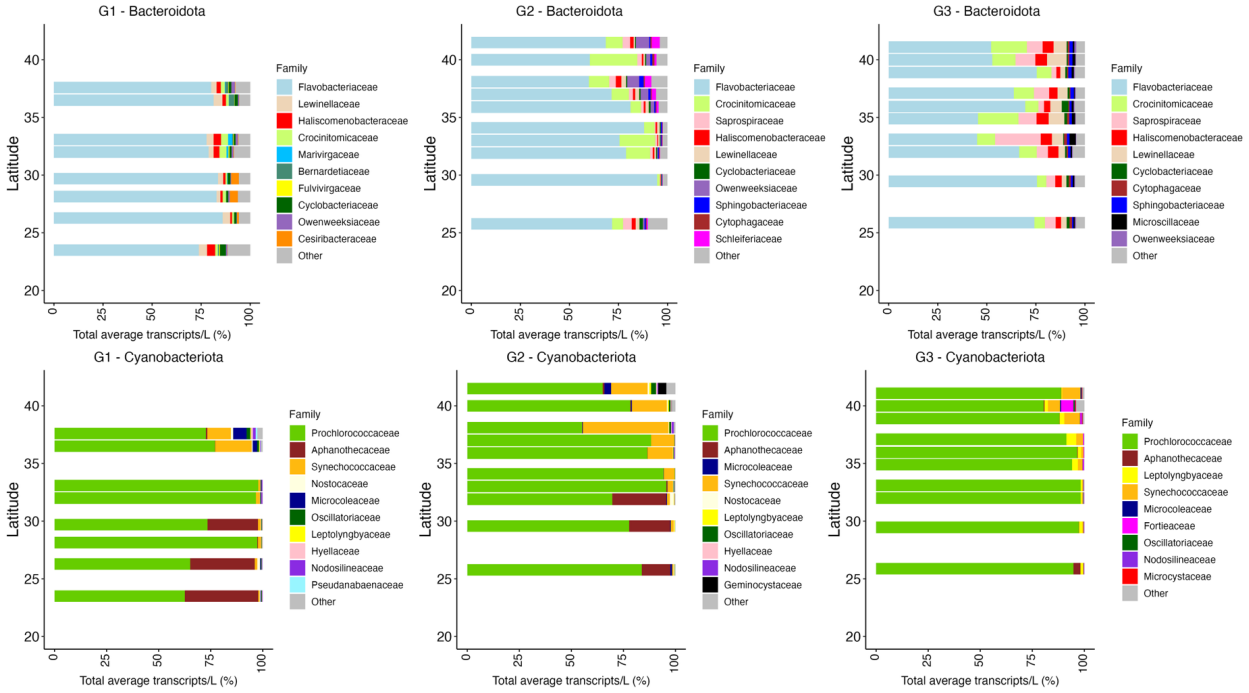
Sup. 2.3. Amplicon sequence variate (AVS) abundance data from 16S rRNA collected across the Gradients cruises. Data adopted from Key et al., 2025. Top panel shows relative ASV abundances of top 12 phylum level bacterial taxonomies across Gradients 1(left), Gradients 2 (middle) and Gradients 3 (right). Bottom panel shows relative abundances of top 12 order level taxonomies within the Proteobacteria phylum across Gradients 1(left), Gradients 2 (middle) and Gradients 3 (right).

Sup 2.4



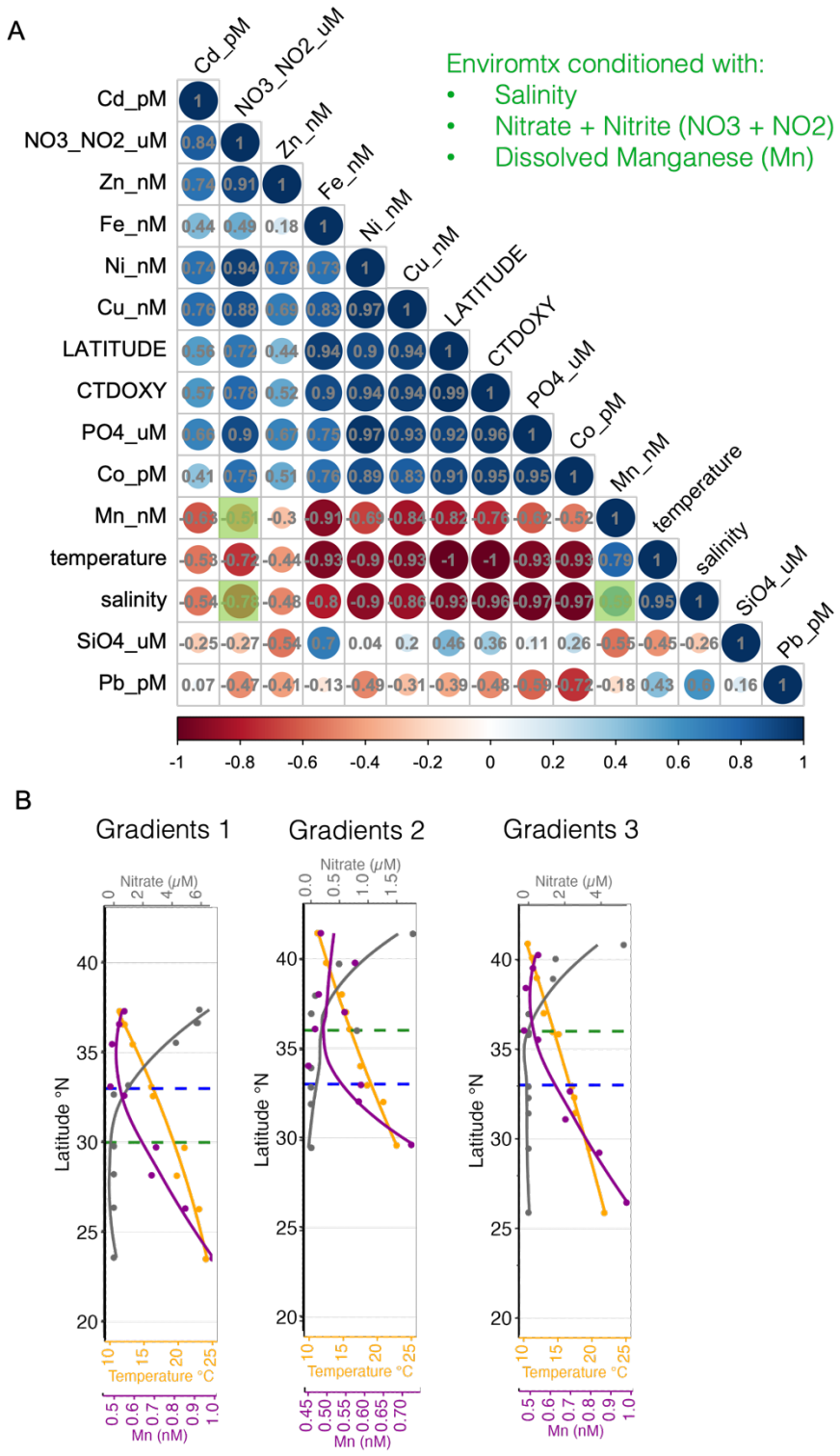
Sup. 2.4. Metatranscriptome abundances in total transcripts/L of taxonomic groups at the order level from three main bacterial phyla, Pseudomonadota (top row), Bacteroidota (middle row) and Cyanobacteriota (bottom row) across Gradients 1 (left), Gradients 2 (middle) and Gradients 3 (right). The top 10 most abundant orders within each phyla are plotted, and the rest are combined into the "Other" category.

Sup 2.5



Sup. 2.5. Metatranscriptome abundances in total transcripts/L of taxonomic groups at the family level from two bacterial phyla, Bacteroidota (top row) and Cyanobacteriota (bottom row) across Gradients 1 (left), Gradients 2 (middle) and Gradients 3 (right). The top 10 most abundant families within each phyla are plotted, and the rest are combined into the "Other" category.

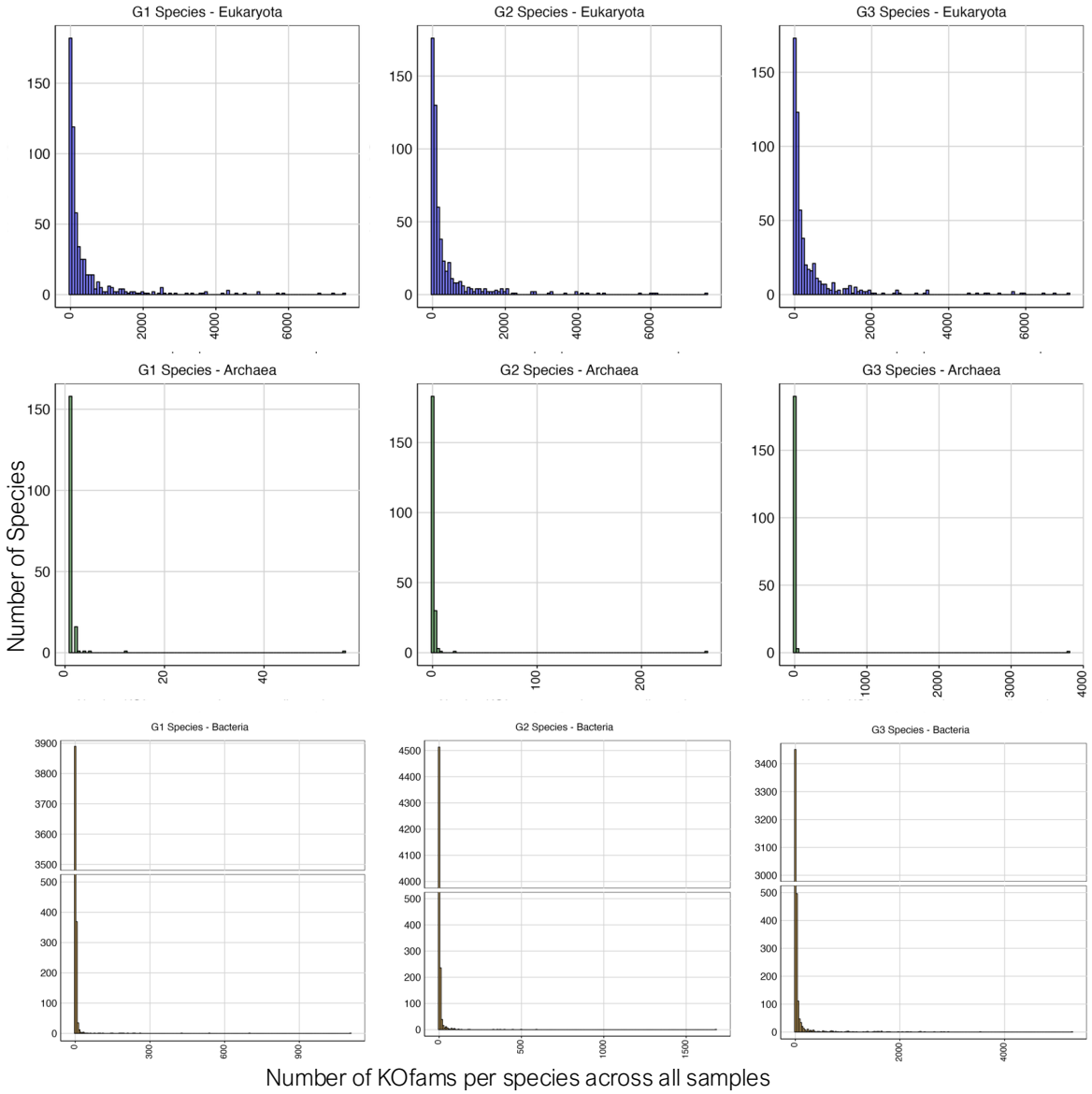
Sup 2.6



Sup. 2.6. Sup. Env correlations. (A) Pearson's correlation between measured abiotic features of the Gradients 1, 2 and 3 transects, data from all three cruises was combined. The color bar indicates the Pearson's correlation coefficient. The following features were measured: latitude in

°N, oxygen collected from CTD casts (CTDOXY), temperature in °C, salinity, Nitrate + Nitrite concentrations in uM (NO3_NO2_uM), Cadmium concentrations in pM (Cd_pM), Zinc concentrations in nM (Zn_nM), Iron concentrations in nM (Fe_nM), Nickel concentrations in nM (Ni_nM), Copper concentrations in nM (Cu_nM), Phosphate concentrations in uM (PO4_uM), Cobalt concentrations in pM (Co_pM), Manganese concentrations in nM (Md_nM), Silica concentrations in uM (Si_uM), Lead concentrations in pM (Pd_pM). (B) Latitudinal distributions of environmental parameters that were used to condition the model for Gradients 1(left), Gradients 2 (middle) and Gradients 3 (right). Showing temperature (orange circles and smoothed trend line), dissolved Manganese (Mn, purple circles and smoothed trend line), and dissolved nitrate and nitrite (Nitrate, gray circles and smoothed trend line). Green dashed line indicates the chlorophyll front (0.15 mg m^{-3}), and blue dashed line indicates the salinity front isohaline (34.82).

Sup 2.7



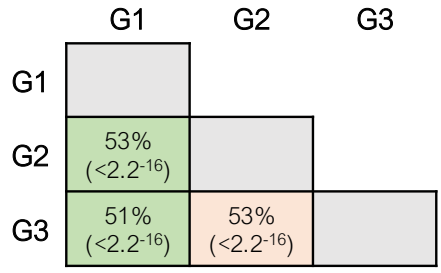
Dataset	Number of species	Number of species with functional annotations (>500 KOfams for PA; >300 KOfams for NS)
G1 NS – bacteria, archaea	4338	4
G1 PA – unicellular eukaryotes	453 (unicellular)	96 (unicellular)
G2 NS - bacteria, archaea	4856	8
G2 PA - unicellular eukaryotes	456 (unicellular)	94 (unicellular)
G3 NS - bacteria, archaea	4284	83
G3 PA - unicellular eukaryotes	452 (unicellular)	95 (unicellular)
Total unique across cruises PA	463 (unicellular)	119 (unicellular)
Total unique across cruises NS	5207	86
Total PA +NS		205

Sup. 2.7. Plots are showing the number of KOfam annotations per species in Gradients 1(left), Gradients 2 (middle) and Gradients 3 (right) separated by eukaryotes (top) archaea (middle) and

bacteria (bottom). Eukaryotic species are taken from the eukaryotic community (polyA selected) fraction of the samples, while the Bacteria and Archaea species are taken from the total community (ribosomally depleted) fraction. Table is showing the total number of species in each dataset, as well as the number of species with sufficient level of functional annotation (> 500 KOfams for eukaryotes and > 300 KOfams for prokaryotes). In the eukaryotic fraction all multicellular organisms were filtered out.

Sup. 2.8

All β_1 (significant and insignificant)

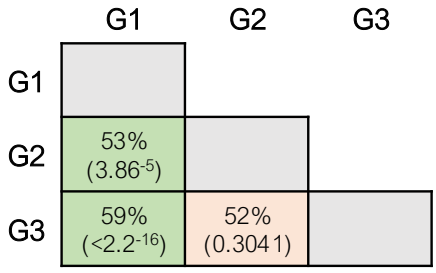


+ve β_1 G2	5044891	5108836
-ve β_1 G2	6265117	5446628
	-ve β_1 G1	+ve β_1 G1

+ve β_1 G3	4173736	4114293
-ve β_1 G3	5339384	4816228
	-ve β_1 G1	+ve β_1 G1

+ve β_1 G3	4378814	4449113
-ve β_1 G3	5941888	4659720
	-ve β_1 G2	+ve β_1 G2

Significant β_1 (p-value <0.1)



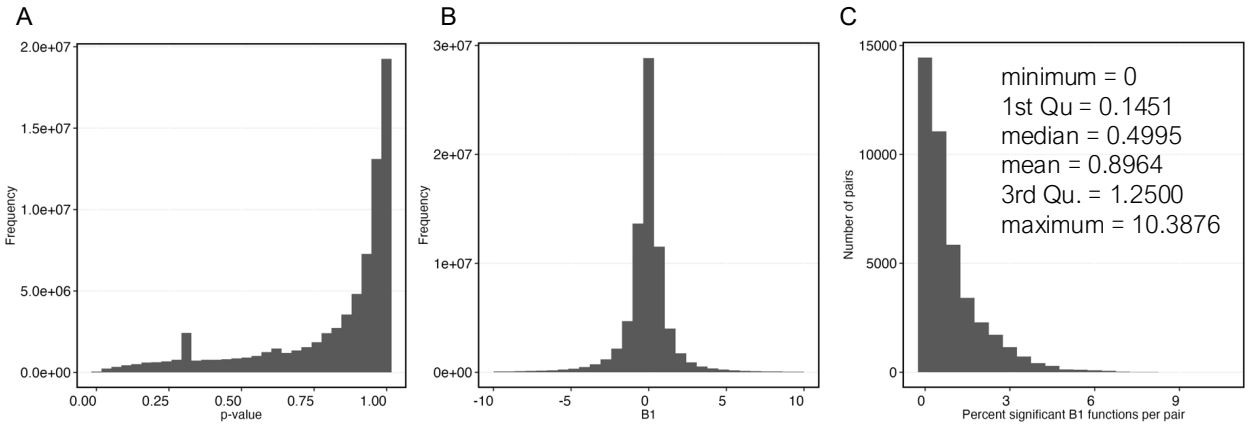
+ve β_1 G2	709	891
-ve β_1 G2	933	883
	-ve β_1 G1	+ve β_1 G1

+ve β_1 G3	557	944
-ve β_1 G3	913	723
	-ve β_1 G1	+ve β_1 G1

+ve β_1 G3	20	30
-ve β_1 G3	42	45
	-ve β_1 G2	+ve β_1 G2

Sup. 2.8. Internal validation of Enviromtx output comparing the direction of β_1 across three cruises. Percent similarity in direction is shown for each cruise combination (top panel) with p-value significance shown in brackets, for all responder/companion/function instances (left), and only significant β_1 (p-val < 0.1) responder/companion/function instances. Tables show corresponding number of KOfams with positive β_1 in one cruise and positive β_1 in another cruise, negative β_1 in one cruise and negative β_1 in another cruise, positive β_1 in one cruise and negative β_1 in another cruise, and negative β_1 in one cruise and positive β_1 in another cruise.

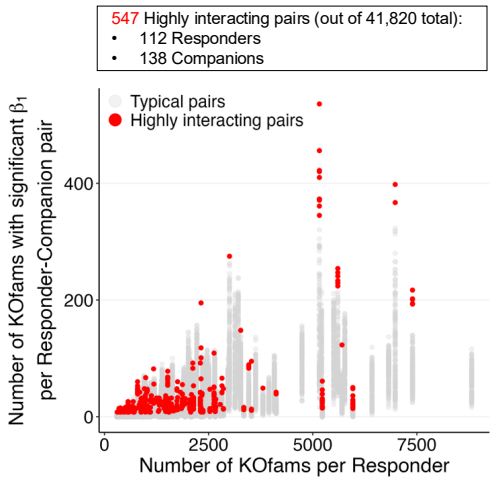
Sup 2.9



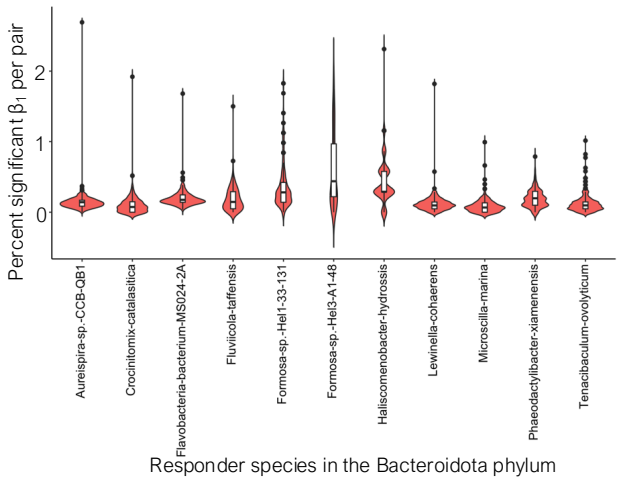
Sup. 2.9. (A) Histogram of p-values for the combined results of Enviromtx output across Gradients 1, 2 and 3. **(B)** Histogram of β_1 values for combined results for Enviromtx output across cruises. **(C)** Histogram showing the distribution of the percent significant β_1 functions per pair.

Sup 2.10

A



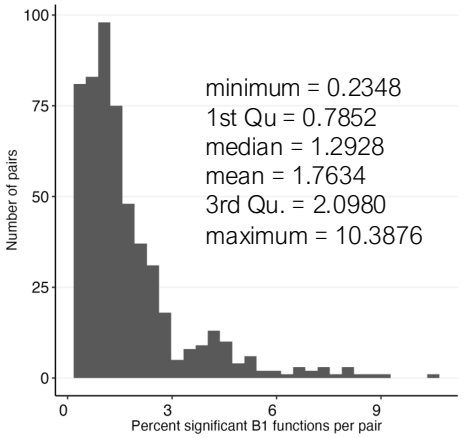
B



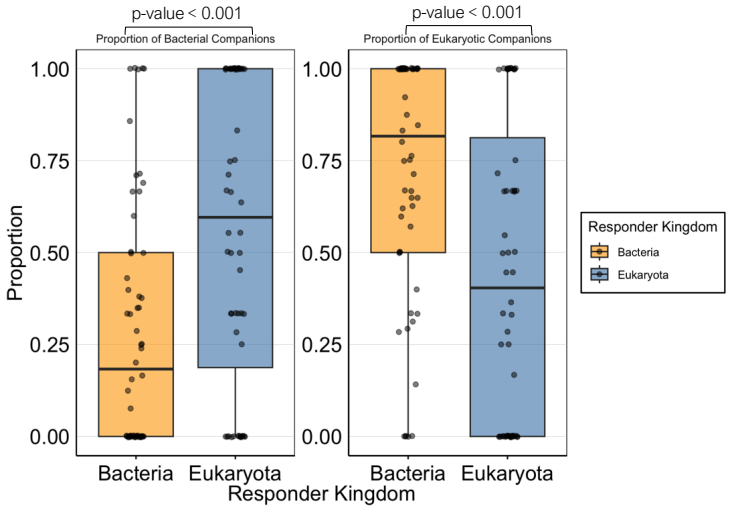
C

Step	Metrics	# pairs retained	# responders	# companions
1. Select top companions per each responder	Based on number of significant functions per pair, find the outliers for each responder (Q3 + 1.5 * IQR).	1376	142	197
2. Remove pairs below median number of significant B1 across entire dataset	Median number of significant B1 per pair across entire dataset = 7	693	123	151
3. Remove pairs with less than the median of enriched pathways across entire dataset	Median number of enriched pathways per pair across entire dataset = 5	547	112	138

D



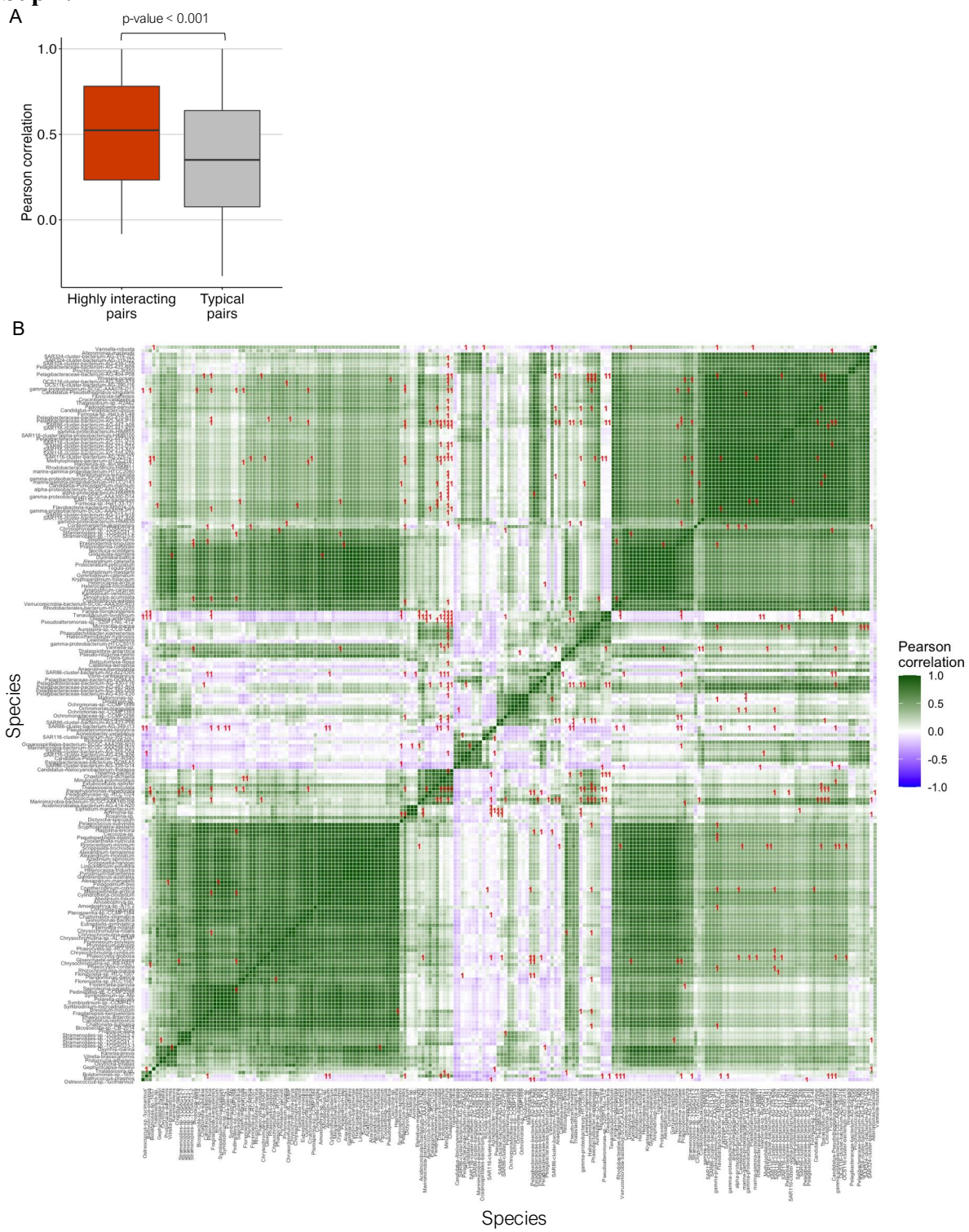
E



Sup. 2.10. Highly interacting pairs. (A) Highly interacting pairs (red dots) are selected as those with outlier number of KOfams with significant β_1 in responder companion pairs per each responder species. All other pairs are colored grey. A total of 547 highly interacting pairs is

identified consisting of 112 responders and 138 companion species. **(B)** Example distribution of the number of significant estimates per pair for each responder in an example phyla (Bacteroidota). Outlier companions for each responder are shown with dots. **(C)** Steps and metrics used to select highly interacting species pairs, along with the number of pairs, responders and companions retained after each step. **(D)** Histogram of the percent significant β_1 KOfams per pair in the highly interacting pairs. **(E)** Proportion of bacteria companions (left) and eukaryotic companions (right) across responders in highly interacting species pairs. Wilcoxson test was used to assess significance of proportion of bacteria and eukaryotic companions across all bacteria vs eukaryotic responders

Sup 2.11

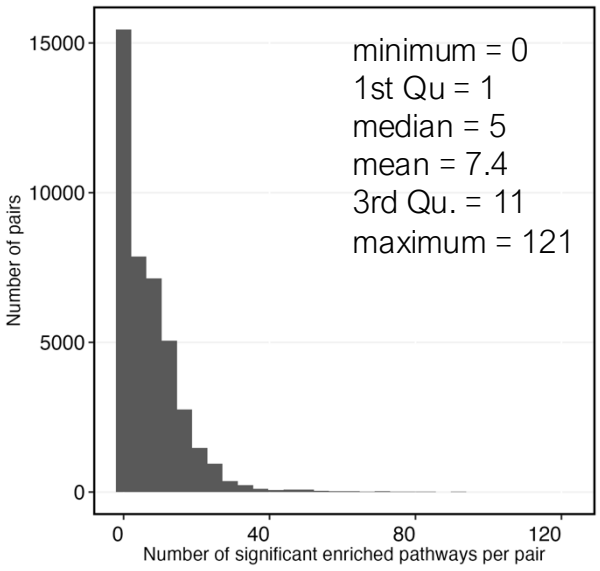


Sup. 2.11. (A) Summary of Pearson's correlation coefficient based on co-occurrence of species abundances in total transcripts/L in highly interacting pairs compared to typical pairs. One-sided t-test was done to determine significance. **(B)** Heatmap of Pearson's correlation for each species pair across the combined dataset of Gradients 1, 2 and 3 based on total transcripts/L for each species. Red crosses indicate which pairs are highly interacting.

Sup. 2.12

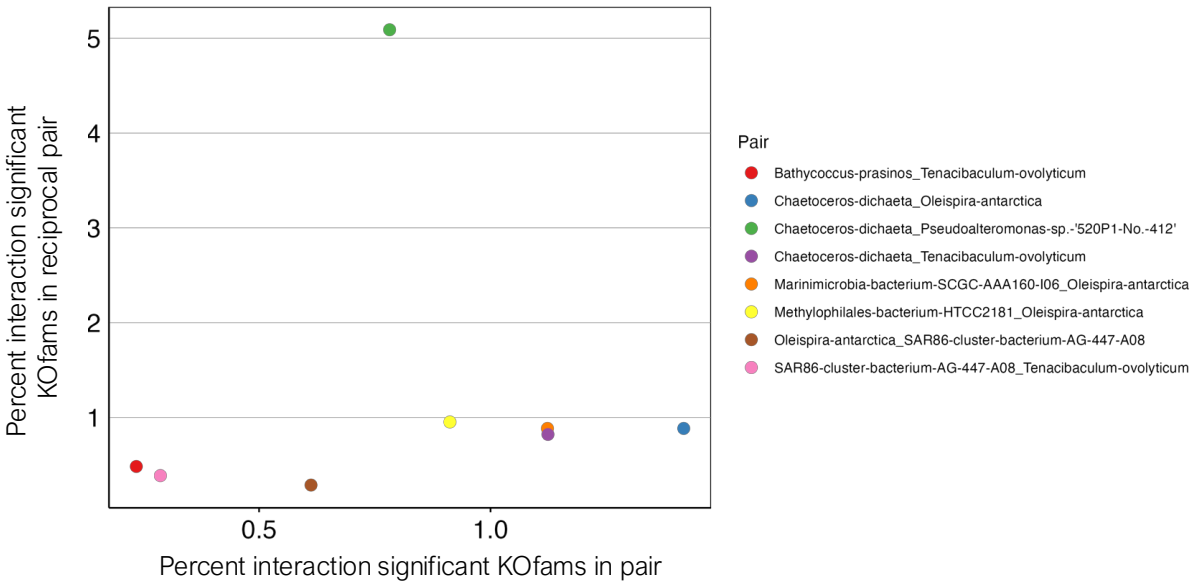
Sup. 2.12. Number of companions belonging to each group, either **(A)** kingdom, or **(B)** phylum per responder species in the highly interacting pairs, separated by bacterial responders on the left and eukaryotic responders on the right.

Sup. 2.13



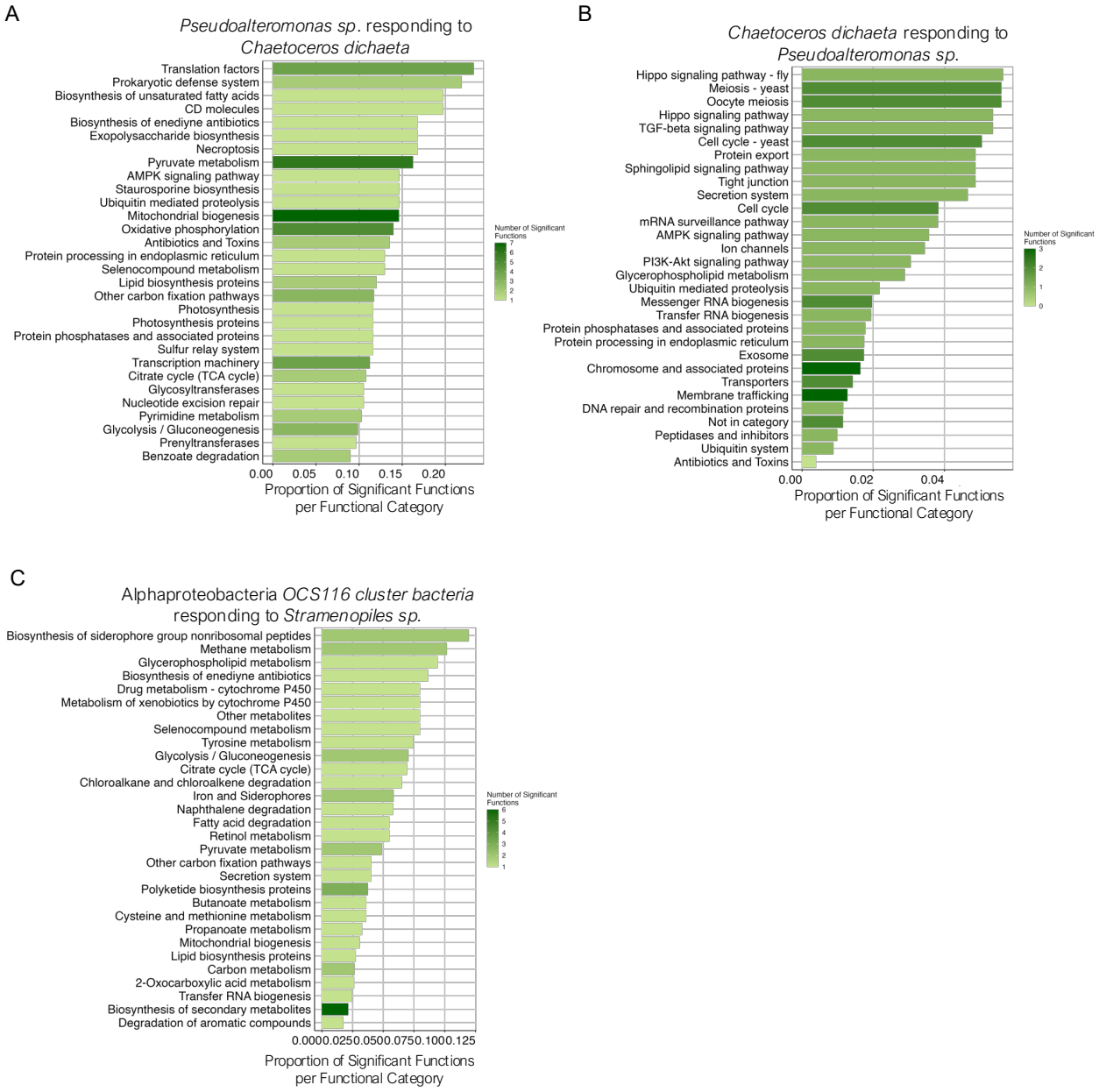
Sup. 2.13. Histogram of number of significantly enriched KEGG pathways based on significant Enviromtx KOfams per pair across Gradients 1, 2 and 3.

Sup 2.14



Sup. 2.14. Reciprocal pairs. Highly interacting pairs that have a reciprocal response, where the responder responds highly to the companion and vice versa. Percent of significant KOfams involved in the interaction is shown for each reciprocal pair. Each species pair is notated as responder_companion, where responder and companion species names are separated by an underscore.

Sup 2.15

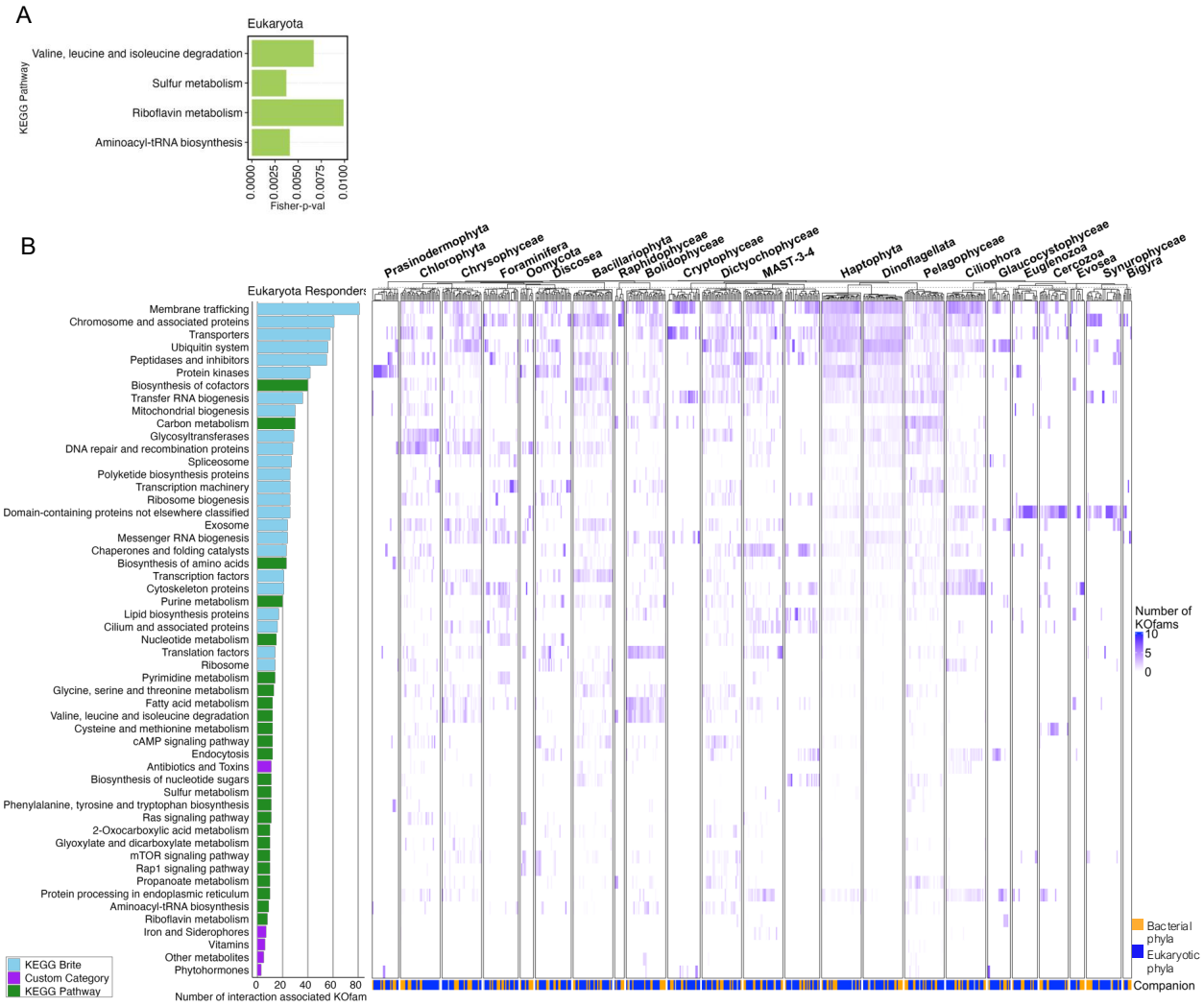


Sup. 2.15. Main functional categories involved in example species-species interactions for three example pairs. For each example pair, the proportion of significant KOfams per functional category induced in the responder is shown, where the color bar shows the total number of significant functions per category. **(A, B)** Reciprocal response between the bacteria *Pseudomonas sp. 520P1 No. 412*, and the eukaryote *Chaetoceros dichaeta.* **(C)** Unidirectional response of the bacteria cluster *OCS116 bacterium AG 390 A19* to *Stramenopiles sp. TOSAG4-2.*

Sup 2.16

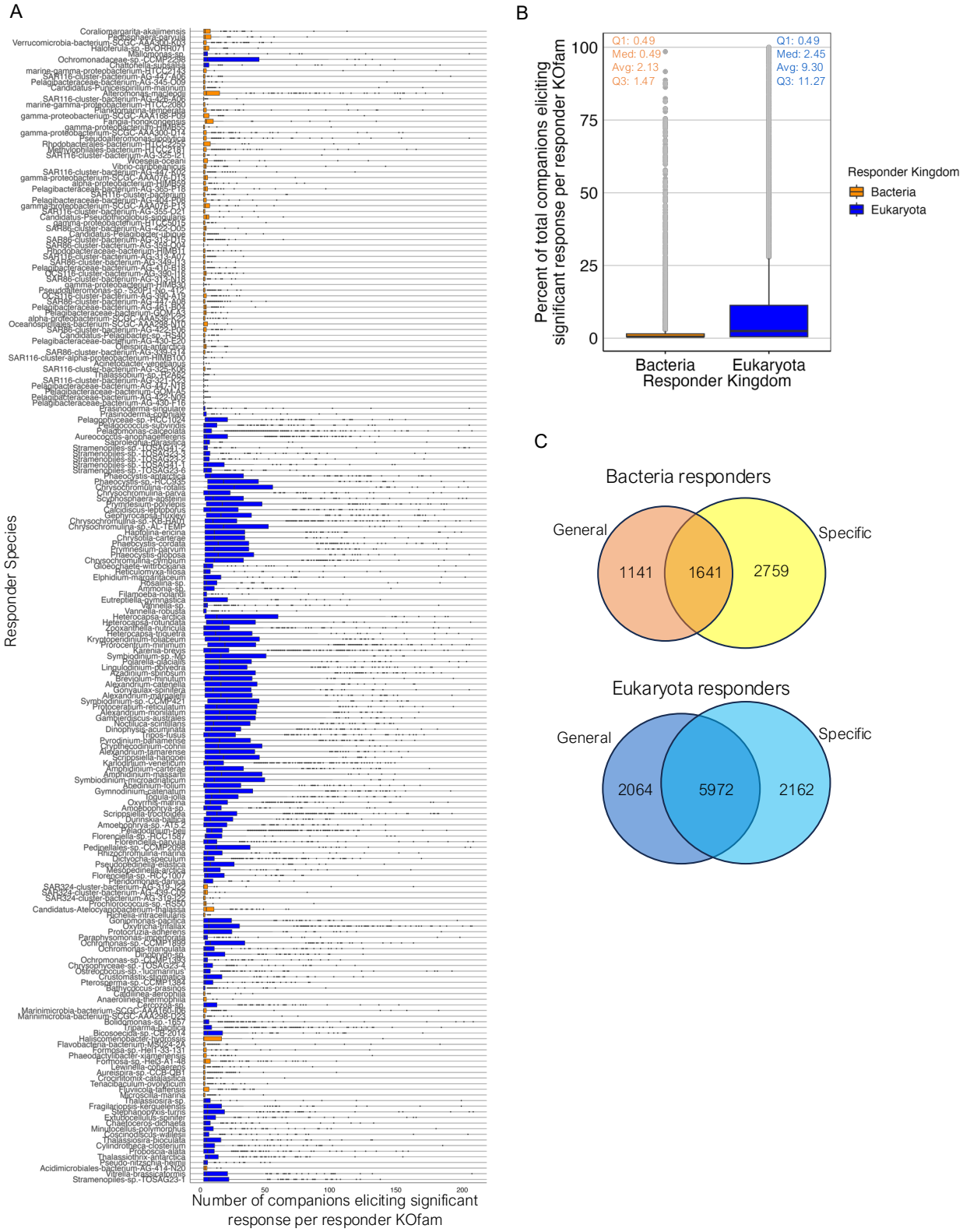
Sup. 2.16. Common functional features used by bacterial responders in interactions. KOfams commonly involved in an interaction response across responder species based on the Wilcoxson rank sum test are grouped into functional categories. **(A)** Enriched KEGG pathways in the set of common interaction response KOfams. **(C)** The number of interaction significant KOfams associated with each functional category across responders is shown in the bar plot on the left, colored by type of KEGG or custom functional category. The heatmap shows on the right shows the number of KOfams that are significantly involved in interactions to all companions separated by responders at phylum level. Within each responder phyla, the heatmap is clustered by response to companion phyla. The kingdom of the companions eliciting the functional response is shown in the colored bar on the bottom.

Sup 2.17



Sup. 2.17. Common functional features used by eukaryotic responders in interactions. KOfams commonly involved in an interaction response across responder species based on the Wilcoxson rank sum test are grouped into functional categories. **(A)** Enriched KEGG pathways in the set of common interaction response KOfams. **(C)** The number of interaction significant KOfams associated with each functional category across responders is shown in the bar plot on the left, colored by type of KEGG or custom functional category. The heatmap shows on the right shows the number of KOfams that are significantly involved in interactions to all companions separated by responders at phylum level. Within each responder phyla, the heatmap is clustered by response to companion phyla. The kingdom of the companions eliciting the functional response is shown in the colored bar on the bottom.

Sup 2.18



Sup. 2.18. Consistent vs. specific response. (A) Number of companions eliciting significant response per each responder KOfam. All responder species are shown separately in the y-axis,

and a bar plots show the distribution of the number of companions eliciting a significant interaction response across all KOfams for each responder. Bars are colored by responder kingdom (orange - bacteria, blue – eukaryotes). **(B)** Showing the distribution of percent companions eliciting a significant interaction response per responder KOfam grouped at kingdom level. **(C)** Venn diagram showing the number of general response and specific response KOfams in bacterial (left) and eukaryotic (right) responders, in response to all companions. The number of KOfams that is always specific, always general or sometimes specific and sometimes general across responders is shown.

Sup. Table 2.1. Summary statistics of assembled metatranscriptome contigs and annotation efficiencies. Eukaryotic community (polyA) summary statistics were adapted from Groussman et al., 2024. Total number of assembled nucleotide contigs for each dataset, along with the number of unique clustered contigs at 99% amino acid identity level. Annotation efficiency for taxonomic and functional annotations including both KOfam and Pfam annotations is stated.

Sup. Table 2.2. List of metatranscriptome samples taken in the Gradients 1, 2 and 3 cruises. Both Eukaryotic (polyA) and Total community (NS) samples are listed. The samples were combined by size fraction and sequencing type (polyA or NS) for analysis. Missing samples and incomplete samples are indicated; those were excluded from the analysis.

Sup. Table 2.3. List of KOfams categories into custom functional categories. Custom functional categories were constructed for Iron and Siderophores, Other metabolites, Vitamins, Antibiotics and Toxins, and Phytohormones based on literature. The KOfams that were included in each category are listed.

Sup. Table 2.4. Species level taxonomic bins used in analysis with Enviromtx. All species bins that had sufficient annotations for analysis with Enviromtx are listed along with taxonomic classification.

Sup. Table 2.5. Highly interacting pairs. List of highly interacting pairs found through Enviromtx, along with basic characteristic for each pair, including taxonomic lineage of the responder and companion taxonomic bins, total number of KOfams with significant B1 in a pair (num_sig_estimate_pair), total number of KOfam annotations in responder (num_unique_funct_resp), percent KOfams with significant B1 in pair (percent_sig_b1_pair), and number of significantly enriched KEGG pathways in pair (num_sig_enrich_path_pair).

Sup. Table 2.6. Enriched KEGG pathways per pair. For each pair analyzed with Enviromtx, list of significantly enriched KEGG pathways and the associated Fisher Test p-values.

References

Agrawal, S., Acharya, D., Adholeya, A., Barrow, C. J., & Deshmukh, S. K. (2017). Nonribosomal Peptides from Marine Microbes and Their Antimicrobial and Anticancer Potential. *Frontiers in Pharmacology*, *8*, 828. https://doi.org/10.3389/fphar.2017.00828

Ahern, O. M., Whittaker, K. A., Williams, T. C., Hunt, D. E., & Rynearson, T. A. (2021). Host genotype structures the microbiome of a globally dispersed marine phytoplankton. *Proceedings of the National Academy of Sciences*, *118*(48), e2105207118. https://doi.org/10.1073/pnas.2105207118

Amin, S. A., Green, D. H., Hart, M. C., Küpper, F. C., Sunda, W. G., & Carrano, C. J. (2009). Photolysis of iron–siderophore chelates promotes bacterial–algal mutualism. *Proceedings of the National Academy of Sciences*, *106*(40), 17071–17076. https://doi.org/10.1073/pnas.0905512106

Amin, S. A., Hmelo, L. R., Van Tol, H. M., Durham, B. P., Carlson, L. T., Heal, K. R., Morales, R. L., Berthiaume, C. T., Parker, M. S., Djunaedi, B., Ingalls, A. E., Parsek, M. R., Moran, M. A., & Armbrust, E. V. (2015). Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. *Nature*, *522*(7554), 98–101. https://doi.org/10.1038/nature14488

Amin, S. A., Parker, M. S., & Armbrust, E. V. (2012). Interactions between Diatoms and Bacteria. *Microbiology and Molecular Biology Reviews*, *76*(3), 667–684. https://doi.org/10.1128/MMBR.00007-12

Ayers, J. M., & Lozier, M. S. (2010). Physical controls on the seasonal migration of the North Pacific transition zone chlorophyll front. *Journal of Geophysical Research: Oceans*, *115*(C5), 2009JC005596. https://doi.org/10.1029/2009JC005596

Azam, F., & Malfatti, F. (2007). Microbial structuring of marine ecosystems. *Nature Reviews Microbiology*, *5*(10), 782–791. https://doi.org/10.1038/nrmicro1747

Bairey, E., Kelsic, E. D., & Kishony, R. (2016). High-order species interactions shape ecosystem diversity. *Nature Communications*, *7*(1), 12285. https://doi.org/10.1038/ncomms12285

Barak-Gavish, N., Frada, M. J., Ku, C., Lee, P. A., DiTullio, G. R., Malitsky, S., Aharoni, A., Green, S. J., Rotkopf, R., Kartvelishvily, E., Sheyn, U., Schatz, D., & Vardi, A. (2018). Bacterial virulence against an oceanic bloom-forming phytoplankter is mediated by algal DMSP. *SCIENCE ADVANCES*.

Bartolek, Z., Van Creveld, S. G., Coesel, S., Cain, K. R., Schatz, M., Morales, R., & Virginia Armbrust, E. (2022). Flavobacterial exudates disrupt cell cycle progression and metabolism of the diatom *Thalassiosira pseudonana*. *The ISME Journal*, *16*(12), 2741–2751. https://doi.org/10.1038/s41396-022-01313-9

Becker, J. W., Hogle, S. L., Rosendo, K., & Chisholm, S. W. (2019). Co-culture and biogeography of *Prochlorococcus* and SAR11. *The ISME Journal*, *13*(6), 1506–1519. https://doi.org/10.1038/s41396-019-0365-4

Bertrand, E. M., & Allen, A. E. (2012). Influence of vitamin B auxotrophy on nitrogen metabolism in eukaryotic phytoplankton. *Frontiers in Microbiology*, *3*. https://doi.org/10.3389/fmicb.2012.00375

Bland, M. J., & Altman, D. G. (1995). *Multiple significance tests: The Bonferroni method*.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bray, N. L., Pimentel, H., Melsted, P., & Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology*, *34*(5), 525–527. https://doi.org/10.1038/nbt.3519

Buchan, A., LeCleir, G. R., Gulvik, C. A., & González, J. M. (2014). Master recyclers: Features and functions of bacteria associated with phytoplankton blooms. *Nature Reviews Microbiology*, *12*(10), 686–698. https://doi.org/10.1038/nrmicro3326

Christie-Oleza, J. A., Sousoni, D., Lloyd, M., Armengaud, J., & Scanlan, D. J. (2017). Nutrient recycling facilitates long-term stability of marine microbial phototroph–heterotroph interactions. *Nature Microbiology*, *2*(9), 17100. https://doi.org/10.1038/nmicrobiol.2017.100

Church, M. J., Björkman, K. M., Karl, D. M., Saito, M. A., & Zehr, J. P. (2008). Regional distributions of nitrogen-fixing bacteria in the Pacific Ocean. *Limnology and Oceanography*, *53*(1), 63–77. https://doi.org/10.4319/lo.2008.53.1.0063

Cirri, E., & Pohnert, G. (2019). Algae−bacteria interactions that balance the planktonic microbiome. *New Phytologist*, *223*(1), 100–106. https://doi.org/10.1111/nph.15765

Coesel, S. N., Van Creveld, S. G., Dugenne, M., Henderikx-Freitas, F., White, A. E., & Armbrust, E. V. (2025). Proportional relationship between transcript concentrations and carbon biomass for open ocean plankton groups. *The ISME Journal*, *19*(1), wraf079. https://doi.org/10.1093/ismejo/wraf079

Coolahan, M., & Whalen, K. E. (2025). A review of quorum-sensing and its role in mediating interkingdom interactions in the ocean. *Communications Biology*, *8*(1), 179. https://doi.org/10.1038/s42003-025-07608-9

Coyne, K. J., Wang, Y., & Johnson, G. (2022). Algicidal Bacteria: A Review of Current Knowledge and Applications to Control Harmful Algal Blooms. *Frontiers in Microbiology*, *13*, 871177. https://doi.org/10.3389/fmicb.2022.871177

Croft, M. T., Lawrence, A. D., Raux-Deery, E., Warren, M. J., & Smith, A. G. (2005). Algae acquire vitamin B12 through a symbiotic relationship with bacteria. *Nature*, *438*(7064), 90–93. https://doi.org/10.1038/nature04056

Di Costanzo, F., Di Dato, V., & Romano, G. (2023). Diatom–Bacteria Interactions in the Marine Environment: Complexity, Heterogeneity, and Potential for Biotechnological Applications. *Microorganisms*, *11*(12), 2967. https://doi.org/10.3390/microorganisms11122967

DiMucci, D., Kon, M., & Segrè, D. (2021). BowSaw: Inferring Higher-Order Trait Interactions Associated With Complex Biological Phenotypes. *Frontiers in Molecular Biosciences*, *8*, 663532. https://doi.org/10.3389/fmolb.2021.663532

Durham, B. P., Boysen, A. K., Carlson, L. T., Groussman, R. D., Heal, K. R., Cain, K. R., Morales, R. L., Coesel, S. N., Morris, R. M., Ingalls, A. E., & Armbrust, E. V. (2019). Sulfonate-based networks between eukaryotic phytoplankton and heterotrophic bacteria in the surface ocean. *Nature Microbiology*, *4*(10), 1706–1715. https://doi.org/10.1038/s41564-019-0507-5

Durham, B. P., Boysen, A. K., Heal, K. R., Carlson, L. T., Boccamazzo, R., Deodato, C. R., Qin, W., Cattolico, R. A., Armbrust, E. V., & Ingalls, A. E. (2022). Chemotaxonomic patterns in intracellular metabolites of marine microbial plankton. *Frontiers in Marine Science*, *9*, 864796. https://doi.org/10.3389/fmars.2022.864796

Durham, B. P., Dearth, S. P., Sharma, S., Amin, S. A., Smith, C. B., Campagna, S. R., Armbrust, E. V., & Moran, M. A. (2017). Recognition cascade and metabolite transfer in a marine bacteria-phytoplankton model system. *Environmental Microbiology*, *19*(9), 3500–3513. https://doi.org/10.1111/1462-2920.13834

Durham, B. P., Johnson, W. M., Bannon, C. C., Bertrand, E. M., Ingalls, A. E., Edwards, B. R., Apprill, A., Boysen, A. K., Bundy, R. M., Chen, H., Ferrer-González, F. X., Fiore, C., Heal, K. R., Kuhlisch, C., Liu, S., Lu, K., Meke, L. E., Pontrelli, S., Vaiyapuri Ramalingam, P., … Kujawinski, E. B. (2025). An ecological framework for microbial metabolites in the ocean ecosystem. *Limnology and Oceanography Letters*, lol2.70046. https://doi.org/10.1002/lol2.70046

Durham, B. P., Sharma, S., Luo, H., Smith, C. B., Amin, S. A., Bender, S. J., Dearth, S. P., Van Mooy, B. A. S., Campagna, S. R., Kujawinski, E. B., Armbrust, E. V., & Moran, M. A. (2015). Cryptic carbon and sulfur cycling between surface ocean plankton. *Proceedings of the National Academy of Sciences*, *112*(2), 453–457. https://doi.org/10.1073/pnas.1413137112

Dutkiewicz, S., Follett, C. L., Follows, M. J., Henderikx-Freitas, F., Ribalet, F., Gradoville, M. R., Coesel, S. N., Farnelid, H., Finkel, Z. V., Irwin, A. J., Jahn, O., Karl, D. M., Mattern, J. P., White, A. E., Zehr, J. P., & Armbrust, E. V. (2024). Multiple biotic interactions establish phytoplankton community structure across environmental gradients. *Limnology and Oceanography*, *69*(5), 1086–1100. https://doi.org/10.1002/lno.12555

Eddy, S. R. (2011). Accelerated Profile HMM Searches. *PLoS Computational Biology*, *7*(10), e1002195. https://doi.org/10.1371/journal.pcbi.1002195

Falkowski, P. G., Fenchel, T., & Delong, E. F. (2008). The Microbial Engines That Drive Earth's Biogeochemical Cycles. *Science*, *320*(5879), 1034–1039. https://doi.org/10.1126/science.1153213

Ferrer-González, F. X., Widner, B., Holderman, N. R., Glushka, J., Edison, A. S., Kujawinski, E. B., & Moran, M. A. (2021). Resource partitioning of phytoplankton metabolites that support bacterial heterotrophy. *The ISME Journal*, *15*(3), 762–773. https://doi.org/10.1038/s41396-020-00811-y

Focardi, A., Bramucci, A. R., Ajani, P., Khalil, A., Raina, J.-B., & Seymour, J. R. (2025). Defining the ecological strategies of phytoplankton associated bacteria. *Nature Communications*, *16*(1), 6363. https://doi.org/10.1038/s41467-025-61523-5

Follett, C. L., Dutkiewicz, S., Forget, G., Cael, B. B., & Follows, M. J. (2021). Moving ecological and biogeochemical transitions across the North Pacific. *Limnology and Oceanography*, *66*(6), 2442–2454. https://doi.org/10.1002/lno.11763

Fondi, M., & Di Patti, F. (2019). A synthetic ecosystem for the multi-level modelling of heterotroph-phototroph metabolic interactions. *Ecological Modelling*, *399*, 13–22. https://doi.org/10.1016/j.ecolmodel.2019.02.012

Fuhrman, J. A. (2009). Microbial community structure and its functional implications. *Nature*, *459*(7244), 193–199. https://doi.org/10.1038/nature08058

Gómez-Consarnau, L., Sachdeva, R., Gifford, S. M., Cutter, L. S., Fuhrman, J. A., Sañudo-Wilhelmy, S. A., & Moran, M. A. (2018). Mosaic patterns of B-vitamin synthesis and utilization in a natural marine microbial community. *Environmental Microbiology*, *20*(8), 2809–2823. https://doi.org/10.1111/1462-2920.14133

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., Di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., … Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, *29*(7), 644–652. https://doi.org/10.1038/nbt.1883

Gradoville, M. R., Farnelid, H., White, A. E., Turk-Kubo, K. A., Stewart, B., Ribalet, F., Ferrón, S., Pinedo-Gonzalez, P., Armbrust, E. V., Karl, D. M., John, S., & Zehr, J. P. (2020). Latitudinal constraints on the abundance and activity of the cyanobacterium UCYN-A and other marine diazotrophs in the North Pacific. *Limnology and Oceanography*, *65*(8), 1858–1875. https://doi.org/10.1002/lno.11423

Groussman, R. D., Blaskowski, S., Coesel, S. N., & Armbrust, E. V. (2023). MarFERReT, an open-source, version-controlled reference library of marine microbial eukaryote functional genes. *Scientific Data*, *10*(1), 926. https://doi.org/10.1038/s41597-023-02842-4

Groussman, R. D., Coesel, S. N., Durham, B. P., Schatz, M. J., & Armbrust, E. V. (2024). The North Pacific Eukaryotic Gene Catalog of metatranscriptome assemblies and annotations. *Scientific Data*, *11*(1), 1161. https://doi.org/10.1038/s41597-024-04005-5

Guidi, L., Chaffron, S., Bittner, L., Eveillard, D., Larhlimi, A., Roux, S., Darzi, Y., Audic, S., Berline, L., Brum, J. R., Coelho, L. P., Espinoza, J. C. I., Malviya, S., Sunagawa, S., Dimier, C., Kandels-Lewis, S., Picheral, M., Poulain, J., Searson, S., … Gorsky, G. (2016). Plankton

networks driving carbon export in the oligotrophic ocean. *Nature*, *532*(7600), 465–470. https://doi.org/10.1038/nature16942

Harvey, E., Yang, H., Castiblanco, E., Coolahan, M., Dallmeyer-Drennen, G., Fukuda, N., Greene, E., Gonsalves, M., Smith, S., & Whalen, K. (2023). Quorum sensing signal disrupts viral infection dynamics in the coccolithophore Emiliania huxleyi. *Aquatic Microbial Ecology*, *89*, 75–86. https://doi.org/10.3354/ame01998

Hawco, N. J., Conway, T. M., Coesel, S. N., Barone, B., Seelen, E. A., Yang, S.-C., Bundy, R. M., Pinedo-Gonzalez, P., Bian, X., Sieber, M., Lanning, N. T., Fitzsimmons, J. N., Foreman, R. K., König, D., Groussman, M. J., Allen, J. G., Juranek, L. W., White, A. E., Karl, D. M., … John, S. G. (2025). Anthropogenic iron alters the spring phytoplankton bloom in the North Pacific transition zone. *Proceedings of the National Academy of Sciences*, *122*(23), e2418201122. https://doi.org/10.1073/pnas.2418201122

Hibbing, M. E., Fuqua, C., Parsek, M. R., & Peterson, S. B. (2010). Bacterial competition: Surviving and thriving in the microbial jungle. *Nature Reviews Microbiology*, *8*(1), 15–25. https://doi.org/10.1038/nrmicro2259

Janda, J. M., & Abbott, S. L. (2021). The Changing Face of the Family *Enterobacteriaceae* (Order: " *Enterobacterales* "): New Members, Taxonomic Issues, Geographic Expansion, and New Diseases and Disease Syndromes. *Clinical Microbiology Reviews*, *34*(2), e00174-20. https://doi.org/10.1128/CMR.00174-20

Juranek, L. W., White, A. E., Dugenne, M., Henderikx Freitas, F., Dutkiewicz, S., Ribalet, F., Ferrón, S., Armbrust, E. V., & Karl, D. M. (2020). The Importance of the Phytoplankton "Middle Class" to Ocean Net Community Production. *Global Biogeochemical Cycles*, *34*(12), e2020GB006702. https://doi.org/10.1029/2020GB006702

Kanehisa, M., & Goto, S. (n.d.). *KEGG: Kyoto Encyclopedia of Genes and Genomes*.

Key, R. (n.d.). *Picophytoplankton Implicated in Productivity and Biogeochemistry in the North Pacific Transition Zone*.

King, K., Bramucci, A. R., Labbate, M., Raina, J.-B., & Seymour, J. R. (2022). Heterogeneous Growth Enhancement of Vibrio cholerae in the Presence of Different Phytoplankton Species. *Applied and Environmental Microbiology*, *88*(17), e01158-22. https://doi.org/10.1128/aem.01158-22

Lambert, B. S., Groussman, R. D., Schatz, M. J., Coesel, S. N., Durham, B. P., Alverson, A. J., White, A. E., & Armbrust, E. V. (2022). The dynamic trophic architecture of open-ocean protist communities revealed through machine-guided metatranscriptomics. *Proceedings of the National Academy of Sciences*, *119*(7), e2100916119. https://doi.org/10.1073/pnas.2100916119

Levering, J., Broddrick, J., Dupont, C. L., Peers, G., Beeri, K., Mayers, J., Gallina, A. A., Allen, A. E., Palsson, B. O., & Zengler, K. (2016). Genome-Scale Model Reveals Metabolic Basis of Biomass Partitioning in a Model Diatom. *PLOS ONE*, *11*(5), e0155038. https://doi.org/10.1371/journal.pone.0155038

Levine, J. M., Bascompte, J., Adler, P. B., & Allesina, S. (2017). Beyond pairwise mechanisms of species coexistence in complex communities. *Nature*, *546*(7656), 56–64. https://doi.org/10.1038/nature22898

Lima-Mendez, G., Faust, K., Henry, N., Decelle, J., Colin, S., Carcillo, F., Chaffron, S., Ignacio-Espinosa, J. C., Roux, S., Vincent, F., Bittner, L., Darzi, Y., Wang, J., Audic, S., Berline, L., Bontempi, G., Cabello, A. M., Coppola, L., Cornejo-Castillo, F. M., … Raes, J. (2015). Determinants of community structure in the global plankton interactome. *Science*, *348*(6237), 1262073. https://doi.org/10.1126/science.1262073

Lin, T., Feng, Y., Miao, W., Wang, S., Bao, Z., Shao, Z., Zhang, D., Wang, X., Jiang, H., & Zhang, H. (2024). Elevated temperature alters bacterial community from mutualism to antagonism with *Skeletonema costatum*: Insights into the role of a novel species, *Tamlana* sp. MS1. *mSphere*, *9*(7), e00198-24. https://doi.org/10.1128/msphere.00198-24

Liu, F., Gledhill, M., Tan, Q.-G., Zhu, K., Zhang, Q., Salaün, P., Tagliabue, A., Zhang, Y., Weiss, D., Achterberg, E. P., & Korchev, Y. (2022). Phycosphere pH of unicellular nano- and micro- phytoplankton cells and consequences for iron speciation. *The ISME Journal*, *16*(10), 2329–2336. https://doi.org/10.1038/s41396-022-01280-1

Lücken, L., Follows, M. J., Bragg, J. G., & Lennartz, S. T. (2025). *The microbial community model MCoM 1.0: A scalable framework for modelling phototroph-heterotrophic interactions in diverse microbial communities*. Biogeosciences. https://doi.org/10.5194/egusphere-2025-2227

Magnúsdóttir, S., Heinken, A., Kutt, L., Ravcheev, D. A., Bauer, E., Noronha, A., Greenhalgh, K., Jäger, C., Baginska, J., Wilmes, P., Fleming, R. M. T., & Thiele, I. (2017). Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nature Biotechnology*, *35*(1), 81–89. https://doi.org/10.1038/nbt.3703

Mayers, T. J., Bramucci, A. R., Yakimovich, K. M., & Case, R. J. (2016). A Bacterial Pathogen Displaying Temperature-Enhanced Virulence of the Microalga Emiliania huxleyi. *Frontiers in Microbiology*, *7*. https://doi.org/10.3389/fmicb.2016.00892

McLaren, M. R., Willis, A. D., & Callahan, B. J. (2019). Consistent and correctable bias in metagenomic sequencing experiments. *eLife*, *8*, e46923. https://doi.org/10.7554/eLife.46923

Meyer, N., Bigalke, A., Kaulfuß, A., & Pohnert, G. (2017). Strategies and ecological roles of algicidal bacteria. *FEMS Microbiology Reviews*, *41*(6), 880–899. https://doi.org/10.1093/femsre/fux029

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., Tosatto, S. C. E., Paladin, L., Raj, S., Richardson, L. J., Finn, R. D., & Bateman, A. (2021). Pfam: The protein families database in 2021. *Nucleic Acids Research*, *49*(D1), D412–D419. https://doi.org/10.1093/nar/gkaa913

Morris, J. J., Lenski, R. E., & Zinser, E. R. (2012). The Black Queen Hypothesis: Evolution of Dependencies through Adaptive Gene Loss. *mBio*, *3*(2), e00036-12. https://doi.org/10.1128/mBio.00036-12

Park, J., Durham, B. P., Key, R. S., Groussman, R. D., Bartolek, Z., Pinedo-Gonzalez, P., Hawco, N. J., John, S. G., Carlson, M. C. G., Lindell, D., Juranek, L. W., Ferrón, S., Ribalet, F., Armbrust, E. V., Ingalls, A. E., & Bundy, R. M. (2023). Siderophore production and utilization by marine bacteria in the North Pacific Ocean. *Limnology and Oceanography*, *68*(7), 1636–1653. https://doi.org/10.1002/lno.12373

Park, J., Heal, K. R., Ingalls, A. E., Groussman, R. D., Bartolek, Z., Armbrust, E. V., & Bundy, R. M. (2025). Efficient cobalamin uptake and cycling contribute to the lack of cobalamins in the surface cobalt-binding ligand pool in the North Pacific. *Limnology and Oceanography Letters*, *10*(4), 547–556. https://doi.org/10.1002/lol2.70019

Paul, C., Mausz, M. A., & Pohnert, G. (2013). A co-culturing/metabolomics approach to investigate chemically mediated interactions of planktonic organisms reveals influence of bacteria on diatom metabolism. *Metabolomics*, *9*(2), 349–359. https://doi.org/10.1007/s11306-012-0453-1

Paul, C., & Pohnert, G. (2011). Interactions of the Algicidal Bacterium Kordia algicida with Diatoms: Regulated Protease Excretion for Specific Algal Lysis. *PLoS ONE*, *6*(6), e21032. https://doi.org/10.1371/journal.pone.0021032

Plichta, D. R., Juncker, A. S., Bertalan, M., Rettedal, E., Gautier, L., Varela, E., Manichanh, C., Fouqueray, C., Levenez, F., Nielsen, T., Doré, J., Machado, A. M. D., De Evgrafov, M. C. R., Hansen, T., Jørgensen, T., Bork, P., Guarner, F., Pedersen, O., Metagenomics of the Human Intestinal Tract (MetaHIT) Consortium, … Nielsen, H. B. (2016). Transcriptional interactions suggest niche segregation among microorganisms in the human gut. *Nature Microbiology*, *1*(11), 16152. https://doi.org/10.1038/nmicrobiol.2016.152

Pinedo-González, P., Hawco, N. J., Bundy, R. M., Armbrust, E. V., Follows, M. J., Cael, B. B., White, A. E., Ferrón, S., Karl, D. M., & John, S. G. (2020). Anthropogenic Asian aerosols provide Fe to the North Pacific Ocean. *Proceedings of the National Academy of Sciences*, *117*(45), 27862–27868. https://doi.org/10.1073/pnas.2010315117

Pollara, S. B., Becker, J. W., Nunn, B. L., Boiteau, R., Repeta, D., Mudge, M. C., Downing, G., Chase, D., Harvey, E. L., & Whalen, K. E. (2021). Bacterial Quorum-Sensing Signal Arrests Phytoplankton Cell Division and Impacts Virus-Induced Mortality. *mSphere*, *6*(3), e00009-21. https://doi.org/10.1128/mSphere.00009-21

Polovina, J. J., Howell, E., Kobayashi, D. R., & Seki, M. P. (2001). The transition zone chlorophyll front, a dynamic global feature defining migration and forage habitat for marine resources. *Progress in Oceanography*, *49*(1–4), 469–483. https://doi.org/10.1016/S0079-6611(01)00036-2

Raina, J.-B., Lambert, B. S., Parks, D. H., Rinke, C., Siboni, N., Bramucci, A., Ostrowski, M., Signal, B., Lutz, A., Mendis, H., Rubino, F., Fernandez, V. I., Stocker, R., Hugenholtz, P., Tyson, G. W., & Seymour, J. R. (2022). Chemotaxis shapes the microscale organization of the ocean's microbiome. *Nature*, *605*(7908), 132–138. https://doi.org/10.1038/s41586-022-04614-3

Ramanan, R., Kim, B.-H., Cho, D.-H., Oh, H.-M., & Kim, H.-S. (2016). Algae–bacteria interactions: Evolution, ecology and emerging applications. *Biotechnology Advances*, *34*(1), 14–29. https://doi.org/10.1016/j.biotechadv.2015.12.003

Rice, P., Longden, I., & Bleasby, A. (2000). EMBOSS: The European Molecular Biology Open Software Suite. *The European Molecular Biology Open Software Suite*.

Rolland, J. L., Stien, D., Sanchez-Ferandin, S., & Lami, R. (2016). Quorum Sensing and Quorum Quenching in the Phycosphere of Phytoplankton: A Case of Chemical Interactions in Ecology. *Journal of Chemical Ecology*, *42*(12), 1201–1211. https://doi.org/10.1007/s10886-016-0791-y

Sarmento, H., & Gasol, J. M. (2012). Use of phytoplankton-derived dissolved organic carbon by different types of bacterioplankton. *Environmental Microbiology*, *14*(9), 2348–2360. https://doi.org/10.1111/j.1462-2920.2012.02787.x

Sarmento, H., Morana, C., & Gasol, J. M. (2016). Bacterioplankton niche partitioning in the use of phytoplankton-derived dissolved organic carbon: Quantity is more important than quality. *The ISME Journal*, *10*(11), 2582–2592. https://doi.org/10.1038/ismej.2016.66

Satinsky, B. M., Gifford, S. M., Crump, B. C., & Moran, M. A. (2013). Chapter Twelve—Use of Internal Standards for Quantitative Metatranscriptome and Metagenome Analysis. In E. F. DeLong (Ed.), *Methods in Enzymology* (Vol. 531, pp. 237–250). Academic Press. https://doi.org/10.1016/B978-0-12-407863-5.00012-5

Seth, E. C., & Taga, M. E. (2014). Nutrient cross-feeding in the microbial world. *Frontiers in Microbiology*, *5*. https://doi.org/10.3389/fmicb.2014.00350

Seyedsayamdost, M. R., Case, R. J., Kolter, R., & Clardy, J. (2011). The Jekyll-and-Hyde chemistry of Phaeobacter gallaeciensis. *Nature Chemistry*, *3*(4), 331–335. https://doi.org/10.1038/nchem.1002

Seymour, J. R., Amin, S. A., Raina, J.-B., & Stocker, R. (2017). Zooming in on the phycosphere: The ecological interface for phytoplankton–bacteria relationships. *Nature Microbiology*, *2*(7), 17065. https://doi.org/10.1038/nmicrobiol.2017.65

Shibl, A. A., Isaac, A., Ochsenkühn, M. A., Cárdenas, A., Fei, C., Behringer, G., Arnoux, M., Drou, N., Santos, M. P., Gunsalus, K. C., Voolstra, C. R., & Amin, S. A. (2020). Diatom modulation of select bacteria through use of two unique secondary metabolites. *Proceedings of the National Academy of Sciences*, *117*(44), 27445–27455. https://doi.org/10.1073/pnas.2012088117

Steinegger, M., & Söding, J. (2018). Clustering huge protein sequence sets in linear time. *Nature Communications*, *9*(1), 2542. https://doi.org/10.1038/s41467-018-04964-5

Sterling, A. R., Holland, L. Z., Bundy, R. M., Burns, S. M., Buck, K. N., Chappell, P. D., & Jenkins, B. D. (2023). Potential interactions between diatoms and bacteria are shaped by trace element gradients in the Southern Ocean. *Frontiers in Marine Science*, *9*, 876830. https://doi.org/10.3389/fmars.2022.876830

Stock, F., Bilcke, G., De Decker, S., Osuna-Cruz, C. M., Van Den Berge, K., Vancaester, E., De Veylder, L., Vandepoele, K., Mangelinckx, S., & Vyverman, W. (2020). Distinctive Growth and Transcriptional Changes of the Diatom Seminavis robusta in Response to Quorum Sensing Related Compounds. *Frontiers in Microbiology*, *11*, 1240. https://doi.org/10.3389/fmicb.2020.01240

Stock, W., Blommaert, L., De Troch, M., Mangelinckx, S., Willems, A., Vyverman, W., & Sabbe, K. (2019). Host specificity in diatom–bacteria interactions alleviates antagonistic effects. *FEMS Microbiology Ecology*, *95*(11), fiz171. https://doi.org/10.1093/femsec/fiz171

Stoecker, D. K., Hansen, P. J., Caron, D. A., & Mitra, A. (2017). Mixotrophy in the Marine Plankton. *Annual Review of Marine Science*, *9*(1), 311–335. https://doi.org/10.1146/annurev-marine-010816-060617

Teeling, H., Fuchs, B. M., Becher, D., Klockow, C., Gardebrecht, A., Bennke, C. M., Kassabgy, M., Huang, S., Mann, A. J., Waldmann, J., Weber, M., Klindworth, A., Otto, A., Lange, J., Bernhardt, J., Reinsch, C., Hecker, M., Peplies, J., Bockelmann, F. D., … Amann, R. (2012). Substrate-Controlled Succession of Marine Bacterioplankton Populations Induced by a Phytoplankton Bloom. *Science*, *336*(6081), 608–611. https://doi.org/10.1126/science.1218344

Tenenbaum, D. (2016). *KEGGREST: Client-side REST access to KEGG. R package version 1.12.3.*

Thingstad, F. T., & Lignell, R. (1997). Theoretical models for the control of bacterial growth rate, abundance, diversity and carbon demand. *Aquatic Microbial Ecology*, *13*, 19–27.

Van Tol, H. M., Amin, S. A., & Armbrust, E. V. (2017). Ubiquitous marine bacterium inhibits diatom cell division. *The ISME Journal*, *11*(1), 31–42. https://doi.org/10.1038/ismej.2016.112

Weissberg, O., Aharonovich, D., Wu, Z., Follows, M. J., & Sher, D. (2025). *Four Potential Mechanisms Underlying Phytoplankton-Bacteria Interactions Assessed Using Experiments and Models*. Microbiology. https://doi.org/10.1101/2025.01.03.631208

Wiener, D., Bartolek, Z., Dunklin, R., & Armbrust, E. V. (n.d.). *Growth phase-specific gene regulation and algicidal interactions between a new A. macleodii strain and the model diatom T. pseudonana*.

Yang, X., Cai, G., Cai, R., Gu, H., Chen, Y., Xie, J., Hu, Z., & Wang, H. (2023). *Redefinition of archetypal phytoplankton-associated bacteria taxa based on globally distributed dinoflagellates and diatoms*. Microbiology. https://doi.org/10.1101/2023.02.13.528248

Chapter 3: Genomes of new heterotrophic bacteria isolated from the Equatorial Pacific reveal potential metabolic exchanges influencing multi-trophic interactions

Zinka Bartolek[1] and Frank X. Ferrer González[1], Veronica Tuerpe[2], Jun Sun[3], Anitra E. Ingalls[1], E. V. Armbrust[1]

[1]School of Oceanography, University of Washington, Seattle, WA, 98195, USA
[2] Biology Department, Seattle University, Seattle, WA, 98122, USA
[3] Department of Biology, University of Washington, Seattle, WA, 98195, USA

Abstract

Heterotrophic bacteria are central to ocean carbon cycling, consuming most of the organic matter produced by phytoplankton, yet their metabolic capabilities in the Equatorial Pacific are still not fully characterized. Here, we present novel isolates and closed genomes of 52 heterotrophic bacteria capable of utilizing phytoplankton-derived organic matter from multiple depths and locations in the Equatorial Pacific. Our isolates range from novel strains and species to a potentially new genus belonging to the Hyphomicrobiales order. Isolate genomes reveal diverse biosynthetic pathways, auxotrophies, and the capacity to produce and modify ecologically relevant metabolites, including vitamins and phytohormones. Genome-based predictions of potential metabolic exchanges between isolates suggest cross-feeding of vitamins, growth factors and signaling molecules. These metabolic exchanges within bacterial communities could influence interactions with phytoplankton, and suggest that higher order interactions may influence the function and structure of microbial communities. The availability of isolates from the Equatorial Pacific provides future opportunities to test metabolic capabilities and potential exchanges in lab studies.

Introduction

Marine heterotrophic bacteria are essential components of microbial networks; they play a central role in global biogeochemical cycles, consuming approximately 85% of the carbon fixed by phytoplankton within days (Moran et al., 2022; Hansell, 2013). The Equatorial Pacific is one of the most biogeochemically dynamic ocean regions, with strong upwelling, high nutrient fluxes and some of the highest primary productivity on Earth (Dai et al., 2023; Pennington et al., 2006). Potential metabolisms of heterotrophic bacteria in the Equatorial Pacific are largely based on metagenome assembled genomes (MAGs) or single-cell amplified genomes (SAGs) obtained directly from environmental samples (Sun & Ward, 2021). Although MAGs and SAGs have greatly advanced our understanding of metabolic diversity and genetic potential of marine bacteria, they are typically incomplete, potentially lacking key metabolic features or genes. Sequencing genomes of cultured isolates results in high quality closed genomes (Wick et al., 2023), and allows for characterization of the full metabolic potential as well as metabolic deficiencies and auxotrophies in each strain. Additionally, closed genomes of isolates provide an opportunity to study a full suite of metabolic interdependencies within bacterial communities both through co-culture lab experiments and modeling approaches.

The functional diversity of marine heterotrophic bacteria is vast, and can be categorized into different functional guilds ranging from oligotrophs that dominate nutrient poor low latitude surface waters to slow growing copiotrophs that grow on recalcitrant substrate at depth (Zakem et al., 2024). A large fraction of marine heterotrophs is attached to particles and phytoplankton, where nutrient levels can exceed surrounding levels by orders of magnitude (Seymour et al., 2017). These environments attract copiotrophic bacterial taxa specialized in degrading complex organic matter. In turn, phytoplankton-derived metabolites promote resource partitioning among heterotrophic bacteria. Bacterial representatives across diverse taxonomic groups appear to use relatively few phytoplankton-derived compounds such as guanosine, proline and *N*-acetyl-d-glucosamine (Ferrer-González et al., 2021). Similarly, uptake of phytoplankton derived metabolites was taxon specific based on measured incorporation of isotope labeled substrates in bacterial cultures (Mayali et al., 2012), and measurements of transcript and metabolite abundances in the environment point towards specific heterotrophic users of liable molecules like DMSP, DHSP, glycine betaine, trehalose and homarine (Durham et al., 2019; Boysen et al., 2021; Boysen et al., 2022). In lab culture experiments, utilization of dissolved organic matter (DOM) by heterotrophs could be predicted for each individual strain based on their metabolic potential, however, in a community of bacteria, DOM utilization became unpredictable, indicating interactions between bacteria can shape the structure of DOM in the ocean (Fuessel et al., 2025).

In addition to their ability to degrade various organic compounds, heterotrophic bacteria can produce and use metabolites that affect the physiology of their hosts as well as mediate interactions within bacterial communities. Different signaling molecules and hormones can influence growth rates of phytoplankton hosts (Amin et al., 2015; Durham et al., 2017; Fei et al., 2020; Shibl et al., 2020). Biosynthetic pathways for synthesis of growth-promoting hormones including indole-3-acetic acid (IAA), gibberellins and brassinolides are present in phytoplankton growth promoting heterotrophic bacteria (Khalil et al., 2024). Trait based comparisons of bacteria genomes revealed unique traits related to interactions with phytoplankton, especially in Alpha- and Gammaproteobacteria, including production of phytohormones, siderophores and different B vitamins (Zoccarato et al., 2022). These genetic traits have also been associated with

cross-feeding within bacterial communities. Heterotrophic bacterial cross-feeding of particular amino acids and B vitamins was predicted through community level metabolic modeling (Giordano et al., 2024). Interestingly, cross feeding of B vitamins such as cobalamin (B12) and thiamine (B1) was most prevalent between strains with partial synthesis pathways, instead of between complete synthesizers and auxotrophs (Arandia-Gorostidi et al., n.d.). However, the effect of cross-feeding and metabolite exchanges within heterotrophic bacterial communities on the activity of metabolites associated with regulating host physiology is still understudied.

Here we report the isolation and metabolic capabilities of 52 novel heterotrophic bacteria for which we obtained closed genomes. These isolates grow on phytoplankton-derived organic matter collected from different locations and depths in the Equatorial Pacific. We examined their metabolic potential and proposed potential metabolic exchanges that could influence interactions both among bacteria and between bacteria and their phytoplankton hosts.

Methods

Bacterial isolation, cultivation and identification

Bacterial isolates were collected on the Gradients 5 cruise (TN412) in the Equatorial Pacific from January 22 to February 18 2023. Seatwater was collected from the CTD rosette or the ship's underway system, with 0.5 L to 2 L of seawater subsampled into acid-washed polycarbonate carboys from depths ranging from 5 m to 1000 m. Seawater was prefiltered with a 100 μm mesh, and concentrated into 1 mL samples by filtering through 0.2 μm polycarbonate filters. Concentrated seawater was plated onto 90 mm petri dish plates in 100 μL aliquots. Plates contained different types of rich media, including ½ ytss (2g of yeast extract, 1.25 g of tryptone, 35 ppt Hawaii sea water, 15 g agar in 1L of distilled water), and 8 different media types with individual carbon or nitrogen sources (1 L of distilled water, 15 g agar, and 2 g of one of the following carbon/nitrogen sources for each media type: succinate, glucose, CAS amino acids, pyruvate, acetate, sarcosine, sucrose, and starch). Plates were incubated at 25°C for 1-4 days until bacterial growth was observed, and then stored at 4°C until further processing. Single colonies from plates composed of the different media types were re-plated onto solid ½ ytss plates, and incubated at 25°C. The process was repeated until single-strain colonies were observed on plates based on visual inspection of color, shape and size. Single colonies were grown in liquid ½ ytss media overnight at 25°C with shaking and preserved in glycerol stocks (1 ml of culture with 10% v/v glycerol) by flash freezing in liquid nitrogen and storing at -80°C. To determine the identity and purity of the isolates individual strains were revived from glycerol stocks by growing on solid ½ ytss plates for 2-7 days. DNA was extracted and the full length 16S rDNA was PCR-amplified using universal primers (27F and 1492R) and Sanger sequenced through Azenta Life Sciences. Strains were taxonomically identified by aligning sequences with BLAST NCBI (Camacho et al., 2009) and taxonomy assigned based on the highest sequence similarity matches to reference sequences.

16S rRNA gene phylogenetic tree of bacterial isolates

A phylogenetic tree based on 16S rDNA sequences was constructed to determine the phylogenetic diversity of isolated strains. The tree included as references the 16S rDNA sequences from 50 marine bacteria and 10 non-marine bacteria from diverse environments including those commonly found on human skin (Sup. Table 3.2) Sequences were trimmed using the R package sangerAnalyseR (Chao et al., 2021) with quality cutoff score of 29 and sliding

window size of 15 (M2CutoffQualityScore=29, M2SlidingWindowSize=15). Trimmed sequences were aligned with MAFT v7.526 (Katoh & Standley, 2013), and aligned sequences were trimmed to the same length (400 bp) in Jalview (Waterhouse et al., 2009). The maximum likelihood phylogenetic tree was built with RAxML (Stamatakis, 2014) with 100 bootstraps (raxmlHPC-PTHREADS-SSE3 -f a -m GTRCATI -p 59867 -B.03 -x 14325 -# 100 -d -k). The phylogenetic tree was visualized in iTOL (Letunic & Bork, 2021).

Genome sequencing, assembly and annotation

Strains selected for whole-genomes sequencing were revived from glycerol stocks by growing at 25°C on solid ½ ytss plates for 2-4 days, then inoculating single colonies into 5 mL liquid ½ ytss plates and incubating overnight at 25°C with shaking. The 52 isolates were sequenced in three different batches (Sup. Table 3.1). High molecular weight DNA was extracted for 17 isolates using the MasterPure Complete DNA and RNA Purification Kit (Biosearch technologies) with an added RNAase step. DNA was cleaned using 1X volume:volume DNA magnetic beads (Sergi Lab Supplies, Seattle WA) with two 80% ethanol washes, then eluted in 20 µL of molecular grade water. DNA was quantified using a Qubit dsDNA Quantitation High Sensitivity kit (Invitrogen, Waltham, MA). Library preparation and sequencing was done in two batches (batch 1 and 2) at the University of Washington Nanopore Sequencing Core (University of Washington, Seattle, WA) using the Nanopore PromethION flow cell (Oxford Nanopore Technologies, Oxford, United Kingdom). High molecular weight DNA was extracted from the remaining 32 isolates using PacBio's Nanobind PanDNA Kit, library preparation was done using Nanopore Library Prep, and sequenced in batch 3 using Nanopore sequencing on the PromethION flowcell (Oxford Nanopore Technologies, Oxford, United Kingdom).

Long read sequences were processed by trimming raw nanopore reads for quality using BBDuk (BBTools suite) (Bushnell, 2014) with parameters qtrim=rl and trimq=20 to remove low-quality bases from both ends. Trimmed reads were then normalized to reduce coverage bias using BBNorm (Bushnell, 2014), retaining a minimum depth of 5X and capping maximum depth at 200X (using parameters min=5 and target=200). Normalized reads were assembled de novo with Flye v2.9.4 (Kolmogorov et al., 2019) with the --nano-raw or –nano-hq settings and an estimated genome size of 5 Mb (--genome-size 5000000). All assemblies were performed within an Apptainer container using 14 threads per run. Assemblies were polished using three iterative rounds of Racon (Vaser et al., 2017), with alignments generated by minimap2 (Li, 2018) (-x map-ont). Final polished assemblies were saved in FASTA format for downstream analyses. Gene prediction and functional annotation were conducted using Prokka v1.14 (Seemann, 2014) with default parameters. Quality of assemblies including GC content was determined using QUAST (Mikheenko et al., 2023), and CheckM v1 (Parks et al., 2015) was used to determine completeness and contamination based on the presence and absence of marker genes.

Phylogenomics and analysis of bacterial metabolism

For each isolate genome, the closest known relative genome was found using BV-BRC Similar Genome Finder (Olson et al., 2023), using both the representative and reference genome option and all public genomes options containing published MAGs and SAGs. Average Nucleotide Identity (ANI) between each isolate and its closest known representative genome relative was calculated with fastANI (Jain et al., 2018) through Kbase (Arkin et al., 2018). Phylogenomic relationships among all isolated strains were assessed by running genome clustering based on pairwise distances and ANI between all selected genomes using the

Neighbor-Joining algorithm in Anvio v8 (Eren et al., 2015). Reconstruction of metabolic pathways and estimation of pathway completeness was done with Anvio v8 through the anvi-estimate-metabolism program using modules from the KEGG MODULES database (Veseli et al., 2025). KEGGDecoder (Graham et al., 2018) was also used to estimate metabolic pathway completeness, with input data based on putative genes annotated with KEGG Orthologs (KOs) though Anvio v8 anvi-run-kegg-kofams program. In addition, secondary metabolite related biosynthetic gene clusters (BGCs) were predicted using antiSMASH (Blin et al., 2017).

Prediction of metabolic exchanges

Metabolic exchanges between bacterial isolates were predicted using Anvio-dev -predict-metabolic-exchanges module (https://github.com/merenlab/anvio/blob/metabolic_interactions/anvio/docs/programs/anvi-predict-metabolic-exchanges.md). Briefly, Anvio contigs-db databases were constructed for each genome and annotated with modelSEED (Seaver et al., 2021) and KEGG (Kanehisa et al., 2025) reference databases and a reaction network was created for each genome. Anvio -predict-metabolic-exchanges was run with the following flags: --use-equivalent-amino-acids ; --maximum-gaps 0; --report-compounds-with-no-prediction. Both the KEGG Pathway Map Walk and the Reaction Network Subset approaches were used to predict which compounds can be produced by one organism and consumed by another.

Results and Discussion

Isolation and sequencing of marine copiotrophic bacteria from the Equatorial Pacific

We sequenced complete genomes for 52 novel bacterial isolates from the Equatorial Pacific. These were selected from a collection of ~400 isolates, out of which 96 were distinct after clustering at 97% 16S rRNA sequence identity (Sup. 3.1). Bacterial isolates originated from a transect along 140°W spanning latitudes 23°N to 4°S and depths from the surface (5m depth) to deep ocean (1000m depth) (Fig. 3.1). To enrich for copiotrophic bacteria capable of growing on phytoplankton-derived metabolites, we used a range of isolation media containing various compounds released by phytoplankton (Methods). The resulting isolates represent 10 bacterial orders, including those generally associated with phytoplankton and particles where nutrient concentrations are higher than surrounding seawater (Fig. 3.1) (Amin et al., 2012; Seymour et al., 2017).

Our use of long-read Oxford Nanopore technology resulted in high quality closed genome assemblies for the 52 chosen isolates. 36 genomes contained 1-3 circular contigs, although some Rhodobacterales had as many as 21 circular contigs of varying sizes (Table 3.1). The elevated number of contigs in Rhodobacterales was restricted to two genera, *Sulfitobacter* and *Loktonella*, which are known to contain numerous extrachromosomal replicons (ECRs) including plasmids and chromids (Petersen & Wagner-Döbler, 2017; Freese et al., 2022). These elements often encode mobilizable functions such as biofilm formation, transporters, secretion systems and antibiotic resistance and contribute to horizontal gene transfer and genome plasticity (Michael et al., 2016). Genome sizes ranged from 2.6 to 6.0 Mbp with CG content varying across orders: on average ~35% GC in Flavobacteriales and Bacillales, ~45% GC in Alteromonadales, ~55% GC in Hyphomicrobiales, Oceanospirillales, and Pseudomonadales, and 60–65% GC in Rhodobacterales and Sphingomonadales (Table 3.1). The genomes were assumed to be >99.94%

complete based on analysis with CheckM (Parks et al., 2015) with low contamination (47 genomes <1% contaminated, 5 genomes <3.5% contaminated; Table 3.1), confirming their suitability for downstream metabolic analyses.

Genomes from the Equatorial Pacific isolates represent novel lineages at the strain, species, and potentially genus level compared to their closest relatives in public databases. Thirty two genomes were classified as novel strains, sharing 95–98% ANI with reference genomes and spanning multiple bacterial orders (Table 3.1). Nineteen genomes represented putative new species (ANI values between 85% and 94%) and included members of the Rhodobacterales (closest relative: *Sulfitobacter dubious*), Alteromonadales (*Idiomarina donghaiensis*, *Glaciecola mesophila*, and *Alteromonas marina*), Oceanospirillales (*Halomonas glaciei*), and Sphingomonadales (*Erythrobacter pelagi*) (Table 3.1). One genome, assigned to the Hyphomicrobiales and most closely related to *Hoeflea halophila*, showed 81% ANI to its nearest reference, indicating it may represent a novel genus (Table 3.1) (Jain et al., 2018; Rodriguez-R et al., 2022). Overall strain diversity was high, with most isolates sharing ANI < 0.8 with each other, except for a few clusters within the Sphingomonales, Oceanospirales and Rhodobacerales (Sup. 3.2).

Metabolic capabilities of Equatorial Pacific isolates

To determine the metabolic capabilities of bacterial isolates based on their genomes, we assessed the completeness of synthesis and degradation pathways for various compounds. We computed KEGG module completeness (Anvio v8) (Sup. 3.3) and the completeness of metabolic pathways (KEGGDecoder) (Fig. 3.2). Functional profiles clustered by taxonomy (Fig 3.2), consistent with the expectation that closely related taxonomic groups share more similar functional capabilities (Konstantinidis & Tiedje, 2005). Most isolates encoded complete pathways for amino acid biosynthesis, including tryptophan, alanine, proline, glycine, glutamine, threonine, lysine and serine, as well as partially complete pathways for valine and isoleucine (Fig 3.2). Distinct auxotrophies were observed in certain lineages: all Flavobacterales lacked complete pathways for ornithine and arginine (Sup. 3.3), while Sphingomonadales were unable to synthesize aspartate and glutamate (Fig 3.2). Auxotrophic Flavobacteria are known to rely on phytoplankton hosts for essential metabolites such as amino acids and cofactors suggesting possible co-evolutionary relationships with eukaryotic hosts or cross-feeding with other bacteria (Van Tol et al., 2017; Bartolek et al., 2022). Pathways for synthesis of several cofactors and vitamins were present across all isolates, including thiamine (B1), riboflavin (B2) and co-enzyme A (Fig 3.2; Sup. 3.3). Phytoplankton depend on bacterial supply for vitamins, with up to 23% of tested phytoplankton species auxotrophic for B1 (Croft et al., 2005). Core metabolic pathways, including the citrate cycle, glycolysis, pentose phosphate pathway and fatty acid biosynthesis were present in all isolates (Sup. 3.3). Finally, type I secretion systems involved in export of enzymes, toxins, and siderophores were partially complete across isolates, consistent with previous reports that marine bacteria retain only subtypes necessary for exporting specific metabolites (Fig. 3.2) (Green & Mecsas, 2016).

Several bacterial orders had distinct metabolic capabilities. Only Rhodobacterales displayed the greatest diversity of distinct features. All members of this order encoded the C-P lyase *phn* operon that enables utilization of phosphonates as a source of dissolved organic phosphorus under phosphate limitation (Sosa et al., 2019), and sulfur-based pathways such as DMSP demethylation, thiosulfate oxidation, and sulfite dehydrogenase (Fig. 3.2). Three

Rhodobacterales strains also encoded a complete phthalate degradation pathway, allowing use of this plasticiser as a carbon source (Iwaki et al., 2012). Rhodobacterales further possessed partial pathways for Type II secretion systems, which facilitate extracellular degradation of complex organic matter (Green & Mecsas, 2016). The Pseudomonadales was the only order that encodes enzymes for denitrification through pathways including nitrate reduction and nitric- and nitrous-oxide reduction (Fig. 3.2). One isolate (i17) likely representing a novel genus most closely related to *Hoeflea halophila,* contained RubisCO and a complete Calvin-Benson-Bassham cycle (CBB), suggesting the capacity for anaplerotic CO_2 fixation (Laguna et al., 2010).

Isolates encoded biosynthesis pathways for several keystone metabolites (Durham et al., 2025) and metabolites mediating interspecies interactions. Two keystone vitamins showed distinct patterns: the biotin (B7) pathway was present in all isolates except the Rhodobacterales isolates (Fig. 3.2; Sup. 3.3), whereas the aerobic cobalamin (B12) pathway was complete only in Rhodobacterales (Fig. 3.2; Sup. 3.3). The B12 transporter was present only in Alteromonadales, Vibrionales and Enterobacterales orders (Fig. 3.2). Antibiotic resistance genes were found in Sphingomondales for β-lactams, while partial vancomycin resistance pathways were present across many isolates. Similarly, Type VI and Type II secretion systems associated with competition and virulence occurred only in Vibro, Enterobacter and some Pseudomonales (Sup. 3.3). Osmoprotectants biosynthetic pathways, including ectoine, were encoded in Oceanospiralles, Rhodobacterales and Pseudomonales (Sup. 3.3). Flavobacterales carried multiple phytoplankton-associated metabolic traits, such as metabolism for phytoplankton exometabolites proline and tryptophan, and siderophore (staphylopherine) biosynthesis, while the chitin degradation pathway was uniquely detected in the *Vibrio* isolate (Fig. 3.2; Sup. 3.3). Partial pathway for Jasmonic acid synthesis which is known to play a role in regulation of phytoplankton growth was observed in a few Flavobacterales (Sup. 3.3) (Wasternack & Strnad, 2018). Collectively these findings show that beyond core metabolism and lineage specific traits, Equatorial Pacific isolates exhibit specialized capacities to produce and metabolize keystone metabolites and compounds central to interspecies interactions.

Metabolic exchanges between Equatorial Pacific isolates

We tested for potential metabolic dependencies and exchanges based on genomic capabilities between bacterial isolates to assess the potential for syntropy or cross-feeding. Since our isolates are copiotrophs commonly associated with phytoplankton and particles, they are likely to come into contact with each other in nutrient rich microenvironments, and in addition to competing with each other, they may have capabilities of exchanging resources and metabolites. We chose one representative from each order with the lowest ANI to a previously sequenced genome, as they are most likely to uncover understudied types of metabolic exchanges (Table 3.1; Sup. 3.4A), and tested whether these six bacteria possess the potential capability of exchanging metabolites with each other. Possible production and consumption pathways for a range of compounds in each bacteria were assessed using Anvio's Predict Metabolic Exchanges function. The same approach was used to assess whether a bacteria had the capability to produce a particular compound or intermediate, and at the same time if another bacteria had the capability to consume or degrade it.

We assessed the genomic potential of the six target isolates to produce-only, consume-only, or both produce and consume metabolites (Sup. 3.4A). We hypothesized that metabolites both produced and consumed by an isolate likely support internal metabolism, those consumed

but not produced indicate potential dependencies and auxotrophies, and those produced but not consumed could mediate interactions with other organisms. Compounds that were primarily both produced and consumed include nucleic acids, N-glycosides, and vitamins and co-factors (Sup. 3.4B). Fatty acyls dominated each of the three categories of compounds and were abundant in the examined bacteria and serve as building blocks for membranes, energy storage, and precursors for various signaling molecules (De Carvalho & Caramujo, 2018). Compounds that were primarily either only produced or only consumed but not both included flavonoids, o-glycosides, phytotoxins and polyketides (Sup. 3.4B). In comparison to compounds that are both produced and consumed, the produced-only or consumed-only compounds had a higher specificity to a particular isolate (Sup. 3.4C), consistent with the high taxonomic and functional diversity of tested bacterial genomes.

Potentially exchanged compounds within the six representative isolates were assessed by looking at the presence of KEGG biosynthesis and degradation pathways and modelSEED reactions between isolate genomes. A potential exchange is defined as the ability of one bacteria to produce a particular compound or intermediate, and another bacteria to modify or degrade it. Classes of compounds that were exchanged across the all 6 tested bacteria were dominated by fatty acids, carbohydrates, lipids, peptides, and terpenoids, however, exchanges of vitamins and cofactors, terpenoids, flavonoids, and hormones were also observed (Sup. 3.5A; Sup 3.5B). We observed exchanges of environmentally relevant metabolites, including those classified as abundant metabolites and keystone metabolites (Sup 3.6) (Durham et al., 2025). Abundant marine metabolites that were exchanged between the isolates included amino acids such as alanine, aspartate, glutamine, glutamate, and serine among others (Sup 3.6). Keystone metabolites that were exchanged contained B vitamins such as niacin (B3), as well as vitamers, which are precursors, moieties, or degradation products of B vitamins. These products can satisfy vitamin requirements as well as mediate interspecies interactions (Durham et al., 2025). Dethiobiotin was exchanged and can be used as an alternative to B7 auxotrophy (Wienhausen et al., 2022), while 4-Amino-5-aminomethyl-2-methylpyrimidine is a B1 precursor that can be used in some organisms instead of B1 (Sup. 3.6) (Paerl et al., 2023). Similarly, coblalamin (B12) remodeling can be done using Alpha-Ribazole and Adenosylcobalamin-5'-phosphate, both of which were potentially exchanged between Equatorial Pacific isolates (Sup. 3.6) (Bannon et al., 2024). Several polyunsaturated fatty acids such as EPA, linolenate, and arachidic acid were potentially exchanged between bacterial isolates; higher tropic levels such as invertebrates are auxotrophic for them and depend on their availability from microalgae and bacterial sources (Nichols, 2003). Additionally, several nitrogen and sulfur osmolytes were potentially exchanged including ectoine, betaine aldehyde and carnitine (Sup. 3.6). A few bacterial toxins were exchanged, including protoanemonin, which is known to disrupt quorum sensing in bacterial communities (Sup. 3.6) (Bobadilla Fazzini et al., 2013). Overall, these potential exchanges highlight core metabolic interdependencies as well as specialized exchange of keystone metabolites that structure marine ecosystems.

Bacterial isolates were able to exchange compounds with the potential to mediate cross-kingdom interactions, including phytohormones that regulate phytoplankton growth and metabolism. These included indole-3-acetic acid (IAA) derivatives, gibberellins, abscisic acid and jasmonic acid (Sup. 3.6), some of which enhance growth of diatoms in cultures (Khalil et al., 2024). Although none of the tested bacterial strains contained a complete pathway for synthesis of IAA from tryptophan, representatives of the Hypomicrobiales, Alteromonasles and Flavobacteriales contained genes for select steps of different synthesis pathways (Fig. 3.4A)

(Labeeuw et al., 2016). Some isolates like the i17 Hyphomicrobiales and i154 Oceanospirillales possessed the ability to take up exogenous IAA and convert it to biologically inactive storage forms of IAA such as IAA-amino acid conjugates (IAA-Leu, IAA-Glu, IAA-Asp, IAA-Ala) via IAA-amodesynthetases, while others including SET2-66 Alteromonadales and SET3-63 Flavobacteriales had the ability to convert it back to its active state (Fig. 3.4; Sup. 3.7) (Hayashi et al., 2021). Exchanges of phytohormones within bacterial communities suggest that interactions between bacteria might shape interactions spanning multiple trophic levels. Additionally, modifications of compounds within bacterial communities that are known to play a role in interactions between bacteria and their eukaryotic hosts indicate a potential prevalence of higher order interactions in marine microbial systems (Bairey et al., 2016).

Conclusion

Isolation and sequencing of novel bacterial strains from the understudied Equatorial Pacific expand our understanding of the metabolic potential of copiotrophic heterotrophic bacteria. Closed genomes enable identification of biosynthetic pathways and auxotrophies for ecologically significant metabolites. While the activity of individual bacteria can be inferred from their metabolic potential, potential exchanges of metabolites, including vitamins and growth factors, can enhance and diversify the metabolic capacity of microbial communities (Fuessel et al., 2025). We tested for possible metabolic exchanges between isolated bacteria and found evidence for the structural modification of phytohormones that influence phytoplankton growth, as well as other keystone metabolites (Durham et al., 2025). This indicates interactions within microbial communities could be an important factor in controlling phototroph-heterotroph associations through higher order interactions. The metabolic exchanges predicted here are putative and based on metabolic potential through KEGG pathways and reactions. Previous studies have found that functional guilds of metabolic interactions cannot be predicted by phylogeny or metabolic capacity alone, highlighting the need for direct measurements of metabolic activity to detect emergent properties of complex bacterial communities (Mayali et al., 2023). The availability of these isolates now provides a platform for testing hypothesized metabolic exchanges under laboratory and environmental conditions, offering a path to uncovering novel emergent properties of microbial communities in the ocean.

Figures and Tables
Figure 3.1

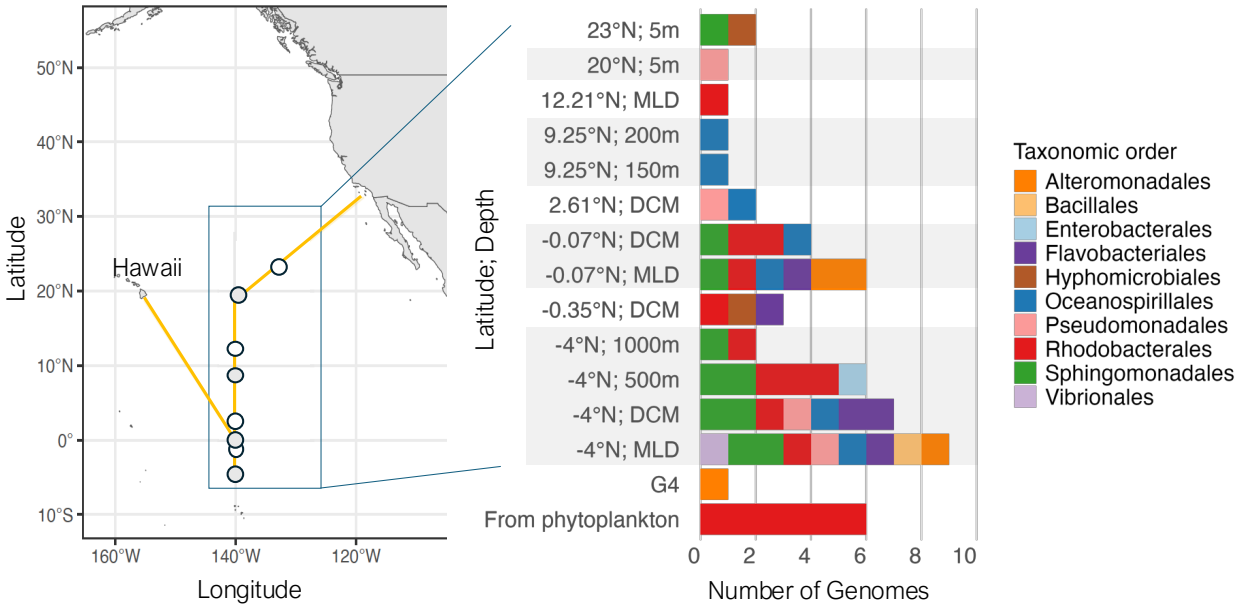


Fig. 3.1. Isolation sites and taxonomic characteristic of sequenced isolates. (left) The cruise track in the Equatorial Pacific is shown in the yellow line, with dots showing sites where the bacteria were isolated from. (right) Taxonomic distribution of isolated strains by site and depth. G4 indicates isolates collected on a different cruise to the same area 2021 (TNT 397). From phytoplankton indicates bacteria that were isolated from phytoplankton cultures collected on the same cruise (Graff van Creveld et al., 2024).

Table 3.1

Genome	# Contigs	GC %	Completeness	Contamination	Size (bp)	# CDS	Closest Species	ANI
Alteromonadales								
SET2-66	1	48.4	99.99	0.02	2606062	2490	Idiomarina donghaiensis	91.06
SET2-64	2	44.55	100	0.11	5168764	4367	Glaciecola mesophila	91.1697
S-E1-4	2	44.81	100	0.2	4853993	4197	Alteromonas marina	92.59
SET2-22	1	44.84	100	0.04	4473462	3956	Alteromonas mediterranea	97.62
Bacillales								
SET1_73	1	36.45	100	0	3803512	3765	Psychrobacillus psychrotolerans	98.237
Enterobacterales								
SET1_71	3	55.23	100	0.04	4789250	4383	Pantoea agglomerans	98.4695
Flavobacteriales								
SET3-63	1	38.11	99.94	0.04	3293458	2883	Dokdonia donghaensis	95.7728
SET3_70	1	36.02	100	0.03	4768663	4057	Maribacter sp	97.3829
SET3_49	1	36.02	100	0.03	4768663	4057	Maribacter sp	97.4446
SET2_72	1	39.82	99.97	0.05	4334862	3715	Leeuwenhoekiella blandensis	97.8348
SET2-31	1	33.95	99.98	0.03	2933880	2674	Croceibacter atlanticus	98.1744
Hyphomicrobiales								
i17	1	58.57	100	0	4764147	4546	Hoeflea halophila	81.9645
SET3_46	3	66.8	100	0.11	4572452	4312	Aurantimonas coralicida	97.1364
Oceanospirillales								
i154	2	55.04	100	0.19	4609007	4379	Halomonas glaciei	86.1623
SET3_2	3	56.82	100	0.03	3937209	3706	Halomonas aquamarina	97.662
SET3_29	1	56.96	99.99	0.05	3462635	3187	Halomonas aquamarina	97.6802
SET1_30	1	56.96	99.99	0.05	3462787	3184	Halomonas aquamarina	97.6996
SET3_7	1	56.96	99.99	0.05	3462629	3190	Halomonas aquamarina	97.701
i78	2	56.98	99.99	0.84	3872877	3637	Halomonas axialensis	98.0926
i162	1	56.74	99.99	0.62	3670685	3437	Halomonas meridiana	98.0948
Pseudomonadales								
SET3_38	2	57.98	100	0	4114437	3792	Marinobacter vinifirmus	96.6779
SET1_67	1	58.44	100	0	3784967	3469	Marinobacter vinifirmus	96.8182
i12	1	58.33	99.99	0.01	3798881	3487	Marinobacter vinifirmus	96.7468
SET1_61	1	57.14	99.98	0.22	4491994	4100	Marinobacter sp	98.4811
Rhodobacterales								
SET2-15	8	67.28	99.98	0.7	4261539	4064	Loktonella cinnabarina	87.8369
SET2_2	8	67.28	99.98	0.71	4261547	4060	Loktanella cinnabarina	87.9538
SET3-66	10	61.82	100	3.1	5014008	4798	Sulfitobacter pseudonitzcchiae	88.68
i34	11	60.76	100	0.99	4374236	4273	Sulfitobacter dubius	89.4672
i182	11	60.76	100	0.95	4374332	4283	Sulfitubacter dubius	89.4686
SET2_79	9	60.85	100	0.22	4259744	4155	Sulfitobacter dubius	89.5244
i1000	11	60.83	100	1.01	4399799	4292	Sulfitobacter dubius	89.5856
i189	21	60.73	100	1.1	4476625	4421	Sulfitobacter dubius	89.5914
SET3_71	11	61.84	100	3.01	5046623	4827	Sulfitobacter pseudonitzschiae	89.9929
SET3_28	10	61.82	100	3.19	5013933	4799	Sulfitobacter pseudonitzschiae	90.019
i179a	9	60.16	100	0.45	4265419	4124	Sulfitobacter dubius	90.0662
plank1	2	59.84	100	0.27	4584759	4331	Thalassobius mediterraneus	94.6493
Prex1	2	59.84	100	0.26	4584788	4331	Thalassobius mediterraneus	94.673
SET3-68	5	60.33	100	0.26	3776382	3631	Sulfitobacter pontiacus	97.09
SET2-6	6	60.35	100	0.05	3607620	3480	Sulfitobacter pontiacus	97.31
SET3_47	5	60.45	100	0.36	3686347	3551	Sulfitobacter pontiacus	97.4136
SET3_59	8	63.68	100	0.68	4807456	4482	Leisingera caerulea	98.0207
Sphingomonadales								
SET1_28	3	63.51	100	0.35	3056370	2920	Erythrobacter pelagi	87.4184
i187	2	63.51	100	0.36	3054197	2915	Erythrobacter pelagi	87.4819
SET3-20	1	64.81	100	0.18	3414984	3294	Citomicrobium bathyomarinum	95.83
SET1-46	2	60.86	100	2.41	3815572	3675	Qipengyuania sp.	97.0388
i62a	1	60.73	100	0.73	3555479	3389	Qipengyuania sp	97.0939
i183	2	64.13	100	0.23	3244173	3168	Erythrobacter citreus	97.8784
SET2_20	1	64.24	100	0.17	2992365	2906	Erythrobacter citreus	97.9005
SET3_52	1	64.56	100	0.02	2918400	2814	Erythrobacter citreus	97.9657
SET3-10	1	64.56	100	0.02	2918398	2812	Erythrobacter citreus	97.9667
i181	1	64.24	100	0.17	2992384	2907	Erythrobacter citreus	98.0012
Vibrionales								
SET1_57	4	44.46	100	3.55	6003807	5619	Vibrio campbellii	96.3258

Table 3.1 Characteristics of Equatorial Pacific bacterial genomes. Bacterial genomes are grouped by order and organized from lowest to highest ANI to closet relative. Total number of circular contigs is shown, along with GC content, completeness and contamination, total size of each genomes and number of coding DNA sequences (CDS). The closest species relative with a genome available in public databases is stated along with the average nucleotide identity (ANI) to the closest relative.

Figure 3.2

Fig. 3.2. Metabolic capabilities of Equatorial Pacific bacteria isolates. Metabolic capabilities of isolates were assessed using KEGGDecoder. Heatmap shows pathway completeness for each KEGG metabolic pathway (y-axis) for every genome (x-axis). The heatmap is clustered by pathway completeness. The color bar on the bottom shows the order level taxonomy of each bacteria isolate.

Figure 3.3

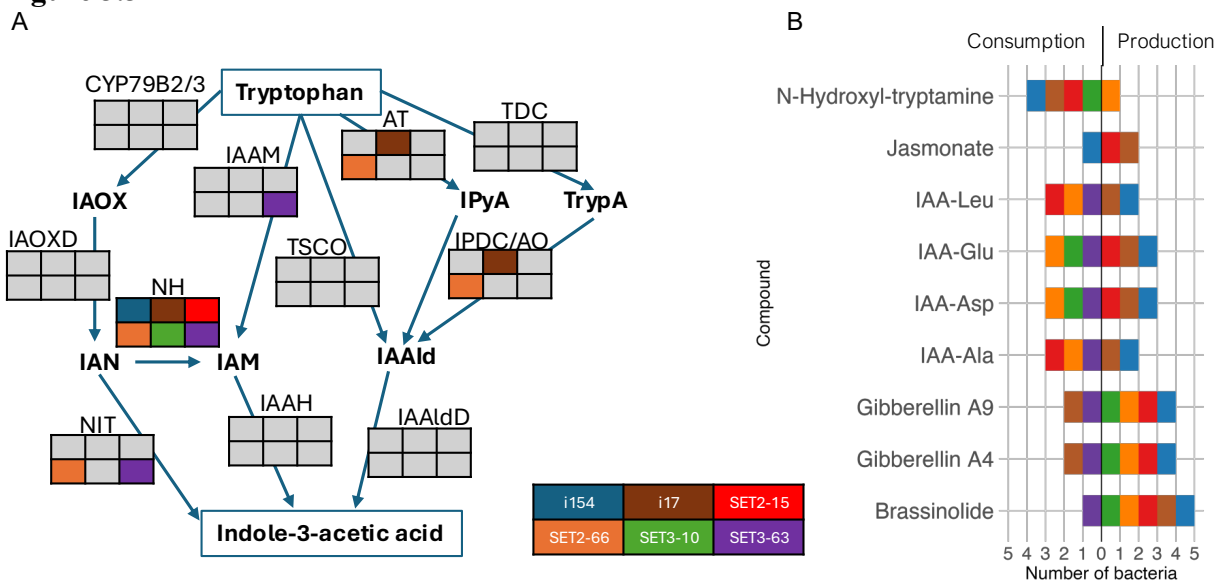
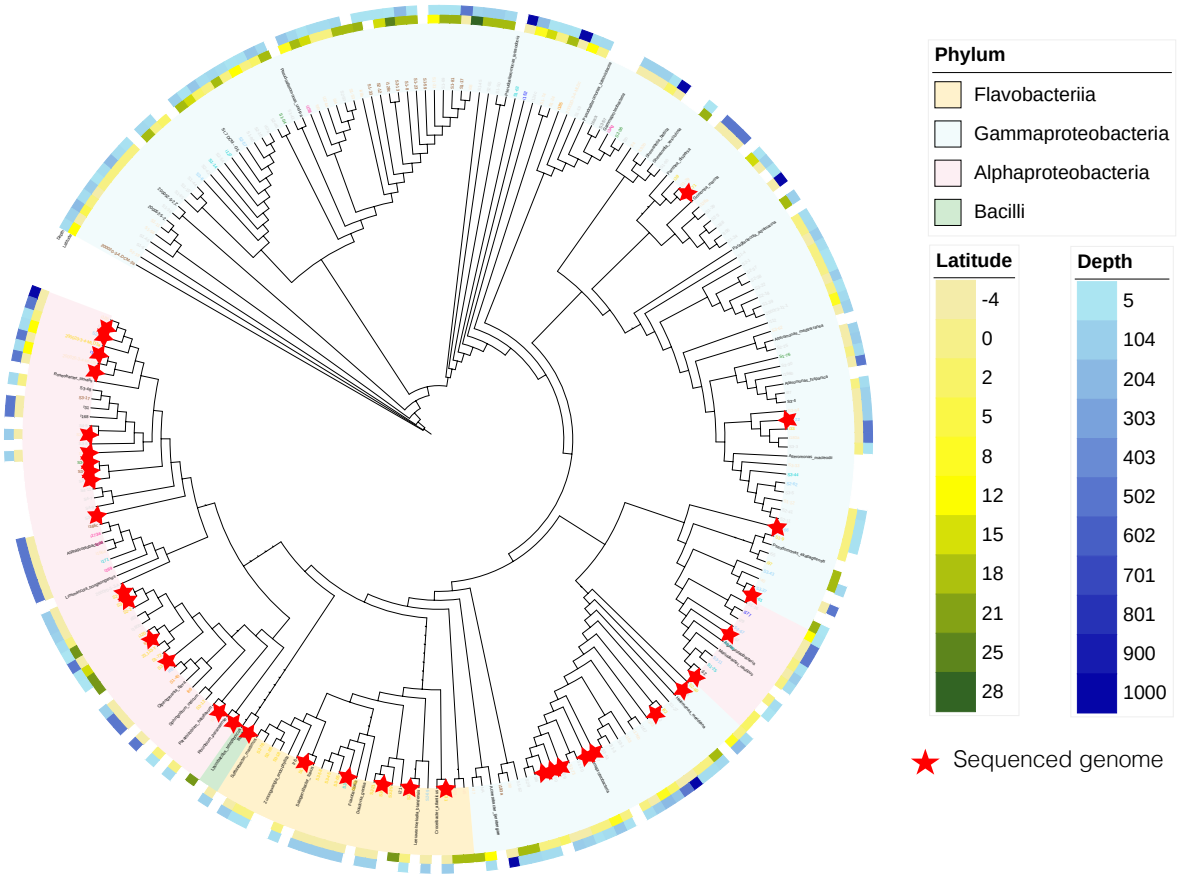


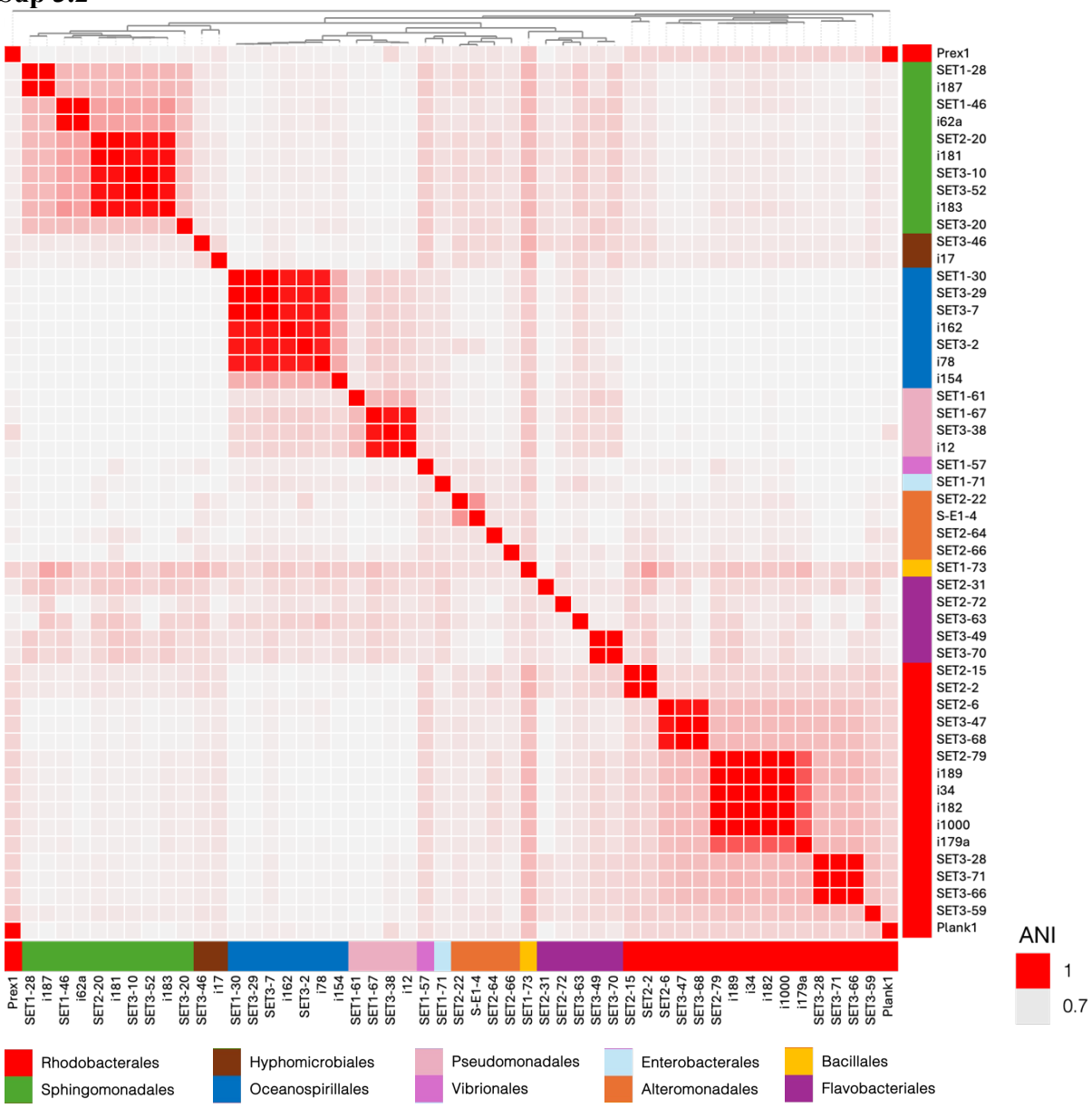
Fig. 3.3. Exchanged compounds between selected bacterial isolates. (A) Pathway for biosynthesis of indole-3-acetic acid (IAA) from tryptophan adapted from Khalil et al., 2024. Intermediate molecules are depicted in bold text. IAOX: Indole-3-acetaldoxime; IAN: Indole-3-acetonitrile; IAM: indole-3-acetamide; IPyA: Indole-3-pyruvate; TrypA: Tryptamine; IAAld: indole-3-acetaldehyde. Enzymes are in un-bold black text. CYP79B2/3: cytochrome P450; IAOxD: Indole-acetaldoxime dehydratase; NIT: Nitrilase; NH: Nitrile hydrolase; IaaM: Tryptophan-2-monooxygenase; IaaH: Indole-3-acetamide hydrolase; TSO: Tryptophan side chain oxidase; IPDC, TDC: Tryptophan decarboxylase; Indole-3-pyruvate decarboxylase; AO: Amine oxidase; IAAldD: indole-3-acetaldehyde dehydrogenase. Colored boxes show which bacterial isolates contain the genes required for each step of the pathway, while grey boxes indicate missing genes. **(B)** Exchanged phytohormones between bacterial isolates. Bar plot is colored by isolate and shows which isolates can produce a compound (right side of bar plot), and which isolates are able to consume the compound (left side of bar plot).

Supplemental Figures and Tables
Sup 3.1



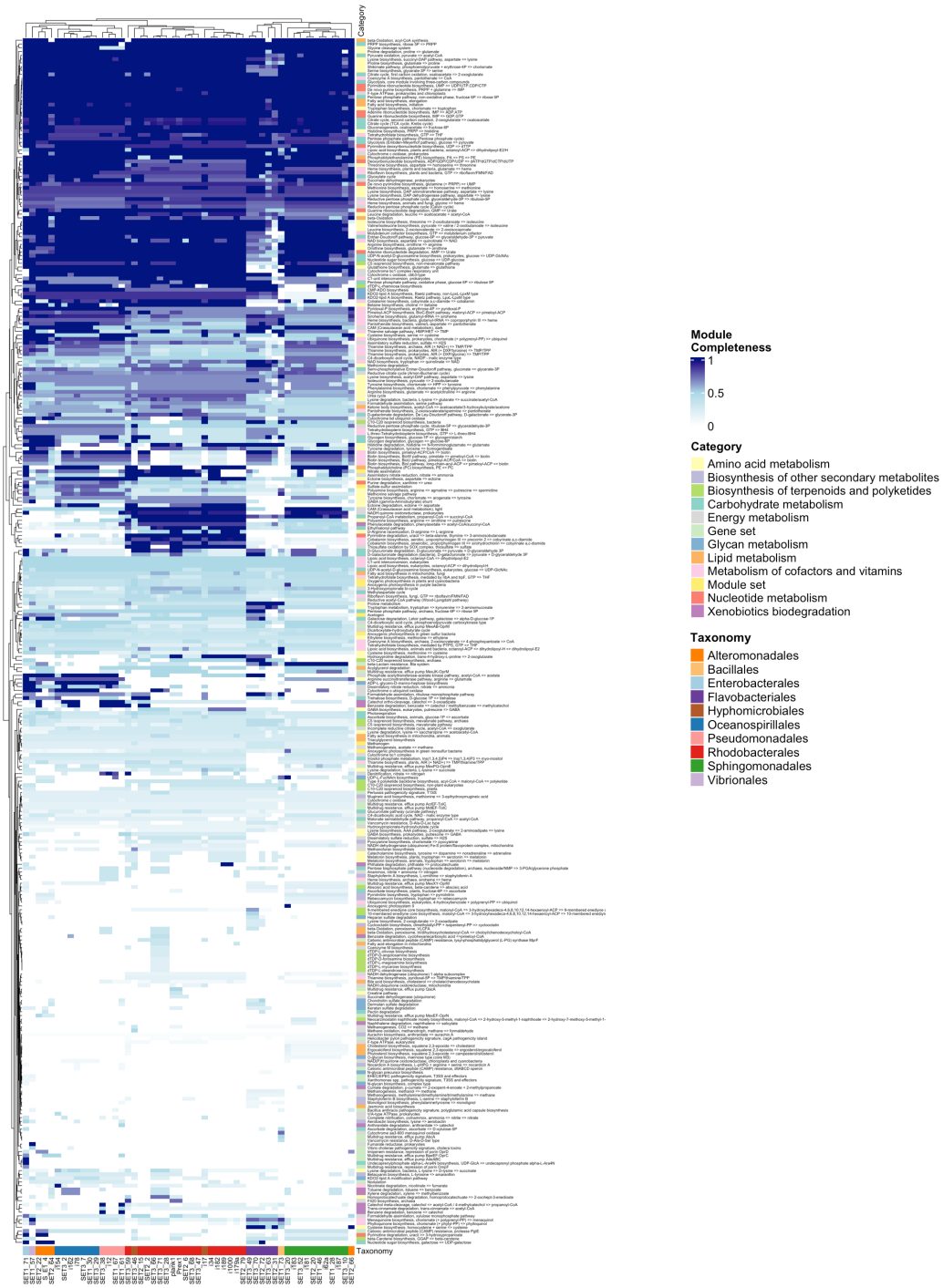
Sup 3.1. Phylogenetic tree of isolated strains based on 16S rRNA sequences. The tree was built using 16S rRNA sequences from all ~400 isolated strains, along with known reference sequences obtained from NCBI. The branches are colored by phylum. The latitude and depth of the isolation site for each isolate is shown in the heatmaps. Red stars indicate the 52 strains that were selected for whole genome sequencing.

Sup 3.2



Sup 3.2. Average nucleotide identity (ANI) across all isolate genomes. All-against-all bacterial isolate ANI heatmap clustered by isolate phylogeny. Color bars on the right and bottom show order level taxonomic classification for each bacteria.

Sup 3.3

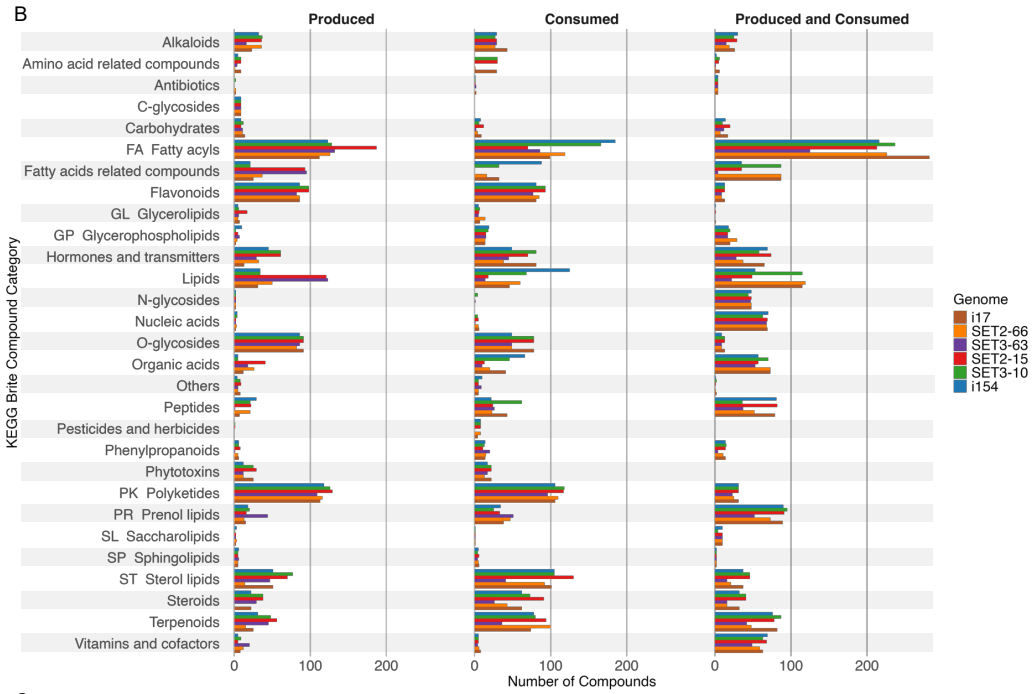


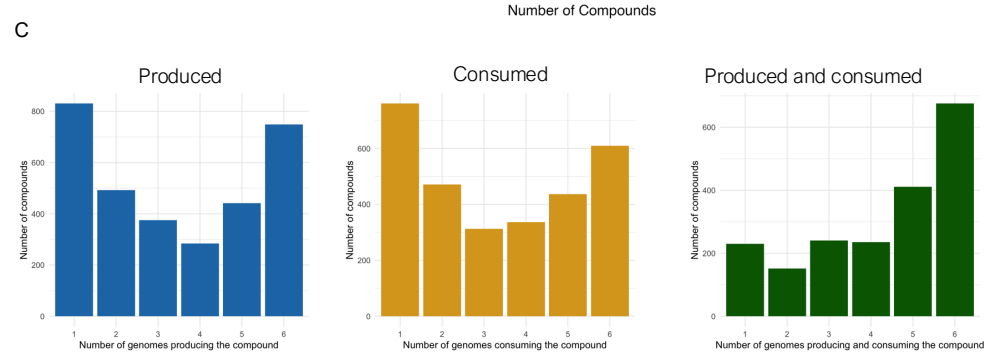
Sup 3.3. Metabolic capabilities of bacteria based on KEGG modules. Heatmap shows module completeness for each KEGG module (y-axis) for every genome (x-axis). The heatmap is clustered by module completeness. The color bar on the bottom shows the order level taxonomy of each bacteria isolate. The color bar on the left shows the KEGG category that each KEGG module belongs to.

Sup 3.4

A

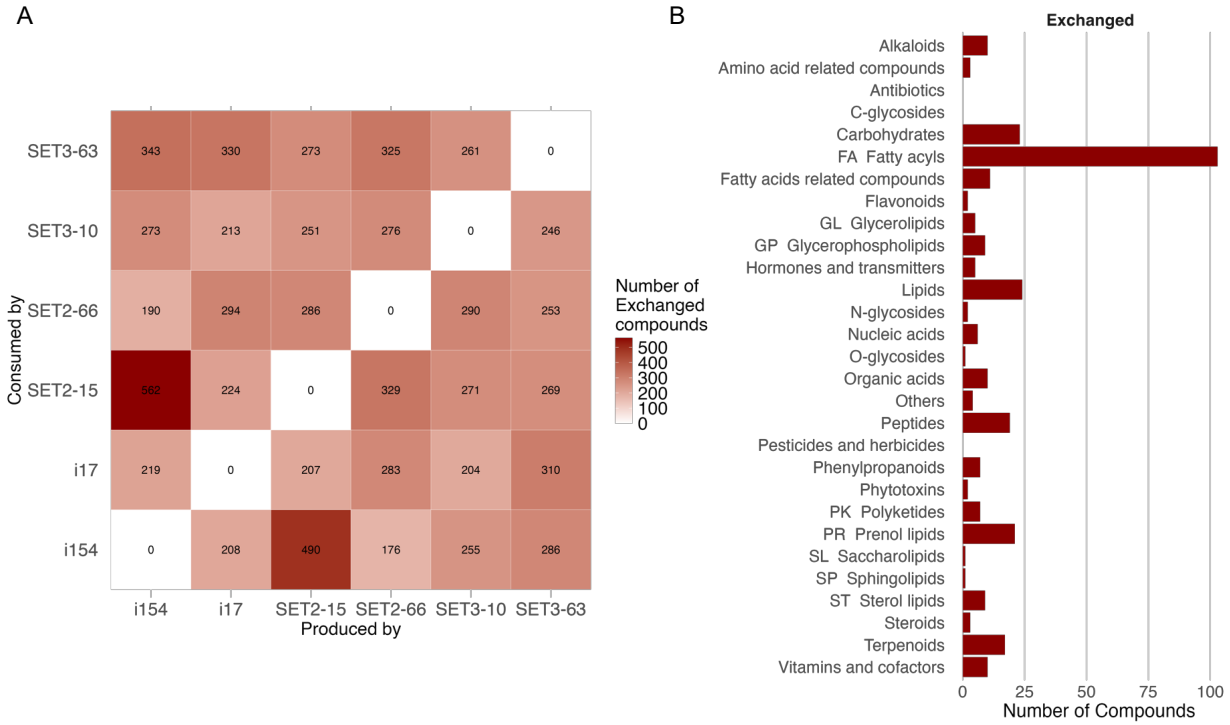
Genome	Taxonomy	Produced only	Consumed only	Produced and consumed
i154	Oceanospirillales	1996	1735	1520
i17	Hyphomicrobiales	1923	1796	1590
SET2_15	Rhodobacterales	1896	1766	1503
SET2_66	Alteromonadales	1685	1492	1387
SET3_10	Sphingomonadales	1793	1628	1336
SET3_63	Flavobacteriales	1475	1402	961





Sup 3.4. Produced and consumed compound for representative isolate genomes. Six genomes were assessed for their capacity to produce only, consume only and both produce and consume compounds based on completeness of KEGG pathway and modelSEED reactions using Anvio's –predict-metabolic-exchanges function. **(A)** Total number of produced only, consumed only and both produced and consumed compounds for each tested genome. **(B)** Number of compounds belonging to different KEGG Brite compound categories that are produced, consumed or both for each isolate. The bars are colored by isolate. **(C)** The number of compounds in each of the produced, consumed or both categories that is unique to one genome or shared between them.

Sup 3.5



Sup 3.5. Exchanged compounds between representative isolate genomes. (A) The total number of exchanged compounds between all genomes tested. Number of compounds produced by one genome and consumed by another are shown. **(B)** All exchanged compounds between any two genomes in the community categorized into KEGG Brite compound categories.

Sup 3.6

Compound	Produced by	Consumed by
Abundant marine metabolites		
Alanine	SET2-15; SET2-66; SET3-63; i154; i17	SET3-10
Aspartate	SET2-66; SET3-10; i154	SET2-15; i17
Citrate	SET2-15; SET2-66; SET3-63; i154; i17	SET3-10
Glutamate	SET3-10; i17	SET2-15; SET2-66; SET3-63; i154
Glutamine	SET2-66; i154	SET2-15; SET3-10; SET3-63; i17
Glycolate	SET2-66; SET3-10; SET3-63	SET2-15; i154; i17
Isethionate	i154	SET2-15; SET3-10; SET3-63; i17
N-Acetyl-D-glucosamine	SET3-63	SET2-15; SET2-66; SET3-10; i154; i17
Serine	SET2-15; SET3-10; i154; i17	SET2-66
Trehalose	SET2-66; i154; i17	SET2-15; SET3-10; SET3-63
Keystone marine metabolites		
4-Amino-5-aminomethyl-2-methylpyrimidine	SET2-15; SET2-66; SET3-63	SET2-15; i154; i17
Adenosylcobalamin-5'-phosphate	i17	SET2-15; i154
Alpha-Ribazole	SET2-15; i154	i17
Arachidic acid	SET2-15; SET3-63	SET2-66; SET3-10; i154; i17
Dethiobiotin	SET2-15; SET2-66; SET3-10; SET3-63	i154
Eicosapentaenoic acid (EPA)	SET2-15; SET3-63	SET2-66; SET3-10; i154; i17
Linolenate	SET2-15; SET3-63	SET2-66; SET3-10; i154; i17
Niacin (B3)	SET2-66; SET3-10; SET3-63; i17	SET2-15; i154
Bacterial toxins and metabolites		
Aklavinone	SET2-15; SET3-10; SET3-63; i154; i17	SET2-66
Premithramycin B	SET3-63	SET2-15; SET2-66; SET3-10; i154; i17
Protoanemonin	SET2-15	SET3-10; SET3-63; i154; i17
Carotenoids		
Lycopene	SET2-15; SET2-66; i154; i17	SET3-10; SET3-63
Zeaxanthin	SET3-63	SET2-15; SET2-66; SET3-10; i154; i17
beta-Carotene	SET2-66; SET3-10; SET3-63; i154	SET2-15; i17
Nitrogen / sulfur osmolytes		
(2R)-3-Sulfolactate	SET2-15; i17	SET2-66; SET3-10; SET3-63; i154
Betaine aldehyde	SET3-10; SET3-63	SET2-15; SET2-66; i154; i17
Carnitine	i154	SET2-15; SET2-66; SET3-10; i17
Ectoine	i154	SET2-15; SET3-10; i17
Gamma-butyrobetaine	SET2-15; SET2-66; SET3-10; i154; i17	SET3-63
Phyto regulation		
Abscisic acid	SET2-15; SET3-10; SET3-63; i154; i17	SET2-15; SET2-66; SET3-10; i154; i17
Brassinolide	SET2-15; SET2-66; SET3-10; i154; i17	SET3-63
Gibberellin A34	SET2-15; SET3-10	SET2-66; SET3-63; i154; i17
Gibberellin A4	SET2-15; SET2-66; SET3-10; i154	SET3-63; i17
Gibberellin A9	SET2-15; SET2-66; SET3-10; i154	SET3-63; i17
IAA-Ala	i154; i17	SET2-15; SET2-66; SET3-63
IAA-Asp	SET2-15; i154; i17	SET2-66; SET3-10; SET3-63
IAA-Asp-N-Glc	SET3-10; i154; i17	SET2-15; SET2-66; SET3-63
IAA-Glu	SET2-15; i154; i17	SET2-66; SET3-10; SET3-63
IAA-Glu-N-Glc	SET3-10; i154; i17	SET2-15; SET2-66; SET3-63
IAA-Leu	i154; i17	SET2-15; SET2-66; SET3-63
Jasmonate	SET2-15; i17	i154
N-Hydroxyl-tryptamine	SET2-66	SET2-15; SET3-10'; i154; i17
Other Carbohydrates		
Alginate	SET2-66; SET3-10; i17	SET3-63
Arginine	SET2-66; i154	SET2-15; SET3-10; i17
Guanosine	SET3-10; SET3-63	SET2-15; SET2-66; i154; i17
Sarcosine	SET2-15; i17	SET2-66; SET3-10; SET3-63; i154
Secoisolariciresinol	SET2-66	SET2-15; SET3-10; SET3-63; i154; i17
Spermidine	SET2-15; SET2-66; SET3-63; i154; i17	SET3-10
Sulfoacetaldehyde	SET3-10; SET3-63	SET2-15; i154; i17
L-Ascorbate	SET3-10; SET3-63	SET2-15; SET2-66; i154; i17
Nicotinamide	SET2-66	SET2-15; SET3-10; SET3-63; i154; i17
Tetrahydrobiopterin	SET3-63	SET2-15; SET2-66; SET3-10; i154; i17
Other Fatty Acyls		
(R)-3-Hydroxybutanoyl-CoA	SET2-15; SET2-66; i154; i17	SET3-10; SET3-63
Butyrate	SET2-15; SET2-66	SET2-66; SET3-10; i154; i17
Butyryl-CoA	SET2-66; SET3-10; SET3-63; i154; i17	SET2-15
D-Mucic¬†acid	SET2-15; i154	i17
Decanoate	SET2-15; SET3-63	SET2-66; SET3-10; i154; i17
Decanoyl-CoA	SET2-66; SET3-10; i154; i17	SET2-15; SET3-63
Oleate	SET3-63	SET2-15; SET2-66; SET3-10; i154; i17
Palmitate	SET3-63	SET2-15; SET2-66; SET3-10; i154; i17
Palmitoyl-CoA	SET2-66; SET3-10; i154; i17	SET2-15; SET3-63
Other Terpenoids		
Androstenedione	SET3-63	SET2-15; SET2-66; SET3-10; i154; i17
Farnesyldiphosphate	SET2-66; SET3-10; i154; i17	SET2-15; SET2-15; SET3-63
Geranyldiphosphate	SET2-66; SET3-10; i154; i17	SET2-15
Solanesyl diphosphate	SET2-15; SET3-10; i154; i17	SET2-66; SET3-63

Sup 3.6. Table of exchanged compounds between representative isolate genomes. Showing example exchanged compounds belonging to ecologically relevant categories. Genomes that have the capability to produce or consume them are listed.

Sup 3.7

Compound	Consumption reaction	Production Reaction
N-Hydroxyl-tryptamine	**rxn32170** N-Hydroxyl-tryptamine <=> (2) H+ + 3-Indoleacetaldoxime	**rxn14082** NADPH + O2 + Tryptamine => H2O + NADP + N-Hydroxyl-tryptamine
Jasmonate	**rxn20909** ATP + L-Isoleucine + Jasmonate => PPi + AMP + (2) H+ + (-)-Jasmonoyl-L-isoleucine	**rxn40507** H2O + Methyl jasmonate <=> H+ + Methanol + Jasmonate
IAA-Leu	**rxn23828** H2O + IAA-Leu <=> L-Leucine + Indoleacetate	**rxn23821** ATP + L-Leucine + Indoleacetate => PPi + AMP + (2) H+ + IAA-Leu
IAA-Glu	**rxn25353** Glucosylated-Glucose-Acceptors + IAA-Glu <=> IAA-Glu-N-Glc + Non-Glucosylated-Glucose-Acceptor	**rxn23823** ATP + L-Glutamate + Indoleacetate => PPi + AMP + (2) H+ + IAA-Glu
IAA-Asp	**rxn25352** Glucosylated-Glucose-Acceptors + IAA-Asp <=> IAA-Asp-N-Glc + Non-Glucosylated-Glucose-Acceptors	**rxn23822** ATP + L-Aspartate + Indoleacetate => PPi + AMP + (2) H+ + IAA-Asp
IAA-Ala	**rxn23826** H2O + IAA-Ala <=> L-Alanine + Indoleacetate	**rxn23820** ATP + L-Alanine + Indoleacetate => PPi + AMP + (2) H+ + IAA-Ala
Gibberellin A9	**rxn21726** S-Adenosyl-L-methionine + Gibberellin A9 <=> S-Adenosyl-homocysteine + GA9 methyl ester	**rxn24334** NADPH + O2 + Gibberellin A12 <=> NADP + CO2 + (4) H+ + Gibberellin A9
Gibberellin A4	**rxn21727** S-Adenosyl-L-methionine + Gibberellin A4 <=> S-Adenosyl-homocysteine + GA4	**rxn24332** NADPH + O2 + Gibberellin A14 <=> NADP + CO2 + (4) H+ + Gibberellin A4
Brassinolide	**rxn27090** Brassinolide + UDP-GLUCOSE <=> UDP + brassinolide-23-O-glucoside	rxn24258 NADPH + O2 + H+ + Castasterone => H2O + NADP + Brassinolide

Sup 3.7. Reactions responsible for the exchange of phytohormones between bacteria isolates. The modelSEED production and consumption reaction is shown for each exchanged phytohormone.

Sup. Table 3.1. Equatorial Pacific bacterial genomes. Total number of circular contigs is shown, along with GC content, completeness and contamination, total size of each genomes and number of coding DNA sequences (CDS). The location and depth of isolation site is stated for each genome. Closest species relative based on 16S rDNA amplicon data as well as based on the genome is stated, including the closest closed genome representative and the closest environmental Metagenome Amplified Genome (MAG) or Single Amplified Genome (SAG) representative. Average nucleotide identity (ANI) to closest relatives is shown.

Sup. Table 3.2. Phylogenetic tree refence sequences. List of reference bacteria and associated 16S rDNA NCBI accession numbers that were included in the 16S rDNA phylogenetic tree of all bacterial isolates.

References

Amin, S. A., Hmelo, L. R., Van Tol, H. M., Durham, B. P., Carlson, L. T., Heal, K. R., Morales, R. L., Berthiaume, C. T., Parker, M. S., Djunaedi, B., Ingalls, A. E., Parsek, M. R., Moran, M. A., & Armbrust, E. V. (2015). Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. *Nature*, *522*(7554), 98–101. https://doi.org/10.1038/nature14488

Amin, S. A., Parker, M. S., & Armbrust, E. V. (2012). Interactions between Diatoms and Bacteria. *Microbiology and Molecular Biology Reviews*, *76*(3), 667–684. https://doi.org/10.1128/MMBR.00007-12

Arandia-Gorostidi, N., Montiel, L., Latorre, F., Balagué, V., Simó, R., Giner, C. R., & Logares, R. (n.d.). *Metagenomic-based network analysis reveals the importance of vitamin cross-feeding in marine microbial assemblages*.

Arkin, A. P., Cottingham, R. W., Henry, C. S., Harris, N. L., Stevens, R. L., Maslov, S., Dehal, P., Ware, D., Perez, F., Canon, S., Sneddon, M. W., Henderson, M. L., Riehl, W. J., Murphy-Olson, D., Chan, S. Y., Kamimura, R. T., Kumari, S., Drake, M. M., Brettin, T. S., … Yu, D. (2018). KBase: The United States Department of Energy Systems Biology Knowledgebase. *Nature Biotechnology*, *36*(7), 566–569. https://doi.org/10.1038/nbt.4163

Bairey, E., Kelsic, E. D., & Kishony, R. (2016). High-order species interactions shape ecosystem diversity. *Nature Communications*, *7*(1), 12285. https://doi.org/10.1038/ncomms12285

Bannon, C., White, P. L., Rowland, E., More, K. J., Gleason, A., Roberts, M., Devred, E., Beazley, L., LaRoche, J., & Bertrand, E. M. (2024). *Seasonal patterns in b-vitamins and cobalamin co-limitation in the Northwest Atlantic*. Microbiology. https://doi.org/10.1101/2024.11.10.622835

Bartolek, Z., Van Creveld, S. G., Coesel, S., Cain, K. R., Schatz, M., Morales, R., & Virginia Armbrust, E. (2022). Flavobacterial exudates disrupt cell cycle progression and metabolism of the diatom *Thalassiosira pseudonana*. *The ISME Journal*, *16*(12), 2741–2751. https://doi.org/10.1038/s41396-022-01313-9

Blin, K., Wolf, T., Chevrette, M. G., Lu, X., Schwalen, C. J., Kautsar, S. A., Suarez Duran, H. G., de los Santos, E. L. C., Kim, H. U., Nave, M., Dickschat, J. S., Mitchell, D. A., Shelest, E., Breitling, R., Takano, E., Lee, S. Y., Weber, T., & Medema, M. H. (2017). antiSMASH 4.0—Improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Research*, *45*(W1), W36–W41. https://doi.org/10.1093/nar/gkx319

Bobadilla Fazzini, R. A., Skindersoe, M. E., Bielecki, P., Puchałka, J., Givskov, M., & Martins Dos Santos, V. A. P. (2013). Protoanemonin: A natural quorum sensing inhibitor that selectively activates iron starvation response. *Environmental Microbiology*, *15*(1), 111–120. https://doi.org/10.1111/j.1462-2920.2012.02792.x

Boysen, A. K., Carlson, L. T., Durham, B. P., Groussman, R. D., Aylward, F. O., Ribalet, F., Heal, K. R., White, A. E., DeLong, E. F., Armbrust, E. V., & Ingalls, A. E. (2021). Particulate Metabolites and Transcripts Reflect Diel Oscillations of Microbial Activity in the Surface Ocean. *mSystems*, *6*(3), e00896-20. https://doi.org/10.1128/mSystems.00896-20

Boysen, A. K., Durham, B. P., Kumler, W., Key, R. S., Heal, K. R., Carlson, L. T., Groussman, R. D., Armbrust, E. V., & Ingalls, A. E. (2022). Glycine betaine uptake and metabolism in marine microbial communities. *Environmental Microbiology*, *24*(5), 2380–2403. https://doi.org/10.1111/1462-2920.16020

Bushnell, B. (2014). BBnorm: Kmer-based error-correction and normalization tool (from the BBTools package). *In SourceForge Repository*. https://sourceforge.net/projects/bbmap/

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, *10*(1), 421. https://doi.org/10.1186/1471-2105-10-421

Chao, K.-H., Barton, K., Palmer, S., & Lanfear, R. (2021). sangeranalyseR: Simple and Interactive Processing of Sanger Sequencing Data in R. *Genome Biology and Evolution*, *13*(3), evab028. https://doi.org/10.1093/gbe/evab028

Croft, M. T., Lawrence, A. D., Raux-Deery, E., Warren, M. J., & Smith, A. G. (2005). Algae acquire vitamin B12 through a symbiotic relationship with bacteria. *Nature*, *438*(7064), 90–93. https://doi.org/10.1038/nature04056

Dai, M., Luo, Y., Achterberg, E. P., Browning, T. J., Cai, Y., Cao, Z., Chai, F., Chen, B., Church, M. J., Ci, D., Du, C., Gao, K., Guo, X., Hu, Z., Kao, S., Laws, E. A., Lee, Z., Lin, H., Liu, Q., … Zhou, K. (2023). Upper Ocean Biogeochemistry of the Oligotrophic North Pacific Subtropical Gyre: From Nutrient Sources to Carbon Export. *Reviews of Geophysics*, *61*(3), e2022RG000800. https://doi.org/10.1029/2022RG000800

De Carvalho, C., & Caramujo, M. (2018). The Various Roles of Fatty Acids. *Molecules*, *23*(10), 2583. https://doi.org/10.3390/molecules23102583

Durham, B. P., Boysen, A. K., Carlson, L. T., Groussman, R. D., Heal, K. R., Cain, K. R., Morales, R. L., Coesel, S. N., Morris, R. M., Ingalls, A. E., & Armbrust, E. V. (2019). Sulfonate-based networks between eukaryotic phytoplankton and heterotrophic bacteria in the surface ocean. *Nature Microbiology*, *4*(10), 1706–1715. https://doi.org/10.1038/s41564-019-0507-5

Durham, B. P., Dearth, S. P., Sharma, S., Amin, S. A., Smith, C. B., Campagna, S. R., Armbrust, E. V., & Moran, M. A. (2017). Recognition cascade and metabolite transfer in a marine bacteria-phytoplankton model system. *Environmental Microbiology*, *19*(9), 3500–3513. https://doi.org/10.1111/1462-2920.13834

Durham, B. P., Johnson, W. M., Bannon, C. C., Bertrand, E. M., Ingalls, A. E., Edwards, B. R., Apprill, A., Boysen, A. K., Bundy, R. M., Chen, H., Ferrer-González, F. X., Fiore, C., Heal, K. R., Kuhlisch, C., Liu, S., Lu, K., Meke, L. E., Pontrelli, S., Vaiyapuri Ramalingam, P., … Kujawinski, E. B. (2025). An ecological framework for microbial metabolites in the ocean ecosystem. *Limnology and Oceanography Letters*, lol2.70046. https://doi.org/10.1002/lol2.70046

Eren, A. M., Esen, Ö. C., Quince, C., Vineis, J. H., Morrison, H. G., Sogin, M. L., & Delmont, T. O. (2015). Anvi'o: An advanced analysis and visualization platform for 'omics data. *PeerJ*, *3*, e1319. https://doi.org/10.7717/peerj.1319

Fei, C., Ochsenkühn, M. A., Shibl, A. A., Isaac, A., Wang, C., & Amin, S. A. (2020). Quorum sensing regulates 'swim-or-stick' lifestyle in the phycosphere. *Environmental Microbiology*, *22*(11), 4761–4778. https://doi.org/10.1111/1462-2920.15228

Ferrer-González, F. X., Widner, B., Holderman, N. R., Glushka, J., Edison, A. S., Kujawinski, E. B., & Moran, M. A. (2021). Resource partitioning of phytoplankton metabolites that support bacterial heterotrophy. *The ISME Journal*, *15*(3), 762–773. https://doi.org/10.1038/s41396-020-00811-y

Freese, H. M., Ringel, V., Overmann, J., & Petersen, J. (2022). Beyond the ABCs—Discovery of Three New Plasmid Types in Rhodobacterales (RepQ, RepY, RepW). *Microorganisms*, *10*(4), 738. https://doi.org/10.3390/microorganisms10040738

Fuessel, J., Miller, S., Hecker, C., Trigodet, F., Dlugosch, L., Brinkhoff, T., Dittmar, T., Eren, A. M., Lennartz, S., & Bunse, C. (2025). *Bacterial interactions shape the molecular composition of dissolved organic matter*. In Review. https://doi.org/10.21203/rs.3.rs-7321332/v1

Giordano, N., Gaudin, M., Trottier, C., Delage, E., Nef, C., Bowler, C., & Chaffron, S. (2024). Genome-scale community modelling reveals conserved metabolic cross-feedings in epipelagic bacterioplankton communities. *Nature Communications*, *15*(1), 2721. https://doi.org/10.1038/s41467-024-46374-w

Graham, E. D., Heidelberg, J. F., & Tully, B. J. (2018). Potential for primary productivity in a globally-distributed bacterial phototroph. *The ISME Journal*, *12*(7), 1861–1866. https://doi.org/10.1038/s41396-018-0091-3

Green, E. R., & Mecsas, J. (2016). Bacterial Secretion Systems: An Overview. *Microbiology Spectrum*, *4*(1), 4.1.13. https://doi.org/10.1128/microbiolspec.VMBF-0012-2015

Hansell, D. A. (2013). Recalcitrant Dissolved Organic Carbon Fractions. *Annual Review of Marine Science*, *5*(1), 421–445. https://doi.org/10.1146/annurev-marine-120710-100757

Hayashi, K., Arai, K., Aoi, Y., Tanaka, Y., Hira, H., Guo, R., Hu, Y., Ge, C., Zhao, Y., Kasahara, H., & Fukui, K. (2021). The main oxidative inactivation pathway of the plant hormone auxin. *Nature Communications*, *12*(1), 6752. https://doi.org/10.1038/s41467-021-27020-1

Iwaki, H., Nishimura, A., & Hasegawa, Y. (2012). Tropicibacter phthalicus sp. Nov., A Phthalate-Degrading Bacterium from Seawater. *Current Microbiology*, *64*(4), 392–396. https://doi.org/10.1007/s00284-012-0085-8

Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., & Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications*, *9*(1), 5114. https://doi.org/10.1038/s41467-018-07641-9

Kanehisa, M., Furumichi, M., Sato, Y., Matsuura, Y., & Ishiguro-Watanabe, M. (2025). KEGG: Biological systems database as a model of the real world. *Nucleic Acids Research*, *53*(D1), D672–D677. https://doi.org/10.1093/nar/gkae909

Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

Khalil, A., Bramucci, A. R., Focardi, A., Le Reun, N., Willams, N. L. R., Kuzhiumparambil, U., Raina, J.-B., & Seymour, J. R. (2024). Widespread production of plant growth-promoting hormones among marine bacteria and their impacts on the growth of a marine diatom. *Microbiome*, *12*(1), 205. https://doi.org/10.1186/s40168-024-01899-6

Kolmogorov, M., Yuan, J., Lin, Y., & Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology*, *37*(5), 540–546. https://doi.org/10.1038/s41587-019-0072-8

Konstantinidis, K. T., & Tiedje, J. M. (2005). Genomic insights that advance the species definition for prokaryotes. *Proceedings of the National Academy of Sciences*, *102*(7), 2567–2572. https://doi.org/10.1073/pnas.0409727102

Labeeuw, L., Khey, J., Bramucci, A. R., Atwal, H., De La Mata, A. P., Harynuk, J., & Case, R. J. (2016). Indole-3-Acetic Acid Is Produced by Emiliania huxleyi Coccolith-Bearing Cells and Triggers a Physiological Response in Bald Cells. *Frontiers in Microbiology*, *7*. https://doi.org/10.3389/fmicb.2016.00828

Laguna, R., Joshi, G. S., Dangel, A. W., Luther, A. K., Tabita, F. R., & Hallenbeck, P. (2010). Integrative Control of Carbon, Nitrogen, Hydrogen, and Sulfur Metabolism: The Central Role of the Calvin-Benson-Bassham Cycle. *Advances in Experimental Medicine and Biology, 675*, 265–271.

Letunic, I., & Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Research*, *49*(W1), W293–W296. https://doi.org/10.1093/nar/gkab301

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18), 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Mayali, X., Samo, T. J., Kimbrel, J. A., Morris, M. M., Rolison, K., Swink, C., Ramon, C., Kim, Y.-M., Munoz-Munoz, N., Nicora, C., Purvine, S., Lipton, M., Stuart, R. K., & Weber, P. K. (2023). Single-cell isotope tracing reveals functional guilds of bacteria associated with the diatom Phaeodactylum tricornutum. *Nature Communications*, *14*(1), 5642. https://doi.org/10.1038/s41467-023-41179-9

Mayali, X., Weber, P. K., Brodie, E. L., Mabery, S., Hoeprich, P. D., & Pett-Ridge, J. (2012). High-throughput isotopic analysis of RNA microarrays to quantify microbial resource use. *The ISME Journal*, *6*(6), 1210–1221. https://doi.org/10.1038/ismej.2011.175

Michael, V., Frank, O., Bartling, P., Scheuner, C., Göker, M., Brinkmann, H., & Petersen, J. (2016). Biofilm plasmids with a rhamnose operon are widely distributed determinants of the 'swim-or-stick' lifestyle in roseobacters. *The ISME Journal*, *10*(10), 2498–2513. https://doi.org/10.1038/ismej.2016.30

Mikheenko, A., Saveliev, V., Hirsch, P., & Gurevich, A. (2023). WebQUAST: Online evaluation of genome assemblies. *Nucleic Acids Research*, *51*(W1), W601–W606. https://doi.org/10.1093/nar/gkad406

Moran, M. A., Ferrer-González, F. X., Fu, H., Nowinski, B., Olofsson, M., Powers, M. A., Schreier, J. E., Schroer, W. F., Smith, C. B., & Uchimiya, M. (2022). The Ocean's labile DOC

supply chain. *Limnology and Oceanography*, *67*(5), 1007–1021. https://doi.org/10.1002/lno.12053

Nichols, D. S. (2003). Prokaryotes and the input of polyunsaturated fatty acids to the marine food web. *FEMS Microbiology Letters*, *219*(1), 1–7. https://doi.org/10.1016/S0378-1097(02)01200-4

Olson, R. D., Assaf, R., Brettin, T., Conrad, N., Cucinell, C., Davis, J. J., Dempsey, D. M., Dickerman, A., Dietrich, E. M., Kenyon, R. W., Kuscuoglu, M., Lefkowitz, E. J., Lu, J., Machi, D., Macken, C., Mao, C., Niewiadomska, A., Nguyen, M., Olsen, G. J., … Stevens, R. L. (2023). Introducing the Bacterial and Viral Bioinformatics Resource Center (BV-BRC): A resource combining PATRIC, IRD and ViPR. *Nucleic Acids Research*, *51*(D1), D678–D689. https://doi.org/10.1093/nar/gkac1003

Paerl, R. W., Curtis, N. P., Bittner, M. J., Cohn, M. R., Gifford, S. M., Bannon, C. C., Rowland, E., & Bertrand, E. M. (2023). Use and detection of a vitamin B1 degradation product yields new views of the marine B1 cycle and plankton metabolite exchange. *mBio*, e00061-23. https://doi.org/10.1128/mbio.00061-23

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., & Tyson, G. W. (2015). CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*, *25*(7), 1043–1055. https://doi.org/10.1101/gr.186072.114

Pennington, J. T., Mahoney, K. L., Kuwahara, V. S., Kolber, D. D., Calienes, R., & Chavez, F. P. (2006). Primary production in the eastern tropical Pacific: A review. *Progress in Oceanography*, *69*(2–4), 285–317. https://doi.org/10.1016/j.pocean.2006.03.012

Petersen, J., & Wagner-Döbler, I. (2017). Plasmid Transfer in the Ocean – A Case Study from the Roseobacter Group. *Frontiers in Microbiology*, *8*, 1350. https://doi.org/10.3389/fmicb.2017.01350

Rodriguez-R, L. M., Conrad, R. E., Viver, T., Feistel, D. J., Lindner, B. G., Venter, F., Orellana, L., Amann, R., Rossello-Mora, R., & Konstantinidis, K. T. (2022). *An ANI gap within bacterial species that advances the definitions of intra-species units*.

Seaver, S. M. D., Liu, F., Zhang, Q., Jeffryes, J., Faria, J. P., Edirisinghe, J. N., Mundy, M., Chia, N., Noor, E., Beber, M. E., Best, A. A., DeJongh, M., Kimbrel, J. A., D'haeseleer, P., McCorkle, S. R., Bolton, J. R., Pearson, E., Canon, S., Wood-Charlson, E. M., … Henry, C. S. (2021). The ModelSEED Biochemistry Database for the integration of metabolic annotations and the reconstruction, comparison and analysis of metabolic models for plants, fungi and microbes. *Nucleic Acids Research*, *49*(D1), D575–D588. https://doi.org/10.1093/nar/gkaa746

Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics*, *30*(14), 2068–2069. https://doi.org/10.1093/bioinformatics/btu153

Seymour, J. R., Amin, S. A., Raina, J.-B., & Stocker, R. (2017). Zooming in on the phycosphere: The ecological interface for phytoplankton–bacteria relationships. *Nature Microbiology*, *2*(7), 17065. https://doi.org/10.1038/nmicrobiol.2017.65

Shibl, A. A., Isaac, A., Ochsenkühn, M. A., Cárdenas, A., Fei, C., Behringer, G., Arnoux, M., Drou, N., Santos, M. P., Gunsalus, K. C., Voolstra, C. R., & Amin, S. A. (2020). Diatom

modulation of select bacteria through use of two unique secondary metabolites. *Proceedings of the National Academy of Sciences*, *117*(44), 27445–27455. https://doi.org/10.1073/pnas.2012088117

Sosa, O. A., Repeta, D. J., DeLong, E. F., Ashkezari, M. D., & Karl, D. M. (2019). Phosphate-limited ocean regions select for bacterial populations enriched in the carbon–phosphorus lyase pathway for phosphonate degradation. *Environmental Microbiology*, *21*(7), 2402–2414. https://doi.org/10.1111/1462-2920.14628

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Sun, X., & Ward, B. B. (2021). Novel metagenome-assembled genomes involved in the nitrogen cycle from a Pacific oxygen minimum zone. *ISME Communications*, *1*(1), 26. https://doi.org/10.1038/s43705-021-00030-2

Van Tol, H. M., Amin, S. A., & Armbrust, E. V. (2017). Ubiquitous marine bacterium inhibits diatom cell division. *The ISME Journal*, *11*(1), 31–42. https://doi.org/10.1038/ismej.2016.112

Vaser, R., Sović, I., Nagarajan, N., & Šikić, M. (2017). Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Research*, *27*(5), 737–746. https://doi.org/10.1101/gr.214270.116

Veseli, I., Chen, Y. T., Schechter, M. S., Vanni, C., Fogarty, E. C., Watson, A. R., Jabri, B., Blekhman, R., Willis, A. D., Yu, M. K., Fernàndez-Guerra, A., Füssel, J., & Eren, A. M. (2025). Microbes with higher metabolic independence are enriched in human gut microbiomes under stress. *eLife*, *12*, RP89862. https://doi.org/10.7554/eLife.89862

Wasternack, C., & Strnad, M. (2018). Jasmonates: News on Occurrence, Biosynthesis, Metabolism and Action of an Ancient Group of Signaling Compounds. *International Journal of Molecular Sciences*, *19*(9), 2539. https://doi.org/10.3390/ijms19092539

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., & Barton, G. J. (2009). Jalview Version 2—A multiple sequence alignment editor and analysis workbench. *Bioinformatics*, *25*(9), 1189–1191. https://doi.org/10.1093/bioinformatics/btp033

Wick, R. R., Judd, L. M., & Holt, K. E. (2023). Assembling the perfect bacterial genome using Oxford Nanopore and Illumina sequencing. *PLOS Computational Biology*, *19*(3), e1010905. https://doi.org/10.1371/journal.pcbi.1010905

Wienhausen, G., Bruns, S., Sultana, S., Dlugosch, L., Groon, L.-A., Wilkes, H., & Simon, M. (2022). The overlooked role of a biotin precursor for marine bacteria—Desthiobiotin as an escape route for biotin auxotrophy. *The ISME Journal*, *16*(11), 2599–2609. https://doi.org/10.1038/s41396-022-01304-w

Zakem, E. J., McNichol, J., Weissman, J. L., Raut, Y., Xu, L., Halewood, E. R., Carlson, C. A., Dutkiewicz, S., Fuhrman, J. A., & Levine, N. M. (2024). *Functional biogeography of marine microbial heterotrophs*. Ecology. https://doi.org/10.1101/2024.02.14.580411

Zoccarato, L., Sher, D., Miki, T., Segrè, D., & Grossart, H.-P. (2022). A comparative whole-genome approach identifies bacterial traits for marine microbial interactions. *Communications Biology*, *5*(1), 276. https://doi.org/10.1038/s42003-022-03184-4

Chapter 4: Ocean Genomes to Explore Microbial Metabolism of Phytoplankton Organic Matter

Frank Xavier Ferrer-González[1] and Zinka Bartolek[1], E. Virginia Armbrust[1], Anitra E. Ingalls[1], Oscar A. Sosa[2]

[1]School of Oceanography, University of Washington, Seattle, WA, USA
[2]Department of Biology, University of Puget Sound, Tacoma, WA, US

Abstract

This course-based undergraduate research experience (CURE) exposes students to concepts, laboratory techniques and data analysis skills at the intersection of genetics and functional genomics through exploration of the metabolism of marine bacteria. Guided by questions, students begin by reading research articles that describe the types of protein functions marine bacteria require to utilize metabolites derived from phytoplankton, the primary producers of the ocean. Students learn wet lab microbiology skills, and data analysis through R programming to conduct a high-throughput screen to test the capacity of a diverse collection of marine bacteria to grow on selected phytoplankton metabolites. Students examine their data collectively to select bacteria isolates for DNA extraction and whole genome sequencing. They learn to navigate online bioinformatics tools and databases to analyze the information and gene annotations in the bacterial genomes. The project culminates with an independent inquiry assignment in which students use genome evidence, the scientific literature, and their knowledge of bacterial gene operons and protein functions to reconstruct a genetic and metabolic pathway that could explain the growth phenotype of their bacterium on marine metabolites. The project incorporates assessments related to group discussions of peer-reviewed research articles, completion of questions based on the articles and lab protocols, weekly lab reports, and a final lab report and research presentation to peers. The project highlights the utility of genomic data to solve interdisciplinary problems and the functions performed by microorganisms in the environment.

Scientific Teaching Context

Learning Goals for the Course

This CURE aims to provide students with a foundation in core concepts and principles at the intersection of genetics and genome science. The project trains students to read and interpret peer-reviewed primary scientific sources, perform research based on authentic questions from ongoing studies, make informed decisions based on lab results, use bioinformatics tools to analyze genomic data they generate, and communicate and write about their findings and contemporary topics in genomics.

By the end of this course, students will be able to:

- Identify common bioinformatics tools and protein databases as resources to study gene sequences and explore the functions of genes in microbial genomes.
- Illustrate how to retrieve and critically assess credible scientific sources, and synthesize the literature to support background research and interpretation of findings.
- Distinguish in the scientific literature evidence derived from genetic research and genomic data versus other lines of evidence.
- Accurately apply basic genetic and genomic concepts to communicate research findings clearly in written, oral, or digital formats.
- Compare information found in genomes to organismal traits or phenotypes by creating visual representations of metabolic pathways based on genomic functional annotations (e.g., KEGG metabolic maps).
- Evaluate how genomic research contributes to interdisciplinary fields such as our understanding of microbial carbon and nitrogen cycling in the ocean.

Selected Learning Objectives

We selected two exemplar lessons from this course to elaborate on. The first lesson focuses on a lab experiment investigating the metabolic interactions between marine bacteria and phytoplankton by testing the growth of 94 bacterial isolates on a diverse range of phytoplankton-derived metabolites using a high-throughput screening assay. The second lesson helps students develop an independent research project, connecting the observed growth phenotypes of bacteria isolates to their metabolic capabilities predicted through genome analysis, using instructor-provided genome assemblies and annotations derived from DNA extracted and sequenced earlier in the course. These lessons were chosen because each offers a distinct set of concepts and skills: lesson one emphasizes collaboration and lab-based studies, while lesson two promotes students to develop original hypotheses, design and carry out analysis methods to test their own questions. Together, both of these lessons connect overarching concepts across the entire course.

<u>Learning goals for exemplar lessons:</u>

Lesson #1: Investigating the metabolic profiles of marine bacteria

- Explain High-Throughput Screening approaches and their applications.
- Summarize the ecological importance of marine bacteria and their metabolite production using primary literature.
- Demonstrate proper use of wet lab tools, including the use of a pin replicator and proper layout of a 96-well plate.
- Grow marine bacteria on defined media, and measure bacterial growth using a 96-well spectrophotometer.
- Process collected bacterial growth data using an R script and plot trends of bacterial growth on different metabolites.
- Use statistical analysis to evaluate bacterial growth data to assess experimental accuracy and reproducibility.

Lesson #2: Data analysis through independent research projects

- Develop an independent research question related to the observed growth phenotype of a sequenced isolate.
- Propose and justify hypotheses that link phenotypic patterns to genomic features.
- Analyze bacteria genomes using tools such as NCBI BLAST and functional annotation platforms.
- Characterize potential metabolic pathways by identifying relevant genes and their functions.
- Interpret genomic and phenotypic data to construct a cohesive scientific explanation.
- Present findings effectively through oral presentations and written reports, emphasizing background, methods, results, and conclusions.

Introduction

Functional genomics and genetic methodologies are rapidly advancing, providing profound insights into various biological processes, from elucidating metabolic disorders to identifying novel drug targets and enhancing organismal traits (Haley & Roudnicky, 2020; Mascher et al., 2024; Przybyla & Gilbert, 2022; Qu et al., 2024). Despite the pivotal role of genomics in science and the vast scale of publicly available genomic data undergraduate biology curricula often lack opportunities for students to engage directly with genomic data, limiting their appreciation of its practical applications and contributing to limited opportunities to develop genomics literacy, even among professionals in genomics-dependent fields (Whitley et al., 2020).

Course-based Undergraduate Research Experiences based on bioinformatics projects offer an effective framework to immerse a large number of students in authentic research and improve understanding of fundamental concepts at the intersection of genetics and functional

genomics (Banta et al., 2012; Jurgensen et al., 2021; Reeves et al., 2018; Shaffer et al., 2014). Likewise, CUREs have been shown to enhance self-confidence in scientific thinking and the development of scientific process skills, increase inclusivity in science for underrepresented populations, and improve persistence in science (Alvarez-Berrios & Haynes, 2023; Bradshaw et al., 2023; Hanauer & Dolan, 2014).

Building on this framework, investigating microbial metabolism provides a compelling context for genomics CUREs, linking gene function to measurable phenotypic traits. This enables students to investigate how genetic function translates into ecological processes. Microbial metabolic traits can be regarded as biochemical phenotypes encoded by genes, frequently organized and regulated within operons. The environmental distribution of these metabolic genes in prokaryotic communities can significantly impact ecosystem level processes (Boon et al., 2014; Imhoff, 2016; Schreier et al., 2023). Genome based approaches, such as single-cell sequencing and metagenomics, enable the exploration of microbial ecology on single cells. However, our functional understanding of genomes predominantly stems from studies on cultivated bacteria, where laboratory experiments and genome sequencing elucidate the functions of previously uncharacterized genetic features (Goldfarb et al., 2025; Ruiz-Perez et al., 2021). Predicting gene function relies on sequence similarity searches against curated databases, which are inherently constrained by prior studies employing genetic and experimental approaches in model organisms. Even in extensively studied organisms, numerous genes remain poorly characterized or unannotated (Moore et al., 2024).

This CURE aims to cultivate in students a foundational understanding of the predictive capabilities and limitations of genomic data in determining gene functions and metabolic properties of organisms. Students explore a diverse collection of marine bacteria to investigate their metabolic phenotypes, focusing on their ability to utilize ecologically relevant metabolites produced by primary oceanic producers (phytoplankton). Through hands-on experience, students collect and analyze bacterial growth data to characterize metabolic phenotypes. In addition, students gain hands-on experience with molecular biology techniques, such as DNA purification, and critically engage with scientific literature to explore complementary methods beyond genomics for understanding gene functions. Central to the course is a multi-week independent research project where the students design and carry out their own investigation and apply core concepts and tools to analyze the genomes of the studied bacteria. They investigate principles of gene organization and sequence homology to identify potential gene functions and metabolic pathways that may explain the observed bacterial growth phenotypes on marine metabolites. The CURE culminates with student presentations, providing a platform to share research findings and demonstrate their grasp of the tools and concepts introduced throughout the course.

Intended Audience

This CURE can be implemented in several biology courses with different student populations. We designed the CURE for a Genetics course directed primarily at second-year students in a four-year undergraduate institution. The majority of students are Biology, Molecular and Cell Biology, Chemistry or Biochemistry majors and some non-majors. To enroll in the course, students must have taken one first-year biology semester course with laboratory and one additional semester of a biology or chemistry course with laboratory. The CURE is also suitable for introductory biology and microbiology courses that aim to introduce students to genome concepts and microbial diversity. The genomic analysis component of the CURE can be adapted to a second- or third-year bioinformatics course.

Term and Context Description

We implemented this CURE project in the laboratory section of a second-year genetics course taught in a 15-week semester. Instruction consisted of three 50 min weekly lectures and one weekly 3 hour and 50 minute laboratory period where students conducted experiments and analyzed data. The lab project was completed in 11 weekly lab periods. For the CURE, students had access to a plate reader, micropipettes, an incubator, and microcentrifuges, in the biology teaching lab. Students had access to a shared computer lab or to their personal computers for data analysis.

Prerequisite Student Knowledge

This CURE project is designed for students previously enrolled in an introductory biology course covering foundational principles of the chemistry of life, genetics, and cell and molecular biology concepts. Students should have at least one semester of biology lab experience in these areas and basic data analysis skills. One additional introductory biology course with lab or an introductory general chemistry course with lab is recommended.

Prerequisite Teacher Knowledge

The course instructor is familiar with the content of a first-year biology course focused on the cell and molecular scales and a second-year, comprehensive genetics course. Familiarity with data analysis, programming and DNA sequence analysis is required to guide students through the analysis of high-throughput screen data and process DNA sequencing data to assemble and annotate genomes. Knowledge of bacterial genome organization, regulation and metabolism is critical to guide the students through their independent inquiry project.

Scientific Teaching Themes

Active Learning:

These lessons use multiple approaches to promote learning. In preparation for CURE labs, students read primary scientific literature relevant to microbial genetics, high-throughput screen approaches, and genomics of marine microorganisms. Students are required to respond to a small set of questions based on these articles before each lab period. At the start of each lab, they discuss their answers in small groups (breakout sessions) to promote learning from peers and to help build a sense of collaboration during the CURE (Cooper et al., 2021). The lab instructor rotates through the small groups to answer questions they had about the readings and then enables a brief discussion about the week's activities where all students can participate. The CURE also promotes a strong sense of student ownership of their work and collaboration (Cavanagh et al., 2018). From start to end, students engage in hands-on laboratory research that takes them from screening diverse bacteria on a suite of marine metabolites, to sharing and analyzing their datasets collectively to decide which bacteria isolate genomes to sequence. In the last stages of the project, students interact and discuss with each other as they analyze their selected genomes and prepare a lab report and a research synthesis presented to the class.

Assessment

We implemented several steps to assess student learning outcomes in the CURE based on the recommendations by (Shortlidge & Brownell, 2016). The main assessments of the CURE included two lab reports, a project proposal and a presentation of research findings. In order to scaffold learning and provide intermediate assessment points, in addition to the main assessments, we designed weekly lab activities that address the CURE learning outcomes and meet the research objectives of the project.

In the design of the weekly lab activities, we considered three features which may contribute to the efficacy and impact of CUREs: collaboration, discovery and relevance, and iteration (Corwin et al., 2015). In preparation for CURE labs, students read primary scientific literature relevant to microbial genetics, high-throughput screen approaches, and genomics of marine microorganisms. Students were required to respond to a small set of questions based on these articles before each lab period. At the start of each lab, they discussed their answers in small groups (breakout sessions) to promote learning from peers and to help build a sense of collaboration during the CURE. To track their research progress and receive feedback throughout the project, students prepared weekly mini-lab reports. These reports included a brief overview of the week's research objectives, a detailed materials and methods section, and a summary of their results or expected outcomes for the following week, supplemented by tables and figures as needed.

Students compiled their weekly reports into two lab reports that required them to communicate their research in a scientific format (Bakshi et al., 2019). The two lab reports assessed their understanding of the basic structure of formal scientific papers, their ability to present lab results clearly and orderly, and to interpret the biological significance of their data. In addition to lab reports, students were required to write a project proposal for their independent study part of the project, evaluating their ability to formulate a scientific question, propose a testable hypothesis and clearly outline methods. At the end of the semester, students delivered a research presentation to demonstrate their ability to synthesize research findings and effectively communicate their scientific work to their peers.

Together, these activities helped assess the ability of students to synthesize a complex biological analysis based on genome data, to make inferences about the properties of microorganisms based on this analysis and their research of relevant scientific literature, and to communicate results describing genetic and metabolic information to their peers. The CURE assessments including lab reports, project proposal, presentation, mini-reports and class participation accounted for 40% of the student's semester course grade, similar to the assignments and learning assessments implemented in the lecture component of the course. This gave students credit for their investment in the research component of the course and highlighted the value of learning objectives and skills we aimed to develop throughout the CURE.

Inclusive Teaching

This course incorporates diverse instructional methods, including think-pair-share (where students discuss key terms in pairs), hands-on activities (such as lab work), open-ended inquiry (through an independent project), inquiry-based learning (by reading and discussing scientific papers), and invited speaker sessions (scientists working on the same bacteria isolates as the students) (Tanner, 2013). Additionally, multiple assessment modes such as group discussions, research article and lab protocol-based questions, weekly lab reports or hand in questions, a final individual project and presentation to ensure that students have varied opportunities to learn and demonstrate their understanding.

Instructors serve as guides and research mentors, promoting an environment where students make independent decisions about their experiments, such as selecting which bacteria to sequence and analyze. Peer discussions are emphasized to mirror the collaborative nature of scientific research and the discourse found in primary journal articles.

Formative feedback is integrated throughout the course to support student learning and skill development. Weekly lab sessions provide structured opportunities for immediate feedback from instructors and peers on experimental techniques, data analysis, and research interpretations. Small group discussions allow students to reflect on assigned research articles, refine their understanding, and incorporate peer feedback before submitting responses. Mini-reports or post-lab questions help students track their progress while receiving regular instructor comments. The mid-semester research report offers comprehensive feedback on scientific knowledge, synthesis and data presentation, guiding improvements before the final report and presentation. Additionally, the research proposal assignment provides early-stage feedback on research question and hypothesis formulation, literature review synthesis, and genome analysis methods, ensuring students are well-prepared for their independent inquiry.

This course makes mentored research accessible to all enrolled students by embedding research-based learning into a structured curriculum (Bangera & Brownell, 2014). Its design allows students with varying levels of experience and diverse academic backgrounds to engage in authentic research, from formulating hypotheses to analyzing gene pathways. Small-group collaborations and instructor mentorship provide consistent support, helping students build confidence in conducting independent investigations.

Course Schedule

We have developed a 11-session lab course with all necessary supporting materials. This comprehensive resource enables instructors to efficiently prepare for classes and teach essential genetics concepts needed to effectively implement the course. Originally part of a genetics class, this course can also be adapted for standalone experiments integrated into existing courses or utilized as an independent unit.

This course-based research project introduces undergraduates to foundational genetic concepts, data analysis techniques, and genome sequencing methodologies, with a focus on the biology of marine microorganisms. Throughout the course, students develop a strong understanding of the gene functions and genetic mechanisms that enable marine bacteria to consume and metabolize phytoplankton-derived compounds by studying novel isolates of heterotrophic marine bacteria.

Phytoplankton in the oceans contribute to roughly half of global net primary production (Behrenfeld et al., 2005). Within minutes to a few days, heterotrophic bacteria assimilate approximately half of the phytoplankton metabolites dissolved in the water column (Moran, Ferrer-González, et al., 2022). Given our limited understanding of these processes, particularly in the context of current and future ocean conditions, identifying the interactions that drive them is crucial for understanding global carbon and nutrient fluxes.

Through this CURE, we are investigating the gene mechanisms that heterotrophic microbes use to support their growth by catabolizing phytoplankton metabolites. Our bacteria isolates were collected from the Equatorial Pacific. Students were given 94 novel bacteria isolates and learned high-throughput techniques on how to culture them on different

phytoplankton associated metabolites. Based on the growth profiles of the isolates, students selected which genomes to sequence. They then extracted the DNA, which was sequenced to generate draft genomes. These genomes were used to study the catabolic mechanisms involved in metabolizing specific compounds, allowing students to make phenotype-to-genotype connections for each isolate. This course design immerses students in scientific exploration with novel marine microorganisms, empowering them to make discoveries that are new not only to themselves but also to the instructors and the broader scientific community.

Exemplar Lesson Plan 1

High throughput growth assay of isolate bacteria on phytoplankton derived metabolites

Before participating in this lab, students should have completed Lab 1, where they explored the diversity of marine bacteria strains and established cultures of each microorganism as part of a library collection. As part of Lab 1, students submitted their first mini-report, documenting their methods, materials, and initial results. This report provides a foundation for understanding the experimental design and prepares students for the high-throughput growth assay conducted in this lab.

The goal of this exemplar lesson is to test the growth of 94 bacteria isolates on a wide range of environmentally relevant metabolites. Within marine bacteria communities, many bacteria groups have overlapping metabolic capabilities, however some species specialize in consuming distinct chemical classes of organic compounds to support their growth. In particular, many marine bacteria can grow on select metabolites produced by marine phytoplankton, creating complex interactions between these trophic groups that influence large scale biogeochemical cycles in the ocean (Boysen et al., 2022; Moran, Kujawinski, et al., 2022; Seymour et al., 2017). In this lesson, we test the types of phytoplankton derived metabolites that support growth of bacteria isolates as a sole carbon and nitrogen source. Isolates are from a bacteria library collected in the Equatorial Pacific and were screened using a high-throughput assay.

Using high-throughput techniques (Karkman et al., 2016; Schroer et al., 2023), students will use the bacteria library that they transferred from single isolate cultures grown on ½ YTSS agar plates to liquid ½ YTSS media during Lab 1. Each student will test the growth of the entire bacteria library on a single metabolite in three replicates. At the end of the growth experiments, the results of the entire class will be pooled and analyzed together to determine the growth of the entire bacteria library on all metabolites of interest.

Every student is given a 96 well plate containing a copy of the bacteria library, where each well contains a different isolate (94 bacteria stains and one control well devoid of bacteria). Students are also given three 96-well plates containing seawater media (Smith & Ferrer-Gonzalez, 2017) supplemented with a single marine metabolite that could support bacteria growth. Every student also prepares a 96-well plate containing only seawater media without marine metabolites as a control plate to measure any background growth. A disposable replicator tool with 96 sterile pins spaced out to match the 96-well plate wells is used to copy the bacteria library into each test plate containing metabolite supplemented media. After transferring the bacteria library to the metabolite plates, the plates are incubated inside an incubator at 22°C for a week. The instructor demonstrate the copying of the bacteria library to metabolite plates and reminded students of using sterile techniques during the procedure.

Bacterial growth in the plates is tracked using a plate reader spectrophotometer by measuring absorbance at 600 nm wavelength. Measurements are taken at the start of the experiment right after transferring the bacteria library to new plates, after 24 h of growth, after 48 h of growth, and finally after one week of growth at the start of the next lab session. Students are trained how to use the plate reader during this lab and are able to access and use it independently afterwards. To save and share results from these growth experiments, each student uploads a file containing growth data over the course of the experiment to a shared google drive for the entire class. During the subsequent lesson (Lab 4), students generate tables of bacterial growth data using R. This way students have access to data from the entire class to make exploratory plots looking for phenotypic growth differences among the bacteria isolates. An example of student-made growth curves of all isolates growing on an example metabolites ornithine, sarcosine, and trehalose compared to control growth in seawater only is shown Fig. 4.1 and Sup 4.1.

Exemplar Lesson Plan 2

Connecting phenotype to genotype through bacteria genome analysis

The goal of this lesson is to connect the observed growth phenotype of isolates tested in previous lessons to their metabolic capabilities through genome analysis (Fig. 4.2). Based on the class growth assay results, students selected several representative bacteria isolates for genome sequencing during Lab 5. In a previous lab session (Lab 2), students extracted DNA from all isolates, and the instructor sent the selected set of isolates (Lab 5) to SeqCenter (Pittsburgh, PA, USA) for paired-end Illumina sequencing. As part of mini report 2, students documented their DNA extraction process, setting the base for linking growth phenotypes to metabolic capabilities through genome analysis. The reads were processed and assembled by the instructor using TrimGalore and Unicycler, and the resulting genome assemblies were annotated with Prokka. Although genome assembly and annotation were beyond the course's scope, the instructor provided students with draft assemblies and genome annotation tables (FASTA files with nucleotide and protein sequences and GFF3 tables) to support their independent analyses (BioProject Accession PRJNA1304954).

This lesson is structured as an independent research project that begins with a short research proposal assignment that prompts students to identify a unique research question that emerges from the growth screen data. To ask their question, each student selects one or two marine metabolites of interest and a bacterium that consumed these to develop their research question. The following are examples of research questions students identified.

Student example research question 1:

> "What specific metabolic pathway and mechanism allows the marine bacteria isolate of unknown taxonomy, SET2-22, to metabolize the free amino acid ornithine as a carbon and nitrogen source?"

Student example research question 2:

> "Can we determine which metabolic pathway is responsible for bacteria isolate SET2-06 growing successfully in the presence of sarcosine?"

The assignment also prompts students to put forward a hypothesis that could explain their observations. To address this, students conduct a literature search to investigate what types of genes, proteins, and biochemical pathways organisms use to transport, transform, and metabolize these substrates. In the research proposal, students provide an individual summary of each research article they found that was relevant to their question. The assignment asks students to summarize at least three papers and to synthesize their findings in one short paragraph. The following are examples of hypotheses described by students in the course.

Student example hypothesis 1:

> "From previous understanding of the major AST pathway for ornithine utilization in *Pseudomonas aeruginosa*, I predict isolate S2-22 is able to catabolize ornithine in a similar pathway as *Pseudomonas aeruginosa.* I predict the catabolic pathway in S2-22 will involve a cluster of several co-local genes to specify a regulatory protein and corresponding enzymes, similar to the AST pathway that is observed in other gram-negative proteobacteria."

Student example hypothesis 2:

> "I hypothesize that the S2_06 marine isolate contains genes coding for all subunits of the tetrameric sarcosine oxidase protein to allow for successful growth in sarcosine since after evaluating multiple BLAST comparisons of our sequenced genome, multiple literature genes with sarcosine subunits had strong relation to our sequenced genome."

The research proposal also asks students to describe the approach they will follow to analyze their bacteria genome. To facilitate this, in Lab 7, the instructor reviews genome concepts and online tools students can use in their analysis. Two concepts covered in the Genetics course associated with the CURE are discussed with students. The first is bacterial gene regulation through operons. This prompts students to predict that genes required for the consumption of a substrate will be co-located in the genome. The second concept the instructor explores with the students is sequence homology aided by BLAST online tools and sequence databases. The instructor first uses articles selected by students through their literature search to point out that the genes and proteins described in published studies are searchable and accessible through resources like the National Center for Biotechnology and Information (NCBI) nucleotide and protein databases. The instructor describes the processes of constructing a small database of reference sequences from these studies. The instructor shows students how to query the NCBI and Uniprot databases with BLAST, how to retrieve sequences by their accessions, and how to compile the sequences in FASTA format. The instructor also guides students through the process of aligning the reference sequences they identified to the gene and proteins in the genomes of their bacterium using BLAST. Students learn how to interpret BLAST parameters and search outputs to assess sequence homology. Finally, to help students connect genes and protein functions to metabolic pathways and networks, the instructor introduces students to the Kyoto Encyclopedia of Genes and Genomes (KEGG) website and database. Students learn how to query KEGG to visualize metabolic pathways and modules associated with the genes and biochemical transformation they discovered in their literature search. This guided exercise gives students the conceptual framework and tools they need to complete their research proposal (Fig. 4.2).

In the final stage of the independent project, students conduct their proposed genome analysis to test their hypothesis (weeks 8 and 9) and synthesize their findings (week 10) into a comprehensive Final Report. During the final week of the course (week 11), students share their findings through a short presentation to the class, showcasing their research process and results (Fig. 4.3).

The final report (Lab report 2) is structured like a formal scientific research paper. Students compose a brief introduction to the project, a methods section describing their bioinformatics approaches, present their results with formal figures and tables (Fig. 4.3), and interpret what these results mean at a biological level in a short discussion. Students are expected to include a list of references and select a proper in-text citation style. Importantly, the Final Report guidelines prompt students to evaluate their hypothesis and results critically and to contextualize their findings by comparing them to published research, including the studies they found in their literature search. Students should be able to answer the following questions: What evidence did you identify in the bacteria genomes that is consistent with protein functions and metabolic pathways used to consume the selected marine metabolite? What results were not consistent with your hypothesis? What may be some reasons that explain why you did not find genomic evidence in favor of your hypothesis?

During the final week of the course (week 11), students deliver a 10-minute presentation, followed by a 5-minute Q&A session, to share their research findings, refine their science communication skills, and demonstrate their understanding of functional and metabolic genome analyses. Presentations must include information about their isolates such as taxonomy, genes for the predicted catabolism of a metabolite, and the marine metabolite studied, as well as a synthesis of the literature that supports their hypothesis about the mechanism enabling bacterial growth on the metabolite. Students will evaluate their genome analysis results, discussing whether the identified gene functions support their hypothesis and suggesting additional experiments or data to strengthen their conclusions. For extra credit, students can reflect on the broader scientific or societal impacts of understanding microbial metabolism. The slides are uploaded to Canvas, and presentations are graded on content (5 pts), hypothesis (10 pts), results discussion (15 pts), and broader impacts (10 pts).

Teaching Discussion

Technical Implementation, Challenges and Possible Modifications

Starting and maintaining cultures of every bacteria isolate in the collection can be a challenging aspect of the first stage of the CURE (Lesson #1). Preparing all the necessary sterile culturing supplies and growth media before the start of the semester facilitated this process. In the first week of the lab, we assigned students a few isolates each that they were responsible for growing. This made the preparation of the culture collection more manageable. A few students volunteered to transfer a small sample of each grown isolate into 96-well plates. At this point, we referred to the collection as a bacteria library and we could clone it easily into other 96-well plates for growth screens. This component of the project can be adapted to any collection of bacteria strains and to a variety of research problems suitable for 96-well plate screening. For example, the class can screen a collection of mutant strains or bacteria agar isolates obtained from a different environment based on the project theme. Similarly, students can instead prepare growth test media to screen bacteria on a different suite of organic compounds. For this project

we screened marine bacteria for growth on common marine metabolites known to support microbial metabolism in the ocean.

The growth screens can be done in any plate reader. Students completed a short training session with laboratory support staff before using the plate reader to ensure they learned how to load plates and configure the instrument for end-point absorbance measurements. The training gave students confidence to read their plates several times that week without supervision. This was necessary to assess growth over time. Depending on the accessibility of the plate reader and research goals, the instructor can assess the data collection plan that students can implement.

We wrote a custom R script shared with students to analyze the 96-well plate-formatted raw data text files downloaded from a SpectraMax M2 Microplate Reader software. The data-processing section of the R script should be adapted to the plate reader data output format available. Students used the R script to obtain summary statistics of the growth (absorbance) data, test for significant differences with a series of t-tests and print graphs to visualize the data. We used R software in this lab because R is the language used primarily in our undergraduate biology courses. The plate reader data processing script can be adapted to Python or to another programming language.

Students extracted high quality DNA from most bacteria isolates successfully using a standard bacterial DNA extraction kit (Zymo Research Quick-DNA Fungal/Bacterial Miniprep Kit). A few isolates yielded low DNA quality and concentration. Nevertheless, this could be sufficient DNA for 16S rRNA gene PCR amplification if the CURE goal is to determine bacterial taxonomic classification. For our purposes, students prioritized isolates with high quality DNA samples with sufficient yield for Illumina paired-end whole-genome sequencing (at least 30 µL of 20 ng/µL measured by Nanodrop). The instructor used a local Linux computer to trim and assemble Illumina reads with Trim Galore for the students (Krueger, 2015), DOI:10.5281/zenodo.5127898; DOI:10.14806/ej.17.1.200) and the Unicycler SPAdes-optimiser (Wick et al., 2017), respectively. If this is not an option, most sequencing vendors can provide a draft genome assembly of the data at an affordable price or the instructor and students can perform genome assemblies on a cloud instance. The instructor annotated assembly contigs with Prokka (Seemann, 2014) and shared with students the resulting GFF3 and fasta files to explore the predicted gene functions. Alternatives to Prokka to annotate and explore bacteria genome sequences include the MicrobeAnnotater (Ruiz-Perez et al., 2021) and Proksee (Grant et al., 2023). The former is a comprehensive genome annotation tool that runs on Python and the latter enables users to upload, assemble, annotate, and visualize sequencing data via a web server. After genome assembly and annotation, the tools and databases students required to analyze the sequences and functions found in their bacteria genomes were readily accessible online (BLAST, KEGG, Uniprot). Processing of raw sequencing reads was done by the instructor and was out of the scope of our class for students (students were given assembled and annotated genomes), however, this process could be adapted as a lesson if the goal of the class were to teach genome processing tools and pipelines.

Observations of Student Learning and Responses to the Lesson Content and Activities

In their independent project literature review (Lesson #2), students encountered a variety of model organisms and approaches used to investigate gene functions. For many students, however, determining if an article contained useful information to test their hypothesis took time. Instructor feedback and peer-to-peer discussions helped students recognize elements that made an article useful. The most useful articles contained diagrams of biochemical pathways or gene

operons that illustrated the parts and connectivity of a metabolic process (Perchat et al., 2018). Articles that readily identified the genome and protein accessions, gene names and descriptive gene functions, were particularly useful for students to curate a set of reference sequences for their project. Some students found it challenging to grasp how to query their bacteria genome using their reference sequences and BLAST. The most taxing task for students was determining if their BLAST search, as constructed, yielded significant results. Often this was because students were not sure if their reference sequences were appropriate, in which case revisiting the relevant article with the instructor was helpful. In other cases, this was because significant BLAST hits corresponded to genes in different contigs from the draft assemblies or with no evidence of co-location as was expected of gene operons that mediated a catabolic pathway. Allowing students to engage autonomously (yet with access to the instructor for feedback) with this analysis, iterating the process between week 7 and week 10, was critical to overcome this learning gap.

By week 9, most students were fully engaged with their genome analysis. They asked questions about how to evaluate their results, compared their results with those of their peers, found more articles with additional genetic or biochemical information relevant to their metabolite, expanded their set of reference gene and protein sequences to refine their analysis, and began to recognize the limitations of their genome-centered approach to answer their question. This stage of the CURE offered several opportunities to discuss with students additional or complementary approaches to genomics (e.g. mutagenesis, gene knockouts, heterologous expression, enzymology) to investigate gene and protein functions. Many of these approaches were the subject of the articles they read. In week 10, many students demonstrated mastery of the assignment and felt confident to use tools like KEGG to synthesize and visualize their genome analysis results as metabolic maps or pathways. This outcome demonstrated advanced learning and the efficacy of the course. Together, the project activities offered students a view of the power and limitations of genome-based approaches and provided basic training and understanding of bacterial genetics and functional genomics concepts applicable to many areas of biological research.

This CURE exposed students to a real-world research project and allowed them to generate new data, work with a novel collection of bacteria isolates, collaborate with their peers, and draw conclusions based on their findings. Towards the end of the course, we deployed a survey (Hanauer & Dolan, 2014) designed to assess how well the CURE promoted student ownership of their work, a factor associated with student retention in the sciences, and their appreciation of its impact. Survey questions focused on cognitive ownership measure on a Likert scale of 1 to 5 (Corwin et al., 2018), where the average score across all cognitive ownership questions was 4.0 out of 5, with a standard deviation of 0.59 (Table 4.1). Previous reports of cognitive ownership scores across other CUREs have an average of 3.93 on a 1 to 5 scale (Zajic et al., 2025), which compares well with the scores of this course.

Figures and Tables
Figure 4.1

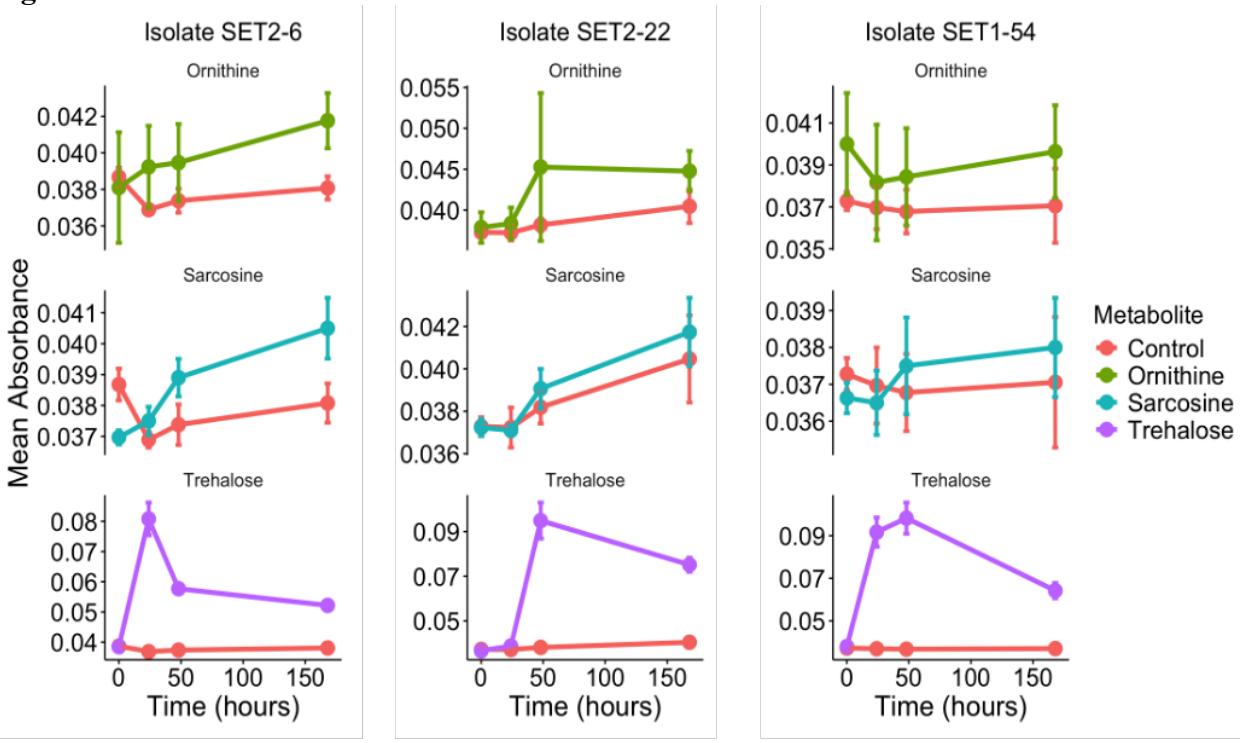


Fig. 4.1. Growth of select isolates to different metabolites over time. An example of student-made growth curves, mean absorbance at 600 nm (OD_{600}) is shown for select isolates grown with added Ornithine, Sarcosine, or Trehalose as the sole carbon source, compared to control growth in seawater over a 168-hour incubation (one week of growth). Error bars represent standard error of the mean (n=3).

Figure 4.2

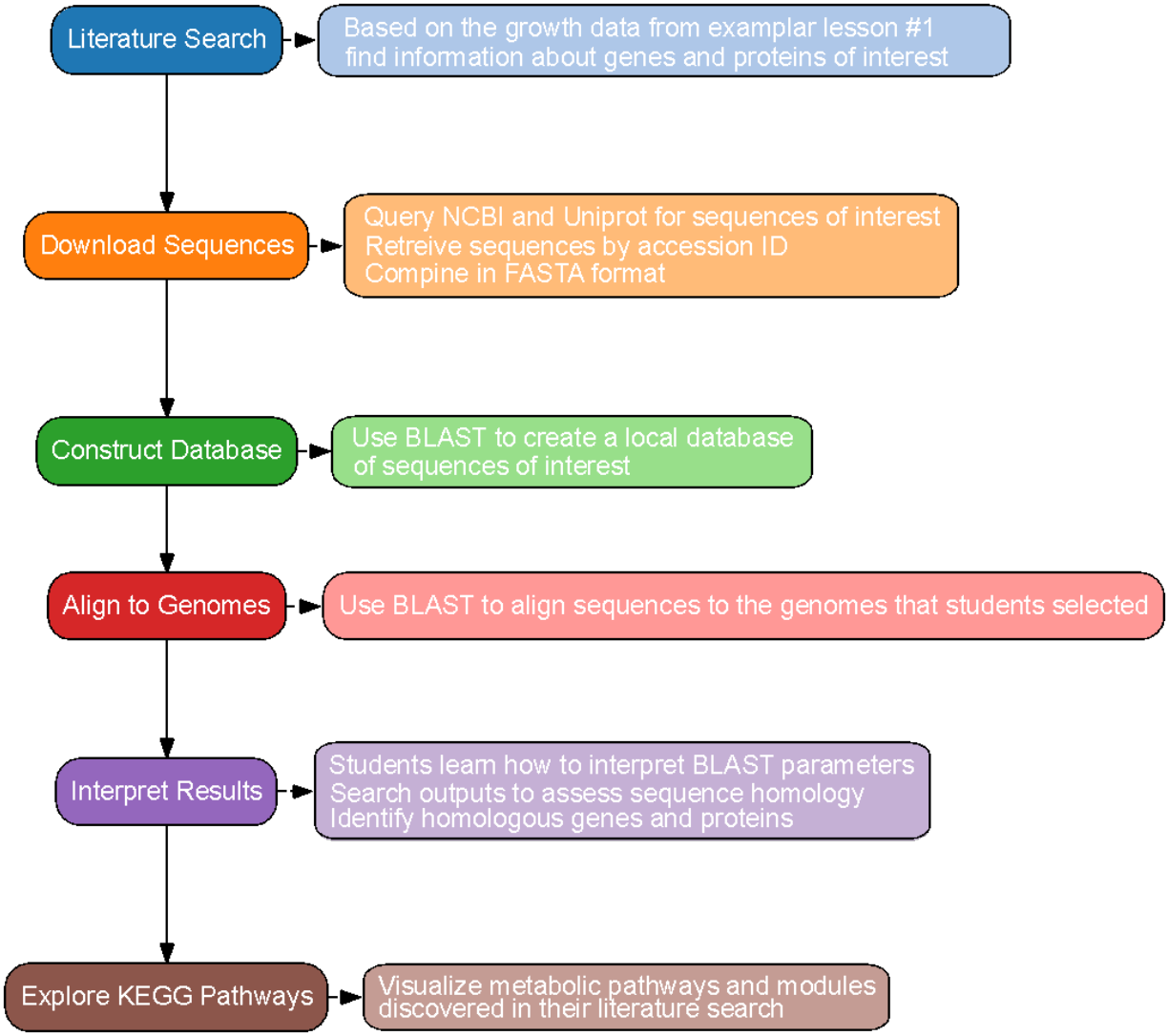


Fig. 4.2. Workflow for student led metabolite catabolic pathway discovery and annotation. Students follow a step-by-step process beginning with literature search (blue box), downloading the gene sequences (orange box), and building a custom BLAST database (green box). Sequences are aligned to genomes of interest (red box), and results are interpreted to identify homologous sequences (purple box). KEGG is used to explore associated metabolite catabolic pathways (brown box).

Figure 4.3

A

Table 1. BLAST results of ornithine degradation pathway in S2-22 compared to reference proteins.

Reference Encoding Gene	Reference Protein ID	S2-22 Protein	S2-22 Protein Function	Query cover (%)	Percent Identity (%)	E-value
AruB	AAC46013	JCFOCAEG_00665	N-succinylarginine dihydrolase	99	63.45	0.0
AruC	AAC46009	JCFOCAEG_03287	Succinylornithine transaminase/acetylornithine aminotransferase	98	64.91	0.0
AruD	ACC46012	JCFOCAEG_03289	N-succinylglutamate 5-semialdehyde dehydrogenase	100	58.9	0.0
AruE	ACC46014	JCFOCAEG_00833	hypothetical protein	97	39.09	7×10^{-5}
AruF	AAC46011	JCFOCAEG_03288	Arginine N-succinyltransferase subunit beta	100	40.18	5×10^{81}
AruG	AAC46011	JCFOCAEG_03288	Arginine N-succinyltransferase subunit beta	98	46.36	4×10^{99}

B

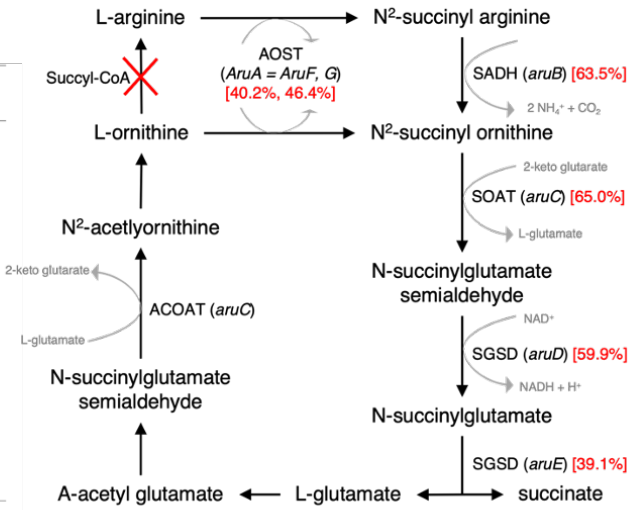


Fig. 4.3. Example of student analysis of metabolite catabolic pathways. This student analysis focuses on isolate S2-22 showing growth on ornithine and analyzes the metabolic pathway of ornithine degradation in S2-22 **(A)** Table showing BLAST results identifying the match between reference proteins and bacterial isolate S2-22 for optimal local alignment and maximum gene match between protein sequences. **(B)** Schematic showing the arginine and ornithine utilization (*aru)* pathway in *Pseudomonas aeruginosa* (adopted from Itoh 1997). Percent identity matches between *P. aeruginosa* reference bacteria and S2-22 are shown in red brackets, and the red X denotes L-ornithine and L-arginine alternative route not utilized by isolate S2-22.

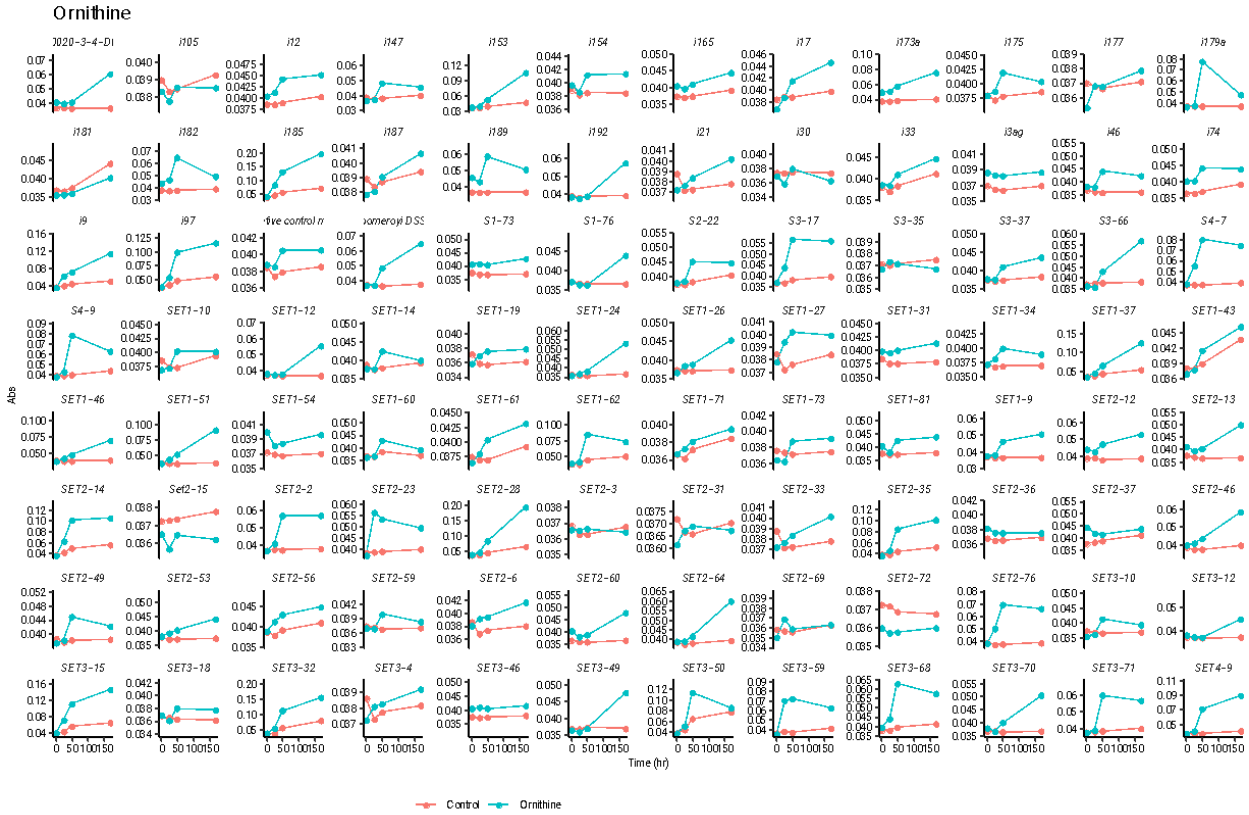
Table 4.1

Question	Average score	STD
My research will help to solve a problem in the world	3.8	0.85
My findings were important to the scientific community	4	0.74
I faced challenges that I managed to overcome in completing my research project	4.7	0.58
I was responsible for the outcomes of my research	4.7	0.76
The findings of my research project gave me a sense of personal achievement	4.4	0.76
I had a personal reason for choosing the research project I worked on	2.8	1.25
The research question I worked on was important to me	3.7	1.15
In conducting my research project, I actively sought advice and assistance	4.4	0.76
Average score across all cognitive ownership questions	4.0	0.59

Table 4.1. Student survey results on cognitive ownership of course material. The student survey was administered at the end of the course and included eight questions assessing cognitive ownership. Students rated each question on a 5-point Likert scale: Strongly agree (5), Agree (4), Neither agree nor disagree (3), Disagree (2) and Strongly disagree (1). The average score across all cognitive ownership question is shown at the bottom of the table.

Supplemental Figures

Sup. 4.1



Sup 4.1. Examples of students growth of isolates in Ornithine. An example of student-made growth curves of 94 isolates, mean absorbance at 600 nm (OD_{600}) is shown for Ornithine as the sole carbon source, compared to control growth in seawater over a 168-hour incubation (one week of growth).

References

Alvarez-Berrios, M. P., & Haynes, G. (2023). Puerto Rican Students Rising in STEM: Findings from a Multicampus Collaborative CURE Program to Promote Student Success. *CBE—Life Sciences Education*, *22*(4). https://doi.org/10.1187/cbe.23-05-0083

Bakshi, A., Webber, A. T., Patrick, L. E., Wischusen, W., & Thrash, C. (2019). The CURE for Cultivating Fastidious Microbes. *Journal of Microbiology & Biology Education*, *20*(1). https://doi.org/10.1128/jmbe.v20i1.1635

Bangera, G., & Brownell, S. E. (2014). Course-Based Undergraduate Research Experiences Can Make Scientific Research More Inclusive. *CBE—Life Sciences Education*, *13*(4), 602–606. https://doi.org/10.1187/cbe.14-06-0099

Banta, L. M., Crespi, E. J., Nehm, R. H., Schwarz, J. A., Singer, S., Manduca, C. A., Bush, E. C., Collins, E., Constance, C. M., Dean, D., Esteban, D., Fox, S., McDaris, J., Paul, C. A., Quinan, G., Raley-Susman, K. M., Smith, M. L., Wallace, C. S., Withers, G. S., & Caporale, L. (2012). Integrating Genomics Research throughout the Undergraduate Curriculum: A Collection of Inquiry-Based Genomics Lab Modules. *CBE—Life Sciences Education*, *11*(3), 203–208. https://doi.org/10.1187/cbe.11-12-0105

Behrenfeld, M. J., Boss, E., Siegel, D. A., & Shea, D. M. (2005). Carbon-based ocean productivity and phytoplankton physiology from space. *Global Biogeochemical Cycles*, *19*(1). https://doi.org/10.1029/2004gb002299

Boon, E., Meehan, C. J., Whidden, C., Wong, D. H.-J., Langille, M. G. I., & Beiko, R. G. (2014). Interactions in the microbiome: Communities of organisms and communities of genes. *FEMS Microbiology Reviews*, *38*(1), 90–118. https://doi.org/10.1111/1574-6976.12035

Boysen, A. K., Durham, B. P., Kumler, W., Key, R. S., Heal, K. R., Carlson, L. T., Groussman, R. D., Armbrust, E. V., & Ingalls, A. E. (2022). Glycine betaine uptake and metabolism in marine microbial communities. *Environmental Microbiology*, *24*(5), 2380–2403. https://doi.org/10.1111/1462-2920.16020

Bradshaw, L., Vernon, J., Schmidt, T., James, T., Zhang, J., Archbold, H., Cadigan, K., Wolfe, J. P., & Goldberg, D. (2023). Influence of CUREs on STEM retention depends on demographic identities. *Journal of Microbiology & Biology Education*, *24*(3). https://doi.org/10.1128/jmbe.00225-22

Cavanagh, A. J., Chen, X., Bathgate, M., Frederick, J., Hanauer, D. I., & Graham, M. J. (2018). Trust, Growth Mindset, and Student Commitment to Active Learning in a College Science Course. *CBE—Life Sciences Education*, *17*(1), ar10. https://doi.org/10.1187/cbe.17-06-0107

Cooper, K. M., Schinske, J. N., & Tanner, K. D. (2021). Reconsidering the Share of a Think–Pair–Share: Emerging Limitations, Alternatives, and Opportunities for Research. *CBE—Life Sciences Education*, *20*(1), fe1. https://doi.org/10.1187/cbe.20-08-0200

Corwin, L. A., Graham, M. J., & Dolan, E. L. (2015). Modeling Course-Based Undergraduate Research Experiences: An Agenda for Future Research and Evaluation. *CBE—Life Sciences Education*, *14*(1), es1. https://doi.org/10.1187/cbe.14-10-0167

Corwin, L. A., Runyon, C. R., Ghanem, E., Sandy, M., Clark, G., Palmer, G. C., Reichler, S., Rodenbusch, S. E., & Dolan, E. L. (2018). Effects of Discovery, Iteration, and Collaboration in Laboratory Courses on Undergraduates' Research Career Intentions Fully Mediated by Student Ownership. *CBE—Life Sciences Education*, *17*(2), ar20. https://doi.org/10.1187/cbe.17-07-0141

Goldfarb, T., Kodali, V. K., Pujar, S., Brover, V., Robbertse, B., Farrell, C. M., Oh, D.-H., Astashyn, A., Ermolaeva, O., Haddad, D., Hlavina, W., Hoffman, J., Jackson, J. D., Joardar, V. S., Kristensen, D., Masterson, P., McGarvey, K. M., McVeigh, R., Mozes, E., … Murphy, T. D. (2025). NCBI RefSeq: Reference sequence standards through 25 years of curation and annotation. *Nucleic Acids Research*, *53*(D1), D243–D257. https://doi.org/10.1093/nar/gkae1038

Grant, J. R., Enns, E., Marinier, E., Mandal, A., Herman, E. K., Chen, C., Graham, M., Van Domselaar, G., & Stothard, P. (2023). Proksee: In-depth characterization and visualization of bacterial genomes. *Nucleic Acids Research*, *51*(W1), W484–W492. https://doi.org/10.1093/nar/gkad326

Haley, B., & Roudnicky, F. (2020). Functional Genomics for Cancer Drug Target Discovery. *Cancer Cell*, *38*(1), 31–43. https://doi.org/10.1016/j.ccell.2020.04.006

Hanauer, D. I., & Dolan, E. L. (2014). The Project Ownership Survey: Measuring Differences in Scientific Inquiry Experiences. *CBE—Life Sciences Education*, *13*(1), 149–158. https://doi.org/10.1187/cbe.13-06-0123

Imhoff, J. (2016). New Dimensions in Microbial Ecology—Functional Genes in Studies to Unravel the Biodiversity and Role of Functional Microbial Groups in the Environment. *Microorganisms*, *4*(2), 19. https://doi.org/10.3390/microorganisms4020019

Jurgensen, S. K., Harsh, J., & Herrick, J. B. (2021). A CURE for Salmonella: A Laboratory Course in Pathogen Microbiology and Genomics. *CourseSource*, *8*. https://doi.org/10.24918/cs.2021.24

Karkman, A., Johnson, T. A., Lyra, C., Stedtfeld, R. D., Tamminen, M., Tiedje, J. M., & Virta, M. (2016). High-throughput quantification of antibiotic resistance genes from an urban wastewater treatment plant. *FEMS Microbiology Ecology*, *92*(3), fiw014. https://doi.org/10.1093/femsec/fiw014

Krueger, F. (2015). Trim Galore!: A wrapper around Cutadapt and FastQC to consistently apply adapter and quality trimming to FastQ files, with extra functionality for RRBS data. *Babraham Institute*.

Mascher, M., Jayakodi, M., Shim, H., & Stein, N. (2024). Promises and challenges of crop translational genomics. *Nature*, *636*(8043), 585–593. https://doi.org/10.1038/s41586-024-07713-5

Moore, L. R., Caspi, R., Boyd, D., Berkmen, M., Mackie, A., Paley, S., & Karp, P. D. (2024). Revisiting the y-ome of *Escherichia coli*. *Nucleic Acids Research*, *52*(20), 12201–12207. https://doi.org/10.1093/nar/gkae857

Moran, M. A., Ferrer-González, F. X., Fu, H., Nowinski, B., Olofsson, M., Powers, M. A., Schreier, J. E., Schroer, W. F., Smith, C. B., & Uchimiya, M. (2022). The Ocean's labile DOC

supply chain. *Limnology and Oceanography*, *67*(5), 1007–1021. https://doi.org/10.1002/lno.12053

Moran, M. A., Kujawinski, E. B., Schroer, W. F., Amin, S. A., Bates, N. R., Bertrand, E. M., Braakman, R., Brown, C. T., Covert, M. W., Doney, S. C., Dyhrman, S. T., Edison, A. S., Eren, A. M., Levine, N. M., Li, L., Ross, A. C., Saito, M. A., Santoro, A. E., Segrè, D., … Vardi, A. (2022). Microbial metabolites in the marine carbon cycle. *Nature Microbiology*, *7*(4), 508–523. https://doi.org/10.1038/s41564-022-01090-3

Perchat, N., Saaidi, P.-L., Darii, E., Pellé, C., Petit, J.-L., Besnard-Gonnet, M., De Berardinis, V., Dupont, M., Gimbernat, A., Salanoubat, M., Fischer, C., & Perret, A. (2018). Elucidation of the trigonelline degradation pathway reveals previously undescribed enzymes and metabolites. *Proceedings of the National Academy of Sciences*, *115*(19). https://doi.org/10.1073/pnas.1722368115

Przybyla, L., & Gilbert, L. A. (2022). A new era in functional genomics screens. *Nature Reviews Genetics*, *23*(2), 89–103. https://doi.org/10.1038/s41576-021-00409-w

Qu, L., Huang, X., Su, X., Zhu, G., Zheng, L., Lin, J., Wang, J., & Xue, H. (2024). Potato: From functional genomics to genetic improvement. *Molecular Horticulture*, *4*(1). https://doi.org/10.1186/s43897-024-00105-3

Reeves, T. D., Warner, D. M., Ludlow, L. H., & O'Connor, C. M. (2018). Pathways over Time: Functional Genomics Research in an Introductory Laboratory Course. *CBE—Life Sciences Education*, *17*(1), ar1. https://doi.org/10.1187/cbe.17-01-0012

Ruiz-Perez, C. A., Conrad, R. E., & Konstantinidis, K. T. (2021). MicrobeAnnotator: A user-friendly, comprehensive functional annotation pipeline for microbial genomes. *BMC Bioinformatics*, *22*(1). https://doi.org/10.1186/s12859-020-03940-5

Schreier, J. E., Smith, C. B., Ioerger, T. R., & Moran, M. A. (2023). A mutant fitness assay identifies bacterial interactions in a model ocean hot spot. *Proceedings of the National Academy of Sciences*, *120*(12). https://doi.org/10.1073/pnas.2217200120

Schroer, W. F., Kepner, H. E., Uchimiya, M., Mejia, C., Rodriguez, L. T., Reisch, C. R., & Moran, M. A. (2023). Functional annotation and importance of marine bacterial transporters of plankton exometabolites. *ISME Communications*, *3*(1). https://doi.org/10.1038/s43705-023-00244-6

Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics*, *30*(14), 2068–2069. https://doi.org/10.1093/bioinformatics/btu153

Seymour, J. R., Amin, S. A., Raina, J.-B., & Stocker, R. (2017). Zooming in on the phycosphere: The ecological interface for phytoplankton–bacteria relationships. *Nature Microbiology*, *2*(7). https://doi.org/10.1038/nmicrobiol.2017.65

Shaffer, C. D., Alvarez, C. J., Bednarski, A. E., Dunbar, D., Goodman, A. L., Reinke, C., Rosenwald, A. G., Wolyniak, M. J., Bailey, C., Barnard, D., Bazinet, C., Beach, D. L., Bedard, J. E. J., Bhalla, S., Braverman, J., Burg, M., Chandrasekaran, V., Chung, H.-M., Clase, K., … Elgin, S. C. R. (2014). A Course-Based Research Experience: How Benefits Change with

Increased Investment in Instructional Time. *CBE—Life Sciences Education*, *13*(1), 111–130. https://doi.org/10.1187/cbe-13-08-0152

Shortlidge, E. E., & Brownell, S. E. (2016). How to Assess Your CURE: A Practical Guide for Instructors of Course-Based Undergraduate Research Experiences. *Journal of Microbiology & Biology Education*, *17*(3), 399–408. https://doi.org/10.1128/jmbe.v17i3.1103

Smith, C., & Ferrer-Gonzalez, F. (2017). *ASM Solution 2 v1*. ZappyLab, Inc. https://doi.org/10.17504/protocols.io.jwacpae

Tanner, K. D. (2013). Structure Matters: Twenty-One Teaching Strategies to Promote Student Engagement and Cultivate Classroom Equity. *CBE—Life Sciences Education*, *12*(3), 322–331. https://doi.org/10.1187/cbe.13-06-0115

Whitley, K. V., Tueller, J. A., & Weber, K. S. (2020). Genomics Education in the Era of Personal Genomics: Academic, Professional, and Public Considerations. *International Journal of Molecular Sciences*, *21*(3), 768. https://doi.org/10.3390/ijms21030768

Wick, R. R., Judd, L. M., Gorrie, C. L., & Holt, K. E. (2017). Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLOS Computational Biology*, *13*(6), e1005595. https://doi.org/10.1371/journal.pcbi.1005595

Zajic, C., Subramanian, K., Adulla, A., Allen, Z., Blitchington, M., Brotzman, A., Carrillo, E., Dhruv, J., Evans, T., Haider, S., Han, S., Hilton, L., Holliday, K., Keairnes, E., Kum, J., Lantz, W., Martin, J., Pierce-Tomlin, M., Plunktett, M., … Dolan, E. (2025). Could instructor talk drive CURE effectiveness? A comparative study of instructor talk in introductory lab courses. *BioRxiv*.

Acknowledgements

Thank you to my advisor, Ginger Armbrust for being the most incredible mentor I could imagine. You have given me all the freedom to explore anything I was interested in and provided a fun and safe space to learn and grow. Your way of thinking and enthusiasm are inspiring.

I would like to thank all my committee members: Anitra, Bob, Amy, Randie and Jodi. My projects would be impossible without you, thank you for teaching me new concepts and asking provoking questions.

I would like to thank all the past and current members of the Armbrust lab and CEG for your friendship and support. Shiri, Frank, Sacha, David, Stephen, Ben, Kathy, Elaina and Mike thanks for being my role models, the best office mates and always there for the smallest and biggest questions.

Thank you to all the undergraduates I've had the chance to mentor, especially Veronica, having your around was one of the most fun parts of grad school, I've learned so much from you.

Everyone in the SCOPE program, thank you for inspiring conversations and cruise experiences.

Thank you to all ocean grad students, especially my friends and cohort members Natalie, Amy, Rita, Evan, Laura C., Laura M., Erin, Jade, and Susan, it was incredible to have you around.

Thank you to all my friends, in Seattle and elsewhere, for all the adventures.

Thank you to my family and Sela, it would be completely impossible without you.