METABOLITE TRANSPORT AND ITS ROLE IN MARINE MICROBIAL INTERACTIONS

by

WILLIAM FRANCIS SCHROER

(Under the Direction of Mary Ann Moran)

ABSTRACT

Surface ocean microbial communities are responsible for half of global net primary production and form the foundation of marine food webs. The structure and function of these marine microbial communities is shaped by a network of metabolite-mediated interactions. Monitoring microbial transcriptomic or proteomic expression are effective ways to identify the metabolites that are exchanged, the community members involved, and the nature of the interactions in question. Particularly valuable genes to monitor are transporters as they are the cell's interface with its surroundings and the mechanism for obtaining extracellular metabolites. Expression-based approaches are limited, however, by the difficulty of accurately assigning functions to many genes, particularly in environmentally relevant microorganisms. In this dissertation I use experimental and computational approaches to characterize the function of genes mediating chemical exchanges in the model marine bacterium *Ruegeria pomeroyi* DSS-3, elucidating the role of metabolites in various interactions. Using an arrayed library of *R. pomeroyi* knockout mutants, I experimentally characterized 13 influx transporters of environmentally important metabolites. These included the transporters of citrate and 3-hydroxybutyrate, which were among the most highly expressed genes during a dinoflagellate bloom in Monterey Bay. The characterized ATP-binding cassette (ABC) transporters informed a

phylogeny-based approach for cognate substrate identification, finding that the ATPase component of the multi-protein transporter most accurately and consistently identifies cognate substrates. This approach was applied to identify taurine transporters in global ocean metagenomes, revealing a strong signal of higher frequency in polar environments. Finally, I characterized a key gene in the biosynthesis pathway of indole-3-acetic acid (IAA), a phytohormone and interdomain signaling molecule. *R. pomeroyi*, though it lacks a complete biosynthesis pathway, can produce IAA when provided the precursor indole-3-acetamide (IAM). A two-way IAM-IAA exchange interaction occurs between *R. pomeroyi* and a diatom when cultured together. Interactions based on the exchange of IAM are common in nature as many environmental bacterial isolates have an incomplete IAA biosynthesis pathway similar to that of *R. pomeroyi*. Together these findings highlight the importance of metabolite transport and trade in microbial communities and demonstrate the value of functional gene annotation for understanding the mechanisms and ecology of microbial exchanges.

INDEX WORDS: Bacterial interactions, Marine bacteria, Metabolites, Membrane transport, Transport proteins, ABC transporter, Indole acetic acid, Indole-3-acetamide, Auxin, Gene annotation

METABOLITE TRANSPORT AND ITS ROLE IN MARINE MICROBIAL INTERACTIONS

by

WILLIAM FRANCIS SCHROER

MS, University of South Carolina, 2017

BS, Allegheny College, 2015

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2023

METABOLITE TRANSPORT AND ITS ROLE IN MARINE MICROBIAL INTERACTIONS

by

WILLIAM FRANCIS SCHROER

Major Professor:	Mary Ann Moran
Committee:	Adrian Burd
	Arthur S. Edison
	Elizabeth L. Harvey
	Jaclyn K. Saunders

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
December 2023

DEDICATION

To my parents and to Caitlin, whose love, support, and encouragement made this possible.

ACKNOWLEDGEMENTS

I am very grateful to the many friends, family, colleagues, and collaborators, without whom this dissertation would not have been possible. The path to my PhD began well before matriculation at UGA. Thank you to my parents, you have always supported me, advocated for me, and made sure I was in environments where I could grow and learn, I would not have made it to this point without them. Thank you to my former teachers and professors for their years of hard work and support, their instruction and inspiration made it possible for me to pursue my dream of earning my doctorate.

At UGA I was immersed in a collaborative and creative environment, and for this I thank my advisor, Mary Ann Moran. Mary Ann has brought together and fostered a tightknit lab group, focused on mutual support, that has become a second family to me. She created an environment in which I had space to take risks, make mistakes, and learn from them, while always knowing I could check in for guidance and support. I also want to thank the Marine Sciences Department. Thank you to Merryl, Chuck, Adrian, Christof, and Daniella for being mentors to me. I am also very grateful for the lifelong friendships I have forged with the graduate students and postdocs of the department. I am grateful for the many amazing opportunities I have been provided. Finally, thank you to my wife, Caitlin, from helping me find my way that first day on campus to helping me in so many ways since then.

TABLE OF CONTENTS

Page

CHAPTER 1

LITERATURE REVIEW AND INTRODUCTION

Metabolite mediated interactions shape the structure and function of the marine microbial communities that control the ocean carbon cycle and represent the foundation of marine food webs. About half of Earth's global net primary production (NPP) occurs in the oceans, and about ½ of this photosynthate is rapidly mineralized to CO_2 by heterotrophic microbial communities in the surface ocean (1-5). This large flux of carbon is comprised of a complex network of metabolite exchanges between phytoplankton and the heterotrophic microbial communities that surround them (6). As primary producers, phytoplankton photosynthesize and fix carbon. Much of this fixed carbon ultimately is released into the water column through a variety of mechanisms including passive leakage, active exudation, sloppy grazing, viral lysis, and senescence (2). These exudates are the primary carbon and energy source supporting the heterotrophic microbial community. Some heterotrophic members enter into mutualistic interactions with phytoplankton, providing inorganic nutrients, vitamins, growth hormones, protection from pathogens, and detoxification services (6-14). Other heterotrophs can be antagonistic or opportunistically antagonistic to the phytoplankton (14, 15). Further, phytoplankton can excrete metabolites that inhibit harmful bacteria and support the growth of beneficial taxa, thus shaping their surrounding microbiome (11). Elucidating the mechanisms and impacts of metabolite-facilitated interactions is critical to our understanding of life and chemistry in the surface ocean and how this system may respond to changing conditions (16).

The diversity of metabolite-mediated microbial interactions is vast, both in the chemical complexity of metabolites themselves and the types of interactions they facilitate. Broadly, these metabolites and their interactions can be divided into three functional classes (6). The most straight forward class are the substrates; substrates are metabolites that are catabolized for growth and energy, and they provide a growth response proportional to the amount of carbon provided. Common substrates include sugars, organic acids, amino acids, as well as the organic sulfur compounds dimethylsulfoniopropionate (DMSP) and dihydroxypropanesulfonate (DHPS) (6, 7, 17-21). Facilitators are metabolites that are not directly catabolized for energy but that are necessary for cellular processes, thus they provide growth effect disproportionate to their carbon content. Classic examples include vitamins and siderophores (22-24). Finally, ecological signals are metabolites that provoke a change in phenotype and are generally not metabolized for growth. Ecological signals include quorum sensing molecules such as homoserine lactones (HSL), as well as growth hormones and antimicrobials (8, 11, 15, 25). These categories, of course, are flexible as some metabolites typically thought of as substrates can also act as ecological signals, and vice versa (26, 27).

Exploration in the field of metabolite mediated interactions is limited by three distinct but intersectional challenges: 1) cryptic fluxes, 2) micron-size spatial scales, 3) and the complex chemical background of seawater. Cryptic fluxes occur rapidly through a small pool, and make direct chemical measurements challenging to interpret (7, 28, 29). For (an oversimplified) example, if a hypothetical population of phytoplankton releases glucose at rate of 10 moles per hour, but the surrounding heterotrophs have the capacity to consume glucose at a rate of 20 moles per hour, glucose will be scavenged to the limit of bacterial transporter uptake almost as soon as it is released. A chemical measurement of glucose in this system would find a vanishing

low concentration, even though the flux, and ecological importance, is great. Spatial scales of nanoliters and microns are where most microbial interactions occur (30), as opposed to the scale of liters and meters where much traditional chemical and biological oceanography are studied. At small scales, much of the water column is devoid of cells and oligotrophic, but with distinct patches or hotspots of nutrients and cells. Each phytoplankton cell, as it excretes metabolites, creates a resource rich region, known as the phycosphere (31). In this region, substrate concentrations can be orders of magnitude greater than in the surrounding seawater (31), enabling bacterial growth and recruitment. Finally, despite a complex chemical background of sea water, the metabolites of greatest importance are typically small, polar, and present at picomolar concentrations (32). They are challenging to differentiate and separate from salts, present at about 35 ppt, and the mix of tens to hundreds of thousands of largely recalcitrant organic compounds (33-35).

Recent advances have made great progress towards overcoming the challenges of microbial-scale oceanography. Advancement in analytical chemistry including benzoyl chloride derivatization (12, 36, 37), cation exchange (38), and ^{13}C NMR (12, 35) have made it possible to identify and quantify many seawater metabolites at nano to picomolar concentration from just a few milliliters of sample (36). Modeling and experimental work with microfluidic devices has helped define the physical, chemical, and biological boundaries of the phycosphere, providing tools and frameworks to run experiments at relevant scales (31, 39). Cryptic fluxes, however, remain the most elusive target. Some new approaches can quantify flux using isotopic labels with time-resolved sampling (19, 28, 40). However, these approaches are time and resource intensive, limiting the extent of implementation. One approach to identify the presence of cryptic fluxes, though not their quantitative rates, is to “ask the bacteria what they are eating” by

monitoring gene expression with transcriptomic or proteomic tools, or by monitoring gene fitness with transposon insertion mutant libraries (41-49). These biosensor type approaches are useful because they are easily scalable, allowing high throughput experiments in both field and lab conditions. Further, in the lab, they can be paired with chemical analyses to provide more context for the magnitude of fluxes (12).

Biosensor type experiments have been used to identify many metabolites across all functional classes and involved in various microbial interactions in many marine environments (8, 43, 46, 48, 50, 51). A group of genes that are particularly useful in identifying metabolite exchanges are bacterial uptake transporters. These transporters are the cell's interface with its environment, and bacteria modulate their expression in response to substrate availability (12, 48, 52, 53). As these expression- and gene fitness-based approaches depend on identified functions, their utility is limited by the accuracy of available gene annotations. For bacteria, gene annotations are assigned by a number of computational tools using various methods to compare target sequences to reference sequences of known function, typically assigning function based on amino acid similarity (54-56). These annotation tools are useful, but fail to annotate or make erroneous annotations for large numbers of genes. These errors occur disproportionately in environmental bacteria that have historically been less well studied (54, 55, 57). Computer science advances are improving the quality of annotation software, but the fundamental need is to expand the reference libraries of experimentally characterized genes, including more diverse enzymes from organisms in more diverse environments. Experimental confirmation of gene function is a time and resource intensive process, requiring construction of bacterial knockout mutant strains or overexpression and purification of the target proteins (58, 59). My dissertation

research focuses on experimentally and computationally annotating marine bacterial genes to elucidate the microbial interactions they facilitate.

In my second chapter I employ an arrayed random barcode transposon insertion library (RB-TnSeq) of the model marine bacterium *Ruegeria pomeroyi* DSS-3, to experimentally characterize the cognate substrates 13 organic influx transporters. The arrayed mutant library allowed for high throughput growth screens, expediting the rate of cognate substrate identification. Four of the transporters characterized were previously hypothesized based on gene expression data (taurine, glucose/xylose, isethionate, and cadaverine/putrescine/spermidine); five were previously hypothesized based on homology to experimentally annotated transporters in other bacteria (citrate, glycerol, N-acetylglucosamine, fumarate/malate/succinate, and DMSP); and four had no previous annotations (thymidine, carnitine, cysteate, and 3-hydroxybutyrate). These bring the total number of experimentally-verified organic carbon influx transporters to 18 of 126 in the *R. pomeroyi* genome. In a longitudinal study of a coastal phytoplankton bloom, expression patterns of the experimentally annotated transporters linked them to different stages of the bloom, and also led to the hypothesis that citrate and 3-hydroxybutyrate were among the most highly available bacterial substrates in this bloom.

My third chapter used the experimental transporter annotations from Chapter 2 to inform a phylogeny based approach to identifying the substrates of transporters in ocean metagenomic data. Gene phylogenies were generated from the component sequences of *R. pomeroyi*'s ABC transporters as well as a curated list of reference transporters. These phylogenies revealed that evolutionary analyses of the ATPase transporter component could most accurately and consistently predict cognate substrates. We applied this approach to environmental sequences using the EcoPhylo workflow in anvi'o. Genes of the taurine transporter ATPase (*tauB*) were

recruited from Tara Oceans metagenomes using a Kyoto Encyclopedia of Genes and Genomes (KEGG) (60) hidden markov model (HMM). Phylogenies of recruited Tara sequences and reference sequences showed that a lineage of environmental *tauB* genes could be readily distinguished from outgroup sequences. But unfortunately, the vast majority of the sequences recruited by the HMM were transporters of other substrates, emphasizing the risk of over-annotation errors using common gene annotation tools.

In my fourth chapter, I focus on a specific metabolite-mediated interaction between *R. pomeroyi* and a diatom. This interaction is mediated by a likely two-way exchange of the auxin type phytohormone indole acetic acid (IAA) and its precursor indole-3-acetamide (IAM). IAA, well studied in plant - microbe interactions, can facilitate both mutualistic and pathogenic interactions, but its roles in marine phytoplankton-bacteria interactions are less well understood. I characterize the function of *R pomeroyi*'s iaaH gene, a key enzyme in IAA biosynthesis, and demonstrate that although *R. pomeroyi* has an incomplete IAA biosynthesis pathway, produces IAA when supplemented with the intermediate IAM. When IAM is provided as a sole carbon source, *R. pomeroyi* prioritizes IAA biosynthesis and immediate release from the cell, demonstrating IAA's role as an excreted signal molecule and suggesting that its release is controlled by the exogenous supply of IAM. Co-culture expression data suggests that IAM-IAA exchanges occur between *R. pomeroyi* and diatom *Thalassiosira pseudonana*. A search of isolate genomes reveals incomplete IAA biosynthesis pathways similar to that in *R. pomeroyi* in many environmental bacteria and suggests that autotroph-bacterial exchange of IAM is common in both marine and terrestrial environments.

The planet and its oceans are undergoing a period of rapid, anthropogenic driven, change (61). The response to these changes of the marine microbial communities that mediate much of

the global carbon cycle and form the foundation of the marine food web has broad implications (16). However, to predict the response of the marine microbiome to climate disturbances we must first understand the mechanisms through with microbial interactions occur and ultimately shape the current structure and function of communities. Dissecting these mechanisms requires interdisciplinary research with high-throughput and scalable tools to meet the pressing timeline of climate action. To this end, I present my dissertation research, a contribution to the broader research effort.

References

1. Field CB, Behrenfeld MJ, Randerson JT, Falkowski P. Primary production of the biosphere: integrating terrestrial and oceanic components. Science. 1998;281:237-40.

2. Moran MA, Ferrer-González FX, Fu H, Nowinski B, Olofsson M, Powers MA, et al. The ocean's labile DOC supply chain. Limnol Oceanogr. 2022;67:1007-21.

3. Anderson TR, Ducklow HW. Microbial loop carbon cycling in ocean environments studied using a simple steady-state model. Aquat Microb Ecol. 2001;26:37-49.

4. Azam F, Fenchel T, Field J, Grey J, Meyer-Reil L, Thingstad F. The ecological role of water-column microbes. Mar Ecol Prog Ser. 1983;10:257-63.

5. Landry MR, Schroer WF. Microbial loops. In: Cochran JK, Bokuniewicz HJ, Yager PL, editors. Encyclopedia of ocean sciences. London: Academic Press; 2019.

6. Moran MA, Kujawinski EB, Schroer WF, Amin SA, Bates NR, Bertrand EM, et al. Microbial metabolites in the marine carbon cycle. Nat Microbiol. 2022;7:508-23.

7. Durham BP, Sharma S, Luo H, Smith CB, Amin SA, Bender SJ, et al. Cryptic carbon and sulfur cycling between surface ocean plankton. Proc Natl Acad Sci USA. 2015;112:453-7.

8. Amin S, Hmelo L, Van Tol H, Durham B, Carlson L, Heal K, et al. Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. Nature. 2015;522:98-101.

9. Morris JJ, Johnson ZI, Szul MJ, Keller M, Zinser ER. Dependence of the cyanobacterium *Prochlorococcus* on hydrogen peroxide scavenging microbes for growth at the ocean's surface. PloS One. 2011;6.

10. Roth-Rosenberg D, Aharonovich D, Luzzatto-Knaan T, Vogts A, Zoccarato L, Eigemann F, et al. *Prochlorococcus* cells rely on microbial interactions rather than on chlorotic resting stages to survive long-term nutrient starvation. mBio 2020;11:10.1128/mbio. 01846-20.

11. 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. Proc Natl Acad Sci USA. 2020;117:27445-55.

12. Ferrer-González FX, Widner B, Holderman NR, Glushka J, Edison AS, Kujawinski EB, et al. Resource partitioning of phytoplankton metabolites that support bacterial heterotrophy. ISME J. 2021;15:762-73.

13. Edwards BR, Bidle KD, Van Mooy BA. Dose-dependent regulation of microbial activity on sinking particles by polyunsaturated aldehydes: Implications for the carbon cycle. Proc Natl Acad Sci USA. 2015;112:5909-14.

14. Beiralas R, Ozer N, Segev E. Abundant *Sulfitobacter* marine bacteria protect *Emiliania huxleyi* algae from pathogenic bacteria. ISME Commun. 2023;3:100.

15. Seyedsayamdost MR, Case RJ, Kolter R, Clardy J. The Jekyll-and-Hyde chemistry of *Phaeobacter gallaeciensis*. Nat Chem. 2011;3:331-5.

16. Cavicchioli R, Ripple WJ, Timmis KN, Azam F, Bakken LR, Baylis M, et al. Scientists' warning to humanity: microorganisms and climate change. Nat Rev Microbiol. 2019;17:569-86.

17. Durham BP, Boysen AK, Carlson LT, Groussman RD, Heal KR, Cain KR, et al. Sulfonate-based networks between eukaryotic phytoplankton and heterotrophic bacteria in the surface ocean. Nat Microbiol. 2019;4:1706-15.

18. Gralka M, Pollak S, Cordero OX. Genome content predicts the carbon catabolic preferences of heterotrophic bacteria. Nat Microbiol. 2023:1-10.

19. Zubkov MV, Tarran GA, Mary I, Fuchs BM. Differential microbial uptake of dissolved amino acids and amino sugars in surface waters of the Atlantic Ocean. J Plankton Res. 2008;30:211-20.

20. Kiene RP, Linn LJ. Distribution and turnover of dissolved DMSP and its relationship with bacterial production and dimethylsulfide in the Gulf of Mexico. Limnol Oceanogr. 2000;45:849-61.

21. Schroer WF, Kepner HE, Uchimiya M, Mejia C, Rodriguez LT, Reisch CR, et al. Functional annotation and importance of marine bacterial transporters of plankton exometabolites. ISME Commun. 2023;3:37.

22. Sañudo-Wilhelmy SA, Cutter LS, Durazo R, Smail EA, Gómez-Consarnau L, Webb EA, et al. Multiple B-vitamin depletion in large areas of the coastal ocean. Proc Natl Acad Sci USA. 2012;109:14041-5.

23. Amin SA, Parker MS, Armbrust EV. Interactions between diatoms and bacteria. Microbiol Mol Biol Rev. 2012;76:667-84.

24. Gregor R, Szabo RE, Vercelli GT, Gralka M, Reynolds R, Qu EB, et al. Widespread B-vitamin auxotrophy in marine particle-associated bacteria. bioRxiv. 2023:2023.10. 16.562604.

25. Gram L, Grossart H-P, Schlingloff A, Kiørboe T. Possible quorum sensing in marine snow bacteria: production of acylated homoserine lactones by Roseobacter strains isolated from marine snow. Appl Environ Microbiol. 2002;68:4111-6.

26. Leveau JH, Gerards S. Discovery of a bacterial gene cluster for catabolism of the plant hormone indole 3-acetic acid. FEMS Microbiol Ecol. 2008;65:238-50.

27. Shemi A, Alcolombri U, Schatz D, Farstey V, Vincent F, Rotkopf R, et al. Dimethyl sulfide mediates microbial predator–prey interactions between zooplankton and algae in the ocean. Nat Microbiol. 2021;6:1357-66.

28. Clifford EL, Varela MM, De Corte D, Bode A, Ortiz V, Herndl GJ, et al. Taurine is a major carbon and energy source for marine prokaryotes in the North Atlantic Ocean off the Iberian Peninsula. Microb Ecol. 2019;78:299-312.

29. Hansell DA. Recalcitrant dissolved organic carbon fractions. Ann Rev Mar Sci. 2013;5:421-45.

30. Stocker R. The 100 μm length scale in the microbial ocean. Aquat Microb Ecol. 2015;76:189-94.

31. Seymour JR, Amin SA, Raina J-B, Stocker R. Zooming in on the phycosphere: the ecological interface for phytoplankton–bacteria relationships. Nat Microbiol. 2017;2:1-12.

32. Weber L, Armenteros M, Soule MK, Longnecker K, Kujawinski EB, Apprill A. Extracellular reef metabolites across the protected Jardines de la Reina, Cuba reef system. Front Mar Sci. 2020:582161.

33. Hertkorn N, Benner R, Frommberger M, Schmitt-Kopplin P, Witt M, Kaiser K, et al. Characterization of a major refractory component of marine dissolved organic matter. Geochim Cosmochim Acta. 2006;70:2990-3010.

34. Metwally H, McAllister RG, Konermann L. Exploring the mechanism of salt-induced signal suppression in protein electrospray mass spectrometry using experiments and molecular dynamics simulations. Anal Chem. 2015;87:2434-42.

35. Holderman NR, Ferrer-González FX, Glushka J, Moran MA, Edison AS. Dissolved organic metabolite extraction from high-salt media. NMR Biomed. 2023;36:e4797.

36. Widner B, Kido Soule MC, Ferrer-González FX, Moran MA, Kujawinski EB. Quantification of amine-and alcohol-containing metabolites in saline samples using pre-extraction benzoyl chloride derivatization and ultrahigh performance liquid chromatography tandem mass spectrometry (UHPLC MS/MS). Anal Chem. 2021;93:4809-17.

37. Johnson WM, Soule MCK, Kujawinski EB. Extraction efficiency and quantification of dissolved metabolites in targeted marine metabolomics. Limnol Oceanogr Meth. 2017;15:417-28.

38. Sacks JS, Heal KR, Boysen AK, Carlson LT, Ingalls AE. Quantification of dissolved metabolites in environmental samples through cation-exchange solid-phase extraction paired with liquid chromatography–mass spectrometry. Limnol Oceanogr Meth. 2022;20:683-700.

39. Lambert B, Raina J, Fernandez V, Rinke C, Siboni N, Rubino F, et al. A microfluidics-based in situ chemotaxis assay to study the behaviour of aquatic microbial communities. Nat Microbiol 2017:1344–9.

40. Clifford EL, De Corte D, Amano C, Paliaga P, Ivančić I, Ortiz V, et al. Mesozooplankton taurine production and prokaryotic uptake in the northern Adriatic Sea. Limnol Oceanogr. 2020;65:2730-47.

41. Ferrer-González FX, Hamilton M, Smith CB, Schreier JE, Olofsson M, Moran MA. Bacterial transcriptional response to labile exometabolites from photosynthetic picoeukaryote *Micromonas commoda*. ISME Commun. 2023;3:5.

42. Moran MA, Satinsky B, Gifford SM, Luo H, Rivers A, Chan L-K, et al. Sizing up metatranscriptomics. ISME J. 2013;7:237-43.

43. Nowinski B, Moran MA. Niche dimensions of a marine bacterium are identified using invasion studies in coastal seawater. Nat Microbiol. 2021;6:524-32.

44. Satinsky BM, Gifford SM, Crump BC, Moran MA. Use of internal standards for quantitative metatranscriptome and metagenome analysis. Methods Enzymol. 531: Elsevier; 2013. p. 237-50.

45. Schreier JE, Smith CB, Ioerger TR, Moran MA. A mutant fitness assay identifies bacterial interactions in a model ocean hot spot. Proc Natl Acad Sci USA. 2023;120.

46. Salazar G, Paoli L, Alberti A, Huerta-Cepas J, Ruscheweyh H-J, Cuenca M, et al. Gene expression changes and community turnover differentially shape the global ocean metatranscriptome. Cell. 2019;179:1068-83.

47. Poretsky RS, Hewson I, Sun SL, Allen AE, Zehr JP, Moran MA. Comparative day/night metatranscriptomic analysis of microbial communities in the North Pacific subtropical gyre. Environ Microbiol. 2009;11:1358-75.

48. Sowell SM, Wilhelm LJ, Norbeck AD, Lipton MS, Nicora CD, Barofsky DF, et al. Transport functions dominate the SAR11 metaproteome at low-nutrient extremes in the Sargasso Sea. ISME J. 2009;3:93-105.

49. Schroer WF. Omics techniques in marine microbiology. In: Cochran JK, Bokuniewicz HJ, Yager PL, editors. Encyclopedia of ocean sciences. London: Academic Press; 2019.

50. Teeling H, Fuchs BM, Becher D, Klockow C, Gardebrecht A, Bennke CM, et al. Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. Science. 2012;336:608-11.

51. Lidbury I, Murrell JC, Chen Y. Trimethylamine N-oxide metabolism by abundant marine heterotrophic bacteria. Proc Natl Acad Sci USA. 2014;111:2710-5.

52. Weiten A, Kalvelage K, Neumann-Schaal M, Buschen R, Scheve S, Winklhofer M, et al. Nanomolar responsiveness of marine *Phaeobacter inhibens* DSM 17395 toward carbohydrates and amino acids. Microb Physiol. 2022;32:108-21.

53. Tang K, Jiao N, Liu K, Zhang Y, Li S. Distribution and functions of TonB-dependent transporters in marine bacteria and environments: implications for dissolved organic matter utilization. PloS One. 2012;7:e41204.

54. Lobb B, Tremblay BJ-M, Moreno-Hagelsieb G, Doxey AC. An assessment of genome annotation coverage across the bacterial tree of life. Microb Genom. 2020;6:e000341.

55. Loewenstein Y, Raimondo D, Redfern OC, Watson J, Frishman D, Linial M, et al. Protein function annotation by homology-based inference. Genome Biol. 2009;10:1-8.

56. Radivojac P, Clark WT, Oron TR, Schnoes AM, Wittkop T, Sokolov A, et al. A large-scale evaluation of computational protein function prediction. Nat Methods. 2013;10:221-7.

57. Dimonaco NJ, Aubrey W, Kenobi K, Clare A, Creevey CJ. No one tool to rule them all: prokaryotic gene prediction tool annotations are highly dependent on the organism of study. Bioinformatics. 2022;38:1198-207.

58. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, et al. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol. 2006;2:2006.0008.

59. Clifton BE, Alcolombri U, Jackson CJ, Laurino P. Ultrahigh-affinity transport proteins from ubiquitous marine bacteria reveal mechanisms and global patterns of nutrient uptake. bioRxiv. 2023:2023.02. 16.528805.

60. Kanehisa M, Furumichi M, Sato Y, Kawashima M, Ishiguro-Watanabe M. KEGG for taxonomy-based analysis of pathways and genomes. Nucleic acids research. 2023;51:D587-D92.

61. Pörtner H-O, Roberts DC, Masson-Delmotte V, Zhai P, Tignor M, Poloczanska E, et al. IPCC special report on the ocean and cryosphere in a changing climate. Cambridge, UK: Cambridge University Press; 2019.

CHAPTER 2

FUNCTIONAL ANNOTATION AND IMPORTANCE OF MARINE BACTERIAL TRANSPORTERS OF PLANKTON EXOMETABOLITES[1]

[1]Schroer, WF; Kepner, HE; Uchimiya, M; Mejia, C; Trujillo Rodriguez, L; Reisch, CR; Moran, MA. 2023. *ISME Communications*. 3(1): 37. Reprinted here with permission of the publisher.

Abstract

Metabolite exchange within marine microbial communities transfers carbon and other major elements through global cycles and forms the basis of microbial interactions. Yet lack of gene annotations and concern about the quality of existing ones remain major impediments to revealing currencies of carbon flux. We employed an arrayed mutant library of the marine bacterium *Ruegeria pomeroyi* DSS-3 to experimentally annotate substrates of organic compound transporter systems, using mutant growth and compound drawdown analyses to link transporters to their cognate substrates. Mutant experiments verified substrates for thirteen *R. pomeroyi* transporters. Four were previously hypothesized based on gene expression data (taurine, glucose/xylose, isethionate, and cadaverine/putrescine/spermidine); five were previously hypothesized based on homology to experimentally annotated transporters in other bacteria (citrate, glycerol, *N*-acetylglucosamine, fumarate/malate/succinate, and dimethylsulfoniopropionate); and four had no previous annotations (thymidine, carnitine, cysteate, and 3-hydroxybutyrate). These bring the total number of experimentally-verified organic carbon influx transporters to 18 of 126 in the *R. pomeroyi* genome. In a longitudinal study of a coastal phytoplankton bloom, expression patterns of the experimentally annotated transporters linked them to different stages of the bloom, and also led to the hypothesis that citrate and 3-hydroxybutyrate were among the most highly available bacterial substrates. Improved functional annotation of the gatekeepers of organic carbon uptake is critical for deciphering carbon flux and fate in microbial ecosystems.

Opening statement

MAM and CRR had the initial idea of generating an arrayed library of *Ruegeria pomeroyi* DSS-3 mutants. CRR, CM, and LTR generated and mapped the *R. pomeroyi* RB-TnSeq library, then they performed the colony isolation, identification and arraying of mutants. WFS, MAM, and CRR developed the idea of using the arrayed library to screen transporter mutants, these three authors were involved in the process of selecting putative influx transporter mutants to be included in the transporter sub-array. CRR, CM, and LTR designed and ran the high throughput growth screens. WFS and HEK designed and ran the high resolution follow up screens and analyzed the data. WFS and MU designed the metabolite drawdown screens, WFS ran the drawdown screen experiments and prepared samples for NMR, MU ran the NMR samples and analyzed data. WFS designed and ran the pooled RB-TnSeq experiments, processed the samples. Primers were designed by WFS with the help of CR and Christa Smith. WFS analyzed the RB-TnSeq data with advice from Jeremy Schreier.

WFS and MAM reanalyzed the Monterey Bay transcriptome data. WFS performed the search for homologous transporters in other members of the Roseobacter group. WFS and MAM drafted the manuscript with valuable feedback from all co-authors.

Introduction

The ocean microbiome plays a central role in mediating carbon and element cycles through its unique ability to process organic carbon dissolved in seawater (1-3). Ultimately, bacteria take up and assimilate as much as half of marine net primary production (NPP) in the form of exometabolites derived from excretion and death of phytoplankton and other microbes (3, 4). Given that current and future controls over this globally significant carbon flux are poorly understood, identification of the metabolites produced and consumed by ocean microbes is vitally important (5).

One approach to unraveling marine metabolite flux is through the application of transcriptomic and proteomic tools by which dynamics of the chemical environment can be gleaned from changes in the expression of microbial genes. Such approaches are easy to scale with advancements in sequencing and data sharing (6, 7) and have successfully addressed metabolite dynamics in various microbial systems such as model communities (8), phytoplankton blooms (9, 10), oligotrophic ocean regions (11, 12), and global-scale ocean surveys (13-15). Transporter genes in particular are of value in such approaches because they are a cell's interface with its environment and their expression can reveal the identity of available metabolites (16). A key limitation to their use, however is a dependence on accurate gene annotation to identify protein function. For most microbial transporters, the cognate substrate is still unknown. Others are annotated computationally based on homology (17-19), yet this is error prone when relationships to experimentally annotated genes are distant (20). Indeed, transporters have a lower rate of successful annotation based on homology than catabolic enzymes (19).

Experimental confirmation of gene function is the gold standard for annotation, but is both time and resource intensive. Moreover, it is largely limited to cultured species for which

genetic systems are available, leaving out much of the diversity represented in environmental bacteria. An alternate approach uses pooled transposon mutants whose fitness under defined selection pressure provides hypotheses of gene function (21-24). This method requires only a minimal genetic system to introduce small DNA fragments (transposons) and a protein that catalyzes insertion of the transposons (transposase) into bacterial cells. A recent high-throughput advancement of this method, termed RB-TnSeq (25, 26), uses unique barcodes that link each transposon insertion site to the specific gene it disrupts, thereby allowing mutant pools to be analyzed for fitness through cost-effective amplicon sequencing. A wide taxonomic range of bacteria have been found amenable to RB-TnSeq library construction, resulting in hypothesis generation for gene functions that include stress response, metabolism, phage resistance, and transport (25, 27-29). Hypotheses can be confirmed experimentally if targeted single-gene mutants are subsequently constructed, as for those predicting membrane proteins (28) and catabolic enzymes (30).

For a small number of well-studied model bacterial species, genome-wide mutant arrays have been constructed through painstaking targeted gene deletions to produce libraries of single-gene knockouts across the genome. Excellent tools for gene annotation, these arrayed libraries are currently available for species such as *Escherichia coli* (31), *Acinetobacter baylyi* (32), *Bacillus subtilis* (33, 34), and *Salmonella enterica* (35). Pooled transposon mutant libraries have been used successfully as the starting material for such arrayed libraries (24, 36, 37), but require individual sequencing of tens of thousands of colonies to determine transposon insertion location.

Recently, a modification of the RB-TnSeq approach was used to create an inexpensive arrayed mutant library of the model marine bacterium *Ruegeria pomeroyi* DSS-3 (38). The

method took advantage of the ease of insertion site identification in RB-TnSeq libraries to map 270,000 barcodes to transposon insertions which were condensed into a library of 4,991 mutants, similar to the approach used recently for the anaerobic gut microbe *Bacteroides thetaiotaomicron* (39). *R. pomeroyi* is known for its ecological association with marine phytoplankton and ability to grow on plankton-derived metabolites (16, 40, 41), but to this point substrates of only four of the 126 putative organic compound influx transporters have been experimentally verified via knockout mutants: choline (42), dihydroxypropanesulfonate (DHPS) (41, 43), ectoine (44), and trimethylamine N-oxide (13). Here we leveraged a set of 156 influx transporter mutants from the *R. pomeroyi* arrayed RB-TnSeq (arrayed-RB-TnSeq) library in high-throughput screens against 70 possible substrates to increase knowledge of transporter function. Resulting gene annotations were then applied to a longitudinal study of *R. pomeroyi* transcriptomes following introduction of the bacterium into Monterey Bay, CA, USA seawater on 35 dates during a dinoflagellate bloom (9) in the manner of an ecological invasion study (45). The 13 newly verified transporter annotations provided insights into the metabolites serving roles as substrates to heterotrophic bacteria during a coastal bloom.

Methods

RB-TnSeq library generation and mapping

Full methods for generating and arraying the *R. pomeroyi* barcoded mutant library are provided in Mejia et al. (38). Briefly, a pool of randomly barcoded transposon mutants was constructed according to Wetmore et al. (26) by conjugating *R. pomeroyi* DSS-3 with *E.coli* WM3064 containing the pKMW7 Tn5 barcoded library (strain APA766). The transposon insertion sites were subsequently linked to the unique barcodes through random barcode transposon-site sequencing (RB-TnSeq) (26). RB-TnSeq was achieved by sequencing on

an Illumina NovaSeq6000 with 150-bp paired-end reads (Novogene Co., Sacramento, CA). Bioinformatic analysis of the reads used custom scripts described in Wetmore et al. (26). To construct the arrayed libraries, individual mutants were isolated on ½ YTSS solid medium amended with 100 μg ml^{-1} kanamycin. Colonies were picked after 2 d (Qpix2 automated colony picker; Molecular Devices, San Jose, CA) and arrayed into 384 well plates containing 80 μl of ½ YTSS + kanamycin broth. Plates were incubated at 30ºC for 2-5 d until visible growth appeared and then replicated. Glycerol was added to a final concentration of 20% and plates were frozen at -80ºC. To identify the mutant locations, barcodes were amplified in each well using a unique combination of indexed 16 forward and 24 reverse location primers, allowing indexing of the well position. These barcode amplicons were re-amplified with dual indexed Illumina adapters, and amplicons from this second round of PCR were pooled within a plate for sequencing (38). Of the 270,510 unique barcodes originally present in the pooled RB-TnSeq library, the condensed RB-TnSeq arrayed library contained 4,991 mutants covering 3,048 of 4,284 protein encoding genes with either 1 or 2 mutants, accounting for 71% of *R. pomeroyi*'s genes. From these, the first 156 putative organic compound influx transporter mutants identified in the arraying procedure were re-arrayed into two 96 well plates for subsequent screening (Table S1).

Growth screens

Mutant cultures were pre-grown overnight in ½ YTSS broth with 50 μg ml^{-1} kanamycin. Screens were performed in L1 minimal medium (46) (dx.doi.org/10.17504/protocols.io.jvccn2w) modified to a salinity of 20 ppt and amended with ammonium (3 mM), kanamycin (50 μg ml^{-1}), and phosphorus as PO_4^{3-} at 36 μM. For the initial screen, overnight cultures of individual mutants (2 μl) were inoculated into 198 μl of modified L1 with a single substrate as the sole carbon source at 8 mM carbon. Plates were incubated at 25ºC with shaking, and optical density

(OD_{600}) was read at intervals of 6-12 h using a SpectraMax M3 (Molecular Devices, San Jose, CA) until cultures entered stationary phase at ~24-48 h. Mutants exhibiting phenotypes in the initial screen were moved to the targeted screen in which 4 replicate 200 µl mutant cultures were prepared by inoculating 2 µl of washed (3x) overnight culture into 96 well plates containing 198 µl modified L1 medium and a single substrate at 8 mM carbon. As a positive control, four wells with the same medium were inoculated with washed overnight cultures of the pooled-RB-TnSeq library, used as a proxy for wild-type *R. pomeroyi* growth but harboring a transposon/kanamycin resistance gene insertion. Cultures were grown at 25°C in a Synergy H1 plate reader (BioTek, Winooski, VT, USA) shaking at 425 rpm for 68-72 h. OD_{600} readings were collected once each hour and corrected to a pathlength of 1 cm assuming a volume of 200 µl.

Mutant defect was identified by comparison to the OD_{600} achieved by the pooled-RB-TnSeq library (n=4; ANOVA and TukeyHSD; $p \leq 0.05$) (Table 1). Mutants with significantly lower OD_{600} on multiple substrates were regrown on rich medium to check for viability, and removed from further consideration if they broadly demonstrated poor growth; one mutant was removed after this viability check (SPO2952).

Metabolite drawdown screen

For each mutant-substrate pair identified from the growth screens, 3 replicate 220 µl cultures were prepared in 96 well plates by inoculating 3 µl of washed (3x) overnight mutant cultures into minimal medium containing the candidate substrate at 8 mM carbon. Cultures were grown shaking at 25°C for 24 h or 36 h, depending on the growth rate supported by the carbon source. At termination, 200 µl of medium were collected and centrifuged at 3,700 rpm for 10 min, and the supernatant was stored at -80°C. Metabolite analysis was performed using a Bruker Avance lll 600 MHz spectrometer (Bruker, Billerica, MA, USA) equipped with a 5-mm TCI

cryoprobe. Samples were prepared with addition of a deuterated phosphate buffer (30 mmol L^{-1}, pH 7.4) and the internal standard 2,2-dimethyl-2-silapentane-5-sulfonate-d_6 (DSS, 1 mmol L^{-1}) (10:1 vol:vol) and transferred to 3 mm NMR tubes (Bruker). Data were acquired by a one dimensional ^{1}H experiment with water suppression (noesypr1d, Bruker) at 298K using TopSpin 3.6.4 (Bruker). For glycerol, a ^{1}H *J*-resolved experiment (jresgpprqf) was used to avoid overlapping background peaks. Spectra were processed using NMRPipe on NMRbox (47, 48), and the processed data were analyzed using Metabolomics Toolbox (https://github.com/artedison/Edison_Lab_Shared_Metabolomics_UGA) and MATLAB R2022a (MathWorks). For quantification of metabolites, spectra were normalized to the DSS standard and peak area for representative peaks was calculated. TopSpin experiment settings, NMRpipe spectra processing parameters, and MATLAB data analysis scripts are available in Metabolomics Workbench (see Data Availability).

Pooled-RB-TnSeq experiment

Minimal medium was prepared for 10 substrates at 8 mM carbon in a 96 well plate, 180 µl per well (n=4). Each well was inoculated with 20 µl of washed (3x) overnight culture of the *R. pomeroyi* pooled-RB-TnSeq library. After growth with shaking at 25^{o}C for 72 h, cultures were serially transferred into fresh media three times. After the third transfer, the full culture volume of each well (200 µl) was transferred to 800 µl of minimal medium with substrate for a final 72 h grow out at 25^{o}C. The libraries averaged 15.5 generations under selection (ranging from 13.5 for xylose to 18.5 for *N*-acetylglucosamine). These 1 ml cultures were then transferred to 1.5 ml tubes, pelleted by centrifugation at 8,000 x g for 3 min, and stored at -80^{o}C until further processing. Genomic DNA was extracted from the cell pellets using the DNEasy blood and tissue kit (Qiagen, Hilden, Germany). PCR amplification of RB-TnSeq barcodes was performed

using primers modified from Wetmore et al. (26) with PhusionHF master mix (Fisher, Pittsburgh, PA). An aliquot of 8 ng of product from each sample was pooled, purified using HiPrep beads (MagBio, Gaithersburg, MD, USA), and sequenced on a NextSeq SE150 Mid Output flow cell (SE150) at the Georgia Genomics and Bioinformatics Core Facility (Athens, Georgia, USA). The initial processing and demultiplexing of sequence data was performed using Perl scripts (MapTnSeq.pl, DesignRandomPool.pl, and MultiCodes.pl) provided in Wetmore et al. (26). Following quality control, an average of 35,090 unique barcodes mapped to insertions that fell within the interior 10 to 90% of *R. pomeroyi* coding sequences. In total, 55 million reads were mapped to insertions in 3,570 genes (out of 4,421 protein-encoding genes in the *R. pomeroyi* genome) with a median of 404,513 mapped reads per sample. Further processing of demultiplexed reads was performed in R v3.6.1. Reads mapping to different insertion sites within the same coding sequence were pooled for subsequent analyses. Two sample T tests with multiple comparison adjustment (FDR) were used to identify enrichment or depletion of normalized reads for each transporter on a given substrate relative to all other substrates tested, a modification of the reference medium method for RB-TnSeq analysis described in Borchert et al. (49). To display enrichment or depletion of multiple transporters and substrates on the same scale, we calculated the fold change of each transporter's normalized abundance for a given substrate relative to the normalized abundance of that transporter on all other substrates.

Transporter expression during a Monterey Bay bloom

Processed *R. pomeroyi* transcriptome data (transcripts per million and Z-scores), metadata, and complete experimental methods are available elsewhere (9). Briefly, on 14 days over 5 weeks, *R. pomeroyi* cells were added to 350 ml of unfiltered Monterey Bay surface water (n=3). *R. pomeroyi* was inoculated at cell numbers equivalent to that of the natural heterotrophic

bacteria. Subsequent sequencing analysis indicated that *R. pomeroyi* transcripts averaged 38% of the bacterial reads in the metatranscriptome datasets (50). Cells were collected by filtration after 90 min and processed for RNAseq analysis.

Homologs in the Roseobacter group

Roseobacter strains with complete genomes were selected based on Simon et al. (51). Phylogenic analysis of the 14 selected strains was carried out with a set of 117 single copy genes using GToTree v1.6.37 (52). *R. pomeroyi* transporter genes with homologs in the other strains were identified by BLASTp using Diamond v2.0.14.152 (53), threshold: $E \leq 10^{-5}$ and identity $\geq$ 60% across the full sequence, and checked with a reciprocal best hits analysis. Data analysis and figure generation was performed with R v3.6.1. Manual checks of gene neighborhoods were performed when BLASTp results showed that multicomponent transporters were missing one or more component genes.

Results and Discussion

From a pooled-RB-TnSeq transposon mutant library of *R. pomeroyi* prepared according to Wetmore et al. (26), colonies were individually arrayed into 384 well plates (Fig. 2.1). The gene disrupted in each arrayed mutant was determined by sequencing the transposon barcode in conjunction with indexed primers that indicated plate column and row (38), creating a library that covers 3,048 of 4284 protein-encoded genes in the *R. pomeroyi* genome (71%). From the genome annotation (54, 55), the first 156 mutants that were predicted to encode for 104 organic compound influx transporter proteins (Table S2.1) were re-arrayed into multi-well plates to facilitate functional screens on 70 compounds (Table S2.2) known to be produced by marine phytoplankton (56).

Growth screens identify 13 transporters of 18 substrates

Initial screens of the 156 mutants identified candidate substrates of transporter genes based on OD_{600} deficits after 24-72 h (n=2). These mutants were transferred to a second round of screening in which each candidate substrate/mutant pair was monitored for growth with hourly OD readings and higher replication (n=4). A positive control treatment consisting of the pooled-RB-TnSeq library approximated wild-type growth (Figs. 2.2a, S2.1). We used mutants of three previously confirmed transporters as positive controls for the screening protocols; these were the *betT* (SPO1087) for choline uptake (42), *hpsKLM* (SPO0591-0593) for DHPS uptake (41, 43), and *uehABC* (SPO1145-1147) for ectoine uptake (44) (Table 1, Fig. S1). One mutant (SPO2952) was excluded from consideration because it exhibited growth defects on multiple structurally dissimilar substrates (alanine, arginine, glycolate, malate, spermidine) in initial growth screens and pooled-RB-TnSeq screens, suggesting a general fitness defect not necessarily related to transport function.

The growth-based screening resulted in substrate predictions for 13 *R. pomeroyi* transporters (Figs 2.2a, S2.1, Table 2.1). Four of these were consistent with target metabolites hypothesized based on previous gene expression data: *xylFGH* (SPO0861-0863) (glucose/xylose) (57), *iseKLM* (SPO2356-2358) (isethionate) (58), *potFGHI* (SPO3466-3469) (polyamines: cadaverine, spermidine, and putrescine) (59), and *tauABC* (SPO0674-0676) (taurine) (60). Four were consistent with target metabolites hypothesized based on *in silico* analysis by the GapMind tool for carbon sources, which combines pathway annotations and gap-filling using mutant fitness data (19): *tctABC* (SPO0184-00186) (citrate), *dctMPQ* (SPO2626-2628) (C4 organic acids succinate, fumarate, and malate), *nagTUVW* (SPO1835, SPO1837-1839) (*N*-acetylglucosamine), and *glpVSTPQ* (SPO0608-0612) (glycerol). One was consistent with a

target metabolite based on homology to an experimentally verified transporter in the closely related species *Roseovarius nubinhibens* (61): *dmdT* (SPO3186) [dimethylsulfoniopropionate (DMSP)] (Table 1). Four were novel annotations with no previous substrate predictions: *cntTUVW* (SPO2995-2998) (carnitine), *cuyTUVW* (SPO3040-3043) (cysteate), *hbtABC* (SPO2571-2573) (3-hydroxybutyrate), and *nupABC* (SPO0376-0379) (thymidine) (Table 2.1). Previous studies had identified all hypothesized substrates as endometabolites in cultured phytoplankton or natural plankton communities or as exometabolites in phytoplankton cultures or seawater (56, 62-69)

Metabolite drawdown confirms gene knockout results

Substrate identifications emerging from the growth screens were further tested in metabolite drawdown experiments. Similar to the design of the growth screens, isolated mutants were inoculated into minimal medium with a single substrate as the sole carbon source (n=3), alongside positive control treatments inoculated with the pooled-RB-TnSeq library as an analog for wild type. Spent media samples were collected at 24 h or, for substrates that supported slower growth, at 36-48 h. Substrate concentration was measured by ^{1}H-NMR and a mutant drawdown defect was defined as significantly higher substrate concentration in the mutant cultures compared to the pooled-RB-TnSeq library (ANOVA and TukeyHSD, $p \leq 0.05$, Table 1). All transporter annotations that had emerged from the growth screens were subsequently upheld in these drawdown screens (Figs. 2.2b, S2.2), consistent with gene disruption reducing or eliminating substrate uptake (Figs. 2.2b, S2.1).

Some transporter mutants, such as *betT*, were completely unable to grow on or draw down the substrate (Figs. S2.1, S2.2). This is the expected pattern if the disrupted transporter is the only system for uptake by *R. pomeroyi*. Alternatively, some mutants, such as *dmdT*, were

capable of partial growth and drawdown, but significantly less than the mutant pool (Figs. 2.2, S2.1, S2.2). This pattern suggests that more than one transporter in the *R. pomeroyi* genome can take up the compound. For example, *dmdT* belongs to the BCCT-type family whose members frequently have low substrate affinity (70), suggesting to us that a second, high-affinity transporter was available when substrates became depleted. Indeed, a recent paper identified a high affinity DMSP transporter in *R. pomeroyi* (SPO2441-2443; *dmpXWV*) that, like *dmdT*, only partially explained observed DMSP uptake (71); this second DMSP transporter now brings the number of experimentally verified transporters to 18. In a mixed result, complete loss of growth and drawdown for fumarate yet partial losses for succinate and malate suggests that *dctMPQ* is the only fumarate transporter system in the *R. pomeroyi* genome while the other C4 organic acids have a second transporter (Figs. 2.2B, S2.2, Table 2.1). We note that while we identified three transporters with multiple substrates (C4 organic acids, glucose/xylose, and multiple polyamines), additional multi-substrate transporters would be missed in our analysis if the other target substrate(s) was not among the 70 screened compounds. Further, the efficacy of growth-based screens for identifying transporter substrates is hampered by poor knowledge of the diversity of metabolites that support heterotrophic growth (56); by the inability to test substrates that don't support growth as a sole carbon source; by substrates that can be taken up by more than one transporter; and by the limited availability of transporter mutant collections that can facilitate matching transporters with their cognate substrates.

Pooled-RB-TnSeq studies are consistent with mutant screens

Another approach to identify substrates of bacterial transporters is to place a pooled-RB-TnSeq library under selection on a single carbon source (25). In this case, transporter mutants that exhibit poor growth are identified as candidate uptake systems. We asked whether the

pooled-RB-TnSeq approach would have been sufficient to recognize the *R. pomeroyi* transporters identified here, saving the effort of arraying the RB-TnSeq library while also providing additional information on catabolic and regulatory genes that may support metabolite utilization.

Mutant abundance was calculated for members of the pooled-RB-TnSeq library following selection for growth on ten substrates used in the growth screens (Fig. 2.3). Selection occurred over four growth dilution cycles of 72 h each. Amplicon sequencing of the pooled library at the beginning and end of selection (26) was used to calculate relative enrichment/depletion for each mutant in the pool as a proxy for fitness. For five substrates, the pooled RB-TnSeq results agreed with results from the arrayed mutant screens, identifying the same transporter systems for DHPS, ectoine, glucose, 3-hydroxybutyrate, and spermidine (n=4; T test, $p \leq 0.05$) (Fig. 2.3). For five other substrates, the known transporter mutant was either not significantly depleted from the mutant pool or significantly enriched, and thus transporters were not correctly identified for DMSP, malate, xylose, cysteate, and *N*-acetylglucosamine. Hypotheses for why these were not identified include the presence of a second transporter in the genome (DMSP, malate), poorer growth leading to weaker selection (xylose), and high substrate concentrations decreasing the need for the transporter substrate binding component (cysteate). In a counterintuitive finding, the *N*-acetylglucosamine transporter mutants (*nagTUVW*) were the most enriched populations in the pool, indicating a fitness gain for cells unable to take up the only provided substrate. We hypothesize that this was due to cross-feeding of an *N*-acetylglucosamine degradation product released by the other mutants and initially used only by the *N*-acetylglucosamine transporter mutant. While these results demonstrate that pooled-RB-

TnSeq mutant libraries are excellent tools for low-cost, high-throughput hypothesis generation, predicted transporter annotations nonetheless require experimental follow-up (28, 30).

Transporter expression reveals the metabolite landscape of a coastal phytoplankton bloom

We used an *R. pomeroyi* gene expression dataset from a natural phytoplankton bloom in Fall 2016 in Monterey Bay, CA, USA (50) to assess the ecological relevance of the verified transporters. On 14 dates over 5 weeks during the decline of a bloom dominated by the dinoflagellate *Akashiwo sanguinea*, *R. pomeroyi* cells were introduced into samples from the natural community for 90 min (9). Metatranscriptomic data from each sample were subsequently mapped to the *R. pomeroyi* genome to identify environmental conditions eliciting transcriptional responses on each sample date. We reanalyzed this dataset in light of the new information on transporter function, with the goal of generating insights into bloom-associated metabolites supporting heterotrophic bacterial growth.

To first evaluate the internal consistency of the expression data, pairwise correlation coefficients were calculated for the individual components of the 14 multi-component transporters across the sample dates. Nine systems had within-transporter correlation coefficients above 0.84 (Pearson correlation, $p \leq 0.05$), confirming coherence in the expression patterns for genes in the same transporter system (Fig. 2.4a). The remaining four had within-transporter correlation coefficients ranging from 0.10 to 0.60; three of these, however, had particularly low expression in Monterey Bay (Fig. 2.4b) that may have affected analytical accuracy.

Expression patterns of the carnitine, choline, taurine, and glycerol transporters were positively related to phytoplankton biomass through the bloom (Pearson correlation, $p \leq 0.05$) (Figs. 2.4c, S2.3), and we hypothesize that these compounds were consistently present in the exometabolite pool. Expression of the C4 organic acid transporter and polyamine transporter had

peaks coinciding with the largest drop in phytoplankton biomass (Fig. 2.4c), and we hypothesize that these compounds leaked from senescing or dead phytoplankton. Transcripts from *R. pomeroyi*'s 126 transporter systems were ranked by their abundance in the transcriptomes [mean transcripts per million (TPM), averaged across components for multi-gene transporters]. Making the assumption that heterotrophic bacterial transporter expression is regulated by substrate detection [admittedly an oversimplification (72)], citrate, 3-hydroxybutyrate, taurine, and DMSP, were among the most important sources of organic carbon to *R. pomeroyi* in this dinoflagellate-dominated bloom (expression ranked in the top 25% of transporters). Conversely, DHPS and cysteate were among the least important (ranked in the bottom 25%) (Figs. 2.4b, S2.3), both of which are found in endometabolites of diatoms and coccolithophores but not dinoflagellates (64). The transporter for 3-hydroxybutyrate was of particular interest for two reasons. First, this newly annotated *hbtABC* is the only confirmed bacterial uptake system for 3-hydroxybutyrate (73). Second, *hbtABC* was the third-most highly expressed *R. pomeroyi* transporter in Monterey Bay. While 3-hydroxybutyrate is well studied as the monomer of the bacterial storage compound polyhydroxybutanoate (74), it is not recognized as an ecologically-important component of the marine dissolved organic carbon pool. This is also the case for citrate, whose transporter was the most highly expressed of all the *R. pomeroyi* transporters (~5-fold higher than the second highest; Fig. 2.4b). Previous research showed that this central metabolite of the tricarboxylic acid (TCA) cycle and precursor for amino acid and cofactor biosynthesis was among the more abundant metabolites measured in phytoplankton cells in North Pacific surface seawater (62). Thus its source in Monterey Bay is likely to be the ongoing phytoplankton bloom.

Orthologous transporters are present in other Roseobacter group members

R. pomeroyi and its relatives in the Roseobacter group are recognized for their high abundance in coastal marine environments (75, 76). The cultured members of this group typically have large, well-regulated genomes capable of diverse metabolisms (77) and are often associated with phytoplankton blooms (78, 79). To determine the distribution of the 17 verified transporters in Roseobacter genomes, 13 other strains having closed genomes and representing a broad sampling of the group's phylogenetic diversity (51) were selected for analysis. Transporters for *N*-acetylglucosamine and carnitine are present only in close relatives of *R. pomeroyi*, consistent with vertical transmission (Fig. 2.5). Transporters for the organic sulfur compounds DHPS, taurine, and isethionate are common in deeply branching strains but retained in few lineages. The transporters for cysteate and ectoine are unique or nearly so to *R. pomeroyi,* suggestive of specialized niche dimensions. Finally, transporters for thymidine, citrate, glycerol, and 3-hydroxybutyrate are well conserved throughout Roseobacter genomes (Fig. 2.5), indicating broad importance of these substrates to the ecology of this group. Patchy distribution of transporter orthologs relative to the Roseobacter phylogeny has been reported previously (80).

Conclusions

Thirteen *R. pomeroyi* transporter annotations were confirmed in a screen of 70 metabolites against 156 transporter mutants representing 104 of the bacterium's 126 organic carbon influx transporter systems. The verified gene functions provided new insights into in a longitudinal dataset of *R. pomeroyi* transcription through a natural phytoplankton bloom, revealing details of the metabolite landscape and generating hypotheses that citrate, 3-hydroxybutyrate, taurine, and DMSP were highly available metabolites during the dinoflagellate-dominated Monterey Bay bloom. Comparative analysis of the verified transporters across Roseobacter genomes revealed, on the one hand, narrow niche dimensions restricted to

subgroups (e.g., *R. pomeroyi* and its closest relatives), and on the other, broad ecological characteristics common across the group and reflecting core ecological roles. As is the case for many marine bacterial taxa (81), the streamlined Roseobacter species that are more numerous in ocean microbial communities are poorly represented in culture collections (82). As such, experimental gene annotation is central for analyzing, or re-analyzing, microbial gene, transcript, and protein data harboring extensive untapped knowledge. For *R. pomeroyi*, this study brings the percent of organic compound influx transporters with identified substrates to 14% of the 126 gene systems that acquire metabolites from the ocean's organic carbon pools.

Acknowledgements

The authors thank C. Smith, C. Sanlatte, and J. Schreier for advice and assistance. This work was supported by Simons Foundation grant 542391 to MAM within the Principles of Microbial Ecosystems Collaborative, and NSF award OCE-2019589 to MAM. This is the Center for Chemical Currencies of a Microbial Planet (C-CoMP) publication #022.

Competing Interests

The authors declare no competing interests.

Data Availability

All growth and RB-TnSeq data are available through BCO-DMO project 884792 (https://www.bco-dmo.org/project/884792). All raw NMR data, processing scripts, and processed files for the metabolite drawdown experiment are available in Metabolomics Workbench with Study ID ST002381 (DOI: http://dx.doi.org/10.21228/M8ST4T).

Contributions

WFS, MAM, and CRR conceived of and designed the research; WFS, HEK, CM, and LTR performed the experiments; MU performed NMR analysis; CM, LTR, and CRR generated

and arrayed the mutant library; WFS and MAM conducted statistical analysis, generated figures, and wrote the manuscript with significant input from all co-authors.

References

1. Azam F, Fenchel T, Field J, Grey J, Meyer-Reil L, Thingstad F. The ecological role of water-column microbes. Mar Ecol Prog Ser. 1983;10:257-63.
2. Anderson TR, Ducklow HW. Microbial loop carbon cycling in ocean environments studied using a simple steady-state model. Aquat Microb Ecol. 2001;26:37-49.
3. Moran MA, Ferrer-González FX, Fu H, Nowinski B, Olofsson M, Powers MA, et al. The ocean's labile DOC supply chain. Limnol Oceanogr. 2022;67:1007-21.
4. Cole JJ, Likens GE, Strayer DL. Photosynthetically produced dissolved organic carbon: An important carbon source for planktonic bacteria. Limnol Oceanogr. 1982;27:1080-90.
5. Cavicchioli R, Ripple WJ, Timmis KN, Azam F, Bakken LR, Baylis M, et al. Scientists' warning to humanity: microorganisms and climate change. Nat Rev Microbiol. 2019;17:569-86.
6. Nayfach S, Roux S, Seshadri R, Udwary D, Varghese N, Schulz F, et al. A genomic catalog of Earth's microbiomes. Nat Biotechnol. 2021;39:499-509.
7. Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y, et al. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. Nat Biotechnol. 2016;34:828-37.
8. Amin S, Hmelo L, Van Tol H, Durham B, Carlson L, Heal K, et al. Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. Nature. 2015;522:98-101.
9. Nowinski B, Moran MA. Niche dimensions of a marine bacterium are identified using invasion studies in coastal seawater. Nat Microbiol. 2021;6:524-32.

10. Teeling H, Fuchs BM, Becher D, Klockow C, Gardebrecht A, Bennke CM, et al. Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. Science. 2012;336:608-11.
11. Poretsky RS, Hewson I, Sun SL, Allen AE, Zehr JP, Moran MA. Comparative day/night metatranscriptomic analysis of microbial communities in the North Pacific subtropical gyre. Environ Microbiol. 2009;11:1358-75.
12. Sowell SM, Wilhelm LJ, Norbeck AD, Lipton MS, Nicora CD, Barofsky DF, et al. Transport functions dominate the SAR11 metaproteome at low-nutrient extremes in the Sargasso Sea. ISME J. 2009;3:93-105.
13. Lidbury I, Murrell JC, Chen Y. Trimethylamine N-oxide metabolism by abundant marine heterotrophic bacteria. Proc Natl Acad Sci USA. 2014;111:2710-5.
14. Tang K, Jiao N, Liu K, Zhang Y, Li S. Distribution and functions of TonB-dependent transporters in marine bacteria and environments: implications for dissolved organic matter utilization. PloS One. 2012;7:e41204.
15. Salazar G, Paoli L, Alberti A, Huerta-Cepas J, Ruscheweyh H-J, Cuenca M, et al. Gene expression changes and community turnover differentially shape the global ocean metatranscriptome. Cell. 2019;179:1068-83.
16. Ferrer-González FX, Widner B, Holderman NR, Glushka J, Edison AS, Kujawinski EB, et al. Resource partitioning of phytoplankton metabolites that support bacterial heterotrophy. ISME J. 2021;15:762-73.
17. Chang Y-C, Hu Z, Rachlin J, Anton BP, Kasif S, Roberts RJ, et al. COMBREX-DB: an experiment centered database of protein function: knowledge, predictions and knowledge gaps. Nucleic Acids Res. 2016;44:D330-D5.

18. Erbilgin O, Rübel O, Louie KB, Trinh M, Raad Md, Wildish T, et al. MAGI: A method for metabolite annotation and gene integration. ACS Chem Biol. 2019;14:704-14.
19. Price MN, Deutschbauer AM, Arkin AP. Filling gaps in bacterial catabolic pathways with computation and high-throughput genetics. PLoS Genet. 2022;18:e1010156.
20. Schnoes AM, Brown SD, Dodevski I, Babbitt PC. Annotation error in public databases: misannotation of molecular function in enzyme superfamilies. PLoS Comput Biol. 2009;5:e1000605.
21. Griffin JE, Gawronski JD, DeJesus MA, Ioerger TR, Akerley BJ, Sassetti CM. High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. PLoS Pathog. 2011;7:e1002251.
22. Van Opijnen T, Camilli A. A fine scale phenotype–genotype virulence map of a bacterial pathogen. Genome Res. 2012;22:2541-51.
23. Chao MC, Abel S, Davis BM, Waldor MK. The design and analysis of transposon insertion sequencing experiments. Nat Rev Microbiol. 2016;14:119-28.
24. Cameron DE, Urbach JM, Mekalanos JJ. A defined transposon mutant library and its use in identifying motility genes in Vibrio cholerae. Proc Natl Acad Sci USA. 2008;105:8736-41.
25. Price MN, Wetmore KM, Waters RJ, Callaghan M, Ray J, Liu H, et al. Mutant phenotypes for thousands of bacterial genes of unknown function. Nature. 2018;557:503-9.
26. Wetmore KM, Price MN, Waters RJ, Lamson JS, He J, Hoover CA, et al. Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons. MBio. 2015;6:e00306-15.

27. Adler BA, Kazakov AE, Zhong C, Liu H, Kutter E, Lui LM, et al. The genetic basis of phage susceptibility, cross-resistance and host-range in *Salmonella*. Microbiology. 2021;167:001126.

28. Jensen HM, Eng T, Chubukov V, Herbert RA, Mukhopadhyay A. Improving membrane protein expression and function using genomic edits. Sci Rep. 2017;7:1-14.

29. Wehrs M, Thompson MG, Banerjee D, Prahl J-P, Morella NM, Barcelos CA, et al. Investigation of Bar-seq as a method to study population dynamics of *Saccharomyces cerevisiae* deletion library during bioreactor cultivation. Microb Cell Fact. 2020;19:1-15.

30. Clark IC, Youngblut M, Jacobsen G, Wetmore KM, Deutschbauer A, Lucas L, et al. Genetic dissection of chlorate respiration in *Pseudomonas stutzeri* PDA reveals syntrophic (per) chlorate reduction. Environ Microbiol. 2016;18:3342-54.

31. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, et al. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol. 2006;2:2006.0008.

32. De Berardinis V, Vallenet D, Castelli V, Besnard M, Pinet A, Cruaud C, et al. A complete collection of single-gene deletion mutants of *Acinetobacter baylyi* ADP1. Mol Syst Biol. 2008;4:174.

33. Koo B-M, Kritikos G, Farelli JD, Todor H, Tong K, Kimsey H, et al. Construction and analysis of two genome-scale deletion libraries for *Bacillus subtilis*. Cell Syst. 2017;4:291-305. e7.

34. Kobayashi K, Ehrlich SD, Albertini A, Amati G, Andersen K, Arnaud M, et al. Essential *Bacillus subtilis* genes. Proc Natl Acad Sci USA. 2003;100:4678-83.

35. Porwollik S, Santiviago CA, Cheng P, Long F, Desai P, Fredlund J, et al. Defined single-gene and multi-gene deletion mutant collections in *Salmonella enterica* sv Typhimurium. PloS One. 2014;9:e99820.

36. Liberati NT, Urbach JM, Miyata S, Lee DG, Drenkard E, Wu G, et al. An ordered, nonredundant library of *Pseudomonas aeruginosa* strain PA14 transposon insertion mutants. Proc Natl Acad Sci USA. 2006;103:2833-8.

37. Ramage B, Erolin R, Held K, Gasper J, Weiss E, Brittnacher M, et al. Comprehensive arrayed transposon mutant library of *Klebsiella pneumoniae* outbreak strain KPNIH1. J Bacteriol. 2017;199:e00352-17.

38. Mejia C, Trujillo Rodriguez L, Poudel R, Ellington A, Rivers A, Reisch C. An arrayed transposon library of *Ruegeria pomeroyi* DSS-3. bioRxiv. 2022:2022.09.11.507510.

39. Shiver AL, Culver R, Deutschbauer AM, Huang KC. Rapid ordering of barcoded transposon insertion libraries of anaerobic bacteria. Nat Protoc. 2021;16:3049-71.

40. Uchimiya M, Schroer W, Olofsson M, Edison AS, Moran MA. Diel investments in metabolite production and consumption in a model microbial system. ISME J. 2022:1-12.

41. Landa M, Burns AS, Roth SJ, Moran MA. Bacterial transcriptome remodeling during sequential co-culture with a marine dinoflagellate and diatom. ISME J. 2017;11:2677-90.

42. Lidbury I, Kimberley G, Scanlan DJ, Murrell JC, Chen Y. Comparative genomics and mutagenesis analyses of choline metabolism in the marine Roseobacter clade. Environ Microbiol. 2015;17:5048-62.

43. Mayer J, Huhn T, Habeck M, Denger K, Hollemeyer K, Cook AM. 2, 3-Dihydroxypropane-1-sulfonate degraded by *Cupriavidus pinatubonensis* JMP134: purification of dihydroxypropanesulfonate 3-dehydrogenase. Microbiology. 2010;156:1556-64.

44. Schulz A, Stöveken N, Binzen IM, Hoffmann T, Heider J, Bremer E. Feeding on compatible solutes: A substrate-induced pathway for uptake and catabolism of ectoines and its genetic control by EnuR. Environ Microbiol. 2017;19:926-46.

45. Mallon CA, Van Elsas JD, Salles JF. Microbial invasions: the process, patterns, and mechanisms. Trends Microbiol. 2015;23:719-29.

46. Guillard R, Hargraves P. *Stichochrysis immobilis* is a diatom, not a chrysophyte. Phycologia. 1993;32:234-6.

47. Maciejewski MW, Schuyler AD, Gryk MR, Moraru II, Romero PR, Ulrich EL, et al. NMRbox: a resource for biomolecular NMR computation. Biophys J. 2017;112:1529-34.

48. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR. 1995;6:277-93.

49. Borchert AJ, Bleem A, Beckham GT. Experimental and analytical approaches for improving the resolution of randomly barcoded transposon insertion sequencing (RB-TnSeq) studies. ACS Synth Biol. 2022;11:2015-21.

50. Nowinski B, Smith CB, Thomas CM, Esson K, Marin R, Preston CM, et al. Microbial metagenomes and metatranscriptomes during a coastal phytoplankton bloom. Sci Data. 2019;6:1-7.

51. Simon M, Scheuner C, Meier-Kolthoff JP, Brinkhoff T, Wagner-Döbler I, Ulbrich M, et al. Phylogenomics of *Rhodobacteraceae* reveals evolutionary adaptation to marine and non-marine habitats. ISME J. 2017;11:1483-99.

52. Lee MD. GToTree: a user-friendly workflow for phylogenomics. Bioinformatics. 2019;35:4162-4.

53. Buchfink B, Reuter K, Drost H-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat Methods. 2021;18:366-8.

54. Rivers AR, Burns AS, Chan L-K, Moran MA. Experimental identification of small non-coding RNAs in the model marine bacterium *Ruegeria pomeroyi* DSS-3. Front Microbiol. 2016;7:380.

55. Moran MA, Buchan A, González JM, Heidelberg JF, Whitman WB, Kiene RP, et al. Genome sequence of *Silicibacter pomeroyi* reveals adaptations to the marine environment. Nature. 2004;432:910-3.

56. Moran MA, Kujawinski EB, Schroer WF, Amin SA, Bates NR, Bertrand EM, et al. Microbial metabolites in the marine carbon cycle. Nat Microbiol. 2022;7:508-23.

57. Wiegmann K, Hensler M, Wöhlbrand L, Ulbrich M, Schomburg D, Rabus R. Carbohydrate catabolism in *Phaeobacter inhibens* DSM 17395, a member of the marine Roseobacter clade. Appl Environ Microbiol. 2014;80:4725-37.

58. Weinitschke S, Sharma PI, Stingl U, Cook AM, Smits TH. Gene clusters involved in isethionate degradation by terrestrial and marine bacteria. Appl Environ Microbiol. 2010;76:618-21.

59. Mou X, Sun S, Rayapati P, Moran MA. Genes for transport and metabolism of spermidine in *Ruegeria pomeroyi* DSS-3 and other marine bacteria. Aquat Microb Ecol. 2010;58:311-21.

60. Gorzynska AK, Denger K, Cook AM, Smits TH. Inducible transcription of genes involved in taurine uptake and dissimilation by *Silicibacter pomeroyi* DSS-3. Arch Microbiol. 2006;185:402-6.

61. Sun L, Curson AR, Todd JD, Johnston AW. Diversity of DMSP transport in marine bacteria, revealed by genetic analyses. Biogeochemistry. 2012;110:121-30.

62. Boysen AK, Carlson LT, Durham BP, Groussman RD, Aylward FO, Ribalet F, et al. Particulate metabolites and transcripts reflect diel oscillations of microbial activity in the surface ocean. mSystems. 2021;6:e00896-20.

63. Dawson H, Heal K, Torstensson A, Carlson L, Ingalls A, Young J. Large diversity in nitrogen-and sulfur-containing compatible solute profiles in polar and temperate diatoms. Integr Comp Biol. 2020;60:1401-13.

64. Durham BP, Boysen AK, Carlson LT, Groussman RD, Heal KR, Cain KR, et al. Sulfonate-based networks between eukaryotic phytoplankton and heterotrophic bacteria in the surface ocean. Nat Microbiol. 2019;4:1706-15.

65. Durham BP, Boysen AK, Heal KR, Carlson LT, Boccamazzo R, Deodato CR, et al. Chemotaxonomic patterns in intracellular metabolites of marine microbial plankton. Front Mar Sci. 2022:1749.

66. Heal KR, Kellogg NA, Carlson LT, Lionheart RM, Ingalls AE. Metabolic consequences of cobalamin scarcity in the diatom *Thalassiosira pseudonana* as revealed through metabolomics. Protist. 2019;170:328-48.

67. Sacks JS, Heal KR, Boysen AK, Carlson LT, Ingalls AE. Quantification of dissolved metabolites in environmental samples through cation-exchange solid-phase extraction paired with liquid chromatography–mass spectrometry. Limnol Oceanogr Methods. 2022;20:683-700.

68. 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. Proc Natl Acad Sci USA. 2020;117:27445-55.

69. Weber L, Armenteros M, Soule MK, Longnecker K, Kujawinski EB, Apprill A. Extracellular reef metabolites across the protected Jardines de la Reina, Cuba reef system. Front Mar Sci. 2020:582161.

70. Chen C, Beattie GA. *Pseudomonas syringae* BetT is a low-affinity choline transporter that is responsible for superior osmoprotection by choline over glycine betaine. J Bacteriol. 2008:2717-25.

71. Li C-Y, Mausz MA, Murphy A, Zhang N, Chen X-L, Wang S-Y, et al. Ubiquitous occurrence of a dimethylsulfoniopropionate ABC transporter in abundant marine bacteria. ISME J. 2023:1-9.

72. Moran MA, Satinsky B, Gifford SM, Luo H, Rivers A, Chan L-K, et al. Sizing up metatranscriptomics. ISME J. 2013;7:237-43.

73. Mierziak J, Burgberger M, Wojtasik W. 3-hydroxybutyrate as a metabolite and a signal molecule regulating processes of living organisms. Biomolecules. 2021;11:402.

74. Findlay RH, White DC. Polymeric beta-hydroxyalkanoates from environmental samples and *Bacillus megaterium*. Appl Environ Microbiol. 1983;45:71-8.

75. González JM, Moran MA. Numerical dominance of a group of marine bacteria in the alpha-subclass of the class Proteobacteria in coastal seawater. Appl Environ Microbiol. 1997;63:4237-42.

76. Giebel H-A, Kalhoefer D, Lemke A, Thole S, Gahl-Janssen R, Simon M, et al. Distribution of Roseobacter RCA and SAR11 lineages in the North Sea and characteristics of an abundant RCA isolate. ISME J. 2011;5:8-19.

77. Luo H, Moran MA. Evolutionary ecology of the marine Roseobacter clade. Microbiol Mol Biol Rev. 2014;78:573-87.

78. González JM, Simó R, Massana R, Covert JS, Casamayor EO, Pedrós-Alió C, et al. Bacterial community structure associated with a dimethylsulfoniopropionate-producing North Atlantic algal bloom. Appl Environ Microbiol. 2000;66:4237-46.

79. Buchan A, LeCleir GR, Gulvik CA, González JM. Master recyclers: features and functions of bacteria associated with phytoplankton blooms. Nat Rev Microbiol. 2014;12:686-98.

80. Newton RJ, Griffin LE, Bowles KM, Meile C, Gifford S, Givens CE, et al. Genome characteristics of a generalist marine bacterial lineage. ISME J. 2010;4:784-98.

81. Stewart EJ. Growing unculturable bacteria. J Bacteriol. 2012;194:4151-60.

82. Buchan A, González JM, Moran MA. Overview of the marine Roseobacter lineage. Appl Environ Microbiol. 2005;71:5665-77.

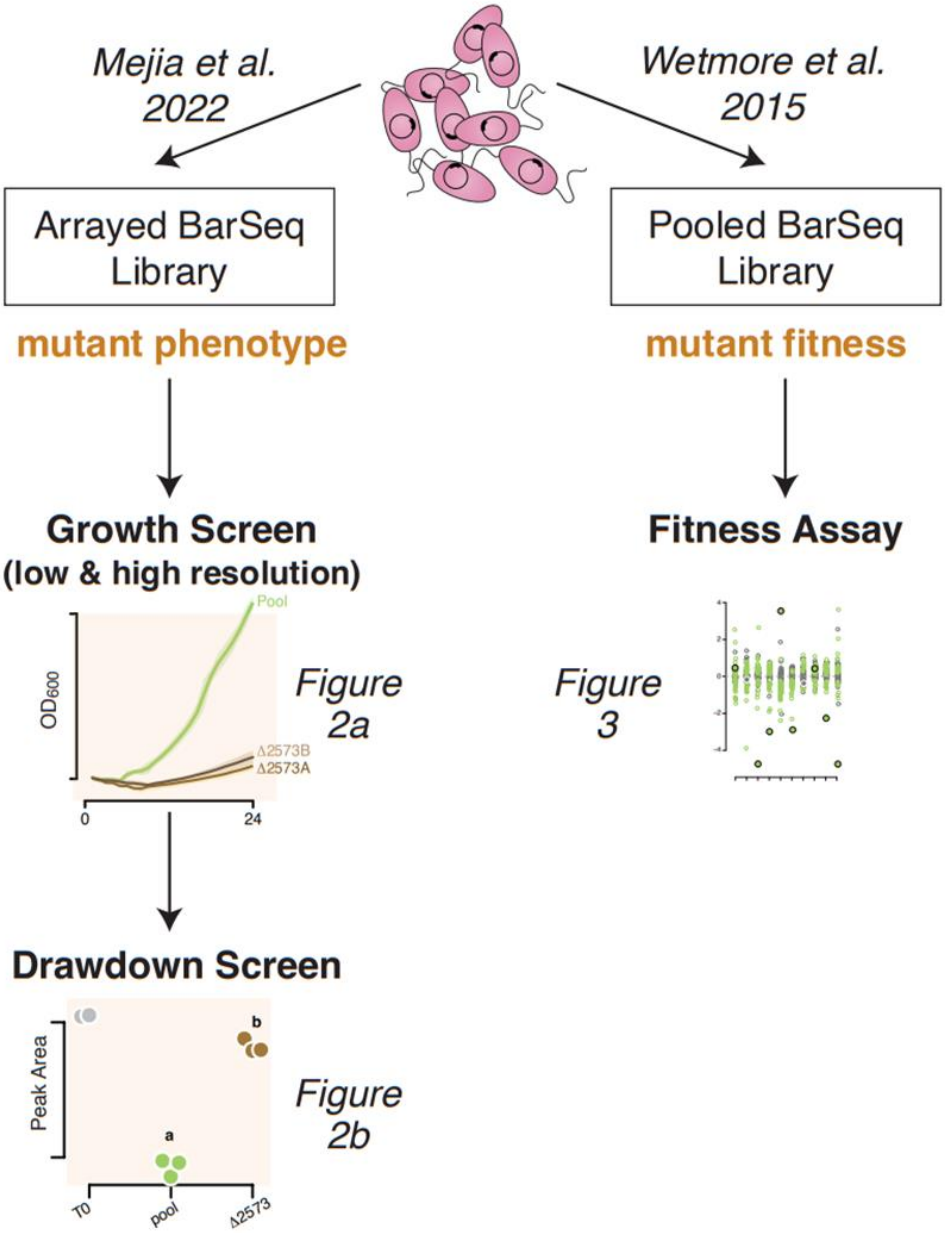


Fig. 2.1. Experimental flow chart for the R. pomeroyi RB-TnSeq library. Right path: Pooled mutant populations (pooled-RB-TnSeq library) are used for gene fitness assays. Left path: Individual transporter mutants from the library (arrayed-RB-TnSeq library) are used to screen for growth and metabolite drawdown.

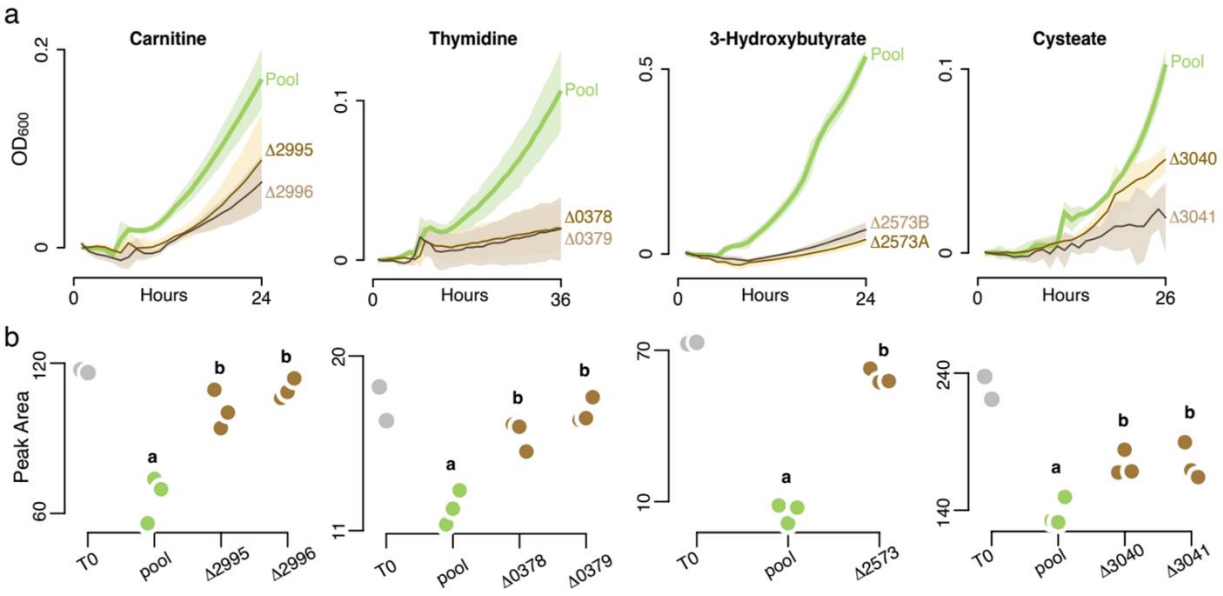


Fig. 2.2. Growth and metabolite drawdown results leading to four novel transporter annotations. a) Growth of transporter mutants compared to growth of the pooled-RB-TnSeq library (an analog for wild-type growth but carrying transposon and resistance gene insertions) on selected marine plankton metabolites. Shaded regions indicate 95% confidence intervals (n=4). Numbers refer to Ruegeria pomeroyi DSS-3 locus tags. b) Substrate concentrations (1H-NMR peak area) after growth of mutants (brown symbols, n=3) or the pooled-RB-TnSeq library (green symbols, n=3), and at inoculation (gray symbols, n=2). Letters that differ indicate that peak areas for the mutants) were significantly higher than for the pooled-RB-TnSeq library (ANOVA, n=3, $p < 0.05$), with a TukeyHSD test carried out when multiple mutants for the same substrate were tested ($p < 0.05$). For full results, see Table 1.

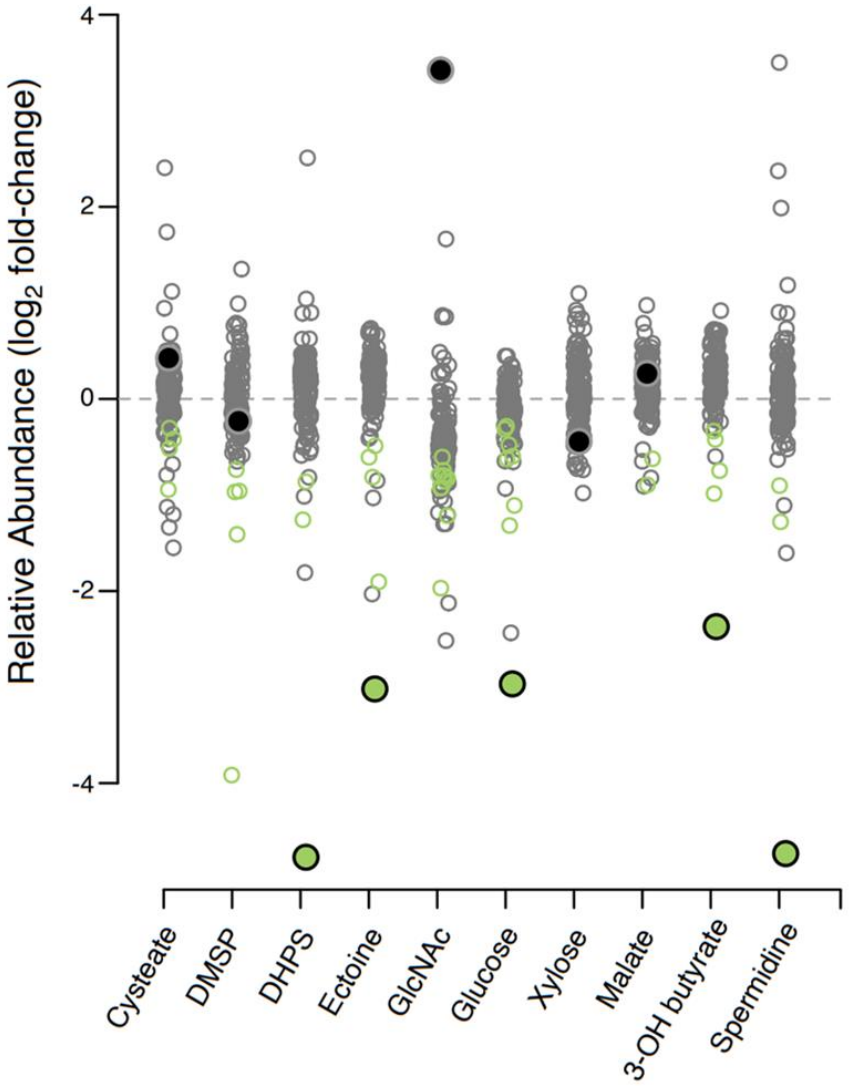


Fig. 2.3. Relative abundance of Ruegeria pomeroyi DSS-3 transporter mutants following selection of the pooled-RB-TnSeq library for growth on 10 metabolites. Green symbols indicate significant mutant depletion (T-test, n=4, Benjamini-Hochberg adjusted $p < 0.05$) and gray symbols indicate non-significant changes. The larger filled symbols indicate the identified transporter for that metabolite as determined from growth and drawdown assays of individual mutants, and is colored green if it was correctly identified, and colored gray if not. Mutant enrichment/depletion for multi-gene transporter systems is plotted as the average of all components.

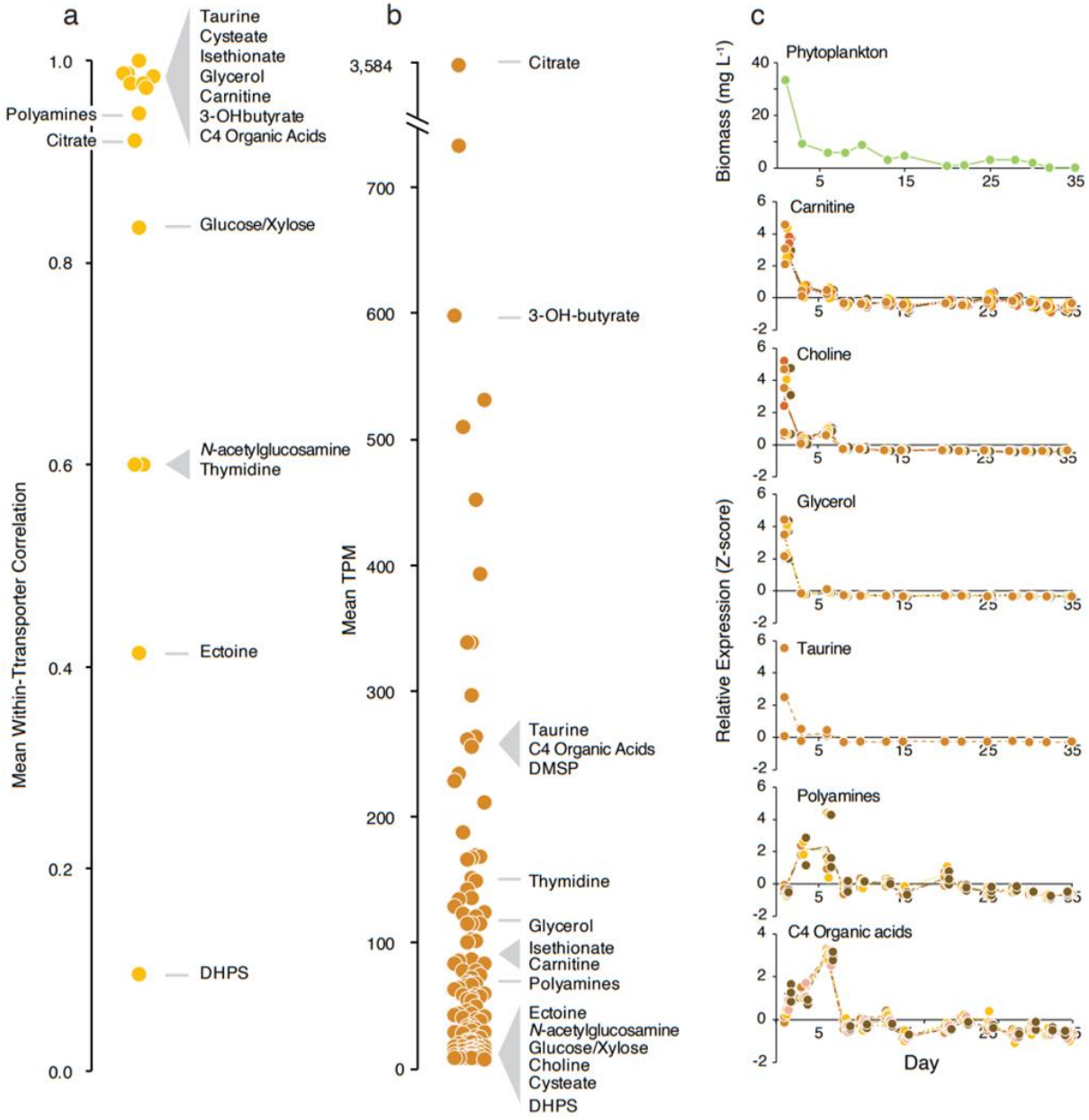


Fig. 2.4. Transporter expression in a Monterey Bay invasion study. a) Mean correlation coefficients of relative expression levels of genes within multi-component transporter systems in Monterey Bay, in Fall, 2016. b) Mean relative expression levels for the 126 R. pomeroyi transporter systems when introduced into Monterey Bay seawater, averaged across 14 dates. c) Relative expression of selected R. pomeroyi DSS-3 transporters in Monterey Bay seawater on

each of 14 dates over 35 days, normalized as Z-scores. Expression of all verified transporters is shown in Fig. S3. Transporters have 1-5 component genes (each colored in different shades of brown) and each component gene has three replicates plotted individually. Lines connect the component mean expression through time. Total phytoplankton biomass (□g C L-1) during the 5-week sampling period is also shown.

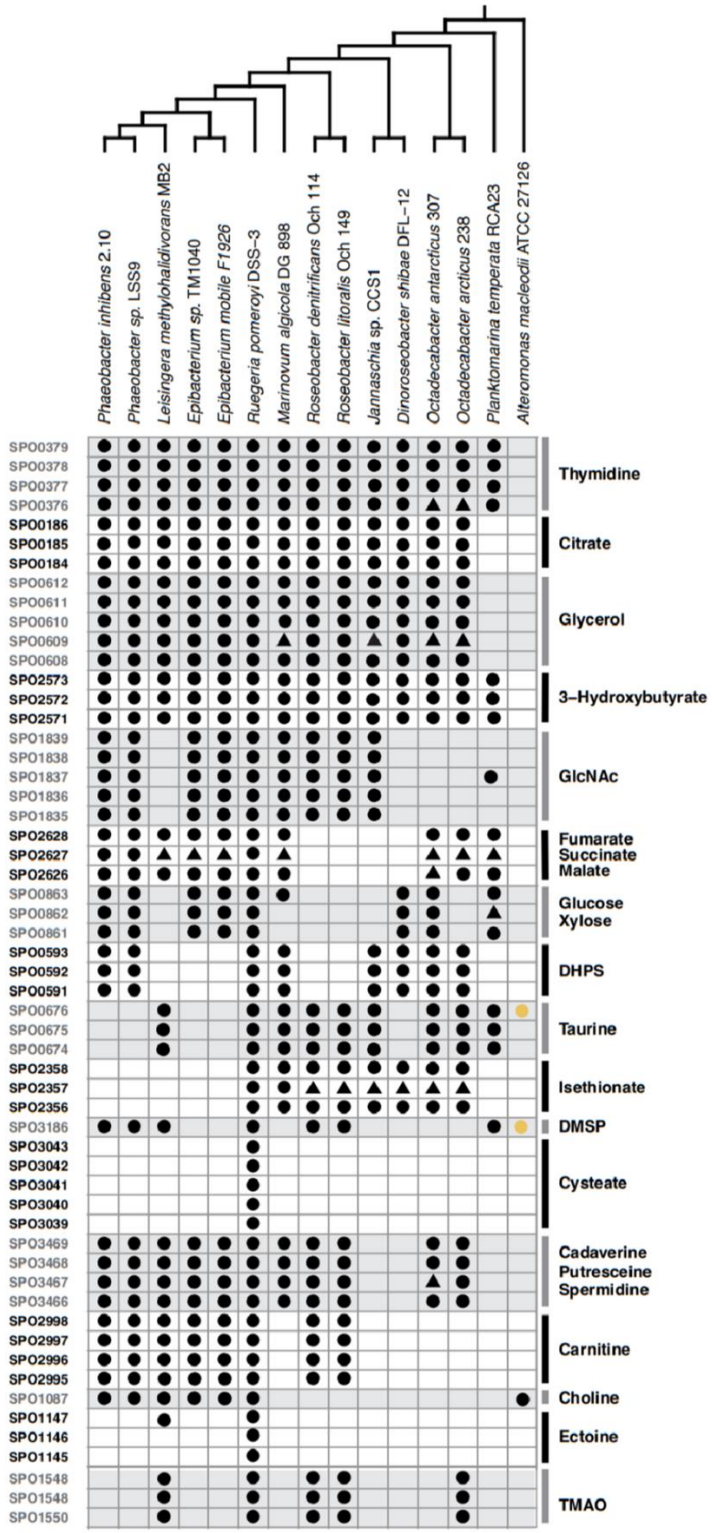


Fig. 2.5. Orthologs of the verified R. pomeroyi DSS-3 transporter systems in Roseobacter group members. Each row indicates a single gene and shading groups the genes that make up multi-

component transporters. Black circles denote orthologs identified by BLASTp using e < 10-5, identity thresholds >60%, and reciprocal best hit analysis. Black triangles denote orthologs of multicomponent transporters that did not meet the BLAST thresholds but were co-located in a transporter operon with components that did. Yellow circles denote potential orthologs in the outgroup A. macleodii ATCC27126 with identities <40%. Species phylogeny is based on analysis of 117 single copy genes.

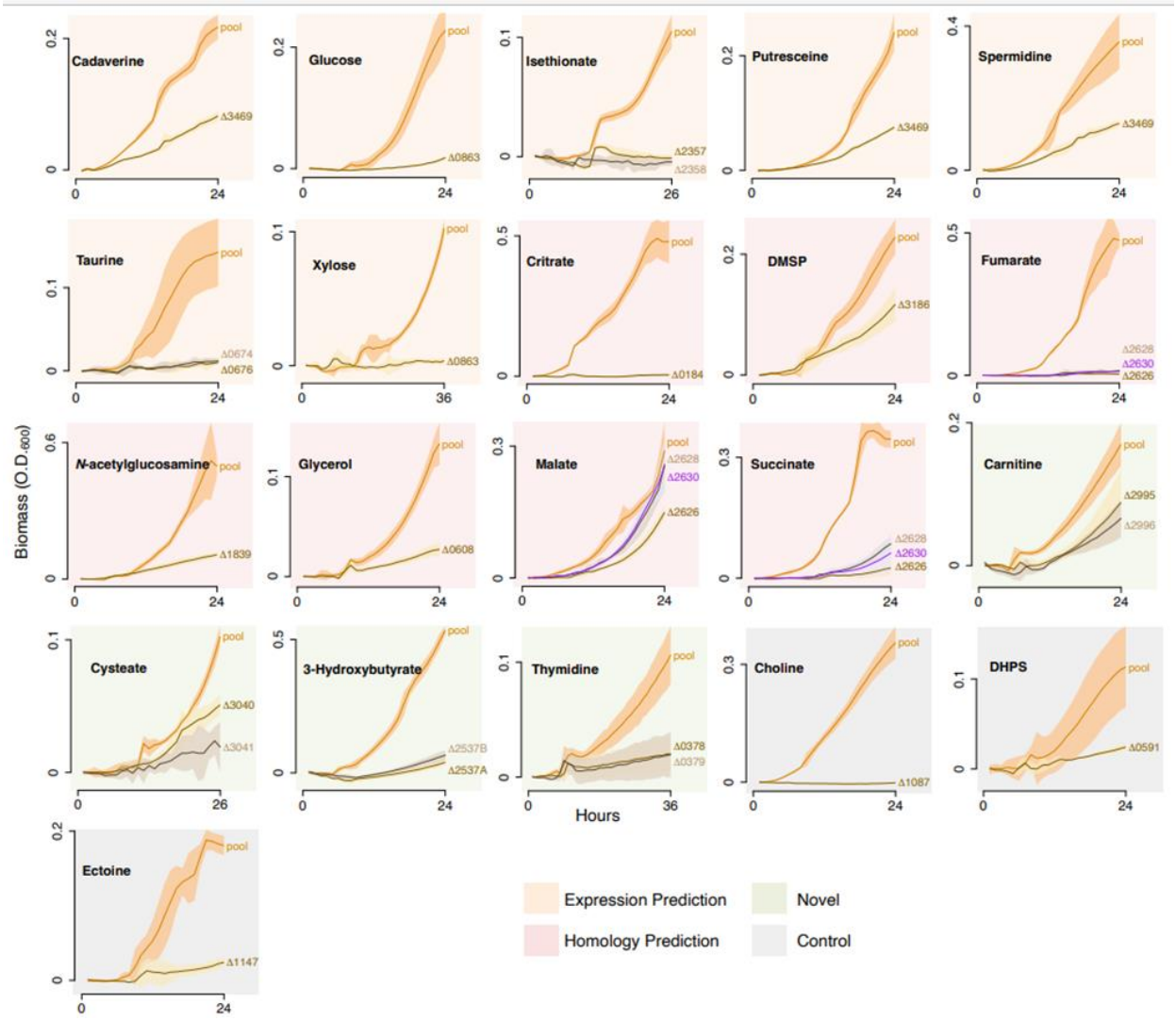


Figure S2.1. Fig. S1. Growth of individual transporter mutants and the pooled-BarSeq library (an analog for wild-type growth) on selected marine plankton metabolites. Substrates are grouped by previous annotation status of the transporter: orange, predicted by expression; pink, predicted by homology; green, novel; gray, positive controls. Shaded regions indicate 95% confidence intervals (n=4). Numbers refer to *Ruegeria pomeroyi* DSS-3 locus tags (Table 1). Significance indications are from ANOVA and TukeyHSD tests ($p < 0.05$) for difference in OD600 between pooled-BarSeq and isolated mutants. DMSP, dimethylsulfoniopropionate; DHPS, dihydroxypropanesulfonate.

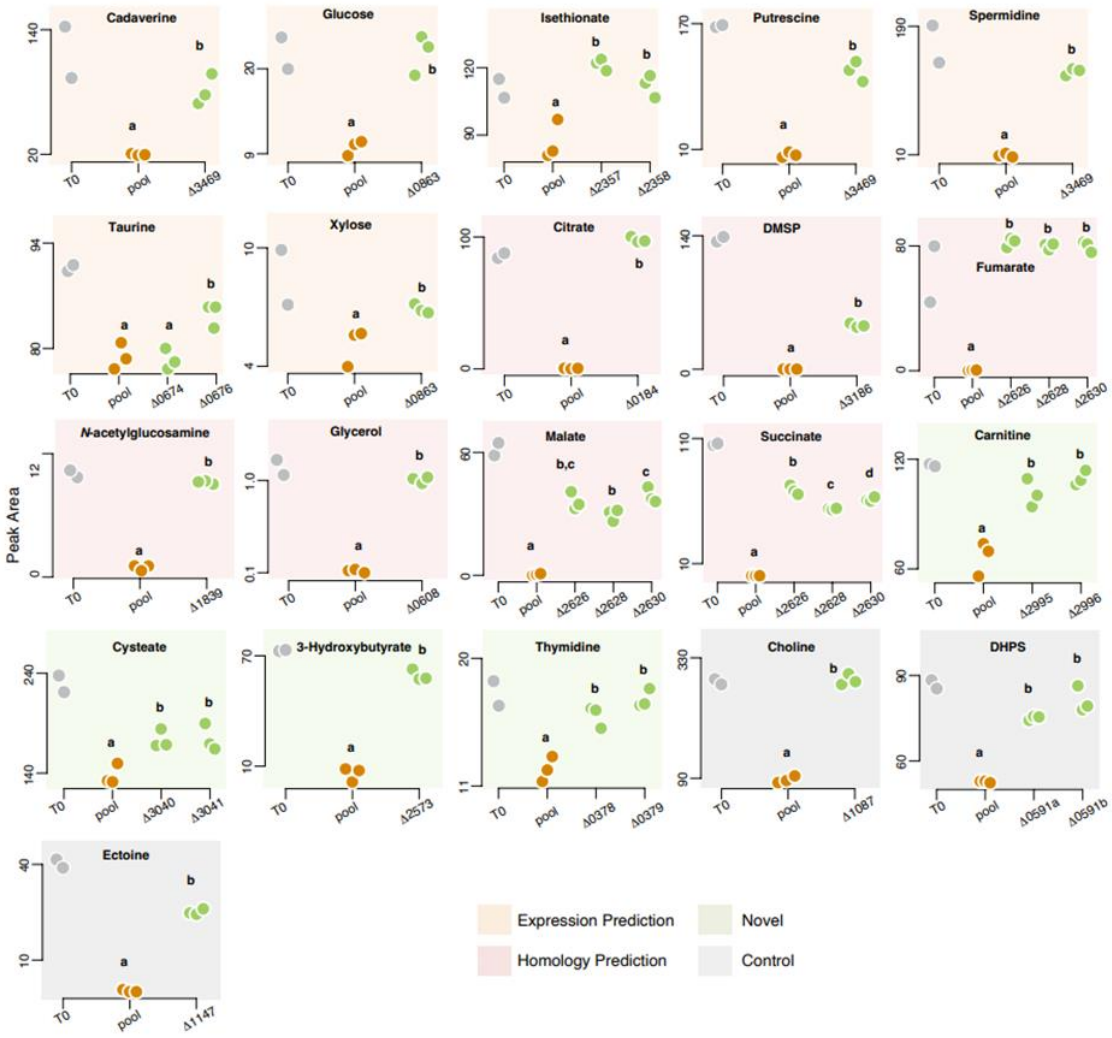


Figure S2.2. Fig. S2. Substrate concentrations during growth of individual transporter mutants and the pooled-BarSeq library (an analog for wild-type growth) on selected marine plankton metabolites. T0, substrate concentration immediately after inoculation (n=2); pool, substrate concentration after BarSeq mutant pool growth (n=3); Δxxxx, substrate concentration after mutant growth, with xxxx indicating Ruegeria pomeroyi DSS-3 locus tag (Table 1) (n=3). Plots are grouped by the previous annotation status of the transporter: orange, predicted by expression; pink, predicted by homology; green, novel; gray, positive control. Significance indications are from ANOVA and TukeyHSD tests ($p < 0.05$) for difference in peak area between BarSeq and isolated mutants. For full results, see Table 1.

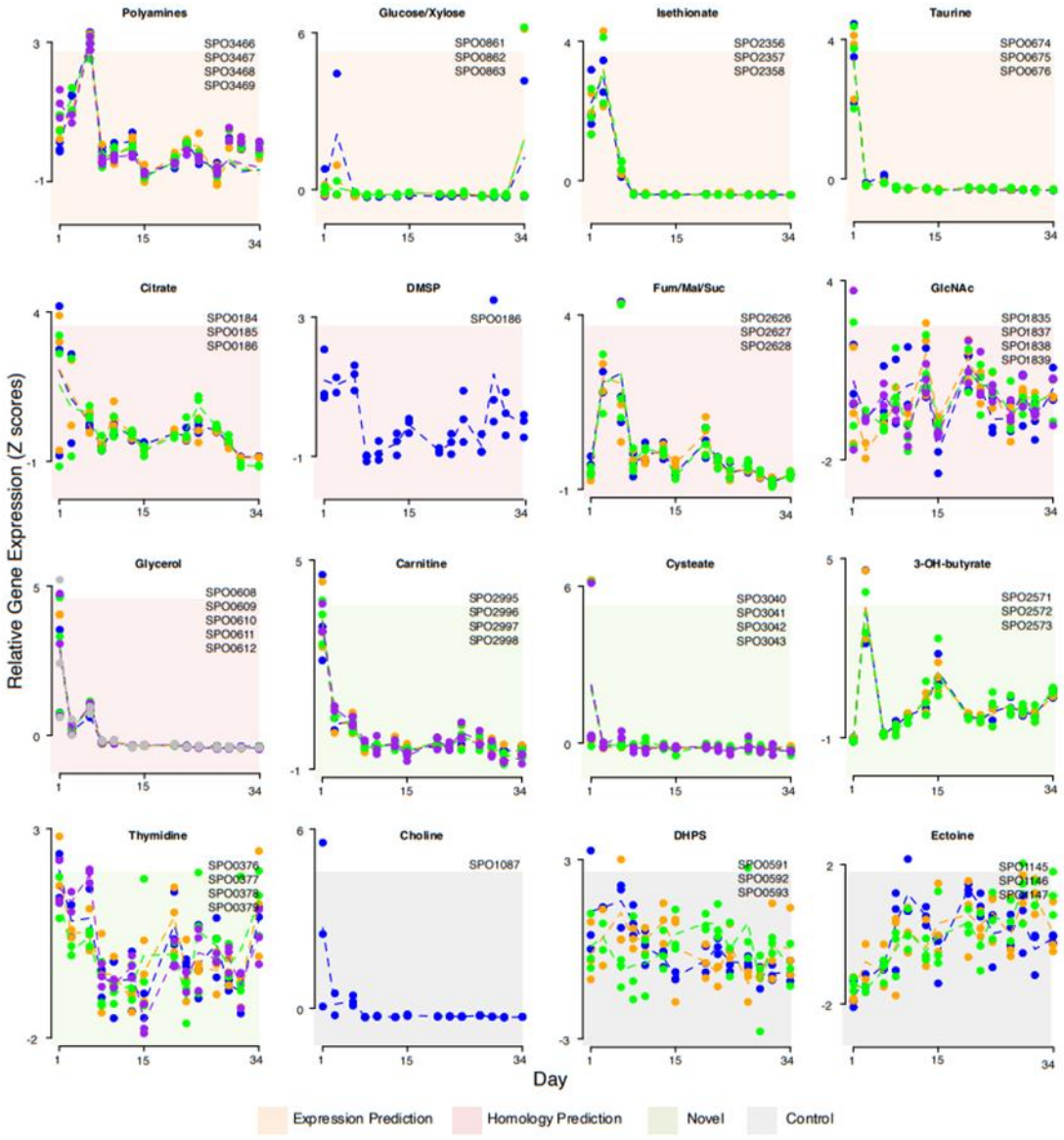


Figure S2.3. Fig. S3. Relative expression of selected R. pomeroyi DSS-3 transporters in Monterey Bay seawater on each of 14 dates over 35 days in Fall 2016, normalized as Z-scores. Transporter component genes (ranging from 1 to 5) with three replicates per component are plotted individually. Lines connect the component mean expression through time.

Table 2.1. Transporter identification based on growth and metabolite drawdown screens. Column headings: Prediction, previous annotation status of the transporter: n = annotation neither known nor hypothesized; h = annotation was hypothesized based on sequence similarity; e = annotation was hypothesized based on gene expression data; c = annotation was known based on a previous *R. pomeroyi* knockout mutant. ΔOD, percent decrease in optical density of the isolated mutant relative to pooled-RB-TnSeq library with associated 95% confidence interval and p value (n=4, ANOVA with TukeyHSD). ΔDrawdown, percent decrease in drawdown by the isolated mutant relative to the pooled-RB-TnSeq library with associated 95% confidence interval and p value (n=3, ANOVA with TukeyHSD). GlcNAc, *N*-acetylglucosamine, N.S., not significant ($p \geq 0.05$). prediction method: homology(h), novel(n), control(c), expression(e).

Trans-porter	Mutant	Substrate	Pred-iction	ΔOD (%)	95% CI (% +/-)	p adj	ΔDraw-down (%)	95% CI (% +/-)	p adj
tctABC	ΔSPO0184	Citrate	h	99.1	12.4	1×10^{-6}	99.3	3.6	5×10^{-12}
nupABC	ΔSPO0378	Thymidine	n	81.1	23.5	1×10^{-5}	63.6	32.5	2×10^{-3}
	ΔSPO0379	Thymidine	n	81.5	23.5	1×10^{-5}	83.0	32.5	6×10^{-4}
hpsKLM	ΔSPO0591	DHPS	c	78.6	30.7	8×10^{-4}	64.8	19.0	1×10^{-4}
glpVSTPQ	ΔSPO0608	Glycerol	h	79.6	13.0	5×10^{-6}	78.8	2.8	2×10^{-12}
tauABC	ΔSPO0674	Taurine	e	91.8	20.6	2×10^{-6}	-3.1	34.2	N.S.
	ΔSPO0676	Taurine	e	92.8	20.6	1×10^{-6}	49.1	34.2	1×10^{-2}
xylFGH	ΔSPO0863	Glucose	e	92.1	9.7	5×10^{-7}	85.6	30.9	2×10^{-3}
	ΔSPO0863	Xylose	e	96.6	5.6	1×10^{-8}	52.8	45.5	3×10^{-2}
betT	ΔSPO1087	Choline	c	100.5	8.9	2×10^{-7}	94.1	10.8	2×10^{-5}

uehABC	ΔSPO1147	Ectoine	c	86.7	6.1	6×10^{-8}	62.1	3.8	9×10^{-3}
nagTUVW	ΔSPO1839	GlcNAc	h	78.4	7.1	3×10^{-7}	92.7	5.0	6×10^{-8}
iseKLM	ΔSPO2357	Isethionate	e	101.1	10.2	2×10^{-9}	96.0	41.9	1×10^{-3}
	ΔSPO2358	Isethionate	e	104.2	10.2	2×10^{-9}	69.8	41.9	5×10^{-3}
hbtABC	ΔSPO2573	3-OH butyrate	n	92.7	3.9	5×10^{-10}	79.8	11.7	5×10^{-5}
dctMPQ	ΔSPO2626	Fumarate	h	99.1	3.3	3×10^{-14}	98.0	7.9	2×10^{-9}
	ΔSPO2626	Malate	h	48.6	25.3	5×10^{-4}	58.1	13.6	4×10^{-6}
	ΔSPO2626	Succinate	h	92.6	11.0	4×10^{-11}	64.9	5.3	2×10^{-9}
	ΔSPO2628	Fumarate	h	96.7	3.3	3×10^{-14}	95.0	7.9	2×10^{-9}
	ΔSPO2628	Malate	h	11.2	25.3	N.S.	47.7	13.6	2×10^{-5}
	ΔSPO2628	Succinate	h	75.2	11.0	7×10^{-10}	50.8	5.3	1×10^{-8}
	ΔSPO2630	Fumarate	h	96.9	3.3	3×10^{-14}	94.9	7.9	2×10^{-9}
	ΔSPO2630	Malate	h	12.7	25.3	N.S.	62.7	13.6	2×10^{-6}
	ΔSPO2630	Succinate	h	81.8	11.0	2×10^{-10}	58.0	5.3	4×10^{-9}
cntTUVW	ΔSPO2995	Carnitine	n	48.1	25.8	1×10^{-3}	69.2	36.7	3×10^{-3}
	ΔSPO2996	Carnitine	n	61.1	25.8	3×10^{-4}	85.7	36.7	9×10^{-4}
cuyTUVW	ΔSPO3040	Cysteate	n	50.3	15.3	2×10^{-5}	38.9	30.7	2×10^{-2}
	ΔSPO3041	Cysteate	n	81.3	15.3	3×10^{-9}	40.0	30.7	2×10^{-2}
dmdT	ΔSPO3186	DMSP	h	48.9	13.4	1×10^{-4}	33.6	2.4	2×10^{-6}
potFGHI	ΔSPO3469	Cadaverine	e	62.3	7.4	8×10^{-7}	61.2	23.5	2×10^{-3}
	ΔSPO3469	Putrescine	e	68.8	10.3	3×10^{-6}	65.1	12.8	2×10^{-4}
	ΔSPO3469	Spermidine	e	63.6	16.8	8.9×10^{-5}	75.2	5.8	2.4×10^{-6}

Table S2.1. Organic compound influx transporter systems in the *Ruegeria pomeryi* DSS-3 genome. The annotations in Column C are taken from the most recent genome annotation and are not necessarily correct with regard to substrate. The annotations in Column G are experimentally determined or confirmed from this study. This table is archived at Zenodo and is publicly available (https://doi.org/10.5281/zenodo.10032011).

Table S2.2. Substrates used in high-throughput and targeted transporter screens. Ability ofsubstrates to support *R. pomeroyi* growth as a sole C source is indicated as follows: yes = supports sufficient growth for screening by OD; poor = supports growth detectable by OD but not with sufficient density for screening by OD; no = does not support growth detectable by OD. This table is archived at Zenodo and is publicly available (https://doi.org/10.5281/zenodo.10032011).

CHAPTER 3

HIGH-CONFIDENCE GLOBAL MAPPING OF BACTERIAL SUBSTRATE UPTAKE POTENTIAL[1]

[1]Schroer, WF; Schechter, M; Saunders, JK; Eren; AM; Moran, MA. To be submitted to *mSystems*

Abstract

One of the most important and most poorly addressed challenges in modern microbial ecology is the ability to assign functions to microbial genes. For many reasons, the investment required for even one successful gene annotation has remained substantial, despite major advances in the technologies that sequence these genes, genomes, transcriptomes, and proteomes from cultured and uncultured microbes in ecosystems that span the globe. Here we address the annotation of bacterial ATP-binding cassette (ABC) transporters, a large and taxonomically distributed class of transporters with which bacteria access organic matter that supports growth and fuels biogeochemical cycles. ABC transporters are multicomponent systems generally comprised of 3-5 component genes. We found that phylogenies based on the transporter's ATPase component were the most effective at predicting transporter cognate substrate. We applied this approach to ocean metagenomic data, using the EcoPhylo workflow in anvi'o to recruit sequences of the taurine transporter ATPase (*tauB*) from Tara Oceans metagenomes. While over 10,000 gene clusters were initially recruited by the KEGG *tauB* HMM our gene phylogenies were able to confidently identify a clade of 359 well supported *tauB* gene clusters. The high number of non *tauB* sequences recruited by the HMM demonstrates the limitations of common sequence similarity based gene annotation tools. The 359 confidently annotated *tauB* gene clusters were dominated by sequences sharing similarity to those of SAR11. The *recA* normalized abundance of reads that mapped to these gene clusters showed a strong polar signal. This work demonstrates how existing annotations can be leveraged to improve the quality of computationally predicted annotations, revealing ecological signal that may otherwise have been obscured.

Opening statement

WFS, MAM, and JKS conceived of the idea for this project. JKS provided WFS with instruction on appropriate software and analysis techniques for this project, designing the work for the project. WFS compiled the transporter reference database and ran phylogenetic analysis. JKS reviewed phylogenetic results, helping with interpretation and context.

MS wrote and developed the EcoPhylo pipeline with the support of AME. Discussions between WFS and MS lead to the idea of applying EcoPhylo to study bacterial uptake transporters in the ocean metagenome. MS ran the EcoPhylo pipeline for the taurine transporter ATPase with reference genomes provided by WFS. WFS included the EcoPhylo recruited gene clusters in maximum likelihood phylogenies and analyzed the recruited read data. WFS wrote the manuscript with critical editing and revision from MAM.

Introduction

The exchange of metabolites within marine microbial communities represents both a significant carbon flux (1-4) and a mechanism that shapes community structure and function (5-8). While direct observation of metabolite exchange among community members is challenging, expression-based approaches allow us to "ask the microbes" about the metabolites they encounter through analysis of transcriptome and proteome expression (9-14). Expression patterns of organic substrate transporter genes are particularly useful for identifying exchanged metabolites (12, 15-17). Serving as a cell's interface with its environment, substrate influx transporter expression generally correlates with substrate availability (18).

Gene expression approaches are powerful and scalable but depend on accurate annotations of gene function. Computational methods assign functional annotations using sequence-based homology, comparing a query gene to genes whose function has been

experimentally verified (19-21). While annotation tools have improved greatly since their introduction (21), their ability to generate useful predictions in studies of environmental microbial communities remains limited by two key factors. First, the reliance on a database of experimentally verified genes introduces taxonomic bias. Only a small fraction of microbial diversity is represented by cultured organisms (22), and only a fraction of cultured organisms have working genetic systems that enable experimental confirmation of gene function. The ten most well studied bacteria are important in human health and disease (19), and indeed gene annotations are more complete for species closely related to these strains than for environmentally relevant species that are more distant taxonomically (19, 23). Second, gene annotations assigned by computational pipelines are frequently incorrect. The most common error is overprediction in which homology-based approaches assign the function identified for a well-studied gene in the same family, though the query gene has evolved different functions (21, 24). Errors such as overprediction can vary by the method of annotation, software used, and microbial taxonomy (19, 23), and unfortunately for ecological studies, homology-based annotation of transporter proteins has a higher error rate than that of metabolic enzymes (25). Recent approaches have gone beyond sequence based homology, integrating genetic tools to improve the quality of functional annotations. The GapMind for Carbon Sources tool, for example, uses gene fitness data to fill gaps in traditional homology based annotations. However, this tool is currently limited to transport and catabolic genes for 62 substrates (25), and does not include many environmentally important marine metabolites (6).

Though there is a great diversity of cellular transport systems, importers of the ATP-binding Cassette (ABC) Superfamily are well studied and widely represented across all domains of life (26, 27). In the model bacterium *Ruegeria pomeroyi* DSS-3, 69 of its 126 organic

substrate influx transporters are the ABC type. Recent studies based on knockout mutants of the model bacterium *R. pomeroyi* experimentally verified the cognate substrates 9 ABC transporters and tracked their expression through a coastal phytoplankton bloom (28, 29). Within the ABC Superfamily, 3 evolutionarily distinct families have been identified (ABC1, ABC2, and ABC3) (30), and almost all bacterial importers belong the ABC2 family (26, 31).

As active transporter systems, ABC importers require energy from ATP to move a substrate across cellular membranes (27). These are multicomponent systems, typically consisting of two, dimerized, transmembrane permease proteins that perform the action of bringing the substrate through the cell membrane, two, also dimerized, ATPase proteins that hydrolyze ATP for energy, and one substrate binding protein that scavenges the substrate from the environment and makes it accessible to the pore of the permease (26). The ATPase component is the most evolutionarily conserved, and is the component most often used for gene phylogeny analysis (26, 30, 32), though some studies have been based on the predicted structural features of the transmembrane permease (31).

Here we use the experimentally verified annotations of *R. pomeroyi* transporters as the foundation of a targeted phylogeny-based approach to obtain high confidence annotations of bacterial transporters in the global ocean. This approach begins with a gene tree built with experimentally-annotated transporter sequences. Manual curation of the tree identified ingroup and outgroup gene clades for each of the nine verified *R. pomeroyi* transporters. Reads from ocean metagenomic data were recruited to the tree using the EcoPhylo workflow in Anvi'o (33) and those mapping to ingroups were assigned the function of the experimentally-verified genes. The focal ingroup of this study is the clade containing the *R. pomeroyi* taurine transporter because of the ubiquity of taurine synthesis by marine bacteria and phytoplankton (5, 34, 35),

and recent estimates that this metabolite supplies ~20% of bacterial and archaeal C demand, ~40% of N demand, and up to 100% of S demand in the Eastern North Atlantic and Mediterranean Sea (36, 37). This targeted phylogenetic approach allowed rapid and confident annotations of taurine transporters in the global ocean by leveraging an experimentally annotated gene from a model marine bacterium representing a taxon widespread in surface seawater.

Methods

Generation of curated reference phylogenies

The transporter classification database (TCDB) is a curated repository of transporter protein sequences that have been confidently annotated (38). Amino acid sequences of all ABC-type importer components were retrieved from TCDB (TC#s: 3.A.1.1 - 3.A.1.34; accessed November 2022). Sequences were filtered for those with comparable structure to *R. pomeroyi's* influx ABC transporters, namely 3 to 5 component genes with at least one each of a substrate binding protein (SBP), permease, and ATPase. Sequences of 9 verified *R. pomeroyi* ABC transporters (28, 29), 9 verified *Pelagibacter ubique* SAR11 transporters (39), and 2 verified transporters from the GapMind database (25) were added to this transporters database, referred to as the reference database (Table S3.1).

For building phylogenetic trees, the reference database was augmented with 60 additional, yet unconfirmed, *R. pomeroyi* transporters. All phylogenetic tree building and alignment software was run on the University of Georgia's GACRC Sapelo2 HPC cluster (Athens, GA, USA). Amino acid sequences were aligned using CLUSTAL OMEGA v1.2.4 (40). Preliminary phylogenetic trees were generated for the ATPase component with amino acid sequences using FastTree 2.1.11 (41). To reduce downstream computational demand, sequences in deeply branching monophyletic clades that did not contain any *R. pomeroyi* transporters were

selectively removed, leaving were 3-5 sequences to represent each lineage in subsequent phylogenies (Table S3.1, S3.2). Final trees were generated using RAxML 8.2.11 (42) using the PROCATLG model; 20 maximum likelihood trees were generated, and bootstrapping (n=100) was performed. Trees were built for each of the three components of the ABC transporters. Visualization was performed in R version 4.2.2 using the ape package (43).

Clades containing verified *R. pomeroyi* transporters were manually examined to assess the support for each clade and the annotations of transporters in neighboring clades. Well supported ingroup clades containing verified transporters of the same substrate from multiple taxa with distinct outgroups were identified for use in targeted phylogeny-based annotation (44).

Read recruitment

To evaluate the ecological and phylogenetic breadth of taurine ATPases across the surface ocean, we ran the anvi'o 'EcoPhylo' workflow (https://anvio.org/m/ecophylo) using Tara Oceans project data (11, 33). Briefly, the EcoPhylo workflow: 1) recovered sequences of putative taurine ATPases from the surface ocean metagenomic samples using the KEGG Orthology (45) HMM K10831 (e-value $<1x10^{-12}$), 2) removed low-quality HMM hits with < 80% alignment coverage, 3) clustered resulting sequences at 94% ANI and identified a representative sequence from each gene cluster using mmseqs2 (46), 4) used the gene cluster representative to calculate a FastTree neighbor joining phylogenetic tree (41) and recruit metagenomic reads (47), and 5) integrated results into the anvi'o interactive interface (48).

Read abundance of *recA* genes as percent of mapped reads in TARA samples was determined from the Ocean Gene Atlas (OGA) "OM-RGCv2+G" database (49) using the HMM PF00154 (e-value $<1x10^{-40}$). Total read counts for each sample were retrieved from Salazar et al.

(11) and used to calculate per sample *recA* coverage. The *recA* coverage was used to normalize *tauB* coverage in TARA samples.

Targeted phylogeny based annotation

The gene phylogeny produced by EcoPhylo contained a clade of 359 recruited sequences, all experimentally verified reference *tauB* genes, and no confirmed outgroup sequences. Sequences from this clade and the two most closely related outgroup clades were extracted. A maximum likelihood phylogeny was built from these sequences along with the curated reference ATPase sequences and 60 additional *R. pomeroyi* ATPase sequences with RAxML 8.2.11 using the parameters provided above.

Results and Discussion

ATPase phylogeny consistently and accurately predicts cognate substrate

A reference list of 216 well-characterized ABC transporter sequences was assembled (Table S3.1) from TCDB (38), GapMind (25), the genome of *Pelagibacter ubique* (9 experimentally verified ABC transporters; (39)), and the genome of *R. pomeroyi* (9 experimentally verified ABC transporters; (28, 29)). This reference transporter dataset represents a broad diversity of bacterial and archaeal ABC influx transporters. Sixty additional uncharacterized *R. pomeroyi* transporters (those without an experimentally verified substrate) (Table S3.1) were also included in the phylogenetic analysis to assist in identification of paraphyletic clades. A comparison of maximum likelihood gene phylogenies made for each transporter component (ATPase, permease, SBP) (Fig. 3.1, S3.1) showed that ATPase phylogenies made the most consistent and accurate predictions of cognate substrate (Fig. 3.1, S3.1, S3.2), which we defined as a well-supported monophyletic clade in which all reference transporters have the same confirmed substrate (44). Among clades containing one of the 9

verified *R. pomeroyi* transporters, ATPase phylogenies were monophyletic for the transporters of taurine (*tauB*), glycerol, N-acetylglucosamine (GlcNAc), and thymidine (Fig. 3.1, S3.1, S3.2). For taurine, the permease (*tauC*) phylogeny was also monophyletic, but the SBP (*tauA*) phylogeny placed *R. pomeroyi*'s transporter on a deeply divergent lineage (Fig. 3.1). For glycerol and GlcNAc, all three transporter components formed monophyletic clades (Fig. S3.1,S3.2). For thymidine and other nucleosides (Fig. S3.1,S3.2), permease and SBP trees were paraphyletic. In the case of the verified transporters for glucose and polyamine, ATPase phylogenies clustered uncharacterized *R. pomeroyi* transporters with the verified transporters. Both of these transporters are capable of taking up multiple substrates (the glucose transporter also takes up xylose, the polyamine transporter takes up putrescine, spermidine, and cadaverine) (28). Additionally, R. pomeroyi likely has redundant transporters for both groups of substrates as well (28), suggesting promiscuity and redundancy could be complicating factors in homology-based clustering. Monophyletic clades could also not be identified for the transporters of DMSP, cysteate, and carnitine, but in this case it was due to the lack of other characterized transporters of these substrates in the reference database. Additional experimental transporter verification is necessary to improve the context for computation-based functional annotation.

The superior performance of ATPase sequences in phylogenetic analysis was unanticipated because this protein component does not directly interact with the substrate (27). Nonetheless, the ATPase component is characterized by highly conserved motifs and has been used previously to study ABC transporter diversity and evolution (26, 30, 32, 50, 51). The component genes of ABC transporter systems are under stabilizing selection, with loss-of-function mutations in any component able to disrupt the transporter function (52-54). This selective pressure leads to overall conservation of ABC transporter sequences and function,

although each component accumulates neutral mutations independently and at differing rates. A study using random mutagenesis of the E. coli maltose SBP found that mutations across several regions of the gene caused no defect of function (53), indicating the susceptibility of these genes to accumulate neutral mutations and remain functional. The ATPase's conserved tertiary structure causes it to accumulate neutral mutations at a lower rate than the permease or SBP components (55). For the permease component, however, the transmembrane pore architecture must be compatible with the cognate substrate promoting sequence conservation between transporters of the same or similar substrates (55). The differential rates of mutation accumulation are clear in a comparison of *R. pomeroyi* and *E. coli* taurine transporter sequences: *tauB* and *tauC* amino acid sequences share 43% and 44% identity, while *tauA* components are only 23% identical. The higher rate of mutation accumulation in the SBP component adds noise to sequence-based phylogeny, decreasing its power.

KEGG HMM recruitment of Tara reads overpredicts tauB sequences

We applied our phylogenetic annotation approach to assess the abundance and distribution of taurine transporters in the Tara oceans metagenomes. Using the EcoPhylo workflow in anvi'o (48), assembled reads from 90 epipelagic ocean Tara Oceans metagenomes (11, 33) and 13 reference genomes (Table S3.3) were recruited to the KEGG HMM (K10831, cutoff $e < 10^{-12}$) for the taurine transporter ATPase (*tauB*). The 31,032 recruited sequences clustered into 10,852 gene clusters (>94% ANI). The gene phylogeny provided by the EcoPhylo pipeline (FastTree) identified a well-defined clade that included 359 TARA gene clusters, all four experimentally verified *tauB* sequences from the reference genomes, and none of the verified outgroups (Fig S3.3). Representative sequences from each gene cluster in the *tauB* clade along with those from gene clusters in the two most closely related outgroup clades were

selected for a more stringent maximum likelihood phylogenetic analysis with the ATPase reference database sequences and 60 *R. pomeroyi* ATPase sequences. The maximum likelihood phylogeny confirmed the placement of representative sequences from the 359 Tara gene clusters within a monophyletic taurine clade, allowing confident annotation as taurine transporter ATPases (Fig. 3.2). Gene cluster sequences in the outgroups were recruited into other clades (Fig. 2).

A FastTree neighbor joining phylogeny including the remaining 97% (10,493) of gene clusters that recruited to the *tauB* HMM with the (cutoff $e < 10^{-12}$) but placed outside the *tauB* clade were distributed across all other lineages of the reference ATPase tree (Fig. S3.4). Although the performance of only a single KO HMM was analyzed here, it illustrates limitations of a standard tool used for bacterial gene annotation to accurately annotate transporter cognate substrates. The phylogeny based approach used here requires additional expertise and is more computationally intensive, but was able to confidently distinguish between *tauB* genes and many thousands of other ATPase sequences from transporters targeting other substrates.

Taurine transporter abundance has a strong polar signal

Taurine is an amino acid derivative produced by a wide variety of marine organisms including phytoplankton (5, 34, 35), bacterioplankton (34, 56), fish (57), and invertebrates (58-61), and high rates taurine efflux from marine invertebrates has been well documented (58-62). Recent field studies in the North Atlantic and Adriatic Sea found that taurine constitutes a major flux of C, N and S, although variable by region and season (36, 37, 62). Given the importance of taurine's carbon in ocean biogeochemical cycles, we identified patterns in the distribution of the taurine ABC transporter across the global oceans. Reads from the 90 Tara samples were mapped to the 359 *tauB* gene cluster representative sequences. For each gene cluster, coverage of

recruited reads was normalized to the coverage of the single copy core gene *recA* (63) to allow per cell estimates of gene abundance. Across samples, normalized *tauB* abundance ranged from 17% of cells to 99%, with a mean of 44%. Polar regions were enriched in *tauB* genes compared to temperate and tropical regions (two sample T-test p=0.008, Fig. 3.3a,b), with a mean of 52% (95%CI: ± 8%) relative to the non-polar mean of 41% (95%CI: ± 2%). The polar samples were characterized by lower sequence diversity relative to non-polar samples (mean of 199 out of 359 gene clusters detected per sample compared to 256 gene clusters per sample for non-polar samples), but had greater read coverage per gene cluster (Fig. 3.3c,d). Gene clusters with the greatest number of mapped reads across all Tara samples were also significantly enriched in polar samples (Fig. 3.4a, Table S3.4)

Each Tara *tauB* sequence was queried against NCBI's clustered nr database using BLAST and taxonomic information retrieved for the top hit (Table S3.4). Across the representative sequences for the 359 *tauB* gene clusters, 233 hit to members of the SAR11 clade (*Candidatus Pelagibacterales*), along with 56 clusters that hit to *Rhodobacterales*, and, represented by 20 or fewer clusters, were unclassified *Alphaproteobacteria* and *Gammaproteobacteria*, *Thiotrichales*, *Pseudomonadales*, and others (Fig. 3.4). The representative sequence with the greatest number of mapped reads across all 90 Tara samples also had high similarity to SAR11 (Fig. 3.5a). SAR11 is an extreme oligotroph with a highly streamlined genome, making it dependent on external sources of several metabolites, one of which is taurine (64). The SAR11 metabolic requirement for taurine uptake paired with the high numerical abundance of this taxon, accounting for up to 50% of surface ocean heterotrophic bacteria (65), explains its dominance in the *tauB* sequence pool.

In marine microbes, taurine functions as an osmolyte, ecological signal, and cryoprotectant (5, 56, 66-68). The latter role in cryoprotection could contribute to the polar bias seen in the frequency of heterotrophic bacterial genomes that carry *tauB*. Further, taurine is a source of reduced C, N, and S for bacterioplankton (36, 37). Given the challenge of directly measuring metabolite exchange among microbes, rigorous identification of bacterial transporter genes in metagenomic and metatranscriptomic datasets provides insights into the metabolites frequently encountered in their environment. Taurine has been identified as a phytoplankton exometabolite and endometabolite (69-71), including in an arctic diatom growing at *in-situ* temperature and light conditions (69). The broad distribution of bacterial taurine transporters suggests its importance to phycosphere interactions.

Conclusion

Gene annotation remains a limiting factor for the advancement of biology in the 'omics age. Here, we demonstrate the efficacy of targeted phylogenetic approaches for identifying transporter cognate substrates in environmental systems. This approach, however, was only made possible because of the experimental verification of transporters in relevant marine bacteria, here specifically the annotation of *R. pomeroyi* and SAR11 transporters (28, 39). Future work to verify transporter annotations in environmental bacteria is necessary, and as new verifications become available they can be integrated into our phylogenetic annotation framework.

References

1. Azam F, Fenchel T, Field J, Grey J, Meyer-Reil L, Thingstad F. The ecological role of water-column microbes. Mar Ecol Prog Ser. 1983;10:257-63.

2. Anderson TR, Ducklow HW. Microbial loop carbon cycling in ocean environments studied using a simple steady-state model. Aquat Microb Ecol. 2001;26:37-49.

3. Moran MA, Ferrer-González FX, Fu H, Nowinski B, Olofsson M, Powers MA, et al. The ocean's labile DOC supply chain. Limnol Oceanogr. 2022;67:1007-21.

4. Cole JJ, Likens GE, Strayer DL. Photosynthetically produced dissolved organic carbon: An important carbon source for planktonic bacteria. Limnol Oceanogr. 1982;27:1080-90.

5. Amin S, Hmelo L, Van Tol H, Durham B, Carlson L, Heal K, et al. Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. Nature. 2015;522:98-101.

6. Moran MA, Kujawinski EB, Schroer WF, Amin SA, Bates NR, Bertrand EM, et al. Microbial metabolites in the marine carbon cycle. Nat Microbiol. 2022;7:508-23.

7. Edwards BR, Bidle KD, Van Mooy BA. Dose-dependent regulation of microbial activity on sinking particles by polyunsaturated aldehydes: Implications for the carbon cycle. Proc Natl Acad Sci USA. 2015;112:5909-14.

8. 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. Proc Natl Acad Sci USA. 2020;117:27445-55.

9. Helbling DE, Ackermann M, Fenner K, Kohler H-PE, Johnson DR. The activity level of a microbial community function can be predicted from its metatranscriptome. ISME J. 2012;6:902-4.

10. Ottesen EA, Young CR, Gifford SM, Eppley JM, Marin III R, Schuster SC, et al. Multispecies diel transcriptional oscillations in open ocean heterotrophic bacterial assemblages. Science. 2014;345:207-12.

11. Salazar G, Paoli L, Alberti A, Huerta-Cepas J, Ruscheweyh H-J, Cuenca M, et al. Gene expression changes and community turnover differentially shape the global ocean metatranscriptome. Cell. 2019;179:1068-83.

12. Ferrer-González FX, Widner B, Holderman NR, Glushka J, Edison AS, Kujawinski EB, et al. Resource partitioning of phytoplankton metabolites that support bacterial heterotrophy. ISME J. 2021;15:762-73.

13. Nowinski B, Moran MA. Niche dimensions of a marine bacterium are identified using invasion studies in coastal seawater. Nat Microbiol. 2021;6:524-32.

14. Landa M, Burns AS, Roth SJ, Moran MA. Bacterial transcriptome remodeling during sequential co-culture with a marine dinoflagellate and diatom. ISME J. 2017;11:2677-90.

15. Teeling H, Fuchs BM, Becher D, Klockow C, Gardebrecht A, Bennke CM, et al. Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. Science. 2012;336:608-11.

16. Lidbury I, Murrell JC, Chen Y. Trimethylamine N-oxide metabolism by abundant marine heterotrophic bacteria. Proc Natl Acad Sci USA. 2014;111:2710-5.

17. Tang K, Jiao N, Liu K, Zhang Y, Li S. Distribution and functions of TonB-dependent transporters in marine bacteria and environments: implications for dissolved organic matter utilization. PloS One. 2012;7:e41204.

18. Weiten A, Kalvelage K, Neumann-Schaal M, Buschen R, Scheve S, Winklhofer M, et al. Nanomolar Responsiveness of Marine *Phaeobacter inhibens* DSM 17395 toward Carbohydrates and Amino Acids. Microb Physiol. 2022;32:108-21.

19. Lobb B, Tremblay BJ-M, Moreno-Hagelsieb G, Doxey AC. An assessment of genome annotation coverage across the bacterial tree of life. Microb Genom. 2020;6:e000341.

20. Loewenstein Y, Raimondo D, Redfern OC, Watson J, Frishman D, Linial M, et al. Protein function annotation by homology-based inference. Genome Biol. 2009;10:1-8.

21. Radivojac P, Clark WT, Oron TR, Schnoes AM, Wittkop T, Sokolov A, et al. A large-scale evaluation of computational protein function prediction. Nat Methods. 2013;10:221-7.

22. Lloyd KG, Steen AD, Ladau J, Yin J, Crosby L. Phylogenetically novel uncultured microbial cells dominate earth microbiomes. MSystems. 2018;3:10.1128/msystems. 00055-18.

23. Dimonaco NJ, Aubrey W, Kenobi K, Clare A, Creevey CJ. No one tool to rule them all: prokaryotic gene prediction tool annotations are highly dependent on the organism of study. Bioinformatics. 2022;38:1198-207.

24. Schnoes AM, Brown SD, Dodevski I, Babbitt PC. Annotation error in public databases: misannotation of molecular function in enzyme superfamilies. PLoS Comput Biol. 2009;5:e1000605.

25. Price MN, Deutschbauer AM, Arkin AP. Filling gaps in bacterial catabolic pathways with computation and high-throughput genetics. PLoS Genet. 2022;18:e1010156.

26. Davidson AL, Dassa E, Orelle C, Chen J. Structure, function, and evolution of bacterial ATP-binding cassette systems. Microbiol Mol Biol Rev. 2008;72:317-64.

27. Higgins CF. ABC transporters: physiology, structure and mechanism–an overview. Res Microbiol. 2001;152:205-10.

28. Schroer WF, Kepner HE, Uchimiya M, Mejia C, Rodriguez LT, Reisch CR, et al. Functional annotation and importance of marine bacterial transporters of plankton exometabolites. ISME Commun. 2023;3:37.

29. Li C-Y, Mausz MA, Murphy A, Zhang N, Chen X-L, Wang S-Y, et al. Ubiquitous occurrence of a dimethylsulfoniopropionate ABC transporter in abundant marine bacteria. ISME J. 2023;17:579-87.

30. Dassa E, Bouige P. The ABC of ABCs: a phylogenetic and functional classification of ABC systems in living organisms. Res Microbiol. 2001;152:211-29.

31. Wang B, Dukarevich M, Sun EI, Yen MR, Saier MH. Membrane porters of ATP-binding cassette transport systems are polyphyletic. J Membr Biol. 2009;231:1-10.

32. Saurin W, Hofnung M, Dassa E. Getting in or out: early segregation between importers and exporters in the evolution of ATP-binding cassette (ABC) transporters. J Mol Evol. 1999;48:22-41.

33. Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, et al. Structure and function of the global ocean microbiome. Science. 2015;348:1261359.

34. Durham BP, Boysen AK, Heal KR, Carlson LT, Boccamazzo R, Deodato CR, et al. Chemotaxonomic patterns in intracellular metabolites of marine microbial plankton. Front Mar Sci. 2022;9:864796.

35. Tevatia R, Allen J, Rudrappa D, White D, Clemente TE, Cerutti H, et al. The taurine biosynthetic pathway of microalgae. Algal Res. 2015;9:21-6.

36. Clifford EL, De Corte D, Amano C, Paliaga P, Ivančić I, Ortiz V, et al. Mesozooplankton taurine production and prokaryotic uptake in the northern Adriatic Sea. Limnol Oceanogr. 2020;65:2730-47.

37. Clifford EL, Varela MM, De Corte D, Bode A, Ortiz V, Herndl GJ, et al. Taurine is a major carbon and energy source for marine prokaryotes in the North Atlantic Ocean off the Iberian Peninsula. Microb Ecol. 2019;78:299-312.

38. Saier Jr MH, Reddy VS, Moreno-Hagelsieb G, Hendargo KJ, Zhang Y, Iddamsetty V, et al. The transporter classification database (TCDB): 2021 update. Nucleic Acids Res. 2021;49:D461-D7.

39. Clifton BE, Alcolombri U, Jackson CJ, Laurino P. Ultrahigh-affinity transport proteins from ubiquitous marine bacteria reveal mechanisms and global patterns of nutrient uptake. bioRxiv. 2023:2023.02. 16.528805.

40. Madeira F, Pearce M, Tivey AR, Basutkar P, Lee J, Edbali O, et al. Search and sequence analysis tools services from EMBL-EBI in 2022. Nucleic Acids Res. 2022;50:W276-W9.

41. Price MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol. 2009;26:1641-50.

42. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics. 2019;35:4453-5.

43. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics. 2019;35:526-8.

44. Saunders JK, Fuchsman CA, McKay C, Rocap G. Complete arsenic-based respiratory cycle in the marine microbial communities of pelagic oxygen-deficient zones. Proc of the National Academy of Sciences. 2019;116:9925-30.

45. Kanehisa M, Furumichi M, Sato Y, Kawashima M, Ishiguro-Watanabe M. KEGG for taxonomy-based analysis of pathways and genomes. Nucleic acids research. 2023;51:D587-D92.

46. Steinegger M, Söding J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nat Biotechnol. 2017;35:1026-8.

47. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357-9.

48. Eren AM, Kiefl E, Shaiber A, Veseli I, Miller SE, Schechter MS, et al. Community-led, integrated, reproducible multi-omics with anvi'o. Nat Microbiol. 2021;6:3-6.

49. Vernette C, Lecubin J, Sánchez P, Sunagawa S, Delmont TO, Acinas SG, et al. The Ocean Gene Atlas v2. 0: online exploration of the biogeography and phylogeny of plankton genes. Nucleic Acids Res. 2022;50:W516-W26.

50. Dean M, Annilo T. Evolution of the ATP-binding cassette (ABC) transporter superfamily in vertebrates. Annu Rev Genomics Hum Genet. 2005;6:123-42.

51. Lane TS, Rempe CS, Davitt J, Staton ME, Peng Y, Soltis DE, et al. Diversity of ABC transporter genes across the plant kingdom and their potential utility in biotechnology. BMC Biotechnol. 2016;16:1-10.

52. Davidson AL, Sharma S. Mutation of a single MalK subunit severely impairs maltose transport activity in Escherichia coli. J Bacteriol. 1997;179:5458-64.

53. Duplay P, Szmelcman S, Bedouelle H, Hofnung M. Silent and functional changes in the periplasmic maltose-binding protein of Escherichia coli K12: I. transport of maltose. J Mol Biol. 1987;194:663-73.

54. Chen J. Molecular mechanism of the Escherichia coli maltose transporter. Curr Opin Struct Biol. 2013;23:492-8.

55. Wilkens S. Structure and mechanism of ABC transporters. F1000Prime Rep. 2015;7.

56. McParland EL, Alexander H, Johnson WM. The osmolyte ties that bind: genomic insights into synthesis and breakdown of organic osmolytes in marine microbes. Front Mar Sci. 2021:732.

57. Chang Y-C, Ding S-T, Lee Y-H, Wang Y-C, Huang M-F, Liu I-H. Taurine homeostasis requires de novo synthesis via cysteine sulfinic acid decarboxylase during zebrafish early embryogenesis. Amino acids. 2013;44:615-29.

58. Allen J, Garrett M. Taurine in marine invertebrates. Adv Mar Biol. 9: Elsevier; 1971. p. 205-53.

59. Manahan DT. Amino acid fluxes to and from seawater in axenic veliger larvae of a bivalve (*Crassostrea gigas*). Mar Ecol Prog Ser. 1989;53:247-55.

60. Trapido-Rosenthal HG, Gleeson RA, Carr WE. The efflux of amino acids from the olfactory organ of the spiny lobster: biochemical measurements and physiological effects. Biol Bull. 1990;179:374-82.

61. Webb KL, Johannes R. Studies of the release of dissolved free amino acids by marine zooplankton 1. Limnol Oceanogr. 1967;12:376-82.

62. Clifford EL, Hansell DA, Varela MM, Nieto-Cid M, Herndl GJ, Sintes E. Crustacean zooplankton release copious amounts of dissolved organic matter as taurine in the ocean. Limnol Oceanogr. 2017;62:2745-58.

63. Howard EC, Henriksen JR, Buchan A, Reisch CR, Bürgmann H, Welsh R, et al. Bacterial taxa that limit sulfur flux from the ocean. Science. 2006;314:649-52.

64. Carini P, Steindler L, Beszteri S, Giovannoni SJ. Nutrient requirements for growth of the extreme oligotroph *'Candidatus Pelagibacter ubique*' HTCC1062 on a defined medium. ISME J. 2013;7:592-602.

65. Morris RM, Rappé MS, Connon SA, Vergin KL, Siebold WA, Carlson CA, et al. SAR11 clade dominates ocean surface bacterioplankton communities. Nature. 2002;420:806-10.

66. Yancey PH. Organic osmolytes as compatible, metabolic and counteracting cytoprotectants in high osmolarity and other stresses. J Exp Biol. 2005;208:2819-30.

67. Huxtable R. Physiological actions of taurine. Physiol Rev. 1992;72:101-63.

68. Ripps H, Shen W. taurine: a “very essential” amino acid. Mol Vis. 2012;18:2673.

69. Dawson HM, Heal KR, Boysen AK, Carlson LT, Ingalls AE, Young JN. Potential of temperature-and salinity-driven shifts in diatom compatible solute concentrations to impact biogeochemical cycling within sea ice. Elem Sci Anth. 2020;8:25.

70. Fiore CL, Longnecker K, Kido Soule MC, Kujawinski EB. Release of ecologically relevant metabolites by the cyanobacterium *Synechococcus elongatus* CCMP 1631. Environ Microbiol. 2015;17:3949-63.

71. Durham BP, Boysen AK, Carlson LT, Groussman RD, Heal KR, Cain KR, et al. Sulfonate-based networks between eukaryotic phytoplankton and heterotrophic bacteria in the surface ocean. Nat Microbiol. 2019;4:1706-15.

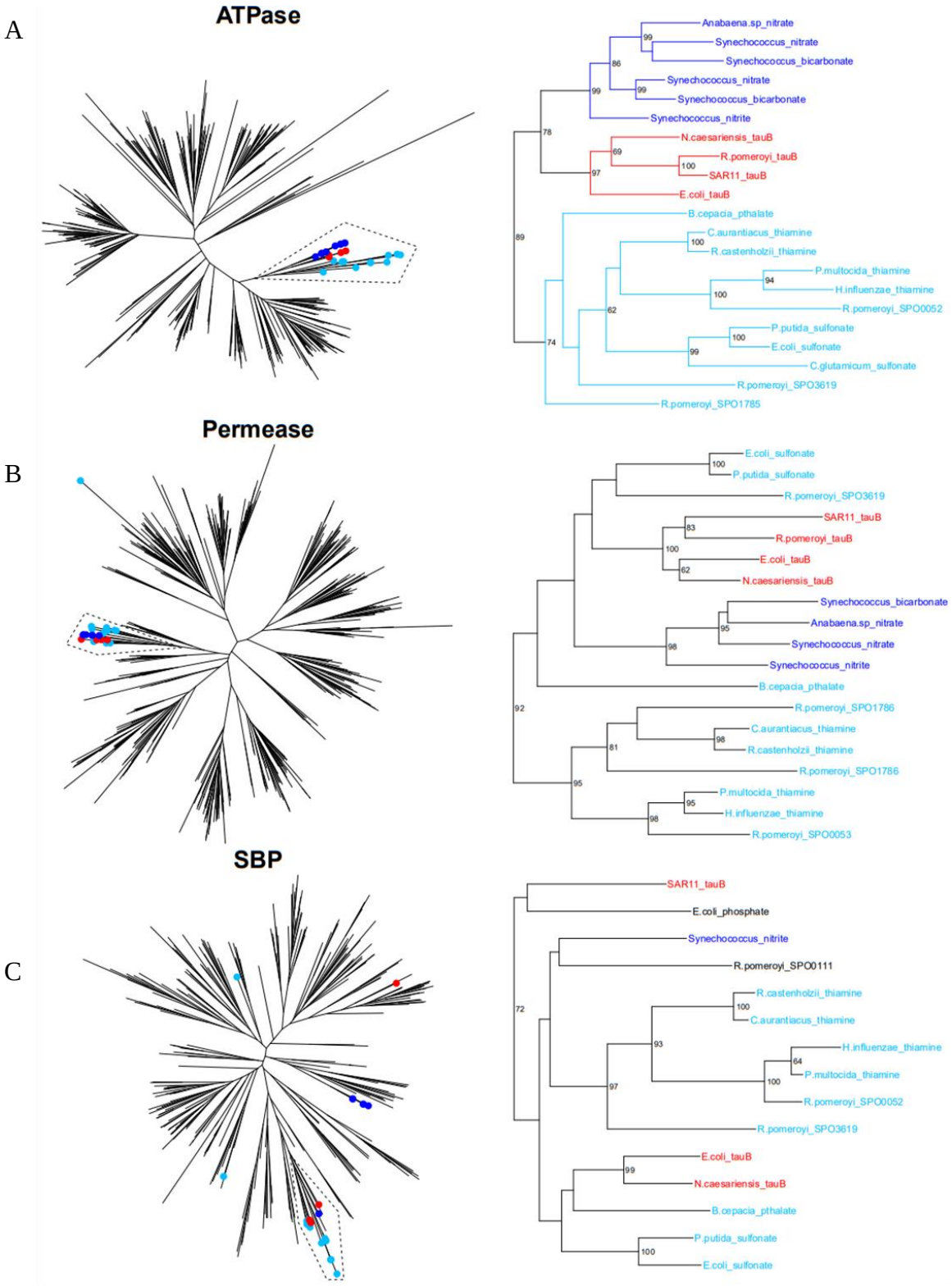


Figure 3.1. Maximum likelihood phylogenies of reference ABC transporters. Left, full trees with dashed lines indicating the clade of interest. Right, detailed view of the clade of interest,

extracted from the full tree. a) Phylogeny of the ATPase transporter component. The confirmed taurine transporter clade (red lines, text, and symbols) is well supported and distinct from two outgroup clades (dark blue/light blue, lines text, and symbols). Clade members defined here are colored consistently in panels b and c to show the differences in phylogenetic reconstruction. b) Phylogeny of the permease transporter component. The confirmed taurine clade (red text, symbols) is well defined, while light blue sequences are no longer in a monophyletic clade. c) Phylogeny of SBP transporter component. The confirmed taurine transporters (red text, symbols) are no longer in a monophyletic clade.

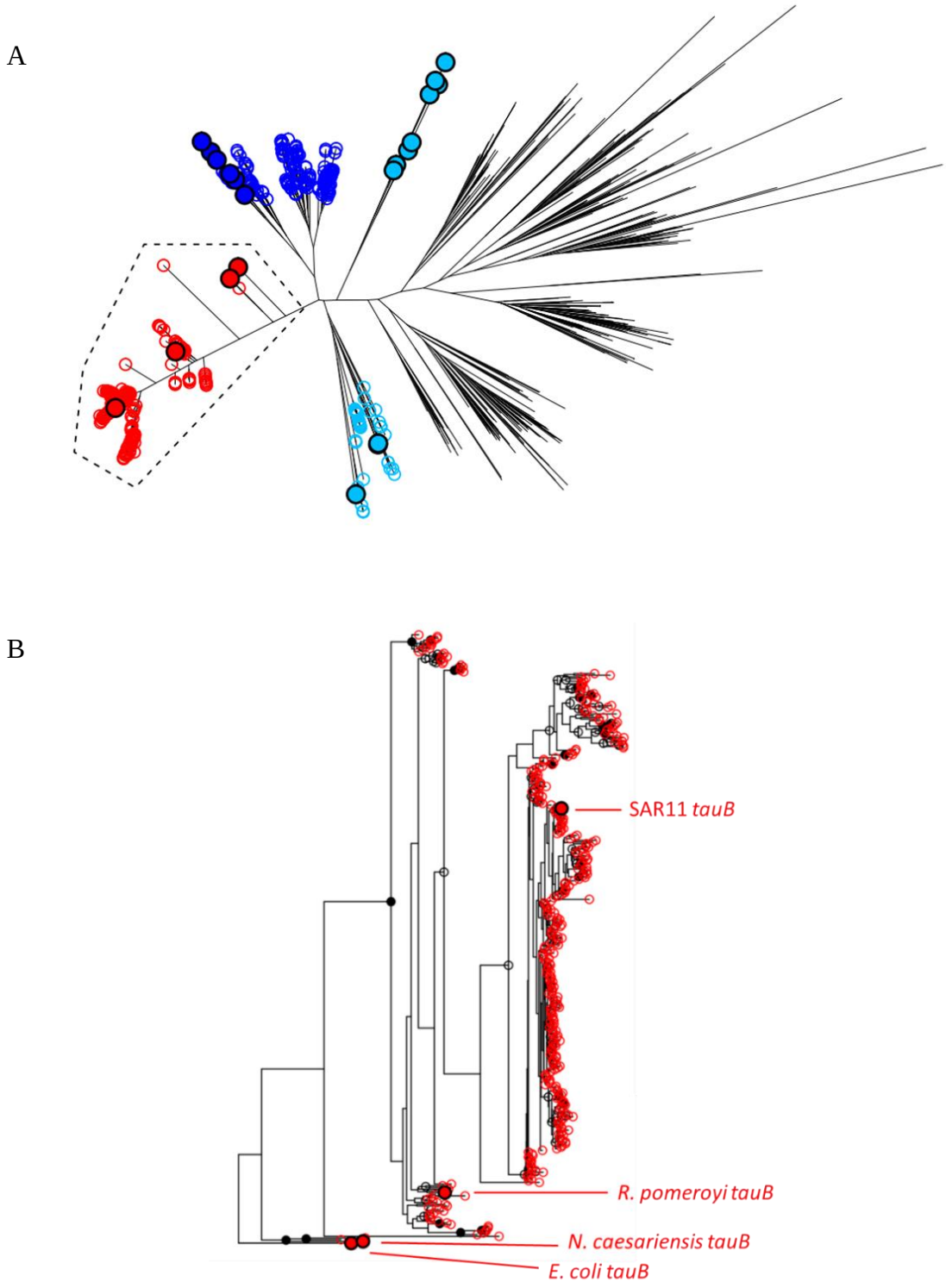


Figure 3.2. Maximum likelihood phylogeny of 341 reference ATPase transporter components, 359 *tauB* gene cluster representative sequences recruited from Tara metagenomes, and 216

outgroup sequences recruited from Tara metagenomes (open circles) by the ecophylo pipeline in anvi'o. Solid symbols with black borders indicate reference sequences of confirmed taurine transporters (red) and the two confirmed outgroup clades (dark blue, light blue) as defined in Fig 1a. a) Full unrooted phylogeny with the taurine transporter clade outlined by dashed lines. b) Detailed view of the taurine transporter clade extracted from the full tree. Open black circles at nodes indicate bootstrap values >60, solid black circles indicate bootstrap values >90.

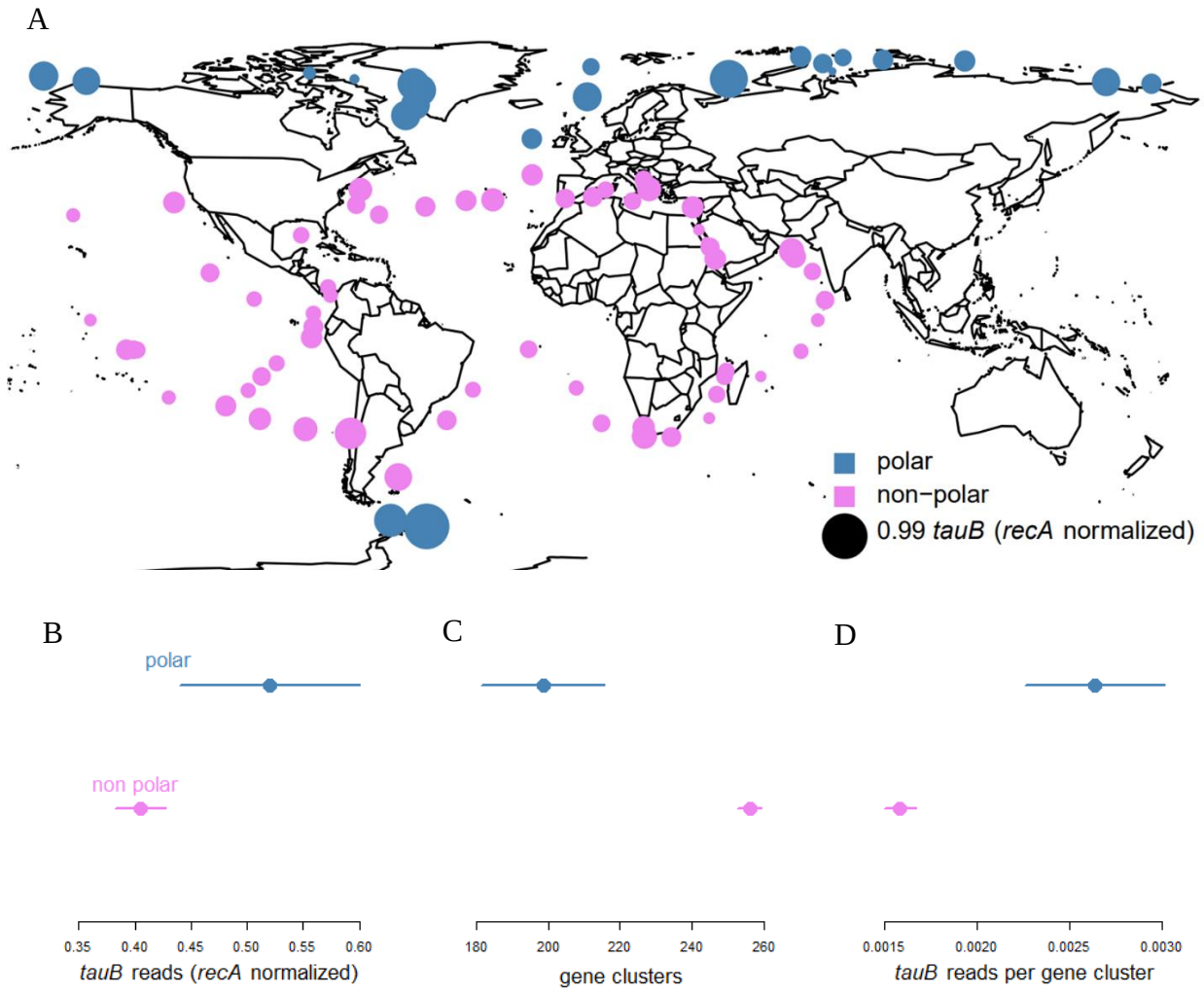


Figure 3.3. A) Map of TARA stations. Symbol size indicates *tauB* read abundance normalized to *recA* read abundance in surface ocean samples. Symbols are colored to indicate polar (blue) or non-polar (pink) sampling stations. B) Mean and 95% CI of normalized *tauB* reads for polar versus nonpolar Tara samples. Polar samples are enriched in *tauB* reads (Two sample T test, p=0.008). C) Mean and 95% CI of the of the number of *tauB* sequences detected in polar versus non-polar Tara samples. Fewer sequences were detected in polar samples. D) Mean and 95% CI of the number of *tauB* reads mapped to each detected sequence in polar and non-polar TARA samples. Polar samples had more reads mapped to each detected sequence.

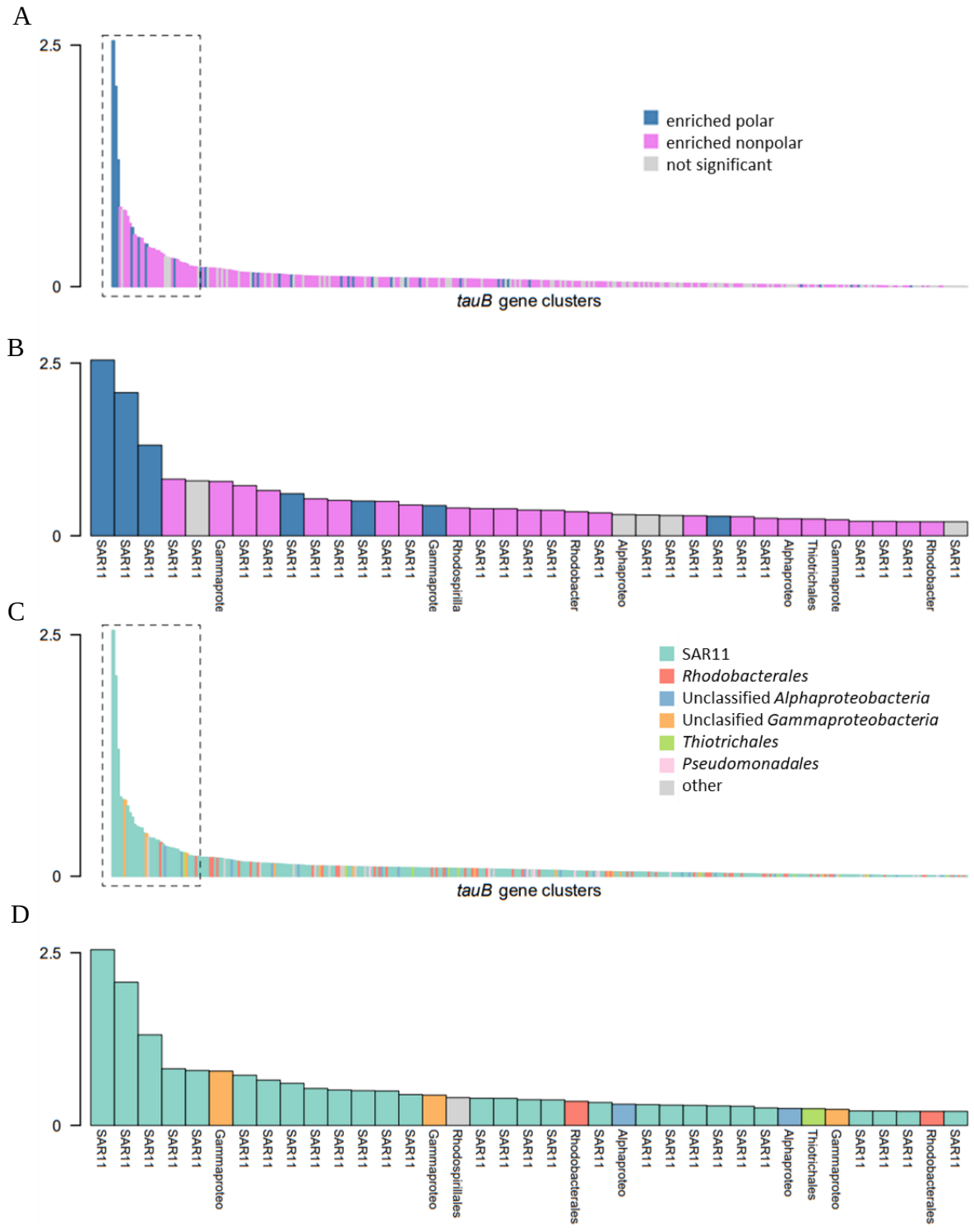


Figure 3.4. A) Total number of normalized Tara *tauB* reads mapped to each gene cluster sequence in descending abundance. Relatively few sequences are responsible for a

disproportionate number of mapped reads, with the top 10% of sequences accounting for ~50% of total reads. Colored indicates polar trend. Blue bars indicate sequences with significantly more reads in polar samples (n=31, Two sample T test p adjusted $\leq$ 0.05). Red bars indicate sequences with significantly more reads in non-polar samples (n= 208, Two sample T test p adjusted $\leq$ 0.05). Grey bars indicate no significant difference between polar and non-polar abundance. Dashes rectangle indicates extent of inset. B) Inset of panel A, depicting the top 37 most abundant gene clusters. C) Same data as A, but colored by the order of top blast hit. The best hits of the most abundant gene clusters are to SAR11 genes. Dashes rectangle indicates extent of inset. D) Inset of panel C, depicting the top 37 most abundant gene clusters.

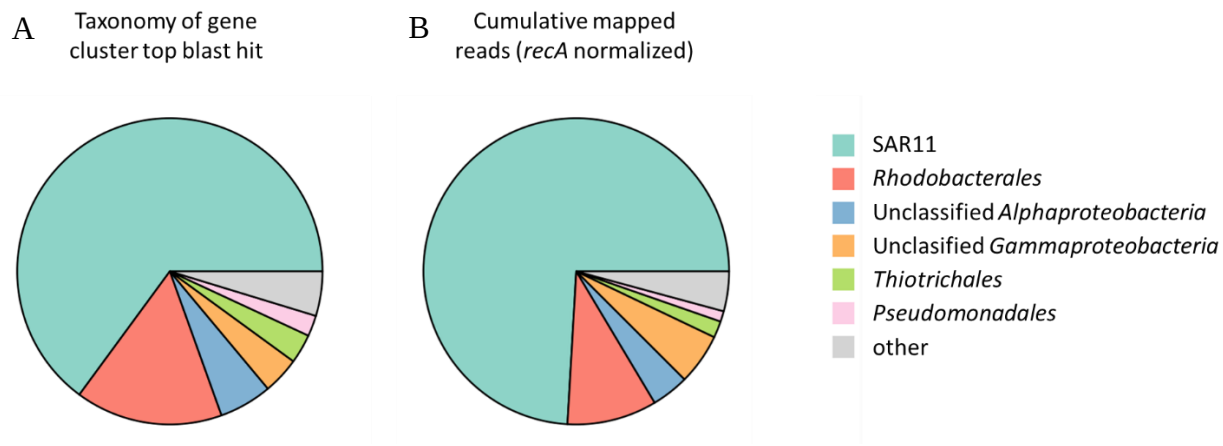


Figure 3.5. Each Tara gene cluster representative sequence was blasted against NCBI's clustered nr database, and taxonomic information was collected from the top blast hit of each sequence. A) Order level taxonomy of top blast hits, SAR11 is the dominant order (total N=359 sequences). B) Total normalized reads mapped to sequences in each order. SAR11 is the dominant order.

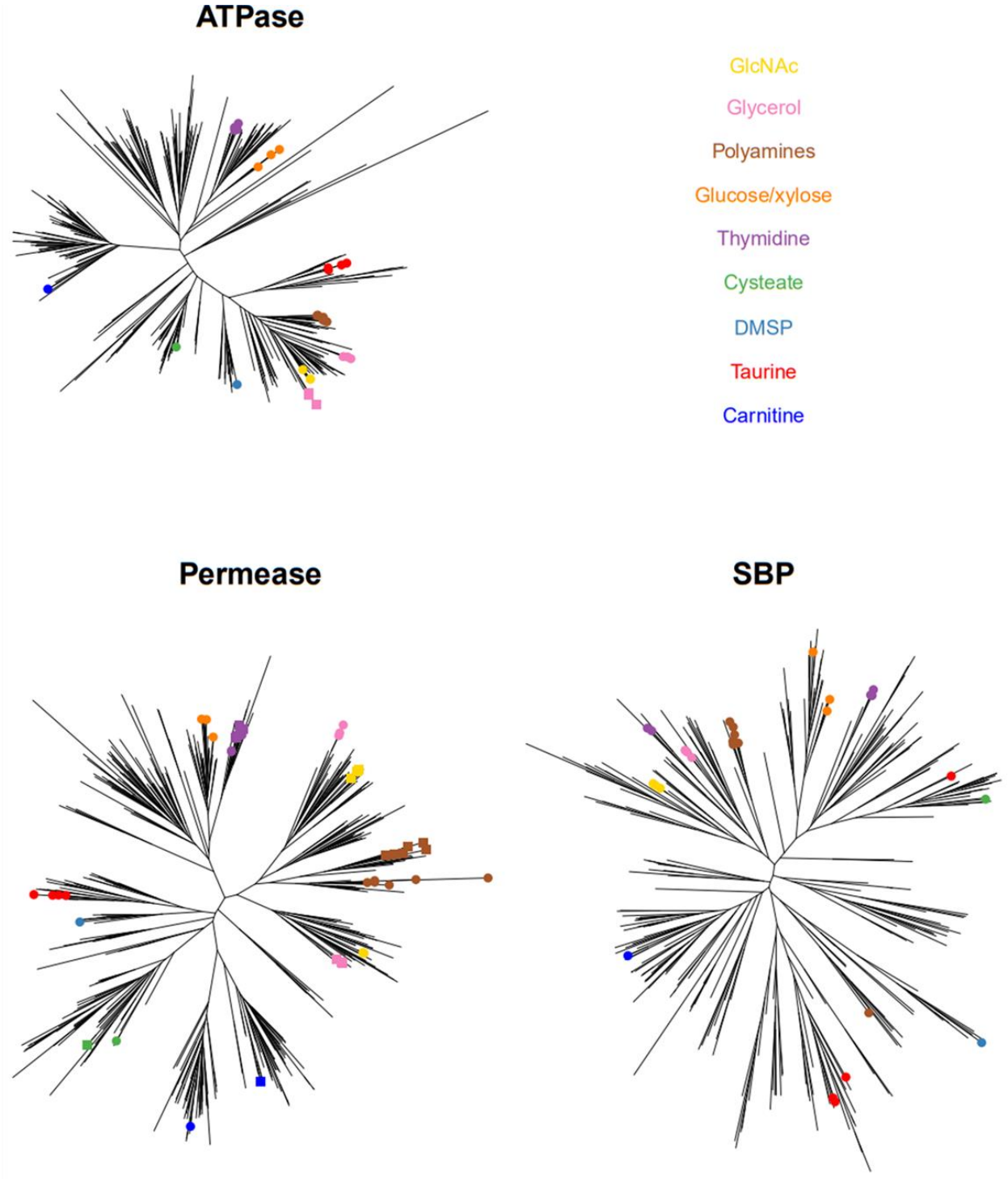


Figure S3.1. Maximum likelihood phylogenies of curated reference ABC transporters. Points indicate the location *R. pomeroyi* transporters of 9 substrates as well as other confirmed transporters of the same substrates.

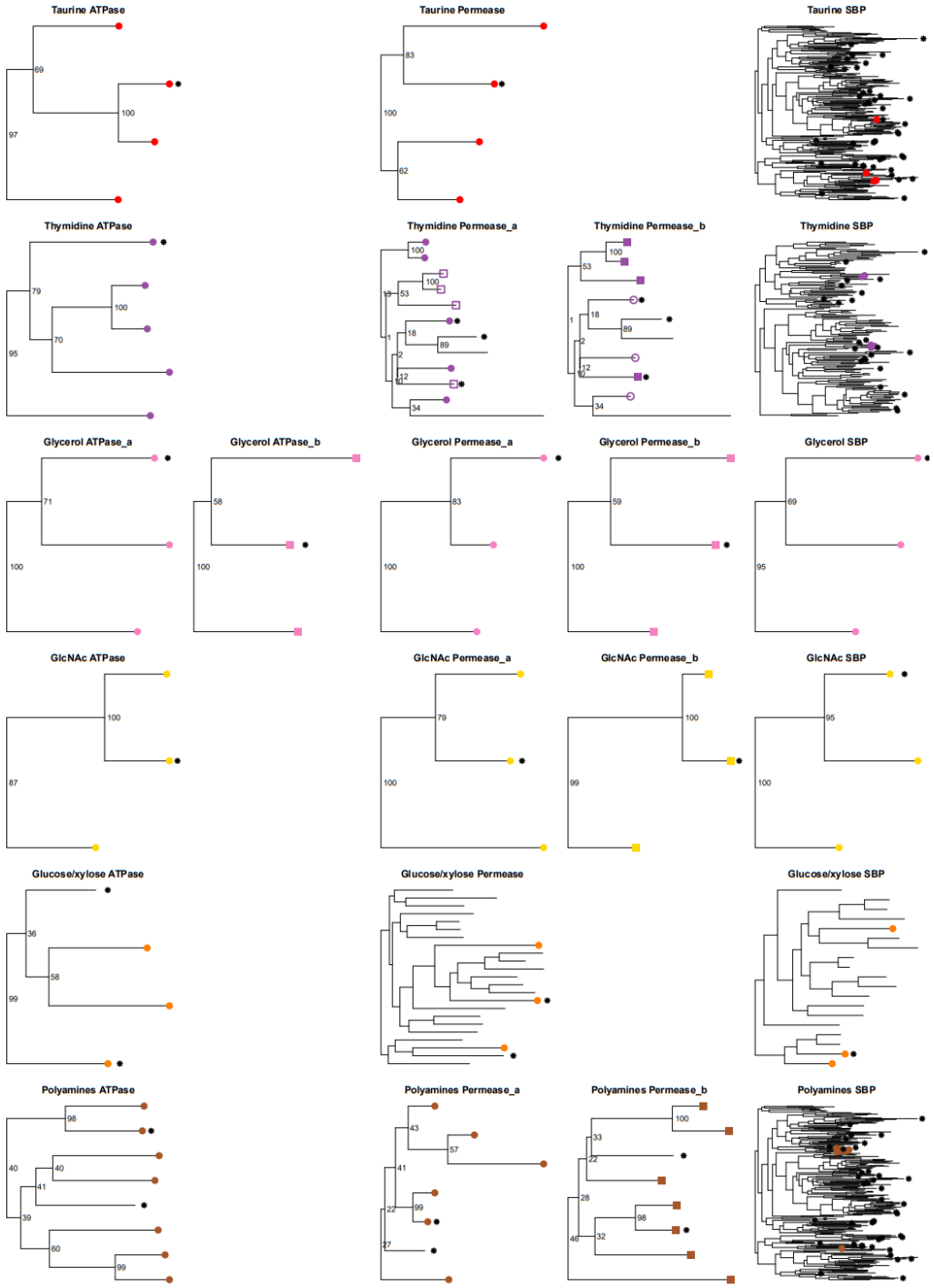


Figure S3.2. Clades extracted from the maximum likelihood phylogenies of curated reference ABC transporters depicted in Fig S1. Each clade extracted from the most recent common ancestor of all verified transporters of a given substrate. Asterisk indicates the position of an R. pomeroyi transporter.

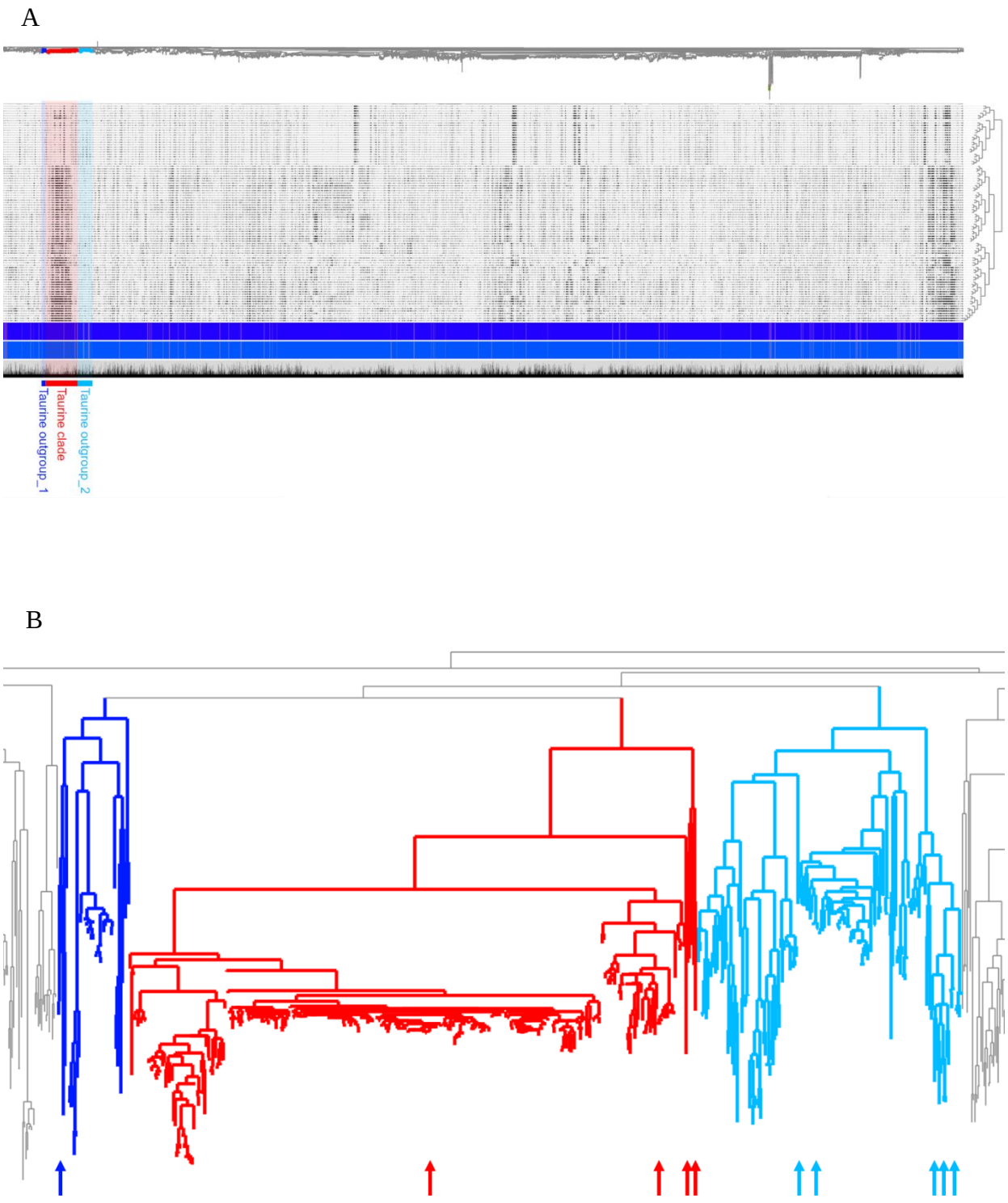


Figure S3.3. A) EcoPhylo output visualized in Anvio Interactive. Rows are recruited representative sequences. Columns are TARA metagenome samples. Red highlighted section is

the putative tauB clade, light and dark blue highlighted sections are outgroup clades. B) Expanded view of the EcoPhylo generated phylogeny (FastTree). Red clade is comprised of putative tauB sequences, light and dark blue highlighted clades are outgroups. Arrow indicate the location of reference sequences of confirmed tauB genes (red) and confirmed outgroups (light and dark blue).

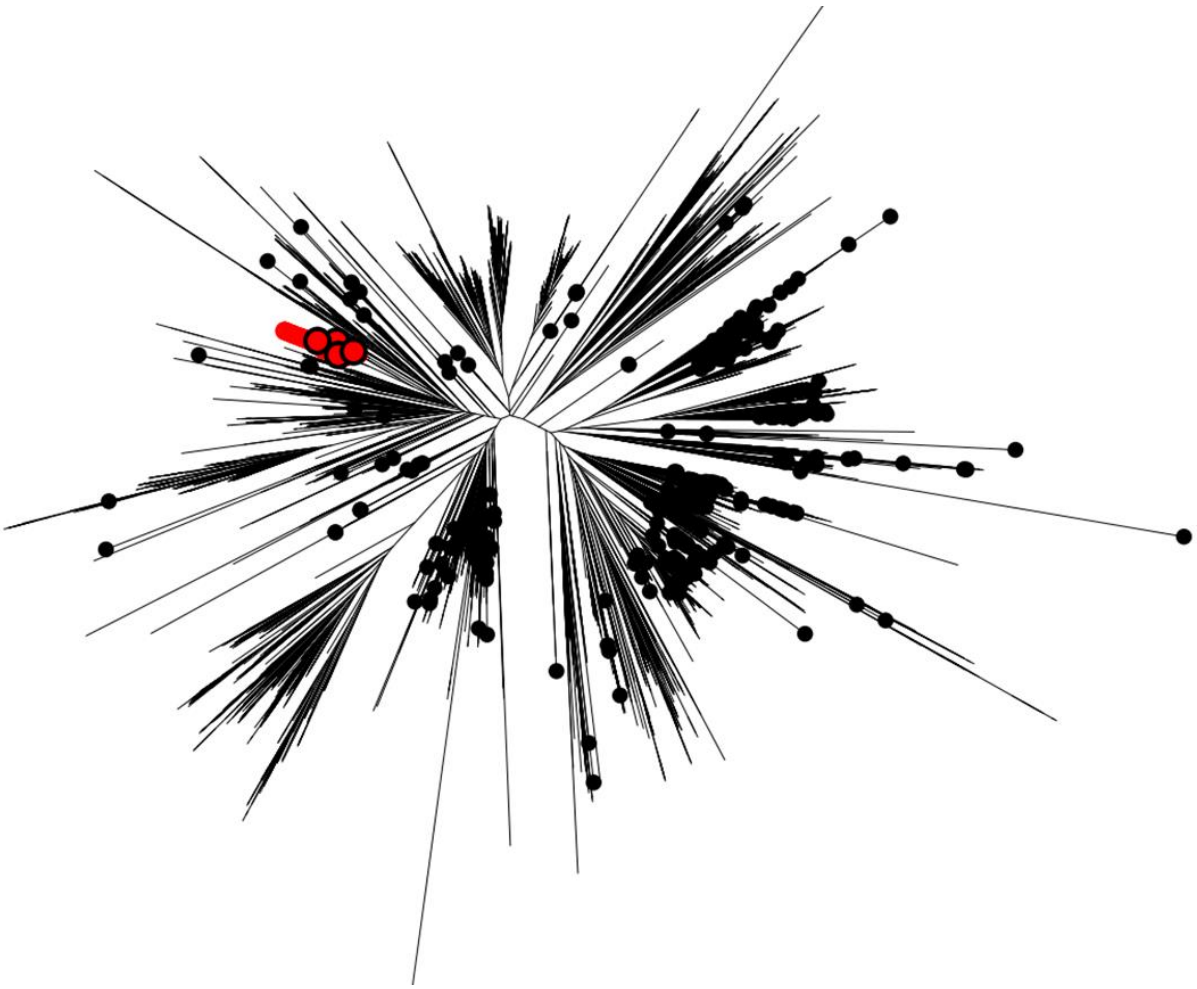


Figure S3.4. FastTree neighbor joining phylogeny of the reference transporter database, all R pomeroyi transporters, 10,852 sequences recruited from TARA oceans metagenomes to the KEGG Ortholog taurine transport system ATPase (K10831) by the ecophylo pipeline in Anvio. Red symbols indicate the location of tauB sequences. Black symbols indicate the location of non-taurine transporters from the reference database and *R. pomeroyi*. Recruited TARA sequences (no tip label) have been placed in all transporter lineages.

Suplemental Table S3.1. Transporter reference database plus R. pomeroyi ABC transporters: components, IDs, names, sources, sequences. This table is archived at Zenodo and is publicly available (https://doi.org/10.5281/zenodo.10032011).

Suplemental Table S3.2. TCDB ABC transporter components not included in the reference database and reason for exclusion, IDs, names, sequences. This table is archived at Zenodo and is publicly available (https://doi.org/10.5281/zenodo.10032011).

Suplemental Table S3.3. Reference Genomes included in the EcoPhylo analysis of Tara metagenomes. This table is archived at Zenodo and is publicly available (https://doi.org/10.5281/zenodo.10032011).

Suplemental Table S3.4. Information on tauB gene clusters from Tara metagenomes. Gene cluster ID, representative sequence, top blast hit results and taxonomy. Identification of gene clusters with polar trend. Normalized reads recruited to each gene cluster from each Tara metagenome. This table is archived at Zenodo and is publicly available (https://doi.org/10.5281/zenodo.10032011).

CHAPTER 4

CHARACTERIZING A METERED AUXIN RECOGNITION CIRCUIT IN MARINE PHYCOSPHERE BACTERIA[1]

[1]Schroer, WF; Uchimiya, M; Hamilton, MM; Smith, CR; Olofsson, M; Moran, MA. To be submitted to *ISME J*.

Abstract

Interactions within marine microbial communities in the surface ocean shape the rate and fate of primary production, some of which have been linked to the exchange of the auxin type phytohormone indole acetic acid (IAA). The roles of IAA have been well studied in plant - microbe interactions where it can facilitate both mutualistic and pathogenic interactions, but its roles in marine phytoplankton-bacteria interactions are less well understood. Here we characterize the function of *iaaH*, a key gene in IAA biosynthesis, in the model marine bacterium *Ruegeria pomeroyi* DSS-3. We demonstrate that *R. pomeroyi*, although its IAA biosynthesis pathway is incomplete, can produce IAA when supplemented with the intermediate indole-3-acetamide (IAM). Unexpectedly, when IAM is provided as a sole carbon source, *R. pomeroyi* takes up IAM, converts it to IAA, and releases the IAA outside the cell. This occurs in an exact 1:1 stoichiometric ratio with the supplied IAM, and thus bacterial release of IAA release is controlled by the exogenous supply of IAM. Bacterial conversion of IAM to IAA occurs at the expense of growth on IAM, even though it is a robust substrate for *R. pomeroyi*, and suggests highly-regulated transport and cellular fate of IAM. A search of isolate genomes reveals the prevalence of incomplete IAA biosynthesis pathways similar to that in *R. pomeroyi* in many environmental bacteria and suggests that autotroph-bacteria exchange of IAM is common in both marine and terrestrial environments. Further, strains that have an incomplete pathway and a putative IAM transporter are significantly less likely to be pathogenic than those with a complete IAM pathway. This work reemphasizes the importance of IAA in the marine microbial loop and underscores the significance of pathway intermediates as fine-tuned ecological signals.

Opening statement

Previous work by MU, MAM, WFS, and MO identified the diel expression pattern of the *iaaH* gene in *Ruegeria pomeroyi* DSS-3. MAM and WFS conceived of the ideas for follow up experiments. WFS and CRS designed and performed the transcriptome experiments. WFS designed and performed the RB-TnSeq experiments. WFS, MAM, and MU designed the time series experiments. WFS ran the experiments and prepared samples for analysis. MU ran NMR samples and analyzed the data. WFS designed and performed bioinformatic analysis with critical input and feed back from MMH. Einat Segev and Shady Amin both provided valuable feedback for optimizing culture conditions. WFS wrote the manuscript with critical editing and revision from MAM.

Introduction

Surface ocean microbial communities are responsible for half of global net primary production and form the foundation of marine food webs (1-4). Yet the processes that shape the structure, function, and resiliency of marine microbial communities remain poorly understood. Recent research has pointed to the importance of interactions among community members, including those that occur within the phycosphere microenvironment (5-12), a diffusive boundary layer around a phytoplankton cell where the concentration of metabolites and heterotrophic bacteria can be orders of magnitude greater than in surrounding bulk seawater (13).

The phycosphere is ecologically analogous to the well-studied rhizosphere of terrestrial plants (13, 14), and indeed this similarity was the inspiration for its name (15). The soil surrounding plant roots is similarly characterized by high concentrations of metabolites and abundant heterotrophic bacteria (14, 16). In both systems, a photoautotroph releases organic carbon into the surrounding microenvironment through processes that include passive leakage,

active exudation, and cell death (3, 16). Heterotrophic community members rely on these metabolites as a source of carbon and energy. Microbes can also enter into mutualistic interactions mediated by metabolite exchanges that provide benefits to the photoautotroph, including nutrients, detoxification enzymes, vitamins, and cofactors (9, 10, 17-22). Further, there is evidence that photoautotrophs are able influence the structure of their associated microbial communities through release of selective metabolites that increase populations of beneficial members or inhibit antagonists (10).

Within this tangle of metabolite exchange, the phytohormone indole acetic acid (IAA) is of particular interest as it is an auxin type growth hormone that has roles in both mutualistic and antagonistic interactions (23, 24). IAA's role as a phytohormone is hypothesized to have evolved in the common ancestor of land plants and phytoplankton, diverging over time with regard to biosynthetic pathways and cellular responses (25-29). In terrestrial plants the mechanisms of IAA action are as hormones and signals that play a variety of roles in regulating tissue growth, light response, root and shoot architecture, and vascular development (29-31). Plants synthesize endogenous IAA, but are also sensitive to exogenous IAA produced by rhizosphere bacteria (23, 32). When at concentrations within a plant's narrow optimal range, exogenous IAA, can stimulate root growth and increase root exudation, benefitting both the plant and its rhizosphere bacteria (30, 33, 34). When at concentrations above optimal, IAA can act as a virulence factor, used by pathogenic bacteria to cause tumors, galls, and necrosis (33, 35).

IAA production and sensitivity has also been observed in marine phytoplankton, including diatoms, coccolithophores, and green algae (18, 26, 28, 36). Like rhizosphere bacteria, phycosphere bacteria produce IAA to facilitate both mutualistic and antagonistic interactions, and phytoplankton have variable dose-dependent responses (18, 36). Despite a shared

evolutionary origin with plants, the genetic IAA response mechanisms in phytoplankton remains unclear (28, 36). Some green algae contain homologs to the IBR5 gene (28, 29) that in plants has been shown to be responsible for IAA sensitivity (37). The diatom *Pseudo-nitzchia multiseries* increases expression of the cyclins 2 and 8 in the presence of IAA producing *Sulfitobacter* SA11, initiating a cell cycle response (18).

Marine phytoplankton and bacteria can synthesize IAA through several pathways (Fig. 1). In the case of model bacterium *Ruegeria pomeroyi* DSS-3, IAA synthesis was hypothesized to occur via the indole-3-acetamide (IAM) pathway, the first step of which is decarboxylation of tryptophan to IAM by the enzyme encoded by *iaaM* (tryptophan 2-monooxygenase), and the second of which is hydrolysis of IAM to IAA by the enzyme encoded by *iaaH* (indole-3-acetamide hydrolase) (Fig. 4.1). The IAM pathway has been well-studied in bacteria that are plant pathogens (32, 33, 38-42), though there are examples of beneficial rhizosphere bacteria using the IAM pathway as well (43, 44). In a recent study, diatom *Pseudo-nitzchia multiseries* was shown to excrete tryptophan that the associated bacterium *Sulfitobacter* SA11 subsequently converted to IAA via the IAM pathway, releasing IAA and stimulating diatom growth (18). When *R. pomeroyi* was grown in co-culture with the diatom *Thalassiosira pseudonana CCMP 1335* increased expression of an *iaaH* homolog was observed (45) and initially thought to be a second example of this tryptophan-to-IAA mutualistic interaction. However, *R. pomeroyi* appeared to have no homolog of *iaaM*, and is only able to carry out the second step of the pathway from IAM to IAA (Fig 4.1). If confirmed, this would suggest an exogenous source of IAM available to the bacterium.

Just recently, incomplete versions of several IAA biosynthesis pathways, including the IAM pathway, were found to be common and widely distributed among rhizosphere bacteria

(46). To this point, there has been no detailed characterization of the ecological implications of incomplete pathways (24, 46), although their high frequency suggests considerable exchange of IAA biosynthesis pathway intermediates among rhizosphere microbes. Given the missing gene in *R. pomeroyi* for conversion of tryptophan to IAM, we hypothesized that the intermediate IAM was provided to the bacterium by the co-cultured diatom.

Here we verify that *R. pomeroyi* imports IAM, confirm the function of the bacterium's IAM hydrolase gene *iaaH* and its regulator, and show that the pathway is indeed incomplete. Unexpectedly, the bacterium stoichiometrically converts IAM to IAA at the expense of growth, excreting all IAA outside the cell. Our data suggest that phytoplankton-derived IAM acts as a metered ecological signal that facilitates interactions with beneficial heterotrophic bacteria while maintaining control of the concentration of IAA. The presence of this partial IAM pathway in ~3,800 bacterial genomes, which includes marine, soil, plant associated, and freshwater isolates, suggests a 'call-and-response' recognition system may be a conserved inter-domain interaction mechanism with deep evolutionary origins (27, 29).

Methods

Growth and drawdown experiments

An isolated mutant with a transposon disruption of the putative *iaaH* (SPO1489) and also a mutant in an adjacent regulatory gene (SPO1488) from an *R. pomeroyi* arrayed library (47). The ability of the mutants to grow on IAM as a sole carbon source was tested following the methods described in Schroer et al. (48). Briefly, overnight cultures of *R. pomeroyi* mutants were grown in ½YTSS with 50 mg ml^{-1} kanamycin (kan50). Experiments were performed in the minimal medium L1 modified to 20 ppt salinity and amended with 3mM NH_4Cl + kan50 ((48, 49), https://doi.org/10.17504/protocols.io.jvccn2w); IAM was provided at 1.2 mM. Mutant

cultures were washed 3 times in substrate-free minimal medium and resuspended cells were inoculated into 96 well plates. A pooled library of mutants, each harboring the transposon and resistant to kanamycin, was used as an analog for WT growth (48). Plates were incubated, shaking at 425 rpm at 25^{o}C for 72 h. Optical density at 600 nm (OD_{600}) was read hourly.

For drawdown experiments, 3 ml of modified L1 containing 12 mM C of IAM, L-tryptophan, or a 1:2 ratio of citrate to IAM were inoculated with 70 ml of washed overnight cultures of *R. pomeroyi* DSS-3 wild-type analog (WT), mutant DSPO1488, or mutant DSPO1489 in triplicate. Samples of 220 ml were collected at 12-24 h intervals for 72 h. OD_{600} was measured, samples were centrifuged at 2250 x g for 10 min at 10^{o}C, and the supernatant was transferred to 1.5 ml microcentrifuge tubes or 96 well plates and stored at -80^{o}C.

Transcriptome experiment

Overnight cultures of *R. pomeroyi* mutants were grown in ½YTSS with kan50, then washed and inoculated into modified L1 medium with either 12 mM C IAM or 12 mM C glucose and grown at 30^{o}C. At early-exponential phase, cultures were centrifuged at 3500 x g for 5 min. The cell pellet was flash frozen in liquid nitrogen and stored at -80^{o}C.

Extraction of mRNA was performed following the protocols described in (45) using the Zymobiomics RNA bacterial miniprep (Zymo, Irvine CA, USA) with internal standards (50). Library construction was performed with NEBNext Ultra 11 Directional RNA Library Prep Kit (New England Biolabs, Ipswich, MA, USA). Illumina Nextseq500 (SE150) sequencing was performed at the Georgia Genomics and Bioinformatics Core (GGBC, Athens, GA, USA). Transcript data were analyzed following the pipeline described in Uchimiya et al. (45). Relative differential expression was determined using DESeq2, and absolute expression was calculated from quantitative normalization to an internal standard (50).

RB-TnSeq

RB-TnSeq methods are described in Schroer et al. (48). Briefly, a pooled barcoded transposon mutant library of *R. pomeroyi* was grown up overnight in ½YTSS + kan50. The culture was washed 3 times and 20 ml were inoculated into 180 ml of modified L1 medium with 8 mM C IAM or 8 mM C glucose. Cultures were grown shaking at 25ºC for 72 h and transferred (10% volume) into fresh medium three times for 72 h, with the final transfer of 200 ml into 800 ml of medium. At termination, samples were centrifuged at 2,250 ? x g for 10 min. Cell pellets were transferred to 1.5 ml microcentrifuge tubes and stored at -80ºC. Genomic DNA was extracted using a DNAeasy kit (Qiagen, Hilden, Germany). Barcode amplification was performed using an indexed primer set (48, 51). Purified PCR product was pooled and sequenced on a NextSeq SE150 Mid Output flow cell (SE150) at the GGBC.

RB-TnSeq analysis was performed as described in Schroer et al. (48). Briefly, the pipeline provided in Wetmore et al. (51) was used to demultiplex barcoded reads. Gene read frequency was normalized within each sample. Mutant defect was calculated as the fold change between each gene's frequency on IAM versus on glucose. Significant enrichment or depletion was determined by T-tests with false discovery rate (FRD) corrected p-values.

NMR metabolomics

Metabolite analysis was performed using a Bruker Avance lll 600 MHz spectrometer (Bruker, Billerica, MA, USA) equipped with a 5-mm TCI cryoprobe. Samples were prepared with addition of a deuterated phosphate buffer (30 mmol L-1, pH 7.4) and the internal standard 2,2-dimethyl-2-silapentane-5-sulfonate-d6 (DSS, 1 mmol L-1) (10:1 (vol : vol)) and transferred to 3 mm NMR tubes (Bruker). Data were acquired by a one dimensional 1H experiment with water suppression (noesypr1d, Bruker) at 298K using TopSpin 3.6.4 (Bruker). Spectra were

processed using NMRPipe on NMRbox (refs), and the processed data were analyzed using Metabolomics Toolbox and MATLAB vR2023a. For metabolite quantification, spectra were normalized to DSS, and peak area for representative peaks was calculated. TopSpin experiment settings, NMRpipe spectra processing parameters, and MATLAB data analysis scripts are available in Metabolomics Workbench.

Analysis of bacterial genomes in IMG

Isolated bacterial genomes were accessed from the JGI's Integrated Microbial Genomes (IMG) database using the IMG workspace interface and built-in tools (52). All genomes were searched for KO assignments (KEGG, (53)) *iaaM* (K00466) and *iaaH* (K21801). Gene phylogenies of putative *iaaH* sequences, confirmed *iaaH* sequences, and outgroup sequences were made using FastTree (54).

Homologs of the putative IAM transporter's ATPase components (SPO1492 – 1493) were initially recruited by blastp of *R. pomeroyi* sequences against the genomes containing *iaaH*, using an E value cutoff of $<10^{-30}$. Sequences were aligned using Clustal Omega (55) and phylogenetic trees were produced with Fast Tree (54). For context, these trees included a set a curated set of ABC transporter ATPases from TCBD (56) and other sources, compiled by Schroer et al. (57), hear after the transporter reference database. The clade containing sequences with high-confidence matches to the *R. pomeroyi* gene was identified manually. Gene neighborhood analysis was performed in R v4.2.2, by determining if genome coordinates of both transporter components were within +/- 10 Kbp of the *iaaH* gene. Assignment of isolate role (pathogen, beneficial, marine, other, or unknown) was performed manually using information from IMG metadata, strain collection libraries, and published literature (Table S7, S8). When

required, role was assigned based on membership of full genome ANI clusters (>96.5% identity) (52), where other members of the given cluster had known roles.

Results and Discussion

iaaH operon genes are highly expressed in response to exogenous IAM

In a transcriptional analysis of *R. pomeroyi* in co-culture with diatom *T. pseudonana* (45), we noted 36-fold upregulation of a gene annotated as an *iaaH* homolog (SPO1489) at the time of peak photosynthesis, potentially mediating the second step in the IAM biosynthesis pathway to IAA (Fig. 4.2a). The *iaaH* homolog is located in an operon with a regulator (SPO1488) and components of an ABC transporter with no existing substrate annotation (SPO1490 – 1493) (Fig. 4.2a). While this finding suggested a possible IAA mediated interaction with the diatom, the *R. pomeroyi* genome lacks an *iaaM* homolog, indicating a partial IAA biosynthesis pathway that synthesizes IAA from IAM but not tryptophan.

We confirmed that the putative *iaaH* and neighboring genes in the operon are indeed expressed in the presence of IAM. Cultures of *R. pomeroyi* grown on minimal medium with either 12 mM C glucose or 12 mM C IAM as sole carbon sources were sampled during early exponential phase. Differential expression between the glucose- and IAM-grown cultures revealed a strong transcriptional response of the *iaaH* operon to IAM (Fig. 4.2b, Table S4.1), confirming an important role for these genes during growth on IAM.

Knockout mutant deficits confirm iaaH *and regulator function*

To verify that the upregulated hydrolase is indeed *iaaH*, knockout mutants of the putative *iaaH* (ΔSPO1489) and adjacent regulator (ΔSPO1488) were screened for defects in growth and IAA production. A positive control consisted of a pooled *R. pomeroyi* RB-TnSeq library (48) which served as an analog for wild-type growth (WT analog) but carried the transposon and

resistance gene insertions present in the knockout mutants. We found that both mutants were unable to grow on IAM as a sole carbon source and were unable to synthesize IAA (Fig. 4.3, S4.1).

The function of the *iaaH* gene has been experimentally verified in 6 rhizosphere bacteria previously (38-43) (Table S4.2). All strains have the complete IAM pathway, and all but one strain are closely related plant pathogens (43). The confirmation of *iaaH* function in *R. pomeroyi* affirms IAA biosynthesis by marine bacteria and suggests an ecological role for incomplete pathways when supplied with appropriate intermediates.

Knockout mutants of two components of the putative IAM uptake transporter (SPO1491 and SPO1493) did not show broad deficits of IAM drawdown or growth. Both mutants did show slight defects of IAM drawdown at the 12 h time point (ANOVA and TukeyHSD, p.adj = 0.01, 0.008, respectively), but the magnitude of the deficit was small (Fig. S4.1). IAM is hydrophobic compound, and it is possible that at the high concentrations provided here it can diffuse through the membrane. Follow up work should assess the function of these transporter mutants at lower concentrations of IAM.

R. pomeroyi's unexpected growth dynamics on IAM

The WT analog was able to grow on IAM as a sole carbon source in minimal medium and synthesize IAA (Fig 4.3), but showed unusual growth and metabolism dynamics. NMR metabolomics of the medium (Fig 4.3c, 4.4) revealed rapid drawdown of IAM concurrent with extracellular release of IAA in the first 24 h of incubation. This conversion of IAM to IAA occurred in an approximate 1:1 molar ratio (Fig. S4.2), indicating that all IAM was converted to IAA and released from the cell, and no growth was observed during this 24 h period. *R.*

pomeroyi initiated growth only when all IAM was exhausted from the medium, with growth supported by uptake of the IAA the bacterium had previously excreted (Fig 4.3b).
The delayed utilization, by *R. pomeroyi*, of the IAA it synthesized was further supported by an experiment with the pooled *R. pomeroyi* RB-TnSeq library (Figure S4.3, Table S4.2) (48). Mutants in *iaaH* operon genes, unable to assimilate or transform IAM, would typically be expected to have decreased fitness when provided with IAM as a sole carbon source. Our finding that, to the contrary, strains with disruptions in *iaaH* operon genes were the among the most fit mutants indicated their early exploitation of the IAA pool released extracellularly by the other members of the mutant pool unaffected in IAM metabolism.

More relevant to a natural marine bacterial community, the RB-TnSeq finding suggests that tight regulation over the use of IAA as a growth substrate is governed by intracellular IAM flux. This results in forfeiture of a carbon source that could support robust growth, although the effect of this loss on fitness depends on the supply rate of IAM and the specificity of the regulation. We asked whether IAM regulation inhibits growth on substrates other than IAM by inoculating *R. pomeroyi* into a mixed substrate pool consisting of citrate and IAM (4 mM C citrate, 8 mM C IAM). The presence of IAM in the medium did not inhibit growth on citrate, with *R. pomeroyi* simultaneously utilizing citrate for growth and drawing down IAM to produce and release IAA. When citrate was depleted, growth ceased and the culture entered a brief lag period while the remaining IAM was converted to IAA (Fig 4.4). Consistent with our earlier experiments using IAM as the only carbon source (Fig. 4.3b), growth on excreted IAA only occurred after all IAM had been converted to IAA (Fig. 4.4). From an evolutionary perspective, the fitness advantage to *R. pomeroyi* of IAM/IAA as a phycosphere interaction signal appears to outweigh their advantage of as growth substrates. A potential mechanism behind this trade-off

could be the establishment of mutualistic interactions between *R. pomeroyi* and the diatom, potentially including increased release of phytoplankton exudates.

R. pomeroyi's incomplete IAM pathway

To confirm that the first step in the IAM pathway, which was not evident based on genomic analysis of *R. pomeroyi*, indeed has no metabolic work around, *R. pomeroyi* was grown on 12 mM C L-tryptophan as a sole carbon source. NMR metabolomics of the extracellular medium indicated no IAA in the medium at any time point (Fig. 4.3). Mutualistic two-way interactions facilitated by the exchange of tryptophan and IAA between photoautotrophs and bacteria have been described in phycosphere (18, 36) and rhizosphere micro-environments (32). However, *R. pomeroyi*'s incomplete pathways suggests a shortcut to the traditional tryptophan to IAA recognition cycle, with phytoplankton-derived IAM acting as the precursor to bacterial IAA synthesis. We hypothesize that this abbreviated recognition circuit may be beneficial to phytoplankton by providing a more targeted interaction. As an essential amino acid, tryptophan would be valuable to a broad diversity of microbes and potentially scavenged by community members incapable of IAA production and mutualism. By releasing IAM, a phytoplankter would select for interactions with a targeted group of microbes capable of transporting IAM.

Incomplete IAM pathways are more common than complete

Hypotheses about the ecological importance of IAM-mediated interactions led us to determine the frequency of bacteria with incomplete pathways similar to that of *R. pomeroyi*. All isolate genomes in the JGI Integrated Microbial Genomes and Microbiomes (IMG) database (52) were searched for homologs of *iaaM* (K00466) and *iaaH* (K21801). Surprisingly, we found that organisms with incomplete pathways were considerably more common than those with complete pathways: of 4,888 isolates containing *iaaM* and/or *iaaH*, only 525 had a complete pathway. Of

the incomplete pathways 454 had only *iaaM*, while 3,910 had only *iaaH*, including *R. pomeroyi* (Fig. 4.5, S4.4, Table S4.3-S4.6). These findings agree with those of Zhang et al (46) who found a high prevalence of incomplete pathways in the genomes of soil bacterial isolates. As the bioinformatic analysis focused on genes in the IAM pathway, it is possible that some isolates have alternative intact IAA biosynthesis pathways. None, however, have a complete indole-3-acetonitrile pathway (Fig. 4.1) that would allow an alternate route to synthesize IAM from tryptophan. Other examples of distributed biochemical pathways in which community members exchange pathway intermediates, each performing distinct reaction steps, have been found in environmental systems for B vitamin biosynthesis (58) and denitrification (59).

The widespread presence of incomplete IAM pathways suggests that the intermediate IAM is a frequently exchanged metabolite in rhizosphere and phycosphere interactions. We speculate that interactions between photoautotrophs and heterotrophic microbes mediated by IAM-IAA exchange are more likely to be beneficial than pathogenic because the photoautotroph ultimately controls IAA concentrations by metering the release of IAM. To investigate this, we characterized the ecological context (plant beneficial, plant pathogen, marine, freshwater, other, unknown) of the 524 bacterial isolates containing complete IAM pathways and a select group 128 isolates with incomplete pathways that had a clearly defined IAM operon similar to *R. pomeroyi* (Table S4.7, S4.8). This added criterion for the incomplete pathways essentially requires an ABC influx transporter, adjacent to *iaaH*, that is homologous to of *R. pomeroyi's* putative IAM transporter (SPO1492, SPO1493). We combined stringent homology and phylogeny-based approaches for gene identification (Methods, Fig. S4.5,S4.6, Table S4.9) with ecological context using information from IMG metadata, strain collection libraries, and published literature (Table S4.7, S4.8). Results showed that isolates with incomplete pathways

and the putative IAM transporter were classified as pathogens significantly less often than isolates with complete pathways (0.8% vs 37%; Fig. 4.6a, Fisher's Exact test, pathogen vs non-pathogen, $p = 2x10^{-16}$), though beneficial strains were frequent in both. Both groups of strains drew broadly from members of the Betaproteobacteria, Alphaproteobacteria, Gammaproteobacteria, and Actinomycetes (Fig. 4.6b).

Other IAA interaction genes are suggested by expression and RB-TnSeq

The gene exhibiting the greatest differential expression when growing on IAM relative to glucose was an efflux DMT transporter (SPO3584) with 1,667 fold higher expression on IAM (Fig 4.2b), which we hypothesize to be the IAA exporter. However, a knockout mutant of this transporter did not show a deficit of IAA release (Fig. S4.1), it is possible that, given the high IAM/IAA concentrations in this system, alternate export mechanisms or diffusion facilitated IAA release. Though IAA exporters are well characterized in plants, a bacterial IAA transporter has yet to be identified (24). Future work to characterize this exporter is important to further understand the mechanism of IAA exchange that facilitates both mutualism and pathogenicity.

Genes annotated for catabolism of phenylalanine and aromatic rings were highly expressed in the transcription data and had low fitness in the RB-TnSeq data when grown on IAM (Fig. S4.3, Table S4.1), and we hypothesize that these mediate catabolism of IAA. The ability to grow on IAA as a sole carbon source is considered relatively uncommon, but in rhizosphere bacteria this trait is associated with plant mutualists (60). An IAA degradation pathway identified in *Pseudomonas putida* and other mutualist bacteria has been shown to be protective, reducing the severity of disease by pathogen derived IAA (60, 61). *R. pomeroyi's* putative IAA degradation genes are not homologs of those previously described in *P. putida*, but are annotated to catalyze similar reactions (aromatic ring breaking, modification of aromatic

sidechains). These findings suggest that, as with IAA biosynthesis, multiple pathways of IAA catabolism have evolved, and similar to the rhizosphere, IAA catabolism in the phycosphere is associated with mutualists.

Conclusion

IAA-mediated interactions play an outsized role in modulating primary production, positively and negatively, in both terrestrial and marine environments. Though the roles of IAA have been a research focus for decades, the complexity of its biosynthesis, response mechanisms, evolutionary origins, and the community interactions are still being unraveled. Here we provide a novel nuance to the tangle of IAA mediated interactions. The characterization of *R. pomeroyi*'s *iaaH* gene and the strict regulation of IAA production and release represents a finely-tuned chemical signal governing phycosphere interactions. Similarly incomplete pathways are common in environmental isolates, suggesting widespread exchange of IAM in soil and marine environments that influences the distribution of pathogenic and mutualistic interactions.

References

1. Azam F, Fenchel T, Field J, Grey J, Meyer-Reil L, Thingstad F. The ecological role of water-column microbes. Mar Ecol Prog Ser. 1983;10:257-63.

2. Anderson TR, Ducklow HW. Microbial loop carbon cycling in ocean environments studied using a simple steady-state model. Aquat Microb Ecol. 2001;26:37-49.

3. Moran MA, Ferrer-González FX, Fu H, Nowinski B, Olofsson M, Powers MA, et al. The ocean's labile DOC supply chain. Limnol Oceanogr. 2022;67:1007-21.

4. Cole JJ, Likens GE, Strayer DL. Photosynthetically produced dissolved organic carbon: An important carbon source for planktonic bacteria. Limnol Oceanogr. 1982;27:1080-90.

5. Nowinski B, Moran MA. Niche dimensions of a marine bacterium are identified using invasion studies in coastal seawater. Nat Microbiol. 2021;6:524-32.

6. Ferrer-González FX, Hamilton M, Smith CB, Schreier JE, Olofsson M, Moran MA. Bacterial transcriptional response to labile exometabolites from photosynthetic picoeukaryote *Micromonas commoda*. ISME Commun. 2023;3:5.

7. Schreier JE, Smith CB, Ioerger TR, Moran MA. A mutant fitness assay identifies bacterial interactions in a model ocean hot spot. Proc Natl Acad Sci USA. 2023;120.

8. Landa M, Burns AS, Roth SJ, Moran MA. Bacterial transcriptome remodeling during sequential co-culture with a marine dinoflagellate and diatom. ISME J. 2017;11:2677-90.

9. Roth-Rosenberg D, Aharonovich D, Luzzatto-Knaan T, Vogts A, Zoccarato L, Eigemann F, et al. *Prochlorococcus* cells rely on microbial interactions rather than on chlorotic resting stages to survive long-term nutrient starvation. MBio. 2020;11:10.1128/mbio. 01846-20.

10. 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. Proc Natl Acad Sci USA. 2020;117:27445-55.

11. Beiralas R, Ozer N, Segev E. Abundant *Sulfitobacter* marine bacteria protect *Emiliania huxleyi* algae from pathogenic bacteria. ISME Commun. 2023;3.

12. Mayali X, Samo TJ, Kimbrel JA, Morris MM, Rolison K, Swink C, et al. Single-cell isotope tracing reveals functional guilds of bacteria associated with the diatom *Phaeodactylum tricornutum*. Nature Commun. 2023;14.

13. Seymour JR, Amin SA, Raina J-B, Stocker R. Zooming in on the phycosphere: the ecological interface for phytoplankton–bacteria relationships. Nat Microbiol. 2017;2:1-12.

14. Berendsen RL, Pieterse CM, Bakker PA. The rhizosphere microbiome and plant health. Trends plant sci. 2012;17:478-86.

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

16. Jones DL. Organic acids in the rhizosphere–a critical review. Plant Soil. 1998;205:25-44.

17. Durham BP, Sharma S, Luo H, Smith CB, Amin SA, Bender SJ, et al. Cryptic carbon and sulfur cycling between surface ocean plankton. Proc Natl Acad Sci USA. 2015;112:453-7.

18. Amin S, Hmelo L, Van Tol H, Durham B, Carlson L, Heal K, et al. Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. Nature. 2015;522:98-101.

19. Morris JJ, Johnson ZI, Szul MJ, Keller M, Zinser ER. Dependence of the cyanobacterium *Prochlorococcus* on hydrogen peroxide scavenging microbes for growth at the ocean's surface. PloS One. 2011;6.

20. Ferrer-González FX, Widner B, Holderman NR, Glushka J, Edison AS, Kujawinski EB, et al. Resource partitioning of phytoplankton metabolites that support bacterial heterotrophy. ISME J. 2021;15:762-73.

21. Moran MA, Kujawinski EB, Schroer WF, Amin SA, Bates NR, Bertrand EM, et al. Microbial metabolites in the marine carbon cycle. Nat Microbiol. 2022;7:508-23.

22. Edwards BR, Bidle KD, Van Mooy BA. Dose-dependent regulation of microbial activity on sinking particles by polyunsaturated aldehydes: Implications for the carbon cycle. Proc Natl Acad Sci USA. 2015;112:5909-14.

23. Spaepen S, Vanderleyden J. Auxin and plant-microbe interactions. Cold Spring Harb Perspect Biol. 2011;3.

24. Duca DR, Glick BR. Indole-3-acetic acid biosynthesis and its regulation in plant-associated bacteria. Appl Microbiol Biotechnol. 2020;104:8607-19.

25. Morffy N, Strader LC. Old Town Roads: Routes of auxin biosynthesis across kingdoms. Curr Opin Plant Biol. 2020;55:21-7.

26. Ohtaka K, Hori K, Kanno Y, Seo M, Ohta H. Primitive auxin response without TIR1 and Aux/IAA in the charophyte alga *Klebsormidium nitens*. Plant Physiol. 2017;174:1621-32.

27. Carrillo-Carrasco VP, Hernandez-Garcia J, Mutte SK, Weijers D. The birth of a giant: evolutionary insights into the origin of auxin responses in plants. EMBO J. 2023;42.

28. Lau S, Shao N, Bock R, Jürgens G, De Smet I. Auxin signaling in algal lineages: fact or myth? Trends plant sci. 2009;14:182-8.

29. De Smet I, Voß U, Lau S, Wilson M, Shao N, Timme RE, et al. Unraveling the evolution of auxin signaling. Plant Physiol. 2011;155:209-21.

30. Schroeder MM, Gomez MY, McLain N, Gachomo EW. *Bradyrhizobium japonicum* IRAT FA3 alters *Arabidopsis thaliana* root architecture via regulation of auxin efflux transporters PIN2, PIN3, PIN7, and ABCB19. Mol Plant Microbe Interact. 2022;35:215-29.

31. Woodward AW, Bartel B. Auxin: regulation, action, and interaction. Ann Bot. 2005;95:707-35.

32. Duca D, Lorv J, Patten CL, Rose D, Glick BR. Indole-3-acetic acid in plant–microbe interactions. Antonie Van Leeuwenhoek. 2014;106:85-125.

33. Spaepen S, Vanderleyden J, Remans R. Indole-3-acetic acid in microbial and microorganism-plant signaling. FEMS Microbiol Rev. 2007;31:425-48.

34. Patten CL, Blakney AJ, Coulson TJ. Activity, distribution and function of indole-3-acetic acid biosynthetic pathways in bacteria. Crit Rev Microbiol. 2013;39:395-415.

35. Kunkel BN, Harper CP. The roles of auxin during interactions between bacterial plant pathogens and their hosts. J Exp Bot. 2018;69:245-54.

36. 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.

37. Strader LC, Monroe-Augustus M, Bartel B. The IBR5 phosphatase promotes *Arabidopsis* auxin responses through a novel mechanism distinct from TIR1-mediated repressor degradation. BMC Plant Biol. 2008;8:1-15.

38. Mazzola M, White FF. A mutation in the indole-3-acetic acid biosynthesis pathway of *Pseudomonas syringae* pv. *syringae* affects growth in *Phaseolus vulgaris* and syringomycin production. J Bacteriol. 1994;176:1374-82.

39. Clark E, Manulis S, Ophir Y, Barash I, Gafni Y. Cloning and characterization of iaaM and iaaH from *Erwinia herbicola* pathovar *gypsophilae*. Phytopathology. 1993;83:234-40.

40. Yang S, Zhang Q, Guo J, Charkowski AO, Glick BR, Ibekwe AM, et al. Global effect of indole-3-acetic acid biosynthesis on multiple virulence factors of *Erwinia chrysanthemi* 3937. Appl Environ Microbiol. 2007;73:1079-88.

41. Yamada T, Palm CJ, Brooks B, Kosuge T. Nucleotide sequences of the *Pseudomonas savastanoi* indoleacetic acid genes show homology with *Agrobacterium tumefaciens* T-DNA. Proc Natl Acad Sci USA. 1985;82:6522-6.

42. Cerboneschi M, Decorosi F, Biancalani C, Ortenzi MV, Macconi S, Giovannetti L, et al. Indole-3-acetic acid in plant–pathogen interactions: a key molecule for in planta bacterial virulence and fitness. Res Microbiol. 2016;167:774-87.

43. Bellés-Sancho P, Liu Y, Heiniger B, von Salis E, Eberl L, Ahrens CH, et al. A novel function of the key nitrogen-fixation activator NifA in beta-rhizobia: Repression of bacterial auxin synthesis during symbiosis. Front Plant Sci. 2022;13.

44. Theunis M, Kobayashi H, Broughton WJ, Prinsen E. Flavonoids, NodD1, NodD2, and nod-box NB15 modulate expression of the y4wEFG locus that is required for indole-3-acetic acid synthesis in *Rhizobium sp. strain* NGR234. Mol Plant Microbe Interact. 2004;17:1153-61.

45. Uchimiya M, Schroer W, Olofsson M, Edison AS, Moran MA. Diel investments in metabolite production and consumption in a model microbial system. ISME J. 2022:1-12.

46. Zhang P, Jin T, Kumar Sahu S, Xu J, Shi Q, Liu H, et al. The distribution of tryptophan-dependent indole-3-acetic acid synthesis pathways in bacteria unraveled by large-scale genomic analysis. Molecules. 2019;24:1411.

47. Mejia C, Trujillo Rodriguez L, Poudel R, Ellington A, Rivers A, Reisch C. An arrayed transposon library of *Ruegeria pomeroyi* DSS-3. bioRxiv. 2022:2022.09.11.507510.

48. Schroer WF, Kepner HE, Uchimiya M, Mejia C, Rodriguez LT, Reisch CR, et al. Functional annotation and importance of marine bacterial transporters of plankton exometabolites. ISME Commun. 2023;3:37.

49. Guillard R, Hargraves P. *Stichochrysis immobilis* is a diatom, not a chrysophyte. Phycologia. 1993;32:234-6.

50. Satinsky BM, Gifford SM, Crump BC, Moran MA. Use of internal standards for quantitative metatranscriptome and metagenome analysis. Methods Enzymol. 531: Elsevier; 2013. p. 237-50.

51. Wetmore KM, Price MN, Waters RJ, Lamson JS, He J, Hoover CA, et al. Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons. MBio. 2015;6:e00306-15.

52. Chen I-MA, Chu K, Palaniappan K, Ratner A, Huang J, Huntemann M, et al. The IMG/M data management and analysis system v. 7: content updates and new features. Nucleic Acids Res. 2023;51:723-32.

53. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28:27-30.

54. Price MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol. 2009;26:1641-50.

55. Madeira F, Pearce M, Tivey AR, Basutkar P, Lee J, Edbali O, et al. Search and sequence analysis tools services from EMBL-EBI in 2022. Nucleic acids research. 2022;50:W276-W9.

56. Saier Jr MH, Reddy VS, Moreno-Hagelsieb G, Hendargo KJ, Zhang Y, Iddamsetty V, et al. The transporter classification database (TCDB): 2021 update. Nucleic Acids Res. 2021;49:D461-D7.

57. Schroer WS, M; Saunders, J; Eren, M; Moran, MA. High-Confidence Global Mapping of Bacterial Substrate Uptake Potential. Doctoral Dissertation. 2023;Ch. 3.

58. Wienhausen G, Bruns S, Sultana S, Dlugosch L, Groon L-A, Wilkes H, et al. The overlooked role of a biotin precursor for marine bacteria-desthiobiotin as an escape route for biotin auxotrophy. ISME J. 2022;16:2599-609.

59. Mosley OE, Gios E, Close M, Weaver L, Daughney C, Handley KM. Nitrogen cycling and microbial cooperation in the terrestrial subsurface. ISME J. 2022;16:2561-73.

60. Leveau JH, Gerards S. Discovery of a bacterial gene cluster for catabolism of the plant hormone indole 3-acetic acid. FEMS Microbiol Ecol. 2008;65:238-50.

61. Zúñiga A, Poupin MJ, Donoso R, Ledger T, Guiliani N, Gutiérrez RA, et al. Quorum sensing and indole-3-acetic acid degradation play a role in colonization and plant growth promotion of *Arabidopsis thaliana* by *Burkholderia phytofirmans* PsJN. Mol Plant Microbe Interact. 2013;26:546-53.

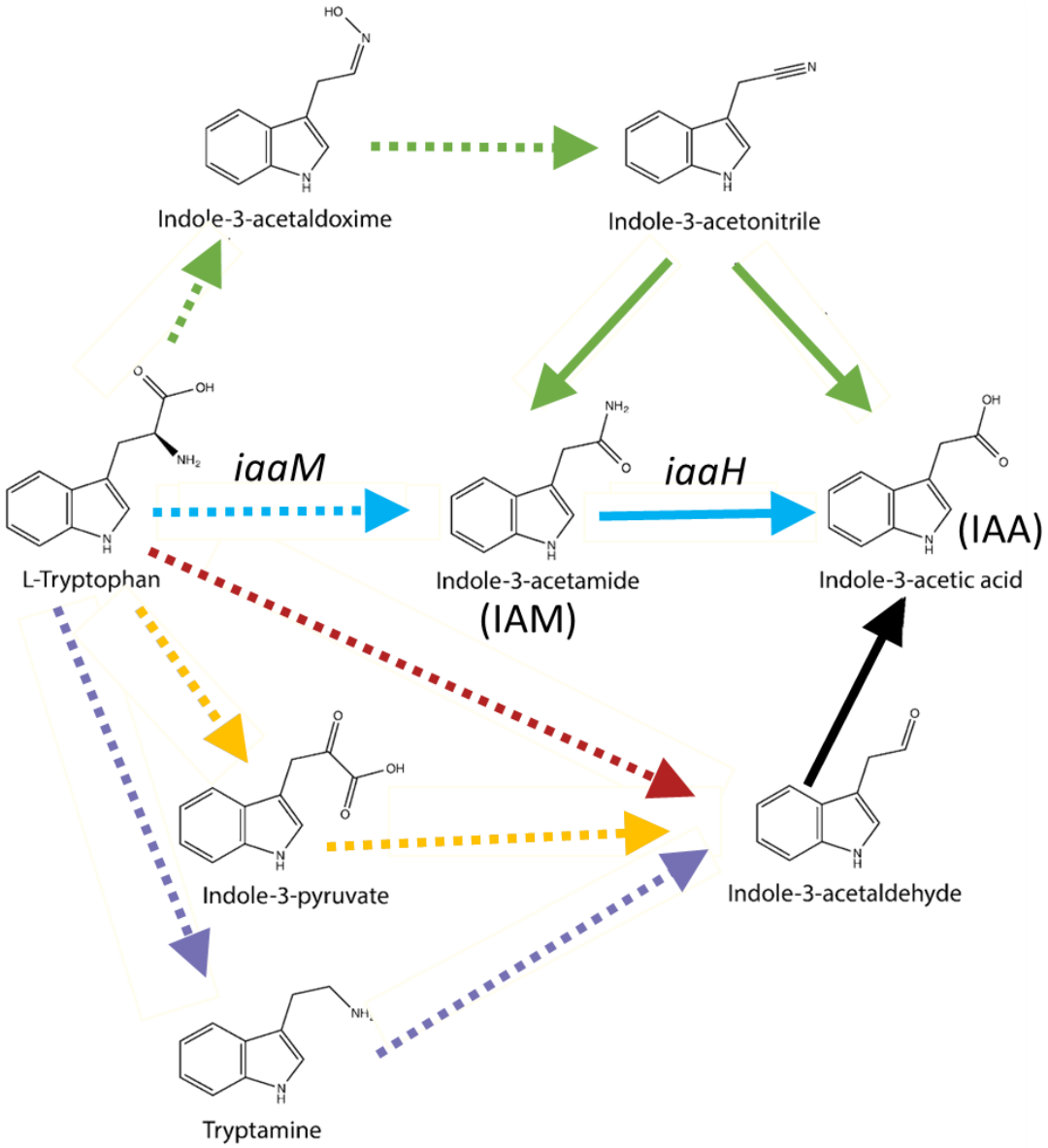


Figure 4.1. Five tryptophan dependent IAA biosynthesis pathways. Pathway names: Indole-3-acetonitrile, green; Indole-3-acetamide (IAM), blue; tryptophan sidechain, maroon; indole-3-pyruvate, orange; tryptamine, purple; final step of multiple pathways, black. Solid arrows indicate the presence of a homolog in *R. pomeroyi's* genome, dashed arrows represent genes without a homolog in *R. pomeroyi's* genome, based on KEGG ortholog and manual annotation. Modified from Amin et al. (2015).

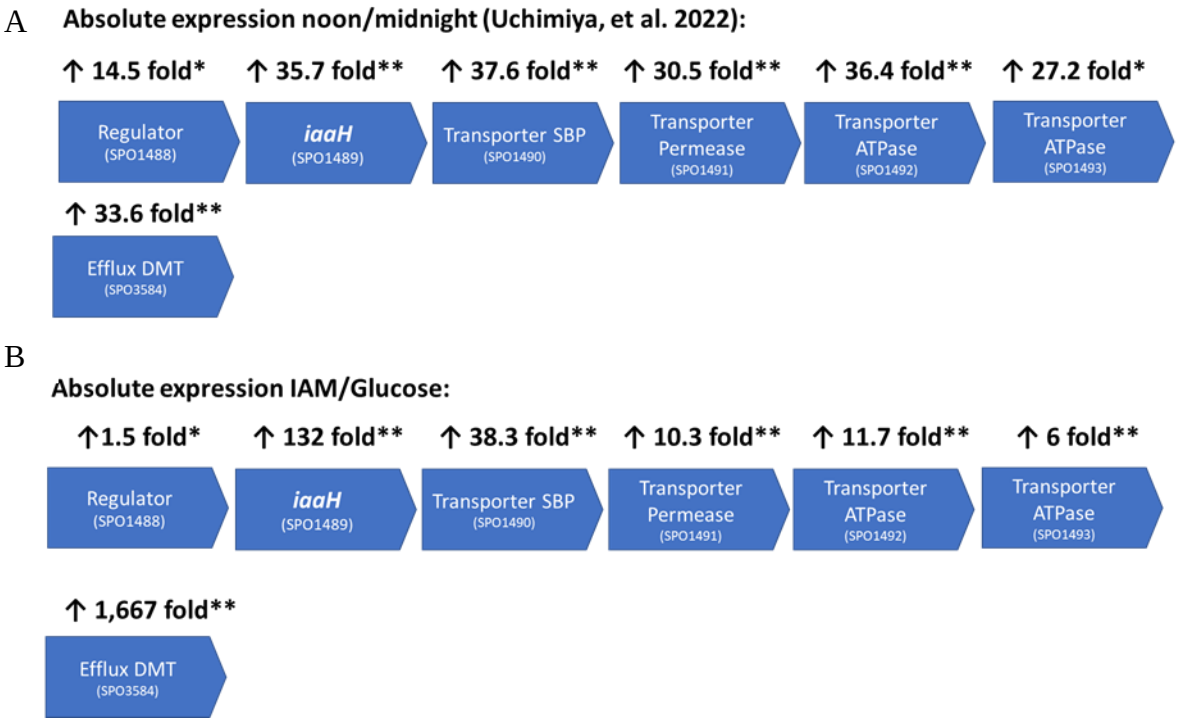


Figure 4.2. Absolute differential expression of *R. pomeroyi* of genes in the *iaaH* operon, including a regulator, IAM hydrolase, and putative IAM transporter, and for a putative IAA exporter. A) Diel expression in a co-culture with T. pseudonana, shown for noon vs midnight (Uchimiya et al., 2022). B) Differential expression when grown on IAM vs glucose as sole carbon sources. * adjusted $p < 0.01$, ** adjusted $p < 0.001$.

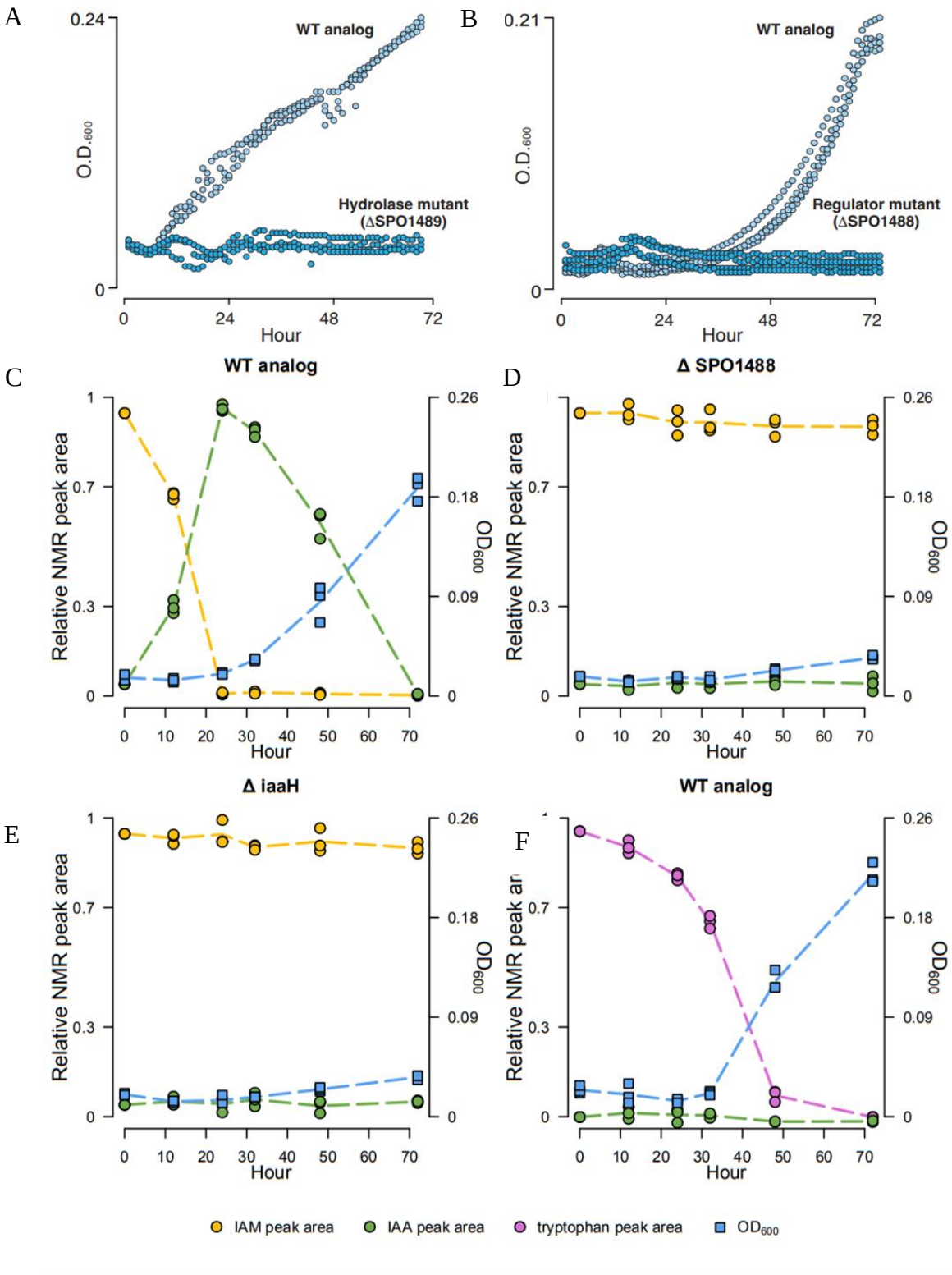


Figure 4.3. Mutants of *iaaH* (SPO1489) and the *iaaH* operon regulator (SPO1488) and WT analog (pooled_RB-TnSeq) with IAM as the sole carbon source. A-B) Growth curves measured

as OD_{600} show severe growth defect of both mutants, relative to WT analog. C-F) Paired NMR exometabolite analysis and OD_{600} of cultures grown with either IAM or tryptophan as sole carbon source. Values presented are relative IAM peak area (yellow circles and lines), relative IAA peak area (green circles and lines), relative tryptophan peak area (purple circles and lines), and OD_{600} (blue squares and lines). C) In growth on IAM, WT analog draws down of IAM, synthesizes IAA, and releases it extracellularly. When IAM is fully depleted, released IAA is taken back into the cell and growth begins. D-E) The IAM hydrolase and regulator mutants are unable to consume IAM and release IAA. F) During WT analog growth on tryptophan, no IAA release was detected.

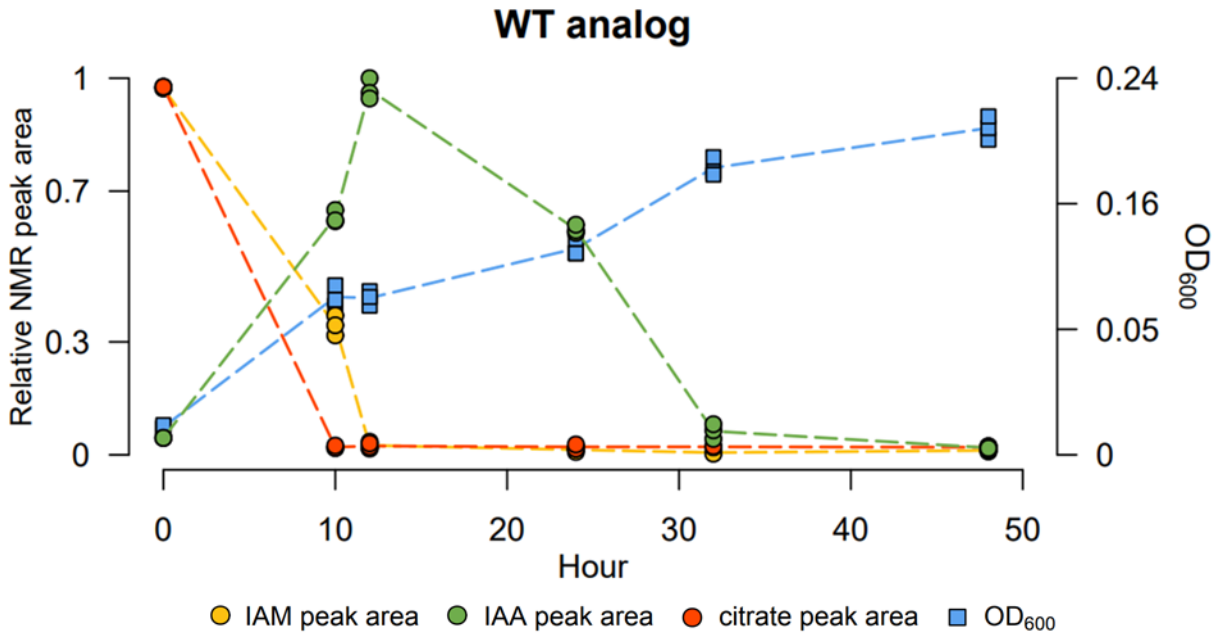


Figure 4.4. Paired NMR exometabolite analysis and OD_{600} of WT analog grown with either IAM or tryptophan as sole carbon source. Values presented are relative IAM peak area (yellow circles and lines), relative IAA peak area (green circles and lines), relative citrate peak area (red circles and lines), and OD_{600} (blue squares and lines). Starting concentrations were 4 mM C citrate and 8 mM C IAM, relative NMR peak area of each is scaled to 1. *R. pomeroyi* simultaneously uses citrate for growth and draws down IAM, producing and releasing IAA. Growth stalls when citrate is depleted while remaining IAM is converted to IAA, when IAM is depleted growth on IAA resumes.

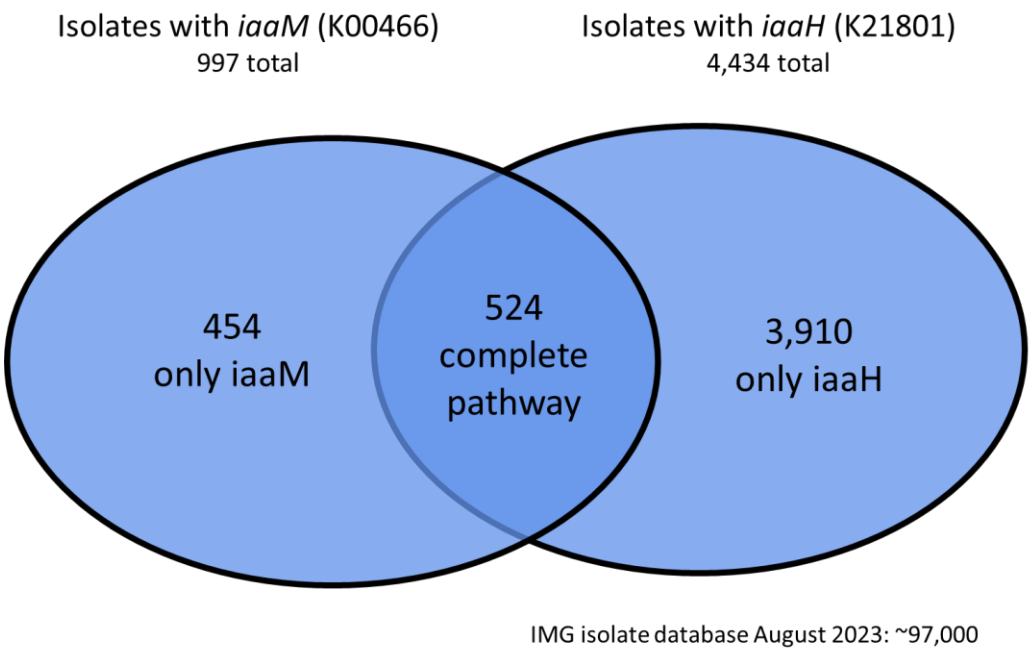
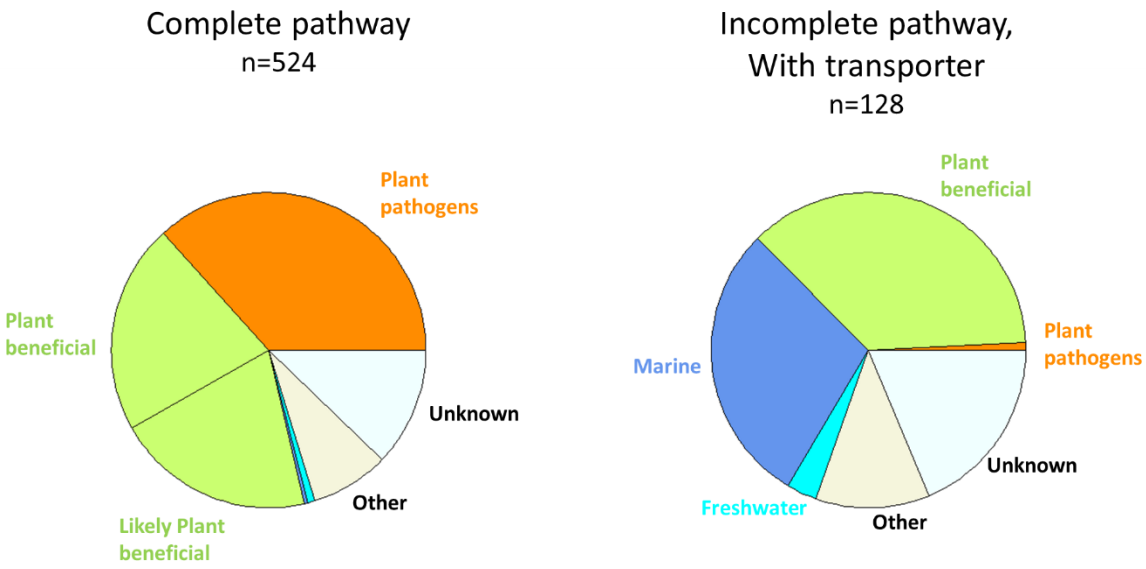


Figure 4.5. The majority of isolates with IAM pathway genes have incomplete pathways, with the most common condition the presence of *iaaH* only, similar to *R. pomeroyi*. Isolate information and KO functional assignments were retrieved from JGI's IMG workspace.

A



B

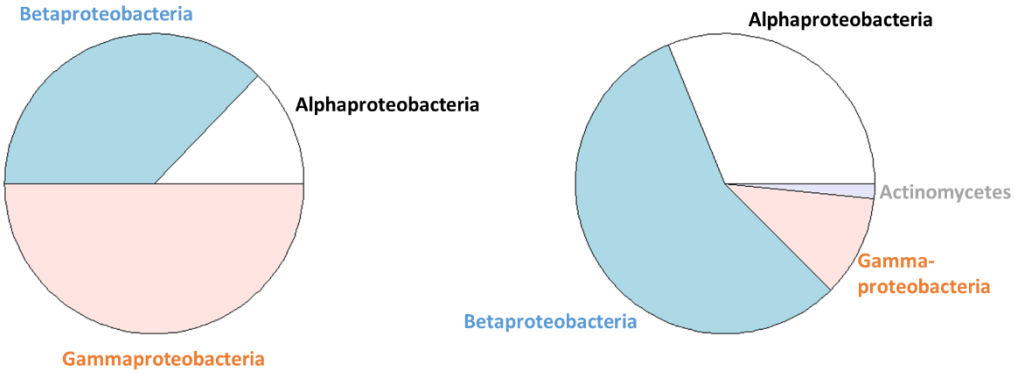


Figure 4.6. A) Ecological context of isolates containing the complete IAM pathway verses those with *iaaH* and a gene neighborhood similar to that of *R. pomeroyi*'s operon, including homologs of the putative IAM transporter components. B) Class level taxonomic breakdown of each group; *R. pomeroyi* is a member of the Alphaproteobacteria.

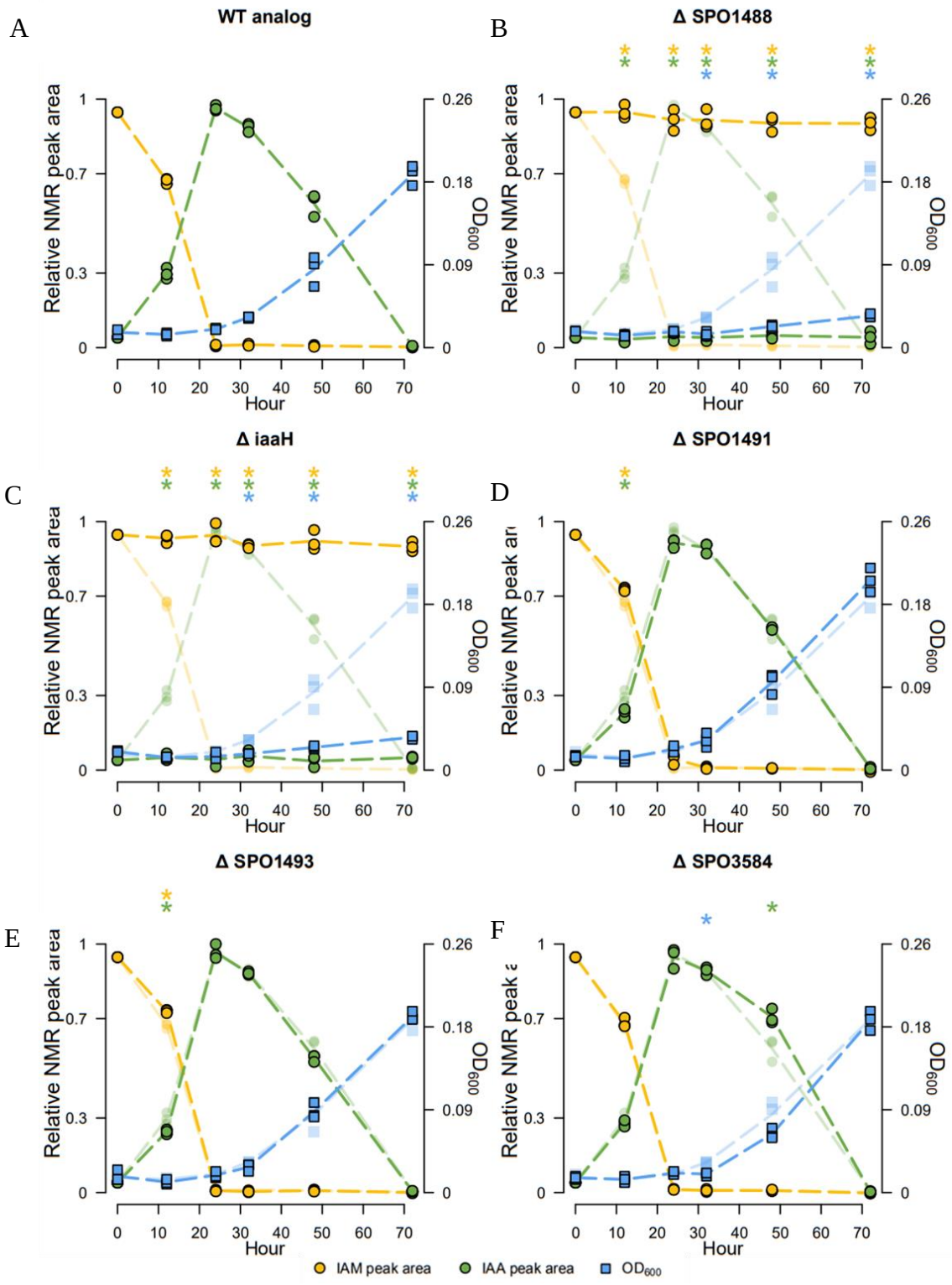


Figure S4.1. Paired NMR exometabolite analysis and OD_{600} of cultures grown with IAM as sole carbon source. Values presented are relative IAM peak area (yellow circles and lines), relative

IAA peak area (green circles and lines), and OD_{600} (blue squares and lines). A) WT analog growth on IAM, reproduced from Fig. 3. B-F) Isolated mutants grown on IAM, transparent lines and points show the WT analog values for comparison. Asterisks above plot indicate a significant difference (ANOVA and TukeyHSD, p.adj < 0.05) between the WT analog values and mutant values at the corresponding time point. The color of asterisk indicates the measured viable that differed (eg. blue asterisk indicates significant difference in OD_{600} between WT analog and mutant).

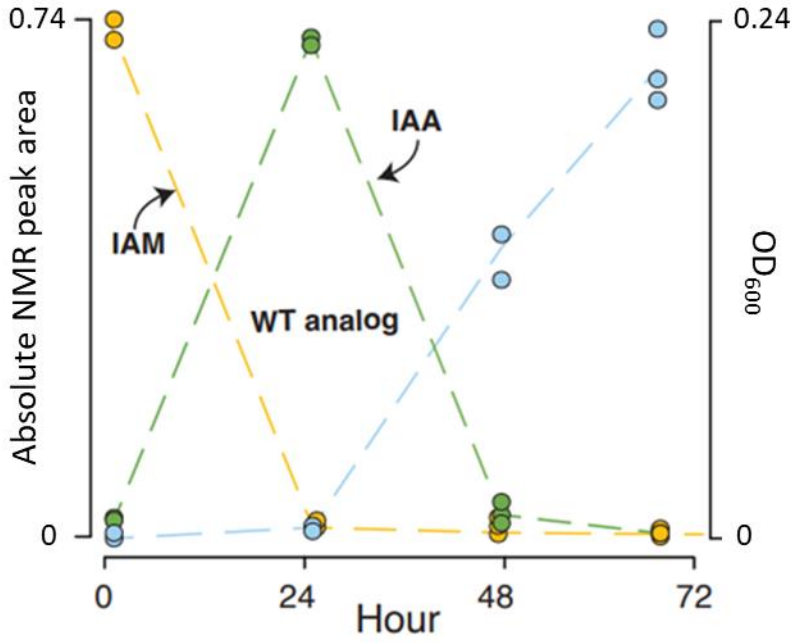


Figure S4.2. Paired NMR exometabolite analysis and OD_{600} of cultures grown with IAM as sole carbon source. Values presented are relative IAM peak area (yellow circles and lines), relative IAA peak area (green circles and lines), and OD_{600} (blue circles and lines). From lower time resolution experiment prior the NMR data shown in Figs 3,4,S1. Here absolute NMR peak area is provided, as opposed to relative peak area in prior figures. Absolute peak areas of IAM and IAA show that IAA is produced in an approximately 1:1 molar ratio from IAM.

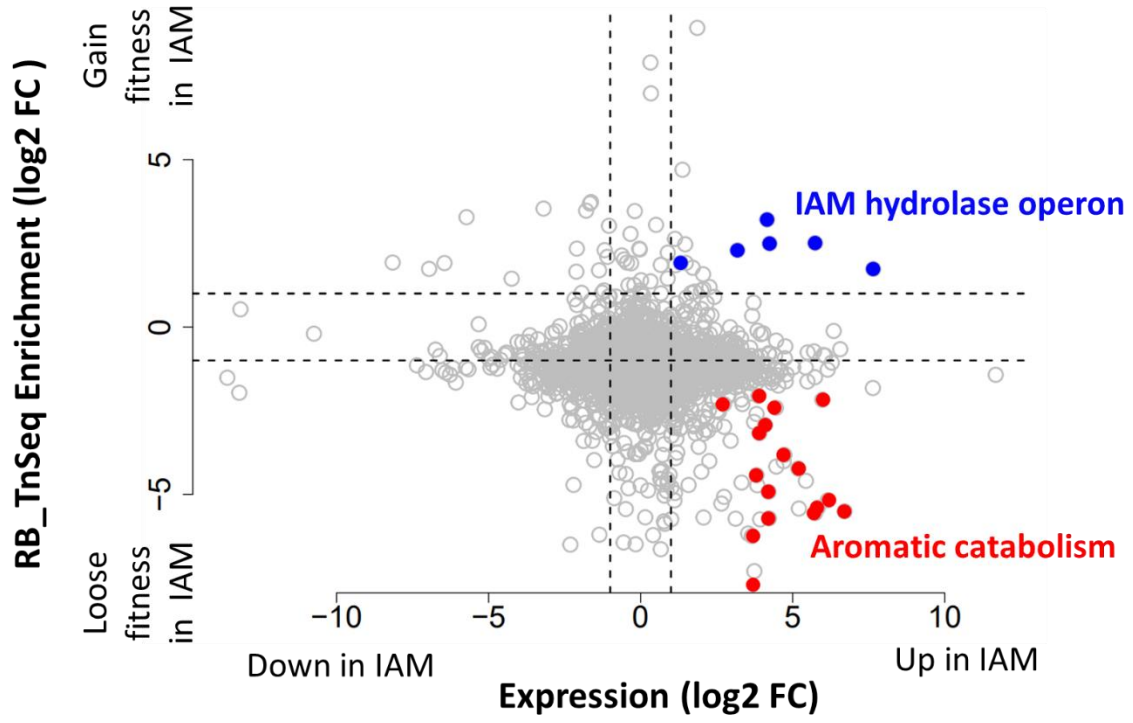


Figure S4.3. RB-TnSeq enrichment (Y axis) and transcript expression (x axis) for the *R. pomeroyi* BarSeq pool grown in IAM versus glucose. Log_2 RB_TnSeq enrichment represents relative fitness of each mutant between the two conditions, with positive values indicating the transposon mutant had increased fitness in IAM, and negative values indicating the mutant had decreased fitness in IAM. Log_2 differential expression represents genes that change in relative transcription when growing on IAM vs glucose, with positive values indicating increased expression in IAM, and negative values indicating genes with decreased expression in IAM. Blue circles indicate the genes of the *iaaH* operon. Red circles indicate genes involved with catabolism of aromatic compounds.

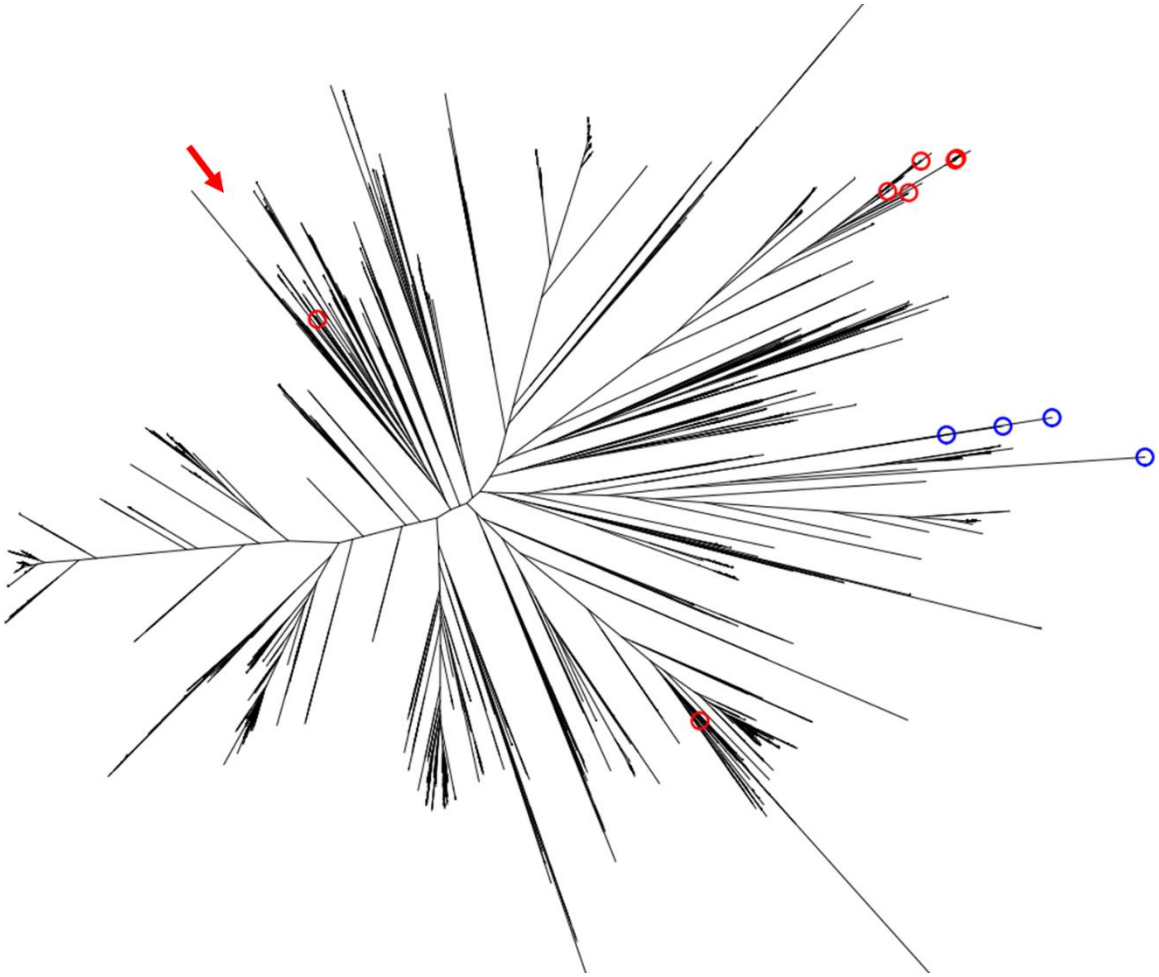


Figure S4.4. Gene Phylogeny (FastTree) of all iaaH genes. Red circles indicate confirmed iaaH genes. Blue Circles indicate outgroup sequences. Red arrow indicated the location of *R. pomeroyi*'s iaaH gene.

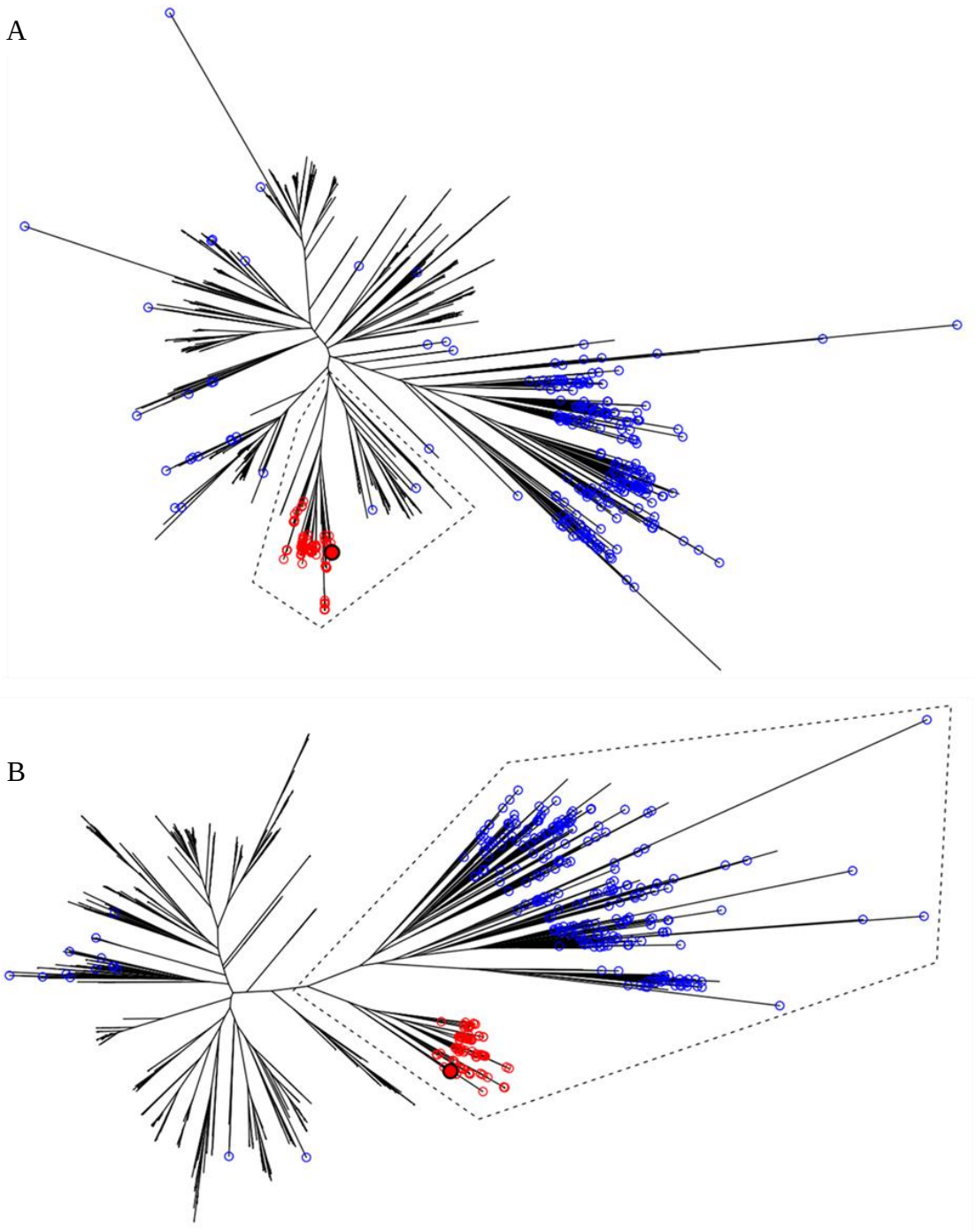


Figure S4.5. Gene phylogeny (FastTree) of homologs of putative IAM transporter ATPase components SPO1492 (a) and SPO1493 Red circle denote confidently annotated homologs of *R.*

pomeroyi's components, these genes meet fout criteria 1) BlastP e-val < 10^{-30}, 2) gene in well defined clade with *R. pomeroyi* SPO1492/SPO1493, distinct from outgroups, 3) both components of each transporter present in their corresponding clades, 4) gene in gene neighborhood with iaaH homolog. Solid red points indicate the position of *R. pomeroyi* SPO1492/SPO1493 sequences. Blue points indicate verified outgroup transporters from a transporter reference database. Dashed line indicates the extracted clades shown in Fig. S6.

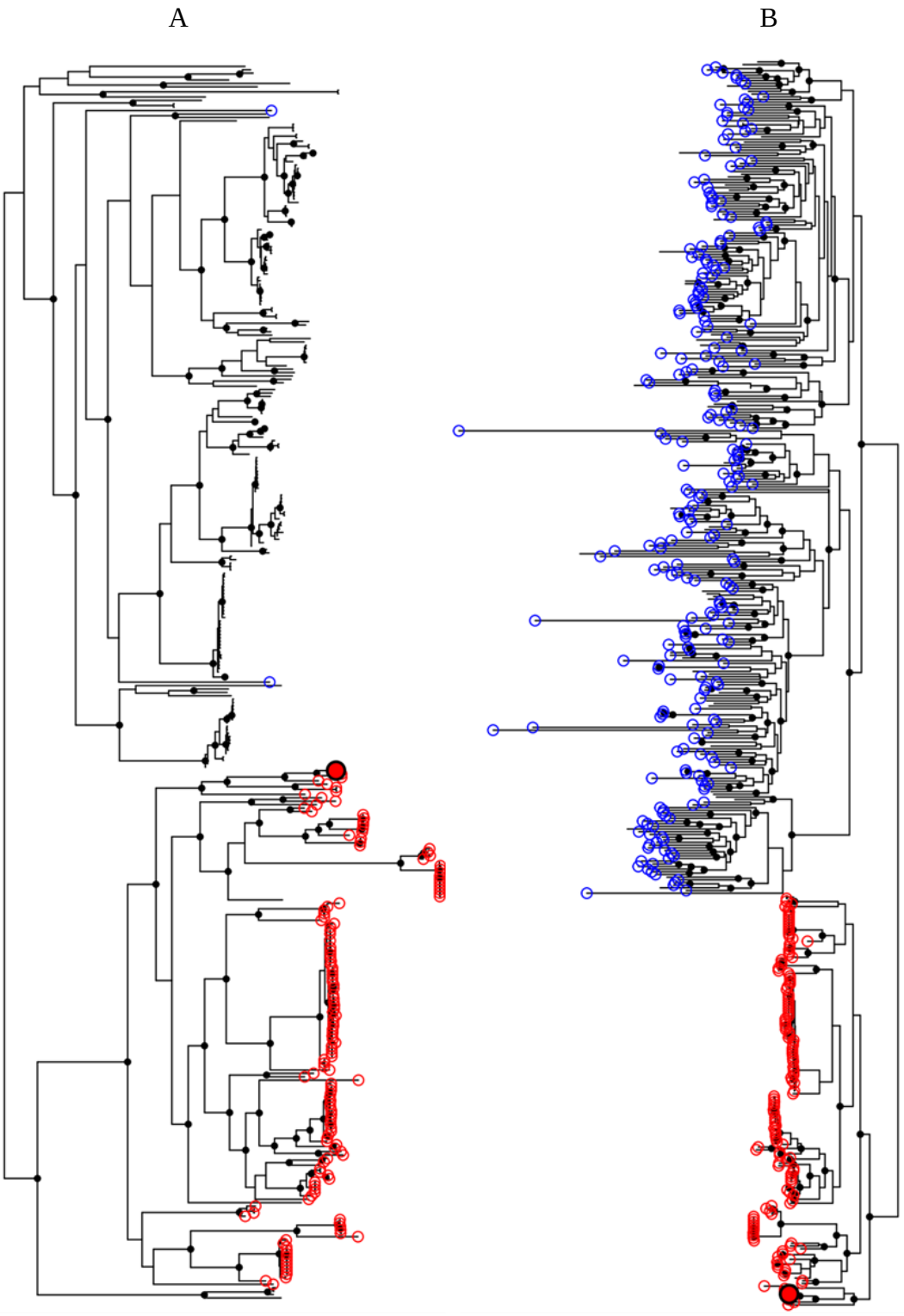


Figure S4.6. Gene phylogeny (FastTree, extracted clades) of homologs of putative IAM transporter ATPase components SPO1492 (a) and SPO1493 (b), see Fig. 4.5 for full trees. Black

circles at nodes indicate FastTree predicted bootstrap values > 0.9. Red circle denote confidently annotated homologs of *R. pomeroyi*'s components, these genes meet three criteria 1) BlastP e-val < 10^{-30}, 2) gene in well defined clade with *R. pomeroyi* SPO1492/SPO1493, distinct from outgroups, 3) both components of each transporter present in their corresponding clades, 4) gene in gene neighborhood with iaaH homolog Solid red points indicate the position of *R. pomeroyi* SPO1492/SPO1493 sequences. Blue points indicate verified outgroup transporters from a transporter reference database.

K00466) and *iaaH* (K21801

Table S4.1. Combined R. pomeroyi expression and RB_TnSeq data, represented as fold change between growth on IAM and growth on Glucose as sole carbon sources. This table is archived at Zenodo and is publicly available (https://doi.org/10.5281/zenodo.10032011).

Table S4.2. Confirmed *iaaH* sequences from literature with source information and amino acid sequence. This table is archived at Zenodo and is publicly available (https://doi.org/10.5281/zenodo.10032011).

Table S4.3. All IMG isolate genomes containing *iaaH* (K00466). This table is archived at Zenodo and is publicly available (https://doi.org/10.5281/zenodo.10032011).

Table S4.4. All IMG isolate genomes containing *iaaM* (K21801). This table is archived at Zenodo and is publicly available (https://doi.org/10.5281/zenodo.10032011).

Table S4.5. All *iaaH* genes in the IMG database. This table is archived at Zenodo and is publicly available (https://doi.org/10.5281/zenodo.10032011).

Table S4.6. All *iaaM* genes in the IMG database. This table is archived at Zenodo and is publicly available (https://doi.org/10.5281/zenodo.10032011).

Table S4.7. All isolate genomes lacking iaaM, containing iaaH, and containing transporter homologs for SPO1492 and SPO1493. From IMG, with metadata, and manually assigned ecological context. This table is archived at Zenodo and is publicly available (https://doi.org/10.5281/zenodo.10032011).

Table S4.8. All isolate genomes containing iaaM and iaaH. From IMG, with metadata, and manually assigned ecological context. This table is archived at Zenodo and is publicly available (https://doi.org/10.5281/zenodo.10032011).

Table S4.9. All homologs of transporter ATPase components SPO1492 and SPO1493. This table is archived at Zenodo and is publicly available (https://doi.org/10.5281/zenodo.10032011).

CHAPTER 5

SUMMARY

Marine microbial communities play a crucial role in major steps of the ocean carbon cycle (1-4). The fate of a substantial portion of marine carbon is determined by the fluxes of individual metabolites between individual microbes in a complex network that dominates the surface ocean (3). To understand the nature and extent of these exchanges it is vital to understand the genes mediating the uptake and utilization of key metabolites. Here I annotate, and gain ecological insight from, a suite of genes identified in the model marine bacterium *Ruegeria pomeroyi* DSS-3 that are involved in microbe-microbe interactions.

In Chapter 2, an arrayed random barcode transposon insertion library (RB-TnSeq) of *R. pomeroyi* was used for high throughput characterization of thirteen organic influx transporters. Four of the transporters characterized were previously, with their cognate substrates hypothesized based on gene expression data; five were previously with their substrates hypothesized based on homology to experimentally annotated transporters in other bacteria; and four had no previous annotations. This work brings the total number of experimentally-verified organic carbon influx transporters to 18 of 126 in the *R. pomeroyi* genome. In a longitudinal study of a coastal phytoplankton bloom, expression patterns of the experimentally annotated transporters linked them to different stages of the bloom, and led to the hypothesis that citrate and 3-hydroxybutyrate were among the most highly available bacterial substrates in this bloom.

Chapter 3 builds on the findings of Chapter 2, using *R. pomeroyi's* verified transporter annotations to develop a phylogeny-based approach to identifying the substrates of transporters in environmental metagenomic data. Gene phylogenies were generated from the component sequences of *R. pomeroyi*'s ABC transporters as well as a curated list of reference transporters. I found that gene phylogenies constructed from the ATPase transporter component could most accurately and consistently predict cognate substrates. We applied this approach to environmental sequences, recruiting taurine transporter ATPase (*tauB*) genes from Tara Oceans metagenomes. Phylogenies of recruited Tara sequences and reference sequences showed that a lineage of environmental *tauB* genes could be readily distinguished from outgroup sequences. Unfortunately, most of the sequences recruited by more standard approaches (sequence identity only, no gene phylogenies) were transporters of other substrates, emphasizing the risk of significant errors when using common gene annotation tools.

In Chapter 4, I focus on the role of the auxin type phytohormone indole acetic acid (IAA) in phycosphere interactions. IAA, well studied in plant - microbe interactions, can facilitate both mutualistic and pathogenic interactions, but its roles in marine phytoplankton-bacteria interactions are less well understood. I characterize the function of *R pomeroyi*'s *iaaH* gene, a key enzyme in IAA biosynthesis, and demonstrate that although *R. pomeroyi* has an incomplete IAA biosynthesis pathway, it produces IAA when supplemented with the intermediate indole-3-acetamide (IAM). When IAM is provided as a sole carbon source, *R. pomeroyi* prioritizes IAA biosynthesis and immediate release from the cell, demonstrating IAA's role as an exogenous signal molecule and suggesting that its release is controlled by the supply of IAM from other microbial cells. Co-culture expression data suggests that IAM is made available to *R. pomeroyi* during co-culture with the diatom *Thalassiosira pseudonana*, and when returned to seawater

would be available to *T. pseudonana*. A search of isolate genomes reveals an incomplete IAA biosynthesis pathway similar to that in *R. pomeroyi* is common in environmental bacteria, suggesting that autotroph-bacterial exchange of IAM is prevalent in marine environments.

The nuances and complexities of microbial interaction are vast, but recent advances in genetic and sequencing tools, chemical methods, and computational resources have made it possible to begin disentangling these networks of interactions (5). My dissertation research provides a framework for how interdisciplinary methods can be brought together to better understand the mechanisms of microbial interaction.

References

1. Azam F, Fenchel T, Field J, Grey J, Meyer-Reil L, Thingstad F. The ecological role of water-column microbes. Mar Ecol Prog Ser. 1983;10:257-63.

2. Anderson TR, Ducklow HW. Microbial loop carbon cycling in ocean environments studied using a simple steady-state model. Aquat Microb Ecol. 2001;26:37-49.

3. Moran MA, Ferrer-González FX, Fu H, Nowinski B, Olofsson M, Powers MA, et al. The ocean's labile DOC supply chain. Limnol Oceanogr. 2022;67:1007-21.

4. Cole JJ, Likens GE, Strayer DL. Photosynthetically produced dissolved organic carbon: An important carbon source for planktonic bacteria. Limnol Oceanogr. 1982;27:1080-90.

5. Moran MA, Kujawinski EB, Schroer WF, Amin SA, Bates NR, Bertrand EM, et al. Microbial metabolites in the marine carbon cycle. Nat Microbiol. 2022;7:508-23.

APENDIX A

CHAPTER 2 SUPPLEMENTAL MATERIAL[1]

[1]Schroer, WF; Kepner, HE; Uchimiya, M; Mejia, C; Trujillo Rodriguez, L; Reisch, CR; Moran, MA. 2023. *ISME Communications*. 3(1): 37. Reprinted here with permission of the publisher.

Dataset S2.1. High Resolution Growth Screen of *Ruegeria pomeroyi* Transporter Mutants Data. This dataset is archived at BCO-DMO and is publicly available (https://www.bco-dmo.org/dataset/894179, 10.26008/1912/bco-dmo.894179.1).

Dataset S2.2. High throughput growth screening of the marine bacterium *Ruegeria pomeroyi* DSS-3 knockout mutants on 70 environmentally relevant marine substrates. This dataset is archived at BCO-DMO and is publicly available (https://www.bco-dmo.org/dataset/904246, 10.26008/1912/bco-dmo.904246.1).

Dataset S2.3. *Ruegeria pomeroyi* RB-TnSeq Transposon Mutant Library Screen Experimental Data. This dataset is archived at BCO-DMO and is publicly available (https://www.bco-dmo.org/dataset/893256, DOI:10.26008/1912/bco-dmo.893256.1).

APENDIX B

CHAPTER 4 SUPPLEMENTAL MATERIAL[1]

[1]Schroer, WF; Uchimiya, M; Hamilton, MM; Smith, CR; Olofsson, M; Moran, MA. To be submitted to *ISME J*.

Dataset S4.1. Optical density (600nm) and NMR normalized peak high for growth of *R. pomeroyi* WT analog (RB-TnSeq, pool) and mutants grown on IAM, L-tryptophan, or a mix of IAM and citrate, ~12 hr sampling resolution. This dataset is archived at Zenodo and is publicly available (https://doi.org/10.5281/zenodo.10037727).

Dataset S4.2. Optical density (600nm) and for growth of *R. pomeroyi* WT analog (RB-TnSeq, pool) and mutants grown on IAM, 1 hour sampling resolution. This dataset is archived at Zenodo and is publicly available (https://doi.org/10.5281/zenodo.10037727).

Dataset S4.3. *Ruegeria pomeroyi* RB-TnSeq Transposon Mutant Library Screen Experimental Data. This dataset is archived at BCO-DMO and is publicly available (https://www.bco-dmo.org/dataset/893256, DOI:10.26008/1912/bco-dmo.893256.1).