UNIVERSITY OF CALIFORNIA SAN DIEGO

Genomics-Based Investigations of Algae-Bacteria Interactions

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy

in

Oceanography

by

Rachel E. Diner

Committee in charge:

Andrew Allen, Chair
Eric Allen
Lihini Aluwihare
Douglas Bartlett
Bianca Brahamsha
Susan Golden

2019

The Dissertation of Rachel E. Diner is approved, and it is acceptable in quality and in form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

_____

Chair

University of California San Diego

2019

DEDICATION

To Elie and Wylie Diner for giving me perspective, purpose, and above all, love.

TABLE OF CONTENTS

LIST OF TABLES

Chapter 1

Chapter 4

LIST OF FIGURES

Chapter 1

Chapter 2

Chapter 3

Chapter 4

ACKNOWLEDGMENTS

Many people helped make this dissertation possible through their support, assistance, and guidance. I would like to first gratefully acknowledge my advisor and committee chair, Dr. Andrew Allen. Andy's passion for science and for life is truly contagious. In addition to giving me the opportunity of a lifetime- to get a Ph.D. at Scripps- he consistently provided positive encouragement, constructive critiques, and overwhelming support for my professional and personal endeavors. Andy has inspired me to conduct excellent science, to think creatively, and to push the boundaries of what is possible. Perhaps most importantly, his unfailing belief in me, even in the face of challenging life transitions such as becoming a parent, has given me the courage to believe in myself.

Thank you to my committee members for providing excellent feedback and advice throughout my time at UCSD. Each of you has been an incredible inspiration to me, and will continue to be so for the rest of my career. I truly appreciate the time taken from your busy schedules to mentor me and give me personal and professional advice. Thank you to Peter Franks who provided me with inspiration, motivation, and laughs throughout my time at SIO. Thank you to Moselio Schaecter for teaching me about the magic of microbes and the joys of writing about them.

I'd like to thank the amazing research community at the J. Craig Venter Institute. Phil Weyman taught me how to be a microbiologist, and I wouldn't be the scientist I am today without his countless hours of training, mentorship, and support. Three amazing women, Lisa Zeigler Allen, Sinem Beyhan, and Anna Edlund, have been incredible role models and friends to me, helping me through challenging life events and inspiring me to be a badass woman

scientist like them. Additionally, Bogumil Karas and Erin Bertrand had a tremendously positive impact on me, scientifically and personally.

Thank you to the members of the Allen Lab, past and present, who are and have always been like a family to me. I can't say enough about each and every one of you- thank you for all of the amazing times: Hong Zheng, Ariel Rabines, Flip McCarthy, Sarah Smith, John McCrow, Jeff McQuaid, Tyler Coale, Maxine Tan, Mark Moosburner, Angela Zoumplis, Bethany Kolody, Jernej Turnsek, Sarah Schwenck, Patrick Brunson, and the numerous "young folk." Thank you to my Marine Biology cohort at SIO and all of the amazing colleagues I have met over the years.

Thank you to my collaborators at the Southern California Coastal Water Research project, particularly Joshua Steele and John Griffith, for training, inspiration, and great conversation. Thank you to the vibrant Scripps community, and the support staff who work tirelessly to help us in our academic journeys.

Thank you to two of my dearest and most inspiring female scientist friends: Sarah Smith and Nati Delherbe. You've made and continue to make this life journey amazing for me. I look forward to many more years of friendship and nerdiness.

Lastly, I would like to thank my family. I love science immensely, but without my family, I am absolutely certain it would mean nothing to me. Thank you to my parents, Hy and Patti Dorfman, my sister Andi Dorfman, may mother and father-in-law Judy and Bruce Diner, and my brother-in-law Ben Diner. Your support means so much to me.

Thank you to my amazing supportive husband Elie, for always believing in me and helping me to believe in myself. I would not be here without you, and I wouldn't want to be anywhere else. And thank you to our beautiful son Wylie for making my life feel complete.

Chapter 1, in full, is a reprint of the material as it appears in **Diner, RE**, Schwenck, SM, McCrow, JP, Zheng, H, & Allen, AE (2016) Genetic Manipulation of Competition for Nitrate between Heterotrophic Bacteria and Diatoms. *Frontiers in Microbiology*, 7, 880. The dissertation author was the primary investigator and author of this material.

Chapter 2, in full, is a reprint of the material as it appears in **Diner, RE**, Bielinski, VA, Dupont CP, Allen, AE, Weyman, PW (2016) Refinement of the Diatom Episome Maintenance Sequence and Improvement of Conjugation-based DNA Delivery Methods. *Frontiers in Biotechnology and Bioengineering.* 4, 65. The dissertation author was the primary investigator and author of this material.

Chapter 3, in full, is a reprint of the material as it appears in **Diner, RE**, Noddings, CM, Lian, NC, Kang, AK, McQuaid, JB, Jablanovic, J, Espinoza, JL, Nguyen, NA, Anzelmatti, MA, Jansson, J, Bielinski, VA., Karas, BJ, Dupont, CL, Allen, AE, and Weyman, PD (2017) Diatom Centromeres Suggest a Mechanism for Nuclear Gene Acquisition. *Proceedings of the National Academy of Sciences*, *114*(29), E6015-E6024. The dissertation author was the primary investigator and author of this material.

Chapter 4, in part, is currently being prepared for submission for publication of the material in **Diner, RE,** Rabines, A., Zheng, H., Allen, A.E. High-resolution taxonomic grouping reveals interactions between pathogenic *Vibrio* species and the planktonic community. The dissertation author was the primary investigator and author of this material.

VITA

2006 Bachelor of Science, Biology, University of Georgia

2009 Juris Doctor, University of San Diego School of Law

2013 Master of Science, Biology with a concentration in Marine Biology, Romberg Tiburon Center for Environmental Studies, San Francisco State University

2019 Doctor of Philosophy, Oceanography, Scripps Institution of Oceanography, University of California San Diego

PUBLICATIONS

Diner, RE, Noddings, CM, Lian, NC, Kang, AK, McQuaid, JB, Jablanovic, J, Espinoza, JL, Nguyen, NA, Anzelmatti, MA, Jansson, J, Bielinski, VA., Karas, BJ, Dupont, CL, Allen, AE, and Weyman, PD (2017) Diatom Centromeres Suggest a Mechanism for Nuclear Gene Acquisition. *Proceedings of the National Academy of Sciences*, *114*(29), E6015-E6024.

Diner, RE, Bielinski, VA, Dupont CP, Allen, AE, Weyman, PW (2016) Refinement of the Diatom Episome Maintenance Sequence and Improvement of Conjugation-based DNA Delivery Methods. *Frontiers in Biotechnology and Bioengineering.* 4, 65.

Diner, RE, Schwenck, SM, McCrow, JP, Zheng, H, & Allen, AE (2016) Genetic Manipulation of Competition for Nitrate between Heterotrophic Bacteria and Diatoms. *Frontiers in Microbiology*, 7, 880.

Cavole, LM, Demko, AM, **Diner, RE**, Giddings, A, Koester, I, Pagniello, CMLS, Paulsen, M-L, Ramirez-Valdez, A, Schwenck, SM, Yen, KY, Zill, ME, Franks, PJS (2016) Biological Impacts of 2013-2015 Warm-Water Anomalies in the Northeast Pacific: Winners, Losers and the Future. *Oceanography* 29(2), 273-265.

Karas, BJ, **Diner, RE,** Lefebvre SC, McQuaid, J, Phillips, APR, Noddings, CM, Brunson, JK, Valas RE, Deerinck, TJ, Jablanovic, J, Gillard, JTF, Beeri, K, Ellisman, MH, Glass, JI, Hutchinson III, CA, Smith, HO, Venter, JC, Allen, AE, Dupont, CL, Weyman, PD (2015) Designer diatom episomes delivered by bacterial conjugation. *Nature Communications,* 6, 6925.

Diner, RE, Benner, I, Passow, U, Komada, T, Carpenter, EJ, Stillman, JH (2015) Negative effects of ocean acidification on calcification vary within the coccolithophore genus *Calcidiscus*. *Marine Biology*, 1-19, 1287-1305.

Benner, I, **Diner, RE**, Lefebvre, SC, Li, D, Komada, T, Carpenter, EJ, & Stillman, JH (2013) *Emiliania huxleyi* increases calcification but not expression of calcification-related genes in long-term exposure to elevated temperature and *p*CO2. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1627), 20130049.

ABSTRACT OF THE DISSERTATION

Genomics-Based Investigations of Marine Algae-Bacteria Interactions

by

Rachel E. Diner

Doctor of Philosophy in Oceanography

University of California San Diego, 2019

Professor Andrew E. Allen, Chair

Interactions between eukaryotic algae and bacteria play an important role in natural ecosystems. Defining the details of these interactions enables a better understanding of organismal distribution and evolution, and also presents an opportunity to further human well-being via biotechnology and protect human health. In this thesis I utilize genomic techniques to elucidate interactions between bacteria and algae in the laboratory and the field. I demonstrate dynamic carbon and nitrogen-dependent interactions between model marine algae and bacteria in a newly developed genetically tractable model laboratory system. I then describe how horizontal gene transfer (HGT) from bacteria to diatoms can be used as a molecular tool for diatom genetic manipulation. The low-GC content of transferred DNA

sequences enables autonomous replication as a diatom episome, effectively expanding the diatom's gene repertoire and providing opportunities for nuclear genome integration. Lastly, I discuss associations between pathogenic species of *Vibrio* bacteria along the San Diego coast and their abundant algal counterparts. I report the first quantitative survey of pathogenic *Vibrio* species in San Diego coastal waters, which are abundant during summer months and possess genes associated with human virulence. When examining the ecological interactions of these species, traditional grouping of diatoms at a high taxonomic level has led to conflicting reports of associations with pathogenic *Vibrio* species. I show that high-resolution taxonomic grouping at the genus level or lower, based on 18S amplicon sequencing, reveals specific interactions that may have important consequences for *Vibrio* ecology and human health, yet would have been overlooked in previous studies. Together, these chapters demonstrate how new molecular tools, including next-generation sequencing, can be used to gain a deeper understand of microbial interactions that are ecologically important on a global scale and also important to human health and well-being.

INTRODUCTION

Microbes inhabit virtually every region of the planet, from the deepest ocean basins to the clouds in our atmosphere. Attached and free-living, dynamic microbial ecosystems drive nutrient and biogeochemical cycling on global scales and form the base of nearly all food webs[1–3]. Furthermore, human health and well-being is profoundly impacted by microbes. Microbial pathogens cause up to a quarter of all global deaths each year[4]. Individual microbes rarely exist in isolation but as part of complex systems that form and change across myriad spatial and temporal scales, interacting with each other and with multicellular counterparts.

Marine microbes, including bacteria, viruses, and unicellular photosynthetic and heterotrophic protists, are present in virtually every biological niche of the ocean. In the euphotic zone, phytoplankton interact frequently with heterotrophic bacteria in ways that are biogeochemically important on a global scale, yet poorly understood[5–7]. Eukaryotic algae known as phytoplankton contribute substantially to global carbon cycling via photosynthesis and primary productivity, supporting higher trophic levels and producing about 50% of Earth's oxygen[8]. Diatoms alone are responsible for up to 40% of the ocean's primary productivity. Heterotrophic bacteria, which constitute the majority of oceanic biomass[9–11], live around, on, and occasionally within these phytoplankton, relying on them for essential carbon and nutrients[12–14].

Given the great importance of algae-bacteria interactions, relatively little is known about how they work at the molecular level. Recent work has highlighted many types of relationships (e.g. mutualistic, predatory, competitive, commensal) and exchanges of small molecules and metabolites between these groups (reviewed in [6]). For example, laboratory studies have shown that bacteria can stimulate phytoplankton growth via hormone signaling[15]

and provide or compete for essential nutrients such as nitrogen[16,17], sulfonate compounds[18], and vitamin B12[19]. They can also protect against high levels of reactive oxygen species[20], help certain phytoplankton function during prolonged period of darkness[21], or in some cases turn algicidal[22]. These studies are valuable, but probe only a limited number of the expansive potential metabolic exchanges and interactions on limited temporal and spatial scales. Furthermore, it is unclear if and to what extent these interactions occur in natural environments.

In this thesis, I explore a myriad of new relationships between algae, primarily diatoms, and heterotrophic bacteria. My research approach includes both controlled laboratory studies exploring relationships between two individual species and natural environments comprised of complex communities relevant to human health. A common theme in my research is the use of genomic information, at the species and whole-community level, to elucidate these novel interactions. From molecular genetic engineering to next-generation sequencing (including RNA-sequencing, CHiP-sequencing, amplicon sequencing, and shotgun metagenomic sequencing) these novel tools and organismal information enable unprecedented explorations in to what cells do, who they interact with, and how.

In Chapter 1 of my thesis, I used microbial co-culturing, genetic manipulation, and transcriptomics to examine how a ubiquitous marine gammaproteobacterial species, *Alteromonas macleodii*, can compete with diatoms (specifically, the model diatom *Phaeodactylum tricornutum*) for nitrate in a carbon-dependent manner. I also use nitrate reductase knockout mutants of both diatoms and bacteria to demonstrate that diatoms and bacteria likely exchange nitrogen substrates under certain conditions. A large accomplishment of this research was developing a genetically tractable model system for studying diatom-

bacteria interactions. This research was published in Diner et al. 2016 (*Frontiers in Microbiology*).

In chapters 2 and 3 of my thesis, I examined the horizontal transfer of genes from bacteria to diatoms via bacterial conjugation, and explored the implications of this transfer in terms of diatom nuclear gene acquisition and genetic tool development. After helping develop a method of delivering DNA into the diatoms *P. tricornutum* and *Thalassiosira pseudonana* using bacterial conjugation from the common laboratory model bacterium *E. coli* (Karas, Diner et al. 2015), we discovered an interesting feature of this DNA delivery system: foreign DNA containing yeast centromere and origin of replication sequences (an artifact of the cloning process) allowed delivered plasmids to replicate autonomously in the diatom nucleus.

In chapter 2, I demonstrated that low GC content was the characteristic feature responsible for stable maintenance of the foreign DNA, and that the yeast centromere and origin of replication do not possess orthologous functions in the diatom. I also presented advancements in technical methods advancing the use of bacterial conjugation as a transgene delivery tool for diatoms. This research was published in Diner et al. 2016 (*Frontiers in Biotechnology and Bioengineering*).

In chapter 3, I describe the sequence identity of native diatom centromeres, the first description of centromeres in the stramenopile lineage. I also demonstrate that DNA sequence similarity to native diatom centromeres allows DNA from many different sources, including bacterial conjugative plasmids and natural diatom plasmids, to become established as part of the diatom nuclear genome repertoire after being delivered by *E. coli* bacterial conjugation, essentially "hijacking" the diatom DNA replication machinery. This research was published in Diner et al. 2017 (*Proceedings of the National Academy of Sciences*)

In chapter 4, I examined associations between pathogenic *Vibrio* bacteria along the San Diego coast and their abundant algal counterparts. I report the first quantitative survey of pathogenic *Vibrio* species in San Diego coastal waters, which are abundant during summer months and possess genes associated with human virulence. When examining the ecological interactions of these species, traditional grouping of diatoms at a high taxonomic level has led to conflicting reports of associations with pathogenic *Vibrio* species. I show that high-resolution taxonomic grouping at the genus level or lower, based on 18S amplicon sequencing, reveals specific interactions that may have important consequences for *Vibrio* ecology and human health, yet would have been overlooked in previous studies.

REFERENCES FOR THE INTRODUCTION

1. Martiny, J. B. H., Bohannan, B. J. M., Brown, J. H., Colwell, R. K., Fuhrman, J. a, Green, J. L., Horner-Devine, M. C., Kane, M., Krumins, J. A., Kuske, C. R., Morin, P. J., Naeem, S., Ovreås, L., Reysenbach, A.-L., Smith, V. H. & Staley, J. T. Microbial biogeography: putting microorganisms on the map. *Nat. Rev. Microbiol.* **4**, 102–112 (2006).

2. Thompson, L. R., Sanders, J. G., McDonald, D., Amir, A., Ladau, J., Locey, K. J., Prill, R. J., Tripathi, A., Gibbons, S. M., Ackermann, G., Navas-Molina, J. A., Janssen, S., Kopylova, E., Vázquez-Baeza, Y., González, A., Morton, J. T., Mirarab, S., Xu, Z. Z., Jiang, L., Haroon, M. F., Kanbar, J., Zhu, Q., Song, S. J., Kosciolek, T., Bokulich, N. A., Lefler, J., Brislawn, C. J., Humphrey, G., Owens, S. M., Hampton-Marcell, J., Berg-Lyons, D., McKenzie, V., Fierer, N., Fuhrman, J. A., Clauset, A., Stevens, R. L., Shade, A., Pollard, K. S., Goodwin, K. D., Jansson, J. K., Gilbert, J. A., Knight, R., Agosto Rivera, J. L., Al-Moosawi, L., Alverdy, J., Amato, K. R., Andras, J., Angenent, L. T., Antonopoulos, D. A., Apprill, A., Armitage, D., Ballantine, K., Bárta, J., Baum, J. K., Berry, A., Bhatnagar, A., Bhatnagar, M., Biddle, J. F., Bittner, L., Boldgiv, B., Bottos, E., Boyer, D. M., Braun, J., Brazelton, W., Brearley, F. Q., Campbell, A. H., Caporaso, J. G., Cardona, C., Carroll, J. L., Cary, S. C., Casper, B. B., Charles, T. C., Chu, H., Claar, D. C., Clark, R. G., Clayton, J. B., Clemente, J. C., Cochran, A., Coleman, M. L., Collins, G., Colwell, R. R., Contreras, M., Crary, B. B., Creer, S., Cristol, D. A., Crump, B. C., Cui, D., Daly, S. E., Davalos, L., Dawson, R. D., Defazio, J., Delsuc, F., Dionisi, H. M., Dominguez-Bello, M. G., Dowell, R., Dubinsky, E. A., Dunn, P. O., Ercolini, D., Espinoza, R. E., Ezenwa, V., Fenner, N., Findlay, H. S., Fleming, I. D., Fogliano, V., Forsman, A., Freeman, C., Friedman, E. S., Galindo, G., Garcia, L., Garcia-Amado, M. A., Garshelis, D., Gasser, R. B., Gerdts, G., Gibson, M. K., Gifford, I., Gill, R. T., Giray, T., Gittel, A., Golyshin, P., Gong, D., Grossart, H. P., Guyton, K., Haig, S. J., Hale, V., Hall, R. S., Hallam, S. J., Handley, K. M., Hasan, N. A., Haydon, S. R., Hickman, J. E., Hidalgo, G., Hofmockel, K. S., Hooker, J., Hulth, S., Hultman, J., Hyde, E., Ibáñez-Álamo, J. D., Jastrow, J. D., Jex, A. R., Johnson, L. S., Johnston, E. R., Joseph, S., Jurburg, S. D., Jurelevicius, D., Karlsson, A., Karlsson, R., Kauppinen, S., Kellogg, C. T. E., Kennedy, S. J., Kerkhof, L. J., King, G. M., Kling, G. W., Koehler, A. V., Krezalek, M., Kueneman, J., Lamendella, R., Landon, E. M., Lanede Graaf, K., LaRoche, J., Larsen, P., Laverock, B., Lax, S., Lentino, M., Levin, I. I., Liancourt, P., Liang, W., Linz, A. M., Lipson, D. A., Liu, Y., Lladser, M. E., Lozada, M., Spirito, C. M., MacCormack, W. P., MacRae-Crerar, A., Magris, M., Martín-Platero, A. M., Martín-Vivaldi, M., Martínez, L. M., Martínez-Bueno, M., Marzinelli, E. M., Mason, O. U., Mayer, G. D., McDevitt-Irwin, J. M., McDonald, J. E., McGuire, K. L., McMahon, K. D., McMinds, R., Medina, M., Mendelson, J. R., Metcalf, J. L., Meyer, F., Michelangeli, F., Miller, K., Mills, D. A., Minich, J., Mocali, S., Moitinho-Silva, L., Moore, A., Morgan-Kiss, R. M., Munroe, P., Myrold, D., Neufeld, J. D., Ni, Y., Nicol, G. W., Nielsen, S., Nissimov, J. I., Niu, K., Nolan, M. J., Noyce, K., O'Brien, S. L., Okamoto, N., Orlando, L., Castellano, Y. O., Osuolale, O., Oswald, W., Parnell, J., Peralta-Sánchez, J. M., Petraitis, P., Pfister, C., Pilon-Smits, E.,

Piombino, P., Pointing, S. B., Pollock, F. J., Potter, C., Prithiviraj, B., Quince, C., Rani, A., Ranjan, R., Rao, S., Rees, A. P., Richardson, M., Riebesell, U., Robinson, C., Rockne, K. J., Rodriguezl, S. M., Rohwer, F., Roundstone, W., Safran, R. J., Sangwan, N., Sanz, V., Schrenk, M., Schrenzel, M. D., Scott, N. M., Seger, R. L., Seguinorlando, A., Seldin, L., Seyler, L. M., Shakhsheer, B., Sheets, G. M., Shen, C., Shi, Y., Shin, H., Shogan, B. D., Shutler, D., Siegel, J., Simmons, S., Sjöling, S., Smith, D. P., Soler, J. J., Sperling, M., Steinberg, P. D., Stephens, B., Stevens, M. A., Taghavi, S., Tai, V., Tait, K., Tan, C. L., Taş, N., Taylor, D. L., Thomas, T., Timling, I., Turner, B. L., Urich, T., Ursell, L. K., Van Der Lelie, D., Van Treuren, W., Van Zwieten, L., Vargas-Robles, D., Thurber, R. V., Vitaglione, P., Walker, D. A., Walters, W. A., Wang, S., Wang, T., Weaver, T., Webster, N. S., Wehrle, B., Weisenhorn, P., Weiss, S., Werner, J. J., West, K., Whitehead, A., Whitehead, S. R., Whittingham, L. A., Willerslev, E., Williams, A. E., Wood, S. A., Woodhams, D. C., Yang, Y., Zaneveld, J., Zarraonaindia, I., Zhang, Q. & Zhao, H. A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* **551**, 457–463 (2017).

3. Azam, F. & Malfatti, F. Microbial structuring of marine ecosystems. *Nat. Rev. Microbiol.* **5**, 782–91 (2007).

4. Fauci, A. S. & Morens, D. M. The Perpetual Challenge of Infectious Diseases. *N. Engl. J. Med.* **366**, 454–461 (2012).

5. Cole, J. J. Bacteria and Algae in Aquatic Ecosystems. *Ecosystems* 291–314 (1982).

6. Amin, S. A., Parker, M. S. & Armbrust, E. V. Interactions between diatoms and bacteria. *Microbiol. Mol. Biol. Rev.* **76**, 667–84 (2012).

7. Ramanan, R., Kim, B. H., Cho, D. H., Oh, H. M. & Kim, H. S. Algae-bacteria interactions: Evolution, ecology and emerging applications. *Biotechnol. Adv.* **34**, 14–29 (2016).

8. Field, C. B., Behrenfeld, M. J., Randerson, J. T. & Falkowski, P. Primary productivity of the biosphere: an integration of terrestrial and oceanic components. *Science.* **281**, 237–240 (1998).

9. Pomeroy, L. R. The Ocean's Food Web, A Changing Paradigm. *Bioscience* **24**, 499–504 (1974).

10. Whitman, W. B., Coleman, D. C. & Wiebe, W. J. Prokaryotes: the unseen majority. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 6578–6583 (1998).

11. Fuhrman, J. A. Dominance of bacterial biomass in the Sargasso Sea and its ecological implications. *Mar. Ecol. Prog. Ser.* **57**, 207–217 (1989).

12. Buchan, A., LeCleir, G. R., Gulvik, C. a & González, J. M. Master recyclers: features and functions of bacteria associated with phytoplankton blooms. *Nat. Rev. Microbiol.* **12**, 686–698 (2014).

13. Sarmento, H. & Gasol, J. M. Use of phytoplankton-derived dissolved organic carbon by different types of bacterioplankton. *Environ. Microbiol.* **14**, 2348–2360 (2012).

14. Sonnenschein, E. C., Syit, D. A., Grossart, H. P. & Ullrich, M. S. Chemotaxis of Marinobacter adhaerens and its impact on attachmentto the diatom Thalassiosira weissflogii. *Appl. Environ. Microbiol.* **78**, 6900–6907 (2012).

15. Amin, S. A., Hmelo, L. R., van Tol, H. M., Durham, B. P., Carlson, L. T., Heal, K. R., Morales, R. L., Berthiaume, C. T., Parker, M. S., Djunaedi, B., Ingalls, a. E., Parsek, M. R., Moran, M. a. & Armbrust, E. V. Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. *Nature* **533**, 98–101 (2015).

16. Diner, R. E., Schwenck, S. M., McCrow, J. P., Zheng, H. & Allen, A. E. Genetic Manipulation of Competition for Nitrate between Heterotrophic Bacteria and Diatoms. *Front. Microbiol.* **7**, 880 (2016).

17. Le Chevanton, M., Garnier, M., Lukomska, E., Schreiber, N., Cadoret, J.-P., Saint-Jean, B. & Bougaran, G. Effects of Nitrogen Limitation on Dunaliella sp.–Alteromonas sp. Interactions: From Mutualistic to Competitive Relationships. *Front. Mar. Sci.* **3**, 1–11 (2016).

18. Durham, B. P., Sharma, S., Luo, H., Smith, C. B., Amin, S. a, Bender, S. J., Dearth, S. P., Van Mooy, B. a S., Campagna, S. R., Kujawinski, E. B., Armbrust, E. V. & Moran, M. A. Cryptic carbon and sulfur cycling between surface ocean plankton. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 453–457 (2014).

19. Bertrand, E. M., McCrow, J. P., Moustafa, A., Zheng, H., McQuaid, J. B., Delmont, T. O., Post, A. F., Sipler, R. E., Spackeen, J. L., Xu, K., Bronk, D. A., Hutchins, D. A. & Allen, A. E. Phytoplankton–bacterial interactions mediate micronutrient colimitation at the coastal Antarctic sea ice edge. *Proc. Natl. Acad. Sci.* **112**, 9938–9943 (2015).

20. Sher, D., Thompson, J. W., Kashtan, N., Croal, L. & Chisholm, S. W. Response of Prochlorococcus ecotypes to co-culture with diverse marine bacteria. *ISME J.* **5**, 1125–32 (2011).

21. Biller, S. J., Coe, A., Roggensack, S. E. & Chisholm, W. Heterotroph Interactions Alter Prochlorococcus Transcriptome. **3**, 1–18 (2018).

22. Mayali, X. & Azam, F. Algicidal bacteria in the sea and their impact on algal blooms. *J. Eukaryot. Microbiol.* **51**, 139–44 (2004).

CHAPTER 1

Genetic manipulation of competition for nitrate between heterotrophic bacteria and diatoms

Synopsis

This chapter is an original research project that uses microbial co-culturing, genetic manipulation, and transcriptomics to examine how a ubiquitous marine gammaproteobacterial species, *Alteromonas macleodii*, can compete with diatoms (specifically, the model diatom *Phaeodactylum tricornutum*) for nitrate in a carbon-dependent manner. I also use nitrate reductase knockout mutants of both diatoms and bacteria to demonstrate that diatoms and bacteria likely exchange nitrogen substrates under certain conditions. A large accomplishment of this research was developing a genetically tractable model system for studying diatom-bacteria interactions.

This chapter is presented as a paper. "Genetic manipulation of competition for nitrate between heterotrophic bacteria and diatoms" was published as a research article in *Frontiers in Microbiology* in 2016.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 09 June 2016
doi: 10.3389/fmicb.2016.00880

Genetic Manipulation of Competition for Nitrate between Heterotrophic Bacteria and Diatoms

Rachel E. Diner[1,2], *Sarah M. Schwenck*[1,2], *John P. McCrow*[2], *Hong Zheng*[2] *and Andrew E. Allen*[1,2]*

[1] *Integrative Oceanography Division, Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA, USA,* [2] *Microbial and Environmental Genomics Group, J. Craig Venter Institute, La Jolla, CA, USA*

Diatoms are a dominant group of eukaryotic phytoplankton that contribute substantially to global primary production and the cycling of important elements such as carbon and nitrogen. Heterotrophic bacteria, including members of the gammaproteobacteria, are commonly associated with diatom populations and may rely on them for organic carbon while potentially competing with them for other essential nutrients. Considering that bacterioplankton drive oceanic release of CO_2 (i.e., bacterial respiration) while diatoms drive ocean carbon sequestration vial the biological pump, the outcome of such competition could influence the direction and magnitude of carbon flux in the upper ocean. Nitrate availability is commonly a determining factor for the growth of diatom populations, particularly in coastal and upwelling regions. Diatoms as well as many bacterial species can utilize nitrate, however the ability of bacteria to compete for nitrate may be hindered by carbon limitation. Here we have developed a genetically tractable model system using the pennate diatom *Phaeodactylum tricornutum* and the widespread heterotrophic bacteria *Alteromonas macleodii* to examine carbon-nitrogen dynamics. While subsisting solely on *P. tricornutum* derived carbon, *A. macleodii* does not appear to be an effective competitor for nitrate, and may in fact benefit the diatom; particularly in stationary phase. However, allochthonous dissolved organic carbon addition in the form of pyruvate triggers *A. macleodii* proliferation and nitrate uptake, leading to reduced *P. tricornutum* growth. Nitrate reductase deficient mutants of *A. macleodii* (Δ*nasA*) do not exhibit such explosive growth and associated competitive ability in response to allochthonous carbon when nitrate is the sole nitrogen source, but could survive by utilizing solely *P. tricornutum*-derived nitrogen. Furthermore, allocthonous carbon addition enables wild-type *A. macleodii* to rescue nitrate reductase deficient *P. tricornutum* populations from nitrogen starvation, and RNA-seq transcriptomic evidence supports nitrogen-based interactions between diatoms and bacteria at the molecular level. This study provides key insights into the roles of carbon and nitrogen in phytoplankton-bacteria dynamics and lays the foundation for developing a mechanistic understanding of these interactions using co-culturing and genetic manipulation.

Keywords: diatoms, bacteria, nitrate, competition, genetic manipulation, transcriptomics, *Phaeodactylum, Alteromonas*

OPEN ACCESS

Edited by:
Xavier Mayali,
Lawrence Livermore National Laboratory, USA

Reviewed by:
Song Xue,
Dalian University of Technology, China
Daniel Jonathan Sher,
University of Haifa, Israel

***Correspondence:**
Andrew E. Allen
aallen@ucsd.edu

Specialty section:
This article was submitted to Aquatic Microbiology, a section of the journal Frontiers in Microbiology

Received: *01 April 2016*
Accepted: *25 May 2016*
Published: *09 June 2016*

Citation:
Diner RE, Schwenck SM, McCrow JP, Zheng H and Allen AE (2016) Genetic Manipulation of Competition for Nitrate between Heterotrophic Bacteria and Diatoms. Front. Microbiol. 7:880. doi: 10.3389/fmicb.2016.00880

INTRODUCTION

Diatoms as well as bacteria are important drivers of oceanic biogeochemical cycling, and frequently occupy overlapping ecological niches. Diatoms are often the dominant primary producers in nutrient rich ecosystems, such as coastal upwelling regions, and can form dense and extensive blooms. Marine bacteria constitute the majority of oceanic biomass (Pomeroy, 1974; Fuhrman, 1989; Whitman et al., 1998), and heterotrophic bacteria utilize and rely on phytoplankton-derived organic matter for survival and growth (Azam and Malfatti, 2007; Sarmento and Gasol, 2012; Buchan et al., 2014). Diatoms and bacteria are subject to frequent environmental fluctuations in availability of essential nutrients such as nitrogen (N) and carbon (C). Nitrate (NO_3^-) in particular often reaches limiting concentrations during phytoplankton blooms (Falkowski and Oliver, 2007). Certain classes of heterotrophic bacteria, such as gammaproteobacteria, are consistently found in phytoplankton-associated microbial communities, and may potentially compete with diatoms for scarce nutrients while simultaneously relying on them for organic C (Buchan et al., 2014). While population dynamics of phytoplankton and bacteria under different environmental conditions have been extensively examined, outside of a few recent studies (e.g., Durham et al., 2014; Amin et al., 2015; Smriga et al., 2016) relatively little is known regarding the cellular, metabolic, or genetic basis for different types of interactions (Bell and Mitchell, 1972; Amin et al., 2012). This is particularly true for competitive interactions (Amin et al., 2012). Laboratory model systems and new experimental approaches can enable hypothesis-testing and lead to new discoveries regarding interactions between diatoms and heterotrophic bacteria in productive microbial ecosystems and the associated influence on C and nutrient cycling.

A common regulator of primary productivity in marine ecosystems is availability of NO_3^-. When estimating primary productivity and characterizing the magnitude of the biological pump it is assumed that inorganic N is converted into particulate organic matter entirely by phytoplankton (Dugdale and Goering, 1967; Bronk et al., 1994). Diatoms are excellent competitors for NO_3^-, and have evolved efficient assimilation, storage and associated recycling systems (Serra et al., 1978; Dortch, 1990; Lomas and Gilbert, 2000; Allen et al., 2006, 2011). The emerging laboratory model diatom *Phaeodactylum tricornutum* has been used in many studies to investigate diatom N utilization, as well as responses to many other environmental variables such as iron (Fe) and phosphorus (P) (Yongmanitchai and Ward, 1991; Geider et al., 1993; Allen et al., 2008; Jiao et al., 2011; Matthijs et al., 2015; Morrissey et al., 2015). These studies have been facilitated by development of a variety of tools for genetic manipulation in *P. tricornutum* (Siaut et al., 2007; Karas et al., 2015; Weyman et al., 2015; Nymark et al., 2016). Notably, mutant strains of *P. tricornutum* that are deficient in ability to utilize NO_3^- have been crucial for understanding N uptake and storage, and impacts on cellular physiology (Levitan et al., 2014). The sequencing of the *P. tricornutum* genome (Bowler et al., 2008) revealed that about 6% of *P. tricornutum* genes appear to be bacterial in origin, including a NAD(P)H dependent assimilatory NO_2^-, reductase, and are possibly the result of horizontal gene transfer. This suggests a historically intimate relationship between diatoms and bacteria, which might also have significant evolutionary implications.

Heterotrophic bacteria also play a large role in N cycling and remineralization (Zehr and Ward, 2002; Azam and Malfatti, 2007). They are known to utilize a variety of sources for satisfying their N requirements, including ammonium (NH_4^+), NO_3^-, urea, free amino acids, and various other organic N compounds. In some studies, NH_4^+ and organic N sources such as amino acids have been shown to satisfy the bulk of heterotrophic bacterial N demand, while other organic N sources and inorganic N such as NO_3^- appeared to play a more minor role (Wheeler and Kirchman, 1986; Keil and Kirchman, 1991). However, a large number of heterotrophic bacteria possess pathways for utilizing NO_3^- and are able to grow using NO_3^- as a sole N source. Studies examining the molecular ecology of heterotrophic bacterial nitrate reductase genes (*nasA*) and their functionality have suggested that bacterial NO_3^- utilization is globally widespread and may play an important role in inorganic N cycling in several ecosystems (Allen et al., 2001; Jiang et al., 2015). Further, heterotrophic bacteria have been shown to satisfy between 10 and 50% of their total N demand with dissolved inorganic nitrogen (NH_4^+ and NO_3^-), and can account for between 10 and 40% of total water column NO_3^- uptake (Allen et al., 2005). Stable isotope probing (SIP) experiments with $^{15}NO_3^-$ have also shown that heterotrophic bacteria in natural assemblages, including members of the *Alteromonas* genera, can and do take up NO_3^- (Wawrik et al., 2012). Methods based on sorting heterotrophic and autotrophic cells with flow cytometry following $^{15}NO_3^-$ incubation have also documented significant levels of heterotrophic bacterial NO_3^- utilization (Bradley et al., 2010a,b,c; Lomas et al., 2011). However, the role that bacterial NO_3^- assimilation plays in shaping microbial communities and regulating NO_3^- flux in pelagic ecosystems remains poorly understood, though it may have important consequences for understanding the biological pump.

To gain a deeper understanding of N-related interactions between diatoms and bacteria, we developed a model co-culture system using the diatom *P. tricornutum* CCMP 632 and the marine heterotrophic bacteria *Alteromonas macleodii*. *A. macleodii* represents an excellent model for investigating these interactions because *Alteromonas* sp. are ecologically important members of the gammaproteobacteria class, and have been shown to be amenable to genetic manipulation (Kato et al., 1998; Weyman et al., 2011). They are relatively large (~1-2 μm), rod-shaped motile bacteria that are capable of utilizing a variety of C and N sources (López-Pérez and Rodríguez-Valera, 2014; Pedler Sherwood et al., 2015). Ecologically, they are frequently associated with nutrient and particle-rich environments and are commonly found as active and dominant members of phytoplankton-associated bacterial assemblages (Buchan et al., 2014; López-Pérez and Rodríguez-Valera, 2014). *Alteromonas* bacteria have previously been shown to interact with individual eukaryotic and prokaryotic phytoplankton species. These interactions range from impairing algal growth, sometimes by algicidal means (Kato et al., 1998; Mayali and Azam, 2004;

Aharonovich and Sher, 2016) to effects that are either neutral or beneficial to algal growth in co-culture (Morris et al., 2008, 2011; Sher et al., 2011; Aharonovich and Sher, 2016). When concentrated organic matter is available, *Alteromonas* bacteria have been shown to be among the most rapidly dividing heterotrophic prokaryotes, and can reach high population densities (Shi et al., 2012; Pedler et al., 2014). *A. macleodii,* the designated type species for the *Alteromonas* genus, is distributed globally and is exceptionally diverse genetically (García-Martínez et al., 2002; Ivars-Martínez et al., 2008; López-Pérez et al., 2012). The strain selected for this study, *A. macleodii* ATCC27126, is capable of utilizing NO_3^- as a sole N source in minimal (Aquil) media supplemented with a dissolved organic carbon (DOC) source, solidifying this strain as an excellent candidate for the model system employed in this study.

Through the use of targeted genetic manipulation we have gained new insights into the mechanisms governing physiological processes related to nutrient exchange and competition between diatoms and bacteria, particularly interactions involving C and N. We examined model diatoms and bacteria that are both capable of NO_3^- utilization. Leveraging new and existing genetic tools available for each organism in the model system presented here, we created both diatom and bacterial mutants lacking the ability to utilize NO_3^-. We then examined the response of these strains in co-culture under varying C and N availability scenarios. Additionally, we conducted transcriptional profiling experiments in order to identify molecular responses of diatoms to the bacteria as well as to gain insights into the physiological status of each partner.

MATERIALS AND METHODS

Strains and Culturing Conditions

Axenic cultures of *P. tricornutum* strain CCMP 632 were obtained from the Provasoli-Guillard National Center for Culture of Marine Algae and Microbiota. Axenic cultures were confirmed via microscopy (light and DAPI staining), in addition to regular plating on marine bacterial growth media. *P. tricornutum* monocultures and *A. macleodii* co-cultures were grown in Aquil artificial seawater media (ASW), with modified concentrations of NO_3^- added as sodium nitrate (Fisher Bioreagents, Waltham MA, USA) or NH_4^+ added as ammonium chloride (Fisher Bioreagents, Waltham MA, USA). Media was microwaved to ∼95°C two times prior to cooling, addition of nutrients and filter sterilization (0.2 μm bottle-top filters, Thermo Fisher Scientific, Waltham MA, USA). Experiments were conducted at 18°C with a light intensity of 170 μmol photons m^{-2} s^{-1} and a 14/10 h light/dark cycle. In order to minimize variability resulting from diel effects, measurements and sampling for each experiment occurred at the same time each day, ∼6 h after light cycle onset.

A. macleodii strain ATCC27126 monocultures were grown routinely either on Zobell 2216 marine broth (MB) 1% agar plates, or in MB liquid medium at 28°C. *A. macleodii* growth was also supported in half-strength marine broth media, and in Aquil ASW supplemented with 5 mM pyruvate. Liquid cultures were grown shaking at 225 rpm. Prior to co-culturing experiments, overnight cultures of *A. macleodii* culture were centrifuged at 6000 x g for 3 min; the supernatant was discarded, and cells were subsequently gently washed 3 times in the experimental seawater media. *E. coli* used for cloning was cultured in LB broth (Amresco, Solon OH, USA) or on LB agar plates at 37°C. Antibiotic concentrations for selective bacterial growth were provided as 100 μg/ml Kanamycin and 100 μg/ml Ampicillin for *A. macleodii* and 50 μg/ml Kanamycin, 100 μg/ml Ampicillin, 10 μg/ml Chloramphenicol, 10 μg/ml Tetracycline, or 10 μg/ml Spectinomycin for *E. coli* as needed.

Genetic Manipulation of *A. macleodii* and *P. tricornutum*

Previous studies genetically manipulating *Alteromonas* bacteria focused on either undesignated species (Kato et al., 1998) or species other than *A. macleodii* (Weyman et al., 2011 focused on the "deep ecotype" which was reclassified as *A. mediterranea*), making this study the first to genetically manipulate this widespread species. The genome sequence of *A. macleodii* strain ATCC27126 was obtained from the JGI IMG data base and the DNA sequence of the single copy nasA gene was identified and used to design the knockout (KO) construct. The *A. macleodii* Δ*nasA* line was engineered using SacB-mediated homologous recombination, as in Weyman et al. (2011). Gibson assembly was used to construct the suicide plasmid pRED16 (Supplementary Figure 1A), which contains an origin of replication from source plasmid pBBR1-MCS5 (incapable of replication in *A. macleodii* ATCC27126), an origin of transfer, a SacB gene conferring toxicity to sucrose, and 2 1-kb regions homologous to the *A. macleodii nasA* gene flanking a kanamycin resistance cassette (Kovach et al., 1995; Gibson et al., 2009; Weyman et al., 2011). This plasmid was assembled and transformed into *E. coli*, which was mated overnight with the *A. macleodii* WT strain. Transconjugants were dilution-plated to obtain Kanamycin resistant single colonies, and then streaked onto 5% sucrose plates to select for double-crossover recombinants. These were again plated to single colonies, which were screened by PCR to amplify regions specific to *A. macleodii* strain ATCC27126 (to confirm sole presence of this strain as the primers do not amplify in *E. coli*), and regions spanning both junctions of the genome insert, as well as the entire insert (data not shown). All colonies screened were identified as *A. macleodii* strain ATCC27126 and were positive for the KO insert. The KO phenotype (inability to utilize NO_3^-) was confirmed by plating transconjugant colonies and the WT strain on seawater-agar plates containing pyruvate as a C source, either NO_3^- or NH_4^+ as a N source, and X-gal solution (Takara, Meadow View CA, USA) to better visualize the phenotypic effect. The WT strain was able to grow on either N source, while the knockout strain displayed growth on NH_4^+ but not on NO_3^- (Supplementary Figure 1B).

P. tricornutum nitrate reductase knockout (NRKO) lines were constructed using Transcription activator-like effector nuclease (TALEN) genetic manipulation, as in Weyman et al. (2015). Using the JGI *P. tricornutum* genome (http://genome.jgi.doe.gov/Phatr2/Phatr2.home.html), the sequence encoding nitrate reductase (NR) (Phatr2 ID: 54983, Phatr3 ID: J54983) was identified and activity was eliminated by interrupting the sequence with a phleomycin-resistance cassette

suitable for downstream selection. Transformation of the NRKO plasmids was accomplished by microparticle bombardment (BIO-RAD PDS-1000/He Biolistic Particle Delivery System).

Experimental Design

To address baseline physiological and transcriptional profiles of diatom and diatom-bacteria co-cultures, 1L batch cultures were grown in autoclaved 2L Erlenmeyer flasks. Three treatments with three biological replicates each were examined: *P. tricornutum* monocultures, *P. tricornutum–A. macleodii* WT co-cultures, and *P. tricornutum–A. macleodii* Δ*nasA* co-cultures. Prior to starting the experiments, both phytoplankton and bacteria were cultured together in the relevant experimental conditions for >7 *P. tricornutum* generations (~1 week), and inoculated during *P. tricornutum* exponential phase. Sampling was conducted daily for spectrophotometric measurement of NO_3^-, and for cell counts via flow cytometry (see below). Samples for additional physiological parameters and RNA-seq transcriptomic analysis were also collected during *P. tricornutum* exponential and stationary phase (**Figure 1**). Physiological parameters included *in vivo* fluorescence, pH, Chl-a, Fv/Fm, and dissolved and particulate organic C and N. The initial NO_3^- concentration was 300 μM. Diatom specific growth rates (μ) were calculated using cell densities obtained via flow cytometry on day 2 and day 5 of the experiment. For comparison to flow cytometry results, *P. tricornutum* cells were also counted on a counting chamber and bacteria colony forming units (CFU) ml^{-1} were determined for the *A. macleodii* (see below). Every second day, the *P. tricornutum* monocultures were plated on MB and co-cultures were plated on MB with Kanamycin in order to confirm lack of culture cross-contamination.

Several small-scale experiments were performed to examine the effects of C and N concentration on co-culture population dynamics, interactions between bacteria and the *P. tricornutum* NRKO line, and *A. macleodii* growth on multiple media types. For these experiments, 20ml cultures and co-cultures were grown in sterile glass tubes in triplicate and 300 μL samples were collected regularly and preserved with paraformaldehyde (PFA) for subsequent processing via flow cytometry (see below). Four hundred microlitre samples were also collected in order to measure NO_3^- concentration in the DOC addition experiment (see below). *A. macleodii* growth on different media sources was evaluated by growing overnight cultures of *A. macleodii* WT, gently rinsing the cells two times in nutrient free Aquil ASW, and resuspending in 1 ml of the same media prior to inoculating media treatments with 1:1000 dilution of the bacteria. Cells were grown in the following media treatments: Aquil ASW with and without 300 μM NO_3^- (Aq and Aq-N, respectively), MB, and expired media from a *P. tricornutum* stationary culture filtered through a 0.2 μm filter (PtF).

In DOC addition experiments, co-cultures that had been maintained semi-continuously were used for inoculations. For both the *A. macleodii* WT and Δ*nasA*-*P. tricornutum* co-cultures, five different DOC-addition treatments were established (each in triplicate): DOC added either at the time of inoculation (Day 0) or on days 2, 4, 6, and 8 of the experiment. A no DOC-addition control was included for a total of 6 treatments. Allochthonous DOC was added as pyruvate at a concentration of 5 mM. Media contained 300 μM final concentration of NO_3^-. As a result of pyruvate interference with spectrophotometer measurement of NO_3^- samples were sent for autoanalyzer NO_3^- analysis (see below).

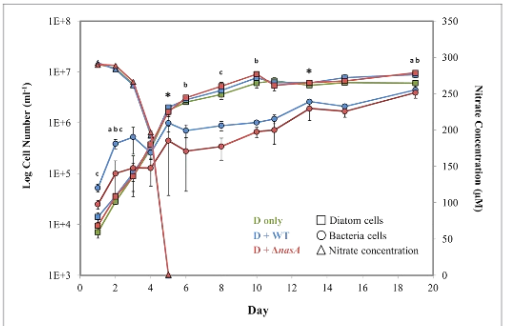

FIGURE 1 | Log number of bacteria and diatom cells (left vertical axis) determined by flow cytometry and nitrate concentrations (right vertical axis) determined by spectrophotometry during the baseline experiment. Squares = diatom cell numbers, circles = bacterial cell numbers, and triangles = nitrate concentrations (μM). Green markers = *P. tricornutum* monocultures (also noted as D only), blue markers = *P. tricornutum–A. macleodii* WT co-cultures (also noted as D + WT), and red markers = *P. tricornutum–A. macleodii* Δ*nasA* co-cultures (also noted as D + Δ*nasA*). Exponential growth stage was defined as days 1–6, and stationary growth phase as days 7–19. * indicates the two sampling points for physiology and transcriptomics. [a] indicates that diatom cell numbers were significantly different between *P. tricornutum* monocultures and *A. macleodii* WT co-cultures. [b] indicates that diatom cell numbers were significantly different between *P. tricornutum* monocultures and *A. macleodii* Δ*nasA* co-cultures. [c] indicates that bacteria cell numbers were significantly different between the WT and Δ*nasA* co-cultures. All data points represent the average of $n = 3$ replicates, and error bars are standard deviation. Some differences in cell numbers and nitrate concentrations (e.g., diatom monoculture nitrate concentrations, which are overlapped by the Δ*nasA* co-culture measurements) are difficult to discern, so average non-transformed cell numbers and nitrate concentrations have been listed in Supplementary Table 2.

To address the impacts of varying C and NO_3^- concentrations on *P. tricornutum* and *A. macleodii* WT population dynamics, co-cultures maintained semi-continuously from the baseline experiment were used to inoculate experiments using nine different media types. NO_3^- concentrations of 50 μM, 300 μM, and 1 mM and DOC concentrations of 0 μM, 50 μM, and 1 mM were tested in a factorial design. Twenty milliliter co-cultures were grown in sterile glass tubes in triplicate and 500 μL samples were collected for flow cytometry analysis on days 2, 4, 6, and 31. Co-cultures containing 300 μM NO_3^- and no DOC addition were cultured semi-continuously for 5 subsequent transfers to determine steady-state growth rates of *P. tricornutum* in co-culture. To account for variability in cell concentration, μ was determined using the highest growth rate observed between any 2 subsequent sampling points in each transfer experiment.

For the *P. tricornutum* NRKO co-culturing experiments, three different co-cultures were grown in both NH_4^+ and NO_3^- amended Aquil ASW; *P. tricornutum* NRKO only, *P. tricornutum* NRKO + WT *A. macleodii*, and *P. tricornutum* NRKO + WT *A. macleodii* + DOC. The *P. tricornutum* NRKO with and without *A. macleodii* were acclimated in NH_4^+ amended media after which they were inoculated into both media types and pyruvate was added to the relevant treatments. N was added as 880 μM final concentration of the relevant source. Since some of the experiments included DOC amendments, we conducted an experiment to examine the impact of such DOC addition on *P. tricornutum* growth. *P. tricornutum* WT monocultures and *P. tricornutum* WT monocultures + 5 mM pyruvate cultures were established in 20 ml cultures as above, and run in triplicate, and media contained 880 μM final concentration of NO_3^-.

Sample Collection and Processing: Cell Numbers, Physiology, and NO_3^- Drawdown

Cell densities of diatoms and bacteria were determined by flow cytometry, and in the case of the baseline experiment, also by manual counting methods. *P. tricornutum* cells were counted manually on either a Sedgwick-Rafter hemocytometer (when cell densities were low) or an Improved-Neubauer hemocytometer (IN-Cyto, Chungnam-do, Korea). *A. macleodii* CFU were determined by dilution-plating onto MB agar plates. (Both WT and Δ*nasA* lines of *A. macleodii* were also plated on MB kanamycin plates to confirm the sole presence of Δ*nasA* lines. Samples for staining and cell counting via flow cytometry were preserved by adding PFA to a final concentration of 0.5%. Samples were incubated at 4°C prior to flash freezing and storage at −80°C until processing. Flow cytometry analysis was conducted on a BD FACS Aria II using the Bacteria Kit for flow cytometry (Thermo Fisher Scientific, Waltham MA, USA) for quantifying the number of diatom and bacteria cells in each sample. After the addition of beads to samples, SYBR green I DNA stain was added, effectively staining all three populations (diatoms, bacteria, and beads). Bacterial and diatom populations were quantified simultaneously, and typical flow cytometer settings were forward scatter (FSC) = 200, side scatter (SSC) = 250, FSC PMT = 550, SYBR Green (SYBR Grn) = 530, Yellow fluorescence (YFP) = 335, with the following thresholds: SSC = 200, FSC = 200.

Nitrate levels were measured either via spectrophotometer or autoanalyzer (in DOC addition experiments). For spectrophotometric measurement, samples were prepared by centrifuging whole sample in Eppendorf tubes at 8000g for 10 min. Supernatant was then recovered, and stored at −20°C until processing. NO_3^- values were determined by generating a standard curve using dilutions of 880 μM NO_3^- media, using only curves with >99% precision (Collos et al., 1999; Johnson and Coletti, 2002). Samples were then measured in triplicate. As allochthonous DOC addition interfered with the spectrophotometric readings, NO_3^- and NO_2^- analysis for DOC addition experiments was conducted using a Lachat QuikChem 8500 autoanalyzer as in Parsons et al. (1984).

Concentrations of total and dissolved organic C and N were determined using a Total Organic Carbon (TOCL) analyzer (Shimadzu) paired with an ASI-L autosampler (Shimadzu). Prior to sample collection, 24-ml glass vials were combusted at 450°C and vial caps were acid rinsed for >24 h. Five to ten milliliters samples were stored at −20°C prior to dilution with milliQ and processing. Standard curves for determining C and N concentrations were generated using automated dilution and sampling of 1000 ppm potassium hydrogen phthalate for C, and 1000 ppm potassium nitrate for N. Samples for determining C and N in the dissolved fraction were collected by gentle filtration through a 0.2 μm syringe filter. Particulate values were determined by subtracting dissolved values from total values. Chlorophyll A (Chl a) concentrations were measured by 90% acetone extraction. Ten milliliters of the sample was gently filtered onto GF/F filters, flash frozen in in liquid N_2, and stored at −20°C before overnight acetone extraction and measurement on a 10AU fluorometer (Turner). Culture Fv/Fm was measured using a pam-fluorometer (WALZ). pH was measured using an InLab Expert pH probe (Mettler Toledo), calibrated using 4, 7, and 10 pH standards (Orion Application Solutions).

Sample Collection and Processing: Transcriptomics

Transcriptomics samples were collected by gentle filtration onto 47 mm 0.2um polycarbonate filters (Whatmann), followed by flash freezing in liquid N_2 and storage at −80°C prior to processing. Total RNA was isolated using Trizol reagent (Thermo Fisher Scientific, Waltham MA, USA). The TURBO DNA-free Kit (Thermo Fisher Scientific, Waltham MA, USA) was used to digest genomic DNA. RNA was subsequently purified further using the Agencourt RNAClean XP kit (Beckman Coulter, Carlsbad CA, USA). The quality of RNA was evaluated using an Agilent 2100 Bioanalyzer (also used for subsequent quality analyses). Ribosomal RNA was removed using Ribo-Zero Magnetic kits (Epicentre, San Diego CA, USA) with a modification of the removal solution, using a mixture of the plant, bacterial, and human/mouse/rat Removal Solutions in a ratio of 2:1:1. Following mRNA enrichment via rRNA removal, RNA quality was further inspected via bioanalyzer. The Ovation RNA-Seq System V2 (NuGEN Technologies, Inc.) was used for first and second strand cDNA synthesis and amplification, followed by evaluation of cDNA quality via bioanalyzer. cDNA was sheared using the S2/E210 focused-ultrasonicator (Covaris) with a target size of 300 bp, confirmed by bioanalyzer. Libraries for sequencing were constructed using the Ovation® Ultralow System V2 (NuGEN Technologies, Inc.), and the quality of libraries verified by bioanalyzer prior to sequencing. Libraries were quantified using qPCR and a Library Quantification Kit (Kapa Biosystems), prior to sequencing on an Illumina NextSEQ500 DNA sequencer.

Paired Illumina reads were filtered for Illumina primer contamination and quality trimmed to Phred score 33 and a minimum length of 30 prior to read mapping. Reads were mapped to target genome contigs of *P. tricornutum* (http://genome.jgi.doe.gov/Phatr2/Phatr2.home.html) and

A. macleodii ATCC 27126 using BWA MEM alignment (Li, 2013). Raw read counts were calculated for each gene using featureCounts (Liao et al., 2014) based on gene models for *A. macleodii* (CP003841) and *P. tricornutum* (Phatr3, http://protists.ensembl.org/Phaeodactylum_tricornutum/Info/Index). Additional gene level *de novo* functional annotation was generated for *P. tricornutum* via KEGG, KO, KOG, Pfam, and TIGRfam assignments. RPKMs were computed using library mapped reads and lengths of CDS for each gene. Biological triplets were used to quantitatively estimate differential expression using edgeR (Robinson et al., 2010) to assign normalized fold-change and Benjamini-Hochberg adjusted *p*-values for each gene. Raw read counts for each gene were used in all edgeR analyses. Sequencing data generated as part of this study has been deposited at NCBI.

Statistical Analyses

In the baseline experiment, statistical analyses were conducted to examine potential differences in diatom cell numbers between the monoculture and co-culture treatments, bacteria cell numbers between co-culture treatments, and to explore whether any physiological parameters were different between treatments during the exponential or stationary phase. One-way ANOVAs were performed followed by Tukey HSD *post-hoc* tests. Statistical significance was assumed at $p \leq 0.05$. All statistical analyses were conducted using R (version 2.14.2), and were performed on raw data (i.e., not transformed).

RESULTS

Baseline Physiology and Transcriptomics in Diatom-Bacteria Co-Cultures

A baseline experiment was conducted initially to examine cell growth, culture physiology, N drawdown, and diatom gene expression in diatom-bacteria co-cultures. In this experiment, N was provided as NO_3^- and no DOC was added. Cell counts obtained via flow cytometry were used to examine population dynamics of *P. tricornutum* and the *A. macleodii* strains in co-culture, which were compared to manual cell counts. Diatom growth stages for this experiment were defined generally as exponential (Day 1–5) and stationary (Day 6–19). The presence of WT or Δ*nasA A. macleodii* in *P. tricornutum* cultures did not significantly affect growth rate (Supplementary Table 1). Maximum diatom cell densities were typically similar among treatments during diatom exponential phase, but were higher in bacteria-containing cultures during stationary phase (**Figure 1**, Supplementary Table 2). These differences were significant between *P. tricornutum* monocultures and the *A. macleodii* Δ*nasA* co-cultures on day 2 ($p < 0.01$), day 6 ($p < 0.05$), day 10 ($p < 0.05$), and day 19 ($p < 0.01$), and between *P. tricornutum* monocultures and the *A. macleodii* WT line on day 1 ($p < 0.05$), day 2 ($p < 0.005$), and day 19 ($p < 0.05$; **Figure 1**). Although the maximum *P. tricornutum* cell densities estimated with manual counts were slightly lower than those calculated via flow cytometry, cell counts obtained from the two methods were similar (Supplementary Figure 2).

Both *A. macleodii* strains were able to grow in co-culture with *P. tricornutum*, as indicated by a gradual increase in cell abundance following *P. tricornutum* exponential stage (**Figure 1**). This growth increase did not occur when *A. macleodii* was grown in Aquil ASW in the absence of *P. tricornutum* (Supplementary Figure 3). The *A. macleodii* WT strain maintained higher numbers than the Δ*nasA* line, though patterns of growth were similar for both strains (**Figure 1**). These differences were significant ($p < 0.05$) on days 1, 2, and 8 of the experiment. After an initial increase in bacterial cell number in both bacteria co-culture treatments, bacterial numbers either declined (WT) or plateaued (Δ*nasA*) during the *P. tricornutum* exponential phase (**Figure 1**). Subsequently, cell numbers of both strains increased during the diatom stationary phase beginning on day 5. Manual CFU counts of the bacteria showed generally the same pattern of growth, however the growth decrease during *P. tricornutum* exponential phase was more dramatic and overall CFU ml^{-1} were lower following this phase in the experiment (Supplementary Figure 2). This difference in cell counts due to methodology has been observed in prior studies (Singleton et al., 1982; Mouriño-Pérez et al., 2003), where CFU counts were lower than direct counts under low DOC conditions. Likely the CFU counts reflect only viable, culturable cells while flow cytometry counts represent all cells including dormant, active, and recently dead cells. Neither *A. macleodii* strain was observed to physically attach to *P. tricornutum* at any point of the *P. tricornutum* growth cycle (qualitative observation, data not shown).

Nitrate concentrations decreased rapidly as *P. tricornutum* cell concentrations increased, and were undetectable by day 5 in all treatments (**Figure 1**). *P. tricornutum* cell numbers continued to increase exponentially even after the complete depletion of NO_3^- in the media (**Figure 1**). NO_3^- drawdown was similar among treatments with and without bacteria. Samples for cell physiology that were collected during the diatom exponential (Day 5) and stationary phase (Day 13), including Chl-a, pH, and Fv/Fm, showed no significant differences between treatments (Supplementary Table 3). Organic C and N were also evaluated for both the dissolved and the particulate culture fractions, and no significant differences between the treatments were observed (Supplementary Table 3).

Baseline Transcriptomic Analysis

Whole-genome transcriptome analyses were conducted at exponential (Day 5) and stationary (Day 13) sampling points. In general, a very low percentage of sequenced and mapped reads were associated with the *A. macleodii* genome, with slightly more observed in the diatom stationary samples than in the exponential samples (**Table 1**). The large majority of sequenced reads (>98% in all co-culture treatments) mapped to the *P. tricornutum* genome (**Table 1**). As a result, analyses of *A. macleodii* gene expression are not included in this study. Genes were considered to be significantly differentially expressed (DE) when the adjusted $p < 0.05$, and only genes with differential expression of > 0.75 fold are discussed. All data reported in this paper are deposited in the NCBI sequence read archive (BioProject accession no. PRJNA319251; BioSample accession nos. SAMN04884450- SAMN04884467).

TABLE 1 | Sequencing data collected for transcriptomic sampling points, including the total number of raw reads, the total number of trimmed reads, the percentage of trimmed reads that mapped to either the a genome or a gene model belonging to *P. tricornutum* or *A. macleodii*, and the percentage of the total mapped reads that corresponded to the *A. macleodii* and *P. tricornutum* genomes.

	Raw Reads	Trimmed Reads	% Reads mapped to a genome	% Reads mapped to a gene model	% Mapped reads: *A. macleodii*	% Mapped reads: *P. tricornutum*
EXPONENTIAL SAMPLING POINT						
P. tricornutum Only	1.0E + 07	9.9E + 06	84.18	58.29	0.00	100.00
P. tricornutum + *A. macleodii* WT	7.7E + 06	7.6E + 06	80.74	35.78	0.21	99.79
P. tricornutum + *A. macleodii* Δ*nasA*	9.2E + 06	9.0E + 06	82.78	36.60	0.28	99.72
TOTAL	2.7E + 07	2.7E + 07				
STATIONARY SAMPLING POINT						
P. tricornutum Only	2.9E + 07	2.8E + 07	88.16	43.52	0.00	100.00
P. tricornutum + *A. macleodii* WT	3.7E + 07	3.6E + 07	88.61	43.96	1.70	98.30
P. tricornutum + *A. macleodii* Δ*nasA*	3.0E + 07	3.0E + 07	88.92	33.69	1.90	98.10
TOTAL	9.5E + 07	9.4E + 07				

No *P. tricornutum* genes were significantly differentially expressed (DE) between any of the treatment during the diatom exponential sampling point, however, during the stationary sampling point many genes were differentially expressed between the *P. tricornutum* monoculture and either or both of the *P. tricornutum*-bacteria co-cultures. A set of 34 genes were significantly DE between axenic cultures and both bacterial co-cultures (**Figure 2**). The gene most highly upregulated in *P. tricornutum* in response to the bacteria (>5.9 fold in both co-cultures) is a putative voltage-gated ion channel (Phatr2 ID: 49093, Phatr3 ID: 302957) that has been shown to be involved in NO_3^- sensing and transport (see discussion). Other upregulated genes include a putative ferredoxin-dependent bilin reductase (Phatr2 ID: 33770, Phatr3 ID: 303606), and a putative fatty acid desaturase (Phatr2 ID: 46830, Phatr3 ID: 306355). Several of the genes that were downregulated in *P. tricornutum* in response to bacteria were related to cellular information storage and processing such as transcriptional regulation and replication, including two different putative heat shock protein transcription factors. The two genes most highly downregulated in the presence of bacteria were a short chain dehydrogenase (Phatr2 ID: 13001, Phatr3 ID: 306282), which was downregulated 8.8 and 6.5-fold in the Δ*nasA* and WT co-cultures, respectively, and a fatty acid hydroxylase (Phatr2 ID: N/A, Phatr3 ID: 308140), which was downregulated 5.7 and 3.9-fold in the Δ*nasA* and WT co-cultures, respectively (**Figure 2**).

Several transcripts were either upregulated or downregulated in both co-cultures, but the difference was only significant for one of the co-cultures. An additional 9 genes were DE between *P. tricornutum* monocultures and the *P. tricornutum*–*A. macleodii* WT co-culture treatments (Supplementary Table 4). These include a putative NO_2^- transporter (Phatr2 ID: 13076, Phatr3 ID: 308281), and a putative membrane associated NO_3^- transporter (Phatr2 ID: 26029, Phatr3 ID: 307720), which exhibited a >2.5-fold difference. Both of these putative genes were upregulated in the *P. tricornutum* monoculture compared to the co-cultures, and had a larger expression change in the WT bacteria co-culture than the Δ*nasA* co-culture (Supplementary Table 4). There were 99 genes significantly DE in *P. tricornutum* monocultures compared to *P. tricornutum*-*A. macleodii* Δ*nasA* co-cultures (Supplementary Table 4), including upregulation of a putative glutamine synthetase gene (Phatr2 ID: 51092, Phatr3 ID: 306624), a putative tryptophan/tyrosine permease (Phatr2 ID: 45852, Phatr3 ID: 310088), and a putative ferredoxin nitrite reductase (Phatr2 ID: 12902, Phatr3 ID: 308097). A putative sugar transporter (Phatr3 ID: 49722, Phatr2 ID: 311238) was downregulated in diatom monocultures, as well as additional heat shock transcription factors, including one that was downregulated >19-fold (Phatr2 ID: 48554, Phatr3 ID: 304737).

We also identified and compared expression of genes related to NH_4^+ utilization and transport. 17 putative genes were identified (Supplementary Table 5), and none were significantly DE between any treatments at a given sampling point.

Growth Physiology of *A. macleodii* in Multiple Media Types

In all media types, *A. macleodii* WT showed a rapid increase in cell number between inoculation on day 1 and the next measured time-point on day 3 (Supplementary Figure 3). Following this initial increase, cell numbers either decreased in all aquil-based media (Aq, Aq-N, and PtF media), or experienced a modest decrease followed by little change (MB media) (Supplementary Figure 3). The highest cell density occurred when cells were grown in MB media, reaching ~1×10^8 cells ml^{-1}. *A. macleodii* WT in PtF media reached a higher maximum cell density (2.2×10^6 cells ml^{-1}) than the Aq and Aq-N treatments (8.6×10^5 and 8.9×10^5 cells ml^{-1}, respectively).

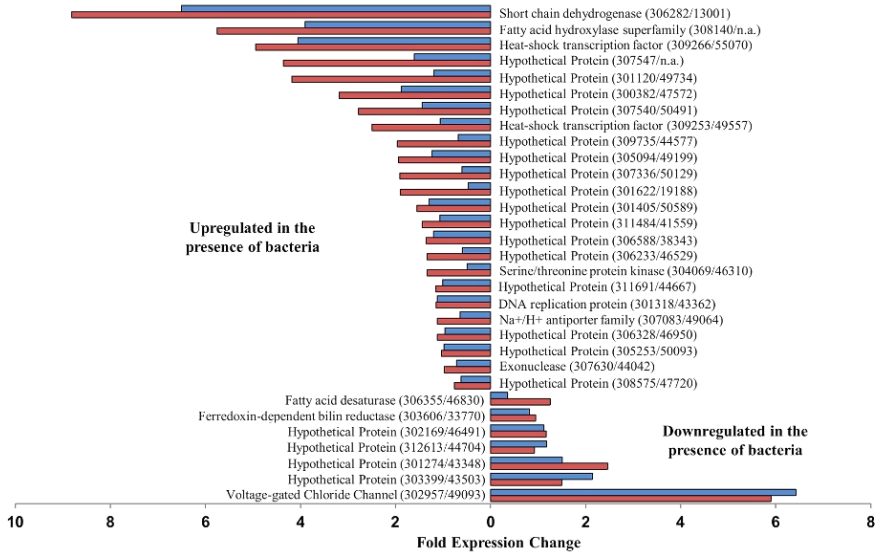

FIGURE 2 | Putative *P. tricornutum* genes that were significantly differentially expressed between *P. tricornutum* monocultures and both WT and Δ*nasA* bacterial co-cultures, and fold expression change. Blue bars = DE in WT co-cultures, red bars = DE in Δ*nasA* co-cultures. Gene descriptions are adjacent to relevant data bars, with gene IDs displayed as (Phatr3 ID/Phatr2 ID).

Competition between Diatoms and Bacteria for NO_3^-

We examined whether bacteria could impede diatom growth by competing for NO_3^- if sufficient DOC was present for bacterial growth. We hypothesized that the presence of DOC prior to NO_3^- depletion could enable bacterial competition for NO_3^-, but that competition would not occur if (1) DOC is not sufficient for bacterial growth, or (2) bacteria are unable to utilize the NO_3^- in the media. We tested this by culturing the diatoms with WT bacteria capable of utilizing NO_3^- and Δ*nasA* bacteria unable to utilize NO_3^- as a N source. We then added DOC at multiple time-points and included a no DOC addition control to better understand any competitive interactions observed, measuring diatom and bacteria growth and NO_3^- in the media (indicative of biological NO_3^- drawdown). In co-cultures containing Δ*nasA* bacteria, diatom growth and NO_3^- drawdown were similar in all treatments regardless of whether and when DOC was added (**Figures 3B,D,F**). Numbers of Δ*nasA* bacteria increased slightly when DOC was added prior to the depletion of NO_3^- from the media, but were not affected when DOC was added on days after NO_3^- depletion. In co-cultures containing diatoms and WT bacteria, when DOC was added early in the experiment (Day T_o or Day 2), bacteria cell numbers increased dramatically and diatom cell numbers reached lower maximum cell densities (**Figures 3A,C**). Diatom cell numbers with DOC added on Day $T_o$were > 6 times lower than in the no DOC addition control. Media NO_3^- was also depleted earlier in these cultures (**Figures 3E,F**), indicating a relationship between bacterial growth, NO_3^- drawdown, and diatom growth impairment. When DOC was added on later time points (Day 4, 6, and 8), diatom growth and NO_3^- drawdown were similar to the no DOC addition control. Bacteria growth increased slightly when DOC was added on Day 4, but DOC addition on subsequent days had no effect.

P. tricornutum cultures with and without the addition of pyruvate (final concentration 5 mM, as used in other DOC addition experiments) did not differ significantly in cell numbers as measured at 3 different time points (Supplementary Figure 4). Cell numbers were determined on days 4, 7, and 15 of the experiment.

Population Dynamics of NRKO Diatoms and WT Bacteria in Co-Culture

We explored whether bacteria can provide diatoms with a useable N source and potentially "rescue" N-starved diatoms. To do this, we cultured the NRKO diatom strain (lacking the ability to use NO_3^-) in media containing either NH_4^+ or NO_3^- as the sole N source, in the presence of absence WT *A. macleodii*, and with the bacteria and a DOC addition at T_o (**Figure 4**). The NRKO diatom was able to grow in all treatments with NH_4^+ as the provided N source, and in the absence of added DOC grew similarly with and without the addition of *A. macleodii* (**Figures 4A,B**). In NH_4^+ media amended with DOC, bacteria numbers were much higher, and NRKO diatom cell densities were lower (**Figures 4C,D**), an observation similar to the DOC addition experiment described

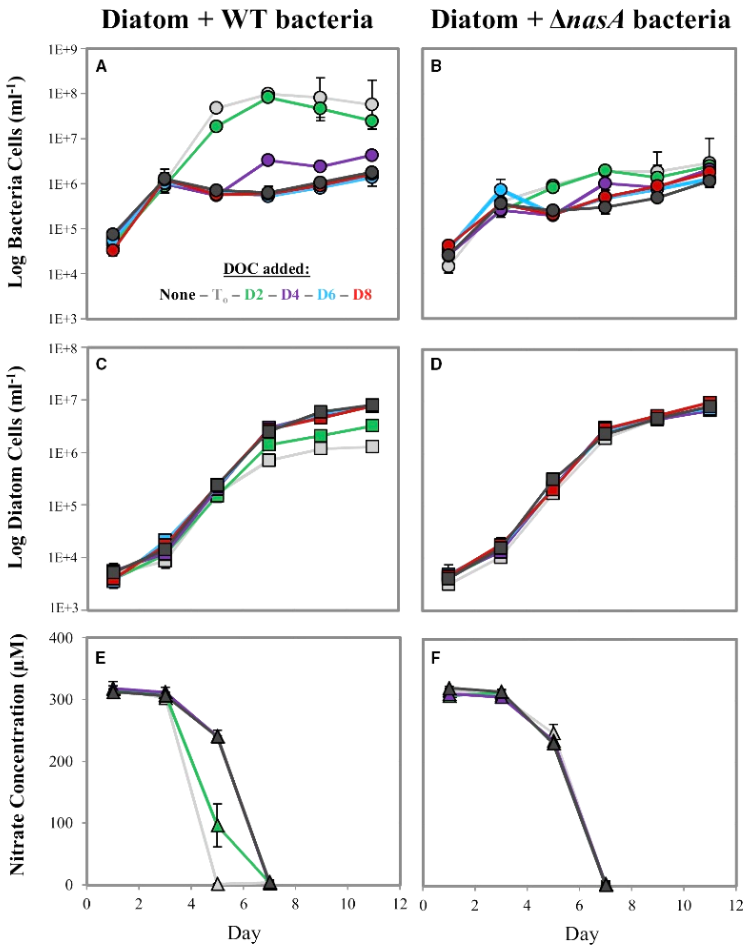

FIGURE 3 | Log cell numbers per ml and nitrate concentrations in DOC addition experiments. (A,C,E) are co-cultures containing the WT *A. macleodii* strain, and **(B,D,F)** contain the *A. macleodii* Δ*nasA* strain. Colors represent the day of the experiment on which DOC was added: black = control, no DOC added, gray = day of inoculation (T_0), green = day 2, purple = day 4, blue = day 6, red = day 8.

above. With NO_3^- as the sole N source and without DOC addition, the diatom NRKO strains displayed little detectable growth with and without *A. macleodii* addition. However, the NRKO diatom-WT bacteria co-culture to which DOC was added displayed a much different growth response; diatom density increased linearly ($R^2 = 0.96$) throughout the experiment, with a maximum cell density of 3.0×10^6 cells ml^{-1} measured on day 19 of the experiment (**Figures 4A,B**). Bacterial cell densities in this treatment peaked on day 9 of the experiment, and subsequently declined.

Effect of Various NO_3^- and DOC Concentrations on Diatom-Bacteria Population Dynamics

To better understand the influence of N and C concentration on diatom-bacteria growth dynamics, we conducted a factorial experiment in which diatom-WT bacteria co-cultures were grown in 9 different media types encompassing a range of NO_3^- and DOC levels. In general, higher NO_3^- levels resulted in higher numbers of both diatoms and bacteria. DOC concentrations

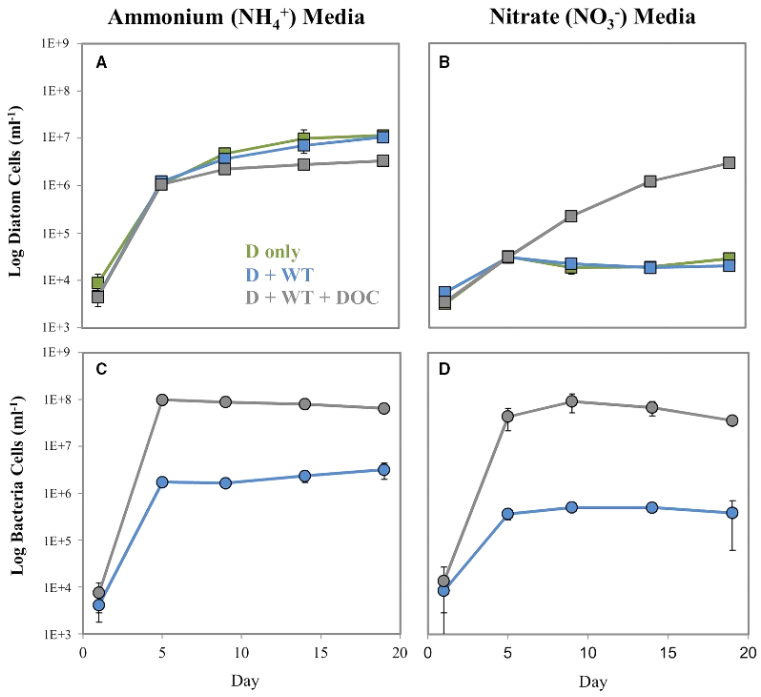

FIGURE 4 | Log cell numbers per ml of *P. tricornutum* NRKO (A,B) and *A. macleodii* WT bacteria (C,D) in co-culture, grown on NH_4^+ or NO_3^- and as the sole nitrogen source. Green markers = *P. tricornutum* NRKO monoculture (also noted as D only), blue = *P. tricornutum* NRKO + *A. macleodii* WT (also noted as D + WT), and gray = *P. tricornutum* NRKO + *A. macleodii* WT + DOC (also noted as D + WT + DOC).

had a strong effect on bacterial cell numbers, but less of an effect on diatom concentrations except for at low NO_3^-. At the lowest NO_3^- concentration tested (50 μM), diatom cell numbers decreased with increasing DOC concentration (**Figure 5A**) while bacterial cell numbers increased (**Figure 5D**). However, both diatom and bacteria concentrations were low at low NO_3^- level regardless of DOC concentration compared to the higher NO_3^- treatments. At the higher NO_3^- levels (300 μM and 1 mM), *P. tricornutum* concentrations were higher and similar to each other across DOC treatments (**Figures 5B,C**). In almost all treatments, *A. macleodii* concentrations were higher during the *P. tricornutum* exponential phase compared to the stationary sampling point on day 31 (**Figures 5D–F**). The reverse was true with the diatoms; cell numbers increased between day 6 and day 31 (**Figures 5A–C**). Final *A. macleodii* cell densities measured on day 31 were positively correlated with both DOC and NO_3^- concentration: lower concentrations were observed at the lower levels of both NO_3^- (50 and 300 μM) and DOC (0 and 50 μM), and highest concentrations were observed in the high NO_3^-, high DOC cultures and (**Figures 5D–F**).

In media containing 300 μM NO_3^- and no DOC addition (the same conditions as the baseline experiment), *P. tricornutum* maintained in semi-continuous cultures displayed consistent growth rates across 5 rounds of culture transfer, indicating steady-state growth. The average $\mu = 1.42$, with a range of $\mu = 1.25$ to 1.62 (Supplementary Table 6).

DISCUSSION

A Continuum: Diatom-Bacteria NO_3^- Utilization and the Role of Organic DOC Availability

Since phytoplankton and bacteria often co-occur in the same environments, the concept of competition for nutrients has been of interest for decades (Bratbak and Thingstad, 1985; Thingstad et al., 1993; Logan et al., 1995; Grossart and Simon, 2007; Amin et al., 2012). New relationships between phytoplankton and bacteria are being discovered and investigated with increasing frequency (Durham et al., 2014; Amin et al., 2015; Bertrand et al., 2015; Smriga et al., 2016), in part due to recent methodological

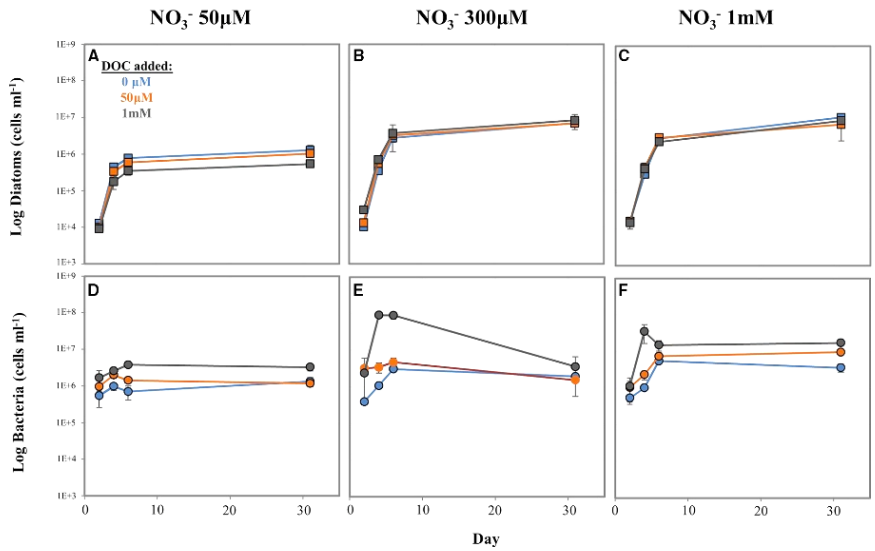

FIGURE 5 | Log cell numbers per ml of *P. tricornutum* and *A. macleodii* grown in co-culture under variable nitrate and dissolved organic carbon (DOC) concentrations. Diatom cell numbers are shown in **(A–C)** and bacterial numbers shown in **(D–F)**. Blue = 0 µM DOC added, orange = 50 µM DOC added, and gray = 1 mM DOC added.

advances. In particular, *Alteromonas* bacteria have been shown to have positive, negative, and seemingly neutral relationships with individual phytoplankton species. It is unclear, however, what role competition for nutrients may play in these observed interactions Understanding the heterogeneous landscape in which these interactions occur and the resulting impacts on global biogeochemical cycles, particularly carbon and nitrogen cycling, is an active area of research (Stocker, 2012; Worden et al., 2015; Smriga et al., 2016).

Many phytoplankton-bacteria interactions likely fall somewhere along a commensal-competitive continuum. Phytoplankton release increasing amounts of organic matter when nutrients become limited, stimulating the growth of bacteria which may compete with them for the same limiting nutrients, an apparent paradox (Bratbak and Thingstad, 1985; Bertrand et al., 2015). Studies examining acquisition of inorganic phosphate have found that bacteria can compete with phytoplankton, and that this potential competition can be nutrient concentration dependent (Bratbak and Thingstad, 1985; Thingstad et al., 1993), and bacteria-phytoplankton population dynamics may be dependent on what species are present as well as nutrient concentrations (Puddu et al., 2003; Grossart and Simon, 2007). Interestingly, while inorganic N availability is also a major driver in ocean productivity and biogeochemical cycling, and both phytoplankton and many bacteria can utilize it, few studies have examined these potentially important interactions (Dugdale and Goering, 1967; Allen et al., 2001; Amin et al., 2012; Jiang et al., 2015). This is despite findings that bacterial *nasA* genes are common, diverse, and highly expressed during phytoplankton blooms, especially for particular bacterial classes and genera (including *Alteromonas*) (Allen et al., 2012; Jiang et al., 2015). Recent advancements in microbiology, including genetic manipulation and the introduction of powerful next generation sequencing technology, call for an examination of these relationships with an aim to understand the underlying complex cellular mechanisms.

Utilizing both WT and Δ*nasA* bacteria in co-culture with diatoms presents the opportunity to explore how the bacteria and diatom in this model system impact one another and are impacted by NO_3^- availability and utilization. NO_3^- was drawn down quickly in the experiments, and the diatom population continued to increase in all co-cultures even after the complete depletion of NO_3^- from the media, which is consistent with reports of the ability of diatoms to rapidly accumulate and store N present in the environment (Cermeño et al., 2011). During exponential phase with no DOC added to the system, *P. tricornutum* growth rate, cell number, and other aspects of *P. tricornutum* physiology, including growth rate, Chl a concentration, Fv/Fm, and culture pH were not affected by the presence of either bacterial strain (**Figure 1**, Supplementary Figure 2, Supplementary Tables 1, 3). At this sampling point, bacterial numbers began to increase after an initial decrease in the experiment. One possible explanation for these observations is that during the exponential phase, the diatom and bacteria co-exist in a commensal relationship whereby the diatom is not affected by the presence of bacteria, but the bacteria are able to

benefit from diatom organic matter. However, it is also possible that low bacterial cell numbers and biomass obscure any positive or negative effect that the bacteria may be having on the diatoms on a smaller scale, and as a result differences are not observable based on the methods we used in this study. Later in the diatom's stationary phase, when nitrogen is limited, cultures containing bacteria reached higher cell densities than the *P. tricornutum* monocultures while bacterial numbers also continued to increase, suggesting a potential cooperative relationship. We hypothesize that the increase in diatom cell number is the result of bacteria providing diatoms with a viable nitrogen source in exchange for DOC (discussed further below).

Previous studies have examined growth effects of *Alteromonas* bacteria in co-culture with phytoplankton. *Alteromonas* bacteria species cultured with eukaryotic phytoplankton have been shown to exhibit algicidal (thought to be the result of secreted dissolved substances) and non-algicidal effects (reviewed in Mayali and Azam, 2004). When cultured with the prokaryotic cyanobacteria from the *Prochlorococcus* genus, the presence of *Alteromonas* sometimes provide the algae with benefits by protecting them from oxidative stress (Morris et al., 2008, 2011) while in other cases they can cause growth inhibition (Sher et al., 2011; Aharonovich and Sher, 2016). In prior non-*Alteromonas* co-culturing studies, dynamics between phytoplankton and bacteria have been shown to manifest late in the growth cycle (Grossart and Simon, 2007; Wang et al., 2014), though often the result is commensal bacteria turning algicidal, possibly to relieve nutrient stress. Immediate growth increases in diatom populations before and during exponential phase resulting from the presence of bacteria have also been observed (Grossart and Simon, 2007; Amin et al., 2015). These results are for the most part different from what we observed in our study, which may be due to the differences in the species of both phytoplankton and bacteria examined (even within the genus *Alteromonas*, species are quite diverse and were not always identified in these studies). Despite being observed in a prokaryote-prokaryote co-culturing system, the possibility that *Alteromonas* bacteria may protect phytoplankton from oxidative stress is interesting and could be examined in the future using the model system developed in this study.

While the presence of the bacteria alone did not hinder diatom growth, when DOC was added early in the diatom growth phase (prior to the depletion of NO_3^- from the media) the bacteria could acquire NO_3^- from the media, making it unavailable to the diatoms. Our data suggest that depending on the availability of allochthonous DOC, bacteria that have the ability to utilize NO_3^- in the environment are able to effectively compete for NO_3^-. This pattern was not observed in co-cultures containing Δ*nasA* bacteria, illustrating how bacteria that are unable to use NO_3^- cannot take advantage of surplus organic C in NO_3^- cultures, and in the case of Δ*nasA* appear to be limited by diatom C production. The WT bacteria generally maintained higher cell numbers than the Δ*nasA* bacteria, possibly suggesting that the WT bacteria were limited by C while the Δ*nasA* bacteria were limited by N. However, the addition of DOC to Δ*nasA* bacteria co-cultures did result in a slight bacterial growth increase (**Figure 3B**). Thus, it is possible that a co-limitation scenario may also arise in low concentrations of both bioavailable N and C.

When diatoms interact with bacteria that can utilize NO_3^-, the concentration of both NO_3^- and DOC present may impact their dynamics in complex ways. To explore this, we tested the effects of various NO_3^- and DOC levels on population dynamics of *P. tricornutum–A. macleodii* co-cultures. We observed that *P. tricornutum* cell densities were largely regulated by NO_3^- concentrations rather than DOC level or bacteria cell densities. An exception was at a low NO_3^- concentrations (50 μM), where bacterial accumulation of NO_3^- linked to increased bacteria cell numbers appeared to reduce diatom growth and final cell densities. This trend was not observed in the higher NO_3^- treatments, where diatom cell densities were similar between 50 μM and 1 mM DOC concentrations despite an increase in bacteria cell density. This suggests that at low NO_3^- levels, bacterial growth made possible by DOC availability can have a negative effect on diatom cell numbers. Based on a similar pattern observed in the DOC addition study discussed above, where NO_3^- drawdown was correlated with high bacterial and low diatom cell numbers in high DOC conditions, we hypothesize that in this experiment fast bacterial acquisition of NO_3^- in the media made NO_3^- limiting for diatom growth. At higher NO_3^- concentrations this effect was not observed, which could be explained by the diatoms having sufficient opportunity to acquire and store N since more was available. The diatoms may also be able to utilize nitrogen derived from dead or growth-arrested bacteria during stationary phase. This is supported by the observation that bacterial abundance peaked during the exponential phase of *P. tricornutum* regardless of NO_3^- level or DOC addition and subsequently declined, with the exception of the 1 mM NO_3^-/50 μM DOC treatment.

Bacterial cell numbers were strongly affected by DOC concentration. Without the addition of DOC, bacterial cell densities generally increased with increasing NO_3^- and corresponding increases in diatom cell density. This likely reflects the link between diatom population and DOC availability; higher diatom density as a result of higher N concentrations leads to higher total DOC available for bacteria utilization. It is also possible that changes in DOC composition may affect bacterial growth dynamics. At higher DOC concentrations, bacteria could utilize NO_3^- in the media to reach higher cell densities than with ambient DOC alone, a result consistent with other DOC addition experiments conducted in this study. By late stationary phase (Day 31), bacteria cell densities were variable. This may be the result of complex N and C recycling dynamics following initial uptake by bacteria and diatoms, and further studies may help to elucidate how cellular responses of diatoms and bacteria lead to the population changes we observed.

Exchange of Nitrogen Substrates between Diatoms and Bacteria

Our findings suggest that the diatoms and bacteria in our model system are able to exchange nitrogen substrates with each other. A large portion of N consumed by phytoplankton is ultimately released as dissolved organic nitrogen (DON) in oceanic, coastal, and estuarine environments (Wheeler and Kirchman, 1986; Bronk et al., 1994; Berman and Bronk, 2003), and is a valuable source of N for marine bacteria. Previous studies on *P. tricornutum* have shown that both organic C and N are

released by *P. tricornutum*, and that concentrations increase after the cells enter stationary phase (Chen and Wangersky, 1996; Pujo-Pay et al., 1997). The bacteria in the present study were able to survive and grow in co-culture with the diatom using diatom-derived organic C, and bacteria concentrations increased after the onset of diatom stationary phase, which is consistent with utilization of diatom derived organic matter. We further observed that the Δ*nasA* bacteria strain in co-culture grew despite the inability to use NO_3^-, which was the only N source provided in the ASW media. This leaves diatom-derived N as the most plausible source for bacterial growth.

While the paradigm for phytoplankton-bacteria relationships is typically that of bacteria utilizing phytoplankton-derived organic matter, and in some cases exchanging various substrates to facilitate this acquisition, few studies have examined if and how diatoms may use bacterial-derived N. Given the high rate of bacterial turnover in the ocean, this could potentially represent an important N source, especially under N limiting conditions. Diatoms are known to utilize a variety of organic and inorganic N substrates (Bronk et al., 1994, 2006; Waser et al., 1998). One recent study suggests that bacteria may use NH_4^+ as a diatom signaling molecule (Amin et al., 2015). Our results show that the NRKO diatom lines could survive and grow normally in NH_4^+ but not NO_3^-. The addition of *A. macleodii* to the diatom cultures without a coincident DOC media amendment resulted in low bacterial cell numbers and did not increase NRKO diatom cell numbers, suggesting that either the bacteria do not provide the diatom with useable N substrates in this physiological state, or that the amount supplied is not sufficient for diatom growth. Potentially, the bacteria could provide the diatoms with useable N, but the amount produced by the low cell numbers observed was not enough to detect diatom population recovery using our methods. Our study does not address other possible interactions at the cellular level (e.g., metabolic shifts perhaps observable using metabolomics or transcriptomics), which may clarify whether N is being provided by bacteria in this co-culture. When DOC was added to the NRKO co-cultures in NO_3^- media, both bacteria and diatom growth increased substantially. This strongly suggests that the bacteria supply the diatoms with a N substrate, perhaps made possible by the large bacterial numbers resulting from the DOC addition, followed by nitrogen release upon the onset of carbon-limited stationary phase. Alternatively or concurrently, after reaching maximum cell density early in the experiment the bacteria may subsequently die allowing *P. tricornutum* to recycle some of the organic N from the dead bacterial cells. Some bacteria were detected in the *P. tricornutum* NRKO cultures without the addition of *A. macleodii*. The addition of DOC to *A. macleodii* amended cultures resulted in high bacterial densities similar to what was observed in the *P. tricornutum* WT–*A. macleodii* co-cultures (**Figure 3**), thus we believe *A. macleodii* growth was responsible for the observed increase in *P. tricornutum* NRKO cell numbers. Even if this is not the case (i.e., the growth of other bacteria present contributed to the high bacterial cell numbers), our results strongly suggest that the high abundance of bacteria due to DOC addition was linked to diatom recovery, and our study presents a proof of concept that diatoms can utilize bacterial-derived N. Though the scope of this study does not directly address the specific bacterial N source, the transcriptomic analysis conducted during the baseline experiment elucidates N-related diatom cellular pathways that may be influenced by bacteria, such as those involved in NO_3^- and NH_4^+ acquisition and metabolism.

Further Insights into "Bacterial Responsive" Genes Revealed by RNA-seq

Using RNA-seq, we were able to identify several putative *P. tricornutum* genes that were responsive to the presence of bacteria, some of which suggest N-related interactions. A recent study examined the response of two *Prochlorococcus* strains in co-culture with *A. macleodii* bacteria, and they also found that several algal genes were bacterial-responsive (Aharonovich and Sher, 2016). The study examines prokaryotic rather than eukaryotic gene expression, and is thus difficult to compare. However, it demonstrates the value of transcriptome analyses in developing hypotheses about microbial interactions and, along with our study, lays the framework for a robust analysis of interactions involving *A. macleodii* bacteria with both prokaryotic and eukaryotic algae in a model laboratory system.

All differential gene expression in our study was observed in the stationary samples, long after NO_3^- had been depleted from the media (**Figure 1**), and cells were potentially N stressed. One discernable pattern is related to downregulation of multiple N transporters (NO_2^- and NO_3^-) in diatom cultures containing bacteria (**Figure 2**, Supplementary Table 4). One of these transporters (Phatr2ID: 49093, Phatr3 ID: 302957) was one of the mostly highly significantly DE putative genes identified in the data set. Orthologs of this ion transporter in *Arabidopsis* has been shown to bind to and sense NO_3^- (Huang et al., 1996; Liu et al., 1999), suggesting a possible role in diatom NO_3^- acquisition. Inorganic N transporters including those DE in our dataset are commonly upregulated during nitrogen stress conditions (Levitan et al., 2014; Alipanah et al., 2015; JGI genome annotation, Allen, 2006). Downregulation in cultures containing bacteria suggests that bacteria are contributing to alleviation of diatom N stress, which is further supported by the higher diatom cell numbers observed in co-cultures during stationary phase and also our finding that under certain N stress conditions (i.e., high DOC present) bacteria have the ability to provide diatoms with N in forms other than NO_3^-. In Levitan et al. (2014), expression of glutamine synthetase II (GSII) followed a similar expression pattern of downregulation under N-stress, and we also observed significant downregulation of this gene in bacteria-containing co-cultures. Another gene related to diatom N metabolism, ornithine cyclodeaminase (Phatr2: 54222, Phatr3: 305662), is involved in the diatom Urea cycle and was downregulated compared to *P. tricornutum* monocultures (Allen et al., 2011). While these two genes were only significantly DE between the *P. tricornutum* monoculture and the Δ*nasA* bacterial co-cultures (Supplementary Table 4), they were downregulated in both co-cultures indicating a common bacterial-responsive pattern.

Some NH_4^+ transporters have also been shown to be upregulated during N stress, while others are downregulated or unaffected (Levitan et al., 2014; Alipanah et al., 2015).

Examination of NH_4^+ acquisition and transport genes in our dataset did not reveal any significant DE between *P. tricornutum* monocultures and co-cultures (Supplementary Table 5). In Amin et al. (2015), it was suggested that bacteria provide NH_4^+ to diatoms as a signaling molecule, which was partially supported by the upregulation of NH_4^+ transport genes in bacteria, as well as an increase in NH_4^+ measured in co-cultures. However, NH_4^+ transport genes were not DE in the diatom they examined, which may suggest that the cellular impacts of this potential exchange are not apparent at the transcriptional level in diatoms.

Several other bacterial responsive genes, including the many unannotated hypothetical proteins in our dataset, may be interesting candidates for further investigation. Several heat shock transcription factors were upregulated in diatom-bacteria co-cultures compared to monocultures. These are transcriptional regulators of heat shock proteins involved in cellular stress responses (Sorger, 1991). High expression may indicate a stress response caused by the bacteria, however, little is known about the regulation of this complex pathway in diatoms, and many non-transcriptional steps are involved in heat shock protein expression and regulation (Sorger, 1991). Upregulation of other putative genes in co-cultures may play a role in exchange of important metabolites or intracellular signaling pathways. These include a putative sugar transporter (Phatr2 ID: 49722, Phatr3 ID: 311238) which was >2 fold upregulated in *P. tricornutum* in both co-cultures (Supplementary Table 4), and the two putative genes that were most highly upregulated and significantly DE in both co-cultures: a putative short chain dehydrogenase (Phatr2 ID: 13001, Phatr3 ID: 306282) and a putative fatty acid hydroxylase (Phatr3 ID: 308140). Short chain dehydrogenases in particular have been shown to serve as molecular links between nutrient signaling and plant hormone biosynthesis in *Arabidopsis* (Cheng, 2002). Further analysis using our model system may allow determination of the role of such genes in diatom-bacteria interactions.

Using the genetically tractable model system developed in this study, we have described mechanisms of interaction between diatoms and bacteria that may be of global biogeochemical significance. Our data strongly suggests bidirectional exchange of N substrates between diatoms and bacteria, and revealed putative diatom genes and pathways that may be impacted by the presence of bacteria and involved in N exchange. Furthermore, we have demonstrated that under certain environmental conditions (i.e., high DOC), marine bacteria are able to effectively compete with diatoms for NO_3^-, which may influence predictions of primary productivity and nutrient utilization by phytoplankton in the ocean and associated estimates of C export via the biological pump.

AUTHOR CONTRIBUTIONS

RD and AA designed research; RD, SS, and HZ performed research; RD, SS, and JM analyzed data; and RD and AA wrote the paper.

FUNDING

Funding was provided to AA by the Gordon and Betty Moore Foundation (GBMF3828 and GBMF5006), the US Department of Energy (DE-SC0008593) and the National Science Foundation (OCE-1136477). Funding was provided to RD by the UCSD/SIO Center for Marine Biodiversity and Conservation.

ACKNOWLEDGMENTS

We would like to acknowledge the following people for their contributions to this research project: Phillip Weyman for assistance with *A. macleodii* culturing and genetic engineering, Erin Bertrand for assistance with co-culturing techniques, Chris Dupont for assisting with organic carbon and nitrogen analysis, Mark Novotny for flow cytometry assistance, and Flip McCarthy for helping with data interpretation and engineering the *P. tricornutum* NRKO line. We would also like to thank Karen Beeri for sequencing assistance.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fmicb.2016.00880

REFERENCES

Aharonovich, D., and Sher, D. (2016). Transcriptional response of *Prochlorococcus* to co-culture with a marine *Alteromonas*: differences between strains and the involvement of putative infochemicals. *ISME J.* doi: 10.1038/ismej.2016.70. [Epub ahead of print].

Alipanah, L., Rohloff, J., Winge, P., Bones, A. M., and Brembu, T. (2015). Whole-cell response to nitrogen deprivation in the diatom *Phaeodactylum tricornutum*. *J. Exp. Bot.* 66, 6281–6296. doi: 10.1093/jxb/erv340

Allen, A. E. (2006). Transcript annotation: JGI Genome Portal *Phaeodactylum tricornutum*. Available online at: http://genome.jgi.doe.gov/annotator/servlet/jgi.annotation.Annotation?pDb=Phatr2&pStateVar=View&pProteinId=26029&pViewType=protein&pDummy=699949276669001 (Accessed March 15, 2016).

Allen, A. E., Booth, M. G., Frischer, M. E., Verity, P. G., Zehr, J. P., and Zani, S. (2001). Diversity and detection of nitrate assimilation genes in marine bacteria. *Appl. Environ. Microbiol.* 67, 5343–5348. doi: 10.1128/AEM.67.11.5343-5348.2001

Allen, A. E., Booth, M. G., Verity, P. G., and Frischer, M. E. (2005). Influence of nitrate availability on the distribution and abundance of heterotrophic bacterial nitrate assimilation genes in the Barents Sea during summer. *Aquat. Microb. Ecol.* 39, 247–255. doi: 10.3354/ame039247

Allen, A. E., Dupont, C. L., Oborník, M., Horák, A., Nunes-Nesi, A., McCrow, J. P., et al. (2011). Evolution and metabolic significance of the urea cycle in photosynthetic diatoms. *Nature* 473, 203–207. doi: 10.1038/nature10074

Allen, A. E., Laroche, J., Maheswari, U., Lommer, M., Schauer, N., Lopez, P. J., et al. (2008). Whole-cell response of the pennate diatom *Phaeodactylum tricornutum* to iron starvation. *Proc. Natl. Acad. Sci. U.S.A.* 105, 10438–10443. doi: 10.1073/pnas.0711370105

Allen, A. E., Vardi, A., and Bowler, C. (2006). An ecological and evolutionary context for integrated nitrogen metabolism and related signaling pathways in marine diatoms. *Curr. Opin. Plant Biol.* 9, 264–273. doi: 10.1016/j.pbi.2006.03.013

Allen, L. Z., Allen, E. E., Badger, J. H., McCrow, J. P., Paulsen, I. T., Elbourne, L. D., et al. (2012). Influence of nutrients and currents on the genomic composition of microbes across an upwelling mosaic. *ISME J.* 6, 1403–1414. doi: 10.1038/ismej.2011.201

Amin, S. A., Hmelo, L. R., van Tol, H. M., Durham, B. P., Carlson, L. T., Heal, K. R., et al. (2015). Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. *Nature* 522, 98–101. doi: 10.1038/nature14488

Amin, S. A., Parker, M. S., and Armbrust, E. V. (2012). Interactions between diatoms and bacteria. *Microbiol. Mol. Biol. Rev.* 76, 667–684. doi: 10.1128/MMBR.00007-12

Azam, F., and Malfatti, F. (2007). Microbial structuring of marine ecosystems. *Nat. Rev. Microbiol.* 5, 782–791. doi: 10.1038/nrmicro1747

Bell, W., and Mitchell, R. (1972). Chemotactic and growth responses of marine bacteria to algal extracellular products. *Biol. Bull.* 143, 265–277. doi: 10.2307/1540052

Berman, T., and Bronk, D. a. (2003). Dissolved organic nitrogen: a dynamic participant in aquatic ecosystems. *Aquat. Microb. Ecol.* 31, 279–305. doi: 10.3354/ame031279

Bertrand, E. M., McCrow, J. P., Moustafa, A., Zheng, H., McQuaid, J. B., Delmont, T. O., et al. (2015). Phytoplankton–bacterial interactions mediate micronutrient colimitation at the coastal Antarctic sea ice edge. *Proc. Natl. Acad. Sci. U.S.A.* 112, 9938–9943. doi: 10.1073/pnas.1501615112

Bowler, C., Allen, A. E., Badger, J. H., Grimwood, J., Jabbari, K., Kuo, A., et al. (2008). The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* 456, 239–244. doi: 10.1038/nature07410

Bradley, P. B., Lomas, M. W., and Bronk, D. A. (2010a). Inorganic and organic nitrogen use by phytoplankton along chesapeake bay, measured using a flow cytometric sorting approach. *Estuaries Coasts* 33, 971–984. doi: 10.1007/s12237-009-9252-y

Bradley, P. B., Sanderson, M. P., Frischer, M. E., Brofft, J., Booth, M. G., Kerkhof, L. J., et al. (2010b). Inorganic and organic nitrogen uptake by phytoplankton and heterotrophic bacteria in the stratified Mid-Atlantic Bight. *Estuar. Coast. Shelf Sci.* 88, 429–441. doi: 10.1016/j.ecss.2010.02.001

Bradley, P. B., Sanderson, M. P., Nejstgaard, J. C., Sazhin, A. F., Frischer, M. E., Killberg-Thoreson, L. M., et al. (2010c). Nitrogen uptake by phytoplankton and bacteria during an induced *Phaeocystis pouchetii* bloom, measured using size fractionation and flow cytometric sorting. *Aquat. Microb. Ecol.* 61, 89–104. doi: 10.3354/ame01414

Bratbak, G., and Thingstad, T. F. (1985). Phytoplankton-bacteria interactions: an apparant paradox? Analysis of a model system with both competition and commensalism. *Mar. Ecol. Prog. Ser.* 25, 23–30. doi: 10.3354/meps025023

Bronk, D. A., Glibert, P. M., and Ward, B. B. (1994). Nitrogen uptake, dissolved organic nitrogen release, and new production. *Science* 265, 1843–1846. doi: 10.1126/science.265.5180.1843

Bronk, D. A., See, J. H., Bradley, P., and Killberg, L. (2006). DON as a source of bioavailable nitrogen for phytoplankton. *Biogeosci. Discuss.* 3, 1247–1277. doi: 10.5194/bgd-3-1247-2006

Buchan, A., LeCleir, G. R., Gulvik, C. A., and González, J. M. (2014). Master recyclers: features and functions of bacteria associated with phytoplankton blooms. *Nat. Rev. Microbiol.* 12, 686–696. doi: 10.1038/nrmicro3326

Cermeño, P., Lee, J.-B., Wyman, K., Schofield, O., and Falkowski, P. G. (2011). Competitive dynamics in two species of marine phytoplankton under non-equilibrium conditions. *Mar. Ecol. Prog. Ser.* 429, 19–28. doi: 10.3354/meps09088

Chen, W., and Wangersky, P. J. (1996). Production of dissolved organic carbon in phytoplankton cultures as measured by high-temperature catalytic oxidation and ultraviolet photo- oxidation methods. *J. Plankton Res.* 18, 1201–1211. doi: 10.1093/plankt/18.7.1201

Cheng, W.-H. (2002). A unique short-chain dehydrogenase/reductase in *Arabidopsis* glucose signaling and abscisic acid biosynthesis and functions. *Plant Cell Online* 14, 2723–2743. doi: 10.1105/tpc.006494

Collos, Y., Mornet, F., Sciandra, A., Waser, N., Larson, A., and Harrison, P. J. (1999). An optical method for the rapid measurement of micromolar concentrations of nitrate in marine phytoplankton cultures. *J. Appl. Phycol.* 11, 179–184. doi: 10.1023/A:1008046023487

Dortch, Q. (1990). The interaction between ammonium and nitrate uptake in phytoplankton. *Mar. Ecol. Prog. Ser.* 61, 183–201. doi: 10.3354/meps061183

Dugdale, R. C., and Goering, J. J. (1967). Uptake of new and regenerated forms of nitrogen in primary productivity. *Limnol. Oceanogr.* 12, 196–206. doi: 10.4319/lo.1967.12.2.0196

Durham, B. P., Sharma, S., Luo, H., Smith, C. B., Amin, S. A., Bender, S. J., et al. (2014). Cryptic carbon and sulfur cycling between surface ocean plankton. *Proc. Natl. Acad. Sci. USA.* 112, 453–457. doi: 10.1073/pnas.1413137112

Falkowski, P. G., and Oliver, M. J. (2007). Mix and match: how climate selects phytoplankton. *Nat. Rev. Microbiol.* 5, 813–819. doi: 10.1038/nrmicro1751

Fuhrman, J. A. (1989). Dominance of bacterial biomass in the Sargasso Sea and its ecological implications. *Mar. Ecol. Prog. Ser.* 57, 207–217. doi: 10.3354/meps057207

García-Martínez, J., Acinas, S. G., Massana, R., and Rodríguez-Valera, F. (2002). Prevalence and microdiversity of *Alteromonas macleodii*-like microorganisms in different oceanic regions. *Environ. Microbiol.* 4, 42–50. doi: 10.1046/j.1462-2920.2002.00255.x

Geider, R. J., La Roche, J., Greene, R. M., and Olaizola, M. (1993). Response of the photosynthetic apparatus of *Phaeodactylum tricornutum* (*Bacillariophycea*) to nitrate, phosphate, or iron starvation. *J. Phycol.* 29, 755–766. doi: 10.1111/j.0022-3646.1993.00755.x

Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., and Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* 6, 343–345. doi: 10.1038/nmeth.1318

Grossart, H., and Simon, M. (2007). Interactions of planktonic algae and bacteria: effects on algal growth and organic matter dynamics. *Aquat. Microb. Ecol.* 47, 163–176. doi: 10.3354/ame047163

Huang, N. C., Chiang, C. S., Crawford, N. M., and Tsay, Y. F. (1996). CHL1 encodes a component of the low-affinity nitrate uptake system in *Arabidopsis* and shows cell type-specific expression in roots. *Plant Cell* 8, 2183–2191. doi: 10.1105/tpc.8.12.2183

Ivars-Martínez, E., D'Auria, G., Rodríguez-Valera, F., Sánchez-Porro, C., Ventosa, A., Joint, I., et al. (2008). Biogeography of the ubiquitous marine bacterium *Alteromonas macleodii* determined by multilocus sequence analysis. *Mol. Ecol.* 17, 4092–4106. doi: 10.1111/j.1365-294X.2008.03883.x

Jiang, X., Dang, H., and Jiao, N. (2015). Ubiquity and diversity of heterotrophic bacterial nasA genes in diverse marine environments. *PLoS ONE* 10:e0117473. doi: 10.1371/journal.pone.0117473

Jiao, N., Herndl, G. J., Hansell, D. A., Benner, R., Kattner, G., Wilhelm, S. W., et al. (2011). The microbial carbon pump and the oceanic recalcitrant dissolved organic matter pool. *Nat. Rev. Microbiol.* 9, 555. doi: 10.1038/nrmicro2386-c5

Johnson, K. S., and Coletti, L. J. (2002). *In situ* ultraviolet spectrophotometry for high resolution and long-term monitoring of nitrate, bromide and bisulfide in the ocean. *Deep. Res. Part I Oceanogr. Res. Pap.* 49, 1291–1305. doi: 10.1016/S0967-0637(02)00020-1

Karas, B. J., Diner, R. E., Lefebvre, S. C., McQuaid, J., Phillips, A. P. R., Noddings, C. M., et al. (2015). Designer diatom episomes delivered by bacterial conjugation. *Nat. Commun.* 6:6925. doi: 10.1038/ncomms7925

Kato, J., Amie, J., Murata, Y., Kuroda, A., Mitsutani, A., and Ohtake, H. (1998). Development of a genetic transformation system for an Alga-Lysing Bacterium. *Appl. Environ. Microbiol.* 64, 2061–2064.

Keil, R., and Kirchman, D. (1991). Contribution of dissolved free amino acids and ammonium to the nitrogen requirements of heterotrophic bacterioplankton. *Mar. Ecol. Prog. Ser.* 73, 1–10. doi: 10.3354/meps073001

Kovach, M. E., Elzer, P. H., Hill, D. S., Robertson, G. T., Farris, M. A., Roop, R. M., et al. (1995). Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. *Gene* 166, 175–176. doi: 10.1016/0378-1119(95)00584-1

Levitan, O., Dinamarca, J., Zelzion, E., Lun, D. S., Guerra, L. T., Kim, M. K., et al. (2014). Remodeling of intermediate metabolism in the diatom *Phaeodactylum tricornutum* under nitrogen stress. *Proc. Natl. Acad. Sci. U.S.A.* 112, 412–417. doi: 10.1073/pnas.1419818112

Liao, Y., Smyth, G. K., and Shi, W. (2014). FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930. doi: 10.1093/bioinformatics/btt656

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv: 1303.3997 [q-bio.GN].

Liu, K. H., Huang, C. Y., and Tsay, Y. F. (1999). CHL1 is a dual-affinity nitrate transporter of *Arabidopsis* involved in multiple phases of nitrate uptake. *Plant Cell* 11, 865–874. doi: 10.1105/tpc.11.5.865

Logan, B. E., Passow, U., Alldredge, A. L., Grossartt, H.-P., and Simont, M. (1995). Rapid formation and sedimentation of large aggregates is predictable from coagulation rates (half-lives) of transparent exopolymer particles (TEP). *Deep Sea Res. Part II Top. Stud. Oceanogr.* 42, 203–214. doi: 10.1016/0967-0645(95)00012-F

Lomas, M. W., Bronk, D. A., and van den Engh, G. (2011). Use of flow cytometry to measure biogeochemical rates and processes in the ocean. *Ann. Rev. Mar. Sci.* 3, 537–566. doi: 10.1146/annurev-marine-120709-142834

Lomas, M. W., and Gilbert, P. M. (2000). Comparisons of nitrate uptake, storage, and reduction in marine diatoms and flagellates. *J. Phycol.* 36, 903–913. doi: 10.1046/j.1529-8817.2000.99029.x

López-Pérez, M., Gonzaga, A., Martin-Cuadrado, A.-B., Onyshchenko, O., Ghavidel, A., Ghai, R., et al. (2012). Genomes of surface isolates of *Alteromonas macleodii*: the life of a widespread marine opportunistic copiotroph. *Sci. Rep.* 2:696. doi: 10.1038/srep00696

López-Pérez, M., and Rodriguez-Valera, F. (2014). "The family *Alteromonadaceae*." in *The Prokaryotes,* eds E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, and F. Thompson (Berlin; Heidelberg: Springer), 69–92. doi: 10.1007/978-3-642-38922-1_233

Matthijs, M., Fabris, M., Broos, S., Vyverman, W., and Goossens, A. (2015). Profiling of the early nitrogen stress response in the diatom *Phaeodactylum tricornutum* reveals a novel family of RING-domain transcription factors. *Plant Physiol.* 170, 489–498. doi: 10.1104/pp.15.01300

Mayali, X., and Azam, F. (2004). Algicidal bacteria in the sea and their impact on algal blooms. *J. Eukaryot. Microbiol.* 51, 139–144. doi: 10.1111/j.1550-7408.2004.tb00538.x

Morris, J. J., Johnson, Z. I., Szul, M. J., Keller, M., and Zinser, E. R. (2011). Dependence of the cyanobacterium *Prochlorococcus* on hydrogen peroxide scavenging microbes for growth at the ocean's surface. *PLoS ONE* 6:e16805. doi: 10.1371/journal.pone.0016805

Morris, J. J., Kirkegaard, R., Szul, M. J., Johnson, Z. I., and Zinser, E. R. (2008). Facilitation of robust growth of Prochlorococcus colonies and dilute liquid cultures by "helper" heterotrophic bacteria. *Appl. Environ. Microbiol.* 74, 4530–4534. doi: 10.1128/AEM.02479-07

Morrissey, J., Sutak, R., Paz-Yepes, J., Tanaka, A., Moustafa, A., Veluchamy, A., et al. (2015). A novel protein, ubiquitous in marine phytoplankton, concentrates iron at the cell surface and facilitates uptake. *Curr. Biol.* 25, 364–371. doi: 10.1016/j.cub.2014.12.004

Mouriño-Pérez, R. R., Worden, A. Z., and Azam, F. (2003). Growth of *Vibrio cholerae* O1 in red tide waters off California. *Appl. Environ. Microbiol.* 69, 6923–6931. doi: 10.1128/AEM.69.11.6923-6931.2003

Nymark, M., Sharma, A. K., Sparstad, T., Bones, A. M., and Winge, P. (2016). A CRISPR/Cas9 system adapted for gene editing in marine algae. *Sci. Rep.* 6:24951. doi: 10.1038/srep24951

Parsons, T. R., Maita, Y., and Lalli, C. M. (1984). *A Manual of Chemical and Biological Methods for Seawater Analysis*. Oxford: Pergamon.

Pedler, B. E., Aluwihare, L. I., and Azam, F. (2014). Single bacterial strain capable of significant contribution to carbon cycling in the surface ocean. *Proc. Natl. Acad. Sci. U.S.A.* 111, 7202–7207. doi: 10.1073/pnas.1401887111

Pedler Sherwood, B., Shaffer, E. A., Reyes, K., Longnecker, K., Aluwihare, L. I., and Azam, F. (2015). Metabolic characterization of a model heterotrophic bacterium capable of significant chemical alteration of marine dissolved organic matter. *Mar. Chem.* 177, 357–365. doi: 10.1016/j.marchem.2015.06.027

Pomeroy (1974). The ocean's food web, a changing paradigm. *Bioscience* 24, 499–504. doi: 10.2307/1296885

Puddu, A., Zoppini, A., Fazi, S., Rosati, M., Amalfitano, S., and Magaletti, E. (2003). Bacterial uptake of DOM released from P-limited phytoplankton. *FEMS Microbiol. Ecol.* 46, 257–268. doi: 10.1016/S0168-6496(03)00197-1

Pujo-Pay, M., Conan, P., and Raimbault, P. (1997). Excretion of dissolved organic nitrogen by phytoplankton assessed by wet oxidation and 1 5 N tracer procedures. *Mar. Ecol. Prog. Ser.* 153, 99–111. doi: 10.3354/meps153099

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616

Sarmento, H., and Gasol, J. M. (2012). Use of phytoplankton-derived dissolved organic carbon by different types of bacterioplankton. *Environ. Microbiol.* 14, 2348–2360. doi: 10.1111/j.1462-2920.2012.02787.x

Serra, J. L., Llama, M. J., and Cadenas, E. (1978). Nitrate utilization by the diatom *Skeletonema costatum*: II. regulation of nitrate uptake. *Plant Physiol.* 62, 991–994. doi: 10.1104/pp.62.6.991

Sher, D., Thompson, J. W., Kashtan, N., Croal, L., and Chisholm, S. W. (2011). Response of *Prochlorococcus* ecotypes to co-culture with diverse marine bacteria. *ISME J.* 5, 1125– 1132. doi: 10.1038/ismej.2011.1

Shi, Y., McCarren, J., and Delong, E. F. (2012). Transcriptional responses of surface water marine microbial assemblages to deep-sea water amendment. *Environ. Microbiol.* 14, 191–206. doi: 10.1111/j.1462-2920.2011.02598.x

Siaut, M., Heijde, M., Mangogna, M., Montsant, A., Coesel, S., Allen, A., et al. (2007). Molecular toolbox for studying diatom biology in *Phaeodactylum tricornutum*. *Gene* 406, 23–35. doi: 10.1016/j.gene.2007.05.022

Singleton, F. L., Attwell, R., and Jangi, S. (1982). Effects of temperature and salinity on *Vibrio cholerae*. *Appl. Environ. Microbiol.* 44, 1047–1058.

Smriga, S., Fernandez, V. I., Mitchell, J. G., and Stocker, R. (2016). Chemotaxis toward phytoplankton drives organic matter partitioning among marine bacteria. *Proc. Natl. Acad. Sci. U.S.A.* 113, 1576–1581. doi: 10.1073/pnas.1512307113

Sorger, P. K. (1991). Heat shock factor and the heat shock response. *Cell* 65, 363–366. doi: 10.1016/0092-8674(91)90452-5

Stocker, R. (2012). Marine microbes see a sea of gradients. *Science* 338, 628–633. doi: 10.1126/science.1208929

Thingstad, T. F., Skjoldal, E. F., and Bohne, R., a (1993). Phosphorus cycling and algal bacterial competition in Sandsfjord, Western Norway. *Mar. Ecol. Prog. Ser.* 99, 239–259. doi: 10.3354/meps099239

Wang, H., Tomasch, J., Jarek, M., and Wagner-Dobler, I. (2014). A dual-species co-cultivation system to study the interactions between *Roseobacters* and dinoflagellates. *Front. Microbiol.* 5:311. doi: 10.3389/fmicb.2014.00311

Waser, N. A., Yu, Z., Tada, K., Paul, J., Turpin, D. H., and Calvertl, S. E. (1998). Nitrogen isotope fractionation during nitrate, ammonium and urea uptake by marine diatoms and coccolithophores under various conditions of N availability. *Mar. Ecol. Prog. Ser.* 169, 29–41. doi: 10.3354/meps169029

Wawrik, B., Boling, W. B., van Nostrand, J. D., Xie, J., Zhou, J., and Bronk, D. A. (2012). Assimilatory nitrate utilization by bacteria on the West Florida Shelf as determined by stable isotope probing and functional microarray analysis. *FEMS Microbiol. Ecol.* 79, 400–411. doi: 10.1111/j.1574-6941.2011.01226.x

Weyman, P. D., Beeri, K., Lefebvre, S. C., Rivera, J., McCarthy, J. K., Heuberger, A. L., et al. (2015). Inactivation of Phaeodactylum tricornutum urease gene using transcription activator-like effector nuclease-based targeted mutagenesis. *Plant Biotechnol. J.* 13, 460–470. doi: 10.1111/pbi.12254

Weyman, P. D., Smith, H. O., and Xu, Q. (2011). Genetic analysis of the *Alteromonas macleodii* [NiFe]-hydrogenase. *FEMS Microbiol. Lett.* 322, 180–187. doi: 10.1111/j.1574-6968.2011.02348.x

Wheeler, P. A., and Kirchman, D. L. (1986). Utilization of inorganic and organic nitrogen by bacteria in marine systems. *Limnol. Oceanogr.* 31, 998–1009. doi: 10.4319/lo.1986.31.5.0998

Whitman, W. B., Coleman, D. C., and Wiebe, W. J. (1998). Prokaryotes: the unseen majority. *Proc. Natl. Acad. Sci. U.S.A.* 95, 6578–6583. doi: 10.1073/pnas.95.12.6578

Worden, A. Z., Follows, M. J., Giovannoni, S. J., Wilken, S., Zimmerman, A. E., and Keeling, P. J. (2015). Rethinking the marine carbon cycle: factoring in the multifarious lifestyles of microbes. *Science* 347:1257594. doi: 10.1126/science.1257594

Yongmanitchai, W., and Ward, O. P. (1991). Growth of and Omega-3 fatty acid production by *Phaeodactylum tricornutum* under different culture conditions. *Appl. Environ. Microbiol.* 57, 419–425.

Zehr, J. P., and Ward, B. B. (2002). Nitrogen cycling in the ocean: new perspectives on processes and paradigms. *Appl. Environ. Microbiol.* 68, 1015–1024. doi: 10.1128/AEM.68.3.1015-1024.2002

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Chapter 1, in full, is a reprint of the material as it appears in **Diner, RE**, Schwenck, SM, McCrow, JP, Zheng, H, & Allen, AE (2016) Genetic Manipulation of Competition for Nitrate between Heterotrophic Bacteria and Diatoms. *Frontiers in Microbiology*, 7, 880. The dissertation author was the primary investigator and author of this material.

CHAPTER 2

Refinement of the diatom episome maintenance sequence and improvement of conjugation-based DNA delivery methods

Synopsis

In this chapter, I demonstrate that low GC content is the characteristic feature responsible for stable maintenance of foreign DNA episomes in diatom nuclei, and that yeast centromere and origins of replication do not possess orthologous functions in the diatom *P. tricornutum*. I also present advancements in technical methods advancing the use of bacterial conjugation as a transgene delivery tool for diatoms.

This chapter is presented as a paper. “Refinement of the diatom episome maintenance sequence and improvement of conjugation-based DNA delivery methods” was published as a research article in *Frontiers in Bioengineering and Biotechnology* in 2016.

frontiers in Bioengineering and Biotechnology

METHODS
published: 08 August 2016
doi: 10.3389/fbioe.2016.00065

Refinement of the Diatom Episome Maintenance Sequence and Improvement of Conjugation-Based DNA Delivery Methods

*Rachel E. Diner[1,2], Vincent A. Bielinski[3], Christopher L. Dupont[1], Andrew E. Allen[1,2] and Philip D. Weyman[3]**

[1] *Microbial and Environmental Genomics Group, J. Craig Venter Institute, La Jolla, CA, USA,* [2] *Integrative Oceanography Division, Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA, USA,* [3] *Synthetic Biology and Bioenergy Group, J. Craig Venter Institute, La Jolla, CA, USA*

Conjugation of episomal plasmids from bacteria to diatoms advances diatom genetic manipulation by simplifying transgene delivery and providing a stable and consistent gene expression platform. To reach its full potential, this nascent technology requires new optimized expression vectors and a deeper understanding of episome maintenance. Here, we present the development of an additional diatom vector (pPtPBR1), based on the parent plasmid pBR322, to add a plasmid maintained at medium copy number in *Escherichia coli* to the diatom genetic toolkit. Using this new vector, we evaluated the contribution of individual yeast DNA elements comprising the 1.4-kb tripartite *CEN6-ARSH4-HIS3* sequence that enables episome maintenance in *Phaeodactylum tricornutum*. While various combinations of these individual elements enable efficient conjugation and high exconjugant yield in *P. tricornutum*, individual elements alone do not. Conjugation of episomes containing *CEN6-ARSH4* and a small sequence from the low GC content 3′ end of *HIS3* produced the highest number of diatom exconjugant colonies, resulting in a smaller and more efficient vector design. Our findings suggest that the *CEN6* and *ARSH4* sequences function differently in yeast and diatoms, and that low GC content regions of greater than ~500 bp are a potential indicator of a functional diatom episome maintenance sequence. Additionally, we have developed improvements to the conjugation protocol including a high-throughput option utilizing 12-well plates and plating methods that improve exconjugant yield and reduce time and materials required for the conjugation protocol. The data presented offer additional information regarding the mechanism by which the yeast-derived sequence enables diatom episome maintenance and demonstrate options for flexible vector design.

Keywords: diatom, bacteria, conjugation, genetic tools, episome, DNA delivery, DNA replication, *Phaeodactylum*

OPEN ACCESS

Edited by:
Antonio Trincone, Consiglio Nazionale delle Ricerche, Italy

Reviewed by:
Aldo Nicosia, Institute for Coastal Marine Environment (CNR), Italy
Senjie Lin, University of Connecticut, USA

***Correspondence:**
Philip D. Weyman
pweyman@jcvi.org

Specialty section:
This article was submitted to Marine Biotechnology, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *30 June 2016*
Accepted: *21 July 2016*
Published: *08 August 2016*

Citation:
Diner RE, Bielinski VA, Dupont CL, Allen AE and Weyman PD (2016) Refinement of the Diatom Episome Maintenance Sequence and Improvement of Conjugation-Based DNA Delivery Methods. Front. Bioeng. Biotechnol. 4:65. doi: 10.3389/fbioe.2016.00065

INTRODUCTION

Diatoms play a critical role in marine ecosystems and global carbon cycling. They are also excellent candidates for bioproduction of valuable commercial compounds and renewable energy resources, as they display rapid growth rates across a range of environmental conditions. However, the development of diatom molecular tools to enhance understanding and enable genetic manipulation

of diatom cellular capabilities is nascent and lags far behind molecular tool development in other prominent model species such as *Escherichia coli* (bacteria) and *Saccharomyces cerevisiae* (yeast). The diatom *Phaeodactylum tricornutum* has emerged as an important model for examining diatom biological processes and commercial potential and as a test strain to develop new genetic tools that can be expanded into other diatom species (Apt et al., 1996; Lopez et al., 2005; Siaut et al., 2007; Bozarth et al., 2009). *P. tricornutum* was the first diatom to be genetically transformed *via* biolistic particle delivery (Apt et al., 1996) and the first to be genetically modified using RNAi-mediated gene silencing (De Riso et al., 2009). Complete gene knockout mutations are also possible using TALEN and CRISPR/CAS9 technology (Daboussi et al., 2014; Weyman et al., 2014; Nymark et al., 2016). Recently, *P. tricornutum* was shown to stably maintain engineered diatom episomes delivered to the diatoms *via* bacterial conjugation, which represents a major advancement in diatom tool development but requires further optimization (Karas et al., 2015).

Diatom episomes potentially allow for consistent and predictable protein expression, perhaps for entire biochemical pathways. Conjugation-based episome delivery may facilitate simplified delivery of genome editing components (e.g., CRISPR/CAS9), allow complementation of genetic mutants in reverse genetics approaches, and provide a platform for overexpressing proteins of interest for functional or subcellular localization studies. Episomes avoid complications associated with random integration of transgenic DNA into diatom nuclear chromosomes, such as multiple or partial expression cassette insertions, interruption of non-target genes, and insertion position expression effects (Dunahay et al., 1995; Falciatore et al., 1999; Miyagawa et al., 2009). Episomes can be introduced into diatoms *via* bacterial conjugation, which is the most efficient method of transgene delivery established to date. Episomes can also be delivered by biolistic particle delivery (confirmed in our lab, data not shown), electroportation, and PEG-mediated transformation (Karas et al., 2015). The multiple methods of episome delivery make them a versatile addition to the diatom molecular toolkit. Our conjugation protocol (Table S1 in Supplementary Material) employs a two-plasmid system in which a mobilizable "cargo plasmid" is transferred to the diatom cell by a conjugative plasmid (pTA-MOB) that itself cannot be mobilized due to deletion of the origin of transfer (*oriT*) (Strand et al., 2014). Diatom episomes must be capable of replication and segregation during division in both the donor bacteria (if conjugation is the preferred delivery method) and in the diatom recipient. In model species of bacteria and yeast, mechanisms of replication are well understood (Murray and Szostak, 1983; Del Solar et al., 1998) and can be varied to accommodate a range of experimental goals. In diatom recipients, the only genetically tested examples of episome maintenance rely on a foreign 1.4-kb DNA sequence derived from yeast, which permits maintenance in both centric and pennate diatoms (Karas et al., 2015). However, very little is known about how the foreign DNA sequence enables maintenance in diatoms.

The DNA sequence *CEN6-ARSH4-HIS3* (CAH), originally derived from a yeast artificial chromosome, consists of three contiguous individual elements, which function in yeast as a centromere (*CEN6*), an origin of replication (*ARSH4*), and a selectable marker (*HIS3*). The *CEN6* sequence, we used in this study, is a 117-bp region of very low GC content (14%); in yeast, this sequence is sufficient for complete centromere function during both meiosis and mitosis (Newlon, 1988; Cottarel et al., 1989). *ARSH4* is a well characterized yeast replication origin enabling initiation of DNA replication (Stinchcomb et al., 1980). Replication origins [also called autonomously replicating sequences (Ars)] are typically found throughout eukaryotic genomes, and the yeast Ars contains a small consensus sequence (Newlon and Theis, 1993; Wyrick et al., 2001; Nieduszynski et al., 2006; Siow et al., 2012; Tagwerker et al., 2012). Only one Ars is required for maintenance of smaller plasmids, while larger plasmids (e.g., greater than 160–300 kb) are more easily assembled and maintained using multiple Ars sequences (Muller et al., 2012; Karas et al., 2013). *ARSH4* is a 388-bp region of low GC content (32%). In addition to the *CEN6* and the *ARSH4*, the final component is a selectable marker that provides a gene to complement yeast histidine (His) auxotrophy, often used for positive selection of successful transformants (Weinstock and Strathern, 1993). The *HIS3* sequence of 872-bp is 45% GC content, which is slightly lower than the average 47% GC content of the *P. tricornutum* genome (Weinstock and Strathern, 1993; Bowler et al., 2008). The CAH region was initially included in diatom episomes to enable yeast-based plasmid assembly methods and was later found to be the element responsible for diatom episome maintenance (Karas et al., 2015). It was unclear, which of these elements individually or in combination were essential for diatom plasmid maintenance, or, given the role of the CAH in supporting stable diatom episome maintenance, whether these elements performed similar functional roles in diatoms as in yeast.

In this study, we explored the sequence requirements for diatom episome maintenance and developed new molecular tools and protocols to facilitate the transfer of episomes from *E. coli* to the diatom *P. tricornutum*. Our objectives were threefold: (1) develop a new diatom episome with medium copy number in *E. coli* for stable and effective genetic engineering, (2) identify elements and characteristics of the CAH sequence that enable diatom episome maintenance, and (3) improve upon the current methods for bacterial transfer of DNA to diatoms. To accomplish these goals, we engineered a cargo vector based on the plasmid pBR322 (Sutcliffe, 1979) that can be efficiently transferred into and maintained in *P. tricornutum*. Using this vector, we identified sub-sequences from CAH that were sufficient for episome maintenance, which permitted new insights into understanding plasmid maintenance requirements and construction of a smaller cargo vector. Finally, we developed modifications of the conjugation protocol, which boost efficiency while saving time and materials and are compatible with high-throughput methods for multiple conjugations.

MATERIALS AND METHODS

Strains and Growth Conditions

The diatom *P. tricornutum* strain CCMP 632 was obtained from the Provasoli–Guillard National Center for Culture of Marine Algae

and Microbiota (NCMA) and was cultured in filter-sterilized (0.2 μm bottle-top filters, Thermo Fisher Scientific, Waltham, MA, USA) L1 artificial seawater medium with Aquil as a base (Price et al., 1989) as described by NCMA (http://ncma.bigelow.org/) (Table S2 in Supplementary Material). Plates for culturing the diatoms on solid medium consisted of 50% full strength L1 medium and 50% melted 2% agar in milli-Q water (autoclave sterilized), both brought to a temperature of 50°C separately prior to mixing. The resulting plates are 1/2 × L1 and 1% final agar concentration (1/2 × L1-agar plates). Cultures were maintained at 18°C with a 60 μmol photons m^{-2} s^{-1} light intensity, and a 14/10-h light/dark cycle. Diatom cultures were xenic, however, bacterial numbers were kept low by regular maintenance with antibiotics. Diatom cell numbers were counted using disposable improved-neubauer haemocytometers (IN-Cyto, Chungnam-do, Korea). Antibiotics for selection of diatom exconjugants were provided as 20 μg mL^{-1} Phleomycin (Phleo) on agar plates or Zeocin (50 μg mL^{-1}) in liquid cultures.

Escherichia coli TransforMax EPI300 cells (Epicenter, Madison, WI, USA) were used for plasmid cloning and conjugative transfer of plasmids to diatoms (Table S2 in Supplementary Material). Cultures were grown in LB broth (Amresco, Solon, OH, USA) at 37°C in a shaking (225 rpm) incubator. Antibiotics as needed were provided as 20 μg mL^{-1} gentamicin (Gm) in water solvent, 100 μg mL^{-1} ampicillin (Amp), 50 μg mL^{-1} kanamycin (Kan), 5 μg mL^{-1} tetracycline (Tet) in ethanol solvent, 10 μg mL^{-1} chloramphenicol (Cm) in ethanol solvent.

Plasmid Design and Construction

All plasmids were constructed by Gibson Assembly of amplified PCR products (Gibson et al., 2009). Templates and primers used for all plasmid assemblies are listed in Tables S3,S4 in Supplementary Material. PCR products were amplified using PrimeStar Max DNA polymerase mastermix (Takara Clontech, Mountain View, CA, USA) using the manufacturer's recommended protocol and were confirmed by agarose gel electrophoresis. PCR products were purified using Qiagen PCR purification kits (Qiagen, Hilden, Germany) and quantified prior to Gibson assembly using a Nanodrop Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). Assembled plasmids were transformed into *E. coli* by electroporation using 1 mm cuvettes and the preset bacterial transformation protocol (Bio-Rad Laboratories, Irvine, CA, USA). Individual colonies were isolated by selective plating on LB + Amp + Tet. PCR screening of colonies was performed using the primers ptrepinsertscrn1 and ptrepinsertscrn2 (Table S3 in Supplementary Material) followed by plasmid purification using the Qiagen miniprep kit (Qiagen, Hilden, Germany) and various diagnostic restriction enzyme digests (New England Biolabs, Ipswich, MA, USA) to verify plasmid construction.

We used the plasmid backbone from pBR322 for all plasmids tested in this study. Plasmid pBR322 is derived from a natural variant of the ColE1 plasmid called MB1 (Bolivar et al., 1977) and contains an *oriT* [originally called basis of mobilization (*bom*)] that is compatible with RP4-mediated conjugation systems (Finnegan and Sherratt, 1982) (Table S2 in Supplementary Material). We replaced the pBR322 *oriT* with the *oriT* from RP4 to ensure maximum transfer efficiency, as this *oriT* has been shown to enable conjugation into diatoms previously (Karas et al., 2015). Two base plasmids were constructed to serve as positive and negative controls for further experiments: pPtPBR1, which contains the CAH sequence (**Figure 1A**), and pPtPBR2, which lacks CAH (**Figure 1B**). To construct pPtPBR1, the pBRR322 backbone lacking the pBR322 *oriT* was combined with an insert containing the RP4 *oriT* and the ShBle cassette amplified from pPtPuc3-Km using primers Insert-F and Insert-R (Table S3 in Supplementary Material). To construct pPtPBR2, the pBRR322 backbone lacking the pBR322 *oriT* was combined with an insert containing the RP4 *oriT* and the ShBle cassette using primers Insert-F and Insert-R (Table S3 in Supplementary Material).

The plasmids pPtPBR3 through pPtPBR12 were constructed by two-piece Gibson assembly of a vector PCR product and an insert PCR product. Plasmid sources for both vector and insert products, as well as primers used, are listed in Table S4 in Supplementary Material. Plasmids pPtPBR3 (*CEN6* only), pPtPBR4 (*ARSH4* only), pPtPBR5 (*HIS3* only), and pPtPBR6 (*CEN6-ARSH4*) were constructed using previously assembled and verified pUC-based versions already containing the specified portions of the CAH region as templates (Table S4 in Supplementary Material). Inserts containing the RP4 *oriT*, the ShBle cassette, and the specified portion of the CAH sequence were then amplified by PCR using the Insert-F and Insert-R primers. These inserts were combined with the pBR322 backbone vector with the *oriT* site omitted as described in the construction of pPtPBR1 and pPtPBR2 above, using primers pBR322-F and pBR322-R.

The remaining plasmids were constructed using plasmids constructed above as templates (Table S4 in Supplementary Material). Plasmid pPtPBR7 has both *CEN6* and ARS4 elements, but, in contrast to pPtPBR6, has these elements located on different parts of the plasmid. The plasmid pPtPBR7 was constructed using pPtPBR3, containing only *CEN6*, as a backbone template and pPtPBR1 as an insert template to amplify the ARS4, which was inserted between the tetracycline resistance cassette (TetR) and the *Rop* gene (**Figure 1B**). A similar procedure was used to construct pTPBR8, which contains two sets of *CEN6*-ARS4 regions on separate sides of the plasmid in the same location as the pPtPBR7 elements (**Figure 1B**). For this, pPtPBR6 was used for both a template and insert (Table S4 in Supplementary Material). Plasmids pPtPBR9 through pPtPBR12 were constructed by amplifying the pPtPBR1 backbone using Gibson assembly primers, omitting particular regions. Plasmid pPtPBR9 contains only the ARS4 and *HIS3* sequences. Plasmid pPtPBR10, 11, and 12 contain the *CEN6* and ARS4 elements along with portions of the beginning of the *HIS3* sequence (100, 200, and 300 bp, respectively).

Plasmid Sequencing and Deposit

The plasmids pPtPBR1, pPtPBR2, and pPtPBR11 constructed in this study were fully sequenced (Eurofins Genomics, Louisville, KY, USA). Primers for all sequencing and PCR amplifications were obtained from Integrated DNA Technologies (Coralville, IA, USA) and are listed in Table S3 in Supplementary Material. Sequences for the plasmids are available in Table S7 in Supplementary Material and sequences as well as annotations are deposited in NCBI Genbank with the following accession numbers: KX523201

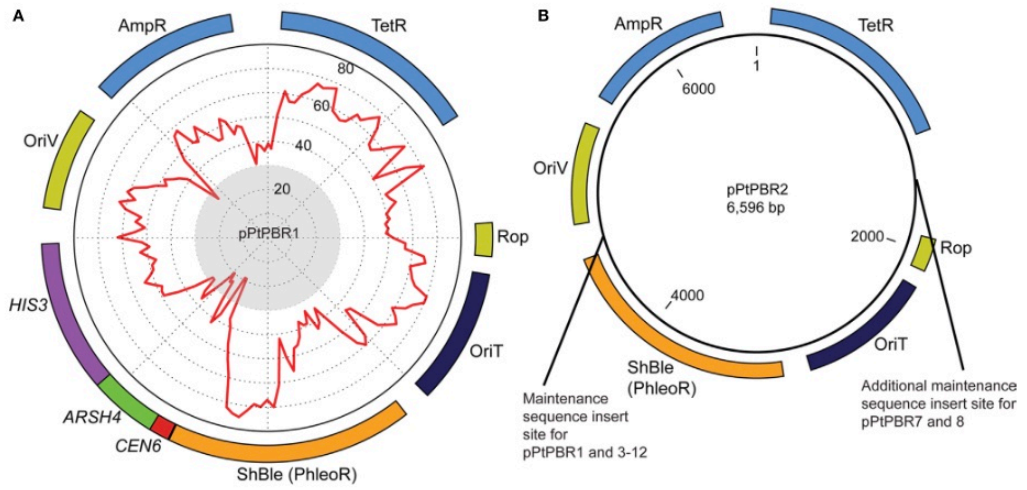

FIGURE 1 | Plasmid maps of (A) pPtPBR1 and (B) pPtPBR2, which both contain Ampicillin (AmpR) and Tetracyline (TetR) resistance cassettes for *E. coli* selection, a diatom ShBle cassette for diatom selection on phleomycin (PhleoR), an origin of transfer enabling conjugative transfer of episomes from bacteria into diatoms (derived from the plasmid pRL2948a), and an origin of replication (OriV) and the rop gene, which reduces plasmid copy number. The plasmid pPtPBR1 **(A)** contains the *CEN6-ARSH4-HIS3* episome maintenance sequence. Percent GC content for each region of the plasmid is graphically displayed as a red line on a graph inside the plasmid and was constructed using a 100-bp window advancing every 50-bp. Regions with a GC content of less than 30% fall within the gray central circle of the graphic. The pPtPBR2 plasmid **(B)** contains no CAH region. The pPtPBR3-12 plasmids share a common backbone with pPtPBR2 with additional modifications indicated in the **(B)**.

(pPtPBR1), KX523202 (pPtPBR2), and KX523203 (pPtPBR11). Plasmids were also deposited with AddGene.

Conjugation of Episomes into Diatoms and Plasmid Rescue

We used the conjugation protocol described in Karas et al. (2015) and also tested various plating and culturing modifications in an effort to improve the conjugation protocol for various applications. A detailed protocol of the original method and modifications presented in this study are presented in Table S1 in Supplementary Material. To examine the sequence elements required to produce an optimal number of exconjugants, conjugations were conducted in triplicate using the previously established protocol (Karas et al., 2015). Pre-plated diatom cultures grown on 1/2 × L1-agar plates were resuspended in L1 medium and adjusted to a final cell concentration of about 5.5×10^8 cells mL^{-1}. Overnight *E. coli* cultures containing the RP4-derived conjugative plasmid pTA-MOB (Strand et al., 2014), which contains no *oriT*, and the desired cargo plasmid/episome for delivery into the diatoms were diluted 1:50 into fresh medium containing appropriate antibiotics and grown to an OD_{600} of 0.8–1.0. Cultures (50 mL) were then centrifuged for 10 min at 3500*g*, supernatant was removed, and the *E. coli* cells were resuspended in 600 µL SOC medium. 200 µL each of *E. coli* and *P. tricornutum* concentrated culture were mixed, plated on 1/2 × L1-agar plates containing 5% LB medium (made with an L1 seawater base) and incubated at 30°C in the dark for 90 min. Plates were then moved to typical diatom culturing conditions and allowed to recover for 2 days, prior to resuspending cells in 1.5 mL of L1 medium and plating a portion of the reaction on selective 1/2 × L1-agar plates supplemented with Phleo for exconjugant selection. To confirm the presence of episomes, DNA was extracted from the exconjugant diatoms using the protocol described in Karas et al. (2015) and then transformed into *E. coli* using electroporation to effectively "rescue" replicating plasmids (Karas et al., 2015). If DNA was maintained as an episome in the diatom and later transformed into *E. coli*, it would subsequently replicate in *E. coli* as well since the necessary elements for maintenance are present on the plasmid. Alternatively, if selection in diatom exconjugants is not due to maintanence of an episome, *E. coli* colonies will not be able to be recovered after transformation of exconjugant DNA. To ensure absence of *E. coli* carry-over contamination, colonies were passaged multiple times on 1/2 × L1-agar plates and were also patched onto LB agar plates and incubated overnight at 37°C; no *E. coli* growth was observed.

For optimization experiments, various modifications were made to the conjugation protocol. The cargo plasmid pPtPBR11 was used in optimization experiments, as this plasmid had high exconjugant yield and a smaller size than pPtPBR1 (see Section "Results" and "Discussion"). For development of a high-throughput option, 12-well Costar (Corning, Corning, NY, USA) plates were utilized, either in lieu of or in conjunction with standard

100 mm Petri dishes (Denville Scientific, South Plainfield, NJ, USA). We first scaled down the Karas et al. (2015) protocol reaction volume by roughly fivefold. Melted 1/2 × L1-agar-5% LB (~3.5 mL) was pipetted into each well of a 12-well cell culture plate and allowed polymerize (Table S1 in Supplementary Material). Liquid *P. tricornutum* cultures were harvested by centrifugation for 10 min at 3,000*g* (room temperature) and adjusted to 5×10^8 cells mL^{-1}. 50 µL of the cell suspension was pipetted into the center of each agar "well" and lightly spread around with a loop or by rotating the plate to cover an area covering about 80% of the agar, leaving a gap between the cell suspension and the well edges. Plates were then dried in a laminar flow hood until no visible liquid remained and, subsequently, incubated for 96 h in the standard diatom culturing conditions (see above). On the day of conjugation, an overnight culture of *E. coli*-containing plasmids pPtPBR11 and pTA-MOB was diluted 1:50 in fresh liquid LB medium with antibiotics (10 µg mL^{-1} Tet, 100 µg mL^{-1} Amp, and 20 µg mL^{-1} Gent) incubated at 37°C until reaching an OD_{600} of 0.8. We, then, harvested bacterial cells by centrifugation (10 min at 3,000*g*, room temperature) and concentrated 100-fold by resuspension in SOC medium. 50 µL of either SOC alone (negative conjugation control) or bacterial suspension was added on top of the dried algal culture and plates rotated by hand to allow the bacteria to cover the algal area. The plates were dried in a laminar flow hood for ~10 min until no visible liquid remained and were covered and transferred to 30°C for 90 min (dark). They were then returned to typical diatom culturing conditions. After a 48-h recovery period, the cells in each well resuspended in 500 µL L1 liquid medium, and the entire reaction was replated on standard 100 mm Petri dishes containing 30 mL of 1/2 × L1-agar supplemented with 20 µg mL^{-1} Phleo. These plates were allowed to dry briefly in a laminar flow hood and then incubated for ~7–10 days in diatom culturing conditions until colonies were large enough to count.

To address the effect of pre-conjugation diatom plating conditions, we tested various reduced time intervals between plating and conjugation. Two 12-well 1/2 × L1-agar-5% LB agar plates were prepared as described above and plated with increasing number of cells on 4 successive days before conjugation, assuming a doubling time of ~1 doubling per day under these growth conditions. Therefore, on t = Day 4 (with t = Day 0 being the day of conjugation) cells were harvested from liquid culture and adjusted to 1×10^8 cells mL^{-1} in L1 medium, and a 50 µL aliquot of this suspension (5×10^6 cells) was plated onto each of 6 wells (3 for pPtPBR11 and 3 SOC negative controls). On t = Day 3, a total of 1×10^7 cells were plated onto each of an additional 6 wells, followed by 2×10^7 cells on t = Day 2 and 4×10^7 cells on t = Day 1. Plate cultures were grown under standard diatom growth conditions. On t = Day 0, the overnight *E. coli* pPtPBR11/pTA-MOB culture was diluted (1:50) in fresh LB and grown to 0.8 OD_{600} with shaking at 37°C. We concentrated 25 mL by centrifugation (as above) and resuspended in 250 µL SOC medium. 50 µL of this suspension was added to each of the experimental treatment wells containing diatom cultures, and 50 µL of sterile SOC medium was added to the remaining wells (negative controls). After drying, these conjugations were carried out as described above with a 48-h recovery period followed by replating onto 100 mm 1/2 × L1-agar plates supplemented with 20 µg mL^{-1} Phleo. Diatom colonies were counted after ~10 days.

We also tested recovery time required before plating diatom conjugation reactions onto selective medium. For this experiment, 12 wells of 1/2 × L1-agar-5% LB were plated with 4×10^7 cells on the day before conjugation and grown overnight in diatom culturing conditions. On the day of the conjugation, *E coli* pPtPBR11/pTA-MOB was cultured and prepared as described above. 50 µL of SOC was added to 3 wells, while 50 µL of bacterial suspension was added on top of the other 9 wells. The conjugation plate was incubated at 30°C for 90 min, after which 4 wells (1 SOC control, 3 pPtPBR11/pTA-MOB) were immediately resuspended into 500 µL L1 liquid medium, while the plate with the remaining reactions was returned to diatom growth conditions for recovery. The four cell suspensions were then plated on 100 mm 1/2 × L1-agar plates supplemented with 20 µg mL^{-1} Phleo and incubated at 18°C after drying (t = Day 0 treatment). The next day (24 h after the incubation at 30°C for 90 min), 4 more wells (1 SOC control, 3 pPtPBR11/pTA-MOB) were resuspended and replated on selective medium (t = Day + 1), and the final four wells resuspended and replated on t = Day + 2 (48 h after the incubation at 30°C for 90 min). The final day 2 plating was similar to the optimized protocol previously described and thus served as a positive control (Karas et al., 2015). Selection plates were incubated for 7–10 days in diatom culturing conditions, after which colonies were counted.

Plasmid Maintenance

Three plasmids (pPtPBR1, 6, and 8) were tested for maintenance following the protocol in Karas et al. (2015). Following conjugation and selection on Phleo 1/2 × L1-agar plates, 2 colonies of pPtPBR1, 2 colonies of pPtPB8, and 1 colony of pPtPBR6 were transferred to liquid L1 medium containing no antibiotics. Cultures were transferred to fresh medium weekly for 30 days and then plated on non-selective 1/2 × L1-agar plates to obtain single colonies (Figure S1A in Supplementary Material). Between 45 and 100 colonies from each treatment were patched onto non-selective 1/2 × L1-agar plates, and, after about 1 week of growth, each patch was repatched onto both non-selective medium as a positive control and selective medium to determine how many retained the transgene or, minimally, the antibiotic resistance cassette (Figures S1A,B in Supplementary Material).

RESULTS

Sequence Elements Enabling Efficient Plasmid Maintenance in *P. tricornutum*

We aimed to determine which combinations of sub-sequences within the yeast *CEN6-ARSH4-HIS3* were sufficient to allow episome maintenance. We found that the newly engineered plasmid pPtPBR1, containing the full CAH sequence, resulted in greater than 150-fold more diatom exconjugants than the negative control plasmid pPtPBR2 lacking the CAH sequence (**Figure 2A**). Plasmids from pPtPBR1 were effectively rescued in *E. coli* (Table S5 in Supplementary Material), confirming their successful maintenance as episomes in diatoms. Furthermore, these plasmids were maintained in non-selective medium at

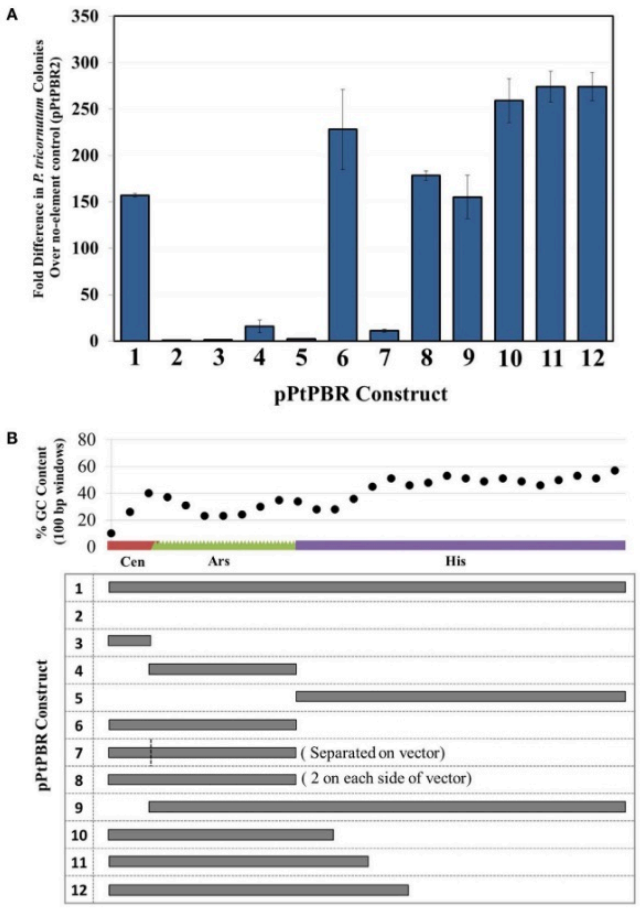

FIGURE 2 | **(A)** Diatom colony yields, presented as fold difference from the negative control plasmid pPtPBR2 (no maintenance elements), for plasmids containing various potential maintenance sequences. *N* = 3, error bars = SD, **(B)** relative size and GC content analysis of possible maintenance sequences contained in plasmids pPtPBR1-12.

levels similar to those previously reported in yeast and diatoms (Karas et al., 2015) (Figure S1B in Supplementary Material). Conjugation with pPtPBR2 yielded a very small number of diatom colonies that could not be recovered *via E. coli* rescue. Thus, pPtPBR1 and pPtPBR2 function similarly to the previously described plasmids containing or lacking the CAH sequence, respectively (Karas et al., 2015). In this study and in the previous study, we observed that conjugation of an episome containing the CAH sequence resulted in high diatom colony numbers. Additionally, these episomes were shown to be maintained in diatoms and could be rescued in *E. coli*. Thus, high diatom colony numbers resulting from a conjugation relative to the pPtPBR2 control suggest the successful transfer and maintenance of an episome.

When the individual CAH sequence elements were tested separately for episome maintenance ability, conjugation efficiency was very low. The numbers of diatom colonies emerging after conjugation with *E. coli* strains bearing episomes containing only *CEN6* or *HIS3* were similar to the no element control pPtPBR2 (**Figure 2A**). Episomes containing only the *ARSH4* sequence for maintenance resulted in a slightly higher number of diatom colonies, ~16-fold higher than the no element control, though 67-fold less than the pPtPBR1 positive control. When plasmid DNA was extracted from some of the few *P. tricornutum* lines obtained

after conjugation with for *E. coli* bearing *CEN6*- or *HIS3*-only containing plasmids, no *E. coli* colonies were obtained, suggesting the absence of episomes in the diatoms and chromosomal integration of the marker. Episome rescue with DNA extracted from episomes containing only *ARSH4* resulted in a small number of *E. coli* colonies (Table S5 in Supplementary Material). Plasmids that contained the *CEN6-ARSH4* region without the *HIS3* (pPtPBR6) and the *ARSH4-HIS3* region without the *CEN6* (pPtPBR9) yielded a large number of *P. tricornutum* colonies, similar in number to the control plasmid pPtPBR1. One colony of pPtPBR6 was tested for maintenance during passage without selection for 30 divisions and was found to be maintained at levels similar to pPtPBR1 and to those previously reported (Karas et al., 2015) (Figure S1B in Supplementary Material). Plasmid pPtPBR7, which contained *CEN6* and *ARSH4* that were spatially separated by 3 kb (**Figure 1B**) resulted in a similar number of exconjugant *P. tricornutum* colonies as the pPtPBR4, containing the *ARSH4* only. Plasmid pPtPBR8, which contained two *CEN6-ARSH4* regions on different parts of the episome yielded a large a number of colonies similar to the pPtPBR1 positive control (**Figure 2A**) and the pPtPBR6 plasmid containing one copy of the *CEN6-ARSH4* region. Of the two pPtPBR8 colonies tested for plasmid maintenance during serial passage in the absence of selection, one was maintained stably while the other was maintained poorly compared to prior results (Figure 1B in Supplementary Material).

The 3′ region of the *HIS3* element contained a relatively high GC content compared to the first 200-bp of the 5′ region of *HIS3* adjacent to the *CEN6-ARSH4* sequence, which were low in GC content (**Figure 2B**). To examine whether this distinction plays a role in diatom episome maintenance, we tested variations of the CAH sequence with modified *HIS3* sequences. The plasmids pPtPBR10–12 contained the *CEN6-ARSH4* regions as well as some of the low GC content sequence adjacent to the *ARSH4* sequence but lacked the relatively high GC content portion of the *HIS3* sequence (**Figure 2B**). Each of these constructs produced similar numbers of diatom colonies to each other and more than the pPtPBR1 positive control containing the full CAH sequence (**Figure 1B**).

Efficiency of Modified Conjugation Protocols

Culturing diatoms on 1/2 × L1-agar supplemented with 5% LB medium led to faster diatom growth and a higher number of exconjugant colonies after adjusting for diatom cell number (Figure S2 in Supplementary Material). After 4 days of growth on either 1/2 × L1-agar only or 1/2 × L1-agar-5% LB plates, diatom cell numbers were more than 2 times higher on LB supplemented plates: 1.59×10^8 cells per plate on 1/2 × L1-agar and 3.59×10^8 cells/plate when supplemented with 5% LB (Figure S2 in Supplementary Material). After resuspending diatom cells prior to conjugation, cell numbers were adjusted for each culture to 5×10^8 cells mL^{-1}. Following the standard conjugation protocol, after plating on selective medium, there were more than two times as many exconjugant colonies in the treatments where diatoms were supplemented with 5% LB agar (Figure S2 in Supplementary Material; Table S6 in SupplementaryMaterial). This experiment was repeated independently, also in triplicate, and similar patterns were observed (Table S6 in Supplementary Material).

We developed a high-throughput method of conjugation based on plating diatoms on 5% LB medium in 12-well plates, followed by the direct addition of *E. coli* culture on top of the diatoms (**Figure 3**). Using the standard timing for diatom pre-plating and post-conjugation recovery (plated 4 days prior to conjugations and plated on selective medium 2 days after the 90-min incubation at 30°C), and after replating each entire conjugation reaction (for each of three replicates) on one 100 mm 1/2 × L1-agar plate containing Phleo 20 μg mL^{-1}, an average of 146 diatom colonies were obtained per plate (Table S6 in Supplementary Material).

When testing recovery time after conjugation using the high-throughput protocol, plating cells immediately after the 30°C step (t = Day 0) resulted in only 5 colonies total (across 3 plates), whereas incubating conjugation reactions for 24 h resulted an average of 79 diatom colonies (Table S6 in Supplementary Material). Allowing for the full 48-h incubation before plating on selective medium resulted in a recovery of 3 times as many colonies, an average of 242 colonies, as the 24-h recovery time (Table S6 in Supplementary Material). There was little difference in the number of diatom exconjugants when the diatoms were plated at t = Day 4, 3, 2, or 1 prior to the conjugation (Table S6 in Supplementary Material). Furthermore, the number of exconjugant colonies was not affected when the recovery step was conducted on 1/2 × L1-agar plates with the antibiotics kanamycin or chloramphenicol in addition to phleomycin to ensure the *E. coli* culture was killed (Table S6 in Supplementary Material).

DISCUSSION

The pPtPBR Plasmids: A New Series of Diatom Conjugation Vectors with Medium Copy Number in *E. coli*

For a robust genetic model system, a variety of plasmid vectors are often developed and optimized depending on the ultimate use, including efficient cloning of the gene(s) of interest in *E. coli*. Important factors to consider include cloning insert and vector size, ease of cloning, restriction sites, and plasmid copy number. Vectors based on bacterial artificial chromosomes (BACs), such as the p0521s plasmid used in Karas et al. (2015), contain a BAC backbone that can maintain hundreds of kilobases of sequence in *E. coli* at low copy. While these plasmids can be used to clone large DNA fragments, they are inconvenient to work with due the large vector size and low copy number. Smaller plasmids such as pPtPuc3 (a derivative of the pUC19 plasmid) are easier to clone and can be advantageous when over-expressing proteins in *E. coli*. However, diatom genes of interest may be toxic to donor *E. coli* cells when highly expressed, and the metabolic energy required to maintain many plasmids can cause growth impairments compared to lower copy alternatives (Jones et al., 2000). Furthermore, high-copy number plasmids may become unstable leading to unwanted sequence modifications in *E. coli* (Green and Sambrook, 2012). To address some of these vector issues, the plasmid pBR322 was developed to be a medium copy number alternative and has since undergone several improvements, becoming a widely

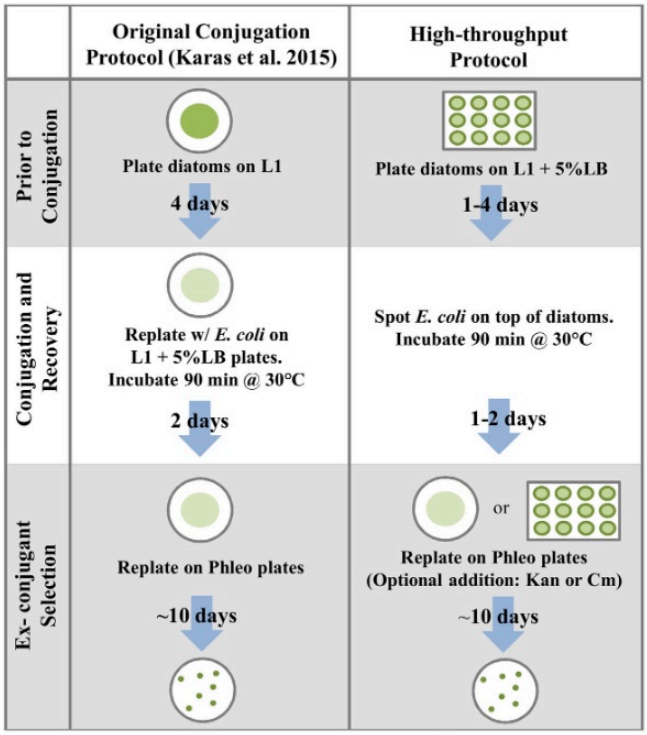

FIGURE 3 | Comparison of the protocol established in Karas et al. (2015) to the high-throughput conjugation protocol developed in this study.

popular plasmid vector (Sutcliffe, 1979; Watson, 1988). Plasmid pBR322 shares the MB1 origin of replication with pUC19 but also contains the *Rop* gene, which regulates plasmid copy number to lower levels (Sutcliffe, 1979).

We successfully modified this plasmid to create the pPtPBR plasmids, derivatives capable of transfer to *P. tricornutum via* bacterial conjugation. We constructed the derivative plasmid pPtPBR1 (**Figure 1A**), which we showed can be stably maintained as an episome in *P. tricornutum* (Figure S1B in Supplementary Material). While dissecting the yeast-derived sequence required for episome maintenance in *P. tricornutum*, we found that pPtPBR derivatives containing the *CEN6*, *ARSH4*, and a truncated version of the *HIS3* sequence encompassing the low GC portion in a contiguous sequence resulted in a higher number of exconjugant colonies than versions containing the entire *HIS3* sequence (**Figure 2**). As a result, we developed episomes that are smaller, potentially easier to assemble, and more efficient in terms of exconjugant yields. We also constructed pPtPBR2 (**Figure 1B**), which does not have the ability to replicate as an episome and can be used as a negative control in conjugation experiments.

One major obstacle to genetic engineering of algae *via* biolistics is random genomic integration of transgenes and markers, leading to differences in gene expression and phenotypes between clones obtained from the same experiment. In addition, mapping of the integration sites is time-consuming, with possible epigenetic effects on transcription adding a layer of complexity. With the diatom episomes described herein, a mechanism is provided for the introduction of expression cassettes and markers at ploidy equivalent to native chromosome levels, while reducing the possible effects of gene disruption and deletion upon random genomic integration. Some potential applications include more meaningful and reproducible measurements of promoter and terminator strengths and greater consistency in controlling inducible promoters. Additionally, the episome allows for what could be a stable platform for the delivery of heterologous genes and pathways, as at least 50 kb of DNA can be maintained within the episome (Karas et al., 2015). Phenotypes resulting from introduction of heterologous genes into *P. tricornutum* could furthermore be calibrated more efficiently when screened across a large number of clones. Finally, the episome and the methods

developed in this study could enable large-scale screening of genomic libraries in forward genetics approaches.

Yeast Centromere and Origin of Replication Sequences Do Not Function Orthologously in *P. tricornutum*

This study begins to investigate the role that yeast artificial chromosome maintenance sequences play in diatom episome maintenance. The identity and characteristics of elements required for native diatom chromosomal replication and maintenance are unknown, but our results here suggest that these requirements differ from yeast episome maintenance. In yeast, both a centromere sequence and a sequence functioning as an origin of replication are necessary for a plasmid to be stably maintained (Murray and Szostak, 1983). In early work developing yeast artificial chromosomes, some initial cloned sequences of yeast centromeres could act alone to maintain plasmids because the sequence also contained a weak replication origin, but both functions are still required. Multiple centromeres lead to unstable dicentric plasmids, while multiple replication origins tend to enhance maintenance especially for large, cloned, and high GC sequences (Karas et al., 2013). We found that the yeast centromere and yeast replication origin sequences, when cloned individually, do not enable efficient diatom episome maintenance (**Figure 2A**). When the yeast centromere and yeast replication origin sequences were cloned on the same plasmid, but spatially separated, they did not allow episome maintenance in *P. tricornutum* (**Figure 2A**). Such a situation in yeast should lead to a stable plasmid; therefore, we conclude that these sequences do not function the same in *P. tricornutum* and yeast, and that the proximity of these yeast sequences is important for achieving the high levels of diatoms colonies after conjugation resulting from episomal maintenance.

It is possible that the *ARSH4* element alone enables episome maintenance in *P. tricornutum*, as we were able to rescue a small number of *E. coli* colonies after transformation of *P. tricornutum*-extracted DNA. One possible explanation for this is that the episomes containing *ARSH4* only are inefficiently maintained and slow-growing, and there is strong selection for chromosomal integration of the selectable marker at some point during colony formation. Thus, pPtPBR4 exconjugant *P. tricornutum* colonies may be composed of a mixture of chromosomal integrants and inefficiently maintained episomes. This could explain the smaller number of *E. coli* colonies rescued using DNA extracted from these colonies. Alternatively, it is possible that rearrangements of the plasmid (e.g., insertions of genomic DNA, recombination) may be responsible for our observation; however, this is unlikely given that pPtPBR2, 3, and 5 *P. tricornutum* exconjugant lines did not yield a single *E. coli* colony during attempts at episome rescue. Our experiments did not rule out the possibility that the *ARSH4* sequence alone could still contribute to diatom episome maintenance, whether as an origin of replication or by another mechanism. Regardless, while the mechanism of inefficient episomal maintenance by *ARSH4*-only plasmids is still unclear, episomes containing only *ARSH4* are not practical, as substantially fewer exconjugant lines are produced compared to the CAH elements combined.

Conjugative transfer of a plasmid containing two sets of adjacent *CEN6-ARSH4* sequences on different parts of the plasmid resulted in a similar number of diatom exconjugant colonies as the pPtPBR1 plasmid containing the entire CAH sequence. In yeast episomes, two replication origin sequences are not problematic and are in fact required for maintenance of large plasmids. However, two centromeric regions on a single centromeric plasmid creates an unstable dicentric plasmid (Mann and Davis, 1983). Here, we observe that two copies of a yeast *CEN6* sequence on different parts of the plasmid is at least as efficient in terms of diatom colony yield as only one copy of the *CEN6*. This lends further evidence to the possibility that the sequences are functioning differently in diatoms than in yeast.

Low GC Content Appears to Drive Episome Maintenance in Diatoms

An emerging pattern while testing the efficiency of pPtPBR1 through pPtPBR9 was that sequences of low GC DNA longer than 100 bp were a common feature of episomes that could be maintained in *P. tricornutum*. For example, while *CEN6*, *ARSH4*, and *HIS3* were each individually inefficient in establishing episomes, combinations of *CEN6-ARSH4* (513-bp) and *ARSH4-HIS3* (588-bp) allowed for episome maintenance and maximal conjugation efficiency. One hypothesis stemming from this result is that a critical length of low GC sequence for episome maintenance is at least 500 bp since the 388-bp *ARSH4* sequence was not sufficient to establish robust episomes. The hypothesized GC content threshold required for episome maintenance is unknown, but the fact that the *CEN6*, *ARSH4*, and 5′ region of *HIS3* are all under ~32% GC (**Figures 1A** and **2A**) may point to this level as critical. While GC content alone may define maintenance function in a sequence, alternatively, there may be specific sequence motifs or patterns responsible for episome maintenance that happen to occur more frequently in low GC content sequences. We could not, however, identify any such patterns in our analyses of the sequences.

Phaeodactylum tricornutum is capable of producing histidine, and there is no predicted functional role for the *HIS3* gene in diatoms. We observed that the 3′ region of the *HIS3* sequence contained a relatively high GC content compared to the first 200 bp adjacent to the *ARSH4* sequence, which was low in GC content (**Figure 2B**). To strengthen our hypothesis that low GC content plays some role in episome maintenance, we examined variations of the CAH sequence with modified *HIS3* sequences. Dissecting the *HIS3* sequence revealed that only a portion of this sequence was necessary for episome replication. The plasmids pPtPBR10-12 contained the *CEN6-ARSH4* regions as well as some of the low GC content sequence adjacent to the *ARSH4* sequence but lacked the relatively high GC content portion of the *HIS3* sequence. These three constructs resulted in a high number of exconjugant colonies, and, in fact, yielded more than pPtPBR1 containing the full CAH sequence. It is unclear why the removal of the 3′, high GC region of the *HIS3* sequence would increase conjugation efficiency. Possible explanations include: (1) the high-GC content region of the *HIS3* sequence is superfluous, and a smaller episome is more effectively transferred and maintained,

(2) the high-GC region of *HIS3* directly decreases conjugation efficiency, regardless of plasmid size, or (3) the position of the CAH sequence in the episome in relation to adjacent sequences is altered, which may affect efficiency.

Low GC content has previously been observed to play an important role in episome and chromosome maintenance in eukaryotic organisms. Centromeres of the protist parasite *Plasmodium*, the causative agent of malaria, have been identified and used to construct artificial *Plasmodium* chromosomes (Bowman et al., 1999; Iwanaga et al., 2010, 2012). These centromere regions consist of 2.3–3.5 kb of extremely low GC content DNA (less than 2% GC), which is considerably lower than the average for the native nuclear chromosomes (Iwanaga et al., 2010, 2012). There was no clear origin of replication identified for the *Plasmodium* artificial chromosomes, and the authors hypothesized that the centromere origin functioned as both the centromere and replication origin. Although the sequences required for *P. tricornutum* episome maintenance were smaller (more than ~500 bp) than the *Plasmodium* sequences, they were similarly lower in GC than the nuclear chromosome average; diatom episome maintenance sequences were ~28–32% GC relative to 47% for *P. tricornutum* nuclear chromosomes (Bowler et al., 2008; Karas et al., 2013). It is possible that the mechanism is similar in that the same sequence can serve both as an origin of replication and as a centromere. Low GC was also found to be a defining feature of centromeres in the red alga *Cyanidioschyzon merolae* (Maruyama et al., 2008; Kanesaki et al., 2015). In this organism, each of the 20 chromosomes was found to have a distinct low GC region that functioned as a centromere and recruited the centromeric histone CENP-A (Kanesaki et al., 2015). These studies support the possibility that low GC content may play an important role in episome maintenance and, perhaps, in diatom DNA replication in general.

Modified Conjugation Protocols Reduce Time and Materials Required, and Increase Exconjugant Yield

Our goals in testing variations of the conjugation protocol were twofold: (1) to improve exconjugant yield and (2) to eliminate the unnecessary expense of time and materials. A major benefit of using conjugative DNA transfer for diatom genetic manipulation is that the method is widely accessible and does not require expensive equipment such as a biolistic particle delivery system. We sought to make it even more accessible and effective and to provide a method to perform the conjugations in high-throughput format. We found that inclusion of 5% LB medium in the 1/2 × L1-agar plates used to grow the diatoms before conjugation improved exconjugant yield compared to growth on unsupplemented 1/2 × L1-agar plates after normalization for equal diatom cell counts. Additionally, we were able to reduce both time and material required to obtain exconjugants. A multi-well plate-based high-throughput method, where there is no resuspension of diatoms prior to conjugation and recovery, led to a reduced usage of materials and time transferring cultures while still leading to a high number of exconjugants. We identified optimal and sufficient plating protocols for the diatoms both before conjugation and after selection, as well as *E. coli* OD_{600} measurements. These data will allow users of the conjugation protocol to optimize conjugation procedures for their experiments based on the time and material resources available to them and the number of required diatom exconjugant colonies.

Culturing diatoms on 1/2 × L1-agar plates supplemented with 5% LB medium led to faster diatom growth and also a higher number of exconjugants after adjusting for cell number (Figure S2 in Supplementary Material). We hypothesize that the improved growth and conjugation efficiency observed are a result of the LB supplement to diatom growth, which may have led to changes in diatom cell physiology. Though *P. tricornutum* is generally thought to be autotrophic, there have been studies suggesting that these diatoms may display some heterotrophic tendencies (Cerón García et al., 2005; Hayward, 2009; Ukeles and Rose, 1976). The possibility of growing diatoms directly on 5% LB plates prior to conjugation eliminates the need to resuspend and replate diatoms prior to the conjugation procedure, which could possibly cause unnecessary stress to the cells and avoids an additional plating step that requires time and materials. Based on our results, it may also increase conjugation efficiency.

Exconjugant colonies were successfully obtained using diatoms plated the day before conjugation. The finding in Karas et al. (2015) that liquid diatom cultures could also be used prior to conjugation but with lower efficiency suggests that plating diatoms prior to conjugation does provide a benefit in terms of total colony number. Here, we show that the number of days prior to conjugation that diatoms are plated is not important, at least when plating on 1/2 × L1 plates containing 5% LB, meaning the entire protocol can be completed faster and with equal efficiency. Recovery time prior to plating exconjugants on selection was an important factor in exconjugant colony yield but could still be shortened compared to the previously published protocol (Karas et al., 2015). Recovery for 2 days prior to selective plating resulted in the highest number of colonies; however, the 50–100 diatom colonies resulting from a recovery period of only 1 day would be more than ample for many applications (e.g., a protein expression or promoter validation assay). When conjugating libraries of sequence variants shorter recovery time may increase library diversity, as many of the colonies on the t = Day 2 recovery plate could be clones of earlier conjugants and not novel and unique colonies.

CONCLUSION

We tested the individual functional elements of the yeast-derived *CEN6-ARSH4-HIS3* to identify the region of the sequence that allowed it to confer episomal maintenance in diatoms. We found that low GC fragments of the *CEN6-ARSH4-HIS3* sequence that were contiguous and greater than 500-bp were required to support robust episome maintenance. We also further optimized the conjugation processes for *P. tricornutum* and developed a higher throughput small-scale protocol. Culturing diatoms on 1/2 × L1-agar supplemented with 5% LB medium led to faster diatom growth and a higher number of exconjugant colonies after adjusting for diatom cell number. Small-scale conjugations in 12-well plates could be performed to reduce materials

required for conjugation when large numbers of colonies are not required. By refining conjugation methodology and elucidating additional features of episome maintenance sequences, this study contributes toward future advances in diatom molecular biology.

AUTHOR CONTRIBUTIONS

RD and PW designed research; RD, VB, and PW performed research; RD, VB, and PW analyzed data; and RD, VB, CD, AA, and PW wrote the paper.

ACKNOWLEDGMENTS

We would like to acknowledge the following people for their contributions to this research project: Bogumil Karas for discussion and critiques, and Quynh Nguyen for assistance with cloning and genetically screening pPtPBR constructs.

FUNDING

Funding for this work was provided by the Gordon and Betty Moore Foundation (GBMF5007 to PW and CD, GBMF3828 and GBMF5006 to AA), the U.S. Department of Energy (DE-SC0008593) to AA and CD, and the National Science Foundation (OCE-1136477) to AA. Additional funding was provided to RD by UCSD/SIO Center for Marine Biodiversity and Conservation.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at http://journal.frontiersin.org/article/10.3389/fbioe.2016.00065

REFERENCES

Apt, K. E., Kroth-Pancic, P. G., and Grossman, A. R. (1996). Stable nuclear transformation of the diatom *Phaeodactylum tricornutum*. *Mol. Gen. Genet.* 252, 572–579. doi:10.1007/s004380050264

Bolivar, F., Rodriguez, R. L., Greene, P. J., Betlach, M. C., Heyneker, H. L., Boyer, H. W., et al. (1977). Construction and characterization of new cloning vehicles II. A multipurpose cloning system. *Gene* 2, 95–113. doi:10.1016/0378-1119(77)90000-2

Bowler, C., Allen, A. E., Badger, J. H., Grimwood, J., Jabbari, K., Kuo, A., et al. (2008). The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* 456, 239–244. doi:10.1038/nature07410

Bowman, S., Lawson, D., Basham, D., Brown, D., Chillingworth, T., Churcher, C. M., et al. (1999). The complete nucleotide sequence of chromosome 3 of *Plasmodium falciparum*. *Nature* 400, 532–538. doi:10.1038/22964

Bozarth, A., Maier, U. G., and Zauner, S. (2009). Diatoms in biotechnology: modern tools and applications. *Appl. Microbiol. Biotechnol.* 82, 195–201. doi:10.1007/s00253-008-1804-8

Cerón Garcí´a, M. C., Sánchez Mirón, A., Fernández Sevilla, J. M., Molina Grima, E., and Garcí´a Camacho, F. (2005). Mixotrophic growth of the microalga *Phaeodactylum tricornutum*. *Process Biochem.* 40, 297–305. doi:10.1016/j.procbio.2004.01.016

Cottarel, G., Shero, J. H., Hieter, P., and Hegemann, J. H. (1989). A 125 bp *CEN6* DNA fragment is sufficient for complete meiotic and mitotic centromere functions in *Saccharomyces cerevisiae*. *Trends Genet.* 5, 322–324. doi:10.1016/0168-9525(89)90119-4

Daboussi, F., Leduc, S., Maréchal, A., Dubois, G., Guyot, V., Perez-Michaut, C., et al. (2014). Genome engineering empowers the diatom *Phaeodactylum tricornutum* for biotechnology. *Nat. Commun.* 5, 3831. doi:10.1038/ncomms4831

De Riso, V., Raniello, R., Maumus, F., Rogato, A., Bowler, C., and Falciatore, A. (2009). Gene silencing in the marine diatom *Phaeodactylum tricornutum*. *Nucleic Acids Res.* 37, e96. doi:10.1093/nar/gkp448

Del Solar, G., Giraldo, R., Jesus Ruiz-Echevarria, M., Espinoza, M., and Diaz-Orejas, R. (1998). Replication and control of circular bacterial plasmids. *Microbiol. Mol. Biol. Rev.* 62, 434–464.

Dunahay, T. G., Jaruis, E. E., and Roessler, P. G. (1995). Genetic transformation of the diatoms *Cyclotella cryptica* and *Navicula saprophila*. *J. Phycol.* 31, 1004–1012. doi:10.1111/j.0022-3646.1995.01004.x

Falciatore, A., Casotti, R., Leblanc, C., Abrescia, C., and Bowler, C. (1999). Transformation of nonselectable reporter genes in marine diatoms. *Mar. Biotechnol.* 1, 239–251. doi:10.1007/PL00011773

Finnegan, J., and Sherratt, D. (1982). Plasmid ColE1 conjugal mobility: the nature of *bom*, a region required in *cis* for transfer. *Mol. Gen. Genet.* 185, 344–351. doi:10.1007/BF00330810

Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., and Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* 6, 343–345. doi:10.1038/nmeth.1318

Green, M. R., and Sambrook, J. (2012). *Molecular Cloning*, 4th Edn, Vol. 1. Cold Spring Harbor: Cold Spring Harbor Laboratory Press, 251.

Hayward, J. (2009). Studies on the growth of *Phaeodactylum tricornutum* III. The effect of iron on growth. *J. Mar. Biol. Assoc. U.K.* 48, 295–302. doi:10.1017/S0025315400034494

Iwanaga, S., Kato, T., Kaneko, I., and Yuda, M. (2012). Centromere plasmid: a new genetic tool for the study of *Plasmodium falciparum*. *PLoS ONE* 7:e333267. doi:10.1371/journal.pone.0033326

Iwanaga, S., Khan, S. M., Kaneko, I., Christodoulou, Z., Newbold, C., Yuda, M., et al. (2010). Functional identification of the *Plasmodium* centromere and generation of a *Plasmodium* artificial chromosome. *Cell Host Microbe* 7, 245–255. doi:10.1016/j.chom.2010.02.010

Jones, K. L., Kim, S. W., and Keasling, J. D. (2000). Low-copy plasmids can perform as well as or better than high-copy plasmids for metabolic engineering of bacteria. *Metab. Eng.* 2, 328–338. doi:10.1006/mben.2000.0161

Kanesaki, Y., Imamura, S., Matsuzaki, M., and Tanaka, K. (2015). Identification of centromere regions in chromosomes of a unicellular red alga, *Cyanidioschyzon merolae*. *FEBS Lett.* 589, 1219–1224. doi:10.1016/j.febslet.2015.04.009

Karas, B. J., Diner, R. E., Lefebvre, S. C., McQuaid, J., Phillips, A. P. R., Noddings, C. M., et al. (2015). Designer diatom episomes delivered by bacterial conjugation. *Nat. Commun.* 6, 6925. doi:10.1038/ncomms7925

Karas, B. J., Molparia, B., Jablanovic, J., Hermann, W. J., Lin, Y.-C., Dupont, C. L., et al. (2013). Assembly of eukaryotic algal chromosomes in yeast. *J. Biol. Eng.* 7, 30. doi:10.1186/1754-1611-7-30

Lopez, P. J., Desclés, J., Allen, A. E., and Bowler, C. (2005). Prospects in diatom research. *Curr. Opin. Biotechnol.* 16, 180–186. doi:10.1016/j.copbio.2005.02.002

Mann, C., and Davis, R. W. (1983). Instability of dicentric plasmids in yeast. *Proc. Natl. Acad. Sci. U. S. A.* 80, 228–232. doi:10.1073/pnas.80.1.228

Maruyama, S., Matsuzaki, M., Kuroiwa, H., Miyagishima, S.-Y., Tanaka, K., Kuroiwa, T., et al. (2008). Centromere structures highlighted by the 100%-complete *Cyanidioschyzon merolae* genome. *Plant Signal. Behav.* 3, 140–141. doi:10.1186/1741-7007-5-28.140

Miyagawa, A., Okami, T., Kira, N., Yamaguchi, H., Ohnishi, K., and Adachi, M. (2009). Research note: high efficiency transformation of the diatom *Phaeodactylum tricornutum* with a promoter from the diatom *Cylindrotheca fusiformis*. *Phycol. Res.* 57, 142–146. doi:10.1111/j.1440-1835.2009.00531.x

Muller, H., Annalura, N., Schwerzmann, J. W., Richardson, S. M., Dymonds, J. S., Cooper, E. M., et al. (2012). Assembling large DNA segments in yeast. *Methods Mol. Biol.* 852, 133–150. doi:10.1007/978-1-61779-564-0

Murray, A. W., and Szostak, J. W. (1983). Construction of artificial chromosomes in yeast. *J. Chem. Inf. Model.* 305, 189–193. doi:10.1017/CBO9781107415324.004

Newlon, C. S. (1988). Yeast chromosome replication and segregation. *Microbiol. Mol. Biol. Rev.* 52, 568–601.

Newlon, C. S., and Theis, J. F. (1993). The structure and function of yeast ARS elements. *Curr. Opin. Genet. Dev.* 3, 752–758. doi:10.1016/S0959-437X(05)80094-2

Nieduszynski, C. A., Knox, Y., and Donaldson, A. D. (2006). Genome-wide identification of replication origins in yeast by comparative genomics. *Genes Dev.* 20, 1874–1879. doi:10.1101/gad.385306

Nymark, M., Sharma, A. K., Sparstad, T., Bones, A. M., and Winge, P. (2016). A CRISPR/Cas9 system adapted for gene editing in marine algae. *Sci. Rep.* 6, 24951. doi:10.1038/srep24951

Price, N. M., Harrison, G. I., Hering, J. G., Hudson, R. J., Nirel, P. M. V., Palenik, B., et al. (1989). Preparation and chemistry of the artificial algal culture medium aquil. *Biol. Oceanogr.* 6, 443–461. doi:10.1080/01965581.1988.10749544

Siaut, M., Heijde, M., Mangogna, M., Montsant, A., Coesel, S., Allen, A., et al. (2007). Molecular toolbox for studying diatom biology in *Phaeodactylum tricornutum*. *Gene* 406, 23–35. doi:10.1016/j.gene.2007.05.022

Siow, C. C., Nieduszynska, S. R., Müller, C. A., and Nieduszynski, C. A. (2012). OriDB, the DNA replication origin database updated and extended. *Nucleic Acids Res.* 40, 682–686. doi:10.1093/nar/gkr1091

Stinchcomb, D. T., Thomas, M., Kelly, J., Selker, E., and Davis, R. W. (1980). Eukaryotic DNA segments capable of autonomous replication in yeast. *Proc. Natl. Acad. Sci. U. S. A.* 77, 4559–4563. doi:10.1073/pnas.77.8.4559

Strand, T. A., Lale, R., Degnes, K. F., Lando, M., and Valla, S. (2014). A new and improved host-independent plasmid system for RK2-based conjugal transfer. *PLoS ONE* 9:e90372. doi:10.1371/journal.pone.0090372

Sutcliffe, J. G. (1979). Complete nucleotide sequence of the *Escherichia coli* plasmid pBR322. *Cold Spring Harb. Symp. Quant. Biol.* 43, 77–90. doi:10.1101/SQB.1979.043.01.013

Tagwerker, C., Dupont, C. L., Karas, B. J., Ma, L., Chuang, R. Y., Benders, G. A., et al. (2012). Sequence analysis of a complete 1.66 Mb *Prochlorococcus marinus* MED4 genome cloned in yeast. *Nucleic Acids Res.* 40, 10375–10383. doi:10.1093/nar/gks823

Ukeles, R., and Rose, W. E. (1976). Observations on organic-carbon utilization by photosynthetic marine microalgae. *Mar. Biol.* 37, 11–28. doi:10.1007/BF00386774

Watson, N. (1988). A new revision of the sequence of plasmid pBR322. *Gene* 70, 399–403. doi:10.1016/0378-1119(88)90212-0

Weinstock, K. G., and Strathern, J. N. (1993). Molecular genetics in *Saccharomyces kluyveri*: the *HIS3* homolog and its use as a selectable marker gene in *S. kluyveri* and *Saccharomyces cerevisiae*. *Yeast* 9, 351–361. doi:10.1002/yea.320090405

Weyman, P. D., Beeri, K., Lefebvre, S. C., Rivera, J., McCarthy, J. K., Heuberger, A. L., et al. (2014). Inactivation of *Phaeodactylum tricornutum* urease gene using transcription activator-like effector nuclease-based targeted mutagenesis. *Plant Biotechnol. J.* 13, 460–470. doi:10.1111/pbi.12254

Wyrick, J. J., Aparicio, J. G., Chen, T., Barnett, J. D., Jennings, E. G., Young, R. A., et al. (2001). Genome-wide distribution of ORC and MCM proteins in *S. cerevisiae*: high-resolution mapping of replication origins. *Science* 294, 2357–2360. doi:10.1126/science.1066101

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Chapter 2, in full, is a reprint of the material as it appears in **Diner, RE**, Bielinski, VA, Dupont CP, Allen, AE, Weyman, PW (2016) Refinement of the Diatom Episome Maintenance Sequence and Improvement of Conjugation-based DNA Delivery Methods. *Frontiers in Biotechnology and Bioengineering.* 4, 65. The dissertation author was the primary investigator and author of this material.

CHAPTER 3

Diatom centromeres suggest a mechanism for nuclear DNA acquisition

Synopsis

In this chapter, I describe the sequence identity of native diatom centromeres, the first description of centromeres in the stramenopile lineage. I also demonstrate that DNA sequence similarity to native diatom centromeres allows DNA from many different sources, including bacterial conjugative plasmids and natural diatom plasmids, to become established as part of the diatom nuclear genome repertoire after being delivered by *E. coli* bacterial conjugation, essentially "hijacking" the diatom DNA replication machinery

This chapter is presented as a paper. "Diatom centromeres suggest a mechanism for nuclear DNA acquisition" was published as a research article in *Proceedings of the National Academy of Sciences* in 2017.

Diatom centromeres suggest a mechanism for nuclear DNA acquisition

Rachel E. Diner[a,b], Chari M. Noddings[c], Nathan C. Lian[c], Anthony K. Kang[c], Jeffrey B. McQuaid[a,b], Jelena Jablanovic[b], Josh L. Espinoza[b], Ngocquynh A. Nguyen[c], Miguel A. Anzelmatti Jr.[b], Jakob Jansson[c], Vincent A. Bielinski[c], Bogumil J. Karas[c,1], Christopher L. Dupont[b], Andrew E. Allen[a,b], and Philip D. Weyman[c,2]

[a]Integrative Oceanography Division, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA 92037; [b]Microbial and Environmental Genomics Group, J. Craig Venter Institute, La Jolla, CA 92037; and [c]Synthetic Biology and Bioenergy Group, J. Craig Venter Institute, La Jolla, CA 92037

Edited by James A. Birchler, Division of Biological Sciences, University of Missouri, Columbia, MO, and approved June 13, 2017 (received for review January 17, 2017)

Centromeres are essential for cell division and growth in all eukaryotes, and knowledge of their sequence and structure guides the development of artificial chromosomes for functional cellular biology studies. Centromeric proteins are conserved among eukaryotes; however, centromeric DNA sequences are highly variable. We combined forward and reverse genetic approaches with chromatin immunoprecipitation to identify centromeres of the model diatom *Phaeodactylum tricornutum*. We observed 25 unique centromere sequences typically occurring once per chromosome, a finding that helps to resolve nuclear genome organization and indicates monocentric regional centromeres. Diatom centromere sequences contain low-GC content regions but lack repeats or other conserved sequence features. Native and foreign sequences with similar GC content to *P. tricornutum* centromeres can maintain episomes and recruit the diatom centromeric histone protein CENH3, suggesting nonnative sequences can also function as diatom centromeres. Thus, simple sequence requirements may enable DNA from foreign sources to persist in the nucleus as extrachromosomal episomes, revealing a potential mechanism for organellar and foreign DNA acquisition.

diatom | *Phaeodactylum tricornutum* | episome | centromere | CENH3

Centromeres play a crucial role in the cellular biology of eukaryotes by acting as a genomic site for kinetochore formation and facilitating effective transmission of replicated nuclear DNA to new cells. Centromere-associated proteins are functionally conserved among eukaryote species (1–3). Nearly all eukaryotes studied to date possess a version of a specialized centromeric histone protein (CENH3, also described as centromere protein A, CENP-A), which binds to centromeric DNA and replaces the histone H3 at the site of kinetochore assembly (4–6). Conversely, the centromeric DNA sequences themselves are extremely variable and appear to evolve rapidly, even among similar organisms (7).

There are three general types of eukaryotic centromeres: point centromeres, holocentromeres, and regional centromeres. Point centromeres are uniquely characterized by specific conserved DNA sequences and are found in limited fungal species including the budding yeast *Saccharomyces cerevisiae* and close relatives (8–10). In holocentromeric organisms, the kinetochore forms along the entire length of each chromosome; a notable example is the model organism *Caenorhabditis elegans* (11). Most eukaryotes have regional centromeres, which are commonly found as a single large DNA region on each chromosome [reviewed in Sullivan et al., 2001 (12) and Torras-Llort et al., 2009 (13)]. Regional centromeres are variable in length and sequence even among closely related species; however, there are often predictable genetic features. For example, human centromeres contain large stretches of repetitive satellite DNA, ranging in size from hundreds of kilobases to megabases (12, 14, 15). Centromeres of several plants and the insect model *Drosophila melanogaster* contain large arrays of satellite repeats interspersed with or adjacent to retrotransposons, which can vary substantially in copy number and organization (16). A common feature of centromeric DNA in many eukaryotes is low-GC content. Centromeres of *Schizosaccharomyces pombe* and other yeast species feature an unconserved core of AT-rich DNA sequence often surrounded by inverted repeats (17–20). The centromeres of the protist *Plasmodium* have no apparent sequence similarity besides being 2–4-kb regions of extremely low-GC content (<3%) (21, 22). Likewise, centromere regions of the red algal species *Cyanidioschyzon merolae* contain 2–3 kb of relatively low-GC content but manifest no other apparent pattern (23, 24).

Centromere identification can also be useful for synthetic biology, enabling further discoveries and biotechnology applications. Artificial chromosomes provide a stable platform for introduction and maintenance of multigene constructs necessary for expression of biosynthetic pathways and large complex proteins (25–28). The experimental identification of eukaryotic centromeres has been extremely useful for developing molecular

Significance

Centromeres are genomic sites facilitating chromosome segregation during cell division. We report our discovery of diatom centromeres and the description of centromere identity in the stramenopile protists. We also show that simple requirements for diatom centromeres permit ecologically relevant foreign DNA molecules to function as diatom centromeres by "hijacking" chromosome maintenance features. Because little is known at the molecular level about chromosome maintenance in diatoms, this paper provides experimental data with broad implications for cellular biology studies and biotechnology applications. The ability to maintain circular artificial chromosomes using foreign DNA sequences is unique among organisms with studied centromeres and opens up fascinating evolutionary questions about the mechanisms of nuclear gene acquisition from the multiple endosymbiotic events characterizing the stramenopile lineage.

Author contributions: R.E.D., C.M.N., N.C.L., A.K.K., J.B.M., J. Jablanovic, J.L.E., V.A.B., B.J.K., C.L.D., A.E.A., and P.D.W. designed research; R.E.D., C.M.N., N.C.L., A.K.K., J.B.M., J. Jablanovic, J.L.E., N.A.N., M.A.A., J. Jansson, V.A.B., and P.D.W. performed research; R.E.D., C.M.N., N.C.L., A.K.K., J.B.M., J. Jablanovic, J.L.E., N.A.N., M.A.A., V.A.B., B.J.K., C.L.D., A.E.A., and P.D.W. analyzed data; and R.E.D., J.L.E., C.L.D., A.E.A., and P.D.W. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The sequence reported in this paper has been deposited in the National Center for Biotechnology Information Sequence Read Archive (accession no. PRJNA357294).

[1]Present addresses: Department of Biochemistry, Schulich School of Medicine and Dentistry, Western University, London, ON N6A 5C1, Canada; and Designer Microbes Inc., London, ON N6G 4X8, Canada.

[2]To whom correspondence should be addressed. Email: pweyman@jcvi.org.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1700764114/-/DCSupplemental.

biology tools, particularly in the creation of artificial chromosomes. Circular and/or linear artificial chromosomes based on native centromeres, origins of replication, and in some cases telomeres have been developed for yeast (29), mammalian cells including human cell lines (30), plants (reviewed in ref. 31), and recently the protist *Plasmodium* (32). Despite the great potential for eukaryotic algae in biotechnology, very little is known about algal centromeres, and few resources are available to control gene expression from introduced autonomously replicating genetic constructs. In 1984, autonomously replicating plasmids using chloroplast DNA were described for the green alga *Chlamydomonas reinhardii* (33). However, these vectors were not maintained stably and have not been commonly used. More recently, centromeres have been identified and characterized in the red alga *C. merolae* (23, 24), where each of the 20 chromosomes was found to contain one distinct region recruiting CENH3. However, to our knowledge, these sequences have not yet been used for the construction of artificial chromosomes.

Identifying centromere composition and optimizing artificial chromosome construction would be particularly valuable for diatoms, which are an abundant group of eukaryotic phytoplankton with important ecological significance. Diatom research has facilitated major discoveries in algal physiology and genetics, and several species have been cultivated and genetically manipulated for the development of valuable bioproducts (34–36). In our previous work, we discovered that a region of *S. cerevisiae* DNA containing low-GC content enabled the stable maintenance of autonomously replicating episomes in diatoms (37, 38). The DNA was introduced into the diatoms *Phaeodactylum tricornutum* and *Thalassiosira pseudonana* by bacterial conjugation, also suggesting a previously unexplored mechanism for horizontal gene transfer from bacteria. Diatom nuclear genomes contain large amounts of DNA derived from nonnuclear sources, including foreign sequences such as bacteria and viruses, and prokaryotic and eukaryotic DNA obtained from endosymbiotic events (e.g., mitochondria, chloroplasts, and additional secondary endosymbioses) (39–42). This genetic complexity and rapid evolution contributes to the ecological success of diatoms. Thus, elucidating mechanisms that may facilitate nuclear gene acquisition and episomal maintenance will advance our knowledge of diatom evolution and enable biotechnological innovation.

Here, we identify centromeric regions of diatom chromosomes using forward and reverse genetics approaches and observe that diatom centromeres are characterized by a simple low-GC signal, which is also found in the previously described synthetic diatom episomes (37, 38). Furthermore, we show that nonnuclear diatom DNA and foreign DNA from a variety of sources with similarly low-GC content can mimic a diatom centromere, suggesting a permissive mechanism for nuclear gene acquisition. This study significantly advances the understanding of diatom genomic features, facilitates the development of diatom molecular tools, and suggests a mechanism for diatom acquisition of foreign genetic material.

Results

Identification of Putative Diatom Centromeres in *P. tricornutum* Chromosomes 25 and 26. We hypothesized that a centromeric region of a diatom chromosome would support maintenance of a nuclear episome, as this is a useful experimental method of confirming centromere function for other organisms (32, 43). To identify a diatom centromeric region, we first examined the shortest *P. tricornutum* chromosomes with telomere-to-telomere assembly (25 and 26) (39), which were each previously cloned as five overlapping ~100-kb (76.6–142.6 kb) DNA fragments (44). In our prior studies (37, 38), sequences supporting episome maintenance in *P. tricornutum* were characterized by greatly improved ex-conjugant colony yield compared with plasmids incapable of episome maintenance. Thus, we predicted 100-kb fragments from a single *P. tricornutum* chromosome that supported episome maintenance would yield similarly increased colony numbers in our standard conjugation assay. Out of the five large fragments spanning each chromosome, one fragment from each chromosome produced increased numbers of ex-conjugant diatom colonies: plasmid Pt25-100kb-1 (containing the first ~100-kb fragment of chromosome 25) (Fig. 1 *A* and *C*) and plasmid Pt26-100kb-5 (containing the fifth fragment of chromosome 26) (Fig. 1 *B* and *D*). The plasmid containing Pt25-100kb-1 resulted in 14–32-fold more colonies than plasmids containing other 100-kb fragments from chromosome 25 (Fig. 1 *A* and *C*), and plasmid containing Pt26-100kb-5 resulted in 26–100-fold higher colony numbers than other chromosome 26 fragments (Fig. 1 *B* and *D*).

Both Pt25-100kb-1 and Pt26-100kb-5 fragments encompass regions of low-GC content. We calculated the GC content of the genome in 100-bp windows overlapping by 50 bp and found that windows with the lowest GC content were found on fragments enabling episome maintenance (Fig. 1 *E* and *F*). When calculating GC percentage with larger window sizes (10 kb to 0.5 kb), an obvious dip in GC content was not apparent on chromosomes 25 and 26 (*SI Appendix*, Fig. S1). We quantified the number of 100-bp windows less than or equal to 32% GC within a 3-kb larger window and observed clear peaks for chromosomes 25 and 26 (Fig. 1 *G* and *H*).

To clarify whether these specific chromosomal regions enriched in low-GC content enabled episome maintenance, three 10-kb DNA subsequences of Pt25-100kb-1 were cloned into plasmids otherwise incapable of maintenance (pPtPBR2) (38): one sequence encompassing the bioinformatically identified low-GC region (Pt25-10kb-12) (Fig. 1*E*), and two other randomly selected sequences (Pt25-10kb-6 and Pt25-10kb-9) (*SI Appendix*, Fig. S2). Pt25-10kb-12 conjugation led to 85-fold more colonies than the negative control, whereas the other plasmids showed no increase (*SI Appendix*, Fig. S2). We further tested the low-GC region found on Pt25-10kb-12 by assembling a 1-kb subregion containing the lowest GC content region of chromosome 25 into pPtPBR2. This plasmid, Pt25-1kb, yielded 27-fold more colonies than the empty vector control (*SI Appendix*, Fig. S2). Another plasmid containing the 1-kb region encompassing the lowest GC content region of chromosome 26, Pt26-1kb, resulted in 68-fold more colonies than the empty vector control. Thus, for chromosomes 25 and 26, regions containing the lowest GC content were the only regions supporting episome maintenance. To confirm that these plasmids were maintained in the diatoms over extended periods of time, two clones of Pt25-1kb were passaged for 30 d with and without selection. As *P. tricornutum* was observed to divide about once a day in the experimental conditions, this corresponds to roughly 30 generations. Antibiotic-resistant colonies were recovered at percentages similar to prior studies (37, 38), which correspond to high per-generation segregation efficiencies (*SI Appendix*, Table S1) (32). Plasmids were recovered and confirmed by gel electrophoresis after the passaging period (*SI Appendix*, Fig. S3), demonstrating the stable maintenance of episomes in these lines (i.e., not integrated into native chromosomal DNA).

Identification of Diatom Centromeres Using ChIP-Sequencing and Reverse and Forward Genetics. *P. tricornutum* genomic DNA sequences enabled episome maintenance in the diatom, suggesting these regions were functioning as centromeres. Nearly all eukaryotes previously studied incorporate the centromeric histone CENH3 into centromeric nucleosomes, and we tested this in *P. tricornutum* to confirm centromere functionality. We constructed an episome containing the *CEN6-ARSH4-HIS3* maintenance sequence and a translational fusion of *P. tricornutum* CENH3 and yellow fluorescent protein (YFP) regulated by a *P. tricornutum* promoter and terminator. After transfer to *P. tricornutum* using bacterial conjugation (see *SI Appendix*, *Materials and Methods*), we performed chromatin immunoprecipitation (ChIP) assays on

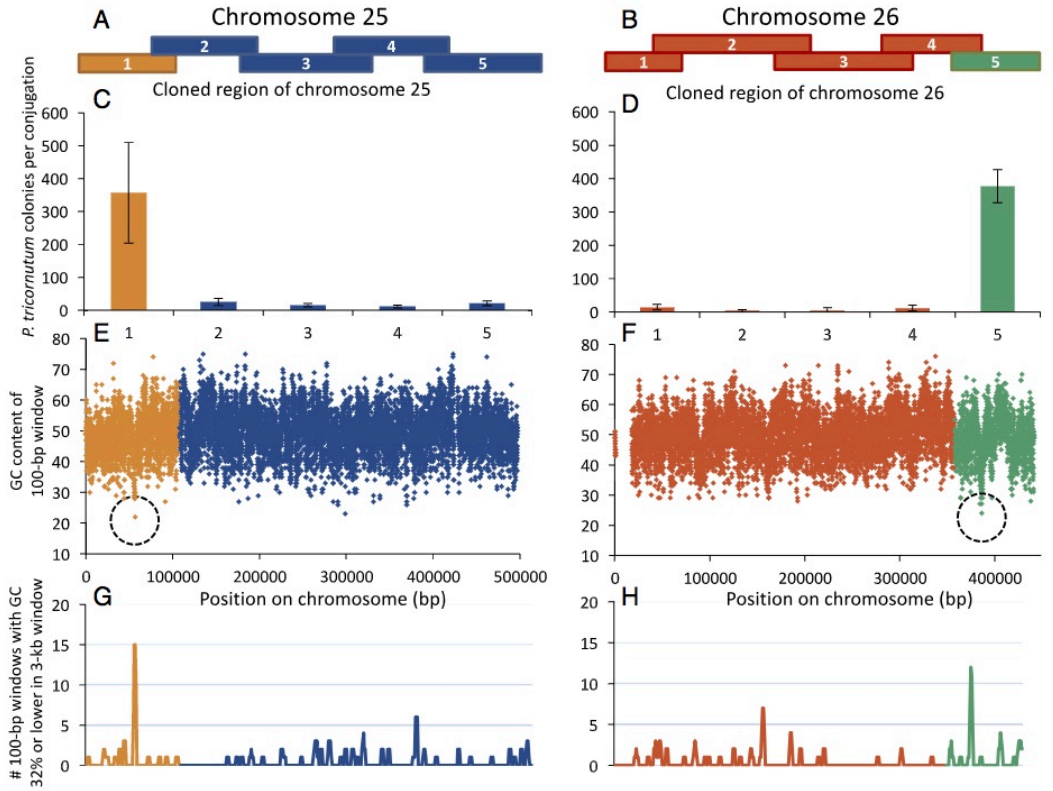

Fig. 1. Regions of *P. tricornutum* chromosomes enriched for low GC support episomal maintenance. (*A* and *B*) Chromosomes 25 and 26 were cloned as five overlapping ~100-kb fragments. (*C* and *D*) Number of resulting *P. tricornutum* colonies per conjugation for episomes containing the indicated region of chromosome 25 or 26. Error bars indicate SD of four independent conjugation reactions for each fragment. (*E* and *F*) GC content was calculated for chromosomes 25 and 26 in 100-bp sliding windows that overlapped by 50 bp. Dashed circles indicate the lowest GC content for the chromosome in a 100-bp window. (*G* and *H*) Number of 100-bp windows with GC content of 32% or lower within a larger sliding 3-kb window that advanced by 1 kb each step.

ex-conjugant lines using GFP epitope antisera, followed by high-throughput DNA sequencing to identify all *P. tricornutum* genome sequences that recruit the centromeric histone.

ChIP-sequencing (ChIP-seq) analysis revealed 25 regions that were enriched for sequence reads (peaks) among the previously reported 33 nuclear chromosome scaffolds (39) (*SI Appendix*, Figs. S4 and S5). The low-GC regions of the chromosomes were specifically enriched for ChIP-seq reads (chromosome 25 shown for reference in Fig. 2; others in *SI Appendix*, Fig. S4). Of the 12 chromosome scaffolds with telomere-to-telomere assembly, all but one (chromosome 11) had ChIP-seq peaks, including chromosomes 25 and 26 (Fig. 3). Two regions recruiting CENH3 were also found within the nonscaffold assemblies ("bottom drawer" sequences) (obtained from the JGI *P. tricornutum* genome website: genome.jgi.doe.gov/Phatr2/Phatr2.home.html) (*SI Appendix*, Fig. S4). A ChIP-seq peak was also identified within the *S. cerevisiae CEN6-ARSH4-HIS3* region on the episome used to express the YFP-CENH3 fusion protein (Fig. 2). No mitochondrial or chloroplast sequences recruited CENH3, which was expected, as these genomes do not contain nucleosomes. Most ChIP-seq peaks on a genome-wide scale colocalized with the presence of at least ten 100-bp windows with GC content less than or equal to 32% GC in a larger 3-kb region (Fig. 2 and *SI Appendix*, Fig. S4).

To verify the ChIP-seq data, we conducted ChIP-qPCR on two regions with ChIP-seq peaks, one in the genome (Pt25-1kb) and one in the episome (*ARSH4*), and a region of genomic and episomal DNA without ChIP-seq peaks as a control (see *Materials and Methods*) (*SI Appendix*, Fig. S6 *A* and *B*). After ChIP, DNA from the low-GC *ARSH4* episomal region was in greater abundance by >50–70-fold compared with the negative control (*SI Appendix*, Fig. S6 *C–E*). Similarly, the Pt25-1kb region was enriched >200–500-fold compared with the genomic DNA negative control (*SI Appendix*, Fig. S6 *C–E*). Thus, ChIP-qPCR confirmed the ChIP-seq results for the CENH3 enriched regions of both episomal and native *P. tricornutum* chromosomal targets.

Of the 25 chromosome scaffolds with ChIP-seq hits, 23 had only one associated ChIP-seq peak that was between 2.4 and 5.6 kb (*SI Appendix*, Fig. S5). Chromosomes 2 and 8 each had two adjacent ChIP-seq peaks (*SI Appendix*, Fig. S5). Both putative centromeres on chromosome 2 (2a and 2b) are contained within a larger direct repeat and separated by a sequencing gap (indicated by Ns in the *P. tricornutum* genome sequence) (*SI Appendix*, Fig. S5). These sequences were highly similar to each other, with ~2.9 kb aligning along the 3.4-kb sequence at >99% sequence identity. The two putative centromeres on chromosome 8 (8a and 8b, respectively) are each partially contained

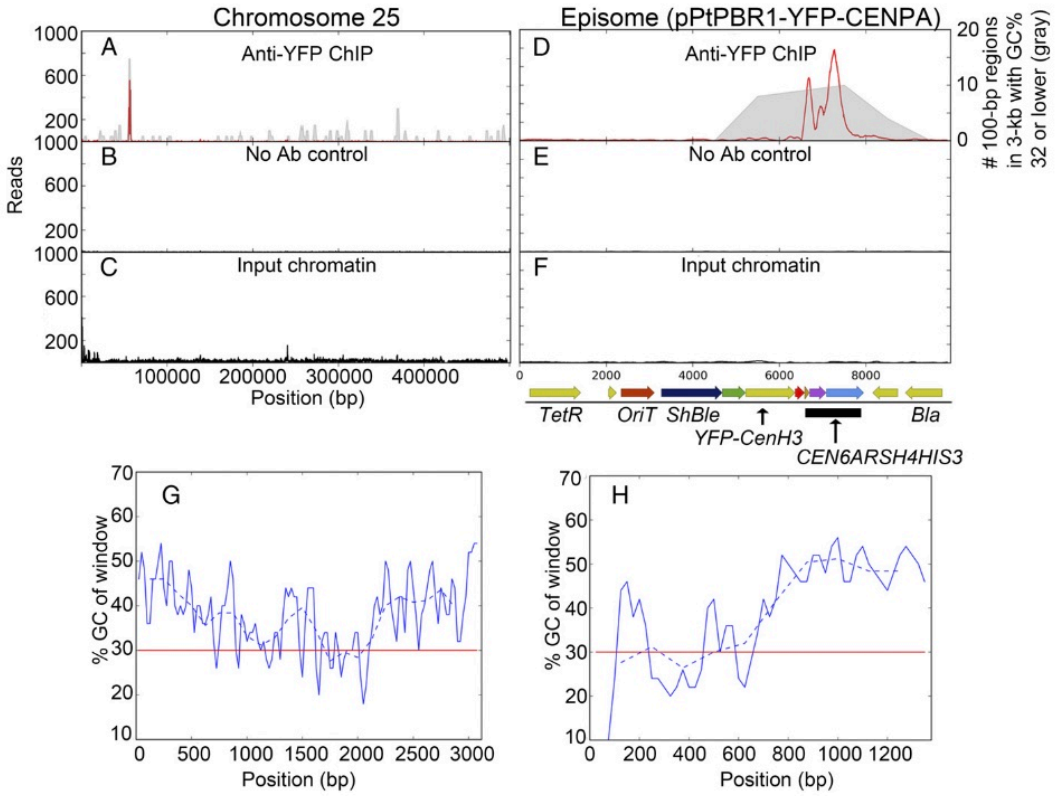

Fig. 2. ChIP-seq and GC data for chromosomee 25 and the episome. For chromosome 25 and the episome, ChIP-seq reads at each position for treatments with the YFP antibody (red) were plotted on the same graph as the number of 100-bp windows with GC 32% or lower in a larger 3-kb window (gray) (*A* and *D*). Graphs of the number of reads for the no-antibody ChIP-seq control (*B* and *E*) and input chromatin (*C* and *F*) were plotted using the same position scale as the anti-YFP ChIP-seq. For the episome, the positions of the genetic features are indicated below the input chromatin (the black bar indicates the *CEN6–ARSH4–HIS3* region). For the peaks identified by the CENH3-YFP ChIP-seq in chromosome 25 and the episome, GC content for 100-bp windows (50 bp overlap, solid blue line) or 250-bp windows (125-bp overlap, dashed blue line), respectively, was plotted with a reference line at 30% in red (*G* and *H*).

within long direct repeats at the 3′ end of the centromere. The 5′ end of 8b is adjacent to a region of unknown sequence (*SI Appendix*, Fig. S5). The 8a and 8b centromere sequences were also highly similar, with alignment across about half of the centromere sequence at 96.5% identity.

Apart from these potentially tandem centromere cases, most *P. tricornutum* centromeres were unique, having no similarity to other centromere sequences, with two exceptions. Predicted centromeres from chromosomes 24 and 29 shared 99.2% sequence identity over the entire 2.4-kb region and differed by only 14 mismatches. Additionally, the centromere from chromosome 30 shared a 1.6-kb region of high identity (97%) to a bottom drawer sequence bd23 × 34, which was one of the two bottom drawer sequences with an associated ChIP-seq hit. Centromeres in *P. tricornutum* were mostly located in intergenic spaces (*SI Appendix*, Fig. S5). Direct repeats were detected in approximately one-third of the centromeres, but the repeat number was low (usually a single sequence found twice) and the repeat period was variable and small (16–400 bp) (*SI Appendix*, Table S2). Genomic coordinates of all predicted centromeres, including ChIP-seq read regions and bioinformatically predicted regions containing low-GC content, are noted in *SI Appendix*, Table S3.

We used forward genetics to test whether sequences in the *P. tricornutum* genome including and in addition to those identified by ChIP-seq could support episomal maintenance. We prepared a *P. tricornutum* genomic library with 2–5-kb inserts using a nonepisome vector (pPtPBR2) and conjugated the library into *P. tricornutum* cells. Episomes were identified by extracting plasmids from antibiotic-resistant *P. tricornutum* colonies and transforming *Escherichia coli*; only DNA maintained as circular episomes in *P. tricornutum* was expected to yield *E. coli* colonies. We amplified and sequenced *P. tricornutum* genomic library inserts from *E. coli* colonies and identified 35 unique insert sequences from 99 recovered plasmids (*SI Appendix*, Table S4). Of these 35 unique insert sequences, 10 mapped to the nuclear genome chromosomal scaffolds and 1 mapped to the unscaffolded bottom drawer assemblies. Eighteen sequences mapped to the chloroplast genome, and 6 mapped to the mitochondrial genome.

Reverse genetics was used to functionally test whether the sequences identified by ChIP-seq and the *P. tricornutum* forward genetics library could maintain episomes. Forty sequences, including all ChIP-seq peaks, potential ChIP-seq artifacts, and *P. tricornutum* forward genetic library sequences including selected mitochondrial and chloroplast DNA sequences, were

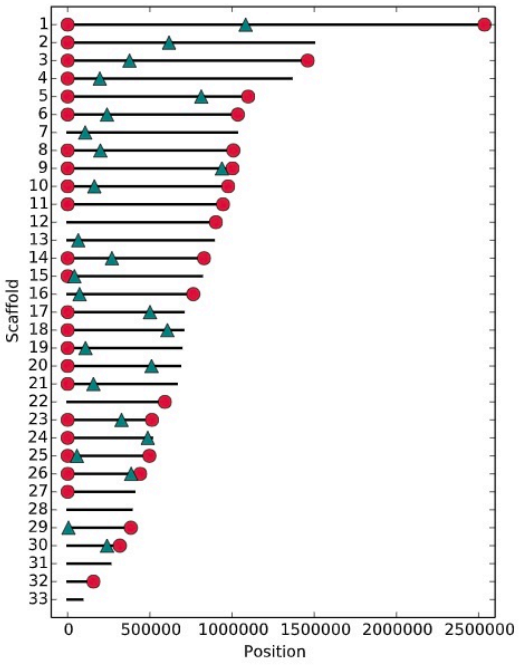

Fig. 3. Centromere and telomere locations on *P. tricornutum* chromosomal-scale scaffolds. Telomeres (crimson circles) and centromeres (teal triangles) are annotated on each scaffold.

cloned into the nonepisomal plasmid pPtPBR2 (see *Materials and Methods*). Most plasmids containing ChIP-seq identified sequences resulted in 7–162-fold more diatom ex-conjugant colonies than the pPtPBR2 negative control (*SI Appendix*, Fig. S7). We also tested random regions of chromosome 1 as negative controls (Test-37, -38, and -39) and regions suspected to be ChIP-seq mapping artifacts based on high read counts in both input and anti-YFP immunoprecipitation treatments (Test-4, -10, and -16). Both classes of sequences were unable to support episome maintenance; ex-conjugant numbers were similar to the negative control and much lower than the positive control pPtPBR1 (*SI Appendix*, Fig. S7). Ex-conjugant colony numbers following conjugation with the pPtPBR1 positive control (containing *CEN6-ARSH4-HIS3*) were not notably different from the episomes containing putative *P. tricornutum* centromeres. One insert sequence from chromosome 11 contained a region of GC content similar to, but slightly higher than, the centromeres (Test-40). However, this region contained no ChIP-seq peak and was unable to maintain an episome (*SI Appendix*, Fig. S7).

We also tested the *P. tricornutum* regions recovered from the forward genetic screen for the ability to maintain episomes. All chloroplast and mitochondrial DNA sequences, the bottom drawer sequence, and 8 of the 10 nuclear genome sequences contained low-GC content of 28–41% (*SI Appendix*, Table S4) across the entire insert region. These 8 nuclear genome sequences and the bottom drawer sequence mapped to identical regions as the ChIP-seq peaks (*SI Appendix*, Fig. S4). The two remaining inserts (Test-18 and Test-20) from the nuclear genome had GC content typical of the *P. tricornutum* nuclear genomic DNA (47%) and did not map to a ChIP-seq peak (*SI Appendix*, Table S3). We retested whether the two high-GC nuclear genome inserts as well as two sequences each from the chloroplast (Test-33 and Test-34) and the mitochondrion (Test-35 and Test-36) could support episome maintenance. Both mitochondrial and both chloroplast sequences supported episomes (*SI Appendix*, Table S4); however, the high-GC nuclear sequences did not, and we predict that their appearance in the library was likely due to plasmid carryover from the initial conjugation (*SI Appendix*, Table S4).

To further examine the minimum sequence size required for centromere function, we combined the information obtained in our initial screening of chromosome 25 with ChIP-seq identification of the full chromosome 25 centromere sequence to test the functionality of various small subsequences (*SI Appendix*, Fig. S8 and Table S5). We designed a series of 19 constructs to test sequences ranging in size from 198 to 1,040 bp. These sequences included a systematic minimization of centromere 25, particularly the 1-kb region (Pt25-1kb) shown above to maintain episomes, as well as a test of low-GC regions of sequentially smaller sizes. When we broke the ∼3-kb region that recruited CENH3 into three equal parts of 1,040 bp, only the middle third supported episomal maintenance. This middle third contained the majority of the Pt25-1kb insert and almost all of the lowest GC 500-bp region. When we further dissected the Pt25-1kb insert, only sequences >500 bp could maintain episomes, and all functional sequences encompassed low-GC DNA sequences, particularly the downstream region of the lowest 500-bp region (*SI Appendix*, Fig. S8).

Foreign DNA Sequences Examined for Episome Maintenance. Because the *CEN6–ARSH4–HIS3* sequence from *S. cerevisiae* supported episome maintenance in *P. tricornutum*, we hypothesized that other foreign DNA sequences with similarly low-GC composition could as well. Deletion analysis of the *CEN6–ARSH4–HIS3* region previously revealed that low-GC regions of >∼500 bp enabled maintenance. To test this pattern in the present study, we examined 24 sequences from *Mycoplasma mycoides* JCVI Syn1.0 (NCBI accession no. CP002027) of various sizes (0.5–1 kb) and GC content (15–50%) for their ability to maintain diatom episomes. All sequences of less than 28% GC content regardless of the size resulted in high numbers of ex-conjugant colonies consistent with episome maintenance (Fig. 4). Most sequences of 28% and 30% GC also resulted in large numbers of *P. tricornutum* ex-conjugant colonies with two exceptions that produced colony numbers similar to the negative control: a 500-bp 28% GC fragment (1.3-fold below control), and a 500-bp 30% GC fragment (1.2-fold above control) (Fig. 4). Additionally, one 700-bp 30% GC fragment produced only 3.3-fold more colonies than the control, a relatively low colony increase. The fragments containing either 40% or 50% GC content sequences produced ex-conjugant colony numbers similar to the negative control. Thus, with a few exceptions (Fig. 5), DNA sequences of ∼30% GC or lower were required and sufficient to support *P. tricornutum* episomes.

The above results suggest that many sequences of at least 500 bp (the smallest fragment tested) of low-GC DNA could maintain an episome in *P. tricornutum*, including sequences with environmental relevance. We examined whether a marine bacterial conjugative plasmid could support episome maintenance by searching the *Alteromonas macleodii* conjugative plasmid pAMDE1 for low-GC content regions (*SI Appendix*, Fig. S9). We then identified and cloned two 500-bp regions, AM-1 and AM-2, with 26.2% and 28.8% GC, respectively; conjugation of plasmids containing either region yielded 6–17-fold more ex-conjugant *P. tricornutum* colonies than the pPtPBR2 negative control with no maintenance sequence elements (*SI Appendix*, Fig. S9). We also tested whether regions of plasmids previously isolated from the diatom *Cylindrotheca fusiformis* (45, 46) could support episomes in *P. tricornutum*. Two plasmids, pCF1 and pCF2, containing low-GC, 560-bp regions (28.9% and 28.4% GC, respectively)

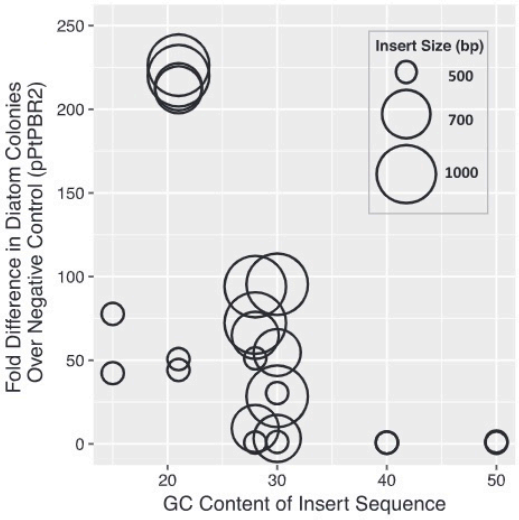

Fig. 4. Maintenance of episomes containing *M. mycoides* DNA sequences. Inserts of various GC content (15%, 21%, 27%, 30%, 40%, and 50% GC) from *M. mycoides* were tested for the ability to support episomal maintenance in *P. tricornutum*. The number of diatom ex-conjugant colonies obtained after conjugation is shown as the fold increase in colony numbers over the pPtPBR2 negative control for plasmids containing inserts of different size and GC content. The size of each circle represents the size of the insert sequence tested: large circles, 1,000 bp; medium circles, 700 bp; and small circles, 500 bp. The center of the circle indicates the data point plotted.

were constructed (see *SI Appendix, Materials and Methods*), and each yielded 7–12-fold more *P. tricornutum* ex-conjugant colonies than the pPtPBR2 negative control (*SI Appendix*, Fig. S9).

We examined maintenance properties of episomes supported by foreign DNA sequences to identify whether these plasmids were stable over time. *P. tricornutum* ex-conjugant lines were maintained with and without antibiotics for 30 d (see *Materials and Methods*). For all clones, as well as all other experiments where conjugations resulted in a high number of ex-conjugant diatom colonies relative to the negative control, episomes were successfully recovered in *E. coli*, confirming their stable extrachromosomal maintenance. In lines containing plasmids with two different *Mycoplasma* inserts (Myco-15–500bp-2 and Myco-21–500bp-2), between 45% and 75% of cells retained the episome without selection (per-generation segregation efficiency of 97–99%), and 91–95% retained it with selection (>99% segregation efficiency). *A. macleodii* and *C. fusiformis* DNA-containing sequences, with the exception of colony 8 of the AM-2 plasmid, were maintained with retention rates between 24% and 84% in the absence of antibiotics and 77% and 93% with antibiotics, corresponding to segregation efficiencies of >95% and >99%, respectively (*SI Appendix*, Table S1 and Fig. S3). In colony 8 containing the AM-2 episome, only 3% of cells retained the episome after passaging without antibiotic, with 74% retained with antibiotics. Although much lower than the other constructs, this still corresponds to segregation efficiencies of 89% without antibiotics and 99% with antibiotics. With this exception, retention rates of foreign DNA plasmids were similar to maintenance of episomes containing the *CEN6–ARSH4–HIS3* sequence (37, 38) and the native *P. tricornutum* centromere sequence from chromosome 25 (*SI Appendix*, Table S1). Episomes maintained with selection, a scenario more similar to native chromosomes containing essential genes, were maintained at much higher levels in all lines (*SI Appendix*, Table S1).

Bioinformatic Analysis of Episome-Supporting Sequences. Our results indicate that 500-bp sequences can maintain episomes. Thus, we searched within these sequences for the 500-bp subregion with the lowest GC content (*SI Appendix*, Table S6). When viewed together based on ability to maintain an episome, all inserts from the native diatom "Test" series and all foreign DNA inserts examined (including *M. mycoides*, *C. fusiformis*, and *A. macleodii* plasmid pAMDE1 source DNA) indicated a clear pattern of low-GC content supporting episome maintenance regardless of whether the source was foreign or native (Fig. 5). However, three *M. mycoides* DNA sequences (28-500-2, 30-500-2, and 30-700-2) that were predicted to be maintained based on low average GC content produced low numbers of ex-conjugant colonies after conjugation. This suggested that additional signals besides average GC content might be important.

Native *P. tricornutum* centromeres do not have repeats or other structures, and attempts to identify a conserved sequence motif using BLAST (47) and MEME (48) were unsuccessful, so we examined k-mer use to determine if very short sequences were overrepresented in DNA fragments supporting episomes. We chose a k-mer length of 6 because it was the longest string that could still be well-represented in a sequence of 500 bp. We identified unique 6 mers overrepresented in native *P. tricornutum* centromeres by requiring their retention to be statistically significant ($P < 0.001$) compared with randomly selected *P. tricornutum* genomic sequence (47% GC) and randomly generated sequences of 47% GC. Because the overall GC content is lower for centromeric ChIP-seq peaks (39% GC average) compared with the genomic regions (47% GC), we also required the 6 mers to be significantly overrepresented in the centromeres relative to a randomly generated set of 39% GC sequences. This allowed us to identify 6 mers overrepresented in the *P. tricornutum* centromeres that were unexplained by GC content difference from the genomic DNA (*SI Appendix*, Table S7). We then examined the recruitment of this set of centromere-enriched 6 mers in two sets of *Mycoplasma* fragments. One set contained the two 28% GC sequences and one 30% GC sequence that did not support

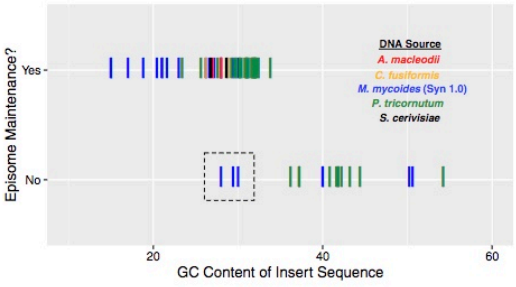

Fig. 5. Relationship between GC content and episome maintenance. The 500-bp subregion with the lowest GC content was identified for each insert sequence in the *P. tricornutum* test series (which includes ChIP-seq peaks, forward genetic library sequences, designed negative controls, and potential ChIP-seq artifacts) and foreign DNA inserts (including *M. mycoides*, *C. fusiformis*, and *A. macleodii* plasmid pAMDE1 source DNA). See *SI Appendix*, Table S6 for data included in the figure. This lowest GC content subregion was plotted as a function of whether the DNA could support episomal maintenance in *P. tricornutum*. DNA sequences are colored by the organism from which each insert sequence originated. The black dotted box indicates three sequences from *M. mycoides* that failed to support episomes despite being in the 28–30% GC range.

episome maintenance despite having a sufficiently low average GC content ("Myco-No" set). The second set comprised the remaining nine *Mycoplasma* sequences with 28% and 30% average GC that successfully supported episome maintenance ("Myco-Yes" set). The 6 mers that were overrepresented in the Myco-Yes set were characterized by very low GC content (i.e., the most abundant 6 mers in the Myco-Yes set were composed entirely of A+T bases) (*SI Appendix*, Fig. S10). When we directly calculated the number of consecutive A+T nucleotides in the *Mycoplasma* sequences that supported episome maintenance compared with those that did not, stretches of 6 or more consecutive A+T bases were more frequent in the *Mycoplasma* fragments that supported episome maintenance (i.e., Myco-Yes; *SI Appendix*, Table S8). The lower distribution of consecutive A+T bases in the Myco-No set was also observed compared with a set of randomly generated sequences of 30% GC (*SI Appendix*, Table S8). Thus, the Myco-No samples that failed to support episome maintenance appear to have fewer long stretches composed of A+T residues despite having the same average GC content as fragments that supported episome maintenance in *P. tricornutum*.

Discussion

Features of Predicted Diatom Centromeres. In this study, we identified native diatom centromere sequences with high resolution. Based on previous studies, we hypothesized that low-GC content would be a common characteristic of diatom centromeres. We deconstructed two *P. tricornutum* chromosomes (25, 26) and found that regions with low-GC content appeared to function as centromeres, whereas adjacent regions did not. We subsequently conducted a genome-wide ChIP-seq screen (confirmed with ChIP-qPCR) and a forward genetic screen to identify centromeres and additional sequences enabling episome maintenance and used reverse genetics to test for function. We discovered 25 unique *P. tricornutum* centromeric DNA sequences: 24 among the nuclear genome scaffolds and 1 in the nonscaffolded genome assemblies. Although there may potentially be more centromeres we did not identify here, if our results are a good estimate of diatom chromosome number, with one unique centromere sequence each, we would predict that the diatom genome contains fewer chromosomes than the 33 predicted previously (39). Centromere sequences may be erroneously missing from the genome assembly. Additionally, some of the *P. tricornutum* chromosome-scale scaffolds lacking telomere-to-telomere assembly may not be individual chromosomes but rather partial chromosomes (Fig. 3). For example, the putative centromeres identified by ChIP-seq from chromosomes 24 and 29 were nearly identical (99%), and each of these two centromeres was positioned near a scaffold terminus lacking a telomere. Thus, chromosome-scale scaffolds 24 and 29 may be two arms of a single chromosome. In any case, the identification of centromeric DNA sequences will help to develop a better model of *P. tricornutum* genome organization.

Our findings suggest that *P. tricornutum* possesses small monocentric regional centromeres. ChIP-seq peaks were typically found only once per chromosome and corresponded to centromere sequences that were unique to each chromosome. This sequence variability precludes categorization as point centromeres, and the presence of a single peak (with rare exceptions, described below) rather than recruitment of CENH3 across the entire chromosome indicates the absence of a holocentric chromosome. The regional centromere structure is found in most eukaryotes, including the closest related organisms with identified centromeres: the protist *Plasmodium falciparum* and the red alga *C. merolae* (21–24). Both organisms have similarly sized centromeric DNA regions (~2–4 kb) and also share low-GC content as a characteristic of their centromeres: ~3% relative to the genome average GC of 21.8% in the case of *P. falciparum* (21, 22), and 48.4% relative to the genome average GC of 55% for *C. merolae* (24). Interestingly, *C. merolae*, which is the only other alga with well-characterized centromeres and the closest relation to *P. tricornutum* of the organisms studied, has centromeres with a GC content that is low only compared with the genome average and not intrinsically, similar to *P. tricornutum* centromeres.

Two *P. tricornutum* chromosomes, 2 and 8, appeared to deviate from the monocentric model by having two sequences identified by the ChIP-seq analysis. The regions adjacent to the centromeres on the chromosome scaffolds are unresolved DNA sequences, and both centromere regions contained long direct repeats. Thus, sample processing, sequencing, or assembly error could be responsible for the apparent duplication of the centromere on these chromosomes. Alternatively, these may be true centromeres that have simply been duplicated. The presence of a nearby retrotransposon may support this theory and could also confound PCR assays (*SI Appendix*, Fig. S5). Dicentric chromosomes have been noted in several organisms; however, typically only one of the centromeres is active and the other is inactivated (13, 49–52). The presence of two active centromeres typically leads to chromosomal breakage followed by either cell death or two functional monocentric chromosomes. Chromosomes with multiple functional centromeres have been identified. In human cells, two active centromeres were in close proximity, essentially behaving as a single centromere (53). In rice, recombinant centromeres were found to contain two repetitive arrays; both recruited CENH3, whereas an intervening sequence did not (54). Additionally, tricentric chromosomes were identified in wheat where one of the centromeres was large and presumably dominant, and co-occurring centromeres were smaller and weaker (55).

Like *P. tricornutum*, the diatom *T. pseudonana* can also use the yeast-derived *CEN6–ARSH4–HIS3* sequence to maintain episomes (37), which may suggest an overall similarity in DNA maintenance mechanisms. We analyzed the GC content of the *T. pseudonana* genome and found similar regions of low-GC content that were often found once per chromosome-scale scaffold (*SI Appendix*, Fig. S11). Thus, the ability of the yeast *CEN6–ARSH4–HIS3* sequence to support episomal maintenance in both species may be due to similar requirements for low-GC sequences to function as centromeres. It is remarkable that these diatoms may have such similar centromere features, to the degree that the same sequence can function as a centromere in both organisms, given the ancient evolutionary divergence of the centric and pennate diatom lineages (~90 Mya) (39) and the relatively rapid evolution of centromere sequences and structures observed for other groups of organisms (56, 57). Further CENH3 ChIP-seq experiments in *T. pseudonana* will enable centromere identification and comparison with *P. tricornutum*, including an examination of evolutionary implications.

Simple Centromere Requirements Permit Nuclear Maintenance of Nonnuclear DNA Sequences. In this study, by identifying characteristics of native diatom centromere sequences, we have uncovered a mechanism by which foreign DNA can become part of the nuclear DNA repertoire; nonnuclear DNA can act as a centromere, enabling stable maintenance as an extrachromosomal nuclear episome. Maintaining plasmids could expand the diatom's biochemical and physiological potential provided the new DNA acquired the necessary regulatory features over time and may also facilitate permanent integration into the native nuclear chromosomes through chromosomal rearrangements. We previously observed that DNA sequences from the yeast *S. cerevisiae* could enable episome maintenance in *P. tricornutum* (37, 38), and in this study, we confirmed that this sequence does, in fact, recruit the *P. tricornutum* centromeric histone protein CENH3. The recruitment of this centromere-specific histone protein and subsequent maintenance of the episome in diatoms suggests the foreign DNA sequence is using native diatom DNA replication machinery, essentially functioning as a diatom centromere. There are very few examples in eukaryotes of foreign

DNA recruiting host CENH3 to maintain a chromosome. Human centromeres have previously been shown to function in mouse chromosomes (58), and in a recent example, *Arabidopsis* centromeric repeats were shown to recruit human CENH3 and maintain chromosomes in human cells (59). In both cases, the chromosomes maintained by foreign DNA originally derived from chimeric host-donor DNA chromosomes followed by chromosomal breakage and/or rearrangement, resulting in smaller linear chromosomes or "mini-chromosomes." To our knowledge, there are no examples of immediate nuclear genome establishment (i.e., without chimeric intermediates) and maintenance in the host cell as a plasmid. This contrasts with bacteria, where DNA transfer between bacteria and subsequent plasmid establishment is quite common. Our results suggest that nonnuclear DNA can mimic diatom centromeres and, along with colocalized DNA, can immediately establish circular chromosomes in the diatom genome.

Establishment of centromeres in *P. tricornutum* is governed by an apparently simple rule: a small length of sequence (>500 bp) with a GC content less than ~33%. For foreign DNA sequences examined and for fragments further subdividing the native chromosome 25 centromere, 500 bp was typically a sufficient length for episome maintenance. Despite testing several shorter sequences from *P. tricornutum* chromosome 25 (*SI Appendix*, Fig. S8 and Table S5), only those >500 bp could maintain episomes, a pattern also observed for yeast-derived sequences (38). Notably, 500 bp is a particularly short sequence to enable centromere function compared with previously studied organisms with regional centromeres; most regional centromeres are reported to be thousands of base pairs in length, compared with the relatively small (~125 bp) point centromeres of some yeast species. Although ChIP-seq peaks for centromeres averaged 39% GC over the entire 2–5-kb sequence, each centromeric ChIP-seq peak contained within it a 500-bp region less than ~33% GC. Likewise, all foreign sequences that maintained episomes contained a 500-bp region with GC content beneath this threshold. The exceptions were three *Mycoplasma* sequences with GC less than 33% did not support episomal maintenance. We observed that these sequences had lower frequencies of 6 or more consecutive A+T bases, a pattern which persisted compared with a randomly generated set of 30% GC sequences. Although there may be something unique about these *Mycoplasma* DNA sequences generally, it is also possible that the frequency or spacing of longer contiguous A+T sequences or a similar signal may play an important functional role in diatom centromeric DNA, and sequences of <33% GC content usually, but not always, happen to contain these signals. Alternatively, certain sequences, perhaps consisting of high-GC content stretches, may actually interrupt centromere formation in otherwise low-GC content DNA sequences. Thus, rather than A+T stretches defining centromere function, the key feature may be a lack of interrupting sequences. Each of these hypotheses remains to be examined further in future studies.

The permissiveness of sequences that can function as centromeres in our organism may suggest that de novo centromere formation is quite common in diatoms. Two mechanisms of de novo centromere formation are of particular interest in this study: the acquisition of entirely new centromeres from nonnuclear DNA sources, and the potential presence of neocentromeres already present in the genome that can function as centromeres under certain circumstances. Transfer of novel centromere sequences from intracellular (i.e., chloroplast or mitochondrial genome) or extracellular DNA sources into the nucleus could alter the genome in multiple ways. Nonnuclear "centromeric" DNA sequences accompanied by other DNA could possibly form the basis of entire new chromosomes, similar to what we observed with episomes containing both native and foreign DNA sequences. Alternatively, foreign DNA sequences possessing centromeric DNA similarity could integrate into the nuclear chromosomes. Because multiple active centromeres typically lead to chromosomal instability, this could restructure the nuclear genome landscape by causing either a loss or gain in chromosome number as well as other rearrangements during the DNA repair process (reviewed in ref. 13). An alternative possibility is that after integration, one centromere sequence is silenced, which is one mechanism observed to prevent instability in dicentric chromosomes (60, 61). However, our observation of only one centromere-like DNA sequence per chromosome based on ChIP-seq experiments would suggest that this is not the case, at least in recent diatom evolutionary history. Neocentromeres are nuclear genome sequences distinct from centromere sequences but that can become active centromeres and recruit CENH3 when centromeres are inactivated or absent due to chromosomal rearrangements (reviewed in refs. 62–64). We did not observe any noticeable patterns of increased CENH3 recruitment outside of the centromere regions, and when we functionally tested several regions with elevated-background CENH3 recruitment, we found none of them functioned as centromeres. Our data would suggest that if there are in fact neocentromeres in this species, they do not recruit CENH3 before activation and would need to be discovered by altering the centromeric DNA region directly or through identification in genomes of aberrant phenotypes.

The observation that low-GC content sequences can act as *P. tricornutum* centromeres may help to explain the transfer of DNA from diatom endosymbiont or organelle genomes into the nuclear genome, which represent major sources of diatom nuclear DNA throughout evolutionary history (39–41). *P. tricornutum* chloroplast and mitochondrial genomes are low in GC content (32% average GC for the chloroplast, 35% average GC for the mitochondria), and we identified multiple sequences from each that could maintain nuclear episomes (*SI Appendix*, Table S4). Endosymbiotic gene transfer (EGT) from anciently acquired bacterial-derived organelles, namely the mitochondria (derived from a proteobacterium) and chloroplast (derived from a cyanobacterium), make up the majority of horizontally transferred genes found in eukaryotic genomes (42, 65, 66). DNA transfer from plastid and mitochondrial genomes to the nucleus, which include noncoding nuclear plastid DNA's (NUPTs) and nuclear mitochondrial DNAs (NUMTs), occurred at the time of endosymbiont acquisition and later in evolutionary time and likely occurs quite frequently in present time (65, 67–71). This is particularly true for photosynthetic organisms, where the majority of genes required for plastid function are actually housed in the nuclear genome, with a greatly reduced plastid genome from the original acquired state (72, 73). For example, in *Arabidopsis* nearly 18% of nuclear genes were found to be of cyanobacterial origin (74), whereas similar findings of 6–12% cyanobacterial-derived genes have been found to comprise unicellular-algae nuclear genomes (75–77). In these instances, DNA transferred to the recipient nucleus was already inside the host cell. Diatoms, like other stramenopiles, are the result of serial endosymbiotic events, though the precise details are still debated. However, it is generally accepted that the complicated series of whole-organism engulfments by many algal species, followed by transfer of both organelle and nuclear genes to the new host nucleus, has resulted in exceptional complex chimeric nuclear genomes. As new efforts attempt to solve these genomic puzzles, it is also important to understand exactly how these gene transfers occur, a subject that itself is not well understood. Thus, identifying mechanisms that could facilitate nuclear gene acquisition, such as the one proposed in this study, can shed light on algal diversity and evolution.

Although the majority of lateral gene transfer from bacteria to eukaryotes is thought to be via EGT, most unicellular algal genomes studied to date also contain a surprising amount of bacterial and viral DNA. This is particularly true for the diatom *P. tricornutum*, and another well-studied diatom *T. pseudonana*,

which are thought to possess a surprisingly high number of non-EGT horizontally acquired genes (39, 41). This assessment was based on the uniqueness of foreign genes possessed by each of these diatom species as well as gene-specific molecular phylogenies. Non-EGT–derived foreign genes in the *P. tricornutum* genome were found more frequently (by an order of magnitude) than in other free-living eukaryotes and were estimated to comprise about 5% of the nuclear genome, leading to the suggestion that horizontal transfer of bacterial DNA into diatoms may be quite common (39). Little is known about the acquisition of this DNA and how it ultimately integrates into the nuclear genome. Recent and/or transient endosymbiosis may be a possibility. Additionally, the discovery that diatoms are amenable to bacterial conjugation (37), the method of gene transfer used experimentally in the present study, provides a potential mechanism for exogenous DNA transfer, though this has yet to be demonstrated in a natural setting. Foreign DNA can also enter algal cells through viral infection (78, 79), an emerging area of algal research. The presence of this apparently non-EGT DNA in algal genomes further emphasizes the importance of understanding both how new genetic material gets into eukaryotes and how it stays there.

It is unclear why maintenance of foreign DNA in the form of episomes appears to be well tolerated in *P. tricornutum*. One possibility is that transfer of foreign DNA into diatoms, or intracellular transfer of previously acquired nonnuclear genetic material, is not common enough for a defense system to have evolved (such as the production of restriction enzymes in bacteria to destroy foreign DNA). In contrast to bacteria–bacteria DNA transfer, nonnative genes are unlikely to be expressed from a plasmid transferred to a diatom if they are of bacterial origin. Functional gene expression would only occur in the unlikely event that it acquired diatom transcriptional, translational, and subcellular localization signals through further modification. Thus, it is possible there was not strong selection to evolve defense mechanisms against foreign DNA because they were not detrimental to cell fitness and most events were entirely innocuous. If such permissiveness occurs for maintenance of DNA transferred through extracellular mechanisms, it is likely that it would also apply to DNA transferred to the nucleus intracellularly from organelles to the nucleus.

Conclusions

Identifying and characterizing centromeres is essential for understanding cellular biology, as these are critical features for stable DNA maintenance during cell division. These sequences can also advance synthetic biology through the development of new molecular tools. Here, we have used multiple approaches to characterize the centromeres of the diatom *P. tricornutum*. We found very simple sequence requirements for DNA to function as a centromere, namely a moderately low-GC content of <33% across a small region. Although most sequences with a GC content of <33% allowed episomal maintenance, a few sequences did not; these contained a lower frequency of contiguous A+T base stretches compared with functional sequences, indicating that more specific sequence characteristics could potentially play a role in centromere formation. Based on bioinformatic analyses, we predict that these features of centromere identity may be conserved in the distantly related diatom *T. pseudonana*. Although low-GC content has often been identified as a centromeric DNA feature, the diatom centromeres appear to be unique from many other eukaryotes in that they are not composed of repeat regions or other notable primary structures and that the functional centromere region may be quite small. We also show that these simple requirements allow foreign and nonnuclear DNA sequences with these characteristics to act as centromeres in diatoms, enabling establishment as extrachromosomal nuclear episomes. Diatoms possess nuclear genes acquired from many foreign DNA sources including viruses, bacteria, and other eukaryotes, including nuclear DNA acquired as a result of the ancient endosymbiotic acquisition of mitochondria and chloroplasts. Our findings present a host-permissive mechanism by which DNA derived from either external or intracellular genetic pools can persist in the diatom nucleus by using host replication and maintenance machinery. This may ultimately lead to gene integration into diatom genomes and subsequent evolutionary diversification.

Materials and Methods

A description of the strains used in this study, culturing conditions, and detailed explanations of the methods used can be found in *SI Appendix*. Briefly, we used *P. tricornutum* strain CCMP 632 (synonymous with the genome-sequenced strain CCMP2561) (39) and conducted ex-conjugant selection on phleomycin antibiotic. To conduct ChIP-seq and ChIP-qPCR assays, we constructed the plasmid pPtPBR1-YFP-CENH3 to express a YFP-CENH3 fusion protein. We confirmed protein expression and nuclear localization by confocal microscopy and Western blot analysis, respectively. Sample preparation and data analyses for the ChIP-seq experiments were conducted as previously described (80), using the Illumina sequencing platform. Sequences were deposited to the NCBI Sequence Read Archive (SRA) with accession no. PRJNA357294. For the *P. tricornutum* genomic library, purified PCR products of inserts were sequenced using Sanger DNA sequencing. Bacteria to diatom conjugations and episome maintenance analyses were conducted as previously described (37, 38).

ACKNOWLEDGMENTS. We thank John McCrow for many helpful discussions concerning bioinformatics and Sarah Smith for assistance with ChIP-seq. Funding for this work was provided by Gordon and Betty Moore Foundation Grants GBMF5007 (to P.D.W. and C.L.D.) and GBMF3828 and GBMF5006 (to A.E.A.), US Department of Energy Grant DE- SC0008593 (to A.E.A. and C.L.D.), and National Science Foundation Grants NSF-MCB-1129303 (to C.L.D.) and OCE-1136477 to (A.E.A.).

1. Pluta AF, Mackay AM, Ainsztein AM, Goldberg IG, Earnshaw WC (1995) The centromere: Hub of chromosomal activities. *Science* 270:1591–1594.
2. Westermann S, et al. (2003) Architecture of the budding yeast kinetochore reveals a conserved molecular core. *J Cell Biol* 163:215–222.
3. Cheeseman IM, Desai A (2008) Molecular architecture of the kinetochore-microtubule interface. *Nat Rev Mol Cell Biol* 9:33–46.
4. Earnshaw WC, et al. (2013) Esperanto for histones: CENP-A, not CenH3, is the centromeric histone H3 variant. *Chromosome Res* 21:101–106.
5. Westhorpe FG, Straight AF (2014) The centromere: Epigenetic control of chromosome segregation during mitosis. *Cold Spring Harb Perspect Biol* 7:a015818.
6. McKinley KL, Cheeseman IM (2016) The molecular basis for centromere identity and function. *Nat Rev Mol Cell Biol* 17:16–29.
7. Henikoff S, Ahmad K, Malik HS (2001) The centromere paradox: Stable inheritance with rapidly evolving DNA. *Science* 293:1098–1102.
8. Smith KM, Galazka JM, Phatale PA, Connolly LR, Freitag M (2012) Centromeres of filamentous fungi. *Chromosome Res* 20:635–656.
9. Cleveland DW, Mao Y, Sullivan KF (2003) Centromeres and kinetochores: From epigenetics to mitotic checkpoint signaling. *Cell* 112:407–421.
10. Cottarel G, Shero JH, Hieter P, Hegemann JH (1989) A 125-bp *CEN6* DNA fragment is sufficient for complete meiotic and mitotic centromere functions in *Saccharomyces cerevisiae*. *Trends Genet* 5:322–324.
11. Albertson DG, Thomson JN (1982) The kinetochores of *Caenorhabditis elegans*. *Chromosoma* 86:409–428.
12. Sullivan BA, Blower MD, Karpen GH (2001) Determining centromere identity: Cyclical stories and forking paths. *Nat Rev Genet* 2:584–596.
13. Torras-Llort M, Moreno-Moreno O, Azorín F (2009) Focus on the centre: The role of chromatin on the regulation of centromere identity and function. *EMBO J* 28:2337–2348.
14. Willard HF (1998) Centromeres: The missing link in the development of human artificial chromosomes. *Curr Opin Genet Dev* 8:219–225.
15. Tyler-Smith C, et al. (1993) Localization of DNA sequences required for human centromere function through an analysis of rearranged Y chromosomes. *Nat Genet* 5:368–375.
16. Ma J, Wing RA, Bennetzen JL, Jackson SA (2007) Plant centromere organization: A dynamic structure with conserved functions. *Trends Genet* 23:134–139.
17. Clarke L, Amstutz H, Fishel B, Carbon J (1986) Analysis of centromeric DNA in the fission yeast *Schizosaccharomyces pombe*. *Proc Natl Acad Sci USA* 83:8253–8257.
18. Nakaseko Y, Adachi Y, Funahashi S, Niwa O, Yanagida M (1986) Chromosome walking shows a highly homologous repetitive sequence present in all the centromere regions of fission yeast. *EMBO J* 5:1011–1021.
19. Kapoor S, Zhu L, Froyd C, Liu T, Rusche LN (2015) Regional centromeres in the yeast *Candida lusitaniae* lack pericentromeric heterochromatin. *Proc Natl Acad Sci USA* 112:12139–12144.

20. Lynch DB, Logue ME, Butler G, Wolfe KH (2010) Chromosomal G + C content evolution in yeasts: Systematic interspecies differences, and GC-poor troughs at centromeres. *Genome Biol Evol* 2:572–583.
21. Bowman S, et al. (1999) The complete nucleotide sequence of chromosome 3 of *Plasmodium falciparum*. *Nature* 400:532–538.
22. Iwanaga S, et al. (2010) Functional identification of the *Plasmodium* centromere and generation of a *Plasmodium* artificial chromosome. *Cell Host Microbe* 7:245–255.
23. Maruyama S, et al. (2008) Centromere structures highlighted by the 100%-complete *Cyanidioschyzon merolae* genome. *Plant Signal Behav* 3:140–141.
24. Kanesaki Y, Imamura S, Matsuzaki M, Tanaka K (2015) Identification of centromere regions in chromosomes of a unicellular red alga, *Cyanidioschyzon merolae*. *FEBS Lett* 589:1219–1224.
25. Yu W, Yau YY, Birchler JA (2016) Plant artificial chromosome technology and its potential application in genetic engineering. *Plant Biotechnol J* 14:1175–1182.
26. Kouprina N, Tomilin AN, Masumoto H, Earnshaw WC, Larionov V (2014) Human artificial chromosome-based gene delivery vectors for biomedicine and biotechnology. *Expert Opin Drug Deliv* 11:517–535.
27. Monaco AP, Larin Z (1994) YACs, BACs, PACs and MACs: Artificial chromosomes as research tools. *Trends Biotechnol* 12:280–286.
28. Coudreuse D (2009) Insights from synthetic yeasts. *Yeast* 33:483–492.
29. Murray AW, Szostak JW (1983) Construction of artificial chromosomes in yeast. *Nature* 305:189–193.
30. Harrington JJ, Van Bokkelen G, Mays RW, Gustashaw K, Willard HF (1997) Formation of de novo centromeres and construction of first-generation human artificial microchromosomes. *Nat Genet* 15:345–355.
31. Liu W, Yuan JS, Stewart CN, Jr (2013) Advanced genetic tools for plant biotechnology. *Nat Rev Genet* 14:781–793.
32. Iwanaga S, Kato T, Kaneko I, Yuda M (2012) Centromere plasmid: A new genetic tool for the study of *Plasmodium falciparum*. *PLoS One* 7:e33326.
33. Rochaix JD, van Dillewijn J, Rahire M (1984) Construction and characterization of autonomously replicating plasmids in the green unicellular alga *Chlamydomonas reinhardii*. *Cell* 36:925–931.
34. Lopez PJ, Desclés J, Allen AE, Bowler C (2005) Prospects in diatom research. *Curr Opin Biotechnol* 16:180–186.
35. Bozarth A, Maier UG, Zauner S (2009) Diatoms in biotechnology: Modern tools and applications. *Appl Microbiol Biotechnol* 82:195–201.
36. Fu W, Wichuk K, Brynjólfsson S (2015) Developing diatoms for value-added products: Challenges and opportunities. *N Biotechnol* 32:547–551.
37. Karas BJ, et al. (2015) Designer diatom episomes delivered by bacterial conjugation. *Nat Commun* 6:6925.
38. Diner RE, Bielinski VA, Dupont CL, Allen AE, Weyman PD (2016) Refinement of the diatom episome maintenance sequence and improvement of conjugation-based DNA delivery methods. *Front Bioeng Biotechnol* 4:65.
39. Bowler C, et al. (2008) The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* 456:239–244.
40. Armbrust EV (2009) The life of diatoms in the world's oceans. *Nature* 459:185–192.
41. Armbrust EV, et al. (2004) The genome of the diatom *Thalassiosira pseudonana*: Ecology, evolution, and metabolism. *Science* 306:79–86.
42. Timmis JN, Ayliffe MA, Huang CY, Martin W (2004) Endosymbiotic gene transfer: Organelle genomes forge eukaryotic chromosomes. *Nat Rev Genet* 5:123–135.
43. Clarke L, Carbon J (1980) Isolation of a yeast centromere and construction of functional small circular chromosomes. *Nature* 287:504–509.
44. Karas BJ, et al. (2013) Assembly of eukaryotic algal chromosomes in yeast. *J Biol Eng* 7:30.
45. Jacobs JD, et al. (1992) Characterization of two circular plasmids from the marine diatom *Cylindrotheca fusiformis*: Plasmids hybridize to chloroplast and nuclear DNA. *Mol Gen Genet* 233:302–310.
46. Hildebrand M, et al. (1992) Nucleotide sequence of diatom plasmids: Identification of open reading frames with similarity to site-specific recombinases. *Plant Mol Biol* 19: 759–770.
47. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410.
48. Bailey TL, et al. (2009) MEME Suite: Tools for motif discovery and searching. *Nucleic Acids Res* 37:W206–W208.
49. Stimpson KM, Matheny JE, Sullivan BA (2012) Dicentric chromosomes: Unique models to study centromere function and inactivation. *Chromosome Res* 20:595–605.
50. Neumann P, et al. (2012) Stretching the rules: Monocentric chromosomes with multiple centromere domains. *PLoS Genet* 8:e1002777.
51. Cuacos M, H Franklin FC, Heckmann S (2015) Atypical centromeres in plants-what they *can* tell us. *Front Plant Sci* 6:913.
52. Sato H, Masuda F, Takayama Y, Takahashi K, Saitoh S (2012) Epigenetic inactivation and subsequent heterochromatinization of a centromere stabilize dicentric chromosomes. *Curr Biol* 22:658–667.
53. Sullivan BA, Willard HF (1998) Stable dicentric X chromosomes with two functional centromeres. *Nat Genet* 20:227–228.
54. Wang G, Li H, Cheng Z, Jin W (2013) A novel translocation event leads to a recombinant stable chromosome with interrupted centromeric domains in rice. *Chromosoma* 122: 295–303.
55. Zhang W, Friebe B, Gill BS, Jiang J (2010) Centromere inactivation and epigenetic modifications of a plant chromosome with three functional centromeres. *Chromosoma* 119:553–563.
56. Malik HS, Henikoff S (2002) Conflict begets complexity: The evolution of centromeres. *Curr Opin Genet Dev* 12:711–718.
57. Malik HS, Henikoff S (2009) Major evolutionary transitions in centromere complexity. *Cell* 138:1067–1082.
58. Hadlaczky G, et al. (1991) Centromere formation in mouse cells cotransformed with human DNA and a dominant marker gene. *Proc Natl Acad Sci USA* 88:8106–8110.
59. Wada N, et al. (2016) Maintenance and function of a plant chromosome in human cells. *ACS Synth Biol* 6:301–310.
60. Agudo M, et al. (2000) A dicentric chromosome of *Drosophila melanogaster* showing alternate centromere inactivation. *Chromosoma* 109:190–196.
61. Faulkner NE, Vig B, Echeverri CJ, Wordeman L, Vallee RB (1998) Localization of motor-related proteins and associated complexes to active, but not inactive, centromeres. *Hum Mol Genet* 7:671–677.
62. Burrack LS, Berman J (2012) Neocentromeres and epigenetically inherited features of centromeres. *Chromosome Res* 20:607–619.
63. Stimpson KM, Sullivan BA (2010) Epigenomics of centromere assembly and function. *Curr Opin Cell Biol* 22:772–780.
64. Scott KC, Sullivan BA (2014) Neocentromeres: A place for everything and everything in its place. *Trends Genet* 30:66–74.
65. Martin W, Herrmann RG (1998) Gene transfer from organelles to the nucleus: How much, what happens, and Why? *Plant Physiol* 118:9–17.
66. Ku C, et al. (2015) Endosymbiotic origin and differential loss of eukaryotic genes. *Nature* 524:427–432.
67. Ju YS, et al.; ICGC Prostate Cancer Working Group; ICGC Bone Cancer Working Group; ICGC Breast Cancer Working Group (2015) Frequent somatic transfer of mitochondrial DNA into the nuclear genome of human cancer cells. *Genome Res* 25:814–824.
68. Huang CY, Ayliffe MA, Timmis JN (2003) Direct measurement of the transfer rate of chloroplast DNA into the nucleus. *Nature* 422:72–76.
69. Stegemann S, Hartmann S, Ruf S, Bock R (2003) High-frequency gene transfer from the chloroplast genome to the nucleus. *Proc Natl Acad Sci USA* 100:8828–8833.
70. Martin W (2003) Gene transfer from organelles to the nucleus: Frequent and in big chunks. *Proc Natl Acad Sci USA* 100:8612–8614.
71. Kleine T, Maier UG, Leister D (2009) DNA transfer from organelles to the nucleus: The idiosyncratic genetics of endosymbiosis. *Annu Rev Plant Biol* 60:115–138.
72. Archibald JM (2015) Genomic perspectives on the birth and spread of plastids. *Proc Natl Acad Sci USA* 112:10147–10153.
73. Archibald JM (2009) The puzzle of plastid evolution. *Curr Biol* 19:R81–R88.
74. Martin W, et al. (2002) Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc Natl Acad Sci USA* 99:12246–12251.
75. Moustafa A, Bhattacharya D (2008) PhyloSort: A user-friendly phylogenetic sorting tool and its application to estimating the cyanobacterial contribution to the nuclear genome of *Chlamydomonas*. *BMC Evol Biol* 8:6.
76. Price DC, et al. (2012) *Cyanophora paradoxa* genome elucidates origin of photosynthesis in algae and plants. *Science* 335:843–847.
77. Reyes-Prieto A, Hackett JD, Soares MB, Bonaldo MF, Bhattacharya D (2006) Cyanobacterial contribution to algal nuclear genomes is primarily limited to plastid functions. *Curr Biol* 16:2320–2325.
78. Suttle CA (2007) Marine viruses–Major players in the global ecosystem. *Nat Rev Microbiol* 5:801–812.
79. Brussaard CPD (2004) Viral control of phytoplankton populations–A review. *J Eukaryot Microbiol* 51:125–138.
80. Lin X, et al. (2012) Protocol: Chromatin immunoprecipitation (ChIP) methodology to investigate histone modifications in two model diatom species. *Plant Methods* 8:48.

Chapter 3, in full, is a reprint of the material as it appears in **Diner, RE**, Noddings, CM, Lian, NC, Kang, AK, McQuaid, JB, Jablanovic, J, Espinoza, JL, Nguyen, NA, Anzelmatti, MA, Jansson, J, Bielinski, VA., Karas, BJ, Dupont, CL, Allen, AE, and Weyman, PD (2017) Diatom Centromeres Suggest a Mechanism for Nuclear Gene Acquisition. *Proceedings of the National Academy of Sciences*, *114*(29), E6015-E6024. The dissertation author was the primary investigator and author of this material.

CHAPTER 4

High-resolution taxonomic grouping reveals interactions between pathogenic Vibrio species and the planktonic community

Synopsis

In this chapter I examined associations between pathogenic *Vibrio* bacteria along the San Diego coast and their abundant algal counterparts. I report the first quantitative survey of pathogenic *Vibrio* species in San Diego coastal waters, which are abundant during summer months and possess genes associated with human virulence. When examining the ecological interactions of these species, traditional grouping of diatoms at a high taxonomic level has led to conflicting reports of associations with pathogenic *Vibrio* species. I show that high-resolution taxonomic grouping at the genus level or lower, based on 18S amplicon sequencing, reveals specific interactions that may have important consequences for *Vibrio* ecology and human health, yet would have been overlooked in previous studies.

Introduction

Bacteria in the *Vibrio* genus occur naturally in coastal aquatic environments[1–5]. Many species can cause infection and represent an international human health concern. The disease cholera, caused by specific *Vibrio cholerae* serotypes, affects millions of people annually worldwide, causing thousands of deaths[6,7]. Cases have risen since 2005, most commonly in developing nations since and often acquired through contaminated drinking water. Most *Vibrio* infections in the United States are caused by two species, *Vibrio parahaemolyticus* and *Vibrio vulnificus*[8], resulting in approximately 80,000 illnesses annually and more than 100 deaths. These species primarily infect via ingestion of contaminated seafood, causing gastroenteritis, or wound infections which can lead to rapid necrosis and septicemia[9–13]. *V. vulnificus* has one of the highest mortality rates of any bacterial pathogen, with an estimated 50% mortality rate for U.S. infections[10,11,14]. These are also the causative agent of many "flesh eating bacteria" reports, particularly in the Southeast United States. *V. parahaemolyticus* is one of the most common bacterial causes of human shellfish poisoning[8,15,16]. At least 12 other species can infect humans, and many others are well-known animal pathogens, some a particular threat to aquaculture operations [8].

For many human pathogenic vibrios, the mechanisms of infection are poorly understood. In the best-studied species, *V. cholerae*, the lysogenic cholera toxin phage (the gene *ctxAB* encodes the toxin) is known to play a critical role in human infection[17], though bacteremia infection can occur without it. For other species, genomic comparisons have led to the discovery of genes more common in clinically isolated strains than environmental strains, and other genes putatively involved in pathogenicity such as the *vcgC* gene in *V. vulnificus*[18] and the thermostable hemolysin (*tdh*) gene[19] and TDH-related (*trh)*[20,21] gene in *V.*

parahaemolyticus. Genes associated with virulence and antibiotic resistance can be transferred horizontally between strains, often a result of biotic interactions including contact with other bacterial species and with the abundant marine polysaccharide chitin[22–24]. Thus, non-pathogenic strains and/or species of *Vibrio* have the ability to become pathogenic, and species can quickly acquire traits for environmental persistence or avoiding antibiotic susceptibility in human hosts[25,26]. A further concern is the emergence of several pandemic strains, notably the X serotype of *V. cholerae* and the 03:K6 serotype of *V. parahaemolyticus*[27–29].

Water temperature and salinity are major drivers of *Vibrio* species distribution (reviewed in [30]). *Vibrio* populations, along with human infections, are often highest during warm summer months[31–33]. As global seawater and air temperatures increase world-wide, the metabolic growth capacity and the geographic and temporal range of pathogenic *Vibrio* species is expected to expand[34–36], making *Vibrio*-related human illness an emerging health concern worldwide[34,35,37–40]. Most infectious *Vibrio* strains thrive in warm water temperatures (>20°C) and can also persist during unfavorable environmental conditions (e.g. <5°C) by entering a viable but nonculturable (VBNC) state[41]. *V. cholerae* infections and epidemics have been linked to environmental temperature increases on decadal scales, and have long been considered to be a case-study for understanding the link between environmental conditions and infectious diseases[37,40]. Perhaps as a result of global warming[42,43], infections have recently been reported in new geographic regions including Israel[44],Chile[45], Peru[46], Spain[47] the Baltic Sea[42], and the Pacific Northwest US[9]. Salinity also plays a strong role in species distribution. Among the most common pathogenic species, *V. vulnificus* and *V. cholerae* prefer fresher environments while *V. parahaemolyticus* is more halotolerant[30,48]. However, all species can be found in saline

environments and can be found simultaneously in the same coastal environments (present study).

Biotic interactions between prokaryotic and eukaryotic community members also play an important role in *Vibrio* ecology and pathogenicity, however there are many unresolved relationships. Vibrios are known to attach to and form biofilms on particles and living or dead eukaryotic organisms. Bacterial species that interact with vibrios in these environments may impact virulence and environmental persistence through horizontal gene transfer, population dynamics via viral infection, and growth through competition or cooperation. Perhaps the best-known example of *Vibrio* attachment to marine eukaryotes involves planktonic copepods. Likely due to their copious chitin production (they molt their chitinous exoskeletons frequently throughout their many life-cycle stages), they are often a characteristic feature of ecosystems with robust pathogenic *V. cholerae* populations, and attach to chitinous surfaces of both live and dead copepods in laboratory studies[49–54]. This attachment to chitin provides nutrition, serves as a substrate for biofilm formation and subsequent protection from environmental stressors and predation, and triggers a suite of cellular interactions triggering bacterial competition via the Type VI secretion system (T6SS) and natural competence, which may be the mechanism for how non-virulent populations become virulent[22,53,55–59].

Associations with algae are another likely important but understudied ecological link. Besides serving as the base of coastal food webs and fueling zooplankton (including copepod) abundances, *Vibrio* bacteria have been shown to attach to dead and living live algal cells[54,60]. Phytoplankton typically reside in the same coastal aquatic environments where *Vibrio* thrive,[60–64], generating organic carbon via photosynthesis that enables heterotrophic bacteria growth[65–67]. Blooms of *Vibrio* bacteria often succeed algal blooms, a common dynamic between marine

algae and bacteria; as algae blooms reach stationary phase and die, large amounts of organic carbon and nutrients are released into the environment for potential bacterial use[62,67,68]. Bacteria interact extensively with algae, exchanging vitamins and nutrients[69], algicides[70–72], and DNA[73,74] Additionally, some diatoms including *Thalassiosira* and *Cyclotella spp.* produce chitin filaments extruding from the cell, and in association with the cell wall[75–77], which may have the same physiological consequences as attachment to copepod chitin. Experimentally, Vibrios have been shown to attached to phytoplankton-produced chitin[60], though much remains to be discovered about the mechanisms of these interactions and whether and how often they may occur in natural ecosystems. Furthermore, as with copepods, it is unclear except at a relatively coarse taxonomic level (i.e. typically class, occasionally genus) what algae co-occur with which species of *Vibrio* bacteria.

While possible ecological relationships between these groups of organisms have been observed in multiple studies, conflicting results in different geographic regions complicate the ability to truly understand these relationships and their importance in terms of *Vibrio* ecology and physiology[30,78–84]. Many of these conflicting results likely stem from overly broad taxonomic groupings for both vibrios and other community members, which may mask true interactions. A meta-analysis by Takemura et al. found that across many studies, overall patterns regarding *Vibrio* organismal interactions are highly dependent on taxonomic resolution[30]. Many studies group all *Vibrio* species together, though it is known that particular pathogenic species have distinct preferences: for example, while vibrios as a genus have been positively associated with temperature and salinity[84–86], *V. cholerae* has a broad temperature range but prefers lower salinity. Co-occurring organisms with a low salinity niche would thus be overlooked in a genus-wide analysis, when in fact they could be positively correlated with

the pathogenic species causing the most infections. The frequent practice of grouping algae together based on chlorophyll A concentrations or on pigment quantification also masks species-specific distinctions, such as the ability to produce chitin. Previous research methodologies have limited the taxonomic resolution achievable in certain studies as visual classification can be challenging as well as time and labor intensive. However, high throughput DNA sequencing technologies paired with highly curated databases presents a new opportunity to examine these important interactions at a high-resolution and within a single dataset spanning a wide range of environmental conditions.

In this study, we elucidate ecological links between pathogenic *Vibrio* species and co-occurring prokaryotic and eukaryotic community members by combining *Vibrio* abundance quantified by digital droplet PCR and virulence-associated genes over a year of monthly sampling at three sites with high-resolution taxonomic composition derived from 16S and 18S amplicon sequencing. We observed distinct environmental preferences of each pathogenic species, all three of which were detected at all sites and in high abundances during the summer months, driven primarily by salinity. Concordantly, we identified genus and species-specific interactions between copepods, algae, and *Vibrio* species (some of which contained virulence genes), likely linked to shared environmental preferences and potentially additional biotic interactions. Additionally, we use shotgun metagenomic sequencing of isolate *Vibrio* communities to better characterize the diversity of *Vibrio* bacteria present at these sites, along with closely related and also potentially pathogenic species.

Materials and Methods

Environmental sampling and Vibrio *isolate culturing*

Monthly samples were collected from December 2015 to November 2016 at 3 sites in San Diego County: Los Peñasquitos Lagoon (LPL), the San Diego River (SDR), and the Tijuana River Estuary (TJ) (Figure 1a-d). For intra-site comparisons, two different locations at SDR (SDR1 and SDR2) and TJ (TJ1 and TJ2) were sampled, totaling 5 sampling locations. Temperature and salinity were measured between 12pm and 1pm using a YSI Pro 30 field instrument (YSI Inc.). Additionally, unfiltered water samples were collected in 4 L opaque bottles and processed in lab beginning no more than 2 hours after collection. These samples were kept in a cool area at roughly room temperature rather than at 4 °C to prevent a viable but not culturable (VBNC) state in *Vibrio* bacteria.

Water samples were gently filtered and flash-frozen in the lab for downstream processing. For chlorophyll A quantification, 10-100 ml samples were collected on GF/F filters (Whatman) and stored at -20 °C. Samples were later extracted in 90% acetone overnight and measured on a 10AU fluorometer (Turner), followed by addition of HCL and re-measurement to account for the chlorophyll A degradation product pheophytin[87]. For downstream nucleic acid extractions 50-400 ml samples were filtered onto 0.4 µm polycarbonate filters (Whatman) and stored at 80 °C until processing (details below).

Live *Vibrio* isolate communities were collected at 24 sampling points by filtering 10-100 µl of whole seawater onto 0.45 µm polycarbonate filters, which were transferred to CHROMagar Vibrio (CHROMagar Microbiology) plates and incubated overnight at 37 °C. For two additional samples, TJ1 and TJ2 June, 7 presumed *V. parahaemolyticus* colonies were

isolated and pooled for each sample. These communities (plate examples in Figure 2a-c) were resuspended in 1 ml of either LB broth (Amresco) or Zobell Marine Broth 2216 (HiMedia), depending on sampling salinity and frozen as 15% glycerol stocks at -80 °C. Half of each glycerol stock was pelleted and used for downstream DNA extraction.

DNA and RNA extraction and cDNA synthesis

Nucleic acids were extracted from filter samples using the NucleoMag Plant kit (Macherey-Nagel) for genomic DNA (gDNA) and the NucleoMag RNA kit (Macherey-Nagel) for RNA. Initial resuspension and vortexing of samples in lysis buffer was completed manually, and the remainder of the steps were completed using an epMotion liquid handling system (Eppendorf). RNA was reverse-transcribed into cDNA using the SuperScript III First-strand cDNA Synthesis Sytem (Invitrogen). gDNA was quantified using the Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen) and RNA using the Quant-iT RiboGreen RNA Assay Kit. Nucleic acid integrity was confirmed using an Agilent 2200 TapeStation (Agilent). Duplicate filters were extracted for all RNA and DNA samples with the exception of SDR1 December and April, for which two RNA but only one DNA sample was extracted. Additionally, gDNA was extracted from *Vibrio* isolate pellets using a DNeasy Blood and Tissue Kit (Qiagen), with subsequent quantification and quality control as described above.

Vibrio *digital droplet and end-point PCR*

Select pathogenic *Vibrio* species and virulence genes were quantified using the QX200 digital droplet PCR (ddPCR) System (BioRad), following the manufacturer's protocols and recommended reagents. Previously published assays based on qPCR were optimized for

ddPCR, including running temperature gradients for each target to establish optimum reaction temperature. Results from technical replicates were merged for analysis, and more than 19,000 droplets were measured per sample. Target-specific gBlocks (Integrated DNA Technologies) were used as positive controls for all ddPCR and end-point PCR targets.

Single copy-number gene targets for the species *V. parahaemolyticus, V. vulnificus,* and *V. cholerae* were quantified and used to approximate cell number per 100 ml of sample (Table 1). We targeted *toxR* for *V. parahaemolyticus*[88], *vvhA* for *V. vulnificus*[89], and *ompW* for *V. cholerae*[90]. We also quantified the virulence-associated *V. vulnificus* genes *vcgC*[91] and *pilF*[92]. We were unable to reliably quantify the *V. parahaemolyticus*-associated virulence genes *tdh*[93] and *trh*[94]. In lieu, we used traditional end-point PCR to screen all samples for the presence or absence of the gene using a TruFi DNA Polymerase Kit (Azura) and Vibrio gBlock 1 as a positive control. Samples containing *V. cholerae*, indicated by *ompW* detection, were also screened by traditional PCR for the virulence-associated *ctxA*[95] gene.

Library construction and sequencing

Amplicon libraries were constructed and sequenced using cDNA template for whole community samples to investigate microbial community composition. The V4-5 region of the 16S small subunit ribosomal RNA gene (SSU-rRNA) was targeted to characterize the prokaryotic community and plastid sequences using primers 515F-926R[96] (Table 2). The V9 region of the 18S rRNA gene was targeted for eukaryotic community composition using primers 1389F and 1510R[97] (Table 2).

TruFi DNA Polymerase Kits (Azura) were used for PCR amplifications, followed by 1.5% agarose gel confirmation of the correct amplicon size. Reactions were then purified using

AMPure XP beads (Beckman Coulter Life Sciences) for cleanup and size selection on the epMotion and quantified with the Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen). After pooling samples in equimolar amounts (~ 10 ng μl^{-1}) samples were sequenced using a dual-barcode index on an Illumina MiSeq platform at either the Institute for Genomic Medicine (IGM, University of California, San Diego) or at the UC Davis Genome Center (https://dnatech.genomecenter.ucdavis.edu/) on an with 250-bp paired-end reads for the 16S amplicon and 150-bp paired-end reads for the 18S amplicon. To account for predicted low abundances of Vibrio ASVs in the samples, 16S libraries were sequenced in small runs (~60-85 samples) to obtain a greater sequencing depth.

For shotgun metagenomics libraries, gDNA extracted from *Vibrio* isolate pellets was fragmented to 400bp on an E210 Sonicator (Covaris). Sequencing libraries were prepared using the NEBNext Ultra II DNA Library Prep Kit (New England Biotechnologies), combined into 2 equimolar concentration pools of 13 samples each and sequenced on an Illumina HiSeq4000 at the UC Davis Genome Center with 250-bp paired-end reads.

Bioinformatic and statistical analyses

Amplicon sequence analysis

Demultiplexed sequences were analyzed using the Qiime2[98] pipeline and additional analyses and visualizations were conducted using the R package phyloseq[99]. Sequences were quality filtered, chimeric sequences were removed, and exact amplicon sequence variants (ASVs)[100] were defined using dada2[101]. Replicate samples were merged using the "qiime feature-table group" function. Feature classifiers were trained using the specific primers for

each amplicon and taxonomy was assigned using Silva[102] version 132 for bacterial and archaeal 16S sequences and PR2[103] for 18S sequences.

Alpha and beta diversity metrics for community composition were calculated and statistically compared among samples and groups using Qiime2. Alpha rare-faction plots were generated to determine sampling depth, which was 152K for 37K for 16S and 18S sequences, respectively. All 60 samples were included for the 16S analyses, and 59 were included for 18S analysis as this enabled ASV abundance and alpha-diversity saturation. For alpha diversity, richness determined by Faith's Phylogenentic Diversity[104] (Faith PD) and evenness were compared across group and site (Krustkal-Wallis test), and were also examined for relationships with environmental variables (Spearman's rank correlations). Beta diversity was calculated using both Bray-Curtis dissimilarity and Weighted Unirac methods for comparison, and dissimilarity was calculated between categorical variables site and month by group and also in pair-wise comparisons across all communities using a PERMANOVA test. Interactive and static emperor PCOA plots were generated using Qiime2.

Spearman rank correlations were calculated to explore relationships between environmental variables, *Vibrio* quantification data, and relative abundance of groups of interest in the amplicon sequencing. Data subsets were generated in phyloseq to characterize relative abundance of specific groups. For example, 16S sequences may belong to bacteria, archaea, or eukaryotes (particularly chloroplast sequences). For the purpose of exploring relative abundances of bacteria and archaea in the 16S community, eukaryotic sequences would be excluded from the analysis. Correlations were visualized as correlograms using the corrplot package in R[105].

Shotgun metagenomic analysis of Vibrio *isolate communities*

Quality control was performed by checking sequence quality using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), trimming poor-quality sequences using Trimmomatic[106], then re-assessing quality with a subsequent FastQC analysis. Trimmed reads were then assigned taxonomy using Metaphlan2[107], and visualized using Graphlan[108]. For some samples, raw reads were assembled using metaSPAdes[109], and assembly quality was assessed and compared between samples using the tool QUAST[110]. Sample binning was attempted using Vizbin[111], but provided limited practical information, likely because the organisms are so similar in genome content. The metagenome assembly from the LPL May site was annotated by open reading frame using the J. Craig Venter Institute's RAP pipeline, which incorporates blast queries of sequences against the NCBI database, to explore genes of interest in the context of the assembled contigs. Additional future work with these samples will attempt to further characterize genomic potential of these isolates, and potentially re-assemble bacterial genomes or mobile genetic elements.

Results

Environmental niche of pathogenic Vibrio species

All sites had similar temperatures which peaked in the Summer months of June and July (Figure 1). SDR1, SDR2, and LPL had similar annual salinity profiles, with the exception of low salinities at LPL from March to May, when lagoon closure (an annual event) presumably led to the accumulation of freshwater. The Tijuana Rivers sites had considerable higher salinity throughout the year.

Pathogenic *Vibrio* species were detected at temperatures below 15 °C but were not abundant (i.e. <216 copies/100ml) below 20 °C (Figure 3a-c). Sampling temperatures ranged from 13.2-33 °C. We observed a wide range of salinity, from a nearly freshwater sample of 2.6 ppt to hypersaline conditions >40 ppt. Temperature and salinity were positively correlated across all sites (Figure 3d), likely because of evaporation during high temperatures and a corresponding lack of rainfall in the warmest months. *V. cholerae* were recorded in higher quantities than other species but were abundant only during three low salinity months (March through May) at Los Peñasquitos Lagoon (Figure 3c). More than 280,000 target gene copies per 100 ml were measured in May, while low abundance (<300 copies/100 ml) characterized LPL months with high salinity and all other sites. Spearman correlations, accordingly, revealed a negative association between *V. cholerae* and salinity, but no association with temperature. We observed a broader distribution with *V. vulnificus*, which also exhibited a negative correlation with salinity but was found across more sites and samples. *V. vulnificus* was highest during the summer months and detected at all sites, reaching highest abundances of >13,000 copies/100 ml at both SDR sites during May. There was no correlation with temperature though notably none were detected below 20°C. *V. parahaemolyticus* was detected in 80% (50/60) of samples, and in contrast to the other two species showed no correlation with salinity but a positive correlation with temperature (Figure 3d). The highest concentrations were generally observed at the Tijuana river estuary sites at salinities above 30 ppt, with the highest recorded concentration of ~33k copies/100 ml found at a salinity of 39.9 ppt. Pathogenic species abundance across all samples was significantly associated with chlorophyll A concentrations, a proxy for total algal abundance, (Figure 3d): *Vibrio parahaemolyticus* and *Vibrio vulnificus*

with a positive association and *Vibrio cholerae* with a negative association/ Chlorophyll A was also significantly positively associated with temperature, but not salinity.

Quantifying and detecting virulence-associated genes

We quantified two genes associated with *V. vulnificus* virulence, pilF and *vcgC*, using digital droplet PCR. Copies of *vcvG* were detected at all sites in low abundances, with a peak of 872 copies/100 ml at SDR2 during March (Figure 4a). Copies of *pilF* were most abundant at the San Diego River sites (Figure 4b) during April, May and June, reaching >7,000 copies/100 ml in May at both sites. It was also detected at LPL during these months at slightly lower abundances, and not at all in the TJ sites, though only 4 were tested.

Of the 20 sites where *V. vulnificus* was detected based on the *vvhA* ddPCR assay, virulence genes were detected at 50% of the sites (Figure 4c). 15% of samples were positive for both *pilF* and *vcgC*, while 30% were positive for *pilF* alone and 5% positive for only *vcgC*. The ratio of *pilF* and *vcgC* genes to *vvha*, which putatively represents the number of virulence gene copies present per copy of the species, ranged from 0-2 (Figure 4d).

We determined the presence and absence of the virulence-associated genes *ctxA* for *V. cholerae* and tdh and trh for *V. parahaemolyticus* using traditional PCR (Figure 5). Of the samples that tested positive for *V. cholerae*, only one sample yielded a PCR product using primers designed to detect an 87-bp amplicon (Figure 5a). Notably, this sample (LPL May) had by far the highest abundance of *V. cholerae* of all samples (Figure 2c). The size of the amplicon, however, was considerably larger at ~600-700 bp. Attempts to sequence the amplicon after PCR purification were unsuccessful.

Of the 50 samples that tested positive for *V. parahaemolyticus*, 48 were tested for the presence of *tdh* and *trh*. Of these, 16 (33%) produced bands for the *trh* assay and 13 (27%) produced bands for the *tdh* assay. Similar to the *ctxA* assay, none of the bands were the same size of the predicted amplicon and positive control sequence, which were ~200 bp for both targets (Table 1). Bands in the *trh* assay were typically 650-1000 bp, often with multiple bands amplified. For the *tdh* assay, many samples had single bands at ~850bp, though other samples had multiple bands of varying sizes.

Vibrio *community composition*

The 16S amplicon sequencing revealed 116 distinct *Vibrio* ASVs. Few of the Vibrio ASV's were assigned to a particular species (Figure 6A). ASV's classified as *Vibrio cholerae* were predominant members of the *Vibrio* community at LPL during March and May, corresponding to peak detection of *V. cholerae* by the ddPCR assays (Figure 2c).

Across all samples, *Vibrio* bacteria comprised 0.04%-5.3% of the bacterial-archaeal 16S community. (Figure 6b). The highest community percentages were during July at SDR2 and September at TJ2, the latter occurring simultaneously with a high abundance of *Vibrio parahaemolyticus* (Figure 2a) *Vibrio* community relative abundance did not, however, significantly correspond to any environmental variables or *Vibrio* marker genes (Figure 2d).

Metagenomics of Vibrio *isolate communities*

To gain a better understanding of *Vibrio* diversity in these communities beyond the 16S amplicon, we isolated vibrios using CHROMagar *Vibrio* media and conducted shotgun metagenomic sequencing on these enriched communities. While this approach is not

quantitative because some species may grow faster than others during the culturing stage, it provides a fine-scale genomics-based characterization of *Vibrio* species and strain diversity.

The isolate communities were predominantly comprised of *Vibrio* species (70-100%) (Figure 7a). One exception was the TJ2 site in February, where sequences mapped primarily to *Vibrio* phage vB_VpaM_MAR (85%) with only 15% mapping to *Vibrio* bacteria (Figure 7b, Table 3). We detected 26 unique *Vibrio* species across all samples (Figure 3), as well as the closely related bacterial species *Grimontia hollisae* (formerly classified *V. hollisae*), also a potential human pathogen, which was abundant in multiple samples (Table 3). Other less abundant species detected (up to 7% community abundance, but typically less than 2%) belonged to the Gammaproteobacterial genera *Aeromonas*, *Pseudomonas*, *Shewanella*, and *Photobacterium*.

Vibrio species were detected in all samples profiled. The species *V. parahaemolyticus* and *V. sp. Ex25* were present in 100% of the samples (Table 3), including some samples where *V. parahaemolyticus* was not detected in the ddPCR assay: LPL February, SDR1 February, and SDR2 August. *V. vulnificus and V. cholerae* were detected in 74% and 61% of the samples, respectively. For these species, there was a link between community presence and relative abundance based on metagenomics and ddPCR quantification. Samples where the species were detected, particularly in high relative abundance, corresponded to higher ddPCR detection levels while low or no abundance within a metagenomic sample usually corresponded with low or no ddPCR detection (Table 3). Notably, although we did not quantify the species *V. alginolyticus* via ddPCR, it was detected in 96% of the isolate samples.

For the LPL site in May, which contained all three pathogenic *Vibrio* species detected by ddPCR (Figure 3) we annotated the assembled metagenome and identified genes of interest

in each species. Several genes with a potential role in human health were observed, including genes that may be involved in antibiotic resistance or virulence, though it is unclear whether many of the putative toxins would be involved in human, animal, or microbial virulence (Supplemental Tables 1-9). We also identified putative chitinase genes and chemotaxis genes present in all three pathogenic species, potentially originating from multiple strains of the same species, that may be involved in interactions with other community members.

Prokaryotic diversity, community composition, and association with vibrios

Both site and month significantly impacted 16S community beta-diversity (i.e. whether communities were dissimilar to each other) using both Bray-Curtis and Weighted Unifrac analyses ($p<0.001$ for site and month under both tests), however, clear groupings were not apparent in PCoA plots based on these factors alone (Table 4, Figure 8a-d). Site and the environmental variables temperature, salinity, and chlorophyll A significantly influenced site-specific (i.e. alpha diversity) species richness, though sampling month had no significant effect (Table 4). Chlorophyll A was the only factor that significantly influenced community evenness ($p<0.01$).

In pairwise comparisons, sites and months were typically similar in community composition to nearby sites and temporally close months, respectively. For locations with two sampling sites, those sites were similar to each other ($p>0.05$) in both the Bray-Curtis and Weighted Unifrac analyses, but different from all other sites (Table 5). For example, the SDR1 site was significantly different from every other site except SDR2 (Table 5). Intra-site species richness was similar between the LPL and SDR sites, which differed from the TJ sites, and

species evenness only differed significantly between LPL and TJ1. Months temporally close to each other had more similar 16S microbial communities (Table 6).

Nearly 32K ASV's were classified for 16S sequences using dada2, which were reduced to ~29K ASV's after removing eukaryotic, mitochondrial, and chloroplast sequences. Despite the large number of ASV's, a few major classes dominated bacterial community composition across most sites, including Gammaproteobacteria (encompassing vibrios and related bacteria), Bacteroidia, and Alphaproteobacteria (Figure 9). During some months, the LPL and SDR sites had sizeable populations of Oxyphotobacteria, a class of Cyanobacteria. Other prominent classes include Campylobacteria (of the phylum Epsilonbacteraeota), and Verrucomicrobia. We also plotted relative abundance of the top 15 bacterial genera and observed that these genera covered the majority of the bacterial relative abundance across all samples.

To investigate associations between pathogenic *Vibrio* species and dominant bacterial community members, we used spearman rank correlations to compare ddPCR results with the top 10 bacterial classes and top 15 genera (Figure 9). At the class taxonomic level, *V. parahaemolyticus* and *V. vulnificus* were positively associated with *Verrucomicrobiae*, a class containing mostly bacteria isolated from freshwater, soil, and human feces. *V. vulnificus, V. cholerae,* and the *V. vulnificus* virulence-associated gene *pilF* were positively associated with *Oxyphotobacteria* and negatively associated with Camphylobacteria. This pattern is similar to the negative association between these three marker genes and salinity. Additionally, *V. cholerae* was negatively associated with Bacteroidia and Kirimatiellae.

Individual genera often exhibited associations that were not evident at the level of class. For example, the Alphaproteobacteria and Gammaproteobacteria classes had no significant associations with any of the species or virulence genes assayed (Figure 9). Among the most

abundant Alphaproteobacteria genera, however, a Rhodobacteraceae strain HIMB11 was negatively associated with *V. vulnificus,* and a SAR11 Clade 1a strain was negatively associated with both *V. vulnificus* and *V. parahaemolyticus*. Likewise, the common copiotrophic Gammaproteobacterium *Alteromonas* was positively linked to *V. parahaemolyticus*, while several other Gammaproteobacteria including *Glaciecola*, *Marinibacter*, and uncultured members of the Thiotrichaceae and Thiomicrospiraceae families were negatively associated with multiple marker genes. Additionally, an abundant *Synechococcus* strain of Cyanobacteria was abundant and positively associated with *V. parahaemolyticus.*

Eukaryotic community diversity and community composition

Like the 16S communities, site and month had a significant ($p<0.05$) effect on community dissimilarity between samples in both the Bray-Curtis and Weighted Unifrac analyses (Table 4). We observed a negative correlation between chlorophyll A concentrations and species richness and evenness. Temperature and salinity did not have an effect on species richness, and temperature impacted species evenness while salinity did not. Sites were different from all other sites except those at the same location, and species richness and evenness were significantly different when comparing the LPL and SDR sites to the TJ sites (Table 5). Also, like the 16S communities, months temporally close to each other had more similar 18S microbial communities (Table 7).

While common bacterial groups dominated 16S communities, 18S communities were considerably more diverse, with many rare species. For example, up to >50% of the taxonomic Classes at some sites were present at <5% abundance (Figure 10). Plotting the top 15 most

abundant community members by Class (Figure 10) gives a clearer view of the most common groups, which include numerous algal, animal, and non-photosynthetic protist groups.

Diatoms were the most common Eukaryotic organisms, comprising ~28% of the total 18S reads (Table 8). They were abundant community members at the LPL and SDR sites during some months, but were particularly abundant at the TJ sites, where they frequently represented >75% of the 18S reads. The second most abundant group contained unicellular ciliates of the class Spirotrichea. These two classes had positive (Bacillariophyta) and negative (Spirotrichea) associations with both *V. parahaemolyticus* and temperature. Other abundant community members included the photosynthetic Cryptophyceae and the chitin-producing animals of the class Arthropoda, which were negatively correlated with the *V. vulnificus* virulence-associated gene *pilF*. Additionally, other ciliates and photosynthetic protists including Chrysophyceae, Dinophyceae, and Mamiellophyceae, were present, with both Chrysophyceae and Mamiellophyceae exhibiting a positive correlation with *V. parahaemolyticus* and a negative association with *V. cholerae*.

Like the bacterial-archaeal community, correlations between individual genera and *Vibrio* species did not always reflect trends of the class level. The most dominant 18S genus, the diatom *Chaetoceros*, had a positive correlation with *V. parahaemolyticus*, temperature and salinity, while another abundant diatom genus, *Thalassiosira*, had no correlation with *V. parahaemolyticus* but instead had a positive association with *V. vulnificus* and a negative association with temperature and salinity (Figure 10, Figure 11). Meanwhile, *Cyclotella* diatoms and a poorly characterized group of raphid pennate diatoms had no significant correlations in the comparisons. Investigating correlations with the top 15 diatom genera

(Figure 10), we observed similarly diverse correlations, however many genera were either positively or negatively correlated with both temperature and salinity.

Beyond the diatom genera, poorly characterized Strombidiida ciliates in the Spirotricheae class were negatively correlated with *V. parahaemolyticus, V. vulnificus*, *pilF*, temperature, and chlorophyll A, while other ciliates in the Mesodinium genera had no associations with marker genes or environmental variables. Additionally, photosynthetic Teleaulax algae in Crysophyte class were positively associated with *V. parahaemolyticus.*

We also examined genera-specific correlations with chitin-producing arthropods. Arthropods were one of the most abundant classes making up about 5% of the 18S reads (Table 8), though no single genus was among the 10 most abundant overall. Copepods were the most abundant arthropods, including the genera *Pseudodiaptomas, Canuella, Tigriopis, Sinocalanus,* and *Cyclops*. With the exception of *Pseudodiaptomas*, these genera were significantly negatively associated with salinity and had no relationship to temperature (Figure 11). Additionally, many of them had positive associations with *V. vulnificus, V. cholerae*, and *pilF*. Notably, these genera also made up a small percentage (<1%) of the total 18S reads, because they were typically abundant only at specific sites or months. We also observed many positive correlations between the copepod and diatom genera discussed above.

Discussion

We observed distinct environmental niches amongst *V. cholerae, V. vulnificus*, and *V. parahaemolyticus* related to salinity and temperature (Figure 3a-c). While these environmental factors are known to drive *Vibrio* distribution[30] many studies focus on individual species or on the *Vibrio* genus as a whole, preventing observed shifts in the abundance of multiple species as

the environment and community changes. We detected all three species at all sites, suggesting either a continuous presence at all times, occasionally in undetectable concentrations, or a temporal residence in the sediments or in a VBNC state, until conditions become ideal for proliferation in the water column (i.e. the "everything is everywhere, but the environment selects" ecological theory[112]). Peak abundances among the three species varied with salinity (Figure 3a-c), with *V. cholerae* highest at the lowest salinity sites (~2-5 ppt), *V. vulnificus* highest at moderate salinities (~17-25 ppt), and *V. parahaemolyticus* highest at high salinity sites (>30 ppt). The distribution patterns of these three species are also linked to site, with *V. cholerae* most abundant at LPL, *V. vulnificus* most common at the SDR sites, and *V. parahaemolyticus* most abundant at the TJ sites. It is unclear whether this is because those sites happened to present an ideal ecological niche for a particular species or strain at a given time, or if biotic interactions limit concentrations of species that would otherwise be abundant.

V. cholerae and *V. vulnificus* were significantly associated with low salinity and had no significant association with temperature (Figure 3b,c). They were, however, generally found above 20 °C, a temperature regarded as a threshold for human concern regarding *Vibrio* infections (Blackwell 2008, Oliver 2015). We observed peak abundances during the warm summer months, reflecting previously reported temperature-associated seasonality of these species, though typically a month or two before the peak temperature. As temperature was inversely correlated with salinity, this may represent an intermediate condition where temperatures are warm, but salinity isn't too high. While *V. cholerae* has been reported in high salinity conditions, it is most common in low salinities, and is well-known to contaminate drinking water (reviewed in Takemura et al., 2014). Likewise, *V. vulnificus* grows poorly if at all at salinities higher than 25 ppt and environmentally prefers salinities in the range of 10-18

ppt[113,114] (Kaspar and Tamplin 1993, Oliver 2015). It has previously been suggested that salinity constraints on the *V. vulnificus* niche may be alleviated by increased temperatures[115]. We observed low abundances of *V. vulnificus* at some high salinity, high temperature sights, so generally this trend wasn't apparent in our data, however biotic interactions likely play a role in the observed ecological niche of *Vibrio* species[116].

Conversely, *V. parahaemolyticus* abundance was significantly associated with high temperatures, but not with salinity. This supports *V.* parahaemolyticus as a more halotolerant, though not necessarily halophilic, species. It also corresponds with the observation from the Takemura et al. meta-analysis which found that, in contrast to *V. cholerae*, *V. parahaemolyticus* was more broadly spread out in terms of salinity across a range of 3-35ppt, with a warmer, more narrow temperature range. We detected high abundances of *V. parahaemolyticus* at extremely high salinities (>40 ppt)(Figure 3a), which was out of the reported range in the meta-analysis and for other prior studies we examined. Potentially the strains we observed possess adaptations for life at extremely high salinities. This is supported by the finding that the fundamental ecological niche of many *Vibrio* species, particularly in terms of salinity, is often larger than realistic environmental conditions[116]. The peak abundance of *V. parahaemolyticus* (at TJ2 in September) occurred at 25 °C, a moderate temperature at which other pathogenic species were found at other sites, but with a salinity of 40 ppt. Thus, our findings may suggest that *V. parahaemolyticus*, through tolerating high salinity, can take advantage of fortuitous environmental conditions such as high temperature to proliferate when the other pathogenic species cannot due to salinity constraints. Since this peak abundance was observed during a moderate rather than a high temperature month, other factors may contribute to proliferation. Potentially fortuitous environmental conditions include temporal events such as the demise of

an algal bloom and concurrent release of organic matter, or a storm even bringing fresh nutrients into the site. Alternatively, species-specific predation, parasitism, viral infection, or other environmental conditions may play a role in controlling *V. vulnificus* and *V. cholerae* populations at the high-salinity sites. Our study, however, generally supports the well-established temperature and salinity preferences previously observed in these species in an entirely different geographic region.

The potentially pathogenic species we examined contained a high percentage of virulence genes, though PCR amplification results suggest the virulence genes of San Diego populations may be divergent from other regions (Figure 4, Figure 5). Along the North Carolina Coast, Williams et al. found that 5.3% of the *V. vulnificus* examined possessed the *vcgC* gene and 1.9% of *V. parahaemolyticus* harbored one or both of the virulence genes *tdh* and *trh*[117]. In samples collected from coastal Alaska, *tdh* and *trh* positive strains were isolated from 19% and 26% of the samples collected, respectively[93]. The *vcgC* gene was detected in 20% of our samples, and 33% of the samples were positive for either *tdh*, *trh*, or both. In our *tdh*/*trh* assays detecting presence or absence of the marker gene, the band size was larger than expected, and multiple bands were present at times. This could potentially be caused by binding to a non-specific or non-*V. parahaemolyticus* target, or the *trh* and *tdh* genes may have diverged in San Diego populations. These are known to be highly variable DNA regions[21,118] which may also impact virulence potential. Further examination of these sequences in local strains may shed light on the evolution of these virulence genes and potential functional consequences.

Our study presents the first quantification and ecological analysis of these species in the Southern California region, an area that due to warm coastal seawater conditions and high recreational use. Coastal southern California counties are among the most densely populated

areas in the United States, totaling more than 15 million people with rapidly increasing residential populations. These areas are also popular worldwide as beach tourism destinations. Furthermore, shellfish harvesting occurs recreationally, and historically has represented a common food source for these coastal communities[119]. The popularly eaten Pacific Oyster is found throughout southern California in addition to the native Olympia Oyster[120] and may provide a sustainable food source for the rapidly growing population. Prior to the present study, there was limited information regarding the presence of potentially pathogenic species in Southern California, though water temperature conditions fall well within their known environmental range. A 1987 study isolated pathogenic *V. vulnificus* from sediment in Mission Bay, San Diego[121]. In a more recent study[122], several pathogenic *Vibrio* species including *V. vulnificus, V. parahaemolyticus* (virulent and non-virulent strains), and *V. cholerae* were detected, though not quantified, via PCR in Orange County, California and on Catalina Island off-shore. Additional genotyping conducted for the study proposed here identified *V. parahaemolyticus* and *V. vulnificus* at more sites, including several in San Diego County.

We detected high concentrations of *V. cholerae* at the LPL site, which are regionally concerning as there is no *Vibrio* monitoring system in place in the San Diego region. Only one sample tested positive for the *ctxA* gene, which amplified a larger DNA sequence than expected, however while *ctxA* is required for traditional cholerae infections, *V. cholerae* can infect immunocompromised humans without these virulence genes. Furthermore, *Vibrio* communities lacking these virulence genes can acquire them rapidly via viral infection or other horizontal gene transfer events. Low salinity resulting in high cholerae concentrations occurred in the LPL while the lagoon was closed naturally due to wave and tidal action and sandbar movement,

which consequently leads to a buildup of freshwater in the lagoon from surrounding residential and commercial developments. This is an annual event, and eventually the lagoon is manually dredged and drained in to the surrounding coastal ocean following permit application procedures through the Army Corps of Engineers. While this dredging and draining is intended to prevent anoxic conditions hazardous to human and wildlife health, it may also prevent high concentrations of *V. cholerae* bacteria. However, this comes at the risk of exposing local recreational ocean users to high bacterial concentrations released from the lagoon during dredging. Additionally, we found high concentrations of *V. parahaemolyticus* bacteria at both TJ sites. The Tijuana River Estuary has a long history of persistent pollution, and the surrounding waters of Imperial beach and beyond are often closed due to sewage overflows and high fecal coliform bacterial contamination. These local findings have important policy implications which demonstrate a need for future, potentially continuous, *Vibrio* sampling in the region.

Using shotgun metagenomics, we characterized a diverse *Vibrio* community containing far more species than suggested by 16S data alone, and also identified genes associated with *Vibrio* interspecies interactions and potentially virulence. A similar community-characterization approach was employed in Jesser et al. where the HSP60 amplicon was used to determine relative abundance of different *Vibrio* community members, also revealing a much more diverse community than the 16S alone[123]. We hope to replicate this approach in future studies as our method for determining community composition is not amenable to determining relative abundance. Our samples displayed a wide-range of *Vibrio* and closely-related bacteria (Figure 7a). Of the 26 *Vibrio* species we identified across all samples, several were potential human (e.g. *V. alginolyticus, V. furnissii*, *V. metschnikovii*) or animal pathogens (i.e. *V. harveyi*

and *V. anguillarum*). We also identified what appears to be an active viral infection by the *Vibrio* phage vB_VpaM_MAR at TJ2 in February (Figure 7b). Based on the species identified at this site, it is likely that either *V. parahaemolyticus* or *Vibrio* strain EX25 is the intended target, though it is also possible that reads from the actual infected species have fallen below the detection limit as the result of the infection. This provides interesting information for future studies investigating *Vibrio* viral infections, which we may be able to answer using these preserved *Vibrio* communities. Community characterization of the LPL site in May (Figure 7c), where the highest abundance of *V. cholerae* was detected, confirms the presence of all pathogenic species quantified in this study at the same time, an interesting finding from a human health perspective. The skew towards *V. parahaemolyticus* sequences in the metagenomic sequences compared to the higher abundance of other species in the ddPCR data likely reflects a growth advantage of *V. parahaemolyticus* on CHROMagar *Vibrio* plates during sample isolation.

Our study also focuses on elucidating links between these pathogenic *Vibrio* populations and their surrounding planktonic community at a high-resolution taxonomic level. This is useful for understanding *Vibrio* ecology and distribution, virulence and environmental persistence, and to potentially develop specific bioindicators of *Vibrio* species abundance. We address this by focusing on the three *Vibrio* species most relevant to human health. Since *Vibrio* species occupy distinct ecological niches and unique physiological capabilities, exploring all vibrios as a genus may result in misleading associations. In one example, Turner et al. found that while vibrios were negatively correlated copepods in a particular size fraction (63-200um), the pathogenic species *V. parahaemolyticus* and *V. vulnificus* were actually positively associated with copepods. Our study allows us to explore correlations at both coarse and fine taxonomic

levels, and we found that the relative abundance of *Vibrio* genus bacteria based on 16S amplicon sequencing had no significant relationship to any *Vibrio* species or virulence gene-specific marker gene, temperature, salinity, or chlorophyll A (Figure 3d). However, individual species had many significant associations at broad and specific taxonomic levels.

We observed this "broad grouping" phenomenon while investigating the *Vibrio*-associated prokaryotic microbial community. While the same broad groups of bacteria were present and abundant across most sites, particular genera exhibited more particular associations. These associations are important when trying to understand interspecific competitive or symbiotic interactions. For example, the class Oxyphotobacteria was negatively associated with *V. vulnificus*, *V. cholerae*, and *pilF*. But only one was among the most abundant bacterial genera, *Syneccococus* (with a closest hit of sp. *CC9902)*, which was negatively associated with *V. parahaemolyticus* and temperature. It is possible that many, less abundant genera in Oxyphotobacteria are responsible for the observed associations, but reliance on this class level alone would miss the *Synecoccocus* relationship. One laboratory study examining the response of *Synechococcus* sp. WH8102 to co-culture with *V. parahaemolyticus* found significant transcriptional changes, including evidence of possible phosphate stress and utilization of specific nitrogen sources. Our study puts laboratory studies like this one in a relevant ecological context. Additionally, genera that were positively (e.g. *Alteromonas* or *Thiomicrospoaceae* sp.) or negatively (e.g. *Marinobacterium*, *Glacielcola*, *SAR11 Clade 1a*) associated with pathogenic species, relationships that were unclear at the class level, may be good candidates for investigating symbiotic or competitive interactions, particularly in the context of the broader community.

Our study also addresses the frequently overlooked eukaryotic community, focusing on phytoplankton and copepods. Diatoms are an example of example of an algal group frequently positively associated with *Vibrio* concentrations [78,80,124] but analyzed as a single group. The genetic distance, however, between the model pennate diatom *Phaeodactylum tricornutum* and the model centric diatom *Thalassiosira pseudonana* is equivalent to the difference between a human and a fish[125] (Bowler). This taxonomic distinction is important as particular algal species are known to produce chitin, host distinct microbial communities[126], release unique DOM[127–129], and be alternatively susceptible or immune to bacterial attack[70,71,130,131]. We observed that diatoms as a group, based on the class Bacilliarophyta, were positively associated only with *V. parahaemolyticus*, temperature, salinity, and chlorophyll A (Figure 10b), and not with the other species measured or *Vibrio* ASV relative abundance. This would seem to suggest that diatoms prefer warm, high salinity conditions (as does *V. parahaemolyticus* in our study), possibly leading to the finding that other species are not associated with diatoms. However, looking at the more resolved genus level, this affect appears to be driven by the most abundant diatom genus *Chaetoceros*. In particular, multiple ASVs of one species most closely related to *Chaetoceros pumilus* comprised the majority of *Chaetoceros* diatoms.

A potentially important but poorly studied interaction involves *Vibrio* associations with chitin-producing diatoms. Experimentally, vibrios have been shown to attached to phytoplankton-produced chitin[60], though much remains to be discovered about the mechanisms of these interactions and whether and how often they may occur in natural ecosystems. We observed 2 genera of chitin-producing diatoms in our samples: *Thalassiosira* and *Cyclotella*. In contrast to *Chaetoceros*, *Thalassiosira* was actually negatively associated with temperature and salinity, and positively associated with *V. vulnificus*. The primary contributing species is most

closely related to the model diatom species *Thalassiosira pseudonana*, which has been observed to have algicidal interactions with chitinase-producing bacterium in laboratory studies[72]. Another potentially chitin-producing diatom was also observed- a species most closely related to *Cyclotella striata* (the model species *Cyclotella cryptica* is known to produce chitin, but ability is unknown for this species[75]. Though this genus had no significant correlations with any *Vibrio* markers or environmental variables It was very abundant (>25% of the diatom community) at the two sampling points with the highest *V. vulnificus* concentrations: SDR1 and SDR2 in April. Other factors such as grazing pressures on the diatom or bacteria may influence these broad-scale correlations. Our findings demonstrate that assuming broad associations between diatoms as a group and *Vibrio* species based on any of these examples would lend itself to a false association and preclude potentially important investigations into species or even strain-specific environmental preferences and biotic interactions. Our findings also established an ecologically relevant framework for future laboratory experiments. In particular, for future studies investigating interactions between the low-salinity *Vibrio* species *V. vulnificus* and *V. cholerae* and chitin producing eukaryotes, our study points to *Thalassiosira* and *Cyclotella* spp., both diatom genera with extensive genetic tools available[73,132–136], as excellent model organisms

Of particular interest in *Vibrio* ecology is the interaction between pathogenic species and planktonic copepods. Given the importance of this relationship relatively little is known about what specific type of copepods different species of pathogenic vibrios attach to in the environment, and whether they are living or dead. In a classic study by Huq et al., *V. cholerae* 01 and non-01 serovars were found to attach to living but not dead *Acartia tonsa*, *Eurytemora affinis*, and *Scottolana spp*. copepods from natural samples[137]. A laboratory study investigating

these same copepod species found that *V. cholerae* preferentially attached to *Acartia tonsa* copepods over *Eurytemora affinis*, and that individual *V. cholerae* strains exhibited different attachment efficiencies[138]. While this study was also conducted with live copepods, another study investigated an O1 *V. cholerae* serovar (strain N16961) and two non-O1/O139 *V. cholerae* isolates, finding that all three strains preferentially attached to dead, rather than living, *Tigriopus californicus* copepods, as well as dinoflagellates[54]. It is unclear whether this difference is due to experimental methodology, the copepod species, or the *Vibrio* strains. Associations in field samples are rare and inconclusive: one study found no association between *V. cholerae* and co-occurring *Diaptomus* and *Cyclops* genera copepods[139], while others have reported qualitative associations in field samples, but quantitative significant relationships are poorly understood [50,52]. This suggests that fine-scale taxonomic distinctions likely play a role in *Vibrio*-zooplankton interactions.

In our environmental samples, we observed positive correlations between pathogenic *Vibrio* species and individual copepod genera. Sequences from copepods, and arthropods in general, were far less abundant than diatoms and other eukaryotic groups (Figure 11, Table 8). The most abundant copepod was the species *Pseudodiaptamus inopinus*, an invasive species originating in Asia[140,141], which was not significantly associated with any *Vibrio* species across all samples (Figure 11d), but was highly abundant during the months where the highest levels of *V. cholerae* and *V. vulnificus* were detected at LPL and the SDR sites (Figure 3b,c). Other abundant copepods were the Harpacticoid genera *Canuella* and *Tigriopus*, which were both positively associated with *V. vulnificus* and the virulence-associated gene *pilF*. In laboratory studies, the type IV pilus (containing the *pilF* subunit) has been shown to be involved in chitin attachment to vibrios [142]. *Tigriopus* was also positively associated with *V. cholerae*. Though we

could not obtain species-specific taxonomic resolution for *Tigriopus* in our sample, it is a well-established laboratory model genus with gene-silencing capabilities and full or partially assembled genomes for several species including *T. japonicus*[143], *T. californicus*[144,145], and *T. kingsejongens*[146]. Thus, *Tigriopus* and *Canuella* spp. may be good candidate genera for future laboratory studies involving ecologically relevant *Vibrio*-plankton interactions.

Beyond interactions between individual *Vibrio* species and groups of planktonic organisms, these planktonic organisms also interact with each other. For example, diatoms are a known food source for copepods[147–149] , however, in some instances they have been shown to negatively impact copepod reproduction[150]. This highlights the need to taxonomically characterize the communities and define their interactions where pathogenic vibrios are abundant. Using the diatom and copepod groupings above, we found that several diatom and copepod species were significantly associated with each other. For example, *Canuella* copepods were negatively associated with *Chaetoceros*, but positively associated with *Thalassiosira* and other diatoms genera, while *Tigriopus* was significantly associated only with the less common diatoms. Further analysis of these community networks will provide a more detailed framework of potential interactions.

Conclusion

Our study quantifies for the first time potentially pathogenic *Vibrio* populations in Southern California, finding abundant populations that conform to previously observed temperature and salinity niches as well as additional potentially pathogenic species. High abundances in previously unstudied areas with high potential for human exposure, along with

the detection of multiple genes associated with human infection, suggest that future sampling and risk modelling for these areas may be appropriate.

We also characterized the microbial and eukaryotic communities co-occurring with these individual *Vibrio* species and identified relationships that are apparent at high taxonomic resolution but masked based on the broader groupings applied in previous studies. Ultimately, our characterization of *Vibrio* communities, other community members, and their shared environmental preferences can be used to develop and test new hypotheses about the role of the environment and biotic interactions in *Vibrio* persistence, proliferation, and disease risk.

Acknowledgements

I would like to acknowledge the Southern California Coastal Water Research project for their assistance with environmental sampling and ddPCR analysis. We would also like to thank Rachel Noble, Brett Froelich, and Kelsey Jesser of UNC for their assistance and mentorship throughout this project. We gratefully acknowledge our funding sources, the NSF, Gordon and Betty Moore Foundation, US Department of Energy, and the UCSD Center for Microbiome innovation.

Works Cited:

1. Farmer III, J. J. *The Family Vibrionaceae*. *Prokaryotes* **6**, (Springer, 2006).
2. Joseph, S. W., Colwell, R. R. & Kaper, J. B. Vibrio parahaemolyticus and related halophilic Vibrios. *Crit. Rev. Microbiol.* **10**, 77–124 (1982).
3. Colwell, R. R., Kaper, J. & Joseph, S. W. Vibrio cholerae, Vibrio parahaemolyticus, and other vibrios: occurence and distribution in Chesapeake Bay. *Science* **198**, 394–396 (1977).

4. Thompson, J. R., Randa, M. a, Marcelino, L. a, Tomita-Mitchell, A., Lim, E. & Polz, M. F. Diversity and dynamics of a north atlantic coastal Vibrio community. *Appl Env. Microbiol* **70**, 4103–4110 (2004).

5. Johnson, C. N., Bowers, J. C., Griffitt, K. J., Molina, V., Clostio, R. W., Pei, S., Laws, E., Paranjpye, R. N., Strom, M. S., Chen, A., Hasan, N. a., Huq, A., Noriea, N. F., Grimes, D. J. & Colwell, R. R. Ecology of vibrio parahaemolyticus and vibrio vulnificus in the coastal and estuarine waters of Louisiana, Maryland, Mississippi, and Washington (United States). *Appl. Environ. Microbiol.* **78**, 7249–7257 (2012).

6. World Health Organization. World Health Organization- Cholera. *Cholera* **3**, 69–74 (2019).

7. Centers for Disease Control and Prevention. CDC: Cholera- Vibrio cholerae infection. *Cholera- Vibrio cholerae infection* (2019). Available at: https://www.cdc.gov/cholera/index.html. (Accessed: 1st May 2019)

8. Centers for Disease Control and Prevention. CDC: Vibrio Species Causing Vibriosis. *CDC: Vibrio Species Causing Vibriosis* (2019). Available at: https://www.cdc.gov/vibrio/index.html. (Accessed: 1st May 2019)

9. Daniels, N. & MacKinnon, L. Vibrio parahaemolyticus infections in the United States, 1973–1998. *J. Infect. Dis.* **181**, 1661–1666 (2000).

10. Waldor, M. K., Ali, A., Baker-Austin, C., Alam, M., Oliver, J. D., Martinez-Urtaza, J. & Qadri, F. Vibrio spp. infections. *Nat. Rev. Dis. Prim.* **4**, (2018).

11. Jones, M. K. & Oliver, J. D. Vibrio vulnificus: Disease and Pathogenesis. *Infect. Immun.* **77**, 1723–1733 (2009).

12. Su, Y. C. & Liu, C. Vibrio parahaemolyticus: A concern of seafood safety. *Food Microbiol.* **24**, 549–558 (2007).

13. Oliver, J. D. Wound infections caused by Vibrio vulnificus and other marine bacteria. *Epidemiol. Infect.* **133**, 383–391 (2005).

14. Horseman, M. A. & Surani, S. A comprehensive review of Vibrio vulnificus: An important cause of severe sepsis and skin and soft-tissue infection. *Int. J. Infect. Dis.* **15**, e157–e166 (2011).

15. Scallan, E., Hoekstra, R. M., Angulo, F. J., Tauxe, R. V., Widdowson, M. A., Roy, S. L., Jones, J. L. & Griffin, P. M. Foodborne illness acquired in the United States-Major pathogens. *Emerg. Infect. Dis.* **17**, 7–15 (2011).

16. Letchumanan, V., Chan, K. & Lee, L. Vibrio parahaemolyticus : a review on the pathogenesis , prevalence , and advance molecular identification techniques. *Front. Microbiol.* **5**, 1–13 (2014).

17. Karaolis, D. K. R., Somara, S., Maneval, D. R., Johnson, J. A. & Kaper, J. B. A bacteriophage encoding a pathogenicity island, a type-IV pilus and a phage receptor in cholera bacteria. *Nature* **399**, 375–379 (1999).

18. Rosche, T. M., Yano, Y. & Oliver, J. D. A rapid and simple PCR analysis indicates there are two subgroups of Vibrio vulnificus which correlate with clinical or environmental isolation. *Microbiol. Immunol.* **49**, 381–9 (2005).

19. Honda, T., Iida, T. The pathogenicity of Vibrio parahaemolyticus and the role of thermostable direct haemolysin and related haemolysins. *Rev. Med. Microbiol.* **4**, 106–113 (1993).

20. Honda, T., Abad-Lapuebla, M. A., Ni, Y., Yamamoto, K. & Miwatani, T. Characterization of a new thermostable direct haemolysin produced by a Kanagawa-phenomenon-negative clinical isolate of Vibrio parahaemolyticus. *J. Gen. Microbiol.* **137**, 253–259 (1991).

21. Kishishita, M., Matsuoka, N., Kumagai, K., Yamasaki, S., Takeda, Y., Nishibuchi2, M., Shirai, H., Ito, H., Hirayama, T., Nakamoto, Y., Nakabayashi, N., Kumagai, K., Takeda, Y. & Nishibuchi, M. Sequence Variation in the Thermostable Direct Hemolysin- Related Hemolysin (trh) Gene of Vibrio parahaemolyticus Our previous molecular epidemiologic study with gene probes (. *Appl. Environ. Microbiol.* **58**, 2449–2457 (1992).

22. Meibom, K. L. Chitin Induces Natural Competence in Vibrio cholerae. *Science* **310**, 1824–1827 (2005)

23. Bartlett, Douglas H; Azam, F. Chitin, Cholera, and Competence. *Science* **310**, 1775–1777 (2005).

24. Le Roux, F. & Blokesch, M. Eco-evolutionary Dynamics Linked to Horizontal Gene Transfer in Vibrios. *Annu. Rev. Microbiol* **72**, 89–110 (2018).

25. Beaber, J. W., Hochhut, B. & Waldor, M. K. SOS response promotes horizontal dissemination of antibiotic resistance genes. *Nature* **427**, 72–74 (2004).

26. Cordero, O. X., Wildschutte, H., Kirkup, B., Proehl, S., Ngo, L., Hussain, F., Le Roux, F., Mincer, T. & Polz, M. F. Ecological populations of bacteria act as socially cohesive units of antibiotic production and resistance. *Science* **337**, 1228–31 (2012).

27. Okuda, J., Ishibashi, M., Hayakawa, E., Nishino, T., Takeda, Y., Mukhopadhyay, A. K., Garg, S., Bhattacharya, S. K., Nair, G. B. & Nishibuchi, M. Emergence of a unique O3:K6 clone of Vibrio parahaemolyticus in Calcutta, India, and isolation of strains from the same clonal group from Southeast Asian travelers arriving in Japan. *J. Clin. Microbiol.* **35**, 3150–3155 (1997).

28. González-Escalona, N., Gavilan, R. G., Brown, E. W. & Martinez-Urtaza, J.

Transoceanic Spreading of Pathogenic Strains of Vibrio parahaemolyticus with Distinctive Genetic Signatures in the recA Gene. *PLoS One* **10**, e0117485 (2015).

29. Martinez-Urtaza, J., van Aerle, R., Abanto, M., Haendiges, J., Myers, R. A., Trinanes, J., Baker-Austin, C. & Gonzalez-Escalona, N. Genomic Variation and Evolution of Vibrio parahaemolyticus ST36 over the Course of a Transcontinental Epidemic Expansion. *mBio 8e01425-17.* **8**, 1–17 (2017).

30. Takemura, A. F., Chien, D. M. & Polz, M. F. Associations and dynamics of vibrionaceae in the environment, from the genus to the population level. *Front. Microbiol.* **5**, 1–26 (2014).

31. Centers for Disease Control and Prevention. National Enteric Disease Surveillance: COVIS Annual Summary, 2011-2013. 1–10 (2013).

32. Froelich, B. a., Ayrapetyan, M., Fowler, P., Oliver, J. D. & Noble, R. T. Development of a Matrix Tool for the Prediction of Vibrio Species in Oysters Harvested from North Carolina. *Appl. Environ. Microbiol.* **81**, 1111–1119 (2015).

33. Baker-Austin, C., Trinanes, J., Gonzalez-Escalona, N. & Martinez-Urtaza, J. Non-Cholera Vibrios: The Microbial Barometer of Climate Change. *Trends Microbiol.* **25**, 76–84 (2017).

34. Burge, C. a, Mark Eakin, C., Friedman, C. S., Froelich, B., Hershberger, P. K., Hofmann, E. E., Petes, L. E., Prager, K. C., Weil, E., Willis, B. L., Ford, S. E. & Harvell, C. D. Climate change influences on marine infectious diseases: implications for management and society. *Ann. Rev. Mar. Sci.* **6**, 249–77 (2014).

35. Levy, S. Warming trend, how climate shapes vibrio ecology. *Environ. Health Perspect.* **275**, 729b–0 (2015).

36. Vezzulli, L., Brettar, I., Pezzati, E., Reid, P. C., Colwell, R. R., Höfle, M. G. & Pruzzo, C. Long-term effects of ocean warming on the prokaryotic community: evidence from the vibrios. *ISME J.* **6**, 21–30 (2012).

37. Lipp, E. K., Huq, A. & Colwell, R. R. Effects of Global Climate on Infectious Disease : the Cholera Model Effects of Global Climate on Infectious Disease : the Cholera Model. *Clin. Microbiol. Rev.* **15**, 757–770 (2002).

38. Lindgren, E., Andersson, Y., Suk, J. E., Sudre, B. & Semenza, J. C. Monitoring EU Emerging Infectious Disease Risk Due to Climate Change. *Science* **336**, 418–419 (2012).

39. Harvell, C. D., Kim, K., Burkholder, J. M., Colwell, R. R., Epstein, P. R., Grimes, D. J., Hofmann, E. E., Lipp, E. K., Osterhaus, a D., Overstreet, R. M., Porter, J. W., Smith, G. W. & Vasta, G. R. Emerging marine diseases--climate links and anthropogenic factors. *Science* **285**, 1505–1510 (1999).

40. Colwell, R. R. Global climate and infectious disease: the cholera paradigm. *Science* **274**, 2025–2031 (1996).

41. Colwell, R. R., Brayton, P. R., Grimes, D. J., Roszak, D. B., Huq, S. A. & Palmer, L. M. Viable but non-culturable vibriocholerae and related pathogens in the environment : implification for release of genetically engineerered microorganisms. *Nature* **3**, 817–820 (1985).

42. Baker-Austin, C., Trinanes, J. a., Taylor, N. G. H., Hartnell, R., Siitonen, A. & Martinez-Urtaza, J. Emerging Vibrio risk at high latitudes in response to ocean warming. *Nat. Clim. Chang.* **3**, 73–77 (2012).

43. Vezzulli, L., Grande, C., Reid, P. C., Hélaouët, P., Edwards, M., Höfle, M. G., Brettar, I., Colwell, R. R. & Pruzzo, C. Climate influence on *Vibrio* and associated human diseases during the past half-century in the coastal North Atlantic. *Proc. Natl. Acad. Sci.* 201609157 (2016). doi:10.1073/pnas.1609157113

44. Paz, S., Bisharat, N., Paz, E., Kidar, O. & Cohen, D. Climate change and the emergence of Vibrio vulnificus disease in Israel. *Environ. Res.* **103**, 390–396 (2007).

45. González-Escalona, N., Cachicas, V., Acevedo, C., Rioseco, M. L., Vergara, J. A., Cabello, F., Romero, J. & Espejo, R. T. Vibrio parahaemolyticus diarrhea, Chile, 1998 and 2004. *Emerg. Infect. Dis.* **11**, 129–31 (2005).

46. Martinez-Urtaza, J., Huapaya, B., Gavilan, R. G., Blanco-Abad, V., Ansede-Bermejo, J., Cadarso-Suarez, C., Figueiras, A. & Trinanes, J. Emergence of Asiatic Vibrio diseases in South America in phase with El Niño. *Epidemiology* **19**, 829–837 (2008).

47. Baker-Austin, C., Stockley, L., Rangdale, R. & Martinez-Urtaza, J. Environmental occurrence and clinical impact of Vibrio vulnificus and Vibrio parahaemolyticus: A European perspective. *Environ. Microbiol. Rep.* **2**, 7–18 (2010).

48. Urakawa, H.; Rivera, I. Aquatic Environment. in *The Biology of the Vibrios* 175–189 (American Society of Microbiology, 2006).

49. Vezzulli, L., Pruzzo, C., Huq, A. & Colwell, R. R. Environmental reservoirs of Vibrio cholerae and their role in cholera. *Environ. Microbiol. Rep.* **2**, 27–33 (2010).

50. Huq, A., Sack, R. B., Nizam, A., Longini, I. M., Nair, G. B., Ali, A., Morris, J. G., Khan, M. N. H., Siddique, A. K., Yunus, M., Albert, M. J., Sack, D. A. & Colwell, R. R. Critical Factors Influencing the Occurrence of Vibrio cholerae in the Environment of Bangladesh. **71**, 4645–4654 (2005).

51. Huq, a, Small, E. B., West, P. a, Huq, M. I. & Colwell, R. R. Ecological relationships between Vibrio cholerae and planktonic crustacean Ecological Relationships Between Vibrio cholerae and Planktonic Crustacean Copepods. **45**, 275–283 (1983).

52. Lizárraga-Partida, M. L., Mendez-Gómez, E., Rivas-Montaño, A. M., Vargas-Hernández, E., Portillo-López, A., González-Ramírez, A. R., Huq, A. & Colwell, R. R. Association of Vibrio cholerae with plankton in coastal areas of Mexico. *Environ. Microbiol.* **11**, 201–8 (2009).

53. Erken, M., Lutz, C. & Mcdougald, D. Interactions of Vibrio spp . with Zooplankton. *Microbiol. Spectr.* **3**, VE-0003-2014 (2015).

54. Mueller, R. S., McDougald, D., Cusumano, D., Sodhi, N., Kjelleberg, S., Azam, F. & Bartlett, D. H. Vibrio cholerae strains possess multiple strategies for abiotic and biotic surface colonization. *J. Bacteriol.* **189**, 5348–5360 (2007).

55. Sun, Y., Bernardy, E. E., Hammer, B. K. & Miyashiro, T. Competence and natural transformation in vibrios. *Mol. Microbiol.* **89**, 583–595 (2013).

56. Lutz, C., Erken, M., Noorian, P., Sun, S. & McDougald, D. Environmental reservoirs and mechanisms of persistence of Vibrio cholerae. *Front. Microbiol.* **4**, 1–15 (2013).

57. Hunt, D. E., Gevers, D., Vahora, N. M. & Polz, M. F. Conservation of the chitin utilization pathway in the Vibrionaceae. *Appl. Environ. Microbiol.* **74**, 44–51 (2008).

58. Borgeaud, S., Metzger, L. C., Scrignari, T. & Blokesch, M. The type VI secretion system of Vibrio cholerae fosters horizontal gene transfer. **084112**, (2013).

59. Antonova, E. S. & Hammer, B. K. Genetics of Natural Competence in Vibrio cholerae and other Vibrios. *Microbiol. Spectr.* **3**, 1–18 (2015).

60. Frischkorn, K. R., Stojanovski, A. & Paranjpye, R. Vibrio parahaemolyticus type IV pili mediate interactions with diatom-derived chitin and point to an unexplored mechanism of environmental persistence. *Environ. Microbiol.* **15**, 1416–27 (2013).

61. Amin, S. A., Parker, M. S. & Armbrust, E. V. Interactions between diatoms and bacteria. *Microbiol. Mol. Biol. Rev.* **76**, 667–84 (2012).

62. Kouzuma, A. & Watanabe, K. Exploring the potential of algae/bacteria interactions. *Curr. Opin. Biotechnol.* **33**, 125–129 (2015).

63. Grossart, H. & Simon, M. Interactions of planktonic algae and bacteria: effects on algal growth and organic matter dynamics. *Aquat. Microb. Ecol.* **47**, 163–176 (2007).

64. Goecke, F., Labes, a, Wiese, J. & Imhoff, J. Chemical interactions between marine macroalgae and bacteria. *Mar. Ecol. Prog. Ser.* **409**, 267–299 (2010).

65. Azam, F. & Malfatti, F. Microbial structuring of marine ecosystems. *Nat. Rev. Microbiol.* **5**, 782–91 (2007).

66. Worden, A. Z., Follows, M. J., Giovannoni, S. J., Wilken, S., Zimmerman, a. E. &

Keeling, P. J. Rethinking the marine carbon cycle: Factoring in the multifarious lifestyles of microbes. *Science* **347**, 1257594–1257594 (2015).

67. Grossart, H. P., Czub, G. & Simon, M. Algae-bacteria interactions and their effects on aggregation and organic matter flux in the sea. *Environ. Microbiol.* **8**, 1074–1084 (2006).

68. Worden, A. Z., Seidel, M., Smriga, S., Wick, A., Malfatti, F., Bartlett, D. & Azam, F. Trophic regulation of Vibrio cholerae in coastal marine waters. *Environ. Microbiol.* **8**, 21–29 (2006).

69. Bertrand, E. M., McCrow, J. P., Moustafa, A., Zheng, H., McQuaid, J. B., Delmont, T. O., Post, A. F., Sipler, R. E., Spackeen, J. L., Xu, K., Bronk, D. A., Hutchins, D. A. & Allen, A. E. Phytoplankton–bacterial interactions mediate micronutrient colimitation at the coastal Antarctic sea ice edge. *Proc. Natl. Acad. Sci.* **112**, 9938–9943 (2015).

70. Paul, C. & Pohnert, G. Interactions of the algicidal bacterium Kordia algicida with diatoms: regulated protease excretion for specific algal lysis. *PLoS One* **6**, e21032 (2011).

71. Su, J. Q., Yang, X. R., Zheng, T. L., Tian, Y., Jiao, N. Z., Cai, L. Z. & Hong, H. S. Isolation and characterization of a marine algicidal bacterium against the toxic dinoflagellate Alexandrium tamarense. *Harmful Algae* **6**, 799–810 (2007).

72. Li, Y., Lei, X., Zhu, H., Zhang, H., Guan, C., Chen, Z., Zheng, W., Fu, L. & Zheng, T. Chitinase producing bacteria with direct algicidal activity on marine diatoms. *Sci. Rep.* **6**, 21984 (2016).

73. Karas, B. J., Diner, R. E., Lefebvre, S. C., McQuaid, J., Phillips, A. P. R., Noddings, C. M., Brunson, J. K., Valas, R. E., Deerinck, T. J., Jablanovic, J., Gillard, J. T. F., Beeri, K., Ellisman, M. H., Glass, J. I., Hutchison III, C. a., Smith, H. O., Venter, J. C., Allen, A. E., Dupont, C. L. & Weyman, P. D. Designer diatom episomes delivered by bacterial conjugation. *Nat. Commun.* **6**, 6925 (2015).

74. Diner, R. E., Noddings, C. M., Lian, N. C., Kang, A. K., McQuaid, J. B., Jablanovic, J., Espinoza, J. L., Nguyen, N. A., Anzelmatti, M. A., Jansson, J., Bielinski, V. A., Karas, B. J., Dupont, C. L., Allen, A. E. & Weyman, P. D. Diatom centromeres suggest a mechanism for nuclear DNA acquisition. *Proc. Natl. Acad. Sci.* 201700764 (2017). doi:10.1073/pnas.1700764114

75. Durkin, C. a., Mock, T. & Armbrust, E. V. Chitin in diatoms and its association with the cell wall. *Eukaryot. Cell* **8**, 1038–50 (2009).

76. Brunner, E., Richthammer, P., Ehrlich, H., Paasch, S., Simon, P., Ueberlein, S., van Pée, K.-H. & van Pée, K.-H. Chitin-based organic networks: an integral part of cell wall biosilica in the diatom Thalassiosira pseudonana. *Angew. Chemie Int. Ed.* **48**, 9724–7 (2009).

77. Strojsová, A. & Dyhrman, S. T. Cell-specific beta-N-acetylglucosaminidase activity in cultures and field populations of eukaryotic marine phytoplankton. *FEMS Microbiol. Ecol.* **64**, 351–61 (2008).

78. Turner, J. W., Malayil, L., Guadagnoli, D., Cole, D. & Lipp, E. K. Detection of Vibrio parahaemolyticus, Vibrio vulnificus and Vibrio cholerae with respect to seasonal fluctuations in temperature and plankton abundance. *Environ. Microbiol.* **16**, 1019–1028 (2014).

79. Turner, J. W., Good, B., Cole, D. & Lipp, E. K. Plankton composition and environmental factors contribute to Vibrio seasonality. *ISME J.* **3**, 1082–1092 (2009).

80. Asplund, M. E., Rehnstam-Holm, A. S., Atnur, V., Raghunath, P., Saravanan, V., Härnström, K., Collin, B., Karunasagar, I. & Godhe, A. Water column dynamics of Vibrio in relation to phytoplankton community composition and environmental conditions in a tropical coastal area. *Environ. Microbiol.* **13**, 2738–2751 (2011).

81. Olofsson, M., Asplund, M. E., Karunasagar, I., Rehnstam-Holm, A.-S. & Godhe, A. Prorocentrum micans promote and Skeletonema tropicum disfavours persistence of the pathogenic bacteria Vibrio parahaemolyticus. *Indian J. Geo-Marine Sci.* **42**, 729–733 (2013).

82. Rehnstam-Holm, A. S., Godhe, A., Härnström, K., Raghunath, P., Saravanan, V., Collin, B., Karunasagar, I. & Karunasagar, I. Association between phytoplankton and Vibrio spp. along the southwest coast of India: a mesocosm experiment. *Aquat. Microb. Ecol.* **58**, 127–139 (2010).

83. Hsieh, J. L., Fries, J. S. & Noble, R. T. Vibrio and phytoplankton dynamics during the summer of 2004 in a eutrophying estuary. *Ecol. Appl.* **17**, 102–109 (2007).

84. Hsieh, J. L., Fries, J. S. & Noble, R. T. Dynamics and predictive modelling of Vibrio spp. in the Neuse River Estuary, North Carolina, USA. *Environ. Microbiol.* **10**, 57–64 (2008).

85. Turner, J. W., Good, B., Cole, D. & Lipp, E. K. Plankton composition and environmental factors contribute to Vibrio seasonality. *ISME J.* **3**, 1082–1092 (2009).

86. Froelich, B., Bowen, J., Gonzalez, R., Snedeker, A. & Noble, R. Mechanistic and statistical models of total vibrio abundance in the neuse river estuary. *Water Res.* **47**, 5783–5793 (2013).

87. Holm-Hansen, O., Lorenzen, C. J., Holmes, R. W. & Strickland, J. D. H. Fluorometric determination of chlorophyll. *ICES J. Mar. Sci.* **30**, 3–15 (1965).

88. Taiwo, M., Baker-Austin, C., Powell, A., Hodgson, E., Natås, O. B. & Walker, D. I. Comparison of toxR and tlh based PCR assays for Vibrio parahaemolyticus. *Food Control* **77**, 116–120 (2017).

89. Campbell, M. S. & Wright, A. C. Real-Time PCR Analysis of Vibrio vulnificus from Oysters. *Appl. Environ. Microbiol.* **69**, 7137–7144 (2003).

90. Garrido-Maestu, A., Chapela, M. J., Peñaranda, E., Vieites, J. M. & Cabado, A. G. In-house validation of novel multiplex real-time PCR gene combination for the simultaneous detection of the main human pathogenic vibrios (Vibrio cholerae, Vibrio parahaemolyticus, and Vibrio vulnificus). *Food Control* **37**, 371–379 (2014).

91. Baker-Austin, C., Gore, A., Oliver, J. D., Rangdale, R., McArthur, J. V. & Lees, D. N. Rapid in situ detection of virulent Vibrio vulnificus strains in raw oyster matrices using real-time PCR. *Environ. Microbiol. Rep.* **2**, 76–80 (2010).

92. Baker-Austin, C., Lemm, E., Hartnell, R., Lowther, J., Onley, R., Amaro, C., Oliver, J. D. & Lees, D. PilF polymorphism-based real-time PCR to distinguish Vibrio vulnificus strains of human health relevance. *Food Microbiol.* **30**, 17–23 (2012).

93. Nordstrom, J. L., Vickery, M. C. L., Blackstone, G. M., Murray, S. L. & DePaola, A. Development of a Multiplex Real-Time PCR Assay with an Internal Amplification Control for the Detection of Total and Pathogenic Vibrio parahaemolyticus Bacteria in Oysters. *Appl. Environ. Microbiol.* **73**, 5840–5847 (2007).

94. Ward, L. N. & Bej, A. K. DETECTION OF VIBRIO PARAHAEMOLYTICUS by multiplex. **72**, 2031–2042 (2006).

95. Blackstone, G. M., Nordstrom, J. L., Bowen, M. D., Meyer, R. F., Imbro, P. & DePaola, A. Use of a real time PCR assay for detection of the ctxA gene of Vibrio cholerae in an environmental survey of Mobile Bay. *J. Microbiol. Methods* **68**, 254–259 (2007).

96. Parada, A. E., Needham, D. M. & Fuhrman, J. A. Every base matters: Assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ. Microbiol.* **18**, 1403–1414 (2016).

97. Amaral-Zettler, L. A., McCliment, E. A., Ducklow, H. W. & Huse, S. M. A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA Genes. *PLoS One* **4**, 1–9 (2009).

98. Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C., Ghalith, G. A. Al, Alexander, H., Alm, E. J., Arumugam, M., Bai, Y., Bisanz, J. E., Bittinger, K., Brejnrod, A., Colin, J., Brown, C. T., Callahan, B. J., Mauricio, A., Rodríguez, C., Chase, J., Cope, E., Silva, R. Da, Dorrestein, P. C., Douglas, G. M., Duvallet, C., Edwardson, C. F., Ernst, M., Fouquier, J., Gauglitz, J. M., Gibson, D. L., Gonzalez, A., Huttley, G. A., Janssen, S., Jarmusch, A. K., Kaehler, B. D., Kang, K. Bin, Keefe, C. R., Keim, P., Kelley, S. T., Ley, R., Loftfield, E., Marotz, C., Martin, B., Mcdonald, D., Mciver, L. J., Alexey, V., Metcalf, J. L., Morgan, S. C., Morton, J. T. & Naimey, A. T. QIIME 2 : Reproducible , interactive , scalable , and extensible microbiome data science. *PeerJ Prepr.* (2018). doi:10.7287/peerj.preprints.27295

99. McMurdie, P. J. & Holmes, S. Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS One* **8**, (2013).

100. Callahan, B. J., McMurdie, P. J. & Holmes, S. P. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME J.* **11**, 2639–2643 (2017).

101. Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A. & Holmes, S. P. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).

102. Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J. & Glöckner, F. O. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* **41**, 590–596 (2013).

103. Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., Boutte, C., Burgaud, G., De Vargas, C., Decelle, J., Del Campo, J., Dolan, J. R., Dunthorn, M., Edvardsen, B., Holzmann, M., Kooistra, W. H. C. F., Lara, E., Le Bescot, N., Logares, R., Mahé, F., Massana, R., Montresor, M., Morard, R., Not, F., Pawlowski, J., Probert, I., Sauvadet, A. L., Siano, R., Stoeck, T., Vaulot, D., Zimmermann, P. & Christen, R. The Protist Ribosomal Reference database (PR2): A catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. *Nucleic Acids Res.* **41**, 597–604 (2013).

104. Faith, D. P. Conservation evaluation and phylogenetic diversity. *Biol. Conserv.* **61**, 1–10 (1992).

105. Wei, Taiyun; Simko, V. R package: 'corrplot': Visualization of a Correlation Matrix (version 0.84). *available from https//github.com/taiyun/corrplot* (2017).

106. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

107. Segata, N., Boernigen, D., Tickle, T. L., Morgan, X. C., Garrett, W. S. & Huttenhower, C. Computational meta'omics for microbial community studies. *Mol. Syst. Biol.* **9**, 1–15 (2013).

108. Asnicar, F., Weingart, G., Tickle, T. L., Huttenhower, C. & Segata, N. Compact graphical representation of phylogenetic data and metadata with GraPhlAn. *PeerJ* **3**, e1029 (2015).

109. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. MetaSPAdes: A new versatile metagenomic assembler. *Genome Res.* **27**, 824–834 (2017).

110. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: Quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072–1075 (2013).

111. Laczny, C. C., Sternal, T., Plugaru, V., Gawron, P., Atashpendar, A., Margossian, H. H., Coronado, S., der Maaten, L. V., Vlassis, N. & Wilmes, P. VizBin - An application for reference-independent visualization and human-augmented binning of metagenomic data. *Microbiome* **3**, 1–7 (2015).

112. O'Malley, M. A. The nineteenth century roots of 'everything is everywhere'. *Nat. Rev. Microbiol.* **5**, 647–651 (2007).

113. Kaspar, C. W. & Tamplin, M. L. Effects of Temperature and Salinity on the Survival of Vibrio-Vulnificus in Seawater and Shellfish. *Appl. Environ. Microbiol.* **59**, 2425–2429 (1993).

114. Oliver, J. D. The Biology of Vibrio vulni fi cus. **3**, 1–10 (2015).

115. Randa, M. a, Polz, M. F. & Lim, E. Effects of Temperature and Salinity on Vibrio vulnificus Population Dynamics as Assessed by Quantitative PCR Effects of Temperature and Salinity on Vibrio vulnificus Population Dynamics as Assessed by Quantitative PCR. *Appl. Environ. Microbiol.* **70**, 5469–5476 (2004).

116. Materna, A. C., Friedman, J., Bauer, C., David, C., Chen, S., Huang, I. B., Gillens, A., Clarke, S. a, Polz, M. F. & Alm, E. J. Shape and evolution of the fundamental niche in marine Vibrio. *ISME J.* **6**, 2168–2177 (2012).

117. Williams, T. C., Froelich, B. A., Phippen, B., Fowler, P., Noble, R. T. & Oliver, J. D. Different abundance and correlational patterns exist between total and presumed pathogenic Vibrio vulnificus and V. parahaemolyticus in shellfish and waters along the North Carolina coast. *FEMS Microbiol. Ecol.* **93**, 1–11 (2017).

118. Nilsson, W. B. & Turner, J. W. The thermostable direct hemolysin-related hemolysin (trh) gene of Vibrio parahaemolyticus: Sequence variation and implications for detection and function. *J. Microbiol. Methods* **126**, 1–7 (2016).

119. Hector, S. M. Shellfish Consumption in San Diego. *PCAS Q.* **38**, 105–116 (2002).

120. Crooks, J. A., Crooks, K. R. & Crooks, A. J. Observations of the non-native Pacific oyster (Crassostrea gigas) in San Diego County , California. *Calif. Fish Game* **101**, 101–107 (2015).

121. Kaysner, C. a., Abeyta, C., Wekell, M. M., DePaola, a., Stott, R. F. & Leitch, J. M. Virulent strains of Vibrio vulnificus isolated from estuaries of the United States West Coast. *Appl. Environ. Microbiol.* **53**, 1349–1351 (1987).

122. Dickinson, G., Lim, K. Y. & Jiang, S. C. Quantitative microbial risk assessment of pathogenic vibrios in marine recreational waters of Southern California. *Appl. Environ. Microbiol.* **79**, 294–302 (2013).

123. Jesser, K. J. & Noble, R. T. Vibrio Ecology in the Neuse River Estuary, North

Carolina, Characterized by Next-Generation Amplicon Sequencing of the Gene Encoding Heat Shock Protein 60 (hsp60). *Appl. Environ. Microbiol.* **84**, 1–21 (2018).

124. Main, C. R., Salvitti, L. R., Whereat, E. B. & Coyne, K. J. Community-level and species-specific associations between phytoplankton and particle-associated Vibrio species in Delaware's inland bays. *Appl. Environ. Microbiol.* **81**, 5703–5713 (2015).

125. Bowler, C., Allen, A. E., Badger, J. H., Grimwood, J., Jabbari, K., Kuo, A., Maheswari, U., Martens, C., Maumus, F., Otillar, R. P., Rayko, E., Salamov, A., Vandepoele, K., Beszteri, B., Gruber, A., Heijde, M., Katinka, M., Mock, T., Valentin, K., Verret, F., Berges, J. a, Brownlee, C., Cadoret, J.-P., Chiovitti, A., Choi, C. J., Coesel, S., De Martino, A., Detter, J. C., Durkin, C., Falciatore, A., Fournet, J., Haruta, M., Huysman, M. J. J., Jenkins, B. D., Jiroutova, K., Jorgensen, R. E., Joubert, Y., Kaplan, A., Kröger, N., Kroth, P. G., La Roche, J., Lindquist, E., Lommer, M., Martin-Jézéquel, V., Lopez, P. J., Lucas, S., Mangogna, M., McGinnis, K., Medlin, L. K., Montsant, A., Oudot-Le Secq, M.-P., Napoli, C., Obornik, M., Parker, M. S., Petit, J.-L., Porcel, B. M., Poulsen, N., Robison, M., Rychlewski, L., Rynearson, T. a, Schmutz, J., Shapiro, H., Siaut, M., Stanley, M., Sussman, M. R., Taylor, A. R., Vardi, A., von Dassow, P., Vyverman, W., Willis, A., Wyrwicz, L. S., Rokhsar, D. S., Weissenbach, J., Armbrust, E. V., Green, B. R., Van de Peer, Y. & Grigoriev, I. V. The Phaeodactylum genome reveals the evolutionary history of diatom genomes. *Nature* **456**, 239–44 (2008).

126. Grossart, H.-P., Levold, F., Allgaier, M., Simon, M. & Brinkhoff, T. Marine diatom species harbour distinct bacterial communities. *Environ. Microbiol.* **7**, 860–73 (2005).

127. Chen, W. & Wangersky, P. J. Production of dissolved organic carbon in phytoplankton cultures as measured by high-temperature catalytic oxidation and ultraviolet photo-oxidation methods. *J. Plankton Res.* **18**, 1201–1211 (1996).

128. Grossart, H.-P. & Simon, M. Interactions of planctonic algae and bacteria: Effects on algal growth and organic matter dynamics. *Aquat. Microb. Ecol.* **47**, 163–176 (2007).

129. Bruckner, C. G., Rehm, C., Grossart, H.-P. & Kroth, P. G. Growth and release of extracellular organic compounds by benthic diatoms depend on interactions with bacteria. *Environ. Microbiol.* **13**, 1052–63 (2011).

130. Mayali, X. & Azam, F. Algicidal bacteria in the sea and their impact on algal blooms. *J. Eukaryot. Microbiol.* **51**, 139–44 (2004).

131. Kim, J., Kim, C., Youn, S. & Choi, T. Isolation and Physiological Characterization of a Novel Algicidal Virus Infecting the Marine Diatom Skeletonema costatum. **31**, 186–191 (2015).

132. Traller, J. C., Cokus, S. J., Lopez, D. A., Gaidarenko, O., Smith, S. R., McCrow, J. P., Gallaher, S. D., Podell, S., Thompson, M., Cook, O., Morselli, M., Jaroszewicz, A., Allen, E. E., Allen, A. E., Merchant, S. S., Pellegrini, M. & Hildebrand, M. Genome

and methylome of the oleaginous diatom Cyclotella cryptica reveal genetic flexibility toward a high lipid phenotype. *Biotechnol Biofuels* **9**, 258 (2016).

133. Crenn, K., Duffieux, D. & Jeanthon, C. Bacterial Epibiotic Communities of Ubiquitous and Abundant Marine Diatoms Are Distinct in Short- and Long-Term Associations. *Front. Microbiol.* **9**, 1–12 (2018).

134. Armbrust, E. V., Berges, J. a, Bowler, C., Green, B. R., Martinez, D., Putnam, N. H., Zhou, S., Allen, A. E., Apt, K. E., Bechner, M., Brzezinski, M. a, Chaal, B. K., Chiovitti, A., Davis, A. K., Demarest, M. S., Detter, J. C., Glavina, T., Goodstein, D., Hadi, M. Z., Hellsten, U., Hildebrand, M., Jenkins, B. D., Jurka, J., Kapitonov, V. V, Kröger, N., Lau, W. W. Y., Lane, T. W., Larimer, F. W., Lippmeier, J. C., Lucas, S., Medina, M., Montsant, A., Obornik, M., Parker, M. S., Palenik, B., Pazour, G. J., Richardson, P. M., Rynearson, T. a, Saito, M. a, Schwartz, D. C., Thamatrakoln, K., Valentin, K., Vardi, A., Wilkerson, F. P. & Rokhsar, D. S. The genome of the diatom Thalassiosira pseudonana: ecology, evolution, and metabolism. *Science* **306**, 79–86 (2004).

135. Shrestha, R. P. & Hildebrand, M. Development of a silicon limitation inducible expression system for recombinant protein production in the centric diatoms Thalassiosira pseudonana and Cyclotella cryptica. *Microb. Cell Fact.* **16**, 1–14 (2017).

136. Poulsen, N., Chesley, P. M. & Kröger, N. Molecular genetic manipulation of the diatom Thalassiosira pseudonana (Bacillariophyceae). *J. Phycol.* **42**, 1059–1065 (2006).

137. Huq, A., Small, E. B., West, P. A., Huq, M. I. & Colwell, R. R. Ecological relationships between Vibrio cholerae and planktonic crustacean Ecological Relationships Between Vibrio cholerae and Planktonic Crustacean Copepods. **45**, 275–283 (1983).

138. Rawlings, T. K., Ruiz, G. M. & Colwell, R. R. Association of Vibrio cholerae O1 El Tor and O139 Bengal with the copepods Acartia tonsa and Eurytemora affinis. *Appl. Environ. Microbiol.* **73**, 7926–7933 (2007).

139. de Magny, G. C., Mozumder, P. K., Grim, C. J., Hasan, N. A., Naser, M. N., Alam, M., Sack, R. B., Huq, A. & Colwell, R. R. Role of Zooplankton Diversity in Vibrio cholerae Population Dynamics and in the Incidence of Cholera in the Bangladesh Sundarbans. *Appl. Environ. Microbiol.* **77**, 6125–6132 (2011).

140. Cordell, J. R. & Morrison, S. M. The Invasive Asian Copepod Pseudodiaptomus inopinus in Oregon , Washington , and British Columbia Estuaries. *Estuaries* **19**, 629–638 (1996).

141. Jeffery R. Cordell, C. A. M. and C. A. S. Occurrence of the Asian Calanoid Copepod Pseudodiaptomus inopinus in the Zooplankton of the Columbia River Estuary. *J. Crustac. Biol.* **12**, 260–269 (1992).

142. Williams, T. C., Ayrapetyan, M. & Oliver, J. D. Molecular and Physical Factors That Influence Attachment of Vibrio vulnificus to Chitin. *Appl. Environ. Microbiol.* **81**, 6158–6165 (2015).

143. Lee, J. S., Rhee, J. S., Kim, R. O., Hwang, D. S., Han, J., Choi, B. S., Park, G. S., Kim, I. C., Park, H. G. & Lee, Y. M. The copepod Tigriopus japonicus genomic DNA information (574Mb) and molecular anatomy. *Mar. Environ. Res.* **69**, S21–S23 (2010).

144. Barreto, F. S., Schoville, S. D. & Burton, R. S. Reverse genetics in the tide pool: Knock-down of target gene expression via RNA interference in the copepod Tigriopus californicus. *Mol. Ecol. Resour.* **15**, 868–879 (2015).

145. Barreto, F. S., Watson, E. T., Lima, T. G., Willett, C. S., Edmands, S., Li, W. & Burton, R. S. Genomic signatures of mitonuclear coevolution across populations of Tigriopus californicus. *Nat. Ecol. Evol.* **2**, 1250–1257 (2018).

146. Kang, S., Ahn, D. H., Lee, J. H., Lee, S. G., Shin, S. C., Lee, J., Min, G. S., Lee, H., Kim, H. W., Kim, S. & Park, H. The genome of the Antarctic-endemic copepod, Tigriopus kingsejongensis. *Gigascience* **6**, 1–9 (2017).

147. Chen, Z., Wang, G., Zeng, C. & Wu, L. Comparative study on the effects of two diatoms as diets on planktonic calanoid and benthic harpacticoid copepods. *J. Exp. Zool. Part A Ecol. Integr. Physiol.* **329**, 140–148 (2018).

148. Buffan-Dubau, E., De Wit, R. & Castel, J. Feeding selectivity of the harpacticoid copepod Canuella perplexa in benthic muddy environments demonstrated by HPLC analyses of chlorin and carotenoid pigments. *Mar. Ecol. Prog. Ser.* **137**, 71–82 (1996).

149. Wang, G., Xu, J., Zeng, C., Jia, Q., Wu, L. & Li, S. Pelagic microalgae as suitable diets for the benthic harpacticoid copepod Tigriopus japonicus. *Hydrobiologia* **762**, 81–88 (2015).

150. Ianora, A., Poulet, S. A. & Miralto, A. The effects of diatoms on copepod reproduction: a review. *Phycologia* **42**, 351–363 (2010).

Chapter 4, in part, is currently being prepared for submission for publication of the material in **Diner, RE,** Rabines, A., Zheng, H., Allen, A.E. High-resolution taxonomic grouping reveals interactions between pathogenic *Vibrio* species and the planktonic community. The dissertation author was the primary investigator and author of this material.

Tables

Table 1. Target genes for digital droplet PCR (ddPCR), including the target species, target gene name, primer name, primer or probe sequences, annealing temperature conditions (Ta), amplicon size when reported, and study primers were obtained from.

Assay	Target Species	Target Gene	Primer name	Sequence	Ta (°C)	Amplicon Size (bp)	Study
ddPCR	*V. vulnificus*	*vcgC*	VVC-FW	AAAACTCATTGARCAGTAACGAAA	60	Not reported	Baker-Austin et al. 2010
ddPCR	*V. vulnificus*	*vcgC*	VVC-REV	AGCTGGATCTAAKCCCAATGC	60	Not reported	Baker-Austin et al. 2010
ddPCR	*V. vulnificus*	*vcgC*	VVC-Probe	/5HEX/AATTAAAGC/ZEN/CGTCAAGCCACTTGACTGTAA/3IABkFQ/	60	Not reported	Baker-Austin et al. 2010
ddPCR	*V. vulnificus*	*pilF*	PILF-FW	GATTGACTACGAYCCACACCG	60	Not reported	Baker-Austin et al. 2012
ddPCR	*V. vulnificus*	*pilF*	PILF-REV	GRCGCGCTTGGGTGTAG	60	Not reported	Baker-Austin et al. 2012
ddPCR	*V. vulnificus*	*pilF*	PILF-PROBE	/56-FAM/TGCTCAACC/ZEN/TCGCTAAGTTGGAAATCGATA/3IABkFQ/	60	Not reported	Baker-Austin et al. 2012
ddPCR	*V. parahaemolyticus*	*toxr*	TOXR-FW	GAACCAGAAGCGCCAGTAGT	58	Not reported	Taiwo et al. 2017
ddPCR	*V. parahaemolyticus*	*toxr*	TOXR-REV	AAACAAGCAGTACGCAAATCG	58	Not reported	Taiwo et al. 2017
ddPCR	*V. parahaemolyticus*	*toxr*	TOXR-Probe	/5HEX/TCACAGCAG/ZEN/AAGCCACAGGTGC/3IABkFQ/	58	Not reported	Taiwo et al. 2017
ddPCR	*V. vulnificus*	*vvhA*	VVHA-FW	TGTTTATGGTGAGAACGGTGACA	58	Not reported	Campbell and Wright 2003
ddPCR	*V. vulnificus*	*vvhA*	VVHA- REV	TTCTTTATCTAGGCCCCAAACTTG	58	Not reported	Campbell and Wright 2003
ddPCR	*V. vulnificus*	*vvhA*	VVHA-Probe	/56-FAM/CCGTTAACC/ZEN/GAACCACCCGCAA/3IABkFQ/	58	Not reported	Campbell and Wright 2003
ddPCR	*V. cholerae*	*ompW*	ompW-F	TCAATGATAGCTGGTTCCTCAAC	58	87	Garrido-Maestu et al. 2014
ddPCR	*V. cholerae*	*ompW*	ompW-R	CGATGATAAATACCCAAGGATTGA	58	87	Garrido-Maestu et al. 2014
ddPCR	*V. cholerae*	*ompW*	ompW-Probe	/5HEX/TGGTATGCC/ZEN/AATATTGAAACAACG/3IABkFQ/	58	87	Garrido-Maestu et al. 2014
Traditional PCR	*V. cholerae*	*ctxA*	ctxA-F	TTTGTTAGGCACGATGATGGAT	63	84	Blackstone et al. 2007
Traditional PCR	*V. cholerae*	*ctxA*	ctxA-R	ACCAGACAATATAGTTTGACCCACTAAG	63	84	Blackstone et al. 2007
Traditional PCR	*V. parahaemolyticus*	*tdh*	TDHF (Fw)	TCCCTTTTCCTGCCCCC	61	233	Nordstrom et al. 2007
Traditional PCR	*V. parahaemolyticus*	*tdh*	TDHR (rev)	CGCTGCCATTGTATAGTCTTTATC	61	233	Nordstrom et al. 2007
Traditional PCR	*V. parahaemolyticus*	*trh*	TRHF (Fw)	CCATCMATACCTTTTCCTTCTCC	60	207	Ward and Bej 2006
Traditional PCR	*V. parahaemolyticus*	*trh*	TRHR (REV)	ACYGTCATATAGGCGCTTAAC	60	207	Ward and Bej 2006

Table 2. Primer sequences used to amplify 16S and 18S amplicon libraries. Color coded sequences represent the following: green and red = Illumina sequencing adaptor sequences, black = library specific 8-bp index sequences, orange = linker base pairs, blue = primer sequence for amplicon annealing.

Amplicon	Primer name	Primer sequence
16S	515F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNXXXXXXXXGTGYCAGCMGCCGCGGTAA
16S	926R	CAAGCAGAAGACGGCATACGAGATXXXXXXXXGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCGYCAATTYMTTTRAGTTT
18S	1389F	AATGATACGGCGACCACCGAGATCTACACXXXXXXXXTATGGTAATTGTTTGTACACACCGCCC
18S	1510R	CAAGCAGAAGACGGCATACGAGATXXXXXXXXAGTCAGTCAGGGCCTTCYGCAGGTTCACCTAC

Table 3. Community statistics from shotgun metagenomic sequencing of *Vibrio* communities isolated on CHROMagar *Vibrio* plates compared to ddPCR quantification levels. Relative abundance in the community is reported for the three pathogenic *Vibrio* species quantified in this study, and other abundant community members including all *Vibrio* bacteria combined. ddPCR detection level categories are defined as follows: low = <100 copies/ml, medium = 100-1000 copies/100ml, high = >1000 copies/100ml.

		V. parahaemolyticus		*V. vulnificus*		*V cholerae*		Abundant community members		
Sampling site	Month	% relative abundance of isolate community	ddPCR detection level	relative abundance of isolate community	ddPCR detection level	relative abundance of isolate community	ddPCR detection level	*Vibrio* % community	*Grimontia hollisae* % community	Vibrio phage vB_VpaM_MAR % community
LPL	February	35	0	1	0	0	0	70	22	0
LPL	March	78	Medium	0	0	16	High	97	0	0
LPL	May	52	Medium	15	High	14	High	81	0	1
LPL	July	79	Low	0	0	0	0	100	0	0
LPL	August	34	Medium	3	Low	0	Low	97	0	0
SDR1	February	7	0	0	0	0	0	89	11	0
SDR1	March	79	Medium	0	0	0	Low	100	0	0
SDR1	May	48	High	25	High	0	0	99	0	0
SDR1	July	16	High	0	Medium	0	Low	100	0	0
SDR1	August	6	Medium	0	0	0	0	100	0	0
SDR2	February	63	Medium	0	0	0	0	96	3	0
SDR2	March	66	Medium	1	0	1	0	99	1	0
SDR2	May	5	High	26	High	5	Medium	99	0	0
SDR2	July	14	Medium	0	0	0	Low	100	0	0
SDR2	August	36	0	1	Low	0	0	100	0	0
TJ1	February	58	High	0	0	1	0	99	1	0
TJ1	March	14	Medium	0	Medium	0	0	98	2	0
TJ1	May	14	Medium	0	0	0	0	99	0	0
TJ1	July *	100	Medium	1	0	0	0	100	0	0
TJ1	August	4	Medium	0	0	0	0	100	0	0
TJ2	February	5	Medium	0	0	0	0	15	1	82
TJ2	March	6	Medium	2	Low	0	0	97	2	0
TJ2	May	32	High	0	Low	0	0	99	0	0
TJ2	July *	100	High	0	0	0	0	100	0	0
TJ2	August	56	Medium	0	0	0	0	100	0	0

Table 4. Alpha and beta diversity statistics for 16S and 18S communities. Significant differences (indicated by a p>0.05) are in bold.

Diversity Type	Diversity Metric	Factor	Method name	Sample size	Number of groups	Test metric	Test statistic	p-value
16S Amplicon								
Beta	Weighted Unifrac	Site	PERMANOVA	60	5	pseudo F	4.52	**0.0010**
Beta	Bray-Curtis	Site	PERMANOVA	60	5	pseudo F	2.39	**0.0010**
Beta	Weighted Unifrac	Month	PERMANOVA	60	12	pseudo F	1.92	**0.0010**
Beta	Bray-Curtis	Month	PERMANOVA	60	12	pseudo F	2.32	**0.0010**
Alpha	Faith PD	Temperature	Spearman	60	NA	Spearman	-0.35	**0.0067**
Alpha	Faith PD	Salinity	Spearman	60	NA	Spearman	-0.29	**0.0241**
Alpha	Faith PD	Chlorophyll A	Spearman	60	NA	Spearman	-0.35	**0.0062**
Alpha	Faith PD	Site	Krustkal-Wallis	60	5	H	16.31	**0.0026**
Alpha	Faith PD	Month	Krustkal-Wallis	60	12	H	16.04	0.1396
Alpha	Evenness	Temperature	Spearman	60	NA	Spearman	-0.11	0.4004
Alpha	Evenness	Salinity	Spearman	60	NA	Spearman	-0.03	0.8280
Alpha	Evenness	Chlorophyll A	Spearman	60	NA	Spearman	-0.40	**0.0016**
Alpha	Evenness	Site	Krustkal-Wallis	60	5	H	5.09	0.2786
Alpha	Evenness	Month	Krustkal-Wallis	60	12	H	16.93	0.1099
18S Amplicon								
Beta	Weighted Unifrac	Site	PERMANOVA	59	5	pseudo F	2.80	**0.0010**
Beta	Bray-Curtis	Site	PERMANOVA	59	5	pseudo F	2.12	**0.0010**
Beta	Weighted Unifrac	Month	PERMANOVA	59	12	pseudo F	1.66	**0.0010**
Beta	Bray-Curtis	Month	PERMANOVA	59	12	pseudo F	1.75	**0.0010**
Alpha	Faith PD	Temperature	Spearman	59	NA	Spearman	-0.24	0.0661
Alpha	Faith PD	Salinity	Spearman	59	NA	Spearman	-0.17	0.2107
Alpha	Faith PD	Chlorophyll A	Spearman	59	NA	Spearman	-0.52	**0.0000**
Alpha	Faith PD	Site	Krustkal-Wallis	59	5	H	15.46	**0.0038**
Alpha	Faith PD	Month	Krustkal-Wallis	59	12	H	21.36	**0.0298**
Alpha	Evenness	Temperature	Spearman	59	NA	Spearman	-0.33	**0.0112**
Alpha	Evenness	Salinity	Spearman	59	NA	Spearman	-0.13	0.3275
Alpha	Evenness	Chlorophyll A	Spearman	59	NA	Spearman	-0.72	**0.0000**
Alpha	Evenness	Site	Krustkal-Wallis	59	5	H	13.45	**0.0093**
Alpha	Evenness	Month	Krustkal-Wallis	59	12	H	21.94	**0.0248**

Table 5. Pairwise diversity analysis (alpha and beta) comparisons between sites for 16S and 18S communities. Both Bray-Curtis and Weighted Unifrac methods were employed for beta diversity, as well as both species richness (Faith PD) and evenness for Alpha diversity. Significantly differences between sites (p>0.0.5) are indicated in bold font.

		Beta Diversity - PERMANOVA						Alpha Diversity - KRUSTKAL-WALIS					
		Bray-Curtis			Weighted Unifrac			Faith PD			Evenness		
16S Amplicon													
Group 1	Group 2	pseudo-F	p-value	q-value	pseudo-F	p-value	q-value	H	p-value	q-value	H	p-value	q-value
LPL	SDR1	2.558	**0.002**	0.003	4.295	**0.002**	0.003	0.563	0.453	0.503	2.430	0.119	0.397
LPL	SDR2	2.253	**0.004**	0.005	2.122	**0.041**	0.051	0.750	0.386	0.483	1.080	0.299	0.747
LPL	TJ1	2.502	**0.001**	0.002	6.060	**0.001**	0.003	4.813	**0.028**	0.082	4.083	**0.043**	0.397
LPL	TJ2	2.639	**0.001**	0.002	7.403	**0.001**	0.003	4.563	**0.033**	0.082	3.000	0.083	0.397
SDR1	SDR2	0.626	0.829	0.829	0.962	0.376	0.376	3.630	0.057	0.113	0.270	0.603	0.754
SDR1	TJ1	2.719	**0.001**	0.002	4.428	**0.004**	0.006	11.603	**0.001**	0.007	0.403	0.525	0.751
SDR1	TJ2	3.240	**0.001**	0.002	6.152	**0.001**	0.003	8.670	**0.003**	0.016	0.120	0.729	0.810
SDR2	TJ1	2.981	**0.002**	0.003	5.128	**0.002**	0.003	3.000	0.083	0.139	0.480	0.488	0.751
SDR2	TJ2	3.454	**0.001**	0.002	6.512	**0.001**	0.003	2.430	0.119	0.170	0.653	0.419	0.751
TJ1	TJ2	0.825	0.630	0.700	1.552	0.142	0.158	0.000	1.000	1.000	0.000	1.000	1.000
18S Amplicon													
Group 1	Group 2	pseudo-F	p-value	q-value	pseudo-F	p-value	q-value	H	p-value	q-value	H	p-value	q-value
LPL	SDR1	2.591	**0.001**	0.001	3.703	**0.001**	0.003	2.970	0.085	0.141	2.970	0.085	0.121
LPL	SDR2	2.346	**0.001**	0.001	2.469	**0.001**	0.003	1.080	0.299	0.332	1.920	0.166	0.207
LPL	TJ1	1.824	**0.002**	0.003	3.179	**0.006**	0.008	5.333	**0.021**	0.052	6.453	**0.011**	0.065
LPL	TJ2	2.346	**0.001**	0.001	3.309	**0.004**	0.007	2.613	0.106	0.151	6.163	**0.013**	0.065
SDR1	SDR2	0.847	0.674	0.749	1.532	0.122	0.136	1.367	0.242	0.303	0.004	0.951	0.951
SDR1	TJ1	2.070	**0.001**	0.001	2.817	**0.004**	0.007	7.004	**0.008**	0.039	4.125	**0.042**	0.094
SDR1	TJ2	2.913	**0.001**	0.001	4.059	**0.001**	0.003	6.367	0.012	0.039	3.640	0.056	0.094
SDR2	TJ1	2.444	**0.001**	0.001	3.004	**0.006**	0.008	7.363	**0.007**	0.039	3.853	**0.050**	0.094
SDR2	TJ2	3.327	**0.001**	0.001	3.692	**0.001**	0.003	4.563	**0.033**	0.065	4.563	**0.033**	0.094
TJ1	TJ2	0.519	0.969	0.969	0.574	0.820	0.820	0.213	0.644	0.644	0.030	0.862	0.951

Table 6. Pairwise diversity analysis (alpha and beta) comparisons between months for 16S communities. Both Bray-Curtis and Weighted Unifrac methods were employed for beta diversity, as well as both species richness (Faith PD) and evenness for Alpha diversity. Significantly differences between months (p>0.0.5) are indicated in bold font.

		16S Amplicon											
		Beta Diversity - PERMANOVA						Alpha Diversity - KRUSTKAL-WALIS					
		Bray-Curtis			Weighted Unifrac			Faith PD			Evenness		
Group 1	Group 2	pseudo-F	p-value	q-value	pseudo-F	p-value	q-value	H	p-value	q-value	H	p-value	q-value
Apr	Aug	2.457	**0.015**	0.025	1.315	0.239	0.329	0.011	0.917	0.917	0.098	0.754	0.917
Apr	Dec	3.097	**0.009**	0.023	4.034	**0.013**	0.061	0.273	0.602	0.778	0.273	0.602	0.863
Apr	Feb	2.179	**0.016**	0.026	1.524	0.207	0.318	0.535	0.465	0.767	0.098	0.754	0.917
Apr	Jan	2.333	**0.017**	0.027	1.796	0.189	0.304	1.320	0.251	0.662	0.011	0.917	0.917
Apr	Jul	2.016	**0.025**	0.038	1.761	0.124	0.234	0.273	0.602	0.778	0.011	0.917	0.917
Apr	Jun	1.773	0.069	0.091	1.029	0.343	0.419	0.273	0.602	0.778	0.535	0.465	0.863
Apr	Mar	1.566	0.13	0.153	0.775	0.404	0.460	0.098	0.754	0.803	0.273	0.602	0.863
Apr	May	1.332	0.161	0.183	0.523	0.57	0.603	0.011	0.917	0.917	1.320	0.251	0.752
Apr	Nov	3.223	**0.01**	0.023	4.061	0.003	0.059	2.455	0.117	0.483	6.818	**0.009**	0.099
Apr	Oct	2.999	**0.009**	0.023	3.724	0.008	0.059	0.273	0.602	0.778	0.535	0.465	0.863
Apr	Sep	2.694	0.012	0.023	2.971	**0.011**	0.061	0.098	0.754	0.803	0.273	0.602	0.863
Aug	Dec	4.062	**0.007**	0.023	3.704	**0.019**	0.078	0.884	0.347	0.764	0.098	0.754	0.917
Aug	Feb	3.280	**0.009**	0.023	1.170	0.356	0.421	1.320	0.251	0.662	0.273	0.602	0.863
Aug	Jan	3.423	**0.006**	0.023	1.658	0.14	0.243	2.455	0.117	0.483	0.098	0.754	0.917
Aug	Jul	1.052	0.37	0.394	0.665	0.576	0.603	0.273	0.602	0.778	0.011	0.917	0.917
Aug	Jun	1.636	0.058	0.081	0.718	0.659	0.680	0.535	0.465	0.767	0.884	0.347	0.863
Aug	Mar	2.534	**0.011**	0.023	1.251	0.324	0.411	0.098	0.754	0.803	0.098	0.754	0.917
Aug	May	2.059	**0.009**	0.023	1.653	0.217	0.320	0.011	0.917	0.917	1.320	0.251	0.752
Aug	Nov	2.691	**0.007**	0.023	3.034	**0.021**	0.082	3.938	**0.047**	0.312	6.818	**0.009**	0.099
Aug	Oct	1.642	0.081	0.103	2.320	0.061	0.134	0.098	0.754	0.803	0.273	0.602	0.863
Aug	Sep	0.830	0.645	0.645	1.863	0.139	0.243	0.535	0.465	0.767	0.273	0.602	0.863
Dec	Feb	1.579	0.072	0.093	1.572	0.096	0.198	0.273	0.602	0.778	0.011	0.917	0.917
Dec	Jan	1.162	0.263	0.294	1.268	0.218	0.320	0.884	0.347	0.764	0.011	0.917	0.917
Dec	Jul	3.257	**0.008**	0.023	2.341	**0.038**	0.113	4.811	**0.028**	0.311	0.011	0.917	0.917
Dec	Jun	3.610	**0.012**	0.023	3.875	**0.007**	0.059	0.098	0.754	0.803	0.884	0.347	0.863
Dec	Mar	2.676	**0.012**	0.023	3.197	**0.012**	0.061	0.273	0.602	0.778	0.011	0.917	0.917
Dec	May	3.186	**0.008**	0.023	4.023	**0.007**	0.059	1.844	0.175	0.524	0.273	0.602	0.863
Dec	Nov	3.371	**0.01**	0.023	2.637	**0.023**	0.084	3.938	**0.047**	0.312	3.938	0.047	0.260
Dec	Oct	3.761	**0.01**	0.023	2.106	**0.041**	0.113	1.320	0.251	0.662	0.011	0.917	0.917
Dec	Sep	4.444	**0.01**	0.023	2.831	**0.041**	0.113	0.884	0.347	0.764	0.011	0.917	0.917
Feb	Jan	1.073	0.321	0.347	0.409	0.824	0.824	0.098	0.754	0.803	0.273	0.602	0.863
Feb	Jul	2.607	**0.011**	0.023	1.015	0.369	0.427	3.938	**0.047**	0.312	0.098	0.754	0.917
Feb	Jun	2.402	**0.007**	0.023	1.065	0.357	0.421	0.098	0.754	0.803	0.535	0.465	0.863
Feb	Mar	1.404	0.118	0.142	0.975	0.457	0.494	0.535	0.465	0.767	0.273	0.602	0.863
Feb	May	2.217	**0.013**	0.023	1.936	0.119	0.231	1.844	0.175	0.524	1.320	0.251	0.752
Feb	Nov	2.725	**0.014**	0.024	1.836	**0.033**	0.104	0.273	0.602	0.778	2.455	0.117	0.516
Feb	Oct	2.985	**0.005**	0.023	1.592	0.096	0.198	0.884	0.347	0.764	0.098	0.754	0.917
Feb	Sep	3.475	**0.008**	0.023	1.944	0.056	0.134	0.535	0.465	0.767	0.273	0.602	0.863
Jan	Jul	2.795	**0.013**	0.023	1.316	0.239	0.329	3.938	**0.047**	0.312	0.273	0.602	0.863
Jan	Jun	2.828	**0.01**	0.023	1.391	0.256	0.345	0.273	0.602	0.778	1.320	0.251	0.752
Jan	Mar	2.011	**0.011**	0.023	1.205	0.281	0.364	1.844	0.175	0.524	0.884	0.347	0.863
Jan	May	2.514	**0.008**	0.023	2.124	0.139	0.243	2.455	0.117	0.483	1.844	0.175	0.678
Jan	Nov	3.293	**0.01**	0.023	2.006	**0.014**	0.062	0.098	0.754	0.803	5.771	0.016	0.108
Jan	Oct	3.377	**0.005**	0.023	1.771	**0.031**	0.104	2.455	0.117	0.483	0.535	0.465	0.863
Jan	Sep	3.741	**0.01**	0.023	1.933	0.061	0.134	1.844	0.175	0.524	0.273	0.602	0.863
Jul	Jun	0.829	0.645	0.645	0.696	0.692	0.703	0.535	0.465	0.767	1.320	0.251	0.752
Jul	Mar	2.193	**0.018**	0.028	1.557	0.15	0.254	1.844	0.175	0.524	0.011	0.917	0.917
Jul	May	1.203	0.293	0.322	1.618	0.181	0.299	1.844	0.175	0.524	0.273	0.602	0.863
Jul	Nov	2.590	**0.008**	0.023	2.662	**0.007**	0.059	6.818	**0.009**	0.149	6.818	**0.009**	0.099
Jul	Oct	1.783	0.059	0.081	1.911	0.11	0.220	3.153	0.076	0.417	0.535	0.465	0.863
Jul	Sep	1.326	0.15	0.174	1.457	0.223	0.320	3.153	0.076	0.417	0.535	0.465	0.863
Jun	Mar	1.846	**0.038**	0.056	0.905	0.425	0.475	0.535	0.465	0.767	3.153	0.076	0.385
Jun	May	0.836	0.539	0.556	1.052	0.333	0.415	0.535	0.465	0.767	4.811	0.028	0.170
Jun	Nov	2.591	**0.008**	0.023	3.533	**0.011**	0.061	0.098	0.754	0.803	0.273	0.602	0.863
Jun	Oct	2.065	**0.04**	0.057	2.969	**0.004**	0.059	0.273	0.602	0.778	0.273	0.602	0.863
Jun	Sep	1.688	0.069	0.091	2.249	0.053	0.134	0.273	0.602	0.778	0.011	0.917	0.917
Mar	May	1.661	0.114	0.139	0.795	0.434	0.477	0.535	0.465	0.767	0.011	0.917	0.917
Mar	Nov	2.438	**0.01**	0.023	2.948	**0.007**	0.059	6.818	0.009	0.149	5.771	**0.016**	0.108
Mar	Oct	2.466	**0.005**	0.023	2.502	**0.005**	0.059	0.098	0.754	0.803	1.844	0.175	0.678
Mar	Sep	2.578	**0.011**	0.023	2.039	**0.032**	0.104	0.098	0.754	0.803	2.455	0.117	0.516
May	Nov	3.096	**0.007**	0.023	4.395	**0.008**	0.059	6.818	**0.009**	0.149	6.818	**0.009**	0.099
May	Oct	2.495	**0.008**	0.023	3.864	**0.011**	0.061	0.535	0.465	0.767	5.771	**0.016**	0.108
May	Sep	2.212	**0.027**	0.041	2.734	0.059	0.134	0.884	0.347	0.764	5.771	**0.016**	0.108
Nov	Oct	1.686	0.09	0.112	1.277	0.203	0.318	6.818	**0.009**	0.149	6.818	**0.009**	0.099
Nov	Sep	2.467	**0.008**	0.023	2.235	0.053	0.134	4.811	**0.028**	0.311	6.818	**0.009**	0.099
Oct	Sep	0.970	0.378	0.396	1.174	0.274	0.362	0.011	0.917	0.917	0.098	0.754	0.917

Table 7. Pairwise diversity analysis (alpha and beta) comparisons between months for 18S communities. Both Bray-Curtis and Weighted Unifrac methods were employed for beta diversity, as well as both species richness (Faith PD) and evenness for Alpha diversity. Significantly differences between months (p>0.0.5) are indicated in bold font.

		18S Amplicon											
		Beta Diversity - PERMANOVA						Alpha Diversity - KRUSTKAL-WALIS					
		Bray-Curtis			Weighted Unifrac			Faith PD			Evenness		
Group 1	Group 2	pseudo-F	p-value	q-value	pseudo-F	p-value	q-value	H	p-value	q-value	H	p-value	q-value
Apr	Aug	1.936	**0.033**	0.062	0.793	0.526	0.620	0.240	0.624	0.749	0.060	0.806	0.873
Apr	Dec	2.005	**0.009**	0.034	2.500	**0.008**	0.066	0.540	0.462	0.653	0.540	0.462	0.697
Apr	Feb	1.864	**0.006**	0.034	1.109	0.431	0.558	0.960	0.327	0.653	0.240	0.624	0.824
Apr	Jan	2.094	**0.006**	0.034	1.846	0.059	0.139	2.160	0.142	0.425	0.240	0.624	0.824
Apr	Jul	1.772	**0.056**	0.086	0.731	0.673	0.705	0.540	0.462	0.653	0.000	1.000	1.000
Apr	Jun	1.818	**0.049**	0.079	0.791	0.614	0.677	1.500	0.221	0.551	0.000	1.000	1.000
Apr	Mar	1.599	0.146	0.182	1.361	0.280	0.393	0.540	0.462	0.653	0.060	0.806	0.873
Apr	May	1.472	0.266	0.293	0.747	0.620	0.677	0.060	0.806	0.902	0.960	0.327	0.637
Apr	Nov	2.731	**0.012**	0.034	3.616	**0.006**	0.066	4.860	**0.027**	0.156	3.840	**0.050**	0.206
Apr	Oct	2.193	**0.007**	0.034	2.649	**0.005**	0.066	2.160	0.142	0.425	2.160	0.142	0.443
Apr	Sep	1.871	**0.035**	0.062	1.487	0.154	0.275	0.540	0.462	0.653	0.240	0.624	0.824
Aug	Dec	1.973	**0.009**	0.034	2.623	**0.024**	0.093	0.098	0.754	0.858	1.844	0.175	0.443
Aug	Feb	1.876	**0.002**	0.034	1.303	0.279	0.393	1.320	0.251	0.551	0.884	0.347	0.637
Aug	Jan	2.157	**0.007**	0.034	2.087	**0.036**	0.119	1.320	0.251	0.551	0.884	0.347	0.637
Aug	Jul	1.186	0.284	0.307	0.605	0.771	0.783	0.273	0.602	0.735	0.098	0.754	0.843
Aug	Jun	1.462	0.127	0.168	0.712	0.650	0.692	0.535	0.465	0.653	0.273	0.602	0.824
Aug	Mar	1.935	**0.016**	0.036	1.561	0.131	0.247	0.011	0.917	0.917	0.098	0.754	0.843
Aug	May	1.562	0.106	0.146	0.908	0.482	0.589	0.011	0.917	0.917	0.098	0.754	0.843
Aug	Nov	2.389	**0.006**	0.034	3.219	**0.012**	0.066	5.771	**0.016**	0.119	4.811	0.028	0.156
Aug	Oct	1.306	0.154	0.188	2.053	0.056	0.139	2.455	0.117	0.387	5.771	**0.016**	0.119
Aug	Sep	1.008	0.367	0.391	1.154	0.319	0.430	0.273	0.602	0.735	0.884	0.347	0.637
Dec	Feb	1.216	0.231	0.258	1.892	0.085	0.187	1.320	0.251	0.551	1.320	0.251	0.534
Dec	Jan	1.476	**0.069**	0.104	2.328	**0.032**	0.111	0.273	0.602	0.735	1.320	0.251	0.534
Dec	Jul	1.956	**0.009**	0.034	2.069	**0.014**	0.066	0.273	0.602	0.735	1.844	0.175	0.443
Dec	Jun	2.040	**0.007**	0.034	2.912	**0.007**	0.066	0.273	0.602	0.735	1.320	0.251	0.534
Dec	Mar	1.765	**0.035**	0.062	2.262	**0.009**	0.066	0.884	0.347	0.653	5.771	**0.016**	0.119
Dec	May	2.028	**0.010**	0.034	2.698	**0.010**	0.066	1.320	0.251	0.551	6.818	**0.009**	0.099
Dec	Nov	1.900	**0.012**	0.034	1.751	0.048	0.132	5.771	**0.016**	0.119	2.455	0.117	0.387
Dec	Oct	1.850	**0.026**	0.050	2.049	0.022	0.091	4.811	**0.028**	0.156	0.535	0.465	0.697
Dec	Sep	1.835	**0.013**	0.034	1.267	0.233	0.358	0.011	0.917	0.917	0.535	0.465	0.697
Feb	Jan	1.279	0.167	0.200	0.774	0.626	0.677	0.011	0.917	0.917	0.098	0.754	0.843
Feb	Jul	1.799	**0.018**	0.040	1.159	0.319	0.430	0.535	0.465	0.653	0.098	0.754	0.843
Feb	Jun	1.694	**0.039**	0.066	0.855	0.568	0.646	0.011	0.917	0.917	0.098	0.754	0.843
Feb	Mar	1.468	0.140	0.178	1.317	0.210	0.338	1.320	0.251	0.551	0.273	0.602	0.824
Feb	May	1.642	**0.042**	0.069	0.937	0.491	0.589	1.844	0.175	0.480	1.320	0.251	0.534
Feb	Nov	1.971	**0.026**	0.050	1.987	0.027	0.099	3.938	**0.047**	0.223	3.153	0.076	0.294
Feb	Oct	1.725	**0.007**	0.034	1.285	0.176	0.298	0.273	0.602	0.735	1.844	0.175	0.443
Feb	Sep	1.868	**0.007**	0.034	1.514	0.075	0.171	0.098	0.754	0.858	0.535	0.465	0.697
Jan	Jul	2.013	**0.012**	0.034	1.457	0.245	0.368	3.153	0.076	0.278	0.098	0.754	0.843
Jan	Jun	1.991	**0.025**	0.050	1.438	0.164	0.285	0.011	0.917	0.917	0.011	0.917	0.945
Jan	Mar	1.893	**0.037**	0.064	1.632	0.101	0.208	3.938	0.047	0.223	0.098	0.754	0.843
Jan	May	2.053	**0.005**	0.034	1.900	0.047	0.132	6.818	**0.009**	0.119	0.535	0.465	0.697
Jan	Nov	2.247	**0.010**	0.034	2.904	**0.013**	0.066	4.811	**0.028**	0.156	3.938	**0.047**	0.206
Jan	Oct	2.057	**0.016**	0.036	1.948	0.041	0.123	2.455	0.117	0.387	2.455	0.117	0.387
Jan	Sep	2.196	**0.014**	0.036	1.766	0.096	0.204	0.884	0.347	0.653	0.535	0.465	0.697
Jul	Jun	0.621	0.704	0.704	0.541	0.847	0.847	0.535	0.465	0.653	0.011	0.917	0.945
Jul	Mar	1.706	0.085	0.122	1.415	0.181	0.299	0.011	0.917	0.917	0.535	0.465	0.697
Jul	May	0.787	0.551	0.568	0.922	0.542	0.628	0.098	0.754	0.858	1.844	0.175	0.443
Jul	Nov	2.329	**0.009**	0.034	2.853	**0.004**	0.066	5.771	**0.016**	0.119	4.811	**0.028**	0.156
Jul	Oct	1.403	0.134	0.173	1.666	0.151	0.275	3.153	0.076	0.278	2.455	0.117	0.387
Jul	Sep	1.298	0.215	0.249	0.918	0.487	0.589	0.884	0.347	0.653	0.273	0.602	0.824
Jun	Mar	1.767	0.073	0.107	1.408	0.219	0.344	0.535	0.465	0.653	0.098	0.754	0.843
Jun	May	0.618	0.678	0.688	0.666	0.703	0.725	0.535	0.465	0.653	1.320	0.251	0.534
Jun	Nov	2.291	**0.007**	0.034	3.302	**0.010**	0.066	3.153	0.076	0.278	3.938	0.047	0.206
Jun	Oct	1.474	0.097	0.136	2.153	**0.038**	0.119	0.535	0.465	0.653	1.844	0.175	0.443
Jun	Sep	1.260	0.208	0.245	1.506	0.105	0.210	0.273	0.602	0.735	0.535	0.465	0.697
Mar	May	1.573	0.111	0.150	1.241	0.277	0.393	0.535	0.465	0.653	0.011	0.917	0.945
Mar	Nov	2.485	**0.016**	0.036	2.882	**0.012**	0.066	6.818	**0.009**	0.119	6.818	**0.009**	0.099
Mar	Oct	1.901	**0.013**	0.034	2.093	0.054	0.139	6.818	**0.009**	0.119	6.818	**0.009**	0.099
Mar	Sep	1.780	**0.020**	0.043	1.162	0.341	0.450	0.535	0.465	0.653	5.771	**0.016**	0.119
May	Nov	2.568	**0.011**	0.034	3.399	**0.010**	0.066	6.818	**0.009**	0.119	6.818	**0.009**	0.099
May	Oct	1.711	0.051	0.080	2.305	**0.007**	0.066	6.818	**0.009**	0.119	6.818	**0.009**	0.099
May	Sep	1.381	0.228	0.258	1.606	0.121	0.235	1.844	0.175	0.480	6.818	**0.009**	0.099
Nov	Oct	1.980	**0.007**	0.034	2.302	**0.019**	0.084	0.884	0.347	0.653	0.884	0.347	0.637
Nov	Sep	2.143	**0.009**	0.034	1.629	0.058	0.139	5.771	**0.016**	0.119	4.811	**0.028**	0.156
Oct	Sep	0.984	0.415	0.435	0.984	0.467	0.589	3.153	0.076	0.278	3.938	**0.047**	0.206

Table 8. Number of reads and percent of the total 18S eukaryote community for the top 15 diatom genera and top 10 arthropod genera, with the total for each group.

Top 15 Diatom Genera	Number of reads	Percent of 18S reads	Top 10 Arthopod Genera	Number of reads	Percent of 18S reads
Chaetoceros	572864	9.51	Pseudodiaptomus	171032	2.84
Unknown Raphid-pennate	400271	6.64	Canuella	46937	0.78
Thalassiosira	259223	4.30	Tigriopus	46874	0.78
Cyclotella	178982	2.97	Unknown Maxillopoda	23371	0.39
Unknown Bacillariophyta_X	71066	1.18	Sinocalanus	6259	0.10
Cymbella	66816	1.11	Cyclops	6024	0.10
Pleurosigma	62031	1.03	Paracalanus	3027	0.05
Minutocellus	30973	0.51	Ishizakiella	1364	0.02
Navicula	19098	0.32	Cyclopina	1150	0.02
Unknown Araphid-pennate	11441	0.19	Oithona	1133	0.02
Unknown Polar-centric-Mediophyce	8410	0.14	Remaining genera combined	3338	0.06
Melosira	8337	0.14			
Amphora	8134	0.14			
Pseudogomphonema	8045	0.13			
Cylindrotheca	6263	0.10			
Remaining genera combined	17214	0.29			
TOTAL	1729168	28.70	**TOTAL**	310509	5.16

Figures

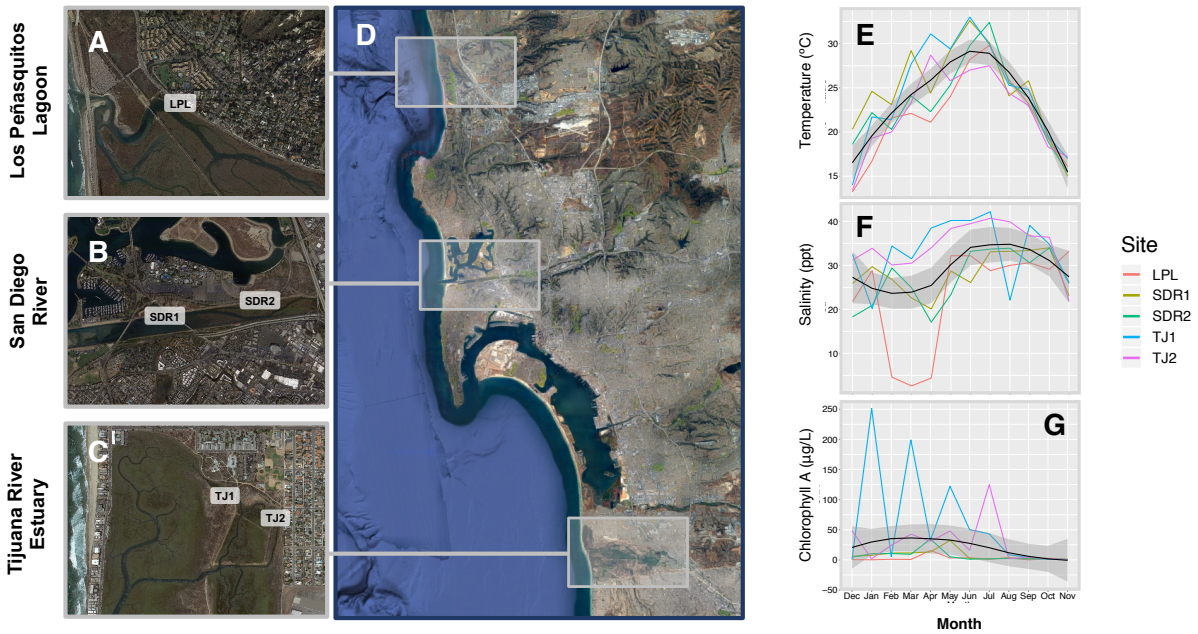

Figure 1. Location of the sampling sites **(A-C),** mapped in the context of the **(D)** San Diego region using Google Earth. Environmental conditions, including **(E)** temperature, **(F)** salinity**,** and **(G)** chlorophyll A, a proxy for photosynthetic organism abundance, were measured monthly at each site for one year from December 2015-November 2016.

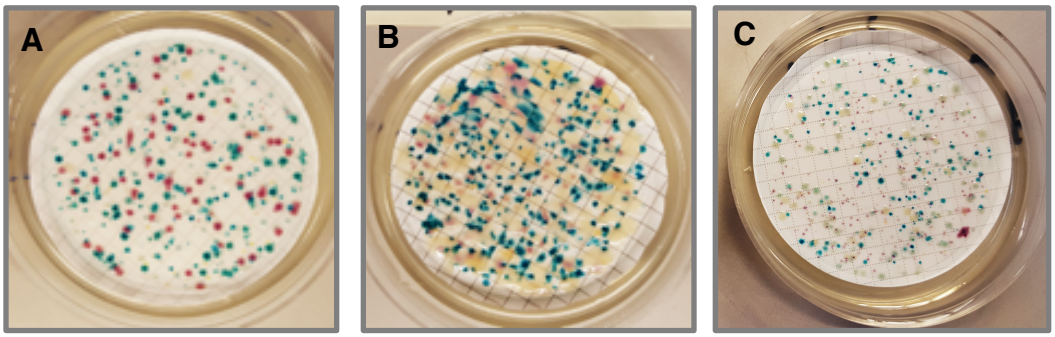

Figure 2. Isolates of putative pathogenic *Vibrio* species on CHROMagar *Vibrio* agar plates (filters are cellulose nitrate membranes, 0.45 µm pore size). **(A)** a 2 ml sample collected from LPL, **(B)** 2 ml volume collected from SDR1, both during May 2016. Both putative *Vibrio parahaemolyticus* (mauve) and putative *Vibrio vulnificus* (dark blue) colonies are evident, as are light blue colonies that may be *Vibrio cholerae*. **(C)** shows a 10 ml sample from TJ1 during December 2015, representing a sample where discerning between colony color is difficult, and illustrating the usefulness of genotyping.

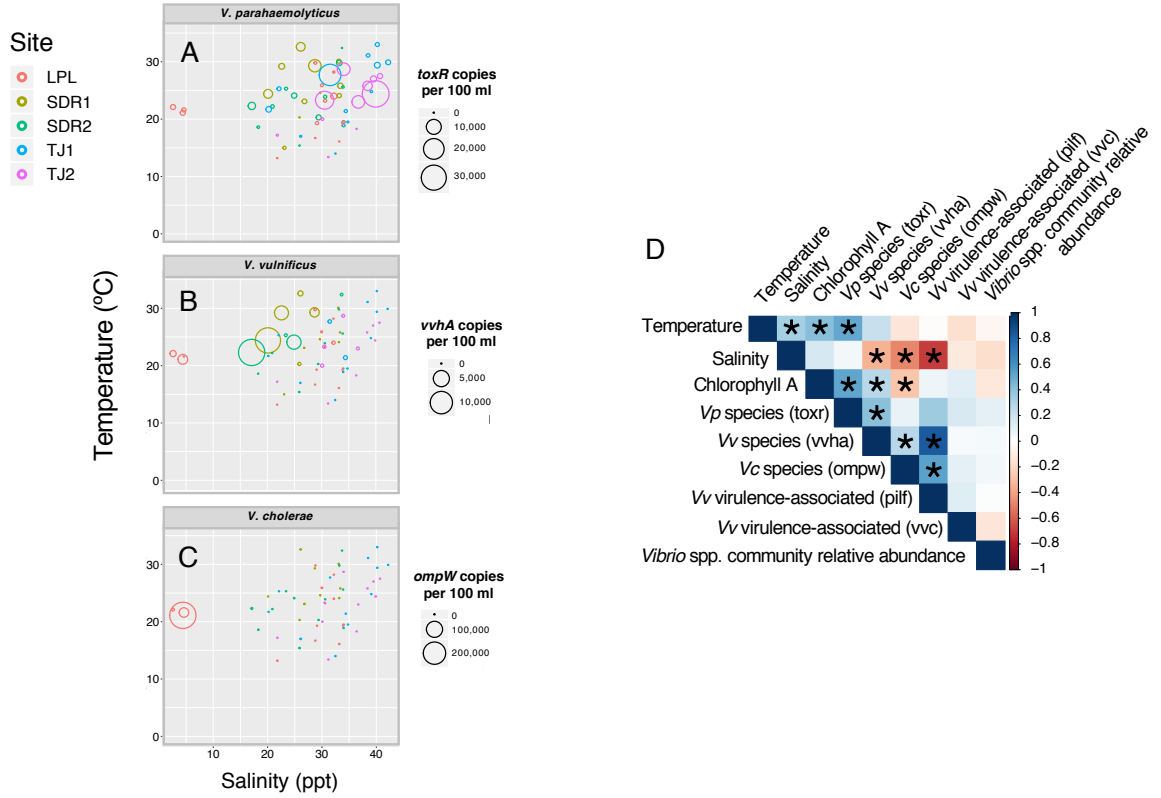

Figure 3. Number of single-genome copy genes (a proxy for cell numbers) per 100 ml detected by digital droplet PCR for **(A)** the *Vibrio parahaemolyticus (Vp)* species-specific gene target *toxR* **(B)** the *Vibrio vulnificus (Vv)* species-specific target *vvhA,* and **(C)** and the *Vibrio cholerae* species-specific target *ompW,* with marker size corresponding to copy number and color corresponding to site, plotted against temperature and salinity. **(D)** Spearman's rank correlation coefficients of associations between environmental variables temperature, salinity, and chlorophyll A, and *Vibrio* species and virulence gene targets. Blue represents a strong positive correlation, while red represents a strong negative correlation, and significant correlations ($p<0.05$) are denoted with an *.

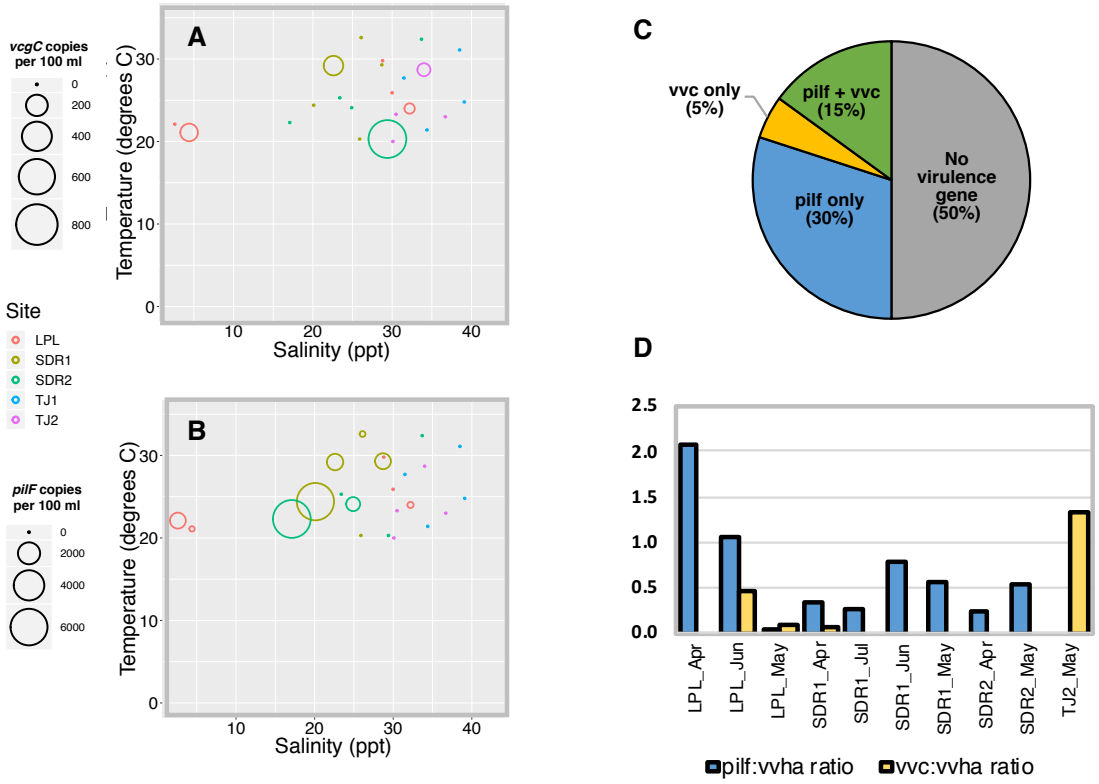

Figure 4. Number of copies detected per 100 ml by digital droplet PCR for the *Vibrio vulnificus* virulence-associated genes **(A)** *vcgC* and **(B)** *pilF,* plotted against temperature and salinity. **(C)** The percent of *V. vulnificus* samples where no virulence gene was detected, either *vcgC* or *pilF* were detected, or both were detected. **(D)** the ratio of the number of *pilF* and *vcgC* copies detected to total *V. vulnificus* determined by *vvhA* copy number.

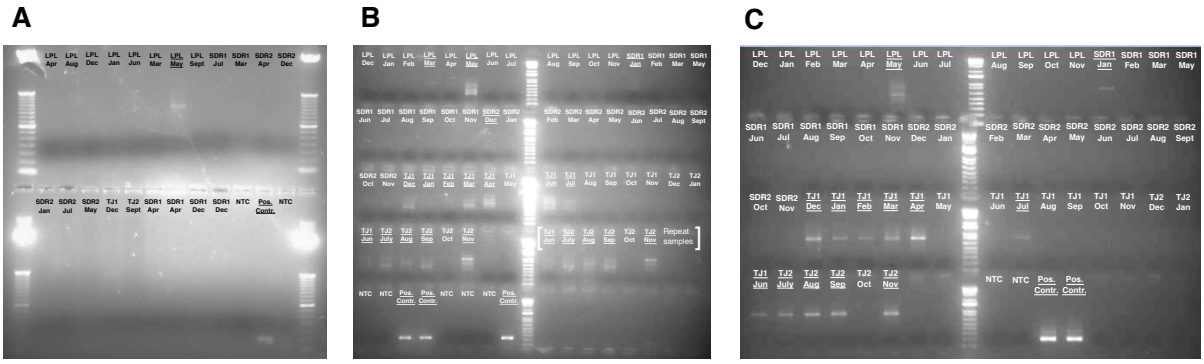

Figure 5. PCR gels depicting presence or absence of the *Vibrio* virulence-associated genes (A) *ctxA* (associated with *Vibrio cholerae)*, (B) *trh*, and (C) *tdh* (both associated with *V. parahaemolyticus*) in DNA from collected environmental samples. The *ctxA* gene (A) was assayed for samples positive for *V. cholerae* as determined by ddPCR assays, and reactions were run with an Invitrogen 50-bp ladder with the following sizes (in bp), smallest to largest: 50, 100,150,200,250,300,350,400,450,500,550,600,650,700,750,800,2,500). The *trh* and *tdh* genes, specific to *V. parahaemolyticus*, were assayed for all samples except SDR1 Dec and SDR1 Apr, and reactions were run with an Invitrogen 1 Kb plus ladder, with the following sizes (in bp), smallest to largest: (100,200,300,400,500,650,850,1,000,1,500,2,000,3,000,4,000,5,000,6,000,7,000,8,000, 10,000,15,000).

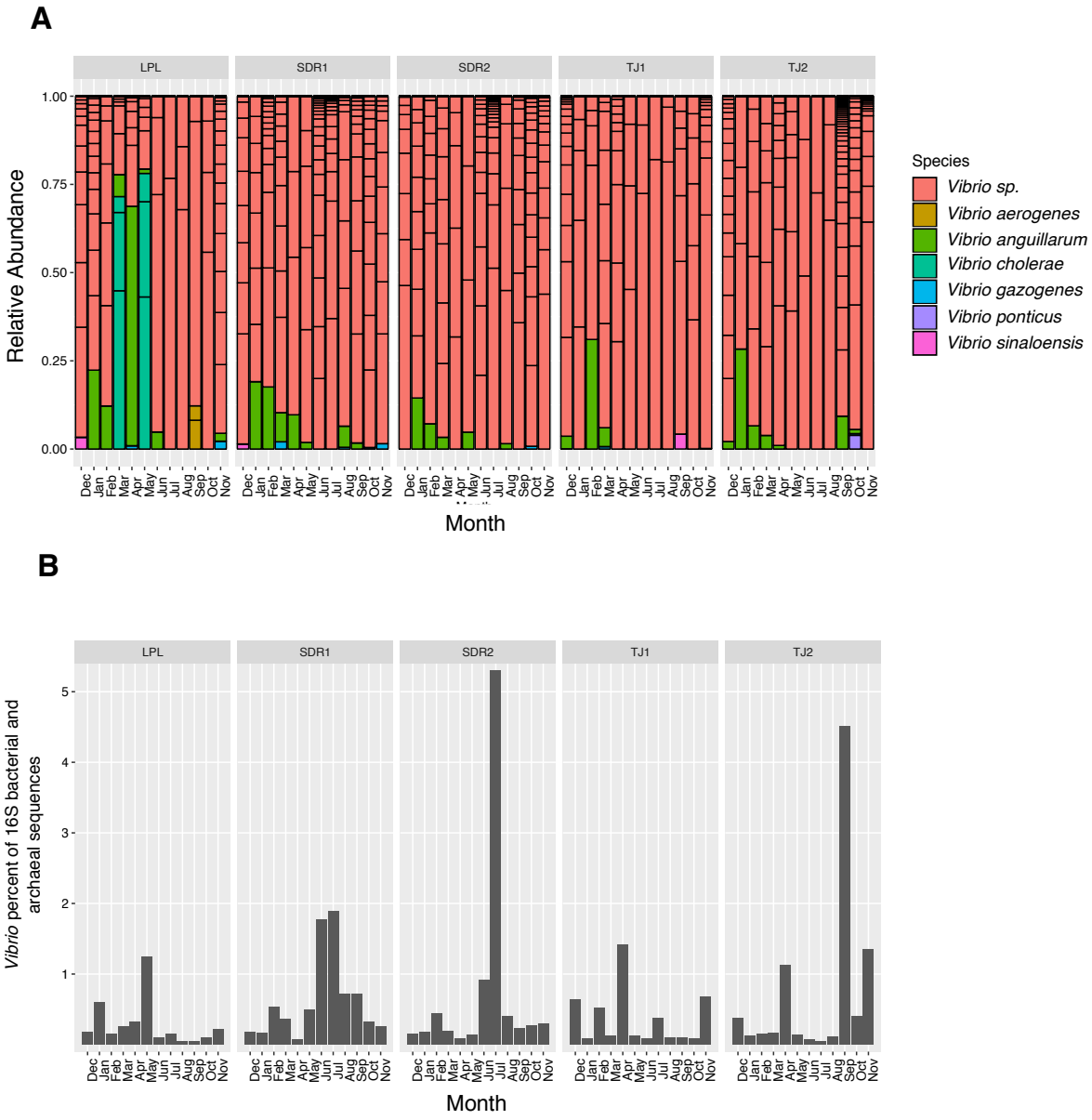

Figure 6. *Vibrio* relative abundance and community composition based on 16S sequences **(A)** The names and relative abundance of each *Vibrio* species in the *Vibrio* community classified through 16S sequencing using the SILVA database. Bars of the same color indicate different ASVs with the same species classification, with size indicating OTU relative abundance **(B)** . The percent of the 16S bacterial and archaeal community comprised of *Vibrio* bacteria for each month and site.

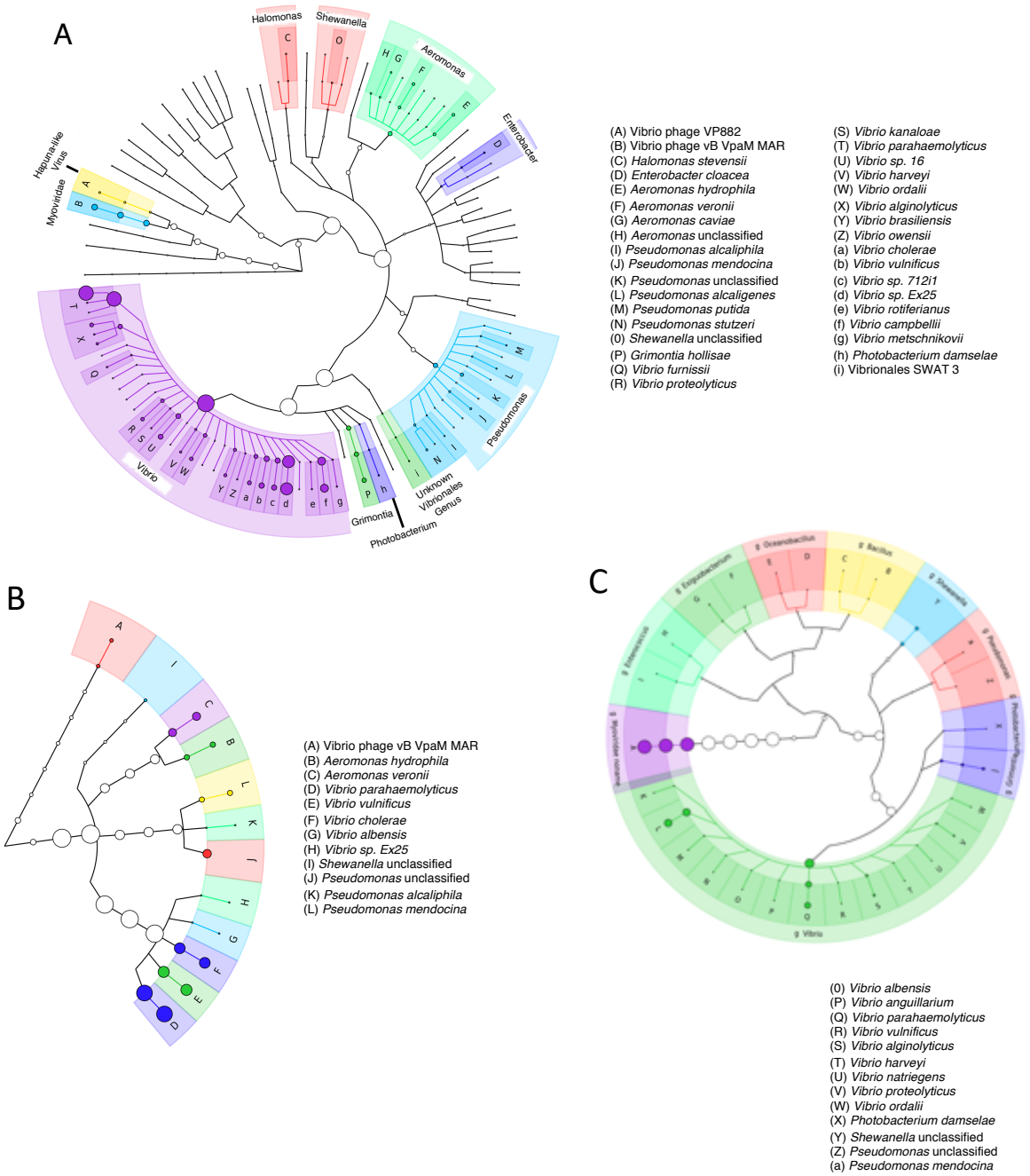

Figure 7. Composition and phylogenetic relationships of bacteria isolated on CHROMagar *Vibrio* plates. **(A)** The most abundant bacterial genera and species from all 26 samples combined, **(B)** Species composition of the LPL May site, which had high abundance of pathogenic *Vibrio* species and **(C)** TJ2 February, which had a high abundance of *Vibrio* bacteria and the Vibrio phage vB VpaM MAR.

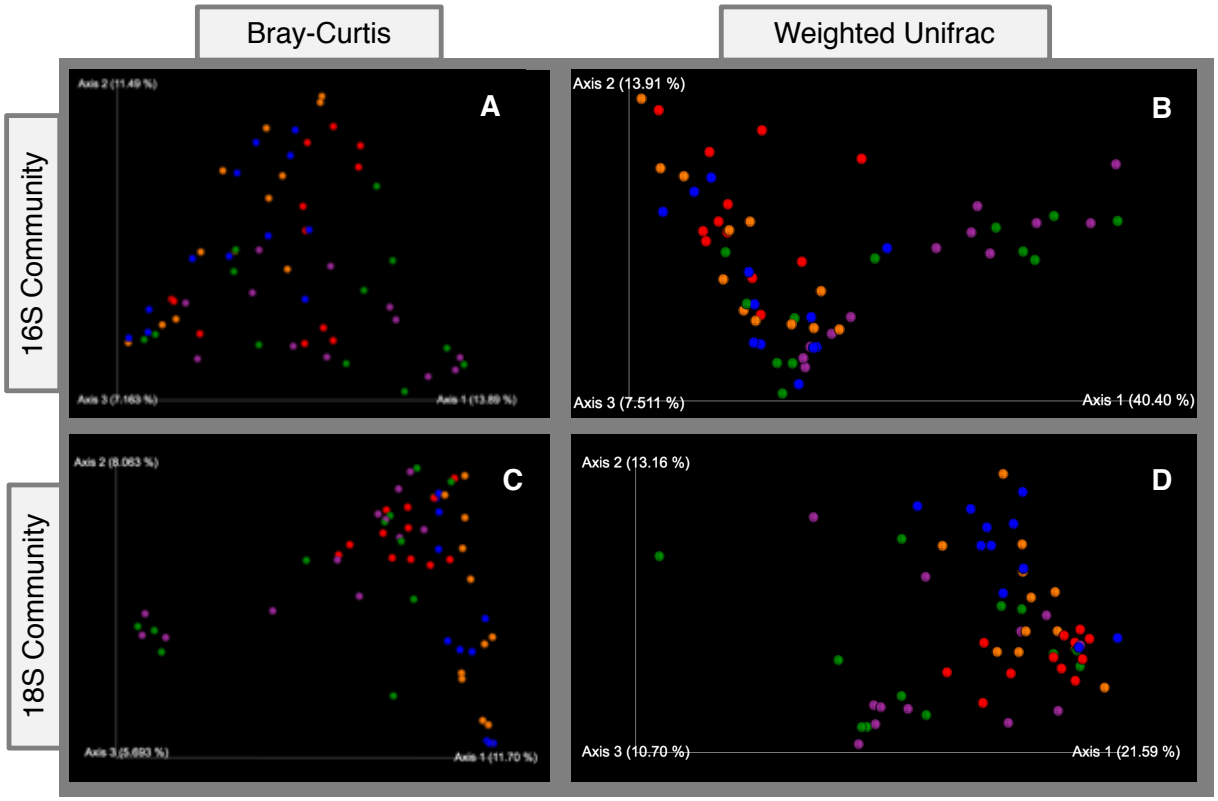

Figure 8. PCoA plots of community dissimilarity based on **(A)** Bray-Curtis dissimilarity of the 16S community, **(B)** Weighted Unifrac analysis of the 16S community, **(C)** Bray-Curtis dissimilarity of the 18S community, and **(D)** Weighted Unifrac analysis of the 18S community. Each dot represents a merged replicate sample for the 60 and 59 samples included in the 16S and 18S diversity analyses, respectively. The dot color indicates the site where the sample was collected: red = LPL, blue = SDR1, orange = SDR2, green = TJ1, and purple = TJ2

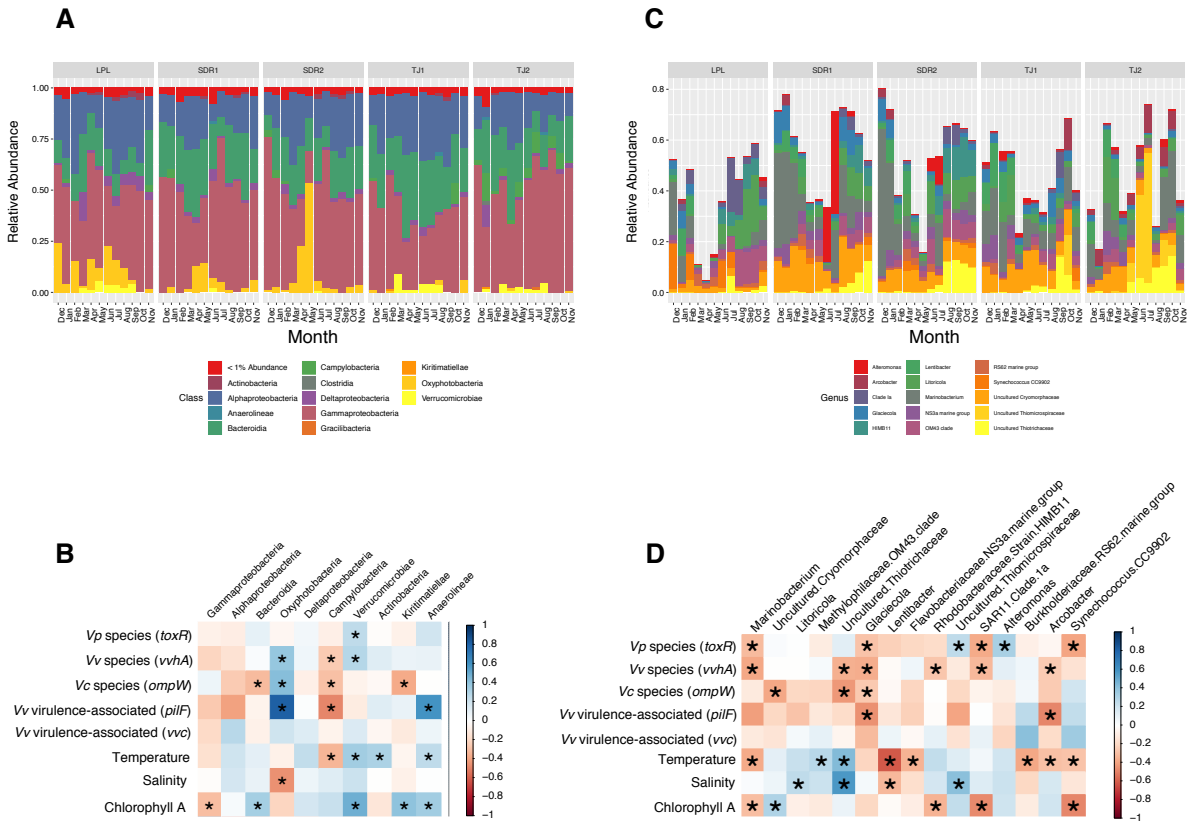

Figure 9. (A) Relative abundance of classes comprising >1% of the bacterial and archaeal 16S community for all sites and months, classified using the SILVA database, and **(B)** Spearman's rank correlations for the top 10 bacterial classes, *Vibrio* marker genes, and environmental factors. Blue indicates a positive correlation while red depicts a negative correlation, with significant ($p<0.05$) correlations denoted by *. **(C)** Top 15 most abundant bacterial genera by month and site, and **(D)** Spearman's rank correlations between these genera, *Vibrio* marker genes, and environmental variables16S

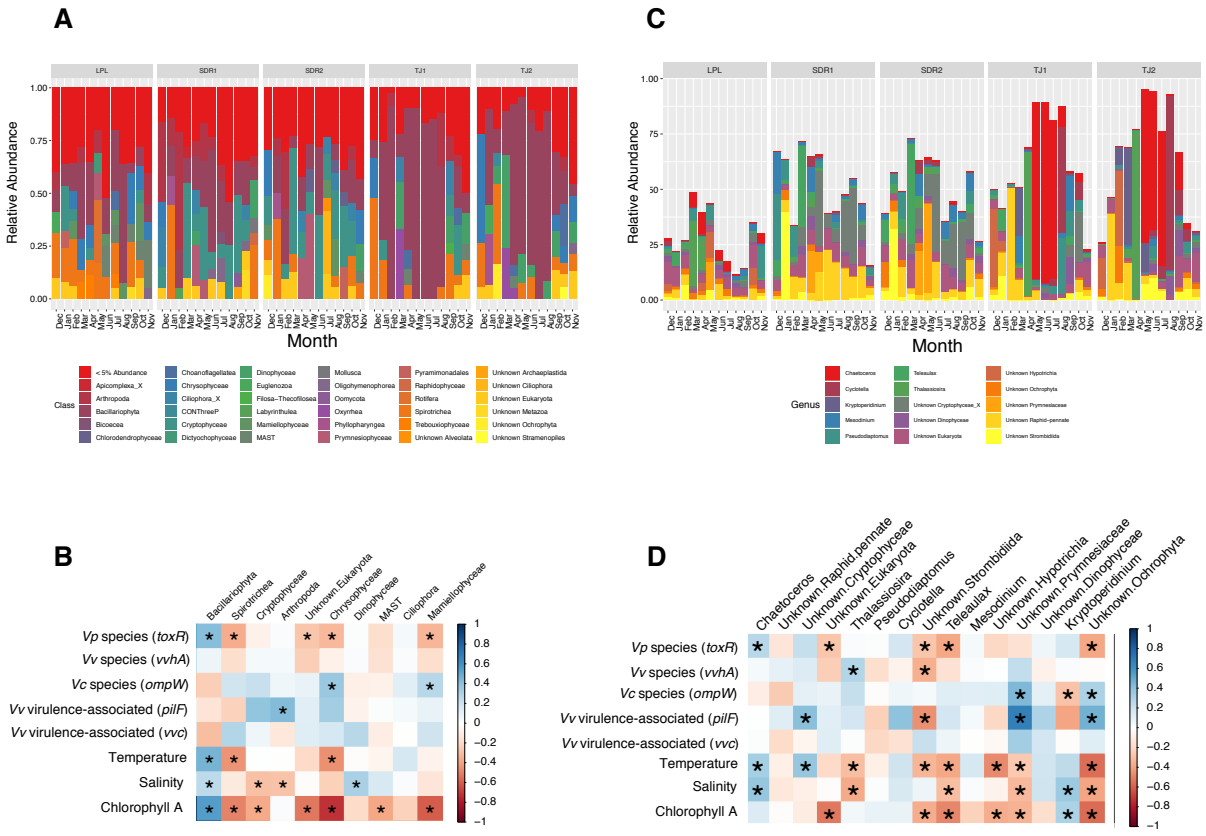

Figure 10. (A) Relative abundance of classes comprising >5% of the eukaryotic 18S community for all sites and months, classified using the PR2 database, and **(B)** Spearman's rank correlations for the top 10 eukaryote classes, *Vibrio* marker genes, and environmental factors. Blue indicates a positive correlation while red depicts a negative correlation, with significant ($p<0.05$) correlations denoted by *. **(C)** Top 15 most abundant eukaryotic genera by month and site, and **(D)** Spearman's rank correlations between these genera, *Vibrio* marker genes, and environmental variables.

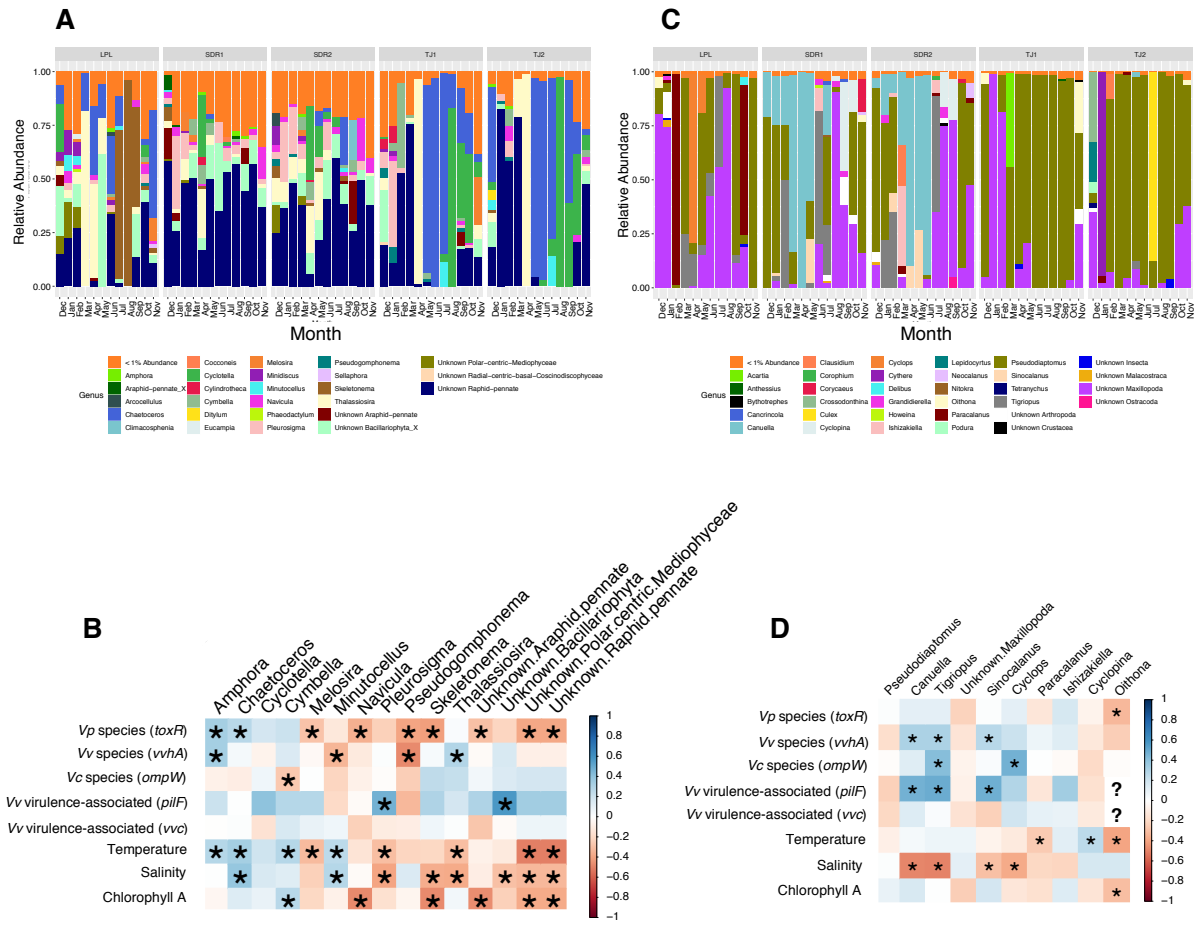

Figure 11. (A) Relative abundance of diatom genera comprising >1% of the eukaryotic 18S community for all sites and months, classified using the PR2 database, and **(B)** Spearman's rank correlations for the top 15 diatom genera, *Vibrio* marker genes, and environmental factors. Blue indicates a positive correlation while red depicts a negative correlation, with significant ($p<0.05$) correlations denoted by *. **(C)** Relative abundance of arthropod genera comprising >1% of the eukaryotic 18S community for all sites and months, classified using the PR2 database, and **(D)** Spearman's rank correlations between the top 10 most abundant arthropod genera, *Vibrio* marker genes, and environmental variables. A "?" denotes samples were the gene marker was not measured

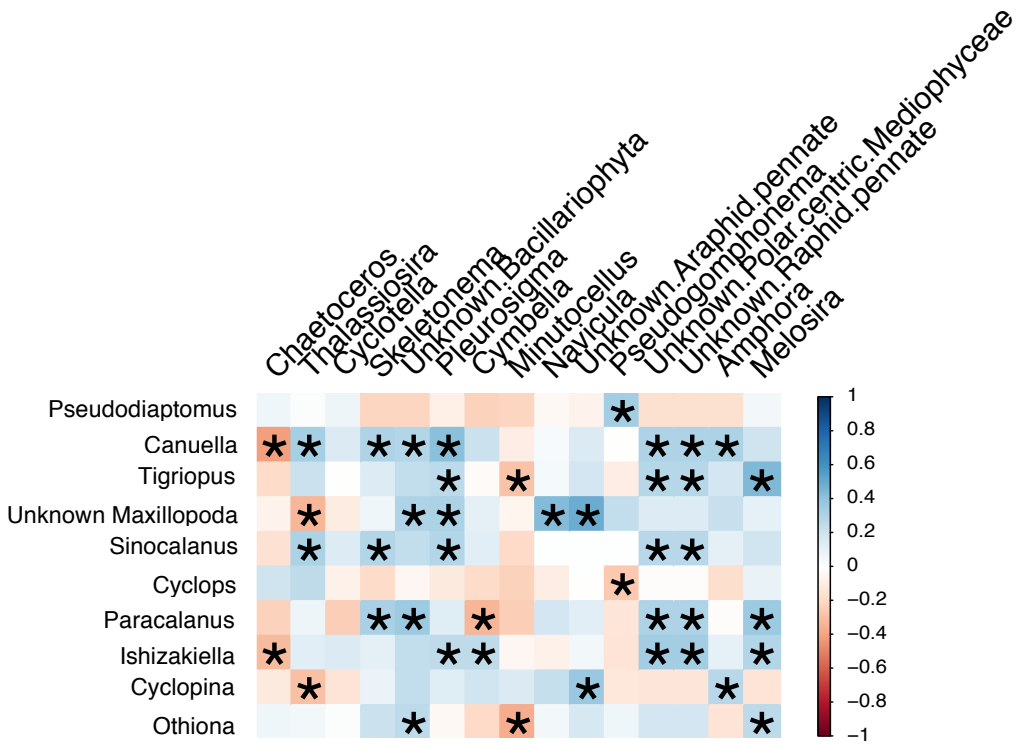

Figure 12. Spearman's rank correlations between the top 15 diatom genera and the top 10 arthropod genera. Blue indicates a positive correlation while red depicts a negative correlation, with significant ($p<0.05$) correlations denoted by *.