UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Spatial and temporal patterns of diversity and community composition in marine molluscan microbiomes

Permalink

https://escholarship.org/uc/item/88r2144d

Author

Neu, Alexander Theodore

Publication Date

2021

Supplemental Material

https://escholarship.org/uc/item/88r2144d#supplemental

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Spatial and temporal patterns of diversity and community composition in marine molluscan microbiomes

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

in

Biology

by

Alexander Theodore Neu

Committee in charge:

      Professor Eric Allen, Co-Chair
      Professor Kaustuv Roy, Co-Chair
      Professor Douglas Bartlett
      Professor Rachel Dutton
      Professor Jonathan Shurin

2021

The dissertation of Alexander Theodore Neu is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2021

DEDICATION

To my wife, Kelsey, for everything.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 5

LIST OF TABLES

CHAPTER 3

CHAPTER 4

CHAPTER 5

LIST OF SUPPLEMENTAL FILES

Dataset 5.1.xlsx

References for Dataset 5.1.docx

ACKNOWLEDGEMENTS

This dissertation is the result of more than five years of hard work with the support and encouragement of so many wonderful people. I would like to start by thanking my advisors, Kaustuv Roy and Eric Allen. Kaustuv, you've always pushed me to be my very best and supported me when I needed it. Your advice to, "go sit on the beach and think about it" has helped me make some of my most important decisions. Eric, you've always been supportive of my most off-the-wall ideas and ambitions and your lab meetings are a crucible of new projects and collaborations. You both took a chance on me, who had barely heard the word "microbiome" before starting my dissertation research, and you gave me all the tools for success. Because of you, I've been able to travel from Alaska to Panama and ask interesting questions about life on Earth. Thank you.

Thanks also to my committee: Doug, Rachel and Jon. Doug, thank you for your support and for always taking the time to talk to me about anything from career advice to the best way to culture anaerobic bacteria. Rachel, your questions and insights into my work have changed the way I speak and think about microbial ecology. Jon, thank you for being willing to both laugh with me and call me out when I need it. I will forever cherish my time on Team Hippo.

Thank you to the members of the Roy Lab, past and present. Elizabeth, you have become like family to me during these past five years and I couldn't have gone through this journey without you. The incredible undergrads: Ian Hughes, Kayla Nedd and Madeleine Tanda, thank you for all of your hard work and the joy you bring to science. To all of our summer students from ENLACE, STARS and various high schools, thank you for rolling with the punches as we all learned together.

Finally, thanks to the thousands of snails, mussels and clams who gave their lives in pursuit of this dissertation. You are the real heroes.

Chapter 1, in full, is a reprint of the material as it appears in Neu, A.T., Allen, E.E. and Roy, K., 2019. Diversity and composition of intertidal gastropod microbiomes across a major marine biogeographic boundary. Environmental Microbiology Reports, 11: 434-447. The dissertation author was the primary investigator and author of this paper.

Chapter 2, in full, is a reprint of the material as it appears in Neu, A.T., Allen, E.E. and Roy, K. 2021. Do host-associated microbes show a contrarian latitudinal diversity gradient? Insights from *Mytilus californianus*, an intertidal foundation host. Journal of Biogeography, 00:1-14. The dissertation author was the primary investigator and author of this paper.

Chapter 3, in full, is a reprint of the material as it appears in Neu, A.T., Hughes, I.V., Allen, E.E. and Roy, K., 2021. Decade-scale stability and change in a marine bivalve microbiome. Molecular Ecology, 30: 1237-1250. The dissertation author was the primary investigator and author of this paper.

Chapter 4, in full, is in review as Neu, A.T., Torchin, M.E., Allen, E.E. and Roy, K. Microbiome divergence of gastropod species separated by the Isthmus of Panama. In review; bioRxiv 2021.07.08.451645. The dissertation author was the primary investigator and author of this paper.

Chapter 5, in full, is in review as Neu, A.T., Allen, E.E. and Roy, K., 2021. Defining and quantifying the core microbiome: challenges and prospects. In review. The dissertation author was the primary investigator and author of this paper.

VITA

2015   Bachelor of Science, Biology, San Diego State University, San Diego, CA

2021   Doctor of Philosophy, Biology, University of California San Diego, La Jolla, CA

PUBLICATIONS

Neu, A.T., Allen, E.E., and Roy, K. The core microbiome: ecological and evolutionary perspectives. In review.

Neu, A.T., Torchin, M.E., Allen, E.E., and Roy, K. Microbiome divergence of geminate species of gastropods separated by the Isthmus of Panama. In review; bioRxiv: 2021.07.08.451645

Neu, A.T., Allen, E.E., and Roy, K. 2021. Do host-associated microbes show a contrarian latitudinal diversity gradient? Insights from *Mytilus californianus*, an intertidal foundation host. *Journal of Biogeography* 00:1-14.

Neu, A.T., Hughes, I.V., Allen, E.E., and Roy, K. 2021. Decade-scale stability and change in a marine bivalve microbiome. *Molecular Ecology* 30: 1237-1250.

Donham, E.M, Hamilton, S.L., Price, N.N., Kram, S., Kelly, E., Johnson, M.D., **Neu, A.T.**, and Smith, J. 2021. Experimental assessment of the impacts of ocean acidification and urchin grazing on benthic kelp forest assemblages. *Journal of Experimental Marine Biology and Ecology* 540: 151548.

Shurin, J.B., Aranguren Riaño, N., Echeverri Lopez, D., Jones, N.T., Laverde-R., O., **Neu**, **A.** and Pedroza Ramos, A. 2020. Accidental re-wilding: Ecosystem effects of the world's largest invasive animal. *Ecology* 101: e02991.

Neu, A.T., Allen, E.E., and Roy, K. 2019. Diversity and composition of intertidal gastropod microbiomes across a major marine biogeographic boundary. *Environmental Microbiology Reports* 11: 434-447.

Podell, S., Blanton, J.M., **Neu**, **A.**, Agarwal, V., Moore, B.S., and Allen, E.E. 2019. Pangenomic Comparison of Globally Distributed *Poribacteria* Associated with Sponge Hosts and Marine Particles. *ISME J* 13: 468-481

ABSTRACT OF THE DISSERTATION

Spatial and temporal patterns of diversity and community composition in marine molluscan microbiomes

by

Alexander Theodore Neu

Doctor of Philosophy in Biology

University of California San Diego, 2021

Professor Eric Allen, Co-Chair
Professor Kaustuv Roy, Co-Chair

Ecological communities have been shown to vary in some predictable ways through space, time, and along environmental gradients, suggesting there may be underlying "rules" in ecology which govern these patterns and processes. However, nearly all our knowledge of these trends comes from studies of large eukaryotes such as plants and animals, and we know very little about how these eukaryotic patterns compare to those of bacteria and archaea, the most dominant life forms on the planet.

The goal of this dissertation is to determine whether ecological patterns that are evident in plants and animals are also applicable to host-associated microbes. First, I investigated whether these microbes exhibit large-scale spatial trends in diversity and community composition that are concordant with those of their eukaryotic hosts. Specifically, I investigated changes in community composition across a marine biogeographic boundary and changes in diversity along a latitudinal transect. Results showed that microbiome compositions varied significantly between geographic sites, but that the identity of the host species played a greater role than geography in determining community composition. Further, microbes associated with the California mussel, *Mytilus californianus*, did not show a traditional latitudinal diversity gradient, and latitudinal diversity patterns varied based on microbial group and host body site. Next, I investigated whether host-associated microbial communities vary over time and in response to environmental change in similar ways to their eukaryotic hosts. I found that over an 11-year period, and in response to environmental change, microbial communities of the bean clam, *Donax gouldii*, significantly differed in composition, but not in richness. Further, I found that microbes did not regularly diverge in concordance with their intertidal gastropod hosts in the ~3.5 million years since the closure of the Isthmus of Panama, though this was dependent on the host taxa and the body site from which the microbes were collected. Finally, I investigated whether current methodologies for determining the core microbiome are guided by ecological and evolutionary principles and identify critical areas for future research. Overall, this dissertation shows that large-scale patterns in host-associated microbial taxa are often context-dependent and distinct from those of their hosts.

INTRODUCTION

For decades, ecologists have worked to uncover whether there are universal "rules" in ecology, which govern ecological patterns and processes (Lawton 1999). This research has shown that properties such species richness and community composition often vary in predictable ways through space, time and along environmental gradients, producing general patterns that may be the product of these universal rules. For example, across large spatial gradients, species richness often decreases with increasing latitude, a pattern known as the latitudinal diversity gradient (LDG), which has been documented in myriad plant and animal taxa (e.g., Hawkins, 2001; Hillebrand, 2004). Community composition, meanwhile, often decreases with increasing distance between communities (e.g., Soininen *et al.*, 2007) and some have identified particular areas, known as biogeographic breaks, where compositional turnover is incredibly high (Valentine 1966). Others have studied how communities are assembled and have outlined rules for this process (Weiher and Keddy 2001). However, for these patterns and processes to be truly universal, they must be broadly applicable across domains of life. Bacteria and archaea, though, are highly distinct from large eukaryotes in a number of ways, and therefore may follow distinctive ecological patterns from larger eukaryotes, though this has been woefully understudied to this point.

The few existing studies investigating large-scale ecological patterns in microbes have almost exclusively focused on free-living taxa, such as those in soils and seawater (e.g., Fuhrman *et al.*, 2008; Chu *et al.*, 2010). In some cases, these bacteria and archaea show patterns similar to those seen in larger eukaryotes. For example, some surveys of free-living marine microbes have shown that these taxa show a traditional LDG (Fuhrman et al. 2008, Ibarbalz et al. 2019) and a pattern of distance-decay (Sunagawa et al. 2015, Milici et al. 2016). In other cases, though,

microbial taxa present patterns significantly different from those in larger eukaryotes. For example, multiple studies of terrestrial soil microbes have shown that these taxa do not tend to show an LDG pattern, but instead their diversity is driven primarily by soil pH (Chu et al. 2010, Adamczyk et al. 2019). These contrasting results suggest that more work is necessary across the diversity of microbial habitats to understand the processes involved in structuring these communities.

Plant and animal hosts often harbor communities of hundreds to thousands of bacterial and archaeal taxa (Chiarello et al. 2020), which have been shown to play significant roles in host digestion, immune function and a number of other processes (McFall-Ngai et al. 2013). Members of these host-associated microbiomes may differ from larger eukaryotes and free-living microbes in (i) level of contact with the external environment, (ii) oxygen and nutrient concentrations, (iii) interaction with the host immune system and (iv) level of reliance on the host to survive and reproduce. Therefore, it is possible that these communities may show unique patterns which reflect different underlying processes from larger eukaryotes or free-living microbes, though this remains largely unexplored.

My dissertation aims to fill this current gap in our understanding of large-scale spatial and temporal patterns, and their underlying ecological processes, in host-associated microbial communities. Marine mollusks provide an excellent study system for this work, as many of the host species are (i) widespread, (ii) abundant in their local environments, (iii) easily and reproducibly collectible and (iv) often have limited mobility during their adult stages.

In Chapter 1 of my dissertation, I analyze the impact of a major marine biogeographic boundary (Pt. Conception, CA, USA) on the diversity and composition of marine gastropod microbiomes. This boundary has been shown to significantly impact the composition of intertidal

animal communities, primarily due to a lack of dispersal from south to north driven by oceanographic currents in the California Bight (Burton 1998). I found that this biogeographic boundary plays a significant role in the community composition of host-associated microbiomes within individual host species, but that host species identity is the primary factor determining microbiome composition. Richness of the microbiome, however, was largely unaffected by Pt. Conception across different host species. The results of this study show the impacts of large-scale spatial differentiation on the intertidal gastropod microbiome and introduce the methods, both laboratory and analytical, that will be used throughout the rest of the dissertation. Chapter 1 is published as: Neu, A.T., Allen, E.E. and Roy, K., 2019. Diversity and composition of intertidal gastropod microbiomes across a major marine biogeographic boundary. Environmental Microbiology Reports, 11: 434-447.

Chapter 2 expands on the large-scale spatial analysis of intertidal molluscan microbiome introduced in Chapter 1, as I determine whether the LDG is present in microbial taxa associated with the California mussel, *Mytilus californianus*, a foundation species in the intertidal. I sampled microbial communities from the gill tissue and shell-surface biofilms of these mussels from 12 sites spanning >24 degrees of latitude, making this the largest latitudinal sampling of host-associated microbiomes to-date. Further, I used diversity metrics analogous to those used in studies of plants and animals to determine latitudinal diversity patterns. Finally, I tested whether individual microbial clades, rather than the aggregate microbial community, showed an LDG signal. The results of this study showed that the gill microbiome presented either a weak, significant LDG pattern in alpha diversity or no pattern at all, while the shell-surface microbiome presented an inverse LDG. Gamma diversity, on the other hand, largely showed a hump-shaped pattern across latitudes. Individual microbial taxa varied in their latitudinal diversity patterns,

suggesting that whole-microbiome analysis may not be the most appropriate level to measure large-scale ecological patterns, which was influential in developing later chapters of this dissertation. Chapter 2 is published as: Neu, A.T., Allen, E.E. and Roy, K. 2021. Do host-associated microbes show a contrarian latitudinal diversity gradient? Insights from *Mytilus californianus*, an intertidal foundation host. Journal of Biogeography, 00:1-14.

Chapter 3 of the dissertation switches focus to temporal changes in the microbiomes of marine mollusks. Here I aimed to determine whether the diversity and composition of the bean clam (*Donax gouldii*) microbiome changed over an 11-year time frame or in response to changes in temperature and phytoplankton abundance. I found that the diversity of these communities was remarkably stable over time, though the composition of the community changed significantly between sampling years. I also showed that samples preserved long-term in ethanol and under refrigeration can be useful for understanding long-term dynamics of host microbiomes and should be explored further. Chapter 3 is published as: Neu, A.T., Hughes, I.V., Allen, E.E. and Roy, K., 2021. Decade-scale stability and change in a marine bivalve microbiome. Molecular Ecology, 30: 1237-1250.

Chapter 4 examines how very large tectonic and environmental changes in the geological past have shaped the microbiomes of marine species today. Specifically, I show how the closure of the Isthmus of Panama ~3.5 mya impacted the microbial communities associated with geminate host species. These geminate hosts were once populations of a single species, which were separated from one another by the closure of the Isthmus and have since diverged into new species in their current environments. I aimed to determine whether a similar phenomenon occurred in the microbiomes of these host species, with geminate hosts containing more similar microbial communities, or whether the microbiomes were more significantly shaped by their

present-day environmental conditions. Chapter 4 is available from bioRxiv (2021.07.08.451645) and is currently under review for publication.

Chapter 5 contains a perspective on the core microbiome, which is a term that has been widely used in the microbial ecology literature but has many operational definitions. Here I present the history of the term, some of the myriad ways it has been used to-date and provide suggestions for integrating ecological and evolutionary theory into the definition and quantification of the core microbiome. This work was informed by much of the research on microbiomes from the previous chapters and is currently under review for publication.

References

Adamczyk, M., F. Hagedorn, S. Wipf, J. Donhauser, P. Vittoz, C. Rixen, A. Frossard, J. P. Theurillat, and B. Frey. 2019. The soil microbiome of Gloria Mountain summits in the Swiss Alps. Frontiers in Microbiology 10:1080.

Burton, R. S. 1998. Intraspecific phylogeography across the Point Conception biogeographic boundary. Evolution 52:734–745.

Chiarello, M., J. C. Auguet, N. A. J. Graham, T. Claverie, E. Sucré, C. Bouvier, F. Rieuvilleneuve, C. X. Restrepo-Ortiz, Y. Bettarel, S. Villéger, and T. Bouvier. 2020. Exceptional but vulnerable microbial diversity in coral reef animal surface microbiomes. Proceedings of the Royal Society B: Biological Sciences 287:20200642.

Chu, H., N. Fierer, C. L. Lauber, J. G. Caporaso, R. Knight, and P. Grogan. 2010. Soil bacterial diversity in the Arctic is not fundamentally different from that found in other biomes. Environmental Microbiology 12:2998–3006.

Fuhrman, J. A., J. A. Steele, I. Hewson, M. S. Schwalbach, M. V. Brown, J. L. Green, and J. H. Brown. 2008. A latitudinal diversity gradient in planktonic marine bacteria. Proceedings of the National Academy of Sciences of the United States of America 105:7774–7778.

Hawkins, B. A. 2001. Ecology's oldest pattern? Trends in Ecology & Evolution 16:470.

Hillebrand, H. 2004. On the Generality of the Latitudinal Diversity Gradient. The American Naturalist 163:192–211.

Ibarbalz, F. M., N. Henry, M. C. Brandao, S. Martini, G. Busseni, H. Byrne, L. P. Coelho, H.

Endo, J. M. Gasol, A. C. Gregory, F. Mahe, J. Rigonaro, M. Royo-Llonch, G. Salazar, I. Sanz-Saez, E. Scalco, D. Soviadan, A. A. Zayed, A. Zingone, K. Labadie, J. Ferland, C. Marec, S. Kandels, M. Picheral, C. Dimier, J. Poulain, S. Pisarev, M. Carmichael, S. Pesant, T. O. Coordinators, M. Babin, E. Boss, D. Iudicone, O. Jaillon, S. G. Acinas, H. Ogata, E. Pelletier, L. Stemmann, M. B. Sullivan, S. Sunagawa, L. Bopp, C. de Vargas, L. Karp-Boss, P. Wincker, F. Lombard, C. Bowler, and L. Zinger. 2019. Global Trends in Marine Plankton Diversity across Kingdoms of Life. Cell 179:1084–1097.

Lawton, J. H. 1999. Are There General Laws in Ecology? Oikos 84:177–192.

McFall-Ngai, M., M. G. Hadfield, T. C. G. Bosch, H. V. Carey, T. Domazet-Lošo, A. E. Douglas, N. Dubilier, G. Eberl, T. Fukami, S. F. Gilbert, U. Hentschel, N. King, S. Kjelleberg, A. H. Knoll, N. Kremer, S. K. Mazmanian, J. L. Metcalf, K. Nealson, N. E. Pierce, J. F. Rawls, A. Reid, E. G. Ruby, M. Rumpho, J. G. Sanders, D. Tautz, and J. J. Wernegreen. 2013. Animals in a bacterial world, a new imperative for the life sciences. Proceedings of the National Academy of Sciences of the United States of America 110:3229–3236.

Milici, M., J. Tomasch, M. L. Wos-oxley, J. Decelle, R. Jáuregui, H. Wang, Z. Deng, I. Plumeier, H. Giebel, T. H. Badewien, M. Wurst, D. H. Pieper, and M. Simon. 2016. Bacterioplankton Biogeography of the Atlantic Ocean: A Case Study of the Distance-Decay Relationship. Frontiers in Microbiology 7:590.

Soininen, J., R. Mcdonald, and H. Hillebrand. 2007. The distance decay of similarity in ecological communities. Ecography 30:3–12.

Sunagawa, S., L. P. Coelho, S. Chaffron, J. R. Kultima, K. Labadie, G. Salazar, B. Djahanschiri, G. Zeller, D. R. Mende, A. Alberti, F. M. Cornejo-Castillo, P. I. Costea, C. Cruaud, F. D'Ovidio, S. Engelen, I. Ferrera, J. M. Gasol, L. Guidi, F. Hildebrand, F. Kokoszka, C. Lepoivre, G. Lima-Mendez, J. Poulain, B. T. Poulos, M. Royo-Llonch, H. Sarmento, S. Vieira-Silva, C. Dimier, M. Picheral, S. Searson, S. Kandels-Lewis, C. Bowler, C. de Vargas, G. Gorsky, N. Grimsley, P. Hingamp, D. Iudicone, O. Jaillon, F. Not, H. Ogata, S. Pesant, S. Speich, L. Stemmann, M. B. Sullivan, J. Weissenbach, P. Wincker, E. Karsenti, J. Raes, S. G. Acinas, and P. Bork. 2015. Structure and function of the global ocean microbiome. Science 348:1261359.

Valentine, J. W. 1966. Numerical analysis of marine molluscan ranges on the extratropical Northeastern Pacific shelf. Limnology and Oceanography 11:198–211.

Weiher, E., and P. Keddy, editors. 2001. Ecological assembly rules: perspectives, advances, retreats. Cambridge University Press.

CHAPTER 1

Diversity and composition of intertidal gastropod microbiomes across a major marine

biogeographic boundary

Brief Report

Diversity and composition of intertidal gastropod microbiomes across a major marine biogeographic boundary

Alexander T. Neu, [1]* Eric E. Allen[2,3] and Kaustuv Roy[1]

[1]*Section of Ecology, Behavior and Evolution, Division of Biological Sciences, University of California San Diego, La Jolla, California, USA.*
[2]*Section of Molecular Biology, Division of Biological Sciences, University of California San Diego, La Jolla, California, USA.*
[3]*Marine Biology Research Division, Scripps Institution of Oceanography, University of California San Diego, La Jolla, California, USA.*

Summary

Marine biogeographic boundaries act as barriers to dispersal for many animal species, thereby creating distinctive faunas on either side. However, how such boundaries affect the distributions of microbial taxa remains poorly known. To test whether biogeographic boundaries influence the diversity and composition of host-associated microbiota, we analysed the microbiomes of three species of common intertidal gastropods at two sites separated by the biogeographic boundary at Point Conception (PtC), CA, using 16S rRNA gene sequencing. Our results show that each host species shows microbiome compositional specificity, even across PtC, and that alpha diversity does not change significantly across this boundary for any of the gastropod hosts. However, for two of the host species, beta diversity differs significantly across PtC, indicating that there may be multiple levels of organization of the marine gastropod microbiome. Overall, our results suggest that while biogeographic boundaries do not constrain the distribution of a core set of microbes associated with each host species, they can play a role in structuring the transient portion of the microbiome.

Introduction

Despite a long history of research on microbes associated with marine invertebrates (e.g. Galli and Giese 1959; Beeson and Johnson 1967; Harris 1993), the diversity, composition and functional roles of host-associated microbial communities are just beginning to be quantified (Littman *et al*., 2009; Green and Barnes, 2010; Pfister *et al*., 2010; Fan *et al*., 2012). Recent studies suggest that microbiomes of marine invertebrates can be involved in vital functions ranging from disease protection in corals (Reshef *et al*., 2006; Vega Thurber *et al*., 2009) and vitamin biosynthesis in sponges (reviewed in Hentschel *et al*., 2012) to nutrient acquisition in chemosymbiotic bivalves (Batstone and Dufour, 2016). However, in spite of their potential importance to host biology, microbiomes of the vast majority of marine species still remain undescribed. Furthermore, how microbiomes vary across different populations of individual host species, especially when such populations are arrayed along major environmental gradients, also remains poorly studied. Experimental studies suggest that elevated seawater temperature as well as ocean acidification can substantially change the composition of microbiomes of invertebrate hosts (Webster *et al*., 2013; Lesser *et al*., 2016), which in turn has the potential to alter the nature of host–microbiome interactions, an important aspect of host health (Apprill, 2017). Thus, better information about population-level variability in the diversity and compositions of microbiomes may be particularly important for better understanding how populations of individual hosts, as well as their microbiomes, are likely to respond to global change.

Marine molluscs are one of the most diverse and ecologically important groups of marine invertebrates, with many species providing a variety of ecosystem services including supporting commercial and artisanal fisheries. Information about the diversity and compositions of microbes associated with molluscan hosts still remains scarce, although microbiomes of some bivalve (Prieur *et al*., 1990; Betcher *et al*., 2012; King *et al*., 2012;

Lokmer *et al*., 2016; Pierce *et al*., 2016; Arfken *et al*., 2017) and gastropod (Dudek *et al*., 2014; Zbinden *et al*., 2014; Aronson *et al*., 2017; Cicala *et al*., 2018) species have been quantified. Moreover, existing analyses of marine gastropod microbiomes have primarily focused on the digestive tract, with only one study including gill-specific microbiome data (Zbinden *et al*., 2014). Microbial communities associated with other parts of the marine gastropod anatomy remain largely unknown. Overall, the data available for marine molluscs are not sufficient to evaluate the relative roles of geographic location and host identity in structuring the microbiome of a host. While studies focusing on microbiomes of multiple species at a particular site show that host identity plays a major role in determining microbiome composition (Davis *et al*., 2013; Aronson *et al*., 2017; Cicala *et al*., 2018), analyses of bivalves in aquaculture settings have shown that rearing the same host in different locations tends to change the relative abundances of a number of microbial phyla between individuals, which suggests that the external environment can also have a substantial effect on the composition of molluscan microbiomes (Trabal *et al*., 2012; Trabal Fernández *et al*., 2014).

In this study, we identify geographical and host-specific components of the microbiomes of rocky intertidal gastropods by sampling two different populations of multiple host species across a major marine biogeographic boundary. Our focal host species include *Lottia gigantea* (Owl limpet), *Littorina keenae* (Rough periwinkle) and *Chlorostoma funebralis* (Black turban snail). All of the species analysed here are algal grazers that reproduce by broadcast spawning and are important components of rocky intertidal communities in the temperate northeastern Pacific. *Lottia gigantea* is a common species on wave-exposed sites from northern California (USA) to Baja California (Mexico). This species is a protandric hermaphrodite, meaning all individuals begin their lives as males and change sex to female as they grow larger and older (Fenberg and Roy, 2012). Adult females of this species are territorial and actively maintain their territories by removing any other potential settlers (Stimson, 1970). They are also thought to 'garden' and graze on microalgae within their territory (Connor, 1986). Individuals of *L. gigantea* are commonly harvested as food items, which typically removes the largest females from the population (known as size-selective harvesting), resulting in lower overall body size in harvested populations (Roy *et al*., 2003). *L. keenae* typically lives in the high intertidal zone from Oregon (USA) to Baja California (Mexico) with some populations inhabiting the splash zone. It is highly abundant at most sites and although primarily an algal grazer, *L. keenae* has been observed to ingest small zooplankton and even portions of the rocky substrate it inhabits (Dahl, 1964). *Chlorostoma funebralis* is also an herbivorous gastropod typically inhabiting the low-to-mid intertidal zone and is present in relatively high densities across its range from British Columbia (Canada) to Baja California (Mexico). *Chlorostoma funebralis* populations are also commonly harvested by humans, historically as food (Erlandson *et al*., 2015), but more recently as bait items (Bednar and Trulio, 2017).

All three of these species range across the major biogeographic boundary at Pt. Conception (PtC) that separates the Californian and Oregonian provinces (Valentine 1966; Supporting Information Fig. S1). This boundary coincides with major changes in coastal oceanography and is characterized by clustering of geographic range end-points of many species (Roy *et al*., 1994). Populations and species on the northern side of this boundary experience colder water temperature, greater upwelling intensity, and higher wave energy (Caldwell *et al*., 1986; Harms and Winant, 1998). In contrast, south of this boundary, the Southern California Countercurrent brings warmer water along the coast and contributes to lower upwelling intensity and wave exposure (Owen, 1980). Thus, populations of the host species at our site north of PtC (Jalama Beach County Park, Lompoc, CA (JB; 34.5131 N, −120.5041 W)) experience very different environmental conditions compared with our site south of this boundary (Scripps Coastal Reserve, La Jolla, CA (SIO; 32.8661 N, 117.2543 W)). Little is known at present about the role of marine biogeographic boundaries in structuring spatial patterns of microbial species richness and composition, but a recent study suggests that richness of free-living bacterial, archaeal and microeukaryotic taxa can change across oceanographic boundaries (Raes *et al*. 2018).

In addition to the three host species mentioned above, we also sampled individuals of *Chlorostoma eiseni* (Banded turban snail), another common rocky intertidal gastropod, at our southern site (see below). This species ranges from Los Angeles, CA (USA) to Baja California (Mexico), and thus does not cross PtC, but is included here, since it is closely related to *C. funebralis* (Hellberg, 1998), thus allowing us to test for phylogenetic conservatism of microbiome compositions in intertidal gastropods.

Results

Sequencing results

A total of 1 958 451 sequences assigned to 15 323 unique operational taxonomic units (OTUs) (at 98% sequence similarity) were retained from 52 samples, giving an average count of 37 662 sequences per sample in the unrarefied dataset. All further analyses were conducted by rarefying all samples to 10 200 sequences, which removed one algal turf sample, two *L. keenae*

samples from SIO, two *L. keenae* samples from JB and one *C. eiseni* from SIO. A Kruskal–Wallis test showed that microbiome alpha diversity, as measured by the Shannon index (H′), was significantly different between the gastropod host species sampled at JB ($X^2 = 10.462$, $p = 0.033$), but not at SIO ($X^2 = 7.745$, $p = 0.171$). Similarly, community evenness, as measured by Pielou's J (Pielou, 1966), was significantly different between the gastropods at JB ($X^2 = 10.499$, $p = 0.033$), but not at SIO ($X^2 = 8.915$, $p = 0.113$). Within each host species, however, neither unique overall diversity (H′) nor community evenness (J) differed significantly between the two sites. Samples collected from JB showed a trend toward higher H′ than SIO, but this was not significant for any of the hosts ($p > 0.05$). Environmental samples (turf algae and water) showed high variability among sites in the calculated diversity metrics (Table 1).

Microbiome composition

Proteobacteria was the most abundant phylum across nearly all species and sites sampled, including algal turf and seawater, making up between 24.6% and 69.6% of the relative abundances (Fig. 1). *Chlorostoma funebralis* samples from SIO provided the only exception to this, with the combined members of the phyla Tenericutes and Bacteroidetes making up between 25% and 50% of sequences. Other abundant phyla include: Fusobacteria, which showed increased abundance in *L. gigantea* [27% (± 27%) and 21% (± 12%) average relative abundance at SIO and JB, respectively] as compared with the other hosts (< 3%), though variation is high among individuals (Fig. 1); Bacteroidetes, which comprised between 4.8% and 26.2% of sequences across all sample types and Verrucomicrobia, which was common in *L. keenae*, particularly at SIO, where its average relative abundance was 20.5% (± 20%) but reaching up to 51% in some individuals. Different classes within the Proteobacteria – Alphaproteobacteria and Gammaproteobacteria – tended to be dominant in different hosts. *Chlorostoma funebralis* and *C. eiseni* were enriched in Alphaproteobacteria, while *L. gigantea* was enriched in Gammaproteobacteria, with *L. keenae* samples showing nearly even distribution of the two classes. At a lower taxonomic level, the Gammaproteobacteria associated with *L. gigantea* were primarily members of the genus *Vibrio*, while the water and algal samples contained undefined Gammaproteobacteria and members of the order Alteromonadales. Further differentiating the microbial communities of each gastropod host was the high abundance of genus *Mycoplasma* in the species *C. funebralis*, *C. eiseni* and *L. gigantea.* This genus was found in low numbers in *L. keenae*, which lives higher in the intertidal, and water samples, but was highly abundant in the three hosts that are mid-intertidal grazers. Archaeal abundance was highly variable, with the phylum Crenarchaeota contributing between 0% and 9% of sequences depending on the host sampled. Using the NCBI 16S rRNA database, nearly all Archaea sequences were matched to the genus *Nitrosopumilus.* Among all the gastropods sampled, *L. keenae* from JB contained the lowest proportion of Archaea, with 0.4% (±1%) average relative abundance. Finally, a total of 196 OTUs, containing 6282 sequences (0.0016%), could not be assigned at any level.

Non-metric multidimensional scaling (NMDS) of Bray–Curtis dissimilarity indicated clustering of microbiomes by host species within the whole dataset (Fig. 2). This was confirmed by a permutational multivariate analysis of variance (PERMANOVA), which tests for significant compositional differences between groups. The homogeneity of dispersions among the samples, an underlying assumption of PERMANOVA, was confirmed by betadisper. Host species-specific clustering at each site was also confirmed by PERMANOVA (Table 2). A significant betadisper value from SIO (Table 2), suggesting heterogeneous dispersion between groups, was further investigated using a Tukey HSD test, which calculates *p* values

Table 1. OTU count, Shannon index and Pielou's evenness measurements (mean ± sd of all individuals samples) for all sites and host species.

	Total OTU count	Average OTU count	Shannon index (H′)	Evenness (J)	*n*
Lottia gigantea SIO	3542	696.9 ± 585.2	4.99 ± 2.50	0.51 ± 0.21	10
Lottia gigantea JB	2846	555.9 ± 363.0	5.16 ± 1.46	0.59 ± 0.12	9
Littorina keenae SIO	3193	805.7 ± 592.3	5.52 ± 1.57	0.59 ± 1.12	6
Littorina keenae JB	2525	721.3 ± 99.9	6.68 ± 0.43	0.71 ± 0.04	6
Chlorostoma eiseni SIO	3475	948.7 ± 350.0	6.48 ± 1.23	0.66 ± 0.10	7
Chlorostoma funebralis SIO	925	563.5 ± 19.1	5.23 ± 0.48	0.63 ± 0.12	2
Chlorostoma funebralis JB	3623	1093.3 ± 422.3	7.07 ± 1.38	0.71 ± 0.09	6
Turf algae SIO	3526	2330.5 ± 27.6	9.71 ± 0.43	0.87 ± 0.03	2
Turf algae JB	1186	697.5 ± 365.6	5.16 ± 1.60	0.56 ± 0.12	2
Water SIO	562	562	1.94	0.83	1
Water JB	1892	1892	8.94	0.83	1

Values are based on data rarefied to 10 200 sequences per sample and calculated using the estimate richness function in the phyloseq package for R.
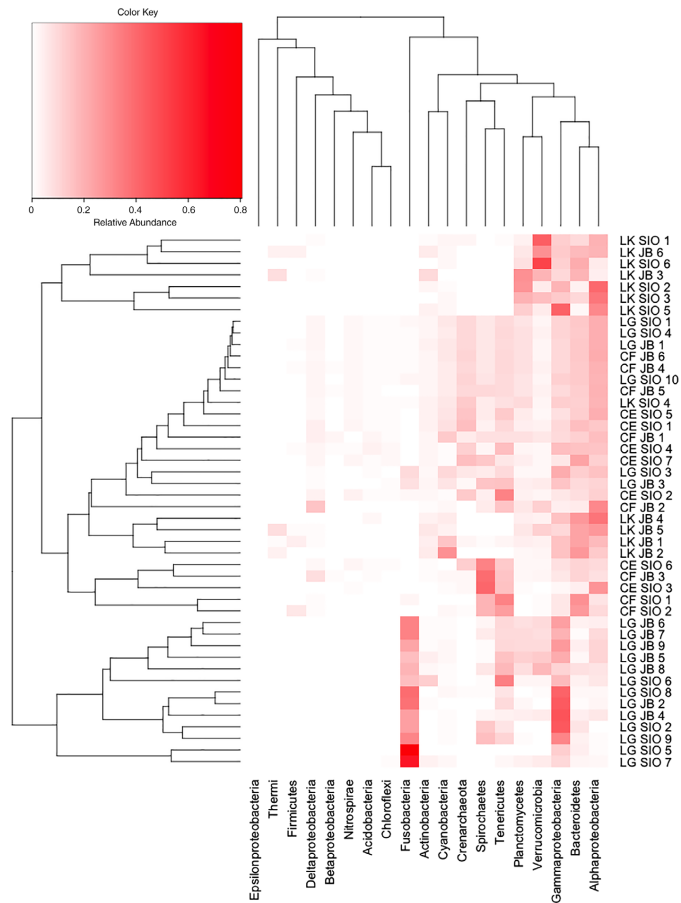
Fig. 1. Heatmap displaying relative abundances of Bacterial and Archaeal phyla in each host species and sampling location. The phylum Proteobacteria is divided into classes for greater resolution. Dendrograms on the top and left sides of the figure were computed using UPGMA clustering of a Bray–Curtis dissimilarity matrix. The dendrogram on the left side shows similarity of individuals of each host species at each site based on their microbiome compositions, while the top shows clusters of major microbial groups based on their compositions across our samples.

for all pairwise comparisons. The latter test indicated that comparisons with *C. funebralis* produced the heterogeneity of dispersions at SIO (all comparisons with *C. funebralis*: $p < 0.05$), likely due to the small sample size of this host. PERMANOVA estimates may be overly conservative if the group with the smaller sample size has lower dispersion values, suggesting that our PERMANOVA results are likely to be robust; nonetheless, they should be interpreted cautiously because of the sample size difference for this particular host (Anderson

and Walsh, 2013). The NMDS results also identify a set of four individuals of *L. gigantea* from both JB and SIO whose microbiomes do not cluster with their conspecifics. This appears to be due to a decrease in the relative abundance of Fusobacteria and subsequent increase in Alphaproteobacteria and Crenarchaeota in these individuals (Fig. 1).

Individual host species differed in their microbial community compositions, with some harbouring more than 1000 unique OTUs at each site (Fig. 3 and Tables 3 and 4).
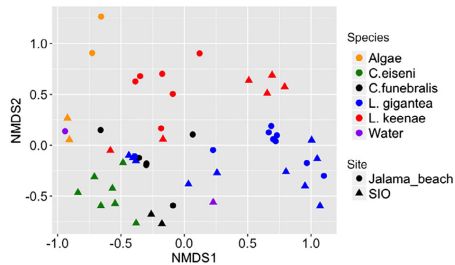
Fig. 2. Plot of the first two axes of NMDS of Bray–Curtis dissimilarity of all samples. Colours differentiate individual host species and shapes represent the two different collection sites. Two-dimensional stress of this plot is 0.12.

Likewise, algal and water samples did not share a majority of OTUs with the gastropod hosts sampled. Those OTUs that were shared, though, were among the most abundant taxa in the community, with unique OTUs being quite rare (Tables 3 and 4). In *L. gigantea* from JB, for example, the 1243 unique OTUs represented only 4.86% of the total sequence count for that population (Table 4). The reason for this difference is not clear and it is possible that some proportion of the low abundance OTUs may be the result of spurious OTUs generated by the OTU picking method (Edgar, 2017).

Microbiomes of individual host species tended to show site-specific differentiation by Bray–Curtis dissimilarity (Fig. 4A–C), which is significant in the case of *L. keenae* and *C. funebralis*, but not for *L. gigantea*, which was also impacted by heterogeneous dispersion (Table 5). NMDS plots of Jaccard distance, which is based on binary presence–absence of OTUs, also showed site-specific differentiation (Fig. 4D–F). This differentiation was significant in *L. keenae* and *C. funebralis*, but not *L. gigantea* (Table 6). Since the Jaccard metric does not take abundance information into account, the results of the Jaccard analysis further confirmed that compositional differences were high between sites and were a significant factor in differentiating these communities. However, results of analyses including *C. funebralis* should be interpreted with caution due to the small sample size of these hosts from SIO. Interestingly, in both the sets of analyses, the

Table 2. PERMANOVA and betadisper results of Bray–Curtis dissimilarity between host species at SIO, JB and across the whole dataset.

Site	Pseudo-F	R^2	P value	betadisper
SIO	2.84	0.393	<0.001*	0.001*
JB	4.23	0.471	<0.001*	0.114
Combined	5.86	0.297	<0.001*	0.092

* indicates significant p value (p < 0.05).

microbiomes of three individuals of *L. gigantea* from SIO and one individual from JB formed a cluster distinct from all the other individuals of this species, suggesting compositional differences between these individuals and their conspecifics (Fig. 4A and D).

Discussion

Host-species specificity in the marine gastropod microbiome

Our results show that although many of the same higher taxonomic groups (phylum and class) of microbes are present in the four different gastropod species analysed here, there is a marked differentiation in their abundances across these co-occurring hosts. Furthermore, compositions of microbial communities associated with each host species are significantly different at lower taxonomic levels (order and genus). Proteobacteria, for example, is the most abundant phylum present in each gastropod species, as well as in the environmental controls. This is consistent with previous studies that have shown members of this phylum to dominate host-associated microbiota in marine animals including limpets, corals, copepods and fish (Bayer *et al*., 2013; Dudek *et al*., 2014; Givens *et al*., 2015; Dorosz *et al*., 2016). However, preferential enrichment of different genera, such as *Mycoplasma* in *C. funebralis* and *Psychrilyobacter* in *L. gigantea,* contributes to significant differences among microbiomes of our four gastropod hosts. The four individuals of *L. gigantea* that cluster together and separately from their conspecifics (Figs. 1 and 4 and Supporting Information Fig. S2) exhibit low relative abundance of Fusobacteria, and instead show an increase in Alphaproteobacteria and Crenarchaeota. At present, the cause of this shift is unclear. Grazing on different algal species is unlikely to be an explanation, since our results show that the microbial communities associated with the algal and water samples differ significantly (Figs. 2 and 3) from those found in the gastropods. While such differentiation is expected – hosts ranging from primates (Yildirim *et al*., 2010) to deep sea sponges (Kennedy *et al*., 2014) have been shown to harbour unique microbial associations that are very different from those in the surrounding environment – further sampling of *L. gigantea* populations and their surrounding environment is needed to better understand the cause and consequence of the switch from Fusobacteria to other microbial taxa in certain individuals of this species.

Clustering of the microbiomes by host species seen here is also consistent with previous studies of vertebrates and invertebrates alike (Yildirim *et al*., 2010; Reveillaud *et al*., 2014). Among the four host species sampled here, the microbiomes of the two closest
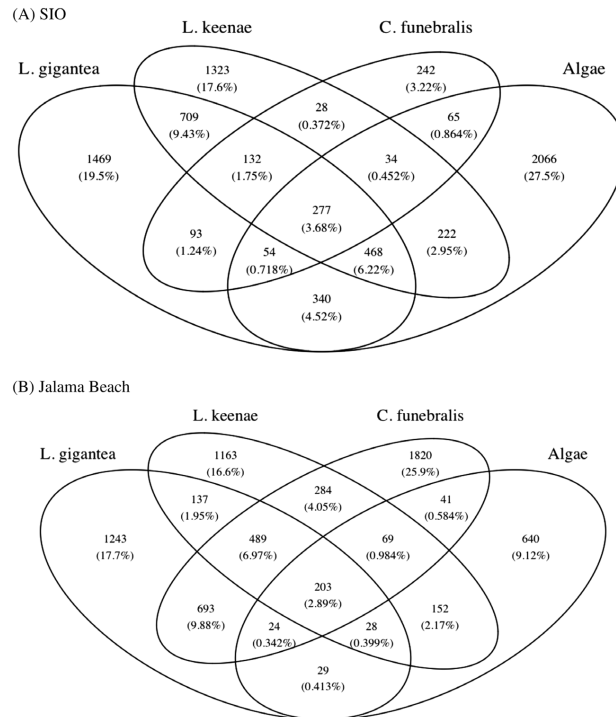
Fig. 3. Venn diagrams showing shared and unique OTUs in each host species at SIO (A) and JB (B). Calculations are based on OTU table rarefied to 10,200 sequences/sample.

Table 3. Number of unique OTUs and sequences present in each gastropod species and environmental control at SIO.

	L. gigantea	*L. keenae*	*C. funebralis*	Algae
Unique OTUs	1469	1323	242	2066
Unique sequence counts	3750	3567	1206	5718
Total Sequence counts	102 000	61 200	20 400	20 400
Percent unique	3.68%	5.83%	5.91%	28.03%

All samples rarefied to 10 200 sequences.

relatives, *C. eiseni* and *C. funebralis* (Hellberg 1998), are more similar to each other than they are to either *L. gigantea* or *L. keenae*, both of whom are distant relatives (Supporting Information Fig. S3). A previous study found evidence for phylosymbiosis, or more similar microbiome composition in more closely related species, in sponges (Easson and Thacker, 2014) and our results for *Chlorostoma* suggest that the pattern of phylosymbiosis may also exist in molluscs. However, future studies with broader sampling of microbiomes of molluscan clades are needed to properly test whether phylosymbiosis is common in this group and whether this pattern is driven by habitat filtering due to host traits or by codivergence of gastropod species and associated microbiota (Mazel *et al*., 2018).

Site specificity in the marine gastropod microbiome

Interestingly, diversity and evenness of microbial communities of a particular host did not significantly differ across the major biogeographic boundary at PtC. This differs from previous studies of host-associated marine microbial

13

Table 4. Number of unique OTUs and sequences present in each gastropod species and environmental control at JB.

	L. gigantea	*L. keenae*	*C. funebralis*	Algae
Unique OTUs	1243	1163	1820	640
Unique sequence counts	4465	4946	4337	1715
Total sequence counts	91 800	61 200	61 200	20 400
Percent unique	4.86%	8.08%	7.09%	8.41%

All samples rarefied to 10 200 sequences.

communities, where alpha diversity significantly differed between collection sites (Luter *et al*., 2015; Pantos *et al*., 2015). Similarly, pelagic microbial diversity has been shown to change across major oceanographic boundaries (Raes *et al*., 2018). Thus, our results suggest that marine biogeographic boundaries may affect the diversity of intertidal gastropod-associated microbial communities differently from that of other groups.

The compositions of the microbiomes of our focal hosts, as measured by Bray–Curtis and Jaccard indices, do vary significantly across PtC, which is consistent with results from other marine invertebrates including corals (Lema *et al*., 2014; Pantos *et al*., 2015; Kellogg *et al*., 2017) and some sponges (Burgsdorf *et al*., 2014) analysed across various dispersal barriers or large geographic distances. In our case, such differentiation is largely driven by composition of rare taxa and changes in the relative abundance of dominant taxa at each site, as has been seen previously in oyster hemolymph microbiomes (Lokmer *et al*., 2016). This is exemplified by PERMANOVA results of the binary Jaccard distances (Table 6) and lack of change in hierarchical clustering between the whole dataset and the 100% core microbiome in *L. gigantea* and *C. funebralis* (Supporting Information Fig. S2). *L. keenae* does appear to exhibit some change in clustering at the 100% core level, suggesting there may be an increased influence of less ubiquitous taxa in separating the two sites. It is likely, then, that a core microbiome exists for each host and that members of this core are distributed across the PtC biogeographic boundary. Spatial distributions of rare taxa, on the other hand, may be constrained by this oceanographic boundary, as suggested by the Jaccard analyses, but further investigation is necessary to test this hypothesis.

Major contributors to the marine gastropod microbiome

Driving part of the difference between species, particularly *L. gigantea*, at the OTU level is members of the phylum Fusobacteria (Fig. 1). Fusobacteria are obligate anaerobes that were first discovered in the 1920s as part of the oral microbiome (Knorr, 1922), but later came to be known as a group with numerous human pathogens (Altshuler and Hyde, 1988; Bennett and Eley, 1993; Zhong *et al*., 2014). However, with the advent of next-generation sequencing and large-scale environmental sampling, they have also been found in the normal gut and oral flora of humans and a number of other animal hosts (Zaura *et al*., 2009; Lyons *et al*., 2017; Schill *et al*., 2017; Smith *et al*., 2017; Cicala *et al*., 2018). While members of this phylum are generally found in low relative abundances in their hosts, Fusobacteria have been shown to comprise greater than 20% of the relative abundance in the tunicate *Ciona intestinalis* (Dishaw *et al*., 2014) and greater than 30% relative abundance in the abalone *Haliotis discus hannai* (Nam *et al*., 2018). They were also found to be the dominant bacterial phylum in the abalone *Haliotis tuberculata* (Gobet *et al*., 2018).

In the gastropods sampled here, the most abundant Fusobacteria OTU was a member of the genus *Psychrilyobacter*. This genus is poorly studied; however, members have been found in the microbiomes of other marine hosts (Fernandez-Piquer *et al*., 2012; Nelson *et al*., 2013). It has been suggested that due to the anaerobic nature of *Psychrilyobacter* sp., these species may form stable biofilms in the mucus layers of the gut, as it is unlikely that they would persist in the plankton (Dishaw *et al*., 2014). *Psychrilyobacter atlanticus* has been shown to utilize a diverse array of carbon sources (Zhao *et al*., 2009), so it is possible that the *Psychrilyobacter* strains found here are aiding in digestion and utilization of carbon from microalgal food. A search of the only currently available *Psychrilyobacter* genome for evidence of carbohydrate-active enzymes concluded that while that strain is likely able to degrade certain oligosaccharides, it does not seem to have the ability to degrade complex algal polysaccharides (Gobet *et al*., 2018). However, our knowledge of the diet or gut physiology of the gastropod hosts sampled here, or the enzymatic activity of the microbiota strains sampled, is too poor to determine if there are any dietary features unique to *L. gigantea* driving this relationship. Furthermore, if *Psychrilyobacter* exists only as biofilms in the anaerobic mucus layers of the gut, it raises the question of how these bacteria are transmitted among different individuals of a broadcast spawning marine invertebrate species.

The phylum Tenericutes was found to be common in *C. funebralis*, with many abundant OTUs belonging to the genus *Mycoplasma*, which is comprised of a large number of human pathogens (Kirkpatrick, 1992). Recently, however, large relative abundances of *Mycoplasma* have
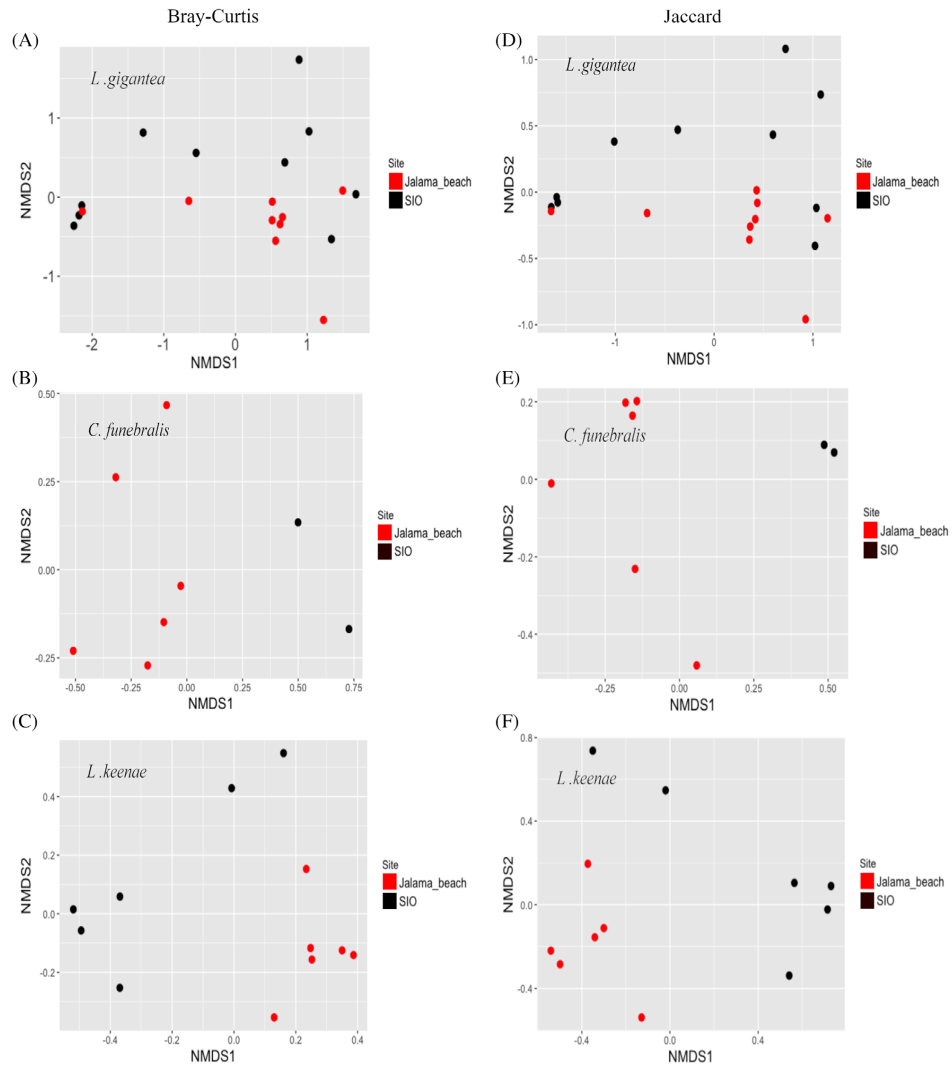
Fig. 4. Plots of the first two axes of NMDS analyses of *L. gigantea* (A, D), *C. funebralis* (B, E) and *L. keenae* (C, F) microbial communities using Bray–Curtis Dissimilarity (A–C) and Jaccard binary distances (D–F). Two-dimensional stresses are 0.09 (A), 0.03 (B), 0.05 (C), 0.08 (D), 0.001 (E) and 0.06 (F) respectively.

been found in the tissues of other marine hosts, including fish (Atlantic salmon: Llewellyn *et al*., 2016 and long-jawed mudsucker: Bano *et al*., 2007), marine molluscs (abalone: Cicala *et al*., 2018 and oysters: Arfken *et al*., 2017) and octocorals (Holm and Heidelberg, 2016). *Mycoplasma*

lack cell walls and require molecules such as sterols and fatty acids provided by the host to grow and persist (Ludwig *et al*., 2010). *Mycoplasma* have also been suggested to be highly host-specific and play important ecological roles in the guts of marine gastropods (Cicala

15

A. T. Neu, E. E. Allen and K. Roy

Table 5. PERMANOVA and betadisper results of Bray–Curtis dissimilarity between populations of *L. keenae, C. funebralis* and *L. gigantea*.

Host species	Pseudo-F	R^2	p value	betadisper
L. keenae	4.00	0.286	<0.007*	0.081
C. funebralis	2.34	0.280	0.023 < p < 0.05*	0.200
L. gigantea	1.78	0.095	0.078 < p < 0.149	0.023*

* indicates significant p value (p < 0.05).

Table 6. PERMANOVA and betadisper results of Jaccard distance between populations of *L. keenae, C. funebralis* and *L. gigantea*.

Host species	Pseudo-F	R^2	p value	betadisper
L. keenae	1.94	0.163	<0.007*	0.081
C. funebralis	2.34	0.280	0.023 < p < 0.05*	0.200
L. gigantea	1.78	0.096	0.074 < p < 0.151	0.020*

* indicates significant p value (p < 0.05).

et al., 2018). This is in line with the findings of Fraune and Zimmer (2008), who showed that hosts that survived on cellulose-based diets had a higher proportion of *Mycoplasma*-like symbionts. Differential abundance and composition of *Mycoplasma* OTUs by host species may indicate commensal relationships between particular strains of *Mycoplasma* and their marine hosts.

Individuals of *L. keenae*, particularly those sampled at SIO, showed enrichment in members of the phylum Verrucomicrobia, which have been found in a number of environmental microbiomes and are nearly ubiquitous in soils (Bergmann *et al*., 2011). *L. keenae* resides high in the splash zone and is regularly exposed to terrestrial inputs. Further, members of the genus *Littorina* have been previously shown to ingest pieces of the substrate on which they reside (Voltolina and Sacchi, 1990), which may increase the number and abundance of soil- and rock-associated microbial taxa in this marine species.

Conclusions

Our results, for the first time, provide information about how microbiomes of intertidal gastropod hosts change across a major biogeographic boundary. They also clearly show that host identity plays a critical role in shaping the marine gastropod microbiome, affecting both microbial diversity and community composition. Even among members of the same bacterial genus, host species show differentiation in the presence and abundance of specific OTUs. Whether this is due to each host differentially ingesting microbes from the local species pool through dietary inputs or physiological differences in the gut environment (e.g. pH and temperature) that nurture different microbial communities is unclear. Furthermore, our results show that within certain hosts, there are some individuals with microbiomes very different from those of their conspecifics. The causes of such outliers remain

unclear at present; further sampling to determine whether these differences are transient or persist over time would be helpful.

Our two sites, separated by a major environmental gradient, play a lesser, although significant, role in shaping the microbiome compositions of individual hosts. Overall diversity did not change significantly between sites, but community composition did. This suggests that there is a significant differentiation in microbial species across the biogeographic boundary at PtC, a pattern long documented for marine metazoans (Valentine 1966). Hosts likely sample functionally similar microbes from their local species pool, which would account for OTU level differentiation, even if microbial profiles look similar at the phylum or class level.

Microbiomes of marine molluscs, despite their ecological and commercial importance, remain understudied relative to other marine invertebrates such as corals and sponges. More information about the diversity and composition of microbiomes of molluscan species and populations, in conjunction with functional and metabolomic analyses, is needed to better understand the interactions between microbes and their molluscan hosts and how such interactions may affect host responses to global change.

Experimental procedures

Sample collection

We sampled multiple individuals of each of our focal species (see Table 1 for counts) at two different rocky intertidal sites, except for *Chlorostoma eiseni*, which was only sampled at the southern site. Our southern site, SIO (32.8661 N, 117.2543 W) is in the Californian biogeographic province while JB (34.5131 N, 120.5041 W) is in the Oregonian province, north of PtC. In order to avoid potential seasonal changes in microbial communities, SIO was sampled on February 23 and JB on February

26, 2017. Immediately after collection, specimens were placed in sterile bags and preserved on ice until they could be returned to the laboratory. Once in the laboratory, individuals were euthanized by freezing and all samples were processed within 24 h of collection. Due to the small body size of *L. keenae*, we removed the shells of each individual, rinsed the tissue in sterile seawater and 90% ethanol and placed the whole animal in RNAlater (Invitrogen, Carlsbad, CA, USA) for storage at −20°C prior to DNA extraction. For all other species, we removed the shells and excised approximately half of the hepatopancreatic tissue, which was rinsed in sterile seawater and 90% ethanol and stored in RNAlater at −20 °C prior to DNA extraction. The remaining tissue ('whole body' samples) was also rinsed with sterile seawater and 90% ethanol and stored in RNAlater at −20 °C. We also sampled hindgut tissue from individuals of *L. gigantea* and preserved them using the same methods. At each site, we also collected one seawater sample and three samples of the dominant algal turf to characterize the microbial communities of the host habitat. Seawater was filtered using a Sterivex 0.22 μm filter (MilliporeSigma, Burlington, MA, USA) within 12 h of collection. Algae samples were stored in RNAlater at −20 °C.

Sample processing

DNA was extracted using a Qiagen PowerLyzer PowerSoil kit (Qiagen GmbH, Hilden, Germany) according to the manufacturer's instructions, but we added a 30 min incubation step at 55 °C after initial bead beating to aid with complete lysis. Bacterial and Archaeal V4 regions of the 16S rRNA gene were amplified using primers 515f (GTGYCAG CMGCCGCGGTAA) and 806rb (ATTAGAWACCCBNG TAGTCC) using the following protocol: 98 °C 1 min, 30 cycles of 98 °C 30s, 54 °C 20s, 72 °C 10s, finished with 72 °C for 5 min. Final amplification volumes were 25 μl and contained 12.5 μl Q5 High-Fidelity 2x Master Mix (New England Biolabs, Ipswich, MA, USA), 7.5 μl ddH2O, 0.5 μl of each primer (10 μM) and 4 μl of template DNA. Reactions were run in triplicate and pooled after visualization on a 1% agarose gel using SybrSafe DNA stain (Invitrogen, Carlsbad, CA, USA). Unique NexteraXT barcode sequence pair (Illumina, Inc., San Diego, CA, USA) were added using the protocol: 98 °C for 30s, 8 cycles of 98 °C for 10s, 60 °C for 20s and 72 °C for 30s, with a final extension at 72 °C for 2 min. Final PCR products were purified with 1 U of Exonuclease I and 1 U Shrimp Alkaline Phosphatase (Exo-SAP; New England Biolabs, Ipswich, MA, USA) heated to 37 °C for 30 min and 85 °C for 15 min followed by Ampure XP bead purification (Beckman Coulter, Inc., Carlsbad, CA) following the manufacturer's instructions. Purified amplicons were submitted for Illumina MiSeq sequencing using v3 chemistry (600 cycle; 2 × 300 bp paired-end output) at the Institute for Genomic Medicine, University of California San Diego. All sequence files were submitted to NCBI Sequence Read Archive and are available under BioProject accession number PRJNA497636.

Sequence processing

Paired-end sequences were merged using PEAR v0.9.10 (Zhang *et al.,* 2014) for further processing in QIIME v1.9.1 (Caporaso *et al*., 2010). Libraries were split and filtered, removing any samples containing a qscore below 25. All primer sequences were removed using CutAdapt v1.12 (Martin, 2011). We used Vsearch v2.3.4 (Rognes *et al*., 2016) to identify chimeric sequences and discarded them from subsequent analyses. All sequences were filtered to a specific length (min. 200 bp, max. 500 bp) and OTUs were picked using the pick_open_reference_otus.py pipeline in QIIME using the Silva 111 database as a reference and clustering OTUs at 98% sequence similarity (Quast *et al*., 2013). Taxonomy was assigned using the UCLUST (Edgar, 2010) classifier with an 80% cutoff value and the SILVA 123 database (Quast *et al*., 2013). Before further analyses, we removed all singleton and doubleton OTUs, samples containing fewer than 1000 sequences (six in total) as well as OTUs matched to mitochondria or plastid sequences.

Whole body samples were compared with samples of the digestive gland in order to determine any differences in community composition or diversity. All of these tests, including Kruskal–Wallis Test of Shannon–Weaver index for diversity and PERMANOVA for composition, were not significant ($p > 0.05$). In order to avoid pseudoreplication, only whole body samples were used to perform subsequent analyses.

Diversity and community composition

All analyses were done using the QIIME v1.9.1 pipeline (Caporaso *et al*., 2010) in conjunction with R v3.4.0 (R Core Team, 2013) packages vegan v2.4-3 (Oksanen *et al*., 2018) and phyloseq v1.20.0 (McMurdie and Holmes, 2013).

We rarefied all samples to 10 200 sequences, the depth of the gastropod sample with the fewest sequences. Seawater from SIO only contained 3600 sequences but is included here to provide environmental information about this site. We used the Shannon index and Pielou's evenness (Pielou, 1966) as our primary alpha diversity metrics, with Wilcoxon Rank-Sum and Kruskal–Wallis tests to investigate variation between sites and species respectively. Bray–Curtis dissimilarity and Jaccard indices were used to explore beta diversity patterns in order to determine whether or not abundance of particular OTUs contributes to compositional differences. We used NMDS in

the phyloseq package of R (McMurdie and Holmes, 2013) to visualize differences in community compositions among species and/or sites and PERMANOVA (Anderson, 2001) to test for differences in community compositions. PERMANOVAs were run on each of 999 rarefied OTU tables and the range of *p* values is reported. We checked all significant PERMANOVA results for heterogeneity of dispersion (using the betadisper function of the vegan package (Oksanen *et al*., 2018) in R with 999 permutations) to determine whether significance was a function of dispersion or a true effect. Venn diagrams were also generated for each site using R package VennDiagram v1.6.20 (Chen and Boutros, 2011). Sequences of OTUs unique to each species at each site in a representative rarefied dataset were counted and used to determine the relative abundances of unique and shared OTUs.

To further explore patterns in the data, we utilized the idea of a 'core' microbiome (Tschöp *et al*., 2009) or set of microbial taxa found across a majority (in this case 100%) of individuals of a host species. Both core and total microbiomes were visualized using unweighted pair group method with arithmetic mean (UPGMA) clustering in R with packages phyloseq (McMurdie and Holmes, 2013) and ape v5.1 (Paradis and Schliep, 2018) to show relationships between samples in an easily comparable way.

To test for phylogenetic conservatism in the microbiome, we plotted pairwise Bray–Curtis dissimilarity values for all species pairs using the make_distance_comparison_plots. py function in QIIME v1.9.1 (Caporaso *et al*., 2010).

References

Altshuler, G., and Hyde, S. (1988) Clinicopathologic considerations of Fusobacteria chorioamnionitis. *Acta Obstet Gynecol Scand* **67**: 513–517.

Anderson, M.J. (2001) A new method for non-parametric multivariate analysis of variance. *Austral Ecol* **26**: 32–46.

Anderson, M.J., and Walsh, D.C.I. (2013) PERMANOVA, ANOSIM, and the mantel test face heterogeneous dispersions: what null hypothesis are you testing? *Ecol Monograph* **83**: 557–574.

Apprill, A. (2017) Marine animal microbiomes: toward understanding host – microbiome interactions in a changing ocean. *Front Mar Sci* **4**: 222.

Arfken, A., Song, B., Bowman, J.S., and Piehler, M. (2017) Denitrification potential of the eastern oyster microbiome using a 16S rRNA gene based metabolic inference approach. *PLoS One* **12**: e0185071.

Aronson, H.S., Zellmer, A.J., and Goffredi, S.K. (2017) The specific and exclusive microbiome of the deep-sea bone-eating snail, Rubyspira osteovora. *FEMS Microbiol Ecol* **93**: fiw250.

Bano, N., DeRae Smith, A., Bennett, W., Vasquez, L., and Hollibaugh, J.T. (2007) Dominance of mycoplasma in the guts of the long-jawed Mudsucker, *Gillichthys mirabilis*, from five California salt marshes. *Environ Microbiol* **9**: 2636–2641.

Batstone, R.T., and Dufour, S.C. (2016) Closely related thyasirid bivalves associate with multiple symbiont phylotypes. *Mar Ecol* **37**: 988–997.

Bayer, T., Neave, M.J., Alsheikh-Hussain, A., Aranda, M., Yum, L.K., Mincer, T., *et al*. (2013) The microbiome of the red sea coral *Stylophora pistillata* is dominated by tissue-associated Endozoicomonas bacteria. *Appl Environ Microbiol* **79**: 4759–4762.

Bednar, C., and Trulio, L. (2017) Reducing human impacts on central Californian intertidal gastropods: are marine protected areas effective? *Molluscan Res* **37**: 148–152.

Bennett, K.W., and Eley, A. (1993) Fusobacteria: new taxonomy and related diseases. *J Med Microbiol* **39**: 246–254.

Bergmann, G.T., Bates, S.T., Eilers, K.G., Lauber, C.L., Caporaso, J.G., Walters, W.A., *et al*. (2011) The under-recognized dominance of Verrucomicrobia in soil bacterial communities. *Soil Biol Biochem* **43**: 1450–1455.

Betcher, M.A., Fung, J.M., Han, A.W., O'Connor, R., Seronay, R., Concepcion, G.P., *et al*. (2012) Microbial distribution and abundance in the digestive system of five shipworm species (Bivalvia: Teredinidae). *PLoS One* **7**: e45309.

Burgsdorf, I., Erwin, P.M., Lopez-Legentil, S., Cerrano, C., Haber, M., Frenk, S., and Steindler, L. (2014) Biogeography rather than association with cyanobacteria structures symbiotic microbial communities in the marine sponge *Petrosia ficiformis*. *Front Microbiol* **5**: 1–11.

Caldwell, P.C., Stuart, D.W., and Brink, K.H. (1986) Mesoscale wind variability near point conception, California during spring 1983. *J Climatol Appl Meteorol* **25**: 1241–1254.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., *et al*. (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**: 335–336.

Chen, H., and Boutros, P.C. (2011) VennDiagram: a package for the generation of highly customizable Venn and Euler diagrams in R. *BMC Bioinformatics* **12**: 35.

Cicala, F., Cisterna-Céliz, J.A., Moore, J.D., and Rocha-Olivares, A. (2018) Structure, dynamics and predicted functional role of the gut microbiota of the blue (*Haliotis fulgens*) and yellow (*H. corrugata*) abalone from Baja California Sur, Mexico. *PeerJ* **6**: e5830.

Connor, V.M. (1986) The use of mucous trails by intertidal limpets to enhance food resources. *Biol Bull* **171**: 548–564.

Dahl, A.L. (1964) Macroscopic algal foods of *Littorina planaxis* Philippi and *Littorina scutulata* Gould (Gastropoda: Prosobranchiata). *Veliger* **7**: 139–143.

Davis, J., Fricke, W.F., Hamann, M.T., Esquenazi, E., Dorrestein, P.C., and Hill, R.T. (2013) Characterization of the bacterial community of the chemically defended Hawaiian sacoglossan Elysia rufescens. *Appl Environ Microbiol* **79**: 7073–7081.

Dishaw, L.J., Flores-Torres, J., Lax, S., Gemayel, K., Leigh, B., Melillo, D., *et al*. (2014) The gut of geographically disparate *Ciona intestinalis* harbors a core microbiota. *PLoS One* **9**: e93386.

Dorosz, J.N.A., Castro-Mejia, J.L., Hansen, L.H., Nielsen, D. S., and Skovgaard, A. (2016) Different microbiomes associated with the copepods Acartia tonsa and *Temora longicornis* from the same marine environment. *Aquat Microb Ecol* **78**: 1–9.

Dudek, M., Adams, J., Swain, M., Hegarty, M., Huws, S., and Gallagher, J. (2014) Metaphylogenomic and potential functionality of the limpet Patella pellucida's gastrointestinal tract microbiome. *Int J Mol Sci* **15**: 18819–18839.

Easson, C.G., and Thacker, R.W. (2014) Phylogenetic signal in the community structure of host-specific microbiomes of tropical marine sponges. *Front Microbiol* **5**: 532.

Edgar, R.C. (2017) Accuracy of microbial community diversity estimated by closed- and open-reference OTUs. *PeerJ* **5**: e3889.

Edgar, R.C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**: 2460–2461.

Erlandson, J.M., Ainis, A.F., Braje, T.J., Jew, N.P., Rick, T. C., Vellanoweth, R.L., *et al* (2015) 12,000 years of human predation on black turban snails (Chlorostoma funebralis) on Alta California's northern Channel Islands. *California Archaeology* **7**: 59–91.

Fan, L., Reynolds, D., Liu, M., Stark, M., Kjelleberg, S., Webster, N.S., and Thomas, T. (2012) Functional equivalence and evolutionary convergence in complex communities of microbial sponge symbionts. *Proc Natl Acad Sci* **109**: E1878–E1887.

Fenberg, P.B., and Roy, K. (2012) Anthropogenic harvesting pressure and changes in life history: insights from a rocky intertidal limpet. *Am Nat* **180**: 200–210.

Fernandez-Piquer, J., Bowman, J.P., Ross, T., and Tamplin, M.L. (2012) Molecular analysis of the bacterial communities in the live Pacific oyster (*Crassostrea gigas*) and the influence of postharvest temperature on its structure. *J Appl Microbiol* **112**: 1134–1143.

Fraune, S., and Zimmer, M. (2008) Host-specificity of environmentally transmitted mycoplasma-like isopod symbionts. *Environ Microbiol* **10**: 2497–2504.

Givens, C.E., Ransom, B., Bano, N., and Hollibaugh, J.T. (2015) Comparison of the gut microbiomes of 12 bony fish and 3 shark species. *Mar Ecol Prog Ser* **518**: 209–223.

Gobet, A., Mest, L., Perennou, M., Dittami, S.M., Caralp, C., Coulombet, C., *et al*. (2018) Seasonal and algal diet-driven patterns of the digestive microbiota of the European abalone *Haliotis tuberculata*, a generalist marine herbivore. *Microbiome* **6**: 60.

Green, T.J., and Barnes, A.C. (2010) Bacterial diversity of the digestive gland of Sydney rock oysters, Saccostrea glomerata infected with the paramyxean parasite, Marteilia sydneyi. *J Appl Microbiol* **109**: 613–622.

Harms, S., and Winant, C.D. (1998) Characteristic patterns of the circulation in the Santa Barbara Channel. *J Geophys Res* **103**: 3041–3065.

Hellberg, M.E. (1998) Sympatric sea shells along the sea's shore: the geography of speciation in the marine gastropod Tegula. *Evolution* **52**: 1311–1324.

Hentschel, U., Piel, J., Degnan, S.M., and Taylor, M.W. (2012) Genomic insights into the marine sponge microbiome. *Nat Rev Microbiol* **10**: 641–654.

Holm, J.B., and Heidelberg, K.B. (2016) Microbiomes of Muricea californica and *M. fruticosa*: comparative analyses of two co-occurring eastern Pacific octocorals. *Front Microbiol* **7**: 917.

Kellogg, C.A., Goldsmith, D.B., and Gray, M.A. (2017) Biogeographic comparison of Lophelia-associated bacterial communities in the western Atlantic reveals conserved core microbiome. *Front Microbiol* **8**: 796.

Kennedy, J., Flemer, B., Jackson, S.A., Morrissey, J.P., O'Gara, F., and Dobson, A.D.W. (2014) Evidence of a putative deep sea specific microbiome in marine sponges. *PLoS One* **9**: e91092.

King, G.M., Judd, C., Kuske, C.R., and Smith, C. (2012) Analysis of stomach and gut microbiomes of the eastern oyster (*Crassostrea virginica*) from coastal Louisiana, USA. *PLoS One* **7**: e51475.

Kirkpatrick, B.C. (1992) Mycoplasma-like organisms—plant and invertebrate pathogens. In *The Prokaryotes*. New York, NY: Springer, pp. 4050–4067.

Knorr, M. (1922) Über die fusispirillare Symbiose, die Gattung Fusobacterium (KB Lehman) und Spirillum sputigenum. *Zbl Bakt I Abt Orig* **87**: 536–545.

Lema, K.A., Willis, B.L., and Bourne, D.G. (2014) Amplicon pyrosequencing reveals spatial and temporal consistency in diazotroph assemblages of the *Acropora millepora* microbiome. *Environ Microbiol* **16**: 3345–3359.

Lesser, M.P., Fiore, C., Slattery, M., and Zaneveld, J. (2016) Climate change stressors destabilize the microbiome of the Caribbean barrel sponge, *Xestospongia muta*. *J Exp Mar Bio Ecol* **475**: 11–18.

Littman, R.A., Willis, B.L., and Bourne, D.G. (2009) Bacterial communities of juvenile corals infected with different Symbiodinium (dinoflagellate) clades. *Mar Ecol Prog Ser* **389**: 45–59.

Llewellyn, M.S., McGinnity, P., Dionne, M., Letourneau, J., Thonier, F., Carvalho, G.R., *et al* (2016) The biogeography of the Atlantic salmon (*Salmo salar*) gut microbiome. *ISME J* **10**: 1280–1284.

Lokmer, A., Goedknegt, M.A., Thieltges, D.W., Fiorentino, D., Kuenzel, S., Baines, J.F., and Mathias Wegner, K. (2016) Spatial and temporal dynamics of pacific oyster hemolymph microbiota across multiple scales. *Front Microbiol* **7**: 1367.

Ludwig, W., Euzéby, J., and Whitman, W.B. (2010) Road map of the phyla Bacteroidetes, Spirochaetes, Tenericutes (Mollicutes), Acidobacteria, Fibrobacteres, Fusobacteria, Dictyoglomi, Gemmatimonadetes, Lentisphaerae, Verrucomicrobia, Chlamydiae, and Planctomycetes. In *Bergey's Manual of Systematic Bacteriology*. Krieg, N.R., *et al.* (eds). New York, NY: Springer, pp. 1–19.

Luter, H.M., Widder, S., Botté, E.S., Abdul Wahab, M., Whalan, S., Moitinho-Silva, L., *et al*. (2015) Biogeographic variation in the microbiome of the ecologically important sponge, *Carteriospongia foliascens*. *PeerJ* **3**: e1435.

Lyons, P.P., Turnbull, J.F., Dawson, K.A., and Crumlish, M. (2017) Phylogenetic and functional characterization of the distal intestinal microbiome of rainbow trout *Oncorhynchus mykiss* from both farm and aquarium settings. *J Appl Microbiol* **122**: 347–363.

Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* **17**: 10.

Mazel, F., Davis, K.M., Loudon, A., and Kwong, W.K. (2018) Is host filtering the Main driver of Phylosymbiosis across the tree of life? *mSystems* **3**: e00097–e00018.

McMurdie, P.J., and Holmes, S. (2013) Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* **8**: e61217.

Nam, B.-H., Jang, J., Caetano-Anolles, K., Kim, Y.-O., Park, J.Y., Sohn, H., *et al*. (2018) Microbial community and functions associated with digestion of algal polysaccharides in the visceral tract of Haliotis discus hannai: insights from metagenome and metatranscriptome analysis. *PLoS One* **13**: e0205594.

Nelson, T.M., Rogers, T.L., Carlini, A.R., and Brown, M.V. (2013) Diet and phylogeny shape the gut microbiota of Antarctic seals: a comparison of wild and captive animals. *Environ Microbiol* **15**: 1132–1145.

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlin, D., et al. (2018) Vegan: Community Ecology Package. R package version 2.5-3. 2018.

Owen, R.W. (1980) Eddies of the California current system: physical and ecological characteristics. In *The California Islands: Proceedings of a Multidiciplinary Symposium*, Santa Barbara, CA: Santa Barbara Museum of Natural History, pp. 237–263.

Pantos, O., Bongaerts, P., Dennis, P.G., Tyson, G.W., and Hoegh-Guldberg, O. (2015) Habitat-specific environmental conditions primarily control the microbiomes of the coral *Seriatopora hystrix*. *ISME J* **9**: 1916–1927.

Paradis, E., and Schliep, K. (2018) ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**: bty633.

Pfister, C.A., Meyer, F., and Antonopoulos, D.A. (2010) Metagenomic profiling of a microbial assemblage associated with the California mussel: a node in networks of carbon and nitrogen cycling. *PLoS One* **5**: e10518.

Pielou, E.C. (1966) The measurement of diversity in different types of biological collections. *J Theor Biol* **13**: 131–144.

Pierce, M.L., Ward, J.E., Holohan, B.A., Zhao, X., and Hicks, R.E. (2016) The influence of site and season on the gut and pallial fluid microbial communities of the eastern oyster, *Crassostrea virginica* (Bivalvia, Ostreidae): community-level physiological profiling and genetic structure. *Hydrobiologia* **765**: 97–113.

Prieur, D., Mével, G., Nicolas, J.L., Plusquellec, A., and Vigneulle, M. (1990) Interactions between bivalve molluscs and bacteria in the marine environment. *Oceangr Mar Biol Annu Rev* **28**: 277–352.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., *et al*. (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41**: D590–D596.

Raes, E.J., Bodrossy, L., van de Kamp, J., Bissett, A., Ostrowski, M., Brown, M.V., *et al*. (2018) Oceanographic boundaries constrain microbial diversity gradients in the South Pacific Ocean. *Proc Natl Acad Sci U S A* **115**: E8266–E8275.

Reshef, L., Koren, O., Loya, Y., Zilber-Rosenberg, I., and Rosenberg, E. (2006) The coral probiotic hypothesis. *Environ Microbiol* **8**: 2068–2073.

Reveillaud, J., Maignien, L., Eren, M.A., Huber, J.A., Apprill, A., Sogin, M.L., and Vanreusel, A. (2014) Host-specificity among abundant and rare taxa in the sponge microbiome. *ISME J* **8**: 1198–1209.

Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahé, F. (2016) VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4**: e2584.

Roy, K., Collins, A.G., Becker, B.J., Begovic, E., and Engle, J.M. (2003) Anthropogenic impacts and historical decline in body size of rocky intertidal gastropods in southern California. *Ecol Lett* **6**: 205–211.

Roy, K., Jablonski, D., and Valentine, J.W. (1994) Eastern Pacific molluscan provinces and latitudinal diversity gradient: no evidence for "Rapoport's rule." *Proc Natl Acad Sci U S A* **91**: 8871–8874.

Schill, W.B., Iwanowicz, D., and Adams, C. (2017) Endozoicomonas dominates the gill and intestinal content microbiomes of *Mytilus edulis* from Barnegat Bay, New Jersey. *J Shellfish Res* **36**: 391–401.

Smith, P., Willemsen, D., Popkes, M., Metge, F., Gandiwa, E., Reichard, M., and Valenzano, D.R. (2017) Regulation of life span by the gut microbiota in the short-lived african turquoise killifish. *Elife* **6**: e27014.

Stimson, J. (1970) Territorial behavior of the owl limpet, *Lottia gigantea*. *Ecology* **51**: 113–118.

Team, R.C. (2013) R: a language and environment for statistical computing.

Trabal Fernández, N., Mazón-Suástegui, J.M., Vázquez-Juárez, R., Ascencio-Valle, F., and Romero, J. (2014) Changes in the composition and diversity of the bacterial microbiota associated with oysters (*Crassostrea corteziensis*, Crassostrea gigas and Crassostrea sikamea) during commercial production. *FEMS Microbiol Ecol* **88**: 69–83.

Trabal, N., Mazón-Suástegui, J.M., Vázquez-Juárez, R., Asencio-Valle, F., Morales-Bojórquez, E., and Romero, J. (2012) Molecular analysis of bacterial microbiota associated with oysters (Crassostrea gigas and *Crassostrea corteziensis*) in different growth phases at two cultivation sites. *Microb Ecol* **64**: 555–569.

Tschöp, M.H., Hugenholtz, P., and Karp, C.L. (2009) Getting to the core of the gut microbiome. *Nat Biotechnol* **27**: 344–346.

Valentine, J.W. (1966) Numerical analysis of marine molluscan ranges on the extratropical northeastern Pacific shelf. *Limnol Oceanogr* **11**: 198–211.

Vega Thurber, R., Willner-Hall, D., Rodriguez-Mueller, B., Desnues, C., Edwards, R.A., Angly, F., *et al*. (2009) Metagenomic analysis of stressed coral holobionts. *Environ Microbiol* **11**: 2148–2163.

Voltolina, D., and Sacchi, C.F. (1990) Field observations on the feeding habits of *Littorina scutulata* Gould and

L. sitkana Philippi (Gastropoda, Prosobranchia) of southern Vancouver Island (British Columbia, Canada). *Hydrobiologia* **193**: 147–154.

Webster, N.S., Negri, A.P., Flores, F., Humphrey, C., Soo, R., Botté, E.S., *et al*. (2013) Near-Future Ocean acidification causes differences in microbial associations within diverse coral reef taxa. *Environ Microbiol Rep* **5**: 243–251.

Yildirim, S., Yeoman, C.J., Sipos, M., Torralba, M., Wilson, B.A., Goldberg, T.L., *et al*. (2010) Characterization of the fecal microbiome from non-human wild primates reveals species specific microbial communities. *PLoS One* **5**: e13963.

Zaura, E., Keijser, B.J., Huse, S.M., and Crielaard, W. (2009) Defining the healthy "core microbiome" of oral microbial communities. *BMC Microbiol* **9**: 259.

Zbinden, M., Marqué, L., Gaudron, S.M., Ravaux, J., Léger, N., and Duperron, S. (2014) Epsilonproteobacteria as gill epibionts of the hydrothermal vent gastropod Cyathermia naticoides (North East-Pacific rise). *Mar Biol* **162**: 435–448.

Zhao, J.S., Manno, D., and Hawari, J. (2009) Psychrilyobacter atlanticus gen. nov., sp. nov., a marine member of the phylum Fusobacteria that produces H2 and degrades nitramine explosives under low temperature conditions. *Int J Syst Evol Microbiol* **59**: 491–497.

Zhong, D., Brower-Sinning, R., Firek, B., and Morowitz, M.J. (2014) Acute appendicitis in children is associated with an abundance of bacteria from the phylum Fusobacteria. *J Pediatr Surg* **49**: 441–446.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Fig. S1. Map showing the locations of the collection sites on each side of the biogeographic boundary at Point Conception, California.

Fig. S2. Unweighted Pair Group Method with Arithmetic Mean (UPGMA) clustering analysis of gastropod microbiomes using Bray–Curtis dissimilarity using the entire dataset (A) and excluding singletons, doubletons and tripletons, as well as any OTUs found in fewer than 100% of individuals of a host species (B).

Fig. S3. Boxplot of Bray–Curtis dissimilarity values between all members of each species/population paired across Point Conception. An asterisk (*) denotes a species/population pair that is significantly different ($p < 0.05$) from the *C. funebralis-C. eiseni* pair by Student's two sample *t*-test with Bonferroni correction.

Supplemental Information
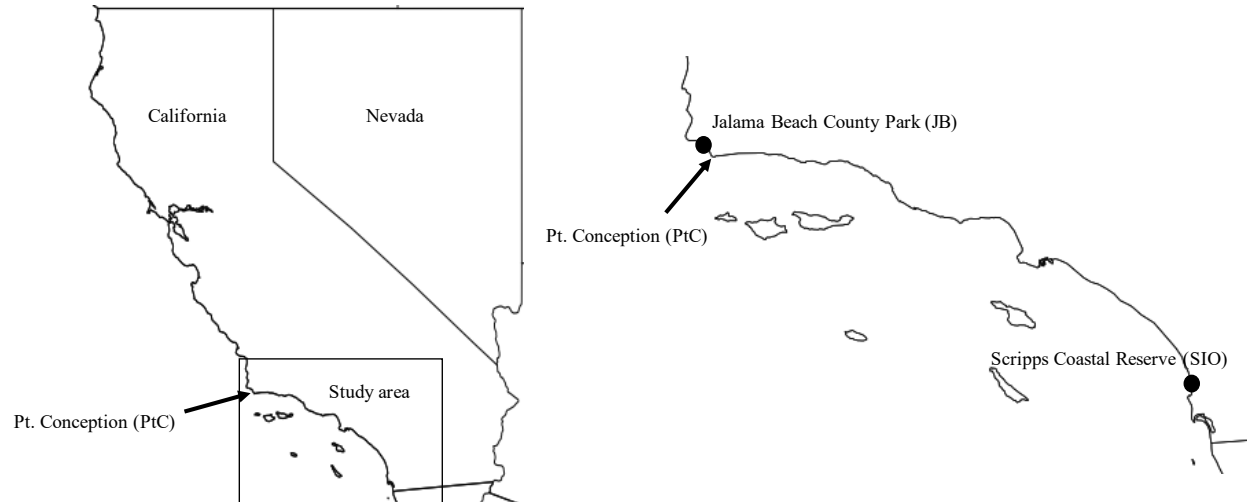
Supplemental Figure S1. Map showing the locations of the collection sites on each side of the biogeographic boundary at Point Conception, California.
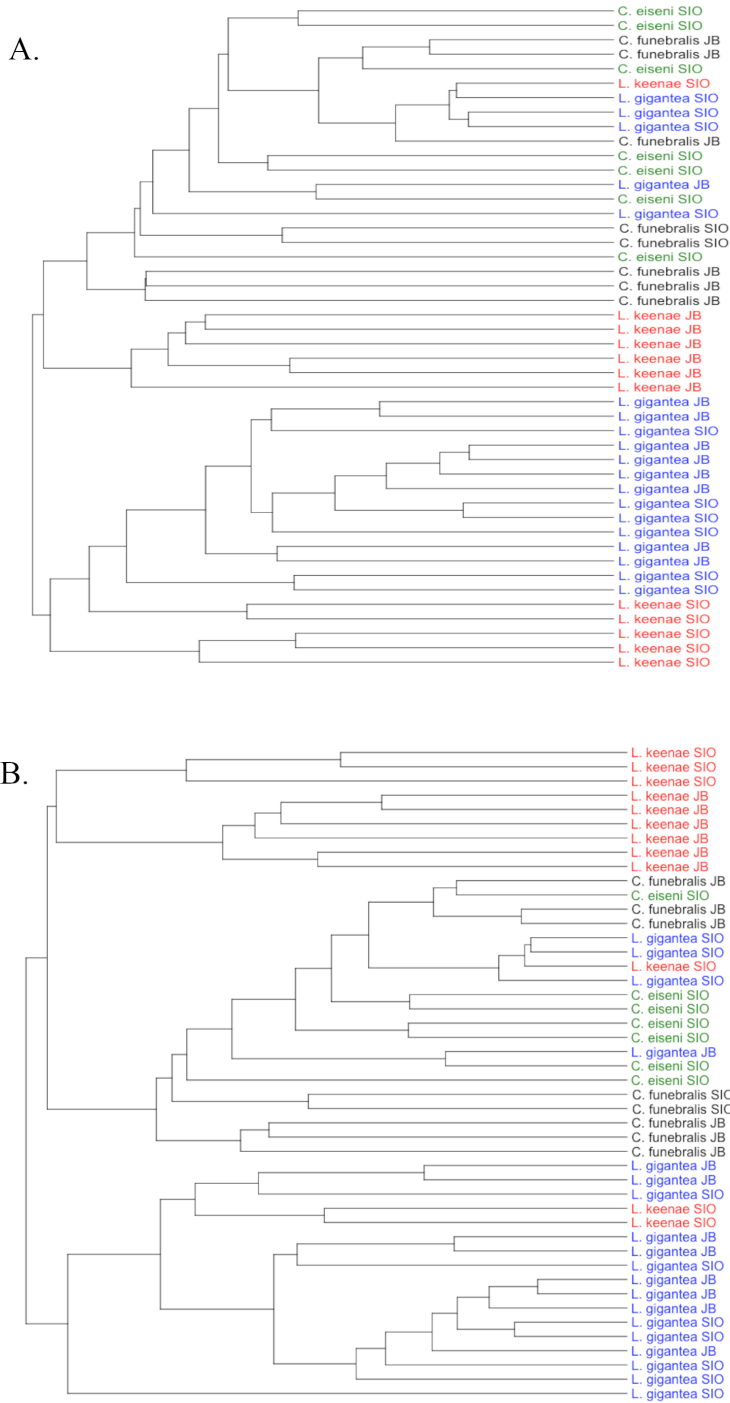
Supplemental Figure S2. Unweighted Pair Group Method with Arithmetic Mean (UPGMA) clustering analysis of gastropod microbiomes using Bray-Curtis dissimilarity using the entire dataset (A) and excluding singletons, doubletons and tripletons, as well as any OTUs found in fewer than 100% of individuals of a host species (B).
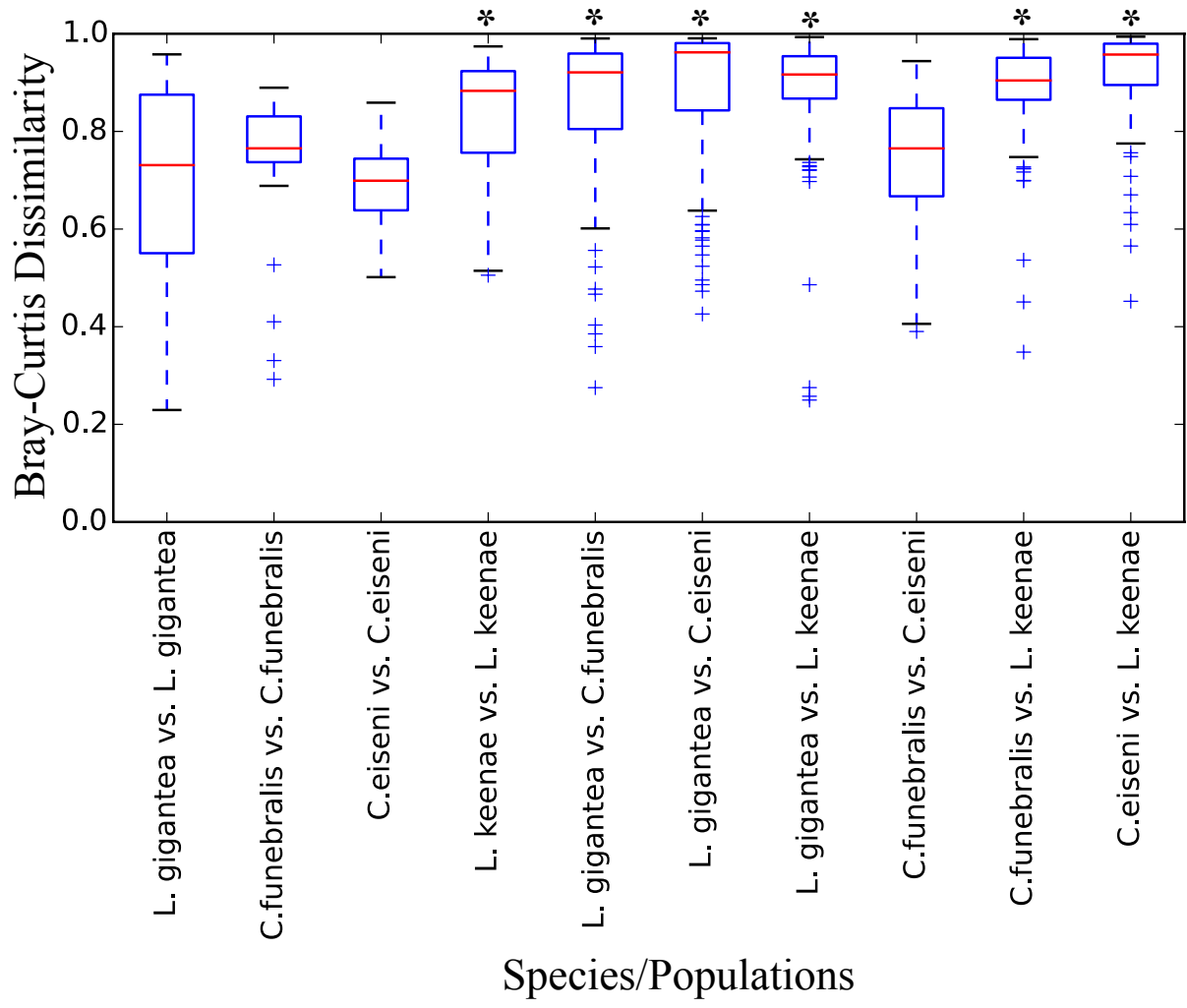
Supplemental Figure S3. Boxplot of Bray-Curtis dissimilarity values between all members of each species/population paired across Point Conception. An asterisk (*) denotes a species/population pair that is significantly different ($p < 0.05$) from the *C. funebralis-C. eiseni* pair by Student's two sample *t*-test with Bonferroni correction.

Acknowledgements

CHAPTER 2

Do host-associated microbes show a contrarian latitudinal diversity gradient? Insights from

Mytilus californianus, an intertidal foundation host

RESEARCH ARTICLE

Journal of Biogeography   WILEY

Do host-associated microbes show a contrarian latitudinal diversity gradient? Insights from *Mytilus californianus*, an intertidal foundation host

Alexander T. Neu[1] | Eric E. Allen[2,3] | Kaustuv Roy[1]

[1]Section of Ecology, Behavior and Evolution, Division of Biological Sciences, University of California San Diego, La Jolla, California, USA

[2]Section of Molecular Biology, Division of Biological Sciences, University of California San Diego, La Jolla, California, USA

[3]Marine Biology Research Division, Scripps Institution of Oceanography, University of California San Diego, La Jolla, California, USA

Correspondence
Alexander T. Neu, Section of Ecology, Behavior and Evolution, University of California San Diego, 9500 Gilman Drive MC0116, La Jolla, CA 92093, USA.
Email: aneu@ucsd.edu

Funding information
UCSD Academic Senate; National Aeronautics and Space Administration; American Museum of Natural History; Sigma Xi

Handling Editor: Rosemary Gillespie

Abstract

Aim: The latitudinal diversity gradient (LDG) is one of the most pervasive large-scale trends in biodiversity. However, whether microbial taxa also conform to this pattern remains poorly known. This study uses the gill and shell-surface microbiomes of the marine bivalve *Mytilus californianus* to test whether the diversity of host-associated microbial communities declines with increasing latitude, as predicted by the LDG.

Location: Northeastern Pacific coast, ranging from Sitka, AK to La Jolla, CA, USA (spanning 24.12 degrees of latitude).

Taxon: Bacteria and Archaea associated with *M. californianus* gills and shell-surfaces.

Methods: We amplified and sequenced the 16S rRNA gene from *M. californianus* gill and shell-surface samples. We used linear and quadratic regressions, accounting for spatial autocorrelation when needed, of average alpha and gamma diversities of the whole microbiome, as well as individual microbial clades, to determine whether these taxa conformed to the traditional LDG. We also used permutational multivariate analysis of variance to determine the level of compositional differentiation among sampling sites.

Results: Gill and shell-surface microbiota show differing latitudinal diversity patterns, with both micro-environments exhibiting high levels of compositional differentiation along latitude. Individual microbial clades within each micro-environment also show different latitudinal trends, most likely due to differing ecologies and life histories.

Main conclusions: *Mytilus californianus*-associated microbial taxa show a flat, hump-shaped or contrarian LDG, suggesting that a steep decline in diversity with increasing latitude may not be a universal pattern across different domains of life.

KEYWORDS
alpha diversity, gamma diversity, host–microbiome, latitudinal diversity gradient, microbial diversity, microbial ecology, microbiome composition, *Mytilus californianus*, rocky intertidal

1 | INTRODUCTION

The latitudinal diversity gradient (LDG), a decrease in species richness from low to high latitudes, is one of the most striking and pervasive large-scale trends in biodiversity. The LDG is likely to be 'ecology's oldest pattern' (Hawkins, 2001), having been the focus of research for over two centuries, and has been documented in the vast majority of plant and animal clades, both marine and terrestrial, although some exceptions exist (e.g. Pyron & Burbrink, 2009; Santelices & Marquet, 1998). However, despite many proposed hypotheses

(Jablonski et al., 2017; Mittelbach et al., 2007), the mechanisms underlying this pattern remain the subject of much discussion and debate (Jablonski et al., 2017; Schemske & Mittelbach, 2017). More importantly, whether latitudinal trends in the diversity of microbial taxa mimic those seen in plants and animals still remains poorly known. Information about how microbial diversity is structured across latitude and other major environmental gradients is essential for establishing whether the LDG is truly a universal pattern across all domains of life or whether it is a phenomenon primarily restricted to macro-eukaryotic clades. Furthermore, such knowledge is essential for testing the generality of existing ecological and evolutionary hypotheses about processes driving large-scale diversity gradients.

A limited number of studies have investigated the relationship between taxon richness and latitude in microbial communities, with mixed results. In soils, for example, latitude likely plays a less dominant role in affecting bacterial diversity and community composition, while pH is a driving factor (Chu et al., 2010; Fierer & Jackson, 2006). Diversity of planktonic marine bacteria, on the other hand, has been shown to exhibit a number of different latitudinal gradients ranging from a traditional LDG (Fuhrman et al., 2008; Ibarbalz et al., 2019) to an inverse LDG (Raes et al., 2017), a steady increase from the Arctic to the Antarctic (Moss et al., 2020) and a mid-latitude peak in average alpha diversity (Ladau et al., 2013; Milici et al., 2016). Far fewer studies have tested for the LDG in microbial taxa that are associated with animal hosts, despite the fact that many of these microbes are known to affect the biological processes of the host, and thus can potentially affect host LDG as well as their responses to climate change and other stressors (Apprill, 2017). These studies of the LDG in host-associated microbiomes do not provide a clear picture either. In humans, for example, a synthesis of multiple datasets found that the gut microbiome exhibited a traditional LDG in phylogenetic diversity (PD) and Shannon's index (H'), but not operational taxonomic unit (OTU) richness (Dikongué & Ségurel, 2017). Furthermore, a meta-analysis from the Earth Microbiome Project found a traditional LDG across plant-associated microbiomes, but a weak inverse LDG in animal-associated communities (Thompson et al., 2017). In targeted studies of fruit fly and sea anemone-associated microbiomes, no significant correlations between latitude and diversity were found (Corby-Harris et al., 2007; Morelan et al., 2019).

Taken at face value, the limited information available suggests that the decrease in species richness from low to high latitude that characterizes the vast majority of plant and animal clades does not universally apply to microbial communities, either free-living or host-associated. However, it is also clear that comparisons of published microbial LDGs with those of eukaryotic clades are not straightforward for a few important reasons—the spatial grain, phylogenetic resolution and geographical scale of the analyses.

1.1 | Sampling and spatial grain

For plants and animals, our knowledge of latitudinal trends in diversity is based almost exclusively on diversity estimates that are derived through interpolations. Sampling of local communities of plants and animals still remains sparse in many parts of the world, especially in the tropics. Because of this lack of information, diversity estimates for individual regions or latitudinal bins are commonly estimated using a range-through approach, where an individual species is assumed to be present everywhere within its geographical range limits, which are generally well constrained (Roy et al., 1998). Obviously, such interpolations tend to inflate estimates of local diversity and also dampen variations in richness within and between regions (Roy & Witman, 2009). Only a few studies have quantified the LDG using direct sampling of local plant and animal communities (e.g. Witman et al., 2004). In contrast, published analyses of microbial diversity patterns predominantly use alpha diversities at a very fine spatial grain (i.e. the average number of taxa observed in each environmental sample or host individual; but see Ray et al., 2019), and the geographical range limits of most microbial taxa still remain poorly constrained (but see Amend et al., 2012; Choudoir et al., 2017). Thus, published LDGs of macro-eukaryotes are based on diversity estimates at a different spatial scale compared to those of microbes as a result of how each of these groups of organisms are sampled in the field.

1.2 | Phylogenetic scale

A second issue is the focal phylogenetic level of analyses. Latitudinal trends in diversities of plants and animals are almost always quantified within individual clades, typically at the species level (Hillebrand, 2004a), but LDGs have also been shown to be present at higher taxonomic levels such as genera and families (e.g. Roy et al., 1996). For example, the literature is replete with examples of the LDG in birds (e.g. Hawkins et al., 2006), mammals (e.g. Rolland et al., 2014), marine molluscs (e.g. Roy et al., 1994) and so on. Microbial LDGs, in contrast, are often computed using the entire microbiome, thus including all microbial taxa in each sample (e.g. Moss et al., 2020) regardless of which family, order, class or phylum they belong to, a phylogenetic breadth that would be analogous to combining the diversity of all known clades of plants or animals. The issue of phylogenetic scale is not only important for comparing microbial latitudinal diversity patterns with those of plants and animals but also for testing hypotheses about the processes driving LDGs. Proposed hypotheses about the origin and maintenance of the LDG are based on differences in diversification rates and ecological processes such as dispersal and species–energy relationships (Jablonski et al., 2017; Mittelbach et al., 2007), most of which are known to be clade-specific (e.g. Hawkins et al., 2012).

1.3 | Geographical scale

The geographical distributions of hosts also present a challenge in analysing latitudinal diversity patterns of host-associated microbial taxa. Ideally, such analyses should be done using a single host to control for potential host effects. However, an exceedingly small number of species have geographical distributions ranging from

the tropics to the poles, thereby limiting the geographical scope of analyses possible using individual hosts. Thus, a more comprehensive understanding of the LDG in host-associated microbes requires analyses of individual hosts over their latitudinal ranges (e.g. Corby-Harris et al., 2007; Morelan et al., 2019) followed by meta-analyses across a number of different hosts (e.g. Thompson et al., 2017).

Given these differences, a better understanding of how the LDG is manifested across different domains of life would require comparative analyses using similar currencies and metrics. Specifically, it would require better information about microbial communities across different hosts and LDGs within specific microbial clades using estimates of both alpha and gamma diversities.

Here we address this issue by comparing latitudinal diversity patterns of individual microbial phyla and classes with the aggregate microbiome-level trends, using both alpha and gamma diversity metrics. We use populations of the California blue mussel (*Mytilus californianus* Conrad, 1837), a foundation species in northeastern Pacific rocky intertidal communities, sampled across more than 24 degrees of latitude (from southeastern Alaska to southern California, USA) as the focal host. *M. californianus* is an excellent system for this type of study as it has a broad range from warm temperate to polar waters, is found in high abundances in wave-exposed rocky intertidal habitats and exhibits high levels of gene flow across its geographical range (Addison et al., 2008), thereby reducing potential genotype-specific effects. We sampled both the gill tissues and shell-surface biofilms of *M. californianus* to compare latitudinal diversity patterns between micro-environments with different levels of exposure to ambient conditions, as the compositions of internal (gut) and external (skin, leaf, etc.) microbial communities have been shown to respond differently to host-derived and environmental factors (Woodhams et al., 2020). We hypothesized that the gill and shell-surface microbiomes of this host, as well as individual microbial clades, would show a decrease in both alpha and gamma diversities with increasing latitude, as has been seen in animals and free-living microbes (e.g. Fuhrman et al., 2008; Hillebrand, 2004b). As our southernmost site

is extratropical (~32.87°N), this would be the expectation regardless of whether these communities exhibit a traditional LDG or the hump-shaped pattern occasionally observed in planktonic marine microbes (Ladau et al., 2013; Milici et al., 2016).

2 | MATERIALS AND METHODS

We sampled *M. californianus* gill tissues at 11 sites ranging from Chilean Memorial, WA to Scripps Institution of Oceanography (SIO), La Jolla, CA (Table 1). Single gill samples were excised from between 6 and 12 individuals collected from a single rock face at each site (Table 1). Gills were then triple-rinsed with 95% EtOH and preserved on dry ice <2 h after sampling. Whole gills were used to determine whether any tissue-associated symbionts were present, as has been previously reported in *Mytilus* species (Schill et al., 2017). We also sampled *M. californianus* shell-surface microbial communities at 11 sites ranging from Pirate's Cove, Sitka, AK to SIO (Table 1). We sampled shell-surface communities in the field by swabbing the exterior surface of a single valve from each individual, using a sterile 152 mm swab (Grenier Bio-One). We took particular care to avoid epibionts, as these most likely have their own unique microbial communities (Bengtsson et al., 2017). Once the surface of the mussel shell was thoroughly swabbed, we returned the swabs to their sterile tubes and placed them on dry ice within 2 h of collection. Gills and shell-surface swabs remained on dry ice until they were returned to the laboratory at the University of California San Diego and frozen at −80°C. We collected samples during August 2017 and August 2018 to minimize any seasonal effects. To account for any potential variation between collection times, we conducted an initial set of analyses using only samples collected in 2017. These analyses produced qualitatively similar results to those derived from the combined 2017–2018 dataset (Supporting Information 1), so we chose to present results from the entire dataset for a more complete latitudinal sampling.

TABLE 1 Name, latitude, longitude and number samples collected from each site

Site name	Site abbreviation	Latitude (°N)	Longitude (°W)	Shell-surface swabs (*n*)	Gill tissues (*n*)
Pirate's Cove, AK	PC	56.99	135.37	8	—
Chilean Memorial, WA	CM	47.96	124.66	9	12
Fogarty Creek, OR	FC	44.84	124.05	10	10
Crescent City, CA	CC	41.75	124.20	8	12
Trinidad Head, CA	TH	41.06	124.15	10	8
Schoolhouse Beach, CA	SC	38.37	123.08	10	9
Moss Landing, CA	ML	36.81	121.79	7	8
Cayucos, CA	CAY	35.45	120.90	11	9
Jalama Beach, CA	JB	34.51	120.50	10	8
El Capitan, CA	EC	34.46	120.03	9	10
Pt. Fermin, CA	PF	33.71	118.30	—	6
La Jolla, CA	SIO	32.87	117.25	11	10

We used the MoBio PowerSoil kit (MoBio Inc.; now part of Qiagen GmbH) to extract DNA from *M. californianus* gill tissue and shell-surface swabs following the manufacturer's instructions, but included an additional 10-min incubation at 55°C after bead beating. We amplified the V4 region of the 16S ribosomal RNA gene using the primers 515f (GTGYCAGCMGCCGCGGTAA; Parada et al., 2016) and 806r (GGACTACNVGGGTWTCTAAT; Apprill et al., 2015) and the two-step PCR protocol described in Neu et al. (2019). We purified the resulting PCR products using an Exo-SAP approach (1 U of Exonuclease I and 1 U of Shrimp Alkaline Phosphate; New England Biolabs), and heated them at 37°C for 30 min followed by 85°C for 15 min. We conducted a final purification step using Ampure XP beads (Beckman Coulter, Inc.), following the manufacturer's instructions. Amplicons were sent for sequencing on the MiSeq platform (Illumina, 2 × 250 bp output) at the DNA Technologies Core, University of California Davis.

We used 'DADA2' 1.10.1 (Callahan et al., 2016) in R 3.5.2 (R Core Team, 2013) for processing raw sequence data (quality scores shown in Figure S1). We first trimmed the forward and reverse reads to 200 and 180 bp, respectively, then filtered, denoised and merged them to consensus sequences of 253 bp. The resulting amplicon sequence variants (ASVs) were assigned taxonomy using the Silva database 138 (McLaren, 2020; Quast et al., 2013) and all sequences matched to chloroplasts and mitochondria were removed. Taxonomic identities of the dominant (>10% relative abundance in at least one sample) gill-associated ASVs within the family Endozoicomonadaceae were further investigated using the National Center for Biotechnology Innovation (NCBI) BLASTn function (https://blast.ncbi.nlm.nih.gov/Blast.cgi). We chose to use ASVs for the majority of our analyses since they have become the preferred taxonomic unit for many microbial ecology studies (Callahan et al., 2017) and showed qualitatively similar results when compared to 97% OTUs in initial testing (Figure S2).

All statistical analyses were conducted in R 3.5.2 (R Core Team, 2013) primarily using the packages 'phyloseq' 1.26.1 (McMurdie & Holmes, 2013), 'vegan' 2.5-6 (Oksanen et al., 2019) and 'picante' 1.8 (Kembel et al., 2010). A representative ASV table was selected after rarefying to 5000 and 4000 sequences per sample in gill and shell-surface communities, respectively (Figure S3). We found that the number of observed ASVs in the rarefied and unrarefied datasets were highly correlated and should therefore present qualitatively similar diversity patterns (Figure S4). Furthermore, rarefied data have been shown to present similar results compared to data standardized using other available approaches (Weiss et al., 2017). Initial analyses performed with 10 unique rarefied ASV tables also produced qualitatively similar results, so we present results of a single rarefaction on each dataset (set.seed = 100) throughout.

We used three different metrics of alpha diversity—observed ASV richness, Shannon's *H'* and Faith's PD (Faith, 1992)—and tested for latitudinal trends using linear regressions. To further determine whether higher taxonomic levels exhibited similar patterns to ASVs, as in some eukaryotic clades (Roy et al., 1996), we calculated the number of taxa present at the phylum, class, order, family and genus

levels and tested for correlations with latitude using linear regressions. It should be noted, however, that a number of ASVs remain unassigned at each taxonomic level due to the resolution currently available for microbial taxa and therefore the number of taxa at each level are likely substantially underestimated. For example, an ASV that could only be assigned to the phylum level was included in the phylum-level analyses, but excluded from analyses at lower taxonomic levels. We used Moran's *I* to evaluate any effects of spatial autocorrelation on the observed diversity patterns using the package 'ape' 5.3 (Paradis & Schliep, 2019). In cases where spatial autocorrelation was detected, we used spatial lag models in the package 'spatialreg' 1.1-5 (Bivand & Piras, 2015) to account for it while testing for latitudinal trends. While it is possible that there are population-specific effects that are unaccounted for in this model, our aim is to quantify the overall latitudinal diversity trend using methods comparable to those commonly used for eukaryotic studies, and therefore we have not accounted for them here. Finally, we used linear regressions of observed ASV richness vs. shell length to investigate whether microbial community richness scales with body size of *M. californianus* individuals, as predicted by the species–area relationship (e.g. Sherrill-Mix et al., 2018).

Following a common approach in macroecology, we used the number of ASVs whose geographical ranges intersect a particular locality as our regional or gamma diversity metric for that locality (Roy & Witman, 2009). Note that this definition of gamma diversity is different from the traditional approach where gamma is estimated by pooling alpha diversities (e.g. Magurran, 1988), but is widely used in previous analyses of LDG (Roy & Witman, 2009). Thus, our results should be directly comparable to many published LDGs of plants and animals. The geographical range of each ASV was computed using a range-through approach where a taxon is assumed to be present at all the sites between its sampled geographical range limits, even if it is not explicitly sampled at those intervening sites. Thus, our gamma diversity metric provides an estimate of the maximum possible richness at a given site (Roy & Witman, 2009), while the alpha diversity metric is likely to be a minimum estimate because of sampling incompleteness (Shade & Gilbert, 2015; Sogin et al., 2006; Weiss et al., 2017). We tested for changes in gamma diversity with latitude using both linear and quadratic regressions. We hypothesized that these microbial communities would exhibit similar alpha and gamma diversity trends, as has been seen in some marine invertebrate communities (Witman et al., 2004).

To determine whether the overall alpha and gamma diversity trends were representative of individual microbial clades within the community, or rather an aggregate pattern, we repeated our alpha and gamma diversity analyses for each of four focal clades: Alphaproteobacteria, Bacteroidota, Gammaproteobacteria and Planctomycetota. These clades were chosen due to their high levels of ASV richness and relative abundance in both gill and shell-surface, making them easily comparable across *M. californianus* micro-environments.

We also hypothesized that *M. californianus* microbiomes would significantly differ in composition among collection sites, as is commonly found in host-associated microbial studies. We tested for

these compositional differences via permutational multivariate analysis of variance (PERMANOVA) of Bray–Curtis dissimilarity (BCD) values, using individual samples, site-level alpha (i.e. all microbial taxa detected at a site by combining all individuals of the host sampled at that site) and range-through data. In addition, we used principal coordinates analysis (PCoA) plots to visualize the differences. For the site-level alpha and gamma diversity comparisons, we determined which microbial phyla (with Proteobacteria split into classes for increased resolution) significantly contributed to increasing diversity at particular latitudes using the gg_envfit function in the 'ggordiplots' package 0.3.0 (Quensen, 2018). To help determine whether spatial compositional turnover, differences in relative abundance or differences in phylogenetic relationships between taxa were driving observed differences, we conducted PERMANOVAs on individual samples using three different beta diversity metrics: (i) the Jaccard index, which calculates dissimilarity between communities based only on the presence or absence of taxa (ii) unweighted UniFrac, which takes phylogenetic relatedness of microbial taxa into account, but not their relative abundance and (iii) weighted UniFrac, which includes phylogenetic relatedness and relative abundance information (Lozupone & Knight, 2005). In addition, PERMANOVAs using BCD and phylum-, class-, order-, family- and genus-level data were used to determine the taxonomic level at which significant compositional differences, if any, occurred. We also conducted PERMANOVAs using BCD and members of each of our four focal microbial clades to determine whether there were significant compositional differences across sites within each of these groups. For all PERMANOVA analyses, we also tested for homogeneity of dispersions, an assumption of PERMANOVA, using the betadisper function in 'vegan' 2.5-6 Oksanen et al. 2019. Finally, to determine whether there was a latitudinal pattern of spatial compositional turnover between communities, we tested for a relationship between the Jaccard dissimilarity of sites and the geographical distance between them.

3 | RESULTS

3.1 | Microbial communities

The gill dataset contained 4,103,234 sequences representing 16,166 ASVs across 102 samples. Four samples were removed after processing due to low sequence counts, leaving 98 samples and 11,089 ASVs in the final rarefied dataset. Gill microbial communities were variable in composition, though Bacteroidota (14.3 ± 8.0% mean relative abundance ± SD) and Alphaproteobacteria (13.0 ± 7.8%) were abundant in samples across all latitudes (Figure 1a). Gammaproteobacteria were the most dominant members of many gill samples, making up 45.7 ± 18.7% of the relative abundance (Figure 1a). This was driven largely by members of the family Endozoicomonadaceae which comprised up to 63.3% of the community in certain individuals. NCBI BLASTn results showed that these taxa were closely related (>94% sequence similarity) to other members of this clade known to be associated with marine bivalves (Table S1).

A total of 4,640,619 sequences were processed from 103 shell-surface samples, generating 19,208 ASVs. All 103 samples as well as 13,737 ASVs were retained in the final, rarefied dataset. Shell-surface microbiomes varied greatly in the relative abundances of the most dominant groups, including Actinobacteriota (6.9 ± 4.1%), Bacteroidota (23.2 ± 5.5%), Alphaproteobacteria (17.8 ± 6.9%), Gammaproteobacteria (15.2 ± 8.0%) and Cyanobacteria (14.2% ± 10.2%) (Figure 1b).

3.2 | Latitudinal trends in alpha versus gamma diversity

3.2.1 | Gill-associated communities

Overall, *M. californianus* gill microbiomes showed a very weak, but significant LDG in both average ASV richness (Figure 2a) and Shannon's H' (Figure 2b), partially in line with our hypothesis, while PD did not show a significant trend with latitude (Figure 2c). However, this LDG pattern was only present at the level of individual ASVs and was not found in microbial genera or any other higher taxa (Figure S5). ASV richness in the gill community was not significantly correlated with shell length, thus ruling out any body size effects (Figure S6). Gamma diversity patterns differed from that of average alpha diversity in the gill environment as it did not show a monotonic latitudinal trend. Instead, it followed a hump-shaped trajectory better described by a quadratic term, with richness peaking at ~37°N and declining towards the edges of the sampled range (Figure 2d). It should be noted, however, that such a quadratic trend may be partially due to the nature of the gamma diversity metric used, which interpolates values based on a range-through assumption and potentially leads to edge effects at the ends of the sampled range.

In contrast to the overall pattern, latitudinal trends in alpha diversity within individual microbial clades varied considerably. Alphaproteobacteria and Gammaproteobacteria showed a weak but significant LDG in ASV richness (Figure 3a,c), similar to the whole community pattern, while Bacteroidota and Planctomycetota richnesses were not significantly correlated with latitude (Figure 3b,d). The average ASV richness of Alphaproteobacteria showed significant spatial autocorrelation (Table 2), but the latitudinal trend remained significant after accounting for it (Z-value = −4.019, $p < 0.001$). Gamma diversity trends of these groups largely mirrored the non-linear trend seen for the whole community, peaking at mid-latitudes with a decline towards both ends of the sampled range (Figure 3e–h). The phylum Planctomycetota, however, was unique in exhibiting an additional gamma diversity peak at ~42°N (Figure 3h).

3.2.2 | Shell-surface communities

Contrary to our hypothesis, shell-surface microbiomes increased in average alpha diversity with increasing latitude, resulting in an inverse LDG in ASV richness (Figure 2e), Shannon's H' (Figure 2f)
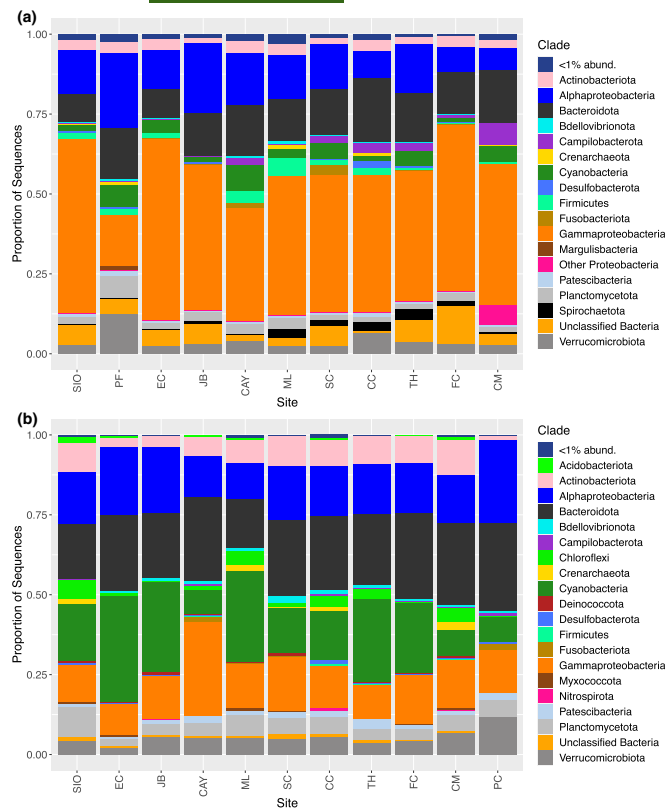
and PD (Figure 2g), which is independent of shell size (Figure S6). This trend was also evident in microbial genera, families and orders, with the strength of the gradient increasing with increased taxonomic resolution (Figure S7; Tables S2 and S3). Taxonomic richness and PD data were significantly spatially autocorrelated (Table 3), but the inverse LDG patterns remained significant after accounting for it (Observed ASVs: Z-value = 3.630, $p < 0.001$; PD: Z-value = 2.184, $p = 0.029$). Shell-surface microbiomes also conformed to the non-linear latitudinal trend in gamma diversity, with increasing total ASV richness from the southern to northern sites, followed by a large drop between the two northernmost collection sites, potentially due to edge effects (Figure 2h).

Individual microbial clades followed the same pattern as the entire community for average alpha diversity, though with a large increase in richness between 41°N and 42°N (Figure 3a–d). Significant spatial autocorrelation was present in three of these clades (Table 2) but did not drive this trend (Alphaproteobacteria: Z-value = 4.547, $p < 0.001$; Bacteroidota: Z-value = 3.545, $p < 0.001$; Planctomycetota: Z-value = 2.199, $p = 0.028$). All of the focal clades showed a hump-shaped latitudinal pattern in gamma diversity, similar to the whole

community pattern (Figure 4e–g), but the phylum Planctomycetota was once again somewhat distinct from the other clades, with peaks at both 42°N and 48°N (Figure 4h).

3.3 | Compositional analyses

The PCoA plots and PERMANOVAs reveal significant compositional differentiation by geographical location for both gill and shell-surface microbiomes (Figure S8). These compositional differences were present at all taxonomic levels, across all diversity metrics tested, and in the four focal microbial clades in both gill and shell-surface environments (Tables S4–S9). Betadisper analyses revealed significant differentiation in dispersion between sites in a number of cases, largely in the shell-surface microbiome (Tables S4–S9). However, a Tukey's HSD test showed that significant differences in dispersion in the shell-surface microbiome were only present when comparing the samples from Cayucos, CA, which had low dispersion, to other samples. Removing the Cayucos samples from our analyses did not qualitatively impact the PERMANOVA results (PERMANOVA: Pseudo-F = 5.741,

WILEY
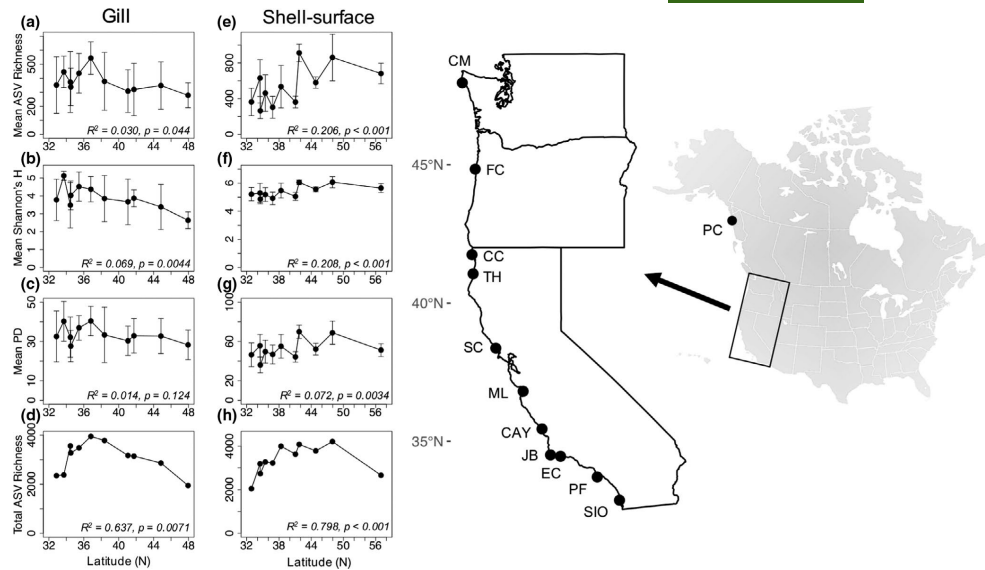
FIGURE 2 Mean ASV richness (a, e), mean Shannon's *H'* (b, f), mean PD (c, g) and total ASV richness (d, h) of *Mytilus californianus* gill (a–d) and shell-surface (e–h) microbiomes across latitude. Error bars reflect the standard error of each metric at each site. Results of linear regressions of alpha diversity versus latitude are presented in the bottom right corners of (a–c, e–g). Results of quadratic regressions of gamma diversity versus latitude are summarized in the bottom right corners of (d, h). The map of North America (top right) shows the location of Pirate's Cove, AK and the black box marks the location of the detailed map of the rest of the collection sites along the west coast of the United States. Map location labels correspond to the site abbreviations in Table 1. ASV, amplicon sequence variant; PD, Phylogenetic diversity

$R^2 = 0.384$, $p < 0.001$; betadisper: $p = 0.145$). Results from the gill microbiome showed that groups with larger sample sizes exhibited greater dispersion, suggesting that our PERMANOVA results may be more conservative than expected (Anderson & Walsh, 2013). In our analyses of site-level compositional differences (i.e. using site-level alpha and range-through approaches), the sites were arrayed in horseshoe patterns on the PCoA plots (Figure 4). In addition, plots of compositional turnover with geographical distance showed that dissimilarity increased with distance, consistent with a distance decay pattern in the shell-surface microbiome, but not the gill. In both micro-environments, however, all pairs of sites showed high (Jaccard index >0.8) levels of dissimilarity (Figure S9).

The microbial taxa driving community differentiation varied greatly between gill and shell-surface. Alphaproteobacteria was the only group shown to significantly impact site-level alpha diversity in gills, with higher diversity in the southern sampling regions (Figure 5a). Using the range-through approach, however, many more taxa emerged as important contributors to diversity, mainly in the region between 35.5°N and 37°N (Figure 5c), in line with the peak in gill gamma diversity (Figure 2d). In shell-surface communities, Verrucomicrobiota and Patescibacteria were the only two phyla shown to drive increased diversity in the northern portion of the range using our site-level metric (Figure 5b). Incorporating

range-through data highlighted six additional groups that significantly increased in diversity in the northern sites (Figure 5d), as expected given the northward increase in average alpha and gamma diversity in shell-surface communities (Figure 2h).

4 | DISCUSSION

4.1 | Latitudinal diversity trends

Mytilus californianus gill microbiomes contained a number of groups common to marine environments including Bacteroidota, Alphaproteobacteria and Gammaproteobacteria. The most common and highly abundant taxa in these samples were members of the family Endozoicomonadaceae, which have been associated with marine invertebrates from corals to tunicates (Neave et al., 2017), as well as the gills of multiple bivalve species (Dubé et al., 2019; Schill et al., 2017). While some members of this family have been predicted to play a role in the health of their hosts (Neave et al., 2017), others have developed into intracellular parasites (Zielinski et al., 2009). The most abundant Endozoicomonadaceae sequences in our dataset were indeed closely related to other bivalve-associated taxa, although their ecological and functional roles remain unknown.
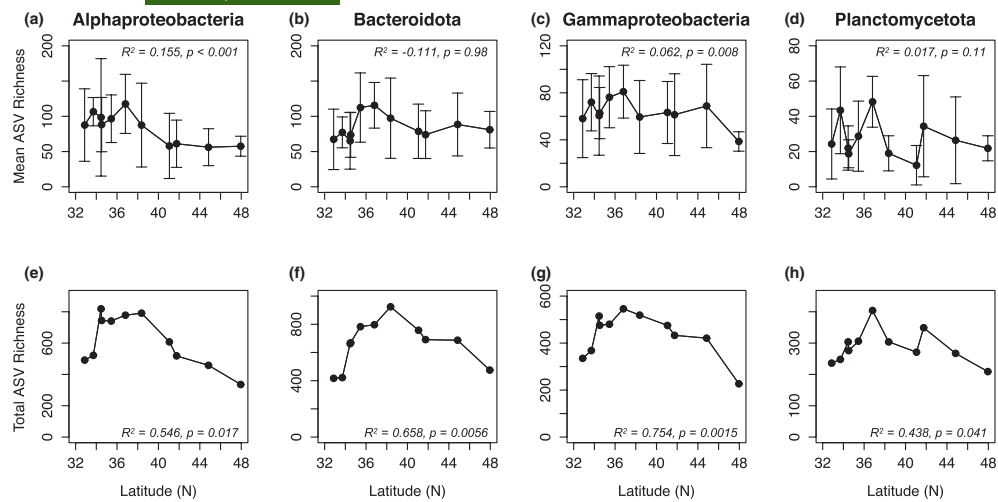
FIGURE 3 Mean ASV richness (a–d) and total ASV richness (e–h) of four diverse bacterial clades including Alphaproteobacteria (a, e), Bacteroidota (b, f), Gammaproteobacteria (c, g) and Planctomycetota (d, h) in *Mytilus californianus* gill microbiomes. Error bars reflect the standard error of each metric at each site. Results of linear regressions of alpha diversity versus latitude are presented in the top right corners of (a–d). Results of quadratic regressions of gamma diversity versus latitude are presented in the bottom right corners of (e–h). ASV, amplicon sequence variant

Clade	Gill		Shell-surface	
	Moran's *I*	*p*-value	Moran's *I*	*p*-value
Alphaproteobacteria	0.0735	<0.001	−0.121	<0.001
Bacteroidota	0.00222	0.443	−0.0607	<0.001
Gammaproteobacteria	−0.0157	0.676	−0.0246	0.267
Planctomycetota	−0.0209	0.161	−0.0752	<0.001

TABLE 2 Results of Moran's *I* analysis of observed amplicon sequence variant richness for the four focal clades from gill and shell-surface communities

Average ASV counts and Shannon's *H'* of gill communities were significantly, but not strongly, correlated with latitude, while PD did not show a significant latitudinal trend. The weak LDG pattern in richness was only detectable at the ASV level and not at higher taxonomic levels. Together, these results suggest that external environmental variables that change with latitude, and have been shown to be strong correlates of the LDG in marine eukaryotes (e.g. Belanger et al., 2012; Roy et al., 1998), are unlikely to be strong drivers of latitudinal variation in microbial diversity in *M. californianus*. Instead, local micro-environmental conditions in the gill mediated by host physiology and diet variability (e.g. Dahlhoff & Menge, 1996) may have a stronger influence on the LDG of the gill microbiome. In addition, the internal body temperature of *M. californianus*, which differs from external water temperatures and does not appear to show a clear latitudinal trend (Helmuth et al., 2002), could also play a role in mediating the diversity of the gill microbiome. Unfortunately, measurements of these potentially relevant variables are not currently available for our samples, which precludes statistical tests

of these hypotheses. Previous studies have also shown a lack of strong correlation between microbiome alpha diversity and latitude, and hypothesized that this may be due to the interacting impacts of multiple local abiotic and biotic factors on microbial diversity (Corby-Harris et al., 2007; Morelan et al., 2019). Finally, the lack of a strong latitudinal trend in whole microbiome diversity may also be partially explained by the conflicting patterns within individual clades. For example, in the gill environment, Alphaproteobacteria and Gammaproteobacteria aligned with the whole community trend and showed weak but significant LDGs in ASV richness while Bacteroidota and Planctomycetota did not. Aggregating the contrasting latitudinal trends across these and other clades could lead to a dampened LDG signal in the whole community dataset and obscure the underlying variation. It is not surprising that these diverse groups present different diversity patterns, as they are characterized by unique life habits and ecologies. Alphaproteobacteria, for example, are highly abundant free-living and particle-associated members of the coastal marine environment, while Planctomycetota are often

34

TABLE 3 Results of Moran's *I* analysis for all alpha diversity indices from gill and shell-surface communities

	Gill		Shell-surface	
Diversity index	Moran's *I*	*p*-value	Moran's *I*	*p*-value
Observed amplicon sequence variants	0.00853	0.186	−0.0685	<0.001
Shannon's *H'*	0.00382	0.401	0.0105	0.128
Phylogenetic diversity	−0.00796	0.670	−0.0795	<0.001
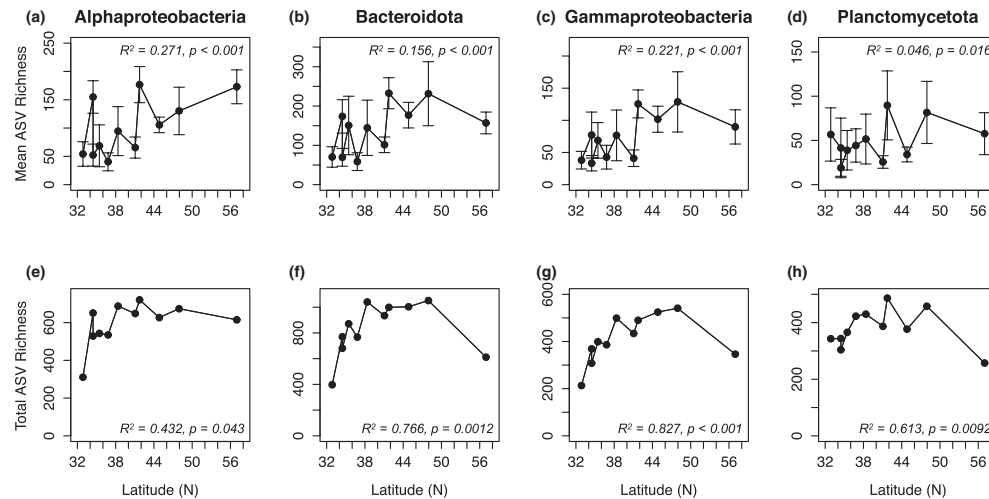
FIGURE 4 Mean ASV richness (a–d) and total ASV richness (e–h) of the four focal bacterial clades including Alphaproteobacteria (a, e), Bacteroidota (b, f), Gammaproteobacteria (c, g) and Planctomycetota (d, h) in *Mytilus californianus* shell-surface microbiomes. Error bars reflect the standard error of each metric at each site. Results of the linear regression of alpha diversity versus latitude are presented in the top right corners of (a–d). Results of quadratic regressions of gamma diversity versus latitude are presented in the bottom right corners of (e–h). ASV, amplicon sequence variant

found on surfaces or in close association with phototrophs such as macroalgae (Wiegand et al., 2018). In addition, each of these clades are likely to be characterized by unique patterns of evolutionary diversification, a parameter that is known to influence the LDG (e.g. Mittelbach et al., 2007). All of this suggests that future analyses of the microbial LDG should investigate trajectories and patterns within individual microbial clades, as is common in eukaryotic systems.

Gill microbiomes did not show a monotonic LDG in gamma diversity, but instead generally peaked around mid-latitudes, a pattern that is not typically observed in larger eukaryotic clades. The whole community gamma diversity trend was broadly conserved across the four individual clades, though the peak shifted slightly in Bacteroidota and a second peak was evident in Planctomycetota. These results support the idea that latitudinal trends in average alpha and gamma diversity need not be concordant (Roy & Witman, 2009), and further highlight the importance of considering both metrics when comparing latitudinal diversity patterns of microbial taxa with those of higher eukaryotes.

The surface biofilms of marine organisms are incredibly dynamic and are impacted by the physical environment of the host as well as interactions, whether mutualistic or antagonistic, between biofilm-forming organisms (Wahl et al., 2012). The biofilm communities on *M. californianus* shell-surfaces varied in their compositions within and between sites, but were largely comprised of Gammaproteobacteria, Alphaproteobacteria, Bacteroidota and Cyanobacteria. These groups have previously been found to dominate the shell-surface microbiomes of other molluscan species and some members of these communities are known to play a significant role in denitrification in the tidepool environment (Pfister et al., 2014).

In contrast to gill, alpha diversity of shell-surface microbiomes generally displayed an inverse LDG—increase in diversity with increasing latitude—across all diversity metrics tested. ASV richness was not correlated with shell size, suggesting that factors other than total available space for settlement drive this pattern. Furthermore, these trends were evident at both the ASV and higher taxonomic levels, as well as within individual clades. This suggests that higher
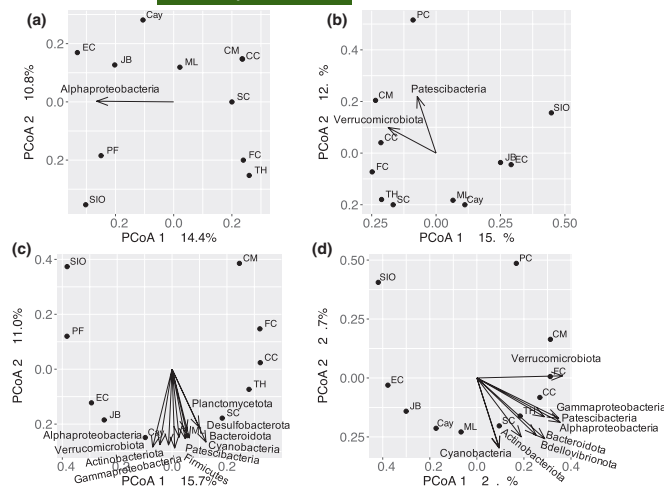
FIGURE 5 Plots of the first two axes of principal coordinates analyses (PCoA) using Bray–Curtis dissimilarity for site-level alpha diversity (a, b) and range-through data (c, d) for *Mytilus californianus* gill (a, c) and shell-surface (b, d) microbiomes. Arrows denote taxa with significantly increased diversity at a particular latitude. Point labels correspond to the site abbreviations in Table 1

taxa may be useful proxies for ASV-level latitudinal diversity trends in shell-surface communities. Such a positive relationship between latitude and diversity has been seen in a meta-analysis of microbiomes of animal hosts (Thompson et al., 2017) as well as in the bacterioplankton of the eastern Indian Ocean (Raes et al., 2017).

Similar to that seen in the gill environment, gamma diversity of shell-surface communities exhibited a nonlinear latitudinal trend, with peaks in the northern latitudes. These results should be interpreted carefully, however, since the large drop between our two northernmost sites could reflect edge effects. The increase in gamma diversity with latitude for the majority of the sampled range, on the other hand, is not explained by such edge effects and may reflect an increase in the size of the regional species pool (bacterioplankton) with increasing latitude. Such an increase would provide an opportunity for more unique taxa to settle on the surface of each shell, leading to greater average alpha diversity and gamma diversity. This may also reflect an increase in the level of physical disturbance (e.g. wave action) with latitude, which could similarly increase diversity. Like in the gill, the exact processes underlying the latitudinal changes in the diversity of shell-surface communities seen here remain unclear, although environmental variables such as temperature and dissolved inorganic nitrogen may play a role, as has been proposed for free-living marine bacterial diversity (Milici et al., 2016; Raes et al., 2017). As discussed above, such hypotheses remain difficult to test given that temperature and other environmental conditions on *M. californianus* shell-surfaces are likely to be very different from that of the surrounding water (Helmuth et al., 2002) and measurements of those parameters are not currently available.

Overall, the latitudinal patterns observed in the gill and shell-surface microbiomes are quite different, and neither clearly mirrored the traditional LDG found in plant and animal taxa, even when using similar currencies and metrics. Marine bivalves, for example, exhibit a steady decrease in species richness with increasing latitude across the range sampled here (Roy et al., 2000; Schumm et al., 2019). The differences observed in the diversity patterns between the two host micro-environments investigated here are likely due to their differential exposure to different environmental and host-derived (e.g. immune function and diet variability) factors, as has been recently shown in a number of other host species (Woodhams et al., 2020). In addition, individual clades within both the gill and the shell-surface communities differed in their diversity patterns, demonstrating that simply using a 'whole microbiome' approach misses a remarkable level of complexity. The divergent patterns outlined in this study should also be useful for discerning the mechanisms involved in determining microbial diversity. Previous studies highlighting patterns contrary to the traditional LDG in specific clades (Krug et al., 2007) and time periods (Song et al., 2020) have shed light on the roles of diversification and extinction rates in structuring the diversity of marine taxa. In a similar way, the contrarian clades and micro-environments associated with the *M. californianus* microbiome may provide a greater mechanistic understanding of microbial diversity patterns.

4.2 | Community composition

Both gill and shell-surface microbiome compositions differed significantly by site according to multiple beta diversity metrics used here. Such differentiation when using the Jaccard coefficient suggests high levels of spatial turnover of ASVs among our sites. Sites were arrayed in a horseshoe pattern in the PCoA plots of both site-level alpha and gamma diversities. Although such a horseshoe has traditionally been thought to be an artefact of dimensionality reduction, recent studies suggest that it is likely to represent a true biological signal. Specifically, a horseshoe may arise when the distance metric used (BCD, in this case) approaches saturation (Morton et al.,

2017). Although geographical distance between sites can increase indefinitely, distance metrics are often bounded (BCD is bounded between 0 and 1), and once saturation (BCD = 1) is achieved, all additional samples along a gradient will be the same ecological distance from one another. Plotting this pattern in two-dimensional space, as in a PCoA plot, results in the horseshoe effect (Morton et al., 2017). Such a pattern suggests that sites are highly distinct compositionally, with large differences between even closely related sites, as demonstrated by the relationship between Jaccard dissimilarity and geographical distance documented here. This was further confirmed by the PERMANOVAs across taxonomic levels, which found that site explained a significant amount of the variation between samples, even at the phylum level, in both gill and shell-surface communities. Whether this consistent shift in community composition is due to changes in the regional species pool, differential filtering by the host at different sites, or both, remains unclear.

5 | CONCLUSIONS

Using samples from different micro-environments within the same host across a large part of its geographical range, we have shown that host-associated microbial taxa do not generally follow the same latitudinal diversity patterns as many larger eukaryotic clades, even when accounting for spatial grain and phylogenetic scales. Differences in host control and environmental conditions between shell-surface and gill likely have a significant impact on latitudinal trends in diversity within each micro-environment. Gill microbiome diversity was variable, but generally presented a weak or non-significant correlation with latitude, while shell-surface communities followed an inverse LDG across multiple taxonomic levels. Individual clades varied in diversity patterns as well, depending on their life history and the micro-environment investigated. However, all of these diversity trends were underlain by steady and significant compositional changes between sites. Overall, our results highlight the many layers of organization within host-associated microbiomes that are revealed through analyses at different spatial and phylogenetic resolutions and open the door for further investigation into the mechanisms driving these latitudinal trends. In addition, they also suggest that the nature of latitudinal trends in diversity most likely differs across the different domains of life. Future studies of microbial LDG, both host-associated and free-living, are needed to better understand the similarities and differences in large-scale biodiversity gradients across the tree of life.

CONFLICTS OF INTEREST

We declare no conflict of interest.

ORCID

Alexander T. Neu https://orcid.org/0000-0003-0833-1704
Eric E. Allen https://orcid.org/0000-0002-1229-8794
Kaustuv Roy https://orcid.org/0000-0003-3693-1045

REFERENCES

Addison, J. A., Ort, B. S., Mesa, K. A., & Pogson, G. H. (2008). Range-wide genetic homogeneity in the California sea mussel (*Mytilus californianus*): A comparison of allozymes, nuclear DNA markers, and mitochondrial DNA sequences. *Molecular Ecology*, *17*(19), 4222–4232. https://doi.org/10.1111/j.1365-294X.2008.03905.x

Amend, A. S., Oliver, T. A., Amaral-Zettler, L. A., Boetius, A., Fuhrman, J. A., Horner-Devine, M. C., Huse, S. M., Welch, D. B. M., Martiny, A. C., Ramette, A., Zinger, L., Sogin, M. L., & Martiny, J. B. H. (2012). Macroecological patterns of marine bacteria on a global scale. *Journal of Biogeography*, *40*(4), 800–811. https://doi.org/10.1111/jbi.12034

Anderson, M. J., & Walsh, D. C. I. (2013). PERMANOVA, ANOSIM, and the Mantel test face heterogeneous dispersions: What null hypothesis are you testing? *Ecological Monographs*, *83*(4), 557–574. https://doi.org/10.1890/12-2010.1

Apprill, A. (2017). Marine animal microbiomes: Toward understanding host–microbiome interactions in a changing ocean. *Frontiers in Marine Science*, *4*, 222. https://doi.org/10.3389/fmars.2017.00222

Apprill, A., McNally, S., Parsons, R., & Weber, L. (2015). Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquatic Microbial Ecology*, *75*(2), 129–137. https://doi.org/10.3354/ame01753

Belanger, C. L., Jablonski, D., Roy, K., Berke, S. K., Krug, A. Z., & Valentine, J. W. (2012). Global environmental predictors of benthic marine biogeographic structure. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(35), 14046–14051. https://doi.org/10.1073/pnas.1212381109

Bengtsson, M. M., Bühler, A., Brauer, A., Dahlke, S., Schubert, H., & Blindow, I. (2017). Eelgrass leaf surface microbiomes are locally variable and highly correlated with epibiotic eukaryotes. *Frontiers in Microbiology*, *8*, 1312. https://doi.org/10.3389/fmicb.2017.01312

Bivand, R., & Piras, G. (2015). Comparing implementations of estimation methods for spatial econometrics. *Journal of Statistical Software*, *63*(18), 1–36. https://doi.org/10.18637/jss.v063.i18

Callahan, B. J., McMurdie, P. J., & Holmes, S. P. (2017). Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME Journal*, *11*(12), 2639–2643. https://doi.org/10.1038/ismej.2017.119

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference

from Illumina amplicon data. *Nature Methods*, *13*, 581–583. https://doi.org/10.1038/nmeth.3869

Choudoir, M. J., Barberán, A., Menninger, H. L., Dunn, R. R., & Fierer, N. (2017). Variation in range size and dispersal capabilities of microbial taxa. *Ecology*, *99*(2), 322–334. https://doi.org/10.1002/ecy.2094

Chu, H., Fierer, N., Lauber, C. L., Caporaso, J. G., Knight, R., & Grogan, P. (2010). Soil bacterial diversity in the Arctic is not fundamentally different from that found in other biomes. *Environmental Microbiology*, *12*(11), 2998–3006. https://doi.org/10.1111/j.1462-2920.2010.02277.x

Corby-Harris, V., Pontaroli, A. C., Shimkets, L. J., Bennetzen, J. L., Habel, K. E., & Promislow, D. E. L. (2007). Geographical distribution and diversity of bacteria associated with natural populations of Drosophila melanogaster. *Applied and Environmental Microbiology*, *73*(11), 3470–3479. https://doi.org/10.1128/AEM.02120-06

Dahlhoff, E. P., & Menge, B. A. (1996). Influence of phytoplankton concentration and wave exposure on the ecophysiology of *Mytilus californianus*. *Marine Ecology Progress Series*, *144*, 97–107. https://doi.org/10.3354/meps144097

Dikongué, E., & Ségurel, L. (2017). Latitude as a co-driver of human gut microbial diversity? *BioEssays*, *39*(3), 1–6. https://doi.org/10.1002/bies.201600145

Dubé, C. E., Ky, C.-L., & Planes, S. (2019). Microbiome of the black-lipped pearl oyster *Pinctada margaritifera*, a multi-tissue description with functional profiling. *Frontiers in Microbiology*, *10*, 1548. https://doi.org/10.3389/fmicb.2019.01548

Faith, D. P. (1992). Conservation evaluation and phylogenetic diversity. *Biological Conservation*, *61*, 1–10. https://doi.org/10.1016/0006-3207(92)91201-3

Fierer, N., & Jackson, R. B. (2006). The diversity and biogeography of soil bacterial communities. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(3), 626–631. https://doi.org/10.1073/pnas.0507535103

Fuhrman, J. A., Steele, J. A., Hewson, I., Schwalbach, M. S., Brown, M. V., Green, J. L., & Brown, J. H. (2008). A latitudinal diversity gradient in planktonic marine bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(22), 7774–7778. https://doi.org/10.1073/pnas.0803070105

Hawkins, B. A. (2001). Ecology's oldest pattern? *Trends in Ecology & Evolution*, *16*(8), 470. https://doi.org/10.1016/S0169-5347(01)02197-8

Hawkins, B. A., Diniz-Filho, J. A. F., Jaramillo, C. A., & Soeller, S. A. (2006). Post-Eocene climate change, niche conservatism, and the latitudinal diversity gradient of New World birds. *Journal of Biogeography*, *33*(5), 770–780. https://doi.org/10.1111/j.1365-2699.2006.01452.x

Hawkins, B. A., McCain, C. M., Davies, T. J., Buckley, L. B., Anacker, B. L., Cornell, H. V., Damschen, E. I., Grytnes, J.-A., Harrison, S., Holt, R. D., Kraft, N. J. B., & Stephens, P. R. (2012). Different evolutionary histories underlie congruent species richness gradients of birds and mammals. *Journal of Biogeography*, *39*(5), 825–841. https://doi.org/10.1111/j.1365-2699.2011.02655.x

Helmuth, B., Harley, C. D. G., Halpin, P. M., O'Donnell, M., Hofmann, G. E., & Blanchette, C. A. (2002). Climate change and latitudinal patterns of intertidal thermal stress. *Science*, *298*(5595), 1015–1017. https://doi.org/10.1126/science.1076814

Hillebrand, H. (2004a). On the generality of the latitudinal diversity gradient. *The American Naturalist*, *163*(2), 192–211. https://doi.org/10.1086/381004

Hillebrand, H. (2004b). Strength, slope and variability of marine latitudinal gradients. *Marine Ecology Progress Series*, *273*(1992), 251–267. https://doi.org/10.3354/meps273251

Ibarbalz, F. M., Henry, N., Brandão, M. C., Martini, S., Busseni, G., Byrne, H., Coelho, L. P., Endo, H., Gasol, J. M., Gregory, A. C., Mahé, F., Rigonato, J., Royo-Llonch, M., Salazar, G., Sanz-Sáez, I., Scalco, E., Soviadan, D., Zayed, A. A., Zingone, A., … Wincker, P. (2019). Global trends in marine plankton diversity across kingdoms of life. *Cell*, *179*, 1084–1097. https://doi.org/10.1016/j.cell.2019.10.008

Jablonski, D., Huang, S., Roy, K., & Valentine, J. W. (2017). Shaping the latitudinal diversity gradient: New perspectives from a synthesis of paleobiology and biogeography. *The American Naturalist*, *189*(1), 1–12. https://doi.org/10.1086/689739

Kembel, S. W., Cowan, P. D., Helmus, M. R., Cornwell, W. K., Morlon, H., Ackerly, D. D., Blomberg, S. P., & Webb, C. O. (2010). Picante: R tools for integrating phylogenies and ecology. *Bioinformatics*, *26*(11), 1463–1464. https://doi.org/10.1093/bioinformatics/btq166

Krug, A. Z., Jablonski, D., & Valentine, J. W. (2007). Contrarian clade confirms the ubiquity of spatial origination patterns in the production of latitudinal diversity gradients. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(46), 18129–18134. https://doi.org/10.1073/pnas.0709202104

Ladau, J., Sharpton, T. J., Finucane, M. M., Jospin, G., Kembel, S. W., O'Dwyer, J., Koeppel, A. F., Green, J. L., & Pollard, K. S. (2013). Global marine bacterial diversity peaks at high latitudes in winter. *The ISME Journal*, *7*, 1669–1677. https://doi.org/10.1038/ismej.2013.37

Lozupone, C., & Knight, R. (2005). UniFrac: A new phylogenetic method for comparing microbial communities. *Applied and Environmental Microbiology*, *71*(12), 8228–8235. https://doi.org/10.1128/AEM.71.12.8228

Magurran, A. E. (1988). *Ecological diversity and its measurement*. Princeton University Press.

McLaren, M. R. (2020). *Silva SSU taxonomic training data formatted for DADA2 (Silva version 138)*. https://doi.org/10.5281/ZENODO.3731176

McMurdie, P. J., & Holmes, S. (2013). Phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One*, *8*(4), e61217. https://doi.org/10.1371/journal.pone.0061217

Milici, M., Tomasch, J., Wos-Oxley, M. L., Wang, H., Jáuregui, R., Camarinha-Silva, A., Deng, Z.-L., Plumeier, I., Giebel, H.-A., Wurst, M., Pieper, D. H., Simon, M., & Wagner-Döbler, I. (2016). Low diversity of planktonic bacteria in the tropical ocean. *Scientific Reports*, *6*(1), 19054. https://doi.org/10.1038/srep19054

Mittelbach, G. G., Schemske, D. W., Cornell, H. V., Allen, A. P., Brown, J. M., Bush, M. B., Harrison, S. P., Hurlbert, A. H., Knowlton, N., Lessios, H. A., McCain, C. M., McCune, A. R., McDade, L. A., McPeek, M. A., Near, T. J., Price, T. D., Ricklefs, R. E., Roy, K., Sax, D. F., … Turelli, M. (2007). Evolution and the latitudinal diversity gradient: Speciation, extinction and biogeography. *Ecology Letters*, *10*(4), 315–331. https://doi.org/10.1111/j.1461-0248.2007.01020.x

Morelan, I. A., Gaulke, C. A., Sharpton, T. J., Vega Thurber, R., & Denver, D. R. (2019). Microbiome variation in an intertidal sea anemone across latitudes and symbiotic states. *Frontiers in Marine Science*, *6*, 7. https://doi.org/10.3389/fmars.2019.00007

Morton, J. T., Toran, L., Edlund, A., Metcalf, J. L., Lauber, C., & Knight, R. (2017). Uncovering the horseshoe effect in microbial analyses. *Msystems*, *2*(1), e00166–e216. https://doi.org/10.1128/mSystems.00166-16

Moss, J. A., Henriksson, N. L., Pakulski, J. D., Snyder, R. A., & Jeffrey, W. H. (2020). Oceanic microplankton do not adhere to the latitudinal diversity gradient. *Microbial Ecology*, *79*(2), 511–515. https://doi.org/10.1007/s00248-019-01413-8

Neave, M. J., Michell, C. T., Apprill, A., & Voolstra, C. R. (2017). Endozoicomonas genomes reveal functional adaptation and plasticity in bacterial strains symbiotically associated with diverse marine hosts. *Scientific Reports*, *7*, 40579. https://doi.org/10.1038/srep40579

Neu, A. T., Allen, E. E., & Roy, K. (2019). Diversity and composition of intertidal gastropod microbiomes across a major marine biogeographic boundary. *Environmental Microbiology Reports*, *11*(3), 434–447. https://doi.org/10.1111/1758-2229.12743

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlin, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., & Solymos, P. (2019). *vegan: Community ecology package*. R package version 2.5-6. https://CRAN.R-project.org/package=vegan

Parada, A. E., Needham, D. M., & Fuhrman, J. A. (2016). Every base matters: Assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environmental Microbiology*, *18*(5), 1403–1414. https://doi.org/10.1111/1462-2920.13023

Paradis, E., & Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, *35*(3), 526–528. https://doi.org/10.1093/bioinformatics/bty63

Pfister, C. A., Gilbert, J. A., & Gibbons, S. M. (2014). The role of macrobiota in structuring microbial communities along rocky shores. *PeerJ*, *2*, e631. https://doi.org/10.7717/peerj.631

Pyron, R. A., & Burbrink, F. T. (2009). Can the tropical conservatism hypothesis explain temperate species richness patterns? An inverse latitudinal biodiversity gradient in the New World snake tribe Lampropeltini. *Global Ecology and Biogeography*, *18*(4), 406–415. https://doi.org/10.1111/j.1466-8238.2009.00462.x

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research*, *41*, D590–D596. https://doi.org/10.1093/nar/gks1219

Quensen, J. (2018). *ggordiplots: Make ggplot versions of Vegan's Ordiplots*. Retrieved from http://github.com/jfq3/ggordiplots

R Core Team. (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Raes, E. J., Bodrossy, L., van de Kamp, J., Bissett, A., & Waite, A. M. (2017). Marine bacterial richness increases towards higher latitudes in the eastern Indian Ocean. *Limnology and Oceanography Letters*, *3*(1), 10–19. https://doi.org/10.1002/lol2.10058

Ray, K. J., Cotter, S. Y., Arzika, A. M., Kim, J., Boubacar, N., Zhou, Z., Zhong, L., Porco, T. C., Keenan, J. D., Lietman, T. M., & Doan, T. (2019). High-throughput sequencing of pooled samples to determine community-level microbiome diversity. *Annals of Epidemiology*, *39*, 63–68. https://doi.org/10.1016/j.annepidem.2019.09.002

Rolland, J., Condamine, F. L., Jiguet, F., & Morlon, H. (2014). Faster speciation and reduced extinction in the tropics contribute to the mammalian latitudinal diversity gradient. *PLoS Biology*, *12*(1), e1001775. https://doi.org/10.1371/journal.pbio.1001775

Roy, K., Jablonski, D., & Valentine, J. W. (1994). Eastern Pacific molluscan provinces and latitudinal diversity gradient: No evidence for "Rapoport's rule". *Proceedings of the National Academy of Sciences of the United States of America*, *91*(19), 8871–8874. https://doi.org/10.1073/pnas.91.19.8871

Roy, K., Jablonski, D., & Valentine, J. W. (1996). Higher taxa in biodiversity studies: patterns from eastern Pacific marine molluscs. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *351*, 1605–1613. https://doi.org/10.1098/rstb.1996.0144

Roy, K., Jablonski, D., & Valentine, J. W. (2000). Dissecting latitudinal diversity gradients: Functional groups and clades of marine bivalves. *Proceedings of the Royal Society B: Biological Sciences*, *267*(1440), 293–299. https://doi.org/10.1098/rspb.2000.0999

Roy, K., Jablonski, D., Valentine, J. W., & Rosenberg, G. (1998). Marine latitudinal diversity gradients: Tests of causal hypotheses. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(7), 3699–3702. https://doi.org/10.1073/pnas.95.7.3699

Roy, K., & Witman, J. D. (2009). Spatial patterns of species diversity in the shallow marine invertebrates: Patterns, processes, and prospects. In J. D. Witman & K. Roy (Eds.), *Marine macroecology* (pp. 101–121). The University of Chicago Press.

Santelices, B., & Marquet, P. A. (1998). Seaweeds, latitudinal diversity patterns, and Rapoport's Rule. *Diversity and Distributions*, *4*(2), 71–75. https://doi.org/10.1046/j.1472-4642.1998.00005.x

Schemske, D. W., & Mittelbach, G. G. (2017). "Latitudinal gradients in species diversity": Reflections on Pianka's 1966 article and a look forward. *The American Naturalist*, *189*(6), 599–603. https://doi.org/10.1086/691719

Schill, W. B., Iwanowicz, D., & Adams, C. (2017). Endozoicomonas dominates the gill and intestinal content microbiomes of *Mytilus edulis* from Barnegat Bay, New Jersey. *Journal of Shellfish Research*, *36*(2), 391–401. https://doi.org/10.2983/035.036.0212

Schumm, M., Edie, S. M., Collins, K. S., Gómez-Bahamón, V., Supriya, K., White, A. E., Price, T. D., & Jablonski, D. (2019). Common latitudinal gradients in functional richness and functional evenness across marine and terrestrial systems. *Proceedings of the Royal Society B: Biological Sciences*, *286*, 20190745. https://doi.org/10.1098/rspb.2019.0745

Shade, A., & Gilbert, J. A. (2015). Temporal patterns of rarity provide a more complete view of microbial diversity. *Trends in Microbiology*, *23*(6), 335–340. https://doi.org/10.1016/j.tim.2015.01.007

Sherrill-Mix, S., McCormick, K., Lauder, A., Bailey, A., Zimmerman, L., Li, Y., Django, J.-B., Bertolani, P., Colin, C., Hart, J. A., Hart, T. B., Georgiev, A. V., Sanz, C. M., Morgan, D. B., Atencia, R., Cox, D., Muller, M. N., Sommer, V., Piel, A. K., … Bushman, F. D. (2018). Allometry and ecology of the bilaterian gut microbiome. *MBio*, *9*(2), e00319–e418. https://doi.org/10.1128/mBio.00319-18

Sogin, M. L., Morrison, H. G., Huber, J. A., Welch, D. M., Huse, S. M., Neal, P. R., Arrieta, J. M., & Herndl, G. J. (2006). Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proceedings of the National Academy of Sciences of the United States of America*, *103*(32), 12115–12120. https://doi.org/10.1073/pnas.0605127103

Song, H., Huang, S., Jia, E., Dai, X. U., Wignall, P. B., & Dunhill, A. M. (2020). Flat latitudinal diversity gradient caused by the Permian—Triassic mass extinction. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(30), 17578–17583. https://doi.org/10.1073/pnas.1918953117

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., Zech Xu, Z., Jiang, L., … Knight, R. (2017). A communal catalogue reveals Earth's multiscale microbial diversity. *Nature*, *551*, 457–463. https://doi.org/10.1038/nature24621

Wahl, M., Goecke, F., Labes, A., Dobretsov, S., & Weinberger, F. (2012). The second skin: Ecological role of epibiotic biofilms on marine organisms. *Frontiers in Microbiology*, *3*, 292. https://doi.org/10.3389/fmicb.2012.00292

Weiss, S., Xu, Z. Z., Peddada, S., Amir, A., Bittinger, K., Gonzalez, A., Lozupone, C., Zaneveld, J. R., Vázquez-Baeza, Y., Birmingham, A., Hyde, E. R., & Knight, R. (2017). Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome*, *5*(1), 27. https://doi.org/10.1186/s40168-017-0237-y

Wiegand, S., Jogler, M., & Jogler, C. (2018). On the maverick Planctomycetes. *FEMS Microbiology Reviews*, *42*(6), 739–760. https://doi.org/10.1093/femsre/fuy029

Witman, J. D., Etter, R. J., & Smith, F. (2004). The relationship between regional and local species diversity in marine benthic communities: A global perspective. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(44), 15664–15669. https://doi.org/10.1073/pnas.0404300101

Woodhams, D. C., Bletz, M. C., Becker, C. G., Bender, H. A., Buitrago-Rosas, D., Diebboll, H., Huynh, R., Kearns, P. J., Kueneman, J., Kurosawa, E., LaBumbard, B. C., Lyons, C., McNally, K., Schliep, K., Shankar, N., Tokash-Peters, A. G., Vences, M., & Whetstone, R. (2020). Host-associated microbiomes are predicted by immune system complexity and climate. *Genome Biology*, *21*, 23. https://doi.org/10.1186/s13059-020-01955-y

Zielinski, F. U., Pernthaler, A., Duperron, S., Raggi, L., Giere, O., Borowski, C., & Dubilier, N. (2009). Widespread occurrence of an intranuclear bacterial parasite in vent and seep bathymodiolin

mussels. *Environmental Microbiology, 11*(5), 1150–1167. https://doi.org/10.1111/j.1462-2920.2008.01847.x

BIOSKETCH

Alexander T. Neu is a PhD candidate in the Section of Ecology, Behavior and Evolution at the University of California San Diego. His research focuses on the microbial communities associated with intertidal molluscs, with specific interest in the evolutionary and biogeographical processes that shape the composition and function of these communities.

Eric E. Allen is a Professor in the Section of Molecular Biology and Marine Biology Research Division at the University of California San Diego. His research focuses on the genomic biology of marine animal microbiomes and environmental microbial communities.

Kaustuv Roy is a Professor in the Section of Ecology, Behavior and Evolution at the University of California San Diego. He is interested in better understanding the physical and biotic processes that determine large-scale biodiversity patterns and the generality of such processes across different domains of life. He also works on understanding the responses of species and communities to anthropogenic global change.

Author contributions: ATN, EEA and KR conceived the ideas; ATN conducted the fieldwork and collected the data with additional material from EEA and KR; ATN, EEA and KR analysed the data; and ATN led the writing with assistance from EEA and KR.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Neu, A. T., Allen, E. E., & Roy, K. (2021). Do host-associated microbes show a contrarian latitudinal diversity gradient? Insights from *Mytilus californianus*, an intertidal foundation host. *Journal of Biogeography*, 00, 1–14. https://doi.org/10.1111/jbi.14243
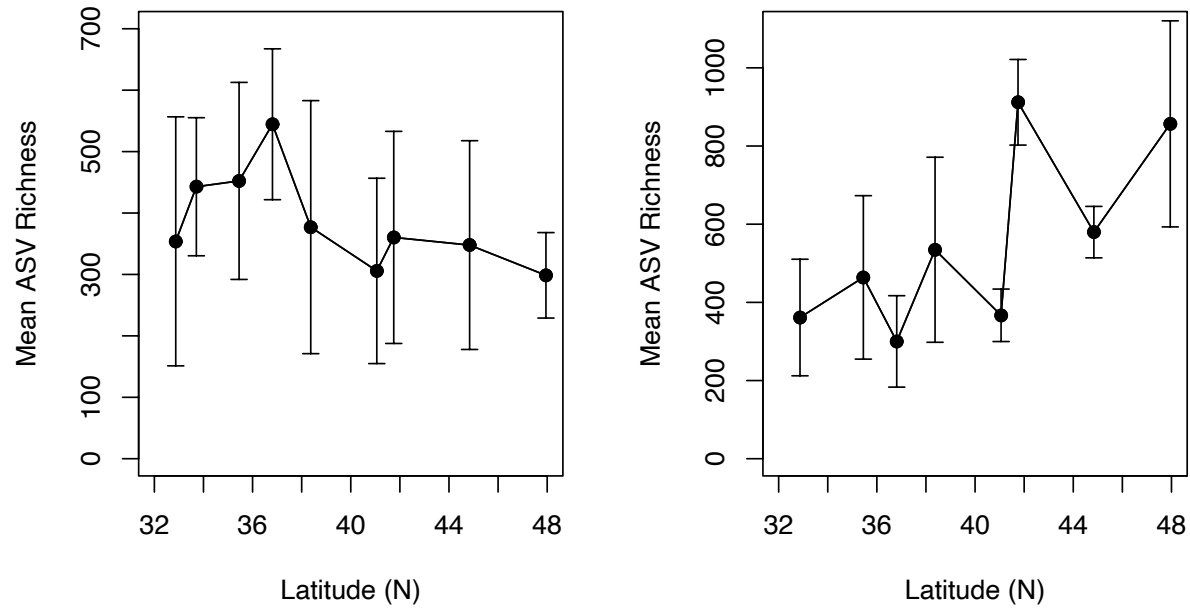
Supplement 1

Supplemental Figure S1. Mean ASV richness of gill (left) and shell-surface (right) microbiomes across latitude, including only those samples collected in 2017. Linear regression results for gill ($R^2 = 0.03$, $p = 0.05$) and shell-surface ($R^2 = 0.344$, $p < 0.001$) are qualitatively similar to those determined using the entire available dataset.
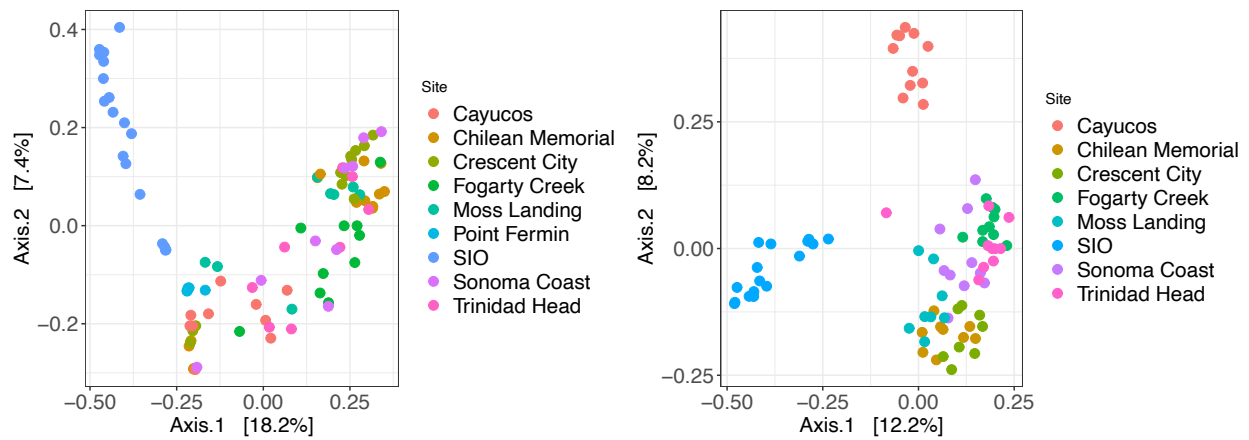
Supplemental Figure S2. PCoA plot of Bray-Curtis Dissimilarity in gill (left) and shell-surface (right) microbiomes, colored by collection site and including only samples collected in 2017. PERMANOVA results – gill: *pseudo-F = 4.493, R² = 0.2924, p < 0.001, betadisper p = 0.731*; shell-surface: *pseudo-F = 5.9397, R² = 0.3597, p < 0.001, betadisper p < 0.001*.

Supplement 2
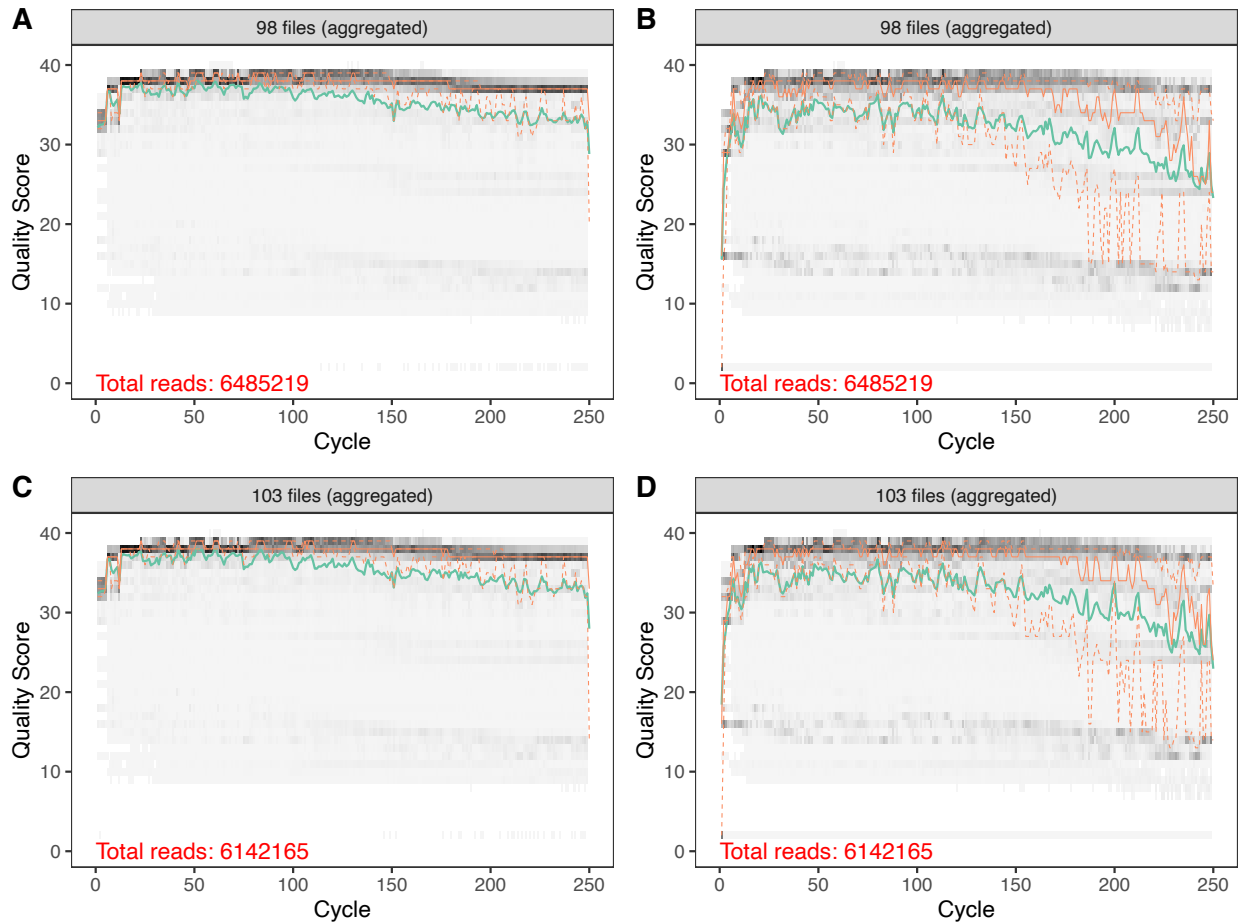
Supplemental Figure S1. Aggregate per-base quality scores of the raw forward (A,C) and reverse (B,D) Illumina MiSeq reads from the gill (A,B) and shell-surface (C,D) microbiomes. The solid green line donates the mean quality score of each base, while the solid orange line represents the median and the dashed orange lines represent the 25th and 75th quantiles, respectively.
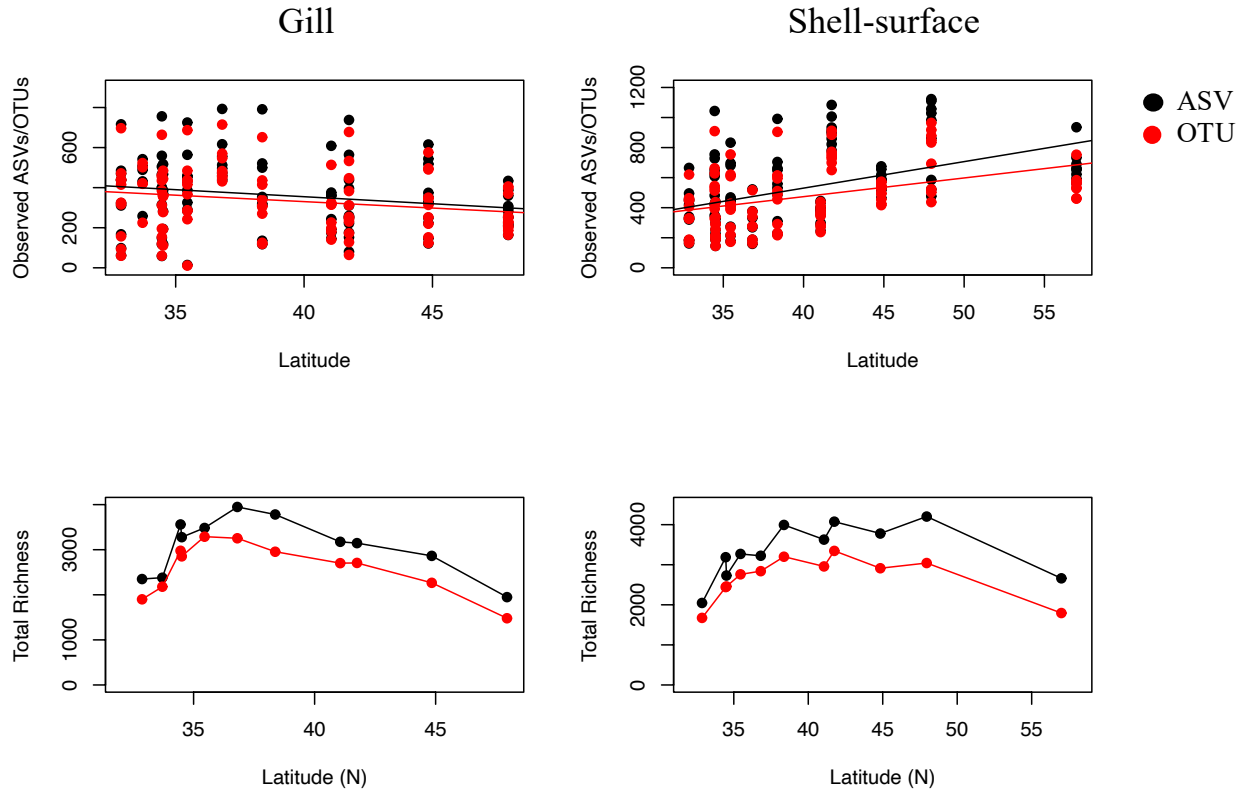
Supplemental Figure S2. Comparisons of average alpha diversity (top) and gamma diversity (bottom) of gill (left) and shell-surface (right) microbiomes when using ASVs (black) and 97% OTUs (red). All ASV and OTU datasets are highly correlated (*Spearman rho > 0.92, p < 0.001*).
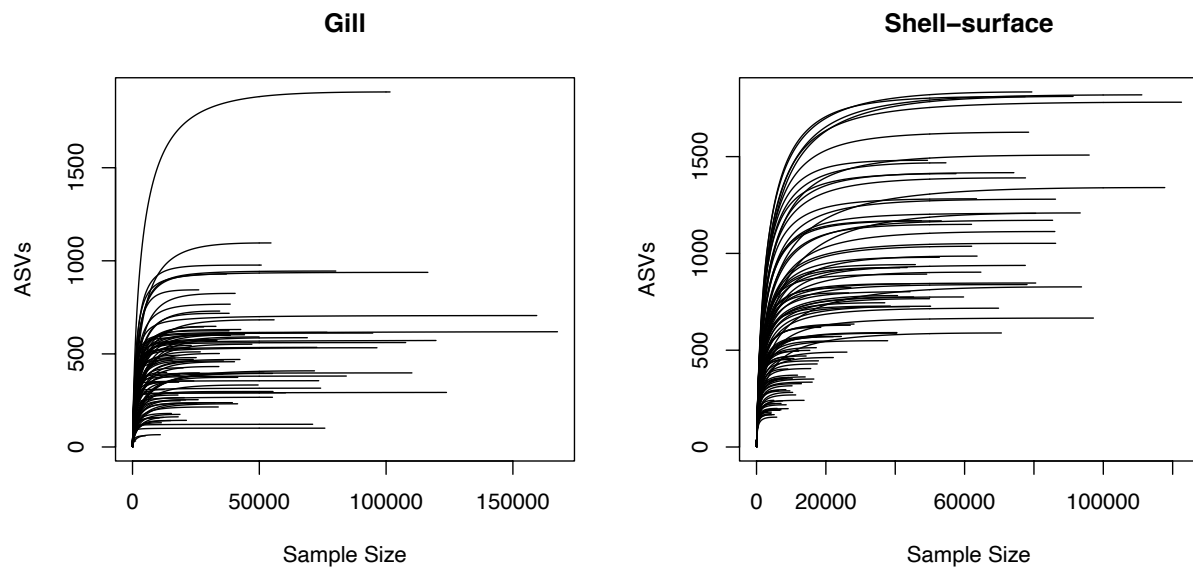
Supplemental Figure S3. Rarefaction curves of gill (left) and shell-surface (right) microbiomes.
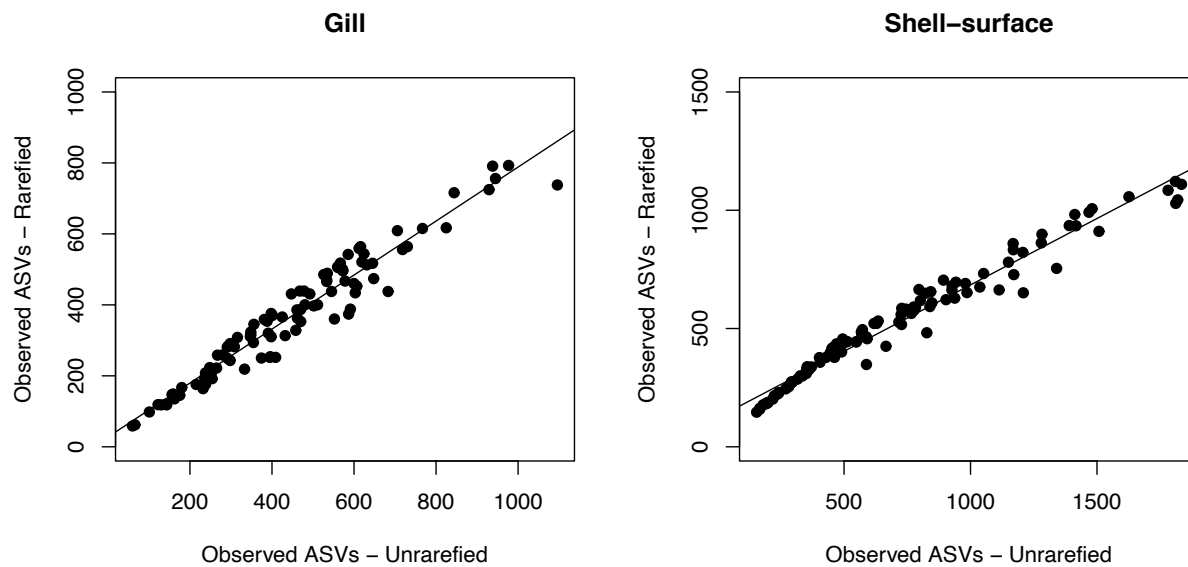
Supplemental Figure S4. Plots of the number of observed ASVs in unrarefied data vs. data rarefied to 5,000 sequences per samples in gill microbiomes (left) and 4,000 sequences per sample in shell-surface microbiomes (right). Linear regression results: gill – $R^2 = 0.938$, $p < 0.001$, shell-surface – $R^2 = 0.965$, $p < 0.001$.
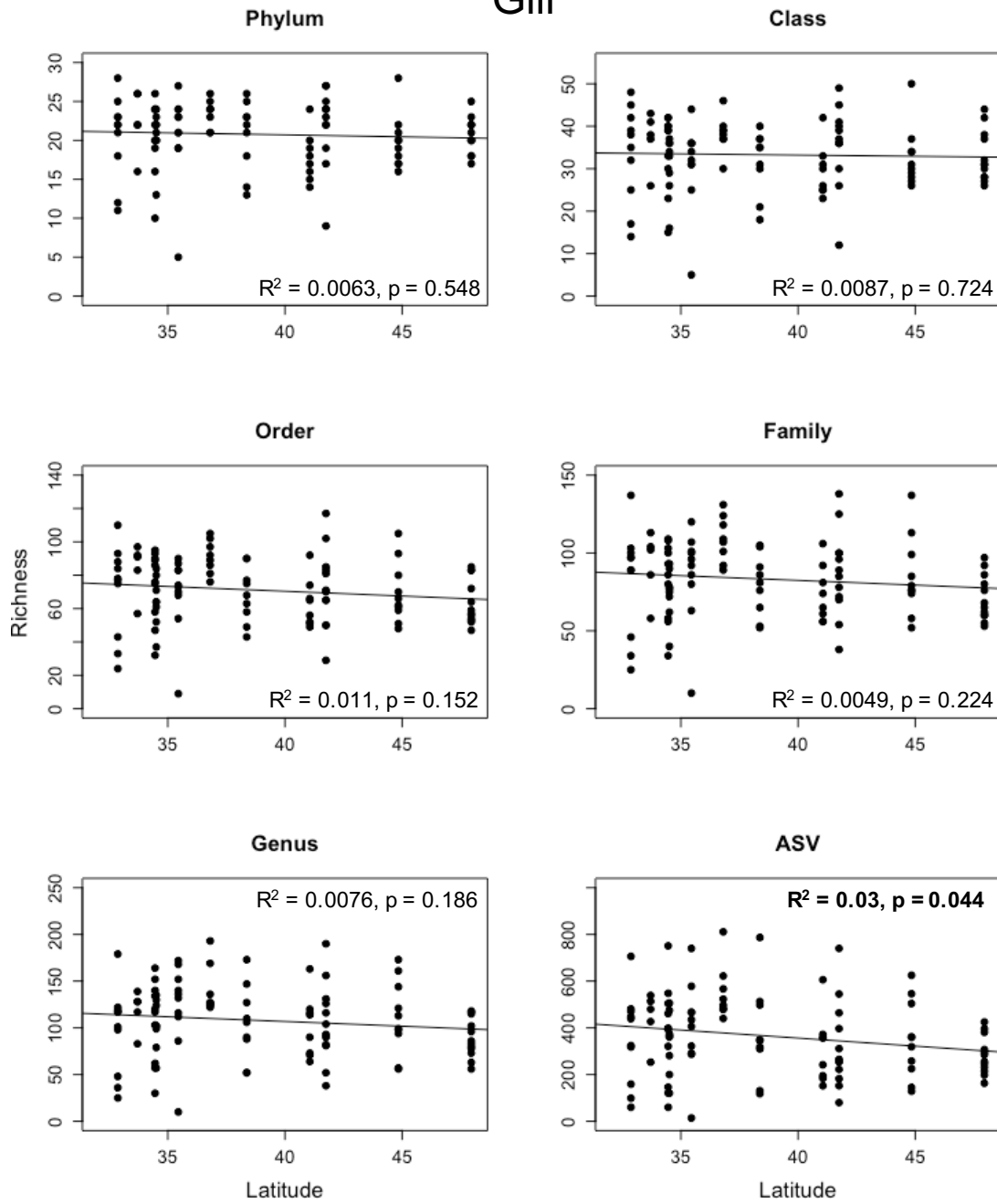
Supplemental Figure S5. Plots of richness across taxonomic levels versus latitude in the gill micro-environment. Results of linear regression are presented in the bottom right (phylum, class, family, order) or top right (genus, ASV) corners of the plots, bold font denotes significant correlation between richness and latitude ($p < 0.05$).
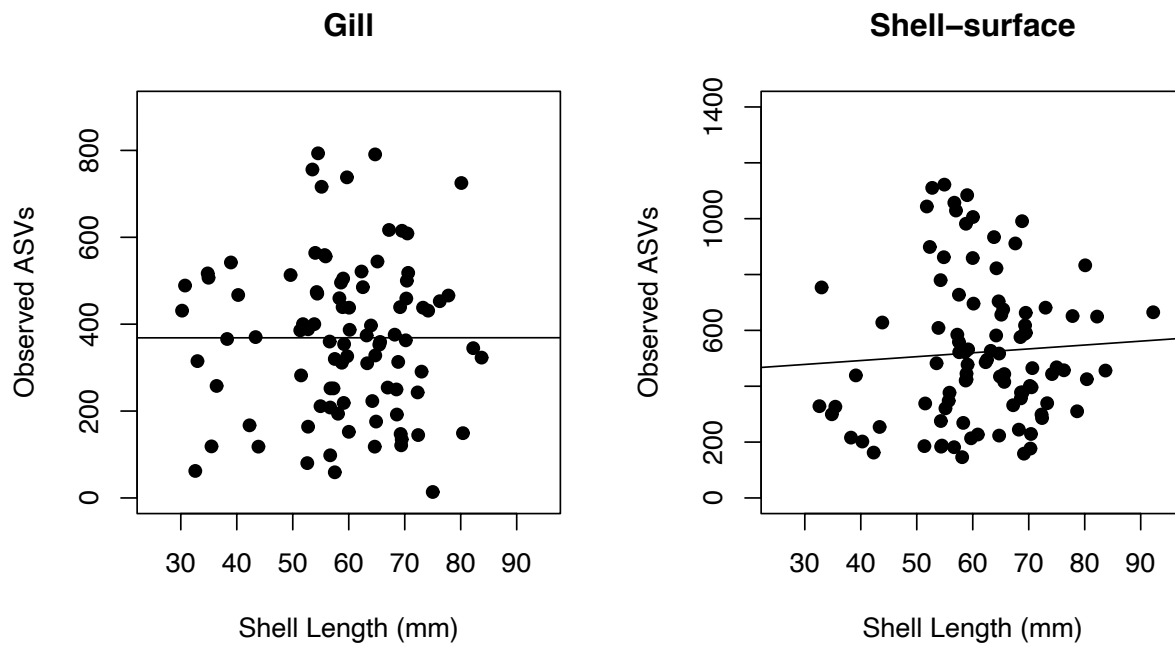
Supplemental Figure S6. Number of observed ASVs, after rarefying, versus shell length of the sampled individual in the gill (left) and shell-surface (right) microbiomes. Linear regression results: gill – $R^2 = -0.011$, $p = 0.998$, shell-surface – $R^2 = -0.0071$, $p = 0.558$.
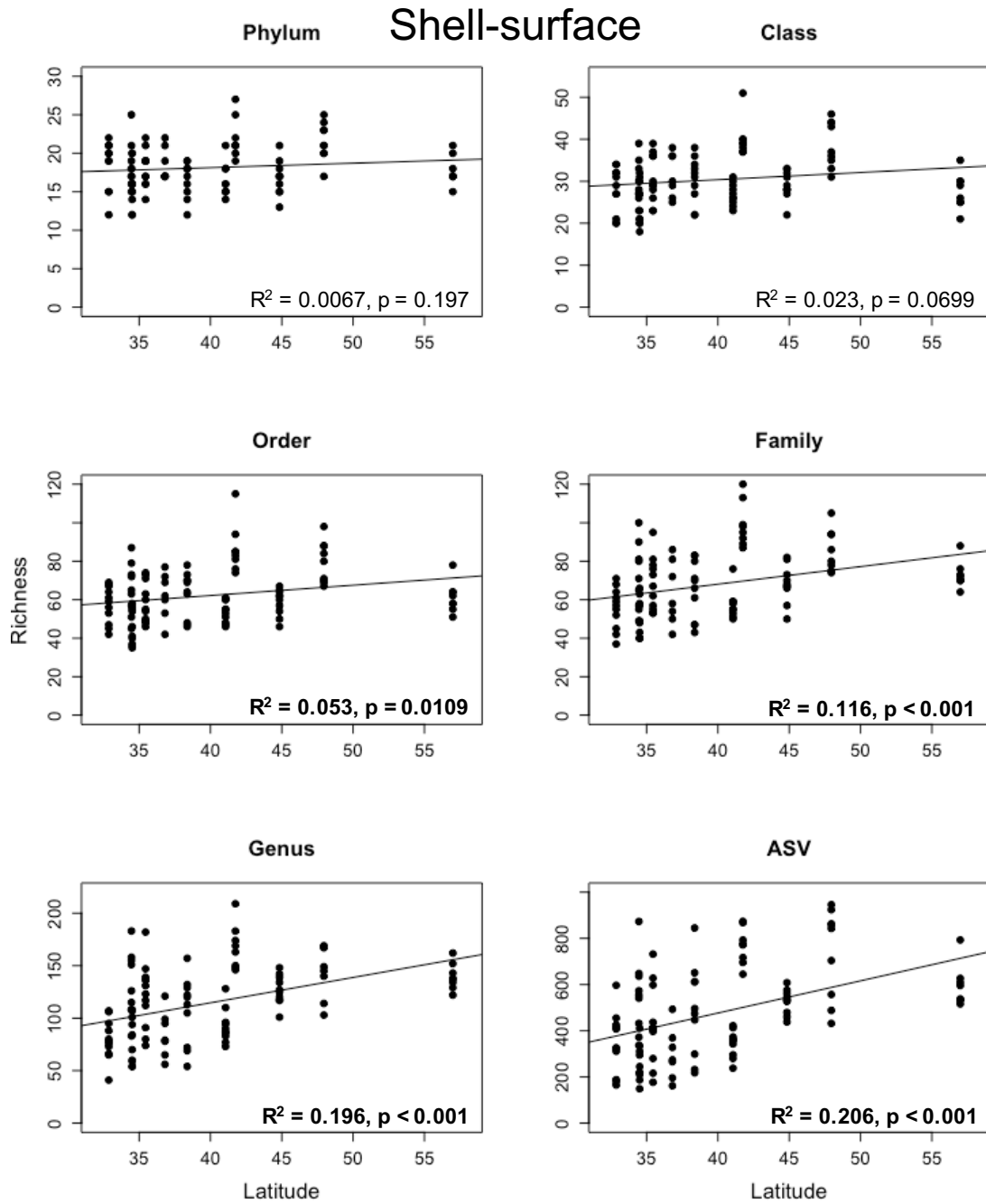
Supplemental Figure S7. Plots of richness across taxonomic levels versus latitude in the shell-surface micro-environment. Results of linear regression are presented in the bottom right corners of the plots, bold font denotes significant correlation between richness and latitude ($p < 0.05$).
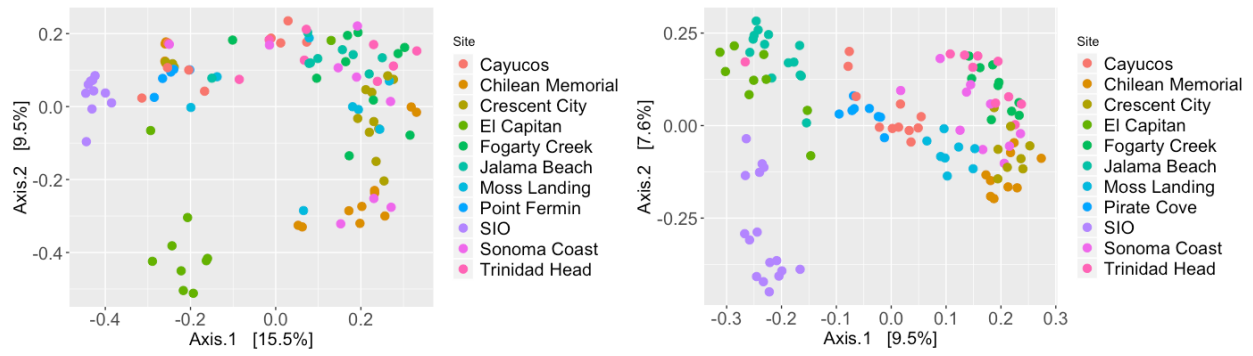
Supplemental Figure S8. PCoA plot of Bray-Curtis Dissimilarity in gill (left) and shell-surface (right) microbiomes, colored by collection site. PERMANOVA results – gill: *pseudo-F = 5.0185, R² = 0.3658, p < 0.001, betadisper p = 0.505*; shell-surface: *pseudo-F = 6.2092, R² = 0.4004, p < 0.001, betadisper p = 0.007.*
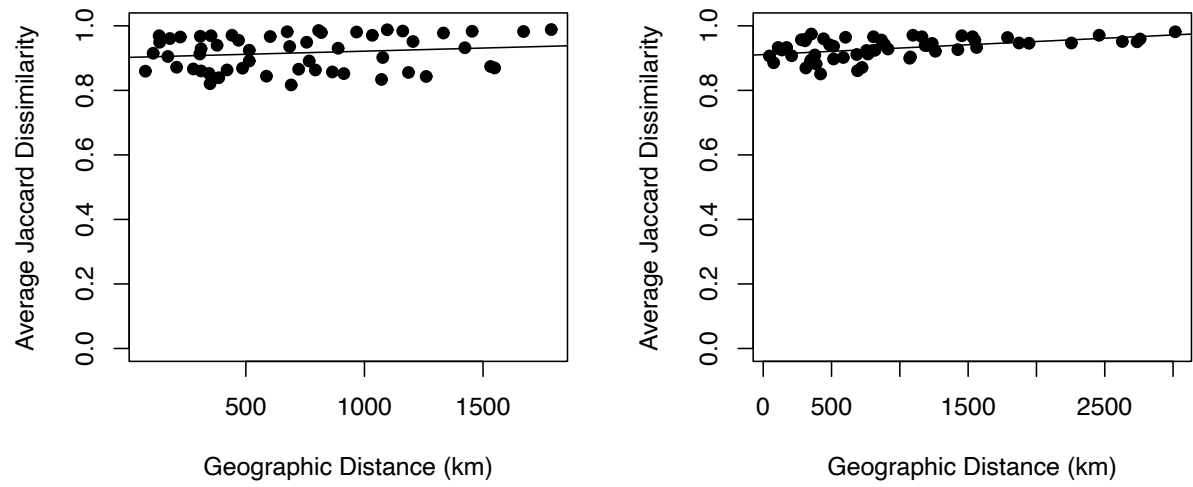
Supplemental Figure S9. Plots of Jaccard dissimilarity, averaged across all individuals at each site, by geographic distance between sites in the gill (left) and shell-surface (right) microbiomes. Linear regression results: gill – $R^2 = 0.00736$, $p = 0.243$, shell-surface – $R^2 = 0.219$, $p < 0.001$.

Supplemental Table S1. NCBI BLASTn results of ASVs assigned to the family Endozoicomonadaceae that comprised >10% relative abundance in at least one *M. californianus* gill sample.

ASV ID	Description	Source	% ident.	Accession
ASV17474	Candidatus Endonucleobacter bathymodioli	*Bathymodiolus brooksi*	94.09%	FM162182.1
ASV17543	Uncultured bacterium clone MeS	*Mytilus edulis*	99.60%	KU936077.1
ASV17549	Uncultured bacterium clone MeS	*Mytilus edulis*	98.41%	KU936077.1
ASV17557	Uncultured bacterium clone MeS	*Mytilus edulis*	99.20%	KU936077.1
ASV17558	Uncultured bacterium clone MeS	*Mytilus edulis*	98.80%	KU936077.1
ASV17570	Uncultured bacterium clone Mus-14	*Mytilus edulis*	99.60%	JN858942.1
ASV17575	Loripes lacteus gill symbiont	*Loripes lacteus*	99.60%	GQ853555.1

Supplemental Table S2. Results of Moran's I analysis of richness from each taxonomic level for gill and shell-surface communities.

Taxonomic level	Gill		Shell-surface	
	Moran's I	p-value	Moran's I	p-value
Phylum	-0.0323	0.366	-0.0681	<0.001
Class	-0.0288	0.316	-0.0757	<0.001
Order	-0.0239	0.650	-0.104	<0.001
Family	-0.0264	0.480	-0.0896	<0.001
Genus	-0.0169	0.775	-0.0887	<0.001

Supplemental Table S3. Results of the spatial autoregressive models for the influence of latitude on taxon richness across taxonomic levels in the shell-surface microbiome.

Taxonomic level	Z value	P value
Phylum	1.136	0.256
Class	1.371	0.170
Order	2.002	0.045
Family	2.739	0.006
Genus	3.601	<0.001

Supplemental Table S4. Results of PERMANOVA and betadisper analyses across taxonomic levels in the gill micro-environment using Bray-Curtis dissimilarity and site as a factor.

Taxonomic Level	Pseudo-F	R^2	P-value	Betadisper p-value
Phylum	2.657	0.226	<0.001	0.15
Class	2.812	0.236	<0.001	0.342
Order	3.268	0.264	<0.001	0.415
Family	2.522	0.217	<0.001	0.082
Genus	2.646	0.225	<0.001	0.075
ASV	4.763	0.344	<0.001	0.475

Supplemental Table S5. Results of PERMANOVA and betadisper analyses across taxonomic levels in the shell-surface micro-environment using Bray-Curtis dissimilarity and site as a factor.

Taxonomic Level	Pseudo-F	R^2	P-value	Betadisper p-value
Phylum	10.488	0.533	<0.001	0.215
Class	10.306	0.529	<0.001	0.903
Order	8.541	0.481	<0.001	0.623
Family	8.732	0.487	<0.001	0.482
Genus	7.635	0.453	<0.001	0.468
ASV	5.776	0.386	<0.001	0.008

Supplemental Table S6. Results of PERMANOVA and betadisper analyses of the focal microbial clades in the gill micro-environment using Bray-Curtis dissimilarity and site as a factor.

Clade	Pseudo-F	R^2	P-value	Betadisper p-value
Alphaproteobacteria	2.935	0.252	<0.001	0.368
Bacteroidetes	2.697	0.237	<0.001	0.037
Gammaproteobacteria	7.238	0.454	<0.001	0.05
Planctomycetota	1.642	0.159	<0.001	0.004

Supplemental Table S7. Results of PERMANOVA and betadisper analyses of the focal microbial clades in the shell-surface micro-environment using Bray-Curtis dissimilarity and site as a factor.

Clade	Pseudo-F	R^2	P-value	Betadisper p-value
Alphaproteobacteria	4.974	0.348	<0.001	0.003
Bacteroidetes	4.753	0.338	<0.001	0.001
Gammaproteobacteria	7.430	0.444	<0.001	0.001
Planctomycetota	3.747	0.287	<0.001	0.001

Supplemental Table S8. Results of PERMANOVA and betadisper analyses of ASVs in the gill micro-environment using various beta diversity metrics and site as a factor.

Diversity Metric	Pseudo-F	R²	P-value	Betadisper p-value
Jaccard	3.385	0.280	<0.001	0.517
Unweighted UniFrac	1.982	0.186	<0.001	0.003
Weighted UniFrac	4.067	0.319	<0.001	0.227

Supplemental Table S9. Results of PERMANOVA and betadisper analyses of ASVs in the shell-surface micro-environment using various beta diversity metrics and site as a factor.

Diversity Metric	Pseudo-F	R^2	P-value	Betadisper p-value
Jaccard	3.682	0.284	<0.001	0.011
Unweighted UniFrac	3.115	0.251	<0.001	0.001
Weighted UniFrac	5.893	0.388	<0.001	0.021

Acknowledgements

CHAPTER 3

Decade-scale stability and change in a marine bivalve microbiome

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Decade-scale stability and change in a marine bivalve microbiome

Alexander T. Neu[1] | Ian V. Hughes[1] | Eric E. Allen[2,3] | Kaustuv Roy[1]

[1]Section of Ecology, Behavior and Evolution, Division of Biological Sciences, University of California San Diego, La Jolla, CA, USA

[2]Section of Molecular Biology, Division of Biological Sciences, University of California San Diego, La Jolla, CA, USA

[3]Marine Biology Research Division, Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA, USA

Correspondence
Alexander T. Neu, Section of Ecology, Behavior and Evolution, University of California San Diego, 9500 Gilman Drive, MC 0116, La Jolla, CA 92093, USA.
Email: aneu@ucsd.edu

Funding information
American Museum of Natural History, Grant/Award Number: Lerner Gray Fund for Marine Research; Sigma Xi; Academic Senate, University of California, San Diego

Abstract
Predicting how populations and communities of organisms will respond to anthropogenic change is of paramount concern in ecology today. For communities of microorganisms, however, these predictions remain challenging, primarily due to data limitations. Information about long-term dynamics of host-associated microbial communities, in particular, is lacking. In this study, we use well-preserved and freshly collected samples of soft tissue from a marine bivalve host, *Donax gouldii*, at a single site to quantify the diversity and composition of its microbiome over a decadal timescale. Site-level measurements of temperature, salinity and chlorophyll *a* allowed us to test how the microbiome of this species responded to two natural experiments: a seasonal increase in temperature and a phytoplankton bloom. Our results show that ethanol-preserved tissue can provide high-resolution information about temporal trends in compositions of host-associated microbial communities. Specifically, we found that the richness of amplicon sequence variants (ASVs) associated with *D.gouldii* did not change significantly over time despite increases in water temperature (+1.6°C due to seasonal change) and chlorophyll *a* concentration (more than ninefold). The phylogenetic composition of the communities, on the other hand, varied significantly between all collection years, with only six ASVs persisting over our sampling period. Overall, these results suggest that the diversity of microbial taxa associated with *D.gouldii* has remained stable over time and in response to seasonal environmental change over the course of more than a decade, but such stability is underlain by substantial turnover in the composition of the microbiome.

KEYWORDS
Donax gouldii, environmental change, host-microbiome, microbial preservation

1 | INTRODUCTION

Predicting how species and communities will respond to anthropogenic global change is an ongoing challenge in ecology. An increase in temperature, for example, may cause shifts in geographical distributions, local or global extinctions, or adaptive evolution of marine species (Hellberg et al., 2001; Poloczanska et al., 2013, 2016). Other impacts such as coastal eutrophication are likely to affect the

duration and composition of phytoplankton blooms, which may affect food web structure (Suikkanen et al., 2013). However, at present our knowledge of biological responses to these environmental perturbations is based predominantly on studies of plants, animals and other eukaryotes. How prokaryotic taxa and communities are likely to respond to such changes remains poorly known (Apprill, 2017; Cavicchioli et al., 2019; Trevathan-Tackett et al., 2019). Understanding how host-associated microbial communities respond

to changes in the environment is particularly important because any alteration of the microbiome has the potential to affect host response to environmental stressors, given its role in regulating many aspects of host function and health, from disease protection to nutrient acquisition (Batstone & Dufour, 2016; Reshef et al., 2006; Vega Thurber et al., 2009). Short-term laboratory experiments have shown that elevated seawater temperature and increased salinity can significantly change the compositions of marine invertebrate microbiomes (Lesser et al., 2016; Minich et al., 2018; Röthig et al., 2016), while the effects of phytoplankton blooms on host-associated microbiomes have not, to our knowledge, been tested explicitly, but have been predicted to significantly impact bivalve microbiome composition (Carrier & Reitzel, 2017).

Short-term experiments, however, do not always provide a comprehensive picture of how natural populations respond to environmental change (Komatsu et al., 2019). To fully understand how microbial communities are impacted by changes in their environment, such experiments need to be supplemented with "natural experiments," which involve observing community responses to environmental perturbations in the wild. However, the use of such natural experiments remains limited, as they typically require long-term sampling of biological communities with paired environmental monitoring. Unfortunately, this means that in most cases we still lack sufficient long-term data to understand how different types of environmental change affect the diversity, composition and function of microbial communities. Exceptions include seawater communities in the Southern California Bight, which were monitored regularly throughout the course of a phytoplankton bloom to assess fine-scale community succession (Needham & Fuhrman, 2016), pelagic marine bacterial communities in the UK that displayed seasonal cycles in diversity, with peaks in the winter months (Gilbert et al., 2012), and forest soil microbiomes that increased in diversity after two decades of warming (Pold et al., 2015). Similar time-series are even rarer for host-associated microbial communities, but monitoring of human faecal microbiota over periods of 5–12 years has shown that individuals display remarkable species- and strain-level stability (Faith et al., 2013; Rajilić-Stojanović et al., 2013), while studies of wild mice (Maurice et al., 2015), fish (Minich et al., 2020) and corals (Sharp et al., 2017; van de Water et al., 2018) have shown that seasonal variability significantly impacts the composition of the internal microbial community in these hosts over the course of 1–2 years.

A potential solution to the problem of long-term sampling of host microbiomes is to use archival samples, such as those in museums or academic institutions. In most cases, such samples were collected for non-microbiome research and preserved in a variety of different ways. Despite the potential importance of these samples for microbial research, at present there is little consensus about whether such long-term preservation of host tissue allows unbiased quantification of the associated microbiome. Available studies about the efficacy of different sample preservation methods for microbiome analyses are primarily focused on the short term (<8 weeks) and provide a variety of recommendations depending on sample type, field conditions, etc. (Blekhman et al., 2016; Gray et al., 2013; Hale et al.,

2015; Song et al., 2016). Nearly all compare their results to samples frozen immediately at −80°C or in liquid nitrogen, which has been the presumed standard. While storage at −80°C with RNAlater has been shown to be effective for longer term (5 years) preservation of faecal samples (Tap et al., 2019), microbial communities sampled via skin swabs can shift dramatically after several months of storage, even at this low temperature (Klymiuk et al., 2016). On the other hand, human stool samples from 13 Type II diabetes patients that were dried and frozen at −20°C for 14 years yielded microbiome profiles that aligned with expectations derived from the American Gut Project (Kia et al., 2016). Similarly, toilet tissue samples from a single human subject, stored dry, produced useful microbiome data after 46 years (Jayasinghe et al., 2017). More importantly, very few studies have investigated whether host tissue preserved in ethanol, the most common preservative used for museum collections, could be used to investigate long-term variations in microbiomes of non-human hosts. The most extensive study on this topic, conducted using trematomid fish tissue collected between 1901 and 2006, showed that only ~10% of samples selected could be successfully used for microbiome analyses (Heindler et al., 2018). However, the authors suggest that many of those samples were probably first stored in formalin before transferring to ethanol (a common practice for museum samples in the past), which may explain the low yield, as formalin is known to cause multiple structural changes in DNA (Heindler et al., 2018). Overall, while there is some indication that samples preserved using a variety of methods may be useful for analysing long-term trends in host microbiomes, such analyses need to consider the potential for preservational biases.

In this study, we use a novel set of tests for preservational biases and samples of the marine bivalve *Donax gouldii* (bean clam, Figure 1a), stored in 95% ethanol at 4°C, to first evaluate the suitability of archival host tissue for long-term microbiome analyses. We then use the data to assess (a) temporal dynamics of the diversity and composition of the microbiome of this species over a decadal time span and (b) how two "natural experiments," a 1.6°C increase in temperature probably due to a seasonal effect, and a spring phytoplankton bloom, impacted the microbiome of this species.

Donax gouldii is a small (<25 mm) marine bivalve found on sandy beaches from central California, USA, to Baja California Sur, Mexico. This species is an important component of sandy beach communities and appears in "boom and bust" cycles, where populations can shift from very high densities (20,000–32,000 individuals/m²; Coe, 1953, 1955; Pohlo, 1967) in some years to vanishingly low numbers (<1 individual/m²) in others (Coe, 1955). These cycles have been of interest to ecologists for more than half a century, but the factor(s) driving them remain unknown. The bacterial community associated with *D.gouldii* was first characterized in 1966 in an attempt to identify potential pathogens involved in the population declines, but no specific candidates could be identified (Beeson & Johnson, 1967). *Donax gouldii* makes an excellent system for monitoring long-term changes in microbiomes because individuals of this host recruit to a narrow (2–5 m) swathe of the beach and, unlike some of its congeners, rarely move with the tide, making it amenable to repeated
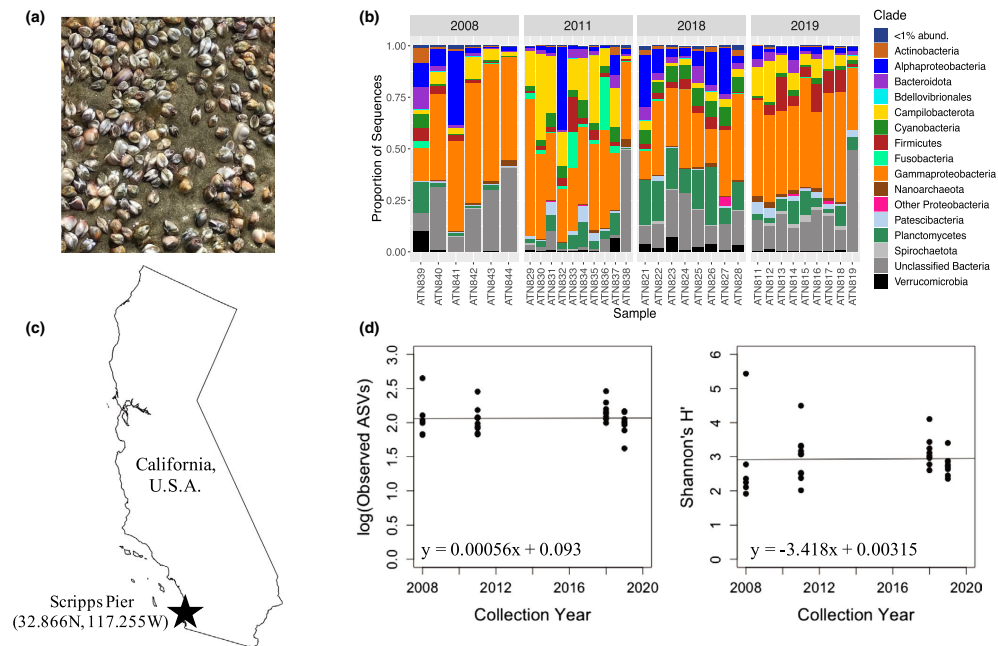
FIGURE 1 Levels of stability in the *Donax gouldii* microbiome. (a) *Donax gouldii* individuals during a "boom" cycle. Individuals are 15–20 mm in length on average. Photo credit: K. Roy. (b) Barplots of the relative abundances of microbial phyla, sorted by collection year. The phylum Proteobacteria is separated into classes for increased resolution. (c) Map of the study area. (d) Linear regression of the logarithm of observed ASV counts (left; R^2 = .00016, p = .944) and Shannon's index (right; R^2 = .00041, p = .911) versus time of collection. Kruskal–Wallis tests showed no significant differences between the years using either the logarithm of ASV counts (χ^2 = 5.828, p = .120) or Shannon's index (χ^2 = 6.015, p = .111). Removal of the high-diversity outlier from 2008 does not significantly impact the overall results

sampling at the same site over multiple years (Coe, 1955). The individuals of *D.gouldii* used here were collected at a single site in La Jolla, California, USA (Figure 1c) during four "boom" periods, 2008, 2011, 2018 and 2019, in conjunction with site-level measurements of multiple environmental variables. Unfortunately, no archival samples from the intervening years are currently available.

2 | MATERIALS AND METHODS

Individuals of *D. gouldii* were collected adjacent to Ellen Browning Scripps Memorial Pier at Scripps Institution of Oceanography, La Jolla, CA, USA (Figure 1c). Archival samples (March 2008 and May 2011 collections) were stored in 95% ethanol at 4°C in a single collecting jar per time point, while samples from March 2018 were stored individually and frozen at −20°C and February 2019 samples were processed fresh. We measured all samples using calipers and selected only individuals of the same size class (15–20 mm shell length) from each collection. Eight individuals were used from the 2008 sampling, nine each from 2011 and 2018 and 10 from 2019.

The entirety of the soft tissue from each animal was removed from the shell using sterilized forceps and processed for microbiome analyses. In 2019, we also collected sand and water samples to provide environmental context for our results. The top 3 cm of sand was collected from three sites within the *D.gouldii* zone using a sterile 50-ml tube, while two 600-ml samples of surface water were collected from shore in autoclaved 1-L bottles and filtered using 0.22-µm Sterivex filters (MilliporeSigma).

We extracted DNA from all *D.gouldii* individuals (n = 36) and environmental samples (n = 5) using the MoBio PowerSoil kit (MoBio Inc.; now part of Qiagen GmbH), following the manufacturer's instructions and including an additional 5-min incubation at 55°C after bead beating. Samples were randomized, extracted by hand and processed several wells apart from one another in order to reduce the impact of well-to-well contamination (Minich et al., 2019). We amplified the V4 region of the 16S ribosomal RNA gene using the two-step PCR (polymersase chain reaction) protocol described in Neu et al. (2019), which is based on the protocol developed by Eloe-Fadrosh et al. (2016). We purified the resulting PCR products using an Exo-SAP approach (1 U of Exonuclease I and 1U

of Shrimp Alkaline Phosphate; New England Biolabs), and heated at 37°C for 30 min followed by 85°C for 15 min. Additional purification was conducted using Ampure XP beads (Beckman Coulter), following the manufacturer's instructions. Amplicons were pooled in equal molarities and sent for sequencing on the MiSeq (Illumina, 2 × 250-bp output) at the DNA Technologies Core, University of California Davis.

All analyses of the resulting data were conducted using R version 3.5.2 (R Core Team, 2013). We merged, filtered, trimmed to 253 bp and denoised paired-end amplicon sequence data using the package DADA2 version 1.10.1 (Callahan et al., 2016). We assigned taxonomy to the resulting amplicon sequence variants (ASVs) using the assignTaxonomy and assignSpecies functions with the SILVA database version 138 (McLaren, 2020; Quast et al., 2013) at an 80% minimum value. All sequences matched to chloroplasts and mitochondria were removed. We merged *D.gouldii*-associated sequence counts and taxonomy to create an ASV table, which was rarefied to 8000 sequences per sample for all analyses. This left 33 *D.gouldii* samples in the final data set. We further merged the 2019 *D.gouldii* samples with sand and water samples, removing any temporal effects, and rarefied to 3000 sequences per sample in order to retain three sand samples and one water sample. Initial analyses of multiple rarefied tables at these sequence depths produced qualitatively similar results, so one representative table from each data set (set.seed =100) was used for the analyses presented here.

Before analysing the data for temporal and environmental trends, we used multiple different metrics to test for preservational biases that could potentially skew our results due to the different ages and storage conditions of these samples. We hypothesized that preservational effects would (a) decrease raw sequence counts, DNA yield and DNA concentration after 16S rRNA PCR in older samples due to DNA degradation. While each of these metrics can differ between samples due to causes other than preservation, systematic preservational bias predicts a specific pattern—a declining trend from the youngest to the oldest samples. (b) Increase homogeneity of samples preserved in the same containers (2008 and 2011 samples were stored as bulk samples in jars) due to cross-contamination compared to fresh samples, and (c) obliterate taxa unique to each specimen in samples preserved long term in ethanol. The last hypothesis is analogous to taphonomic bias in the palaeontological literature (Behrensmeyer et al., 2000) and relies on the assumption that taxa present in older samples, but absent from fresh material, represent real signals and cannot be preservational artefacts. In addition, it is also expected that the number of unique taxa should decrease with decreasing preservation probability (Foote & Raup, 1996), again in our case from the youngest to the oldest samples. It is worth noting that some of our metrics for evaluating preservational biases, by themselves, may have certain limitations. For example, the equimolar pooling of amplicons could obscure smaller differences in sequence counts resulting from preservation. However, large differences in raw sequence counts of individual host-associated microbiomes have been observed in studies using freshly collected samples even when equimolar pooling is used (e.g. Adam et al., 2018;

TABLE 1 Averages of environmental data collected from Ellen Browning Scripps Memorial Pier, La Jolla, CA, for the month of collection

Date	Average SST (°C)	Average salinity (psu)	Average [Chl a], (µg/L)
March 2008	15.9	33.53	1.75
May 2011	17.5	33.56	1.36
March 2018	15.2	33.53	4.00
February 2019	16.0	33.49	36.83

Garcia-Recinos et al., 2019; Neu et al., 2019), so such counts should be useful for identifying larger preservational biases. Similarly, total DNA yield could reflect the condition of host tissue rather than microbial preservation, although degradation of host tissues is likely to also affect recovery of the microbiome. Determining DNA concentration of a sample after 16S rRNA PCR partially avoids this problem but does not capture the relative host to microbial DNA concentration across samples. Because of this, our approach is to use the aggregate trend, rather than individual metrics, to test for preservational biases.

We evaluated these preservational hypotheses by (a) using linear regressions and Kruskal–Wallis tests of raw sequence counts, extracted DNA yield and DNA concentration after 16S rRNA PCR (as measured by Qubit 2.0; ThermoFisher Scientific) versus sampling time to test for significant temporal trends; (b) calculating the homogeneity of dispersions within each year sampled using betadisper in the package VEGAN version 2.5-6 (Oksanen et al., 2019); and (c) constructing a Venn diagram showing the number of ASVs unique to each year using the package VENNDIAGRAM version 1.6.20 (Chen, 2018).

We downloaded time-adjusted sea surface temperature (SST; Rasmussen et al., 2020) and daily raw salinity data from https://shorestations.ucsd.edu/data-sio/ and chlorophyll *a* concentration data from https://www.sccoos.org/data/autoss/timeline/?main=single&station=scripps_pier. We used the average value from the month of collection for further analyses in order to smooth out any short-term and measurement noise (Table 1; Figure S1). Using these data, we found that the environment was largely stable between collection times, with two notable exceptions. The first was a 1.6°C increase in temperature in the 2011 samples when compared to the 2008 samples. This increase was probably due to the month of collection (May vs late February and March), reflecting the natural seasonal variation at this site (Figure S1). Nevertheless, these samples experienced a substantially warmer condition compared to the other time points in this study and thus allow us to test the impact of increased temperature on the bivalve microbiome in a natural setting. The second was a more than nine-fold increase in chlorophyll *a* concentration in 2019, the result of a major phytoplankton bloom (Figure S1). These two events serve as natural experiments to test the impacts of changes in environmental variables on the diversity and composition of the *D.gouldii* microbiome. Note that while other environmental variables are also likely to have changed during our

study interval, reliable site-level measurements are only available for SST, chlorophyll *a* concentration and salinity. Thus, we only focus on those three variables here.

We hypothesized that the diversity of microbial communities would decrease with both increasing SST, as has been previously shown in short-term experiments with marine bivalves (Li et al., 2018), and increasing chlorophyll *a* concentration, as phytoplankton blooms have been shown to decrease bacterial diversity in the water column (Teeling et al., 2012; Wemheuer et al., 2014; Yang et al., 2015). To test these hypotheses, we calculated two different alpha diversity metrics—ASV richness and Shannon's index (*H'*)—using the package PHYLOSEQ version 1.26.1 (McMurdie & Holmes, 2013) and used linear regression to test for temporal and environmental trends. We tested for any discrete differences in alpha diversity between sampling years using Kruskal–Wallis tests. We further conducted a bidirectional stepwise regression using the package MASS version 7.3.51.4 (Venables & Ripley, 2002) in order to fit the most appropriate combination of environmental variables to the regression model. We also used a Kruskal–Wallis test to determine differences between observed ASVs in *D.gouldii* tissues from 2019 and the surrounding sand.

We visualized the compositional differences between collection years and sample types using stacked barplots of the relative abundances of microbial phyla as well as principal coordinates analysis (PCoA) plots using Bray–Curtis dissimilarity (BCD) and unweighted UniFrac (Lozupone & Knight, 2005). The microbial taxa (mainly at the phylum level, although the Proteobacteria were split into classes for increased resolution) primarily responsible for the differentiation between samples were identified and visualized using the gg_envfit function (999 permutations, alpha =.05) in the package GGORDIPLOTS version 0.3.0. We then conducted permutational multivariate analyses of variance (PERMANOVAs) using sampling time or sample type as a factor to test for significant differences in composition between time points and to compare *D.gouldii* samples with the surrounding sand and water. To investigate how changes in each environmental variable, as well as time, affected the compositions of the *D.gouldii* microbial community, we generated a generalized dissimilarity model (GDM), which describes the rate of compositional turnover along environmental gradients (Ferrier et al., 2007), using the package GDM version 1.3.11 (Manion et al., 2018). We ran this model using BCD and plotted the I-splines of each predictor variable that was found to be significantly correlated with dissimilarity. We further estimated the relative impact of each predictor on BCD using the percentage deviance explained by that predictor when the rest were kept constant (Picazo et al., 2020).

To gain a finer scale understanding of which microbial taxa were driving community differentiation between years of sampling, we undertook an indicator species analysis using the package INDICSPECIES version 1.7.8 (De Cáceres & Legendre, 2009). This approach uses relative abundance and occurrence data to determine associations between groups of samples and taxa and test for the significance of such associations using permutation tests. We identified ASVs as indicators strongly correlated with a sampling point if they represented an indicator value ≥0.7 associated with a *p* ≤.05 (Ross et al., 2017, 2018).

Finally, we identified the subset of the microbiome that was stable through time, commonly referred to as the "core microbiome," suggesting a functional or ecological link between host and microbe. We defined the core as those ASVs present in at least 90% of all samples, as has been done previously in studies of this size (e.g. Lawler et al., 2016; Lorenzen et al., 2015).

3 | RESULTS

After DNA extraction, one sample from 2018 was not sequenced due to very low yield. Due to lack of adequate sequence depth, we excluded two samples from 2008, one sample from 2019 (<8000 sequences) and one water sample (<3000 sequences). Negative controls produced fewer than 150 sequences and an *Escherichia coli*-positive control returned 99.8% of sequences matching *E.coli*, indicating minimal contamination. The remaining *D. gouldii* samples produced a total of 497,180 sequences after quality and taxonomic filtering, with an average and standard deviation of 15,066 ± 6131 sequences per sample (Neu et al., 2020). These sequences formed 2195 unique ASVs and represented 48 bacterial and archaeal phyla. After rarefying to 8000 sequences per sample, 2112 ASVs remained and representatives of all 48 phyla were retained. Nearly all samples were dominated by a few taxonomic groups including the Alphaproteobacteria, Gammaproteobacteria, Planctomycetota and Campilobacterota (Figure 1b). The 2019 samples, which included *D.gouldii* individuals as well as seawater and sand, contained 3356 ASVs after rarefying to 3000 sequences per sample. Alpha diversity within *D.gouldii* was substantially lower than that of the surrounding environment (Figure 2a). Environmental samples were also compositionally distinct from *D.gouldii* samples, with each harbouring characteristic microbial taxa (Figure 2b,c). The phylum Campilobacterota, for example, was prevalent in *D.gouldii*, but was not found in the surrounding environment. Conversely, members of the phyla Acidobacteriota and Chloroflexi were abundant in both sand and water but were exceedingly rare in *D.gouldii* (Figure 2c).

Contrary to expectations of DNA degradation, the raw sequence counts, DNA yield and post-PCR DNA concentration all showed decreasing but non-significant trends over time. The directionality of these trends is opposite of the expectation of preservational bias. Raw sequence counts and DNA yield were also not significantly different using a Kruskal–Wallis test (*p* = .111 and *p* = .837, respectively), but post-PCR DNA concentration did show significant differences (*p* = .0027), with 2018 samples presenting the lowest average values (Figure S2), inconsistent with the expectation of DNA degradation over time. Furthermore, no significant differences in dispersion between time points (BCD: *p* = .197; UniFrac: *p* = .262) were evident using homogeneity of dispersion tests, indicating that cross-contamination of individuals in 2008 and 2011 samples is highly unlikely. Finally, a Venn diagram of ASVs by sampling time showed that 2008 and 2011 shared a number of taxa with the other
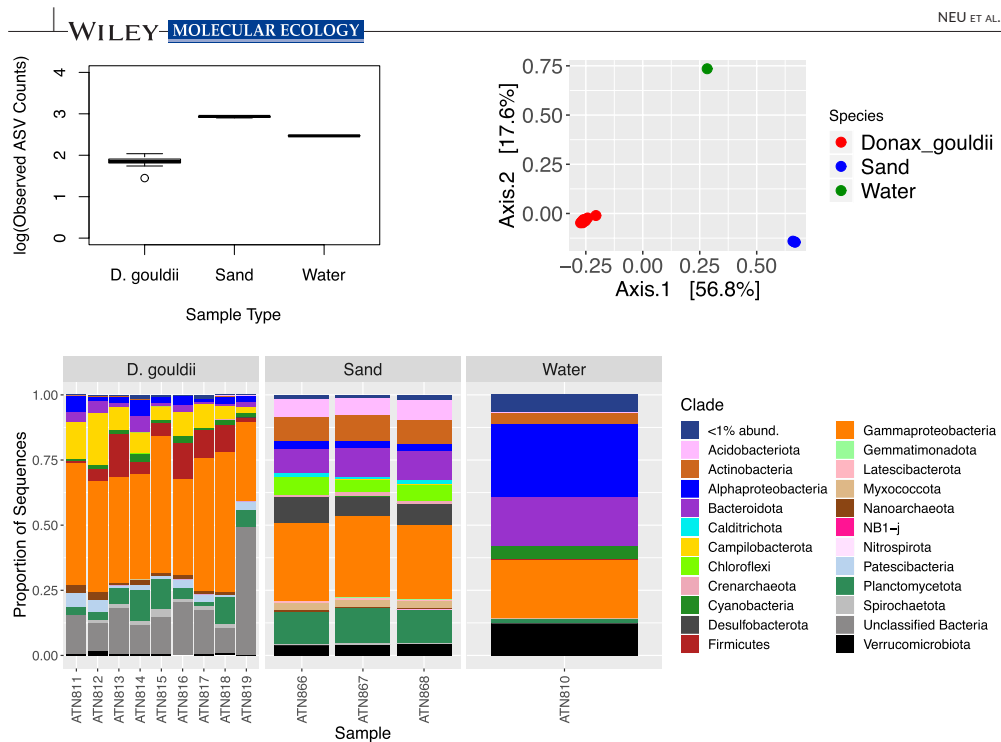
FIGURE 2 Comparison of sand (n = 3) and water (n = 1) samples with *Donax gouldii* (n = 9) collected in 2019. Boxplot of the logarithm of observed ASV counts by sample type (top left; Kruskal–Wallis: χ^2 = 7.91, p = .02), principal coordinates analysis of Bray–Curtis dissimilarity, coloured by sample type (top right; PERMANOVA: R^2 = .699, p < .001) and barplots of the relative abundances of microbial phyla, sorted by sample type (bottom). The phylum Proteobacteria is separated into classes for increased resolution

temporal bins and contained more unique sequences than the most recent samples from 2019 (Figure S3). Samples from 2018, however, had more than twice as many unique ASVs as any other time. Although the reason for the increase in unique ASVs in 2018 is unclear, the results of our other tests for preservational effects suggest that this increase is probably a true biological signal. In aggregate, the metrics used here provide little support for the hypothesis of significant preservational bias in our samples (see Section 2 for details).

The *D.gouldii* microbiomes did not show any significant temporal trends, or significant differences between time points, in the two alpha diversity metrics tested, average Shannon's index or the logarithm of observed ASV counts (Kruskal–Wallis test, p = .111 and .120, respectively; Figure 1d). This stability was a consistent pattern, as no significant correlations were evident between our alpha diversity metrics and any of the environmental variables tested, including SST, salinity and chlorophyll a concentration (Figure 3; Table S1). This was further confirmed by a stepwise linear regression model, which found that no combinations of these variables were significantly correlated with alpha diversity (best-fit model: R^2 = .054, p = .15). These results are contrary to our hypotheses of decreasing alpha diversity

with increasing SST and chlorophyll a concentration and suggest that the alpha diversity of *D.gouldii*-associated microbiomes may not be significantly impacted by external factors.

Despite the high level of stability in alpha diversity, samples from each year were significantly different compositionally, and showed significant divergence from the surrounding environment, as confirmed by PERMANOVAs using BCD and unweighted UniFrac (Figures 2 and 4). Interestingly, the PCoA results using BCD separated the 2011 samples from the other collection years along Axis 1, while the unweighted UniFrac metric separated these groups along Axis 2 (Figure 4). However, both analyses show that the differentiation of 2011 samples from the other years is primarily due to high abundances of members of the phyla Campilobacterota and Fusobacteriota (Figure 4). The differences among other years, on the other hand, are driven by a diverse set of taxa. For the analyses using BCD, these include Gammaproteobacteria, Patescibacteria and Planctomycetota, while a group of low-diversity and very low-abundance (<100 sequences) phyla such as Methylomirabilota and Deinococcota play a role in separating samples along Axis 1 in the analysis using the UniFrac metric

FIGURE 3 The impacts of environmental change on the *Donax gouldii* microbiome. Linear regression of the logarithm of observed ASV counts versus (left column) and Shannon's index (right column) by average temperature (top row), salinity (middle row) and chlorophyll *a* concentration (bottom row) for the month of collection. All $R^2 < .04$ and all $p > 0.05$; full statistics available in Table S1. Data collected from SCCOOS sampling station at Scripps Institution of Oceanography
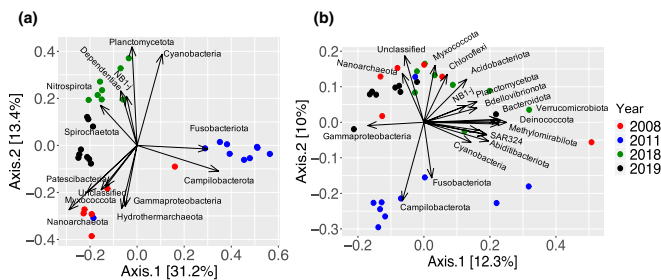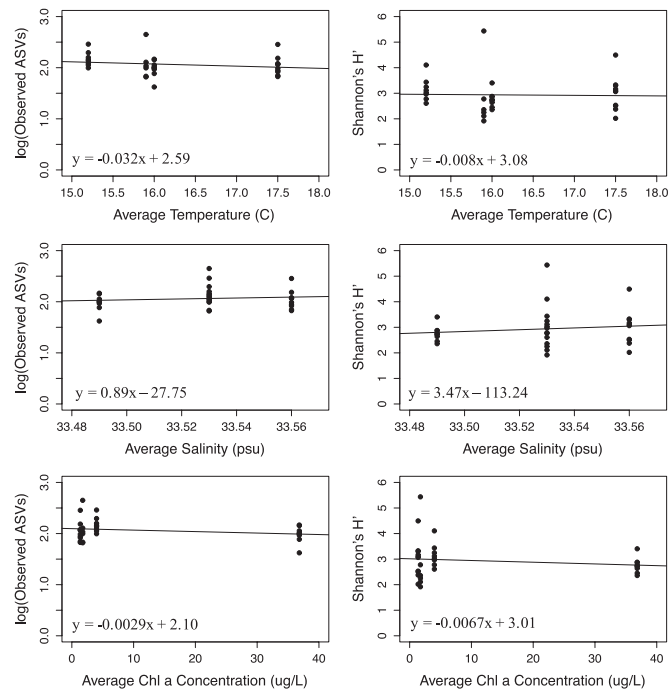
FIGURE 4 PCoA plot of (a) Bray–Curtis dissimilarity (PERMANOVA: $R^2 = .506$, $p < .001$; betadisper: $p = .197$) and (b) Unweighted UniFrac (PERMANOVA: $R^2 = .159$, $p < .001$; betadisper: $p = .262$), coloured by collection year. Arrows represent phyla (or classes of Proteobacteria) significantly contributing to differentiation between samples as determined by gg_envfit. "Unclassified" represents bacteria which could not be assigned taxonomy at the phylum level

(Figure 4). This difference reflects the different properties of the metrics used—BCD primarily reflects changes in relative abundance while unweighted UniFrac provides a more evolutionary perspective, highlighting phylogenetic distances between taxa in individual samples, thus allowing unique low-abundance taxa to play a significant role in differentiation. The distinctiveness of the 2011 samples is probably due to the fact that they were collected in May versus February and March, when SST was 1.6°C higher than at other time points (Table 1). However, the 2008, 2018 and

2019 samples were still significantly different from one another when the 2011 samples were excluded from analysis (Figure S4). Results from the GDM analysis complemented these findings and showed that the observed changes in BCD over time were significantly correlated with changes in SST, salinity and collection year, but not chlorophyll *a* concentration (Figure 5). Overall, these three variables explained 36.01% of the variation in the data, with SST contributing the most to this explanatory power. These results should be interpreted with caution, however, as the limited
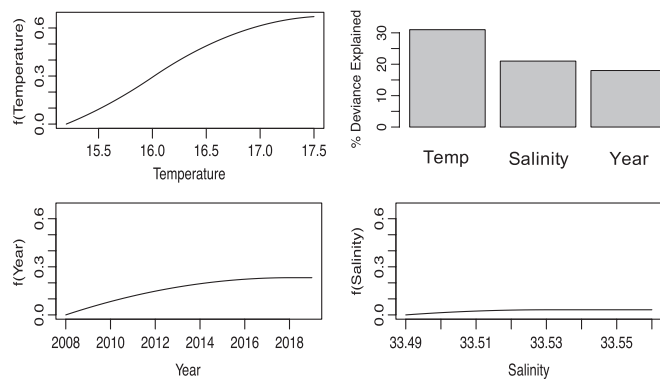
FIGURE 5 I-splines of predictors calculated by the generalized dissimilarity model using Bray–Curtis dissimilarity. The full model explained 36.01% of the deviance in the data. Grey bars (top right) show the relative utility of each environmental variable as a predictor. Chlorophyll *a* concentration was not included here as it was not significant in the model

number of time points and collection month of the 2011 samples (May vs February–March) may contribute to the high explanatory power of SST in this case.

Indicator species analysis showed that 2011 had the largest number of indicator taxa (26 ASVs), while 2008 had the fewest (two ASVs) (Figure S5, Table S2). Phylogenetic affinities of ASVs identified as indicators varied greatly among the years. For example, one of the indicator ASVs in samples from 2008 belonged to the genus *Endozoicomonas*, which primarily comprises taxa that live endosymbiotically with metazoan hosts, while the other could not be assigned at the phylum level. In contrast, indicator ASVs for 2011 were distributed across multiple phyla, including Campilobacterota and Fusobacteriota, which were the main groups separating 2011 samples from the rest in our PCoAs (Figure 4), as well as Firmicutes, Proteobacteria and Bacteroidota. The most significant indicators in 2011 came from the genus *Fusibacter* and family Arcobacteraceae, which include a number of opportunistic pathogens. For 2018, indicator ASVs were mainly members of the order Verrucomicrobiales, which includes a wide array of life strategies (Navarrete et al., 2015), while those for 2019 included members of the copiotrophic genus *Shewanella*, the host-dependent genus *Mycoplasma*, as well as the phylum Spirochaetota and some unassigned taxa.

Indicator ASVs contributed substantially to the community compositions of samples in their respective years. In 2008, for example, the two indicator ASVs contained 20.54 ± 10.58% (mean ± *SD*) of sequences on average, despite comprising <2% of the relative abundance of any other sample (Figure S6). In the 2011 samples, sequences from indicator ASVs accounted for nearly half (49.51 ± 25.13%) of the mean relative abundance (Figure S6), which may further explain why the microbiome compositions of individuals collected that year are so distinct from the other time periods (Figure 4). Similarly, the indicators of 2018 comprised 12.50 ± 7.58% mean relative abundance, although they were scarce (<0.5% mean relative abundance) in other time points (Figure S6). Interestingly, the 2019 indicators made up the lowest proportion of sequences, comprising 8.51 ± 2.54% relative abundance, on average, despite the

more than nine-fold increase in chlorophyll *a* concentration during this time (Figure S6).

Although there was significant differentiation in the composition of *D. gouldii* microbiomes over time and across changing environments, the four sampled time points shared 66 ASVs in common (Figure S3), which may constitute a resident microbiome at this locality. Six of these ASVs were present in >90% of samples sequenced, representing a more stringent "core" microbiome for *D. gouldii* (Lawler et al., 2016; Lorenzen et al., 2015). These included members of the genera *Blastopirellula* and *Rubripirellula*, an ASV within the family Helicobacteraceae, a member of the gammaproteobacterial order HglApr721 and two ASVs which were assigned above the ordinal level (Table S3). These six core ASVs accounted for between 3.3% and 73.2% relative abundance, depending on the sample, with the lowest percentages occurring in 2011 (Figure S7).

4 | DISCUSSION

The lack of significant temporal trends in DNA yield, post-PCR DNA concentration, raw sequence counts or alpha diversity from 2008 to 2019 strongly suggests that storage in 95% ethanol at 4°C was sufficient for the preservation of *D. gouldii* microbiomes. Furthermore, dispersion between samples was not significantly different between years sampled here, indicating that there was as much intra-annual variability in 2008 as there was in 2011, 2018 or 2019. Thus, it is also unlikely that there was any significant homogenization of the microbiome or contamination due to storage conditions, which should have led to lower variability between samples preserved together compared to fresh samples. The largest numbers of unique sequences were present in our 2018 samples, but other archival (2008 and 2011) as well as freshly collected (2019) samples each had substantial numbers of unique ASVs as well. Together, these results show that the temporal trends in the *D. gouldii* microbiome documented here probably represent real biological signals rather than artefacts of long-term

sample preservation. In particular, the presence of large numbers of unique ASVs in our archival samples shows the importance of archival samples in capturing the compositional breadth of host microbiomes, something that cannot be done by sampling single time points.

The alpha diversity of the *D.gouldii* microbiome showed striking stability over the 11-year study period despite substantial variation in both temperature and chlorophyll *a* concentration, which have both been previously shown to decrease microbial diversity in free-living systems (Gilbert et al., 2012; Wemheuer et al., 2014), while temperature has also been shown to decrease diversity in host-associated microbiomes (Li et al., 2018). The mechanism allowing for the maintenance of stable diversity in *D.gouldii* remains unclear, but could reflect a level of control asserted by the host (e.g., an immune response; Foster et al., 2017), competitive exclusion between members of the microbial community, or a physical barrier to further colonization, such as space limitation. It has also been suggested that changes in microbial richness may be driven by host physiology (e.g., gut length), stage of development or diet composition, among other factors (Reese & Dunn, 2018). It is unlikely that any of these properties changed substantially in the sampled individuals of *D.gouldii*. All samples used in this study were between 15 and 20 mm in length, indicating they were all fully developed adults between 1.5 and 3 years of age (Coe, 1955). Diet, however, may have varied between collection times, because *D.gouldii* are filter feeders, but each population sampled was likely to come into contact with similar types of food, including detritus, phytoplankton and bacteria themselves (Coe, 1948) as they were collected from the same location year after year and only occupy a band of the intertidal zone that is 2–5 m wide (Coe, 1955). Samples from 2019 were probably exposed to a greater abundance of phytoplankton due to the ongoing bloom, but this did not have a significant impact on alpha diversity.

The compositions of the *D.gouldii* microbiomes were primarily dominated by clades previously associated with healthy marine bivalves (Pierce et al., 2016), although at different relative abundances. However, they were compositionally distinct from the microbiomes of a number of marine invertebrates collected at an adjacent rocky intertidal site, which were dominated by Fusobacteriota, Verrucomicrobiota and others (Neu et al., 2019). The compositions of *D.gouldii* microbial communities were also significantly different from one another between sampling events. These changes were most strongly correlated with SST, followed by salinity and year of collection. Shifts in microbiome compositions seen here are similar to those in previous temperature stress experiments with marine bivalves, showing an increase in the relative abundance of the family Arcobacteraceae, a member of the phylum Campilobacterota (Lokmer & Wegner, 2015). However, natural changes in SST observed across this study were markedly lower than those used in previous short-term climate change experiments: 4–14°C (Li et al., 2018; Lokmer & Wegner, 2015) versus a range of ~1.6°C from 2008 to 2011. Similarly, differences in salinity have been shown to alter microbiome composition in experiments with multiple hosts including seaweeds (Saha et al., 2020) and corals (Röthig et al., 2016),

although salinity variations in such experiments (8–19 practical salinity units (psu)) are typically much larger than the natural variations seen here (<0.1 psu). Decade-scale microbiome differentiation has been previously shown in the human gut, although nearly 60% of bacterial phylotypes persisted over more than a decade (Faith et al., 2013). *Donax gouldii* microbiomes, on the other hand, showed significant turnover, with only six core ASVs identified. This is probably due to the fact that the same humans were repeatedly sampled in previous studies, while new populations of *D.gouldii* were surveyed at each sampling point due to their relatively short lifespan.

The compositional changes outlined above were largely driven by the proliferation of the indicator ASVs at each time point. In the 2011 samples, the dominance of the indicator ASVs may have been due to their roles as potential pathogens. The genus *Fusibacter*, for example, has been found in association with disease states in marine organisms, such as black band disease in corals (Meyer et al., 2016) and Pacific oyster mortality syndrome (de Lorgeril et al., 2018), while some members of the family Arcobacteraceae are known human pathogens (Collado & Figueras, 2011; Vandenberg et al., 2004) and have been associated with stressed, diseased and dead marine molluscs (King et al., 2019; Lokmer & Wegner, 2015; Madigan et al., 2014). These taxa may have opportunistically colonized *D.gouldii* hosts as a result of increased temperature stress and/or other seasonal effects (including nutrient availability, photoperiod, etc.) experienced by the 2011 samples. The most prominent indicators in the 2019 samples, which were collected during a phytoplankton bloom, included members of the genera *Shewanella* and *Mycoplasma*, which have recently been found in association with a number of healthy marine molluscs (King et al., 2012; Neu et al., 2019). Members of these genera are known to break down complex polysaccharides and may play a role in algal digestion (Ivanova et al., 2001; Nam et al., 2018). Phytoplankton bloom-associated taxa, including members of the SAR11, Flavobacteria, MGII Archaea (Needham & Fuhrman, 2016) and Planctomycetota (Morris et al., 2006; Zeigler Allen et al., 2012) clades, were scarce among the 2019 indicator ASVs, with only one member of the phylum Planctomycetota represented. These taxa were present in the environmental samples collected in 2019, with Planctomycetota and Flavobacteria representing 5.8%–14.9% and 2.4%–3.3% of the relative abundance of sand and water samples, respectively. Furthermore, SAR11 comprised 1.74% of the relative abundance in the water sample. Overall, the environmental samples shared few microbial taxa in common with *D.gouldii*, suggesting that the host may be selectively enriching particular taxa from its environment and sustaining rare microbial diversity not commonly detected in the sand or water column (Troussellier et al., 2017). In the 2008 samples, only one of the indicator ASVs was able to be classified past the level of domain, a member of the genus *Endozoicomonas*. This genus has been of much interest over the past several years due to its association with a wide array of marine hosts (Neave et al., 2017; Schill et al., 2017). Although this particular ASV was found in nearly every sample across the years, its massive proliferation in 2008 is notable and may be due to its role in carbon

sugar transport and protein secretion, which may provide bene-fits to the host (Neave et al., 2017). Lastly, 2018 samples were largely indicated by members of the Verrucomicrobiota, including *Roseibacillus* and *Rubritalea*, two genera commonly found in the water column, sediment and associated with macroalgae (Freitas et al., 2012; Lachnit et al., 2011). These taxa may have been associ-ated with particles in the water column that were filtered through the gills and colonized to form biofilms. Unfortunately, sediment or water samples were not available from 2018 to directly test this hypothesis.

Finally, the six core ASVs identified here were the dominant members of the community in many individuals, except in those from 2011 (relative abundances ranging from 3.3% to 73.2% across all samples). Therefore, there may be a small subset of bacterial taxa that maintain a stable relationship with *D.gouldii*, whether mutualistic, commensal or parasitic. The functional roles of these taxa remain unclear because only two, members of *Blastopirellula* and *Rubripirellula*, could be identified at the genus level and one, a member of the Helicobacteraceae, could be identified at the family level. *Blastopirellula* and *Rubripirellula* are relatively newly described marine genera in the phylum Planctomycetota and con-tain few representative species (Bondoso et al., 2015; Schlesner et al., 2004). Interestingly, both of these genera have been de-scribed previously as members of macroalgal biofilms (Bondoso et al., 2015; Faria et al., 2018; Lage & Bondoso, 2014), while *Rubripirellula* has also been found in association with particles of marine plastics (Kallscheuer et al., 2019). Therefore, members of these genera may be filtered from the environment and form biofilms in the gills and other organs of *D.gouldii*, as these gen-era were also found in surrounding sediments. Members of the family Helicobacteraceae have been recently described in asso-ciation with marine invertebrates including sea stars and anemo-nes where they have been predicted to aid in sulphide oxidation (Murray et al., 2016; Nakagawa et al., 2017). Future studies using metagenomics and/or metatranscriptomics methods are needed to better identify this core component of the *D.gouldii* microbiome and their functional roles.

Together, our results indicate that for some hosts, seasonal or larger scale environmental changes may not lead to major changes in the diversity of their microbiomes, but they may be enough to significantly alter their microbiome compositions. This has potential implications for predicting the impacts of anthropogenic change, as an increase in global mean SST of 0.6–3.2°C is currently pre-dicted for the year 2100 (Aral & Guan, 2016), alongside a decrease in the scale of natural spring phytoplankton blooms (Sommer & Lengfellner, 2008) and increase in the frequency of harmful algal blooms (Hallegraeff, 2010). In this study, a seasonal temperature increase showed the strongest impact on shifts in the microbiome of *D.gouldii.* A major phytoplankton bloom, on the other hand, did not seem to disrupt the microbiome more than normal temporal turnover, as GDM analysis did not identify chlorophyll *a* concen-tration as a significant driver of microbial community composition. Furthermore, samples from 2019 were compositionally distinct from

the environment, suggesting that *D.gouldii* may be able to filter par-ticular taxa from its environment despite an increase in phytoplank-ton density and act as a reservoir for rare or dormant taxa from the surrounding sand and seawater. These samples also had the lowest contribution from indicator ASVs and large contributions from the "core" ASVs. More study on this topic is clearly warranted, however, as it is unclear how host-associated microbiomes respond through-out a phytoplankton bloom or in response to blooms of different species.

5 | CONCLUSION

By leveraging historical samples collected at a single site during mul-tiple years over more than a decade, we have shown that ethanol-preserved samples, a previously untapped resource in microbiome research, can be used to better understand how microbiomes of wild populations of marine hosts respond to environmental changes on longer timescales than would be tractable through laboratory ex-periments. The alpha diversity of microbial communities associated with *D.gouldii* remained surprisingly stable in the face of increases in temperature and chlorophyll *a,* but the underlying community composition was much more dynamic, with high levels of differ-entiation between collections. Specifically, increases in ambient temperature and salinity significantly altered the microbiome com-position, although other environmental factors may have also con-tributed to this shift. Further work uncovering the functional roles of dominant taxa in the community and how these change through time and across environmental regimes will be critical for better un-derstanding how individual hosts will be impacted in the years to come. Such research should better utilize samples that have been freshly preserved in ethanol, common in many museum and labora-tory collections, although the utility of such samples for other taxa or time frames need to be evaluated. The tests for preservational biases proposed here have been well validated by the palaeontologi-cal community and can help in this regard. Hopefully, future stud-ies of archival material will allow us to extend our understanding of host–microbe associations much further in the past than has been possible so far.

AUTHOR CONTRIBUTIONS
A.T.N., E.E.A. and K.R. designed the project. A.T.N. and I.V.H. con-ducted the molecular work. A.T.N. analysed the data, with guid-ance from E.E.A. and K.R. A.T.N. wrote the initial manuscript, with

comments from I.V.H., E.E.A. and K.R. All authors read and approved the final manuscript.

ORCID

Alexander T. Neu 🔟 https://orcid.org/0000-0003-0833-1704

REFERENCES

Adam, E., Bernhart, M., Müller, H., Winkler, J., & Berg, G. (2018). The Cucurbita pepo seed microbiome: Genotype-specific composition and implications for breeding. *Plant and Soil*, *422*(1–2), 35–49. https://doi.org/10.1007/s11104-016-3113-9

Apprill, A. (2017). Marine animal microbiomes: Toward understanding host – Microbiome interactions in a changing ocean. *Frontiers in Marine Science*, *4*, 222. https://doi.org/10.3389/fmars.2017.00222

Aral, M. M., & Guan, J. (2016). Global sea surface temperature and sea level rise estimation with optimal historical time lag data. *Water*, *8*, 519. https://doi.org/10.3390/w8110519

Batstone, R. T., & Dufour, S. C. (2016). Closely related thyasirid bivalves associate with multiple symbiont phylotypes. *Marine Ecology*, *37*(5), 988–997. https://doi.org/10.1111/maec.12310

Beeson, R. J., & Johnson, P. T. (1967). Natural bacterial flora of the bean clam, *Donax gouldi*. *Journal of Invertebrate Pathology*, *9*, 104–110.

Behrensmeyer, A. K., Kidwell, S. M., & Gastaldo, R. A. (2000). Taphonomy and paleobiology. *Paleobiology*, *26*(4), 103–147.

Blekhman, R., Tang, K., Archie, E. A., Barreiro, L. B., Johnson, Z. P., Wilson, M. E., Kohn, J., Yuan, M. L., Gesquiere, L., Grieneisen, L. E., & Tung, J. (2016). Common methods for fecal sample storage in field studies yield consistent signatures of individual identity in microbiome sequencing data. *Scientific Reports*, *6*, 31519. https:// doi.org/10.1038/srep31519

Bondoso, J., Albuquerque, L., Nobre, M. F., Lobo-da-Cunha, A., da Costa, M. S., & Lage, O. M. (2015). *Roseimaritima ulvae* gen. nov., sp. nov. and *Rubripirellula obstinata* gen. nov., sp. nov. two novel planctomycetes isolated from the epiphytic community of macroalgae. *Systematic and Applied Microbiology*, *38*(1), 8–15. https://doi.org/10.1016/j.syapm.2014.10.004

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*, 581–583. https://doi.org/10.1038/nmeth.3869

Carrier, T. J., & Reitzel, A. M. (2017). The hologenome across environments and the implications of a host-associated microbial repertoire. *Frontiers in Microbiology*, *8*, 802. https://doi.org/10.3389/fmicb.2017.00802

Cavicchioli, R., Ripple, W. J., Timmis, K. N., Azam, F., Bakken, L. R., Baylis, M., Behrenfeld, M. J., Boetius, A., Boyd, P. W., Classen, A. T., Crowther, T. W., Danovaro, R., Foreman, C. M., Huisman, J., Hutchins, D. A., Jansson, J. K., Karl, D. M., Koskella, B., Mark Welch, D. B., … Webster, N. S. (2019). Scientists' warning to humanity: Microorganisms and climate change. *Nature Reviews Microbiology*, *17*(9), 569–586. https://doi.org/10.1038/s41579-019-0222-5

Chen, H. (2018). VennDiagram: Generate High-Resolution Venn and Euler Plots. R package version 1.6.20. Retrieved from https:// CRAN.R-project.org/package=VennDiagram

Coe, W. R. (1948). Nutrition, environmental conditions, and growth of marine bivalve mollusks. *Journal of Marine Research*, *7*(3), 586–601.

Coe, W. R. (1953). Resurgent populations of littoral marine invertebrates and their dependence on ocean currents and tidal currents. *Ecology*, *34*(1), 225–229.

Coe, W. R. (1955). Ecology of the bean clam *Donax gouldi* on the Coast of Southern California. *Ecology*, *36*(3), 512–514.

Collado, L., & Figueras, M. J. (2011). Taxonomy, epidemiology, and clinical relevance of the genus Arcobacter. *Clinical Microbiology Reviews*, *24*(1), 174–192. https://doi.org/10.1128/CMR.00034-10.

De Cáceres, M., & Legendre, P. (2009). Associations between species and groups of sites: Indices and statistical inference. *Ecology*, *90*(12), 3566–3574. https://doi.org/10.1890/08-1823.1.

de Lorgeril, J., Lucasson, A., Petton, B., Toulza, E., Montagnani, C., Clerissi, C., Vidal-Dupiol, J., Chaparro, C., Galinier, R., Escoubas, J.-M., Haffner, P., Dégremont, L., Charrière, G. M., Lafont, M., Delort, A., Vergnes, A., Chiarello, M., Faury, N., Rubio, T., …Mitta, G. (2018). Immune-suppression by OsHV-1 viral infection causes fatal bacteraemia in Pacific oysters. *Nature Communications*, *9*, 4215. https:// doi.org/10.1038/s41467-018-06659-3

Eloe-Fadrosh, E. A., Ivanova, N. N., Woyke, T., & Kyrpides, N. C. (2016). Metagenomics uncovers gaps in amplicon-based detection of microbial diversity. *Nature Microbiology*, *1*, 15032. https://doi. org/10.1038/nmicrobiol.2015.32

Faith, J. J., Guruge, J. L., & Charbonneau, M. (2013). The long-term stability of the human gut microbiota. *Science*, *341*(6141), 1237439. Retrieved from: http://www.sciencemag.org/content/341/6141/1237439.short

Faria, M., Bordin, N., Kizina, J., Harder, J., Devos, D., & Lage, O. M. (2018). Planctomycetes attached to algal surfaces: Insight into their genomes. *Genomics*, *110*(5), 231–238. https://doi.org/10.1016/j. ygeno.2017.10.007

Ferrier, S., Manion, G., Elith, J., & Richardson, K. (2007). Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Diversity and Distributions*, *13*(3), 252–264. https://doi. org/10.1111/j.1472-4642.2007.00341.x

Foote, M., & Raup, D. M. (1996). Fossil preservation and the stratigraphic ranges of taxa. *Paleobiology*, *22*(2), 121–140.

Foster, K. R., Schluter, J., Coyte, K. Z., & Rakoff-Nahoum, S. (2017). The evolution of the host microbiome as an ecosystem on a leash. *Nature*, *548*(7665), 43–51. https://doi.org/10.1038/nature23292

Freitas, S., Hatosy, S., Fuhrman, J. A., Huse, S. M., Mark Welch, D. B., Sogin, M. L., & Martiny, A. C. (2012). Global distribution and diversity of marine Verrucomicrobia. *ISME Journal*, *6*(8), 1499–1505. https://doi.org/10.1038/ismej.2012.3

Garcia-Recinos, L., Burrowes, P. A., & Dominguez-Bello, M. (2019). The skin microbiota of Eleutherodactylus frogs: Effects of host ecology, phylogeny, and local environment. *Frontiers in Microbiology*, *10*, 2571. https://doi.org/10.3389/fmicb.2019.02571

Gilbert, J. A., Steele, J. A., Caporaso, J. G., Steinbru, L., Reeder, J., Temperton, B., Huse, S., McHardy, A. C., Knight, R., Joint, I., Somerfield, P., Fuhrman, J. A., & Field, D. (2012). Defining seasonal marine microbial community dynamics. *The ISME Journal*, *6*, 298–308. https://doi.org/10.1038/ismej.2011.107

Gray, M. A., Pratte, Z. A., & Kellogg, C. A. (2013). Comparison of DNA preservation methods for environmental bacterial community samples. *FEMS Microbiology Ecology*, *83*(2), 468–477. https://doi. org/10.1111/1574-6941.12008

Hale, V. L., Tan, C. L., Knight, R., & Amato, K. R. (2015). Effect of preservation method on spider monkey (*Ateles geoffroyi*) fecal microbiota over 8 weeks. *Journal of Microbiological Methods*, *113*, 16–26. https://doi.org/10.1016/j.mimet.2015.03.021

Hallegraeff, G. M. (2010). Ocean climate change, phytoplankton community responses, and harmful algal blooms: A formidable predictive challenge. *Journal of Phycology*, *46*(2), 220–235. https://doi.org/10.1111/j.1529-8817.2010.00815.x

Heindler, F. M., Christiansen, H., Frédérich, B., Dettaï, A., Lepoint, G., Maes, G. E., Van de Putte, A. P., & Morgan, K. (2018). Historical DNA metabarcoding of the prey and microbiome of trematomid fishes using museum samples. *Frontiers in Ecology and Evolution*, *6*, 151. https://doi.org/10.3389/fevo.2018.00151

Hellberg, M. E., Balch, D. P., & Roy, K. (2001). Climate-driven range expansion and morphological evolution in a marine gastropod. *Science*, *292*(5522), 1707–1710. https://doi.org/10.1126/science.1060102

Ivanova, E. P., Sawabe, T., Gorshkova, N. M., Svetashev, V. I., Mikhailov, V. V., Nicolau, D. V., & Christen, R. (2001). *Shewanella japonica* sp. nov. *International Journal of Systematic and Evolutionary Microbiology*, *51*, 1027–1033.

Jayasinghe, T. N., Hilton, C., Tsai, P., Apple, B., Shepherd, P., Cut, W. S., & Sullivan, J. M. O. (2017). Long-term stability in the gut microbiome over 46 years in the life of Billy Apple. *Human Microbiome Journal*, *6*, 7–10. https://doi.org/10.1016/j.humic.2017.09.001.

Kallscheuer, N., Jogler, M., Wiegand, S., Peeters, S. H., Heuer, A., Boedeker, C., Jetten, M. S. M., Rohde, M., & Jogler, C. (2019). Three novel Rubripirellula species isolated from plastic particles submerged in the Baltic Sea and the estuary of the river Warnow in northern Germany. *Antonie Van Leeuwenhoek*, *1–12*, https://doi.org/10.1007/s10482-019-01368-3

Kia, E., Mackenzie, B. W., Middleton, D., Lau, A., Waite, D. W., Lewis, G., Chan, Y.-K., Silvestre, M., Cooper, G. J. S., Poppitt, S. D., & Taylor, M. W. (2016). Integrity of the human faecal microbiota following long-term sample storage. *PLoS One*, *11*(10), e0163666. https://doi.org/10.1371/journal.pone.0163666

King, G. M., Judd, C., Kuske, C. R., & Smith, C. (2012). Analysis of stomach and gut microbiomes of the eastern oyster (*Crassostrea virginica*) from coastal Louisiana, USA. *PLoS One*, *7*(12), e51475. https://doi.org/10.1371/journal.pone.0051475

King, W. L., Jenkins, C., Go, J., Siboni, N., Seymour, J. R., & Labbate, M. (2019). Characterisation of the pacific oyster microbiome during a summer mortality event. *Microbial Ecology*, *77*(2), 502–512. https://doi.org/10.1007/s00248-018-1226-9

Klymiuk, I., Bambach, I., Patra, V., Trajanoski, S., & Wolf, P. (2016). 16S based microbiome analysis from healthy subjects' skin swabs stored for different storage periods reveal phylum to genus level changes. *Frontiers in Microbiology*, *7*, 2012. https://doi.org/10.3389/fmicb.2016.02012

Komatsu, K. J., Avolio, M. L., Lemoine, N. P., Isbell, F., Grman, E., Houseman, G. R., Koerner, S. E., Johnson, D. S., Wilcox, K. R., Alatalo, J. M., Anderson, J. P., Aerts, R., Baer, S. G., Baldwin, A. H., Bates, J., Beierkuhnlein, C., Belote, R. T., Blair, J., Bloor, J. M. G., … Zhang, Y. (2019). Global change effects on plant communities are magnified by time and the number of global change factors imposed. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(36), 17867–17873. https://doi.org/10.1073/pnas.1819027116

Lachnit, T., Meske, D., Wahl, M., Harder, T., & Schmitz, R. (2011). Epibacterial community patterns on marine macroalgae are host-specific but temporally variable. *Environmental Microbiology*, *13*(3), 655–665. https://doi.org/10.1111/j.1462-2920.2010.02371.x

Lage, O. M., & Bondoso, J. (2014). Planctomycetes and macroalgae, a striking association. *Frontiers in Microbiology*, *5*, 267. https://doi.org/10.3389/fmicb.2014.00267

Lawler, S. N., Kellogg, C. A., France, S. C., Clostio, R. W., Brooke, S. D., & Ross, S. W. (2016). Coral-associated bacterial diversity is conserved across two deep-sea Anthothela species. *Frontiers in Microbiology*, *7*, 458. https://doi.org/10.3389/fmicb.2016.00458.

Lesser, M. P., Fiore, C., Slattery, M., & Zaneveld, J. (2016). Climate change stressors destabilize the microbiome of the Caribbean barrel sponge, *Xestospongia muta*. *Journal of Experimental Marine Biology and Ecology*, *475*, 11–18. https://doi.org/10.1016/j.jembe.2015.11.004

Li, Y. F., Yang, N., Liang, X., Yoshida, A., Osatomi, K., Power, D., Batista, F. M., & Yang, J. L. (2018). Elevated seawater temperatures decrease microbial diversity in the gut of *Mytilus coruscus*. *Frontiers in Physiology*, *9*, 839. https://doi.org/10.3389/fphys.2018.00839

Lokmer, A., & Mathias Wegner, K. (2015). Hemolymph microbiome of Pacific oysters in response to temperature, temperature stress and infection. *The ISME Journal*, *9*(3), 670–682. https://doi.org/10.1038/ismej.2014.160

Lorenzen, E., Kudirkiene, E., Gutman, N., Grossi, A. B., Agerholm, J. S., Erneholm, K., Skytte, C., Dalgaard, M. D., & Bojesen, A. M. (2015). The vaginal microbiome is stable in prepubertal and sexually mature Ellegaard Göttingen Minipigs throughout an estrous cycle. *Veterinary Research*, *46*, 125. https://doi.org/10.1186/s13567-015-0274-0

Lozupone, C., & Knight, R. (2005). UniFrac: A new phylogenetic method for comparing microbial communities. *Applied and Environmental Microbiology*, *71*(12), 8228–8235. https://doi.org/10.1128/AEM.71.12.8228

Madigan, T. L., Bott, N. J., Torok, V. A., Percy, N. J., Carragher, J. F., de Barros Lopes, M. A., & Kiermeier, A. (2014). A microbial spoilage profile of half shell Pacific oysters (*Crassostrea gigas*) and Sydney rock oysters (Saccostrea glomerata). *Food Microbiology*, *38*, 219–227. https://doi.org/10.1016/j.fm.2013.09.005

Manion, G., Lisk, M., Ferrier, S., Nieto-Lugilde, D., Mokany, K., & Fitzpatrick, M. C. (2018). gdm: Generalized Dissimilarity Modeling. R package version 1.3.11. https://CRAN.R-project.org/package=gdm

Maurice, C. F., Knowles, S. C. L., Ladau, J., Pollard, K. S., Fenton, A., Pedersen, A. B., & Turnbaugh, P. J. (2015). Marked seasonal variation in the wild mouse gut microbiota. *The ISME Journal*, *9*, 2423–2434. https://doi.org/10.1038/ismej.2015.53

McLaren, M. R. (2020). Silva SSU taxonomic training data formatted for DADA2 (Silva version 138). https://doi.org/10.5281/ZENODO.3731176

McMurdie, P. J., & Holmes, S. (2013). Phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One*, *8*(4), e61217. https://doi.org/10.1371/journal.pone.0061217

Meyer, J. L., Gunasekera, S. P., Scott, R. M., Paul, V. J., & Teplitski, M. (2016). Microbiome shifts and the inhibition of quorum sensing by Black Band Disease cyanobacteria. *ISME Journal*, *10*(5), 1204–1216. https://doi.org/10.1038/ismej.2015.184

Minich, J. J., Morris, M. M., Brown, M., Doane, M., Edwards, M. S., Michael, T. P., & Dinsdale, E. A. (2018). Elevated temperature drives kelp microbiome dysbiosis, while elevated carbon dioxide induces water microbiome disruption. *PLoS One*, *13*(2), e0192772. https://doi.org/10.1371/journal.pone.0192772

Minich, J. J., Petrus, S., Michael, J. D., Michael, T. P., Knight, R., & Allen, E. E. (2020). Temporal, environmental, and biological drivers of the mucosal microbiome in a Wild Marine Fish, *Scomber japonicus*. *Msphere*, *5*, e00401-20

Minich, J. J., Sanders, J. G., Amir, A., Humphrey, G., Gilbert, J. A., & Knight, R. (2019). Quantifying and understanding well-to-well contamination in microbiome research. *Msystems*, *4*(4), e00186-19. https://doi.org/10.1128/msystems.00186-19

Morris, R. M., Longnecker, K., & Giovannoni, S. J. (2006). Pirellula and OM43 are among the dominant lineages identified in an Oregon coast diatom bloom. *Environmental Microbiology*, *8*(8), 1361–1370. https://doi.org/10.1111/j.1462-2920.2006.01029.x

Murray, A. E., Rack, F. R., Zook, R., Williams, M. J. M., Higham, M. L., Broe, M., Kaufmann, R. S., & Daly, M. (2016). Microbiome composition

and diversity of the ice-dwelling sea anemone, *Edwardsiella andrillae*. *Integrative and Comparative Biology*, *56*(4), 542–555. https://doi.org/10.1093/icb/icw095

Nakagawa, S., Saito, H., Tame, A., Hirai, M., Yamaguchi, H., Sunata, T., Aida, M., Muto, H., Sawayama, S., & Takaki, Y. (2017). Microbiota in the coelomic fluid of two common coastal starfish species and characterization of an abundant Helicobacter-related taxon. *Scientific Reports*, *7*, 8764. https://doi.org/10.1038/s41598-017-09355-2

Nam, B.-H., Jang, J., Caetano-Anolles, K., Kim, Y.-O., Park, J. Y., Sohn, H., Yoon, S. H., Kim, H., & Kwak, W. (2018). Microbial community and functions associated with digestion of algal polysaccharides in the visceral tract of *Haliotis discus* hannai: Insights from metagenome and metatranscriptome analysis. *PLoS One*, *13*(10), e0205594.

Navarrete, A. A., Soares, T., Rossetto, R., van Veen, J. A., Tsai, S. M., & Kuramae, E. E. (2015). Verrucomicrobial community structure and abundance as indicators for changes in chemical factors linked to soil fertility. *Antonie Van Leeuwenhoek*, *108*, 741–752. https://doi.org/10.1007/s10482-015-0530-3.

Neave, M. J., Michell, C. T., Apprill, A., & Voolstra, C. R. (2017). Endozoicomonas genomes reveal functional adaptation and plasticity in bacterial strains symbiotically associated with diverse marine hosts. *Scientific Reports*, *7*, 40579. https://doi.org/10.1038/srep40579

Needham, D. M., & Fuhrman, J. A. (2016). Pronounced daily succession of phytoplankton, archaea and bacteria following a spring bloom. *Nature Microbiology*, *1*, 16005. https://doi.org/10.1038/nmicrobiol.2016.5

Neu, A. T., Allen, E. E., & Roy, K. (2019). Diversity and composition of intertidal gastropod microbiomes across a major marine biogeographic boundary. *Environmental Microbiology Reports*, *11*(3), 434–447. https://doi.org/10.1111/1758-2229.12743

Neu, A. T., Hughes, I. V., Allen, E. E., & Roy, K. (2020). 16S rRNA gene sequences from *D. gouldii* and the surrounding environment. NCBI Sequence Read Archive.

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Peter Solymos, M., Stevens, H. H., Szoecs, E., & Wagner, H. (2019). vegan: Community Ecology Package. R package version 2.5-6. https://CRAN.R-project.org/package=vegan

Picazo, F., Vilmi, A., Aalto, J., Soininen, J., Casamayor, E. O., Liu, Y., Wu, Q., Ren, L., Zhou, J., Shen, J., & Wang, J. (2020). Climate mediates continental scale patterns of stream microbial functional diversity. *Microbiome*, *8*, 92. https://doi.org/10.1186/s40168-020-00873-2

Pierce, M. L., Ward, J. E., Holohan, B. A., Zhao, X., & Hicks, R. E. (2016). The influence of site and season on the gut and pallial fluid microbial communities of the eastern oyster, *Crassostrea virginica* (Bivalvia, Ostreidae): Community-level physiological profiling and genetic structure. *Hydrobiologia*, *765*(1), 97–113. https://doi.org/10.1007/s10750-015-2405-z

Pohlo, R. H. (1967). Aspects of the biology of *Donax gouldi* and a note on the evolution in Tellinacea (Bivalvia). *The Veliger*, *9*, 330–337.

Pold, G., Melillo, J. M., & DeAngelis, K. M. (2015). Two decades of warming increases diversity of a potentially lignolytic bacterial community. *Frontiers in Microbiology*, *6*, 480. https://doi.org/10.3389/fmicb.2015.00480

Poloczanska, E. S., Brown, C. J., Sydeman, W. J., Kiessling, W., Schoeman, D. S., Moore, P. J., Brander, K., Bruno, J. F., Buckley, L. B., Burrows, M. T., Duarte, C. M., Halpern, B. S., Holding, J., Kappel, C. V., O'Connor, M. I., Pandolfi, J. M., Parmesan, C., Schwing, F., Thompson, S. A., & Richardson, A. J. (2013). Global imprint of climate change on marine life. *Nature Climate Change*, *3*(10), 919–925. https://doi.org/10.1038/nclimate1958

Poloczanska, E. S., Burrows, M. T., Brown, C. J., Molinos, J. G., Halpern, B. S., Hoegh-Guldberg, O., Kappel, C. V., Moore, P. J., Richardson, A. J., Schoeman, D. S., & Sydeman, W. J. (2016). Responses of marine organisms to climate change across oceans. *Frontiers in Marine Science*, *3*, 62. https://doi.org/10.3389/fmars.2016.00062

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research*, *41*, D590–D596. https://doi.org/10.1093/nar/gks1219

R Core Team. (2013). R: A language and environment for statistical computing.

Rajilić-Stojanović, M., Heilig, H. G. H. J., Tims, S., Zoetendal, E. G., & De Vos, W. M. (2013). Long-term monitoring of the human intestinal microbiota composition. *Environmental Microbiology*, *15*(4), 1146–1159. https://doi.org/10.1111/1462-2920.12023

Rasmussen, L. L., Carter, M. L., Flick, R. E., Hilbern, M., Fumo, J. T., Cornuelle, B. D., Gordon, B. K., Bargatze, L. F., Gordon, R. L., & McGowan, J. A. (2020). A century of Southern California coastal ocean temperature measurements. *Journal of Geophysical Research: Oceans*, *125*(5), e2019JC015673. https://doi.org/10.1029/2019jc015673

Reese, A. T., & Dunn, R. R. (2018). Drivers of microbiome biodiversity: A review of general rules, feces, and ignorance. *MBio*, *9*(4), e01294-18.

Reshef, L., Koren, O., Loya, Y., Zilber-Rosenberg, I., & Rosenberg, E. (2006). The coral probiotic hypothesis. *Environmental Microbiology*, *8*(12), 2068–2073. https://doi.org/10.1111/j.1462-2920.2006.01148.x

Ross, A. A., Doxey, A. C., & Neufeld, J. D. (2017). The skin microbiome of cohabiting couples. *Msystems*, *2*(4), e00043-17. https://doi.org/10.1128/mSystems.00043-17

Ross, A. A., Müller, K. M., Weese, J. S., & Neufeld, J. D. (2018). Comprehensive skin microbiome analysis reveals the uniqueness of human skin and evidence for phylosymbiosis within the class Mammalia. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(25), E5786–E5795. https://doi.org/10.1073/pnas.1801302115

Röthig, T., Ochsenkühn, M. A., Roik, A., Van Der Merwe, R., & Voolstra, C. R. (2016). Long-term salinity tolerance is accompanied by major restructuring of the coral bacterial microbiome. *Molecular Ecology*, *25*(6), 1308–1323. https://doi.org/10.1111/mec.13567

Saha, M., Ferguson, R. M. W., Dove, S., Künzel, S., Meichssner, R., Neulinger, S. C., Petersen, F. O., & Weinberger, F. (2020). Salinity and time can alter epibacterial communities of an invasive seaweed. *Frontiers in Microbiology*, *10*, 2870. https://doi.org/10.3389/fmicb.2019.02870

Schill, W. B., Iwanowicz, D., & Adams, C. (2017). Endozoicomonas dominates the gill and intestinal content microbiomes of *Mytilus edulis* from Barnegat Bay, New Jersey. *Journal of Shellfish Research*, *36*(2), 391–401. https://doi.org/10.2983/035.036.0212

Schlesner, H., Rensmann, C., Tindall, B. J., Gade, D., Rabus, R., Pfeiffer, S., & Hirsch, P. (2004). Taxonomic heterogeneity within the Planctomycetales as derived by DNA-DNA hybridization, description of *Rhodopirellula baltica* gen. nov., sp. nov., transfer of *Pirellula marina* to the genus Blastopirellula gen. nov. as *Blastopirellula marina* comb. nov. an. *International Journal of Systematic and Evolutionary Microbiology*, *54*(5), 1567–1580. https://doi.org/10.1099/ijs.0.63113-0.

Sharp, K. H., Pratte, Z. A., Kerwin, A. H., Rotjan, R. D., & Stewart, F. J. (2017). Season, but not symbiont state, drives microbiome structure in the temperate coral *Astrangia poculata*. *Microbiome*, *5*, 120. https://doi.org/10.1186/s40168-017-0329-8

Sommer, U., & Lengfellner, K. (2008). Climate change and the timing, magnitude, and composition of the phytoplankton spring bloom. *Global Change Biology*, *14*(6), 1199–1208. https://doi.org/10.1111/j.1365-2486.2008.01571.x

Song, S. J., Amir, A., Metcalf, J. L., Amato, K. R., Xu, Z. Z., Humphrey, G., & Knight, R.(2016). Preservation methods differ in fecal microbiome stability, affecting suitability for field studies. *Msystems*, *1*(3), e00021-16. https://doi.org/10.1128/mSystems.00021-16.Editor

Suikkanen, S., Pulina, S., Engström-Öst, J., Lehtiniemi, M., Lehtinen, S., & Brutemark, A. (2013). Climate change and eutrophication induced shifts in Northern summer plankton communities. *PLoS One*, *8*(6), e66475. https://doi.org/10.1371/journal.pone.0066475

Tap, J., Cools-Portier, S., Pavan, S., Druesne, A., Öhman, L., Törnblom, H., Simren, M., & Derrien, M. (2019). Effects of the long-term storage of human fecal microbiota samples collected in RNAlater. *Scientific Reports*, *9*, 601. https://doi.org/10.1038/s41598-018-36953-5

Teeling, H., Fuchs, B. M., Becher, D., Klockow, C., Gardebrecht, A., Bennke, C. M., Kassabgy, M., Huang, S., Mann, A. J., Waldmann, J., Weber, M., Klindworth, A., Otto, A., Lange, J., Bernhardt, J., Reinsch, C., Hecker, M., Peplies, J., Bockelmann, F. D., … Amann, R. (2012). Substrate-controlled succession of marine bacterio-plankton populations induced by a phytoplankton bloom. *Science*, *336*(6081), 608–611. https://doi.org/10.1126/science.1218344

Trevathan-Tackett, S. M., Sherman, C. D. H., Huggett, M. J., Campbell, A. H., Laverock, B., Hurtado-McCormick, V., Seymour, J. R., Firl, A., Messer, L. F., Ainsworth, T. D., Negandhi, K. L., Daffonchio, D., Egan, S., Engelen, A. H., Fusi, M., Thomas, T., Vann, L., Hernandez-Agreda, A., Gan, H. M., … Macreadie, P. I. (2019). A horizon scan of priorities for coastal marine microbiome research. *Nature Ecology and Evolution*, *3*(11), 1509–1520. https://doi.org/10.1038/s41559-019-0999-7

Troussellier, M., Escalas, A., Bouvier, T., Mouillot, D., & Murray, A. E. (2017). Sustaining rare marine microorganisms: Macroorganisms as repositories and dispersal agents of microbial diversity. *Frontiers in Microbiology*, *8*, 947. https://doi.org/10.3389/fmicb.2017.00947

van de Water, J. A. J. M., Voolstra, C. R., Rottier, C., Cocito, S., Peirano, A., Allemand, D., & Ferrier-Pagès, C. (2018). Seasonal stability in the microbiomes of temperate gorgonians and the red coral *Corallium rubrum* across the Mediterranean Sea. *Microbial Ecology*, *75*(1), 274–288. https://doi.org/10.1007/s00248-017-1006-y

Vandenberg, O., Dediste, A., Houf, K., Ibekwem, S., Souayah, H., Cadranel, S., Douat, N., Zissis, G., Butzler, J.-P., & Vandamme, P. (2004). Arcobacter species in humans. *Emerging Infectious Diseases*, *10*(10), 1863–1867. https://doi.org/10.3201/eid1010.040241

Vega Thurber, R., Willner-Hall, D., Rodriguez-Mueller, B., Desnues, C., Edwards, R. A., Angly, F., Dinsdale, E., Kelly, L., & Rohwer, F. (2009). Metagenomic analysis of stressed coral holobionts. *Environmental Microbiology*, *11*(8), 2148–2163. https://doi.org/10.1111/j.1462-2920.2009.01935.x

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S*. New York, NY: Springer.

Wemheuer, B., Güllert, S., Billerbeck, S., Giebel, H. A., Voget, S., Simon, M., & Daniel, R. (2014). Impact of a phytoplankton bloom on the diversity of the active bacterial community in the southern North Sea as revealed by metatranscriptomic approaches. *FEMS Microbiology Ecology*, *87*(2), 378–389. https://doi.org/10.1111/1574-6941.12230

Yang, C., Li, Y., Zhou, B., Zhou, Y., Zheng, W., Tian, Y., Van Nostrand, J. D., Wu, L., He, Z., Zhou, J., & Zheng, T. (2015). Illumina sequencing-based analysis of free-living bacterial community dynamics during an Akashiwo sanguine bloom in Xiamen sea, China. *Scientific Reports*, *5*, 8476. https://doi.org/10.1038/srep08476

Zeigler Allen, L., Allen, E. E., Badger, J. H., McCrow, J. P., Paulsen, I. T., Elbourne, L. D., Thiagarajan, M., Rusch, D. B., Nealson, K. H., Williamson, S. J., Venter, J. C., & Allen, A. E. (2012). Influence of nutrients and currents on the genomic composition of microbes across an upwelling mosaic. *ISME Journal*, *6*(7), 1403–1414. https://doi.org/10.1038/ismej.2011.201

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.
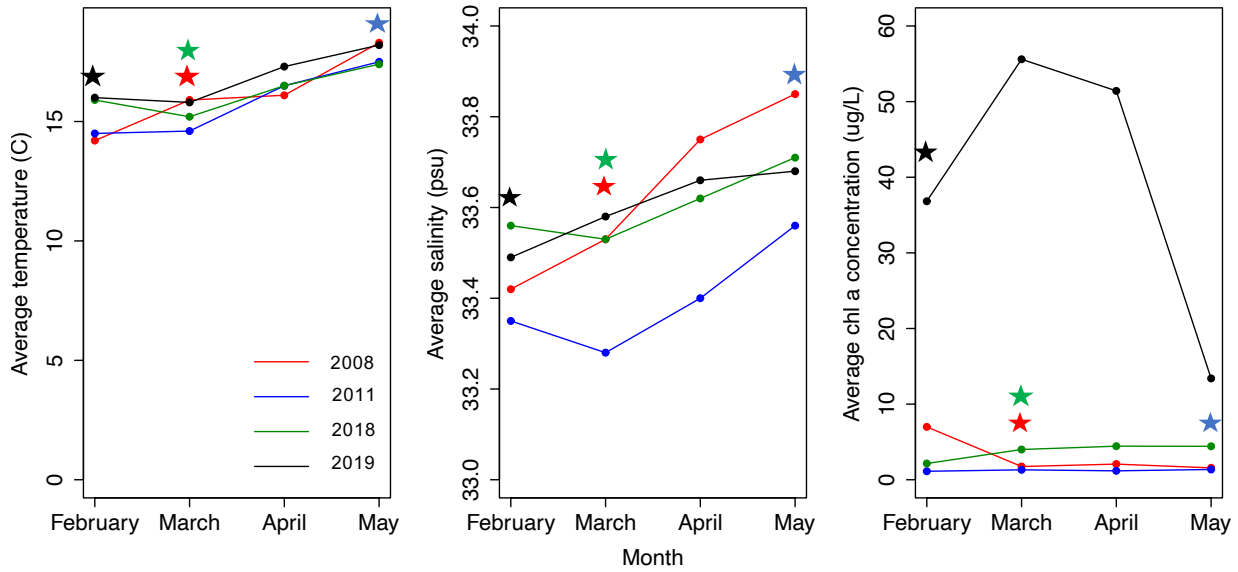
Supplemental Figure S1. Monthly average temperature (left), salinity (middle) and chlorophyll *a* concentration (right) at Ellen Browning Scripps Memorial Pier from February-May of each collecting year. Stars indicate months during which *D. gouldii* were collected and colors correspond to the legend in the left panel.
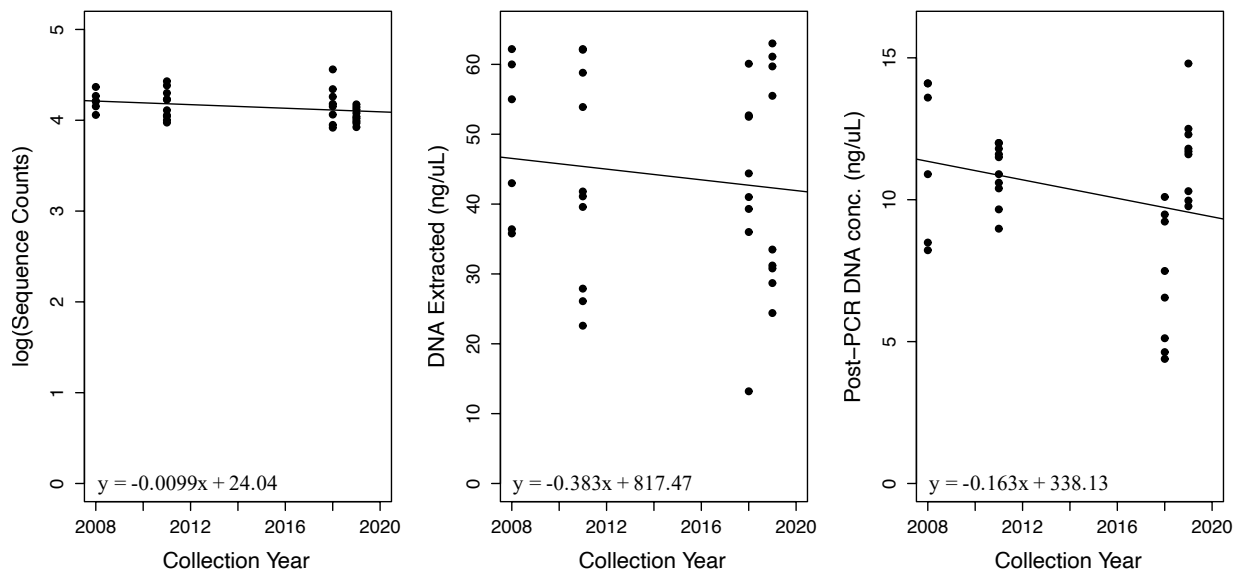
Supplemental Figure S2. Linear regression of the logarithm of raw sequence counts (left; $R^2 = 0.052; p = 0.107$) and concentration of extracted DNA (center; $R^2 = 0.015; p = 0.497$) and DNA concentration after 16S PCR (right; $R^2 = 0.051; p = 0.109$) over time. No discrete differences were found between sampling times for raw sequence counts ($X^2 = 6.006; p = 0.111$) or extracted DNA concentration ($X^2 = 0.853; p = 0.837$) via Kruskal-Wallis test. Post-PCR DNA concentration did show discrete differences between time points, with 2018 samples presenting the lowest average concentration ($X^2 = 14.188; p = 0.0027$).
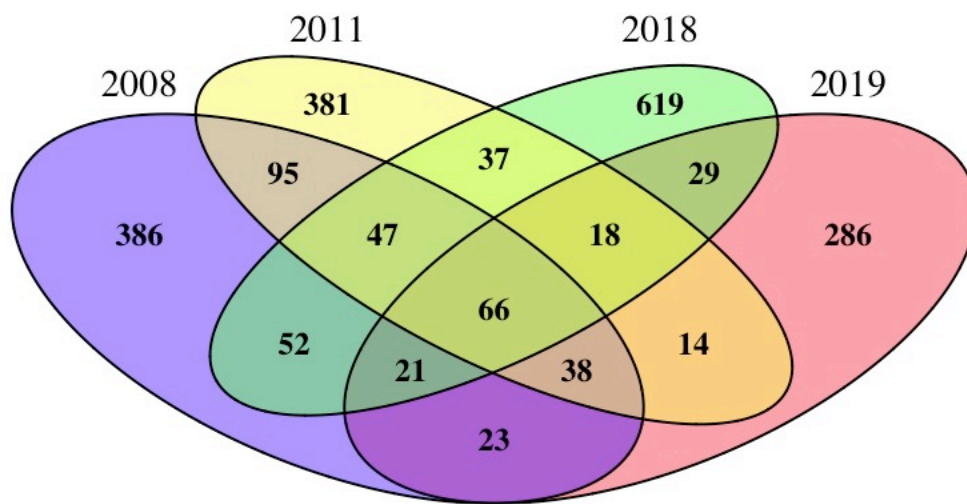
Supplemental Figure S3. Venn Diagram showing shared and unique ASVs between each sampling time point.
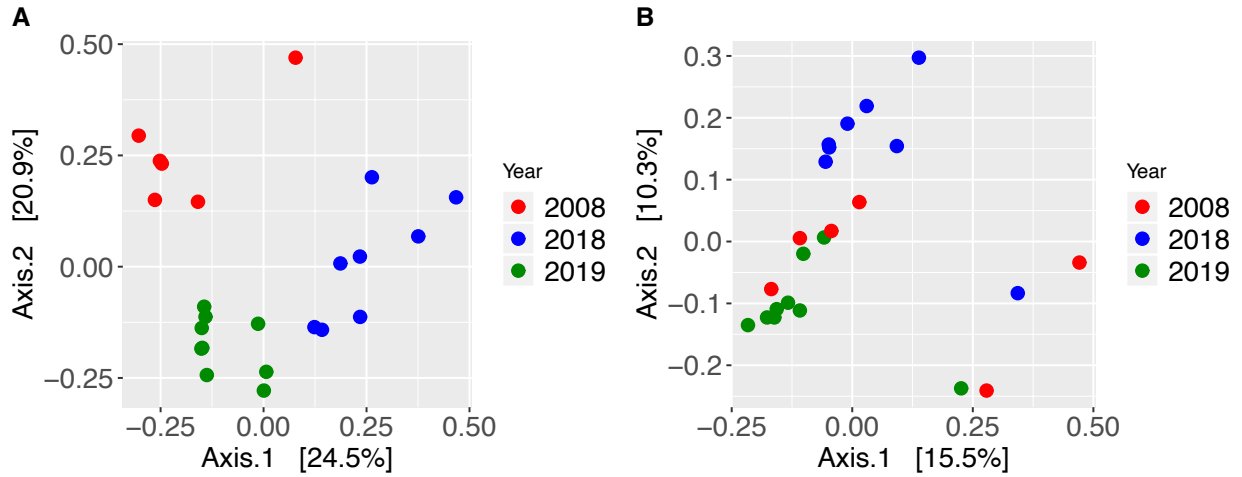
Supplemental Figure S4. Principal coordinate analysis plot of (A) Bray-Curtis dissimilarity (PERMANOVA: $R^2 = 0.334$, $p < 0.001$; betadisper: $p = 0.153$) and (B) unweighted UniFrac (PERMANOVA: $R^2 = 0.166$, $p < 0.001$; betadisper: $p = 0.200$), excluding 2011 samples, colored by collection year.
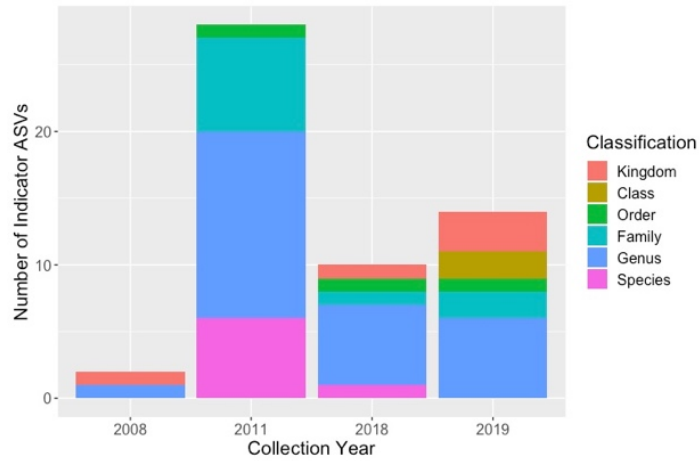
Supplemental Figure S5. Barplot of the number of indicator ASVs identified for each collection year (*indicator value ≥ 0.7, p ≤ 0.05*), colored by lowest taxonomic classification possible using the Silva database v138.
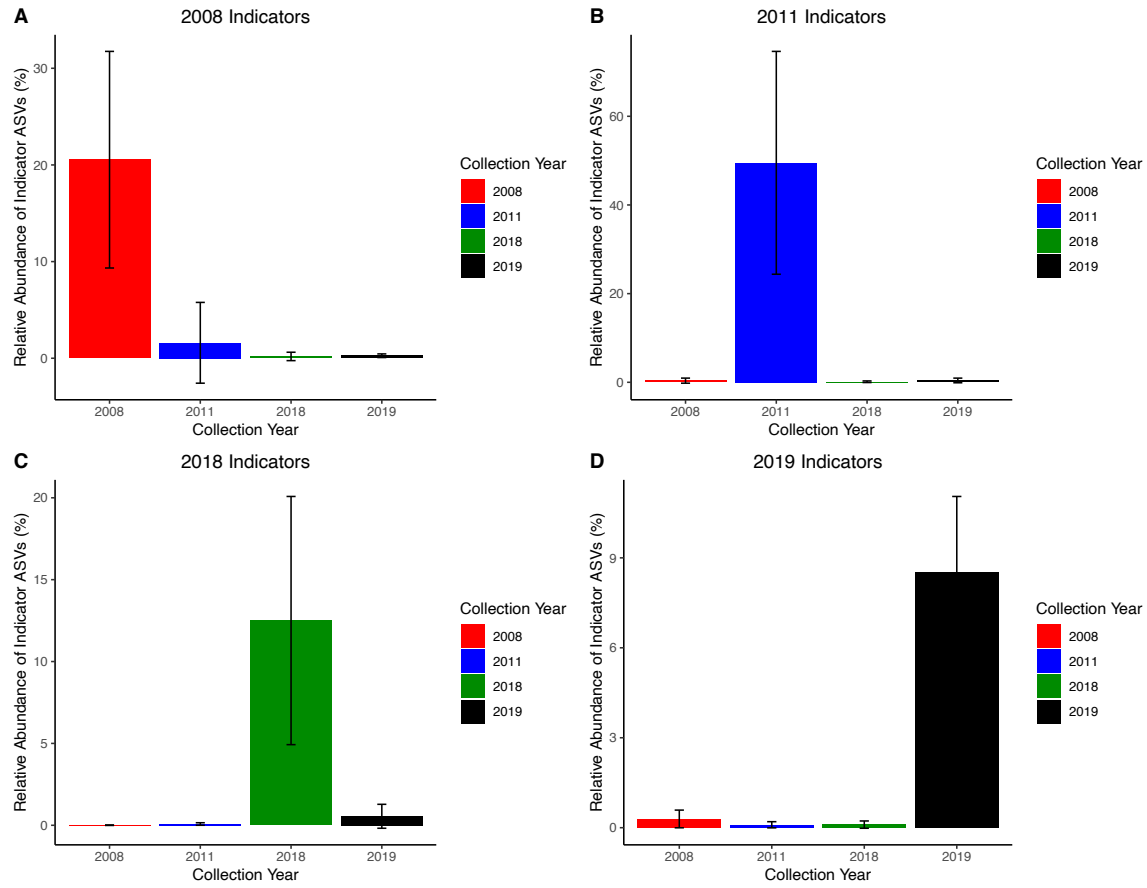
Supplemental Figure S6. The contributions of the indicator ASVs (mean ± s.d.) to *D. gouldii* microbiome composition by collection year.
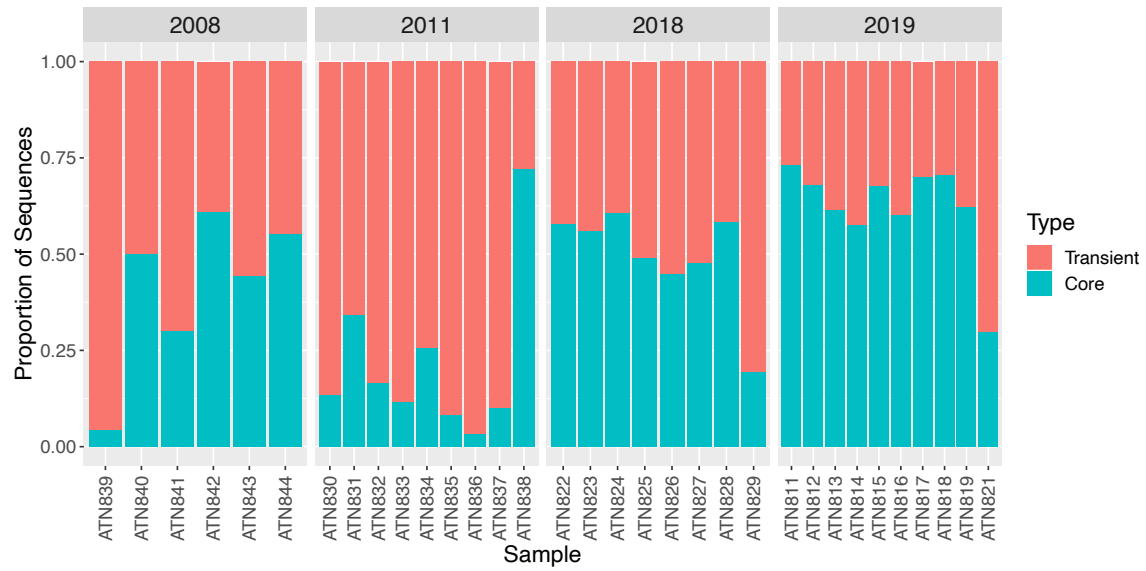
Supplemental Figure S7. Relative abundance of the six core ASVs, present in >90% of samples, in each sample, sorted by collection year.

Supplemental Table S1. Statistical results of linear regression models of alpha diversity metrics versus environmental data.

Year	F-statistic	R^2	p
Log(Observed) vs. SST	0.93	0.0021	0.34
Shannon vs. SST	0.0055	0.032	0.94
Log(Observed) vs. Salinity	0.45	0.018	0.51
Shannon vs. Salinity	0.55	0.014	0.46
Log(Observed) vs. Chl a	1.64	0.020	0.21
Shannon vs. Chl a	0.70	0.0096	0.41

Supplemental Table S2. Indicator ASVs from each collection year.

Year	ASV ID	Indicator Value	p-value	Kingdom	Phylum	Class	Order	Family	Genus	Species
2019	OTU1811	0.998	0.001	Bacteria	Proteobacteria	Gammaproteobacteria	Alteromonadales	Shewanellaceae	Shewanella	
	OTU2048	0.983	0.001	Bacteria	Firmicutes	Bacilli	Mycoplasmatales	Mycoplasmataceae	Mycoplasma	
	OTU914	0.958	0.001	Bacteria	Patescibacteria	Gracilibacteria	JGI_0000069-P22			
	OTU2054	0.933	0.001	Bacteria	Firmicutes	Bacilli	Mycoplasmatales	Mycoplasmataceae	Mycoplasma	
	OTU905	0.93	0.001	Bacteria	Cyanobacteria	Cyanobacteriia				
	OTU2047	0.928	0.002	Bacteria	Firmicutes	Bacilli	Mycoplasmatales	Mycoplasmataceae	Mycoplasma	
	OTU379	0.925	0.001	Bacteria	Bacteroidota	Bacteroidia				
	OTU2028	0.874	0.001	Bacteria	Spirochaetota	Spirochaetia	Spirochaetales	Spirochaetaceae	Spirochaeta_2	
	OTU74	0.816	0.001	Bacteria						
	OTU2025	0.801	0.003	Bacteria						
	OTU1371	0.745	0.003	Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Devosiaceae	Maritalea	
	OTU1280	0.741	0.001	Bacteria	Planctomycetota	Planctomycetes	Planctomycetales	Rubinisphaeraceae		
	OTU84	0.731	0.004	Bacteria						
	OTU1794	0.711	0.004	Bacteria	Proteobacteria	Gammaproteobacteria	Cellvibrionales	Halieaceae	Halioglobus	
2018	OTU54	0.952	0.001	Bacteria						
	OTU1512	0.923	0.001	Bacteria	Proteobacteria	Alphaproteobacteria	Rickettsiales	Anaplasmataceae	Neorickettsia	
	OTU2194	0.842	0.006	Bacteria	Verrucomicrobiota	Verrucomicrobiae	Verrucomicrobiales	Rubritaleaceae	Roseibacillus	
	OTU1261	0.839	0.001	Bacteria	Planctomycetota	Plancomycetes	Planctomycetales			
	OTU2221	0.793	0.002	Bacteria	Verrucomicrobiota	Verrucomicrobiae	Verrucomicrobiales	DEV007		
	OTU2248	0.791	0.001	Bacteria	Verrucomicrobiota	Verrucomicrobiae	Verrucomicrobiales	Rubritaleaceae	Rubritalea	
	OTU2232	0.726	0.008	Bacteria	Verrucomicrobiota	Verrucomicrobiae	Verrucomicrobiales	Rubritaleaceae	Persicirhabdus	
	OTU534	0.707	0.002	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	NS3a_marine_group	
	OTU1686	0.707	0.002	Bacteria	Bdellovibrionota	Oligoflexia	Silvanigrellales	Silvanigrellaceae	Silvanigrella	
	OTU2246	0.707	0.002	Bacteria	Verrucomicrobiota	Verrucomicrobiae	Verrucomicrobiales	Rubritaleaceae	Rubritalea	spongiae
2011	OTU829	0.949	0.001	Bacteria	Firmicutes	Clostridia	Peptostreptococcales-Tissierellales	Fusibacteraceae	Fusibacter	
	OTU831	0.949	0.001	Bacteria	Firmicutes	Clostridia	Peptostreptococcales-Tissierellales	Fusibacteraceae	Fusibacter	
	OTU1704	0.949	0.001	Bacteria	Campilobacterota	Campylobacteria	Campylobacterales	Arcobacteraceae	Halarcobacter	
	OTU1705	0.949	0.001	Bacteria	Campilobacterota	Campylobacteria	Campylobacterales	Arcobacteraceae		
	OTU1706	0.949	0.001	Bacteria	Campilobacterota	Campylobacteria	Campylobacterales	Arcobacteraceae	Poseideonibacter	lekithochrous
	OTU1809	0.949	0.001	Bacteria	Proteobacteria	Gammaproteobacteria	Alteromonadales	Shewanellaceae	Shewanella	
	OTU1810	0.949	0.001	Bacteria	Proteobacteria	Gammaproteobacteria	Alteromonadales	Shewanellaceae	Shewanella	
	OTU1915	0.949	0.001	Bacteria	Proteobacteria	Gammaproteobacteria	Oceanospirillales	Nitrincolaceae	Neptuniibacter	pectenicola
	OTU1776	0.944	0.001	Bacteria	Proteobacteria	Gammaproteobacteria	Alteromonadales	Colwelliaceae	Colwellia	aestuarii
	OTU1707	0.944	0.001	Bacteria	Campilobacterota	Campylobacteria	Campylobacterales	Arcobacteraceae	Poseidonibacter	
	OTU889	0.931	0.001	Bacteria	Fusobacteriota	Fusobacteriia	Fusobacteriales	Fusobacteriaceae	Psychrilyobacter	atlanticus
	OTU883	0.894	0.001	Bacteria	Fusobacteriota	Fusobacteriia	Fusobacteriales	Fusobacteriaceae	Ilyobacter	polytropus
	OTU1777	0.894	0.001	Bacteria	Proteobacteria	Gammaproteobacteria	Alteromonadales	Colwelliaceae	Colwellia	
	OTU1799	0.894	0.001	Bacteria	Proteobacteria	Gammaproteobacteria	Alteromonadales	Psychromonadaceae	Psychromonas	
	OTU1982	0.85	0.001	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales	Vibrionaceae	Photobacterium	damselae
	OTU1917	0.847	0.003	Bacteria	Proteobacteria	Gammaproteobacteria	Oceanospirillales	Marinomonadaceae	Marinomonas	
	OTU1711	0.837	0.002	Bacteria	Campilobacterota	Campylobacteria	Campylobacterales	Arcobacteraceae		
	OTU1910	0.837	0.003	Bacteria	Proteobacteria	Gammaproteobacteria	Oceanospirillales	Halomonadaceae	Cobetia	
	OTU1914	0.837	0.002	Bacteria	Proteobacteria	Gammaproteobacteria	Oceanospirillales	Halomonadaceae	Halomonas	
	OTU930	0.775	0.002	Bacteria	Verrucomicrobiota	Lentisphaeria	P.palmC41			
	OTU1919	0.775	0.001	Bacteria	Proteobacteria	Gammaproteobacteria	Oceanospirillales	Marinomonadaceae	Marinomonas	
	OTU1979	0.752	0.01	Bacteria	Proteobacteria	Gammaproteobacteria	Alteromonadales	Pseudoalteromonadaceae	Pseudoalteromonas	
	OTU1146	0.75	0.001	Bacteria	Planctomycetota	Planctomycetes	Pirellulales	Pirellulaceae		
	OTU371	0.707	0.005	Bacteria	Bacteroidota	Bacteroidia	Bacteroidales	Marinifilaceae		
	OTU474	0.707	0.004	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Aquibacter	
	OTU1452	0.707	0.005	Bacteria	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae		
	OTU1700	0.707	0.003	Bacteria	Campilobacterota	Campylobacteria	Campylobacterales	Arcobacteraceae		
	OTU1964	0.707	0.003	Bacteria	Proteobacteria	Gammaproteobacteria	Nitrosococcales	Methylophagaceae		
2008	OTU1895	0.957	0.002	Bacteria	Proteobacteria	Gammaproteobacteria	Oceanospirillales	Endozoicomonadaceae	Endozoicomonas	
	OTU2020	0.856	0.002	Bacteria						

Supplemental Table S3. Taxonomic classifications of the six core ASVs (occurring in >90% of samples).

ASV ID	Kingdom	Phylum	Class	Order	Family	Genus
OTU1728	Bacteria	Proteobacteria	Gammaproteobacteria	HglApr721		
OTU1726	Bacteria	Proteobacteria	Gammaproteobacteria			
OTU2021	Bacteria					
OTU1215	Bacteria	Planctomycetota	Planctomycetes	Pirellulales	Pirellulaceae	Blastopirellula
OTU1171	Bacteria	Planctomycetota	Planctomycetes	Pirellulales	Pirellulaceae	Rubripirellula
OTU1692	Bacteria	Campilobacterota	Campylobacteria	Campylobacterales	Helicobacteraceae	

Acknowledgements

CHAPTER 4

Microbiome divergence of marine gastropod species separated by the Isthmus of Panama

Abstract

The rise of the Isthmus of Panama ~3.5 mya separated populations of many marine organisms, which then diverged into new geminate sister species currently living in the Eastern Pacific Ocean and Caribbean Sea. However, we know very little about how such evolutionary divergences of host species have shaped their microbiomes. Here, we compared the microbiomes of whole-body and shell-surface samples of geminate species of marine gastropods in the genera *Cerithium* and *Cerithideopsis* to those of congeneric outgroups. Our results show that the effects of the Isthmus on microbiome composition varied among host genera and between sample types within the same hosts. In the whole-body samples, microbiome compositions of geminate species pairs in the focal genera tended to be similar, likely due to host filtering, although the strength of this relationship varied among the two groups and across similarity metrics. Shell-surface communities showed contrasting patterns, with co-divergence between the host taxa and a small number of microbial clades evident in *Cerithideopsis*, but not *Cerithium*. These results suggest that (i) the rise of the Isthmus of Panama affected microbiomes of geminate hosts in a complex and clade-specific manner and (ii) host-associated microbial taxa respond differently to vicariance events than the hosts themselves.

Keywords

Introduction

The formation of the Isthmus of Panama (IP) ~3.5 mya (O'Dea et al. 2016, Jaramillo 2018) led to a mixing of terrestrial faunas between North and South America, commonly known

as the Great American Biotic Interchange (Simpson 1980, Leigh et al. 2014). In the marine environment, however, the rise of the Isthmus separated populations of many species, which then evolved to form geminate species pairs (Jordan 1908), with one currently inhabiting the nutrient-rich upwelling zone of the Eastern Pacific Ocean and the other inhabiting the warmer, more oligotrophic Caribbean Sea (Lessios 2008). Geminate species pairs provide a powerful set of natural evolutionary experiments that have been used to investigate how closely related species evolve under different environmental conditions (Lessios 1990, Marko and Jackson 2001, Thacker 2017) and to better constrain rates of molecular evolution (e.g., Knowlton and Weigt 1998, Miura et al. 2010, O'Dea et al. 2016). However, it remains unclear how the microbiomes of geminate hosts have diverged in response to this vicariance event. Although there is a framework for examining how biogeographic isolation and environmental change may structure the microbiomes of geminate hosts (Wilkins et al. 2019), host-associated microbial divergence across the IP has so far only been studied using eggs of geminate species of echinoids (Carrier et al. 2020).

Tests of concordance between divergence patterns of microbiomes and those of their hosts rely on correlations between the phylogenetic relatedness of host taxa and the similarity of their microbiomes, a pattern termed "phylosymbiosis" (Brooks et al. 2016, Lim and Bordenstein 2020). Under this model, geminate taxa are expected to harbor microbial communities that are significantly more similar to each other than to more distantly related hosts currently living in the same environment. This pattern may arise either as the result of co-phylogenetic divergence between hosts and their associated microbes, host filtering, or both (Lim and Bordenstein 2020). In the first case, geminate hosts and their microbiomes (or a subset thereof) are expected to diverge in concert with one another after the closure of the IP, resulting in co-phylogenetic

relationships between the geminate host clades and associated microbial taxa. In the second case, aspects of the host environment such as morphology, immune function and diet (e.g., nutrient and metabolite availability) may exclude certain microbial taxa while providing highly suitable conditions for others (Mazel et al. 2018). Since closely related hosts, such as geminates, often share many traits, they would be predicted to select for more compositionally similar microbial communities. While they are distinct processes, co-speciation and host filtering are not mutually exclusive pathways to phylosymbiosis, and have been shown to act in concert to generate concordance between host evolutionary history and microbiome composition in some host groups, such as corals (Pollock et al. 2018).

Here we use geminate species from two genera of marine gastropods, *Cerithium* and *Cerithideopsis,* along with congeneric outgroups in each case, to test for phylosymbiosis resulting from vicariance due to the rise of the IP. Among our focal host species, *Cerithium lutosum* and *Cerithium stercusmuscarum* form a well-supported geminate species pair, with *Cerithium atratum* as an outgroup (Miura et al. 2010; Fig. 1). *Cerithideopsis mazatlanica* and *Cerithideopsis valida* diverged in the Eastern Pacific after the rise of the IP and are a geminate clade to *Cerithideopsis pliculosa* in the Caribbean, with the Eastern Pacific *Cerithideopsis montagnei* as a congeneric outgroup (Miura et al. 2010; Fig. 1). Both outgroups were sampled in the same geographic locations as our geminate taxa (Fig. 1). The *Cerithium* species used here are typically found in near-shore environments with close proximity to mangroves. *Cerithium stercusmuscarum* largely occupies rocky shores in the Eastern Pacific, while its geminate *Cerithium lutosum* and the outgroup *Cerithium atratum* are found in coral rubble or sandy flats in the Caribbean. The *Cerithideopsis* species sampled here all occupy shallow, muddy environments and associated mangrove habitats (Miura et al. 2010). Though all of these species

91

forage in the mud, *Cerithideopsis montagnei* differs from its congeners by climbing mangrove roots during high tides.

For each host, we used soft body tissues (hereafter "whole-body") as well as swabs of the shell-surfaces of multiple individuals (Table S1) to test the hypothesis that geminate species pairs would harbor microbiomes that are both compositionally and phylogenetically more similar to one another than to their sympatric congeners.

Methods

Sample collection

We sampled multiple individuals of each of seven different marine gastropod species from our focal genera at four sites across the IP (Figure 1, Table S1). We swabbed the shell-surface of each individual in the field using a sterile 152mm swab (Grenier Bio-One, Frickenhausen, Germany), avoiding any visible patches of mud. At each site, we also sampled the top 3 cm of sediment in a sterile 50 mL conical tube and collected two replicate seawater samples by filtering 500mL of surface seawater through 0.22μm Sterivex filters (MilliporeSigma, Burlington, MA, USA). Samples were transported on ice to the Naos Marine Laboratory at the Smithsonian Tropical Research Institute (STRI), where they were kept at -20°C. Subsequently, the whole-body of each individual was removed from the shell, triple-rinsed with 95% EtOH and preserved at -20°C. Samples were later shipped to the University of California, San Diego for DNA extraction and sequencing.

DNA extraction, amplification and sequencing

We used the Qiagen Blood and Tissue kit (Qiagen GmbH, Hilden, Germany) to extract DNA from our samples, following the manufacturer's instructions. We used the protocol of (Neu

et al. 2021) to amplify and purify the V4 region of the 16S rRNA gene. The resulting amplicons were sequenced on the MiSeq platform (Illumina, 2x250bp output) at the DNA Technologies Core, University of California Davis. In order to verify host identities, we also amplified the cytochrome oxidase *c* subunit I (COI) gene from one individual of each host species using the PCR protocol of (Miura et al. 2010). The PCR products were cleaned using an Exo-SAP approach and Sanger sequenced at Eton Bioscience, Inc. (San Diego, CA, USA).

Sequence processing and statistical analyses

We used 'DADA2' v1.18.0 (Callahan et al. 2016) in R v4.0.3 (R Core Team 2013) to process the 16S rRNA gene sequences. The resulting amplicon sequence variants (ASVs) were assigned taxonomy using the Silva database v138 (Quast et al. 2013) and all sequences not matching Bacteria or Archaea, as well as any chloroplasts and mitochondria, were removed. We aligned the remaining sequences using MAFFT v7.310 (Katoh and Standley 2013) and constructed microbial phylogenies for further analyses with FastTree v2.1.11 (Price et al. 2010) using the GTR + CAT model and midpoint rooting.

For each host genus, we tested for differences in alpha diversity among the hosts using two sets of metrics. First, we used the entire, unrarefied microbial dataset to estimate Hill numbers with q of 0 and 1 for each sample using the package hilldiv v1.5.1 (Alberdi and Gilbert 2019a). Using Hill numbers with $q = 0$ does not account for relative abundance and is equivalent to the number of observed ASVs, while $q = 1$ weights each ASV by its relative abundance and is equivalent to the exponential of Shannon's index (H') (Alberdi and Gilbert 2019b). Second, we generated 100 ASV tables, each rarefied to 3,000 sequences per sample, using the package metagMisc v0.0.4 (Mikryukov 2018). We then calculated the number of observed ASVs and Shannon's H' for each sample from each ASV table and used the average values to account for

93

any differences in individual random draws. In both cases, we tested for differences between host species using Kruskal-Wallis tests and further investigated significant differences using Dunn's tests in the package dunn.test v1.3.5 (Dinno 2017).

All subsequent analyses used both the unrarefied dataset, normalized via total sum scaling (TSS) (McKnight et al. 2019), as well as the 100 rarefied datasets (3,000 sequences/sample), unless otherwise specified. We tested for differences in microbiome composition between host species using permutational multivariate analyses of variance (PERMANOVAs) in the package phyloseq v1.34.0 (McMurdie and Holmes 2013) with Bray-Curtis Dissimilarity (BCD) and unweighted UniFrac Distance (Lozupone et al. 2011) as our beta diversity metrics. We use both of these metrics in our analyses as they emphasize different attributes; BCD includes relative abundance information, while unweighted UniFrac includes a measure of phylogenetic relatedness of the sampled microbial taxa. We also conducted PERMANOVAs on the TSS dataset to determine whether the host-associated microbiomes were significantly different from the microbial communities in the sediment and water samples. We tested for homogeneity of dispersions between species and sample types using the betadisper function in vegan v2.5-7 (Oksanen et al. 2020) for all iterations.

To investigate the role of host phylogenetic distance on microbiome composition more directly, we pooled together all microbial sequences from each individual host species. Since *Cerithium stercusmuscarum* was collected from two sites, we pooled all sequences from this species together but also pooled each site individually. We then clustered the samples using unweighted paired group with arithmetic means (UPGMA), using both BCD and unweighted UniFrac, and the TSS and rarefied datasets. To test the robustness of clustering, we calculated Dunn's Index for each node using the package clValid v0.7 (Brock et al. 2008). We used the

normalized Robinson-Foulds metric (nRF) in the package phangorn v2.5.5 (Schliep 2011) to test

for congruency between the UPGMA plots and the host phylogeny generated from COI

sequences using FastTree v2.1.11 (Price et al. 2010) with the GTR + γ model. For some samples,

relationships varied across different rarefied datasets, so we generated boxplots showing the

BCD and unweighted UniFrac distance between species across all 100 rarefied datasets.

To identify microbial clades contributing to similarities between geminate host species, we

generated network plots using sparse partial least squares discriminant analysis (SPLS-DA) in

the package mixOmics v6.14.0 (Rohart et al. 2017) at taxonomic levels ranging from ASV to

class. This analysis requires a log-ratio transformation and was only conducted on the unrarefied

dataset using a centered log-ratio transformation. Microbial taxa that were positively correlated

with the geminate species (r > 0.6) or negatively correlated with the congeneric outgroups (r < -

0.6) were considered to contribute to phylosymbiosis resulting from the rise of the IP (Demko

2021). To further investigate whether the taxa contributing to the differentiation between the

geminate clades and outgroups also demonstrated potential co-phylogenetic relationships with

the hosts, we used the parafit function in ape v5.4.1 (Paradis and Schliep 2019).

Results

<u>Alpha Diversity</u>

 Cerithium: The whole-body microbiomes of all three species of *Cerithium* showed

similar Hill numbers at q = 0 (Fig. 2a) and observed ASVs after rarefying (Fig. S1a), and there

were no significant differences in total ASV richness between the geminate and outgroup species

(Tables S2, S3). However, at q = 1, the Hill numbers of *C. atratum* were significantly higher

than those of the geminate pair *C. lutosum* and *C. stercusmuscarum* (Fig. 2c, Table S2), as

confirmed by Shannon's H' values in the rarefied dataset (Fig. S2a, Table S3). In the *Cerithium* shell-surface microbiome, no significant differences were found between geminate and outgroup species using any of the alpha diversity metrics (Figs. 2b, 2d, S1b, S2b, Tables S2, S3).

Cerithideopsis: In this group, whole-body samples showed significant differences in Hill numbers at q = 0 (Fig. 2e; Table S2) and marginally significant differences in observed ASV richness (Fig. S1c, Table S3). This was driven by the low diversity of *C. montagnei*, which was significantly different from *C. mazatlanica* and *C. pliculosa* (Dunn's tests: *p < 0.05*), but not *C. valida*. Interestingly, the geminate and outgroup species did not show significant differentiation at q = 1 (Fig. 2f; Table S2) or using Shannon's index (Fig. S2c, Table S3), largely due to the presence of a single *C. montagnei* sample with high diversity (Fig. 2e, Dataset S1). *C. montagnei* showed significantly lower diversity than *C. mazatlanica* and *C. pliculosa* (Dunn's test: *p = 0.0084* and *p = 0.005*, respectively), but not *C. valida*, when this high diversity individual was removed. In the *Cerithideopsis* shell-surface swabs, Hill numbers at q = 0 did not show significant differences between the host species (Fig. 2g; Table S2), again due to a single high diversity sample of *C. montagnei* (Fig. 2g, Dataset S1), though the rarefied observed ASV count showed marginally significant differentiation (Fig. S1d, Table S3). Removing the high diversity sample from the Hill number analyses resulted in significant differences between *C. montagnei* and the higher diversity *C. pliculosa* and *C. mazatlanica* (Dunn's test: *p = 0.0129* and *p = 0.0135*, respectively), but not *C. valida*. When q = 1, significant differentiation (Table S2) was driven by the difference between *C. pliculosa* and *C. montagnei* (Dunn's test: *p = 0.0244*), as *C. mazatlanica* and *C. valida* were not significantly different from either of these hosts (Fig. 2h). These results were confirmed by Shannon's H' from the rarefied dataset (Fig. S2d, Table S3).

Taxonomic Composition

Cerithium: Whole-body samples of individual host species were largely comprised of Actinobacteriota, Alphaproteobacteria, Cyanobacteria and Gammaproteobacteria (Fig. 3a). Interestingly, the most common group found in the geminate species *C. lutosum* and *C. stercusmuscarum* was "Unclassified Bacteria", as determined by the Silva database. A single ASV within this group (ASV10677) comprised, on average, greater than 20% of the relative abundance of the microbiome in these hosts. The closest named relative to this microbe is a member of the phylum Chlorobi, *Chloroherpeton thalassium* (81.89% sequence similarity using NCBI BLASTn). The two populations of *C. stercusmuscarum* whole-body samples were differentiated by the relative abundances of the phyla Crenarchaeota, which was higher at Bique, and Cyanobacteria, which was higher at Punta Culebra (Fig. 3a). *Cerithium* shell-surface microbiomes were comprised mainly of Alphaproteobacteria, Bacteroidota, Cyanobacteria, Gammaproteobacteria and Planctomycetota (Fig. 3b).

Cerithideopsis: Whole-body samples of these host species contained high relative abundances of Alphaproteobacteria, Bacteroidota and Gammaproteobacteria. However, *C. montagnei* showed higher abundance of Actinobacteriota while the geminate clade including *C. mazatlanica*, *C. valida* and *C. pliculosa* had increased relative abundances of Planctomycetota and Verrucomicrobiota (Fig. 3c). *Cerithideopsis* shell-surface microbiomes showed high relative abundances of Alphaproteobacteria, Bacteroidota, Gammaproteobacteria and Planctomycetota (Fig. 3d). However, *C. montagnei* shell-surfaces showed higher levels of Actinobacteriota and a notable decrease in Cyanobacteria compared to the geminate species (Fig. 3d).

Microbial community compositions of sediment and water samples varied among the collection sites, but nearly all were dominated by members of the class Gammaproteobacteria and were compositionally distinct from the microbiomes of all host species (Fig. 3e).

Host evolutionary history and microbiome composition

Results of PERMANOVAs show that in both host genera, the microbiome compositions of individual host species are distinct, regardless of tissue type, normalization method or beta diversity metric used (Tables S4, S5), and significantly different from the environmental samples (Table S6).

Cerithium: UPGMA clustering of whole-body samples was consistent with the phylosymbiosis hypothesis when using BCD and either normalization method, with microbiomes of geminate species *C. lutosum* and *C. stercusmuscarum* being compositionally more similar to each other than to *C. atratum* (Figs. 4a, S3a). However, when *C. stercusmuscarum* individuals were separated by collection site, *C. stercusmuscarum* samples from Punta Culebra were more similar to *C. lutosum* than to the *C. stercusmuscarum* samples from Bique (Fig 5a). Analyses using unweighted UniFrac showed a different pattern, with *C. lutosum* microbiomes clustering closest to those from *C. atratum* with the *C. stercusmuscarum* populations as outgroups (Figs. 4a, 5a, S4a). UPGMA clustering of shell-surface samples from *Cerithium* did not match the host phylogeny, regardless of distance metric or normalization method (Figs. 4b, 5b, S3b, S4b). Instead, microbiome composition appeared to be determined by the local environment, as the Caribbean species *C. lutosum* and *C. atratum* were more similar to each other than to either population of the Eastern Pacific *C. stercusmuscarum*.

Cerithideopsis: When using BCD and the TSS dataset, UPGMA clustering of host microbiome compositions mirrored the host phylogeny (Fig. 4c). However, BCD of samples rarefied to 3,000 sequences produced mixed results depending on which of the 100 resampled ASV tables was used (Fig. 4c). In 71% of the datasets, results were consistent with the hypothesis of phylosymbiosis, but in the other 29% of cases *C. valida* microbiomes were shown

to be more similar to those of *C. montagnei* than to those of its sister species, *C. mazatlanica*, or the geminate *C. pliculosa*. This appears to be driven by the compositional dissimilarity between individuals of *C. valida*, as the BCD between this species and other *Cerithideopsis* encompass a large range of values (Fig. S3c). Clustering the TSS dataset using unweighted UniFrac distance does not support the phylosymbiosis hypothesis. However, the rarefied OTU tables once again presented split patterns, with 86% of the results being consistent with predicted phylosymbiosis (Fig. 4c, S4c). In contrast to the whole-body samples, shell-surface microbiomes of the *Cerithideopsis* geminate clade consistently clustered together, with *C. montagnei* as the outgroup, regardless of the distance metric or normalization method used (Figs. 4d, S3d, S4d).

Unique microbial associates of geminate hosts

Cerithium: Only one ASV (ASV10677) significantly differentiated whole-body samples of *C. lutosum* and *C. stercusmuscarum* from *C. atratum* (Fig. S5). As discussed above, this ASV was the most dominant member of the *C. lutosum* and *C. stercusmuscarum* communities but was not highly abundant in the *C. atratum* microbiome (Fig. S5). In the shell-surface communities, no individual taxa could significantly differentiate the geminate pair from *C. atratum*.

Cerithideopsis: These samples contained several microbial taxa that significantly differentiated the geminate clade of *C. mazatlanica*, *C. valida* and *C. pliculosa* from the outgroup, *C. montagnei*. In the whole-body, two genera and four families were significant drivers of this differentiation (Fig. S6, Table S7), while shell-surface swabs contained four genera, five families, four orders and two classes of microbes which separated the geminates from *C. montagnei* (Fig. S7, Table S7).

Host-microbe co-phylogenetic relationships

Cerithium: We were unable to test for potential co-phylogenetic relationships between *Cerithium* samples and the microbial clades driving differentiation between geminates and the outgroup, as only a single ASV was responsible for this pattern in the whole-body samples, and none differed significantly between the geminate and outgroup shell-surface samples (Table S7).

Cerithideopsis: Whole-body samples of this genus did not show any co-phylogenetic relationships across multiple levels of microbial taxonomy (Table S7). In the shell-surface swabs, however, two families (A4b and Thermoanaerobaculaceae), two orders (SBR1031 and Thermoanaerobaculales) and the class Gammaproteobacteria all showed significant co-phylogenetic relationships with the host (Table S7).

Discussion

Whole-body and shell-surface samples of tropical marine gastropod hosts from the genera *Cerithium* and *Cerithideopsis* house diverse, complex and host-specific microbial communities. The richness of ASVs was surprisingly similar across host genera in both the whole-body and shell-surface samples, with the exception of relatively low diversity in *Cerithideopsis montagnei*. This may reflect the unique ecology of *C. montagnei*, which forages in the mud like the other *Cerithideopsis* species, but climbs mangroves during high tide, potentially reducing its exposure to taxa present in the seawater that surrounds its congeners, resulting in reduced diversity. Differences in microbial diversity between the geminate taxa and their congeneric outgroups are most evident using diversity metrics that account for relative abundances. The geminate pair *Cerithium lutosum* and *C. stercusmuscarum* had significantly lower Hill numbers at q = 1 and Shannon's H' compared to their congeneric outgroup, *C. atratum*. These metrics suggest that fewer taxa dominate the microbiomes of these geminate

species as compared to *C. atratum*. Within *Cerithideopsis*, the trend is reversed, with *C. montagnei* harboring reduced diversity compared to its geminate congeners using these metrics.

Compositionally, the whole-body and shell-surface microbiomes of *Cerithium* and *Cerithideopsis* species were distinct from one another but were comprised of taxa commonly associated with marine invertebrate hosts. Many of the dominant microbial clades in the whole-body samples, including Alphaproteobacteria, Gammaproteobacteria and Planctomycetota, have been found previously in marine molluscan tissues (Neu et al. 2019, 2021). The shell-surfaces of *Cerithium* and *Cerithideopsis* species also contained a number of clades previously found on shells of other marine invertebrates, including Alphaproteobacteria, Bacteroidota and Cyanobacteria (Pfister et al. 2010, Arfken et al. 2017). In contrast, our environmental samples were largely dominated by Gammaproteobacteria, which are commonly found in seawater and intertidal sediment communities (Wang et al. 2012, Sunagawa et al. 2015). Thus the microbiome compositions of all of our host species are distinct from that of the ambient environment, as is common in marine invertebrates (e.g., Chu and Vollmer 2016, Neu et al. 2019).

Similarities of microbiomes of geminate species varied considerably depending on the host genus, the type of sample and the metric used. *Cerithium* whole-body samples conformed to a phylosymbiosis pattern when using BCD, as *C. lutosum* was more similar to its geminate *C. stercusmuscarum* than to *C. atratum* samples from the same environment. However, when using unweighted UniFrac, *C. lutosum* clustered more closely with *C. atratum* than *C. stercusmuscarum*. These results suggest that the relative abundances of certain microbial taxa are more similar between the geminate species than the congeneric outgroup, while microbial taxa occurring on the same side of the Isthmus tend to be closer evolutionarily. A single ASV, ASV10677, was primarily responsible for the similarity between the geminates *Cerithium*

lutosum and *C. stercusmuscarum* observed in the BCD dataset. Although this ASV was present in all three *Cerithium* hosts, its relative abundance was nearly an order of magnitude higher in the geminates when compared to *C. atratum*. ASV10677 was also found in all of our sediment and water samples, but at <1% relative abundance, suggesting selective enrichment in *C. lutosum* and *C. stercusmuscarum*. The reason for this is unclear, though the closest named relative to this taxon is *Chloroherpeton thalassium*, a green sulfur bacterium (Gibson et al. 1984), suggesting a potential role in sulfur cycling. The *Cerithium* geminate species are found associated with mangrove sediments, where sulfur cycling occurs (Varon-Lopez et al. 2014), while *C. atratum* is often found further out on the reef flat.

When the two sampled populations of *Cerithium stercusmuscarum* were analyzed separately, the BCD dataset showed a different pattern. Though the geminate species were separated from the congeneric outgroup, *C. atratum*, *C. stercusmuscarum* populations did not cluster together, potentially due to their positions on opposite sides of the Panama Canal (Fig. 1). This shows that intraspecific, site-level differences in microbiome composition can be greater than the differences observed between species and should be accounted for in future studies.

Shell-surface microbiomes of *Cerithium* did not show any significant geminate signal, regardless of the metric used. Thus, the ambient environment may have a greater impact on shell-surface microbiome composition in this group than evolutionary history, especially as *C. atratum* and *C. lutosum* were collected from the same site.

In the whole-body samples of *Cerithideopsis*, compositional similarities of geminate hosts varied depending on the method used. In the TSS dataset, geminate species clustered together if BCD was used, but not unweighted UniFrac, a result similar to that in *Cerithium*. In our rarefied datasets, the outcomes were even more variable, with the geminate clade clustering

102

together in 71% and 86% of random draws using the BCD and unweighted UniFrac metrics, respectively. This highlights two important points. First, although rarefaction is commonly used, and often recommended (Weiss et al. 2017, McKnight et al. 2019), for standardization of microbiome data, our results show that in some cases, hypothesis testing using rarefied datasets can produce different results than other normalization approaches. Second, using a single rarefied ASV table to test hypotheses, as is common in microbial ecology, can be problematic given the stochastic nature of individual rarefaction draws. Instead, we recommend using multiple iterations to test the robustness of results using rarefied data. Finally, our network analyses identified multiple taxa that helped differentiate the geminate clade from the outgroup, *C. montagnei*, but none of these showed significant co-phylogenetic relationships with the host group. Taken together, our results suggest that in the *Cerithideopsis* whole-body samples, host filtering and/or host ecology likely drive any observed similarity between the microbiomes of the geminate species. As stated above, the climbing behavior of *C. montagnei* may limit its exposure to seawater microbes thus differentiating the microbiome composition of this species from its entirely mud-dwelling relatives.

In contrast to the whole-body samples, shell-surface microbiomes of *Cerithideopsis* species consistently mirrored the host phylogeny. Several microbial clades were responsible for this differentiation between the geminate species and *C. montagnei*, some of which showed co-phylogenetic relationships with these hosts. These include members of the microbial orders Thermoanaerobaculales and SBR1031, and the class Gammaproteobacteria. Thermoanaerobaculales is an order within the phylum Acidobacteriota, recently described from freshwater hot springs (Dedysh and Yilmaz 2018). Little is currently known about this clade, but members have been found in high abundance in intertidal sands (Taylor and Kurtz, Jr. 2020) and

are present in many of our seawater and sediment samples, suggesting that these microbes may be transferred from the environment to shell-surfaces. SBR1031 is an order of nitrifying bacteria within the phylum Chloroflexi previously found in association with marine invertebrates (Cleary et al. 2013, Morrow et al. 2016) and the mangrove rhizosphere (Xie et al. 2021). Since shell-surfaces are sites of active nitrogen cycling and nitrification in the marine environment (Pfister et al. 2010, Arfken et al. 2017), and members of this order were present in our sediment and seawater samples, this association may also constitute an acquisition from the environment. Gammaproteobacteria is a highly diverse clade common in the marine environment and many marine hosts (Sunagawa et al. 2015, Chiarello et al. 2020), so in this case the processes driving a co-phylogenetic pattern with *Cerithideopsis* are unclear. In general, it is surprising that significant co-phylogenetic divergence between geminate hosts and their microbiomes was only evident in shell-surface biofilms of *Cerithideopsis*, as these communities are expected to be more influenced by the external environment than the biology of host (Mazel et al. 2018, Woodhams et al. 2020).

While the formation of the IP led to predictable patterns of evolutionary divergence of marine animal species on either side, our results show that this geologic event affected host-associated microbiomes in a much more complex and context-dependent manner. Depending on the geminate clade and host tissue sampled, we found evidence for (i) phylosymbiosis driven by environmental filtering, (ii) co-phylogenetic divergence of hosts and their microbial associates and (iii) no discernable impact of host evolutionary history on microbiomes of geminate species. These results could reflect multiple causes. First, an expectation of co-divergence between hosts and their microbiomes due to the rise of the IP assumes that populations of individual microbial taxa were originally present in both oceans and subsequently split by the closure of the IP,

leaving representatives on both sides. However, microbiome compositions of marine invertebrate hosts commonly differ substantially between populations of the same host species (e.g., Cleary et al. 2013, Neu et al. 2019) as shown by our samples of *Cerithium stercusmuscarum*. If populations of hosts did not share many microbes prior to being separated by the IP, it is perhaps unsurprising that observed phylosymbiosis patterns are driven by a small number of taxa, rather than being a community-wide trend. Second, for co-phylogenetic divergence of hosts and their microbes to be detectable millions of years after a vicariant event, hosts and microbial taxa must diverge at similar rates. However, due to their short generation times and differences in selective pressures between the Caribbean and Eastern Pacific after the closure of the IP, the divergence rates of microbial taxa in separate oceans were likely much higher than those of their hosts. Such differences in evolutionary rates can decrease the phylosymbiosis signal and create closer evolutionary relationships between microbial taxa currently present on the same side of the IP, as we found in multiple cases. Unfortunately, current data limitations make this difficult to test, as ancestral distributions of microbial taxa remain unknown and calibrating microbial phylogenies also remains challenging (Louca et al. 2018).

Overall, our results suggest that ecological and evolutionary responses of host-associated microbial communities to major vicariant events such as the rise of the Isthmus of Panama, are complex and taxon-specific. Future comparative analyses of microbiomes of other geminate hosts would be useful for better understanding how the microbiome compositions of marine hosts have been shaped by major geological events.

References

Alberdi, A., and M. T. P. Gilbert. 2019a. Hilldiv: An R package for the integral analysis of diversity based on Hill numbers. bioRxiv.

Alberdi, A., and M. T. P. Gilbert. 2019b. A guide to the application of Hill numbers to DNA-based diversity analyses. Molecular Ecology Resources 19:804–817.

Arfken, A., B. Song, J. S. Bowman, and M. Piehler. 2017. Denitrification potential of the eastern oyster microbiome using a 16S rRNA gene based metabolic inference approach. PLoS ONE 12:e0185071.

Brock, G., V. Pihur, S. Datta, and D. S. 2008. clValid: An R package for cluster validation. Journal of Statistical Software 25.

Brooks, A. W., K. D. Kohl, R. M. Brucker, E. J. van Opstal, and S. R. Bordenstein. 2016. Phylosymbiosis: Relationships and Functional Effects of Microbial Communities across Host Evolutionary History. PLoS Biology 14:e2000225.

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

Carrier, T. J., H. A. Lessios, and A. M. Reitzel. 2020. Eggs of echinoids separated by the Isthmus of Panama harbor divergent microbiota. Marine Ecology Progress Series 648:169–177.

Chiarello, M., J. C. Auguet, N. A. J. Graham, T. Claverie, E. Sucré, C. Bouvier, F. Rieuvilleneuve, C. X. Restrepo-Ortiz, Y. Bettarel, S. Villéger, and T. Bouvier. 2020. Exceptional but vulnerable microbial diversity in coral reef animal surface microbiomes. Proceedings of the Royal Society B: Biological Sciences 287:20200642.

Chu, N. D., and S. V. Vollmer. 2016. Caribbean corals house shared and host-specific microbial symbionts over time and space. Environmental Microbiology Reports 8:493–500.

Cleary, D. F. R., L. E. Becking, N. J. de Voogd, A. C. C. Pires, A. R. M. Polónia, C. Egas, and N. C. M. Gomes. 2013. Habitat- and host-related variation in sponge bacterial symbiont communities in Indonesian waters. FEMS Microbiology Ecology 85:465–482.

Dedysh, S. N., and P. Yilmaz. 2018. Refining the taxonomic structure of the phylum Acidobacteria. International Journal of Systematic and Evolutionary Microbiology 68:3796–3806.

Demko, A. M. 2021. Chemical ecology of marine microbial communities: An assessment of bacterial diversity and dynamics in tropical marine sediments. UC San Diego.

Dinno, A. 2017. dunn.test: Dunn's test of Multiple Comparisons Using Rank Sums. R package version 1.3.5.

Gibson, J., N. Pfennig, and J. B. Waterbury. 1984. Chloroherpeton thalassium gen. nov. et spec. nov., a non-filamentous, flexing and gliding green sulfur bacterium. Archives of Microbiology 138:96–101.

Jaramillo, C. A. 2018. Evolution of the Isthmus of Panama: biological, palaeoceanographic and palaeoclimatological implications. Mountains, Climate and Biodiversity:323–338.

Jordan, D. S. 1908. The Law of Geminate Species. The American Naturalist 42:73–80.

Katoh, K., and D. M. Standley. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Molecular Biology and Evolution 30:772–780.

Knowlton, N., and L. A. Weigt. 1998. New dates and new rates for divergence across the Isthmus of Panama. Proceedings of the Royal Society B: Biological Sciences 265:2257–2263.

Leigh, E. G., A. O'Dea, and G. J. Vermeij. 2014. Historical biogeography of the isthmus of panama. Biological Reviews 89:148–172.

Lessios, H. A. 1990. Adaptation and phylogeny as determinants of egg size in echinoderms from the two sides of the Isthmus of Panama. American Naturalist 135:1–13.

Lessios, H. A. 2008. The Great American Schism: Divergence of Marine Organisms After the Rise of the Central American Isthmus. Annual Review of Ecology, Evolution, and Systematics 39:63–91.

Lim, S. J., and S. R. Bordenstein. 2020. An introduction to phylosymbiosis. Proceedings of the Royal Society B: Biological Sciences 287:20192900.

Louca, S., P. M. Shih, M. W. Pennell, W. W. Fischer, L. W. Parfrey, and M. Doebeli. 2018. Bacterial diversification through geological time. Nature Ecology & Evolution 2:1458–1467.

Lozupone, C., M. E. Lladser, D. Knights, J. Stombaugh, and R. Knight. 2011. UniFrac: An effective distance metric for microbial community comparison. ISME Journal 5:169–172.

Marko, P. B., and J. B. C. Jackson. 2001. Patterns of morphological diversity among and within arcid bivalve species pairs separated by the Isthmus of Panama. Journal of Paleontology 75:590–606.

Mazel, F., K. M. Davis, A. Loudon, and W. K. Kwong. 2018. Is Host Filtering the Main Driver of Phylosymbiosis across the Tree of Life? mSystems 3:e00097-18.

McKnight, D. T., R. Huerlimann, D. S. Bower, L. Schwarzkopf, R. A. Alford, and K. R. Zenger. 2019. Methods for normalizing microbiome data: An ecological perspective. Methods in Ecology and Evolution 10:389–400.

McMurdie, P. J., and S. Holmes. 2013. Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS ONE 8:e61217.

Mikryukov, V. 2018. metagMisc: Miscellaneous functions for metagenomic analysis. R package version 0.0.4.

Miura, O., M. E. Torchin, and E. Bermingham. 2010. Molecular phylogenetics reveals differential divergence of coastal snails separated by the Isthmus of Panama. Molecular Phylogenetics and Evolution 56:40–48.

Morrow, K. M., C. L. Fiore, and M. P. Lesser. 2016. Environmental drivers of microbial community shifts in the giant barrel sponge, Xestospongia muta, over a shallow to mesophotic depth gradient. Environmental microbiology 18:2025–2038.

Neu, A. T., E. E. Allen, and K. Roy. 2019. Diversity and composition of intertidal gastropod microbiomes across a major marine biogeographic boundary. Environmental Microbiology Reports 11:434–447.

Neu, A. T., I. V. Hughes, E. E. Allen, and K. Roy. 2021. Decade-scale stability and change in a marine bivalve microbiome. Molecular Ecology 30:1237–1250.

O'Dea, A., H. A. Lessios, A. G. Coates, R. I. Eytan, S. A. Restrepo-Moreno, A. L. Cione, L. S. Collins, A. De Queiroz, D. W. Farris, R. D. Norris, R. F. Stallard, M. O. Woodburne, O. Aguilera, M.-P. Aubry, W. A. Berggren, A. F. Budd, M. A. Cozzuol, S. E. Coppard, H. Duque-Caro, S. Finnegan, G. M. Gasparini, E. L. Grossman, K. G. Johnson, L. D. Keigwin, N. Knowlton, E. G. Leigh, J. S. Leonard-Pingel, P. B. Marko, N. D. Pyenson, P. G. Rachello-Dolmen, E. Soibelzon, L. Soibelzon, J. A. Todd, G. J. Vermeij, and J. B. C. Jackson. 2016. Formation of the Isthmus of Panama. Science Advances 2:e1600883.

Oksanen, J., F. G. Blanchet, M. Friendly, R. Kindt, P. Legendre, D. McGlinn, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, E. Szoecs, and H. Wagner. 2020. vegan: Community Ecology Package. R package version 2.5-7.

Paradis, E., and K. Schliep. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35:526–528.

Pfister, C. A., F. Meyer, and D. A. Antonopoulos. 2010. Metagenomic profiling of a microbial assemblage associated with the california mussel: a node in networks of carbon and nitrogen cycling. PLoS ONE 5:e10518.

Pollock, F. J., R. McMinds, S. Smith, D. G. Bourne, B. L. Willis, M. Medina, R. V. Thurber, and J. R. Zaneveld. 2018. Coral-associated bacteria demonstrate phylosymbiosis and cophylogeny. Nature Communications 9:4921.

Price, M. N., P. S. Dehal, and A. P. Arkin. 2010. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. PLoS ONE 5:e9490.

Quast, C., E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies, and F. O. Glöckner. 2013. The SILVA ribosomal RNA gene database project: improved data processing and

web-based tools. Nucleic Acids Research 41:D590–D596.

R Core Team 2013. R: A language and environment for statistical computing.

Rohart, F., B. Gautier, A. Singh, and K. A. Lê Cao. 2017. mixOmics: an R package for 'omics feature selection and multiple data integration. PLoS Computational Biology 13:e1005752.

Schliep, K. P. 2011. phangorn: Phylogenetic analysis in R. Bioinformatics 27:592–593.

Simpson, G. G. 1980. Splendid isolation: the curious history of South American mammals.

Sunagawa, S., L. P. Coelho, S. Chaffron, J. R. Kultima, K. Labadie, G. Salazar, B. Djahanschiri, G. Zeller, D. R. Mende, A. Alberti, F. M. Cornejo-Castillo, P. I. Costea, C. Cruaud, F. D'Ovidio, S. Engelen, I. Ferrera, J. M. Gasol, L. Guidi, F. Hildebrand, F. Kokoszka, C. Lepoivre, G. Lima-Mendez, J. Poulain, B. T. Poulos, M. Royo-Llonch, H. Sarmento, S. Vieira-Silva, C. Dimier, M. Picheral, S. Searson, S. Kandels-Lewis, C. Bowler, C. de Vargas, G. Gorsky, N. Grimsley, P. Hingamp, D. Iudicone, O. Jaillon, F. Not, H. Ogata, S. Pesant, S. Speich, L. Stemmann, M. B. Sullivan, J. Weissenbach, P. Wincker, E. Karsenti, J. Raes, S. G. Acinas, and P. Bork. 2015. Structure and function of the global ocean microbiome. Science 348:1261359.

Taylor, H. B., and H. D. Kurtz, Jr. 2020. Microbial community structure shows differing levels of temporal stability in intertidal beach sands of the grand strand region of South Carolina. PLoS ONE 15:e0229387.

Thacker, C. E. 2017. Patterns of divergence in fish species separated by the Isthmus of Panama. BMC Evolutionary Biology 17:111.

Varon-Lopez, M., A. C. F. Dias, C. C. Fasanella, A. Durrer, I. S. Melo, E. E. Kuramae, and F. D. Andreote. 2014. Sulphur-oxidizing and sulphate-reducing communities in Brazilian mangrove sediments. Environmental Microbiology 16:845–855.

Wang, Y., H. F. Sheng, Y. He, J. Y. Wu, Y. X. Jiang, N. F. Y. Tam, and H. W. Zhou. 2012. Comparison of the levels of bacterial diversity in freshwater, intertidal wetland, and marine sediments by using millions of illumina tags. Applied and Environmental Microbiology 78:8264–8271.

Weiss, S., Z. Z. Xu, S. Peddada, A. Amir, K. Bittinger, A. Gonzalez, C. Lozupone, J. R. Zaneveld, Y. Vázquez-Baeza, A. Birmingham, E. R. Hyde, and R. Knight. 2017. Normalization and microbial differential abundance strategies depend upon data characteristics. Microbiome 5:27.

Wilkins, L. G. E., M. Leray, A. O'Dea, B. Yuen, R. S. Peixoto, T. J. Pereira, H. M. Bik, D. A. Coil, J. E. Duffy, E. A. Herre, H. A. Lessios, N. M. Lucey, L. C. Mejia, D. B. Rasher, K. H. Sharp, E. M. Sogin, R. W. Thacker, R. V. Thurber, W. T. Wcislo, E. G. Wilbanks, and J. A. Eisen. 2019. Host-associated microbiomes drive structure and function of marine

ecosystems. PLoS Biology 17:e3000533.

Woodhams, D. C., M. C. Bletz, C. G. Becker, H. A. Bender, D. Buitrago-Rosas, H. Diebboll, R. Huynh, P. J. Kearns, J. Kueneman, E. Kurosawa, B. C. Labumbard, C. Lyons, K. McNally, K. Schliep, N. Shankar, A. G. Tokash-Peters, M. Vences, and R. Whetstone. 2020. Host-associated microbiomes are predicted by immune system complexity and climate. Genome Biology 21:23.

Xie, H., J. Chen, L. Feng, L. He, C. Zhou, P. Hong, S. Sun, H. Zhao, Y. Liang, L. Ren, Y. Zhang, and C. Li. 2021. Chemotaxis-selective colonization of mangrove rhizosphere microbes on nine different microplastics. Science of the Total Environment 752:142223.
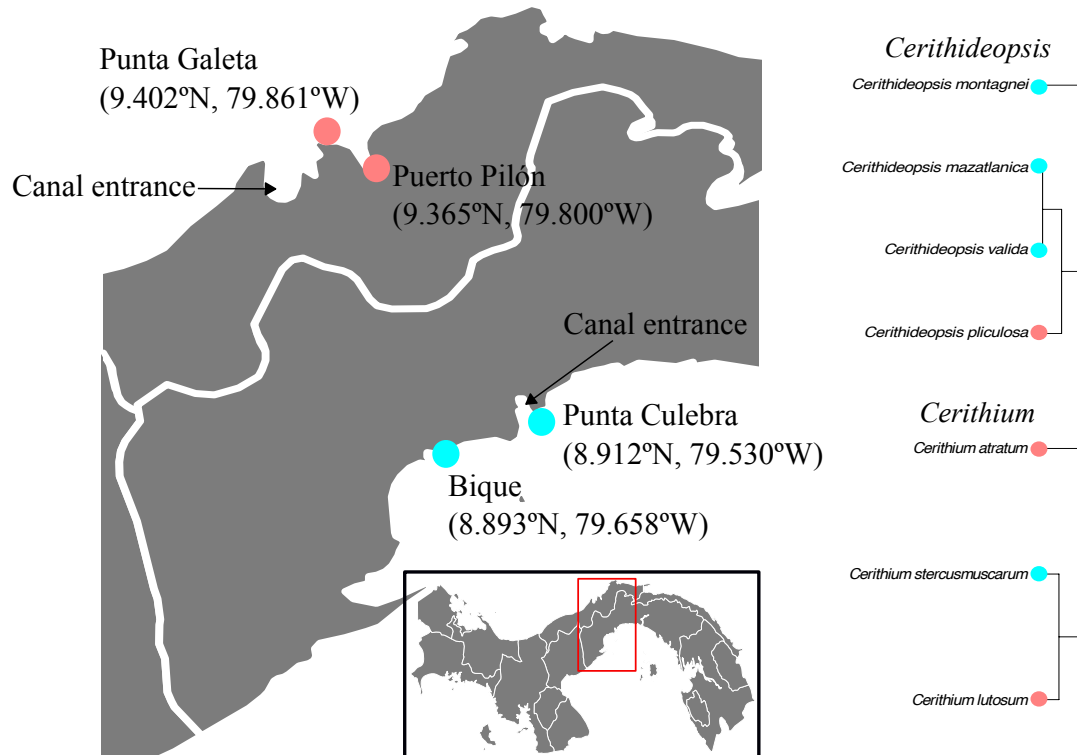
Figure 4.1. Map of sampling area. The map of Panama (bottom inset) shows the location of the detailed map of the sampling area (left). Blue dots indicate sampling sites in the Eastern Pacific, while pink dots indicate sampling sites in the Caribbean. Phylogenies of the species sampled from each host genus (right) have colors to represent the side of the IP each species was sampled from.
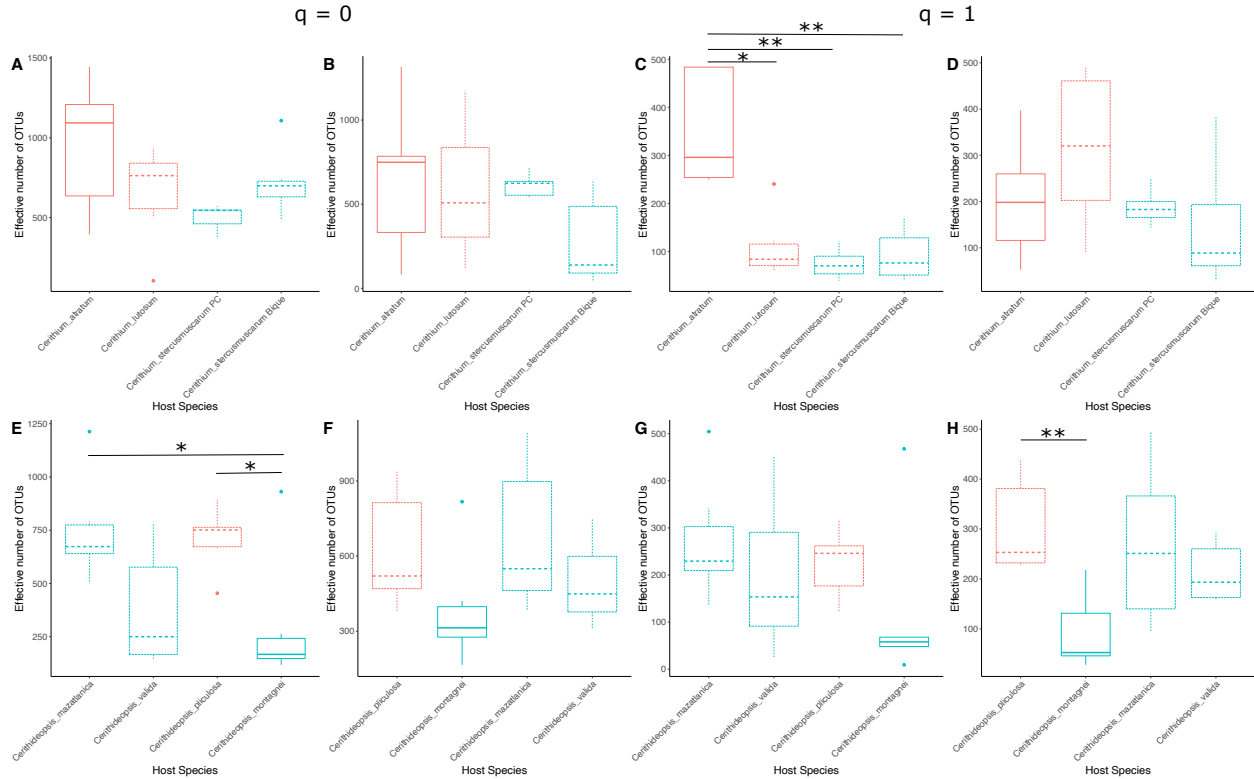
Figure 4.2. Hill numbers of q = 0 (A,B,E,F) and q = 1 (C,D,G,H) from whole-body (A,C,E,G) and shell-surface (B,D,F,H) samples. Blue boxes represent samples collected from the Eastern Pacific and pink boxes represent samples collected from the Caribbean. Dashed boxes represent geminate species and solid boxes represent congeneric outgroups. Asterisks indicate samples which were found to be significantly different by Dunn's test at p < 0.05 (*) or p < 0.005 (**).
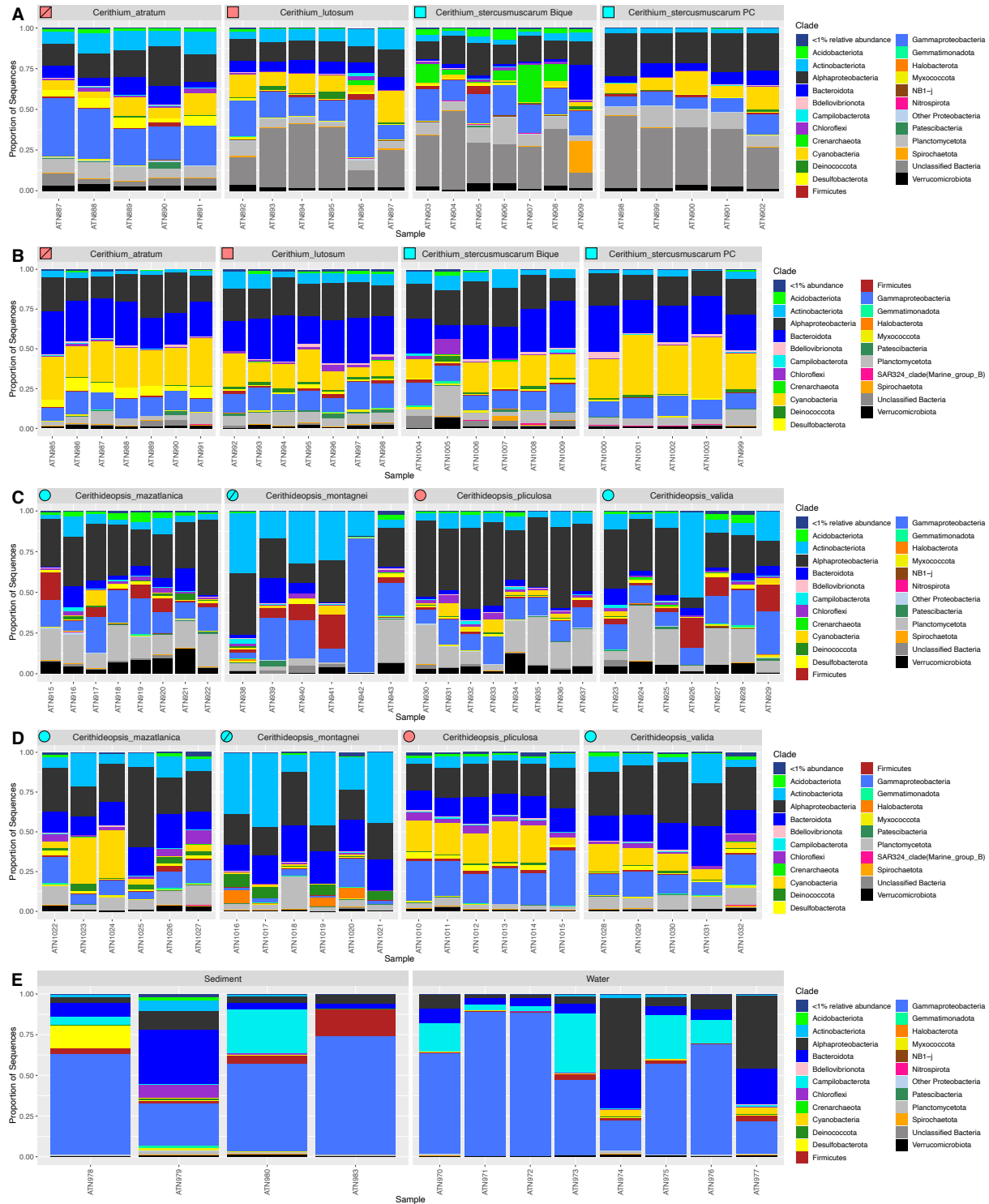
Figure 4.3. Barplots of the relative abundances of microbial phyla in the whole-body (A,C), shell-surface (B,D) and environmental (E) samples. The phylum Proteobacteria is separated into classes for increased resolution. In the top-left corner of each plot, the shape refers to the genus (squares and circles for *Cerithium* and *Cerithideopsis*, respectively), the color refers to the ocean basin of sampling (blue and pink for Eastern Pacific and Caribbean, respectively) and geminate species are open while congeneric outgroups are represented by a diagonal line.
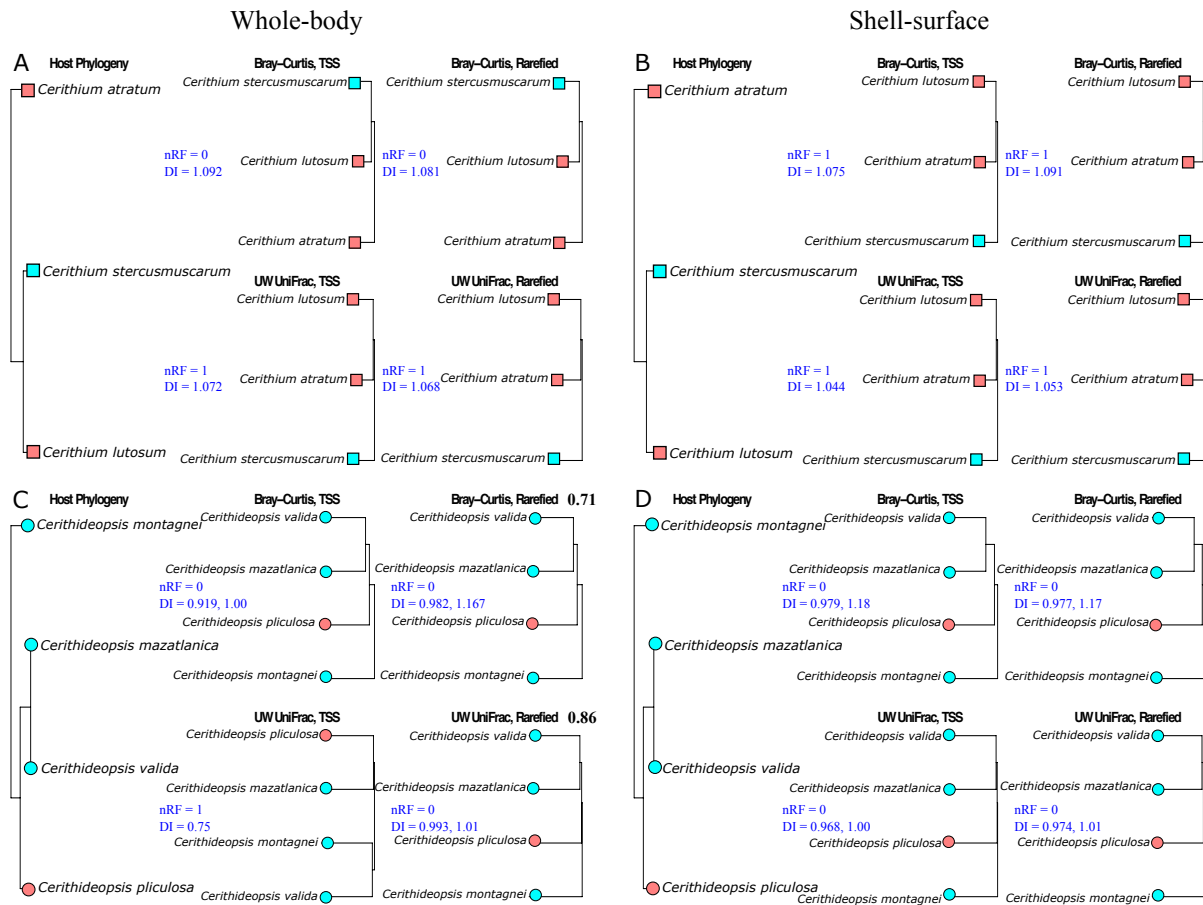
Figure 4.4. Comparing host phylogeny to microbiome composition in the whole-body (A,C) and shell-surface (B,D) communities of *Cerithium* (A,B) and *Cerithideopsis* (C,D) species. For each sample type, we show the host phylogeny (left) and UPGMA plots generated using BCD and TSS (top middle), BCD and rarefying (top right), unweighted UniFrac and TSS (bottom middle) and unweighted UniFrac and rarefying (bottom right). Each UPGMA plot includes the normalized Robinson-Foulds value (nRF) and Dunn's index (DI) in blue. Multiple DI values provide the clustering strength when the UPGMA plot is divided into two and three clusters, respectively. Bolded numbers in (C) indicate the proportion of rarefied datasets which produce the pattern shown. Tip shape refers to the genus (squares and circles for *Cerithium* and *Cerithideopsis*, respectively) and color refers to the ocean basin of sampling (blue and pink for Eastern Pacific and Caribbean, respectively).
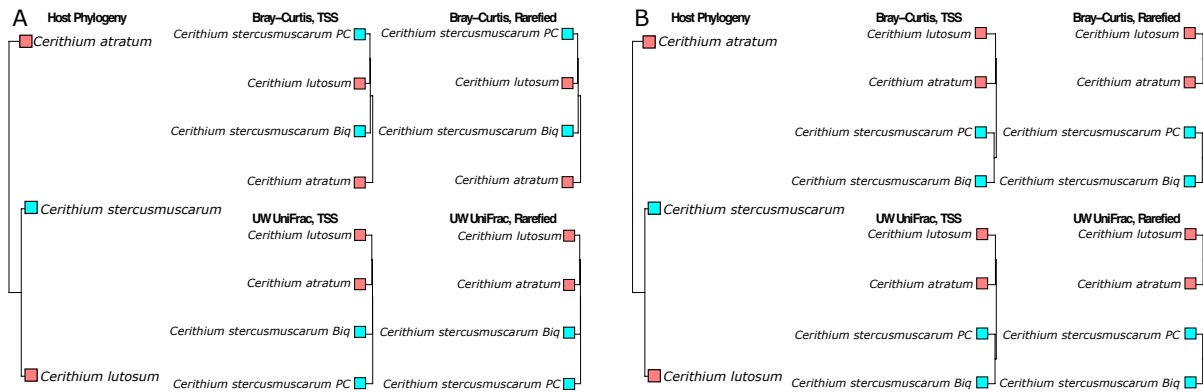
Figure 4.5. Comparing host phylogeny to microbiome composition in the whole-body (A) and shell-surface (B) communities of *Cerithium* species, separated by collection site. For each sample type, we show the host phylogeny (left) and UPGMA plots generated using BCD and TSS (top middle), BCD and rarefying (top right), unweighted UniFrac and TSS (bottom middle) and unweighted UniFrac and rarefying (bottom right). Tip color refers to the ocean basin of sampling (pink and blue for Eastern Pacific and Caribbean, respectively).
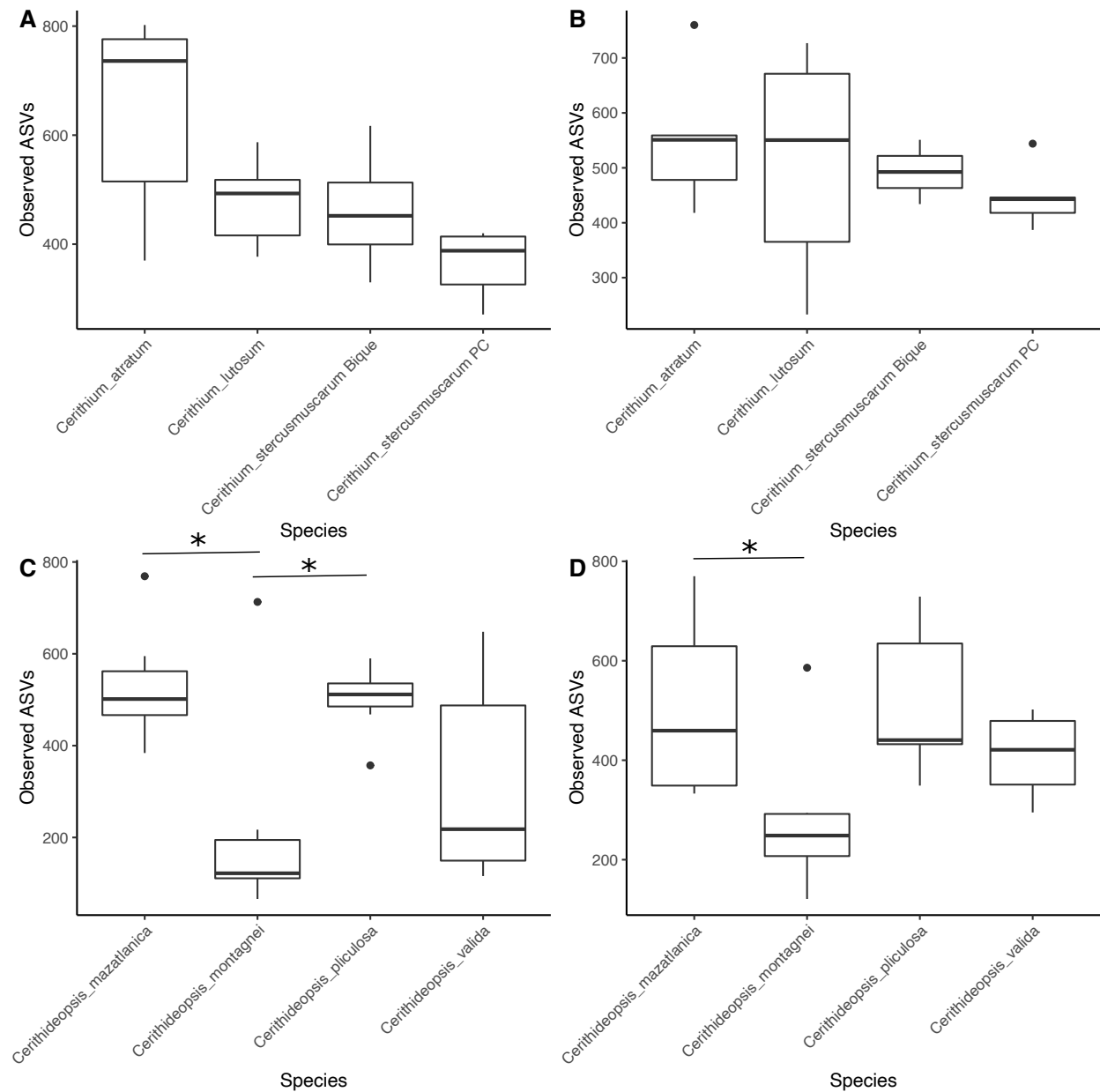
Supplemental Figure 4.1. Observed ASV richness, based on the average of 100 ASV tables rarefied to 3,000 sequences per sample, for *Cerithium* (A,B) and *Cerithideopsis* (C,D) whole-body (A,C) and shell-surface (B,D) samples. Asterisks indicate samples which were found to be significantly different by Dunn's test at $p < 0.05$ (*).
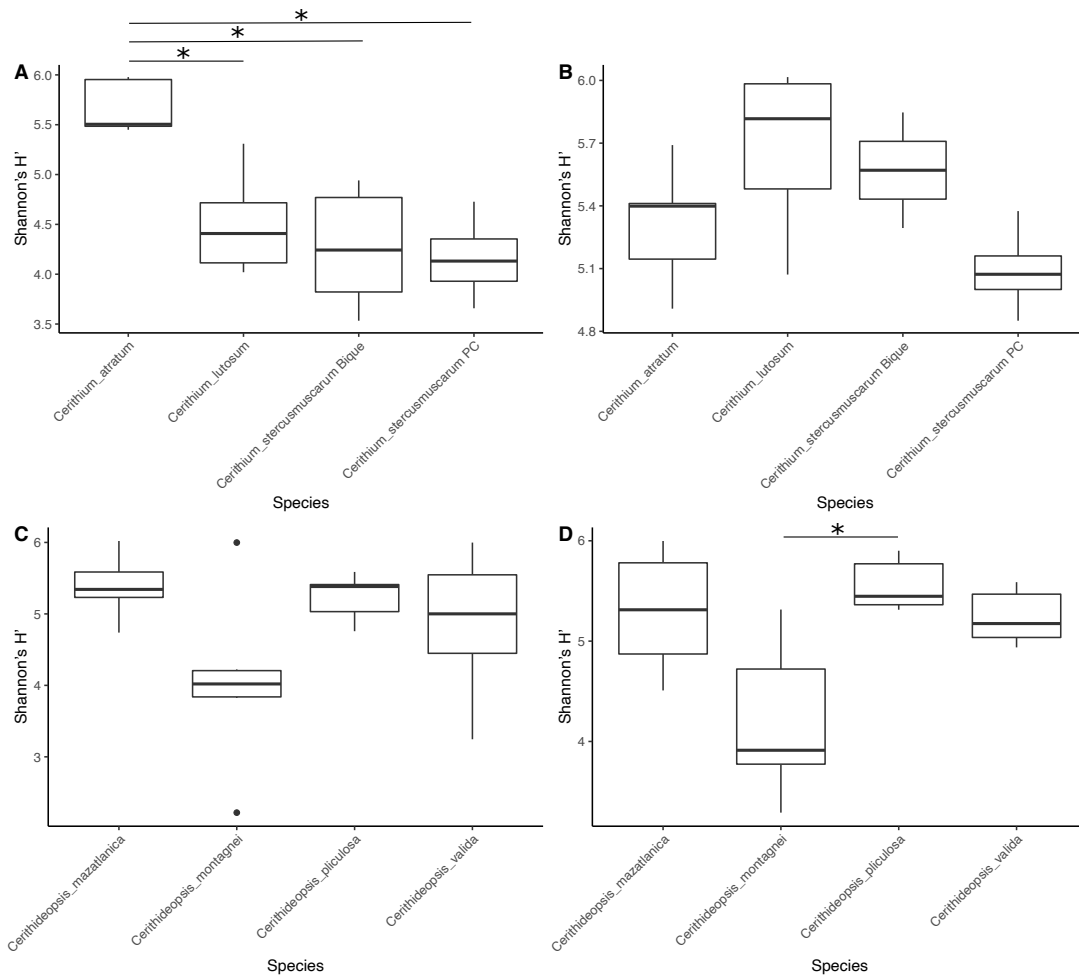
Supplemental Figure 4.2. Shannon's H', based on the average of 100 ASV tables rarefied to 3,000 sequences per sample, for *Cerithium* (A,B) and *Cerithideopsis* (C,D) whole-body (A,C) and shell-surface (B,D) samples. Asterisks indicate samples which were found to be significantly different by Dunn's test at $p < 0.05$ (*).
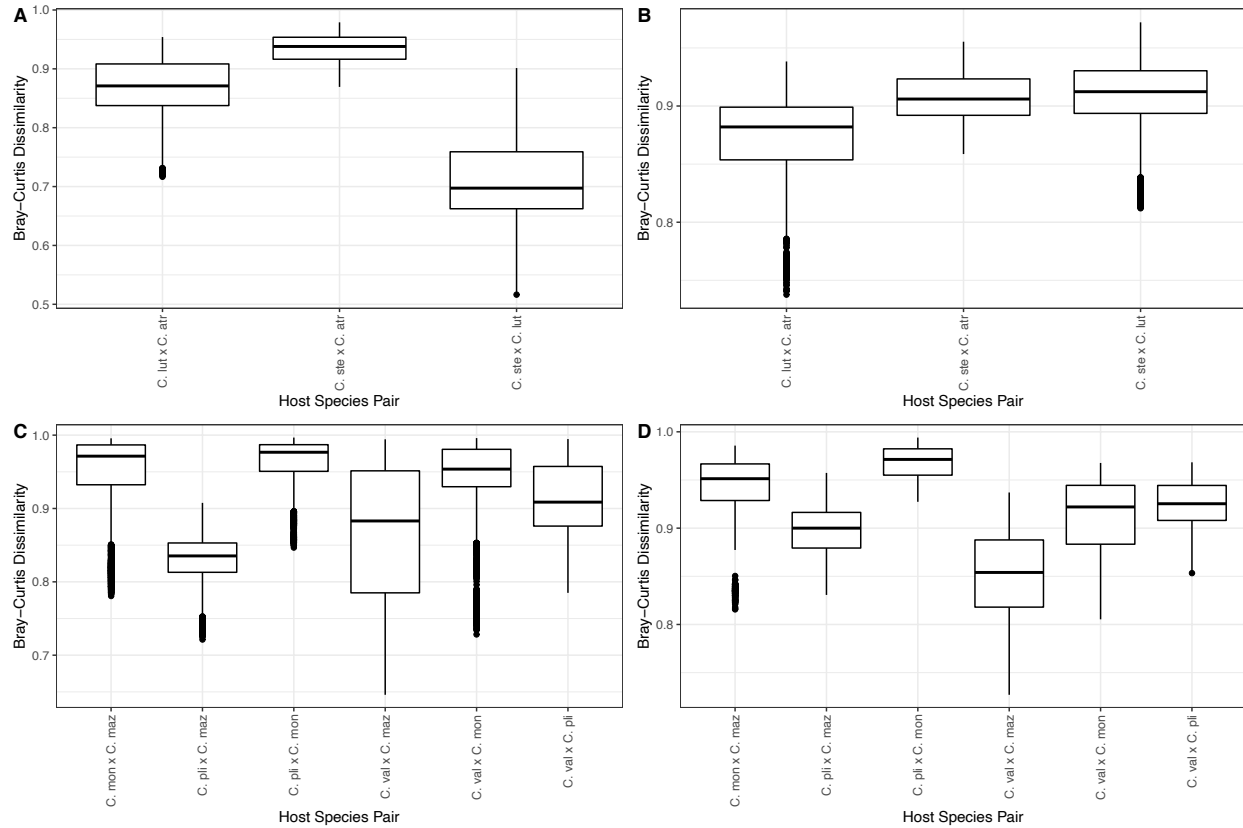
**Supplemental Figure 4.3.** Boxplots of Bray-Curtis dissimilarity between each host species pair in the *Cerithium* (A,B) and *Cerithideopsis* (C,D) whole-body (A,C) and shell-surface (B,D). Boxplots show all values across the 100 datasets rarefied to 3,000 sequences/sample.

118
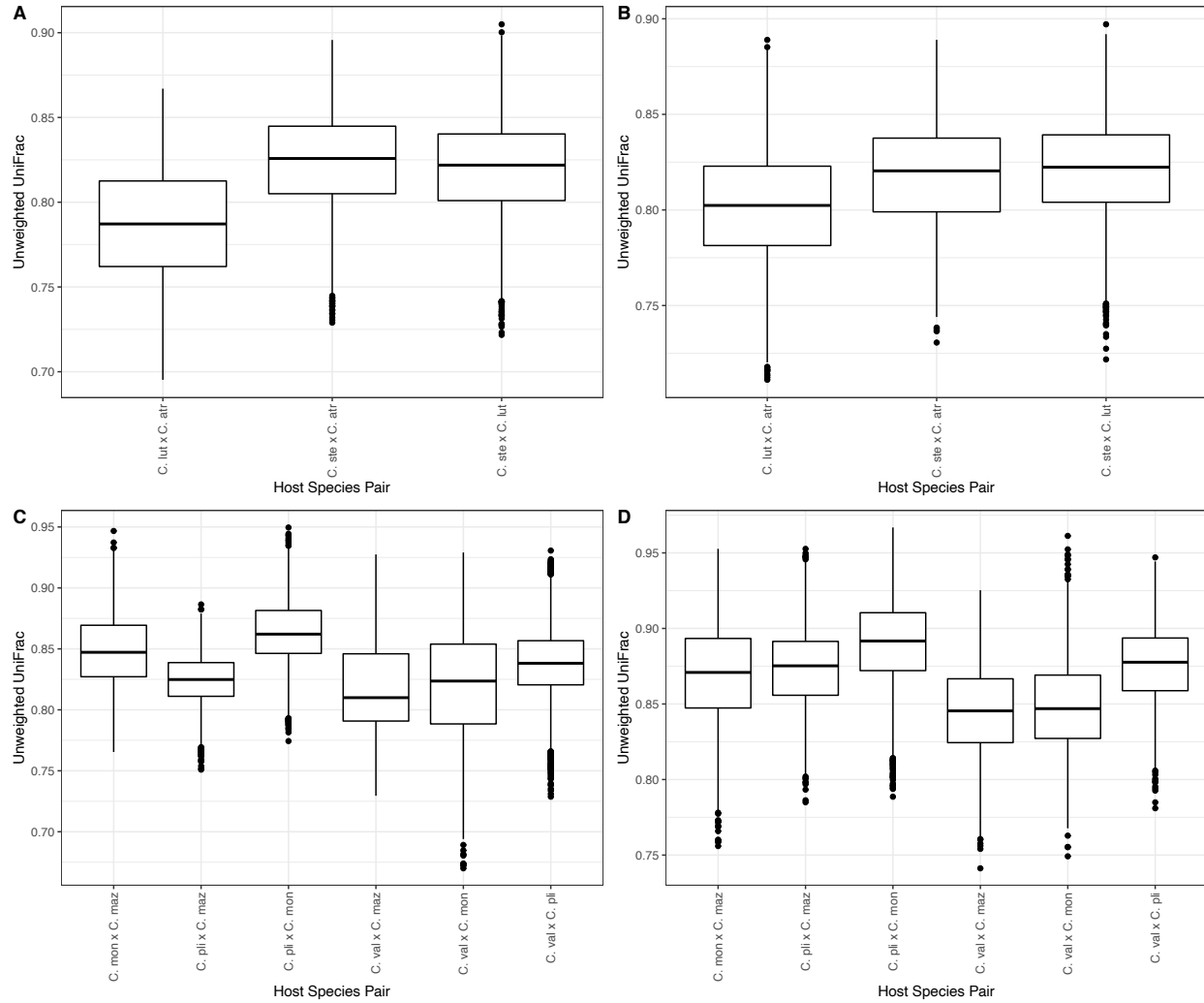
Supplemental Figure 4.4. Boxplots of unweighted UniFrac distance between each host species pair in the *Cerithium* (A,B) and *Cerithideopsis* (C,D) whole-body (A,C) and shell-surface (B,D). Boxplots show all values across the 100 datasets rarefied to 3,000 sequences/sample.
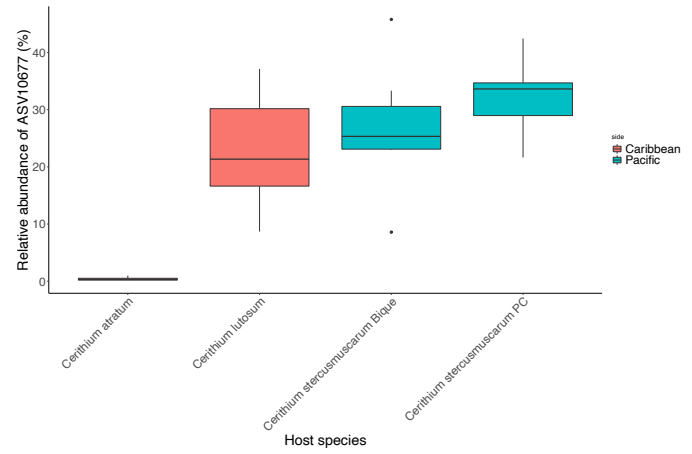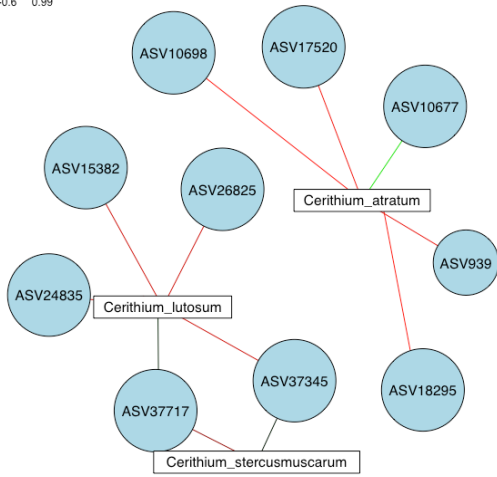
Supplemental Figure 4.5. SPLS-DA network of ASVs (left) and relative abundance of ASV10677 (right) in *Cerithium* whole-body samples.
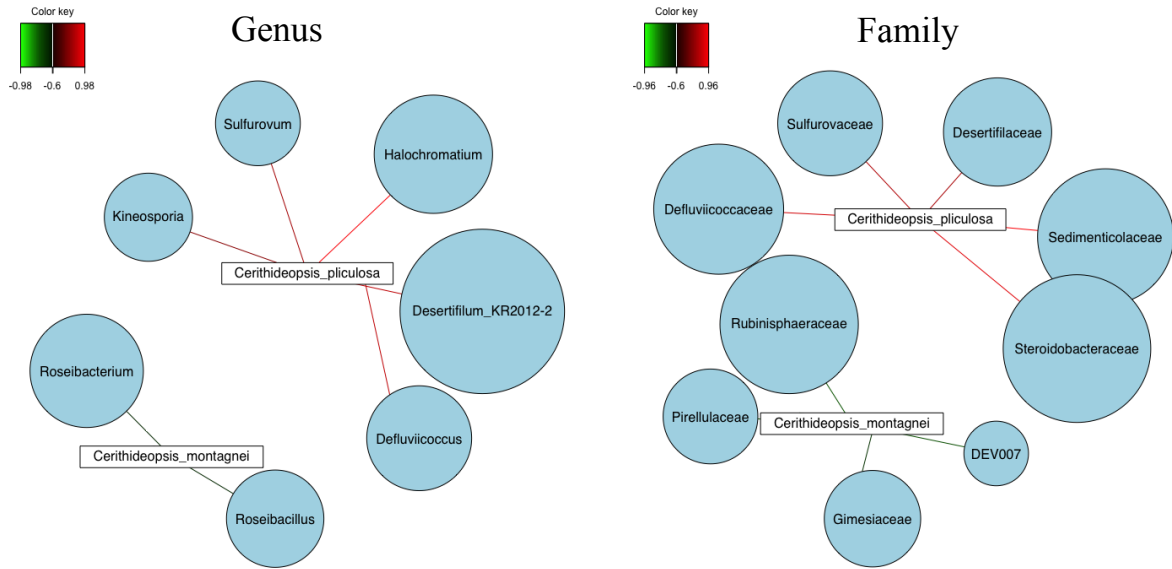
Supplemental Figure 4.6. SPLS-DA networks of microbial genera (left) and families (right) in *Cerithideopsis* whole-body samples.
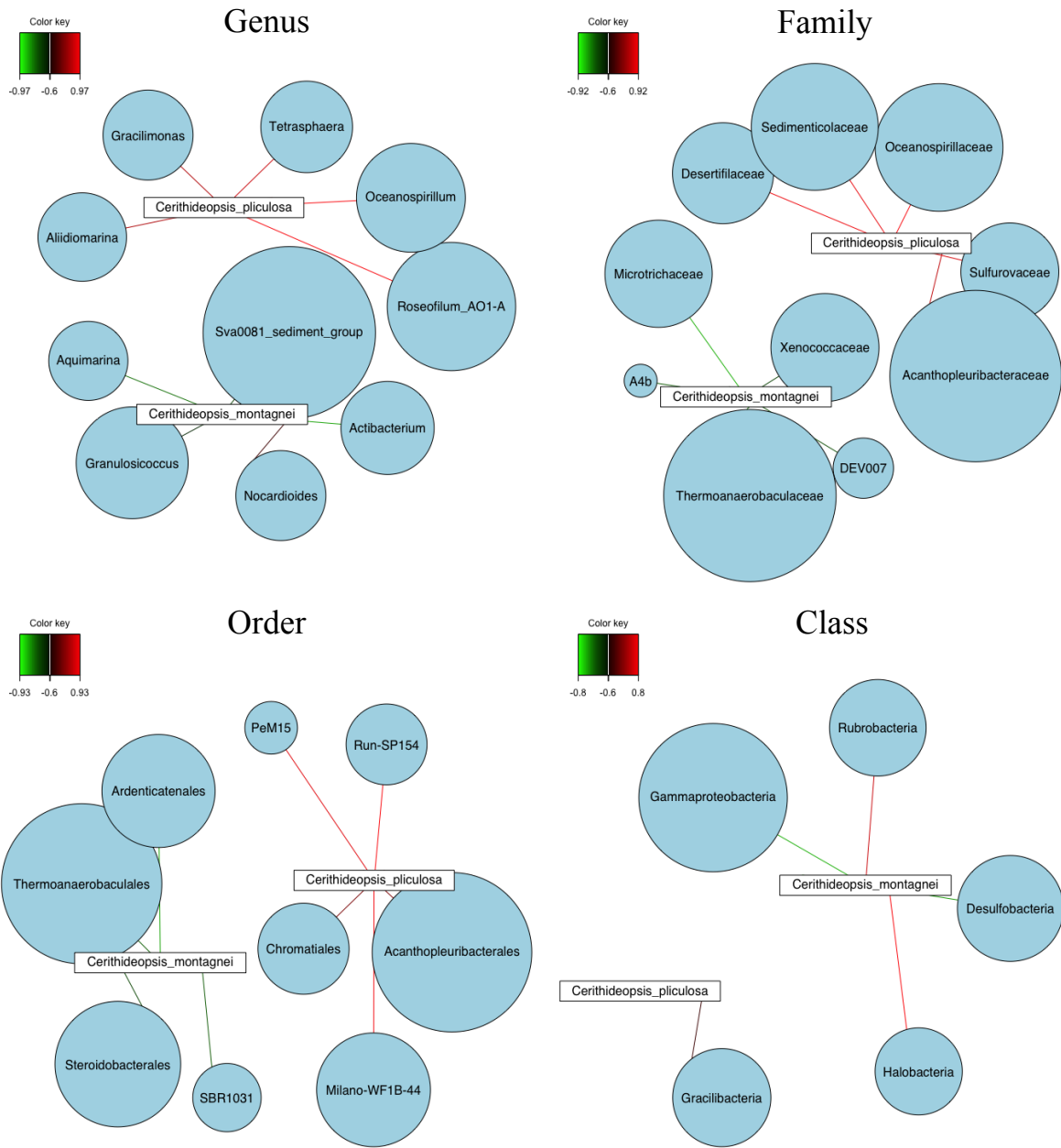
Supplemental Figure 4.7. SPLS-DA networks of microbial genera (top left), families (top right), orders (bottom left) and classes (bottom right) in *Cerithideopsis* shell-surface samples.

Supplemental Table 4.1. The number of whole-body and shell-surface samples collected from each host species at each site.

Host species	Collection site	Whole-body samples (*n*)	Shell-surface samples (*n*)
Cerithium atratum	Punta Galeta	5	7
Cerithium lutosum	Punta Galeta	6	7
Cerithium stercusmuscarum	Bique	7	6
Cerithium stercusmuscarum	Punta Culebra	5	5
Cerithideopsis mazatlanica	Bique	8	6
Cerithideopsis montagnei	Bique	6	6
Cerithideopsis pliculosa	Puerto Pilón	8	6
Cerithideopsis valida	Bique	7	5

Supplemental Table 4.2. Results of Kruskal-Wallis tests comparing Hill numbers at q = 0 and q = 1 across species within each host genus and body site, based on the unrarefied datasets.

Sample Type	Diversity Metric	X^2	p
Cerithium whole-body	q = 0	5.396	0.145
Cerithium whole-body	q = 1	11.858	**0.0079**
Cerithium shell-surface	q = 0	4.326	0.228
Cerithium shell-surface	q = 1	5.172	0.16
Cerithideopsis whole-body	q = 0	9.902	**0.0194**
Cerithideopsis whole-body	q = 1	6.203	0.102
Cerithideopsis shell-surface	q = 0	6.541	0.088
Cerithideopsis shell-surface	q = 1	10.362	**0.0157**

Supplemental Table 4.3. Results of Kruskal-Wallis tests comparing observed ASV richness and Shannon's H' across species within each host genus and body site, based on the average of 100 datasets rarefied to 3,000 sequences/sample.

Sample Type	Diversity Metric	X^2	p
Cerithium whole-body	Observed ASVs	5.483	0.251
Cerithium whole-body	Shannon's H'	6.818	**0.009**
Cerithium shell-surface	Observed ASVs	0.033	0.855
Cerithium shell-surface	Shannon's H'	1.633	0.201
Cerithideopsis whole-body	Observed ASVs	7.531	0.057
Cerithideopsis whole-body	Shannon's H'	6.436	0.092
Cerithideopsis shell-surface	Observed ASVs	7.662	0.053
Cerithideopsis shell-surface	Shannon's H'	9.517	**0.023**

Supplemental Table 4.4. PERMANOVA and betadisper results from all sample types, normalized by total sum scaling, using both Bray-Curtis dissimilarity and unweighted UniFrac distance. A significant betadisper p value in the *Cerithium* whole-body samples was found due to the divergence between the two populations of *C. stercusmuscarum*; though after removing one population from analysis, species were still found to be significantly different from one another (PERMANOVAs: all *p < 0.001*, betadisper: all *p > 0.2*). In the *Cerithideopsis* whole-body samples, a significant betadisper p value is unlikely to impact the interpretation of the PERMANOVA results, as PERMANOVA is robust to heterogeneity of dispersions when sample sizes are similar.

Sample Type	Diversity Metric	F Model	R^2	p	betadisper
Cerithium whole-body	Bray-Curtis	5.226	0.343	**<0.001**	0.969
Cerithium whole-body	Unweighted UniFrac	1.802	0.153	**<0.001**	**0.005**
Cerithium shell-surface	Bray-Curtis	3.537	0.243	**<0.001**	0.510
Cerithium shell-surface	Unweighted UniFrac	1.660	0.131	**<0.001**	0.902
Cerithideopsis whole-body	Bray-Curtis	3.484	0.295	**<0.001**	**<0.001**
Cerithideopsis whole-body	Unweighted UniFrac	2.082	0.200	**<0.001**	0.086
Cerithideopsis shell-surface	Bray-Curtis	4.149	0.396	**<0.001**	0.466
Cerithideopsis shell-surface	Unweighted UniFrac	1.475	0.189	**<0.001**	0.455

Supplemental Table 4.5. PERMANOVA and betadisper results from all sample types, normalized by rarefying to 3,000 sequences per sample, using both Bray-Curtis dissimilarity and unweighted UniFrac distance.

Sample Type	Diversity Metric	F Model	R^2	p	betadisper
Cerithium whole-body	Bray-Curtis	5.392	0.362	**<0.001**	0.278
Cerithium whole-body	Unweighted UniFrac	1.946	0.170	**<0.001**	**<0.001**
Cerithium shell-surface	Bray-Curtis	4.345	0.367	**<0.001**	0.097
Cerithium shell-surface	Unweighted UniFrac	2.151	0.223	**<0.001**	0.218
Cerithideopsis whole-body	Bray-Curtis	3.191	0.277	**<0.001**	**<0.001**
Cerithideopsis whole-body	Unweighted UniFrac	2.204	0.209	**<0.001**	0.054
Cerithideopsis shell-surface	Bray-Curtis	3.971	0.385	**<0.001**	0.397
Cerithideopsis shell-surface	Unweighted UniFrac	1.675	0.209	**<0.001**	0.714

Supplemental Table 4.6. PERMANOVA and betadisper results comparing host species to environmental samples, normalized by total sum scaling, using both Bray-Curtis dissimilarity and unweighted UniFrac distance.

Sample Type	Diversity Metric	F Model	R^2	p	betadisper
Cerithium whole-body	Bray-Curtis	5.483	0.174	**<0.001**	0.283
Cerithium whole-body	Unweighted UniFrac	3.026	0.074	**<0.001**	0.988
Cerithium shell-surface	Bray-Curtis	5.297	0.117	**<0.001**	0.288
Cerithium shell-surface	Unweighted UniFrac	3.014	0.070	**<0.001**	0.205
Cerithideopsis whole-body	Bray-Curtis	6.315	0.126	**<0.001**	0.536
Cerithideopsis whole-body	Unweighted UniFrac	2.667	0.057	**<0.001**	0.725
Cerithideopsis shell-surface	Bray-Curtis	5.283	0.122	**<0.001**	0.417
Cerithideopsis shell-surface	Unweighted UniFrac	2.133	0.053	**<0.001**	0.080

Supplemental Table 4.7. Microbial taxa at each taxonomic level found to be significantly associated with a geminate species group by SPLS-DA analysis. Bolded taxa are those presenting potential cophylogenetic patterns with the host clade using parafit ($p < 0.05$). An "X" indicates that no significant associations were found.

Sample type	ASV	Genus	Family	Order	Class
Cerithium whole-body	ASV10677	X	X	X	X
Cerithium shell-surface	X	X	X	X	X
Cerithideopsis whole-body	X	*Roseibacillus* *Roseibacterium*	DEV007 Gimesiaceae Pirellulaceae Rubinisphaeraceae	X	X
Cerithideopsis shell-surface	X	*Actibacterium* *Aquimarina* *Granulosicoccus* *Sva0081*	**A4b** DEV007 Microtrichaceae **Thermoanaerobaculaceae** Xenococcaceae	Ardenticatenales **SBR1031** Steroidobacterales **Thermoanaerobaculales**	Desulfobacteria **Gammaproteobacteria**

Acknowledgements

Chapter 4, in full, is in review as Neu, A.T., Torchin, M.E., Allen, E.E. and Roy, K. Microbiome divergence of gastropod species separated by the Isthmus of Panama. In review; bioRxiv 2021.07.08.451645. The dissertation author was the primary investigator and author of this paper.

CHAPTER 5

Defining and quantifying the core microbiome: challenges and prospects

Abstract

The term "core microbiome" has become widely used in microbial ecology over the last decade. Broadly, the core microbiome refers to any set of microbial taxa, or the genomic and functional attributes associated with those taxa, that are characteristic of a host or environment of interest. Most commonly, core microbiomes are measured as the microbial taxa shared among two or more samples from a particular environment. Despite the popularity of this term and its growing use, there is little consensus about how a core microbiome should be quantified in practice. Here, we present a brief history of the core microbiome concept and use a representative sample of the literature to review the different metrics commonly used for quantifying the core. Empirical analyses have used a wide range of metrics for quantifying the core microbiome, including arbitrary occurrence and/or abundance cutoff values with the focal taxonomic level of the core ranging from phyla to amplicon sequence variants. However, many of these metrics are susceptible to sampling and other biases. Developing a standardized set of metrics for quantifying the core that consider such biases is necessary for testing specific hypotheses about the functional and ecological roles of core microbiomes.

Keywords

core microbiome; microbiota; microbial ecology; 16S ribosomal RNA gene

Introduction

The search for the core microbiome has become widespread within the field of microbial ecology. In general, a core microbiome can be defined as any set of microbial taxa as well as the associated genomic or functional attributes characteristic of a specific host or environment

(Turnbaugh *et al.*, 2007; Hamady and Knight, 2009; Risely, 2020). This has led to a number of studies focused on the genes (e.g., Turnbaugh *et al.*, 2009), functional pathways (e.g., Jiang *et al.*, 2016), and metabolic profiles (e.g., Stefanini *et al.*, 2017) common to microbial communities in a number of environments. Most commonly, however, the search for the core microbiome involves determining which taxa, if any, are shared among two or more microbial communities in a given host species or environment (Shade and Handelsman, 2012). These shared taxa are hypothesized to represent the most ecologically and/or functionally important microbial associates of that host or environment under the conditions sampled. In fact, it has been suggested that identifying core microbiome components may assist in addressing topics ranging from the maintenance of human oral and gut health (Zaura *et al.*, 2009; Bäckhed *et al.*, 2012) to the responses of organisms to anthropogenic climate change (Ainsworth and Gates, 2016; Hutchins *et al.*, 2019). The potential utility of these core taxa has led researchers to identify core microbiomes in a wide range of environments and hosts, from the skins of frogs (Christian *et al.*, 2018) to the Baltic Sea (Lindh *et al.*, 2017) to activated sludge (Saunders *et al.*, 2016), resulting in a rapid increase in the number of studies that include a core microbiome component over the past decade (Fig. 1).

Although analyses of the core taxonomic microbiome have provided a number of insights into the microbial ecology of a multitude of environments and hosts (Zaura *et al.*, 2009; Lundberg *et al.*, 2012; Ainsworth *et al.*, 2015; Henderson *et al.*, 2015), they often vary in their criteria for quantifying the core. In general, this involves determining the proportion of samples that share a set of microbial taxa, the relative abundances of shared taxa across samples or hosts, or a combination of the two. The taxonomic level used to define the core can also vary, as a core microbiome may be determined at the level of amplicon sequence variants (ASVs, sequences

clustered at 100% sequence similarity), phyla, or anywhere in between. Furthermore, the spatial and temporal scales over which a core microbiome is quantified are also variable, ranging from samples collected at a single site to a global sampling of a particular host taxon, and from samples collected at a single time point to multiple collections spanning days to years.

Here, we explore the diversity of methods that have been used to quantify core microbiomes, determine which practices are most widely used in the literature, benefits and challenges of this methodological variability, as well as implications for understanding the ecological and evolutionary processes that produce and maintain core microbiomes. To do this, we searched Google Scholar using the terms "core microbiome" OR "core microbiota" and limited the search to studies published between 2008 and 2019 (search conducted on March 18, 2020). We selected the first 200 primary research articles (excluding reviews and computational methods) which quantified a core microbiome of bacterial and archaeal taxa using 16S rRNA gene amplicon sequencing, a common method for determining the diversity and composition of prokaryotic communities. We then supplemented these articles with those previously downloaded into our own personal reference libraries, using the same date range and methodological criteria. This resulted in a representative sample of 224 studies published between 2008 and 2019 that involve an analysis of the core microbiome (Fig. 2a, Dataset S1). These studies are distributed across a variety of different plant and animal hosts, environments (e.g., soil, seawater), and industrial processes (e.g., wastewater treatment, industrial fermentation), which allowed us to further explore the methodologies for quantifying core microbiomes across subfields of microbial ecology (Fig. 2b). Although some studies within our sample set also determine a core fungal microbiome or core functional genes (see Dataset S1), we have limited our analysis of the literature to core prokaryotic taxa, as these comprise the

majority of published empirical studies. However, many of the major points discussed below are also applicable to fungal taxa or functional core microbiomes. We also note that most of the studies referenced here focus on the "core microbiota", which includes the microbial taxa within a particular environment, rather than the "core microbiome", which includes the structure and function of the community as well as the abiotic conditions in their particular environment (Berg *et al.*, 2020). While such distinction is useful, the published literature predominately uses the term "microbiome" for both types of analyses (Dataset S1) and our use of the term "core microbiome" here follows that practice.

History and evolution of the core microbiome concept

One of the original goals of the Human Microbiome Project (HMP) was to identify the *core microbiome* or "… *whatever factors are common among the microbiomes of all or the vast majority of humans*," (Turnbaugh *et al.*, 2007, p. 804). Thus, the term core microbiome is, by design, incredibly broad, recognizing that there may be multiple types of characteristics shared among different human microbiomes. These include individual genes, metabolic pathways, microbial taxa, as well as any other mechanisms that may play a role in host-microbiome interactions. Early studies commonly used a taxonomic approach, based on 16S rRNA gene sequencing data, to quantify a core microbiome in lean, healthy humans as well as people with specific phenotypes such as obesity (Turnbaugh *et al.*, 2009) and inflammatory bowel disease (Willing *et al.*, 2010). However, these studies varied greatly in their ability to identify a set of core microbial taxa consistently associated with humans. In the gut, for example, few taxa were shared among individuals, and in some cases no overlapping operational taxonomic units (OTUs), or clusters of sequences with a level of shared nucleotide identity, were found

(Turnbaugh *et al.*, 2009). In contrast, in the oral microbiome, a number of OTUs were identified

as members of the taxonomic core of healthy individuals (Zaura *et al.*, 2009; Eren *et al.*, 2014).

The heterogeneity of the results led to multiple different hypotheses about the core microbiome,

including that (i) cores may exist within certain human populations but not globally, (ii) a core

may be discernible only at a higher taxonomic level such as genera, or that (iii) the core may be

wholly functional, comprised of functional gene clusters rather than individual taxa (outlined in

Hamady and Knight, 2009). While these hypotheses are not mutually exclusive, they do provide

different insights into the nature and functioning of the human core microbiome. Further, these

early studies of human microbiomes provided a framework for subsequent analyses of core

microbiomes in a variety of non-human hosts and environments.

The first comprehensive overview of the conceptual basis for the core microbiome, by

Shade and Handelsman (Shade and Handelsman, 2012), revealed that the search for a core

microbiome was still largely in its discovery phase, relying heavily on Venn diagrams to identify

taxa (usually OTUs) shared among samples. They further outlined additional approaches for

future research, such as inclusion of relative abundances of taxa in identifying cores and

inclusion of a "persistent" or "dynamic" core microbiome measured across timescales (detailed

below). These advances have led to discoveries such as the stability of particular microbial

strains in the guts of individual humans (Faith *et al.*, 2013) and that certain microbial taxa persist

across major transitions, such as the saltwater to freshwater migration in Atlantic salmon (Dehler

et al., 2017). Identification of core microbial taxa can also lead to better understanding of their

functional roles through targeted culturing or other 'omics approaches (such as metagenomics or

meta-transcriptomics), as exemplified in the honeybee system (Sabree *et al.*, 2012; Engel and

Moran, 2013). On the other hand, a few studies across phylogenetically diverse hosts, ranging

from caterpillars to crustaceans, have found that the putative core microbiome was comprised of transient taxa repeatedly acquired from the environment rather than obligate associates of the host (Hammer *et al.*, 2017; Martin *et al.*, 2020). Finally, some authors have proposed moving away from a taxonomic approach altogether to an entirely functional definition of the core microbiome where the core "... *can be defined as a whole set of microbial vehicles, including replicators coding for essential functions for holobiont fitness"* (Lemanceau *et al.*, 2017, p. 584). One proposed solution to this issue is to specify the type of core microbiome of interest, such as a temporal core, ecological core, or functional core (Risely, 2020).

Current methods for identifying the taxonomic core microbiome

Core microbiomes are typically quantified using one of three methods: (i) the occurrence of microbial taxa across multiple samples of the same host/environment, (ii) the relative abundances of microbial taxa across such samples, or (iii) some combination of the two. These metrics can be readily estimated from a standard OTU table containing the number of sequence counts of each OTU (or other units of analysis) detected in each sample. However, as discussed below, these methods have a number of pros and cons (Table 1) and their applications have been highly variable, making it difficult to compare the sizes and compositions of core microbiomes across different hosts and environments.

<u>Occurrence only</u>

In our dataset, 75.4% of core microbiomes were quantified based solely on the occurrence of a taxon (OTU, genus, etc.) within a proportion of the samples collected (Dataset S1). Occurrence data are readily available, easily comparable, and provide information about the

spatial or temporal span over which the microbe and host/environment interact. The proportions of sites, samples, or time points over which a microbe must occur to be considered core, however, is always at the discretion of the author(s). In the most liberal cases from our dataset, a 30% occurrence standard was used, meaning that any OTU detected in at least 30% of samples was considered a core member (Ainsworth *et al.*, 2015; Sweet *et al.*, 2017). Others have used cutoffs between 50 and 99.9%, depending on the study system and number of total samples (Fig. 2c). Most commonly, though, studies required a taxon be observed in 100% of samples to be considered core (Fig. 2c). This approach aims to identify the obligate, or seemingly obligate, relationships between host and microbe. A potential criticism, however, is that a 100% occurrence requirement is likely to miss lower abundance taxa that have fallen below the threshold of detection in one or more samples, but are of functional and/or ecological importance (e.g., Huse *et al.*, 2012). The probability of this occurring depends, in part, on the sequencing depth of the samples, which is discussed in greater detail below. One proposed solution is to quantify the core microbiome using multiple occurrence cutoffs and evaluate how these different values affect the composition of the core. For example, in *Xestospongia* sponges, changing the occurrence cutoff has a relatively small impact on the composition of the core, suggesting many stable associations between host and microbe (Astudillo-García *et al.*, 2017). However, in other cases multiple occurrence cutoffs can result in significantly different numbers of core taxa and alter core composition, showing that individual microbial taxa may have different levels of association with the host (e.g., Turnbaugh *et al.*, 2009; Segata *et al.*, 2016).

Relative abundance only

Nine studies (4%) within our sample set quantified the core microbiome using only relative abundance criteria (Dataset S1). In certain cases, though, occurrence data may have been used but not explicitly stated in the text. In some of these studies, the most abundant set of taxa in each sample type are identified as core members (Brodie *et al.*, 2016; Beckers *et al.*, 2017). In others, taxa considered to be core were those that were preferentially enriched in the host of interest relative to the surrounding environment (Lundberg *et al.*, 2012; Schlaeppi *et al.*, 2014; Edwards *et al.*, 2015; Yeoh *et al.*, 2016). This method assumes that a high, or at least significantly increased, relative abundance in a particular host or environment is evidence of a stronger and more stable association with that system. A criticism of this approach is that some low abundance taxa may still play important functional roles within the host (Jousset *et al.*, 2017; Banerjee *et al.*, 2018). Furthermore, relative abundances of some taxa may change substantially over time (e.g., Neu, Hughes, *et al.*, 2021) so identifying core microbiomes using the abundance distribution of a single temporal snapshot may miss taxa that may be more common at other time points. One way to test for such variability is to use the relationship between the mean and variance of relative abundance of each OTU (e.g., Oh *et al.*, 2016), a trend known as Taylor's power law (Taylor, 1961). While this law has wide applications in ecology and beyond, in the present context, if the temporal or spatial variance of an OTU is greater than the mean, then the taxon should be considered less stable and its inclusion in the core microbiome potentially problematic. Such analyses, however, require replicate samples over time/space, something that is currently lacking in many core microbiome analyses.

Abundance-occurrence

139

While occurrence and abundance have each been used individually to quantify the core microbiome, combining them can provide a more conservative approach. This paired method was used in 11.6% of studies in our dataset (Dataset S1). Many studies using this method set a minimum relative abundance threshold under which a taxon is disqualified from core membership. Then, the remaining taxa are assessed by the number of sites/samples in which they occur, again using an assigned cutoff value. In the studies sampled here, the minimum relative abundance threshold for an OTU to be considered for core membership ranged from 0.001% (Antwis *et al.*, 2018) to 4.5% (Leis and Costa, 2019), while occurrence cutoffs ranged from 50-100% across host species, geographic sites, or treatments (Dataset S1).

Abundance-occurrence relationships have a conceptual basis in the macroecological literature, as it has been shown in larger eukaryotes (Hanski, 1982; Gaston *et al.*, 2000), and more recently in marine bacterioplankton (Lindh *et al.*, 2017), that the local relative abundance of a species is generally positively correlated with the number of sites that species occupies within a region. However, whether this relationship is universally applicable across environments and microbial taxa is still unknown. Recently, though, there have been a few attempts to use the abundance-occurrence relationship to determine the compositions of core microbiomes. For example, Li *et al.* (Li *et al.*, 2013) designed a simple model which visualizes taxa that lie above user-specified abundance and occurrence cutoffs values, using a binomial distribution to account for differences in sequencing depth. Shade and Stopnisek (Shade and Stopnisek, 2019) developed a method using abundance-occurrence relationships and contribution to Bray-Curtis similarity between samples in order to identify potential members of the core microbiome. They further compared the abundance-occurrence distributions of these taxa to a null model (Sullam *et al.*, 2012) in order to determine which core members may be

deterministically selected by the host/environment (Shade and Stopnisek, 2019). While such use of abundance-occurrence relationships has the potential to better constrain the makeup of the core microbiome, it is important to keep in mind the underlying assumptions of such models. The first assumption is that this relationship is likely to be phylogenetically and/or functionally constrained. Abundance-occurrence relationships in plants and animals have been shown to be constrained by phylogeny or functional guilds (Brown, 1995) and different clades of microbial taxa also show different patterns of abundance and occurrence across spatial scales (Neu, Allen, *et al.*, 2021; Neu, Torchin, *et al.*, 2021). Thus, using the abundance-occurrence relationship across the entire microbiome to identify the core, as is the current practice, is likely to be problematic. Instead, it would be better to use this approach to identify core taxa at the level of individual clades and then aggregate across the whole community. Another critical assumption here is that any observed abundance-occurrence relationship of microbes using a small number of samples or populations is truly characteristic of that taxon. In larger eukaryotic communities, the abundance of a species is usually non-uniform across its geographic range (Sexton *et al.*, 2009), and often tends to peak near the center of the distribution (Brown *et al.*, 1995). Whether such an "abundant center" pattern is also true of microbial taxa is currently unknown, but abundances of individual microbial taxa are also known to change along spatial gradients (Choudoir *et al.*, 2016; Neu, Allen, *et al.*, 2021). Thus, abundance-occurrence models are likely to be most useful for constraining the makeup of core microbiomes when they are based on large, representative samples across the geographic distributions of individual taxa, again something that is currently rare.

The Role of Sequencing Depth

The number of microbial sequences obtained from each sample, or the sequencing depth, remains an important consideration in microbiome studies. Multiple studies of environmental microbiomes using massively deep sequencing efforts have shown that rare taxa, often undetected by shallower sequencing, make up a substantial portion of many microbial communities (Sogin *et al.*, 2006; Gibbons *et al.*, 2013). This suggests that maximizing sequencing depth is important for quantifying core microbiomes, especially when using strict occurrence cutoffs. This is also true of functional cores determined using other 'omics datasets, as shallower sequencing is likely to miss low-abundance gene clusters and provide incomplete coverage of the whole community (e.g., Pereira-Marques *et al.*, 2019).

In addition to overall sequencing depth, a related issue is accounting for differences in sequencing depths across a set of samples when quantifying core microbiomes. Some lab-based methods can help to control for sequencing depth and include standardizing the mass or volume of samples prior to DNA extraction and pooling libraries in equal molarity before sequencing. Despite these efforts, though, variation in sequencing depths across different samples in the same study can be substantial and the use of statistical standardization methods to account for such variations has been the subject of much discussion and debate. Some have suggested that data should be rarefied (Weiss *et al.*, 2017; McKnight *et al.*, 2019) to a common sampling depth, typically to the level of the sample with fewest sequences, while others argue that such rarefaction is "inadmissible" and favor approaches that transform or scale sequence counts (McMurdie and Holmes, 2014; Gloor *et al.*, 2017). While these debates have led to the development of new statistical tools to account for sequencing depth, they have primarily focused on analyses of the alpha (e.g., Chiu and Chao, 2016; Alberdi and Gilbert, 2019) or beta (e.g., Martino *et al.*, 2019; Liu *et al.*, 2021) diversities. The performance of such approaches in

the context of quantifying the core microbiome remains poorly explored. Within our dataset, multiple studies rarefied sequence data before determining a core (e.g., Christian *et al.*, 2018; Aira *et al.*, 2019), while others did not rarefy their data at all (e.g., Burgsdorf *et al.*, 2014; Antwis *et al.*, 2018). A recent study found that, all else being constant, the number and identities of core microbiome members can change after rarefying data to four different sequencing depths (Ramakodi, 2021). This suggests that using rarefied data may provide an incomplete or potentially inaccurate picture of the core microbiome, as rarefying can remove tens or hundreds of thousands of sequences from individual samples, especially when variation in sequencing depth across samples is large.

In addition to taxon sampling, sequencing depth is likely to also have an impact on spatial occurrences of individual taxa (Roy and Witman, 2009) and thus affect occurrence-based measures of the core microbiome as well as the use of abundance-occurrence models discussed above. A potential solution here, borrowed from the macroecological literature, is to use a range-through approach where a taxon is presumed to be present everywhere within its geographical range limits even though it may be missing from some samples within that range (Roy and Witman, 2009). For example, in samples collected along a latitudinal transect a taxon would be counted as present in all sites between its northernmost and southernmost point of detection, even if it is not explicitly sampled there (e.g., Neu, Allen, *et al.*, 2021). While some of the observed absences are likely to be real, driven by local environmental conditions or other factors, the range-through metric can provide a hypothesis about the occurrences of individual taxa in a specific sample, which can then be tested by repeated sampling or deeper sequencing. Another approach would be to sequence multiple technical replicates from the same sample to determine the likelihood of a true negative result, as has been done in human genotyping data (Kim *et al.*,

2019), though this creates potential tradeoffs with additional sequencing costs or decreased sequencing depth per individual sample.

Unspecified Definition

We found that more than 18% of the sampled studies did not include a methodology for quantifying the core microbiome in the Methods section. In some cases, the criteria used for determining the core were provided in the Results or Discussion, while in other cases those criteria were unclear or not explicitly stated at all. This practice has also been mentioned by others (Shade and Stopnisek, 2019) and should be strongly discouraged, as it creates a challenge for those aiming to replicate a study or to undertake comparative or meta-analyses of core microbiomes across studies.

Effects of spatial, temporal and phylogenetic scales

Spatial Scale of sampling

Regardless of whether the core microbiome is quantified based solely on occurrences or a combination of occurrence and abundance, the choice of spatial scale over which the host is sampled can have a strong influence on the composition of the core. For example, the diversity, composition and function of the core microbiome of two host populations that are geographically close to each other is likely to be very different from that of two host populations separated by hundreds of kilometers. In our dataset, 67% of studies used samples collected at a single site (Fig. 2e), though in many of these cases, a core was determined between experimental treatments, disease treatments, time points, etc. (Dataset S1). The other 33% of studies analyzed samples collected from at least two distinct sites, as defined by the authors (Fig. 2e). However,

these studies vary greatly in the spatial scale over which these locations were sampled, ranging from two locations within the same bay (Turon *et al.*, 2019) to 23 collections on six continents (Xu *et al.*, 2018) (Fig 2e).

Because of the large variance in spatial scale over which published core microbiomes have been quantified, we have examples of cores that are essentially local (Turon *et al.*, 2019) to those that are "region-specific" (Burgsdorf *et al.*, 2014), and adapted to a host in a specific environment (e.g., Hernandez-Agreda *et al.*, 2016), to those that cover the entire geographic distribution of a host. The citrus tree rhizosphere, for example, was sampled from 23 sites spanning six continents, six climate regimes and seven soil types, effectively covering a large proportion of the areas in which citrus is grown as an agricultural product, providing a core citrus rhizosphere microbiome consisting of 132 genera (Xu *et al.*, 2018).

Industrial and environmental microbiome studies can also utilize large-scale spatial sampling to test for widely distributed core microbiomes. For example, an extensive study of 51 wastewater treatment plants from five countries did not identify any core microbial OTUs present across all samples, but showed that anaerobic digester conditions, rather than geographic location, were significant correlated with microbial community composition (Mei *et al.*, 2017). Similarly, a study of six soil types from 24 locations in eastern Europe showed that only two OTUs were shared across all samples, while many taxa appeared to be core to individual soil types (Pershina *et al.*, 2018).

Temporal Scale of sampling

22.3% of the articles in our sample included some form of temporal replication, ranging from three days (Sweet *et al.*, 2017) to six years (Saunders *et al.*, 2016). Of these, the majority

were sampled over fewer than two months, likely due to the challenges of repeated sampling across a prolonged period of time (Fig. 2f). However, such studies provide unique insight into whether core microbiomes persist in a particular location or host system despite potential changes in diet, development, or environment.

Free-living environmental microbiomes are particularly amenable to this type of repeated temporal sampling. Long-term sampling allows for the determination of how microbial communities fluctuate over time (e.g., 76), as well as the taxa that persist through major events, such as hurricanes in the upper troposphere (DeLeon-Rodriguez *et al.*, 2013) and beach oiling in the Gulf of Mexico (Newton *et al.*, 2013). Such studies can also reveal the level of stability present in the microbiome of a particular environment, as exemplified by a study of soda lakes in British Columbia, which found a highly persistent core microbiome in these lakes over the course of four years (Zorz *et al.*, 2019). Similarly, in long-lived host species such as humans, it is possible to sample the same individual through time to determine the level of community stability within the host. Studies of this type have identified particular OTUs that are able to persist over multi-year periods in the guts and mouths of certain individuals (Caporaso *et al.*, 2011; Martínez *et al.*, 2013).

Long-term studies of hosts with shorter generation times, or where repeated sampling proves difficult, require different approaches. One approach is to sample multiple individuals of an organism across different stages of the life cycle to determine if any core microbial taxa persist throughout. Studies using this method have shown that chickens maintain a large set of core microbial genera throughout their lives (Ding *et al.*, 2017) and Atlantic salmon maintain a relatively small core microbiome across the freshwater-to-saltwater transition (Rudi *et al.*, 2018), while parasites experience high levels of microbiome turnover when moving between hosts

during their life cycle (Jorge *et al.*, 2020). A second method is to sample individuals of the same species at the same site over multiple time horizons. Although this type of sampling cannot determine which microbial taxa persist over time in a specific individual of a given host, any shared OTUs identified using this approach represent a species or population level core microbiome that is repeatedly acquired from the surrounding environment or vertically transmitted from parent to offspring. For example, monthly sampling of six sponge species over the course of three years showed that host species with denser microbial loads harbored larger core communities (Björk *et al.*, 2018). Similarly, populations of the intertidal bivalve *Donax gouldii*, collected at four time points over eleven years, only had six ASVs present across all time points, indicating a small but persistent core (Neu, Hughes, *et al.*, 2021).

Taxonomic resolution of the core microbiome

Determining the taxonomic resolution of the core microbiome begins prior to sequencing, as the selection of 16S rRNA gene region and corresponding PCR primers play a significant role in the composition of the microbiome (Abellan-Schneyder *et al.*, 2021). Certain primer pairs may miss particular microbial groups or provide lower resolution due to limited coverage in taxonomic databases. Further, taxonomic assignments provided by one primer set may not directly correspond to those from a different set, making core microbiome comparisons across studies difficult (Abellan-Schneyder *et al.*, 2021). Sequence length also determines the taxonomic level available for analysis, as shorter reads (e.g., 100 base pairs) provide less specific taxonomic information and, therefore, sequences may not be able to be assigned at lower taxonomic levels. Maximizing sequence length by merging paired-end reads, or generating full-

length 16S rRNA gene amplicons using long-read technologies (e.g., Callahan *et al.*, 2021), are potential ways to increase the taxonomic resolution available for further analysis.

Regardless of sequencing strategy, though, the taxonomic level chosen for core microbiome quantification has important implications for the ecological and functional relevance of the core. For example, core microbiomes identified at the phylum level, as in four of the studies sampled here (Fig. 2d), may offer limited insights about the specific ecological and functional roles of those microbes. On the other hand, in nearly 50% of the studies in our dataset the core was determined using OTUs clustered at the 97% level, reflecting the popularity of this sequence similarity cutoff in delineating microbial OTUs (Fig. 2d). This proportion may be even higher, as bacterial species and phylotypes are often classified at the 97% sequence similarity level but were included as separate groups in these analyses when not explicitly defined. Interestingly, only ten studies included a more stringent cutoff (99% or 100% OTUs) (Fig. 2d) despite the fact that OTUs generated using 100% sequence similarity [also known as ASVs or zero-radius OTUs (zOTUs)] have become a commonly used taxonomic descriptor with the increasing use of denoising applications such as deblur (Amir *et al.*, 2017) and DADA2 (Callahan *et al.*, 2016). This difference could reflect, in part, a lag in the use of these units for core microbiome study, as they were not available until 2016 or 2017 (SI Appendix, Fig. S1). Alternatively, 97% OTUs, or higher taxonomic levels such as genera, may be more appropriate for determining the core microbiome since 100% sequence similarity may be too stringent for delineating functional or ecological differences among sequences (Shade and Stopnisek, 2019). For example, two sequences with a single base-pair variation would be identified as different OTUs using the 100% cutoff but would be collapsed into a single 97% OTU. Further, certain microbial taxa are known to contain multiple copies of the 16S rRNA gene, which can vary

intragenomically by >1% sequence identity (Coenye and Vandamme, 2003; Větrovský and Baldrian, 2013). Thus, using the 100% OTU cutoff may inflate the size of the core microbiome and increase potential redundancy, particularly if core taxa contain many copies of the 16S rRNA gene (e.g., Gammaproteobacteria (Větrovský and Baldrian, 2013)). On the other hand, it has recently been shown that as many as 16 unique microbial strains with diverse temperature preferences and carbohydrate utilization profiles can cluster into a single 97% OTU (Chase *et al.*, 2017). The limited resolution of databases used for taxonomic identification may also decrease the utility of determining the core at lower taxonomic levels, particularly in non-human samples, as many microbial groups are still unknown or uncharacterized and therefore cannot provide much information about community function. In more than 20% of cases, though, the authors quantified core microbiomes using multiple different taxonomic criteria. This is done by either (i) clustering OTUs at different levels of sequence similarity (e.g., Dougal *et al.*, 2013) or (ii) using the levels of taxonomic classification produced by a database (e.g., greengenes (DeSantis *et al.*, 2006) or Silva (Quast *et al.*, 2013)). Multiple classifications have the potential to provide additional information regarding the strength of the association between the microbial taxa and the host/environment. For example, if some functions are conserved at higher taxonomic levels than a microbial genus, they may play a core functional role within a host, but individual OTUs within that genus may have a more limited occurrences and not appear as core members.

Choice of host taxa

Nearly 88% of studies in our sample set focused on microbiomes associated with plant and animal hosts (including various micro-environments within humans and other hosts), rather than environmental or industrial samples. Most frequently, core microbiomes were identified for

149

a single host species, but 42 studies attempted to identify such cores using two or more host species of varying phylogenetic relatedness. In some cases, the phylogenetic relationships among the hosts are known, and a shared core microbiome represents a potential co-phylogenetic history or conserved functional/ecological role across species. For example, 18 "types" of *Symbiodinium* dinoflagellate spanning at least 8 species were found to share three highly abundant microbial OTUs, which were predicted to play various roles in nutrient acquisition and stress tolerance (Lawson *et al.*, 2018). Similar analyses using 36 strains of *Leptocylindrus* diatoms found that only one microbial OTU, a member of the genus *Roseovarius*, was shared by all of the strains (Ajani *et al.*, 2018). Even more broadly, one study tested whether a highly diverse set of sponge hosts (32 species) share a core microbiome, but found no evidence for such conservatism (Schmitt *et al.*, 2012).

Instead of phylogenetic affinities, some studies have tested whether hosts with similar traits (e.g., mode of digestion, diet) share core microbial taxa. Ruminant mammals, for example, have been found to share a number of core genera across a wide geographic range, which are hypothesized to play a role in digestion and fermentation (Henderson *et al.*, 2015; O' Donnell *et al.*, 2017). Similarly, a phylogenetically diverse group of cycad-eating insects were found to share five microbial OTUs in common, at least one of which was predicted to provide functional benefits for digestion of cycad tissue (Salzman *et al.*, 2018).

Synthesis and Future Directions

Determining the most effective way to quantify the core microbiome remains challenging, with some arguing that a taxonomic approach is no longer useful and that a core functional microbiome should be prioritized (Lemanceau *et al.*, 2017). However, despite the

utility in understanding the core functional properties of the microbiome, a taxonomic approach remains highly practical for a few important reasons. First, the per-sample cost of 16S rRNA gene sequencing remains much lower than the cost to generate the high-quality metagenomes and/or meta-transcriptomes necessary for detailed functional profiling. Second, this approach is better suited when macroecological relationships, such as abundance-occurrence models of microbial taxa, are used for constraining the composition of the core microbiome. Finally, a taxonomic core microbiome provides a list of potentially ecologically relevant taxa which can be prioritized for targeted culturing and/or 'omics study. This also allows for the development of testable hypotheses about the roles of these organisms within the microbiome.

However, as shown by this representative sampling of the literature, quantifying a core microbiome is not straightforward. Perhaps more importantly, though, is the fact that these differences in methodology for quantifying the core change its functional definition. For example, abundance-driven metrics prioritize the most dominant members of the community, or those that have most effectively colonized a particular environment. However, this type of core likely overlooks a number of ecologically and/or functionally important but low abundance taxa and may be skewed by high levels of variance. On the other hand, occurrence-based metrics often require taxa to be present in every sample to be counted in the core, which can miss relevant taxa due to inadequate sequencing depth or sampling effects. Relaxing this criterion necessarily requires using an arbitrary cutoff for the number of samples a microbial OTU must be present in to count as part of the core. These different cutoff values include taxa with different levels of association to the host/environment and different ecological roles, thereby changing the way the core is defined. Methods which combine abundance and occurrence have been introduced to overcome some of these challenges (e.g., Li *et al.*, 2013), and are well-supported in

the broader ecological literature, but many still require arbitrary occurrence and abundance cutoffs, which vary widely across studies. Such combined methods which use modeling approaches represent a potentially significant advancement, but their broad applicability to microbial taxa is not yet certain. In addition to specific metrics for quantifying a core, the issue of spatial and taxonomic grain at which the core should be determined also remains fluid. As discussed above, the spatial extent of sampling can have a strong impact on the makeup of the core microbiome at the population or species level, especially for hosts that have large geographic distributions. Similarly, determining the taxonomic level at which the core should be quantified also remains a challenge. Although 97% OTUs are the most commonly used taxonomic units in our dataset, cores are also routinely identified using ASVs, genera or phyla, thus introducing different levels of potential functional and ecological redundancy into the core microbiome and changing its practical definition. One issue that has largely been ignored when quantifying taxonomic cores is the fact that even consistent associations between a host and a group of microbes identified by any of the methods discussed above could still simply reflect repeated acquisitions of microbes from the environment by the host rather than obligate host-microbe relationships where microbes play important functional roles (Hammer *et al.*, 2017). Experiments, quantitative PCR and microscopy techniques can all be used to test whether a putative core microbiome is a stable component of the host physiology (Martin *et al.*, 2020) and/or plays a significant functional role in the host (Hammer *et al.*, 2017) but such approaches remains rare.

In summary, our review of the literature clearly shows that the term "core microbiome" represents different things to different researchers, which makes comparative analyses and meta-analyses of the core microbiome across hosts and environments very difficult, if not impossible.

While a single metric is unlikely to capture all the different aspects of core microbiomes, we hope that the information provided is a useful starting point for the development of measures of the core microbiome that are robust to sampling, sequencing depth and other issues discussed here. Such metrics are necessary not only for testing specific hypotheses about the functional and ecological roles of core microbiomes but also for understanding the general nature of core microbiomes and the ecological and evolutionary processes that generate and maintain these stable associations between certain microbes and their hosts. Finally, we provide a set of recommendations below that could serve as the starting point for achieving this goal.

1. Explicitly define and state the criteria used for determining the core microbiome in the Methods section of the manuscript. As noted in this review, these methods provide important context for interpreting the results but are often not adequately described.

2. When conducting spatial analyses, explicitly distinguish between local, regional and range-wide cores (e.g., Burgsdorf *et al.*, 2014; Hernandez-Agreda *et al.*, 2016). When conducting temporal analyses, explicitly distinguish between short-term, seasonal and multi-year cores. They each require their own contextual definitions but provide important information about the potential spatial and/or temporal stability of any core associations.

3. Sequence as deeply as possible and/or ensure an adequate number of sequencing replicates to determine the core. Standardizing the size of samples (pre-DNA extraction) or molarity (for library pooling) of samples are potential strategies to achieve uniform sequencing depth. In spatial datasets, it may be possible to

compensate for variable sequencing depths using the range-through approach described above.

4. Rarefying samples to a common sequencing depth is best avoided while quantifying the core microbiome, especially when variation in sequencing depth across samples is large. Such an approach can lead to underestimates of core size and inaccurate core composition (Ramakodi, 2021).

5. The use of macroecological null models for constraining the makeup of the core microbiome should be based on adequate spatial sampling coverage, especially for widely distributed microbial taxa. Most existing analyses of core microbiomes do not have enough spatial and/or temporal coverage for computing meaningful macroecological relationships.

References

Abellan-Schneyder, I., Matchado, M.S., Reitmeier, S., Sommer, A., Sewald, Z., Baumbach, J., et al. (2021) Primer, Pipelines, Parameters: Issues in 16S rRNA Gene Sequencing. *mSphere* **6**: e01202-20.

Ainsworth, T.D. and Gates, R.D. (2016) Corals' microbial sentinels. *Science* **352**: 1518–1519.

Ainsworth, T.D., Krause, L., Bridge, T., Torda, G., Raina, J.-B., Zakrzewski, M., et al. (2015) The coral core microbiome identifies rare bacterial taxa as ubiquitous endosymbionts. *ISME J* **9**: 2261–2274.

Aira, M., Pérez-Losada, M., and Domínguez, J. (2019) Microbiome dynamics during cast ageing in the earthworm Aporrectodea caliginosa. *Appl Soil Ecol* **139**: 56–63.

Ajani, P.A., Kahlke, T., Siboni, N., Carney, R., Murray, S.A., and Seymour, J.R. (2018) The microbiome of the cosmopolitan diatom leptocylindrus reveals significant spatial and temporal variability. *Front Microbiol* **9**: 2758.

Alberdi, A. and Gilbert, M.T.P. (2019) A guide to the application of Hill numbers to DNA-based diversity analyses. *Mol Ecol Resour* **19**: 804–817.

Amir, A., McDonald, D., Navas-Molina, J.A., Kopylova, E., Morton, J.T., Xu, Z.Z., et al. (2017) Deblur Rapidly Resolves Single-Nucelotide Community Sequence Patterns. *mSystems* **2**: e00191-16.

Antwis, R.E., Lea, J.M.D., Unwin, B., and Shultz, S. (2018) Gut microbiome composition is associated with spatial structuring and social interactions in semi-feral Welsh Mountain ponies. *Microbiome* **6**: 207.

Astudillo-García, C., Bell, J.J., Webster, N.S., Glasl, B., Jompa, J., Montoya, J.M., and Taylor, M.W. (2017) Evaluating the core microbiota in complex communities: A systematic investigation. *Environ Microbiol* **19**: 1450–1462.

Bäckhed, F., Fraser, C.M., Ringel, Y., Sanders, M.E., Sartor, R.B., Sherman, P.M., et al. (2012) Defining a healthy human gut microbiome: Current concepts, future directions, and clinical applications. *Cell Host Microbe* **12**: 611–622.

Banerjee, S., Schlaeppi, K., and van der Heijden, M.G.A. (2018) Keystone taxa as drivers of microbiome structure and functioning. *Nat Rev Microbiol* **16**: 567–576.

Beckers, B., De Beeck, M.O., Weyens, N., Boerjan, W., and Vangronsveld, J. (2017) Structural variability and niche differentiation in the rhizosphere and endosphere bacterial microbiome of field-grown poplar trees. *Microbiome* **5**: 25.

Berg, G., Rybakova, D., Fischer, D., Cernava, T., Vergès, M.C.C., Charles, T., et al. (2020) Microbiome definition re-visited: old concepts and new challenges. *Microbiome* **8**: 103.

Björk, J.R., O'Hara, R.B., Ribes, M., Coma, R., and Montoya, J.M. (2018) The dynamic core microbiome: Structure, dynamics and stability. *bioRxiv* 137885.

Brodie, J., Williamson, C., Barker, G.L., Walker, R.H., Briscoe, A., and Yallop, M. (2016) Characterising the microbiome of Corallina officinalis, a dominant calcified intertidal red alga. *FEMS Microbiol Ecol* **92**: fiw110.

Brown, J.H. (1995) Macroecology, Chicago: University of Chicago Press.

Brown, J.H., Mehlman, D.W., and Stevens, G.C. (1995) Spatial variation in abundance. *Ecology* **76**: 2028–2043.

Burgsdorf, I., Erwin, P.M., Lopez-Legentil, S., Cerrano, C., Haber, M., Frenk, S., and Steindler, L. (2014) Biogeography rather than association with cyanobacteria structures symbiotic microbial communities in the marine sponge Petrosia ficiformis. *Front Microbiol* **5**: 529.

Callahan, B.J., Grinevich, D., Thakur, S., Balamotis, M.A., and Yehezkel, T. Ben (2021) Ultra-accurate microbial amplicon sequencing directly from complex samples with synthetic long reads. *Microbiome* **9**: 130.

155

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

Caporaso, J.G., Lauber, C.L., Costello, E.K., Berg-Lyons, D., Gonzalez, A., Stombaugh, J., et al. (2011) Moving pictures of the human microbiome. *Genome Biol* **12**: R50.

Chase, A.B., Karaoz, U., Brodie, E.L., Gomez-Lunar, Z., Martiny, A.C., and Martiny, J.B.H. (2017) Microdiversity of an abundant terrestrial bacterium encompasses extensive variation in ecologically relevant traits. *MBio* **8**: e01809-17.

Chiu, C.-H. and Chao, A. (2016) Estimating and comparing microbial diversity in the presence of sequencing errors. *PeerJ* **4**: e1634.

Choudoir, M.J., Doroghazi, J.R., and Buckley, D.H. (2016) Latitude delineates patterns of biogeography in terrestrial Streptomyces. *Environ Microbiol* **18**: 4931–4945.

Christian, K., Weitzman, C., Rose, A., Kaestli, M., and Gibb, K. (2018) Ecological patterns in the skin microbiota of frogs from tropical Australia. *Ecol Evol* **8**: 10510–10519.

Coenye, T. and Vandamme, P. (2003) Intragenomic heterogeneity between multiple 16S ribosomal RNA operons in sequenced bacterial genomes. *FEMS Microbiol Lett* **228**: 45–49.

Dehler, C.E., Secombes, C.J., and Martin, S.A.M. (2017) Seawater transfer alters the intestinal microbiota profiles of Atlantic salmon (Salmo salar L.). *Sci Rep* **7**: 13877.

DeLeon-Rodriguez, N., Lathem, T.L., Rodriguez-R, L.M., Barazesh, J.M., Anderson, B.E., Beyersdorf, A.J., et al. (2013) Microbiome of the upper troposphere: Species composition and prevalence, effects of tropical storms, and atmospheric implications. *Proc Natl Acad Sci U S A* **110**: 2575–2580.

DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., et al. (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* **72**: 5069–5072.

Ding, J., Dai, R., Yang, L., He, C., Xu, K., Liu, S., et al. (2017) Inheritance and establishment of gut microbiota in chickens. *Front Microbiol* **8**: 1967.

Dougal, K., de la Fuente, G., Harris, P.A., Girdwood, S.E., Pinloche, E., and Newbold, C.J. (2013) Identification of a Core Bacterial Community within the Large Intestine of the Horse. *PLoS One* **8**: e77660.

Edwards, J., Johnson, C., Santos-Medellín, C., Lurie, E., Podishetty, N.K., Bhatnagar, S., et al. (2015) Structure, variation, and assembly of the root-associated microbiomes of rice. *Proc Natl Acad Sci U S A* **112**: E911–E920.

Engel, P. and Moran, N.A. (2013) Functional and evolutionary insights into the simple yet specific gut microbiota of the honey bee from metagenomic analysis. *Gut Microbes* **4**: 60–65.

Eren, A.M., Borisy, G.G., Huse, S.M., and Mark Welch, J.L. (2014) Oligotyping analysis of the human oral microbiome. *Proc Natl Acad Sci U S A* **111**: E2875–E2884.

Faith, J.J., Guruge, J.L., and Charbonneau, M. (2013) The long-term stability of the human gut microbiota. *Science* **341**: 1237439.

Gaston, K.J., Blackburn, T.I.M.M., Greenwood, J.D., Gregory, R.D., Quinn, M., and Lawton, J.H. (2000) Abundance-occupancy relationships. *J Appl Ecol* **37**: 39–59.

Gibbons, S.M., Caporaso, J.G., Pirrung, M., Field, D., Knight, R., and Gilbert, J.A. (2013) Evidence for a persistent microbial seed bank throughout the global ocean. *Proc Natl Acad Sci U S A* **110**: 4651–4655.

Gilbert, J.A., Field, D., Swift, P., Newbold, L., Oliver, A., Smyth, T., et al. (2009) The seasonal structure of microbial communities in the Western English Channel. *Environ Microbiol* **11**: 3132–3139.

Gloor, G.B., Macklaim, J.M., Pawlowsky-Glahn, V., and Egozcue, J.J. (2017) Microbiome datasets are compositional: And this is not optional. *Front Microbiol* **8**: 2224.

Hamady, M. and Knight, R. (2009) Microbial community profiling for human microbiome projects: Tools, techniques, and challenges. *Genome Res* **19**: 1141–1152.

Hammer, T.J., Janzen, D.H., Hallwachs, W., Jaffe, S.P., and Fierer, N. (2017) Caterpillars lack a resident gut microbiome. *Proc Natl Acad Sci U S A* **114**: 9641–9646.

Hanski, I. (1982) Dynamics of Regional Distribution: The Core and Satellite Species Hypothesis. *Oikos* **38**: 210–221.

Henderson, G., Cox, F., Ganesh, S., Jonker, A., Young, W., and Janssen, P.H. (2015) Rumen microbial community composition varies with diet and host, but a core microbiome is found across a wide geographical range. *Sci Rep* **5**: 14567.

Hernandez-Agreda, A., Leggat, W., Bongaerts, P., and Ainsworth, T.D. (2016) The microbial signature provides insight into the mechanistic basis of coral success across reef habitats. *MBio* **7**: e00560-16.

Huse, S.M., Ye, Y., Zhou, Y., and Fodor, A.A. (2012) A core human microbiome as viewed through 16S rRNA sequence clusters. *PLoS One* **7**: e34242.

Hutchins, D.A., Jansson, J.K., Remais, J. V., Rich, V.I., Singh, B.K., and Trivedi, P. (2019) Climate change microbiology — problems and perspectives. *Nat Rev Microbiol* **17**: 391–

396.

Jiang, Y., Xiong, X., Danska, J., and Parkinson, J. (2016) Metatranscriptomic analysis of diverse microbial communities reveals core metabolic pathways and microbiomespecific functionality. *Microbiome* **4**: 2.

Jorge, F., Dheilly, N.M., and Poulin, R. (2020) Persistence of a Core Microbiome Through the Ontogeny of a Multi-Host Parasite. *Front Microbiol* **11**: 954.

Jousset, A., Bienhold, C., Chatzinotas, A., Gallien, L., Gobet, A., Kurm, V., et al. (2017) Where less may be more: How the rare biosphere pulls ecosystems strings. *ISME J* **11**: 853–862.

Kim, J., Kim, D., Lim, J.S., Maeng, J.H., Son, H., Kang, H.C., et al. (2019) The use of technical replication for detection of low-level somatic mutations in next-generation sequencing. *Nat Commun* **10**: 1047.

Lawson, C.A., Raina, J.B., Kahlke, T., Seymour, J.R., and Suggett, D.J. (2018) Defining the core microbiome of the symbiotic dinoflagellate, Symbiodinium. *Environ Microbiol Rep* **10**: 7–11.

Leis, M.L. and Costa, M.O. (2019) Initial description of the core ocular surface microbiome in dogs: Bacterial community diversity and composition in a defined canine population. *Vet Ophthalmol* **22**: 337–344.

Lemanceau, P., Blouin, M., Muller, D., and Moënne-Loccoz, Y. (2017) Let the Core Microbiota Be Functional. *Trends Plant Sci* **22**: 583–595.

Li, K., Bihan, M., and Methé, B.A. (2013) Analyses of the Stability and Core Taxonomic Memberships of the Human Microbiome. *PLoS One* **8**: e63139.

Lindh, M. V., Sjöstedt, J., Ekstam, B., Casini, M., Lundin, D., Hugerth, L.W., et al. (2017) Metapopulation theory identifies biogeographical patterns among core and satellite marine bacteria scaling from tens to thousands of kilometers. *Environ Microbiol* **19**: 1222–1236.

Liu, J., Zhang, X., Chen, T., Wu, T., Lin, T., Jiang, L., et al. (2021) A semiparametric model for between-subject attributes: Applications to beta-diversity of microbiome data. *Biometrics* 1–13.

Lundberg, D.S., Lebeis, S.L., Paredes, S.H., Yourstone, S., Gehring, J., Malfatti, S., et al. (2012) Defining the core Arabidopsis thaliana root microbiome. *Nature* **488**: 86–90.

Martin, G.G., Natha, Z., Henderson, N., Bang, S., Hendry, H., and Loera, Y. (2020) Absence of a microbiome in the midgut trunk of six representative Crustacea. *J Crustac Biol* **40**: 122–130.

Martínez, I., Muller, C.E., and Walter, J. (2013) Long-Term Temporal Analysis of the Human

Fecal Microbiota Revealed a Stable Core of Dominant Bacterial Species. *PLoS One* **8**: e69621.

Martino, C., Morton, J.T., Marotz, C.A., Thompson, L.R., Tripathi, A., Knight, R., and Zengler, K. (2019) A Novel Sparse Compositional Technique Reveals Microbial Perturbations. *mSystems* **4**: e00016-19.

McKnight, D.T., Huerlimann, R., Bower, D.S., Schwarzkopf, L., Alford, R.A., and Zenger, K.R. (2019) Methods for normalizing microbiome data: An ecological perspective. *Methods Ecol Evol* **10**: 389–400.

McMurdie, P.J. and Holmes, S. (2014) Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. *PLoS Comput Biol* **10**:.

Mei, R., Nobu, M.K., Narihiro, T., Kuroda, K., Muñoz Sierra, J., Wu, Z., et al. (2017) Operation-driven heterogeneity and overlooked feed-associated populations in global anaerobic digester microbiome. *Water Res* **124**: 77–84.

Neu, A.T., Allen, E.E., and Roy, K. (2021) Do host-associated microbes show a contrarian latitudinal diversity gradient? Insights from Mytilus californianus , an intertidal foundation host. *J Biogeogr* **00**: 1–14.

Neu, A.T., Hughes, I. V., Allen, E.E., and Roy, K. (2021) Decade-scale stability and change in a marine bivalve microbiome. *Mol Ecol* **30**: 1237–1250.

Neu, A.T., Torchin, M.E., Allen, E.E., and Roy, K. (2021) Microbiome divergence of marine gastropod species sepatated by the Isthmus of Panama. *bioRxiv*.

Newton, R.J., Huse, S.M., Morrison, H.G., Peake, C.S., Sogin, M.L., and McLellan, S.L. (2013) Shifts in the Microbial Community Composition of Gulf Coast Beaches Following Beach Oiling. *PLoS One* **8**: e74265.

O' Donnell, M.M., Harris, H.M.B., Ross, R.P., and O'Toole, P.W. (2017) Core fecal microbiota of domesticated herbivorous ruminant, hindgut fermenters, and monogastric animals. *Microbiologyopen* **6**: e509.

Oh, J., Byrd, A.L., Park, M., Kong, H.H., and Segre, J.A. (2016) Temporal Stability of the Human Skin Microbiome. *Cell* **165**: 854–866.

Pereira-Marques, J., Hout, A., Ferreira, R.M., Weber, M., Pinto-Ribeiro, I., Van Doorn, L.J., et al. (2019) Impact of host DNA and sequencing depth on the taxonomic resolution of whole metagenome sequencing for microbiome analysis. *Front Microbiol* **10**: 1277.

Pershina, E. V., Ivanova, E.A., Korvigo, I.O., Chirak, E.L., Sergaliev, N.H., Abakumov, E. V., et al. (2018) Investigation of the core microbiome in main soil types from the East European plain. *Sci Total Environ* **631**–**632**: 1421–1430.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41**: D590–D596.

Ramakodi, M.P. (2021) Effect of Amplicon Sequencing Depth in Environmental Microbiome Research. *Curr Microbiol* **78**: 1026–1033.

Risely, A. (2020) Applying the core microbiome to understand host–microbe systems. *J Anim Ecol* **89**: 1549–1558.

Roy, K. and Witman, J.D. (2009) Spatial Patterns of Species Diversity in the Shallow Marine Invertebrates: Patterns, Processes, and Prospects. In *Marine macroecology*. Witman, J.D. and Roy, K. (eds). Chicago: The University of Chicago Press, pp. 101–121.

Rudi, K., Angell, I.L., Pope, P.B., Vik, J.O., Sandve, S.R., and Snipen, L.G. (2018) Stable core gut microbiota across the freshwater-to-saltwater transition for farmed Atlantic salmon. *Appl Environ Microbiol* **84**: e01974-17.

Sabree, Z.L., Hansen, A.K., and Moran, N.A. (2012) Independent studies using deep sequencing resolve the same set of core bacterial species dominating gut communities of honey bees. *PLoS One* **7**: e41250.

Salzman, S., Whitaker, M., and Pierce, N.E. (2018) Cycad-feeding insects share a core gut microbiome. *Biol J Linn Soc* **123**: 728–738.

Saunders, A.M., Albertsen, M., Vollertsen, J., and Nielsen, P.H. (2016) The activated sludge ecosystem contains a core community of abundant organisms. *ISME J* **10**: 11–20.

Schlaeppi, K., Dombrowski, N., Oter, R.G., Ver Loren van Themaat, E., and Schulze-Lefert, P. (2014) Quantitative divergence of the bacterial root microbiota in Arabidopsis thaliana relatives. *Proc Natl Acad Sci U S A* **111**: 585–592.

Schmitt, S., Tsai, P., Bell, J., Fromont, J., Ilan, M., Lindquist, N., et al. (2012) Assessing the complex sponge microbiota: core, variable and species-specific bacterial communities in marine sponges. *ISME J* **6**: 564–576.

Segata, N., Baldini, F., Pompon, J., Garrett, W.S., Truong, D.T., Dabiré, R.K., et al. (2016) The reproductive tracts of two malaria vectors are populated by a core microbiome and by gender- and swarm-enriched microbial biomarkers. *Sci Rep* **6**: 24207.

Sexton, J.P., McIntyre, P.J., Angert, A.L., and Rice, K.J. (2009) Evolution and ecology of species range limits. *Annu Rev Ecol Evol Syst* **40**: 415–436.

Shade, A. and Handelsman, J. (2012) Beyond the Venn diagram: The hunt for a core microbiome. *Environ Microbiol* **14**: 4–12.

Shade, A. and Stopnisek, N. (2019) Abundance-occupancy distributions to prioritize plant core microbiome membership. *Curr Opin Microbiol* **49**: 50–58.

Sogin, M.L., Morrison, H.G., Huber, J.A., Mark Welch, D., Huse, S.M., Neal, P.R., et al. (2006) Microbial diversity in the deep sea and the underexplored "rare biosphere." *Proc Natl Acad Sci* **103**: 12115–12120.

Stefanini, I., Carlin, S., Tocci, N., Albanese, D., Donati, C., Franceschi, P., et al. (2017) Core microbiota and metabolome of Vitis vinifera L. cv. Corvina grapes and musts. *Front Microbiol* **8**:.

Sullam, K.E., Essinger, S.D., Lozupone, C.A., and Connor, M.P.O. (2012) Environmental and ecological factors that shape the gut bacterial communities of fish: a meta-analysis. *Mol Ecol* **21**: 3363–3378.

Sweet, M.J., Brown, B.E., Dunne, R.P., Singleton, I., and Bulling, M. (2017) Evidence for rapid, tide-related shifts in the microbiome of the coral Coelastrea aspera. *Coral Reefs* **36**: 815–828.

Taylor, L.R. (1961) Aggregation, Variance and the Mean. *Nature* **189**: 732–735.

Turnbaugh, P.J., Hamady, M., Yatsunenko, T., Cantarel, B.L., Duncan, A., Ley, R.E., et al. (2009) A core gut microbiome in obese and lean twins. *Nature* **457**: 480–484.

Turnbaugh, P.J., Ley, R.E., Hamady, M., Fraser-Liggett, C., Knight, R., and Gordon, J.I. (2007) The human microbiome project: exploring the microbial part of ourselves in a changing world. *Nature* **449**: 804–810.

Turon, M., Cáliz, J., Triadó-Margarit, X., Casamayor, E.O., and Uriz, M.J. (2019) Sponges and their microbiomes show similar community metrics across impacted and well-preserved reefs. *Front Microbiol* **10**: 1961.

Větrovský, T. and Baldrian, P. (2013) The Variability of the 16S rRNA Gene in Bacterial Genomes and Its Consequences for Bacterial Community Analyses. *PLoS One* **8**: e57923.

Weiss, S., Xu, Z.Z., Peddada, S., Amir, A., Bittinger, K., Gonzalez, A., et al. (2017) Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome* **5**: 27.

Willing, B.P., Dicksved, J., Halfvarson, J., Andersson, A.F., Lucio, M., Zheng, Z., et al. (2010) A pyrosequencing study in twins shows that gastrointestinal microbial profiles vary with inflammatory bowel disease phenotypes. *Gastroenterology* **139**: 1844-1854.e1.

Xu, J., Zhang, Y., Zhang, P., Trivedi, P., Riera, N., Wang, Y., et al. (2018) The structure and function of the global citrus rhizosphere microbiome. *Nat Commun* **9**: 4894.

Yeoh, Y.K., Paungfoo-Lonhienne, C., Dennis, P.G., Robinson, N., Ragan, M.A., Schmidt, S., and Hugenholtz, P. (2016) The core root microbiome of sugarcanes cultivated under varying nitrogen fertilizer application. *Environ Microbiol* **18**: 1338–1351.

Zaura, E., Keijser, B.J.F., Huse, S.M., and Crielaard, W. (2009) Defining the healthy "core microbiome" of oral microbial communities. *BMC Microbiol* **9**: 259.

Zorz, J.K., Sharp, C., Kleiner, M., Gordon, P.M.K., Pon, R.T., Dong, X., and Strous, M. (2019) A shared core microbiome in soda lakes separated by large distances. *Nat Commun* **10**: 4230.
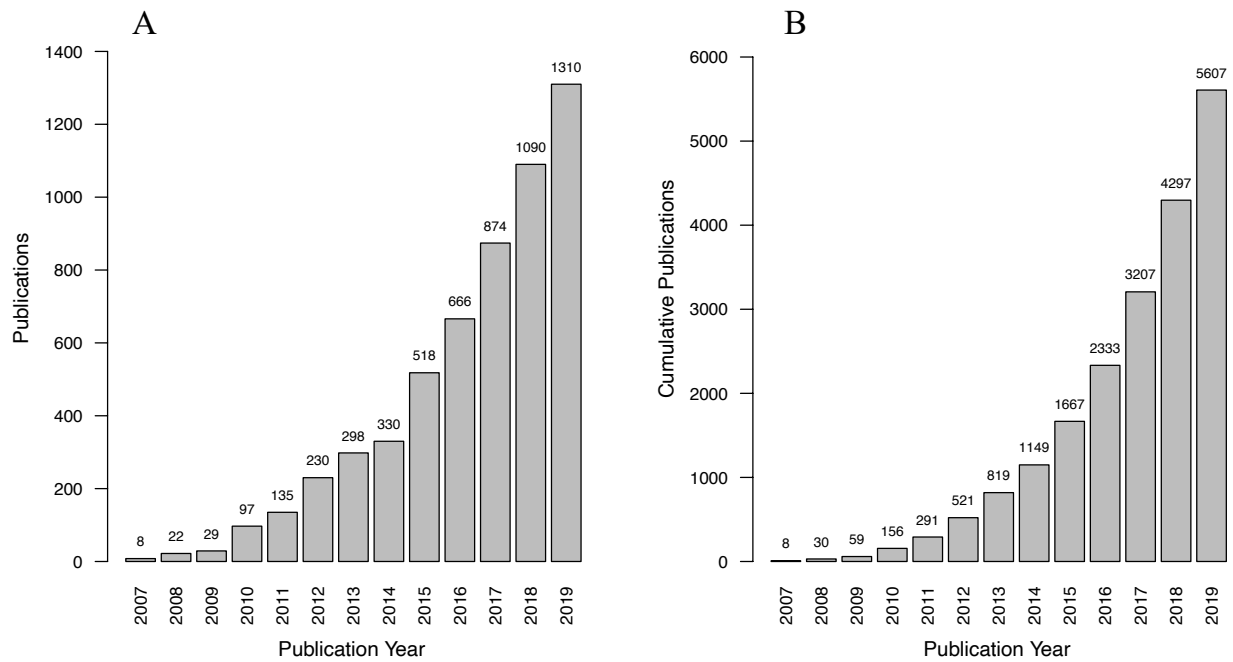
Figure 5.1. Number of publications per year (A) and cumulative (B) in Google Scholar including the terms "core microbiome" or "core microbiota" from the introduction of the term in 2007 to 2019 (search conducted on 12/22/2020).
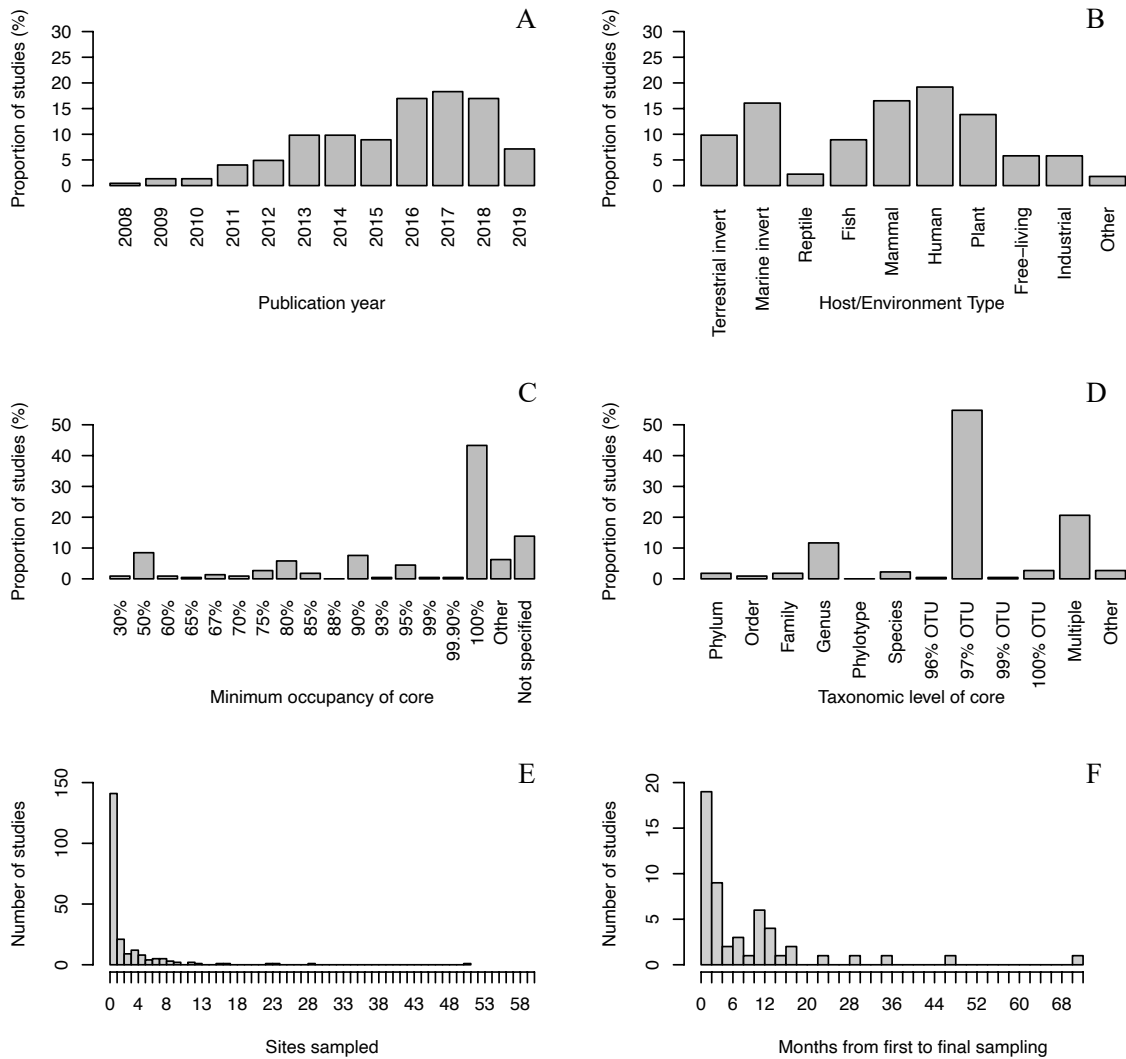
Figure 5.2. Barplots of the proportion of studies in our representative dataset (n = 224) from each publication year (A) and environment (B), as well as the minimum occupancy value (C, those with no explicitly defined occupancy value are counted as "Other") and taxonomic level (D) used to determine the core microbiome. Histograms illustrate the number of sites sampled in each study (E) and the number of months from first sampling to the completion of the project for those studies which included a temporal component (F; n = 51).

Table 5.1. Pros, cons and areas for improvement for the currently available methods of quantifying core microbiomes

Metric Type	Pros	Cons	Areas for improvement
Occurrence only	• Easily calculable • Commonly used in the literature • Include rare taxa	• No abundance information • Arbitrary cutoffs • Can be highly impacted by sampling coverage and sequencing depth	• Maximizing sequencing and replicate sampling • Use multiple occurrence cutoffs • Use range through approach
Relative abundance only	• Easily calculable • Attempt to determine core taxa unique from the surrounding environment	• Impacted by sequencing depth and other processing steps • Core membership could be impacted by mean and variance • Arbitrary cutoffs	• Use Taylor's Power Law approaches to account for stability • Make attempts to ensure uniform sequencing across samples
Abundance-occurrence	• Supported by the broader macroecological literature • New methods (including code) are being developed in this space • Can potentially differentiate the stochastic from the deterministically selected core	• Often use arbitrary cutoffs of abundance and occurrence • Assume that microbial taxa follow similar patterns as plants and animals	• Determine whether macroecological patterns in plants and animals are applicable to microbial taxa • Constrain analyses by phylogenetic or functional group • Conduct large-scale, representative sampling across the range or time period of interest
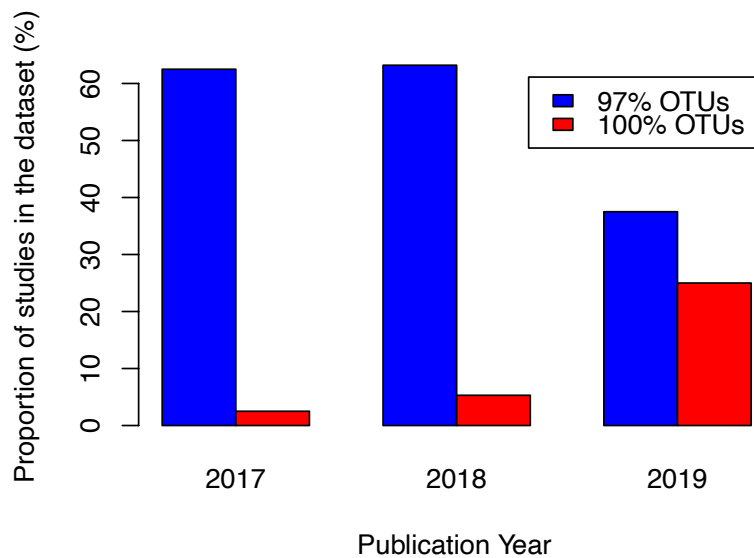
Supplemental Figure 5.1. Proportion of studies in our representative dataset from 2007-2019 which include core microbiomes defined at the 97% OTU and 100% OTU (also called ASVs or zOTUs) levels. Columns in each year do not sum to 100% as other taxonomic levels were also used.

Acknowledgements

We thank the members of the K. Roy and E. Allen labs for helpful discussions on this topic.

Chapter 5, in full, is in review as Neu, A.T., Allen, E.E. and Roy, K., 2021. Defining and quantifying the core microbiome: challenges and prospects. In review. The dissertation author was the primary investigator and author of this paper.