MECHANISMS OF ACCLIMATION AND ADAPTATION IN THE SEA URCHIN *ECHINOMETRA* SP. EZ

by

Remi Nikita Ketchum

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Biological Sciences

Charlotte

2021

Approved by:

____
Dr. Adam Reitzel

____
Dr. Joseph Ryan

____
Dr. Bao-Hua Song

____
Dr. López-Duarte

____
Dr. Daniel Janies

ABSTRACT

REMI N. KETCHUM. Mechanisms of acclimation and adaptation in the sea urchin *Echinometra* sp. EZ. (Under the direction of ADAM M. REITZEL)

Climate change has resulted in warming of coastal aquatic habitats around the world at almost every latitude, threatening ecosystems with a significant loss in biodiversity and occurring at a rate that may exceed species' ability to adapt. Understanding how reef species survive in habitats that currently experience extreme temperatures will help identify the biological processes that will govern future responses to climate change. The Persian/Arabian Gulf experiences the warmest coral reef temperatures on the planet (summer maxima ~35-36°C but can exceed 37°C) and connects to the neighboring Gulf of Oman, which experiences more benign environmental conditions (summer maxima of ~30-32°C). Here, we leverage this unique environmental gradient as a natural laboratory to better understand how the keystone sea urchin *Echinometra* sp. *EZ* copes with thermal extremes. Species survival in extreme habitats is dependent on their ability to acclimate over the course of an organisms' lifetime and adapt over the course of many generations. Two complementary mechanisms for coping with environmental change are shifts in the host-associated microbial community, which can happen on a timescale of hours to days, and classic Darwinian evolution in which selection results in different patterns of alleles between populations over many generations. Here, we identify temperature as a robust predictor of community-level microbial variation and highlight specific bacterial taxa that may be crucial for maintenance of host homeostasis during thermal extremes. Next, we show that while there is a high degree of genetic admixture between all sites and bidirectional gene flow between the two seas, there is also significant population differentiation. We describe nine candidate loci that show evidence of positive selection, including one collagen gene. Finally, we sequence, assemble, and annotate a chromosome-level genome that will add substantial value to future functional genomic datasets. Together, the research composing my dissertation has identified the importance of novel microbiome and genomic variation in the adaptation of a dominant ecosystem engineer to the warmest marine environment on Earth. These integrative results provide a foothold for understanding shared and unique mechanisms for the adaptation of marine species to historic and ongoing climate change.

ACKNOWLEDGMENTS

Adam, you are the best PhD advisor I could have asked for and I cannot thank you enough for everything you have done to support me these last five years. I hope that your boundless patience, scientific curiosity, and kindness has been passed on to me in some way, shape, or form.

Additionally, I am fortunate to have had so many more amazing mentors who have made my work possible. Mr. Tyree, thank you for taking the time to show me how cool science is and for just being kind. Youssef, your depth of knowledge and excitement for all things science is absolutely contagious and your confidence in my ability gave me the assurance that I needed to start graduate school. John, thank you for your time, constant support, continuous mentorship over the past ten years, and your sense of humor. My committee members, Dr. Song, Dr. López-Duarte, Dr. Ryan, and Dr. Janies, thank you for your input, insights, and encouragement. Last but not least, I would also like to thank all the other researchers who have assisted, inspired, and pushed me throughout my graduate career. I also acknowledge and appreciate the support from the Department of Biological Sciences, the Graduate School at UNC Charlotte, and the National Science Foundation.

I am not quite sure how to adequately thank my family and friends, but I will try (in no particular order)! Tess, you are unwaveringly supportive and are my constant reminder to take up more space (and I know you would never *ever* let me hunt for horcruxes all by myself). Caroline, Rory, and Jess, thank you for keeping me sane with long calls, puppy videos, snorts, laughs, and a whole lot of love! Grace, Dain, Gashel, Noura, Whitney, Devin, and Quinton - you have all made graduate school so much more

fun, have always been there to help me hunt for creatures, have provided a place to sleep when I'm extremely stinky, and have supported me at my most irritable (usually before 10AM). Alex and Bob, you both have been a home away from home for the last five years and I love you for it. Grandma, I miss you all the time and wish you had left me contact information for the trolls – how are they doing? Kona, thank you for letting me be your emotional support human, I promise to repay you in stinky treats. Thank you to all my butthead siblings, Anika, Jack, Hannah, and Lily, for pretending to listen while I talk about science. My parents - from you I have learned grit, kindness (yes, *really*), to laugh at myself (daily), that you don't ever *really* need to grow up, and probably many other invaluable lessons. Thank you for supporting me in every way possible. And finally, thank you to my favorite fieldwork buddy, my code-editor, my in-home reviewer #3, and my endlessly supportive partner, Ed. I could not have gotten through a PhD without your 'sense of humor.'

TABLE OF CONTENTS

CHAPTER 2: UNRAVELING THE PREDICTIVE ROLE OF TEMPERATURE IN THE GUT MICROBIOTA OF THE SEA URCHIN *ECHINOMETRA* SP. *EZ* ACROSS SPATIAL AND TEMPORAL GRADIENTS

CHAPTER 3: POPULATION GENOMIC ANALYSES OF THE SEA URCHIN *ECHINOMETRA* SP. *EZ* ACROSS AN EXTREME ENVIRONMENTAL GRADIENT

CHAPTER 4: CHROMOSOME-LEVEL GENOME ASSEMBLY OF THE HIGHLY HETEROZYGOUS SEA URCHIN *ECHINOMETRA* SP. EZ

LIST OF TABLES

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

LIST OF FIGURES

CHAPTER 1

CHAPTER 2

CHAPTER 3

Introduction

Coping with climate change

Climate change has resulted in accelerated changes of coastal aquatic habitats around the world at almost every latitude, threatening ecosystems with a significant loss in biodiversity and occurring at a rate that may exceed species' ability to respond. A central component of climate change is rising temperatures, which are expected to increase the frequency of chronic and acute thermal stress events due to higher baseline temperatures and extreme heat events. Temperature governs all biological processes at every level of biological organization for ectotherms, from changes in molecular functioning to behavioral variation (Somero and Hochachka, 1976). Changes in temperature affect basic cellular functioning through the fluidity of lipid membranes, the conformational activities of proteins, and the stability of DNA duplexes (Hochachka and Somero, 2002). Because thermal effects have such pervasive impacts on cellular functioning, organisms have developed extensive coping mechanisms. Generally, these mechanisms include migration towards thermal optima, phenotypic acclimation (gene expression modulation, epigenetic modifications, or microbial restructuring) and genomic adaptation across species' range. The latter mechanisms are particularly important for organisms such as sessile or nearly sessile marine benthic invertebrates that are unable to move to more favorable environments. As such, marine benthic invertebrates (MBI's) represent important study systems to understand acclimation and adaptation to elevated temperatures.

Two mechanisms for coping with environmental change are shifts in the associated microbial community, which can happen on a timescale of hours to days, and classic Darwinian evolution in which selection results in different patterns of alleles between populations over many generations. In recent years, microbiome research has increased dramatically and unlocked tremendous insights into the role of microbial communities and their relationships with their hosts. The microbiome has been shown to impact host physiology and metabolism, has been linked to several diseases, and can confer a competitive advantage to their hosts when acclimating to different abiotic stressors (Gatesoupe, 1999;Bayer et al., 2008;Hentschel et al., 2012;Ziegler et al., 2017b). In well-studied marine benthic organisms such as corals, the microbiome has been shown to be involved in metabolic cycling, pathogen defense, and thermal tolerance. For example, Ziegler et al. showed that coral microbiomes from the American Samoa are different across thermally variable habitats and highlight specific bacterial taxa that are indicative of the coral host response to thermal stress experiments (Ziegler et al., 2017b). These findings have led to a theory called the Coral Probiotic Hypothesis which suggests that the coral microbiome can change in response to environmental conditions and form the most advantageous coral holobiont composition (Reshef et al., 2006;Voolstra and Ziegler, 2020). This selection of the 'metaorganism' emphasizes the microbiomes' functional importance and the need for better characterization of MBI microbiomes and their response to thermally stressful environments.

The second mechanism is local adaptation, or natural selection operating on genetic diversity over a species' geographic range that results in different phenotypes. Local adaptation results from the interplay between evolutionary forces such as gene flow, natural selection, genetic drift, and mutation (Blanquart et al., 2013). Spatially heterogeneous environments acting on genetically isolated populations are thought to be a prerequisite for the emergence of local adaptation (Blanquart et al., 2013). Indeed, much of our understanding of local adaptation has resulted from taxa whose life history strategies restrict gene flow. In these taxa, genes involved in cell membrane formation, metabolism, protein folding, endoplasmic reticulum function and immune responses are generally thought to be key genes for the thermal stress response (e.g.,Yampolsky et al. (2014); Guzman and Conaco, (2016)). It is currently unclear how species that, over the course of their lifespan, experience a wide temperature range and ongoing gene flow respond to thermal stress on a genomic level.

The sea urchin genus *Echinometra*

Echinometra is the most widespread genus of sea urchin and their geographic boundaries and phylogenetic history are well described in the literature (Lessios, 2006). *Echinometra* forms a monophyletic group with two principle clades; the eastern Pacific to Atlantic clade, and the Indo-West Pacific clade (species level taxonomy is still not complete)(McClanahan and Muthiga, 2007). The Indo-West species formed during the Pleistocene and speciated roughly 1-3 million years ago (McClanahan and Muthiga, 2007). To date, most of the *Echinometra* literature has focused on reconstructing the evolutionary history of speciation events with phylogenies and species-specific fertilization patterns. Perhaps the most well-described group of *Echinometra* urchins is the species complex *E*. sp. A-D which are found on Okinawan reefs.

Comparative studies of the *Echinometra* sp. A-D have shown significant morphological, ecological, and reproductive differences between this species complex uniquely found sympatry in the waters around Okinawa (Matsuoka and Toshihiko, 1991;Metz et al., 1994;Appana et al., 2004;Mita et al., 2004;Rahman et al., 2012;Bronstein and Loya, 2013). Morphological studies of adult coloration, spine morphology, and sperm structure have consistently been used to differentiate these species in the waters around Okinawa as well as in other locations in the Indo-West Pacific (Landry et al., 2003;Bronstein and Loya, 2013). Although these species live sympatrically, previous studies have shown differences in local distribution around the island as well as variation in feeding preferences (Hiratsuka and Uehara, 2007). Reproductive assays comparing these four species have uncovered unique reproductive dynamics in which there are pronounced species-specific fertilization patterns. Experimental hybridization assays have been conducted between *E.* sp. D and *E*. sp. A, *E*. sp. C and *E* sp. A, *E.* sp. D and *E.* sp. B, and *E*. sp. C and *E.* sp. B (Aslan and Uehara, 1997;Rahman et al., 2000;Rahman et al., 2001;Rahman and Uehara, 2004;Rahman et al., 2004;Rahman et al., 2005). In all of these combinations, ova from the former species can be fertilized by sperm of the latter, but the reciprocal crosses are not as successful (Rahman et al., 2012). Single locus studies have shown that *E*. sp. B and *E*. sp. D are *E. mathaei* and *E. oblonga*, respectively (McClanahan and Muthiga, 2007).

In addition to the sister species complex *E*. sp. A-D, *E. lucunter*, *E*. *viridis* and *E. vanbrunti* also have been researched extensively. Following the formation of the Isthmus of Panama, *Echinometra* speciated into *E. vanbrunti* in the Pacific and *E. lucunter* and *E. viridis* in the Caribbean (the two Caribbean species formed after the formation of the

Isthmus; McCartney et al. (2000)). Of these three species, *E. lucunter* is the only species which has eggs that are fertilized poorly by heterospecific sperm. It was hypothesized that the gametic incompatibility evolved only once in the lineage leading to *E. lucunter* because otherwise, the fertilization block would be shared with *E. viridis* (and *E. viridis* and *E. vanbrunti* are interfertile; McCartney et al. (2000)). While there has been substantial effort to understand species differences in both the A-D species complex and the species on either side of the Isthmus of Panama, studies have generally focused on phylogenetic analysis based on short mitochondrial sequences or nuclear regions, reproductive assays, or morphological comparisons. While informative, these types of studies do not provide a mechanistic understanding of the microbial and genomic processes that promote acclimation and adaptation. Further, they do not encompass all of the species within the *Echinometra* genus.

The number of valid species in the *Echinometra* genus and the associated methods for species identification has been debated within the scientific literature for over 180 years. Recently, it was proposed that *Echinometra* urchins from Eilat and Zanzibar should be regarded as a new species (Bronstein and Loya, 2013). Although this species has not yet been formally described, Bronstein and Loya (2013) referred to them as *E.* sp. *EZ* (for Eilat and Zanzibar). This species was also found in the PAG (Ketchum et al., 2018a) where it had been misidentified as *E. mathaei* (Figure 1). *E.* sp. *EZ* is the most abundant sea urchin in the PAG (Figure 2; densities averaging 8.6m^{-2} across eight sites between 2015-2019, [Burt, JA, unpublished data]) and they play a significant role in the health and dynamics of coral reef ecosystems in the region as major bioeroders (Downing and El-Zahr, 1987). Where densities of these sea urchins are low, there is often a proliferation of algae and a community shift towards larger fleshy macro algae (Bauman et al., 2016). The few studies performed on *Echinometra* in the PAG have focused on metal bioaccumulation (Sadiq et al., 1996), reproductive cycles (Alsaffar and Lone, 2000), and spatial variation across different substrates (Bauman et al., 2016). Prior to this dissertation research, there were no studies on the microbial dynamics of the *E*. sp. *EZ* microbiome, their population dynamics, or their capacity for adaptation. Given the ecological importance, large population numbers, and dearth of knowledge on this species in this unique region, *E*. sp. *EZ* represents an excellent model system with which to study the mechanisms of adaptation and acclimation.

The Persian/Arabian Gulf

Understanding how current reef species survive in habitats with extreme thermal conditions will help identify the biological processes that will govern future responses to climate change. The Persian/Arabian Gulf (hereafter PAG) is an extreme environment that experiences the warmest coral reef temperatures on the planet. Temperatures around 34°C for several months annually are common and summer maxima can exceed 37°C (Smith et al., 2017c;Burt et al., 2019). The PAG is connected to the neighboring Gulf of Oman (hereafter GO) by the narrow Strait of Hormuz (42 km wide) which experiences more benign temperatures with mean monthly maximum less than 32°C (Coles, 2003). Modern coastlines in the PAG only formed ~6,000 years ago following the Holocene transgression (Lambeck, 1996) and this system therefore represents an excellent system in which to study the contributions of adaptation and acclimation to a species' persistence to variable environments. Further, oceanographic conditions between the PAG and GO are unique in that the water entering the PAG travels as a surface current along the Iranian coast and the water leaving the PAG travels as a subsurface current (Reynolds, 1993). Residence times in the PAG can last up to three years and may limit exchange of larvae between the two seas (Alosairi et al., 2011). Several studies have shown that the PAG is unique in terms of its biological community structure (Burt et al., 2011;Bauman et al., 2013), but while there have been a few studies exploring the mechanisms of acclimation and adaptation in the region, they have been limited to corals, coral symbionts, and fish.

Given the young age of the PAG and the geographic isolation from the wider Indian Ocean, it is currently unclear if organisms have acclimated or adapted to the extreme PAG environment. One study on cryptobenthic fishes from the two seas found that fishes from the PAG have reduced diversity, abundance, and body condition with those compared to the GO (Brandl et al., 2020). They also found evidence for intraspecific thermal plasticity which was hypothesized to ensure survival in PAG conditions (Brandl et al.). Further, a study of the coral *Platygyra daedalea* showed that DNA methylation is heritable and congruent with parental environment which may suggest epigenetic acclimatization to local conditions (Liew et al., 2020). The only microbiome study on PAG animals was conducted on the coral *Porites lobata* from the PAG and Red Sea (Hadaidi et al., 2017). They found that the bacterial community was highly similar in bleached and healthy corals and that bacterial taxa between samples were almost identical. Although there have been no microbial studies on sea urchins in the PAG, studies from other regions have shown that microbial gut communities in sea urchins may be relatively stable when exposed to elevated temperatures (only shown in one species; Brothers et al. (2018)), vary with habitat and resource availability (Miller et al., 2021), and vary across different sections of the gut (Hakim et al., 2019). Investigating these communities further holds great potential to explicate specific microbes that play a key role in resilience to thermal stress.

The first step towards understanding the processes involved in adaptation to thermal extremes is to understand the connectivity between populations because restricted gene flow can lead to local adaptation. One study on *P. daedalea* and their symbiotic algae found significant population structuring in the coral host between the two seas and a symbiont community dominated by a *Symbiodinium thermophilum* variant which is known to be thermally tolerant (Howells et al., 2016;Smith et al., 2017a). Another study on three species of *Porites* coral found that their symbiont communities were also dominated by *S. thermophilum*, which they hypothesized to contribute to the coral holobionts' thermal tolerance (D'angelo et al., 2015). Population genetic structure has also been found in the sea urchin *Diadema setosum* and several species of fishes (Lessios et al., 2001;Hoolihan et al., 2004;Torquato et al., 2019). While there has been research investigating species persistence to the PAGs' extreme temperatures, these studies have focused on a small number of species with specific life history traits.

In order to fully understand the mechanisms driving persistence of biological communities, it is crucial to study species across the animal tree of life that employ a wide range of life history strategies (i.e., population size, dispersal patterns, mating systems, philopatry, and reproductive timing). Studies conducted in the PAG have so far been largely focused on fish that employ a direct development strategy which typically means they have lower dispersal potential and corals which have lecithotrophic larvae which have a comparatively greater dispersal potential. However, species with planktotrophic larvae, like sea urchins, have the highest dispersal potential because the duration of the pelagic period can range from weeks to months. Species with large dispersal potential should, in theory, have more homogenous population structure and thus, limited local adaptation. Currently, there is a lack of clear insight into how different life history traits affect gene flow, genetic diversity, and subsequent adaptation in the PAG. *Echinometra* sea urchins are common in the PAG, have planktotrophic larvae with pelagic larval durations (PLD) of about 18-30 days (McClanahan and Muthiga, 2007), and are understudied in the region despite their importance for reef ecosystems.

Overview of Dissertation Research and Organization

My dissertation research focused on determining the microbial community dynamics and population genomics of the species *Echinometra* sp. EZ. Through a combination of field collections and sequence-based studies, I investigated four major areas of research to understand the potential contributions of the microbiome and the genome to acclimation and adaptation of this *Echinometra* species in the PAG. The four data chapters for my dissertation are:

Chapter 1: In this chapter, I tested the most effective method to capture difficult to lyse bacterial taxa during DNA extractions in order to develop a ‘gold standard’ methodological approach for microbial studies. I found that the inclusion of a lysozyme and bead-beating step resulted in an increase in taxa such as gram-positive bacteria which have tougher cell walls. I also generated a literature synthesis to highlight the wide variety of approaches used to characterize the microbiome. This variation in methodology is hindering the field because different methods (e.g., DNA extractions and technical variation) can lead to differences in the inferred microbial communities compromising comparisons across studies.

Chapter 2: In this chapter, I use the method developed in chapter one to investigate the role of temperature in the gut microbiome of *E.* sp. EZ. I generated two independent datasets with a high degree of spatial and temporal resolution (seven reefs, two seas, and eight months). I found that temperature plays a critical role in gut microbiome and that increasing temperatures can result in a more disperse microbiome. Further, I described a consistent relationship between specific bacterial taxa (which mostly consisted of Vibrio's) and rising temperatures which may point to these strains playing an important role in the host's stress response.

Chapter 3: In this chapter, I investigated the population genetic structure of *E.* sp. EZ from seven sites along the Arabian Peninsula. I showed that, despite the young age of the PAG and the dispersal potential of *E.* sp. EZ, there were two main populations along this environmental gradient corresponding to each respective sea (PAG and Gulf of Oman). I hypothesized that this population structuring might be a result of the strong selective pressure on species living within the PAG. I also showed evidence of bidirectional admixture occurring between the two seas and implicate a gene related to collagen production as being under putative natural selection. Finally, I described a draft genome that I had assembled for this species and describe extreme heterozygosity present in the genome.

Chapter 4: In this chapter, I improve upon the draft genome generated through chapter three by assembling and annotating a chromosome-level genome for *E.* sp. EZ. The genome contains 21 chromosomes and has an assembly size of 817.8 Mb. The contiguity of this assembly is shown by the scaffold N50 of 39.5 Mb and a BUSCO completeness score of 95.3%. I conducted a comparison of defensome gene family composition in *E*. sp. EZ relative to four other echinoids and found that there are two nuclear receptor genes missing from the *E.* sp. EZ genome and putative regulatory regions of two different key defensome genes are under strong positive selection.

References

Alosairi, Y., Imberger, J., & Falconer, R. A. (2011). Mixing and flushing in the Persian Gulf (Arabian Gulf). *Journal of Geophysical Research: Oceans, 116*(C3). *Bulletin of Marine Science, 67*(2), 845-856.

Appana, S. D., Vuki, V. C., & Cumming, R. L. (2004). Variation in abundance and spatial distribution of ecomorphs of the sea urchins, *Echinometra* sp. nov A and *E.* sp. nov C on a Fijian reef. *Hydrobiologia, 518*(1), 105-110. doi:10.1023/B:HYDR.0000025060.18736.bf

Aslan, L. M., & Uehara, T. (1997). Hybridization and F1 backcrosses between two closely related tropical species of sea urchins (genus *Echinometra*) in Okinawa. *Invertebrate Reproduction & Development, 31*(1-3), 319-324. doi:10.1080/07924259.1997.9672591

Bauman, A. G., Dunshea, G., Feary, D. A., & Hoey, A. S. (2016). Prickly business: abundance of sea urchins on breakwaters and coral reefs in Dubai. *Marine Pollution Bulletin, 105*(2), 459-465. doi:https://doi.org/10.1016/j.marpolbul.2015.11.026

Bauman, A. G., Feary, D. A., Heron, S. F., Pratchett, M. S., & Burt, J. A. (2013). Multiple environmental factors influence the spatial distribution and structure of reef communities in the northeastern Arabian Peninsula. *Marine Pollution Bulletin, 72*(2), 302-312.

Bayer, K., Schmitt, S., & Hentschel, U. (2008). Physiology, phylogeny and in situ evidence for bacterial and archaeal nitrifiers in the marine sponge *Aplysina aerophoba*. *Environmental Microbiology, 10*(11), 2942-2955.

Blanquart, F., Kaltz, O., Nuismer, S. L., & Gandon, S. (2013). A practical guide to measuring local adaptation. *Ecology Letters, 16*(9), 1195-1205.

Brandl, S. J., Johansen, J. L., Casey, J. M., Tornabene, L., Morais, R. A., & Burt, J. A. (2020). Extreme environmental conditions reduce coral reef fish biodiversity and productivity. *Nature Communications, 11*(1), 1-14.

Bronstein, O., & Loya, Y. (2013). The taxonomy and phylogeny of *Echinometra* (Camarodonta: Echinometridae) from the Red Sea and Western Indian Ocean. *PLOS One, 8*(10), e77374. doi:10.1371/journal.pone.0077374

Brothers, C. J., Van Der Pol, W. J., Morrow, C. D., Hakim, J. A., Koo, H., & McClintock, J. B. (2018). Ocean warming alters predicted microbiome functionality in a common sea urchin. *Proceedings of the Royal Society B, 285*(1881), 20180340.

Burt, J. A., Feary, D. A., Bauman, A. G., Usseglio, P., Cavalcante, G. H., & Sale, P. F. (2011). Biogeographic patterns of reef fish community structure in the northeastern Arabian Peninsula. *ICES Journal of Marine Science, 68*(9), 1875-1883.

Burt, J. A., Paparella, F., Al-Mansoori, N., Al-Mansoori, A., & Al-Jailani, H. (2019). Causes and consequences of the 2017 coral bleaching event in the southern Persian/Arabian Gulf. *Coral Reefs, 38*(4), 567-589.

Coles, S. L. (2003). Coral species diversity and environmental factors in the Arabian Gulf and the Gulf of Oman: a comparison to the Indo-Pacific region. *Atoll Research Bulletin*.

D'angelo, C., Hume, B. C., Burt, J., Smith, E. G., Achterberg, E. P., & Wiedenmann, J. (2015). Local adaptation constrains the distribution potential of heat-tolerant Symbiodinium from the Persian/Arabian Gulf. *The ISME Journal, 9*(12), 2551-2560.

Downing, N., & El-Zahr, C. (1987). Gut evacuation and filling rates in the rock-boring sea urchin, *Echinometra mathaei*. *Bulletin of Marine Science, 41*(2), 579-584.

Gatesoupe, F. J. (1999). The use of probiotics in aquaculture. *Aquaculture, 180*(1-2), 147-165.

Guzman, C., & Conaco, C. (2016). Gene expression dynamics accompanying the sponge thermal stress response. *PLOS One, 11*(10), e0165368.

Hadaidi, G., Röthig, T., Yum, L. K., Ziegler, M., Arif, C., Roder, C., . . . Voolstra, C. R. (2017). Stable mucus-associated bacterial communities in bleached and healthy corals of *Porites lobata* from the Arabian Seas. *Scientific Reports, 7*(1), 1-11.

Hakim, J. A., Schram, J. B., Galloway, A. W., Morrow, C. D., Crowley, M. R., Watts, S. A., & Bej, A. K. (2019). The purple sea urchin *Strongylocentrotus purpuratus* demonstrates a compartmentalization of gut bacterial microbiota, predictive functional attributes, and taxonomic co-occurrence. *Microorganisms, 7*(2), 35.

Hentschel, U., Piel, J., Degnan, S. M., & Taylor, M. W. (2012). Genomic insights into the marine sponge microbiome. *Nature Reviews Microbiology, 10*(9), 641-654.

Hiratsuka, Y., & Uehara, T. (2007). Feeding ecology of four species of sea urchins (genus *Echinometra*) in Okinawa. *Bulletin of Marine Science, 81*, 85-100.

Hochachka, P. W., & Somero, G. N. (2002). *Biochemical adaptation: mechanism and process in physiological evolution*: Oxford University Press.

Hoolihan, J., Premanandh, J., D'Aloia-Palmieri, M.-A., & Benzie, J. (2004). Intraspecific phylogeographic isolation of Arabian Gulf sailfish *Istiophorus platypterus* inferred from mitochondrial DNA. *Marine Biology, 145*(3), 465-475.

Howells, E. J., Abrego, D., Meyer, E., Kirk, N. L., & Burt, J. A. (2016). Host adaptation and unexpected symbiont partners enable reef□building corals to tolerate extreme temperatures. *Global Change Biology, 22*(8), 2702-2714.

Ketchum, R. N., DeBiasse, M. B., Ryan, J. F., Burt, J. A., & Reitzel, A. M. (2018). The complete mitochondrial genome of the sea urchin, *Echinometra* sp. *EZ. Mitochondrial DNA Part B, 3*(2), 1225-1227.

Lambeck, K. (1996). Shoreline reconstructions for the Persian Gulf since the last glacial maximum. *Earth and Planetary Science Letters, 142*(1-2), 43-57.

Landry, C., Geyer, L., Arakaki, Y., Uehara, T., & Palumbi, S. R. (2003). Recent speciation in the Indo–West Pacific: rapid evolution of gamete recognition and sperm morphology in cryptic species of sea urchin. *Proceedings of the Royal Society of London B: Biological Sciences, 270*(1526), 1839-1847.

Lessios, H. (2006). Speciation in sea urchins. *Echinoderms: Durham Proceedings of the 12th International Echinoderm Conference* 91-101.

Lessios, H. A., Kessing, B. D., & Pearse, J. S. (2001). Population structure and speciation in tropical seas: global phylogeography of the sea urchin *Diadema*. *Evolution, 55*(5), 955-975.

Liew, Y. J., Howells, E. J., Wang, X., Michell, C. T., Burt, J. A., Idaghdour, Y., &

Aranda, M. (2020). Intergenerational epigenetic inheritance in reef-building corals. *Nature Climate Change, 10*(3), 254-259.

Matsuoka, N., & Toshihiko, H. (1991). Molecular evidence for the existence of four sibling species with the sea-urchin, *Echinometra mathaei* in Japanese waters and their evolutionary relationships. *Zoological Science, 8*(1), 121-133.

McCartney, M. A., Keller, G., & Lessios, H. A. (2000). Dispersal barriers in tropical oceans and speciation in Atlantic and eastern Pacific sea urchins of the genus *Echinometra*. *Molecular Ecology, 9*(9), 1391-1400.

McClanahan, T. R., & Muthiga, N. A. (2007). Chapter 15 Ecology of *Echinometra*. In J. M. Lawrence (Ed.), *Developments in Aquaculture and Fisheries Science* (Vol. 37, pp. 297-317): Elsevier.

Metz, E. C., Kane, R. E., Yanagimachi, H., & Palumbi, S. R. (1994). Fertilization between closely related sea urchins is blocked by incompatibilities during sperm-egg attachment and early stages of fusion. *The Biological Bulletin, 187*(1), 23-34. doi:10.2307/1542162

Miller, P. M., Lamy, T., Page, H. M., & Miller, R. J. (2021). Sea urchin microbiomes vary with habitat and resource availability. *Limnology and Oceanography Letters, 6*(3), 119-126.

Mita, M., Uehara, T., & Nakamura, M. (2004). Speciation in four closely related species of sea urchins (genus *Echinometra*) with special reference to the acrosome reaction. *Invertebrate Reproduction & Development, 45*(3), 169-174. doi:10.1080/07924259.2004.9652588

Palumbi, S. R., Grabowsky, G., Duda, T., Geyer, L., & Tachino, N. (1997). Speciation and population genetic structure in tropical Pacific sea urchins. *Evolution, 51*(5), 1506-1517. doi:doi:10.1111/j.1558-5646.1997.tb01474.x

Rahman, M. A., & Uehara, T. (2004). Interspecific hybridization and backcrosses between two sibling species of Pacific sea urchins (genus *Echinometra*) on Okinawan intertidal reefs. *Zoological Studies, 43*(1), 93-111.

Rahman, M. A., Uehara, T., Arshad, A., Yusoff, F. M., & Shamsudin, M. N. (2012). Absence of postzygotic isolating mechanisms: evidence from experimental hybridization between two species of tropical sea urchins. *Journal of Zhejiang University SCIENCE B, 13*(10), 797-810. doi:10.1631/jzus.B1100152

Rahman, M. A., Uehara, T., & Aslan, L. M. (2000). Comparative viability and growth of hybrids between two sympatric species of sea urchins (Genus *Echinometra*) in Okinawa. *Aquaculture, 183*(1), 45-56. doi:https://doi.org/10.1016/S0044-8486(99)00283-5

Rahman, M. A., Uehara, T., & Pearse, J. S. (2001). Hybrids of two closely related tropical sea urchins (genus *Echinometra*): evidence against postzygotic isolating mechanisms. *The Biological Bulletin, 200*(2), 97-106. doi:10.2307/1543303

Rahman, M. A., Uehara, T., & Pearse, J. S. (2004). Experimental hybridization between two recently diverged species of tropical sea urchins, *Echinometra mathaei* and *Echinometra oblonga*. *Invertebrate Reproduction & Development, 45*(1), 1-14. doi:10.1080/07924259.2004.9652569

Rahman, M. A., Ueharaa, T., & Lawrence, J. M. (2005). Growth and heterosis of hybrids of two closely related species of Pacific sea urchins (Genus *Echinometra*) in Okinawa. *Aquaculture, 245*(1-4), 121-133.

Reshef, L., Koren, O., Loya, Y., Zilber□Rosenberg, I., & Rosenberg, E. (2006). The Coral

Probiotic Hypothesis. *Environmental Microbiology, 8*(12), 2068-2073.

Reynolds, R. M. (1993). Physical oceanography of the Gulf, Strait of Hormuz, and the Gulf of Oman—Results from the Mt Mitchell expedition. *Marine Pollution Bulletin, 27*, 35-59.

Sadiq, M., Mian, A., & Saji, A. (1996). Metal bioaccumulation by sea urchin (*Echinometra mathaei*) from the Saudi coastal areas of the Arabian Gulf: 2. Cadmium, copper, chromium, barium, calcium, and strontium. *Bulletin of Environmental Contamination and Toxicology, 57*(6), 964-971.

Smith, E. G., Hume, B. C., Delaney, P., Wiedenmann, J., & Burt, J. A. (2017). Genetic structure of coral-Symbiodinium symbioses on the world's warmest reefs. *PLOS One, 12*(6), e0180169.

Smith, E. G., Vaughan, G. O., Ketchum, R. N., McParland, D., & Burt, J. A. (2017). Symbiont community stability through severe coral bleaching in a thermally extreme lagoon. *Scientific Reports, 7*(1), 2428. doi:10.1038/s41598-017-01569-8

Somero, G., & Hochachka, P. (1976). Biochemical Adaptations to Temperature. In *Adaptation to environment: essays on the physiology of marine animals* (pp. 125-190): Butterworths London.

Torquato, F., Range, P., Ben□Hamadou, R., Sigsgaard, E. E., Thomsen, P. F., Riera, R., . .

. Marshell, A. (2019). Consequences of marine barriers for genetic diversity of the coral□specialist yellowbar angelfish from the Northwestern Indian Ocean. *Ecology and Evolution*.

Voolstra, C. R., & Ziegler, M. (2020). Adapting with Microbial Help: Microbiome flexibility facilitates rapid responses to environmental change. *BioEssays, 42*(7), 2000004.

Yampolsky, L. Y., Zeng, E., Lopez, J., Williams, P. J., Dick, K. B., Colbourne, J. K., & Pfrender, M. E. (2014). Functional genomics of acclimation and adaptation in response to thermal stress in *Daphnia*. *BMC Genomics, 15*(1), 1-12.

Ziegler, M., Seneca, F. O., Yum, L. K., Palumbi, S. R., & Voolstra, C. R. (2017). Bacterial community dynamics are linked to patterns of coral heat tolerance. *Nature Communications, 8*, 14213. doi:10.1038/ncomms14213

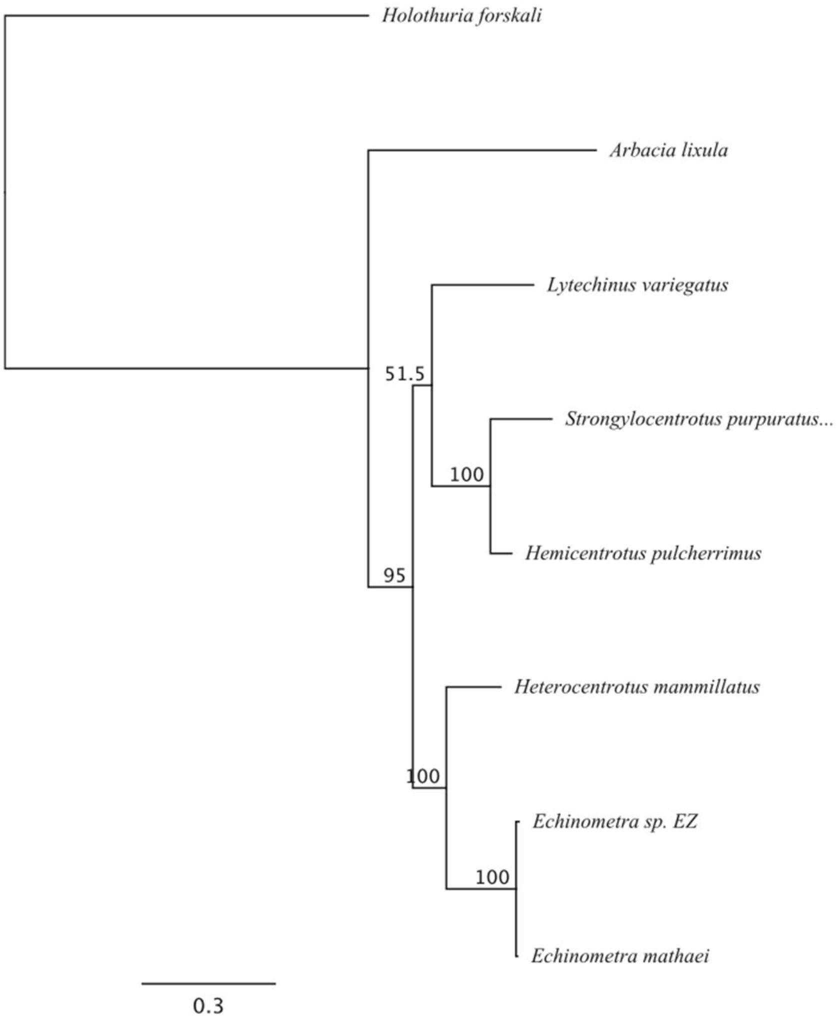


Figure 1: The maximum likelihood (ML) tree generated using Geneious v11.1.4 with six Echinoidea species: *Strongylocentrotus purpuratus* (accession number: X12631.1), *Lytechinus variegatus* (NC_037785.1), *Heterocentrotus mammillatus* (NC_034768.1), *Hemicentrotus pulcherrimus* (NC_023771.1), *Echinometra mathaei* (NC_034767.1), *Arbacia lixula* (NC_001770.1), and one Holothuroidea as an outgroup: *Holothuria forskali* (FN562582.1). The ML tree was generated with an alignment of the whole mitogenome sequences of all species, using the GTR þ G model. The numbers above the branches specify bootstrap percentages (1000 replicates).

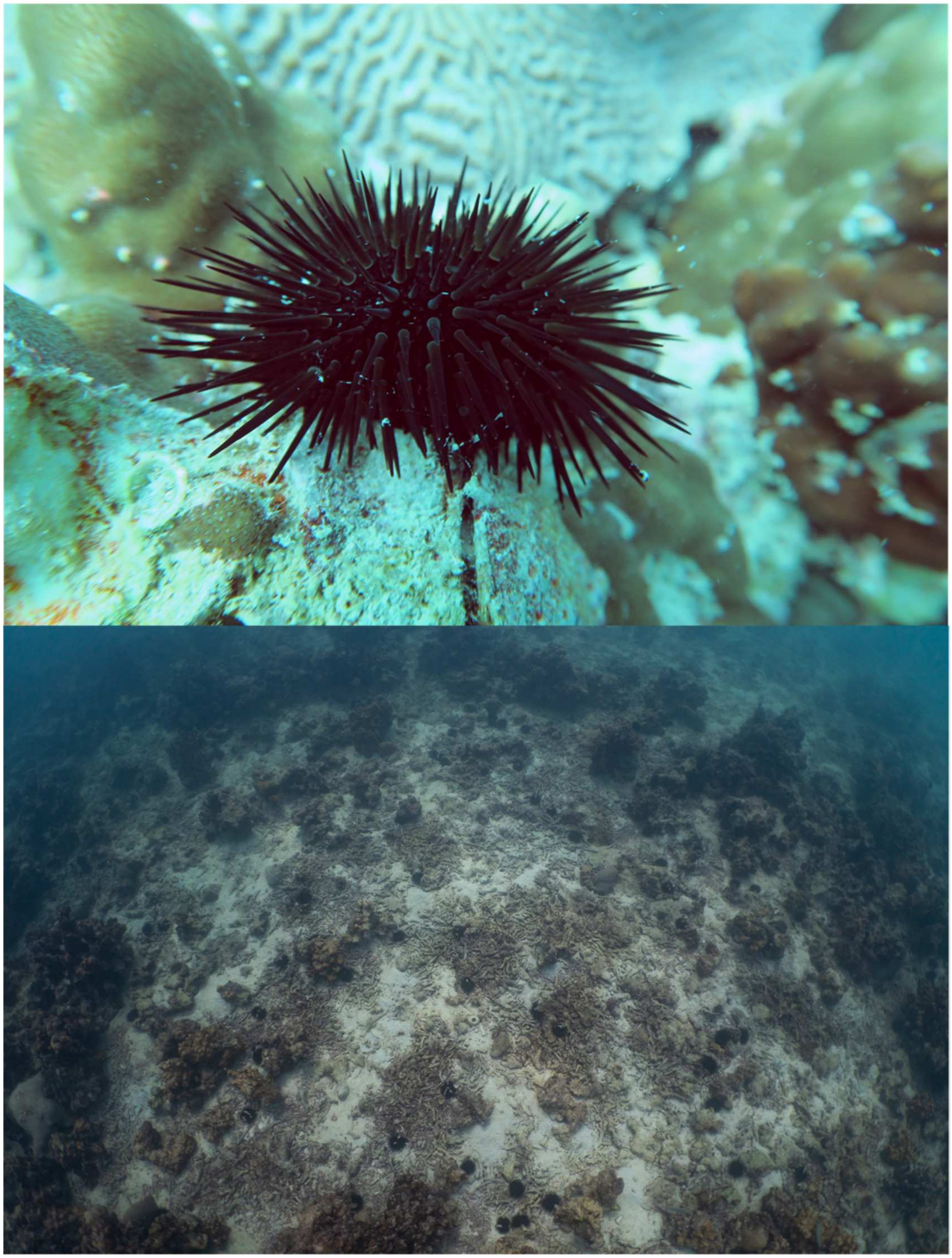


Figure 2: Top: *E.* sp. EZ in the Persian/Arabian Gulf. Photo credit: Grace O. Vaughan. Bottom: *E.* sp. EZ in the Persian/Arabian Gulf after a bleaching event. Photo credit: Noura Al-Mansoori.

CHAPTER 1

DNA EXTRACTION METHOD PLAYS A SIGNIFICANT ROLE WHEN DEFINING BACTERIAL COMMUNITY COMPOSITION IN THE MARINE INVERTEBRATE *ECHINOMETRA MATHAEI*

Remi N. Ketchum, Edward G. Smith, Grace O. Vaughan, Britney L. Phippen, Dain McParland, Noura Al-Mansoori, Tyler J. Carrier, John A. Burt, Adam M. Reitzel

Citation

Ketchum, Remi N., et al. "DNA extraction method plays a significant role when defining bacterial community composition in the marine invertebrate *Echinometra mathaei." Frontiers in Marine Science* 5 (2018): 255.

Abstract

The microbial assemblages of marine organisms play fundamental biological roles in their eukaryotic hosts. Studies aimed at characterizing this diversity have increased over the last decade and with the availability of high-throughput sequencing, we are now able to characterize bacteria that were non-culturable and, therefore, went undetected. With the number of marine microbiome studies growing rapidly, it is increasingly important to develop a set of "best practices" in order to accurately represent the bacterial communities present, and correct for biases. To address this, we sampled the gut communities of the pan-tropical echinoid *Echinometra mathaei* from two environmentally distinct populations along the Arabian Peninsula. We used three common DNA extraction procedures and compared inferred bacterial diversity from each method through 16S ribosomal RNA (rRNA) gene amplicon sequencing. Our results show that the addition of a bead-beating and lysozyme step more effectively capture traditionally difficult to lyse taxa, such as gram-positive bacteria. Further, DNA

extraction method plays an important role in estimates of Shannon diversity, with diversity indices significantly higher in both sites combined when a lysozyme and bead beating step was used. Finally, we conducted a literature synthesis to highlight the current diversity of approaches used to characterize the microbiome of marine invertebrates and found that the inclusion of a lysozyme treatment is uncommon (2% of surveyed studies), despite the importance of this step in recovery of rare OTUs as shown in our study.

Introduction

Bacteria associate with animals and can play a significant role in their development, immunity, metabolism, and physiology (McFall-Ngai et al., 2013;Bordenstein and Theis, 2015;Theis et al., 2016). In marine environments, the functional role of the microbiome has been associated with diverse processes including cycling of essential nutrients, the passage of trace minerals, production of secondary metabolites, and vitamin synthesis (reviewed by Bourne et al., 2016;Webster and Reusch, 2017). There has been an increasing amount of research focused on characterizing microbial community dynamics and shifts, and understanding how the functional role of the microbiome may co-vary with changes in the environment (Kohl and Carey, 2016;Carrier and Reitzel, 2017). For example, specific bacterial communities were found to be tightly correlated with thermal tolerance in a coral host (Ziegler et al., 2017b), expression levels of innate immunity genes in the threespine stickleback (Small et al., 2017), and geographic and depth gradients in a sponge species (Reveillaud et al., 2014). The Earth Microbiome Project (EMP) resulted in a set of recommended approaches for characterization of microbial taxa based on sequence analysis. While sponge microbial research has recently adopted the EMP protocol (Thomas et al., 2016;Moitinho-Silva et al., 2017), studies of associated bacterial communities of other marine invertebrate taxa currently lack a standardized approach. Sequence-based methods for describing associated microbes are sensitive to methodological steps and these differences may limit the ability to compare between studies or species. It is well established in human and environmental samples that bias can be introduced at each methodological step and significant research has gone into elucidating the role that methodology plays on interpretation of bacterial communities (McOrist et al., 2002;Yuan et al., 2012;Albertsen et al., 2015;de Bruin and Birnboim, 2016;Pollock et al., 2018). The variation in methodology that currently exists can be confounding in many ways, including differences in inferred microbial communities due to methods (species differ in efficacy for DNA extraction), information differences (different regions of 16S rRNA or metagenomics), and technical variation (PCR replicates and purifications). However, the majority of this research has been conducted on the human microbiome, activated sludge, and soil (Miller et al., 1999;McOrist et al., 2002;Vanysacker et al., 2010;Yuan et al., 2012). The role of methodology in the microbial communities of marine organisms remains comparatively understudied.

While the impact of each step in a sequence-based analysis of the microbiome is likely important, extraction and amplification of nucleic acids is central in characterizing a mixed bacterial community. DNA extractions should, as accurately as possible, characterize the bacterial diversity present. For example, methods for DNA extraction in corals influence the DNA yield and inferred microbial community (Weber et al., 2017). Exclusion of particular steps (e.g., mechanical and/or chemical lysis) (Lesser and Walker,

1992) in extraction protocols may result in unsampled portions of the microbial community due to difficult to lyse bacteria (e.g., gram-positive bacteria or endospores), which may play important functional roles in their hosts. Extraction protocols that result in a greater portion of the bacterial community being represented are important for identifying common, core, and rare associated microbes.

Echinoderms are dominant members of benthic marine habitats throughout the world's oceans. Over the last few decades, research using histological approaches has reported subcuticular bacteria in all classes of echinoderms (e.g., Holland and Nealson, 1978;Kelly et al., 1995). The localization and apparent specificity of these bacteria has suggested that echinoderms select particular bacteria for these associations, but any specific functions have only rarely been described (Walker and Lesser, 1989;Lesser and Walker, 1992). Efforts to culture echinoderm-associated bacteria have been either unsuccessful or difficult to repeat (discussed in Kelly et al., 1995). Recent studies using sequence-based approaches have reported diverse microbial communities that differ between species and developmental stage for asteroids (Galac et al., 2016) and echinoids (Carrier and Reitzel, 2018). Shifts in the microbial community associated with echinoid spines was reported for adults cultured under different temperature and pH conditions (Webster et al., 2016). Microbial communities of adult digestive systems of holothuroids (Gao et al., 2014), and echinoids (Meziti et al., 2007;Nelson et al., 2010;Hakim et al., 2015;Hakim et al., 2016a) have identified specific microbial communities located in different regions of the digestive system. Together, these previous studies suggest that echinoderms associate with unique microbial communities that may shift in response to different environmental conditions.

Here, we determined the microbial community in the hind-gut of *Echinometra mathaei*, a keystone urchin species (Uthicke et al., 2013), from two reef populations along the Arabian Peninsula. One population, inside the Persian/Arabian Gulf (Miller et al.), Dhabiya, experiences extreme environmental conditions, while the other, inside the Gulf of Oman (GO), Al Aqah, experiences cooler thermal maxima and a more narrow range of temperatures (Smith et al., 2017b). Sampling from these two sites allowed us to determine the effect of DNA extraction method in conspecifics across two contrasting environments in this species' contiguous geographic range. We use three different nucleic acid extraction techniques to compare bacterial communities and elucidate specific bacterial taxa that may be missed depending on the methods used. We then compare our approach with a literature survey of the methodologies currently used in studies with other marine invertebrates in order to place our results in the context of the broader community studying these microbial associations.

Materials and Methods

Sample collections

Adult *Echinometra mathaei* were sampled in November 2016 from two reef sites: Al Aqah (in the Gulf of Oman, GO) and Dhabiya (in the Persian/Arabian Gulf, PAG, see Figure 1.1). The sampling sites were chosen because they experience vastly different environmental conditions and they were as geographically distant (approximately 270 miles apart) as was possible given our sampling capabilities. Fifteen individuals were taken from each reef and placed in a 100 qt cooler filled with ~40 L of seawater until tissue extractions (within 1-2 hours). For gut extractions, individual urchins were then cut in half with a sterile knife, and a fragment of intestine located closest to the anus, and its contents, were removed with forceps and placed in RNAlater (Ambion). After one hour (to allow RNAlater to infiltrate the tissue), tubes were placed in a -20°C freezer.

DNA extraction

Prior to extraction, each sample tube was homogenized using a vortex (3x at 15 s) and mixed with a wide bore 100 µL pipette tip to ensure that each solution was uniform. This initial step reduced or eliminated bias that could occur simply through extraction of different sections of the sample. An aliquot of approximately 20 µL of the original sample (gut tissue and contents) was placed in three separate tubes, one for each different extraction method. Each method used the DNeasy Blood & Tissue Kit (QIAGEN). All extraction materials were autoclaved, UV sterilized, and DNA extractions and PCRs were performed under a PCR hood.

Method 1. The first extraction method used only the DNA extraction kit and followed the manufacturer's protocols for extracting animal tissue, with slight modification. We added 40 µL of proteinase K and 220 µL of AL buffer (instead of the recommended 20 µL and 200 µL, respectively) as recommended by the User-Developed Protocol for DNeasy Blood and Tissue Kit when samples are preserved in RNAlater. These modifications were kept consistent throughout each of the other two extraction methods.

Method 2. In addition to the kit described above, we added a bead beating step after addition of Buffer ATL and before placing sample on a heat block. We added 0.5 mL of 0.5 mm zirconia-silica beads (Fisher Scientific) and used the Bead Beater (BioSpec Products) for 40 s at 2,400-3,800 strokes/min and repeated this a total of three times.

Method 3. In addition to the kit and the bead beating step, Method 3 included the addition of 180 µL enzymatic lysis buffer; 20 mM TrisCL (pH 8), 2 mM sodium EDTA, 1.2% Triton, 20 mg/ml lysozyme (as described in DNeasy Blood and Tissue Manual). The buffer was added in place of Buffer ATL, incubated at 37°C and followed by the bead beating step (the extraction then proceeded as per the guidelines in the QIAGEN manual). Samples that did not contain sufficient DNA (< 2ng/µL) were either extracted again or concentrated using an ethanol precipitation. All samples were then quantified using a Qubit dsDNA High Sensitivity Assay Kit on the Qubit ® 2.0 Fluorometer and visualized using a 2% agarose gel. Prior to PCR amplification, all samples were normalized to a concentration of 1 ng/µL.

PCR amplification and sequencing

PCRs were performed in triplicate 25 µL reactions and pooled per individual sample. The 16S rRNA V3/V4 region was amplified using the universal V3/V4 PCR primers (forward: 5'- CTACGGGNGGCWGCAG-3'; reverse: 5'-GACTACHVGGGTATCTAATCC-3') (Klindworth et al., 2013). Each PCR contained 5 ng of DNA, 1.5 µL of forward primer (final concentration of 10 mM), 1.5 µL of reverse primer (final concentration of 10 mM), 12 µL of KAPA Biosciences HiFi HotStart Ready Mix, and 5 µL of water for a final volume of 25 µL. PCR cycling conditions were 95°C for 3 min, followed by 20 cycles of 95°C for 30 s, 50°C for 30 s, and 72 for 30 s, with a final extension time of 5 min at 72°C. Successful amplification was visualized at 35 cycles on a 2% agarose gel.

MiSeq indexing adaptors were added according to the Illumina 16S Metagenomic Sequencing Library Preparation protocol and an AxyPrep Mag™ PCR Clean-up Kit (Axygen Biosciences, Corning). 16S rRNA gene amplicon libraries were sequenced along with a 30% PhiX control at the University of North Carolina Charlotte sequencing facility on the Illumina MiSeq platform using 2x300 bp paired-end reads. As an extra quality control measure for potential contamination in the reagents, we sequenced different types of no template controls: two blanks that were processed through the DNA extraction column to identify potential contaminants from the kit, as well as sequencing one no template control that was only a water sample added to the original PCR reaction to identify potential contaminants in other parts of the procedure.

Sequence data processing

We merged the paired-end reads using PEAR v.0.9.10 with default settings, and used Trimmomatic v.0.36 for moderate quality trimming (threshold quality=20; minimum length=50) (Bolger et al., 2014;Zhang et al., 2014b). Chimeras were identified using UCHIME and removed using the filter_fasta.py command in QIIME v.1.9.1 (Caporaso et al., 2010;Edgar et al., 2011). All subsequent analysis was performed in QIIME v.1.9.1. Merged, trimmed, non-chimeric sequences were clustered into Operational Taxonomic Units (OTUs) at 97% sequence similarity using open-reference picking with the UCLUST algorithm. OTUs with fewer than 10 sequence reads were removed. One DNA sample from the PAG resulted in few sequence reads (maximum was 2,432 reads which was lower than the blanks) regardless of DNA extraction method and was therefore removed from later analyses, resulting in a total of 87 samples. We also filtered out any OTUs that were present in a higher abundance within the three blank samples than within our biological samples. This resulted in deletion of 228 OTUs comprising 11,916 sequence reads from all three blank samples combined and deletion of 28,101 sequence reads from all 87 samples combined. This was less than 0.8% of the total reads resulting from the biological samples. We also removed chloroplast sequences as they are assumed to be brought into the gut through ingestion. Taxonomy was assigned to the remaining OTUs through comparison with the Greengenes database 13_5. OTU counts were rarefied to 17,903 sequence reads (the lowest number of sequences over all samples included in the analysis) using the single_rarefaction.py command and the rarefied OTU table was used in downstream analysis.

16S rRNA microbial community analysis

Taxonomically annotated sequences were used to create bacterial community composition stacked plots at the family level to compare between methods and sites. Venn diagrams were generated using Venny 2.1.0 (Oliveros, 2007) to compare the number of shared and unique OTUs between methods and sites. PCoA plots were generated using Euclidean distance and Shannon and Simpson diversity indices were calculated and processed in QIIME. A Kruskal-Wallis test was used to identify OTUs that were represented significantly differently between methods. A nonparametric, two sample t-test was run using QIIME to test significance between alpha diversity indices for all pairwise comparisons between methods using the default number of Monte Carlo permutations (999) and the Bonferroni correction method. Nonmetric multidimensional scaling (MDS) was used to visualize differences between methods in Dhabiya, PAG.

Variations between methodologies from both collection sites were examined using an analysis of similarity (ANOSIM), a robust method which does not require normally distributed data or balanced replicates between groupings (Voss et al., 2007). Nonmetric MDS ordinations and ANOSIM were conducted using Primer 7 (Clarke, 1993).

Literature search

We searched the Web of Science (access dates 20/02/2018 - 24/02/2018) for publications reporting empirical research on the associated bacterial community for marine invertebrates. Our intention for these searches was not to be exhaustive but to identify and compare methodological approaches primarily related to DNA extraction from tissues. Literature searches were completed using two common "topics" search terms, 'marine' and 'microbio*', and a third specific search term for four groups of animals, 'coral', 'sponge', 'echino*', and 'mollus*'. We replaced the search term 'microbio*' with 'bacteria', which resulted in considerably more articles (e.g., 4.6X more for 'coral'), but our preliminary comparisons suggested many were largely unrelated to our area of interest and thus we determined using 'microbio*' was a more effective search term. Results for these four searches first included publications over all of the years in the database (1900 - 2017) to compare publication trends (Supplemental Figure 1.1). To restrict our search for a more in-depth literature comparison, we then reviewed all publications in these results for the years 2015-2017. We removed publications that were not relevant, typically because they were not specific to the taxonomic group (e.g., water samples), did not study microbes from live organisms (e.g., paleontology or non-living structures like shells), or did not use next generation or clone-library sequence-based methodology (e.g., culture-based studies). For the pruned list of publications, we categorized the methodology of each approach based on DNA extraction method, PCR strategy, and sequencing technology. A number of publications summarized their methods when using a commercial kit using phrases like "following manufacturer's protocol" or similar. For these, we categorized the methods using the default steps for DNA extraction without additional steps recommended elsewhere in the manual.

Results

Bacterial community composition between methods and sites

Sequencing of the V3/V4 region of the 16S rRNA resulted in a total of 6,554,465 reads from 93 samples: 90 *E. mathaei* samples (15 unique samples from Dhabiya (PAG), 15 unique samples from Al Aqah (GO); 30 total samples extracted using three different methods) and three blanks. DNA extraction Method 1 used the DNeasy Blood and Tissue Kit (QIAGEN), Method 2 added a bead beating step to this kit protocol, and Method 3 added the bead beating step and lysozyme to the kit protocol. After quality filtering, our pipeline retained 87 total samples with 3,469,193 sequence reads. To evaluate differences in the bacterial community composition, sequences were categorized taxonomically to the family level, where possible. We found family level taxonomic assignment to be the most informative, as genus and species level classification was often unsuccessful and resulted in many unknown strains. Samples from both geographic sites and all three methods were composed of similar principal bacterial families, although varying in levels of abundance (Figure 1.2). For example, almost all 87 samples contained Bacteriodales (range: ~2-38%, average number of reads: 3,203, standard deviation: 1442.4), Flavobacteriales (~5-24%, 2,269, 850.8), Desulfobulbaceae (~1-11%, 1,076, 449.3), and Fusobacteriaceae (~1-26%, 2,078, 1620.8), with a similar proportion of reads classifying as unassigned (~4-19%, 2,191, 609.7). Individuals from the PAG had a total of 304 OTUs while those from the GO had 271, with the PAG containing 50 unique OTUs, and the GO containing 17 (Figure 1.3). The location where samples were collected explained 63.7% of the variation in the samples (PC1 vs PC2) (Figure 1.4). For this reason, we separated our data based on location in order to describe the differences driven solely by methodology. Method 3 contains 11 unique OTUs while both Method 1 and 2 contain only 4. Further, Method 2 and 3 share more OTUs (n=23) than Method 1 and 2 (n=7) or Method 1 and 3 (n=8).

Inside the PAG, there were 13 bacterial families that differed significantly in abundance between the methodologies used (Tables 1.1 and 1.2). For every bacterial family whose abundance differed significantly between methods, Methods 2 and 3 contained a higher number of mean sequence reads than Method 1 for all but one bacteria (Acidimicrobiales; C111) (Table 1.1 and 1.2). Notably, five out of 13 of these taxa were gram-positive bacteria, groups known to have thick cell walls: Gaiellales, Turicibacteraceae, Staphylococcaceae, Pseudonocardiaceae, and Acidimicrobiales. Moreover, two of these 13 taxa were cyanobacteria; Xenococcaceae, and Pseudanabaenaceae, with differential representation between methods. The most significant difference (Kruskal-Wallis, $P < 0.001$) between methods was the Pseudomonadaceae. Lastly, the remaining five bacterial taxa consisted of Hyphomicrobiaceae, Amoebophilaceae, Rhodospirillales, Ellin6529, and

Methylobacteriaceae. Inside the GO, there were four significant differences between methods; two cyanobacteria (Cyanobacteriaceae and Xenococcaceae), one gram-positive bacteria (C111) and Pseudomonadaceae.

Shannon diversity was higher inside the PAG than the GO, independent of methodology (Figure 1.5). Shannon diversity for both sites was significantly higher in Method 3 than Method 1 (nonparametric t-test, *P*=0.018). Simpson diversity showed a similar trend to Shannon diversity when comparing methodologies, although this difference was not significant (nonparametric t-test, *P*=0.123) (Supplemental Figure 1.2). Multivariate statistics revealed an overall significant community difference between Methods 1 and 3 in both sites combined (ANOSIM, *P*=0.011, R=0.084). Although the differences between Methods 1 and 2 (ANOSIM*, P*=0.421, R=0.001) and Method 2 and 3 (ANOSIM, *P*=0.122, R=0.031) were not statistically significant, the non-metric MDS plot showed near uniform directionality between methods in the PAG (Figure 1.6). This was indicative of consistent changes in the abundance of certain taxa (see Tables 1.1 and 1.2) between methods that were driving the vectors connecting individuals in a consistent direction relative to the other two methods. Further, the effect that method had on community composition was heavily determined by individual (as seen in Figures 1.5 and 1.6), with some individuals displaying a greater variation in vector length between methods. This suggests that these individuals contain a higher abundance of difficult to lyse bacteria.

Literature analysis

Our comparison in methodologies for characterizing *E. mathaei*'s microbiome indicated that bead beating and lysozyme significantly influenced the community composition of microbiota. Therefore, we surveyed the literature to assess the commonality of these approaches in studies of marine invertebrate microbiomes and to better understand these results relative to the broader research community. Our literature search for manuscripts using our search terms (see Materials and Methods) at Web of Science resulted in 129 articles for corals, 167 for sponges, 14 for echinoderms, and 50 for molluscs, beginning in the early to mid-1990s (Supplemental Figure 1.1).

In general, publications increased over time and were generally highest in most recent years. The review of articles from 2015-2017 returned from our search resulted in substantially fewer that were relevant for a comparison of methodologies for sequence-based microbiome comparisons in these groups of marine invertebrates (15 of 67 for coral, 36 of 91 for sponge, 2 of 15 for mollusc*, 2 of 5 for echino*). From these 55 retained publications, comparisons of DNA extraction approach (53 total, two studies lacked sufficient details) resulted in 20 different commercial kits used in 44 studies and 9 studies used non-kit extractions [e.g., CTAB procedure as in (Fiore et al., 2013), and bead-beating method (Taylor et al., 2004)]. We then categorized the methods using commercial kits depending on if there was a column-based purification step, bead vortexing or beating step, a lysozyme digest, or a combination. Of the 53 studies, 44 used a column-based step, 29 used bead vortexing or beating, and only one specifically listed a lysozyme step. Next, we assessed the sequencing techniques and genetic region for comparing microbial taxa. A majority of studies utilized high throughput sequencing (32 Illumina MiSeq or HiSeq, 16 Roche pyrosequencing, 1 Ion Torrent, 1 PacBio), with the remaining five using either

Sanger-sequenced clone libraries or gel-based analysis (e.g., DGGE). While metagenomic methods are increasing in frequency, almost all studies (49 of 55) we identified in our search used 16S rRNA for comparisons of microbial communities. The specific variable region(s) of 16S rRNA that were used varied across studies. Lastly, only two studies (Marcelino and Verbruggen, 2016;Nakagawa et al., 2017) explicitly stated in their methods that negative control samples were used in their analysis.

Discussion

In the past, microbial community studies were restricted to culture-based approaches that isolated microorganisms as pure cultures and performed downstream analysis on these cultures. This is a limiting approach as most bacteria in natural environments cannot be cultured (Su et al., 2015). One of the major advances in modern sequencing technologies is the ability to use phylogenetic markers, such as the conserved 16S rRNA gene, to identify bacterial members of the microbiome that are nonculturable or rare (Lagkouvardos et al., 2016). However, using molecular methods that exclude certain bacteria, for example those with robust cell walls not lysed using traditional approaches, will underestimate microbial diversity and limit our interpretation of bacterial communities. It is vital that the microbial research community develops a set of best practices that most accurately represent the bacterial communities present (Miller et al., 1999;Yuan et al., 2012;Albertsen et al., 2015;Larsen et al., 2015;Burbach et al., 2016). Our goal in this study was to examine how DNA extraction approaches may influence the characterization of microbiome diversity through an empirical study of a sea urchin collected from different geographic areas and how these methods vary through literature-based comparisons.

Here we show that DNA extraction method plays an important role in estimates of Shannon diversity for bacterial communities associated with gut samples from the urchin *E. mathaei* in a site- and individual-specific manner. Within the GO, method did not result in significant differences in Shannon diversity metrics', although there were still specific families of bacteria that are found in significantly different proportions between methods. However, in the PAG, we observed the maximum Shannon diversity in Method 3, and this was significantly higher than diversity estimates from Method 1. These site-specific effects could be a result of methodology playing a more important role in identifying a more diverse community, as the PAG has higher Shannon indices than the GO, regardless of method. Further, we found the role methodology has on bacterial communities is specific to the individual. In other words, there are particular individuals that demonstrate larger community shifts when using Method 3. This variation could be a result of these individuals containing more difficult to lyse bacteria than others. Although Simpson diversity indices are generally higher using Method 3 for PAG samples, the difference between methods was not significant. This is not surprising, as Simpson diversity is heavily weighted by dominant OTUs (Hill et al., 2003), and it is the rare OTUs that are driving the differences between methods (Supplemental Figure 1.2). These results indicate that it may be challenging to compare microbial communities between studies that used different methods and highlights the importance of using the most robust method possible, especially when studying an uncharacterized system. As shown in our literature analysis, where DNA extraction techniques vary substantially, these biases could be influential in community characterization.

Of the bacterial taxa that were better captured with Method 3, there were five gram-positive bacteria in the PAG and one in the GO. These results were not surprising as gram-positive bacteria have a thick peptidoglycan layer in their cell wall that makes them more difficult to lyse than most gram-negative bacteria. Lysozymes are effective at hydrolyzing the glycosidic linkages in the peptidoglycan layer of the cell walls (Mehta et al., 2015). Gram-positive bacteria are common in many different environments and play crucial ecological roles. Actinobacteria, for example, are secondary metabolite producers and play an important role in organic matter turnover and the carbon cycle (Zhang et al., 2014a). There was also an increased abundance of two cyanobacteria and Pseudomonadaceae after lysozyme treatment in both sampling sites as well. Cyanobacteria have a resilient cell wall that is thicker and has a more highly crosslinked peptidoglycan layer than many gram-negative bacteria (Mehta et al., 2015). Although Pseudomonadaceae is gram-negative, *Pseudomonas aeruginosa* (the most well-described species within the Pseudomonadaceae family) is difficult to lyse without EDTA and lysozyme (Eagon and Carson, 1965). Further, we saw a general trend of increasing mean sequence reads from Method 1 to 3, with Method 3 consistently outperforming the other two methods (Table 1.1 and 1.2). It is also important to note that there were no OTUs that were represented significantly higher in Method 1 or 2 than Method 3. These results emphasize that although bead beating does help lyse hardy bacteria, it does not do so as effectively as when lysozyme and bead beating are combined. Microbial studies that are not using bead beating and lysozyme may be missing key microbial players.

The significantly different bacterial families that we present here typically comprised <0.2% of an individual's microbial community extracted using Method 3, although Xenococcaceae did contribute approximately 2.8% of the individual's community. While these families are not the most abundant taxa in our dataset, they may play important ecological or physiological roles. For example, several studies have found that rare species are disproportionately active compared to their abundant counterparts and therefore, should not be overlooked (Dimitriu et al., 2010;Debroas et al., 2015;Jousset et al., 2017). In one study, similar sized communities with a more diverse rare microbiome were shown to have higher respiration than communities with a smaller overall rare microbiome (Dimitriu et al., 2010). It is also important to note that although some of these taxa may be present in relatively low abundances in this dataset, other communities may be composed of larger abundances of difficult to lyse bacteria and therefore, method may play a more influential role. There is no way to assess the relative abundances of difficult to lyse bacteria *a priori*. Therefore, choosing the most rigorous methodology is recommended as a default choice to ensure accurate bacterial community representation, as well as allow for cross comparisons between studies.

Method-based biases can limit comparisons of microbial communities between studies that used different methods, thereby decreasing the potential to synthesize across groups of organisms. Many animal species or phylogenetic groups may have specific protocols for DNA extraction (e.g., inclusion of CTAB for molluscs and corals) to increase yield of nucleic acids for downstream analysis. Similarly, individual bacterial species have different optimal lysis conditions. Our survey of recently published studies on the microbial communities of marine invertebrates showed that while most studies used a commercial kit for DNA extraction, the methodological steps varied. Specifically, the inclusion of bead vortexing or beating was part of about 50% of studies and a lysozyme step was found in less than 5% of studies, which could result in reduced representation of particular groups of bacteria. The impact of these variations in the kit or the protocol remains sparsely reported in the literature for animal-associated bacteria of marine invertebrates (Weber et al., 2017). Moreover, many of these kits contain multiple protocols for DNA extraction depending on the type of tissue or organismal group. Future reporting by authors on which specific protocol was used would be beneficial for later utility in meta-analyses of the field.

In addition to rigorous methodology, microbial studies also need to account for bacterial contamination. Contamination can be introduced throughout the experimental process, from collection to sample processing. A significant amount of research has investigated the potential for contamination from DNA extraction kits and our data support the need to account for this contamination (Salter et al., 2014;Weiss et al., 2014;Glassing et al., 2016). With three blank samples (two of which were run through a DNA extraction column), we had roughly 14,000 total sequence reads, highlighting the prominence of contaminants. It is possible that these contaminants were also introduced at different steps of the extraction methodology (e.g., pipette tips or sample tubes). However, the blanks that were run through the extraction column had >60% more sequence reads and more OTUs (138 compared to 400) than the blank that was only used during the PCR step. Although removal of all the OTUs that are present in blank and biological samples may remove some biological signal (if contamination is a result of the biological sample contaminating the blanks), it is important to consider the ramifications of not accounting for the presence of contaminants. Indeed, in our review of the literature, we only identified two studies that definitely used a blank in their study. It is also important to note that each extraction kit may contain a different set of contaminants (the "kitome"), making it crucial to run and account for blanks in parallel with the biological samples (Salter et al., 2014;Weiss et al., 2014;Glassing et al., 2016). Our review of publications over the last three years showed that 20 different commercial kits were used to extract DNA. This potential for bias from these kits and variation between lots of the same kit will likely add a number of rare OTUs to the characterization of the microbial community. These bacteria would not only increase the derived diversity of the microbiome but could also result in shared, rare microbes between studies of species that happen to use the same kits. Thus, although kit-based DNA extractions may increase repeatability of methods, studies that do not account for the potential of contaminating DNA from particular bacteria will include bacteria not specific to the biological system under study.

In summary, we show that DNA extraction methods that incorporate a bead beating and lysozyme step more accurately characterize this bacterial community. While we observe significant differences in community composition between DNA extraction methods, this is only one step in a multi-step procedure. Sample collection, primer pair selection, PCR conditions, and sequencing approach each have their own propensity for introducing bias that also merit further study to determine to what extent each step effects the inferred microbial diversity. Further, we describe the current state of marine microbial methodology in invertebrates and highlight the importance of developing a standardized DNA extraction protocol for marine microbial community analysis. Finally, it is important that studies select rigorous extraction methods to most accurately sample the community and increase consistency between studies.

Acknowledgements

We thank Karen Lopez (UNC Charlotte) for her technical assistance with library preparation and sequencing for the sea urchin samples. We also thank New York University Abu Dhabi for supportive funding and the Environmental Agency – Abu Dhabi for permits.

References

Albertsen, M., Karst, S. M., Ziegler, A. S., Kirkegaard, R. H., & Nielsen, P. H. (2015). Back to Basics – The influence of DNA extraction and primer choice on phylogenetic analysis of activated sludge communities. *PLOS One, 10*(7), e0132783. doi:10.1371/journal.pone.0132783

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

Bordenstein, S. R., & Theis, K. R. (2015). Host biology in light of the microbiome: ten principles of holobionts and hologenomes. *PLOS Biology, 13*(8), e1002226. Retrieved from https://doi.org/10.1371/journal.pbio.1002226

Bourne, D. G., Morrow, K. M., & Webster, N. S. (2016). Insights into the coral microbiome: underpinning the health and resilience of reef ecosystems. *Annual Review of Microbiology, 70*(1), 317-340. doi:10.1146/annurev-micro-102215-095440

Burbach, K., Seifert, J., Pieper, D. H., & Camarinha□Silva, A. (2016). Evaluation of DNA extraction kits and phylogenetic diversity of the porcine gastrointestinal tract based on Illumina sequencing of two hypervariable regions. *MicrobiologyOpen, 5*(1), 70-82. doi:10.1002/mbo3.312

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., . . . Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods, 7*, 335. doi:10.1038/nmeth.f.303

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). doi:10.3389/fmicb.2017.00802

Carrier, T. J., & Reitzel, A. M. (2018). Convergent shifts in host-associated microbial communities across environmentally elicited phenotypes. *Nature Communications, 9*(1), 952. doi:10.1038/s41467-018-03383-w

Clarke, K. R. (1993). Non-parametric multivariate analyses of changes in community structure. *Australian Journal of Ecology, 18*(1), 117-143. doi:10.1111/j.1442-9993.1993.tb00438.x

de Bruin, O. M., & Birnboim, H. C. (2016). A method for assessing efficiency of bacterial cell disruption and DNA release. *BMC Microbiology, 16*(1), 197. doi:10.1186/s12866-016-0815-3

Debroas, D., Hugoni, M., & Domaizon, I. (2015). Evidence for an active rare biosphere within freshwater protists community. *Molecular Ecology, 24*(6), 1236-1247. doi:10.1111/mec.13116

Dimitriu, P. A., Lee, D., & Grayston, S. J. (2010). An evaluation of the functional significance of peat microorganisms using a reciprocal transplant approach. *Soil Biology and Biochemistry, 42*(1), 65-71. doi:10.1016/j.soilbio.2009.10.001

Eagon, R. G., & Carson, K. J. (1965). Lysis of cell walls and intact cells of *Pseudomonas aeruginosa* by ethylenediamine tetraacetic acid and by lysozyme. *Canadian Journal of Microbiology, 11*(2), 193-201. doi:10.1139/m65-025

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., & Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics, 27*(16), 2194-2200. doi:10.1093/bioinformatics/btr381

Fiore, C. L., Jarett, J. K., & Lesser, M. P. (2013). Symbiotic prokaryotic communities from different populations of the giant barrel sponge, *Xestospongia muta*. *MicrobiologyOpen, 2*(6), 938-952. doi:10.1002/mbo3.135

Galac, M. R., Bosch, I., & Janies, D. A. (2016). Bacterial communities of oceanic sea star (Asteroidea: Echinodermata) larvae. *Marine Biology, 163*(7), 162. doi:10.1007/s00227-016-2938-3

Gao, F., Li, F., Tan, J., Yan, J., & Sun, H. (2014). Bacterial community composition in the gut content and ambient sediment of sea cucumber *Apostichopus japonicus* revealed by 16S rRNA gene pyrosequencing. *PLOS One, 9*(6), e100092. doi:10.1371/journal.pone.0100092

Glassing, A., Dowd, S. E., Galandiuk, S., Davis, B., & Chiodini, R. J. (2016). Inherent bacterial DNA contamination of extraction and sequencing reagents may affect interpretation of microbiota in low bacterial biomass samples. *Gut Pathogens, 8*(1), 24. doi:10.1186/s13099-016-0103-7

Hakim, J. A., Koo, H., Dennis, L. N., Kumar, R., Ptacek, T., Morrow, C. D., . . . Watts, S. A. (2015). An abundance of Epsilonproteobacteria revealed in the gut microbiome of the laboratory cultured sea urchin, *Lytechinus variegatus*. *Frontiers in Microbiology, 6*(1047). doi:10.3389/fmicb.2015.01047

Hakim, J. A., Koo, H., Kumar, R., Lefkowitz, E. J., Morrow, C. D., Powell, M. L., . . . Bej, A. K. (2016). The gut microbiome of the sea urchin, *Lytechinus variegatus*, from its natural habitat demonstrates selective attributes of microbial taxa and predictive metabolic profiles. *FEMS Microbiology Ecology, 92*(9), fiw146-fiw146. doi:10.1093/femsec/

Holland, N. D., & Nealson, K. H. (1978). The fine structure of the echinoderm cuticle and the subcuticular bacteria of echinoderms. *Acta Zoologica, 59*(3-4), 169-185. doi:10.1111/j.1463-6395.1978.tb01032.x

Jousset, A., Bienhold, C., Chatzinotas, A., Gallien, L., Gobet, A., Kurm, V., . . . Hol, W. H. G. (2017). Where less may be more: how the rare biosphere pulls ecosystems strings. *The ISME Journal, 11*, 853. doi:10.1038/ismej.2016.174

Kelly, M. S., Barker, M. F., McKenzie, J. D., & Powell, J. (1995). The incidence and morphology of subcuticular bacteria in the echinoderm fauna of New Zealand. *The Biological Bulletin, 189*(2), 91-105. doi:10.2307/1542459

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., & Glöckner, F. O. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Research, 41*(1). doi:10.1093/nar/gks808

Kohl, K. D., & Carey, H. V. (2016). A place for host–microbe symbiosis in the comparative physiologist's toolbox. *The Journal of Experimental Biology, 219*(22), 3496-3504. doi:10.1242/jeb.136325

Lagkouvardos, I., Joseph, D., Kapfhammer, M., Giritli, S., Horn, M., Haller, D., & Clavel, T. (2016). IMNGS: A comprehensive open resource of processed 16S rRNA microbial profiles for ecology and diversity studies. *Scientific Reports, 6*, 33721. doi:10.1038/srep33721

Larsen, A. M., Mohammed, H. H., & Arias, C. R. (2015). Comparison of DNA extraction protocols for the analysis of gut microbiota in fishes. *FEMS Microbiology Letters, 362*(5), fnu031-fnu031. doi:10.1093/femsle/fnu031

Lesser, M. P., & Walker, C. W. (1992). Comparative study of the uptake of dissolved amino acid in sympatric brittle stars with and without endosymbiotic bacteria. *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry, 101*(1-2), 217-223. doi:https://doi.org/10.1016/0305-0491(92)90182-Q

Marcelino, V. R., & Verbruggen, H. (2016). Multi-marker metabarcoding of coral skeletons reveals a rich microbiome and diverse evolutionary origins of endolithic algae. *Scientific Reports, 6*, 31508. doi:10.1038/srep31508

McFall-Ngai, M., Hadfield, M. G., Bosch, T. C. G., Carey, H. V., Domazet-Lošo, T., Douglas, A. E., . . . Wernegreen, J. J. (2013). Animals in a bacterial world, a new imperative for the life sciences. *Proceedings of the National Academy of Sciences, 110*(9), 3229-3236. doi:10.1073/pnas.1218525110

McOrist, A. L., Jackson, M., & Bird, A. R. (2002). A comparison of five methods for extraction of bacterial DNA from human faecal samples. *Journal of Microbiological Methods, 50*(2), 131-139. doi:https://doi.org/10.1016/S0167-7012(02)00018-0

Mehta, K. K., Evitt, N. H., & Swartz, J. R. (2015). Chemical lysis of cyanobacteria. *Journal of Biological Engineering, 9*, 10. doi:10.1186/s13036-015-0007-y

Meziti, A., Kormas, K. A., Pancucci-Papadopoulou, M.-A., & Thessalou-Legaki, M. (2007). Bacterial phylotypes associated with the digestive tract of the sea urchin *Paracentrotus lividus* and the ascidian *Microcosmus* sp. *Russian Journal of Marine Biology, 33*(2), 84-91. doi:10.1134/s1063074007020022

Miller, D. N., Bryant, J. E., Madsen, E. L., & Ghiorse, W. C. (1999). Evaluation and optimization of DNA extraction and purification procedures for soil and sediment samples. *Applied and Environmental Microbiology, 65*(11), 4715-4724.

Miller, P. M., Lamy, T., Page, H. M., & Miller, R. J. (2021). Sea urchin microbiomes vary with habitat and resource availability. *Limnology and Oceanography Letters, 6*(3), 119-126.

Moitinho-Silva, L., Nielsen, S., Amir, A., Gonzalez, A., Ackermann, G. L., Cerrano, C., . . . Thomas, T. (2017). The sponge microbiome project. *GigaScience, 6*(10), 1-7. doi:10.1093/gigascience/gix077

Nakagawa, S., Saito, H., Tame, A., Hirai, M., Yamaguchi, H., Sunata, T., . . . 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*(1), 8764. doi:10.1038/s41598-017-09355-2

Nelson, L., Blair, B., Murdock, C., Meade, M., Watts, S., & Lawrence, A. L. (2010). Molecular analysis of gut microflora in captive-raised sea urchins (*Lytechinus variegatus*). *Journal of the World Aquaculture Society, 41*(5), 807-815. doi:10.1111/j.1749-7345.2010.00423.x

Oliveros, J. C. (2007). VENNY: An interactive tool for comparing lists with Venn Diagrams.

Pollock, J., Glendinning, L., Wisedchanwet, T., & Watson, M. (2018). The madness of microbiome: Attempting to find consensus “best practice” for 16S microbiome studies. *Applied and Environmental Microbiology, 84*, e02627-02617. doi:10.1128/aem.02627-17

Reveillaud, J., Maignien, L., Eren, A. M., Huber, J. A., Apprill, A., Sogin, M. L., & Vanreusel, A. (2014). Host-specificity among abundant and rare taxa in the sponge microbiome. *The ISME Journal, 8*, 1198. doi:10.1038/ismej.2013.227

Salter, S. J., Cox, M. J., Turek, E. M., Calus, S. T., Cookson, W. O., Moffatt, M. F., . . . Walker, A. W. (2014). Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biology, 12*(1), 87. doi:10.1186/s12915-014-0087-z

Small, C. M., Milligan-Myhre, K., Bassham, S., Guillemin, K., & Cresko, W. A. (2017). Host genotype and microbiota contribute asymmetrically to transcriptional variation in the threespine stickleback gut. *Genome Biology and Evolution, 9*(3), 504-520. doi:10.1093/gbe/evx014

Smith, E. G., Hume, B. C. C., Delaney, P., Wiedenmann, J., & Burt, J. A. (2017). Genetic structure of coral-Symbiodinium symbioses on the world's warmest reefs. *PloS one, 12 (6)*, e0180169. doi:10.1371/journal.pone.0180169

Su, X., Sun, F., Wang, Y., Hashmi, M. Z., Guo, L., Ding, L., & Shen, C. (2015). Identification, characterization and molecular analysis of the viable but nonculturable *Rhodococcus biphenylivorans*. *Scientific Reports, 5*, 18590. doi:10.1038/srep18590

Taylor, M. W., Schupp, P. J., Dahllöf, I., Kjelleberg, S., & Steinberg, P. D. (2004). Host specificity in marine sponge-associated bacteria, and potential implications for marine microbial diversity. *Environmental Microbiology, 6*(2), 121-130. doi:10.1046/j.1462-2920.2003.00545.x

Theis, K. R., Dheilly, N. M., Klassen, J. L., Brucker, R. M., Baines, J. F., Bosch, T. C.

G., . . . Bordenstein, S. R. (2016). Getting the hologenome concept right: An eco-evolutionary framework for hosts and their microbiomes. *mSystems, 1*(2), e00028-00016.

Thomas, T., Moitinho-Silva, L., Lurgi, M., Björk, J. R., Easson, C., Astudillo-García, C., . . . Webster, N. S. (2016). Diversity, structure and convergent evolution of the global sponge microbiome. *Nature Communications, 7*, 11870. doi:10.1038/ncomms11870

Uthicke, S., Soars, N., Foo, S., & Byrne, M. (2013). Effects of elevated pCO2 and the effect of parent acclimation on development in the tropical Pacific sea urchin *Echinometra mathaei*. *Marine Biology, 160*(8), 1913-1926. doi:10.1007/s00227-012-2023-5

Vanysacker, L., Declerck, S. A. J., Hellemans, B., De Meester, L., Vankelecom, I., & Declerck, P. (2010). Bacterial community analysis of activated sludge: an evaluation of four commonly used DNA extraction methods. *Applied Microbiology and Biotechnology, 88*(1), 299-307. doi:10.1007/s00253-010-2770-5

Voss, J. D., Mills, D. K., Myers, J. L., Remily, E. R., & Richardson, L. L. (2007). Black band disease microbial community variation on corals in three regions of the wider Caribbean. *Microbial Ecology, 54*(4), 730-739. doi:10.1007/s00248-007-9234-1

Walker, C. W., & Lesser, M. P. (1989). Nutrition and development of brooded embryos in the brittlestar *Amphipholis squamata*: do endosymbiotic bacteria play a role? *Marine Biology, 103*(4), 519-530. doi:10.1007/bf00399584

Weber, L., DeForce, E., & Apprill, A. (2017). Optimization of DNA extraction for advancing coral microbiota investigations. *Microbiome, 5*(1), 18. doi:10.1186/s40168-017-0229-y

Webster, N. S., Negri, A. P., Botté, E. S., Laffy, P. W., Flores, F., Noonan, S., . . . Uthicke, S. (2016). Host-associated coral reef microbes respond to the cumulative pressures of ocean warming and ocean acidification. *Scientific Reports, 6*, 19324. doi:10.1038/srep19324

Webster, N. S., & Reusch, T. B. H. (2017). Microbial contributions to the persistence of coral reefs. *The ISME Journal, 11*, 2167. doi:10.1038/ismej.2017.66

Weiss, S., Amir, A., Hyde, E. R., Metcalf, J. L., Song, S. J., & Knight, R. (2014). Tracking down the sources of experimental contamination in microbiome studies. *Genome Biology, 15*(12), 564. doi:10.1186/s13059-014-0564-2

Yuan, S., Cohen, D. B., Ravel, J., Abdo, Z., & Forney, L. J. (2012). Evaluation of methods for the extraction and purification of DNA from the human microbiome. *PloS One, 7*(3). doi:10.1371/journal.pone.0033865

Zhang, H., Ding, W., He, X., Yu, H., Fan, J., & Liu, D. (2014). Influence of 20–Year Organic and inorganic fertilization on organic carbon accumulation and microbial community structure of aggregates in an intensively cultivated sandy loam soil. *PloS One, 9*(3). doi:10.1371/journal.pone.0092733

Zhang, J., Kobert, K., Flouri, T., & Stamatakis, A. (2014). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics, 30*(5), 614-620. doi:10.1093/bioinformatics/btt593

Ziegler, M., Seneca, F. O., Yum, L. K., Palumbi, S. R., & Voolstra, C. R. (2017).

Bacterial community dynamics are linked to patterns of coral heat tolerance. *Nature Communications, 8*, 14213. doi:10.1038/ncomms14213

Table 1.1: All bacterial families that are significantly different between the three methods in the Persian/Arabian Gulf.

Persian/Arabian Gulf: Dhabiya	***FDR P-value***	***Method 1: Mean Read Count***	***Method 2: Mean Read Count***	***Method 3: Mean Read Count***	***Gram***
Pseudomonadales; Pseudomonadaceae	0	0	0.2	28.3	Negative
Rhizobiales; Hyphomicrobiaceae	0	12.6	31.3	191.7	Negative
Gaiellales; unclassified	0	0.5	2.9	26.1	Positive
Chroococcales; Xenococcaceae	0.001	6.4	98.1	426.6	Negative
Pseudanabaenales; Pseudanabaenaceae	0.002	3.1	12.9	38.4	Negative
Cytophagales; Amoebophilaceae	0.01	0.4	2.9	8.4	Negative
Rhodospirillales; unclassified	0.012	4.6	11.4	37.8	Negative
Turicibacterales; Turicibacteraceae	0.012	0	0.1	2.9	Positive
Bacillales; Staphylococcaceae	0.014	0	1	6.6	Positive
Chloroflexi; Ellin6529; unclassified	0.017	1	3.1	22.1	Negative
Actinomycetales; Pseudonocardiaceae	0.024	0.7	0.9	4.3	Positive
Rhizobiales; Methylobacteriaceae	0.032	0	0.1	1.3	Negative
Acidimicrobiales; unknown	0.032	3.8	6	20.4	Positive

Table 1.2: All bacterial families that are significantly different between the three methods in the Gulf of Oman.

Gulf of Oman: Al Aqah	***FDR P-value***	***Method 1: Mean Read Count***	***Method 2: Mean Read Count***	***Method 3: Mean Read Count***	***Gram***
Pseudomonadales; Pseudomonadaceae	0	0	0	8.9	Negative
Chroococcales; Cyanobacteriaceae	0.015	5.7	15.9	28.7	Negative
Chroococcales; Xenococcaceae	0.016	0.9	5.4	7.9	Negative
Acidimicrobiales; C111	0.019	5.5	5.1	12.7	Positive

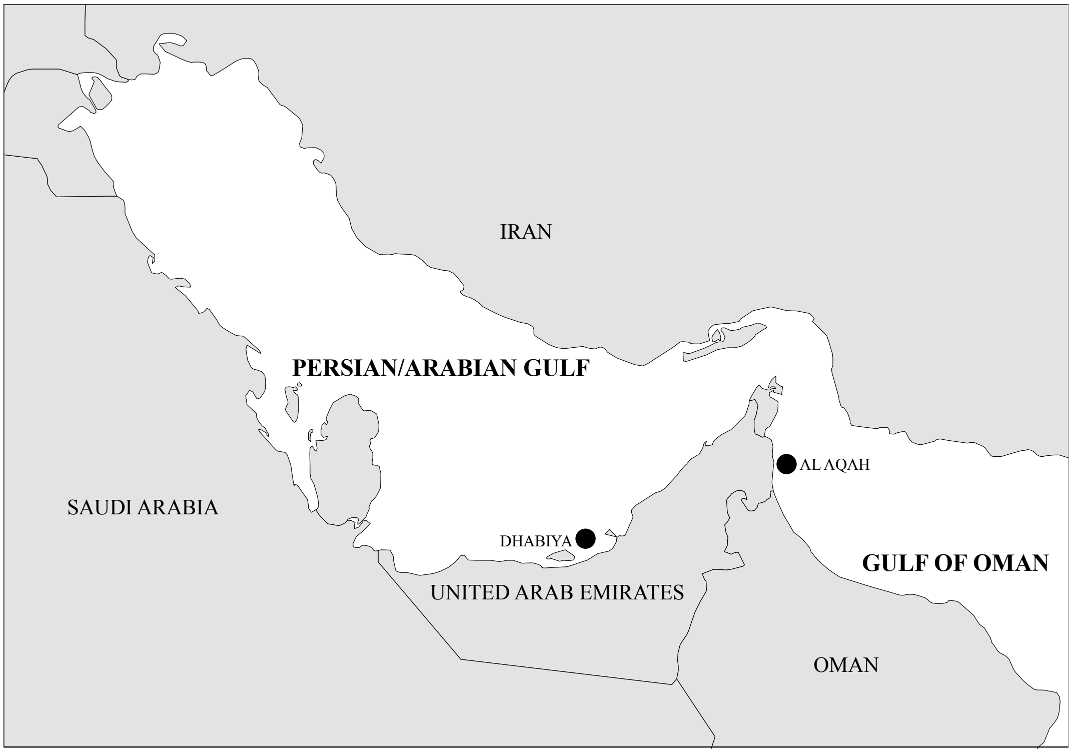


Figure 1.1: Map of Arabian Peninsula indicating the location of the two sampling sites.

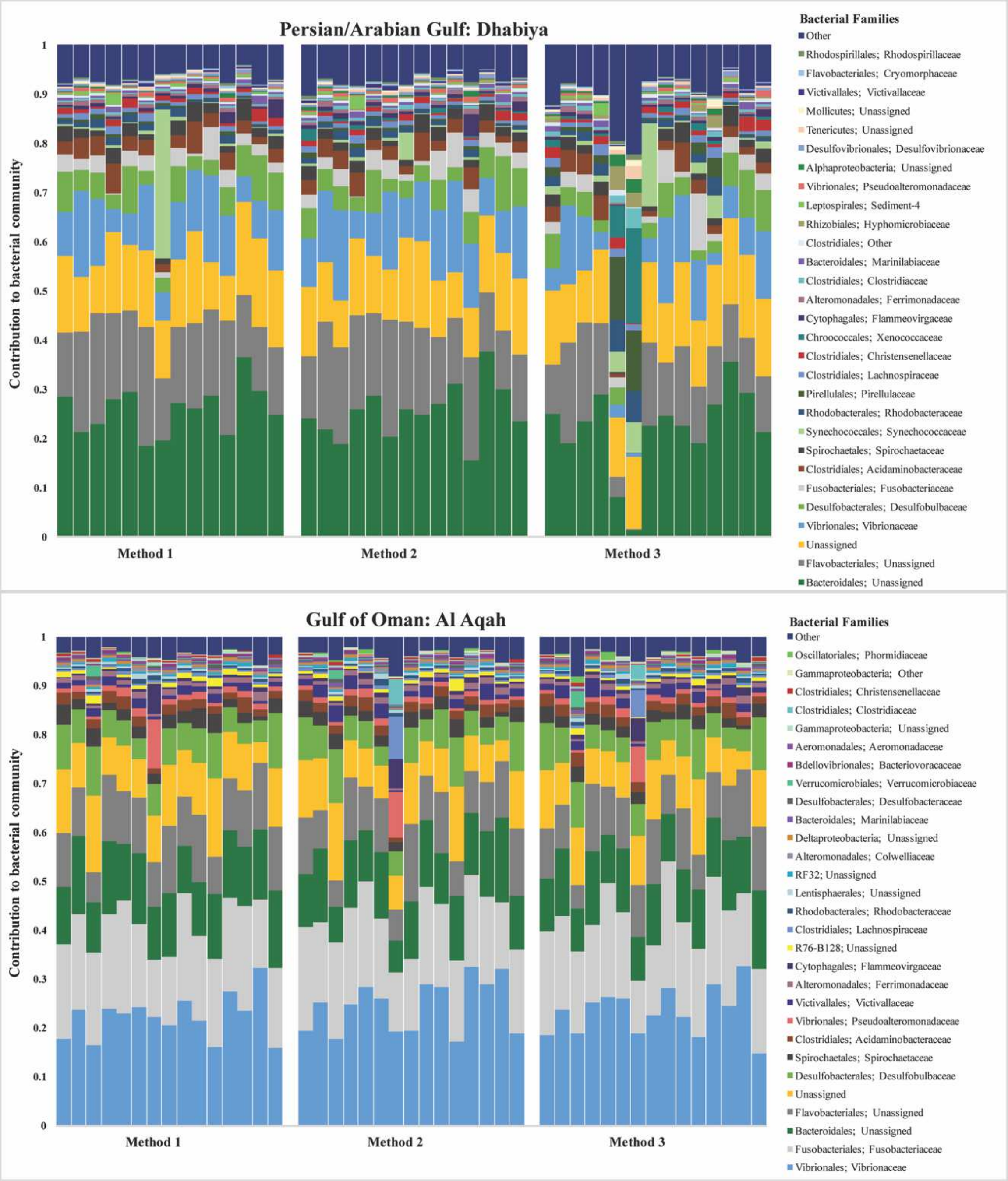


Figure 1.2: Bacterial community composition of gut contents from *Echinometra mathaei* from two populations; Persian/Arabian Gulf (on top), Gulf of Oman (on bottom). This taxonomy stacked column plot is characterized to the family level, where possible. Each color represents one of the 29 most abundant families and the remaining taxa are grouped into the category ‘other.’ Samples are grouped by method.

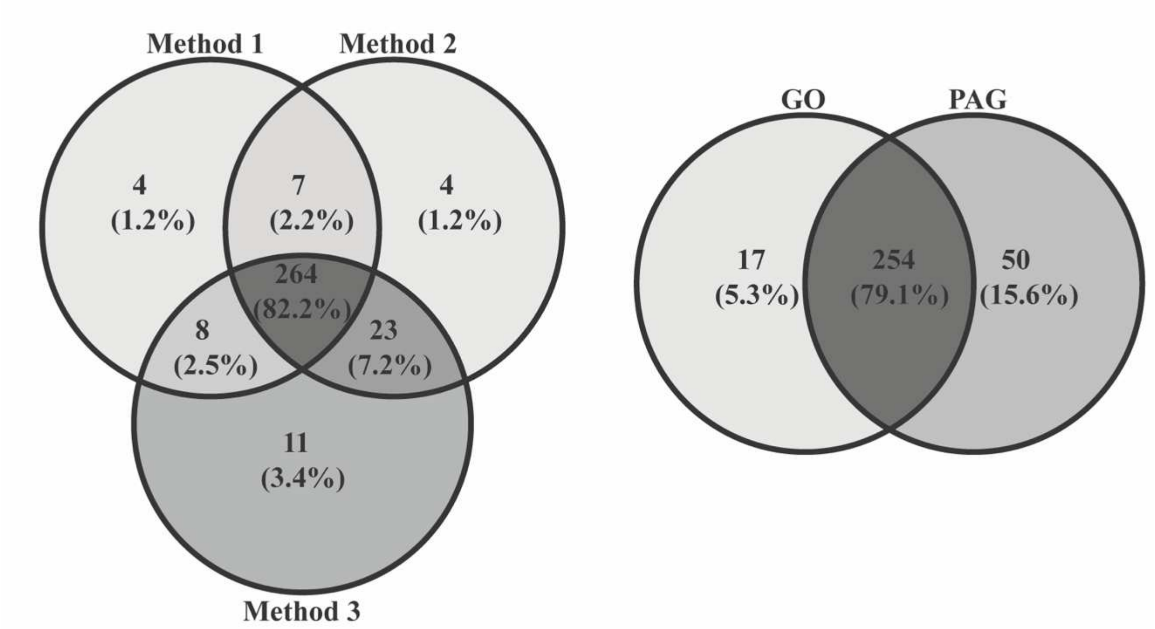


Figure 1.3: Venn diagrams comparing number of OTUs within and between the PAG and the GO, as well as within and between the three different methods.

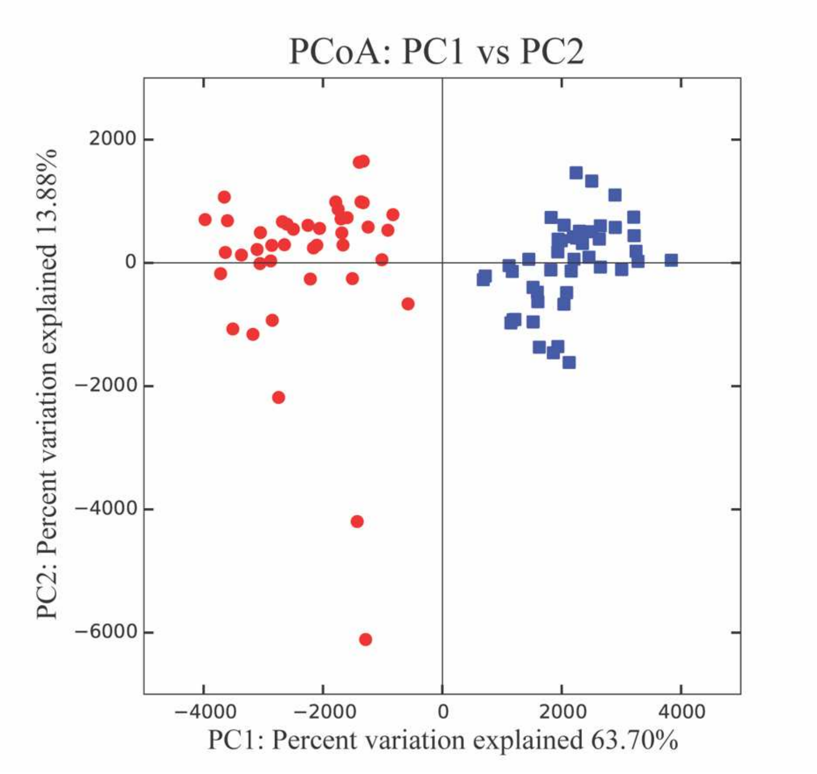


Figure 1.4: PCoA plot showing PC1 vs PC2 using Euclidean distance (Bray-Curtis distance yielded nearly identical results). Red circles indicate samples collected from the PAG, while blue squares indicate samples collected from the GO.

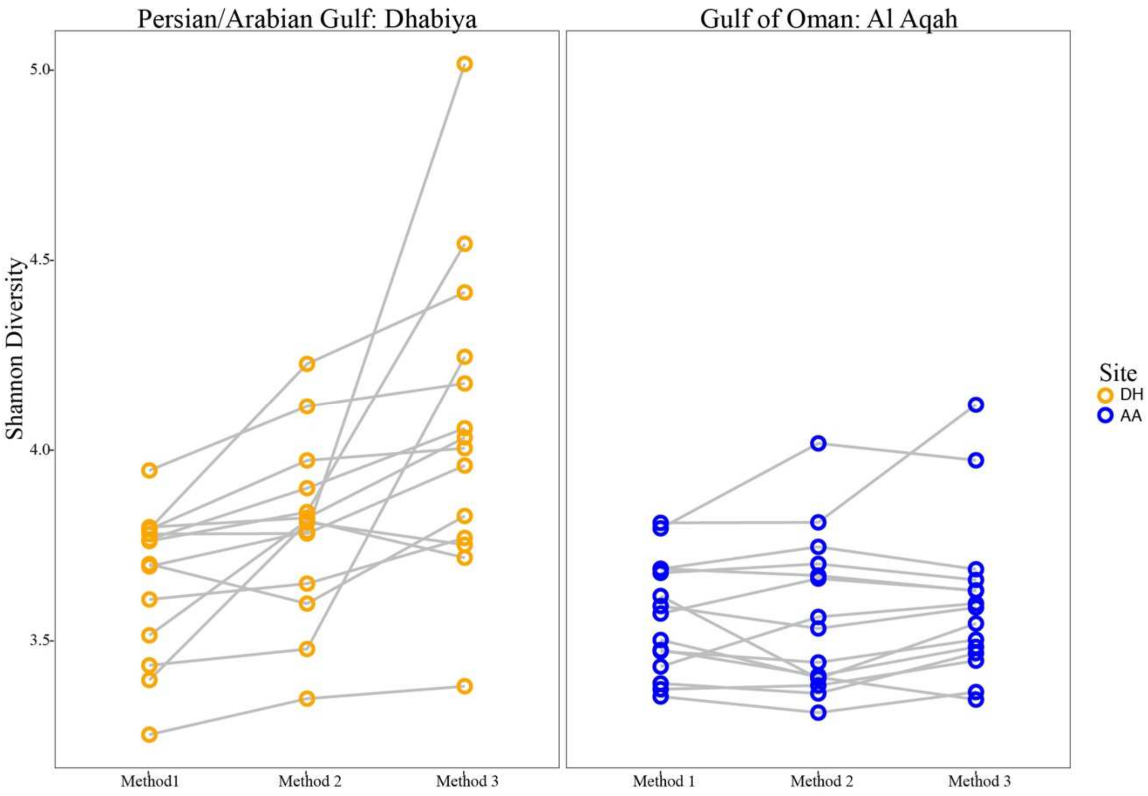


Figure 1.5: Shannon Diversity indices of Gulf of Oman samples (right) and Persian/Arabian Gulf samples (left). Each node corresponds to an individual sample and each line connects one individual's tissue subjected to the three DNA extraction methods.

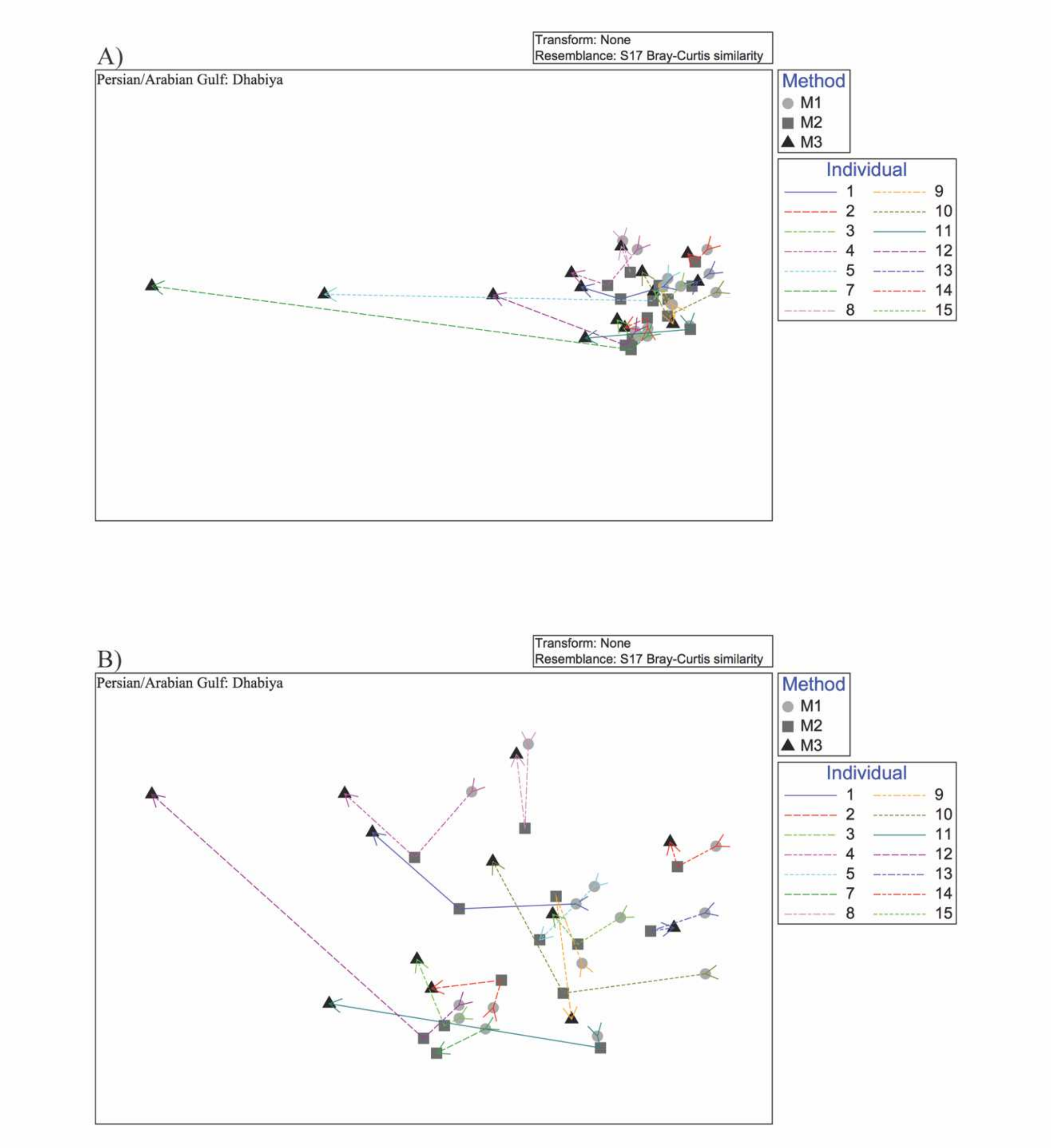


Figure 1.6: Non-metric multidimensional scaling plot based on Bray-Curtis similarities of the microbial method comparison data depicting Method 1 (circle), Method 2 (square), and Method 3 (triangle). Lines connect individual urchin samples. The upper plot (a) consists of all samples within the Persian/Arabian Gulf site, Dhabiya, while the lower plot (b) does not contain the two outliers (Sample ID: DH7, DH5) for a more in-depth view.

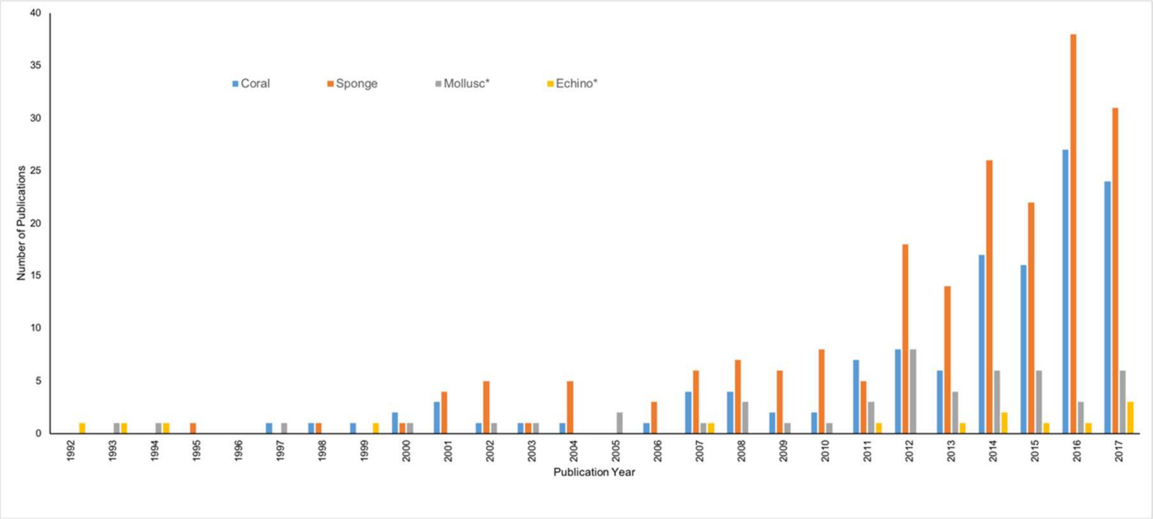


Figure S1.1: Bar chart showing growth in number of publications from 1992 to 2017 in the four invertebrates; coral (in blue), sponge (in orange), mollusc*(in gray), and echino*(in yellow).

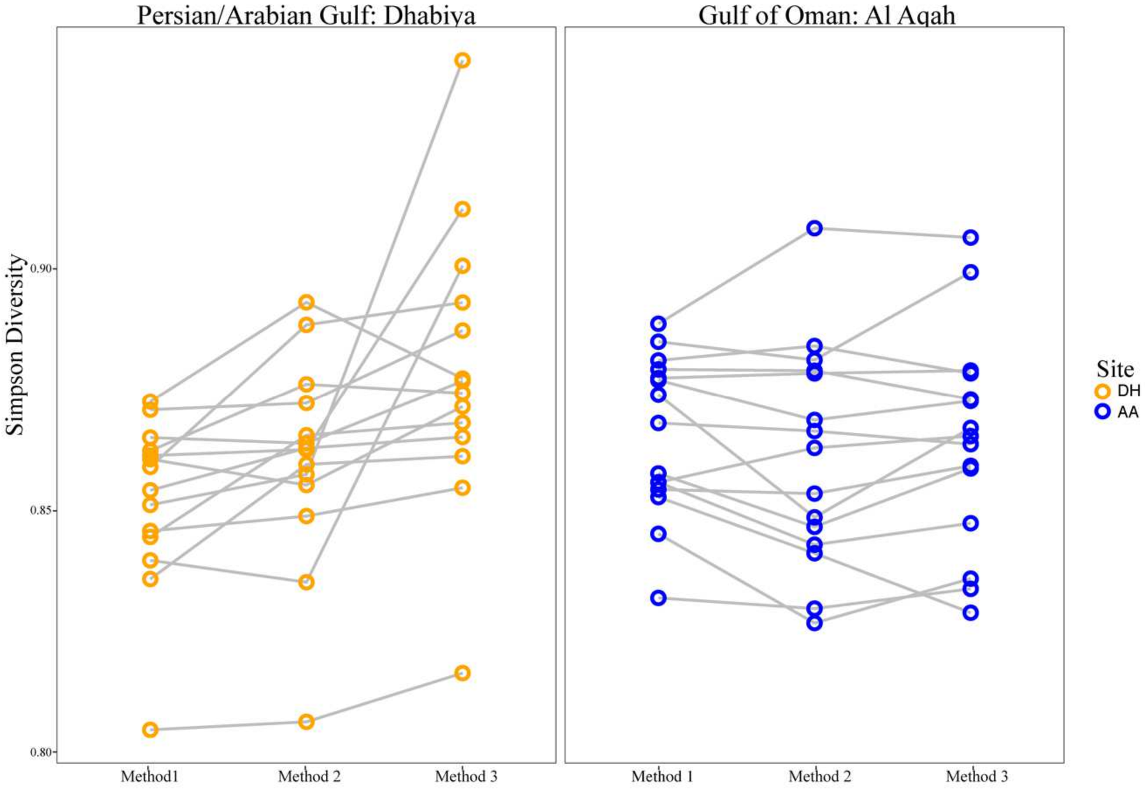


Figure S1.2: Simpson Diversity indices of Gulf of Oman samples (right) and Persian/Arabian Gulf samples (left). Each node corresponds to an individual sample and each line connects one individual's tissue subjected to the three DNA extraction methods.

CHAPTER 2

UNRAVELING THE PREDICTIVE ROLE OF TEMPERATURE IN THE GUT MICROBIOTA OF THE SEA URCHIN *ECHINOMETRA* SP. *EZ* ACROSS SPATIAL AND TEMPORAL GRADIENTS

Remi N. Ketchum, Edward G. Smith, Grace O. Vaughan, Dain McParland, Noura Al-Mansoori, John A. Burt, Adam M. Reitzel

Citation

Ketchum, Remi N., et al. "Unraveling the predictive role of temperature in the gut microbiota of the sea urchin *Echinometra* sp. *EZ* across spatial and temporal gradients." *Molecular Ecology* 30.15 (2021): 3869-3881.

Abstract

Shifts in microbial communities represent a rapid response mechanism for host organisms to respond to changes in environmental conditions. Therefore, they are likely to be important in assisting the acclimatization of hosts to seasonal temperature changes as well as to variation in temperatures across a species' range. The Persian/Arabian Gulf is the world's warmest sea, with large seasonal fluctuations in temperature (20 °C - 37 °C) and is connected to the Gulf of Oman which experiences more typical oceanic conditions (<32 °C in the summer). This system is an informative model for understanding how symbiotic microbial assemblages respond to thermal variation across temporal and spatial scales. Here, we elucidate the role of temperature on the microbial gut community of the sea urchin *Echinometra* sp. *EZ* and identify microbial taxa that are tightly correlated with the thermal environment. We generated two independent datasets with a high degree of geographic and temporal resolution. The results show that microbial communities vary across thermally variable habitats, display temporal shifts that correlate with temperature, and can become more disperse as temperatures rise. The relative abundances of several of the same ASVs significantly correlate with temperature in both independent datasets despite the >300 km distance between the furthest sites and the extreme seasonal variations. Notably, over 50% of the ASVs identified from the two datasets belonged to the family Vibrionaceae. Together, our results identify temperature as a robust predictor of community-level variation and putatively highlight specific microbial taxa involved in the response to thermal environment.

Introduction

The microbiome is important in shaping organismal biology in a wide range of eukaryotic species and has been shown to play critical roles in host physiological functions and susceptibility to disease (Gatesoupe, 1999;Bayer et al., 2008;Hentschel et al., 2012). Research describing the distribution, structure, and function of the microbiome has flourished in the last decade, and in marine habitats these studies have generally focused on species important for ecosystem function, including corals (Hadaidi et al., 2017;Ziegler et al., 2019), sponges (Reveillaud et al., 2014), macroalgae (Thurber et al., 2012), seagrasses (Hurtado-McCormick et al., 2019), sea urchins (Carrier et al., 2020), and mangroves (Lin et al., 2019;Trevathan-Tackett et al., 2019). These studies have highlighted dynamic microbial responses to environmental variables that have been implicated in acclimatization of the host to abiotic stressors, either through changes in the abundance of bacteria or by colonization of beneficial bacteria (Reshef et al., 2006;Voolstra and Ziegler, 2020). Given the importance of the microbiome and the fundamental role it plays in overall holobiont function (Bordenstein and Theis, 2015;Pita et al., 2018), understanding the factors that drive microbial change is crucial. This is especially true at a time in which historically rapid climate change is occurring (Konopka, 2009).

Sea surface temperatures are predicted to increase 1-3 °C by 2100 (Collins et al., 2013), representing a significant challenge to marine organisms. Research has unequivocally shown that many marine species respond to increases in temperature through altered physiological functioning (Shama et al., 2016), behavioral variation (Shraim et al., 2017;D'Agostino et al., 2020), disease susceptibility (Howells et al., 2020), and genomic and epigenomic modifications (Liew et al., 2020;Popovic and Riginos, 2020). There are, however, fewer studies characterizing holobiont-associated microbial dynamics in response to elevated temperatures. To date, studies have shown that temperature influences microbial composition in corals (Ziegler et al., 2017a;Wang et al., 2018), sponges (Erwin et al., 2012;Vargas et al., 2020), oysters (Lokmer and Wegner, 2015), anemones (Mortzfeld et al., 2016), and mussels (Li et al., 2019), among others (Brothers et al., 2018). These studies typically focus on a single time-point across a geographic range or temporal variability at a single site (Ward et al., 2017;Woo et al., 2017;Li et al., 2018); rarely are these approaches combined. Further, as almost every step from sample collection to data analysis has been shown to introduce bias, assessing consistent trends using data from many studies that were not processed in the same manner is inherently unreliable (Pollock et al., 2018). Using an approach that includes multiple datasets processed using the same methodology would facilitate a robust characterization of how environmental variables drive microbial dynamics and elucidate conserved microbial responses to temperature.

To this end, we generated two independent datasets investigating the relationship between temperature and the gut microbiota of the sea urchin *Echinometra* sp. *EZ*. We sampled the gut microbiota because it is integrated with host metabolic and immune systems and is a key regulator of host physiology (Sepulveda and Moeller, 2020). *E*. sp. *EZ* are found along an extreme environmental gradient between the Persian/Arabian Gulf (herein the PAG) and the Gulf of Oman (herein the GO) which represents an informative system to understand how environmental variables impact the microbiome. *E*. sp. *EZ* is the most abundant sea urchin in the PAG (densities average $8.6m^{-2}$ across eight sites; Burt JA, unpublished data) and they play a significant role in the health and dynamics of coral ecosystems in the region as major bioeroders (Downing and El-Zahr, 1987). The PAG experiences daily mean summer temperatures regularly >35 °C and extremes exceeding 37 °C (Smith et al., 2017c;Burt et al., 2019) while temperatures in the GO are more typical of oceanic conditions (<32 °C in the summer, Coles (2003)). For our first dataset, we sampled in August and February from six reefs located in the PAG and GO. For the second dataset, we sampled from one reef in the PAG across eight months. We describe community-level differentiation across spatial and temporal gradients, test the hypothesis that rising temperatures result in increased community dispersion, and explore the dynamics between temperature and key microbial taxa.

Materials and Methods

Sample collections

Two sampling strategies were implemented in this study, each generating an independent dataset. First, adult *Echinometra* sp. *EZ* (Ketchum et al., 2018a) were sampled in August 2017 and February 2018 from six sites along the Arabian Peninsula (Table S2.1, Supporting Information) for a total of 183 samples (Figure 2.1). This dataset is referred to as the "summer-winter spatial series." Second, *E.* sp. *EZ* adults were sampled approximately every other month from March 2017 to February 2018 (Figure S2.1, Supporting Information) from Saadiyat reef in the PAG, for a total of 120 samples. This dataset is referred to as the "temporal series." For both sampling strategies we collected 15-17 individuals at each site and/or time point and placed urchins into a 100 L cooler filled with seawater until tissue extractions were performed (within 0-2 h). All of the collection sites are shallow (<7 m) and there are no known thermoclines. The water column is well mixed and a previous study has shown that the difference between bottom temperature and surface temperature is only 0.2 °C in the summer (Paparella et al., 2019). Urchins were cut in half with sterile scissors and a fragment of intestine closest to the anus, and its contents, were removed with sterile forceps and placed in RNA*later* (Ambion). Tubes were then stored in -20 °C after one hour to allow the RNA*later* in infiltrate the tissue.

Environmental variables

The summer-winter spatial series involved six sites along the Arabian Peninsula. We used NOAA's Environmental Research Division Data Access Program (ERDDAP) website to collect sea surface temperature data. The temperature data was downloaded using a bounding box that covered the study area on the day of collection at 12:00:00 UTC (temperature was averaged over one day). In addition, a temperature logger (Onset Hobo Tidbit V2) was deployed on the reef substrate at Saadiyat reef which recorded at

60-minute intervals for the temporal series (Figure S2.1, Supporting Information). To check the accuracy of the data collected from ERDDAP, a Pearson correlation was used to test for a significant correlation between temperature collected using the two different approaches (Figure S2.2, Supporting Information). Chlorophyll concentrations were obtained from MODIS AQUA (https://oceancolor.gsfc.nasa.gov) level 3 monthly averaged data, and salinity data was obtained from a numeric ocean model: the 1/12 Global Hybrid Coordinate Ocean Model (HYCOM; https://www.hycom.org/data/glbu0pt08/expt-91pt2) at 12:00:00 UTC on the day of collections.

DNA extraction and PCR amplification

Total DNA from urchin gut and its contents was extracted according to the optimized protocol (Method 3) described in Ketchum et al (2018b). Briefly, the DNeasy Blood & Tissue Kit (QIAGEN) was used on ~20 µL of homogenized sample according to the manufacturers' protocols for extracting animal tissue, with a few modifications. First, we added our sample to 180 µL of enzymatic lysis buffer (instead of Buffer ATL); 20 mM TrisCL (pH 8), 2 mM sodium EDTA, 1.2% Triton, 20 mg/ml lysozyme (as described in DNeasy Blood and Tissue Manual). The sample was then incubated for 40 minutes at 37 °C. Next, we added 0.5 mL of 0.5 mm zirconia-silica beads (Fisher Scientific) and used a Bead Beater (BioSpec Products) for 40 s at 2,400-3,800 strokes/min and repeated this for a total of three times. Finally, as recommended by the User-Developed Protocol for samples preserved in RNA*later*, we used 40 µL of proteinase K and 220 µL of AL buffer instead of 20 and 200 µL, respectively. These additional steps have been shown to more effectively capture traditionally difficult to lyse taxa, such as gram-positive bacteria (Ketchum et al., 2018).

All samples were then quantified using a Qubit dsDNA High Sensitivity Assay Kit on a Qubit® 2.0 Fluorometer and visualized using a 2% agarose gel. Samples that did not contain sufficient DNA were reextracted and only samples that had a concentration greater than 2 ng/µL were used downstream (n=318). All extraction materials were autoclaved and UV sterilized. DNA extractions and PCRs were performed in a sterile hood. DNA extractions were normalized to a concentration of 1 ng/µL prior to PCR amplification. PCRs were performed in triplicate 25 µL reactions and pooled per individual sample to avoid PCR bias. PCR amplification was performed using the universal V3/V4 PCR primers (forward: 5′- CTACGGGNGGCWGCAG-3′; reverse: 5′-GACTACHVGGGTATCTAATCC-3′) (Klindworth et al., 2013) and followed the protocol described in Ketchum et al (2018).

To account for potential contamination in the reagents and extraction kits, four blanks were sequenced. One of these blanks was processed through the DNA extraction column with water as the input to identify kit contaminants and three of the blanks were water samples that we ran through PCR and subsequently sequenced.

Sequencing and sequence data processing

MiSeq indexing adaptors were added following the Illumina 16S Metagenomic Sequencing Library Preparation protocol and an AxyPrep Mag™ PCR Clean-up Kit (Axygen Biosciences, Corning) was implemented. 16S rRNA gene amplicon libraries were sequenced on the Illumina MiSeq platform using 2 x 300 bp paired-end reads with a

30% PhiX control at the University of North Carolina Charlotte sequencing facility. All 322 (318 microbial samples and four blanks) samples were spread across two sequencing runs with 15 replicate samples on both runs to account for potential run-specific variation. PERMANOVA analyses revealed no significant run-specific effect based on weighted or unweighted unifrac distance ($R^2 = 0.161766$, *p*-value=0.996, $R^2 = 0.4529$, *p*-value=0.996, respectively).

Raw reads and quality information were imported into QIIME2 v.2020.2 (Bolyen et al., 2019). Each sequencing run was independently run through DADA2 (Callahan et al., 2016) with a p-trim-left-f of 17 and a p-trim-left-r of 21 to remove adapters/primers, and a quality filter of p-trunc-len-f of 280 and p-trunc-len-r of 220. The two sequencing runs were then combined into one dataset so that it could be filtered and taxonomically annotated more efficiently. The Naïve Bayes classifier was trained on the region of the target sequences and taxonomy was assigned on the combined dataset using Silva 132 reference sequences, clustered at 99% similarity (Quast et al., 2012). As the taxonomical annotation was performed on the combined dataset, there is direct correspondence between ASVs in the spatial and temporal dataset. Sequences matching to Archaea, chloroplast, mitochondria, or sequences that were present in blanks were filtered from the combined dataset. Using the feature-table rarefy command within QIIME2, the dataset was then rarified to 10 400 sequences per sample and samples which had fewer than 10 400 sequences were removed (a total of five samples were removed; A-AF-6, A-DB-6, F-MS-3, A-RG-11, F-MS-19) unless specific downstream software required a nonrarefied dataset. This dataset was subsequently split back into the summer-winter spatial series and the temporal series.

16S microbial community analysis

Significant differences in microbial community composition were tested with PERMANOVA using the adonis function in *vegan* v.2.5-6 (Oksanen et al., 2013) on the rarified feature table. The statistical significance of environmental factors (temperature, salinity, and chlorophyll concentration) was analyzed using the function envfit within *vegan* and ordination was performed using NMDS based on Bray-Curtis dissimilarity. Pearson correlation was used to test for significant correlations between temperature and salinity in the two datasets (Figure S2.3 and S2.4, Supporting Information). Feature tables were imported into ampvis2 (Andersen et al., 2018) to run Principal Component Analysis (PCA) using Bray-Curtis dissimilarities, a Hellinger transformation, and the "filter_species" flag set to zero. The temporal series was divided into three groups based on the temperature on the day of collection: summer (33-34 °C), intermediate (26-32 °C), and winter (20-24 °C) to reveal large-scale patterns. We calculated alpha diversity metrics, including the Shannon diversity index and observed features metric, through the QIIME2 core diversity metrics plugin. We then tested for significant differences in alpha diversity indices with a Kruskal-Wallis test and applied the Benjamini-Hochberg false discovery rate correction for multiple comparisons (Thissen et al., 2002).

Next, we tested the hypothesis that thermally stressful conditions may result in an increase in microbial community dispersion as the host becomes less able to regulate their microbiome (in line with the Anna Karenina Principle (Zaneveld et al., 2017). In the summer-winter spatial series, we hypothesized that dispersion would increase in August compared to February for the PAG sites due to the uncharacteristically hot summer of 2017 (Paparella et al., 2019), which was likely physiologically stressful for urchins. For the temporal series, we hypothesized that dispersion would increase in summer 2017. To test these hypotheses, analysis of multivariate homogeneity of group dispersion was quantified by conducting permutation tests based on Bray-Curtis dissimilarities and applying Tukey's HSD (PERMDISP2, Anderson et al (2006)) using the package *vegan* v.2.5-6.

In order to elucidate specific microbial signatures that associate with temperature, we used *selbal* which outperforms other methods commonly used in microbiome research by selecting the smallest number of variables with a higher discrimination accuracy (Susin et al., 2020). Prior to running *selbal*, feature tables were filtered to remove ASVs that were not consistently present in the data (ASVs that were found in less than 20% of samples were removed). The analysis of microbiome communities is challenging due to the compositional aspect of these data, as the relative nature of ASV abundances can lead to spurious correlations. To circumvent these issues, *selbal* uses balances, or relative abundances of two groups of taxa, which preserves the principles of compositional data analysis. *Selbal* uses an algorithm that starts with a scan for two taxa whose balances (or log ratios) most closely associate with the response variable, in this case temperature. Once these two taxa are selected, the algorithm then sequentially adds new taxa to the balance such that the predictive power is improved. This process continues until there are no new variables that can improve the optimization or when the maximum number of components are reached. *Selbal* was run on both independent datasets with an "n.fold" or "number of folds in the cross-validation procedure" of 5, 10 iterations, and the "covar" flag set to NULL (as recommended when working with a continuous variable). For the summer-winter spatial series, the temperature data consisted of the output from the

NOAA ERDDAP website. For the temporal series, temperature was derived from the HOBO logger and averaged over the day of sampling. The raw *selbal* output can be found in Figure S2.5 and S2.6, Supporting Information. ASVs that *selbal* found to be predictive of temperature were then extracted from the feature tables and their abundances were used to generate bubble plots (Zorz, 2019). We performed a search using the Nucleotide Basic Local Alignment Search Tool (BLASTn) on the ASVs identified by *selbal* against NCBI's Nucleotide collection (nr/nt) database in order to identify ASVs to species level, where possible. An e-value cutoff of 10^{-8} was used and only BLAST hits with 100% identity were retained.

To explore how ASVs that were predictive of temperature from the *selbal* output fit into the context of the wider microbial network, we conducted network analysis using SpiecEasi (Sparse inverse covariance estimation for ecological association inference, Kurtz et al. (2015)) v1.1.1 on the two datasets (Supplemental Methods, Figure S2.7 and S2.8, Supporting Information). Further, we used Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt v2.0.0, Douglas et al. (2020)) to characterize microbial pathways enriched during warmer conditions (Supplemental Methods, Figure S2.9 and S2.10, Supporting Information).

DEICODE (Martino et al., 2019), a robust Aitchison PCA, was implemented within QIIME2 to identify ASVs responsible for the differences between the PAG and the GO by looking for ASVs which drive the clustering along PC 1. The unrarefied summer-winter spatial series dataset was used with a --p-min-feature-count of 10 and --p-min-sample-count of 500. The ordination file output from DEICODE was then exported and the top ten and bottom ten feature loadings were input into Rstudio and plotted in a heatmap.

Figures were created in R using ggplot2 and illustrations were further stylized in Adobe Illustrator.

Results

Study design

Our study design implemented two sampling strategies and resulted in two independent datasets. The summer-winter spatial series consisted of 183 samples collected from six different reefs in both August and February (Figure 2.1). This dataset covers an extreme environmental gradient across sites as well as two thermally distinct months; August 2017 (summer) and February 2018 (winter). The temporal series consisted of 120 samples across eight months from one sampling site (PAG-SA; Figure 1) with large seasonal temperature variation (~17 °C).

Differences in microbiota composition and diversity in the summer-winter spatial series

The summer-winter spatial series showed differences in overall microbiota composition, as measured by Bray-Curtis dissimilarities. There was clear differentiation between the PAG and the GO along PC 1 with PAG-MS samples clustered in between samples collected from the two seas (Figure 2.2A). Samples collected from PAG-MS in February were differentiated from the rest of the dataset along PC 2. Samples collected in August were differentiated from those collected in February along PC 3 (Figure 2.2B). PERMANOVA analyses showed a significant effect of gulf ($R^2 = 0.1335$, *p*-

value<0.001), site ($R^2 = 0.1757$, *p*-value<0.001), month ($R^2 = 0.0634$, *p*-value<0.001), gulf by month ($R^2 = 0.0367$, *p*-value<0.001), and site by month ($R^2 = 0.1134$, *p*-value<0.001) on community composition (Table S2.2, Supporting Information). Envfit analysis showed a significant correlation between NMDS ordination of the microbial community structure for temperature ($R^2 = 0.1850$, *p*-value<0.003), chlorophyll concentration ($R^2 = 0.4581$, *p*-value<0.003), and salinity ($R^2 = 0.5185$, *p*-value<0.003, Table S2.3 and S2.4, Supporting Information). The NMDS ordination plot showed that these three environmental variable vectors were orthogonal to one another (Figure S2.11, Supporting Information). We used DEICODE to identify the main taxa driving the differences along PC 1 (i.e., between the two seas). We found that ASVs classified as *Spirochaeta* (ASV8874), *Desulfotalea* (ASV14864), Bacteroidia (ASV11638 and ASV13084), *Roseimarinus* (ASV5463), and Marinifilaceae (ASV4027) were more abundant in the GO and two different *Vibrio* taxa (ASV12189 and ASV1103), *Propionigenium* (ASV331), and *Photobacterium* (ASV16623) were more abundant in the PAG (Figure S2.12, Supporting Information).

There were significant differences in the Shannon Diversity Index (Figure 2.2C) when comparing August to February for PAG-DH, PAG-MS, and GO-AF. For the number of observed features (or ASVs, Figure 2.2D), there were significant differences between PAG-DH, PAG-MS, PAG-RG and GO-AF (p-value < 0.05, Kruskal-Wallis). Notably, diversity metrics were higher in August than February for samples from all PAG sites and, inversely, lower in August than February for samples from the GO. Permutation tests of multivariate dispersion showed that dispersion was significantly higher in August than February in PAG-MS (Figure S2.13, Supporting Information).

Specific ASVs correlate with increasing temperature across a wide geographic range and between summer and winter

In the *selbal* analysis, we identified ASVs that were associated with temperature between collection months at each site in the summer-winter spatial series. The numerator and denominator in the *selbal* output contains taxa whose relative abundances increase and decrease, respectively, with increasing temperatures. *Selbal* analyses determined that the optimal number of variables was seven. With these seven ASVs, we obtained a R^2 value of 0.945 for the regression model (Figure 2.3A). The relative abundance of four of the seven ASVs increased with increasing temperatures (uncultured *Vallitalea* sp. [ASV14750], *Propionigenium* sp. [ASV331], and two *Photobacterium* sp. [ASV9510 and ASV10617], see Figure 2.3B) and three of the ASVs decreased with increasing temperature (*Roseimarinus* sp. [ASV16721], *Vibrio* sp.[ASV4517], and Shewanellaceae sp. [ASV4983]).

Differences in microbiota composition and diversity in the temporal series

The temporal series showed differences in overall microbiota composition, as measured by Bray-Curtis dissimilarities (Figure 2.4A). There was clear differentiation on PC1 between samples grouped by temperature at the time of collection. The major axis of community variation (eigenvalues from PC 1, Figure 2.4B) was significantly correlated with the average temperature on the day of collection (R = 0.76, *p*-value < 2.2e-16; Figure S2.14, Supporting Information). PERMANOVA analysis showed a significant effect of season (summer, winter, and intermediate; R^2 = 0.12272, *p*-value<0.001) and month (R^2 = 0.13967, *p*-value<0.001) on community composition (Table S2.5, Supporting Information). Envfit analysis showed a significant correlation between NMDS ordination of the microbial community structure for temperature ($R^2 = 0.7124$, *p*-value<0.003), with weaker correlations for chlorophyll concentration ($R^2 = 0.1951$, *p*-value<0.003), and salinity ($R^2 = 0.1732$, *p*-value<0.003, Table S2.6 and S2.7, Supporting Information). The NMDS ordination plot showed that the salinity vector was orthogonal to the temperature and chlorophyll vectors which were overlapping (Figure S2.15, Supporting Information). There were significant differences between specific months for both Shannon Diversity Indices (Figure 2.4C) and observed features (Figure 2.4D, Supporting Information). For the Shannon Diversity Index, February was significantly lower than March, May, and July and January was significantly lower than May and July. Further, July is significantly higher than November, August is significantly lower than July and May, and May is significantly higher than both November and September. For observed features, February was significantly lower than March, May, July, and November. No significant differences in alpha diversity metrics were found when comparing between the three different temperature groups after a Benjamini & Hochberg correction. Permutation tests of multivariate dispersion analyses revealed that dispersion was significantly higher in August than all other months except for March (Figure S2.16, Supporting Information). No significant differences occurred between the other months.

Specific ASVs correlate with increasing temperature in a sampling dataset with high temporal resolution

The *selbal* analysis identified ASVs that were associated with temperature between collection months in our temporal series. *Selbal* identified that the optimal number of variables was eight and these eight ASVs resulted in a R^2 value of 0.966 for the regression model (Figure 2.5A). The relative abundance of four of the ASVs increased as temperature rose (*Propionigenium* sp. [ASV331], uncultured *Vallitalea* sp. [ASV14750], and 2 ASVs belonged to *Vibrio* [ASV12030 and ASV1103], see Figure 2.5B) and the relative abundance of the other four ASVs decreased (two *Vibrio* spp. [ASV679 and ASV4517], *Roseimarinus* sp. [ASV5483], and *Photobacterium* sp. [ASV4592]). While there were several ASVs which were taxonomically labelled as *Vibrio* spp., they all represent unique sequence variants.

Consistent responses to temperature in both independent datasets

Three identical ASVs were identified as correlating with temperature in the two independent *selbal* analyses (these are denoted in Figure 2.5B by an asterisk in front of the taxonomic ID). These three ASVs were *Propionigenium* sp. (ASV331), uncultured *Vallitalea* sp. (ASV14750), and *Vibrio* sp. (ASV4517). The *Vibrio* sp. strain was further classified through a BLAST search as *Vibrio chagasii* (100% identity score, expect value 0.0, accession number: MT269630.1). *V. chagasii* became relatively less abundant as temperature increased. We were unable to confidently classify any additional ASVs to the species level. Beyond these specific ASVs, there was also taxonomic redundancy in the two analyses. Although ASV identifiers were different, strains of both *Roseimarinus* spp. and *Photobacterium* spp. were identified in both *selbal* analyses. However, while the abundance trend for *Roseimarinus* sp. was consistent across datasets, *Photobacterium* spp. was identified as a numerator in the summer-winter spatial series and as a denominator in the temporal series.

SpiecEasi network analysis retained all of the ASVs output by *selbal*. The *selbal* ASVs did not consistently co-occur with each other in either of the two datasets. Further, they were not identified as keystone microbes based on their hubbiness (Figure S2.7 and S2.8, Supporting Information). PICRUSt2 analysis showed that both datasets were enriched for pathways related to cell wall machinery in the warmer sampling months. Further, both datasets showed an enrichment in sucrose biosynthesis and degradation-related pathways for samples collected in the cooler months (Figure S2.9 and S2.10, Supporting Information).

Discussion

All multicellular organisms associate with a diverse microbiome that contributes to the physiology, development, and fitness of their host (Voolstra and Ziegler, 2020). A fundamental question in animal-microbe interactions is how the structure and function of the microbiome is influenced by environmental variables and which variables are the main drivers of microbial variation. It is particularly important to understand temperature-related microbial dynamics as historically rapid climate change is occurring. While assessing community-level changes across environmental gradients is informative, extracting specific microbial taxa that correlate with environmental variables is crucial for building predictive models for diagnosis of, for example, dysbiosis, stress responses, or disease states (Rivera-Pinto et al., 2018). To this end, we generated two independent datasets that spanned an extreme thermal gradient with high temporal resolution in order to assess both community-level changes, as well as elucidate temperature-predictive microbial taxa.

In the summer-winter spatial series, the majority of the variation in the data was explained by collection site where PC 1 differentiated the PAG from the GO. This microbial differentiation was congruent with previous analyses on the genetic structure of the host species, which showed two populations, one in the PAG and one in the GO (Ketchum et al., 2020). The Musandam collection site is located within the Strait of Hormuz and is geographically situated at the connection between the two seas. The intermediate geographic location is mirrored in the PCA where the samples from Musandam are located between samples from the PAG and GO on PC 1. Additionally, this site is thermally divergent from the two seas so it unclear whether this differentiation on PC 1 is a result of geographic proximity or environmental conditions. Differences in microbiota composition were also identified according to the month of sampling. This variation is likely due to seasonal temperature changes of about 20 °C in the PAG (Coles, 2003) and 10 °C in the GO (Coles, 1997). While there is differentiation in the ordination data which correlates with salinity and chlorophyll concentration, these variables do not follow the same pattern as temperature and their ordination vectors are orthogonal to each other. This makes it unlikely that salinity and chlorophyll concentration are responsible for the differentiation between August and February. It is also possible that the differentiation between August and February is a result of shifting dietary patterns, although this would likely strongly correlate with chlorophyll concentrations and therefore would have been identified in our analysis. Interestingly, measures of microbial diversity were higher and generally more variable in August than February for all PAG sites but in the GO sites this pattern was not evident. This increase in alpha diversity in the PAG may be an adaptive mechanism which allows the urchins to meet their energy and nutrient demands during warmer months or could simply be a result of optimal growth temperatures for a wider variety of microbiota. In our temporal series, we saw a similar pattern in which temperature was a likely driver of community composition; samples clustered by the temperature at which they were collected and we found a significant correlation between temperature and community composition. While other factors like salinity and chlorophyll concentration may contribute to this differentiation (ordination vectors between temperature and chlorophyll concentration were overlapping and temperature and salinity were weakly correlated) temperature was the single most contributing factor with respect to community composition and ordination. Our alpha diversity analyses reveal a temporal oscillation where diversity is generally higher in warmer months, however this relationship is not always significant. Overall, temperature was a robust predictor of community-level variation in both independent datasets.

With both datasets, we tested for the Anna Karenina Principle (AKP) (Zaneveld et al., 2017) which states that stressed organisms may host more stochastic microbiomes than healthy organisms due to their host being unable to regulate their microbiome. In the summer of 2017, there was a mass coral bleaching event that occurred between July and August in the southern PAG and reef-bottom temperatures were among the hottest on record with benthic organisms spending about two months at temperatures exceeding 34 °C (Burt et al., 2019). Therefore, for the summer-winter spatial series, we hypothesized that the 2017 coral bleaching event would be physiologically stressful for urchins and would result in an increase in microbial dispersion in August compared to February. Similarly, we hypothesized that July and August would be physiologically stressful and would result in an increase in dispersion in the temporal series. For the summer-winter spatial series we only found evidence for a significant increase in dispersion in Musandam between August and February. For the temporal series, we found that August was significantly more disperse than all the other months except for March. It is unclear why the summer-winter spatial series did not highlight an increase in dispersion in August for the southern PAG sites while the temporal series did. It is possible that the differences in dispersion for the southern PAG sites were swamped by the difference between August and February for Musandam (dispersion was higher in August for all sites, although not significant). There may also be a different abiotic stressor which we have not measured that is only affected the Musandam populations (e.g., runoff, sewage, or industrial waste). Alternatively, Dhabiya and Ras Ghanada may have been more protected from thermal extremes than PAG-SA. However, this pattern was not evident in Burt et al. (2019), where temperatures on Saadiyat reef paralleled those from Dhabiya and Ras Ghanada. With the increased resolution in the temporal dataset, we were able to show that dispersion was highest during the peak of the coral bleaching event. However, it is unclear why dispersion in March was also quite high. There may be another stressor affecting dispersion in March that is unaccounted for. Finally, we did not observe an increase in dispersion in July when temperatures began to steadily increase. It is possible that we collected urchins in July before they became physiologically stressed or that there is a time-lag between temperature change and microbial shifts. Here we provide some evidence for the AKP, however, future work is needed to ascertain at what temperature *E.* sp. *EZ* becomes physiologically compromised to better understand the relationship between dispersion and thermal stress in sea urchins.

In order to go beyond community-level descriptions, we used *selbal* to identify key microbial signatures whose balances were predictive of temperature. We performed this analysis on both independent datasets and found striking patterns. In addition to the taxonomic redundancy in the ASVs that were identified, three ASVs were identified by *selbal* in both datasets. The consistency of these ASV trends across the two datasets were also reflected in the putative functional profiles associated with temperature. The presence of the same ASVs across datasets may point to a consistent microbial biomarker that is responsible for the maintenance of host homeostasis or is opportunistically proliferating (and may be pathogenic) in response to temperature change.

Over 50% of the ASVs identified from the two *selbal* analyses were strains of Vibrionaceae, highlighting a consistent temperature-dependent response from this family of Proteobacteria. Vibrios, as well as *Photobacterium* spp. (which belongs to the Vibrionaceae family), are found in aquatic habitats throughout the world and occupy a wide variety of ecological niches, sometimes as beneficial symbionts (McFall-Ngai and Ruby, 1991;Thompson et al., 2004) or as potential pathogens (Cervino et al., 2008;Fabbro et al., 2012;Newton et al., 2012). Species in these taxonomic groups have been observed in several species of sea urchins (Beleneva and Kukhlevskii, 2010;Hakim et al., 2016a;Yao et al., 2019) and their abundance has been shown to increase in response to temperature in the coral *Pocillopora damicornis* (Tout et al., 2015). We did not see a pattern in which increased temperatures consistently resulted in an increase in Vibrionaceae strains, rather we found that some strains increased and some decreased in relative abundance. Together, these results point to an interesting relationship between

Vibrionaceae spp. and temperature that warrants further investigation due to the great variability in phenotypic and pathogenic profiles within this family.

Two other ASVs were identified in both datasets as predictive of temperature and were taxonomically labelled as an uncultured strain of *Vallitalea* sp. (from the Lachnospiraceae family) and *Propionigenium* sp., both of which increase in relative abundance when temperature increases. Additionally, *Roseimarinus* spp. were identified in both analyses, although the ASVs differed. *Vallitalea* is a relatively poorly described genus with only three species described to date. These species were isolated from hydrothermal systems (Aissa et al., 2014;Schouw et al., 2018;Sun et al., 2019), which may indicate extreme thermal tolerance. *Vallitalea guaymasensis* has been classified as a bacterial indicator species associated with the coral *Porites lutea* from the Gulf of Thailand and Andaman Sea (Pootakham et al., 2017). However, a study on the conspecific *P. lobata* from the PAG and GO did not find *V. guaymasensis* in their microbiome data (Hadaidi et al., 2017). *Propionigenium* has been identified as one of the most abundant bacterial taxa in the guts of five different sea urchins (Yao et al., 2019) and is likely involved in the metabolism of carbohydrates, amino acids, and lipids (Hakim et al., 2016a). *Propionigenium* has also been shown to be involved in a variety of different host health benefits, not limited to sea urchins. For example, it has been associated with the modulation of the lifespan of the killifish *Nothobranchius furzeri* (Smith et al., 2017d). Unlike *Vallitalea* sp. and *Propionigenium* sp., *Roseimarinus*' abundance decreases as temperature increases. Although the exact biological function that *Roseimarinus* spp. play in its host is unknown, it has been isolated in other marine microbial studies and has been shown to decrease in relative abundance as temperature increases in the mussel *Mytilus galloprovincialis* (Li et al., 2019). Although beyond the scope of this study, it would beneficial to determine whether these changes in microbial abundances are a direct consequence of temperature change on the microbes (i.e., changes in relative abundance of microbes across season or site is simply due to different optimal growth temperatures) or indirect selection of microbial abundance by the host requiring different microbes in response to changing metabolic needs.

Assessing how environmental variables drive microbial diversity underpins our understanding of the relationships between hosts and their microbiome. Our sampling strategy has allowed us to characterize the microbial gut community across a wide geographic and temporal span and implicate temperature as a regulator of community composition. We take this further by identifying bacterial taxa whose abundances correlate with temperature and find a consistent signature response in the two independent datasets. As the PAG is the warmest sea in the world, it is a highly informative model for our understanding of the microbial response to thermal extremes as well as for predicting microbial shifts in response to climate change. The observed patterns presented in this study align well with the idea that acclimatization through restructuring of the microbial community constitutes a dynamic environmental response mechanism. We identified several key microbial taxa that may either represent opportunistic pathogens or be crucial for the maintenance of host homeostasis during thermal extremes.

Acknowledgements

The field work was carried out using the Marine Biology Core Technology Platform resources at New York University Abu Dhabi, and this support is greatly appreciated. We thank the Environmental Agency – Abu Dhabi for permits.

References

Aissa, F. B., Postec, A., Erauso, G., Payri, C., Pelletier, B., Hamdi, M., . . . Fardeau, M.-L. (2014). *Vallitalea pronyensis* sp. nov., isolated from a marine alkaline hydrothermal chimney. *International Journal of Systematic and Evolutionary Microbiology, 64*(4), 1160-1165.

Andersen, K. S., Kirkegaard, R. H., Karst, S. M., & Albertsen, M. (2018). ampvis2: an R package to analyse and visualise 16S rRNA amplicon data. *bioRxiv*, 299537.

Anderson, M. J., Ellingsen, K. E., & McArdle, B. H. (2006). Multivariate dispersion as a measure of beta diversity. *Ecology Letters, 9*(6), 683-693.

Bayer, K., Schmitt, S., & Hentschel, U. (2008). Physiology, phylogeny and in situ evidence for bacterial and archaeal nitrifiers in the marine sponge *Aplysina aerophoba*. *Environmental Microbiology, 10*(11), 2942-2955.

Beleneva, I., & Kukhlevskii, A. (2010). Characterization of *Vibrio gigantis* and *Vibrio pomeroyi* isolated from invertebrates of Peter the Great Bay, Sea of Japan. *Microbiology, 79*(3), 402-407.

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., . . . Asnicar, F. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology, 37*(8), 852-857.

Bordenstein, S. R., & Theis, K. R. (2015). Host biology in light of the microbiome: ten principles of holobionts and hologenomes. *PLOS Biology, 13*(8), e1002226. Retrieved from https://doi.org/10.1371/journal.pbio.1002226

Brothers, C. J., Van Der Pol, W. J., Morrow, C. D., Hakim, J. A., Koo, H., &

McClintock, J. B. (2018). Ocean warming alters predicted microbiome functionality in a common sea urchin. *Proceedings of the Royal Society B, 285*(1881), 20180340.

Burt, J. A., Paparella, F., Al-Mansoori, N., Al-Mansoori, A., & Al-Jailani, H. (2019). Causes and consequences of the 2017 coral bleaching event in the southern Persian/Arabian Gulf. *Coral Reefs, 38*(4), 567-589.

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*(7), 581.

Carrier, T. J., Lessios, H. A., & Reitzel, A. M. (2020). Eggs of echinoids separated by the Isthmus of Panama harbor divergent microbiota. *Marine Ecology Progress Series, 648*, 169-177.

Cervino, J., Thompson, F., Gomez□Gil, B., Lorence, E., Goreau, T., Hayes, R., . . . Bartels, E. (2008). The *Vibrio* core group induces yellow band disease in Caribbean and Indo□Pacific reef□building corals. *Journal of Applied Microbiology, 105*(5), 1658-1671.

Coles, S. (1997). Reef corals occurring in a highly fluctuating temperature environment at Fahal Island, Gulf of Oman (Indian Ocean). *Coral Reefs, 16*(4), 269-272.

Coles, S. L. (2003). Coral species diversity and environmental factors in the Arabian Gulf and the Gulf of Oman: a comparison to the Indo-Pacific region. *Atoll Research Bulletin*.

Collins, M., T., D. Q., F. Stocker, Plattner, G.-K., Tignor, M., Allen, S. K., . . . Midgley,

P. M. (2013). In Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. *Cambridge University Press*, 1029–1136.

D'Agostino, D., Burt, J. A., Reader, T., Vaughan, G. O., Chapman, B. B., Santinelli, V., . . . Feary, D. A. (2020). The influence of thermal extremes on coral reef fish behaviour in the Arabian/Persian Gulf. *Coral Reefs, 39*(3), 733-744.

Douglas, G. M., Maffei, V. J., Zaneveld, J. R., Yurgel, S. N., Brown, J. R., Taylor, C. M., . . . Langille, M. G. (2020). PICRUSt2 for prediction of metagenome functions. *Nature Biotechnology, 38*(6), 685-688.

Downing, N., & El-Zahr, C. (1987). Gut evacuation and filling rates in the rock-boring sea urchin, *Echinometra mathaei*. *Bulletin of Marine Science, 41*(2), 579-584.

Erwin, P. M., Pita, L., López-Legentil, S., & Turon, X. (2012). Stability of sponge-associated bacteria over large seasonal shifts in temperature and irradiance. *Applied and Environmental Microbiology, 78*(20), 7358-7368.

Fabbro, C., Celussi, M., Russell, H., & Del Negro, P. (2012). Phenotypic and genetic diversity of coexisting Listonella anguillarum, *Vibrio harveyi* and *Vibrio chagassi* recovered from skin haemorrhages of diseased sand smelt, *Atherina boyeri*, in the Gulf of Trieste (NE Adriatic Sea). *Letters in Applied Microbiology, 54*(2), 153-159.

Gatesoupe, F. J. (1999). The use of probiotics in aquaculture. *Aquaculture, 180*(1-2), 147-165.

Hadaidi, G., Röthig, T., Yum, L. K., Ziegler, M., Arif, C., Roder, C., . . . Voolstra, C. R.

(2017). Stable mucus-associated bacterial communities in bleached and healthy corals of *Porites lobata* from the Arabian Seas. *Scientific Reports, 7*(1), 1-11.

Hakim, J. A., Koo, H., Kumar, R., Lefkowitz, E. J., Morrow, C. D., Powell, M. L., . . . Bej, A. K. (2016). The gut microbiome of the sea urchin, *Lytechinus variegatus*, from its natural habitat demonstrates selective attributes of microbial taxa and predictive metabolic profiles. *FEMS Microbiology Ecology, 92*(9), fiw146-fiw146. doi:10.1093/femsec/

Hentschel, U., Piel, J., Degnan, S. M., & Taylor, M. W. (2012). Genomic insights into the marine sponge microbiome. *Nature Reviews Microbiology, 10*(9), 641-654.

Howells, E. J., Vaughan, G., Work, T. M., Burt, J., & Abrego, D. (2020). Annual outbreaks of coral disease coincide with extreme seasonal warming. *Coral Reefs, 39*, 771-781.

Hurtado-McCormick, V., Kahlke, T., Petrou, K., Jeffries, T., Ralph, P. J., & Seymour, J. R. (2019). Regional and microenvironmental scale characterization of the Z*ostera muelleri* seagrass microbiome. *Frontiers in Microbiology, 10*, 1011.

Ketchum, R. N., DeBiasse, M. B., Ryan, J. F., Burt, J. A., & Reitzel, A. M. (2018). The complete mitochondrial genome of the sea urchin, *Echinometra* sp. *EZ. Mitochondrial DNA Part B, 3*(2), 1225-1227.

Ketchum, R. N., Smith, E. G., DeBiasse, M. B., Vaughan, G. O., McParland, D., Leach, W. B., . . . Reitzel, A. M. (2020). Population genomic analyses of the sea urchin *Echinometra* sp. *EZ* across an extreme environmental gradient. *Genome Biology and Evolution*.

Ketchum, R. N., Smith, E. G., Vaughan, G. O., Phippen, B. L., McParland, D., Al

Mansoori, N., . . . Reitzel, A. M. (2018). DNA extraction method plays a significant role when defining bacterial community composition in the marine invertebrate *Echinometra mathaei*. *Frontiers in Marine Science, 5*, 255.

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., & Glöckner, F. O. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Research, 41*(1). doi:10.1093/nar/gks808

Konopka, A. (2009). What is microbial community ecology? *The ISME Journal, 3*(11), 1223-1230.

Kurtz, Z. D., Müller, C. L., Miraldi, E. R., Littman, D. R., Blaser, M. J., & Bonneau, R. A. (2015). Sparse and compositionally robust inference of microbial ecological networks. *PLoS Computional Biology, 11*(5), e1004226.

Li, Y.-F., Xu, J.-K., Chen, Y.-W., Ding, W.-Y., Shao, A.-Q., Liang, X., . . . Yang, J.-L. (2019). Characterization of gut microbiome in the mussel *Mytilus galloprovincialis* in response to thermal stress. *Frontiers in Physiology, 10*, 1086.

Li, Y.-F., Yang, N., Liang, X., Yoshida, A., Osatomi, K., Power, D., . . . Yang, J.-L. (2018). Elevated seawater temperatures decrease microbial diversity in the gut of *Mytilus coruscus*. *Frontiers in Physiology, 9*, 839.

Liew, Y. J., Howells, E. J., Wang, X., Michell, C. T., Burt, J. A., Idaghdour, Y., & Aranda, M. (2020). Intergenerational epigenetic inheritance in reef-building corals. *Nature Climate Change, 10*(3), 254-259.

Lin, X., Hetharua, B., Lin, L., Xu, H., Zheng, T., He, Z., & Tian, Y. (2019). Mangrove sediment microbiome: adaptive microbial assemblages and their routed biogeochemical processes in Yunxiao mangrove national nature reserve, China. *Microbial Ecology, 78*(1), 57-69.

Lokmer, A., & Wegner, K. M. (2015). Hemolymph microbiome of Pacific oysters in response to temperature, temperature stress and infection. *The ISME Journal, 9*(3), 670-682.

Martino, C., Morton, J. T., Marotz, C. A., Thompson, L. R., Tripathi, A., Knight, R., & Zengler, K. (2019). A novel sparse compositional technique reveals microbial perturbations. *MSystems, 4*(1).

McFall-Ngai, M. J., & Ruby, E. G. (1991). Symbiont recognition and subsequent morphogenesis as early events in an animal-bacterial mutualism. *Science, 254*(5037), 1491-1494.

Mortzfeld, B. M., Urbanski, S., Reitzel, A. M., Künzel, S., Technau, U., & Fraune, S. (2016). Response of bacterial colonization in *Nematostella vectensis* to development, environment and biogeography. *Environmental Microbiology, 18*(6), 1764-1781.

Newton, A., Kendall, M., Vugia, D. J., Henao, O. L., & Mahon, B. E. (2012). Increasing Rates of Vibriosis in the United States, 1996–2010: Review of Surveillance Data From 2 Systems. *Clinical Infectious Diseases, 54*(suppl_5), S391-S395. doi:10.1093/cid/cis243

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'hara, R., . . . Wagner, H. (2013). Package 'vegan'. *Community ecology package, version, 2*(9), 1-295.

Paparella, F., Xu, C., Vaughan, G. O., & Burt, J. A. (2019). Coral bleaching in the Persian/Arabian Gulf is modulated by summer winds. *Frontiers in Marine Science, 6*, 205.

Pita, L., Rix, L., Slaby, B. M., Franke, A., & Hentschel, U. (2018). The sponge holobiont in a changing ocean: from microbes to ecosystems. *Microbiome, 6*(1), 46.

Pollock, J., Glendinning, L., Wisedchanwet, T., & Watson, M. (2018). The madness of microbiome: Attempting to find consensus "best practice" for 16S microbiome studies. *Applied and Environmental Microbiology, 84*, e02627-02617. doi:10.1128/aem.02627-17

Pootakham, W., Mhuantong, W., Yoocha, T., Putchim, L., Sonthirod, C., Naktang, C., . . . Tangphatsornruang, S. (2017). High resolution profiling of coral-associated bacterial communities using full-length 16S rRNA sequence data from PacBio SMRT sequencing system. *Scientific Reports, 7*(1), 1-14.

Popovic, I., & Riginos, C. (2020). Comparative genomics reveals divergent thermal selection in warm□and cold□tolerant marine mussels. *Molecular Ecology, 29*(3), 519-535.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., . . . Glöckner, F. O. (2012). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research, 41*(D1), D590-D596.

Reshef, L., Koren, O., Loya, Y., Zilber□Rosenberg, I., & Rosenberg, E. (2006). The coral probiotic hypothesis. *Environmental Microbiology, 8*(12), 2068-2073.

Reveillaud, J., Maignien, L., Eren, A. M., Huber, J. A., Apprill, A., Sogin, M. L., &

Vanreusel, A. (2014). Host-specificity among abundant and rare taxa in the sponge microbiome. *The ISME Journal, 8*, 1198. doi:10.1038/ismej.2013.227

Rivera-Pinto, J., Egozcue, J., Pawlowsky-Glahn, V., Paredes, R., Noguera-Julian, M., & Calle, M. (2018). Balances: a New Perspective for Microbiome Analysis. In (Vol. 3). *mSystems*: August.

Schouw, A., Vulcano, F., Roalkvam, I., Hocking, W. P., Reeves, E., Stokke, R., . . . Steen, I. H. (2018). Genome analysis of *Vallitalea guaymasensis* strain L81 isolated from a deep-sea hydrothermal vent system. *Microorganisms, 6*(3), 63.

Sepulveda, J., & Moeller, A. H. (2020). The effects of temperature on animal gut microbiomes. *Frontiers in Microbiology, 11*.

Shama, L. N., Mark, F. C., Strobel, A., Lokmer, A., John, U., & Mathias Wegner, K. (2016). Transgenerational effects persist down the maternal line in marine sticklebacks: gene expression matches physiology in a warming ocean. *Evolutionary Applications, 9*(9), 1096-1111.

Shraim, R., Dieng, M. M., Vinu, M., Vaughan, G., McParland, D., Idaghdour, Y., & Burt, J. A. (2017). Environmental Extremes Are Associated with Dietary Patterns in Arabian Gulf Reef Fishes. *Frontiers in Marine Science, 4*, 285.

Smith, E. G., Vaughan, G. O., Ketchum, R. N., McParland, D., & Burt, J. A. (2017). Symbiont community stability through severe coral bleaching in a thermally extreme lagoon. *Scientific Reports, 7*(1), 2428. doi:10.1038/s41598-017-01569-8

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

Sun, Y.-T., Zhou, N., Wang, B.-J., Liu, X.-D., Jiang, C.-Y., Ge, X., & Liu, S.-J. (2019). *Vallitalea okinawensis* sp. nov., isolated from Okinawa trough sediment and emended description of the genus *Vallitalea*. *International Journal of Sytematic and Evolutionary Microbiology, 69*(2), 404-410.

Susin, A., Wang, Y., Lê Cao, K.-A., & Calle, M. L. (2020). Variable selection in microbiome compositional data analysis. *NAR Genomics and Bioinformatics, 2*(2), lqaa029.

Thissen, D., Steinberg, L., & Kuang, D. (2002). Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive rate in multiple comparisons. *Journal of Educational and Behavioral Statistics, 27*(1), 77-83.

Thompson, F. L., Iida, T., & Swings, J. (2004). Biodiversity of Vibrios. *Microbiology and Molecular Biology Reviews, 68*(3), 403-431.

Thurber, R. V., Burkepile, D. E., Correa, A. M., Thurber, A. R., Shantz, A. A., Welsh, R., . . . Rosales, S. (2012). Macroalgae decrease growth and alter microbial community structure of the reef-building coral, *Porites astreoides*. *PLOS One, 7*(9), e44246.

Tout, J., Siboni, N., Messer, L. F., Garren, M., Stocker, R., Webster, N. S., . . . Seymour, J. R. (2015). Increased seawater temperature increases the abundance and alters the structure of natural *Vibrio* populations associated with the coral *Pocillopora damicornis*. *Frontiers in Microbiology, 6*, 432.

Trevathan-Tackett, S. M., Sherman, C. D., Huggett, M. J., Campbell, A. H., Laverock, B., Hurtado-McCormick, V., . . . Ainsworth, T. D. (2019). A horizon scan of priorities for coastal marine microbiome research. *Nature Ecology & Evolution*, 1-12.

Vargas, S., Leiva, L., & Wörheide, G. (2020). Short-term exposure to high-water temperature causes a shift in the microbiome of the common aquarium sponge *Lendenfeldia chondrodes*. *bioRxiv*.

Voolstra, C. R., & Ziegler, M. (2020). Adapting with microbial help: Microbiome flexibility facilitates rapid responses to environmental change. *BioEssays, 42*(7), 2000004.

Wang, L., Shantz, A. A., Payet, J. P., Sharpton, T. J., Foster, A., Burkepile, D. E., & Vega Thurber, R. (2018). Corals and their microbiomes are differentially affected by exposure to elevated nutrients and a natural thermal anomaly. *Frontiers in Marine Science, 5*, 101.

Ward, C. S., Yung, C.-M., Davis, K. M., Blinebry, S. K., Williams, T. C., Johnson, Z. I., & Hunt, D. E. (2017). Annual community patterns are driven by seasonal switching between closely related marine bacteria. *The ISME Journal, 11*(6), 1412-1422.

Woo, S., Yang, S.-H., Chen, H.-J., Tseng, Y.-F., Hwang, S.-J., De Palmas, S., . . . Yum, S. (2017). Geographical variations in bacterial communities associated with soft coral *Scleronephthya gracillimum*. *PLOS One, 12*(8), e0183663.

Yao, Q., Yu, K., Liang, J., Wang, Y., Hu, B., Huang, X., . . . Qin, Z. (2019). The composition, diversity and predictive metabolic profiles of bacteria associated with the gut digesta of five sea urchins in Luhuitou fringing reef (northern South China Sea). *Frontiers in Microbiology, 10*, 1168.

Zaneveld, J. R., McMinds, R., & Thurber, R. V. (2017). Stress and stability: applying the Anna Karenina principle to animal microbiomes. *Nature Microbiology, 2*(9), 1-8.

Ziegler, M., Grupstra, C. G., Barreto, M. M., Eaton, M., BaOmar, J., Zubier, K., . . . Voolstra, C. R. (2019). Coral bacterial community structure responds to environmental change in a host-specific manner. *Nature Communications, 10*(1), 1-11.

Ziegler, M., Seneca, F. O., Yum, L. K., Palumbi, S. R., & Voolstra, C. R. (2017). Bacterial community dynamics are linked to patterns of coral heat tolerance. *Nature Communications, 8*(1), 1-8.

Zorz, J. (2019). https://jkzorz.github.io/2019/06/05/Bubble-plots.html.

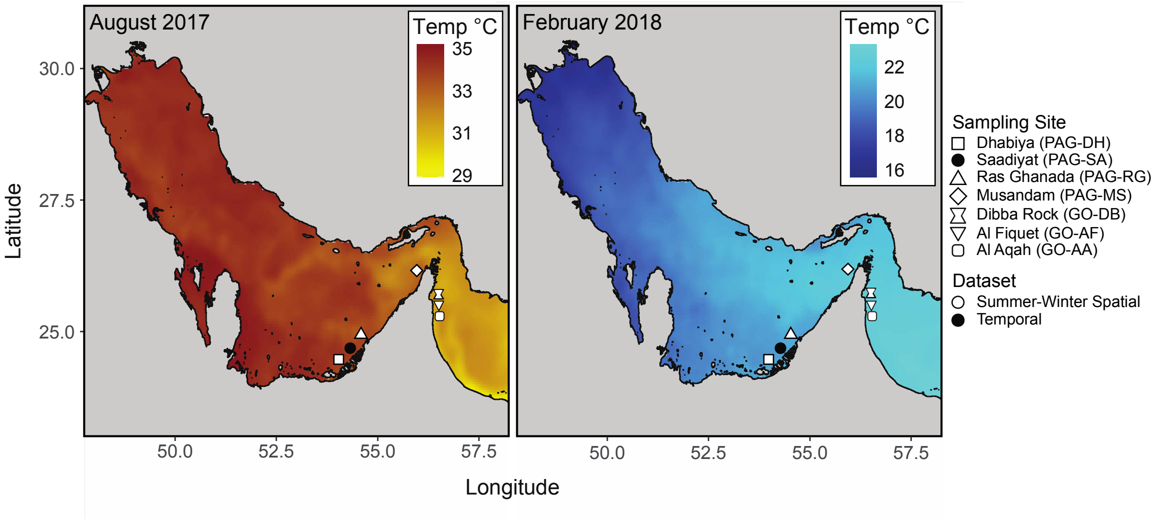


Figure 2.1: Sampling map of all collection sites in August 2017 and February 2018. The six sampling sites for the summer-winter spatial series are shown in white and the one sampling site for the temporal series is shown in black. Temperature data was downloaded from NOAA's Environmental Research Division Data Access Program (ERDDAP) website (temperatures were averaged over one day and extracted from Aug 15, 2017 and Feb 15, 2018 at 12:00:00 UTC for this plot).

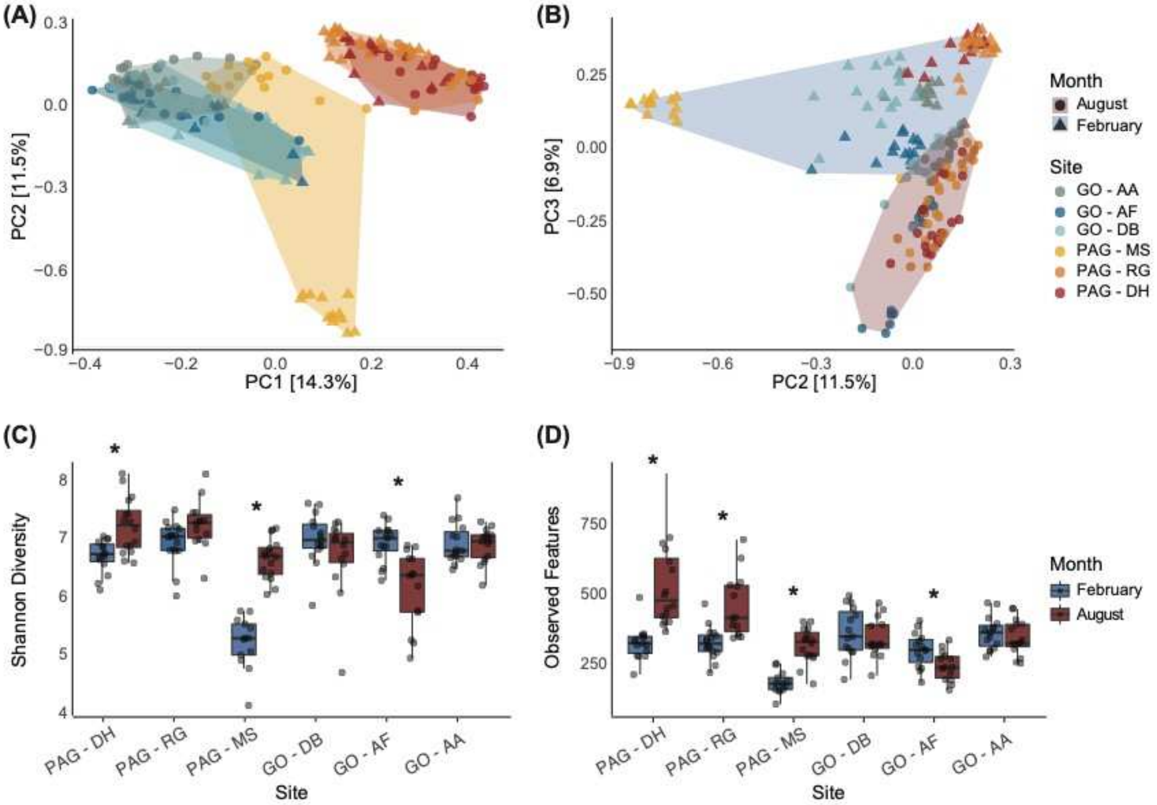


Figure 2.2: Microbial diversity from the summer-winter spatial series. **A)** Principal Component Analysis based on Bray-Curtis dissimilarities after a Hellinger Transform of bacterial communities from three sites within the PAG and three sites within the GO. The shaded color represents collection sites and the shape of the point represents month of collection. **B)** PCA showing principal component two and three and shaded according to month of collection. **C)** Box plots of Shannon Diversity metrics for all collection sites and split by month of collection. **D)** Box plots of observed features for all collection sites and split by month of collection. Asterisks denote significant differences in Shannon Diversity or number of observed features when comparing between month of collection within each site, respectively.

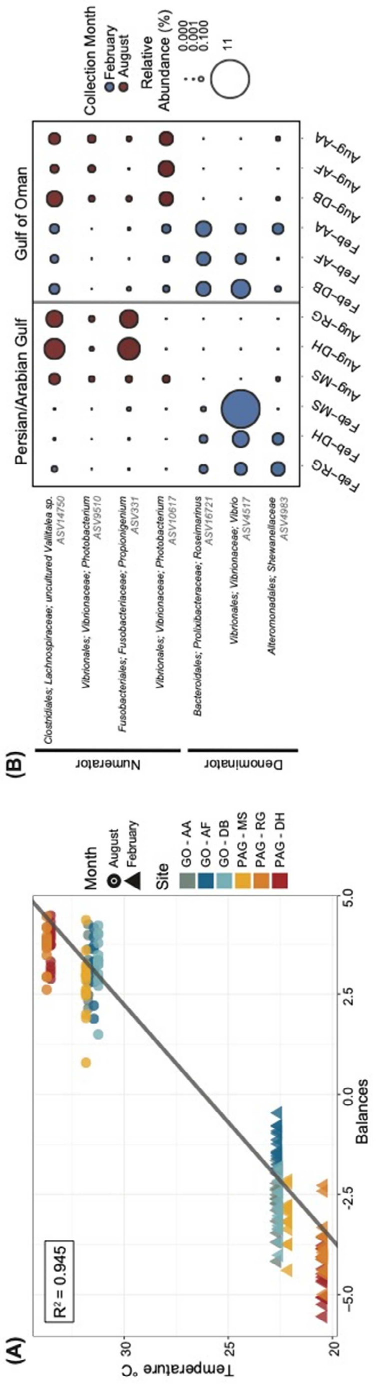


Figure 2.3: **(A)** Regression model of the seven variables that define the balance for the summer-winter spatial series. The shape of the points represents month of collection and color represents the site where the samples were collected. **(B)** The variables that define the balance are taxonomically annotated in the bubble plot and their position in the balance is labelled "Numerator" or "Denominator." The size of the bubble corresponds to relative abundance in the count table and the color of the bubble represents the month the samples were collected.

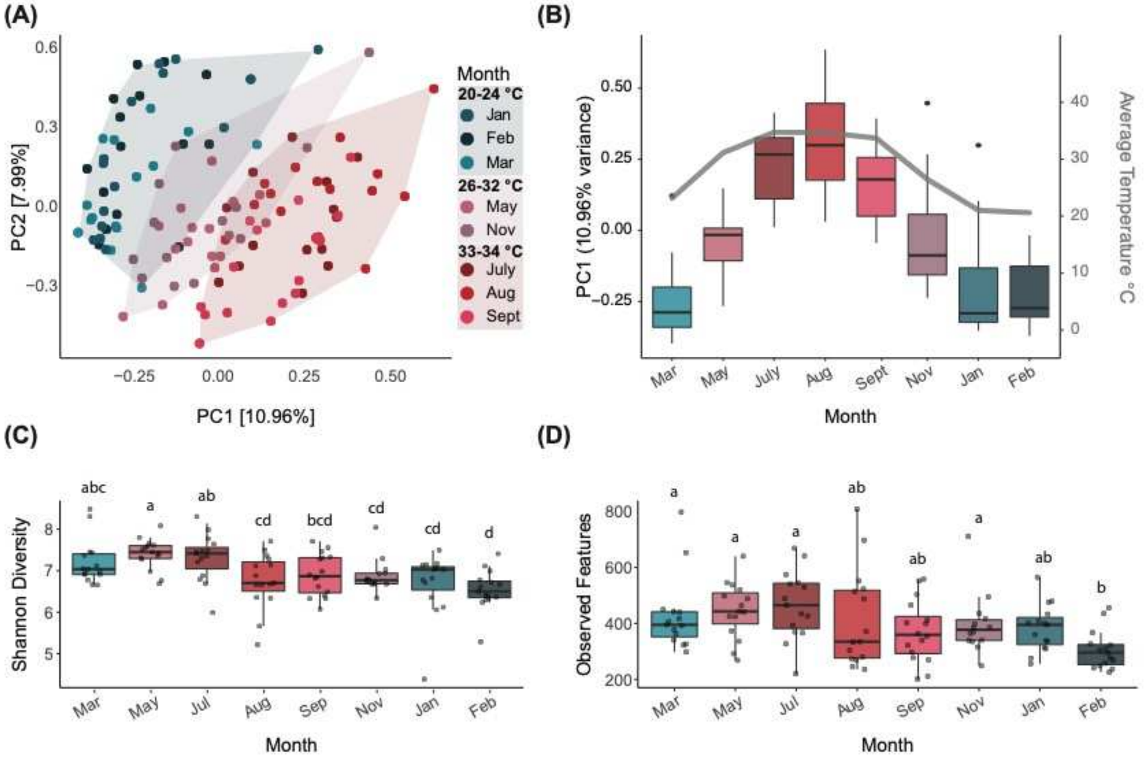


Figure 2.4: Microbial diversity of the temporal series. **A)** PCA based on Bray-Curtis dissimilarities after a Hellinger Transform of bacterial communities across eight months from Saadiyat reef, in the PAG. The color of the point represents collection month and the shaded regions represent the temperatures at the time of collection. **B)** The eigenvalues from the first principal component were extracted and plotted alongside the average temperature on the day of collection**. C)** Box plots of Shannon Diversity metrics for all collection months. **D)** Box plots of observed features for all collection months. For plots C and D, boxes that do not share similar letters denote statistical significance ($p<0.05$, Kruskal-Wallis test with a Benjamini & Hochberg correction).

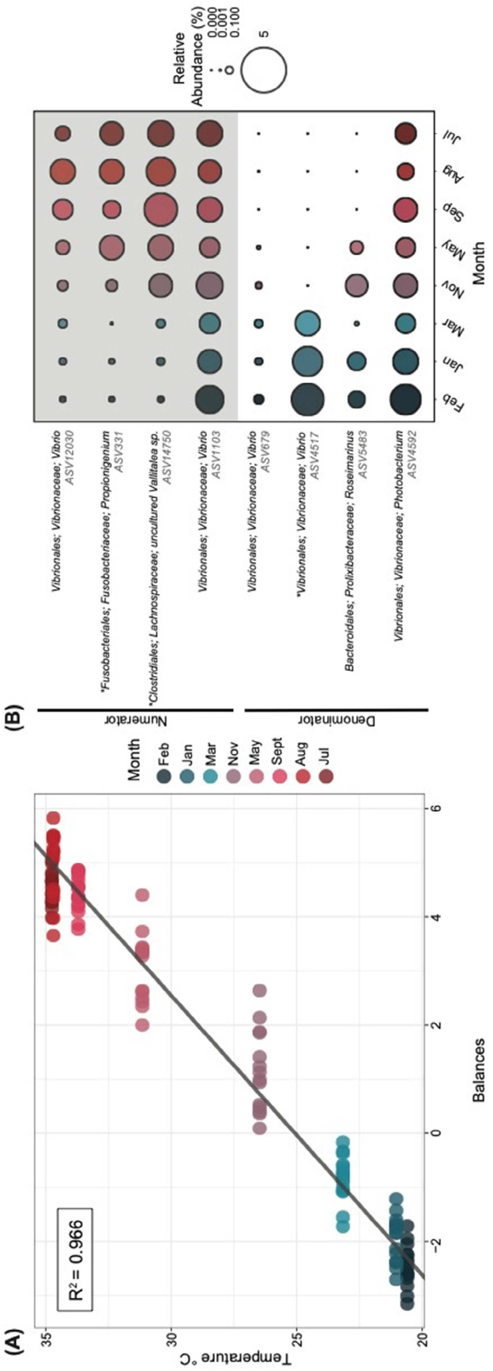


Figure 2.5: **(A)** Regression model of the eight variables that define the balance for the temporal series. The color of the points represents the month at which they were collected. **(B)** The variables that define the balance are taxonomically annotated in the bubble plot and their position in the balance is labelled “Numerator” or “Denominator.” The size of the bubble corresponds to relative abundance in the count table and the color of the bubble represents month of collection. ASVs which were retained from both datasets in *selbal* analysis are denoted by an asterisk.

Table S2.1: Sampling site names and ID, coordinates of sampling sites, and temperatures in August 2017 and February 2018 based off of NOAA's Environmental Research Division Data Access Program (ERDDAP) website.

Sampling Site	**Site ID**	**Latitude**	**Longitude**	**Temp Aug '17**	**Temp Feb '18**
Dhabiya	PAG-DH	24° 21' 55.8"	54° 6' 2.8794"	33.53	20.44
Saadiyat	PAG-SA	24° 35' 56.4"	54° 25' 17.4"	33.76	20.51
Ras Ghanada	PAG-RG	24° 50' 53.394"	54° 41' 25.065"	33.72	20.41
Musandam	PAG-MS	26° 10' 29.7264"	56° 10' 24.6216"	31.83	22.12
Dibba Rock	GO-DB	25° 36' 11.286"	56° 20' 54.5172"	31.24	22.55
Al Fiquet	GO-AF	25° 33' 45.6624"	56° 21' 12.5208"	31.46	22.6
Al Aqah	GO-AA	25° 29' 34.605"	56° 21' 48.6828"	31.73	22.64

Table S2.2: Statistical output for the PERMANOVA analyses of the summer-winter spatial series microbiome composition based on Bray-Curtis dissimilarities with 999 permutations.

Adonis Model	df	Sum of Squares	Mean Squares	F.Model	R^2	*p*-value
Gulf	1	6.721	6.7207	46.424	0.13349	0.001***
Site	4	8.846	2.2115	15.276	0.17570	0.001***
Month	1	3.191	3.1907	22.040	0.06337	0.001***
Gulf*Month	1	1.847	1.8471	12.759	0.03669	0.001***
Site*Month	4	5.711	1.4277	9.862	0.11343	0.001***
Residuals	166	24.032	0.1448		0.47732	
Total	177	50.347			1.00000	

Table S2.3: PERMANOVA analyses of the rarified summer-winter spatial series microbiome composition based on Bray-Curtis dissimilarities with 999 permutations. Temperature data was collected from NOAA's ERDDAP website on the day of collections at 12:00:00 UTC, chlorophyll concentration was obtained from MODIS AQUA level 3 monthly averaged data, and salinity data was obtained from a numeric ocean model: the 1/12 Global Hybrid Coordinate Ocean Model (HYCOM) at 12:00:00 UTC on the day of collections.

Adonis Model	df	Sum of Squares	Mean Squares	F.Model	R^2	*p*-value
Temperature	1	3.162	3.1621	15.364	0.06281	0.001***
Chlorophyll Concentration	1	5.614	5.6138	27.277	0.11150	0.001***
Salinity	1	5.761	5.7605	27.990	0.11442	0.001***
Residuals	174	35.810	0.2058		0.71128	
Total	177	50.347			1.00000	

Table S2.4: Results generated by EnvFit in *vegan v.2.5-6* for the rarified summer-winter spatial series. Temperature data was collected from NOAA's ERDDAP website on the day of collections at 12:00:00 UTC, chlorophyll concentration was obtained from MODIS AQUA level 3 monthly averaged data (4 km), and salinity data was obtained from a numeric ocean model: the 1/12 Global Hybrid Coordinate Ocean Model (HYCOM) at 12:00:00 UTC on the day of collections. The significance was determined based on 999 permutations and Bonferroni's correction was used to adjust the *p*-values for the multiple test problem. Due to the proximity of the reefs to the coast, we were sometimes unable to retrieve values/data for certain locations. In these cases, we used data from the nearest pixel with available data (always within 10 km from the exact collection location).

	NMDS1	NMDS2	R^2	*p*-value
Temperature	0.71251	-0.70167	0.1850	0.003 ***
Chlorophyll Concentration	0.32214	0.94669	0.4581	0.003 ***
Salinity	0.86985	0.49332	0.5185	0.003 ***

Table S2.5: Statistical output for the PERMANOVA analyses of the temporal series microbiome composition based on Bray-Curtis dissimilarities with 999 permutations.

Adonis Model	df	Sum of Squares	Mean Squares	F.Model	R^2	*p*-value
Season	2	2.9944	1.49719	9.3171	0.12272	0.001***
Month	5	3.4079	0.68159	4.2415	0.13967	0.001***
Residuals	112	17.9976	0.16069		0.73761	
Total	119	24.3999			1.00000	

Table S2.6: PERMANOVA analyses of the rarified temporal series microbiome composition based on Bray-Curtis dissimilarities with 999 permutations. Temperature data was collected from the HOBO logger, chlorophyll concentration was obtained from MODIS AQUA level 3 monthly averaged data (4km), and salinity data was obtained from a numeric ocean model: the 1/12 Global Hybrid Coordinate Ocean Model (HYCOM) at 12:00:00 UTC on the day of collection. Importantly, data was not available for the month of July 2017 and so we removed these samples for this analysis.

Adonis Model	df	Sum of Squares	Mean Squares	F.Model	R^2	*p*-value
Temperature	1	2.0563	2.05629	11.8942	0.09663	0.001***
Chlorophyll Concentration	1	0.7534	0.75340	4.3572	0.03540	0.001***
Salinity	1	1.0069	1.00692	5.8235	0.04732	0.001***
Residuals	101	17.4636	0.17176		0.82065	
Total	104	21.2802			1.00000	

Table S2.7: Results generated by EnvFit in *vegan v.2.5-6* for the rarified temporal series. Temperature data was collected from the HOBO logger, chlorophyll concentration was obtained from MODIS AQUA level 3 monthly averaged data (4km), and salinity data was obtained from a numeric ocean model: the 1/12 Global Hybrid Coordinate Ocean Model (HYCOM) at 12:00:00 UTC on the day of collection. The significance was determined based on 999 permutations and Bonferroni's correction was used to adjust the *p*-values for the multiple test problem. Please note that for this series, we removed samples collected in July for this analysis as we were unable to collect chlorophyll data for this month.

	NMDS1	NMDS2	R^2	*p*-value
Temperature	0.13227	0.99121	0.7124	0.003 ***
Chlorophyll Concentration	0.10873	0.99407	0.1951	0.003 ***
Salinity	-0.23572	-0.97182	0.1732	0.003 ***

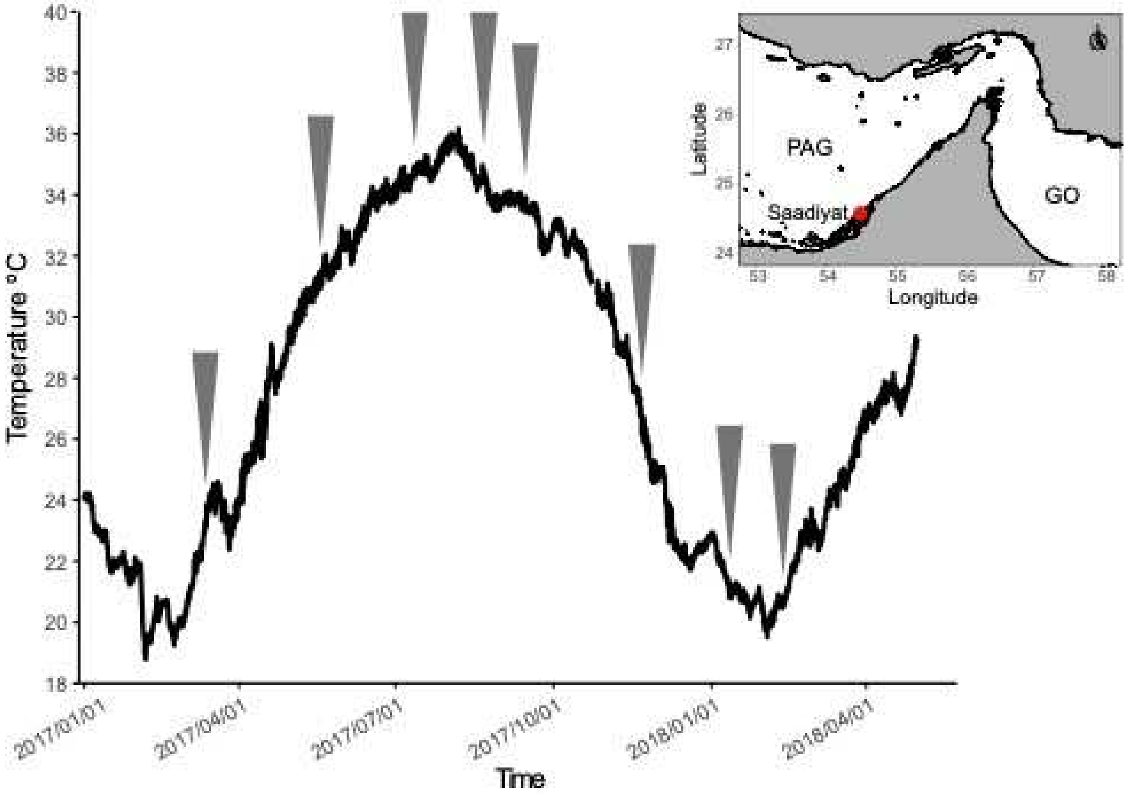


Figure S2.1: Temperature plot with HOBO logger data from Saadiyat reef within the PAG. Gray arrows indicate the dates where samples were collected. Exact dates include: 2017/03/12, 2017/05/18, 2017/07/12, 2017/08/21, 2017/09/14, 2017/11/21, 2018/01/11, and 2018/02/11.

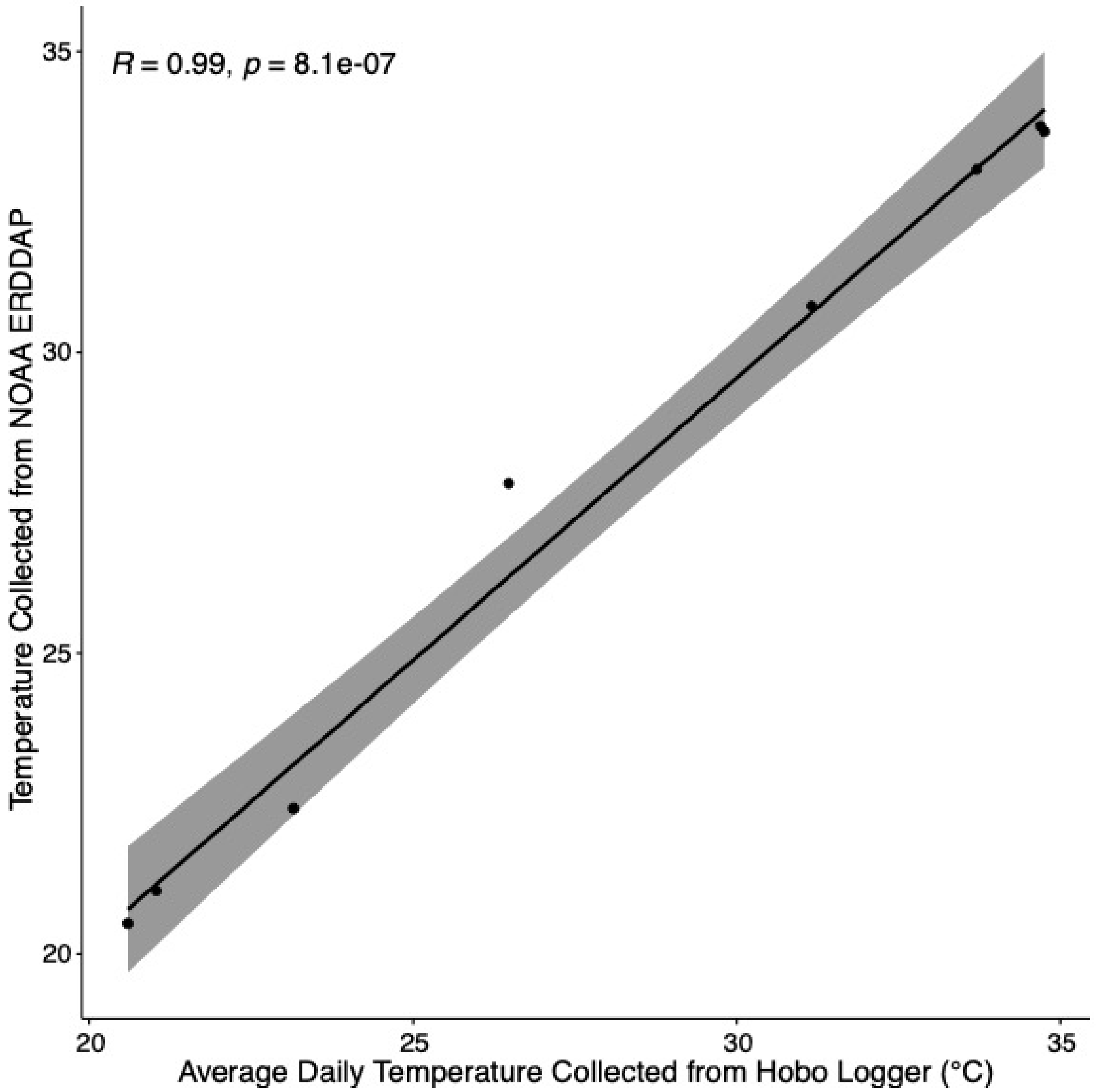


Figure S2.2: Pearson correlation between the average daily temperatures collected from Saadiyat with the Hobo Logger and the temperatures collected from the NOAA ERDDAP website on the days that urchins were sampled for the temporal series. While we have logger data for the temporal series (collected only from Saadiyat), we do not have logger data for the summer-winter spatial series. Therefore, in order to check the accuracy of the NOAA ERDDAP temperatures, we performed a Pearson correlation.

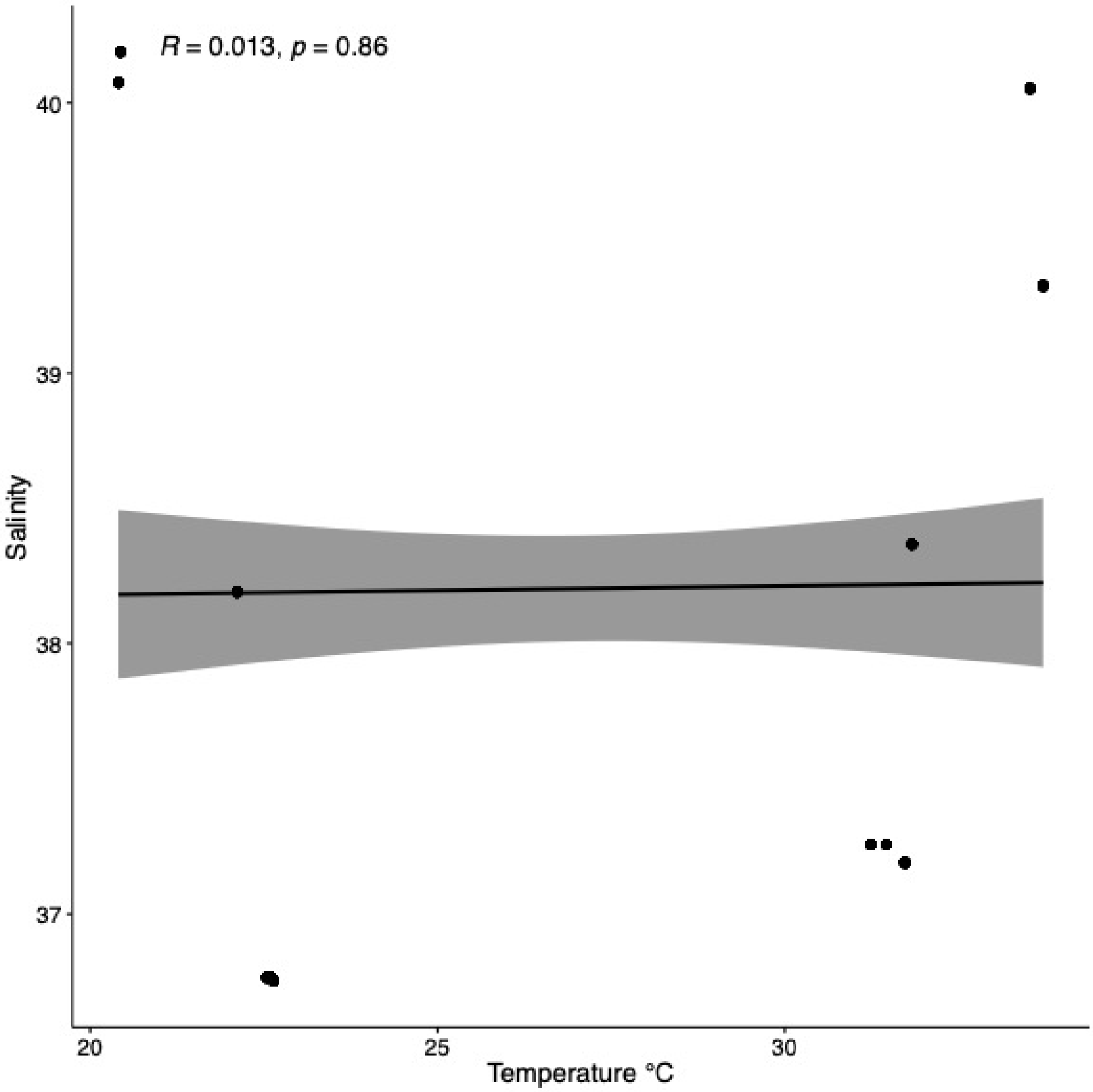


Figure S2.3: Pearson correlation between temperature and salinity for the summer-winter spatial series. Temperature data was collected from NOAA's ERDDAP website on the day of collections at 12:00:00 UTC and salinity data was obtained from a numeric ocean model: the 1/12 Global Hybrid Coordinate Ocean Model (HYCOM) at 12:00:00 UTC on the day of collections.

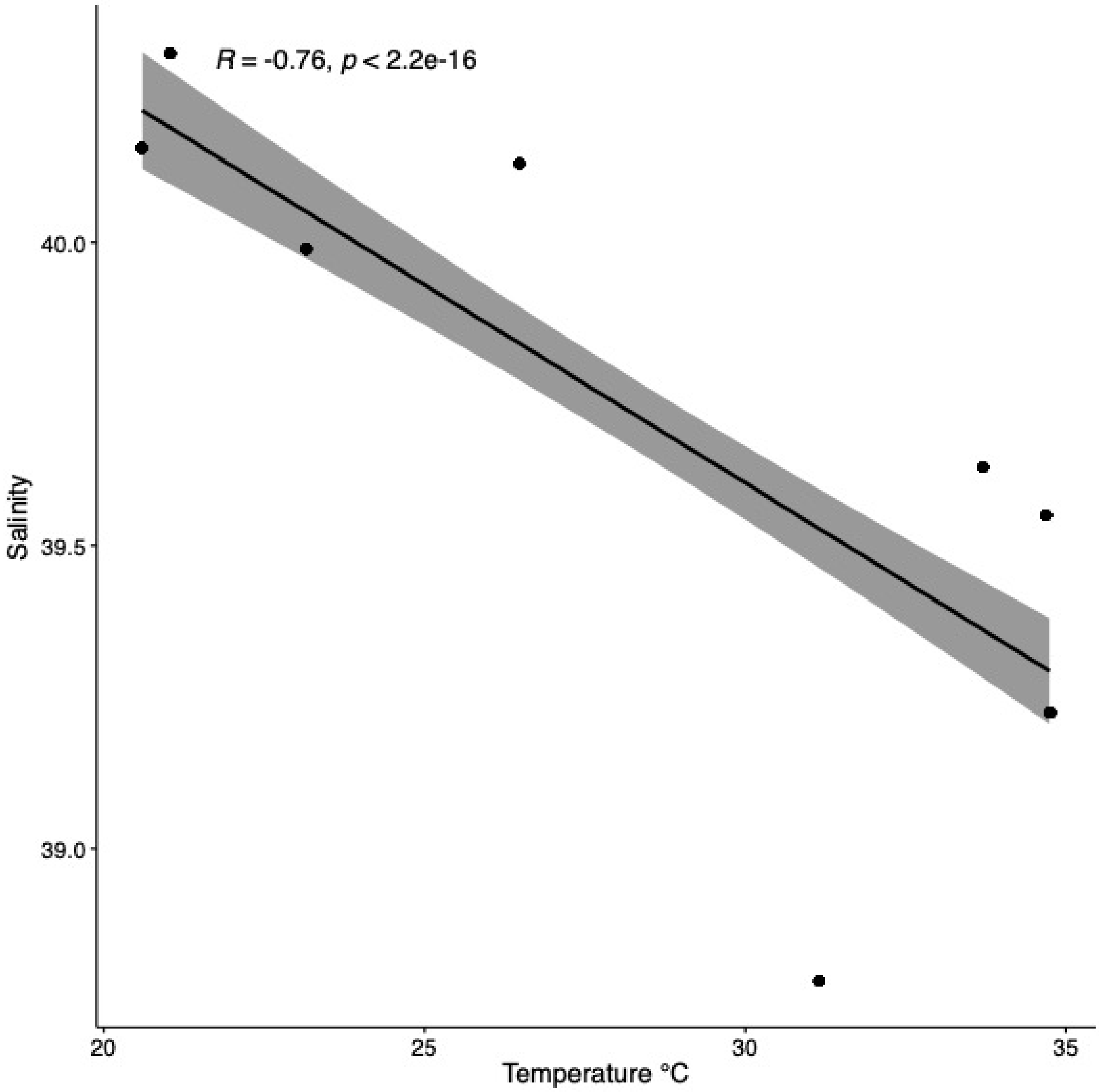
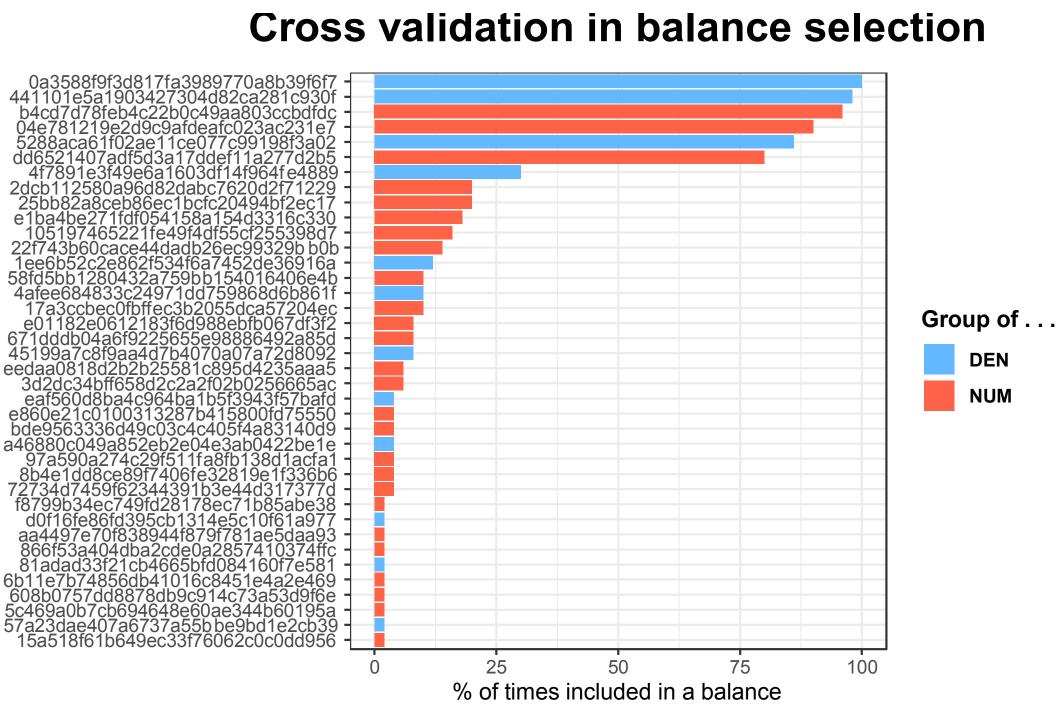
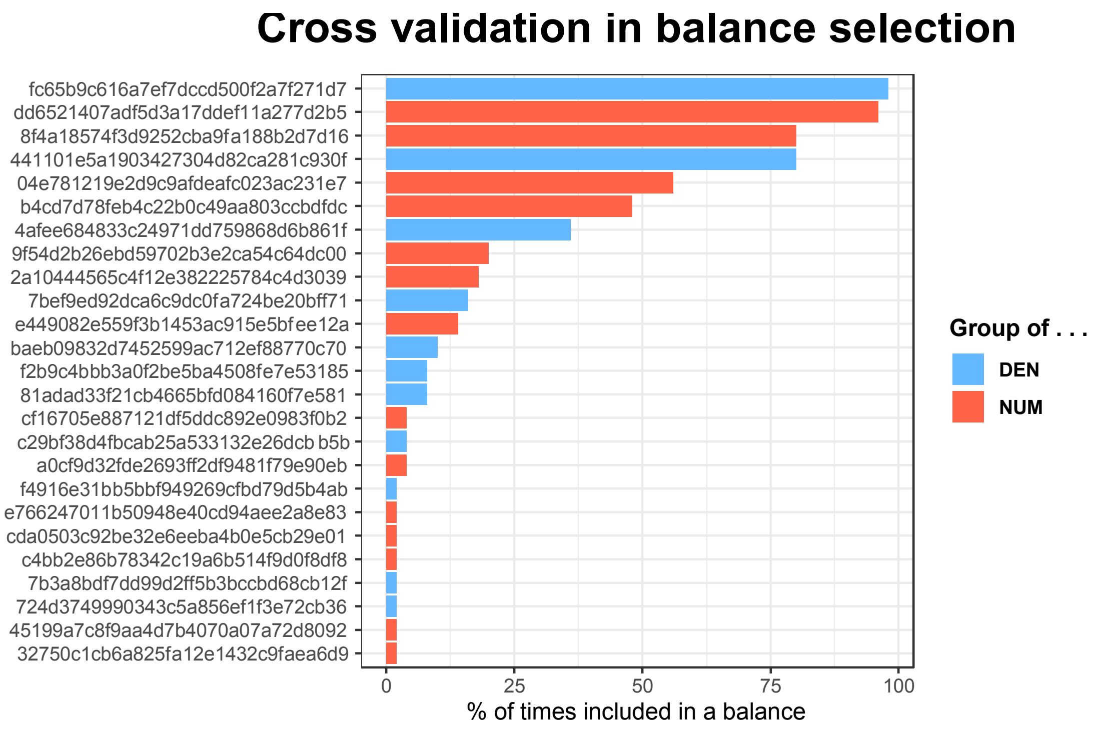


Figure S2.4: Pearson correlation between temperature and salinity for the temporal series. Temperature data was collected from the HOBO logger and salinity data was obtained from a numeric ocean model: the 1/12 Global Hybrid Coordinate Ocean Model (HYCOM) at 12:00:00 UTC on the day of collection.



	%	Global	BAL 1	BAL 2	BAL 3
0a3588f9f3d817fa3989770a8b39f6f7	100	🟦	🟦	🟦	🟦
441101e5a1903427304d82ca281c930f	98	🟦	🟦	🟦	🟦
b4cd7d78feb4c22b0c49aa803ccbdfdc	96	🟥	🟥	🟥	🟥
04e781219e2d9c9afdeafc023ac231e7	90	🟥	🟥	🟥	🟥
5288aca61f02ae11ce077c99198f3a02	86	🟦	🟦	🟦	🟦
dd6521407adf5d3a17ddef11a277d2b5	80	🟥	🟥	🟥	🟥
2dcb112580a96d82dabc7620d2f71229	20		🟥		
e1ba4be271fdf054158a154d3316c330	18		🟥		
105197465221fe49f4df55cf255398d7	16	🟥		🟥	
22f743b60cace44dadb26ec99329bb0b	14				🟥
17a3ccbec0fbffec3b2055dca57204ec	10				🟥
45199a7c8f9aa4d7b4070a07a72d8092	8	🟦		🟦	
FREQ	–	–	0.1	0.08	0.06

Figure S2.5: *Selbal* output for the summer-winter spatial series. The top figure shows ASV IDs, the color represents if the variables are included as numerators (red) or denominators (blue), and the x axis represents the percentage of times that ASV is included in the balances. The bottom figure is a summary of the cross validation procedure where rows represent the ASVs included in either the global balance and the three most frequent balances in the cross validation procedure. The second column is the percentage of times each ASV has appeared in the cross validation procedure and the last row is the proportion of times the most repeated balances have appeared.



	%	Global	BAL 1	BAL 2	BAL 3
fc65b9c616a7ef7dccd500f2a7f271d7	98	🟦	🟦	🟦	🟦
dd6521407adf5d3a17ddef11a277d2b5	96	🟥	🟥	🟥	🟥
441101e5a1903427304d82ca281c930f	80	🟦	🟦		🟦
8f4a18574f3d9252cba9fa188b2d7d16	80	🟥	🟥		🟥
04e781219e2d9c9afdeafc023ac231e7	56	🟥	🟥		
b4cd7d78feb4c22b0c49aa803ccbdfdc	48				🟥
4afee684833c24971dd759868d6b861f	36	🟦	🟦		
9f54d2b26ebd59702b3e2ca54c64dc00	20	🟥	🟥		
2a10444565c4f12e382225784c4d3039	18				🟥
7bef9ed92dca6c9dc0fa724be20bff71	16				🟦
FREQ	–	–	0.16	0.14	0.14

Figure S2.6: *Selbal* output for the temporal series. The top figure shoes ASV IDs, the color represents if the variables are included as numerators (red) or denominators (blue), and the x axis represents the percentage of times that ASV is included in the balances. The bottom figure is a summary of the cross validation procedure where rows represent the ASVs included in either the global balance and the three most frequent balances in the cross validation procedure. The first second column is the percentage of times each ASV has appeared in the cross validation procedure and the last row is the proportion of times the most repeated balances have appeared.

Bacterial co-occurrence networks

Supplemental Methods and Results

To explore how ASVs that were predictive of temperature from the *selbal* output fit into the context of the wider microbial network, we conducted network analysis using SpiecEasi (Sparse inverse covariance estimation for ecological association inference) v1.1.1 on the two datasets. SpiecEasi first performs a transformation for compositionality correction and then an estimation of the interaction graph from the transformed data using sparse inverse covariance selection. Rarified datasets were filtered to exclude ASVs that only occurred in 30% or less of samples. SpiecEasi was then run using the neighborhood selection framework (the MB method Meinshausen and Bühlmann (2006)) with the parameters set to: lambda.min.ratio=1e-2, nlambda=20, pulsar.params=list(thresh=0.05, rep.num=999). Regression coefficients were extracted and the negative edge weights (which indicated inverse trends between ASVs) were excluded. The regression coefficients were used as edge weights to generate the bacterial co-occurrence network using igraph (Csardi and Nepusz, 2006), as described in (Alfiansah et al., 2020). Louvain clustering was performed to extract network modules. Finally, five taxa with the highest probability of being keystone species were extracted based off of their hubbiness score. Hubs are nodes in the network that have a high number of links compared to other nodes in the network.

For the summer-winter spatial series, 218 ASVs were retained after filtering for low sample coverage. These ASVs were comprised predominantly of Bacteroidales, Clostridiales, Desulfobacterales, Alteromonadales, and Vibrionales. Louvain clustering generated 20 bacterial co-occurrence modules. ASVs with the highest hubbiness score (and therefore the most likely to be keystone microbes) were ASV3689, ASV11638, ASV8807, ASV3638, ASV5005 which corresponded to microbes from the order Bacteroidales, an unclassified Bacteria, Desulfobacterales, Kiritimatiellales, and Bacteroidales, respectively. For the temporal series, 307 ASVs were retained after filtering for low sample coverage and these consisted predominantly of Bacteroidales, Clostridiales, Cytophagales, Desulfobacterales, Rhodobacterales, Alteromonadales, and Vibrionales. Louvain clustering generated 22 bacterial co-occurrence modules. ASVs with the highest hubbiness score were ASV10996, ASV11472, ASV1679, ASV9510, and ASV4477. The first four of these ASVs were Vibrionales and the last was Clostridiales.

For both datasets, the ASVs output by *selbal* were retained in this analysis but did not co-occur with each other in either of the two datasets. Further, they were not identified as keystone microbes. However, building co-occurrence networks is an active area of research and warrants caution when drawing conclusions about biological interactions because high false discovery rates are common across large multivariate datasets (Knight et al., 2018). All of the microbes described in this method are common within sea urchin gut communities (Hakim et al., 2016b;Schwob et al., 2020) so it is challenging to place the *selbal* ASVs within groups of microbes belonging to opportunistic taxa or pathogenic taxa, for example. In order to move beyond correlational analysis and address causation (i.e., exact function of the microbes), functional assays are warranted.

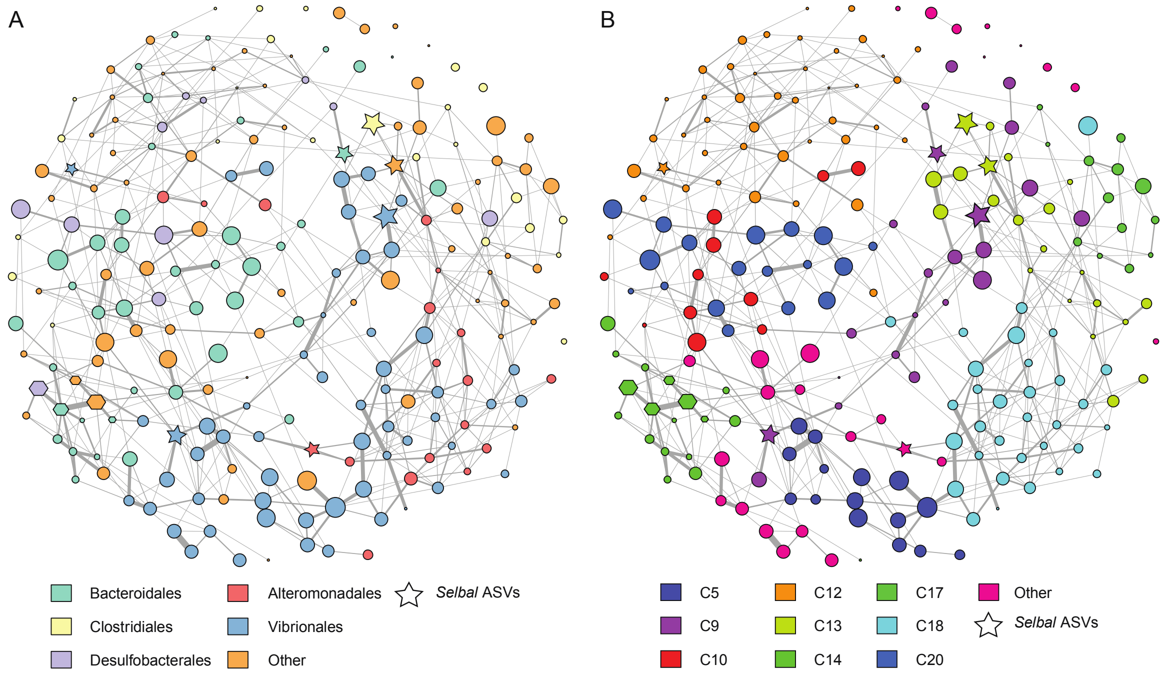


Figure S2.7: Bacterial co-occurrence networks generated by SpiecEasi for the summer-winter spatial series. Node size corresponds to the average sequence proportion of each ASV and the edge width corresponds to the strength of the association between ASVs. ASVs identified through *selbal* as predictive of temperature are denoted with a star. Keystone ASVs (determined through their hubbiness score) are denoted with a polygon. **A.** Bacterial co-occurrence network with the node color indicating ASV taxonomy (Norderhaug et al.). B. Network modules detected by Louvain clustering are shown in different colors and modules that contained less than 10 ASVs were grouped into "Other" for better visualization.

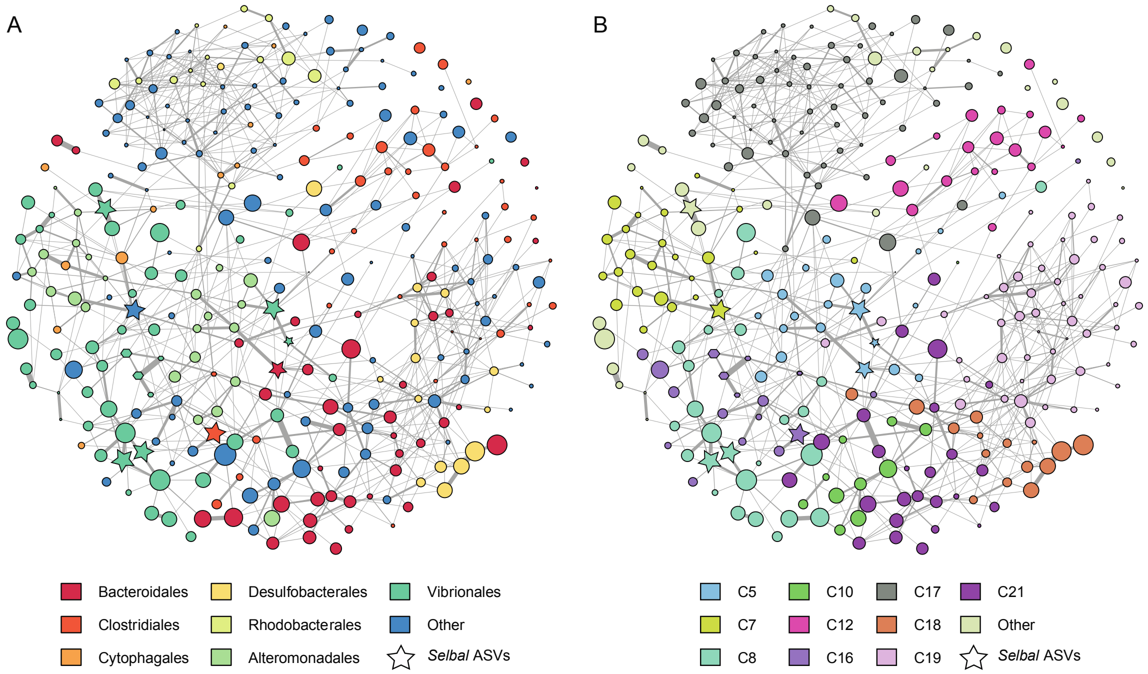


Figure S2.8: Bacterial co-occurrence networks generated by SpiecEasi for the temporal series. Node size corresponds to the average sequence proportion of each ASV and the edge width corresponds to the strength of the association between ASVs. ASVs identified through *selbal* as predictive of temperature are denoted with a star. Keystone ASVs (determined through their hubbiness score) are denoted with a polygon. **A.** Bacterial co-occurrence network with the node color indicating ASV taxonomy (Norderhaug et al.). B. Network modules detected by Louvain clustering are shown in different colors and modules that contained less than 10 ASVs were grouped into “Other” for better visualization.

PICRUSt2

Supplemental Methods and Results

Functional predictions of the bacterial communities were determined through the Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt v2.0.0). Rarified datasets were filtered to remove features that only occurred in 10% or less of samples. For the temporal series, we further filtered the dataset to remove samples from thermally intermediate months (May and November) so that we could compare the functional profiles of cooler months (January, February, and March) to warmer months (July, August, and September). PICRUSt2 was then run using the picrust2_pipeline.py script which incorporates sequence placement, hidden-state prediction of genomes, metagenome prediction, and pathway-level pathways. Any ASV with a NSTI value higher than 2 was classified as uncharacterized phyla and was discarded. Finally, we identified the predicted microbial metabolic pathway using the MetaCyc database (Caspi et al., 2018) with the add_descriptions.py script. Differentially abundant ASVs were identified using ANOVA-like differential expression (ALDEx2, Gloor (2015)) analysis with a Wilcoxon test. *P*-values were adjusted with the Benjamini-Hochberg FDR multiple-test correction. Comparisons with a p-value < 0.05 were retained. For the summer-winter spatial series we compared samples collected in August to those collected in February.

For the summer-winter spatial series, 164 out of 10,434 ASVs were removed because they had a NSTI value higher than two and PICRUSt2 predicted a total of 342 MetaCyc pathways. ALDEx2 analysis found 58 significantly differentially abundant pathways for the summer-winter spatial series. For samples collected in August, there was an enrichment of MetaCyc pathways related to bacterial cell wall biosynthesis, aromatic compound degradation, biotin biosynthesis, fermentation, and methanogenesis. For samples collected in February, there was an enrichment in pathways related to sucrose biosynthesis, nucleotide degradation, fermentation, and aerobactin biosynthesis. Notably, two of the pathways (PWY-5647 and PWY-6210) that were enriched in February are implicated in the degradation of compounds related to pesticides, plasticizers, dyes and herbicides. This could be a result of run-off as January through March is the wettest time of year in the UAE.

For the temporal series, 78 out of 8,581 ASVs were removed and PICRUSt2 predicted 365 MetaCyc pathways. Of these 365 pathways, 57 were significantly differentially abundant. ALDEx2 revealed an enrichment of MetaCyc pathways related to LPS synthesis, synthesis of lipophilic components of the cytoplasmic membrane, metabolic pathways, bacterial cell wall biosynthesis, and polyamine biosynthesis for the samples collected in July/August/September. For the samples collected in January/February/March, pathways related to sucrose biosynthesis and amino acid, aromatic compound, and nucleotide degradation were enriched. Interestingly, both datasets showed enrichment for pathways related to cell wall machinery in the warmer sampling months. Further, both datasets showed an enrichment in sucrose biosynthesis and degradation-related pathways for samples collected in the cooler months. This suggests that the microbial communities from both datasets experience similar shifts in the predicted functional profile of the microbiome in response to temperature and potentially other seasonal variation (e.g., rainfall).

Although methods for predicting functional profiles of microbial communities based on taxonomic composition can sometimes be informative and are a promising avenue for future research, PICRUSt2 analysis should be evaluated with caution. Predictions generated by PICRUSt are limited by the currently available genomes and the predictions are highly biased towards taxa linked to human health (Sun et al., 2020). Further, the downstream analyses required are still heavily debated and generally rely on tests which are inappropriate for compositional datasets (Gloor et al., 2017;Lin and Peddada, 2020). Despite these limitations, temperature variation can lead to distinct functional compositions of the microbiome and by comparing PICRUSt2 results generated from our two independent datasets we highlight predicted metabolomic changes.

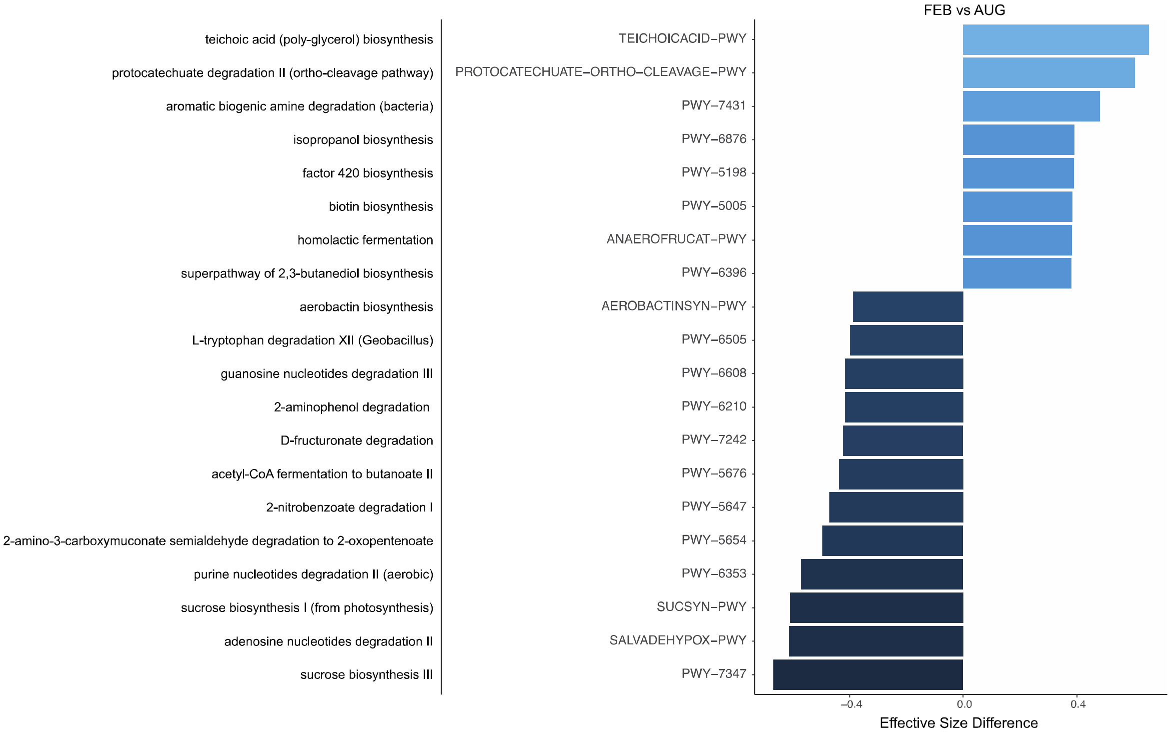


Figure S2.9: Function differentiation between urchin gut microbiota collected in February versus August (summer-winter spatial series) determined using PICRUSt v2.0.0. Details of the top 20 MetaCyc pathways significantly enriched between the two months (*p* -value < 0.05 after a Benjamini-Hochberg FDR correction).

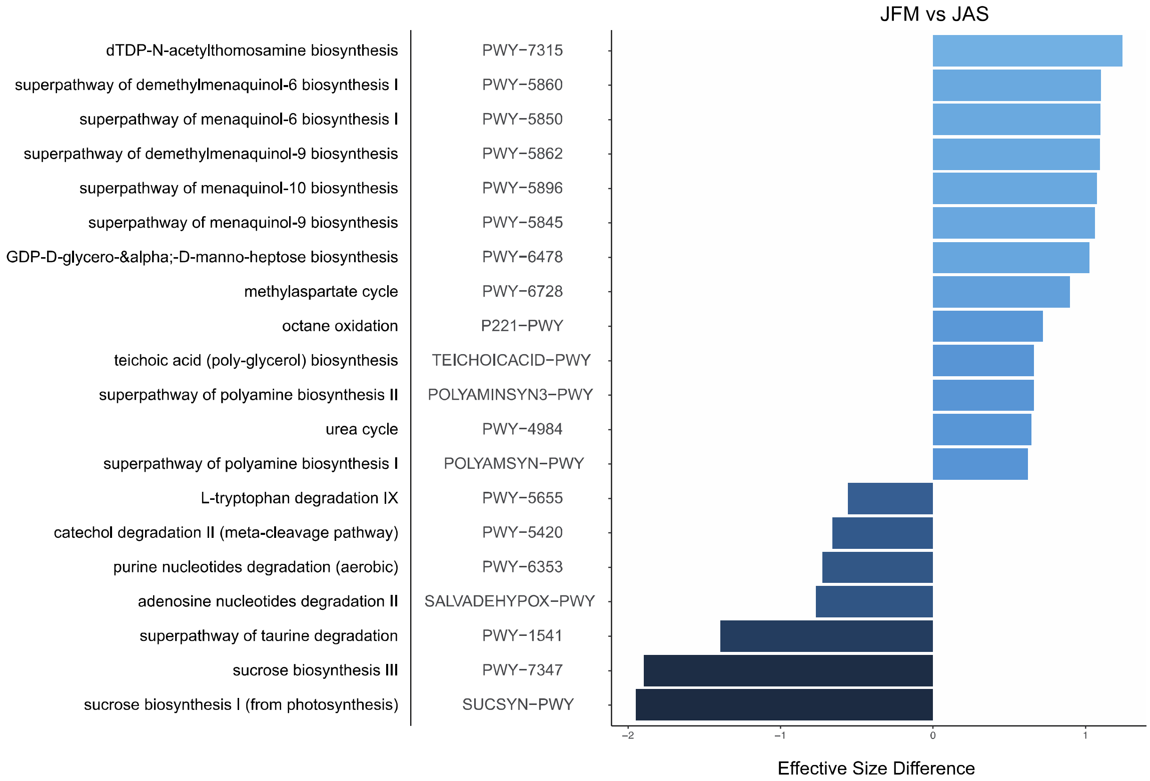


Figure S2.10: Function differentiation between urchin gut microbiota collected in cool months (January, February, March) and warm months (July, August, September) for the temporal series. Details of the top 20 MetaCyc pathways significantly enriched between the two months (*p* -value < 0.05 after a Benjamini-Hochberg FDR correction).

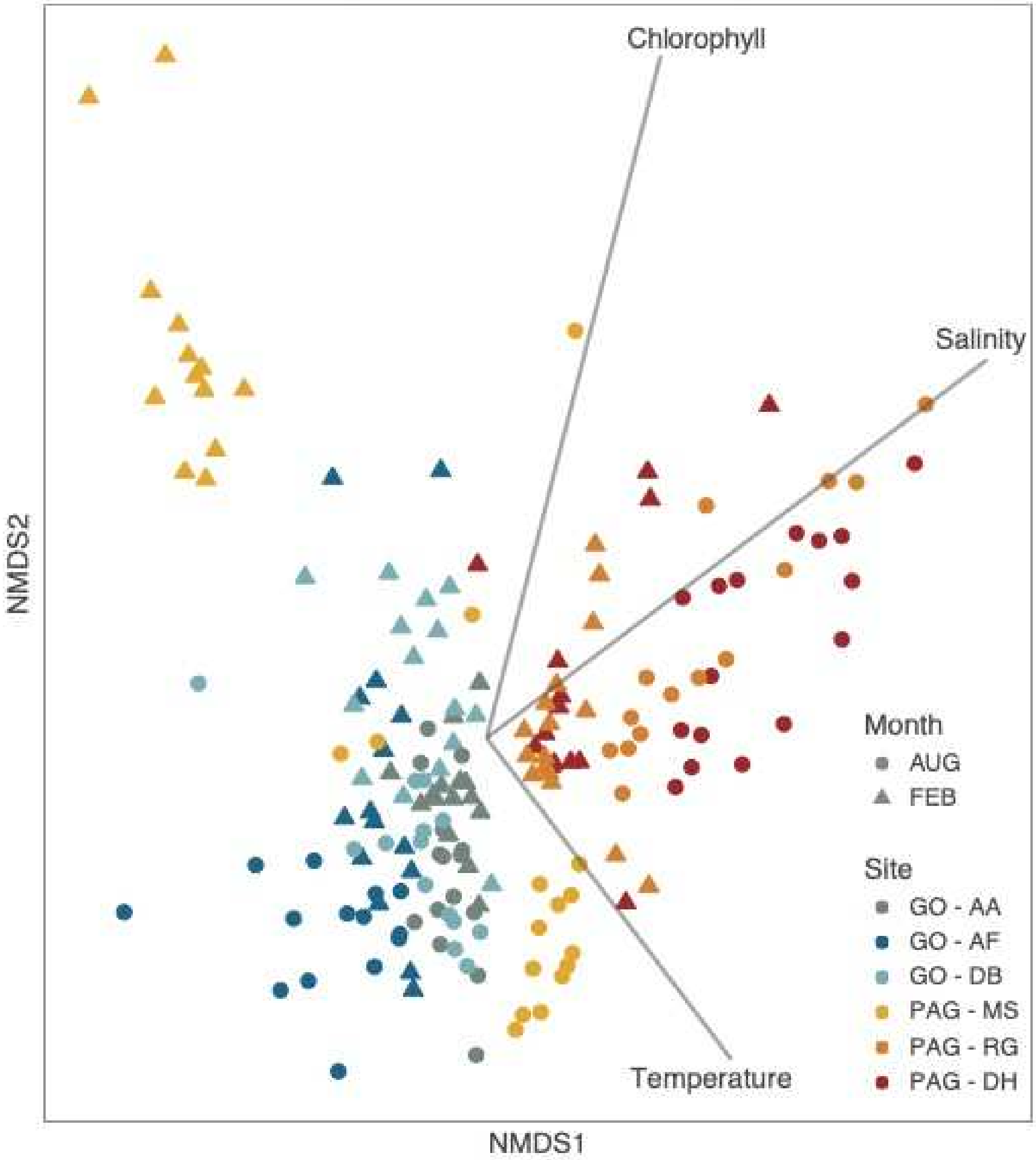


Figure S2.11: Ordination was performed using NMDS based on Bray-Curtis dissimilarity for the summer-winter spatial series. Temperature data was collected from NOAA's ERDDAP website on the day of collections at 12:00:00 UTC, chlorophyll concentration was obtained from MODIS AQUA level 3 monthly averaged data (4 km), and salinity data was obtained from a numeric ocean model: the 1/12 Global Hybrid Coordinate Ocean Model (HYCOM) at 12:00:00 UTC on the day of collections. The three environmental parameters are displayed as arrows, with the length proportional to the correlation between the variable and the ordination. Temperature explained 6.3% of the dissimilarities in the microbiota (PERMANOVA, *p*-value<0.001), while chlorophyll concentration and salinity accounted for 11.1% and 11.4% (PERMANOVA, *p*-value<0.001), respectively (Table S2.3 and S2.4, Supporting Information).

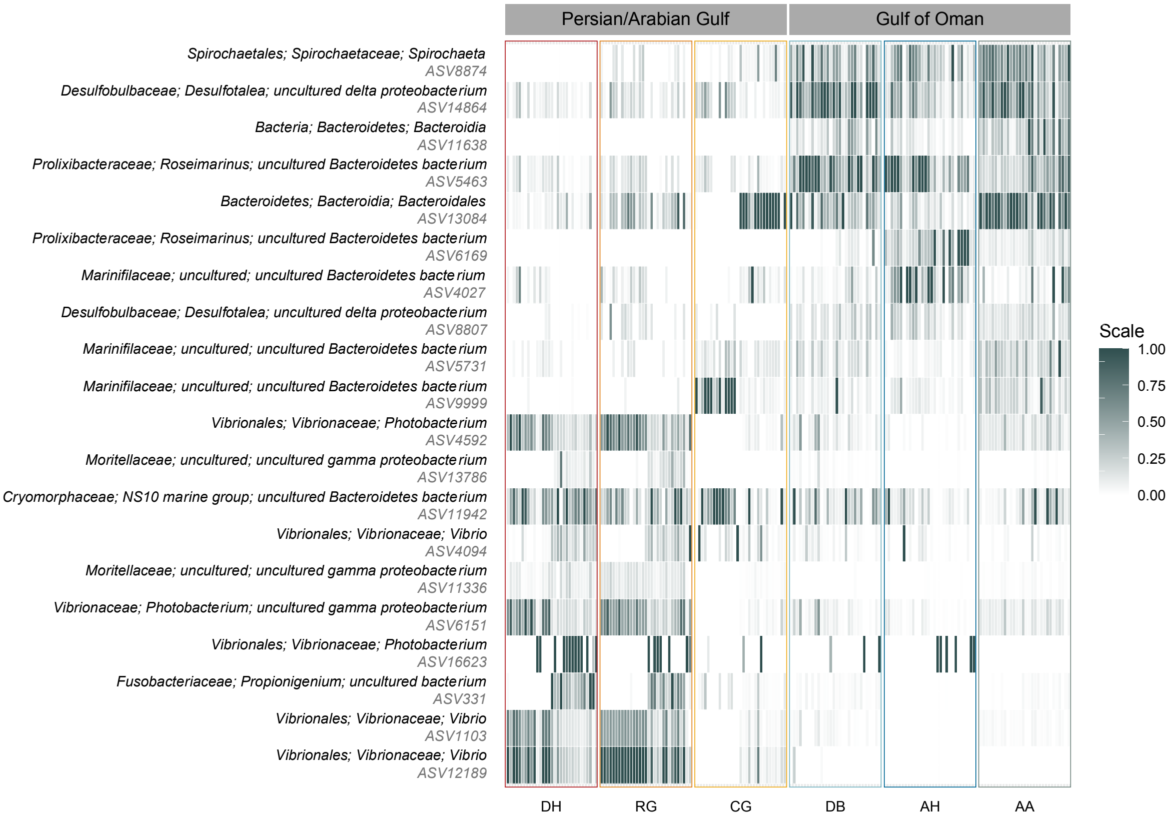


Figure S2.12: Heat map of taxa that are associated with differentiation across PC 1, which separates the PAG from the GO. Results were generated in DEICODE.

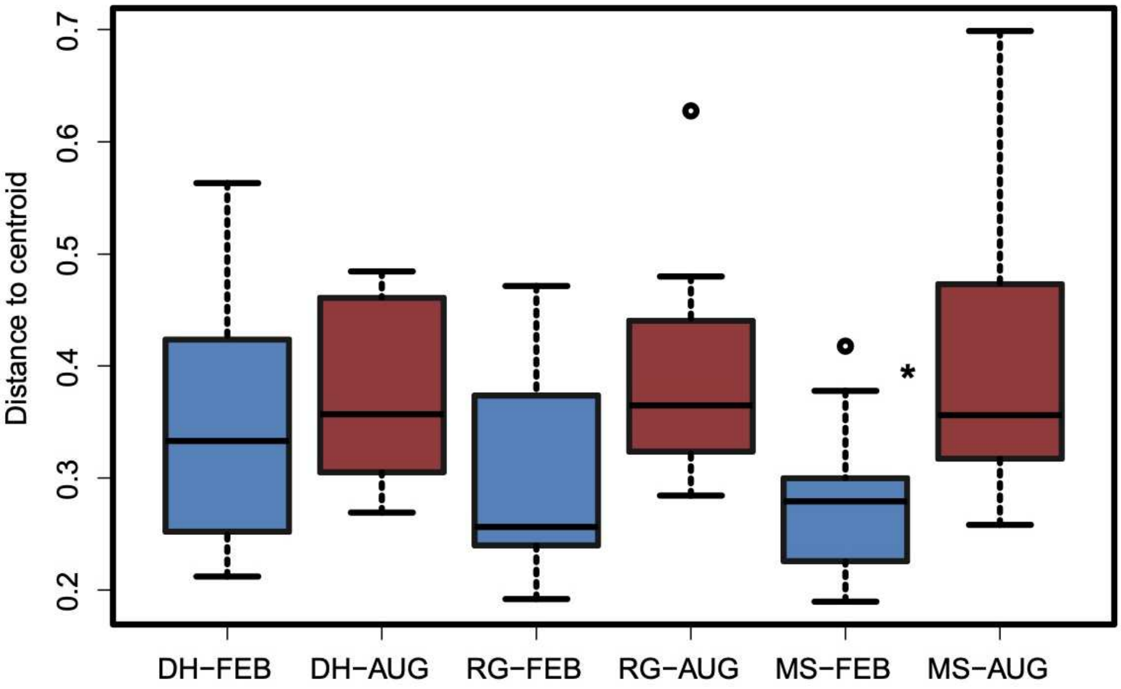


Figure S2.13: Box plot representation of the dispersion values for the summer-winter spatial series comparing February (blue) to August (red) for each respective site. Significance was tested between months within sites (i.e., we are not making comparisons *between* collection sites) and denoted with an asterisk. Data was obtained by the PERMADISP function.

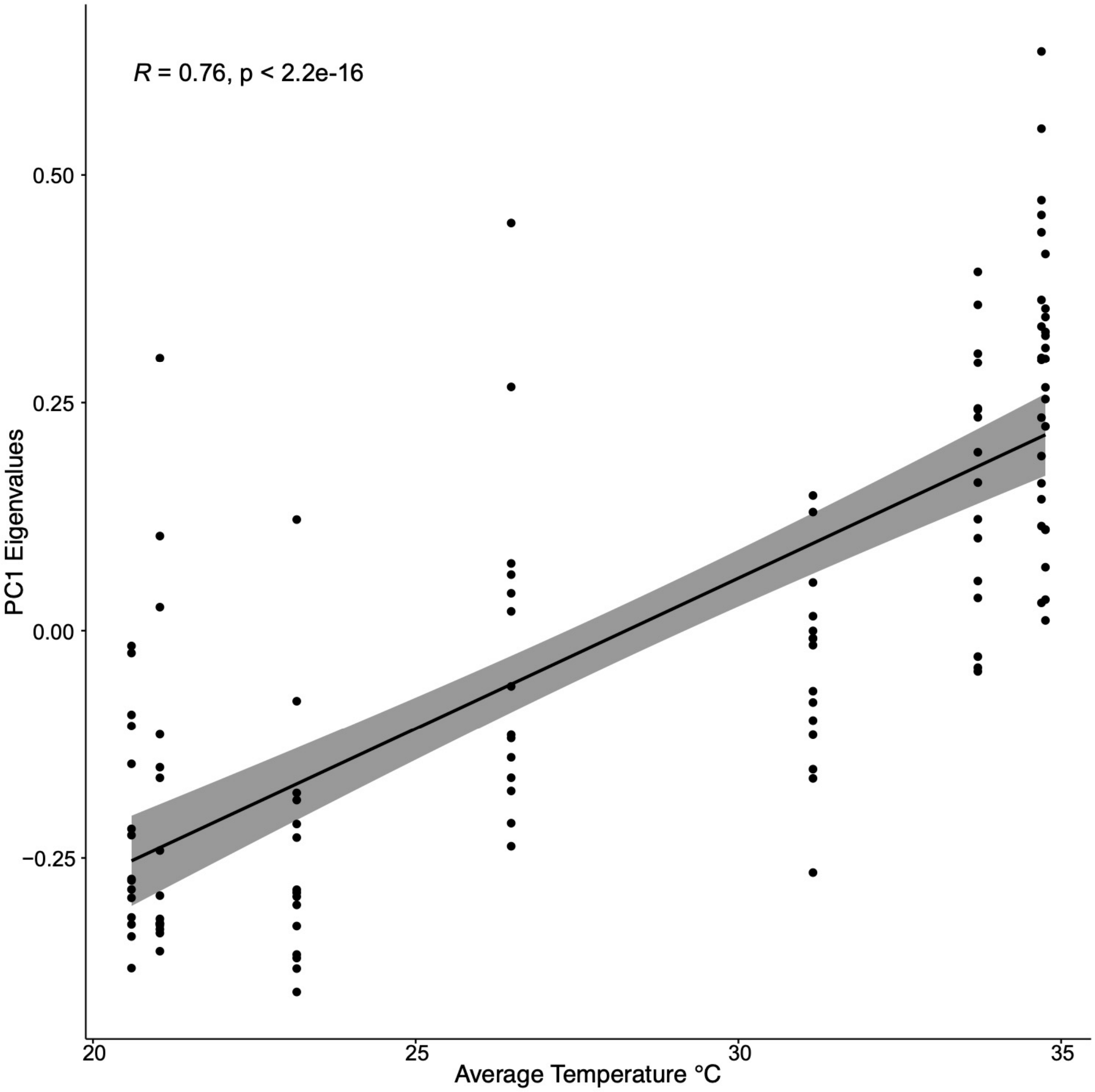


Figure S2.14: Pearson correlation between average temperature on the day of collection from Saadiyat reef and PC 1 eigenvalues for the temporal series.

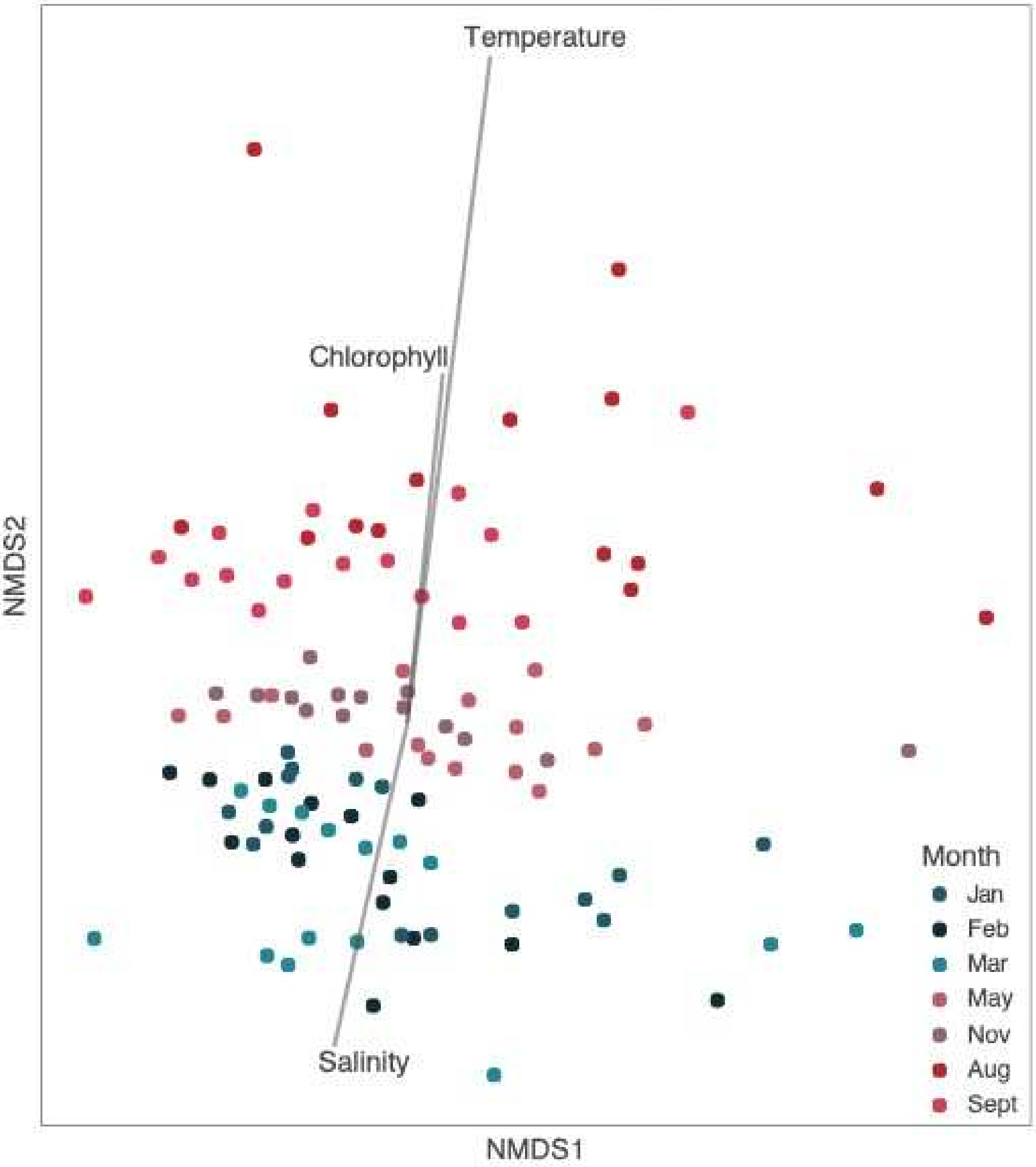


Figure S2.15: Ordination was performed using NMDS based on Bray-Curtis dissimilarity for the temporal series. Temperature data was collected from the HOBO logger, chlorophyll concentration was obtained from MODIS AQUA level 3 monthly averaged data (4 km), and salinity data was obtained from a numeric ocean model: the 1/12 Global Hybrid Coordinate Ocean Model (HYCOM) at 12:00:00 UTC on the day of collection. The month of July was removed for this analysis as there was no chlorophyll data available. The three environmental parameters are displayed as arrows, with the length proportional to the correlation between the variable and the ordination. Temperature explained 7.1% of the dissimilarities in the microbiota (PERMANOVA, *p*-value<0.001), while chlorophyll concentration and salinity accounted for 1.9% and 1.7% (PERMANOVA, *p*-value<0.001), respectively (Table S2.6 and S2.7, Supporting Information).

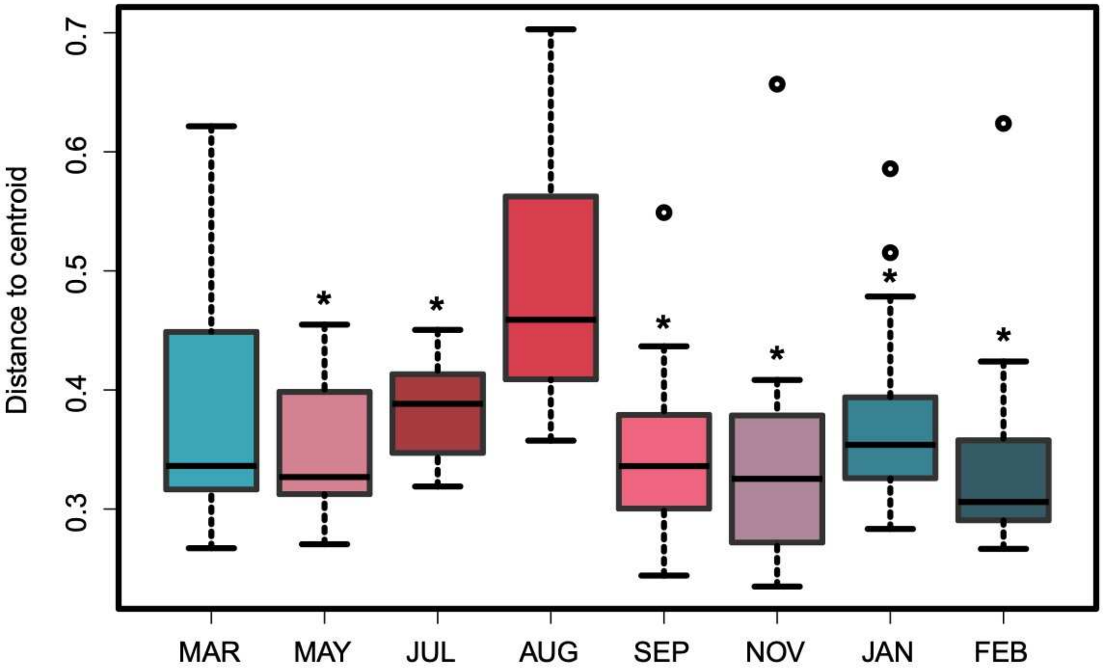


Figure S2.16: Box plot representation of the dispersion values for the temporal series by the PERMADISP function. Significance was tested for all combinations of months and only comparisons between August were significant. Therefore, all significant comparisons of dispersion between August and the other months are denoted with an asterisk.

CHAPTER 3

POPULATION GENOMIC ANALYSES OF THE SEA URCHIN *ECHINOMETRA* SP. *EZ* ACROSS AN EXTREME ENVIRONMENTAL GRADIENT

Remi N. Ketchum, Edward G. Smith, Melissa B. DeBiasse, Grace O. Vaughan, Dain McParland, Whitney B. Leach, Noura Al-Mansoori, Joseph F. Ryan, John A. Burt, Adam M. Reitzel

Citation

Ketchum, Remi N., et al. "Population genomic analyses of the sea urchin *Echinometra* sp. *EZ* across an extreme environmental gradient." *Genome Biology and Evolution* 12.10 (2020): 1819-1829.

Abstract

Extreme environmental gradients represent excellent study systems to better understand the variables that mediate patterns of genomic variation between populations. They also allow for more accurate predictions of how future environmental change might affect marine species. The Persian/Arabian Gulf is extreme in both temperature and salinity while the adjacent Gulf of Oman has conditions more typical of tropical oceans. The sea urchin *Echinometra* sp. *EZ* inhabits both of these seas and plays a critical role in coral reef health as a grazer and bioeroder, but, to date, there have been no population genomic studies on this or any urchin species in this unique region. *E*. sp. *EZ*'s life history traits (e.g., large population sizes, large reproductive clutches, and long life spans), in theory, should homogenize populations unless non-neutral processes are occurring. Here, we generated a draft genome and a restriction-site associated DNA sequencing dataset from seven populations along an environmental gradient across the

Persian/Arabian Gulf and the Gulf of Oman. The estimated genome size of *E*. sp. *EZ* was 609 Mb and the heterozygosity was amongst the highest recorded for an echinoderm at 4.5%. We recovered 918 high quality SNPs from 85 individuals which we then used in downstream analyses. Population structure analyses revealed a high degree of admixture between all sites, although there was population differentiation and significant pairwise F_{ST} values between the two seas. Preliminary results suggest migration is bidirectional between the seas and nine candidate loci were identified as being under putative natural selection, including one collagen gene. This study is the first to investigate the population genomics of a sea urchin from this extreme environmental gradient and is an important contribution to our understanding of the complex spatial patterns that drive genomic divergence.

Introduction

The mechanisms governing evolutionary divergence are not well understood in marine systems, where clear barriers to gene flow are uncommon and organisms generally exhibit high dispersal capabilities (Palumbi et al., 1997;DeFaveri et al., 2013;Kelley et al., 2016;Oleksiak, 2016;Takeuchi et al., 2019). Indeed, this propensity for dispersal coupled with large population sizes has been predicted to homogenize genetic variation and dampen the effects of genetic drift, respectively (Lande, 1980;Waples, 1998;Xuereb et al., 2018). Conversely, an increasing number of studies have shown signatures of genetic differentiation across small geographic scales as a result of oceanographic currents (Lal et al., 2017), organismal behavior that may favor local retention (Miller et al., 2001), and local adaptation due to environmental heterogeneity

(Gleason and Burton, 2016). Species with ranges spanning environmentally heterogenous ecosystems are highly informative study systems for advancing our understanding of complex population dynamics, as well as assessing the capacity of organisms for adaptation to changing environments (Reitzel et al., 2013;Gleason and Burton, 2016).

The Persian/Arabian Gulf (hereafter PAG) is an example of an extreme marine environment, which is separated from the Gulf of Oman and the wider Indian Ocean by the narrow (42 km) Strait of Hormuz (Burt et al., 2019). The PAG has the world's warmest sea with daily mean summer temperatures regularly >35 °C and extremes exceeding 37 °C (Smith et al., 2017c;Burt et al., 2019). These conditions surpass climate change predictions for the Indo-Pacific in the next century (Hoegh-Guldberg et al., 2014). The neighboring Gulf of Oman (hereafter GO) experiences much lower summer temperatures which are typically <32 °C (Coles, 2003). In addition to extreme thermal conditions, the PAG also experiences higher salinity than the GO (40-42 PSU vs. 37 PSU, respectively) (Burt et al., 2008;Bauman et al., 2013). To date, several studies have shown population structuring between the two seas, including in the brain coral, *Platygyra daedalea* (Howells et al., 2016;Smith et al., 2017b), the sea urchin *Diadema setosum* (Lessios et al., 2001), and several species of fishes (Hoolihan et al., 2004;Giles et al., 2014;Torquato et al., 2019). The majority of studies in this region focused on corals and fishes (Vaughan and Burt, 2016) and the only work on sea urchins used short mitochondrial sequences. To better understand these diverse ecosystems, it is crucial to examine species with different life history traits (i.e., population size, dispersal patterns, mating systems, philopatry, and reproductive timing) as these traits may mediate the effects of gene flow and genetic drift (Whiteley et al., 2004).

Sea urchins are critical ecosystem engineers around the world, particularly in shallow coastal habitats, where their grazing plays an important role in bioerosion and algal control (Downing and El-Zahr, 1987;McClanahan and Muthiga, 2007). Due to their importance, urchins have been used as study systems to determine the potential for adaptation to stressful environmental conditions (Pespeni et al., 2011;Pespeni et al., 2012;Kelly et al., 2013;Pespeni and Palumbi, 2013). To date, there have been no population genomic studies of any species of sea urchin within the entire northeastern Arabian region (although there have been population genetic studies based on mitochondrial regions, see: Bronstein and Loya, 2013 and Lessios, et al., 2001). This represents a critical knowledge gap as sea urchins are highly abundant in the PAG (densities averaging 8.6m^{-2} across eight sites between 2015-2019, [Burt, JA, unpublished data]) and they play a significant role in the health and dynamics of coral reef ecosystems in the region as major bioeroders (Downing and El-Zahr, 1987). The most abundant sea urchin in the PAG is *Echinometra* sp. *EZ,* previously thought to be *Echinometra mathaei* (Ketchum et al., 2018a). *Echinometra* are common in shallow water (1-3 meters depth) and have been found in waters up to 20m deep (McClanahan and Muthiga, 2007). The seasonal reproductive patterns of regional *Echinometra* sea urchins are not yet well understood; however, one study in the northern PAG showed peak spawning in June (Alsaffar and Lone, 2000). The larvae feed on phytoplankton and although the pelagic larval duration (PLD) for this species is unknown, congeners have PLDs of a few weeks [e.g., *E. vanbrunti* (18 days) and *E. viridis* (30 days) (McClanahan and Muthiga, 2007)]. On a large spatial scale, these larvae behave as passive particles and their transport is governed by oceanographic current patterns (Rahman et al., 2014). Taken together, these traits make *E*. sp. *EZ* an excellent study organism to better understand how different life history strategies can drive molecular evolution, resulting in different patterns of population divergence in the PAG.

In this study, we used restriction-site associated DNA sequencing (RAD-seq) to characterize patterns of genetic diversity and population structure of *E.* sp. *EZ.* We collected samples from seven sites spanning >500 km from the southern PAG into the western GO. We generated a draft genome to use as a reference for our RAD-seq analyses. We performed an initial outlier analysis to identify SNPs under potential selection and characterized historical gene flow to understand migration patterns. This study contributes to our understanding of genetic differentiation in marine invertebrates in environmentally divergent habitats and how this may pertain to a changing climate.

Materials and Methods

Draft genome assembly

Sample collection

A gonadal tissue sample from a single *E*. sp. *EZ* adult from Dhabiya reef in the southern PAG (24°21’55.8”N 54°06’02.9”E) was collected, preserved in RNA*later,* and subsequently stored at -20 °C.

DNA extraction and sequencing

Total genomic DNA was extracted from the gonadal tissue sample using the DNeasy Blood and Tissue Kit (Qiagen). DNA quality was visualized on an agarose gel and concentration was determined with a 2000 Nanodrop spectrophotometer (ThermoFisher Scientific, Waltham, MA). High-quality DNA was submitted for PCR-free library preparation and whole genome sequencing on one lane on an Illumina HiSeq3000 (150 bp paired-end reads) and one lane on a NextSeq500 (150 bp paired-end reads) at the University of Florida Interdisciplinary Center for Biotechnology Research.

DNA read processing and genome assembly

Approximately 302 million paired end (PE) reads were obtained from the HiSeq and NextSeq sequencing. We performed adaptor trimming and quality filtering using Trimmomatic v0.36 (Bolger et al., 2014) with a phred quality score of 33. Leading and trailing bases with a quality score below three were removed, a 4-base wide sliding window was used to cut where the average quality per base dropped below 15, and reads that were less than 36 bp long were removed. This was followed by error correction with Allpaths-LG version v44,837 (Gnerre et al., 2011). To estimate genome size, we generated a frequency histogram for a *k*-mer length of 21 using Jellyfish v2.2.6 (Marçais and Kingsford, 2011). This histogram was then analyzed using GenomeScope to obtain estimates for genome size, as well as heterozygosity and duplication levels (Vurture et al., 2017). Mitochondrial reads were removed using FastqSifter v1.1.1 (Ryan, 2015a) with the *E. mathaei* mitochondrial genome as a reference (GenBank Accession Number: NC034767.1). We performed *de novo* genome assemblies using Platanus v.1.2.4 (Kajitani et al., 2014) with default parameters and *k*-mer lengths ranging from 45-99. A custom Perl script, plat.pl (Ohdera and Ryan, 2018), was used to invoke the Platanus commands for assembly, scaffolding, and gap closing. We then used the sub-optimal assemblies (*k*-mer = 45, 64, 99) to construct artificial mate-pair libraries for five insert sizes (2,000, 3,000, 5,000, 7,000, 10,000) with MateMaker v1.0 (Ryan, 2015b). SSPACE Standard v3.0 (Boetzer et al., 2010) was subsequently used to scaffold the optimal assembly (generated using *k*-mer= 85) using the previously generated artificial mate-pair libraries. We removed contigs smaller than 200 bp for our RAD-seq analysis (however, the uploaded final assembly, available on Dryad (https://doi.org/10.5061/dryad.c59zw3r40, still contains these reads). The commands and parameters used for the genome assembly are available in a github repository (Ketchum, 2020). We checked completeness of the genome using CEGMA v2.5 (Parra et al., 2007) and BUSCO v2.01 (Simão et al., 2015) through the gVolante web server (Nishimura et al., 2017).

Restriction-site associated DNA sequencing and data processing

Sample collection

Ten to 15 *E.* sp. *EZ* individuals were collected between 2017 and 2018 from seven sites along the northeastern Arabian Peninsula, for a total of 94 samples (Figure 3.1; Supplemental Table 3.1). Four of the sites were in the environmentally extreme PAG and three of the sites were in the GO. Gonadal tissue samples were preserved in RNA*later,* and subsequently stored at -20 °C.

DNA extraction and sequencing

DNA was extracted using the DNeasy Blood and Tissue Kit (Qiagen) following the manufacturer’s protocol for DNA extraction from tissues. DNA concentrations were normalized to 1ng/ul for a total of 50ng per reaction. Library preparation and sequencing of RAD markers was performed by Floragenex Inc. (Eugene, Oregon) using the restriction enzyme *sbfI* and Illumina 100 bp single-end sequencing.

Read processing

The dDocent pipeline (Puritz et al., 2014) was used for read mapping, SNP calling, and SNP filtering. First, raw reads were demultiplexed into separate files according to individual indices and quality filtered using “process_radtags” within *Stacks* v1.46 (flags: -e *sbfI* -c -q -r), which removes any reads with an uncalled base and discards reads with low quality scores (Catchen et al., 2013). RADtags were aligned to the draft genome using BWA-mem v0.7.17 (Li and Durbin, 2009;Li, 2013) with default parameters. Samtools v1.7 (Li et al., 2009) was used to sort and filter out any alignments that had a mapping quality <30 and FreeBayes v1.1.0 (Garrison and Marth, 2012) was used to call SNPs using default parameters.

SNP filtering

Preliminary filtering of variants was performed with VCFtools v0.1.14 (Danecek et al., 2011). We used the following parameters in the dDocent SNP filtering pipeline: 1) quality score ≥ 30; 2) minimum depth for a genotype call ≥ 3; 3) individuals with ≥ 50% missingness were removed; 4) a genotype call rate of 95% was applied across all individuals; 5) minimum mean depth ≥ 20; 6) population specific call rate ≥ 90%; 7) minor allele frequency (Douglas et al.) ≥ 0.05; 8) removed loci with an allele balance ≤ 0.25 or ≥ 0.75; 9) removed loci above a mean depth of ≥ 475; and 10) kept SNPs that had frequencies that were not statistically different from Hardy-Weinberg equilibrium (cutoff = 0.25, alpha = 0.01, applied on a per population basis). After filtering, a total of 918 SNPs remained in 85 sea urchins (number of individuals per site: DH = 11, SA = 8, RG = 11, MS = 11, DB = 15, AF = 14, AA = 15). The commands and parameters used in these analyses are available in a github repository (Ketchum, 2020). This filtered VCF file was used for downstream analysis.

Summary statistics, population differentiation, and structure

Pairwise genetic differentiation (F_{ST}) between populations and their significance was calculated in Arlequin v3.5.2.2 (Excoffier et al., 2005) with 10,000 permutations. Genetic diversity statistics within populations, including observed heterozygosity (H_{O}), expected heterozygosity (H_{E}), and the nucleotide diversity of variable sites (*P*i) were estimated using "populations" in *Stacks* v1.46.

We used STRUCTURE v2.3.4 (Pritchard et al., 2000), which implements a Bayesian clustering algorithm and ignores geographic proximity, to estimate the most likely number of genetic clusters. The number of clusters (K) was set from 1 to 10 with 20 independent runs for each fixed number of K. Each run included a burn-in period of 100,000 iterations, followed by 100,000 iterations of the Monte Carlo Markov Chain (MCMC) algorithm. The admixture model was run with correlated allele frequencies. To identify the most probable number of groups (K) that best fit the data, we used STRUCTURE HARVESTER (Earl, 2012), which implements the Evanno method to determine the optimal value of K depending on the ΔK value. The program CLUMPP v1.1.2 (Jakobsson and Rosenberg, 2007) was used to align the 20 repetitions of the K

value with the highest likelihood. These results were then visualized as bar plots using a custom R script.

To statistically test for population structure, we used principal component analysis (PCA) within the smartpca program in EIGENSOFT v6.0.1 (Patterson et al., 2006). We then used the twstats program within EIGENSOFT to perform a formal statistical test for population structure by calculating the significance of each eigenvector with a Tracey-Widom test.

Estimating historical relationships

Treemix v1.13 (Pickrell and Pritchard, 2012) was used to understand historical patterns of gene flow between populations. Treemix leverages allele frequencies to generate a maximum likelihood tree for a set of populations and then connects branches in the tree with edges (or migration events) to explain excess covariance and improve model fit. The Al Aqah population was chosen as the outgroup because it is most geographically distant from the PAG. The PAG is roughly 14,000 years old (Lambeck, 1996) and we therefore hypothesized that the populations in the GO are ancestral to those in the PAG. Consideration should be taken when interpreting these results as it is possible that Al Aqah is not sufficiently genetically distinct from the other populations and accuracy has been shown to decrease when outgroups are not present in the data (Pickrell and Pritchard, 2012). We ran Treemix for 0 - 10 migrations using the parameters -bootstrap -noss -k 500. Migration edges were plotted until 99.8% of the variance in ancestry between populations was explained by the model. The consistency of migration edges was visually evaluated by running Treemix with 30 total replicates for each added migration edge number. Further, each of the 30 replicates was run using a different, randomly generated seed. We present results from one seed that had the highest likelihood for each number of migration edges.

Detection of loci under putative selection

To identify outlier loci, we used a Mahalanobis distance-based approach in the R package *pcadapt* (Luu et al., 2017)*,* which has been shown to be robust to a high degree of admixture and does not assume prior knowledge of population structure. Population structure was inferred using PCA, and putative outliers were detected with respect to how they relate to population structure. The number of principal components (K) was defined by running a PCA with $K = 1 – 20$, and applying Cattell's graphical rule (Cattell, 1966) to the screeplot of eigenvalues to determine the optimal number of principal components, as recommended by Luu et al., 2017. Finally, we used the R package *qvalue* to generate a list of candidate outlier SNPs using the *q*-value procedure at a false discovery rate (FDR) of $\alpha = 0.1$ (meaning that 10% of the SNPs are expected to be false positives).

We performed a search using the Nucleotide Basic Local Alignment Search Tool (BLASTn) on the genomic scaffolds which contained outlier loci against NCBI's Nucleotide collection (nr/nt) database in order to functionally annotate the identified outliers. An *E*-value cutoff of 10^{-8} was used and only outlier SNPs which were within 1 kb of the BLAST hit were retained.

Results

Draft genome

The optimal assembly (*k*-mer = 85) resulted in an genome assembly with 4,487,317 scaffolds, measuring a total of 1.59 Gb. The assembly had an N50 of 1,006 bp and a mean coverage of ~27x. We recovered 60% (16% complete and 44% partial) of the core eukaryotic genes and 75% (37% complete and 38% partial) of the core metazoan genes with CEGMA and BUSCO, respectively. The low recovery rates for conserved genes in the *E.* sp. *EZ* genome are due to the fragmented nature of the genome, which is likely a direct consequence of high heterozygosity and repeat content (25% of the genome is comprised of repeat regions). The estimated genome size was 609 Mb, the heterozygosity rate was 4.54%, and the duplication levels were 0.6%.

RAD sequencing

RAD sequencing of 94 *E.* sp. *EZ* individuals resulted in 347,439,950 total sequences, of which 287,973,771 (82.9%) were retained after initial quality filtering steps. Of the 59,446,179 discarded reads, 0.05%, 1.6%, and 15.4% were discarded due to low quality, ambiguous RAD-tags, and ambiguous barcodes, respectively. After mapping RADtags to the draft genome assembly and filtering for mapping quality, 378,775 SNPs were called. After stringent SNP filtering, 918 SNPs from 85 individuals across seven populations were kept. These 918 SNPs were used in downstream analysis unless a program did not allow triallelic SNPs, in which case we used a reduced VCF file containing 901 SNPs. This low SNP retention rate was likely due to extremely high heterozygosity and an abundance of repeat regions in the *E*. sp. *EZ* genome (Gautier et al., 2013).

Population genetic diversity and structure

Estimates of H_O and H_E across 901 SNPs was consistent across the seven sampling sites (H_O = 0.2076 – 0.2343, H_E = 0.2376 – 0.2572, Table 3.1). Nucleotide diversity ranged from 0.2491 to 0.2670 and was similar to $H_{E.}$ The only significant pairwise F_{ST} values were found when comparing populations from inside the PAG to those in the GO (Table 3.2). The highest significant F_{ST} values (0.02514, *P*-value=0.00000 and 0.02189, *P*-value=0.00000) were found when comparing Al Fiquet (GO) to Saadiyat (S-PAG) and Dhabiya (S-PAG), respectively; these PAG sites are the most geographically distant from the GO. When all samples in each respective Gulf were pooled, the F_{ST} between the PAG and the GO was 0.0057.

To further characterize population structure, we used STRUCTURE, smartpca*,* and a Tracey-Widom test. The Evanno method, which evaluates the second order rate of change of the likelihood function with respect to ΔK*,* identified K=2 (with ΔK = 5.2236, see Supplemental Figure 3.1) as the optimum number of populations from the STRUCTURE output (Figure 3.2; K = 3 and K = 4 are available in Supplemental Figure 3.2). The STRUCTURE plots showed population structure between the PAG and the GO, although a high degree of admixture resulted in all individuals with identities corresponding to both seas. We used smartpca to generate a principal component analysis (PCA) plot and we applied Cattell's graphical rule (Cattell, 1966) to the associated scree plots, which indicated that the optimal number of principal components was one (Figure 3.2). In other words, the majority of the variation in the data was explained by the first principal component and all subsequent axes only served to explain random variation. As the Evanno method cannot formally test for K = 1, we used the Tracey-Widom test to calculate the significance of eigenvectors (generated in smartpca) and subsequently, the number of populations within the dataset. We found that only the first eigenvector was significant (*P*-value = 3.37e-06) and explained 2.46% of the total genetic variation. All population structure analyses showed a slight degree of population differentiation.

Historical relationships

We ran TreeMix with 85 sea urchin samples from seven populations to identify patterns of divergence and add migration edges to the phylogenetic model. The proportion of variance began to asymptote at 0.999 when 6 migration edges were fit (Supplemental Figure 3.3). The consistency of these runs were evaluated using 30 independent runs of TreeMix with all 10 migration edges. Across all iterations, 95.1% of the total variance was explained by the graph model without any migration edges. The phylogenetic tree shows separation between the two seas, recapitulating the results found through population structure analyses (Figure 3.3). The first migration edge showed a migration event from Saadiyat (SA) to Dhabiya (DH), which are both located in the southern PAG (Figure 3.3). This result was consistent across all 30 replicates. Residual plots showed that as more migration edges were added, the proportion of variance in relatedness between populations explained by the models continued to increase. At six migration edges, there were more vectors moving from the PAG into the GO than from the GO into the PAG (the phylogenetic network at six migration edges were consistent across all 30 replicates). However, there were migration edges moving both in and out of the PAG, consistent with previous results that reveal a high degree of admixture between these seas (all migration events shown in Supplemental Figure 3.4).

Candidate loci under selection

Outlier detection in *pcadapt* was performed by retaining loci correlated with the first principal component axis after a 10% FDR correction. Of the 918 SNPs generated across 85 individuals, we identified nine candidate outliers. Of these nine outliers, we eliminated the loci that did not have a clear BLAST match (*E*-values cut off of 10^{-8}), and those whose associated genomic scaffold had a query length less than 5,000 bp. This resulted in one outlier on scaffold 0012050 (length of query: 7,820 bp) that was located in an *E.* sp. *EZ COLP5α* (5α collagen-like chain) (Exposito et al., 1995) gene, which has a clear ortholog in *Strongylocentrotus purpuratus* (the purple sea urchin; Accession number: AC165428.1). In adult sea urchins, this gene is expressed in mineralized regions and in the adult mutable collagenous tissues (Cluzel et al., 2001). However, the exact function of this gene has not been well characterized (Exposito et al., 1995).

Discussion

We sequenced and assembled a draft genome for the sea urchin, *E.* sp. *EZ* and generated a population genomic dataset consisting of seven populations from two environmentally distinct seas. With these genomic tools, we analyzed population dynamics of individuals living across dramatically divergent thermal and salinity environments. This study contributes to the growing body of literature characterizing population dynamics in environmentally extreme marine systems, and represents the first genome-wide investigation of a sea urchin from northeastern Arabia. Our findings are relevant to predicting how species will respond to future and ongoing climate change.

Echinometra is a pantropical genus with geographical distributions across the Indo-Pacific, Caribbean, and Atlantic. *Echinometra* is widely studied and recognized for their distinct patterns of population structure and speciation dynamics (McCartney et al., 2000;Landry et al., 2003;Lessios, 2006;Bronstein and Loya, 2013). The availability of the genomic-level sequence data will be a useful tool to explore the unique ecological and reproductive dynamics of this genus as well as a tool for comparative genomics with other sea urchins, as there are only a few sea urchin genomes available (Cameron et al., 2015). Until a recently published mitochondrial genome became available (Ketchum et al., 2018a), *E*. sp. *EZ* was misidentified as *E. mathaei*, highlighting the importance of genomic resources for taxonomic identification and associated studies.

A notable characteristic of the *E*. sp. *EZ* genome was the high frequency of polymorphisms (the estimated heterozygosity was approximately 4.5%). This is likely a result of either large population size or an elevated mutation rate. In the sea squirt, *Ciona savignyi*, high heterozygosity (4.49%) was shown to be driven by a large effective population size, not elevated mutation rates (Small et al., 2007). This level of heterozygosity is also comparable with the *Strongylocentrotus purpuratus* genome, which required the sequencing of large BAC clones to parse haplotypes (Sodergren et al., 2006). The high genomic variation results in a challenging genome assembly as it is difficult to distinguish between reads that are from duplicated but diverged sections of the genome or highly heterozygous homologs (Sodergren et al., 2006). This problem is further aggravated by repeat sequences and short read sequencing data. These variables resulted in a genome assembly that was too highly fragmented to perform gene annotations. However, our assembly was a valuable resource for our RAD-seq analyses.

Oceanographic circulation patterns, selective pressures exerted by temperature and salinity extremes, effective population sizes, and dispersal capabilities are all factors that may govern population structure and gene flow. It is often assumed that the life history traits of many marine invertebrates (i.e., long pelagic larval durations and large effective population sizes) should result in a lack of genetic structuring between geographically distant populations (Waples, 1998;Casteleyn et al., 2009). Modern coastlines in the PAG were formed only ~6,000 years ago following the Holocene transgression (Lambeck, 1996) and our two most distant sites are circa 500 km apart. Therefore, it would be reasonable to assume that marine organisms inhabiting the PAG and GO may represent one panmictic population. However, our data suggest weak but significant population structuring between the two seas.

The F_{ST} results shown in Table 3.2 indicate population differentiation between the PAG and the GO sites. Indeed, the only significant F_{ST} values were found when comparing sites within the PAG to sites within the GO. The only F_{ST} values that were not significant between the two seas were found when comparing Dibba Rock to Musandam and to Ras Ghanada (the two PAG sites closest to GO). No significant F_{ST} values were found when comparing between sites within the same sea. These findings are congruent with studies on other marine organisms in the region, which also describe significant population differentiation between the two seas. One study on the sea urchin *Diadema setosum* used mitochondrial DNA to investigate population structure around the Arabian Peninsula and found that $F_{\mathrm{ST}} = 0.05$ (Lessios et al., 2001). A study on the coral *Platygyra daedalea* analyzed the ITS region and found F_{ST} values ranging from 0.051 to 0.29 (Smith et al., 2017b). Finally, a study on the yellowbar angelfish, *Pomacanthus* *maculosus*, generated a SNP dataset with 10,225 SNPs and found that the F_{ST} between the two seas was 0.015 (Torquato et al., 2019). The F_{ST} values calculated in these studies all indicate more population structure than what was found in *E.* sp. *EZ* with the exception of some pairwise comparisons between specific sampling locations. The differences in the magnitude of values may be due to sequencing different gene regions and using different sequencing approaches. Although F_{ST} calculations were comparatively lower in *E*. sp. *EZ*, despite the fact that urchins are expected to be a high gene flow species (e.g., large population size, high larval dispersal capabilities, and reproductive output), we were still able to detect weak but significant population differentiation. Together, these studies support a hypothesis of a consistent genetic break for many species between the PAG and GO.

Through multiple analyses we revealed the presence of two populations of *E.* sp. *EZ*, each corresponding to their respective Gulf. Interestingly, the STRUCTURE analysis revealed that assignment probabilities of individuals varied greatly within and between populations. For example, in every collection site in the PAG except Saadiyat, there are some individuals whose assignment probabilities more closely associate with the GO than the PAG. This same subtle pattern can also be seen for individuals from the GO whose assignment probabilities more closely resemble individuals from the PAG. These results are in contrast to a similar study on the coral *P. daedalea* by Howells, et al. 2016 where one site within the PAG and one site within the GO were sampled. They also found that the most likely number of populations was two. However, the assignment probabilities of individuals were more clearly differentiated and there was little evidence of admixture. This could be a result of different larval characteristics (e.g., lecithotrophic versus planktotrophic larvae and associated differences in PLD) as the degree of genetic exchange and subsequent population structure in these two species relies on larval migration.

The patterns of admixture in *E*. sp. *EZ* between the two seas could be a result of a high degree of unidirectional or biased migration events. Oceanographic models have shown reduced mixing between the seas through subsurface outflow that prevents the transport of buoyant larvae, as well as long residence times of about 1-3 years for seawater in the PAG (Alosairi et al., 2011). However, our preliminary findings suggest that migration occurs bidirectionally between seas, although there were more migration vectors moving from the PAG to the GO. Alternatively, this population structure could be a result of genetic drift or the extreme environmental conditions of the PAG, which may act as a selective pressure on urchins and other species. Future research should employ demographic models to explore these hypotheses as well as explore other possible demographic events (e.g., range expansions, founder events, and population bottlenecks) which may have or may continue to shape allele frequency patterns between populations. Our outlier analysis was preliminary and only resulted in nine significant outliers under putative selection. We could only functionally annotate one of these outliers; *COLP5α*. It is currently unclear what the exact role of this gene is in sea urchins but it has been implicated in collagen formation. Collagen genes have been shown to respond transcriptionally to thermal stress in other marine invertebrates (DeSalvo et al., 2010;Kenkel et al., 2013). Further studies are warranted to generate a more suitable dataset for investigating genes under natural selection and to better understand the main drivers of this population differentiation.

Our study contributes to recent efforts to characterize the population dynamics of organisms across extreme environmental gradients. The most striking result from this analysis was the presence of population structuring given that the young age of the PAG, the dispersal capability of *E*. sp. *EZ*, and the large effective population sizes, should all act to homogenize population differentiation. The RAD-seq dataset and *E.* sp. *EZ* draft genome assembly presented here will provide a platform for future studies on this ecologically important and understudied species.

Acknowledgements

The field work was carried out using Core Technology Platform resources at New York University Abu Dhabi. We thank the Environmental Agency – Abu Dhabi for permits.

References

Alosairi, Y., Imberger, J., & Falconer, R. A. (2011). Mixing and flushing in the Persian Gulf (Arabian Gulf). *Journal of Geophysical Research: Oceans, 116*(C3).

Alsaffar, A. H., & Lone, K. P. (2000). Reproductive cycles of *Diadema setosum* and *Echinometra mathaei* (Echinoidea: echinodermata) from Kuwait (northern Arabian Gulf). *Bulletin of Marine Science, 67*(2), 845-856.

Bauman, A. G., Feary, D. A., Heron, S. F., Pratchett, M. S., & Burt, J. A. (2013). Multiple environmental factors influence the spatial distribution and structure of reef communities in the northeastern Arabian Peninsula. *Marine Pollution Bulletin, 72*(2), 302-312.

Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D., & Pirovano, W. (2010). Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics, 27*(4), 578-579.

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

Bronstein, O., & Loya, Y. (2013). The taxonomy and phylogeny of *Echinometra* (Camarodonta: Echinometridae) from the Red Sea and Western Indian Ocean. *PLOS One, 8*(10), e77374. doi:10.1371/journal.pone.0077374

Burt, J., Bartholomew, A., & Usseglio, P. (2008). Recovery of corals a decade after a bleaching event in Dubai, United Arab Emirates. *Marine Biology, 154*, 27-36. doi:10.1007/s00227-007-0892-9

Burt, J. A., Paparella, F., Al-Mansoori, N., Al-Mansoori, A., & Al-Jailani, H. (2019).

Causes and consequences of the 2017 coral bleaching event in the southern Persian/Arabian Gulf. *Coral Reefs, 38*(4), 567-589.

Cameron, R. A., Kudtarkar, P., Gordon, S. M., Worley, K. C., & Gibbs, R. A. (2015). Do echinoderm genomes measure up? *Marine Genomics, 22*, 1-9.

Casteleyn, G., Evans, K. M., Backeljau, T., D'hondt, S., Chepurnov, V. A., Sabbe, K., & Vyverman, W. (2009). Lack of population genetic structuring in the marine planktonic diatom *Pseudo-nitzschia pungens* (Bacillariophyceae) in a heterogeneous area in the Southern Bight of the North Sea. *Marine Biology, 156*(6), 1149-1158.

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Molecular Ecology, 22*(11), 3124-3140.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*(2), 245-276.

Cluzel, C., Lethias, C., Humbert, F., Garrone, R., & Exposito, J.-Y. (2001). Characterization of fibrosurfin, an interfibrillar component of sea urchin catch connective tissues. *Journal of Biological Chemistry, 276*(21), 18108-18114.

Coles, S. L. (2003). Coral species diversity and environmental factors in the Arabian Gulf and the Gulf of Oman: a comparison to the Indo-Pacific region. *Atoll Research Bulletin*.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., . . . Sherry, S. T. (2011). The variant call format and VCFtools. *Bioinformatics, 27*(15), 2156-2158.

DeFaveri, J., Jonsson, P. R., & Merilä, J. (2013). Heterogeneous genomic differentiation in marine Threespine Sticklebacks: Adaptation along an environmental gradient. *Evolution, 67*(9), 2530-2546. doi:10.1111/evo.12097

DeSalvo, M. K., Sunagawa, S., Voolstra, C. R., & Medina, M. (2010). Transcriptomic responses to heat stress and bleaching in the elkhorn coral *Acropora palmata*. *Marine Ecology Progress Series, 402*, 97-113.

Downing, N., & El-Zahr, C. (1987). Gut evacuation and filling rates in the rock-boring sea urchin, *Echinometra mathaei*. *Bulletin of Marine Science, 41*(2), 579-584.

Earl, D. A. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources, 4*(2), 359-361.

Excoffier, L., Laval, G., & Schneider, S. (2005). Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics, 1*, 117693430500100003.

Exposito, J. Y., Boute, N., Deleage, G., & Garrone, R. (1995). Characterization of two genes coding for a similar four□cysteine motif of the amino□terminal propeptide of a sea urchin fibrillar collagen. *European Journal of Biochemistry, 234*, 59-65.

Garrison, E., & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *arXiv preprint arXiv:1207.3907*.

Gautier, M., Gharbi, K., Cezard, T., Foucaud, J., Kerdelhué, C., Pudlo, P., . . . Estoup, A. (2013). The effect of RAD allele dropout on the estimation of genetic variation within and between populations. *Molecular Ecology, 22*(11), 3165-3178. doi:10.1111/mec.12089

Giles, J. L., Ovenden, J. R., AlMojil, D., Garvilles, E., Khampetch, K.-o., Manjebrayakath, H., & Riginos, C. (2014). Extensive genetic population structure in the Indo–West Pacific spot-tail shark, *Carcharhinus sorrah*. *Bulletin of Marine Science, 90*, 427-454.

Gleason, L. U., & Burton, R. S. (2016). Genomic evidence for ecological divergence against a background of population homogeneity in the marine snail *Chlorostoma funebralis*. *Molecular Ecology, 25*(15), 3557-3573. doi:10.1111/mec.13703

Gnerre, S., MacCallum, I., Przybylski, D., Ribeiro, F. J., Burton, J. N., Walker, B. J., . . . Sykes, S. (2011). High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proceedings of the National Academy of Sciences, 108*(4), 1513-1518.

Hoegh-Guldberg, O., Cai, R., Poloczanska, E. S., Brewer, P. G., Sundby, S., Hilmi, K., . . . Stone, D. A. (2014). “The Ocean” in *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part B: Regional Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, C. B. Field et al., Eds. (Cambridge Univ. Press, Cambridge, 2014). pp. 1655–1731.

Hoolihan, J., Premanandh, J., D’Aloia-Palmieri, M.-A., & Benzie, J. (2004). Intraspecific phylogeographic isolation of Arabian Gulf sailfish *Istiophorus platypterus* inferred from mitochondrial DNA. *Marine Biology, 145*(3), 465-475.

Howells, E. J., Abrego, D., Meyer, E., Kirk, N. L., & Burt, J. A. (2016). Host adaptation and unexpected symbiont partners enable reef□building corals to tolerate extreme temperatures. *Global Change Biology, 22*(8), 2702-2714.

Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics, 23*(14), 1801-1806.

Kajitani, R., Toshimoto, K., Noguchi, H., Toyoda, A., Ogura, Y., Okuno, M., . . . Maruyama, H. (2014). Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Research, 24*(8), 1384-1395.

Kelley, J. L., Brown, A. P., Therkildsen, N. O., & Foote, A. D. (2016). The life aquatic: advances in marine vertebrate genomics. *Nature Reviews Genetics, 17*(9), 523.

Kelly, M. W., Padilla□Gamiño, J. L., & Hofmann, G. E. (2013). Natural variation and the capacity to adapt to ocean acidification in the keystone sea urchin *Strongylocentrotus purpuratus*. *Global Change Biology, 19*(8), 2536-2546.

Kenkel, C. D., Meyer, E., & Matz, M. V. (2013). Gene expression under chronic heat stress in populations of the mustard hill coral (*Porites astreoides*) from different thermal environments. *Molecular Ecology, 22*(16), 4322-4334. doi:10.1111/mec.12390

Ketchum, R. N. (2020). *https://github.com/remiketchum/GBE_Ketchum_et_al_2020*.

Ketchum, R. N., DeBiasse, M. B., Ryan, J. F., Burt, J. A., & Reitzel, A. M. (2018). The complete mitochondrial genome of the sea urchin, *Echinometra* sp. *EZ. Mitochondrial DNA Part B, 3*(2), 1225-1227.

Lal, M. M., Southgate, P. C., Jerry, D. R., Bosserelle, C., & Zenger, K. R. (2017). Swept away: ocean currents and seascape features influence genetic structure across the 18,000 Km Indo-Pacific distribution of a marine invertebrate, the black-lip pearl oyster *Pinctada margaritifera*. *BMC Genomics, 18*, 66.

Lambeck, K. (1996). Shoreline reconstructions for the Persian Gulf since the last glacial maximum. *Earth and Planetary Science Letters, 142*(1-2), 43-57.

Lande, R. (1980). Genetic variation and phenotypic evolution during allopatric speciation. *The American Naturalist, 116*(4), 463-479.

Landry, C., Geyer, L., Arakaki, Y., Uehara, T., & Palumbi, S. R. (2003). Recent speciation in the Indo–West Pacific: rapid evolution of gamete recognition and sperm morphology in cryptic species of sea urchin. *Proceedings of the Royal Society of London B: Biological Sciences, 270*(1526), 1839-1847.

Lessios, H. (2006). Speciation in sea urchins. *Echinoderms: Durham Proceedings of the 12th International Echinoderm Conference* 91-101.

Lessios, H. A., Kessing, B. D., & Pearse, J. S. (2001). Population structure and speciation in tropical seas: global phylogeography of the sea urchin *Diadema*. *Evolution, 55*(5), 955-975.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997.*

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics, 25*(14), 1754-1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., . . . Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics, 25*(16), 2078-2079.

Luu, K., Bazin, E., & Blum, M. G. (2017). pcadapt: an R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources, 17*, 67-77.

Marçais, G., & Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics, 27*(6), 764-770.

McCartney, M. A., Keller, G., & Lessios, H. A. (2000). Dispersal barriers in tropical oceans and speciation in Atlantic and eastern Pacific sea urchins of the genus *Echinometra*. *Molecular Ecology, 9*(9), 1391-1400.

McClanahan, T. R., & Muthiga, N. A. (2007). Chapter 15 Ecology of *Echinometra*. In J. M. Lawrence (Ed.), *Developments in Aquaculture and Fisheries Science* (Vol. 37, pp. 297-317): Elsevier.

Miller, L. M., Kallemeyn, L., & Senanan, W. (2001). Spawning-site and natal-site fidelity by northern pike in a large lake: mark–recapture and genetic evidence. *Transactions of the American Fisheries Society, 130*(2), 307-316.

Nishimura, O., Hara, Y., & Kuraku, S. (2017). gVolante for standardizing completeness assessment of genome and transcriptome assemblies. *Bioinformatics, 33*(22), 3635-3637.

Ohdera, A., & Ryan, J. F. (2018). Ohdera et al 2018. https://github.com/josephryan/Ohdera_et_al_2018.

Oleksiak, M. F. (2016). Marine genomics: insights and challenges. In: Oxford University Press.

Palumbi, S. R., Grabowsky, G., Duda, T., Geyer, L., & Tachino, N. (1997). Speciation and population genetic structure in tropical Pacific sea urchins. *Evolution, 51*(5), 1506-1517. doi:doi:10.1111/j.1558-5646.1997.tb01474.x

Parra, G., Bradnam, K., & Korf, I. (2007). CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics, 23*(9), 1061-1067.

Patterson, N., Price, A. L., & Reich, D. (2006). Population Structure and Eigenanalysis. *PLOS Genetics, 2*(12), e190. doi:10.1371/journal.pgen.0020190

Pespeni, M. H., Barney, B. T., & Palumbi, S. R. (2012). Differences in the regulation of growth and biomineralization genes revealed through long□term common□garden acclimation and experimental genomics in the purple sea urchin. *Evolution, 67*(7), 1901-1914.

Pespeni, M. H., Garfield, D. A., Manier, M. K., & Palumbi, S. R. (2011). Genome-wide polymorphisms show unexpected targets of natural selection. *Proceedings of the Royal Society B: Biological Sciences, 1732*, 1412-1420.

Pespeni, M. H., & Palumbi, S. R. (2013). Signals of selection in outlier loci in a widely dispersing species across an environmental mosaic. *Molecular Ecology, 22*(13), 3580-3597.

Pickrell, J. K., & Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLOS Genetics, 8*(11), e1002967.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics, 155*(2), 945-959.

Puritz, J. B., Hollenbeck, C. M., & Gold, J. R. (2014). dDocent: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. *PeerJ, 2*, e431.

Rahman, M. A., Yusoff, F. M., Arshad, A., & Uehara, T. (2014). Effects of delayed metamorphosis on larval survival, metamorphosis, and juvenile performance of four closely related species of tropical sea urchins (Genus *Echinometra*). *The Scientific World Journal, 2014*.

Reitzel, A., Herrera, S., Layden, M., Martindale, M., & Shank, T. (2013). Going where traditional markers have not gone before: utility of and promise for RAD sequencing in marine invertebrate phylogeography and population genomics. *Molecular Ecology, 22*(11), 2953-2970.

Ryan, J. F. (2015a). FastqSifter. *https://github.com/josephryan/FastqSifter*.

Ryan, J. F. (2015b). matemaker. *https://github.com/josephryan/matemaker*.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics, 31*(19), 3210-3212.

Small, K. S., Brudno, M., Hill, M. M., & Sidow, A. (2007). Extreme genomic variation in a natural population. *Proceedings of the National Academy of Sciences, 104*(13), 5698-5703.

Smith, E. G., Hume, B. C. C., Delaney, P., Wiedenmann, J., & Burt, J. A. (2017). Genetic structure of coral-Symbiodinium symbioses on the world's warmest reefs. *PLOS One, 12 (6)*, e0180169. doi:10.1371/journal.pone.0180169

Smith, E. G., Vaughan, G. O., Ketchum, R. N., McParland, D., & Burt, J. A. (2017). Symbiont community stability through severe coral bleaching in a thermally extreme lagoon. *Scientific Reports, 7*, 2428. doi:10.1038/s41598-017-01569-8

Sodergren, E., Weinstock, G. M., Davidson, E. H., Cameron, R. A., Gibbs, R. A., Angerer, R. C., . . . Burke, R. D. (2006). The genome of the sea urchin *Strongylocentrotus purpuratus*. *Science, 314*(5801), 941-952.

Takeuchi, T., Masaoka, T., Aoki, H., Koyanagi, R., Fujie, M., & Satoh, N. (2019). Divergent northern and southern populations and demographic history of the pearl oyster in the western Pacific revealed with genomic SNPs. *Evolutionary Applications, 13*(4), 837-853.

Torquato, F., Range, P., Ben□Hamadou, R., Sigsgaard, E. E., Thomsen, P. F., Riera, R., .

.

& Marshell, A. (2019). Consequences of marine barriers for genetic diversity of the coral□specialist yellowbar angelfish from the Northwestern Indian Ocean. *Ecology and Evolution*.

Vaughan, G. O., & Burt, J. A. (2016). The changing dynamics of coral reef science in Arabia. *Marine Pollution Bulletin, 105*(2), 441-458.

Vurture, G. W., Sedlazeck, F. J., Nattestad, M., Underwood, C. J., Fang, H., Gurtowski, J., & Schatz, M. C. (2017). GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics, 33*(14), 2202-2204.

Waples, R. S. (1998). Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species. *Journal of Heredity, 89*(5), 438-450.

Whiteley, A. R., Spruell, P., & Allendorf, F. W. (2004). Ecological and life history characteristics predict population genetic divergence of two salmonids in the same landscape. *Molecular Ecology, 13*(12), 3675-3688.

Xuereb, A., Benestan, L., Normandeau, É., Daigle, R. M., Curtis, J. M., Bernatchez, L.,

& Fortin, M. J. (2018). Asymmetric oceanographic processes mediate connectivity and population genetic structure, as revealed by RAD seq, in a highly dispersive marine invertebrate (*Parastichopus californicus*). *Molecular Ecology, 27*(10), 2347-2364.

Table 3.1: Summary of genetic diversity statistics for seven populations of *E.* sp. *EZ.* S-PAG is southern PAG, N-PAG is northern PAG and GO is the Gulf of Oman. Variant sites are the number of total SNPs; % Polymorphic loci is the proportion of SNPs in this population; Num Indv is the mean number of individuals per locus in this population; H_O is the mean observed heterozygosity per population; H_E is the mean expected heterozygosity per population; *P*i is mean nucleotide diversity.

Pop ID	**Location**	**Variant Sites**	**% Polymorphic Loci**	**Num Indv**	H_O	H_E	***P*i**
Dhabiya	**S-PAG**	900	89.5556	10.8278	0.2266	0.2472	0.2592
Saadiyat	**S-PAG**	901	89.1232	10.8169	0.2076	0.2376	0.2491
Ras Ghanada	**S-PAG**	901	94.1176	14.7714	0.2151	0.2471	0.2558
Musandam	**N-PAG**	901	94.4506	13.8224	0.2343	0.2572	0.2670
Dibba Rock	**GO**	900	91.5556	10.7744	0.2163	0.2499	0.2621
Al Fiquet	**GO**	900	84.2222	7.9744	0.2172	0.2391	0.2552
Al Aqah	**GO**	901	95.3385	14.8058	0.2323	0.2554	0.2649

Table 3.2: F_{ST} values from 918 SNPs across seven populations of *E.* sp. *EZ.* Values in bold were significant ($P < 0.05$).

	Dhabiya	**Saadiyat**	**Ras Ghanada**	**Musandam**	**Dibba Rock**	**Al Fiquet**	**Al Aqah**
Dhabiya	—	—	—	—	—	—	—
Saadiyat	0.01241	—	—	—	—	—	—
Ras Ghanada	0.01208	0.01114	—	—	—	—	—
Musandam	0.01168	0.00665	0.00658	—	—	—	—
Dibba Rock	**0.01479**	**0.01931**	0.0062	0.00906	—	—	—
Al Fiquet	**0.02189**	**0.02514**	**0.01529**	**0.01496**	0.00625	—	—
Al Aqah	**0.01578**	**0.01996**	**0.01244**	**0.01349**	0.00275	0.00251	—

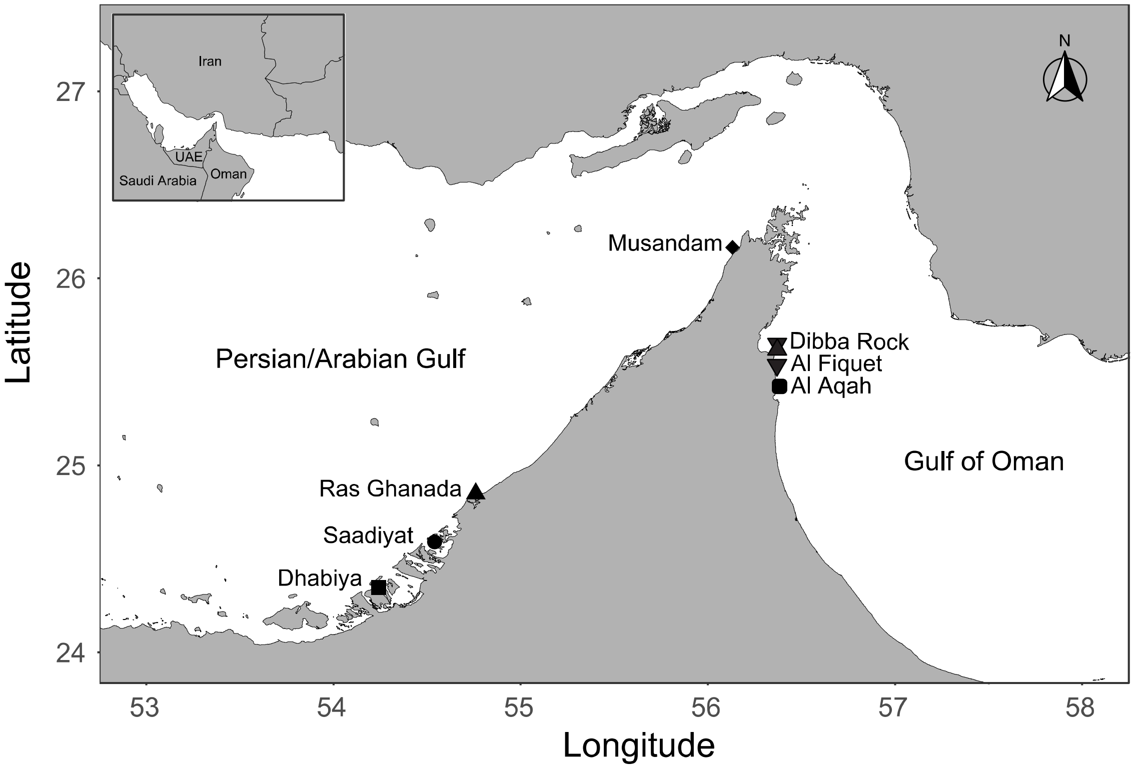


Figure 3.1: Map of the seven *E.* sp. *EZ* sampling locations.

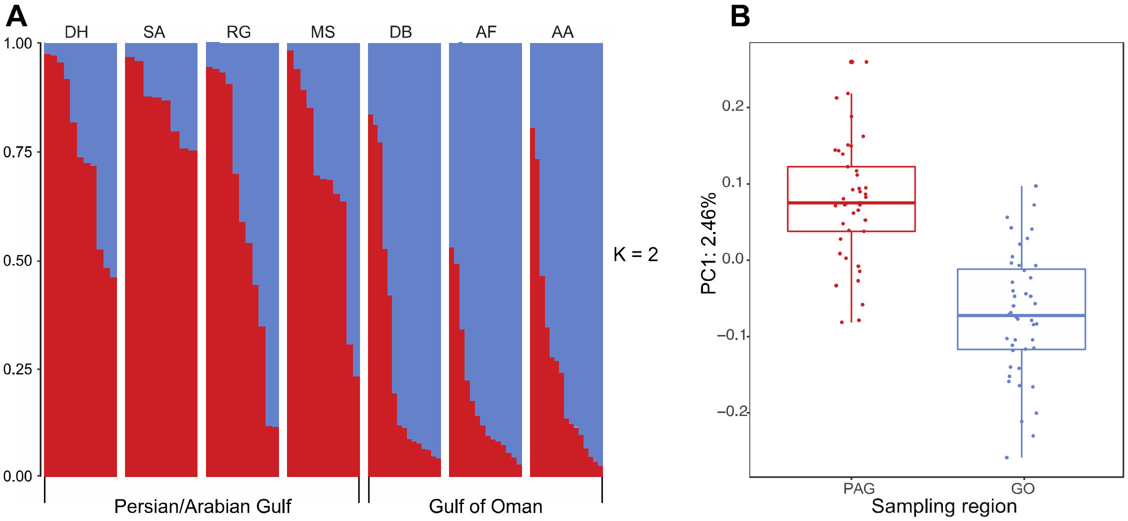


Figure 3.2: **A:** plot of the individual ancestry inference for $K = 2$ based on all 918 loci. The population abbreviations are as follows: DH = Dhabiya, SA = Saadiyat, RG = Ras Ghanada, MS = Musandam, DB = Dibba Rock, AF = Al Fiquet, and AA = Al Aqah. **B:** Box plot of eigenvalues for 85 individuals explained by principal component one, generated in the smartpca package.

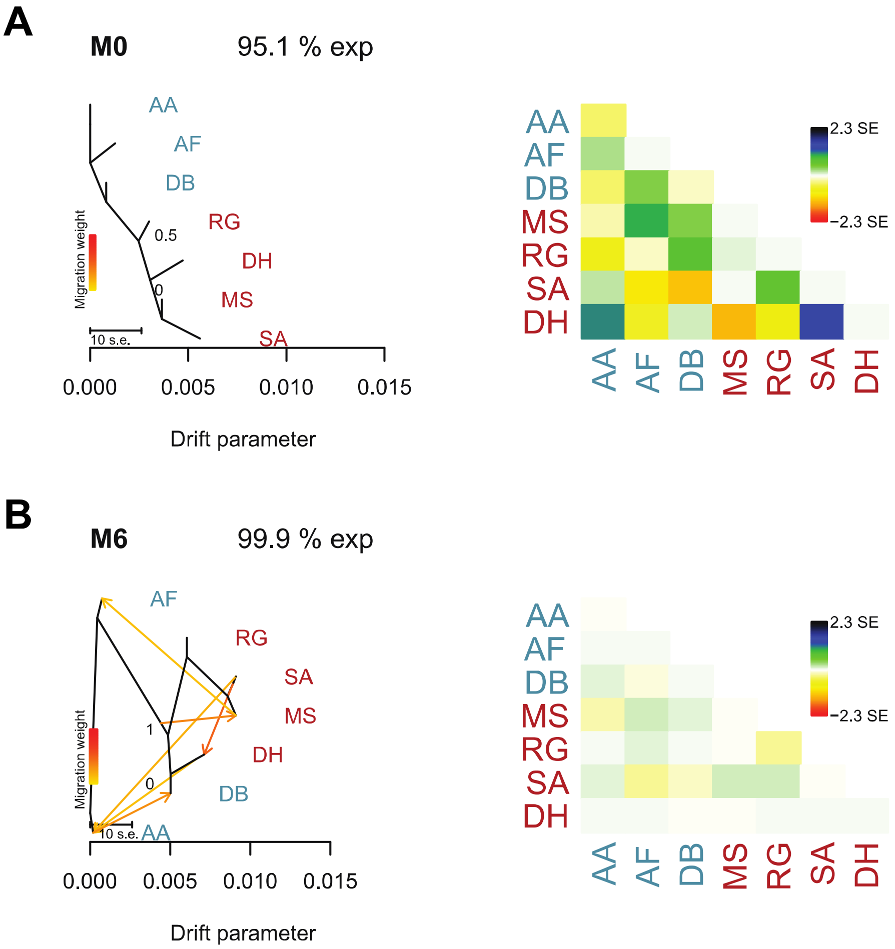


Figure 3.3: Phylogenetic network of the inferred relationships between seven populations of *E*. sp. *EZ*. The population abbreviations are: DH = Dhabiya, SA = Saadiyat, RG = Ras Ghanada, MS = Musandam, DB = Dibba Rock, AF = Al Fiquet, and AA = Al Aqah. Population abbreviations were colored based on their Gulf of origin (PAG = red, GO = blue). Migration edges are colored according to percent ancestry received from the donor population and s.e. represents the standard error of migration rates. **A:** M0 represents a phylogram with no migration edges. **B:** M6 contains six migration edges. Next to each phylogenetic network are the corresponding residual plots.

Table S3.1: Sampling site names and ID, as well as the coordinates of sampling sites.

Sampling Site	Site ID	Latitude	Longitude
Dhabiya	DH	24.36549992	54.10079998
Saadiyat	SA	24.59899996	54.42149998
Ras Ghanada	RG	24.84816545	54.69029588
Musandam	MS	26.174924	56.173506
Dibba Rock	DB	25.60313453	56.34847672
Al Fiquet	AF	25.562684	56.353478
Al Aqah	AA	25.4929462	56.36352265

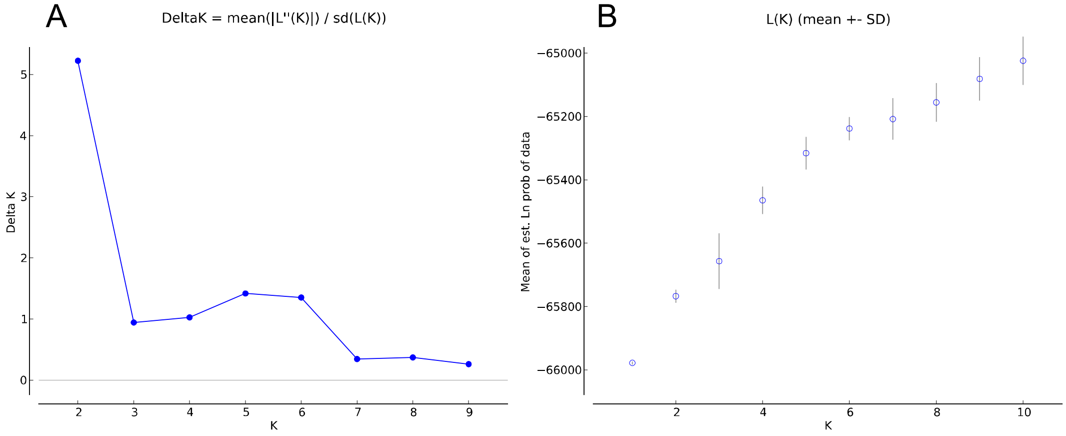


Figure S3.1: STRUCTURE HARVESTER output from STRUCTURE analysis which reveals the most probable number of populations. **A.** Delta *K* is related to the second order rate of change of the log probability of data with respect to the number of clusters and is a good predictor of the number of populations (*K* = 2). **B.** Log probability of data L(*K*) as a function of K.

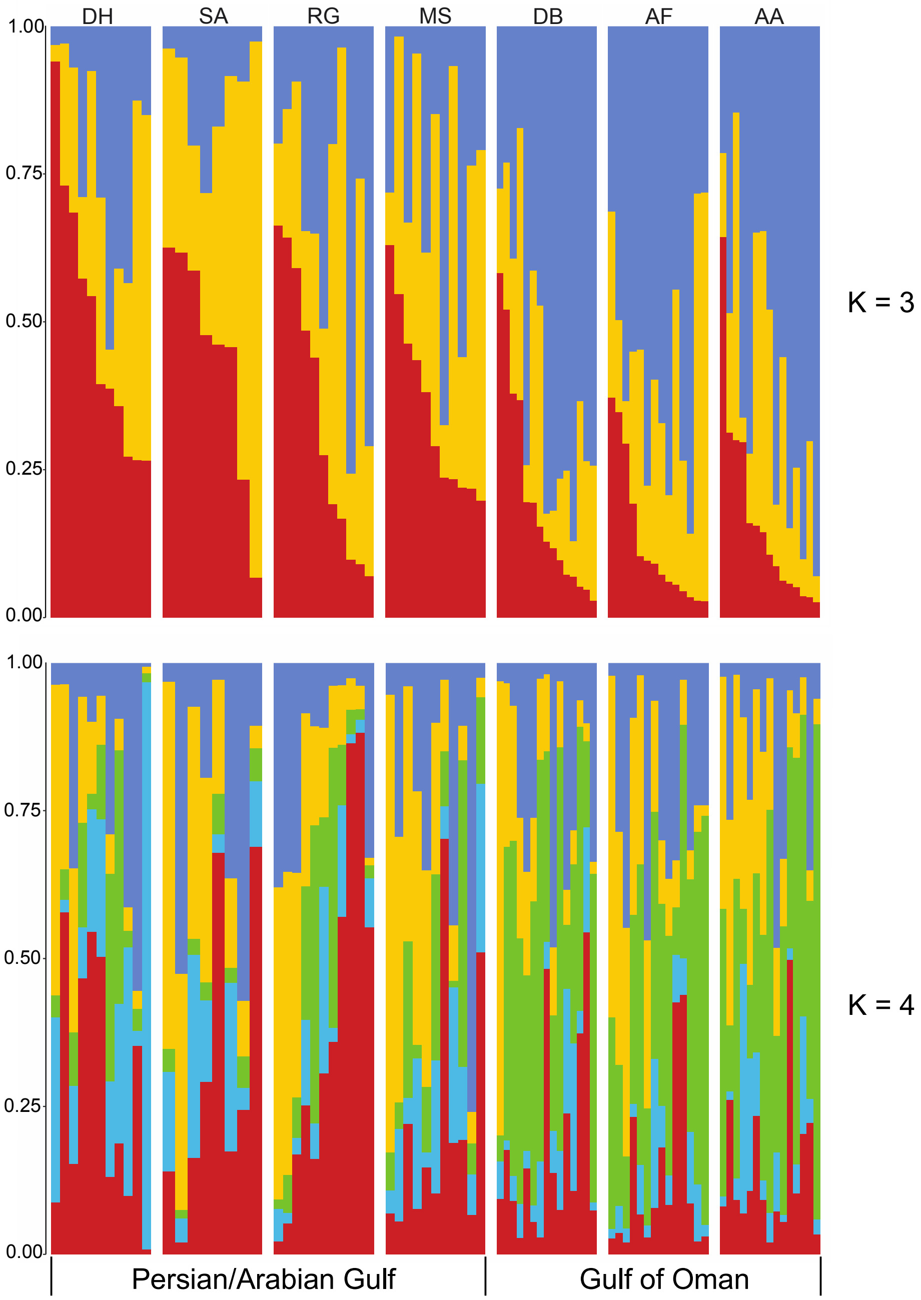


Figure S3.2: STRUCTURE plots of the individual ancestry inference for K = 3 and K = 4 based on all 918 loci. The population abbreviations are as follows: DH = Dhabiya, SA = Saadiyat, RG = Ras Ghanada, MS = Musandam, DB = Dibba Rock, AF = Al Fiquet, and AA = Al Aqah.

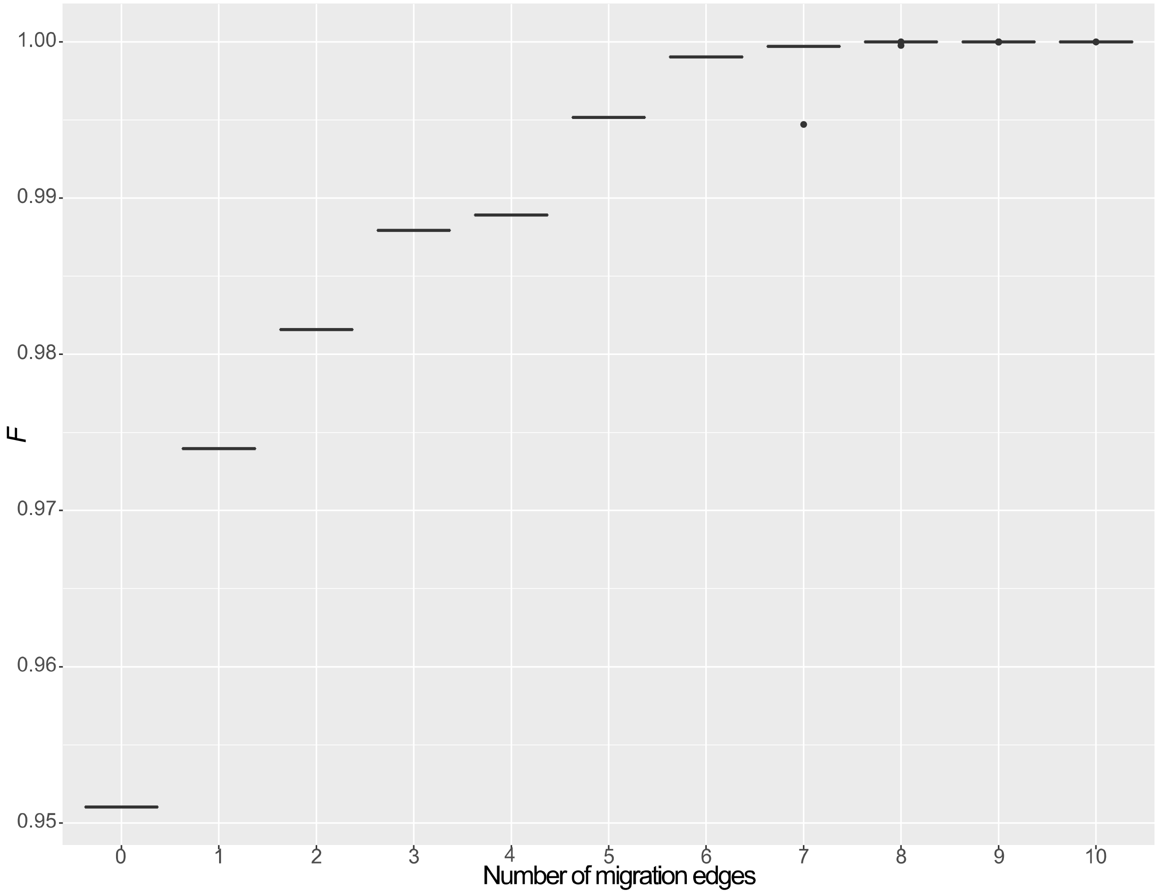


Figure S3.3: Box plot showing the fraction of variance in relatedness between populations accounted for by phylogenetic models with 0 to 10 migration edges. The fraction begins to asymptote near 0.998 at six migration edges.

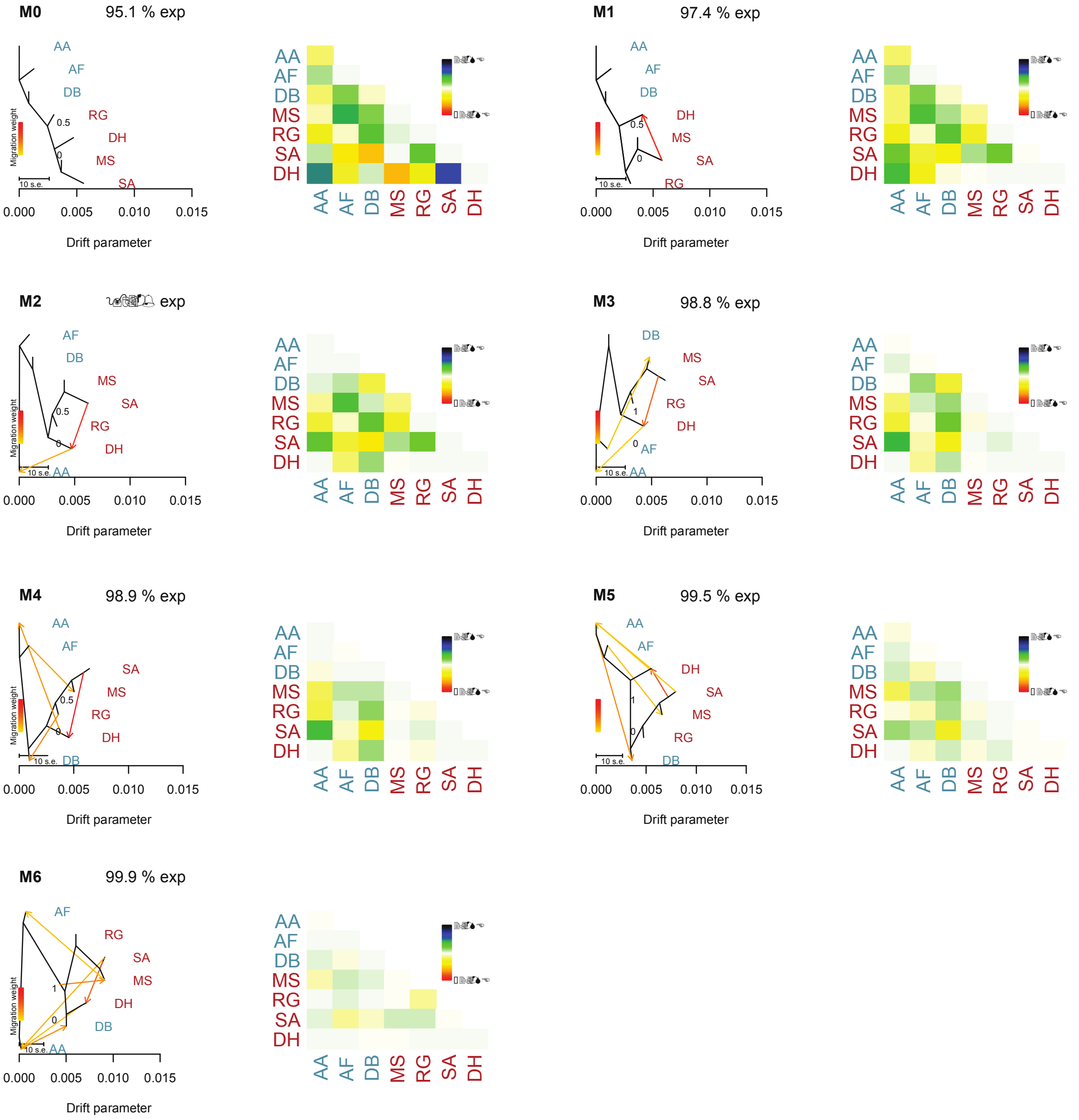


Figure S3.4: Phylogenetic network of the inferred relationships between seven populations of *E.* sp. *EZ*. The population abbreviations are: DH = Dhabiya, SA = Saadiyat, RG = Ras Ghanada, MS = Musandam, DB = Dibba Rock, AF = Al Fiquet, and AA = Al Aqah. Population abbreviations were colored based on their Gulf of origin (PAG = red, GO = blue) and s.e. represents the standard error of migration rates. Migration edges are colored according to percent ancestry received from the donor population. M0 represents a phylogram with no migration edges and each increasing number (e.g., M1-6) represents an additional migration edge. Next to each phylogenetic network are the corresponding residual plots.

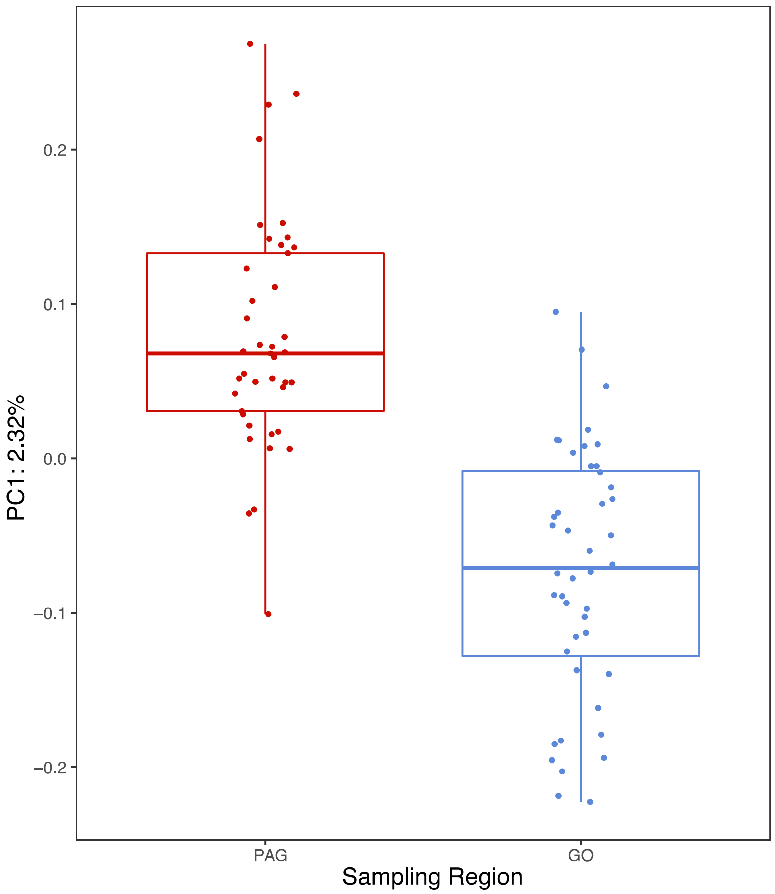


Figure S3.5: As our analyses does not account for linkage disequilibrium (LD), we tested the impact of LD on our dataset by using the *–-indep-pairwise* command in Plink (Purcell et al. 2007) with a 10kb window, a step size of 1, and a pairwise r^2 threshold of 0.2. The LD-pruned SNP dataset containing 691 SNPs in linkage equilibrium were analyzed in smartpca to verify that population structure was not a result of LD. We used a Tracey-Widom test to calculate the significance of each eigenvector and found that PC1 was significant with a *P*-value = 0.000727, indicating that there are two populations in the dataset.

CHAPTER 4

CHROMOSOME-LEVEL GENOME ASSEMBLY OF THE HIGHLY HETEROZYGOUS SEA URCHIN *ECHINOMETRA* SP. EZ

Remi N. Ketchum, Phillip L. Davidson, Edward G. Smith, John A. Burt, Joseph F. Ryan, Greg A. Wray, Adam M. Reitzel

Abstract

Echinometra is the most widespread genus of sea urchin and plays an important role as a major bioeroder on the reefs they inhabit. Most studies to date have focused on reconstructing their phylogenies and describing species-specific fertilization patterns. Genetic data on this genus is limited and generally only include a few selected gene sequences for any species. Here, we present a chromosome-level genome assembly based on 10x Genomics, PacBio, and Hi-C sequencing for *E.* sp. EZ from the Persian/Arabian Gulf. The genome is assembled into 210 scaffolds totaling 817.8 Mb with an N50 of 39.5 Mb. From this assembly we were able to determine that the *E*. sp. EZ genome consists of 2n = 42 chromosomes. BUSCO analysis showed that 95.3% of BUSCO genes were complete and ab initio and transcript-informed gene modeling and annotation identified 29,898 genes. Due to this species' distribution in high temperature and salinity environments, we compared gene families and transcription factors associated with environmental stress response ("defensome") with four other echinoid species with existing genomic resources. While the number of defensome genes was broadly similar for all species, we identified strong signatures of selection as well as losses of transcriptions factors important for environmental response. This genome will provide

key insights into comparative genomics and speciation biology as well as serve as a useful tool for the scientific community.

Introduction

The expansion of sequencing technologies, particularly long-read sequencing, has dramatically improved the assembly of genomes for all species, particularly those with high heterozygosity and/or repeat content. These advances are especially impactful for species with few closely related reference genomes available and species with large population sizes where nucleotide diversity is typically high. Both of these factors are typical for the majority of marine invertebrates. Within the phylum Echinodermata, the purple sea urchin, *Strongylocentrotus purpuratus*, was the first genome to be sequenced (Sodergren et al., 2006). This genome not only revealed a number of critical expansions of gene families as well as conservation of core genes in developmental regulatory networks, but also provided an essential tool for future studies of cis-regulation and genetic adaptation. Since the release of this genome, the genomes for other echinoderms have been published (Long et al., 2016;Hall et al., 2017;Kinjo et al., 2018) and genomic resources such as SpBase (Cameron et al., 2009), Echinobase (Kudtarkar and Cameron, 2017), and EchinoDB (Janies et al., 2016) have been made available for data sharing and analyses. More recently, the genomes of *Lytechinus variegatus* (Davidson et al., 2020) and *Lytechinus pictus* (Warner et al., 2021) have been assembled to chromosome-level. These genomic resources for echinoderms provide an important resource to expand the toolkit used to address hypotheses in marine biology, cell and developmental biology, and evolutionary biology.

A chromosome-level reference assembly provides information beyond areas focused on cellular and molecular biology by also enabling the investigation of fundamental evolutionary questions. While transcriptomes have provided the tools necessary to understand coding sequence evolution across taxa, the rapid generation of reference genomes has the power to expand upon this research because it includes noncoding regions and information on genome architecture. For example, a well-resolved genome can provide the tools necessary to investigate genomic variation such as structural rearrangements and copy number variants and perform scans for selection in coding and noncoding regions, quantitative trait loci (QTL) mapping, or taxonomic species delineation (Ekblom and Wolf, 2014). Given the progress made in sea urchin genome assemblies, it is necessary to expand on the taxonomic representation of chromosome-level assemblies in order to draw inferences regarding evolutionary processes associated with the closest known relatives of chordates.

Echinometra is a widespread genus of pantropical sea urchins that have been the focus of many genetic, ecological, and reproductive studies. Their distribution includes the Indo-Pacific, Caribbean, and Atlantic and they are often the most prevalent urchins in the reefs they inhabit (McClanahan and Muthiga, 2007;Bronstein and Loya, 2013). Although widely studied, the species-level taxonomy for this genus is unfinished and while some studies have referenced eight species, there is more likely to be at least nine species of *Echinometra* (Bronstein and Loya, 2013;Ketchum et al., 2018a). Bronstein and Loya (2013) were the first to describe *Echinometra* species EE and ZE (these abbreviations reference their collection site: Eilat and Zanzibar, respectively), which had been historically misidentified as *E. mathaei,* and were inferred to be a single species (hence the combined name *E*. sp. EZ). The misidentification for this new species also occurred in the Persian/Arabian Gulf (herein the PAG) and Gulf of Oman but has since been corrected (Ketchum et al., 2018a;Ketchum et al., 2018b;Ketchum et al., 2020). Although a highly fragmented draft genome was generated for *E*. sp. EZ (Ketchum et al., 2020), given the ecological importance and extensive genetic (Matsuoka and Toshihiko, 1991;Palumbi et al., 1997;A. et al., 2000;McCartney and Lessios, 2004;Bronstein and Loya, 2013) and reproductive assays on this genus (Metz et al., 1994;Aslan and Uehara, 1997;Rahman et al., 2000;Rahman et al., 2001;Mita et al., 2004), a chromosome-level genome assembly would be instrumental to answering many fundamental biological questions in ecology and evolution.

Here we present the chromosome-level assembly for *E.* sp. EZ and the associated gene annotation set. This new assembly is a marked improvement on the draft genome assembled in 2020. We highlight some of the insights from this genome with case studies in gene family evolution and whole genome analyses of selection.

Materials and Methods

Tissue collection and DNA/RNA extraction

One *E.* sp. EZ adult was collected from Dhabiya Reef (lat-long: 24.36549992, 54.10079998) in December 2017 for isolation of DNA for genome assembly. Gonadal tissue was dissected from this individual and immediately placed on dry ice and sent to the Dovetail Genomics Center who performed the subsequent DNA extraction using the Qiagen Genomic-tip DNA isolation method.

One *E.* sp. EZ adult was collected from Saadiyat Reef (lat-long: 24.59899996, 54.42149998) in December 2017 for isolation of RNA for the generation of a transcriptome. Gonadal tissue was dissected from a single individual and immediately placed on dry ice. RNA was extracted using the RNAqueous Total RNA Isolation Kit. Genomic DNA was sent to DHMRI (Kannapolis, NC, USA) for library preparation and sequencing with an Illumina HiSeq2500. Total RNA was quantified using the the Quant-iT RiboGreen RNA Assay Kit (Thermo Fisher) and RNA integrity assessed using an Agilent Bioanalyzer (Santa Clara, CA USA). RNA sequencing libraries were generated using the Illumina TruSeq RNA Library Prep RNA Kit following the manufacturer's protocol and quantitated using qPCR and fragments visualized using an Agilent Bioanalyzer. The library was run on one flow cell for a 125 bp paired end sequencing run.

Sequencing approach

Dovetail Genomics Center performed all the library construction and sequencing. The assembly and scaffolding was done through a joint effort between the authors and Dovetail. Library construction, sequencing, assembly, and scaffolding are described in detail below.

10X library prep and sequencing

Whole genome sequencing libraries were prepared with 1.0 ng of genomic DNA using the Chromium Genome Library and Gel Bead Kit v.2 (10X Genomics, cat. 120262).

Link-read based technology using 10X Genomics Chromium was performed according to manufacturer's instructions with one modification. Briefly, in order to create Gel Bead-in-Emulsions (GEMs), gDNA was combined with Master Mix, a library of Genome Gel Beads, and partitioning oil on a Chromium Genome Chip. The GEMs were isothermally amplified with primers containing an Illumina Read 1 sequencing primer, a unique 16-bp 10x bar-code and a 6-bp random primer sequence, and bar-coded DNA fragments were recovered for Illumina library construction._The amount and fragment size of the post-GEM DNA was quantified prior to using a Bioanalyzer 2100 with an Agilent High sensitivity DNA kit (Agilent, cat. 5067-4626). The GEM amplification product was sheared on an E220 Focused Ultrasonicator (Covaris, Woburn, MA) to approximately 350bp (55 seconds at peak power = 175, duty factor = 10, and cycle/burst = 200). Next, the sheared GEMs were converted to a sequencing library following the 10X standard operating procedure and the library was quantified by qPCR with a Kapa Library Quant kit (Kapa Biosystems-Roche). Finally, the library was sequenced on a partial lane (472M reads) of the NovaSeq6000 sequencer (Illumina, San Diego, CA) with paired-end 150 bp reads.

PacBio library prep and sequencing

A Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA) was used to quantify DNA samples. The PacBio SMRTbell libraries (~20kb) for PacBio Sequel were constructed using SMRTbell Template Prep Kit 1.0 (PacBio, Menlo Park, CA, USA) using the manufacturer recommended protocol. The Sequel Binding Kit 2.0 (PacBio) was used to bind the pooled library to the polymerase and then loaded onto the PacBio Sequel using the MagBead Kit V2 (PacBio). Sequencing was performed on one PacBio Sequel SMRT cells (Instrument Control Software Version 5.0.0.6235, Primary Analysis Software Version 5.0.0.6236 and SMRT Link Version 5.0.0.6792).

Dovetail Hi-C library prep and sequencing

A Dovetail HiC library was prepared in a similar manner as described in Lieberman-Aiden et al. (2009). For each library, chromatin was fixed with formaldehyde in the nucleus and then extracted. This was then digested with DpnII and the 5' overhangs were filled with biotinylated nucleotides. After the free blunt ends were ligated, crosslinks were reversed and the DNA was purified from protein. Biotin that was not internal to ligated fragments was removed and DNA was then sheared to a mean fragment size of 350 bp. Sequencing libraries were generated using NEBNext Ultra enzymes and Illumina-compatible adapters and biotin-containing fragments were isolated using streptavidin beads before the PCR enrichment of each library. Finally, the libraries were sequenced on an Illumina HiSeq X to produce 200 million 2x150 bp paired end reads.

Genome assembly

We ran CANU v.2.0 (Koren et al., 2017) with the following parameters to generate a *de novo* assembly using the PacBio reads: genomeSize=1300, minReadLength=3000, corOutCoverage=300, useGrid=False. To align the 10x sequences to the CANU assembly, we used longRanger v.2.2.2 (Marks et al., 2019) and then Pilon to polish the CANU assembly with the 10x reads. In order to reduce the percentage of duplication (due to the high levels of heterozygosity, see below), we ran PurgeHaplotigs

(Roach et al., 2018) multiple times with the percent cutoff for identifying a contig as a haplotig ranging from 35 to 55%. A cutoff of 40% yielded the most complete assembly and this genome assembly was subsequently run through the PurgeHaplotigs ‘clip’ option that identifies and trims overlapping contig ends.

The purged genome assembly and the Dovetail HiC library reads were input into HiRise. HiRise is a software pipeline designed specifically for using proximity ligation data to scaffold genome assemblies; Putnam et al. (2016). The Dovetail HiC library sequences were aligned to the draft input assembly using the modified SNAP read mapper (http://snap.cs.berkeley.edu). HiRise analyzed the separations of Dovetail HiC read pairs mapped within the draft scaffolds to produce a likelihood model for genomic distance between read pairs. The model was used to identify and break misjoins, to score potential joins, and make joins above a threshold.

Genome prediction and annotation

Prior to gene prediction and annotation, genomic repetitive elements were identified with RepeatModeler v1.0.11 (Smit and Hubley, 2008-2015) to generate a *E*. sp. EZ-specific repeat element library. Repetitive regions were soft-masked prior to gene prediction and annotation using RepeatMasker v4.0.8 (Smit et al., 2013-2015) with the -s and -xsmall flags. Transcriptome sequences were cleaned and trimmed using Trimmomatic v0.38 (Bolger et al., 2014) with a phred quality score of 33. Leading and trailing bases with a quality score <3 were removed, a 4-base wide sliding window was used to cut where the average quality per based dropped below 15, and reads that were <36 bp long were removed. BWA-mem v0.7.17 (Li, 2013) was used to align the transcriptome sequences to the genome with default parameters. Aligned RNA-seq data and *S. purpuratus* v5.0 protein models (https://www.ncbi.nlm.nih.gov/assembly/GCF_000002235.5, accessed February 23, 2021) were input into the BRAKER v2.1.5 (Hoff et al., 2018) pipeline which relies on GeneMark v4.62 (Lomsadze et al., 2005) and Augustus v3.4.0 (Stanke et al., 2006) to generate gene models.

To identify transposable elements (TE) in the *E.* sp. EZ genome, a master TE file with 26,470 TEs was generated by combining TE's from RepeatModeler and the default set of TE's that comes from Maker. Potential TEs were inspected with BlastP v2.5.0+ (Altschul et al., 1990) against the TE master file and the *S. purpuratus* v5.0 gene models. We then filtered gene models by removing those that 1) had a hit to the TE master file but no hit to the S. *purpuratus* gene models, 2) had a highly significant hint (BLASTX E-value <1e-20) to a TE but a weak hit to the *S. purpuratus* gene models (BLASTX E-value>1), and 3) had a hit to the *S. purpuratus* gene models and was labelled as a retrotransposon.

Assembled protein models were functionally annotated using BlastP v2.5.0+ with three protein databases: *S. purpuratus* v4.2, UniProt Knowledgebase Swiss-Prot protein models v2021-03, and RefSeq invertebrate protein models with *S. purpuratus* excluded. Lastly, BUSCO v5 (Manni et al., 2021) was used to measure the completeness of the genome assembly using the metazoan database.

PFAM analysis

We conducted a comparison of gene family representation of *E.* sp. EZ with other echinoids to compare the gene annotations and identify potential expansions and contractions across species. We focused this comparison on genes and gene families related to environmental responses. The list of genes included for these comparisons were from the chemical defensome developed from *S. purpuratus* (Goldstone et al., 2006). We performed hidden Markov model (HMM) searches using HMMER (Finn et al., 2011) and PFAM profiles which represent the chemical defensome in five species of sea urchins: *E.* sp. EZ, *L. variegatus* (Davidson et al., 2020), *S. purpuratus* (v5.0 gene models), *Heliocidaris erythrogramma* (unpublished data), and *Heliocidaris tuberculata* (unpublished data).

Positive selection

We used *adaptiPhy* (Berrio et al., 2020), which infers regions of the genome that are targets of branch-specific positive selection, to identify specific noncoding regions of the *E.* sp. EZ genome under selection. Our analyses compared *E.* sp. EZ to the *H. erythrogramma*, and *H. tuberculata* genomes with *L. variegatus* used as the outgroup. First, we curated a list of 25 transcription factors based on their roles in the defensome, general stress response, as well as metabolism, cell growth, immunity, and wound healing. Using a whole-genome alignment between all four species, we extracted 25 kb upstream and downstream of the transcription start site for each of the 25 genes. We used the sliding window approach in bedtools to split these regions into 300 bp fragments and excluded any coding sequences. We then identified one-to-one orthologous sequences shared across all urchins and ran *adaptiPhy* to test for selection.

After filtering for gaps and alignments lengths of at least 75 bp in each species, a *P*-value was calculated based on a likelihood ratio test comparing models of branch-specific positive selection and neutrally evolving sequence, and then adjusted using the false discovery rate (FDR) estimation. Finally, we calculated the zeta statistic (ratio of substitution rate of query to neutral substitution rate) to visualize the strength of selection. The branch substitution rates for the zeta scores were calculated separately in *phyloFit* (Hubisz et al., 2011) for the query and reference sequences. A zeta statistic that is less than one suggests negative selection and a statistic greater than 1 indicates positive selection.

Results and Discussion

Genome assembly and annotation

The initial *E.* sp. EZ draft genome (Ketchum et al., 2020) that was generated using Illumina short reads had an assembly size of 1,589 Mb, was comprised of 4,487,317 scaffolds, and had a BUSCO complete score of 34.8% (Table 1). The improved assembly generated with PacBio long reads produced a genome assembly size of 817.5 Mb with 3,800 scaffolds and an overall BUSCO completeness score of 94.4%. The final, chromosome-level genome assembly that included the Hi-C reads and was assembled with HiRise software produced an assembly that was 817.8 Mb in length with 210 scaffolds. The longest scaffold was 65.8 Mb and the scaffold N50 length was 39.5 Mb. The contact map is indicative of the presence of 21 chromosomes in *E.* sp. EZ (Figure 4.1). This is consistent with estimates for other *Echinometra* species based on karyotyping (Uehara et al., 2020). The assembly is highly complete with very little duplicated or missing content and has a BUSCO complete score of 95.3%. The high contiguity and BUSCO scores are in line with other recently published sea urchin genomes (Davidson et al., 2020;Warner et al., 2021). Finally, we were able to identify 29,898 genes in this assembly.

A notable challenge in assembling this genome was the high heterozygosity (Figure 4.2). The estimated heterozygosity was 4.4% which is extremely high but consistent with the draft genome estimates (4.5%). These levels of heterozygosity are comparable to the *S. purpuratus* genome (4-5%) but are higher than the *L. variegatus* genome (2.9%). Prior to running PurgeHaplotigs, the percentage of core duplicated genes as predicted by BUSCO was 71.1%, indicating that regions of the genome where allelic variation exceeded CANU thresholds were assembled into multiple haplotigs rather than a single haplotype-fused representation of the genome. With the application of PurgeHaplotigs, we were able to reduce the duplication levels to 2.8%. Despite the higher heterozygosity, these duplication levels are within the ranges observed in the *Lytechinus* assemblies (0.6-4.76% in *L. variegatus* and *L. pictus*, respectively). As such, these results indicate that we have generated a haploid assembly.

Chemical defensome

Adaptive evolution can occur rapidly in species as they respond to new environments or as they encounter environmental challenges during range expansion (Chen et al., 2018). Gene family expansions offer a potential mechanism for rapid adaptation to novel environments (Kondrashov, 2012;Meerupati et al., 2013). The *E.* sp. EZ individual sampled for this genome is hypothesized to have undergone rapid adaptation considering that the PAG is a thermally extreme environment (warmest reef temperatures on the planet; Smith et al. (2017);Burt et al. (2019)) and the young age of the PAG (modern coastlines formed 3,000-6,000 years ago; Riegl and Purkis (2012)). Therefore, we hypothesized that key gene families underwent expansion events and would be evident in the *E.* sp. EZ genome when comparing to four other sea urchin species. We investigated several different gene families included in the chemical defensome; nuclear receptors, biotransformation enzymes, stress response, and transporter genes. Genome analyses of five species of sea urchin showed that the number of chemical defensome genes ranged from 811 in *E.* sp. EZ to 948 in *H. tuberculata* (Figure 4.3). The variability in the number of genes could be a result of missing annotations, however, this is likely to be a random process so is unlikely to only effect one subfamily of genes. Each gene subfamily is represented in each species and although the number of genes in each subfamily varies across species, they are relatively consistent, with the exception of the nuclear receptors. BLAST queries confirmed that the nuclear receptor ROR-alpha-like and the nuclear receptor subfamily 2 group E member 1 (*nr2e1*) genes are missing from the genome assembly. The NR ROR-alpha-like gene performs a diverse array of biological functions which includes regulation of glucose and inflammation cytokines, and free fatty acid metabolism (Zueva et al., 2018). *Nr2e1* is a gene involved in human retinal development and also regulates adult neural stem cell proliferation (O'Loghlen et al., 2015). Adaptive loss of function mutations or gene loss events are likely to occur more frequently than amino acid substitutions and have been documented in a wide variety of organisms (Borges et al., (2015);Wang et al., (2006)). It is currently unclear what functions *nr2e1* performs in sea urchins or why these two nuclear receptors are missing from *E.* sp. EZ's genome. Although we do not see evidence of gene expansion events in the defensome, it appears that gene loss may instead be driving adaptation. There may also be other variables linked to their thermal tolerance capacity such as sequence divergence, expression variation, or expansions in other untested gene families.

Positive selection

We tested whether there is evidence of selection acting on regulatory sequences associated with transcription factors associated with stress responses. Transcription factors regulate the expression of large suites of downstream genes and thus are potentially important targets of selection for adaptation. Previous studies of selection for transcription factors involved in developmental gene regulatory networks (GRNs) have revealed particular cis-regulatory regions under positive selection, which may explain differences in developmental pathways (Erkenbrack and Davidson, 2015). We used *adaptiPhy* to identify elevated rates of sequence divergence associated with noncoding sequences 25 kb upstream and downstream of 25 key transcription factors that are broadly involved in mediating responses to environmental variation. The divergence in these regions are compared to the background rates of sequence evolution across the phylogeny to identify branch-specific selection. Ten out of 25 genes were discarded during filtering due to a lack of long systemic fragments or possible duplications. *AdaptiPhy* returned 1,350 sites across the four species of sea urchins compared in this analysis. We detected evidence of selection in most of the remaining 10 genes but there were two genes that stood out as being under strong selective pressure exclusively in *E.* sp. EZ (Figure 4.4). The first gene is the nuclear receptor 1h6b (nr1h6b) whose biological function is not clearly defined, but is predicted to be involved in the regulation of biotransformation enzymes and transporters based on the sequence similarity with vertebrate *nr1h* subfamily (e.g., FXR, LXR) (Goldstone et al., 2006). The second of these genes, hypoxia inducible factor 1 subunit alpha (*hif1a*), is a transcription factor that serves as a gene expression modulator in response to hypoxia (Wenger, 2002). In the sea urchin *Paracentrotus lividus*, HIF1A is hypothesized to be a target of diatom polyunsaturated aldehydes (PUAs) which negatively affect reproduction and development of benthic organisms who feed on these algae (Varrella et al., 2016). In the sea anemone *Exaiptasia pallida, hif1a* expression changes rapidly and significantly early on in the heat stress response (Cleves et al., 2020). A link between temperature and HIF-1 activity has also been documented in a species of carp (Rissanen et al., 2006) and Atlantic salmon (Olsvik et al., 2013) where temperature stress resulted in impaired binding activity of HIF-1. These findings point towards a critical relationship between *hif1a* and the stress response. Further, our observation of selection in the noncoding sequences neighboring *hif1a* could be indicative of altered expression of this gene in *E.* sp. EZ, which in turn could potentially impact stress response genes under the regulatory control of this transcription factor.

Echinometra is a widespread genus of sea urchin that play important ecological roles in tropical benthic communities. This is the first chromosome-level genome assembly for a species from this genus and represents a crucial advance with implications across multiple fields. Firstly, it will allow comparisons across multiple sea urchin genera to elucidate fundamental evolutionary patterns. Secondly, the completeness of the genome will further ongoing research investigating the role of adaptation in these urchins, particularly with respect to identifying structural variation across ecological gradients.

Lastly, the *Echinometra* genus is a classically studied example of both allopatric and sympatric speciation and therefore this new genome assembly will complement past studies and may help elucidate genomic underpinnings of speciation.

Acknowledgements

The fieldwork was carried out using the Core Technology Platform resources at New York University Abu Dhabi. We thank the Environmental Agency – Abu Dhabi for permits.

References

A., M. M., G., K., & A., L. H. (2000). Dispersal barriers in tropical oceans and speciation in Atlantic and eastern Pacific sea urchins of the genus *Echinometra*. *Molecular Ecology, 9*(9), 1391-1400. doi:doi:10.1046/j.1365-294x.2000.01022.x

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology, 215*(3), 403-410.

Aslan, L. M., & Uehara, T. (1997). Hybridization and F1 backcrosses between two closely related tropical species of sea urchins (genus *Echinometra*) in Okinawa. *Invertebrate Reproduction & Development, 31*(1-3), 319-324. doi:10.1080/07924259.1997.9672591

Berrio, A., Haygood, R., & Wray, G. A. (2020). Identifying branch-specific positive selection throughout the regulatory genome using an appropriate proxy neutral. *BMC Genomics, 21*, 1-16.

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

Borges, R., Khan, I., Johnson, W. E., Gilbert, M. T. P., Zhang, G., Jarvis, E. D., ... & Antunes, A. (2015). Gene loss, adaptive evolution and the co-evolution of plumage coloration genes with opsins in birds. *BMC Genomics*, *16*(1), 1-14.

Bronstein, O., & Loya, Y. (2013). The taxonomy and phylogeny of *Echinometra* (Camarodonta: Echinometridae) from the Red Sea and Western Indian Ocean. *PLOS One, 8*(10), e77374. doi:10.1371/journal.pone.0077374

Burt, J. A., Paparella, F., Al-Mansoori, N., Al-Mansoori, A., & Al-Jailani, H. (2019). Causes and consequences of the 2017 coral bleaching event in the southern Persian/Arabian Gulf. *Coral Reefs, 38*(4), 567-589.

Cameron, R. A., Samanta, M., Yuan, A., He, D., & Davidson, E. (2009). SpBase: the sea urchin genome database and web site. *Nucleic Acids Research, 37*(suppl_1), D750-D754.

Chen, Y., Shenkar, N., Ni, P., Lin, Y., Li, S., & Zhan, A. (2018). Rapid microevolution during recent range expansion to harsh environments. *BMC Evolutionary Biology, 18*(1), 1-13.

Cleves, P. A., Krediet, C. J., Lehnert, E. M., Onishi, M., & Pringle, J. R. (2020). Insights into coral bleaching under heat stress from analysis of gene expression in a sea anemone model system. *Proceedings of the National Academy of Sciences, 117*(46), 28906-28917.

Davidson, P. L., Guo, H., Wang, L., Berrio, A., Zhang, H., Chang, Y., . . . Wray, G. A. (2020). Chromosomal-level genome assembly of the sea urchin *Lytechinus variegatus* substantially improves functional genomic analyses. *Genome Biology and Evolution*. doi:10.1093/gbe/evaa101

Eide, M., Zhang, X., Karlsen, O. A., Goldstone, J. V., Stegeman, J., Jonassen, I., & Goksøyr, A. (2021). The chemical defensome of five model teleost fish. *Scientific Reports, 11*(1), 1-13.

Erkenbrack, E. M., & Davidson, E. H. (2015). Evolutionary rewiring of gene regulatory network linkages at divergence of the echinoid subclasses. *Proceedings of the National Academy of Sciences, 112*(30), E4075-E4084.

Finn, R. D., Clements, J., & Eddy, S. R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research, 39*(suppl_2), W29-W37.

Goldstone, J., Hamdoun, A., Cole, B., Howard-Ashby, M., Nebert, D., Scally, M., . . . Stegeman, J. (2006). The chemical defensome: environmental sensing and response genes in the *Strongylocentrotus purpuratus* genome. *Developmental biology, 300*(1), 366-384.

Hall, M. R., Kocot, K. M., Baughman, K. W., Fernandez-Valverde, S. L., Gauthier, M. E., Hatleberg, W. L., . . . Shoguchi, E. (2017). The crown-of-thorns starfish genome as a guide for biocontrol of this coral reef pest. *Nature, 544*(7649), 231-234.

Hoff, K. J., Lomsadze, A., Stanke, M., & Borodovsky, M. (2018). BRAKER2: incorporating protein homology information into gene prediction with GeneMark-EP and AUGUSTUS. *Plant and Animal Genomes XXVI.*

Hubisz, M. J., Pollard, K. S., & Siepel, A. (2011). PHAST and RPHAST: phylogenetic analysis with space/time models. *Briefings in Bioinformatics, 12*(1), 41-51.

Long KA, Nossa CW, Sewell MA, Putnam NH, & Ryan JF. Low coverage sequencing of three echinoderm genomes: the brittle star *Ophionereis fasciata*, the sea star *Patiriella regularis*, and the sea cucumber *Australostichopus mollis*. GigaScience. 2016 Dec 1;5(1):s13742-016.

Janies, D. A., Witter, Z., Linchangco, G. V., Foltz, D. W., Miller, A. K., Kerr, A. M., . . . Wray, G. A. (2016). EchinoDB, an application for comparative transcriptomics of deeply-sampled clades of echinoderms. *BMC Bioinformatics, 17*(1), 1-6.

Ketchum, R. N., DeBiasse, M. B., Ryan, J. F., Burt, J. A., & Reitzel, A. M. (2018). The complete mitochondrial genome of the sea urchin, *Echinometra* sp. *EZ*. *Mitochondrial DNA Part B, 3*(2), 1225-1227.

Ketchum, R. N., Smith, E. G., DeBiasse, M. B., Vaughan, G. O., McParland, D., Leach, W. B., . . . Reitzel, A. M. (2020). Population genomic analyses of the sea urchin *Echinometra* sp. *EZ* across an extreme environmental gradient. *Genome Biology and Evolution*.

Ketchum, R. N., Smith, E. G., Vaughan, G. O., Phippen, B. L., McParland, D., Al Mansoori, N., . . . Reitzel, A. M. (2018). DNA extraction method plays a significant role when defining bacterial community composition in the marine invertebrate *Echinometra mathaei*. *Frontiers in Marine Science, 5*, 255.

Kinjo, S., Kiyomoto, M., Yamamoto, T., Ikeo, K., & Yaguchi, S. (2018). HpBase: A genome database of a sea urchin, *Hemicentrotus pulcherrimus*. *Development, Growth & Differentiation, 60*(3), 174-182.

Kondrashov, F. A. (2012). Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proceedings of the Royal Society B: Biological Sciences, 279*(1749), 5048-5057.

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., & Phillippy, A. M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Research, 27*(5), 722-736.

Kudtarkar, P., & Cameron, R. A. (2017). Echinobase: an expanding resource for echinoderm genomic information. *Database, 2017*.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997*.

Lieberman-Aiden, E., Van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., . . . Dorschner, M. O. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science, 326*(5950), 289-293.

Lomsadze, A., Ter-Hovhannisyan, V., Chernoff, Y. O., & Borodovsky, M. (2005). Gene identification in novel eukaryotic genomes by self-training algorithm. *Nucleic Acids Research, 33*(20), 6494-6506.

Manni, M., Berkeley, M. R., Seppey, M., Simao, F. A., & Zdobnov, E. M. (2021). BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *arXiv preprint arXiv:2106.11799*.

Marks, P., Garcia, S., Barrio, A. M., Belhocine, K., Bernate, J., Bharadwaj, R., . . . Fehr, A. (2019). Resolving the full spectrum of human genome variation using Linked-Reads. *Genome Research, 29*(4), 635-645.

Matsuoka, N., & Toshihiko, H. (1991). Molecular evidence for the existence of four sibling species with the sea-urchin, *Echinometra mathaei* in Japanese waters and their evolutionary relationships. *Zoological Science, 8*(1), 121-133.

McCartney, M. A., & Lessios, H. A. (2004). Adaptive evolution of sperm bindin tracks egg incompatibility in neotropical sea urchins of the genus *Echinometra*. *Molecular Biology and Evolution, 21*(4), 732-745.

McClanahan, T. R., & Muthiga, N. A. (2007). Chapter 15 Ecology of *Echinometra*. In J. M. Lawrence (Ed.), *Developments in Aquaculture and Fisheries Science* (Vol. 37, pp. 297-317): Elsevier.

Meerupati, T., Andersson, K.-M., Friman, E., Kumar, D., Tunlid, A., & Ahren, D. (2013). Genomic mechanisms accounting for the adaptation to parasitism in nematode-trapping fungi. *PLOS Genetics, 9*(11), e1003909.

Metz, E. C., Kane, R. E., Yanagimachi, H., & Palumbi, S. R. (1994). Fertilization between closely related sea urchins is blocked by incompatibilities during sperm-egg attachment and early stages of fusion. *The Biological Bulletin, 187*(1), 23-34. doi:10.2307/1542162

Mita, M., Uehara, T., & Nakamura, M. (2004). Speciation in four closely related species of sea urchins (genus *Echinometra*) with special reference to the acrosome reaction. *Invertebrate Reproduction & Development, 45*(3), 169-174. doi:10.1080/07924259.2004.9652588

O'Loghlen, A., Martin, N., Krusche, B., Pemberton, H., Alonso, M. M., Chandler, H., . . . Gil, J. (2015). The nuclear receptor NR2E1/TLX controls senescence. *Oncogene, 34*(31), 4069-4077.

Olsvik, P. A., Vikeså, V., Lie, K. K., & Hevrøy, E. M. (2013). Transcriptional responses to temperature and low oxygen stress in Atlantic salmon studied with next-generation sequencing technology. *BMC Genomics, 14*(1), 1-21.

Palumbi, S. R., Grabowsky, G., Duda, T., Geyer, L., & Tachino, N. (1997). Speciation and population genetic structure in tropical Pacific sea urchins. *Evolution, 51*(5), 1506-1517. doi:doi:10.1111/j.1558-5646.1997.tb01474.x

Putnam, N. H., O'Connell, B. L., Stites, J. C., Rice, B. J., Blanchette, M., Calef, R., . . . Sugnet, C. W. (2016). Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Research, 26*(3), 342-350.

Rahman, M. A., Uehara, T., & Aslan, L. M. (2000). Comparative viability and growth of hybrids between two sympatric species of sea urchins (Genus *Echinometra*) in Okinawa. *Aquaculture, 183*(1), 45-56. doi:https://doi.org/10.1016/S0044-8486(99)00283-5

Rahman, M. A., Uehara, T., & Pearse, J. S. (2001). Hybrids of two closely related tropical sea urchins (genus *Echinometra*): evidence against postzygotic isolating mechanisms. *The Biological Bulletin, 200*(2), 97-106. doi:10.2307/1543303

Riegl, B. M., & Purkis, S. J. (2012). Coral reefs of the Gulf: adaptation to climatic extremes in the world’s hottest sea. In *Coral reefs of the Gulf* (pp. 1-4): Springer.

Rissanen, E., Tranberg, H. K., Sollid, J., Nilsson, G. r. E., & Nikinmaa, M. (2006). Temperature regulates hypoxia-inducible factor-1 (HIF-1) in a poikilothermic vertebrate, crucian carp (*Carassius carassius*). *Journal of experimental biology, 209*(6), 994-1003.

Roach, M. J., Schmidt, S. A., & Borneman, A. R. (2018). Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics, 19*(1), 1-10.

Smit, A., & Hubley, R. (2008-2015). RepeatModeler Open-1.0. Retrieved from http://www.repeatmasker.org

Smit, A., Hubley, R., & Green, P. (2013-2015). RepeatMasker Open-4.0. Retrieved from http://www.repeatmasker.org

Smith, E. G., Vaughan, G. O., Ketchum, R. N., McParland, D., & Burt, J. A. (2017). Symbiont community stability through severe coral bleaching in a thermally extreme lagoon. *Scientific Reports, 7*(1), 2428. doi:10.1038/s41598-017-01569-8

Sodergren, E., Weinstock, G. M., Davidson, E. H., Cameron, R. A., Gibbs, R. A., Angerer, R. C., . . . Burke, R. D. (2006). The genome of the sea urchin *Strongylocentrotus purpuratus*. *Science, 314*(5801), 941-952.

Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., & Morgenstern, B. (2006). AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Research, 34*(suppl_2), W435-W439.

Uehara, T., Shingaki, M., Taira, K., Arakaki, Y., & Nakatomi, H. (2020). Chromosome studies in eleven Okinawan species of sea urchins, with special reference to four species of the Indo-Pacific *Echinometra*. In *Biology of echinodermata* (pp. 119-129): CRC Press.

Varrella, S., Romano, G., Costantini, S., Ruocco, N., Ianora, A., Bentley, M. G., & Costantini, M. (2016). Toxic diatom aldehydes affect defence gene networks in sea urchins. *PLOS One, 11*(2), e0149734.

Vurture, G. W., Sedlazeck, F. J., Nattestad, M., Underwood, C. J., Fang, H., Gurtowski, J., & Schatz, M. C. (2017). GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics, 33*(14), 2202-2204.

Wang, X., Grus, W. E., & Zhang, J. (2006). Gene losses during human origins. *PLoS Biology*, *4*(3), e52.

Warner, J. F., Lord, J. W., Schreiter, S. A., Nesbit, K. T., Hamdoun, A., & Lyons, D. C. (2021). Chromosomal-Level Genome Assembly of the Painted Sea Urchin

Lytechinus pictus: A Genetically Enabled Model System for Cell Biology and Embryonic Development. *Genome Biology and Evolution, 13*(4), evab061.

Wenger, R. H. (2002). Cellular adaptation to hypoxia: O2-sensing protein hydroxylases, hypoxia-inducible transcription factors, and O2-regulated gene expression. *The FASEB journal, 16*(10), 1151-1162.

Zueva, K. J., Lumme, J., Veselov, A. E., Kent, M. P., & Primmer, C. R. (2018). Genomic signatures of parasite-driven natural selection in north European Atlantic salmon (*Salmo salar*). *Marine Genomics, 39*, 26-38.

Table 4.1: Comparison of the published *E*. sp. EZ draft genome, the assembly prior to running HiRise, and the final chromosome-level genome assembly.

	Draft Genome (Ketchum et al., 2020)	Genome assembly prior to HiRise	Chromosome-level Genome
Assembly Size	1,589 Mb	817.5 Mb	817.8 Mb
No. scaffolds	4,487,317	3,800	210
N50 scaffold length	1,006 bp	352,130 bp	39.5 Mb
Longest scaffold	66,286 bp	1.9 Mb	65.8 Mb
BUSCO Complete	34.8%	94.44%	95.28%
BUSCO Duplicated	6.7%	4.0%	2.8%
BUSCO Fragmented	45.5%	2.8%	2.0%
BUSCO Missing	19.7%	2.7%	2.7%

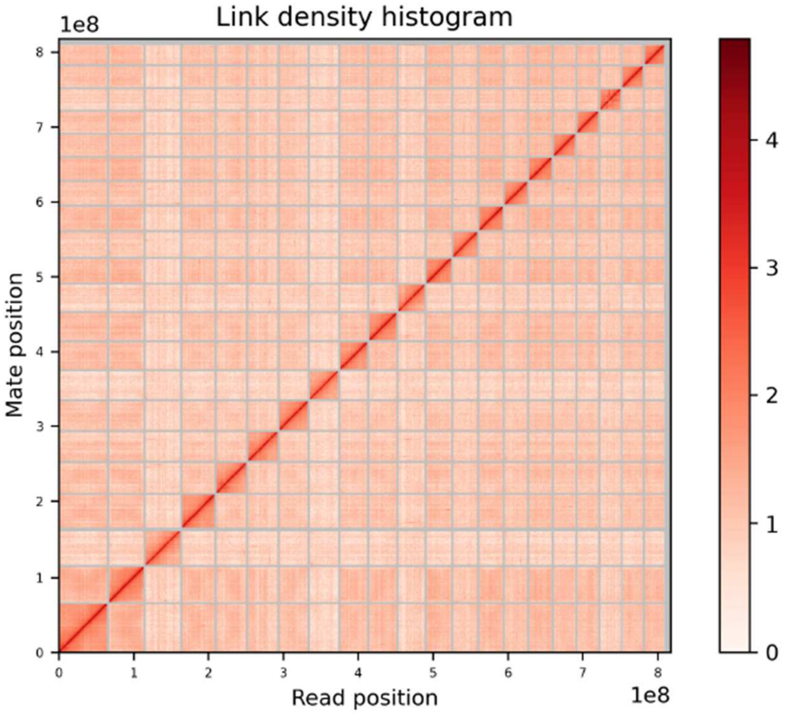


Figure 4.1: Link density histogram mapped with Hi-C reads for the *E*. sp. *EZ* chromosome-level assembly. The x and y axes represent the mapping positions of the first and second read, respectively, grouped into bins. The color of the squares represents the density of read pairs within that bin.

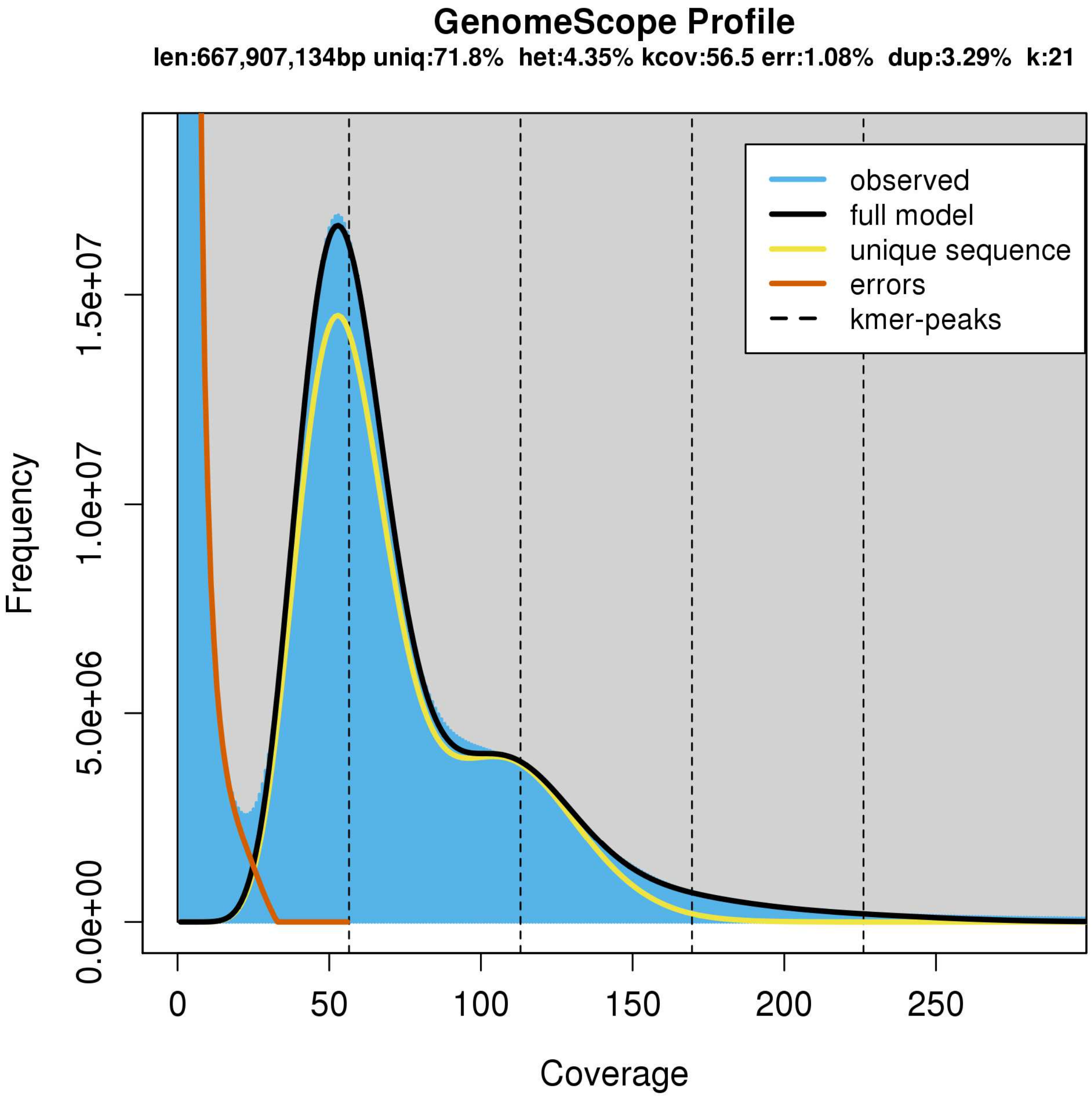


Figure 4.2: GenomeScope (Vurture et al., 2017) profile of *E*. sp. *EZ* based on the chromosome-level assembly and a kmer length of 21. This plot was generated from linked-read sequencing (10x Genomics). The sample was sequenced on a partial lane (472M reads/1.5Gb of genome) on the NovaSeq 6000 sequences (Illumina, San Diego, CA) with paired-end 150 bp reads.

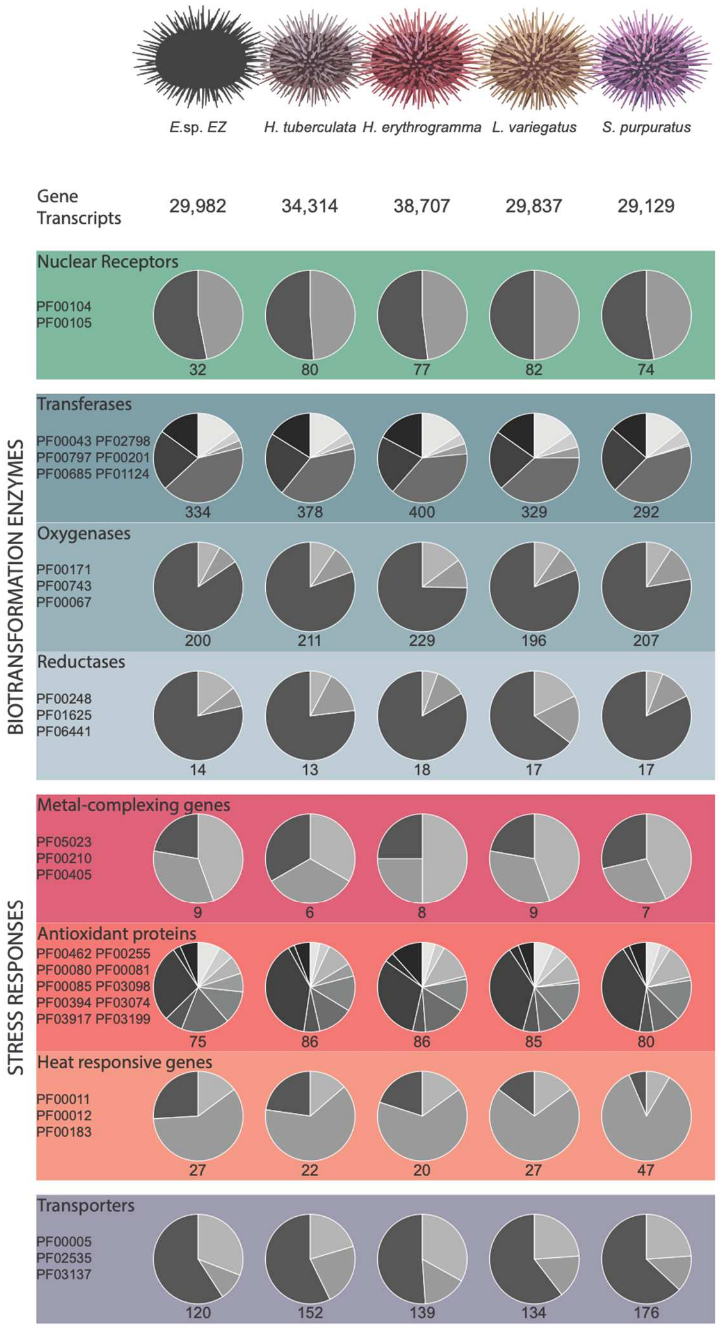


Figure 4.3: Chemical defensome genes in five species of sea urchin: *E.* sp. EZ, *L. variegatus*, *S. purpuratus*, *H. erythrogramma*, and *H. tuberculata*. Genes were identified by using HMMER searches with PFAM profiles. The gene families were then grouped by their role in the chemical defensome. The colors in the disk represents the PFAM ID and the size of the disk slice represents the number of genes in each respective gene family group. The total number of genes in each group family is represented underneath the disk. PFAM groups included in this analysis were from Goldstone et al. (2006) and the grouping of PFAMs into categories followed Eide et al. (2021).

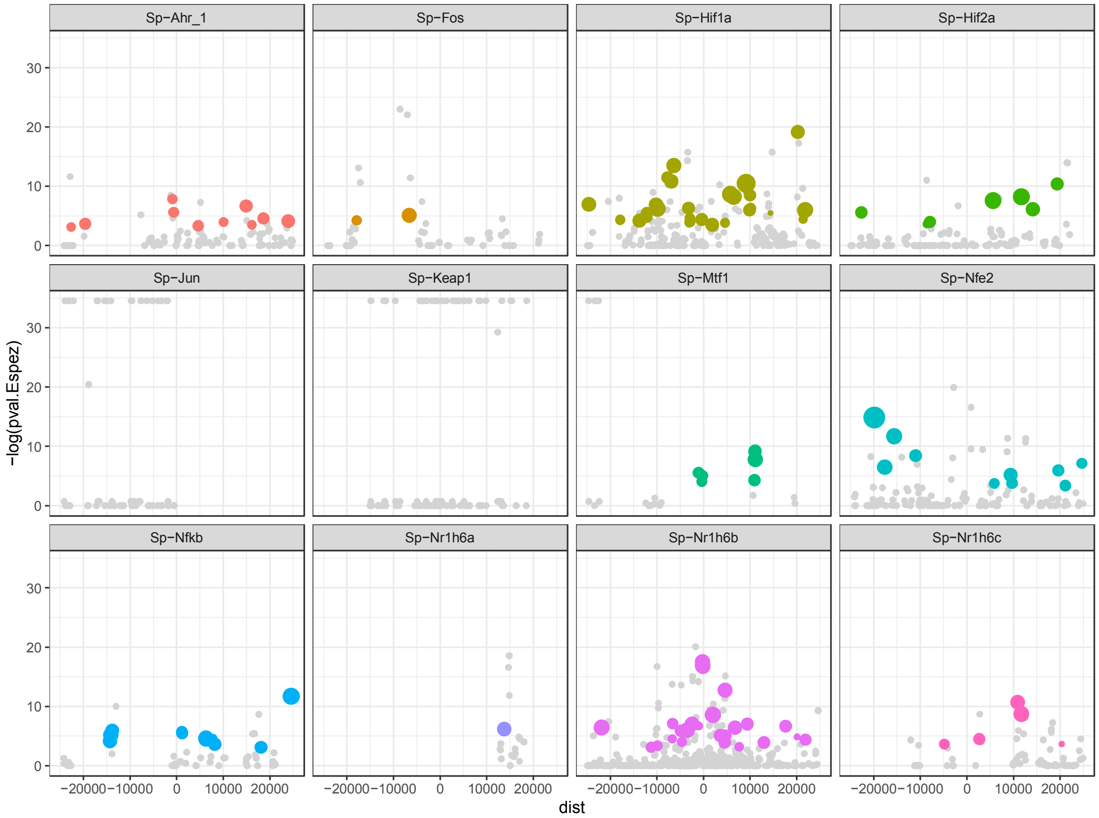


Figure 4.4: Positive selection results based on comparing *E.* sp. EZ to *H. erythrogramma* and *H. tuberculata* with *L. variegatus* as the outgroup. The plots presented includes sites that show evidence of positive selection exclusively in *E.* sp. EZ genes. The title of each individual plot is the corresponding gene name in the *S. purpuratus* genome and the *E.* sp. EZ split gene models have been merged for visualization purposes. The x axis is the distance of the site to the transcription start site of the gene (negative is upstream and positive is downstream) and the y axis is the -log(pval). The size of the colored circles is proportional to the zeta score (ratio of substitution rate of query to neutral substitution rate).

OVERALL CONCLUSIONS

This dissertation uses an integrative approach to better understand the mechanisms of thermal adaptation and acclimation of the sea urchin *Echinometra* sp. *EZ* along the Arabian Peninsula. The Persian/Arabian Gulf is extreme in both temperature and salinity while the adjacent Gulf of Oman has conditions more typical of tropical oceans. The first two chapters explore the relationship between *E*. sp. *EZ*'s gut microbiome across methodologies, reefs, seasons, and seas. The third chapter investigates *E.* sp. *EZ*'s population genetics and mines the genome for signatures of thermal selection. The fourth chapter focuses on describing the chromosome-level genome and performing comparative genomics between five species of sea urchin. Cumulatively, I show the importance of thermal environment across multiple levels of biological organization, provide useful tools for the invertebrate genomics community, and are able to make predictions about the future of *E*. sp. *EZ* under ongoing climate change.

In Chapter One, I showed that the addition of a bead-beating and lysozyme step during DNA extractions would more effectively capture difficult to lyse bacterial taxa, such as gram-positive bacteria. Further, I generated a literature synthesis of the different methodologies currently being used in the field of microbiology and highlight that the inclusion of a lysozyme treatment is uncommon, despite the importance of this step. Studies aimed at characterizing microbial diversity have increased exponentially in the last decade and it is therefore increasingly important to develop a 'gold standard' methodological approach that best represents the bacterial communities present.

In Chapter Two, I use the aforementioned approach to elucidate the role of temperature on the microbial gut community of *E*. sp. *EZ* collected across a wide geographic and temporal range. I show that microbial communities vary across thermally variable habitats and seasons and can become more disperse as temperatures rise. Importantly, I highlight several ASVs whose relative abundances correlate with temperature in both independent datasets and largely belonged to the family Vibrionaceae. This may represent a mechanism of acclimatization through restructuring of the microbiome, although further studies are warranted to explore the functional role of these taxa in their hosts.

In Chapter Three, I describe the population structure and dynamics of *E*. sp. *EZ* along the Arabian Peninsula. Population structure analysis revealed two main populations in each respective sea with a high degree of admixture between all sites. Given the young age of the PAG, it is surprising that enough evolutionary time has passed for population structure to develop. It is possible that the strength of thermal selection has driven this divergence. Finally, I show evidence for selection on genes related to collagen production. This work is an important contribution to our understanding of the complex environmental variables that drive genomic divergence.

In Chapter Four, I annotate and describe the chromosome-level genome of *E.* sp. *EZ*. I perform PFAM analysis to identify potential expansions and contractions across five sea urchin species and investigate noncoding regions associated with key defensome genes for evidence of positive selection. I did not find evidence of any gene expansion or contraction events in these urchins. However, I did find that two nuclear receptor genes were missing in the *E.* sp. EZ genome. Further, I found that several noncoding regions neighboring defensome genes show signatures of strong positive selection exclusively in the *E.* sp. EZ genome. One of these is the hypoxia inducible factor 1 subunit alpha which is a transcription factor that serves as a gene expression modulator in response to hypoxia and has a temperature-dependent relationship in other marine species. This genome will facilitate evolutionary comparisons across an increased breadth of taxonomic representation and help answer fundamental questions in the field of speciation biology.

The Arabian Peninsula is a natural laboratory for studying species' response to climate change as reef temperatures currently surpass climate change predictions for the Indo-Pacific in the next century. Sea urchins play a critical role in coral reef ecosystems and while previous work has focused on speciation dynamics in the *Echinometra* genus, very little is known about within species population connectivity, adaptation, and microbial communities of these keystone organisms. By leveraging this unique study system and ecologically important species, I provide new genomic insights into high gene flow marine organisms and through the use of cutting-edge analytical tools, I demonstrate the utility of robust statistical approaches in identifying microbial community response to environmental variables. Taken together, this dissertation represents several key advances in our understanding of acclimation and adaptation across extreme environmental gradients.