GENOMIC INSIGHT INTO THE GUT MICROBIOME OF THE SEA URCHINS *LYTECHINUS VARIEGATUS* AND *STRONGYLOCENTROTUS PURPURATUS* REVEALED DISTINCT COMMUNITY COMPOSITIONS AND THEIR METABOLIC PROFILES

by

JOSEPH ANTOINE HAKIM

ASIM K. BEJ, COMMITTEE CHAIR
JASON G. LINVILLE
CASEY D. MORROW
STEPHEN A. WATTS
THANE R. WIBBELS

A DISSERTATION

Submitted to the graduate faculty of The University of Alabama at Birmingham,
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Biology

BIRMINGHAM, ALABAMA

2019

GENOMIC INSIGHT INTO THE GUT MICROBIOME OF THE SEA URCHINS *LYTECHINUS VARIEGATUS* AND *STRONGYLOCENTROTUS PURPURATUS* REVEALED DISTINCT COMMUNITY COMPOSITIONS AND THEIR METABOLIC PROFILES

JOSEPH ANTOINE HAKIM

BIOLOGY

ABSTRACT

For over 500 million years, our planet's self-replicating prokaryotes have colonized multicellular host organisms, forging complex interdependent relationships under the selective pressures of the natural environment. Recently, the use of high-throughput sequencing (HTS) technology targeting metacommunity DNA has uncovered the unprecedented diversity and metabolic processes of these communities. Such information has extensively been investigated in the guts of higher bilaterian organisms, such as mammals including humans. One of the early bilaterian organisms that have been linked to humans are the sea urchins, the gut microbiota of which have not been sufficiently studied. Thus, the objective of this dissertation was to explore the structure and function of the microbiota in the gut systems of two sea urchins, *Lytechinus variegatus* (green) and *Strongylocentrotus purpuratus* (purple), which are dominant grazers of marine seagrass and algae in their natural coastal marine habitats. These organisms possess primitive deuterostome gut systems whereupon ingestion, the mucous-producing cells in the pharynx envelop their feed and microbiota forming "gut digesta" that remains separate from the "gut

tissue” while transiting through the gut lumen. In this study, we have used metagenomics and bioinformatics tools to reveal the microbial communities in the gut tissue and gut digesta in the green and purple sea urchins. The results showed unique abundances of Epsilonproteobacteria, primarily representing *Arcobacter* in the gut tissue of the green, whereas *Sulfurimonas* and *Arcobacter* were observed in the purple sea urchin. However, the gut digesta was dominated by Gammaproteobacteria representing *Vibrio* in the green, and *Psychromonas* in the purple sea urchins. Functional metagenomics analysis further supported the gut digesta as the primary location for the microbial metabolism of macromolecules in both green and purple sea urchins. Moreover, genes in the reduction and fixation of nitrogen into ammonia, and subsequent assimilation into organic molecules primarily for the synthesis of nucleotides and amino acids were abundant in the metagenomes. The results of this study revealed the distinct gut microbial community composition and metabolic processes benefitting two ecologically and evolutionarily significant marine invertebrates, and their potential impact at various trophic levels of their natural habitats.

Keywords: Echinoderm, Deuterostome, High-Throughput Sequencing, Gut Digesta, 16S rRNA, Shotgun Metagenomics

DEDICATION

It is with the utmost sincerity that I dedicate this dissertation to my family. Thank you for believing in me, and for all of your support and encouragement. To my dad, Antoine, and my mom, Katherine, I always know to look to you for a healthy dose of advice, support, and reassurance. Being away during my graduate study was difficult, but, because of you, I never felt like I was alone. To my older brother, Steve, your encouragement and advice got me through some of the most difficult times, and to my younger brother, Tony, your unwavering optimism was always welcomed. Thank you both for making the time to write, play, and record all songs that kept me smiling these past few years.

In addition to my family, I extend this dedication to my friends and loved ones. To those I have had all along, and to those I met during this graduate study, thank you for sticking it out with me. Thank you for your empathy, for listening to all of my aches, pains, and frustrations, and importantly, thank you for matching my enthusiasm whenever the experiments or analyses would go as planned. Thank you for every thoughtful suggestion, proofread, and comment to all my ideas, drafts, and theories. I am forever indebted to you all, for every selfless hour of your time spent on me. Thank you to Matthew Whitehurst, for being an incredible friend, bassist, and someone to humor my crazy ideas. Finally, to Dr. Hyunmin Koo, M.S., Ph.D., who I was fortunate to share this lab space with, thank you for all of the memories and the amazing collaborations.

ACKNOWLEDGMENTS

I would like to thank my mentor Dr. Asim K. Bej for his guidance, support, and insight throughout my dissertation study. I would also like to thank my committee members, Drs. Stephen A. Watts, Casey D. Morrow, Thane R. Wibbels, and Jason G. Linville for all their support and advice during my dissertation study.

My dissertation progress would not have been possible without the UAB Department of Biology for their graduate teaching assistantship support; the UAB School of Medicine grant support from Dr. Casey D. Morrow of the Department of Cell, Developmental and Integrative Biology towards my graduate research assistantship; the UAB Heflin Center for Genomic Sciences and UAB Microbiome Resource Core facility for all high-throughput sequencing; and the UAB Cheaha HPC and HTC grid for data analysis support.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table *Page*

CHAPTER V: SHOTGUN METAGENOMICS REVEALED DIFFERENCES IN THE MICROBIOTA WITH KEY METABOLIC ATTRIBUTES EMPHASIZING NITROGEN FIXATION IN SEA URCHIN GUT DIGESTA

CHAPTER V: SHOTGUN METAGENOMICS REVEALED DIFFERENCES IN THE MICROBIOTA WITH KEY METABOLIC ATTRIBUTES EMPHASIZING NITROGEN FIXATION IN SEA URCHIN GUT DIGESTA

LIST OF FIGURES

Figures *Page*

CHAPTER I: GENERAL INTRODUCTION

CHAPTER II: THE GUT MICROBIOME OF THE SEA URCHIN, *LYTECHINUS VARIEGATUS*, FROM ITS NATURAL HABITAT DEMONSTRATES SELECTIVE ATTRIBUTES OF MICROBIAL TAXA AND PREDICTIVE METABOLIC PROFILES

CHAPTER III: THE PURPLE SEA URCHIN *STRONGYLOCENTROTUS PURPURATUS* DEMONSTRATES A COMPARTMENTALIZATION OF GUT BACTERIAL MICROBIOTA, PREDICTIVE FUNCTIONAL ATTRIBUTES, AND TAXONOMIC CO-OCCURRENCE

CHAPTER IV: COMPARISON OF GUT MICROBIOTA IN NATURALLY OCCURRING AND LABORATORY AQUACULTURE SEA URCHIN *LYTECHINUS VARIEGATUS* REVEALED DIFFERENCES IN THE COMMUNITY COMPOSITION AND PREDICTED FUNCTIONS

CHAPTER V: SHOTGUN METAGENOMICS REVEALED DIFFERENCES IN THE MICROBIOTA WITH KEY METABOLIC ATTRIBUTES EMPHASIZING NITROGEN FIXATION IN SEA URCHIN GUT DIGESTA

CHAPTER V: SHOTGUN METAGENOMICS REVEALED DIFFERENCES IN THE MICROBIOTA WITH KEY METABOLIC ATTRIBUTES EMPHASIZING NITROGEN FIXATION IN SEA URCHIN GUT DIGESTA

LIST OF ABBREVIATIONS

ANOSIM	Analysis of Similarity
ASV	Amplicon Sequence Variant
bp	Base Pair
BLAST	Basic Local Alignment Search Tool
CoNet	Co-Occurrence Network
DADA2	Divisive Amplicon Denoising Algorithm 2
DNA	Deoxyribonucleic Acid
dNTP	Deoxyribonucleotide Triphosphate
EMBL	European Molecular Biology Laboratory
FDR	False Discovery Rate
Ga	Giga-annum
GG	Greengenes Database
KEGG	Kyoto Encyclopedia of Genes and Genomes
KO	Kyoto Encyclopedia of Genes and Genomes Orthology
kPa	Kilopascal
MDS	Multi-Dimensional Scaling
MG-RAST	Metagenomics RAST server
NCBI	National Center for Biotechnology Information
NGS	Next Generation Sequencing
HTS	High-Throughput Sequencing

LCA	Least Common Ancestor
LDA	Linear discriminant analysis
LEfSe	Linear discriminant analysis Effect Size
LTP	All-Species Living Tree Project
Ma	Mega-annum
NSTI	Nearest Sequenced Taxon Index
nt	Nucleotide
OTU	Operational Taxonomical Unit
PBS	Phosphate-Buffered Saline
PCoA	Principal Coordinates Analysis
PCR	Polymerase Chain Reaction
PhyloToAST	Phylogenetic Tools for Analysis of Species-level Taxa
PICRUSt	Phylogenetic Investigation of Communities by Reconstruction of Unobserved States
QIIME	Quantitative Insights into Microbial Ecology
RAST	Rapid Annotations using Subsystem Technology
RDP	Ribosomal Database Project
RNA	Ribonucleic Acid
rRNA	Ribosomal Ribonucleic Acid
RTL	Research and Testing Laboratories
SD	Standard Deviation
SILVA-ACT	SILVA Alignment, Classification and Tree Service
SRA	Sequence Read Archive

STAMP	Statistical Analysis of Metagenomic Profiles
UV	Ultraviolet
V4	16S rRNA hypervariable region 4
WGS	Whole Genome Sequencing

CHAPTER I: GENERAL INTRODUCTION

Prologue: The Co-Evolved Invisible Community

The accounts and illustrations by Charles Darwin in his book, *On the Origin of Species*, offered insights into the diversity of life through natural selection (Darwin, 1859). Such diversity often constitutes an interconnected network of interactive organisms which co-evolved during the planet's geologic time scale (Ehrlich & Raven, 1964, Nisbet & Sleep, 2001). Although the chronology of this process has widely been observed in the fossil records of *macro*-organisms, another group of living microscopic organisms has also been closely integrated within this network, playing a crucial role throughout the co-evolutionary processes (Knoll, 2015). These microscopic organisms represent primarily prokaryotic microorganisms, including bacteria that thrived for ~4.2 Ga. often referred to as the "unseen majority" (Whitman *et al.*, 1998, Nisbet & Sleep, 2001). The prokaryotes are single-celled and self-replicating organisms, encompassing approximately 4-6 × 10^{30} cells, or 2-5 orders of magnitude more than the total number plant and animal cells combined (Whitman *et al.*, 1998, Griffin, 2007, Kallmeyer *et al.*, 2012, Charlop-Powers *et al.*, 2014). They often assemble into diverse communities that function interdependently within their niche, performing critical metabolisms and participate in the biogeochemical cycles at the local and global scale. These microorganisms have been associated with multicellular eukaryotic organisms for over 500 million years, a

relationship that was forged under the selective pressures of the natural environment (Bäckhed *et al.*, 2005, Cho & Blaser, 2012, Shade & Handelsman, 2012, Amato, 2013, Blekhman *et al.*, 2015, Rosshart *et al.*, 2017, Limborg & Heeb, 2018). Together, the microbiota and their host collectively referred to as the “holobiont” (Zilber-Rosenberg & Rosenberg, 2008, Rosenberg & Zilber-Rosenberg, 2018), have been suggested to influence aspects of co-evolutionary processes and diversification (Ley *et al.*, 2008, Limborg & Heeb, 2018). The recent advancement of powerful genomics techniques has made it possible to better understand these microbial communities at the highest possible resolution and within an evolutionary context. In addition, such genomics technologies have provided an outlook of the fundamental processes such as barrier maintenance, pathogen protection, and vitamin and nutrient supply occurring in extant organisms (Figure 1-1), including soft-bodied organisms that radiated in the Pre-Cambrian era to the later bilaterians with distinct mouth-to-anus digestive tract morphologies (Dishaw *et al.*, 2014, Thursby & Juge, 2017). Of the bilaterians, the deuterostomes represent an evolutionary and developmental milestone that included echinoderms such as the sea urchins grouped with higher chordates (McClay, 2011). As compared to the more complex gut system of higher chordates such as humans, the sea urchins, which emerged on our planet ~ 450 Ma., would maintain relatively simple gut systems in a straightforward model (McClay, 2011). By investigating the gut microbial communities of the sea urchin, and particularly those occurring in their natural habitat, the fundamental metabolic processes occurring by microbial communities in a “primitive” gut

system can be evaluated, as they relate to the nutritional need of these evolutionarily significant organisms and perhaps transcending to higher chordates.

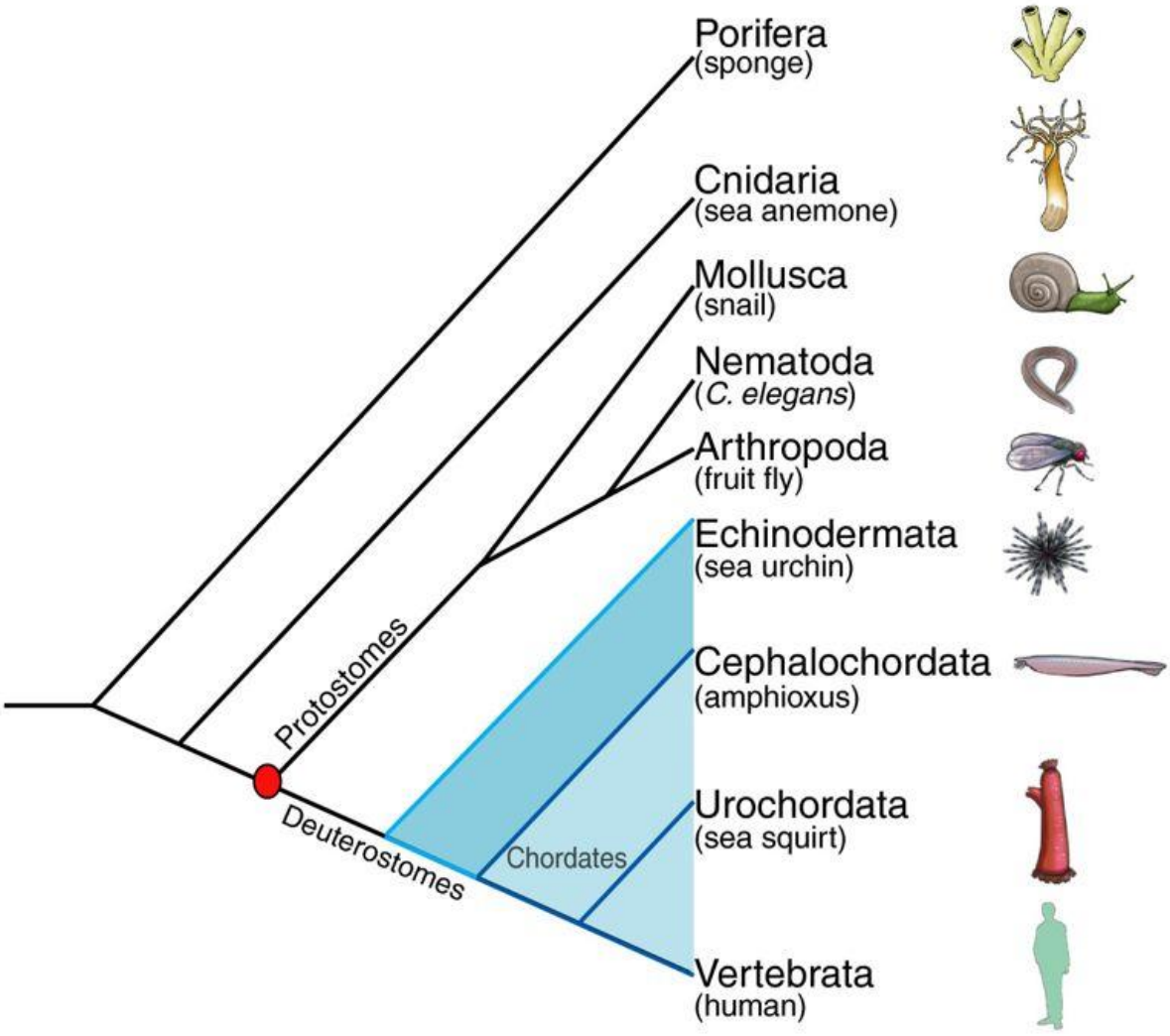


Figure 1-1: A schematic diagram of a phylogenetic tree showing the general evolutionary relationships of metazoans, including the divergence of the protostomes (red dot) and the deuterostomes (blue shade) showing Echinodermata (darker blue) clustered with the chordates (lighter blue). Note that the divergence times and tree branches are not to scale. Figure from “Novel genes dramatically alter regulatory network topology in amphioxus” by Zhang *et al.*, 2008, *Genome Biology*, 9(8), R123. Copyright 2008 by *Zhang et al.*; licensee BioMed Central Ltd. 20082004. Reprinted with permission.

Overview and General Description

Microbial communities colonize plants and animals alike, performing valuable metabolic processes. Some notable examples of this host-microbe relationship include the organic and inorganic chemical transformations conducted by rhizosphere-associated microbes (Reinhold-Hurek *et al.*, 2015) and the anaerobic degradation of plant cellulose by ruminal bacteria (Söllinger *et al.*, 2018). Their contributions are particularly significant in the gut environment of various host species and have been studied considerably in vertebrates such as humans (Cani, 2018), where they aid in digestion, contribute metabolites, impact immune system development, resist pathogen colonization, and perpetuate essential biochemical cycles (Cho & Blaser, 2012, Foster *et al.*, 2017). In humans, shortly after birth, microorganisms colonize the gut and undergo dynamic population shifts throughout the various stages of life (Nagpal *et al.,* 2018). Besides normal microflora, studies have reported altered gut microbial community compositions in various disease states in humans, such as obesity, gastrointestinal cancers, metabolic and autoimmune diseases, neuropsychiatric and cardiometabolic disorders, and inflammatory bowel diseases to name a few (Rodriguez *et al.*, 2015; Wang *et al.*, 2018). However, in marine organisms, the microbial communities in the gut are shaped by a relatively broad range of factors such as (1) the transfer of endosymbionts from food, (2) the internal biotic and abiotic conditions such as nutrient availability, pH, temperature, and oxygen saturation promoting certain microorganisms, and (3) the host's diet (Bjorndal, 1980, Lawrence *et al.*, 2013, Troussellier *et al.*, 2017). It has been suggested that

although the abundance and diversity of specific microbial taxa may vary geographically, metabolism of key macromolecules necessary for the sustenance of all organisms remain same (Cho & Blaser, 2012).

Traditionally, the study of these microbial communities has required a culture-dependent approach, which may underestimate diversity by failing to account the approximately 98% microorganisms that cannot be grown in the laboratory (Whitman *et al.*, 1998, Reveillaud *et al.*, 2014, Locey & Lennon, 2016, Jousset *et al.*, 2017, Tarnecki *et al.*, 2017). However, recent advancements of high-throughput sequencing (HTS) technologies have made it possible to develop comprehensive evaluations of the microbial communities inhabiting a specific location, popularly referred to as the "microbiota." The use of this technology alongside bioinformatics tools onto the collective microbial community DNA, or the "microbiome," has uncovered an unprecedented taxonomic diversity and underlying genetic potential of these microorganisms in various environments. Investigations of the importance of HTS technology in unraveling the structure-function relationship of the microbial communities at the highest possible resolution, regarding their significance to the host and their impact to the environment, have increased over a relatively short period of time (Figure 1-2). By using this genomics approach on the primitive gut systems of two popular sea urchin model organisms, the green *Lytechinus variegatus* and purple *Strongylocentrotus purpuratus*, we will determine the shared aspects of taxonomic profiles and their associated functional attributes, as it relates to the

health of this evolutionary and ecologically significant organism, including their environmental impact.

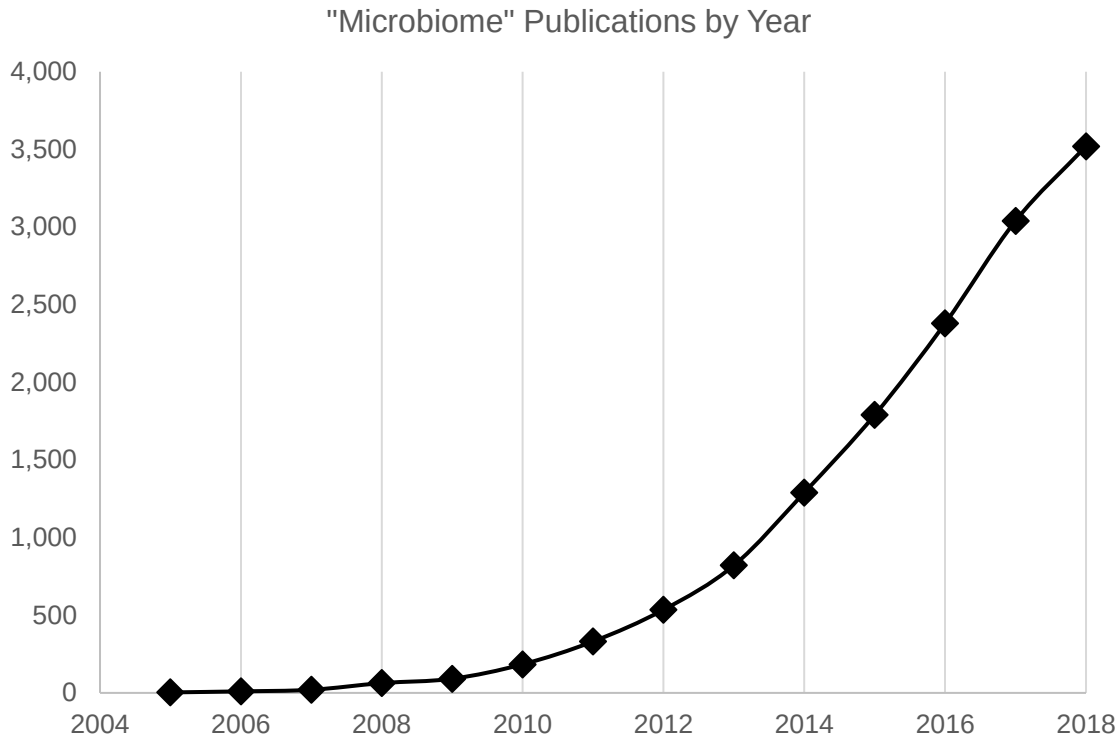


Figure 1-2: The number of microbiome publications by year shown as a scatterplot with smoothed lines. The data was acquired by searching the keyword “microbiome” in the article titles through Google Scholar. The Y-axis depicts the number of matches, and the X-axis depicts the year. Data plotted using Microsoft Excel Software (Microsoft, WA, USA).

The Sea Urchin Biology and Ecology

Sea urchins are invertebrate deuterostomes and share many fundamental biological and physiological process with higher chordates such as humans (Sodergren *et al.*, 2006), and these animals have been used as model organisms for embryology, aging, evolutionary and cell biology research (Davidson *et al.*, 2002, Bodnar & Coffman, 2016). In North America, the green sea urchin *Lytechinus variegatus* (order Temnopleuroida, family Toxopneustidae) are found

along the Gulf of Mexico and Eastern Coast of the United States, where it tempers sea grass growth as a dominant grazer of nearshore seagrass beds (Albright *et al.*, 2012). In its natural habitat in the Saint Joseph Bay Aquatic Preserve, Florida (Figure 1-3), the green sea urchin affects the biomass of various marine and coastal organisms of economic and ecological significance (Ellis *et al.*, 2011). This temperate marine environment maintains a yearly water temperature between 13.3 - 30 °C (average = 22 °C) (www.nodc.noaa.gov/), and represents a rich microbiota that inhabit the water column and benthic zone, including plants and other marine organisms at various trophic levels (Skoog *et al.*, 1999, Welsh, 2000, Deming & Carpenter, 2008, Felder & Camp, 2009, Kellogg *et al.*, 2009, Erwin *et al.*, 2011, Koo *et al.*, 2014).

Figure 1-3: The collection site of the green *Lytechinus variegatus* sea urchins in their seagrass meadow habitat in Port Saint Joseph, Florida, Aquatic Preserve (29.80 °N 85.36 °W) (red arrow) Curtesy of Google Earth.

On the other hand, the purple sea urchin *Strongylocentrotus purpuratus* (order Camarodonta, family Strongylocentrotidae) grazes algae, kelp, and seagrass along the shoreline and within rocky tide pools on the Pacific Northwest Coast, ranging from Alaska to Baja Mexico (Watanabe & Harrold, 1991, Ebert *et al.*, 1994, Tegner & Dayton, 2000, Davidson & Grupe, 2015). Their intertidal habitat along the shoreline of Oregon (Figure 1-4) maintains a comparatively colder yearly temperature of 10 - 12.8 °C (average = 11.7 °C) (www.nodc.noaa.gov/) and accommodates tufted algae and various invertebrates, including a variety of interspecies associations and dynamic population patterns that shape their ecosystems (Dethier, 1984, Metaxas & Scheibling, 1993).

Figure 1-4: The collection site of the purple *Strongylocentrotus purpuratus* sea urchins in their Intertidal pool habitat at Cape Arago, Oregon (43.30 °N 124.40 °W) (red arrow) Curtesy of Google Earth.

Gut Anatomy and Physiology of the Sea Urchin

Both the green and purple sea urchins represent a simple gut system (Figure 1-5), and though they are considered omnivorous, they are mainly herbivorous in their natural habitat (Ziegler *et al.*, 2010). Digestion of their food begins at the Aristotle's Lantern mastication apparatus that is comprised of five tooth-like structures of pentameral symmetry and is used for scraping algae from rock surfaces, benthic feeding and excavating, puncturing of seagrass and algal cell walls, and mechanical digestion (Lasker & Giese, 1954). After ingestion, specialized mucus-producing cells in the pharynx and esophagus envelop their feed into a gut digesta pellet separate from the gut tissue that is accompanied by an enrichment of microbiota and remains intact during transit through the gut lumen and upon egestion into their environment (Lasker & Giese, 1954, Holland & Ghiselin, 1970, De Ridder *et al.*, 1982, Hakim *et al.*, 2015). The gut digesta then traverses into the gut system that is comprised of outcrops (festoons). The proximal gut tissue contains exocrine cells producing digestive enzymes, along with a siphon organ to shunt water and avoid enzyme dilution (Ziegler *et al.*, 2010). The siphon reconnects at the distal gut, where the majority of the absorptive epithelial cells are found for the uptake of dietary nutrients (Claereboudt & Jangoux, 1985, Holland, 2013). The basal surface of the gut tissue is in contact with the inner test coelomic fluid that is oxygenated (Thorsen, 1998), and the internal environment of the pelleted digesta therein has been suggested as anaerobic environment (Meziti *et al.*, 2007). The oxygen saturation levels in the gut system creates an interface of oxic and anoxic microniches that

may shape microbial community profiles and their associated metabolic qualities (Brune *et al.*, 2000). The nutrients acquired from their diet will mobilize into the gut wall for immediate storage and into the coelomic fluid to be allocated into the gonads to be stored in the nutritive phagocytes as glycogen and peptide components to supply developing egg and sperm cells during gametogenesis (Pearse & Cameron, 1991).

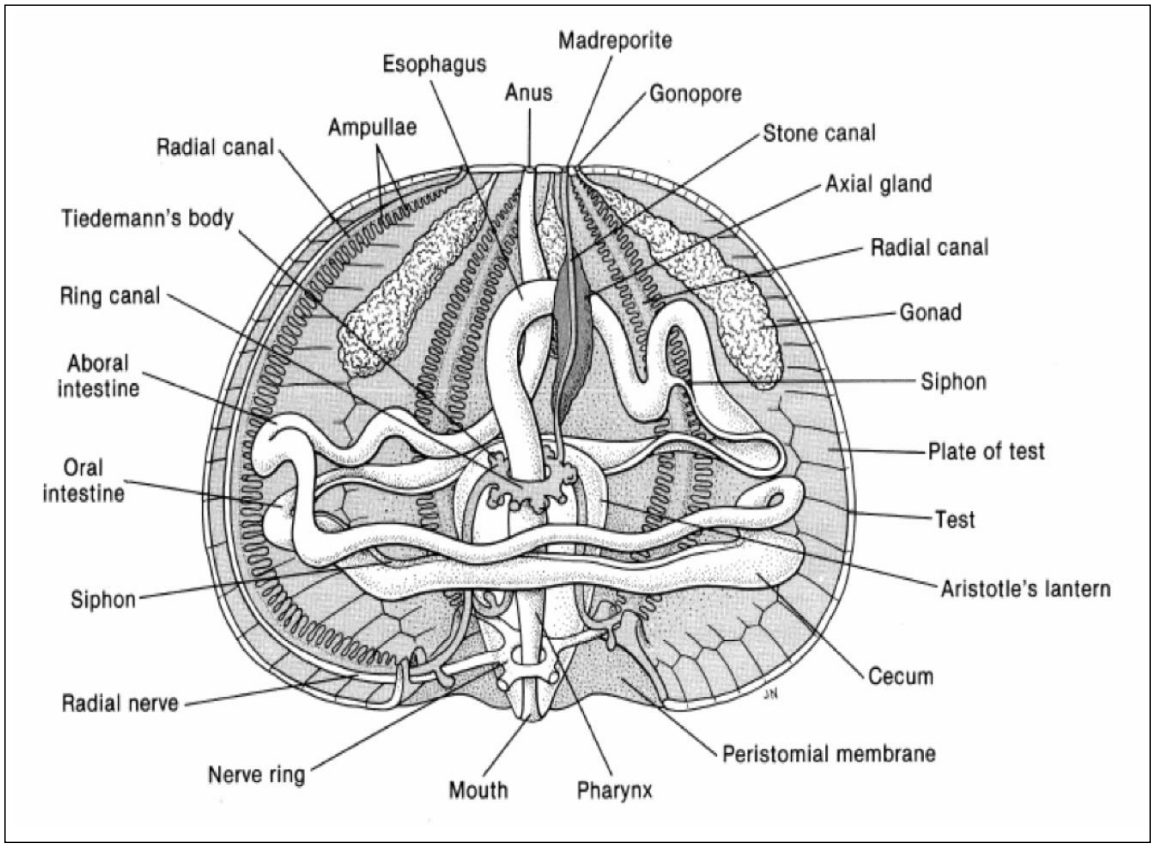


Figure 1-5: A schematic diagram of the general internal anatomy of sea urchins. Figure adapted from W.M. Reid in Invertebrate Zoology by E.E. Ruppert and R.D. Barnes, Copyright 1994 by Saunders College Publishing, Philadelphia. Reprinted with permission.

Gut Microbiota in Sea Urchins

Innate digestive enzymes have been examined in the gut system of sea urchins, with galactosidases and glucosidases implicated in some carbohydrate metabolisms (Klinger, 1984, Klinger *et al.*, 1986). However, the complete profile for the digestion of their naturally encountered diet rich in structural polysaccharides, as well as protein and lipid digestion, have been scarcely observed in sea urchins, implicating bacteria in the digestive processes of this organism (Lawrence *et al.*, 2013). Early culture-dependent studies on sea urchin microbiota focused on the bacteria of the gut digesta of *Strongylocentrotus purpuratus* and their involvement in the digestion of structural polysaccharides in their algal diet (Lasker & Giese, 1954). Later examinations have suggested *Vibrio* to be a dominant gut inhabitant of *Echinus esculentus* (Unkles, 1977), with others assigning this genera with alginolytic activity in the guts of sea urchins *Strongylocentrotus nudus* and *Strongylocentrotus intermedius* (Sawabe *et al.*, 1995). Other metabolisms attributed to *Vibrio* in the digestive processes of sea urchins included gelatin digestion (Beleneva & Kukhlevskii, 2010), and protein metabolism including nitrogen fixation and amino acid assimilation (Fong & Mann, 1980, Guerinot & Patriquin, 1981a, Guerinot & Patriquin, 1981b). Various anaerobic processes have also been suggested to occur by the sea urchin gut bacteria, such as fermentation of sugars by *Vibrio* and the utilization of inorganic compounds as alternate electron acceptors for energy (Meziti *et al.*, 2007). Moreover, nitrogen fixation in the absence of oxygen by the bacteria in the gut has been implied by Guerinot and Patriquin (1981a,b) in the guts of sea urchins

Diadema antillarum, *Echinometra launter*, *Tripneustes ventricosus*, *Strongylocentrotus drobachiensis* and *Strongylocentrotus droebachiensis* (Guerinot & Patriquin, 1981a, Guerinot & Patriquin, 1981b, Guerinot *et al.*, 1982). Such anaerobic metabolisms are suggested to accompany the compartmentalization of bacteria based on oxygen saturation of the gut digesta as an anaerobic microniche (Meziti *et al.*, 2007). With the recent advances in HTS targeting the microbial metacommunity DNA directly, it has become possible to evaluate these gut microbial communities comprehensively and at a high taxonomic resolution.

Sea Urchins and their Environmental Impact

The sea urchin influences neighboring seagrass and algae communities by their grazing activity, and effects other marine organisms sharing their ecosystem (Pearse, 2006, Zhadan *et al.*, 2017). In addition to tempering marine vegetation, their egesta has been considered a key component in the detritus pathway as well as a nutrient-rich food source for marine organisms in the benthic, pelagic, and coastal regions (Sauchyn *et al.*, 2011, Jensen *et al.*, 2018). Sea urchins have been shown to incorporate on average 25% of their ingested material, with the remaining portion released into their environment as fecal particulate organic matter (POM) in the mucous enveloped egesta (Mamelona & Pelletier, 2005). Importantly, the microbiota associated with the sea urchin egesta has been suggested to contribute to the biogeochemical processes of their marine environment, with bacteria performing key chemical transitions such

as increasing the lipid, nitrogen, and organic carbon content (Sauchyn *et al.*, 2011). Such metabolisms have been investigated in other marine invertebrates as they contribute to the dynamic food web in their habitat (Wotton & Malmqvist, 2001, Distel *et al.*, 2002, Reynolds *et al.*, 2007, Fiore *et al.*, 2010, Weigel & Erwin, 2017).

Nitrogen Cycles in the Sea Urchin Gut and their Habitat

The marine environment comprises a dynamic biogeochemistry, particularly in the accumulation of nitrogen in its various organic and inorganic forms (Yagi *et al.*, 2010). Nitrogen is critical in the structure of some biological molecules, such as amino acids and nucleotides, and its bioavailability is often limited as nearly 80% of the global nitrogen budget is in an unreactive dinitrogen state (Mulholland & Lomas, 2008). However, primary producers such as algae and seagrass require a biologically accessible form of nitrogen (Ryther & Dunstan, 1971, Herbert, 1999), and heterotrophic marine organisms depend on external sources of assimilated organic nitrogen for their normal function, particularly in herbivores consuming diets disproportionately high in carbohydrates (Sterner & Hessen, 1994, Kneip *et al.*, 2007). Marine microbiota play a significant role in the nitrogen redox cycles in the coastal habitat (Herbert, 1999, Morrison *et al.*, 2017). Specifically, nitrate and nitrite reducing marine bacteria may utilize these metabolic processes often yielding higher energy as compared to carbohydrate fermentation (Atlas & Bartha, 1998), and nitrogen-fixing bacteria utilizing elemental nitrogen may generate ammonia, which is a

preferred source of assimilation into its organic form. Such processes have been linked to the bacteria in the sea urchin gut system for the retention of dietary and marine nitrogen and its assimilation into organic molecules such as amino acids and nucleotides in the host (Fong & Mann, 1980, Guerinot & Patriquin, 1981a, Guerinot & Patriquin, 1981b), with these pathways continuing in the egesta into their environment (Sauchyn *et al.*, 2011). However, excess accumulation of ammonia are toxic to sea urchins (Siikavuopio *et al.*, 2004), indicating both endogenous and microbially-mediated detoxification of the compounds to be an important biochemical process. As such, determining the microbial communities and their functions will help elucidate the contributions of key taxa and functional genes in these metabolisms in the gut of the sea urchin, and as they relate to the continued biochemical processes post-egestion in their marine habitat.

Laboratory Aquaculture of Sea Urchins

Sea urchins are frequently collected from their natural habitat and cultured in the laboratory setting as model animals for various basic and applied research purposes, such as developmental biology, toxicology, immunology, biological aging, evolutionary and cell biology (Di Bernardo & Di Carlo, 2017). The potential for the commercial use of the sea urchin in the aquaculture industry has also been proposed (McBride, 2005, Heflin *et al.*, 2016), as a popular food delicacy due to their gonads, and their potential utility in co-culture with detritivores that may benefit from their egesta as a high energy food source (Brown *et al.*, 2011, Yokoyama, 2013, Harris & Eddy, 2015, Jensen *et al.*, 2018). The laboratory

aquaculture conditions include modulations of their diet, by preparing formulated feeds that are optimized for their nutritional requirements (Hammer *et al.*, 2006), or through various combinations of algae and seagrasses from their habitats (McBride, 2005, Schram *et al.*, 2018). However, the changes in their diet from the natural habitat to a laboratory-formulated feed may reshape the microbiota and their metabolic roles in the gut ecosystem and potentially impact the host's nutrition and health. In contrast, the natural marine environment is a reservoir of diverse local microbial communities. Additionally, the natural marine environment undergoes dynamic abiotic fluctuations including temperature, pH, salinity, and photoperiod that may modulate gut microbial profiles compared to the regulated laboratory environment conditions (Vellend, 2010, Costello *et al.*, 2012, Martínez *et al.*, 2015, Apprill, 2017, de la Calle, 2017). Application of HTS has allowed us to compare and understand the impact of laboratory conditions in shaping the gut microbial communities once transferred and maintained under aquaculture conditions to the laboratory. This approach also helps to determine those bacteria that are both consistent and likely retained as key inhabitants of the gut system between the two environments.

High-Throughput Sequencing and Bioinformatics

Due to the complex life strategies of the near-one-trillion species of microbes inhabitting the planet (Locey & Lennon, 2016), it is postulated that less than 2% of the total microbiota can be cultured by traditional laboratory microbiology culture techniques. These culture-dependent methods can also lead

to over estimations of the metabolic input of certain microbial taxa, and a failure to account for rare taxa that may be crucial in the community structure, stability, and biogeochemical cycling of the microbiota (Reveillaud *et al.*, 2014, Jousset *et al.*, 2017, Tarnecki *et al.*, 2017). However, the rapid progress in HTS and concurrent bioinformatics analysis techniques have made it possible to achieve microbial community structure, distribution, and metabolic capacity using a genomics approach termed “metagenomics.” The targeted metagenomics approach incorporates universal DNA oligonucleotide primers and the polymerase chain reaction (PCR) designed to amplify a specific region of the purified DNA from a microbial community sample. Typically, this region is a phylogenetically informative gene, such as the conserved bacterial 16S rRNA genes coding for the small ribosomal subunit (Woese, 1987). This gene is comprised of hypervariable regions of DNA (V1-9) that are flanked by conserved regions, and utilizing HTS on the V4 region in particular has been suggested to offer adequate variation to confidently delineate phylogeny and assign taxonomic identity of microbiota at the genus level (Kozich *et al.*, 2013).

Alternatively, the shotgun metagenomics method involves the sheering and sequencing of purified microbial community DNA (*i.e.* metagenome) and individual sequencing of DNA fragments that encompass taxonomically informative genes as well as other coding sequences to achieve biological functions across the genome (Sharpton, 2014, Quince *et al.*, 2017). For both approaches, the resultant nucleotide read data per sample are then analyzed by bioinformatics tools to assign taxonomic identities and their distribution patterns,

including functional gene information (Langille *et al.*, 2013, Keegan *et al.*, 2016). Some popular databases used to assign taxonomic identities include SILVA (Pruesse *et al.*, 2007), the All-Species Living Tree Project (LTP) (Yilmaz *et al.*, 2013), Ribosomal Database Project (RDP) (Wang *et al.*, 2007), Greengenes (GG) (McDonald *et al.*, 2012), and European Molecular Biology Laboratory (EMBL) Nucleotide Sequence Database (Baker *et al.*, 2000), and oftentimes incorporating multiple databases may provide added validation to support the taxonomic distribution determined by using one of these databases alone (Balvočiūtė & Huson, 2017).

Hypothesis, Specific Objectives, and Brief Descriptions of the Dissertation Research

The overall hypothesis of this dissertation is that the microbial community structure and function of the green sea urchin *Lytechinus variegatus* and purple sea urchin *Strongylocentrotus purpuratus* from two distinct habitats will manifest microbial taxa from their environments into the gut tissue and gut digesta, demonstrating a commonality of functional attributes likely benefiting the host, and contributing to crucial marine biogeochemical processes.

Based on the hypothesis, the overall objectives of this dissertation are to utilize targeted metagenomics of the collective microbial 16S rRNA genes to investigate the microbial composition and predicted functional attributes in the gut ecosystem of the (1) green *Lytechinus variegatus* and (2) purple *Strongylocentrotus purpuratus* sea urchins from their natural habitats.

Additionally, we will (3) utilize the targeted metagenomics data from the naturally occurring green sea urchin for comparison with the gut tissue and gut digesta microbiome of the laboratory aquaculture counterparts, to show the persistence of microbiota in the gut system of this commonly used model organism. We will then utilize the shotgun metagenomics approach to the gut digesta of the naturally occurring green and purple sea urchins, to (4) gain insight into the taxa comprising the microbial communities in the gut digesta and elaborate the metabolisms and biochemical cycles occurring in the uniquely encapsulated environment, with an emphasis on the nitrogen metabolic cycle; (5) and lastly we will establish a comprehensive taxonomic profile of the naturally occurring and laboratory aquaculture green sea urchin gut ecosystem, to verify the reported microbial identities and their distribution using multiple taxonomic databases. The specific objectives and their brief descriptions are as follows:

1. *Investigate the composition and predicted metabolic functions of gut microbial communities in green sea urchins* Lytechinus variegatus *collected from their seagrass meadow habitat in the Gulf of Mexico.*

This objective has been elaborated in Chapter 2 of this dissertation, in which we used targeted metagenomics of the green sea urchin's gut tissue, pharynx tissue, gut digesta, and egesta microbial communities, including the seagrass and seawater samples collected form Saint Joseph Bay Aquatic Preserve, Florida (29.80 °N 85.36 °W). This objective was accomplished by using the targeted metagenomics approach, incorporating the 250 bp paired-end kits

on an Illumina MiSeq HTS platform (Kozich *et al.*, 2013, Kumar *et al.*, 2014). Then, the raw HTS data was processed and analyzed using multiple bioinformatics tools. We were able to summarize the relative abundance of taxa per sample, showing Epsilonproteobacteria (Campylobacteraceae) to dominate the gut tissue compared to the gut digesta that showed *Vibrio*, *Propionigenium*, *Photobacterium,* and Flavobacteriales to be abundant. We validated a low intra-sample variation between biological replicates, and verified the hypothesized microbial community compartmentalization in the gut system, separate from the pharynx, seagrass and seawater microbial communities using beta diversity analyses. By assigning functional categories to the observed taxa in gut system, we showed that the metabolisms of macromolecules such as carbohydrate, amino acid, and lipid to be heightened in the gut digesta compared to energy metabolisms in the gut tissue. We also found nitrogen metabolisms in the gut tissues likely performed by Epsilonproteobacteria.

2. *Determine the microbial composition and their predicted metabolic functions in the gut ecosystem of purple sea urchins* Strongylocentrotus purpuratus *collected from their intertidal pool habitat on the North-American Pacific Coast.*

This objective has been elaborated in Chapter 3 of this dissertation. Similar to the approach in Chapter 2, we used targeted metagenomics HTS of the 16S rRNA gene (V4 region) and bioinformatics tools to determine the microbial communities of the gut tissue, pharynx, digesta and egesta of the purple sea urchin *Strongylocentrotus purpuratus* and the algae and sea water in

their intertidal pool habitat near Cape Arago, Oregon (43.30 °N 124.40 °W), including the predicted functional attributes in the gut system. In support of the results from Chapter 2, we showed a noteworthy abundance of Epsilonproteobacteria likely performing energy metabolisms in the gut tissue, compared to Gammaproteobacteria, Fusobacteria, and Bacteroidetes performing carbohydrate and amino acid metabolisms in the gut digesta. Beta diversity analysis also showed these gut microbial communities to be distinct from the pharynx, algae, and seawater communities. Additionally, we have measured the effect size of those taxa contributing most to the observed differences in the gut tissue and digesta, including the key taxa contributing most to the gut microbial community structure through co-occurrence network analysis.

3. *Compare the gut microbial communities and their predicted functional profiles in the naturally occurring and laboratory aquaculture green sea urchin* Lytechinus variegatus*.*

This objective has been elaborated in Chapter 4 of this dissertation. To accomplish the objective, we used the 16S rRNA gene sequence datasets generated from the gut ecosystem of the naturally occurring sea urchins described in Chapter 2 for comparison with a complement dataset generated from green sea urchins collected from the same location and transported to the University of Alabama at Birmingham where they were held for six months in a recirculating saltwater tank system, and fed *ad libitum* once every 24-48 h a formulated feed (Hammer *et al.*, 2006, Hakim *et al.*, 2015). We applied an

updated bioinformatics approach for enhanced beta diversity and species resolution and showed an abundance of *Arcobacter* spp. performing energy metabolisms in the gut tissues of both laboratory aquaculture and naturally occurring sea urchins. Additionally, though the gut digesta of both groups showed similar higher-level phylogenetic assignments, the laboratory group showed a higher abundance of *Vibrio*, whereas the naturally occurring group had a more diverse taxonomic distribution that included *Photobacterium*, *Propionigenium*, and Flavobacteriales. The predicted metabolic profiles of the gut digesta of both groups showed categories related to the digestion of macromolecules including carbohydrate, amino acid, and lipids compared to the gut tissue. However, a preferential enrichment of these metabolisms was observed in the naturally occurring digesta compared to the laboratory aquaculture counterparts, potentially due to the nutritional profile of their naturally encountered *Thalassia testudinum* food source performed by a more diverse microbial community. The results of this objective have shown that while the gut tissue maintains a similar microbial profile, the gut digesta will show differences in the diversity and richness between their natural habitat and the laboratory aquaculture conditions.

4. *Verify the taxonomic identities and metabolic qualities using shotgun metagenomics of the microbiota occurring in the green and purple sea urchin gut digesta.*

This objective has been elaborated in Chapter 5 of this dissertation. To accomplish these objectives, we have applied shotgun metagenomics on the purified microbial DNA from the gut digesta of Chapter 2 and 3 as the primary location for the digestion of their naturally encountered food and other biochemical cycles such as nitrogen retention and assimilation into amino acids and nucleotides. We have shown that the gut digesta of green and purple sea urchins had taxa distribution consistent with a broad range of invertebrates including echinoderms, with a high abundance of Gammaproteobacteria. This included *Vibrio* in both sea urchins and *Psychromonas* in purple sea urchins that could be correlated with their role in the digestion and metabolism of their carbohydrate-rich diet. Overall, the heightened metabolic functions performed in the digesta of these sea urchins were related to amino acid and carbohydrate digestion, including the genetic components for the reduction and fixation of nitrogen into ammonia and assimilation onto glutamine and asparagine, benefitting sea urchin nutrition during gut transit and influencing various trophic levels upon egestion.

5. *Provide a comprehensive taxonomic profile and distribution in the gut ecosystem of the naturally occurring and laboratory aquaculture green sea urchin.*

The objective of this research is elaborated in Chapter 6 of this dissertation. To do this, we have used the SILVA ACT: Alignment, Classification and Tree Service (www.arb-silva.de/aligner) to assign taxonomic identities to the representative sequences of the naturally occurring and laboratory aquaculture green sea urchin gut ecosystem from Chapter 4 using the Least Common Ancestor (LCA) approach to the following taxonomic sequence databases: SILVA, LTP, GG, RDP, or EMBL. By utilizing multiple databases, we have strengthened the results from Chapter 4**,** supporting Epsilonproteobacteria, and specifically *Arcobacter*, as the dominant taxon in the gut tissues of both groups. The relative abundance distribution patterns in the gut digesta were also supported, distinguishing an abundance of *Vibrio* in the laboratory aquaculture compared to the naturally occurring group. Moreover, we have shown the usefulness of utilizing multiple databases to establish a currently up-to-date, reliable and comprehensive taxonomic distribution.

Overall, this dissertation provides insights into the gut microbial community profile and functional capacity of two geographically distinct sea urchin species, to help understand the selective enrichment and distribution of specific microbial taxa into the gut environment, including their role in the digestion and health of their evolutionarily significant host, and ecological impact

upon egestion into their environment. Incorporating modern technological advancements of HTS and powerful bioinformatics tools has allowed us to establish a baseline microbial community of these two organisms through a genomic approach. Taken together, we show the unique enrichment and common presence of Epsilonproteobacteria in the gut systems of the green and purple sea urchins, along with their likely role of energy metabolisms in a compartmentalized gut ecosystem. For the gut digesta, Gammaproteobacteria appeared to be consistent irrespective of habitat, with variations occurring at the genus/species level, namely showing *Vibrio* in the green and *Psychromonas* in the purple sea urchins from their natural habitat. However, both the naturally occurring green and purple sea urchins showed common taxa that included *Propionigenium* and Flavobacteriales, with comparable assigned functions that included the digestion of dietary macromolecules. By comparing the naturally occurring and laboratory aquaculture green sea urchins, we have determined the effect of laboratory aquaculture conditions onto the gut microbiota of this popular model animal, as it likely pertains to host health. Additionally, by utilizing shotgun metagenomics of the gut digesta of the naturally occurring green and purple sea urchins, we have shown the underlying microbial genes involved in macromolecule digestion, and specifically, those involved in the reduction of nitrogen into ammonia and subsequent assimilation into amino acids and other organic molecules such as nucleotides. Moreover, by aligning the 16S rRNA gene sequences of the naturally occurring and laboratory aquaculture green sea urchin gut microbial communities to multiple taxonomic databases, we have

provided a comprehensive overview of taxa and their distribution for the research community, and further supported the presence of *Arcobacter* in the gut tissue. This dissertation will contribute to the growing body of knowledge of gut microbial communities in host organisms, utilizing modern genomics approaches to elaborate aspects of microbial community structure and function in this evolutionary significant sea urchin deuterostome.

CHAPTER II: THE GUT MICROBIOME OF THE SEA URCHIN, *LYTECHINUS VARIEGATUS*, FROM ITS NATURAL HABITAT DEMONSTRATES SELECTIVE ATTRIBUTES OF MICROBIAL TAXA AND PREDICTIVE METABOLIC PROFILES

by

JOSEPH A. HAKIM, HYUNMIN KOO, RANJIT KUMAR, ELLIOT J. LEFKOWITZ, CASEY D. MORROW, MICKIE L. POWELL, STEPHEN A. WATTS, AND ASIM K. BEJ

FEMS Microbiology Ecology, Sep. 2016, Vol. 92(9), PMID: 27368709

ABSTRACT

In this paper, we describe the microbial composition and their predictive metabolic profile in the sea urchin *Lytechinus variegatus* gut ecosystem along with samples from its habitat by using NextGen amplicon sequencing and downstream bioinformatics analyses. The microbial communities of the gut tissue revealed a near-exclusive abundance of Campylobacteraceae, whereas the pharynx tissue consisted of Tenericutes, followed by Gamma-, Alpha- and Epsilonproteobacteria at approximately equal capacities. The gut digesta and egested fecal pellets exhibited a microbial profile comprised of Gammaproteobacteria, mainly *Vibrio*, and Bacteroidetes. Both the seagrass and surrounding sea water revealed Alpha- and Betaproteobacteria. Bray-Curtis distances of microbial communities indicated a clustering profile with low intrasample variation. Predictive metagenomics performed on the microbial communities revealed that the gut tissue had high relative abundances of metabolisms assigned to the KEGG-Level-2 designation of energy metabolisms compared to the gut digesta, which had higher carbohydrate, amino acid and lipid metabolisms. Overall, the results of this study elaborate the spatial distribution of microbial communities in the gut ecosystem of *L. variegatus*, and specifically a selective attribute for Campylobacteraceae in the gut tissue. Also, the predictive functional significance of bacterial communities in uniquely compartmentalized gut ecosystems of *L. variegatus* has been described.

INTRODUCTION

The sea urchin, *Lytechinus variegatus*, inhabits the eastern coast of the United States, ranging through the Gulf of Mexico from NC, USA to the northern coast of Brazil (Hendler et al.1995; Watts, McClintock and Lawrence 2013). This species is typically found in shallow nearshore meadows of turtlegrass *Thalassia testudinum*, upon which they graze and ingest the leaves, including the associated epibionts and microbiota (Moore et al.1963; Zieman 1982; Beddingfield and McClintock 2000). Although seasonal variations occur, seagrass leaves consist of carbohydrates (45%–60% dry wt., with a majority being insoluble at 35%–45% dry wt.), proteins (10%–15% dry wt.) and low levels of lipids (<5% dry wt.) (Zieman 1982; Pradheeba et al.2011). Once consumed by the sea urchin, ingested materials (ingesta) will receive a mucosal contribution from the pharynx, and envelop the ingesta in the form of a spherical pellet, herein referred to as gut digesta. This unique digestive feature demonstrates a physical compartmentalization of the ingesta from the surface of the gut tissue, and is considered to be an advantageous digestive strategy for this animal (Brooks and Wessel 2003; Ziegler et al.2010). Importantly, gut digesta formation is accompanied by an apparent microbial enrichment that is distinctively different from the microbial community of the gut tissue (Hakim et al.2015). With this distribution of microbial communities between the gut digesta and the gut tissue, an allocation of defined metabolic profiles would be expected, enriched with those metabolic genes represented by the heightened microbial taxa. Additionally, pellets representing the gut digesta remain intact even after

egestion (egested fecal pellets) (Sauchyn, Lauzon-Guay and Scheibling 2011; Holland 2013), and are predicted to undergo microbial-driven molecular transitions (Sauchyn, Lauzon-Guay and Scheibling 2011; Hakim et al.2015). These egested fecal pellets have been acknowledged as an enriched source of nutrient to organisms at various trophic levels in the hydrosphere (Johannes and Satomi 1966; Koike, Mukai and Nojima 1987; Sauchyn, Lauzon-Guay and Scheibling 2011), and therefore sea urchin grazing of marine seagrass and other available sea vegetation has been identified as a major factor in nutrient cycling within benthic marine communities (Eklöf et al.2008; Miyata 2010).

Distribution of microbial communities in such a unique and compartmentalized ecosystem suggests either (i) a selective attribute of the host to promote the growth of key microbial members, (ii) a microbial life strategy selecting the host gut environment as conducive to growth and division or (iii) a complex combination of both circumstances (Bäckhed et al.2005; Shade and Handelsman 2012). Recently, the gut microbial community of the sea urchin *L. variegatus* cultured in the laboratory and fed with a formulated diet was described using next generation sequencing (NextGen) technology and bioinformatics analyses (Hakim et al.2015). This study established a baseline bacterial profile in the gut lumen, gut digesta and egested fecal pellets, including the feed and culture environment. Additionally, the study showed that although a diverse microbial community exists in their feed and surrounding culture environment, a select group of microbial taxa is differentially enriched in the gut lumen and gut digesta (Hakim et al.2015). Given the importance of the sea urchin gut microbiota

to the host as it pertains to digestive physiology, as well as the nutritional benefit provided by bacteria enveloped in the egested fecal pellets to the ecosystem they inhabit, it is imperative to map the microbial profiles of the gut as they occur in nature, including comparisons of the microbial communities of the marine environment to the sea urchin's own microbiota. These comparisons would define the selectiveness of the sea urchin gut ecosystems, if such selection exists, for preferred bacterial taxa in their natural habitat. The role of microbiota in digestion of the ingested food within the gut lumen and gut digesta has not been confirmed in sea urchins. However, based on the suggested role of gut microbiota in other organisms, it is likely that there is an intimate association with the digestion process (Guerinot and Patriquin 1981; Nelson et al.2010). With the advent of bioinformatics tools (Phylogenetic Investigation of Communities by Reconstruction of Unobserved States, PICRUSt v.1.0.0) which utilize the 16S rRNA NextGen sequencing data, it is possible to predict the functional attributes of microbiota residing in various components of the sea urchin digestive system (Langille et al.2013).

In this study, we collected fresh specimens of the sea urchin *L. variegatus* from shallow-water seagrass beds located in the northern Gulf of Mexico. We have identified the microbiota occurring in the lumen of the gut and pharynx, the gut digesta and the egested fecal pellets, as well as the surrounding sea water and natural seagrass (turtlegrass, *T. testudinum*) with high taxonomic coverage using a culture-independent NextGen Illumina MiSeq sequencing technology and bioinformatics tools. In addition, we have used the PICRUSt v.1.0.0 on 16S rRNA

gene sequence datasets and determined the predictive functional profile of microbial communities in the naturally occurring sea urchin gut microbiome.

MATERIALS AND METHODS

Collection of Lytechinus variegatus and sample collection

Lytechinus variegatus (*n* = 3) were collected in October 2014 from within 1 m^2 of each other in the Saint Joseph Bay Aquatic Preserve, Florida (29.80°N 85.36°W), placed in a clean plastic cooler containing sea water collected from the same location, and transported to the laboratory at the University of Alabama at Birmingham. Oceanic water conditions were recorded as 20 ± 2 °C, with a pH of 7.8 ± 0.2 and salinity of 28 ± 1 ppt. Leaves of the seagrass *Thalassia testudinum* were harvested by excision at the urchin collection site and placed in plastic bags for further microbiota extraction. Fresh sea water samples were collected within the top 1 m of the collection site and placed in sterile containers. Three adult sea urchins were dissected for the study (UR1 d = 55 mm, wet wt. = 31.59 g; UR2 d = 55 mm, wet wt. = 38.58 g; and UR3 d = 60 mm, wet wt. = 46.77 g). Tissue extraction and environmental sample preparation for NextGen began 7 ± 1 h following collection. Prior to dissection, the sea urchins were placed in individual containers containing sea water from the sample site. From these containers, egested fecal pellets were collected upon their release, to ensure that the released fecal pellets (egesta) were appropriately collected from each sea urchin, without the contamination of another sea urchin's fecal pellets. The gut tissue and pharynx tissue were collected as described in Hakim et al. (2015). The gut

digesta was voided and collected from the gut tissue by gentle shaking in autoclaved (121°C for 20 min at 103.42 kPa) sea water. The microbiota of the sea water (water) (n = 3) from Saint Joseph Bay Aquatic Preserve, Florida, was vacuum-filtered through Millipore 0.22 µm filtration paper (EMD Millipore Corporation, Danvers, MA, USA). The seagrass collected from the area of sea urchin grazing were minced using a sterile scalpel.

Metacommunity DNA purification and generation of 16S rRNA amplicon library

Microbial genomic DNA was isolated using the Fecal DNA isolation kit from Zymo Research (Irvine, CA, USA; catalog no. D6010) according to the manufacturer's instructions. An amplicon library was prepared from metacommunity DNA, using unique bar-coded oligonucleotide primers through PCR to amplify the hyper variable region 4 (V4) of the 16S rRNA gene (Kozich et al.2013; Kumar et al.2014). Oligonucleotide primers were adapted from the standard protocols of the Earth Microbiome Project (www.earthmicrobiome.org; Caporaso et al.2011, 2012), and were as follows: forward primer (515F) V4: 5′-AATGATACGGCGACCACCGAGATCTACACTATGGTAATTGTGTG-CCAGCMGCCGCGGTAA-3′; and reverse primer modified from 806R to include uniquely bar-coded 5′ region and adaptor sequence V4: 5′-CAAGAGAAGACGGCATAC-GAGATNNNNNNAGTCAGTCAGCCGGACTACHVGGGTWTCTAAT-3′ (Eurofins Genomics, Inc., Huntsville, AL, USA) (Kumar et al.2014). PCR amplification was set up as follows: 10 µL of 5× Reaction Buffer; 1.5 µL (200 µM) of each of the

dNTPs; 2 µL (1.5 µM) of each oligonucleotide primer solution; 1.5 µL (5 U) of the LongAmp® enzyme kit (New England Biolabs, Ipswich, MA, USA; catalog no. E5200S); 30 µL (2–5 ng µL–1) of the template DNA; and 3 µL of sterile H2O for a total reaction volume of 50 µL. The PCR cycling parameters were as follows: initial denaturation at 94°C for 1 min; 32 cycles of amplification with each cycle consisting of 94°C for 30 s, 50°C for 1 min, 65°C for 1 min; followed by final extension of 65°C for 3 min and a final hold at 4°C. The resultant PCR reaction was electrophoresed on a 1.0% (w/v) Tris-borate-EDTA/agarose gel, and the PCR product (~380 bp predicted product size) was visualized using UV illumination. The amplified DNA band was excised and purified from the agarose matrix using QIAquick Gel Extraction Kit according to the manufacturer's instructions (Qiagen Inc., Venlo, Limburg; catalog no. 28704).

NextGen sequencing by Illumina and bioinformatics

In preparation for NextGen sequencing, PicoGreen dye (Life Technologies, Grand Island, NY, USA) was used to quantitate the samples, which were adjusted to a concentration of 4 nM (Kumar et al.2014). To sequence the PCR products, the single lane flowcell NextGen sequencing Illumina MiSeq platform (Kozich et al.2013; Kumar et al.2014) was used, incorporating the 250 bp paired-end kits from Illumina specific to the V4 region of the 16S rRNA gene. Raw sequence data were demultiplexed and converted to FASTQ format (Cock et al.2010), evaluated for quality and filtered using the FASTX toolkit (Gordon and Hannon 2010). The overlapping regions of the paired-end reads (~245 bp)

were merged using the ‘fastq_mergepairs’ module of USEARCH (Edgar 2010). Read pairs with <50 bp overlap and/or over 20 mismatching nucleotides were discarded, and chimeras were removed using ‘identify_chimeric_seqs.py’ module of USEARCH (Edgar 2010). Read quality was assessed before and after filtering using FASTQC (Andrews 2010). All NextGen raw sequence data files have been deposited to NCBI SRA for public access (the accession number is SRP076869).

The following steps were performed using the Quantitative Insights Into Microbial Ecology microbiome analysis package (QIIME, v1.8.0) (Lozupone et al.2007; Caporaso et al.2010; Navas-Molina et al.2013; Kumar et al.2014). First, sequences were grouped into operational taxonomic units (OTUs) at a similarity threshold of 97% using UCLUST (Edgar 2010), and taxonomic assignments (to the species level when possible) were achieved through the Ribosomal Database Project (RDP) classifier (Wang et al.2007), trained using the Greengenes (v13.8) 16S rRNA database (DeSantis et al.2006; McDonald et al.2011), at a 60% confidence threshold (Wang et al.2007). OTUs observed to be <0.0005% abundant were removed. To summarize taxa abundance at different hierarchical levels (phylum, class, order, family and genus), biological replicates in the resultant OTU table were grouped according to analysis of similarity (ANOSIM) of the weighted ($R = 0.7251$; $P = 0.001$) Unifrac distances calculated between each sample (Lozupone and Knight 2005; Hamady, Lozupone and Knight 2010), conducted at 999 permutations using QIIME (v1.8.0). The top 100 most resolved taxa from this this data were used to construct stacked column bar charts using Microsoft Excel (Microsoft, Seattle, WA, USA). Alpha diversity was estimated

using the observed OTUs, Shannon (Shannon 1948; Hill et al.2003; Marcon et al.2014), and Simpson (Simpson 1949; Hill et al.2003) diversity was measured using QIIME (v1.8.0).

To construct the multidimensional-scaling (MDS) plots (Kruskal and Wish 1978; Clarke 1993; Clarke and Gorley 2001), the quality assessed unfiltered OTU table was used, and all samples were rarefied to the median OTU value (87 225) to account for variation in read depth (QIIME, v1.8.0) (Gotelli and Colwell 2011). Those samples with OTUs totaling less than the median value were maintained at their original value, and included in the Beta-diversity analysis (de Cárcer et al.2011). Beta diversity was visualized using PRIMER-6 software (Primer-E Ltd, Plymouth Marine Laboratory, Plymouth UK, v6.1.2), by standardizing and square root-transforming the subsampled OTUs from each sample (Clarke and Gorley 2001), and MDS plots were generated according to the Bray–Curtis similarity values (Bray and Curtis 1957). These similarity values were also used to construct the complete-linkage hierarchical clustering dendrogram (Krebs 1999; Clarke and Gorley 2001; Dawyndt, De Meyer and De Baets 2005) used in the heatmap, to show triplicate sample clustering according to the Bray–Curtis values (Bray and Curtis 1957). A heatmap was constructed by merging sample replicates, followed by filtering out those taxa with <1% (<0.01) from the top 100 OTUs, using the ‘heatmap.2’ function of the R package (available at http://CRAN.R-project.org/package=gplots). The dendrograms were created using complete-linkage clustering of the compositional data using the Bray–Curtis values in the R package. An additional dendrogram to show

intrasample variation in the heatmap was created using Bray–Curtis similarity and complete-linkage clustering using PRIMER-6 (v6.1.2) software.

PICRUSt (v1.0.0) and STAMP (v2.1.3) analysis for predictive metagenomics

PICRUSt v1.0.0 (Langille et al.2013) was used to determine the predictive metagenomes of the microbial populations from each sample. PICRUSt (v1.0.0) extrapolates known metabolic characteristics based on the associated phylogeny from the sequence data achieved after NextGen sequencing of the variable V4 region of the 16S rRNA gene (Langille et al.2013; Lang, Eisen and Zivkovic 2014). To achieve a predicted metagenome of each sample, the FASTA file created after quality assessment, chimeric trimming and filtering of sequences (<0.0005%) was used. OTUs corresponding to the sequence file were closed-referenced picked against the Greengenes (v13.5) database at a 97% identity, as suggested in PICRUSt (v1.0.0). The resultant OTU table was then supplemented with *de novo* OTUs present at >100 in any sample, as determined by open-reference OTU picking against the Greengenes (v13.5) database. The resultant OTU table was normalized, and metagenomes were predicted by referencing the assigned Greengenes Ids to the Kyoto Encyclopedia of Genes and Genomes (KEGG) Orthology (KO) Database (Kanehisa and Goto, 2000; Kanehisa et al.2014), using the 'predict_metagenomes.py' module of PICRUSt (v1.0.0). The predicted metagenomes were collapsed into hierarchical categories (KEGG-Level-2 and 3), and the relative abundances of the gut digesta (n = 3) and the gut tissue (n = 3) metagenome categories were calculated and graphed using

Microsoft Excel software (Microsoft). The KEGG-Level-2 functional categories were used for two-group box-plot analysis in Statistical Analysis of Metagenomic Profiles (STAMP v2.1.3) (Parks et al.2014). Variance was calculated using a two-sided Welch's t-test, which does not assume equal variance (Welch 1938), along with the Benjamini–Hochberg false discovery rate (FDR) (Benjamini and Hochberg 1995) statistic for multiple test corrections (Parks et al.2014; Zhao et al.2015). Confidence intervals were set to 95% (0.95). The P-value for the total variance between the two groups is listed in the box plots. The P-value for a significant difference between the mean relative abundances of each group was recognized as $P < 0.05$.

RESULTS

Total Illumina sequence reads, quality trimming and OTU designation

NextGen sequencing of the V4 region of the 16S rRNA gene using the Illumina MiSeq platform resulted in a total of 1,761,403 raw sequence reads from the 18 samples of the study (sea urchin samples $n = 3$; water $n = 3$; seagrass $n = 3$) (Table 2-1). Quality assessment and chimeric trimming produced 1,362,092 sequence reads, and removal of those unique sequences accounting for <0.0005% resulted in a total of 1,294,253 reads. Of these reads, the gut tissues (260,551 reads) expressed a combined total of 467 OTUs, and the pharynx tissue (238,729 reads), a total of 1,893 OTUs. The gut digesta samples (302,287 reads) displayed a total of 6,740 OTUs, and the egested fecal pellet samples (351,170 reads), a total of 10,172 OTUs. Of the environmental contributions to

the gastrointestinal microbiota, the sea water (11,508 reads) produced 914 total OTUs, and the seagrass (130,008 reads) produced 2271. Rarefaction curves reached or approached a plateau, indicating sufficient sampling depth (data not shown).

Microbial diversity across different samples

Relative abundances of the top 100 resolved taxa determined through the RDP classifier using Greengenes (v13.8) at a 60% confidence threshold are elaborated in Figure 2-1. Biological replicates were merged on the basis of ANOSIM of the weighted UniFrac distances (R = 0.7251; P = 0.001). From the gut tissues collected from the three sea urchins, microorganisms belonging to phylum Proteobacteria constituted the highest relative abundance. At the class level, Epsilonproteobacteria accounted for the highest abundance, revealing family Campylobacteraceae to be heightened (93%). In the gut tissue, genus level taxa assignments could not be achieved using QIIME (v1.8.0). Subsequent BLAST (Altschul et al.1990; Morgulis et al.2008) search of the overrepresented sequence (253 bp) corresponding to family Campylobacteraceae, and occurring across all samples, revealed uncultured *Arcobacter* sp. (identity: 91%, E-value: 1.82e–87), *Arcobacter bivalviorum* (identity: 91%, E-value: 2.00e–89), *Sulfuricurvum* sp. (identity: 90%, E-value: 4.00e–86) and other uncultured bacterium clone (identity: 90%, E-value: 2.00e–89). Additionally, *Candidatus Hepatoplasma* (~5%), of phylum Tenericutes, were observed. The pharynx tissue revealed a high abundance of Tenericutes (~60%), showing class

Mollicutes to be the most resolved taxa. Proteobacteria were also observed, with class Alpha-, Beta-, Epsilon- and Gammaproteobacteria showing as prevalent (Figure 2-1).

The gut digesta and egested fecal pellets of the sea urchins revealed similar microbial profiles, and were dominated by phyla Proteobacteria (~50%) with class Gammaproteobacteria appearing as heightened, followed by Alpha-, Delta-, and Epsilonproteobacteria at approximately equal prevalence. The heightened genera were *Vibrio*, *Photobacterium*, *Propionigenium* and *Ferrimonas*. Bacteroidetes were also observed, with the classes Flavobacteria, Cytophagia and Bacteroidia represented. *Persicobacter* and *Tenacibaculum* were the observed genera from Bacteroidetes.

From the surrounding environment in the Gulf of Mexico, the grazed upon seagrass revealed a high abundance of Cyanobacteria, followed by Proteobacteria and Bacteroidetes. Positive identification from within Cyanobacteria revealed order Streptophyta to be the most represented taxa, though resolution past the order level could not be achieved. However, the occurrence of this order was insignificant in the gut microbial community, and therefore their role in digestive process, if any, is unknown. Both the seagrass and the water samples revealed phylum Proteobacteria, particularly Rhodobacteraceae (class Alphaproteobacteria). The water samples displayed phylum Bacteroidetes, with a fairly high relative abundance of class Flavobacteria (~25%), with family Flavobacteraceae and Cryomorphaceae represented.

Table 2-1: Sample statistics corresponding to the 18 samples of the study, determined after NextGen sequencing of the V4 variable region of the 16S rRNA gene using the Illumina MiSeq platform. Included are the raw sequences prior to analysis, followed by the quality-assessed sequence read counts and the resultant unique observations (unfiltered). This is followed by the filtered count (in which those sequences occurring at less than 0.0005% were removed), as well as the resultant unique observations. The Shannon and Simpson diversity values corresponding to each sample are also included. For Shannon diversity, a value much higher than 0 would be considered more diverse, and for Simpson diversity, a value closer to 1 would be considered more diverse.

Sample	Raw Sequences	Quality Assessment	Unique Observations (Unfiltered)	Filtered Reads	Unique Observations (Filtered)	Shannon	Simpson
Egested Fecal Pellet 1	204,150	149,503	16,030	133,208	3,431	8.33	0.97
Egested Fecal Pellet 2	158,186	126,394	10,306	117,384	3,367	8.14	0.97
Egested Fecal Pellet 3	138,390	108,855	9,538	100,578	3,374	8.46	0.98
Gut Digesta 1	132,132	105,608	7,550	99,468	2,554	6.94	0.95
Gut Digesta 2	143,149	118,643	5,712	114,053	2,004	5.76	0.93
Gut Digesta 3	113,415	92,589	5,131	88,766	2,182	7.20	0.96
Gut Tissue 1	89,881	74,363	1,205	73,264	184	0.83	0.19
Gut Tissue 2	125,274	108,088	1,627	106,493	173	0.52	0.09
Gut Tissue 3	95,653	81,861	1,089	80,794	110	0.60	0.12
Pharynx Tissue 1	140,831	104,981	3,416	101,762	667	4.14	0.67
Pharynx Tissue 2	94,593	68,311	2,637	65,931	672	4.18	0.76
Pharynx Tissue 3	99,972	72,323	1,678	71,036	554	2.21	0.38
Seagrass 1	230	175	70	128	41	5.66	0.97
Seagrass 2	696	525	222	470	177	7.23	0.99
Seagrass 3	208,197	137,462	7,006	129,410	2,053	6.87	0.88
Water 1	494	228	104	195	76	6.01	0.97
Water 2	12,175	9,271	1,026	8,621	568	5.83	0.91
Water 3	3,985	2,912	439	2,692	270	5.73	0.92
Total 18	**1,761,403**	**1,362,092**	**74,786**	**1,294,253**	**22,457**		

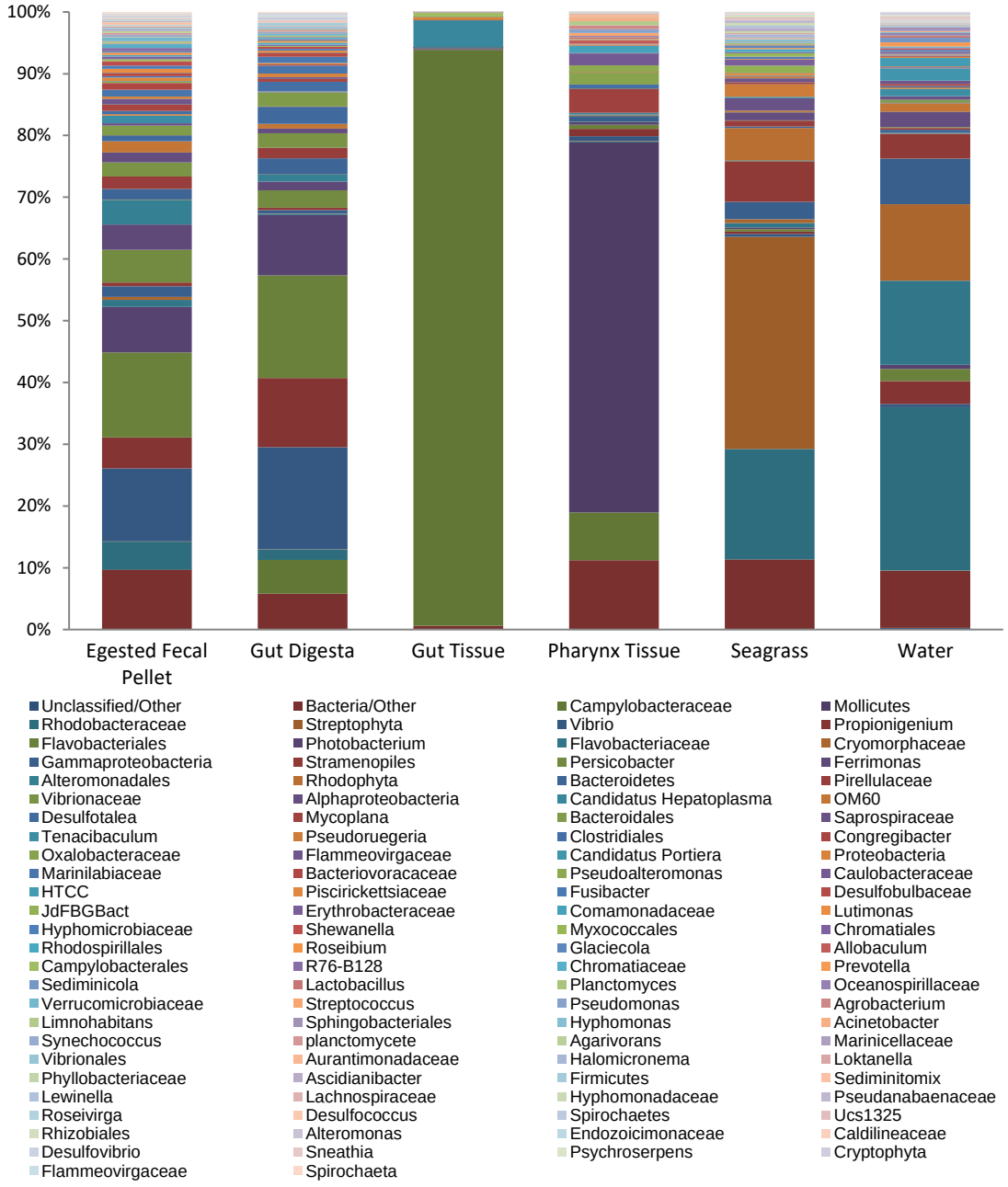


Figure 2-1: Stacked column bar graph of the top 100 most resolved taxa (to the genus level where possible) across all samples are presented. Replicates (*n* = 3) were merged, and OTUs were left untrimmed. Proteobacteria was found to be considerably abundant across all samples, as well as Bacteroidetes in the gut digesta, egested fecal pellets and water samples. The seagrass contained a high abundance of Cyanobacteria, and class Mollicutes of phylum Tenericutes dominated the pharynx tissue. Family Campylobacteraceae was determined to be the most abundant taxa in the gut tissue. In the gut digesta and egested fecal pellets, *Vibrio*, *Propionigenium*, Flavobacteriales, and *Photobacterium* were most abundant. Relative abundances were calculated in QIIME (v1.8.0), and graphed using Microsoft Excel software (Microsoft).

Statistical analysis

The Shannon diversity and Simpson diversity indices were calculated for the 18 samples of the study, revealing the gut tissue to have the least diversity of observed OTUs (Table 2-1). The pharynx tissue followed with a moderate diversity, and a high diversity was observed to be associated with the gut digesta and egested fecal pellets. The water and seagrass samples also displayed a high microbial diversity. MDS plot analysis revealed observable clustering of biological replicates from the sea urchin (Figure 2-2). Three gut tissue samples clustered together at 20% Bray–Curtis similarity, as did the pharynx tissue samples. The gut digesta and fecal pellet samples also clustered together. The seagrass and water displayed more intrasample variability, and did not cluster as predictably as the other sea urchin samples. Filtering for heatmap analysis (<1% of the top 100 most resolvable taxa) resulted in 42 taxa (Figure 2-3). This analysis revealed the gut tissue to be unique in microbial composition, separate from the other samples of the study, with a minor similarity to the pharynx tissue as depicted through complete-linkage hierarchical clustering (Figure 2-3). The OTU data of the gut digesta and egested fecal pellets were shown to have a similar microbial profile, and the sea water and seagrass showed a moderate similarity through complete-linkage (Figure 2-3). Dendrogram analysis of sample replicates using Bray–Curtis distances through complete-linkage hierarchical clustering supported both ANOSIM as well as the sample type grouping for the heatmap analysis.

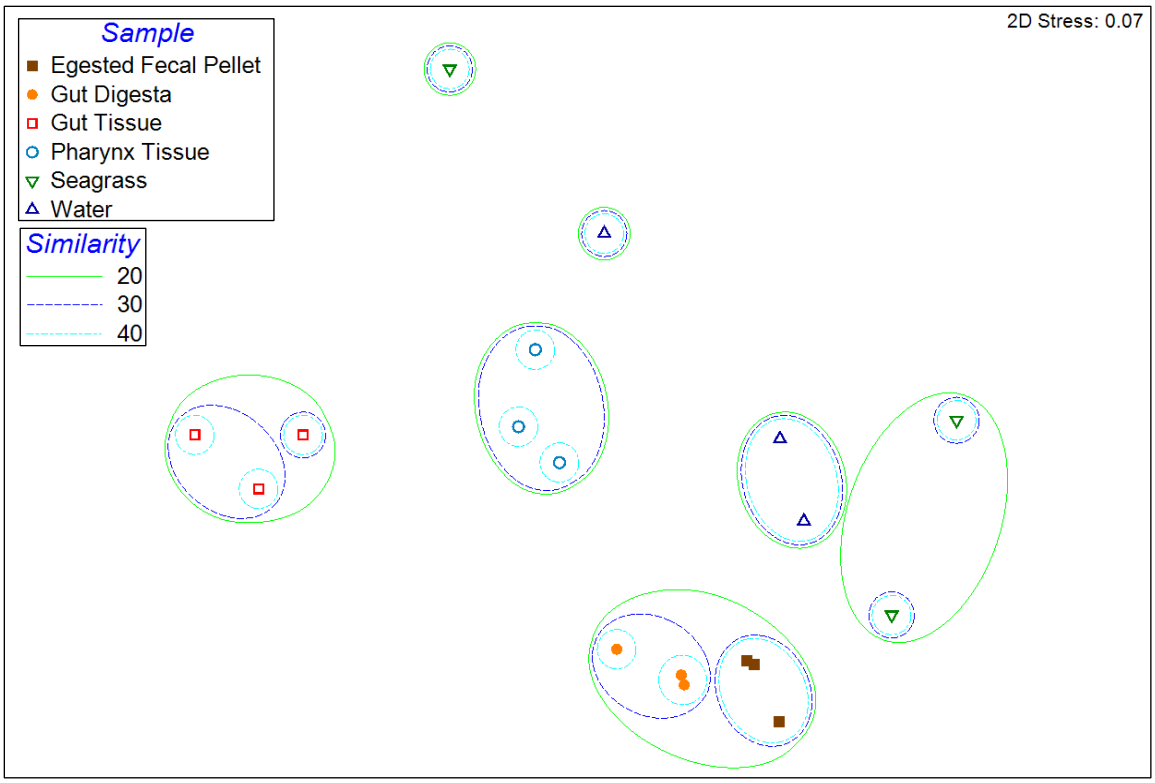


Figure 2-2: 2D multidimensional-scaling (MDS) graph produced by PRIMER-6 (v6.1.2) using subsampled OTU data generated through QIIME (v1.8.0). Overlay of clusters were generated according to Bray–Curtis similarity, followed by complete-linkage clustering with similarity thresholds set at 10% intervals from 20%–40%. The samples obtained from the sea urchin microbiome showed distinct clustering patterns, with low intrasample (*n* = 3) variation. Similarity = Bray–Curtis similarity (scaled to 100).

Predicted metagenomes based on PICRUSt (v1.0.0)

The microbial populations of those samples likely influencing the digestive processes occurring in the gut of the sea urchin were subjected to PICRUSt (v1.0.0) analysis to determine the metabolic processes predicted to be occurring in grouped populations, and compare those processes across compartmentalized microbial populations using the Welch's t-test of two groups (Figure 2-4). The KEGG-Level-2 categories considered were as follows: carbohydrate metabolism, amino acid metabolism, lipid metabolism and energy

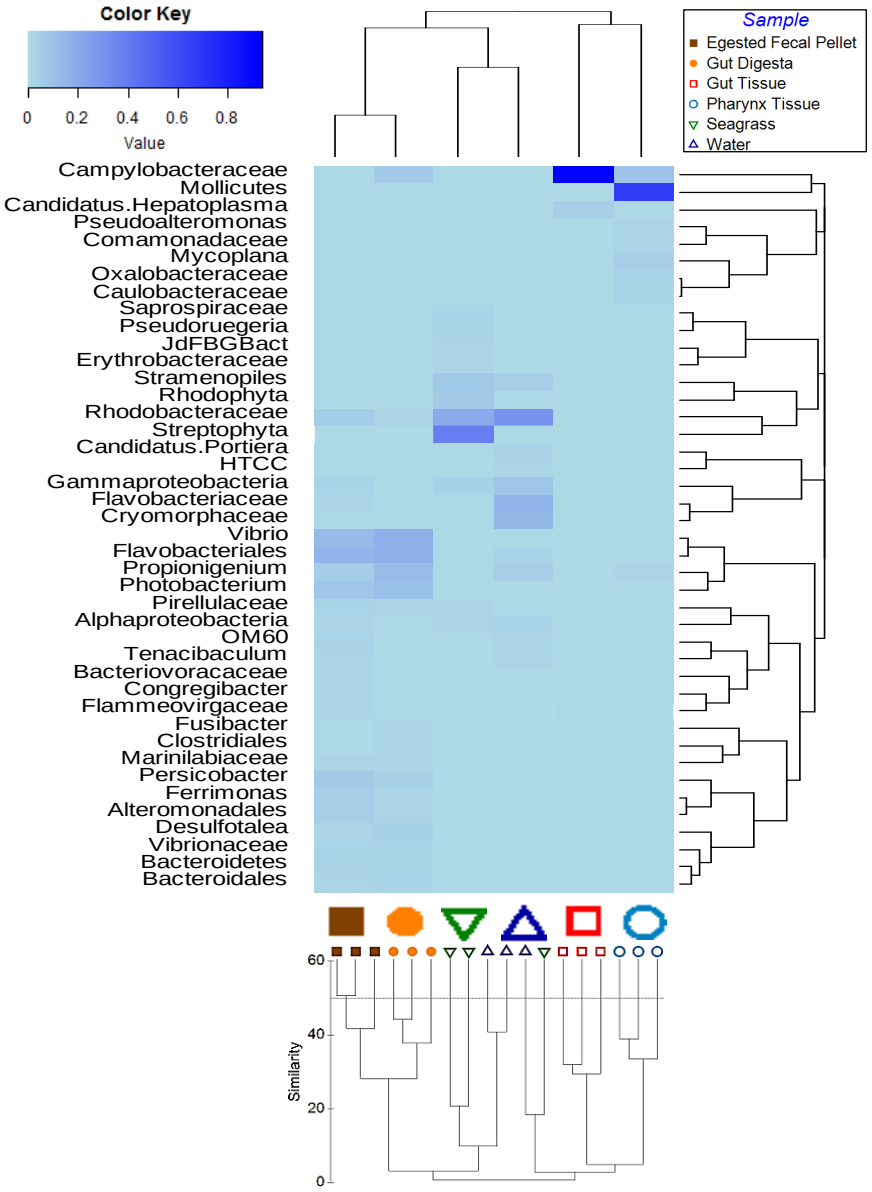


Figure 2-3: Heatmap generated using OTU data of the top most resolved taxa as determined using QIIME (v1.8.0), and filtered to 42 taxa by including only those taxa representing >1% of the total abundance. The rows correspond to bacterial taxa, and the columns represent the six different sample types (merged replicates) of this study. Dendrograms were created using complete-linkage hierarchical clustering of the compositional data. Heatmap was generated using the 'heatmap.2' function in R package (available at http://CRAN.R-project.org/package=gplots). Below is a complete-linkage (furthest neighbor) hierarchical clustering graph produced by PRIMER-6 (v6.1.2) using subsampled OTU data generated through QIIME (v1.8.0). Replicates corresponding to the sea urchin gut system possessed a low intrasample variation, and the environmental sample replicates showed slight intrasample variation. Similarity index is depicted on the left of the graph below, and similarity to a value of 60 is shown. Similarity = Bray–Curtis similarity (scaled to 100).

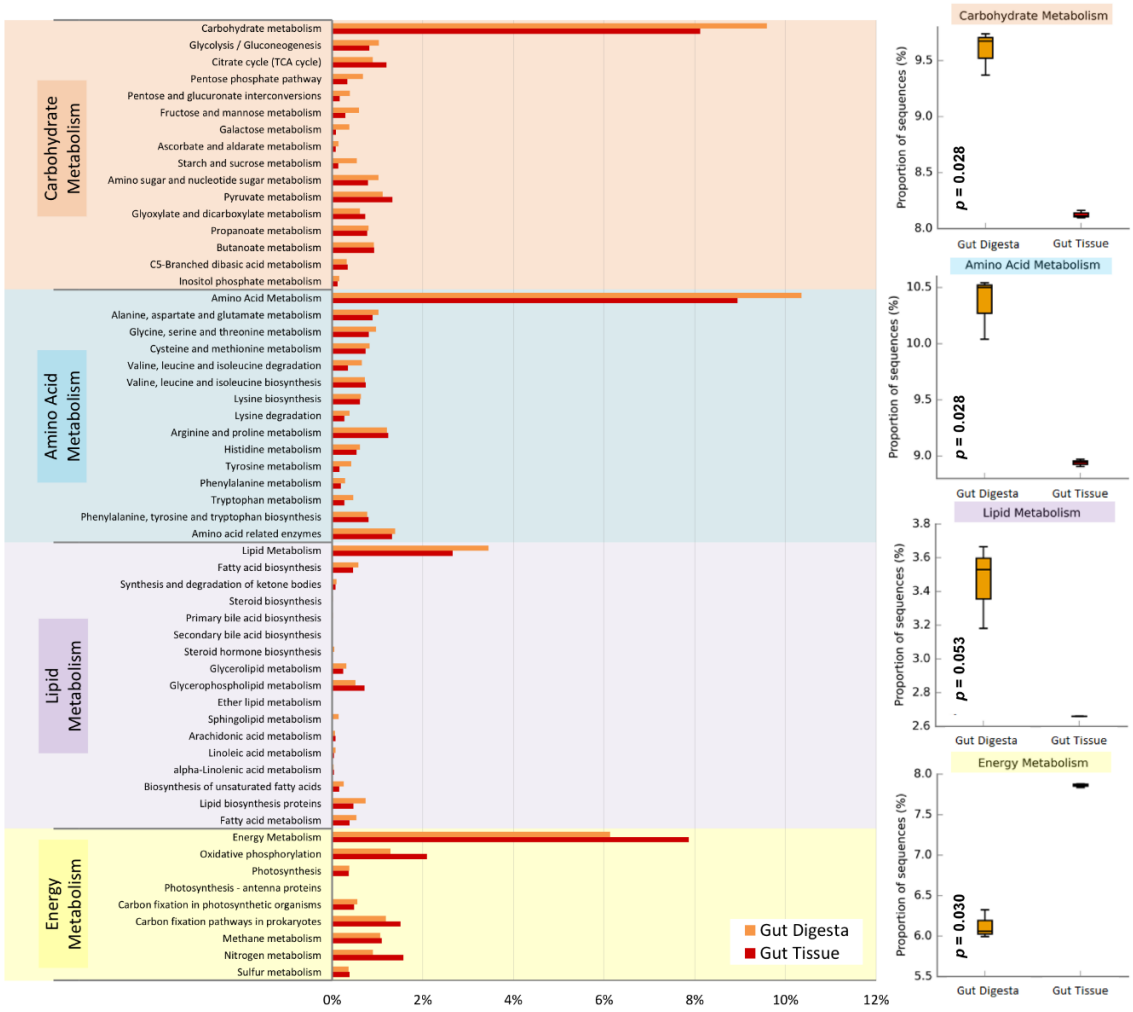


Figure 2-4: PICRUSt (v1.0.0) analysis of predicted metagenomes generated by using the 16S rRNA gene data of the gut tissue (*n* = 3) and gut digesta (*n* = 3) samples. OTUs were picked using closed-reference picking, as suggested by PICRUSt (v1.0.0), and merged with open-reference picked *de novo* OTUs (occurring at >100 per sample) which included each OTU's respective representative Greengenes ID. KEGG pathways were assigned (KO IDs) using the 'predict_metagenomes.py' module, and collapsed into hierarchical KEGG pathways (KEGG-Level-2 and 3). (A) The mean relative abundance of KEGG-Level-2 metadata categories are listed along with the associated KEGG-Level-3 pathways. (B) Box plots of the KEGG-Level-2 category of carbohydrate metabolism, amino acid metabolism, lipid metabolism and energy metabolism were generated using STAMP (v2.1.3) analytical software according to two-group statistics, using a two-sided Welch's t-test (not assuming equal variance) along with Benjamini–Hochberg FDR. Confidence intervals were selected as 95% (i.e.: 0.95), and P-value of each KEGG-Level-2 two-group analyses is listed in the respective box plot. Relative abundance data were graphed using Microsoft Excel software (Microsoft), and box plots generated using STAMP (v2.1.3).

metabolism. The groups compared included the gut digesta and gut tissues. The gut digesta revealed a heightened relative abundance of the KEGG-Level-2 categories corresponding to carbohydrate (9.59%; P = 0.028) and amino acid (10.39%; P = 0.028) metabolisms, as well as a marginally heightened abundance of lipid (3.45%; P = 0.053) metabolisms compared to the gut tissue. Conversely, the gut tissue displayed high abundances of pathways related to energy (7.87%; P = 0.030) metabolism (Figure 2-4). Investigations of the KEGG-Level-3 subcategories from carbohydrate metabolism assigned to the gut digesta showed various sugar metabolic processes, such as the pentose phosphate pathway, pentose glucoronate interconversion pathway and glycolysis. Starch, sucrose, fructose, mannose, galactose and amino acid sugar metabolisms were also pronounced as compared to the gut tissue. KEGG-Level-3 categories corresponding to amino acid metabolism showed the gut digesta to exceed the gut tissue in the degradation or biosynthesis of multiple amino acids (alanine, aspartate, glutamate, glycine, serine, threonine, cysteine, methionine, valine, leucine, isoleucine, lysine, histidine, tyrosine, phenylalanine and tryptophan). KEGG-Level-3 categories of lipid metabolisms also showed the gut digesta to contain pathways related to the biosynthesis of fatty acids, unsaturated fatty acids and glycerolipids, among others. From the gut tissue, KEGG-Level-3 pathways of energy metabolisms revealed relative abundances that surpassed the gut digesta in oxidative phosphorylation, prokaryotic carbon fixation, and methane, nitrogen and sulfur metabolisms.

DISCUSSION

Previous studies on the sea urchin digestive tract microbiota have described a distinct and pervasive bacterial profile that will colonize the apparent gut epithelium tissue, recognizably different from the bacterial community of the ingested feed, marine environment and pelleted gut digesta and egested fecal pellets (Guerinot and Patriquin 1981; Meziti et al.2007; Lawrence, Lawrence and Watts 2013). The microbial profile of the laboratory-cultured sea urchin *Lytechinus variegatus*, studied previously by our lab, revealed a near-exclusive occurrence of the order Campylobacterales associated with the gut tissue, of which oligotyping depicted *Arcobacter* sp., *Sulfuricurvum* sp. and *Arcobacter bivalviorum* (each with BLAST identities ≥90%) (Hakim et al.2015). The aforementioned microbial distribution was corroborated in the naturally occurring sea urchins of the current study, as the highly abundant representative sequences were assigned to Campylobacteraceae (93%) (Figure 2-1), and also determined to be *Arcobacter* sp., *Sulfuricurvum* sp. and *A. bivalviorum*. ANOSIM and subsequent OTU Bray–Curtis MDS cluster analysis revealed the microbial community of each gut tissue sample (n = 3) to cluster together (Figure 2-2), away from all other samples of the study. These results support a distinct gut tissue-associated bacterial profile, as reported in previous studies (Guerinot and Patriquin 1981; Meziti et al.2007). This is to note that the sample replicates used in this study exhibited a sufficient number of sequences to generate a reliable taxonomical identification of microbial communities, as well as their predictive functional attributes, which was supported by the statistical analyses of the

current study (Table 2-1; Fig 2-2). Previously, three or less replicate samples were reported in NextGen sequencing analyses comparing microbial composition in various ecosystems showing adequate power that was verified by relevant statistical methods (Hong et al.2015; Manzari et al.2015; Sha et al.2016). Since the heightened taxa of the gut tissue in this study showed a high BLAST similarity to *Arcobacter*, it is important to note that phylotypes related to *Arcobacter* have been observed in other marine invertebrates, such as the Chilean oyster *Tiostrea chilensis* (Romero et al.2002), shrimp *Rimicaris exoculata* (Durand et al.2010) and the hydrothermal vent-dwelling gastropod *Alviniconcha aff. Hessleri* (Suzuki et al.2005). To a much lesser extent, members of the representative class (Epsilonproteobacteria) have been observed in the cecum of mice (Gu et al.2013), and in humans (Eppinger et al.2004; Larsen and Dai 2015), in which both pathogenic and non-pathogenic roles have been described. However, it is unlikely that the Campylobacteraceae taxa observed in this study are antagonistic to the sea urchin *L. variegatus*, considering the near-dominant and consistent presence of these bacteria signifying host-selection (Hakim et al.2015), and the described non-detrimental association of this taxa described in the aforementioned marine invertebrates. The pharynx tissue also revealed a discernable microbial composition from the other sample types, and was found to consist of a high prevalence of Mollicutes (60%), belonging to phylum Tenericutes (Figures 2-1 and 2-2). Members of Mollicutes have previously been observed in the stomach of the eastern oyster, *Crassostrea virginica* (King et al.2012), as well as the terrestrial isopod, *Porcellio scaber*

(Wang et al.2004). The unique group-specific clustering of the microbial communities of the gut and pharynx tissues of *L. variegatus* was further evidenced through heatmap analysis (Figure 2-3).

The microbiota of the gut digesta and egested fecal pellets revealed *Vibrio* to be noticeably abundant in the naturally occurring sea urchin. Interestingly, Vibrio was not found to be significantly associated with the gut and pharynx tissues (<1%) (Figure 2-1). This taxon has previously been identified in various sea urchins (Unkles 1977; Guerinot et al.1982), particularly *L. variegatus* (Nelson et al.2010; Hakim et al.2015), which have been described to play a role in nitrogen fixation, as well as protein assimilation in gonadal tissues (Fong and Mann 1980; Guerinot et al.1982). Certain *Vibrio* spp. are considered stable and common associates of many marine invertebrates, and have been observed in copepod species from the Gulf of Maine (Moisander, Sexton and Daley 2015), various crabs (*Carcinus maenas* and *Hemigrapsus sanguineus*), mussels (*Mytilus edulis*) and zooplankton (Preheim et al.2011). The representative class (Gammaproteobacteria) has also been observed in a number of marine and freshwater fish (Roeselers et al.2011), with studies showing significant abundances of *Vibrio* in zebrafish (*Danio rerio*) (Roeselers et al.2011), the abalone (*Haliotis discus hannai*) (Tanaka et al.2004) and the farmed marine turbot fish (*Scophthalmus maximus*) (Xing et al.2013). Additionally, in this study, *Photobacterium* of family Vibrionaceae were found to be abundant in the gut digesta and egested fecal pellets. Members of this genus, specifically *Photobacterium lipolyticum*, have been identified in the herbivorous sea urchin

Paracentrotus lividus (Meziti et al.2007; Yeruham et al.2015) and other marine organisms (Gomez-Gil et al.2011), with a potential role in lipolytic activity (Seo et al.2005; Yoon et al.2005). Lastly, Flavobacterales, commonly found in the marine environment (Smith et al.2013), were the likely source of high abundance in the gut digesta and egested fecal pellets in the naturally occurring sea urchins of this study (Figure 2-1).

Considering the nutritive profile of seagrass, and the limitations of the innate gut digestive enzymes of the sea urchin, it has been proposed that the bacteria of the gastrointestinal tract contribute to the digestion of complex sugars and cellulose (Lasker and Giese 1954; García-Tello and Baya 1973; Unkles 1977; Becker et al.2009), as well as the metabolism or synthesis of necessary biomolecules for protein and lipid incorporation (Tysskt et al.1961; Fong and Mann 1980; Lawrence, Lawrence and Watts 2006; Arafa et al.2012). The gastrointestinal tract of this animal represents a physical compartmentalization of the gut digesta from the gut tissue, and are each accompanied by uniquely dissimilar microbial profiles. This would indicate disparate functional profiles along with those divergent microbial populations, suggesting an allocation of microbial metabolisms and digestive responsibilities occurring in the compartmentalized gut ecosystem of this animal. By using PICRUSt (Figure 2-4), a bioinformatics program that has shown considerable efficacy in the predictions of metabolic functions of microbial communities (Langille et al.2013), we were able to predict carbohydrate metabolic pathways, seeming to be essential for the digestion of starch and cellulose from turtlegrass consumed within its natural

habitat. This is supported by a previous study that reported bacterial digestion of complex carbohydrates from alginate within the sea urchin gut (Sawabe et al.1995; Nelson et al.2010), identifying members of *Vibrio*—a genus that appeared heightened in the gut digesta revealed in this study (Figure 2-1). PICRUSt analysis also showed protein and lipid metabolisms to be elevated in the gut digesta as compared to the gut tissue (Figure 2-4). The reliance on gut microbial communities for protein metabolisms in the sea urchin has been previously addressed by Fong and Mann (1980), who demonstrated that suppressed microbial growth in *Strongylocentrotus droebachiensis* by antibiotic treatment resulted in a significantly reduced incorporation of essential amino acids into the gonadal tissue. Also, from the current study, the observation of pathways related to lipid metabolism in the gut digesta is likely to be conducted by the microbial community, and such metabolisms have been supported by previous studies of the microbial involvement in lipid and fatty acid biosynthesis (Leo and Parker, 1966) in sea urchins *Psammechinus miliaris* (Cook et al.2000), *Paracentrotus lividus* (Arafa et al.2012) and other marine invertebrates (Phillips 1984).

Conversely, the bacteria of the gut tissue surpassed the gut digesta in the KEGG-Level-2 category of energy metabolism, which encompasses the subcategories of oxidative phosphorylation, carbon fixation, methane, nitrogen and sulfur metabolisms. Such metabolisms denote interdependency between the bacterial members associated with the gut tissue and other microbial communities of the sea urchin gut ecosystem (such as the gut digesta)

(Fischbach and Sonnenburg 2011). Bacteria colonizing the mucosal layer of the gastrointestinal tract in humans and other animals have been implicated in a myriad of roles in the development, maintenance and homeostatic preservation of the digestive tract, often outcompeting transient microbiota, as a product of a co-evolution of host and microbe (Rakoff-Nahoum et al.2004; Tlaskalová-Hogenová et al.2011; Schluter and Foster 2012; Wu and Wu 2012). Considering the near-exclusive abundance of Campylobacteraceae in the gut tissue, and the substantial relative abundance of energy metabolisms performed by this community, it appears that a mutualistic relationship may be occurring between the sea urchin *L. variegatus* and its selected gut mucosal resident, though the mechanisms of selection and the supposed benefit, if one exists, are unclear at this time.

In summary, the results of this study have revealed the gut microbial communities, with the highest taxonomic coverage, in the sea urchin *L. variegatus* from their natural habitat of the Gulf of Mexico. An enrichment of Campylobacteraceae (93%) was observed in the gut lumen, and is similar to that observed in *L. variegatus* held in culture and fed formulated diets. Given the near-exclusive abundance of Campylobacteraceae in the gut tissue, one can predict that the high energy metabolism observed by PICRUSt (v1.0.0) analysis is attributed by the members of this taxonomic group. In contrast, OTUs assigned to *Vibrio*, *Photobacterium*, *Propionigenium* and Flavobacteria were found to occur with high abundance in the gut digesta and egested fecal pellets. The innate digestive enzymes capable of processing complex carbohydrates, proteins

and lipids from diet (turtlegrass) are scarcely reported in sea urchins, and consequently, studies have implicated the bacteria in the gut ecosystem in executing such necessary metabolic processes (Leo and Parker 1966; Phillips 1984; Schlosser et al.2005; Arafa et al.2012). This indicates the bacterial community within the mucosally enveloped ingesta of the naturally occurring sea urchin likely to be involved, as determined by PICRUSt (v1.0.0), in the metabolism of carbohydrates, amino acids and lipids. Further metagenomics analyses would help substantiate the metabolic attributes of the gut microbiota in benefitting health and digestive physiology of the animal, as well as the community in their inhabited ecosystem.

FUNDING

We thank the UAB Institutional Animal Care and Use Committee for their support in this research. Also, we thank Edward Partridge, MD (Comprehensive Cancer Center; grant P30AR050948), Robert Kimberly, MD (Center for Clinical Translational Science; grant UL1TR000165), Michael Saag, MD (Center for AIDS Research; grant 5P30AI027767) and David Allison, PhD (UAB Nutrition Obesity Research Center (NORC); grant NIH P30DK056336) for providing support with the center grants and core facilities for this work. This work was supported by the Microbiome Resource at the UAB, School of Medicine, Comprehensive Cancer Center (P30AR050948), Center for AIDS Research (5P30AI027767) and Center for Clinical Translational Science (UL1TR001417). Animal husbandry was

supported in part by the Aquatic Animal Research Core, part of the UAB NORC (NIH P30DK056336).

ACKNOWLEDGEMENTS

We would like to thank Dr. Peter Eipers of the Department of Cell, Developmental and Integrative Biology, and Dr. Michael Crowley of the Heflin Center for Genomics Sciences at the University of Alabama at Birmingham (UAB), for their assistance in NextGen sequencing for this study. We would also like to thank UAB Institutional Animal Care and Use Committee (IACUC), and the Biology Department at UAB for logistics and graduate tuition and stipend support.

CONFLICT OF INTERESTS

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

REFERENCES

1. Altschul SF, Gish W, Miller W, et al. Basic local alignment search tool. *J Mol Biol*. 1990;215:403–10.

2. Andrews SF. FASTQC. A quality control tool for high throughput sequence data. 2010. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (5 July 2016, date last accessed)

3. Arafa S, Chouaibi M, Sadok S, et al. The influence of season on the gonad index and biochemical composition of the sea urchin *Paracentrotus lividus* from the Golf of Tunis. *Sci World J.* 2012;2012:1–8.

4. Bäckhed F, Ley RE, Sonnenburg JL, et al. Host-bacterial mutualism in the human intestine. *Science*. 2005;307:1915–20.

5. Becker PT, Samadi S, Zbinden M, et al. First insights into the gut microflora associated with an echinoid from wood falls environments. *Cah Biol Mar.* 2009;50:343.

6. Beddingfield SD, McClintock JB. Demographic characteristics of *Lytechinus variegatus* (Echinoidea:Echinodermata) from three habitats in a North Florida Bay, Gulf of Mexico. *Mar Ecol*. 2000;21:17–40.

7. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Roy Stat Soc B Met*. 1995;57:289–300.

8. Bray JR, Curtis JT. An ordination of the upland forest communities of southern Wisconsin. *Ecol Monogr*. 1957;27:325–49.

9. Brooks JM, Wessel GM. A diversity of yolk protein dynamics and function. *Recent Dev Cell Res*. 2003;1:1–30.

10. Caporaso JG, Kuczynski J, Stombaugh J, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*. 2010;7:335–6.

11. Caporaso JG, Lauber CL, Walters WA, et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *P Natl Acad Sci USA*. 2011;108:4516–22.

12. Caporaso JG, Lauber CL, Walters WA, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* 2012;6:1621–24.

13. Clarke KR. Non-parametric multivariate analyses of changes in community structure. *Aust J Ecol*. 1993;18:117–43.

14. Clarke KR, Gorley RN. Primer V5 (Plymouth Routines in Multivariate Ecological Research): User Manual/tutorial. Plymouth, UK: Primer-E; 2001.

15. Cock PJ, Fields CJ, Goto N, et al. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Res*. 2010;38:1767–71.

16. Cook EJ, Bell MV, Black KD, et al. Fatty acid compositions of gonadal material and diets of the sea urchin, *Psammechinus miliaris*: trophic and nutritional implications*. J Exp Mar Biol*. 2000;255:261–74.

17. Dawyndt P, De Meyer H, De Baets B. The complete linkage clustering algorithm revisited. *Soft Comput*. 2005;9:385–92.

18. de Cárcer DA, Denman SE, McSweeney C, et al. Evaluation of subsampling-based normalization strategies for tagged high-throughput sequencing data sets from gut microbiomes. *Appl Environ Microb*. 2011;77:8795–8.

19. DeSantis TZ, Hugenholtz P, Larsen N, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microb*. 2006;72:5069–72.

20. Durand L, Zbinden M, Cueff-Gauchard V, et al. Microbial diversity associated with the hydrothermal shrimp *Rimicaris exoculata* gut and occurrence of a resident microbial community. *FEMS Microbiol Ecol*. 2010;71:291–303.

21. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*. 2010;26:2460–1.

22. Eklöf J, De la Torre-Castro M, Gullström M, et al. Sea urchin overgrazing of seagrasses: a review of current knowledge on causes, consequences, and management. *Estuar Coast Shelf S*. 2008;79:569–80.

23. Eppinger M, Baar C, Raddatz G, et al. Comparative analysis of four Campylobacterales. *Nat Rev Microbiol*. 2004;2:872–85.

24. Fischbach MA, Sonnenburg JL. Eating for two: how metabolism establishes interspecies interactions in the gut. *Cell Host Microbe*. 2011;10:336–47.

25. Fong W, Mann K. Role of gut flora in the transfer of amino acids through a marine food chain. *Can J Fish Aquat Sci.* 1980;37:88–96.

26. García-Tello P, Baya A. Acerca de la posible función de bacterias agaroliticas aisladas del erizo blanco (*Loxechinus albus* (Mol.)) *Mus Nac Hist Nac Pub Oc*. 1973;15:3–8.

27. Gomez-Gil B, Roque A, Rotllant G, et al. *Photobacterium swingsii* sp. nov., isolated from marine organisms. Int J Syst Evol Micr. 2011;61:315–9.

28. Gordon A, Hannon G. Fastx-toolkit. FASTQ/A short-reads pre-processing tools. 2010. http://hannonlab.cshl.edu/fastx_toolkit/ (21 March 2016, date last accessed)

29. Gotelli NJ, Colwell RK. Estimating species richness. In: Magurran AE, McGill BJ, editors. Biological Diversity: Frontiers in Measurement and Assessment. Oxford: Oxford University Press; 2011. pp. 39–54.

30. Gu S, Chen D, Zhang JN, et al. Bacterial community mapping of the mouse gastrointestinal tract. *PLoS One*. 2013;8:e74957.

31. Guerinot M, Patriquin D. N2-fixing vibrios isolated from the gastrointestinal tract of sea urchins. Can J Microbiol. 1981;27:311–7.

32. Guerinot M, West P, Lee J, et al. *Vibrio diazotrophicus* sp. nov., a marine nitrogen-fixing bacterium. *Int J Syst Bacteriol.* 1982;32:350–7.

33. Hakim JA, Koo H, Dennis LN, et al. An abundance of Epsilonproteobacteria revealed in the gut microbiome of the laboratory cultured sea urchin, *Lytechinus variegatus*. *Front Microbiol*. 2015;6:1047.

34. Hamady M, Lozupone C, Knight R. Fast UniFrac: facilitating high-throughput phylogenetic analyses of microbial communities including analysis of pyrosequencing and PhyloChip data. *ISME J*. 2010;4:17–27.

35. Hendler G, Kier PM, Miller JE, et al. Sea Stars, Sea Urchins, and Allies: Echinoderms of Florida and the Caribbean. Vol. 390. Washington, DC: Smithsonian Institution Press; 1995.

36. Hill TC, Walsh KA, Harris JA, et al. Using ecological diversity measures with bacterial communities. *FEMS Microbiol Ecol*. 2003;43:1–11.

37. Holland ND. Digestive System. In: Lawrence JM, editor. Sea Urchins: Biology and Ecology. Vol. 38. UK: Elsevier; 2013. p. 119.

38. Hong C, Si Y, Xing Y, et al. Illumina MiSeq sequencing investigation on the contrasting soil bacterial community structures in different iron mining areas. *Environ Sci Pollut R*. 2015;22:10788–99.

39. Johannes R, Satomi M. Composition and nutritive value of fecal pellets of a marine crustacean. *Limnol Oceanogr*. 1966;11:191–7.

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

41. Kanehisa M, Goto S, Sato Y, et al. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res*. 2014;42:D199–205.

42. King GM, Judd C, Kuske CR, et al. Analysis of stomach and gut microbiomes of the eastern oyster (*Crassostrea virginica*) from coastal Louisiana, USA*. PLoS One*. 2012;7:e51475.

43. Koike I, Mukai H, Nojima S. The role of the sea urchin, *Tripneustes gratilla* (Linnaeus), in decomposition and nutrient cycling in a tropical seagrass bed. *Ecol Res*. 1987;2:19–29.

44. Kozich JJ, Westcott SL, Baxter NT, et al. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microb*. 2013;79:5112–20.

45. Krebs CJ. Ecological Methodology. Vol. 620. CA, USA: Benjamin/Cummings Menlo Park; 1999.

46. Kruskal J, Wish M. Quantitative Applications in the Social Sciences: Multidimensional Scaling. Vol. 11. Beverly Hills, CA: Sage; 1978.

47. Kumar R, Eipers P, Little RB, et al. Getting started with microbiome analysis: sample acquisition to bioinformatics*. Curr Protoc Hum Genet*. 2014;82:18.8.1–29.

48. Lang JM, Eisen JA, Zivkovic AM. The microbes we eat: abundance and taxonomy of microbes consumed in a day's worth of meals for three diet types. *PeerJ*. 2014;2:e659.

49. Langille MG, Zaneveld J, Caporaso JG, et al. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nat *Biotechnol*. 2013;31:814–21.

50. Larsen PE, Dai Y. Metabolome of human gut microbiome is predictive of host dysbiosis. *GigaScience*. 2015;4:1–16.

51. Lasker R, Giese AC. Nutrition of the sea urchin, *Strongylocentrotus purpuratus*. *Biol Bull.* 1954;106:328–40.

52. Lawrence J, Lawrence A, Watts S. Feeding, digestion, and digestibility of sea urchins. In: Lawrence JM, editor. Sea Urchins: Biology and Ecology. Vol. 38. UK: Elsevier; 2013. pp. 135–54.

53. Lawrence JM, Lawrence AL, Watts SA. Feeding, digestion, and digestibility. In: Lawrence JM, editor. Edible Sea Urchins: Biology and Ecology. Vol. 37. UK: Elsevier; 2006. p. 135.

54. Leo RF, Parker PL. Branched-chain fatty acids in sediments. *Science*. 1966;152:649–50.

55. Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microb*. 2005;71:8228–35.

56. Lozupone CA, Hamady M, Kelley ST, et al. Quantitative and qualitative β diversity measures lead to different insights into factors that structure microbial communities. *Appl Environ Microb*. 2007;73:1576–85.

57. Manzari C, Fosso B, Marzano M, et al. The influence of invasive jellyfish blooms on the aquatic microbiome in a coastal lagoon (Varano, SE Italy) detected by an Illumina-based deep sequencing strategy. *Biol Invasions*. 2015;17:923–40.

58. Marcon E, Scotti I, Hérault B, et al. Generalization of the partitioning of Shannon diversity. *PLoS One.* 2014;9:e90289.

59. McDonald D, Price MN, Goodrich J, et al. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J*. 2011;6:610–8.

60. Meziti A, Kormas KA, Pancucci-Papadopoulou M-A, et al. Bacterial phylotypes associated with the digestive tract of the sea urchin *Paracentrotus lividus* and the ascidian *Microcosmus* sp. *Russ J Mar Biol*. 2007;33:84–91.

61. Miyata T. Reducing overgrazing by sea urchins by market development. *Bull Fish Res Agen*. 2010;32:103–7.

62. Moisander PH, Sexton AD, Daley MC. Stable associations masked by temporal variability in the marine copepod microbiome. *PLoS One*. 2015;10:e0138967.

63. Moore HB, Jutare T, Bauer J, et al. The biology of *Lytechinus variegatus*. *B Mar Sci*. 1963;13:23–53.

64. Morgulis A, Coulouris G, Raytselis Y, et al. Database indexing for production MegaBLAST searches. *Bioinformatics*. 2008;24:1757–64.

65. Navas-Molina JA, Peralta-Sánchez JM, González A, et al. Advancing our understanding of the human microbiome using QIIME. *Method Enzymol*. 2013;531:371–444.

66. Nelson L, Blair B, Murdock C, et al. Molecular Analysis of gut microflora in captive-raised sea urchins (*Lytechinus variegatus*) *J World Aquacult Soc*. 2010;41:807–15.

67. Parks DH, Tyson GW, Hugenholtz P, et al. STAMP: statistical analysis of taxonomic and functional profiles. *Bioinformatics*. 2014;30:3123–4.

68. Phillips NW. Role of different microbes and substrates as potential suppliers of specific, essential nutrients to marine detritivores. *B Mar Sci.* 1984;35:283–98.

69. Pradheeba M, Dilipan E, Nobi E, et al. Evaluation of seagrasses for their nutritional value. *IJMS*. 2011;40:105.

70. Preheim SP, Boucher Y, Wildschutte H, et al. Metapopulation structure of Vibrionaceae among coastal marine invertebrates. *Environ Microbiol*. 2011;13:265–75.

71. Rakoff-Nahoum S, Paglino J, Eslami-Varzaneh F, et al. Recognition of commensal microflora by toll-like receptors is required for intestinal homeostasis. *Cell*. 2004;118:229–41.

72. Roeselers G, Mittge EK, Stephens WZ, et al. Evidence for a core gut microbiota in the zebrafish. *ISME J*. 2011;5:1595–608.

73. Romero J, Garcia-Varela M, Laclette JP, et al. Bacterial 16S rRNA gene analysis revealed that bacteria related to *Arcobacter* spp. constitute an abundant and common component of the oyster microbiota (*Tiostrea chilensis*) *Microbial Ecol*. 2002;44:365–71.

74. Sauchyn LK, Lauzon-Guay J-S, Scheibling RE. Sea urchin fecal production and accumulation in a rocky subtidal ecosystem. *Aquat Biol*. 2011;13:215–23.

75. Sawabe T, Oda Y, Shiomi Y, et al. Alginate degradation by bacteria isolated from the gut of sea urchins and abalones. *Microb Ecol*. 1995;30:193–202.

76. Schlosser SC, Lupatsch I, Lawrence JM, et al. Protein and energy digestibility and gonad development of the European sea urchin *Paracentrotus lividus* (Lamarck) fed algal and prepared diets during spring and fall. *Aquacult Res*. 2005;36:972–82.

77. Schluter J, Foster KR. The evolution of mutualism in gut microbiota via host epithelial selection. *PLoS Biol*. 2012;10:e1001424.

78. Seo HJ, Bae SS, Lee J-H, et al. *Photobacterium frigidiphilum* sp. nov., a psychrophilic, lipolytic bacterium isolated from deep-sea sediments of Edison Seamount. *Int J Syst Evol Micr*. 2005;55:1661–6.

79. Sha Y, Liu M, Wang B, et al. Gut bacterial diversity of farmed sea cucumbers *Apostichopus japonicus* with different growth rates. *Microbiology*. 2016;85:109–15.

80. Shade A, Handelsman J. Beyond the Venn diagram: the hunt for a core microbiome. *Environ Microbiol*. 2012;14:4–12.

81. Shannon CE. A mathematical theory of communication. *Bell Syst Tech J.* 1948;27:379–423.

82. Simpson EH. Measurement of diversity. *Nature*. 1949;163:688.

83. Smith MW, Allen LZ, Allen AE, et al. Contrasting genomic properties of free-living and particle-attached microbial assemblages within a coastal ecosystem. *Front Microbiol*. 2013;4:120.

84. Suzuki Y, Sasaki T, Suzuki M, et al. Novel chemoautotrophic endosymbiosis between a member of the Epsilonproteobacteria and the hydrothermal-vent gastropod *Alviniconcha aff. hessleri* (Gastropoda: Provannidae) from the Indian Ocean. *Appl Environ Microb*. 2005;71:5440–50.

85. Tanaka R, Ootsubo M, Sawabe T, et al. Biodiversity and in situ abundance of gut microflora of abalone (*Haliotis discus hannai*) determined by culture-independent techniques. *Aquaculture*. 2004;241:453–63.

86. Tlaskalová-Hogenová H, Štěpánková R, Kozáková H, et al. The role of gut microbiota (commensal bacteria) and the mucosal barrier in the pathogenesis of inflammatory and autoimmune diseases and cancer: contribution of germ-free and gnotobiotic animal models of human diseases. *Cell Mol Immunol*. 2011;8:110–20.

87. Tysskt C, Mailloux M, Brisou J, et al. Sur la Microflore Normale de L'oursin Violet Des Cotes Algéroises (*Paracentrotus lividus*, Ltmk) *Archives de l'Institut Pasteur d'Algérie.* 1961;39:271.

88. Unkles S. Bacterial flora of the sea urchin *Echinus esculentus*. *Appl Environ Microb*. 1977;34:347–50.

89. Wang Q, Garrity GM, Tiedje JM, et al. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microb.* 2007;73:5261–7.

90. Wang Y, Stingl U, Anton-Erxleben F, et al. "*Candidatus Hepatoplasma crinochetorum*," a new, stalk-forming lineage of Mollicutes colonizing the midgut glands of a terrestrial isopod. *Appl Environ Microb*. 2004;70:6166–72.

91. Watts SA, McClintock JB, Lawrence JM. Lytechinus. In: Lawrence JM, editor. Sea Urchins: Biology and Ecology. Vol. 38. UK: Elsevier; 2013. pp. 475–86.

92. Welch BL. The significance of the difference between two means when the population variances are unequal. *Biometrika*. 1938;29:350–62.

93. Wu H-J, Wu E. The role of gut microbiota in immune homeostasis and autoimmunity. *Gut Microbes*. 2012;3:4–14.

94. Xing M, Hou Z, Yuan J, et al. Taxonomic and functional metagenomic profiling of gastrointestinal tract microbiome of the farmed adult turbot (*Scophthalmus maximus*) *FEMS Microbiol Ecol.* 2013;86:432–43.

95. Yeruham E, Rilov G, Shpigel M, et al. Collapse of the echinoid *Paracentrotus lividus* populations in the Eastern Mediterranean—result of climate change? *Sci Rep*. 2015;5:1–6.

96. Yoon J-H, Lee J-K, Kim Y-O, et al. *Photobacterium lipolyticum* sp. nov., a bacterium with lipolytic activity isolated from the Yellow Sea in Korea. *Int J Syst Evol Micr*. 2005;55:335–9.

97. Zhao W, Wang Y, Liu S, et al. The dynamic distribution of porcine microbiota across different ages and gastrointestinal tract segments*. PloS One*. 2015;10:1–13.

98. Ziegler A, Mooi R, Rolet G, et al. Origin and evolutionary plasticity of the gastric caecum in sea urchins (Echinodermata: Echinoidea) *BMC Evol Biol*. 2010;10:1–32.

99. Zieman JC. Ecology of the seagrasses of South Florida: a community profile. Technical report. University of Virginia Deptartment of Environmental Sciences; 1982.

CHAPTER III: THE PURPLE SEA URCHIN *STRONGYLOCENTROTUS PURPURATUS* DEMONSTRATES A COMPARTMENTALIZATION OF GUT BACTERIAL MICROBIOTA, PREDICTIVE FUNCTIONAL ATTRIBUTES, AND TAXONOMIC CO-OCCURRENCE

by

JOSEPH A. HAKIM, JULIE B. SCHRAM, AARON W.E. GALLOWAY, MICHAEL R. CROWLEY, CASEY D. MORROW, STEPHEN A. WATTS, AND ASIM K. BEJ

MDPI Microorganisms, Jan. 2019, Vol. 7(2)*,* PMID: 30691133

Format adapted and errata corrected for dissertation

ABSTRACT

The sea urchin *Strongylocentrotus purpuratus* (order Camarodonta, family Strongylocentrotidae) can be found dominating low intertidal pool biomass on the southern coast of Oregon, USA. In this case study, three adult sea urchins were collected from their shared intertidal pool, and the bacteriome of their pharynx, gut tissue, and gut digesta, including their tide pool water and algae, was determined using targeted high-throughput sequencing (HTS) of the 16S rRNA genes and bioinformatics tools. Overall, the gut tissue demonstrated *Arcobacter* and *Sulfurimonas* (Epsilonproteobacteria) to be abundant, whereas the gut digesta was dominated by *Psychromonas* (Gammaproteobacteria), *Propionigenium* (Fusobacteria), and Flavobacteriales (Bacteroidetes). Alpha and beta diversity analyses indicated low species richness and distinct microbial communities comprising the gut tissue and digesta, while the pharynx tissue had higher richness, more closely resembling the water microbiota. Predicted functional profiles showed Kyoto Encyclopedia of Genes and Genomes (KEGG) Level-2 categories of energy metabolism, membrane transport, cell motility, and signal transduction in the gut tissue, and the gut digesta represented amino acid, carbohydrate, vitamin and cofactor metabolisms, and replication and repair. Co-occurrence network analysis showed the potential relationships and key taxa, such as the highly abundant *Arcobacter* and *Propionigenium*, influencing population patterns and taxonomic organization between the gut tissue and digesta. These results demonstrate a trend of microbial community integration,

allocation, predicted metabolic roles, and taxonomic co-occurrence patterns in the *S. purpuratus* gut ecosystem.

INTRODUCTION

The purple sea urchin *Strongylocentrotus purpuratus* (order Camarodonta, family Strongylocentrotidae) inhabits the rocky tide pools along the North-East Pacific from Alaska to Baja Mexico. *S. purpuratus* is primarily herbivorous, which tempers the growth of marine vegetation and plays an important role in shaping the dynamic population patterns in their marine ecosystem [1,2,3,4,5,6]. The low intertidal tide pools on the southern Oregon coast are dominated by *S. purpuratus* and are interspersed with mosaics of tufted algae and invertebrate assemblages representing multiple phyla [7]. The microhabitats of these tide pools are influenced by the feeding activity of the inhabiting sea urchins [7]. The sea urchins present unique digestive physiology in a straightforward model and offer an evolutionary context to fundamental biological and physiological processes occurring in higher deuterostome organisms [8]. In general, the pharynx is enclosed within the Aristotle's Lantern, which is a pentamerally symmetric mastication apparatus of five tooth-like structures that assist in scraping and releasing intracellular nutrients from algae [9]. The pharynx tissue contains specialized mucus cells that contribute to the formation of a mucous envelope of ingested feed [10], forming individual pellets of gut digesta [11]. This gut digesta pellet formation has been considered an evolved digestive strategy of this organism, likely as a result of water flow dynamics in the gut lumen

environment [12]. The role of gut bacteria in host health and digestion have been of interest beginning with the work of Lasker and Giese [9], who isolated gut bacteria from the gut digesta of *S. purpuratus*, showing the potential for these bacteria to digest polysaccharides from algal sources. In a separate study, bacteria isolated from the sea urchins *S. intermedius* and *S. nudus* demonstrated a similar algynolytic activity [13]. The importance of gut bacteria in sea urchin host health was further supported in *S. droebachiensis*, in which microbial suppression through antibiotics showed a reduced capacity for host incorporation of essential amino acids [14].

The microbial communities of the sea urchin gut pellets also play an important role in the biogeochemical cycles of the marine environment. In previous studies, it has been shown that the microbial community composition and their metabolic processes in the gut digesta remains stable following egestion into the environment [15,16,17]. For example, studies examining the chemical composition of *S. droebachiensis* egesta through flash combustion have shown increased in lipid, nitrogen, and organic carbon, and decreases in the carbon: nitrogen ratio, indicating the metabolic importance of the bacterial communities in the degradation and transformation of the contents within the pellets into a nutrient-rich food source for nearby marine organisms [17,18,19,20]. Additionally, it has been suggested that urchin gut microbiota are responsible for differences in algal digestion and synthesis of essential long chain fatty acids in both *S. purpuratus* and *S. droebachiensis* [20]. The pelleted egesta has also been considered as a mode for the dispersion of sea urchin gut

microbiota into their environment [17,21]. Most of the early studies of the potential role of microbial communities in digestive processes of sea urchin gut ecosystem were conducted by culture-dependent methods [9]. However, recent advancement of the culture-independent method of high-throughput sequencing (HTS) of 16S rRNA genes from the metacommunity DNA has been shown to provide gut microbial community composition with high taxonomic coverage, including their potential metabolic functions [22,23]. Recently, the application of HTS on the V4 hypervariable segment of the 16S rRNA gene of the gut bacteriome of *Lytechinus variegatus* from the U.S. Gulf of Mexico [15,16,24] demonstrated distinct microbial community compositions between the gut tissue and gut digesta. Specifically, representative taxa from class Epsilonproteobacteria (assigned as *Arcobacter*/*Sulfuricurvum* through the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST)) dominated the gut tissue, whereas Gammaproteobacteria (namely *Vibrio*) were heightened in the digesta [15]. Additionally, predictive functional profiling of these compartmentalized microbial communities showed energy metabolisms such as oxidative phosphorylation, carbon fixation, nitrogen, methane, and sulfur metabolisms to be heightened in the gut tissue, compared to carbohydrate, amino acid, and lipid metabolisms in the digesta [16].

Such HTS technology and bioinformatics analyses applied to the gut ecosystem of the naturally occurring sea urchin *S. purpuratus* can help establish a comprehensive microbial community composition and provide crucial information into the gut bacterial taxa and likely functions performed as they

relate to host health and digestion. In this study, we have elaborated the microbial profiles of the gut tissue, pharynx tissue, and mucous enveloped gut digesta of *S. purpuratus* collected from their natural rocky tide pool habitat on the coastline of Oregon, USA. In addition, samples of the tide pool seawater and adjacent algal community were collected and also analyzed for comparison with the gut tissue and digesta. We used HTS and bioinformatics tools to analyze the community composition, patterns of microbial taxa allocation in the gut environment, and the predicted metabolic functions of the bacterial microbiota in the gut ecosystem. These data were further refined using Phylogenetic Tools for Analysis of Species-level Taxa (PhyloToAST v1.4.0) [25] alongside Quantitative Insights into Microbial Ecology (QIIME v1.9.1) [26] to condense redundant taxonomic groups. This allowed increased resolution of microbial taxonomic groups to the species level and enhanced beta diversity inference. Additionally, the keystone taxa (herein “key” taxa) of the gut ecosystem were elaborated through topological analysis of Co-occurrence Network inferences (CoNet v1.1.1) [27,28,29] based on criteria described in Berry and Widder [30,31]. The results of this baseline case study demonstrate the microbial composition and associated functional capacity within the compartmentalized gut system of this evolutionarily and ecologically significant purple *S. purpuratus* sea urchin species.

MATERIALS AND METHODS

Collection and Sample Preparation of S. Purpuratus

Adult *S. purpuratus* sea urchins (UR; *n* = 3) were collected from within the same natural rocky tide pool habitat at Cape Arago, Oregon (43°18′14.3″N 124°24′05.1″W) in September 2016, from within a 1 m^2 sampling plot (Figure 3-1), under permit is: #20366 (Oregon Department of Fish and Wildlife). Sea urchins were measured and sexed, which was followed by tissue dissection performed at the Oregon Institute of Marine Biology (OIMB) in Coos County, Oregon. For each sea urchin, an incision was made into the test area surrounding the Aristotle's Lantern mastication structure using sterilized instruments, and the test was cut radially to expose the internal digestive tissue. The pharynx, which was enclosed by the Aristotle's Lantern, was separated from the gut tissue and collected. The remaining digestive tissue (gut tissue) was gently rinsed with sterile phosphate buffered saline water (1x PBS, pH 7.4) (Fisher Scientific, Hampton, NH, USA), and the contents (gut digesta) were collected. The whole gut tissue was collected separately from the voided gut digesta. Replicate seawater samples (water; *n* = 3) (1 L) from each tide pool was vacuum filtered separately through 0.22 µm filter paper (EMD Millipore Corporation, Danvers, MA, USA). The grazed-upon algal communities (algae; *n* = 3) immediately surrounding the sea urchins were also collected as the general food source and used in this study. A total of 15 samples (pharynx, *n* = 3; gut tissue, *n* = 3; gut digesta, *n* = 3; water, *n* = 3; and algae, *n* = 3) were placed into 95% (v/v) ethanol [32], flash frozen in liquid nitrogen, and shipped to the

University of Alabama at Birmingham (UAB), where they were preserved at −20 °C until used. Research performed under the Institutional Animal Care and Use Committee (IACUC-10043).

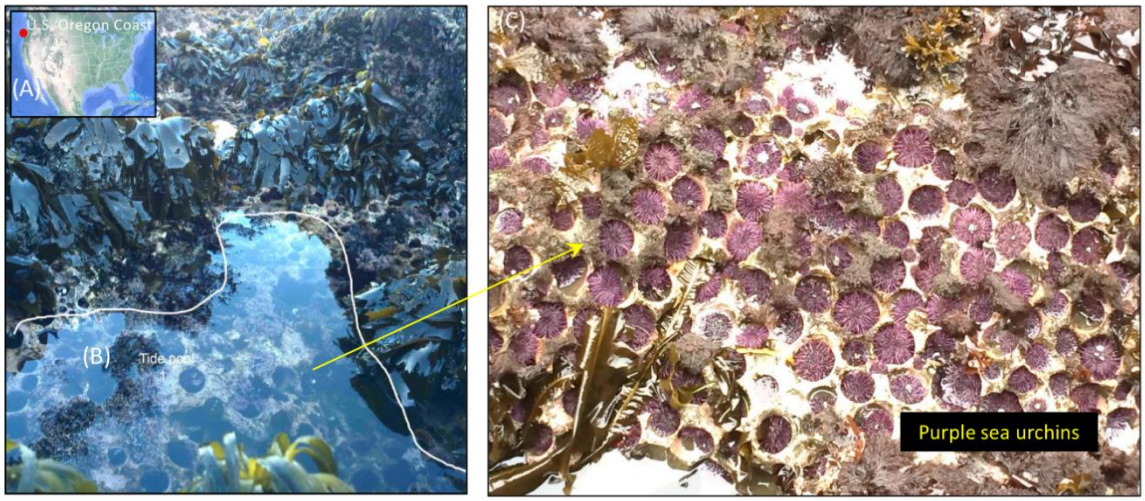


Figure 3-1: Sample collection site of *S. purpuratus* (purple sea urchins) from their natural rocky tide pool habitat along the coast of Oregon (43°18′14.3″N 124°24′05.1″W). (A) Satellite image of the collection site (red marker) provided through Google Earth Pro (v.7.3.2.5491) (Data SIO, NOAA, US Navy NGA, GEBCO, Image Landsat/Copernicus; US Dept. of State Geographer; image date: December 2015). (B) Overview of the tide pool collection site (labeled as tide pool 1) showing naturally occurring sea urchins. (C) Sea urchin congregates with the algal food source in view. Photographs by J.B Schram.

Community DNA Extraction, Illumina MiSeq Sample Preparation, and High-Throughput Sequencing

The metacommunity DNA from each sample was purified using the Fecal DNA isolation kit (Zymo Research, Irvine, CA, USA; catalog no. D6010), and an amplicon library of the metacommunity V4 hypervariable region (V4) of the 16S rRNA gene was created using uniquely barcoded DNA oligonucleotide primers adapted from the Earth Microbiome Project (www.earthmicrobiome.org) [33,34,35]. These primers consisted of the upstream nucleotide sequence for

hybridization to the Illumina MiSeq flow-cell surface (underlined), a “pad” region (italicized), and a “linker” region (bolded). The forward primer (515F) for the V4 segment was: 5′-AATGATACGGCGACCACCGAGATCTACACTATGGTAATTGTGTGCCAGCMGCCGCGGTAA-3′. The reverse primer for the V4 segment (modified from 806R) also included a unique barcode (6 N’s) region and was as follows: 5′-CAAGAGAAGACGGCATACGAGATNNNNNNAGTCAGTCAGCCGGACTACHVGGGTWTCTAAT-3′ (Eurofins Genomics, Inc., Huntsville, AL, USA) [35,36]. Polymerase chain reaction (PCR) amplification was performed using the LongAmp Taq PCR Kit (New England Biolabs, Ipswich, MA, USA; catalog no. E5200S) at a total reaction volume of 50 µL with the following reagents: 10 µL of 5× Reaction Buffer; 1.5 µL of each dNTPs (200 µM); 2 µL of each oligonucleotide primer (1.5 µM); 1.5 µL of LongAmp® enzyme (5 U); 30 µL of template DNA (2–5 ng/µL); and 3 µL of sterile H2O. The PCR proceeded with an initial denaturation at 94 °C for 1 min followed by 32 cycles of amplification of which each cycle consisted of denaturation at 94 °C for 30 sec, primer annealing at 50 °C for 1 min, and primer extension at 65 °C for 1 min, followed by the final extension at 65 °C for 3 min and a final hold at 4 °C. An amplicon fragment of approximately 380 bases was visualized through an ultraviolet (UV) transilluminator (Photodyne, Inc., Los Angeles, CA, USA) and excised with a sterile scalpel following electrophoresis through a 1.0% (w/v) Tris-borate-EDTA (TBE)/agarose gel [37]. The excised DNA fragments were purified using the QIAquick Gel Extraction Kit (Qiagen Inc., Venlo, Limburg; catalog no. 28704). PicoGreen dye (Life

Technologies, Grand Island, NY, USA) was used to quantify each sample to adjust the concentration to 4 nM [35]. HTS was performed using the Illumina MiSeq platform [35,36], incorporating the 250 base paired-end kits from Illumina specific to the V4 region of the 16S rRNA gene.

Quality Assessment and Filtering

The raw sequence reads generated by HTS on the Illumina MiSeq platform were demultiplexed and converted to FASTQ format [38]. The read quality was evaluated using FastQC [39], and quality reads with 80% of bases at Q score >33 were retained for downstream analysis using the “fastx_trimmer” command from the FASTX Toolkit [35,40]. Then, the paired-ends were merged using USEARCH [41], and pairs with <50 base overlap and/or over 20 mismatching nucleotides were filtered. Read quality was again assessed after filtering using FASTQC, chimeric sequences were identified and removed using USEARCH [41]. Additionally, with the newly established bioinformatics techniques presented in the QIIME2 package (v2018.11) [42], an alternative approach to filtering and merging the paired-end sequence data was implemented. To do this, a “denoising” strategy was used based on the Poisson distribution through the Divisive Amplicon Denoising Algorithm program (DADA2, v1.10) [43,44]. This was performed utilizing the “qiime dada2 denoise-paired” module on the demultiplexed sequence data, with a truncation set at 250 bases for the forward and reverse reads. The sequence reads corresponding to each

sample of this study have been deposited in NCBI SRA for public access (Bioproject number PRJNA504890).

Taxonomic Distribution

The resultant quality assessed sequence files were processed using QIIME (v1.9.1) [26] along with PhylotoAST (v1.4.0) [25] to condense redundant operational taxonomic units (OTUs) [45]. First, OTUs were selected at a 97% sequence similarity threshold using the default UCLUST algorithm option in QIIME (v1.9.1) [41]. Representative OTU sequences were then selected using the “most_abundant” option, and taxonomy was assigned to the representative sequences at a 60% confidence threshold using the Ribosomal Database Project (RDP) classifier [46], trained with the GreenGenes reference database (v13.8) [47,48]. At this stage, OTUs occurring at less than 0.0005% average abundance across all samples in the study were filtered [49,50,51,52,53]. Then by using the PhyloToAST (v1.4.0) workflow, the species-level resolution was enhanced using the “assign_taxonomy_by_blast _result.py” command to assign taxonomy through BLAST [45] to the GreenGenes (v13.8) database, and redundant OTUs were merged through the “condense_workflow.py” command [25]. Variation in the read-depth was accounted for by subsampling of the condensed OTU table using both the median and minimum read count values across all samples as described in de Carcer et al. [54] through the “single_rarefaction.py” command in QIIME (v1.9.1), and both subsampled OTU tables were assessed for downstream analysis. Additionally, for the top 100 taxa determined in the rarefied

OTU table, the representative sequences were extracted and aligned to multiple databases using the SILVA ACT: Alignment, Classification and Tree Service (www.arb-silva.de/aligner) [55]. For this analysis, the SSU (Small Sub-Unit) category was selected, and a minimum similarity identity was set to 0.9, with 20 neighbors per query sequence. Sequences below an identity threshold of 70% were discarded. For taxonomic identification, the least common ancestor (LCA) method was used, and the databases selected included GreenGenes [47,48], Ribosomal Database Project (RDP) [56], and SILVA [57]. Lastly, for the alternative QIIME2 (v2018.11) method, the denoised and merged sequence data was used to generate representative sequences with the “qiime feature-table tabulate-seqs” command. These representative sequences, herein referred as amplicon sequence variants (ASVs), were assigned taxonomic identities through the “qiime feature-classifier” command utilizing the “classify-sklearn” option [58] against the GreenGenes (v13.8) database.

Alpha Diversity

The PhyloToAST (v1.4.0) condensed and minimum-count subsampled OTU table (herein, rarefied OTU table) was used to determine the taxonomic distribution and alpha diversity metrics of each sample. The rarefied OTU table was merged according to biological replicates and used to create the relative abundance graph of phyla represented at >1% abundant, as well as the top 100 most resolved taxonomic identities across all sample groups using Microsoft Excel Software (Seattle, WA, USA). Taxa represented at >1% in the gut system

(gut tissue and digesta) were also visualized, with standard deviations calculated through STAMP (v2.1.3) [59]. Shannon [60,61,62], and Simpson [61,63] diversity measurements were determined through the “alpha_diversity.py” command in QIIME (v1.9.1). These values were plotted as a kernel density estimator-smoothed histogram using the “diversity.py” command through PhyloToAST (v1.4.0), to show both the diversity value and the range of underlying data points (density) for each sample group. Kruskal–Wallis H-tests were performed for the five groups to show the alpha diversity variation between groups at a significance value of $p = 0.1$ [64].

Beta Diversity

For beta diversity, the rarefied OTU table was used to determine the Bray–Curtis distance matrix values [65]. These values were also used to calculate significant grouping among biological replicates ($n = 3$) through an analysis of similarity (ANOSIM) and multivariate analysis of variance (Adonis) of groups, both set at 999 permutations, determined through the QIIME (v.1.9.1) “compare_categories.py” module utilizing the Vegan (v2.4.3) R package implementation of the statistical methods [66,67,68]. Additionally, this ANOSIM and Adonis analysis was performed on each OTU table generated in this study (5 total), which included the unfiltered OTU table, filtered OTU table (<0.0005%), PhyloToAST (v1.4.0) condensed OTU table, condensed median, and condensed minimum subsampled OTU tables. Visualization of beta diversity trends was performed using Plymouth Routines in Multivariate Ecological Research

(PRIMER-6) software (Primer-E Ltd, Plymouth Marine Laboratory, Plymouth UK, v6.1.2) [67]. In PRIMER-6, a 2D multidimensional scale (MDS) plot was generated using the Bray–Curtis distance matrices, to show variation between each sample, along with an overlay of Bray–Curtis similarity values [69]. A dendrogram was also generated based on clustering by group average [69]. A 2D MDS and dendrogram cluster analysis was also performed on the top 100 OTUs and the remaining rare OTUs, to show the contributions of both the heightened and rare taxa to the observed sample community diversity and cluster patterns. In addition to the Bray–Curtis based analyses described above, the rarefied OTU table was used to determine the weighted and unweighted Unifrac distances [70], which was calculated through the "beta_diversity_through_plots.py" module of QIIME (v1.9.1), and used to generate the 3D principal coordinates analysis (PCoA) plots through the "PCoA.py" command in PhyloToAST (v1.4.0). These values were also used to calculate the ANOSIM and Adonis metrics for group analyses as previously described, and uploaded into PRIMER-6 to generate the dendrogram based on group average. Heatmap analysis was performed using the rarefied OTU table in R (v3.3.2), incorporating the heatmap.2 function from gplots (v3.0.1) package [71]. In brief, the associated sample group dendrogram was created through the Vegan (v2.4.3) package [68] using the Bray–Curtis distance metric of the grouped biological replicate count data and clustered according to the group average algorithm. Microbial taxa represented at <1% of the total dataset were filtered from the heatmap. A color palette was selected using the RColorBrewer

package [72], and the relative abundances were shown for each taxon across all sample groups (black bar lines). Linear discriminant analysis (LDA) effect size (LEfSe) analysis was used to determine the taxa contributing to the effect size between the compartmentalized gut microbial communities of the gut tissue (n = 3) and digesta (n = 3) [73]. This analysis was performed through the Hutlab Galaxy web application (huttenhower.sph.harvard.edu/galaxy/), and incorporated the non-parametric Kruskal-Wallis sum-rank test for significant differential abundance set at a significance of p = 0.05 [64], followed by LDA to estimate effect size at log(10) values [73,74]. The results were plotted to show those taxa that demonstrated an LDA of ±3 for effect size.

Predicted Functional Analysis

The functional capacity associated with the microbial communities of the gut tissue and digesta was determined using the Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt v1.1.2) package [75] and analyzed in STAMP (v2.1.3) [59]. For this analysis, an OTU table was constructed by the “pick_closed_reference_otus.py” strategy to ensure representative taxonomic information through the GreenGenes (v13.8) database as suggested in PICRUSt [47,48,75]. The resultant OTU table was normalized by copy number, and Kyoto Encyclopedia of Genes and Genomes (KEGG) Orthology (KO) Ids were predicted along with the weighted Nearest Sequenced Taxon Index (NSTI) values for the confidence of predictions using the “predict_metagenomes.py” command. The assigned functional categories were

then collapsed into levels (KEGG-Level-2, and 3) through the "categorize_by_function.py" command. The KEGG-Level-2 and 3 profiles were uploaded into STAMP (v2.1.3) for two-group scatter plot analysis, to determine the metabolic categories that are preferentially enriched in each group at two levels of hierarchical classification. Additionally, LEfSe analysis [73] was performed on the KO Ids using an LDA score of ± 2.4, again utilizing the non-parametric Kruskal–Wallis sum-rank test for significant differential abundance set at a significance of $p = 0.05$ [64] and LDA for effect size using log(10) values [73,74], to demonstrate those categories contributing most to functional profile dissimilarity.

Co-Occurrence Analysis of Microbial Taxa

Significant co-occurrence patterns occurring between the microbial communities of the gut tissue and the gut digesta were determined using Co-occurrence Network inference (CoNet v1.1.1) [27,28,29]. To do this, the rarefied OTU data were uploaded into Cytoscape (v3.6.0) [76] through the CoNet (v1.1.1) plugin with taxa assigned to sample type (gut tissue and gut digesta). Links between higher level taxa were not explored and a parent-child exclusion was applied. Taxonomic entries with a cumulative group sum of 200 and at least 2/3 of samples containing non-zero values were kept [27,28,29,77]. Significant co-occurrences between taxa were determined by utilizing the Pearson [78,79], Spearman [80], Bray–Curtis [65], Kullback–Leibler [81], and mutual information similarity [82], with a 10^{-8} pseudo-count [27,28,29,77]. The 200 highest (most

positive) and lowest (most negative) edges were selected and merged by the union approach using the mean value [77]. The multi-edge scores were shuffled row-wise at 100 permutations (for null distributions), followed by bootstrapping at 100 permutations (for randomizations). The p-values of the multi-edges assigned to node pairs were merged using the Brown method [83], with unstable edges filtered out, and the corrected significance value (*q*-value) was determined with a threshold set at $p < 0.05$ for significance [27,28,29].

The final network was constructed in Cytoscape (v3.6.0) using the radial layout algorithm in the yFiles plugin (v1.0) [84], and topological parameters were determined by NetworkAnalyzer (v2.7) [85] using an undirected approach. Node sizes were scaled to their group abundance, colored according to phylum (class for Proteobacteria), and assigned a shape according to group membership (circle for gut tissue, “-gut”; diamond for gut digesta; “-dig”). The edges were scaled by the *q*-value and colored according to their positive (co-presence; green) and negative (co-exclusion; red) association. Based on the topological features determined through NetworkAnalyzer (v2.7), those nodes tending to have a high degree (number of edges), closeness centrality, and low betweenness centrality have been referred to as key taxa as described by Berry and Widder et al. [30] [77,86,87]. These features were plotted as a scatter plot (y = closeness centrality; x = betweenness centrality; node size is scaled to degree) through Microsoft Excel Software (Seattle, WA, USA). The top 10 nodes based on their closeness centrality values were selected as likely key taxa.

RESULTS

Environmental Conditions and Sea Urchin Measurements

The tide pool location sea water conditions were determined to have a salinity of 30.19 ppt with a 7.7 pH. The dissolved oxygen content of the tide pool water was determined to be 59% and the temperature was determined to be 13.1 °C. The sea urchins of this study weighed between 38.98–53.35 g and had a mean diameter of 5.1–5.6 cm, a height of 2.6–2.8 cm, and a spheroid volume of 38.4–41.6 cm^3. The sexes of the sea urchins for the study were determined as UR1 = F, UR2 = F, and UR3 = M.

Environmental Conditions and Sea Urchin Measurements

The total sequences generated through Illumina MiSeq-based HTS of the bacterial 16S rRNA gene of the 15 samples of the study generated a total of 1,714,746 forward and reverse reads (Table 3-1). Quality checking and trimming using the FASTX Toolkit, followed by merging of the forward and reverse sequences resulted in 1,249,827 total reads. Grouping of biological replicate data (n = 3) showed the following total sequence read counts: algae (340,438), gut digesta (221,684), gut tissue (231,854), pharynx (194,640), and water (261,211). Clustering of sequences into OTUs and taxa assignment revealed a total of 44,664 unique assignments. Filtering of rare OTUs occurring at less than 0.0005% reduced the number of OTUs to 4290 unique observations across all samples. Condensing of redundant taxonomic IDs through PhyloToAST (v1.4.0) showed a total of 776 OTUs. Rarefication of the condensed OTU table to the

Table 3-1: Sequence reads, operational taxonomic unit (OTU) count, and alpha diversity of each sample of the study. The table shows the sequence read count (1) before and (2) after quality checking and filtering based on low quality reads using FASTX Toolkit, the (3) unique unfiltered (Unfilt.) OTU observances following chimera removal, (4) the filtered (Filt.) OTUs after removal of rare OTUs (<0.0005% abundant in all samples), (5) the resultant condensed (Cond.) OTU count following the merging of redundant taxonomic information through Phylogenetic Tools for Analysis of Species-level Taxa (PhyloToAST) (v1.4.0), (6) the OTU count following subsampling to the median (Med.) value (77,806), and (7) the minimum (Min.) value (49,641). Also shown are the (8) Shannon and (9) Simpson diversity indices corresponding to each sample determined using the condensed OTU table data.

Sample	Raw Reads	Trimmed Reads	Unfilt. OTUs	Filt. OTUs	Cond. OTUs	Cond. OTUs Med.	Cond. OTUs Min.	Shannon Diversity	Simpson Diversity
Algae 1	154,561	121,543	4,985	1,345	268	259	243	5.359	0.9376
Algae 2	137,323	103,116	2,973	855	226	220	204	4.796	0.934
Algae 3	160,926	115,779	4,745	1,383	329	325	302	4.9626	0.9156
Gut Digesta 1	76,329	60,871	2,133	580	155	155	151	3.1936	0.826
Gut Digesta 2	74,249	61,267	2,144	599	119	119	113	2.476	0.5991
Gut Digesta 3	123,640	99,546	3,056	611	126	121	106	2.9289	0.7891
Gut Tissue 1	128,539	90,412	4,384	1,094	328	322	301	4.1236	0.844
Gut Tissue 2	68,644	51,895	2,311	898	273	273	273	3.8787	0.8436
Gut Tissue 3	123,735	89,547	4,765	1,418	403	397	368	4.8038	0.925
Pharynx Tissue 1	107,276	72,954	5,663	1,558	430	430	418	6.0719	0.9642
Pharynx Tissue 2	86,515	58,281	4,417	1,488	431	431	430	5.9389	0.9632
Pharynx Tissue 3	92,987	63,405	4,918	1,489	402	402	397	6.2246	0.9697
Water 1	123,154	81,885	4725	1679	504	504	487	5.2838	0.8797
Water 2	137,044	93,713	6127	1087	403	400	386	5.7289	0.9546
Water 3	119,824	85,613	5,608	1,406	400	400	377	4.2286	0.8211
Summary	total = 1,714,746	total = 1,249,827	total = 44,664	total = 4290	total = 776	total = 776	total = 776	avg. = 4.6666	avg. = 0.8778

median read count value (77,806) and the minimum read count value (49,641) both maintained 776 unique observations. Through the alternative strategy utilizing DADA2 (v1.10) implemented in QIIME2 (v2018.11), a total of 1134 unique features were determined (ASVs) representing a total of 467,866 reads across all samples, which were subsequently assigned to 371 taxonomic identities when collapsed to the species level (data not shown).

Taxonomic Distribution across Samples

Taxonomic distribution across all samples showed the gut tissue represented a uniquely heightened amount of Epsilonproteobacteria in the order of Campylobacterales, namely family Camplylobacteraceae (*Arcobacter*) (~20%) and Helicobacteraceae (*Sulfurimonas*) (~12%) as compared to the other samples of the study (Figure 3-2A,B). Members of Firmicutes (*Tissierella_Soehngenia*) were observed in the gut tissue and appeared to be present in the pharynx tissue and the environmental samples, particularly the water. Also observed were Bacteroidetes (Flavobacteriales, ~2%), as well as Gammaproteobacteria (*Psychromonas*, ~7%), Deltaproteobacteria (*Desulfotalea*, ~5%), and Fusobacteria (*Propionigenium*, ~5%), to lesser degrees of abundance.

The dominant microbial taxa in the gut digesta were observed to be members of *Psychromonas* (~40%), *Propionigenium* (~15%), and class Flavobacteriales (~25%). Compared to the gut tissue, these taxa comprised a large relative abundance (~80%) of the bacterial microbiota observed in the gut digesta. The gut digesta also included members of phylum Bacteroidetes (3%),

class Gammaproteobacteria identified as Vibrionaceae and *Vibrio* (~1 and ~2% respectively), and *Desulfotalea* (~5%) at noticeable relative abundances (Figure 3-3).

The pharynx tissue presented many of the same bacterial taxa observed in the water and to a lesser extent the algae samples. Of these shared taxa, *Fusobacterium* (~10%) was observed at the highest abundance, which was followed by the families S27-7 (Bacteroidales) and Gemellaceae, and genus *Prevotella*. The presence of these bacteria in the gut tissue and gut digesta were negligible (<1%). However, the pharynx tissue also included members of *Tissierella_Soehngenia* (~6%), *Sulfurimonas* (~3%), and *Desulfotalea* (~1%) which were observed in the gut tissue and digesta.

The algae samples showed Saprospiraceae (~15%), Rhodophyta (~10%), and *Stramenopiles* (~9%) to be heightened, the presence of which was negligible in the other samples in the present study. We also observed the genera *Maribacter* and *Octadecabacter* at equal capacities (~4%).

From the alternative strategy utilizing ASVs, the taxonomic distribution was in concert with the OTU picking strategy, with only slight variations in relative abundance. In the gut tissue, a slightly higher relative abundance of *Arcobacter* (~22%) and *Sulfurimonas* (~14%) was determined through the ASV method compared to the OTU picking strategy. A variation in relative abundances of heightened taxa was also observed in the gut digesta, where *Psychromonas* was more highly represented at ~50%, whereas *Propionigenium* (~13%) and Flavobacteriales (~14%) were marginally less abundant. The pharynx tissue was

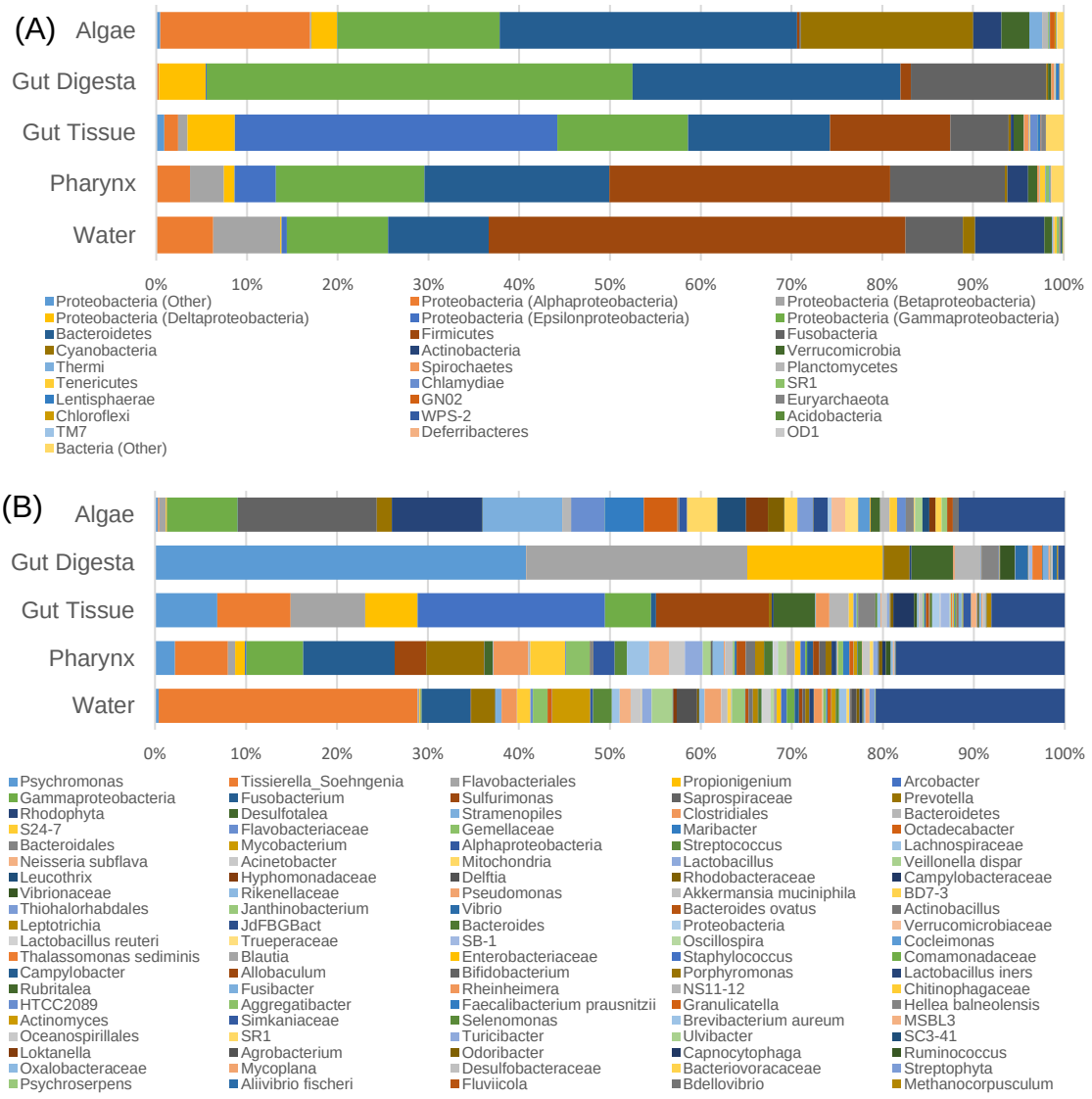


Figure 3-2: Taxonomic distribution of microbial communities in the gut ecosystem and rocky tide pool environment of the sea urchin *S. purpuratus*. (A) The relative abundance of phyla (class for Proteobacteria) represented at >1% are shown, with phyla <1% grouped as “Other.” (B) The top 100 taxa at the most resolvable level across all samples were also visualized, and taxa not included as the top 100 were assigned as “Other.” OTUs were picked at 97% similarity threshold, filtered at <0.0005%, condensed using PhyloToAST (v1.4.0), and subsampled to minimum OTU count (rarefied OTU table). Taxonomic identities were determined by using the GreenGenes (v13.8) database, and the color code corresponds to each taxon observed across the gut and environmental samples. Grouping of biological replicates (n = 3) was supported by an analysis of similarity (ANOSIM) and multivariate analysis of variance (Adonis) (p < 0.001). Relative abundance plot was created through Microsoft Excel Software (Seattle, WA, USA).

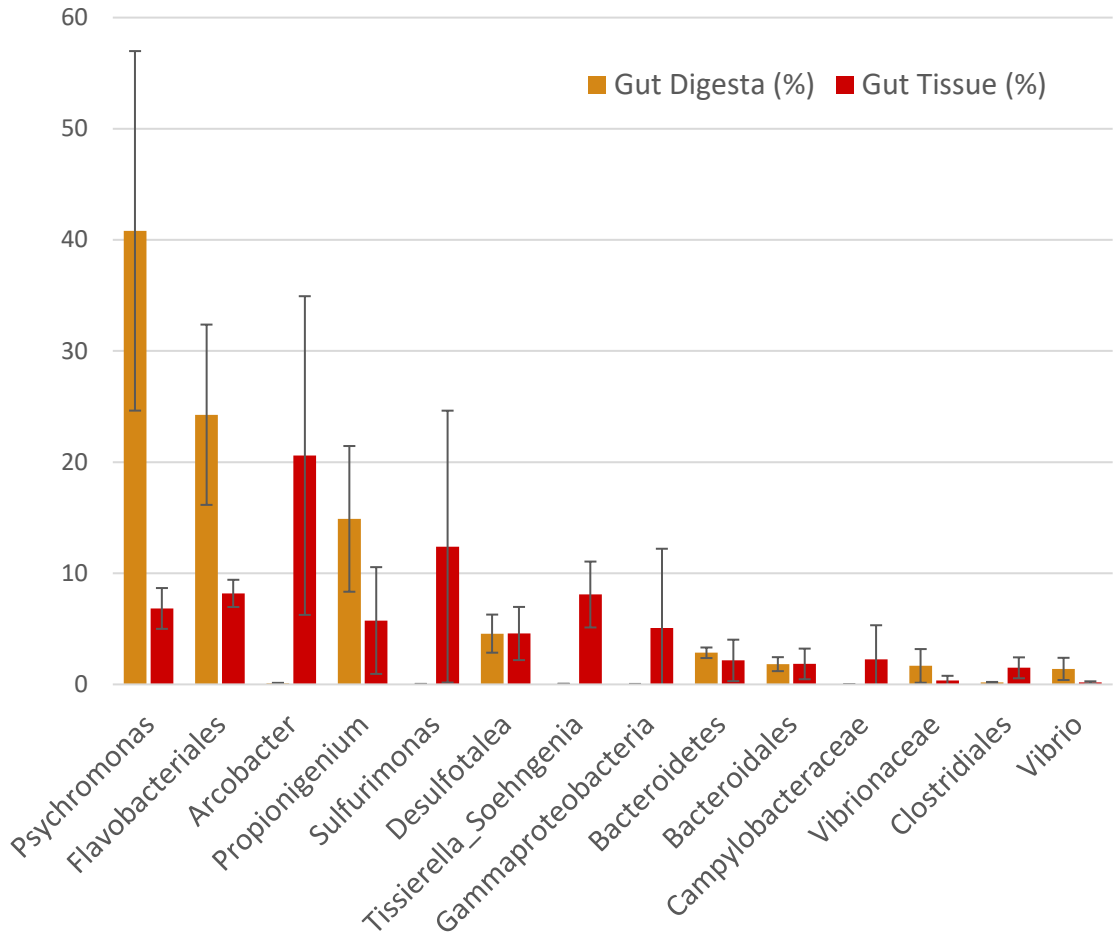


Figure 3-3: Comparison of the observed taxa between the gut tissue (*n* = 3) and gut digesta (*n* = 3) using the rarefied OTU table data. Taxa observed at <1% were filtered from the graph. Standard deviation and relative abundances were determined through STAMP (v2.1.3), and the graph was generated through Microsoft Excel Software (Seattle, WA, USA).

also consistent between the two strategies, with a slightly higher abundance of *Fusobacterium* (12%) and *Sulfurimonas* (~7%). The ASV method was able to resolve an extra phylogenetic level of one heightened feature in the pharynx tissue, which was classified to phylum Gammaproteobacteria (~6%) through the OTU picking method, but was determined as order Legionellales (~10%) through the ASV approach. For the water samples, *Tissierella_Soengenia* (~32%)

comprised the highest abundance, followed by Fusobacterium (~7%), and for the algae samples, Saprospiraceae, Rhodophyta, and *Stramenopiles* were confirmed through the ASV method, at slightly higher relative abundances compared to the OTU picking method. Additionally, the results of the alignment of the representative sequences corresponding to the top 100 taxa determined in the rarefied OTU table through SILVA ACT: Alignment, Classification and Tree Service (www.arb-silva.de/aligner) have been elaborated (Table 3-S1).

Alpha Diversity

Alpha diversity measures performed on the rarefied OTU table showed the highest Shannon diversity in the pharynx tissues (avg = 6.08 ± 0.08 SEM), followed by the water (5.08 ± 0.44 SEM) and algae (5.04 ± 0.17 SEM), with the gut tissue (avg = 4.27 ± 0.28 SEM) and gut digesta (avg = 2.87 ± 0.21 SEM) showing the lowest diversity. Simpson diversity showed a similar trend in the sea urchin gut samples, with the pharynx tissue showing the highest diversity (avg = 0.976 ± 0.002 SEM), and the gut tissue (0.871 ± 0.027 SEM) and gut digesta (0.738 ± 0.070 SEM) showing the lowest. The water (0.885 ± 0.039 SEM) and algae (0.929 ± 0.007 SEM) microbial profiles were more diverse than the gut tissue and digesta, but less diverse than the pharynx tissue (Table 3-1). Kruskal–Wallis H-test analysis through PhyloToAST (v1.4.0) showed significant differences between the Shannon (p = 0.017) and Simpson (p = 0.027) alpha diversity values across the 5 groups of the study. The kernel density smoothed histograms plotted using PhyloToAST (v1.4.0) visualized the alpha diversity

values, with density representing the intra-sample variation in the group (high density = low variation between the diversity of samples in the group). For both Shannon (Figure 3-4A) and Simpson (Figure 3-4B) diversity, the highest density corresponded to the pharynx tissue samples, indicating low intra-sample variation in the group. Shannon diversity showed the broadest diversity value range in the water (Figure 4A), and Simpson showed the broadest range in the gut digesta (Figure 3-4B). For Shannon diversity, the gut digesta and gut tissue samples had distinct histogram peaks. For Simpson diversity, the range of alpha diversity measures for the gut digesta was broader, indicating a minimal density peak, and the gut tissue showed a low but noticeable peak (Figure 3-4A,B).

Beta Diversity

Microbial taxonomic distribution patterns determined through Bray–Curtis metrics across all samples revealed the gut tissues to cluster together at >50% and the gut digesta >60% (Figure 3-5A). These two sample groups maintained a Bray–Curtis similarity >40%. The pharynx tissues group demonstrated low intrasample variation (Bray–Curtis similarity >60%) and had a microbial community structure that was the most similar to the water samples (Bray–Curtis similarity >40%, Figure 3-5B). The algae samples clustered together at a value >60% but were least similar to the other samples of the study. The observed cluster patterns were strengthened when only the top 100 OTUs were plotted

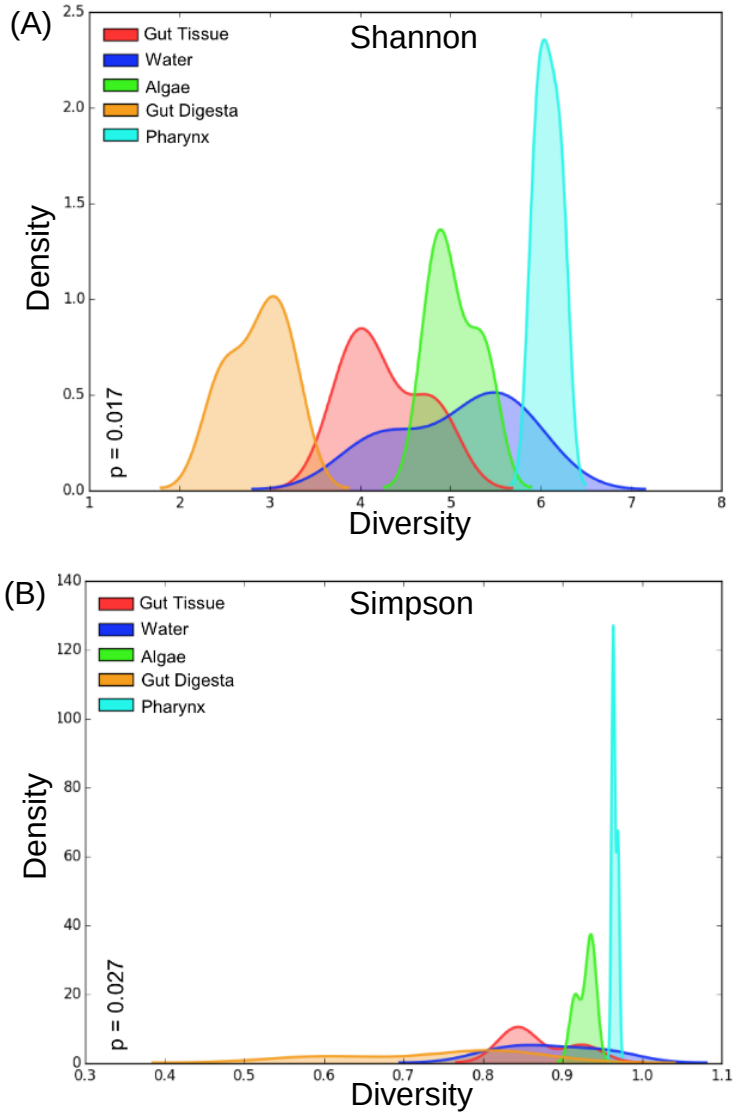


Figure 3-4: Per-group alpha diversity measurements calculated across all samples in the study. (A) Shannon and (B) Simpson alpha diversity histograms were smoothed by kernel density estimation. The Kruskal–Wallis H-tests were performed for the five groups and showed a significance value of $p = 0.017$ for the Shannon and $p = 0.027$ for Simpson diversity measurements, indicating significant differences between each group's alpha diversity. The X-axis shows the diversity value of Shannon (values much greater than 0 are more diverse) and Simpson (values closer to 1 are more diverse). The histogram values of each sample were smoothed through kernel estimation to show the range of sample data points within each group. The Y-axis depicts the density function, which denotes the distribution of data points falling within this range (higher peak represents more clustered data points). Relevant p-values are listed in each graph. Plots were generated using the "diversity.py" command through PhyloToAST (v1.4.0).

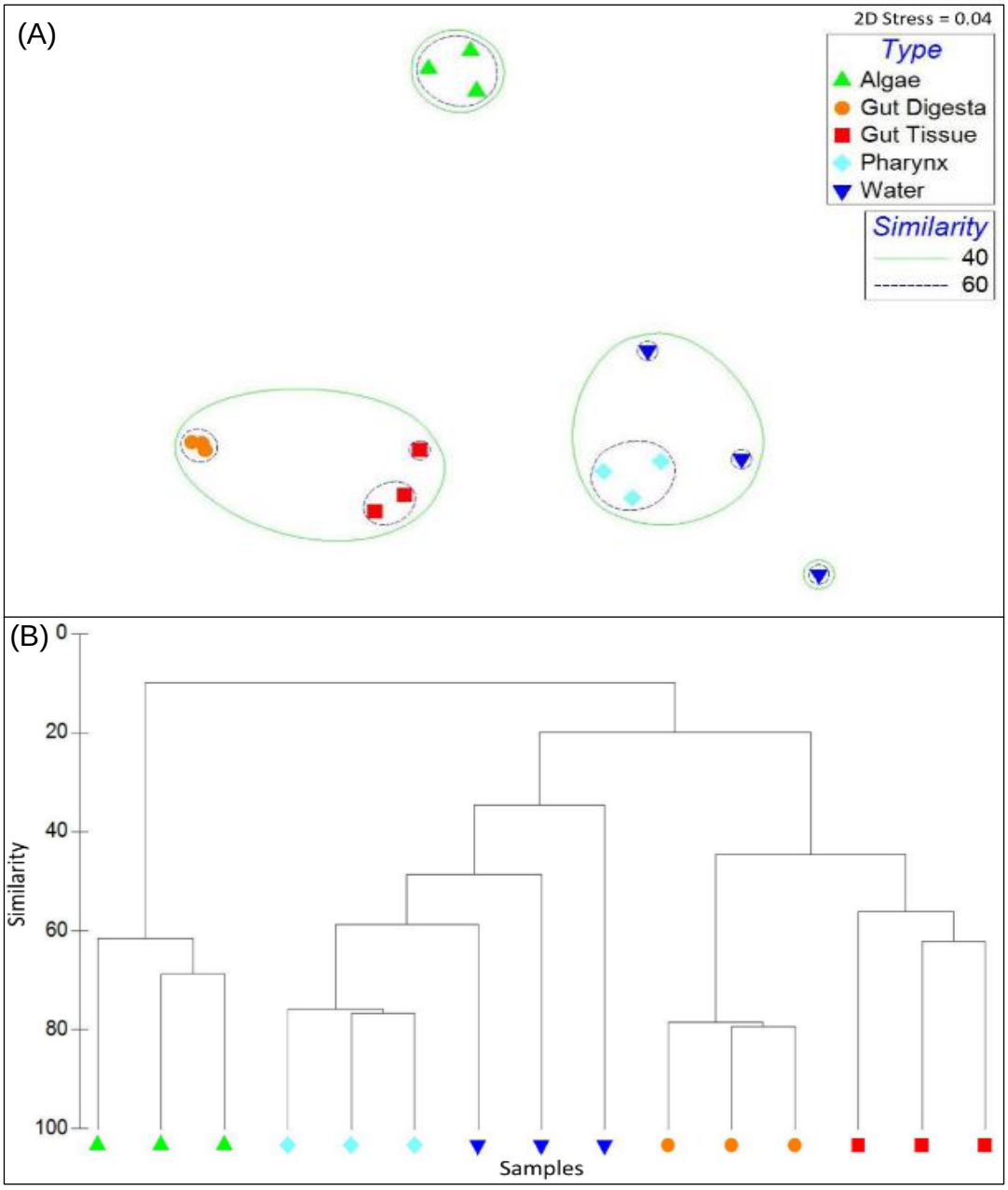


Figure 3-5: Beta diversity analysis of microbial communities observed across all samples in the study using Bray–Curtis similarity metrics determined for the rarefied OTU table. (A) A 2D multidimensional scale (MDS) plot analysis was performed to show sample cluster patterns based on observed OTUs, with a 40% and 60% Bray-Curtis similarity overlay, and the stress value (2D Stress = 0.04) was indicated. (B) Dendrogram analysis was also performed and each sample's cluster patterns were based on group average. The OTU table was pretreated via standardization by the total and log transformation prior to Bray-Curtis analysis. Figure legends are shown in the 2D MDS plot. Data was generated and plotted through PRIMER-6 software (Primer-E Ltd, Plymouth Marine Laboratory, Plymouth UK, v6.1.2).

through 2D MDS (Figure 3-S1A,B). For the rare taxa (the taxa not included in the top 100 OTUs), although low intrasample variation was observed, the gut tissue group was clustered nearer the pharynx and water samples (Figure 3-S1C,D). Although the general trends of within- and between-group sample similarity were supported through the weighted (Figure 3-S2A,B) and unweighted (Figure 3-S3A,B) Unifrac 3D PCoA and dendrogram analysis, there were slight variations in cluster patterns between the two Unifrac approaches. Through the weighted Unifrac method, the gut tissue and gut digesta microbial communities clustered closer together as compared to the unweighted method, which showed the gut tissue samples to more closely resemble the pharynx. However, in concert with the Bray–Curtis method, both methods demonstrated the pharynx tissue to more closely resemble the water samples, and the algae samples maintained a divergent cluster pattern away from the other samples of the study.

Similarity trends between sample groups were elaborated by heatmap analysis, which also depicted the relative abundance associated with each taxon contributing to the group diversity across the grouped biological replicate samples (Figure 3-6). Dendrogram analysis across grouped biological replicates for the heatmap analysis revealed the gut digesta and gut tissue to have a more similar microbial ecology composition, and likewise for the water and pharynx groups. The algal food source was the least similar to the other samples of the study, confirming the similarity trends observed in the 2D MDS plot analysis (Figure 3-5A). LEfSe analysis of the gut tissue and digesta at an LDA score of ±3 showed the taxa contributing most to the dissimilarity (effect size) of the gut

tissue to be *Arcobacter*, *Sulfurimonas*, and *Tissierella_Soehngenia*, whereas the gut digesta revealed *Psychromonas*, Flavobacteriales and *Vibrio* (Figure 3-7). The effect size of *Propionigenium*, which was noticeably abundant in the gut tissue (5.7% ± 4.8%) and digesta (14.9% ± 9.6%), was not observed in LDA analysis, likely due to overlapping standard deviation values (as observed in Figure 3-3).

Table 3-2: Grouping statistics performed on each OTU table generated in the study. Results show both ANOSIM and Adonis measurements of the initial OTU table (unfiltered; Unfilt.), followed by filtering of those taxa represented at <0.0005% in the study (filtered; Filt.), condensing (Cond.) using PhyloToAST (v1.4.0) and rarefying to the median (Med.) and minimum (Min.) of the condensed OTU table file. Analysis was performed on grouped biological sample replicates (pharynx, *n* = 3; gut tissue, *n* = 3; gut digesta, *n* = 3; water, *n* = 3; and algae, *n* = 3).

Diversity Measure	Unfilt. OTUs	Filt. OTUs	Cond. OTUs	Cond. OTUs Med.	Cond. OTUs Min
ANOSIM (R)	0.93185	0.93185	0.94074	0.94074	0.94222
Adonis (R^2)	0.69518	0.71145	0.74894	0.74951	0.75688

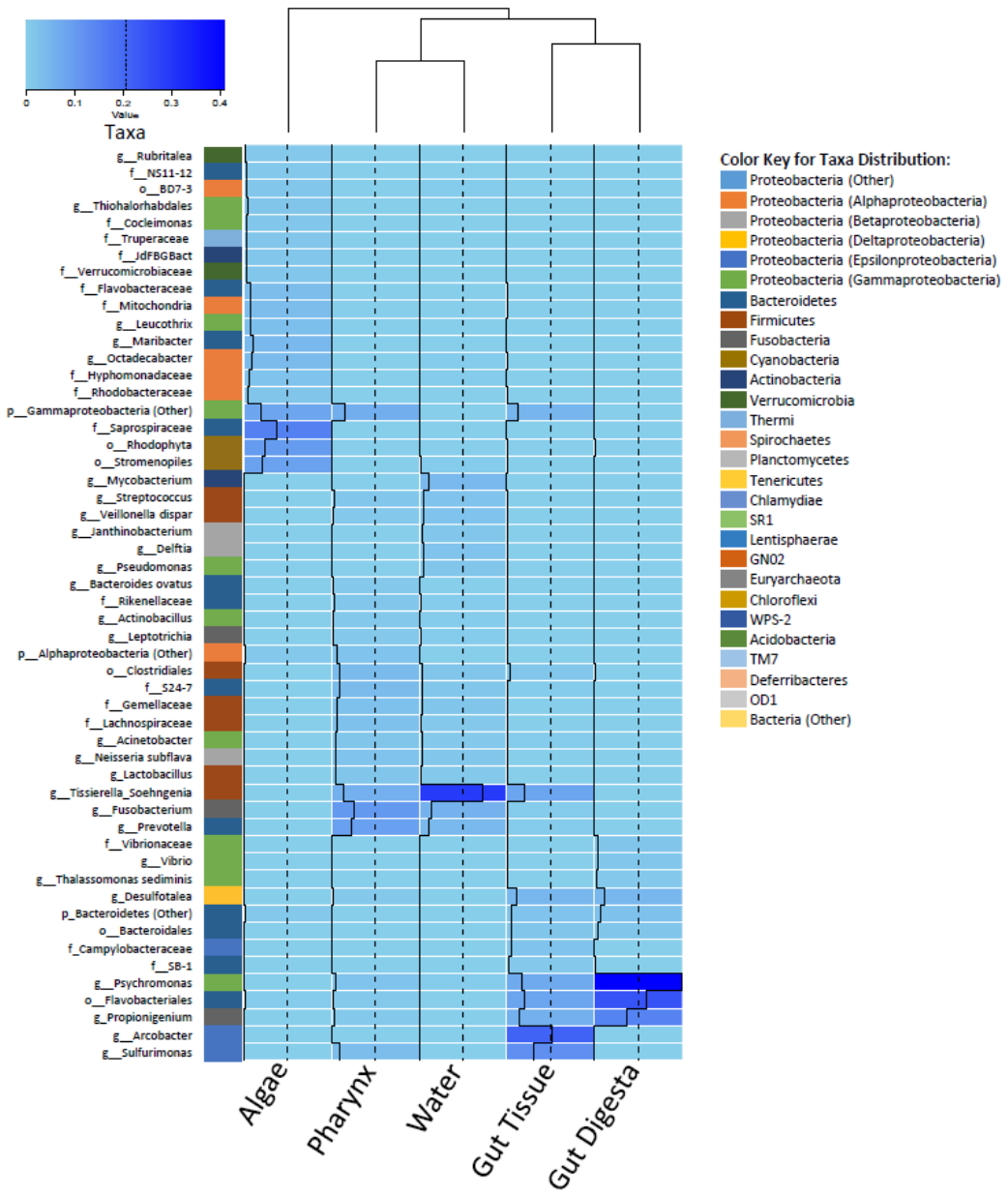


Figure 3-6: Heatmap of the top 53 taxa at the highest resolution, determined using the rarefied OTU table and generated using R (v3.3.2). The heatmap.2 function from the gplots (v3.0.1) package was used. Sample dendrogram was generated using Vegan (v2.4.3), employing the Bray-Curtis metric of the grouped biological replicate count data. Color palette selected using the RColorBrewer package and set from “sky blue” for less abundant, to “blue” for more abundant (shown in color key). Relative abundance values of each taxon are also indicated through a trace line (black). The associated table includes the most resolvable taxonomic assignment according to the GreenGenes (v13.8) database, which is color-coded to the phylum level assignments (class for Proteobacteria) as indicated in the key and corresponding to the relative abundances in the Figure 2A relative abundance graphs.

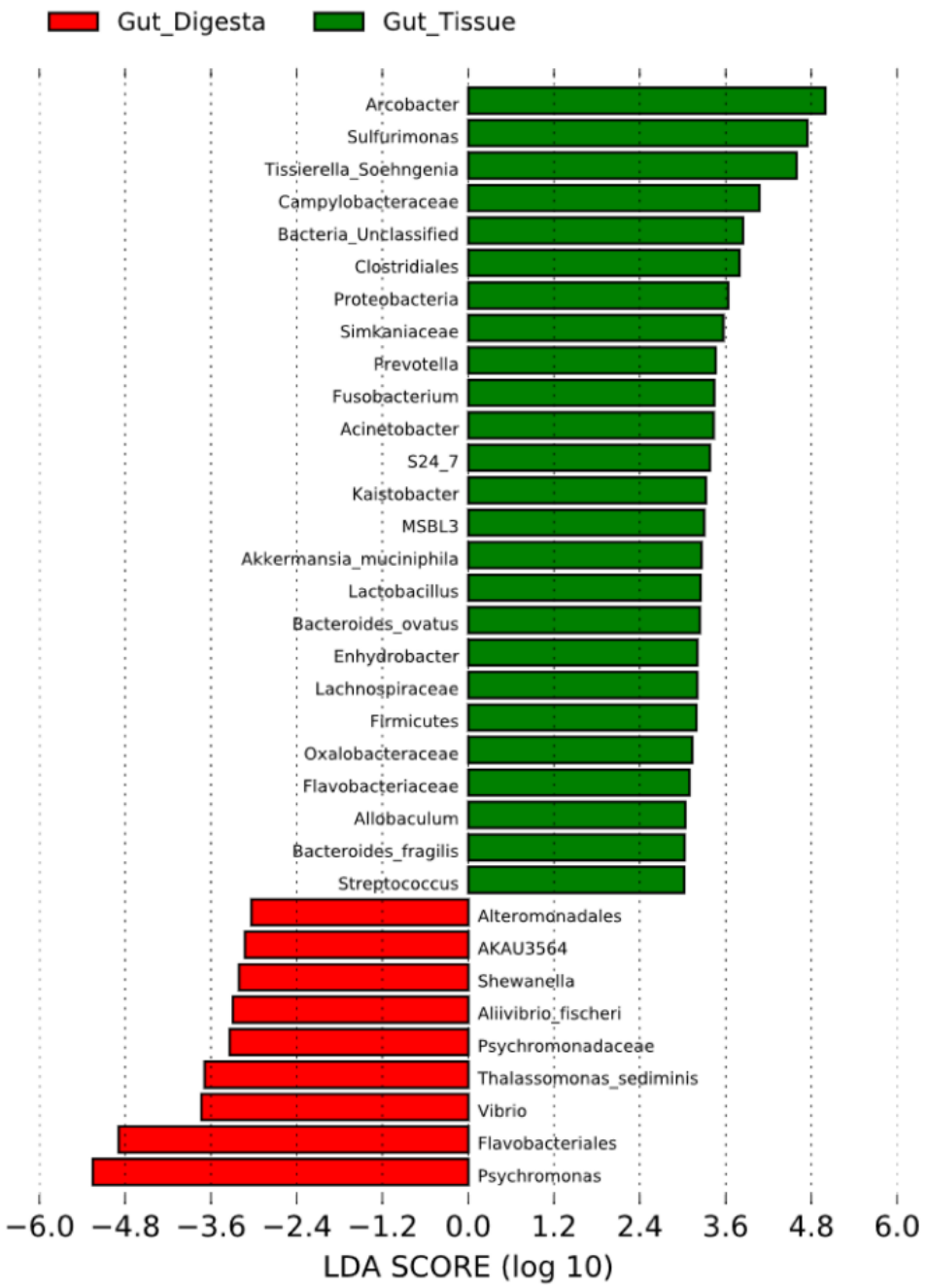


Figure 3-7: Linear discriminant analysis (LDA) effect size (LEfSe) performed on the microbial community relative abundance data at the of the gut tissue (*n* = 3) and gut digesta (*n* = 3). Grouped data were first analyzed using the Kruskal–Wallis test with a significance set to 0.05 to determine if the data was differentially distributed between groups, and those taxa that were differentially distributed were used for LDA model analysis to rank the relative abundance difference between groups. The LDA for significance was set to ±3, and the log(10) transformed score is shown to demonstrate the effect size. Data were analyzed and prepared through Hutlab Galaxy provided through the Huttenhower lab. The gut tissue group is shown as green, and the gut digesta group as red.

Predicted Functional Capacity

The predicted functional capacity of the microbial communities of the digestive system was performed using PICRUSt (v1.1.2), showing an average NSTI value of 0.139 (range from 0.105 - 0.179). Scatter plot analysis using STAMP (v2.1.3) of the KEGG-Level-2 categories showed a preferential abundance of energy metabolism in the gut tissue, as well as membrane transport, cell motility, and signal transduction (Figure 3-8A). For the gut digesta, amino acid metabolism, carbohydrate metabolism, metabolism of cofactors and vitamins, and replication and repair categories were observed. KEGG-Level-3 observances showed a preferential abundance of oxidative phosphorylation, carbon fixation, methane, and nitrogen metabolisms in the gut tissue, and categories related to the transporter and motility-related categories (2-component system, chemotaxis, bacterial motility proteins, and flagellar assembly) (Figure 3-8B). The gut digesta displayed categories related to pyrimidine metabolisms, as well as peptidases and amino acid enzymes, including arginine and proline metabolisms. Other categories that were enriched in the digesta included starch and sucrose metabolism, pentose phosphate pathway, glycolysis/ gluconeogenesis, and ubiquinone and turpenoid-quinone biosynthesis necessary for electron transport. The LEfSe analysis of KO Ids (highest metabolic resolution) contributing most to the effect size difference between the gut tissue and gut digesta were determined and listed along with their associated metabolic definitions (Figure 3-9).

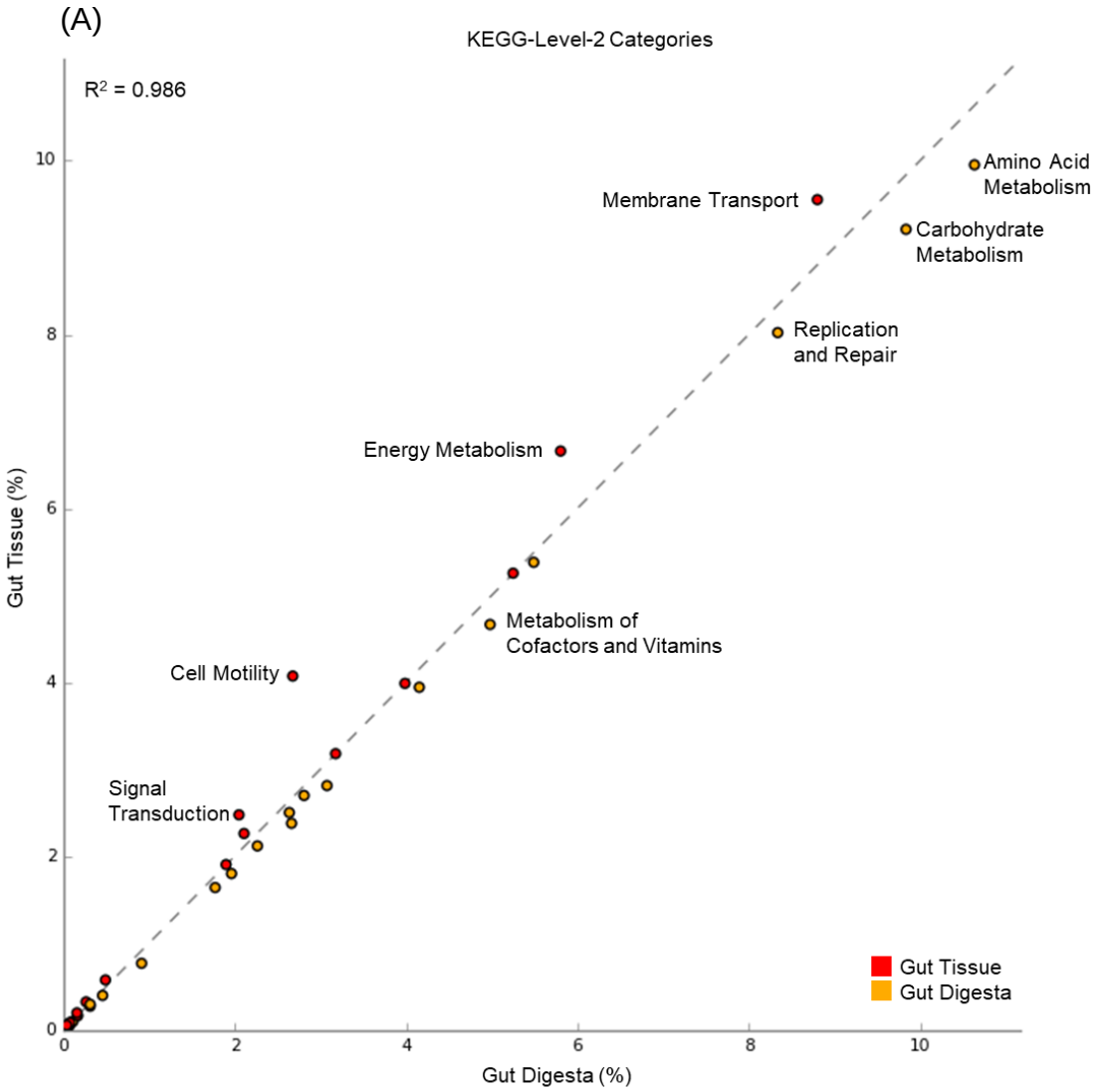
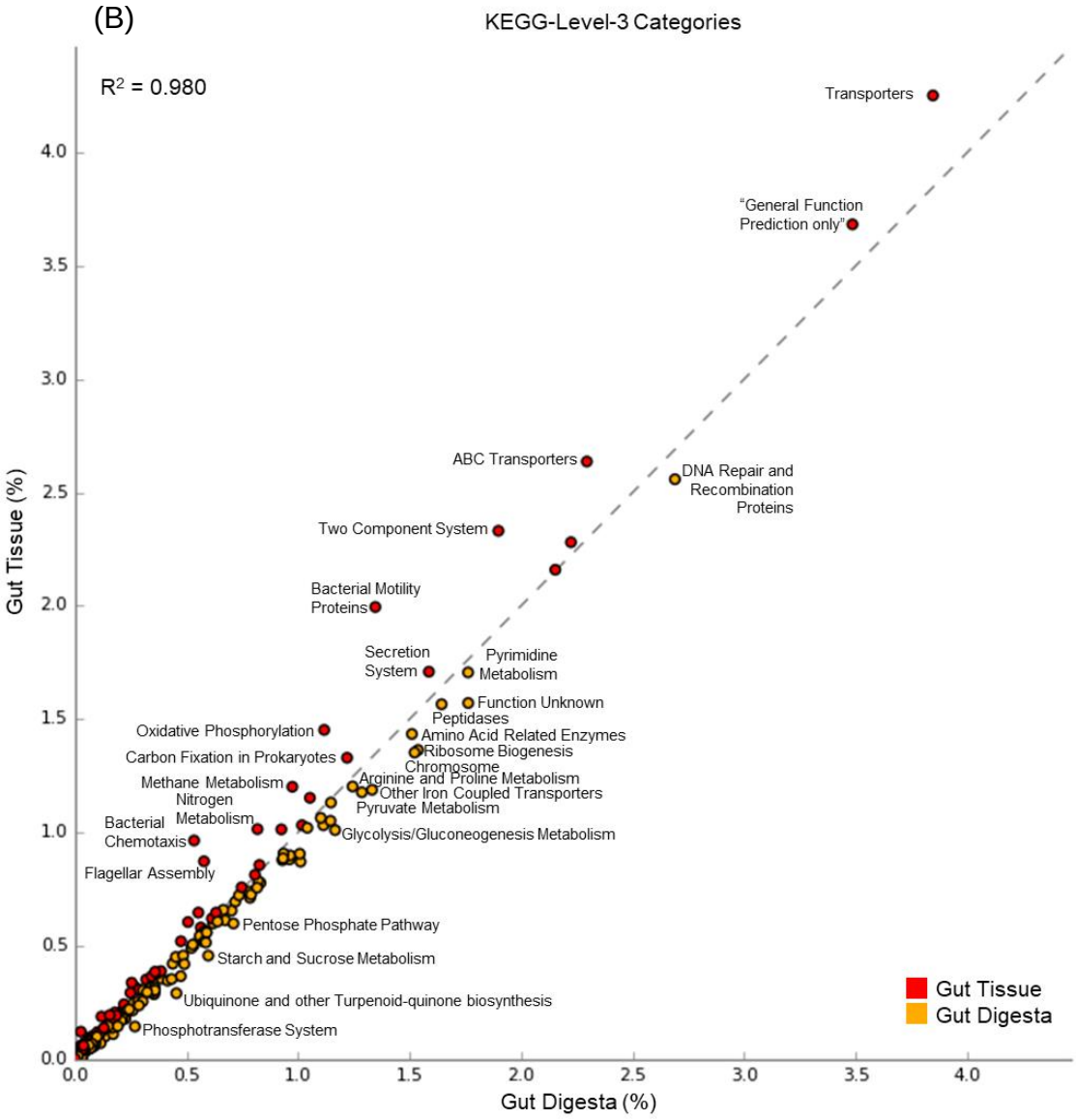


Figure 3-8: Scatter plot analysis of the predicted KEGG Orthology (KO) metabolic functions determined through Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt v1.1.2) performed on the gut tissue (*n* = 3) and gut digesta (*n* = 3). Biological replicates were grouped, and analysis was performed for the (A) KEGG-Level-2 and (B) KEGG-Level-3 hierarchical functional categories. The linear regression value calculated for the



(Figure 3-8 continued) two groups is shown for each scatter plot graph. Preferentially enriched categories for the gut tissue are shown as red, and for the gut digesta as brown. Those categories with clearly preferentially abundant categories have been labeled. Data were analyzed and visualized using STAMP (v2.1.3) analytical software.

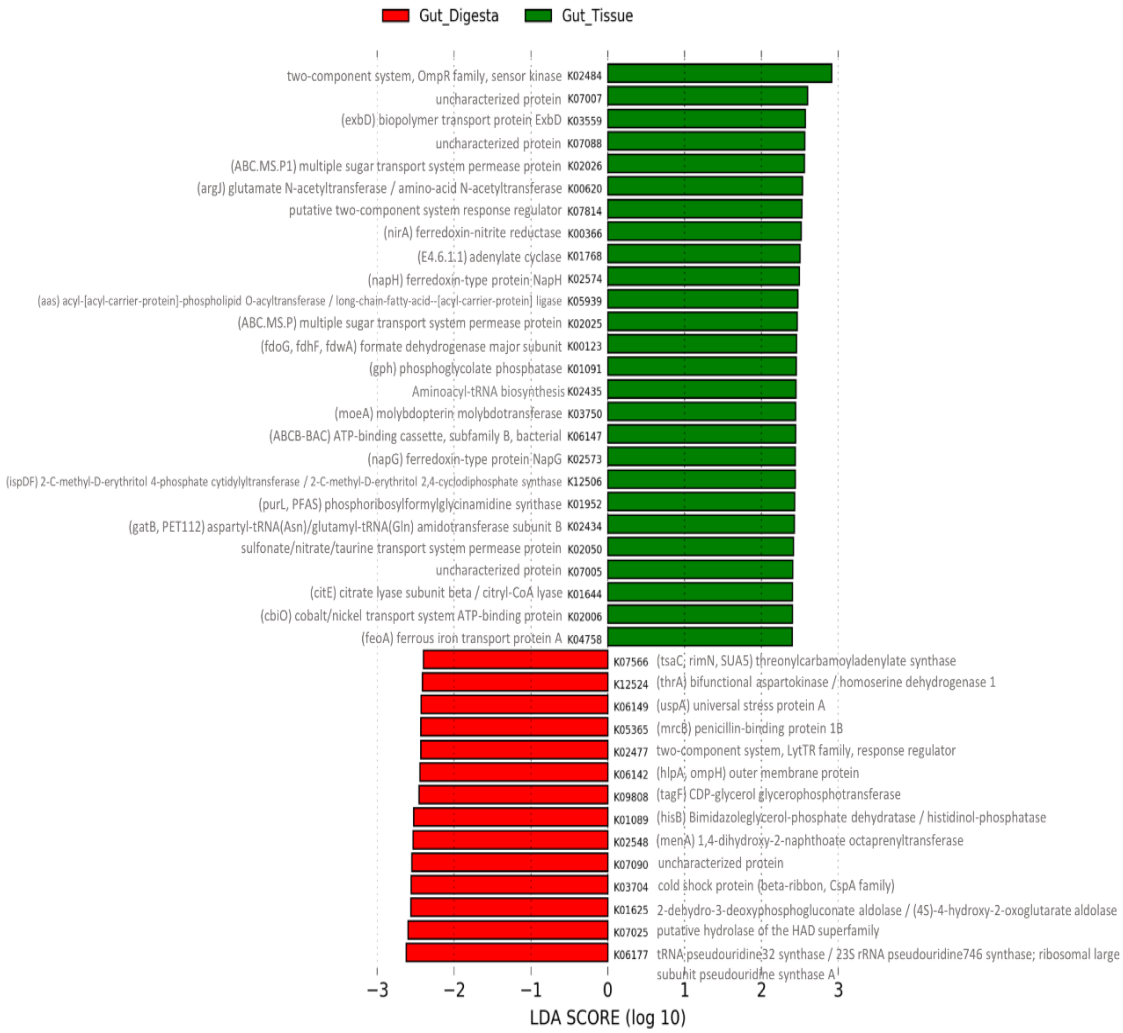


Figure 3-9: Linear discriminant analysis (LDA) effect size (LEfSe) performed on the KEGG Orthology (KO) metabolic functions determined through PICRUSt (v1.1.2) for the gut tissue (*n* = 3) and gut digesta (*n* = 3). The KO Ids were determined through the “predict_metagenomes.py” command in PICRUSt (v1.1.2). Grouped data were analyzed using the Kruskal–Wallis with a significance set to 0.05, and the significantly differentially distributed KO Ids were used for LDA model analysis ranking the relative abundance significance, at an LDA threshold showing entries ranking at ± 2.4. The log(10) transformed score is shown as the effect size. Data were analyzed and prepared through Hutlab Galaxy provided through the Huttenhower lab. The gut tissue group is shown as green, and the gut digesta group as red.

Co-Presence, Co-Exclusion, and Key Taxa in the Gut Environment

The resultant network generated through CoNet (v1.1.1) in Cytoscape (v3.6.0) yielded 71 nodes and 294 edges elucidating possible interactions occurring between taxa representing the distinct microbial communities of the sea urchin gut environment (Figure 3-10A). Analysis of network properties using NetworkAnalyzer (v2.7) showed an average network centralization of 0.128, the characteristic path length of 2.856, average number of neighbors at 8.282, with a network density of 0.118 and network heterogeneity of 0.452. Scatter plot analysis demonstrated the trends of closeness centrality plotted against betweenness centrality, along with the nodes scaled to the degree (Figure 3-10B). The top 10 candidate key taxa based on the topological qualities of taxonomic nodes described in Berry and Widder et al. [30] were ranked by their closeness centrality, and showed *Propionigenium*, *Moritella*, SB-1 (Bacteroidetes), Desulfobacteraceae and *Desulfovibrio* in the gut digesta, and Rhodobacteraceae, Rhodophyta, Vibrionaceae, *Arcobacter* and *Bacilli* in the gut tissue (Table 3-3). Of these taxa, the gut digesta showed *Propionigenium* to have the highest degree of associations with the gut tissue taxa (17 total), with the majority of these associations shown as co-presence (14 total). The gut tissue showed Rhodobacteraceae to have the highest degree (9 total), and most of these associations were co-exclusion (8 total). The gut tissue also showed the highly abundant *Arcobacter* to be a likely key taxon, revealing a degree of 9, with the majority of these associations as co-presence (7 total).

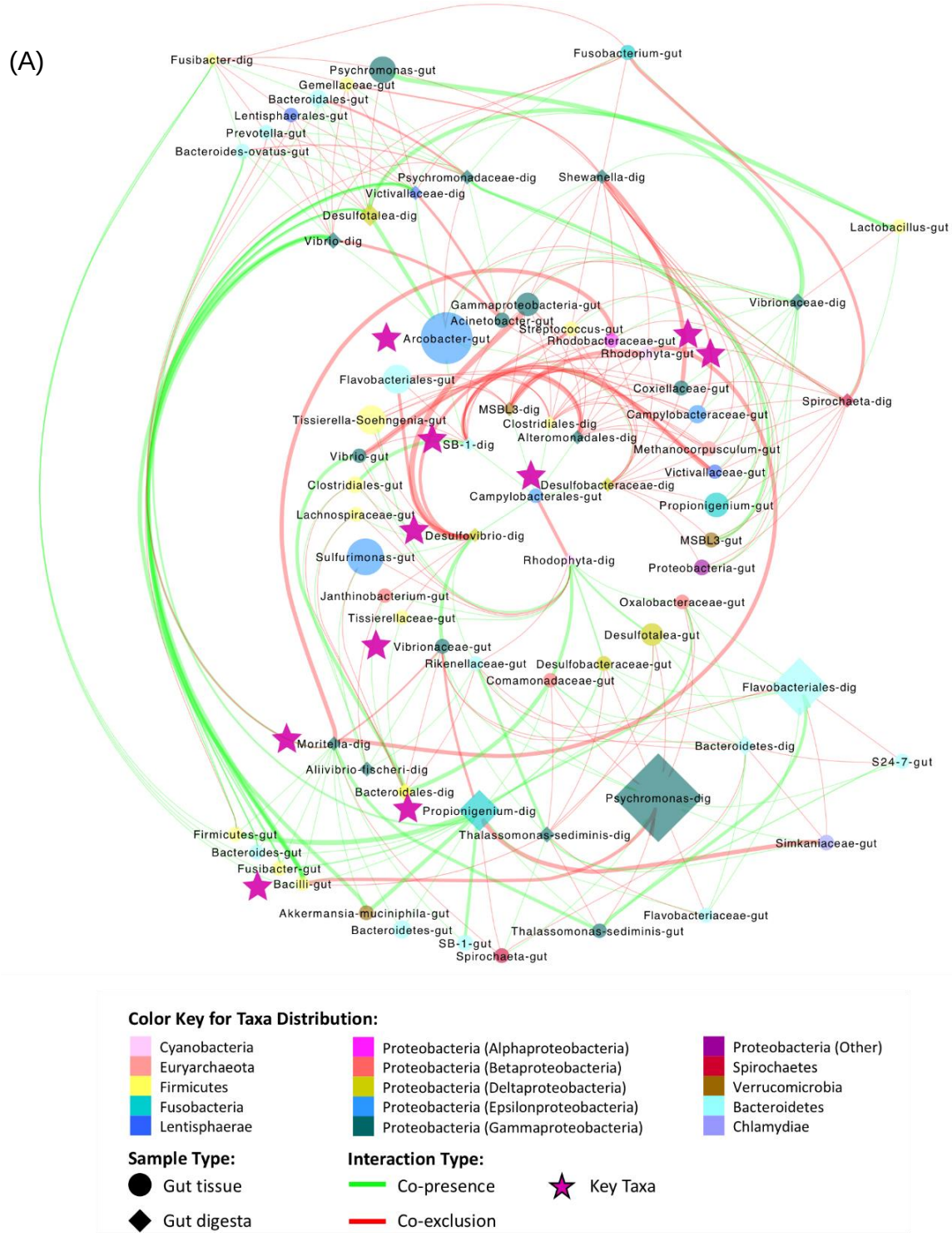
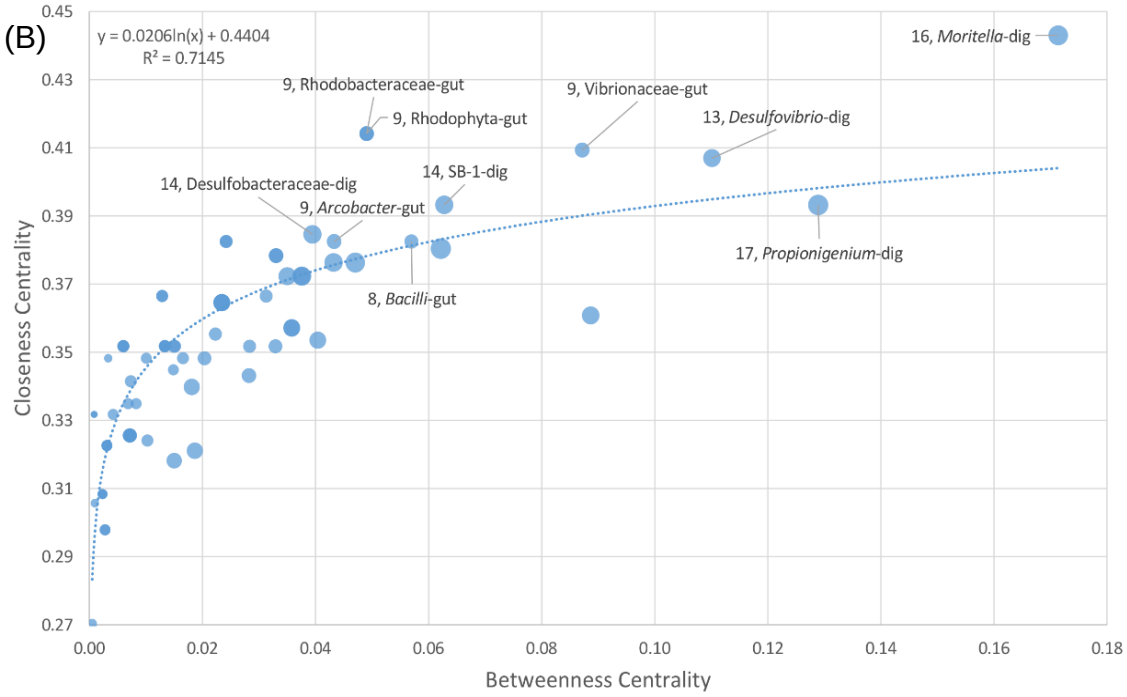


Figure 3-10: Co-occurrence patterns between taxonomic entries of the gut tissue and gut digesta, determined through Co-occurrence Network inference (CoNet v1.1.1), and analyzed through Cytoscape (v3.6.0). Taxonomic entries with a cumulative row sum of 200 or above with 2/3 of samples showing non-zero value



(Figure 3-10 continued) were used through an ensemble approach that incorporated the Pearson, Spearman, Bray–Curtis, Kullback–Leibler, and mutual information metrics. The top and bottom 200 edges were selected and merged by the union method. (A) The network analysis shows the edges represented by the *q*-value (merged with the Brown method at $p < 0.05$ for each metric) and are shown as green (co-presence) and red (co-exclusion), with the nodes representing taxa were scaled according to relative abundance and colored according to the phyla (class for Proteobacteria) assignments. The final network was arranged using the yFiles (v1.0) Cytoscape (v3.6.0) add-on radial layout, and taxonomic entries shown at the highest resolution are denoted with the sample type (circle for gut tissue, "-gut"; diamond for gut digesta; "-dig"). (B) Scatter plot analysis was performed using topological metrics determined by NetworkAnalyzer (v2.7), to demonstrate patterns of key (keystone) species between taxonomic entries of the gut tissue and gut digesta based on closeness and betweenness centrality, as well as the degree (number of co-presence and co-exclusion edges). Linear regression between closeness and betweenness centrality was shown as logarithmic (R2 value = 0.7145), and the top 10 entries ranked by closeness centrality are depicted. Note, the taxa Rhodophyta and Rhodobacteraceae had the same closeness and betweenness centrality measurements, and their corresponding plot is indistinguishable. Linear regression determined through Microsoft Excel Software (Seattle, WA).

Table 3-3: Candidate key taxa resulting from CoNet (v1.1.1) analysis between the gut tissue and gut digesta microbial communities. The group assignment is shown (gut tissue, *n* = 3; gut digesta, *n* = 3), along with the phylum and taxon at the highest resolution. The average abundance (Av. Ab.) was determined for each taxon based on the respective group (gut tissue or digesta as indicated). The entries were ranked according to closeness centrality (Clos. Cent.), highest to lowest (a feature of keystoneness), and then by degree (edges; Deg.), highest to lowest. Also shown is the betweenness centrality (Bet. Cent.), and the total, positive (co-presence; Pos. Deg.) and negative (co-exclusion; Neg. Deg.) degree values

Sample Type	Phylum	Taxon	Av. Ab.	Clos. Cent.	Bet. Cent.	Deg.	Pos. Deg.	Neg. Deg.
Gut Digesta	Fusobacteria	*Propionigenium*	14.89%	0.39	0.13	17	14	4
Gut Digesta	Proteobacteria (Gammaproteobacteria)	*Moritella*	0.15%	0.44	0.17	16	10	6
Gut Digesta	Bacteroidetes	SB-1	0.38%	0.39	0.06	14	3	11
Gut Digesta	Proteobacteria (Deltaproteobacteria)	Desulfobacteraceae	0.24%	0.38	0.04	14	2	12
Gut Digesta	Proteobacteria (Deltaproteobacteria)	*Desulfovibrio*	0.24%	0.41	0.11	13	6	7
Gut Tissue	Proteobacteria (Alphaproteobacteria)	Rhodobacteraceae	0.34%	0.41	0.05	9	1	8
Gut Tissue	Cyanobacteria	Rhodophyta	0.22%	0.41	0.05	9	1	8
Gut Tissue	Proteobacteria (Gammaproteobacteria)	Vibrionaceae	0.36%	0.41	0.09	9	5	5
Gut Tissue	Proteobacteria (Epsilonproteobacteria)	*Arcobacter*	20.59%	0.38	0.04	9	7	2
Gut Tissue	Firmicutes	*Bacilli*	0.16%	0.38	0.06	8	6	2

DISCUSSION

Overall, the *S. purpuratus* gut ecosystem exhibited high abundances of *Arcobacter* and *Sulfurimonas* in the gut tissues. These taxa belong to phylum Epsilonproteobacteria, and specifically within the order Campylobacterales, members of which are known to be chemolithoautotrophic [88]. Both *Arcobacter* and *Sulfurimonas* have been implicated as marine sulfur-oxidizing bacteria, with members of *Arcobacter* forming oxidized sulfur filaments in response to geothermally produced sulfide in hydrothermal vents [89,90], and *Sulfuromonas* has been observed to oxidize sulfur in anoxic deep-sea hydrothermal sediments [91,92]. Additionally, Epsilonproteobacteria have been reported in the gut systems of marine organisms, with *Arcobacter* previously observed in the deep-sea vent-dwelling shrimp *Rimicaris exoculata* [93] and the marine Chilean oyster *Tiostrea chilensis* [94], and *Sulfurimonas* in the stomach of the hydrothermal vent crab *Xenograpsus testudinatus* [95] and as symbionts in marine gastropod mollusks *Alviniconcha* [96,97]. These bacteria have been implicated as potential symbionts, assisting in the oxidation of sulfur in the highly sulfidic environments [98]. Previous studies using Illumina MiSeq HTS of the V4 region of the collective 16S rRNA genes conducted on the gut environment of the sea urchin *L. variegatus* from the Gulf of Mexico similarly revealed Campylobacterales to be near-exclusively abundant in the gut tissue [15,16], of which further analysis of the highly-represented Campylobacterales sequence showed a ~90-91% similarity match to *Arcobacter*/*Sulfuricurvum* through NCBI BLAST [15].

In contrast to the gut tissue, the microbial composition of gut digesta was dominated by *Psychromonas* (phylum Gammaproteobacteria). This halophilic genus is known to occur in cold marine environments (~4 °C) [99], and members are capable of hydrolyzing starch and other insoluble sugars [100,101,102,103], as well as producing ω-3 polyunsaturated fatty acids [99]. It has recently been shown that the total ω-3 fatty acids of the gut contents (23.26 ± 6.88; mean ± SD) of experimentally fed *S. purpuratus* are higher than the total ω-3 fatty acids of their algal diets (13.41 ± 6.32); the dominant driver of this pattern was from the ω-3 eicosapentaenoic acid (EPA; 20:5ω-3) [20], but the mechanism for this enrichment was unknown. The psychrophilic capabilities of *Psychromonas* are consistent with the temperature of the Oregon Pacific Coast seawater temperature from which *S. purpuratus* was sampled, which was recorded at 13.1 °C, within the range of optimum growth Psychromonas spp. [100]. Also observed in the gut digesta were *Propionigenium* of phylum Fusobacteria, members of which are anaerobic, are capable of fermenting succinate to propionate [104] and have been implicated in carbohydrate metabolisms that include cellulose, producing short-chain fatty acids [105]. *Propionigenium* is involved in a myriad of host health benefits, including its association in the modulation of the lifespan of the Turquoise Killifish (*Nothobranchius furzeri*) [105]. Order Flavobacteriales of phylum Bacteroidetes were also observed, which have been identified in the intestines of shrimp *Litopenaeus stylirostris* raised in aquaculture and clear waters [106]. Importantly, members of Flavobacterales have been investigated with consideration of their metabolic capability to break-down alginate from

brown seaweeds common in both temperate and polar coastal environments [107]. Lastly, the genus *Desulfotalea* of Deltaproteobacteria was observed at approximately equal relative abundances in the gut digesta and gut tissue (~5% in each). Certain members of *Desulfotalea* are adapted to extremely cold environments, and these bacteria are capable of sulfur reduction [108,109].

Shannon and Simpson's diversity measured for all samples showed the microbial communities of the gut environment to have low diversity, whereas the pharynx and environmental samples had a high diversity value. Such high microbial diversity in the marine environment indicates a rich microbial community and is expected due to the fluctuations of abiotic conditions and nutrients, spatial dispersion of microbes through the marine environment, and dynamic host-species populations that shape the microbiota to the environment [21]. The reduction of species richness from the environment to the gut tissue and digesta indicated a more preferential microbial community comprised of specific taxa in the gut system.

Compartmentalization of microbial community profiles in the gut environment of *S. purpuratus* was demonstrated through Bray-Curtis based 2D MDS plot and dendrogram cluster analysis and was supported through heatmap and ANOSIM/Adonis analysis. Grouping based on sample type was found to be significant through ANOSIM/Adonis statistics, indicating the consistency of microbial taxa across biological replicates representing each group (pharynx, gut tissue, gut digesta, water, and algae). 2D MDS plot analysis supported this low intra-sample variation within groups, and dendrogram analysis showed the gut

tissues to cluster nearer the gut digesta, and the pharynx nearer the water samples. 2D MDS plot and dendrogram analysis of the top 100 OTUs strengthened the trends observed when using the rarefied OTU table. Additionally, an intra-sample variation of the rare taxa was not markedly transformed when determined through 2D MDS plot and dendrogram analysis. However, analysis using rare taxa slightly altered the cluster patterns of groups, showing the gut tissue to cluster nearer the water and pharynx, suggesting shared rare microbial taxa between these groups. Interestingly, although the weighted and unweighted Unifrac approaches did support the significant within-group similarities observed through Bray–Curtis based 2D MDS and dendrogram analysis, the between-group cluster patterns were slightly transformed. For weighted Unifrac, which measures the distance between samples using the relative abundances and phylogenetic relationships of the presented microorganisms [70], the gut tissue and digesta samples clustered closer together. However, through the unweighted method, which considers the presence/absence of taxa and their phylogeny [70], the gut tissue clustered with the water and pharynx samples. Such differences suggest an underlying phylogenetic relatedness between those bacterial microbiota colonizing the gut system based on their abundance. For all ordination analyses performed in this study, the microbial communities of the algae remained the least similar to all samples of the study.

Heatmap analyses supported the 2D MDS plot and dendrogram analysis, showing the contribution of Epsilonproteobacteria (*Arcobacter* and *Sulfurimonas*)

to the unique microbial profile of the gut tissue, and *Psychromonas* in the gut digesta. LEfSe analysis demonstrated the significant effect size of certain key bacterial microbiota contributing to microbial community dissimilarity of the gut tissue and digesta, showing *Arcobacter*, *Sulfurimonas*, and *Tissierella_Soehngenia* in the gut tissue, and *Psychromonas*, Flavobacteriales, and *Vibrio* in the gut digesta. Although *Propionigenium* was found to be more heightened in the gut digesta (14.9% ± 9.6%) than the gut tissue (5.7% ± 4.8%), this taxon did not contribute to a high LDA score. This is likely due to the overlapping standard deviation of the relative abundance values of this taxon (Figure 3-3). Thus, these beta diversity analyses demonstrate the contribution of all taxonomic groups (heightened and rare taxa) in shaping both intra-sample and group-wise microbial profile similarity in the sea urchin *S. purpuratus*, as well as the specific taxa contributing to the unique cluster patterns and effect size of the compartmentalized gut tissue and digesta bacteria.

Public repositories of taxonomic information, such as the GreenGenes (v13.8) database used in this study, often possess repetitive entries, resulting in OTU tables with many redundant taxonomic assignments [25]. By using PhyloToAST (v1.4.0), such bias in the in the representation of 16S rRNA gene and associated taxonomic information can be improved, by binning highly similar taxonomic OTU sequences and assigning taxonomy to a higher resolution (genus/species when possible) through BLAST. Other strategies to improving the diversity measures of OTU tables generated from a targeted HTS approach have been proposed, such as the LULU algorithm that utilizes a post-clustering

strategy to curate an OTU table based on co-occurrence patterns to help eliminate spuriously generated OTUs from the final dataset [110]. In this study, the initial OTU table generated following quality checks resulted in 44,664 unique assignments. However, condensing of redundant taxonomic IDs through PhyloToAST (v1.4.0) reduced the number of unique observations to 776, without any loss of read-count data when compared to the filtered OTU table. Additionally, subsampling of the condensed OTU table to the minimum read count (49,641) increased the beta diversity significance between biological replicate groups. Thus, the combined use of PhyloToAST (v1.4.0) and subsampling to the minimum value, enhanced the resolution of taxa, as well as increased the robustness of the beta analyses measures.

The popularity of targeted high throughput sequencing of microbial 16S rRNA genes has spurred the development of multiple bioinformatics tools, techniques, and approaches, designed to ensure the quality and reliability of the generated results. Traditionally, following the quality assessment and filtering of raw sequence read data, OTUs are chosen based on clustering highly similar sequences at a designated threshold, such as the 97% pairwise similarity described in our analysis, and commonly used as a proxy for species based on the 16S rRNA gene [111,112]. This OTU picking approach will often generate an initially high amount of unique observations, spuriously inflate the alpha diversity, and skew beta diversity measures, due to PCR and/or sequencing errors generating a high number of very rare OTUs [49]. Therefore, an additional filtration step is often necessary, such as the removal of OTUs occurring at

<0.0005% average abundance across all samples as applied in our study and others [50,51], and described for Illumina MiSeq generated sequence read data [49,52]. Alternative approaches to this strategy include a current practice of selecting ASVs as the operational units, which are generated by coupling a denoising step on the raw data that employs a parametric error model to define the frequency of observed sequence variants as it relates to real biological sequence data [43,44]. One such ASV selection tool, DADA2 (v1.10) as implemented in the recently developed QIIME2 package (v2018.11), utilizes a Poisson model to determine these repeated sequence variants [43,44], and was used as an alternative approach to our OTU method. For our OTU-based approach, the initial unfiltered OTU table generated a high number of unique observances, which were further filtered to eliminate potential spurious sequence reads, and then condensed through a BLAST approach with PhyloToAST (v1.4.0) and rarefied to the minimum sequence value. As previously stated, this produced 776 unique OTUs with 49,641 reads across all samples. After using DADA2 (v1.10) through QIIME2 (v2018.11), a total of 1134 unique ASVs with an average read count of ~31,191 were generated across all samples, producing a quantitatively comparable sequence read count to the rarefied OTU table used in our downstream analyses. Additionally, following taxonomic assignment of the representative sequences generated for both the OTUs and ASVs using GreenGenes (v13.8), a similar relative abundance distribution was determined at comparable levels of phylogenetic resolution. Some exceptions included the additional resolution of one OTU assigned as phylum Gammaproteobacteria in

the pharynx tissue, which was further resolved to order Legionellales through the ASV method. However, the similarities of relative abundance are consistent with the most current literature comparing the two strategies [52]. Lastly, it should be noted that although DADA2 denoising may be advantageous in determining real biological sequences associated with rare taxa, this may come at the cost of determining “false positives” as described elsewhere [52].

Predicted functional profiles generated for the gut tissue and digesta at KEGG-Level-3 showed enrichment of metabolic qualities related to energy metabolisms, including oxidative phosphorylation, carbon fixation, methane, and nitrogen metabolisms in the gut tissue. The gut tissue also included heightened motility-related categories, suggesting a potential role of these functional attributes in the colonization of the gut tissue lumen [113], potentially through abiotic factors driving microbial colonization into the more oxygenated sea urchin gut tissue surface [114]. In the gut digesta, KEGG-Level-3 categories related to protein metabolisms (e.g., peptidases and amino acid enzymes, arginine and proline metabolisms) and carbohydrate metabolisms (e.g., starch and sucrose metabolism, pentose phosphate pathway, glycolysis/gluconeogenesis, and ubiquinone and other terpenoid-quinone biosynthesis) suggest a potential role of these bacteria in the digestive physiology of the purple sea urchin. These results are consistent with previous culture-dependent studies of sea urchin gut bacteria in digestion [9] and reflect potential anaerobic metabolisms likely to be performed in this mucous-sequestered niche [115]. LEfSe analysis at the KO Id level indicated the categories that contributed most to the differences between the

distinct microbial communities of the gut system and supported the observed KEGG-Level-2 and 3 categories. The gut tissue presented seven categories related to membrane transporters, such as multiple sugar transport system permease proteins, cobalt/nickel transport system ATP-binding proteins and ferrous iron transport protein A, including ABC transport proteins such as the sulfonate/nitrate/taurine transport system permease protein category. Categories related to nitrogen metabolism, and specifically nitrate reduction, were also observed as heightened. Such nitrate reductive activity occurring in the sea urchin gut has been previously suggested to occur by the gut microbiota of sea urchins [14,116,117]. The gut digesta represented categories related to amino acid metabolism related to threonine and histidine biosynthesis, vitamin metabolisms such as Menaquinone (vitamin K2) production, and categories related to carbohydrate metabolism in the pentose phosphate pathway. These categories support the trends observed in the scatter plot analysis and offer insight into the functional qualities that accompany the life strategies of the gut microbes that may drive their distribution in the compartmentalized gut system. Lastly, although the calculated NSTI values (avg. = 0.139) for the data in this study indicated adequate confidence in the functional predictions of non-human-associated microbial communities based on taxonomic inference alone [75], a shotgun metagenomics approach targeting the metacommunity DNA of a microbiome sample would offer more reliable insight into those genes likely involved in host health and digestion.

We used CoNet (v1.1.1) analysis to identify theoretical modeling of relationships occurring between taxa in the distinctly compartmentalized gut tissue and gut digesta microbial communities [27,28,29]. Although taxa that were noticeably abundant in this study were represented (such as *Propionigenium* in the gut digesta and *Arcobacter* in the gut tissue), there were many low-abundance taxa representing likely key taxa. For example, *Moritella* (Gammaproteobacteria, <1% in the gut digesta) were observed to have a high closeness centrality relative to betweenness centrality and a high degree of edges (16 total) with 10 positive associations to taxa in the gut digesta. Interestingly, two taxa identified as Desulfobacteraceae and *Desulfovibrio* to the highest resolution were included in the top 10 likely key taxa in the gut digesta, were sulfur-reducing bacteria (<1% in the gut digesta) and represented a total degree of 14 and 13 respectively, with the majority of these indicated as co-exclusion relationships. These taxa belong to the phylum Deltaproteobacteria and represent species known to utilize sulfate as electron acceptors [118]. In the gut tissue, Rhodobacteraceae (Alphaproteobacteria) and order Rhodophyta (classified in phylum Cyanobacteria according to the GreenGenes v13.8 database) revealed the same pattern of degree (nine each), closeness and betweenness centrality measures. Additionally, *Arcobacter*, which was highly abundant in the gut tissue (~ 20%), was identified as a key taxon with a degree of nine (with seven positive associations). This genus is comprised of species capable of utilizing elemental sulfur, hydrogen sulfide, and thiosulfate as terminal electron acceptors [119], alluding to a biogeochemical basis for the observed

ecological relationships as suggested previously [120]. However, the relationship between microbial taxa of the gut tissue and gut digesta, including the role or key status of particular microbial taxa will require further verification.

In summary, the results of this study provide insight into the gut bacterial microbiota of *S. purpuratus* grown *in situ*, specifically elaborating (1) the taxonomic distribution, (2) the predicted functional categories assigned to the gut-associated bacterial communities, and (3) key taxa likely involved in maintaining the distribution patterns of gut microbiota through co-occurrence relationships in this evolutionarily and ecologically significant deuterostome. The gut environment demonstrated an allocation of microbial communities in the gut tissue, with a heightened abundance of Epsilonproteobacteria (namely *Arcobacter* and *Sulfurimonas*), and in the gut digesta, a higher abundance of Gammaproteobacteria (such as the psychrophilic genus *Psychromonas*). This trend of microbial compartmentalization has been previously observed in the laboratory [15,24] and naturally occurring [16*] L. variegatus* sea urchin bacteriome from the Gulf of Mexico. In those studies, certain microbial taxa were found to be consistent between the gut systems of both laboratory-raised and naturally occurring organisms, with the naturally occurring organisms showing slightly higher species diversity and richness in the gut system. More specifically, a near-exclusive abundance of *Arcobacter* (Epsilonproteobacteria) was observed in the gut tissue, whereas *Vibrio* (Gammaproteobacteria), a genus common to halophilic temperate marine environments, was most dominant in the gut digesta. Other bacterial microbiota found consistent in the gut digesta of *L. variegatus*

included *Propionigenium* and taxa assigned as family Rhodobacteraceae. For both the laboratory-raised and naturally occurring sea urchins, beta diversity analysis of the pharynx showed cluster patterns that diverged from the gut tissue and digesta and more closely resembled that of their environment. Interestingly, many of the same trends of microbial integration *in L. variegatus* were also observed in the naturally occurring sea urchin *S. purpuratus* gut system of this study. This included an abundance of Epsilonproteobacteria such as *Arcobacter* and *Sulfurimonas* in the gut tissue, Gammaproteobacteria such as *Psychromonas* and *Vibrio*, as well as other common species such as *Propionigenium* in the gut digesta, and a pharynx microbial community resembling the environment. Such selective enrichment and compartmentalization of bacteria in both *S. purpuratus* and *L. variegatus* despite geographical separation support an essential role of specific bacterial taxa to their hosts' health and digestion, and could be further supported by future studies establishing the gut microbiome of laboratory-raised counterparts and/or sea urchins collected at different time-points and locations. Lastly, whether this trend of bacterial enrichment into the sea urchin gut system is the result of (1) naturally occurring microbes in adjacent seawater finding a suitable habitat in the urchin gut environment to flourish, or (2) a host-mediated selective integration of key bacterial microbiota, remains to be verified. However, the similarities identified in taxa across geographical scales suggests that this could be an interesting avenue for future study.

FUNDING

The following are acknowledged for their support of the Microbiome Resource at the University of Alabama at Birmingham, School of Medicine: Comprehensive Cancer Center (P30AR050948), Center for Clinical Translational Science (UL1TR000165), the Heflin Center for Genomic Sciences and the UAB Microbiome Center. Research supported in part by NIH NORC Nutrition Obesity Research Center (P30DK056336) (S.A.Watts). Animal care and use for the experiments conducted in this study was approved by the Oregon Institute of Marine Biology, University of Oregon, and the collection of sea urchins was conducted under the collection permit issued by the Oregon Department of Fish and Wildlife (is: #20366), and the Institutional Animal Care and Use Committee (IACUC-10043; 17 Jan, 2014–current).

ACKNOWLEDGEMENTS

We would like to thank Peter Eipers of the Department of Cell, Developmental and Integrative Biology at the University of Alabama at Birmingham (UAB), for his assistance in high-throughput sequencing for this study. We would also like to thank the Biology Department at UAB for logistics and graduate tuition and stipend support.

CONFLICT OF INTERESTS

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

REFERENCES

1. Ebert T.A., Schroeter S.C., Dixon J.D., Kalvass P. Settlement patterns of red and purple sea urchins (*Strongylocentrotus franciscanus* and *S. purpuratus*) in California, USA. Mar. Ecol. Prog. Ser. 1994;111:41–52. doi: 10.3354/meps111041.

2. Dethier M.N. Disturbance and recovery in intertidal pools: Maintenance of mosaic patterns. Ecol. Monogr. 1984;54:99–118. doi: 10.2307/1942457.

3. Lafferty K.D., Kushner D. Population regulation of the purple sea urchin (*Strongylocentrotus purpuratus*) at the California Channel Islands. In: Browne D.R., Mitchell K.L., Chaney H.W., editors. Proceedings of the 5th California Islands Symposium. Minerals Management Service; Santa Barbara, CA, USA: 2000. pp. 379–381. Publication 99-0038.

4. Watanabe J.M., Harrold C. Destructive grazing by sea urchins *Strongylocentrotus* spp. in a central California kelp forest: Potential roles of recruitment, depth, and predation. Mar. Ecol. Prog. Ser. 1991:125–141. doi: 10.3354/meps071125.

5. Metaxas A., Scheibling R.E. Community structure and organization of tidepools. Mar. Ecol. Prog. Ser. 1993:187–198. doi: 10.3354/meps098187.

6. Tegner M.J., Dayton P.K. Ecosystem effects of fishing in kelp forest communities. ICES J. Mar. Sci. 2000;57:579–589. doi: 10.1006/jmsc.2000.0715.

7. Davidson T.M., Grupe B.M. Habitat modification in tidepools by bioeroding sea urchins and implications for fine-scale community structure. Mar. Ecol. 2015;36:185–194. doi: 10.1111/maec.12134.

8. Sea Urchin Genome Sequencing Consortium The genome of the sea urchin *Strongylocentrotus purpuratus*. Science. 2006;314:941–952. doi: 10.1126/science.1133609.

9. Lasker R., Giese A.C. Nutrition of the sea urchin, *Strongylocentrotus purpuratus*. Biol. Bull. 1954;106:328–340. doi: 10.2307/1538767.

10. Holland N.D., Ghiselin M.T. A comparative study of gut mucous cells in thirty-seven species of the class Echinoidea (Echinodermata) Biol. Bull. 1970;138:286–305. doi: 10.2307/1540213.

11. De Ridder C., Jangoux M. Digestive system: Echinoidea. In: Jangoux M., Lawrence J.M., editors. Echinoderm Nutrition. A.A. Balkema Publ.; Rotterdam, The Netherlands: 1982. pp. 213–234.

12. Ziegler A., Mooi R., Rolet G., De Ridder C. Origin and evolutionary plasticity of the gastric caecum in sea urchins (Echinodermata: Echinoidea) BMC Evol. Biol. 2010;10:313. doi: 10.1186/1471-2148-10-313.

13. Sawabe T., Oda Y., Shiomi Y., Ezura Y. Alginate degradation by bacteria isolated from the gut of sea urchins and abalones. Microb. Ecol. 1995;30:193–202. doi: 10.1007/BF00172574.

14. Fong W., Mann K.H. Role of gut flora in the transfer of amino acids through a marine food chain. Can. J. Fish. Aquat. Sci. 1980;37:88–96. doi: 10.1139/f80-009.

15. Hakim J.A., Koo H., Dennis L.N., Kumar R., Ptacek T., Morrow C.D., Lefkowitz E.J., Powell M.L., Bej A.K., Watts S.A. An abundance of Epsilonproteobacteria revealed in the gut microbiome of the laboratory cultured sea urchin, *Lytechinus variegatus*. Front. Microbiol. 2015;6:1047. doi: 10.3389/fmicb.2015.01047.

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

17. Sauchyn L.K., Lauzon-Guay J.S., Scheibling R.E. Sea urchin fecal production and accumulation in a rocky subtidal ecosystem. Aquat. Biol. 2011;13:215–223. doi: 10.3354/ab00359.

18. Sauchyn L.K., Scheibling R.E. Degradation of sea urchin feces in a rocky subtidal ecosystem: Implications for nutrient cycling and energy flow. Aquat. Biol. 2009;6:99–108. doi: 10.3354/ab00171.

19. Sauchyn L.K., Scheibling R.E. Fecal production by sea urchins in native and invaded algal beds. Mar. Ecol. Prog. Ser. 2009;396:35–48. doi: 10.3354/meps08296.

20. Schram J.B., Kobelt J.N., Dethier M.N., Galloway A.W.E. Trophic transfer of macroalgal fatty acids in two urchin species: Digestion, egestion, and tissue building. Front. Ecol. Evol. 2018;6:83. doi: 10.3389/fevo.2018.00083.

21. Troussellier M., Escalas A., Bouvier T., Mouillot D. Sustaining rare marine microorganisms: Macroorganisms as repositories and dispersal agents of microbial diversity. Front. Microbiol. 2017;8:947. doi: 10.3389/fmicb.2017.00947.

22. Ley R.E., Turnbaugh P.J., Klein S., Gordon J.I. Microbial ecology: Human gut microbes associated with obesity. Nature. 2006;444:1022–1023. doi: 10.1038/4441022a.

23. Turnbaugh P.J., Ley R.E., Mahowald M.A., Magrini V., Mardis E.R., Gordon J.I. An obesity-associated gut microbiome with increased capacity for energy harvest. Nature. 2006;444:1027–1031. doi: 10.1038/nature05414.

24. Brothers C.J., Van Der Pol W.J., Morrow C.D., Hakim J.A., Koo H., McClintock J.B. Ocean warming alters predicted microbiome functionality in a common sea urchin. Proc. R. Soc. B. 2018;285:20180340. doi: 10.1098/rspb.2018.0340.

25. Dabdoub S.M., Fellows M.L., Paropkari A.D., Mason M.R., Huja S.S., Tsigarida A.A., Kumar P.S. PhyloToAST: Bioinformatics tools for species-level analysis and visualization of complex microbial datasets. Sci. Rep. 2016;6:29123. doi: 10.1038/srep29123.

26. Caporaso J.G., Kuczynski J., Stombaugh J., Bittinger K., Bushman F.D., Costello E.K., Fierer N., Gonzalez Peña A., Goodrich J.K., Gordon J.G., et al. QIIME allows analysis of high-throughput community sequencing data. Nat. Meth. 2010;7:335–336. doi: 10.1038/nmeth.f.303.

27. Faust K., Sathirapongsasuti J.F., Izard J., Segata N., Gevers D., Raes J., Huttenhower C. Microbial co-occurrence relationships in the human microbiome. PLOS Comput. Biol. 2012;8:e1002606. doi: 10.1371/journal.pcbi.1002606.

28. Faust K., Raes J. Microbial interactions: From networks to models. Nat. Rev. Microbiol. 2012;10:538–550. doi: 10.1038/nrmicro2832.

29. Faust K., Raes J. CoNet app: Inference of biological association networks using Cytoscape. F1000Research. 2016;5:1519. doi: 10.12688/f1000research.9050.1.

30. Berry D., Widder S. Deciphering microbial interactions and detecting keystone species with co-occurrence networks. Front. Microbiol. 2014;5:219. doi: 10.3389/fmicb.2014.00219.

31. Banerjee S., Schlaeppi K., Heijden M.G. Keystone taxa as drivers of microbiome structure and functioning. Nat. Rev. Microbiol. 2018;16:567–576. doi: 10.1038/s41579-018-0024-1.

32. Song S.J., Amir A., Metcalf J.L., Amato K.R., Xu Z.Z., Humphrey G., Knight R. Preservation methods differ in fecal microbiome stability, affecting suitability for field studies. mSystems. 2016;1:e00021-16. doi: 10.1128/mSystems.00021-16.

33. Caporaso J.G., Lauber C.L., Walters W.A., Berg-Lyons D., Lozupone C.A., Turnbaugh P.J., Fierer N., Knight R. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc. Natl. Acad. Sci. USA. 2011;108(Suppl. 1):4516–4522. doi: 10.1073/pnas.1000080107.

34. Caporaso J.G., Lauber C.L., Walters W.A., Berg-Lyons D., Huntley J., Fierer N., Owens S.M., Betley J., Fraser L., Bauer M., et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J. 2012;6:1621–1624. doi: 10.1038/ismej.2012.8.

35. Kumar R., Eipers P., Little R.B., Crowley M., Crossman D.K., Lefkowitz E.J., Morrow C.D. Getting started with microbiome analysis: Sample acquisition to bioinformatics. Curr. Protoc. Hum. Genet. 2014;82:18.8.1–18.8.29. doi: 10.1002/0471142905.hg1808s82.

36. Kozich J.J., Westcott S.L., Baxter N.T., Highlander S.K., Schloss P.D. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Appl. Environ. Microbiol. 2013;79:5112–5120. doi: 10.1128/AEM.01043-13.

37. Ausubel F.M. In: Current Protocols in Molecular Biology. Ausubel F.M., Brent R., Kingston R.F., Moore D.D., Seidman J.G., Smith J.A., Struhl K., editors. Publishing Associates and Wiley-Interscience; New York, NY, USA: 1987.

38. Cock P.J.A., Fields C.J., Goto N., Heuer M.L., Rice P.M. The Sanger FASTQ File Format for Sequences with Quality Scores, and the

Solexa/Illumina FASTQ Variants. Nucleic Acids Res. 2009;38:1767–1771. doi: 10.1093/nar/gkp1137.

39. Andrews S. FastQC: A quality Control Tool for High Throughput Sequence Data. [(accessed on 12 January 2019)];2010 Available online: https://www.bioinformatics.babraham.ac.uk/projects/fastqc.

40. Gordon A., Hannon G.J. “FASTX-Toolkit,” FASTQ/A Short-Reads pre-Processing Tools. [(accessed on 12 January 2019)];2012 Available online: http://hannonlab.cshl.edu/fastx_toolkit.

41. Edgar R.C. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 2010;26:2460–2461. doi: 10.1093/bioinformatics/btq461.

42. Bolyen E., Rideout J.R., Dillon M.R., Bokulich N.A., Abnet C., Al-Ghalith G.A., Alexander H., Alm E.J., Arumugam M., Asnicar F., et al. QIIME 2: Reproducible, interactive, scalable, and extensible microbiome data science. PeerJ. 2018 doi: 10.7287/peerj.preprints.27295v1.

43. Callahan B.J., McMurdie P.J., Holmes S.P. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. ISME J. 2017;11:2639–2643. doi: 10.1038/ismej.2017.119.

44. Callahan B.J., McMurdie P.J., Rosen M.J., Han A.W., Johnson A.J.A., Holmes S.P. DADA2: High-resolution sample inference from Illumina amplicon data. Nat. Methods. 2016;13:581. doi: 10.1038/nmeth.3869.

45. Blaxter M., Mann J., Chapman T., Thomas F., Whitton C., Floyd R., Abebe E. Defining operational taxonomic units using DNA barcode data. Philos. Trans. Royal Soc. B. 2005;1462:1935–1943. doi: 10.1098/rstb.2005.1725.

46. Wang Q., Garrity G.M., Tiedje J.M., Cole J.R. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl. Environ. Microbiol. 2007;73:5261–5267. doi: 10.1128/AEM.00062-07.

47. Desantis T.Z., Hugenholtz P., Larsen N., Rojas M., Brodie E.L., Keller K., Huber T., Dalevi D., Hu P., Andersen G.L. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl. Environ. Microbiol. 2006;72:5069–5072. doi: 10.1128/AEM.03006-05.

48. McDonald D., Price M.N., Goodrich J., Nawrocki E.P., DeSantis T.Z., Probst A., Anderson G.L., Knight R., Hugenholtz P. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary

analyses of bacteria and archaea. ISME J. 2011;6:610–618. doi: 10.1038/ismej.2011.139.

49. Bokulich N.A., Subramanian S., Faith J.J., Gevers D., Gordon J.I., Knight R., Mills D.A., Caporaso J.G. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. Nat. Methods. 2013;10:57–59. doi: 10.1038/nmeth.2276.

50. De Souza R.S.C., Okura V.K., Armanhi J.S.L., Jorrín B., Lozano N., Da Silva M.J., Imperial J. Unlocking the bacterial and fungal communities assemblages of sugarcane microbiome. Sci. Rep. 2016;6:28774. doi: 10.1038/srep28774.

51. Pechal J.L., Schmidt C.J., Jordan H.R., Benbow M.E. Frozen: Thawing and its effect on the postmortem microbiome in two pediatric cases. J. Forensic Sci. 2017;62:1399–1405. doi: 10.1111/1556-4029.13419.

52. Nearing J.T., Douglas G.M., Comeau A.M., Langille M.G. Denoising the Denoisers: An independent evaluation of microbiome sequence error-correction approaches. PeerJ. 2018;6:e5364. doi: 10.7717/peerj.5364.

53. Altschul S.F., Gish W., Miller W., Myers E.W., Lipman D.J. Basic local alignment search tool. J. Mol. Biol. 1990;215:403–410. doi: 10.1016/S0022-2836(05)80360-2.

54. De Cárcer D.A., Denman S.E., McSweeney C., Morrison M. Evaluation of subsampling-based normalization strategies for tagged high-throughput sequencing datasets from gut microbiomes. Appl. Environ. Microbiol. 2011;77:8795–8798. doi: 10.1128/AEM.05491-11.

55. Pruesse E., Peplies J., Glöckner F.O. SINA: Accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics. 2012;28:1823–1829. doi: 10.1093/bioinformatics/bts252.

56. Cole J.R., Wang Q., Cardenas E., Fish J., Chai B., Farris R.J., Kulam-Syed-Mohideen A.S., McGarrell D.M., Marsh T., Garrity G.M., et al. The Ribosomal Database Project: Improved alignments and new tools for rRNA analysis. Nucleic Acids Res. 2008;37(Suppl. 1):D141–D145. doi: 10.1093/nar/gkn879.

57. Pruesse E., Quast C., Knittel K., Fuchs B.M., Ludwig W., Peplies J., Glöckner F.O. SILVA: A comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res. 2007;35:7188–7196. doi: 10.1093/nar/gkm864.

58. Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V., et al. Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. 2011;12:2825–2830.

59. Parks D.H., Tyson G.W., Hugenholtz P., Beiko R.G. STAMP: Statistical analysis of taxonomic and functional profiles. Bioinformatics. 2014;30:3123–3124. doi: 10.1093/bioinformatics/btu494.

60. Shannon C.E. A mathematical theory of communication. Bell Syst. Tech. J. 1948;27:379–423. doi: 10.1002/j.1538-7305.1948.tb01338.x.

61. Hill T.C., Walsh K.A., Harris J.A., Moffett B.F. Using ecological diversity measures with bacterial communities. FEMS Microbiol. Ecol. 2003;43:1–11. doi: 10.1111/j.1574-6941.2003.tb01040.x.

62. Marcon E., Scotti I., Hérault B., Rossi V., Lang G. Generalization of the partitioning of Shannon diversity. PLoS ONE. 2014;9:e90289. doi: 10.1371/journal.pone.0090289.

63. Simpson E.H. Measurement of diversity. Nature. 1949;163:688. doi: 10.1038/163688a0.

64. Kruskal W.H., Wallis W.A. Use of ranks in one-criterion variance analysis. J. Am. Stat. Assoc. 1952;47:583–621. doi: 10.1080/01621459.1952.10483441.

65. Bray J.R., Curtis J.T. An ordination of the upland forest communities of southern Wisconsin. Ecol. Monograph. 1957;27:325–349. doi: 10.2307/1942268.

66. Anderson M.J. A new method for non-parametric multivariate analysis of variance. Austral Ecol. 2001;26:32–46. doi: 10.1111/j.1442-9993.2001.01070.pp.x.

67. Clarke K.R. Non-parametric multivariate analyses of changes in community structure. Austral Ecol. 1993;18:117–143. doi: 10.1111/j.1442-9993.1993.tb00438.x.

68. Oksanen J., Kindt R., Legendre P., O'Hara B., Stevens M.H.H., Oksanen M.J., Suggests M. The vegan package. Community Ecol. Package. 2007;10:631–637.

69. Clarke K.R., Gorley R.N. PRIMER v6: User Manual/Tutorial (Plymouth Routines in Multivariate Ecological Research) PRIMER-E Ltd.; Plymouth, UK: 2006.

70. Lozupone C., Lladser M.E., Knights D., Stombaugh J., Knight R. UniFrac: An effective distance metric for microbial community comparison. ISME J. 2011;5:169–172. doi: 10.1038/ismej.2010.133.

71. gplots: Various R Programming Tools for Plotting Data. [(accessed on 22 October 2018)]; Available online: cran.r-project.org/web/packages/gplots.

72. RColorBrewer: ColorBrewer Palettes. [(accessed on 22 October 2018)]; Available online: cran.r-project.org/web/packages/RColorBrewer.

73. Segata N., Izard J., Waldron L., Gevers D., Miropolsky L., Garrett W.S., Huttenhower C. Metagenomic biomarker discovery and explanation. Genome Biol. 2011;12:R60. doi: 10.1186/gb-2011-12-6-r60.

74. Fisher R.A. The use of multiple measurements in taxonomic problems. Ann. Hum. Genet. 1936;7:179–188. doi: 10.1111/j.1469-1809.1936.tb02137.x.

75. Langille M.G., Zaneveld J., Caporaso J.G., McDonald D., Knights D., Reyes J.A., Clemente J.C., Burkepile D.E., Thurber R.L.V., Knight R., et al. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nat. Biotechnol. 2013;31:814. doi: 10.1038/nbt.2676.

76. Shannon P., Markiel A., Ozier O., Baliga N.S., Wang J.T., Ramage D., Amin N., Schwikowski B., Ideker T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13:2498–2504. doi: 10.1101/gr.1239303.

77. Choma M., Bárta J., Šantrůčková H., Urich T. Low abundance of Archaeorhizomycetes among fungi in soil metatranscriptomes. Sci. Rep. 2016;6:38455. doi: 10.1038/srep38455.

78. Galton F. Regression towards mediocrity in hereditary stature. J. R. Anthropol. Inst. 1886;15:246–263. doi: 10.2307/2841583.

79. Pearson K. Note on regression and inheritance in the case of two parents. Proc. R. Soc. Lond. 1895;58:240–242. doi: 10.1098/rspl.1895.0041.

80. Spearman C. The proof and measurement of association between two things. Am. J. Psychol. 1904;15:72–101. doi: 10.2307/1412159.

81. Kullback S., Leibler R.A. On information and sufficiency. Ann. Math. Stat. 1951;22:79–86. doi: 10.1214/aoms/1177729694.

82. Cover T.M., Thomas J.A. Elements of Information Theory. Volume 1. John Wiley and Sons, Inc.; New York, NY, USA: 1991. Information Theory and Statistics; pp. 279–335.

83. Brown M.B. 400: A method for combining non-independent, one-sided tests of significance. Biometrics. 1975;31:987–992. doi: 10.2307/2529826.

84. Wiese R., Eiglsperger M., Kaufmann M. yFiles—Visualization and Automatic Layout of Graphs. In: Jünger M., Mutzel P., editors. Graph Drawing Software (Mathematics and Visualization) Springer; Berlin/Heidelberg, Germany: 2004. pp. 173–191.

85. Assenov Y., Ramírez F., Schelhorn S.E., Lengauer T., Albrecht M. Computing topological parameters of biological networks. Bioinformatics. 2007;24:282–284. doi: 10.1093/bioinformatics/btm554.

86. Ma B., Wang H., Dsouza M., Lou J., He Y., Dai Z., Brookes P.C., Xu J., Gilbert J.A. Geographic patterns of co-occurrence network topological features for soil microbiota at continental scale in eastern China. ISME J. 2016;10:1891. doi: 10.1038/ismej.2015.261.

87. Wang Z., Bafadhel M., Haldar K., Spivak A., Mayhew D., Miller B.E., Tal-Singer R., Johnston S.L., Ramsheh M.Y., Barer M.R. Lung microbiome dynamics in chronic obstructive pulmonary disease exacerbations. Eur. Respir. J. 2016;47:1082–1092. doi: 10.1183/13993003.01406-2015.

88. Wang Y., Sheng H.F., He Y., Wu J.Y., Jiang Y.X., Tam N.F.Y., Zhou H.W. Comparison of the levels of bacterial diversity in freshwater, intertidal wetland, and marine sediments by using millions of Illumina tags. Appl. Environ. Microbiol. 2012;78:8264–8271. doi: 10.1128/AEM.01821-12.

89. Wirsen C.O., Sievert S.M., Cavanaugh C.M., Molyneaux S.J., Ahmad A., Taylor L., Delong E.F., Taylor C.D. Characterization of an autotrophic sulfide-oxidizing marine *Arcobacter* sp. that produces filamentous sulfur. Appl. Environ. Microbiol. 2002;68:316–325. doi: 10.1128/AEM.68.1.316-325.2002.

90. Sievert S.M., Hügler M., Taylor C.D., Wirsen C.O. Sulfur Oxidation at Deep-Sea Hydrothermal Vents. In: Dahl C., Friedrich C.G., editors. Microbial Sulfur Metabolism. Springer; Berlin/Heidelberg, Germany: 2008. pp. 238–258.

91. Inagaki F., Takai K., Kobayashi H., Nealson K.H., Horikoshi K. *Sulfurimonas autotrophica* gen. nov.; sp. nov.; a novel sulfur-oxidizing ε-proteobacterium isolated from hydrothermal sediments in the Mid-

Okinawa Trough. Int. J. Syst. Evol. Microbiol. 2003;53:1801–1805. doi: 10.1099/ijs.0.02682-0.

92. Campbell B.J., Engel A.S., Porter M.L., Takai K. The versatile ε-proteobacteria: Key players in sulphidic habitats. Nat. Rev. Microbiol. 2006;4:458–468. doi: 10.1038/nrmicro1414.

93. Durand L., Zbinden M., Cueff-Gauchard V., Duperron S., Roussel E.G., Shillito B., Cambon-Bonavita M.A. Microbial diversity associated with the hydrothermal shrimp *Rimicaris exoculata* gut and occurrence of a resident microbial community. FEMS Microbiol. Ecol. 2009;71:291–303. doi: 10.1111/j.1574-6941.2009.00806.x.

94. Romero J., Garcia-Varela M., Laclette J., Espejo R. Bacterial 16S rRNA gene analysis revealed that bacteria related to *Arcobacter* spp. constitute an abundant and common component of the oyster microbiota (*Tiostrea chilensis*) Microb. Ecol. 2002;44:365–371. doi: 10.1007/s00248-002-1063-7.

95. Ho T.W., Hwang J.S., Cheung M.K., Kwan H.S., Wong C.K. Dietary analysis on the shallow-water hydrothermal vent crab *Xenograpsus testudinatus* using Illumina sequencing. Mar. Biol. 2015;162:1787–1798. doi: 10.1007/s00227-015-2711-z.

96. Urakawa H., Dubilier N., Fujiwara Y., Cunningham D.E., Kojima S., Stahl D.A. Hydrothermal vent gastropods from the same family (Provannidae) harbour ε-and γ-proteobacterial endosymbionts. Environ. Microbiol. 2005;7:750–754. doi: 10.1111/j.1462-2920.2005.00753.x.

97. Beinart R.A., Sanders J.G., Faure B., Sylva S.P., Lee R.W., Becker E.L., Gartman A., Luther G.W., Seewald J.S., Fisher C.R. Evidence for the role of endosymbionts in regional-scale habitat partitioning by hydrothermal vent symbioses. Proc. Natl. Acad. Sci. USA. 2012;109:E3241–E3250. doi: 10.1073/pnas.1202690109.

98. Beinart R., Gartman A., Sanders J., Luther G., Girguis P. The uptake and excretion of partially oxidized sulfur expands the repertoire of energy resources metabolized by hydrothermal vent symbioses. Proc. R. Soc. B. 2015;282:20142811. doi: 10.1098/rspb.2014.2811.

99. Nichols D.S. Prokaryotes and the input of polyunsaturated fatty acids to the marine food web. FEMS Microbiol. Lett. 2003;219:1–7. doi: 10.1016/S0378-1097(02)01200-4.

100. Kawasaki K., Nogi Y., Hishinuma M., Nodasaka Y., Matsuyama H., Yumoto I. *Psychromonas marina* sp. nov.; a novel halophilic, facultatively

psychrophilic bacterium isolated from the coast of the Okhotsk Sea. Int. J. Syst. Evol. Microbiol. 2002;52:1455–1459. doi: 10.1099/ijs.0.01919-0.

101. Groudieva T., Grote R., Antranikian G. *Psychromonas arctica* sp. nov.; a novel psychrotolerant, biofilm-forming bacterium isolated from Spitzbergen. Int. J. Syst. Evol. Microbiol. 2003;53:539–545. doi: 10.1099/ijs.0.02182-0.

102. Breezee J., Cady N., Staley J. Subfreezing growth of the sea ice bacterium "*Psychromonas ingrahamii*" Microb. Ecol. 2004;47:300–304. doi: 10.1007/s00248-003-1040-9.

103. Hosoya S., Jang J.-H., Yasumoto-Hirose M., Matsuda S., Kasai H. *Psychromonas agarivorans* sp. nov.; a novel agarolytic bacterium. Int. J. Syst. Evol. Microbiol. 2009;59:1262–1266. doi: 10.1099/ijs.0.003244-0.

104. Schink B. The Prokaryotes. Springer; New York, NY, USA: 2006. The genus *Propionigenium*; pp. 955–959.

105. Smith P., Willemsen D., Popkes M., Metge F., Gandiwa E., Reichard M., Valenzano D.R. Regulation of life span by the gut microbiota in the short-lived African turquoise killifish. eLife. 2017;66:e27014. doi: 10.7554/eLife.27014.

106. Cardona E., Gueguen Y., Magré K., Lorgeoux B., Piquemal D., Pierrat F., Noguier F., Saulnier D. Bacterial community characterization of water and intestine of the shrimp *Litopenaeus stylirostris* in a biofloc system. BMC Microbiol. 2016;16:157. doi: 10.1186/s12866-016-0770-z.

107. Thomas F., Barbeyron T., Tonon T., Génicot S., Czjzek M., Michel G. Characterization of the first alginolytic operons in a marine bacterium: From their emergence in marine Flavobacteriia to their independent transfers to marine Proteobacteria and human gut Bacteroides. Environ. Microbiol. 2012;14:2379–2394. doi: 10.1111/j.1462-2920.2012.02751.x.

108. Rabus R., Ruepp A., Frickey T., Rattei T., Fartmann B., Stark M., Bauer M., Zibat A., Lombardot T., Becker I., Amann J., et al. The genome of *Desulfotalea psychrophila*, a sulfate-reducing bacterium from permanently cold Arctic sediments. Environ. Microbiol. 2004;6:887–902. doi: 10.1111/j.1462-2920.2004.00665.x.

109. Lockhart C., Fraga D. Characterization of the arginine kinase from *Desulfotalea psychrophila* LSv54: The effects of environmental conditions and catalytic domain sequence variations on enzymatic turnover. FASEB J. 2007;21:A299.

110. Frøslev T.G., Kjøller R., Bruun H.H., Ejrnæs R., Brunbjerg A.K., Pietroni C., Hansen A.J. Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. Nat. Commun. 2017;8:1188. doi: 10.1038/s41467-017-01312-x.

111. Konstantinidis K.T., Tiedje J.M. Genomic insights that advance the species definition for prokaryotes. Proc. Natl. Acad. Sci. USA. 2005;102:2567–2572. doi: 10.1073/pnas.0409727102.

112. Stackebrandt E., Goebel B.M. Taxonomic note: A place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. Int. J. Syst. Evol. Microbiol. 1994;44:846–849. doi: 10.1099/00207713-44-4-846.

113. Tamar E., Koler M., Vaknin A. The role of motility and chemotaxis in the bacterial colonization of protected surfaces. Sci. Rep. 2016;6:19616. doi: 10.1038/srep19616.

114. Thorsen M.S. Microbial activity, oxygen status and fermentation in the gut of the irregular sea urchin *Echinocardium cordatum* (Spatangoida: Echinodermata) Mar. Biol. 1998;132:423–433. doi: 10.1007/s002270050408.

115. Meziti A., Kormas K.A., Pancucci-Papadopoulou M.A., Thessalou-Legaki M. Bacterial phylotypes associated with the digestive tract of the sea urchin *Paracentrotus lividus* and the ascidian *Microcosmus* sp. Russ. J. Mar. Biol. 2007;33:84–91. doi: 10.1134/S1063074007020022.

116. Guerinot M.L., Fong W., Patriquin D.G. Nitrogen fixation (acetylene reduction) associated with sea urchins (*Strongylocentrotus droebachiensis*) feeding on seaweeds and eelgrass. J. Fish. Res. Board Can. 1977;34:416–420. doi: 10.1139/f77-067.

117. Guerinot M.L., Patriquin D.G. N2-fixing vibrios isolated from the gastrointestinal tract of sea urchins. Can. J. Microbiol. 1981;27:311–317. doi: 10.1139/m81-048.

118. Achá D., Hintelmann H., Yee J. Importance of sulfate reducing bacteria in mercury methylation and demethylation in periphyton from Bolivian Amazon region. Chemosphere. 2011;82:911–916. doi: 10.1016/j.chemosphere.2010.10.050.

119. Roalkvam I., Drønen K., Stokke R., Daae F.L., Dahle H., Steen I.H. Physiological and genomic characterization of *Arcobacter anaerophilus* IR-1 reveals new metabolic features in Epsilonproteobacteria. Front. Microbiol. 2015;6:987. doi: 10.3389/fmicb.2015.00987.

120. Becker P.T., Samadi S., Zbinden M., Hoyoux C., Compère P., De Ridder C. First insights into the gut microflora associated with an echinoid from wood falls environments. Cah. Biol. Mar. 2009;50:343–352.

CHAPTER IV: COMPARISON OF GUT MICROBIOTA IN NATURALLY OCCURRING AND LABORATORY AQUACULTURE SEA URCHIN *LYTECHINUS VARIEGATUS* REVEALED DIFFERENCES IN THE COMMUNITY COMPOSITION AND PREDICTED FUNCTIONS

by

JOSEPH A. HAKIM, CASEY D. MORROW, STEPHEN A. WATTS, AND ASIM K. BEJ

Submitted to *PLoS One*

Format adapted and errata corrected for dissertation

ABSTRACT

In this study, the gut microbiota of the naturally occurring sea urchin *Lytechinus variegatus* from the Gulf of Mexico (ENV) was compared with sea urchins fed with formulated diet and maintained in the laboratory aquaculture conditions (LAB) by using targeted metagenomics of the 16S rRNA gene and bioinformatics tools. Overall, the gut tissue of both groups showed a high abundance of Campylobacteraceae (>88%) that were assigned as *Arcobacter* spp. (~90% sequence similarity) through the Basic Local Alignment Search Tool (BLAST). In contrast, the microbial composition of the gut digesta from both groups had taxa from Gammaproteobacteria, particularly *Vibrio* spp. to be most abundant. However, the ENV group showed a higher overall taxonomic richness and abundance of *Propionigenium*, *Photobacterium*, and taxa assigned to order Flavobacterales, whereas the LAB group showed less richness and noticeable abundance of *Arcobacter*, *Agarivorans,* and *Shewanella*. Predicted functional analysis of the microbiota in the gut tissue of both groups showed a trend of energy-related metabolisms as compared to the gut digesta. In contrast, amino acid, carbohydrate, and lipid metabolisms although heightened in the gut digesta of both groups, these categories were more prominent in the ENV group. The outcome of this study revealed that although the gut tissue maintained a similar microbial profile, there were marked differences in the diversity and richness in the gut digesta of *L. variegatus* from their natural habitat than in the laboratory aquaculture conditions.

INTRODUCTION

Diverse organisms from all domains of life cohabit in the nearshore marine ecosystems of North America where they thrive by modulating their community structure with the fluctuating biotic and abiotic factors, food sources and other perturbances, such as natural calamities or human activities [1-3]. Besides these environmental factors, the resiliency of the ecosystem is also contingent on the composition of microbial communities and their metabolic processes that help sustain some of the crucial trophic functions [2]. However, oftentimes organisms are transferred from their natural habitat to laboratory aquaculture conditions for various scientific studies. These organisms, unlike their natural habitats, are often fed with a formulated feed consisting of balanced nutrients. Such controlled laboratory conditions have been linked to shaping the gut microbiota in various model organisms such as *Mus musculus* [4], *Danio rerio* [5], *Caenorhabditis elegans* [6, 7], and *Drosophila* [8]. Similarly, the green sea urchin *Lytechinus variegatus* (order Temnopleuroida, family Toxopneustidae) are often collected from their natural habitats, maintained in the laboratory and used for various basic and applied research [9]. In addition, the potential for the commercial use of sea urchins in the aquaculture industry has also been investigated [10, 11].

L. variegatus is primarily a herbivore echinoderm found along the South-Eastern Coast and into the Gulf of Mexico [12]. Although considered omnivorous, this sea urchin primarily grazes upon seagrass, algae, and decomposed materials [12-15]. *L. variegatus* represents a uniquely compartmentalized deuterostome gut system, in which the ingested food along with the natural

microbial communities from their habitat is encapsulated in the pharynx by a thick mucus layer often referred as gut digesta. These microbial communities in the digesta remain separated from that of the gut tissue during their passage through the gut lumen until egested [16-19]. Previously reported culture-dependent studies of sea urchins from their habitat have shown representative bacteria from the gut capable of alginolytic activities [20]. In addition, gelatin, protein, and amino acid assimilation by the gut microbiota have also been elaborated in sea urchins in the laboratory environment both through *in vitro* [21-23] and *in situ* approaches [24]. However, the culture conditions of the laboratory aquaculture sea urchins are normally fed with a diet consisting of combinations of naturally occurring algal and seagrass sources [10, 25] or a formulated feed optimized for their nutritional requirements and health [26]. These conditions contrast the food in their natural habitat, including the associated diverse marine microorganisms [27-34]. Thus, the changes in their diet from the natural habitat to a laboratory-formulated feed could potentially reshape the microbiota and their metabolic roles in the gut ecosystem.

Previously we have reported a comprehensive taxonomic profile of the microbial community, and their predictive metabolic functions, in the gut ecosystem of naturally occurring and laboratory aquaculture *L. variegatus* by using 16S rRNA gene-targeted high-throughput amplicon sequencing (HTS) and bioinformatics tools [17, 18]. Although noticeable variations of microbial community compositions in other laboratory animals have been reported [35], such changes have not been established in the sea urchin. Thus, the objective of

this study was to compare the microbial communities and metabolic profiles of naturally occurring and laboratory aquaculture sea urchins using the metagenomics approach and bioinformatics tools.

The microbial communities of the sea urchin gut pellets also play an important role in the biogeochemical cycles of the marine environment. In previous studies, it has been shown that the microbial community composition and their metabolic processes in the gut digesta remains stable following egestion into the environment [15,16,17]. For example, studies examining the chemical composition of *S. droebachiensis* egesta through flash combustion have shown increased in lipid, nitrogen, and organic carbon, and decreases in the carbon: nitrogen ratio, indicating the metabolic importance of the bacterial communities in the degradation and transformation of the contents within the pellets into a nutrient-rich food source for nearby marine organisms [17,18,19,20]. Additionally, it has been suggested that urchin gut microbiota are responsible for differences in algal digestion and synthesis of essential long chain fatty acids in both *S. purpuratus* and *S. droebachiensis* [20]. The pelleted egesta has also been considered as a mode for the dispersion of sea urchin gut microbiota into their environment [17,21]. Most of the early studies of the potential role of microbial communities in digestive processes of sea urchin gut ecosystem were conducted by culture-dependent methods [9]. However, recent advancement of the culture-independent method of high-throughput sequencing (HTS) of 16S rRNA genes from the metacommunity DNA has been shown to provide gut microbial community composition with high taxonomic coverage,

including their potential metabolic functions [22,23]. Recently, the application of HTS on the V4 hypervariable segment of the 16S rRNA gene of the gut bacteriome of *Lytechinus variegatus* from the U.S. Gulf of Mexico [15,16,24] demonstrated distinct microbial community compositions between the gut tissue and gut digesta. Specifically, representative taxa from class Epsilonproteobacteria (assigned as *Arcobacter*/*Sulfuricurvum* through the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST)) dominated the gut tissue, whereas Gammaproteobacteria (namely *Vibrio*) were heightened in the digesta [15]. Additionally, predictive functional profiling of these compartmentalized microbial communities showed energy metabolisms such as oxidative phosphorylation, carbon fixation, nitrogen, methane, and sulfur metabolisms to be heightened in the gut tissue, compared to carbohydrate, amino acid, and lipid metabolisms in the digesta [16].

Such HTS technology and bioinformatics analyses applied to the gut ecosystem of the naturally occurring sea urchin *S. purpuratus* can help establish a comprehensive microbial community composition and provide crucial information into the gut bacterial taxa and likely functions performed as they relate to host health and digestion. In this study, we have elaborated the microbial profiles of the gut tissue, pharynx tissue, and mucous enveloped gut digesta of *S. purpuratus* collected from their natural rocky tide pool habitat on the coastline of Oregon, USA. In addition, samples of the tide pool seawater and adjacent algal community were collected and also analyzed for comparison with the gut tissue and digesta. We used HTS and bioinformatics tools to analyze the

community composition, patterns of microbial taxa allocation in the gut environment, and the predicted metabolic functions of the bacterial microbiota in the gut ecosystem. These data were further refined using Phylogenetic Tools for Analysis of Species-level Taxa (PhyloToAST v1.4.0) [25] alongside Quantitative Insights into Microbial Ecology (QIIME v1.9.1) [26] to condense redundant taxonomic groups. This allowed increased resolution of microbial taxonomic groups to the species level and enhanced beta diversity inference. Additionally, the keystone taxa (herein “key” taxa) of the gut ecosystem were elaborated through topological analysis of Co-occurrence Network inferences (CoNet v1.1.1) [27,28,29] based on criteria described in Berry and Widder [30,31]. The results of this baseline case study demonstrate the microbial composition and associated functional capacity within the compartmentalized gut system of this evolutionarily and ecologically significant purple *S. purpuratus* sea urchin species.

MATERIALS AND METHODS

Sample Description and the High-Throughput Sequencing Data

The high-throughput amplicon sequencing data of the 16S rRNA gene (V4 hypervariable region) of the laboratory aquaculture (LAB) and naturally occurring (ENV) *Lytechinus variegatus* were obtained from previous studies conducted in our laboratory [17, 18]. Briefly, the LAB group consisted of adult sea urchins (*n* = 2) that were collected from Saint Joseph Bay Aquatic Preserve, Florida (29.80° N 85.36° W), and transported to the University of Alabama at Birmingham (UAB) where they were held for 6 months in a recirculating saltwater tank system prior

to tissue collection. The sea urchins were fed *ad libitum* once every 24-48 h with a formulated feed consisting of 6% lipid, 28% protein, and 36% carbohydrate relative percentages [26], and the aquaria conditions were maintained at 22 ± 2°C with a pH of 8.2 ± 0.2 and salinity of 32 ± 1 parts per thousand (ppt.) [17]. The ENV group was also comprised of adult sea urchins (*n* = 3), that were collected from the same location (29.80° N 85.36° W). The water conditions were recorded as 20 ± 2 °C with a pH of 7.8 ± 0.2 and salinity of 28 ± 1 ppt. and tissue collection were performed 7 ± 1 h after sea urchin collection [18]. Both investigations utilized the Illumina MiSeq high throughput sequencing platform with the 250 bp paired-end kits from Illumina targeting the V4 hypervariable region of the 16S rRNA gene [36, 37]. The resultant demultiplexed and FASTQ formatted [38] paired-end sequence data used for this study were downloaded from the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA), found under Bioproject #PRJNA291441 and #PRJNA326427 for the LAB and ENV groups, respectively. The Biosample Ids were SAMN03944319 - SAMN03944322 (LAB group) and SAMN05277845 - SAMN05277850 (ENV group). The subgroups for the LAB group were relabeled for this study as LAB.Gut.Tissue (*n* = 2) and LAB.Gut.Digesta (*n* = 2), and for the ENV group the samples were relabeled as ENV.Gut.Tissue (*n* = 3) and ENV.Gut.Digesta (*n* = 3).

Quality Assessment and Filtering

Initial sequence quality of the paired-end data was assessed using FastQC [39]. Quality reads with 80% of bases at a Q score of >33 were kept for

downstream analysis using the “fastx_trimmer” command from the FASTX Toolkit [37, 40]. Paired-ends that had <50 bp overlap and/or over 20 mismatching nucleotides were filtered from the analysis, and the remaining paired-ends were merged through USEARCH [41]. FastQC was again used to check read quality, and chimeric sequences were removed using USEARCH [41].

Taxonomic Distribution

The quality-checked and merged sequence files were analyzed using Quantitative Insights into Microbial Ecology (QIIME; v1.9.1) [42] along with Phylogenetic Tools for Analysis of Species-level Taxa (PhyloToAST; v1.4.0) [43]. Operational Taxonomic Unit (OTU) sequences were clustered at a 97% sequence similarity using UCLUST by the default parameters in QIIME (v1.9.1) [41]. Representative sequences were selected by the “most_abundant” option, and OTUs < 0.0005% were filtered. Then, the “assign_taxonomy_by_blast _result.py” command through PhyloToAST (v1.4.0) was used to assign taxonomy to the Greengenes (v13.8) database [44, 45]. Redundant OTUs were merged through the “condense_workflow.py” command [43]. Variation in the read-depth was accounted for by subsampling of the condensed OTU table to the minimum read count value using “single_rarefaction.py” in QIIME (v1.9.1). The resultant PhyloToAST (v1.4.0) condensed and minimum-count subsampled OTU table was analyzed through rarefaction plot analysis of OTU count against sequence depth and used for all downstream analyses. The taxonomic distribution was visualized through relative abundance graphs generated in Microsoft Excel

Software (Seattle, WA, USA), showing the top 100 most resolved taxonomic identities across all sample groups. The representative sequence corresponding to the highly abundant Campylobacteraceae OTU determined in this study was further analyzed by using NCBI Basic Local Alignment Search Tool (BLAST) [46] against the non-redundant nucleotide collection (nr/nt) database optimized for highly similarity sequences via MEGABLAST (https://blast.ncbi.nlm.nih.gov/).

Alpha Diversity

Alpha diversity was calculated for all samples by the “alpha_diversity.py” command in QIIME (v1.9.1), specifically to establish the Shannon [47] and Simpson [48] diversity measurements. These values were also visualized using “diversity.py” command from PhyloToAST (v1.4.0) to show the diversity value and the range of underlying data points (density) for the subgroups as kernel density estimator-smoothed histograms. For both diversity measures, a two-sample t-test was performed between the LAB and ENV gut tissue, and separately the gut digesta samples, using the “compare_alpha_diversity.py” command through QIIME (v1.9.1).

Beta Diversity

For weighted Unifrac analysis [49], the representative sequences corresponding to the rarefied OTU data were aligned through the “align_seqs.py” command in QIIME (v1.9.1) using the PyNAST algorithm [50], and used to construct a phylogenetic tree by using “make_phylogeny.py” with the default

FastTree tree building method [51]. Then, the “beta_diversity.py” command was used to calculate the weighted Unifrac distances [49]. The distance matrices were then used as the input for the “principal_coordinates.py” command, in order to determine the principal coordinate analysis (PCoA) plot points for each sample. The distance matrix and phylogenetic tree data were used to generate the weighted Unifrac 3D PCoA plots using the “make_emperor.py” in QIIME (v1.9.1) and visualized in Emperor Software [52]. The resultant weighted Unifrac tree file was also visualized as a dendrogram through the Interactive Tree of Live (iTol, v4.3.3) web service [53]. Analyses of similarity (ANOSIM) [54] and Adonis [55] calculations were performed between grouped biological replicates based on the weighted Unifrac values using the “compare_categories.py” command through QIIME (v1.9.1). To determine the effect size of key taxa in the compartmentalized gut ecosystem, the OTU data of all the gut tissue and gut digesta samples were further analyzed by the Linear Discriminant Analysis (LDA) Effect Size (LEfSe) program as implemented in the Hutlab Galaxy web application (huttenhower.sph.harvard.edu/galaxy/) using default parameters [56]. Briefly, the non-parametric Kruskal-Wallis sum-rank test was used between classes to determine significant differential abundance set at a significance of p = 0.05 [57], followed by the pairwise Wilcoxon signed-rank test between the subclasses at a significance of p = 0.05 [58]. The resultant data was used for LDA analysis using the log(10) values at an inclusion threshold of ± 3.6 [56, 59]. Those taxa with a significant effect size were also listed in a table format, to show the LDA effect size and average relative abundance in each group with standard

deviations determined through the Statistical Analysis of Metagenomic Profiles (STAMP, v2.1.3) program [60].

Predicted Functional Analysis

The predicted functional capacity of gut microbial communities was determined through Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt, v1.1.3) [61]. To accomplish this, the representative sequences were used to create an OTU table using the "pick_closed_reference_otus.py" strategy against the Greengenes (v13.8) database [44, 45, 61]. The OTU data was then normalized by copy number as suggested in PICRUSt (v1.1.3). Using the "predict_metagenomes.py" command with the Kyoto Encyclopedia of Genes and Genomes (KEGG) database, the associated KEGG Ortholog (KO) Ids were determined along with the weighted Nearest Sequenced Taxon Index (NSTI) values [61-63]. The functional categories were collapsed by the "categorize_by_function.py" command into KEGG-Level-2 and KEGG-Level-3 categories and uploaded into STAMP (v2.1.3) for group analysis [60]. First, samples were grouped by biological replicate as LAB.Gut.Tissue (n = 2), ENV.Gut.Tissue (n = 3), LAB.Gut.Digesta (n = 2), and ENV.Gut.Digesta (n = 3). The KEGG-Level-2 categories corresponding to the metabolism of primary macromolecules such as carbohydrate, amino acid, lipid, and energy metabolisms were selected for boxplot analysis. Significant differences were determined by an analysis of variance (ANOVA) with a Tukey-Kramer post-hoc test set at 0.95, and p-values were corrected by a Benjamini-

Hochberg False Discovery Rate (FDR) set at a q-value of 0.05 for significance [64, 65].

RESULTS

Read Quality and Sample Statistics

The raw sequence count generated from paired-end Illumina MiSeq 16S rRNA gene HTS across all samples of the study totaled 1,214,823 reads, which yielded 1,058,526 following quality checking (Table 4-1). A total of 432 OTUs were identified following quality filtering and the PhyloToAST (v1.4.0) workflow. The rarefaction plot analysis of OTUs at the rarefied sequence depth showed a plateau, indicating a sufficient sequence saturation for downstream microbial community analysis (data not shown).

Taxonomic Distribution across Samples

Overall, the gut tissue of both the LAB and ENV *L. variegatus* groups was consistently dominated by Proteobacteria, showing an almost exclusive abundance of Epsilonproteobacteria (Fig 4-1). At the highest achievable taxonomic resolution through the described bioinformatics methods, these taxa were identified as Campylobacteraceae, which comprised >88% of the relative abundance in all gut tissues. Further analysis by using the NCBI BLAST alignment of the representative sequence provided additional resolution to this

Table 4-1: Sample statistics determined for the sequence reads, OTU table, and alpha diversity values of each *Lytechinus variegatus* sample of the study. The Shannon and Simpson alpha diversity values for each sample are also shown. The group assignments are indicated as ENV = naturally occurring *L. variegatus*, and LAB = laboratory aquaculture *L. variegatus*.

Sample	Raw Sequence Reads	Trimmed Sequence Reads	Condensed and Rarefied OTU Count	Shannon Diversity	Simpson Diversity
ENV.Gut.Digesta.1	132,132	104,704	300	4.90	0.93
ENV.Gut.Digesta.2	143,149	117,939	265	4.17	0.90
ENV.Gut.Digesta.3	113,415	92,147	321	5.24	0.95
ENV.Gut.Tissue.1	89,881	74,366	90	0.75	0.21
ENV.Gut.Tissue.2	125,274	108,093	93	0.44	0.11
ENV.Gut.Tissue.3	95,653	81,873	68	0.55	0.15
LAB.Gut.Digesta.1	120,424	108,902	217	3.08	0.73
LAB.Gut.Digesta.2	176,771	161,991	193	2.81	0.76
LAB.Gut.Tissue.1	90,693	83,722	100	0.30	0.06
LAB.Gut.Tissue.2	127,431	112,702	149	0.69	0.13
Summary	Total = 1,214,823	Total = 1,058,526	Total = 432	Avg. = 2.29	Avg. = 0.49

taxon. From the top 100 assigned identities, 34% were determined to be related to Uncultured *Arcobacter* sp., 8% to *Arcobacter* sp., 7% as *Arcobacter bivalviorum*, and 3% to *Sulfuricurvum* sp., all with an E-value < 5E-83 and percent identity > 89.76% (Supplementary Table 4-1). The ENV gut tissue showed a noticeable abundance of phylum Tennericutes, identified as *Candidatus_Hepatoplasma* (~2.5 - 7.5%), that were negligibly abundant in the LAB group (<1%).

The gut digesta of both the LAB and ENV groups showed taxa assigned to Gammaproteobacteria to be the most abundant. From this class, *Vibrio* was found to be more dominant in the LAB digesta (~40 - 50%) as compared to the ENV digesta group (~10 - 22%). However, the LAB gut digesta showed a unique abundance of taxa that were not noticeable in the ENV digesta, which included *Agarivorans* (~2 - 24%) and *Shewanella algae* (~2 - 3%) from Gammaproteobacteria, Rhodobacteraceae from Alphaproteobacteria (~3 - 7%), and Epsilonproteobacteria that were resolved to *Arcobacter* through the described methods (~11 - 40%). In contrast, the ENV digesta showed Gammaproteobacteria such as *Photobacterium* (~9 - 11%), as well as *Propionigenium* of Fusobacteria (~9 - 12%), and a noticeable abundance of Flavobacteriales (~17 - 19%)

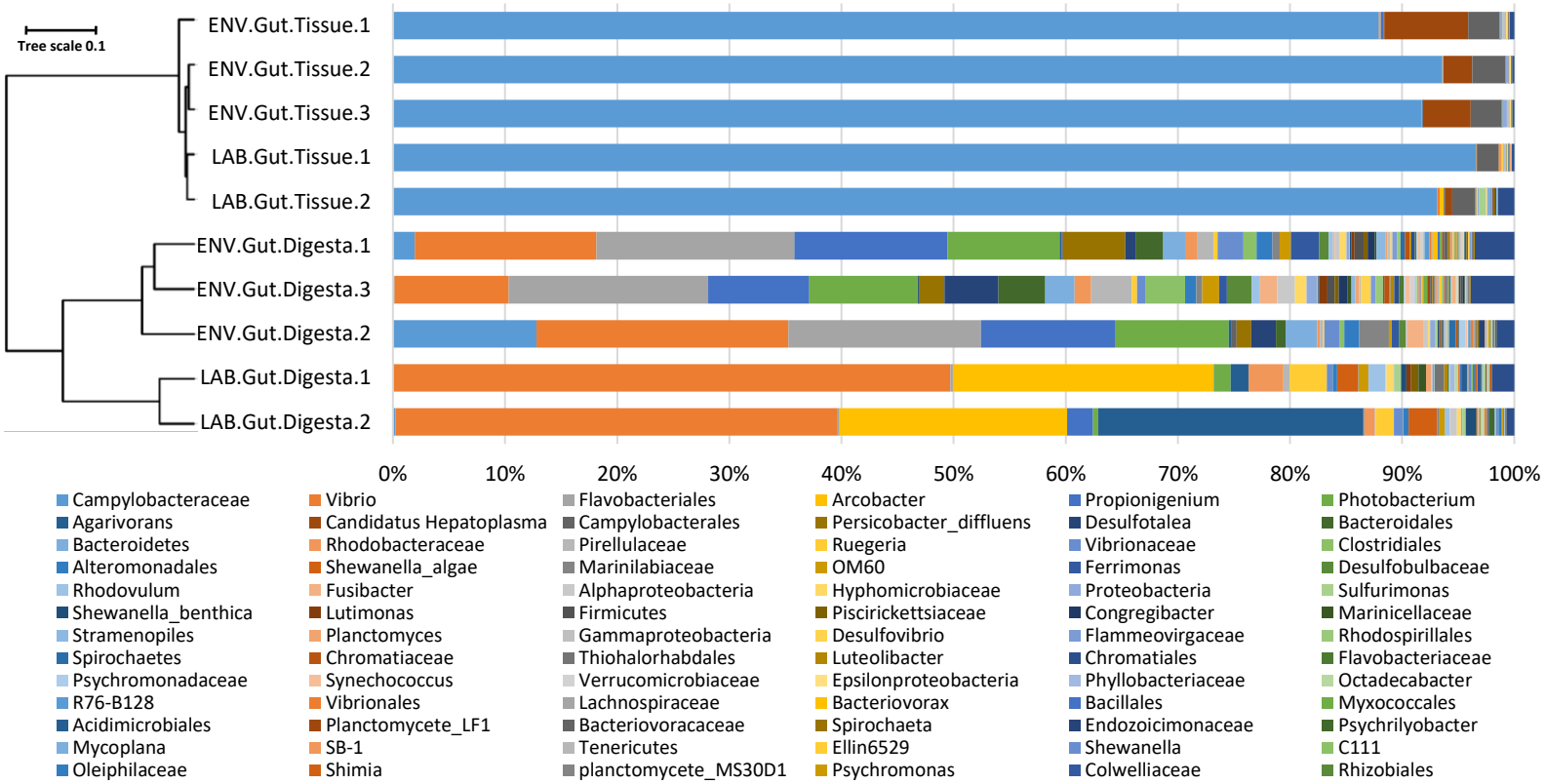


Figure 4-1**:** Relative abundance bar graphs showing the top 100 taxa at the most resolved level in the gut ecosystem of the green sea urchin *Lytechinus variegatus* gut ecosystem. Taxa not included in the top 100 were listed as "Other." Taxonomic identities were based on their assignment through the Greengenes (v13.8) database as determined by the Quantitative Insights into Microbial Ecology (QIIME, v1.9.1) and Phylogenetic Tools for Analysis of Species-level Taxa (PhyloToAST, v1.4.0) analyses. The dendrogram was created based on the weighted Unifrac metric rendered through the Interactive Tree of Live (iTol, v4.3.3), and shows a grouping of sample types. The group assignments are indicated as ENV = naturally occurring *L. variegatus*, and LAB = laboratory aquaculture *L. variegatus*. The relative abundance plot was created using Microsoft Excel Software (Seattle, WA, USA).

Alpha Diversity

For both the Shannon and Simpson alpha diversity statistics, the LAB gut tissue and digesta showed less taxonomic diversity as compared to the ENV group (Table 4-1; Fig 4-2). Overall, the LAB and ENV gut tissue had the least number of OTUs compared to the gut digesta samples. The ENV gut digesta had the highest diversity and OTU count in the study, followed by the LAB gut digesta that showed a comparatively moderate alpha diversity and OTU count. A paired t-test comparison between the alpha diversity values of the gut tissues from the LAB and ENV groups showed no significant ($p > 0.05$) differences using the Shannon ($p = 0.58$) and Simpson ($p = 0.227$) metrics (Fig 4-2). However, a comparison between the LAB and ENV digesta showed a significant difference of the Shannon ($p = 0.03$) and Simpson ($p = 0.003$) values between the two groups. The plotted kernel-density smoothed histograms showed the density (variation between the diversity of samples in the group), with the highest peak, observed in the LAB gut digesta for both diversity metrics.

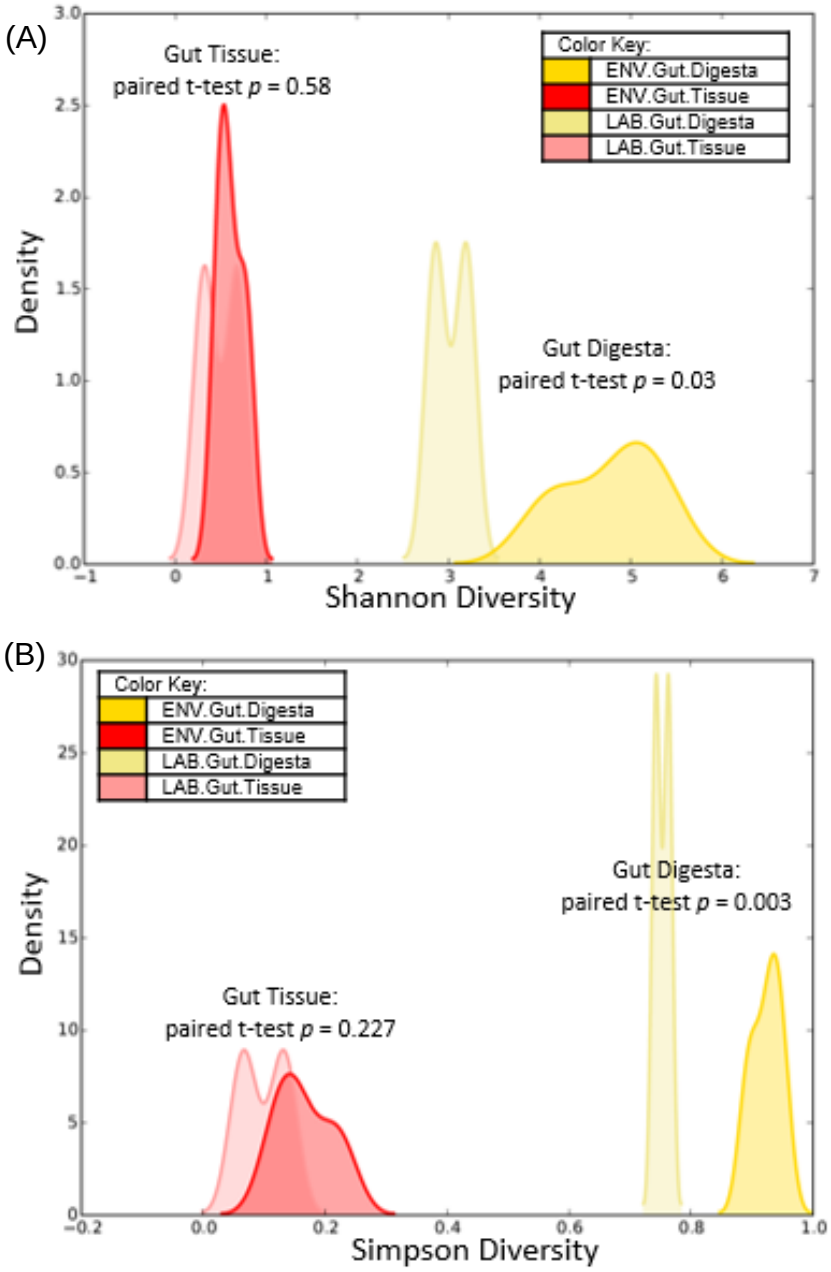


Figure 4-2: Histograms generated based on the (A) Shannon and (B) Simpson alpha diversity values calculated across each *Lytechinus variegatus* sample. Biological replicates were grouped, and paired t-tests were performed between groups for both diversity metrics (indicated in the figure). All histogram plot-points were rendered with smoothed kernel density estimation using the "diversity.py" of PhyloToAST (v1.4.0). The X-axis corresponds to the diversity value and the Y-axis depicts the density function. The group assignments are indicated as ENV = naturally occurring *L. variegatus*, and LAB = laboratory aquaculture *L. variegatus*. Sample groups are color coded as follows: ENV.Gut.Digesta (*n* = 3, brown); LAB.Gut.Digesta (*n* = 2, light brown); ENV.Gut.Tissue (*n* = 3, red); LAB.Gut.Tissue (*n* = 2, light red).

Beta Diversity

The 3D PCoA plots using the weighted Unifrac distances across all samples showed the gut tissue from both the LAB and ENV groups to cluster strongly together (Fig 4-3). For the gut digesta, distinct sub-clustering according to group assignment was observed. These cluster patterns were also elaborated in the dendrogram (Fig 4-1). ANOSIM and Adonis also supported the low within-group variation shown by the cluster patterns, revealing an R and R^2 value of 0.777 and 0.969, respectively (p = 0.001), and indicating significant grouping based on biological replicates.

LEfSe analysis performed between the collective gut tissue and gut digesta showed those taxa contributing most to the effect size (Fig 4-4). For the gut tissue samples, the highest effect size was attributed to the abundant Campylobacteraceae taxon (LDA score = 5.65). This was followed by *Candidatus_Hepatoplasma* and Campylobacterales. For the gut digesta, *Vibrio* showed the highest effect size (LDA score = 5.11). This taxon was more abundant in the LAB digesta (44.66 ± 5.04) as compared to the ENV digesta (16.21 ± 4.99). This was followed by Flavobacteriales, *Propionigenium*, and *Photobacterium*, which were noticeably abundant in the ENV digesta, as well as *Agarivorans* and Rhodobacteraceae that were more heightened in the LAB digesta. Few taxa that were represented at low abundances were indicated by the LEfSe analysis, specifically Pirellulaceae, Alteromonadales, Marinilabiaceae, and OM60 (phylum Gammaproteobacteria; order Alteromonadales), all of which were negligently abundant in the gut tissue samples.

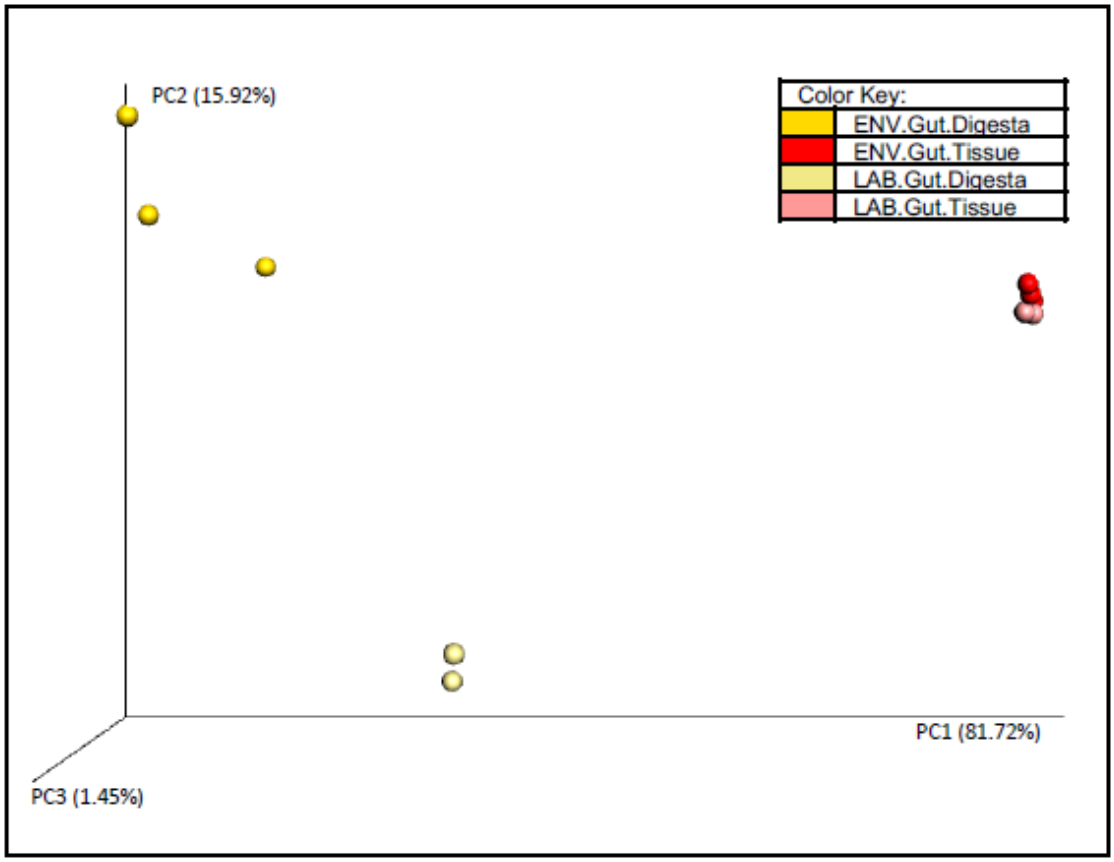
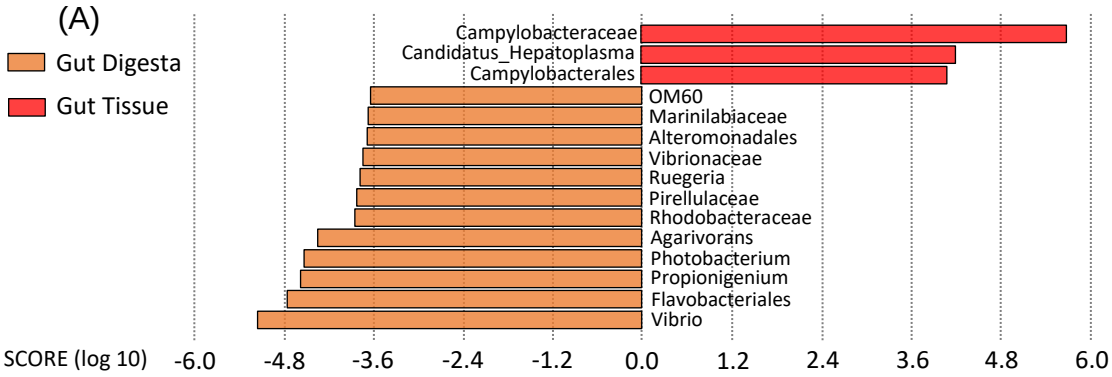


Figure 4-3: 3D Principle Coordinate Analysis (PCoA) plot based on the weighted Unifrac distances. The plot was generated using QIIME (v1.9.1), and the percent variation is shown on each axis. Analysis of similarity (ANOSIM) (R = 0.777) and Adonis (R^2 = 0.969) was performed across all four groups, and the results indicated significant clustering based on group assignment (p = 0.001). The group assignments are indicated as ENV = naturally occurring *L. variegatus*, and LAB = laboratory aquaculture *L. variegatus*. Sample groups are color coded as follows: ENV.Gut.Digesta (n = 3, brown); LAB.Gut.Digesta (n = 2, light brown); ENV.Gut.Tissue (n = 3, red); LAB.Gut.Tissue (n = 2, light red).



Phyla	Taxa	Class	Effect Size	p-value	ENV Gut Tissue	LAB Gut Tissue	ENV Gut Digesta	LAB Gut Digesta
Epsilonproteobacteria	Campylobacteraceae	Gut Tissue	5.65	0.01	91.63 ±2.44	95.14 ±1.81	5.05 ±5.69	0.11 ±0.10
Tenericutes	*Candidatus Hepatoplasma*	Gut Tissue	4.17	0.01	4.77 ±2.07	0.35 ±0.31	0.01 ±0.01	0.00 ±0.00
Epsilonproteobacteria	Campylobacterales	Gut Tissue	3.99	0.01	2.30 ±0.06	1.62 ±0.06	0.15 ±0.17	0.02 ±0.01
Gammaproteobacteria	OM60	Gut Digesta	3.62	0.01	0.00 ±0.00	0.00 ±0.00	0.93 ±0.53	0.68 ±0.21
Bacteroidetes	Marinilabiaceae	Gut Digesta	3.65	0.01	0.00 ±0.00	0.01 ±0.00	1.32 ±1.01	0.15 ±0.07
Gammaproteobacteria	Alteromonadales	Gut Digesta	3.69	0.01	0.00 ±0.00	0.00 ± 0.00	1.23 ±0.16	0.51 ±0.00
Planctomycetes	Pirellulaceae	Gut Digesta	3.82	0.01	0.00 ±0.00	0.04 ±0.04	1.78 ±1.40	0.30 ±0.22
Alphaproteobacteria	Rhodobacteraceae	Gut Digesta	4.08	0.01	0.01 ±0.01	0.17 ±0.17	1.27 ±0.68	4.82 ±2.16
Gammaproteobacteria	*Agarivorans*	Gut Digesta	4.43	0.01	0.00 ±0.00	0.02 ±0.02	0.18 ±0.05	12.91 ±11.33
Gammaproteobacteria	*Photobacterium*	Gut Digesta	4.53	0.01	0.02 ±0.02	0.00 ±0.00	10.19 ±0.47	0.96 ±0.48
Fusobacteria	*Propionigenium*	Gut Digesta	4.58	0.03	0.21 ±0.19	0.01 ±0.01	23.01 ±3.97	2.38 ±2.33
Bacteroidetes	Flavobacteriales	Gut Digesta	4.75	0.01	0.03 ±0.03	0.00 ±0.00	11.34 ±0.58	0.00 ±0.00
Gammaproteobacteria	*Vibrio*	Gut Digesta	5.11	0.01	0.04 ±0.03	0.13 ±0.12	16.21 ±4.99	44.66 ±5.04

Figure 4-4: Linear Discriminant Analysis (LDA) Effect Size (LEfSe) analysis performed on the taxonomic data of the *Lytechinus variegatus* samples at the highest resolution. (A) The effect size was visualized as a bar graph of two classes, one class representing the Gut.Tissue samples (*n* = 5; green bars) that comprised subclass LAB.Gut.Tissue (*n* = 2) and subclass ENV.Gut.Tissue (*n* = 3), and the other representing the Gut.Digesta samples (*n* = 5; red bars) that comprised the LAB.Gut.Digesta (*n* = 2) and ENV.Gut.Digesta (*n* = 3). The values shown on the X-axis correspond to the log(10) effect size values at an inclusion threshold of ± 3.6. (B) The table inset lists the taxa that had a significant effect size. The phylum information (class for Proteobacteria) for each taxon is given, including the class that the effect size favored and the *p*-value of significance. For each taxon, the average abundance and standard deviation determined through Statistical Analysis of Metagenomic Profiles (STAMP, v2.1.3) is provided for each of the four subclasses. LEfSe analysis was conducted and plotted using the Huttenhower LEfSe Galaxy web application (huttenhower.sph.harvard.edu/galaxy/)

Predicted Functional Analysis

The NSTI values calculated through PICRUST (v1.1.3) showed an average value of 0.14 (ranging from 0.08-0.20). Overall, the trends of KEGG-Level-2 categories were consistent among sample replicates, irrespective of habitat (Fig 4-5). The LAB and ENV gut tissues showed heightened energy metabolism when compared to the gut digesta. In contrast, the gut digesta of both groups showed a heightened abundance of amino acid, carbohydrate, and lipid metabolisms as compared to the gut tissues. Moreover, these metabolic categories were noticeably more enriched in the ENV digesta as compared to the LAB digesta.

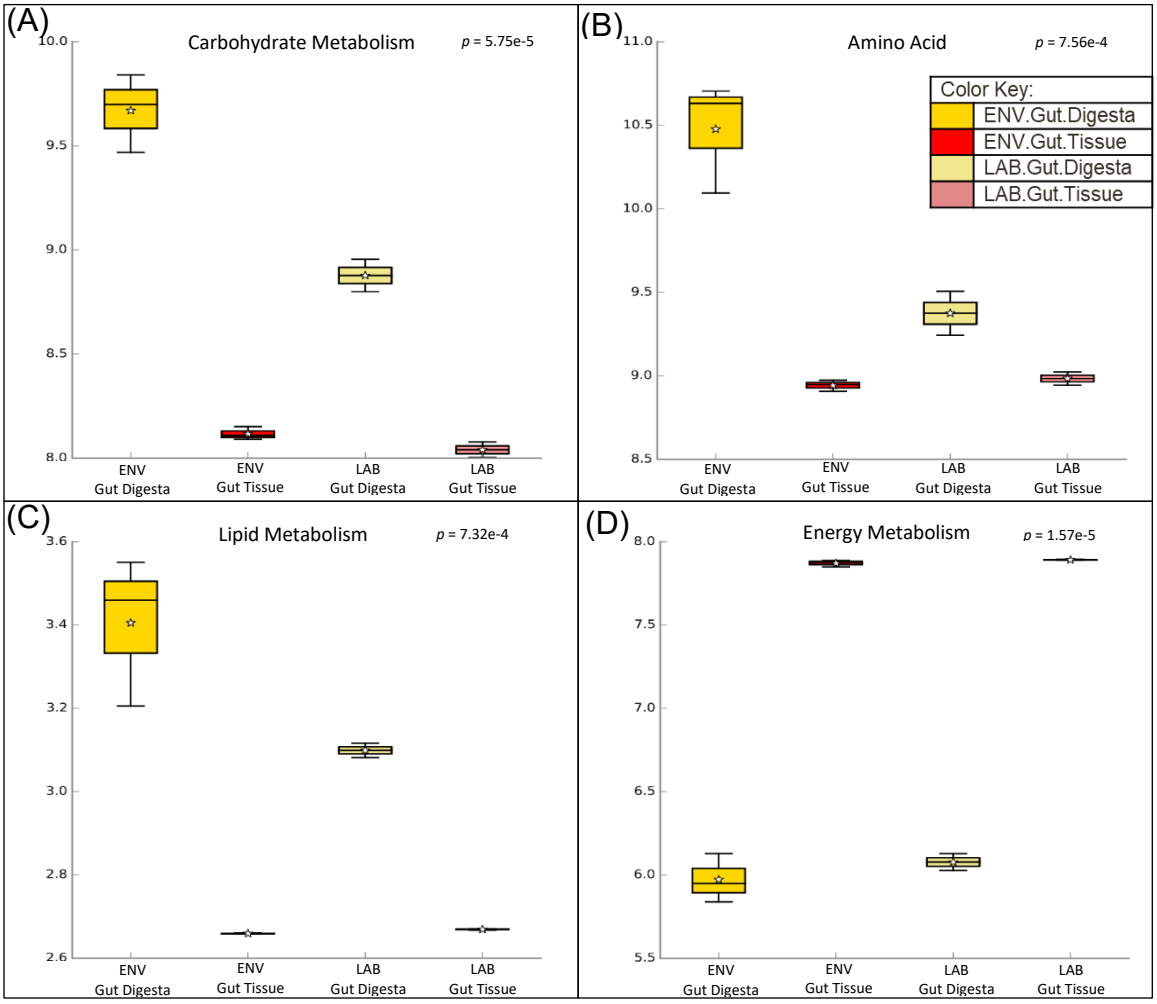


Figure 4-5: Boxplot analysis of the predicted functional profiles corresponding to the microbial communities of the *Lytechinus variegatus* gut ecosystem. The KEGG-Level-2 hierarchical functional categories were created through the "categorize_by_function.py" command using the KEGG Orthology (KO) gene data determined by Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUST v1.1.3). Boxplots for (A) carbohydrate, (B) amino acid, (C) lipid, and (D) energy KEGG-Level-2 categories were generated using STAMP (v2.1.3). Multiple group analysis was performed by Analysis of Variance (ANOVA) with a Tukey–Kramer post hoc set at 0.95 and a Benjamini–Hochberg False Discovery Rate (FDR) correction q-value set at 0.05 for significance. The group assignments are indicated as ENV = naturally occurring *L. variegatus*, and LAB = laboratory aquaculture *L. variegatus*. Sample groups are color coded as follows: ENV.Gut.Digesta (*n* = 3, brown); LAB.Gut.Digesta (*n* = 2, light brown); ENV.Gut.Tissue (*n* = 3, red); LAB.Gut.Tissue (*n* = 2, light red). The black line through each boxplot represents the median value, and the star represents the mean. The adjusted *p*-values for each metabolic category are listed in each boxplot.

DISCUSSION

A comparative outlook of the microbiota in the compartmentalized gut ecosystem of *L. variegatus* using the rarified HTS data revealed that *Arcobacter* spp. belonging to Epsilonproteobacteria were the dominant taxon in the gut tissues of both LAB and ENV samples (Fig 4-1; Supplementary Table 4-1). Besides *Arcobacter* found in *L. variegatus* [17, 18, 66], other members of the Epsilonproteobacteria are commonly associated with other marine Echinoderms, such as unculturable *Helicobacter* in sea stars *Patiria pectinifera* and *Asterias amurensis* [67], and *Sulfurospirillum* and *Sulfuricurvum* in the sea cucumber [68]. In contrast, a comparison between the LAB and ENV gut digesta showed noticeable differences in microbial community composition. Overall, both LAB and ENV gut digesta showed *Vibrio* to be dominant, which is consistent with previously reported sea urchins *Strongylocentrotus droebachiensis*, *Tripneustes ventricosus* [23], *Strongylocentrotus intermedius*, *Strongylocentrotus nudus* [69], *Echinus esculentus* [70], *Lytechinus variegatus* [66], *Paracentrotus lividus* [71], *Hemicentrotus pulcherrimus* [72], and *Strongylocentrotus purpuratus* [73] gut microbiota studies. However, the higher relative abundance of *Vibrio* in the LAB group could be due to the relatively higher salt concentration (32 ± 1 ppt.), pH (~8.4), and temperature (22 ± 2°C) used in the laboratory aquaculture as compared to the conditions of their natural habitat (salinity = 28 ± 1 ppt.; pH = 7.8 ± 0.2; temperature = 20 ± 2°C) [74]. Other differences included a prevalence of *Photobacterium* in the ENV group, with certain species suggested to perform lipid metabolisms [75-77], as reported in another sea urchin species, *P. lividus* [71].

Additionally, the prevalence of the strictly anaerobic genus *Propionigenium* of phylum Fusobacteria in the ENV groups may be due to a comparatively higher fraction of non-digestible carbohydrates in seagrass compared to the laboratory-formulated feed. *Propionigenium* spp. generate the fatty acid propionate from succinate [78-80], and may benefit the sea urchin through such health-related effects as mitigation of inflammation and anti-carcinogenic activity as described in humans [79]. Additionally, previously reported taxonomic co-occurrence network (CoNet) analysis [81] has attributed keystone status to *Propionigenium* in the gut digesta of sea urchins *S. purpuratus*, suggesting their dominant influence in structuring the gut microbial communities [73, 82].

Although the gut tissues in both groups had a comparable taxonomic diversity, the differences in the alpha diversity observed between the LAB and ENV gut digesta could be due to the diverse bacterial taxa in the nearshore Gulf of Mexico marine habitat (Fig 4-2). Additionally, fluctuations of the abiotic factors such as pH, temperature, photoperiod, and salinity could also promote differences in the microbial diversity on a temporal scale [83-87]. Moreover, it has been reported that diet plays a significant role in the gut microbial composition in a wide range of organisms [88, 89]. Although *L. variegatus* generally grazes seagrass in their habitat, alternate food sources such as detritus materials, various algae and their epibionts [27, 28]. In contrast, the defined laboratory aquaculture conditions could have contributed to the low alpha diversity in the gut digesta of the LAB group. Although a similar reduction of the alpha diversity has been reported [8, 90, 91], this is not a universal trend in laboratory-maintained

organisms as observed in the zebrafish [5] and baboon [92]. Thus, in our study, whether the replacement of the diet from their natural habitat to the defined nutritionally balanced feed used *ad libitum* for the LAB group restructured the gut microbial community would need further investigation.

The ANOSIM and Adonis statistics (p = 0.001) supported distinct cluster patterns of the four groups comprising the gut tissue and gut digesta of the LAB (n = 2) and ENV (n = 3) sea urchins biological replicates used in this study. The microbial communities of the gut tissue and gut digesta showed unique cluster patterns of biological replicates as determined by beta diversity analysis (Fig 4-3). Notably, the gut tissues of both the LAB and ENV groups clustered strongly together, indicating a gut tissue-specific microbiota likely maintained following the transition to the laboratory aquaculture environment. Although the LAB and ENV gut digesta groups showed common higher taxonomic classifications, such as Gammaproteobacteria, there were differences at the lower taxonomic level. Such results suggest that habitat may influence the microbial community composition in the gut digesta environment. The LEfSe analysis between the gut tissue and gut digesta across both groups predicted Campylobacteraceae and *Vibrio* to contribute to the uniquely compartmentalized microbial ecology in the gut ecosystem (Fig 4-4).

Predicted functional analysis of the compartmentalized gut microbial communities through PICRUSt (v1.1.3) indicated carbohydrate, amino acid, and lipid metabolisms as the dominant KEGG-Level-2 categories in the gut digesta as compared to the gut tissue in both LAB and ENV groups (Fig 4-5). Such results

indicate that the gut digesta is the primary location for the microbial-driven metabolism of environmental and laboratory-prepared dietary macromolecules [93-95]. Moreover, these metabolic categories were enriched in the ENV gut digesta, suggesting a higher metabolic capacity. Conversely, the gut tissues of both groups showed energy metabolism to be significantly heightened as compared to the gut digesta. This category includes nitrogen and sulfur metabolisms, which have been attributed to the microbial communities of other sea urchins [22, 23, 71, 96, 97]. Additionally, *Arcobacter* spp. have been described as chemolithoautotrophic bacterium [98] performing crucial biochemical processes in the marine environment such as sulfur oxidation in hydrothermal vents [99] and nitrogen metabolisms [100]. However, whether these metabolisms are of any benefit to their host's health and nutrition, including the specific metabolic input of the dominant Epsilonproteobacteria of the gut tissue remains to be clarified.

In conclusion, *L. variegatus* maintained a distinct microbial community representing primarily *Arcobacter spp.* in the gut tissues. Predictive functional roles indicated that this taxon is involved in the energy metabolisms irrespective of the laboratory aquaculture conditions (LAB) or natural habitat (ENV). While a comparison between the gut digesta of the LAB and ENV groups showed distinct taxa at the most resolved level, consistencies were observed at the phylum or the class level. Similar observations have been reported in a number of other "domesticated" animals, including hydra [101], fruit flies [102], birds [35, 103, 104], zebrafish [5], and mice [105]. Additionally, the metabolisms of

macronutrients in the gut digesta were consistently higher than the gut tissues in both LAB and ENV groups. These results indicate that the gut digesta are potentially the primary location in the compartmentalized gut ecosystem of *L. variegatus* where the maximum microbial energy metabolisms occur. Although our results provide an insight into the structure and predictive metabolic functions of microbial communities in the laboratory aquaculture *L. variegatus*, future studies would elaborate the significance of the close association of Epsilonproteobacteria with the gut tissues. Future studies may give insight into the potential role of habitat-specific or laboratory formulated diet in restructuring the microbiota in the gut digesta following ingestion and mucous-encapsulation. Moreover, the role of the high energy gut digesta contributing to the host's nutrition and ecological impact at various trophic levels could also be of interest in future studies.

FUNDING

The following are acknowledged for their support of the Microbiome Resource at the University of Alabama at Birmingham: School of Medicine, Comprehensive Cancer Center (P30AR050948), Center for AIDS Research (5P30AI027767), Center for Clinical Translational Science (UL1TR000165) and Heflin Center. Animal husbandry supported in part by NIH P30DK056336.

ACKNOWLEDGEMENTS

We would like to thank Dr. Peter Eipers of the Department of Cell, Developmental and Integrative Biology, and Dr. Michael Crowley of the Heflin Center for Genomics Sciences at the University of Alabama at Birmingham (UAB), for their assistance in high-throughput sequencing for this study. We would also like to thank the Biology Department at UAB for logistics, and acknowledge the UAB School of Medicine grant support for C.D.M to fund the graduate research assistantship position of J.A.H.

CONFLICT OF INTERESTS

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

REFERENCES

1. Elmqvist T, Folke C, Nyström M, Peterson G, Bengtsson J, Walker B, et al. Response diversity, ecosystem change, and resilience. Front Ecol Env. 2003;1(9):488-94.

2. Levin SA, Lubchenco J. Resilience, robustness, and marine ecosystem-based management. Bioscience. 2008;58(1):27-32.

3. Palumbi SR, McLeod KL, Grünbaum D. Ecosystems in action: lessons from marine ecology about recovery, resistance, and reversibility. BioScience. 2008;58(1):33-42.

4. Rosshart SP, Vassallo BG, Angeletti D, Hutchinson DS, Morgan AP, Takeda K, et al. Wild mouse gut microbiota promotes host fitness and improves disease resistance. Cell. 2017;171(5):1015-28. e13.

5. Roeselers G, Mittge EK, Stephens WZ, Parichy DM, Cavanaugh CM, Guillemin K, et al. Evidence for a core gut microbiota in the zebrafish. ISME J. 2011;5(10):1595.

6. Dirksen P, Marsh SA, Braker I, Heitland N, Wagner S, Nakad R, et al. The native microbiome of the nematode *Caenorhabditis elegans*: gateway to a new host-microbiome model. BMC Biology. 2016;14(1):38.

7. Zhang F, Berg M, Dierking K, Félix M-A, Shapira M, Samuel BS, et al. *Caenorhabditis elegans* as a model for microbiome research. Frontiers in microbiology. 2017;8:485.

8. Staubach F, Baines JF, Künzel S, Bik EM, Petrov DA. Host species and environmental effects on bacterial communities associated with *Drosophila* in the laboratory and in the natural environment. PloS one. 2013;8(8):e70749.

9. Di Bernardo M, Di Carlo M. The Sea Urchin Embryo: A Model for Studying Molecular Mechanisms Involved in Human Diseases and for Testing Bioactive Compounds. Sea Urchin - From Environment to Aquaculture and Biomedicine: IntechOpen; 2017.

10. McBride SC, editor Sea Urchin Aquaculture. American Fisheries Society Symposium; 2005; Eureka, California, USA: Publication of the University of California Sea Grant Extension Program.

11. Heflin LE, Makowsky R, Taylor JC, Williams MB, Lawrence AL, Watts SA. Production and economic optimization of dietary protein and carbohydrate in the culture of Juvenile Sea Urchin *Lytechinus variegatus*. Aquaculture. 2016;463:51-60.

12. Albright R, Bland C, Gillette P, Serafy JE, Langdon C, Capo TR. Juvenile growth of the tropical sea urchin *Lytechinus variegatus* exposed to near-future ocean acidification scenarios. *J Exp Mar Biol Ecol*. 2012;426:12-7.

13. Hendler G. Sea stars, sea urchins, and allies: echinoderms of Florida and the Caribbean1995.

14. Watanabe JM, Harrold C. Destructive grazing by sea urchins *Strongylocentrotus* spp. in a central California kelp forest: potential roles of recruitment, depth, and predation. Marine Ecology Progress Series. 1991:125-41.

15. Watts SAM, James B.; Lawrence, John M. . Chapter 31 - Lytechinus. Lawrence JM, editor. UK: Elsevier; 2013.

16. De Ridder C, Jangoux M, Lawrence J. Digestive System: Echinoidea. Echinoderm Nutrition. 1982:213-34.

17. Hakim JA, Koo H, Dennis LN, Kumar R, Ptacek T, Morrow CD, et al. An abundance of Epsilonproteobacteria revealed in the gut microbiome of the laboratory cultured sea urchin, *Lytechinus variegatus*. *Front Microbiol*. 2015;6.

18. Hakim JA, Koo H, Kumar R, Lefkowitz EJ, Morrow CD, Powell ML, et al. The gut microbiome of the sea urchin, *Lytechinus variegatus*, from its natural habitat demonstrates selective attributes of microbial taxa and predictive metabolic profiles. *FEMS Microbiol Ecol*. 2016;92(9):fiw146.

19. Holland ND, Ghiselin MT. A comparative study of gut mucous cells in thirty-seven species of the class Echinoidea (Echinodermata). Biol Bull. 1970;138(3):286-305.

20. Lasker R, Giese AC. Nutrition of the sea urchin, *Strongylocentrotus purpuratus*. Biol Bull. 1954;106(3):328-40.

21. Beleneva I, Kukhlevskii A. Characterization of *Vibrio gigantis* and *Vibrio pomeroyi* isolated from invertebrates of Peter the Great Bay, Sea of Japan. Microbiology. 2010;79(3):402-7.

22. Guerinot ML, Patriquin D. The association of N2-fixing bacteria with sea urchins. *Mar Biol*. 1981;62(2):197-207.

23. Guerinot M, Patriquin D. N2-fixing vibrios isolated from the gastrointestinal tract of sea urchins. *Can J Microbiol*. 1981;27(3):311-7.

24. Fong W, Mann K. Role of gut flora in the transfer of amino acids through a marine food chain. *Can J Fish Aquat Sci*. 1980;37(1):88-96.

25. Brothers CJ, Van Der Pol WJ, Morrow CD, Hakim JA, Koo H, McClintock JB. Ocean warming alters predicted microbiome functionality in a common sea urchin. Proc R Soc Lond B Biol Sci. 2018;285(1881):20180340.

26. Hammer H, Hammer B, Watts S, Lawrence A, Lawrence J. The effect of dietary protein and carbohydrate concentration on the biochemical composition and gametogenic condition of the sea urchin *Lytechinus variegatus*. J Exp Mar Bio Ecol. 2006;334(1):109-21.

27. Beddingfield SD, McClintock JB. Food Resource Utilization in the Sea Urchin *Lytechinus variegatus* in Contrasting Shallow-Water Microhabits of Saint Joseph Bay, Florida. Gulf of Mexico Science. 1999;17(1):3.

28. Beddingfield SD, McClintock JB. Demographic characteristics of *Lytechinus variegatus* (Echinoidea: Echinodermata) from three habitats in a North Florida Bay, Gulf of Mexico. Mar Ecol. 2000;21(1):17-40. doi: 10.1111/brv.12004.

29. Deming JW, Carpenter SD. Factors influencing benthic bacterial abundance, biomass, and activity on the northern continental margin and deep basin of the Gulf of Mexico. Deep Sea Research Part II: Topical Studies in Oceanography. 2008;55(24-26):2597-606.

30. Erwin PM, Olson JB, Thacker RW. Phylogenetic diversity, host-specificity and community profiling of sponge-associated bacteria in the northern Gulf of Mexico. PloS one. 2011;6(11):e26806.

31. Felder DL, Camp DK. Gulf of Mexico Origin, Waters, and Biota: Biodiversity. College Station, Texas: Texas A&M University Press; 2009.

32. Kellogg CA, Lisle JT, Galkiewicz JP. Culture-independent characterization of bacterial communities associated with the cold-water coral *Lophelia pertusa* in the northeastern Gulf of Mexico. Appl Environ Microbiol. 2009;75(8):2294-303.

33. Koo H, Mojib N, Thacker RW, Bej AK. Comparative analysis of bacterial community-metagenomics in coastal Gulf of Mexico sediment microcosms following exposure to Macondo oil (MC252). Antonie Van Leeuwenhoek. 2014;106(5):993-1009.

34. Skoog A, Biddanda B, Benner R. Bacterial utilization of dissolved glucose in the upper water column of the Gulf of Mexico. Limnol Oceanogr. 1999;44(7):1625-33.

35. Hird SM. Evolutionary biology needs wild microbiomes. Frontiers in microbiology. 2017;8:725.

36. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Appl Environ Microbiol. 2013;79(17):5112-20.

37. Kumar R, Eipers P, Little RB, Crowley M, Crossman DK, Lefkowitz EJ, et al. Getting started with microbiome analysis: sample acquisition to bioinformatics. Curr Protoc Hum Genet. 2014:18.8. 1-.8. 29.

38. Cock PJ, Fields CJ, Goto N, Heuer ML, Rice PM. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. Nucleic Acids Res. 2009;38(6):1767-71.

39. Andrews S. FastQC: a quality control tool for high throughput sequence data 2010 [cited 2019]. Available from: https://www.bioinformatics.babraham.ac.uk/projects/fastqc.

40. Gordon A, Hannon G. Fastx-toolkit. FASTQ/A short-reads pre-processing tools. 2010 [cited 2019]. Available from: http://hannonlab.cshl.edu/fastx_toolkit.

41. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 2010;26(19):2460-1.

42. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods. 2010;7(5):335.

43. Dabdoub SM, Fellows ML, Paropkari AD, Mason MR, Huja SS, Tsigarida AA, et al. PhyloToAST: Bioinformatics tools for species-level analysis and visualization of complex microbial datasets. Sci Rep. 2016;6:29123.

44. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol. 2006;72(7):5069-72.

45. McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, et al. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. The ISME journal. 2012;6(3):610.

46. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403-10.

47. Shannon CE. A mathematical theory of communication. Bell System Technical Journal. 1948;27(3):379-423.

48. Simpson EH. Measurement of diversity. Nature. 1949;163(4148):688.

49. Lozupone C, Lladser ME, Knights D, Stombaugh J, Knight R. UniFrac: an effective distance metric for microbial community comparison. The ISME journal. 2011;5(2):169.

50. Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R. PyNAST: a flexible tool for aligning sequences to a template alignment. Bioinformatics. 2009;26(2):266-7.

51. Price MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol. 2009;26(7):1641-50.

52. Vázquez-Baeza Y, Pirrung M, Gonzalez A, Knight R. EMPeror: a tool for visualizing high-throughput microbial community data. Gigascience. 2013;2(1):16.

53. Letunic I, Bork P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics. 2006;23(1):127-8.

54. Clarke KR. Non-parametric multivariate analyses of changes in community structure. Austral Ecol. 1993;18(1):117-43.

55. Oksanen J, Kindt R, Legendre P, O'Hara B, Stevens MHH, Oksanen MJ, et al. The vegan package. Community Ecology Package. 2007;10:631-7.

56. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, et al. Metagenomic biomarker discovery and explanation. Genome Biol. 2011;12(6):R60.

57. Kruskal WH, Wallis WA. Use of ranks in one-criterion variance analysis. J Am Stat Assoc. 1952;47(260):583-621.

58. Wilcoxon F. Individual comparisons by ranking methods. Biometr Bull. 1945;1(6):80-3.

59. Fisher RA. The use of multiple measurements in taxonomic problems. Ann Eugen. 1936;7(2):179-88.

60. Parks DH, Tyson GW, Hugenholtz P, Beiko RG. STAMP: statistical analysis of taxonomic and functional profiles. Bioinformatics. 2014;30(21):3123-4.

61. Langille MG, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, et al. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nat Biotechnol. 2013;31(9):814-21.

62. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. Nucleic acids research. 2015;44(D1):D457-D62.

63. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J Mol Biol. 2016;428(4):726-31.

64. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol. 1995;57(1):289-300.

65. Tukey JW. Comparing individual means in the analysis of variance. Biometrics. 1949;5(2):99-114.

66. Nelson L, Blair B, Murdock C, Meade M, Watts S, Lawrence AL. Molecular Analysis of gut microflora in captive-raised sea urchins (*Lytechinus variegatus*). J World Aquac Soc. 2010;41(5):807-15.

67. Nakagawa S, Saito H, Tame A, Hirai M, Yamaguchi H, Sunata T, et al. Microbiota in the coelomic fluid of two common coastal starfish species and characterization of an abundant *Helicobacter*-related taxon. Sci Rep. 2017;7(1):8764.

68. Enomoto M, Nakagawa S, Sawabe T. Microbial communities associated with holothurians: presence of unique bacteria in the coelomic fluid. Microbes Environ. 2009:1203190369-.

69. Sawabe T, Oda Y, Shiomi Y, Ezura Y. Alginate degradation by bacteria isolated from the gut of sea urchins and abalones. Microb Ecol. 1995;30(2):193-202.

70. Unkles S. Bacterial flora of the sea urchin *Echinus esculentus*. Appl Environ Microbiol. 1977;34(4):347-50.

71. Meziti A, Kormas KA, Pancucci-Papadopoulou M-A, Thessalou-Legaki M. Bacterial phylotypes associated with the digestive tract of the sea urchin *Paracentrotus lividus* and the ascidian *Microcosmus* sp. Russ J Mar Biol. 2007;33(2):84-91.

72. Kim D, Baik KS, Hwang YS, Choi J-S, Kwon J, Seong CN. *Vibrio hemicentroti* sp. nov., an alginate lyase-producing bacterium, isolated from the gut microflora of sea urchin (*Hemicentrotus pulcherrimus*). Int J Syst Evol Microbiol. 2013;63(10):3697-703.

73. Hakim JA, Schram JB, Galloway AW, Morrow CD, Crowley MR, Watts SA, et al. The Purple Sea Urchin *Strongylocentrotus purpuratus* Demonstrates a Compartmentalization of Gut Bacterial Microbiota, Predictive Functional Attributes, and Taxonomic Co-Occurrence. Microorganisms. 2019;7(2):35.

74. Farmer J, Hickman-Brenner F. The Genera *Vibrio* and *Photobacterium*. 3 ed. Dworkin M, Falkow S, Rosenberg E, Schleifer K-H, Stackebrandt E, editors: Springer Science & Business Media; 2006. 508-63

75. Gomez-Gil B, Roque A, Rotllant G, Peinado L, Romalde JL, Doce A, et al. *Photobacterium swingsii* sp. nov., isolated from marine organisms. Int J Syst Evol Microbiol. 2011;61(2):315-9.

76. Seo HJ, Bae SS, Lee J-H, Kim S-J. *Photobacterium frigidiphilum* sp. nov., a psychrophilic, lipolytic bacterium isolated from deep-sea sediments of Edison Seamount. Int J Syst Evol Microbiol. 2005;55(4):1661-6.

77. Yoon J-H, Lee J-K, Kim Y-O, Oh T-K. *Photobacterium lipolyticum* sp. nov., a bacterium with lipolytic activity isolated from the Yellow Sea in Korea. Int J Syst Evol Microbiol. 2005;55(1):335-9.

78. De Vadder F, Kovatcheva-Datchary P, Zitoun C, Duchampt A, Bäckhed F, Mithieux G. Microbiota-produced succinate improves glucose homeostasis via intestinal gluconeogenesis. Cell Metab. 2016;24(1):151-7.

79. Reichardt N, Duncan SH, Young P, Belenguer A, Leitch CM, Scott KP, et al. Phylogenetic distribution of three pathways for propionate production within the human gut microbiota. The ISME journal. 2014;8(6):1323.

80. Schink B, Pfennig N. *Propionigenium modestum* gen. nov. sp. nov. a new strictly anaerobic, nonsporing bacterium growing on succinate. Arch Microbiol. 1982;133(3):209-16.

81. Faust K, Raes J. Microbial interactions: from networks to models. Nat Rev Microbiol. 2012;10(8):538.

82. Berry D, Widder S. Deciphering microbial interactions and detecting keystone species with co-occurrence networks. Front Microbiol. 2014;5:219.

83. Apprill A. Marine animal microbiomes: toward understanding host–microbiome interactions in a changing ocean. Front Mar Sci. 2017;4:222.

84. Costello EK, Stagaman K, Dethlefsen L, Bohannan BJ, Relman DA. The application of ecological theory toward an understanding of the human microbiome. Science. 2012;336(6086):1255-62.

85. de la Calle F. Marine microbiome as source of natural products. Microb Biotechnol. 2017;10(6):1293.

86. Martínez I, Stegen JC, Maldonado-Gómez MX, Eren AM, Siba PM, Greenhill AR, et al. The gut microbiota of rural papua new guineans: composition, diversity patterns, and ecological processes. Cell Rep. 2015;11(4):527-38.

87. Vellend M. Conceptual synthesis in community ecology. Q Rev Biol. 2010;85(2):183-206.

88. David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, et al. Diet rapidly and reproducibly alters the human gut microbiome. Nature. 2014;505(7484):559.

89. Reese AT, Dunn RR. Drivers of microbiome biodiversity: a review of general rules, feces, and ignorance. mBio. 2018;9(4):e01294-18.

90. Gall CA, Scoles GA, Magori K, Mason KL, Brayton KA. Laboratory colonization stabilizes the naturally dynamic microbiome composition of field collected *Dermacentor andersoni* ticks. Microbiome. 2017;5(1):133.

91. Ng SH, Stat M, Bunce M, Simmons LW. The influence of diet and environment on the gut microbial community of field crickets. Ecol Evol. 2018;8(9):4704-20.

92. Tsukayama P, Boolchandani M, Patel S, Pehrsson EC, Gibson MK, Chiou KL, et al. Characterization of wild and captive baboon gut microbiota and their antibiotic resistomes. mSystems. 2018;3(3):e00016-18.

93. Sauchyn LK, Scheibling RE. Degradation of sea urchin feces in a rocky subtidal ecosystem: implications for nutrient cycling and energy flow. Aquat Biol. 2009;6:99-108.

94. Sauchyn LK, Scheibling RE. Fecal production by sea urchins in native and invaded algal beds. Mar Ecol Prog Ser. 2009;396:35-48.

95. Sauchyn LK, Lauzon-Guay J-S, Scheibling RE. Sea urchin fecal production and accumulation in a rocky subtidal ecosystem. Aquat Biol. 2011;13(3):215-23.

96. Becker PT, Samadi S, Zbinden M, Hoyoux C, Compère P, De Ridder C. First insights into the gut microflora associated with an echinoid from wood falls environments. Cah Biol Mar. 2009;50(4):343.

97. Tanrattanapitak N. Bacterial Community in Gut Contents of the Sea Urchin *Diadema setosum* (Leske, 1778) and the Ambient Sediments from Sichang Island using Metagenomics Approaches. NU Int J Sci. 2018;15(1):117-25.

98. Wang Y, Sheng H-F, He Y, Wu J-Y, Jiang Y-X, Tam NF-Y, et al. Comparison of the levels of bacterial diversity in freshwater, intertidal wetland, and marine sediments by using millions of illumina tags. Appl Environ Microbiol. 2012;78(23):8264-71.

99. Wirsen CO, Sievert SM, Cavanaugh CM, Molyneaux SJ, Ahmad A, Taylor L, et al. Characterization of an autotrophic sulfide-oxidizing marine *Arcobacter* sp. that produces filamentous sulfur. Appl Environ Microbiol. 2002;68(1):316-25.

100. Pati A, Gronow S, Lapidus A, Copeland A, Del Rio TG, Nolan M, et al. Complete genome sequence of *Arcobacter nitrofigilis* type strain (CI T). Stand Genomic Sci. 2010;2(3):300.

101. Fraune S, Bosch TC. Long-term maintenance of species-specific bacterial microbiota in the basal metazoan Hydra. Proc Natl Acad Sci. 2007;104(32):13146-51.

102. Cox CR, Gilmore MS. Native microbial colonization of *Drosophila melanogaste*r and its use as a model of *Enterococcus faecalis* pathogenesis. Infect Immun. 2007;75(4):1565-76.

103. Scupham AJ, Patton TG, Bent E, Bayles DO. Comparison of the cecal microbiota of domestic and wild turkeys. Microb Ecol. 2008;56(2):322-31.

104. Xenoulis PG, Gray PL, Brightsmith D, Palculict B, Hoppes S, Steiner JM, et al. Molecular characterization of the cloacal microbiota of wild and captive parrots. Vet Microbiol. 2010;146(3-4):320-5.

105. Wilson KH, Brown RS, Andersen GL, Tsang J, Sartor B. Comparison of fecal biota from specific pathogen free and feral mice. Anaerobe. 2006;12(5-6):249-53.

CHAPTER V: SHOTGUN METAGENOMICS REVEALED DIFFERENCES IN THE MICROBIOTA WITH KEY METABOLIC ATTRIBUTES EMPHASIZING NITROGEN FIXATION IN SEA URCHIN GUT DIGESTA

by

JOSEPH A. HAKIM, CASEY D. MORROW, STEPHEN A. WATTS, MICHAEL R. CROWLEY, AND ASIM K. BEJ

Submitted to *Scientific Reports*

Format adapted and errata corrected for dissertation

ABSTRACT

In this paper, we describe the composition and key metabolic genes of the microbial community in the mucous-enveloped gut digesta of green (*Lytechinus variegatus*) and purple (*Strongylocentrotus purpuratus*) sea urchins by using the shotgun metagenomics approach. Both the green and purple urchins showed high relative abundances of Gammaproteobacteria at 30% and 60%, respectively. However, the green urchins represented Alphaproteobacteria as significantly higher (20%) compared to the purple urchins (2%). At the genus level, *Vibrio* was dominant in both green (~9%) and purple (~10%), whereas *Psychromonas* was prevalent in purple urchins (~24%). Preferential enrichment of *Roseobacter* and *Ruegeria* were found in the green urchins, whereas purple urchins revealed a higher abundance of *Shewanella, Photobacterium,* and *Bacteroides* (q-value < 0.01). Analysis of key metabolic genes revealed the KEGG-Level-2 categories of amino acids (~20%), nucleotides (~5%), cofactors and vitamins (~6%), energy (~5%), and carbohydrates (~13%) metabolisms in both urchins. Moreover, genes in the assimilatory nitrogen reduction pathway were abundant in the metagenome of both urchins. Overall, the results from this study describe the differences of the microbial community and genes for the metabolic processes in the nutrient-rich sea urchin gut digesta, suggesting their importance to the host and their environment.

INTRODUCTION

The intertidal and nearshore marine ecosystems of North America harbor a diverse community of invertebrate, vertebrate, and microorganisms, along with primary producers such as drift and benthic macroalgae, kelp forests, and seagrass meadows that constitute a dynamic aquatic food web[1-3]. These ecosystems are enriched with inorganic and organic nutrients that are produced and utilized by the resident organisms. Among invertebrates, the sea urchins found in the nearshore coastal waters worldwide play a crucial role in the energy flow and nutrient cycling at various trophic levels[4]. In North America, the green sea urchin, *Lytechinus variegatus* (order Temnopleuroida, family Toxopneustidae), are generally found along the South Eastern Coast and into the Gulf of Mexico[5], whereas the purple sea urchin, *Strongylocentrotus purpuratus* (order Echinoida, family Strongylocentrotidae), inhabits the U.S. Pacific coast from Alaska to Baja Mexico[6]. Although the sea urchins are omnivores, they mostly graze upon algae, kelp, seagrass, and decomposed materials[5-10]. The grazing activity by increased densities of sea urchins often severely limits seagrass and macroalgal biomass, resulting in a barren and depauperate ecosystem[11,12]. However, the grazing enables the sea urchins to metabolize and transform the ingested seagrass and macroalgal biomass into rich organic nutrients. Thus, despite their potentially damaging effect, sea urchins also play an important ecological role in the structuring the communities in their habitats[4,11,13-16].

The efficacy of the digestive process of the carbohydrate-rich seagrass and macroalgal biomass by sea urchins has been of interest due to their unique organization and digestive enzymes described to be largely absent in the gut lumen[17]. Both green and purple urchins possess compartmentalized deuterostome gut system in a relatively straightforward model[18]. Normally, the ingested foods are masticated by the Aristotle's Lantern apparatus, which then enter the pharynx and esophagus, where a thick mucus layer produced by specialized mucous-producing cells in the gut tissue envelops them. This gut digesta are physically separated from the luminal surface of the gut tissue and will remain intact throughout their passage and upon egestion[17]. Additionally, this compartmentalization has been observed to occur with an enrichment of bacteria[19], which has recently been supported by reports demonstrating these microbial communities to be distinct from the gut tissue of both the purple[20] and the green[21,22] urchins. Previous studies described the crucial role of these microbial communities in the digestion and metabolism of the ingested seagrass and algae[17,19,23,24]. However, sea urchins are remarkably inefficient in assimilating these nutrients from the digesta[25]. As a result, a large quantity of mucous-enveloped high-energy egesta consisting of residual nutrients along with the microbiota is released in the ecosystem. These gut egesta are considered to be an important source of nutrients to marine organisms such as fish, crustaceans, shellfish, and other echinoderms[14,26-28]. The nutritional benefit of the green urchin egesta has also been shown to enhance the growth and the taste quality of shrimp *Litopenaeus vannamei* as compared to laboratory-formulated

feed alone[13]. One important aspect of these nutrient transitions involves the assimilation of nitrogen into amino acids and nucleotides, which are essential macromolecules for all living organisms[29]. Although various forms of inorganic and organic nitrogen exist in the marine environment, nearly 80% of the global nitrogen budget exists in its unreactive diatomic form[30]. However, the productivity of primary producers is dependent upon the availability of sufficient amounts of fixed nitrogen such as ammonia[31,32]. Moreover, assimilated organic nitrogen such as amino acids is crucial for marine heterotrophic animals to fulfill their nutritional requirements[33]. It has been suggested that the bacterial community in the sea urchin gut digesta play a vital role in the metabolism of nucleotides and amino acids through nitrogen-fixation[4,21,23,31,34].

The objective of this study was to determine the taxonomic composition and metabolic profile of the microbial communities in the mucous-encapsulated gut digesta of the green and purple urchins using shotgun metagenomics. The study elaborates the signature genes for the carbohydrate, amino acid, and nucleotide metabolisms, with an emphasis on the nitrogen cycle.

MATERIALS AND METHODS

Microbial Metacommunity DNA for High-Throughput Sequencing

The collection of the green sea urchins, *L. variegatus*, from the Gulf of Mexico Coast of Florida (Florida Coast) and the purple sea urchins, *S. purpuratus*, from the Pacific Coast of Oregon (Oregon Coast), and the purified metacommunity DNA of their gut digesta used in this study were obtained from

previous investigations by our laboratory[20,22]. Briefly, two groups of metacommunity DNA of gut digesta were prepared, one group representing three (*n*=3) pooled biological replicates of the *L. variegatus* gut digesta, and the other group consisted of three (*n*=3) biological replicates of *S. purpuratus*. Then, each of the two pooled DNA samples was further aliquoted into two technical replicates and subjected to shotgun sequencing on an Illumina HiSeq 2500 sequencing platform at the UAB Heflin Center for Genomic Science (http://www.uab.edu/hcgs/), using the paired-end (2 X 100) protocol described elsewhere[35]. The number of raw sequences generated by the Illumina HiSeq HTS has been described in Table 5-1.

MG-RAST Quality Checking and Sequence Read Processing

Initial sequence read quality was assessed using FastQC (v.0.11.2)[36], and host DNA was filtered from the raw sequence data using Bowtie2 (v2.3.4.3)[37] against the most recent *Lytechinus variegatus* (assembly Lvar_2.2) and *Strongylocentrotus purpuratus* (assembly Spur_4.2) genome assembly from the National Center for Biotechnology Information (NCBI). Host-filtered paired-end sequences were then uploaded to Metagenomic Rapid Annotations using Subsystems Technology (MG-RAST; v4.0.3)[38], and processed using the default MG-RAST pipeline. Briefly, low-quality sequences were trimmed using the Dynamic Trim tool from SolexaQA[39], with the lowest PHRED quality score selected at 15, and sequences were trimmed if five nucleotides at most fell below this threshold. Taxonomic classifications were assigned using the NCBI

Reference Sequence (RefSeq) Database[40], and functional categories were assigned through the Kyoto Encyclopedia of Genes and Genomes (KEGG) Orthology (KO) Database[41,42], with the minimum sequence length set at 15 nucleotides at an E-value of 10^{-5} for a sequence match. The assigned taxonomic count data representing domain Bacteria (Bacteriome) were extracted and subsampled to the minimum value across all samples (RefSeq = 646,758). The KO data was also subsampled to the minimum count value (KO = 206,021). Subsampling was performed using the "single_rarefaction.py" command from Quantitative Insights into Microbiology (QIIME, v1.9.1)[43].

Taxonomic Distribution

The rarefied taxonomic relative abundance was visualized and plotted using Microsoft Excel Software (Microsoft, Seattle, WA, USA). First, a relative abundance stacked column bar graph was generated to show the distribution of phyla (class for Proteobacteria) across all samples, and entries represented at <1% abundant across all samples were listed as "Other." This phylum level relative abundance information was also used to create a heatmap table using the conditional formatting option in Microsoft Excel Software. Colors were selected as "red" for less abundant, "yellow" for intermediate abundance, and "green" for high abundance. The relative abundance distribution at the genus level was also plotted as stacked column bar format, showing the top 50 taxa with the remaining taxa listed as "Other." This data was also used for extended error bar analysis in Statistical Analysis of Metagenomic Profiles (STAMP;

v2.1.3). For each urchin gut digesta, technical replicates were grouped and normalized, and a two-sided Welch's t-test[44] was performed with *p*-values corrected using the Benjamini-Hochberg False Discovery Rate (FDR) approach[45]. The resultant *q*-values were set at < 0.01 to indicate a significant differential abundance of genera between gut digesta groups.

Alpha and Beta Diversity of Taxa

Shannon[46] and Simpson[47] alpha diversity measures were performed at the most resolvable Operational Taxonomic Unit (OTU) level (species where possible) using the "alpha_diversity.py" command through QIIME (v1.9.1). Beta diversity was performed through heatmap analysis at the genus level. To do this, a heatmap was constructed in R (v3.3.2) using the heatmap.2 function from gplots (v3.0.1) package[48]. The sample dendrograms were constructed using Vegan (v2.5-3) package according to the Bray-Curtis metric[49], and taxa represented at <1% were filtered out. The RColorBrewer package[50] was used to select the color palette at blue for more abundant, and sky blue for least abundant, and a trace line was plotted (black bar lines) to additionally show the percentage distribution.

Functional Analysis through KEGG Orthology

Functional analysis was performed using the rarefied KEGG Orthology (KO) functional categories across all samples. First, the KO data were collapsed into their respective KEGG-Level-2 hierarchical category and plotted as relative

abundance bar graphs using Microsoft Excel software (Microsoft, Seattle, WA, USA). The heightened categories (>1%) were shown, and the remaining low-abundant KEGG-Level-2 categories were listed as “Other.” Each included KEGG-Level-2 category was then ranked by average relative abundance within its higher KEGG-Level-1 category of metabolism, genetic information, environmental and cellular processing. The KEGG-Level-1 category of metabolism was further analyzed to show the preferential abundance of KEGG-Level-3 KEGG function map Ids derived from the KEGG Pathway Database between the green and purple urchin gut digesta samples. To do this, the KEGG map Ids were grouped according to technical replicate and rendered as a scatter plot based on their relative proportion in the KEGG-Level-1 category of metabolism using STAMP (v2.1.3) analytical software[51]. Moreover, the metabolic pathways related to the KEGG-Level-3 pathway of nitrogen metabolism was further investigated using KEGG Mapper[52] at the highest level of functional resolution (KO) and reconstructed as it relates to amino acid metabolism. The sequence counts assigned to each KO category in the pathway were listed in bar graphs through Microsoft Excel software (Microsoft, Seattle, WA, USA), and included alongside the direction of the indicated reactions.

RESULTS

Sequence Statistics

The total number of raw sequences were subjected to host DNA removal, followed by the quality-checked sequences from *L. variegatus* and *S. purpuratus*

that were used for downstream bioinformatics analysis are listed in Table 5-1. The taxonomic assignments from the quality checked sequences and subsequent KO functional categories from both the green and purple urchin samples are also listed in Table 5-1.

Taxonomic Distribution of the Bacteriome

For relative abundance distribution at the phylum (class for Proteobacteria) level, the green urchin digesta samples showed Gammaproteobacteria (~30%) and Alphaproteobacteria (~20%) to be heightened (Figure 5-1A). Comparatively, the purple urchin digesta samples showed Gammaproteobacteria to be most abundant at ~60%, whereas Alphaproteobacteria was only represented at ~2%. Of the commonly found phyla across both sea urchin digesta samples, the green urchins showed a slightly higher relative abundance of Bacteroidetes, Cyanobacteria, Actinobacteria, Planctomycetes, Verrucomicrobia, Beta- and Epsilonproteobacteria, and the purple urchins had a slightly heightened Firmicutes, Fusobacteria, and Deltaproteobacteria (Figure 5-1A; Table 5-2).

At the genus level, both the green and purple urchins gut digesta showed a heightened distribution of *Vibrio* (Figure 5-1B). However, the purple urchin showed uniquely heightened enrichment of *Psychromonas*. Of the commonly observed genera between the two groups, extended error bar analysis indicated *Roseobacter* and *Ruegeria* as significantly heightened in the green urchins compared to the purple urchins (q-value < 0.01) (Supplementary Figure 5-1).

Table 5-1: Sample statistics determined following high-throughput sequencing on the Illumina HiSeq platform, sequence processing, taxonomic and functional assignments. The total sequences 1) prior to and 2) after host sequence removal are shown and following 3) quality checking through MG-RAST (v4.0.3). The sequences assigned to each domain through RefSeq are shown and include: 4) Archaea, 5) Bacteria, 6) Eukaryota, and 7) Viruses. High-quality sequences that did not receive a taxonomic identity were listed as 8) “Other Seqs”. Alpha diversity was determined for each sample based on the 9) Shannon and 10) Simpson metrics as implemented through QIIME (v1.9.1) at the species level of taxonomic resolution. 11) The number of sequences assigned to a functional gene through KEGG Orthology (KO) was also indicated.

Sample	Total Seqs	Host-Removed Seqs	QC Seqs	Archaea	Bacteria	Eukaryota	Viruses	Other Seqs	Shannon	Simpson	KO Functions
LV.GD1	11,179,611	1,903,485	1,879,183	4,114	646,758	22,505	285	3	8.99	0.996	206,021
LV.GD2	12,989,418	2,191,267	2,162,089	4,738	747,440	25,869	367	4	8.99	0.996	238,974
SP.GD1	11,280,351	6,232,967	6,082,068	44,711	3,323,484	43,539	2,634	4	7.47	0.965	1,103,493
SP.GD2	12,994,228	7,169,738	6,976,796	51,113	3,823,945	49,481	3,164	12	7.47	0.965	1,265,954

Table 5-2: Phylum (class for Proteobacteria) level heatmap table of the green and purple sea urchin gut digesta based on the taxonomic assignments determined through RefSeq. Phyla represented at <1% across all samples were listed as "Other." The minimum and maximum relative abundance values were used to establish the gradient color scale, with 0.56% set at the minimum and 61.68% at the maximum. The gradient percentage values are indicated as follows: red = low; yellow = intermediate; green = high relative abundance. The table was generated using Microsoft Excel Software (Microsoft, Seattle, WA, USA) using the conditional formatting option, and samples are indicated as follows: LV.GD = green sea urchin *Lytechinus variegatus* gut digesta; SP.GD = purple sea urchin *Strongylocentrotus purpuratus* gut digesta.

Phylum	**LV.GD1**	**LV.GD2**	**SP.GD1**	**SP.GD2**
Actinobacteria	2.99%	2.95%	0.69%	0.68%
Bacteroidetes	16.29%	16.21%	12.89%	12.84%
Cyanobacteria	2.60%	2.62%	0.62%	0.60%
Firmicutes	5.85%	5.85%	6.13%	6.10%
Fusobacteria	0.89%	0.87%	3.07%	3.13%
Planctomycetes	4.96%	5.02%	0.56%	0.56%
Proteobacteria (Alphaproteobacteria)	20.45%	20.61%	2.56%	2.55%
Proteobacteria (Betaproteobacteria)	3.17%	3.22%	1.60%	1.61%
Proteobacteria (Deltaproteobacteria)	4.65%	4.75%	5.61%	5.60%
Proteobacteria (Epsilonproteobacteria)	2.52%	2.50%	0.68%	0.66%
Proteobacteria (Gammaproteobacteria)	29.83%	29.55%	61.64%	61.68%
Verrucomicrobia	1.74%	1.78%	0.75%	0.75%
Other	4.06%	4.06%	3.19%	3.25%

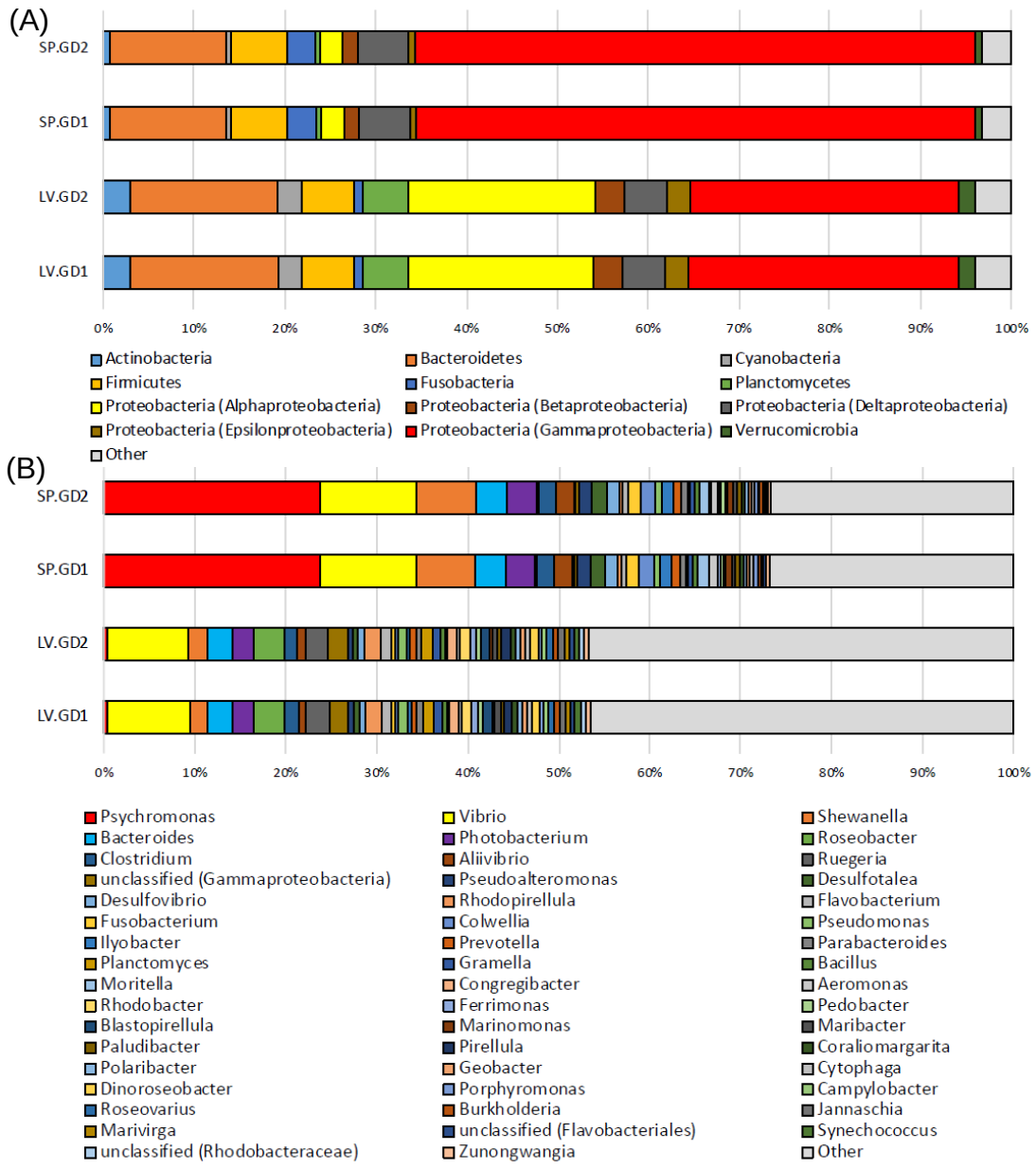


Figure 5-1: Relative abundance stacked column bar graphs illustrating the distribution of taxa assigned to domain Bacteria (Bacteriome) through RefSeq as implemented in MG-RAST (v4.0.3). (A) The phylum level (class for Proteobacteria) distribution is shown, and phyla with an average abundance of <1% were listed as “Other.” (B) The relative abundance of the top 50 genera across all samples was plotted, and the low represented taxa were categorized as “Other.” Bar graphs generated using Microsoft Excel Software (Microsoft, Seattle, WA, USA), and samples are indicated as follows: LV.GD = green sea urchin *Lytechinus variegatus* gut digesta; SP.GD = purple sea urchin *Strongylocentrotus purpuratus* gut digesta.

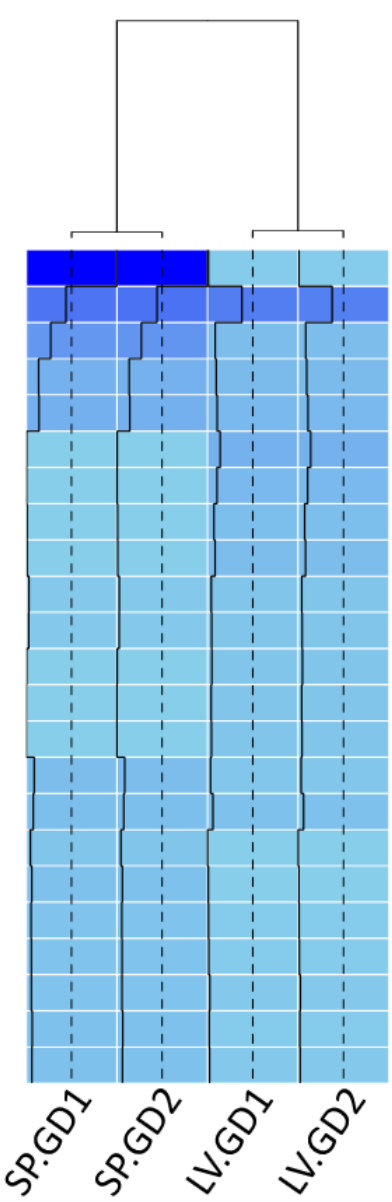
Conversely, the purple urchin digesta showed *Shewanella, Photobacterium,* and *Bacteroides* to be significantly enriched compared to the green urchins (q-value < 0.01). The two-group Welch's t-test results for each of the top 50 genera, including the p-values, Benjamini-Hochberg FDR corrected q-values, and effect size differences have been elaborated in Supplementary Table 5-1.

Alpha and Beta Diversity

Alpha diversity of the samples in this study showed the green urchin digesta to have a higher Shannon and Simpson diversity as compared to the purple urchin digesta samples (Table 5-1). Beta diversity demonstrated the technical replicates to cluster within a 97% Bray-Curtis similarity, and the beta diversity of the digesta across all samples was shown to be >60% as shown through dendrogram cluster analysis (Figure 5-2). Heatmap analysis further elaborated the beta diversity and the contribution of heightened genera (>1%) to the observed variation of sample types (Figure 5-2).

Functional Categories using the KEGG-Level-1 and Level-2 Orthology

The relative distribution of genes assigned showed the KEGG-Level-1 category of metabolism (global overview of carbohydrate, lipid, amino acids, energy, co-factor and vitamins, nucleotides, secondary metabolites, and terpenoid/polyketides, chemical structure transformation) to be the most heightened category, represented at approximately 60% for the green and 58%



Taxa	SP.GD1 (%)	SP.GD2 (%)	LV.GD1 (%)	LV.GD2 (%)
Psychromonas	23.77	23.81	0.31	0.32
Vibrio	10.46	10.58	9.09	8.99
Shewanella	6.45	6.48	2.09	2.07
Photobacterium	3.30	3.28	2.36	2.34
Bacteroides	3.42	3.46	2.61	2.67
Roseobacter	0.15	0.16	3.39	3.39
Ruegeria	0.12	0.11	2.56	2.57
Rhodopirellula	0.27	0.26	1.77	1.81
Unclass. Gammap.	0.27	0.28	2.06	2.04
Pseudomonas	0.75	0.70	1.07	1.06
Flavobacterium	0.70	0.68	1.17	1.14
Planctomyces	0.16	0.16	1.21	1.22
Rhodobacter	0.07	0.07	1.04	1.03
Congregibacter	0.07	0.07	1.12	1.11
Aliivibrio	2.12	2.12	0.87	0.87
Clostridium	1.89	1.87	1.51	1.52
Moritella	1.13	1.12	0.10	0.10
Colwellia	1.51	1.50	0.29	0.29
Ilyobacter	1.33	1.35	0.34	0.33
Fusobacterium	1.39	1.43	0.41	0.41
Desulfovibrio	1.40	1.41	0.69	0.68
Desulfotalea	1.62	1.62	0.53	0.53
Pseudoalteromonas	1.53	1.52	0.62	0.62

Figure 5-2: Heatmap analysis of the genera comprising domain Bacteria (Bacteriome) assigned through RefSeq using MG-RAST (v4.0.3). The analysis was performed using R (v3.3.2) with the heatmap.2 function from the gplots (v3.0.1) package. The sample dendrogram was constructed based on the Bray-Curtis similarity value through Vegan (v2.4.3), and the gradient of relative abundance was illustrated using RColorBrewer package as “blue” for more abundant and “sky blue” for less abundant. The trace lines (black) were generated to further elaborate the relative abundance of taxa. Samples are indicated as follows: LV.GD = green sea urchin *Lytechinus variegatus* gut digesta; SP.GD = purple sea urchin *Strongylocentrotus purpuratus* gut digesta.

for the purple urchin samples. This was followed by genetic information processing (green = ~18% and purple = ~18%), and environmental information processing (green = ~14% and purple = ~16%). From the heightened KEGG-Level-1 category of metabolism, the KEGG-Level-2 categories showed amino acid and carbohydrate metabolism to be the most prevalent in both the green and purple urchin digesta (Figure 5-3). This was followed by the metabolism of cofactors, vitamins, nucleic acid, and lipids.

The analysis of metabolic functions from within the KEGG-Level-2 categories showed amino acids metabolism to be most prevalent (Figure 5-4). Furthermore, the KEGG-Level-3 pathway assigned to alanine, aspartate, and glutamate (KEGG map Id: 00250) was prevalent in both the green and purple urchin digesta (Figure 5-4). However, when compared between the two urchin digesta, the purple urchins were higher than the green urchins. In addition, glycine, serine, and threonine metabolism (KEGG map Id: 00260) were demonstrated at higher abundance in the green urchins compared to purple urchins. Other heightened amino acid metabolism categories included arginine and proline (KEGG map Id: 00330) that favored the purple urchins, whereas valine, leucine and isoleucine degradation (KEGG map Id: 00190) in the green urchins. Categories involved in energy metabolism were also observed, showing nitrogen metabolism (KEGG map Id: 00910) to be more enriched in the purple urchins compared to the green urchins. In green urchins, the citrate cycle (KEGG map Id: 00020) and oxidative phosphorylation were enriched as compared to the purple urchins.

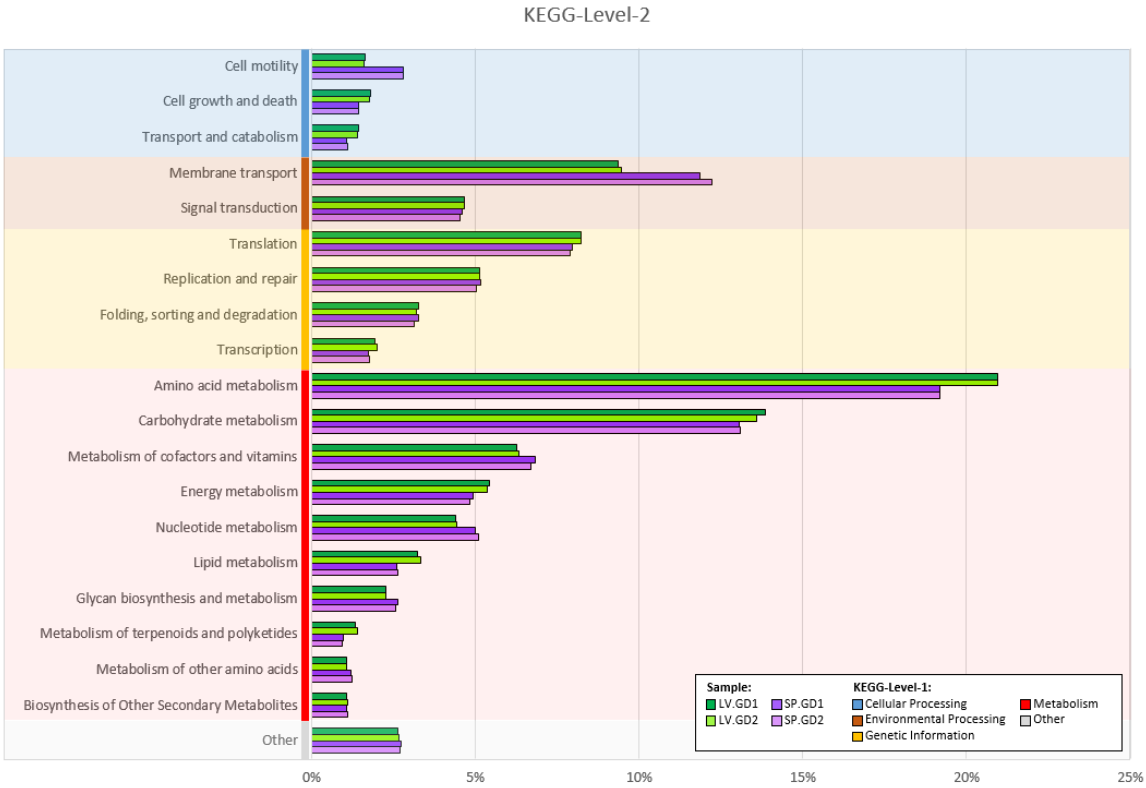


Figure 5-3: Relative abundance bar graphs generated for each sample based on the KEGG-Level-2 categories determined through KEGG Orthology (KO) data following the MG-RAST (v4.0.3) workflow. KEGG-Level-2 categories were binned into their respective KEGG-Level-1 broad hierarchical functional category and ranked in decreasing order from top to bottom based on their average abundance. Each KEGG-Level-1 category was color coded as follows: cellular processing = blue; environmental processing = brown; genetic information = orange; metabolism = red; and “other” = grey. The bar graphs for each sample have been color-coded and indicated as follows: SP.GD = purple sea urchin *Strongylocentrotus purpuratus* gut digesta; SP.GD1 = purple and SP.GD2 = light purple; LV.GD = green sea urchin *Lytechinus variegatus* gut digesta; LV.GD1 = green and LV.GD2 = light green.

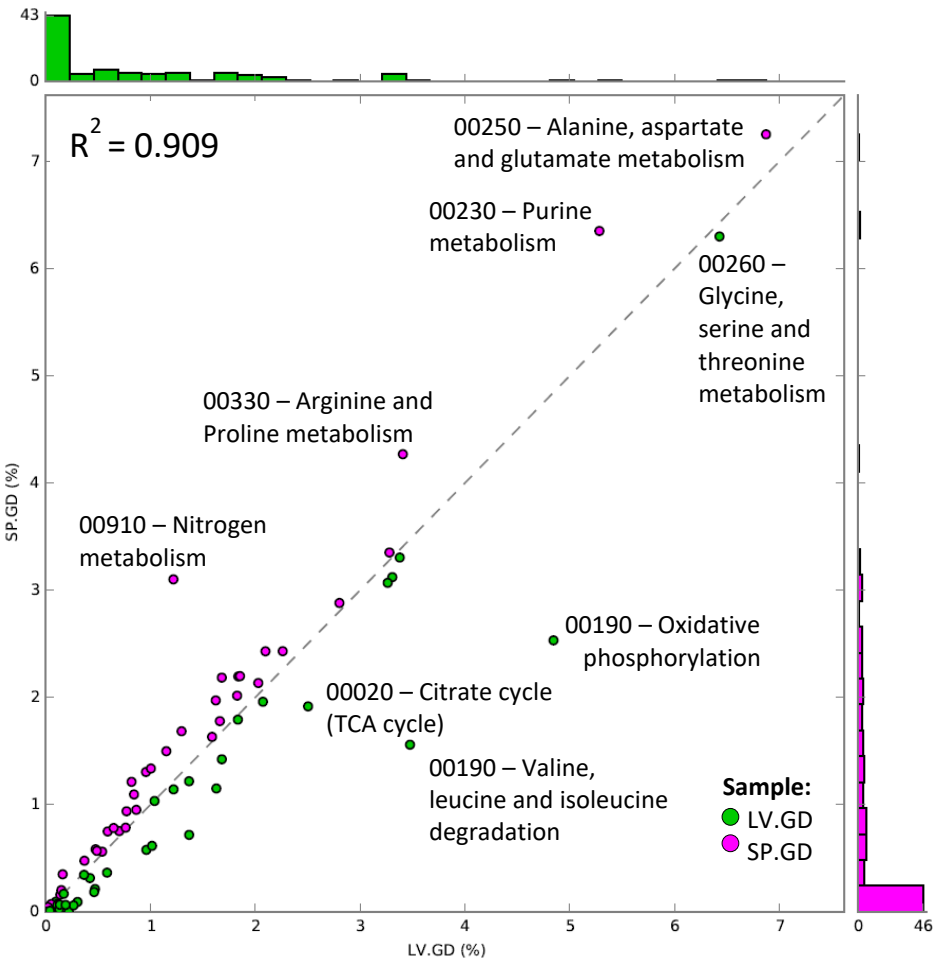


Figure 5-4: Relative abundance scatter plot analysis of the KEGG map Ids derived from the KEGG-Level-1 category of “metabolism”. KEGG map Id data was retrieved from the KEGG-Level-3 functional categories from MG-RAST (v4.0.3) and was uploaded into STAMP (v2.1.3). Technical replicates were grouped, and count data normalized as the relative proportion of each KEGG map Id per group from the KEGG-Level-1 category of metabolism. The X-axis and Y-axis show the relative abundance of each KEGG map Id per group, including the histograms to show the number of functional entries falling at the specified abundance on the scatter plot. Those KEGG map Ids that were noticeably enriched in one group were indicated by their KEGG pathway name and Id number. The regression analysis (R^2) was also shown in the plot. Sample groups and color code are indicated as LV.GD = green sea urchin *Lytechinus variegatus* gut digesta (green); SP.GD = purple sea urchin *Strongylocentrotus purpuratus* gut digesta (purple).

Genes for the Assimilatory Reduction and Fixation of Nitrogen into Ammonia

An elaboration of the nitrogen metabolism pathway as it relates to amino acid metabolism through KEGG Mapper showed pathways involved in ammonia (NH_3) production to be abundant (Figure 5-5A and 5-5B). In the nitrate reduction pathway, the *nasA* assimilatory nitrate reductase catalytic subunit (KO number: K00372) and *napA* periplasmic nitrate reductase (KO number: K02567) were heightened, as well as nitrite reduction *nirB* (KO number: K00362) and *nrfA* (KO number: K03385) corresponding to nitrite reductase (NADH) large subunit and nitrite reductase (cytochrome c-552), respectively (Figure 5-5A). Nitrogen fixation was also represented by the *nif* cluster, which included the nitrogenase iron protein (*nifH*; KO number: K02588), nitrogenase molybdenum-iron protein alpha (*nifD*; KO number: K02586) and beta chain (*nifK*; KO number: K02591), nitrogenase molybdenum-cofactor synthesis protein (*nifE*; KO number: K02587), nitrogen fixation protein (*nifB*; KO number: K02585), nitrogenase molybdenum-iron protein (*nifN*; KO number: K02592), and nitrogen fixation homocitrate synthase (*nifV*; KO number: K02594) at a noticeable abundance, particularly in the purple urchins. The other genes in ammonia production that were noticeably abundant included glutamate dehydrogenase, and aspartate ammonia-lyase, which are elaborated as part of the ammonia assimilatory pathway.

Genes for Ammonia Assimilation into Glutamine and Asparagine

Genes involved in the assimilation of ammonia into various amino acids, namely glutamine and asparagine were observed as part of the nitrogen

metabolism pathway (Figure 5-5B). For the assimilation of ammonia to generate glutamine, glutamine synthetase (*glnA;* KO number: K01915) was abundant in both the green and purple urchins. Similarly, genes involved in the transition of glutamine to glutamate were also heightened, which included glutamate synthase ferredoxin (*gltS*; KO number: K00284), glutaminase (*glsA;* KO number: K01425), and both the glutamate synthase (NADPH) large chain (*gltB*; KO number: K00265) and small chain (*gltD*; KO number: K00266). Also included in this pathway was asparagine synthase (*asnB*; KO number: K01953), which hydrolyzes glutamine and aspartate to produce glutamate and asparagine. In the generation of ammonia from glutamate, glutamate dehydrogenase (*gudB*, *rocG*; KO number: K00260), and glutamate dehydrogenase (NADP+) (*gdhA*; KO numbers: K00261 and K00262) were abundant. For the assimilation of ammonia into asparagine, aspartate-ammonia ligase (*asnA*; KO number: K01914) was represented in both the green and purple urchins. Asparagine synthase (*asnB*; KO number: K01953) was also included in this pathway and denoted by an asterisk due to its dual function in generating both asparagine and glutamate from glutamine as indicated above. Additionally, L-asparaginase (*ansA* and *ansB;* KO number: K01424), which is involved in the synthesis of aspartate from asparagine, was represented in the continuation of this cycle. Lastly, aspartate ammonia-lyase (*aspA*; KO number: K01744) was also represented, which generates ammonia and fumarate from L-aspartate.

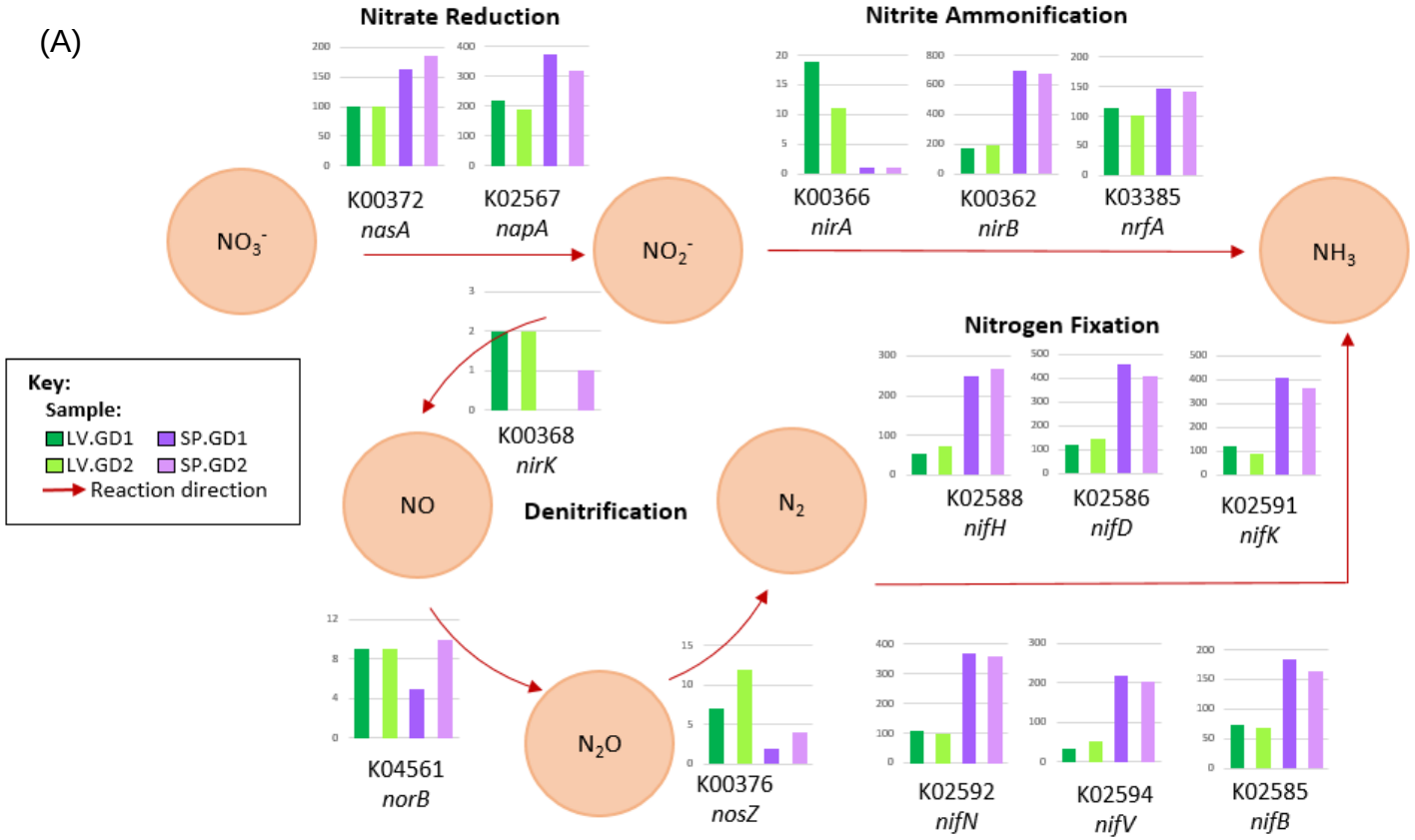
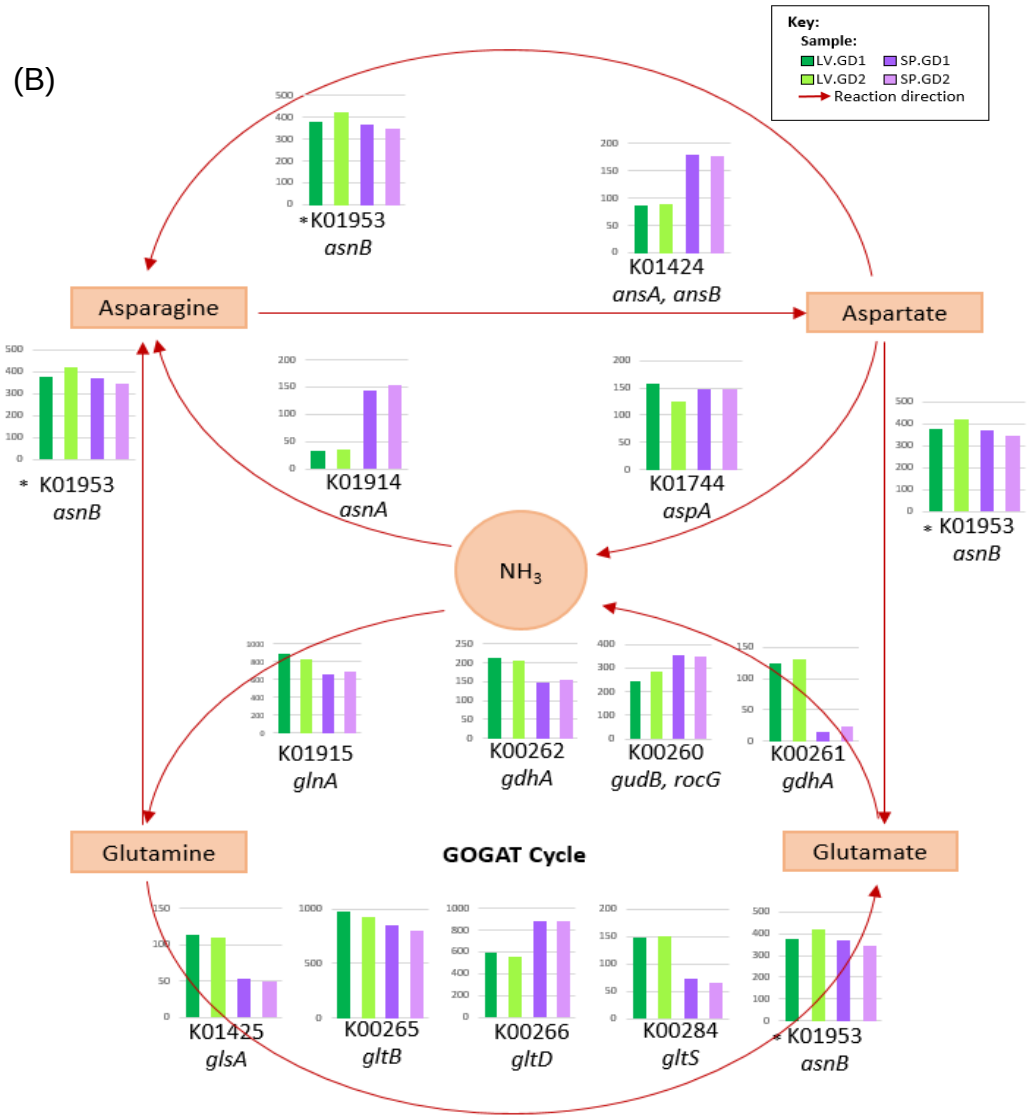


Figure 5-5: The KEGG Orthology (KO) functional categories comprising the KEGG-Level-3 pathway of nitrogen metabolism (00910) were mapped. The reaction directions were indicated by the arrows. (A) Inorganic nitrogen metabolism was mapped to show the substrate and product as it pertains to nitrate (NO_3^-) reduction, nitrite (NO_2^-) ammonification, denitrification to nitric oxide (NO), nitrous oxide (N_2O), and nitrogen gas (N_2), and nitrogen fixation into ammonia (NH_3); (B) The assimilation of ammonia into the amino acids glutamine and asparagine are also shown,



(Figure 5-5 continued) including the subsequent transitions of the amino acids. For each gene, the count data is presented as bar graphs for each gut digesta, including the KO number and the gene name assigned through the KEGG database. Each category was color-coded according to their sample indicated in the color key as follows: SP.GD = purple sea urchin *Strongylocentrotus purpuratus* gut digesta; SP.GD1 = purple and SP.GD2 = light purple; LV.GD = green sea urchin *Lytechinus variegatus* gut digesta; LV.GD1 = green and LV.GD2 = light green. The metabolic pathways were elaborated using the information provided through KEGG Mapper as implemented in MG-RAST (v4.0.3). *The KO number K01953 corresponding to the *asnB* gene is listed multiple times due to its function in the generation of glutamate and asparagine from glutamine and aspartate.

DISCUSSION

The overall bacterial community composition in both the green and purple sea urchin gut digesta at the phylum (*e.g.* Proteobacteria, Bacteroidetes, and Firmicutes) and the class levels (*e.g.* Gammaproteobacteria, Alphaproteobacteria) resulted from our study showed a similar trend in the gut of other Echinoderms such as sea cucumbers[53,54], sea stars[55], and brittle stars[56], as well as in a broad range of marine invertebrates such as sponges[57], tunicates[58], marine copepods[59], and corals[60]. In addition, previous studies in our laboratory also had similar microbial taxa in these sea urchin gut digesta based on high-throughput amplicon sequencing of the 16S rRNA gene[20,22]. Thus, both shotgun and 16S rRNA amplicon-based metagenomics approaches portrayed a commonality in bacterial taxa in Echinoderms and other marine invertebrates.

At the genus level, the high prevalence of the psychrophilic bacterium, *Psychromonas* (Gammaproteobacteria), within the purple urchin digesta could be due to the colder yearly water temperatures (10 - 12.8 °C; average = 11.7 °C) in the Oregon Coast compared to the Florida Coast (13.3 - 30 °C; average = 22 °C) (www.nodc.noaa.gov/dsdt/cwtg/npac.html). In contrast, the presence of a high relative abundance of *Vibrio* in both sea urchins could be due to their tolerance to a wider range of temperatures[22,61,62]. Earlier studies have indicated *Vibrio* to be commonly found in sea urchins from diverse marine habitats[19,63-65]. Such ubiquity of *Vibrio* in sea urchins has been implicated in metabolism carbohydrate-rich algae and seagrass as their primary food source[19,63,64]. Similarly, the metabolic

benefit of the *Psychromonas* in the purple urchin could also be carbohydrate-related metabolism and ω-3 polyunsaturated fatty acids (PUFA) production[16,66-69].

The use of shotgun metagenomic sequencing in this study revealed an in-depth identification of the genes within the KEGG-Level-1 Orthology reference hierarchy of “metabolism” (K09100). Carbohydrate and amino acid metabolisms identified from KEGG-Level-2 indicated a high genetic potential of digestion and nutrient assimilation from the natural food source by the microbial communities in both green and purple urchin digesta. These results support and expand our knowledge beyond previously predicted metabolic profiles determined by our laboratory in the gut digesta of both green and purple urchins by using Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt) analysis[20,22].

All animals require fixed nitrogen, which is normally conducted by nitrogen-fixing bacteria for the synthesis of amino acid and nucleotides[70]. In particular, bacteria in the gut ecosystems of herbivores play a vital role in generating bioavailable forms of nitrogen by recycling nitrogen from organic molecules, reducing marine nitrate/nitrite, or fixing elemental nitrogen into ammonia[34]. Such metabolisms have been reported in diverse marine invertebrates[34], including bivalves[70-72] and sea urchins[73,74]. Similar to the results of our study, the nitrate/nitrite reduction pathways and nitrogen fixation have also been described in the marine sponge *Hymeniacidon heliophila,* with emphasis on the assimilation into organic molecules[75]. Importantly, previous studies on sea urchin gut bacteria have linked these microbial-mediated nitrogen metabolic

processes benefitting the host as well as nutrient enrichment at the trophic levels. Specifically, bacteria of the sea urchin gut system have been suggested to play important roles in the synthesis of essential amino acids[23], and other studies have suggested nitrogenase-positive *Vibrio* isolates from the gut to provide a source of fixed nitrogen to the sea urchin host[73,74]. Moreover, studies performed on kelp-feeding *Strongylocentrotus droebachiensis* sea urchin aggregates have demonstrated an increase in the organic nitrogen fraction of their egesta at a rate of 0.21 g nitrogen m^2/day. This activity indicates that the egesta are an energy-rich substrate to neighboring marine organisms at various trophic levels[14]. In our study, we obtained new insights into the genes involved in the fixation of nitrogen and assimilatory reduction of nitrate and nitrite to ammonia in the sea urchins gut digesta, as compared to the studies reported by other investigators using non-genomic approaches[23,73,74]. Specifically, the genes involved in the assimilation pathway of ammonia to amino acids and nucleotides by the synthesis of glutamine and asparagine indicates that the gut digesta is enriched with essential nutrients. The metabolic pathway for the assimilation of excess ammonia may help in reducing the toxic effect to the sea urchins and the inhibition of the nitrate and nitrite reduction pathway conducted by the microbiota[76-79]. This uniquely compartmentalized mucous-enveloped gut digesta formed shortly after the ingestion of food functions to maintain the necessary anaerobic environment for carbohydrate fermentation, and the utilization of alternative inorganic electron acceptors in the nitrogen reduction pathways[24].

In conclusion, the shotgun metagenomics of microbial community composition in the gut digesta of green and purple urchins revealed taxa distribution with high richness at the phylum level, which is consistent with other echinoderms as well as a broad range of marine invertebrates. The high abundance of *Vibrio* in both green and purple urchins, and *Psychromonas* in purple urchins could be associated with their role in the digestion and metabolism of their carbohydrate-rich diet. Overall, this study provides new insights into the structure of the microbial communities and genes for the metabolism of the key biological macromolecules in the nutrient-rich gut digesta of sea urchins from two separate coastal ecosystems of North America. Future studies focusing on the functionality of these genes would help understand the metabolic processes conducted by the microbial communities in the gut digesta and how these processes benefit the sea urchin nutrition during their passage through the gut lumen and impact trophic levels following egestion.

CONFLICT OF INTERESTS

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

ACKNOWLEDGEMENTS

We would like to thank Dr. Peter Eipers of the Department of Cell, Developmental and Integrative Biology at the University of Alabama at

Birmingham (UAB) for his assistance in the high-throughput sequencing for this study. Graduate Research Assistant funding to J.A.H. was provided from grant support to C.D.M. by the UAB School of Medicine. We would also like to thank the Biology Department at UAB for logistics.

FUNDING

The following are acknowledged for their support of the Microbiome Resource at the University of Alabama at Birmingham, School of Medicine: Comprehensive Cancer Center (P30AR050948), Center for Clinical Translational Science (UL1TR000165), the Heflin Center for Genomic Sciences and the UAB Microbiome Center. Research supported in part by NIH NORC Nutrition Obesity Research Center (P30DK056336) (S.A.W.). Animal care and use for the experiments of this study were conducted under the Institutional Animal Care and Use Committee (IACUC-10043; 17 Jan 2014–current).

AUTHOR CONTRIBUTIONS

AJ.A.H. and A.K.B. conceived the study, analysis, and interpretation of the data, and wrote the first draft of the manuscript. S.A.W. and C.D.M. critically reviewed the manuscript, edited and helped interpret the data. C.D.M. and M.R.C. conducted the high-throughput sequencing, helped data processing, and provided critical comments on the manuscript.

REFERENCES

1 Steneck, R. S., Vavrinec, J. & Leland, A. V. Accelerating trophic-level dysfunction in kelp forest ecosystems of the western North Atlantic. *Ecosystems* **7**, 323-332 (2004).

2 van der Zee, E. M. *et al.* How habitat-modifying organisms structure the food web of two coastal ecosystems. *Proc. Biol. Sci.* **283**, 20152326 (2016).

3 Zieman, J. C. & Zieman, R. T. The ecology of the seagrass meadows of the west coast of Florida: a community profile. (Fish and Wildlife Service, Washington, DC (USA); Virginia Univ., 1989).

4 Sauchyn, L. K. & Scheibling, R. E. Fecal production by sea urchins in native and invaded algal beds. *Mar Ecol Prog Ser.* **396**, 35-48 (2009).

5 Albright, R. *et al.* Juvenile growth of the tropical sea urchin *Lytechinus variegatus* exposed to near-future ocean acidification scenarios. *J. Exp. Mar. Biol. Ecol.* **426**, 12-17 (2012).

6 Ebert, T. A., Schroeter, S. C., Dixon, J. D. & Kalvass, P. Settlement patterns of red and purple sea urchins (*Strongylocentrotus franciscanus* and *S. purpuratus*) in California, USA. *Mar Ecol Prog Ser.*, 41-52 (1994).

7 Davidson, T. M. & Grupe, B. M. Habitat modification in tidepools by bioeroding sea urchins and implications for fine-scale community structure. *Mar. Ecol.* **36**, 185-194 (2015).

8 Ellis, S. L. *et al.* Four regional marine biodiversity studies: approaches and contributions to ecosystem-based management. *PLoS One* **6**, e18997 (2011).

9 Tegner, M. & Dayton, P. Ecosystem effects of fishing in kelp forest communities. *ICES J. Mar. Sci.* **57**, 579-589 (2000).

10 Watanabe, J. M. & Harrold, C. Destructive grazing by sea urchins *Strongylocentrotus* spp. in a central California kelp forest: potential roles of recruitment, depth, and predation. *Mar Ecol Prog Ser.*, 125-141 (1991).

11 Pearse, J. S. Ecological role of purple sea urchins. *Science* **314**, 940-941 (2006).

12 Zhadan, P. M., Vaschenko, M. A. & Almyashova, T. N. in *Sea Urchin-From Environment to Aquaculture and Biomedicine* (ed Maria Agnello) (InTech, 2017).

13 Jensen, K. E. *et al.* The value of sea urchin, *Lytechinus variegatus*, egesta consumed by shrimp, *Litopenaeus vannamei*. *J. World Aquac. Soc.,* 1-8 (2018)

14 Sauchyn, L. K., Lauzon-Guay, J.-S. & Scheibling, R. E. Sea urchin fecal production and accumulation in a rocky subtidal ecosystem. *Aquatic Biology* **13**, 215-223 (2011).

15 Sauchyn, L. K. & Scheibling, R. E. Degradation of sea urchin feces in a rocky subtidal ecosystem: implications for nutrient cycling and energy flow. *Aquatic Biology* **6**, 99-108 (2009).

16 Schram, J. B., Kobelt, J. N., Dethier, M. N. & Galloway, A. W. Trophic transfer of macroalgal fatty acids in two urchin species: Digestion, egestion, and tissue building. *Front. Ecol. Evol.* **6**, 83 (2018).

17 Lawrence, J. M., Lawrence, A. L. & Watts, S. A. in *Dev. Aquacult. Fish. Sci.* Vol. 38 135-154 (Elsevier, 2013).

18 Ziegler, A., Mooi, R., Rolet, G. & De Ridder, C. Origin and evolutionary plasticity of the gastric caecum in sea urchins (Echinodermata: Echinoidea). *BMC Evol. Biol.* **10**, 313 (2010).

19 Lasker, R. & Giese, A. C. Nutrition of the sea urchin, *Strongylocentrotus purpuratus*. *Biol. Bull.* **106**, 328-340 (1954).

20 Hakim, J. A. *et al.* The Purple Sea Urchin *Strongylocentrotus purpuratus* Demonstrates a Compartmentalization of Gut Bacterial Microbiota, Predictive Functional Attributes, and Taxonomic Co-Occurrence. *Microorganisms* **7**, 35 (2019).

21 Hakim, J. A. *et al.* An abundance of Epsilonproteobacteria revealed in the gut microbiome of the laboratory cultured sea urchin, *Lytechinus variegatus*. *Front. Microbiol.* **6** (2015).

22 Hakim, J. A. *et al.* The gut microbiome of the sea urchin, *Lytechinus variegatus*, from its natural habitat demonstrates selective attributes of microbial taxa and predictive metabolic profiles. *FEMS Microbiol. Ecol.* **92**, fiw146 (2016).

23 Fong, W. & Mann, K. Role of gut flora in the transfer of amino acids through a marine food chain. *Can. J. Fish. Aquat. Sci.* **37**, 88-96 (1980).

24 Meziti, A., Kormas, K. A., Pancucci-Papadopoulou, M.-A. & Thessalou-Legaki, M. Bacterial phylotypes associated with the digestive tract of the

sea urchin *Paracentrotus lividus* and the ascidian *Microcosmus* sp. *Russ. J. Mar. Biol.* **33**, 84-91 (2007).

25 Vadas, R. L. Preferential feeding: an optimization strategy in sea urchins. *Ecological Monographs* **47**, 337-371 (1977).

26 Johannes, R. & Satomi, M. Composition and nutritive value of fecal pellets of a marine crustacean. *Limnol. Oceanogr.* **11**, 191-197 (1966).

27 Koike, I., Mukai, H. & Nojima, S. The role of the sea urchin, *Tripneustes gratilla* (Linnaeus), in decomposition and nutrient cycling in a tropical seagrass bed. *Ecological research* **2**, 19-29 (1987).

28 Wotton, R. S. & Malmqvist, B. Feces in Aquatic Ecosystems: Feeding animals transform organic matter into fecal pellets, which sink or are transported horizontally by currents; these fluxes relocate organic matter in aquatic ecosystems. *AIBS Bulletin* **51**, 537-544 (2001).

29 Fowler, D. *et al.* The global nitrogen cycle in the twenty-first century. *Phil. Trans. R. Soc. B.* **368**, 20130164 (2013).

30 Mulholland, M. R. & Lomas, M. W. in *Nitrogen in the Marine Environment* (eds D.G. Capone, D.A. Bronk, M.R. Mulholland, & E.J. Carpenter) Ch. 7, 303-384 (Academic Press, 2008).

31 Herbert, R. Nitrogen cycling in coastal marine ecosystems. *FEMS Microbiol. Rev.* **23**, 563-590 (1999).

32 Ryther, J. H. & Dunstan, W. M. Nitrogen, phosphorus, and eutrophication in the coastal marine environment. *Science* **171**, 1008-1013 (1971).

33 Kneip, C., Lockhart, P., Voß, C. & Maier, U.-G. Nitrogen fixation in eukaryotes - new models for symbiosis. *BMC Evol. Biol.* **7**, 55 (2007).

34 Fiore, C. L., Jarett, J. K., Olson, N. D. & Lesser, M. P. Nitrogen fixation and nitrogen transformations in marine symbioses. *Trends Microbiol.* **18**, 455-463 (2010).

35 Kumar, R. *et al.* Identification of donor microbe species that colonize and persist long term in the recipient after fecal transplant for recurrent *Clostridium difficile*. *NPJ Biofilms Microbiomes* **3**, 12 (2017).

36 Andrews, S. *FastQC: a quality control tool for high throughput sequence data*, <https://www.bioinformatics.babraham.ac.uk/projects/fastqc> (2010).

37 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357 (2012).

38 Meyer, F. *et al.* The metagenomics RAST server–a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **9**, 386 (2008).

39 Cox, M. P., Peterson, D. A. & Biggs, P. J. SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics* **11**, 485 (2010).

40 O'Leary, N. A. *et al.* Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733-D745 (2015).

41 Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **44**, D457-D462 (2015).

42 Mao, X., Cai, T., Olyarchuk, J. G. & Wei, L. Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics* **21**, 3787-3793 (2005).

43 Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335 (2010).

44 Welch, B. L. The generalization of Student's problem when several different population variances are involved. *Biometrika* **34**, 28-35 (1947).

45 Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B Stat. Methodol.* **57**, 289-300 (1995).

46 Shannon, C. E. A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379 - 423, doi:10.1002/j.1538-7305.1948.tb01338.x (1948).

47 Simpson, E. H. Measurement of diversity. *Nat. Biotechnol.* **163** (1949).

48 Warnes, M. G. R., Bolker, B., Bonebakker, L. & Gentleman, R. Package ‘gplots’. *Various R Programming Tools for Plotting Data* (2016).

49 Bray, J. R. & Curtis, J. T. An ordination of the upland forest communities of southern Wisconsin. *Ecological Monographs* **27**, 325-349 (1957).

50 Neuwirth, E. *RColorBrewer: ColorBrewer palettes. R package version 1.1-2*, <https://cran.r-project.org/web/packages/RColorBrewer/index.html> (2014).

51 Parks, D. H., Tyson, G. W., Hugenholtz, P. & Beiko, R. G. STAMP: statistical analysis of taxonomic and functional profiles. *Bioinformatics* **30**, 3123-3124 (2014).

52 Kanehisa, M., Goto, S., Sato, Y., Furumichi, M. & Tanabe, M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* **40**, D109-D114 (2011).

53 Gao, F., Li, F., Tan, J., Yan, J. & Sun, H. Bacterial community composition in the gut content and ambient sediment of sea cucumber *Apostichopus japonicus* revealed by 16S rRNA gene pyrosequencing. *PLoS One* **9**, e100092 (2014).

54 Plotieau, T., Lavitra, T., Gillan, D. C. & Eeckhaut, I. Bacterial diversity of the sediments transiting through the gut of *Holothuria scabra* (Holothuroidea; Echinodermata). *Mar. Biol.* **160**, 3087-3101 (2013).

55 Jackson, E. W., Pepe-Ranney, C., Debenport, S. J., Buckley, D. H. & Hewson, I. The microbial landscape of sea stars and the anatomical and interspecies variability of their microbiome. *Front. Microbiol.* **9**, 1829 (2018).

56 Morrow, K. M., Tedford, A. R., Pankey, M. S. & Lesser, M. P. A member of the Roseobacter clade, *Octadecabacter* sp., is the dominant symbiont in the brittle star *Amphipholis squamata*. *FEMS Microbiol. Ecol.* **94**, fiy030 (2018).

57 Webster, N. S. & Taylor, M. W. Marine sponges and their microbial symbionts: love and other relationships. *Environment. Microbiol.* **14**, 335-346 (2012).

58 Dishaw, L. J. *et al.* The gut of geographically disparate *Ciona intestinalis* harbors a core microbiota. *PLoS One* **9**, e93386 (2014).

59 Moisander, P. H., Sexton, A. D. & Daley, M. C. Stable associations masked by temporal variability in the marine copepod microbiome. *PLoS One* **10**, e0138967 (2015).

60 Bourne, D. G. *et al.* Coral reef invertebrate microbiomes correlate with the presence of photosymbionts. *ISME J.* **7**, 1452 (2013).

61 Percival, S. L. & Williams, D. W. in *Microbiology of Waterborne Diseases: Microbiological Aspects and Risks* (ed Marylynn V. Yates Steven L. Percival, David W. Williams, Rachel M. Chalmers, Nicholas F. Gray) Ch. 12, 237-248 (Elsevier, 2014).

62 Yumoto, I. *et al.* Characterization of a facultatively psychrophilic bacterium, *Vibrio rumoiensis* sp. nov., that exhibits high catalase activity. *Appl. Environ. Microbiol.* **65**, 67-72 (1999).

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

64 Sawabe, T., Oda, Y., Shiomi, Y. & Ezura, Y. Alginate degradation by bacteria isolated from the gut of sea urchins and abalones. *Microb Ecol.* **30**, 193-202 (1995).

65 Unkles, S. Bacterial flora of the sea urchin *Echinus esculentus*. *Appl. Environ. Microbiol.* **34**, 347-350 (1977).

66 Arts, M. T., Ackman, R. G. & Holub, B. J. "Essential fatty acids" in aquatic ecosystems: a crucial link between diet and human health and evolution. *Can. J. Fish. Aquat. Sci.* **58**, 122-137 (2001).

67 Hosoya, S., Jang, J.-H., Yasumoto-Hirose, M., Matsuda, S. & Kasai, H. *Psychromonas agarivorans* sp. nov., a novel agarolytic bacterium. *Int. J. Syst. Evol. Microbiol.* **59**, 1262-1266 (2009).

68 Monroig, Ó., Tocher, D. & Navarro, J. Biosynthesis of polyunsaturated fatty acids in marine invertebrates: recent advances in molecular mechanisms. *Mar. Drugs* **11**, 3998-4018 (2013).

69 Nichols, D. S. Prokaryotes and the input of polyunsaturated fatty acids to the marine food web. *FEMS Microbiol. Lett.* **219**, 1-7 (2003).

70 Petersen, J. M. *et al.* Chemosynthetic symbionts of marine invertebrate animals are capable of nitrogen fixation. *Nat. Microbiol.* **2**, 16195 (2017).

71 Distel, D. L., Morrill, W., MacLaren-Toussaint, N., Franks, D. & Waterbury, J. *Teredinibacter turnerae* gen. nov., sp. nov., a dinitrogen-fixing, cellulolytic, endosymbiotic gamma-proteobacterium isolated from the gills of wood-boring molluscs (Bivalvia: Teredinidae). *Int. J. Syst. Evol. Microbiol.* **52**, 2261-2269 (2002).

72 Reynolds, L. K., Berg, P. & Zieman, J. C. Lucinid clam influence on the biogeochemistry of the seagrass *Thalassia testudinum* sediments. *Estuaries Coast.* **30**, 482-490 (2007).

73 Guerinot, M. & Patriquin, D. N2-fixing vibrios isolated from the gastrointestinal tract of sea urchins. *Can. J. Microbiol.* **27**, 311-317 (1981).

74 Guerinot, M. L. & Patriquin, D. The association of N2-fixing bacteria with sea urchins. *Mar. Biol.* **62**, 197-207 (1981).

75 Weigel, B. L. & Erwin, P. M. Effects of reciprocal transplantation on the microbiome and putative nitrogen cycling functions of the intertidal sponge, *Hymeniacidon heliophila*. *Sci. Rep.* **7**, 43247 (2017).

76 Atlas, R. & Bartha, R. *Microbial Ecology: Fundamentals and Applications*. 4 edn, (Benjamin cummings publishing company Inc. Addison Wesley Longman Inc, 1998).

77 Kennedy, C. *et al.* Ammonium sensing in nitrogen fixing bacteria: functions of the *glnB* and *glnD* gene products. *Plant and Soil* **161**, 43-57 (1994).

78 Siikavuopio, S. I., Dale, T., Foss, A. & Mortensen, A. Effects of chronic ammonia exposure on gonad growth and survival in green sea urchin *Strongylocentrotus droebachiensis*. *Aquaculture* **242**, 313-320 (2004).

79 Cavanaugh, C. M., McKiness, Z. P., Newton, I. L. & Stewart, F. J. in *The Prokaryotes* Vol. 1 (eds Dworkin M. *et al.*) 475-507 (Springer, 2006).

CHAPTER VI: HIGH-THROUGHPUT AMPLICON SEQUENCING OF THE METACOMMUNITY DNA OF THE GUT MICROBIOTA OF LABORATORY AQUACULTURE AND NATURALLY OCCURRING GREEN SEA URCHINS *LYTECHINUS VARIEGATUS*

by

JOSEPH A. HAKIM, CASEY D. MORROW, STEPHEN A. WATTS, AND ASIM K. BEJ

Submitted to *Data in Brief*

Format adapted and errata corrected for dissertation

ABSTRACT

The high-throughput amplicon sequence (HTS) dataset presented in this article was generated on an Illumina MiSeq sequencing platform by targeting the V4 region of the 16S rRNA gene of the microbial-metacommunity DNA of the gut tissue and the gut digesta of naturally occurring (*n* = 3) and laboratory aquaculture (*n* = 2) green sea urchins, *Lytechinus variegatus*. The HTS dataset was quality checked and filtered, which resulted in 88% sequence reads for downstream analyses. The applicability of these sequences was evaluated by using bioinformatics tools to generate operation taxonomic units (OTUs), which were then used to determine the taxonomic profiles of *L. variegatus* gut ecosystem. The resultant OTU data was verified for saturation by using rarefaction analysis at a 3% sequence variation. Further, the OTUs were randomly subsampled to the minimum sequence count value. Then, the FASTA-formatted representative sequences were assigned taxonomic identities through multiple databases using the SILVA ACT: Alignment, Classification and Tree Service (www.arb-silva.de/aligner). The sequence data can be accessed from https://www.ncbi.nlm.nih.gov/bioproject/ under the BioProject IDs PRJNA291441 and PRJNA326427.

SPECIFICATIONS TABLE*

Subject area	*Biology*
More specific subject area	*Metagenomics*
Type of data	*Figures and Tables*
How data was acquired	*Illumina MiSeq platform with 250 paired-end kits.*
Data format	*Raw, analyzed*
Experimental factors	*Laboratory aquaculture (LAB) Lytechinus variegatus (n=2) were gathered from Port Saint Joseph, Florida (29.80° N 85.36° W), and held in the laboratory aquaculture condition, fed with a formulated diet for 6 months prior to investigation. Naturally occurring (ENV) L. variegatus (n=3) were collected from the same location and sample preparation began immediately upon arrival to the University of Alabama at Birmingham (UAB)*
Experimental features	*Targeted high-throughput sequencing of the microbial metacommunity 16S rRNA gene (V4 hypervariable regions) using the Illumina MiSeq sequencing platform with 250 paired-end kits.*
Data source location	*L. variegatus collected from Port Saint Joseph, Florida, USA (29.80° N 85.36° W) located in the Gulf of Mexico. L. variegatus were maintained in laboratory aquaculture condition at the UAB Biology Department, 1300 University Blvd., Birmingham, AL 35294, USA. Microbial metacommunity DNA was prepared and sequenced at the UAB Department of Genetics, Heflin Center Genomics Core, School of Medicine, the University of Alabama at Birmingham, 705 South 20th Street, Birmingham, AL 35294, USA.*
Data accessibility	*Raw data corresponding to the 10 samples from this study are available at the NCBI's BioSample database following this link: http://www.ncbi.nlm.nih.gov/sra/?term=sea+urchin+gut+microbiome* *For the LAB group, the BioProject number is PRJNA291441 and the BioSample IDs are SAMN03944319, SAMN03944320, SAMN03944321, SAMN03944322. For the ENV group, the BioProject number is PRJNA326427 and the BioSample IDs are SAMN05277844, SAMN05277845, SAMN05277846, SAMN05277847, SAMN05277848, SAMN05277849, and SAMN05277850.*

* The format of the table presented as required by *DIB* publication.

VALUE OF THE DATA

- The HTS datasets would help establish knowledge of the source, distribution, and selection of the microbial community by a diverse species of marine echinoderms, and other marine invertebrates.
- The potential applicability of the HTS datasets presented in this study provides an insight into the modulation of gut microbiota of model organisms fed a standard formulated reference diet in laboratory aquaculture conditions.
- This dataset can be used to compare the gut microbial metabolic processes for the nutritional benefit of other sea urchins, as well as the contribution of macromolecules through the high energy gut digesta in various trophic levels.
- To the best of our knowledge, these datasets provide the first insights into the microbiota of the green sea urchin *L. variegatus* gut ecosystem at a high coverage using the Illumina MiSeq HTS platform.
- Access to the raw files achieved through HTS of the sea urchin microbiome permits researchers to apply their own bioinformatics analyses, based on their exploratory goals.

DATA

The OTUs resulting from the quality-checked and filtered 16S rRNA-targeted HTS dataset was visualized by rarefaction analysis, which indicated that the total quality sequences from each sample are approaching saturation when constructed at a 3% sequence variation (Figure 6-1). The microbial community

composition revealed the *L. variegatus* LAB and ENV gut tissue showed a comparable profile, with a near-exclusive abundance of Epsilonproteobacteria (Figure 6-2). However, the gut digesta between the LAB and ENV showed distinct taxonomic profiles, with a noticeably higher diversity in the ENV group.

EXPERIMENTAL DESIGN, MATERIALS AND METHODS

Sample Description

The sea urchins used in this study were collected from Saint Joseph Bay Aquatic Preserve of the U.S. Gulf of Mexico (29.80° N 85.36° W). For the laboratory aquaculture (LAB) group [1], adult sea urchins (n = 2) were kept in a recirculating saltwater tank system for six months, and fed a formulated feed *ad libitum* once every 24-48 h that consisted of 6% lipid, 28% protein, and 36% carbohydrate relative percentages [2]. The aquaria were maintained at 22 ± 2°C with a pH of 8.2 ± 0.2 and salinity of 32 ± 1 ppt. For the naturally occurring (ENV) group [3], adult sea urchins (n = 3) were collected from within the same 1 m^2 area and transported to the laboratory at the University of Alabama at Birmingham (UAB) for sample collection. Water conditions were recorded as 20 ± 2°C with a pH of 7.8 ± 0.2 and salinity of 28 ± 1 ppt. For both groups, the Illumina MiSeq high throughput-sequencing (HTS) platform was used with the 250 bp paired-end kits targeting the V4 hypervariable region [4, 5]. The paired-end raw sequence data were demultiplexed and formatted into FASTQ files [6]. The raw data were deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under BioProject #PRJNA291441 and

#PRJNA326427 for the LAB and ENV group, respectively. The paired-end sequence data for the gut microbial communities can be accessed under the following NCBI BioSample Ids: SAMN03944319 - SAMN03944322 (LAB group) and SAMN05277845 - SAMN05277850 (ENV group). Subgroups for the laboratory-fed group are as follows: LAB.Gut.Tissue (*n* = 2), LAB.Gut.Digesta (*n* = 2), ENV.Gut.Tissue (*n* = 3), and ENV.Gut.Digesta (*n* = 3).

Quality Assessment and Filtering

The raw and demultiplexed paired-end data was initially assessed by FastQC [7], and only reads showing 80% of bases at a Q score of >33 were retained by using the “fastx_trimmer” command from the FASTX Toolkit [5, 8] and merged using USEARCH [9]. Paired-end reads with <50 base overlap and/or >20 mismatching nucleotides were filtered from the analysis, and chimeric sequences were removed using USEARCH [9].

Taxonomic Distribution and Alpha Diversity

The merged sequence data was analyzed using Quantitative Insights into Microbial Ecology (QIIME; v1.9.1) along with Phylogenetic Tools for Analysis of Species-level Taxa (PhylotoAST; v1.4.0) [10, 11]. The initial Operational Taxonomic Units (OTUs) were clustered at a 97% similarity through UCLUST in QIIME (v1.9.1) [9], and representative sequences were established by the “most_abundant” option. Then, OTUs with < 0.0005% average abundance across all samples were filtered. The redundant OTUs were merged by using the

“condense_workflow.py” command through PhyloToAST (v1.4.0) [11]. The resultant OTUs per sample were plotted against the filtered sequence read counts as rarefaction curves, and the data was subsampled to the minimum value using “single_rarefaction.py” in QIIME (v1.9.1). The representative sequences were then assigned taxonomy using the SILVA ACT: Alignment, Classification and Tree Service (www.arb-silva.de/aligner), which utilizes the SILVA Incremental Aligner (SINA; v1.2.11) to align rRNA gene sequences and classify based on Least Common Ancestor (LCA) methods [12]. For this, the SSU (Small Sub-Unit) option selected at a minimum similarity of 0.7 with 20 neighbors per query sequence, and the databases selected were as follows: SILVA database [13], Ribosomal Database Project (RDP) [14], All-Species Living Tree (LTP) project [15], Greengenes (GG) [16, 17], and European Molecular Biology Laboratory (EMBL) [18]. Biological replicates were validated and merged according to their sub-group assignment based on significant Analysis of Similarity (ANOSIM) [19] and Adonis [20] measurements (p = 0.001) using the weighted Unifrac distances [21] calculated for each sample. The top 25 taxa at the highest resolution from each database were combined and plotted as relative abundance graphs using Microsoft Excel Software (Seattle, WA, USA). The resultant taxonomic data was summarized for each database as a table, to show the total number of OTUs that were assigned a taxonomy, including the proportion that was resolved to the family and genus level.

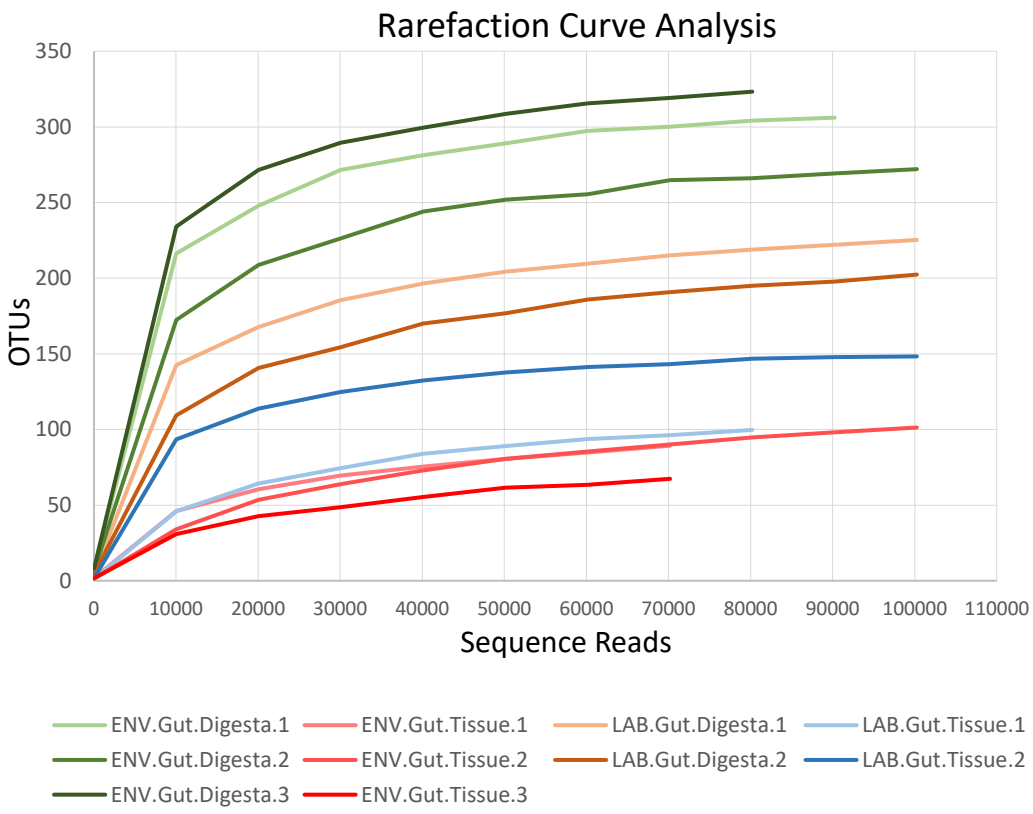


Figure 6-1: Rarefaction curve analysis of the HTS data showing the number of OTUs (Y-axis) plotted against a total number of sequences (X-axis) per sample. OTUs were determined by using the PhyloToAST (v1.4) taxonomy condensing workflow, which is integrated into QIIME (v1.9.1). Samples were rarefied to the minimum sequence count across all samples for downstream bioinformatics analysis. Data were plotted using Microsoft Excel Software (Seattle, WA, USA).

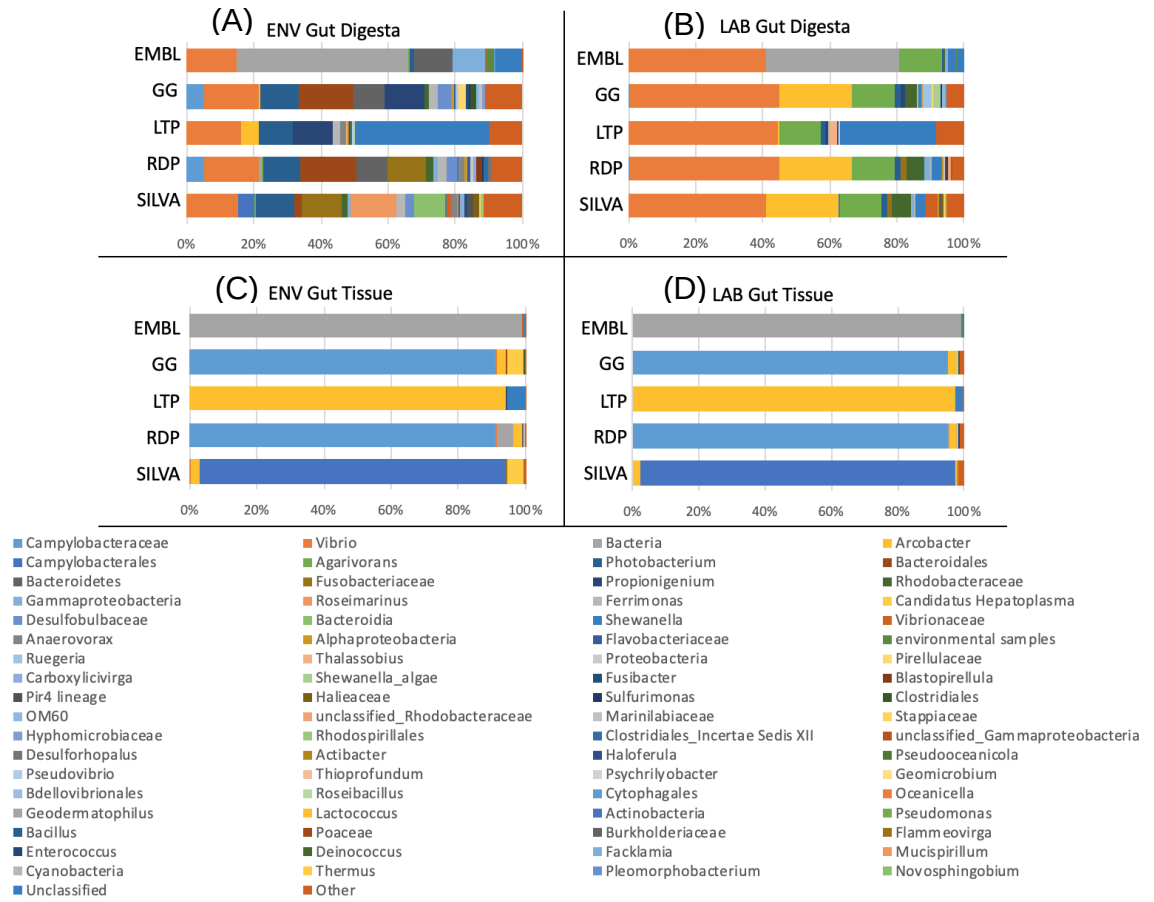


Figure 6-2: Relative abundance distribution of taxa at the highest resolution determined for the merged biological replicates of the study using multiple taxonomic databases. The FASTA-formatted representative sequences determined by the PhyloToAST (v1.4) workflow integrated into QIIME (v1.9.1) were aligned to multiple databases using the SILVA ACT: Alignment, Classification and Tree Service (www.arb-silva.de/aligner). Taxonomic assignments were performed using the SSU (Small Subunit) category and the Least Common Ancestor (LCA) method with the following databases: SILVA, Ribosomal Database Project (RDP), The All-Species Living Tree (LTP) project, Greengenes (GG), and the European Molecular Biology Laboratory (EMBL). Sequences aligned with a similarity threshold below 70% were discarded. The top 25 taxa from each database were merged based on their common taxonomic assignments at the specific level of classification.

Table 6-1: Statistical analysis of the representative sequences aligned to multiple databases using the SILVA ACT: Alignment, Classification and Tree Service (www.arb-silva.de/aligner). Taxonomic assignments were performed using the SSU (Small Subunit) category and the Least Common Ancestor (LCA) method with the following databases: SILVA, Ribosomal Database Project (RDP), The All-Species Living Tree (LTP) project, Greengenes (GG), and the European Molecular Biology Laboratory (EMBL). Sequences aligned with a similarity threshold below 70% were discarded. For each database, the total number of uniquely assigned sequences were determined, and the fraction of those assignments to the family and the genus level were listed.

Level	SILVA	RDP	LTP	GG	EMBL
family	234	193	128	219	18
	82.98%	80.08%	100.00%	76.04%	54.55%
genus	167	147	121	132	12
	59.22%	61.00%	94.53%	45.83%	36.36%
Total Unique	282	241	128	288	33

ACKNOWLEDGMENTS

We would like to thank Dr. Peter Eipers of the Department of Cell, Developmental and Integrative Biology, and Dr. Michael Crowley of the Heflin Center for Genomic Sciences at the University of Alabama at Birmingham (UAB), for their assistance in high-throughput sequencing for this study. Graduate Research Assistant funding to J.A.H. was provided from grant support to C.D.M by the UAB School of Medicine. The following are acknowledged for their support of the Microbiome Resource at the University of Alabama at Birmingham: School of Medicine, Comprehensive Cancer Center (P30AR050948; C.D.M.), Center for AIDS Research (5P30AI027767; C.D.M.), Center for Clinical Translational Science (UL1TR000165; C.D.M.) and Heflin Center. Animal husbandry supported in part by NIH P30DK056336 (S.A.W.).

REFERENCES

1. J.A. Hakim, H. Koo, L.N. Dennis, R. Kumar, T. Ptacek, C.D. Morrow, E.J. Lefkowitz, M.L. Powell, A.K. Bej, S.A. Watts. An abundance of Epsilonproteobacteria revealed in the gut microbiome of the laboratory cultured sea urchin, *Lytechinus variegatus*. Front. Microbiol. 6 (2015) 1047.

2. H. Hammer, B. Hammer, S. Watts, A. Lawrence, J. Lawrence. The effect of dietary protein and carbohydrate concentration on the biochemical composition and gametogenic condition of the sea urchin *Lytechinus variegatus*. J. Exp. Mar. Bio. Ecol. 334(1) (2006) 109-21.

3. J.A. Hakim, H. Koo, R. Kumar, E.J. Lefkowitz, C.D. Morrow, M.L. Powell, S.A. Watts, A.K. Bej. The gut microbiome of the sea urchin, *Lytechinus variegatus*, from its natural habitat demonstrates selective attributes of microbial taxa and predictive metabolic profiles. FEMS Microbiol Ecol. 92(9) (2016) fiw146.

4. J.J. Kozich, S.L. Westcott, N.T. Baxter, S.K. Highlander, P.D. Schloss. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Appl. Environ. Microbiol. 79(17) (2013) 5112-20.

5. R. Kumar, P. Eipers, R.B. Little, M. Crowley, D.K. Crossman, E.J. Lefkowitz, C.D. Morrow. Getting started with microbiome analysis: sample acquisition to bioinformatics. Curr. Protoc. Hum. Genet. 82(1) (2014) 18.8.1–18.8.29.

6. P.J. Cock, C.J. Fields, N. Goto, M.L. Heuer, P.M. Rice. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. Nucleic Acids Res. 38(6) (2009) 1767-71.

7. S. Andrews. FastQC: a quality control tool for high throughput sequence data 2010 [cited 2019]. Available from: https://www.bioinformatics.babraham.ac.uk/projects/fastqc.

8. A. Gordon, G. Hannon. Fastx-toolkit. FASTQ/A short-reads pre-processing tools. 2010 [cited 2019]. Available from: http://hannonlab.cshl.edu/fastx_toolkit.

9. R.C. Edgar. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 26(19) (2010) 2460-1.

10. J.G. Caporaso, J. Kuczynski, J. Stombaugh, K. Bittinger, F.D. Bushman, E.K. Costello, N. Fierer, A.G. Peña, J.K. Goodrich, J.I. Gordon, G.A. Huttley, S.T. Kelley, D. Knights, J.E. Koenig, R.E. Ley, C.A. Lozupone, D. McDonald, B.D. Muegge, M. Pirrung, J. Reeder, J.R. Sevinsky, P.J. Turnbaugh, W.A. Walters, J. Widmann, T. Yatsunenko, J. Zaneveld, R. Knight. QIIME allows analysis of high-throughput community sequencing data. Nat. Methods. 7(5) (2010) 335.

11. S.M. Dabdoub, M.L. Fellows, A.D. Paropkari, M.R. Mason, S.S. Huja, A.A. Tsigarida, P.S. Kumar. PhyloToAST: Bioinformatics tools for species-level analysis and visualization of complex microbial datasets. Sci Rep. 6 (2016) 29123.

12. E. Pruesse, J. Peplies, F.O. Glöckner. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics. 28(14) (2012) 1823-9.

13. C. Quast, E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies, F.O. Glöckner. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 41(D1) (2012) D590-D6.

14. J.R. Cole, Q. Wang, J.A. Fish, B. Chai, D.M. McGarrell, Y. Sun, C.T. Brown, A. Porras-Alfaro, C.R. Kuske, J.M. Tiedje. Ribosomal Database Project: data and tools for high throughput rRNA analysis. Nucleic Acids Res. 42(D1) (2013) D633-D42.

15. P. Yilmaz, L.W. Parfrey, P. Yarza, J. Gerken, E. Pruesse, C. Quast, T. Schweer, J. Peplies, W. Ludwig, F.O. Glöckner. The SILVA and “all-species living tree project (LTP)” taxonomic frameworks. Nucleic Acids Res. 42(D1) (2013) D643-D8.

16. T.Z. DeSantis, P. Hugenholtz, N. Larsen, M. Rojas, E.L. Brodie, K. Keller, T. Huber, D. Dalevi, P. Hu, G.L. Andersen. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl. Environ. Microbiol. 72(7) (2006) 5069-72.

17. D. McDonald, M.N. Price, J. Goodrich, E.P. Nawrocki, T.Z. DeSantis, A. Probst, G.L. Andersen, R. Knight, P. Hugenholtz. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. ISME J. 6(3) (2012) 610.

18. C. Kanz, P. Aldebert, N. Althorpe, W. Baker, A. Baldwin, K. Bates, P. Browne, A. van den Broek, M. Castro, G. Cochrane, K. Duggan, R. Eberhardt, N. Faruque, J. Gamble, F. Garcia Diez, N. Harte, T. Kulikova, Q. Lin, V. Lombard, R. Lopez, R. Mancuso, M. McHale, F. Nardone, V.

Silventoinen, S. Sobhany, P. Stoehr, M.A. Tuli, K. Tzouvara, R. Vaughan, D. Wu, W. Zhu, R. Apweiler. The EMBL nucleotide sequence database. Nucleic Acids Res. 33(suppl_1) (2005) D29-D33.

19. K.R. Clarke. Non-parametric multivariate analyses of changes in community structure. Austral Ecol. 18(1) (1993) 117-43.

20. J. Oksanen, R. Kindt, P. Legendre, B. O'Hara, G.L. Simpson, P. Solymos, M.H.H. Stevens, H Wagner. The vegan package. Community Ecology Package.10 (2007) 631-7.

21. C. Lozupone, M.E. Lladser, D. Knights, J. Stombaugh, R. Knight. UniFrac: an effective distance metric for microbial community comparison. ISME J. 5(2) (2011) 169.

CHAPTER VII: SUMMARY

In this dissertation, the gut microbiota and their metabolic genes in the ecologically and evolutionary significant green *Lytechinus variegatus* and purple *Strongylocentrotus purpuratus* sea urchins were determined by using high-throughput sequencing (HTS) technologies and bioinformatics tools. Through amplicon HTS of the V4 region of the 16S rRNA gene, we were able to show the distribution of the microbiota in the compartmentalized gut ecosystems between these two geographically isolated sea urchins species. Overall, we found that the dominant phyla included Epsilonproteobacteria in the gut tissue and Gammaproteobacteria in the gut digesta, whereas the distribution of genera appeared to be habitat-specific. This compartmentalization also demonstrated an allocation of microbial metabolic qualities linked to energy metabolisms in the gut tissue, whereas the metabolism of the key dietary macromolecules was prevalent in the gut digesta. A comparison of the naturally occurring to laboratory aquaculture green sea urchins elaborated the differences in the gut microbial communities. Moreover, by using shotgun metagenomics we have discovered potential mechanisms, and specifically those involving nitrogen reduction and assimilation into organic compounds performed by the microbiota in the gut system of naturally occurring sea urchins, as they relate to the nutritional benefit to the host, the local food web, and global biogeochemical cycles.

In Chapter 2, we elaborated the microbial community composition and predicted functional attributes in the naturally occurring green sea urchin *L. variegatus* gut ecosystem in the temperate nearshore seagrass meadows of Saint Joseph Bay Aquatic Preserve, Florida. Through 16S rRNA gene (V4 region) amplicon HTS, it was determined that Epsilonproteobacteria identified as *Arcobacter* and *Sulfuricurvum* through BLAST analysis were nearly exclusive in the gut tissue, whereas the gut digesta showed *Vibrio*, *Propionigenium*, *Photobacterium*, and Flavobacteriales to be prevalent. Beta diversity analysis showed the low intra-sample variation of microbial communities across biological replicates, with a distinct cluster pattern of the gut tissue away from all other samples of the study. The gut digesta and egesta subclustered together, indicating retention of gut-assembled microbiota likely continuing their metabolic processes once released into the environment. The pharynx, water, and seagrass samples clustered away from the gut system samples supporting a distinct gut microbial community profile. Predictive functional analysis showed an allocation of microbial community metabolisms in the gut tissue and digesta. Specifically, the gut microbiota revealed a prevalence of the KEGG-Level-2 category of energy metabolism, which includes oxidative phosphorylation, methane, nitrogen, sulfur, and prokaryotic carbon fixation pathways. Conversely, the gut digesta displayed carbohydrate, amino acid, and lipid metabolism pathways as heightened. These functional classifications revealed a potential supportive role of compartmentalization in the green sea urchin host's health through specific allocated community functions.

In Chapter 3, we established the microbial profile of the purple *Strongylocentrotus purpuratus* sea urchin gut ecosystem from its intertidal pool habitat at Cape Arago, Oregon, including its surrounding environment as the habitat-specific source of integration of microbes into the gut. Similar to Chapter 2, this study utilized 16S rRNA gene (V4 region) amplicon HTS and showed Epsilonproteobacteria to be noticeable in the gut tissue and Gammaproteobacteria in the gut digesta. This included *Arcobacter* and *Sulfurimonas* in the gut tissue resembling that of the green sea urchin (*Arcobacter* and *Sulfuricurvum*). In the gut digesta, *Propionigenium*, *Vibrio,* and Flavobacteriales were observed similar to the green sea urchin. However, the psychrophilic *Psychromonas* (Gammaproteobacteria) was a dominant genus in the purple sea urchin digesta, likely due to the cold yearly temperature of that habitat. Beta diversity by cluster analyses elaborated the compartmentalization of microbial communities between the gut tissue and digesta, away from the other samples of the study similar to Chapter 2. Predicted functional profiles indicated the gut tissue to have a preferential abundance of KEGG-Level-2 category of energy metabolism that included nitrogen, methane, oxidative phosphorylation, and carbon fixation pathways. The gut digesta showed amino acid metabolisms that included KEGG-Level-3 categories defined by the KEGG Orthology database as peptidases and amino acid enzymes, and arginine and proline metabolisms, as well as carbohydrate metabolisms that included starch and sucrose metabolism, pentose phosphate pathway, glycolysis/gluconeogenesis, and ubiquinone and another terpenoid-quinone biosynthesis. Such results further

supported the role of the gut digesta microbiota in the digestive physiology of the purple sea urchin as observed in the green sea urchin. Lastly, co-occurrence network analysis of the gut tissue and digesta indicated the theoretical modeling of co-presence and co-exclusion relationships between compartmentalized taxa. Specifically, the key taxa shown included *Arcobacter* in the gut tissue and *Propionigenium* in the gut digesta. Both genera had a high proportion of positive associations with other taxa in the gut system, indicating a positive influence in sustaining the gut microbial community structure. Taxa associated with sulfur reducing bacteria (SRB) were also shown as key taxa in the gut digesta, alluding to a biogeochemical basis for the observed community relationships. These results supported the findings from Chapter 2, which demonstrated a similar trend of a predicted microbial community function in the green sea urchins *L. variegatus*, and supports the hypothesis of a shared functional profile across geographically distinct habitats in the allocated gut environment of the green and purple sea urchins.

In Chapter 4, we compared 16S rRNA gene (V4 region) HTS datasets corresponding to the gut microbial communities of the naturally occurring green sea urchins from Chapter 2 to a laboratory aquaculture green *L. variegatus* sea urchin group established previously in our laboratory (Hakim *et al.*, 2015), using current bioinformatics tools and taxonomic database information. The laboratory aquaculture group was held at defined physiochemical parameters and fed a formulated diet *ad libitum* once every 24-48 h consisting of 6% lipid, 28% protein,

and 36% carbohydrate relative percentages for 6 months (Hammer et al., 2006). We showed that in both datasets, the gut tissue comprised *Arcobacter*, likely performing energy metabolisms in a compartmentalized gut ecosystem irrespective of habitat. Additionally, compositional differences between the gut digesta and similarly the egesta were observed between the two groups. This included a higher proportion of *Vibrio* and a lower overall species richness in the laboratory aquaculture sea urchins, whereas the naturally occurring group showed a more diverse community profile as described in Chapter 2. A comparison of predicted metabolic profiles supported the observed trends in Chapter 2. However, the gut digesta of the naturally occurring group possessed a higher proportion of carbohydrate, amino acid, and lipid metabolisms, suggesting these metabolisms to be crucial for the digestion of their naturally encountered *Thalassia testudinum* food source. The results of this study provided a direct comparison of gut microbial communities of the popular green sea urchin *L. variegatus* model organism when fed a formulated feed in a laboratory aquaculture environment.

In Chapter 5, we applied shotgun metagenomics on the microbial DNA purified from the gut digesta samples of the green and purple sea urchins in Chapters 2 and 3, and determined the taxonomic identities and metabolic functions, including key microbial-driven biochemical cycles as they related to the nutritional benefit of the host, local food web, and global biogeochemistry. We showed that both green and purple sea urchins had a relatively high abundance

of Gammaproteobacteria, with the green sea urchins showing more Alphaproteobacteria. From Gammaproteobacteria, *Vibrio* was a dominant genus in both sea urchins, whereas *Psychromonas* was found only in purple sea urchins. In addition, statistical analysis supported an enrichment of *Roseobacter* and *Ruegeria* in the green sea urchins compared to the purple sea urchins, which showed *Shewanella*, *Photobacterium*, and *Bacteroides* to be prevalent. Functional analysis using the KEGG Orthology database of both sea urchins revealed the broad KEGG-Level-1 category of “metabolism” to be significantly heightened, specifically showing KEGG-Level-2 categories of amino acids (~20%), carbohydrate (~13%), nucleotides (~5%), cofactors and vitamins (~6%), and energy (~5%) metabolic processes to be dominant. Both sea urchins showed genes comprising the genetic pathway for the reduction and fixation of nitrogen to be enriched, including the assimilation into organic compounds such as glutamine and asparagine. This microbial-driven process represents a crucial pathway for the incorporation of biologically accessible nitrogen into amino acids and nucleotides into the host through digestion, and into the environment as it impacts the local community and global nitrogen cycle.

In Chapter 6, we further analyzed the 16S rRNA gene metagenomics data from Chapter 4, to elaborate the microbial community profiles in the gut tissue and gut digesta of the naturally occurring and laboratory aquaculture green sea urchins. Whereas Greengenes, Silva, and RDP were able to assign taxonomic identities to more representative sequence reads, LTP was able to resolve the

highly abundant Epsilonproteobacteria taxon in the gut tissues of both groups to the genus level, specifically *Arcobacter*. Therefore, in addition to providing comprehensive taxonomic profiles of gut microbial communities in this model organism from the natural environment and when used in the laboratory, we also show the benefits of combining multiple taxonomic databases to support the observed taxonomic distribution and potentially provide added phylogenetic resolution.

Conclusion

This dissertation research utilized genomics techniques and bioinformatics tools to gain an insight into the microbiota and their metabolic genes of the uniquely compartmentalized gut ecosystem of sea urchins. Although the bacterial communities were the focus of this dissertation, future studies that include Archaea and Eukaryota may offer a more comprehensive microbial community structure. The outcome of our research will help to expand our knowledge of the role of various factors that shapes gut microbiota, provide nutrition to the host through metabolic processes, their impact in the nearshore marine environment, and co-evolution through geologic time of sea urchins, a deuterostome bilaterian echinoderm with ~450 Ma. evolutionary history.

GENERAL LIST OF REFERENCES

Albright, R. *et al.* Juvenile growth of the tropical sea urchin *Lytechinus variegatus* exposed to near-future ocean acidification scenarios. *J Exp Mar Bio Ecol.* **426**, 12-17 (2012).

Amato, K. R. Co-evolution in context: the importance of studying gut microbiomes in wild animals. *Microbiome Science and Medicine* **1** (2013).

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

Atlas, R. & Bartha, R. *Microbial Ecology: Fundamentals and Applications*. 4 edn, (Benjamin cummings publishing company Inc. Addison Wesley Longman Inc, 1998).

Bäckhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A. & Gordon, J. I. Host-bacterial mutualism in the human intestine. *Science* **307**, 1915-1920 (2005).

Baker, W. *et al.* The EMBL nucleotide sequence database. *Nucleic Acids Res.* **28**, 19-23 (2000).

Balvočiūtė, M. & Huson, D. H. SILVA, RDP, Greengenes, NCBI and OTT - how do these taxonomies compare? *BMC Genomics* **18**, 114 (2017).

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

Bjorndal, K. A. Nutrition and grazing behavior of the green turtle *Chelonia mydas*. *Mar Biol.* **56**, 147-154 (1980).

Blekhman, R. *et al.* Host genetic variation impacts microbiome composition across human body sites. *Genome Biol.* **16**, 191 (2015).

Bodnar, A. G. & Coffman, J. A. Maintenance of somatic tissue regeneration with age in short-and long-lived species of sea urchins. *Aging Cell* **15**, 778-787 (2016).

Brown, N., Eddy, S. & Plaud, S. Utilization of waste from a marine recirculating fish culture system as a feed source for the polychaete worm, *Nereis virens*. *Aquaculture* **322**, 177-183 (2011).

Brune, A., Frenzel, P. & Cypionka, H. Life at the oxic–anoxic interface: microbial activities and adaptations. *FEMS Microbiol. Rev.* **24**, 691-710 (2000).

Cani, P. D. Human gut microbiome: hopes, threats and promises. *Gut* **67**, 1716-1725, doi:10.1136/gutjnl-2018-316723 (2018).

Charlop-Powers, Z., Owen, J. G., Reddy, B. V. B., Ternei, M. A. & Brady, S. F. Chemical-biogeographic survey of secondary metabolism in soil. *Proc Natl Acad Sci.* **111**, 3757-3762 (2014).

Cho, I. & Blaser, M. J. The human microbiome: at the interface of health and disease. *Nat Rev Genet.* **13**, 260 (2012).

Claereboudt, M. & Jangoux, M. Conditions de digestion et activité de quelques polysaccharidases dans le tube digestif de l'oursin Paracentrotus lividus (Echinodermata). *Biochem Syst Ecol.* **13**, 51-54 (1985).

Costello, E. K., Stagaman, K., Dethlefsen, L., Bohannan, B. J. & Relman, D. A. The application of ecological theory toward an understanding of the human microbiome. *Science* **336**, 1255-1262 (2012).

Darwin, C. *On the origin of species by means of natural selection, or preservation of favoured races in the struggle for life.* (J. Murray, London, 1859).

Davidson, E. H. *et al.* Arguments for sequencing the genome of the sea urchin *Strongylocentrotus purpuratus*. *Bethesda, MD, USA: National Human Genome Research Institute* (2002).

Davidson, T. M. & Grupe, B. M. Habitat modification in tidepools by bioeroding sea urchins and implications for fine-scale community structure. *Mar Ecol.* **36**, 185-194 (2015).

de la Calle, F. Marine microbiome as source of natural products. *Microb Biotechnol.* **10**, 1293 (2017).

De Ridder, C., Jangoux, M. & Lawrence, J. in *Echinoderm Nutrition* (eds Michel Jangoux & JM Lawrence) Ch. 10, 213-234 (A.A. Balkema, 1982).

Deming, J. W. & Carpenter, S. D. Factors influencing benthic bacterial abundance, biomass, and activity on the northern continental margin and deep basin of the Gulf of Mexico. *Deep Sea Res Part 2 Top Stud Oceanogr.* **55**, 2597-2606 (2008).

Dethier, M. N. Disturbance and recovery in intertidal pools: maintenance of mosaic patterns. *Ecol Monogr.* **54**, 99-118 (1984).

Di Bernardo, M. & Di Carlo, M. in *Sea Urchin - From Environment to Aquaculture and Biomedicine* (ed Maria Agnello) (*IntechOpen*, 2017).

Dishaw, L. J. *et al.* The gut of geographically disparate *Ciona intestinalis* harbors a core microbiota. *PLoS One* **9**, e93386 (2014).

Distel, D. L., Morrill, W., MacLaren-Toussaint, N., Franks, D. & Waterbury, J. *Teredinibacter turnerae* gen. nov., sp. nov., a dinitrogen-fixing, cellulolytic, endosymbiotic gamma-proteobacterium isolated from the gills of wood-boring molluscs (Bivalvia: Teredinidae). *Int J Syst Evol Microbiol.* **52**, 2261-2269 (2002).

Ebert, T. A., Schroeter, S. C., Dixon, J. D. & Kalvass, P. Settlement patterns of red and purple sea urchins (*Strongylocentrotus franciscanus* and *S. purpuratus*) in California, USA. *Mar Ecol Prog Ser.*, 41-52 (1994).

Ehrlich, P. R. & Raven, P. H. Butterflies and plants: a study in coevolution. *Evolution* **18**, 586-608 (1964).

Ellis, S. L. *et al.* Four regional marine biodiversity studies: approaches and contributions to ecosystem-based management. *PLoS One* **6**, e18997 (2011).

Erwin, P. M., Olson, J. B. & Thacker, R. W. Phylogenetic diversity, host-specificity and community profiling of sponge-associated bacteria in the northern Gulf of Mexico. *PLoS One* **6**, e26806 (2011).

Felder, D. L. & Camp, D. K. *Gulf of Mexico Origin, Waters, and Biota: Biodiversity*. Vol. 1 (Texas A&M University Press, 2009).

Fiore, C. L., Jarett, J. K., Olson, N. D. & Lesser, M. P. Nitrogen fixation and nitrogen transformations in marine symbioses. *Trends Microbiol.* **18**, 455-463 (2010).

Fong, W. & Mann, K. Role of gut flora in the transfer of amino acids through a marine food chain. *Can J Fish Aquat Sci.* **37**, 88-96 (1980).

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

Griffin, D. W. Atmospheric movement of microorganisms in clouds of desert dust and implications for human health. *Clin Microbiol Rev.* **20**, 459-477 (2007).

Guerinot, M. & Patriquin, D. N2-fixing vibrios isolated from the gastrointestinal tract of sea urchins. *Can. J. Microbiol.* **27**, 311-317 (1981a).

Guerinot, M., West, P., Lee, J. & Colwell, R. *Vibrio diazotrophicus* sp. nov., a marine nitrogen-fixing bacterium. *Int J Syst Evol Microbiol.* **32**, 350-357 (1982).

Guerinot, M. L. & Patriquin, D. The association of N2-fixing bacteria with sea urchins. *Mar Biol.* **62**, 197-207 (1981b).

Hakim, J. A. *et al.* An abundance of Epsilonproteobacteria revealed in the gut microbiome of the laboratory cultured sea urchin, *Lytechinus variegatus*. *Front Microbiol.* **6** (2015).

Hammer, H., Hammer, B., Watts, S., Lawrence, A. & Lawrence, J. The effect of dietary protein and carbohydrate concentration on the biochemical composition and gametogenic condition of the sea urchin *Lytechinus variegatus*. *J Exp Mar Bio Ecol.* **334**, 109-121 (2006).

Harris, L. G. & Eddy, S. D. in *Echinoderm Aquaculture* Ch. 1, 3-24 (John Wiley & Sons, Inc 2015).

Heflin, L. E. *et al.* Production and economic optimization of dietary protein and carbohydrate in the culture of Juvenile Sea Urchin *Lytechinus variegatus*. *Aquaculture* **463**, 51-60 (2016).

Herbert, R. Nitrogen cycling in coastal marine ecosystems. *FEMS Microbiol. Rev.* **23**, 563-590 (1999).

Holland, N. D. in *Dev Aquacult Fish Sci.* Vol. 38 119-133 (Elsevier, 2013).

Holland, N. D. & Ghiselin, M. T. A comparative study of gut mucous cells in thirty-seven species of the class Echinoidea (Echinodermata). *Biol Bull.* **138**, 286-305 (1970).

Jensen, K. E. *et al.* The value of sea urchin, *Lytechinus variegatus*, egesta consumed by shrimp, *Litopenaeus vannamei*. *J World Aquac Soc.* (2018).

Jousset, A. *et al.* Where less may be more: how the rare biosphere pulls ecosystems strings. *ISME J.* **11**, 853-862 (2017).

Kallmeyer, J., Pockalny, R., Adhikari, R. R., Smith, D. C. & D'Hondt, S. Global distribution of microbial abundance and biomass in subseafloor sediment. *Proc Natl Acad Sci.* **109**, 16213-16216 (2012).

Keegan, K. P., Glass, E. M. & Meyer, F. in *Microbial Environmental Genomics (MEG)* 207-233 (Springer, 2016).

Kellogg, C. A., Lisle, J. T. & Galkiewicz, J. P. Culture-independent characterization of bacterial communities associated with the cold-water coral *Lophelia pertusa* in the northeastern Gulf of Mexico. *Appl Environ Microbiol.* **75**, 2294-2303 (2009).

Klinger, T., Hsieh, H., Pangallo, R., Chen, C. & Lawrence, J. M. The effect of temperature on feeding, digestion, and absorption of *Lytechinus variegatus* (Lamarck)(Echinodermata: Echinoidea). *Physiol Zool.* **59**, 332-336 (1986).

Klinger, T. S. Activities and kinetics of digestive α-and β-glucosidase and β-galactosidase of five species of echinoids (Echinodermata). *Comp Biochem Physiol A Physiol.* **78**, 597-600 (1984).

Kneip, C., Lockhart, P., Voß, C. & Maier, U.-G. Nitrogen fixation in eukaryotes - new models for symbiosis. *BMC Evol. Biol.* **7**, 55 (2007).

Knoll, A. H. Paleobiological perspectives on early microbial evolution. *Cold Spring Harb Perspect Biol.* **7**, a018093 (2015).

Koo, H., Mojib, N., Thacker, R. W. & Bej, A. K. Comparative analysis of bacterial community-metagenomics in coastal Gulf of Mexico sediment microcosms following exposure to Macondo oil (MC252). *Antonie Van Leeuwenhoek* **106**, 993-1009 (2014).

Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K. & Schloss, P. D. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol.* **79**, 5112-5120 (2013).

Kumar, R. *et al.* Getting started with microbiome analysis: sample acquisition to bioinformatics. *Curr Protoc Hum Genet.*, 18.18. 11-18.18. 29 (2014).

Langille, M. G. *et al.* Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotechnol.* **31**, 814-821 (2013).

Lasker, R. & Giese, A. C. Nutrition of the sea urchin, *Strongylocentrotus purpuratus*. *Biol Bull.* **106**, 328-340 (1954).

Lawrence, J. M., Lawrence, A. L. & Watts, S. A. in *Dev Aquacult Fish Sci.* Vol. 38 135-154 (Elsevier, 2013).

Ley, R. E. *et al.* Evolution of mammals and their gut microbes. *Science* **320**, 1647-1651 (2008).

Limborg, M. & Heeb, P. Special Issue: Coevolution of Hosts and Their Microbiome. *MDPI Genes* (2018).

Locey, K. J. & Lennon, J. T. Scaling laws predict global microbial diversity. *Proc Natl Acad Sci.* **113**, 5970-5975 (2016).

Mamelona, J. & Pelletier, É. Green urchin as a significant source of fecal particulate organic matter within nearshore benthic ecosystems. *J Exp Mar Bio Ecol.* **314**, 163-174 (2005).

Martínez, I. *et al.* The gut microbiota of rural papua new guineans: composition, diversity patterns, and ecological processes. *Cell Rep.* **11**, 527-538 (2015).

McBride, S. C. in *Am Fish Soc Symp.* 179-208 (Publication of the University of California Sea Grant Extension Program).

McClay, D. R. Evolutionary crossroads in developmental biology: sea urchins. *Development* **138**, 2639-2648 (2011).

McDonald, D. *et al.* An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J.* **6**, 610 (2012).

Metaxas, A. & Scheibling, R. E. Community structure and organization of tidepools. *Mar Ecol Prog Ser.*, 187-198 (1993).

Meziti, A., Kormas, K. A., Pancucci-Papadopoulou, M.-A. & Thessalou-Legaki, M. Bacterial phylotypes associated with the digestive tract of the sea urchin *Paracentrotus lividus* and the ascidian *Microcosmus* sp. *Russ J Mar Biol.* **33**, 84-91 (2007).

Morrison, J. M. *et al.* Microbial communities mediating algal detritus turnover under anaerobic conditions. *PeerJ* **5**, e2803 (2017).

Mulholland, M. R. & Lomas, M. W. in *Nitrogen in the Marine Environment* (eds D.G. Capone, D.A. Bronk, M.R. Mulholland, & E.J. Carpenter) Ch. 7, 303-384 (Academic Press, 2008).

Nisbet, E. & Sleep, N. The habitat and nature of early life. *Nature* **409**, 1083 (2001).

Pearse, J. & Cameron, R. in *Reproduction of Marine Invertebrates: Echinoderms and Lophophorates* Vol. 6 (eds AC Giese, JS Pearse, & VB Pearse) 513–662 (Boxwood Press, 1991).

Pearse, J. S. Ecological role of purple sea urchins. *Science* **314**, 940-941 (2006).

Pruesse, E. *et al.* SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* **35**, 7188-7196 (2007).

Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J. & Segata, N. Shotgun metagenomics, from sampling to analysis. *Nat Biotechnol.* **35**, nbt. 3935 (2017).

Reinhold-Hurek, B., Bünger, W., Burbano, C. S., Sabale, M. & Hurek, T. Roots shaping their microbiome: global hotspots for microbial activity. *Annu Rev Phytopathol.* **53**, 403-424 (2015).

Reveillaud, J. *et al.* Host-specificity among abundant and rare taxa in the sponge microbiome. *ISME J.* **8**, 1198-1209 (2014).

Reynolds, L. K., Berg, P. & Zieman, J. C. Lucinid clam influence on the biogeochemistry of the seagrass *Thalassia testudinum* sediments. *Estuaries Coast.* **30**, 482-490 (2007).

Rodríguez, J. M. *et al.* The composition of the gut microbiota throughout life, with an emphasis on early life. *Microb Ecol Health Dis.* **26**, 26050 (2015).

Rosenberg, E. & Zilber-Rosenberg, I. The hologenome concept of evolution after 10 years. *Microbiome* **6**, 78 (2018).

Rosshart, S. P. *et al.* Wild mouse gut microbiota promotes host fitness and improves disease resistance. *Cell* **171**, 1015-1028. e1013 (2017).

Ruppert, E. E. & Barnes, R. D. *Invertebrate Zoology*. 6 edn, Vol. 1 (Philadelphia: Saunders College Publishing, 1994).

Ryther, J. H. & Dunstan, W. M. Nitrogen, phosphorus, and eutrophication in the coastal marine environment. *Science* **171**, 1008-1013 (1971).

Sauchyn, L. K., Lauzon-Guay, J.-S. & Scheibling, R. E. Sea urchin fecal production and accumulation in a rocky subtidal ecosystem. *Aquat Biol.* **13**, 215-223 (2011).

Sawabe, T., Oda, Y., Shiomi, Y. & Ezura, Y. Alginate degradation by bacteria isolated from the gut of sea urchins and abalones. *Microb Ecol.* **30**, 193-202 (1995).

Schram, J. B., Kobelt, J. N., Dethier, M. N. & Galloway, A. W. E. Trophic Transfer of Macroalgal Fatty Acids in Two Urchin Species: Digestion, Egestion, and Tissue Building. *Front Ecol Evol.* **6**, doi:10.3389/fevo.2018.00083 (2018).

Shade, A. & Handelsman, J. Beyond the Venn diagram: the hunt for a core microbiome. *Environ Microbiol.* **14**, 4-12 (2012).

Sharpton, T. J. An introduction to the analysis of shotgun metagenomic data. *Front Plant Sci.* **5**, 209 (2014).

Siikavuopio, S. I., Dale, T., Foss, A. & Mortensen, A. Effects of chronic ammonia exposure on gonad growth and survival in green sea urchin *Strongylocentrotus droebachiensis*. *Aquaculture* **242**, 313-320 (2004).

Skoog, A., Biddanda, B. & Benner, R. Bacterial utilization of dissolved glucose in the upper water column of the Gulf of Mexico. *Limnol Oceanogr.* **44**, 1625-1633 (1999).

Sodergren, E. *et al.* The genome of the sea urchin *Strongylocentrotus purpuratus*. *Science* **314**, 941-952 (2006).

Söllinger, A. *et al.* Holistic Assessment of Rumen Microbiome Dynamics through Quantitative Metatranscriptomics Reveals Multifunctional Redundancy during Key Steps of Anaerobic Feed Degradation. *MSystems* **3**, e00038-00018 (2018).

Sterner, R. W. & Hessen, D. O. Algal nutrient limitation and the nutrition of aquatic herbivores. *Annu Rev Ecol Syst.* **25**, 1-29 (1994).

Tarnecki, A. M., Burgos, F. A., Ray, C. L. & Arias, C. R. Fish Intestinal Microbiome: Diversity and Symbiosis Unraveled by Metagenomics. *J Appl Microbiol.* (2017).

Tegner, M. & Dayton, P. Ecosystem effects of fishing in kelp forest communities. *ICES J Mar Sci.* **57**, 579-589 (2000).

Thorsen, M. S. Microbial activity, oxygen status and fermentation in the gut of the irregular sea urchin *Echinocardium cordatum* (Spatangoida: Echinodermata). *Mar Biol.* **132**, 423-433 (1998).

Troussellier, M., Escalas, A., Bouvier, T. & Mouillot, D. Sustaining rare marine microorganisms: macroorganisms as repositories and dispersal agents of microbial diversity. *Front Microbiol.* **8**, 947 (2017).

Unkles, S. Bacterial flora of the sea urchin *Echinus esculentus*. *Appl Environ Microbiol.* **34**, 347-350 (1977).

Vellend, M. Conceptual synthesis in community ecology. *Q Rev Biol.* **85**, 183-206 (2010).

Wang, B., Yao, M., Lv, L., Ling, Z. & Li, L. The human microbiota in health and disease. *Engineering* **3**, 71-82 (2017).

Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol.* **73**, 5261-5267 (2007).

Watanabe, J. M. & Harrold, C. Destructive grazing by sea urchins *Strongylocentrotus* spp. in a central California kelp forest: potential roles of recruitment, depth, and predation. *Mar Ecol Prog Ser.*, 125-141 (1991).

Weigel, B. L. & Erwin, P. M. Effects of reciprocal transplantation on the microbiome and putative nitrogen cycling functions of the intertidal sponge, *Hymeniacidon heliophila*. *Sci Rep.* **7**, 43247 (2017).

Welsh, D. T. Nitrogen fixation in seagrass meadows: regulation, plant–bacteria interactions and significance to primary productivity. *Ecol Lett.* **3**, 58-71 (2000).

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

Woese, C. R. Bacterial evolution. *Microbiol Rev.* **51**, 221 (1987).

Wotton, R. S. & Malmqvist, B. Feces in Aquatic Ecosystems: Feeding animals transform organic matter into fecal pellets, which sink or are transported horizontally by currents; these fluxes relocate organic matter in aquatic ecosystems. *AIBS Bulletin* **51**, 537-544 (2001).

Yagi, J. M. *et al.* Subsurface cycling of nitrogen and anaerobic aromatic hydrocarbon biodegradation revealed by nucleic acid and metabolic biomarkers. *Appl Environ Microbiol.* **76**, 3124-3134 (2010).

Yilmaz, P. *et al.* The SILVA and “all-species living tree project (LTP)” taxonomic frameworks. *Nucleic Acids Res.* **42**, D643-D648 (2013).

Yokoyama, H. Growth and food source of the sea cucumber *Apostichopus japonicus* cultured below fish cages - Potential for integrated multi-trophic aquaculture. *Aquaculture* **372–375**, 28-38, doi:https://doi.org/10.1016/j.aquaculture.2012.10.022 (2013).

Zhadan, P. M., Vaschenko, M. A. & Almyashova, T. N. in *Sea Urchin - From Environment to Aquaculture and Biomedicine* (ed Maria Agnello) (*IntechOpen*, 2017).

Zhang, Q. *et al.* Novel genes dramatically alter regulatory network topology in amphioxus. *Genome Biol.* **9**, R123 (2008).

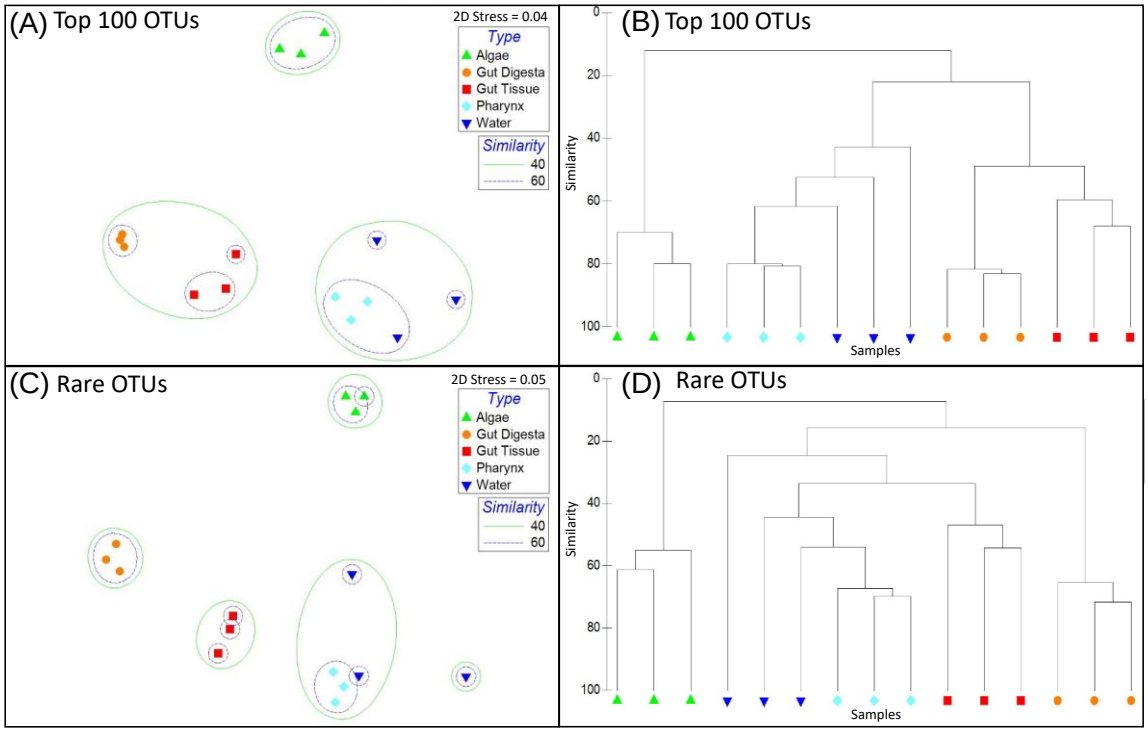
Ziegler, A., Mooi, R., Rolet, G. & De Ridder, C. Origin and evolutionary plasticity of the gastric caecum in sea urchins (Echinodermata: Echinoidea). *BMC Evol. Biol.* **10**, 313 (2010).

Zilber-Rosenberg, I. & Rosenberg, E. Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. *FEMS Microbiol. Rev.* **32**, 723-735 (2008).

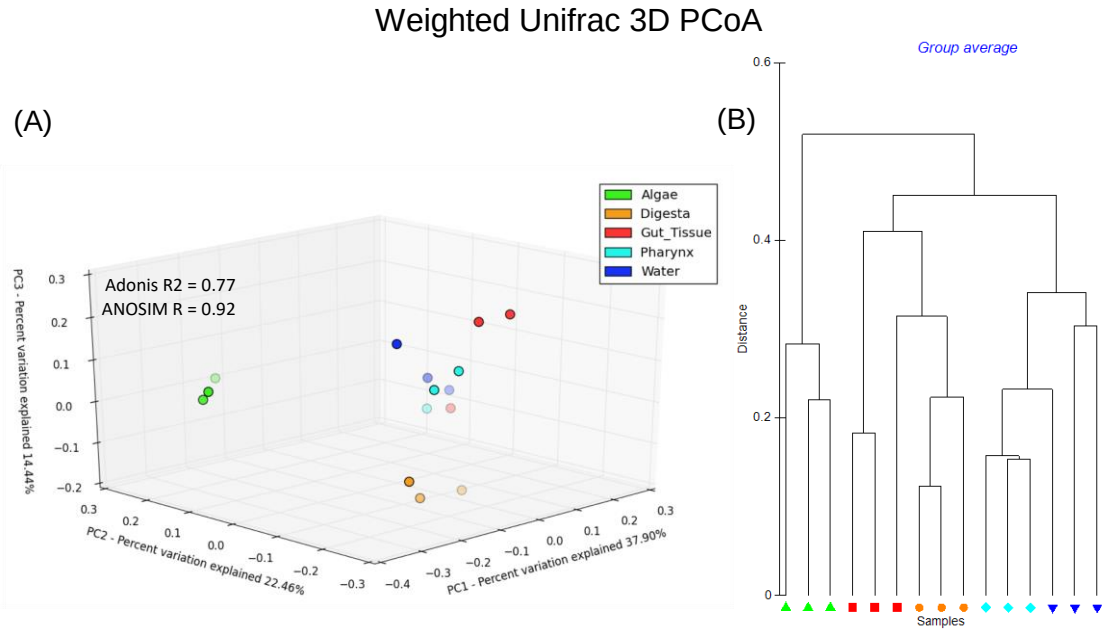
APPENDIX A

SUPPLEMENTARY MATERIALS FOR “CHAPTER III: THE PURPLE SEA URCHIN *STRONGYLOCENTROTUS PURPURATUS* DEMONSTRATES A COMPARTMENTALIZATION OF GUT BACTERIAL MICROBIOTA, PREDICTIVE FUNCTIONAL ATTRIBUTES, AND TAXONOMIC CO-OCCURRENCE”

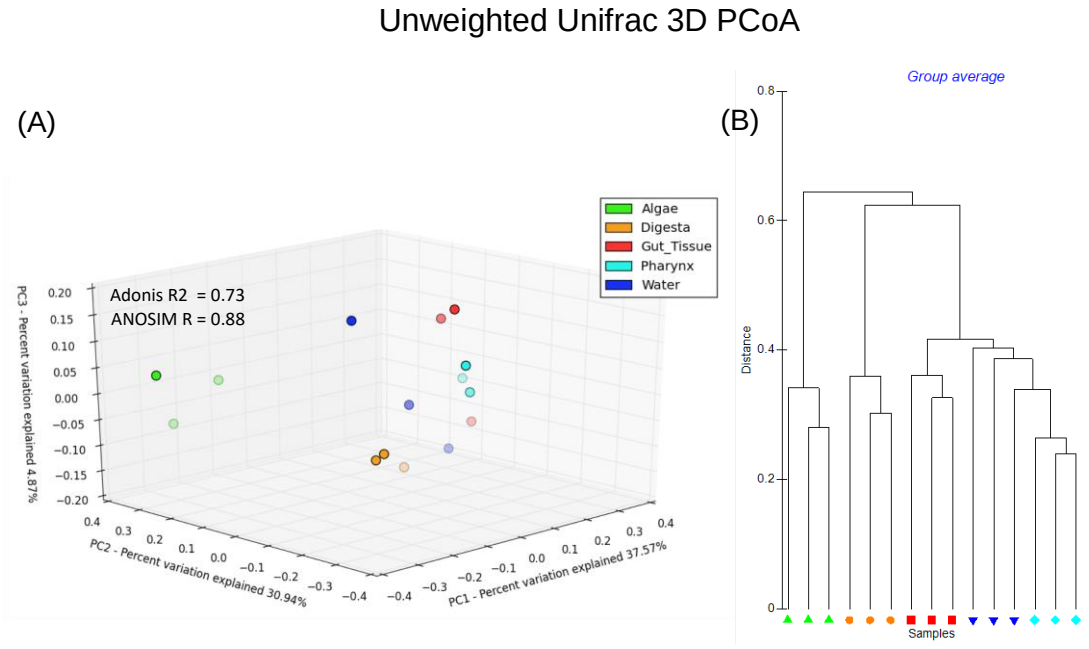
Supplementary Figure 3-S1: 2D multidimensional scale (MDS) plot and dendrogram analysis performed using the Bray-Curtis metrics calculated for the top 100 and rare taxa across all samples of the study. (A) 2D MDS and (B) dendrogram analysis of the top 100 OTUs, and (C) 2D MDS and (D) dendrogram analysis of the rare OTUs (non-top 100) determined across all samples of the study. The rarefied OTU table was pre-treated prior to calculating the Bray-Curtis similarity, by standardizing each OTU by the total sequence count per sample (to reduce the large disparities between samples), and then by log transforming the data transform to lessen the dominant contribution of highly abundant OTUs. The overlay of similarity observed in the 2D MDS plots is shown as 40% and 60% Bray-Curtis similarity. Figure legends are shown in the 2D MDS plots, including the 2D Stress values. Data was generated and plotted through PRIMER-6 software (Primer-E Ltd, Plymouth Marine Laboratory, Plymouth UK, v6.1.2)



Supplementary Figure 3-S2: 3D principle coordinates analysis (PCoA) plot and dendrogram analysis performed using the weighted Unifrac metrics calculated for the rarefied OTU table data across all samples of the study. (A) The weighted Unifrac distances were determined through QIIME (v1.9.1) using the "beta_diversity_through_plots.py" module, and these values were used to construct the 3D PCoA plot using the "PCoA.py" command in PhyloToAST (v1.4.0). The percent variation explained is listed on the X, Y, and Z axes, and the ANOSIM and Adonis measures were calculated between groups using the "compare_categories.py" command in QIIME (v1.9.1) which are listed in the plot. (B) The Unifrac distance values were uploaded into PRIMER-6 (Primer-E Ltd, Plymouth Marine Laboratory, Plymouth UK, v6.1.2), and used to generate the dendrogram using group average clustering. Figure legend corresponding to each sample type is listed to the top right of the PCoA plot.



Supplementary Figure 3-S3: 3D principle coordinates analysis (PCoA) plot and dendrogram analysis performed using the unweighted Unifrac metrics calculated for the rarefied OTU table data across all samples of the study. (A) The unweighted Unifrac distances were determined through QIIME (v1.9.1) using the “beta_diversity_through_plots.py” module, and these values were used to construct the 3D PCoA plot using the “PCoA.py” command in PhyloToAST (v1.4.0). The percent variation explained is listed on the X, Y, and Z axes, and the ANOSIM and Adonis measures were calculated between groups using the “compare_categories.py” command in QIIME (v1.9.1) which are listed in the plot. (B) The Unifrac distance values were uploaded into PRIMER-6 (Primer-E Ltd, Plymouth Marine Laboratory, Plymouth UK, v6.1.2), and used to generate the dendrogram using group average clustering. Figure legend corresponding to each sample type is listed to the top right of the PCoA plot.



Supplementary Table 3-S1: Top 100 representative sequences aligned to multiple databases. For each of the top 100 taxa determined in this study, the representative sequence was extracted, and used to achieve a more comprehensive taxonomic profile using the SILVA ACT: Alignment, Classification and Tree Service (www.arb-silva.de/aligner). For the parameters, the SSU (Small Sub-Unit) category was selected, the minimum identity was set to 0.9, with 20 neighbors per query sequence. Sequences below a threshold of 70% were discarded. The Least Common Ancestor (LCA) method was used for taxonomic identification against the GreenGenes, Ribosomal Database Project (RDP), and SILVA databases. Classification parameters were kept at the default, which included a minimum kmer length of 10, LCA-quorum of 0.8, and kmer search candidates set to 1000. The “search-no-fast” method was also selected. Shown in the table is the taxonomic identity assigned in this study through the described methods, alignment score, base pair score, percent identity, and quality score. This is followed by the taxonomic identities as determined through GreenGenes, RDP, and SILVA databases by the LCA method.

Assigned Taxonomic (GreenGenes)	Seq Score	BP Score	Ident	qual	LCA GreenGenes	LCA RDP	LCA SILVA
Psychromonas	0.97	71	98	96	Psychromonas	Psychromonas	Psychromonas
Tissierella Soehngenia	0.95	65	92.9	95	Tissierellaceae	Tissierella	Family XI
Flavobacteriales	0.96	65	92.5	95	Bacteroidales	Bacteroidales	Roseimarinus
Propionigenium	0.95	69	95.6	94	Propionigenium	"Fusobacteriaceae"	Fusobacteriaceae
Arcobacter	0.93	66	92.1	93	Arcobacter	Arcobacter	Arcobacter
Gammaproteobacteria	0.98	72	98.4	98	Gammaproteobacteria	Gammaproteobacteria	Gammaproteobacteria
Fusobacterium	0.97	70	96	96	Fusobacterium	Fusobacterium	Fusobacterium
Sulfurimonas	0.95	67	95.7	95	Sulfurimonas	Sulfurimonas	Sulfurimonas
Saprospiraceae	0.97	71	87.4	96	Saprospiraceae	"Saprospiraceae"	Saprospiraceae

Prevotella	0.99	69	90.9	99	Prevotella	Alloprevotella	Alloprevotella
Rhodophyta	0.97	69	88.9	96	Rhodophyta	Unclassified Chloroplast	Chloroplast
Desulfotalea	0.95	68	94.1	94	Desulfobulbaceae	Desulfobulbaceae	Desulfobulbaceae
Stramenopiles	0.98	66	97.6	98	Stramenopiles	Bacillariophyta	Chloroplast
Clostridiales	0.92	70	88.9	92	Unclassified	Unclassified	Ruminococcaceae UCG-013
Bacteroidetes	0.98	66	89.8	97	Bacteroidales	"Bacteroidales"	Bacteroidales
S24-7	0.99	71	94.8	99	S24-7	Unclassified "Porphyromonadaceae"	Muribaculaceae
Flavobacteriaceae	0.98	68	97.2	97	Flavobacteriaceae	Flavobacteriaceae	Flavobacteriaceae
Gemellaceae	0.99	65	93.3	99	Streptococcus	Bacilli	Bacilli
Maribacter	0.99	70	98.8	98	Maribacter	Maribacter	Flavobacteriaceae
Octadecabacter	0.99	69	96.4	98	Rhodobacteraceae	Rhodobacteraceae	Rhodobacteraceae uncultured
Bacteroidales	0.91	68	90.1	91	Bacteroidales	"Bacteroidetes"	Bacteroidia
Mycobacterium	0	70	100	100	Mycobacterium	Mycobacterium	Mycobacterium
Alphaproteobacteria	0.98	72	88.5	97	Unclassified	Unclassified	Unclassified
Streptococcus	0.96	67	98	96	Streptococcus	Streptococcus	Streptococcus
Lachnospiraceae	0.95	64	93.7	94	Clostridiales	Unclassified Lachnospiraceae	Lachnospiraceae
Neisseriasubflava	0	70	100	100	Neisseria	Neisseria	Neisseria
Acinetobacter	0.99	72	99.6	98	Acinetobacter	Acinetobacter	Acinetobacter

mitochondria	0.89	54	76	88	Unclassified	Unclassified	Unclassified
Lactobacillus	0.96	67	97.2	95	Lactobacillus	Lactobacillus	Lactobacillus
Veillonella dispar	0.96	70	97.2	96	Veillonella dispar	Veillonella	Veillonella
Leucothrix	0.97	72	97.6	96	Leucothrix	Gammaproteobacteria	Leucothrix
Hyphomonadaceae	0.99	69	95.7	99	Hyphomonadaceae	Hyphomonadaceae	Hyphomonadaceae
Delftia	0	72	100	100	Delftia	Delftia	Delftia
Rhodobacteraceae	0.97	68	98.4	97	Rhodobacteraceae	Rhodobacteraceae	Rhodobacteraceae
Campylobacteraceae	0.95	66	89.4	95	Arcobacter	Arcobacter	Campylobacterales
Vibrionaceae	0.95	68	96.8	95	Vibrionaceae	Vibrio	Vibrio
Rikenellaceae	0.96	68	92.5	95	Rikenellaceae	Alistipes	Alistipes
Pseudomonas	0.99	72	98	99	Pseudomonas	Pseudomonas	Pseudomonas
Akkermansia muciniphila	0.97	71	94.1	97	muciniphila	Akkermansia	Akkermansia
BD7-3	0.98	71	87	98	BD7-3	Unclassified Alphaproteobacteria	Micavibrionales
Thiohalorhabdales	0.99	72	96.4	98	Thiohalorhabdales	Arenicellaceae	Arenicellaceae
Janthinobacterium	0	71	100	100	Oxalobacteraceae	Massilia	Massilia
Vibrio	0.98	69	98	97	Vibrio	Vibrio	Vibrio
Bacteroidesovatus	0.98	69	98	98	Bacteroides	Bacteroides	Bacteroides
Actinobacillus	0.97	70	98.4	97	Actinobacillus parahaemolyticus	Unclassified Pasteurellaceae	Actinobacillus
Leptotrichia	0.98	69	91.7	97	Leptotrichia	Leptotrichia	Leptotrichia

JdFBGBact	0.97	71	91.7	97	JdFBGBact	Iamiaceae	uncultured
Bacteroides	0.98	67	94.5	97	Bacteroides	Bacteroides	Bacteroides
Proteobacteria	0.99	69	90.2	99	Desulfocapsa	Desulfocapsa	Desulfocapsa
Verrucomicrobiaceae	0.98	68	92.1	98	Verrucomicrobiaceae	Verrucomicrobiaceae	Roseibacillus
Lactobacillus reuteri	0.98	70	99.2	98	Lactobacillus	Lactobacillus	Lactobacillus
Trueperaceae	0.92	65	84.6	92	Unclassified	Bacteria	Truepera
SB-1	0.99	70	96.8	98	SB-1	Prolixibacteraceae	Prolixibacteraceae
Oscillospira	0.99	73	95.7	99	Oscillospira	Unclassified Ruminococcaceae	Ruminiclostridium 9
Cocleimonas	0.99	71	96	99	Cocleimonas	Cocleimonas	Thiotrichaceae
Thalassomonas sediminis	0.98	72	96.1	97	Colwelliaceae	Colwelliaceae	Colwelliaceae
Blautia	0.96	69	96.4	95	Blautia	Blautia	Blautia
Enterobacteriaceae	0.88	65	91.7	88	Enterobacteriaceae	Enterobacteriaceae	Enterobacteriaceae
Staphylococcus	0	70	100	100	Staphylococcus	Staphylococcus	Staphylococcus
Comamonadaceae	0.98	73	99.2	98	Limnohabitans	Comamonadaceae	Burkholderiaceae
Campylobacter	0.98	67	99.2	98	Campylobacter	Campylobacter	Campylobacter
Allobaculum	0.98	71	90.9	97	Allobaculum	Erysipelotrichaceae	Dubosiella
Bifidobacterium	0.96	73	98	96	Bifidobacterium	Bifidobacterium	Bifidobacterium
Porphyromonas	0.95	64	96.4	94	Porphyromonas	Porphyromonas	Porphyromonas
Lactobacillus iners	0.95	67	96.8	95	Lactobacillus iners	Lactobacillus	Lactobacillus

Rubritalea	0.96	69	94.5	95	Rubritalea	Rubritalea	Rubritalea
Fusibacter	0.98	67	95.2	98	Fusibacter	Fusibacter	Fusibacter
Rheinheimera	0.96	73	94.9	95	Rheinheimera	Rheinheimera	Rheinheimera
NS11-12	0.98	64	88.1	97	NS11-12	"Bacteroidetes"	NS11-12 marine group
Chitinophagaceae	0.97	68	92.1	97	Chitinophagaceae	Unclassified Chitinophagaceae	uncultured
HTCC2089	0.96	73	94.9	96	HTCC2089	Unclassified Gammaproteobacteria	KI89A clade
Aggregatibacter	0.99	70	99.6	99	Aggregatibacter	Pasteurellaceae	Aggregatibacter
Faecalibacterium prausnitzii	0	71	100	100	Ruminococcaceae	Ruminococcaceae	Subdoligranulum
Granulicatella	0	71	100	100	Granulicatella	Granulicatella	Granulicatella
Helleabalneolensis	0.99	67	97.6	99	Hyphomonadaceae	Hyphomonadaceae	Hyphomonadaceae
Actinomyces	0	71	100	100	Actinomyces	Actinomyces	Actinomyces
Simkaniaceae	0.99	65	91.7	99	Simkaniaceae	Simkania	Bacteria
Selenomonas	0.96	68	97.6	96	Selenomonas	Selenomonas	Selenomonas 3
Brevibacterium aureum	0	71	100	100	Brevibacterium	Brevibacterium	Brevibacterium
MSBL3	0.97	71	87.7	96	MSBL3	Unclassified Bacteria	MSBL3
Oceanospirillales	0.96	74	90.9	95	Gammaproteobacteria	Gammaproteobacteria	Gammaproteobacteria
SR1	0.99	71	95.7	99	SR1	SR1_incertae_sedis	Absconditabacteriales (SR1)
Turicibacter	0.95	62	95.3	95	Lachnospiraceae	Firmicutes	Firmicutes

Ulvibacter	0.99	67	98.4	99	Ulvibacter	Ulvibacter	Ulvibacter
SC3-41	0.99	71	94.9	98	SC3-41	Unclassified Iamiaceae	Sva0996 marine group
Loktanella	0.99	70	99.2	98	Rhodobacteraceae	Roseovarius	Ascidiaceihabitans
Agrobacterium	0.98	72	99.2	98	Agrobacterium	Rhizobium	Rhizobium
Odoribacter	0.97	65	87.7	97	Odoribacter	Odoribacter	Odoribacter
Capnocytophaga	0	68	100	100	Capnocytophaga	Capnocytophaga	Capnocytophaga
Ruminococcus	0.95	73	92.1	94	Ruminococcaceae	Ruminococcaceae	Ruminococcaceae uncultured
Oxalobacteraceae	0.97	71	97.6	97	Oxalobacteraceae	Oxalobacteraceae	Noviherbaspirillum
Mycoplana	0	70	100	100	Caulobacteraceae	Brevundimonas	Brevundimonas
Desulfobacteraceae	0.97	72	92.9	96	Desulfobacteraceae	Desulfobacteraceae	Desulfobacteraceae
Bacteriovoracaceae	0.97	71	94.1	96	Bacteriovoracaceae	Bacteriovoracaceae	Peredibacter
Streptophyta	0.98	69	98.4	98	Streptophyta	Streptophyta	Chloroplast
Psychroserpens	0	68	100	100	Flavobacteriaceae	Flavobacteriaceae	Flavobacteriaceae
Aliivibriofischeri	0	70	100	100	Vibrionaceae	Vibrio	Vibrio
Fluviicola	0.98	72	96.4	97	Fluviicola	Cryomorphaceae	Crocinitomicaceae
Bdellovibrio	0.99	73	88.5	99	Bdellovibrio	Deltaproteobacteria	Bdellovibrio
Methanocorpusculum	0.98	74	97.3	98	Methanocorpusculum	Methanocorpusculum	Methanocorpusculum

APPENDIX B

SUPPLEMENTARY MATERIALS FOR “CHAPTER IV: COMPARISON OF GUT MICROBIOTA IN NATURALLY OCCURRING AND LABORATORY AQUACULTURED SEA URCHIN *LYTECHINUS VARIEGATUS* REVEALED DIFFERENCES IN THE COMMUNITY COMPOSITION AND PREDICTED FUNCTIONS”

Supplementary Table 4-1: The maximum and total scores, E-value, percent similarity and sequence accession number of the top 100 taxa identities determined by the Basic Local Alignment Search Tool (BLAST) alignment of the representative sequence assigned to highly abundant Campylobacteraceae of the *Lytechinus variegatus* samples. The BLAST alignment was conducted using the nucleotide collection (nr/nt) database through Megablast optimized for highly similar sequences (https://blast.ncbi.nlm.nih.gov/).

Description	Max Score	Total Score	Query Cover	E-value	Per. Ident	Accession
Uncultured bacterium clone CP_Otu217 16S ribosomal RNA gene, partial sequence	335	335	100%	5.00E-88	90.51%	KY275704.1
Arcobacter bivalviorum strain LMG 26154 chromosome, complete genome	335	1341	100%	5.00E-88	90.51%	CP031217.1
Arcobacter sp. W129-99 partial 16S rRNA gene, strain W129-99	335	335	100%	5.00E-88	90.51%	LT904750.1
Arcobacter bivalviorum strain SCAU-004 16S ribosomal RNA gene, partial sequence	335	335	100%	5.00E-88	90.51%	MF155896.1
Arcobacter bivalviorum strain LPB305 16S ribosomal RNA gene, partial sequence	335	335	100%	5.00E-88	90.51%	MG397006.1
Arcobacter bivalviorum strain CECT7835T 16S ribosomal RNA gene, partial sequence	335	335	100%	5.00E-88	90.51%	MG195891.1
Uncultured *Arcobacter* sp. clone 99 16S ribosomal RNA gene, partial sequence	335	335	100%	5.00E-88	90.51%	KY595132.1
Arcobacter sp. F155-44 partial 16S rRNA gene, strain F155-44	335	335	100%	5.00E-88	90.51%	LT629994.1
Arcobacter sp. F161-33 partial 16S rRNA gene, strain F161-33	335	335	100%	5.00E-88	90.51%	LT629993.1
Arcobacter sp. F155-33 partial 16S rRNA gene, strain F155-33	335	335	100%	5.00E-88	90.51%	LT629992.1

Uncultured *Arcobacter* sp. clone TST2N70 16S ribosomal RNA gene, partial sequence	335	335	100%	5.00E-88	90.51%	KX119557.1
Uncultured *Arcobacter* sp. clone TST2N59 16S ribosomal RNA gene, partial sequence	335	335	100%	5.00E-88	90.51%	KX119556.1
Uncultured *Arcobacter* sp. clone TST2N60 16S ribosomal RNA gene, partial sequence	335	335	100%	5.00E-88	90.51%	KX119555.1
Uncultured *Arcobacter* sp. clone TST1N74 16S ribosomal RNA gene, partial sequence	335	335	100%	5.00E-88	90.51%	KX119531.1
Uncultured *Arcobacter* sp. clone TST1N64 16S ribosomal RNA gene, partial sequence	335	335	100%	5.00E-88	90.51%	KX119530.1
Uncultured bacterium clone I3Q1XXJ02CKAN1 16S ribosomal RNA gene, partial sequence	335	335	100%	5.00E-88	90.51%	KP952681.1
Uncultured bacterium clone I3Q1XXJ02B6EDT 16S ribosomal RNA gene, partial sequence	335	335	100%	5.00E-88	90.51%	KP952498.1
Uncultured bacterium clone I3Q1XXJ01A2FM8 16S ribosomal RNA gene, partial sequence	335	335	100%	5.00E-88	90.51%	KP952291.1
Uncultured bacterium clone I3Q1XXJ01ASHTJ 16S ribosomal RNA gene, partial sequence	335	335	100%	5.00E-88	90.51%	KP951781.1
Uncultured bacterium clone IZ1RPV404EBV4Y 16S ribosomal RNA gene, partial sequence	335	335	100%	5.00E-88	90.51%	KP944254.1
Uncultured *Arcobacter* sp. clone 16G-8 16S ribosomal RNA gene, partial sequence	335	335	100%	5.00E-88	90.51%	KC918188.1
Uncultured *Arcobacter* sp. clone 16G-7 16S ribosomal RNA gene, partial sequence	335	335	100%	5.00E-88	90.51%	KC918187.1
Uncultured *Arcobacter* sp. clone 16G-5 16S ribosomal RNA gene, partial sequence	335	335	100%	5.00E-88	90.51%	KC918185.1
Uncultured *Arcobacter* sp. clone 16G-4 16S ribosomal RNA gene, partial sequence	335	335	100%	5.00E-88	90.51%	KC918184.1
Uncultured *Arcobacter* sp. clone 16G-30 16S ribosomal RNA gene, partial sequence	335	335	100%	5.00E-88	90.51%	KC918183.1

Uncultured *Arcobacter* sp. clone 16G-29 16S ribosomal RNA gene, partial sequence	335	335	100%	5.00E-88	90.51%	KC918181.1
Uncultured *Arcobacter* sp. clone 16G-28 16S ribosomal RNA gene, partial sequence	335	335	100%	5.00E-88	90.51%	KC918180.1
Uncultured *Arcobacter* sp. clone 16G-27 16S ribosomal RNA gene, partial sequence	335	335	100%	5.00E-88	90.51%	KC918179.1
Uncultured *Arcobacter* sp. clone 16G-25 16S ribosomal RNA gene, partial sequence	335	335	100%	5.00E-88	90.51%	KC918177.1
Uncultured *Arcobacter* sp. clone 16G-23 16S ribosomal RNA gene, partial sequence	335	335	100%	5.00E-88	90.51%	KC918175.1
Uncultured *Arcobacter* sp. clone 16G-21 16S ribosomal RNA gene, partial sequence	335	335	100%	5.00E-88	90.51%	KC918173.1
Uncultured *Arcobacter* sp. clone 16G-20 16S ribosomal RNA gene, partial sequence	335	335	100%	5.00E-88	90.51%	KC918172.1
Uncultured *Arcobacter* sp. clone 16G-2 16S ribosomal RNA gene, partial sequence	335	335	100%	5.00E-88	90.51%	KC918171.1
Uncultured *Arcobacter* sp. clone 16G-17 16S ribosomal RNA gene, partial sequence	335	335	100%	5.00E-88	90.51%	KC918169.1
Uncultured *Arcobacter* sp. clone 16G-15 16S ribosomal RNA gene, partial sequence	335	335	100%	5.00E-88	90.51%	KC918168.1
Uncultured *Arcobacter* sp. clone 16G-13 16S ribosomal RNA gene, partial sequence	335	335	100%	5.00E-88	90.51%	KC918166.1
Uncultured *Arcobacter* sp. clone 16G-11 16S ribosomal RNA gene, partial sequence	335	335	100%	5.00E-88	90.51%	KC918164.1
Uncultured *Arcobacter* sp. clone 16G-10 16S ribosomal RNA gene, partial sequence	335	335	100%	5.00E-88	90.51%	KC918163.1
Arcobacter sp. 0609ALT48R8-KH partial 16S rRNA gene, isolate 0609ALT48R8-KH	335	335	100%	5.00E-88	90.51%	HF952659.1
Arcobacter sp. AK19 partial 16S rRNA gene, strain AK19	335	335	100%	5.00E-88	90.51%	HE653971.1

Uncultured bacterium clone HS100 16S ribosomal RNA gene, partial sequence	335	335	100%	5.00E-88	90.51%	JX391436.1
Arcobacter sp. AK11 partial 16S rRNA gene, strain AK11	335	335	100%	5.00E-88	90.51%	FR870464.1
Arcobacter bivalviorum partial 16S rRNA gene, strain F118-4	335	335	100%	5.00E-88	90.51%	HE565358.1
Arcobacter bivalviorum partial 16S rRNA gene, strain F118-2	335	335	100%	5.00E-88	90.51%	HE565357.1
Uncultured *Arcobacter* sp. clone Liv16S-L231 16S ribosomal RNA gene, partial sequence	335	335	100%	5.00E-88	90.51%	JN087474.1
Uncultured *Arcobacter* sp. clone Liv16S-L219 16S ribosomal RNA gene, partial sequence	335	335	100%	5.00E-88	90.51%	JN087467.1
Arcobacter bivalviorum strain F4 16S ribosomal RNA, partial sequence	335	335	100%	5.00E-88	90.51%	NR_116730.1
Uncultured bacterium clone B9_10.2_2 16S ribosomal RNA gene, partial sequence	335	335	100%	5.00E-88	90.51%	FJ717129.1
Uncultured bacterium clone H12_10.2_2 16S ribosomal RNA gene, partial sequence	335	335	100%	5.00E-88	90.51%	FJ717121.1
Uncultured bacterium clone A11_10.2_2 16S ribosomal RNA gene, partial sequence	335	335	100%	5.00E-88	90.51%	FJ717100.1
Uncultured epsilon proteobacterium clone PI_4b8c 16S ribosomal RNA gene gene, partial sequence	335	335	100%	5.00E-88	90.51%	AY580422.1
Uncultured bacterium clone I3Q1XXJ02B658K 16S ribosomal RNA gene, partial sequence	331	331	100%	6.00E-87	90.16%	KP951028.1
Arcobacter sp. strain SCAU-007 16S ribosomal RNA gene, partial sequence	329	329	100%	2.00E-86	90.12%	MF155899.1
Uncultured bacterium clone I3Q1XXJ01ALEQC 16S ribosomal RNA gene, partial sequence	329	329	100%	2.00E-86	90.12%	KP953530.1

Uncultured bacterium clone G250WV301BV5KC 16S ribosomal RNA gene, partial sequence	329	329	100%	2.00E-86	90.16%	KF344543.1
Uncultured bacterium clone G250WV301A47CA 16S ribosomal RNA gene, partial sequence	329	329	100%	2.00E-86	90.16%	KF344504.1
Uncultured bacterium clone G250WV301BGLNY 16S ribosomal RNA gene, partial sequence	329	329	100%	2.00E-86	90.16%	KF344498.1
Uncultured bacterium clone G250WV301BHHL1 16S ribosomal RNA gene, partial sequence	329	329	100%	2.00E-86	90.16%	KF344449.1
Sulfuricurvum sp. L14 16S ribosomal RNA gene, partial sequence	329	329	100%	2.00E-86	90.12%	JX399886.1
Sulfuricurvum sp. L2 16S ribosomal RNA gene, partial sequence	329	329	100%	2.00E-86	90.12%	JX399884.1
Uncultured *Arcobacter* sp. clone 16G-9 16S ribosomal RNA gene, partial sequence	329	329	100%	2.00E-86	90.12%	KC918189.1
Uncultured *Arcobacter* sp. clone 16G-22 16S ribosomal RNA gene, partial sequence	329	329	100%	2.00E-86	90.20%	KC918174.1
Uncultured *Arcobacter* sp. clone 16G-19 16S ribosomal RNA gene, partial sequence	329	329	100%	2.00E-86	90.12%	KC918170.1
Uncultured *Arcobacter* sp. clone 16G-12 16S ribosomal RNA gene, partial sequence	329	329	100%	2.00E-86	90.12%	KC918165.1
Uncultured bacterium clone G250WV301BJ7J9 16S ribosomal RNA gene, partial sequence	329	329	100%	2.00E-86	90.16%	JX919780.1
Uncultured bacterium clone G250WV301A64P2 16S ribosomal RNA gene, partial sequence	329	329	100%	2.00E-86	90.16%	JX919407.1
Uncultured *Arcobacter* sp. clone Liv16S-L246 16S ribosomal RNA gene, partial sequence	329	329	100%	2.00E-86	90.12%	JN087480.1
Uncultured bacterium gene for 16S rRNA, partial sequence, clone: SRWH-BA07	329	329	100%	2.00E-86	90.12%	AB546063.1
Uncultured bacterium clone I3Q1XXJ01A3XE4 16S ribosomal RNA gene, partial sequence	327	327	100%	8.00E-86	90.12%	KP953351.1

Uncultured bacterium clone I3Q1XXJ02CBUNT 16S ribosomal RNA gene, partial sequence	327	327	100%	8.00E-86	90.12%	KP952832.1
Uncultured *Arcobacter* sp. clone 16G-26 16S ribosomal RNA gene, partial sequence	327	327	99%	8.00E-86	90.12%	KC918178.1
Uncultured *Arcobacter* sp. clone 16G-1 16S ribosomal RNA gene, partial sequence	327	327	100%	8.00E-86	90.12%	KC918162.1
Uncultured bacterium clone G7DUZBG01AEIH7 16S ribosomal RNA gene, partial sequence	327	327	100%	8.00E-86	90.12%	JX923959.1
Uncultured bacterium clone G250WV301ADHDT 16S ribosomal RNA gene, partial sequence	326	326	100%	3.00E-85	89.80%	KF344910.1
Uncultured bacterium clone I3Q1XXJ01A3EA9 16S ribosomal RNA gene, partial sequence	324	324	100%	1.00E-84	89.76%	KP949269.1
Uncultured bacterium clone G7DUZBG01AJSFT 16S ribosomal RNA gene, partial sequence	324	324	100%	1.00E-84	89.76%	KF352829.1
Uncultured bacterium clone G7DUZBG01A11SY 16S ribosomal RNA gene, partial sequence	324	324	100%	1.00E-84	89.76%	KF348739.1
Uncultured bacterium clone G7DUZBG01AE1IY 16S ribosomal RNA gene, partial sequence	324	324	100%	1.00E-84	89.76%	KF346738.1
Sulfuricurvum sp. L12 16S ribosomal RNA gene, partial sequence	324	324	100%	1.00E-84	89.76%	JX399885.1
Uncultured *Arcobacter* sp. clone 16G-24 16S ribosomal RNA gene, partial sequence	324	324	100%	1.00E-84	89.72%	KC918176.1
Uncultured bacterium clone G7DUZBG01BIQOQ 16S ribosomal RNA gene, partial sequence	324	324	100%	1.00E-84	89.76%	JX923246.1
Uncultured bacterium clone 16S_PCR_2H05 16S ribosomal RNA gene, partial sequence	324	324	100%	1.00E-84	89.72%	EF462744.1
Uncultured bacterium clone Sm3-24 16S ribosomal RNA gene, partial sequence	324	324	100%	1.00E-84	89.72%	EF582442.1
Uncultured bacterium clone I3Q1XXJ02CAQWU 16S ribosomal RNA gene, partial sequence	322	322	100%	4.00E-84	89.76%	KP952853.1

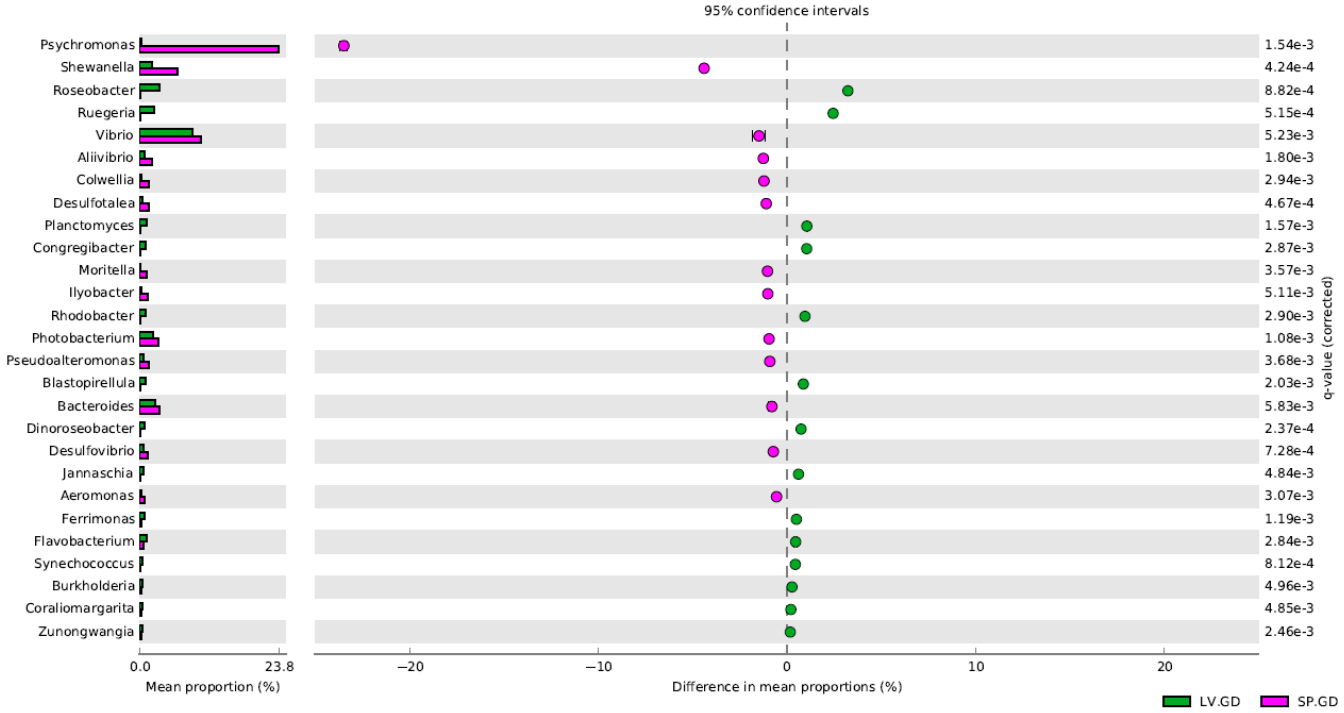
Uncultured bacterium clone I3Q1XXJ01BCMJ8 16S ribosomal RNA gene, partial sequence	322	322	100%	4.00E-84	89.72%	KP952296.1
Uncultured bacterium clone IZ1RPV403CXNVC 16S ribosomal RNA gene, partial sequence	322	322	100%	4.00E-84	89.72%	KP946747.1
Uncultured bacterium clone I3Q1XXJ01A6UZO 16S ribosomal RNA gene, partial sequence	320	320	100%	1.00E-83	89.45%	KP952416.1
Uncultured bacterium clone G250WV301BNGWS 16S ribosomal RNA gene, partial sequence	320	320	100%	1.00E-83	89.45%	KF344553.1
Uncultured bacterium clone G250WV301A56MJ 16S ribosomal RNA gene, partial sequence	320	320	100%	1.00E-83	89.45%	KF344452.1
Uncultured bacterium clone G250WV301AMKTW 16S ribosomal RNA gene, partial sequence	320	320	100%	1.00E-83	89.45%	KF344451.1
Uncultured bacterium clone I3Q1XXJ01A4PNI 16S ribosomal RNA gene, partial sequence	318	318	96%	5.00E-83	90.16%	KP948886.1
Uncultured bacterium clone OTU33285_AL241_97981 16S ribosomal RNA gene, partial sequence	318	318	100%	5.00E-83	89.37%	KP928658.1
Uncultured bacterium clone G7DUZBG01BSP15 16S ribosomal RNA gene, partial sequence	318	318	100%	5.00E-83	89.41%	KF349220.1
Uncultured bacterium clone MD12g11_18542 16S ribosomal RNA gene, partial sequence	318	318	100%	5.00E-83	89.33%	JQ378555.2
Uncultured bacterium clone 3M34_048 16S ribosomal RNA gene, partial sequence	318	318	100%	5.00E-83	89.33%	JQ287456.1
Uncultured bacterium clone 3M34_047 16S ribosomal RNA gene, partial sequence	318	318	100%	5.00E-83	89.33%	JQ287455.1
Uncultured bacterium clone 3M34_025 16S ribosomal RNA gene, partial sequence	318	318	100%	5.00E-83	89.33%	JQ287445.1
Uncultured bacterium clone 7M24_080 16S ribosomal RNA gene, partial sequence	318	318	100%	5.00E-83	89.33%	JQ287348.1

Uncultured bacterium clone 7M24_040 16S ribosomal RNA gene, partial sequence	318	318	100%	5.00E-83	89.33%	JQ287316.1
Uncultured bacterium clone 7M24_017 16S ribosomal RNA gene, partial sequence	318	318	100%	5.00E-83	89.33%	JQ287299.1

APPENDIX C

SUPPLEMENTARY MATERIALS FOR “CHAPTER V: SHOTGUN METAGENOMICS REVEALED DIFFERENCES IN THE MICROBIOTA WITH KEY METABOLIC ATTRIBUTES EMPHASIZING NITROGEN FIXATION IN SEA URCHIN GUT DIGESTA”

Supplementary Figure 5-1: Extended error bar analysis conducted on the top 50 genera determined in the sea urchin gut digesta. The analysis was performed using Statistical Analysis of Metagenomic Profiles (STAMP; v2.1.3) on the top 50 genera from the Bacteriome, determined through RefSeq as implemented in MG-RAST (v4.0.3). Technical replicates were grouped and normalized according to relative abundance. A two-sided Welch's t-test was performed between groups, and *p*-values were corrected using the Benjamini-Hochberg False Discovery Rate (FDR) correction (*q*-values). Genera showing significant differences between groups (*q*-value < 0.01) were listed, along with the relative abundance bar graphs (left) and the difference in mean proportions (center). Taxa identified as "Other" or "Unclassified" were excluded from the figure for better representation at the genus level. Groups are indicated as follows: LV.GD = green sea urchin *Lytechinus variegatus* gut digesta; SP.GD = purple sea urchin *Strongylocentrotus purpuratus* gut digesta. Taxonomic data generated through RefSeq as implemented in MG-RAST (v4.0.3).



Supplementary Table 5-1: The relative abundances and two-group statistical results of the top 50 genera determined in the green and purple sea urchin gut digesta have been elaborated. Technical replicates were grouped and normalized by relative abundance. A two-sided Welch's t-test was performed between groups, and *p*-values were corrected using the Benjamini-Hochberg False Discovery Rate (FDR) correction (*q*-values). Taxa identified as "Other" and "Unclassified" are also indicated. Shown is the relative abundance per group including the standard deviation, the *p*-values, corrected *p*-values (*q*-values), and difference between means (effect size). Groups are indicated as follows: LV.GD = green sea urchin *Lytechinus variegatus* gut digesta; SP.GD = purple sea urchin *Strongylocentrotus purpuratus* gut digesta. Taxonomic data generated through RefSeq as implemented in MG-RAST (v4.0.3).

Genus	**LV.GD relative abundance (%)**	**LV.GD standard deviation (%)**	**SP.GD relative abundance (%)**	**SP.GD standard deviation (%)**	***p*-values**	***q*-values**	**effect size**
Aeromonas	0.291	0.003	0.839	0.001	1.14E-03	3.07E-03	-0.548
Aliivibrio	0.875	0.000	2.121	0.002	5.31E-04	1.80E-03	-1.246
Bacillus	0.636	0.003	0.638	0.001	5.46E-01	5.46E-01	-0.002
Bacteroides	2.642	0.029	3.436	0.020	3.43E-03	5.83E-03	-0.794
Blastopirellula	0.950	0.000	0.078	0.001	6.36E-04	2.03E-03	0.873
Burkholderia	0.521	0.005	0.249	0.002	2.62E-03	4.96E-03	0.272
Campylobacter	0.685	0.013	0.107	0.003	1.10E-02	1.52E-02	0.578
Clostridium	1.512	0.006	1.880	0.014	9.98E-03	1.45E-02	-0.367
Colwellia	0.289	0.001	1.502	0.004	1.15E-03	2.94E-03	-1.213
Congregibacter	1.119	0.005	0.069	0.001	1.01E-03	2.87E-03	1.050

Coraliomargarita	0.569	0.000	0.357	0.001	2.66E-03	4.85E-03	0.212
Cytophaga	0.553	0.016	0.281	0.002	3.46E-02	3.76E-02	0.271
Desulfotalea	0.528	0.002	1.620	0.002	9.16E-06	4.67E-04	-1.091
Desulfovibrio	0.687	0.004	1.405	0.005	1.14E-04	7.28E-04	-0.718
Dinoroseobacter	0.793	0.001	0.041	0.001	9.28E-06	2.37E-04	0.752
Ferrimonas	0.780	0.005	0.276	0.005	2.11E-04	1.19E-03	0.504
Flavobacterium	1.156	0.011	0.691	0.010	1.17E-03	2.84E-03	0.465
Fusobacterium	0.409	0.003	1.406	0.019	9.50E-03	1.42E-02	-0.997
Geobacter	0.499	0.010	0.375	0.004	2.98E-02	3.38E-02	0.124
Gramella	0.918	0.017	0.442	0.002	2.17E-02	2.63E-02	0.476
Ilyobacter	0.331	0.005	1.342	0.013	2.61E-03	5.11E-03	-1.010
Jannaschia	0.669	0.003	0.053	0.001	2.37E-03	4.84E-03	0.616
Maribacter	0.626	0.009	0.344	0.000	2.12E-02	2.64E-02	0.282
Marinomonas	0.272	0.001	0.748	0.012	1.60E-02	2.03E-02	-0.476
Marivirga	0.494	0.005	0.211	0.000	1.16E-02	1.55E-02	0.283
Moritella	0.100	0.001	1.126	0.005	1.68E-03	3.57E-03	-1.026
Paludibacter	0.400	0.001	0.564	0.005	1.38E-02	1.80E-02	-0.163
Parabacteroides	0.598	0.014	0.782	0.004	3.43E-02	3.80E-02	-0.184
Pedobacter	0.555	0.007	0.498	0.001	6.72E-02	6.85E-02	0.057

Photobacterium	2.348	0.009	3.294	0.010	2.12E-04	1.08E-03	-0.945
Pirellula	0.900	0.015	0.040	0.001	1.08E-02	1.53E-02	0.859
Planctomyces	1.216	0.003	0.160	0.001	4.30E-04	1.57E-03	1.057
Polaribacter	0.602	0.012	0.322	0.001	2.45E-02	2.90E-02	0.279
Porphyromonas	0.361	0.004	0.471	0.009	2.50E-02	2.90E-02	-0.110
Prevotella	0.643	0.019	0.838	0.007	3.81E-02	4.05E-02	-0.196
Pseudoalteromonas	0.623	0.000	1.526	0.003	1.66E-03	3.68E-03	-0.904
Pseudomonas	1.062	0.006	0.722	0.026	3.90E-02	4.06E-02	0.340
Psychromonas	0.314	0.001	23.790	0.016	3.93E-04	1.54E-03	-23.476
Rhodobacter	1.032	0.005	0.073	0.001	1.25E-03	2.90E-03	0.959
Rhodopirellula	1.791	0.017	0.265	0.000	6.96E-03	1.14E-02	1.526
Roseobacter	3.390	0.001	0.156	0.003	1.04E-04	8.82E-04	3.234
Roseovarius	0.721	0.008	0.052	0.000	7.33E-03	1.17E-02	0.669
Ruegeria	2.569	0.006	0.118	0.004	3.03E-05	5.15E-04	2.451
Shewanella	2.083	0.011	6.470	0.015	4.16E-05	4.24E-04	-4.386
Synechococcus	0.563	0.003	0.115	0.002	1.12E-04	8.12E-04	0.448
Vibrio	9.043	0.050	10.519	0.058	2.98E-03	5.23E-03	-1.476
Zunongwangia	0.425	0.002	0.245	0.003	8.21E-04	2.46E-03	0.179
unclassified (Flavobacteriales)	0.478	0.005	0.211	0.001	8.11E-03	1.25E-02	0.266

unclassified (Gammaproteobacteria)	2.051	0.010	0.276	0.005	3.86E-04	1.64E-03	1.775
unclassified (Rhodobacteraceae)	0.638	0.002	0.035	0.001	2.48E-04	1.15E-03	0.602
Other	46.692	0.080	26.820	0.068	3.55E-05	4.53E-04	19.872