814-769-9443

No guts, no reef: understanding the microbiome, genetic diversity, and phylogenetic relationships of the Puerto Rican long-spined black sea urchin, *Diadema antillarum*, a keystone species of the Caribbean reef

Alejandro J. Mercado Capote

A dissertation submitted in partial fulfillment of the requirements for the degree of

thesis?

MASTER OF SCIENCE

in

BIOLOGY

UNIVERSITY OF PUERTO RICO

MAYAGÜEZ CAMPUS

2020

Approved by:

Juan Carlos Martinez Cruzado, Ph.D.	Date
President, Graduate Committee	
Audrey Majeske, Ph.D.	Date
Member, Graduate Committee	

delete space

Taras Olesyk, Ph.D.

Member, Graduate Committee

Date

Nikolaos V. Schizas, Ph.D

Member, Graduate Committee

Date

Ana Velez, M.S.

~~Internal Director of~~ Biology Department

Interim

Date

Rosa I. Roman Perez, PhD

Representative Office of Graduate Studies

Documentation format?
(one page → 1 sheet??

1. Resumen
2. Table of Figures
   All figures must have a title, like the appendix, tho do not have one.

THIS PAGE IS INTENTIONALLY LEFT BLANK

Abstract

In this paper we describe the gut microbiome of the Puerto Rican keystone sea urchin *Diadema antillarum*. We also used next generation sequencing of the *Cytochrome B* region to examine the phylogenetic relationships among the different samples. The gut microbial communities were mostly populated with Proteobacteria, in which a large proportion were in the class Alphaproteobacteria, followed by classes Betaproteobacteria and Gamaproteobacteria. Firmicutes, Clostrodiales and Tenericutes were other represented phyla in the gut tissue samples of *D. antillarum*. Within Tenericutes, only several animals were harboring the species *Candidatus hepatoplasma*. There is reason to suspect that all these bacteria do not represent a threat to the sea urchin except for the Firmicutes. Clostridium are known highly virulent bacteria that could potentially be the main cause behind the great *Diadema antillarum* die off. Bioinformatic and statistical analysis using pairwise chi-squared analysis also reveal the impact that anthropocentric activities can impact the microbiome. We understand that current strength and relative position on the island also play a role on defining the microbiome. The sea urchins placed upstream have a statistically different composition that animals placed downstream. These findings paint a clearer picture of the Puerto Rican *D. antillarum* population that can aid in the efforts to restore this animals' numbers premortality.

on different page

Resumen

Acknowledgements

Foremost, I would like to express my genuine gratitude to my committee members for the continuous support I received throughout my academic career at ~~Universidad de~~ Mayaguez (at) Puerto Rico. I want to thank my advisor Prof. Audrey Majeske for her incredible insights, persistence, motivation, and communication. Her guidance was essential throughout the entire life of the project and this project would have not been possible without her. I want to extend my thanks to Prof. Juan Carlos Martinez Cruzado for his guidance and patience from the very beginning. My sincere thanks to Prof. Taras Olesyk who was an essential member that helped the entire project through his invaluable knowledge, and experience. Also, Prof. Nikolaos Schizas who provided meaningful suggestions and guidance through the entire process. Lastly, Professor Heidy Morales who was there at the seeding of this project.

I want to also extend my thanks to my colleagues who helped me throughout my career. To Mrs. Stephanie Castro Marquez for helping me and the project with her ideas, comments, and doing the immaculate sequencing lab work. I would like to thank my friend Edmundo Torres, who helped me solve bioinformatic errors and gave many comments and feedback. I would also like to thank Raul Mojica Soto-Albors for helping me with the impeccable lab work at the earlier stages of the project. My previous instructor and friend Mr. Walter Wolfsberger who aided in the bioinformatics and gave excellent insights. Also, my friend David Repollet for helping me with useful comments, feedback, and urchin collection. All the friends, students and family that helped me with the collection, Mariana Torres Gonzales, and Victor Martinez. ~~I want to thank my parents for all the support during this period.~~ Thanks to my students as well for the curiosity in my projects. Among many other classmates and lab partners. Also, I want to thank my parents for all the support during this period.

I would like to thank the Sequencing Facility at Escuela de Medicina de Ponce who conducted sequencing as well. Finally, I would like to thank the University of Mayaguez of Puerto Rico and the department of biology for providing me with the facilities and knowledge that lead to this project and its findings.

List of Content

Table of Figures

Introduction

The importance of Diadema antillarum

Sea urchins are a marine model that have been used extensively for scientific investigations in ecology, toxicology, aquaculture, development, molecular and cell biology, pathology, genetics and many other fields (Agnello 2017; Bianchini et al 2005; Bianchini et al 2007; Bielmyer et al 2005; Buttino et al 2016; Defilippo, J et al 2008; Zhang 2019; Varrella et al 2016; Gambardella, Chiara, et al. 2018; Gambardella et al 2015). Sea urchins appeared about 520 million years ago, before the Cambrian explosion, and they represent an important midpoint between the vertebrates and invertebrates. The apparent homology with vertebrate genomes has been used extensively in multiple comparative genomic studies: there are at least 7,077 sea urchin genes conserved in humans associated with many different illnesses (Sodergren et. al. 2006; Agnello 2017).

Sea urchins also play an important role in the ecosystem. The long-spined sea urchin *Diadema antillarum* is a keystone benthic community member that helps keep the balance of the fragile and decaying coral reef marine ecosystem by grazing on low-nutrient algae allowing for coral reef spawning (Carperter 1988; Carperter & Edmund 2006) and providing complex spaces for predator-prey dynamics with its long spines (Alvarez-Filip 2009). Previous studies have indicated that since the 1970s, the Caribbean reef system has declined consistently and scientists have correlated this with several key events in recent history: the loss of the dominant reef-building *Acropora* corals, the mass mortality of the grazing urchin, *Diadema antillarum*, and the 1998 El Nino Southern Oscillation-induced worldwide coral bleaching event (Alvarez-Filip 2009). These events are correlated, and are partially caused by global climate change, hurricanes, and anthropocentric factors (Hughes 1994, Precht & Aronson 2006, Mumby et al. 2007). The chronic decline in the Caribbean reef system has led to a marine ecosystem phase shift towards macro-algal dominance, resulting in a system with low nutritional value that hinders severely coral growth and recovery (Carpenter 1988; Edmunds & Carpenter 2001; Carperter & Edmund 2006 ; Precht & Aronson 2006). The role of the sea urchin is essential as it physically removes this macroalgae cover and provides the substrate necessary for coral settlement (Furman & Heck 2009) however high numbers of sea urchin can also be damaging to the ecosystem (Edmunds 2001). Many scientists agree that a phase-shift in which algae is dominant with detrimental consequences for the survival of juvenile coral one study and places in which *D. antillarum* has made a recovery it has been observed that the coral reef ecosystem have made a recovery by reducing macroalgae cover (Furman & Heck 2009; Chiappone et al. 2002). *D. antillarum* is a keystone marine species that provides complex topographical three-dimensional space to the ecosystem that is required to maintain fish density (Hay & Taylor 1985).

History

In the early 1980's *D. antillarum* suffered a mass mortality event due to an unknown pathogen (Lessios et al. 1984; Lessios 1988, 1995, 2004, 2015), and since no other organism was affected during that time, it was caused likely by a species-specific pathogen (Defilippo 2018). This mortality appeared to spread through surface currents (Lessios 1988) and the population numbers were depleted, and more than 90% of the population failed to recover in the next two

decades. Before this event, the coral reefs at places like Discovery Bay, Jamaica were characterized by having coral as much as 90% of the substratum and by 1990 the coral cover was reduced to less than 5% (Edmunds & Carpenter 2001).

While *D. antillarum* populations have recovered in some places, the ecosystems in other places have shifted accordingly, resulting in less macroalgal cover and enhanced coral cover and recruitment (Mumby 2007, Blanco et al 2010; Lessios 2016). Due to the nature of this sea urchin's reproductive strategy, ie., broadcast spawning – where individuals will release gametes into their environment, the population must be high for the species to be successful (Allele Effect, Petersen and Levitan 2001; Feehan et al 2016). Consequently, this species has yet to recover to its historic levels because they lack the numbers to successfully reproduce (Lessios 2016; Lessios 2005, Chiappone et al. 2013). Currently in Puerto Rico, the *D. antillarum* populations appear to be stable even though there is little evidence for recovery trends back to the pre-mass mortality densities (Rodriguez-Barreras 2018, Tuohy & Weil 2020).

As continued anthropogenic global climate change will result in extinctions, reduced species diversity, drastic changes in ecosystems (Blois et al. 2013; Moritz and Agudo 2013), and an increased likelihood of disease outbreaks (Burge et al. 2014; Harvell et al. 1999), we must continue to survey population recovery across the Caribbean if we want to design appropriate conservation and management practices for this momentous creature (Quintero 2014, Ripple 2017). Most of the research with populations of *D. antillarum* were conducted in the wake of the massive mortality event (Lessios et al. 1984; Lessios 1979; Lessios 1981; Lessios et al. 2001) which leaves us with important questions that remain unanswered, e.g. what can be said about the Caribbean gene flow and genetic diversity and what is the current pathogenic exposures of this species.

Microbiome & Genetic Flow

Some of these research questions can be explored by studying the microbiome of *D. antillarum*, which have been classified as secondary endocrine and digestive organs that have important roles. These roles include increasing energy extraction from food sources and increasing energy absorption, as well as interacting with molecules that have been shown to cause disease in humans (Baohong et. al 2018; Ahmadmehrabi & Tang 2017). The functional properties of the microbiome are attributed to the numerous genes found inside the metagenome that provides the host with unique and specific biochemical pathways (Baohong et. al 2018). In humans, it has been found that healthy individuals can have variations between each other and in some cases have different functional metabolic levels between subjects considered sedentary and professional athletes (Ahmadmehrabi 2017; Barton et al 2017). In addition, it has been speculated that the microbiome is a source of genetic flow and immunity (Baohong et. al 2018).

Animals usually have digestive compartments that contain specific environments that help the organism efficiently extract nutrients (Ceja-Navarro et. al 2019; Ishaq & Wright 2012). Recent sequencing studies have shown that three species of sea urchin, *Lytechinus variegatus*, *Strongylocentrotus purpuratus* and *Paracentrotus lividus* have intestinal microbiomes that are uniquely compartmentalized ecosystems, which have arisen due to active selection between the

host and their commensal microorganisms (Hakim et al.2015; Hakim et al.2016; Hakim et al 2019 Meziti et al 2007).

Besides monitoring their population levels over the years since the massive mortality event in the 1980s, few genetic studies (Lessios 1979, Lessios 1981, Lessios et al 1984, Lessios et al 2001) have been undertaken to monitor the population gene flow and genetic diversity of *D. antillarum* following a presumed Caribbean wide population bottleneck effect. As scientists are beginning to understand the ecological, immune, and health benefits of the intestinal microbiome in a variety of organisms, it is important to study the microbiome of *D. antillarum,* as it is an essential keystone species for the Caribbean reef system (**Figure 1**). The ecological role of *D. antillarum* is simple but important, and this organism could not perform that role without its microbiome that helps it degrade low nutritional macro-algae cover into nutrients. Also, microbiomes can give us insight into the evolutionary history of the organism in question (Moeller 2016) and even provide the context of the functional role of microbiomes and how they work.

For this experiment we collected Diadema specimens across the island of Puerto Rico and collected its gut contents. We identified the taxonomy of the gut microbiota using NextGen Illumina MiSeq sequencing technology and bioinformatic tools including QIIME2. In addition, to understand the animal's phylogeny a Cytochrome B amplicon was sequenced and compared. By understanding the gene flow, genetic diversity, gut microbiome and metapopulation phylogeny of *D. antillarum* we can understand more about their ecological role both present and past which is important information for conservation efforts if another massive mortality were to occur or prevented.

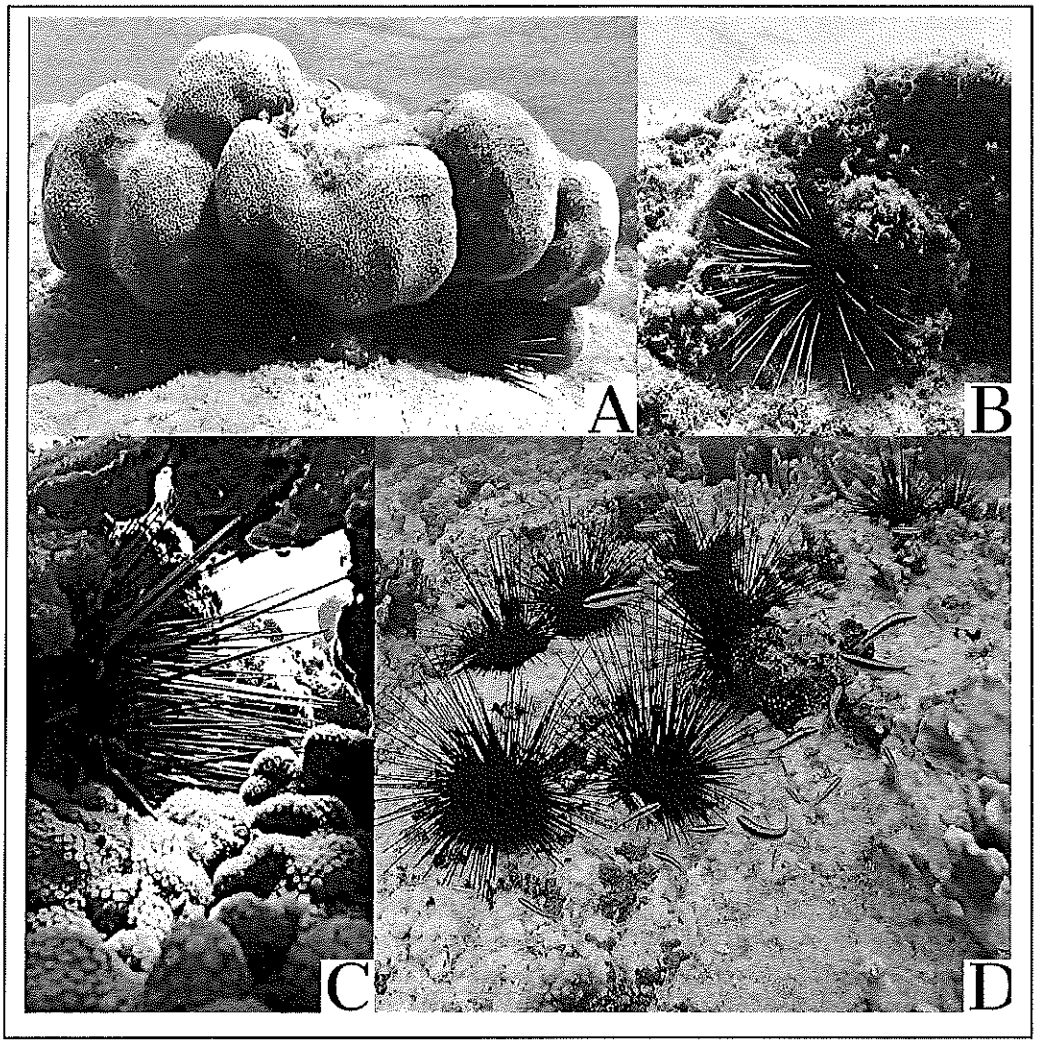


Figure 1. The Puerto Rican subtidal reef habitat of the long-spined black sea urchin, *Diadema antillarum*. Here, the sea urchins are shown hiding in different reef structures, hiding from predators, and finding food (A, B, C). The sea urchins move along the reef shelf to forage for food while other animals use their long spines to protect themselves from predators (D).

[similar to page 12]

Methods

Ethical Statement

D. antillarum habituates in fragile marine ecosystems and the procedures described here were done with extensive precaution to reduce organismal stress and environmental impact. The collection methods to work with this vulnerable species were approved by Departamento de Recursos Naturales of Puerto Rico (O-VS-PVS15AG-00047-01082018).

Animal and Demographic Collection

Sea urchins were collected across the island based on December 2018. The sampled specimens were chosen independent of gender or size. To collect each animal, a diving knife was used to carefully separate *D. antillarum* from the environment by "scooping" the specimens from the place that they were found. Sea urchins were collected across eight different cardinal locations from different municipalities of Puerto Rico representing: north, east, south and west. The municipalities were chosen based on accessibility and include: Ceiba, Culebra, Guánica, Guayama, Isabella, Luquillo, Ponce and Rincon. A total of 44 specimens were collected across the island (**Figure 2**).

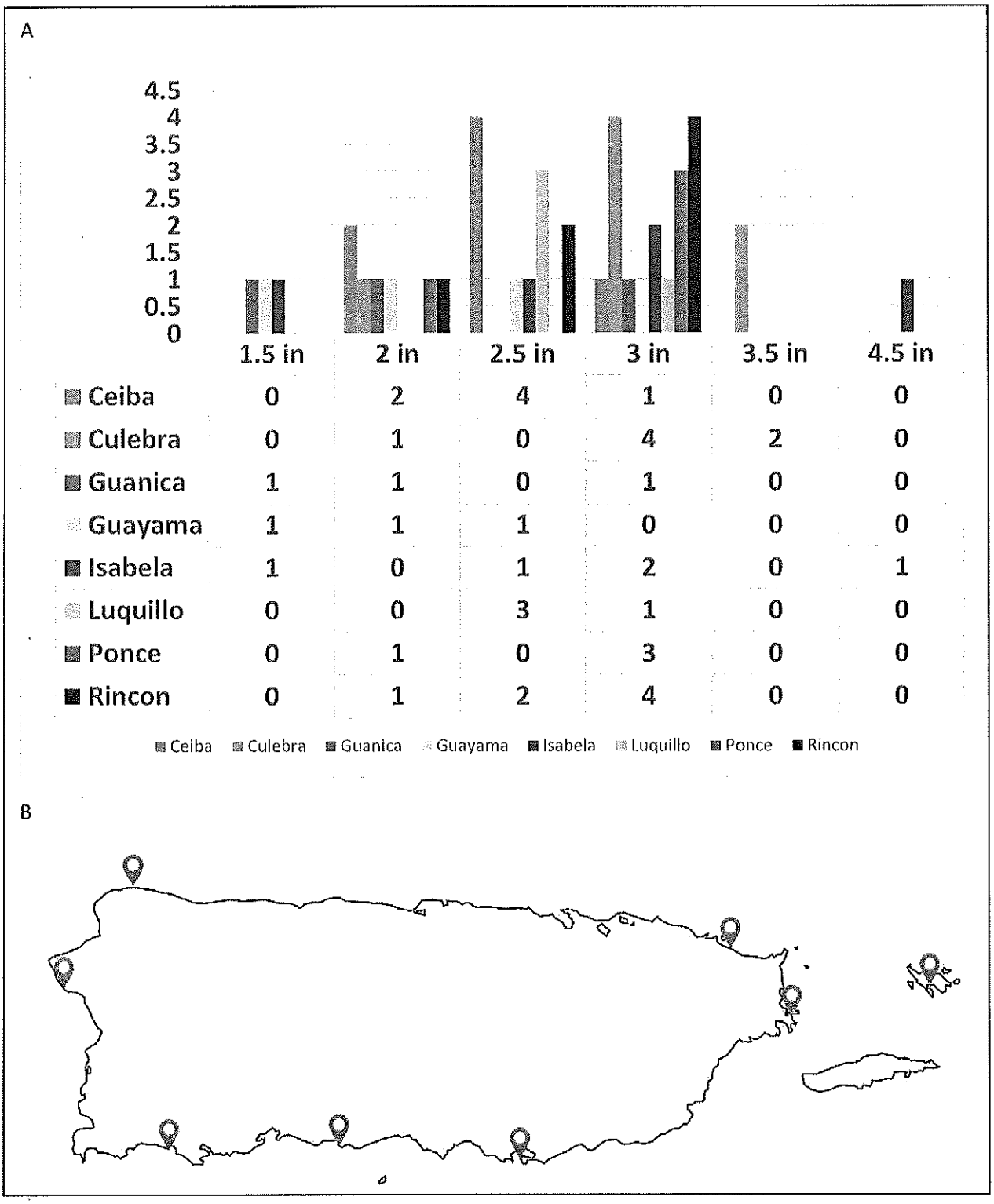


Figure 2. *Sea urchin collection distribution by location and size.* A total of 44 sea urchins were collected from the eight named municipalities in Puerto Rico (A). The number of collected animal samples, one per animal, are shown according to the size of the animal, as given by the diameter in inches (in). The locations of each collection site are indicated with a red open teardrop pin (B),

which includes: Ceiba (18°13'07.8"N 65°36'15.4"W), Culebra (18°18'08.8"N 65°18'33.8"W), Guanica (17°56'05.2"N 66°57'25.6"W), Guayama (17°55'51.3"N 66°09'41.0"W; 17°55'47.3"N 66°09'32.1"W), Isabella (18°30'56.8"N 67°06'00.6"W), Luquillo (18°23'15.3"N 65°43'10.6"W), Ponce (17°57'50.5"N 66°36'35.9"W; 17°58'20.7"N 66°37'04.5"W; 17°57'54.5"N 66°36'28.1"W) and Rincon (18°20'35.2"N 67°15'36.5"W). A total of 44 samples were collected from Rincon (n=10), Guanica (n=3), Ponce(n=3), Isabella (n=5), Luquillo (n=5), Culebra (n=7), Ceiba (n=9), Guayama (n=2).

Once the specimens were physically separated from the environment, they were placed in a diving bag to transport outside of the water. Outside of the water, the specimens were placed in sea water to reduce harm until the individuals were measured and prepared for gut tissue collection (**Figure 3**).

Additional data was collected during animal sampling. The diameter of each animal was measured and recorded in inches. The relative surface water current was recorded as calm, medium or strong according to the cardinal sample location site, which is defined as calm in the south facing the Caribbean Sea, strong to the north facing the Atlantic Ocean and medium in the east and west, according to these positions between the Caribbean Sea and Atlantic Ocean.

Tissue Collection

Sea urchin gut samples were collected at each site to avoid overstressing the animal, potentially risk contamination, or even potentially change microbiome composition. Once the individuals were measured their spines were cut out and an incision was made in the test surrounding the peristomial membrane using ethanol sterilized scissors dissecting around the mouth (**Figure 3C**). The peristomial membrane, along with the nested mouth (Artistotle's Lantern) was lifted from the animal and the gut was collected using ethanol sterilized tweezers (**Figure 3D**). The samples were placed in a 1.5mL tube and held on ice while in immediate transition for storage in a -20 C freezer.

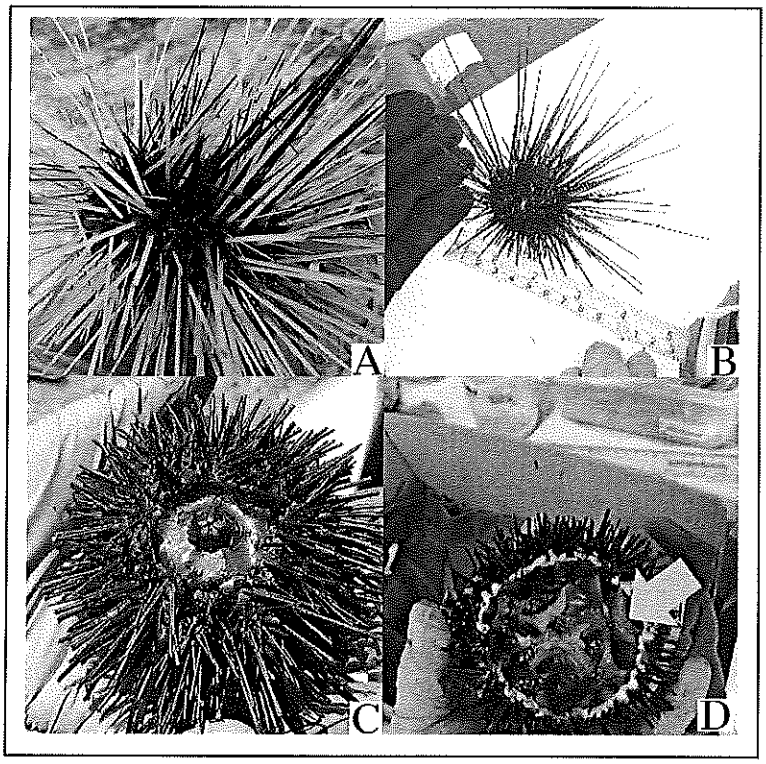


Figure 3. Animal measurement and sample collection. Relative specimen size is shown in (A). The diameter of the body test size was measured in inches (B). Sea urchin spines were cut prior to dissection. An incision was made through the peristomal membrane (C) to gain access to the inner wall and remove the Aristotle's lantern (mouth structure), prior to making a continuous circular incision to split the animal into two sections for sampling (D). A sample of the intestine tissue (grey arrow in D) was placed in a sterile tube and stored on ice during transit to the lab.

DNA Extraction

The Sigma-Aldrich GenElute Stool DNA Isolation Kit was used to isolate DNA from the samples according to the manufacturer's protocol. The DNA was eluted into 50 µL of the company provided buffer and immediately placed in the -20 °C freezer for long-term storage.

DNA Sequencing - 16S V4 Region

Samples were prepared for 16S rRNA targeted sequencing in the variable V3 and V4 region of the gene using the 16S Metagenomic Sequencing Library Preparation kit. Sequencing was performed on an Illumina MiSeq sequencing system that generated raw paired end reads. Sample preparation and sequencing were performed at the Ponce School of Medicine core sequencing facility.

DNA Sequencing - Cytochrome B Region

A subset of the 44 DNA samples (n=21) were subjected to PCR amplification targeting Cytochrome B. To construct the primers, a mitochondrial DNA sequence alignment from 27 closely related species to *D. antillarum* was used to design primers targeting Cytochrome B (Chunxia 2016, Bronstein 2019). The aligned sequences were found to be mostly conserved, at the 14,988 – 16,070 bp region in alignment. This segment of the conserved region was used to design the forward (14,968 - 14,987) and reverse (16,071-16,093) primers, which resulted in a 16,093 base pair fragment and were as follows: forward (5'-GGT|CCA|TTA|CGA|AAG|GAA|CA-3') and reverse (5'-AAT|CTT|TTT|TTC|TAG|GGT|ACA|TA-3'). Each PCR included 50 ng of template

DNA, 0.5 mM of each primer, 20 mM of each deoxyribonucleotide, 25 mM MgCl2, 1x company supplied buffer, 0.025 U of Q5 High Fidelity DNA polymerase (New England Biolabs), 80% DMSO, and 25mM BSA. A total reaction volume of 25 µL was employed for each sample. The PCR program was 95 °C for 5 min followed by 35 cycles of 95 °C for 30 secs, 46°C for 30 sec, 72 °C for 1:30min, with a final extension of 72 °C for 5 min and a hold at 4 °C. Amplicons were electrophoresed through 1.5% agarose 1 x TAE buffer (2 M Tris, 1 M Glacial Acetic Acid, 0.5 M EDTA pH 8). Amplified samples were held in the -20 °C freezer until transportation on ice to the OU Genomics Lab at Oakland University in Rochester, MI. PCR amplicons were purified using Ampure XP beads (Beckman Coulter) in a 0.5X dilution ratio to maximize recovery of the large amplicon size. A library of nucleotide fragments was generated per sample using the Nextera DNA Flex Library Prep Kit with a starting quantity of 100 ng, and according to the manufacturer's instructions. A unique adapter sequence was added to each sample prior to pooling. Pooled and indexed samples were diluted to a loading concentration of 200 pM prior to a final library dilution according to the iSeq100 sequencing system protocol. Samples were sequenced on an Illumina iSeq100 sequencer that generated raw paired end sequences.

Bioinformatic Analysis - Raw Data

The sample data along with commands used for this study can be found at (https://github.com/mercadocapote/diadema_ajmc2020). Two different sequencers were employed to generate the data using separate DNA aliquots from the same animal sample, that targeted the ribosomal 16S and Cytochrome B genes. Additionally, there were sample metadata files generated by the researchers at the site of collection which described location, size, and habitat of each specimen. The metadata sample files are in the ".tsv" format.

Bioinformatic Analysis - QIIME2

The QIIME2 microbiome bioinformatics platform was used to perform the microbiome analysis of the 16S rRNA samples. All of the information pertaining to the installation procedures and how to use the software can be found at docs.qiime2.org. A virtual machine was installed using a Virtual Box Image previously installed with QIIME2 found in the online official documentation in QIIME2. The pipeline analysis was employed using the previously available protocols that largely followed the description given by Estaki (2020) but also according to the research described by Bolyen et al. (2019), Hall & Beiko (2018), Hakim et al (2016, 2015).

The data generated by Illumina's Casava software was first imported into QIIME2. Then a quality filtering process was applied based on quality scores. Quality control of sequences was then completed using a Deblur workflow with a trim length of 220 bps, which resulted in sequences that are referred to as sub-operational taxonomic units (OTUs), or more commonly called in the QIIME2 documentation as features (Estaki 2020). This workflow included the removal of chimera and rare reads that accounted for <0.0005% of all the reads. A feature table was generated as an output of this workflow. A phylogenetic tree was then generated using the fragment-insertion tree building method described by Janseen et al. (2018) to conduct diversity analysis with Faith's phylogenetic diversity (Faith 1992) and UniFrac (Lozupone & Knight 2005). The Greengenes 16s rRNA reference database (McDonald et al. 2012) was used to identify the taxa and build a rooted

phylogenetic tree. The feature table was filtered to only contain samples present in the phylogenetic tree. The sampling depth was evaluated using alpha rarefaction plots to determine if the within-sample diversity is fully reached. The sampling depth was taken from the Feature Table created in the quality filtering process and was p-max-depth 5677 which was the median frequency of features found in the samples. Using the pipeline action core-metrics-phylogenetic we rarefied the Feature Table to a p-sampling-depth of 2603 based on information on the Feature Table (Vaquez-Baeza, Pirrung, Gonzalez, & Knight 2013). This pipeline generates several alpha diversity metrics such as: Shannon's diversity index, a quantitative measure of community richness (Shannon & Weaver 1949); Observed features or OTUs; Evenness, a measure of community richness (Pielou 1966); Faith's Phylogenetic Diversity, a qualitative measure of community richness incorporating phylogenetic relationships (Faith 1992). Also this pipeline generates several beta diversity metrics such as: Jaccard distance, a qualitative measure of community dissimilarity (Jaccard 1908); Bray-Curtis distance, a quantitative measure of community dissimilarity (Sørensen 1948); unweighted UniFrac Distance, a qualitative measure of community dissimilarity with phylogenetic relationships (Lozupone & Knight 2005); weighted UniFrac distance, a quantitative measure of community dissimilarity with phylogenetic relationships (Lozupone, Hamady, Kellley, & Knight 2007). The next pipeline was to test alpha diversity and the distribution of features with boxplots and Kurskal-Wallis test (Estaki 2020). The beta diversity was tested using principal coordinate (PCoA) plots generated in the previous pipeline. The next test conducted was PERMANOVA (Anderson & Walsh 2013) per different metadata category. Finally, the data was divided into the various categories in the metadata (including municipality location, cardinal location, cardinal surface current, reef habitat, and animal size) and then the taxonomic bar plots were visualized for each category or group and all samples and transferred to Microsoft Excel files. Pearson chi-squared statistical analysis was performed on log transformed data ($\log_{10}(x+1)$) for different metadata groupings that compared the microbial profiles at the phylum level of classification.

Bioinformatic Analysis - Geneious Prime

The Geneious Prime bioinformatics platform was used to perform the Cytochrome B gene analysis. All information pertaining to the use of the platform can be found at www.geneious.com. The samples were first imported and merged into paired reads using BBMerge paired Read Merger Version 38.37 (Bushnell 2017, BBDuk Geneious plugin) using the default settings and merge rate in the high setting. Paired reads were then Trimmed using BBDuk Adapter/Quality Trimming Version 38.37 (Bushnell, BBDuk Geneious plugin) using the default settings for Nextera DNA adapters with a Kmer length of 27, maximum substitutions of 1, minimum quality of 30, minimum overlap of 20 and minimum length of 100bps. Then sequences were error corrected and normalized using BBNorm error correction and read normalization version 38.37 (BBNorm, Geneious plugin) using default settings. The duplicate sequences were removed using Dedupe Duplicate Read Remover version 38.37 (Geneious plugin) on default settings and Kmer seed length of 31. Chimeric reads were removed using UCHIME v4.2.40 Chimeric sequence detection using a reference database by Robert Edgar (2011) using default settings and Multiple Sea Urchin Mitochondrion Genome Alignments (Chunxia 2016, Bronstein 2019). Then, the samples were processed through a Velvet 1.2.1 de novo assembly using de Bruijn graphs to

produce assembled contigs. These contigs were then edited to only contain sequence lengths that were close to our desired Cytochrome B amplicon 1,140 bps, so anything below 900 bps and above 2,000 bps was discarded. The contigs were then aligned using Geneious Global Alignment and a cost matrix of 51% similarity (5/-3). A consensus sequence was generated using a threshold of 0% (Majority) from each of the aligned edited contigs. Finally, with the consensus sequence a phylogenetic tree was generated using Geneious Tree Builder and Global alignment using a cost matrix of 51% similarity, with a genetic distance model of Tamura-Nei and method of Neighbor-Joining tree.

Results

Raw 16S V3 and V4 Region

The total number of raw sequences generated from the Illumina MiSeq platform targeting the 16S ribosomal gene in the DNA from 44 sea urchin gut tissue samples collected from eight different locations in Puerto Rico was 1,483,248. The Deblur quality filtering pipeline produced 246,261 features in 43 samples. After removing rare-OTUs or features the data retained 111,929 (45%) features in 43 samples at the sampling depth of 2603. One sample was removed for low quality reads. Of these, taxonomic identification was assigned to 181,713 reads, with 64, 186 unidentified reads. The unidentified reads were not subjected to further statistical analysis. Rarefaction curves were reach suggesting sufficient sampling depth (data not shown).

Microbial diversity across the animal samples

Relative abundances for the sea urchin microbiome taxonomic values was determined using the RDP classifier Greengenes (v13.8) and QIIME2. The gut tissues collected showed an overabundance of organisms from the Proteobacteria phylum classification. Of the identified reads, 99.27% were identified as Proteobacteria (180,399 reads), followed by Firmicutes (952 reads), Bacteroidetes (344 reads) and Tenericutes (18 reads). Besides the number of reads that were classified at the phylum level of Proteobacteria (3241), at the class level, the largest represented group was Alphaproteobacteria (7178 reads), followed by Betaproteobacteria (1217 reads) and then Gammaproteobacteria (916 reads). Other represented groups within Proteobacteria include the class Epsilonproteobacteria, the order Rhizobiales, the families Alcaligenaceae and Bradyrhizobiaceae, the genus Helicobacteraceae and species *Rhizobium daejeonense* (represented in n=4 individuals). The largest represented group within the Firmicutes was the class Clostridiales (1253 reads), and other represented groups within this phylum include the families Lachnospiraceae and Ruminococcaceae. For the phylum Tenericutes, only the species *Candidatus hepatoplasma* was represented.

The type of bacteria found in the samples differ by sample locations and only bacteria from the phyla Proteobacteria and Bacteroidetes were represented in samples from all eight collection sites (**Figure 4**). Within Proteobacteria, bacteria from the classes Alpha-, Beta-, and Gammaproteobacteria were also represented in gut tissue samples from all collection sites. Within Alphaproteobacteria, the order Rhizobiales and the family Bradyrhizobiaceae were represented in most sample location sites, with a few exceptions, e.g., samples classified in the order Rhizobiales were not found in samples collected from Guayama and Guanica, and samples classified in the family Bradyrhizobiaceae were not found in samples collected from Culebra and Guanica. The Gram-negative nitrogen fixing bacteria *Rhizobium daejeonense*, from the order Rhizobiales were identified in four different animal samples collected from three sites, including Rincon, Luquillo (n=2) and Culebra. The family Alcaligenaceae from the phylum Betaproteobacteria were represented in five animal samples collected from Rincon, Isabella, Culebra (n=2) and Ceiba. Bacteria from the phylum Gammaproteobacteria found in samples representing all collection sites were not further classified into more specific taxonomic groups. Bacteria from the class Epsilonproteobacteria as well as those classified further into the genus Helicobacteraceae were

found in multiple animals but were restricted to the animal samples collected in Ceiba (n=4). Within the phylum Firmicutes, bacteria were classified at different taxonomic levels, and at various locations. For example, one animal sample from Guanica was harboring bacteria classified as Firmicutes, whereas the class Clostridiales was found in different animal samples collected at all municipalities. Within Clostridiales, the families Lachnospiraceae and Ruminococcaceae were represented in samples from Rincon, Luquillo and Ceiba. Within the phylum Tenericutes, no bacteria was classified at the higher taxonomic levels, but the species, *Candidatus hepatoplasma*, within the class Mollicutes was found in multiple animal samples at most of the municipality collection sites, except for Ponce, Isabella, and Guayama.

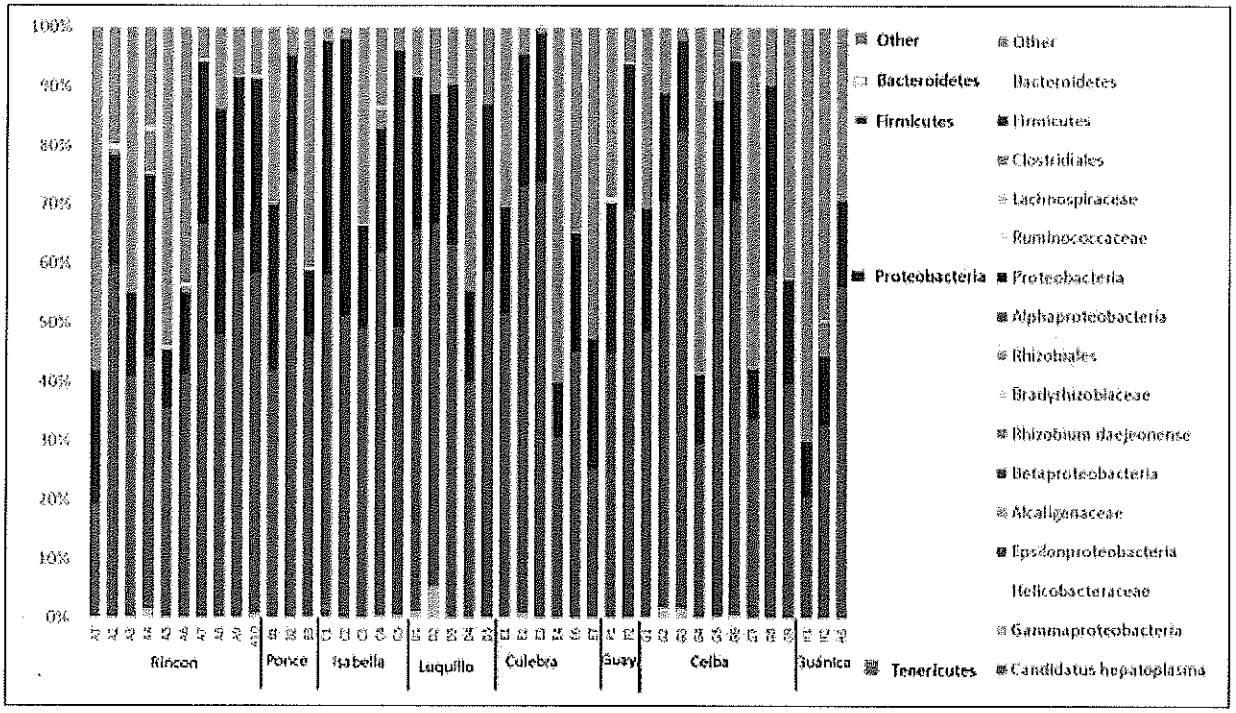


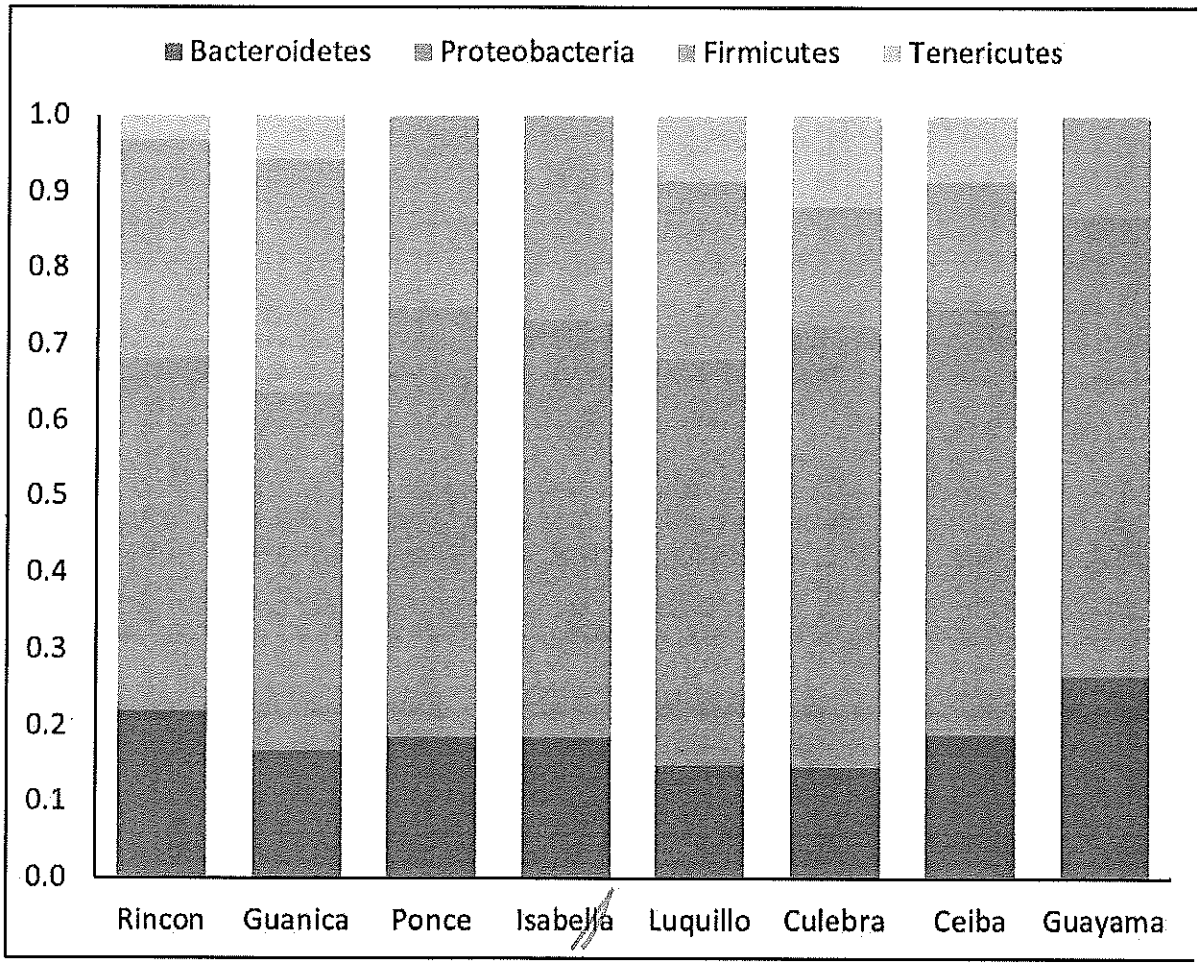
Figure 4. *Taxonomic classification of urchin intestine*. Intestinal microbiome of *D. antillarum* (n=44) collected in the coastal waters of eight different municipalities across Puerto Rico. Microbiome data was generated using 16S metagenomic sequencing. Bars indicate the percentage of microbiota present in each animal sample, in which "Other" refers to the unidentified feature data or OTUs. Samples from the different animals were labeled as A – H according to the different municipality locations and the order of collection, e.g., 10 different animal samples were collected in Rincon (A1 – A10).

Statistical Analysis of Log Transformed data

Overall chi-squared statistical analysis performed on the log transformed data comparing the microbial profiles contained within the phyla (Bacteriodetes, Proteobacteria, Firmicutes and Tenericutes) to animal samples grouped by municipality, cardinal location, cardinal surface current, and animal size resulted in no significant differences between these groupings (results not shown). Pairwise comparisons resulted in significant differences between taxa and groupings pertaining to municipalities, cardinal location, cardinal surface current, and size. For example,

bacteria represented in the samples from Guayama were statistically different from all other municipalities (**Figure 5**).

A



B

	Rincon	Guanica	Ponce	Isabella	Luquillo	Culebra	Ceiba	Guayama
Rincon	-	0.4360	**0.0068**	**0.0098**	**0.0385**	**0.0001**	**0.0042**	**1.24E-06**
Guanica		-	**0.0487**	**0.0802**	0.2671	**0.0013**	**0.0256**	**2.94E-05**
Ponce			-	0.9845	**0.0325**	**0.0006**	**0.0089**	**0.0055**
Isabella				-	**0.03**	**0.0002**	**0.0038**	**0.0011**
Luquillo					-	0.1829	0.3265	**2.34E-18**
Culebra						-	0.5091	**1.1E-31**

Figure 5. *Gut microbiome of D. antillarum by municipalities of Puerto Rico.* Relative taxonomic values are represented in the bar graph (A) by location municipalities. A total of 44 samples were collected from Rincon (n=10), Guanica (n=3), Ponce(n=3), Isabella (n=5), Luquillo (n=5), Culebra (n=7), Ceiba (n=9), and Guayama (n=2). Bars indicate the relative proportions of microbiota found in each specimen. Pairwise Pearson chi-square analyses was used to test differences between the

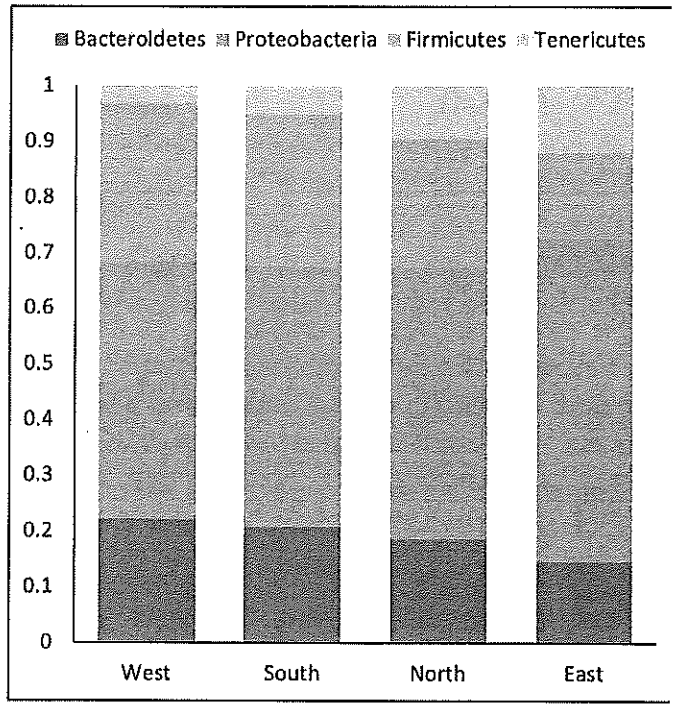
And F [Find Isabella and replace with Isabela]

microbiota profiles according to municipality. P-values are shown in (B) and bolded text indicates significant differences.

There was a greater proportion of bacteria in the phyla Bacteriodetes and Proteobacteria, and no bacteria from the phylum Tenericutes in the samples from Guayama. Similarly, samples from Ponce and Isabella did not include bacteria from the phylum Tenericutes, yet other proportions making up the microbiota profile for these locations were different from Guayama. In addition, the microbiota profile for Luquillo, Culebra and Ceiba were similar to each other, as were the profiles for Rincon and Guanica, however, these two data sets were mostly different from each other in the pairwise comparisions, with the exception of the profiles between Luquillo and Guanica.

When municipalities were grouped by cardinal location, the microbiota profile of the East (Ceiba and Culebra) was significantly different from that of the West (Rincon), as well as the South (Guanica, Ponce and Guayama), in which East and West had the highest level of significance (**Figure 6**; p-value 3.1278E-05). There were no differences in the microbial profiles between other pairwise comparisons including West vs. South, West vs. North, North vs. South and North vs. East.

A



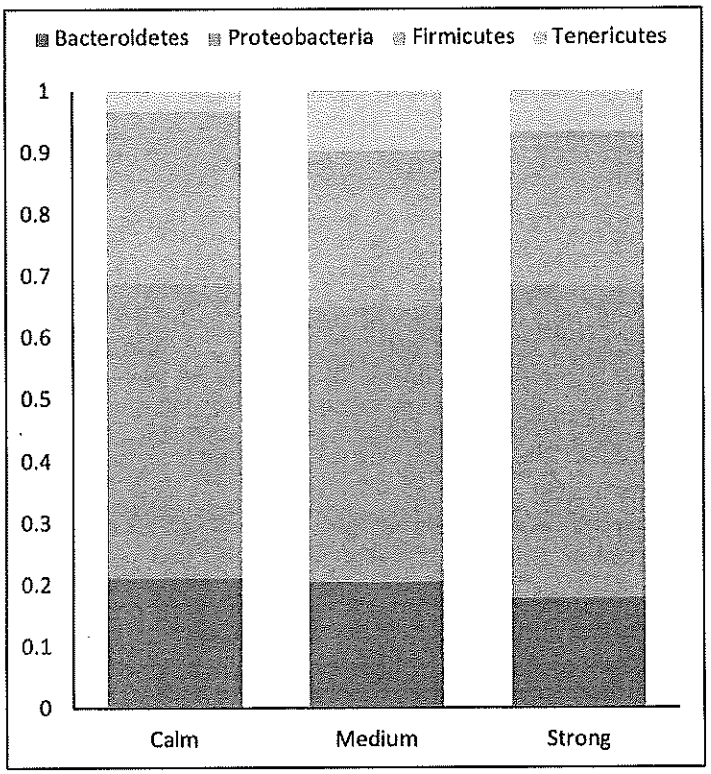
B

	South	North	East
West	0.8391	0.1049	**3.1278E-05**
South		0.3701	**0.0005**
North			0.0692

Figure 6. *Gut microbiome of D. antillarum by cardinal location in Puerto Rico.* Bars indicate the percentage of microbiota that was present in each sample organized by the respective cardinal grouping (A). The bars were generated by QIIME2 analysis which were then log transformed and graphed without unknown taxa. Cardinal groupings included west (n=10), south (n=8), north (18) and east (n=7). Collection sites are categorized into west (Rincon), south (Guanica, Ponce, Guayama), North (Luquillo, Isabella) and East (Ceiba, Culebra). A pairwise for the Pearson chi-square test was conducted to test differences in taxonomic grouping and collection site by cardinal location (B). The p-values are outlined in the bottom table and bolded text indicates significant differences.

When the animal samples were grouped according to cardinal surface current at the collection site, there were no significant differences between the profiles of microbiota and the groupings into calm, medium and strong (**Figure 7**).

A



B

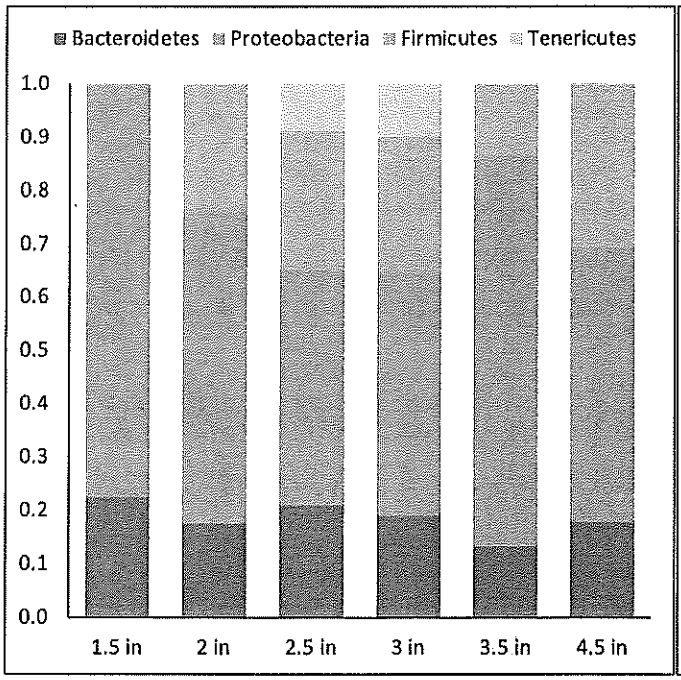
	Medium	Strong
Calm	0.18	0.42
Medium		0.47

Figure 7. *Gut microbiome of D. antillarum by surface water currents.* Surface water currents were categorized by calm in the south facing the Caribbean Sea, strong to the north facing the Atlantic Ocean and medium strength in the east and west being in between both bodies of water. The bars in the graph (A) indicate the relative percentage of microbiota present in each sample based on the current classification that was generated using QIIME2, log transforming data and then graphed without unknown taxa. The corresponding animal sample numbers are calm (n=8), medium (n=26) and strong (n=10). A pairwise statistical analysis was performed using Pearson Chi-squared analysis (B). The p-values are shown in the bottom table.

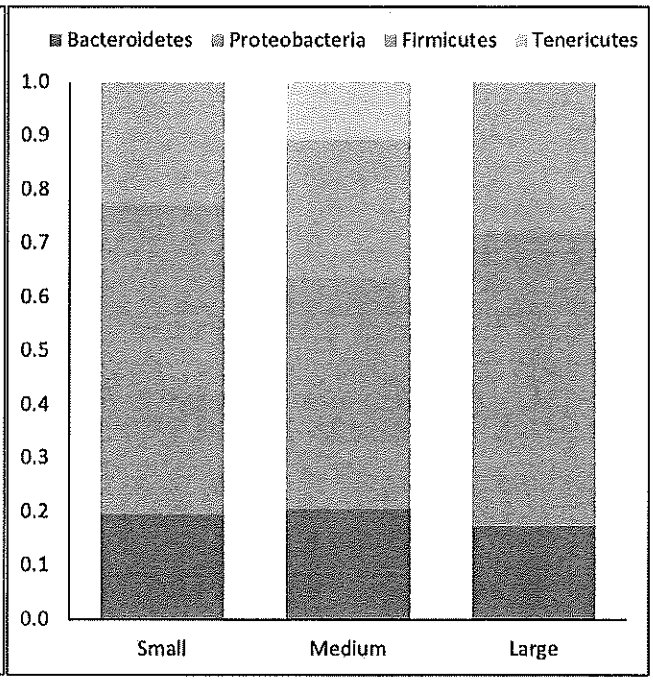
When the animals were grouped by size and relative proportion, there were differences between the profiles of microbiota and these data groupings (**Figure 8A, C**). For example, while the microbial profiles for the sea urchins measured at 2.5 in and 3 in were similar, these were significantly different from all other profiles of sea urchins in the other size categories, including 1.5 in, 2 in, 3.5 in and 4.5 in. Apart from this, all other microbial profiles were significantly different from each other (**Figure 8C**). When the animal sample taxonomic data were grouped into relative proportion, including Small (1.5 - 2 in), Medium (2.5 - 3 in) and Large (3.5 - 4.5 in), the microbiota profile of the Medium animals was significantly different from that of the Small, as well as the Large animals (**Figure 8B, D**). For these comparisons, the largest significant difference was between the Medium and Large animals (**Figure 8D**; p-value 1.552E-25).

Note: in. inch must use period to avoid confusion with the word in
Check Appendices as well.

A



B



C

	Two	Two Point Five	Three	Three Point Five	Four Point Five
One Point Five	0.0000	0.0000	0.0000	0.0003	0.0000
Two		0.0044	0.0059	0.0077	0.5073
Two Point Five			0.9419	0.0000	0.0000
Three				0.0000	0.0000
Three Point Five					0.0002

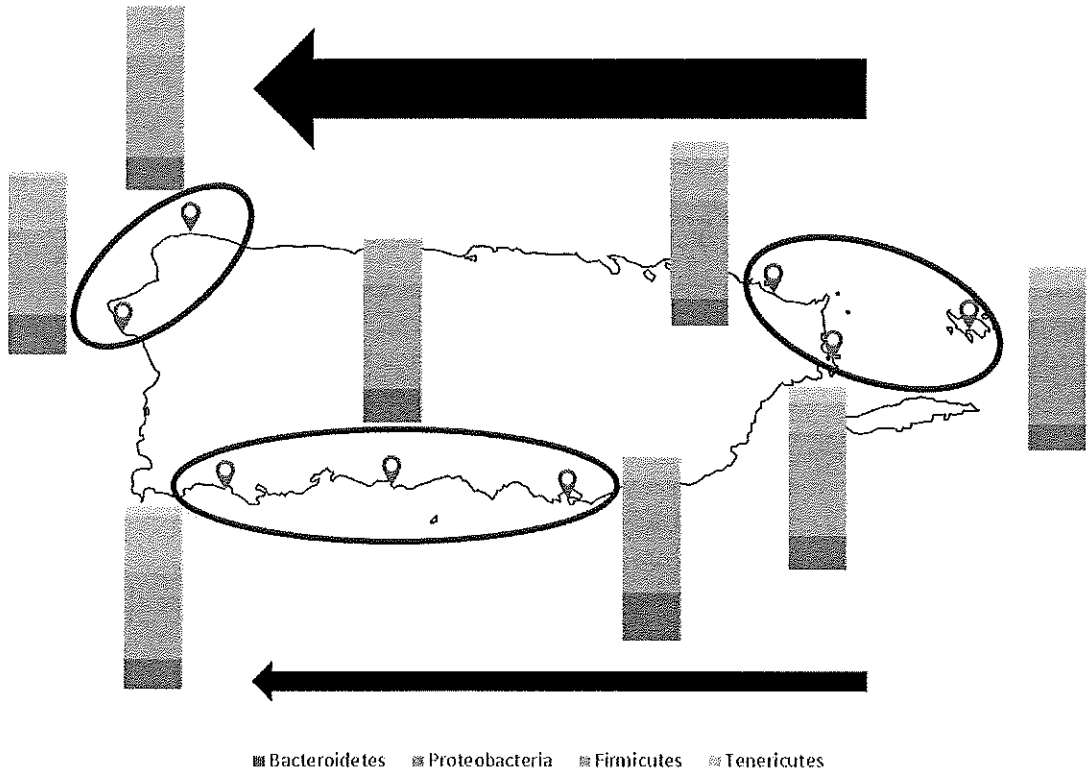
D

	Medium	Large
small	0.0017	0.7097
medium		0.0000

Figure 8. *Gut microbiota of D. antillarum based on size and relative proportion.* In these figures, is depicted the relative taxonomic values for the different sizes in inches (A). Animal body diameters were measured and placed in the following categories one point five (n=3), two (n=6), two point five (n=15), three (n=16), three point five (n=3), four point five (n=1). Animals in the proportion category were placed into categories relative to their size, namely small (1.5 - 2 in), medium (2.5 - 3 in) and large (3.5 - 4.5 in). The animal sample numbers include small (n=8), medium (n=31) and large (n=4). A pairwise statistical analysis was performed using Pearson Chi-squared analysis for the size (C) and the proportion (D).

An additional chi-squared analysis was performed by grouping the taxonomic microbial data into groups based on island surface current, which is known to sweep across the island from the east towards the west, with a stronger current sweeping across the north of the island (**Figure 9**). The previous analysis of grouping the taxonomic microbial data samples into the collection sites according to North, South, East and West indicated that there was no difference in the microbial profiles for samples from the North vs. South, but there was a highly significant difference between microbial profiles for samples from the East vs. West. Thus, we grouped taxonomic microbial data samples according to East (Luquillo, Ceiba and Culebra), West (Rincon and Isabella) and South (Guanica, Ponce and Guayama). Pairwise chi-squared analyses indicated that there was a significant difference between the microbial profiles in the samples representing the East vs. West. Although not significant, there was a trend indicating that the microbial profiles between samples representing East vs. South were also different.

A



B

West-South	0.79821168
West-East	0.00819983
South-East	0.05602085

Figure 9. *Gut microbiome of D. antillarum by sampling location and island wide currents.* The bars in the map represent the relative taxonomic profile of each municipality (A). Animal samples were collected from Rincon (n=10), Guanica (n=3), Ponce (n=3), Isabella (n=5), Luquillo (n=5),

Culebra (n=7), Ceiba (n=9), Guayama (n=2). Samples were grouped into island wide currents (red circles) East (n=21), South (n=8), and West (n=15). Surface water current strength and direction is indicated using the thickness of the arrows and direction of the arrow, respectively. In the top arrow the current corresponds to the north Atlantic Ocean water currents which are stronger than the bottom calmer waters of the Caribbean Sea. P-values are shown in the bottom table (B).

A heatmap was generated to visualize the overall results of the different metadata groupings by municipalities, cardinal location, cardinal surface current, size, and reef habitat (**Figure 10**) in comparison to the taxonomic microbial data by phyla. As represented in **Figure 10**, there is no clustering of the taxonomic data among the samples pertaining to Proteobacteria and Bacteroidetes, which is similar to the data shown in **Figure 4**. In addition, Firmicutes are also largely dispersed across the different samples; however, this refers to reads that were classified at the class level Clostridiales. Within Clostridiales, the family Lachnospiraceae was represented in the sample labeled as A4 collected in Rincon, and the family Ruminococcaceae was represented in samples labeled as A9 (Rincon), D3 (Luquillo), and G1 (Ceiba). Finally, taxonomic data pertaining to the phyla Tenericutes was clustered for several samples, which actually refer to the six samples (A1, D4, E4, E6, G7, and H2) that contained bacteria of the species *Candidatus hepatoplasma*.

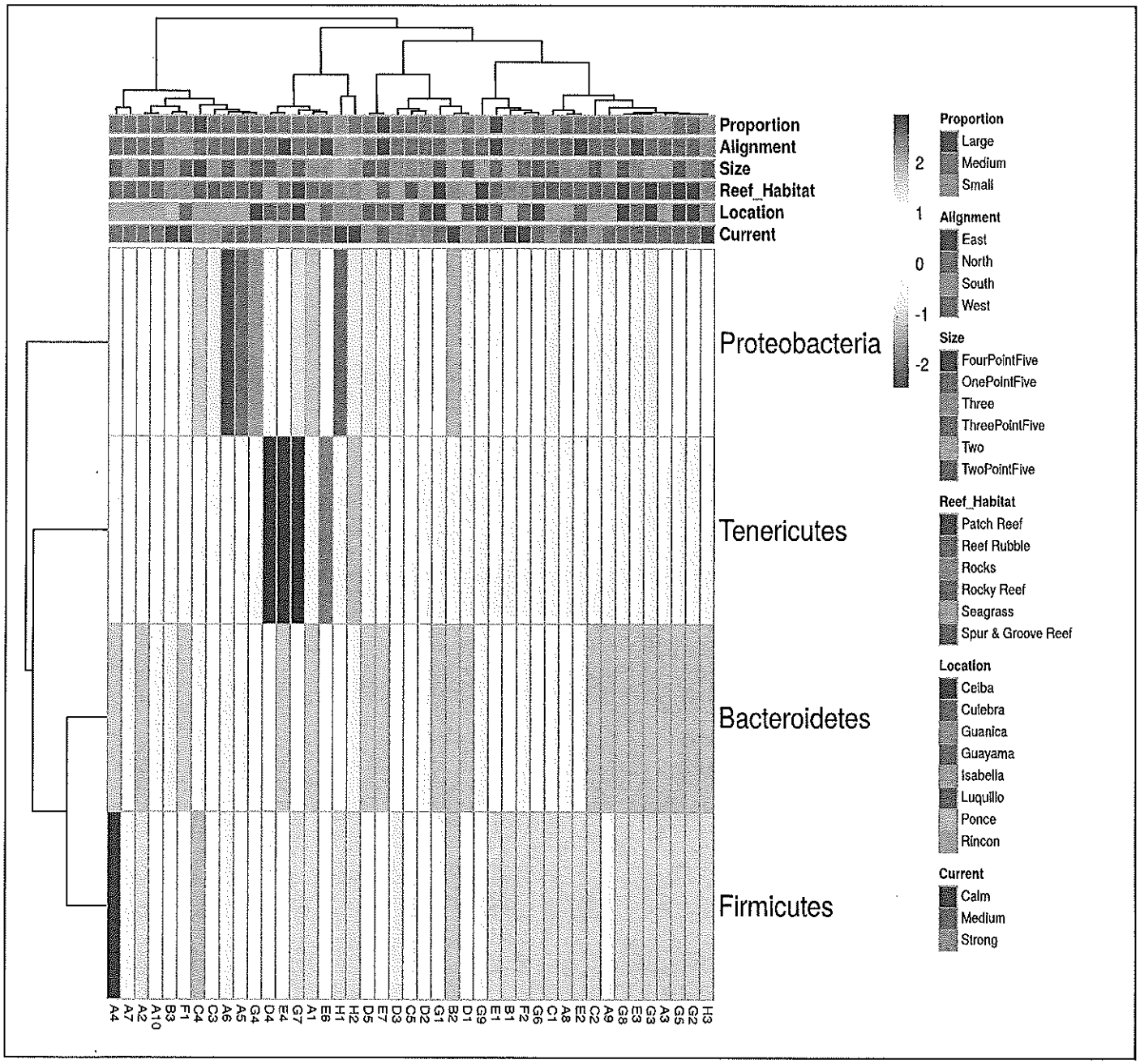


Figure 10. *Heatmap generated using ClustVis*. This image shows a heatmap generated using multiviance analysis that includes the metadata and relative taxonomic representation of each sample (Mersalu & Vilo 2015). The top rows represent the different metadata categories for each sample which are the columns represented by the identifiers at the bottom. In the following rows the taxonomic groups are represented. The relative presence in each sample is indicated by the intensity of the color. To the left and at the top appears a hierarchical category of the groups.

Population genetic diversity across the animal samples

Analysis of population genetic diversity in *D. antillarum* from eight municipalities across Puerto Rico according to *Cytochrome B* indicate there is substantial gene flow between animals at the different cardinal location sites (**Figure 11**). According to the phylogenetic tree shown in **Figure 11** (left), there is no formation of clades restricted to locations pertaining to animal samples characterized as North (Luquillo and Isabella), South (Guanica, Ponce and Guayama) East (Ceiba and Culebra) and West (Rincon). Rather, clades have formed between animals at various cardinal locations (including animals collected at sites labeled as West and North, South and North, South and West, West and East, and East and North) with no clear animal migration pattern between these locations. The phylogenetic tree is also paired with the *16S* sequencing results by taxonomic classification **Figure 11** (right) according to the animal sample shown in the tree. To test for patterns of *16S* taxonomic diversity between the four major clades in the phylogenetic tree (**Figure 11** left; see asterisks), Simpson and Shannon diversity indices were calculated to test differences between these clades. Results from these analyses indicated there were no significant differences between the species represented in the samples between clades and the abundance of these species between the clades (results not shown).

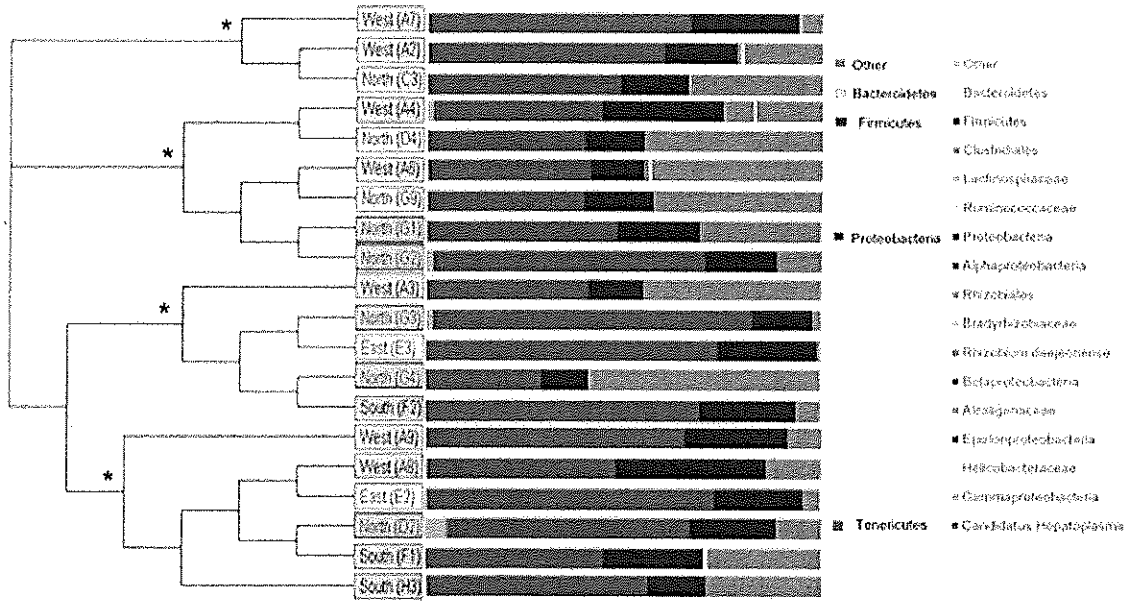


Figure 10. *Phylogenetic tree of Cytochrome B (left) with 16S sequence results (right) by individual animals.* Animals are labeled at tree nodes according to cardinal location site of collection and animal identification (see **Figure 4**). A Tamura-Nei Neighbor-Joining phylogenetic tree was generated using Geneious Prime software and the taxonomic classification was generated in QIIME2. There are four major clades represented in the tree, each marked with an asterisk.

Discussion

It is well documented that the coral reef ecosystem has been in decline for decades and that this has disastrous consequences for the biodiversity that inhabits this ecosystem. The great *D. antillarum* die off in the 1980's had serious consequences that also contribute to this ecosystem decay. We know very little of what caused the great die off, but it is believed that a water borne pathogen was responsible for the event (Lessios et al., 1984). In 1987, an experiment conducted by Bauer found bacterial isolates from *D. antillarum's* digestive tract and gonad tissues which were confirmed to be strains of *Clostridia perfringes* and *Clostridia sordellii*. It is well documented that *Clostridia* strains are commonly associated with polluted estuarine water (Daily et al., 1981; Watkins and Cabelli, 1985). In this experiment (Bauer 1987), the Clostridium bacteria were found to be highly infectious especially at a higher temperature. The correlation between temperature and pathogenicity has been studied in recent investigations involving the sea urchin *L. variegatus* (Brothers et. al. 2018) and it has been proposed that perhaps at higher temperature sea urchins experience a destabilization of the microbiome (dysbiosis). In this same study also, there is some evidence of an increase of virulence of opportunistic organisms inside the microbiome of the *L. variegatus* as it has been reported that some bacteria respond with pathogenicity with increasing temperature. In another sea urchin *Heliocidaris erythrogramma*, elevated temperatures have been observed to cause a degradation of the coelomocyte concentrations which impacts overall immunity (Brothers et. al. 2016). In the case of our study, it is probable that Clostridium is the number one suspect of the great D. antillarum die off. This bacterium reproduces by spores and can travel large distances with this apparatus (Bauer 1987). Perhaps this added synergistic effects of elevated sea temperatures and a highly virulent strain of Clostridium could present a challenge for the survival of *D. antillarum* in the wild.

Previous studies done on the sea urchin digestive tissue have found distinct bacteria not found in outer environment or ingested feed (Guerinot and Patriquin 1981; Hakim et al., 2015). Bacteroidetes and Proteobacteria have been found in the guts of the mussel *Mytolus galloprovincialis* (Li etal 2019), a small abalone *Haliotisdiversicolor* (Zhao et al., 2018), a crab *Callinectes sapidus* (Givens et al., 2013), a sea cucumber *Holothuria glaberrima* (Pagan-Jimenez 2019) and more importantly other sea urchins such as *Lytechinus variegatus* (Hakim et al., 2015) and *Strongylocentrotus purpuratus* (Hakim et al 2019). This suggest that the Bacteroidetes and Proteobacteria found in *D. antillarum* are not antagonistic to the host. This is because there is a near-dominant and consistent presence of these bacteria that can support a host-selection mechanism and there is evidence for a non-detrimental association for other orders of bacterium in sea urchins along with bioinformatical predictions of the role of this bacterium that show it might be related to symbiont-host energy metabolism (Hakim et al. 2015). Bacteroides have been found in polluted waters as well (Daily et al., 1981) however they do seem to be also associated with marine microbiomes (Li etal 2019). An important result was not finding Campylobacteraceae or Vibrio which are both bacterium that have been found in various sea urchins (Guerinot et al.1982; Hakim et al.2015). This could suggest that either these bacteria do not play an important role in the microbiome of *D. antillarum* or the organism is experiencing a form of dysbiosis which negatively affects the relationships between the symbiont and host.

The pairwise comparison between the different categorical findings suggest various clues about how the life of the Puerto Rican *D. antillarum* population. The size and proportion grouping categories show significance with potentially small and medium specimens. This suggest that there might be a maturation of the microbiome, however more research is needed for this. Also, who is responsible for the adaptation the microbiome or the sea urchin?

Other findings in the location-based and current-based categories paint a different picture. Guayama is inherently different from the rest of the other municipalities possibly due to the high level of anthropocentric activity. In the cardinal categories, east seems to be most different from the rest of the cardinal points, this might be due perhaps to the currents in the island originate from the eastern region and drag across the island. However, we did not find any statistical significance between the current categories based on strength. Based on this we grouped together the locations into three to form the island wide current category and we had a statistical significance between east and west. This finding suggests that the drag of the current across the island makes the west population different from the east population. Could this suggest that the specimens found in the east are healthier than the other sides that get bombarded with island debris. Also, the south and east categories were almost statistically significant that perhaps can be addressed in future bioinformatic studies.

Having accurate information about the genetic diversity, gene flow, and microbiome of key species of this environment is essential information to conserve the species if another disastrous mortality event were to occur. ~~Here~~ In this study, we unlocked the mysteries of one of the most important microbiomes of the coral reef ecosystem helping us understand the ecosystem a little bit further. In summary, the results of this study reveal the genetic diversity of the sea urchin *D. antillarum* via the sequencing of its microbiome. The sea urchin has a highly compartmentalized gastrointestinal tract along with unique microbial profiles for each compartment that can indicate a specific functional profile for that bacteria, as shown in previous studies (Hakim et. al 2016). The bacteria of the gastrointestinal tract play a huge role in the digestion of complex sugars and cellulose especially considering both the physiological limitations of the sea urchin gut and the low nutritional value of seagrass (Arafa et al. 2012). Further studies can reveal the functional genome of some of these bacteria which can then reveal more about the function in the sea urchin and can be a source of antibiotics or synthetic functions via genetic engineering. Also it is important to note that anthropocentric induced climate change along with the prevalence of disease in coral and other important benthic coral reef community members hinders the positive effect of keystone animals such as *D. antillarum* on the coral reef ecosystem (Burge et. al 2014; Carperter & Edmund 2006) thus taking care of that ecosystem is our priority.

Audrey said something about the unknowns - are these areas for further research

Add observations regarding color