Nanopore Sequencing for the Metagenomic Analysis of the Gut Microbiota of the Millipede *Anadenobolus monilicornis*

by

Orlando J. Geli-Cruz

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

MASTER OF SCIENCE
IN
BIOLOGY

University of Puerto Rico
Mayagüez Campus
2019

Approved by:

_____	_____
Alex Van Dam, Ph.D. Graduate Committee Chair	Date
_____	_____
Matías J. Cafaro, Ph.D. Graduate Committee Member	Date
_____	_____
Carlos Santos-Flores, Ph.D. Graduate Committee Member	Date
_____	_____
Raúl E. Macchiavelli, Ph.D. Representative of Graduate Studies	Date
_____	_____
Ana Vélez-Díaz, M.S. Department Director	Date

Abstract

Millipedes act as important components of leaf litter decomposition in terrestrial ecosystems, thanks in part to the microbial diversity in their guts. However, millipedes and their gut microbiota are understudied, compared to other arthropods. For this reason, we designed a protocol for the extraction, sequencing, and shotgun metagenomic analysis of the gut of *Anadenobolus monilicornis*. We collected specimens of *A. monilicornis* from different municipalities of Puerto Rico. We extracted their guts and sequenced the DNA with Oxford Nanopore Technologies' MinION nanopore sequencer, then analyzed the data using the programs Phylosift and MEGAN6, and the MG-RAST online server. From our two successful samples from the municipalities of Mayagüez and Rincón, we obtained a total of 87,110 and 99,749 reads, respectively. We found that many of our assigned bacteria reads were annotated to the phyla Proteobacteria, Bacteroidetes, and Firmicutes; the Mayagüez sample had much more Chlamydiae representation, however. Sampled eukaryote phyla include Arthropoda, Streptophyta, and Chordata. Judging by the fact that some of the reads were annotated as belonging to Chordata, coupled with the lack of Nematoda and Ascomycota reads, we concluded that some missannotation may have occurred. We would need a larger sample size to better identify the intestinal microbial taxa, as we were setback by the landfall of Hurricane Maria in 2017. Considering our small sample size, however, we were able to develop an initial fingerprint of the millipede intestinal microbiota using shotgun metagenomics and nanopore sequencing.

Resumen

Los milpiés actúan como componentes importantes de la descomposición de la hojarasca en los ecosistemas terrestres, debido en parte a la diversidad microbiana que se encuentra en sus intestinos. Sin embargo, los milpiés y su microbiota intestinal no han sido estudiados tanto como en otros grupos de artrópodos. Por esta razón, desarrollamos un protocolo de extracción, secuenciación y análisis metagenómico del intestino de *Anadenobolus monilicornis*. Colectamos especímenes de *A. monilicornis* de diferentes municipios de Puerto Rico. Extrajimos sus intestinos y secuenciamos su ADN usando el secuenciador de nanoporos MinION de *Oxford Nanopore Technologies*, para luego analizar los datos usando los programas Phylosift y MEGAN6, y el servidor en línea MG-RAST. De las dos muestras exitosas, una de Mayagüez y otra de Rincón, obtuvimos un total de 87,110 y 99,749 lecturas de ADN, respectivamente. Encontramos que la mayoría de las lecturas asignadas a bacterias pertenecían a los filos Proteobacteria, Bacteroidetes, y Firmicutes; la muestra de Mayagüez tuvo mucha más representación de Chlamydiae. Algunos filos eucarióticos muestreados incluyen Arthropoda, Streptophyta, Chordata. Juzgando por el hecho de que obtuvimos lecturas asignadas a Chordata, además de la falta de lecturas de Nematoda y Ascomycota, llegamos a la conclusión que alguna anotación incorrecta de lecturas pudo haber ocurrido. Necesitaríamos una muestra más grande para mejor determinar los taxones microbianos intestinales, ya que nuestros planes fueron afectados por el paso del Huracán María de 2017. Considerando nuestro tamaño de muestreo pequeño, sin embargo, logramos desarrollar un perfil inicial de la microbiota intestinal de un milpiés usando metagenómica “shotgun” y secuenciación de nanoporos.

To my parents, Orlando and Josefina, and to my sister, María, for always having my back throughout my academic journey. To Marlys Massini and Beatriz Romero, for their advice and all the time they took to help me adjust to graduate studies and to the Entomology Lab. To the original housemates, for all the good times we had before going our separate ways. To the Biology Department staff and to the graduate students from Biology and across RUM, for being some of the friendliest, funniest and sweetest people I've met.

Acknowledgements

I would like to thank Marlys Massini, Beatriz Romero and Ruben Irizarry for taking part in the sampling and DNA extraction processes. I would like to thank Alex Van Dam, Matías Cafaro and Carlos Santos-Flores for making this project possible. I would also like to thank PSC staff Tom Maiden, Rick Costa, and Roberto Gomez for their help with installing albacore on the Bridges system. Finally, laboratory work was funded by a NIH PR-INBRE Grant Contract #5P20GM102475 awarded to AVD, and the bioinformatics was funded by an NSF-XSEDE Grant TG-BIO170059 awarded to AVD.

Table of Contents

List of Tables viii
List of Figures ix
1. Introduction 1
2. Literature Review 4
2.1 Metagenomics 4
2.2 Soil Arthropod Microbiota 5
2.3 Developments in Sequencing 5
2.4 Nanopore - MinION Sequencing 6
3. Materials and Methods 8
3.1 Millipede Sampling 8
3.2 Gut and DNA Extraction 8
3.3 DNA Fragmentation 9
3.4 NEBNext FFPB DNA Repair 10
3.5 NEBNext Ultra II End Repair / dA-Tailing module 10
3.6 Ligation of Barcode Adapter 11
3.7 Barcoding PCR 11
3.8 Ligation of Sequencing Adapter 11
3.9 SpotOn Flow Cell Prep 122
3.10 Quality filtering and de-multiplexing 12
3.11 Metagenomic Analyses 12
4. Results 14
4.1 Quality filtered reads and summary statistics 14
4.2 Bacterial Reads 188
4.3 Eukaryotic reads 211
4.4 Metabolism annotation 233
5. Discussion 266
6. Conclusion 31
7. Literature Cited 32

List of Tables

Table 1. Summary Statistics for the quality-filtered *A. monilicornis* sample data 15

Table 2. Comparison of annotated reads from the mayor bacterial phyla sampled from *A. monilicornis* guts, based on the program used 20

Table 3. Comparison of annotated reads from the mayor eukaryote phyla sampled from A. monilicornis guts, based on the program used 22

List of Figures

Fig 1. Read length and Phred quality scores for the Mayagüez sample **14**

Fig 2. Read length and Phred quality scores for the Rincón sample **15**

Fig 3. Sampled phyla representation from the Mayagüez and Rincón sample combined, using MEGAN6. Phyla marked with darker shades of green had a larger number of assigned reads **16**

Fig 4. Rarefaction plot for the Mayagüez sample (blue) and the Rincón sample (yellow), created using MEGAN6 ..... **17**

Fig 5. Rarefaction plot for the Mayagüez and Rincón samples combined, created using MEGAN6 ..... **17**

Fig 6. Phylosift results of Bacterial diversity chart from the Mayagüez sample (left) and the Rincón sample (right) **19**

Fig 7. Phylum representation chart for the Mayagüez sample (left) and the Rincón sample (right), created by uploading the sample data to MG-RAST **19**

Fig 8. Eukaryotic diversity chart for the Mayagüez sample (left) and the Rincón sample (right), created using Phylosift data. Eukaryotes represented 12% of the sampled reads in the Mayagüez sample, and 4% in the Rincón sample..... **21**

Fig 9. KO metabolism analysis for the Mayagüez sample (left) and the Rincón sample (right), obtained from MG-RAST **23**

Fig 10. Metabolic pathways for the Mayagüez sample, made by uploading a KO list to iPath3 ..... **24**

Fig 11. Metabolic pathways for the Rincón sample, made by uploading a KO list to iPath3 ..... **25**

Introduction

Millipedes are a group of arthropods belonging to the class Diplopoda and the subphylum Myriapoda. With around 12,000 described species worldwide, they are a diverse group found in a variety of habitats from humid rainforests to xeric deserts (Crawford et al., 1983; Taylor, 1982; Vélez, 2014). Millipedes are one of many soil-inhabiting decomposers, alongside other macroinvertebrates such as earthworms and isopods (Kitz et al., 2015; Pitz & Sierwald, 2010; Snyder & Hendrix, 2008). They are considered in some ecosystems as one of the more important components of terrestrial litter decomposition and nutrient cycling (Kitz et al., 2015; Pitz & Sierwald, 2010; Snyder & Hendrix, 2008). However, they are a comparatively understudied arthropod group and have received little attention to elucidate their interactions with their gut microbial community, despite their significance in nutrient cycling at the soil leaf litter interface (Brewer et al., 2012; Sierwald & Bond, 2007).

In comparison, the microbial composition of the gut of other leaf litter macroinvertebrates has received some attention from the scientific community. Termites, for example, are known to harbor diverse clades of bacteria from the phyla Bacteroidetes, Firmicutes and Spirochaetes in their guts, depending on the diet and gut compartments studied (Rossmassler et al., 2015). The guts of terrestrial isopods, which also have a similar ecological role to millipedes, are well represented by Proteobacteria (Dittmer et al., 2016). In a similar fashion, millipedes have been shown to host certain microorganisms in their guts, which aid in digestive processes (Szabo et al., 1990). Dilution plating techniques have shown that some millipede species harbor an abundant diversity of proteobacteria and actinobacteria (Byzov, 2006).

There have been a small number of microbial community surveys of the gut of millipedes. In some species, the most dominant bacteria were found to belong to the Enterobacteriaceae family; in addition, ascomycetes were the most common yeast strains found (Konig, 2006). A desert-dwelling millipede species possesses gut bacteria that can degrade cellulose, contributing to nutrient cycling in deserts (Taylor, 1982). Certain species from the millipede orders Julida, Spirobolida, and Spirostreptida harbor an association between methanogenic archaea and ciliate protozoa in their hindguts, contributing to methane production (Sustr et al., 2014). The diversity of bacteria and other microorganisms that occur in millipede guts might be of interest to field ecologists and microbiologists alike, as the interactions between these organisms affects both soil nutrient recycling and organic matter decomposition (David, 2014; Knapp et al., 2009). A full genetic or metagenomic approach to these problems, however, has yet to take off.

For this study, we will be focusing on the microbiota that inhabits *Anadenobolus monilicornis*'s digestive tract. *A. monilicornis* is a species native to the Caribbean which has also been introduced to Florida (U.S.A.), where it is treated as a pest (Gabel et al., 2006; Shelley, 2014). It is considered the most common millipede in the karst zones of Puerto Rico (Vélez, 2014). Previous microbiota studies have been morsphohology based for *A. monilicornis*; specifically, Contreras & Cafaro (2013) conducted a morphometric study of the protozoa *Enterobryus luteovirgatus*, which forms a commensalistic relationship with the millipede. Beyond this study, very little is known about microbes in this species.

To identify the microbial diversity inside *A. monilicornis*'s gut, we will utilize a shotgun metagenomic analysis. Metagenomic sequencing of environmental samples, most notably from microbial communities (Thomas et al., 2012), can lead to the discovery of microorganisms that are otherwise difficult or impossible to culture in a laboratory setting (Qi et al., 2009). Several

metagenomic sequencing studies have successfully revealed complex host-symbiont relationships, such as those that occur in deep-sea tube worms (Robidart, 2006), termites (Warnecke et al., 2007) and *Daphnia* species (Qi et al., 2009). Shotgun metagenomics, which consider the genomes of all organisms within a sample, has allowed scientists to determine taxa from samples such as honey (Bovo et al. 2018) and even ethanol used to preserve insects (Linard et al. 2016). In general, metagenomic studies involving arthropods have focused mostly on insects (Engel et al., 2012; Muturi et al., 2017; Yun et al., 2014), successfully revealing their microbial diversity. However, there are few comprehensive metagenomics sequencing surveys of the microbiota of other non-insect arthropods, millipedes included (Bouchon et al., 2016; Degli & Martinez, 2017).

To analyze the DNA of *A. monilicornis* and its microbiota, nanopore sequencing will be used. This relatively new technique works by identifying the order of nucleotides in a DNA sequence as it passes through individual nanopore channels. Nanopore sequencing has already proven useful for real-time analysis (Jain et al., 2016). Specifically, we will be using the MinION, a portable nanopore sequencer developed by Oxford Nanopore Technologies (ONT). Our primary objective is to develop an initial fingerprint of the millipede gut microbiota via a shotgun metagenomic analysis, using nanopore sequencing. This will allow us to accomplish our secondary objective: to determine the major components of microbial taxon diversity within the gut of *A. monilicornis,* whether they be bacterial, eukaryotic, or archaeal.

Literature Review

Metagenomics

Metagenomic studies involve DNA sequencing, comparisons between genomic datasets and a variety of computer programs to manage the sequencing data. This process of extracting, analyzing and comparing genetic material directly from natural habitats has led to several unique findings. These include the identification of several microorganisms that are difficult to culture or cannot be cultured in a laboratory setting (Qi et al., 2009) and horizontal gene exchange, or the transfer of genes between populations (Guo et al., 2015) or even between species (Jain et al., 2002). Analysis of this kind has been done on microbial communities associated with different types of animals, including many invertebrates: Robidart (2006) studied the bacteria *Endoriftia persephone* and its metabolic versatility, as well as its symbiotic relationship with its host, *Riftia pachyptila*, a deep-sea tube worm that lives around hydrothermal vents. This study was successful in overcoming the challenge of studying an organism in the relatively inaccessible and extreme biological system that is the deep sea. Qi et al. (2009) studied symbionts in three *Daphnia* species and found similarities between the symbiont communities in all three species. Up to 123 distinct bacterial genera were found to be living in and on these planktonic arthropods. Woyke et al. (2010) analyzed the symbiotic bacteria that live in *Olavius algarvensis*, a mouthless, gutless worm that solely depends on its symbionts to survive. The study found that the bacterial symbionts do not share an obligate relationship with the worm. However, the bacteria "may be in transition to an obligate stage" (Woyke et al., 2010), possibly hinting at a new symbiotic bacteria species emerging in the future.

Soil Arthropod Microbiota

Arthropod gut microbiota has been studied extensively in the past. Most of these studies have focused on insects, including bees (Engel et al., 2012), moths (Chen et al., 2016), mosquitoes (Muturi et al., 2017), and several others, showing the great microbial diversity hidden within their digestive tracts (Yun et al., 2014). Soil arthropods with a similar ecological role to millipedes have also been studied: Warnecke et al. (2007) and Rossmassler et al. (2015) studied termites, insects whose gut microbiota are found nowhere else on Earth (Brune, 2006) and could serve as "potential sources of biochemical catalysts for efforts aimed at converting wood into biofuel" (Warnecke et al., 2007). Warnecke et al. focused on an arboreal species related to *Nasutitermes ephrata* and *N. corniger* and found bacterial diversity from 12 different phyla and several gene modules relevant to cellulose hydrolysis. On the other hand, Rossmassler et al. found a varying representation of bacterial phyla based on the diet of the termite studied, whether it was wood, soil, humus or detritus-based. Dittmer et al. (2016) studied the microbiota of terrestrial isopod species and found different microbial communities between different populations of the same species. Finally, although there are few comprehensive microbial surveys of the millipede gut, some have found no major differences in bacterial diversity when comparing specimens with different diets (Knapp et al., 2009). Judging by the different discoveries done concerning other arthropods' gut microbiota, millipede gut microbiota studies could help broaden our understanding of arthropod microbiota diversity and host-microbiota interactions.

Development in Sequencing

Sequencing procedures and technology have been steadily developing over the years, with different techniques used for studying bacterial-animal relations: First generation sequencing, as its name suggests, refer to the first sequencing techniques devised and include "Sanger

sequencing" and "Gilbert sequencing"; both of these were a major improvement from previous techniques, which were laborious and required radioactive materials (Liu et al., 2012). Second and next-generation sequencing were devised later and improved further upon DNA sequencing. "Pyrosequencing", a type of second-generation sequencing, has been used to study human gastrointestinal microbiota; Andersson et al. (2008) managed to produce high-fidelity taxonomic classifications of the bacterial diversity within the human gut. "Illumina sequencing", a type of next-generation sequencing, has been used to study animal-bacterial symbiosis via genome-centric approaches, allowing researchers to identify the metabolic pathways and genes that are most relevant to the symbiotic relationship (Kumar & Blaxter, 2011). Illumina sequencing has also been used to assemble the mitochondrial genomes of two millipede species, whose gene order was discovered to be "novel among known myriapod mitochondrial genomes" (Dong et al., 2016). Nanopore sequencing is a next-generation sequencing technique, which relies on the electrical detection of single DNA strands in contact with a series of pores. This technique allows for the sequencing of long strands of DNA without sacrificing the quality of the fragments (Loman & Quinlan, 2014). Oxford Nanopore Technologies has developed the MinION, a portable nanopore sequencer that can be used in field and lab alike; the MinION sequencer has even been tested in microgravity, with good results (McIntyre et al., 2016).

Nanopore - MinION Sequencing

The Oxford Nanopore Technologies (ONT) MinION is a relatively new sequencing device which has already garnered good results: Laver et al. (2015) tested the MinION in its early-access phase by re-sequencing bacterial genomes with different nucleotide compositions. While they found the error rate limiting, the MinION proved to be able to generate extremely long reads and improve genome assembly contiguity (Laver et al., 2015). Batovska et al. (2017) used the MinION

to successfully detect the presence of the Ross River Virus (RRV) in infected mosquitoes with minimal sample purification. These results vouch for the ability of the MinION to do real-time sequencing, which could push it to become an ideal in-field device for biosecurity surveillance (Batovska et al., 2017). Schmidt et al. (2017) generated a nanopore sequencing dataset for *Solanum pennellii*, a wild tomato species, using MinION SpotOn Flow Cells. After some "polishing" with Illumina short read data, the resulting genome assembly had a gene completeness of 96.5%, more than the *S. pennellii* reference genome used, proving that the long-read sequencing data could be used to "affordably sequence Gigabase-sized diploid plant genomes" (Schmidt et al., 2017).

Materials and Methods

Millipede Sampling

Anadenobolus monilicornis millipedes were collected from the Puerto Rican municipalities of Mayagüez & Rincón. The millipedes were kept in small glass containers with moist filter paper without food for 24 hours (Mayagüez samples), and ten days (Rincón samples). This was done in order to eliminate intestinal contents, so as not to sequence ingested organisms.

Gut and DNA extraction

The gut and DNA extraction work were done in the Symbiosis laboratory at the University of Puerto Rico, Mayagüez Campus. Following workstation and lab material sterilization with 10% bleach, the head and the last two or three segments of the abdomen of the specimens were cut and removed with a scalpel. The abdomen was cut to facilitate gut extraction. The guts were removed and placed in 2mL tissue disruption tubes, where they were liquified by manually shaking the tubes. 0

We followed the Qiagen Fast DNA Tissue Kit (cat. No. 51404) protocol to purify the DNA samples from the specimens. The buffer for the Qiagen Fast DNA Tissue Kit was prepared before use as the protocol specified: 40 mL of ethanol were added to the AW1 and AW2 Buffer concentrates, and 25mL of isopropanol were added to the Buffer MVL concentrate. The tubes were spun down briefly via vortex mixer and set on a block heater at 56°C for one hour. The tubes were briefly spun down every 10 minutes for that hour. 265µL of the Buffer MVL mixture (200µL of AVE, 40µL of VXL, 20µL of Proteinase K, 1µL of DX Reagent and 4µL of RNase A) were added to the tubes.

For all samples, the mixtures were moved to a QIAamp Mini Spin Column and centrifuged for one minute at 15,000 rpm. The spin column was then placed in a clean 2mL collection tube, while the previous tube and filtrate were discarded. 500µL of the AW1 Buffer were added to the spin column before being centrifuging, again for one minute at 15,000 rpm; the spin column was placed in another 2mL collection tube, and the previous tube discarded. 500µL AW2 Buffer were added to the spin column before centrifuging. The spin column was then added into a new collection tube, which was centrifuged again for two minutes, and later placed in a clean 1.5mL microcentrifuge tube. 50µL of nuclease-free water was added directly into the spin column, which was left subsequently for one minute at room temperature and later centrifuged for one minute. This last step was repeated once to increase yield. After this step, the Oxford Nanopore Technologies (ONT) 1D PCR barcoding genomic DNA (SQK-LSK108) for version R9 chemistry procedure was followed, with some minor alterations.

DNA Fragmentation

A master mix of 14µL of Fragmentase buffer and 2µL of 10X NEBNext® dsDNA Fragmentase® (NEB cat. No. M0348s) was mixed first. In new tubes, we added 32µL of the samples and 8µL of the master mix to each. The new tubes were vortexed for two seconds and spun down; they were then placed on a thermocycler for five minutes at 37℃ followed by approximately five minutes at 4℃. In order to heat kill the Fragmentase, 5µL of EDTA was added and placed on a thermocycler for 15 minutes at 65℃ followed by 10 minutes at 5℃. We aimed to produce 5,000-30,000Kb DNA fragments. DNA quality was verified using 2µL of each sample mixed with 3µL of loading dye and then added to a 1X electrophoresis gel set to 66V for 30 minutes.

Leftover enzymes were cleaned via Agencourt® Ampure® XP beads: 50µL of samples of each sample were added to 90µL of Ampure XP beads, mixed 10 times by pipetting. The mixture was left at room temperature for five minutes, then placed on a magnetic rack for two minutes. The cleared solution was then aspirated out. The process was repeated but with 200µL of 70% ethanol followed by aspiration. Finally, 48µL of nuclease-free water was added, and aspirated out into new 1.5 mL tubes and carried forward in the protocol.

NEBNext FFPE DNA Repair

DNA was repaired via the NebNext FFPE DNA repair kit. 5.5µL of nuclease-free water, 6.5µL of FFPE DNA repair buffer, 2µL of NEBNext FFPE DNA repair mix (NEB cat No. M6630) and 53.5µL of the sample DNA were mixed. The samples were transferred to 0.2mL tubes for and placed in a thermocycler programmed to 20℃ for 15 minutes, followed by 4℃ for 10 minutes. The process ended with the previous Ampure XP beads cleaning procedure.

NEBNext Ultra II End Repair / dA-Tailing module

We used the NEBNext® Ultra™ II End Repair/dA-Tailing Module (NEB cat No. E7546). 5µL of nuclease-free water, 3µL of NEBNext Ultra II End Prep Enzyme mix, 7µL of NEBNext Ultra II End Prep Reaction buffer and 45µL of sample DNA were placed in 0.2mL tubes. The tubes were transferred to a thermocycler programmed for 20℃ for 20 minutes, followed by 65℃ for 5 minutes and finally 4℃ for a few minutes. The process ended with another round of Ampure XP bead cleanup as before.

Ligation of Barcode Adapter

Ligation of Barcode Adapters was performed with NEB Blunt/TA Ligase Master Mix (NEB cat No. M0367). 20µL of ONT ligation adapter (ONT cat No. Ligation Sequencing Kit SQK-LSK108), 50µL of NEB Blunt/TA Ligase master mix and 30µL of sample DNA were mixed by inversion in a 1.5mL tube. The tube was left at room temperature for 10 minutes, followed by AMPure XP beads cleaning procedure. The finished samples were then transferred to PCR tubes.

Barcoding PCR

2µL of PCR barcode from the PCR Barcoding Kit (ONT cat No. SQK-PBK004), 2µL of 10ng/µL adapter ligated template, 50µL of NEB LongAmp Taq 2X Master Mix (NEB cat No. M0287), and 46µL of nuclease-free water were mixed. The samples were placed on a thermocycler using the following cycling conditions: 95℃ for three minutes for initial denaturation, 95℃ for 15s for denaturalization, 62℃ for 15s for annealing and 4℃ on hold. The process ended with the Ampure XP beads cleanup procedure as before. DNA quality was verified again by adding 1µL of sample DNA into the NanoDrop spectrophotometer, which we used with the permission of Audrey Majeske, Ph.D.

Ligation of Sequencing Adapter

20µL of Adapter Mix, 50µL of Blunt/TA Ligation Master Mix and 30µL of end-prepped DNA were mixed. After ten minutes at room temperature, another round of Ampure XP bead cleanup was performed. The finished samples were then transferred to Eppendorf DNA LoBind tubes.

SpotOn Flow Cell Prep

We followed the ONT for the SpotOn Flow Cell version R9 chemistry (ONT cat No. FLO-MIN 107 R9). After extracting the buffer from inside the flow cell's priming port by pipetting, we mixed 480µL of Running Buffer with Fuel (RBF) mix with 520µL of nuclease-free water and added 800µL of this mixture into the priming port via pipette. 35µL of RBF with 2.5µL of nuclease-free water, 25.5µL of Library Loading Bead kit (ONT cat No. EXP-LLB001) and 12µL of the DNA library were mixed. 200µL of the priming mixture (RBF & nuclease-free water) was loaded into the flow cell via the priming port by pipetting, while 75µL of the sample were loaded via the sample port in a dropwise fashion. The MinION was connected to a local MacBook, and the MinKNOW software program was accessed to start a sequencing run for 48 hours. Having obtained the data in the form of shotgun single long reads, we used the MinKNOW software to acquire and analyze the sequencing data obtained from the MinION. The libraries were then sequenced again on a second Flow Cell.

Quality filtering and de-multiplexing

We used the MinKNOW software program for initial quality filtering of reads obtained from the second Flow Cell. The HDF5-formatted data from the nanopore sequencer was moved from the MacBook to the Pittsburgh Supercomputing Center's (PSC) Bridges Supercomputer. Within the PSC and using the Anaconda and Python environments, we installed the albacore basecaller v-2.1.3 (Oxford Nanopore Technologies, 2017) to separate the different barcodes and convert the data to FASTQ format (Oxford Nanopore Technologies, 2017). Read length and Phred quality scores were calculated using pauvre. We used KmerGenie to predict k values for our datasets in order to attempt optimizing the genome assembly process (Chikhi & Medvedev, 2014). We ran velvetg and velveth to attempt a *de novo* genome assembly (Zerbino & Birney, 2008). This

assembly was unsuccessful. We also tried Canu (Koren et al., 2017), but it also did not produce any scaffolds. The ONT data did not have enough depth of coverage to produce a *de novo* assembly.

Metagenomic Analyses

To summarize the diversity and relative abundance of the community of microbes sequenced, we used a variety of metagenomic classification programs for our long-read data. We chose to use programs that should work well with shotgun long-read data produced by the ONT MinION sequencer. We used Phylosift to reconstruct a phylogenetic tree and place our organisms found in the sample (Darling et al., 2014). We used BLASTn to align the sequenced data to the NCBI database (Altschul et al., 1990). We imported the data to MEGAN6 to do taxonomic, functional and comparative analysis of said data (Huson et al., 2007). We used the default LCA parameters, changing the minimum bit score ("Min Score") to 115.0, the LCA algorithm to "longReads" and the percent to cover to 80.0. Finally, we uploaded our data to the MG-RAST server to analyze the metagenome and annotate the genes to their respective organisms and to metabolic processes; to compliment the latter, we also uploaded our sample data to GenomeNet to obtain KO lists for the samples. KO lists were then uploaded to iPath3 to analyze the sampled metabolic pathways (Meyer et al., 2008; Yamada et al., 2011).

Results

Quality filtered reads and summary statistics

We were able to de-multiplex the data from the two millipede gut samples via albacore, to which we will be referring to as the Mayagüez and Rincón samples. Read length and Phred quality scores were calculated using pauvre (Fig 1 & 2). For the Mayagüez sample, we obtained a total of 87,110 quality-filtered reads; for the Rincón sample, a total of 99,749 reads. For more summary statistics, refer to Table 1. Phylosift matched 298 reads (261 bacterial, 36 eukaryotic) for the Mayagüez sample, and 48 reads (45 bacterial, 2 eukaryotic) for the Rincón sample. MG-RAST assigned taxonomic groups to 1,277 total reads for the Mayagüez sample, and 780 total reads for the Rincón sample. Finally, the MEGAN6 analyses were able to utilize 4,698 reads in total of which 3,847 were unassigned for the Mayagüez sample (Fig 3), and 5,626 reads in total with 4,495 reads being unassigned for the Rincón sample (Fig 4).

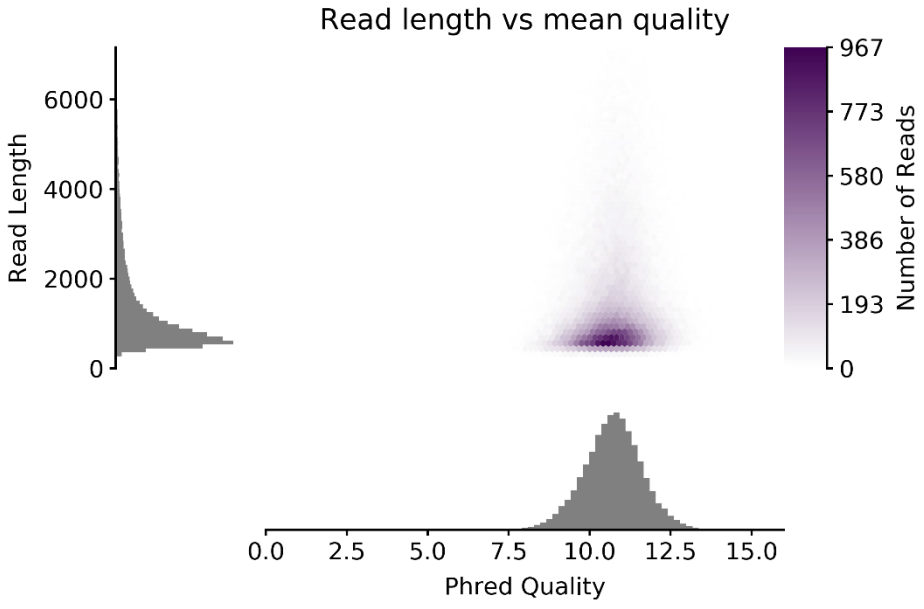

Fig 1. Read length and Phred quality scores for the Mayagüez sample.

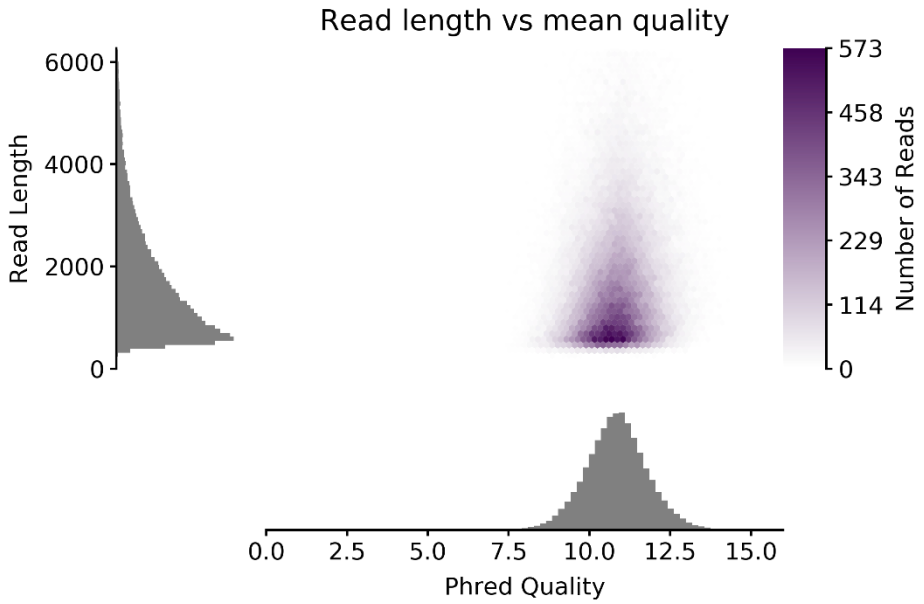

Fig 2. Read length and Phred quality scores for Rincón sample.

Table 1. Summary statistics for the quality-filtered *A. monilicornis* sample data.

	Mayagüez sample	**Rincón sample**
Reads	87,110	99,749
Base Pairs	132,196,067	176,209,113
Mean Length	1,517.6	1,766.5
Median Length	959.0	1,388.0
Min. Length	187	137
Max. Length	18,142	14,296

Through MEGAN6 we produced a summary taxonomic tree of the phyla sampled showing the distribution of reads across phyla (Fig 3). With MEGAN6 we also created rarefaction plots for the two separate samples and for both samples combined; the plot for the separate samples starts to plateau around 20 phyla for the Mayagüez sample, and around 15 phyla for the Rincón sample (Fig 4). Finally, the combined sample plot starts to plateau around 14 phyla (Fig 5).

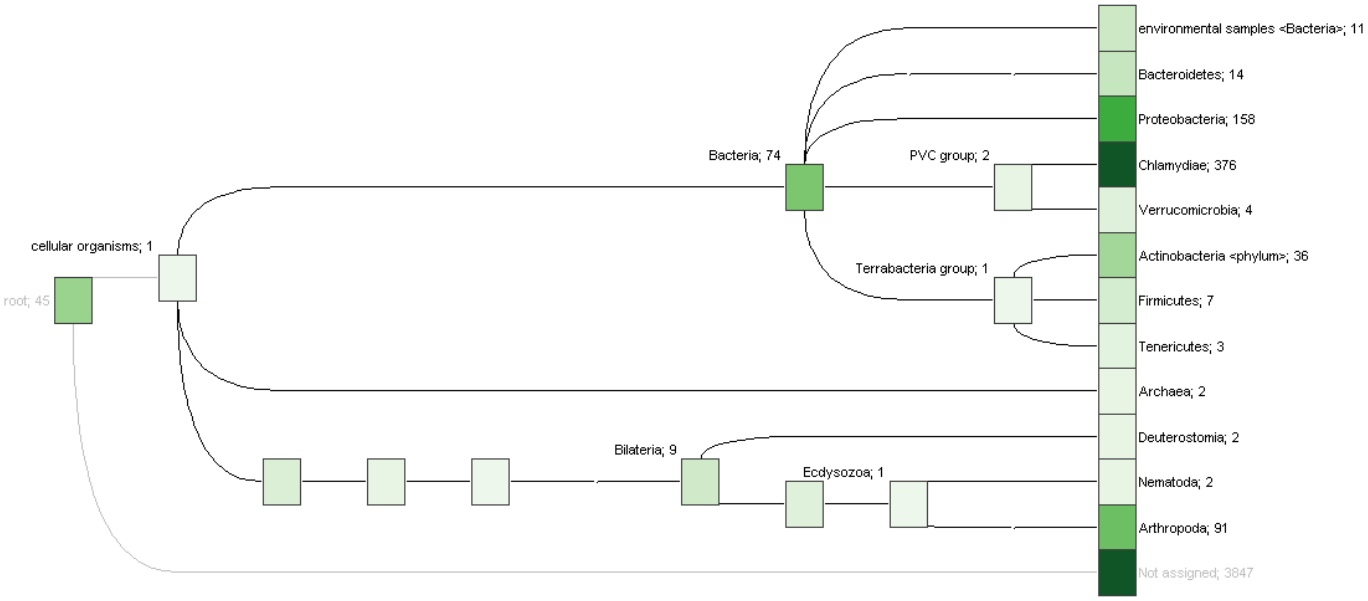

Fig 3. Sampled phyla representation from the Mayagüez sample, using MEGAN6. Phyla marked with darker shades of green had a larger number of assigned reads.

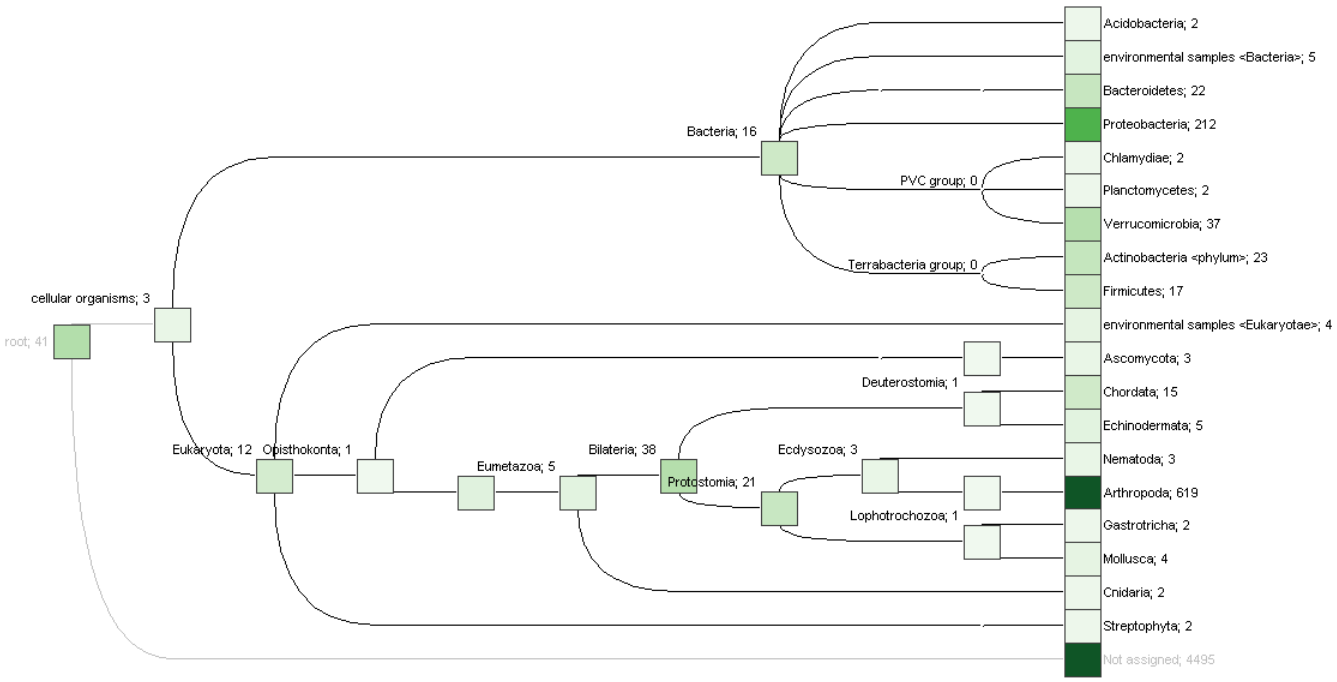

Fig 4. Sampled phyla representation from the Rincón sample, using MEGAN6. Phyla marked with darker shades of green had a larger number of assigned reads.

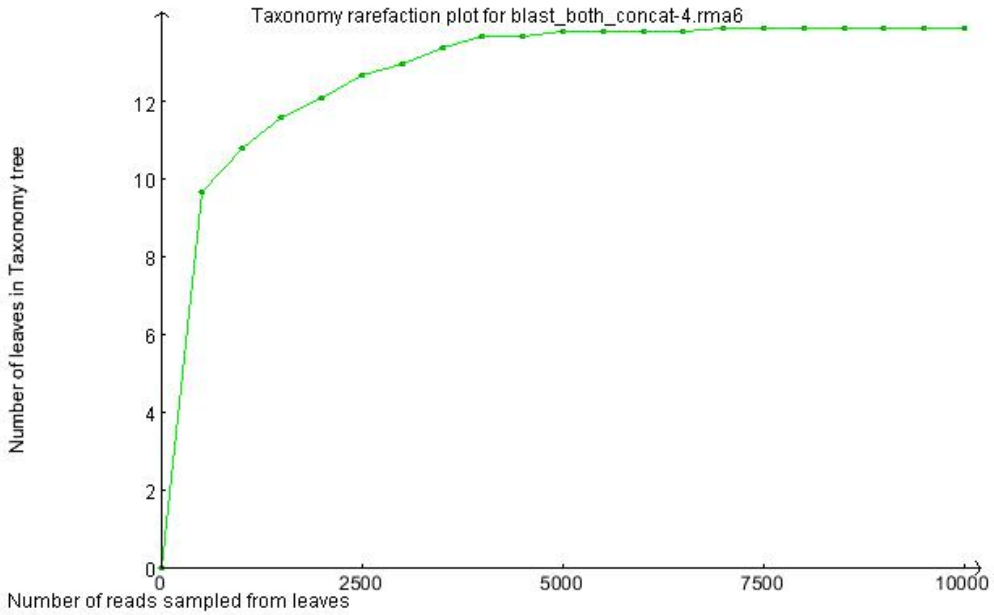

Fig 5. Rarefaction plot for the two *A. monilicornis* samples, created using MEGAN6.

Bacterial Reads

Across the three metagenomics summary analyses (Phylosift, MEGAN6 and MG-RAST), the two samples showed similar annotated bacterial phyla: both samples had Proteobacteria, Firmicutes, and Bacteroidetes representation. Many bacterial reads from the Mayagüez sample were assigned to the phylum Chlamydiae in all three programs used, with a total of 187 reads for Phylosift (Fig 6), 673 reads for MG-RAST (Fig 7) and 376 reads for MEGAN6 (Fig 3). Most bacterial reads from the Rincón sample were assigned to the phyla Bacteroidetes and Proteobacteria, with a total of 15 and 10 reads for Phylosift (Fig 6), 147 and 204 reads for MG-RAST (Fig 4), and 22 and 212 reads for MEGAN6 (Fig 4), respectively. Phylosift indicated that Bacteria represented 87% of the annotated reads for the Mayagüez sample, and 96% for the Rincón sample (Fig 6).

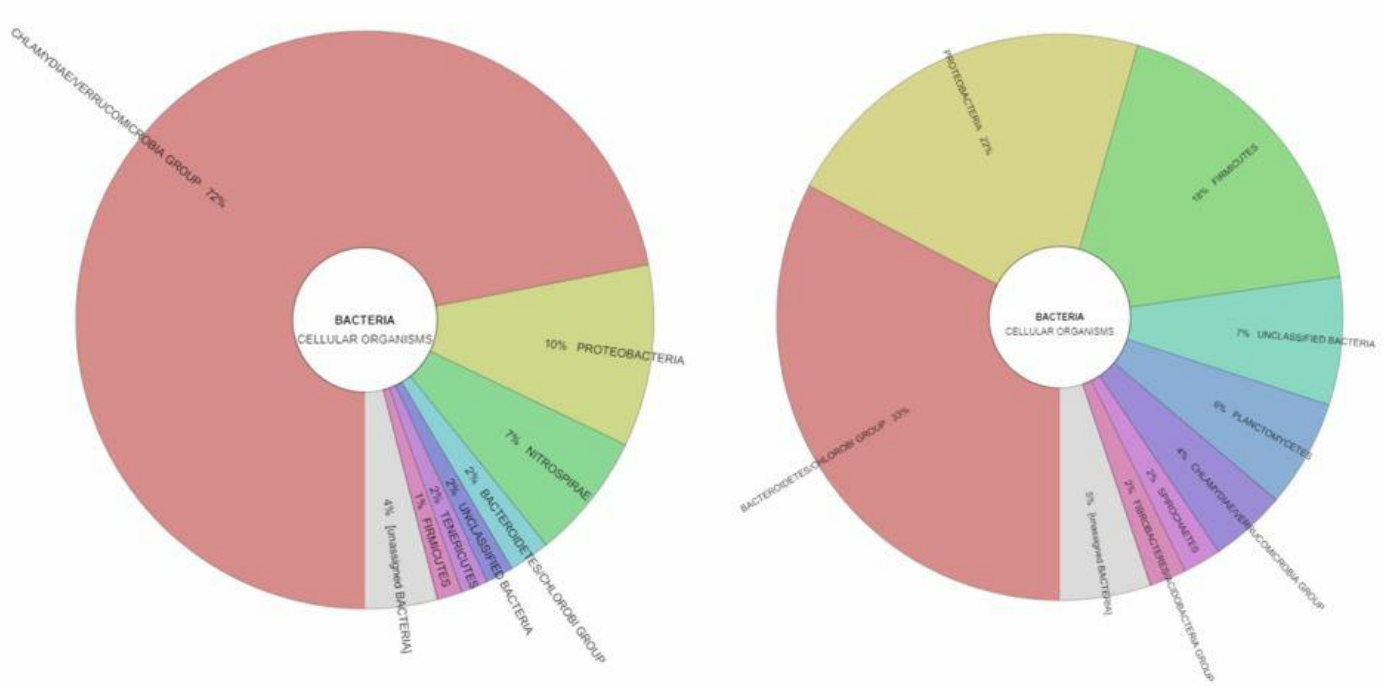

Fig 6. Phylosift results of Bacterial diversity chart from the Mayagüez sample (left) and the Rincón sample (right).

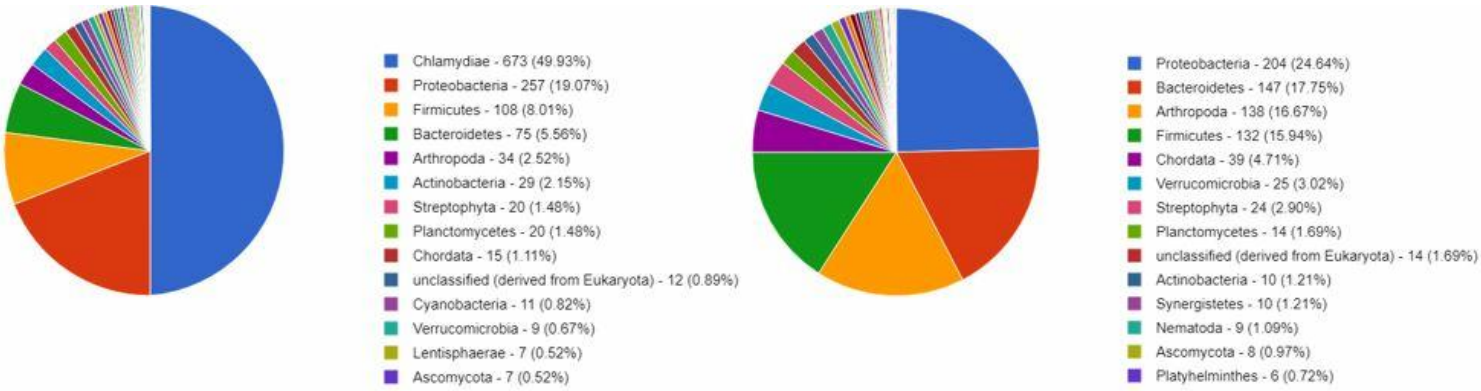

Fig 7. Phylum representation chart for the Mayagüez sample (left) and the Rincón sample (right), created by uploading the sample data to MG-RAST.

Table 2. Comparison of annotated reads from the mayor bacterial phyla sampled from *A. monilicornis* guts, based on the program used.

	Mayagüez sample			Rincón sample		
Phylum (Bacteria)	**Phylosift**	**MEGAN6**	**MG-RAST**	**Phylosift**	**MEGAN6**	**MG-RAST**
Chlamydia	187	376	673	2	2	0
Proteobacteria	27	158	257	10	212	204
Bacteroidetes	6	14	75	15	22	147
Firmicutes	4	7	108	8	17	132
Verrucomicrobia	0	4	9	0	37	25
Actinobacteria	0	36	29	0	23	10
Planctomycetes	0	0	20	3	2	14

Eukaryotic reads

According to Phylosift, the two samples had roughly the same number of reads for the protist phyla Alveolata and Stramenopiles (Fig 8). The MG-RAST analysis showed a majority of eukaryotic reads represented by Arthropoda, Streptophyta, and Chordata; the Mayagüez sample had 34 reads for Arthropoda, 20 for Streptophyta, and 15 for Chordata; the Rincón sample had 138 reads for Arthropoda, 39 for Chordata, and 24 for Streptophyta (Fig 7). Finally, MEGAN6 showed most eukaryotic reads belonged to Arthropoda, with 91 reads for the Mayagüez sample and 619 reads for the Rincón sample (Fig 3 & 4).

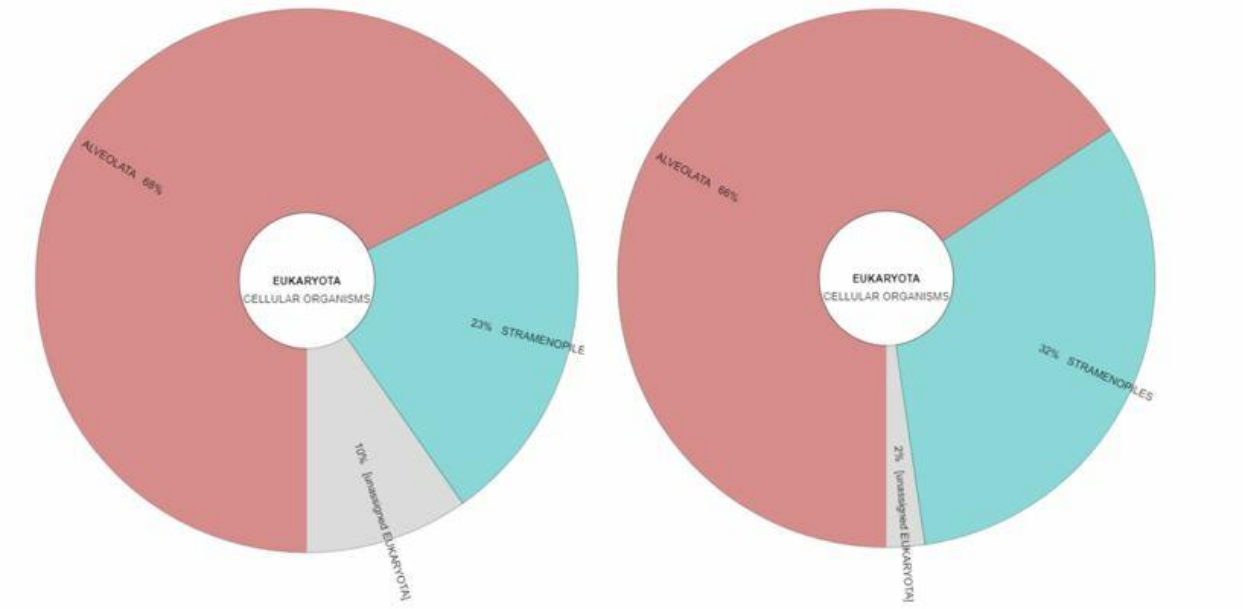

Fig 8. Eukaryotic diversity chart for the Mayagüez sample (left) and the Rincón sample (right), created using Phylosift data. Eukaryotes represented 12% of the sampled reads in the Mayagüez sample, and 4% in the Rincón sample.

Table 3. Comparison of annotated reads from the mayor eukaryote phyla sampled from *A. monilicornis* guts, based on the program used.

	Mayagüez sample			Rincón sample		
Phylum (Eukaryota)	**Phylosift**	**MEGAN6**	**MG-RAST**	**Phylosift**	**MEGAN6**	**MG-RAST**
Arthropoda	0	91	34	0	619	138
Chordata	0	2	15	0	15	39
Nematoda	0	2	0	0	3	9
Streptophyta	0	0	20	0	2	24
Ascomycota	0	0	7	0	3	8
Alveolata	24	0	0	1	0	0
Stramenopiles	8	0	0	1	0	0

Metabolism annotation

MG-RAST analysis showed that most of the annotated metabolic reads belonged to core cellular metabolism, followed by genetic and environmental metabolic pathways (Fig 9). The metabolic pathways, created via iPath3, can be seen in detail in Figures 10 and 11**.**

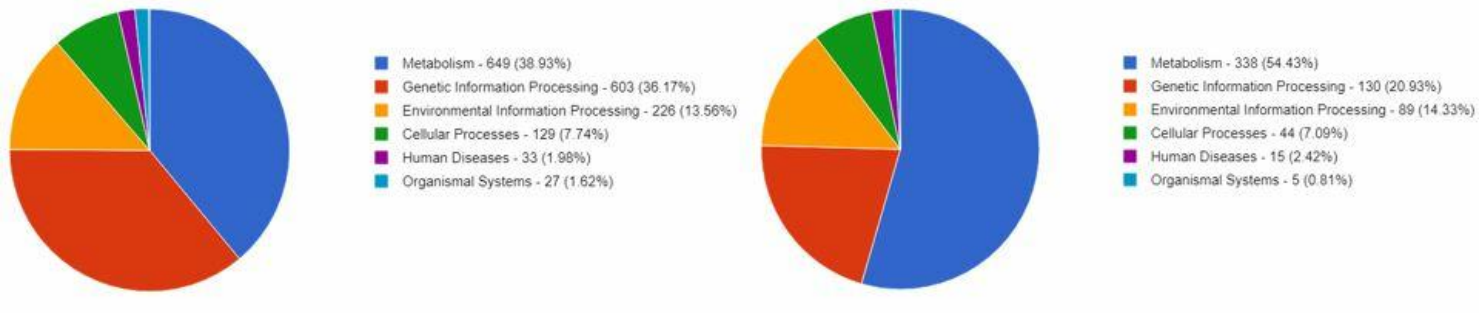

Fig 9. KO metabolism analysis for the Mayagüez sample (left) and the Rincón sample (right), obtained MG-RAST.

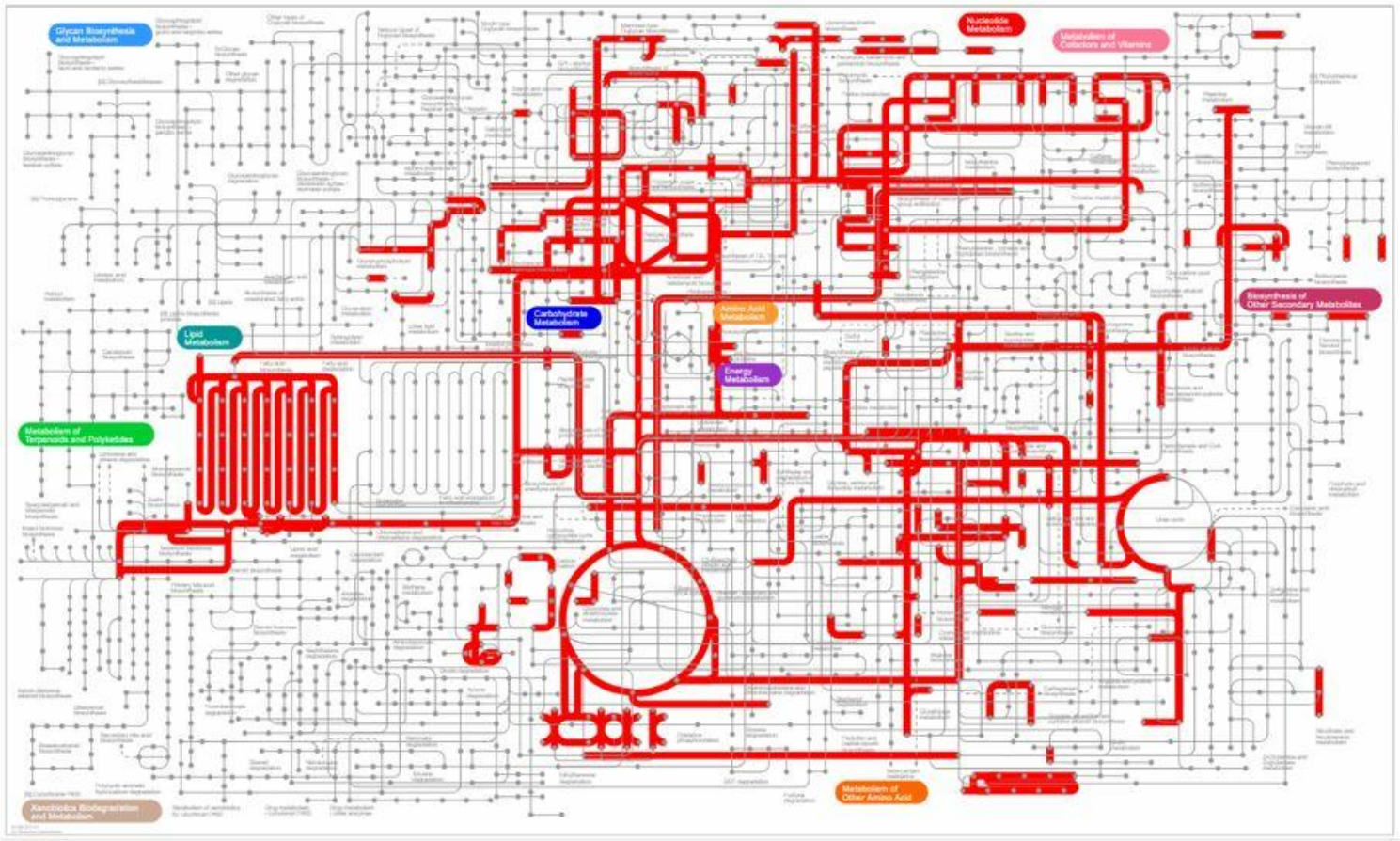

Fig 10. Metabolic pathways for the Mayagüez sample, made by uploading a KO list to iPath3.

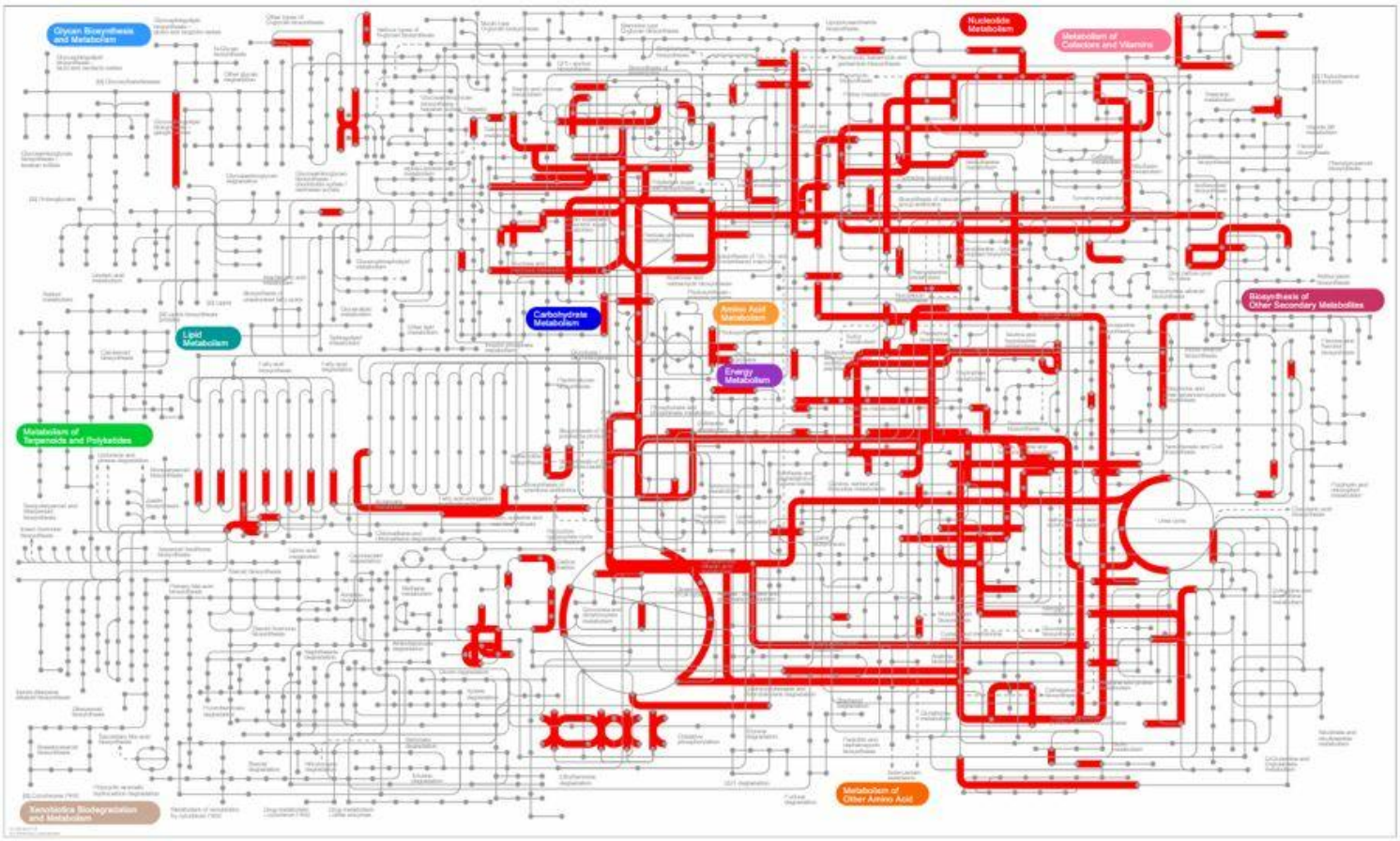

Fig 11. Metabolic pathways for the Rincón sample, made by uploading a KO list to iPath3.

Discussion

Our results show a preliminary fingerprint of the microbial diversity within the gut of *A. monilicornis*, with many reads annotated to bacteria (Tables 1 & 2). In comparison, a termite intestinal microbiota study by Rossmassler et al. (2015) found that across five termite species and three gut compartments, most of the genes under study were annotated to bacterial taxa. Many of our bacterial reads were annotated to phyla such as Proteobacteria and Bacteroidetes in both of our samples (**Table 2**). The sample from Mayagüez, however, showed a large number of bacterial reads that were annotated to the phylum Chlamydiae (**Table 2**). Since most arthropods are not known to "play a role in the epidemiology of chlamydial infections" (Corsaro & Greub 2006), it is possible that the Chlamydiae within the Mayagüez millipede may have been associated with eukaryotic symbionts. Indeed, some of the annotated Chlamydiae reads belonged to the family Parachlamydiaceae, a group commonly associated with amoebae (Greub & Raoult 2002). However, we did not get a notable amount of Amoebozoa reads from our samples. Due to this fact and the amoebozoans' relatively small genome size, we cannot determine whether these findings reflect a commonplace interaction between Chlamydiae and microscopic eukaryotes within *A. monilicornis's* intestine, or if other factors were involved.

Many arthropod metagenomic studies have focused on the 16S rRNA gene to determine bacterial composition. Termite microbiota has been extensively analyzed through 16S rRNA metagenomics, for example. Rossmassler et al. (2015) found bacterial diversity belonging to phyla such as Bacteroidetes and Firmicutes. This study found discrepancies between the taxa annotated through the 16S rRNA and the protein-coding gene analyses used, which they believe was due to "lack of appropriate reference genomes in public databases" (Rossmassler et al. 2015). Bourguignon et al. (2018), who also studied termites, found a high diversity of Firmicutes among

other bacterial phyla within their termites' guts. Interestingly, they found colony-offspring and colony-colony transmission of symbionts across the termites studied, which has shaped the evolution of the termite gut microbiome (Bourguignon et al. 2018). In terms of other insects, Chen et al. (2018) found many annotated reads belonging to genera within Firmicutes, Proteobacteria, and Actinobacteria through their 16S rRNA analysis of the gut microbiota of silkworms in conjunction with shotgun metagenomics. Through our shotgun metagenomic approach, we obtained many reads that were annotated to bacteria (Table 2). In conjunction with 16S rRNA metagenomics and specialized databases, we could potentially describe a greater proportion of the bacterial diversity within the millipede gut down to the genus or species level with higher accuracy with increasing sequencing depth greatly.

Our shotgun metagenomics approach allowed us to annotate bacterial and eukaryotic reads and identify many microbial phyla (Table 1 & 2). There are many studies that have taken this approach with good results. For example, Paula et al. (2016) were able to determine the gut contents of insect predators by comparing their shotgun data to different DNA reference databases. These included the mitogenomes of potential insects to be found, which were downloaded and sequenced for the study (Paula et al. 2016). The use of reference genomes of taxa expected to appear in the data may be of use in order to better determine microbial taxa in our millipede gut samples. The inclusion of the host reference genome in order to confirm the origin of the sample, as in Paula et al. (2016) and Bovo et al. (2018), could also be useful. Furthermore, Paula et al. (2016) express concern over the use of low number of reads for analyses, as these reads can potentially be "generated from sequencing or bioinformatics errors" (Paula et al. 2016). Since we obtained a relatively small number of annotated reads, any identification of taxa below the phylum level on our part could be prone to these errors.

Proteobacteria, Bacteroidetes and Firmicutes were some of the most well represented bacterial phyla in both of our samples in terms of relative abundance (**Table 2**). These findings are comparable with the microbiota found in the guts of termites, which harbor a varying representation of the same phyla as well as Actinobacteria and Planctomycetes, as in Rossmassler et al. (2015). Dittmer et al. (2016) found Proteobacteria as one of the most abundant phyla in the terrestrial isopod species they studied. Both termites and isopods share a similar ecological role to millipedes in the form of terrestrial nutrient recycling. As such, we would expect similarities in gut microbiota representation across the three arthropod groups.

In the present study, we sampled the microorganisms across the entire gut of the *A. monilicornis* specimens. Studies have shown differences in bacterial representation across the different gut compartments of different termite species. It was found that the representation of bacterial phyla in the termites was in part dependent on the gut section sampled; for example, the P1 hindgut compartment was dominated by Firmicutes in most of the termite species studied (Rossmassler et al., 2015). Nardi et al. (2016) studied the millipede *Cylindroiulus caeruleocinctus* and found that the highest microbial density could be found in the hindgut. Gut microbial diversity could potentially vary between different populations of millipedes. Dittmer et al. (2016) examined the terrestrial isopod *Armadillidium vulgare* finding differences in microbial communities based on localities. Differences in microbial representation could also be due to changes in diet as in Rossmassler et al. (2015), where the phylum Bacteroidetes was most abundant in wood-feeding termites (Rossmassler et al., 2015). In contrast, Knapp et al. (2009), who studied the effect of different diets on the gut microbiota of the alpine millipede *Cylindroiulus fulviceps*, found no significant microbial diversity changes between the samples. In the case of this study, our sample size is too small to be able to determine any differences in bacterial representation. We added an

additional variable to our study in the form of starvation time of the millipedes. The Mayagüez millipede sample was starved for approximately 24 hours, whereas the Rincón millipede sample was starved for ten days. Due to the different variables and conditions prior to gut extraction, some questions remain to be answered by the present study: Did starving cause differences in gut microbial diversity? Or were there differences in soil microbial composition between the two sites? Or a combination of both? A greater sample size and a more fleshed out experimental design could have allowed us to answer these questions with some degree of statistical significance.

Several additional questions arise from this study. Dittmer et al. (2016) found that 70% of the gut microbial taxa found in the isopod *Armadillidium vulgare* were also detected in feces and in the soil, suggesting that an important fraction of the microbiota may be acquired from environmental sources. The question as to how *A. monilicornis* acquires most of its microbiota, whether from the soil it inhabits, or horizontal transmission from other millipedes, or a combination of both remains unanswered. Crawford et al. (1983) found that the lumen bacteria of the desert millipede *Orthoporus ornatus* virtually disappeared after molting. Does microbial diversity representation change across *A. monilicornis's* lifespan, before and after molting? These questions, however, are beyond the scope of this study and require broader geographic and temporal sampling.

It is very odd that we did not get many annotated Nematoda reads, with Phylosift unable to obtain any nematode reads whatsoever (**Table 3**). We expected to find more, since we visually identified nematodes inside the extracted guts before sequencing them. We also expected a larger number of reads for the fungal phylum Ascomycota; according to the work of Byzov et al. (1993), ascomycetes are the most common yeast strains found on certain species of millipedes. However, we obtained a small number of Ascomycota reads compared to other phyla (**Table 3**). With our

current data, we are not able to determine why this happened, although it was most likely due to under sampling.

Some of the reads may have been misannotated as well. For example, the MG-RAST and MEGAN6 analyses annotated a small number of the eukaryote reads in both samples as belonging to Chordata (Fig 4, Fig 7), and MG-RAST annotation for metabolism returned reads related to human diseases (Fig 9). In addition, the Phylosift analysis annotated the eukaryotic reads as belonging to the protists Alveolata and Stramenopiles; specifically, the aquatic species *Thalassiosira* sp., *Kryptoperidinium foliaceum* and *Durinskia baltica*. This could be interpreted as misannotation of these eukaryotic reads to the organisms and molecular processes with the closest genetic resemblance that are available in the database. *Thalassiosira*, for example, is notable for being one of the first marine phytoplankton genera whose genome was sequenced (Armbrust et al., 2004); perhaps most of the protist reads databased belong to a small number of genera or species. Errors in annotation have been a problem since the advent of sequencing, and according to Schnoes et al. (2009) it has "increased from 1993 to 2005" for public databases in particular. Indeed, studies have reported error rates as high as 90% for protein and rRNA sequences, for example, in databases such as GenBank and TrEMBL (Schnoes et al., 2009; Tripp et al., 2011). Though this topic is beyond the scope of this study, we believe finding techniques to consistently circumvent these annotation errors in public databases is of upmost importance to the omics fields.

Conclusion

In our present study, a greater sample size could have allowed us to answer more biology-related questions. We planned to extract and sequence more millipede guts, but due to drawbacks concerning Hurricane Maria, we only managed to study two samples. This is a shame, as we were hoping to be able to distinguish populations based on metagenomic profiles, and to compare the microbial representation of the millipede intestine with that of other soil arthropods. This study nevertheless showed what to expect from sequencing the gut of a millipede using nanopore sequencing. We hope to be able to continue studying millipede gut metagenomics or to encourage more studies from the community of metagenomics researchers to continue further work. Future directions could focus on a more standardized sampling and extraction protocol, to be able to secure a larger sample size, and consequently compile a greater amount of the microbial taxa found in the gut. With more samples and sequencing depth, we could better determine any differences in microbial diversity between populations across Puerto Rico and delve deeper into the ecological importance of millipedes in the biomes they inhabit.

Literature Cited

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

Andersson, A. F., Lindberg, M., Jakobsson, H., Backhed, F., Nyrén, P., & Engstrand, L. (2008). Comparative analysis of korean human gut microbiota by barcoded pyrosequencing. *PLoS ONE*, *3*(7), e2836. https://doi.org/10.1371/journal.pone.0022109

Armbrust, E., Berges, J., Bowler, C., Green, B., Martinez, D., Putnam, N., Zhou, S., Allen, A., Apt, K., Bechner, M., Brzezinski, M., Chaal, B., Chiovitti, A., Davis, A., Demarest, M., Detter, J., Glavina, T., Goodstein, D., Hadi, M., Hellsten, U., Hildebrand, M., Jenkins, B., Jurka, J., Kapitonov, V., Kroger, N., Lau, W., Lane, T., Larimer, F., Lippmeier, J., Lucas, S., Medina, M., Montsant, A., Obornik, M., Parker, M., Palenik, B., Pazour, G., Richardson, P., Rynearson, T., Saito, M., Schwartz, D., Thamatrakoln, K., Valentin, K., Vardi, A., Wilkerson, F. & Rokhsar, D. (2004). The genome of the diatom Thalassiosira pseudonana: Ecology, evolution, and metabolism. *Science*, *306*(5693), 79–86. https://doi.org/10.1126/science.1101156

Batovska, J., Lynch, S. E., Rodoni, B. C., Sawbridge, T. I., & Cogan, N. O. (2017). Metagenomic arbovirus detection using MinION nanopore sequencing. *Journal of Virological Methods*. https://doi.org/10.1016/j.jviromet.2017.08.019

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

Bouchon, D., Zimmer, M., & Dittmer, J. (2016). The terrestrial isopod microbiome: An all-in-one toolbox for animal-microbe interactions of ecological relevance. *Frontiers in Microbiology*, *7*(SEP), 1–19. https://doi.org/10.3389/fmicb.2016.01472

Bourguignon, T., Lo, N., Dietrich, C., Sobotnik, J., Sidek, S., Roisin, Y., Brune, A. & Evans, T. (2018). Rampant host switching shaped the termite gut microbiome. *Current Biology*, 28(4), 649-654.

Bovo, S., Ribani, A., Utzeri, V., Schiavo, G., Bertolini, F., & Fontanesi, L. (2018). Shotgun metagenomics of honey DNA: Evaluation of a methodological approach to describe a multikingdom honey bee derived environmental DNA signature. PLoS ONE, 13(10), e0205575.

Brewer, M. S., Sierwald, P., & Bond, J. E. (2012). Millipede taxonomy after 250 years : Classification and taxonomic practices in a mega-diverse yet understudied arthropod group. *PLoS One*, *7*(5), e37240.

Brune, A. (2006). *Symbiotic associations between termites and prokaryotes*. (M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer, & E. Stackebrandt, Eds.), *The prokaryotes, handbook of the biology of Bacteria* (3rd ed., Vol. 1). New York: Springer. https://doi.org/10.1007/0-387-30741-9_17

Byzov, B. (2006). Gut Microbiota of Millipedes. In *Konig & Varma, A. Gut microorganisms of termites and other invertebrates.* (pp. 89–114). Heidelberg: Springer.

Byzov, B., Thanh, V., & Babjeva, I. (1993). Yeasts associated with soil invertebrates. *Biology and Fertility of Soils*, *16*, 183–187.

Chen, B., Teh, B., Sun, C., Hu, S., Lu, X., Boland, W., & Shao, Y. (2016). Biodiversity and Activity of the Gut Microbiota across the Life History of the Insect Herbivore Spodoptera littoralis. *Scientific Reports*, *6*, 29505. https://doi.org/10.1038/srep29505

Chen, B., Yu, T., Xie, S., Kaiqian, D., Liang, X., Lan, Y., Sun, C., Lu, X. & Shao, Y. (2018). Comparative shotgun metagenomic data of the silkworm Bombyx mori gut microbiome. *Scientific Data*, 5, 180285.

Chikhi, R., & Medvedev, P. (2014). Informed and automated k-mer size selection for genome assembly. *Bioinformatics*, *20*, 31–37.

Contreras, K., & Cafaro, M. (2013). Morphometric studies in Enterobryus luteovirgatus sp. nov. (Ichthyosporea: Eccrinales) associated with yellowbanded millipedes in Puerto Rico. *Acta Protozoologica*, *52*, 291–297.

Corsaro, D. & Greub, G. (2006). Pathogenic potential of novel Chlamydiae and diagnostic approaches to infections due to these obligate intracellular bacteria. *Clinical Microbiology Reviews*, 19(2), 283-297.

Crawford, C., Minion, G., & Boyers, M. (1983). Intima morphology, bacterial morphotypes, and effects of annual molt on microflora in the hindgut of the desert millipede, Orthoporus ornatus (Girard) (Diplopoda:Spirostreptidae). *International Journal of Insect Morphology and Embriology*, *12*, 301–312.

Darling, A. E., Jospin, G., Lowe, E., Matsen, F. A., Bik, H. M., & Eisen, J. A. (2014). PhyloSift: phylogenetic analysis of genomes and metagenomes. *PeerJ*, *2*, e243. https://doi.org/10.7717/peerj.243

David, J. (2014). The role of litter-feeding macroarthropods in decomposition processes: a reappraisal of common views. *Soil Biology and Biochemistry, 76*, 109–118.

Degli, M., & Martinez, E. (2017). The functional microbiome of arthropods. *PLoS ONE*, *12*. https://doi.org/https://doi.org/10.1371/journal.pone.0176573

Dittmer, J., Lesobre, J., Moumen, B., & Bouchon, D. (2016). Host origin and tissue microhabitat shaping the microbiota of the terrestrial isopod Armadillidium vulgare. *FEMS Microbiology Ecology*, *92*(5), fiw063. https://doi.org/10.1093/femsec/fiw063

Dong, Y., Zhu, L., Bai, Y., Ou, Y., & Wang, C. (2016). Complete mitochondrial genomes of two flat-backed millipedes by next-generation sequencing (Diplopoda, polydesmida). *ZooKeys*, *637*, 1–20. https://doi.org/10.3897/zookeys.637.9909

Engel, P., Martinson, V., & Moran, N. (2012). Functional diversity within the simple gut microbiota of the honey bee. *PNAS*, *109*, 11002–11007.

Gabel, K., Hunsberger, A., Mannion, C., Buss, L., & Buss, E. (2006). Yellow-banded Millipede, *13*. Retrieved from http://trec.ifas.ufl.edu/mannion/pdfs/Yellow-bandedMillipede.pdf

Greub, G. & Raoult, D. (2002). Parachlamydiaceae: Potential emerging pathogens. Emerging Infectious Diseases, 8(6), 625-630.

Guo, J., Wang, Q., Wang, X., Wang, F., Yao, J., & Zhu, H. (2015). Horizontal gene transfer in an acid mine drainage microbial community. *BMC Genomics*, *16*(1), 1–11. https://doi.org/10.1186/s12864-015-1720-0

Huson, D., Auch, A., Qi, J., & Schuster, S. (2007). MEGAN analysis of metagenomic data. *Genome Research*, *17*, 377–386.

Jain, M., Olsen, H. E., Paten, B., & Akeson, M. (2016). The Oxford Nanopore MinION: Delivery of nanopore sequencing to the genomics community. *Genome Biology*, *17*(1), 1–11. https://doi.org/10.1186/s13059-016-1103-0

Jain, R., Rivera, M. C., Moore, J. E., & Lake, J. A. (2002). Horizontal gene transfer in microbial genome evolution. *Theoretical Population Biology*, *61*(4), 489–495. https://doi.org/10.1006/tpbi.2002.1596

Kitz, F., Steinwandter, M., Traugott, M., & Seeber, J. (2015). Increased decomposer diversity accelerates and potentially stabilises litter decomposition. *Soil Biology and Biochemistry*, *83*, 138–141. https://doi.org/10.1016/j.soilbio.2015.01.026

Knapp, B., Seeber, J., Podmirseg, S., Rief, A., Meyer, E., & Insam, H. (2009). Molecular fingerprinting analysis of gut microbiota of Cylindroiulus fulviceps (Diplopoda). *Pedobiologia*, *52*, 325–336.

Konig, H. (2006). Bacillus species in the intestine of termites and other soil invertebrates. *Journal of Applied Microbiology*, *101*, 620–627.

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

Kumar, S., & Blaxter, M. L. (2011). Simultaneous genome sequencing of symbionts and their hosts. *Symbiosis*, *55*(3), 119–126. https://doi.org/10.1007/s13199-012-0154-6

Laver, T., Harrison, J., O'Neill, P. A., Moore, K., Farbos, A., Paszkiewicz, K., & Studholme, D. J. (2015). Assessing the performance of the Oxford Nanopore Technologies MinION. *Biomolecular Detection and Quantification, 3,* 1–8. https://doi.org/10.1016/j.bdq.2015.02.001

Linard, B., Arribas, P., Andújar, C., Crampton-Platt, A., & Vogler, A. (2016). Lessons from genome skimming of arthropod-preserving ethanol. *Molecular Ecology Resources*, 16(6), 1365-1377.

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L. & Law, M. (2012). Comparison of next-generation sequencing systems. *Journal of Biomedicine and Biotechnology, 2012*, 1–11. https://doi.org/doi:10.1155/2012/251364

Loman, N. J., & Quinlan, A. R. (2014). Poretools: A toolkit for analyzing nanopore sequence data. *Bioinformatics*, *30*(23), 3399–3401. https://doi.org/10.1093/bioinformatics/btu555

McIntyre, A. B. R., Rizzardi, L., Yu, A. M., Alexander, N., Rosen, G. L., Botkin, D. J., Stahl, S., John, K. K., Castro-Wallace, S. L., McGrath, K., Burton, A. S., Feinberg, A. P. & Mason, C. E. (2016). Nanopore sequencing in microgravity. *Npj Microgravity*, *2*, 1–9. https://doi.org/10.1038/npjmgrav.2016.35

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R., Wilke, A., Wilkening, J. & Edwards, R. (2008). The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Informatics*, *9*(386). https://doi.org/10.1186/1471-2105-9-386

Muturi, E., Ramirez, J., Rooney, A., & Kim, C. (2017). Comparative analysis of gut microbiota of mosquito communities in central Illinois. *PLoS Neglected Tropical Diseases, 11*. https://doi.org/10.1371/

Nardi, J., Bee, C., & Taylor, S. (2016). Compartmentalization of microbial communities that inhabit the hindguts of millipedes. *Arthropod Structure & Development*, *45*, 462–474.

Oxford Nanopore Technologies. (2017). Albacore basecaller from Oxford Nanopore. Retrived from https://nanoporetech.com/about-us/news/new-basecaller-now-performs-raw-basecalling-improved-sequencing-accuracy

Oxford Nanopore Technologies. (2017). New basecaller now performs `raw basecalling', for improved sequencing accuracy. Retrieved from https://nanoporetech.com/about-us/news/new-basecaller-now-performs-raw-basecalling-improved-sequencing-accuracy

Paula, D., Linard, B., Crampton-Platt, A., Srivathsan, A., Timmermans, M., Sujii, E., Pires, C., Souza, L., Andow, D. & Vogler, A. (2016). Uncovering trophic interactions in arthropod predators through DNA shotgun-sequencing of gut contents. PLoS ONE, 11(9), e0161841. doi:10.1371/journal.pone.0161841

Pitz, K., & Sierwald, P. (2010). Phylogeny of the millipede order Spirobolida (Arthropoda: Diplopoda: Helminthomorpha). *Cladistics*, *26*, 497–525.

Qi, W., Nong, G., Preston, J. F., Ben-Ami, F., & Ebert, D. (2009). Comparative metagenomics of Daphnia symbionts. *BMC Genomics*. https://doi.org/10.1109/TCSI.2002.800838

Robidart, J. C. (2006). Metagenomics of the Riftia pachyptila symbiont. *UC San Diego Electronic Theses and Dissertations*. University of California, San Diego. Retrieved from https://escholarship.org/uc/item/0zh707vn

Rossmassler, K., Dietrich, C., Thompson, C., Mikaelyan, A., Nonoh, J., Scheffrahn, R., Sillam-Dusses, D. & Brune, A. (2015). Metagenomic analysis of the microbiota in the highly compartmented hindguts of six wood- or soil-feeding higher termites. *Microbiome*, *3*(56), 111–118.

Schmidt, M., Vogel, A., Denton, A., Istace, B., Wormit, A., van de Geest, H., Bolger, M., Alseekh, S., Mass, J., Pfaff, C., Schurr, U., Chetelat, R., Maumus, F., Aury, J-M., Fernie, A., Zamir, D., Bolger, A. & Usadel, B. (2017). Reconstructing the gigabase plant genome of Solanum pennellii using nanopore sequencing. *BioRxiv*, 1–23. https://doi.org/10.1101/129148

Schnoes, A. M., Brown, S. D., Dodevski, I., & Babbitt, P. C. (2009). Annotation error in public databases: Misannotation of molecular function in enzyme superfamilies. *PLoS Computational Biology*, *5*(12). https://doi.org/10.1371/journal.pcbi.1000605

Shelley, R. (2014). A consolidated account of the polymorphic Caribbean milliped, Anadenobolus monilicornis (Porat, 1876) (Spirobolida: Rhinocricidae), with illustrations of the holotype. *Insecta Mundi*, *0378*, 1–12.

Sierwald, P., & Bond, J. E. (2007). Current status of the myriapod class Diplopoda (Millipedes): Taxonomic diversity and phylogeny. *Annual Review of Entomology*, *52*(1), 401–420. https://doi.org/10.1146/annurev.ento.52.111805.090210

Snyder, B., & Hendrix, P. (2008). Current and potential roles of soil macroinvertebrates (earthworms, millipedes, and isopods) in ecological restoration. *Restoration Ecology*, *16*, 629–636.

Sustr, V., Chronakova, A., Semanova, S., Tajovsky, K., & Simek, M. (2014). Methane production and methanogenic Archaea in the digestive tracts of millipedes (Diplopoda). *PLoS ONE*, *9*, 7. https://doi.org/10.1371/journal.pone.0102659

Szabo, I., Nasser, E., Striganova, B., Rakhmo, Y., Jager, K., & Heydrich, M. (1990). Interactions among millipedes (Diplopoda) and their gut bacteria. *Ber. Nat.-Med Verein Innsbruck, 10*, 289–296.

Taylor, E. (1982). Role of aerobic microbial populations in cellulose digestion by desert millipedes. *Applied and Environmental Microbiology*, *44*(2), 281–291.

Thomas, T., Gilbert, J., & Meyer, F. (2012). Metagenomics – a guide from sampling data analysis. *Microbial Informatics and Experimentation*, *2*, 3.

Vélez, M. (2014). Los gongolies, gungulenes o milpiés (Clase Diplopoda). *In Joglar, R., Santos-Flores, C. & Torres-Pérez, J. Biodiversidad de Puerto Rico: Invertebrados*, 240–253.

Warnecke, F., Luginbühl, P., Ivanova, N., Ghassemian, M., Richardson, T. H., Stege, J. T., Cayouette, M., McHardy, A., Djordjevic, G., Aboushadi, N., Sorek, R., Tringe, S., Podar, M., Martin, H., Kunin, V., Dalevi, D., Madejska, J., Kirton, E., Platt, D., Szeto, E., Salamov, A., Barry, K., Mikhailova, N., Kyrpides, N., Matson, E., Ottesen, E., Zhang, X., Hernández, M., Murillo, C., Acosta, L., Rigoutsos, I., Tamayo, G., Green, B., Chang, C., Rubin, E., Mathur, E., Robertson, D., Hugenholtz, P. & Leadbetter, J. R. (2007). Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature*, *450*(7169), 560–565. https://doi.org/10.1038/nature06269

Woyke, T., Teeling, H., Ivanova, N., Huntemann, M., Richter, M., Gloeckner, F., Boffelli, D., Anderson, I., Barry, K., Shapiro, H., Szeto, E., Kyrpides, N., Mussmann, M., Amann, R., Bergin, C., Ruehland, C., Rubin, E. & Dubilier, N. (2010). Metagenomic analysis of microbial symbionts in a gutless worm. *Lawrence Berkeley National Laboratory*, January 2014, 1–51.

Retrieved from

http://escholarship.org/uc/item/42w8m413%5Cnhttp://escholarship.org/uc/item/42w8m413.pdf

Yamada, T., Letunic, I., Okuda, S., Kanehisa, M., & Bork, P. (2011). iPath2.0: Interactive pathway explorer. *Nucleic Acids Research*, *39*, 412–415.

Yun, J., Roh, S., Whon, T., Jung, M., Kim, M., Park, D., Yoon, C., Nam, Y., Kim, Y., Choi, J., Kim, J., Shin, N., Kim, S., Lee, W. & Bae, J. (2014). Insect gut bacterial diversity determined by environmental habitat, diet, developmental stage, and phylogeny of host. *Applied and Environmental Microbiology*, *80*, 5254–5264.

Zerbino, D. R., & Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*, *18*(5), 821–829. https://doi.org/10.1101/gr.074492.107