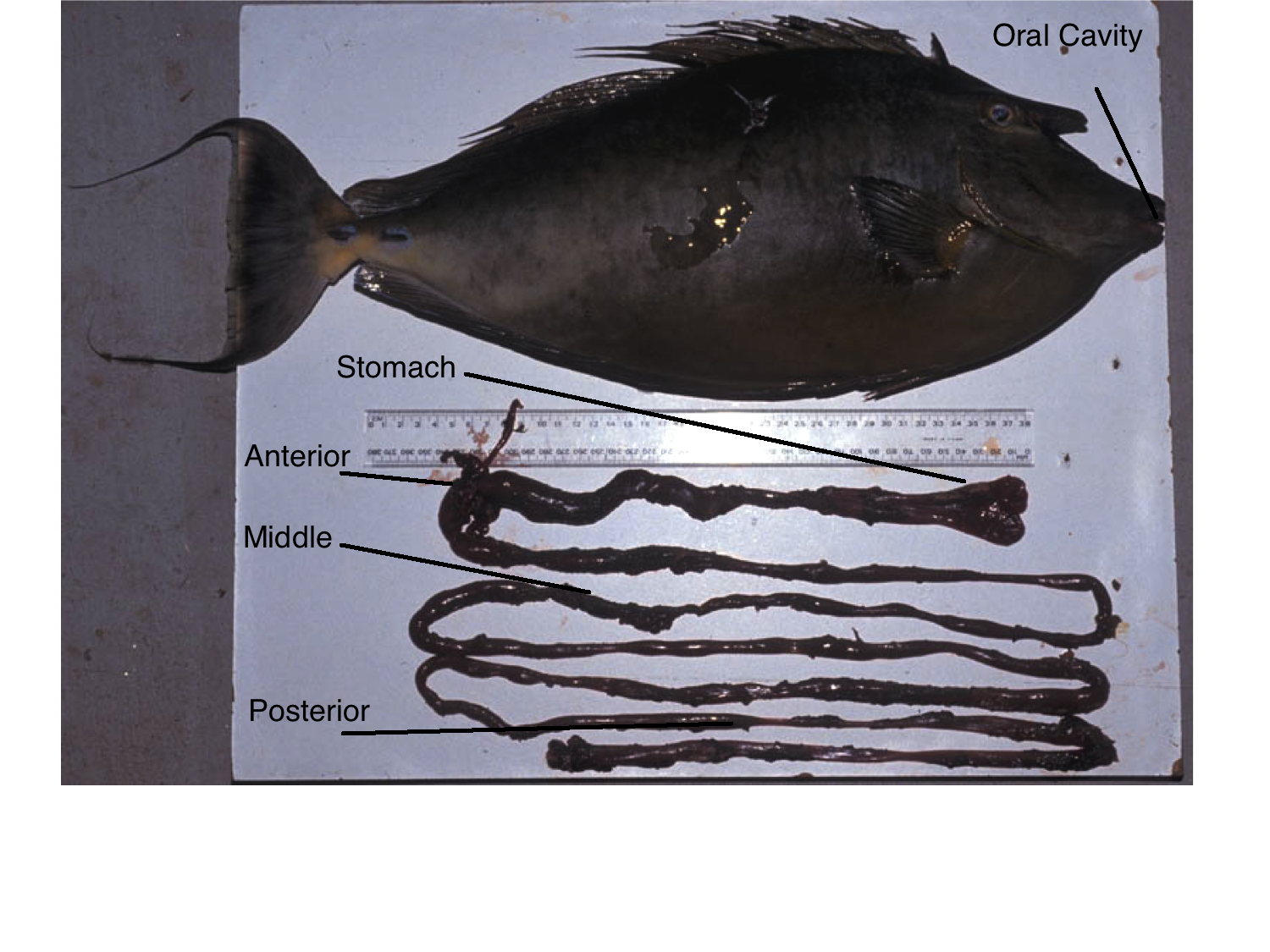
**Methods**

Sixty-two samples were collected from surgeonfish off of the coast of Oahu, Hawaii in July 2013, and another four samples were collected from Oahu earlier in 2011. *Naso lituratus* was the only species of surgeonfish collected from Oahu in 2011. The fish species that were collected in July 2013 were: *Naso lituratus, Naso unicornis*, *Acanthurus leucoparieus, Acanthurus triostegus, Acanthurus nigrofuscus* and *Zebrasoma flavescens.*

|  |
| --- |
| Surgeonfish species collected in July of 2011 and 2013 from O’ahu, Hawai’i |

|  |  |  |
| --- | --- | --- |
| Sample # | Fish Species | Number of replicates collected |
| 2,6,10,13 | *Naso unicornis* | 4 |
| 12 | *Acanthurus leucoparieus* | 1 |
| 11,7 | *Acanthurus nigrofuscus* | 2 |
| 9 | *Acanthurus triostegus* | 1 |
| 1,3 | *Naso lituratus* | 2 |
| 8 | *Zebrasoma flavescens* | 1 |

**Table 1) Information pertaining to surgeonfish samples that were collected from Oahu, Hawai’i in June 2013. The sample ID number that each fish species was assigned is on the left, the species name is in the center column, and the number of replicates collected is on the right.**

**Figure 1) Dissection of *Naso lituratus* gastrointestinal tract, the oral cavity is labeled along with the stomach, anterior, middle, and posterior sections of the gut. Photo taken from presentation by Kendall Clements.**

Samples were taken from the intestinal regions of the surgeonfish gut. An incision was made and fluid from the gut was collected and placed into 1.5 ml microcentrifuge tubes. Sections of the gastrointestinal tract were taken from the stomach, the anterior, the middle, and the posterior. The samples were then transported to Flagstaff, Arizona and kept frozen at -20 °C until DNA was extracted.

**DNA Extraction**

**PCR Conditions**

All Polymerase Chain Reaction (PCR) reactions were performed in 96-well plates on the BIO RAD iCycler thermal cycler in the Environmental Genetics and Genomics Laboratory (Enggen) .

In order to confirm the presence of bacteria in samples, the V4 hypervariable region of the 16S rRNA gene was amplified with the primers 515F and 806R. When running PCR reactions, 1 𝝻l of DNA template was added to each well in a row of a 96-well plate, along with .8 𝝻l of combined primers at 1mM, .6 𝝻l of mgcl2 at 1.5 mM, 2.6 𝝻l of water, and 5 𝝻l of Phusion master mix, which consisted of 0.01 U/uL Phusion Hot Start II Polymerase Thermo Fisher, 1X Phusion HF buffer (Thermo Fisher), 3.0mM MgCl2, 6% glycerol, and 200 uM dNTPs. PCR reagents were loaded into rows of a 96-well plate, 1 𝝻l of DNA template was added to each well, and the 96-well plate was then sealed, vortexed, spun down and briefly centrifuged. A 10:1 dilution was performed by transferring 1 𝝻l of of the previous PCR reaction to a new well with 9 𝝻l of master mix, Mgcl2, H2O, and combined primers, the plate was then sealed, vortexed for five seconds and spun down in a centrifuge. PCR reactions prior to sequencing were performed on the Illumina Miseq with the following conditions: 95C for two minutes, followed by 30 cycles of 95C for two minutes, 55C for one minute, 60C for four minutes, followed by a 10C hold until the plate was taken off of the thermocycler block and stored at 4C until gel electrophoresis could be performed. PCR amplification products were visualized using a .8% agarose gel using 1X sodium boric acid as a running buffer. During gel electrophoresis, agarose gels were ran for 15 minutes at 250 volts and 150 mA, with a KAPA 10 kb DNA ladder.

**Illumina Sequencing**

Prior to sequencing, all samples were normalized to one ng/𝝻l following quantification by Picogreen fluorescence. In order to construct a library we amplified the V4 domain of bacterial 16S DNA using primers 515F ( 5’-GTGCCAGCMGCCGCGGTAA-3’) and 806R ( 5’- GGACTACHVGGGTATCTAAT-3’) both of these primers were modified to contain an Illumina adapter region for sequencing on the Illumina Miseq, and on the forward 515F primer a Hamming error-correcting barcode was added to enable multiplexing of sequences. All samples were amplified in triplicate and combined prior to purification. Prior to sequencing, a bead cleanup was performed to remove any unincorporated primers. Purified samples were then run on an Illumina Miseq for a 2 X 150 paired end run.

**Data Analysis**

After samples were run on the Illumina Miseq, raw Illumina fastq files were demultiplexed, assigned to sample ids, quality filtered, and analyzed using QIIME 1.8.0 (Caporaso *et al*, 2010). A metadata mapping file was created with sample IDs matching one of several metadata categories. Subsampled open reference (Rideout, He *et al*. 2014)

Operational Taxonomic Units (OTUs) were assigned using the QIIME implementation of uclust (Edgar 2010), with the Greengenes 13\_8 reference sequence set (McDonald, Price et al. 2011) using a threshold of 97% pairwise identity. A sampling depth of 21,159 sequences per sample was chosen for rarefaction. Initial alpha diversity metrics, beta diversity metrics, and taxanomic summary plots were computed using the core\_diversity\_analyses.py script in QIIME. Due to the Ribosomal Database Project taxonomy classifier having a default confidence threshold of .8, taxa summary plots failed to show *Epulopiscium* as a genus, and showed the correct taxonomic assignment to the class level. Sequences from the taxonomic class *Clostridia* were identified in the OTU map and filtered from the representative set sequence file, and a new *Clostridia* only sequence file was then created. The *Clostridia* only sequence filewas searched against the NCBI database reference collection using BLAST against known *Epulopiscium fishelsoni* sequences. Next, the confidence level for the RDP taxonomy assignment classifier was changed to .5, which confirmed the presence of *Epulopiscium* as a genus in taxonomic summary plots.

**Third collection trip May 2014**

A third collection trip was conducted on May 22 - May 30th of 2014 to go to Oahu, Hawai’i in order to collect samples off of Coconut Island in Kaneohe Bay.Twenty-one individual fish were collected and dissected from various points off of Coconut Island. Surgeonfish were collected from the dates of May 23rd, 2014 - May 29th, 2014. Fish were collected from areas around the island such as the Tamashoiro Market, the reef pool off of Coconut island, the Sampan Channel, the Lanai channel, and the Electric Beach off of Oahu, Hawai’i. Samples were collected mostly in the morning hours from 8:30 am to 10:30 am and dissection was performed in the afternoon.

|  |
| --- |
| Surgeonfish species collected in May of 2014 from O’ahu Hawai’i |

|  |  |  |
| --- | --- | --- |
| Sample ID | Fish Species | Number of replicates collected |
| 3Ab | *Acanthurus blochii* | 1 |
| Al11,Al12 | *Acanthurus leucoparieus* | 2 |
| An9, An10 | *Acanthurus nigrofuscus* | 2 |
| Ao20, Ao20 | *Acanthurus olivaceus* | 2 |
| At14,At4 | *Acanthurus triostegus* | 2 |
| Ax16 | *Acanthurus xanthopterus* | 1 |
| Nl17,Nl18, Nl19, | *Naso lituratus* | 4 |
| Nu15, Nu1 | *Naso unicornis* | 2 |
| Zf5 | *Zebrasoma flavescens* | 1 |
| Zv6 | *Zebrasoma veliferum* | 1 |

**Table 2) Information on the species and the quantity of surgeonfish collected from Hawai’i in May 2014. The left column displays the sample ID names that fish were given, the center column displays the species of fish collected, and the right column displays the number of replicate samples that were collected.**

**Direct to PCR Amplification**

The direct to PCR approach(Flores *et al*,2012) was used to extract, and directly amplify DNA from the surgeonfish gastrointestinal tract. Swabs were placed into 1.6 ml deep wells of an Axygen scientific 96 well plate. 250 ul of Extract-N-AMP Extraction solution was placed into each well. The plate was then sealed using a sealing mat and heated in a water bath at 90 - 95 C for 10 minutes. followed by centrifugation at 2,500 x g. Extract-N-Amp Plant Dilution Solution was then added to the wells at a ratio of 1:1. Five 𝝻l of Extract-N-AMP Ready Mix was used, along with .5 𝝻l of 515 Forward(5-GTGCCAGCMGCCGCGGTAA-3) - 806 Reverse (GGACTACHVGGGTATCTAAT) primers each, 2.5 ul of PCR grade water and 2 𝝻l of Extract-N-AMP sample solutions from the PCR plates.

**Plant DNA extraction in 96 - well plates**

Due to the Extract N Amp kit yielding products that were not pure enough for next generation sequencing, another plant protocol was used to extract DNA in 96 well plates.

First, a small amount of SiC particles were added to each well in the plate, then a small amount of cotton swab was added to each well over the SiC particles. Following that 40 𝝻l of .25 M NaOH was added to each well, and sealed with strip caps. The material in the plate was then ground using a mixer mill at 28 Hz for 30 seconds. A denature program was then used on a thermal cycler with the following protocol: 4 minutes at 95 C, 1 minute at 4C, and 10C forever; the plate was then centrifuged briefly at 2,000 rpm. Next, 130 𝝻l of .1M Tris-HCl pH 8.0 was added to each well and sealed with foil; the plate was then centrifuged for 10 minutes at 3500 rpm. Next 150 ul of supernatant was added to a clean 96 well plate. Next, 15 𝝻l 3M NaOAc, pH 5.2 was added to each well. Following this, 120 𝝻l of isopropanol was added to each well, the plate was sealed, and mixed by inversion. The plate was then centrifuged at 3,500 rpm for 60 minutes. After centrifugation, the supernatant was discarded and the plate was inverted on a paper towel and centrifuged for 10 seconds at 200 rpm. Following this, 100 ul of 70% EtOH was added to each well and the plate was sealed. Next the plate was centrifuged for 20 minutes at 3,500 rpm. Then, the plate was centrifuged at 200 rpm for 10 seconds. Following brief centrifugation, the plate was dried by leaving it unsealed for 6 minutes in a vacuum centrifuge at 60 C. Finally, 20 ul TE was added to each well for the final step.

PCR reactions were performed in 96-well plates on the BIO RAD iCycler thermal cycler in Enggen . The V4 hypervariable region of the 16S rRNA region was used with the primers being 515 forward to 806 Reverse. Reaction volumes were 10 𝝻l for each PCR reaction, reagents and their concentrations were .8 𝝻l of combined primers at 1 mM, .6 𝝻l of mgcl2  at 1.5 mM, 2.6 𝝻l of water, 5 𝝻l of a master mix that consisted of 0.01 U/uL KAPA2G Fast DNA Polymerase, 1X KAPA2G Fast buffer, 3mM mgcl2, 6% glycerol, and 200 𝝻M dNTPs. All 16S rRNA PCR reactions prior to sequencing and on the Illumina Miseq instrument were performed with the following conditions: 95C for two minutes, followed by 30 cycles of 95C for two minutes, 55C for one minute, 60C for four minutes, followed by a 10C hold until the plate was taken off of the block.

PCR amplification products were visualized using gel electrophoresis using a one percent agarose gel with SYBR Safe nucleic acid stain. Electrophoresis gels were ran for 15 minutes at 250 volts and 150 mA, with a KAPA 10 kb ladder.

**Bead Cleanup (**Andrew Krohn, July 2013. Adapted from Rohland & Reich, 2012)

Bead cleanups were performed with .1 % carboxyl-modified Sera-Mag magnetic Speed-beads and a magnetic 96 well plate. Approximately equal volume of bead solution as the sample was added to each well of a 96 well plate. The solution was pipetted up and down multiple times to properly mix the solution. The plate was then spun down briefly at approximately 1,200 rpm. The plate was covered and placed on a 96 well magnetic plate for ten minutes. The plate was then kept on the 96 well magnetic plate for two to five minutes, then the liquid in the wells was disposed of. Next 100 uL of 70% EtOH was added, after 30 seconds the EtOH was pipetted off and new EtOH was added, again the EtOH was pipetted off. The plate was then dried at 60 C in a vacuum centrifuge for 2- 4 minutes or until the wells were dry. Next, 20 ul of 10 mM Tris-Cl pH 8.8 was added to each well and the plate was vortexed and spun down, then placed on the magnet for approximately two minutes.

Finally, the elutant was transferred to a new well of a 96 well plate and stored at -20C until samples could be sequenced.

|  |
| --- |
| Surgeonfish species collected in June 2011,June 2013 and May of 2014 from O’ahu, Hawai’i |

|  |  |  |
| --- | --- | --- |
| Sample ID | Fish Species | Number of individuals collected |
| 3Ab | *Acanthurus blochii* | 1 |
| Al11,Al12,12 | *Acanthurus leucoparieus* | 3 |
| An9, An10,11,7 | *Acanthurus nigrofuscus* | 4 |
| Ao20, Ao20 | *Acanthurus olivaceus* | 2\* |
| At14,At4, 9 | *Acanthurus triostegus* | 3\* |
| Ax16 | *Acanthurus xanthopterus* | 1 |
| Nl17,Nl18, Nl19, Nl8,1,3 | *Naso lituratus* | 6 |
| Nu15, Nu1, 2, 6, 10, 13 | *Naso unicornis* | 6 |
| Zf5,8 | *Zebrasoma flavescens* | 2 |
| Zv6 | *Zebrasoma veliferum* | 1 |

**Table 3)** **List of surgeonfish species collected in both Hawai’ian collection trips. Sample IDs are in the left column ,the name of the species is in the center column, number of replicate fish collected is in the right column. Numbers with an asterisk mean that more fish were caught, but due to poor sample quality were not included in further analysis.**

**Data Analysis**

The V4 hypervariable region the 16S rRNA gene was amplified using barcoded PCR with 515F and 806R primers that were developed as protocol for the Earth Microbiome Project (Caporaso, Lauber et al. 2012). A metadata mapping file was created with sample IDs matching one of several metadata categories. After samples were run on the Illumina Miseq, raw Illumina fastq files were demultiplexed, quality filtered with a phred quality score of 19 and analyzed using QIIME 1.8.0 (Caporaso *et al*, 2010).

Sequences were then chimera filtered using usearch 61, and non chimeric sequences were filtered out from the demultiplexed output. Subsampled open reference (Rideout, He *et al*. 2014) Operational Taxonomic Units (OTUs) were assigned using the QIIME implementation of uclust (Edgar 2010), with the Greengenes 13\_8 reference sequence set ( McDonald, Price et al. 2011) using a threshold of 97% pairwise identity. If a sequence failed to hit the reference database, the sequence was assigned an OTU, and any forthcoming sequences could be compared to that new reference OTU. Sequences that were assigned to a new OTU were included in downstream analysis, but no taxonomy would be assigned to them. Output reference sequences of open reference OTU picking were then aligned using PyNAST (Caporaso, Bittinger *et al*, 2010), filtering was applied with a lanemask, which defines when a position in a DNA sequence should be filtered out when building a phylogenetic tree. A phylogenetic tree was built using FastTree (Price, Dehal et al. 2010), in order to calculate phylogenetic diversity metrics. Next, taxonomy was assigned to each sequence using the RDP taxonomy classifier (Wang, Garrity *et al*, 2007). Following taxonomy assignment, a “raw” OTU table was created combining the open reference OTU table and the taxonomy assignment file created as an output from RDP taxonomic assignment. A phylogenetic tree was then generated using the pynast aligned open reference sequences. Finally, in order to filter out any observations in the OTU table that may have been attributed to sequencing error, observations that were observed only once or twice were filtered out of the OTU table. A sampling depth of 21,282 sequences per sample was chosen for rarefaction. Initial alpha diversity metrics, beta diversity metrics, and taxanomic summary plots were computed using the core\_diversity\_analyses.py script in QIIME. Both distance matrices and OTU tables were filtered on a per individual and per species basis using the scripts filter\_distance\_matrix.py for distance matrices and filter\_samples\_from\_otu\_table.py to filter OTU tables. Both OTU tables and distance matrices were filtered on a per species and per individual basis. In order to compare the within individual and between individual Unifrac distances in a sample category, make\_distance\_boxplots.py was performed with 999 Monte Carlo permutations using a two sample t-test.

**Combination of samples from collection trips in July 2011 ,July 2013 and May 2014 for downstream sequence analysis.**

After sequences were processed and data was analyzed separately from the June 2011, June 2013, and May 2014 collection trips, sequences from all runs were combined. In total 163 surgeonfish gut samples were analyzed; 10 oral cavity, 24 stomach,39 anterior, 39 middle, and 51 posterior samples were analyzed. Demultiplexed sequences were concatenated, metadata mapping files were combined from both sequencing runs. Sequences were chimera filtered using usearch 61 and subsampled open reference OTU picking was performed using the Greengenes 13\_8 reference set using a threshold of 97% identity. Output sequences of subsampled open reference OTU picking were aligned via PyNAST, a phylogenetic tree was created using FastTree. Taxonomy was assigned to sequences using the RDP taxonomy assigner at a confidence level of .8. OTUs that were only observed once or twice in the OTU table filtered out, in order to filter observations that could be sequencing error. For downstream bioinformatics analysis, QIIME was used. Rarefaction sampling depth was set to 21,282 sequences per sample. The core\_diversity\_analysis.py script was used to analyze alpha diversity, beta diversity, and taxonomic summary plots. OTU tables and distance matrices were filtered on a per individual and per species basis.