8-2021

A Comparison of Metagenomic Sequencing Using Targeted 16S and Whole Genome Shotgun NGS on Microbial DNA Samples

Adam Daniel Parker
Montclair State University

Abstract

The Oyster Creek Nuclear Generating Station (OCNGS; Lacey Township, New Jersey, USA), affects the surrounding aquatic environment as the outflow water is approximately 5^{o}C warmer than ambient water temperature. A metagenomic analysis was performed to assess microbial biodiversity at 4 sites located in Barnegat Bay, New Jersey, USA possibly in response to thermal gradients. A total of twelve samples from four sites was examined using Next Generation Sequencing (NGS). These represented the outflow and intake of the OCNGS, as well as bay area and river control sites. In addition, we compared targeted (16S) and Whole Genome Shotgun (WGS) methods. The microbiome analysis package QIIME2 and The Metagenomics RAST server (MG-RAST) were used to taxonomically identify bacterial composition and to compare the taxonomic makeup of sites. The sites where the higher temperatures were recorded showed a decrease in diversity compared to other sites. The OCNGS outflow site showed the lowest taxonomic diversity compared to all other sites. The comparison between targeted and WGS found the same overall trends in terms of the most abundant taxa identified. However, WGS identified more individuals at all levels of taxonomy.

A COMPARISON OF METAGENOMIC SEQUENCING USING TARGETED 16S AND WHOLE GENOME SHOTGUN NGS ON MICROBIAL DNA SAMPLES

by

Adam Daniel Parker

A Master's Thesis Submitted to the Faculty of

Montclair State University

In Partial Fulfillment of the Requirements

For the Degree of

Master of Science

August 2021

College of Science and Mathematics

Department of Biology

Thesis Committee:

[redacted]

Dr. Robert Meredith
Thesis Sponsor

[redacted]

Dr. Paul Bologna
Committee Member

[redacted]

Dr. John Gaynor
Committee Member

A COMPARISON OF METAGENOMIC SEQUENCING USING TARGETED 16S AND WHOLE GENOME SHOTGUN NGS ON MICROBIAL DNA SAMPLES

A THESIS

Submitted in partial fulfillment of the requirements

For the degree of Master of Science

by

Adam Daniel Parker

Montclair State University

Montclair, NJ

2021

Acknowledgements

Thesis Advisor Dr. Robert Meredith

Thesis committee member Dr. John Gaynor

Thesis committee member Dr. Paul Bologna

Dr. Dena Restaino

The Madelon Wehner Research Fund

Financial support supplied by NSF grants DEB-1556701, DBI-1725932 to Dr. Meredith

Montclair State University, Biology Department

Table of Contents

List of Tables

List of Figures

Appendices

Introduction

The identification of bacteria has been an important area of research for the past 150 years (Jordan, 1894). Traditionally, morphological identification based on observation was the main method (Phumudzo *et al.*, 2013). Key problems have been found with this methodology such as being highly time consuming, the results can be ambiguous (due to variability of culture) and subjective nature of the observer (Phumudzo *et al.*, 2013). Recent advances in technology have led to the development of genetic identification techniques such as Next Generation Sequencing (NGS), which aid in the identification and discrimination of bacterial taxa (Dowd *et al.*, 2008). Microbiome scientific studies have developed the concept of DNA barcoding, where a genetic sequence library focusing on species and specific portions of the genome is used to identify unknown organisms (Bukin *et al.*, 2019). DNA barcoding involves using polymerase chain reaction (PCR) to amplify a target region, which can then be nucleotide sequenced for comparison to a known bacterium. The most targeted region for bacteria has been the 16S gene which encodes the rRNA 30S subunit region (Fig. 1). This gene is *ca.* 1600 base pairs long and has nine hypervariable regions (Bukin *et al.*, 2019). The more conserved regions are useful at determining higher-level taxonomic rankings; the more variable regions help to identify genus and species (Bukin *et al.*, 2019)

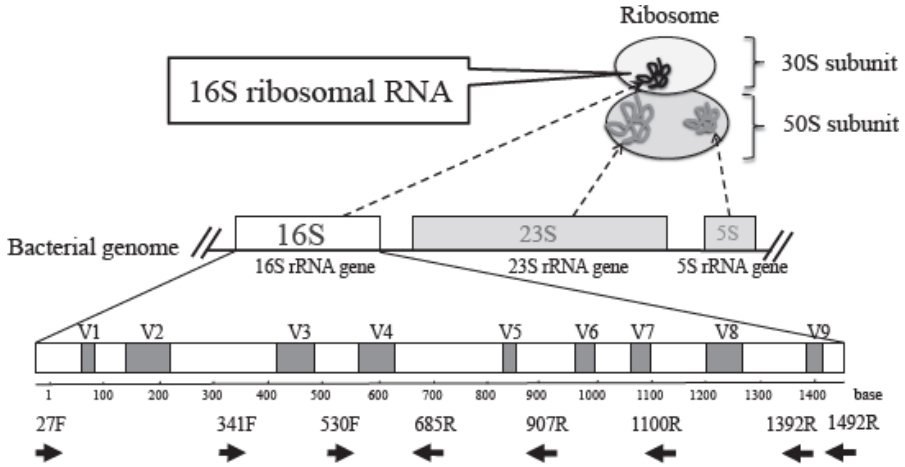


***Figure 1**. The 16S rRNA gene, the target of bacterial identification due to its nine conserved yet variable regions. Image from Fukuda K., et al. 2016.*

The most commonly sequenced region of 16S for bacterial identification is the V3/V4 16S region (Jovel *et al.*, 2016)(Ranjan *et al.*, 2016). However, there is a growing interest

in Whole Genome Shotgun (WGS) metagenomics as a more precise and inclusive method to capture more taxa by sequencing all available genomic DNA in a sample (Brooks *et al.*, 2015). Several studies have found bias in the 16S amplification method arising from PCR amplification, DNA extraction protocol, sequencing artifacts, DNA sample crossover, and primer design (Brooks *et al.*, 2015)(Hansen *et al.*, 1998)(Acinas *et al.*, 2005). Some of these biases can be overcome, for example, by altering extraction methods (triple DNA extraction) or reducing PCR cycles to avoid chimera formation (Brooks *et al.*, 2015). Another bias is from the interactions between DNA from different bacterial species during PCR, where observed proportions of bacteria are amplified or suppressed by the presence of other bacteria. This can be characterized as the difference in ability to utilize resources in PCR, due to a synergistic (if presence bacterium B increases the observed proportion of bacterium A) or antagonistic (if presence bacterium B decreases the observed proportion of bacterium A) interaction between bacterial DNA (Brooks *et al.*, 2015).

Whole Genome Shotgun (WGS) sequencing can potentially reduce the PCR biases mentioned above due to the absence of a targeted PCR step. WGS sequencing is a method by which random fragments of the whole genome are directly sequenced (Ong *et al.*, 2013). The main advantage of WGS is the elimination of the competitive bias of targeted PCR. Other advantages of WGS are the ability to go beyond the genus-level taxonomic assignments that are generally the greatest level of resolution in the targeted 16S approach (Hillmann *et al.*, 2018). Additionally, only rough estimates of functional profiles can be identified using a targeted approach (Gilbert *et al.*, 2018). This is due to the constraint of having to use 16S databases and not being able to use functional databases (e.g., UniProt, eggnog) to identify individuals. Another criticism of the targeted 16S approach is that, due to being able to sequence at the most 2 x 300 bp on the NGS Illumina platform, only a small portion of the 16S gene can be targeted using targeted PCR (Bukin *et al.*, 2019). One reason for potential differences in either composition or resolution of the taxa (in the results of 16S compared to WGS) is that instead of covering a small variable region, we would be able to cover multiple areas of the genome simultaneously. This could potentially lead to a potentially higher degree of resolution and accuracy in taxonomic classification. While WGS may present a solution to some of the 16S amplification method issues currently, there is not as much public data of whole genome sequences in comparison to 16S databases. Therefore, if the area of the genome that is sequenced using WGS has not yet been identified, we would be unable to assign it taxonomically.

Benthic DNA

The main purpose of this study is to compare the biodiversity between different environments. Biodiversity is of interest to ecological research as it is vital in order to preserve a healthy environment (Chapin *et al.*, 2000). Therefore, accurate analysis of

biodiversity and the potential changes by human influence would be highly beneficial. Benthic microbial communities are critical members of the aquatic ecosystem, as they play various roles in organic matter decomposition, nutrient cycling, and bioremediation. In addition, benthic samples normally contain a much higher concentration of bacteria compared to the water column above. Both factors make it the ideal target for research as an indicator of changes in biodiversity. Utilizing molecular biology techniques such as DNA extraction, Next Generation Sequencing (NGS), and phylogenetic analysis we can observe the composition of bacteria present. This study also presents an opportunity for a secondary objective, to compare the NGS methods of targeted 16S amplification and WGS sequencing.

Methods

Study site

The Oyster Creek Nuclear Generating Station (OCNGS; Lacey Township, New Jersey, USA), a nuclear power plant located near Oyster Creek in New Jersey, USA (Figure 2) was commissioned in December 1969, and permanently decommissioned in September 2018. At the time of the shutdown, this plant was the longest operating nuclear power plant in the United States. The OCNGS requires approximately 360 million gallons of water per day for the purposes of circulating, cooling, and dilution (Gallagher, 2018). This is acquired from the surface water of the South Branch of Forked River. This water is used to convey heat from the reactor core to drive the steam turbines in a closed loop (World Nuclear Association, 2020). Water is also used to remove and dump surplus heat from this circuit, by cooling the closed system to condense the steam. The heat transferred is considered surplus waste and is discharged into a body of water (World Nuclear Association, 2020). This leads to a rise in temperature in the surrounding area where this water is discharged (in this case Oyster Creek). In the case of OCNGS, the water temperature of the outflow is approximately 5°C higher (thermal loading) than the surrounding areas (Table 1). This represents a unique opportunity to compare benthic community composition driven by differences in environmental conditions. We investigated whether this environmental difference has an influence on bacterial diversity using both the WGS and targeted 16S sequencing techniques.

***Figure 2**. The Power plant location (green) in relation to benthic sample collection sites. Forked River (blue) represents the intake of the Power plant, while Oyster Creek (red) represents the outflow. Created using GPS Visualizer, https://www.gpsvisualizer.com.*

Four sites were identified for collection (Figure 3) with three duplicates for comparison and normalization. The sites were chosen, as they represent different environments within the coastal ecosystem. The Oyster Creek site was selected, as it was the closest accessible site for sample collection where the water temperature is affected by the outflow. Forked River represents the nearest river region for comparison, while Sunrise Beach was chosen as a nearby bay area control. The Traders Cove site was chosen as an additional control region for comparison to these, while far enough away to not be affected by the power plant (28.269km/17.566mi), it is still close enough to be a good comparison in terms of general geographical biodiversity expected.

***Figure 3**. The locations of sampling sites for benthic sample collection. Created using GPS Visualizer, https://www.gpsvisualizer.com.*

Sample collection

The samples were collected between 3-14-2018 and 3-18-2018 supervised by Drs. John Gaynor and Dena Restaino, Montclair State University, NJ. Four sites were identified for collection with subsampling of three independent replicates. Each sediment sample was collected using a benthic grabber (Ekman Benthic Sampler (Model #196-B15). Three sterile Falcon 50-ml conical tubes were collected from each site and were stored on ice until they were returned to the lab. All tubes were stored at -80ºC until they could be extracted for DNA. GPS coordinates and water chemistries such as salinity in parts per thousand (ppt), temperature, and dissolved oxygen were recorded at all four sampling sites to include as metadata values for comparison in the bioinformatics analysis. These values were also defined as a category by grouping the data, such as salinity which was divided into 3 categorical groups (15-20ppt, <15ppt and >20ppt). This enabled us to compare the environmental variables in more general terms to try and see the overall trends.

DNA extraction

After being thawed on ice, a total of the 3.5g of each benthic sediment sample was transferred into a 15ml sterile tube and resuspended in 10 ml of extraction buffer (100 mM Tris-HCl [pH 8.0], 100 mM EDTA [pH 8.0], 100 mM Na_2HPO_4 [pH 8.0], 1.5 M NaCl). It was then incubated at 37°C for 30 min with vigorous shaking (250 rpm) using a floor mounted shaking incubator. Next, 100 mg of lysozyme (100 mg/ml) and 100 µl of Pronase (20 mg/ml) were added to the sample. This mixture was then incubated at 37°C

for 1 h with gentle shaking in an incubator oven with a rotating spit. Proteinase K (50 µl) was added to the mix followed by incubation: 30 min at 37°C and 30 min at 55°C. Sodium dodecyl sulfate (SDS) (20%, 1.5 ml) and 1 ml of 20% N-laurylsarcosine were then added to each sample. These samples were incubated at 65°C for 2 h and slowly rotated. Samples were extracted twice with an equal volume of phenol, twice with an equal volume of phenol-chloroform-isoamyl alcohol (50:49:1), and twice with an equal volume of chloroform-isoamyl alcohol (24:1). The aqueous phase was precipitated with isopropanol (at 0.6 by volume) at room temperature for 1 h. The precipitation was then transferred into 1.5ml Eppendorf tubes. These were spun down to pellet the DNA via centrifugation at 20,000 x g. After removal of the supernatant the DNA was then washed in 70% cold ethanol, dried, and resuspended in sterile distilled water. Each isolated DNA sample was transferred to a sterile 1.5ml Eppendorf tube (50 µL volume per tube). The concentration of DNA was then measured spectrophotometrically using the NanoDrop1000 (Thermo Fisher) and fluorometrically using the Qubit 3.0 (Invitrogen) (see appendix for measurement values). These extractions were then stored at -20^{O}C until library preparation.

16S Amplification

The 16S NGS library preparation for works optimally with DNA at a concentration of 5ng/µL. Therefore, all samples were normalized using 10 mM Tris pH 8.5, to ~5ng/µL, except for FR1, FR2 and FR3 as they were under 5ng/µL, so those samples were left undiluted. The extracted DNA was used to make PCR targeted V3/V4 16S next generation sequencing libraries using Illumina's 16S Metagenomic Sequencing Library Preparation (Preparing 16S Ribosomal RNA Gene Amplicons for the Illumina MiSeq System, protocol part # 15044223 Rev. B). This protocol targets the 16S rRNA gene variable region V3/V4 using a targeted primer (Figure 4, 5' end, in red) attached to a universal adapter (Figure 4, non-bold, 3' end sequence).

16S Amplicon PCR Forward Primer =
5'-
TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG-3'
16S Amplicon PCR Reverse Primer =
5'-
GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAA
TCC-3'

Figure 4*. The primer nucleotide sequence used to amplify the 16S rRNA gene amplicon. At the 5' end in red is the target specific sequence, at the 3' end is the universal adapter sequence. Adapted from Illumina's 16S Metagenomic Sequencing Library Preparation, Preparing 16S Ribosomal RNA Gene Amplicons for the Illumina MiSeq System (protocol part # 15044223 Rev. B).*

PCR reactions were carried out on the Veriti thermal cycler (Thermofisher, cat. 4375786) in a 0.2ml, 96 well plate. The PCR mixture contained 2.5µL of DNA, 5µL of forward

primer (1µM), 5µL of reverse primer (1µM) and 12.5µL of I-5™ 2X High-Fidelity Master Mix (Molecular Cloning Laboratories (MCLAB)). All samples used the following temperature regime: 2min at 95ºC, 25 cycles of 95ºC 30sec, 55ºC for 30sec, 72ºC for 30sec, final extension at 72ºC for 5min. PCR cleanup was then performed using Beckman Coulter AMPure XP beads (Peffers *et al.,* 2014). After bringing the beads to room temperature 20µL were added to each PCR product on the 96 well plate. After mixing via pipetting up and down for 10 times, they were incubated at room temperature for 5 minutes. The 96 well plate was placed on the magnetic stand for 2 minutes, the supernatant was then removed and discarded. Using a freshly prepared 80% ethanol solution, a wash was applied twice by pipetting 200µL of the solution into each well and then removing. The 96 well plate was then removed from the magnetic stand and 52.5µL of 10 mM Tris pH 8.5 was used to resuspend the beads. After a 2-minute incubation the 96 well plate was placed back on the magnetic stand until the supernatant cleared. The supernatant containing our purified DNA was then transferred to a new sterile 96 well plate.

16S Library Preparation

Unique sequencing indices (Illumina Nextera, index kit A) were used to enable identification of individual samples in the sequencing data. The dual indices were incorporated into each 16S PCR amplicon sample via a limited PCR, the components of which were 5µL of purified PCR product, 5µL Nextera XT index primer 1, 5µL of Nextera XT index primer 2, 25µL of I-5™ 2X High-Fidelity Master Mix (Molecular Cloning Laboratories (MCLAB)), and 10µL of PCR grade water. The index PCR conditions were, 3min at 95ºC, 8 cycles of 95ºC for 30sec, 55ºC for 30sec, 72ºC for 30sec and a final extension of 72ºC for 5min. Final sample cleanup was performed using Beckman Coulter AMPure XP beads (Greenwald *et al.*, 2019) using the same purification method as the 16S PCR amplicons (described above), with the alteration of using 56µL of beads and 27.5µL of 10mM Tris pH 8.5 to resuspend the beads. The purified libraries (16S amplicons with dual indices and Illumina sequencing adapters incorporated) were then quantified using an Invitrogen™ Qubit™ 3 Fluorometer and normalized to 4nM using the equation in figure 5.

$$\frac{concentration\ in\ ng/\mu L}{660\ g/mol \times\ average\ library\ size} \times 10^6 = concentration\ in\ nM$$

Figure 5*. The equation used to convert concentration in ng/µL to concentration in nM. Adapted from, Preparing 16S Ribosomal RNA Gene Amplicons for the Illumina MiSeq System, protocol part # 15044223 Rev. B*

The libraries were pooled at an equal volume of 10µL in a new sterile 1.7ml microcentrifuge tube and using the MiSeq System Denature and Dilute Libraries Guide (Illumina, Document # 15039740 v10) were prepared for analysis on the MiSeq

(Illumina) instrument. In a new sterile 1.7ml microcentrifuge tube, 5µL of the pooled libraries (4nM) were combined with 5µL of 0.2M NaOH for 5 minutes to convert dsDNA to ssDNA. Immediately after the 5 minutes, 990µL of 4^{O}C chilled HT1 (Hybridization Buffer, Illumina MiSeq reagent kit, V3-600) was used to dilute the denatured libraires to 20pM, the libraries were then further diluted to a loading concentration of 4pM. PhiX (PhiX Control v3, Illumina) was added to the pooled libraries to account for sequence similarity and optimize cluster differentiation. This was prepared by adding 2µL of PhiX (PhiX Control v3, Illumina) to 3µL of 10mM Tris pH 8.5 to make a 4nM solution, this was then denatured and diluted to the same 4pM concentration using the same method as the libraries. The PhiX and sample libraries were then combined in a new 1.7ml microcentrifuge tube by adding 30µL of the PhiX library to 570µL of the pooled sample libraries. Lastly, the combined libraries tube was placed on a heating block at 96^{O}C for 2mins, after which it was placed on ice until ready to be pipetted into the MiSeq reagent cartridge (MiSeq Reagent Kit v3 (600-cycle)) for analysis on the MiSeq instrument.

Running the samples on the MiSeq

Using the Illumina experiment manager software, a sample sheet was constructed with the relevant information (unique dual index sequence for each sample, experiment type, reagent cartridge type, number of runs) required by the MiSeq instrument to analyze the sample. This was uploaded to the MiSeq system software and after loading the flow cell and reagent buffer (MiSeq Reagent Kit v3 (600-cycle)) the combined libraries were pipetted into the reagent cartridge which was then also loaded into the MiSeq instrument. After completion of the MiSeq analysis, raw sequence data of the 16S libraries was demultiplexed using the onboard analysis of the MiSeq. This produced a set of fastq files for each sample corresponding to the index used in library preparation. The resulting fastq files produced by the MiSeq were uploaded via basespace (online server, www.basespace.illumina.com) to a user account, these files were then downloaded and analyzed using the microbiome analysis package, QIIME2.

Analysis of the 16S sequencing data using QIIME2

Import and quality filter

Microbiome bioinformatics were performed using QIIME2 2021.2 (Bolyen *et al.*, 2019). To analyze the data QIIME2 (version 2021.2 1614815453) was installed in addition to Oracle VM VirtualBox, on a Dell laptop with a Windows® 10 operating system. The import tool was used to add sequences and convert fastq data into a .qza file, which can then be analyzed (Bolyen *et al.*, 2019). The q2-demux plugin was then used and it provides an interactive summary of the data so we can see the quality in Q values. This is achieved by converting the data from an artifact (.qza) to a visualization file (.qzv). The dada2 plugin (Callahan *et al.*, 2016) was used to filter sequences by quality. This plugin filters the data according to values we decide, based on the previous plugin (Callahan *et al.*, 2016). We trimmed data using the following settings: -trim-left-f 10, trim-left-r 10, trunc-len-f 250, trunc-len-r 250. In order to create a phylogenetic alignment the mafft-

fasttree plugin was used, this constructs a tree of alignment using the combined sequences from the previous plugin (Katoh *et al.*, 2002).

Alpha and beta diversity

We used the diversity core-metrics-phylogenetic plugin to produce alpha and beta diversity analysis, this uses the phylogenetic tree created in conjunction with the feature table to produce a wide array of files. In addition, we used the p-sampling-depth command to normalize the output to a frequency of 80,000 rarefied sequences from each sample. This resulted in artifacts for alpha diversity (diversity within one sample) to be used by other plugins, such as faith_pd_vector.qza, shannon_vector.qza. Files are also created which are already in the visualization format, these are all beta diversity (diversity between two or more samples) analysis output and all use the emperor (Vázquez-Baeza *et al.*, 2013) visualization software. The alpha-group-significance plugin was then used. This takes the output of the previous plugin to create a visualization, using boxplots of the faith_pd_vector. This was used to test associations between metadata columns and alpha diversity (Faith, 1992). To test the differences, the beta-group-significance plugin uses beta diversity to test associations between an internal group and an external group, in this case location. This compares similarity between samples from different locations (Lozupone and Knight, 2005). The alpha-rarefaction plugin was used to explore sampling depth (if the richness of the samples has been fully observed or sequenced) against alpha diversity by specifying various values that pick random amplified sequence variants (ASVs) to represent the sample. After testing multiple depths, we selected a count of 4000 random ASV picks, as this was the minimum number without samples dropping out of the analysis due to low total frequencies closer to the minimum sampling depth than the maximum sampling depth. This produces a .qzv file that has two plots, one where the richness can be analyzed to see if samples reached a plateau indicating a close to maximum level of sequencing depth has been reached. The other plot shows which samples remain after the alpha-rarefaction filtering.

Taxonomic classification

To classify each sequence by taxa, the feature-classifier sklearn (Pedregosa *et al.*, 2011) was used. This plugin takes the sequences created and compares them to a known sequence database, in the form of a file (classifier.qza). The output is now a taxonomic breakdown of the sample data in an artifact form. The classifier was constructed using the database silva-138-99-tax-515-806.qza (Quast *et al.*, 2013) trained on the 16S sequence specific primer set used during the library preparation. To use the output file, the metadata tabulate plugin was used. This plugin tabulates the output file into a tubular format, which can then be used by other plugins. The QIIME taxa bar plot was used to construct an interactive bar plot of the taxonomy, which can then be separated according to 7 levels of taxonomy. If multiple samples are included, they can be viewed side by side. The QIIME2 analysis of 16S data were all viewed using QIIME2view

(https://view.qiime2.org/) this can only display the .qzv file type. Please note the full script and metadata can be found in the appendix (Figure 1 and Table 1).

Whole Genome Shotgun (WGS) Library Preparation

WGS libraries were constructed for each sample using the Illumina protocol (Nextera DNA Flex Library Prep; Reference Guide (document # 1000000025416 v07)). We also included a positive control (ZymoResearch, ZymoBIOMICS Microbial Community DNA Standard), which contained a mixture of genomic DNA of ten microbial strains. As the Nextera DNA Flex Library Prep method is compatible with DNA input of 1-500ng no additional normalization was required. Therefore, we used the same 5ng/µL DNA that we previously used to prepare the 16S libraries. A total of 30µL of each samples DNA was added separately to a 96 well plate. After vortexing at max speed 132µL of BLT (Bead-linked transpososomes, Illumina Nextera DNA Flex kit) was added to 132µL of TB1 (Tagmentation Buffer 1, Illumina Nextera DNA Flex kit) to create a tagmentation master mix. A total of 20µL of this master mix was added to each sample in the 96 well plate, this mixed using a multichannel pipette and sealed using adhesive PCR Plate Seals (Thermofisher, cat. AB0558). This was then run on the Veriti thermal cycler (Thermofisher, cat. 4375786) using a program of 55^{O}C for 15 minutes, 10^{O}C hold. After this 10µ of TSB (Tagmentation Stop Buffer, Illumina Nextera DNA Flex kit) was added to each sample and pipetted to resuspend the BLT. The plate was placed on the magnetic stand for 3 minutes, supernatant was then removed and discarded. The plate was removed from the magnetic stand and then washed twice by adding 100µL of TWB (Tagmentation Wash Buffer, Illumina Nextera DNA Flex kit), pipetting to resuspend the beads, placing the plate back on the magnetic stand for 3 minutes and removing and discarding the supernatant by pipetting. After the second wash the plate was removed from the magnetic stand and 100µL of TWB was added using a pipette while mixing to resuspend the beads. The plate was placed back on the magnetic stand until the PCR master mix was prepared. The PCR master mix contained 240µL of EPM (Enhanced PCR Mix, Illumina Nextera DNA Flex kit) and 240µL of PCR grade water. The supernatant was removed and discarded from the 96 well plate that remained on the magnetic stand the plate was then removed from the magnetic stand and 40µL of the PCR master mix was added by pipetting while mixing to resuspend the beads. The plate was centrifuged at 280 x g for 3 seconds and dual indices were then added using 5µL of i7 adapter and 5µL of i5 adapter. The plate was again centrifuged at 280 x g for 3 seconds and placed on the Veriti thermal cycler (Thermofisher, cat. 4375786), the following program was then run 68^{O}C for 3 mins, 98^{O}C for 3mins, 8 cycles of 98^{O}C for 45 sec, 62^{O}C for 30 sec, 68^{O}C for 2 mins, a final stage of 68^{O}C for 1 min followed by a 10^{O}C hold. After the completion of the PCR the plate was centrifuged at 280 x g for 3 seconds and placed on the magnetic stand. From the supernatant 45µL was transferred from each well to the corresponding well of a new 96 well plate. To each well 45µL of vortexed SPB (Sample Purification Beads,

Illumina Nextera DNA Flex kit) were added using a pipette to mix. The plate was incubated for 5mins and then placed on the magnetic stand until the liquid was clear. To a new 96 well plate 15µL of SPB were added to each well, 125µL of the supernatant from the plate on the magnetic stand was then added to the corresponding wells of the new plate (containing 15µL of SPB) using a pipette to mix and incubating for 5 mins. After discarding the first plate from the magnetic stand, the new plate was then placed on the magnetic stand for 5 mins and the supernatant was removed and discarded. A wash was then performed twice by adding 200µL of 80% ethanol to the wells then removing by pipetting and discarding. After the second wash the plate was removed from the magnetic stand and 32µL of RSB (Resuspension Buffer, Illumina Nextera DNA Flex kit) was added using a pipette to mix and resuspend the beads. After a 2min incubation the plate was placed back on the magnetic stand for 2mins, 30µL of the supernatant was then transferred to a new 96 well plate. The samples were then pooled by adding 5µL of each library into a new sterile 1.7ml microcentrifuge tube, the pooled library was then quantified using an Invitrogen™ Qubit™ 3 Fluorometer. To validate the WGS library, the pooled sample was run on the Agilent 2100 Bioanalyzer to confirm the presence of libraries and approximate size (in base pairs) distribution. This base pair value was used to convert the concentration in ng/µL to nM by utilizing the equation shown in figure 5, the pooled libraries were then diluted to 4nM.

Using the MiSeq System Denature and Dilute Libraries Guide (Illumina, Document # 15039740 v10) the pooled sample libraries were prepared for analysis on the MiSeq (Illumina) instrument. In a new sterile 1.7ml microcentrifuge tube, 5µL of the WGS pooled libraries (4nM) were combined with 5µL of 0.2M NaOH for 5 minutes to convert dsDNA to ssDNA. Immediately after the 5 minutes, 990µL of 4^{O}C chilled HT1 (Hybridization Buffer, Illumina MiSeq reagent kit, V3-600) was used to dilute the denatured libraires to 20pM, the libraries were then further diluted to a loading concentration of 12pM. Due to the expected diversity of fragments being sequenced, a Phix spike was not used. Lastly, the combined libraries tube was placed on a heating block at 96^{O}C for 2mins, after which it was placed on ice until ready to be pipetted into the MiSeq reagent cartridge (MiSeq Reagent Kit v3 (600-cycle)) for analysis on the MiSeq instrument.

Running the samples on the Illumina MiSeq System

Using the Illumina experiment manager software, a sample sheet was constructed with the relevant information (unique dual index sequence for each sample, experiment type, reagent cartridge type, number of runs) required by the MiSeq instrument to analyze the sample. This was uploaded to the MiSeq system software and after loading the flow cell and reagent buffer (MiSeq Reagent Kit v3 (600-cycle)) the combined libraries were pipetted into the reagent cartridge which was then also loaded into the MiSeq instrument.

After completion of the MiSeq analysis, raw sequence data of the WGS libraries was demultiplexed using the onboard analysis of the MiSeq. This produced a set of fastq files for each sample corresponding to the index used in library preparation. The resulting fastq files produced by the MiSeq were uploaded via basespace (online server, Illumina) to a user account, these files were then downloaded and analyzed using The Metagenomics RAST server (MG-RAST).

Analysis of the WGS sequencing data using MG-RAST

To analyze the WGS data, fastq files and metadata were uploaded to https://www.mg-rast.org/. It was at this point that we could see that Oyster Creek sample 2 had failed during the sequencing, it was therefore omitted from all further analysis. The WGS fastq files were joined using the MG-RAST tool, prior to submission to the pipeline. Also submitted to MG-RAST was the targeted 16S for comparison to the WGS fastq files. The dataset was submitted to version 4.0.3 of the MG-RAST (Meyer *et al.*, 2008) pipeline with the options of dereplication and dynamic trimming selected. A minimum quality was chosen as a Q score of 15 and sequences were screened for *H. sapiens* using NCBI v36. The initial sequence statistics were calculated using DRISEE (Keegan *et al.*, 2012) and Jellyfish (Marçais and Kingsford, 2011). Adapter trimming using Skewer (Jiang *et al.*, 2014) was performed followed by denoising and normalization using fastq-mcf (Aronesty, 2013). Removal of sequencing artifacts and host DNA contamination was carried out using Bowtie2 (Langmead and Salzberg, 2012). RNA feature identification or gene calling used the plugin SortMeRNA (Kopylova, Noé and Touzet, 2012) RNA similarity search used Blat (Kent, 2002) followed by gene calling using protein coding features using FragGeneScan (Rho, Tang and Ye, 2010). Amino acid sequence clustering was performed using the plugin CD-HIT (Fu *et al.*, 2012). The protein similarity search used BLAT and the M5NR database (Wilke *et al.*, 2012). This database is particularly useful as it combines source databases from Genbank (NCBI), IMG (JGI), KEGG, PATRIC (VBI), RefSeq (NCBI), SEED, SwissProt (UniProt), TrEMBL (UniProt), eggNOG, COG (eggNOG), GO, KO (KEGG), NOG (eggNOG) and Subsystems (SEED). The following steps were then performed by MG-RAST scripts; protein similarity annotation, RNA similarity annotation, merge, and index similarities, annotate and index similarities, feature abundance profile, LCA abundance profile, data source profile, extract features with no similarity hits, abundance profile load, abundance profile build and load and summary statistics. The output of MG-RAST was an interactive analysis for each sample that could then be combined using the server analysis tools.

Results

Environmental differences of sample sites

We observed that Oyster Creek has a higher temperature than all other samples (Table 1), approximately 4.6ºC above the mean of all other sites combined. In addition, Oyster Creek had the lowest salinity in parts per thousand compared to all other sites. Both Forked River and Sunrise Beach had the same bottom type (sand), while Oyster Creek and Traders Cove had the same bottom type (mud).

Table 1. Metadata recorded at all sample sites.

Sample site	Mean temperature (Celsius)	Mean salinity (ppt)	Mean Dissolved oxygen (mg/L)	Mean depth (meters)	Bottom type
Forked River	5.3	25.5	13.01	1.5	Sand
Oyster Creek	10	14.2	13.12	0.9	Mud
Sunrise Beach	5.1	16.3	12.06	1.1	Sand
Traders Cove	5.6	21.9	13.44	2	Mud

16S data analysis of bacterial diversity

The results of alpha diversity (the variance within one sample) are shown in Figures 6, 7 and 8. This is a qualitative measure of community richness, that incorporates phylogenetic relationships between the features using metadata. The phylogenetic relationships were analyzed using Faith's Phylogenetic Diversity (PD)(Faith, 1992) and compared using the pairwise Kruskal-Wallis tests with Benjamini-Hochbery false discovery rate corrections for multiple samples. Faith's PD is defined as the sum of the branch lengths of a phylogenetic tree, connecting all species in the target assemblage (Pellens and Grandcolas, 2016). The alpha diversity in relation to location name, produced a significant p-value of 0.044 when comparing all groups. However, when using a pairwise comparison the corrected p-value (q-value) was not significant for any of the pairwise calculations.

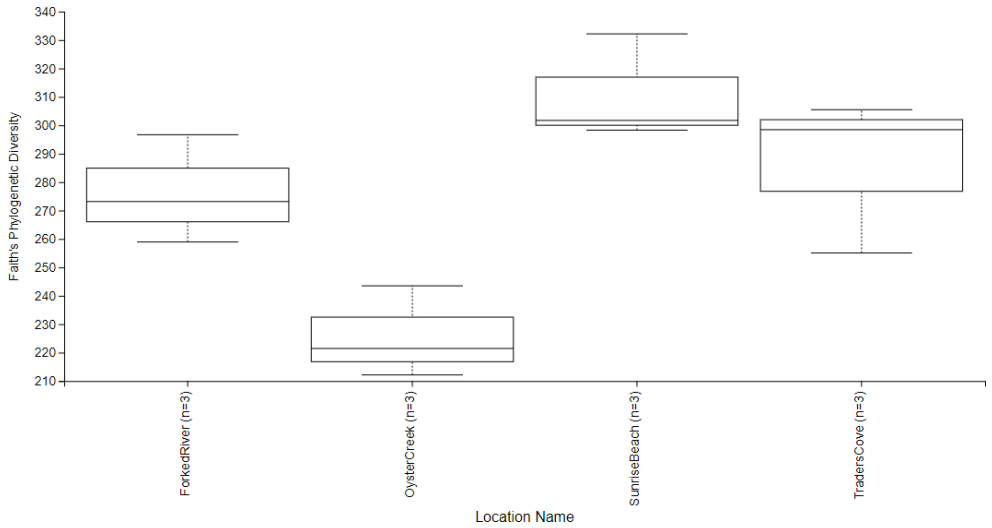


***Figure 6**. Boxplot of Faith's Phylogenetic Diversity, Alpha diversity between location names. This tests associations between metadata columns (sample site) and alpha diversity. The y-axis represents the sum of all branch lengths of a phylogenetic tree connecting all species. Constructed using QIIME2 view.*

We also used QIIME2 to calculate Beta diversity (the variance between multiple samples) of the 16S data. Shown in Figure 7 is a Bray-Curtis distance plot viewed using EMPeror (Vázquez-Baeza *et al.*, 2013). This principal component analysis plot is a quantitative measure of community dissimilarity, calculated by a compositional dissimilarly in ASV counts between samples. This analysis shows a value on each axis for explained variance in relation to a principal component. The locations are indicated by different colors and the bottom type at the sample site is indicated by shape. As shown all Forked River and Sunrise Beach samples are very closely aligned showing they are similar in terms of beta diversity of ASVs present, they are also both same bottom type (sand). All Traders Cove samples are the furthest away from all other samples.

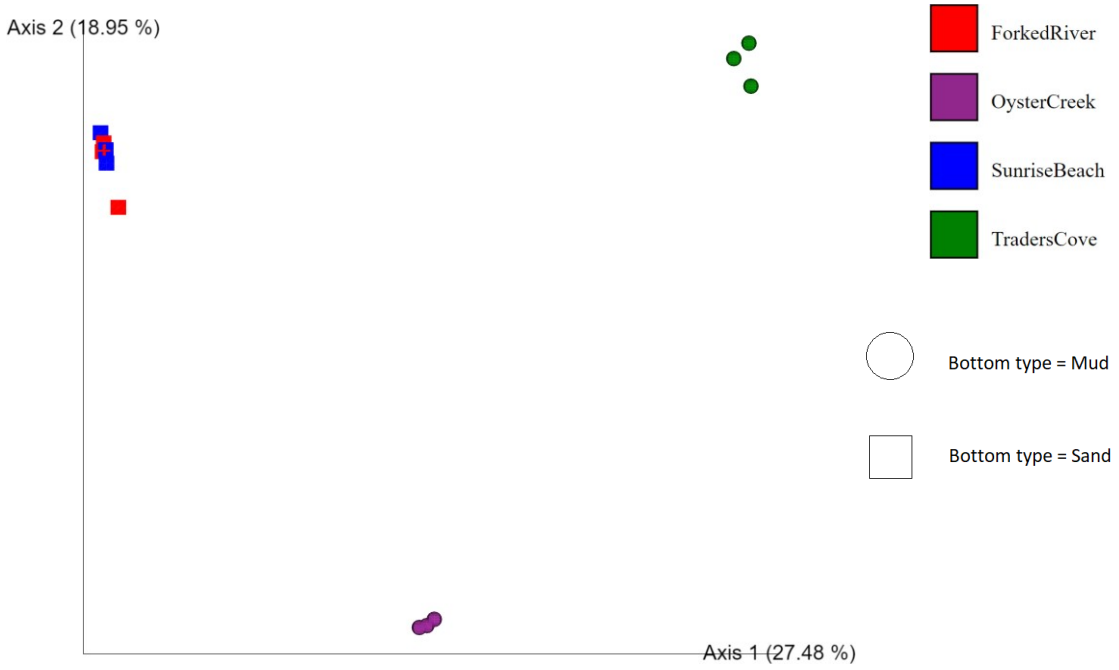


***Figure 7**. Bray-Cutis emperor plot of beta diversity viewed using Emperor, a web browser enabled tool of 3D visualization. Axes represent variation explained by a principal component. Different locations are indicated by different colors (Forked River=red, Oyster Creek=purple, Sunrise Beach=blue, Traders Cove=green). Bottom type is indicated by shape (sand=square, mud=circle).*

To further explore bottom type, we conducted a further beta diversity analysis, a PERMANOVA test. Shown in Figure 7 mud bottom type was compared to sand. Using 999 permutations a p-value of 0.02 was found (Table 1) pairwise analysis also produced a significant result of a q-value of 0.004.

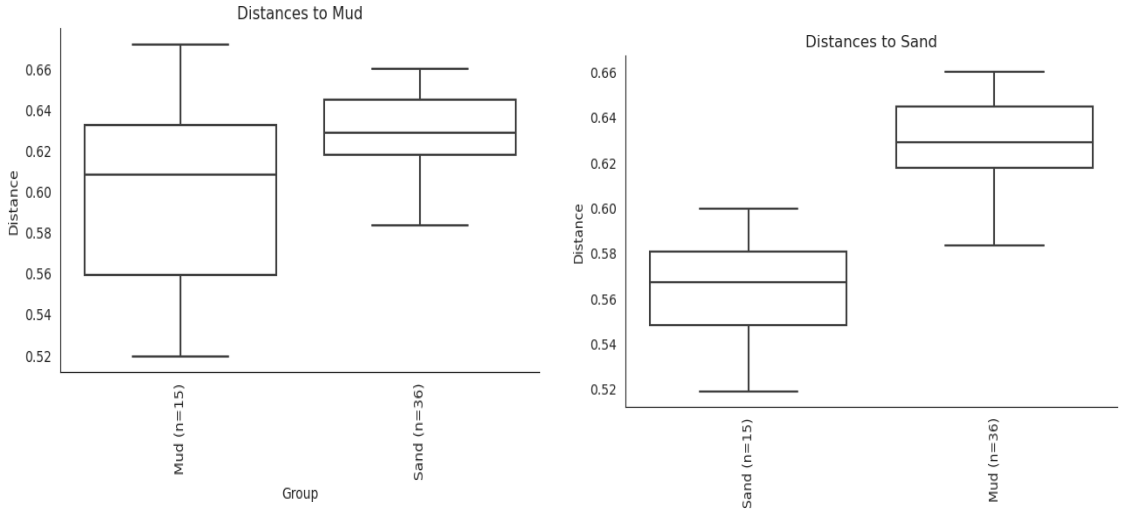


***Figure** 7. PERMANOVA plot comparing sand and mud bottom type in all samples, constructed using QIIME2 view.*

Table 2. PERMANOVA results for bottom type

method name	PERMANOVA
test statistic name	pseudo-F
sample size	12
number of groups	2
test statistic	1.95023
p-value	0.002
number of permutations	999

Table 3. Pairwise PERMANOVA results for bottom type

Group 1	Group 2	Sample size	Permutations	pseudo-F	p-value	q-value
Mud	Sand	12	999	1.95023	0.004	0.004

We used QIIME2 to create an alpha rarefaction plot (shown in Figure 8) of the 16S data, using Shannon alpha diversity index as a metric. This calculates the number of different ASVs and the similarity of frequency in a sample. This shows us if the maximum richness has been reached by our sequencing, indicated by reaching a plateau. Using 4000 iterations of subsampling in the 16S data and comparing to Shannon diversity, a plateau was reached for all sample sites.

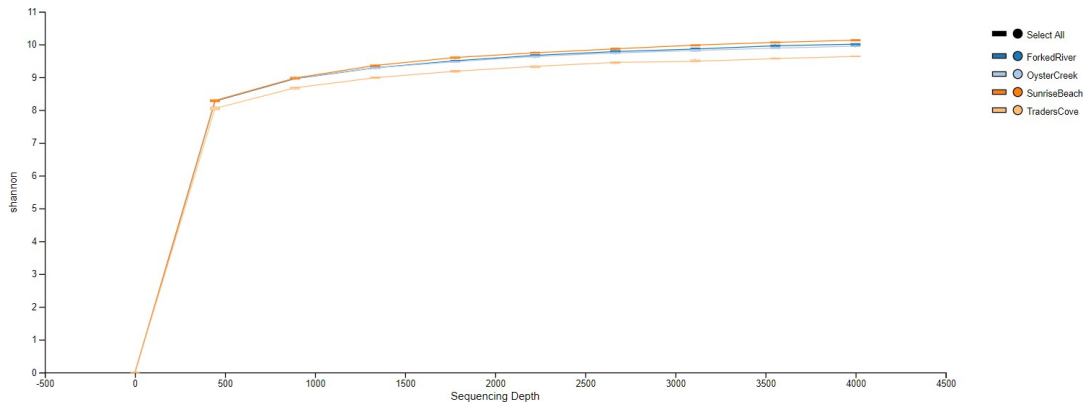


***Figure 8**. The alpha rarefaction plot, comparing sequencing depth using 4000 iterations of subsampling of the 16S data and Shannon diversity. Shown are the values for all sample sites separated by color, constructed using QIIME2 view.*

16S data analysis of taxonomy

Shown in Figure 9 is the taxonomic analysis of all samples at the phylum level (only the top 10 in relative abundance for each sample site are shown), using the 16S data and analyzed in QIIME2. In total 41 separate bacterial phyla were identified across all sample sites with Traders Cove 2 showing the greatest diversity with 37 separate phyla identified. Forked River 2 and Oyster Creek sites 1 and 3 showed the least diversity with 32 phyla identified for each sample site. To account for the variation in phyla count between samples, this was expressed as relative abundance in comparison to the total bacteria identified at the phylum level for each sample site. The most abundant phyla in each sample are very similar, with Proteobacteria being the most abundant, accounting for approximately 45% of the population across all sites. The other most abundant phyla are Bacteroidetes, Actinobacteria, Planctomycetes, Epsilonbacteraeota, Chloroflexi, Cyanobacteria, Acidobacteria and Firmicutes. When combined with Proteobacteria, these make up approximately 95% of all sample's composition

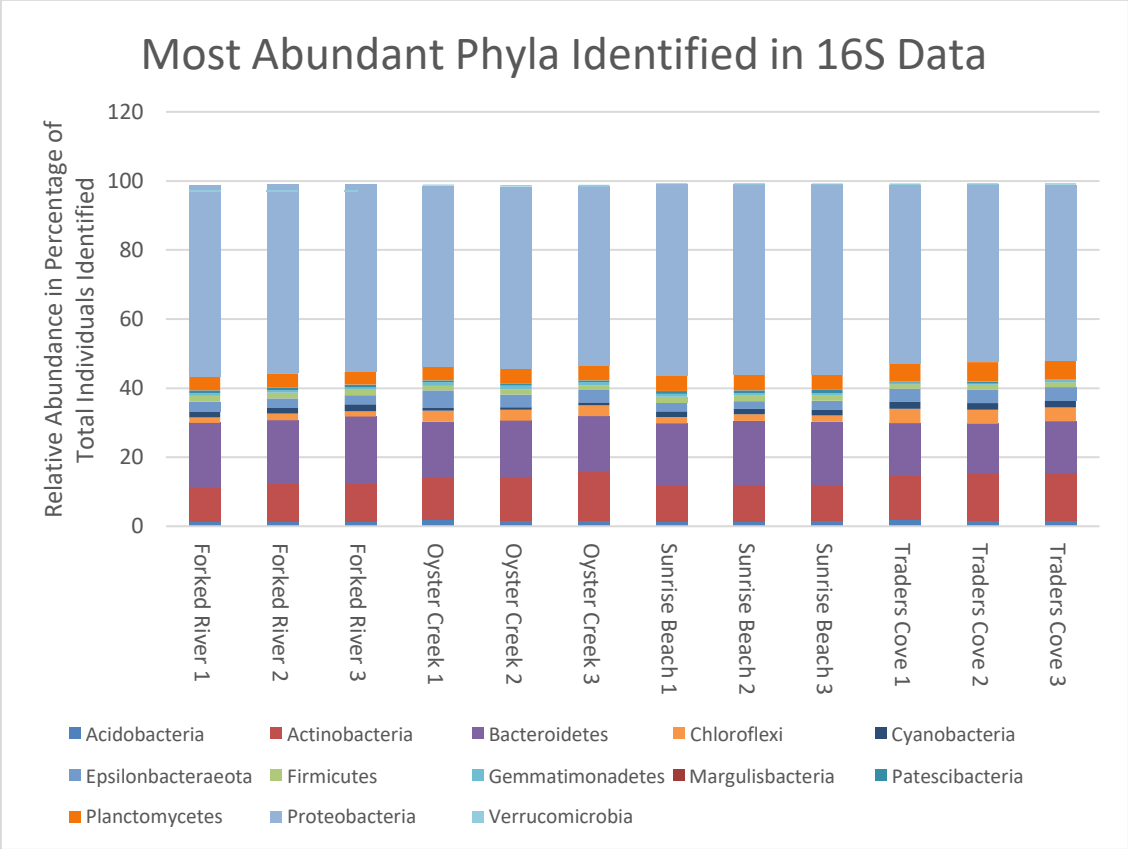


***Figure 9**. The results of taxonomic classification of 16S data according to phylum. A total of 41 separate phyla were identified across all sample sites, shown is the top 10 most abundant according to each sample site. Each phylum is represented by a different color. Constructed using Excel® with CSV data output from QIIME2.*

Using the same 16S dataset at the species level, relative abundance of species after classification is shown in Figure 10. Seven hundred and eighty-seven different species were identified. Although Traders Cove 2 showed the most diversity as a single sample, the mean of all Sunrise Beach samples showed the greatest richness at a combined location with 459 species. Similarly, although Forked River 3 showed the least diversity with 345 species, the mean of all Oyster Creek samples showed the least richness at a combined location with 368 species.

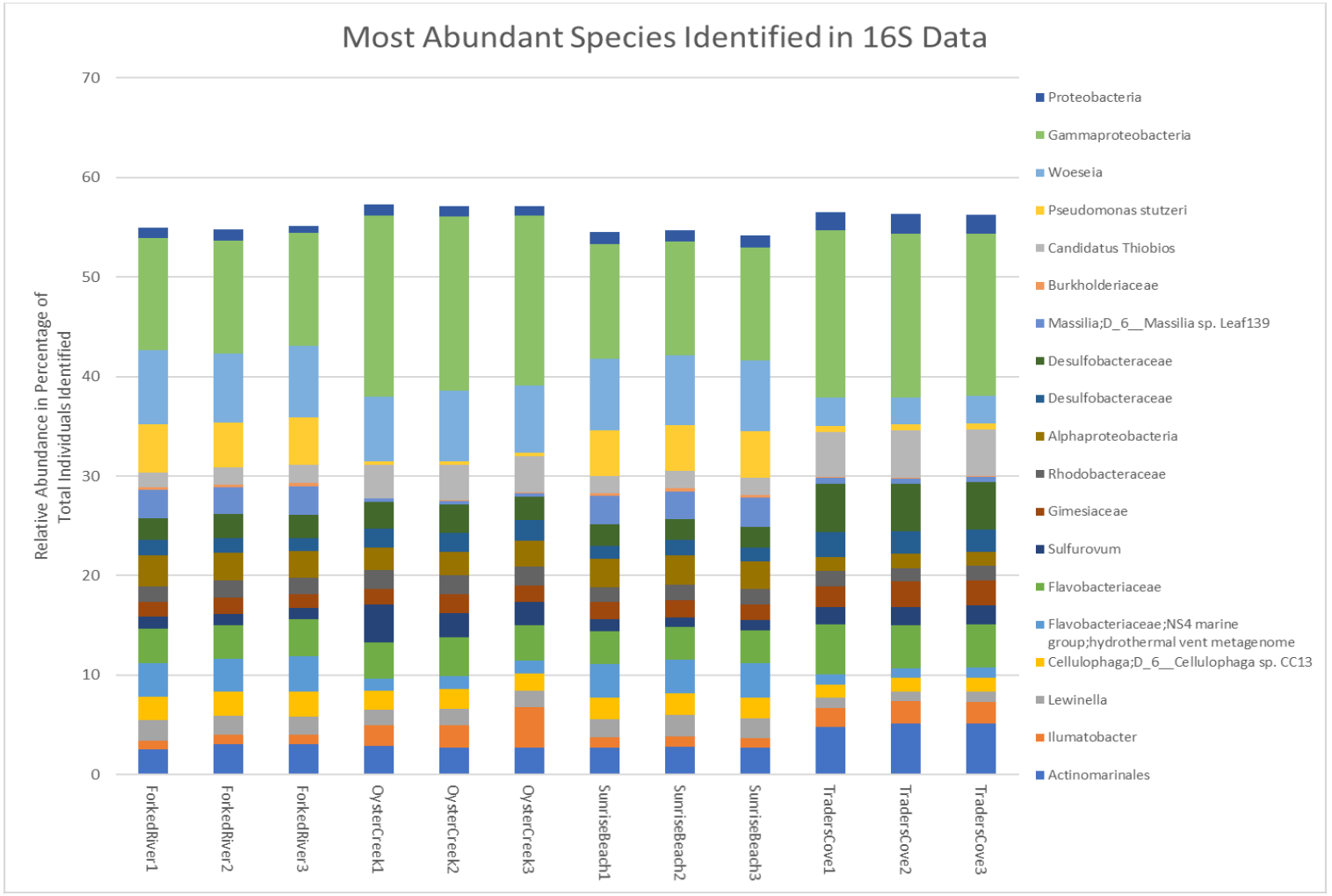


***Figure 10**. The results of taxonomic classification of 16S data according to Species. A total of 787 separate species were identified across all sample sites, shown is the top 10 most abundant according to each sample site. Each species is represented by a different color. Constructed using Excel® with CSV data output from QIIME2.*

WGS data MG-RAST taxonomic analysis

After analyzing the WGS data using MG-RAST we used the taxonomic data to construct Figure 11, this shows the 10 most abundant phyla identified. To account for the variation in phyla count between samples, this was expressed as relative abundance in comparison to the total bacteria identified at the phylum level for each sample site. We identified 48 separate phyla in all samples. As the sample Oyster Creek 2 had failed to sequence, it was not included in any MG-RAST analysis.

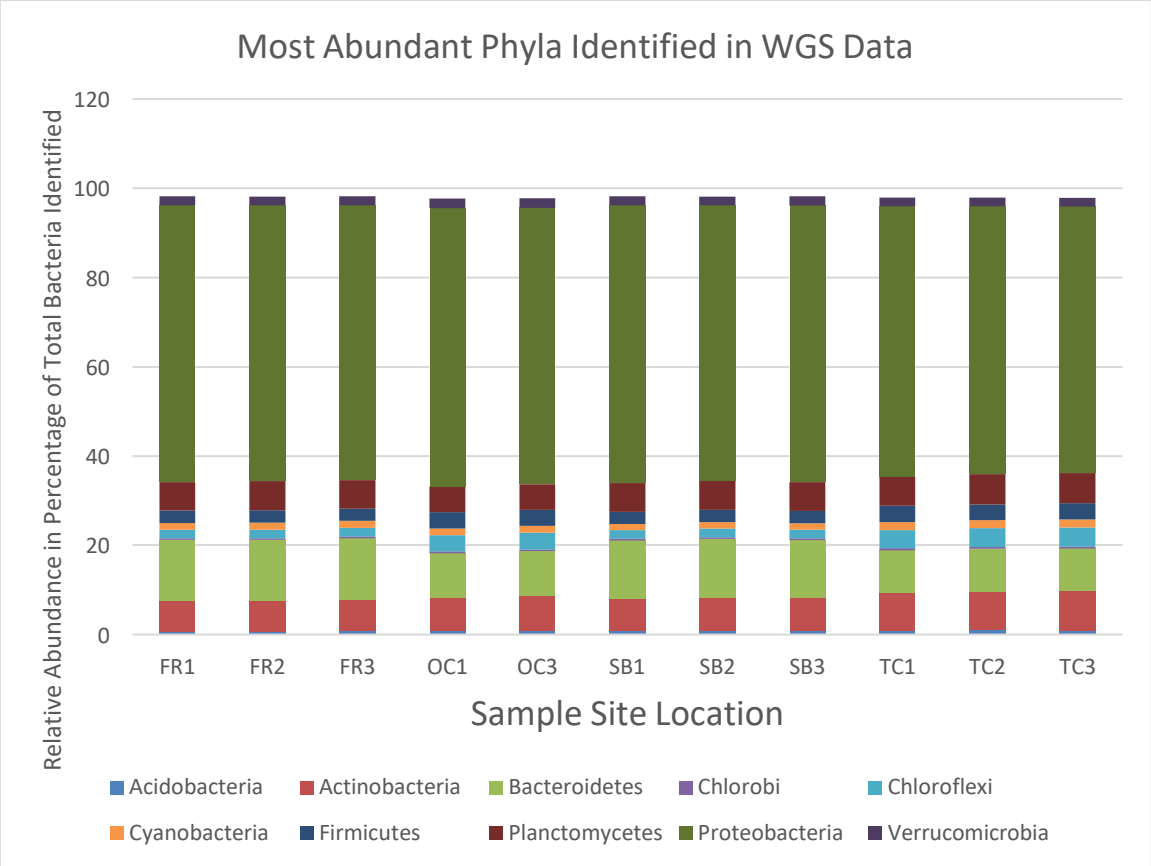


***Figure 11**. WGS data analyzed using MG-RAST showing the top 10 most abundant at the Phylum level, different colors indicate different Phyla identified. TC=Traders Cove, FR=Forked River, SB=Sunrise Beach, OC=Oyster Creek. Constructed using Excel® with taxonomic CSV data output from MG-RAST (WGS).*

Shown in Table 2 is the total number of individuals identified at each taxonomic level, both for WGS and 16S data. The number of individuals identified using WGS data are greater across all sites compared to targeted 16S data, at all taxonomic levels. Except for Oyster Creek 2 which we could not compare due to lack of sequencing data. Traders Cove 3 was an anomaly at all taxonomic levels compared to Traders Cove 1 and 2, in that we identified far less individuals. This is despite having good concentration values for the

DNA extraction and NGS library preparation, also the sequencing yield was comparable to Traders Cove 1 and 2. Sample site Oyster Creek 2 is omitted from the WGS data due to failure to sequence. Traders Cove 3 WGS data was only able to be obtained up to the Genus level.

Table 3. Number of individuals identified at all levels of taxonomic analysis, for both 16S and WGS. Constructed using Excel® from the taxonomic data output of QIIME2 (16S data) and MG-RAST (WGS data).

Taxonomic level							
Phylum	Class	Order	Family	Genus	Species	Sample Site	
33	74	183	286	375	428	Forked River 1	16S Data
32	76	187	286	370	417	Forked River 2	
33	69	164	247	310	345	Forked River 3	
32	73	180	252	313	350	Oyster Creek 1	
33	73	178	268	352	395	Oyster Creek 2	
32	72	176	246	319	359	Oyster Creek 3	
35	78	202	305	403	461	Sunrise Beach 1	
33	78	196	303	404	462	Sunrise Beach 2	
35	77	190	301	397	455	Sunrise Beach 3	
33	77	183	282	372	419	Traders Cove 1	
37	85	208	323	446	511	Traders Cove 2	
33	79	185	284	373	429	Traders Cove 3	
47	143	284	483	941	2040	Forked River 1	WGS data
47	151	289	482	941	2038	Forked River 2	
46	150	277	473	931	2009	Forked River 3	
47	127	234	415	840	1818	Oyster Creek 1	
46	124	227	411	841	1818	Oyster Creek 3	
48	152	299	513	979	2138	Sunrise Beach 1	
47	157	300	508	966	2083	Sunrise Beach 2	
46	147	283	479	925	2018	Sunrise Beach 3	
45	144	271	460	906	1982	Traders Cove 1	
46	144	264	450	895	1992	Traders Cove 2	
46	148	290	496	886	No data	Traders Cove 3	

To compare MG-RAST and QIIME2 we took the total number of different species identified from each analysis. As shown in figure 12, MG-RAST identified the most species using WGS data (Sunrise Beach 2080) QIIME2 identified the most species using the 16S method (Sunrise Beach 459). Oyster Creek has the lowest richness across all analyses in both WGS (1818) and 16S targeted (368).

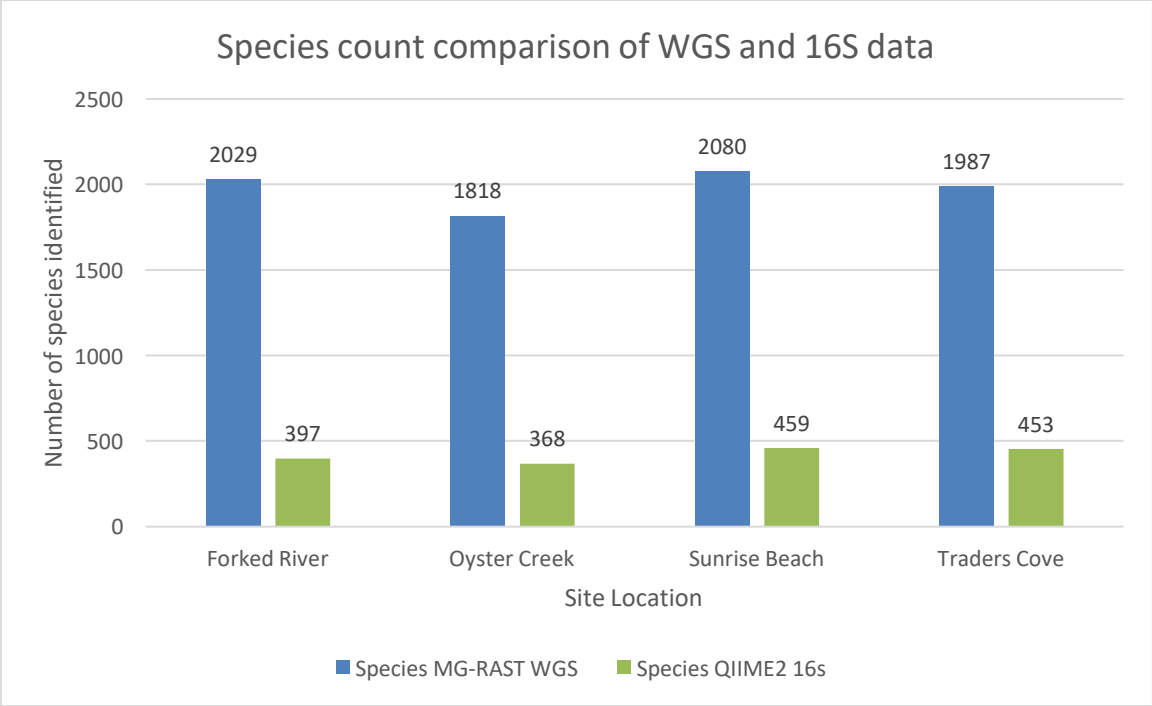


***Figure 12**. Comparison of WGS data MG-RAST analysis and 16S QIIME2 analysis at species level, using the mean of the sites. Oyster Creek had the least number of species identified using both WGS and 16S. While Sunrise Beach had the greatest number of species identified. Constructed using Excel® from the taxonomic data output of QIIME2 (16S data) and MG-RAST (WGS data).*

Shown in Figure 13 is a comparison of *Verrucomicrobia* and *Deinococcus-Thermus* relative abundance across all sample sites in both WGS and 16S data. Oyster Creek 3 sample site shows the greatest relative abundance of *Verrucomicrobia* in both WGS and 16S data. Traders Cove samples 1, 2 and 3 show the greatest abundance of *Deinococcus-Thermus* in the WGS data, while Forked River shows the greatest abundance in the 16S data.

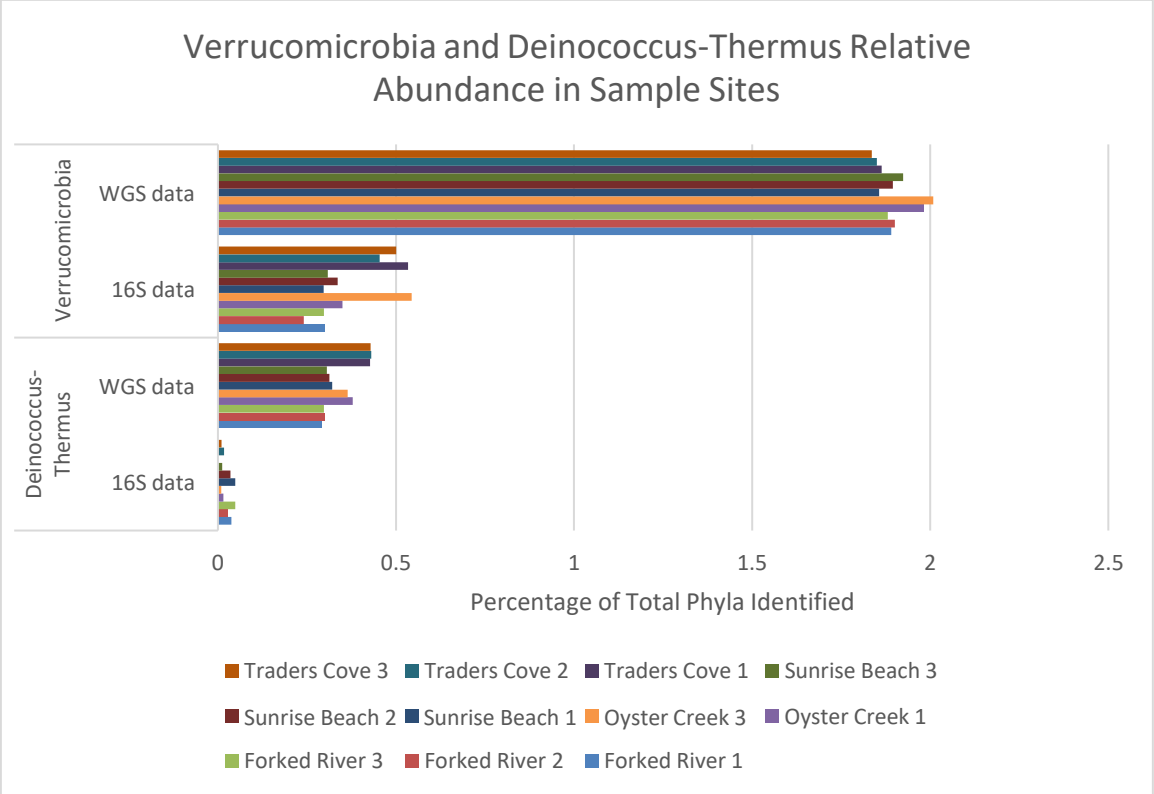


Figure 13. Comparison of *Verrucomicrobia* and *Deinococcus-Thermus* relative abundance in 16S and WGS data across all sample sites. Expressed as a percentage of total phyla identified. Constructed using Excel® from the taxonomic data output of QIIME2 (16S data) and MG-RAST (WGS data).

Shown in Figure 14 is a comparison of *Betaproteobacteriales* relative abundance across all sample sites in both WGS and 16S data. Oyster Creek sample sites 1 and 3 shows the least relative abundance of *Betaproteobacteriales* in both WGS and 16S data. Traders Cove samples 1, 2 and 3 show the greatest abundance of *Betaproteobacteriales* in the WGS data, while the Sunrise Beach 3 sample shows the greatest abundance in the 16S data.

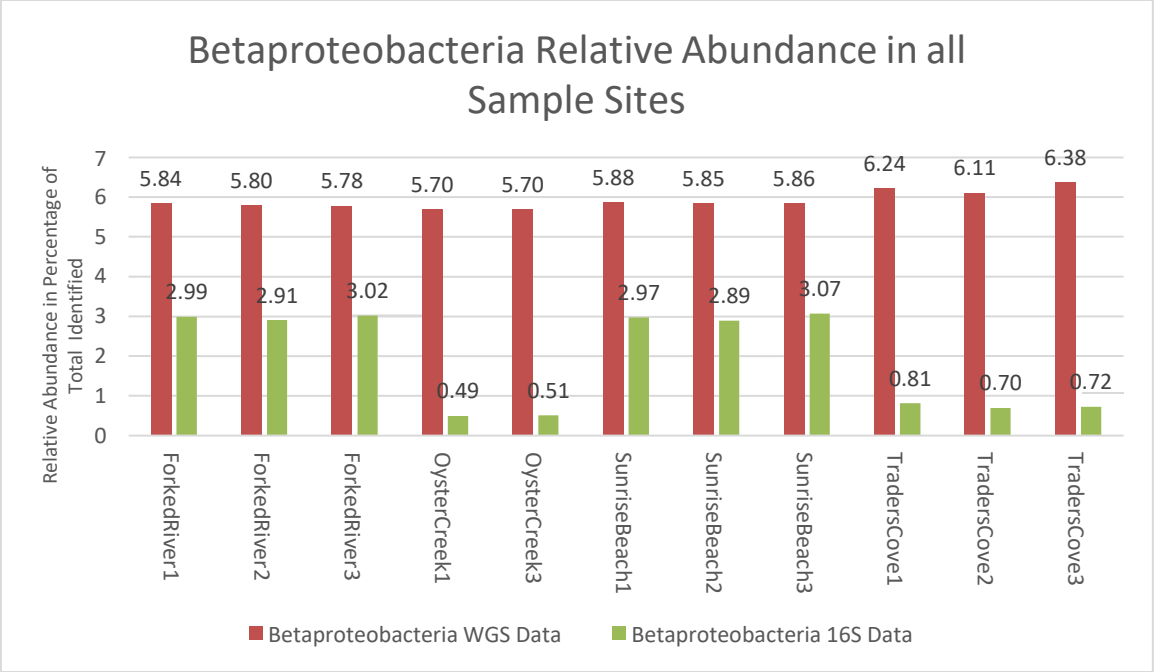


Figure 14. Comparison of *Betaproteobacteriales* relative abundance in 16S and WGS data across all sample sites. Expressed as a percentage of individuals identified. Constructed using Excel® from the taxonomic data output of QIIME2 (16S data) and MG-RAST (WGS data).

Shown in Figure 15 is a comparison of *Planctomycetes* Relative Abundance in all samples in both WGS and 16S data. Oyster Creek 1 and 3 samples show the least relative abundance of *Planctomycetes* in the WGS data. Forked River 3 shows the least relative abundance in the 16S data.

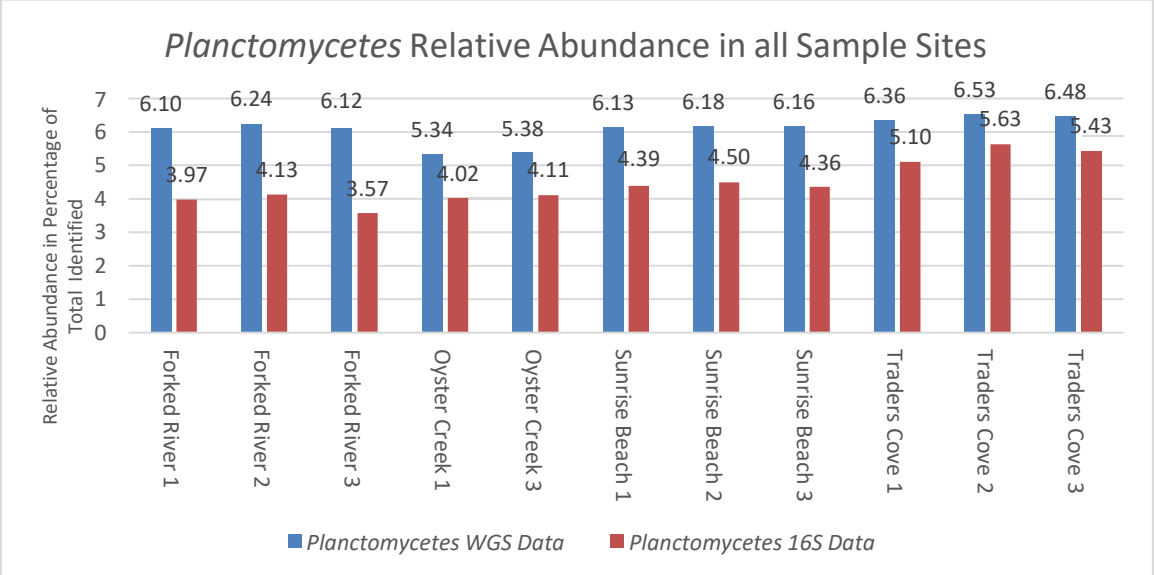


Figure 15. Comparison of *Planctomycetes* relative abundance in 16S and WGS data across all sample sites. Expressed as a percentage of total phyla identified. Constructed using Excel® from the taxonomic data output of QIIME2 (16S data) and MG-RAST (WGS data).

Shown in Figure 16 is a comparison of Cyanobacteria Relative Abundance compared to all phyla identified for each sample in both 16S and WGS data. Oyster Creek 1 and 3 sample sites show the least relative abundance of Cyanobacteria in the 16S data. Sunrise Beach samples 1, 2 and 3 show the least relative abundance in the WGS data.

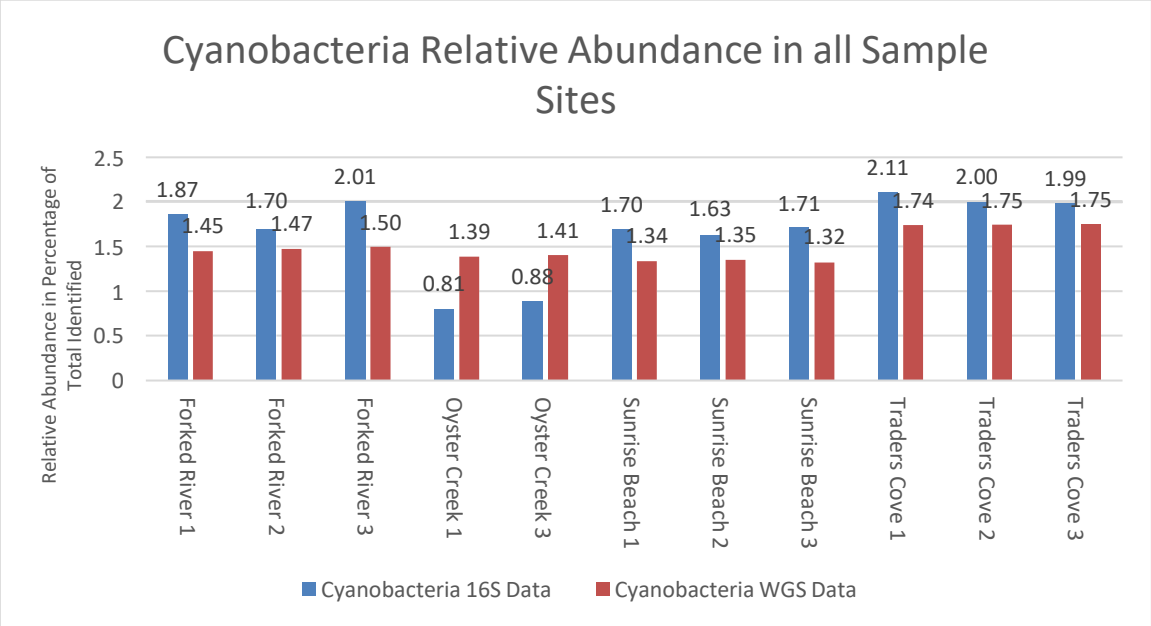


Figure 16. Comparison of *Cyanobacteria* relative abundance in 16S and WGS data across all sample sites. Expressed as a percentage of total phyla identified. Constructed using Excel® from the taxonomic data output of QIIME2 (16S data) and MG-RAST (WGS data).

Discussion

In comparison to all other sites, Oyster Creek has the highest mean temperature and lowest salinity. We found that three variables were significant when comparing all groups, while only one was significant when using a corrected pairwise comparison. Location (Fig. 6) was significant when comparing all groups (p-value 0.044), however pairwise comparison showed no significance when corrected using q-value in all the sites (all above 0.05). Salinity was analyzed as a category in parts per thousand either between 15-20, <15 or >20. Analysis found that a comparison of all groups was significant with p-value of 0.022 but pairwise was not (all above 0.05). Water temperature as a category above or below 10^{O}C was significant across all groups (p-value 0.013) and pairwise (q-value 0.013). This shows that although location and salinity have significance when looking at all groups as a large dataset, the pairwise comparisons were not significant. For example, Oyster Creek as a location was significantly different in terms of the diversity of ASVs when compared to the whole dataset, but when comparing to just one location, it was not. Temperature differences indicate that the sites above 10^{O}C are

significantly different in diversity to the whole dataset as well as to sites where the temperature was below 10^{O}C. It should also be noted that Faith's PD does not consider species abundance, just their presence or absence.

Bray-Curtis distance graph was constructed using beta diversity (Figure 7) analysis of 16S data in QIIME2, this compares community dissimilarity where 0 indicates all ASVs are shared and 1 indicates no ASVs are shared. This data was analyzed using a principal component analysis plot, this enables us to view the percentage of variation that can be explained by the principal components. Both Forked River and Sunrise Beach show highly similar diversity in ASVs as indicated by very tight clustering on both axes. While Traders Cove and Oyster Creek are dissimilar to all other groups, the replicates within these groups are clustered closely together indicating similar diversity in relation to the principal components. The fact Traders Cove is showing the most dissimilarity is somewhat expected as although it shares many of the same environmental factors it is the furthest geographically from all other sites. The similarity of the bottom type (sand) at these sites may account for the highly similar diversity between Forked River and Sunrise beach. To evaluate if bottom type was playing a role we conducted a further beta diversity analysis, a PERMANOVA test. This showed that bottom type was in fact significant in terms of diversity between samples indicated by a p-value of 0.02. A recent publication by Boey *et al.,* 2021, showed that from sandy to muddy sediment type communities were sensitive to changes, with significant changes at only a 3% increase in mud. In addition nitrogen cycling was found to be more prevalent in muddier sediments, with mud content being a strong environmental driver of diversity (Boey *et al.*, 2021).

To verify if we had effectively analyzed these samples with enough sequencing depth, we constructed an alpha rarefaction plot (Fig 8) using Shannon(Shannon and Weaver, 1949) as a metric. This indicates that further depth of sequencing (more sequence reads per sample) would not provide any additional value. Therefore, further sequencing would not serve to change our resultant data as the maximum richness (identified ASV) was reached by our sequencing effort.

The phylum level WGS and 16S data show a similar pattern between samples as the most abundant phyla are common in all samples. In addition, these phyla are at a similar relative abundance level across all samples. Despite different library preparation techniques and bioinformatic analysis, Oyster Creek remains the lowest separate species identified count in both WGS (1818) and 16S (368). Therefore, this is the least diverse of all the sites at the species level. The fact Oyster Creek shows the lowest diversity at the species level also supports our hypothesis that environmental changes, in particular temperature as indicated by alpha diversity analysis (appendix Figure 1, Table 4), are affecting the diversity at the Oyster Creek location.

It should also be noted that while MG-RAST used the M5NR database, QIIME2 used the silva-138-99-515-806 (02-Nov-2020 15:08:59) (Quast *et al.*, 2013) database. A previous study, where 16S and WGS data were compared, found that WGS was able to provide much more data in terms of taxa prediction and abundance estimation (Khachatryan *et al.*, 2020). They also found that 16S were often missing taxa and had a high level of false-positive rates. Additionally, while SILVA is a 16S specific database, the M5NR database also uses functional genomic data to identify, as well as being a pool of multiple databases (source databases from Genbank (NCBI), IMG (JGI), KEGG, PATRIC (VBI), RefSeq (NCBI), SEED, SwissProt (UniProt), TrEMBL (UniProt), eggNOG, COG,(eggNOG) GO, KO (KEGG), NOG (eggNOG) and Subsystems (SEED)). All of these could be factors in explaining why far more species were identified in all samples using the WGS method compared to the 16S method.

Comparison to previous studies

A previous study by Hicks et al. (2018), used three temperatures $6^{O}C$, $12^{O}C$, and $18^{O}C$, to analyze the effect on sediment bacterial composition. They found that specific responses to temperature were present in *Verrucomicrobia*, with a decrease in abundance, as temperature increased(Hicks *et al.*, 2018). While thermophilic bacteria from the phylum *Deinococcus-Thermus* were only found in the highest mean temperatures (Hicks *et al.*, 2018). We did not find this same pattern in our data (Figure 13).

It was also found by Hicks et al. (2018) that increasing mean temperature led to a decrease in Betaproteobacteria abundance. The same trend was found both our 16S data and to a lower extent in our WGS data (Figure 14) where Oyster Creek, which had the highest mean temperature, also had the lowest relative abundance of Betaproteobacteria. Betaproteobacteria oxidize ammonia to nitrite as the first initial step in nitrification, as estuaries are nitrogen limited this is considered a significant role in nitrogen cycling (Bernhard *et al.*, 2005). Therefore, any decrease in abundance of Betaproteobacteria could negatively affect the nitrogen cycle. A measure of Nitrogen levels at the sample sites may provide evidence that a decrease in abundance of Betaproteobacteria is affecting the nitrogen cycling.

In the same study (Hicks *et al.*, 2018) found *Planctomycetes* was also affected according to temperature changes by reduced abundance. We constructed figure 15, using WGS and 16S data, this shows the mean relative abundance of *Planctomycetes* for all sites. As shown, we also found a decrease in the relative abundance of *Planctomycetes* at the higher temperature site of Oyster Creek in the WGS data. However, in the 16S data Forked River 2 showed the lowest relative abundance. Interestingly, Traders Cove was the highest abundance of *Planctomycetes* in both analyses and is also the furthest from the power plant geographically.

It should be noted that Hicks *et al.*, (2018) used operational taxonomic units (OTUs) as opposed to ASVs which were used in this study. Also, this study only used the V4 16S region in sequencing analysis as opposed to our V3 and V4 regions. These could both be factors in the differences found between these two studies.

Cyanobacteria are generally considered to favor higher temperatures (Thomas and Litchman, 2016). However, we found in our 16S data that Cyanobacteria are lower in relative abundance compared to all other sample sites. A potential explanation may be the availability of nutrients; in particular Nitrogen, as this has been found to affect Cyanobacteria growth rates (Thomas and Litchman, 2016).

Overall, we found that there are significant differences between these sites when comparing alpha diversity to temperature, indicated by a p-value of 0.013 (appendix Table 4). In terms of beta diversity, we found that bottom type may be playing a role in community dissimilarity as the samples of sand bottom type were very closely aligned in the Bray-Curtis analysis. This was further supported by the PERMANOVA test which produced significant a p-value when testing beta diversity in relation to bottom type. At the species level, Oyster Creek showed less diversity than all other sites in both WGS and 16S analysis. The link between lack of diversity and temperature seems to support our hypothesis that the thermal loading of the Power Plant is affecting biodiversity. However, it is not changing the entire composition of the bacteria found, as Oyster Creek still shares similar biodiversity and relative abundance to all sites at the phylum level. When we looked at individual differences in the biodiversity, using past studies as a reference, our conclusions were mixed as the WGS and 16S data did not show the same patterns. Further analysis of the differences in bacterial composition, especially in relation to function, may help clarify the impact of the temperature change, driven by the power plant.

Bibliography

Acinas, Silvia G., Sarma-Rupavtarm, Ramahi, Klepac-Ceraj, Vanja, Polz, Martin F. (2005) 'PCR-induced sequence artifacts and bias: Insights from comparison of two 16s rRNA clone libraries constructed from the same sample', *Applied and Environmental Microbiology*, 71(12), pp. 8966–8969. doi: 10.1128/AEM.71.12.8966-8969.2005.

Aronesty, E. (2013) 'Comparison of Sequencing Utility Programs', *The Open Bioinformatics Journal*, 7(1), pp. 1–8. doi: 10.2174/1875036201307010001.

Bernhard, Anne E, Donn, Thomas, Giblin, Anne E, Stahl, David A (2005) 'Loss of diversity of ammonia-oxidizing bacteria correlates with increasing salinity in an estuary system', 7, pp. 1289–1297. doi: 10.1111/j.1462-2920.2005.00808.x.

Boey, Jian Sheng, Mortimer, Redmond, Couturier, Agathe, Worrallo, Katie, Handley, Kim M (2021) 'Estuarine microbial diversity and nitrogen cycling increase along sand–mud gradients independent of salinity and distance', *Environmental Microbiology*. Wiley Online Library.

Bolyen, Evan, Rideout, Jai Ram, Dillon, Matthew R., Bokulich, Nicholas A., Abnet, Christian C., Al-Ghalith, Gabriel A., Alexander, Harriet, Alm, Eric J., Arumugam, Manimozhiyan, Asnicar, Francesco, Bai, Yang, Bisanz, Jordan E., Bittinger, Kyle, Brejnrod, Asker, Brislawn, Colin J., Brown, C. Titus, Callahan, Benjamin J., Caraballo-Rodríguez, Andrés Mauricio, Chase, John, Cope, Emily K., Da Silva, Ricardo, Diener, Christian, Dorrestein, Pieter C., Douglas, Gavin M., Durall, Daniel M., Duvallet, Claire, Edwardson, Christian F., Ernst, Madeleine, Estaki, Mehrbod, Fouquier, Jennifer, Gauglitz, Julia M., Gibbons, Sean M., Gibson, Deanna L., Gonzalez, Antonio, Gorlick, Kestrel, Guo, Jiarong, Hillmann, Benjamin, Holmes, Susan, Holste, Hannes, Huttenhower, Curtis, Huttley, Gavin A., Janssen, Stefan, Jarmusch, Alan K., Jiang, Lingjing, Kaehler, Benjamin D., Kang, Kyo Bin, Keefe, Christopher R., Keim, Paul, Kelley, Scott T., Knights, Dan, Koester, Irina, Kosciolek, Tomasz, Kreps, Jorden, Langille, Morgan G. I., Lee, Joslynn, Ley, Ruth, Liu, Yong-Xin, Loftfield, Erikka, Lozupone, Catherine, Maher, Massoud, Marotz, Clarisse, Martin, Bryan D., McDonald, Daniel, McIver, Lauren J., Melnik, Alexey V., Metcalf, Jessica L., Morgan, Sydney C., Morton, Jamie T., Naimey, Ahmad Turan, Navas-Molina, Jose A., Nothias, Louis Felix, Orchanian, Stephanie B., Pearson, Talima, Peoples, Samuel L., Petras, Daniel, Preuss, Mary Lai, Pruesse, Elmar, Rasmussen, Lasse Buur, Rivers, Adam, Robeson, Michael S., Rosenthal, Patrick, Segata, Nicola, Shaffer, Michael, Shiffer, Arron, Sinha, Rashmi, Song, Se Jin, Spear, John R., Swafford, Austin D., Thompson, Luke R., Torres, Pedro J., Trinh, Pauline, Tripathi, Anupriya, Turnbaugh, Peter J., Ul-Hasan, Sabah, van der Hooft, Justin J. J., Vargas, Fernando, Vázquez-Baeza, Yoshiki, Vogtmann, Emily, von Hippel, Max, Walters, William, Wan, Yunhu, Wang, Mingxun, Warren, Jonathan, Weber, Kyle C., Williamson, Charles H. D., Willis, Amy D., Xu, Zhenjiang Zech, Zaneveld, Jesse R., Zhang, Yilong, Zhu, Qiyun, Knight, Rob, Caporaso, J. Gregory (2019) 'Author Correction: Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2', *Nature Biotechnology*, 37(9), pp. 1091–1091. doi: 10.1038/s41587-019-0252-6.

Brooks, J. Paul, Edwards, David J., Harwich, Michael D., Rivera, Maria C., Fettweis, Jennifer M., Serrano, Myrna G., Reris, Robert A., Sheth, Nihar U., Huang, Bernice, Girerd, Philippe, Strauss, Jerome F., Jefferson, Kimberly K., Buck, Gregory A. (2015) 'The truth about metagenomics: Quantifying and counteracting bias in 16S rRNA studies Ecological and evolutionary microbiology', *BMC Microbiology*, 15(1), pp. 1–15. doi: 10.1186/s12866-015-0351-6.

Bukin, Yu S., Galachyants, Yu P., Morozov, I. V., Bukin, S. V., Zakharenko, A. S., Zemskaya, T. I. (2019) 'The effect of 16s rRNA region choice on bacterial community metabarcoding results', *Scientific Data*. The Author(s), 6, pp. 1–14. doi: 10.1038/sdata.2019.7.

Callahan, Benjamin J, McMurdie, Paul J, Rosen, Michael J, Han, Andrew W, Johnson, Amy Jo A, Holmes, Susan P (2016) 'DADA2: High-resolution sample inference from Illumina amplicon data', *Nature Methods*, 13(7), pp. 581–583. doi: 10.1038/nmeth.3869.

Chapin, F. Stuart, Zavaleta, Erika S., Eviner, Valerie T., Naylor, Rosamond L., Vitousek, Peter M., Reynolds, Heather L., Hooper, David U., Lavorel, Sandra, Sala, Osvaldo E., Hobbie, Sarah E., Mack, Michelle C., Díaz, Sandra (2000) 'Consequences of changing biodiversity', *Nature*, 405(6783), pp. 234–242. doi: 10.1038/35012241.

Dowd, Scot E., Sun, Yan, Secor, Patrick R., Rhoads, Daniel D., Wolcott, Benjamin M., James, Garth A., Wolcott, Randall D. (2008) 'Survey of bacterial diversity in chronic wounds using

Pyrosequencing, DGGE, and full ribosome shotgun sequencing', *BMC Microbiology*, 8, pp. 1–15. doi: 10.1186/1471-2180-8-43.

Faith, D. P. (1992) 'Conservation evaluation and phylogenetic diversity', *Biological Conservation*. Elsevier, 61(1), pp. 1–10. doi: 10.1016/0006-3207(92)91201-3.

Fu, Limin, Niu, Beifang, Zhu, Zhengwei, Wu, Sitao, Li, Weizhong (2012) 'CD-HIT: accelerated for clustering the next-generation sequencing data', *Bioinformatics*, 28(23), pp. 3150–3152. doi: 10.1093/bioinformatics/bts565.

Gallagher, M. P. (2018) *'Permanent Cessation of Operations at Oyster Creek Nuclear Generating Station,'*.

Gilbert, Jack A., Blaser, Martin J., Caporaso, J. Gregory, Jansson, Janet K., Lynch, Susan V., Knight, Rob, Hillmann, Callahan, Benjamin J, Wong, Joan, Heiner, Cheryl, Oh, Steve, Theriot, Casey M, Gulati, Ajay S, McGill, Sarah K, Dougherty, Michael K, de Goffau, Marcus C., Lager, Susanne, Salter, Susannah J., Wagner, Josef, Kronbichler, Andreas, Charnock-Jones, D. Stephen, Peacock, Sharon J., Smith, Gordon C. S., Parkhill, Julian (2018) 'Title: Evaluating the information content of shallow shotgun metagenomics Running Title: Evaluating shallow shotgun metagenomics Benjamin Hillmann', *bioRxiv*, 24(4), pp. 851–853. doi: 10.1101/320986.

Greenwald, William W., Li, He, Benaglio, Paola, Jakubosky, David, Matsui, Hiroko, Schmitt, Anthony, Selvaraj, Siddarth, D'Antonio, Matteo, D'Antonio-Chronowska, Agnieszka, Smith, Erin N., Frazer, Kelly A. (2019) 'Subtle changes in chromatin loop contact propensity are associated with differential gene regulation and expression', *Nature Communications*, 10(1), pp. 1–17. doi: 10.1038/s41467-019-08940-5.

Hansen, Martin Christian, Tolker-Nielsen, Tim, Givskov, Michael, Molin, Søren (1998) 'Biased 16S rDNA PCR amplification caused by interference from DNA flanking the template region', *FEMS Microbiology Ecology*, 26(2), pp. 141–149. doi: 10.1016/S0168-6496(98)00031-2.

Hicks, Natalie, Liu, Xuan, Gregory, Richard, Kenny, John, Lucaci, Anita, Lenzi, Luca, Paterson, David M., Duncan, Katherine R. (2018) 'Temperature driven changes in benthic bacterial diversity influences biogeochemical cycling in coastal sediments', *Frontiers in Microbiology*. Frontiers Media S.A., 9(AUG). doi: 10.3389/fmicb.2018.01730.

Hillmann, Benjamin, Al-ghalith, Gabriel A, Shields-cutler, Robin R, Zhu, Qiyun, Gohl, Daryl M, Beckman, Kenneth B, Knight, Rob, Knights, Dan (2018) 'crossm Metagenomics', *mSystems*, 3(6), pp. 1–12.

Jiang, Hongshan, Lei, Rong, Ding, Shou Wei, Zhu, Shuifang (2014) 'Skewer: A fast and accurate adapter trimmer for next-generation sequencing paired-end reads', *BMC Bioinformatics*, 15(1), pp. 1–12. doi: 10.1186/1471-2105-15-182.

Jordan, E. O. (1894) 'THE IDENTIFICATION OF THE TYPHOID FEVER BACILLUS.', *Journal of the American Medical Association*, XXIII(25), pp. 931–935. doi: 10.1001/jama.1894.02421300005001c.

Jovel, Juan, Patterson, Jordan, Wang, Weiwei, Hotte, Naomi, O'Keefe, Sandra, Mitchel, Troy, Perry, Troy, Kao, Dina, Mason, Andrew L., Madsen, Karen L., Wong, Gane K.S. (2016) 'Characterization of the gut microbiome using 16S or shotgun metagenomics', *Frontiers in Microbiology*, 7(APR), pp. 1–17. doi: 10.3389/fmicb.2016.00459.

Katoh, Kazutaka, Misawa, Kazuharu, Kuma, Kei-ichi, Miyata, Takashi (2002) 'MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform', *Nucleic Acids Research*, 30(14), pp. 3059–3066. doi: 10.1093/nar/gkf436.

Keegan, Kevin P, Trimble, William L, Wilkening, Jared, Wilke, Andreas, Harrison, Travis, D'Souza, Mark, Meyer, Folker (2012) 'A Platform-Independent Method for Detecting Errors in Metagenomic Sequencing Data: DRISEE', *PLOS Computational Biology*. Public Library of Science, 8(6), p. e1002541. Available at: https://doi.org/10.1371/journal.pcbi.1002541.

Kent, W. J. (2002) 'BLAT---The BLAST-Like Alignment Tool', *Genome Research*, 12(4), pp. 656–664. doi: 10.1101/gr.229202.

Khachatryan, Lusine, de Leeuw, Rick H., Kraakman, Margriet E.M., Pappas, Nikos, te Raa, Marije, Mei, Hailiang, de Knijff, Peter, Laros, Jeroen F.J. (2020) 'Taxonomic classification and abundance estimation using 16S and WGS—A comparison using controlled reference samples', *Forensic Science International: Genetics*. Elsevier Ireland Ltd, 46. doi: 10.1016/j.fsigen.2020.102257.

Kopylova, E., Noé, L. and Touzet, H. (2012) 'SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data', *Bioinformatics*, 28(24), pp. 3211–3217. doi: 10.1093/bioinformatics/bts611.

Langmead, B. and Salzberg, S. L. (2012) 'Fast gapped-read alignment with Bowtie 2', *Nature Methods*, 9(4), pp. 357–359. doi: 10.1038/nmeth.1923.

Lozupone, C. and Knight, R. (2005) 'UniFrac: A new phylogenetic method for comparing microbial communities', *Applied and Environmental Microbiology*, 71(12), pp. 8228–8235. doi: 10.1128/AEM.71.12.8228-8235.2005.

Marçais, G. and Kingsford, C. (2011) 'A fast, lock-free approach for efficient parallel counting of occurrences of k-mers', *Bioinformatics*, 27(6), pp. 764–770. doi: 10.1093/bioinformatics/btr011.

Meyer, Folker, Paarmann, D., D'Souza, M., Olson, R., Glass, E. M., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R., Wilke, A., Wilkening, J., Edwards, R. A. (2008) 'The metagenomics RAST server - A public resource for the automatic phylogenetic and functional analysis of metagenomes', *BMC Bioinformatics*, 9, pp. 1–8. doi: 10.1186/1471-2105-9-386.

Ong, Swee Hoe, Kukkillaya, Vinutha Uppoor, Wilm, Andreas, Lay, Christophe, Ho, Eliza Xin Pei, Low, Louie, Hibberd, Martin Lloyd, Nagarajan, Niranjan (2013) 'Species Identification and Profiling of Complex Microbial Communities Using Shotgun Illumina Sequencing of 16S rRNA Amplicon Sequences', *PLoS ONE*, 8(4), pp. 1–8. doi: 10.1371/journal.pone.0060811.

Pedregosa, F. Varoquaux, G. Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M. Prettenhofer, P., and Weiss, R. Dubourg, V. Vanderplas, J., Passos, A., Cournapeau, D. Brucher, M., Perrot, M., Duchesnay, (2011) 'Scikit-learn: Machine Learning in Python', *Journal of Machine Learning Research*, 12, pp. 2825–2830. Available at: https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf?source=post_page----------.

Peffers, M. J., Liu, X. and Clegg, P. D. (2014) 'Transcriptomic profiling of cartilage ageing', *Genomics Data*, 2, pp. 27–28. doi: 10.1016/j.gdata.2014.03.001.

Pellens, R. and Grandcolas, P. (2016) *Phylogenetics and Conservation Biology: Drawing a Path*

into the Diversity of Life. doi: 10.1007/978-3-319-22461-9_1.

Phumudzo, Tshikhudo, Ronald, Nnzeru, Khayalethu, Ntushelo, Fhatuwani, Mudau (2013) 'Bacterial species identification getting easier', *African Journal of Biotechnology*, 12(41), pp. 5975–5982. doi: 10.5897/ajb2013.12057.

Quast, Christian, Pruesse, Elmar, Yilmaz, Pelin, Gerken, Jan, Schweer, Timmy, Yarza, Pablo, Peplies, Jörg, Glöckner, Frank Oliver (2013) 'The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools', *Nucleic Acids Research*, 41(D1), pp. 590–596. doi: 10.1093/nar/gks1219.

Ranjan, Ravi, Rani, Asha, Metwally, Ahmed, McGee, Halvor S., Perkins, David L. (2016) 'Analysis of the microbiome: Advantages of whole genome shotgun versus 16S amplicon sequencing', *Biochemical and Biophysical Research Communications*, 469(4), pp. 967–977. doi: https://doi.org/10.1016/j.bbrc.2015.12.083.

Rho, M., Tang, H. and Ye, Y. (2010) 'FragGeneScan: Predicting genes in short and error-prone reads', *Nucleic Acids Research*, 38(20), pp. 1–12. doi: 10.1093/nar/gkq747.

Shannon, C. E. and Weaver, W. (1949) 'The mathematical theory of com-munication', *Urbana: University of Illinois Press*, 96.

Thomas, M. K. and Litchman, E. (2016) 'Effects of temperature and nitrogen availability on the growth of invasive and native cyanobacteria', *Hydrobiologia*. Springer International Publishing, 763(1), pp. 357–369. doi: 10.1007/s10750-015-2390-2.

Vázquez-Baeza, Yoshiki, Pirrung, Meg, Gonzalez, Antonio, Knight, Rob (2013) 'EMPeror: A tool for visualizing high-throughput microbial community data', *GigaScience*, 2(1), pp. 2–5. doi: 10.1186/2047-217X-2-16.

Wilke, Andreas, Harrison, Travis, Wilkening, Jared, Field, Dawn, Glass, Elizabeth M., Kyrpides, Nikos, Mavrommatis, Konstantinos, Meyer, Folker (2012) 'The M5nr: A novel non-redundant database containing protein sequences and annotations from multiple sources and associated tools', *BMC Bioinformatics*, 13(1). doi: 10.1186/1471-2105-13-141.

World Nuclear Association (2020) *Cooling Power Plants*, *https://world-nuclear.org/*. Available at: https://world-nuclear.org/information-library/current-and-future-generation/cooling-power-plants.aspx#:~:text=Most nuclear power (and other,going to a cooling tower.&text=The cooling in the tower,of some of the water. (Accessed: 7 July 2021).

Appendix

Appendix List 1. *Commands used in QIIME. The script written for use with QIIME2 and the 16S fastq file data. Also required was a sample manifest showing the filepath locations of all fastq files and a metadata file containing environmental information of each sample.*

```
qiime tools import \
 --type 'SampleData[PairedEndSequencesWithQuality]' \
  --input-path /home/parkera/16S_manForkedRiver_blade  \
```

```
  --output-path /home/parkera/paired-end-demux_ForkedRiver1_only.qza  \
  --input-format PairedEndFastqManifestPhred33V2
qiime demux summarize \
  --i-data /home/parkera/paired-end-demux_ForkedRiver1_only.qza \
  --o-visualization /home/parkera/paired-end-demux_ForkedRiver1_only.qzv
qiime dada2 denoise-paired \
  --i-demultiplexed-seqs /home/parkera/paired-end-demux_ForkedRiver1_only.qza \
  --p-trim-left-f 10 \
  --p-trim-left-r 10 \
  --p-trunc-len-f 250 \
  --p-trunc-len-r 250 \
  --p-n-threads 0 \
  --o-table /home/parkera/table-dada2_ForkedRiver1_only.qza \
  --o-representative-sequences /home/parkera/rep-seqs-dada2_ForkedRiver1_only.qza \
  --o-denoising-stats /home/parkera/denoising-stats-dada2_ForkedRiver1_only.qza
qiime metadata tabulate \
  --m-input-file /home/parkera/denoising-stats-dada2_ForkedRiver1_only.qza \
  --o-visualization /home/parkera/denoising-stats-dada2_ForkedRiver1_only.qzv
qiime feature-table summarize \
  --i-table /home/parkera/table-dada2_ForkedRiver1_only.qza \
  --o-visualization /home/parkera/table-dada2_ForkedRiver1_only.qzv \
qiime feature-table tabulate-seqs \
  --i-data /home/parkera/rep-seqs-dada2_ForkedRiver1_only.qza \
  --o-visualization /home/parkera/rep-seqs-dada2_ForkedRiver1_only.qzv
qiime feature-classifier classify-sklearn \
  --i-classifier /home/parkera/classifier.qza \
  --i-reads /home/parkera/rep-seqs-dada2_ForkedRiver1_only.qza \
  --o-classification                    /home/parkera/Classification_taxonomy_rep-seqs-
dada2_ForkedRiver1_only.qza
qiime metadata tabulate \
  --m-input-file  /home/parkera/Classification_taxonomy_rep-seqs-dada2_ForkedRiver1_only.qza
\
  --o-visualization                     /home/parkera/Classification_taxonomy_rep-seqs-
dada2_ForkedRiver1_only.qva
qiime feature-table filter-samples \
  --i-table /home/parkera/table-dada2_ForkedRiver1_only.qza \
  --p-min-frequency 80000 \
  --o-filtered-table /home/parkera/table-dada2_ForkedRiver1_only_80k_Filter.qza
qiime taxa barplot \
  --i-table /home/parkera/table-dada2_ForkedRiver1_only_80k_Filter.qza \
  --i-taxonomy /home/parkera/Classification_taxonomy_rep-seqs-dada2_ForkedRiver1_only.qza \
  --m-metadata-file /home/parkera/benthic16SOnly.tsv \
  --o-visualization /home/parkera/taxa-bar-plots_ForkedRiver1_only.qzv
qiime feature-table merge \
  --i-tables /home/parkera/ForkedRiver1/table-dada2_ForkedRiver1_Only_80k_Filter.qza \
  --i-tables /home/parkera/ForkedRiver2/table-dada2_ForkedRiver2_Only_80k_Filter.qza \
  --i-tables /home/parkera/ForkedRiver3/table-dada2_ForkedRiver3_Only_80k_Filter.qza \
  --i-tables /home/parkera/OysterCreek1/table-dada2_OysterCreek1_Only_80k_Filter.qza \
  --i-tables /home/parkera/OysterCreek2/table-dada2_OysterCreek2_Only_80k_Filter.qza \
```

```
  --i-tables /home/parkera/OysterCreek3/table-dada2_OysterCreek3_Only_80k_Filter.qza \
  --i-tables /home/parkera/SunriseBeach1/table-dada2_SunriseBeach1_Only_80k_Filter.qza \
  --i-tables /home/parkera/SunriseBeach2/table-dada2_SunriseBeach2_Only_80k_Filter.qza \
  --i-tables /home/parkera/SunriseBeach3/table-dada2_SunriseBeach3_Only_80k_Filter.qza \
  --i-tables /home/parkera/TradersCove1/table-dada2_TradersCove1_Only_80k_Filter.qza \
  --i-tables /home/parkera/TradersCove2/table-dada2_TradersCove2_Only_80k_Filter.qza \
  --i-tables /home/parkera/TradersCove3/table-dada2_TradersCove3_Only_80k_Filter.qza \
  --o-merged-table /home/parkera/table-dada2_all_80k_Filter.qza
qiime feature-table merge-seqs \
  --i-data /home/parkera/ForkedRiver1/rep-seqs-dada2_ForkedRiver1_Only.qza \
  --i-data /home/parkera/ForkedRiver2/rep-seqs-dada2_ForkedRiver2_Only.qza \
  --i-data /home/parkera/ForkedRiver3/rep-seqs-dada2_ForkedRiver3_Only.qza \
  --i-data /home/parkera/OysterCreek1/rep-seqs-dada2_OysterCreek1_Only.qza \
  --i-data /home/parkera/OysterCreek2/rep-seqs-dada2_OysterCreek2_Only.qza \
  --i-data /home/parkera/OysterCreek3/rep-seqs-dada2_OysterCreek3_Only.qza \
  --i-data /home/parkera/SunriseBeach1/rep-seqs-dada2_SunriseBeach1_Only.qza \
  --i-data /home/parkera/SunriseBeach2/rep-seqs-dada2_SunriseBeach2_Only.qza \
  --i-data /home/parkera/SunriseBeach3/rep-seqs-dada2_SunriseBeach3_Only.qza \
  --i-data /home/parkera/TradersCove1/rep-seqs-dada2_TradersCove1_Only.qza \
  --i-data /home/parkera/TradersCove2/rep-seqs-dada2_TradersCove2_Only.qza \
  --i-data /home/parkera/TradersCove3/rep-seqs-dada2_TradersCove3_Only.qza \
  --o-merged-data /home/parkera/rep-seqs-dada2_all.qza
qiime phylogeny align-to-tree-mafft-fasttree \
  --i-sequences /home/parkera/rep-seqs-dada2_all.qza \
  --o-alignment /home/parkera/aligned-rep-seqs_all.qza \
  --o-masked-alignment /home/parkera/masked-aligned-rep-seqs_all.qza \
  --o-tree /home/parkera/unrooted-tree_all.qza \
  --o-rooted-tree /home/parkera/rooted-tree_all.qza
qiime diversity core-metrics-phylogenetic \
  --i-phylogeny /home/parkera/rooted-tree_all.qza \
  --i-table /home/parkera/table-dada2_all_80k_Filter.qza \
  --p-sampling-depth 80000 \
  --m-metadata-file /home/parkera/16S_sample-metadata.tsv \
  --output-dir /home/parkera/core-metrics-results_all
qiime diversity alpha-group-significance \
  --i-alpha-diversity /home/parkera/core-metrics-results_all/faith_pd_vector.qza \
  --m-metadata-file /home/parkera/16S_sample-metadata.tsv \
  --o-visualization /home/parkera/core-metrics-results_all/faith-pd-group-significance.qzv
qiime diversity alpha-group-significance \
  --i-alpha-diversity /home/parkera/core-metrics-results_all/evenness_vector.qza \
  --m-metadata-file /home/parkera/16S_sample-metadata.tsv \
  --o-visualization /home/parkera/core-metrics-results_all/evenness-group-significanceall.qzv
qiime diversity beta-group-significance \
  --i-distance-matrix                                    /home/parkera/core-metrics-
results_all/unweighted_unifrac_distance_matrix.qza \
  --m-metadata-file /home/parkera/16S_sample-metadata.tsv \
  --m-metadata-column location_name \
  --o-visualization/home/parkera/core-metrics-results_all/unweighted-unifrac-location-name-
significance.qzv \
  --p-pairwise
qiime diversity alpha-rarefaction \
```

```
--i-table /home/parkera/table-dada2_all_80k_Filter.qza \
--i-phylogeny /home/parkera/rooted-tree_all.qza \
--p-max-depth 4000 \
--m-metadata-file /home/parkera/16S_sample-metadata.tsv \
--o-visualization /home/parkera/core-metrics-results_all/alpha-rarefaction.qzv
```

***Appendix Table 1**. the metadata file information of each sample containing environmental variables associated with each site.*

sample-id	location name	Power Plant Closing	Bottom type	Dissolved oxygen (mg/L)	Dissolved oxygen (%)	Dissolved oxygen categorical percent	Water temperature C	Water temperature categorical C	Salinity (ppt)	Salinity categorical (ppt)	depth (m)	time of day	month	day	year
#q2:types	categorical	categorical	categorical	numeric	numeric	categorical	numeric	categorical	numeric	categorical	numeric	categorical	categorical	categorical	categorical
ForkedRiver1	ForkedRiver	Before	Sand	13.01	103.5	<110	5.3	<10	25.5	>20	1.5	afternoon	3	18	2018
ForkedRiver2	ForkedRiver	Before	Sand	13.01	103.5	<110	5.3	<10	25.5	>20	1.5	afternoon	3	18	2018
ForkedRiver3	ForkedRiver	Before	Sand	13.01	103.5	<110	5.3	<10	25.5	>20	1.5	afternoon	3	18	2018
OysterCreek1	OysterCreek	Before	Mud	13.12	119.5	>110	10	>10	14.2	<15	0.9	afternoon	3	14	2018
OysterCreek2	OysterCreek	Before	Mud	13.12	119.5	>110	10	>10	14.2	<15	0.9	afternoon	3	14	2018
OysterCreek3	OysterCreek	Before	Mud	13.12	119.5	>110	10	>10	14.2	<15	0.9	afternoon	3	14	2018
SunriseBeach1	SunriseBeach	Before	Sand	12.06	106.6	<110	5.1	<10	16.3	15-20	1.1	afternoon	3	14	2018
SunriseBeach2	SunriseBeach	Before	Sand	12.06	106.6	<110	5.1	<10	16.3	15-20	1.1	afternoon	3	14	2018
SunriseBeach3	SunriseBeach	Before	Sand	12.06	106.6	<110	5.1	<10	16.3	15-20	1.1	afternoon	3	14	2018
TradersCove1	TradersCove	Before	Mud	13.44	107.1	<110	5.6	<10	21.9	>20	2	late-morning	3	18	2018
TradersCove2	TradersCove	Before	Mud	13.44	107.1	<110	5.6	<10	21.9	>20	2	late-morning	3	18	2018
TradersCove3	TradersCove	Before	Mud	13.44	107.1	<110	5.6	<10	21.9	>20	2	late-morning	3	18	2018

***Appendix Table 2**. The DNA extraction readings, the index used for each sample and the output (yield and % reads identified past filter (PF)) statistics for both 16S and WGS sequencing.*

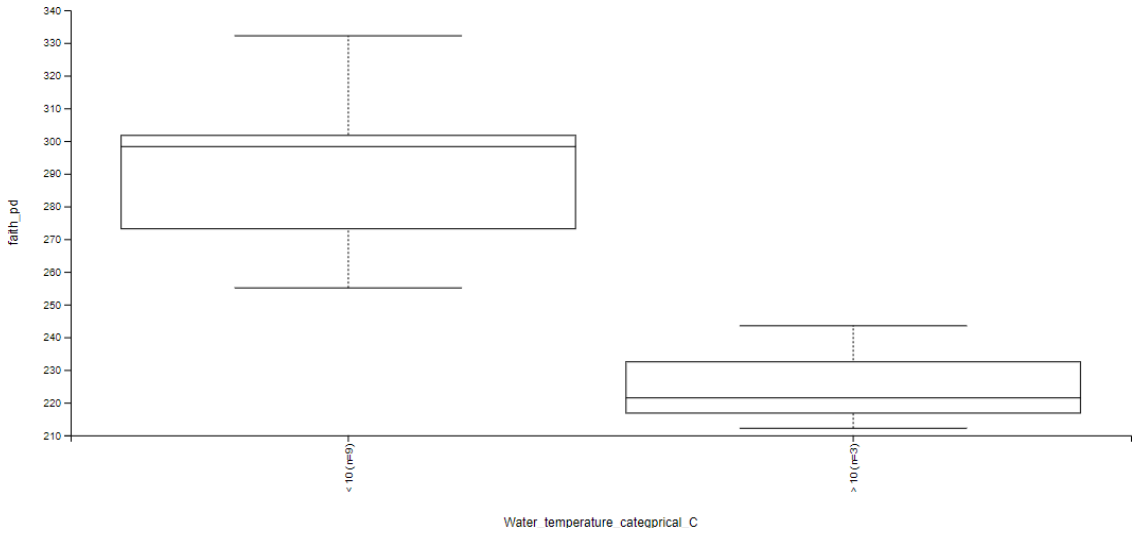
	Extracted DNA values		Shotgun data				16s data				
Sample	ng/uL	260/280	INDEX 1 (I7)	INDEX 2 (I5)	% READS IDENTIFIED (PF)	Yield	LIBRARY NAME	INDEX 1 (I7)	INDEX 2 (I5)	% READS IDENTIFIED (PF)	Yield
Water	-0.19	0.28									
FR1	4.15	2.31	CGAGGCTG	TATCCTCT	7.9262	1.00 Gbp	ForkedRiver1	TAAGGCGA	CTCTCTAT	5.7106	297.57Mbp
FR2	3.95	1.82	GGACTCCT	TATCCTCT	8.3032	1.05 Gbp	ForkedRiver2	CGTACTAG	CTCTCTAT	4.6648	243.08Mbp
FR3	4.6	1.87	TAGGCATG	TATCCTCT	7.2654	920.56 Mbp	ForkedRiver3	AGGCAGAA	CTCTCTAT	2.812	146.53Mbp
OC1	8.84	1.73	CGAGGCTG	GTAAGGAG	3.425	433.97 Mbp	OysterCreek1	TAAGGCGA	AGAGTAGA	2.8777	149.95Mbp
OC2	9.29	2.06					OysterCreek2	CGTACTAG	AGAGTAGA	4.0657	211.86Mbp
OC3	8.97	1.83	GCTCATGA	GCGTAAGA	3.2189	407.85 Mbp	OysterCreek3	AGGCAGAA	AGAGTAGA	3.289	171.39Mbp
SB1	4.85	1.85	CTCTCTAC	GTAAGGAG	12.6742	1.61 Gbp	SunriseBeach1	TAAGGCGA	GCGTAAGA	8.0518	419.57Mbp
SB2	5.14	1.74	AAGAGGCA	GTAAGGAG	10.3186	1.31 Gbp	SunriseBeach2	CGTACTAG	GCGTAAGA	7.7212	402.34Mbp
SB3	3.93	3.05	GCTCATGA	GTAAGGAG	6.5929	835.34 Mbp	SunriseBeach3	AGGCAGAA	GCGTAAGA	8.2986	432.43Mbp
TC1	7.8	1.64	CGAGGCTG	ACTGCATA	5.264	666.97 Mbp	TradersCove1	TCCTGAGC	CTCTCTAT	5.786	301.5 Mbp
TC2	8.08	1.77	GGACTCCT	ACTGCATA	5.7549	729.17 Mbp	TradersCove2	TCCTGAGC	TATCCTCT	11.9399	622.17Mbp
TC3	9.96	1.69	TAGGCATG	ACTGCATA	4.4152	559.42 Mbp	TradersCover3	TCCTGAGC	AGAGTAGA	6.6897	348.59Mbp
POS			CTCTCTAC	ACTGCATA	10.6754	1.35 Gbp					

***Appendix Table 3**. The alpha diversity analysis of location using Faith's phylogenetic diversity.*

Kruskal-Wallis (all groups)				
	Result			
H	8.076923077			
p-value	0.044448403			
Kruskal-Wallis (pairwise)				
		H	p-value	q-value
Group 1	Group 2			

ForkedRiver (n=3)	OysterCreek (n=3)	3.857143	0.049535	0.074302
	SunriseBeach (n=3)	3.857143	0.049535	0.074302
	TradersCove (n=3)	0.428571	0.512691	0.512691
OysterCreek (n=3)	SunriseBeach (n=3)	3.857143	0.049535	0.074302
	TradersCove (n=3)	3.857143	0.049535	0.074302
SunriseBeach (n=3)	TradersCove (n=3)	0.428571	0.512691	0.512691

.



Appendix Figure 1. *Analysis of 16S data, boxplot of temperature as a category (<10 or >10) using Faith's Phylogenetic Diversity, Alpha diversity between sample sites. This tests associations between metadata columns (sample site) and alpha diversity. Constructed using QIIME2 view*

Appendix Table 4. *Analysis of 16S data, statistical analysis of temperature as a category (<10 or >10) using Faith's Phylogenetic Diversity, Alpha diversity between sample sites. Constructed in Excel® using data output from QIIME2.*

Kruskal-Wallis (all groups)				
	Result			
H	6.230769231			
p-value	0.012554919			
Kruskal-Wallis (pairwise)				
		H	p-value	q-value
Group 1	Group 2			
< 10 (n=9)	> 10 (n=3)	6.230769	0.012555	0.012555