

College of Saint Benedict and Saint John's University

DigitalCommons@CSB/SJU

Honors Theses, 1963-2015

Honors Program

2015

Comparison of Bacterial Flora in Fish Mucus and Lake Water

Robert W. Thomas

College of Saint Benedict/Saint John's University

Follow this and additional works at: https://digitalcommons.csbsju.edu/honors_theses



Part of the [Biology Commons](#)

Recommended Citation

Thomas, Robert W., "Comparison of Bacterial Flora in Fish Mucus and Lake Water" (2015). *Honors Theses, 1963-2015*. 85.

https://digitalcommons.csbsju.edu/honors_theses/85

This Thesis is brought to you for free and open access by DigitalCommons@CSB/SJU. It has been accepted for inclusion in Honors Theses, 1963-2015 by an authorized administrator of DigitalCommons@CSB/SJU. For more information, please contact digitalcommons@csbsju.edu.

Comparison of Bacterial Flora in Fish Mucus and Lake Water

Robert William Thomas
Under the Guidance of Barbara May
Summer 2014

Abstract:

Although commonly known for their ability to cause deadly infectious diseases, there are populations of bacteria (identified as normal flora) that symbiotically live on and amongst larger host organisms and positively impact survival of the host. In previous research, culturing methods suggest that freshwater fish (i.e. bluegill and northern pike) maintain a normal flora population and that this flora aids in protection and the health of the fish. Our study used Ion torrent PGM sequencing to identify a larger, more complete population of normal flora in 2 different lake samples from three different species of fish. Identifying and comparing the normal flora populations in the lakes and different fish species would provide us with a better understanding of a fish's interactions with normal bacterial flora. Data presented here suggests that fish mucus specializes its flora as compared to the aquatic environment in which they live. The diversity of the aquatic samples is significantly higher as compared to all fish samples. Fish tend to harbor particular phyla including types of *Proteobacteria*. Upon comparison of mucus flora between crappie and bluegill (regardless of lake), data suggests significant differences in flora. Some of this difference may be accounted for in the aquatic environment from which they were sampled. However, even within a single aquatic environment, there are some significant differences between bacterial floras. This suggests that certain fish may specialize their flora even while in the same environment.

Introduction:

Microorganisms are important players in an ecosystem

Scientific research has shown that microorganisms, namely bacteria, have numerous vital roles that impact the ecosystem. They serve as vital players in ecological food webs.

They participate in nutrient cycles that allow humans and other animals to thrive on a once inhospitable planet. Finally, they form tight relationships that help maintain order in an ecosystem.

Microorganisms are vital to the food web

The global oceans are thought to contain $\sim 2 \times 10^6$ species of microorganisms (Curtis et al. 2002). These microorganisms perform certain biogeochemical processes vital to life.

For example, phytoplankton, composed of bacteria, protists and other microorganisms, perform the majority of primary production in the Ocean and nearly half of the net

primary production on Earth (Field et al. 1998). Primary producers are the foundation for the food web, providing organic carbon vital for survival of the entirety of the food web. In addition, viroplankton (viruses of the oceans), heterotrophic prokaryotes and protists, together, contribute to organic matter as a diet for others (Azam et al. 1983; Pernthaler 2005; Pomeroy et al. 2007). Hence, microorganisms are the foundation for complex organisms to thrive.

Microorganisms were the first to inhabit the planet

The planet Earth was formed approximately 4.5 billion years ago. The Earth remained barren and lifeless for another 800 million years. Approximately 3.7 billion years ago, the first signs of life in the form of prokaryotes were formed. It is thought that these prokaryotes, namely the phototrophs *Cyanobacteria*, were the organisms responsible for altering the planet to the habitable planet of today (Lal 2008). Not only did *Cyanobacteria* produce vast amounts of organic material, but also high levels of atmospheric oxygen, necessary for life as described today (Biello 2009). Since their arrival, the evolution, and delineation of the three major domains of living organisms appeared: archaea, bacteria, and eukaryotes.

Microorganisms are vital for nutrient cycles

Bacteria provide the fertile soil necessary for farming, agriculture, and landscapes across the planet. Bacteria perform nitrogen fixation in the soil which converts unusable nitrogen gas to nitrogen sources vital for plant and animal nutrition. More than 90 percent of all nitrogen fixation is conducted by these bacteria (Encyclopedia Britannica, 2015). This soil, rich in nitrogen, allows for crops, forests, and other landscapes to flourish.

Microorganisms interact with other living organisms

Bacteria and other microorganisms can establish one of three different relationships with other living organisms in their environment. Bacteria can form a mutualistic relationship in which both groups (the microorganism and its host) provide beneficial services for one another. This is demonstrated in the relationship between humans and its gut bacteria. Humans cannot fully digest all of the contents that enter the gastrointestinal tract. Bacteria in the gut, however, can help digest what the host cannot. Humans benefit from the bacteria aiding in digestion, and the bacteria benefits from nutrients provided by the human (Sears 2005).

Bacteria can also maintain a commensal relationship with other organisms. In this relationship, either the bacteria or the host benefit without harming nor benefitting the other. For example, *Mycobacterium vaccae* is able to draw on propane as its energy resource and as a result, co-metabolize cyclohexane to its corresponding alcohol, cyclohexanol. Subsequently, certain *Pseudomonas* species are able to use cyclohexanol and act as the beneficiaries of this commensal association (Hogan 2012). *M. vaccae* does not benefit from its relationship with *Pseudomonas* but is not harmed.

Bacteria also have parasitic relationships with other organisms. In this type of relationship, the bacterium or microorganism benefits while the host is harmed. For example, *Clostridium tetani* is a bacterium that typically enters the body through a wound or breach in the skin. Once inside, it proliferates and begins to release a toxin. This toxin interferes with the release of neurotransmitters at neuromuscular junctions. This leads to uncontrollable muscle spasms and may induce seizure (Brüggemann et al. 2003).

Bacteria are best known for their pathogenic capabilities and are a cause of concern for the health not only of humans but animals on the planet.

These three different relationships are important and help scientists to better understand the natural ecosystems and their evolution. Bacteria engage in all of these relationships and constitute much of the normal flora of many living organisms.

Understanding the basis for these interactions helps one understand common relationships as well as what might be altered in these relationships should the flora change.

Bacteria are part of the normal flora of living organisms:

Humans have many different bacterial populations living on and within the body. These bacteria perform one of the three types of relationships mentioned above and constitute what is referred to as normal flora. Some of the most well-known bacteria found inside and on humans include: *Staphylococcus epidermidis*, *Micrococcus luteus*, and

Haemophilus parainfluenzae (Davis 1996). The human bacterial flora has been heavily studied and researched. The identification and roles associated with these bacteria continues to be studied within the Human Microbiome Project (HMP)

<http://hmpdacc.org/>. This project, initiated in 2008, seeks to generate resources that enable the comprehensive characterization of the human microbiome and analysis of its role in human health and disease. The goals of the project include:

1. Develop a reference set of 3,000 isolate microbial genome sequences.
2. Initiate 16S & mWGS (metagenomic whole genome shotgun sequencing) studies to generate an estimate of the complexity of the microbial community at each body site, in hopes to determine whether there is a "core" microbiome at each site.
3. Demonstrate projects to determine the relationship between disease and changes in the human microbiome.

4. Develop new tools and technologies for computational analysis, establishment of a data analysis and coordinating center (DACC), and resource repositories.
5. Examine of the ethical, legal and social implications (ELSI) to be considered in the study and application of the metagenomic analysis of the human microbiota.

As a result of this project, certain relationships between humans and bacteria have been elucidated. For example, one determined beneficial relationship between the two can be seen in the colon of a human. Bacteria like *Bifidobacteria* produce a range of antimicrobial agents that protect the body from unwanted gram-positive and gram-negative bacteria. These *Bifidobacteria* also provide protection by occupying reception sites and consuming nutrients that possible pathogens might use (Rastall 2004). A determined harmful relationship between humans and bacteria is seen in the disease, Tuberculosis. Tuberculosis is caused by the bacterium, *Mycobacterium tuberculosis*. Although the bacteria can live latent in the host, a parasitic relationship can develop between the bacteria and the human. The bacterium causes macrophages to engulf it in order to rid the body of the bacterium. However, the bacterium feeds off of the nutrients in the macrophage. The macrophage then dies and releases the proliferating bacteria into the respiratory airways and blood stream. This can cause coughing, fever, and chest pain in the human (Click et al. 2012). These relationships and the mechanisms behind them are important to understand for health related reasons, but more importantly to understand the linkages of the microscopic world to the macroscopic world.

The role of bacterial flora in a freshwater aquatic ecosystem:

Not only are microorganisms shown to be important for humans, as elucidated in the human microbiome project, but microorganisms play vital roles in the aquatic

ecosystems. Aquatic ecosystems include oceans, lakes, rivers, streams, estuaries, and wetlands. Each one of these aquatic ecosystems contains a fragile network built around relationships between living organisms. Of particular interest, freshwater fish are known to harbor bacterial populations in numerous parts of their body including their gills, gut, and mouth (Hashizume 2005). This includes the mucosal layer on the surface of the fish scales, secreted by fish. It is believed that these bacterial populations benefit the fish by providing protection from the fish's aquatic environment (Maria 2012). Studies using culture isolation of bacteria have supported this belief. One such study shed light on the relationships between bacteria and fish in Lake Biwa, the largest freshwater lake in Japan. In this study, *Lepomis macrochirus* (common name, the bluegill) were collected and a rubber spatula was used to scrape off superficial mucus for bacterial analysis. Water from the lake was also collected to compare to the mucus flora. After having isolated and cultured the bacteria, and conducting 16S sequencing, observations suggested differences between fish mucus flora and lake water flora. Of particular interest, nearly all of the isolates identified from the mucus could metabolize glucose and only half of the isolates identified from the lake water could do the same (Hashizume 2005). Although the relevance of this physiological difference between flora on the fish and lake has not been clarified, the data suggests this relationship could be a subject for further study.

Although some additional studies similar to the one conducted on Lake Biwa have been conducted to examine bacterial populations on and within fish, the fish microbiome is relatively poorly understood. As an example, previous studies like the one described above have utilized culturing techniques to identify bacterial populations in several

populations. Although culturing provides a glimpse of some of the bacterial flora, it is not entirely thorough. Certain bacterial strains cannot be easily cultured on agarose plates. Depending on the nutrient requirements for each microorganism, many are incapable of replicating on the agar of choice. This thesis attempts to obtain a more thorough and comprehensive understanding. This experiment used a form of next generation sequencing technology to identify the 16s rRNA region of all bacteria in the population. The fish mucus was chosen because it is in direct contact with the lake environment (like human skin with the outside environment). It is particularly interesting to study because skin diseases are relatively more common in fish than in terrestrial vertebrates and are one of the primary disease conditions presented to the aquatic animal practitioner (Groff 2001). This thesis aimed to initiate a thorough collection and analysis of bacterial normal flora on freshwater fish mucus to better understand these relationships and their role in the ecosystem.

Traditional analysis of bacterial populations: isolation and culture

Identification of a bacterial species using culture techniques involves a set of traditional processes. First, a bacterium is isolated and grown on nutrient agar plates. It is important that these agar plates contain the appropriate nutrients for the growth of the microorganisms isolated. Should nutrient conditions be appropriate for energy and nutrient requirements, the bacteria proliferate. As cell numbers exponentially increase, an observable colony forms that is visible to the naked eye. A sample of the colony is then removed. Identification of the cells forming this colony can occur with several techniques: macroscopic, microscopic, physiological, and/or molecular characterization. Macroscopic study of bacteria involves the technique as described above. Colony

structure is often unique to particular microorganisms as it forms on the plate. This can include variations in color, shape, margin structure, and texture of the colony. Often this characterization is not enough for full identification of the species. Further microscopic investigation of bacteria involves the visualization of the bacterium microscopically. A wet mount or varieties of staining techniques are used to determine certain features of bacterium. Similar to macroscopic characterization including the addition of microscopic evaluation, often times, bacterial species can still not yet be identified. The addition of physiological investigation of bacteria aims to identify and categorize microbes at the species level by observing the types of biochemical processes performed in order to sustain life. In the laboratory, this involves culture of the isolated bacterium with a variety of different media that enable elucidation of physiological capabilities. Upon analysis of macroscopic, microscopic, and physiological characteristics, many bacterial species can be identified. However, the most current and straightforward mechanism of species identification is through molecular characterization and identification of the 16S nucleotide sequence. The sample is used in congruence with DNA extraction, PCR amplification of the 16S rRNA region on the bacterial genome, gel electrophoresis for PCR verification, and sequencing of this sample. Upon collecting the sequence, this is compared to a database of known sequences to compare and identify the particular bacterial species as described below.

Advantages of 16S sequence analysis

All cells require proteins for cellular function and ribosomes are vital for their production during the process of translation. The ribosome is composed of both ribosomal proteins and ribosomal RNA. In bacteria, there are three types of ribosomal RNA: 23S, 16S, and

5S. The 5S rRNA acts as the physical transducer of information in the ribosome. It communicates between different functional centers of the ribosome and helps to coordinate ribosomal events (Dinman 2005) The 23S rRNA contains the peptidyl transferase component and acts as the ribozyme (Bocchetta 1998). The 16S rRNA is responsible for recognizing and binding to the Shine-Dalgarno sequence in front of each bacterial operon to induce translation. This sequence helps recruit the ribosome to the mRNA to initiate protein synthesis by aligning the ribosome with the start codon (Czernilofsky et al. 1975). Due to the ribosome's conserved nature in bacteria (they all require protein synthesis) and its slow rate of evolution, the 16S DNA sequence that codes for this rRNA has been adopted for identification of different bacterial species. This region has been the housekeeping genetic marker used by scientists to identify and perform phylogenies. There are several reasons for this:

1. the 16S DNA that encodes for 16S rRNA is present in the genome of all bacteria. Ribosomes cannot translate mRNA or make proteins without it.
2. the 16s rRNA base pair length of 1500 is long enough to perform informatics in a laboratory setting (Janda 2007).
3. the function of the 16S rRNA gene has not changed over time, suggesting that random sequence changes are a more accurate measure of time.

This is helpful in determining evolutionary changes and for characterization of different species of bacteria. Importantly, mutations and increased diversity at the 16S rRNA region allows for clarification between differing species of bacteria. In this experiment, a 200 base pair region of the 1500 nucleotide 16S DNA sequence is variable enough so as to identify different bacterial species (Clarridge 2004). Bacteria with a 97% homology at this 200 base pair region are accepted to be within the same bacterial species. As

researchers have continued to identify different bacteria and their associated 16S DNA sequence, this data has been collected in various bacterial databases such as NCBI (National Center for Biotechnology Information) <http://www.ncbi.nlm.nih.gov/>, green genes <http://greengenes.lbl.gov/cgi-bin/nph-index.cgi>, and RDP (ribosomal database project) <http://rdp.cme.msu.edu/>. Then, after a researcher has amplified and sequenced the desired segment, they can use a program like BLAST (Basic Local Alignment Search Tool) which uses heuristic algorithms to compare a new 16S sequence to a characterized set of 16S sequences contained in one of these databases. BLAST works by detecting local alignments between sequences that coincide. BLAST analysis begins with a three letter “query word.” These letters will represent three amino acids or nucleotides, in a certain order (for example, the nucleotides ATC in that order). The BLAST search subsequently looks for the number of times in which this query word appears and places it along the sequence. The algorithm also looks for closely related query words in which one letter is different. Then, each query is scored to determine which database is “in the neighborhood” of the query sample. If the sequence has a close enough identity, it is identified as that particular bacteria. Sequences with identity scores greater than 97% are resolved at the species level, between 95% and 97% at the genus level, between 90% and 95% at the family level, between 85% and 90% at the order level, between 80% and 85% at the class level, and below this to the phylum level (77% and 80%).

Each database uses similar BLAST heuristic algorithms as the one described above with subtle differences. Due to these subtle differences in algorithms, there is no universal technique for identifying bacterial species via 16s rRNA sequencing. Authors vary widely in their use of acceptable criteria for identifying a species match (Janda

2007). Thus, different databases are used based on the type of information that is required for the experiment. Researchers and publishers accept any of these three databases.

However, instead of using traditional Sanger sequencing approaches to culture, amplify and identify each bacterial species individually, a more modern technique, next generation sequencing, has flooded the market as a vital tool for bacterial population studies.

Current methodologies for identification of bacterial populations:

Next generation sequencing uses the basic method of 16S sequencing, except on a much larger scale. A sample (i.e. a fecal sample, mucus sample, water or soil sample) is collected and its DNA is extracted without culture. The advantages to extracting, amplifying, and identifying the DNA by means of high throughput sequencing are numerous. First, bacteria require specific nutrients in order to proliferate. The agars used for culturing cannot supply some microbes with their required nutrients. Thus, high throughput sequencing has the means to identify these particular microorganisms not commonly replicated under laboratory conditions. Second, it requires far less time than traditional phenotypic culture-based identification (Jordan 2009). Pyrosequencing determines sequence reads by using a “sequence by synthesis” approach in which individual bases are added one at a time and are incorporated by DNA Polymerase. The bases that are added evoke a certain wavelength of light that tells the computer the identity of the base in the sequence. It is much faster because time is not needed for the bacteria to grow on agar plates. The classification of bacteria is also faster due to the use of optics and machinery.

A current method that uses the chemistry of “sequence by synthesis” is Ion Torrent PGM technology. Instead of using wavelength technology, this type of next generation sequencing uses a change in pH as an indicator as nucleotides are added during the amplification process. As an example and described in Figure 1, each piece of extracted DNA is first separated into its own “well” where the 16S sequence will be amplified. During the amplification process, a specific nucleotide (deoxyribonucleotide triphosphate; dNTP) is added to the well containing the strand of DNA. If the introduced dNTP is complementary to the leading template nucleotide (in this case, a 16S sequence), it is incorporated into the growing complementary strand. One dNTP is added at a time. The changes in pH that occur when the nucleotide is added, allow researchers to determine if that base, and how many thereof, was added to the sequence read. The dNTPs are then washed away, and the process is repeated cycling through different dNTPs. This produces the complementary sequence that is then used for downstream analysis. (www.mrdnalab.com, Shallowater, TX, USA)

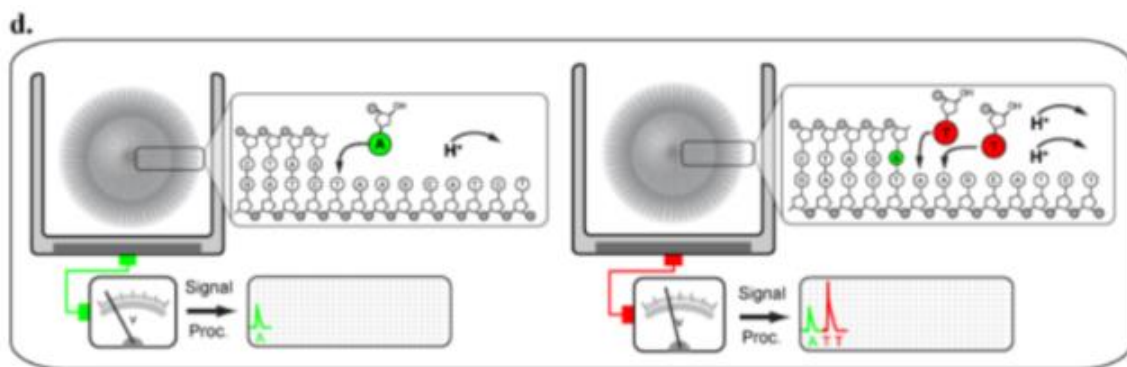


Figure 1. Representation of Ion Torrent PGM Sequencer. This figure shows the detection of pH changes as correct dNTPs are added to template strands.

Thesis plan: Determination and Comparison of Bacterial Populations in Fish Mucus and their Environments

Due to the new technologies available and the relatively unexplored questions regarding the normal flora of fish, this thesis has initiated the scientific exploration of the bacterial population in fish mucus. This research is useful because it allows for a more thorough and comprehensive understanding of the aquatic relationships between fish and their environment. From this research and collection of data, many research questions will be identified for current and future studies to better understand the ecological role and the relationship between bacteria and fish. This research provides the scientific community with the bacterial populations that inhabit fish mucus and will let researchers further predict their function and role in this environment. This thesis aims to address whether fish have similar bacterial populations as their aquatic environment. In addition, the answers to the following questions could be ascertained:

1. Does the bacterial flora differ on the mucosal layer between same species of fish within the same lake?
2. Does the bacterial flora differ between the lake environment and the mucosal layer of the fish?
3. Does the bacterial flora differ between fish of the same species within separate lakes?
4. Once differences and similarities are identified, why are they similar or different?

To address these questions, several different species of fish were caught and sampled between May and June of 2014. Bluegills (*Lepomis macrochirus*), black crappie (*Pomoxis nigromaculatus*), largemouth bass (*Micropterus salmoides*), and lake water were sampled from East Gemini Lake on the St. John's University campus in Collegeville, MN. Northern pike (*Esox Lucius*), bluegills (*Lepomis macrochirus*), black crappie (*Pomoxis nigromaculatus*), largemouth bass (*Micropterus salmoides*), Yellow

Perch (*Perca flavescens*), walleye (*Sander vitreus*), and lake water were sampled from Fireside Lake in Island Lake, WI. These are lakes of similar latitude and trophic level, but not in the same waterway. This allowed for examination of bacterial populations in similar fish, in a similar aquatic ecosystem but with distinct settings. 16S bacterial population data was collected via ion torrent technology from both bodies of water and from the fish that inhabit them. Over 700,000 sequences were collected in attempts to better understand the relationship between bacterial populations and fish.

Materials and Methods:

Sampling and collection of bacterial populations from fish mucous:

Sterile swabs were prepared in 15ml tubes, with 2ml of sterile water added to soak the swab. Fish were caught using hook and line methods. When fish were caught via traditional pole and line methods, they were swabbed from gill to tail on both sides of the fish as seen in figure 2. The fish was not touched previous to swabbing (gloves were used if touching the fish anywhere on its body was necessary). The gill and anus were avoided and not swabbed to select only bacteria along the fish scales. During collection, the swab was rotated to pick up maximum amount of mucus.

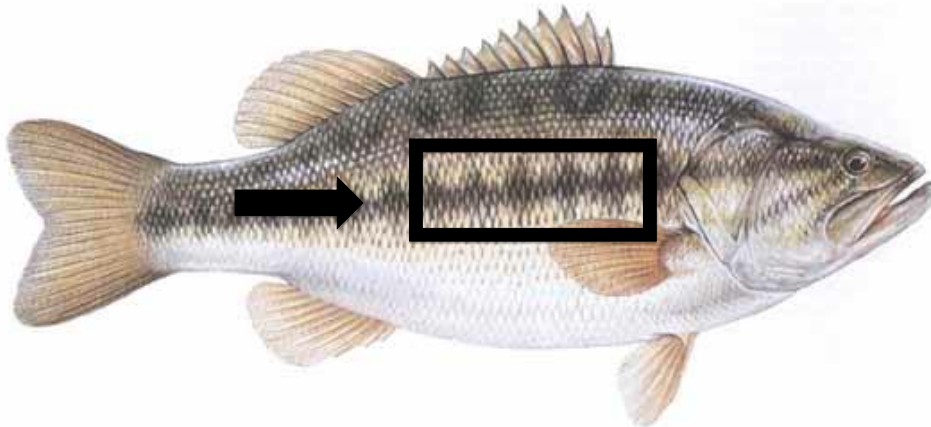


Figure 2. Representation of swab area.

This figure shows the exact spot each fish was swabbed in order to obtain a sample. The swab was moved from gill to tail along the lateral line whilst it was being rotated.

In addition to sampling fish, lake water samples were collected to compare the bacterial roster of the lake water for to that of the fish. Information was logged for each collected fish sample: date of sample collected, secchi disk value at the appropriate date, length (in) of the fish, picture, and general observations regarding sampling. Samples were labeled in the following format: each fish was given a 6-character log name FSH-LK-n (FSH =

Species of fish, LK = lake sample was taken from, n = number of fish of that species from that lake). For example, LMB-FS-2 would identify a largemouth bass (LMB) sampled from Fireside Lake (FS) and identified as sample two (2) of that particular species of fish taken from the lake. Lake water was logged in a similar fashion is described for the fish. Samples were stored at 4°C until returned to the lab. Upon returning to the lab, the 15ml tube containing the swab was vortexed and 1 ml was removed from the 15 ml tube and placed into a 1.5 ml eppendorf tube. The tube containing the sample was centrifuged to pellet the DNA. If the DNA was not immediately extracted from the samples/lakewater, the pellet was frozen for later extraction.

DNA Extraction:

DNA was extracted using the ZR Fungal/Bacterial DNA MicroPrep™ Kit, protocol, and materials provided by the ZYMO Research Company

(<http://www.zymoresearch.com/dna/genomic-dna/bacterial-fungal-dna/zr-fungal-bacterial-dna-kits/zr-fungal-bacterial-dna-miniprep>). Samples were eluted with sterile water. DNA samples were stored at -10°C frozen prior to normalization in preparation for shipping.

DNA Normalization for sequencing:

Equal concentrations were prepared for shipping and further downstream analysis using the NanoDrop 2000c UV-Vis Spectrophotometer. Concentrations of 10 µg/ml were collected with a 260/280 ratio of approximately 1.8 were recorded and prepared for ion torrent sequencing.

Ion Torrent Sequencing of bacterial populations:

Once normalized, DNA samples were processed by MR DNA Lab (www.mrdnalab.com, Shallowater, TX, USA) using ion torrent PGM pyrosequencing. The 16S rRNA gene V4 variable region PCR primers 515/806 (Caporaso et al 2011) were used in a single-step 30 cycle PCR using the HotStarTaq Plus Master Mix Kit (Qiagen, USA) in accordance with the following conditions: 94°C for 3 minutes, followed by 28 cycles of 94°C for 30 seconds, 53°C for 40 seconds and 72°C for 1 minute, and a final elongation step at 72°C for 5 minutes. Sequencing was performed at MR DNA on an Ion Torrent PGM adhering to the manufacturer's guidelines. Sequence data were processed using a proprietary analysis pipeline (MR DNA, Shallowater, TX, USA). Sequences were denoised by being stripped of barcodes and primers. Sequences of less than 150bp were removed. Sequences with indistinct base calls and/or homopolymer runs exceeding 6bp were also removed. Operational taxonomic units (OTUs) were created and chimeras were removed. OTUs were defined by clustering at 3% divergence (97% similarity). Final OTUs were taxonomically classified using BLASTn against a curated GreenGenes database (DeSantis et al 2006).

Data and Statistical Analysis:

The amount of data received from MR DNA was incredibly dense with information. Approximately 1.5 million 16S sequences were collected that grouped into over 8,000 OTUs. Each read had a length of approximately 200 bp.

Although additional data was collected, it was decided that some species did not have replication across lakes, and these were excluded from the study. For this reason, Northern Pike (*Esox Lucius*), Yellow Perch (*Perca flavescens*), and Walleye (*Sander vitreus*) were dropped from this thesis and the current study.

All data and statistical analyses were performed using the metagenomeSeq (<http://bioconductor.org/packages/release/bioc/manuals/metagenomeSeq/man/metagenomeSeq.pdf>) and limma (<http://www.bioconductor.org/packages/release/bioc/vignettes/limma/inst/>) packages for the R statistical computing environment (R Core Development Team, 2015) with guidance from the metagenomeSeq user manual. Briefly, OTU counts, taxonomy, and sample names provided by MR DNA were loaded into R and subsequently into the metagenomeSeq package. OTU counts were then normalized. Normalization is required due to varying depths of coverage across samples. As identified in Table 1 variation in coverage existed especially in comparison to the environmental and fish samples (the environmental samples produced about 3.5 times the amount of reads that the fish samples produced).

Table 1. Total number of reads received per type of sample. Table shows the type of sample examined and the total number of reads for samples within each type produced.

Sample type	eDNA	largemouth bass	bluegill	black crappie
Number of reads	600,300	251,331	228,348	216,111

Normalization enables the ability to scale all of the counts across different samples so that they can be directly compared despite the fact that some samples have way more reads than others. So this vast difference in reads between lake and fish was accounted for. After counts were normalized, a principal component analysis plot and rarefaction plot were generated. Then, OTUs with less than a total of 75 reads across all samples were removed and the filtered data were renormalized. This filtered, normalized dataset was then used to fit a zero-inflated Gaussian mixture model that estimated coefficients for each species x lake combination. This model was then reoriented in

terms of several contrasts that enabled the identification of differentially abundant OTUs as a function of species, lake, and various combinations of species and lake using limma. Once significant OTUs were identified for each contrast, additional figures including heat maps and histograms were generated.

Results:

Lake x Fish comparison:

The histogram shown in Figure 3 highlights similarities, subtle differences, and diversity between fish and aquatic bacteria at the phylum level. The data suggests that the diversity of bacterial species in the two lakes sampled differs from the fish bacterial populations at the phyla level. It indicates that the lakes are more diverse than the fish (because there are more phyla represented in their bars on the histogram). The lakes also appear to be very similar despite being from different geographical locations.

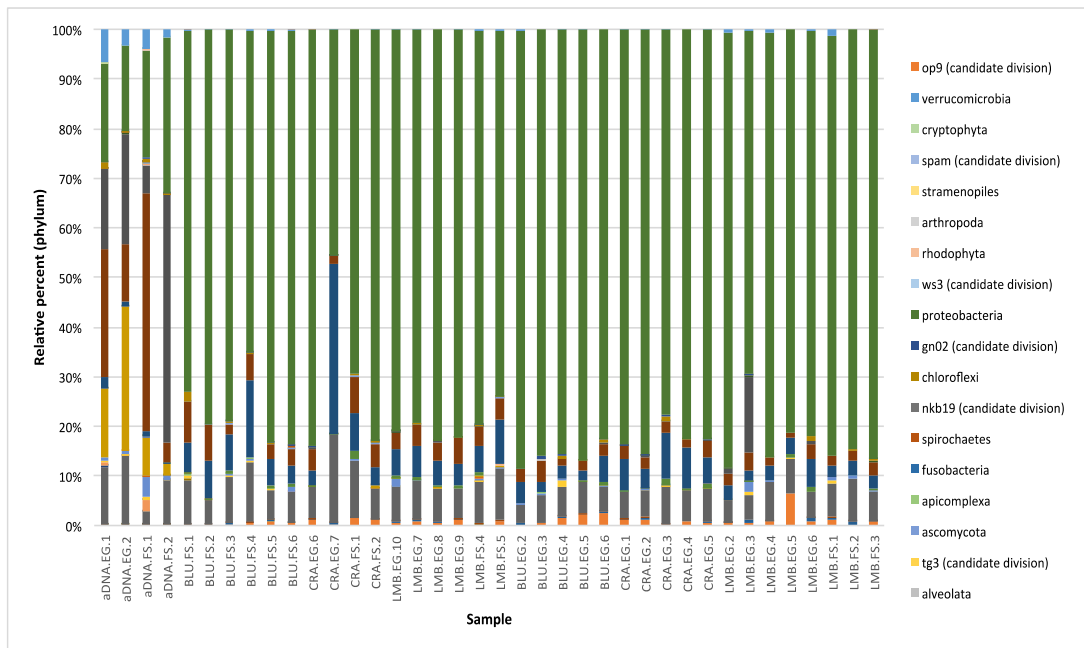


Figure 3. Stacked Column histogram of OTU phylum percentage per sample. This figure shows the relative abundances of bacteria (phylum) per observed sample. Data prior to normalization was analyzed. On the x-axis are each sample collected from East Gemini and Fireside Lake. Included are all samples collected from the lake (aDNA), bluegill (BLU), largemouth bass (LMB), and crappie (CRA) mucous.

Raw data from the histogram was filtered into R to identify the statistically significant OTUs between the aquatic and fish bacterial populations. Upon normalization and statistical analysis, the 38 samples produced 7,877 OTUs. From this, a sampling curve was produced (Figure 4 plots all samples and their relative depth of coverage versus the

number of detected features (OTUs)). The Michaelis Menten line represents the asymptotic association that should be expected if all samples were sufficiently sampled. In comparison of the aquatic lake and fish samples, the diversity appears to be much higher in the lake samples. While this data may indicate that all OTUs have not been identified for the samples, it is interesting to note the relationship between the aquatic samples and the fish. These results align with the histogram as seen in Figure 3 that also shows a larger variation in diversity in the lake samples as compared to the fish samples.

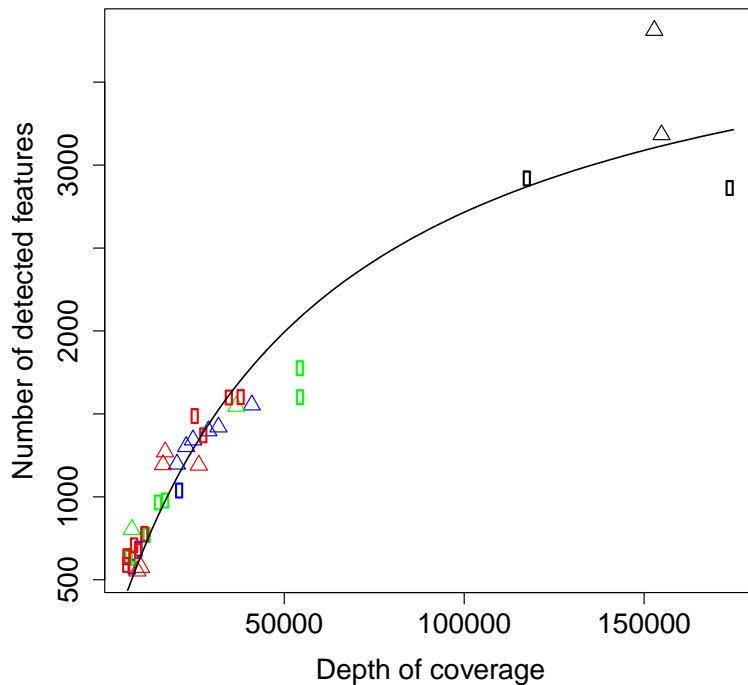


Figure 4. Number of OTUs detected (y-axis) as a function of number of reads per sample (x-axis). Fireside samples are represented by triangles, and circles represent East Gemini samples. Black shapes represent eDNA samples, blue represents bluegill samples, green represents crappie samples, and red represents largemouth bass samples. A Michaelis-Menten curve was fit to these data.

To continue to examine the diversity in the samples (and its variation between the fish and aquatic samples), a principal component plot data was constructed from normalized data. . The first and second components were used and accounted for 40.31% of the total variance within the samples. Figure 5 further exemplifies that the fish sample bacterial flora cluster very closely together suggesting that they don't account for much of the variance observed. The lakes, however, are very spread out and discrete along both the first principal component and the second principal component. This suggests that the aquatic lake samples account for the majority of the variance observed in the samples.

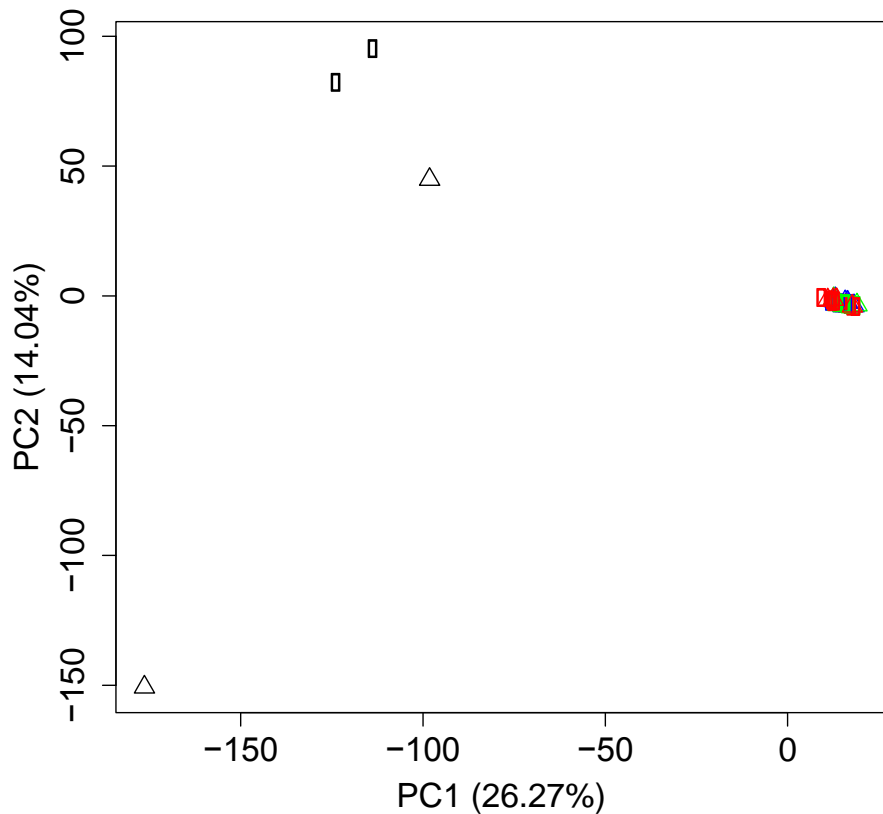


Figure 5. Principal component analysis (PCA) in which samples were ordinated in OTU space. The proportion of variance explained by the first two principal components is shown in parentheses. Fireside samples are represented by triangles, and circles represent East Gemini samples. Black shapes represent eDNA samples, blue represents bluegill samples, green represents crappie samples, and red represents largemouth bass samples.

Due to the lake samples increased diversity, it was important to identify all significant OTUs in the flora of lakes versus the fish samples. Approximately 2,500 OTUs were identified as significant between the lake and fish bacterial samples. Table 1 shows significance of OTUs per contrast. Each column adds up to 6,060. This represents the number of OTUs after sequences with reads under 75 were removed. For the eDNA v Fish contrast (lake vs fish), 633 OTUs were significant for fish in this contrast.

Conversely, 1,918 OTUs were significant for eDNA. As figure 6 suggests, the lakes accounted for many of the significant OTU differences. This can be seen in the vast shades of red on the left of the heat map. Red indicates overrepresented (significant) OTUs. When matched up to the appropriate OTU on the right hand side, OTU number and phylogeny were studied. Most of the significant OTUs (on the left hand side pertaining to the lake samples) represent the phylum *Cyanobacteria*. The very bottom right of the figure shows a cluster of OTUs that seems to be significant for the fish. When matched up to the appropriate OTU on the right hand side, it was determined that these OTUs represent the phylum *Proteobacteria*.

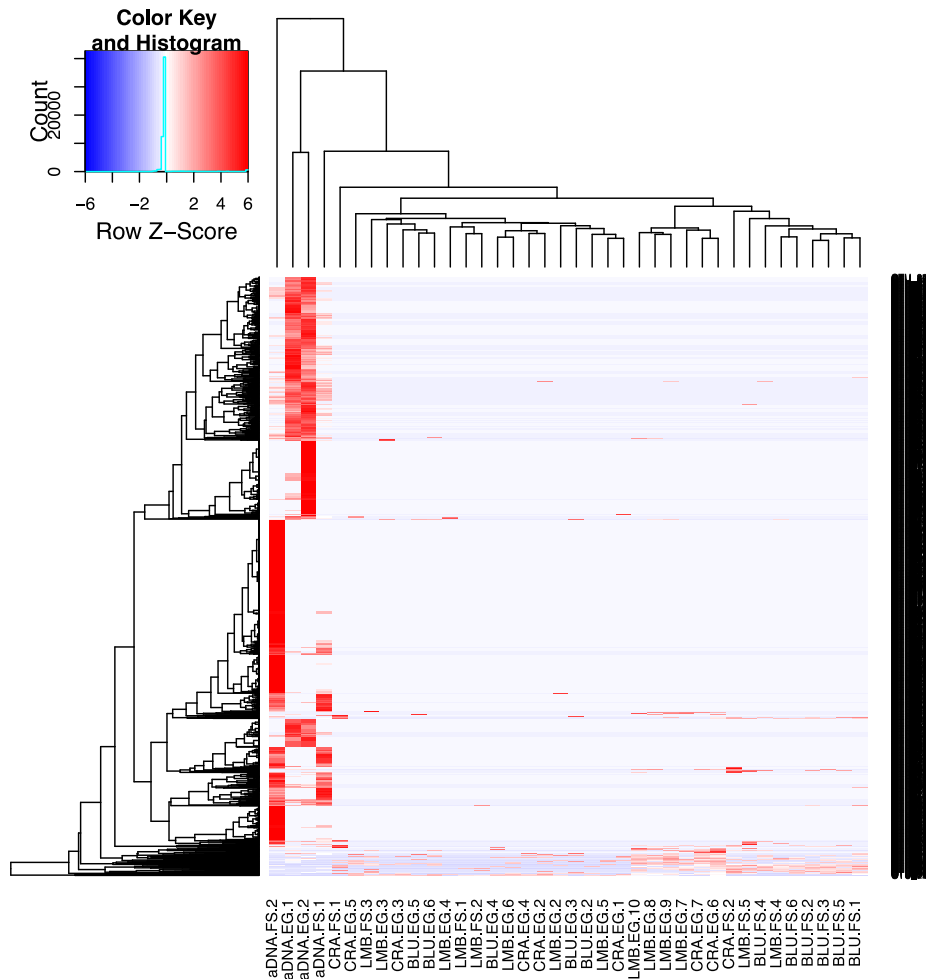


Figure 6. Heat map showing all significant OTUs between lakes and fish. OTU-wise (rows) and sample-wise (columns) hierarchical clustering was performed on a Euclidian distance matrix generated from normalized counts. The data are row-scaled and the color key is given on the upper left hand panel of the figure.

Of the 633 OTUs that were significant for fish approximately 89% of them belonged to the phylum *Proteobacteria*. Of the 1,918 OTUs that were significant for eDNA, approximately 35% of the OTUs accounted for *Proteobacteria*. This means that even though both environments harbored *Proteobacteria*, within the phyla, different populations within the phyla appeared to live in the water as compared to the fish. The second largest phylum containing significant OTUs in the aquatic samples was *Cyanobacteria*, which accounted for approximately 16% of the significant OTUs. Fish contained 0% significant OTUs for *Cyanobacteria*.

Species x Species Comparison:

After having compared the bacterial populations between lake and fish samples, the differences in bacterial populations between two fish species (bluegill and crappie) were investigated. There were 105 OTUs determined to be significant in this contrast. Another histogram was created to visually show differences between these two different species.

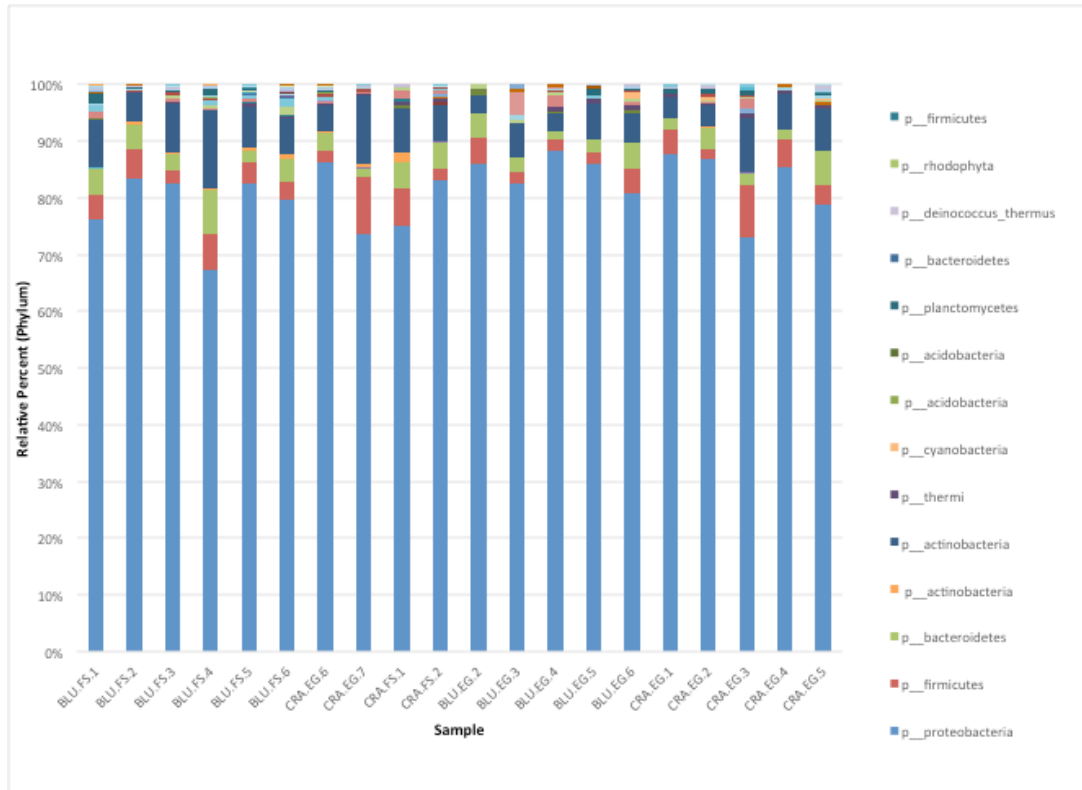


Figure 7. Stacked column histogram of OTU phylum percentage per bluegill/crappie sample. This figure shows the relative abundances of bacteria (phylum) per observed sample. Normalized data was analyzed. On the x-axis are all samples collected that were either a bluegill or a crappie.

According to Figure 7, there does not appear to be much variation between bluegill and crappie flora at the phylum level. Each bar on the histogram has 4 distinct similar areas. Each bar contains a significant amount of *Proteobacteria* (light blue), *Firmicutes* (light red), followed by *Bacteroidetes* (light green), then *Actinobacteria* (dark blue). The phyla on the very top of the histogram are variable, but constitute little percentage of the

overall phyla. Because differences were difficult to distinguish in Figure 7, a heat map between the significant OTUs for this contrast was constructed.

Figure 8 shows a heat map and comparison of the 105 significant OTUs between crappie and bluegill. The dendrograms on the top and left sides of the heat map clustered the samples and OTUs based on the similarity of normalized OTU counts. As described by the top dendrogram, the bluegill samples tended to cluster on the right of the figure and the crappie samples tended to cluster on the left. This indicates a difference in bacterial flora between the bluegill and crappie species. Then, within the species clusters, it appears that they are clustered based on which lake they came from. This clustering indicates a possible lake effect. This means that even though this is a contrast of crappie x bluegill, there still seems to be clustering based on similar normalized OTU abundances between the same species in different lakes. This can be seen in Figure 8 on the left hand side. Most crappie cluster here, but the crappie from Fireside (FS) and the crappie from East Gemini (EG) cluster separately.

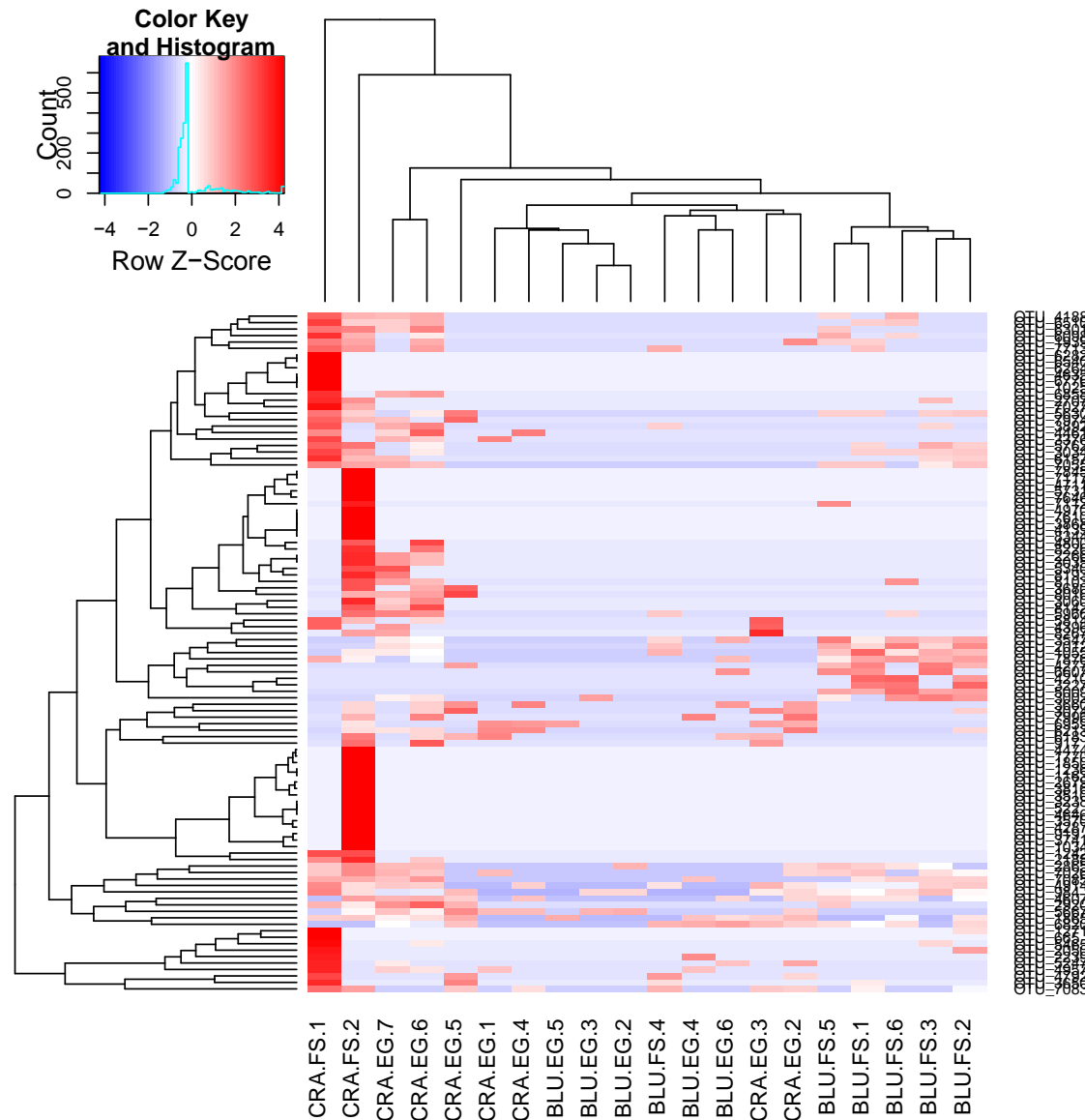


Figure 8. Heat map showing significant OTUs between crappie and bluegill. OTU-wise (rows) and sample-wise (columns) hierarchical clustering was performed on a Euclidian distance matrix generated from normalized counts. The data are row-scaled and the color key is given on the upper left hand panel of the figure.

Of the 11 OTUs that were significant for bluegill, 100% of them pertained to the *Proteobacteria* phylum. All 11 of these OTUs situate themselves in the middle of the heat map (the cluster of red on the right in the center). Of the 94 OTUs that were significant for crappie, 76% of them pertained to the *Proteobacteria* phylum and 9% of

them pertained to the *Actinobacteria* phylum. This means even though both contain significant OTUs pertaining to *Proteobacteria*, they contain different types of *Proteobacteria*. This also means that crappie contain a more diverse set of significant OTUs. It should be noted that the two crappie on the left hand side of the heat map are from Fireside Lake. These appear to be the most variable in terms of significant OTUs. More samples are required to determine if this is the norm for crappie flora in this lake.

Discussion:

Summary of Results

Our results suggest that there is a significant difference between the normal bacterial flora in the environmental lake and fish mucus. This is supported by the sample variation plot, principal component analysis, and histogram data. The sample variation curve shows the immense gap there is between number of OTUs from the lake samples and the number of OTUs from the fish samples. This suggests a more diversified flora in the lake samples. Observing the principal component analysis (PCA) supports this. The PCA supports the fact that the lake samples are more diverse because they account for most of the variance in the OTUs. Then species to species were compared. Contrasts for all species were made, but to begin, a comparison of the black crappie and bluegill bacterial flora was analyzed. The histogram showed similar bacterial phyla across both crappie and bluegill samples. When statistically significant OTUs were identified and examined in a heat map, differences in the flora became visible. It showed both a difference between lake and fish biota as described before, but it also showed a difference between species because the species clustered separately.

Taxonomy and Biological Significance

The PCA plot, rarefaction curve, and raw data histogram all suggest that the fish are not variable amongst each other. Because the fish samples have so fewer OTUs and are less diverse than the lake samples, one mechanism to explain this phenomenon is that the flora of the fish is specialized. The fish did not appear to transiently acquire all bacteria from the lake water and instead, maintain only a limited amount of normal flora. The significant OTUs for the fish in the lake vs. fish contrast were dominated by the phylum

Proteobacteria. It represented approximately 89% of the significant OTUs in the mucus microbiota. Certain genera of *Proteobacteria* found in the fish samples include *Methylobacterium*, *Herbaspirillum*, and *Sphingomonas*.

Methylobacteria are known for their ability to produce poly- β -hydroxybutyrate, which degrades short-chain fatty acids. This behavior is known to inhibit the growth of potential pathogens (Halet et al. 2007). This could be one possible reason as to why fish mucus harbors *Proteobacteria* and would demonstrate a mutualistic relationship between the normal flora and the fish surface. The genus *Herbaspirillum* constitutes the greatest percentage of significantly different bacteria found in the fish samples collected. Although this is the greatest percentage genus, not much is known about its impact in aquatic environments. This genus is a gram-negative and recent studies show evidence that identify this particular genus as a possible contaminant in DNA extraction kits (Salter et al. 2014). However, more studies need to be done on these recent findings in order to support this hypothesis.

Significant OTUs from the lake samples expressed phyla such as *Proteobacteria* (although less than the fish), and *Cyanobacteria*. The genus most readily expressed within the *Cyanobacteria* as part of the lake flora is *Chroococcidiopsis*. This bacterium is gram-negative and performs oxygenic photosynthesis (Cumbers 2014). It is logical that bacteria in the water perform photosynthesis as part of the aquatic food web and it is also logical that these photosynthetic bacteria do not seem to be in the normal flora of the fish slime. Fish exist higher in the food web as primary and secondary consumers and therefore may not directly benefit by harboring photosynthetic microorganisms as part of

their mucus flora. This difference further suggests a specialization of flora that are maintained on the fish mucus.

Significant OTUs for the bluegill in the bluegill x crappie contrast all pertained to the phylum *Proteobacteria*. The top genera expressed in this phylum were *Herbaspirillum* and *Aquabacterium* which both constituted approximately 17% of the significant OTUs for bluegill. *Aquabacterium* are rod-shaped, gram-negative bacteria (Kalmbach et al. 1999). Not much known about the importance of this type of microbe in aquatic environments. One study has identified this genus as a predominant member of various European drinking water distribution system biofilms (Bachman 2006)

The majority of significant OTUs for the crappie in the bluegill x crappie contrast also pertained to the phylum *Proteobacteria*. The top genera for these significant OTUs were *Pseudomonas* and *Phenylobacterium* Each accounted for approximately 10% of the significant OTUs for crappie. *Pseudomonas* is a gram-negative rod shaped bacterium (Palleroni 2010). Pseudomonads are common components of the nonpathogenic flora of fish (Bly et al. 1997). Certain species of them have also shown probiotic effects on fish pathogens (Korkea-aho et al. 2011). *Phenylobacterium* is also a gram-negative rod shaped bacterium. This genus is known for its extremely limited nutritional spectrum. It can even live on artificial compounds such as chloridazon, which is an active ingredient in herbicides. (Lingens et al. 1985). However, not much is known in terms of their impact on aquatic ecosystems.

This contrast of species x species should be better examined in future research to make sure the determined flora is true for the species and not just due to the lake from which the sample was taken.

In conclusion, this study gives insight into the normal flora of fish slime and aquatic flora. Aquatic flora is much more diverse than that of fish flora. Because fish flora is less diverse and more prone to harbor certain bacteria (*Proteobacteria*), it is much more specialized. Finally, when comparing between species as seen in the heat map and histogram between black crappie and bluegill, there seems to be a difference between species because the top genus' within *Proteobacteria* are different for bluegill and crappie within the contrast. Table 1 shows the types of contrasts, which were made during this analysis. More in depth research should be done on the other contrasts mentioned in the table. This extended research will validate the differences between lake and fish flora, but more importantly, better explain the subtle differences in flora between species.

Table 2. Number of significant OTUs per contrast. Significant OTUs pertaining to the object on the right of the contrast have red font. Significant OTUs that are significant for the left of the contrast have blue font.

Contrast	eDNAvFis h	EGvFS	eDNAeg_v s_eDNAfs	LMBvC	LMBvB G	BGvC	EGlmb_v_ FSlmb	EGbg_v_ Sbg	EGc_v_ FSc
Significant	633	789	917	89	22	94	48	66	139
Not Significant	3509	4513	4001	5964	6002	5955	5998	5937	5902
Significant	1918	758	1142	7	36	11	14	57	19

Acknowledgements:

There are several people to thank for the successful completion of this investigation. First, thank you to Dr. Barbara May for advising on this project and for providing insight into microbiology, ecology, and proper lab technique for handling of microbes. Second, thank you to Dr. Robert Page for assisting with the bioinformatics aspect of the project. Third, thank you to Dr. Jen Schaeffer and Dr. Robert Page for serving on this project's review committee. Fourth, thank you to Nate Ley for assisting with the summer research, which lead to the development of this thesis. Finally, thank you to the CSB/SJU Biology Department and Honors Thesis Program for making this investigation possible.

References:

- Azam F, Fenchel T, Field JG et al. (1983) The ecological role of water-column microbes in the sea. *Marine Ecology Progress Series*, 10, 257–263.
- Bachmann, R., & Edyvean, R. (2006). AFM study of the colonisation of stainless steel by *Aquabacterium commune*. *International Biodeterioration & Biodegradation*, 58(3/4), 112-118. doi:10.1016/j.ibiod.2006.06.008
- Biello, D. (2009). The Origin of Oxygen in Earth's Atmosphere. *Scientific American*.
- Bly, J.E., Quiniou, S.M.-., Lawson, L.A. and Clem, L.W. (1997) Inhibition of *Saprolegnia* pathogenic for fish by *Pseudomonas fluorescens*. *J Fish Dis* 20, 35–40.
- Bocchetta, M., Xiong, L., & Mankin, A. S. (1998). 23S rRNA positions essential for tRNA binding in ribosomal functional sites. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 95(7), 3525.
- Brüggemann, H., Bäumer, S., Fricke, W. F., Wiezer, A., Liesegang, H., Decker, I., & ... Gottschalk, G. (2003). The genome sequence of *Clostridium tetani*, the causative agent of tetanus disease. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 100(3), 1316.
- Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ et al (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *PNAS* 15: 4516-4522.
- Clarridge, J. E. (2004). Impact of 16S rRNA Gene Sequence Analysis for Identification of Bacteria on Clinical Microbiology and Infectious Diseases. *Clinical Microbiology Reviews*, 17(4), 840-862. doi:10.1128/CMR.17.4.840-862.2004
- Click, E. S., Moonan, P. K., Winston, C. A., Cowan, L. S., & Oeltmann, J. E. (2012). Relationship Between *Mycobacterium tuberculosis* Phylogenetic Lineage and Clinical Site of Tuberculosis. *Clinical Infectious Diseases*, 54(2), 211-219. doi:10.1093/cid/cir788
- Cumbers, John, and Lynn J. Rothschild. 2014. "Salt tolerance and polyphyly in the cyanobacterium *Chroococcidiopsis* (Pleurocapsales)." *Journal Of Phycology* 50, no. 3: 472-482. *Academic Search Premier*, EBSCOhost.
- Curtis TP, Sloan WT, Scannell JW (2002) Estimating prokaryotic diversity and its limits. *Proceedings of the National Academy of Sciences of the United States of America*, 99, 10494–10499.

- Czernilofsky, A., Kurland, C., & Stöffler, G. (1975). 30S Ribosomal proteins associated with the 3'-terminus of 16S RNA. *FEBS Letters*, 58(1-2), 281-284. doi:10.1016/0014-5793(75)80279-1
- Davis CP. Normal Flora. In: Baron S, editor. Medical Microbiology. 4th edition. Galveston (TX): University of Texas Medical Branch at Galveston; 1996. Chapter 6.
- DeSantis TZ, Hugenholtz P, Larsen N, M R, Brodie EL, Keller K *et al* (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and Environmental Microbiology* 72: 5069-5072.
- Dinman, J. D. (2005). 5S rRNA: Structure and Function from Head to Toe. *International Journal of Biomedical Science : IJBS*, 1(1), 2-7
- Field CB, Behrenfeld MJ, Randerson JT, Falkowski P (1998) Primary production of the biosphere: integrating terrestrial and oceanic components. *Science*, 21, 237-240.
- Groff J.M., Cutaneous biology and diseases of fish. *Vet Clin North Am Exot Anim Pract*. 2001 May;4(2) 321-411, v-vi. PubMed PMID: 11480359.
- Halet D, Defoirdt T, Van Damme P, Vervaeren H, Forrez I *et al*. (2007) Poly- β -hydroxybutyrate-accumulating bacteria protect gnotobiotic *Artemia franciscana* from pathogenic *Vibrio campbellii*. *FEMS Microbiol Ecol* 60: 363-369. [10.1111/j.1574-6941.2007.00305.x](https://doi.org/10.1111/j.1574-6941.2007.00305.x) PubMed: [17391334](https://pubmed.ncbi.nlm.nih.gov/17391334/) [PubMed]
- Hashizume, T., Takai, C., Naito, M., & Morisaki, H. (2005). Characteristics of the Mucus Layer on the Surface of the Bluegill (*Lepomis macrochirus*) and the Bacterial Flora in the Mucus. *Microbes and Environments*, 20(1), 69-80.
- Hogan, C. (2012). Commensalism. Retrieved from <http://www.eoearth.org/view/article/171918>
- Janda, J. M., & Abbott, S. L. (2007). 16S rRNA Gene Sequencing for Bacterial Identification in the Diagnostic Laboratory: Pluses, Perils, and Pitfalls. *Journal of Clinical Microbiology*, 45(9), 2761-2764. doi:10.1128/JCM.01228-07
- Jordan JA, Jones-Laughner J, Durso MB. 2009. Utility of pyrosequencing in identifying bacteria directly from positive blood culture bottles. *J. Clin. Microbiol.* 47:368-372
- Kadota K, Nishiyama T, Shimizu K (2012) A normalization strategy for comparing tag count data. *Algorithms Mol Biol* 7: 1-13 [PMC free article] [PubMed]

- Kalmbach, S.; Manz, W.; Wecke, J.; Szewzyk, U. (1999). "Aquabacterium gen. Nov., with description of *Aquabacterium citratiphilum* sp. nov., *Aquabacterium parvum* sp. nov. And *Aquabacterium commune* sp. nov., three in situ dominant bacterial species from the Berlin drinking water system". *International Journal of Systematic Bacteriology* **49** (2): 769. doi:[10.1099/00207713-49-2-769](https://doi.org/10.1099/00207713-49-2-769)
- Korkea-aho, T. L., Heikkinen, J., Thompson, K. D., von Wright, A., & Austin, B. (2011). Pseudomonas sp. M174 inhibits the fish pathogen Flavobacterium psychrophilum. *Journal Of Applied Microbiology*, *111*(2), 266-277. doi:10.1111/j.1365-2672.2011.05044.x
- Lal, Ashwini Kumar. 2008. "Origin of Life." *Astrophysics & Space Science* 317, no. 3/4: 267-278. *Academic Search Premier*, EBSCOhost (accessed April 10, 2015).
- Lingens, F.; Blecher, R.; Blecher, H.; Blobel, F.; Eberspacher, J.; Frohner, C.; Gorisch, H.; Gorisch, H.; Layh, G. (1985). "Phenylobacterium immobile gen. nov., sp. nov., a Gram-Negative Bacterium That Degrades the Herbicide Chloridazon". *International Journal of Systematic Bacteriology* **35** (1): 26–39. doi:[10.1099/00207713-35-1-26](https://doi.org/10.1099/00207713-35-1-26)
- Liu, Z., Desantis, T. Z., Andersen, G. L., & Knight, R. (2008). Accurate taxonomy assignments from 16S rRNA sequences produced by highly parallel pyrosequencers. *Nucleic Acids Research*, *36*(18), E120-E120. doi:10.1093/nar/gkn491
- Malaria Parasites. (n.d.). Retrieved February 23, 2015, from <http://www.cdc.gov/malaria/about/biology/parasites.html>
- María Ángeles Esteban, "An Overview of the Immunological Defenses in Fish Skin," *ISRN Immunology*, vol. 2012, Article ID 853470, 29 pages, 2012.
- nitrogen-fixing bacteria. (2015). In *Encyclopædia Britannica*. Retrieved from <http://www.britannica.com/EBchecked/topic/967311/nitrogen-fixing-bacteria>
- Palleroni, N. J. (2010). "The Pseudomonas Story". *Environmental Microbiology* **12** (6): 1377–1383.
- Pernthaler J (2005) Predation on prokaryotes in the water column and its ecological implications. *Nature Reviews Microbiology*, *3*, 537–546.
- Pomeroy LR, Williams PJI, Azam F, Hobbie JE (2007) The microbial loop. *Oceanography*, *20*, 28–33.
- Rastall R.A. (2004). Bacteria in the gut: friends and foes and how to alter the balance. *J. Nutr.* 134:2022S–2026S.

Salter, S. J., Cox, M. J., Turek, E. M., Calus, S. T., Cookson, W. O., Moffatt, M. F., & ... Walker, A. W. (2014). Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biology*, *12*(1), 87-98. doi:10.1186/s12915-014-0087-z

Sears, C. L. (2005). A dynamic partnership: Celebrating our gut flora. *Anaerobe*, *11*(5), 247-251. doi:10.1016/j.anaerobe.2005.05.001