# BioSeq: Bioinformatics Inquiry through Sequencing

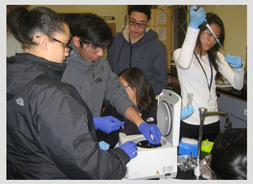
**Elise Gan**, Quin Bottom-Johnson, Matt Cassar, Jack Reid, Hannah Voelker,
Kristin Harrington, Mark Hartman, Matthew Fierman, Prof. Donna Slonim, Prof. David Walt

SEPA SCIENCE EDUCATION PARTNERSHIP AWARD
Supported by the National Institutes of Health

## Introduction

Bioinformatics research facilitated by next-generation DNA sequencing (NGS) is becoming increasingly commonplace, but exposure to these important technologies remains out of reach for younger students. The Bioinformatics Inquiry through Sequencing (BioSeq) program serves as an introduction to NGS and bioinformatics by providing high school students with hands-on exposure to genuine, research-based activities. We have developed guided in-class modules that ask open-ended research questions and provide facilities and supervision for students to develop their own projects. BioSeq has already successfully reached more than 150 high school students, and initial results indicate strong knowledge and attitudinal gains across all levels of learners.

## Goals



- Bring sequencing and bioinformatics into the classroom
- Enable independent student and teacher research
- Provide training and development to educators

## Experimental Modules

- Microbiome Portrait
- Food Microbiome
- Water Microbiome
- Genetics of Race
- Personal Genomics
- Mutations Investigation

## Research Team

Quin Bottom-Johnson, Matt Cassar, Elise Gan, Jack Reid, Hannah Voelker – Undergraduate Science Mentors
Dr. Mark Hartman – Lab Supervisor
Kristin Harrington – Lab Technician
Dr. Matthew Fierman – Program Administrator
Prof. Donna Slonim – Co-PI
Prof. David Walt – Principal Investigator

## Acknowledgements

This project was supported by the Office Of The Director, National Institutes Of Health under Award Number R25OD010547-01. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## DNA Quantification: Qubit

Qubit fluorometric quantification is used to show how much DNA there is in a sample. The procedure is performed after PCR samples are cleaned up. The concentration of the DNA is found and measured in μg/mL. The following results were found for both samples after each round of PCR:

After PCR 1
- Mystic “MS” – 4.70 μg/mL
- Cold Tap “TS” – 2.32 μg/mL

After PCR 2
- Mystic “MS” – 13.1 μg/mL
- Cold Tap “TS” – 9.64 μg/mL

# Microbiome of Water

## Background

Microbes can be found in virtually all sources of water. Further understanding the population dynamics of water microbiomes can improve the standards of human living. A better sense for how microbes grow and interact in various water systems could have a wide range of applications. For example, it is important to know whether or not bottled water and other sources of drinking water are safe. Not all water is safe and studying the microbiomes of water can help investigate water-borne diseases. Through this experimental module students can learn about what organisms live in our most important resource.

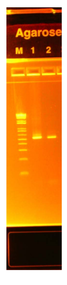
## Purpose

- Sequence microbial DNA from water samples using NGS
- Compare microbiomes of different water sources
- Use Jaccard similarity to measure similarity of samples

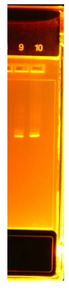
## Materials and Methods

Samples were collected from the Mystic River, the cold tap from Tisch Library, and the hot tap from 200 Boston Ave. The water samples were filtered through syringe filters and treated with chemicals to release the DNA. The GenElute Water RNA/DNA Purification Kit (Sigma-Aldrich) was used to isolate the DNA, and PCR amplification was performed on the samples. Following PCR, AMPure beads (Agilent) were used to clean the PCR samples. A second round of PCR was performed to index the samples prior to sequencing. Again, AMPure beads were used to clean the PCR samples. Before sequencing, the Bioanalyzer was used to confirm the quality, size, and amount of DNA in each sample.

## DNA Quantification: Gel Electrophoresis

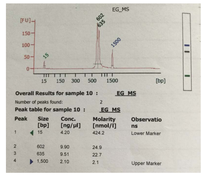


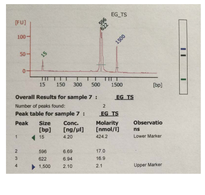

After PCR 1, the samples were analyzed using gel electrophoresis. On the left, in the column labelled “M,” the ladder can be seen. Well “1” is a positive control sample which shows that the primers used in PCR should have no problem binding to the samples. Well “2” shows the sample “TS” which is the cold tap sample. On the right, well “9” shows the sample “MS” taken from the Mystic River and well “10” is another positive control.




## DNA Quantification: Bioanalyzer

The Bioanalyzer performs capillary electrophoresis to ensure there is sufficient DNA for sequencing. The Bioanalyzer results were consistent with both gel electrophoresis and Qubit.

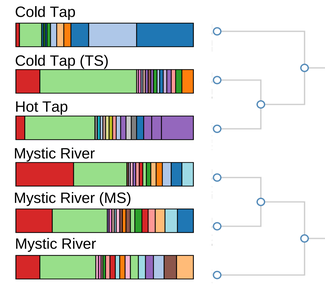



The picture on the left shows the analysis of the Mystic “MS” sample while the picture on the right shows the analysis of the Cold Tap “TS” sample. The concentrations of DNA at indicated sizes are shown at the bottom of each image.

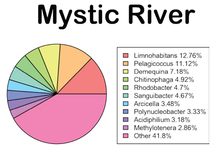
## Results: Dendrogram

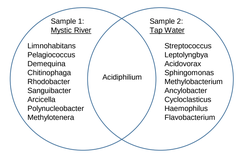
This dendrogram shows the distribution of bacteria within each sample discussed (MS and TS) and also other samples processed. The sixteen most prevalent bacteria in each sample are shown and the samples are clustered according to their similarity. Green and red bars refer to the Other and Unclassified categories of bacteria in each sample.




## Results: Sample Comparisons

The pie charts below show the ten most prevalent bacteria in each sample. The sample on the left, “MS,” is from the Mystic River and the sample on right, “TS,” is cold tap water from Tisch Library. Similar to the dendrogram, the pie charts have an “other” section, shown in pink, that refers to many small amounts of varied bacteria in the samples. The two samples can be compared using Jaccard similarity. The Venn diagram on the right helps calculate Jaccard similarity. Of the prevalent bacteria, there are nineteen different kinds. As seen on the Venn diagram, there is only one prevalent bacteria that the two samples share. Because of this, the Jaccard similarity between the two samples is 1/19 or 5.3%


## Conclusion

In this experiment, samples from different sources were collected and compared. From the results of the experiment, Jaccard similarity was calculated and it was shown that water samples from two different locations have two almost completely different microbiomes when it comes to prevalent bacteria. From the dendrogram, it can be seen that water samples from similar sources have more similar microbiomes. This experiment demonstrates the practical applicability of NGS methods to analyze water microbiomes.