

Using Microbiome Analysis to Identify Factors Influencing the Presence of Human Pathogens in Alder Creek

S. Ronk, L. Thompson, M. Thompson, A. Gunn



Abstract

Bacteria form the foundation of the food web, and identification of bacteria present can reveal a great deal about an environment. Previous research used 16S RNA sequencing over eight sites that were selected along Alder Creek to identify bacterial strains present in the stream. Through this, bacterial human pathogens were identified.

Variables such as ecological diversity, bacterial diversity, the percent of the population composed of the dominant bacterial species, and stream bank altitude were evaluated for effects on the presence of human pathogens within the first four sites. The Akaike Information Criterion (AIC) was used in the model selection process and a general linear model was used to confirm statistical significance.

The study yielded that, within the first four sites, lower bacterial diversity, along with a higher presence of a dominant bacterial species, correlated with a higher presence of human pathogens. The nature of this relationship remains unclear. The use of AIC was limited by the number of sample sites.

The need for additional data could be ameliorated by supplementing classroom instruction with data collection. This would provide students exposure to field work and the opportunity to gain hands-on learning experience. The data collected could provide valuable information on using microbiome analysis in order to identify the variables influencing pathogenic bacteria in Alder Creek.



Figure 1: Map of sample sites along Alder Creek (Richardson et al. 2016), stars indicate sites included in this preliminary analysis.

Introduction

Microorganisms are the foundation of the food web. Bacteria and Archaea compose more active biomass than plants or animals⁷. In a natural ecosystem, removing a plant or animal from that ecosystem can have serious impacts on the equilibrium of that environment. Just as with plants or animals, removing a bacterial species, or changing the composition of bacterial species in an environment, can have cascading impacts that extend to the top of the food chain^{1,6}. This is why understanding the dynamic of microbial interactions in an ecosystem, and the variables influencing the composition of microbes, is important.

The mechanisms of stream ecology on a molecular level are poorly understood. This holds especially true when considering the relationship between humans and bacteria in local streams. Urbanization can have drastic impacts in microbial composition in streams.² Variables such as biological waste, detergent run-off, and alterations in surrounding habitat can influence microbial composition. A change in microbial composition results in changes in stream health. Changes in stream health can result in a stream that can no longer support species that are native to the stream. These changes are deleterious to human beings, as these changes may result in water that is no longer safe for human consumption, while also depleting fisheries that can have drastic impacts on local economies that are dependent on them.

Alder Creek is an ideal water source to investigate the effects that urbanization has on microbial composition in water. This creek possesses stretches that have minimal human impact, while also possessing downstream stretches that take tortuous courses through man-made culverts under roads representing regions that are under heavy influence of urbanization. This significant contrast of surrounding environment will allow for a comparison of how urbanization can influence the composition of micro flora when compared to a control with minimal human impact.

Model selection is the determination of how specific variables, or explanatory variables, influence another particular variable, or response variable. Often, a single predictor variable is selected and evaluated against a response variable. This paradigm in scientific investigation limits the ability to identify correlations that may be pertinent to a particular response, but not predicted by researchers. By establishing a large data set that documents variables, such as; biodiversity, streambed, and nearby human influences, characterizing a stream and its influencing variables is possible.

Bacterial strains can be identified using 16S RNA sequencing. Previous 16S RNA sequencing has characterized the microbial composition of the stream. This process is repeated yearly. By comparing these results to the dataset characterizing the streams environment, correlations to microbial content relating to urbanization, or other disturbances, can be established.

In order to establish these correlations, the statistical modeling software R will be used in conjunction with the model selection function Akaike Information Criterion, or AIC. By using this function, it is possible to identify correlations that may not be predicted or understood.³ By using these functions with "R", there is no limit to the number of variables that can be used, assuming requirements for degrees of freedom is met. This will allow for the research to hone in on variables of interest that result in dangerous, or pathogenic, bacteria in local streams. Using this method we can establish multiple correlations to pathogenic bacteria, and establish a web of influence that humans have on stream health on a microbial health, while addressing how these influences result in stream health that influences human health.

Methods

Characterizing Plant-life:

A 2x8 meter grid extending from the bank of the stream was established outward from the stream bank in one-meter increments. Species count of plant life was conducted, and species data was binned according to the grid points in which it was found.

Biodiversity:

A species count was conducted of the surrounding wildlife. This species count was combined with the plant count data. Using the Shannon diversity index, a biodiversity figure was acquired of the surrounding area of the sample site.

Dissolved Oxygen, pH, Temperature:

Dissolved oxygen, Temperature, and pH was measured. measurements were recorded at a depth of 6-8 inches.

Statistics:

Statistics were done using "R" statistical programming software. AIC was used in model selection. General linear model established significance.

Results

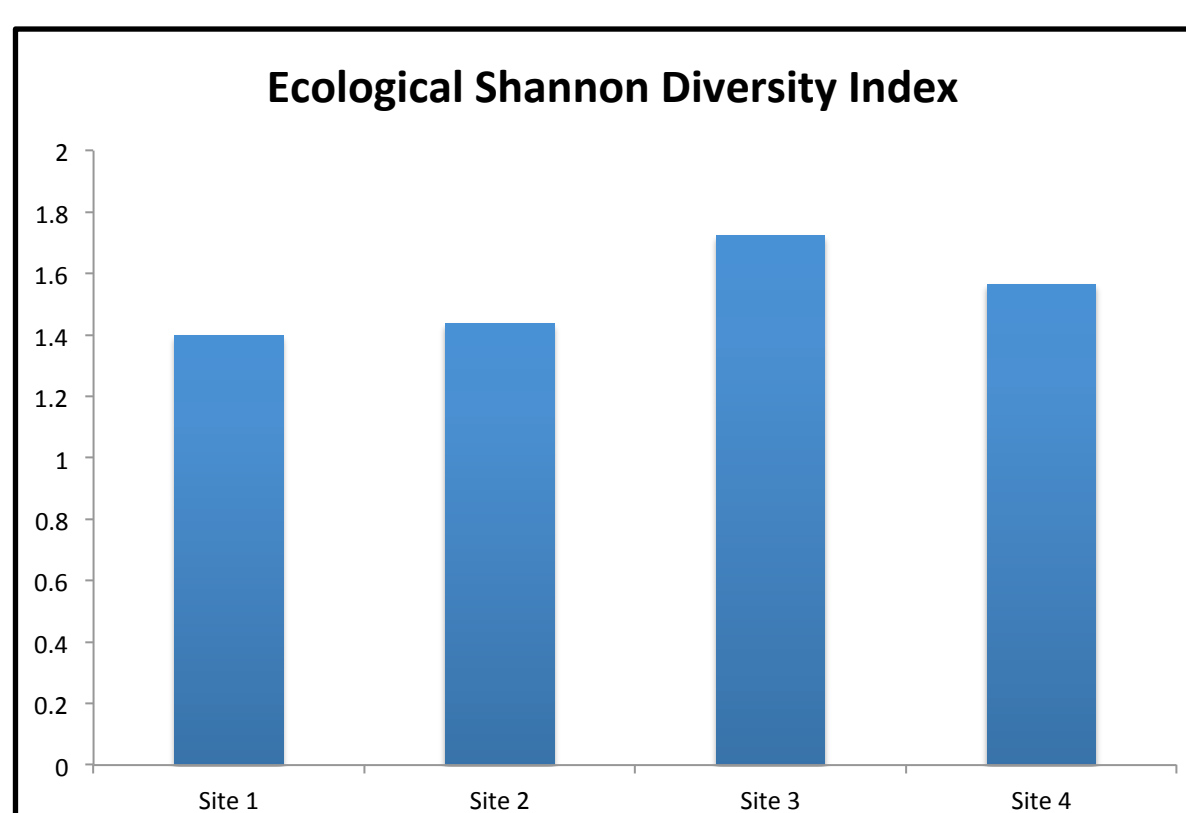


Figure 2: Total Biodiversity at sample sites. Biodiversity was calculated using Shannon Diversity Index.

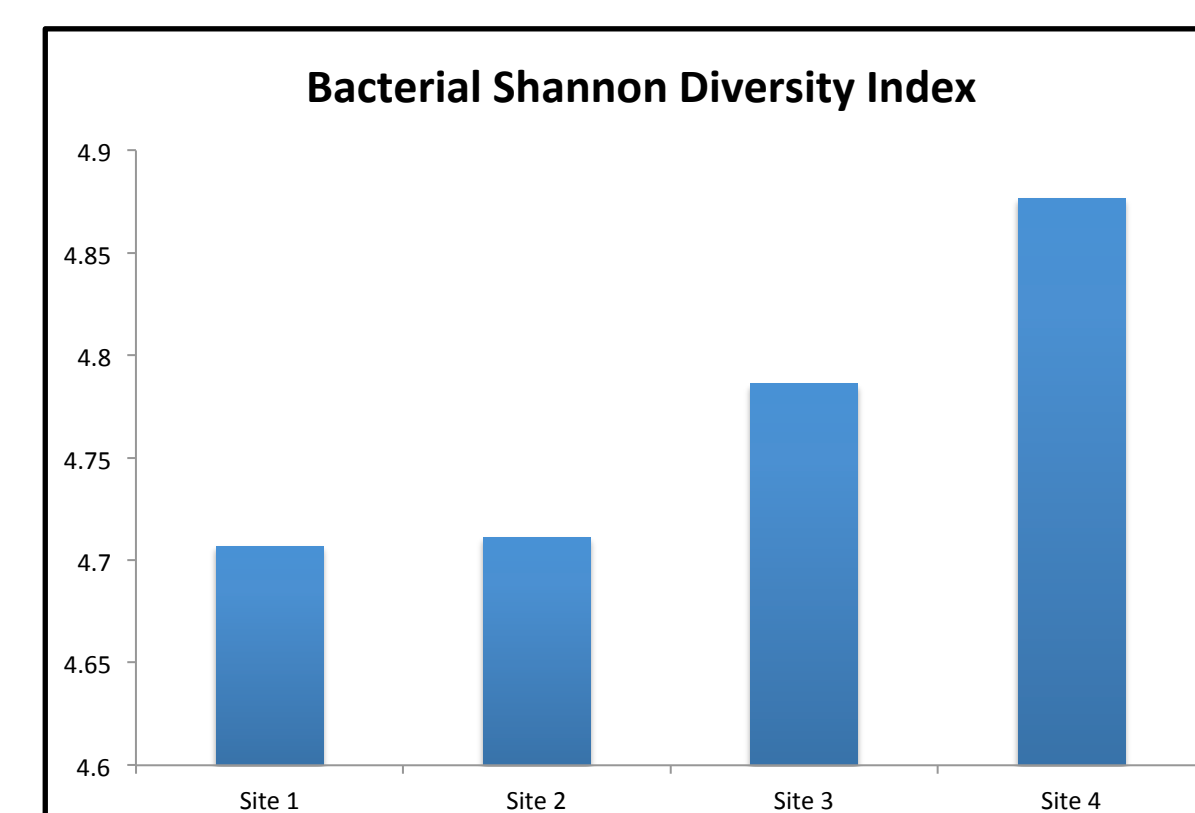


Figure 3: Bacterial strains were identified using 16S RNA sequencing. Diversity determined using Shannon Diversity.

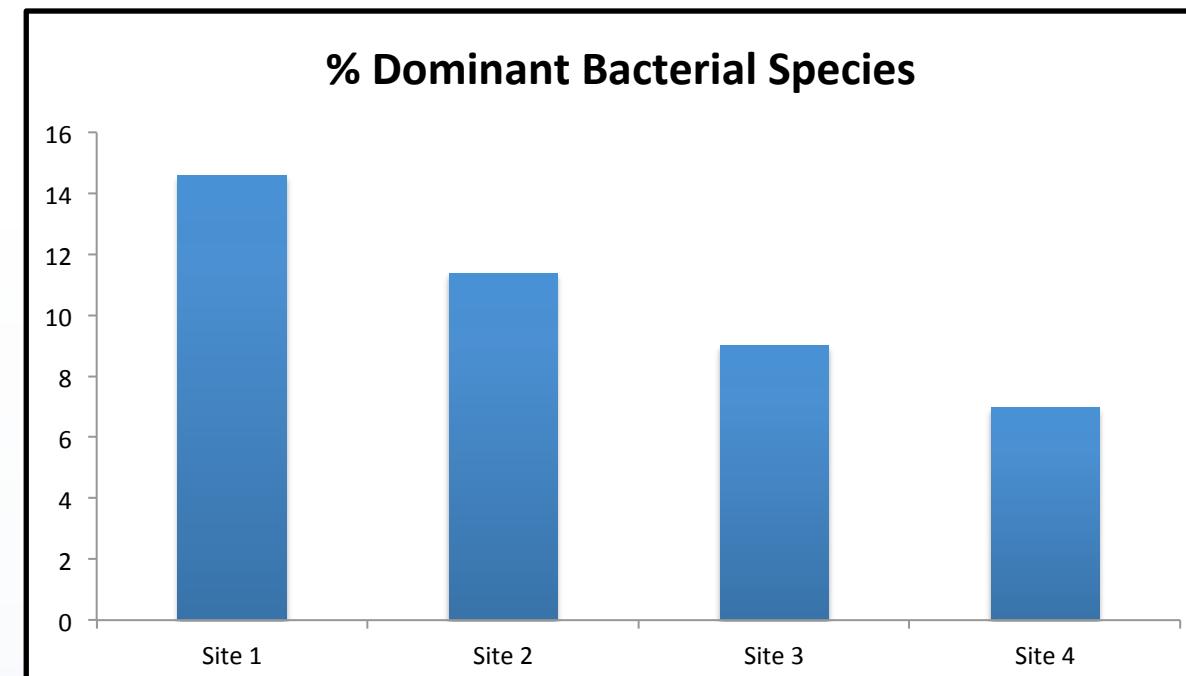


Figure 4: Diversity of environment near sample site. Diversity was calculated using the Shannon Diversity index

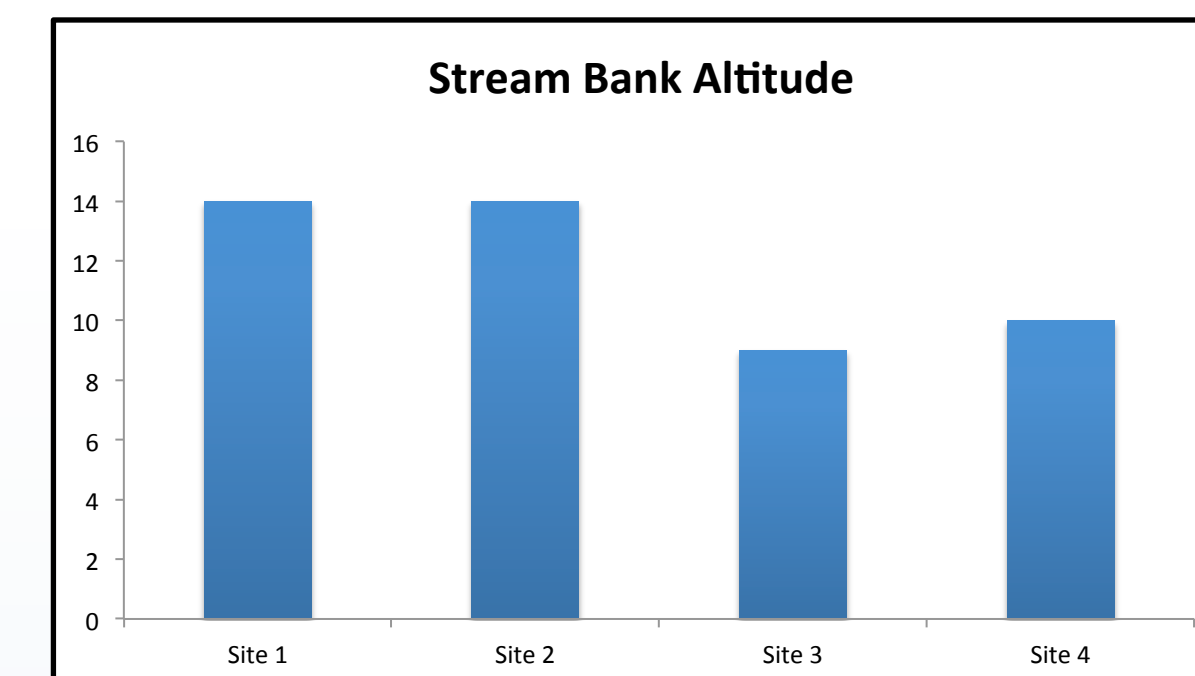


Figure 5: Altitude of stream bank. Altitude was determined using a barometric altimeter.

Ecological diversity differed among the four different sample sites (figure 2). The lowest ecological diversity was noted closest to Lake Swano, at site 1, and increased up to site 3, halfway between Huntley and Harriman Street. There was a slight drop in ecological diversity at site 4, which is located next to Harriman Street. There were no statistically significant trends in this data set. Bacterial diversity was the lowest at site 1, closest to Lake Swano while increasing progressively up through site 4 (figure 3). A generalized linear model determined that there was a significant relationship with bacterial diversity and the presence of human pathogens (p=.05). Specifically, that a low bacterial diversity related to a higher presence of human pathogens. The percentage of dominant bacterial species (figure 4) was calculated using the percentage that the most dominant bacteria composed of the total number of bacteria present in the sample. The highest percent of composition of the dominant species was noted at site 1, and that percentage dropped sequentially as the distance from Lake Swano increased. A general linear model yielded a statistically significant trend when concerning the presence of human pathogens and the higher presence of a dominant species (p=.05). Specifically, that a higher percent of a dominant species results in a higher presence of human pathogens. Stream bank altitude (figure 5) yielded no significant trend.

The presence of bacteria that are pathogenic to humans was calculated by identifying pathogens and dividing the total by the total number of bacteria (figure 6). Bacteria that are opportunistic pathogens were excluded. The presence of human pathogens was highest at site 1. Sequentially, the presence of human pathogens changed from 4-6%, however there was no statistically significant difference in these numbers.

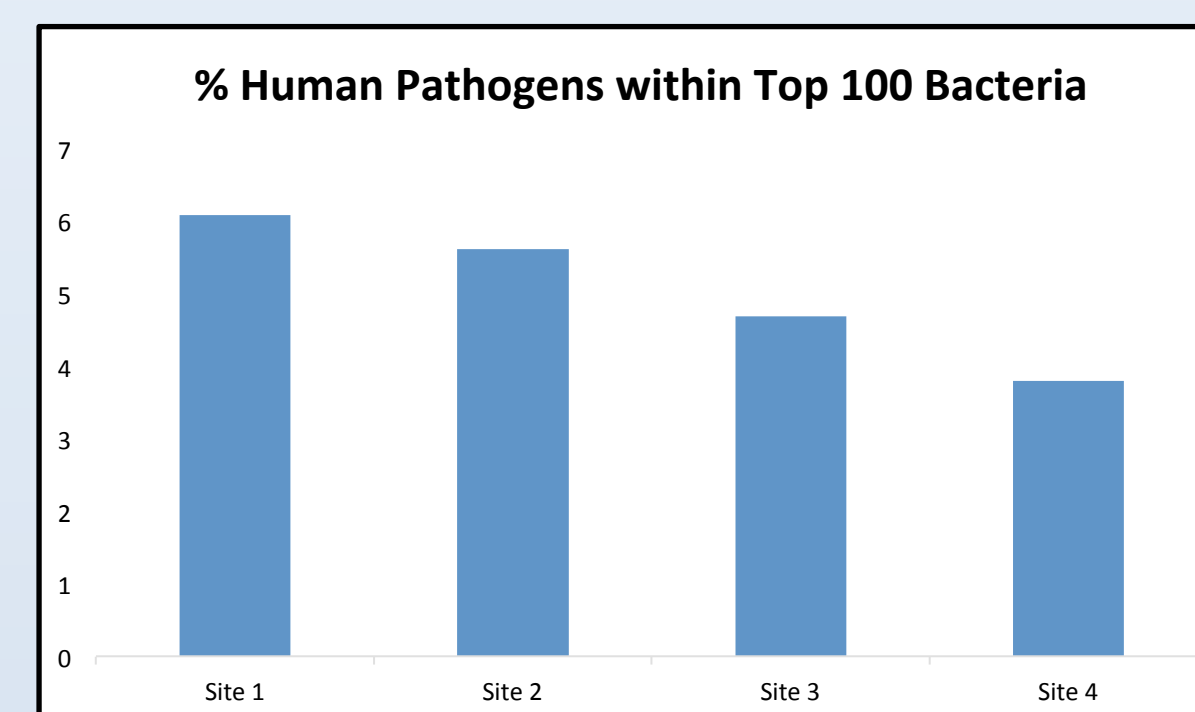


Figure 6: Number of bacteria pathogenic represented as a percentage. Bacteria were identified using 16S RNA

```
Start: AIC=-24.86
Data$Pathogens ~ Data$of.dominant.bacteria + Data
$BacterialH

          Df Sum of Sq  RSS   AIC
<none>                0.001786 -24.8566
- Data$of.dominant.bacteria  1  0.12255 0.124333  -9.8843
- Data$BacterialH          1  0.17114 0.172922  -8.5648

Call:
lm(formula = Data$Pathogens ~ Data$of.dominant.bacteria
+ Data$BacterialH)

Coefficients:
(Intercept)  Data$of.dominant.bacteria  Data
$BacterialH      37.1890          0.1454         -7.0593
```

Figure 7: An AIC print out in R

Figure 7 shows an AIC analysis run in the program R. The model consisting of both percentage of dominant bacteria and bacterial diversity represents the highest value through the AIC analysis. Further evaluation of the data shows that a combination of these variables using a generalized linear model presents a significance of .01. Further analysis using the generalized linear model determined that the separation of percentage of dominant bacteria and bacterial diversity yielded a p-value of .0283 and .0203, respectively.

Discussion

Monitoring the health of streams is important, both in consideration for human health and when considering local fisheries. Using the microbiome as an indicator for characterizing stream behavior and health is a relatively novel concept employed here. Microbes can provide invaluable information on stream health in several ways. Most apparently is the identification of microbes that are pathogenic to humans. Less evident is that it can provide indirect information about the composition of the stream. For example, if the stream possesses iron-reducing bacteria, than this would suggest the presence of iron in the stream. With data exploration, variables influencing the presence of certain bacteria, such as those pathogenic to humans, can be identified.

How the presence of bacteria pathogenic to humans is related to bacterial diversity and the percent of dominant bacteria at the sample site remains unclear. The intermediate disturbance hypothesis dictates that a moderate amount of disturbance will yield the highest biological diversity in a given environment.⁸ With too much disturbance, only the most resilient species will survive, or all species will die. With too little disturbance, a few species will dominate and supplant all other species in an environment.⁸

The least impacted sample sites possessed a lower bacterial diversity, while also possessing a dominant species that composes more of the total microbe population. Sites downstream with higher disturbance demonstrated higher bacterial diversity and a lower percent of the dominant species of the total population. This reflects the intermediate disturbance hypothesis. When applying the intermediate disturbance hypothesis, it would seem likely that a higher bacterial diversity would translate into a higher presence of human pathogens due to an increased ability of human pathogens to supplant native bacteria species. However, this is unclear in this dataset. There may be an unidentified variable influencing the presence of pathogens. It is also likely that more data is required, as this is only four sites during one year. Future research could confirm, or refute, this idea.

This process of stream monitoring was limited by the number of sites surveyed. This need could be satisfied supplementing classroom instruction with field work gathering data. This would not only satisfy the need for more data, but it would also give students an opportunity to participate in research and learn techniques that are used in field ecology. In just this project the following techniques were used: plant identification, Shannon diversity index, identification of amphibians, the use of pH/DO2/Temp meter, the use of "R", biometry, and data set construction and management in excel.

The approach used regarding data compilation and statistics is flexible. The software used, "R", is free and it can work with an infinite sized data set. This will allow for a year by year comparison of how human or ecological disturbances, such as floods or road work, effect the microbiome in a stream. Using "R" also allows for easy future modification of the research question. For example, if nitrogen fixing bacteria became of interest in the future, then only simple modifications to the inputs into "R" are needed. The only limit to changing the research question is the amount of data that is gathered each year, specifically that the data of interest was collected. By utilizing data collection to supplement classroom instruction, a great deal more data can be collected, during multiple points throughout the year.

AIC, used with "R", allows for a rapid evaluation of explanatory variables influencing a selected response variable, a process known as model selection. This means that explanatory variables influencing a response variable can be evaluated much quicker. This, however, is not the only benefit to an AIC analysis. It also allows for the evaluation of multiple variables on a response variable. This means that, in a hypothetical situation, while a couple variable individually may not yield statistically relevant results, but when combined together they do, an AIC analysis will identify this. This results in a faster, much more efficient means to evaluate correlations in collected data.

Together, this approach to evaluation of stream health could prove a powerful and flexible method, while also providing educational benefit to students. The limit to this approach is the amount of sites sampled, however this can be ameliorated using classroom students to collect data. This not only satisfies the need for data, but also provides educational material in the field which students can use. In conclusion, using 16S RNA analysis in conjunction with classroom education and statistical analysis processes could prove useful in the identification of pathogenic stream bacteria and the factors that influence their presence and abundance

References

- Berg, M. P., E. T. Kiers, G. Driessen, M. Van Der Heijden, B. W. Kooij, F. Kuenen, M. Liefing, H. A. Verhoef, and J. Ellers. Adapt or disperse: understanding species persistence in a changing world. *Global Change Biology* 16: 587-598. (2010)
- Cydzik-Kwiatkowska, A., & Zielińska, M. Bacterial communities in full-scale wastewater treatment systems. *World Journal of Microbiology and Biotechnology*, 32(4), 66. (2014)
- Jaffer, A. J. maximum likelihood Angular Resolution of Multiple Sources, Circuits Systems and Computers. Nineteenth Asilomar Conference on, pp. 68-72, 1985, ISSN 1058-6393. (1985)
- Richardson, S., Kalisch, J., Richardson, J., Gunn, A., Microbiome analysis of Alder Creek. Poster presented at: GCH Summer Research Program. 2016 Aug 23. Aberdeen, WA.
- Sivonen K, Kononen K, Carmichael WW, Dahlem AM, Rinehart KL, Kiviranta J, Niemela SI. Occurrence of the hepatotoxic cyanobacterium *Nodularia spumigena* in the Baltic Sea and structure of the toxin. *Appl. Environ. Microbiol.* 55 (8): 1990-5. PMC 202992 Freely accessible. PMID 2506812 (1989)
- Van der Putten, W. H., et al. Plant-soil feedbacks: the past, the present and future challenges. *Journal of Ecology* 101: 265-276 (2013)
- Whitman, W. B., D. Coleman, and W. Wiebe Prokaryotes: The unseen majority *PNAS* 1998 95 (12) 6578-6583 (1998)
- Wilkinson, D. M. The Disturbing History of Intermediate Disturbance. *Oikos*. 84 (1): 145-7. JSTOR 3546874. doi: 10.2307/3546874 (1999)

Acknowledgements

This project was funded by the Rose Foundation for Communities and the Environment, and was carried out as part of the GHC Fish Lab Summer Research Program.

