Stafford Maracle 20005033

Shannon Edie 20008148

Logan Lafreniere

Yukthi Wickramarachchi

Katy Waters

BIOL 432 Computational Biology

TA: Ying Chen

Submitted: 02 March 2020

Project Proposal

## Overview

Environmental DNA (eDNA), or the sequencing of DNA fragments from environmental samples such as from lakes and rivers, has recently become a popular topic in ecology. eDNA has applications in environmental assessment, conservation and monitoring of wildlife, discovery of novel niche species, and more.

Using the techniques we have learned in BIOL 432, we hope to explore an eDNA dataset to identify species, investigate species richness, and compare diversity amongst sampled sites.

## Data Description

Our data utilizes environmental DNA (eDNA) to analyze species diversity in freshwater ecosystems.

Using data from Lim (2016) obtained from DryadDatasets, our project will consist of water-based environmental DNA (eDNA) samples with universal metazoan primers applied. 42 samples were collected from each of two freshwater reservoirs. eDNA is a cutting edge technique primarily used to assess freshwater ecosystems diversity and health. This technique involves isolating and sequencing DNA segments within water samples using DNA primers created to target specific species or our case to target any metazoan species. DNA segments tend to below quality due to the nature of eDNA largely composing of detritus or excrement. This leaves us with a data set of 84 fragmented DNA strands in FASTA format (a string of nucleic acid bases) from unknown metazoan species.

## Focal questions

1. **What species are we working with?** Working with eDNA and universal primers our nucleotide sequences may belong to any metazoan species. Narrowing down each sample to the family or genus level, with hopes to the species level, will allow us to broadly determine the diversity of species among our samples. We will do this using BLAST and Rentrez in R to compare unknown sequences with NCBI GenBank database.
2. After classifying our samples to the family level or lower we can then ask: **How diverse is our dataset?** This can be done by creating a Shannon index to determine species richness and relative abundance among the sample population. This method may have limitations due to some sequences only being identified at the family level, and with some of our sequences being identified to a species level. We will consult with our TA for recommendations if problems arise.
3. Adding to our understanding of the sample populations diversity, we will then question the relatedness of the species in these reservoirs. **What is the weighted phylogenetic distribution of our sample population?** By creating a weighed phylogenetic tree with the raw FASTA data we can determine the variation in the sequences collected by using eDNA. Another tree may be created using the results from question one to sort samples according to determined taxonomy.
4. Finally, we may attempt to **compare the distribution, density, diversity of the eDNA samples found between the two reservoirs.** Frequently, eDNA is used for environmental assessment and monitoring. As such, comparison amongst sites is an apt exploration of the powers of eDNA.

## Remarks

Our focal questions may shift or be expanded upon as we progress through the course material in BIOL 432, depending on what new skills and tools we learn to evaluate our data.

Additionally, we may further explore current research related to eDNA evaluation, as well as the rest of the methods described in the paper which we sourced the data from, for other inspiration and ideas on what to explore with our dataset.

## References

Lim, N.K.M., Y.C. Tay, A. Srivathsan, J.W.T. Tan, J.T.B. Kwik, B. Baloglu, R. Meier, D.C.J. Yeo. 2016. Next-generation freshwater bioassessment: eDNA metabarcoding with a conserved metazoan primer reveals species-rich and reservoir-specific communities. Royal Society Open Science, Article-journal 3(11) 1-12.