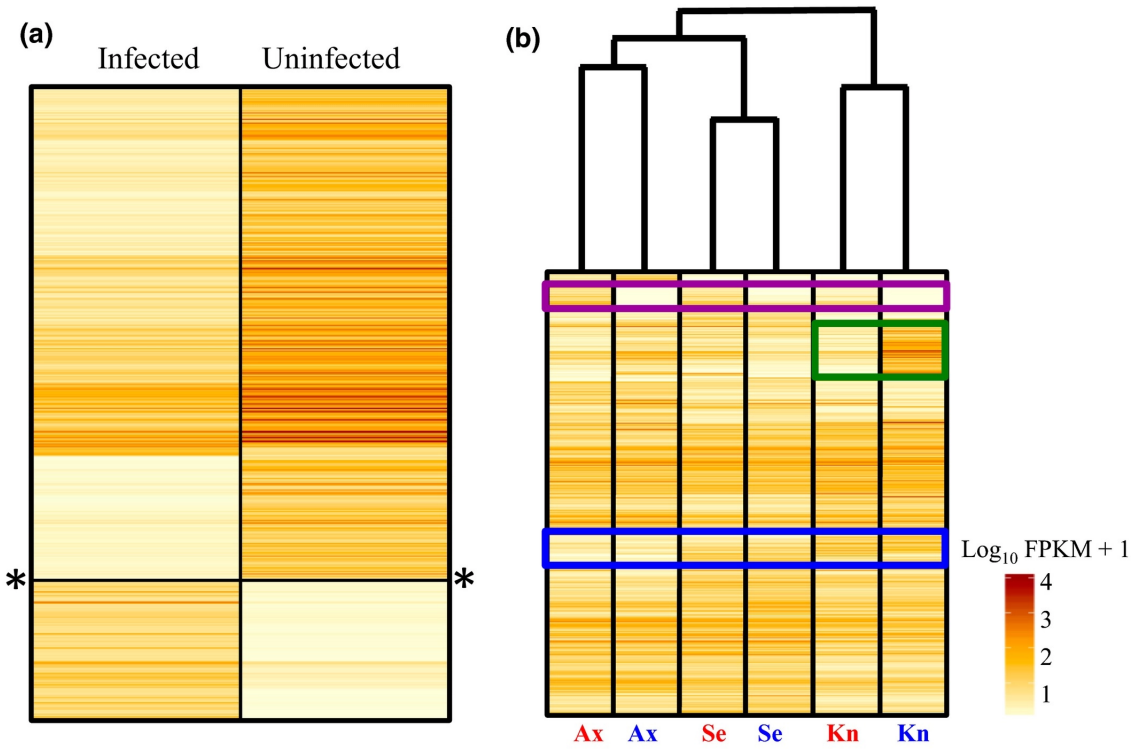
Introduction

[Bankers, et al 2017]

Bankers, L, Fields, P, McElroy, KE, Boore, JL, Logsdon, JM, Neiman, M. Genomic evidence for population-specific responses to co-evolving parasites in a New Zealand freshwater snail. Mol Ecol. 2017; 26: 3663– 3675. https://doi.org/10.1111/mec.14146

> *Co-evolving interactions between hosts and parasites create strong selection that can promote rapid, specific evolution. That said, we still only have a limited knowledge of how coevolution works in host-parasite relationships on a genomic level. Here, Bankers, et al, analyzes gene expression and sequence mutation to bridge that gap in knowledge between Potamopyrgus antipodarum and Microphallus sp., a sterilizing trematode parasite. They found systematic downregulation of many genes in infected P. antipodarum in relation to their uninfected counterparts, but the genes experiencing this change in expression differ depending on source location of each snail population, which is consistent with populaiton-level coevolution turning into population-specific host-parasite interactions.*



*"Heatmaps representing significantly differentially expressed transcripts (higher FPKM = higher expression). Significance was assessed with a FDR of 5% and Benjamini-Hochberg multiple test correction. (a) Inclusive analysis: 1,408 significantly differentially expressed transcripts in infected vs. uninfected snails. Transcripts above the black line demarcated with an astricks are downregulated in infected snails (1,057), and transcripts below this line are upregulated in infected snails (351). (b) Within-lake analysis: dendrogram and heatmap representing the 6,228 significantly differentially expressed transcripts across infection status and lakes. “Ax” = Alexandrina, “Se” = Selfe, “Kn” = Kaniere, red = infected, blue = uninfected. Snails

clustered by lake of origin rather than infection status, indicating spatially structured gene expression patterns. Highlighted regions of heatmap contain examples of: expression differences between infected and uninfected snails within a lake (green), lake-specific expression regardless of infection status (downregulation in Ax, blue) and across-lake upregulation in infected snails (purple)"(Bankers 2017)*

This figure shows the relative expression levels of many genes across populations from differing origins. While panel (a) shows an all-inclusive analysis, panel (b) breaks the analysis down by source lake, highlighting the changes in host-parasite relationship and evolution trajectory. This figure shows that there is a clear downregulation in the expression of certain genes, marked with an asterick, in uninfected snails.

Materials and Methods

Data

I was able to obtain the full transcriptome and the CuffDiff outputs from the author of the paper, which will greatly help my process. The transcriptome is the full transcriptome for Potamopyrgus antipodarum, and it can be found in this project's GitHub repository or on the Neiman Lab website. The CuffDiff output zip folders are available through my project repository only, as it includes unpublished data. The CuffDiff outputs are essentially estimates of the number of fragments of each transcript and gene in the transcriptome. This helps determine the gene count of genes of interest, giving us the ability to track the expression of our target genes. When we compare the CuffDiff output to the reference genome, which is the genome of Potamopyrgus antipodarum found in Lake Alexandrina in New Zealand, we can look for signatures of selection and determine what genes are being upregulated (or downregulated) in infected versus uninfected populations. This process is outlined below.

Here's the shasum value for the transcriptome - 741a6a2e0e337034a5a3957b7fc2088d7e39e459 transcriptome

Methods

Initiallizing the CummeRbund program

First, I need to download R and CummeRbund. From there, I have a record of the exact commands that worked in the previous project. In R, I need to enter the CuffDiff output directory and start the cummeRbund program using the |library()| command. By using the |cuff| command set, I will be able to see the cuffset. I will need to follow this path for both figure panels separately.

Analyzing significantly differentially expressed genes

Next, I will filter down to significantly differentially expressed genes by using the |sig| command set, specifying that I'm looking for 'genes' with a p value of 0.05 or less. To make a heat map of the results, thus creating figure panel A, I will use the |h| and |cdHeatMap| commands and create the map and save it as a png file with specified dimensions. Finally, I will create the dendogram above the heat map by using the |den| command set. A record of the actual command line is saved in the README.md file under data/ in the repository.

Prior Data Acquisition

Thankfully, I already have access to the CuffDiff outputs. If I didn't, I could have used Cufflinks to run the transcriptome file in the CuffDiff program. Cufflinks is commonly used for transcriptome assembly and differential expression analysis for RNA-seq data, and it is freely avialble online, here. I already have the data from this, but if you don't, you could simply follow these commands:

```
cuffdiff [options]* <transcripts.gtf>
<sample1_replicate1.sam[,…,sample1_replicateM.sam]>
<sample2_replicate1.sam[,…,sample2_replicateM.sam]> …
[sampleN.sam_replicate1.sam[,…,sample2_replicateM.sam]]
```

Reflection

Unfortunately, I faced a lot of issues in originally acquiring the data. Originally, I was getting my data from the lab database, but I was accidentally removed, and we ran into a lot of trouble getting my account back. I was thankfully able to recover my data, but once I had the data, I had what I'm assuming were merge conflicts in GitHub. I eventually figured out that I needed to start a new branch, tell Git to ignore certain files, and reconfigure my entire project folder because of corrupt data and conflicts in what was in my local repository versus the shared remote repository. I finally figured out my problem through searching StackOverflow and trying just about everything I found, resigning myself to starting over. From what I can tell, I have solved the issues, and everything is set/ready to use.