HW02\_bkretzle\_EG

Bailey Kretzler

10/18/2021

## Background

*Arcartia tonsa* is a marine copepod species with a global distribution. *A.tonsa* are important members of marine ecosystems, as they are both consumed and consumers. As oceans warm, these marine copepods can serve as indicators of overall ecosystem health and plasticity. The health of *A.tonsa* has implications for the economic viability of many fisheries, making it important to study how the species is impacted by environmental changes. One thousand *A.tonsa* individuals were collected from a coastal ecosystem in Groton, CT and raised for three generations prior to exposure to treatments. Then copepods were experimentally evolved for 20 generations in ambient environments (line - AA), or high temperature high CO2 environments (line - HH) with 4 replicates each. After 20 generations, copepods were reciprocally transplanted to either ambient conditions or high temperature high CO2 conditions for 3 generations, resulting in 4 distinct treatments (AAHH, AAAA, HHHH,HHAA) with 4 replicates. After one (F1) and three (F3) generations, 20 individuals were collected from each replicate for all treatments and used to extract RNA. RNA was sequenced using 150 bp paired end reads on the Illumina NovaSeq 6000 platform for RNAseq. Sequencing was conducted by Novogene who also handled the library preparation.

The goal of the experiment was to a) understand how gene expression is impacted by either the environment in which copepods were evolved or transplanted and b) to assess the impact of time in transplant on gene expression. **I hypothesize that the environment in which copepods were evolved (“line”) will have the strongest impact on gene expression, but that this will degrade the longer copepods are in the transplant environment (“environment”).**

## Bioinformatics Pipeline

Sequencing data was received in the form of .fastq files containing quality scores for the reads. FastQC was used to view this data and decide how to trim it. Files were trimmed using trimmomatic which removed the paired then removed adapters, filtered sequences for a minimum quality score of 20 for forward and reverse reads and a minimum length of 36. Data was again visualized with FastQC to confirm imporvement, after which the transcriptome was assembled using Trinity. Using this transcriptome assembly, reads were then mapped and trancripts were quantified using Salmon. The file generated by Salmon is a count matrix for each transcripts in all samples.

The count file and associated metadata were imported to R and counts per transcript were analyzed using the DESeq2 package in R. The counts matrix was loaded in and counts were rounded. After exploring the data, a DESeq object was created with the model set to design = ~line + environment + line:environment to capture the impact of line, environment, and the interaction between the two on gene expression. To remove genes with low expression, genes with less than 10 reads per sample were omitted. The DESeq model was then created using the DESeq function, and was used to plot the principal component analysis. Two additional models were created to test for the impact of each term in the model on gene expression using the likelihood ratio test (LRT). Specifically the models were design = ~ environment + line and design = ~ line + environment. These models were also used to determine which genes were most significantly impacted by each term in the model. Genes were extracted from the model using the results function, ordered based on adjusted pvalue and filtered to contain only those that were significant. Significant genes could then be graphed and used to determine where the models overlap to plot a venn diagram of genes unique and shared among models.

## Results

For the F1 generation, the average number of reads per gene and sample were 11931 and 18166143 respectively. For the F3 generation the average number of reads per gene and sample were 11221 and 17727744 respectively. When comparing the principal component analysis for both generations, the F3 shows more clustering by environment whereas the F1 shows more clustering by line (figure 1). No genes were shared between F1 and F3 generations when looking at the top five most significant genes for each model. In the F3 generation there is a strong impact of the environment in which copepods were evolved (line) on differential gene expression (figure 2), with 1465 differential expressed genes being unique to the impact of line. When compared to the F1 generation (figure 3), this is nearly a 10 fold increase. There is also a nearly two fold increase from F1 to F3 in differential expressed genes due the transplant environment (figure 2,3). Further, when comparing the F1 to F3 generation, we see a strong decline in genes attributed to the interaction of line and environment. Importantly, 49,54, and 105 genes are consistently differentially expressed due to the copepods’ line, environment or the interaction of the two across the F1 and F3 generations. The directionality and magnitude of the differential expression does, however, vary between the F1 and F3 generations. The significant increases in gene expression both as a result of the transplant environment and the evolved environment indicate an ability to adapt to new environments, but also some genetic memory of the evolved environment.

## Conclusions

From my analysis it is apparent that copepods are capable of adapting to sudden warming events, as indicated by the change in differntially expressed genes as a result of transplant environment from F1 to F3. Additionally, there is some legacy effect of the enviornment in which copepods were experimentally evolved as indicate by the large number of differentially expressed genes in the line category in the F3 generation. The impact of the line on gene expression indicated that 20 generations of experimental evolution does significantly alter gene expression even after transplanting. Further as the model showed significant differences in gene expression between both line and environment, we know that high temperature and high environments do alter the physiology of copepods. Initally we see strong indication of interaction in the F1 generation, however the number of interaction associated DEGs are less prevalent in the F3 generation. All these results combined could indicate adaptive capacity in the expression of some genes, whereas others are less flexible. It is apparent that time in a given environment has an impact on gene expression; in the future it would be useful to understand the expression levels in the generations leading up to transplant as well as in more generations post transplant (e.g. up to F10). It would also be useful to understand how environment is impacting phenotypic traits of copepods like body size (per Rice et.al. 2014), to connect gene expression with the physical nature of the organism.

## References

Rice, Edward, et al. “Impact of Climate Change on Estuarine Zooplankton: Surface Water Warming in Long Island Sound Is Associated with Changes in Copepod Size and Community Structure.” Estuaries and Coasts, vol. 38, no. 1, 2015, pp. 13–23.