Epigenetics responses to excess heat and CO_2 selection in the copepod *Acartia tonsa*

Thomas O'Leary

Background

Acartia tonsa introduction

A world wide distribution and a huge population size make the copepod, *Acartia tonsa*, a particularly interesting species to study in relation to epigenetic changes in response to the environment. *A. tonsa* lives and thrives in a broad range of habitats, including large variation in temperature, salinity, and pH (Chen and Hare). There large population size and ecological importance maekt ehaeljsfkjoie In this experiment, natural collections of *A. tonsa* were subjected to 25 generations of selection under heat and carbon dioxide stress. Methylation of the genome was quantified before and after experimental treatment. Here, I will limit my focus to the effect of heat stress and combined heat and carbon dioxide stress on *Acartia tonsa*. The aims of this study are to (i) describe the epigenetic response to heat stress and (ii) how the additional stressor of carbon dioxide affects that response, and (iii) to identify the genes associated with the loci that are differentially methylated.

Selection experiment, sample collection, library preparation, and sequencing

Natural collections of *Acartia tonsa* were kept in a common garden for three generations, before being split for 25 generations in four separate treatment groups: (i) a control with ambient temperature and ambient CO_2 levels, (ii) a high temperture treatment with ambient CO_2 levels, (iii) a high CO_2 treatment with ambient temperature and, (iv) a combined high temperature and high CO_2 treatment.[1] There were four replicates of each treatment, with approximately 3,000–5,000 individuals per replicate. Whole genomic DNA was extracted from pools of individuals[2] and collected at generation zero (*i.e.* after three generations of common garden conditions) and generation 25. Reduced representation bisulfite sequencing (RRBS) library preparation was used to quantify methlyation frequencies, where all unmethylated cytosines are converted to thymines. The samples were 150 bp sequenced on an Illumina.[3]

Bioinformatics Pipeline

The quality of the raw reads were accessed using FastQC v0.11.7 (Andrews et al.) and low quality sequence data was trimmed using Trimmomatic v0.33 (Bolger et al.). The cleaned reads were vizualized again with FastQC and then mapped to the *A. tonsa* reference genome[4] with Bismark v0.22.1 (Krueger and Andrews). The reference genome needed to be modified from its original version to reflect the RRBS treatment into two versions: one with the cytosines converted to thymines for the forward reads (Read 1) and the other with guanines converted to adenines for the reverse reads (Read 2). Differential methylation was called at and false discovery rate of 0.05 and a minimum relative fold-change of 10%[5] using the package methylKit v1.13.1 (Akalin et al.) R v3.6.1 (R Core Team). The methylation data was pre-filtered to include only loci that had at least 20 reads and to exculed loci in the highest 2.5% of per base read coverage[6]. Other analysis and visualization was carried out in R using ggplot2 and dplyr v1.2.1 (Wickham).

[1] Absolute or relative temperature or CO_2 levels in each treatment not stated.

[2] Not sure what the approximate number of individuals pooled for each extraction

[3] Exact type of Illumina machine unknown (*e.g.* HiSeq X).

[4] Unknown *Acartia tonsa* reference genome.

[5] We required a minimum relative fold-change of 10% for differential methylation in order to exclude statistically significant but minor relative fold-changes that may have little-to-no biological significance.

[6] The pre-filtering was done to ensure that there was enough data to allow for an accurate estimate of the percent methylation (low-end limit) and remove data that might be the product of procedural sequencing error (high-end limit).

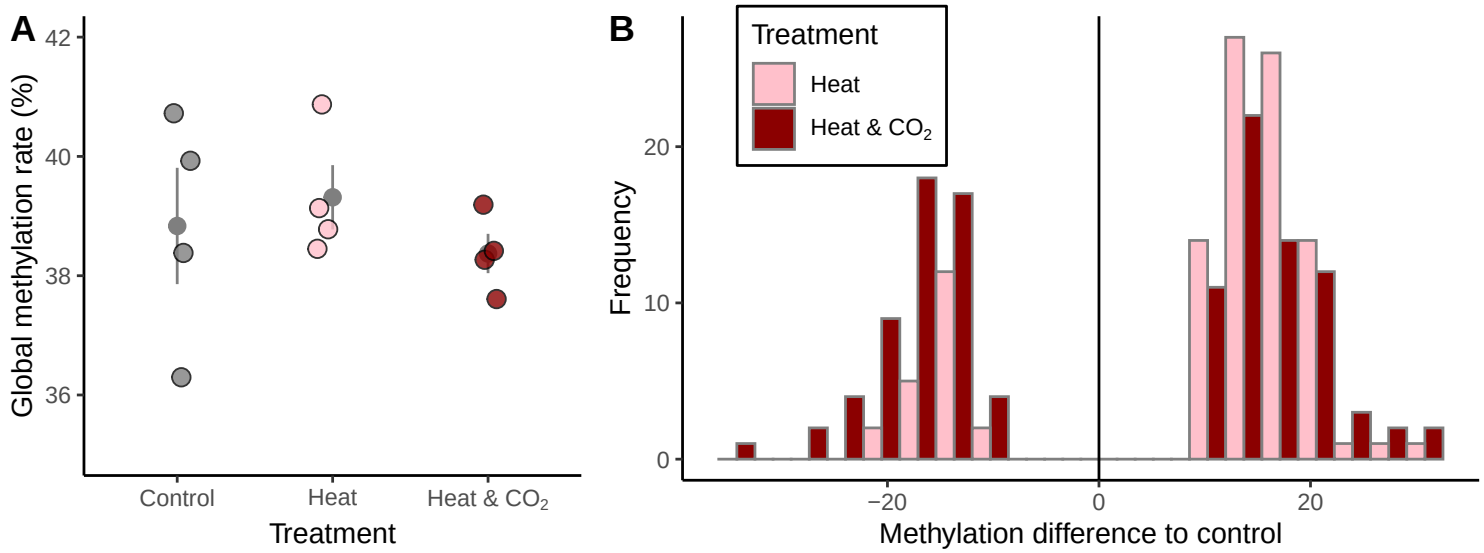

Figure 1: **(A)** Mean per site methylation rate (%) across all sites included in the analysis. **(B)** Frequency histogram of the percent difference in methylation of only the differenitally methylated loci (DML) for each treatment, Heat (Pink) and Heat & CO_2 (Dark Red), compared to the control.

Results

The average mapping rate across all samples was 45.2%, with averages for the subset of groups I analyzed of 43.2%, 33.9%, and 31.0% for the generation 25 control, high temperature only, and combined high temperature and high carbon dioxide groups respectively. The global methylation rate was not drastically different between the three treatment groups after 25 generations, with a mean per site methylation rate of all three groups between 38% and 39% (Fig. 1 **A**). However, there were still individual loci that showed differential methylation between the experimental treatments and the control. We identified 105 differentially methylated loci (DML) in response to heat stress, with 84 hypermethylated loci and 21 hypomethylated loci (Fig. 1 **B**). Under combined heat stress and excess CO_2, 121 DML were idenified, with 66 of those hypermethylated loci and 55 hypomethylated loci (Fig. 1 **B**). There was no overlap between the lists of DML. Those DML mapped to 9 unique differentially methylated genes (DMGs) for the heat stress and 13 DMGs for combined the heat and CO_2 stress, of which 5 DMGs were overlapping between the groups.

Conclusion

The number of identified loci

It should be noted that here we identify DMGs only with respect to the pairwise comparisons of high temperature and combined high temperature and high CO_2 to the control treatment. This approach does not include the full model of treatments and time points. Pairwise comparisons used in this way, ignoring the full experimental design, may lead to an increase in false positives, so care must be taken when interpreting the lists of DML. Futher complicating this analysis is the fact that this data was taken from pooled individuals and all tissue types, creating greater noise in the data and potentially masking important individual and tissue type epigenetic variation, a drawback of the study system. Future directions must involve incorporating transcriptomic and genomic sequencing data with this epigenetic data. Despite finding several DMGs, it is not know how that may or may not affect the expression of a gene without the corresponding transcriptomic data set. Along the same lines, the epigenetic and transcriptomic data together do not tell the full story either. Adaptive allele frequency changes in regulatory and coding regions may result in differential gene expression.

To understand if their really has been adaptive change in these experimental pools of *A. tonsa*, you must measure certain fitness related phenotpyes, like survival, fecundity, and growth. And do a reciprocal transplant to see if the control pools fair worse under those same environmental stressors.

References

Akalin, Altuna, et al. "MethylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles." *Genome Biology*, vol. 13, no. 10, 2012, doi:10.1186/gb-2012-13-10-R87.

Andrews, Simon, et al. *FastQC*. Babraham Institute, Jan. 2012, https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Bolger, Anthony M., et al. "Trimmomatic: A flexible trimmer for Illumina sequence data." *Bioinformatics*, vol. 30, no. 15, 2014, pp. 2114–20, doi:10.1093/bioinformatics/btu170.

Chen, Gang, and Matthew P. Hare. "Cryptic diversity and comparative phylogeography of the estuarine copepod Acartia tonsa on the US Atlantic coast." *Molecular Ecology*, vol. 20, no. 11, 2011, pp. 2425–41, doi:10.1111/j.1365-294X.2011.05079.x.

Krueger, Felix, and Simon R. Andrews. "Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications." *Bioinformatics*, vol. 27, no. 11, 2011, pp. 1571–72, doi:10.1093/bioinformatics/btr167.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, 2019, https://www.r-project.org/.

Wickham, Hadley. *tidyverse: Easily Install and Load the 'Tidyverse'*. 2017, https://cran.r-project.org/package=tidyverse.