The epigenetic response of the copepod *Acartia tonsa* to excess heat and CO_2 selection

Thomas O'Leary

Background

Acartia tonsa introduction

A worldwide 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). Expected increases in sea surface temperature and decreases in pH (from absorbed atmospheric carbon dioxide) driven by climate change make *A. tonsa* a particularly interesting species to study in the context of heat and carbon dioxide stress (Peck et al.). In this experiment, natural collections of *A. tonsa* were subjected to 25 generations of selection under heat and CO_2 stress treatments. 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 *A. tonsa* after 25 generations of selection. The aims of this study are to (i) describe the epigenetic response to heat stress and (ii) describe 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 (18°C) and ambient CO_2 levels (400 ppm), (ii) a high temperature treatment (22°C) with ambient CO_2 levels, (iii) a high CO_2 (2000 ppm) treatment with ambient temperature and, (iv) a combined high temperature and high CO_2 treatment. There were four replicates of each treatment, with approximately 3,000–5,000 individuals per replicate. Whole genomic DNA was extracted from pools of ~30 individuals and collected at generation zero and generation 25. Reduced representation bisulfite sequencing (RRBS) library preparation was used to quantify methylation frequencies, where all unmethylated cytosines are converted to thymines. The samples were 150 bp sequenced on an Illumina HiSeq X.

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 visualized again with FastQC and then mapped to the *A. tonsa* reference genome (Jørgensen et al.) 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%[1] using the package methylKit v1.13.1 (Akalin et al.) in 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 exclude loci in the highest 2.5% of per base read coverage[2]. All other analysis and visualization was carried out in R using ggplot2 and dplyr v1.2.1 (Wickham).

Results

The mean mapping rate across all samples was 45.2%, with group means for the subsets I analyzed of 43.2%, 33.9%, and 31.0% for the generation 25 control, high temperature only, and combined high temperature and high CO_2 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 identified, with 66 hypermethylated loci and 55 hypomethylated loci (Fig. 1 **B**). The total number of identified DML was only slightly higher in the

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

[2] 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).

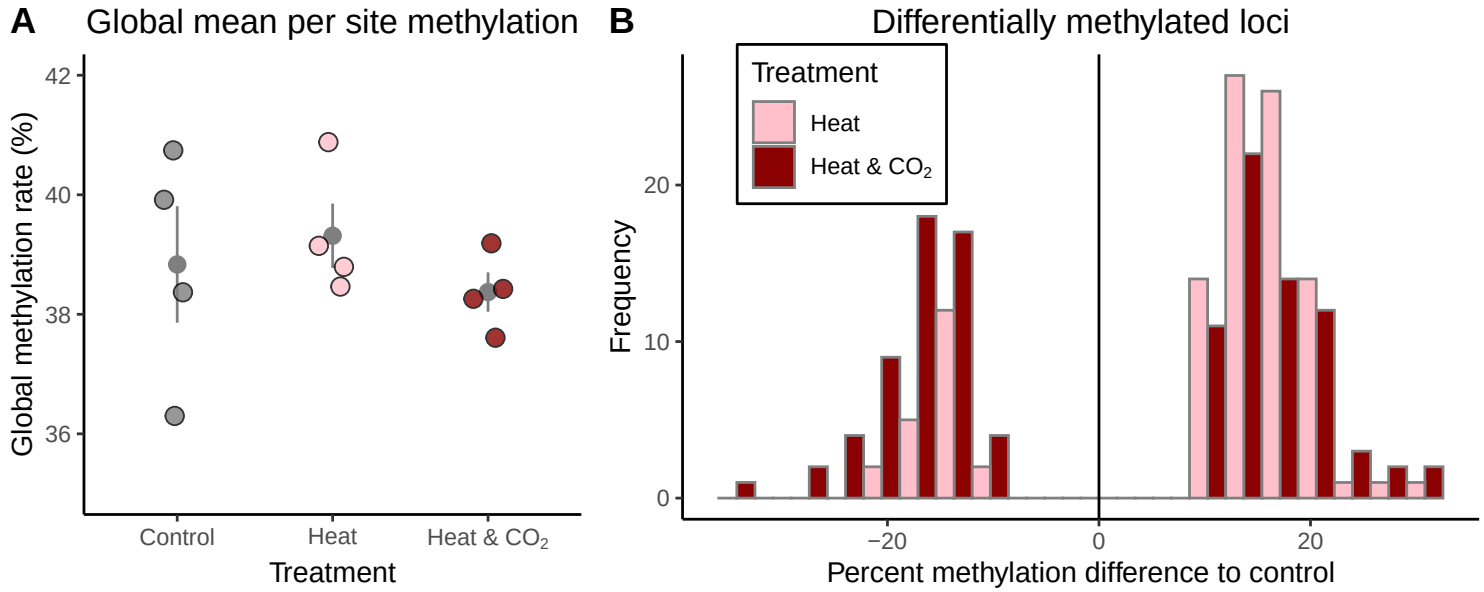

Figure 1: **(A)** Mean per site methylation rate (%) across all sites for all three treatments included in the analysis. **(B)** Frequency histogram of the percent difference in methylation of only the differenitally methylated loci (DML) for each treatment, Heat (pink) and Heat & CO_2 (dark red), compared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

combined Heat & CO_2 treatment, but there is a significant difference in the distribution of the two treatments (Kolmogorov-Smirnov test: $p = 7.1 \times 10^{-4}$), with the heat stress only treatment having disproportionately more hypermethylated loci compared to the relatively even distribution of hyper-/hypomethylated loci in the combined heat & excess CO_2 treatment (Fig 1. **B**).

There was no overlap between the DML identified in each treatment. 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 five DMGs were overlapping between the groups. The gene ontology of the identified DMGs indicate that these genes are involved in a variety of biological processes. Three of the five shared DMGs, Retrovirus-related Pol polyprotein from transposon opus, Gypsy retrotransposon integrase-like protein 1, and Endogenous retrovirus group K member 5 Gag polyprotein are involved in biological processes such as DNA integration, recombination, and RNA mediated transposition.

Conclusion

In as little as 25 generations of experimental selection, *A. tonsa* seem to be altering their methylome in response to environmental heat and CO_2 stress. Although there is little difference in the global methylation rate (Fig. 1 **A**), there are individual loci responding to the environment with a slightly greater methylation response under the combined heat & CO_2 stress than just heat alone (Fig. 2 **B**). The different distributions of methylation patterns perhaps reflects a more concerted hypermethylation response under heat-only stress. What this means biologically depends largely on the specific genes and loci being differentially methylated. The shared gene ontology categories of DNA integration, recombination, and RNA mediated transposition could reflect a targeted trend of differential methylation of genes that contribute to increasing genetic diversity in the population through recombination and transposition. However, more work would need to be done to understand the role of the specific genes and their effect on the genome in *A. tonsa* to be able to speculate with any confidence.

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 generations. 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 identified DML. One further possible complication of this analysis is the fact that this data was taken from several pooled whole-bodied individuals, leading to greater noise and potentially masking important individual and tissue type epigenetic variation. Future directions should 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, even the epigenetic and transcriptomic data together do not tell the full story. Adaptive allele frequency changes in regulatory and coding regions may result in differential gene expression. And finally, to understand if these identified changes are truly adaptive to the environment, we must measure certain fitness related phenotypes, like survival, fecundity, and growth and do reciprocal transplants 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.

Jørgensen, Tue Sparholt, et al. "The genome and mRNA transcriptome of the cosmopolitan calanoid copepod acartia tonsa dana improve the understanding of copepod genome size evolution." *Genome Biology and Evolution*, vol. 11, no. 5, 2019, pp. 1440–50, doi:10.1093/gbe/evz067.

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.

Peck, Nadine, et al. "Interactive effects of temperature and salinity on population dynamics of the calanoid copepod Acartia tonsa." *Journal of Plankton Research*, vol. 37, no. 1, 2014, pp. 197–210, doi:10.1093/plankt/fbu093.

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.