Transcriptomics of experimental evolution in cold-adapted copepods

Background

Copepods are highly abundant marine animals that form the basis of the ocean food chain. Many marine ecosystem functions depend upon copepods as both a major source of food for forage fish and as a driver of spatiotemporal variability in carbon and nutrient cycling through their consumption of phytoplankton (Steinberg & Landry, 2017). The cold-adapted copepod species *Acartia hudsonica* is found primarily along the New England coast and in estuaries that flow into the Atlantic Ocean (deMayo, *et al.*, 2023). It is unclear what capacity this keystone species may have to adapt to rapid ocean warming and acidification in the coming decades, but the downstream ecological consequences of a mass die-off would no doubt be severe. In this study, we examined differential gene expression in *A. hudsonica* subjected to experimentally induced ocean warming and acidification in the lab over the course of 11 generations. Our ultimate aim is to evaluate the extent to which genetic adaptation and phenotypic plasticity will allow *A. hudsonica* and other ecologically foundational copepods to persist in a changing climate, and what level of fitness costs they may incur. Previous work has begun to explore multigenerational adaptation to warming and acidification in copepods, but the combined effect of these stressors remains poorly understood (Langer, *et al.*, 2019).

We collected samples from Long Island Sound and waited three generations before introducing the experimental treatments (OA = acidification, OW = warming, OWA = both, AM = ambient conditions). In each treatment, we used three replicate vessels, containing ~4000 individuals each. 50 adults were sampled from the F0, F2, F4, and F11 generations, and RNA was extracted using a modified TRIzol extraction protocol. Libraries were prepared at Novogene under standard TruSeq3 protocols, and sequenced on the Illumina Novaseq 6000 platform, producing paired-end 150 bp reads for downstream bioinformatics analyses, with a total of 38 samples (generation x treatment combinations) and >6Gb per sample. Additionally, to explore the effects of model parameter choice on our final results, we filtered transcripts by several different read depth criteria and compared the output of downstream analyses for each filter.

Bioinformatics Pipeline

We assessed quality scores of the raw reads using 'fastp' (Chen, 2023), trimmed the reads, and filtered those that were too short (<35 bp) or too low-quality (QC <20). Average reads per sample before filtering was 45.3M (range: 37.9-55.9M) and was then reduced to 44.5M (37.3-54.9M) after filtering, well within acceptable bounds for diploid eukaryotes. On average, 98.1% of reads passed this step. A reference transcriptome was generated using the Trinity *de novo* assembler v2.13.2 (Grabherr, *et al.*, 2011), and quality of the reference was assessed with BUSCO v5.2.2 (Manni, *et al.*, 2021), which found a complete score of 96.9%. Reads were mapped to the reference transcriptome with 'Salmon' (Patro, *et al.*, 2017). Mapping rates for all samples were 91-96%, with the exception of one replicate for the F11 AM treatment, which only achieved a 68% mapping rate.

We analyzed the mapped data with the R package 'DESeq2' (Love, *et al.*, 2014), which uses negative binomial generalized linear models to statistically test for differential gene expression. DESeq2 automatically reduces multiple testing by eliminating genes that don't have a sufficient number of reads to identify differential expression. However, a pre-filtering step can be useful for reducing the size of the dataset, increasing modeling speed, and simplifying data visualizations (Love, *et al.*, 2023). Recommended read depth filtering levels remove genes with

counts < 15 in more than 75% of samples. To explore the effects of this parameter choice on the results and their biological interpretation, we repeated the gene expression analyses with four additional filter settings: counts < 5, 10, 25, and 35 in 75% of samples (Table 1).

Results

We found significant divergence in gene expression between the four experimental conditions, with reasonable consistency between replicates, suggesting distinctive responses to the differing environmental conditions. After transforming the data with variance stabilization for better plotting, we conducted a PCA of gene expression patterns for the different generations and conditions (Fig. 1). In F0, we see clustering of the points by treatment group, whereas they occupy overlapping PCA space in generations F2 and F4. In F11, the OWA treatment group is separate from the AM control group, but these results are difficult to interpret without F11 data for the OW and OA groups. Using the baseline DESeq2 filtering threshold of <15, the model identified a total of 6,808 genes that were significantly differentially expressed between treatment groups. As expected, number of DEGs decreased as more stringent filtering thresholds were applied, ranging from 7,816 DEGs with the least stringent criterion to 5,898 with the most stringent criterion, a decrease of approximately 25%. Yet, despite the variation in total number of DEGs, the relative proportions of genes differentially expressed between the treatment groups was constant across all filtering thresholds, as seen in Fig. 2. The OW condition, both in isolation and in interaction with the other treatments, showed the only consistent drop in DEGs.

We created a heatmap of the 100 most differentially expressed genes (Fig. 3), and we primarily see genes downregulated in ambient conditions and upregulated in response to warming and/or acidification. OW showed the highest differential upregulation in comparison to the other groups, followed by OWA. However, the heatmap does identify several genes with comparatively unusual expression patterns that would be interesting candidates for further functional exploration. When looking at relative degrees of expression, there are in fact significant differences depending on which filtering threshold is used, potentially leading to different biological conclusions.

Conclusions

In this study, we have demonstrated differential transcriptomic responses to ocean warming, acidification, and a combination of both. We have identified candidate genes that are highly differentially expressed across one or more conditions, the exact functions of which should be investigated, as they may be the basis for climate adaptation in this species. We have also shown a notable change in gene expression patterns over the course of multiple generations of experimental evolution. Differences among the F0 generation can likely be attributed to plasticity, while differences across generations are more likely to indicate genetic adaptation via selection. Taken together, the results suggest that *A. hudsonica* may have considerable plastic and evolutionarily adaptive capabilities that will help it persist through changing climate conditions. One particularly notable drawback of this study is that due to logistical data collection issues, we do not have data for the OWA condition for the F2 generation, nor do we have data for OA and OW for F11. Additionally, we found that gene expression patterns were significantly impacted by the different parameters chosen for the DESeq2 model. Future work should therefore either test a suite of parameter values and verify robustness of conclusions or be able to justify choosing a particular read depth threshold with reference to the biology of the organism or the particulars of the sequencing method employed.

Figures

Filter setting	Counts
Before filtering	67916
Counts < 5 in 75% of samples	33252
Counts < 10 in 75% of samples	28196
Counts < 15 in 75% of samples	25260
Counts < 25 in 75% of samples	21763
Counts < 35 in 75% of samples	19571

Table 1. Number of genes in dataset before and after different filters were applied prior to analyzing differential gene expression.

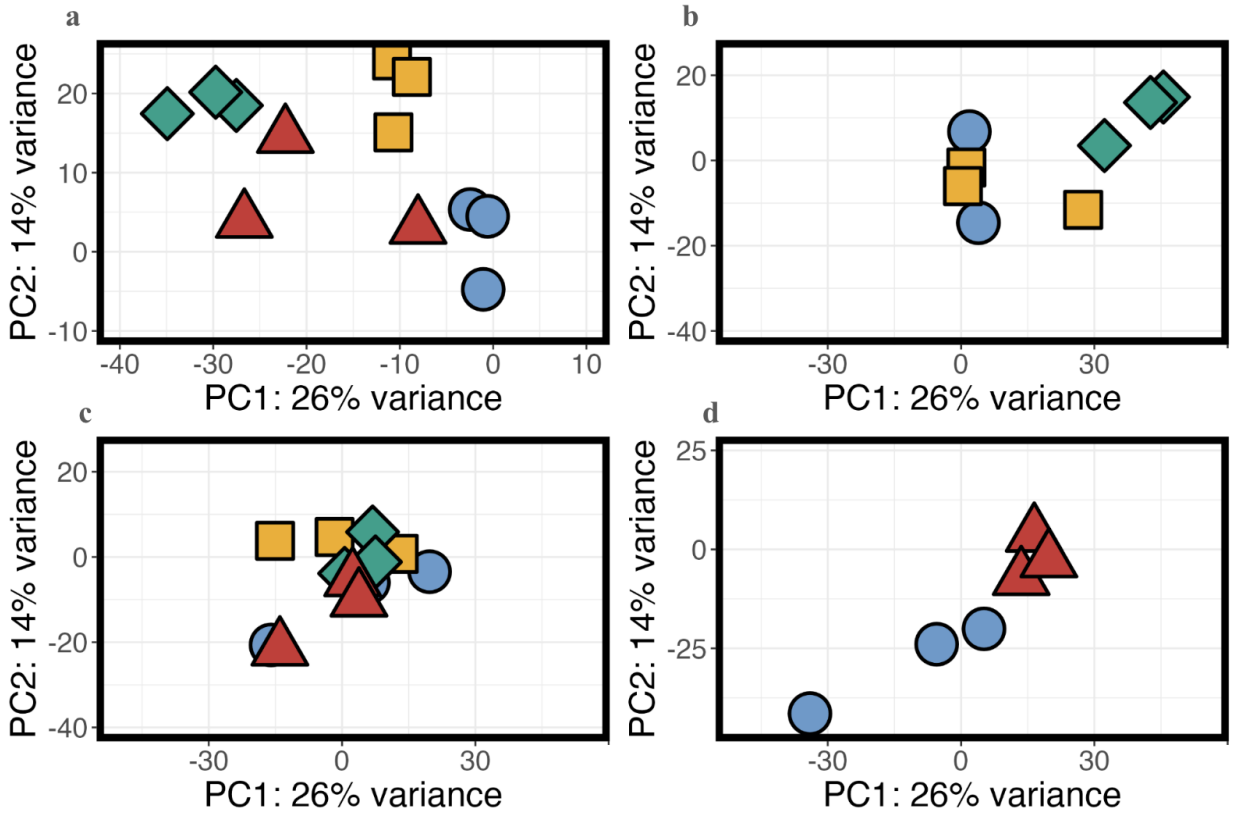

Figure 1. PC1 and PC2 from PCA of global gene expression patterns. (a) F0 generation, (b) F2 generation, (c) F4 generation, (d) F11 generation. Blue circles = ambient conditions, yellow squares = ocean acidification, green diamonds = ocean warming, red triangles = ocean warming and acidification.

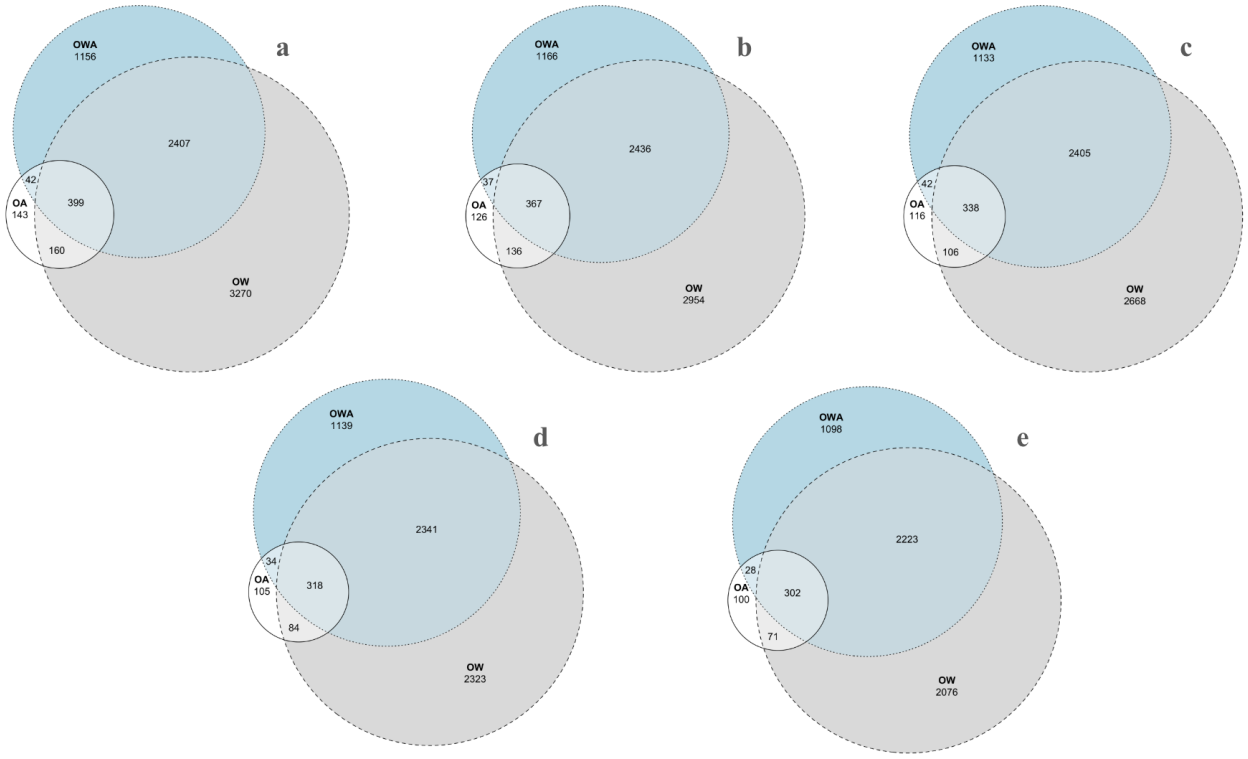

Figure 2. Venn Euler diagrams of overlapping differentially expressed gene counts for the three treatment conditions in the F0 generation: OA = ocean acidification, OW = ocean warming, OWA = ocean warming and acidification. Diagrams depict results of DEG analysis after application of different read depth filters. Genes were removed if they had less than (a) 5, (b) 10, (c) 15, (d) 25, or (e) 35 reads in more than 75% of samples. Numbers of DEGs for each condition changes, but relative degrees of overlap are unaffected.

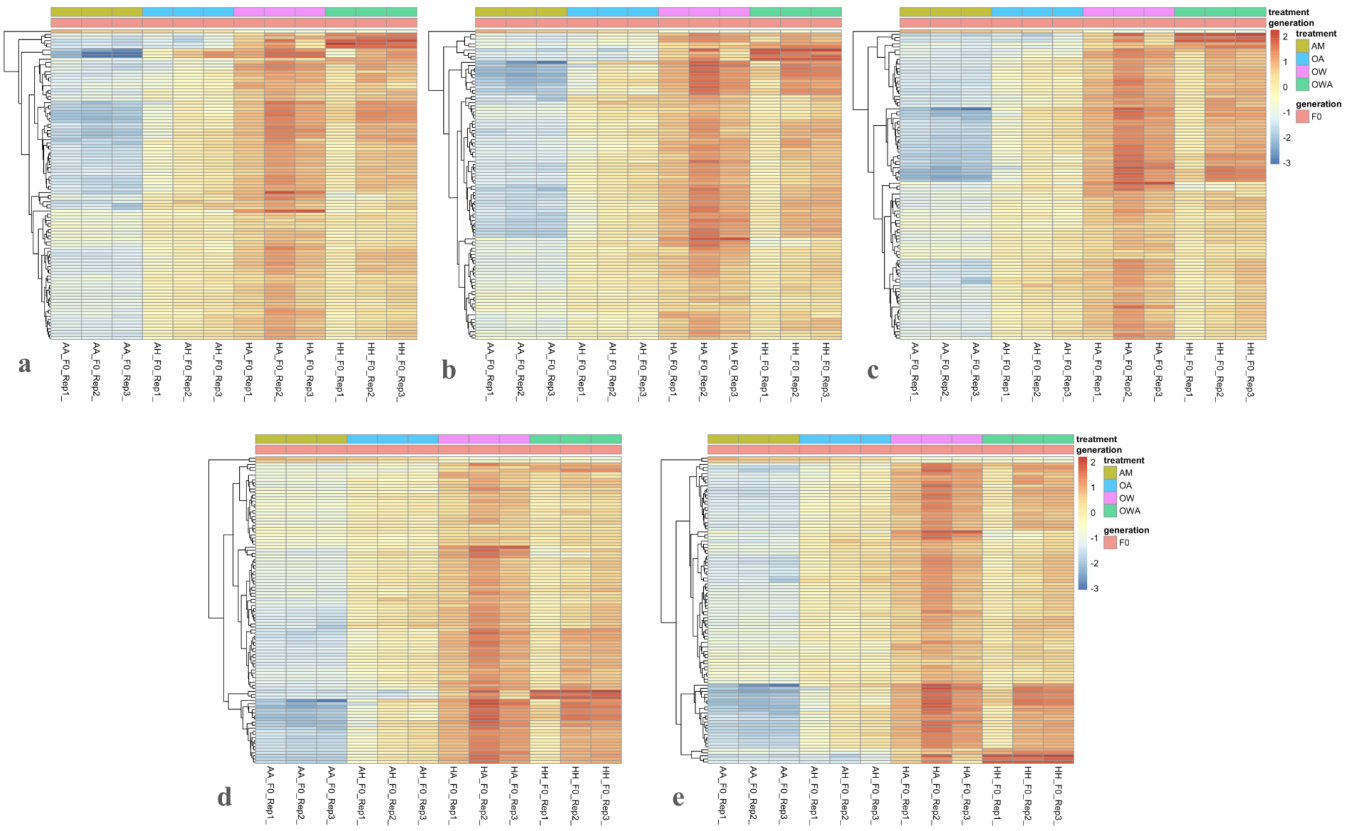

Figure 3. Heat maps of differential gene expression for the F0 generation across the four experimental conditions: AM = ambient (control), OA = ocean acidification, OW = ocean warming, OWA = ocean warming and acidification. Each column represents one replicate.

References

Chen, S. (2023). Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *iMeta 2*(e107). https://doi.org/10.1002/imt2.107.

deMayo, J.A., Brennan, R.S., Pespeni, M.H., Finiguerra, M., Norton, L., Park, G., Baumann, H., Dam, H.G. (2023). Simultaneous warming and acidification limit population fitness and reveal phenotype costs for a marine copepod. *Proceedings of the Royal Society B, 290*(2006). https://doi.org/10.1098/rspb.2023.1033.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., *et al*. (2011). Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nature Biotechnology, 29*(7), 644-52. https://doi.org/10.1038/nbt.1883.

Langer, J.A.F., Meunier, C.L., Ecker, U., Horn, H.G. (2019). Acclimation and adaptation of the coastal calanoid copepod *Acartia tonsa* to ocean acidification: a long-term laboratory investigation. *Marine Ecology Progress Series, 619*, 35-51. https://doi.org/10.3354/meps12950.

Love, M.I., Anders, S., Huber, H. (2023). Analyzing RNA-seq data with DESeq2. Bioconductor. https://bioconductor.org/packages/release/bioc/vignettes/DESeq2/inst/doc/DESeq2.html.

Love, M.I., Huber, W., Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology, 15*, 550. https://doi.org/10.1186/s13059-014-0550-8.

Manni, M., Berkeley, M.R., Seppey, M., Simão, F.A., Zdobnov, E.M. (2021). BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Molecular Biology and Evolution, 38*(10), 4647–4654. https://doi.org/10.1093/molbev/msab199.

Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods, 14*, 417-419. https://doi.org/10.1038/nmeth.4197.

Steinberg, D.K. & Landry, M.R. (2017). Zooplankton and the Ocean Carbon Cycle. *Annual Review of Marine Science, 9*, 413-444. https://doi.org/10.1146/annurev-marine-010814-015924.

Supplemental: Data, scripts, and results available at https://github.com/noraheaphy/ecological_genomics_2023/tree/main