### *Supplemental Materials for Molecular indicators of warming and other climate stressors in larval Pacific cod*

Laura H Spencer1,2, Emily Slesinger3, Ingrid Spies1, Benjamin Laurel3, Thomas Hurst3

1. Resource Ecology and Fisheries Management Division, Alaska Fisheries Science Center, 7600 Sand Point Way NE, Seattle, WA 98115, USA
2. University of Washington School of Aquatic and Fishery Sciences, 1122 NE Boat St, Seattle, WA 98195
3. Fisheries Behavioral Ecology Program, Alaska Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 3020 SE Marine Science Dr, Newport, OR 97365, USA

### **Detailed Genomics Methods & Results**

#### **RNA Extraction, Sequencing, and Data Processing**

RNAlater-preserved larval fish were homogenized and then RNA was isolated following the TRIzol™ Reagent protocol for total RNA (Invitrogen Inc., Carlsbad, CA) which included cell lysis, impurities removal, inhibition of RNAse activity, and total RNA extraction using the phase separation method from cell debris. A Bioanalyzer 2100 (Agilent, Santa Ana, CA) was used to check RNA integrity and concentration. Non-directional RNA-Seq libraries were prepared with the NEBNext® Ultra II RNA Library Prep Kit for Illumina (New England BioLabs Inc., Ipswich, MA), and paired-end sequencing was conducted on one lane of a NovaSeq 6000 Sequencing System (Illumina, Inc., San Diego, CA) with 150 bp read length.

#### **Bioinformatics**

Raw sequence data were demultiplexed and trimmed using Cutadapt v3.5 [(Martin, 2011)](https://paperpile.com/c/n1vINc/CDU3E) to remove Illumina adapters, flanking N bases, reads less than 50 bp, and low-quality ends from reads using minimum quality scores of 15 and 10 for the 5’ and 3’ ends, respectively. Raw and trimmed data were inspected using FastQC [(Andrews, 2010)](https://paperpile.com/c/n1vINc/Dw2V) and MultiQC [(Ewels et al., 2016)](https://paperpile.com/c/n1vINc/QtTz). Reads were aligned to the Atlantic cod (*Gadus morhua*) genome assembly v3 (gadMor3.0, Genbank accession GCA\_902167405.1) using STAR v2.7.10a [(Dobin et al., 2013)](https://paperpile.com/c/n1vINc/OdOxH). The number of fragments aligning to gene coding regions of the *G. morhua* genome was quantified using featureCounts v2.0.3 [(Liao et al., 2014)](https://paperpile.com/c/n1vINc/Hic0) with settings -p --countReadPairs to count paired-end fragments, -t exon to only count reads overlapping with exons, and -C and -B to exclude chimeras and singletons, respectively. Multimapped reads were also excluded from gene counts. Gene functions were identified by querying coding sequences of the *G. morhua* genome, derived from gene annotations published along with the *G. morhua* genome, against the Uniprot/Swissprot database [(UniProt Consortium, 2021)](https://paperpile.com/c/n1vINc/DS4C) using blastx from blast v2.11.0 (e-value < 1–10) [(Camacho et al., 2009)](https://paperpile.com/c/n1vINc/hMth).

#### **Gene expression analysis**

Gene expression analyses were performed in R v4.1.2 using RStudio interface v2021.09.1 [(R Core Team, 2021; RStudio Team, 2020)](https://paperpile.com/c/n1vINc/yzhr+hiGD). Unless otherwise specified, significance thresholds were alpha = 0.05 and representations of spread in data were one standard deviation. Gene counts were filtered to remove outlier samples and low-frequency genes. Outlier samples were identified using principal component analysis (PCA), which was performed on variance-stabilizing transformed counts of the top 500 genes using vsd and plotPCA from *DESeq2* v1.34.0 [(Love et al., 2014)](https://paperpile.com/c/n1vINc/oa9J). Genes with mean count <10 across all samples or those with counts <30 across at minimum 10% of the samples were discarded, and differences in the number of remaining fragments per sample among treatments was tested using ANOVA.

#### **Differential gene expression analysis**

Differentially expressed genes (DEGs) among treatments were identified using *DESeq2* with default settings [(Love et al., 2014)](https://paperpile.com/c/n1vINc/oa9J+nW0B). DEGs were determined among the control (6°C-360 µatm) and five experimental treatments (Table 2). Outliers were handled using *DESeq2*’s built-in replacement method, where Cook's Distance was used to identify influential outliers and then the original count values were replaced with trimmed mean, and using an additional iterative Leave-One-Out (iLOO) approach to identify and remove DEGs with outlier samples (n=887 genes) [(George et al., 2015)](https://paperpile.com/c/n1vINc/l39EZ). To control for tank effects, we removed genes from the DEG lists that were also differentially expressed among tanks within treatments (n=126 genes) using alpha=0.01 to ensure sizable tank effects.

#### **Functional Analyses**

Gene sets of interest were characterized by Uniprot Keyword and Gene Ontology (GO) enrichment analyses. Two enrichment analyses were performed for each of the five pairwise treatment contrasts to determine the functions of genes that were upregulated (L2FC > 0.5) and downregulated (L2FC < -0.5) in response to experimental treatments. For all gene sets, genes were filtered for those that map to the Uniprot/Swissprot database [(UniProt Consortium, 2021)](https://paperpile.com/c/n1vINc/DS4C), and enriched Uniprot Keywords and GO terms were identified by entering UniprotID’s into the Gene-Enrichment and Functional Annotation Tool from DAVID v2021 [(Sherman et al., 2022)](https://paperpile.com/c/n1vINc/ZiDBe) to identify enriched biological processes, which were defined as those with at minimum three genes contributing to each process, and modified Fisher Exact *p*-values (EASE Scores) <0.05.

#### **RNA-Seq Data Processing Results**

A total of ~4.53B reads remained after discarding ~1.0M (0.02%) during initial quality-filtering, with a per-sample mean of 60.4M ±8.3M reads passing filters. On average 49.0M ±6.8M reads uniquely aligned to the *G. morhua* genome for each sample (81.1% ±1.8% alignment rate), and 45.4M ±6.4M were assigned to gene-coding regions (75.2%±2.1). Initial examination of gene counts using PCA identified one sample from each of the control and acidifiedtreatments as outliers, which were removed from the dataset and resulted in 11-14 samples for each treatment (Table 1). In total, we detected 32,484 genes in the *G. morhua* genome, but after removing low frequency genes (totaling 0.03% of fragments), 21,076 genes remained, 19,424 of which mapped to genes in the Uniprot/Swissprot database (e-value<1e-10). Two-way ANOVA indicated that the number of aligned fragments retained for analyses did not differ among temperature (F(2,67)=1.18, *p*-value=0.31), or pCO2 (F(2,67)=1.40, *p*-value=0.24), nor was there significant interaction (F(2,67)=1.18, *p*-value=0.18). After removing 6,664 genes that were associated with fish length, 14,412 remained in the final dataset that was used for comparative analyses. An additional 710 genes were removed during the differential expression analysis as they were strongly influenced by outliers (n=606) and by tank-specific expression (n=104).

#### **Additional References**

[Andrews, S. (2010). A quality control tool for high throughput sequence data [Online].](http://paperpile.com/b/n1vINc/Dw2V) <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. and Madden, T. L. (2009). BLAST+: architecture and applications. BMC Bioinformatics 10, 421.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21.

Ewels, P., Magnusson, M., Lundin, S. and Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 32, 3047–3048.

Liao, Y., Smyth, G. K. and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923–930.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17, 10–12.

##### 

##### 

##### Figure S1: Biplots of PC1 against PC3, PC4, PC5, and PC6 from the PCA constructed from all analyzed genes excluding those associated with fish length (n=14,412). Points are color-coded by treatment. Y-axis labels show the percent variation explained by each PC.

##### 

##### Figure S2: Comparison of the number of differentially expressed genes (DEGs) that responded to cold temperature (3°C), warm temperature (10°C), and high pCO2 (1,560µatm) as single stressors (TOP) and combined stressors (BOTTOM) relative to control conditions (6°C, 360µatm), highlighting the overlap among each stressor. In parenthesis are the percent of all analyzed genes that differed in response to each single stressor.

##### 

##### Table S1: Mean+sd of abs(LFC) of genes differentially expressed in treatments compared to control conditions.

|  |  |  |
| --- | --- | --- |
| Treatment | Mean abs(LFC), upregulated | Mean abs(LFC), downregulated |
| Acidification | 0.383 +/- 0.31 | 0.382 +/- 0.195 |
| Warm | 0.526 +/- 0.300 | 0.413 +/- 0.268 |
| Warm+Acidification | 0.487 +/- 0.291 | 0.399 +/- 0.243 |
| Cold | 0.505 +/- 0.83 | 0.359 +/- 0.228 |
| Cold+Acidification | 0.577 +/- 1.018 | 0.387 +/- 0.254 |

##### Figure S3: Enriched Gene Ontology biological processes in genes that were expressed at lower levels in response to temperature and pCO2 stressors alone (Cold=3°C, Warm=10°C, OA=pCO2 1,560μatm), and when combined. All responses are compared to the control conditions (6°C, pCO2 360μatm). Point size indicates the mean effect size (fold change) of differentially expressed genes involved in each process.

##### 

##### Figure S4: Enriched Gene Ontology biological processes in genes that were expressed at higher levels in response to temperature and pCO2 stressors alone (Cold=3°C, Warm=10°C, OA=pCO2 1,560μatm), and when combined. All responses are compared to the control conditions (6°C, pCO2 360μatm). Point size indicates the mean effect size (fold change) of differentially expressed genes involved in each process.

##### 

##### 

##### Table S2: Enrichment results and gene-specific details. *See separate Excel spreadsheet; or* [Enrichment Results spreadsheet](https://docs.google.com/spreadsheets/d/1A-00rgU401KHrx807ViII7xFMB1TutEBPrg2z6jH4IA/edit?usp=sharing).

##### https://lh7-rt.googleusercontent.com/docsz/AD_4nXeG0A2S3K4ZQzDpjY76Zi9LOw9gczYNn-J_h0Dr3VYqCjMn4bOZJNJs-w6zYRT1dgC0dQOtUAaMawPYu3W_TH_an_lgohihGopbMyWlxNCo0ysb3yRE-rotv6o3SCoUaP9D8WaJytMs86J8_GI5cuA?key=y-RkjHAFGdFY8mvBT9Vgqno4Figure S5.

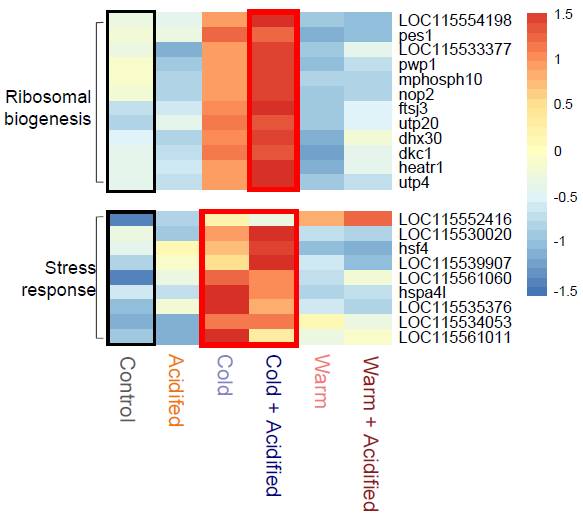
Expression of differentially expressed genes (rows = genes) involved in lipid metabolism, immune function, and vision, which were affected by warming (outlined in red) compared to control (outlined in black). The red-blue gradient indicates the z-score of expression values (treatment means) standardized across treatments for each gene, where red and blue indicate higher and lower mean expression, respectively. Gene IDs refer to the gadMor3.0 assembly.

##### 

##### Figure S6: Heat map showing expression of cell adhesion genes (rows = genes) that were differentially expressed in warm treatments (outlined in red) compared to control (outlined in black). The red-blue gradient indicates the z-score of expression values (treatment means) standardized across treatments for each gene, where red and blue indicate higher and lower mean expression, respectively. Gene IDs refer to the gadMor3.0 assembly.

##### https://lh7-rt.googleusercontent.com/docsz/AD_4nXfzB9v9nWO6Q6Ld6ppxD7hftceXL-hRsJJ2O7uFUAveABfmnVrkqDEizCQ-2ODJx5vfIIKonZdBPprQUHwCL_hJh6ChvuhbtZid4yP10dsWnD1-hGPTANxRGo2xQT3gluSbeN9D8gAgzENkLSIhgCw?key=y-RkjHAFGdFY8mvBT9Vgqno4Figure S7.

Expression of genes (rows = genes) involved in digestion, blood coagulation, and electron transport, which were affected by acidification as a single stressor or when combined with suboptimal temperatures (outlined in red) compared to control (outlined in black). The red-blue gradient indicates the z-score of expression values (treatment means) standardized across treatments for each gene, where red and blue indicate higher and lower mean expression, respectively. Gene IDs refer to the gadMor3.0 assembly.



**Figure S8**.

Expression of genes (rows = genes) involved in protein synthesis (ribosomal biogenesis and stress

response chaperone proteins) were affected by cold temperature (outlined in red) relative to control (outlined in black). The red-blue gradient indicates the z-score of expression values (treatment means) standardized across treatments for each gene, where red and blue indicate higher and lower mean expression, respectively.

Gene IDs refer to the gadMor3.0 assembly.

##### 

##### 

##### Figure S9: Word cloud generated from gene ontology (GO) biological processes associated with immune system genes that were differentially expressed in warm treatment. Word size indicates prevalence (larger = more frequent). Purple and orange colors highlight the most frequent words. The following general words were removed due to presence in many processes: process, regulation, negative, positive, response, pathway, and production.

##### 

##### 

##### 

##### 

##### 

##### 

##### 

##### 

##### 

##### 

##### 

##### 

##### 

##### Figure S10: Wordcloud generated from all biological processes associated with genes that were upregulated and contributed to the enriched biological process Cell Adhesion.

##### 

##### Figure S11: Expression level of heat shock proteins that were differentially expressed at higher levels in cold treatments. Points represent gene counts for individual fish.

##### 

##### 

##### Genes expressed non-linearly with temperature

Given the large effect of temperature on gene expression, we performed deeper analyses on the processes that become more or less active away from control temperature (6°C). Genes were interrogated for those with expression values fitting 2nd order non-linear patterns across temperature [(Bogan, 2022; Bogan et al., 2023)](https://paperpile.com/c/n1vINc/TEDOR+B6yAs). Using edgeR [(Robinson et al., 2010)](https://paperpile.com/c/n1vINc/n07LT), gene counts were normalized by library size using normalization factors calculated by the weighted trimmed mean of M-values method (TMM), and scaled to log2-counts per million (CPM). Negative binomial generalized linear models were constructed from log2-CPM counts to test for main effects of temperature and temperature2. Separate linear and nonlinear regression models were then fit to log2-CPM normalized counts using lm(), and likelihood ratio tests (LRTs, using lrtest() and FDR-adjusted p-values <0.01) determined the best fit model for each gene (linear, or 2nd-order non-linear). A conservative set of genes responding non-linearly to temperature was then identified as those with expression affected by temperature2 as a main effect, for which the non-linear model predicted expression, and the LRT ratio supported non-linear expression. Nonlinear gene sets were further filtered for those that also were identified as differentially expressed among any two treatments from the *DESeq2* analysis.

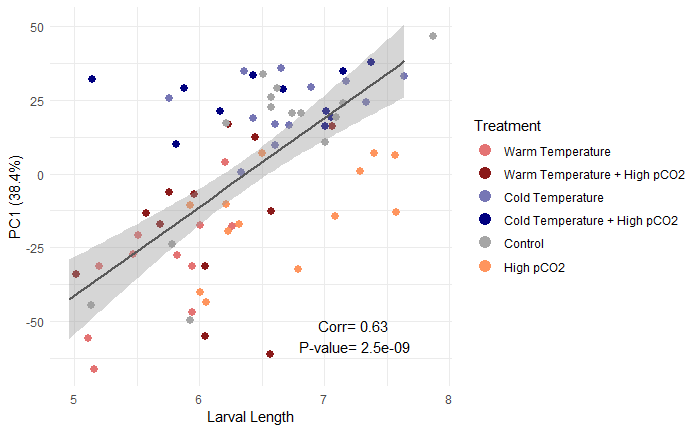
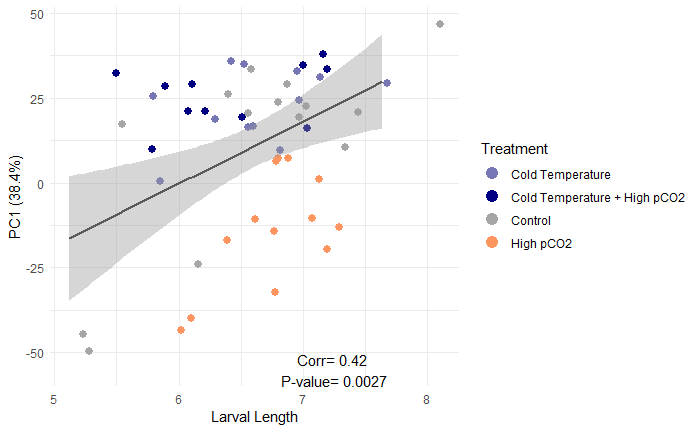
##### Figure S12. (Left) Enriched Uniprot Keyword biological processes in genes that responded non-linearly to temperature, where purple (i.e. convex) and green (i.e. concave) indicate processes that were most and least active in control temperature (6°C), respectively. Point size indicates the -Log10 transformed p-value (all p-values < 0.05), such that larger points are more significantly enriched. (Right) Average expression of genes that were highest (n=207, light purple) and lowest (n=80, light green) at control temperature (6°C). Expression levels shown here are only for larval cod held at ambient pCO2.

We identified a conservative set of genes that fit 2nd-order nonlinear models across temperature (n=287). 207 genes responded to temperature in a convex shape such that expression was generally highest at control temperature and decreased in cold and/or warm temperatures (Figure S9, purple). The convex-pattern genes were enriched for lipid and fatty acid metabolism, heme biosynthesis, and porphyrin biosynthesis. Another 80 genes had the opposite pattern – a concave shape with temperature such that expression was generally lowest at control temperature (Supplemental Figure X, green), which were enriched for processes relating to transcription regulation, apoptosis, and phagocytosis.

#### Exploring size-effects on gene expression

Larvae reared in warming were smaller at the time of sampling than the other two treatments (no difference in size by pCO2) (Supplemental Figure 9). While they were at the same developmental stage (post-yolk absorption, pre-flexion, feeding), and were similar in age using degree days post-hatch (110 DDPH) compared to the control conditions (114 DDPH), we explored the impact of length on gene expression.

##### Figure S13: Distribution of fish lengths at the time of sampling larval fish used in RNASeq analysis.



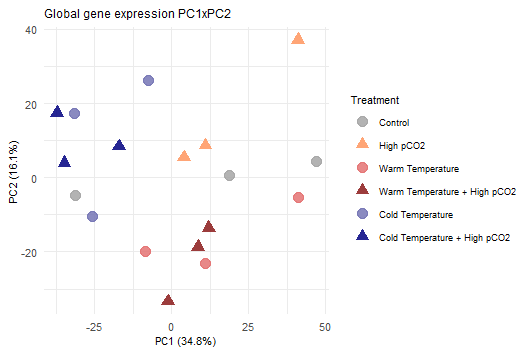
##### Figure S14: PC1 from the global expression PCA analysis, plotted against larval length along with the Pearson correlation coefficient and P-value calculated using all treatments (left) and without the warming treatment (right).

To first explore the relationship between length and gene expression, we examined the correlation between length and the global PCA principal components (Figure S10). PC1 correlated strongly with length when all treatments were included in the correlation, and weakly but still significantly when the warming treatment was omitted. We then performed PCA analyses using subsets of the samples (n=3, 7 per treatment) to minimize the discrepancies in larval length.

|  |  |
| --- | --- |
| **Table S3:** Larval length mean+sd when n=3 per treatment, range = 5.6-6.4 mm. | |
| Treatment | Mean Length |
| control conditions | 5.9±0.4 |
| OA | 6.2±0.2 |
| Cold | 6.2±0.2 |
| Cold+OA | 6.1±0.2 |
| Warm | 6.0±0.4 |
| Warm+OA | 6.1±0.2 |

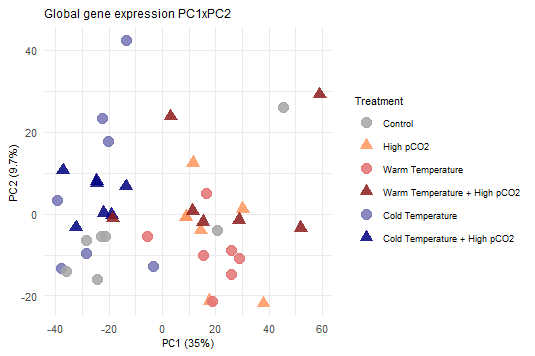
When n=3 per treatment to restrict larval size to a narrow range and ensure similar size distribution (Table S3), PC1 and PC2 are significant components, samples still cluster by treatment in multivariate space, and PC1 correlates with length (-0.60, P-value=0.009).

##### Figure S15: PCA constructed from n=3 fish per treatment with narrow size range.



When n=7 per treatment to maximize the number of samples used in the analysis while maintaining a relatively narrow size range (Supplemental Table 2), PC1, PC2 and PC3 are significant components, samples still cluster by treatment in multivariate space, and PC1 correlates with length (-0.52, P-value=0.0004).

|  |  |  |
| --- | --- | --- |
| **Table S4:** Larval length mean±sd and number of differentially expressed genes when n=7 random samples per treatment within the length range = 5.7-6.8 mm. | | |
| Treatment | Mean Length | No. DEGs compared to control conditions |
| control conditions | 6.3±0.5 | NA |
| OA | 6.4±0.3 | 3,398 |
| Cold | 6.5±0.3 | 491 |
| Cold+OA | 6.3±0.5 | 406 |
| Warm | 5.9±0.2 | 2,119 |
| Warm+OA | 6.0±0.2 | 1,682 |



##### 

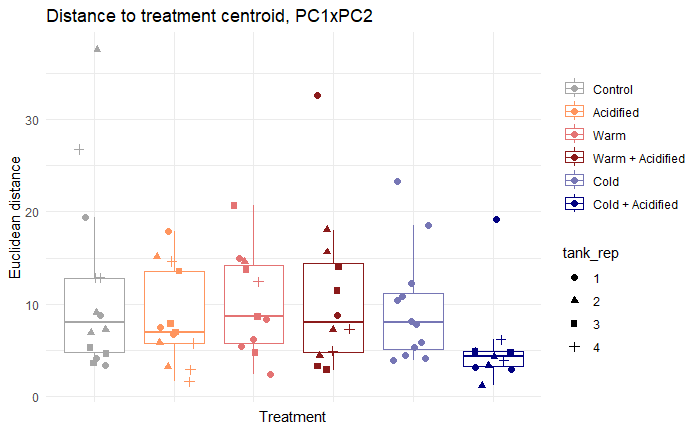
##### Figure S16: PCA constructed from n=7 fish per treatment with narrow size range.

##### Figure S17: Enriched biological processes in differentially expressed genes when n=7 randomly selected samples / larval fish per treatment within a narrow length range 5.7-6.8 mm.

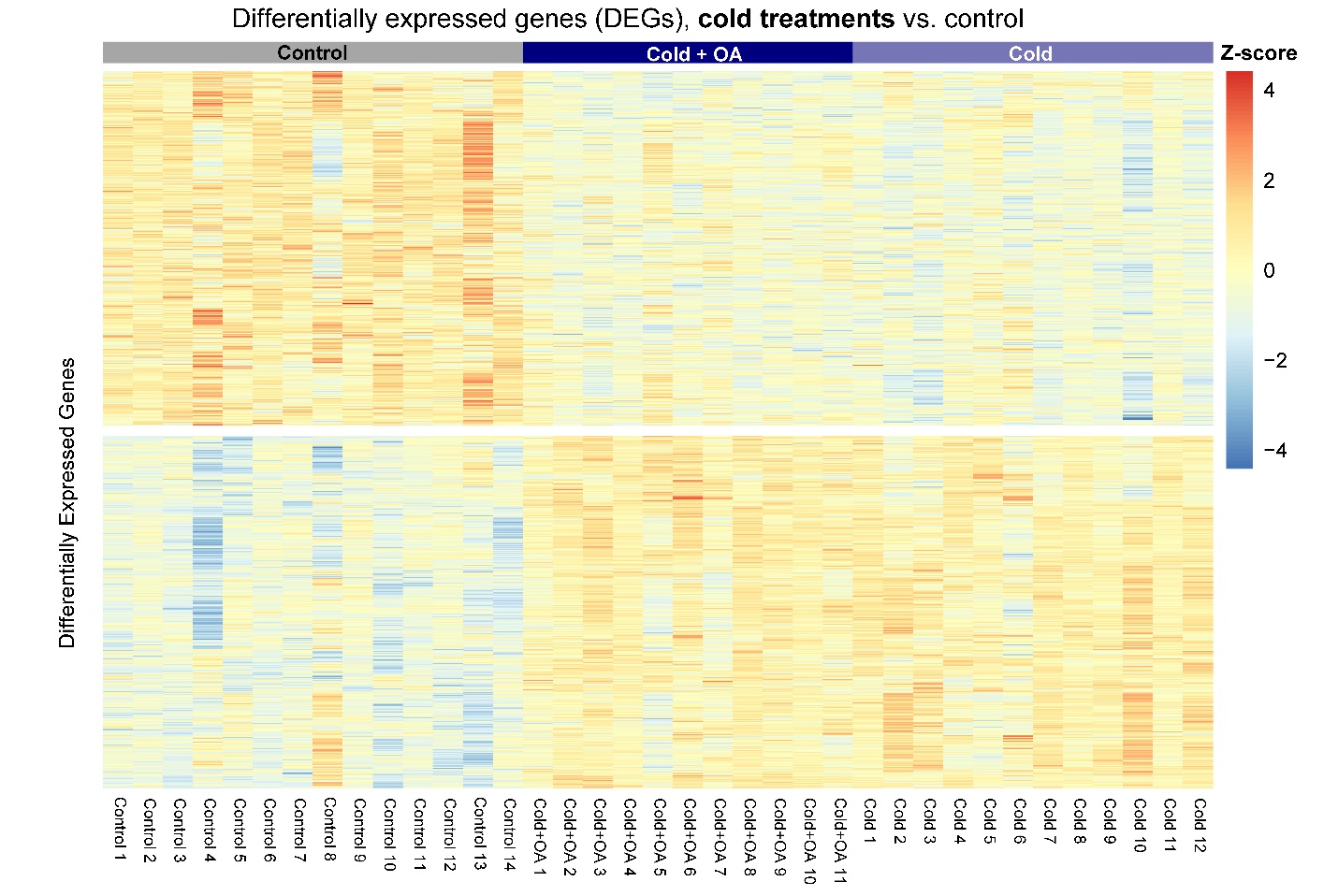
I re-ran the full differential expression analysis using n=7 random samples per treatment within the size range 5.7-6.8 mm. There were fewer differentially expressed genes overall in response to each treatment (Supplemental Table 2), but the biological processes enriched in differentially expressed genes were similar to those identified by the analysis using the full dataset. However, given the high correlation among larval length and PC1 in the full analysis and the restricted analysis, which explains a majority of variation in multivariate space in most of the PCA’s, I performed a separate differential expression analysis to identify and remove genes with expression strongly associated with length. Using *DESeq2* I identified genes with expression dependent on larval length. Using edgeR I tested for effects of length as a continuous explanatory variable. Before looking at the consensus set of genes among *DESeq2* and edgeR, I wanted to ensure that the linear-response relationship wasn’t simply due to the warm temperature treatment having unique expression levels. Therefore, I re-ran the analysis without the 10C group to identify genes with linear response to length across the other temperature groups. For a gene to be removed from the analysis because it was considered length-dependent, it had to 1) be associated with size as per DESeq2, 2) respond linearly to length in the all-temperature analysis, and 3) respond linearly to length without the warm treatment. To examine the function of genes that are associated with length in these larval fish I performed an enrichment analysis on the discarded genes. Of the 6,664 genes associated with length 5,680 were annotated with a Uniprot ID. Nine Biological Process Uniprot Keywords were enriched (Table S3), and included those relating to ion transport, neurogenesis, biological rhythms (related to circadian rhythm), apoptosis, and host-virus interaction.

##### Table S5: Biological Process Uniprot Keywords enriched in genes removed from analysis because they were associated with larval fish length.

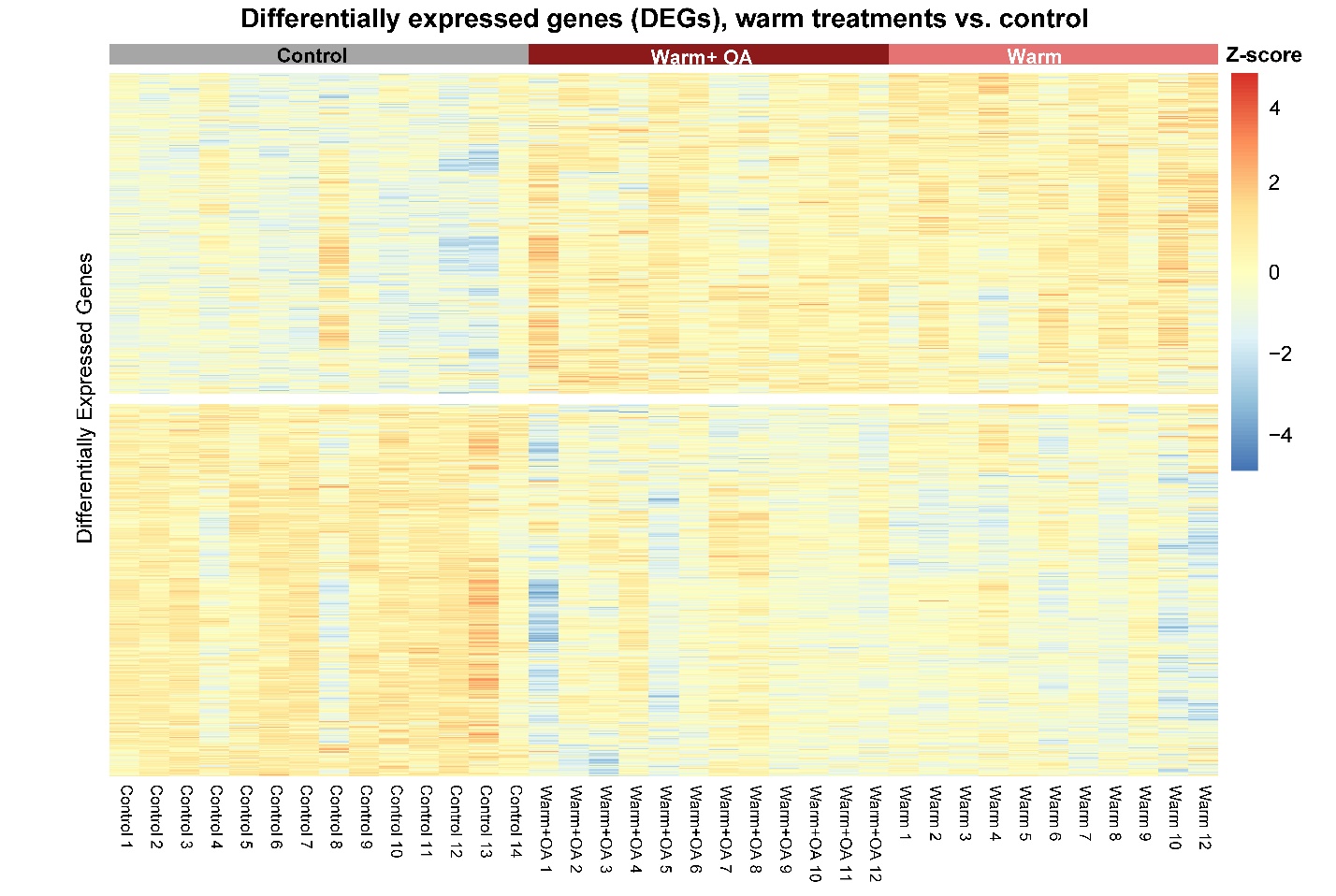
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Uniprot Keyword** | **Biological Process** | **No. of genes** | **P-Value** | **FDR** |
| KW-0229 | DNA integration | 12 | 0.000000028 | 0.0000032 |
| KW-0633 | Potassium transport | 22 | 0.000016 | 0.00092 |
| KW-0233 | DNA recombination | 15 | 0.000054 | 0.0020 |
| KW-0406 | Ion transport | 54 | 0.0015 | 0.042 |
| KW-0524 | Neurogenesis | 32 | 0.0069 | 0.15 |
| KW-0109 | Calcium transport | 14 | 0.0085 | 0.16 |
| KW-0090 | Biological rhythms | 15 | 0.025 | 0.40 |
| KW-0053 | Apoptosis | 32 | 0.037 | 0.52 |
| KW-0945 | Host-virus interaction | 16 | 0.047 | 0.59 |



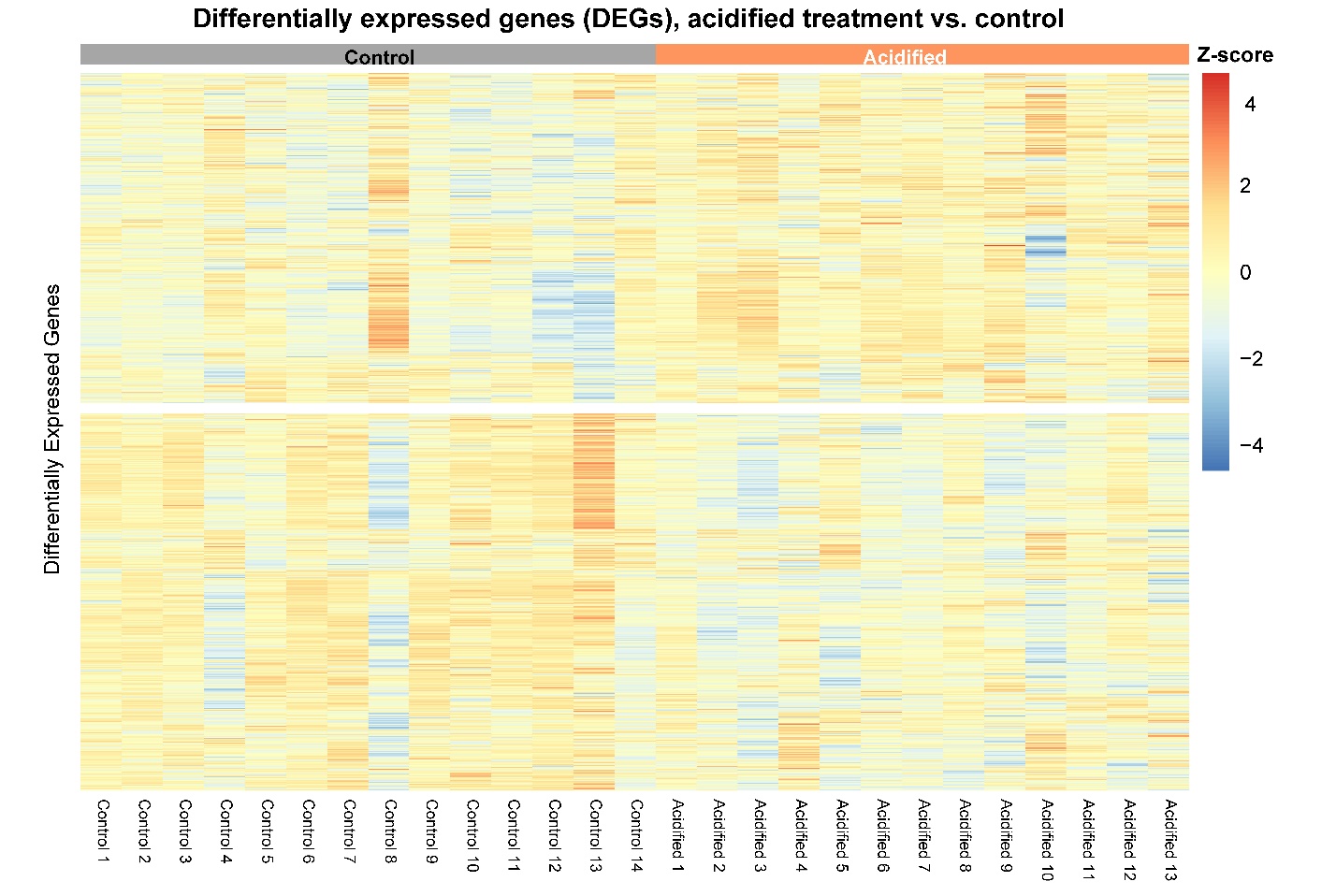
**Figure S18:** Euclidean distances from each sample to treatment centroids (means) in global gene expression PCA (PC1 x PC2, n=14,412 genes). Point shapes indicate the replicate tanks from which larvae were sampled.



**Figure S19:** Heatmap of the genes that were differentially expressed in the cold or cold + OA treatments compared to control conditions (2,257 unique genes). Columns represent individual larval cod, rows represent individual genes, and color indicates the relative expression value (Z-score) with dark red values indicating highest expression levels and dark blue indicating lowest expression values. Note that not all shown genes were differentially expressed in both cold treatments.



**Figure S20:** Heatmap of the genes that were differentially expressed in the warm or warm + OA treatments compared to control conditions (4,626 unique genes). Columns represent individual larval cod, rows represent individual genes, and color indicates the relative expression value (Z-score) with dark red values indicating highest expression levels and dark blue indicating lowest expression values. Note that not all shown genes were differentially expressed in both warm treatments.



**Figure S21:** Heatmap of the genes that were differentially expressed in the acidified treatment compared to control conditions (1,990 genes). Columns represent individual larval cod, rows represent individual genes, and color indicates the expression value (Z-score) with dark red values indicating highest expression levels and dark blue indicating lowest expression values.