

# Evolutionary and acclimatory shifts in gene expression of *Eurytemora affinis* copepods reared in saline and freshwater conditions during laboratory experiments from 2011-2014

**Website:** <https://www.bco-dmo.org/dataset/883426>

**Data Type:** Other Field Results, experimental

**Version:** 1

**Version Date:** 2022-11-10

## Project

» [Evolutionary Responses to Global Changes in Salinity and Temperature](#) (Evolutionary genomics of a copepod)

Contributors	Affiliation	Role
<a href="#">Lee, Carol E.</a>	University of Wisconsin (UW-Madison)	Principal Investigator
<a href="#">Posavi, Marijan</a>	University of Wisconsin (UW-Madison)	Student
<a href="#">Gerlach, Dana Stuart</a>	Woods Hole Oceanographic Institution (WHOI BCO-DMO)	BCO-DMO Data Manager

## Abstract

To explore mechanisms of freshwater adaptation and distinguish between adaptive (evolutionary) and acclimatory (plastic) responses to salinity change, we examined genome-wide patterns of gene expression between ancestral saline and derived freshwater populations of the *Eurytemora affinis* species complex, reared under two different common-garden conditions (0 vs. 15 PSU). These data include the RNA-seq Illumina short paired end reads (101 base pairs) of 10 freshwater and 12 saline copepods *Eurytemora affinis* samples. The freshwater copepod samples were collected in Lake Michigan (Racine Harbor), while the saline copepods were collected in Baie de L'Isle Verte, St. Lawrence marsh, Quebec, Canada. These data have important implications for our understanding of the evolutionary and physiological mechanisms of range expansions by some of the most widespread invaders in aquatic habitats.

## Table of Contents

- [Coverage](#)
- [Dataset Description](#)
  - [Methods & Sampling](#)
  - [Data Processing Description](#)
- [Data Files](#)
- [Related Publications](#)
- [Related Datasets](#)
- [Parameters](#)
- [Instruments](#)
- [Project Information](#)
- [Funding](#)

## Coverage

**Spatial Extent:** N:48.003889 E:-69.425278 S:42.729444 W:-87.778889

**Temporal Extent:** 2011-09-28 - 2014-03-18

## Methods & Sampling

### Study objectives

The goal of this study was to explore evolutionary shifts in gene expression between ancestral saline and freshwater invading populations of the *Eurytemora affinis* (copepod) species complex on a genome-wide scale.

To explore mechanisms of freshwater adaptation and distinguish between adaptive (evolutionary) and acclimatory (plastic) responses to salinity change, laboratory experiments were conducted using both

ancestral saline and derived freshwater populations of *Eurytemora affinis*. Then RNA-seq data -- Illumina short paired-end (PE) reads (101 base pairs) of 10 freshwater and 12 saline *E.affinis* samples -- were used to answer the following questions:

- (1) What are the patterns of evolutionary shifts in gene expression between the ancestral saline and freshwater invading populations?
- (2) What are the plastic (acclimatory) changes in gene expression between salinities (0 PSU vs. 15 PSU conditions) within each of the saline and freshwater populations?
- (3) Are the magnitude and direction of plasticity in gene expression correlated with evolutionary responses?
- (4) Has plasticity in gene expression evolved following freshwater invasions?

### Collection of ancestral populations

The copepods were collected using a plankton net mesh size of 50  $\mu\text{m}$  in diameter, from a depth of 1-4 meters from near the shore. The freshwater copepods were collected in April-May 2006 by throwing the plankton net off the dock in Racine Harbor, Lake Michigan in Wisconsin, USA (42.729444 N, 87.778889 W). The saline copepods were collected by small boats near the shore in Baie de L'Isle Verte, St. Lawrence marsh, Quebec, Canada (48.003889 N, 69.425278 W) in May-June 2006. Collected samples were transported to the laboratory where *Eurytemora affinis* individuals were identified and sampled under the microscope.

### Laboratory cultures and experiments

Four inbred lines of *Eurytemora affinis* (two each from the two populations) were generated through full-sibling mating for 30 generations. Two independent saltwater inbred lines (SW1 and SW2) were derived from the ancestral saline population in Baie de L'Isle Verte (Canada) and reared at their native salinity of 15 PSU. The two freshwater inbred lines (FW1 and FW2) were derived from the freshwater invading population in Racine Harbor (USA) and reared in Lake Michigan water (0 PSU, conductivity 300  $\mu\text{S}/\text{cm}$ ). In addition, reciprocal F1 crosses between freshwater and saline inbred lines were created and reared to test for allele-specific expression by comparing gene expression in parental lines and their F1 crosses.

Two replicate common-garden reaction norm experiments, each consisting of a 2  $\times$  2 factorial design, were performed to compare patterns of the gene expression of the FW and SW inbred lines (see Materials and Methods in Posavi et al. 2020). Total RNA from whole bodies of 50 copepods (25 females and 25 males) per sample was extracted using Trizol reagent (Ambion RNA) and Qiagen RNeasy Mini Kit for purification (Qiagen cat. no. 74104). Extracted and purified RNA samples were stored at -80 degrees Celsius until sequencing. The strand-specific Illumina RNA-seq libraries (Parkhomchuk et al., 2009) of polyA purified mRNA were constructed using the TruSeq RNA Sample Prep kit (Illumina). Three biological replicates per inbred line were sequenced using the Illumina HiSeq 2000 platform in the Institute for Genome Sciences at the University of Maryland School of Medicine and generated 101-bp-long paired-end read data.

These data have important implications for understanding the evolutionary and physiological mechanisms of range expansions by some of the most widespread invaders in aquatic habitats.

### Problem report

One replicate of the FW1 inbred line was excluded because of bacterial infection

### Additional information

~ Detailed methods, results, and figures can be found in Posavi et al. (2020) (see *Related Publications* section).

~ The sequence data can be viewed under NCBI BioProject PRJNA278152 (see *Related Datasets*).

### Data Processing Description

To assess the taxonomic composition and ensure the provenance of the sample, the RNA-seq reads were screened by the Institute for Genome Sciences' QC pipeline against a local installation of the NCBI nucleotide database. After that, each sample was run through the data processing pipeline to detect the presence of adaptor sequences or low read quality using FastQC (Andrews, 2010). Reads were trimmed with Trimmomatic version 032 (Bolger et al., 2014). On average,  $3.5 \times 10^7$  paired-end (101 bp) reads per sample passed these filtering steps.

To quantify transcript (gene) expression levels, an expectation maximization approach employing the RSEM (RNA-seq by Expectation Maximization) package (Li & Dewey, 2011) was used. The *Eurytemora affinis* complex (Atlantic clade, aka *E. carolleae*) draft genome served as the reference genome. Automated gene annotation of this genome was conducted at the Baylor College of Medicine Human Genome Sequencing Center within the i5K pilot project (using Maker2.2 following methods of Holt & Yandell, 2011) resulting in 29,783 gene models. For additional details on gene annotation, see Eyun et al. (2017).

To improve the automated gene annotation, the Cufflinks Tuxedo protocol (Trapnell et al., 2012) was used. The merged gene annotation file was used as input into RSEM to (a) build reference transcript sequences using the prepare-reference and (b) align RNA-seq reads to the reference transcripts and estimate gene and transcript abundances (using rsem-calculate-expression).

To map RNA-seq reads to the *E. affinis* complex genome, Bowtie2 was employed resulting in 16–30 million mapped paired-end reads. To verify the annotation of DE genes, the manual annotation of the *E. affinis* complex genome (using the Web Apollo platform on the i5k Workspace, [https://i5k.nal.usda.gov/Eurytemora\\_affinis](https://i5k.nal.usda.gov/Eurytemora_affinis)) was performed.

Structural gene annotation, generated by merging transcriptomes of 22 RNA samples, resulted in 37,827 putative genes. To increase the power to detect differential expression, genes with less than one count-per million (CPM) in at least two samples were filtered out, as were transcripts with best BLASTx matches to bacteria (n = 3), and transcripts without BLASTx hits. These filtering steps left 14,082 putative genes remaining for the differential gene expression analysis, all of which mapped to the *E. affinis* genome assembly. Subsequently, the normalization on 14,082 genes was performed, using the Trimmed Mean of M values method (Robinson & Oshlack, 2010), available in Bioconductor's edgeR package for R software (Chen et al., 2018; Robinson et al., 2009).

To identify significant differences in gene expression between saline and freshwater inbred lines (Goal 1) and between salinities (0 and 15 PSU) (Goal 2), statistical analyses using a generalized linear model (GLM) were performed. To detect DE genes, a negative binomial generalized linear model was used that accommodated the complex designs of the common-garden experiments (Bioconductor's edgeR package with function glmQLFit with option robust= TRUE). To conduct the test for each genotype (inbred line) and salinity combination, the read counts were modeled as the result of the fixed effects of genotype (inbred line effect), salinity (0 and 15 PSU), batch, and genotype-by-salinity interactions. For multiple hypothesis testing, we adjusted p-values using the Benjamini and Hochberg (1995) method with a false discovery rate (FDR) threshold of 0.05.

## Software

(See also Related Publications section below)

- Bowtie2 (Langmead & Salzberg, 2012)
- BLASTx (Mukhopadhyay et al., 2017)
- Cufflinks Tuxedo protocol (Trapnell et al., 2012)
- edgeR (Chen et al., 2018; Robinson et al., 2009)
- FastQC (Andrews, 2010)
- Maker2.2 (Holt & Yandell, 2011)
- RSEM (Li & Dewey, 2011)
- Trimmomatic (Bolger et al., 2014)

## BCO-DMO Processing

- Converted date to Y-M-D format
- Added information about inbred lines for better readability

[ [table of contents](#) | [back to top](#) ]

---

## Data Files

File
<b>gene_expression.csv</b> (Comma Separated Values (.csv), 8.64 KB) MD5:d28111ecea446ecf40dec8018e1dc61d
Primary data file for dataset ID 883426

[ [table of contents](#) | [back to top](#) ]

---

## Related Publications

Andrews S. (2010). FastQC: a quality control tool for high throughput sequence data. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>  
Software

Benjamini, Y., & Hochberg, Y. (2000). On the Adaptive Control of the False Discovery Rate in Multiple Testing With Independent Statistics. *Journal of Educational and Behavioral Statistics*, 25(1), 60–83.

doi:[10.3102/10769986025001060](https://doi.org/10.3102/10769986025001060)

Methods

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120. doi:[10.1093/bioinformatics/btu170](https://doi.org/10.1093/bioinformatics/btu170)  
*Software*

Chen, Y., McCarthy, D. J., Ritchie, M., Robinson, M., & Smyth, G. K. (2018). edgeR: differential expression analysis of digital gene expression data. User's guide. <https://www.bioconductor.org/packages/devel/bioc/vignettes/edgeR/inst/doc/edgeRUsersGuide.pdf>  
*Software*

De Wit, P., Pespeni, M. H., Ladner, J. T., Barshis, D. J., Seneca, F., Jaris, H., Therkildsen, N. O., Morikawa, M., & Palumbi, S. R. (2012). The simple fool's guide to population genomics via RNA-Seq: an introduction to high-throughput sequencing data analysis. *Molecular Ecology Resources*, 12(6), 1058–1067. Portico. <https://doi.org/10.1111/1755-0998.12003>  
*Methods*

Eyun, S., Soh, H. Y., Posavi, M., Munro, J. B., Hughes, D. S. T., Murali, S. C., Qu, J., Dugan, S., Lee, S. L., Chao, H., Dinh, H., Han, Y., Doddapaneni, H., Worley, K. C., Muzny, D. M., Park, E.-O., Silva, J. C., Gibbs, R. A., Richards, S., & Lee, C. E. (2017). Evolutionary History of Chemosensory-Related Gene Families across the Arthropoda. *Molecular Biology and Evolution*, 34(8), 1838–1862. <https://doi.org/10.1093/molbev/msx147>  
*Related Research*

,  
*Methods*

Holt, C., & Yandell, M. (2011). MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics*, 12(1). <https://doi.org/10.1186/1471-2105-12-491>  
*Software*

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), 357–359. doi:[10.1038/nmeth.1923](https://doi.org/10.1038/nmeth.1923)  
*Software*

Li, B., & Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, 12(1). doi:[10.1186/1471-2105-12-323](https://doi.org/10.1186/1471-2105-12-323)  
*Software*

Mukhopadhyay, C.S. and Choudhary, R.K (2017) BLASTx. In C.S. Mukhopadyay, R.K. Choudhary, & M.A. Iquebal (Eds.) *Basic Applied Bioinformatics* (Chapter 14, p.103-108). Hoboken, New Jersey: John Wiley & Sons, Inc (Wiley) <https://isbnsearch.org/isbn/9781119244417>  
*Software*

Parkhomchuk, D., Borodina, T., Amstislavskiy, V., Banaru, M., Hallen, L., Krobitch, S., Lehrach, H., & Soldatov, A. (2009). Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic Acids Research*, 37(18), e123–e123. <https://doi.org/10.1093/nar/gkp596>  
*Methods*

Poelchau, M., Childers, C., Moore, G., Tsavatapalli, V., Evans, J., Lee, C.-Y., Lin, H., Lin, J.-W., & Hackett, K. (2014). The i5k Workspace@NAL—enabling genomic data access, visualization and curation of arthropod genomes. *Nucleic Acids Research*, 43(D1), D714–D719. <https://doi.org/10.1093/nar/gku983>  
*Software*

Posavi, M., Gulisija, D., Munro, J. B., Silva, J. C., & Lee, C. E. (2020). Rapid evolution of genome-wide gene expression and plasticity during saline to freshwater invasions by the copepod *Eurytemora affinis* species complex. *Molecular Ecology*, 29(24), 4835–4856. Portico. <https://doi.org/10.1111/mec.15681>  
*Results*

,  
*Methods*

Robinson, M. D., & Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology*, 11(3), R25. <https://doi.org/10.1186/gb-2010-11-3-r25>  
*Methods*

Robinson, M. D., & Smyth, G. K. (2007). Moderated statistical tests for assessing differences in tag abundance. *Bioinformatics*, 23(21), 2881–2887. <https://doi.org/10.1093/bioinformatics/btm453>  
*Methods*

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2009). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), 139–140. <https://doi.org/10.1093/bioinformatics/btp616>  
*Software*

The UniProt Consortium (2016). UniProt: the universal protein knowledgebase. *Nucleic Acids Research*, 45(D1), D158–D169. doi:[10.1093/nar/gkw1099](https://doi.org/10.1093/nar/gkw1099)

*Methods*

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., Pimentel, H., Salzberg, S. L., Rinn, J. L., & Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols*, 7(3), 562–578. <https://doi.org/10.1038/nprot.2012.016>

*Software*

[ [table of contents](#) | [back to top](#) ]

---

## Related Datasets

### IsDerivedFrom

University of Maryland School of Medicine. *Eurytemora affinis* Transcriptome or Gene expression. 2015/03. In: BioProject [Internet]. Bethesda, MD: National Library of Medicine (US), National Center for Biotechnology Information; 2011-. Available from: <http://www.ncbi.nlm.nih.gov/bioproject/PRJNA278152>. NCBI:BioProject: PRJNA278152.

### IsSupplementedBy

Lee, C. (2020). *Eurytemora affinis* complex (Atlantic clade). i5k Workspace@NAL. [https://i5k.nal.usda.gov/Eurytemora\\_affinis](https://i5k.nal.usda.gov/Eurytemora_affinis)

[ [table of contents](#) | [back to top](#) ]

---

## Parameters

Parameter	Description	Units
Record_num	Record number	unitless
Culture_sample_date	Date of sampling for the inbred line and F1 crosses	unitless
Sample_name	Sample name indicating the E.affinis line, salinity PSU, and replicate number	unitless
Organism	Organism that was raised and studied (copepod Eurytemora affinis)	unitless
Sex	Sex of copepods comprising the sample; male, female, or pooled	unitless
Female_inbred_line	Identifier of inbred line for the females in the sample; VA = SW1 (saline inbred line 1), VE = SW2 (saline inbred line 2) , RA = FW1 (freshwater inbred line 1), RB = FW2 (freshwater inbred line 2)	unitless
female_ancestor_location	Home location of the female ancestor	unitless
Male_inbred_line	Identifier of inbred line for the males in the sample; VA = SW1 (saline inbred line 1), VE = SW2 (saline inbred line 2) , RA = FW1 (freshwater inbred line 1), RB = FW2 (freshwater inbred line 2)	unitless
male_ancestor_location	Home location of the male ancestor	unitless
Experimental_culture	Culture identification of inbred lines	unitless
Description	Experiment description	unitless
Salinity_conditions_rearing	Salinity conditions in which the copepods were reared	unitless
PSU_culture	Practical salinity unit value for the culture water (or water in which the copepods were reared)	PSU
BioProject	NCBI BioProject identifier	unitless
SRA	NCBI Sequence Read Archive identifier for sample accession	unitless
BioSample	NCBI BioSample number	unitless

[ [table of contents](#) | [back to top](#) ]

## Instruments

<b>Dataset-specific Instrument Name</b>	Illumina HiSeq 2000
<b>Generic Instrument Name</b>	Automated DNA Sequencer
<b>Dataset-specific Description</b>	Libraries were sequenced on an Illumina HiSeq platform at the University of Maryland, School of Medicine, Institute for Genome Sciences.
<b>Generic Instrument Description</b>	General term for a laboratory instrument used for deciphering the order of bases in a strand of DNA. Sanger sequencers detect fluorescence from different dyes that are used to identify the A, C, G, and T extension reactions. Contemporary or Pyrosequencer methods are based on detecting the activity of DNA polymerase (a DNA synthesizing enzyme) with another chemoluminescent enzyme. Essentially, the method allows sequencing of a single strand of DNA by synthesizing the complementary strand along it, one base pair at a time, and detecting which base was actually added at each step.

<b>Dataset-specific Instrument Name</b>	
<b>Generic Instrument Name</b>	Plankton Net
<b>Generic Instrument Description</b>	A Plankton Net is a generic term for a sampling net that is used to collect plankton. It is used only when detailed instrument documentation is not available.

[ [table of contents](#) | [back to top](#) ]

---

## Project Information

### Evolutionary Responses to Global Changes in Salinity and Temperature (Evolutionary genomics of a copepod)

**Coverage:** St. Lawrence estuary, Gulf of Mexico, Great Lakes, Baltic Sea

NSF Award Abstract:

Drastic changes in the global water cycle and increases in ice melt are causing the freshening of Northern coastal seas. The combination of both reduced salinity and increased temperature will likely act in concert to reduce populations of estuarine and marine organisms. Data indicate that reduced salinity and high temperature would each increase the energy costs as well as reduce survival and reproduction of the common copepod *Eurytemora affinis*. This project will examine the joint effects of salinity reduction and temperature increase on the evolutionary responses of populations of *E. affinis* in the wild, as well as in selection experiments in the laboratory. This study will provide novel insights into responses of organisms to climate change, as no study has analyzed the joint impacts of salinity and temperature on evolutionary responses, and relatively few studies have examined the impacts of declining salinity. In general, how selection acts at the whole genome level is not well understood, particularly for non-model organisms. As a dominant estuarine copepod, *E. affinis* is among the most important species sustaining coastal food webs and fisheries in the Northern Hemisphere, such as salmon, herring, and anchovy. Thus, insights into its evolutionary responses with changing climate have important implications for sustainability of fisheries and food security. Two graduate students from historically underrepresented groups will be trained during this project. The project will have additional societal benefits, including development of educational modules for K-12 students and international collaboration.

This study will address the following questions: (1) To what extent could populations evolve in response to salinity and temperature change, and what are the fitness and physiological costs? (2) How will populations respond to the impacts of salinity-temperature interactions? (3) Do wild populations show evidence of natural selection in response to salinity and temperature? To analyze the evolutionary responses of *E. affinis* populations to the coupled impacts of salinity and temperature, the investigator will perform laboratory selection experiments and population genomic surveys of wild populations. Selection experiments constitute powerful tools for determining the rate, trajectory, and limits of adaptation. During laboratory selection, evolutionary shifts in fitness-related traits and genomic expression will be examined, as well as genomic signatures of selection in response to low salinity and high temperature selection regimes. The investigator will also conduct population genomic sequencing of *E. affinis* populations that reside along salinity and temperature gradients in the St. Lawrence and Baltic Sea, and identify genes that show signatures of selection. The project will determine whether the loci that show signatures of selection in the wild populations are the same as those favored during laboratory selection. This reproducibility will provide greater confidence that the genes involved in adaptation to salinity and/or temperature have been captured.

[ [table of contents](#) | [back to top](#) ]

---

## Funding

Funding Source	Award
<a href="#">NSF Division of Ocean Sciences (NSF OCE)</a>	<a href="#">OCE-1658517</a>

[ [table of contents](#) | [back to top](#) ]