AN ABSTRACT OF THE THESIS OF

Kenya S. Bynes for the degree of Master of Science in Fisheries Science presented on May 26, 2021.

Title: Influence of Oceanographic Conditions on Copepod (*Calanus pacificus*) Genetic Expression

Abstract approved: _____

Michael A. Banks

RNAseq data among copepods (*Calanus pacificus*) were sampled and analyzed in relation to oceanographic and climatic conditions to determine if new information derived from such an approach might assist applications that use copepods as ecosystem indicators to predict fishery yields. These samples were taken as part of long-term studies conducted by NOAA and Oregon State University scientists that have surveyed biological and physical conditions fortnightly for over twenty years to better understand changing ocean conditions. Plankton tows from two sampling periods (2016-2017 and 2017-2018) taken at station 5 of the Newport Hydrographic Line enabled replicated study and comparison of upwelling (active and relaxed), against downwelling (active and relaxed) conditions, during alternate states of the Pacific Decadal Oscillation, and El Nino Southern Oscillation. Substantial information on differential expression was evident with the range of log fold change among mean counts observed including significant observations for most comparisons. Although evidence for similar biological process among comparisons of common conditions were apparent, these could only be inferred and assumed from reviewed database of function for a diverse assemblage of transcript tags from arthropods. Because of this limitation and that the ocean did not occupy a productive state at all during the total study it cannot be determined if gene expression in *C. pacificus* enhanced previous forecasts due to the lack of yield forecasted during the years of this study.

Influence of Oceanographic Conditions on Copepod (*Calanus pacificus*) Genetic Expression

by

Kenya S. Bynes

A THESIS

submitted to

Oregon State University

in partial fulfillment of the

requirements for the

degree of

Master of Science

Presented May 26, 2021

Commencement June 2022

Master of Fisheries Science thesis of Kenya S. Bynes presented on May 26, 2021

APPROVED:

Major Professor, representing Fisheries Science

Head of the Department of Fisheries, Wildlife and Conservation Sciences

---

Dean of the Graduate School

I understand that my thesis will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my thesis to any reader upon request.

Kenya S. Bynes, Author

ACKNOWLEDGEMENTS

I would like to acknowledge the people and institutions that made the following research possible. I would like to thank and show gratitude to the funding agencies of NOAA Living Marine Resources Cooperative Science Center and Cooperative Institute of Marine Resources Studies (PMEL) that believed in my research. I would also like to thank the Center for Genome Research and Biocomputing (CGRB) for teaching me the bioinformatic skills necessary to complete the research as well as the Illumina sequencing. I am very grateful to Jennifer Fisher and Sam Zeman for their help in collecting the samples. Dave Jacobson was tremendous in his assistance for the completion of this research by teaching me RNA isolation and preparation of the RNAlater®. Kym Jacobson and Bill Peterson were integral in teaching me about the morphology and life cycle of the *Calanus pacificus.* Heather Auld also shared her bioinformatic knowledge to help me with the code to help produce the bioinformatics analysis.

TABLE OF CONTENTS

LIST OF FIGURES

Figure Page

LIST OF TABLES

LIST OF APPENDICES

Chapter 1: Introduction

Research in fisheries oceanography focuses on integrated assessment of the "health" of the ecosystems supporting fished species, and assessment of the effects of climate variability and future climate change on growth, survival, and the abundance of fish entering a targeted population (Peterson et al. 2014). To evaluate the health of an ecosystem, ecosystem indicators are often used. Environmental variability has been shown to affect bioenergetics of lower trophic levels which ultimately impact the abundance of higher trophic levels (Peterson et al. 2014). In addition, the link between copepods and exploited fish stocks have previously been described in the match-mismatch hypothesis (Hays et al. 2005). Copepods are small, planktonic crustaceans that have previously been utilized as valuable ecosystem indictors (Roncalli et al. 2017) owing to their ubiquity, lack of commercial exploitation, and strong environmental influence on their population structure and distribution (Hays et al. 2005). For example, research on the relationship between plankton community composition and the strength of salmon and other fishery indexes has provided promising associations (Peterson et al. 2014). As a result, the notion that 'fat' and 'skinny' copepods provide reliable indicators for 'good' and 'bad' fishery yields and that these fluxes associate with positive and negative phases for the Pacific Decadal Oscillation (PDO) has become increasingly accepted (Losee et al., 2014). In recent years, however, rigor of these linkages has been questioned because of a run of unprecedented oceanographic conditions, such as the marine heat waves of 2014 and 2019. The primary objective of my research centers on the applicability of an alternate approach to provide additional information for fishery yield predictions. An overall longer-term goal is to determine which genes, expressed under different environmental (climatic) conditions relate to alternate nutrition states of plankton, and if such information has potential to improve fishery index predictions. Briefly, my study could be summarized as an evaluation of gene expression among copepods as a candidate method to provide complimentary information to commonly applied fishery indices such as the Peterson stop-light assay (Peterson et al. 2014). The stop-light assay is unique to the California Current ecosystem and conveys the characterization of different oceanographic conditions on juvenile salmon survival along with their predicted effect on the future catch. The stop-light assay receives its name from the color-coded table that uses rank scores to define the effect on the juvenile salmon population. The lower scores are shown in

green depicting a good or positive effect with slightly higher scores shown in yellow indicating neutral effect. The highest scores are shown in red indicating a bad or negative effect on the juveniles which will lead to a lower future yield (NOAA 2021). The connection of copepods to primary production and changing environmental conditions indicate the assessment of their 'productivity-status' may provide immediate reflection of the impact of changing bottom-up effects.

The environmental conditions of interest include upwelling, downwelling, the El Nino Southern Oscillation, and the Pacific Decadal Oscillation. Upwelling is an oceanographic condition in which winds blow from the north causing Ekman transport of surface water away from the shore and brings high nutrient colder water from the bottom of the ocean and fuels highly productive states. Down-welling is the reverse process of upwelling in which southerly winds cause surface water to move towards the coast, stack up and sink to the bottom and produce lower nutrition and less productive states. The El Nino Southern Oscillation is an environmental condition that is described as a conjunction of surface wind change from the easterly winds to westerly winds and a warming of the ocean surface, by above-average sea surface temperatures, in the central and eastern tropical Pacific Ocean while rainfall tends to become reduced over Indonesia and increases over the tropical Pacific Ocean. The oceanographic condition of the Pacific Decadal Oscillation has a similar climatic pattern to an El Nino Southern Oscillation. The Pacific Decadal Oscillation differs from an El Nino Southern Oscillation by occurring in the Pacific Basin instead of the equator and alternating between warm and cold phase on a ten-year cycle. These phases are sometimes extended and last up to 20 years.

The initial research focus was to develop a genetic assay to assess the relative occurrence of transcripts related to stress response, lipid, carbohydrate and protein synthesis and metabolism of *Calanus pacificus* copepods in relation to alternate environmental conditions. One reason for the focus of nutritional value and stress response is due to the nutritional effects and implications resulting from the consumption of copepods of differential nutritional value by juvenile stages of commercially important fish (anchovy, herring, other forage fish, cod, and flounder). Another reason for the focus on nutritional content is that the lipid status of zooplankton species composition has a close association with Pacific Decadal Oscillation (PDO) which has been successfully used as a predictor of fishery yields (Peterson et al. 2014). In brief, positive phases

of the PDO (a period in which warm temperatures are present) predict nutrient poor plankton (colloquially known as the southern/celery-type copepod assemblage) and low fishery yields, while a negative phases PDO (a period in which cold temperatures are present) indicate lipid rich plankton (colloquially known as the northern/cheeseburger-type copepod assemblage) and a high fishery yield (Peterson et al. 2014).

A potential breakdown of this linkage may have arisen from unprecedented persistent warmer temperature conditions observed in the California Current over the last several years that has raised questions as to whether these PDO/fishery yield predictions will remain consistent. Thus, the primary objective of my research was to determine if alternate approaches such as assessment of the relative occurrence of alternate transcripts among copepods sampled from different environmental conditions (eg. active and relaxed states of upwelling or downwelling) might yield information about the response of ecosystem indicators such as copepods to new and unprecedented conditions.

From summer 2016- early 2018, the project covered here has profiled RNA-Seq data collected from *C. pacificus* copepods sampled along the Newport Hydrographic Line (NHL) under different environmental conditions that have arisen naturally, paying attention to typical upwelling and downwelling periods. Findings were evaluated by apportioning RNA-seq observations for samples taken under different oceanographic/seasonal states:

1. Active upwelling against relaxed upwelling (temporary down-welling during an overall upwelling period) evaluation. This evaluation is important due to the typical condition of upwelling occurring along the NHL in the summer months from the influx of the cold water of the California Current. Gene expression profiles under upwelling will serve as the baseline for all gene expression findings. In relation to in situ gene expression, it is expected that transcripts relating to lipid production and energy will be more highly upregulated in active upwelling than in relaxed upwelling (temporary down-welling) or during states of the persistent increased temperature anomaly.

2. Seasonal evaluations focusing on longer-term upwelling (summer) against down-welling conditions (winter). Evaluating the gene expression during different localized oceanographic states such as upwelling and downwelling, is necessary to describe the underlying factors for the correlation of fishery and zooplankton abundances with these

conditions. It is expected that transcripts relating to lipid & energy synthesis will be upregulated during a period of overall upwelling (relaxed and active upwelling). Target transcripts for lipid synthesis will be down-regulated during down-welling conditions and stress transcripts will be up-regulated.

Chapter 2: Materials and Methods

Sample Collection

C. pacificus samples were captured along the NHL at station NH5 that is 9.26 kilometers offshore of Newport, Oregon from July 2016-Feburary 2018 approximately every two weeks, with sampling occurring only during conditions of safe weather and no mechanical issues. The zooplankton were collected using a 0.5m diameter, 202-µm plankton net (Fisher et al. 2015). The plankton net was dragged through the water column from a few meters above the sea floor (~50m depth) to the sea surface; the volume was estimated with a TKS flowmeter (Fisher et al. 2015). The samples were immediately placed into RNAlater® to preserve in situ gene expression of the copepods. The environmental conditions were collected along with the samples of zooplankton. Temperature, salinity, oxygen, and fluorescence was measured by a CTD (Conductivity, Temperature, and Depth) Rosette.

Identification of Copepods & RNA Extraction Preparation

Prior to sequencing, the samples were sorted to retrieve the *C. pacificus* from the other species of zooplankton captured in the plankton tow. The sample of the desired date was poured from the original 50 ml FALCON® tube to a large petri dish (100x15mm). Within the large petri, the *C. pacificus* were morphologically identified (Frost 1974) and removed from the large petri dish into a smaller petri dish with RNAlater®. Within the small petri dish, the morphological identification was verified using a dissecting microscope and to determine length more precisely. Iced RNAlater® was used throughout the sorting process as the medium. Forceps and disposable pipettes were stored in clean RNAlater. Each disposable pipette was labeled with the sample identification and used only for that sample to prevent cross-sample contamination. Portions of the sample that have been sorted were stored in a separate FALCON® tube labeled with the original sample and marked. Approximately 10-20 individuals, with a size of at least 50µm, were selected, and placed into a single microcentrifuge tube with TRIZol and stored at -80°C.

RNA extraction

Following storage at -80°C samples were thawed and homogenized for the sequencing procedure using the protocol of Zhang et al. (2013). Homogenization was achieved through use of additional TRIzol and a micro-pestle to break down copepod carapaces and tissue (of the 10-20

pooled copepods), rinsing down the micro-pestle with TRIzol upon removal. Then, 40μl of chloroform was added to the sample to induce phase separation. The mixture was vortexed for 30 seconds. The mixture was centrifuged for 20 minutes at 10,200g at 4°C. The centrifugation resulted in a layered solution. The supernatant was extracted and placed into another tube. Within the new tube, phenyl chloroform was added in equal parts, which is approximately 120µl, to separate the nucleic acids from lipids and proteins. To mix the mixture thoroughly, the mixture vortexed for 30 seconds. Then, the centrifuge was used at 10,200g at 4°C for 5 minutes to facilitate the separation of the phases. After separation is complete, the supernatant was extracted for placement in another tube. Then, ethanol (100%) was added to the supernatant in equal parts, approximately 95µl. The new mixture was vortexed for 10 seconds to mix the supernatant and ethanol, then centrifuged for 10 seconds to remove any additional excess from the sides of the tube. The Direct-zol RNA MiniPrep Protocol was utilized to continue the preparation of the RNA (Instruction Manual Direct-zol RNA Mini Prep, 2012). The mixture was placed into a column and centrifuged for 30 seconds at room temperature at 10,200g. After using the centrifuge, 400µl of RNA Wash Buffer was added to the mixture. Another cycle of centrifugation occurred with the same conditions as previously mentioned above to remove contaminants from the RNA. Then, a master mix of 80µl, comprised of 75µl of Digestion Buffer and 5µl of DNase, was added to each sample, and the sample incubated for 15 minutes. Upon the completion of incubation, 400µl of Pre-Wash Buffer was added to the mixture and placed into another centrifugation cycle. The column was transferred to a new tube, and 700µl of RNA Wash Buffer was added to mixture. A centrifugation cycle proceeded under the room temperature condition for 2 minutes at 10,200g. The column was transferred to a new tube and proceeded to be eluted with 50µl of DNase/RNase Free Water. Another centrifugation cycle proceeded at room temperature for 40 seconds at 10,200g. The columns were removed, and the tubes sealed before transport to the Center for Genome Research and Biocomputing (CGRB) for sequencing.

Sequencing & Bioinformatic Analysis

The sequencing for this project was completed at the CGRB using paired-end high-throughput sequencing on the Illumina High-Seq 3000. RNA-Seq was more effective than genetic microarrays at sequencing the numerous amounts of copepods than at one time without sacrificing reliability (Reidy et al. 1995). RNA-Seq was the chosen method due the lower false positive rate as compared with microarrays and better detection of lower expressed transcripts

(Mortazavi et al. 2008). The paired-end sequencing produced 150bp reads for each strand. Paired-end sequencing Quality filtering was conducted on the raw reads using Skewer v.0.2.2 (Jiang et al. 2014). The quality filtering was conducted at a Q score of 20 for a 99.7% accuracy rate. Then, a de novo transcriptome assembly was conducted with Trinity v.2.6.6 (Trapnell et al. 2012). The transcriptome was mapped and quantified with Salmon v.0.10.0. An index was created with the assembled de novo reads with Salmon. Then, the raw reads were mapped to the assembled transcriptome. Salmon used the index and read map to quantify the raw reads to the map (Patro et al. 2017). The result is a quantification file which contains the gene identification number generated from Trinity, read length, and number of reads mapped to the gene. The quantification files are necessary to conduct differential gene expression analysis. Prior to conducting differential gene expression analysis, a tx2gene file was created using the transcript-to-gene-map created from the Trinity assembly. The transcript-to-gene-map was loaded into Excel to switch the columns containing the gene and transcript/transcript-isoform names and saved as a text file. After creating the tx2gene file, a sample file was created using Excel. The sample file contained the name of quant file for each sample in the first column with the corresponding sample condition in the second column and year in the third column. Upon the creation of these files, the R package 'readr' was used to import the sample and tx2gene file. The quantification counts from Salmon were imported via 'tximport' package to R. Then, DESeq2 v.1.18 (Love et al. 2014) was used to calculate differential expression in the form of log fold change, generate MA plots, conduct Principal Component Analysis (PCA) and visualize the scatter of samples on the first two principal components. Upwelling and downwelling oceanographic conditions were compared for all evaluations. The oceanographic states and environmental conditions are described in Table 1.

Functional annotation

To conduct functional annotations, the transcripts determined significant based on the generalized linear model (GLM) in DESeq2 were sorted and ranked based on log fold change from largest to smallest. These transcripts were placed in a list, and a R script was utilized to pull only these significant transcripts from the corresponding Trinity assembly to generate a new fasta file (see appendix A). The new fasta file composed of only the significant transcripts from DESeq2 were blasted (National Center for Biotechnology Information, Johnson et al. 2008) against a database of reviewed arthropod proteins retrieved from Uniprot database (UniProt

Consortium 2021). This database contained 6,193,248 reviewed protein sequences. The matches had to meet a threshold e-value of 0.05. Top hits (up to five for both upregulated and downregulated) were recorded along with other parameters in tables under their corresponding graphs sorted by log fold change rank. Some comparisons did not have any hits to this database, and thus no details could be recorded. My study did not query any non-protein-coding databases.

Evaluation of alternate oceanographic states

Sampling dates and oceanographic states are detailed in Table 1. The first four figures (Figures 1-4) depict the environmental conditions at the time of sampling which were recorded from the rosette and reported via the Newportal Blog from the Northwest Fisheries Science Center (NOAA Fisheries, 2021) Gene expression findings were evaluated at four different levels:

1. **Seasonal evaluation** was conducted for each year. Samples taken during upwelling conditions (both active and relaxed) were pooled and compared against pooled samples taken during downwelling.
2. **Active Alternate States** compared active upwelling and active downwelling samples collected within the same year.
3. **Oceanographic state intensity** compared samples taken within the same year of active versus relaxed states from both upwelling and downwelling states.
4. **Gross pool** combined data across years to compare the four pooled upwelling samples with the four pooled downwelling samples.

Table 1a: All samples used in the evaluation are present along with respective oceanographic condition of the copepods sampled at that date. The categories of PDO and ENSO are arbitrary for the purpose of DESeq2 categories. The categories of no, neutral, weak, and strong are the same for PDO and ENSO. The no category is defined by the conditions with numerical values of 0-0.1. The weak category is defined by the numerical values between 0.11-0.49. The neutral category is defined by 0.5-0.59. The strong category is defined by 0.6-1.2.

Sample Date	**Oceanographic Condition**	**PDO**	**ENSO**
		Temp|Value|Category	**Temp|Value|Category**
8/1/2016	Active upwelling	warm |0.52 |neutral	cold |-0.6|strong
9/20/2016	Relaxed upwelling	warm |0.45 | weak	cold |-0.7|strong
12/6/2016	Active downwelling	warm |1.17 |strong	cold |-0.6 |strong
2/17/2017	Relaxed downwelling	warm|0.7 |strong	cold |-0.3 |weak
8/8/2017	Active Upwelling	warm| 0.09 |no	cold |-0.1|weak
9/6/2017	Relaxed upwelling	warm |0.32 |weak	cold |-0.4|weak

12/4/2017	Active downwelling	warm |0.5 |neutral	cold |-1.0 |strong
2/27/2018	Relaxed downwelling	warm |0.32 |weak	cold |-0.8 |strong

Table 1b: Analysis evaluations descriptions stating which samples were incorporated in each evaluation.

Analysis Comparison	Anaylsis	Samples Pooled
Seasonal Evaluation	Seasonal 2016-2017: upwelling vs downwelling	8/1/2016 & 9/20/2016 vs. 12/6/2016 & 2/17/2017
Seasonal Evaluation	Seasonal 2017-2018: upwelling vs downwelling	8/9/2017 & 9/6/2017 vs. 12/4/2017 & 2/28/2018
Active Alternate States	2016: active upwelling vs active downwelling	8/1/2016 vs. 12/6/2016
Active Alternate States	2017: active upwelling vs active downwelling	8/9/2017 vs. 12/4/2017
Oceanographic State Intensity	In-Year 2016: active upwelling vs relaxed upwelling	8/1/2016 vs. 9/20/2016
Oceanographic State Intensity	In-Year 2016-2017: active downwelling vs relaxed downwelling	12/6/2016 vs. 2/17/2017
Oceanographic State Intensity	In-Year 2017: active upwelling vs relaxed upwelling	8/9/2017 vs. 9/6/2017
Oceanographic State Intensity	In-Year 2017-2018: active downwelling vs relaxed downwelling	12/4/2017 vs. 2/27/2018
Climatic Evaluation	Pooled Downwelling: active downwelling vs relaxed downwelling	12/6/2016 & 12/4/2017 vs. 2/17/2017 & 2/27/2018
Climatic Evaluation	Pooled Upwelling: active upwelling vs relaxed upwelling	8/1/2016 & 8/9/2017 vs. 9/20/2016 & 9/6/2017

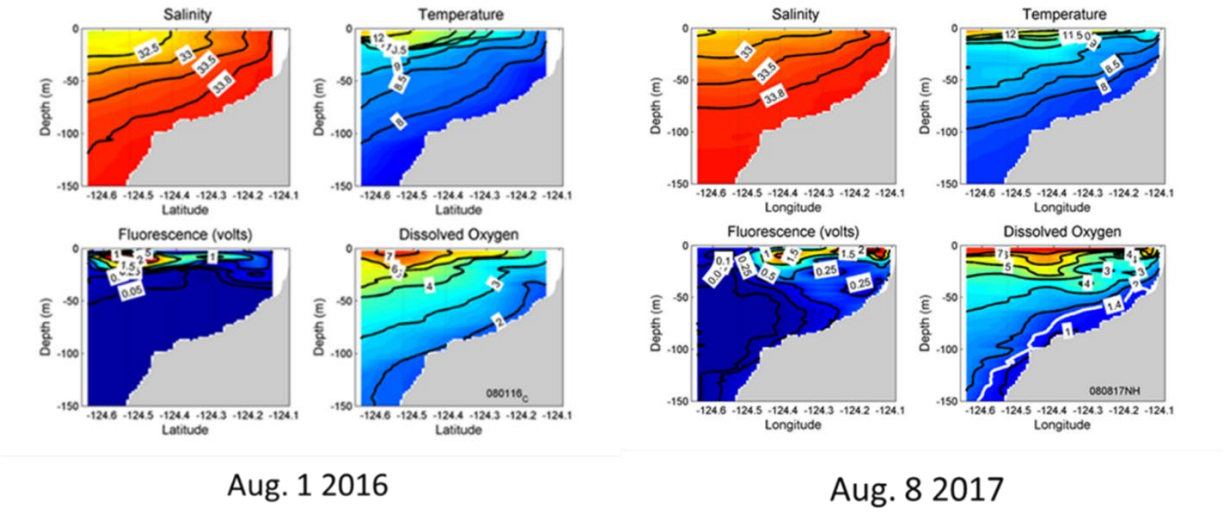

Figure 1: Environmental conditions at the time of active upwelling sampling were taken with a rosette. The figure depicts salinity, temperature, fluorescence, and dissolved oxygen during active upwelling. The x axis for all charts in figures 1 through 4 should be labeled Longitude (some have Latitude in error).

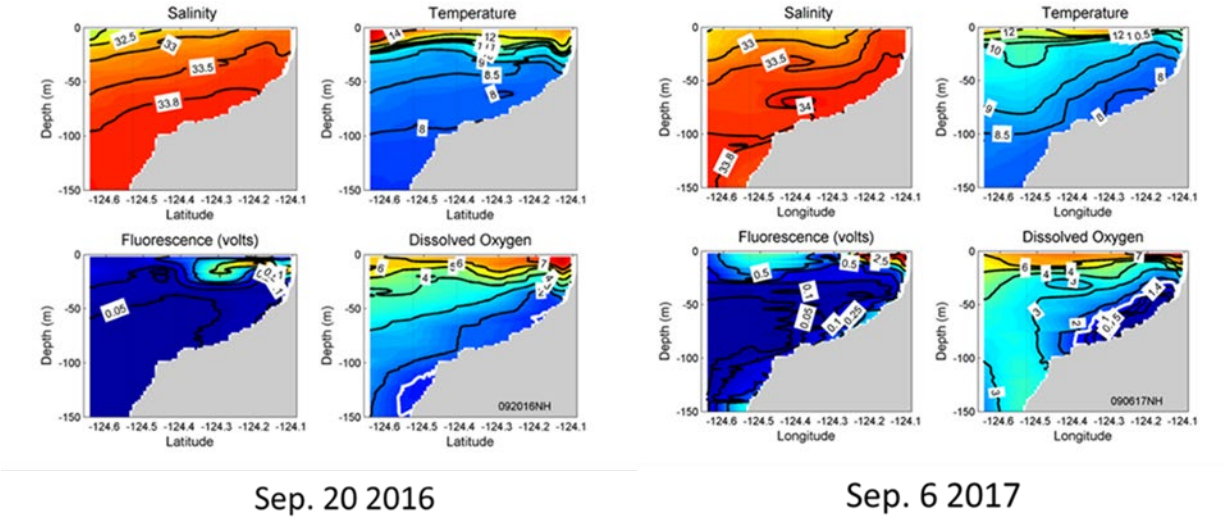

Figure 2: Environmental conditions taken at the time of relaxed upwelling sampling with a rosette. The figure depicts salinity, temperature, fluorescence, and dissolved oxygen during relaxed upwelling. The x axis for all charts in figures 1 through 4 should be labeled Longitude (some have Latitude in error).

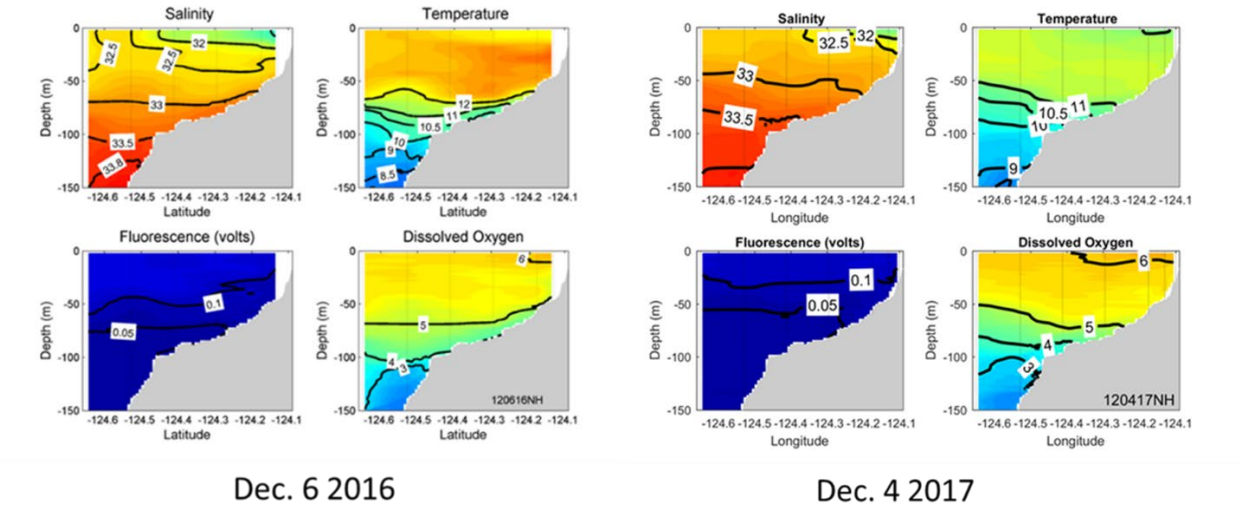

Figure 3: Environmental conditions taken at the time of active downwelling sampling with a rosette. The figure depicts salinity, temperature, fluorescence, and dissolved oxygen during active downwelling. The x axis for all charts in figures 1 through 4 should be labeled Longitude (some have Latitude in error).

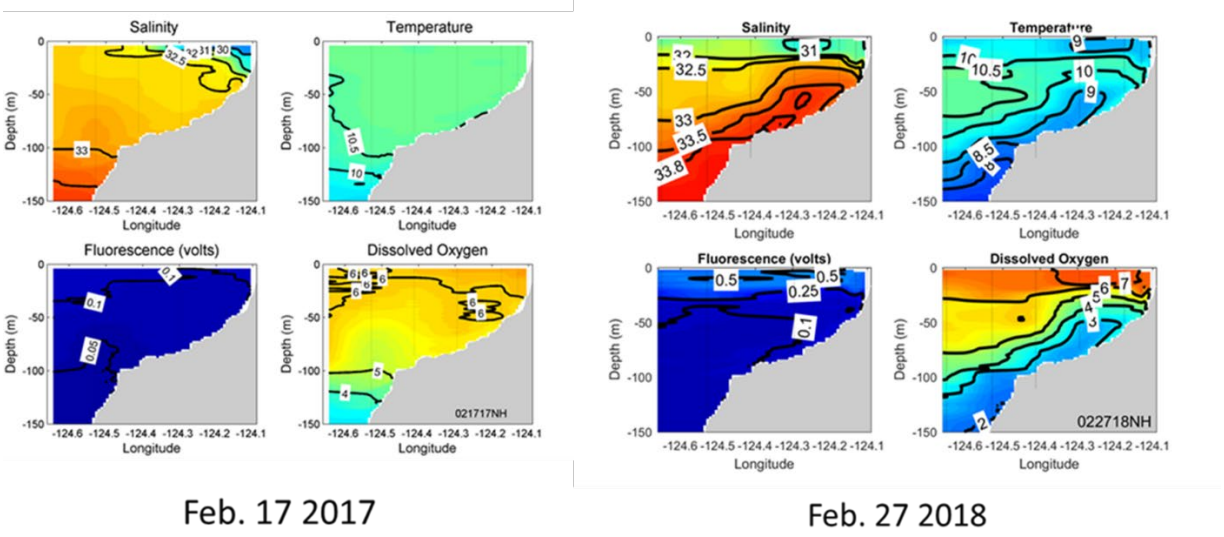

Figure 4: Environmental conditions taken at the time of relaxed downwelling sampling with a rosette. The figure depicts salinity, temperature, fluorescence, and dissolved oxygen during relaxed downwelling. The x axis for all charts in figures 1 through 4 should be labeled Longitude (some have Latitude in error).

Chapter 3: Results

Evidence for differential gene expression across all levels of comparison among samples from 2016-2018 are visualized using MA plots. These graphs plot log fold change in gene expression against mean normalized counts. RNAseq tags within the normalized distribution are depicted in black and those representing significant expression ($\alpha<0.05$) appear in red. Blast match details ranking the most extreme cases among significant log fold expression differences are presented in tables following MA plots. These tables reflect the transcripts that could be annotated using the Uniprot database. Some of the transcripts yielded no matches that met the e-value threshold of 0.05, and were not recorded as annotations. Inference on relationship to oceanographic and climatic state are best described, drawn, and discussed at the four levels of comparison as detailed in methods.

Seasonal Comparison

The Northeastern Pacific generally occupies an upwelling state in spring and through summer followed by a transition to downwelling in late fall and winter. Differential expression among RNAseq tags in this first analysis compared pooled upwelling (active – 8/1/16 and relaxed - 9/20/16) against pooled downwelling samples (active – 12/6/16 and relaxed – 2/7/17) (Figure 5). Most extreme among significant differential expressions evident were for upregulation of tags associated with genes involved in chitin degradation and oxygen transport. Those most extreme for downregulation were for aerobic respiration (Table 2).

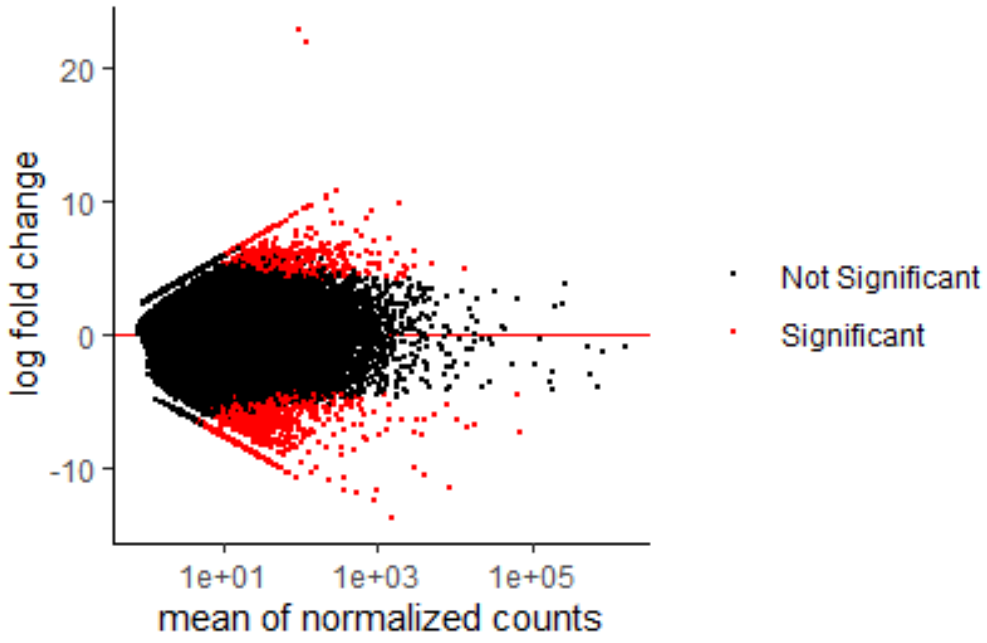

Figure 5: Log fold change among pooled upwelling verses downwelling seasons in 2016-2017.

Table 2: Top ranking (log fold change) among significant RNAseq tags for upwelling verses downwelling seasons in 2016-2017 following blast searches against reviewed arthropod proteins retrieved from Uniprot database. LFC = Log Fold Change, GO = Gene Ontology.

Rank	P-value	LFC	Transcript Trinity ID	Gene	Protein	GO Term (Biological)	Function	Regulation
6	2.21E-4	-10.422	135928_c6_g2_i1	mt:CoI	Cytochrome c oxidase subunit 1	GO:0009060 GO:0005507 GO:0005743 GO:0016021 GO:0070469	Aerobic respiration/electron transport coupled proton transport	Down
51	3.73E-3	-8.847	123290_c0_g1_i2	MT-CYB	Cytochrome b	GO:0009055 GO:0008121 GO:0016021 GO:0022904 GO:0045275 GO:0046872	Respiratory electron transport chain	Down
37	8.02E-3	-8.488	133552_c0_g1_i1	N/A	Sarcoplasmic calcium-binding protein	GO:0005509	Fast Muscle Contraction	Down
69	3.33E-3	-8.649	142085_c0_g1_i1	ND3	NADH-ubiquinone oxidoreductase chain 3	GO:0008137 GO:0016021 GO:0031966 GO:0070469	NADH dehydrogenase/Electron transport, Respiratory chain	Down
1	1.18E-10	23.584	139250_c3_g2_i1	ctcA	Endochitinase A	GO:0006032 GO:0004568 GO:0005576 GO:0005618 GO:0005886 GO:0006032 GO:0008061 GO:0031225	Chitin catabolic process/ polysaccharide catabolic process	Up
9	7.67E-4	8.974	131863_c0_g1_i5	N/A	Hemocyanin C chain	GO:0046872	Oxygen Transport	Up

Similar seasonal variation in between pooled upwelling (active condition -Aug. 9, 2017 and relaxed condition – Sept. 6, 2017) and downwelling (active condition -Dec. 4, 2017 and relaxed condition – Feb. 28, 2018) states were examined for the next year, 2017-2018 season (Figure 6). Most extreme among significant differential expressions evident were for genes involved with oxygen transport and chitin degradation (endochitinase 11) for downregulation as an immune response and an upregulation in chitin degradation (endochitinase 33) to achieve organismal growth and biosynthesis of highly reducing polyketide synthases (Table 3).

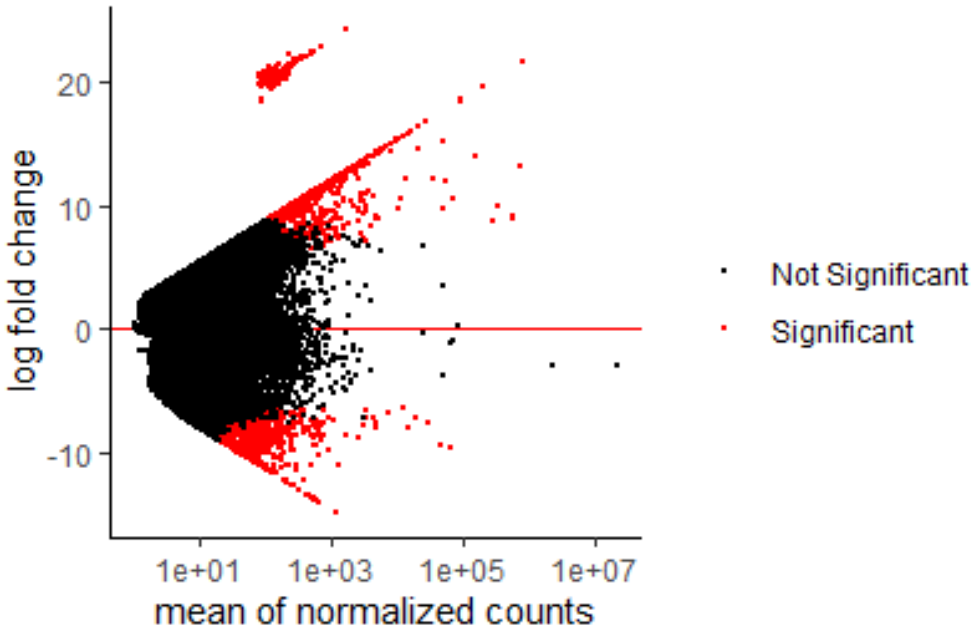

Figure 6: Log fold change among pooled upwelling verses downwelling seasons in 2017-2018.

Table 3: Top ranking (log fold change) among significant RNAseq tags for upwelling verses downwelling seasons in 2017-2018 following blast searches against reviewed arthropod proteins retrieved from Uniprot database. LFC = Log Fold Change, GO = Gene Ontology.

Rank	P-value	LFC	Transcript Trinity ID	Gene	Protein	GO Term (Biological)	Function	Regulation
11	1.21E-3	-13.081	157237_c5_g1_i1	N/A	Hemocyanin subunit 2	GO:0005344 GO:0005615 GO:0016491 GO:0046872	Oxygen transport in hemolymph of arthropods	Down
18	2.54E-3	-12.235	168255_c1_g2_i5	chi11	Endochitinase 11	GO:0000272 GO:0004568 GO:0005576 GO:0006032 GO:0008061 GO:0009405	Chitin degradation/infection process by direct penetration of the host cuticle.	Down
66	7.81E-3	-10.078	164539_c0_g1_i13	N/A	Hemocyanin units G and H	GO:0005344 GO:0016491 GO:0046872	Oxygen transport in hemolymph of arthropods	Down
76	5.03E-1	-3.275	153164_c0_g1_i1	HCB	Hemocyanin B chain	GO:0005344 GO:0005507 GO:0005615 GO:0016491 GO:0031404	Oxygen transport in hemolymph of arthropods	Down
196	3.00E-6	19.839	137348_c0_g1_i1	N/A	Sarcoplasmic calcium-binding protein, beta chain	GO:0005509	Fast Muscle Contraction	Up
223	2.96E-6	16.673	141680_c0_g1_i7	COII	Cytochrome c oxidase subunit 2	GO:0004129 GO:0005507 GO:0005743 GO:0016021 GO:0070469	last enzyme of mitochondrial electron transport chain driving oxidative phosphorylation	Up
269	2.54E-3	12.935	141676_c2_g1_i1	srdA	Highly reducing polyketide synthase	GO:0004315 GO:0006633 GO:0016491	Highly reducing polyketide synthase/ biosynthesis of sordarial	Up
440	4.71E-3	11.274	143982_c1_g1_i3	chit33	Endochitinase 33	GO:0000272 GO:0004568 GO:0005576 GO:0006032 GO:0008061 GO:0009405	Chitin degradation/ morphogenetic role during apical growth, cell division and differentiation	Up
507	7.40E-3	10.987	139332_c2_g8_i1	srdG	Epoxide hydrolase	GO:0016787	Highly reducing polyketide synthase/ biosynthesis of sordarial	Up

Active Alternate States

The seasonal evaluation of the gene expression of *C. pacificus* considered above included active and relaxed conditions within upwelling and downwelling states and compared them as seasons replicated over two sampling periods (2016-2017 and 2017-2018). The active states comparison considered below was to make a direct comparison of response to gene expression changes between active upwelling phase and active downwelling phase that occurred *within the same season*.

For the 2016-2017 sampling period, the active upwelling state (8/1/16) is contrasted against the active downwelling state (12/6/16). Differential gene expression among these active alternate oceanographic conditions depict a significant expression of the metabolism of cysteine (Figure 7). The cysteine metabolic process will lead to energy biosynthesis.

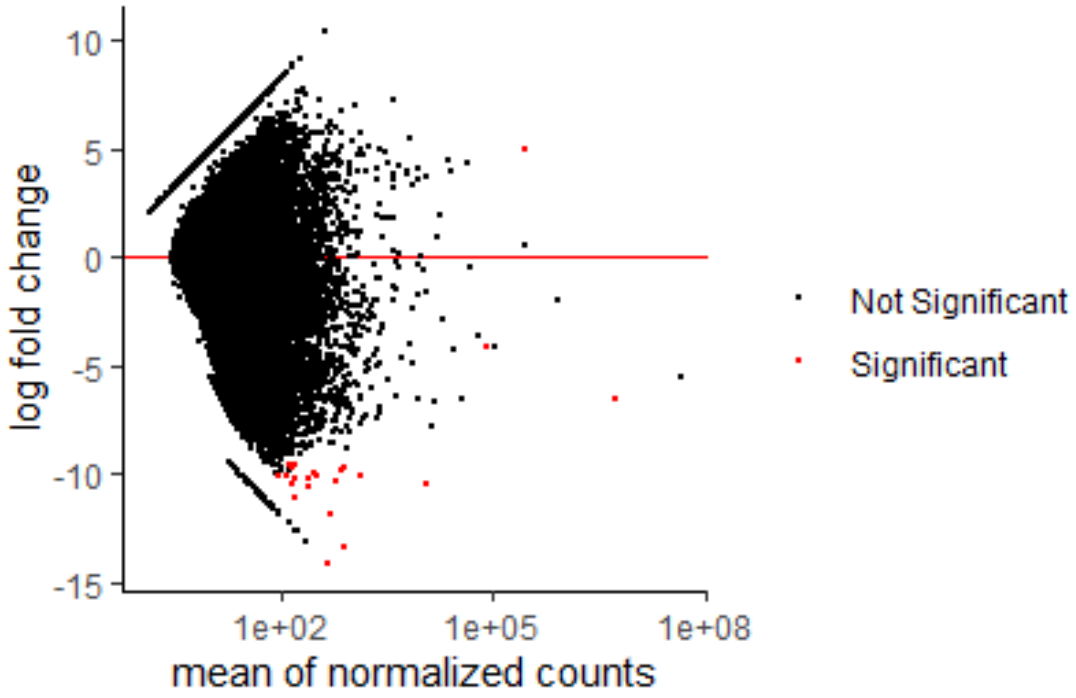

Figure 7: Log fold change between active upwelling against active downwelling.

For the 2017-2018 sampling alternate active states contrasting active upwelling (8/8/17) against active downwelling (12/4/17), it was shown that there were more significantly differentially expressed genes in this comparison than 2016-2017 comparison (Figure 8). The significantly up

or downregulated genes were involved in muscle movement and oxygen circulation. Table 4 describes transcripts able to be annotated through the reviewed Uniprot arthropod database.

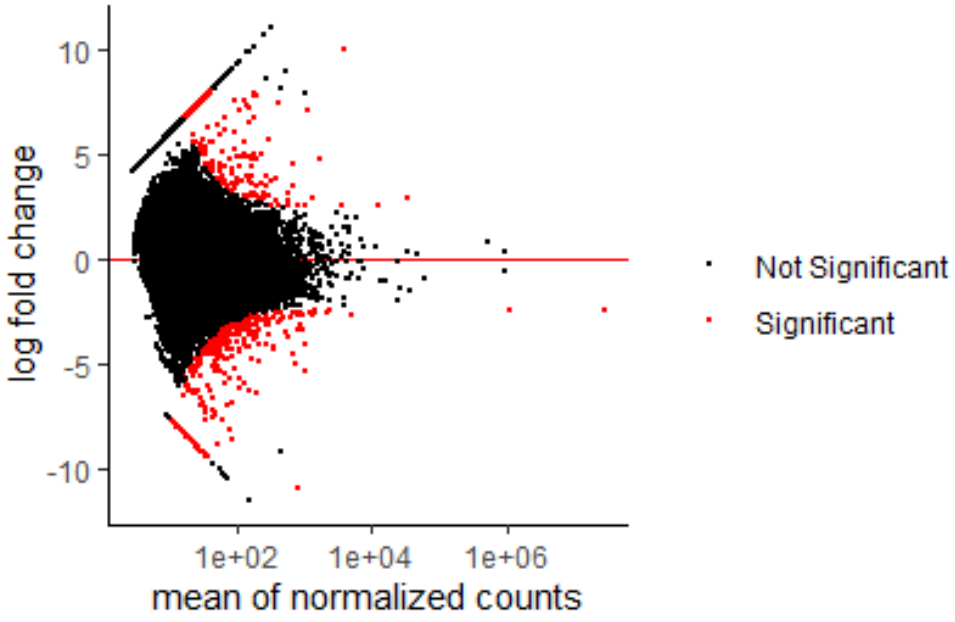

Figure 8: Log fold change between active upwelling against active downwelling states is shown above for the seasonal evaluation of 2017-2018. Significantly expressed genes are in red with a p-value of 0.05

Table 4: Top ranking (log fold change) among significant RNAseq tags for comparison between active upwelling against active downwelling states is shown above for the seasonal evaluation of 2017-2018. LFC = Log Fold Change, GO = Gene Ontology.

Rank	P-value	L FC	Transcript Trinity ID	Gene	Protein	GO Term (Biological)	Function	Regulation
2	3.99E-3	-7.645	11201_c0_g1_i2	N/A	Sarcoplasmic calcium-binding protein, beta chain	GO:0005509	Fast Muscle Contraction	Down
60	4.95E-3	-4.246	84545_c0_g1_i1	ODHCY	Hemocyanin G-type	GO:0016491 GO:0005344 GO:0015671 GO:0046872	Oxygen transport in hemolymph of arthropods	Down
39	8.05E-3	7.978	87780_c0_g1_i1	N/A	Hemocyanin, beta-C chain	GO:0016491 GO:0005344 GO:0015671 GO:0046872	Oxygen transport in hemolymph of arthropods	Up

Oceanographic State Intensity Comparison

Intensity of the oceanographic state may influence the gene expression of *C. pacificus*. We conducted a comparison of the active states against the relaxed states of both upwelling and downwelling within the same year to decipher if there were any difference in gene expression.

For the 2016-year, the active upwelling state (8/1/16) was compared against the relaxed upwelling state (9/20/16) to identify differences in gene expression (Figure 9). It was determined that expression in some genes were significantly altered. The genes with significantly altered expression were involved in muscle movement and regulation as well as oxygen transport and oxidative phosphorylation. (Table 5).

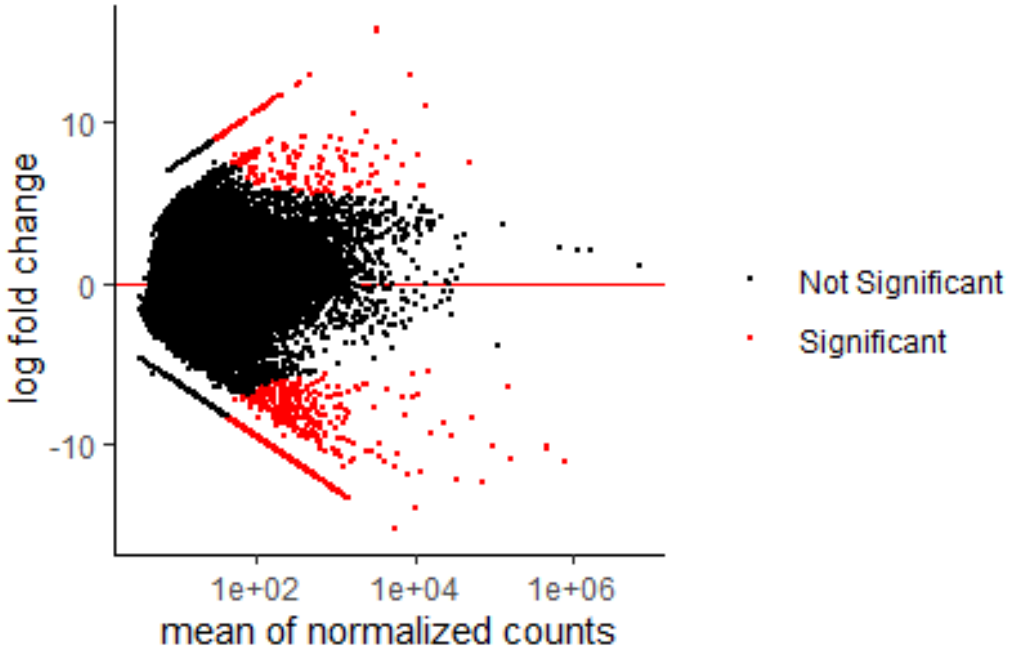

Figure 9: The log fold change of active upwelling compared to relaxed upwelling. The red dots represent an alpha of 0.05.

Table 5: Top ranking (log fold change) among significant RNAseq tags for comparison between active upwelling state against relaxed upwelling states within 2016. LFC = Log Fold Change, GO = Gene Ontology.

Rank	P-value	LFC	Transcript Trinity ID	Gene	Protein	GO Term (Biological)	Function	Regulation
31	1.47E-3	-11.300	102873_c3_g1_i1	nodZ	Cytochrome P450 monooxygenase	GO:0005506 GO:0020037 GO:0016705 GO:0004497 GO:0016491	Oxygen-reduction /heme binding	Down
54	7.49E-3	-10.577	104829_c2_g1_i5	srdI	FAD-linked oxidoreductase	GO:0016491 GO:0071949 GO:0050660	oxidation-reduction process	Down
57	7.36E-3	-8.694	104887_c2_g2_i4	KLH2	Hemocyanin 2	GO:0005344 GO:0005615 GO:0016491 GO:0046872	Oxygen transport	Down
64	6.23E-3	-10.817	105229_c1_g3_i1	N/A	Tropomyosin Lep s 1.0101 above.	GO:0006937 GO:0042803	regulate of muscle contraction	Down
72	9.53E-3	-10.266	105580_c2_g1_i2	N/A	Sarcoplasmic calcium-binding protein 1	GO:0005509	Fast Muscle Contraction	Down

The intensity of the downwelling states that occurred between December (active downwelling occurred on 12/6/16) and February (relaxed downwelling occurred on 2/17/17) of 2016-2017 sampling period were compared to determine any differences in gene expression (Figure 10). The most extreme among significant change in expressed genes were downregulated and involved in energy metabolism and immune response. The significant genes found in this comparison could not all be fully annotated using the Uniprot reviewed database for arthropods.

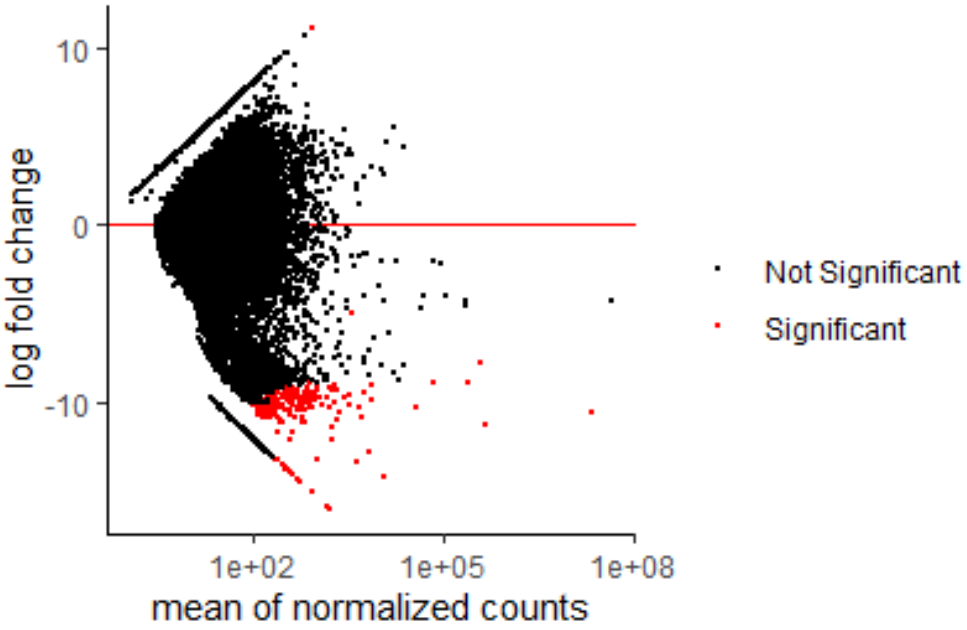

Figure 10: The log fold change depicted is describing the oceanographic condition of active downwelling against the condition of relaxed downwelling. The red dots represent significant genes of 0.05 or less.

The upwelling intensity comparison for the 2017-2018 sampling period contrasted active upwelling (8/8/17) against relaxed upwelling (9/6/17) and yielded several significant genes that were both upregulated and downregulated (Figure 11). The most extreme cases involved upregulation of genes associated with muscle movement and regulation and aerobic respiration, and downregulation of genes involved in oxygen transportation, aerobic respiration, and chitin degradation (Table 6).

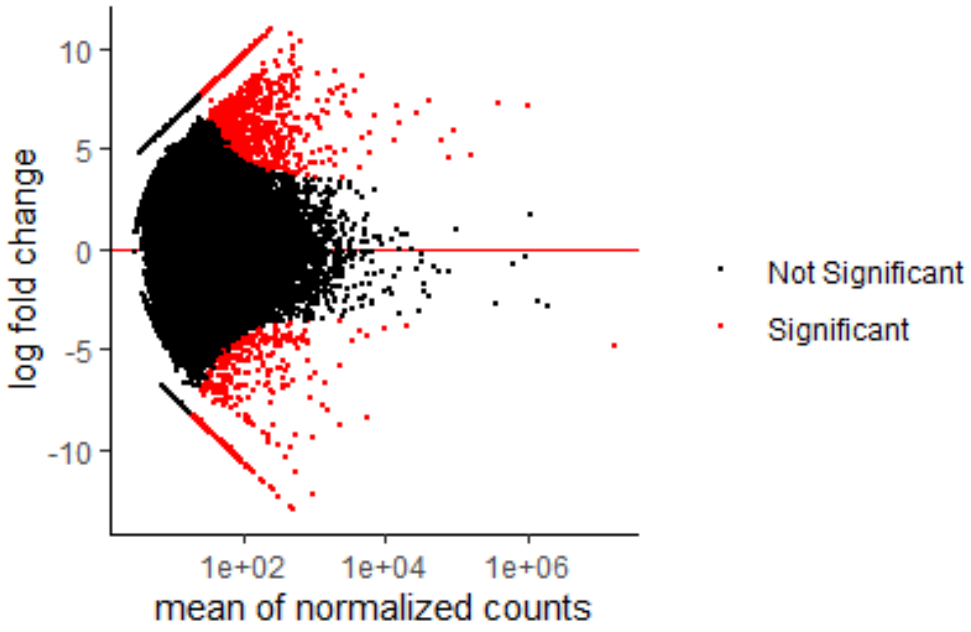

Figure 11: Log fold change between the oceanographic conditions of active upwelling against relaxed upwelling of 2017-2018. The red dots represent significantly expressed genes with a p-value of 0.05.

The downwelling states of the 2017-2018 year occurred during December of 2017 and February of 2018. The active (12/4/17) and relaxed (2/27/18) downwelling states were contrasted against one another to determine any changes in gene expression of *C. pacificus* (Figure 12). There were not many genes that were significantly differentially expressed. Of the genes that were significantly differentially expressed, the genes were involved in the processes of cellular respiration and energy metabolism. The identity of significant genes found in this comparison could not be determined from annotations using the Uniprot reviewed database for arthropods.

Table 6: Top ranking (log fold change) among significant RNAseq tags for comparison between Top ranking (log fold change) among significant RNAseq tags for comparison between active upwelling state against relaxed upwelling states within 2017-2018. LFC = Log Fold Change, GO = Gene Ontology.

Rank	P-value	LFC	Transcript Trinity ID	Gene	Protein	GO Term (Biological)	Function	Regulation
19	1.77E-3	10.260	121944_c0_g1_i1	N/A	Hemocyanin 2-c chain	GO:0016491 GO:0005344 GO:0015671	Oxygen transport in hemolymph of arthropods	Up
29	3.14E-3	8.356	118537_c0_g1_i1	N/A	Sarcoplasmic calcium-binding protein	GO:0005509	Fast Muscle Contraction	Up
43	6.88E-3	6.170	123574_c2_g2_i5	N/A	Tropomyosin Lep s 1.0101	GO:0006937 GO:0042803	Central in calcium dependent regulation of muscle contraction	Up
57	3.78E-3	5.841	120308_c1_g2_i2	N/A	Venom serine protease Bi-VSP	GO:0004252 GO:0006508 GO:0010952 GO:0008233	arthropod prophenoloxidase-activating factor triggering phenoloxidase cascade	Up
70	5.70E-3	5.511	124508_c3_g1_i3	ND5	NADH-ubiquinone oxidoreductase chain 5	GO:0070469 GO:0008137	Electron transport/ respiratory chain NADH dehydrogenase	Up
73	1.11E-3	-10.992	136106_c4_g4_i1	N/A	Hemocyanin A chain	GO:0005344 GO:0005615 GO:0016491 GO:0046872	Oxygen transport in hemolymph of arthropods	Down
80	5.74E-3	-9.801	130167_c0_g1_i1	N/A	Hemocyanin AA6 chain	GO:0005344 GO:0005507 GO:0005615 GO:0016491 GO:0031404	Oxygen transport in hemolymph of arthropods	Down
110	6.04E-3	-9.765	131388_c2_g2_i4	ctcA	Endochitinase A	GO:0000272 GO:0004568 GO:0005576 GO:0005618 GO:0005886 GO:0006032 GO:0008061 GO:0031225	Carbohydrate metabolism/ Chitin degradation	Down
115	8.77E-4	-7.447	135660_c0_g1_i1	COII	Cytochrome c oxidase subunit 2	GO:0004129 GO:0005507 GO:0005743 GO:0016021 GO:0070469	Aerobic respiration/ electron transport chain	Down

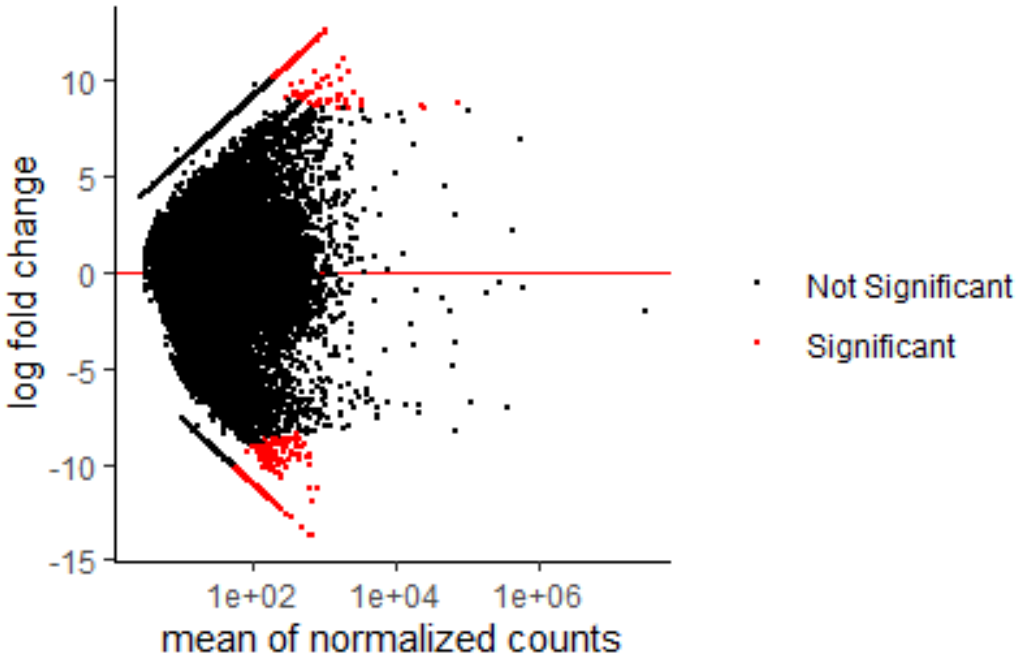

Figure 12: The log fold change of the active downwelling contrasted against relaxed downwelling. The red dots represent significantly expressed genes with a p-value of 0.05.

Gross Pool

As a final investigation, samples were pooled across all sampling years (2016-2017) to determine most extreme differences among expression patterns within the two primary oceanographic states. For downwelling, this included sampling dates from both active (12/6/16 and 12/4/17) and relaxed (2/1/17 and 2/27/18) intensities (Figure 13). The most extreme significant differences included downregulation of tags associated with aerobic respiration and upregulation of tags associated with egg production and oxygen circulation (Table 7).

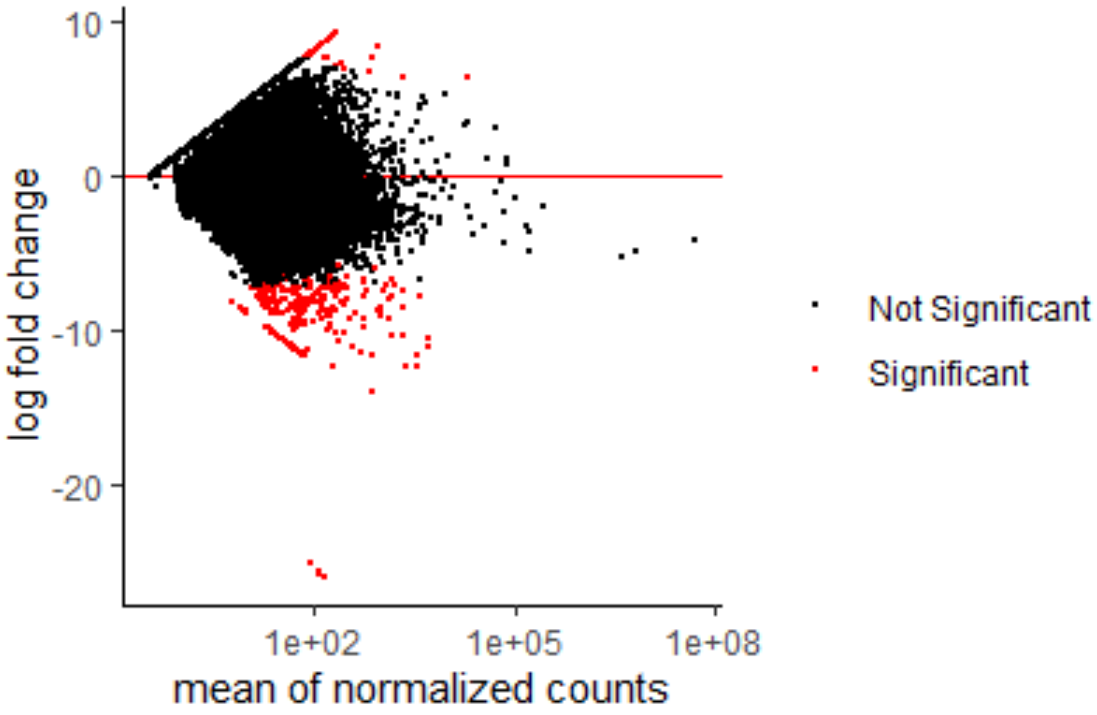

Figure 13: Log fold change among downwelling samples from all sampling years 2016-2018.

Table 7: Top ranking (log fold change) among significant RNAseq tags for comparison among pooled active and relaxed downwelling intensities for all sampling years. LFC = Log Fold Change, GO = Gene Ontology.

Rank	P-value	LFC	Transcript Trinity ID	Gene	Protein	GO Term (Biological)	Function	Regulation
39	1.77E-14	-25.269	182493_c2_g2_i1	MT-CYB	Cytochrome b	GO:0005743 GO:0008121 GO:0016021 GO:0022904 GO:0045275 GO:0046872	respiratory electron transport chain	Down
58	1.12E-14	-25.162	216021_c1_g1_i1	mt:CoI	Cytochrome c oxidase	GO:0004129 GO:0005751 GO:0006123 GO:0009060 GO:0015990 GO:0016021 GO:0020037 GO:0022904 GO:0045277 GO:0046872	Electron transport coupled proton transport	Down
104	7.64E-4	-9.906	221890_c1_g1_i4	MT-CYB	Cytochrome b	GO:0005743 GO:0008121 GO:0016021 GO:0022904 GO:0045275 GO:0046872	respiratory electron transport chain	Down
113	8.76E-3	-9.461	192021_c1_g1_i2	ND5	NADH-ubiquinone oxidoreductase chain 5	GO:0005743 GO:0008137 GO:0016021 GO:0070469	Respiratory electron transport chain	Down
19	1.96E-4	10.690	170171_c0_g1_i2	COII	Cytochrome c oxidase	GO:0004129 GO:0005507 GO:0005743 GO:0016021 GO:0070469	Electron transport chain	Up
22	4.22E-3	8.163	170569_c4_g2_i1	HCD	Hemocyanin D chain	GO:0005344 GO:0005507 GO:0005615 GO:0016491 GO:0031404	Oxygen transport in hemolymph of arthropods	Up
24	7.09E-3	7.575	181539_c1_g1_i15	mt:CoI	Cytochrome c oxidase	GO:0004129 GO:0005751 GO:0006123 GO:0009060	Electron transport coupled proton transport	Up
25	5.34E-3	7.559	171555_c0_g2_i1	Astl	Astacin-like metalloendopeptidase	GO:0004222 GO:0005737 GO:0005886 GO:0007155 GO:0008233	Oocyte-specific oolemmal receptor involved in sperm and egg adhesion and fertilization	Up
130	6.50E-3	7.196	171589_c1_g1_i5	HCA	Hemocyanin A chain	GO:0005344 GO:0005507 GO:0005615 GO:0016491 GO:0031404	Oxygen transport in hemolymph of arthropods	Up

Similar to the previous gross pool, this comparison utilizes samples across all years. All upwelling samples including the sampling dates from both active upwelling (8/1/16 and 8/8/17) and relaxed upwelling (9/20/16 and 9/6/17) (Figure 14). Most extreme among significant differences included both up- and down regulation of tags associated with egg production, aerobic respiration, oxygen circulation, and immune response (Table 8).

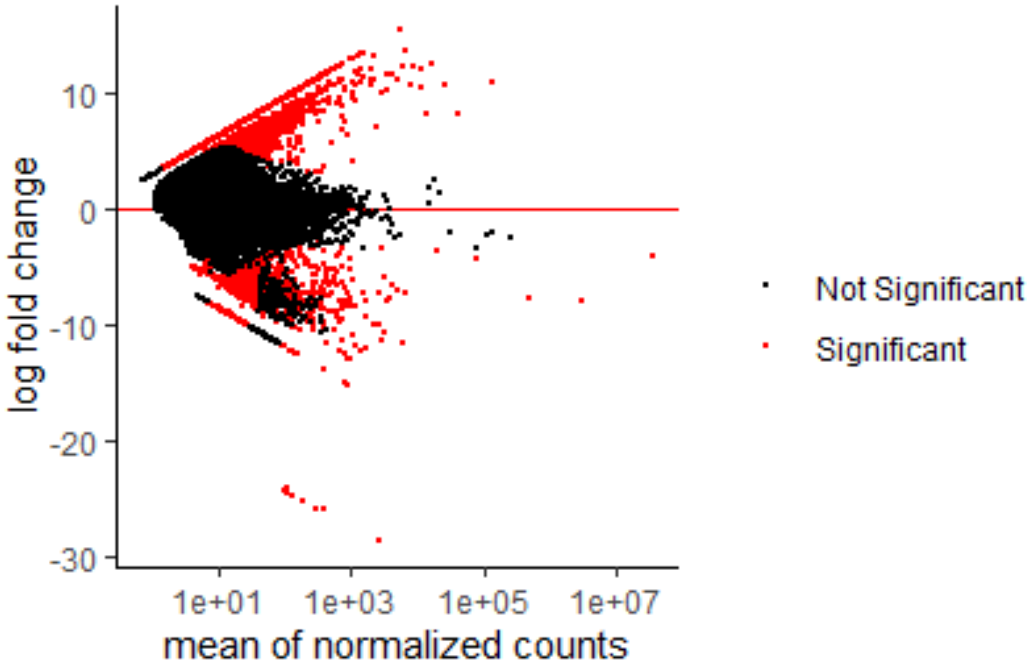

Figure 14: Log fold change among upwelling samples from all sampling years 2016-2018.

Table 8: Top ranking (log fold change) among significant RNAseq tags for comparison among pooled active and relaxed upwelling intensities for all sampling years (2016-2017). LFC = Log Fold Change, GO = Gene Ontology.

Rank	P-value	LFC	Transcript Trinity ID	Gene	Protein	GO Term (Biological)	Function	Regulation
38	5.69E-02	-9.569	65044_c0_g1_i1	N/A	Venom serine protease Bi-VSP	GO:0004252 GO:0005576 GO:0016504 GO:0046872 GO:0090729	arthropod prophenoloxidase-activating factor triggering phenoloxidase cascade.	Down
40	6.40E-3	-9.481	59702_c0_g1_i1	ND3	NADH-ubiquinone oxidoreductase chain 3	GO:0008137 GO:0016021 GO:0031966 GO:0070469	NADH dehydrogenase (Complex I)/ transfers electrons from NADH to the respiratory chain	Down
84	7.17E-3	-9.387	106979_c0_g1_i1	N/A	Hemocyanin A-type, units Ode to Odg	GO:0005344 GO:0016491 GO:0046872	Oxygen transport in hemolymph of arthropods	Down
115	1.70E-3	-8.441	23234_c0_g1_i1	N/A	Hemocyanin AA6 chain	GO:0005344 GO:0005507 GO:0005615 GO:0016491 GO:0031404	Oxygen transport in hemolymph of arthropods	Down
130	2.71E-3	-6.393	107252_c0_g1_i2	KLH1	Hemocyanin 1	GO:0005344 GO:0005615 GO:0016491 GO:0046872	Oxygen transport in hemolymph of arthropods	Down
13	1.14E-6	8.452	100645_c0_g1_i1	ND5	NADH-ubiquinone oxidoreductase chain 5	GO:0005743 GO:0008137 GO:0016021 GO:0070469	NADH dehydrogenase (Complex I)/ transfers electrons from NADH to the respiratory chain	Up
28	2.17E-3	7.860	101548_c0_g1_i1	srdE	Short-chain dehydrogenase	GO:0000140 GO:0004806 GO:0005783 GO:0005811 GO:0006654 GO:0016021 GO:0019433	phosphatidic acid biosynthetic process/ triglyceride catabolic process	Up
120	3.96E-3	7.611	105164_c0_g1_i1	Astl	Astacin-like metalloendopeptidase	GO:0004222 GO:0005737 GO:0005886 GO:0007155 GO:0008233 GO:0008270 GO:0009566 GO:0010954 GO:0030133 GO:0060468 GO:0060473 GO:0070001 GO:0070002 GO:2000360	Oocyte-specific oolemmal receptor involved in sperm and egg adhesion and fertilization.	Up
140	4.06E-3	7.527	104600_c0_g1_i1	URE G_04 748	Sphingomyelinase D	GO:0005576 GO:0009405 GO:0016042 GO:0046872 GO:0050290	lipid catabolic process/ pathogenesis/membrane destabilization/host cell penetration/pulmonary inflammation and cutaneous lesions	Up
141	8.74E-3	7.245	105156_c0_g1_i1	ODH CY	Hemocyanin G-type, units Oda to Odg	GO:0005344 GO:0016491 GO:0046872	Oxygen transport in hemolymph of arthropods	Up

Principal Component Analysis (PCA)

Although a total of 81% of the variance is explained by the first and second principal components, evidence for expected clustering among samples from common environmental conditions was scarce. The closest clustering among samples collected under similar conditions was observed between the two active upwelling samples. These samples are close along PC1 and essentially identical along PC2. One relaxed upwelling sample also clusters nearby, but overall there is a lack of strong clustering between different states of upwelling (active – 8/1/16 and 8/8/17 and relaxed – 9/20/16 and 9/6/17). In general, upwelling conditions are concentrated toward the upper portion of the y-axis and both active and relaxed states of downwelling (active – 12/6/16 and 12/4/17 and relaxed – 2/17/17 and 2/27/18) were clustered toward the lower end of the y-axis. Notably, the two active downwelling samples fall at the opposite extremes of PC1, indicating that the most substantial differences in gene expression among any of the eight samples occur between those two samples collected under ostensibly similar environmental conditions.

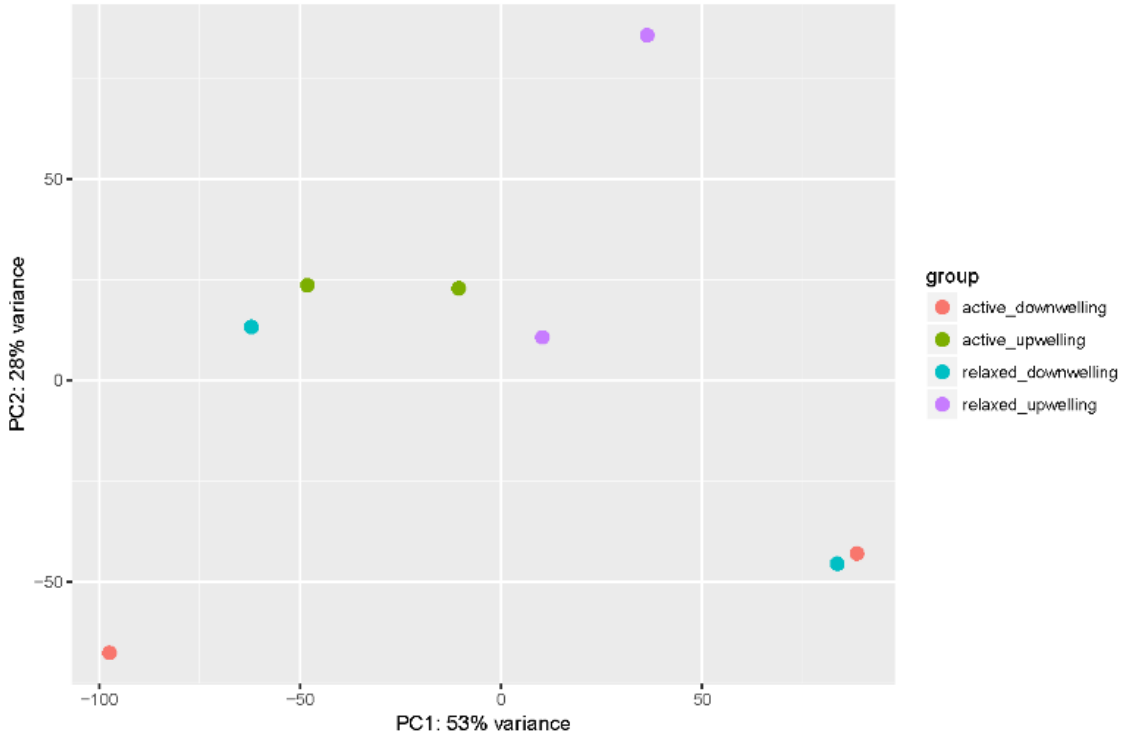

Figure 15: PCA of all samples showing the variance distribution expression intensity under different oceanographic state. Please note the greatest variation is between the two active downwelling samples.

Chapter 4: Discussion

Overall RNAseq findings from perspectives presented here were diverse and varied, making it difficult to draw inference on consistent relationships among the common environmental conditions considered. This, in part may result from difficulty in consistently annotating high ranking tags that had significant expression differences in the absence of any annotated genome available for *C. pacificus*, and generally limited annotation among genomes of related copepods. The best approach applied here was blasting findings against a database of reviewed arthropod proteins retrieved from Uniprot database. This meant that top ranking RNAseq tags detailed among tables were sparsely sampled across a range of the rank that may include up to ~600 tags, as a function of which RNAseq tags attained a blast hit. To accommodate this difficulty, much of this discussion will focus on overall trends from inferred function among those tags identified.

Seasonal Comparison

Although not immediately apparent in comparing figures 5 and 6 and tables 2 and 3, the seasonal evaluation demonstrated differential gene expression for similar biological processes (for example, energy metabolism) occurring during both the 2016-2017 and 2017-2018 sampling period for the comparison of upwelling against downwelling seasons. This was expected because similar environmental states are expected to yield similar overall responses from *C. pacificus.* Among key differences were that tags associated with aerobic respiration were downregulated in 2016-2017 and those associated with oxygen transport upregulated, while in 2017-2018 tags associated with oxygen transport were downregulated and organismal growth was upregulated.

Active Alternate States

The evaluation of the active alternate states for the 2016-2017 sampling period did not meet the threshold of significance (p-value of 0.01) necessary to determine full annotation, thus only the active alternate states sample for 2017-2018 is available. Active upwelling (8/8/17) against active downwelling (12/4/17) demonstrated a downregulation of muscle movement and oxygen transport throughout the copepods (Figure 11, Table 6). However, there was an upregulation of oxygen transport as well. This can be attributed to different steps in the hemocyanin pathway being upregulated while another is downregulated. The upregulated step has a greater impact due to its higher log fold change. It is unclear how the upregulation or downregulation of specific step informs the overall availability of adequate food sources.

Oceanographic State Intensity

Oceanographic state intensity contrasts active and relaxed states of the same oceanographic condition such as upwelling or downwelling. Oceanographic intensity of the downwelling in both 2016-2017 and 2017-2018 could not be fully annotated because transcripts did not meet the necessary threshold of significance of 0.01. The upwelling comparison, however, did meet the threshold enabling annotation of transcripts for both 2016 and 2017. The 2016 comparison showed a downregulation of muscle movement, oxygen transport and aerobic respiration from active upwelling (8/1/16) to relaxed upwelling (9/20/16). This suggests *C. pacificus* exhibits lower energy output during the period of relaxed upwelling than during active upwelling (Figure 12, Table7). This difference is likely due to reduced availability of food sources as the relaxed upwelling process provides fewer nutrients to the surface (Pascual et al. 2003). Similar physiological responses were observed in 2017 (Figure 14, Table 8). Downregulation of the aerobic respiration was present as well as oxygen transport; suggesting some similar energy conservation was underway. The difference in 2017 occurred in the upregulation of transcripts involved in muscle contraction regulation, muscle movement, and immune response of phenoloxidase cascade which produces a lethal toxin to parasites (Sin et. al, 2015). The

upregulation of these biological processes suggests that less energy conservation occurred, due to a need for an immune response to protect against parasitic organisms.

Gross Pool

When all the downwelling samples were pooled together and evaluated (Figure 17, Table 10), the most obvious downregulation occurred among transcripts associated with Complex I and III of the mitochondrial electron transport chain (Garrett, Grisham, 2016) while other portions of the this biological process were upregulated. This result suggests *C. pacificus* was not starting the process of ATP synthesis and completed a previous process of ATP synthesis. Elements of Complex V (electron transport via oxidative phosphorylation) as described in Berg, Tymoczko, and Stryer (2002) of the electron transport chain was both downregulated and upregulated. Cytochrome c oxidase is capable of either catalyzing cellular respiration or regulating cellular respiration (Brown and Bren, 2008). However, it is unclear in whether cytochrome c oxidase was catalyzing cellular respiration or limiting cellular respiration. A more in-depth study on marine copepods would be necessary to determine how cytochrome c oxidase behaves in marine copepods under various environmental conditions. Upregulation of oxyhemocyanins in the form of hemocyanin D and hemocyanin A indicated increased oxygen transport throughout the organisms. In addition to the increased oxygen transport, tags that associated with Astacin-like metalloendopeptidase that have function associated with oocyte oolemma receptor functions were upregulated. This is indicative of sperm and egg adhesion which contributes to gamete production. It was expected that gamete production would begin around relaxed downwelling to prepare for spawning during spring months in which conditions would be favorable to juveniles due to growth and molting due to increased temperature (Vidal, 1980). However, this same egg production process was not expected to be observed in the pooled upwelling evaluation in which only relaxed upwelling and active upwelling were included (Figure 18, Table 13). An upregulation of egg production was observed during upwelling periods, which further supports previous observations off the California coast that *C. pacificus* are biphasic (Ohman et. al, 1998). For the pooled upwelling samples, upregulation of Complex I in the electron transport chain was observed indicating the start of cellular respiration (Garrett, Grisham, 2016). Catabolism of triglycerides and lipids signals the deconstruction of the cellular membrane. The deconstruction

of the cellular membrane could be a symptom of gamete production, but it is unclear. Within the pooled upwelling evaluation, downregulation of oxygen transport was observed which could indicate a reduction in food availability or food sources that are not as nutritious (Pascual et al. 2003). Immunity response in the form of phenoloxidase cascade was downregulated in this evaluation which could be indicative of a successful immune response after a viral or parasitic infection (Sin et. al. 2015).

Principal Component Analysis (PCA)

It was expected that common oceanographic upwelling (active – 8/1/16 and 8/8/17 and relaxed – 9/20/16 and 9/6/17) and downwelling (active – 12/6/16 and 12/4/17 and relaxed – 2/17/17 and 2/27/18) states would cluster together in PCA. Although a large component of the variance is apportioned to the first (53%) and second (28%) axes, the resulting distribution patterns differed from expectations. It was expected that upwelling samples would cluster together with particularly close clustering among the active upwelling samples. A similar phenomenon was expected among the downwelling samples. Although the abiotic conditions were similar for the active downwelling samples, there may be other underlying factors causing greater dispersion than expected, such as difference in food availability, given that a whole year separates many of the common sampling conditions. Among weaker evidence for clustering, it is notable that closest neighbors among common conditions were for active upwelling, from each of the two sampling periods (2016-17 and 2017-18).

Chapter 5: Conclusions

I began with an expectation that this study would be able to provide insight that could potentially enhance the Peterson et al. (2014) stoplight assay (NOAA 2021). Primary to this expectation was that upregulation of RNAseq tags associated with long-chain fatty acids and energy synthesis pathways along with downregulation of heat shock proteins (indicative of lower stress) would correlate with high fishery yields for chinook salmon (*Oncorhynchus tshawytscha)* and coho (*Oncorhynchus kisutch*) salmon. Yet there is a lack of direct evidence for any of these specific pathways among evidence presented above. Increase or decrease of RNAseq tags associated with oxygen transport, however, could be indicative of changes in food source availability and quality (Pascual et al. 2003) and were evident throughout this study. Oxygen transport was downregulated (seasonal evaluation 2017-2018, active alternate states 2017-2018, oceanographic state intensity 2016, oceanographic state intensity 2017, and gross pool of upwelling conditions) overall suggesting decreased food availability or quality which does correlate with the high mean rank (poor forecast) observed for juvenile catch of chinook observed in 2016, 2017, and 2018 (NOAA 2021). Juvenile coho salmon has an identical correlation, except for 2018 which indicated a low mean rank (good forecast) for future years. Also, the immunity responses observed in the *C. pacificus* may be an indicator of parasites and virus infections that could affect the food chain negatively, leading by observing the poor yield of ichthyoplankton during 2016, 2017, and 2018 and lower catch of the juvenile salmon owing to lower quality of food source due to infection. However, more studies would have to be conducted to confirm this possibility. The egg production observed in the pooled upwelling and pooled downwelling samples does not appear to indicate or correlate with positive forecasting due to the overall forecast of each year from 2016-2018 being a poor year. The egg production observed may only be indicative of the reproductive health of the copepod population, and thus may not reflect overall ecosystem health.

Another reason why my study may not have revealed any consistent evidence for upregulation of RNAseq tags associated with long-chain fatty acids and energy synthesis pathways is that the study was conducted during a time frame in which there was a complete lack of positive forecasts for the salmon returns. To overcome this, the study should be continued to determine

baseline observations during positive forecasts yield allowing contrast of those results against the results presented here. Thus, a longer-term study would need to be conducted using a mixture of positive forecasts and negative forecasts for salmon returns to determine how well the *C. pacificus* transcriptome can enhance the predictions of the stoplight assay.

It should also be noted that more studies should be conducted to characterize and annotate the *C. pacificus* genome. That improved annotation would be useful in accounting for variation in copepod subpopulation structure and sampling bias and would permit more accurate analysis of RNAseq data such as that considered here. Within this study, the *C. pacificus* transcriptome was annotated using the available reviewed proteomes of other crustaceans. Perspectives from such an approach are limited because it the databases do not include certain proteins that may only be present in copepods or differ functionally among *Calanus* copepods.

References

Berg, J.M., Tymoczko, J.L., Stryer, L. (2002) Section 18.4, A proton gradient powers the synthesis of ATP. Biochemistry, 5th edition. New York: W H Freeman. Available from: https://www.ncbi.nlm.nih.gov/books/NBK22388/

BioBam Bioinformatics (2019) OmicsBox – Bioinformatics made easy (Version 1.2.4.). Retrieved March 3, 2019, from https://www.biobam.com/omicsbox.

Bowman, S.E., Bren, K.L. (2008) "The chemistry and biochemistry of heme c: functional bases for covalent attachment". Natural Product Reports. 25 (6): 1118–30. doi:10.1039/b717196j

Fisher, J.L., Peterson, W.T., Rykaczewski, R. (2015) The impact of El Nino events on the pelagic food chain in the northern California Current. Global Change Biology, 21, 4401-4414. doi:10.1111/gcb.13054

Frost, B.W., (1974) *Calanus marshallae*, a new species of calanoid copepod closely allied to the sibling species *C. finmarchicus* and *C. glacialis*. Marine Biology, Berlin 26(1):77-99, figs. 1-21, tabs. 1-5. (12-vii-1974)

Garrett, R., Grisham, C.M. (2016) Biochemistry. Boston: Cengage. p. 687. ISBN 978-1-305-57720-6.

Hays, G., Richardson, A., Robinson, C. (2005) Climate change and marine plankton. Trends in Ecology & Evolution, 20(6), 337-344. doi:10.1016/j.tree.2005.03.004

Jiang, H., Lei, R., Ding, S. et al. (2014) Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. BMC Bioinformatics 15, 182. https://doi.org/10.1186/1471-2105-15-182

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., Madden, T.L. (2008) NCBI BLAST: a better web interface. Nucleic Acids Res. 2008 Jul 1;36(Web Server issue):W5-9. doi: 10.1093/nar/gkn201. Epub 2008 Apr 24. PMID: 18440982; PMCID: PMC2447716.

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

Losee, J. P., Miller, J. A., Peterson, W. T., Teel, D. J., & Jacobson, K. C. (2014). Influence of ocean ecosystem variation on trophic interactions and survival of juvenile coho and Chinook salmon. Canadian Journal of Fisheries and Aquatic Sciences, 71(11), 1747–1757. https://doi.org/10.1139/cjfas-2014-0043

NOAA (2021) https://www.fisheries.noaa.gov/west-coast/science-data/ocean-ecosystem-indicators-pacific-salmon-marine-survival-northern Last updated by Northwest Fisheries Science Center on 04/15/2021.

Fisheries, NOAA (2021) The Newportal Blog - Northwest Fisheries Science Center NOAA https://www.fisheries.noaa.gov/content/newportal-blog-northwest-fisheries-science-center

Pascual, C., Gaxiola, G., Rosas, C. (2003) "Blood metabolites and hemocyanin of the white shrimp, *Litopenaeus vannamei*: The effect of culture conditions and a comparison with other crustacean species". Marine Biology. 142 (4): 735. doi:10.1007/s00227-002-0995-2

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(4), 417–419. https://doi.org/10.1038/nmeth.4197

Peterson W.T., Fisher, J.L., Peterson, J.O., Morgan, C.A., Burke, B.J., Fresh, K.L. (2014) Applied fisheries oceanography: Ecosystem indicators of ocean conditions inform fisheries management in the California Current. Oceanography 27(4):80–89, http://dx.doi.org/10.5670/oceanog.2014.88.

Pundir, S., Martin, M.J., Odonovan, C. (2016) UniProt Tools. Current Protocols in Bioinformatics. doi:10.1002/0471250953.bi0129s53

Reidy, M., Timm, E. Jr, Stewart, C. (1995) Quantitative RT-PCR for measuring gene expression. Biotechniques, 18(1), 70-74. Retrieved May 16, 2018.

Roncalli, V., Christie, A.E., Sommer, S.A., Cieslak, M.C., Hartline, D.K., Lenz, P.H. (2017) A deep transcriptomic resource for the copepod crustacean *Labidocera madurae*: A potential indicator species for assessing near shore ecosystem health. PLoS ONE 12(10): e0186794. https://doi.org/10.1371/journal.pone.0186794

Sin, Y.W., Kenny, N.J., Qu, Z., Chan, K.W., Chan, K.W., Cheong, S.P., Leung, R.W., Chan, T.F., Bendena,, W.G., Chu, K.H., Tobe, S.S., Hui, J.H. (2015) Identification of putative ecdysteroid and juvenile hormone pathway genes in the shrimp *Neocaridina denticulata*. General and comparative endocrinology, 214, 167–176. https://doi.org/10.1016/j.ygcen.2014.07.018

Trapnell, C., Roberts, A., Goff, L. et al. (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 7, 562–578. https://doi.org/10.1038/nprot.2012.016

UniProt Consortium (2021) UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res. 2021 Jan 8;49(D1):D480-D489. doi: 10.1093/nar/gkaa1100. PMID: 33237286; PMCID: PMC7778908.

Vidal, J. (1980) Physioecology of zooplankton. I. Effects of phytoplankton concentration, temperature, and body size on the growth rate of *Calanus pacificus* and *Pseudocalanus* sp.. Marine Biology. 56, 111–134 https://doi.org/10.1007/BF00397129

Zhang, H., Finiguerra, M., Dam, H., Lin, S. (2013) An improved method for achieving high-quality RNA for copepod transcriptomic studies. Journal of Experimental Marine Biology and Ecology, 446, 57-66. doi:10.1016/j.jembe.2013.04.021

Zhang, X., Yuan, J., Li, F., Xiang, J. (2021) Chitin synthesis and degradation in crustaceans: a genomic view and application. Marine Drugs, 19(3), 153. https://doi.org/10.3390/md19030153

APPENDICES

Appendix A: Command-Line Workflow

Quality Filtering: Skewer

Quality filtering excludes noise within the dataset. The contamination is limited after quality filtering and improves accuracy of downstream anaylsis. Quality filtering is performed at a Q score of 20. The Q score of 20 is equivalent to 99.7% accuracy. The filter removes low quality base calls with 1 in 100 failure rate.

Script:

Base:

skewer -y <barcode> -q <Q score> -o <output name added> <paired-reads1.fastq> <paired-reads2.fastq>

Example:

skewer -y ATCACG -q 20 -o trim_index lane7-s013-indexRPI1-ATCACG-2_S13_L007_R1_001.fastq lane7-s013-indexRPI1-ATCACG-2_S13_L007_R2_001.fastq

Assembly: Trinity

Organizes reads into a transcriptome. The output is a transcriptome. This is a de novo assembly.

Script:

Base:

/local/cluster/trinityrnaseq-2.6.6/Trinity --seqType fq --left R1.fastq --right R2.fastq --output trinity_out_dir --CPU 6 --max_memory 20G

Trinity --seqType fq --left R1.fastq --right R2.fastq --output trinity_out_dir --CPU 6 --max_memory 20G

Example:

/local/cluster/trinityrnaseq-2.6.6/Trinity --seqType fq --left seasonal_16-17_R1.fastq --right seasonal_16-17_R2.fastq --output seasonal_16-17_trinity_out_dir --CPU 6 --max_memory 20G

Trinity --seqType fq --left seasonal_16-17_R1.fastq --right seasonal_16-17_R2.fastq --output seasonal_16-17_trinity_out_dir --CPU 6 --max_memory 20G

Map/Quantify

It will create an index in the form of a directory. Thus, do not look for the file, and simply use the directory name to continue quasi-mapping.

Program: Salmon

Index Script:

Base:

salmon index -t {transcriptome} -i {index name}

Example:

salmon index -t Trinity.fasta -i UATrinity_index

Quantification Script:

Base:

salmon quant -i {index} -l A -1 {fastq file paired 1} -2{fastq file paired 2} -p {thread #} --validateMappings –gcBias -o {directory/name of output}

Example:

salmon quant -i UATrinity_index -l A -1 s013_trim_R1.fastq -2 s013_trim_R2.fastq -p 8 --validateMappings --gcBias -o quants/Sept_20_16_sample_gcBias_quant

Appendix B: R Workflow

DEseq2 Workflow

1. Load Libraries into R
    - library(readr)
    - library(DESeq2)
    - library(ggplot2)
    - library(tximport)
2. Transfer corresponding trinity.fasta.gene_trans_map using WinSCP to performed assembly
3. Open gene_trans_map in Excel to switch the columns. Labels with isoforms are to be the first column. Labels without isoforms are to be the second column. Insert a header that reads Transcript for column A and Gene for column B.
4. Save the file as tx2gene suffix.
5. Create sample text file
    - Use Excel
    - Column A = quant file name
    - Column B = condition (active or relaxed) (strong PDO, weak PDO, no PDO)
    - Column C = year
    - Save as text file
6. Set a working directory

- Should have tx2gene file, sample file, and quant files

7. Load tx2gene file
    - tx2gene <- read.table(file.path(“ <file path> /tx2gene.txt“), header = TRUE
8. Load sample file
    - samples <- read.table(file.path(" <file path>", "samples.txt"), header=TRUE)
9. Identify the file path for quantification files
    - files <- file.path("<file path> ", samples$Sample)
10. Name the quant files to be imported
    - names(files) <- paste0(c("quant.sf"))
11. Import quant file data
    - txi <- tximport (files, type ="salmon", tx2gene= tx2gene)
12. Begin DESeq2 – Identify the data and compare conditions (do this for each condition)
    - ddsTxi <- DESeqDataSetFromTximport(txi,

      colData = samples,

      design = ~Condition_1)
13. Remove low counts (example uses the number 10 but could be any number)
    - keep <- rowSums(counts(ddsTxi)) >= 10

ddsTxi <- ddsTxi[keep,]

14. Remove levels with no samples (if needed)
    - ddsTxi$Condition <- droplevels(ddsTxi$Condition)
15. Create dds object
    - ddsTxi <- DESeq(ddsTxi)

16. Get results (do this for each condition)

- res_condition_1 <- results(ddsTxi)
- resOrdered_condition_1 <- res_condition_1[order(res_condition_1$pvalue),]

17. See results in table

- res_condition_1 (within R)
- resOrdered_condition_1 (within R)
- write.csv(as.data.frame(res_condition_1), file = "Feb_replicate_C1results.csv")
- write.csv(as.data.frame(resOrdered_condition_1), file = "Feb_replicate_pvalue_ordered_C1results.csv")

18. Graph Log Fold Change with Shrinkage

- resultsNames(ddsTxi)
- resLFC_C1 <- lfcShrink(ddsTxi, coef= "Condition_1_weak_PDO_vs_strong_PDO", type=c("normal"))
- resLFC_ashr_C1 <- lfcShrink(ddsTxi, coef = "Condition_1_weak_PDO_vs_strong_PDO", type =c("ashr"))
- plotMA(resLFC_C1, ylim=c(-5,5), alpha=0.05)
- plotMA(resLFC_ashr_C1, ylim=c(-2,2), alpha=0.05)

19. Plot Principal Component Analysis

- ddsTxi_transform <- estimateSizeFactors(ddsTxi)
- se <- SummarizedExperiment(log2(counts(ddsTxi_transform, normalized = TRUE) + 1), colData = colData(ddsTxi_transform))
- plotPCA(DESeqTransform(se), intgroup=c("Condition_1"))

Extracting Significant Transcripts Workflow

1. Load libraries

```
library(Biostrings)

library(here)
```

```
# From: https://bioinformatics.stackexchange.com/questions/595/how-can-longest-isoforms-per-gene-be-extracted-from-a-fasta-file
```

2. Read in wanted sample IDs

```
wanted_ids <- read.table(here("C:/Users/Kenya Bynes/Documents/2021/Masters Degree/Research/Bioinformatics/Active_States16-17", "active_states16-17_downregulated_sig_C1genes.txt"),
                         header = TRUE)
```

3. read your fasta in as Biostrings object

```
fasta.s <- readDNAStringSet(here("C:/Users/Kenya Bynes/Documents/2021/Masters Degree/Research/Bioinformatics/Active_States16-17", "active_states_16-17_Trinity.fasta"))

names.fasta <- names(fasta.s)
```

4. extract only the relevant gene and isoform id - IN THIS CASE everything is by a '_', THIS MIGHT NOT WORK FOR ALL GENOMES
5. split by " " so we get the full name

```
gene.iso <- sapply(names.fasta,function(j) cbind(unlist(strsplit(j,'\\ '))[c(1,2)]))
```

6. make a dataframe

```
gene.iso.df <- data.frame(t(gene.iso))
colnames(gene.iso.df) <- c('genes.t','len')
```

7. Again, now by "_"

```
plz <- sapply(gene.iso.df$genes.t, function(j) cbind(unlist(strsplit(j,'\\_'))))
plz.df <- data.frame(t(plz))
colnames(plz.df) <- c('gene.t','zero', "one", "two", "isoform")
```

8. Put whole gene name back together

```
plz.df$gene <- paste(plz.df$gene.t, plz.df$zero,
                     plz.df$one, plz.df$two, sep = "_")
```

9. Make new dataframe

```
iso.info <- data.frame(whole_name = rownames(gene.iso.df),
                       gene = plz.df$gene,
                       isoform = plz.df$isoform)
```

10. Get length

```
iso.info$width <- width(fasta.s)
rownames(iso.info) <- iso.info$whole_name
```

11. split data.frame into list with entry for each gene

```
iso.info.split <- split(iso.info,iso.info$gene)
```

12. optional to keep all the information but really you just need indices

```
gene.iso.df.split.best <- lapply(iso.info.split,function(x) x[order(x$width)[1],])
```

13. pull out the longest isoform ID for each gene (in case of a tie just take the first one)

```
best.id <- sapply(iso.info.split,function(x) row.names(x)[order(x$width)[1]])
```

14. Now filter these ID's for down regulated genes

```
best.id.f <- best.id[which(names(best.id) %in% wanted_ids$Downregulated)]
```

15. subset your original reads with the subset

```
fasta.s.best <- fasta.s[best.id.f]
```

16. export new fastafile containing longest isoform per gene

```
writeXStringSet(fasta.s.best, filepath=here("C:/Users/Kenya Bynes/Documents/2021/Masters Degree/Research/Bioinformatics/Active_States16-17", "active_states16-17_downregulated_c1_Trinity-filtered.fasta"))
```

17. Repeat with upregulated significant tags and subsequent conditions.

MA plot Workflow

1. Load Libraries

```
library(ggplot2)
library(dplyr)
```

2. Set working directory

```
setwd("C:/Users/Kenya Bynes/Documents/2020/Research/Bioinformatics/Seasonal2016-2017")
```

3. Load csv file from DESeq2

```
seasonal1617results <- read.csv(file = "C:/Users/Kenya
Bynes/Documents/2020/Research/Bioinformatics/Seasonal2016-
2017/active_vs_relaxed_seasonal16-17_results_res.csv", header = TRUE, row.names = 1)
```

4. Remove columns for log fold change and p-value and place them into a table.

```
df <- data.frame(log10FC = (seasonal1617results$log2FoldChange),
                 logpv = (seasonal1617results$pvalue))
```

5. Graph with ggplot2

```
ggplot(df, aes(x=log10FC, y=logpv)) +
  geom_point(size=0.5)


dat <- seasonal1617results %>%
  dplyr::mutate(logFC = log2FoldChange,
                logpv = pvalue,
                meanfoldchange = baseMean,
                signif = if_else(pvalue < 0.05, 'Significant', 'Not Significant'))
```

```
ggplot(dat, mapping = aes(x = meanfoldchange,
                          y = logFC,
                          color = signif)) +
  theme_bw() +
  theme(panel.border = element_blank(), panel.grid.major = element_blank(), panel.grid.minor = element_blank(), axis.line = element_line(color = "black")) +
  geom_hline(yintercept = 0, color = 'red') +
  geom_point(size=0.5) +
  scale_color_manual(values = c('black', 'red'), name = '') +
  # coord_cartesian(xlim = c(0,1000)) +
  xlab(label = 'mean of normalized counts')+
  ylab(label = 'log fold change')+
  scale_x_log10(labels = scales::scientific_format())
```