2016

Temperature Effects on the Proteome of *Ciona intestinalis*

Chelsea Lopez

University of Rhode Island, cstephens@my.uri.edu

Follow this and additional works at: <https://digitalcommons.uri.edu/theses>

Recommended Citation

Lopez, Chelsea, "Temperature Effects on the Proteome of *Ciona intestinalis*" (2016). *Open Access Master's Theses*. Paper 1081.

<https://digitalcommons.uri.edu/theses/1081>

This Thesis is brought to you for free and open access by DigitalCommons@URI. It has been accepted for inclusion in Open Access Master's Theses by an authorized administrator of DigitalCommons@URI. For more information, please contact digitalcommons-group@uri.edu.

TEMPERATURE EFFECTS ON THE PROTEOME OF
CIONA INTESTINALIS

BY
CHELSEA LOPEZ

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
IN
BIOLOGICAL AND ENVIRONMENTAL SCIENCES

UNIVERSITY OF RHODE ISLAND

2016

MASTER OF SCIENCE
OF
CHELSEA LOPEZ

APPROVED:

Thesis Committee:

Major Professor	Steven Irvine
	Niall Howlett
	Joanna Norris

Nasser H. Zawia
DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND
2016

ABSTRACT

Ciona intestinalis is a solitary ascidian that lives in temperate waters around the globe. *C. intestinalis* in Rhode Island exhibit normal embryonic development at water temperatures between 10° C and 18° C. This thesis is interested in what changes occur in the ovary tissue at temperatures above 18° C and below 10° C. A preliminary study to assess the reproductive fitness of *C. intestinalis*, found that *C. intestinalis* raised at 22 °C had lower embryo viability when compared to the normal temperature range. 22 °C was chosen to mimic the high temperature predicted in some climate change models (IPCC, 2014). During the coldest part of the winter when temperatures drop below 8 °C, feeding slows, the production of gametes ceases, and much of the ovary tissue is resorbed (Dybern, 1965). Ovary samples were collected from animals reared at 18 °C, 22 °C, and from a non-temperature controlled tank at 8 °C. The samples were processed and sequenced by mass spectrometry. The data showed a decrease in the number of proteins produced between 18 °C and 22 °C samples. It also showed that there were more proteins upregulated in 22 °C compared to the 18 °C samples. Over the winter, the number of proteins also decreases when compared to 18 °C, and metabolic pathways were downregulated. Due to the lack of methods for ecological studies dealing with proteomic data, new methods were developed for both tissue and data processing.

ACKNOWLEDGMENTS

I would like to acknowledge my major professor, Steven Irvine, for allowing me to work on the project and all his help along the way. In addition, I would like to thank my committee for taking the time to read my thesis, and for all the support you have given me. Thanks to the undergraduates in my lab for helping to monitor the animals and helping with anything I needed. Thank you to the Experimental Program to Stimulate Competitive Research (EPSCoR) and the Rhode Island Science and Technology Advisory Council (STAC) for funding this research and my research assistantship. Lastly, I want to thank my friends and family for supporting me through all the long days and all the stress. “This material is based upon work conducted at a Rhode Island NSF EPSCoR research facility, the Genomics and Sequencing Center, supported in part by the National Science Foundation EPSCoR Cooperative Agreement #EPS-1004057.”

PREFACE

This thesis consists of two chapters. Each chapter is on a different project. Chapter 1 is written in the form of a scientific paper with an introduction, materials and methods, results, and discussion sections. Chapter 2 is similar to the first chapter and uses many of the same techniques with comparisons to data from chapter 1. However, this project poses a different research question. A third project is described in Appendix 2 and is unrelated to the other chapters.

TABLE OF CONTENTS

ABSTRACT	ii
ACKNOWLEDGMENTS	iii
PREFACE.....	iv
TABLE OF CONTENTS.....	v
LIST OF TABLES	vi
LIST OF FIGURES	vi
CHAPTER 1	1
Ocean Warming effects on <i>Ciona intestinalis</i> Proteome	1
CHAPTER 2	36
Changes in the <i>Ciona intestinalis</i> Proteome due to Overwintering	36
APPENDICES	69
BIBLIOGRAPHY	34, 67

LIST OF TABLES

TABLE	PAGE
Table 1. Five proteins deemed significant by 2 Tailed student T test performed by Proteomics core at Rhode Island Hospital.	24
Table 2. Proteins associated with signaling and transcription factors for both up and down regulated proteins.	28
Table 3. Proteins unique to the 8 °C samples	45
Table 4. Proteins unique to the 18 °C samples	47
Table 5. Difference between samples in number of proteins in each BLAST2GO functional category	55
Table 6. KEGG terms (protein Pathways) found only in cold-water samples	57
Table 7. KEGG terms (protein Pathways) found only in 18° C samples.....	59
Table 8. KEGG terms (protein Pathways) found in both of the samples	63

LIST OF FIGURES

FIGURE	PAGE
Figure 1. Fold changes between means of protein quantifications at 22 °C vs. 18 °C.	18
Figure 2. Reproductive stress test methods diagram.....	20
Figure 3. Data from the reproductive stress test.....	22
Figure 4. Heat map of peak intensity measurements	26
Figure 5. Work flow for the mass spectrometer data analysis	28
Figure 6. Bar chart of the GO data	33
Figure 7. Pie chart showing the number of unique proteins at each temperature	66

CHAPTER 1

Ocean Warming Effects on the *Ciona intestinalis* Proteome

Abstract

A four-degree increase in global ocean temperature is predicted by some climate change models (IPCC, 2014) and even small increases in temperature have been shown to have negative effects on animal physiology. A preliminary fertilization study to assess the reproductive physiology of *Ciona intestinalis* found that when reared at 22 °C they had lower embryo viability when compared to those raised at 18 °C (unpublished study). For this experiment, *C. intestinalis* were reared in two groups. Group 1 was reared at 18 °C (maximum temperature for normal development) and Group 2 at 22 °C (anticipated temperature due to climate change). Once sexually mature, ovaries were dissected and frozen. Protein was extracted from the tissue and sent to be sequenced on a tandem mass-spectrometer. There were about 151 upregulated proteins and 17 downregulated proteins out of the 1616 total proteins identified at a 1% False Discovery Rate in the ovary samples. Analysis of the data generated a set of GO terms (gene ontology) for the up and down regulated proteins using BLAST2GO. Gene ontology was used to identify the function of selected proteins. PANTHER, an online resource, was used to determine protein pathways while Ghost KOALA was used to visualize these pathways. The analysis revealed that the ubiquitin-mediated proteolysis pathway was overrepresented in the data, suggesting that protein damage was significantly increased in the high temperature condition.

Introduction

Ocean warming can have a profound effect on the physiology of marine organisms. It can have dramatic effects on metabolism and reproduction in addition to causing an overall stress response (Dorts, et al., 2012). In most organisms, exposure to both chronic and acute stressors decreases reproductive capacity (Einarsson, et al., 2008). It is predicted in some climate change models that ocean temperature is expected to rise at least 4 degrees Celsius in the next 100 years (IPCC, 2014). Marine organisms often use temperature as a cue to spawn or begin migration. These temperature cues usually correspond with seasonal changes. These seasonal changes include changes in sunlight and therefore, changes in food availability to provide the necessary energy for reproduction. Lower temperature is also important in triggering diapause, which happens over winter (Caceres, 1997). Because of climate change, marine organisms will experience higher temperatures, and it will be important to assess the physiological effects.

The organism used for this study was *C. intestinalis*, a type of solitary ascidian. Most populations of *C. intestinalis* are found in cold temperate water, and are present on six of the seven continents (Dybern, 1965). In the population local to Rhode Island, the highest water temperature they are likely to encounter is 25 °C in the summer. They survive and reproduce best at a maximum temperature of around 18 °C (Berrill, 1947). Cold-water *C. intestinalis* have a one to two year life span. The first generation will spawn during the summer months producing many offspring that will grow to maturity by the end of summer. This second generation will go dormant over the winter months in order to reproduce in the spring. The first generation will often not

survive the winter. The third generation that was spawned at the end of summer will remain in juvenile form, which is more resilient to the cold temperatures (Dybern, 1965). Most of the Mediterranean population dies off in the summer, then those that remain spawn in the spring and the fall (Dybern, 1965). If the cold-water animals were to be subjected to a temperature over 25 °C for an extended period, the population may die back, mimicking the life cycle of the Mediterranean population.

C. intestinalis is an invasive species that can out-compete local fauna for space. It is a common fouling organism often found on boats and fishing equipment (Carver, et al., 2006). *C. intestinalis* is a filter feeder eating plankton and algae floating in the water column. When the water temperature is decreased below a certain limit *C. intestinalis* lowers its food intake by slowing the beating of its cilia (Petersen, et al., 1999). If the water temperature never reaches this lower limit, *C. intestinalis* will continue to feed and this will decrease the food supply of the other filter feeding local fauna.

C. intestinalis is also a model organism for developmental genetics, and its genome has been completely sequenced making it a good candidate for proteomics studies (Endo, et al., 2011). It is important to have a fully sequenced and annotated genome when performing a proteomics study, allowing for a larger base of predicted proteins to map peptides.

Due to new orbitrap tandem mass spectrometry technology, it is possible to assess which proteins an organism is expressing globally. This gives information about the biological processes that the animal is performing (Tomanek, 2011). Instead of first separating proteins on a 2D PAGE gel, shotgun proteomics uses total tissue lysate for

analysis. After the lysate is prepared, a tryptic digest is performed to break the proteins into peptides. The peptides are sequenced. The sequences obtained should be representative of the proteins present in the tissue (Zhang, et al., 2013). By comparing differences in protein quantity using peak intensity data given by the mass spectrometer, it is possible to compare the expression level of individual proteins among experimental samples (Zhang, et al., 2013). If there is a large fold change in peak intensity between the two samples, and consistency between the biological replicates, then this protein is considered upregulated or downregulated. Knowing which proteins are up and down regulated informs which biological processes are being modified in response to the increase in temperature. Overall, this work is attempting to discover the differences between the proteomes of the animals reared at normal and elevated temperatures. The hope is to attribute these differences to changes in physiology.

Materials and Methods

Study Species:

The species used for this study is *Ciona intestinalis*, sp. B. All of the animals that were spawned for use in this experiment came from Point Judith Marina on Point Judith Pond, in Wakefield, Rhode Island.

Set up of Experiment:

At the URI Graduate School of Oceanography (GSO) research aquarium, two 40-gallon tanks were used to house the animals for the experiment. Any food or water that was added to the tanks was added via a head tank in order to keep the water and

food levels in the tanks consistent. Raw seawater flowed from the head tank into each tank at a rate of 1 liter per minute. The Group 1 tank was set to 18 °C while the Group 2 tank was set at 22 °C. Each tank was fitted with a chiller unit and pump used to chill the seawater to the appropriate temperature. The pump was attached to keep the flow rate consistent between the tanks. A heater unit was also used to help maintain a constant temperature. Temperatures and flowrates from each tank were recorded daily to insure that the conditions were kept consistent between tanks and that the water was at the desired temperature. After the summer, the rate of growth slowed and it was necessary to supplement the food given to the animals. A mixture of algae, and zooplankton was added once per day into the head tank to insure that all the tanks received equal amounts of food.

Spawning and Plating:

Two sexually mature *C. intestinalis* (larger than 5 cm) were collected from Snug Harbor, Point Judith Pond in early July of 2015. They were placed in the dark for one to two hours, and then placed under direct light, causing them to release gametes (Lambert & Brandt, 1967) (Whittingham, 1967). The eggs were collected from both animals and placed into separate dishes. Each was cross-fertilized with the sperm from the other. After a fifteen-minute time interval to allow fertilization to occur, the eggs were washed in filtered seawater to remove any excess sperm. They were incubated at 15 °C for 24 hours. After 24 hours, the eggs began to hatch and the free-swimming larva emerged. The larva and remaining eggs were transferred to a new dish containing a PVC or acrylic plate and fresh filtered seawater. After another 12 hours, the free-swimming larvae settled on the plates and started to metamorphose into

juveniles (reabsorb tail, become sessile, and begin to feed). The plates containing the juveniles were incubated at 15 °C for up to a week, at which point the number of juveniles settled on the plates was determined using a dissecting microscope. Plates containing 5 or more juveniles were placed into a 15-gallon tank at 15 °C. This tank was placed into an incubator and, over the course of a week, brought up to 18 °C. Half of the plates were carefully transferred to tank one and the other half were transferred to tank two. Tank two was gradually raised over the course of two weeks to its final temperature of 22 °C.

Collection and reproductive stress test:

After 3 months, the animals reached sexual maturity and reached about 5-6 cm in length. Four animals (two from tank 1 and two from tank 2) were collected and spawned. The eggs from the Group 1 animals were collected and placed in separate labeled dishes. The sperm from each animal was used to cross-fertilize the eggs. The same was done for Group 2. The fertilized eggs were then allowed to grow at both a normal temperature and a stress temperature. Group 1 embryos were placed at 18 °C and 22 °C. Group 2 embryos were placed at 22 °C and 26 °C (see figure 2). After 24 hours, the larvae were fixed and scored for normal development (Morgan, 1945). This test was used to score the reproductive ability of the animals, as well as the viability of the offspring. This was performed until six animals from tank one and 6 animals from tank two were collected and spawned.

Dissection and Tissue Preparation:

Directly after spawning, the adult animals were placed in clean dishes containing filtered seawater. The ovary and a portion of the testis were collected from each

animal. Each tissue type was removed and immediately frozen in liquid nitrogen.

After freezing, the labeled tubes were stored at -80 °C.

Lysate Preparation:

Ten of the ovary samples were chosen based on spawning test results (5 from tank one and 5 from tank two), this insured the samples used were from reproductively healthy individuals. The frozen ovary tissue samples were thawed on ice and washed with ice cold Phosphate Buffered Saline (PBS). The ovary tissue was transferred to a sterile glass douncer where it was homogenized with 9 M Urea buffer (9M urea and HEPES, pH 8) for 1 minute and returned to ice. The homogenate was transferred to a clean 2 ml tube. The samples were sonicated for 5 seconds and placed on ice for 30 seconds; this step was completed 6 times for a total of 30 seconds of sonication time. The samples were spun in a micro-centrifuge at 15 °C for 15 minutes in order to pellet out insoluble proteins. The lysate was sent to the Proteomics Core of the COBRE Center for Cancer Research and Development at Rhode Island Hospital, in Providence, RI. The Proteomics Core performed a tryptic digest and ran the samples through a High Performance Liquid Chromatography column, which feeds into the Q Exactive (ThermoFisher) tandem mass spectrometer. After sequencing was complete for all samples, the hospital staff analyzed the raw data and sent a list of identified proteins along with statistical data on each sample. (Refer to Appendix 1 for detailed proteomics methods).

Data Analysis:

The LC-MS/MS raw data was analyzed to identify proteins at a 1% false discovery rate. Three statistical tests were run on peak area intensity data. First a q-

value test for multiple hypothesis testing identified only five statistically upregulated proteins, using a significance q-value <0.05 (see Appendix 1 for statistical method). To identify more putatively differentially regulated proteins, a Student's two-tailed t-test and a Mann-Whitney U-test were performed for all proteins. Protein IDs with p-values < 0.05 by either test were examined for mean peak area intensity fold changes, between Group One and Group Two greater than that of the protein identified by the stringent q-value test with the lowest fold change (HSP71, UniProt ID F6W8I3_CIOIN). This criterion gave a value of 2 or greater for $10 \times \log_{10}(\text{mean fold change})$, which was considered the minimum cutoff value for upregulated proteins. Likewise, a $10 \times \log_{10}(\text{mean fold change})$ value of -2 or less was considered downregulated.

The differentially regulated proteins were uploaded as accession numbers to UniProtKB database in order to get a .fasta file containing the sequences. The same was done for the down-regulated proteins. These .fasta files were run through the BLAST2GO software. This software runs a BLAST search, followed by an InterPro search. This data is used to map the GO terms to the proteins. Finally, an annotation step chooses the highest scoring GO terms for each protein. After this data has been processed, graphs of the biological process were generated. The graphs show how many proteins map to each of the GO terms. Proteins that map together have similar functions and are often found in the same pathway.

The sequences were run through an online server called Ghost KOALA (<http://www.kegg.jp/ghostkoala/>). This server is part of online-based software that takes sequence information and assigns KEGG (Kyoto Encyclopedia of Genes and

Genomes) (<http://www.kegg.jp/kegg/>) values to each protein. KEGG is a database that stores information on protein pathways. Having the KEGG data allows for more in-depth analysis of individual biological pathways because it can pinpoint the exact location of the protein in the pathway.

The Overrepresentation Test on the PANTHER classification system (<http://pantherdb.org/tools/compareToRefList.jsp>) was used to pick out overrepresented proteins from the data. The statistically significant pathways were then searched for in the KEGG database using the Ghost KOALA data obtained in the previous step.

Results and Discussion

Setting up the Experiment:

Two tanks were set up at the URI GSO facility because it has access to raw seawater. One of the tanks was set to 18 °C (Group 1) and the other was set to 22 °C (Group 2). 18 °C is the upper limit of healthy development for *C. intestinalis* larva from the Rhode Island population. This was used as the control temperature. 22 °C is an increase of 4 degrees from the control, which increase corresponds to some climate change models, and was used as the experimental temperature (IPCC, 2014).

Reproductive Stress Test:

The *C. intestinalis* collected from the two tanks were spawned by placing them in the dark for 2 hours then placing them into bright light. The eggs were fertilized with sperm from another animal grown at the same temperature. These embryos were incubated at a stress temperature (22 °C for the Group 1 animals and 26 °C for the

Group 2 animals) and a control temperature (18 °C and 22 °C respectively) (Figure 2). Larvae were counted, sorted, and scored for normal development. If there were more than 20 larvae/embryos, only a random subset was scored. There were four Group 1 (18 °C) and four Group 2 (22 °C) animals tested, each with a subset of eggs at the control and at the stressed temperature. The number of normally developing offspring was averaged together by temperature. Figure 3 shows the averaged counts of normally developed larvae as a percentage of the total. The Group 1 animals produced a higher percentage of normal larvae in both 1a (18 °C) and 1b (22 °C) than the Group 2 animals in either 2a (22 °C) or 2b (26 °C), indicating a reduction in reproductive success at the higher temperature. In addition to having fewer normal larvae overall at the higher temperature, there is less difference between 2a (control) and 2b (stressed) than between 1a (control) and 1b (stressed). One possible reason for the less dramatic decrease in normally developing larvae in the higher temperature Group 2 is likely due to the group 2 animals already being stressed. Therefore stressing them further does not provide a substantial difference in response. A T-test between 1b and 2a (both incubated at 22 °C) had a p value of 0.081, which although not significant when using 95% confidence rating, suggests that more replicates would show a statistically significant decrease in reproductive success in the group reared at the higher temperature.

Protein Sequencing Results:

After dissecting out the ovarian tissue from five replicate animals at each condition, tissue lysates were prepared and sent for shotgun LC-MS/MS. Ovary tissue was used because it should be informative of what is happening to female

reproduction. It was also used because it is possible to dissect ovaries cleanly, without gut contamination.

1,616 proteins were identified in both conditions at a 1% false discovery rate. Based on 2-tailed Q tests, allowing for multiple hypothesis testing, 5 proteins were considered significantly differentially expressed between the control and stressed conditions (see Table 1, $P < 0.05$). Using a heat map (See Figure 4) generated using the peak intensity data for each biological replicate; it was possible to choose all proteins with at least three biological replicates up or down regulated and no more than two with contradictory expression in the complementary condition. This resulted in 168 proteins, which were considered consistent among the biological replicates. Of these proteins, 151 were upregulated ($10 \log_{10}$ fold change > 2) and 17 were downregulated ($10 \log_{10} < -2$). This indicates that more genes were being activated than down regulated due to the increased temperature.

Data Analysis:

The LC-MS/MS data was received as an Excel file. The Excel file contained columns for protein name, peptide sequence, and peak intensity for each sample, fold change ratios, q values, and accession numbers for NCBI and UniProtKB. The UniProt accession numbers were used on the UniProtKB (<http://www.uniprot.org/uploadlists/>) website to obtain a .fasta file containing the sequences for all of the proteins. This was also done for the up and downregulated lists. UniProtKB was used because it had more protein sequence data available than the NCBI database (<http://www.ncbi.nlm.nih.gov/>). It has a feature that allows batch

conversion of accession numbers into sequences that can be downloaded into one file. The fasta sequences are needed for the BLAST2GO program and for Ghost KOALA.

BLAST2GO (<https://www.blast2go.com/>) is a program that gives the gene ontology (GO) information for proteins. GO terms are a way of expressing gene function in a way that is comparable among species by using a consistent vocabulary. Due to the nature of GO terms they can be coded making them readable by many different computer programs (The Gene Ontology Consortium , 2015). GO terms also allow different analyses to be performed on a data set. The main purpose of running the BLAST2GO software is to get an idea of the types of genes that are present in the data set and how they group together by functionality. Having this information is useful for identifying a specific function or group of genes. It is also useful for comparing the up and down regulated lists to the total list in order to identify functional groups over or under represented in the differentially regulated proteins (Figure 6). This chart can be used to identify GO terms (or protein types/functions) that are more abundant either in the up or down regulated lists. Once this is identified, the individual proteins can be further investigated to see if an interesting pattern emerges. The groups that contain more than 20 % of the proteins are single organism metabolic process, single organism cellular process, single multicellular, organism process, response to stress, positive regulation of metabolic process, and anatomical structure morphogenesis. The categories that contain only upregulated proteins single organism signaling, regulation of localization, positive regulation of metabolic process, positive regulation of cellular process, negative regulation of cellular processes, macromolecule localization, and cellular localization

From the BLAST2GO software, transcription factors, signaling molecules, stress proteins, and heat shock proteins were identified (Table 2). Transcription factors can give insight into which genes are activated due to the increase in temperature. In addition, signaling molecules are used to activate or turn off certain pathways. Knowing which pathways are activated allows for the identification of the ligand or molecule that triggered the signal. Stress proteins and heat shock proteins are expected to appear during times of stress, such as the increase in temperature (Kultz, 2005). This makes them good indicators that the 4 degree increase in temperature is actually causing a stress response.

After finding the GO terms for the proteins, PANTHER (<http://www.pantherdb.org/geneListAnalysis.do>) was used as a tool for performing an overrepresentation test. This test compares the list of up and down regulated proteins (or whatever list of proteins and fold change data that needs processing) to the whole genome database for a chosen organism. The over representation test works via an algorithm that counts the number of proteins in a genome that map to a particular pathway or function. It determines the degree of difference between those expected by random chance and what was found in the list. The software can compare individual proteins as well as KEGG pathways. Knowing what proteins and pathways are over or underrepresented is a good way to find important trends in the data. The reason for using PANTHER instead of other software is that it was user friendly and had the whole genome annotations for *C. intestinalis* built into the program. Many of the other free online-based applications limit the species available for use. In the up and down

regulated list, the ubiquitin mediated proteolysis pathway was significantly overrepresented, according to the application default criteria.

After running the list through PANTHER, it was run through Ghost KOALA (<http://www.kegg.jp/ghostkoala/>), which provides KEGG annotations for each of the proteins and maps them to individual pathways. Ghost KOALA was designed for analyzing large datasets, thus allowing all the data to be input at once rather than searching one at a time. After running the list through the program, it generates a list of all the proteins in each pathway and a KEGG pathway diagram. The diagram shows where the proteins from the reference list are located in comparison to all other proteins mapped in the pathway. From the PANTHER data, it was determined that there was one significant overrepresented pathway, as mentioned above, which was then searched in the Ghost KOALA pathway data for a match.

Findings

It was found that there were some differences in the proteome of the animals grown at a higher temperature. For example, there was an upregulation of heat shock and other stress related proteins. Ubiquitin mediated proteolysis was the pathway that PANTHER designated as overrepresented. Ubiquitination is a process where ubiquitin tags are added to proteins in order to mark them for degradation by the proteasome. During times of heat stress, proteins can begin to unfold or misfold. Those that do not regain their native conformation are often marked for degradation. This pathway also plays a role in degradation of damaged or short-lived proteins, which often have regulatory functions (Bachmair, et al., 1986). Heat and oxidative stress can cause protein damage, which is a probable cause for the upregulation in this pathway

(Somero & Hofmann, 1995). It could be related to the upregulation in signaling molecules, which are short-lived and degraded in the same pathway. An upregulation in signaling molecules is a sign that other pathways are being signaled to perform a function due to the temperature difference (Table 2).

Summary

In summary, this work endeavored to determine if there was a noticeable difference in the proteome of *C. intestinalis* when grown at elevated temperatures. It was also attempting to determine what was causing decreased reproductive success when water temperatures are at their local high range. Though no proteins could be exclusively linked to reproduction, many proteins involved in stress were detected. This work also provides a method for analyzing shotgun proteomics data for non-model organisms (see figure 5), and for using proteomics for studies with an ecological aspect.

Future work:

In order to validate this data, subsequent research may involve measuring transcript levels by RT-PCR. This will be used to validate certain up and down regulated proteins among the biological replicates; making clear some of the biological variation and giving a better picture of what is going on in the organism. In addition, the uncharacterized proteins that are found to be consistently up or down regulated using the heat-map might be characterized using BLAST and a program called Phyre 2 (<http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index>). Phyre 2 looks at the amino acid sequence and picks out homologous protein domains based on structure and folding patterns. Knowing the biological function of the peptides contained in the protein helps to inform the biological process that it belongs in.

Therefore, it can be grouped with other proteins in the pathway thereby confirming that the process is in fact upregulated.

Figure 1. Fold changes between the means of protein quantifications at 22 °C vs. 18 °C sorted from highest to lowest. $10 \times \log_{10}$ of fold change values were plotted against the sequential numbering of the protein. Inflection points for up or down regulated proteins were identified visually to include the five proteins with differential fold changes with p-values <0.05 by 2-tailed unpaired t test (see Appendix I). The fold change values at these inflection points were used to determine which proteins were up or down regulated. The top graph is based on the raw data and the bottom graph is based off only those proteins that were consistent among the 5 biological replicates.

Figure 1.

Figure 2. Diagram of the different groups in the reproductive stress test.

The offspring from each set of parents was split into two treatments, a control (1A and 2A) and an experimental (1B and 2B). Each treatment was incubated at different temperatures.

Figure 2

Figure 3. Data from the reproductive stress test shows that there is a noticeable difference in the viability and health of the offspring spawned from the temperature stressed animals.

Graph of the % normal development of the embryos spawned from the Group 1 and Group 2 animals. The Group 1 offspring incubated at the control temperature of 18 °C grew normally while the Group 1 offspring incubated at the stress temperature of 22 °C had fewer normally developed embryos. The Group 2 offspring incubated at the control temperature of 22 °C showed fewer normally developed embryos than either of the Group 1 tests. The Group 2 offspring incubated at the stress temperature of 26 °C had almost the same number of normally developing embryos as the control temperature.

P value:

Control (1a vs. 2a)	0.007
Stressed (1b vs. 2b)	0.087

Figure 3.

Table 1. Five proteins deemed significant by 2-tailed Student T test allowing for multiple hypotheses.

These proteins showed a significant increase, between the two conditions, in the number of protein detected by the mass spectrometer.

Table 1.

Protein name	Identified Peptide Sequence	Fold change R22 vs R18	q-values R22 vs R18
PREDICTED: uncharacterized protein LOC104265360 [C. intestinalis]	R.GVFLYYYFPK.S	169.4019173	0.00875107
PREDICTED: sec1 family domain-containing protein 1 [C. intestinalis]	K.VLIYDNAGQDIITPLFSVADLR.N	96.43522677	0.02055127
PREDICTED: succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial [C. intestinalis]	R.LGANSLLDLVVFGFGR.A	235.7674098	0.03162674
PREDICTED: tubulin beta chain [C. intestinalis]	R.AILVDLEPGTMDSVR.S	5.675911654	0.03210064
PREDICTED: heat shock cognate 71 kDa protein isoform X2 [C. intestinalis]	K.NQVAMNPTNTVFDAK.R	1.589488348	0.03602449

Figure 4. Heat map showing peak intensities for proteins with 3 or more measurements from each condition either up or down regulated.

Figure 4.

Figure 5. Work flow for the mass spectrometer data analysis

This is a flow chart for how to analyze the raw data. Two paths can be taken. Path one takes the ids from the excel file and turns them into sequences. This is fed into BLAST2GO, which gives gene ontology information about each protein. This gives insight into its function. This is used when looking for a specific class of protein. The sequences are also used as input for Ghost KOALA, which gives protein pathway information. The other path uses PANTHER to determine interesting pathways and then Ghost KOALA is used to visualize all the proteins in that pathway. Both processes can be performed simultaneously using different software packages.

Figure 5.

Table 2. Up and down-regulated signaling, transcription factor, heat shock, and stress-related proteins.
This table was created from the BLAST2GO data.

Table 2

Protein type	Accession Num.	Protein Name
	Upregulated	
Signaling	F6YR71_CIOIN	peptidyl-prolyl cis-trans isomerase FKBP4-like
	F6Y5B7_CIOIN	ubiquitin-conjugating enzyme E2-17 kDa
	F6Y4V4_CIOIN	ubiquitin-conjugating enzyme E2-17 kDa
	F6YTX7_CIOIN	TPA: zinc finger protein, partial
	F6SV14_CIOIN	sarcoplasmic/endoplasmic reticulum calcium ATPase 2-like isoform X2
	Q0GTF8_CIOIN	insulin-like 1 precursor
	F6S3B3_CIOIN	transcription factor protein
	F6U1X7_CIOIN	calcium-binding protein 39
	F6WST8_CIOIN	uncharacterized protein LOC100179648
	F6RZG4_CIOIN	PREDICTED: dipeptidyl peptidase 1-like, partial
	H2Y2N2_CIOIN	STAM-binding protein-like
	F7BCV4_CIOIN	2-oxoisovalerate dehydrogenase subunit alpha, mitochondrial-like
	F7ACX7_CIOIN	gelsolin
	F6TEQ0_CIOIN	ras-related protein O-RAL isoform X2
	F6RLS6_CIOIN	ADP-ribosylation factor-like protein 6
	F6VG31_CIOIN	nodal modulator 1
	F6TJA4_CIOIN	signal peptidase complex subunit 3-like isoform X1
	F6WUJ1_CIOIN	structural maintenance of chromosomes protein 3-like, partial
	H2Y1U8_CIOIN	ras-related protein Rab-27B
	H2Y3H5_CIOIN	ras-related protein RAP-1B homologue

	F6X159_CIOIN	mitogen-activated protein kinase kinase
	F7AHT0_CIOIN	H3L-like histone, partial
	H2XKD5_CIOIN	protein transport protein Sec23A
	H2Y164_CIOIN	aarF domain-containing protein kinase 4
Transcription factors	F6S3B3_CIOIN	transcription factor protein
	H2XRX3_CIOIN	LIM domain-binding protein 3-like
	H2XZX2_CIOIN	TPA: zinc finger protein, partial
Stress	F6SV14_CIOIN	sarcoplasmic/endoplasmic reticulum calcium ATPase 2-like isoform X2
	F6Y5B7_CIOIN	ubiquitin-conjugating enzyme E2-17 kDa
	F6YR71_CIOIN	peptidyl-prolyl cis-trans isomerase FKBP4-like
	F6YGK4_CIOIN	insulin-like growth factor 2 mRNA-binding protein 1
	F6QWS2_CIOIN	protein disulfide-isomerase A4-like isoform X2
	F6SV14_CIOIN	sarcoplasmic/endoplasmic reticulum calcium ATPase 2-like isoform X2
	F6QB99_CIOIN	electron transfer flavoprotein-ubiquinone oxidoreductase, mitochondrial
	H2XRX3_CIOIN	LIM domain-binding protein 3-like
	H2XTS1_CIOIN	bacterial stress protein UspA-like (possible contaminant)
	F6X159_CIOIN	mitogen-activated protein kinase kinase
Heat Shock Protein	H2XWU8_CIOIN	heat shock cognate 71 kDa protein isoform X2
	F6W8I3_CIOIN	heat shock protein 67B2-like
		Down regulated
Signaling	F7BBW9_CIOIN	NADH-ubiquinone oxidoreductase 75 kDa subunit, mitochondrial-like
Stress	F6YF53_CIOIN	DAZ-associated protein 1-like

Figure 6. Bar chart of the GO data.

The number of proteins per GO term in the upregulated set was divided by the proteins per GO terms for the entire list of identified proteins; the same was also done for the downregulated proteins. This allows for comparison of the difference in number of proteins between the up and down regulated lists.

Figure 6.

Bibliography

- Bachmair, A., Finley, D. & Varshavsky, A., 1986. In vivo half-life of a protein is a function of its amino-terminal residue. *Science*, Volume 234, pp. 179-186.
- Berrill, N. J., 1947. THE DEVELOPMENT AND GROWTH OF CIONA. *Journal of the Marine Biological Association of the United Kingdom*, 26(4), pp. 616-625.
- Brockington, S. & Clarke, A., 2001. The relative influence of temperature and food on the metabolism of a marine invertebrate. *Journal of Experimental Marine Biology and Ecology*, 258(1), pp. 87-99.
- Carver, C., Mallet, A. & Vercaemer, . B., 2006. *Biological Synopsis of the Solitary Tunicate C. intestinalisintestinalis*, Dartmouth, Nova Scotia: Fisheries and Oceans Canada Bedford Institute of Oceanography .
- Doney, S. C., 2011. Climate Change Impacts on Marine Ecosystems. *Annual Review of Marine Science*, Volume 4, pp. 11-37.
- Dorts, J. et al., 2012. Evidence that elevated water temperature affects the reproductive physiology of the European bullhead *Cottus gobio*. *Fish Physiology and Biochemistry*, 38(2), pp. 389-399
- Dybern, B. I., 1965. The Life Cycle of *C. intestinalisintestinalis* (L.) f. *typica* in Relation to the Environmental Temperature. *Oikos*, pp. 109-131.
- Einarsson, S., Brandt, Y., Lundeheim, N. & Madej, A., 2008. Stress and its influence on reproduction in pigs: a review. *Acta Vet Scand.*, 50(48), pp. 1-8.
- Endo, T. et al., 2011. CIPRO 2.5: *C. intestinalisintestinalis* protein database, a unique integrated repository of large-scale omics data, bioinformatic analyses and curated annotation, with user rating and reviewing functionality. *Nucleic Acids Res n.d.*, pp. D807-D814.
- Kanehisa, M., Sato, Y., and Morishima, K. (2016) BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* 428, 726-731
- Kultz, D., 2005. Molecular and evolutionary basis of the cellular stress response. *Annu. Rev. Physio.* 67:225-257.
- Iannelli, F., Pesole, G. & Sordin, P., 2007. Mitogenomics reveals two cryptic species in *C. intestinalisintestinalis*. *Trends in Genetics*, 23(9), p. 419-422.

- Lambert, C. C. & Brandt, C. L., 1967. The Effect of Light on the Spawning of *Ciona intestinalis*. *Biological Bulletin*, 132(2), pp. 222-228.
- Morgan, T. H., 1945. Normal and Abnormal Development of The Eggs of *Ciona*. *Journal of Experimental Zoology*, 100(3), p. 407–416.
- NOAA, 2015. *Water Temperature Table of the Northern Atlantic Coast*. [Online] Available at: <https://www.nodc.noaa.gov/cwtg/natl.html>
- PANTHER version 10: expanded protein families and functions, and analysis tools. Huaiyu Mi, Sagar Poudel, Anushya Muruganujan John T Casagrande and Paul D. Thomas, *Nucl. Acids Res.* (2016) doi: 10.1093/nar/gkv1194
- Petersen, J. K., Mayre, S. & Knudsen, M. Å., 1999. Beat frequency of cilia in the branchial basket of the ascidian *C. intestinalis* in relation to temperature and algal cell concentration. *Marine Biology*, 133(2), pp. 185-192.
- S. Götz et al. "High-throughput functional annotation and data mining with the Blast2GO suite", *Nucleic Acids Research*, Vol. 36, June, 2008, pp. 3420-3435.
- Serafinia, L., Hanna, J. B., Kültzb, . D. & Tomaneka, . L., 2011. The proteomic response of sea squirts (genus *Ciona*) to acute heat stress: A global perspective on the thermal stability of proteins. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics*, p. 322–334.
- Somero, G. & Hofmann, G., 1995. Evidence for protein damage at environmental temperatures: seasonal changes in levels of ubiquitin conjugates and hsp70 in the intertidal mussel *Mytilus trossulus*. *Journal of Experimental Biology*, Volume 198, pp. 1509-1518.
- The Gene Ontology Consortium , 2015. Gene Ontology Consortium: Going Forward. *Nucl Acids Res*, Volume 43.
- Tomanek, L., 2011. Environmental Proteomics: Changes in the Proteome of Marine Organisms in Response to Environmental Stress, Pollutants, Infection, Symbiosis, and Development.. *Annu. Rev. Mar. Sci*, Volume 3, p. 373–99.
- Whittingham, D. G., 1967. Light-Induction of Shedding of Gametes in *Ciona intestinalis* and *Molgua manhattensis*. *Biological Bulletin* , 132(2), pp. 292-298.
- Zhang, Y. et al., 2013. Protein Analysis by Shotgun/Bottom-up Proteomics. *Chem Rev*, Volume 113, p. 2343–94.

CHAPTER 2

Changes in the *Ciona intestinalis* Proteome Due to Overwintering

Abstract

C. intestinalis in Rhode Island exhibit normal gametogenesis at water temperatures between 10 °C and 18 °C. During the coldest part of the winter when the water is below 8 °C, their feeding slows, the production of gametes ceases and much of the ovary tissue is resorbed and used as an energy source (Dybern, 1965). Ovary samples were collected from animals raised at 18 °C and from a non-temperature controlled tank at 8 °C. The samples were sequenced using liquid chromatography-tandem mass spectrometry. The data showed a decrease in the number of proteins produced by the overwintering animals. The goal of this chapter is to find the differences in physiology that occur during dormancy by looking at the protein pathways that are differentially expressed.

Introduction

Ciona intestinalis is a solitary ascidian found on most rocky shores in the Northern Hemisphere. *C. intestinalis* prefers the cold temperate waters of protected bays often growing on the side of docks and other manmade structures. *Ciona intestinalis*, which translates to “column of intestines”, uses a branchial basket filled with beating cilia to filter algae and other particles from the water column (Carver, et al., 2006). As water temperature increases, so does the beat frequency of the cilia, increasing the amount of water and food passed over the branchial basket (Petersen, et

al., 1999). A decrease in water temperature therefore decreases the intake of food and many of the biological processes that require energy are downregulated.

Dormancy is defined as some type of metabolic and/or developmental depression due to an ecological cue, such as temperature, and encompassing a wide range of physiological changes (Caceres, 1997). During dormancy, *C. intestinalis* is not actively producing gametes and no mature gametes are present in the reproductive tissue. This generally occurs during January, February, and March in the North Atlantic, when the water temperatures are at their lowest, but can vary based on local water temperature. As the water temperature rises, the adult animals that survived the winter spawn and die in early-mid spring. This produces a new population that will spawn at the end of summer and then overwinter to spawn and die in the next spring (Carver, et al., 2006). Those populations that do not experience extreme cold or hot temperatures, like those in Southern England (5° C- 20 °C) and California may spawn continuously because they do not reach a critically lower temperature limit or large fluctuations in water temperature (Carver, et al., 2006). The local population in Rhode Island can experience temperature as low as 2 °C and as high as 25 °C over the course of the year (NOAA, 2015).

Traditionally, 2D gels were used when comparing protein expression between two environmental samples (Wright, et al., 2012). Only those spots that differ between the two samples are sent for sequencing on a mass spectrometer. This method however, may not give a clear picture of all the processes that are affected by the change in environment. Due to new mass spectrometry technology, it is possible to sequence a larger number of proteins at once, allowing a more complete proteome to

be sequenced in one run (Wright, et al., 2012). Having a more complete proteome allows for an in depth analysis of biological pathways that are being affected (Tomanek, 2011). The main benefit of using proteomics, rather than transcriptomics for this study, is that it shows the protein expression directly, which may not be accurately reflected by the transcriptome (Horgan & Kenny, 2011). Therefore, it is possible to predict which physiological processes are altered during dormancy. I hypothesize that the expression level of certain genes in the ovarian tissue will decrease due to the lack of active reproduction, more specifically those dealing directly with metabolism.

Methods

Set up:

C. intestinalis were spawned and settled on petri dishes. They were transported to the University of Rhode Island Bay Campus (GSO) where they were suspended in outdoor cylindrical tanks pumped with raw seawater. The tanks were covered to decrease the amount of light to prevent an overgrowth of algae in the tanks, which could cause hypoxic events.

Tissue Collection, Dissection, and Lysate Preparation:

Five animals were collected at 8 °C in late February, and the ovaries were removed. The tissue was frozen in 1.5 ml tubes using liquid nitrogen. The frozen ovary tissue samples were thawed on ice and washed with ice cold Phosphate Buffered Saline (PBS) in order to remove excess saltwater from the tissue. The ovary tissue was transferred to a sterile glass douncer, homogenized with 9 M Urea buffer (pH 8) for 1

minute, and returned to ice. The homogenate was transferred to a clean 2 ml tube. The samples were sonicated for 5 seconds at 10% amplitude (Fisher Scientific sonic demembrator model 500) and placed on ice for 30 seconds; this step was repeated 6 times for a total of 30 seconds of sonication time. The samples were centrifuged at 14000 g at 15 °C for 15 minutes in order to pellet out insoluble proteins. The lysate was frozen and sent to the COBRE Center for Cancer Research and Development at Rhode Island Hospital. Here a tryptic digest was performed and the sample was run through a High Performance Liquid Chromatography column, which feeds into the Q Exactive (ThermoFisher) tandem mass spectrometer. Refer to Appendix 1 for detailed proteomics methods.

Data Analysis

The data was received as an Excel file containing a list of the proteins identified in each sample. From this file, a text file was created containing the UniProtKB (<http://www.uniprot.org/uploadlists/>) accession numbers and fold change ratio for each protein. First the accession numbers were searched in UniProtKB to obtain full sequences for each protein. The sequences were uploaded to the Ghost KOALA (<http://www.kegg.jp/ghostkoala/>) online server, which returned KEGG pathway information (known as KEGG terms) for 77% of the 219 unique proteins. The KEGG terms were compiled into three lists. List one contained unique terms in the cold-water (8 °C) samples, list two contained unique terms in the 18 °C samples, and list three contained terms found in both samples. In addition, the sequences from UniProtKB were uploaded to the BLAST2GO software. This software performs GO annotation analysis.

Results

Because the 18 °C and 8 °C samples were not run on the same HPLC column, it was not possible to reliably compare protein quantification between the two conditions. Therefore, I compared the samples using the presence or absence of individual proteins. There were 178 proteins unique to the 18 °C samples (Table 3) and 41 proteins unique to the 8 °C samples (Table 4) (Figure 7). BLAST2GO was used to determine the functions of the individual proteins. The number of proteins that fell into each of the functional categories is listed in Table 5. The two highlighted rows in the table show the two categories in which there was the greatest or least difference in number of mapped proteins. The category with the highest difference is "response to stimulus", which is defined as any process that results in a change in state or activity of a cell or an organism (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a stimulus. The process begins with detection of the stimulus and ends with a change in state or activity of the cell or organism (Carbon, et al., 2009). This general category encompasses any response to a hormone, chemical or stimulus of any kind. MAP/microtubule affinity-regulating kinase, Ras-related C3 botulinum toxin substrate 1, and Niemann-Pick C2 protein are a few examples of proteins that fall into this category. The category with the least difference was negative regulation of biological process, which is defined as any process that stops, prevents, or reduces the frequency, rate or extent of a biological process. Biological processes are regulated by many means; examples include the control of gene expression, protein modification or interaction with a protein or substrate molecule (Carbon, et al., 2009). Glucose-6-phosphate 1-dehydrogenase,

calcium-transporting ATPase, and tubulin alpha chain are a few examples of proteins in this category.

The Ghost KOALA results are listed in Tables 6, 7, and 8. Table 6 lists the KEGG terms (or protein pathways) that are unique to the 8 °C samples, Table 7 shows those unique to the 18 °C samples, and Table 8 shows those present in both. There are 166 pathways listed for 18 °C, as opposed to 9 pathways listed for 8 °C, including more at 18 °C in the category of signal transduction, body systems, and metabolism.

Discussion

The sequence data was analyzed using BLAST2GO and Ghost KOALA. The purpose of this experiment was to determine the difference in the proteome of the reproductive tissues of *C. intestinalis* when it is actively reproducing compared to when it is reproductively dormant. As the temperature and food availability decrease, the ovary shrinks in size due to both resorption of gametes and the reallocation of energy stored for reproduction. This energy is used elsewhere and resorbed tissue is not regenerated until favorable conditions return in the spring (Carver, et al., 2006) (Airi, et al., 2014). Below 10 °C for the Rhode Island population, *C. intestinalis* likely decreases its food intake and therefore must get its energy from other sources to include breaking down of the ovary (Petersen, et al., 1999) (Dybern, 1965).

The number of proteins detected in 8 °C samples is lower than of 18 °C samples (1291 proteins detected in 8 °C samples compared to 1399 proteins in 18 °C samples). This is to be expected; when the animal is not actively reproducing, it does not need to produce proteins necessary to generate mature gametes. In addition, many

of the cellular processes in the ovary cells shut down when not in active use in order to conserve energy, including cell signaling, digestion, and general metabolism (based on pathway analysis, see tables). Some associated proteins found only in the 18 °C data are adenosine a signaling molecule in the FAS signaling pathway, microsomal triglyceride transfer protein that is involved in the digestion and absorption of lipids, and phosphoenolpyruvate carboxykinase, a part of pyruvate metabolism. In many organisms, diapause or dormancy is accompanied by a decrease in metabolism in order to slow depletion of energy stores during a time of low food abundance (Caceres, 1997).

Proteins present in the transport pathways remained consistent between the samples, except for protein export and RNA transport pathways, for which representative proteins were unique to the 8 °C samples (see table 6). Therefore, it is possible that the components from the ovary cells that are not in current use are sent to other parts of the body. In addition, many of the proteins expressed in the 8 °C samples played some role in RNA production or processing. The increase in RNA processing is likely due to the time of year that the animals were harvested. Since they were collected toward the end of winter right before the water temperatures began to rise, the animals may have begun increasing the amount of transcription in order to prepare for the production of eggs (Berrill, 1947). Although the water temperature was beginning to rise, the temperature is still below 10 °C which is the lower limit for reproduction in *C. intestinalis* (Berrill, 1947). In addition, no gametes were found in the reproductive tract or organs at the time of dissection.

Based on the pathways assigned by Ghost KOALA to the proteins from the 8 °C data, the active systems were the nervous system, excretory system, endocrine system, and the circulatory system, meaning that proteins were detected in these categories. There were fewer proteins present in these categories in the 8 °C samples than the 18 °C. During dormancy, it is expected that many non-essential functions will be downregulated in order to conserve energy. In conclusion, there was an overall decrease in the number of proteins detected in the 8 °C samples and many non-essential systems were turned off including egg production. For future experiments, more samples should be taken at varying points during dormancy in order to understand the metabolic decline better, and to see if there was an increase in energy storage prior to metabolic decline. In addition, taking measurements of food intake and respiration will verify the extent of depression in metabolism occurs over the winter. This will help to clarify the process of dormancy in *C. intestinalis*.

Table 3 and 4. Proteins unique to each sample

These tables were created by taking the raw data from each of the 5 samples and removing duplicate proteins then comparing this list to the list of 18 °C samples to the 8 °C samples. Any duplicate between lists were removed leaving proteins found only in one list or the other.

Table 3. Proteins in the 8 °C list but not in 18 °C list

UniProt #	Protein family
F6V934	C-TERMINAL-BINDING PROTEIN (PTHR10996:SF189)
F6WJ25	MITOCHONDRIAL IMPORT RECEPTOR SUBUNIT TOM40 HOMOLOG 1-RELATED (PTHR10802:SF7)
F6VKZ4	PROTEIN MAK16 HOMOLOG (PTHR23405:SF7)
F6ZM86	PROTEIN TRANSPORT PROTEIN SEC61 SUBUNIT ALPHA-LIKE 1-RELATED (PTHR10906:SF14)
F6W141	MITOCHONDRIAL RHO GTPASE (PTHR24072:SF206)
F6QQ49	EXPORTIN-5 (PTHR11223:SF7)
F6XA69	VALACYCLOVIR HYDROLASE (PTHR10992:SF972)
Q7YT70	CDC42 HOMOLOG (PTHR24072:SF215)
F6VRH3	PERIODIC TRYPTOPHAN PROTEIN 2 HOMOLOG (PTHR19858:SF1)
H2XXH6	RAS-RELATED PROTEIN M-RAS (PTHR24070:SF321)
F6X4B8	CALCIUM-BINDING MITOCHONDRIAL CARRIER PROTEIN ARALAR1 (PTHR24089:SF532)
F6WWD5	ATPASE FAMILY AAA DOMAIN-CONTAINING PROTEIN 3 (PTHR23075:SF10)
F6S4U2	GUANINE NUCLEOTIDE-BINDING PROTEIN-LIKE 3 HOMOLOG (PTHR11089:SF58)
F6T949	ATP-DEPENDENT ZINC METALLOPROTEASE YME1L1 (PTHR23076:SF80)
F6YGC1	ATP-DEPENDENT RNA HELICASE DDX27-RELATED (PTHR24031:SF438)
F7ARF4	CYTOCHROME P450 4V2 (PTHR24291:SF67)
F7B5A5	IMPORTIN-9 (PTHR10997:SF50)
F6Z3S0	ARGININE--TRNA LIGASE, CYTOPLASMIC (PTHR11956:SF9)
H2XVM9	BRAHMA ASSOCIATED PROTEIN 155 KDA (PTHR12802:SF68)
F6QS20	26S PROTEASOME NON-ATPASE REGULATORY SUBUNIT 4 (PTHR10223:SF3)
F6UDJ0	PROTEIN RRP5 HOMOLOG (PTHR23270:SF11)
F6R407	MYB-BINDING PROTEIN 1A (PTHR13213:SF3)
H2XSA5	U6 SNRNA-ASSOCIATED SM-LIKE PROTEIN LSM6 (PTHR11021:SF6)
F7B7H5	CATALASE (PTHR11465:SF29)
F6XDT3	ADENOSINE KINASE (PTHR10584:SF224)
F6Y8C3	N-ACETYLTRANSFERASE 10 (PTHR10925:SF8)

H2XPN9	CCAAT/ENHANCER-BINDING PROTEIN ZETA (PTHR12048:SF1)
F6YN07	NNT-1 (PTHR10160:SF26)
F6Q021	H/ACA RIBONUCLEOPROTEIN COMPLEX SUBUNIT 1 (PTHR23237:SF9)
F6WWT1	EUKARYOTIC TRANSLATION INITIATION FACTOR 2 SUBUNIT 2 (PTHR23001:SF13)
F6YQR4	NUCLEOLAR COMPLEX PROTEIN 2 HOMOLOG (PTHR12687:SF9)
F6XTJ3	NUCLEOLAR GTP-BINDING PROTEIN 1 (PTHR11702:SF32)
F7ABF6	AP-2 COMPLEX SUBUNIT ALPHA (PTHR22780:SF29)
F6QZB5	INTEGRATOR COMPLEX SUBUNIT 11 (PTHR11203:SF45)
F6UE24	PROTEIN MAGO NASHI HOMOLOG (PTHR12638:SF3)
H2XX77	COLD SHOCK DOMAIN-CONTAINING PROTEIN CG9705 (PTHR12962:SF6)
F7A1V5	ATP-DEPENDENT RNA HELICASE DDX23-RELATED (PTHR24031:SF525)
F6UMP0	PROTEASOMAL UBIQUITIN RECEPTOR ADRM1 (PTHR12225:SF1)
F6Q662	ACTIN
F6XQ45	28S rRNA (CYTOSINE-C(5))-METHYLTRANSFERASE
H2XPK9	SERINE/ARGININE-RICH SPLICING FACTOR 6-LIKE

Table 4. Proteins in 18 °C list not in 8 °C list

UniProt #	Protein family
F7B2X5	ILE-2 (PTHR12223:SF41)
F6YR71	FKB-6 (PTHR10516:SF341)
F7BDH1	PROGRAMMED CELL DEATH 6-INTERACTING PROTEIN (PTHR23030:SF26)
F6YDF4	DELTA-1-PYRROLINE-5-CARBOXYLATE DEHYDROGENASE 1, ISOFORM A-RELATED (PTHR42862:SF1)
F7AEM1	PROTEIN DISULFIDE-ISOMERASE A5 (PTHR18929:SF153)
F6VN74	COATOMER SUBUNIT ZETA-RELATED (PTHR11043:SF12)
Q8HIN6	CYTOCHROME C OXIDASE SUBUNIT 2 (PTHR22888:SF14)
H2Y0Y4	CORONIN (PTHR10856:SF33)
F6SHN0	BETA-HEXOSAMINIDASE SUBUNIT BETA (PTHR22600:SF27)
H2XZE3	HILLARIN, ISOFORM A (PTHR11039:SF51)
H2XW02	SMALL NUCLEAR RIBONUCLEOPROTEIN SM D2 (PTHR12777:SF2)
F6S4Q0	ACETYL-COA ACETYLTRANSFERASE, MITOCHONDRIAL (PTHR18919:SF114)
F6Q1Z4	SUBFAMILY NOT NAMED (PTHR11311:SF26)
H2XKA1	SMALL NUCLEAR RIBONUCLEOPROTEIN E (PTHR11193:SF3)
F6VYQ8	STEROL 4-C-METHYLTRANSFERASE STRM-1 (PTHR10108:SF1029)
H2XXI4	FI01422P-RELATED (PTHR11461:SF233)
F6YWN2	NADH DEHYDROGENASE [UBIQUINONE] 1 ALPHA SUBCOMPLEX SUBUNIT 7 (PTHR12485:SF2)
H2Y2P2	ANGIOPOIETIN-RELATED PROTEIN 7 (PTHR19143:SF319)
F6SSX2	CYTOSOL AMINOPEPTIDASE (PTHR11963:SF23)
Q8I7E3	ACTIN-RELATED PROTEIN 2/3 COMPLEX SUBUNIT 4 (PTHR22629:SF2)
F7AHS5	60 KDA SS-A/RO RIBONUCLEOPROTEIN (PTHR14202:SF2)
F6ZU97	ALPHA-AMINOADIPIC SEMIALDEHYDE SYNTHASE, MITOCHONDRIAL (PTHR11133:SF16)
H2XYB2	MICOS COMPLEX SUBUNIT MIC10 (PTHR21304:SF4)
Q7YT79	RAS-RELATED PROTEIN RAC1-RELATED (PTHR24072:SF205)
F6PP61	PROTEASOME SUBUNIT BETA TYPE-4 (PTHR11599:SF72)
F7AFF3	PROTEIN YIPF (PTHR21236:SF13)
H2XLZ2	INACTIVE PEPTIDYL-PROLYL CIS-TRANS ISOMERASE SHUTDOWN-RELATED (PTHR10516:SF368)
H2XW16	SUBFAMILY NOT NAMED (PTHR35169:SF2)

F7BFN9	SUBFAMILY NOT NAMED (PTHR11311:SF26)
F6QBL6	INTEGRIN BETA-PS-RELATED (PTHR10082:SF57)
F6XG50	OLIGORIBONUCLEASE, MITOCHONDRIAL (PTHR11046:SF6)
H2Y1Z9	ALANINE--TRNA LIGASE, CYTOPLASMIC (PTHR11777:SF23)
F6XCZ8	TUBA7L PROTEIN (PTHR11588:SF218)
F6PTJ9	PROTEIN PHOSPHATASE 1G (PTHR13832:SF381)
H2XXZ9	SUBFAMILY NOT NAMED (PTHR23408:SF19)
F6S9Y0	UDP-GLUCOSE 4-EPIMERASE (PTHR43725:SF5)
F6VTS4	FI01422P-RELATED (PTHR11461:SF233)
F6VWW5	SUBFAMILY NOT NAMED (PTHR24044:SF399)
F6ZLN5	LEUCINE-RICH REPEAT-CONTAINING PROTEIN 47 (PTHR10947:SF7)
H2XWK0	SH3 DOMAIN-BINDING GLUTAMIC ACID-RICH-LIKE PROTEIN (PTHR12232:SF13)
F7BHX3	PROTEIN GM973 (PTHR21937:SF5)
F7B397	CALNEXIN 14D-RELATED (PTHR11073:SF22)
F6VJQ1	POLYUBIQUITIN-B (PTHR10666:SF198)
F6S4F7	RAC SERINE/THREONINE-PROTEIN KINASE (PTHR24356:SF248)
H2XJF4	SERINE HYDROXYMETHYLTRANSFERASE, CYTOSOLIC (PTHR11680:SF18)
F7AHD2	GST-40 (PTHR11571:SF169)
F6YB01	NUCLEOSIDE DIPHOSPHATE KINASE (PTHR11349:SF78)
H2Y1Q1	GUANINE NUCLEOTIDE-BINDING PROTEIN G(O) SUBUNIT ALPHA (PTHR10218:SF248)
H2XK44	CHROMOBX PROTEIN (CHCB2) (PTHR22812:SF111)
F6TEQ0	RAS-RELATED PROTEIN RAL-A (PTHR24070:SF343)
H2XRW4	NADH DEHYDROGENASE [UBIQUINONE] IRON-SULFUR PROTEIN 8, MITOCHONDRIAL (PTHR10849:SF23)
F6U9F4	ACONITATE HYDRATASE (PTHR11670:SF46)
F6QC45	TRANSMEMBRANE EMP24 DOMAIN-CONTAINING PROTEIN 2 (PTHR22811:SF89)
F6U1L2	SHORT-CHAIN SPECIFIC ACYL-COA DEHYDROGENASE, MITOCHONDRIAL (PTHR43884:SF16)
H2Y169	SEPTIN-7 (PTHR18884:SF86)
F6S1S7	DRAB5 (PTHR24073:SF609)
F7BC32	NADH DEHYDROGENASE [UBIQUINONE] 1 BETA SUBCOMPLEX SUBUNIT 7 (PTHR20900:SF1)

F7AR17	SUBFAMILY NOT NAMED (PTHR12461:SF67)
F7ACJ3	TBC-12 (PTHR22957:SF430)
H2Y1U8	EG:80H7.4 PROTEIN (PTHR24073:SF738)
F6ZED0	PROGRAMMED CELL DEATH PROTEIN 5 (PTHR10840:SF1)
H2XMT5	CYTOCHROME B-C1 COMPLEX SUBUNIT 8 (PTHR12119:SF4)
F6RZU6	PROTEIN BUD31 HOMOLOG (PTHR19411:SF6)
H2XX36	ATP SYNTHASE F(0) COMPLEX SUBUNIT B1, MITOCHONDRIAL (PTHR12733:SF5)
F6WAG8	ENOYL-COA HYDRATASE, MITOCHONDRIAL (PTHR11941:SF111)
F6XH64	NUCLEAR AUTOANTIGENIC SPERM PROTEIN (PTHR15081:SF10)
F6VLB7	GLYOXALASE DOMAIN-CONTAINING PROTEIN 4 (PTHR10374:SF28)
F6UN42	PHOSPHATIDYLETHANOLAMINE-BINDING PROTEIN 1 (PTHR11362:SF45)
H2XP47	GUANINE NUCLEOTIDE-BINDING PROTEIN SUBUNIT GAMMA (PTHR13809:SF39)
H2XP91	MESENCEPHALIC ASTROCYTE-DERIVED NEUROTROPHIC FACTOR HOMOLOG (PTHR12990:SF13)
F7A1Y9	TRANSMEMBRANE 9 SUPERFAMILY MEMBER 4 (PTHR10766:SF61)
F6PWH2	PROSTAMIDE/PROSTAGLANDIN F SYNTHASE (PTHR28630:SF9)
Q589R6	LIM AND SH3 DOMAIN PROTEIN LASP (PTHR11039:SF50)
F7BH23	PROGRAMMED CELL DEATH PROTEIN 6 (PTHR10183:SF367)
F6UHT2	VERY LONG-CHAIN SPECIFIC ACYL-COA DEHYDROGENASE, MITOCHONDRIAL (PTHR43884:SF10)
H2XTH7	ISOBUTYRYL-COA DEHYDROGENASE, MITOCHONDRIAL (PTHR43831:SF1)
F7BBW9	NADH-UBIQUINONE OXIDOREDUCTASE 75 KDA SUBUNIT, MITOCHONDRIAL (PTHR11615:SF194)
H2Y2A4	THIOREDOXIN REDUCTASE 3 (PTHR43256:SF4)
F6ZOW4	NADH DEHYDROGENASE [UBIQUINONE] 1 ALPHA SUBCOMPLEX SUBUNIT 12 (PTHR12910:SF6)
F6VIC3	AT11025P (PTHR10809:SF65)
F6Y403	CARBOXYPEPTIDASE Y (PTHR11802:SF117)

F6Z3G2	SUBFAMILY NOT NAMED (PTHR12461:SF75)
F6TAA1	SUBFAMILY NOT NAMED (PTHR24044:SF399)
H2XQF1	VACUOLAR PROTEIN SORTING-ASSOCIATED PROTEIN VTA1 HOMOLOG (PTHR12741:SF27)
H2XZW2	MITOCHONDRIAL IMPORT INNER MEMBRANE TRANSLOCASE SUBUNIT TIM10 (PTHR11038:SF20)
H2XQ34	SOLUBLE NSF ATTACHMENT PROTEIN (PTHR13768:SF15)
F6T060	SUBFAMILY NOT NAMED (PTHR43272:SF13)
F6TF08	CYTOCHROME P450 18A1 (PTHR24300:SF232)
F7B9N2	MAJOR VAULT PROTEIN (PTHR14165:SF13)
F6SHG8	NADH DEHYDROGENASE [UBIQUINONE] IRON-SULFUR PROTEIN 2, MITOCHONDRIAL (PTHR11993:SF32)
F6W4W6	ADSEVERIN (PTHR11977:SF73)
F6TWF6	ALPHA-AMINOADIPIC SEMIALDEHYDE DEHYDROGENASE (PTHR43521:SF1)
H2Y0C3	U6 SNRNA-ASSOCIATED SM-LIKE PROTEIN LSM5 (PTHR20971:SF1)
H2XL87	FI01544P (PTHR24073:SF622)
H2XSU1	PROTEASOME SUBUNIT BETA TYPE-1 (PTHR11599:SF96)
F6ZZF2	PHOSPHOENOLPYRUVATE CARBOXYKINASE [GTP], MITOCHONDRIAL (PTHR11561:SF10)
H2XQN0	NK LYSIN-LIKE PROTEIN (PTHR11480:SF59)
F7AX62	GLUCOSE-6-PHOSPHATE 1-DEHYDROGENASE (PTHR23429:SF11)
F6PUC9	DEHYDROGENASE/REDUCTASE SDR FAMILY MEMBER 11 (PTHR43115:SF4)
H2Y1B2	RAS-RELATED PROTEIN RAB-2A (PTHR24073:SF681)
F6W5P6	SUBFAMILY NOT NAMED (PTHR11699:SF261)
F6ZPP4	ATP SYNTHASE-COUPPLING FACTOR 6, MITOCHONDRIAL (PTHR12441:SF12)
F6QZ03	LP01981P-RELATED (PTHR22811:SF81)
F6WZ22	SUBFAMILY NOT NAMED (PTHR10529:SF293)
H2XT58	SUBFAMILY NOT NAMED (PTHR24346:SF33)
F6ZJB9	MICOS COMPLEX SUBUNIT MIC60 (PTHR15415:SF10)
H2XNI9	NADH DEHYDROGENASE [UBIQUINONE] IRON-SULFUR PROTEIN 4, MITOCHONDRIAL (PTHR12219:SF19)
F6XJL4	SUBFAMILY NOT NAMED (PTHR31964:SF67)
F7AFI4	ERYTHROCYTE BAND 7 INTEGRAL MEMBRANE PROTEIN (PTHR10264:SF109)

F7A8C2	ATPASE, H+ TRANSPORTING, LYSOSOMAL V1 SUBUNIT B2-RELATED (PTHR43389:SF5)
F7AKV1	SUBFAMILY NOT NAMED (PTHR11977:SF59)
F7AMH7	CALPAIN-9 (PTHR10183:SF330)
H2Y3G1	LATENT-TRANSFORMING GROWTH FACTOR BETA-BINDING PROTEIN 3 (PTHR24034:SF69)
F6WQG9	BETA-UREIDOPROPIONASE (PTHR43674:SF1)
F6Y6F6	FAD-DEPENDENT OXIDOREDUCTASE DOMAIN-CONTAINING PROTEIN 2 (PTHR43539:SF12)
F6TWU9	PYRUVATE CARBOXYLASE, MITOCHONDRIAL (PTHR43778:SF4)
F6UAX7	FI08027P-RELATED (PTHR23139:SF89)
H2XNA4	CALCIUM-TRANSPORTING ATPASE SARCOPLASMIC/ENDOPLASMIC RETICULUM TYPE (PTHR42861:SF24)
F6Z3E3	UDP-GALACTOPYRANOSE MUTASE (PTHR21197:SF1)
F6Q7A6	ALDEHYDE DEHYDROGENASE TYPE III, ISOFORM Q (PTHR43570:SF6)
F7BBJ9	PYRUVATE DEHYDROGENASE E1 COMPONENT SUBUNIT BETA, MITOCHONDRIAL (PTHR11624:SF89)
F7A6C5	MYOSIN BINDING SUBUNIT, ISOFORM O (PTHR24179:SF28)
F6TUG7	NADH DEHYDROGENASE [UBIQUINONE] 1 BETA SUBCOMPLEX SUBUNIT 6 (PTHR15083:SF1)
H2XWY3	K01C8.1 (PTHR43419:SF9)
F6S5A6	DIPEPTIDASE B, ISOFORM A (PTHR11963:SF27)
F6V2S0	TRANSGELIN (PTHR18959:SF68)
H2XTB8	APOPTOSIS INHIBITOR 5 (PTHR12758:SF32)
H2XY10	ADENYLYL CYCLASE-ASSOCIATED PROTEIN (PTHR10652:SF14)
F6Q106	GLUTARYL-COA DEHYDROGENASE, MITOCHONDRIAL (PTHR42807:SF1)
F6XQV0	MICROSOMAL TRIACYLGLYCEROL TRANSFER PROTEIN (PTHR13024:SF5)
F6X863	GLUTATHIONE PEROXIDASE (PTHR11592:SF52)
H2XV39	STATHMIN, ISOFORM D (PTHR10104:SF26)
F6Q5V5	SPECTRIN BETA CHAIN (PTHR11915:SF414)
F6PKL1	IMPORTIN SUBUNIT ALPHA (PTHR23316:SF48)
F6Q2H6	GOLGI APPARATUS PROTEIN 1 (PTHR11884:SF5)

H2XMP3	EPIDIDYMAL SECRETORY PROTEIN E1 (PTHR11306:SF26)
F6YKT9	ACYL-COA DEHYDROGENASE FAMILY MEMBER 9, MITOCHONDRIAL (PTHR43884:SF7)
F6ZV29	RING CANAL KELCH PROTEIN (PTHR24412:SF308)
F6ZE91	LACTOYLGLUTATHIONE LYASE (PTHR10374:SF35)
H2Y050	SERINE PROTEASE HTRA2, MITOCHONDRIAL (PTHR22939:SF117)
F6Z215	CYTOCHROME C OXIDASE SUBUNIT 6C (PTHR12916:SF3)
H2XWU8	THIOSULFATE SULFURTRANSFERASE RDL1, MITOCHONDRIAL (PTHR42757:SF5)
F7AVY6	ATP-BINDING CASSETTE SUB-FAMILY F MEMBER 1 (PTHR19211:SF67)
H2Y1S7	PROTEIN DISULFIDE-ISOMERASE A6 (PTHR18929:SF127)
F7B6S8	S-FORMYLGLUTATHIONE HYDROLASE (PTHR10061:SF2)
F6SDE2	LD38919P (PTHR10677:SF34)
F6PSX9	GRANULINS (PTHR12274:SF6)
F6PIL6	Uncharacterized protein
F6QHG3	Uncharacterized protein
F6QXG1	PROTEIN TRANSPORT PROTEIN SEC31A
F6R3I9	INOSITOL HEXAKISPHOSPHATE AND DIPHOSPHOINOSITOL-PENTAKISPHOSPHATE KINASE 2
F6R3X1	ECHINODERM MICROTUBULE-ASSOCIATED PROTEIN-LIKE 2 ISOFORM X1
F6SGH3	SERINE/ARGININE-RICH SPLICING FACTOR 1A
F6SGW2	TRANSMEMBRANE PROTEIN VC16 PRECURSOR
F6SMG7	FIBROPELLIN-1-LIKE
F6TSY3	DNAJ HOMOLOG SUBFAMILY C MEMBER 10
F6U3B3	ADP-RIBOSYLATION FACTOR-LIKE
F6UMT8	CYSTATHIONINE GAMMA-LYASE ISOFORM X1
F6VB79	LOW-DENSITY LIPOPROTEIN RECEPTOR-RELATED PROTEIN 2-LIKE
F6VZ12	NON-MUSCLE CALDESMON ISOFORM X1
F6WZT7	U4/U6 SMALL NUCLEAR RIBONUCLEOPROTEIN PRP31
F6YF15	UNCHARACTERIZED PROTEIN
F6YF21	UNCHARACTERIZED PROTEIN
F6YX55	UNCHARACTERIZED PROTEIN
F7ACZ6	BRAIN ACID SOLUBLE PROTEIN 1
F7AFW2	RNA POLYMERASE II DEGRADATION FACTOR 1
F7AGD6	UNCHARACTERIZED PROTEIN

F7BPZ5	YTH DOMAIN-CONTAINING FAMILY PROTEIN 3
H2XJV7	NUCLEOPORIN NUP211-LIKE
H2XKR2	ACTIN-RELATED PROTEIN 3-LIKE
H2XMW7	UNCHARACTERIZED PROTEIN
H2XT84	EH DOMAIN-CONTAINING PROTEIN 3-LIKE
H2XUE9	UNCHARACTERIZED PROTEIN
H2XUN1	UNCHARACTERIZED PROTEIN
H2XWA3	CYTOPLASMIC DYNEIN 1 HEAVY CHAIN 1
H2XWW6	UNCHARACTERIZED PROTEIN
H2XZ26	NEURAL CELL ADHESION MOLECULE 1-LIKE
H2XZC8	SUCRASE-ISOMALTASE, INTESTINAL-LIKE

Table 5. BLAST2GO results.

This table shows the number of proteins from each sample that fall into each functional category. The percentage of proteins that fell into each category was then taken. There are more proteins in the 18° C samples overall. The top and bottom rows show the smallest and largest difference respectively. The largest difference was in response to stimulus, Any process that results in a change in state or activity of a cell or an organism (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a stimulus. The process begins with detection of the stimulus and ends with a change in state or activity of the cell or organism (Carbon, et al., 2009). The smallest difference was in negative regulation of biological processes, Any process that stops, prevents, or reduces the frequency, rate or extent of a biological process. Biological processes are regulated by many means; examples include the control of gene expression, protein modification or interaction with a protein or substrate molecule (Carbon, et al., 2009).

Table 5.

GO term	# of seq 8 °C	# of seq 18 °C	percent 8 °C	percent 18 °C
negative regulation of biological process	8	14	36.36	63.64
cellular component organization or biogenesis	17	33	34.00	66.00
localization	14	31	31.11	68.89
metabolic process	33	89	27.05	72.95
cellular process	33	90	26.83	73.17
regulation of biological process	15	44	25.42	74.58
biological regulation	15	52	22.39	77.61
multicellular organismal process	5	21	19.23	80.77
positive regulation of biological process	4	17	19.05	80.95
signaling	4	18	18.18	81.82
single-organism process	19	89	17.59	82.41
developmental process	5	24	17.24	82.76
response to stimulus	6	34	15.00	85.00

Table 6. KEGG terms unique to the cold-water data set.

Over all there are far fewer pathways represented in the cold-water data than in the control samples. This is likely due to dormancy and the lack of active reproduction. During dormancy metabolism slows and many non-essential functions are turned off. Numbers in parentheses are the number of proteins assigned to the pathway.

Table 6.

KEGG Pathways Unique to 8 °C Data
00760 Nicotinate and nicotinamide metabolism (1)
03013 RNA transport (3)
03015 mRNA surveillance pathway (1)
03060 Protein export (1)
04011 MAPK signaling pathway - yeast (1)
04139 Regulation of mitophagy - yeast (1)
04214 Apoptosis - fly (1)
04330 Notch signaling pathway (1)
04961 Endocrine and other factor-regulated calcium reabsorption (1)

Table 7. KEGG terms unique to the 18° C sample.

This list contains far more pathways than the cold-water data set. Therefore, it can be implied that more pathways are functioning when the animal is actively reproducing. Numbers in parentheses are the number of proteins assigned to the pathway.

Table 7.

KEGG Pathways unique to 18° C Data	
00010 Glycolysis / Gluconeogenesis (4)	04611 Platelet activation (4)
00020 Citrate cycle (TCA cycle) (4)	04612 Antigen processing and presentation (1)
00030 Pentose phosphate pathway (1)	04614 Renin-angiotensin system (1)
00040 Pentose and glucuronate interconversions (1)	04620 Toll-like receptor signaling pathway (2)
00052 Galactose metabolism (2)	04630 Jak-STAT signaling pathway (1)
00053 Ascorbate and aldarate metabolism (2)	04650 Natural killer cell mediated cytotoxicity (1)
00061 Fatty acid biosynthesis (1)	04660 T cell receptor signaling pathway (1)
00062 Fatty acid elongation (1)	04662 B cell receptor signaling pathway (2)
00071 Fatty acid degradation (8)	04664 Fc epsilon RI signaling pathway (2)
00072 Synthesis and degradation of ketone bodies (1)	04666 Fc gamma R-mediated phagocytosis (4)
00190 Oxidative phosphorylation (13)	04668 TNF signaling pathway (1)
00240 Pyrimidine metabolism (3)	04670 Leukocyte transendothelial migration (3)
00250 Alanine, aspartate and glutamate metabolism (1)	04713 Circadian entrainment (1)
00260 Glycine, serine and threonine metabolism (4)	04722 Neurotrophin signaling pathway (2)
00270 Cysteine and methionine metabolism (2)	04723 Retrograde endocannabinoid signaling (1)
00280 Valine, leucine and isoleucine degradation (7)	04724 Glutamatergic synapse (1)
00300 Lysine biosynthesis (1)	04725 Cholinergic synapse (2)
00310 Lysine degradation (6)	04726 Serotonergic synapse (1)
00330 Arginine and proline metabolism (4)	04727 GABAergic synapse (1)
00340 Histidine metabolism (2)	04728 Dopaminergic synapse (2)
00362 Benzoate degradation (2)	04730 Long-term depression (1)
00410 beta-Alanine metabolism (4)	04914 Progesterone-mediated oocyte maturation (2)
00450 Selenocompound metabolism (2)	04915 Estrogen signaling pathway (3)
00460 Cyanoamino acid metabolism (1)	04916 Melanogenesis (1)
00480 Glutathione metabolism (5)	04917 Prolactin signaling pathway (1)
00511 Other glycan degradation (1)	04918 Thyroid hormone synthesis (2)
00520 Amino sugar and nucleotide sugar metabolism (3)	04919 Thyroid hormone signaling pathway (1)

00531 Glycosaminoglycan degradation (1)	04920 Adipocytokine signaling pathway (3)
00561 Glycerolipid metabolism (2)	04921 Oxytocin signaling pathway (2)
00590 Arachidonic acid metabolism (2)	04922 Glucagon signaling pathway (3)
00603 Glycosphingolipid biosynthesis - globo series (1)	04923 Regulation of lipolysis in adipocytes (2)
00604 Glycosphingolipid biosynthesis - ganglio series (1)	04924 Renin secretion (1)
00620 Pyruvate metabolism (7)	04931 Insulin resistance (2)
00625 Chloroalkane and chloroalkene degradation (1)	04932 Non-alcoholic fatty liver disease (NAFLD) (12)
00627 Aminobenzoate degradation (1)	04933 AGE-RAGE signaling pathway in diabetic complications (2)
00640 Propanoate metabolism (3)	04962 Vasopressin-regulated water reabsorption (2)
00650 Butanoate metabolism (3)	04964 Proximal tubule bicarbonate reclamation (1)
00670 One carbon pool by folate (1)	04966 Collecting duct acid secretion (1)
00680 Methane metabolism (2)	04971 Gastric acid secretion (1)
00720 Carbon fixation pathways in prokaryotes (4)	04972 Pancreatic secretion (3)
00770 Pantothenate and CoA biosynthesis (1)	04973 Carbohydrate digestion and absorption (1)
00830 Retinol metabolism (1)	04975 Fat digestion and absorption (1)
00900 Terpenoid backbone biosynthesis (1)	05010 Alzheimer's disease (13)
00903 Limonene and pinene degradation (1)	05012 Parkinson's disease (13)
00930 Caprolactam degradation (1)	05030 Cocaine addiction (1)
00980 Metabolism of xenobiotics by cytochrome P450 (1)	05032 Morphine addiction (1)
00982 Drug metabolism - cytochrome P450 (1)	05034 Alcoholism (1)
00983 Drug metabolism - other enzymes (1)	05100 Bacterial invasion of epithelial cells (3)
01120 Microbial metabolism in diverse environments (13)	05120 Epithelial cell signaling in Helicobacter pylori infection (2)
01210 2-Oxocarboxylic acid metabolism (1)	05130 Pathogenic Escherichia coli infection (3)
01212 Fatty acid metabolism (5)	05131 Shigellosis (3)
01230 Biosynthesis of amino acids (6)	05132 Salmonella infection (3)
01521 EGFR tyrosine kinase inhibitor resistance (1)	05133 Pertussis (2)
01522 Endocrine resistance (1)	05134 Legionellosis (1)

01524 Platinum drug resistance (2)	05140 Leishmaniasis (1)
02020 Two-component system (1)	05142 Chagas disease (American trypanosomiasis) (2)
02024 Quorum sensing (1)	05145 Toxoplasmosis (3)
03320 PPAR signaling pathway (3)	05146 Amoebiasis (2)
04012 ErbB signaling pathway (1)	05152 Tuberculosis (2)
04013 MAPK signaling pathway - fly (1)	05160 Hepatitis C (1)
04020 Calcium signaling pathway (1)	05161 Hepatitis B (1)
04022 cGMP-PKG signaling pathway (4)	05162 Measles (1)
04024 cAMP signaling pathway (4)	05164 Influenza A (3)
04062 Chemokine signaling pathway (3)	05203 Viral carcinogenesis (2)
04066 HIF-1 signaling pathway (2)	05204 Chemical carcinogenesis (1)
04070 Phosphatidylinositol signaling system (1)	05210 Colorectal cancer (2)
04071 Sphingolipid signaling pathway (3)	05211 Renal cell carcinoma (2)
04142 Lysosome (3)	05212 Pancreatic cancer (3)
04150 mTOR signaling pathway (2)	05213 Endometrial cancer (1)
04151 PI3K-Akt signaling pathway (4)	05214 Glioma (1)
04152 AMPK signaling pathway (3)	05215 Prostate cancer (1)
04210 Apoptosis (2)	05218 Melanoma (1)
04270 Vascular smooth muscle contraction (1)	05221 Acute myeloid leukemia (1)
04360 Axon guidance (3)	05222 Small cell lung cancer (2)
04370 VEGF signaling pathway (2)	05223 Non-small cell lung cancer (1)
04380 Osteoclast differentiation (2)	05230 Central carbon metabolism in cancer (3)
04510 Focal adhesion (4)	05231 Choline metabolism in cancer (2)
04512 ECM-receptor interaction (1)	05322 Systemic lupus erythematosus (1)
04514 Cell adhesion molecules (CAMs) (2)	05323 Rheumatoid arthritis (1)
04520 Adherens junction (1)	05412 Arrhythmogenic right ventricular cardiomyopathy (ARVC) (1)
04540 Gap junction (2)	05416 Viral myocarditis (1)
04550 Signaling pathways regulating pluripotency of stem cells (1)	Glycan biosynthesis and metabolism

Table 8. KEGG terms that appear in both data sets.

Terms that appear in both sets are assumed to have no significant change in the pathway function. Numbers in parentheses are the number of proteins assigned to the pathway.

Table 8.

KEGG Pathways in Both Data Sets	
00230	Purine metabolism (1)
00380	Tryptophan metabolism (5)
00630	Glyoxylate and dicarboxylate metabolism (4)
00970	Aminoacyl-tRNA biosynthesis (1)
01100	Metabolic pathways (41)
01110	Biosynthesis of secondary metabolites (12)
01130	Biosynthesis of antibiotics (14)
01200	Carbon metabolism (10)
03008	Ribosome biogenesis in eukaryotes (1)
03018	RNA degradation (1)
03040	Spliceosome (7)
03050	Proteasome (2)
04010	MAPK signaling pathway (2)
04014	Ras signaling pathway (4)
04015	Rap1 signaling pathway (5)
04068	FoxO signaling pathway (2)
04072	Phospholipase D signaling pathway (2)
04141	Protein processing in endoplasmic reticulum (6)
04144	Endocytosis (5)
04145	Phagosome (7)
04146	Peroxisome (1)
04211	Longevity regulating pathway (1)
04212	Longevity regulating pathway - worm (1)
04213	Longevity regulating pathway - multiple species (1)
04260	Cardiac muscle contraction (2)
04261	Adrenergic signaling in cardiomyocytes (2)
04310	Wnt signaling pathway (1)
04530	Tight junction (2)
04721	Synaptic vesicle cycle (2)
04810	Regulation of actin cytoskeleton (5)
04910	Insulin signaling pathway (2)
05014	Amyotrophic lateral sclerosis (ALS) (1)
05016	Huntington's disease (12)
05110	Vibrio cholerae infection (1)
05166	HTLV-I infection (2)
05168	Herpes simplex infection (1)
05169	Epstein-Barr virus infection (1)
05200	Pathways in cancer (5)
05205	Proteoglycans in cancer (4)

05220 Chronic myeloid leukemia (1)
05410 Hypertrophic cardiomyopathy (HCM) (1)
05414 Dilated cardiomyopathy (1)

Figure 7.
Pie chart showing the number of unique
proteins at each temperature.

Figure 7.

Bibliography

- Airi, V. et al., 2014. Reproductive Efficiency of a Mediterranean Endemic Zooxanthellate Coral Decreases with Increasing Temperature along a Wide Latitudinal Gradient. *PLoS One*, 9(3), p. e91792.
- Brockington, S. & Clarke, A., 2001. The relative influence of temperature and food on the metabolism of a marine invertebrate. *Journal of Experimental Marine Biology and Ecology*, 258(1), pp. 87-99.
- Carbon S, Ireland A, Mungall CJ, Shu S, Marshall B, Lewis S, AmiGO Hub, Web Presence Working Group. AmiGO: online access to ontology and annotation data. *Bioinformatics*. Jan 2009;25(2):288-9.
- Carver, C., Mallet, A. & Vercaemer, . B., 2006. *Biological Synopsis of the Solitary Tunicate C. intestinalisintestinalis*, Dartmouth, Nova Scotia: Fisheries and Oceans Canada Bedford Institute of Oceanography .
- Doney, S. C., 2011. Climate Change Impacts on Marine Ecosystems. *Annual Review of Marine Science*, Volume 4, pp. 11-37.
- Dybern, B. I., 1965. The Life Cycle of *C. intestinalisintestinalis* (L.) f. typica in Relation to the Environmental Temperature. *Oikos*, pp. 109-131.
- Endo, T. et al., n.d. CIPRO 2.5: *C. intestinalisintestinalis* protein database, a unique integrated repository of large-scale omics data, bioinformatic analyses and curated annotation, with user rating and reviewing functionality. *Nucleic Acids Res n.d.*, pp. D807-D814.
- The Gene Ontology Consortium. Gene Ontology Consortium: going forward. (2015) *Nucl Acids Res* **43** Database issue D1049–D1056
- Horgan, R. P. & Kenny, L. C., 2011. SAC review ‘Omic’ technologies: genomics, transcriptomics, proteomics, and metabolomics. *The Obstetrician & Gynaecologist*, Volume 13, p. 189–195.
- Iannelli, F., Pesole, G. & Sordin, P., 2007. Mitogenomics reveals two cryptic species in *C. intestinalisintestinalis*. *Trends in Genetics*, 23(9), p. 419–422.
- James, P. & Siikavuopio, S., 2011. *A Guide to the Sea Urchin Reproductive Cycle and Staging Sea Urchin Gonad Samples*. [Online]
Available at: http://www.nofima.no/filearchive/guide-to-sea-urchins_lowres.pdf
- NOAA, 2015. *Water Temperature Table of the Northern Atlantic Coast*. [Online]
Available at: <https://www.nodc.noaa.gov/cwtg/natl.html>

Petersen, J. K., Mayre, S. & Knudsen, M. Å., 1999. Beat frequency of cilia in the branchial basket of the ascidian *C. intestinalis* in relation to temperature and algal cell concentration. *Marine Biology*, 133(2), pp. 185-192.

Serafinia, L., Hanna, J. B., Kültzb, . D. & Tomaneka, . L., 2011. The proteomic response of sea squirts (genus *Ciona*) to acute heat stress: A global perspective on the thermal stability of proteins. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics*, p. 322–334.

Song, J. L. & Wessel, G. M., 2005. How to make an egg: transcriptional regulation in oocytes. *International Society of Differentiation*, Volume 73, p. 1–17.

Tomanek, L., 2011. Environmental Proteomics: Changes in the Proteome of Marine Organisms in Response to Environmental Stress, Pollutants, Infection, Symbiosis, and Development.. *Annu. Rev. Mar. Sci*, Volume 3, p. 373–99.

Wright, P., Noirela, J., Owa , . S.-Y. & Fazelib, A., 2012. A review of current proteomics technologies with a survey on their widespread use in reproductive biology investigations. *Theriogenology*, 77(4), p. 738–765.

APPENDICES

Appendix 1

Proteomics Core Rhode Island Hospital Protocol

(How the samples were processed by the mass spectrometer facility)

Methods

LC-MS/MS analysis

LC/MS was performed as described previously (Nguyen, et al., 2009). Tryptic peptides were analyzed by a fully automated proteomic technology platform (Yu & Salomon, 2009), (Yu & Salomon, 2010). The nanoLC-MS/MS experiments were performed with an Agilent 1200 Series Quaternary HPLC system (Agilent Technologies, Santa Clara, CA) connected to a “Q Exactive Plus” mass spectrometer (Thermo Fisher Scientific, Waltham, MA). The lyophilized tryptic peptides were reconstituted in buffer A (0.1 M acetic acid) at a concentration of 1 $\mu\text{g}/\mu\text{l}$ and 5 μl was injected for each analysis. The electrospray ion source was operated at 2.0 kv in a split flow configuration, as described previously (Elias & Gygi, 2007). The Q Exactive was operated in the data dependent mode using a top-9 data dependent method. Survey full scan MS spectra (m/z 400-1800) were acquired at a resolution of 70,000 with an AGC target value of 3×10^6 ions or a maximum ion injection time of 200 ms. Peptide fragmentation was performed via higher-energy collision dissociation (HCD) with the energy set at 28 NCE. The MS/MS spectra were acquired at a resolution of 17,500, with a targeted value of 2×10^4 ions or a maximum integration time of 250 ms. The under fill ratio, which specifies the minimum percentage of the target value likely to be reached at maximum fill time, was defined as 1.0%. The ion selection abundance

threshold was set at 8.0×10^2 with charge state exclusion of unassigned and $z=1$, or 6-8 ions and dynamic exclusion time of 20 seconds.

Data analysis

Peptide spectrum matching of MS/MS spectra from whole cell lysate tryptic digest samples was performed against a human-specific database (UniProt; downloaded 2/1/2013) using MASCOT v. 2.4 (Matrix Science, Ltd, London W1U 7GB UK). A concatenated database containing 144,156 “target” and “decoy” sequences was employed to estimate the false discovery rate (FDR). Msconvert from ProteoWizard (v. 3.0.5047), using default parameters and with the MS2Deisotope filter on, was employed to create peak lists for Mascot. Mascot database searches were performed with the following parameters: trypsin enzyme cleavage specificity, 2 possible missed cleavages, 7 ppm mass tolerance for precursor ions, 20 mmu mass tolerance for fragment ions. Search parameters permitted variable modification of methionine oxidation (+15.9949 Da), and static modification of carbamidomethylation (+57.0215 Da) on cysteine. The resulting peptide spectrum matches (PSMs) were reduced to sets of unique PSMs by eliminating lower scoring duplicates. Mascot results were filtered by Mowse Score (>20). Peptide assignments from the database search were filtered down to 1% false discovery rate (FDR) by a logistic spectral score, as previously described (Ficarro, et al., 2005), (Yu, et al., 2009).

Quantitation of relative peptide abundance

Relative quantification of peptide abundance was performed via calculation of selected ion chromatograms (SIC) peak areas. Retention time alignment of individual replicate analyses was performed as previously described (Demirkan, et al.,

2011). Peak areas were calculated by inspection of SICs using in-house software programmed in R 3.0 based on the Scripps Center for Metabolomics' XCMS package (version 1.40.0). This approach performed multiple passes through XCMS's central wavelet transformation algorithm (implemented in the centWave function) over increasingly narrower ranges of peak widths, and used the following parameters: mass window of 10 ppm, minimum peak widths ranging from 2 to 20 seconds, maximum peak width of 80 seconds, signal to noise threshold of 10, and detection of peak limits via descent on the non-transformed data enabled. For cases when centWave did not identify an MS peak, we used the getPeaks function available in XCMS to integrate in a pre-defined region surrounding the maximum intensity signal of the SIC. SIC peak areas were determined for every peptide that was identified by MS/MS. In the case of a missing MS/MS for a particular peptide, in a particular replicate, the SIC peak area was calculated according to the peptide's isolated mass and the retention time calculated from retention time alignment. A minimum SIC peak area equivalent to the typical spectral noise level of 1000 was required of all data reported for label-free quantitation. Individual SIC peak areas were normalized to the peak area of the exogenously spiked synthetic peptide DRVYHPF added prior to reversed-phase elution into the mass spectrometer. *P*-values were calculated from five replicates. To select peptides that show a statistically significant change in abundance between control and treatment, two-tailed unpaired Student's *t* tests and *q*-values for multiple hypothesis tests were calculated based on the determined *p*-values using the R package QVALUE as previously described (Storey, 2003), (Storey & Tibshirani, 2003).

Nguyen, V.; Cao, L.; Lin, J. T.; Hung, N.; Ritz, A.; Yu, K.; Jianu, R.; Ulin, S. P.; Raphael, B. J.; Laidlaw, D. H.; Brossay, L.; Salomon, A. R., A new approach for quantitative phosphoproteomic dissection of signaling pathways applied to T cell receptor activation. *Mol Cell Proteomics* **2009**, 8, (11), 2418-31.

Yu, K.; Salomon, A. R., PeptideDepot: flexible relational database for visual analysis of quantitative proteomic data and integration of existing protein information. *Proteomics* **2009**, 9, (23), 5350-8.

Yu, K.; Salomon, A. R., HTAPP: high-throughput autonomous proteomic pipeline. *Proteomics* **2010**, 10, (11), 2113-22.

Elias, J. E.; Gygi, S. P., Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods* **2007**, 4, (3), 207-14.

Ficarro, S. B.; Salomon, A. R.; Brill, L. M.; Mason, D. E.; Stettler-Gill, M.; Brock, A.; Peters, E. C., Automated immobilized metal affinity chromatography/nano-liquid chromatography/electrospray ionization mass spectrometry platform for profiling protein phosphorylation sites. *Rapid Commun Mass Spectrom* **2005**, 19, (1), 57-71.

Yu, K.; Sabelli, A.; DeKeukelaere, L.; Park, R.; Sindi, S.; Gatsonis, C. A.; Salomon, A., Integrated platform for manual and high-throughput statistical validation of tandem mass spectra. *Proteomics* **2009**, 9, (11), 3115-25.

Demirkan, G.; Yu, K.; Boylan, J. M.; Salomon, A. R.; Gruppuso, P. A., Phosphoproteomic profiling of in vivo signaling in liver by the mammalian target of rapamycin complex 1 (mTORC1). *PloS One* **2011**, 6, (6), e21729.

Storey, J. D., The positive false discovery rate: A Bayesian interpretation and the q-value. *Annals of Statistics* **2003**, 31, (6), 2013-35.

Storey, J. D.; Tibshirani, R., Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* **2003**, 100, (16), 9440-5.

Appendix 2

Stem Cell Gene Expressed During Reproduction in *Diplosoma listerianum*

Abstract

Diplosoma listerianum is a colonial ascidian, which can reproduce both sexually and asexually. It has been theorized that the asexual reproductive process uses stem cells to form the asexual bud tissue (Brown & Swalla, 2012). For this project, animals were collected, dissected and RNA extracted. *DlOct2* (proposed name for gene) was isolated from the RNA and cloned. The cloned gene was sent for Sanger sequencing which returned a clean sequence. This is the first time this gene has been successfully cloned from *Diplosoma listerianum*. We hypothesize that *DlOct2*, a stem cell associated gene, will be expressed in the bud and precursor tissue. In the future, the DNA sequence from *DlOct2* will be used to design a probe that will be used for *in situ* hybridization trials in order to locate the site of expression.

Introduction

Diplosoma listerianum is an invasive species of tunicate that is believed to have originated on the Northwest coast of North America. It is now found all over the world in warm, temperate and cold waters. It is likely spread by encrusting on shipping vessels or in ballast water (Lambert, 2001). It continues to spread to new places due to its rapid growth, and can displace local species. It smothers bivalves and other animals in the filter feeding community (Dijkstra, et al., 2007). Each colony contains many small animals about 2 mm in length, but can be up to 20 cm across. In the United Kingdom, scientists received the unique opportunity to see how

introduction happens first hand. In 2005, *Diplosoma listerianum* was not present in the UK, but by 2009, it became the dominant species in the fouling community (Vance, Lauterbatch and Wahl, 2009). This showed that it only takes a few years for *Diplosoma listerianum* to dominate a new habitat and oust local organisms. This is why it is important to understand how this organism spreads by looking into its reproductive capabilities.

In other colonial tunicates such as *Botryllid* ascidians, there are multi and pluripotent stem cells that contribute to both regeneration of the colony as well as reproduction (Kurn, Rendulic and Tiozzo, 2011). However, the molecular mechanisms that control these processes are not well understood. Therefore, this study will provide more insight into the genetic base for the claim that stem cells give rise to bud tissue (Brown and Swalla, 2012). The gene used as a marker was *DlOct2*, which is a transcription factor believed to be associated with stem-cells (Pan, et al., 2002). *DlOct2* is part of the POUgene family.

Another reason for using *Diplosoma listerianum* to study this phenomenon is that the didemnids have a unique method of semi-conservative budding that is not used by other ascidians (N.J.Berrill, 1935). There is controversy about which tissues give rise to buds in didemnids. For example, it could be the epithelial tissue surrounding the esophagus or the epicardial tissue adjacent to the esophagus. By performing *in situ* hybridization using stem cell genes, it may be possible to pinpoint the tissue from which the buds originate. Due to the small size and delicate nature of the animal, it makes sectioning techniques difficult, making histology impractical. Overall, this research is important to understand the role that stem cells play in

reproduction and how the *DlOct2* gene is associated with stem cells in *Diplosoma listerianum*.

Methods

Study Organism:

Diplosoma listerianum is a colonial ascidian, which is a member of the fouling community and an invasive species to Rhode Island. The animals used in this study came from Point Judith Pond, Rhode Island. These animals are clear and fragile making them difficult to find and collect, but making RNA extraction and dissection of individual animals easy.

Collection and Dissection:

Diplosoma listerianum specimens were collected from the side of docks or from plates suspended in the water. The specimens were removed from the substrate and brought back to the University of Rhode Island main Campus where they were kept alive in a closed system tank for up to 2 weeks before they were dissected. A colony was placed into a Petri dish of filtered seawater and using a stereoscope individual zooids were dissected. The upper tunic was removed and the animals were detached from their lower tunic using sharp tweezers and placed into a 1.5 ml tube, which was kept on ice. After the 1.5 ml tube was full, it was flash frozen in either liquid nitrogen or an ethanol/dry ice bath at -80 °C for future isolation of RNA. Another 1.5 ml tube was used to collect individual zooids, which were fixed in 4% formaldehyde in PBS and stored at -20 °C in ethanol, for *in situ* hybridization.

RNA Preparation:

The frozen samples were homogenized with a cell lysis buffer in 1.5 ml tubes with a plastic sterile pestle. Total RNA was extracted by adding one volume of phenol /chloroform to the homogenate, vortexed for 1 minute, and spun in a microcentrifuge at 13,000 g for 5 min. The aqueous phase was transferred to a new 1.5 ml tube and extracted again. Two volumes of 100% ethanol were added to the tube, which was placed in the -20 °C freezer for 10 minutes. The sample was spun at 13,000 g for 10 minutes to precipitate the RNA. The 100% ethanol was removed and 70% ethanol was added to wash the pellet. It was again spun for 5 minutes the ethanol removed and the pellet allowed to air dry. After drying, nuclease-free water was added to suspend the RNA. DNase was added to eliminate any DNA molecules. The PolyAtract mRNA Isolation System III from Promega was used to extract and purify the mRNA from the total RNA. cDNA was made from mRNA using the Protoscript reverse transcriptase kit from New England Biolabs with a poly-T anchor primer.

PCR and Gel Purification:

The cDNA was used as the template for PCR using degenerative primers for *DlOct2*, which were designed using the sequence from *Ciona intestinalis*. The primer sequences used were 5'- ACNCAGGGNGATGTNGGNC and 5'- GNCGNCGNTTGCARAACCA. The PCR product was run on an agarose gel to confirm the presence of a band at the expected size. Expected size was determined by locating where the primers would fall in the *C. intestinalis* genome and determining the number of nucleotides that fell between them giving an approximate band size. The band was excised and the DNA purified using MinElute Gel Extraction Kit from

Qiagen. Another PCR was run using the purified gel band as template to amplify the specific sequence in order to increase the amount of DNA for later steps. A phenol chloroform extraction followed by ethanol precipitation (same as in above method) purified the sample, and a portion was sent to the Genome Sequencing Center (GSC) at University of Rhode Island for Sanger sequencing. The sequence was run through NCBI BLAST to confirm that it contained the expected gene.

Transformation:

The PCR product was cloned into pUC18 vector using a blunt end ligation and transformed into E. coli strain Top 10. Pure Yield Plasmid Miniprep System from Promega was used to purify the plasmid, which was sent for sequencing at the GSC.

Sequence Alignment:

The Sequence returned from the sequencing reaction was trimmed to get rid of vector and primer sequence. BLAST was used to check that the sequence was an Oct homolog. *Ciona*, frog and human homologs taken from the NCBI database (see table 2) were used for the alignment (U.S. National Library of Medicine, 2016). ExPASy Translator was used to translate the nucleotide sequence into amino acid sequence. The UniProt UGENE program was used to align the sequences (see figure 1).

Results and Discussion

Diplosoma listerianum was collected and the individual zooids dissected, RNA extracted, and cDNA synthesized. *DlOct2* was isolated from the cDNA, cloned, and the plasmid sent for Sanger sequencing. This is the first successful attempt at isolating and cloning a stem cell regulatory gene from *Diplosoma listerianum*. Its

genome has yet to be sequenced and its genes lack functional classification or expression data. Due to the anatomy of *Diplosoma listerianum*, it is easy to dissect and homogenize. Therefore, the methods used in this project can be used to isolate other genes from this organism.

The *DlOct2* fragment isolated is about 120 amino acids long and is about 360 base pairs in length (Figure 2). When our *DlOct2* sequence is run through the BLAST database, Oct 2 or POU2f2 (the name of the human homolog) is the top hit. Oct 2 is found in mammalian B-cells and is a transcription factor regulating immunoglobulin genes in mammals (Corcoran, et al., 1993). On the POU family evolutionary tree, POU 2 diverged directly before POU 5 (Gold, et al., 2014). Oct 4 or POU5f1 codes for a transcription factor that activates genes involved in inducing pluripotent stem cells in mammals (Pan, et al., 2002). However, it is unclear if uro-chordates also have this gene or if the POU genes that it possess have different functionality or classification than those found in mammals (Gold, et al., 2014).

The sequence alignment shows small areas that are highly conserved which are consistent with the previous alignments (Rosenfeld, 1991). The reason for choosing the genes in the alignment was to get a good sample of the different POU genes including POU2f2, POU2f1, POU5f1, POU3f4, and POU4. This covers almost all of the POU groups. By having this wide range, it shows that sequence similarity is close between *DlOct2* and POU2f2/POU2f1 but it is also similar to POU5f1.

Future work:

The next step on this project is to perform *in situ* hybridization trials to test for the expression of *DlOct2*. The sequence that we obtained will be used to design a

probe for use in the *in situ* trials. The hypothesis is that the bud tissue will show expression of the gene because of its role as a stem cell specification gene. If no expression is found in the bud tissue, it may occur in other reproductive tissues or precursors like epicardial tissue or epithelium of the esophagus. Stem cells also play a role in regeneration of damaged tissue, and therefore the gene might be expressed in wound healing (Laird, et al., 2008).

Figure 1. A multiple sequence alignment for the POU gene family including *DlOct2*.

The sequences used in the alignment were downloaded from the NCBI database (excluding *DlOct2*). Homologs of different POU genes were used for the alignment. A representative of each of the 5 classes of POU genes and from human, *Ciona*, and Frog. The alignment shows that *DlOct2* aligns best with Oct 2.

Gene Name Percent Difference

diplo oct

Ciona pou2 KH2012:KH.C4.85.v1.A.ND1-1

gi|333360875|ref|NP_001193954.1| POU domain, class 2, transcription factor 2 isoform 1 [Homo sapiens]

sp|P14859|PO2F1_HUMAN POU domain, class 2, transcription factor 1 OS=Homo sapiens

gi|148235592|ref|NP_001090728.1| POU domain, class 3, transcription factor 4 [Xenopus tropicalis]

gi|42560248|ref|NP_002692.2| POU domain, class 5, transcription factor 1 isoform 1 [Homo sapiens]

gi|189441854|gb|AAI67702.1| pou5f1.1 protein [Xenopus tropicalis]

Ciona Pou4 KH2012:KH.C2.42.v2.A.SL1-1

gi|110347455|ref|NP_004566.2| POU domain, class 4, transcription factor 2 [Homo sapiens]

0%

40%

40%

41%

41%

46%

48%

54%

56%

1

2

3

4

5

6

7

8

Q

G

D

V

G

L

A

T

G

Q

L

S

G

-

-

-

N

D

F

S

Q

T

T

I

S

R

F

E

A

L

N

L

S

F

K

N

M

C

K

L

R

P

L

L

E

K

W

L

Q

G

D

V

G

L

A

M

G

C

L

Y

G

-

-

-

N

D

F

S

Q

T

T

I

S

R

F

E

A

L

N

L

S

F

K

N

M

I

K

L

K

P

M

L

E

R

W

L

Q

G

D

V

G

L

A

M

G

K

L

Y

G

-

-

-

N

D

F

S

Q

T

T

I

S

R

F

E

A

L

N

L

S

F

K

N

M

C

K

L

K

P

L

L

E

K

W

L

Q

A

D

V

G

L

A

L

G

T

L

Y

G

-

-

-

N

V

F

S

Q

A

T

I

C

R

F

E

A

L

Q

L

S

F

K

N

M

C

K

L

K

P

L

L

N

K

W

L

Q

A

D

V

G

L

T

L

G

V

L

F

G

-

-

-

K

V

F

S

Q

T

T

I

C

R

F

E

A

L

Q

L

S

F

K

N

M

C

K

L

R

P

L

L

Q

K

W

V

Q

A

D

V

G

Y

A

L

G

V

L

Y

G

-

-

-

K

M

F

S

Q

T

T

I

C

R

F

E

S

L

Q

L

S

F

K

N

M

C

Q

L

K

P

F

L

E

R

W

L

Q

N

D

V

G

Q

A

L

A

N

L

K

L

P

G

V

G

S

L

S

Q

S

T

I

C

R

F

E

S

L

T

L

S

H

N

N

M

V

A

L

K

P

I

L

I

T

W

L

Q

A

D

V

G

S

A

L

A

N

L

K

I

P

G

V

G

S

L

S

Q

S

T

I

C

R

F

E

S

L

T

L

S

H

N

N

M

I

A

L

K

P

I

L

Q

A

W

L

1

2

3

4

5

6

7

8

N

Y

A

N

K

M

V

K

K

A

E

G

G

E

M

V

E

V

P

M

D

-

-

-

P

S

I

T

P

G

R

K

R

K

R

T

S

I

D

Q

K

R

R

G

V

L

L

D

A

N

N

V

L

R

E

S

K

D

V

V

I

P

T

P

I

Q

L

-

-

-

D

P

I

S

I

S

R

K

R

K

R

T

S

I

D

A

K

K

R

T

T

L

N

D

A

E

T

M

S

V

D

S

S

L

P

S

-

P

N

Q

L

S

S

P

S

L

G

F

D

G

L

P

G

R

R

R

K

R

T

S

I

E

T

N

V

R

F

A

L

N

D

A

E

N

L

S

S

D

S

S

L

S

-

P

S

A

L

N

S

P

G

I

-

-

E

G

L

S

R

R

R

K

R

T

S

I

E

T

N

I

R

V

A

L

E

E

A

D

S

S

T

G

S

P

T

S

I

D

-

-

-

-

-

-

-

K

I

A

A

Q

G

R

K

R

K

R

T

S

I

E

V

S

V

K

G

V

L

E

E

A

D

N

N

E

N

L

Q

E

I

C

K

-

-

-

-

-

-

-

A

E

T

L

V

Q

A

R

K

R

-

-

-

N

R

V

R

G

N

L

V

E

A

E

N

N

D

N

L

Q

E

L

I

N

R

E

-

-

-

-

-

-

-

Q

V

I

A

Q

T

R

K

R

K

R

-

-

-

N

I

E

N

I

V

K

G

T

L

E

K

A

E

E

H

R

R

K

M

E

N

-

-

-

-

-

-

-

-

A

A

M

G

E

K

K

R

K

R

T

S

I

A

A

P

E

K

R

S

L

E

E

A

E

K

S

H

R

E

K

L

T

K

P

-

-

-

-

-

-

-

-

E

L

F

N

G

A

E

K

K

R

-

-

-

-

-

A

A

P

E

K

R

S

L

1

2

3

4

5

6

7

8

D

D

M

F

N

K

N

P

K

P

T

S

D

E

L

H

N

I

A

E

E

C

C

L

D

K

E

V

V

R

V

D

K

Q

F

E

R

N

P

K

P

S

S

E

D

L

Q

A

I

A

E

E

C

K

M

E

K

E

V

V

R

V

E

K

S

F

L

A

N

Q

K

P

T

S

E

E

I

L

L

I

A

E

Q

L

H

M

E

K

E

V

I

R

V

E

K

S

F

L

E

N

Q

K

P

T

S

E

E

I

T

M

I

A

D

Q

L

N

M

E

K

E

V

I

R

V

E

T

H

F

L

K

C

P

K

P

A

A

Q

E

I

S

S

L

A

D

S

L

Q

L

E

K

E

V

V

R

V

E

N

L

F

L

Q

C

P

K

P

T

L

Q

Q

I

S

H

I

A

Q

Q

L

G

L

E

K

D

V

V

R

V

E

S

Y

F

M

K

C

S

K

P

G

A

Q

E

M

V

Q

I

A

K

E

L

N

M

D

K

D

V

V

R

V

E

A

Y

F

A

V

Q

P

R

P

S

S

E

K

I

A

A

I

A

E

K

L

D

L

K

K

N

V

V

R

V

E

A

Y

F

A

I

Q

P

R

P

S

S

E

K

I

A

A

I

A

E

K

L

D

L

K

K

N

V

V

R

V

Figure 1.

Table 1. BLAST scores of the 5 sequences used in the alignment.

All the sequences were BLASTed against *DlOct2* using the BLAST server.

Sequence name	Max score	Total score	Query cover	E value	Ident
LOC778732 POU domain, class 2, transcription factor 3 [<i>C. intestinalis</i>]	146	146	100%	1.00E-45	61%
sp P14859 PO2F1_HUMAN POU domain, class 2, transcription factor 1 OS= <i>Homo sapiens</i>	127	127	100%	3.00E-38	56%
gi 148235592 ref NP_001090728.1 POU domain, class 3, transcription factor 4 [<i>Xenopus tropicalis</i>]	119	137	100%	3.00E-36	53%
gi 42560248 ref NP_002692.2 POU domain, class 5, transcription factor 1 isoform 1 [<i>Homo sapiens</i>]	109	124	100%	1.00E-32	51%
gi 189441854 gb AAI67702.1 pou5f1.1 protein [<i>Xenopus tropicalis</i>]	101	101	100%	1.00E-29	47%
<i>C. intestinalis</i> Pou4 KH2012:KH.C2.42.v2.A.SL1-1	73.6	73.6	100%	1.00E-19	39%
gi 110347455 ref NP_004566.2 POU domain, class 4, transcription factor 2 [<i>Homo sapiens</i>]	72.8	72.8	100%	2.00E-19	39%

Figure 2. DNA sequence of the cloned gene

```
CAGGGGGATGTGGGGCTCGCTATTGGTCAGTT
GTCTGGCAACGATTTTCAGTCAAACAACATAT
CGAGATTCGAAGCTTTAAACTTGAGTTTTAAA
AATATGTGCAAAC TGCGCCCTCTTCTTGAAAA
GTGGCTAAACTATGCAAACAAAATGGTGAAGA
AAGCAGAAGGAGGAGAAATGGTTGAGGTTCC
TATGGATCCTTCAATAACTCCTGGTAGGAAGC
GGAAGAAACGTACAAGCATTGATCAAAAGCG
GAGAGGCGTGCTTGATGACATGTTTAACAAGA
ATCCGAAACCGACCTCGGACGAGCTTCACAAT
ATAGCAGAGGAGTGTTGCTTGGATAAAGAAGT
TGTTTCGAGTC
```

Bibliography

- Aboobaker, A. A., 2011. Planarian stem cells: a simple paradigm for regeneration. *Trends in Cell Biology*, pp. 304-311.
- Anon., 2015. *UniPathway*. [Online]
Available at: www.grenoble.prabi.fr/obiwarehouse/unipathway?oscode=CIOIN
[Accessed 2015].
- Bay-Nouailhat, W., n.d. *Fauna and Flora*. [Online]
Available at: <http://www.european-marine-life.org/32/photo-diplosoma-listerianum-wb.php>
- Brown, F. D. & Swalla, B. J., 2012. Evolution and development of budding bt stem cells: Ascidian coloniality as a case study. *Developmental Biology*, pp. 151-162.
- Caceres, C. E., 1997. Dormancy in Invertebrates. *Invertebrate Biology*, 116(4), pp. 371-383.
- Dijkstra, J., Harris, L. G. & Westerman, E., 2007. Distribution and long-term temporal patterns of four invasive colonial ascidians in the Gulf of Maine. *Journal of Experimental Marine Biology and Ecology*, 342(1), pp. 61-68.
- Gold, D. A., Gates, R. D. & Jacobs, D. K., 2014. The Early Expansion and Evolutionary Dynamics of POUClass Genes. *Mol. Biol. Evol.*, 31(12), p. 3136–3147.
- Horie, T. et al., 2011. Ependymal cells of chordate larvae are stem- like cells that form the adult nervous system. *Nature*, pp. 525-529.
- Kurn, U., Rendulic, S., Tiozzo, S. & lauzon, R. L., 2011. Asexual Propagation and Regeneration in Colonial Ascidians. *Biol. Bull.*, pp. 43-61.
- Lambert, G., 2001. A Global Overview of Ascidian Introductions and Their Possible Impact on the Endemic Fauna. In: *The Biology of Ascidians*. s.l:Springer Japan, pp. 249-257.
- Martin, et al., 2016. *ANISEED*. [Online] Available at:
<http://www.aniseed.cnrs.fr/aniseed/> [Accessed 2016].
- U.S. National Library of Medicine, 2016. *National Center for Biotechnology Information*. [online] Available at: <http://www.ncbi.nlm.nih.gov/> [Accessed 2016].
- N.J.Berrill, 1935. Studies in Tunicate Development. PArt IV. Asexual Reproduction. *PHil. Trans. R. Soc. Lond. B*, pp. 327-379.

Pan, G. J., Chang, Z. Y., Scholer, H. R. & Pei, D., 2002 . Stem cell pluripotency and transcription factor Oct4. *Cell Research*, Volume 12, p. 321–329.

Rosenfeld, M. G., 1991. POU-domain transcription factors: POU-er-ful developmental regulators. *GENES & DEVELOPMENT*, Volume 5, pp. 897-907 .

Skold, H. N. et al., 2011. Patern of Cell Proliferation During Budding in the Colonial Ascidian *Diplosoma listerianum*. *Biol. Bull.*, pp. 126-136.

Stephen F. Altschul, Thomas L. Madden, Alejandro A. Schäffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", *Nucleic Acids Res.* 25:3389-3402.

Stephen F. Altschul, John C. Wootton, E. Michael Gertz, Richa Agarwala, Aleksandr Morgulis, Alejandro A. Schäffer, and Yi-Kuo Yu (2005) "Protein database searches using compositionally adjusted substitution matrices", *FEBS J.* 272:5101-5109.

Vance, T. et al., 2009. Rapid invasion and ecological interactions of *Diplosoma listerianum* in the North Sea, UK. *Marine Biodiversity Records*, pp. 1-5.