

Rochester Institute of Technology

**RIT Digital Institutional Repository**

---

Theses

---

4-27-2020

## **Transcriptomic Analysis and Developmental Neural Transcript Identification in the Brittle Star *Ophioplocus esmarki***

Alexandria Shumway  
ajs7696@rit.edu

Follow this and additional works at: <https://repository.rit.edu/theses>

---

### **Recommended Citation**

Shumway, Alexandria, "Transcriptomic Analysis and Developmental Neural Transcript Identification in the Brittle Star *Ophioplocus esmarki*" (2020). Thesis. Rochester Institute of Technology. Accessed from

This Thesis is brought to you for free and open access by the RIT Libraries. For more information, please contact [repository@rit.edu](mailto:repository@rit.edu).



Transcriptomic Analysis and Developmental  
Neural Transcript Identification in the Brittle  
Star *Ophioplocus esmarki*

By  
Alexandria Shumway

A Thesis Submitted in Partial Fulfillment of the Requirements for the  
Degree of Master of Science in Bioinformatics

Thomas H. Gosnell School of Life Sciences  
College of Science

Rochester Institute of Technology

Rochester, NY

April 27, 2020

**Rochester Institute of Technology**

**Thomas H. Gosnell School of Life Sciences**

**Bioinformatics Program**

To: Head, Thomas H. Gosnell School of Life Sciences

The undersigned state that Alexandria Juliana Shumway, a candidate for the Master of Science degree in Bioinformatics, has submitted her thesis and has satisfactorily defended it.

This completes the requirements for the Master of Science degree in Bioinformatics at Rochester Institute of Technology.

**Name**

**Date**

---

Dr. Hyla Sweet, Ph.D.

Thesis Advisor

---

---

Dr. Michael Osier, Ph.D.

Committee Member

---

---

Dr. Andre Hudson, Ph.D.

Committee Member

---

## 1 ABSTRACT

*Ophioplocus esmarki* is one species within a family of brittle stars that includes an abbreviated mode of development with a non-feeding, vitellaria larva. This development contrasts with the ancestral mode that produces a feeding, ophiopluteus larva. This project aims to complete functional annotation of the *O. esmarki* transcriptome, to provide a comparison of gene classification in both the vitellaria and juvenile stages of development, and to identify developmental neural transcripts through ortholog searches, and verify their identify through phylogenetic analysis. During my undergraduate research, Illumina sequencing was performed at the University of Rochester Genomics Center. The samples underwent RNA isolation, quality checks and were then assembled through Trinity, FastQC, and Trimmomatic tools. Functional annotation was performed using Kyoto Encyclopedia of Genes and Genomes (KEGG), and EuKaryotic Orthologous Groups (KOG), and Gene Ontology (GO) tools. The graduate portion of the research then focused on identifying neural transcripts of interest. To begin, candidate transcripts from the model sea urchin, *Strongylocentrotus purpuratus*, were identified and run against the *de novo* transcriptome using a local tblastn search to find similar sequences in the *O. esmarki* juvenile sample. The transcript identities were then confirmed with the ortholog assignment tool in eggNOG-mapper. Through phylogenetic tree analysis, the identity of the transcripts was then validated by comparing the conserved domains within other species. The significance of this research will provide a greater understanding of *O. esmarki* through both stages of development, while also determining the key neural transcripts shared with other species.

## 2 ACKNOWLEDGEMENTS

I would like to thank Jason Myers for his sequencing work for this project, and the assistance with assembling the *de novo* transcriptome. The Thomas H. Gosnell School of Life Sciences also provided a great amount of support by funding the transcriptome generation. A big thank you as well, to the System Administrator, Spencer Richman, who worked to always keep BLAST updated for me on the server. A big thank you as well to the entire Bioinformatics program for their support. I would also like to say thank you to my committee for always supporting me during this project, as well as all of my goals. Without their time, effort, and advice, this project would not be what it is today. Most of all, I would like to say thank you to Dr. Sweet for always pushing me to be my best and supporting me all of these years. I could not have asked for a better mentor to guide me through my undergraduate and graduate career. Your kind words have always helped to push me over every obstacle during this project and throughout college. Under your guidance I have learned so much and gained so many important skills. Thank you for all of your support and time to get me to where I am now.

### **3 TABLE OF CONTENTS**

#### **1 ABSTRACT 3**

#### **2 ACKNOWLEDGEMENTS 4**

#### **3 TABLE OF CONTENTS 5**

#### **4 LIST OF FIGURES 7**

#### **5 LIST OF TABLES 8**

#### **6 INTRODUCTION 9**

6.1 Brittle Star Developmental Modes, Developmental Stages, and Neural Development 9

6.2 Genes of Interest 14

6.3 Functional Annotation 15

6.4 Orthology Assignment 18

6.5 Bootstrapping 19

6.6 Project Objective 19

#### **7 METHODS 21**

7.1 RNA Preparation and Sequencing 21

7.2 Functional Annotation 22

7.2.1 KEGG

7.2.2 KOG

7.2.3 GO

7.3 Identification of Transcripts 24

7.4 Phylogenetic Tree 26

#### **8 RESULTS 26**

- 8.1 Transcriptome Assembly Statistics 26
- 8.2 Evaluation of Transcriptome Completeness 27
- 8.3 General Functional Annotation Comparisons 28
- 8.4 Specific Functional Annotation Comparisons 33
- 8.5 Neural Transcript and Ortholog Identification 36
- 8.6 Phylogenetic Tree Analysis of Neural Transcripts 39

## **9 DISCUSSION 46**

- 9.1 Transcriptome and Functional Annotation 47
- 9.2 Identification of Neural Transcripts 53
- 9.3 Future Work 56

## **10 REFERENCES**

## 4 LIST OF FIGURES

Figure 1. Phylogenetic tree of the Echinodermata phylum.

Figure 2. Ancestral and Abbreviated Modes of Development.

Figure 3. Neural Staining.

Figure 4. Assembly pipeline of *O. esmarki* vitellaria and juvenile RNA sample.

Figure 5. BUSCO results for vitellaria and juvenile samples.

Figure 6. KEGG, KOG, GO functional annotation outputs.

Figure 7. Venn diagram results of KEGG, KOG, and GO functional annotation.

Figure 8. KEGG Nervous System Venn Diagram.

Figure 9. Phylogenetic trees of Fox and Sox family members.

Figure 10. Phylogenetic trees of Arnt and Otx sequences.

Figure 11. Phylogenetic trees of full Arnt and Otx sequences.

Figure 12. Phylogenetic trees of other 15 neural transcripts.

## 5 LIST OF TABLES

Table 1. List of 28 candidate neural transcripts.

Table 2. Initial statistics from the transcriptome of vitellaria and juvenile samples.

Table 3. List of transcription factors and growth factors unique to the vitellaria and juvenile transcriptomes.

Table 4. List of neural transcripts with tblastn and eggNOG mapper results.

Supplemental Table 1. KEGG IDs unique to the vitellaria and juvenile stages.

Supplemental Table 2. KEGG cytokine and growth IDs unique to the vitellaria and juvenile stages.

Supplemental Table 3. KEGG nervous system IDs unique to the vitellaria and juvenile stages.

Supplemental Table 4. KOG IDs unique to the vitellaria and juvenile stages.

Supplemental Table 5. GO IDs unique to the vitellaria and juvenile stages.

Supplemental Table 6. List of SMART conserved protein domains in neural transcripts.

## 6 INTRODUCTION

### Brittle Star Developmental Modes, Developmental Stages, and Neural Development

With over 7,000 living species of marine organisms, the Phylum Echinodermata contains morphologically diverse, larval nervous systems while still providing many similarities in its organization. Echinoderms are developmentally unique in their five-fold radial symmetry, ability to reproduce asexually, and their utilization of a water vascular system, all which are very interesting in the field of developmental biology. Composing this phylum is five main classes: Asterozoa (sea stars), Echinozoa (sea urchins and sand dollars), Ophiurozoa (brittle stars), Crinozoa (sea lilies), and Holothurozoa (sea cucumbers) (Figure 1). In each of these five classes, an indirect form of development occurs through a larval stage. The larval stage is bilaterally symmetrical and swims in the plankton. Within the larval stages, the juvenile forms with five-fold symmetry. The mechanism of this transition from bilateral symmetry to five-fold symmetry is one of the fundamental questions in echinoderm biology. The nervous system controlling the bilateral larval stage must have a transition to a nervous system controlling the five-fold symmetry of the juvenile/adult.

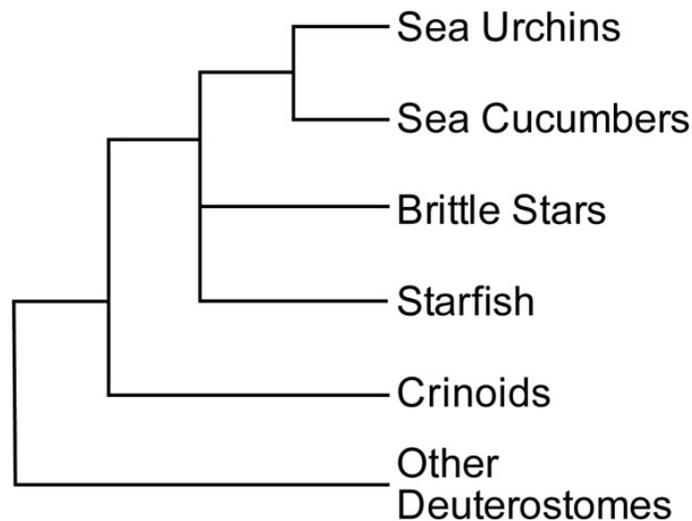


Figure 1. Phylogenetic tree of the Echinodermata phylum (Paul and Smith, 1984; Littlewood, et al. 1997; Harmon, 2005)

Two major types of development occur in echinoderms, the ancestral mode and the abbreviated mode, which is derived from the ancestral form. The development of many echinoderms features the use of an ancestral, feeding larva. Within the Class Ophiuroidea, the ophiopluteus larva takes several weeks to metamorphose into the juvenile and is believed to be the ancestral mode of development (Figure 2 A-C) (MacBride, 1907). Four pairs of arms are formed within the ophiopluteus and a single ciliary band extends throughout the larva for swimming and feeding (MacBride, 1907). The bilaterally symmetrical nervous system forms within the larva along the ciliary band and digestive system. After, the nervous system of the juvenile stage begins to develop radial and podial nerves in five-fold symmetry (Hirokawa, et al. 2008; Dupont, et al. 2009). Through evolution, many echinoderms have undergone changes from a feeding, larval stage to develop a more abbreviated form of development (Brooks and Grave, 1899). In the ophiuroids, this alternative mode of development includes a different larva known as the vitellaria. It is nonfeeding and metamorphoses into a juvenile after only a few days (Figure 2 D-F) (Sweet, et al. 2019). In contrast to the ancestral mode of development, the vitellaria larva does not have arms and it contains 3-5 ciliary bands used only for swimming (Sweet, et al. 2019).

These stages contain vast morphological differences and nervous system complexities (Hinman and Burke, 2018). The nervous system of echinoderms is made up of neurons and interconnected axons that connect mostly with the larval mouth, digestive tract, and ciliary bands

(Hinman and Burke, 2018). Echinoderms are structured with the neurons in extremities which control motor function locally, without much integration from the central nervous system (Cobb, 1987). Despite many similarities in echinoderm larvae, the neural organization does change based on the differences in feeding and locomotion (Strathmann, 1975). These different neuronal subtypes in the larval nervous systems reflect different ranges of neurotransmitters employed, thus resulting in different levels of expression (Burke, et al. 2006). Due to its diverse morphological history, the evolutionary origins of the nervous system have been notoriously difficult to understand (Hinman and Burke, 2018). Thus, the echinoderm nervous system is perhaps one of the most misunderstood and least well studied of any phyla (Hinman and Burke, 2018). Although progress has been made in recent years to further understand neurogenesis in echinoderms, sea urchins and sea stars are most often researched (Hinman and Burke, 2018). Studied less often is the Ophiuroidea class, which contains the *Ophioplocus esmarki* brittle star, the main focus of this study.

In *O. esmarki*, an abbreviated mode of development can be studied and compared to the ancestral mode of development to view the evolutionary changes in neural formation. This abbreviated mode of development is found in five other families of brittle stars, but very little is known about it. The *O. esmarki* species was chosen for this unique developmental mode and its relative accessibility. With no genome or transcriptome currently available for *O. esmarki*, an investigation into this mode of development is very interesting in developmental biology. So much is still unknown about echinoderm neurogenesis and an investigation into this less studied mode of development provides new information to major evolutionary questions.

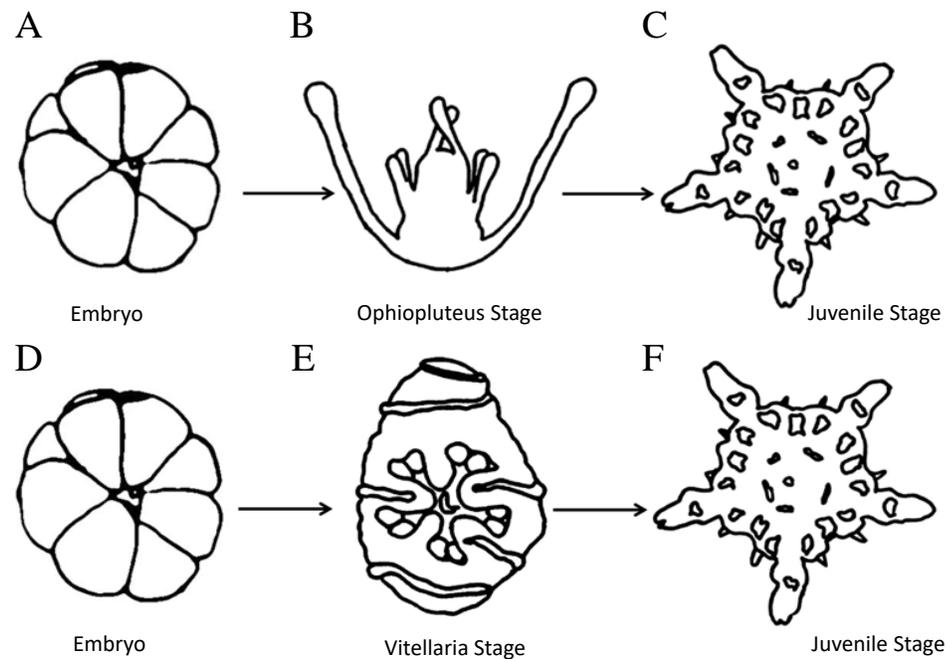


Figure 2. Ancestral and abbreviated modes of development in brittle stars. Stages A-C show the ancestral mode from the embryo (A) to the ophiopluteus larva (B), and the juvenile (C). Stages D-F show the abbreviated mode from the embryo (D) to the vitellaria larva (E), and the juvenile (F).

The nervous system of the vitellaria and the juvenile stages of *O. esmarki* is patterned differently (Figure 3) (Sweet, et al. 2019). The vitellaria stage has a larval set of neurons in bilateral symmetry which are most likely involved in swimming and sensory function (Figure 3A). In contrast, the juvenile has a nervous system in 5-fold-symmetry that controls the motor and sensory functions (Figure 3B). In the abbreviated mode of development, there are no neurons related to the larval digestive system since the vitellaria is non-feeding (Sweet, et al. 2019). However, the juvenile stages in both modes are shown to develop similarly, with the ring nerve, radial nerves, and nerves for the tube feet forming after the water vascular system (Sweet, et al. 2019). It is hypothesized that both stages of development would include some similar transcripts

because both stages have differentiated neurons. However, the larval nervous system is highly modified, while the juvenile nervous system develops in a similar way as in most other brittle stars. These findings support the idea that the larval and juvenile nervous systems evolve independently and are subject to different evolutionary pressures (Burke, 2011; Sweet, et al. 2019). Thus, expression at different levels and locations would occur to generate different neural patterns in each stage. Specifically, neural transcripts would be expressed in a bilaterally symmetrical pattern in the vitellaria, and a 5-fold pattern in the juvenile. Within the vitellaria larva, the 5-fold juvenile systems form. This includes the juvenile nervous system. Thus, we also hypothesized that the vitellaria would express genes in earlier stages of neural development, while the juvenile expresses differentiation genes representing later development.

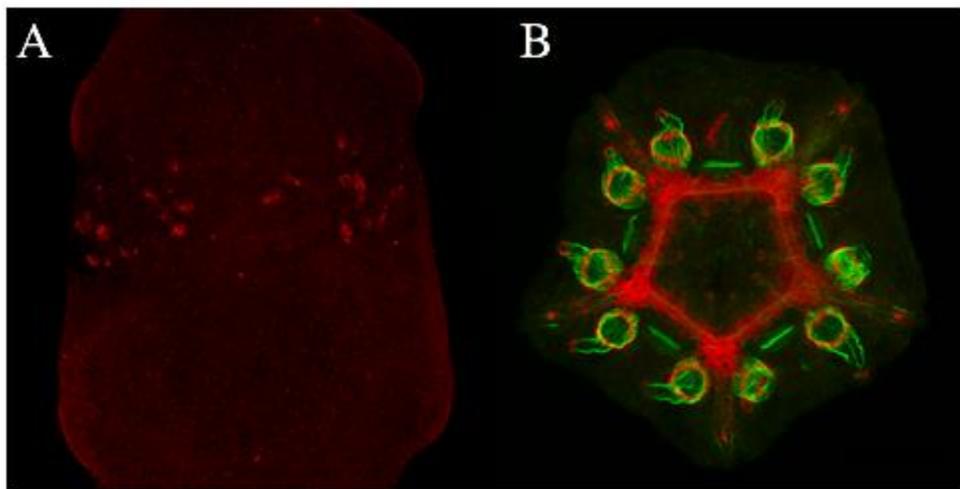


Figure 3. Neural Staining. A confocal image with synaptotagmin staining (red) in the vitellaria (A) and juvenile structures (B), exhibiting the different structures produced during development and the contrasting forms of neural development, from Sweet et al. (2019).

### Genes of Interest

To examine neural development in *O. esmarki*, candidate genes were picked based on papers focused on expression and neurogenesis in the model sea urchin *Strongylocentrotus purpuratus*. Due to the lack of reference transcriptome or genome, the sea urchin was used as a model organism. Although *S. purpuratus* may not be the closest related organism, it is the most published echinoderm that had its genome sequenced back in 2006. Due to the substantial amount of literature surrounding this organism that focused on neurogenesis and neural patterning in the larvae, similar known transcripts of interest were chosen for *O. esmarki*. The 28 neural transcripts were selected from: Howard-Ashby (2006), McClay (2018), Vokes (2007), and Burke (2006), which focus on both development and the nervous system of *S. purpuratus*. The majority of the selected neural transcripts were taken from Burke (2006), as shown in Table 1. This paper identifies several developmental neural genes from the genome of the model sea urchin and provides a list of genes expressed during specific stages of development. In the Burke paper, the expression data of the candidate regulatory genes was focused on locations in and outside the two main neurogenic regions, the embryonic apical ectoderm or ciliated band of *S. purpuratus*. The first grouping of neural transcripts occurs exclusively in the neurogenic regions of the apical ectoderm and the ciliated band. Two major genes in this category include *Achaete-scute*, which is known for its role in neurogenesis, and *Hbn*, which is required for the development of the brain in *Drosophila* embryos. The rest of the group is composed of *Ngn*, *NeuroD1*, and *Engrailed*, which are well-known pro-neural genes that are expressed later in development. The other chosen transcripts from the Burke paper came from a second group primarily located in the apical ectoderm and ciliary band, but not limited to expression in these areas. Genes in this category are critical neural factors used in development throughout the

ectoderm and used for purposes such as eye development, rapid cell division, or general nervous system in other organisms (Burke, et al. 2006).

The other chosen papers worked to validate the previously chosen transcripts and add other transcripts that play a vital role in neural development. Howard-Ashby (2006) added the transcripts *Emx* and *Hox7*, which have peak expression during early embryogenesis in the posterior region for *Hox7* and near the apical oral ectoderm for *EMX*. The Vokes (2007) paper introduced the important role *Glia* plays in the hedgehog mediated neural patterning, making this neural transcript crucial in a much later stage of development. The final paper, McClay (2018), identified six pro-neural transcription factors involved in early neurogenesis in the sea urchin *Lytechinus variegatus*. This paper mentioned many of the previously chosen transcripts, but also added a new focus on *Sip1* due to its role as a pro-neural gene involved in development (Burke, et al. 2006; Howard-Ashby, et al. 2006; Vokes, et al. 2007; McClay, et al. 2018). By identifying these transcripts, we can set a foundation for comparing neural development between developmental stages of the brittle star (vitellaria vs juvenile) and between developmental modes of the brittle star (ancestral ophiopluteus larva vs abbreviated vitellaria larva).

### Functional Annotation

Functional annotation is used to identify the orthologous genes and orthologous patterns through the use of different public databases. These orthologous genes, or orthologs, are genes in other species that have branched by speciation from a single gene of their last common ancestor. They play an influential role in newly sequenced transcriptomes since orthologs tend to have equivalent functions among different species. For this project, three types of functional annotation (KEGG, KOG and GO) were used to provide different information and categorization of gene function, allowing for varying amounts of relevancy to this project.

Kyoto Encyclopedia of Genes and Genomes (KEGG) has a pathway-based assignment of orthologs known as KEGG Orthology (KO) (Kanehisa, et al. 2000). Each KO ID represents a single orthology group that is linked to the KEGG pathway for a gene product. Molecular functions are kept in the KO database and associated with specific ortholog groups, which can then be extended to other organisms with experimental evidence. These KO IDs are manually defined in KEGG, assigning only a limited number of genes based on the available number of organisms. Once a KO ID is assigned, pathways are constructed to further interpret molecular, biological, and cellular functions. We hypothesize that the KEGG functional annotation will help discern specific transcription factors and growth factors involved in the vitellaria and juvenile stages (Kanehisa, et al. 2000).

The second form of functional annotation relies on computational identification of orthologs through EuKaryotic Orthologous Groups (KOG)/ Clusters of Orthologous Groups (COG) (Tatusov, et al. 2003). The COG database attempts to do a phylogenetic classification of proteins that are encoded in complete genomes, while the KOG database is restricted to eukaryotic genomes. The tool first detects repetitive domains through the use of RPS-BLAST and masks them. With these common and repetitive domains masked, the tool ensures a more robust classification and prevents categorizing non-orthologous proteins together. The known and predicted functions of KOGs are then classified into 26 different categories: RNA processing and modification, chromatin structure and dynamics, energy production and conversion, cell cycle control, cell division, and chromosome partitioning, amino acid transport and metabolism, nucleotide transport and metabolism, carbohydrate transport and metabolism, coenzyme transport and metabolism, lipid transport and metabolism, translation, ribosomal structure and biogenesis, transcription, replication, recombination and repair, cell wall/membrane/envelope

biogenesis, cell motility, posttranslational modification, protein turnover, and chaperones, inorganic ion transport and metabolism, secondary metabolites biosynthesis, transport and catabolism, general function, function unknown, signal transduction mechanisms, intracellular trafficking, secretion, and vesicular transport, defense mechanisms, extracellular structures, nuclear structure, and cytoskeleton. Most relevant to this project based on development is the transcription factors category of functional annotation. We hypothesize that this group could help identify the 28 neural candidates in each of the transcriptomes, to see if they are expressed (Tatusov, et al. 2003).

Gene Ontology (GO) works by describing gene products into the three main ontologies of biological processes, cellular components, and molecular functions (Ashburner, et al. 2000). The GO Consortium uses manual and automated methods to annotate genes using GO terms. The annotation must go back to another database or source of literature and provide evidence to support the newly attributed GO term. Using a limited vocabulary in the GO Consortium, any evidence found in references or databases can then be used to support the provided annotation (Ashburner, et al. 2000). The distinct categories of gene products would allow for a more specialized look into the processes and functions of these transcriptomes. We hypothesize that the GO functional annotation will help discern which function the nervous system, morphogenesis, and development pathways play in the vitellaria and juvenile stages.

We hypothesize that the different forms of functional annotation will bring varying amounts of clarity to gene function and the developmental pathways used during the vitellaria and juvenile stages of the brittle star. The KOG annotation is the least commonly used of the three forms of functional annotation and provides an overall smaller number of fairly specific IDs, while, in contrast, the GO annotation is widely used and provides a large number of broad

IDs. Of the most interest to this project are the GO categories of development and embryonic development. The functional annotation method KOG would provide information on the role of transcription factors and signal transduction. But KEGG would provide a differing view of IDs by evaluating their role in major pathways and provide a new categorization of applicable terms. The most important features for this project are the KEGG IDs involving the nervous system, development, morphogenesis, and cell differentiation. With the differing categories of functional annotation, the IDs will help to elucidate any differences in gene function for the contents of the two transcriptomes.

### Ortholog Assignment

In order to compare transcripts of interest to other species, orthologs had to be identified through tblastn and eggNOG-mapper tools. Orthologs are genes in different species that evolved from a common ancestral gene by speciation and typically retain the same function. To identify orthologs two major techniques were used. First, tblastn was run as a well-known method to directly compare protein sequences to translated nucleotides. Tblastn works by using sequence databases and then calculating its statistical significance (Altschul, et al. 1990). In contrast, the second method, eggNOG-mapper v1, is less well known and provides ortholog assignments for large sets of sequences based on pre-computed eggNOG clusters and phylogenies (Huerta-Cepas, et al. 2016). Orthologs are inferred based on pre-computed phylogenies that are associated with the location where the seed orthology was first identified using one-to-one and one-to-many orthology searches. The predicted gene names are then transferred from orthologs to the query for final assignment (Huerta-Cepas, et al. 2016). Both assignment tools were used to provide validation from different methodology using both an ortholog prediction tool and a BLAST

search-based method. Through the use of both forms of neural transcript identification a higher level of confidence can be placed on the ortholog identifications.

### Bootstrapping

Bootstrapping methods are used to establish a level of confidence in the orientation and branching of phylogenetic trees. When assembling phylogenetic trees, the construction of the phylogeny can be done through methods such as: maximum likelihood, neighbor-joining, or minimum-evolution. These methods of reconstruction work to estimate the most probable tree using estimates drawn from the data, the distance between each pair of taxa, or the smallest sum of branch length. To add an extra measure of robustness, a bootstrapping method is added to provide a higher level of confidence for a specific tree formation. Each bootstrapping simulation will choose new data at random with replacement, to provide as many phylogenetic reconstructions from the data as possible. The number of times that the same branch is selected when repeating this phylogenetic construction, will add to the confidence in the final tree formation. The percent bootstrap values are displayed for 1000 re-samplings that take place. A bootstrap value <50% would have lower confidence and would be condensed on a phylogenetic tree. Higher bootstrapping values would be displayed on the tree to provide information about branches of increased confidence (Tu et al., 2006).

### Project Objective

Currently, there is no published transcriptome or genome for the *Ophioplocus esmarki* brittle star. For our purposes, a *de novo* transcriptome is more useful than a genome for examining developmental pathways, because it represents the RNAs expressed at the stage of

collection. The transcripts also lack introns and other regulatory regions of the genes. Through the use of the transcriptome, an analysis of transcripts and gene function can be analyzed at different stages of development. For further insight to the major functions of genes, multiple forms of functional annotation are used to compare between developmental stages. The *de novo* assembly in this project is used for further identification of specific candidate neural transcripts of interest identified in the model sea urchin *S. purpuratus* (Burns, et al. 2013). With a transcriptome available for *S. purpuratus*, this model organism is used as an informative guide in the search for related candidate neural transcripts. Once identified within *O. esmarki*, the conserved domains of the transcripts of interest and other species containing these candidate orthologs such as *Drosophila melanogaster*, *Homo sapiens*, and *S. purpuratus* are used to validate the identity of *O. esmarki* transcripts through clustering on phylogenetic trees.

In summary, the purpose of this project is to provide a transcriptomic analysis of candidate neural transcripts through functional annotation and phylogenetic tree analysis in *Ophioplocus esmarki* brittle star. The hypothesis of the transcriptome analysis portion of the project is that the functional annotation will show differences between the two stages of development. Once the annotation of *O. esmarki* is complete, the specific goals for this project are to identify and validate candidate neural transcripts through the use of phylogenetic trees. The hypothesis is that the brittle star vitellaria and juvenile will express similar neural transcripts as in the model sea urchin, and that there will be differences in neural transcripts between the two developmental stages. This study should elucidate the contents of the *O. esmarki* transcriptome and give a better understanding of its shared neural transcripts with other species. The identification and validation of candidate neural transcripts will form the basis of future studies on the development of the nervous system in the vitellaria and juvenile brittle star, and

ultimately a comparison of neural development in the ancestral and abbreviated modes of development.

## **Materials and Methods**

### RNA Preparation and Sequencing

Adult brittle stars were obtained by Marinus Scientific in Long Beach, CA. Vitellaria larvae and juvenile *Ophioplocus esmarki* were collected and treated with TRIzol reagent. For each sample, ~400ul of embryos were collected and 4ml of Trizol was added. The mixtures were ground up with a pestle and lysed 20x with a pipettor. The RNA was then sent to the University of Rochester Genomics Center and isolated with the Trizol RNA Extraction protocol. The addition of 0.2mL chloroform was put in each tube and centrifuged for 15 minutes at 4°C. The samples were then precipitated by adding 0.5 mL of isopropanol, centrifuging for 10 minutes at 4°C, and removing the supernatant. Each sample had 1 mL of 75% ethanol added and was then centrifuged for another 5 minutes each at 4°C. The supernatant was removed, and the pellet was left to dry before resuspension in RNase-free water. Both samples were then incubated in a heat block for 10-15 minutes at 55-60°C.

After quality assessment of the RNA yield, the samples were prepared for paired-end Illumina HiSeq2500v4 sequencing (Figure 4), through the University of Rochester Genomics Research Center. A TruSeq mRNA-Seq Library was arranged by Dr. Jason Myers with in-line controls, Phix control, sample QC, Library QC/quantification, and pool normalization. The Illumina data were then evaluated and shortened through Trimmomatic by removing lower quality reads identified through FastQC quality assessment, as well as any Illumina-specific sequences from the file. FastQC works to provide quality control checks on the raw sequence

data, to quickly inspect any major problems. Through the basic statistics section, sequence quality can be better observed before continuing with further analysis (Bolger, et al. 2014).

The transcript sequences were then *de novo* assembled through Trinity without a reference model. Candidate coding regions were identified within the transcript sequences using TransDecoder (Grabherr, et al. 2011). This program then identified ORFs through Tophat & Cufflinks and annotated them based their similarity to sequences in the protein databases SwissProt and Pfam (Ghosh, et. al. 2016). The quality of the assembly was then evaluated with BUSCO v3 against the metazoan dataset to assess the completeness of the *de novo* transcriptome (Simão, et. al 2015).

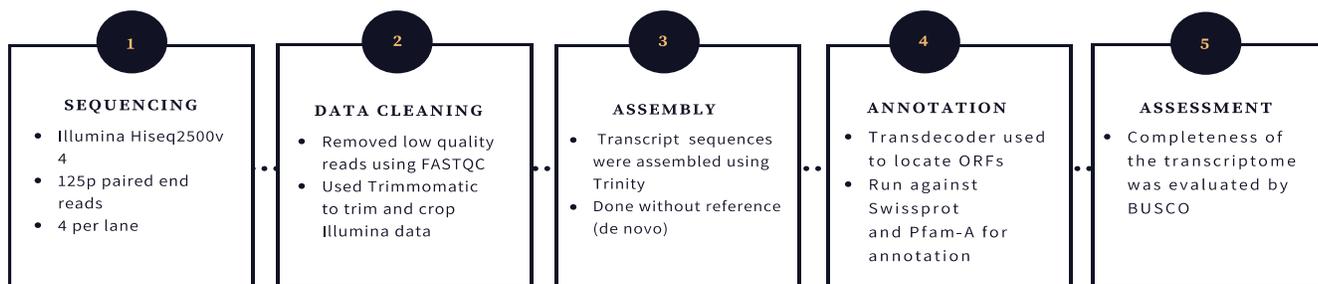


Figure 4. Assembly pipeline of *O. esmarki* vitellaria and juvenile RNA samples (Dylus et. al, 2017).

### Functional Annotation

The output of the Trinity program produced a FASTA file of 'gene' clusters based on similar sequence content. Due to the large clusters of reads, the protein output was divided into separate files of under 5000 contigs each so it could be run through annotation tools. The first form of functional annotation used was the KAAS website for KEGG annotation (Kanehisa,

2000). The FASTA file was uploaded in nucleotide format using BLAST mode. *S. purpuratus* was used as the reference organism due to its close ancestry to *O. esmarki* and its usefulness as a model organism. The bi-directional best hit (BBH) method was used to assign orthologs and the representative data set was restricted to Eukaryotes (Moriya, et al. 2007). The results were then uploaded and analyzed in an R script based on K numbers to produce a bar chart based on each factor class and its frequency.

Another type of functional annotation was performed using the WebMGA (web services for metagenomic analysis) server through the RPSBLAST 2.2.15 program on the NCBI KOG database for eukaryotic proteins (Wu, et al. 2011). The protein FASTA file was uploaded and run with an e-value of 0.001. The results were then uploaded and analyzed in an R script based on each factor class and its frequency.

eggNOG-mapper v1 was used with the DIAMOND protein database to produce GO IDs for this form of functional annotation (Huerta-Cepas, et al. 2016). The taxonomic scope automatically adjusted to perform functional transfer on all orthologs rather than just a selected clade. All orthologs were used to prioritize coverage over precision if it was restricted to one-on-one coverage. The Gene Ontology evidence used non-electronic terms to prioritize coverage, rather than using experimental terms. The output resulted in a list of functional annotation IDs and the specific GO IDs were extracted from the file and arranged from most abundant to least. The most frequently seen IDs were then displayed in two bar charts using R for both the vitellaria and juvenile samples. The proportion of the top GO IDs displayed the most frequent molecular, cellular, and protein functions of each sample.

### Identification of Transcripts

To identify neural transcripts of interest in *Ophioplocus esmarki*, known candidate transcripts were taken from sea urchins, *L. variegatus* and *S. purpuratus*. These 28 neural transcripts were selected from Howard-Ashby, et al (2006), McClay, et al. (2018), Vokes, et al. (2007), and Burke, et al. (2006) due to their focus on the genome and transcriptome of sea urchins (Table 1). The protein sequences were downloaded from Echinobase (Cary, et al. 2018) and run against the *O. esmarki* juvenile transcriptome using a local tblastn search to identify potential orthologs (Delroisse, et al. 2015). The sequences with the lowest e-values were recorded along with the percent identity. The eggNOG-mapper v2 was then used to predict orthologs for the transcriptome to validate the tblastn findings. A protein FASTA file was uploaded and the taxonomic scope was automatically adjusted by query and annotations were set to be transferred from any ortholog. The e-value was recorded from each query and suggested orthologs were provided from *S. purpuratus* and other chordates that were then downloaded. The FASTA files from other species were then run against *S. purpuratus* using blastp to validate the match.

Table 1. List of 28 candidate neural transcripts in *S. purpuratus* and *L. variegatus* compiled from Howard-Ashby, et al (2006), McClay, et al. (2018), Vokes, et al. (2007), and Burke, et al. (2006).

Gene ID	Gene Name	Transcription Factor Family	References
SPU_028148	Sp-Ac-Sc (achaete-scute)	bHLH	Burke, et al. (2006)
SPU_023177	Sp-Hbn (homeobrain)	hbox-paired	Burke, et al. (2006); Howard-Ashby, et al (2006)
SPU_007147	Sp-Ngn (neurogenin)	bHLH	Burke, et al. (2006); McClay, et al. (2018)
SPU_024918	Sp-NeuroD1 (neuroD)	bHLH	Burke, et al. (2006)
SPU_020975	Sp-Engrailed	hbox	Burke, et al. (2006); Howard-Ashby, et al (2006)
SPU_014289	Sp-Rx (retinal anterior hbx)	hbox-paired	Burke, et al. (2006); Howard-Ashby, et al (2006)
SPU_022820	Sp-SoxB1	HMG	Burke, et al. (2006); McClay, et al. (2018)
SPU_025113	Sp-SoxB2	HMG	Burke, et al. (2006)
SPU_004217	Sp-SoxD	HMG	Burke, et al. (2006)
SPU_010424	Sp-Otx (orthodenticle)	hbox	Burke, et al. (2006); Howard-Ashby, et al (2006)
SPU_021608	Sp-Hes (hairy-related)	bHLH	Burke, et al. (2006)
SPU_027969	Sp-FoxJ1	forkhead	Burke, et al. (2006)
SPU_025590	Sp-FoxM	forkhead	Burke, et al. (2006)
SPU_018908	Sp-Six3 (sine oculis)	hbox-atypical	Burke, et al. (2006); Howard-Ashby, et al (2006)
SPU_002603	Sp-SoxC (Sox4/11/22/24)	HMG	Burke, et al. (2006); McClay, et al. (2018)
SPU_002592	Sp-Emx (empty spiracles)	Hox	Burke, et al. (2006); Howard-Ashby, et al (2006)
SPU_016449	Sp-Hnf6 (oncut2)	hbox	Burke, et al. (2006)
SPU_019290	Sp-Otp (orthopedia)	hbox	Burke, et al. (2006); Howard-Ashby, et al (2006)
SPU_002815	Sp-DLX (distal-less)	hbox-other	Burke, et al. (2006); Howard-Ashby, et al (2006)
SPU_008936	Sp-Tlx1 (tail-less)	nuclear receptor	Burke, et al. (2006)
SPU_014418	Sp-FoxD	forkhead	Burke, et al. (2006)
SPU_000129	Sp-Arnt	aryl hydrocarbon receptor	Burke, et al. (2006)
SPU_023941	Sp-Myt1 (myelin TF1)	zinc finger	Burke, et al. (2006)
SPU_027603	Sp-Gmfb (Glia)	zinc finger	Vokes, et al. (2007)
SPU_002634	Sp-Hox7	hbox	Howard-Ashby, et al (2006)
SPU_028583	Sp-Zic2 (odd-paired)	zinc finger	Burke, et al. (2006)
SPU_007599	Sip-GlassL	zinc finger	Burke, et al. (2006)
SPU_026620	Sp-Sip1	zinc finger	McClay, et al. (2018)

## Phylogenetic Trees

To verify the identity of the transcripts of interest, the sequences were compared to conserved domains in other species. Orthologs were gathered from NCBI for *Drosophila melanogaster*, *Homo sapiens*, *S. purpuratus*, and other outgroups when necessary. Each FASTA file was run through both SMART (Letunic & Bork, 2017) and InterPro (Hunter, et al 2009) to identify commonly conserved domains across species. The sequences of the conserved domains were used to generate phylogenetic trees through MEGA7 (Kumar, et al. 2016). Protein alignments were first built and aligned using MUSCLE. A gap penalty was set to -2.9 and gap extend was set to 0. The alignment was then saved and uploaded to construct a new Maximum Likelihood Tree using the neighbor-joining method with 1000 bootstrap replications. A Jones-Taylor-Thornton (JTT) model was used for substitutions with a Nearest-Neighbor-Interchange (NNI). The output was exported as a PDF and colored labels were added to distinguish different species. For the larger families of genes that encode transcription factors, combined phylogenetic trees were created to better identify the clustering of transcripts.

## **Results**

### Transcriptome Assembly Statistics

To continue with confidence, the *O. esmarki* transcriptomes were evaluated for completeness through preliminary statistics (Table 2). The N50 scores produced were high enough to give ample confidence in the assembly contiguity. The similar statistics for genes, percent GC, and contig length were also consistent with two samples from the same species. The ‘gene’ cluster number listed in row one is based on the grouping of Trinity transcripts based on shared sequence content. Between the two samples, there is only a slight variation in the total

number of 'genes' and transcripts, with the vitellaria sample showing a smaller number in both categories. The percent GC is approximately 39% for both samples, but the N50 does vary. The N50 of 671 bases in the juvenile sample means that 50% of the genome/transcript can be described using contigs greater than or equal to 671 bases. This does not necessarily mean that half of the transcripts are of base length 671 or greater. The N50 score provides a summary of assembly contiguity, in which a higher number would represent how few contigs of large length are needed to cover the transcriptome. For the vitellaria sample, the contig N50 and average contig length are higher than the juvenile. The N50 for the vitellaria is 951 bases and the average contig is ~618 bases. While this preliminary analysis does give a promisingly high N50 score for the transcriptomes, secondary quality analysis was assessed with Benchmarking Universal Single-Copy Orthologs (BUSCO) to provide full confidence in the completeness of the *de novo* assembly.

Table 2. Initial statistics from the transcriptome of juvenile and vitellaria samples.

	Juvenile	Vitellaria
Total Trinity 'genes':	375684	317883
Total Trinity transcripts:	650202	579917
Percent GC:	39.70	39.84
Contig N50:	671	951
Average contig length:	525.29	618.86

#### Evaluation of Transcriptome Completeness

To assess imperative secondary metabolites, BUSCO (Figure 5) was run against the metazoan dataset as a secondary method of validating transcriptome quality. With both samples containing over 80% BUSCO completeness scores, we can conclude that the transcriptomes have

a high-quality assembly. The BUSCO analysis for the vitellaria sample had 84.8% completeness and 96.8% for the juvenile sample (Figure 5). The total completeness score was composed of the complete single-copy and complete duplicated samples when run against the metazoan dataset for important metabolites. Both samples had less than 2% of the gene content missing. However, the vitellaria did have a much higher fragmented section of the total gene content at 13.4%, as compared to only 2.4% in the juvenile. The results supported the prior N50 results, concluding that the transcriptome assembly was of good quality for both samples, but the higher quality transcriptome (juvenile) was used for further neural identification.

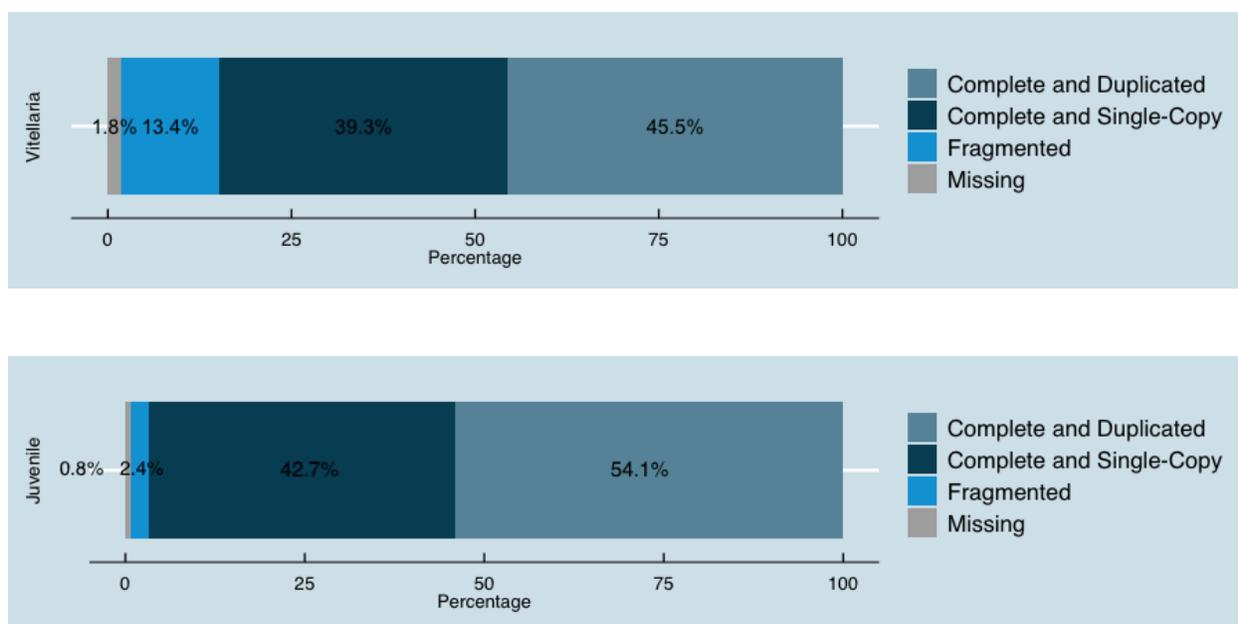


Figure 5. BUSCO results for the juvenile and vitellaria samples.

### General Functional Annotation Comparisons

Using three types of functional annotation, the comparison between the vitellaria and the juvenile stages yielded very similar outcomes (Figure 6). For each functional annotation tool, the means of both samples were not seen to have statistically significant differences in the overall ID

comparison. The results are categorized in major factor classes and represented in relative abundance or percentage of the total number of IDs given for the specific type of functional annotation. For the KEGG functional annotation, both samples appear to have similar percentages for each factor class (Figure 6A-B). The only minimal difference is in the percentage of KEGG category J (signal transduction) that appears to be lower in the juvenile sample at just under 15%. Of the previously specified categories of interest, signal transduction (with a subsection of cytokines and growth factors) was listed as potentially informative. This is consistent with the hypothesis that the KEGG functional annotation will help discern differences in growth factors in the vitellaria and juvenile stages (see section on Specific Functional Annotation Analysis). The other primary area of interest, KEGG category F (transcription) did not have any noticeable difference between the two stages, but by looking deeper into the transcription category, specific neural transcripts may also be identified to further address the hypothesis that there are different transcription factors expressed at the different stages (see section on Specific Functional Annotation Analysis).

For the KOG annotation, one difference lies in the abundance of KOG category J (Translation, ribosomal structure and biogenesis). In the juvenile sample (Figure 6C), the abundance of KOG IDs is almost at 0.05%, while the vitellaria sample (Figure 6D), shows almost half that. A slight difference is also apparent in KOG category C (Energy production and conversion), with elevated abundance in the juvenile sample. The apparent differences in the KOG categories between the vitellaria and juvenile samples are consistent with the idea that the different stages have different subsets of gene expression that each transcriptome would provide different levels of abundance for functional annotation IDs due to its different developmental

stage. However, the apparent differences in the categories of translation and energy production were unexpected results from the original focus on transcription and signal transduction.

For the Gene Ontology comparison, the results for both the juvenile and vitellaria samples appear to be fairly proportional (Figure 6E-F). However, a difference lies in the percent of IDs in the GO category G (morphogenesis) for each sample. As previously hypothesized, morphogenesis remained a major category of interest in the comparison of both transcriptomes but GO category B (development) stayed proportional in each sample. The nervous system was also not a prevalent category to provide an informative comparison. Due to the large number of factor classes present in GO, any categories below a certain percentage of IDs were removed from the chart, including the nervous system. For GO category G (morphogenesis), there is ~3% of the total IDs in the juvenile sample (Figure 6E), while in the vitellaria (Figure 6F) it remains at just above 2% of the total. The GO category N (protein metabolism) for the vitellaria appears to be lower than the juvenile.

To further investigate the overall differences between the juvenile and the vitellaria samples, a two-sample t-test was performed. A null hypothesis was established that the two samples would have no difference between the means. In order to accept this hypothesis with 95% confidence, the p-values for each type of functional annotation would have to be less than 0.05. However, the p-values for each type of functional annotation were 0.991 for KEGG, 0.998 for KOG, and 0.930 for GO. With all of these p-values ~0.9, we were unable to reject the null hypothesis. Therefore, the slight differences shown between each of the samples are not statistically significant enough to provide a difference in sample means.

The preliminary comparison of KEGG, KOG, and GO IDs between the vitellaria and juvenile samples showed some potential differences that could be explored further. For example, the KEGG category J (signal transduction) is lower in the juvenile sample than in the vitellaria. The KOG category J (Translation, ribosomal structure and biogenesis) and category C (Energy production and conversion) are also higher in the juvenile sample. The GO category G (morphogenesis) and N (protein metabolism) are also higher in the juvenile than in vitellaria.

Overall, with the comparison of functional annotation methods, the relative abundance of each ID category remains quite similar for each sample. Initially each functional annotation method had different categories of interest to this project, with KEGG providing categories for the nervous system, development, and morphogenesis, KOG providing information on transcription factors and signal transduction, and GO IDs involving development. To get a better understanding of the differences in the transcriptomes between the two developmental stages for these categories of interest, the functional annotation methods were compared using Venn diagrams that were generated in R (Figure 7).

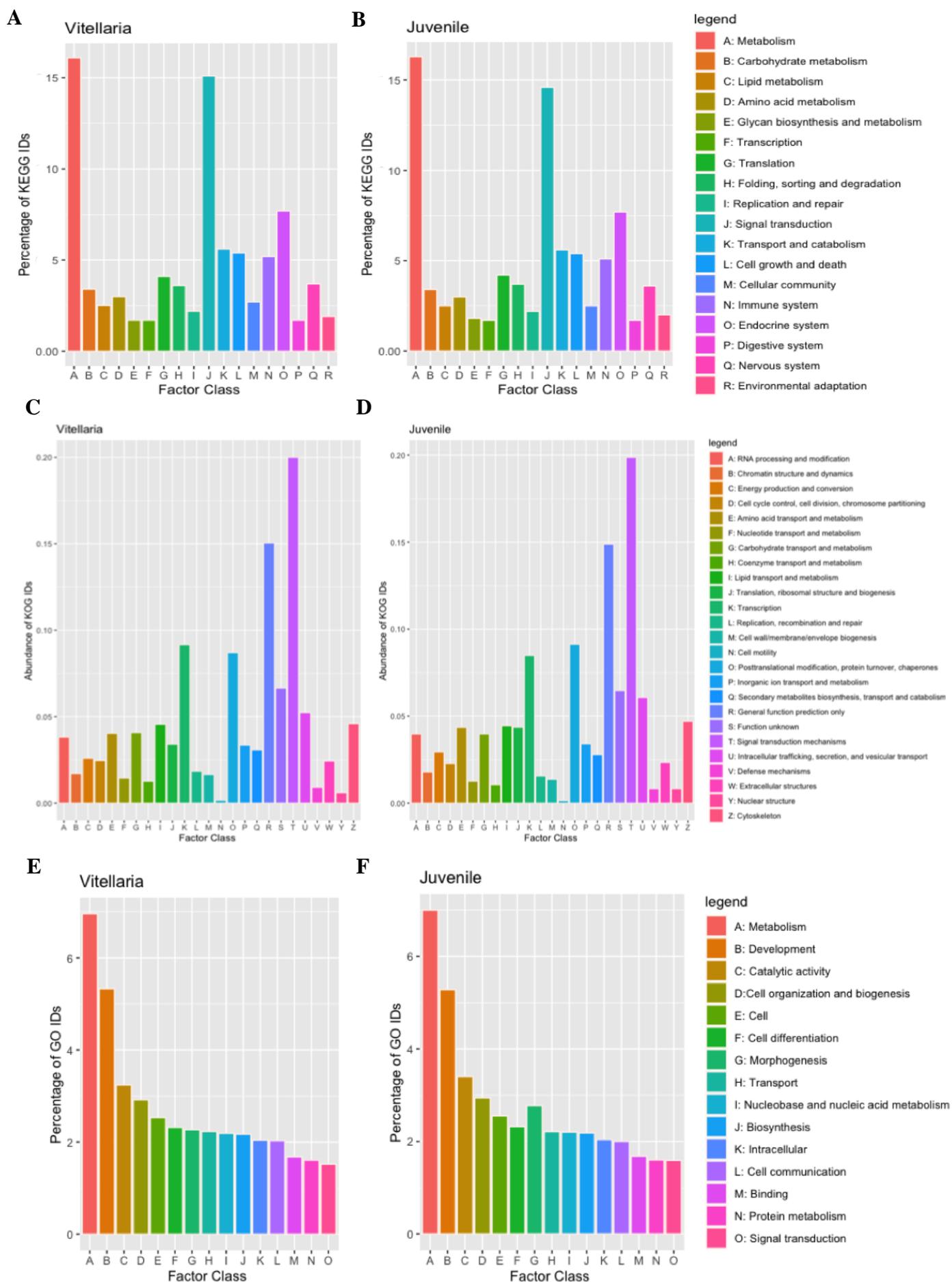


Figure 6. KEGG (A-B), KOG(C-D), GO (E-F) functional annotation outputs in each row. The left column is composed of the vitellaria samples, and the right has the juvenile samples.

### Specific Functional Annotation Comparisons

Due to the proportional results when comparing the general functional annotation methods between the two samples, the IDs unique to each developmental and neural stage were analyzed. KEGG was found to contain unique IDs for transcription and growth factors, KOG contained unique IDs for transcription factors of interest, and GO contained a large number of broad IDs. Venn diagrams were composed to show the overlapping functional annotation in both the vitellaria and juvenile samples (Figure 7). The purple section of the Venn diagrams shows the IDs that each form of functional annotation has in common with both stages of development. The sections that are light blue represent the unique IDs to the juvenile stage of development and the pink represents unique IDs for the vitellaria.

The first form of functional annotation in Figure 7A was KEGG annotation. A total of 4,837 KEGG IDs were associated with both the vitellaria and juvenile modes of development. KEGG Orthology contains a total of 23,318 IDs which means the total amount identified represents 20% of the total IDs. The vitellaria stage had 419 unique KEGG IDs associated only with it, while the juvenile had 363 unique KEGG IDs. Some IDs of interest in the vitellaria stage, include *Zic2*, *Myt1*, *Glia*, *Hox7*, *Glass*, and *DLX2*, which are all particularly of interest to this project. Also uniquely shown during this stage of development were *Six4*, *FoxI*, *Gli3*, and *ArntL*, which are all family members of the 28 candidate neural transcripts as previously tested. Whereas, unique to the juvenile stage was *SoxC*, which was also from the candidate neural transcript list for this project. But shared between the two stages were several candidate transcripts of interest, such as *Achaete-scute*, *Engrailed*, *Emx*, *Zic2*, *Ngn*, *Arnt*, *Six3*, *FoxD*,

FoxJ1, and SoxD. Other developmentally relevant transcripts are included in the table for transcription factors (Supplementary Table 1) and growth factors (Supplementary Table 2).

In Figure 7B, the results of KOG functional annotation produced 130 IDs unique to the vitellaria and 41 IDs unique to the juvenile. In common, the two shared 4395 identified KOG IDs. The total number of KOG IDs listed in this form of annotation is 4395, meaning 100% of the IDs were represented in this search. The unique KOG IDs are shown in Supplementary Table 3. Developmentally relevant transcripts found uniquely in the vitellaria stage include the transcription factors BSH, Caudal, and DLX (each of which contains a HOX domain) and Nanos (which is involved in the specification of the germ line). However, the juvenile stage was not shown to have any unique transcripts with major relevancy to this project.

In Figure 7C, the GO terms identified were 14,489, with fairly equal numbers of terms in each stage of development. The vitellaria had 1,219 unique GO terms associated with it, and the juvenile had 1,493. GO provides the largest number of IDs available, with 70,344 total, and the 14,489 identified in these samples make up 20% of the total consortium. The GO ID categories encompassed very broad functions and several taxa unrelated to brittle stars. Overall, the results proved to be less meaningful than the functional annotation methods, KEGG and KOG. The GO IDs that were shown to be developmental unique are shown in Supplementary Table 5.

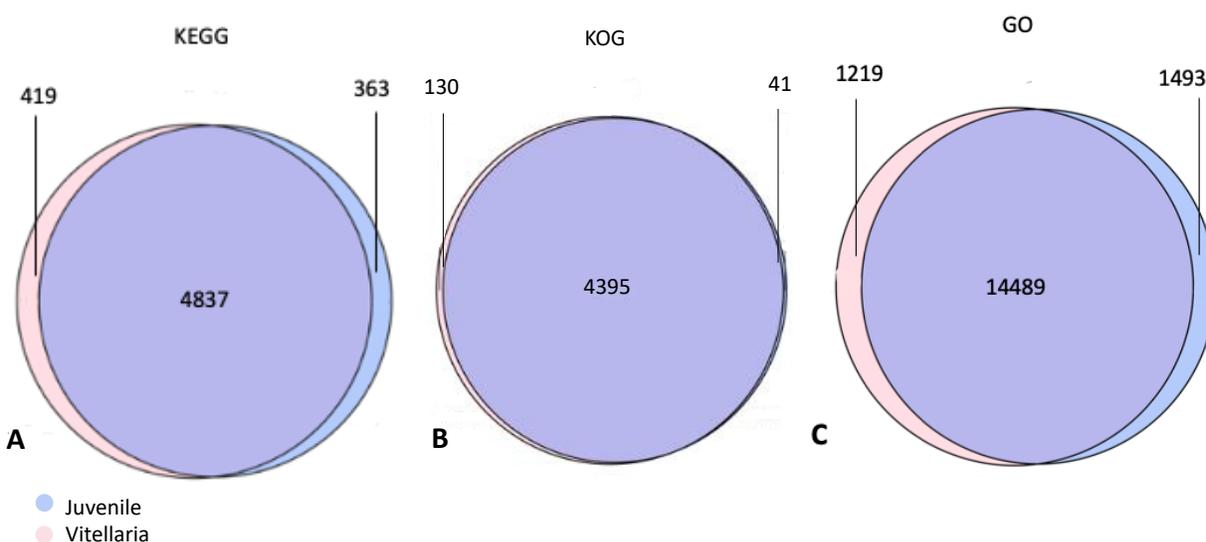


Figure 7. Venn diagram results of KEGG (A), KOG (B), and GO (C) functional annotation.

A major category of interest was the KEGG subsection of the nervous system, which allowed for a further categorical analysis into the unique neural IDs for each transcriptome (Figure 8). Upon initial inspection, the vitellaria was shown to have ten unique IDs and the juvenile had three unique IDs, but 162 Nervous System KEGG IDs in common. This supports the hypothesis that both stages of development would include some similar transcripts because both stages have differentiated neurons. After looking at the different nervous system KEGG IDs (Supplementary Table 3), one transcription factor shown to be unique to the vitellaria was *ArntL*, which is found in the same family as *Arnt*. The function of this transcript is primarily involved in brain and muscle development, implying increased activity during this stage of development. However, the three unique juvenile KEGG IDs only showed basic metabolic pathway function and were not as informative for this project. The results supported the hypotheses about the

presence of overlapping transcripts and unique transcripts that would occur when comparing the two stages of development in the same organism.

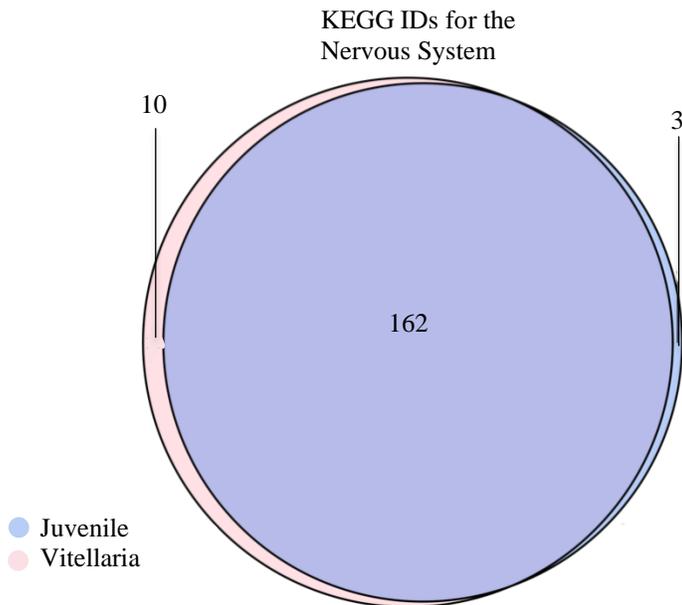


Figure 8. Venn diagram of KEGG Nervous System.

### Neural Transcript and Ortholog Identification

The identification of the 28 neural transcripts was first obtained through a *tblastn* search with the juvenile *O. esmarki* transcriptome and then validated through eggNOG-mapper (Table 3). All transcripts were identified through *tblastn* and 24 were further supported through identification with eggNOG-mapper. The *S. purpuratus* IDs of candidate neural transcripts were taken from Echinobase and run against the juvenile transcriptome to find the best possible match with the lowest e-value (Cary, et al. 2018). The results were then validated through the ortholog finder in eggNOG mapper, which produced the closest orthology in *O. esmarki* or other related species. The corresponding e-value was also produced from the eggNOG mapper results. The

alternative species provided have highlighted IDs such as *Saccoglossus kowalevskii* (Acorn worm) in blue, *Lepisosteus oculatus* (Spotted gar) in yellow, and *Astyanax mexicanus* (Cave fish) in pink. In these cases, the neural transcripts of interest matched more closely with orthologs from alternative species than with the sea urchin, but still had the same identified neural transcript. The alternative species were checked through BLAST to confirm their positive identification of the attributed neural transcript. Of the 28 neural transcripts, four had unexpected results from eggNOG mapper. In Table 3, gray boxes with the names of DLX, Hbn, Hox7, and Glass were all shown to be in conflict with the tblastn identifications. Starting with DLX, eggNOG mapper showed a match with the transcription factor NK2-3/5 at that location of the *O. esmarki* transcriptome. Hbn and Hox7 were both also incorrectly identified as different *S. purpuratus* genes, *Aristaless* and *Hox8* respectively. Lastly, Glass was never identified through eggNOG mapper for any section of the transcriptome. Because of the uncertainty in identity, these four transcripts were removed from further analysis. However, the identities of the remaining transcripts were successfully validated through the findings of eggNOG mapper.

Table 3. List of neural transcripts with tblastn and eggNOG mapper results. The neural transcripts in gray boxes were removed from further analysis due to uncertainty in transcript identity. The alternative species provided have highlighted IDs such as *Saccoglossus kowalevskii* (Acorn worm) in blue, *Lepisosteus oculatus* (Spotted gar) in yellow, and *Astyanax mexicanus* (Cave fish) in pink.

	Name	Echinobase ID	tblastn	e-value	eggNOG mapper	e-value
1	Achaete-scute	SPU_028148	DN98138_c0_g10_i1	3.31E-38	NP_001158485	5.10E-45
2	Hbn	SPU_023177	DN96584_c0_g3_i6	3.97E-26	SPU_025302 (Aristaless)	3.50E-39
3	Ngn	SPU_007147	DN111089_c1_g1_i3	7.12E-44	DN111089_c1_g1_i3	3.20E-27
4	NeuroD1	SPU_024918	DN82457_c0_g1_i1	3.13E-60	SPU_024918	2.20E-42
5	Engrailed	SPU_020975	DN111079_c3_g2_i1	7.69E-48	DN111079_c3_g2_i1.p1	2.00E-31
6	Rx	SPU_014289	DN31536_c0_g1_i1	3.13E-38	NP_001158375	7.70E-32
7	SoxB1	SPU_022820	DN87323_c1_g3_i2	1.00E-81	SPU_022820	4.90E-89
8	SoxB2	SPU_025113	DN91268_c4_g3_i1	9.00E-91	SPU_025113	9.80E-72
9	SoxD	SPU_004217	DN107517_c1_g1_i1	2.30E-122	SPU_004217	7.10E-117
10	Otx	SPU_010424	DN113378_c5_g1_i1	3.94E-55	ENSAMXP00000021108	1.90E-35
11	Hes	SPU_021608	DN86689_c0_g4_i2	1.03E-47	DN86689_c0_g4_i2	7.60E-38
12	FoxJ1	SPU_027969	DN73901_c0_g1_i1	1.00E-69	NP_001158438.1	2.30E-48
13	FoxM	SPU_025590	DN107113_c4_g1_i1	3.00E-58	SPU_025590	4.10E-61
14	Six3	SPU_018908	DN95223_c2_g6_i1	1.70E-114	NP_001158378.1	2.40E-105
15	SoxC	SPU_002603	DN94804_c1_g1_i1	3.00E-87	SPU_002603	1.30E-82
16	Emx	SPU_002592	DN101615_c0_g1_i2	1.09E-66	DN101615_c0_g1_i2	4.40E-65
17	Onecut	SPU_016449	DN94224_c6_g1_i1	9.10E-140	DN94224_c6_g1_i1.p1	1.60E-114
18	Otp	SPU_019290	DN92812_c3_g2_i1	2.29E-79	NP_001158374.1	1.80E-75
19	DLX	SPU_002815	DN34428_c0_g1_i2	1.27E-17	XP_006815459.1 (NK2-3/5)	3.90E-24
20	Tailless	SPU_008936	DN109743_c1_g1_i2	0	DN109743_c1_g1_i2	8.70E-143
21	FoxD	SPU_014418	DN94750_c8_g1_i3	1.00E-70	SPU_027648	9.20E-59
22	Arnt	SPU_000129	DN109334_c3_g1_i4	0	DN109334_c3_g1_i4	4.60E-198
23	Myt1	SPU_023941	DN109244_c1_g1_i5	3.29E-100	DN109244_c1_g1_i5	1.00E-10
24	Glia	SPU_027603	DN99405_c2_g1_i6	3.00E-28	ENSLOCP00000017971	3.70E-41
25	Hox7	SPU_002634	DN111079_c3_g1_i1	5.00E-28	SPU_021309 (HOX8)	1.90E-45
26	Zic2	SPU_028583	DN110993_c1_g1_i2	2.23E-43	NP_001158430	2.60E-108
27	Glass	SPU_007599	DN97386_c3_g1_i2	5.00E-35	NO MATCH	N/A
28	Sip1	SPU_026620	DN101938_c6_g1_i1	1.00E-32	SPU_026620	3.50E-30

### Phylogenetic Tree Analysis of Neural Transcripts:

To further validate the identities of the remaining 24 neural transcripts of interest, phylogenetic trees of conserved domains were made in MEGA7 (Figures 9, 10, 11). Of the 24 identified neural transcripts of interest, 22 phylogenetic trees of conserved protein domains supported branching of the echinoderm orthologs. Once the full-length sequences were used for the remaining two trees, their branching was supported as well. The formation of all 24 trees supported the confidence in transcript identification by clustering most closely to the model sea urchin.

The sequences used are shown in Supplementary Table 6 with conserved domain regions predicted using SMART. Because multiple Fox and Sox family members are within our transcript list (Tables 1, 2), combined trees of the different family members were produced (Figure 9). The first tree (Figure 9A) shows a lower clustering of FoxD, with 95% confidence in the clustering of all four FoxD sequences among the species. FoxD orthologs from *O. esmarki* and *S. purpuratus* (both echinoderms) are also clustered with a 93% bootstrapping confidence level. Above this cluster is FoxJ1 and FoxM, which starts at a 30% confidence due to the similarity of the sequences. FoxM orthologs from *O. esmarki* and *S. purpuratus* are clustered with an 88% bootstrapping confidence level and FoxJ1 also has a 93% confidence between the same echinoderm orthologs. The FoxM sequence of *D. melanogaster* acted as an outgroup, leaving the rest of the FoxM sequences to cluster with an 89% confidence rate.

Figure 9B shows the clustering of the Sox family transcripts, SoxB1, SoxB2, SoxD, and SoxC. The upper clustering of the tree shows a 99% confidence level with the eight ortholog sequences of SoxB1 and SoxB2 clustering by species due to their similarity. The SoxB1 sequence of *D. melanogaster* acted as an outgroup from the rest of the cluster. Below this is the

cluster of SoxC at a 63% confidence level with the four different species of the SoxC sequences aligning together. The individual alignments in this cluster show an alignment of *H. sapiens* and *S. purpuratus* with only a 32% confidence and *O. esmarki* appearing as the next branch over with a 42% confidence. Finally, the 0.00 branch lengths and identical conserved regions of *O. esmarki* and *S. purpuratus* gave the SoxD orthologs a 99% confidence level.

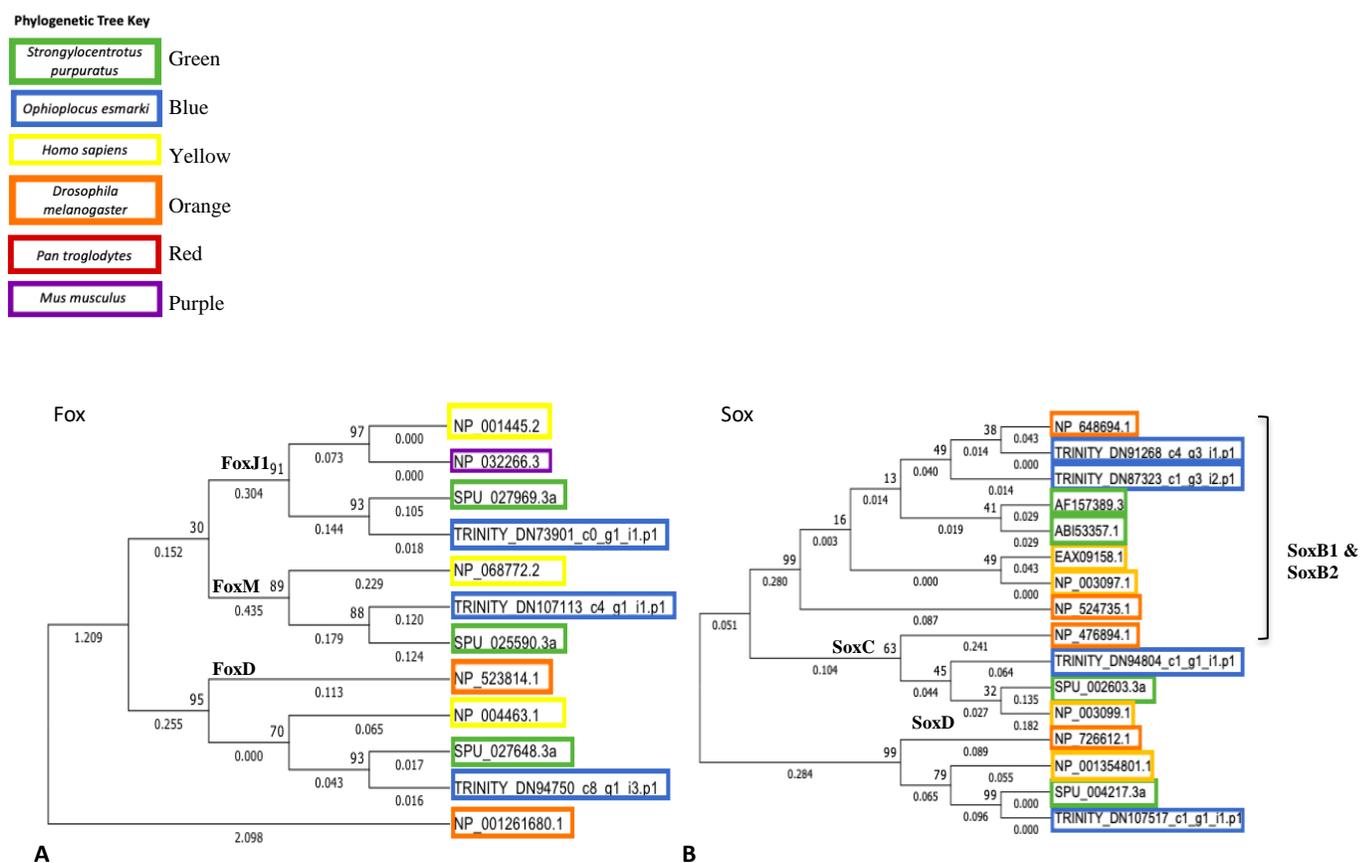


Figure 9. Phylogenetic trees of the combined Fox (Figure 9A) and Sox (Figure 9B) neural transcripts. Bootstrap values are shown at each branch. Branch lengths are shown along each lineage.

Of the 24 alignments, only two, Arnt and Otx, had a confidence level below 50% (Figure 10). These trees showed *S. purpuratus* clustering with *O. esmarki* only 31% of the time in Arnt (Figure 10A) and 47% of the time with Otx (Figure 10B). Of the neural transcripts of interest,

these two are the only ones with a bootstrapping confidence under 60%, giving lower confidence to the neural transcript identification. The *D. melanogaster* sequences typically become the outgroup in the alignments; however, in Otx the *H. sapiens* sequence becomes the outgroup. With both Arnt and Otx, the branch lengths are very small with most at 0.000 or very close to it.

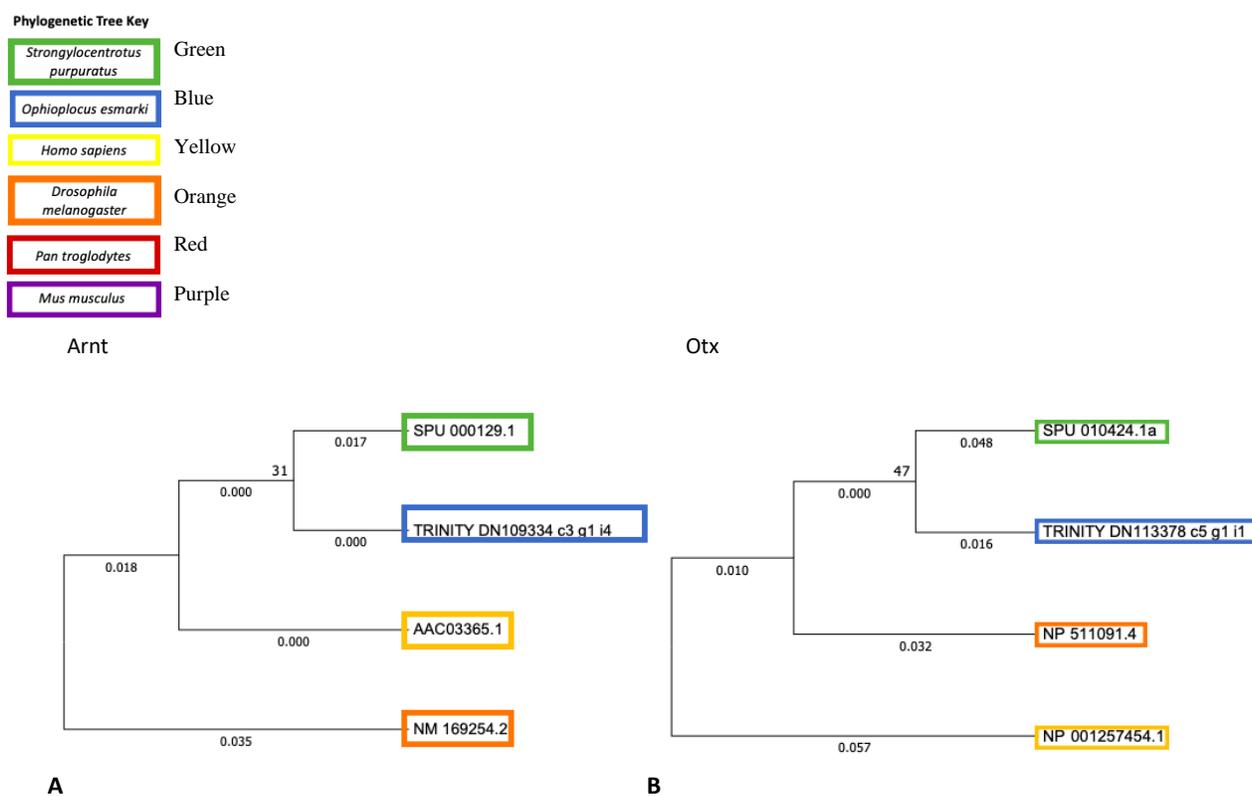


Figure 10. Phylogenetic trees of the Arnt and Otx Sequences. Bootstrap values are shown at each branch. Branch lengths are shown along each lineage.

To further examine the closeness of the branches in Otx and Arnt, the alignment was re-constructed to see if a full alignment could produce different results. Using the whole transcript sequence, the confidence levels rose to much higher bootstrapping numbers. These trees showed *S. purpuratus* clustering with *O. esmarki* 96% of the time in Arnt (Figure 11A) and 97% of the time with Otx (Figure 11B). The *D. melanogaster* sequences served as the outgroup in the alignments. By observing the closeness of the branch lengths and repeating the alignments with the full sequences, the bootstrapping results gave a high enough confidence level to support a successful transcript identification.

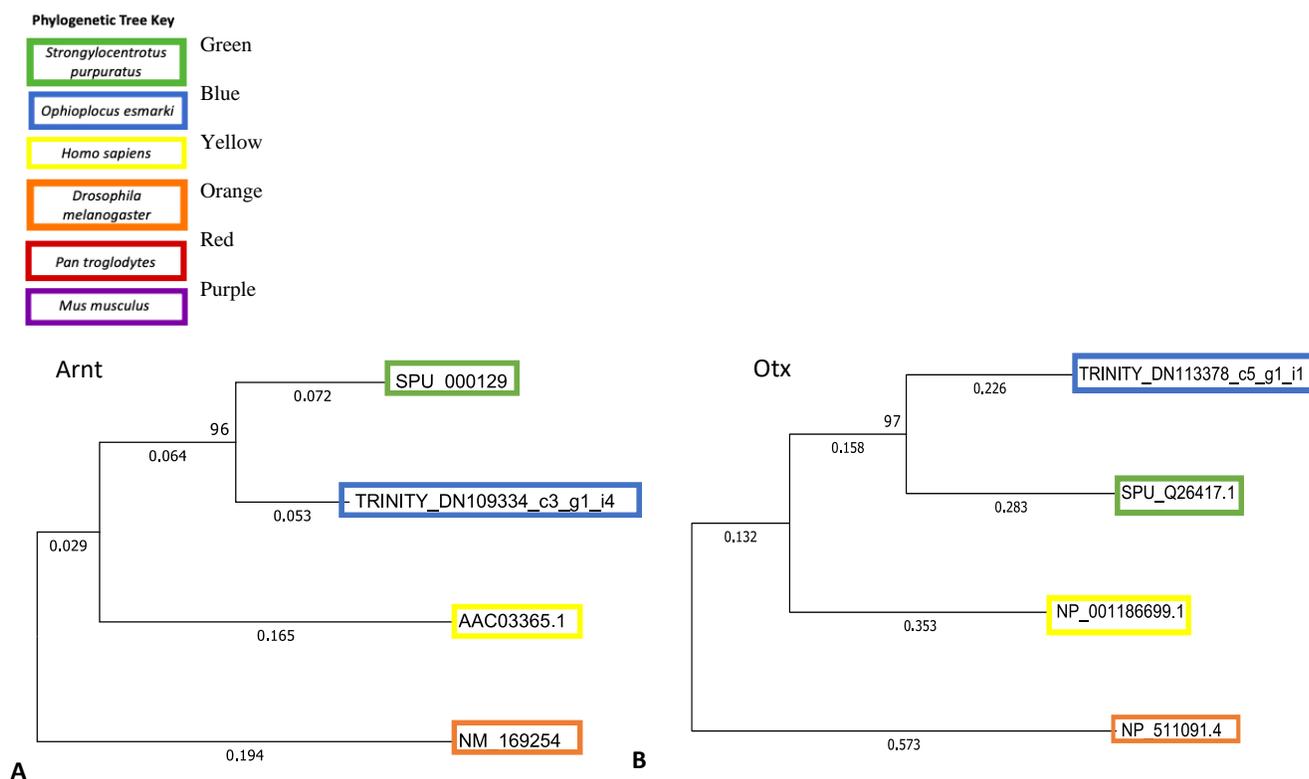


Figure 11. Phylogenetic trees of the full Arnt and Otx neural transcripts. Bootstrap values are shown at each branch. Branch lengths are shown along each lineage.

The phylogenetic trees of the remaining 15 neural transcripts were combined (Figure 12) due to their bootstrapping confidence levels being above 60% and the general pattern of results with the ortholog clustering of the echinoderms. Both *Engrailed* and *Hes* have a 99% confidence level with *S. purpuratus* clustering with *O. esmarki* through bootstrapping, and *D. melanogaster* as the outgroup. *Tailless* also follows this structure, but with a closely followed 98% confidence level for the grouping of *S. purpuratus* and *O. esmarki*. *Emx* and *Six3* follow the same structure, with both having *D. melanogaster* as the outgroup and showing a 96% confidence level with *S. purpuratus* clustering with *O. esmarki* through bootstrapping. *Sip1* and *Otp* also share a 91% confidence interval for *S. purpuratus* clustering with *O. esmarki*, but the outgroup for *Otp* differs from the other trees. While *Sip1* shows a similar structure of the human sequence being a branch over from the sister taxa of the *S. purpuratus* and *O. esmarki*, in contrast, *Otp* shows it as an outgroup. With branch lengths  $\sim 0.03$  for both *H. sapiens* and *D. melanogaster*, the tree exhibits that both sequences are similar. *NeuroD1* shows an 86% confidence interval with *S. purpuratus* and *O. esmarki*, and the outgroup is *Mus musculus*, due to the lack of common ortholog in *H. sapiens*. *Myt1*, *Achaete-scute*, and *Zic2* follow with an 86%, 84%, and 81% confidence, respectively, for the branching of *S. purpuratus* and *O. esmarki*. All three trees follow a similar structure as the majority of other trees, with the ortholog of *D. melanogaster* as an outgroup. *Rx* provides a different tree structure with no outgroup provided, and both the *D. melanogaster* and *H. sapiens* sequences appearing to be fairly similar and cluster together. However, the more confident alignment at 74% confidence level is the branching of the *Rx* orthologs of *S. purpuratus* and *O. esmarki*. *Glia* also appears to follow this structure, but with the sequences of *Pan troglodytes* and *D. melanogaster* clustering together at a lower confidence than the 68% of *S. purpuratus* and *O. esmarki*. Both *Onecut2* and *Ngn* follow a similar tree structure with a 64%

and 65% bootstrapping confidence level, respectively, as well as both trees having *D. melanogaster* as an outgroup.

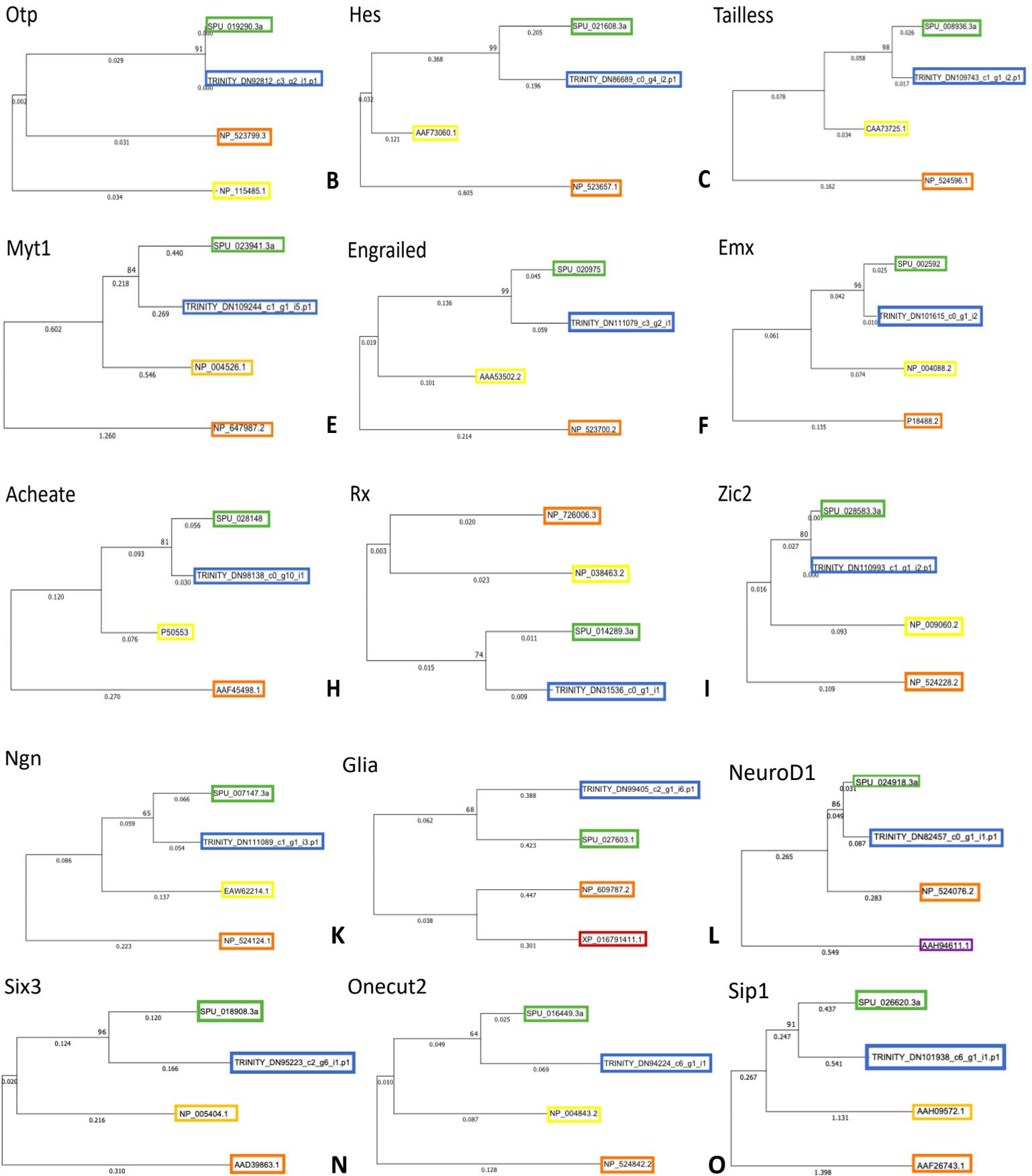


Figure 12. Phylogenetic trees of the 15 remaining neural transcripts. Bootstrap values are shown at each branch. Branch lengths are shown along each lineage.

## Discussion

Overall, the goal of this project was to provide a transcriptomic analysis of candidate neural transcripts through functional annotation and phylogenetic tree analysis of 24 neural transcripts in the brittle star *O. esmarki*. This species was used because it includes a unique, but common, form of development that has not been well studied. Specifically, the embryos have a rapid development to the juvenile stage, including modifications to the larval nervous system and the rapid development of the juvenile system within the larva. The comparison between these two morphologically different stages of development were thought to show similar transcripts, but at different levels and locations. It was hypothesized that both forms of development would include some similar transcripts because both stages have differentiated neurons. The vitellaria would express genes in earlier stages of neural development, while the juvenile expresses differentiation genes representing later development. The hypotheses involving the differences in the transcriptomes of the larval and juvenile stages were addressed through the use of various functional annotation methods (KEGG, KOG, and GO) and their different categories of transcription, growth factors, the nervous system, and development. After the annotation of *O. esmarki* transcriptomes, the identification and validation of chosen neural transcripts were completed to provide a better understanding of the shared orthologs with other species. By validating the 24 different neural transcripts, the developmental and neurological information in the two stages of development were able to give a better understanding for future studies.

### Transcriptome and Functional Annotation

The transcriptomes were first assessed by standard initial statistics and BUSCO to ensure their quality. From the initial statistics provided (Table 2), the N50 scores were high and showed promising results of the relatively large contigs needed to compose the transcriptome. However, this N50 score has been known to summarize assembly contiguity, but does not provide full confidence in the completeness of the transcriptome (Simão, et al. 2015). To provide a secondary method of validation and further analyze the quality of the results, a BUSCO analysis ran major metabolites against the metazoan dataset. An 84.8% completeness score was given for the vitellaria sample and 96.8% for the juvenile sample (Figure 5). Typically, a eukaryotic assembly above 80-85% completeness has been shown to have a good quality assembly and gave enough confidence to move forward with this *de novo* assembly (Simão, et al. 2015).

After the two measures of validation for the *de novo* transcriptomes, both samples were able to be analyzed with confidence, despite the minor differences in BUSCO results. One difference that occurred was a higher level of fragmented BUSCO results in the vitellaria at 13%, compared to 2.4% in the juvenile sample. One reason for this increase could be due to divergent or complex structures, keeping transcripts from being predicted in full, as well as an increase in alternative splice sites for this sample (Simão, et al. 2015). The initial assembly statistics also show an average contig length of 525.29 bases for the juvenile, as compared to a larger 618.86 bases for the vitellaria sample. The increase in length could imply that the initial assembly for the vitellaria did have a longer average contig length, leading to more transcripts ending up outside of the range of alignment to the BUSCO profiles (Simão, et al. 2015).

An interesting finding lies in the category of complete and duplicated BUSCOs for each sample, which were quite high. The number of complete and duplicated BUSCOs was at 54.1%

for the juvenile sample and 45.5% for the vitellaria sample. With such a large percentage of the total BUSCOS showing up as duplicated, this could imply sequence duplication in the genome. However, as shown in Cary (2019), analysis of echinoderm genomes does show a high proportion of duplicated BUSCOs, especially when compared to other non-vertebrate deuterostomes. As shown in this paper, the brittle star *Ophiothrix spiculata* has ~30-35% fraction of complete and duplicated BUSCO results (Cary, et al. 2019). This is consistent with the possibility that a duplication arose in the ancestor to the brittle star lineage but is also consistent with an increased presence of genetic heterozygosity and intra-species variation as the samples were fertilized in the wild with unknown and possibly heterogeneous paternal contribution.

With a high enough assembly quality to continue analysis, the overall comparison of the functional annotation methods between the two *O. esmarki* transcriptomes appeared to be fairly proportional (Figure 6). These similar results were expected, since both transcriptomes were composed of RNA of the same species at two stages of development. The overall content of gene function should have similar overall results due to the comparison of broad categories such as metabolism or cell growth. The largest potential difference between these functional annotation comparisons lies in the percent of GO IDs in the morphogenesis category for each sample. While GO category G (morphogenesis) is ~3% of the total IDs in the juvenile sample (Figure 6A), in the vitellaria (Figure 6B) it remains at just above 2% of the total. There was a higher level of morphogenesis IDs in the juvenile sample, but cell differentiation and cell organization levels are still similar during both stages of development implying major morphogenetic change is not occurring. With the overall functional annotation methods providing proportional results between the two stages of development, the annotation methods were sub-divided into categories of interest to evaluate any distinctive IDs. With the differing categories of functional annotation,

the IDs help to elucidate any differences in gene function for the contents of the two transcriptomes.

To compare overall ID abundance in each stage of development, the results from the three functional annotation tools were used in a t-test. While the t-test was able to provide a comparison of overall mean differences in the two samples, the specific categorical differences within functional annotation categories are unable to be compared with just two samples. These putative differences could be further evaluated with added replicates that could use statistical testing, such as t-test among the various replicates for the individual category. From these results shown, the extremely high p-values leave us unable to reject the null hypothesis that there are significant differences between the two stages in any of the functional annotation methods. For each of the observed dissimilarities in each functional annotation category, the only conclusion to be drawn is the differences do not impact the overall means of the samples.

The unique functional annotation IDs from KEGG, KOG, and GO were compared to observe any changes in function during the two stages. This comparison served to evaluate the hypothesis that both stages of development would express genes during different points in neural development. This is because as the vitellaria stage progresses to the juvenile stage, the genes involved in developing the juvenile nervous system would be activated and then turned off as the juvenile begins to express neural differentiation genes. As shown in Supplementary Tables 1, 2, 3, and 4, the functional annotation IDs unique to each stage represent the currently activated genes during each stage of development.

When evaluating the 419 unique KEGG IDs (Figure 7) for the vitellaria sample, the transcription factor category was composed of 25 different markers. KEGG was thought to provide a differing view of IDs by evaluating their role in major pathways and provide a new

categorization of applicable terms. The most important features for this project were thought to be the KEGG IDs involving the nervous system, development, morphogenesis, and cell differentiation. While unable to find Glass in the juvenile sample through both identification methods, a KEGG ID for this specific neural transcript was listed uniquely in the vitellaria stage. By locating this transcript in another stage of development, the proposed answer for its failure to be identified was confirmed. Glass was present in the organism but expressed at an earlier stage of development shown by its KEGG identification. Hox7 was also unable to be found at the juvenile stage, but remained listed as a suggested KEGG ID. The presence of Hox7 as a KEGG ID for the juvenile sample was a promising result due to our detection of Hox8. While Echinobase has listed that Hox7 and Hox8 have been mislabeled in the NCBI database, the false identification could also be due to the close relation of the two transcripts in the same family. DLX2 was also seen during this stage, and although its presence has been seen with relatively low expression in samples it does have a KEGG ID during the vitellaria stage (Burke, et al. 2006). Due to the lack of identification in the juvenile transcriptome, the annotation implies the expression levels may be higher earlier in development.

Other homeobox and zinc finger proteins are listed as KEGG IDs for the vitellaria stage, with many being from the same families of the neural transcripts of interest that were previously identified in this project. Uniquely listed during this stage of development were Six4, FoxI, Gli3, and ArntL, which are all family members of the 28 candidate neural transcripts as previously tested. The other large category of IDs were zinc finger proteins, such as Zic2, Myt1, or Glia. The vitellaria showed zinc finger KEGG IDs for Znf362, Zeb2, GliS1, and Zfp1. The other transcription factors listed in this category were not seen in the previously reviewed forms of

literature, such as *Ovo-L*, *Thrb*, or *Tbr1*, which are focused on the cytoskeleton, hormone receptors, or metabolic processes.

For the 17 unique transcriptome factors in the juvenile KEGG IDs, *SoxC* was the only neural transcript shown exclusively to this stage of development. Also, in this stage were familiar families of transcripts such zinc fingers, which included neural transcripts such as *Zeb1*, *Osr*, and *Krab*. The few others listed were less documented transcription factors that mainly dealt with general nucleic acid binding. Of the 205 shared transcription factors, *Achaete-scute*, *Engrailed*, *Emx*, *Zic2*, *Ngn*, *Arnt*, *Six3*, *FoxD*, *FoxJ1*, and *SoxD* were all listed from the neural transcripts of interest for this project. Other *Sox*, *Hox*, *Fox*, and *Gli* family transcription factors were also included in the shared KEGG IDs between the two modes of development. Overall, the differences in KEGG IDs for each stage of development support the hypothesis that differences in development are associated with differences in gene function. This also helps to support the conclusion that the inability to identify *Glass*, *DLX*, and *Hox7* during the juvenile stage of development could have been due to their presence in an earlier stage of development. The KEGG IDs found support the hypothesis that the KEGG functional annotation would help discern specific transcription factors and growth factors involved in the vitellaria and juvenile stages.

For the unique KOG IDs in both stages of development, the unique IDs contrast to KEGG with more over-arching terms, providing general support for the hypothesis. KOG was thought to provide information on the role of transcription factors and signal transduction. While the commonly held KOG IDs of forkheads, zinc-fingers, helix-loop-helix transcription factors, and homeobox transcription factors encompass all of the previously chosen neural transcripts, some IDs relating to transcription are still unique. The vitellaria stage contains the transcription

factor Bsh, which is a brain-specific homeobox and Cdx, which is a caudal-type homeobox that both regulate DNA binding transcription factor activity (Cary, et al. 2018). Also listed was a transcription factor for “DLX and related proteins with zinc binding and HOX domains”. DLX failed to be identified correctly with high confidence in the juvenile stage, most likely due to the low levels of expression of DLX during this stage. Since DLX did also appear in the vitellaria KEGG annotation, it can be assumed that DLX would most likely show higher expression earlier on in development, which caused its failure to be identified through eggNOG-mapper. Another interesting transcription factor at this stage is Nanos, which has highly regulating binding during every step of transcription that is essential for germ-line success. In *S. purpuratus*, Nanos knockdown larvae develop guts, skeletal systems and larval shape. Nanos is required for this formation or the coelomic pouches will not form and the larvae will not continue to develop. This protein sequence is extremely diverse in other echinoderms and could provide greater understanding to development and gene regulation for future studies (Oulhen & Wessel, 2014). These results supported the hypothesis that KOG could help identify neural candidates in each of the transcriptomes, to see if they are expressed. With KOG identifying many families of the transcripts of interest, new developmental transcription factors of future interest, and the DLX transcript that was unable to be identified with confidence, this form of functional annotation supported the understanding of the expression of candidate neural transcripts.

The final form of functional annotation, GO, provided more IDs than the other two methods combined, but also led to an overwhelming number of annotations. With some identification leading to attributes in plants, protists, and other highly unrelated taxa, the results of this form of annotation were not found to be particularly useful. Initially many of the categories of development (included in Supplementary Table 5), morphogenesis, and cell

differentiation piqued the interest of this project as we hoped to further elucidate the contents of gene function in these categories. However, GO provided no distinct gene products to focus on and instead listed many types of varied gene function with no narrowed focus on this particular organism or the different stages of development.

The initial reasoning for completing three different types of functional annotation was in anticipation for one or more of the tools providing information too broad or narrow for the scope of this project. Initially it was hypothesized that the GO functional annotation would help to discern which function the nervous system, morphogenesis, and development pathways played in the vitellaria and juvenile stages, but the overwhelming number of taxa and functions provided lacked any substantial information for this project. While KEGG and KOG provided new data and identification, the GO functional annotation method was too broad for the further analysis of the two stages of *O. esmarki* neural development. However, as previously hypothesized the different forms of functional annotation did bring varying amounts of clarity to gene function during the vitellaria and larval stages of the brittle star. The functional annotation also showed the differences in neural transcripts between the two developmental stages. These presence of unique IDs in each stage helped to show the transcription factors expressed during that point in neural development.

### Identification of Neural Transcripts

The phylogenetic analysis of the conserved domains of 24 neural transcripts was able to provide validation through the clustering with similar species and gaining high bootstrapping confidence levels of over 60%. When searching the juvenile transcriptome with tblastn and eggNOG-mapper, four transcripts were not able to be found with certainty, which include Hbn, Glass, DLX, and Hox7. The first transcript Hbn, was from Burke (2006), where the model

organism *S. purpuratus* was used and Hbn was present with the oral ganglia associated with the larval mouth. The reason this neural transcript may not have shown up in the *O. esmarki* transcriptome, may be due to the lack of a mouth in the vitellaria larva, and thus the lack of homologous larval oral ganglia. Also shown in Burke (2006) was the neural transcript DLX, which was shown to have exceptionally low RNA expression levels that may not have been picked up during the scope of the juvenile developmental stage, as it was not shown in adults in the Burke (2006) paper. The expression in the gastrula, early larva, and late larva were only 0.04%, 0.02%, and 0.01% respectively. However, DLX failed to be detected in adult tissues. This could explain why DLX was shown as a KEGG and KOG ID in the vitellaria but failed to appear in actual tblastn and eggNOG-mapper searches of the juvenile transcriptome. The associated IDs being identified in the vitellaria stage does suggest that it would be found by a tblastn and eggNOG-mapper search of the vitellaria transcriptome. For Hox7, its incorrect identification as Hox8 could be due to a common mistake in the NCBI database. The gene described as Hox8, with NCBI accession D85419, is actually Hox7 (SPU\_002634) (Cary, et al. 2018). Due to this error, eggNOG mapper may not be corrected on the changes in the database that arose following the misidentification, or the two neural transcripts could just be very similar within the family causing misidentification (Cary et al., 2018). Finally, the Glass neural transcript was never found in the eggNOG mapper search of the juvenile transcriptome, which could be due to poor annotation of the gene. However, the presence of the KEGG ID for Glass in the vitellaria sample supports the possibility that it could be present in an earlier stage of development.

Other identified transcripts were successfully confirmed through tblastn and eggNOG mapper, but initially lacked certainty after receiving less than a 50% bootstrapping confidence

level during phylogenetic analysis. The conserved domains of both Arnt and Otx (Figure 10) originally had a confidence score of 31% and 47% respectively, giving the tree analysis less confidence. When looking at the alignments for Arnt, it appears that sequences of the four orthologs used for the analysis contain only four different amino acids among them, making the sequences all extremely similar to each other. When looking at Figure 10A, the branch lengths are identical for the first three species and only a 0.035 branch length difference occurs when looking at the Arnt ortholog in *D. melanogaster*. The vast similarities in this conserved domain make a high bootstrapping score impossible, due to the ability for the almost identical branches to move among themselves. The same problem occurs in Otx, but with eight different amino acids among the orthologs of the four species, leading to a slightly higher bootstrapping confidence, but still falling short due to the similarity in the conserved regions of the orthologs. After doing the alignments again with the full sequences (Figure 11), the alignment scores greatly increased to 96% and 97%, giving a much higher confidence to the identified neural transcripts.

Another concern was based in the alternative species that were provided for neural transcript matches. One reason for this identification could be due to a lack of annotation in the eggNOG-mapper database for certain transcripts for *S. purpuratus*. *Saccoglossus kowalevskii* (Acorn worm), *Lepisosteus oculatus* (Spotted gar), and *Astyanax mexicanus* (Cave fish) have been used as model organisms and are highly annotated. Of the 24 transcripts identified, six were matched to the acorn worm, due to its close relationship to echinoderms. The spotted gar and cave fish each had one other transcript assigned, with the remaining 16 aligning to *S. purpuratus*. The alternative organisms had matching transcript names to the prospective candidate transcript that was being searched. The matching transcript name from the alternative species and the

matching identification for species and transcript in tblastn gave the confidence to proceed with phylogenetic analysis.

Overall, the chosen neural transcripts give an opportunity to develop hypotheses about possible functions in the nervous system. Most predictions will need to be verified with full length cDNAs and expression studies to determine conserved gene functions. However, phylogenetic analysis of the sequences of these neural transcripts revealed that *O. esmarki* and *S. purpuratus* are the closest orthologs. Most sequences have a shared ortholog with *H. sapiens*, reflecting their shared deuterostome heritage. In a few of the phylogenetic trees, the closest homolog was actually *D. melanogaster*, which could imply possible vertebrate-specific diversification of the sequence. Overall, there was a high conservation of neural transcripts across species that provides the opportunity to better understand neural development (Howard-Ashby et al., 2006). These results supported the hypothesis that the vitellaria and juvenile stages of development in *O. esmarki* will express similar neural transcripts as in the model sea urchin, and that there will be differences in neural transcripts between the two developmental stages.

### Future Work

For future work a subset of neural transcripts could be examined for tissue specific expression by *in situ* hybridization. In McClay (2018), neurogenesis in *S. purpuratus* is shown in three different domains. In *O. esmarki*, the expectation would be expression of these transcripts in apical ectoderm and ciliary band neurogenic regions, but no expression in the gut because *O. esmarki* does not have a functional gut until the juvenile stage. With future analysis focused on Nodal, BMP, FGF, and Wnt, downstream patterning of the nervous system could be monitored and compared to the previous McClay studies that include a gut system (McClay et al., 2018).

In contrast, following the Burke (2006) paper would lead to further studies in the area of developmental stages of neurogenesis. Through the study of the early, mid, and late stages of neurogenesis and the differentiation of neurons, this research would provide new lines of investigation into neural development. A greater comparison would be made by analyzing distinct body plans and separate nervous systems to provide more understanding to the evolutionary questions of different modes of development in *O. esmarki* (ancestral ophiopluteus larva vs abbreviated vitellaria larva) in brittle stars.

To build off of this project, gene expression analysis could be completed with added sample replicates. With added replication, a heat map or volcano plot could be analyzed to see the up and downregulated expression levels of specific neural transcripts. The house-keeping gene, actin, could also be used to analyze expression levels. If the expression of this gene was consistent in both samples, it would give the data more confidence to rule out any experimental or sampling variability. RNA from additional stages of development could also be analyzed to examine earlier developmental pathways. If other stages were analyzed, transcripts that were not able to be identified in the vitellaria and juvenile stages could be studied further. The vitellaria and juvenile transcriptomes will provide an opportunity to examine the development of other tissues beyond the nervous system and to uncover more about the evolution of the abbreviated mode of development.

**Literature Cited:**

- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. "Basic local alignment search tool." *J. Mol.* 1990. vol. 215, 403-410.
- Ashburner, M., Ball, C., Blake, J. et al. "Gene ontology: tool for the unification of biology." *Nat Genet.* May 2000;25(1):25-9. doi:10.1038/75556.
- Bolger, A., Lohhse, M., Usadel, B. "Trimmomatic: a flexible trimmer for Illumina sequence data." *Bioinformatics*, vol. 30, 2014, pp. 2114-20. doi:10.1093/bioinformatics/btu170
- Brooks, W. K., and C. Grave. "Ophiura brevispina". *Mem. Natl. Acad. Sci.* 1899. vol. 8 pp. 83-100.
- Burke, R., Angerer, L., Elphick, M. et al. "A genomic view of the sea urchin nervous system." *Dev. Biol.* 2006. 300, 434–460. <http://dx.doi.org/10.1016/j.ydbio.2006.08.007>.
- Burke, R. "Deuterostome neuroanatomy and the body plan paradox." *Evol. Dev.* Jan 2011; 13: 110–115. <https://doi.org/10.1111/j.1525-142X.2010.00460.x>.
- Burns, G., Thorndyke, M., Peck, L., Clark, M. "Transcriptome Pyrosequencing of the Antarctic Brittle Star *Ophionotus victoriae*." *Marine Genomics*, vol. 9, 2013, pp. 9–15., doi:10.1016/j.margen.2012.05.003.
- Cary G., Cameron R., and Hinman V. "EchinoBase: Tools for Echinoderm Genome Analyses." *Methods Mol Biol.* 2018.
- Cary G., Cameron R., and Hinman V. "Genomic Resources for the Study of Echinoderm Development and Evolution." *Methods in Cell Biology*, vol 151. 2019.
- Delroisse, J., Ortega-Martinez, O., Dupont, S., Mallefet, J., Flammang, P. "De Novo Transcriptome of the European Brittle Star *Amphiura filiformis* Pluteus Larvae." *Marine Genomics*, vol. 23, 2015, pp. 109–121., doi:10.1016/j.margen.2015.05.014.

- Ding, J., Zhao, L., Chang, Y. et al. “Transcriptome sequencing and characterization of Japanese scallop *Patinopecten yessoensis* from different shell color lines.” *PloS one*. vol.10, no. 2 13 Feb. 2015, doi:10.1371/journal.pone.0116406.
- Dupont, S., Thorndyke, W., Thorndyke, M., Burke, R. “Neural Development of the Brittle star *Amphiura filiformis*.” *Development Genes and Evolution*. vol. 219, no. 3, 2009, pp. 159–166., doi:10.1007/s00427-009-0277-9.
- Dylus, D., Czarkwiani, A., Blowes, L., Elphick, M., Oliveri, P. “Developmental transcriptomics of the brittle star *Amphiura filiformis* reveals gene regulatory network rewiring in echinoderm larval skeleton evolution.” *Genome Biology*, vol. 19, no. 26, 2017, pp 1-17. <https://doi.org/10.1186/s13059-018-1402-8>
- Ghosh, S. and Chan, C. “Analysis of RNA-Seq Data Using TopHat and Cufflinks.” *Plant Bioinformatics Methods in Molecular Biology*, 2016, pp. 339–361., doi:10.1007/978-1-4939-3167-5\_18.
- Grabherr, M., Haas, B., Yassour., et al. “Full-length transcriptome assembly from RNA-Seq data without a reference genome.” *Nature Biotechnology*, vol. 29,7. 15 May. 2011, pp. 644-52. doi:10.1038/nbt.1883
- Harmon M. “The Position of the Ophiuroidea Within the Phylum Echinodermata.” Master’s thesis. University of South Florida, Biology Department; 2005.
- Hirokawa, T., Komatsu, M., and Nakajima, Y. “Development of the Nervous System in the Brittle Star *Amphipholis kochii*.” *Development Genes and Evolution*, vol. 218, no. 1, 2007, pp. 15–21., doi:10.1007/s00427-007-0196-6.
- Hinman, V., and Burke, R. “Embryonic Neurogenesis in Echinoderms.” *Developmental Biology*, vol. 7, no. 4, 2018, doi:10.1002/wdev.316.

- Howard-Ashby, M., Materna, S., Brown., T. et al. “Gene Families Encoding Transcription Factors Expressed in Early Development of *Strongylocentrotus purpuratus*.” *Developmental Biology*, vol. 300, no. 1, 2006, pp. 90–107., doi:10.1016/j.ydbio.2006.08.033.
- Huerta-Cepas, J., Szklarzyk, D., Forslund., K. et al. “eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences.” *Nucleic Acids Research*, vol. 44, no. D1, 4 January 2016, pp. D286–D293, <https://doi.org/10.1093/nar/gkv1248>
- Hunter, S., Apweiler, R., Attwood, T., et al. “InterPro: the integrative protein signature database.” *Nucleic Acids Research*, vol. 37, Database issue, 2009, pp.D211-5, doi:10.1093/nar/gkn785
- Kanehisa, M., and Goto, S. “KEGG: Kyoto Encyclopedia of Genes and Genomes.” *Nucleic Acids Research*, vol. 28, no. 1, 2000, pp. 27–30., doi:10.1093/nar/28.1.27.
- Kumar S, Stecher G, and Tamura K. “MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets.” *Molecular Biology and Evolution*, vol. 33, 2016, pp.1870-1874
- Letunic, I and Bork, P. “20 years of the SMART protein domain annotation resource.” *Nucleic Acids Research*, doi:10.1093/nar/gkx922
- Littlewood D, Smith A, Clough K, Emson R: “The interrelationships of the echinoderm classes: morphological and molecular evidence.” *Biol J Linnean Soc*, vol. 61, 1997, pp. 409–438.
- MacBride, E. W. “The development of *Ophiothrix fragilis*.” *J. Cell Sci.* vol. 51, pp. 557-606.

- Mi, H., Muruganujan, A., Casagrande, J., Thomas, P. “Large-Scale gene function analysis with the PANTHER classification system.” *Nature Protocols*, vol. 8, no. 8, 2013, pp. 1551–1566. doi: 10.1038/nprot.2013.092.
- Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A., and Kanehisa, M. “KAAS: an automatic genome annotation and pathway reconstruction server.” *Nucleic Acids Res.* vol. 35, 2007, W182-W185. doi: 10.1093/nar/gkm321.
- Oulhen, N and Wessel, G . “Every which way--nanos gene regulation in echinoderms.” *Genesis*. vol. 52, no. 3, 2014, pp. 279-86. doi:10.1002/dvg.22737
- Paul C, Smith A. “The early radiation and phylogeny of echinoderms.” *Biol Rev.* vol 59,1984, pp. 443–481.
- Rowe, M., and Elphick, R. “The Neuropeptide Transcriptome of a Model Echinoderm, the Sea Urchin *Strongylocentrotus purpuratus*.” *General and Comparative Endocrinology*, vol. 179, no. 3, 2012, pp. 331–344., doi:10.1016/j.ygcen.2012.09.009.
- Simão, F., Waterhouse, R., Ioannidis, P., Kriventseva, E., and Zdobnov, E., “BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs.” *Bioinformatics*, vol. 31, no. 19, 2015, pp. 3210-2., doi: 10.1093/bioinformatics/btv351
- Strathmann, R. “Larval Feeding in Echinoderms.” *American Zoologist*, vol. 15, no. 3, 1975, pp. 717–730., doi:10.1093/icb/15.3.717.
- Sweet, H., Doolin, M., Yanowiak, C., Coots, A., Freyn, A., Armstrong, J., Spiecker, B. “Abbreviated Development of the Brooding Brittle Star *Ophioplocus esmarki*.” *The Biological Bulletin*, vol. 236, no. 2, 2019, pp. 75–87., doi:10.1086/701916.
- Tatusov, R., Fedorova, N., Jackson, J. et al. “The COG database: an updated version includes eukaryotes.” *BMC bioinformatics*, vol. 4, no. 41, 2003., doi:10.1186/1471-2105-4-41

Tu, Q., Brown, C.T., Davidson, E.H., Oliveri, P., “Sea urchin Forkhead gene family:Phylogeny and embryonic expression.” *Dev. Biol.* vol. 300, 2006, pp. 49–62.

doi:10.1016/j.ydbio.2006.09.031

Wu, S., Zhu, Z., Fu, L., Niu, B., Li, W. “WebMGA: a Customizable Web Server for Fast Metagenomic Sequence Analysis.” *BMC Genomics*, vol. 12, no. 1, 2011.,

doi:10.1186/1471-2164-12-444.

## Supplementary

Supplemental Table 1. Transcription factor KEGG IDs unique to the vitellaria and juvenile stages.

Juv IDs	Juv Function	Vit IDs	Vit Function
K07294	NR1C1; peroxisome proliferator-activated receptor alpha	K08362	THRB; thyroid hormone receptor beta
K09210	KLF15; krueppel-like factor 15	K08704	NR2A4; hepatocyte nuclear factor 4
K09220	KRAB; KRAB domain-containing zinc finger protein	K09214	GL; glass
K09453	SNAPC4; snRNA-activating protein complex subunit 4	K23480	ZNF362 384; zinc finger protein 362/384
K09434	ERF; ETS domain-containing transcription factor ERF	K04686	PITX2; paired-like homeodomain transcription factor 2
K142242	NFKBIZ; NF-kappa-B inhibitor zeta	K10172	BRA; brachyury protein
K09103	EBF; early B-cell factor ZEB1; zinc finger homeobox protein 1	K10174	TBR1; T-box brain protein 1
K09168	NFIA; nuclear factor I/A	K02296	ARNTL; aryl hydrocarbon receptor nuclear translocator-like protein 1
K09322	MEOX; homeobox protein MOX	K23560	ZEB2; zinc finger homeobox protein 2
K15608	PAX2; paired box protein 2	K09324	GSC; homeobox protein goosecoid
K09448	TEAD; transcriptional enhancer factor	K23581	SOX4; transcription factor SOX4
K09268	SOX11_12; transcription factor SOX11/12 (SOX group C)	K23194	GATAD2; transcriptional repressor p66
K15603	TCF4 12; transcription factor 4/12	K09216	OVOL; ovo
K09064	MYOD1; myogenic factor 3	K06230	GLI3; zinc finger protein GLI3
K09215	OSR; odd-skipped	K23195	CTCF; transcriptional repressor CTCF
K09378	ATBF1; AT-binding transcription factor 1	K09307	HOX 7; homeobox protein HoxA/B7
		K08561	NR6A1; germ cell nuclear factor
		K09232	GLIS1 3; zinc finger protein GLIS1/3
		K23317	MTF1; metal regulatory transcription factor 1
		K18488	DLX2; homeobox protein DLX2
		K09329	PRRX; paired mesoderm homeobox protein
		K09401	FOXI; forkhead box protein I
		K06053	RBPSUH; recombining binding protein suppressor of hairless
		K17441	ZFPM1; zinc finger protein ZFPM1
		K15615	SIX4; homeobox protein SIX4

Supplemental Table 2. KEGG cytokine and growth factor IDs unique to the vitellaria and juvenile stages.

Juv IDs	Juv IDs	Vit IDs	Vit IDs
ko:K05478	TNFSF15; tumor necrosis factor ligand superfamily member 15	ko:K05477	TNFSF14; tumor necrosis factor ligand superfamily member 14
		ko:K04666	NODAL; nodal
		ko:K05502	BMP1; bone morphogenetic protein 1
		ko:K04668	LEFTY; left-right determination factor
		ko:K05449	VEGFC_D; vascular endothelial growth factor C/D

Supplemental Table 3. KEGG nervous system IDs unique to the vitellaria and juvenile stages.

Juv IDs	Juv Function	Vit IDs	Vit Function
K01115	phospholipase D1/2	K00461	ALOX5; arachidonate 5-lipoxygenase
K03938	NADH dehydrogenase (ubiquinone) Fe-S protein 5	K04603	GRM1; metabotropic glutamate receptor 1
K13806	DAGL; sn1-specific diacylglycerol lipase	K00502	TPH1_2; tryptophan 5-monoxygenase
		K05869	CAMK4; calcium/calmodulin-dependent protein kinase IV
		K04534	GNAO, G-ALPHA-O; guanine nucleotide-binding protein G(o) subunit alpha
		K02296	ARNTL, BMAL1, CYC; aryl hydrocarbon receptor nuclear translocator-like protein 1
		K14387	SLC5A7, CHT1; solute carrier family 5 (high affinity choline transporter), member 7
		K04373	RPS6KA; ribosomal protein S6 kinase alpha-1/2/3/6
		K05704	SRC; tyrosine-protein kinase Src
		K05036	SLC6A3, DAT; solute carrier family 6 (neurotransmitter transporter, dopamine) member 3

Supplemental Table 4. KOG IDs unique to the vitellaria and juvenile stages.

Juvenile IDs		Vitellaria IDs		Vitellaria IDs (Continued)		Vitellaria IDs (Continued)	
KOG0023	Alcohol dehydrogenase	KOG0005	Ubiquitin-like protein	KOG2143	Uncharacterized conserved protein	KOG3268	Predicted E3 ubiquitin ligase
KOG0462	Elongation factor-type GTP-binding protein	KOG0171	Mitochondrial inner membrane protease, subunit IMP1	KOG2149	Uncharacterized conserved protein	KOG3276	Uncharacterized conserved protein
KOG0590	Checkpoint kinase and related serine/threonine protein kinases	KOG0218	Mismatch repair MSH3	KOG2184	Tuftelin-interacting protein TIP39, contains G-patch domain	KOG3297	DNA-directed RNA polymerase subunit E
KOG0724	Zuotin and related molecular chaperones	KOG0283	WD40 repeat-containing protein	KOG2257	N-acetylglucosaminyltransferase complex, subunit PIG-P, required for phosphatidylinositol biosynthesis	KOG3307	Molybdopter in converting factor subunit 2
KOG0805	Carbon-nitrogen hydrolase	KOG0310	Conserved WD40 repeat-containing protein	KOG2316	Predicted ATPase (PP-loop superfamily)	KOG3322	Ribonuclease s P/MRP protein subunit
KOG0974	Conserved domain family	KOG0322	G-protein beta subunit-like protein GNB1L,	KOG2350	Zn-finger protein joined to JAZF1	KOG3335	Predicted coiled-coil protein

			contains WD repeats				
KOG1546	Metacaspase involved in regulation of apoptosis	KOG0325	Lipoyltransferase	KOG2373	Predicted mitochondrial DNA helicase twinkle	KOG3339	Predicted glycosyltransferase
KOG1733	Mitochondrial import inner membrane translocase, subunit TIM13	KOG0348	ATP-dependent RNA helicase	KOG2464	Serine/threonine kinase (haspin family)	KOG3355	Mitochondrial sulfhydryl oxidase involved in the biogenesis of cytosolic Fe/S proteins
KOG1754	40S ribosomal protein S15/S22	KOG0433	Isoleucyl-tRNA synthetase	KOG2477	Uncharacterized conserved protein	KOG3438	DNA-directed RNA polymerase, subunit L
KOG1801	tRNA-splicing endonuclease positive effector (SEN1)	KOG0456	Aspartate kinase	KOG2486	Predicted GTPase	KOG3755	SATB1 matrix attachment region binding protein
KOG1946	RNA polymerase I transcription factor UAF	KOG0491	Transcription factor BSH, contains HOX domain	KOG2518	5'-3' exonuclease	KOG3870	Uncharacterized conserved protein
KOG2057	Predicted equilibrative nucleoside transporter protein	KOG0611	Predicted serine/threonine protein kinase	KOG2571	Chitin synthase/hyaluronan synthase (glycosyltransferases)	KOG3917	Beta-1,4-galactosyltransferase B4GALT7/SQV-3
KOG2212	Alpha-amylase	KOG0680	Actin-related protein - Arp6p	KOG2599	Pyridoxal/pyridoxine/pyridoxamine kinase	KOG3931	Uncharacterized conserved protein
KOG2289	Rhomboid family proteins	KOG0721	Molecular chaperone (DnaJ superfamily)	KOG2661	Predicted integral membrane protein	KOG3936	Nitroreductases
KOG2404	Fumarate reductase, flavoprotein subunit	KOG0722	Molecular chaperone (DnaJ superfamily)	KOG2688	Transcription-associated recombination protein - Thp1p	KOG3965	Predicted glycerate kinase
KOG2419	Phosphatidylserine decarboxylase	KOG0744	AAA+-type ATPase	KOG2710	Rho GTPase-activating protein	KOG3978	Predicted membrane protein
KOG2583	Ubiquinol cytochrome c reductase, subunit QCR2	KOG0769	Predicted mitochondrial carrier protein	KOG2800	Conserved developmentally regulated protein	KOG4023	Uncharacterized conserved protein
KOG2767	Translation initiation factor 5 (eIF-5)	KOG0779	Protease, Ulp1 family	KOG2859	DNA repair protein, member of the recA/RAD51 family	KOG4058	Uncharacterized conserved protein
KOG3047	Predicted transcriptional regulator UXT	KOG0809	SNARE protein TLG2/Syntaxin 16	KOG2868	Decapping enzyme complex component DCP1	KOG4079	Putative mitochondrial ribosomal protein mRpS2

KOG3058	Uncharacterized conserved protein	KOG0826	Predicted E3 ubiquitin ligase involved in peroxisome organization	KOG2873	Ubiquinol cytochrome c reductase assembly protein CBP3	KOG4134	DNA-dependent RNA polymerase I
KOG3137	Peptide deformylase	KOG0827	Predicted E3 ubiquitin ligase	KOG2896	UV radiation resistance associated protein	KOG4141	DNA repair and recombination protein RAD52/RAD22
KOG3160	Gamma-interferon inducible lysosomal thiol reductase	KOG0848	Transcription factor Caudal, contains HOX domain	KOG2911	Uncharacterized conserved protein	KOG4160	BPI/LBP/CE1P family protein
KOG3332	N-acetylglucosaminyl phosphatidylinositol de-N-acetylase	KOG0850	Transcription factor DLX and related proteins with LIM Zn-binding and HOX domains	KOG2940	Predicted methyltransferase	KOG4174	Uncharacterized conserved protein
KOG3465	Signal recognition particle, subunit Srp9	KOG0884	Similar to cyclophilin-type peptidyl-prolyl cis-trans isomerase	KOG2942	Uncharacterized conserved protein	KOG4352	Fas-mediated apoptosis inhibitor FAIM
KOG3485	Uncharacterized conserved protein	KOG0894	Ubiquitin-protein ligase	KOG2978	Dolichol-phosphate mannosyltransferase	KOG4372	Predicted alpha/beta hydrolase
KOG4078	Putative mitochondrial ribosomal protein mRpS35	KOG1019	Retinoblastoma pathway protein LIN-9/chromatin-associated protein Aly	KOG2986	Uncharacterized conserved protein	KOG4380	Carnitine deficiency associated protein
KOG4092	Mitochondrial F1F0-ATP synthase, subunit f	KOG1045	Uncharacterized conserved protein HEN1/COR12/MBDA2	KOG2994	Uracil DNA glycosylase	KOG4382	Uncharacterized conserved protein
KOG4114	Cytochrome c oxidase assembly protein PET191	KOG1071	Mitochondrial translation elongation factor EF-Tsmt, catalyzes nucleotide exchange on EF-Tumt	KOG3004	Meiotic chromosome segregation protein	KOG4478	Protein tyrosine kinase
KOG4118	Uncharacterized conserved protein	KOG1204	Predicted dehydrogenase	KOG3009	Predicted carbohydrate kinase, contains PfkB domain	KOG4487	Uncharacterized conserved protein
KOG4119	G protein gamma subunit	KOG1210	Predicted 3-ketosphinganine reductase	KOG3059	N-acetylglucosaminyltransferase complex, subunit PIG-C/GPI2, required for	KOG4491	Predicted membrane protein

					phosphatidylinositol biosynthesis		
KOG4244	Failed axon connections (fax) protein/glutathione S-transferase-like protein	KOG1223	Isochorismate synthase	KOG3127	Deoxycytidylate deaminase	KOG4509	Uncharacterized conserved protein
KOG4394	RNase P subunit that is not also a subunit of RNase MRP, involved in pre-tRNA processing	KOG1228	Uncharacterized conserved protein	KOG3131	Uncharacterized conserved protein	KOG4510	Permease of the drug/metabolite transporter (DMT) superfamily
KOG4506	Uncharacterized conserved protein	KOG1230	Protein containing repeated kelch motifs	KOG3134	Predicted membrane protein	KOG4541	Nuclear transport receptor exportin 4 (importin beta superfamily)
KOG4543	Uncharacterized conserved protein	KOG1297	Uncharacterized conserved protein	KOG3156	Uncharacterized membrane protein	KOG4567	GTPase-activating protein
KOG4662	NADH dehydrogenase subunit 3 and related proteins	KOG1433	DNA repair protein RAD51/RHP55	KOG3164	Uncharacterized proteins of PilT N-term./Vapc superfamily	KOG4576	Sulfite oxidase, heme-binding component
KOG4679	Uncharacterized protein PSP1	KOG1478	3-keto sterol reductase	KOG3189	Phosphomannomutase	KOG4595	Uncharacterized conserved protein
KOG4685	tRNA splicing endonuclease SEN2	KOG1545	Voltage-gated shaker-like K <sup>+</sup> channel KCNA	KOG3192	Mitochondrial J-type chaperone	KOG4602	Nanos and related proteins
KOG4769	Cytochrome c oxidase, subunit I	KOG1568	Mitochondrial inner membrane protease, subunit IMP2	KOG3197	Predicted hydrolases of HD superfamily	KOG4621	Uncharacterized conserved protein
KOG4794	Thymosin beta	KOG1617	CDP-alcohol phosphatidyl transferase/Phosphatidylglycerol-phosphate synthase	KOG3222	Inosine triphosphate pyrophosphatase	KOG4653	Uncharacterized conserved protein
KOG4845	NADH dehydrogenase	KOG1774	Small nuclear ribonucleoprotein E	KOG3246	Sentrin-specific cysteine protease (Ulp1 family)	KOG4714	Nucleoporin
KOG4852	Uncharacterized conserved protein	KOG1949	Uncharacterized conserved protein	KOG3266	Predicted glycine cleavage system H protein	KOG4734	Uncharacterized conserved protein
				KOG4739	Uncharacterized protein	KOG4784	Uncharacterized conserved protein

				KOG4746	Small nuclear RNA activating complex (SNAPc), subunit SNAP43	KOG4814	Uncharacterized conserved protein
				KOG4751	DNA recombinational repair protein BRCA2	KOG4841	Dolichol-phosphate mannosyltransferase, subunit 3

Supplemental Table 5. Developmental GO IDs unique to the vitellaria and juvenile stages.

Juvenile IDs	Juvenile Function	Juvenile IDs (Continued)		Vit IDs	Vit Function	Vit IDs (Continued)	
GO:0000905	sporocarp development involved in asexual reproduction	GO:2000733	regulation of glial cell-derived neurotrophic factor receptor signaling pathway involved in ureteric bud formation	GO:0001545	primary ovarian follicle growth	GO:0097118	neuroligin clustering involved in postsynaptic membrane assembly
GO:0001713	ectodermal cell fate determination	GO:2000734	negative regulation of glial cell-derived neurotrophic factor receptor signaling pathway involved in ureteric bud formation	GO:0001828	inner cell mass cellular morphogenesis	GO:1901163	regulation of trophoblast cell migration
GO:0002317	plasma cell differentiation	GO:2000793	cell proliferation involved in heart valve development	GO:0001830	trophectodermal cell fate commitment	GO:1901165	positive regulation of trophoblast cell migration
GO:0002572	pro-T cell differentiation	GO:2000800	regulation of endocardial cushion to mesenchymal transition involved in heart valve formation	GO:0001923	B-1 B cell differentiation	GO:1901490	regulation of lymphangiogenesis
GO:0003138	primary heart field specification	GO:2000802	positive regulation of endocardial cushion to mesenchymal transition involved in heart valve formation	GO:0002206	gene conversion of immunoglobulin genes	GO:1901491	negative regulation of lymphangiogenesis
GO:0003172	sinoatrial valve development	GO:2000979	positive regulation of forebrain neuron differentiation	GO:0002314	germinal center B cell differentiation	GO:2000004	regulation of metanephric S-shaped body morphogenesis
GO:0003185	sinoatrial valve morphogenesis	GO:2001035	regulation of tongue muscle cell differentiation	GO:0002507	tolerance induction	GO:2000005	negative regulation of metanephric S-shaped body morphogenesis
GO:0003257	positive regulation of transcription from RNA polymerase II promoter involved in myocardial precursor cell differentiation	GO:2001037	positive regulation of tongue muscle cell differentiation	GO:0002565	somatic diversification of immune receptors via gene conversion	GO:2000006	regulation of metanephric comma-shaped body morphogenesis

GO:0007387	anterior compartment pattern formation	GO:0003310	pancreatic A cell differentiation	GO:0002574	thrombocyte differentiation	GO:2000007	negative regulation of metanephric comma-shaped body morphogenesis
GO:0007462	R1/R6 cell fate commitment	GO:0003322	pancreatic A cell development	GO:0003129	heart induction	GO:2000024	regulation of leaf development
GO:0007496	anterior midgut development	GO:0007347	regulation of preblastoderm mitotic cell cycle	GO:0003142	cardiogenic plate morphogenesis	GO:2000039	regulation of trichome morphogenesis
GO:0007510	cardioblast cell fate determination	GO:0007375	anterior midgut invagination	GO:0003175	tricuspid valve development	GO:2000137	negative regulation of cell proliferation involved in heart morphogenesis
GO:0009553	embryo sac development	GO:0007376	cephalic furrow formation	GO:0003186	tricuspid valve morphogenesis	GO:2000172	regulation of branching morphogenesis of a nerve
GO:0009640	photomorphogenesis	GO:0007385	specification of segmental identity, abdomen	GO:0003195	tricuspid valve formation	GO:2000173	negative regulation of branching morphogenesis of a nerve
GO:0010051	xylem and phloem pattern formation	GO:0009901	anther dehiscence	GO:0003199	endocardial cushion to mesenchymal transition involved in heart valve formation	GO:2000347	positive regulation of hepatocyte proliferation
GO:0010093	specification of floral organ identity	GO:0009908	flower development	GO:0003243	circumferential growth involved in left ventricle morphogenesis	GO:2000722	regulation of cardiac vascular smooth muscle cell differentiation
GO:0010103	stomatal complex morphogenesis	GO:0009956	radial pattern formation	GO:0003250	regulation of cell proliferation involved in heart valve morphogenesis	GO:2000723	negative regulation of cardiac vascular smooth muscle cell differentiation
GO:0010150	leaf senescence	GO:0009960	endosperm development	GO:0003251	positive regulation of cell proliferation involved in heart valve morphogenesis	GO:2000764	positive regulation of semaphorin-plexin signaling pathway involved in outflow tract morphogenesis
GO:0010187	negative regulation of seed germination	GO:0010029	regulation of seed germination	GO:0003285	septum secundum developmen	GO:0097118	neuroligin clustering involved in postsynaptic membrane assembly
GO:0019101	female somatic sex determination	GO:0010260	animal organ senescence	GO:0003290	atrial septum secundum morphogenesis	GO:1901163	regulation of trophoblast cell migration
GO:0021519	spinal cord association neuron specification	GO:0010305	leaf vascular tissue pattern formation	GO:0003357	noradrenergic neuron differentiation	GO:1901165	positive regulation of trophoblast cell migration
GO:0021541	ammon gyrus development	GO:0010374	stomatal complex development	GO:0003379	establishment of cell polarity involved in gastrulation cell migration	GO:1901490	regulation of lymphangiogenesis
GO:0021547	midbrain-hindbrain boundary initiation	GO:0010376	stomatal complex formation	GO:0003386	amphid sensory organ development	GO:1901491	negative regulation of lymphangiogenesis
GO:0021556	central nervous system formation	GO:0010589	leaf proximal/distal pattern formation	GO:0003387	neuron differentiation involved in amphid sensory organ development	GO:2000004	regulation of metanephric S-shaped body morphogenesis
GO:0021586	pons maturation	GO:0014708	regulation of somitomeric trunk muscle development	GO:0003388	neuron development involved in amphid	GO:2000005	negative regulation of metanephric S-

					sensory organ development		shaped body morphogenesis
GO:0021703	locus ceruleus development	GO:0014709	positive regulation of somitomeric trunk muscle development	GO:0003389	retrograde extension	GO:2000006	regulation of metanephric comma-shaped body morphogenesis
GO:0021894	cerebral cortex GABAergic interneuron development	GO:0021718	superior olivary nucleus development	GO:0003390	dendrite development by retrograde extension	GO:2000007	negative regulation of metanephric comma-shaped body morphogenesis
GO:0021902	commitment of neuronal cell to specific neuron type in forebrain	GO:0021722	superior olivary nucleus maturation	GO:0003391	amphid sensory organ dendrite retrograde extension	GO:2000024	regulation of leaf development
GO:0021905	forebrain-midbrain boundary formation	GO:0021759	globus pallidus development	GO:0007382	specification of segmental identity, maxillary segment	GO:2000039	regulation of trichome morphogenesis
GO:0021917	somatic motor neuron fate commitment	GO:0021775	smoothened signaling pathway involved in ventral spinal cord interneuron specification	GO:0007463	R2/R5 cell fate commitment	GO:2000137	negative regulation of cell proliferation involved in heart morphogenesis
GO:0021918	regulation of transcription from RNA polymerase II promoter involved in somatic motor neuron fate commitment	GO:0021776	smoothened signaling pathway involved in spinal cord motor neuron cell fate specification	GO:0019102	male somatic sex determination	GO:0007487	analia development
GO:0021920	regulation of transcription from RNA polymerase II promoter involved in spinal cord association neuron specification	GO:0021830	interneuron migration from the subpallium to the cortex	GO:0021540	corpus callosum morphogenesis	GO:0007495	visceral mesoderm-endoderm interaction involved in midgut development
GO:0022012	subpallium cell proliferation in forebrain	GO:0021853	cerebral cortex GABAergic interneuron migration	GO:0021560	abducens nerve development	GO:0010214	seed coat development
GO:0022018	lateral ganglionic eminence cell proliferation	GO:0021870	Cajal-Retzius cell differentiation	GO:0021598	abducens nerve morphogenesis	GO:0014041	regulation of neuron maturation
GO:0045656	negative regulation of monocyte differentiation	GO:0030222	eosinophil differentiation	GO:0021599	abducens nerve formation	GO:0014042	positive regulation of neuron maturation
GO:0045658	regulation of neutrophil differentiation	GO:0031133	regulation of axon diameter	GO:0021658	rhombomere 3 morphogenesis	GO:0014043	negative regulation of neuron maturation
GO:0045660	positive regulation of neutrophil differentiation	GO:0033278	cell proliferation in midbrain	GO:0035142	dorsal fin morphogenesis	GO:0021679	cerebellar molecular layer development
GO:0046665	amnioserosa maintenance	GO:0035154	terminal cell fate specification, open tracheal system	GO:0035144	anal fin morphogenesis	GO:0021732	midbrain-hindbrain boundary maturation
GO:0048296	regulation of isotype switching to IgA isotypes	GO:0035292	specification of segmental identity, trunk	GO:0035263	genital disc sexually dimorphic development	GO:0021750	vestibular nucleus development
GO:0048298	positive regulation of isotype switching to IgA isotypes	GO:0035318	imaginal disc-derived wing hair outgrowth	GO:0035469	determination of pancreatic left/right asymmetry	GO:0021754	facial nucleus development
GO:0048334	regulation of mesodermal cell fate determination	GO:0035462	determination of left/right asymmetry in diencephalon	GO:0035630	bone mineralization involved in bone maturation	GO:0021896	forebrain astrocyte differentiation
GO:0048335	negative regulation of mesodermal cell fate determination	GO:0035788	cell migration involved in	GO:0035675	neuromast hair cell development	GO:0021897	forebrain astrocyte development

			metanephros development				
GO:0048369	lateral mesoderm morphogenesis	GO:0035789	metanephric mesenchymal cell migration	GO:0035678	neuromast hair cell morphogenesis	GO:0021966	corticospinal neuron axon guidance
GO:0048370	lateral mesoderm formation	GO:0035793	positive regulation of metanephric mesenchymal cell migration by platelet-derived growth factor receptor-beta signaling pathway	GO:0039007	pronephric nephron morphogenesis	GO:0021972	corticospinal neuron axon guidance through spinal cord
GO:0048371	lateral mesodermal cell differentiation	GO:0035854	eosinophil fate commitment	GO:0039008	pronephric nephron tubule morphogenesis	GO:0021985	neurohypophysis development
GO:0048437	floral organ development	GO:0042001	hermaphrodite somatic sex determination	GO:0042482	positive regulation of odontogenesis	GO:0022004	midbrain-hindbrain boundary maturation during brain development
GO:0048438	floral whorl development	GO:0042004	feminization of hermaphrodite soma	GO:0042488	positive regulation of odontogenesis of dentin-containing tooth	GO:0030710	regulation of border follicle cell delamination
GO:0048443	stamen development	GO:0042479	positive regulation of eye photoreceptor cell development	GO:0042637	catagen	GO:0030862	positive regulation of polarized epithelial cell differentiation
GO:0048444	floral organ morphogenesis	GO:0042673	regulation of retinal cone cell fate specification	GO:0042662	negative regulation of mesodermal cell fate specification	GO:0033085	negative regulation of T cell differentiation in thymus
GO:0048449	floral organ formation	GO:0042701	progesterone secretion	GO:0042665	regulation of ectodermal cell fate specification	GO:0033335	anal fin development
GO:0048466	androecium development	GO:0043366	beta selection	GO:0042666	negative regulation of ectodermal cell fate specification	GO:0033337	dorsal fin developmen
GO:0048539	bone marrow development	GO:0045315	positive regulation of compound eye photoreceptor development	GO:0048696	regulation of collateral sprouting in absence of injury	GO:0042682	regulation of compound eye cone cell fate specification
GO:0048618	post-embryonic foregut morphogenesis	GO:0045656	negative regulation of monocyte differentiation	GO:0048698	negative regulation of collateral sprouting in absence of injury	GO:0042683	negative regulation of compound eye cone cell fate specification
GO:0060461	right lung morphogenesis	GO:0048620	post-embryonic hindgut morphogenesis	GO:0048893	afferent axon development in lateral line nerve	GO:0042695	thelarche
GO:0060486	Clara cell differentiation	GO:0048621	post-embryonic digestive tract morphogenesis	GO:0048894	efferent axon development in a lateral line nerve	GO:0043704	photoreceptor cell fate specification
GO:0060510	type II pneumocyte differentiation	GO:0048632	negative regulation of skeletal muscle tissue growth	GO:0048899	anterior lateral line development	GO:0044333	Wnt signaling pathway involved in digestive tract morphogenesis
GO:0060677	ureteric bud elongation	GO:0048653	anther development	GO:0048929	efferent axon development in posterior lateral line nerve	GO:0045163	clustering of voltage-gated potassium channels
GO:0060776	simple leaf morphogenesis	GO:0048672	positive regulation of collateral sprouting	GO:0048933	afferent axon development in posterior lateral line nerve	GO:0045464	R8 cell fate specification
GO:0060838	lymphatic endothelial cell fate commitment	GO:0048734	proboscis morphogenesis	GO:0051884	regulation of timing of anagen	GO:0045625	regulation of T-helper 1 cell differentiation
GO:0060849	regulation of transcription involved in lymphatic endothelial cell fate commitment	GO:0048825	cotyledon development	GO:0051885	positive regulation of timing of anagen	GO:0045626	negative regulation of T-helper 1 cell differentiation

GO:0060873	anterior semicircular canal development	GO:0048885	neuromast deposition	GO:0051891	positive regulation of cardioblast differentiation	GO:0045679	regulation of R8 cell differentiation
GO:0060875	lateral semicircular canal development	GO:0055026	negative regulation of cardiac muscle tissue development	GO:0055018	regulation of cardiac muscle fiber development	GO:0045680	negative regulation of R8 cell differentiation
GO:0060913	cardiac cell fate determination	GO:0055060	asymmetric neuroblast division resulting in ganglion mother cell formation	GO:0055020	positive regulation of cardiac muscle fiber development	GO:0045804	negative regulation of eclosion
GO:0060957	endocardial cell fate commitment	GO:0060010	Sertoli cell fate commitment	GO:0060031	mediolateral intercalation	GO:0045805	positive regulation of eclosion
GO:0060981	cell migration involved in coronary angiogenesis	GO:0060032	notochord regression	GO:0060036	notochord cell vacuolation	GO:0046595	establishment of pole plasm mRNA localization
GO:0060995	cell-cell signaling involved in kidney development	GO:0060222	regulation of retinal cone cell fate commitment	GO:0060040	retinal bipolar neuron differentiation	GO:0048054	R2/R5 cell differentiation
GO:0061034	olfactory bulb mitral cell layer development	GO:0060366	lambdoid suture morphogenesis	GO:0060138	fetal process involved in parturition	GO:0048088	regulation of male pigmentation
GO:0061037	negative regulation of cartilage development	GO:0060367	sagittal suture morphogenesis	GO:0060447	bud outgrowth involved in lung branching	GO:0048092	negative regulation of male pigmentation
GO:0061067	negative regulation of dauer larval development	GO:0060458	right lung development	GO:0060480	lung goblet cell differentiation	GO:0048391	intermediate mesoderm formation
GO:0072106	regulation of ureteric bud formation	GO:0061193	taste bud development	GO:0060481	lobar bronchus epithelium development	GO:0048392	intermediate mesodermal cell differentiation
GO:0072107	positive regulation of ureteric bud formation	GO:0061194	taste bud formation	GO:0060502	epithelial cell proliferation involved in lung morphogenesis	GO:0048677	axon extension involved in regeneration
GO:0072156	distal tubule morphogenesis	GO:0061195	fungiform papilla development	GO:0061055	myotome development	GO:0060503	bud dilation involved in lung branching
GO:0072204	cell-cell signaling involved in metanephros development	GO:0061217	regulation of mesonephros development	GO:0061149	BMP signaling pathway involved in ureter morphogenesis	GO:0060599	lateral sprouting involved in mammary gland duct morphogenesis
GO:0072223	metanephric glomerular mesangium development	GO:0061289	Wnt signaling pathway involved in kidney development	GO:0061150	renal system segmentation	GO:0060618	nipple development
GO:0072262	metanephric glomerular mesangial cell proliferation involved in metanephros development	GO:0061290	canonical Wnt signaling pathway involved in metanephric kidney development	GO:0061151	BMP signaling pathway involved in renal system segmentatio	GO:0060658	nipple morphogenesis
GO:0072286	metanephric connecting tubule development	GO:0061300	cerebellum vasculature development	GO:0061155	pulmonary artery endothelial tube morphogenesis	GO:0060744	mammary gland branching involved in thelarche
GO:0072574	hepatocyte proliferation	GO:0061301	cerebellum vasculature morphogenesis	GO:0061332	Malpighian tubule bud morphogenesis	GO:0060775	planar cell polarity pathway involved in gastrula mediolateral intercalation
GO:0072575	positive regulation of hepatocyte proliferation	GO:0061445	endocardial cushion cell fate commitment	GO:0061341	non-canonical Wnt signaling pathway involved in heart development	GO:0060796	regulation of transcription involved in primary germ layer cell fate commitment
GO:0080086	stamen filament development	GO:0070171	negative regulation of tooth mineralization	GO:0061346	planar cell polarity pathway involved in heart morphogenesis	GO:0060803	BMP signaling pathway involved in mesodermal cell fate specification

GO:0090050	positive regulation of cell migration involved in sprouting angiogenesis	GO:0071691	cardiac muscle thin filament assembly	GO:0061347	planar cell polarity pathway involved in outflow tract morphogenesis	GO:0060807	regulation of transcription from RNA polymerase II promoter involved in definitive endodermal cell fate specification
GO:0090095	regulation of metanephric cap mesenchymal cell proliferation	GO:0072027	connecting tubule development	GO:0061348	planar cell polarity pathway involved in ventricular septum morphogenesis	GO:0060832	oocyte animal/vegetal axis specification
GO:0090096	positive regulation of metanephric cap mesenchymal cell proliferation	GO:0072060	outer medullary collecting duct development	GO:0061349	planar cell polarity pathway involved in cardiac right atrium morphogenesis	GO:0060994	regulation of transcription from RNA polymerase II promoter involved in kidney development
GO:0090133	mesendoderm migration	GO:0072061	inner medullary collecting duct development	GO:0061350	planar cell polarity pathway involved in cardiac muscle tissue morphogenesis	GO:0061043	regulation of vascular wound healing
GO:2000793	cell proliferation involved in heart valve development	GO:0090134	cell migration involved in mesendoderm migration	GO:0061354	planar cell polarity pathway involved in pericardium morphogenesis	GO:0060503	bud dilation involved in lung branching
GO:2000800	regulation of endocardial cushion to mesenchymal transition involved in heart valve formation	GO:0090171	chondrocyte morphogenesis	GO:0072096	negative regulation of branch elongation involved in ureteric bud branching	GO:0070586	cell-cell adhesion involved in gastrulation
GO:2000802	positive regulation of endocardial cushion to mesenchymal transition involved in heart valve formation	GO:0090249	regulation of cell motility involved in somitogenic axis elongation	GO:0072097	negative regulation of branch elongation involved in ureteric bud branching by BMP signaling pathway	GO:0070787	conidiophore development
GO:2000979	positive regulation of forebrain neuron differentiation	GO:0097374	sensory neuron axon guidance	GO:0072099	anterior/posterior pattern specification involved in ureteric bud development	GO:0070791	cleistothecium development
GO:2001035	regulation of tongue muscle cell differentiation	GO:1900140	regulation of seedling development	GO:0072100	specification of ureteric bud anterior/posterior symmetry	GO:0071109	superior temporal gyrus development
GO:2000589	regulation of metanephric mesenchymal cell migration	GO:1900238	regulation of metanephric mesenchymal cell migration by platelet-derived growth factor receptor-beta signaling pathway	GO:0072101	specification of ureteric bud anterior/posterior symmetry by BMP signaling pathway	GO:0071893	BMP signaling pathway involved in nephric duct formation
GO:2000591	positive regulation of metanephric mesenchymal cell migration	GO:2000607	negative regulation of cell proliferation involved in mesonephros development	GO:0072116	pronephros formation	GO:0072013	glomus development
GO:2000606	regulation of cell proliferation involved in mesonephros development	GO:2000702	regulation of fibroblast growth factor receptor signaling pathway involved in ureteric bud formation	GO:0072138	mesenchymal cell proliferation involved in ureteric bud development	GO:0072023	thick ascending limb development
GO:2000703	negative regulation of fibroblast growth factor receptor signaling pathway involved in ureteric bud formation	GO:2000733	regulation of glial cell-derived neurotrophic factor receptor signaling pathway involved in	GO:0072139	glomerular parietal epithelial cell differentiation	GO:0072024	macula densa development

			ureteric bud formation				
GO:2000589	regulation of metanephric mesenchymal cell migration			GO:0072149	glomerular visceral epithelial cell fate commitment	GO:0072051	juxtaglomerular apparatus development
				GO:0072158	proximal tubule morphogenesis	GO:0072069	DCT cell differentiation
				GO:0097116	gephyrin clustering involved in postsynaptic density assembly	GO:0072191	ureter smooth muscle development
				GO:0097118	neuroligin clustering involved in postsynaptic membrane assembly	GO:0072192	ureter epithelial cell differentiation
				GO:1901163	regulation of trophoblast cell migration	GO:0072193	ureter smooth muscle cell differentiation
				GO:1901165	positive regulation of trophoblast cell migration	GO:0072198	mesenchymal cell proliferation involved in ureter development
				GO:1901490	regulation of lymphangiogenesis	GO:0072199	regulation of mesenchymal cell proliferation involved in ureter development
				GO:1901491	negative regulation of lymphangiogenesis	GO:0072200	negative regulation of mesenchymal cell proliferation involved in ureter development
				GO:2000004	regulation of metanephric S-shaped body morphogenesis	GO:0072206	metanephric juxtaglomerular apparatus development
				GO:2000005	negative regulation of metanephric S-shaped body morphogenesis	GO:0072227	metanephric macula densa development
				GO:2000006	regulation of metanephric comma-shaped body morphogenesis	GO:0072229	metanephric proximal convoluted tubule development
				GO:2000007	negative regulation of metanephric comma-shaped body morphogenesis	GO:0072233	metanephric thick ascending limb development
				GO:2000024	regulation of leaf development	GO:0072237	metanephric proximal tubule development
				GO:2000039	regulation of trichome morphogenesis	GO:0072240	metanephric DCT cell differentiation

Supplementary Table 6. List of SMART conserved protein domains in neural transcripts.

	Domain	<i>D. melanogaster</i>	<i>H. sapiens</i>	<i>S. purpuratus</i>	<i>O. esmarki</i>
Achaete	Helix	NARERNRVKQVNNGFSQLRQ HIPAAVIADLSNGRRGIGPGA	NERERNRVKLVNLGF ATLREHVPGAANKK	NERERNRVKLVNHGFAN LRQQLPNGANNKKMSKV	NERERNRVKLVNMGFANL RQQLPNGVNNKKMSKVET

	loop helix domain	NKKLSKVSTLKMAVEYIRRL QKVLHE	MSKVETLRSAVEYIR ALQQLLDE	ETLRSAVSYIRQLQLLDE	LRSAVEYIRQLQTLLE
Arnt	Helix loop helix domain	CEIERRRRNKMTAYITELSDM VPTCSALARKPKDLTILRMAV AHMKALRGTGNT	SEIERRRRNKMTAYIT ELSDMVPTCSALARK PDKLTILRMAVSHMK SLRGTGNT	SEIERRRRNKMTAYITELS DMVPSCSALARKPKDLTI LRMAVSHMKSLRGTGNT	SEIERRRRNKMTAYITELSD MVPTCSALARKPKDLTILR MAVSHMKSLRGTGNT
Emx	Homeodomain	PKRIRTAFFSPSPLLKLEHAFES NQYVVGAERKALANLNLE TQVKVWFQNRRTKHKRMQQ ED	PKRIRTAFFSPSPLLRL ERAFEKNHYVVGAEER KQLAGLSLSETQVK VWFQNRRTKYKRQK LEE	PKRIRTAFFSPSPLLRL ENA FEKNHYVVGAEERKQLAA SLNLTETQVKVWFQNRRT KYKRKRIKSEE	PKRIRTAFFSPSPLLRL EQAF EKNHYVVGAEERKQLAASL NLTETQVKVWFQNRRTKY KRKIAEE
Engrailed	Homeodomain	EKRPRTAFFSSEQLARLKREFN ENRYLTERRRQQLSSELGLNE AQIKIWFQNKRAKIKKSTGSK	DKRPRTAFTAELQRLK LKAEFQANRYITEQR RQTLAQELSLNESQIK IWFQNKRAKIKKATGI K	EKRPRTAFFSASQLQRLKQ EFQSQNYL TEQRRRALAK ELTSESQIKIWFQNKRAK IKKATGLK	EKRPRTAFFSAAQLQRLKQE FQSQNYL TEQRRRGLAKEL KLNESQIKIWFQNKRAKIK KASGVK
FoxD	Forkhead	LVKPPYSYIALITMAILQSPHK KLTLSGICDFIMSRFPYKDK FPAWQNSIRHNLNLCDFIKV PREPGNPGKGNFWTLDPLAE DMFDNGS	LVKPPYSYIALITMAILQSPKRLTLSEICEFI SGRFPYREKFPWQ NSIRHNLNLCDFIKV PREPGNPGKGNFWTLDPEADMFNDNGS	SVKPPYSYIALITMSILQSP QKRLTLSGICEFIMNRFY YREKFPVWQNSIRHNLNLCDFIKV PREPGNPGKGNFWTLDPEADMFNDNGS	LVKPPYSYIALITMSILQSP QKRLTLSGICEFIMNRFY YR EKFPVWQNSIRHNLNLCDFIKV PREPGNPGKGNFWTLDPEADMFNDNGS
FoxJ1	Forkhead	HVKPPYSYATLICMAMQASK ATKITLSAIYKWITDNFCYFR HADPTWQNSIRHNLNLCDFIKV KVPREKDEPGKGGFWRIDPQ YAEERLLSGA	HVKPPYSYATLICMAMQASKATKITLSAIYKWITDNFCYFRHADPT WQNSIRHNLNLCDFIKV KVPREKDEPGKGGFWRIDPQYAEERLLSGA	SIKPPYSYATLIWMAMKE SKKHKITLSSYKWITENF KYVQVADPSWQNSIRHNLNLCDFIKV KVPREKDEPGKGGFWRIDPAHADELENGV	YVKPPYSYATLIWMAMKDK SKKNKITLSAIYKWITSNFK YVQVADPSWQNSIRHNLNLCDFIKV KVPREKDEPGKGGFWRIDPAHA
FoxM	Forkhead	RLHVSNIPIFRDPDLRAMFG QFGTILDVEIIFNERGSKGFGF VTFANSNDAERARERLHGTVEGRKIEVN	SERPPYSYAMAMQFAI NSTERKRMTLKDITWIEDHFPYFKHIAKPG WKNSIRHNLNLCDFIKV VRETSANGKVSFWTI HPSANRYLTLD	KERPPYSYSSLIQFAISSAP EGKTLRDVYFWIETHFP YFRTAKLGWKN SIRHNLNLCDFIKV VREAPSPPGQPAFWTLRPGTVVRLPERKV	HERPPYSYSSLIQFAISTAPS GRMTLREIYHWIEIHPYFR TAKLGRWNSIRHNLNLCDFIKV FIREPPVGHGQPAFWTLRPGTVVRLPEK
Glia	Actin depolymerization factor	NEVLEELKKFRFSKSKNNAAL ILKVDREKQTVLDEFIDDISV DELQDTLPGHQPRYVIYTYK MVHDDQRISYPMCFIFYPTRD SQIELQMMYACTKSALQREV DLTRVYEIRELDELTEEWLKA KLK	MKVDKDRQMVVLEE EFQNSPEELKMEPE RQPRFVVYSYKYVHD DGRVSYPLCFIFSSPV GCKPEQQMMYAGSK NRLVQTAELTKVFEIR TTDDLTEAWLQEKLSFER	PEIDELVKKFRFRKEKNN AAVLKIDTSRLMVILDEQ YEDMTPDELQELPEHLR RYVLYSYCRHDDGRISY PLCFIFHPQGTKAELAM MYSGCITLHRRTGITKYF ELSDLEEFTEEWLKKKLG	EELKAKLKKFRFRKEKTNA AIVKIDKETQKVIQDDSFEE EEDLEGMDPETLSDELPAH VPRFVAYSICYHDDGRIS YPLVLIHCAPAGCSTELQV MYAGSRNNLVNEAKMTK VFEVRNPEEITEEWLKKSLA
Hes	Helix Loop Helix Domain	GVIEKRRRDRINSSLTTELKRL VPSAYEKQGS AKLEKAEILQLTVEHLKSLQSKTLD	PIMEKRRRARINESLS QLKTLILDALKKDSR HSKLEKADILEMTVK HLRNLQRAQMT	HLTERKRRARINDSLLQL KSMVFPVIKKDISRHPKM EKADILEMTVRYLKDVTPEQG	HLMERKRRARINDSLLQLK SLVFPVTRKEIDRHPKLEK ADILEMTVRHIQELQKHSNA
NeuroD1	Helix Loop Helix Domain	NDRERNRMHNLNDALEKLR VTLPSLPEETKLTKEILRFAH NYIFALEQVLES	NARERNRMHGLNAA LDNLRVVPCYSKTQ KLSKIETLTAPALPLT DPSARRSASMATSLS NTNHP	NDRERNRMHNLNDALEKLR VTLPSLPEETKLTKEILRFAH NYIFALEQVLES	NDRERNRMHNLNDALEKLR VTLPSLPEETKLTKEILRFAH NYIFALEQVLES
Ngn	Helix Loop Helix Domain	NDRERNRMHNLNDALEKLR VTLPSLPEETKLTKEILRFAH NYIFALEQVLES	NDRERNRMHNLNDALEKLR VTLPSLPEETKLTKEILRFAH NYIFALEQVLES	NDRERNRMHNLNDALEKLR VTLPSLPEETKLTKEILRFAH NYIFALEQVLES	NDRERNRMHNLNDALEKLR VTLPSLPEETKLTKEILRFAH NYIFALEQVLES

Onecut2	CUT DNA-binding domain	ANSSDMEEINTKDLAQRISAE LKRY SIPQAIFAQRVLCRSQG TLDLLRNPKPWSKLSGRET FRRMWKWLQEPEFQRMSALR MAA	ATSGQLEEINTKEVA QRITAE LKRY SIPQAIF AQRVLCRSQGTLSDL LRNPKPWSKLSGRET TFRRMWKWLQEPEF QRMSALRLAA	DGGQAGEEINTKEIAARV TSELKRY SIPQAVFAQRV LCRSQGTLSDLLRNPKPW SKLSGRETFRRMWKWL QEPEFQRMSALRLAG	DGANAGEEINTKEVAARV TSELKRY SIPQAVFAQRV CRSQGTLSDLLRNPKPWSK LKSGETFRRMWKWLAEP EFQRMSALRLAA
Otp	Homeodomain	QKRHRTRFTPAQLNELERCFS KTHYPDIFMREEIAMRIGLTS RVQVWFQNRRAKWKRRKKT T	QKRHRTRFTPAQLNELER LERSFAKTHYPDIFMR EELALRIGLTSERVQV WFQNRRAKWKRRKKT TT	QKRHRTRFTPAQLNELER NFAKTHYPDIFMREEIAM RVGLTESRVQVWFQNR AKWKRRKKT	QKRHRTRFTPAQLNELERN FAKTHYPDIFMREEIAMRV GLTESRVQVWFQNRRAK KWKRRKKT
Otx	Homeodomain	QRRERTTFTRAQLDVLEALFG KTRYPDIFMREEVALKINLPE SRVQVWFKNRRACRQQLQ QQ	QRRERTTFTRSQLDV LEALFAKTRYPDIFMR EEVALKINLPESRVQV WFKNRRACRQQQQ SG	QRRERTTFTRAQLDVLET LFSRTRYPDIFMREEVAM KINLPESRVQVWFKNRR KCRQQQQQ	QRRERTTFTRAQLDVLEAL FSKTRYPDIFMREEVALKIN LPESRVQVWFKNRRACR QQAQQQ
Rx	Homeodomain	HRRNRTTFTTYQLHELERA FEKSHYDPVYSREELAMKVNLP EVRVQVWFQNRRAKWRQ KSE	HRRNRTTFTTYQLHE LERAFEKSHYDPVYS REELAGKVNLPESRV QVWFQNRRAKWRQ EKLE	HRRNRTTFTTYQLHELER AFEKSHYDPVYSREEL KVNLPESRVQVWFQNR AKWRQEKME	HRRNRTTFTTYQLHELERA FEKSHYDPVYSREELAIK VNLPESRVQVWFQNRRA KWRQEKME
Six3	Homeodomain	EQKTHCFKERTSLLREWYL QDPYPNPTKKRELAQATGLN PTQVGNWFKNRRQRDR AKN	EQKTHCFKERTSLL REWYLQDPYPNPSK RELAQATGLTPTQV NWFKNRRQRDR AKN	EQKTHCFKERTSLLREW YLQDPYPNPTKKRELAQA TGLTPTQVGNWFKNRR RDRAAAKN	EQKTHCFKERTSLLREWY LQDPYPNPTKKRELAQATG LTPTQVGNWFKNRRQRDR AAAAKN
SoxC	High mobility group	HIKRPMAFMVWSQMERRKI CERTPDLHNAEISKELGRRWQ LLSKDDKQPYIEAEKLRK LH MIEYPNYKY	HIKRPMAFMVWSKI ERRKIMEQSPDMHNA EISKRLGKRWKMLKD SEKIPFIREAERLRK LH MADYPDYKY	HIKRPMAFMVWSQIERR RIMETTPDMHNAEISKRL GRRWKTLDVAKSPYVE EAERLRL LHMAQYPDYKY	HVKRPMAFMVWSQIERR KIMEQTPDMHNAEISKRLG RRWKTLDVAKSPYVEEA ERLRLHMQEFPDYKY
SoxB1	High mobility group	RVKRPMAFMVWSRGQRK MASDNPKMHNSEISKRLGAQ WKDLSESEKRPFIDEAKRLRA VHMKEHPDYKY	RVKRPMAFMVWSR GQRKMAQENPKMH NSEISKRLGAEWKVM SEAEKRPFIDEAKRLR ALHMKEHPDYKY	RVKRPMAFMVWSRGQ RRKLSQENPKMHNSEISK RLGAEWKLLSEDEKRPFI DEAKRLRA VHMKEHPDYKY	RVKRPMAFMVWSRGQR RKMAQENPKMHNSEISK LGAEWKLLTEEQKRPFI DEAKRLRAVHMKEHPDYKY
SoxB2	High mobility group	HIKRPMAFMVWSRGQRK MAQDNPKMHNSEISKRLGAQ WKDLSESEKRPFIDEAKRLR ALHMKEHPDYKY	RVKRPMAFMVWSR GQRKMAQENPKMH NSEISKRLGAEWKLLS ETEKRPFIDEAKRLRA LHMKEHPDYKY	HVKRPMAFMVWSRGQ RRKLAQENPKMHNSEISK RLGAEWKLLSEDDKRPFI DEAKRLRALHMKEHPDY KY	HVKRPMAFMVWSRGQR RKMAQENPKMHNSEISK LGAEWKLLTEEQKRPFI DEAKRLRALHMKEHPDYKY
SoxD	High mobility group	HIKRPMAFMVWAKDERRKI LKACPDHNSNISKILGARW KAMSADKQPYEEQSRLSK LHMEQHPDYRY	HIKRPMAFMVWAK DERRKILQAFDPMHN SNISKILGSRWKSMSN QEKQPYEEQARLSK IHLEKYPNYKY	HIKRPMAFMVWAKEER RKILARHPDMHNSNISKIL GSKWKTMSNAEKQPYEE EQARLSKAHLEKYPDYK Y	HIKRPMAFMVWAKEERR KILARHPDMHNSNISKILGS KWKTMSNAEKQPYEEQA RLSKAHLEKYPDYKY
Tailless	c4 zinc finger	HVPCVKCRDHSSGKHYGIYA CDGCAGFFKRSIRRSRYVCK SQKQGLCVVDKTHRNQCRAC RLRKCCEVGMNKD	DIPCKVCGDRSSGKH YGVYACDGCAGFFKRS SIRRNRTYVCKSGNQ GGCPVDKTHRNQCR ACRLKCKLEVNMNK D	DIPCKVCGDRSSGKHYGV YACDGCAGFFKRSIRRN TYVCKNRSGGPCPVDK THRNQCRACRLKCKLQV DMNKD	DIPCKVCGDRSSGKHYGVY ACDGCAGFFKRSIRRNRTY VCKNRNSGPCPIDKTHRNQ CRA CRLKCKLQVDMNKD
Zic2	Zinc finger	KTCNKVHSMHEIVTHLTV E HVGGEPTTHACFWVGC SRN GRPFKAKYKLVNHIR VHTGEKPFACHPGCGKVF ARSENLIK IHKRTHTGK PFKCEHEGCDR RFANSS DRKHHSHVHTSDKP YNCRINGCDKSYTHPSS LRKH MKVH	LSNPKKSCNKTFTSTM HELVTHSVVEHVG GGPEQSNHVCWEECPRE EQKPFKAKYKLVNHIR VHTGEKPFACHPGCGK VFARSENLIK IHKRTHT GKPFKCEHEGCDR RFAN SSDRKHHSHVHTSDKP YNCRINGCDKSYTHPSS LRKH MKVH	LSCLWIDQDLPEPRKPCN KTFTTMHEIVTHITVEHV GGPEQTNHTCFWQNC SR EQKPFKAKYKLVNHIR VHTGEKPFACHPGCGK VFARSENLIK IHKRTHT GKPFKCEHEGCDR RFAN SSDRKHHSHVHTSDKP YNCRINGCDKSYTHPSS LRKH MKVH	LSCLWIDQEQEPRKPCNK TFTTMHEIVTHITVEHV GGPEQTNHTCFWQNC SREQPFKAKYKLVNHIR VHTGEKPFACHPGCGK VFARSENLIK IHKRTHT GKPFKCEHEGCDR RFAN SSDRKHHSHVHTSDKP YNCRINGCDKSYTHPSS LRKH MKVH

Of the 28 neural transcripts, Glass, Hbn, Hox7, and DLX have been omitted since they were unable to be successfully identified. Sip1 and Myt1 also failed to have a commonly occurring conserved domain across all four species, so the full sequence was used for alignment, taking it out of Table 4. For sequences highlighted in yellow, *Mus musculus* was used as an alternative outgroup and highlighted in blue is the alternative outgroup, *Pan troglodytes*.