23

Comparative Genomic Analysis of Geoduck Clam (*Panopea generosa*)

https://doi.org/10.5281/zenodo.7996007

Abstract

Pacific geoduck (*Panopea generosa*) is a subtidal clam species with a range from Alaska to Baja California. In this study, five transcriptomic libraries from three tissue types (gonad, heart, ctenidia) and two different life stages (larvae, juvenile) were assembled and annotated with biological ontological information. A particular emphasis in this study were reproductive genes and the gonad library, as the most likely location for reproductive candidate genes for gene editing. In addition, a comparative genomic approach was used to look for homologous genes across the *Venerida* clade. This effort represents an establishment of an important genomic resource for Pacific geoduck that will be valuable in the improvement of sustainable aquaculture.

---

1. Introduction

1.1 Geoduck Biology, Ecology and Economic Importance

Farming of geoduck clams, *Panopea generosa*, in the cold, nutrient-rich, and clean waters of the Pacific Northwest is a long-standing tradition and important cultural, economic and ecological part of the coastal communities, (Feldman *et al.,* 2004). Geoduck clams live deep in the intertidal zone (-2.0 feet and below) but have been observed as deep as 360 feet (WDFW, 2023). Geoduck clams are large, long-lived and fecund. They are reproductive as males as early as 2 years old and are often considered market size at 5 years old. In the wild, geoducks are found between 18-80 feet deep and take 15 years to reach full maturation (~7 lbs). The oldest recorded living geoduck was 173 years old and can be sexually mature up to 50 years (Edge *et al.,* 2021). Due to their long-lives and sexual asynchronicity, they have a low effective population size (Vadopalas, Davis & Friedman, 2015).

Geoduck clams are an increasingly important fishery and aquaculture product for the Eastern Pacific coast of the US from Baja California to Alaska. The geoduck industry consists of a small number of private operators committed to harvesting, processing, and marketing their product. Geoduck meat is sold primarily outside of the US; siphon meat goes to Japan and Taiwan while body meat is sold in California and the East Coast, (Cheney & Mumford 1986). Geoduck aquaculture is considered the most economically important clam fishery in North America (Hoffman et al., 2000), bringing in $24.5 million in sales and over $1 million in state revenue in 2013 (Washington Sea Grant, 2015). Geoduck aquaculture also supports local oyster farming as well due to its high price per square acre. Recent evidence also suggests that geoduck aquaculture gear can support the recovery of the threatened cockle species in Washington State (Dimond et al., 2022).

Geoducks are primary consumers of phytoplankton by filter feeding. As filter feeders, geoducks provide essential ecosystem services as well by removing algae, organic matter and excess nutrients from the water column (Cubillo, Ferreira, Pearce et al., 2018). In addition, when geoducks are harvested excess nutrients such as nitrogen and phosphorus are removed from the marine ecosystem. Geoducks are a

keystone species in the subtidal zones by removing organic matter from the environment and providing a food source for the declining populations of sea otters and crabs.

Wild populations have been threatened by overharvesting and poaching (KUOW, 2015). Due to concerns of genetic mixing between hatchery and wild populations, most aquaculture of geoduck is done using wild broodstock, but the potential for genetic mixing still exists. Farmed geoducks may become reproductively mature as early as two years old (Vadapolas et al. 2015). Due to these concerns, a key area of research focuses on the reproductive biology of geoduck clams. The use of triploid geoduck clams may help alleviate reproductive maturation in farmed clams, akin to research done on ploidy in Pacific oysters (Allen & Downing, 1986). Due to issues with triploidy, there is a need to better understand the reproductive genes responsible for sexual maturation to provide resources for future work developing sterility approaches. This work could include the development of gene knock-down strategies.

The main objective of this study is to build annotated reference transcriptome libraries. There were three specialized tissue types: gonad, ctenidia, and heart. There were two tissue samples from pooled larvae and pooled juveniles as well. In addition, this study also leveraged tools and resources from previously published genome and transcriptome studies on clams from Venerida.

1.2 Geoduck Genomics and Proteomics

The publication of a fully annotated juvenile *P. generosa* reference genome by Putnam et al., (2022), along with previous studies focusing on geoduck genomics and gene expression in response to environmental stress, contributes to a comprehensive understanding of geoduck and mollusk genomics while shedding light on the role of DNA methylation in environmental acclimatization. Previous work on geoduck genomics focuses on gene expression in response to environmental stress. Work on the *P. globose* juvenile transcriptome exposed to chronic and acute thermal stress demonstrated that there were similar gene expression patterns between stress and non-stressed animals, (Juarez *et al.,* 2018). In the same study there was also a high degree of expression genes related to DNA repair and transcription regulation in chronically exposed juveniles where protective genes against oxidative stress were highly expressed in acutely exposed juveniles. Timmins-Schiffman *et al.,* (2017), published the first proteomic study of three maturation stages in males and female geoduck clams using gonad proteins. They showed that gonad proteins became increasingly divergent between males and females as maturation progressed. Spencer et al., (2019) investigated sex-specific broodstock response and differential gene expression in *P. generosa* in response to low pH. Temperature and dissolved oxygen increases corresponded to differences in protein abundance patterns such as heat shock protein 90-α. In larvae, Timmins-Schiffman *et al.,* (2020) looked at the proteomics of larval *P. generosa* with ciliate infection to investigate the molecular underpinnings of the innate immune response of the larvae to a pathogen. Ciliate response proteins included many associated with ribosomal synthesis and protein translation, suggesting the importance of protein synthesis during larval immune response. In juvenile *P. generosa,* Gurr *et al.,* (2020; 2022) conditioned the animals before testing them with elevated pCO2 (~2400 utam). Following the secondary exposure, neither elevated or ambient pCO2 altered juvenile respiration rates, indicating ability for metabolic recovery under subsequent conditions.

Recently, Putnam *et al.,* (2022), published an annotated *P. generosa* reference genome as part of a larger common-garden ocean acidification study. They looked at the role of DNA methylation on environmental acclimatization. Functional enrichment analysis of differentially methylated genes revealed regulation of signal transduction that influences cell growth, proliferation, tissue and skeletal formation,

and cytoskeletal change. Putnam's work, as well as this study, will greatly aid in the collective understanding of not only geoduck genomics but overall mollusks genomics. In this study, five RNA-seq transcriptome libraries from three geoduck tissue types (gonad, heart, ctenidia) and two different life stages (larvae, juvenile) were assembled and annotated with biological Gene Ontology information. A particular emphasis in this study were reproductive genes and the gonad library, as the most likely location for reproductive genes.

1.3 Comparative Species Genomics

In addition to the five geoduck RNA-seq libraries described in this study, there is also value in leveraging publicly available data for a comparative clam species transcriptome study. Mun *et al.,* (2017), published a transcriptome of the Manila clam (*Ruditapes philippinarum*) as part of a greater effort in selective breeding and disease control. They reported 41,275 annotated sequences in the *de novo* whole transcriptome assembly of *R. philippinarum* across three different tissues (foot, gill and adductor muscle). Wang *et al.,* (2016), was the first to publish an annotated transcriptome of the hard clam *Mercenaria mercenaria*. It was part of the work investigating the parasite QPX in hard clams. A *de novo* assembly was constructed and a consensus transcriptome of 62,980 sequences were functionally annotated. A total of 3,131 transcripts were identified as differentially expressed in healthy versus infected tissues. Comparative analysis of annotated genes can reveal the conserved molecular mechanisms between mollusks, such as genes with high homology expressed across the Venerida clade.

Genome resources for clams in the Venerida clade are more abundant than transcriptome or proteomic resources. A reference genome is available in *R. philippinarum* (Mun *et al.,* 2017) and *M. mercenaria* (Wang *et al.,* 2016) as part of the research mentioned above. In order to leverage even more clam genomic resources, the genomes of *Spinsula solida*, *Mactra quadrangularis* and *Archivesica marissincia* were also compared to *P. generosa* for functional analysis. The assembly of the surf clam, *S. solida*, was based on Hi-C data generated as part of the Darwin Tree of Life Project. The other surf clam, *Mactra quadrangularis* (or *Mactra veneriformis*), also has a recently assembled genome as part of the efforts of Sun *et al.,* (2022). Low natural yields in *M. quadrangularis* in China lead to this recent effort to better understand surf clam genomic resources. Using Hi-C assembly, a total of 29,315 protein-coding genes were predicted. From this study, a genome-level phylogenetic tree was constructed demonstrating that *M. quadrangularis* and *R. philippinarum* diverged around 231 million years ago. In the Northern quahog clam, *Mercenaria mercenaria*, Farhat *et al.,* (2022) published the first publicly available genome. Due to high environmental variability on the East Coast of the US, and a desire to understand mass mortality events, the genome of *M. mercenaria* needed to be assembled. Genome annotation yielded 34,728 predicted protein-coding genes, the most of all the Venerida so far. Using these previously published genomic resources by running a comparative genomic analysis, will provide an important resource for future comparative work.

1.4 Reproductive Genes

Limited studies in bivalve genomics have investigated sexual maturation through differentially gene expression in various tissue types. One study, Dheilly *et al.,* (2012), looked at the basis of sex differentiation in Pacific oysters (*Crassostrea gigas*) using a microarray assay. Gene expression was studied in the gonad over a yearly reproductive cycle. There were 2,482 genes found to be differentially expressed between males and females during gametogenesis. The expression of 434 genes could be

localized to the germ cells or somatic cells of the gonad and between the sexes. Maturation analysis processes like this study can reveal the conserved and diverged genes between males and female gonads.

1.5 Summary

Transcriptomics is an important field of study that provides insight into the complex gene expression patterns of various organisms. Comparative transcriptomics allows for a deeper understanding of the differences and similarities in gene expression between different species. In addition to a functional annotation of the geoduck transcriptome, the focus on this study will be to investigate and characterize five different geoduck tissue types (gonad, heart, ctenidia, larvae and juvenile). We compared the transcriptomes of Manila clams (*R. philippinarum*), Mercenaria clams (*M. mercenaria*) and Pacific oysters (*Crassostrea gigas*), against the geoduck (*P. generosa*) transcriptome, focusing on the most commonly expressed and overexpressed Gene Ontology (GO) terms and genes. Genomes of five clam species were also compared to the *P. generosa* genome, looking for genes with high homology. As more clam species are sequences and genomes assembled, the overall gap in knowledge will decrease and more functional applications for aquaculture can be developed.

---

2. Methods

2.1 Genome Annotation

There is a reference genome of juvenile *P. generosa* recently published by Putnam *et al.,* (2022). They used the Proximo Hi-C process (Phase Genomics) resulting in 18 chromosome scaffolds containing 1.42 Gpb of sequence (64.53% of the corrected assembly). Juicebox correction resulted in a scaffold N50 of 57,743,597 bp. Genome annotation identified 34,947 genes and 236,960 coding sequence regions which corresponds to 38,326 mRNA features. Genome feature tracks included genes, exons, introns, repetitive sequences, and CG motifs (Roberts *et al.,* 2020). Annotation yielded 16,899 tRNAs with a mean and median length of 75 bp in the range of 53-314 bp. CG content was determined to be 33.78% and a total of 15,712,294 CG motifs are present in the genome. The assembled genome is available on the National Center for Biotechnology Information website (NCBI) under GCA_902825435.1. Sequences were annotated by comparing contiguous sequences to the UniProtKB/Swiss-Prot database (http://uniprot.org) using the BLASTn algorithm (Altschul *et al.,* 1997) with a 1.0E-20 e-value threshold. Based on the Swiss-Prot values, there were 14,672 protein coding sequences in the *P. generosa* genome that had gene ontology characterization information such as GO enrichment analysis.

2.2 Library Construction and Sequencing

Total RNA was extracted from adult, juvenile and pooled larvae of *P. generosa*. The adult tissue was isolated using the PAXgene Tissue RNA Kit (Qiagen) based on manufacturer's instructions. The adult tissue was separated into three different types by function: gonad, ctenidia and heart. Five RNA-seq libraries were constructed from pooled mRNA and sequenced at the University of Washington High Throughput Genomics Unit (HTGU) on the Illumina Hi-Seq 2000 platform (Illumina, San Diego, CA, USA). Each library was run on a single lane. Raw sequence reads were quality trimmed using Trim Galore v0.4.0, and the sequence data was quality assessed using FastQC (Andrews, 2010).

2.3 Geoduck Sequence Analysis

Sequences were annotated by comparing contiguous sequences to the UniProtKB/Swiss-Prot database (http://uniprot.org) using the BLASTn algorithm with a 1.0E-20 e-value threshold. Genes were then classified according to their biological processes that were determined by their Gene Ontology (GO) information and are classified into one or more of 72 parent categories (GO slims). The full dataset is available in Supp. Table 1. Genes were classified into their RNA-seq library (ctenidia, gonad, heart, juvenile or larvae) and any gene with a transcript per million (tpm) greater than zero was removed from further analysis. Gene Ontology terms were then characterized relative to all Gene Ontology terms (GO slims) present in the *P. generosa* genome. A new term was calculated by taking the proportion of a single GO slims present in an RNA-seq library over the proportion of that same GO slim present in the entire *P. generosa* gene set. We coined this new term "Gene Ontology Proportional Value" where, for example, when the proportional value equals one then that GO slim in the RNA-seq library is representative to the GO slim in the entire gene repertoire.

2.4 Comparative Species Genomics

M. mercenaria and *R. philippinarum* transcriptomes were annotated by comparing sequences to the entire gene list of *P. generosa*, and given Gene Ontology Proportion Values for inter-species comparison. Transcriptome libraries were downloaded off of NCBI. (*R. philippinarum* GenBank accession number: GCA_026571515.1 and *M. mercenaria* accession number: GCF_021730395.1). Gene lists were annotated by using BLASTn with E-value of 1E-20 and associated Gene Ontology terms classified and counted by transcriptome library (either *M. mercenaria* or *R. philippinarum*). Gene Ontology terms were then classified using the same process as the geoduck sequence analysis above. Instead, the Gene Ontology Proportional Value was calculated using the proportion of GO slims present in either *M. mercenaria* or *R. philippinarum* transcriptome libraries relative to the proportion of that same GO slim present in the entire *P. generosa* transcriptome library.

In order to see if there was a functional difference between genes with high homology, five different clam genomes in the *Venerida* family were annotated. The clams genomes were *Archivesica marissinica* (GenBank: GCA_014843695.1), *Mactra quadrangularis* (GCA_025267735.1), *Mercenaria mercenaria* (GCF_021730395.1), *Ruditapes philippinarum* (GCA_026571515.1) and *Spinsula solida* (GCA_947247005.1). All five clam genomes were annotated by comparing contiguous sequences to the *P. generosa* gene database using the BLASTn algorithm with a 1.0E-20 e-value threshold.

2.5 Characterization of Reproductive Genes

Characterization of reproductive genes in *P. generosa* was done in two ways. The first way was to gather a list of genes expressed in the gonad. This produced a list of reproductive genes in the *P. generosa* adult gonad that code for proteins related to the "reproductive process" functions (Supplemental Table 2).

Dheilly *et al.* (2012), investigated the temporal variation of gene expression during oyster gonad differentiation and development in *C. gigas*. The genes identified in their oyster study are differentiated by sex and stage of development: somatic tissues and oocytes. Differentially expressed oyster gonad and oocyte genes were pulled from Supp. Table 3 from Dheilly *et al.,* (2012). Genbank accession numbers were gathered from clusters 1-10 and were annotated against the *P. generosa* gene database using BLASTn algorithm with a 1.0E-10 e-value threshold. The focus of this study will be the female

reproductive genes from developmental stage 0 to stage 3 and compared to the reproductive genes found in the RNA-seq gonad library.

---

3. Results

3.1 Geoduck Genomic Analysis

After leveraging the genomic resources from *Putnum et al, 2022* to compare to the Swiss-Prot database, a fully annotated genome linked transcriptome was produced. This genome has 34, 947 annotated protein coding sequences. Of those, 2,180 are expressed only in the juveniles. The RNA-seq library for larvae geoduck returned 19,449 genes with 868 genes found only in the larvae transcriptome library. In the heart library, there were 17,479 genes representing only 371 genes unique to the heart. In the ctenidia, there were 17, 479 genes representing 340 genes found only in the ctenidia tissue library. Most crucially for this study on reproduction, the gonad RNA-seq library had the fewest genes with only 13,682. Of those genes, 119 were found uniquely expressed in the gonad. The full annotation of the 5 RNA-seq libraries are found in Supp. Table 1. A pairwise comparison between RNA-seq libraries was produced as from the unique genes list above, represented as count of biological gene ontology terms (Supp. Figure 1).

Expanding out from gene characterization, the Gene Ontology Proportion Value is descriptive of the abundance of biological Gene Ontology processes per library relative to the entire *P. generosa* genome (Figure 1). In the *P. generosa* genome, there are 17, 611 GO slims representing 34, 947 genes. In the juvenile, there are 17, 277 GO slims representing 19, 449 genes. In the larvae, there are 16,632 GO slims representing 19,449 genes.In the heart there are 16, 021 GO slims representing 17, 602 genes. In the ctenidia, there are 15, 911 GO slims representing 17, 479 genes. Finally, in the gonad there are 14, 715 GO slims representing 13, 682 genes.

3.2 Comparative Species Genome analysis

Annotations from the *M. mercenaria* and *R. philippinarum* transcript libraries against the *P. generosa* gene list revealed a strong phylogenetic correlation between *P. generosa* and *M. mercenaria*. Of the 34,947 mRNA features in the *P. generosa* transcriptome, 5,099 (14.6%) were found in *M. mercenaria* transcriptome. In contrast, only 657 (1.8%) mRNA features were found in the *R. philippinarum* transcriptome, (Table 1). The transcript libraries of *M. mercenaria* and *R. philippinarum* were annotated with Gene Ontology Proportion Values. In *M. mercenaria* transcript libraries there are 7,027 GO slims while in *R. philippinarum* there are only 1,390 GO slims. The top 20 GO slim terms ranked by their Gene Ontology Proportional Values (Supp. Table 3). The most abundant GO slim categories were: anatomical structural development, signaling and cell differentiation, (Supp. Table 3). For example, of the 456 genes shared between *P. generosa* and *M. mercenaria,* the most abundant were related to anatomical structural development while only 67 genes related to anatomical structural development were shared between *P. generous* and *R. philippinarum*.

Annotations from other clam genomes, against the *P. generosa* genome, further highlighted genomic homology. The *M. mercenaria* genome is 1.9 Gb and the CG content was determined to be 34.5% (Table 2). Of the 34, 947 genes in *P. generosa*, 8,736 genes were matched in *M. mercenaria*, the most matches of all five genomes. The *R. philippinarum* genome is 1.4 Gb large and the CG content was

determined to be 32%. 2,263 genes found in *P. generosa* were matched in *R. philippinarum*. The *A. marissinica* genome is 1.5 Gb and the CG content was determined to be 39%. Of those genes in *P. generosa*, 3,268 were matched to *A. marissinica*. The *S. solida* is a smaller genome at 932 Mb. *S. solida* CG content was determined to be 35.5% and 1,629 genes were matched to the *P. generosa* genome. The *M. quadrangularis* annotated genome is the smallest of the five with a size of only 979 Mb, a CG content of 33% and only returning 874 gene matches to the *P. generosa* gene set.

3.3 Characterization of Reproductive Genes

Characterization of reproductive genes in *P. generosa* was done in two ways. The first way was to gather a list of genes expressed in the gonad. This produced a list of reproductive genes in the *P. generosa* adult gonad that coding for proteins related to the "reproductive process" functions (Supplemental Table 2). Of the 34, 947 genes in the *P. generosa* gene list, 640 genes are involved with the reproductive process (1.8% reproductive).

Leveraging Dheilly *et al.,* (2012) results from 32 individual gonad samples of *C. gigas*, they identified 2,482 genes differentially expressed between gametogenesis stages. Of those differentially expressed, 511 genes were found to be expressed in major expression stage 0 of development (neither male or female). Of those 511 oyster reproductive genes, 7 geoduck reproductive genes were found in the *P. generosa* genome. Key genes from stage 0 are Ttn, BMP2 and ZNF107. In the next developmental stage, major expression stages 1-3 for females, there were 197 genes differentially expressed in oysters. Of those 197 genes, 5 candidate genes were found in *P. generosa*. The only candidate gene from stages 1-3 is Rusf1. In the later developmental stages 2-3, Dheilly et al. 2012 found 312 candidate reproductive genes and of those 312, we found 13 candidate genes. Key candidate genes from this developmental stage are SMC5, Cep57, KCTD5, ATF7IP, and TPST1. In the final developmental stage 3, Dheilly *et al.,* (2012) reported 222 reproductive genes. Our annotation revealed 14 *P. generosa* reproductive genes including STOX1, SPPL3, Wdr20, and Rere. There were also two *P. generosa* genes found in cluster 9 (female and male differentially expressed gametogenesis stages) that were specific to the female gonad tissue. Those genes were SUMO3 and ARHGAP11A. The full results from this analysis are found in Table 3, including the *Dheilly et al 2012* cluster information as well as the results from the male reproductive candidate gene investigation.

Table 1. *NCBI BLASTn transcriptome annotation:* Transcriptomes are gathered off NCBI and blasted against the *P. generosa* (GCA_902825435.1). Total hits represent genes matched from *M. mercenaria* or *R. philippinarum* to the P. generosa gene list. *M. mercenaria* has almost x10 the amount of genes matched to *P. generosa* than *R. philippinarum.*

Species	total hits	GenBank Accession
Mercenaria mercenaria	5,099	GCF_021730395.1
Ruditapes Philippinarum	657	GCA_026571515.1

Table 2. *Comparative Genome Annotation Summary: M. mercenaria* has the largest shared genes (total hits) with *P. generosa* as well as the largest genome. *M. quadrangularis* has the fewest shared genes with *P. generosa*, as well as one of the smallest genomes.

Species	total hits	genome size (Mb)	GenBank Accession
M. mercenaria	8,736	1,900	GCF_021730395.1
M. quadrangularis	874	979	GCA_025267735.1
R. philippinarum	2,263	1,400	GCA_026571515.1
A. marissinica	3,268	1,500	GCA_014843695.1
S. solida	1,629	932	GCA_947247005.1

Table 3. *Characterization of Reproductive Development Genes: P. generosa* reproductive gene annotation summary using Dheilly *et al.,* (2012) gene clusters (1-10) and major expression stages. *P. generosa* hits are the genes matched from *C. gigas* reproductive candidate genes by either predominantly female or male expression stages.

Female Reproduction

Major Expression Stage	*P. generosa* hits	*Dheilly et al. 2012* cluster	Reproductive Candidate Genes
Stage 0	7	1	Ttn, BMP2, ZNF107
Stage 1-3	5	4	Rusf1, Foxl1
Stage 2-3	13	3	SMC5, Cep57, KCTD5, ATF7IP, TPST1
Stage 3	14	2	STOX1, SPPL3, Wdr20, Rere
Stage 1-3	7	9	SUMO3, ARHGAP11A

Male Reproduction

Major Expression Stage	*P. generosa* hits	*Dheilly et al 2012* cluster	Reproductive Candidate Genes
Stage 0	7	1	Ttn, BMP2, ZNF107
Stage 1-3	4	6	PDS5B, CREM, ELAVL2
Stage 3	8	5	Spag6

Stage 1-2	2	10	CSRP3, PCR3
Stage 1-3	7	9	CBX3, PIPOX, CBX3, mcm7
Stage 2-3	1	8	H2A. F/Z
Stage 3	13	7	DGKE

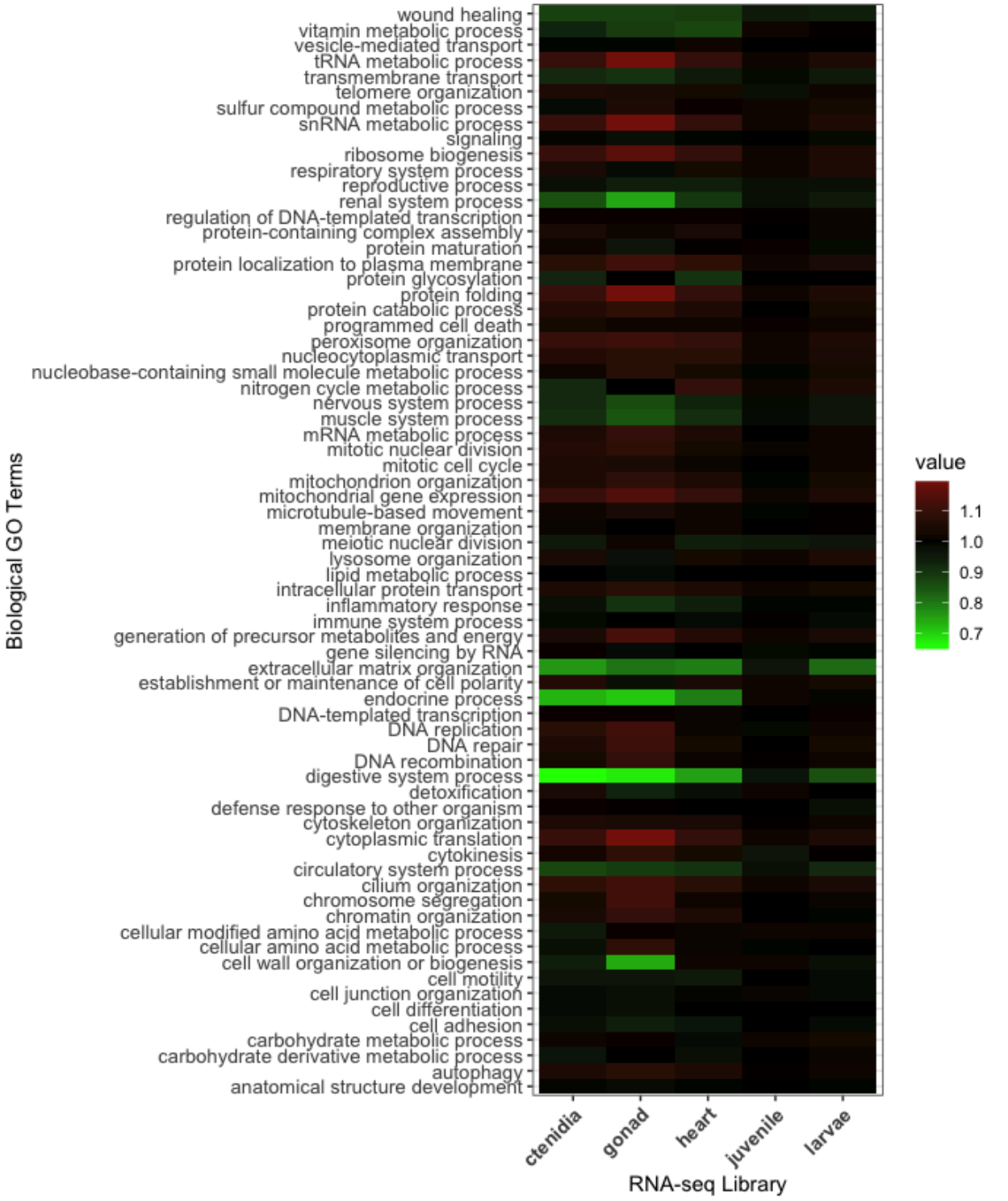


Figure 1: *Relative abundance of gene ontology terms compared to the entire geoduck genome.* Where value = Gene Ontology proportion value. Values highlighted in green or red represent an under or over abundant Biological Gene Ontology Term relative to the entire geoduck gene repertoire respectively.

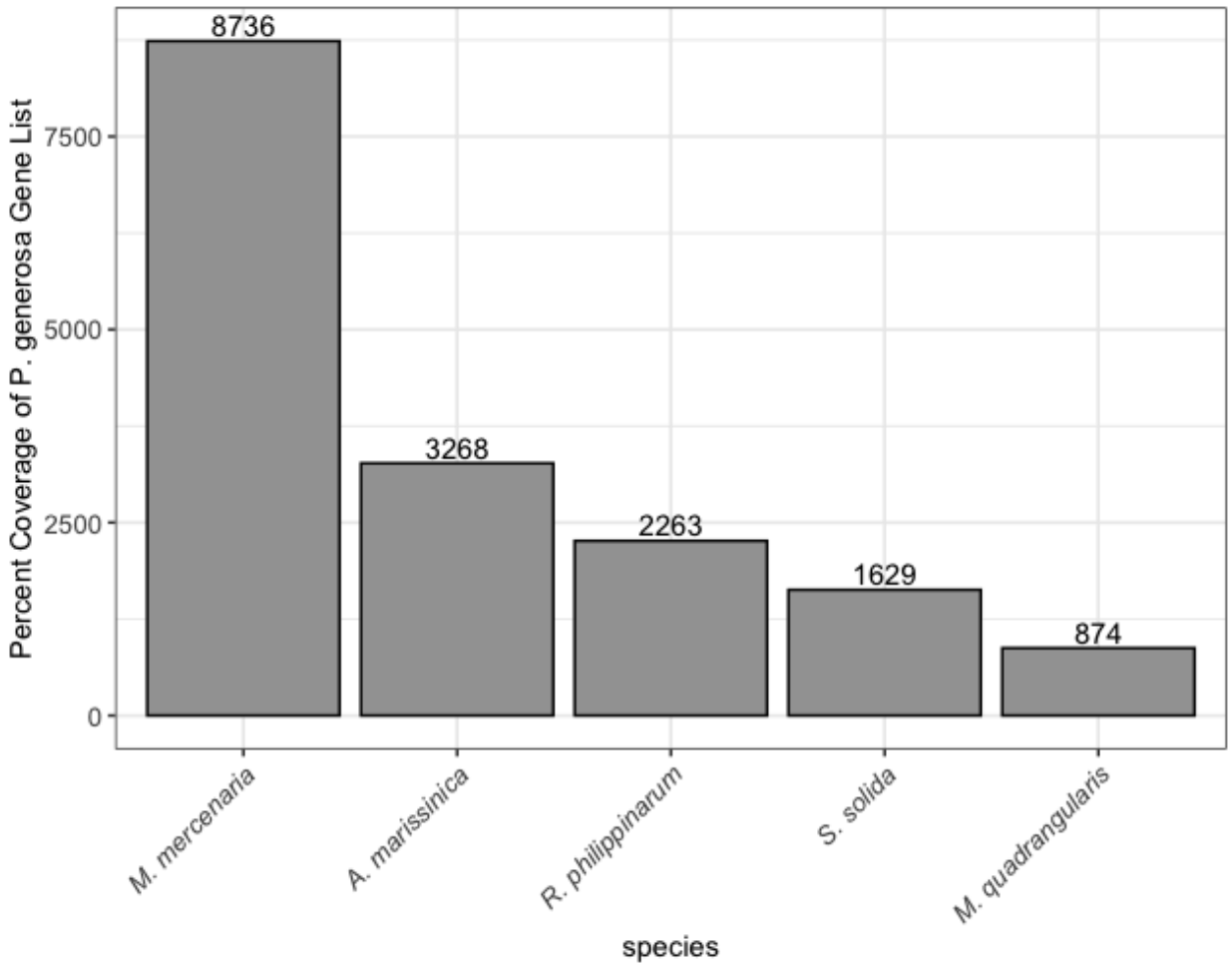


Figure 2: Percent Coverage of *P. generosa* Gene list represented in the 5 clam genome libraries of *Mercenaria mercenaria, Archivesica marissinica, Ruditapes philippinarum, Spinsula solida*, and *Mactra quadrangularis*

4. Discussion

4.1 *Geoduck Genomics*

The generation of five fully characterized RNA-seq libraries, provides comprehensive insights into gene expression patterns, ultimately advancing our understanding of Pacific geoduck reproductive biology and gene expression profiles. These transcript libraries are beneficial to describing individual tissue functions as the unique sets of genes found in each tissue type are consistent with their specialized functions. For example, the heart tissue had a unique set of genes involved in muscle and renal system processes, which are important for its role in circulation (Drake *et al.,* 2012). The ctenidia tissue had a unique set of genes involved in DNA repair, which is likely important for its role in filtering and cleaning the water that the geoduck lives in (Juárez *et al.,* 2018). The larvae library had a unique set of genes involved in nucleobase-containing small molecule metabolic processes and cytoskeleton organization. The

juvenile library was the largest with 67.01% genes expressed relative to the entire gene list of geoduck, representing almost all of the relevant biological processes (Figure 2).

We found that each tissue type exhibited a distinct set of expressed genes, indicating unique functional roles. These genes are categorized into their biological gene ontology (GO) processes by their Gene Ontology Proportional Value into over abundant or under abundant relative to the entire *P. generosa* genome. The most commonly overabundant GO processes distinct to all tissue types were related to metabolism, gene expressions and protein making, (Figure 1). Indicating that metabolism is more important at the specialized tissue level for geoduck. The most common under abundant GO processes were related to specialized system processes such as renal, endocrine, digestive and circulatory systems. This is unsurprising as digestive biological processes are going to be underexpressed in tissue types that don't specialize in digestion. Most of the overabundant GO processes unique to specific tissue types were functional to their specific needs. For example, in the gonad, the top overabundant GO processes were related to mitochondria and metabolism while in the larvae, the top overabundant GO processes were related to movement and energy allocation systems. For the heart, the top overabundant GO processes were related to nitrogen cycle metabolism, which would agree with its main function as a mechanical pump as well as its electrical activity (Drake *et al.,* 2012). For the ctenidia, the top overabundant GO processes were related to cytoplasmic translation. Cytoplasmic translation refers to the process by which mRNA molecules are translated into proteins in the cytoplasm of the cell, rather than in ribosomes. This process is important in many cells, including those in the ctenidia of oysters, because it allows for rapid and efficient protein synthesis, (Mclean & Whiteley, 1974). The juvenile tissue types did not have any highly overabundant genes due to their later developmental stage.

Investigating gene patterns between RNA-seq libraries is helpful for identifying tissue specific gene expression. One notable finding from our inter-library analysis is that ctenidia and juvenile tissues had the most genes expressed in these two tissues, not expressed in other tissues, between two tissue types, with a total of 665 shared genes. These two tissues share common functional roles, possibly related to growth and development. Conversely, gonad and ctenidia had the least shared genes with only 20, mainly related to tRNA metabolic processes, indicating that these tissues have very different gene expression profiles and that tRNA metabolism is a highly conserved suite of genes, (Galli, 1981). Overall, our results provide a foundation for future studies aimed at understanding the molecular basis of geoduck physiology and development. By identifying tissue-specific gene expression patterns, we can begin to unravel the complex molecular networks that underlie geoduck biology.

4.2 *Comparative Species Genomics*

Investigating the annotated transcriptome libraries shared between *P. generosa* and *M. mercenaria* or *R. philippinarum,* further illuminate foundational gene expression of *P. generosa*. Genes involved with translation, microtubule based movement, and cilium organization are all highly expressed in both the *M. mercenaria* and the *R. philippinarum* libraries. Interestingly, there is a unique group of conserved genes in *M. mercenaria* and *P. generosa*, but not found in *R. philippinarum*. The genes were related to metabolism and cellular organization. These differences, as well as *M. mercenaria* having more shared genes than *R. philippinarum*, may be due *M. mercenaria* being a closer phylogenetic relation to *P. generosa* than *R. philippinarum*, (Chen *et al.,* 2011). The inter-species analysis is valuable in determining gene orthologs important for reproduction investigations. Cross species library comparison is also useful for future

studies on gene function in marine bivalves outside of reproductive control, such as studies involving disease tolerance.

In the *P. generosa* genome from Putnam *et al.,* (2022), they found the geoduck genome to be almost 2 times larger in size than oyster genomes with twice as many putative chromosomes. This trend reversed when comparing *P. generosa* to other clam genomes. *R. philippinarum, M. mercenaria* and *A. marissinica* all have genomes larger than *P. generosa* and an additional chromosome. Interesting, regardless of genome size, the relative number of genes was approximately the same (~30,000). The GC content was also highly conserved across species with 32-33%, (except for *A. marissinica* at 39%). Comparing the percent coverage of *P. generosa* gene list to the query sequences of the other five clam genomes, *M. marissinica , R. philippinarum,* and *S. solida* have shared gene lists with *P. generosa* of 3268, 2263 and 1629 respectively. *M. mercenaria* has the most shared genes with *P. generosa* with 8736 and *M. quadrangularis* the least with only 874. (Figure 3). Looking at the phylogenetic relationship between *M. mercenaria* and *R. philippinarum* (Chen *et al.* 2011) reveals that *R. philippinarum* more basal than *M. mercenaria*. As more genomes are fully annotated and made available on NCBI, then comparative studies like this will be more robust.

4.3 *Characterization of Reproductive Genes*

A key emphasis of this study was to describe reproductive genes in geoduck. Our first approach, using the gonad RNA-seq library, produced a moderately large set of genes (n = 640) with gene expression patterns related to the reproductive process. This is also consistent with previous studies (Timmins-Schiffman, 2017) that have shown that reproductive tissues often have unique gene expression profiles. Leveraging the results of Dheilly *et al.,* (2012), provided a list of reproductive genes found in geoducks (74) that are linked to a major expression stage by sex and developmental stage. Looking at highly conserved reproductive genes across species gives us confidence that these are homologous genes for controlling reproduction in Bivalvia. Dheilly *et al.,* (2012), described reproductive genes related to mitosis and meiosis regulation including centromere proteins and kinesin related proteins. In *P. generosa* transcriptional dataset, reproductive genes related to centromere and kinesin related proteins were: Kinesin-like protein 6, KIF18A, and KIF9, inner centromere protein, centromere protein zw10 homolog, and centromere associated proteins S, E and X. Dheilly *et al.,* (2012) identified genes associated with the female specific processes such as oogenesis. Those included: *vitellogenin, cd63, mitotic apparatus, p62, forkhead box L2* and *caveolin*. In *P. generosa*, reproductive genes related to oogenesis were: *putative vitellogenin receptor (protein yolkless), forkhead box protein C1 and J3,* and *RecQ-mediated genome instability protein 1 (M-caveolin).*

Between the two approaches for identifying key reproductive genes in *P. generosa*, there are two genes found using both: Foxl1 and Cep57. The Fox genes, which code for forkhead class transcription factors, are classical orthologs involved in sex determination/differentiation, (Broquard *et al.,* 2021.) *Dheilly et al. 2012* found forkhead box L2 genes in *C. gigas.* Many studies on bivalves have found this gene to be involved in reproduction and have been found in the *C. gigas* and the pearl oyster *Pinctada fucata*, as well as other bivalves (Matusumoto *et al.,* 2013). The Cep57 gene, which codes for centrosome- and midbody-associated proteins, is not commonly studied for its implications in reproduction in bivalves. A similar ortholog, Cep55 has been documented to be involved in embryonic development in zebrafish (Jeffery *et al.,* 2015), and should be the focus of study in bivalve reproduction going forward. In

particular, genes related to vitellogenin, caveolin, foxhead class transcription factors, and Cep55/57 all appear to be highly conserved across Bivalvia and are very closely related to reproductive development.

5. Conclusion

Our results provide a foundation for future studies aimed at understanding the molecular basis of geoduck (*Panopea generosa*) physiology and development. Comparative analysis with *Mercenaria* clam and Manila clam transcript libraries provided insights into gene orthologs and conserved functions across species. Annotated RNA-seq libraries facilitated the identification of tissue-specific genes, including a significant number of previously undiscovered reproductive genes in geoduck. Specifically, Foxhead and Cep55/57 genes are key genes found. Each tissue type exhibited distinct sets of expressed genes, reflecting their specialized functions. Overabundant biological gene ontology (GO) terms in all tissue types were related to metabolism, gene expression, and protein synthesis, highlighting the importance of these processes at the tissue level. The study also focused on identifying reproductive genes for potential gene editing efforts, highlighting key genes involved in sex determination, oogenesis, and embryonic development. Overall, these findings lay the groundwork for future studies investigating geoduck physiology, development, and reproductive control, as well as broader investigations into gene function and disease tolerance in marine bivalves. By identifying these tissue-specific gene expression patterns, we can begin to unravel the complex molecular networks that underlie geoduck biology.

Work Cited

Washington Department of Fish and Wildlife (2023). Geoduck Clam (*Panopea generosa*). *WDFW*. https://wdfw.wa.gov/species-habitats/species/panopea-generosa#conservation

Edge, D. C., Reynolds, D. J., Wanamaker, A. D., Griffin, D., Bureau, D., Outridge, C., Stevick, B. C., Weng, R., & Black, B. A. (2021). A Multicentennial Proxy Record of Northeast Pacific Sea Surface Temperatures From the Annual Growth Increments of *Panopea generosa*. Paleoceanography and Paleoclimatology, 36(9), e2021PA004291. https://doi.org/10.1029/2021PA004291

Vadopalas, B., Davis, J. P., & Friedman, C. S. (2015). Maturation, Spawning, and Fecundity of The Farmed Pacific Geoduck *Panopea Generosa* in Puget Sound, Washington. Journal of Shellfish Research, 34(1), 31–37. https://doi.org/10.2983/035.034.0106

Cheney, Daniel P., Thomas F. Mumford, J. (1986). Shellfish and Seaweed Harvests of Puget Sound. The University of Washington Press.

A Hoffmann, A Bradbury, and CL Goodwin, Modeling geoduck, *Panopea abrupta* (Conrad, 1849) Population Dynamics. I. Growth: Journal of Shellfish Research [J. Shellfish Res.], Vol. 19, No. 1, pp. 57-62, Jun 2000.

Washington Sea Grant. (2015). Shellfish Aquaculture in Washington State. December 1–84. https://wsg.washington.edu/shellfish-aquaculture

Dimond, J. L., Crim, R. N., Unsell, E., Barry, V., & Toft, J. E. (2022). Population genomics of the basket cockle *Clinocardium nuttallii* in the southern Salish Sea: Assessing genetic risks of stock enhancement for a culturally important marine bivalve. Evolutionary Applications, 15(3), 459–470. https://doi.org/10.1111/eva.13359

Cubillo, A. M., Ferreira, J. G., Pearce, C. M., Marshall, R., Cheney, D., & Hudson, B. (2018). Ecosystem Services Of Geoduck Farming in South Puget Sound, USA: a modeling analysis. Aquaculture International, 26(6), 1427–1443. https://doi.org/10.1007/s10499-018-0291-x

KUOW - Poaching 100-Year-Old Geoducks For Big Money. (2015). Retrieved June 7, 2023, from https://www.kuow.org/stories/poaching-100-year-old-geoducks-big-money/

Allen, S. K., & Downing, S. L. (1986). Performance of triploid Pacific oysters, *Crassostrea gigas* (Thunberg). I. Survival, growth, glycogen content, and sexual maturation in yearlings. Journal of Experimental Marine Biology and Ecology, 102(2–3), 197–208. https://doi.org/10.1016/0022-0981(86)90176-0

Putnam, H. M., Trigg, S. A., White, S. J., Spencer, L. H., Vadopalas, B., Natarajan, A., Hetzel, J., Jaeger, E., Soohoo, J., Gallardo-Escárate, C., Goetz, F. W., & Roberts, S. B. (2022). Dynamic DNA Methylation Contributes To Carryover Effects And Beneficial Acclimatization In Geoduck Clams. BioRxiv, 2022.06.24.497506. https://doi.org/10.1101/2022.06.24.497506

Juárez, O. E., Lafarga-De la Cruz, F., Leyva-Valencia, I., López-Landavery, E., García-Esquivel, Z., Díaz, F., Re-Araujo, D., Vadopalas, B., & Galindo-Sánchez, C. E. (2018). Transcriptomic and Metabolic Response To Chronic And Acute Thermal Exposure Of Juvenile Geoduck Clams *Panopea globosa*. Marine Genomics, 42, 1–13. https://doi.org/10.1016/j.margen.2018.09.003

Timmins-Schiffman, E. B., Crandall, G. A., Vadopalas, B., Riffle, M. E., Nunn, B. L., & Roberts, S. B. (2017). Integrating Discovery-driven Proteomics and Selected Reaction Monitoring to Develop a Noninvasive Assay for Geoduck Reproductive Maturation. Journal of Proteome Research, 16(9), 3298–3309. https://doi.org/10.1021/acs.jproteome.7b00288

Timmins-Schiffman, E., Guzmán, J. M., Elliott Thompson, R., Vadopalas, B., Eudeline, B., & Roberts, S. B. (2020). Dynamic response in the larval geoduck (*Panopea generosa*) proteome to elevated pCO2. *Ecology and Evolution*, *10*(1), 185–197. https://doi.org/10.1002/ece3.5885

Spencer, L. H., Horwith, M., Lowe, A. T., Venkataraman, Y. R., Timmins-Schiffman, E., Nunn, B. L., & Roberts, S. B. (2019). Pacific geoduck (*Panopea generosa*) resilience to natural pH variation. Comparative Biochemistry and Physiology - Part D: Genomics and Proteomics, 30, 91–101. https://doi.org/10.1016/j.cbd.2019.01.010

Gurr, S. J., Trigg, S. A., Vadopalas, B., Roberts, S. B., & Putnam, H. M. (2022). Acclimatory Gene Expression Of Primed Clams Enhances Robustness To Elevated pCO2. Molecular Ecology, 31(19), 5005–5023. https://doi.org/10.1111/mec.16644

Gurr, S. J., Vadopalas, B., Roberts, S. B., & Putnam, H. M. (2020). Metabolic Recovery And Compensatory Shell Growth Of Juvenile Pacific Geoduck *Panopea generosa* Following Short-Term Exposure To Acidified Seawater. Conservation Physiology, 8(1). https://doi.org/10.1093/conphys/coaa024

Mun, S., Kim, Y. J., Markkandan, K., Shin, W., Oh, S., Woo, J., Yoo, J., An, H., & Han, K. (2017). The whole-genome and transcriptome of the Manila clam (*Ruditapes philippinarum*). Genome Biology and Evolution, 9(6), 1487–1498. https://doi.org/10.1093/gbe/evx096

Wang, K., del Castillo, C., Corre, E., Espinosa, E. P., & Allam, B. (2016). Clam focal and systemic immune responses to QPX infection revealed by RNA-seq technology. BMC Genomics, 17(1), 1–22. https://doi.org/10.1186/s12864-016-2493-9

Dheilly, N. M., Lelong, C., Huvet, A., Kellner, K., Dubos, M. P., Riviere, G., Boudry, P., & Favrel, P. (2012). Gametogenesis in the Pacific oyster *Crassostrea gigas*: A microarrays-based analysis identifies sex and stage specific genes. *PLoS ONE*, *7*(5). https://doi.org/10.1371/journal.pone.0036353

Sun, Y., Liu, X., Xie, X., Bai, Y., Wang, S., Teng, W., Li, D., Li, H., Yu, Z., Zhang, M., Zhou, Z., Liu, X., Nie, H., Du, S., Li, X., Li, Q., & Wang, Q. (2022). A high-quality chromosome-level genome assembly of the bivalve mollusk *Mactra veneriformis*. G3: Genes, Genomes, Genetics, 12(11). https://doi.org/10.1093/g3journal/jkac229

Farhat, S., Bonnivard, E., Pales Espinosa, E., Tanguy, A., Boutet, I., Guiglielmoni, N., Flot, J. F., & Allam, B. (2022). Comparative analysis of the *Mercenaria mercenaria* genome provides insights into the diversity of transposable elements and immune molecules in bivalve mollusks. BMC Genomics, 23(1). https://doi.org/10.1186/S12864-021-08262-1

Roberts S, White S, Trigg S, Putnam H. 2020 Panopea_generosa_genome. http://dx.doi.org/10.17605/OSF.IO/YEM8n.

Andrews S (2010) FastQC: A Quality Control Tool for High Throughput Sequence Data.

Drake, K. J., Sidorov, V. Y., McGuinness, O. P., Wasserman, D. H., & Wikswo, J. P. (2012). Amino acids as metabolic substrates during cardiac ischemia. In Experimental Biology and Medicine (Vol. 237, Issue 12, pp. 1369–1378). NIH Public Access. https://doi.org/10.1258/ebm.2012.012025

McLean, K. W., & Whiteley, A. H. (1974). RNA synthesis during the early development of the pacific oyster, *Crassostrea gigas*. Experimental Cell Research, 87(1), 132–138. https://doi.org/10.1016/0014-4827(74)90534-5

Galli, G., Hofstetter, H. & Birnstiel, M. Two conserved sequence blocks within eukaryotic tRNA genes are major promoter elements. Nature 294, 626–631 (1981). https://doi.org/10.1038/294626a0

Broquard, C., Saowaros, S. ard, Lepoittevin, M., Degremont, L., Lamy, J. B., Morga, B., Elizur, A., & Martinez, A. S. (2021). Gonadal transcriptomes associated with sex phenotypes provide potential male and female candidate genes of sex determination or early differentiation in *Crassostrea gigas*, a sequential hermaphrodite mollusc. BMC Genomics, 22(1), 609. https://doi.org/10.1186/s12864-021-07838-1

Matsumoto, T., Masaoka, T., Fujiwara, A., Nakamura, Y., Satoh, N., & Awaji, M. (2013). Reproduction-related genes in the pearl oyster genome. Zoological Science, 30(10), 826–850. https://doi.org/10.2108/zsj.30.826

Jeffery, J., Neyt, C., Moore, W., Paterson, S., Bower, N. I., Chenevix-Trench, G., Verkade, H., Hogan, B. M., & Khanna, K. K. (2015). Cep55 regulates embryonic growth and development by promoting Akt stability in zebrafish. FASEB Journal, 29(5), 1999–2009. https://doi.org/10.1096/fj.14-265090

Chen, J., Li, Q., Kong, L., & Zheng, X. (2011). Molecular phylogeny of venus clams (*Mollusca, Bivalvia, Veneridae*) with emphasis on the systematic position of taxa along the coast of mainland China. *Zoologica Scripta*, *40*(3), 260–271. https://doi.org/10.1111/j.1463-6409.2011.00471.x

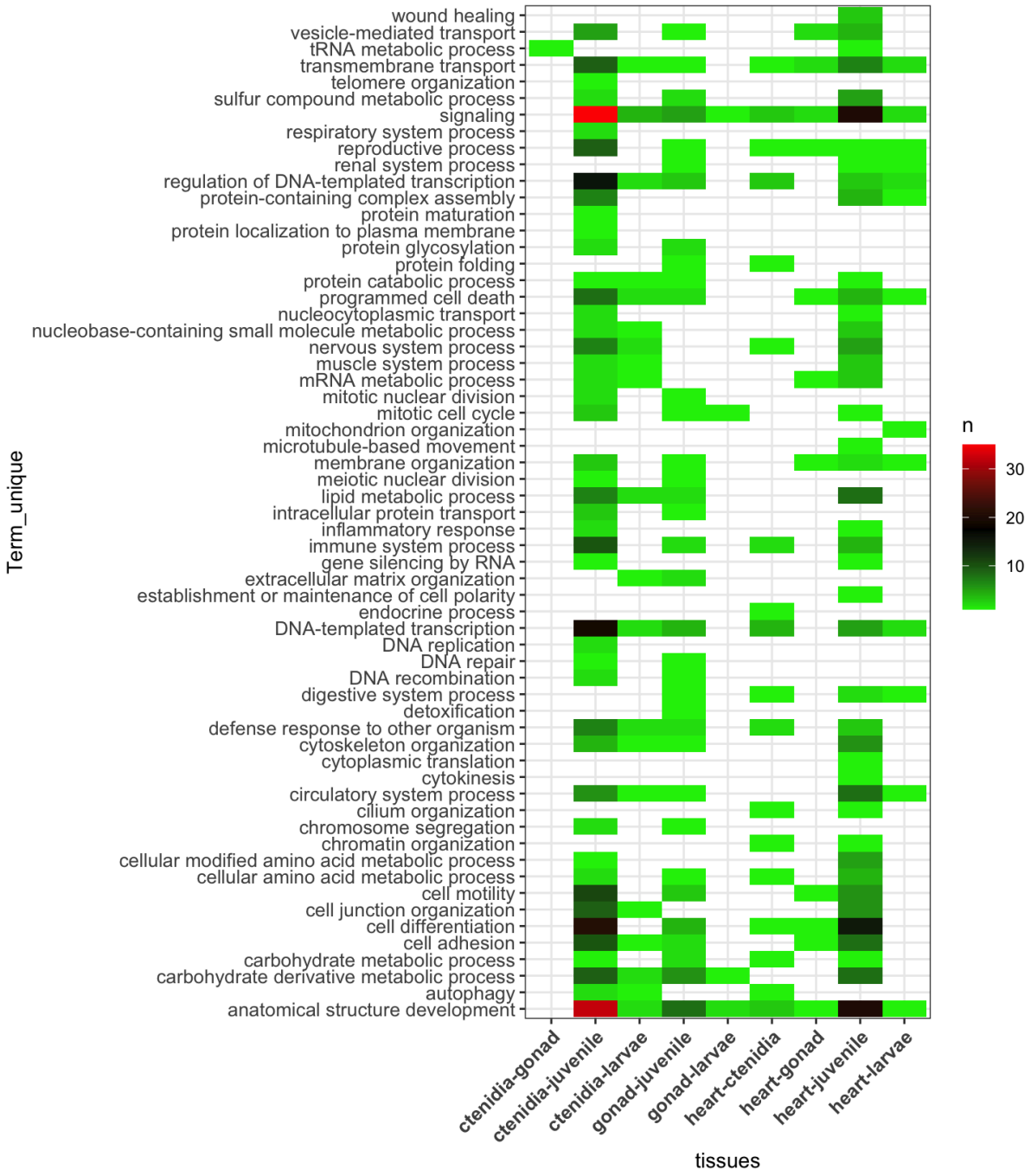
---

Supplemental Figures

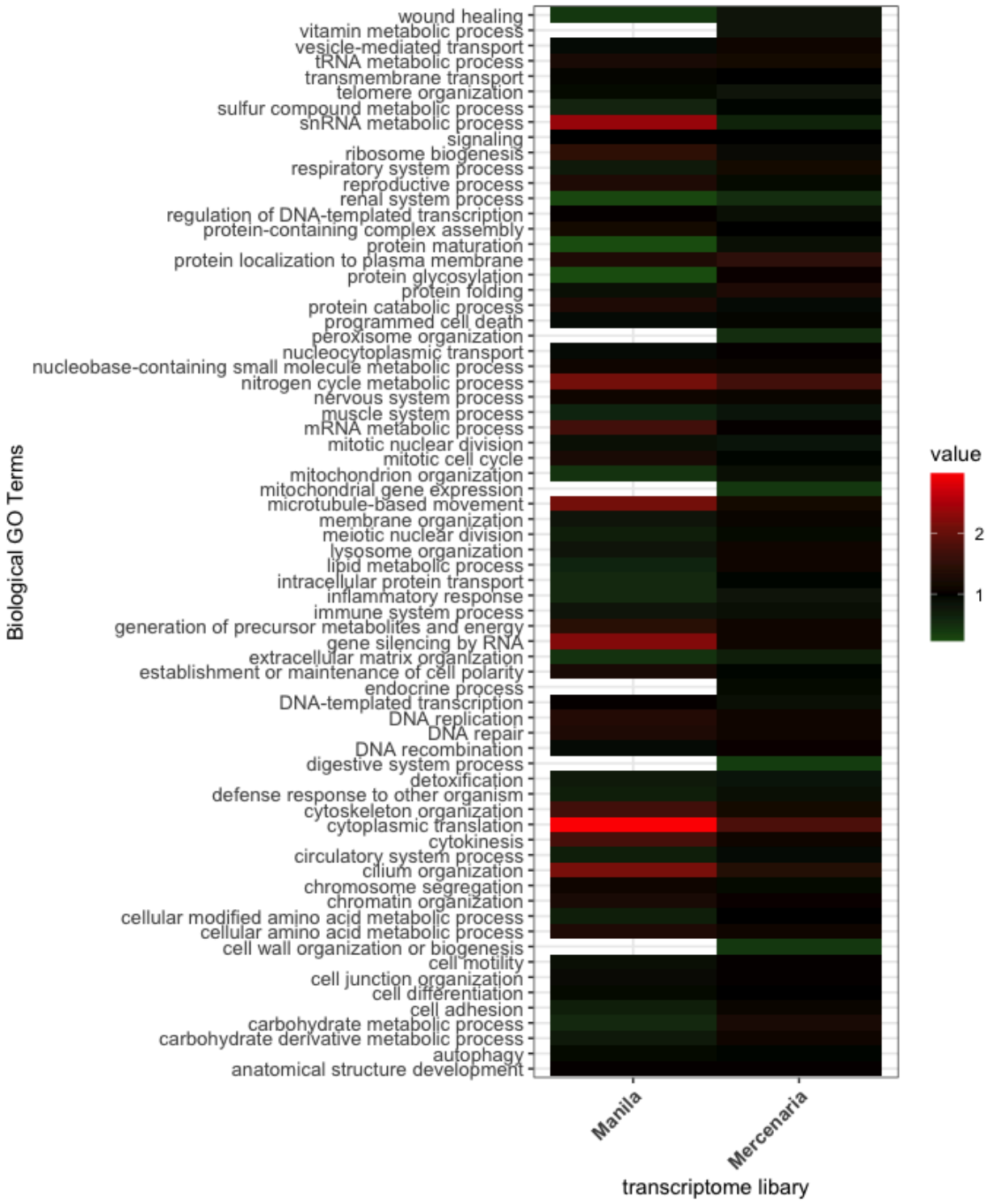
Supplemental Table 1: Full *P. generosa* annotated transcriptome with biological gene ontology information

https://github.com/course-fish546-2023/olivia-geoduck/blob/main/output/tables/Supp_%10Table_1.xlsx

Supplemental Figure 1: Heatmap of comparative RNA-seq libraries with biological gene ontology (GO) terms



Supplemental Figure 2: Comparative RNA-seq libraries with biological gene ontology (GO) terms of Manila clams and Mercenaria clams



Supplemental Table 2: Reproductive genes in the RNA-seq library "gonad" where biological gene ontology terms = "reproductive process"

https://github.com/course-fish546-2023/olivia-geoduck/blob/main/output/tables/Supp_Table_2.xlsx

Supplemental Table 3: Top 20 biological gene ontology terms by clam transcript library

order	transcriptome	Unique Goslim Term	n
1	Geoduck	anatomical structure development	1128
2	Geoduck	signaling	962
3	Geoduck	cell differentiation	785
4	Geoduck	DNA-templated transcription	422
5	Geoduck	immune system process	413
6	Geoduck	regulation of DNA-templated transcription	382
7	Geoduck	cell motility	354
8	Geoduck	programmed cell death	329
9	Geoduck	reproductive process	318
10	Geoduck	cytoskeleton organization	305
11	Geoduck	vesicle-mediated transport	294
12	Geoduck	protein-containing complex assembly	278
13	Geoduck	cell adhesion	232
14	Geoduck	lipid metabolic process	212
15	Geoduck	nervous system process	210
16	Geoduck	transmembrane transport	206
17	Geoduck	mitotic cell cycle	200
18	Geoduck	defense response to other organism	185
19	Geoduck	carbohydrate derivative metabolic process	179
20	Geoduck	protein catabolic process	178
1	Manila	anatomical structure development	67
2	Manila	signaling	57
3	Manila	cell differentiation	41
4	Manila	cytoskeleton organization	36
5	Manila	DNA-templated transcription	23

6	Manila	protein-containing complex assembly	22
7	Manila	regulation of DNA-templated transcription	20
8	Manila	cell motility	19
9	Manila	immune system process	19
10	Manila	mitotic cell cycle	18
11	Manila	reproductive process	18
12	Manila	transmembrane transport	18
13	Manila	protein catabolic process	17
14	Manila	nervous system process	16
15	Manila	programmed cell death	15
16	Manila	vesicle-mediated transport	14
17	Manila	mRNA metabolic process	13
18	Manila	cilium organization	12
19	Manila	microtubule-based movement	12
20	Manila	DNA repair	10
1	Mercenaria	anatomical structure development	456
2	Mercenaria	signaling	370
3	Mercenaria	cell differentiation	300
4	Mercenaria	cell motility	146
5	Mercenaria	cytoskeleton organization	144
6	Mercenaria	DNA-templated transcription	136
7	Mercenaria	immune system process	135
8	Mercenaria	vesicle-mediated transport	128
9	Mercenaria	regulation of DNA-templated transcription	126
10	Mercenaria	reproductive process	118
11	Mercenaria	programmed cell death	116
12	Mercenaria	protein-containing complex assembly	115
13	Mercenaria	cell adhesion	103
14	Mercenaria	lipid metabolic process	94
15	Mercenaria	nervous system process	86

16	Mercenaria	carbohydrate derivative metabolic process	83
17	Mercenaria	mitotic cell cycle	83
18	Mercenaria	cell junction organization	75
19	Mercenaria	protein catabolic process	71
20	Mercenaria	transmembrane transport	71

Supplemental Table 4: Genbank accession numbers from *Dheilly et al. 2012*
https://github.com/ocattau/code-for-Pgenerosa/blob/main/output/oyster_gonad_clusters.csv