Proteomic assay development for reproductive maturation state determination in Geoduck: towards a non-invasive proteomic assay for improving aquaculture

Authors: Emma, Grace, Brent, Michael Riffle, Steven

Affiliation

Target Journal
Journal of Proteome Research
submit preprint to biorxiv

Abstract

Geoduck clams are an increasingly important aquaculture product in the Pacific Northwest of the United States, while also holding essential ecological roles in the ecosystem. These long-lived clams are very fecund, but hatchery production of larvae is hindered by the inability to sex a live geoduck, thus resulting in asynchronous spawning of broodstock and unequal sex ratios. During geoduck reproductive maturation, the physiology of the gonad changes from sexually undifferentiated connective tissue to sex-specific, mature reproductive cells that can be released into the water column. We applied proteomics tools to uncover the cellular mechanisms underlying these physiological changes in the gonad. Sex- and stage-specific proteins were characterized using data dependent acquisition (DDA), or whole proteome profiling, on gonad tissue. Gonad proteomes became increasingly divergent between males and females as maturation progressed. The DDA data were leveraged to develop biomarkers of geoduck sex and maturation stage, analyzed with selected reaction monitoring (SRM) in gonad and hemolymph. The SRM assay yielded a reduced suite of peptides that can be used as an efficient assay to non-lethally determine geoduck sex and maturation stage pre-spawning. This is one of a few examples of cutting-edge proteomics being used to develop applicable tools for the aquaculture industry.

Introduction

Geoduck clams play important ecological and economic roles in the Pacific Northwest of the United States, leading to an increased demand for more efficient hatchery practices. [Can we get an FAO statistic about geoduck?] The fishery has an annual worth of approximately $40 million USD (Khan 2006), with X% of production exported to China and other countries. The high market value of geoduck clams has led to black market trade. These market demands have led to increases in hatchery production of geoduck seed in recent years, but the efficiency of production is stymied by an inability to determine individual clam sex and reproductive stage before spawning is attempted.

Geoduck spawning can be unsuccessful because of mismatched spawn timing within a group of broodstock. In addition to varied spawn timing, aquaculturists are unable to determine the sexes of geoduck. There is no distinguishing coloration, size, or other morphological feature between the sexes. Geoduck gametogenesis, the development of the oocytes in females and spermatocytes in males, begins in September, with spawning events occurring in March through July (Goodwin and Pease, 1989). It is also known that male geoduck clams mature sooner than females, however both sexes are not reproductively mature by approximately one year (Andersen, 1971). Currently, the only way to determine the sex of geoduck clams is to lethally dissect their gonads and examine a tissue sample, or wait for spawning, which occurs when females and males release their gametes for reproduction. Because of this, sex ratios of brood animals prior to spawning are unknown and maximum productivity in hatcheries cannot be attained.

In addition to not achieving maximum productivity, hatcheries run the risk of having low genetic diversity in larval cohorts. This could pose problems both for raising robust seed within the hatchery and for generating geoduck for restoration purposes. Inbreeding can be minimized if sex ratios are determined pre-spawning to maximize male-female pairing. Genetic diversity aids in a population’s resilience to different environmental conditions due to a greater variability in phenotypes (Hughes and Stachowicz, 2004). Furthermore, if cross-breeding between hatchery and wild stocks ever occurs, the diversity of the wild stock will be minimized (Waples, 1991).

Since male and female geoduck gonads are morphologically and functionally different at reproductive maturity, molecular assays to track these early biochemical changes could help predict which animals to spawn and when. Significant changes in both gene and protein expression are expected as gonad tissue differentiates in preparation of spawning. Thus, proteomic assays designed to detect these expression changes could lead to non-lethal, informative, and efficient tools for selecting broodstock.

Here we apply data dependent acquisition and selected reaction monitoring to characterize the geoduck gonad proteome and develop peptide assays that differentiate among sexes and maturation stages. Various proteomics studies of aquatic organisms have revealed large differences in protein abundance between organisms of different sex or reproductive status. Proteome profiles of Pacific oyster hemocytes revealed patterns of differential protein abundance between females of differing oocyte quality, as measured by larval production (Corporeau et al., 2012). Significant protein abundance differences were detected in the liver tissue of non-reproductively active male and female threespine stickleback (Viitaniemi & Leder 2011) as well as in gonad tissue of Persian sturgeon (Keyvanshokooh et al., 2009) and zebrafish (Groh et al., 2011). These previous studies support the utility of detecting reproductively-linked molecular level differences in protein expression between different sexes and reproductive stages.

have your final paragraph touch on how this study demonstrates that nonlethal proteomic assays on gonads can reveal sex diff, maturation state, and decrease chance of inbreeding.

This study combined histological methods with cutting-edge proteomics technology to develop biomarkers of geoduck sex and reproductive stage.

Methods

Tissue Sampling

Geoduck clams were collected in November 2014 from Nisqually Reach, Washington (latitude:47 08.89 , longitude:122 47.439 WGS84). Clams were collected at depths between 9 to 14 meters from a sandy substrate. Gonad tissue and hemolymph from geoduck clams at early (n=3), mid (n=3), and late stage (n=3) gonad maturation, from both males and females were characterized for protein expression. Female reproductive maturation stages were categorized as early (no secondary oocytes, or oocytes that measure ~5-15µ), middle (secondary oocytes ~50-70µ), and late (secondary oocytes ~65-85µ). Male reproductive maturation stages were characterized as early (mostly somatic cells and ~5% spermatid composition per acinus), middle (about equal parts somatic cells and reproductive tissue and ~50% spermatid composition per acinus), and late (very little somatic cells and ~75-90% spermatid composition per acinus). Details of gonadal maturation classification and histological details have been previously described (Crandall and Roberts 2016). The corresponding samples examined as part of this study include fG03, fG04, fG08, mG02, mG07, mG09, fG34, fG35, fG38, mG41, mG42, mG46, fG51, fG69, fG70, mG65, mG67, mG68. Hemolymph samples were collected from the same individuals, except for the early stage males and females and mid-stage female 34, all of which had poor quality hemolymph samples and were therefore substituted with samples from clams of the same sex and stage. For the hemolymph, there were only two biological replicates for mid female (fH25 and fH35) and mid male (mH42 and mH46).

Protein preparation

Hemolymph protein content was determined using Pierce's BCA assay and a volume containing 50 µg of protein for each sample was evaporated to ~20 µl and then resuspended in 100 µl of 6M urea in 50 mM NH4HCO3 for sonication. Gonad tissue from each of the eighteen geoduck clams was sonicated in 50 mM NH4HCO3 and 6 M urea. An aliquot (100 µl) of the sonicated gonad tissue was used for protein isolation and 100 µg was used for protein digestion (Pierce's BCA assay). Protein digestion for both tissues followed the protocol outlined in Timmins-Schiffman et al. (2013). Briefly, each sample was incubated with TCEP buffered at pH 8.8 (1 hr, 37°C). Samples were alkylated with iodoacetamide (IAM; 1hr, 20°C) followed by a 1 hr incubation with dithiothreitol to absorb any remaining IAM. To each sample, NH4HCO3 and HPLC grade methanol were added to dilute urea and to increase solubilization of membrane proteins. Samples were digested overnight with trypsin at 37°C. Digested samples were evaporated and reconstituted in 5% acetonitrile (ACN) + 0.1% trifluoroacetic acid (TFA) (100 µl) and pH was decreased to < 2. Desalting of the samples was done using Macrospin columns (sample capacity 0.03-300 µg; The Nest Group, Southborough, MA, USA) following the manufacturer's specifications. Dried peptides were reconstituted in 100 µl of 5% ACN + 0.1% formic acid.

Global Proteome Analysis: Data-Dependent Acquisition (DDA) LC-MS/MS

Data-dependent acquisition (DDA), also referred to as shotgun proteomics, was performed to complete a whole proteome comparison of different geoduck sexes and maturation stages across all eighteen individuals for the gonad tissue (Figure 1). Liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS) was accomplished on a Q-Exactive-HF (Thermo) on technical triplicates for each sample. The analytical column was 20 cm long and packed with C18 beads (Dr. Maisch, 0.3 µm) with a flow rate of 0.3 µl/min. Chromatography was accomplished with an increasing ratio of solvent A (ACN + 0.1% formic acid): solvent B (water + 0.1% formic acid). The solvent gradient consisted of: 0-1 minutes 2% A; 1-60 minutes 5% A; 60-61 minutes 35% A; 61-71 minutes 80% A; 71-90 minutes 2% A. Quality control standards (Pierce Peptide Retention Time Calibration mixture (PRTC) + bovine serum albumin peptides (BSA)) were analyzed throughout the experiment to ensure consistency of peptide detection and elution times.

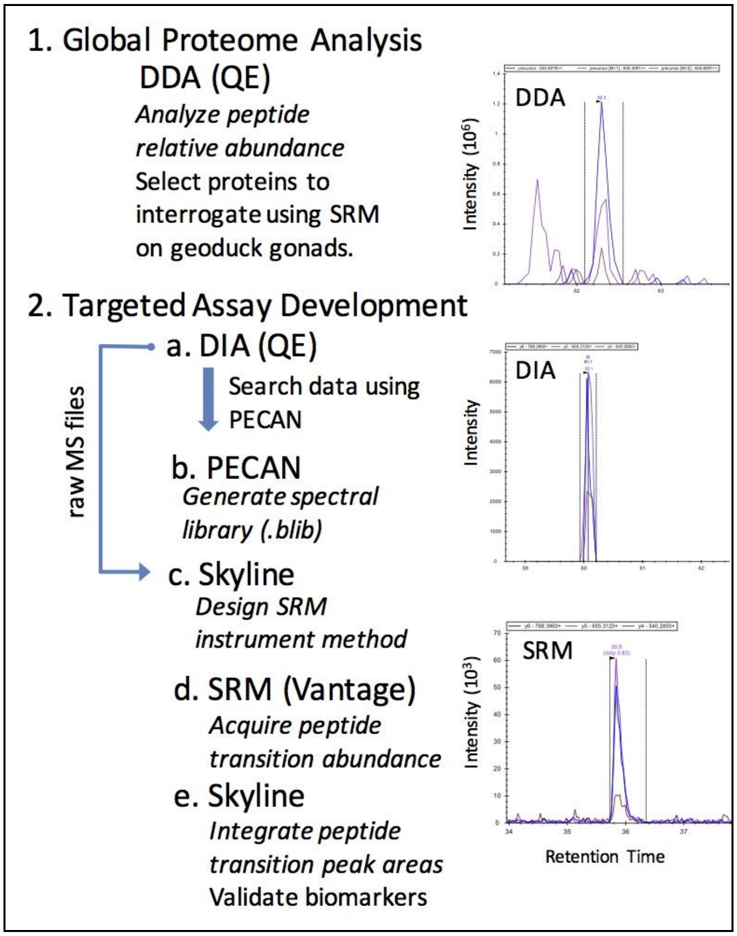

1. Global Proteome Analysis
DDA (QE)
Analyze peptide relative abundance
Select proteins to interrogate using SRM on geoduck gonads.
2. Targeted Assay Development
raw MS files
a. DIA (QE)
Search data using PECAN
b. PECAN
Generate spectral library (.blib)
c. Skyline
Design SRM instrument method
d. SRM (Vantage)
Acquire peptide transition abundance
e. Skyline
Integrate peptide transition peak areas
Validate biomarkers
DDA
Intensity (10^6)
DIA
Intensity
SRM
Intensity (10^3)
Retention Time

Figure 1. Illustration of experimental setup and workflow for mass spectrometry data acquisition and analysis (adapted from Timmins-Schiffman et al., in review). MS experimental workflow: 1. Data dependent acquisition (DDA) was performed on the Q-Exactive-HF (QE) to assess the global proteomic differences in gonad tissue between sexes and maturation stages. 2. Targeted assay development followed the steps of (a) data independent acquisition (DIA) on gonad tissue was also completed on the QE to create spectral libraries for selected reaction monitoring (SRM) method development; (b&c) spectral libraries were analyzed in PECAN using Skyline to select optimal transitions and design an instrument method for SRM analyses; (d) SRM was completed on the TSQ Vantage for geoduck peptide transitions in gonad and hemolymph; (e) Peptide transition detection and quantification was performed in Skyline. The chromatograms of peptide KEEELIDYMVKQ (from protein 130261_c0_seq1|m.17926) were collected using the 3 different MS approaches (DDA, DIA, and SRM) from the same late-stage female. Black vertical lines indicate peak integration boundaries, and colored peaks represent the different transitions (i.e. peptide fragments) collected.

DDA protein identification and quantification

Gonad peptides were identified and proteins inferred using a proteome derived from a *de novo* assembled transcriptome of a male and female geoduck clam gonad tissue libraries (NCBI Bioproject Accession #PRJNA316216) (Del Rio-Portilla et al unpublished). Briefly, reads were assembled using Trinity (ref) and deduced protein sequences determined using the Transdecoder algorithm within Trinity. The raw mass spectrometry data (PRIDE Accession #PXD003127) was searched against the protein database (fasta = Supplemental file 1) using Comet v 2015.01 rev.2 specifying trypsin as the cleaving enzyme, two missed cleavages allowed, peptide mass tolerance of 20 ppm, cysteine modification of 57 Da (from the IAM) and methionine modification of 15.999 Da (from oxidation) (parameter file - supp file 2) to find peptide spectral matches (Eng et al., 2012 & 2015). Protein inference and match probability were found using the Trans-Proteomic Pipeline with a probability cut-off of 0.9 (Pedrioli 2010; Deutsch et al., 2015, version downloaded November, 2014). Protein identifications were considered true matches when the probability of the match was at least 0.9 and at least 2 independent spectra were associated with the protein across all samples.

Non-metric multidimensional scaling analysis (NMDS) was used to determine the similarity of technical replicates (supp file 3) using the vegan package (Oksanen et al., 2016) in R v. 3.2.3 (R Core Team, 2015). NMDS was performed on log-transformed data using a Bray-Curtis dissimilarity matrix. As technical replicates clustered closely together and showed less variability than biological replicates (supp file 4), spectral counts were averaged across each sample (n=18). Normalized spectral abundance factors (NSAF) were calculated on the averaged spectral counts for all high confidence, detected proteins as a proxy for protein expression (Florens et al., 2006). Proteomics differences between sexes and maturation stages were explored with two methods: 1) NMDS and analysis of similarity (ANOSIM) were used to compare the entire proteomic profiles in multivariate space and 2) QSpec (Choi et al., 2008) was used to determine significant differences at the individual protein level. NMDS was performed on NSAF data followed by ANOSIM in the vegan package in R to determine the differences between different stages within sex, and between maturational stages across sex. Differentially abundant proteins among the six conditions were identified using QSpec (Choi et al., 2008). Spectral counts were summed across technical replicates to create input files for QSpec (http://www.nesvilab.org/qspec.php/) with data normalized by abundance. Nine comparisons were analyzed for differentially abundant proteins: Early male (EM) vs. early female (EF); midstage male (MM) vs. midstage female (MF); late male (LM) vs. late female (LF); EM vs. MM; EM vs. LM; MM vs. LM; EF vs. MF; EF vs. LF; MF vs. LF. Proteins were considered significantly different if the absolute value of the log fold change was $\geq$ 0.5 and the absolute value of the z-statistic was $\geq$2.

The gonad transcriptome was translated and protein coding sequences were predicted using Transdecoder (http://transdecoder.sf.net). The proteome was annotated with UniProt-KB/Swissprot accession numbers with BLASTp (Uniprot trembl database downloaded April 28, 2015, www.uniprot.org).

- GoSLIM
- Compare to dheilly-
- Gigaton

Proteins that were annotated with Uniprot IDs of proteins known to be involved in egg and sperm interactions were searched against the non-redundant protein database in NCBI (BLASTp) constrained by species selected for comparison. These proteins were a zona pellucida sperm-binding protein, a polycystic kidney disease and receptor for egg jelly-related protein, and a receptor for egg jelly protein (supp file 7). The geoduck sequences were directly compared with several marine invertebrates that have sequenced genomes: *Nematostella vectensis*, *Acropora digitifera*, *Strongylocentrotus purpuratus*, *Crassostrea gigas*, *Lottia gigantea*, *Octopus bimaculoides*, and *Danio rerio* (as the out-group). The top hit (below the e-value cut-off of 1E-10) was downloaded for each species and aligned with the same proteins from other invertebrates in Clustal Omega (Sievers et al., 2011). Phylogenetic trees were constructed using Phylip 3.695 Drawgram from the Clustal Omega Newick tree output using default parameters (use branch lengths if present; branch length scaling = automatic; depth/breadth of tree = 0.53; stem length/tree depth = 0.05; character height/tip space = 0.333; ancestral nodes = weighted) (Felsenstein 1989 & 2005). Full length sequences were used for phylogenetic trees, but sequence alignments trimmed to the sequence length represented in all the species were used for percent identity (calculated in Clustal Omega).

Gene Ontology (GO) enrichment analysis was performed for each sex and maturation stage on 1) all the proteins detected in the sexo-stage and 2) the proteins that were uniquely detected in that sex-stage in the gonad tissue. Briefly, a p-value was calculated representing the likelihood that a GO term would be as represented as it was (or more) by chance among the set of tested proteins, given the annotation of the geoduck gonad proteome (3,627 proteins). A p-value cutoff of 0.01 was used to ascribe statistical significance to GO terms representing biological process.

GO terms were first assigned directly to the geoduck gonod protein names resulting from the *de novo*-assembled transcriptome (see methods). This was done by assigning GO terms associated with Uniprot-KB BLAST hits (see methods) to the geoduck protein names that produced that hit. The geoduck protein names present in the FASTA file used to search the data by mass spectrometry were then used to look up GO assignments and perform GO analysis.

A p-value representing the statistical significance of the representation of a GO term in a set of proteins was calculated using the hypergeometric distribution using the following formula:

$$P(I) = \frac{\binom{A}{I} * \binom{T-A}{B-I}}{\binom{T}{B}}$$

Where A = total number of proteins submitted that have a GO annotation, B = the total number of proteins in the background proteome annotated with the given GO term (or any of its descendants), I (intersection of A and B) = the total number of submitted proteins annotated with the given GO term (or any of its descendants), and T = the total number of annotated proteins in the proteome background.

Then, the p-value describing the chance of having an intersection of size I or larger by chance may be computed as:

P-value = $\sum_{i=I}^{\min(A,B)} P(i)$, where min (A,B) is the minimum of values A and B.

The p-value is then corrected for multiple hypothesis testing using the Bonferroni method by multiplying the p-value by the number of GO terms tested (setting resulting values over 1 to 1). The number of GO terms tested equals the number of GO terms found (and all ancestors) for the submitted set of proteins.

To generate the GO graph images, a p-value was calculated for every GO term represented in the protein set and for all ancestor terms up to the root term. The result is a complete directed acyclic graph (DAG) that represents a subset of the whole GO DAG where every term has an associated p-value. This DAG was then filtered by removing all childless terms that had an associated p-value > 0.01. The resulting DAG was then filtered using the same method, and this process repeated until no childless terms remained with a p-value > the cutoff. The final result is a filtered subset of the GO DAG that contains no leaf nodes with a p-value greater than the cutoff, but where a given term is guaranteed to have all of its ancestor terms, even if those terms have a p-value greater than the cutoff. Having the ancestor terms present is critical to visualization as they provide context for interpreting the results.

A web application for performing this analysis for the geoduck gonad proteome is available for public use at http://yeastrc.org/compgo_geoduck/pages/goAnalysisForm.jsp and can be explored with our input files (supp file 5). The source code of the web application and script used to assigned Uniprot GO annotations to geoduck gonad accession strings is available at https://github.com/yeastrc/compgo-geoduck-public.

Targeted Assay Development: Selected Reaction Monitoring (SRM)

A subset of proteins were chosen for development of a suite of targeted assays using selected reaction monitoring (SRM) on the mass spectrometer (MS). Based on the data-dependent acquisition, proteins that were detected only in one of these stages; early stage males (EM), early stage females (EF), late stage males (LM), or late stage females (LF) were screened for usable peptide transitions in SRM in Skyline in gonad tissue.

Data independent acquisition (DIA) was used to generate spectral libraries for biomarker development in the gonad tissue (Figure 1). Equal amounts of isolated peptides from the three biological replicates for EM, EF, LM, and LF used in the DDA experiment (described above) were pooled in equal quantities for DIA on the Q-Exactive HF (Thermo). Each sample included a spiked-in internal quality control peptide standard (375 fmol PRTC + BSA; Pierce, hereafter referred to as “QC”). Sample injections for all DIA experiments included 1 ug protein plus the QC in a 2 µl injection. An analytical column (27cm) packed with 3 µm C18 beads (Dr. Maisch) and a 3 cm trap with 3 µm C12 beads (Dr. Maisch) were used for chromatography. Technical replicate DIA spectra were collected in 4 m/z isolation width windows spanning 125 m/z ranges each (400-525, 525-650, 650-775, 775-900; Panchaud et al., 2009). For each method, a gradient of 5-80% ACN over 90 minutes was applied for peptide spectra acquisition. Raw data can be accessed via ProteomeXchange (http://www.proteomexchange.org/) under identifier PXD004921. MSConvert (Chambers et al., 2012) was used to generate mzML files from the raw DIA files.

In order to generate spectral libraries for targeted method development, Peptide Centric Analysis was completed with the software program PECAN (Ting et al., in review; Ting et al., 2015). Input files included the list of peptides generated for SRM (n=212), as described above, and the mzML files generated from the raw DIA files. PECAN correlates a list of peptide sequences with the acquired DIA spectra in order to locate the peptide-specific spectra within the acquired DIA dataset. A background proteome of the *in silico* digested geoduck gonad proteome was used.

The PECAN .blib output file was then imported into Skyline daily v. 3.5.1.9706 (MacLean et al., 2010) to select peptide transitions and create MS methods that would target specific peptides and transitions. Peptide transitions are the reproducible fragments of peptides that are generated during the MS2 scan in a MS. Peptides reliably fragment in the same way in the MS, therefore transitions are a robust and consistent signal of a peptide's presence (Wolf-Yadlin et al., 2007). Peptide transitions were selected if peak morphology was uniform and consistent across the MS2 scans for technical replicates. Peptides were selected for targeted analysis if they had ≥ 3 good quality transitions and there were ≥ 2 peptides per protein. A maximum of 4 transitions per peptide were selected for targeted analysis and no more than 3 peptides per protein were selected. The final list included 25 transitions for EM biomarkers, 22 for EF, 133 for LM, and 52 for LF. This transition list was divided between two method files for the final SRM analyses to provide adequate dwell time on individual transitions in order to accurately detect and measure all peptides desired (Picotti and Aebersold, 2012).

Selected reaction monitoring (SRM) was carried out on a Thermo Vantage for all eighteen geoduck gonad samples used in the original DDA analysis. Samples were prepared as described above for DIA (1 µg of protein per 3 µl injection). A new C18 trap (2 cm) and C18 analytical column (27.5 cm) were used and each sample was analyzed in triplicate across two MS experiments to cover the entire peptide transition list (n=212). Raw data can be accessed in the PeptideAtlas (http://www.peptideatlas.org/PASS/PASS00943) under accession PASS00943.

In addition to the peptides selected from the gonad proteome, transitions from five proteins were added to the assay for hemolymph SRM analysis. These proteins were selected based on gonad proteome annotation to include 1) proteins that would likely be circulating in the hemolymph and 2) gonad proteins that had homology with the mussel hemolymph proteome (CITE). The proteins are vitellogenin (2 proteins), glycogen synthase (2 proteins), and glycogenin-1. PECAN was run as described above to use Skyline to select hemolymph protein transitions. For the new hemolymph proteins, the minimums for peptides and transitions could not be met for every protein but were still included in the analysis (n=40 transitions). These transitions, and the previous gonad transitions, were analyzed across the eighteen geoduck hemolymph samples in two technical replicates, as described above for the gonad. Some of these peptides yielded no data in the hemolymph, resulting in a dataset of 171 peptide transitions. Raw data can be accessed in PeptideAtlas (http://www.peptideatlas.org/PASS/PASS00942) under accession PASS00942.

Acquired SRM data were analyzed in Skyline for peptide transition quantification. A Skyline documents that were used to analyze the monitored peptide transitions can be found on Panorama for the gonad dataset and the hemolymph datasets (panoramaweb.org/labkey/geoduckrepro.url). Peak presence was determined based on consistency of retention time (verified by spiked in QC peptides) and peak morphology.

All peptide transition peak intensities were exported from Skyline for automated peak selection and peak integration analysis. QC internal standard transitions were monitored for consistency across runs by calculating coefficient of variation (CV) of transition peak area across injections. Peak intensities for the geoduck peptide transitions were normalized by dividing by the averaged intensities for the 6 QC peptide transitions that had the lowest CV in the gonad data (CV < 13) and the 6 with the lowest CV in the hemolymph data (CV < 10).

NMDS and ANOSIM were performed on the QC-normalized SRM dataset as described above for the DDA dataset. An initial NMDS showed that technical replicates clustered together well and that variation was lower within biological replicates compared to between replicates (supp file 6), therefore technical replicate peak intensities were averaged for the rest of the analysis. ANOSIM was performed using grouping by sex and reproductive stage alone, as well as by a combined sex-stage factor. Coefficients of variation were calculated for combined technical replicates using the raster package in R.

Eigenvector loadings were calculated for the gonad and hemolymph data using the vegan package in R. For each dataset, the top 20 transitions with the combination of lowest p-value and highest loading value were selected as the most "significant" biomarkers. Heatmaps of the log-normalized transition intensities for these biomarker transitions were made using pheatmap in R (CITE), clustering rows (peptide transitions) and columns (samples) using euclidean distance and the average clustering method.

Results

Generic Proteome Annotation.

Proteome

Three proteins were selected for phylogenetic analysis: zona pellucida sperm-binding protein, polycistic kidney disease receptor for egg jelly-related protein, and receptor for egg jelly. The first two proteins were represented multiple times within the geoduck gonad proteome and only the longest sequence entry was selected for this analysis. The geoduck protein sequences were most similar to other molluscs - the geoduck sequence clustered most closely with Pacific oysters for the polycistic kidney disease receptor protein and receptor for egg jelly; it clustered most closely with the giant owl limpet for the zona pellucida sperm-binding protein (Figure). In the percent identity matrix analysis, sequence similarity was highest between geoduck and oyster for all three proteins (Figure).

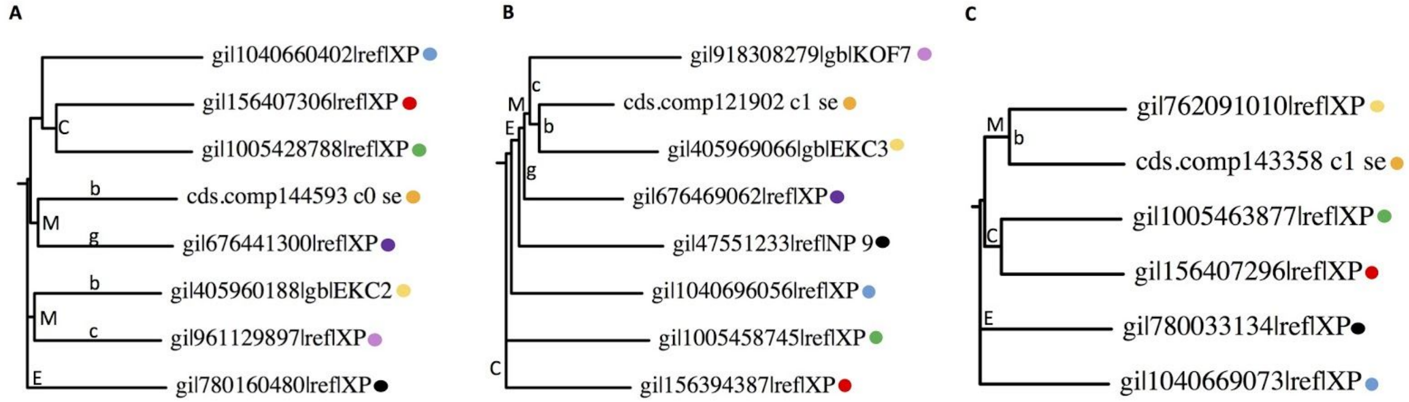


Figure X. Phylogenetic trees of sequences of A) receptor for egg jelly 7, B) polycystic kidney disease and receptor for egg jelly-related protein, and C) zona pellucida-sperm binding protein. Uniprot accession numbers are given for each protein, except for geoduck, which is indicated by the protein ID from the current sequencing effort. The trees include sequences from zebrafish (blue), the coral Acropora digitifera (green), the anemone Nematostella vectensis (red), the oyster Crassostrea gigas (yellow), geoduck (orange), the limpet Lottia gigantea (purple), the octopus Octopus bimaculoides (pink) and the urchin Strongylocentrotus purpuratus (black).

"M" indicates the molluscan clade (with "b", "g", and "c" on the branches for bivalvia, gastropoda, and cephalopoda), "C" the cnidarians, and "E" the echinoderm.

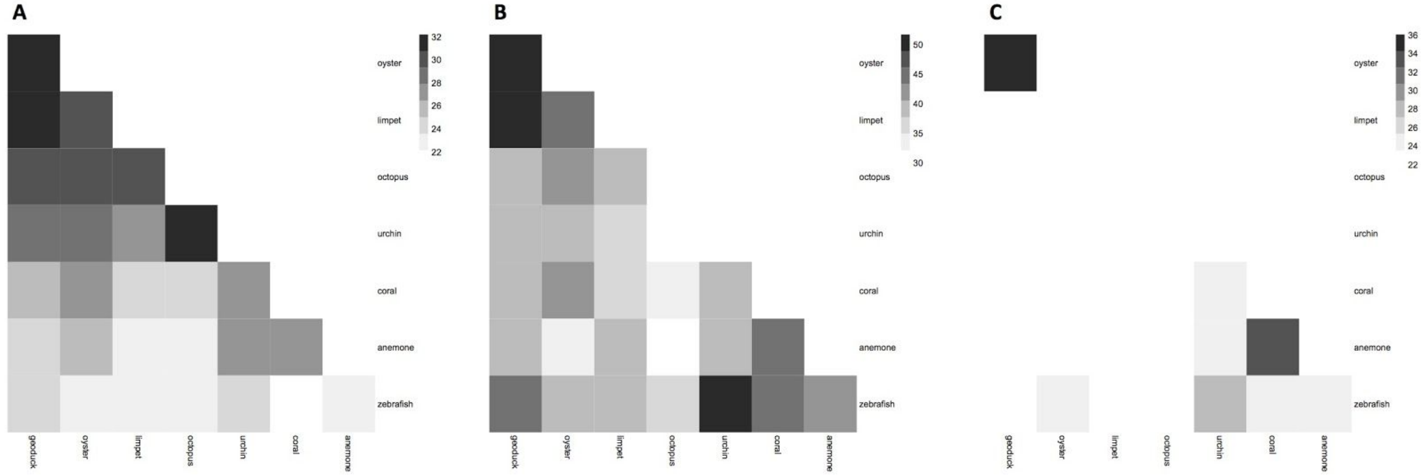


Figure Y. Percent identity heatmaps of sequence similarity among geoduck, oyster, urchin, octopus, anemone, coral, and zebrafish for A) receptor for egg jelly 7, B) polycystic kidney disease and receptor for egg jelly-related protein, and C) zona pellucida-sperm binding protein.

DDA Proteomics

The mass spectrometry data (PRIDE Accession #PXD003127) used in conjunction with a deduced gonad transcriptome identified 3,627 proteins detected with high confidence across males and females of early, middle, and late maturation stages (supp file 8). Female and male gonad proteomic profiles were more similar in early stage maturation, with proteomes diverging as reproductive maturity advanced (NMDS figure). Proteomic profiles were significantly different between sexes (R = 0.4122, p = 0.002) and maturation stages (R = 0.3025, p = 0.004).

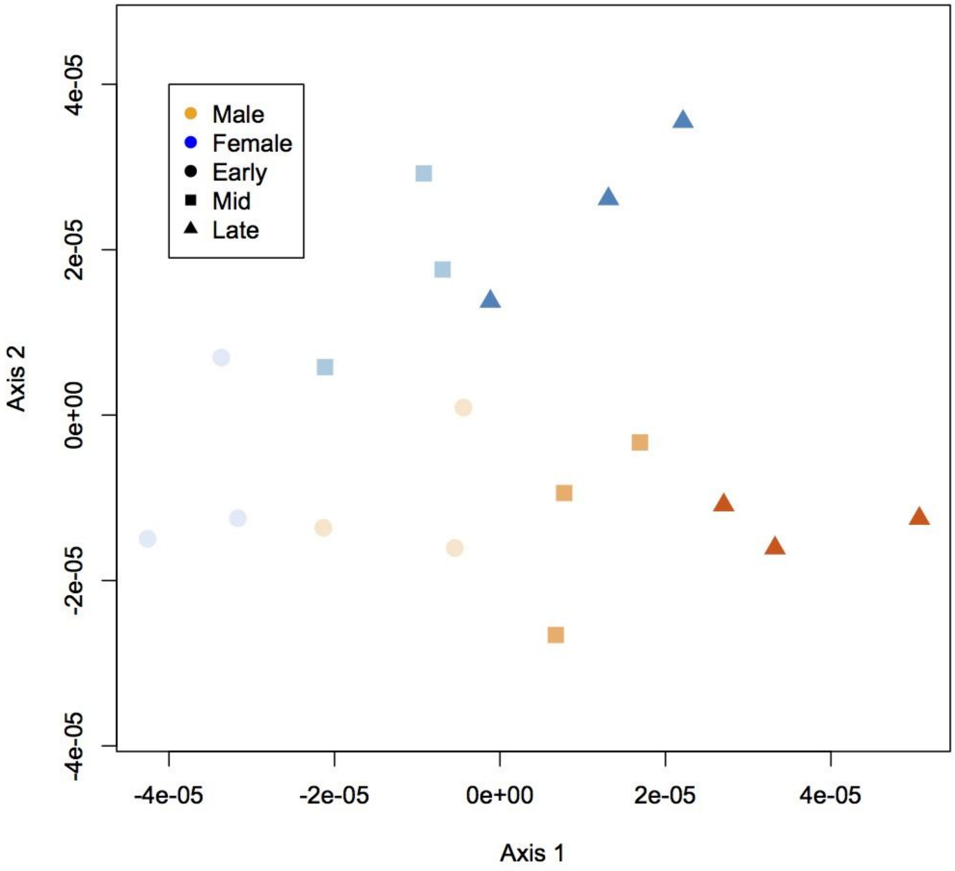
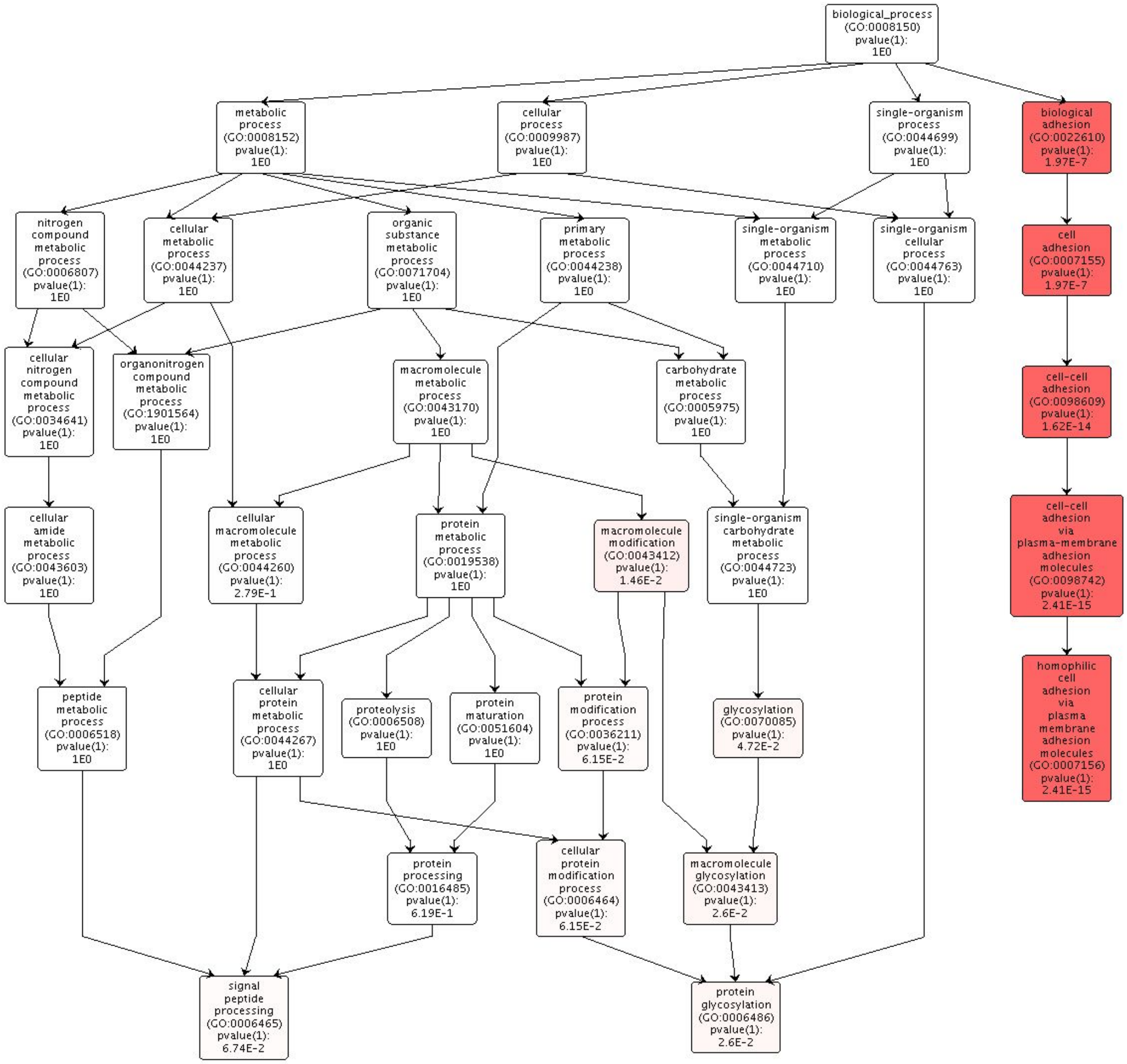


Figure. Non-metric multidimensional scaling plot (NMDS) of geoduck gonad whole proteomic profiles generated by data-dependent analysis. Gonad proteomes differ among clams by both sex (male = orange, female = blue) and stage (early = circles, mid = squares, late = triangles; p<0.05).

In addition to the proteins that were unique to reproductive stage, many more were differentially abundant in comparisons between stage and sex based on analysis with QSpec. Between sexes, there were 387 differentially abundant proteins in early females (EF) vs. early males (EM), 625 in midstage females (MF) vs. midstage males (MM), and 1035 in late females (LF) vs. late males (LM). Between stages in females, there were 445 differentially abundant proteins for EF vs. MF, 414 for MF vs. LF, and 878 for EF vs. LF. There were higher numbers of differentially abundant proteins between reproductive stages for males: 625 for EM vs. MM, 671 for MM vs. LM, and 1011 for EM vs. LM (supp file 8).

NEED SOME BETTER LEAD

Enrichment analysis revealed sex- and stage-specific biological processes related to physiological status of the gonad. Enrichment of biological processes was analyzed for two datasets for each sex and stage: 1) all proteins detected for that sex and stage and 2) proteins uniquely detected in a given sex and stage. Three hundred five proteins were unique to the female proteomic profile and 522 were unique to males. The number of proteins unique to a specific maturation stage increased with maturity from 24 and 20 in early females and males, respectively, to 161 and 145 in late stage females and males. “Translation” and related biological processes were enriched in all sex-stage combinations (supp file 9). In early stage females, proteins involved in phosphorylation were enriched, in midstage female proteins associated with regulation of hydrolase activity and carbohydrate metabolism were enriched (supp file 10 for early and mid stage clams, figure for late stage). In proteins unique to late-stage females, the enriched processes included cell adhesion, protein glycosylation, DNA replication, and signal peptide processing (figure). Early male unique proteins were enriched for small molecule metabolic process and oxidation-reduction process, midstage males had more proteins involved in ATP metabolic process, and late-stage males expressed proteins enriched in cellular macromolecule complex assembly, nucleotide metabolic process, and were highly enriched for microtubule-based movement (figure).


biological_process (GO:0008150) pvalue(1): 1E0
metabolic process (GO:0008152) pvalue(1): 1E0
cellular process (GO:0009987) pvalue(1): 1E0
single-organism process (GO:0044699) pvalue(1): 1E0
biological adhesion (GO:0022610) pvalue(1): 1.97E-7
nitrogen compound metabolic process (GO:0006807) pvalue(1): 1E0
cellular metabolic process (GO:0044237) pvalue(1): 1E0
organic substance metabolic process (GO:0071704) pvalue(1): 1E0
primary metabolic process (GO:0044238) pvalue(1): 1E0
single-organism metabolic process (GO:0044710) pvalue(1): 1E0
single-organism cellular process (GO:0044763) pvalue(1): 1E0
cell adhesion (GO:0007155) pvalue(1): 1.97E-7
cellular nitrogen compound metabolic process (GO:0034641) pvalue(1): 1E0
organonitrogen compound metabolic process (GO:1901564) pvalue(1): 1E0
macromolecule metabolic process (GO:0043170) pvalue(1): 1E0
carbohydrate metabolic process (GO:0005975) pvalue(1): 1E0
cell-cell adhesion (GO:0098609) pvalue(1): 1.62E-14
cellular amide metabolic process (GO:0043603) pvalue(1): 1E0
cellular macromolecule metabolic process (GO:0044260) pvalue(1): 2.79E-1
protein metabolic process (GO:0019538) pvalue(1): 1E0
macromolecule modification (GO:0043412) pvalue(1): 1.46E-2
single-organism carbohydrate metabolic process (GO:0044723) pvalue(1): 1E0
cell-cell adhesion via plasma-membrane adhesion molecules (GO:0098742) pvalue(1): 2.41E-15
peptide metabolic process (GO:0006518) pvalue(1): 1E0
cellular protein metabolic process (GO:0044267) pvalue(1): 1E0
proteolysis (GO:0006508) pvalue(1): 1E0
protein maturation (GO:0051604) pvalue(1): 1E0
protein modification process (GO:0036211) pvalue(1): 6.15E-2
glycosylation (GO:0070085) pvalue(1): 4.72E-2
homophilic cell adhesion via plasma membrane adhesion molecules (GO:0007156) pvalue(1): 2.41E-15
protein processing (GO:0016485) pvalue(1): 6.19E-1
cellular protein modification process (GO:0006464) pvalue(1): 6.15E-2
macromolecule glycosylation (GO:0043413) pvalue(1): 2.6E-2
signal peptide processing (GO:0006465) pvalue(1): 6.74E-2
protein glycosylation (GO:0006486) pvalue(1): 2.6E-2

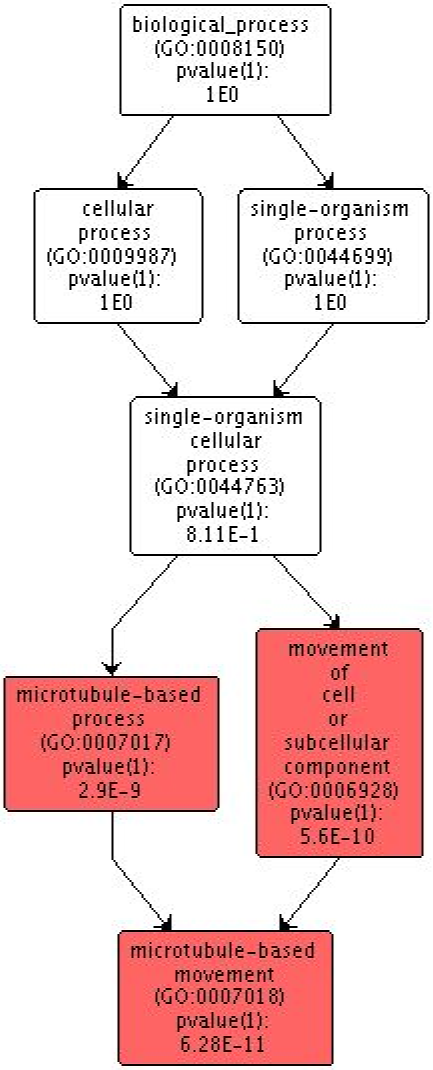

biological_process (GO:0008150) pvalue(1): 1E0
cellular process (GO:0009987) pvalue(1): 1E0
single-organism process (GO:0044699) pvalue(1): 1E0
single-organism cellular process (GO:0044763) pvalue(1): 8.11E-1
microtubule-based process (GO:0007017) pvalue(1): 2.9E-9
movement of cell or subcellular component (GO:0006928) pvalue(1): 5.6E-10
microtubule-based movement (GO:0007018) pvalue(1): 6.28E-11

Figure. Maps of enriched Gene Ontology biological processes for A) late stage female geoduck and B) late stage male geoduck. The complete set of enrichment maps can be found in supplemental file 10. Red processes are from the protein set that was uniquely detected in a sex-stage. Darker colors represent higher significance of enrichment.

SRM

SRM was applied to create peptide assays that could resolve geoduck sexes and maturation stages for better broodstock management. DDA is suited to global discovery of proteomic patterns, but it suffers from lack of reproducibility and relative insensitivity to low abundance peptides. When a MS is in SRM mode, it reproducibly measures the entire signal of specific peptides of interest. Peptide transitions in the gonad (n=212) and hemolymph (n=171) derived from proteins detected in a single sex-maturation stage from early male (EM), early female (EF), late male (LM), and late female (LF) proteomes were measured across all three maturation stages and both sexes (supp file 11). Coefficients of variation were reasonably stable across all biological replicates, although many transitions still have high CVs, indicating a need for assay optimization (supp file 12).

In an ordination plot, the gonad SRM data resolved males and females better in late-stage samples than the early- or mid-stage (Figure NMDSA). There was significant separation based on gonad SRM data for sex (R = 0.2373, p=0.003), stage (R=0.2189, p=0.007), and a combined sex-stage factor (R=0.4132, p=0.001). The hemolymph SRM data only resolved the late female group from the rest of the geoduck (Figure NMDSB). There was significant separation of the clam hemolymph proteomic profiles based on sex (R=0.1384, p=0.043) and sex-stage (R=0.4892, p=0.001), but not by stage alone (R=0.1435, p=0.065).

The peptide transitions that drive the differences among the geoduck sex-stage groups in the NMDS plots are likely the most informative for differentiating groups and have the best potential as biomarkers in a pared down assay. Along Axis 1 in the gonad NMDS plot (Figure NMDS) peptides that are “pushing” the separation of the male maturation stages are from the proteins flap endonuclease, IQ domain-containing protein K, WD repeat-containing protein on Y chromosome, tetratricopeptide repeat protein 18, spectrin alpha chain, and four uncharacterized proteins (Figure heat map). Along Axis 2 for the gonad data, the proteins that play the most significant role in the separation of the female stages are WD repeat-containing protein on Y chromosome, tetratripcopeptide repeat protein 18, spectrin alpha chain, centrosomal protein of 70 kDa, and four uncharacterized proteins. From the hemolymph SRM NMDS (Figure), the main peptide drivers along Axis 1 (separating late females from all other geoduck) are from the proteins flap endonuclease, vitellogenin, and one uncharacterized protein (Figure heat map). In a cluster analysis of just these significant transitions, sex and stage groups from the gonad data cluster well and there is a reasonable male-female separation in the hemolymph data (Figure heat map).

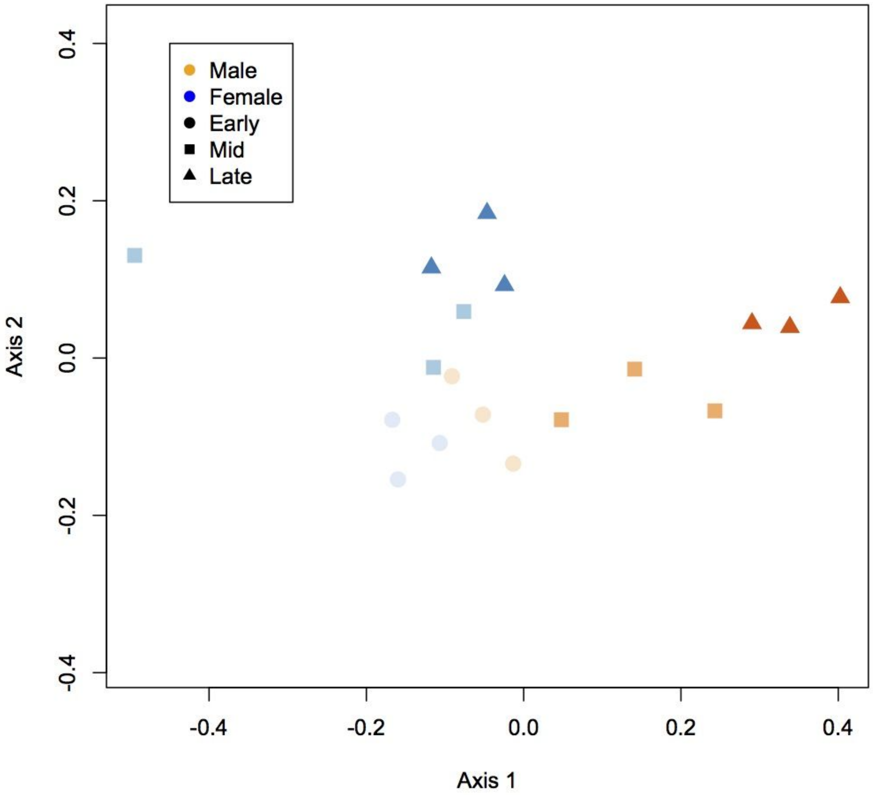


Figure A. NMDS of geoduck gonad peptide transition abundance generated by selected reaction monitoring. Gonad proteomes differ among clams by both sex (male = orange, female = blue) and stage (early = circles, mid = squares, late = triangles; p<0.05).

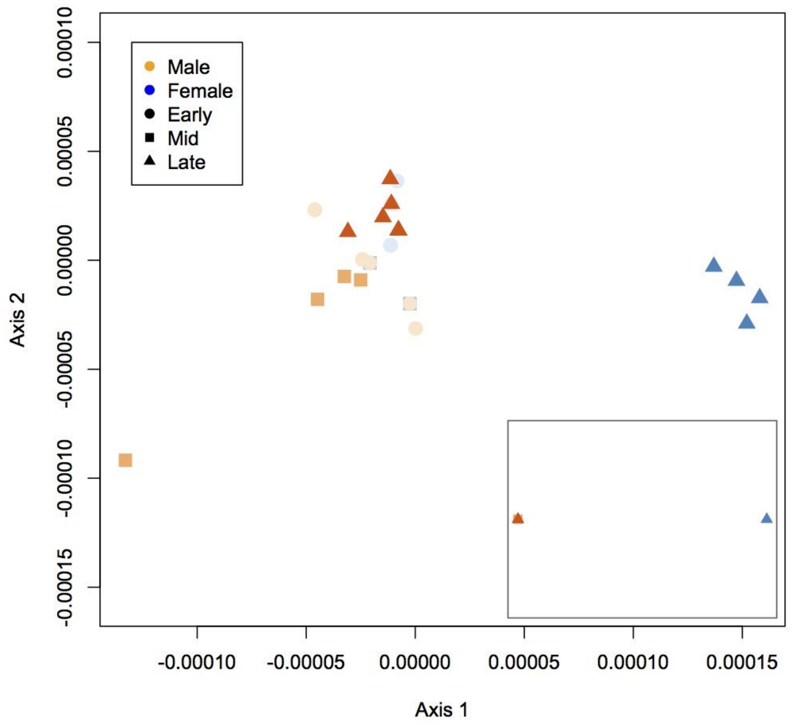


Figure B. NMDS of geoduck hemolymph peptide transition abundance generated by selected reaction monitoring. The main plot contains a point for each technical and biological replicate; the inset (lower righthand corner) shows the same data in a NMDS with the technical replicates averaged. Hemolymph proteomes differ among clams by sex (male = orange, female = blue; p<0.05).

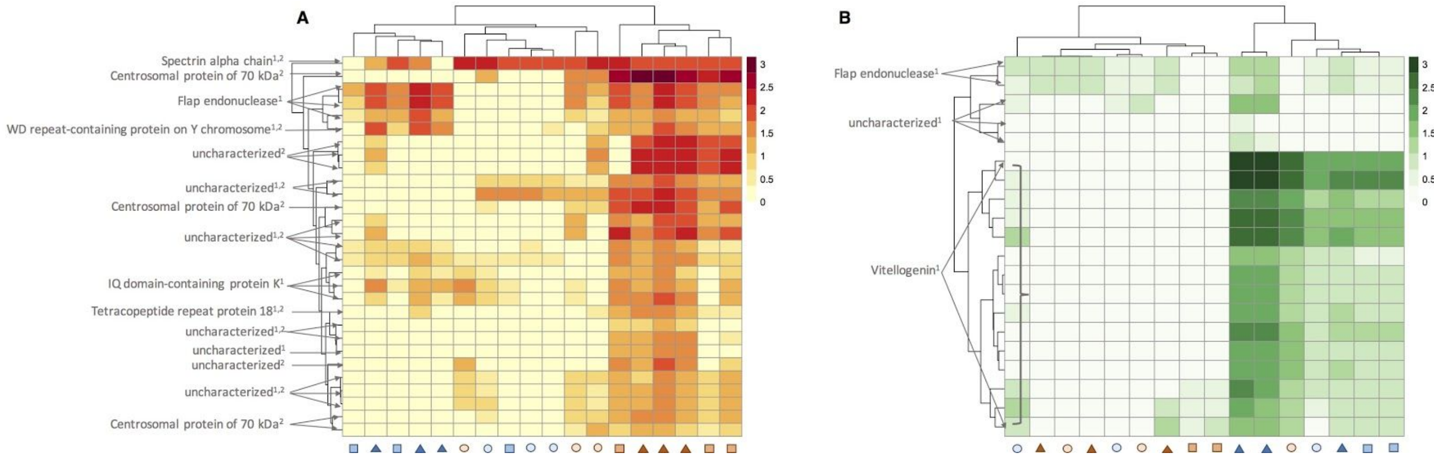


Figure. Heatmaps of log-normalized peptide transition intensities for the top 20 most significant transitions (based on NMDS analysis) across all samples (early/circles, mid/squares, and late/triangles males/orange and females/blue) for A) gonad and B) hemolymph. Each row (peptide transition) is labeled by its corresponding protein annotation in gray. Superscripts indicate for which NMDS axis a transition is significant.

Discussion -

Intro paragrpah - -limited use in ecological systems - omics
Biology sex stage unique
Informative biomarks gonad - early late male female
Hemolymph works

Proteomics has the capacity to uncover molecular physiological processes underlying functional phenotypic change, such as the sexual maturation of reproductive tissue. We have characterized the geoduck clam gonad proteome throughout reproductive maturation for both males and females. Our data-dependent analysis (DDA) approach yielded 3,627 detected proteins across both sexes and three maturation stages. Based on the DDA data, 41 proteins from early and late stage male and female clams were selected for selected reaction monitoring (SRM), or targeted proteomics, to develop biomarker assays of geoduck reproductive stage based on 96 peptides. These proteins yielded 212 peptide transitions that were candidates for biomarker development in gonad and 171 in hemolymph. The gonad peptide transitions could be decreased to a set of 29 transitions that would accurately predict a) if a geoduck is male or female and b) how far it is from reproductive maturity. In the hemolymph, 20 peptide transitions can accurately differentiate late female geoduck from others. These molecular tools can directly address the geoduck production problem of asynchronous spawning due to different maturation stages and unknown sexes.

Transcriptome/Proteome
[Steven's space!}

A phylogenetic analysis of three sperm-egg interaction proteins revealed that sequence similarity between related taxonomic groups is generally higher than between more distantly related groups. However, across the three proteins, the highest percent identity between two species is 50%, suggesting high levels of sequence divergence at the protein level. In a comparison of four yeast species, a gene involved in sporulation (i.e. gamete production) was identified as a fast-evolving gene and had only 32% nucleotide sequence identity and 13% amino acid identity across species (Kellis et al. 2003). [CITE protein sequence divergence between species?] These particular proteins were selected because they are considered "fast evolving" since they are involved in the species-specific process of sperm and egg recognition [CITE Swanson paper]. These proteins are the foundation of maintaining the integrity of a species by avoiding hybridization. None of these proteins were detected in the DDA experiment (the sequence data were taken from the translated transcriptome), so they are either expressed at low levels or are not translated until gametes are released into the water column. [proteins not translated until gametes released?]

Immunogenetics
January 2005, Volume 56, Issue 10, pp 683–695
Protein synthesis during hormonally induced meiotic maturation and fertilization in starfish oocytes

Possible roles of sex steroids in the control of reproduction in bivalve molluscs

The relationships between early ionic events, the pattern of protein synthesis, and oocyte activation in the surf clam, *Spisula solidissima*

DDA

There are clear proteomic profile differences in the geoduck gonad by both sex and maturation stage and these differences elucidate the biochemical pathways underlying tissue reproductive specialization. More proteins are differentially abundant between early and late stages of reproductive maturity, compared to early and mid-stages, for both males and females, reflecting that as maturation progresses, gonadal tissue becomes highly specialized to male or female reproduction. For example, the zona pellucida protein (cds.comp134923_c0_seq3|m.23445) and vitellogenin (cds.comp144315_c0_seq1|m.50928) become significantly more abundant as female geoduck become more mature.

The sex- and stage-specific proteomes were enriched for specific biological processes, further demonstrating tissue specialization. In females, the early-stage gonad proteome was enriched in proteins involved in protein phosphorylation, the mid-stage was enriched in regulation of hydrolase activity and carbohydrate metabolism, and the late stage was enriched in protein glycosylation, signal peptide processing, DNA replication, and cell adhesion. These shifts in biological pathways clarify the mechanisms behind what is already known about bivalve tissue morphology during maturation. The enrichment in the process "protein phosphorylation" likely represents the activation of translated proteins as the gonad tissue undergoes structural and functional changes. The enrichment of carbohydrate metabolism reflects the increased mobilization of carbohydrates (most likely glycogen, Beukema & De Bruin 1977, Li et al. 2000) to support the energy-intensive process of gonad maturation. As geoduck females enter the late stages of maturation, more proteins are detected that are related to cell replication (the "DNA replication" group) as well as cell adhesion, which is an essential biological process in maturing oocytes across taxa (CITE). These changes represent the diversity in function of female gonad tissue as it undergoes reproductive maturation.

The male tissue similarly demonstrated increased complexity through the enrichment analysis: early male gonad proteome was enriched only in translation processes, the mid-stage male proteome was enriched in ATP metabolic process, and the late stage for nucleotide metabolic process, microtubule-based movement, and cellular macromolecular complex assembly. Translation, which was enriched in every stage in both sexes, signifies the physiological need in the gonad for a larger suite of proteins as it changes form and function. In mid-stage males, increases in proteins involved in ATP metabolic processes suggest a greater metabolic demand in the tissue as it matures. [CITE something about male reproductive tissue and metabolism?] In late-stage males, the enriched processes clearly reflect an increase in meiosis during the last steps of sperm formation. [CITE meiosis and spermatogenesis] These results clearly reflect increased cellular energy metabolism and meiosis during the development of mature, motile sperm.

The phenotypic expression of these proteomic changes are reflected in gonad histology, which reveals egg and sperm development and tissue restructuring as maturation proceeds. This annual process of restructuring tissue to create mature gametes is incredibly energy intensive, which is reflected in the proteomics data. For geoduck, gonad production is almost constantly occurring, with just 1-2 months between spawning and re-initiation of gonadal development (Sloan and Robinson 1984) and in some populations two spawning events occur (Calderon-Aguilera et al, 2013; Marshall et al., 2012).

We identified 3,627 high confidence proteins across both sexes and all maturation stages, reflecting the dynamism of the tissue physiology as well as the suitability of our tissue- and species-specific protein identification database. These two reasons - one biological and one bioinformatic - allowed us to detect, identify, and explore many more proteins than previous, similar studies. Throughout geoduck sexual maturation (which lasts approximately X days), the gonad goes from undifferentiated connective tissue to a highly specialized reproductive tissue that can release fully mature sperm and eggs. These dramatic phenotypic changes required for reproductive maturity are realized via changes at the molecular level that prompt the biochemical shifts necessary to create mature gonad. In sequencing the gonad transcriptome to use as our protein identification database, we were able to identify many more peptides, allowing for a fuller characterization of the dynamic maturation process (i.e. Brautigaum et al., XXX;

Timmins-Schiffman et al., in press). In these other studies only this many proteins were identified because of other experimental choices that were made...

[direct comparisons with other invertebrate reproduction proteomes]

Even with a highly specific protein identification database, the DDA method excludes many important, relatively low abundance proteins from analysis (yeast proteome citation). The ten most abundant proteins (as measured by total spectral count) for each sex-maturation stage accounted to 14-26% of the total spectral counts across all proteins for a given sex-stage. Many of these proteins are “housekeeping” proteins, such as actin and myosin, and their dominance in the peptide mixture is likely masking many informative, low abundance peptides. Similarly, in zebrafish, highly abundant vitellogenin (also highly abundant in the MF and LF in this study) made it difficult to detect lower abundant proteins using DDA (Groh et al., 2011). Vitellogenin can be an informative biomarker, by masking the detection of other proteins that may have a more nuanced expression profile, it and other highly abundant proteins can obscure the finer-scale changes that occur.

SRM

DDA analysis yields biologically relevant results, but for practical purposes of biomarker development we do not need the abundance profiles of >3000 proteins. Selected reaction monitoring allows for the selection and accurate detection of only the most informative proteins' peptides to address a specific hypothesis. We leveraged the DDA dataset to create informative SRM assays of geoduck sex and maturation stage in both gonad and hemolymph. Both of these tissues have the potential for being the basis of non-lethal assays for broodstock assessment in aquaculture.

In the gonad tissue, we measured 212 peptide transitions that could differentiate males and females as well as early and late maturation stages. Since we selected these proteins based on data analysis, and not assumptions as to which pathways would be informative, they represent functional groups that do not appear to be directly related to reproduction. In contrast, a simple assay to determine female sea turtle reproductive status was developed for vitellogenin alone based on sequence homology in closely related species (CITE). The sea turtle study followed a different route for SRM development in a non-model organism, however their goal and method led to an assay with a more limited application. Our assay is based on proteins that are involved in calcium ion binding, phosphorylation, meiosis, DNA replication and repair, and membrane structure. Additionally, seven of these informative proteins are unannotated, so we cannot even guess at their function based on homology. A subset of the peptide transitions from these proteins (n=29) were deemed highly informative based on our analysis and would be good candidates for a more streamlined SRM assay. Reproductive maturation is a complex process and involves suites of interacting proteins to achieve the end goal of mature gametes. By applying an unbiased method in the first step of our biomarker development (DDA) we were able to detect and include these informative peptides in our assay.

Hemolymph is another option for a non-lethal tissue sample and proved to have differentially abundant peptides based on sex and reproductive stage. We measured the peptide transitions developed for gonad along with some others that we suspected would be circulating in the hemolymph and related to reproductive maturation. A similar method was used to discover protein biomarkers of prostate cancer in the blood based on original detection in the prostate (Klee et al. 2012). Even if the same proteins were circulating in the hemolymph as were detected in the gonad, they may not be informative because 1) the proteins could be fragmented in ways that make them undetectable with our assay or 2) the protein abundance could be changing out of sync with reproductive changes. We were able to collect data on 171 transitions (other transitions were undetectable in hemolymph) that statistically differentiated late-stage females from the other clams. Peptide transitions from 16 proteins were reliably detected in both gonad and hemolymph, however, only two proteins had transitions that were highly informative in both tissues: flap endonuclease and an uncharacterized protein. The inclusion of vitellogenin in our assay confirmed its presence in geoduck hemolymph during later stages of reproductive maturity in females, but since it was such a strong physiological signal it dominated our statistical analysis. Since we were limited to the proteins from the gonad DDA dataset for our assay development, there are likely additional circulating proteins that would expand our hemolymph assay and make it informative beyond distinguishing late-stage females.

Conclusions

Acknowledgements

Sonia, Jarrett, Lindsay, GSIT, Jimmy, Priska, Yuval, Nick

MR is supported by the National Institute of General Medical Sciences of the National Institutes of Health (P41 GM103533) and the University of Washington Proteomics Resource (UWPR95794).

Supplemental files

1. Protein identification database derived from gonad transcriptome sequencing.
2. Parameter file used in Comet searches for the DDA data.
3. NMDS of the whole proteomic profiles (DDA data) for the three biological replicates and corresponding three technical replicates for each sex and maturation stage. Blue points represent female proteomes and orange represent male. Shapes and shade (light to

dark) represent the different stages: circles for early, squares for mid, and triangles for late.

4. Box plot of the coefficients of variation for protein spectral counts from DDA data A) across technical replicates for each geoduck biological replicate and B) across biological replicates for each sex-maturation stage. CVs were calculated in R using the raster package. X-axis labels correspond to the geoduck sex and maturation stage, indicated by e.g. “EF3” represents early female sample 3. The boxes represent the upper and lower quartiles of the data distribution; horizontal black line represents median value; “whiskers” extend to the greatest and least values, excluding outliers; open circles represent outliers (more or less than 3/2 times the upper or lower quartiles).
5. Input files for running the enrichment analysis. Each tab in the spreadsheet has a list of geoduck protein IDs for each sex and stage. The tabs labeled e.g. “LM” are all the proteins detected in the late male gonad proteome whereas the tabs labeled e.g. “LM only” are the proteins that were detected exclusively in the late male proteome.
6. NMDS of gonad SRM data for the 3 biological replicates and corresponding three technical replicates for each sex and maturation stage. Blue points represent female proteomes and orange represent male. Shapes and shade (light to dark) represent the different stages: circles for early, squares for mid, and triangles for late.
7. Identifying information for the three proteins used for phylogenetic comparison of egg-sperm interacting proteins. The geoduck protein ID is provided, along with additional geoduck protein IDs that had the same Uniprot annotation. The Uniprot ID and protein annotation are provided, along with the NCBI accession numbers of the proteins used for the comparison from the other species.
8. All identified proteins from the DDA experiment with Uniprot annotations (e-value cut-off of 1E-10), total spectral counts for each technical replicate, calculated normalized spectral abundance factor (NSAF) for combined technical replicates, and indication of significantly differentially abundant proteins by pairwise comparison. Columns containing spectral count data have headers “SpC” followed by the biological replicate number, sex, maturation stage, and technical replicate (for example, “SpC 3FE” is technical replicate 1 from early female 3). Notation for NSAF is similar to SpC. The last 9 columns in the sheet have headers such as “EFvLF” (comparison between early and late females) and asterisks in the cells that correspond to proteins that were differentially abundant for each given comparison.
9. Summary of enrichment analysis report output for GO biological processes for both sexes and all stages. For each sex-maturation stage combination, a list of significantly enriched GO terms is provided with GO number, name, and significance value in parentheses.
10. Enrichment plots for early and mid-maturation stage geoduck (EM, EF, MM, MF). Processes colored in blue were enriched from the set of all proteins detected in the given

sex-stage, while red processes are from the protein set that was uniquely detected in a sex-stage. Darker colors represent higher significance of enrichment.

11. Raw Skyline output (in the tab "Skyline output) and peak intensities normalized by QC peptide abundance ("Normalized Intensities") for the gonad and hemolymph SRM data.
12. Integrated peak area coefficient of variation values across all gonad peptide transitions (n=212) and hemolymph peptide transitions (n=171) A) across technical replicates for each geoduck biological replicate in gonad tissue, B) across biological replicates for each sex-maturation stage in gonad tissue, C) across technical replicates for each geoduck biological replicate in hemolymph, and D) across biological replicates for each sex-maturation stage in hemolymph. CVs were calculated in R using the raster package. X-axis labels correspond to the geoduck sex and maturation stage, indicated by e.g. "EF3" represents early female sample 3. The boxes represent the upper and lower quartiles of the data distribution;horizontal black line represents median value;"whiskers" extend to the greatest and least values, excluding outliers; open circles represent outliers (more or less than 3/2 times the upper or lower quartiles).

Other figures/tables from my capstone: (Grace)

Females						
Stage 1	**Stage 2**	**Stage 3**	**Stage 4**	**Stage 5**	**Stage 6**	**Stage 7**
Geo-01	Geo-15	Geo-48 **	Geo-31	Geo-25	Geo-37	Geo-57
Geo-03	Geo-18	Geo-55	Geo40	Geo-34	Geo-50	Geo-61
Geo-04	Geo-19	Geo-64		Geo-35	Geo-51	
Geo-05	Geo-21	Geo-66		Geo-38	Geo-69	
Geo-06	Geo-22			Geo-39 **	Geo-70	
Geo-08	Geo-23			Geo-44		
Geo-14	Geo-24			Geo-45		
Geo-29				Geo-58		
Geo-30						

**Geoduck tissue specimens that were incorrectly fixed. There are likely a few more than those marked, but were not discernible from the correctly fixed tissues.

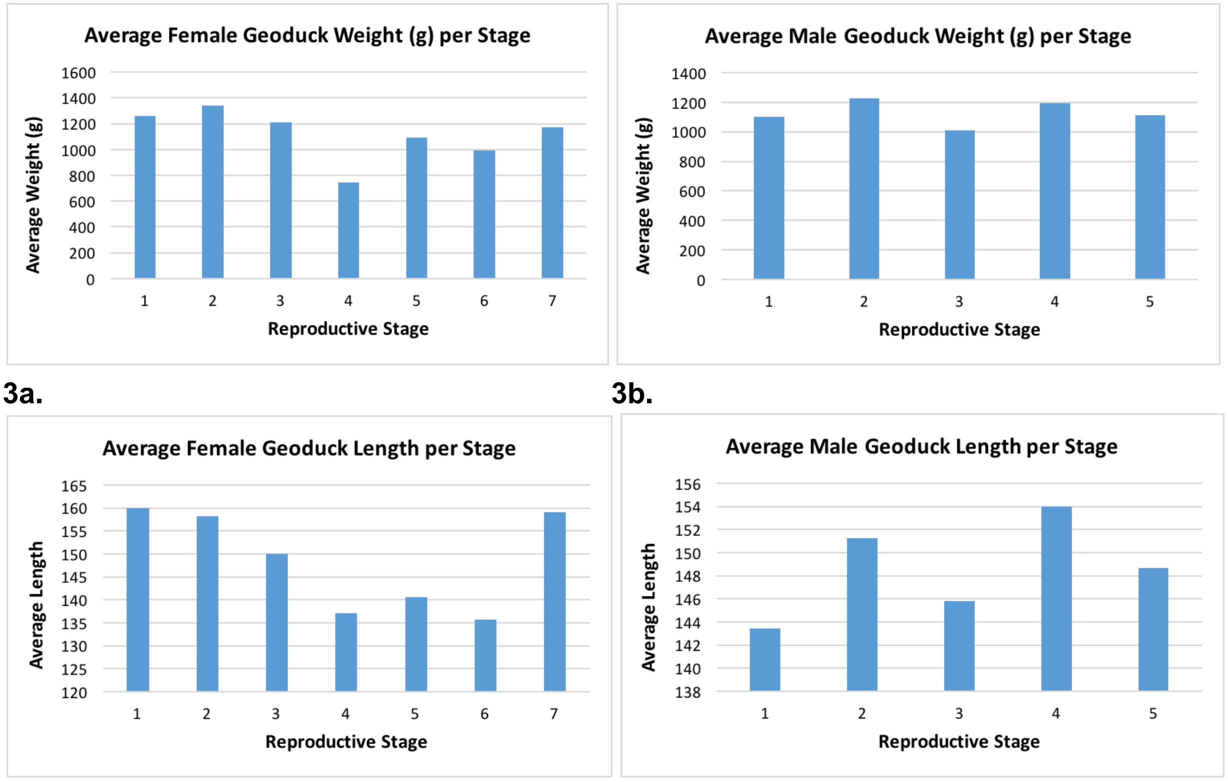
Table 3. Results of female sex-determination of histology slides and placement into reproductive stages. The number after "Geo_" represents the order in which the geoduck was sampled from week one.

Males				
Stage 1	**Stage 2**	**Stage 3**	**Stage 4**	**Stage 5**
Geo-02	Geo-10	Geo-41	Geo-33	Geo-52
Geo-07	Geo-11	Geo-42	Geo-43	Geo-62

Geo-09	Geo-13	Geo-46	Geo-49	Geo-63
Geo-12	Geo-17	Geo-47	Geo-53 **	Geo-65
Geo-16	Geo-20	Geo-54	Geo-56	Geo-67
Geo-26	Geo-27		Geo-59	Geo-68
Geo-28	Geo-32		Geo-60	
	Geo-36			

** Geoduck tissue specimens that were incorrectly fixed. There are likely a few more than those marked, but were not discernible from the correctly fixed tissues.

Table 4. Results of male sex-determination of histology slides and placement into reproductive stages. The number after "Geo_" represents the order in which the geoduck was sampled from week one.



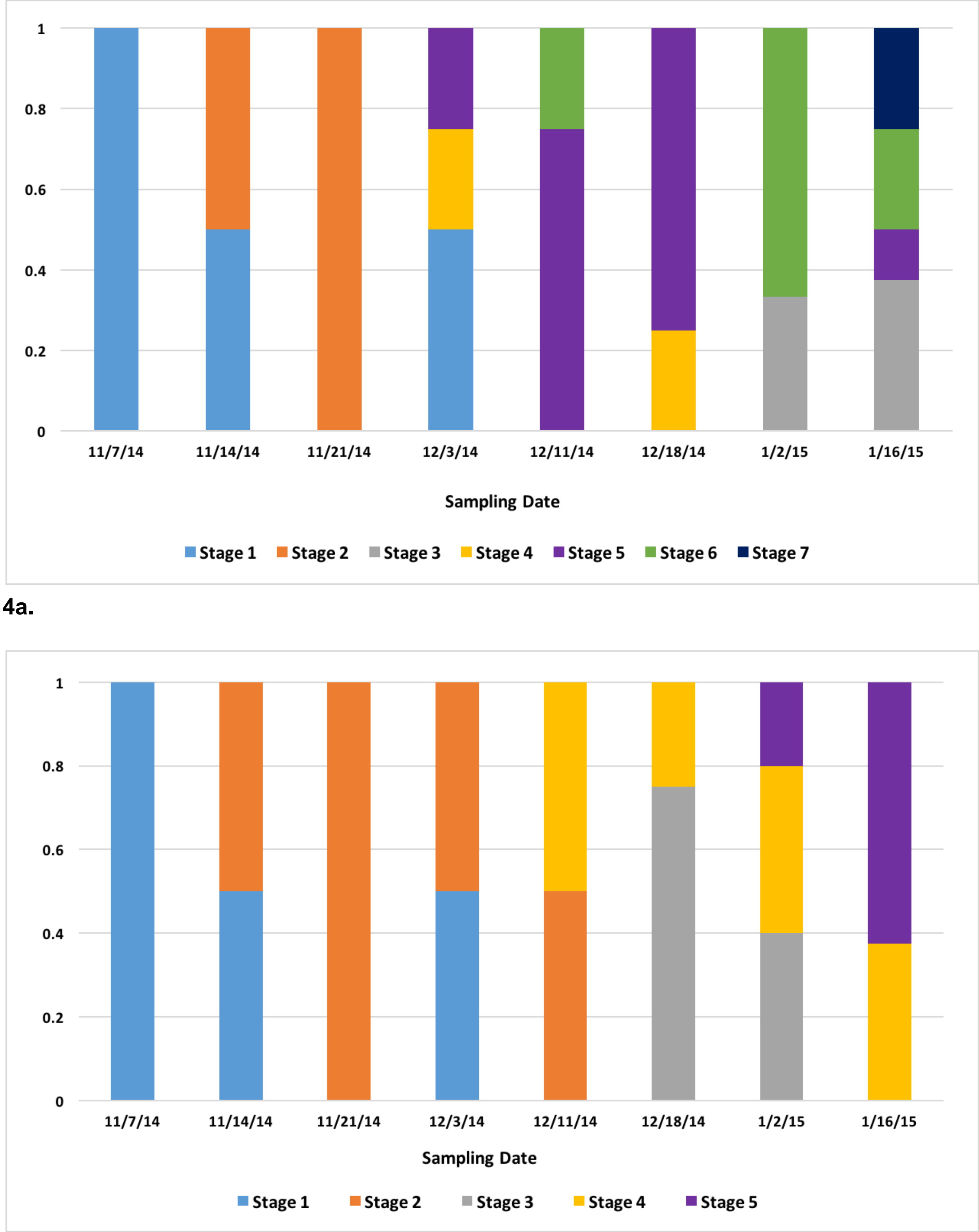
3a.

3b.

3c.

3d.

Figure 3. The average geoduck weights per reproductive stage in females (**3a**) and males (**3b**) and the average geoduck lengths per reproductive stage in females (**3c**) and males (**3d**). There is no clear correlation between average weight or length and the reproductive stage for either sex.



4a.

4b.

Figure 4a and b. The proportion of geoducks at each reproductive stage across the sampling weeks for females (**4a**) and males (**4b**). There was a general trend of increasing reproductive ripeness for both sexes as time progressed. However, there are also some variations. In both sexes, Stage 4 geoducks appear before Stage 3. Also, Stage 1 geoducks reappear for both sexes in December.