

Article

Integrating discovery-driven proteomics and selected reaction monitoring to develop a non-invasive assay for geoduck reproductive maturation

Emma B. Timmins-Schiffman, Grace A Crandall, Brent
Vadopalas, Michael E. Riffle, Brook L. Nunn, and Steven Roberts

J. Proteome Res., **Just Accepted Manuscript** • DOI: 10.1021/acs.jproteome.7b00288 • Publication Date (Web): 21 Jul 2017

Downloaded from <http://pubs.acs.org> on August 3, 2017

Just Accepted

"Just Accepted" manuscripts have been peer-reviewed and accepted for publication. They are posted online prior to technical editing, formatting for publication and author proofing. The American Chemical Society provides "Just Accepted" as a free service to the research community to expedite the dissemination of scientific material as soon as possible after acceptance. "Just Accepted" manuscripts appear in full in PDF format accompanied by an HTML abstract. "Just Accepted" manuscripts have been fully peer reviewed, but should not be considered the official version of record. They are accessible to all readers and citable by the Digital Object Identifier (DOI®). "Just Accepted" is an optional service offered to authors. Therefore, the "Just Accepted" Web site may not include all articles that will be published in the journal. After a manuscript is technically edited and formatted, it will be removed from the "Just Accepted" Web site and published as an ASAP article. Note that technical editing may introduce minor changes to the manuscript text and/or graphics which could affect content, and all legal disclaimers and ethical guidelines that apply to the journal pertain. ACS cannot be held responsible for errors or consequences arising from the use of information contained in these "Just Accepted" manuscripts.

Integrating discovery-driven proteomics and selected reaction monitoring to develop a non-invasive assay for geoduck reproductive maturation

Emma B. Timmins-Schiffman¹, Grace A. Crandall², Brent Vadopalas², Michael E. Riffle³, Brook L. Nunn¹, Steven B. Roberts²*

¹ Department of Genome Sciences, University of Washington, Seattle, WA 98105

² School of Aquatic and Fishery Sciences, University of Washington, Seattle, WA 98105

³ Department of Biochemistry, University of Washington, Seattle, WA 98105

*Corresponding author:

Steven B. Roberts, School of Aquatic and Fishery Sciences, University of Washington, 1122 NE

Boat Street, Seattle WA 98105. email: sr320@uw.edu

Abstract

Geoduck clams (*Panopea generosa*) are an increasingly important fishery and aquaculture product along the eastern Pacific coast from Baja California, Mexico to Alaska. These long-lived clams are highly fecund, though sustainable hatchery production of genetically diverse larvae is hindered by the lack of sexual dimorphism, resulting in asynchronous spawning of broodstock, unequal sex ratios, and low numbers of breeders. Development of assays of gonad physiology could indicate sex and maturation stage, as well as be used to assess the status of natural populations. Proteomic profiles were determined for three reproductive maturation stages in both male and female clams using data dependent acquisition (DDA) of gonad proteins. Gonad proteomes became increasingly divergent between males and females as maturation progressed. The DDA data was used to develop targets analyzed with selected reaction monitoring (SRM) in gonad tissue as well as hemolymph. The SRM assay yielded a suite of indicator peptides that can be used as an efficient assay to determine geoduck gonad maturation status. Application of SRM in hemolymph samples demonstrates this procedure could effectively be used to assess reproductive status in marine mollusks in a non-lethal manner.

Key words: reproduction, gonad, geoduck, hemolymph, mollusk, proteomics, SRM, targeted proteomics

Introduction

Wild Pacific geoduck clams (*Panopea generosa*), like other suspension feeders, provide vital ecosystem services as both primary consumers of phytoplankton and biodepositors.¹ Additionally, commercial geoduck fisheries provide significant social benefits as the most economically important clam fishery in North America.² Commercial farming of geoduck now generates over 20 million dollars in annual sales and geoduck are among the most valuable farmed shellfish on a per acre basis.³ The rapid development of geoduck aquaculture close to wild populations of conspecifics has raised concerns over farmed-wild interbreeding and the subsequent loss of genetic diversity. To address these concerns, geoduck hatcheries are currently encouraged to use large numbers of wild broodstock to maximize genetic diversity in farmed geoducks. An ongoing impediment to fulfilling this conservation goal is the absence of nonlethal identification methods for sex and maturation stage determination in broodstock. Geoducks are not sexually dimorphic, thus sex is not accurately determined until spawning, which can be induced through increase in temperature and microalgae ration. Further, at a given event, only a few of the broodstock mature and spawn. Highly skewed sex ratios in some populations and the inability to determine when to induce spawning based on maturation state present additional challenges.

To date, only a few studies have investigated molecular or biochemical analyses to identify sex or reproductive stages in marine invertebrates. For example, studies on vertebrate steroids explored these organic molecules as an analyte for tracking mollusk reproduction processes and stages.⁴ There is suggestive evidence of steroidal metabolism and response in mollusks, but these pathways remain poorly understood in terms of biochemical mechanism and biological function.⁵ Further research on gonad lipids⁶, specific gene transcripts^{7,8}, and

1
2
3 RNA/DNA ratios ⁹ have clarified some of the physiology underlying molluscan sexual
4 maturation. A comprehensive study of the Pacific oyster gonadal gene expression across
5 reproductive stages further explained the molecular mechanisms at play ¹⁰; however, none of
6 these studies translated physiological discovery into an applied assay.
7
8
9
10
11

12
13 Previous studies have demonstrated the promise of protein based assays for
14 characterization of reproductive status in bivalves. Arcos et al. ¹¹ explored the use of enzyme-
15 linked immunoassay assays (ELISA) to quantify proteins associated with vitellogenesis to
16 determine oyster oocyte maturation. Although the two proteins targeted by Arcos et al. ¹¹
17 allowed distinction of males and females and differentiation of female reproductive stages, the
18 assay relies on a single physiological measure and cannot be adapted to a non-lethal assay. ⁴ The
19 complexity of molluscan reproductive maturation at the molecular level has been proven by
20 global profiling studies ^{10,12} that reveal possibilities for an informative assay that leverages the
21 involvement of diverse biochemical pathways.
22
23
24
25
26
27
28
29
30
31
32

33
34 Here, we apply mass spectrometry (MS)-based proteomics technology to provide a large-
35 scale, unbiased approach for examining relative abundances of all proteins present in gonad
36 tissue across several maturation stages of both female and male geoducks. Whole proteome
37 characterization via data dependent acquisition (DDA) provided basic information on
38 gametogenesis, gonad maturation, and informed peptide selection for the determination of sex
39 and maturation stage in subsequent downstream MS analyses. Selected peptides underwent
40 targeted proteomic analyses via selected reaction monitoring (SRM) to more accurately quantify
41 protein abundance. Together, these data provide fundamental insight into marine mollusk
42 reproductive biology and demonstrate effective non-invasive peptide quantification on
43 circulating hemolymph.
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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-, mid-, and late-stage gonad maturation, from both males and females, were characterized histologically. Female reproductive maturation stages were categorized as follows: early-stage (no secondary oocytes, or oocytes that measure $\sim 5\text{-}15\mu$), mid-stage (secondary oocytes $\sim 50\text{-}70\mu$), and late-stage (secondary oocytes $\sim 65\text{-}85\mu$). Male reproductive maturation stages were characterized as follows: early-stage (mostly somatic cells and $\sim 5\%$ spermatid composition per acinus), mid-stage (about equal parts somatic cells and reproductive tissue and $\sim 50\%$ spermatid composition per acinus), and late-stage (very little somatic cells and $\sim 75\text{-}90\%$ spermatid composition per acinus). Details of gonadal maturation classification and histological details have been previously described.¹³ Gonad tissue samples were taken from 3 early-stage females (fG03, fG04, fG08), 3 mid-stage females (fG34, fG35, fG38), 3 late-stage females (fG51, fG69, fG70), 3 early-stage males (mG02, mG07, mG09), 3 mid-stage males (mG41, mG42, mG46), and 3 late-stage males (mG65, mG67, mG68) with identification codes corresponding to those reported in Crandall et al.¹³ Hemolymph tissue samples were taken from 3 early-stage females (fG18, fG29, fG30), 2 mid-stage females (fG25, fG35), 3 late-stage females (fG51, fG69, fG70), 3 early-stage males (mG17, mG20, mG028), 2 mid-stage males (mG42, mG46), and 3 late-stage males (mG65, mG67, mG68). All tissue samples corresponded to clams with a known sex and maturation stage, detailed in the histology in Crandall et al.¹³

Protein Preparation

Gonad tissue from the eighteen geoduck clams was sonicated in lysis buffer (50 mM NH_4HCO_3 , 6 M urea). Gonad tissue protein homogenate content and hemolymph protein content were quantified using Pierce BCA Protein Assay Kit (Thermo Fisher Scientific, Waltham, MA; catalog #23225). Protein (100 μg - gonad samples, 50 μg - hemolymph) was evaporated to ~20 μl and resuspended in 100 μl of lysis buffer followed by sonication. Protein digestion for gonad tissue samples and hemolymph followed the protocol outlined in Timmins-Schiffman et al.¹⁴ Briefly, each sample was incubated with 200 mM tris(2-carboxyethyl)phosphine buffered at pH 8.8 (1 hr, 37°C). Samples were alkylated with 200 mM iodoacetamide (IAM; 1hr, 20°C) followed by an incubation (1 hr) with 200 mM dithiothreitol to absorb any remaining IAM. To each sample, NH_4HCO_3 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) was performed to examine gonad proteomic profiles across samples (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 (20 cm long) was packed in-house with C18 beads (Dr. Maisch GmbH HPLC, Germany, 0.3 μ m) with a flow rate of 0.3 μ l/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 of 2-5% solvent A; 1-60 minutes 5-35% solvent A; 60-61 minutes 35-80% solvent A; 61-70 minutes 80% solvent A; 71-90 minutes 80-2% solvent 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.

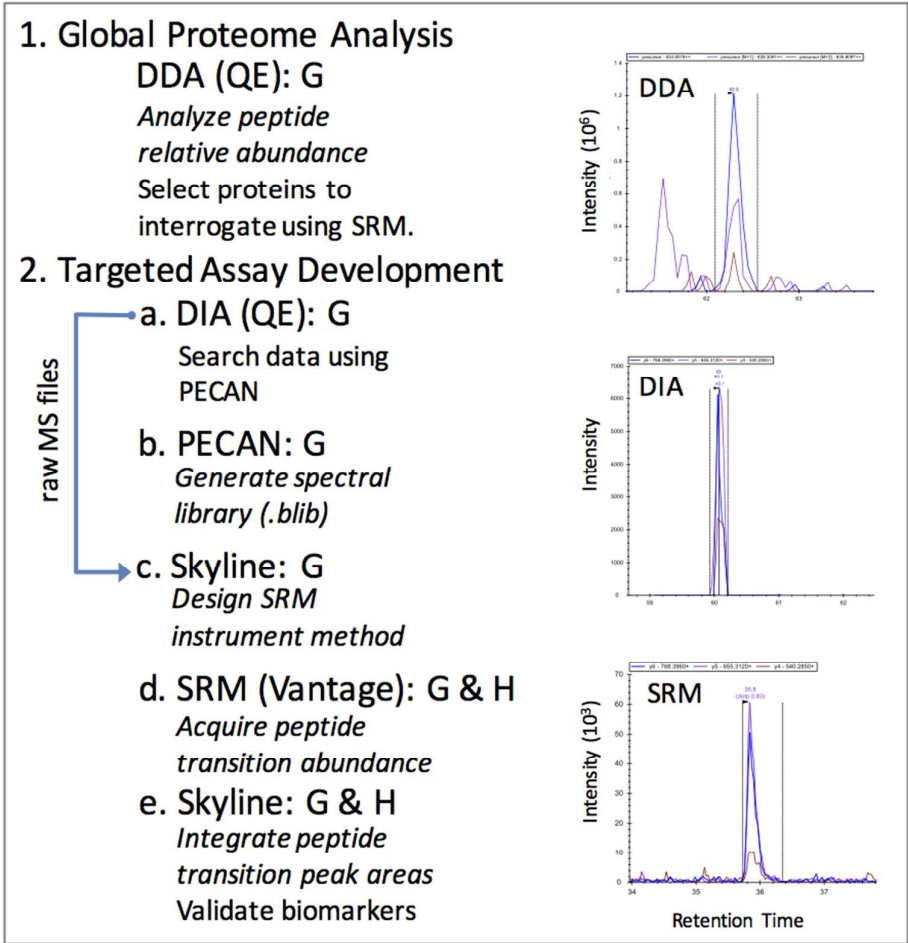


Figure 1. Illustration of experimental setup and workflow for mass spectrometry data acquisition and analysis. MS experimental workflow: 1. Data dependent acquisition (DDA) was performed on the Q-Exactive-HF (QE) to assess the global proteomic differences in gonad (G) tissue between sexes and maturation 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 (G & H); (e) Peptide transition detection and quantification was performed in Skyline. The chromatograms of peptide KEEELIDYMKQ (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 Trinity¹⁵ and deduced protein sequences determined using the Transdecoder algorithm within Trinity. Raw mass spectrometry data (PRIDE Accession #PXD003127) was searched against the protein sequences using Comet v 2016.01 rev.2.^{16,17} Parameters used with Comet included concatenated decoy search, 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) to find peptide spectral matches (all parameters can be found in Supporting information 2). Protein inference and match probability were found using the Trans-Proteomic Pipeline.^{18,19} Data files were combined and normalized spectral abundance factor (NSAF)²⁰ was calculated in Abacus²¹ to determine consistent protein inferences across replicates. Protein identifications were considered true matches when the probability of the match was at least 0.8611 (corresponding to error rate, which approximates FDR, of 0.01) 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 (Supporting Information 3) using the vegan package²² in R v. 3.2.3²³. NMDS was performed on log-transformed data using a Bray-Curtis dissimilarity matrix (Supporting Information 2). As technical replicates clustered closely together and showed less variability than biological replicates (Supporting Information 3), NSAF was averaged across each sample (n=18). Proteomic 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) Fisher's exact test 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 sexes and maturation stages. Differentially abundant proteins among the six conditions were identified using Fisher's exact test in Excel. Spectral counts were summed across technical replicates to compare biological replicates. Nine comparisons were analyzed for differentially abundant proteins: Early-stage male (EM) vs. early-stage female (EF); mid-stage male (MM) vs. mid-stage

female (MF); late-stage male (LM) vs. late-stage female (LF); EM vs. MM; EM vs. LM; MM vs. LM; EF vs. MF; EF vs. LF; MF vs. LF. All p-values resulting from the test were multiplied by 9 as a Bonferroni correction for multiple comparisons. Proteins with a corrected p-value ≤ 0.05 were considered differentially abundant.

Enrichment Analysis

We used enrichment analysis to determine biological processes (represented by detected proteins) predominant in each gonad sample. GO enrichment analysis was performed for each maturation stage (for each sex) on 1) unique proteins and 2) all detected proteins. In both cases, the entire detected proteome was used as the background. Enrichment analysis was also performed on sets of differentially abundant proteins. Code used to perform enrichment analysis is available in a corresponding GitHub repository (<https://github.com/yeastrc/compgo-geoduck-public>) as is the underlying code for the end user web-interface

(http://yeastrc.org/compgo_geoduck/pages/goAnalysisForm.jsp).

Briefly, a p-value was calculated describing the enrichment of a GO term in a set of tested proteins above what would be expected by chance, 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 gonad protein names resulting from the *de novo*-assembled transcriptome. This was done by assigning GO terms associated with Uniprot-KB Swiss-Prot BLASTp hits. 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:

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

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.

Complete directed acyclic graphs (DAG) that represent a subset of the whole GO DAG were generated to visualize enriched biological processes. These DAGs were then filtered by removing all childless terms that had an associated p-value ≥ 0.01 . The resulting DAGs were then filtered using the same method, and this process repeated until no childless terms remained with a p-value above 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. Proteins that contribute to specific enriched processes for all enrichment analyses are included in Supporting information 4 and visualizations of enrichments for full and unique proteomes for each sex and stage are visualized in a single DAG for each sex and stage (Supporting Information 3). Venn diagrams of enriched processes were produced in Venny²⁴.

Targeted Assay Development: Selected Reaction Monitoring (SRM)

A subset of proteins was chosen for development of a suite of targeted assays using selected reaction monitoring (SRM) on the mass spectrometer. Based on the data dependent acquisition, proteins that were detected in only one gonadal stage (early-stage males (EM), early-stage females (EF), late-stage males (LM), or late-stage females (LF)) or were at considerably higher abundance in one of these stages were screened for usable peptide transitions in SRM in Skyline daily v. 3.5.1.9706²⁵.

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; Pierce). Sample injections for all DIA experiments included protein (1 µg) plus PRTC (50 fmol) in a 2 µl injection. An analytical column (27cm) packed with C18 beads (3 µm; Dr. Maisch) and a trap (3 cm) with C12 beads (3 µm; Dr. Maisch) were used for chromatography. Technical replicate DIA spectra were collected in 4 m/z isolation width windows spanning 125 m/z ranges each²⁶ (400-525, 525-650,

650-775, 775-900). 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 under identifier PXD004921. MSConvert²⁷ was used to generate mzML files from the raw DIA files.

Peptide Centric Analysis was completed with the software program PECAN to generate spectral libraries for targeted method development.²⁸ Input files included the list of peptides generated for SRM (n=217), 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 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 output file (.blib) was imported into Skyline 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. Peptides reliably fragment in the same way in the mass spectrometer, therefore transitions are a robust and consistent signal of a peptide's presence.²⁹ 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. This transition list was divided between two method files for the final SRM analyses to provide adequate dwell time on individual transitions to accurately detect and measure all peptides desired.³⁰

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 μ g of protein per 3 μ l 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=228, 217 of which yielded quantifiable data). Raw data can be accessed in the PeptideAtlas under accession PASS00943.

Hemolymph from early-, mid-, and late-stage males and females was also assayed for sex- and stage-specific biomarkers. An augmented SRM assay was used on the hemolymph that included all the peptide transitions analyzed in the gonad with an additional 40 transitions from five new proteins. 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.³¹ The additional peptides for SRM analysis of hemolymph were from vitellogenin (2 proteins: cds.comp100108_c2_seq1|m.5995, cds.comp144315_c0_seq1|m.50928), glycogen synthase (2 proteins: cds.comp140343_c0_seq2|m.34696, cds.comp141785_c0_seq1|39776), and glycogenin-1 (cds.comp140645_c0_seq3|m.35608). PECAN generated spectral libraries, as described above, and Skyline software selected hemolymph protein transitions to target during SRM analyses. For the new hemolymph proteins, the minimum targets for peptides (≥ 3) and transitions (≥ 2) as described for gonad SRM could not be met for every protein, but were still included in the analysis. These transitions, and the previous gonad transitions (total of n = 254), were analyzed across sixteen 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 that were reliably detected. Raw data can be accessed in PeptideAtlas under accession PASS00942.

Acquired SRM data in gonad and hemolymph were analyzed in Skyline for peptide transition quantification. 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). Accurate Peak detection was determined based on consistency of retention time (verified by spiked in PRTC peptides and correlation with DIA peptide retention times) and peak morphology.

All peptide transition peak intensities were exported from Skyline for automated peak selection and peak integration analysis. PRTC internal standard transitions were monitored for consistency across runs by calculating the 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 PRTC 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 PRTC-normalized SRM dataset as described above for the DDA dataset, except data were $\log(x+1)$ transformed for the gonad SRM dataset. An initial NMDS showed that technical replicates clustered together well and that variation was lower within biological replicates compared to between replicates, therefore technical replicate peak intensities were averaged for the rest of the analysis (Supporting Information 3). 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 raster³² 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 for each MDS axis were selected as biomarkers. Heatmaps of the log-normalized transition intensities for these biomarkers were made using pheatmap³³ in R, clustering rows (peptide transitions) and columns (samples) using euclidean distance and the

average clustering method. Receiver operating characteristic (ROC) curves were calculated for each of the 40 transitions that were determined to be the main drivers distinguishing sexes and/or stages for gonad and hemolymph. ROCs and areas under the curve (AUC) were calculated in the pROC package³⁴ in R as described in Whiteaker et al.³⁵ For each transition that demonstrated increased abundance in one or two maturation stages for a sex, sensitivity, specificity and false positive rate were calculated.³⁵

Results

DDA Proteomics

The mass spectrometry data (PRIDE Accession #PXD003127) interpreted with a species- and tissue-specific transcriptome yielded 3,651 proteins inferred with high confidence across all gonad tissue samples (total spectral count across all replicates > 1) (Supporting information 4). Female and male gonad proteomic profiles were more similar in early-stage maturation, with proteomes diverging as reproductive maturity advanced (Figure 2). Proteomic profiles were significantly different between sexes ($R = 0.4167$, $p = 0.001$) and maturation stages ($R = 0.3494$, $p = 0.002$).

Unique proteomic profiles include the set of proteins that were detected in a specified sex or specific stage within sex in the DDA analysis. The female proteomic profile across all stages contained 156 unique proteins whereas 144 proteins were unique to males. The number of proteins unique to a specific maturation stage increased with maturity from 16 and 18 in early-stage females and males, respectively, to 132 and 104 in late-stage females and males (Figure 3). Enriched GO biological processes for the unique proteins and all detected proteins for each sex and stage are summarized in Table 1. Only the most specific GO terms (i.e., farthest down on the

DAG) are listed; parent terms were frequently enriched as well and can be seen in the DAGs (Supporting Information 3).

Table 1. Enriched biological process GO terms for the unique proteins and all detected proteins in the gonad for each sex and stage. If no GO terms are listed, no enriched processes were identified

Gonad Stage	Detected Proteome GO terms	Unique Proteome GO terms
Early-stage Female	Translation	
Mid-stage Female	Translation	
Late-stage Female	Macromolecule biosynthetic process	Homophilic cell adhesion via plasma membrane adhesion molecules
Early-stage Male	Translation; Small molecule metabolic process	
Mid-stage Male	Translation; Nucleoside metabolic process; Purine-containing compound metabolic process	
Late-stage Male	Translation; Microtubule-based process; ribonucleoside metabolic process; ribonucleoside triphosphate metabolic process	Microtubule-based process

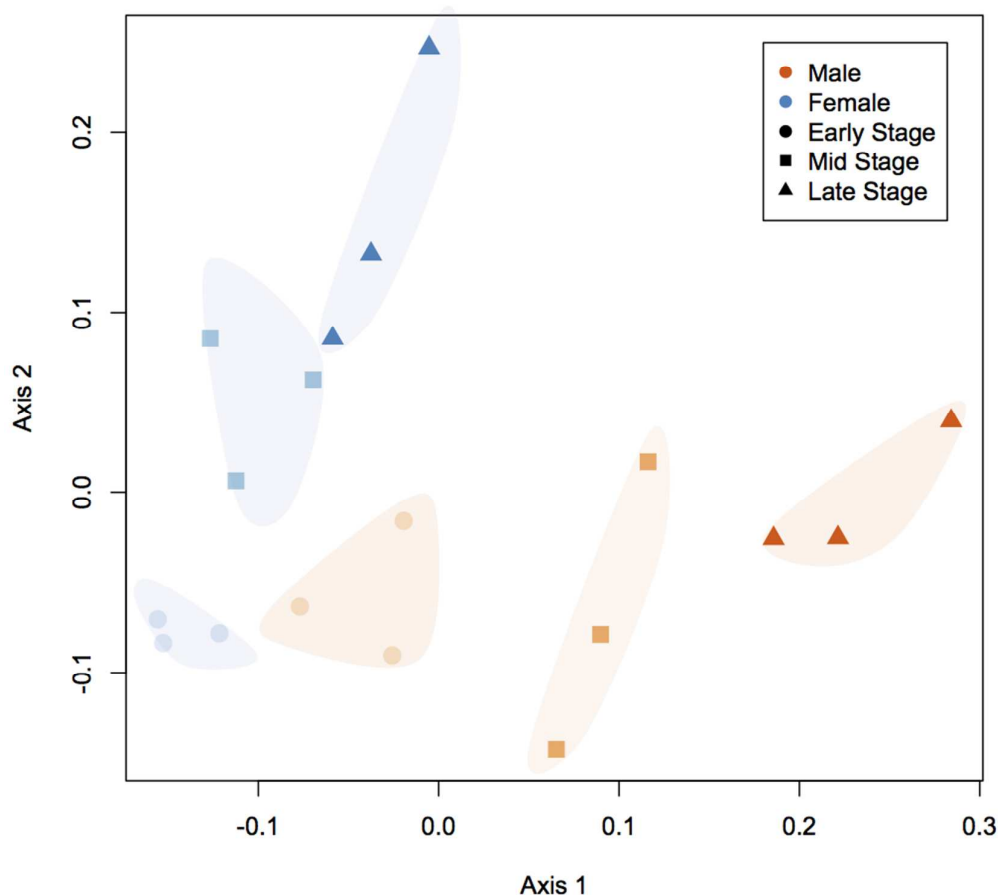


Figure 2. Non-metric multidimensional scaling plot (NMDS) of geoduck gonad whole proteomic profiles generated by data dependent acquisition. Gonad proteomes differ among clams by both sex (male = orange, female = blue) and stage (early-stage = circles, mid-stage = squares, late-stage = triangles; $p < 0.05$).

Numerous proteins were differentially abundant in comparisons between reproductive stage and sex based on analysis with a Fisher's exact test with Bonferroni correction for multiple comparisons. The number of differentially abundant proteins within a stage and between sexes

increased from 102 between early-stage males and females to 1109 between late-stage males and females (Figure 4, Supporting information 4). Enrichment analysis revealed which processes were over-represented in the sets of differentially abundant proteins, when compared to the entire proteome (Figure 4 and Supporting information 4).

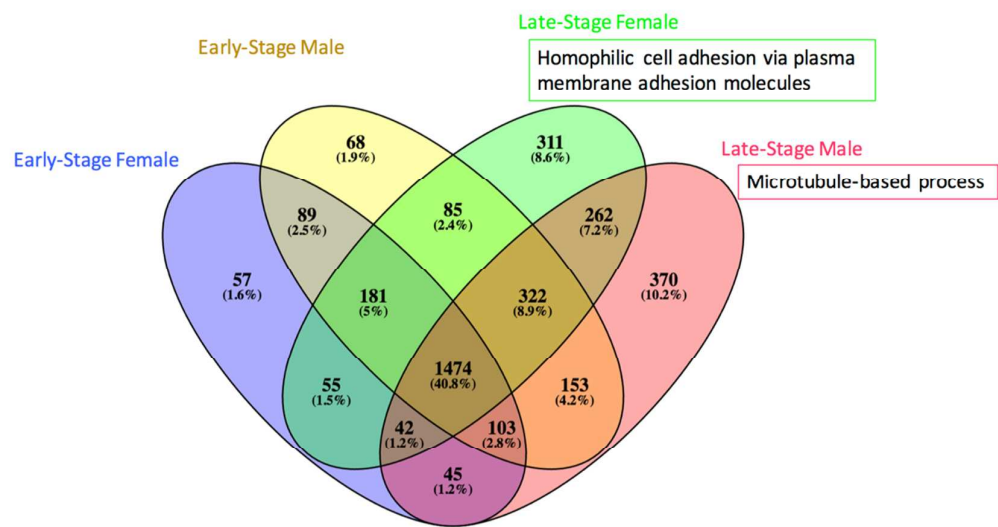
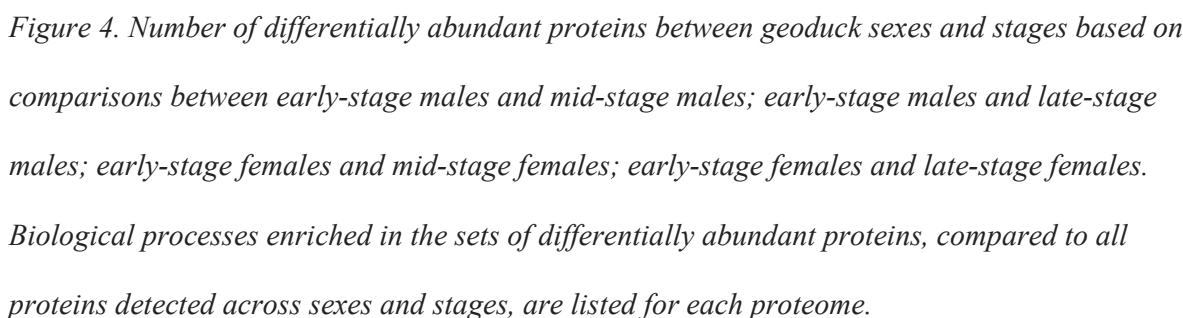


Figure 3. Number of proteins detected in sex and stage specific proteomes for early-stage female (blue), early-stage male (yellow), late-stage female (green), and late-stage male (red). Biological processes enriched in the sex- and stage-specific proteomes compared to all proteins detected across sexes and stages are listed for each proteome.



SRM was applied to create peptide assays that could resolve geoduck sex and maturation stages. Peptide transitions in the gonad (n=217) and hemolymph (n=171) derived from proteins detected in a single sex-maturation stage from one of the following: early-stage male (EM), early-stage female (EF), late-stage male (LM), and late-stage female (LF) proteomes were measured across 18 geoduck gonad samples and 16 hemolymph samples (Supporting information 4). Many transitions had stable coefficients of variation across all biological replicates, although an average of 3% of the peptide transitions had CVs > 100 across technical replicates for gonad samples and 10% were >100 for the hemolymph samples, indicating a need for assay optimization (Supporting Information 3).

1
2
3 In an ordination plot, the gonad SRM data resolved males and females better in late-stage
4 samples than the early- or mid-stage (Supporting Information 3). There was significant
5 separation based on gonad SRM data for sex ($R = 0.2435$, $p=0.002$), stage ($R=0.1984$, $p=0.003$),
6 and a combined sex-stage factor ($R=0.3959$, $p=0.001$). The hemolymph SRM data resolved the
7 late-stage female and mid-stage male groups from the rest of the geoduck (Supporting
8 Information 3). There was significant separation of the hemolymph proteomic profiles based on
9 stage ($R=0.1772$, $p=0.048$) and sex-stage ($R=0.2749$, $p=0.028$), but not by sex alone
10 ($R=0.01786$, $p=0.37$).
11
12

13 Peptide transitions that directed the observed spread of the geoduck sex-stage data along
14 Axis 1 and Axis 2 in the NMDS plots (determined by eigenvector analysis) were considered key
15 analytes for differentiating groups. These peptide transitions represent a starting point for
16 biomarker assay optimization. Peptides that drove the separation of the male maturation stages
17 along Axis 1 in the gonad NMDS plot (Supporting information 4) are from a set of 7 proteins: IQ
18 domain-containing protein K (cds.comp133232_c0_seq1|m.21239), armadillo repeat-containing
19 protein (cds.comp140515_c2_deq1|m.35054), flap endonuclease
20 (cds.comp130260_c0_seq1|m.17926), dynein heavy chain (cds.comp144210_c0_seq2|m.50335),
21 spectrin alpha chain (cds.comp141647_c0_seq1|m.39026), and four uncharacterized proteins
22 (cds.comp135264_c2_seq1|m.24100, cds.comp133422_c0_seq1|m.21474,
23 cds.comp131654_c1_seq1|m.19394, cds.comp144596_c0_seq1|m.53297) (Figure 5). Proteins
24 that played a significant role in the separation of the female stages along Axis 2 for the gonad
25 data included centrosomal protein of 70 kDa (cds.comp133063_c0_seq1|m.21029), WD repeat-
26 containing protein on Y chromosome (cds.comp131715_c0_seq1|m.19472), tetratricopeptide
27 repeat protein 18 (cds.comp139151_c0_seq5|m.31757), and four uncharacterized proteins
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

(cds.comp131654_c1_seq1|m.19394, cds.comp139654_c0_seq1|m.32810, cds.comp135264_c2_seq1|m.24100, cds.comp140866_c0_seq1|m.36276). Many of these 40 gonad transitions had ROC AUC >0.7, sensitivity of 1, and false positive rate < 0.5, suggestive of good candidates for SRM biomarkers (Supporting information 4).³⁵ Analysis of the hemolymph SRM NMDS (Supporting Information 3) reveals that the main peptide drivers along Axis 1 (separating late-stage females and mid-stage males) were from six proteins: flap endonuclease (cds.comp130260_c0_seq1|m.17926), vitellogenin (cds.comp144315_c0_seq1|m.50928), spectrin alpha chain (cds.comp141647_c0_seq1|m.39026), vesicle-fusing ATPase (cds.comp144180_c0_seq1|m.39026), and two uncharacterized proteins (cds.comp138018_c0_seq1|m.29238, cds.comp133422_c0_seq1|m.21474) (Figure 6). A cluster analysis of the 40 transitions that were primary drivers of individual variation in the NMDS exposed a strong cluster of the females in the hemolymph data (Figure 6). Only the transitions from vitellogenin peptides were analyzed with ROC since they were the only analytes that demonstrated a strong correlation with a specific sex and stage (mid- and late-female). All 10 transitions had AUC > 0.7, sensitivity of 1, and false positive rate < 0.3, verifying that these are good candidates for hemolymph peptide biomarkers of late females (Supporting information 4).³⁵

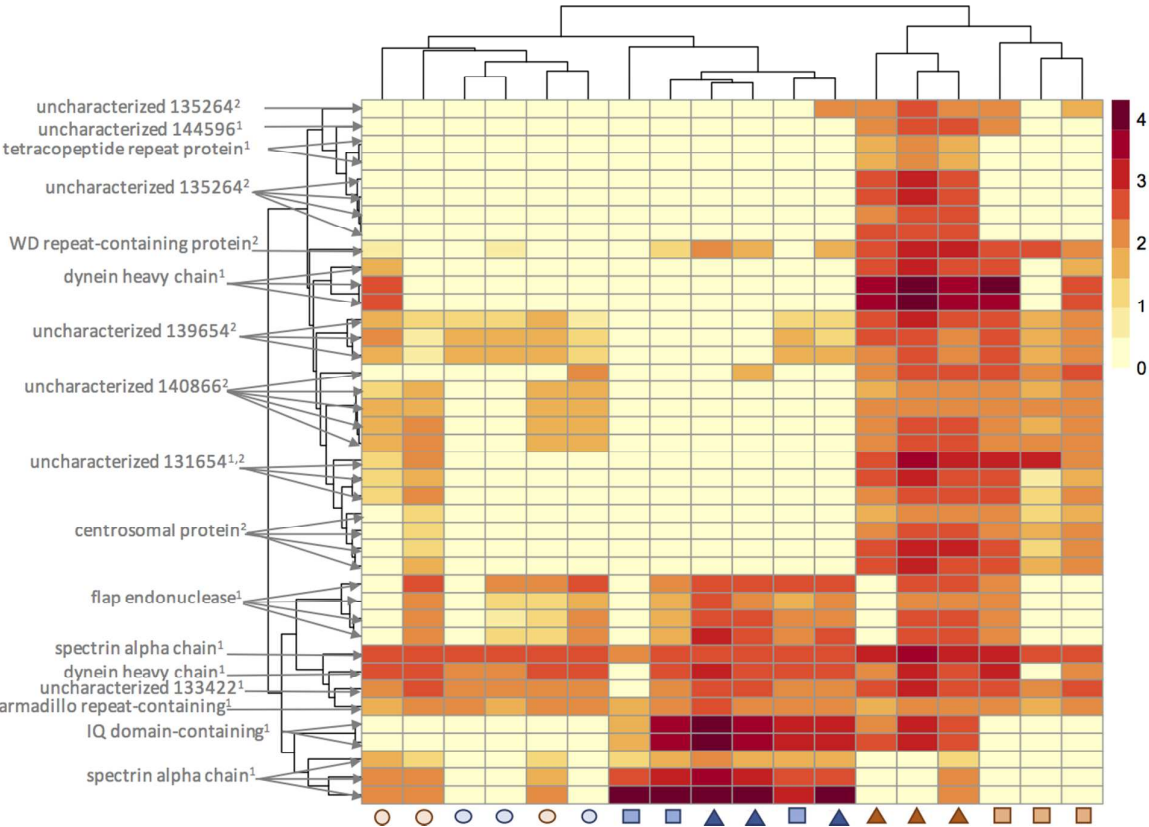


Figure 5. Heatmap of log-normalized peptide transition intensities for 40 significant transitions (based on NMDS analysis) across all samples (early-stage/circles, mid-stage/squares, and late-stage/triangles males/orange and females/blue) gonad. Each row (peptide transition) is labeled by its corresponding protein annotation in gray. Superscripts indicate for which NMDS axis a transition is significant.

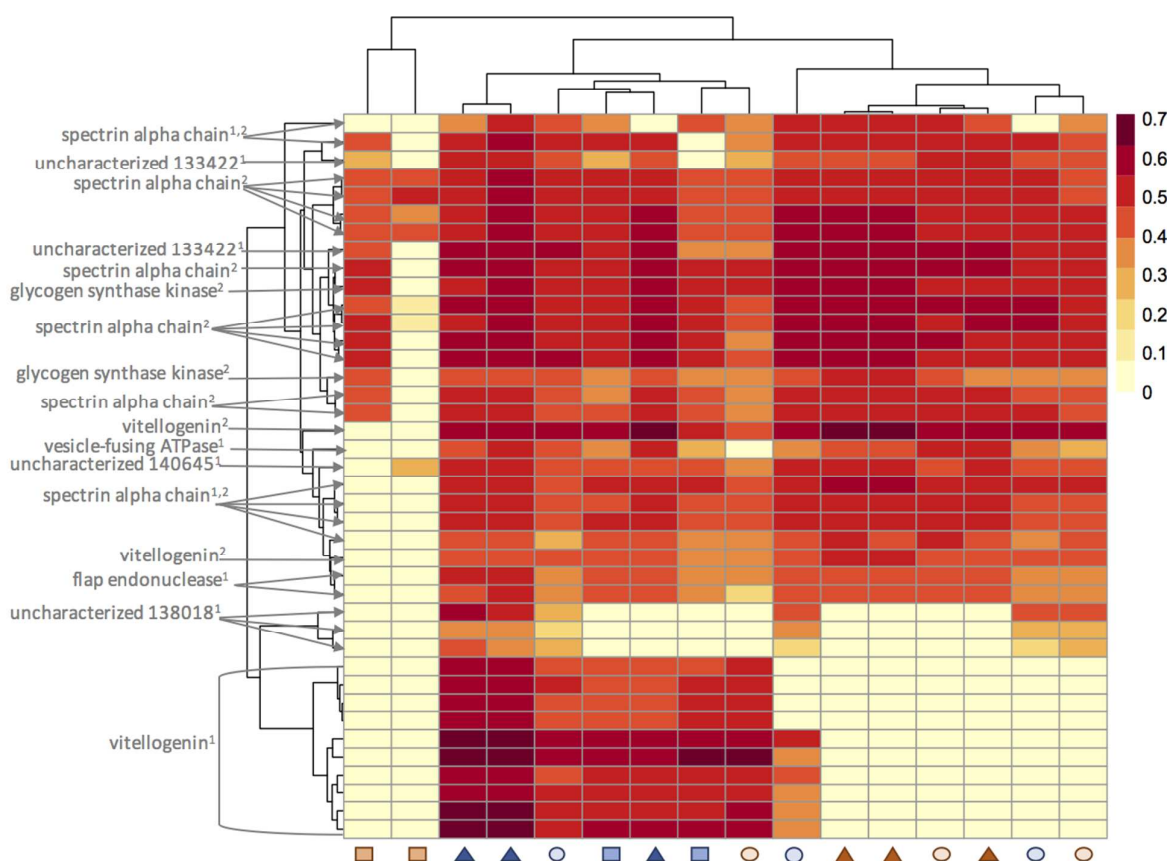


Figure 6. Heatmap of log-normalized peptide transition intensities for 40 significant transitions (based on NMDS analysis) across all samples (early-stage/circles, mid-stage/squares, and late-stage/triangles males/orange and females/blue) hemolymph. Each row (peptide transition) is labeled by its corresponding protein annotation in gray. Superscripts indicate for which NMDS axis a transition is significant.

Discussion

Proteomics has the capacity to uncover physiological processes underlying functional phenotypic change, including the 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 3,651 detected proteins across both sexes

1
2
3 and three maturation stages. This is a significant escalation in the understanding of proteomic
4 responses in maturation stages of marine mollusks. Based on the DDA data, 27 proteins (with a
5
6 corresponding 85 peptides) from early- and late-stage male and female clams were chosen for
7
8 selected reaction monitoring (SRM). Peptide transitions were detected from the selected peptides
9
10 as candidates for biomarker development in gonad (n=217) and in hemolymph (n=171).
11
12
13
14
15
16

17 *DDA*

18
19
20 There are clear proteomic profile differences in the geoduck gonad by both sex and
21
22 maturation stage and these differences reveal the biochemical pathways underlying tissue
23
24 specialization. More proteins were differentially abundant between early and late stages of
25
26 reproductive maturity, compared to early and mid-stages, for both males and females, reflecting
27
28 that, as maturation progresses, gonad tissue becomes more specialized, expressing a more
29
30 diverse array of regulatory and structural proteins. Gonad goes from undifferentiated connective
31
32 tissue to specialized structures that regulate gametogenesis and release of gametes. These
33
34 phenotypic changes required for reproductive maturity are evident with histology¹³ and are
35
36 realized via changes at the protein level. For example, the zona pellucida (ZP) containing protein
37
38 (cds.comp134923_c0_seq3|m.23445) and vitellogenin (cds.comp144315_c0_seq1|m.50928)
39
40 were more abundant in the female gonad as geoduck matured. The zona pellucida surrounds
41
42 vertebrate oocytes and its invertebrate homolog is the vitelline envelope; the vitelline envelope is
43
44 likely the origin of the zona pellucida (ZP) containing protein and the site of sperm-oocyte
45
46 recognition and binding. Vitellogenin is an egg yolk protein precursor that is closely linked to
47
48 gametogenesis in bivalves. Increasing levels of vitellogenin are correlated with advancement in
49
50 reproductive maturation in oysters^{36,37,38}, clams³⁹, and mussels⁴⁰.
51
52
53
54
55
56
57
58
59
60

1
2
3 In females, enrichment analysis of unique proteins and all detected proteins revealed that
4 the early-stage gonad proteome was enriched in proteins involved in translation and cytoskeleton
5 structure and proteins associated with purine metabolism and translation were predominant
6 during the mid-stage. These protein-level changes likely represent alterations to cell morphology
7 and shifts in protein translation required to effect gonad structural and functional changes.
8
9

10 The late-stage female gonad proteome was enriched in proteins associated with cell
11 adhesion and translation/phosphorylation. Cell adhesion is an essential biological process in
12 maturing oocytes in vertebrates.^{41,42} The strong cell adhesion signal in late-stage females was
13 driven in large part by the many cadherin and protocadherin proteins detected in a single female.
14 However, twelve protocadherins were detected at significantly higher abundance in more than a
15 single late-stage female when compared to early- and mid-stages. Further, the cell adhesion
16 proteins laminin, talin, and vinculin were less abundant in late-stage females compared to other
17 females. The shift from laminin, talin, and vinculin to cadherins during female gonad
18 development may signal a change in cell-cell interactions as oocytes form. There was also strong
19 evidence for the importance of protein translation and phosphorylation. Proteins with homology
20 to several protein kinases were all detected solely in late-stage females. This surge in
21 phosphorylating proteins is likely a reflection of an increase in protein activation in the later
22 stages of gonad maturation.
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44

45 Enriched processes in late-stage male geoduck gonad proteome including “kinesin
46 complex” and many microtubule-related functions reflect an increase in cell replication during
47 the last steps of sperm formation. Several proteins implicated in mitosis and meiosis were
48 detected at significantly increased abundance in late-stage males compared to early- and mid-
49 stage: dynein, intraflagellar transport protein, kinesin, and alpha and beta tubulin. All of these
50
51
52
53
54
55
56
57
58
59
60

1
2
3 proteins are instrumental in the cell division and replication and their importance at this
4
5 maturation stage suggests the presence of rapidly dividing cells. In mussels off the Atlantic coast
6
7 of Spain, meiotic division begins approximately four months before the onset of spawning, while
8
9 mitosis reconstructs the gonad during and after spawning.⁴³ These pieces of proteomic evidence
10
11 point towards increased cellular energy metabolism and cell division during the development of
12
13 mature, motile sperm.
14
15

16
17 Even with a highly specific protein identification database, the DDA method excludes
18
19 many important, relatively low abundance proteins from analysis.⁴⁴ The ten most abundant
20
21 proteins (as measured by total spectral count) for each sex-maturation stage accounted for 6-28%
22
23 of the total spectral counts across all proteins for a given sex-stage. Many of these proteins are
24
25 “housekeeping” proteins, such as actin and myosin, and their dominance in the peptide mixture
26
27 was likely masking many informative, low abundance peptides. Similarly, in zebrafish, highly
28
29 abundant vitellogenin (also highly abundant in the mid- and late-stage females in this study)
30
31 made it difficult to detect lower abundance proteins using DDA.⁴⁵
32
33

34
35 The most abundant protein across all replicates likely plays an essential role in
36
37 reproductive maturation. The protein is glutamine gamma-glutamyltransferase, or
38
39 transglutaminase, and it has total spectral counts ranging from 259 to 1878 per MS analysis. The
40
41 annotation to this protein was only recently updated in Uniprot, so it was originally annotated as
42
43 “uncharacterized” in our BLAST search. Transglutaminase is typically associated with protein
44
45 cross-linking during wound healing in crustaceans⁴⁶ and bivalves⁴⁷ and also serves as a
46
47 biomarker for summer mortality in bivalves.⁴⁸ However, gene expression of this protein
48
49 transcript has also been linked to vitellogenesis in the shrimp *Panaeus monodon* where it is
50
51 hypothesized to mediate ovarian maturation via a hormone-receptor interaction⁴⁹. The high
52
53
54
55
56
57
58
59
60

1
2
3 abundance of transglutaminase in our geoduck dataset suggests a key role in invertebrate
4
5 reproductive maturation. It is significantly more abundant in early-stage males and females
6
7 compared to late- and mid-stages as well as more abundant in late-stage females than males,
8
9 which is consistent with the shrimp study where its gene transcript was at highest abundance in
10
11 early vitellogenesis⁴⁹.
12
13
14

15 16 17 *SRM* 18

19
20 Selected reaction monitoring allows for the detection and absolute quantification of only
21
22 the most informative proteins' peptides to address a specific hypothesis. We leveraged the DDA
23
24 dataset to create informative SRM assays of geoduck sex and maturation stage in both gonad and
25
26 hemolymph. In the gonad tissue, we measured 217 peptide transitions that could differentiate
27
28 males and females as well as early and late maturation stages. Our assay is based on proteins that
29
30 are involved in calcium ion binding, phosphorylation, meiosis, DNA replication and repair, and
31
32 membrane structure. Additionally, seven of these informative proteins were unannotated. These
33
34 unknown proteins may represent taxon-specific reproductive proteins and warrant further
35
36 characterization. Reproductive proteins are some of the fastest-evolving, resulting in many
37
38 species-specific proteins.⁵⁰
39
40
41
42

43
44 A subset of the peptide transitions (n=40) from these proteins were deemed highly
45
46 informative based on our analysis in gonad samples and would be good candidates for a more
47
48 streamlined SRM assay. These transitions correspond to eight annotated proteins and six
49
50 unannotated proteins. The annotated proteins are implicated in cytoskeleton structure (spectrin
51
52 alpha chain, dynein heavy chain), DNA replication and repair (centrosomal protein of 70 kDa,
53
54 flap endonuclease), and proteins that were identified by motifs and thus yield little functional
55
56
57
58
59
60

1
2
3 information (WD repeat-containing protein on Y chromosome, IQ domain containing protein K,
4 armadillo repeat-containing protein, and tetracopeptide repeat protein 18). The 40 transitions
5
6 were almost uniformly at higher abundance in late stage males and females, compared to earlier
7
8 reproductive stages.
9
10

11
12 The peptide-based assay developed from the evidence found in the gonad tissue was
13 applied to circulating hemolymph. We were able to collect data on 171 peptide transitions that
14
15 were derived from gonad tissue but detectable in the hemolymph, including all the transitions
16
17 that were added specifically to the hemolymph assay. Peptide transitions from 16 proteins were
18
19 reliably detected in both gonad and hemolymph, however, only three proteins had transitions that
20
21 were highly informative in both tissues: spectrin alpha chain, flap endonuclease, and an
22
23 uncharacterized protein. The inclusion of vitellogenin in our assay confirmed its presence in
24
25 geoduck hemolymph during later stages of reproductive maturity in females. Vitellogenin is an
26
27 informative proteomic biomarker for oyster and sea turtle female reproductive stage^{38,51}, and our
28
29 analysis proves that its sensitivity and specificity in geoduck hemolymph make it a good
30
31 biomarker for detecting late stage females. In the oyster study, vitellogenin fragments were
32
33 among the most abundant proteins detected and were particularly accumulated at greater
34
35 abundances in oocytes associated with low production of larvae.³⁸ In this case, it was
36
37 hypothesized that greater accumulation of vitellogenin fragments in the poor versus high quality
38
39 oyster oocytes was indicative of the enzymatic breakdown of the protein during oocyte ageing.³⁸
40
41
42
43
44
45
46
47

48 The full suite of peptide transitions was not detected in hemolymph, likely due to
49
50 physiological differences between the tissues in which the assay was designed (gonad) and
51
52 applied (hemolymph). Biomarkers of tissue-specific changes can be difficult to detect in
53
54 circulating fluid since hemolymph is not the primary site of physiological change. Proteins and
55
56
57
58
59
60

1
2
3 their peptide fragments can yield physiologically relevant assays, even for organ-specific
4
5 changes ⁵², but may not be detectable due to different fragmentation patterns or delay in
6
7 abundance changes between the site of physiological change and the hemolymph or blood. In
8
9 geoduck hemolymph, 98 transitions were detected in the gonad but not in the hemolymph, but
10
11 119 transitions were detected in both tissues. Flap endonuclease and an uncharacterized protein,
12
13 detected in both tissues, were informative for late-stage female identification in the hemolymph
14
15 assay with flap endonuclease also an informative biomarker in gonad (Figures 5 & 6). The
16
17 hemolymph results emphasize both the flexibility and limitations of a peptide-based SRM assay
18
19 when designed for measurement of physiological change in a specific organ or tissue.
20
21
22
23

24 Through characterization of the geoduck gonad proteome and development of targeted
25
26 peptide-based assays for reproductive maturity, we not only exponentially increased genomic
27
28 resources available for this species, but we also provide an effective approach for non-lethal
29
30 detection of sex and maturation stage in this species. The gonad peptide transitions could be
31
32 optimized to a set of 40 transitions that would accurately predict a) geoduck sex and b)
33
34 reproductive maturation status. In the hemolymph, vitellogenin peptide biomarkers can
35
36 differentiate late-stage females from other geoduck. Together these molecular tools can directly
37
38 address the geoduck production problem of asynchronous spawning due to different maturation
39
40 stages and unknown sexes.
41
42
43
44
45
46
47

48 **Acknowledgements**

49
50 This work was supported and funded by a grant from the University of Washington
51
52 Royalty Research Fund (to SR), a Training Grant from the National Institutes of Health (to ETS)
53
54 (T32 HG00035), and the University of Washington's Proteomics Resource (UWPR95794). We
55
56
57
58
59
60

would like to thank Jimmy Eng and Priska von Haller for their help in data acquisition and analysis; Sonia Ting, Jarrett Egerton, Lindsay Pino, Yuval Boss, Brian Searle, and Nick Shulman for assistance with data analysis; the Genome Sciences Information and Technology group for their technical support; and Michael MacCoss for providing support and space to complete the work. ETS and BLN would like to thank IJE and TAN for their ongoing inspiration.

Supporting Information.

S-1 Cover page for Supporting Information.

S-2 Parameter file used in Comet searches for the DDA data and R script for creating NMDS of DDA NSAF data. All other NMDS were executed similarly.

S-3 Plots demonstrating technical replication for DDA and SRM, and DAGs for enrichment analysis.

S-4 The first tab of the workbook contains the sample-specific probabilities (“protein probability”) associated with $FDR = 0.01$, as well as the total proteins inferred for each sample (“number entries”) and the total spectra identified at $FDR = 0.01$. Each sample is referred to with its sex-stage and sample number, as well as technical replicate (for example, “EF3.1 SpC” is technical replicate 1 from early-stage female 3). In the second tab, 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” preceded by the biological replicate number, sex, maturation stage, and technical replicate. Notation for NSAF is

similar to notation for SpC. The 9 columns in the sheet have headers such as “EFvLF” (comparison between early- and late-stage females) have asterisks in the cells that correspond to proteins that were differentially abundant for each given comparison. The last two columns (“NMDS Gonad” and “NMDS Hemolymph”) have asterisks in cells that correspond to the proteins with peptide transitions that contribute significantly to the SRM NMDS plot distributions. The third tab contains GO biological processes enriched in proteins that were differentially abundant between stages within a sex (e.g. early- vs. mid-stage female) or between sexes within a stage (e.g. LF vs. LM). Only the most specific GO terms in the GO hierarchy are listed. No terms are listed when a comparison was not made. The fourth through seventh tabs contain raw Skyline output (in the tab “Skyline output”) and peak intensities normalized by PRTC peptide abundance (“Normalized Intensities”) for the gonad and hemolymph SRM data. The last two tabs contain the results of the receiver operating characteristic (ROC) curve analysis for gonad and hemolymph. For each transition that demonstrated a strong correlation with at least one sex and maturation stage (see Figures 5 and 6), the area under the curve (AUC) was calculated for a primary and secondary sex and stage association. Additionally, sensitivity (true positive rate), specificity, and false positive rate (1-specificity) were calculated for each transition according to Whiteaker et al.³⁵

References

- (1) Newell, R. I. Ecosystem influences of natural and cultivated populations of suspension-feeding bivalve mollusks: A review. *J. Shellfish Res.* **2004**, *23* (1), 51–62.
- (2) Hoffmann, A.; Bradbury, A.; Goodwin, C. L. Modeling geoduck, *Panopea abrupta* (Conrad, 1849) population dynamics. I. Growth. *J. Shellfish Res.* **2000**, *19* (1), 57–62.

- (3) Wellman, K.; Cuyno, L.; Schenken, W.; Bond, A.; Eaton, G.; McCoy, T. *The Economic Impact of Shellfish Aquaculture in Washington, Oregon and California*; Technical Report for Northern Economics: Anchorage, AK, April 2013.
- (4) Scott, A. P. Do mollusks use vertebrate sex steroids as reproductive hormones? II. Critical review of the evidence that steroids have biological effects. *Steroids* **2013**, 78 (2), 268–281.
- (5) Fernandes, D.; Denise, F.; Barbara, L.; Cinta, P. Biosynthesis and metabolism of steroids in molluscs. *J. Steroid Biochem. Mol. Biol.* **2011**, 127 (3-5), 189–195.
- (6) Palacios, E.; Racotta, I. S.; Arjona, O.; Marty, Y.; Le Coz, J. R.; Moal, J.; Samain, J. F. Lipid composition of the pacific lion-paw scallop, *Nodipecten subnodosus*, in relation to gametogenesis. *Aquaculture* **2007**, 266 (1-4), 266–273.
- (7) Anantharaman, S.; Craft, J. A. Annual variation in the levels of transcripts of sex-specific genes in the mantle of the common mussel, *Mytilus edulis*. *PLoS One* **2012**, 7 (11), e50861.
- (8) Ciocan, C. M.; Cubero-Leon, E.; Minier, C.; Rotchell, J. M. Identification of reproduction-specific genes associated with maturation and estrogen exposure in a marine bivalve *Mytilus edulis*. *PLoS One* **2011**, 6 (7), e22326.
- (9) Li, Q.; Qi, L.; Makoto, O.; Katsuyoshi, M. Seasonal biochemical variations in Pacific oyster gonadal tissue during sexual maturation. *Fish. Sci.* **2000**, 66 (3), 502–508.
- (10) Dheilly, N. M.; Lelong, C.; Huvet, A.; Kellner, K.; Dubos, M.-P.; Riviere, G.; Boudry, P.; Favrel, P. Gametogenesis in the Pacific oyster *Crassostrea gigas*: a microarrays-based analysis identifies sex and stage specific genes. *PLoS One* **2012**, 7 (5), e36353.
- (11) Arcos, F. G.; Ibarra, A. M.; del Carmen Rodríguez-Jaramillo, M.; García-Latorre, E. A.; Celia, V.-B. Quantification of vitellin/vitellogenin-like proteins in the oyster *Crassostrea corteziensis* (Hertlein 1951) as a tool to predict the degree of gonad maturity. *Aquac. Res.*

- 2009, 40 (6), 644–655.
- (12) Li, Y.; Siddiqui, G.; Wikfors, G. H. Non-lethal determination of sex and reproductive condition of Eastern oysters *Crassostrea virginica* Gmelin using protein profiles of hemolymph by Proteinchip® and SELDI-TOF-MS technology. *Aquaculture* **2010**, 309 (1–4), 258–264.
- (13) Crandall, G.; Roberts, S. Reproductive Maturation in Geoduck clams (*Panopea generosa*) FigShare.
https://figshare.com/articles/Reproductive_Maturation_in_Geoduck_clams_Panopea_generosa_/3205975/1. (accessed April 29, 2016).
- (14) Timmins-Schiffman, E.; Nunn, B. L.; Goodlett, D. R.; Roberts, S. B. Shotgun proteomics as a viable approach for biological discovery in the Pacific oyster. *Conserv Physiol* **2013**, 1 (1), cot009.
- (15) Grabherr, M. G.; Haas, B. J.; Yassour, M.; Levin, J. Z.; Thompson, D. A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.; Chen, Z.; Mauceli, E.; Hacohen, N.; Gnirke, A.; Rhind, N.; di Palma, F.; Birren, B.W.; Nusbaum, C.; Lindblad-Toh, K.; Friedman, N.; Regev, A. Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nature Biotechnology* **2011**, 29, 644-652.
- (16) Eng, J.K.; Jahan, T.A.; Hoppmann, M.R. Comet: An open-source MS/MS sequence database search tool. *Proteomics* **2012**, 13, 22-24.
- (17) Eng, J.K.; Hoopmann, M.R.; Jahan, T.A.; Egertson, J.D.; Noble, W.S.; MacCoss, M.J. A Deeper Look into Comet – Implementation and Features. *Journal of the American Society for Mass Spectrometry* **2015**, 11, 1865-1874.
- (18) Pedrioli, P. G. A. Trans-Proteomic Pipeline: A Pipeline for Proteomic Analysis. In

- Methods in Molecular Biology*; 2009; pp 213–238.
- (19) Deutsch, E. W.; Mendoza, L.; Shteynberg, D.; Slagel, J.; Sun, Z.; Moritz, R. L. Trans-Proteomic Pipeline, a standardized data processing pipeline for large-scale reproducible proteomics informatics. *Proteomics Clin. Appl.* **2015**, *9* (7-8), 745–754.
- (20) Florens, L.; Carozza, M. J.; Swanson, S. K.; Fournier, M.; Coleman, M. K.; Workman, J. L.; Washburn, M. P. Analyzing chromatin remodeling complexes using shotgun proteomics and normalized spectral abundance factors. *Methods* **2006**, *40* (4), 303–311.
- (21) Damien Fermin, Vekatesha Basrur, Anastasia K. Yocum, and Alexey I. Nesvizhskii. Abacus: A computational tool for extracting and pre-processing spectral count data for label-free quantitative proteomic analysis. *Proteomics* **2011**, *11* (7), 1340–1345.
- (22) Oksanen, J.; Guillaume Blanchet, F.; Friendly, M.; Kindt, R.; Legendre, P.; McGlinn, D.; Minchin, P. R.; O'Hara, R. B.; Simpson, G. L.; Solymos, P.; Stevens, M.H.H.; Szoecs, E.; Wagner, H. vegan: Community Ecology Package. R package version 2.4-1. **2016**.
<https://CRAN-R-project.org/package=vegan>
- (23) R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, **2015**. <https://www.R-project.org>
- (24) Oliveros, J. C. Venny. An interactive tool for comparing lists with Venn's diagrams. <https://bioinfogp.cnb.csic.es/tools/venny/index.html> (accessed April 5, 2017).
- (25) MacLean, B.; Tomazela, D. M.; Shulman, N.; Chambers, M.; Finney, G. L.; Frewen, B.; Kern, R.; Tabb, D. L.; Liebner, D. C.; MacCoss, M. J. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **2010**, *26* (7), 966–968.
- (26) Panchaud, A.; Scherl, A.; Shaffer, S. A.; von Haller, P. D.; Kulasekara, H. D.; Miller, S.

- I.; Goodlett, D. R. Precursor acquisition independent from ion count: how to dive deeper into the proteomics ocean. *Anal. Chem.* **2009**, *81* (15), 6481–6488.
- (27) Chambers, M. C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D. L.; Neumann, S.; Gatto, L.; Fischer, B.; Pratt, B.; Egertson, J.; et al. A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol.* **2012**, *30* (10), 918–920.
- (28) Ting, Y. S.; Egertson, J. D.; Payne, S. H.; Kim, S.; MacLean, B.; Käll, L.; Aebersold, R.; Smith, R. D.; Noble, W. S.; MacCoss, M. J. Peptide-Centric Proteome Analysis: An Alternative Strategy for the Analysis of Tandem Mass Spectrometry Data. *Mol. Cell. Proteomics* **2015**, *14* (9), 2301–2307.
- (29) Wolf-Yadlin, A.; Hautaniemi, S.; Lauffenburger, D. A.; White, F. M. Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (14), 5860–5865.
- (30) Picotti, P.; Aebersold, R. Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nat. Methods* **2012**, *9* (6), 555–566.
- (31) Moreira, R.; Pereiro, P.; Canchaya, C.; Posada, D.; Figueras, A.; Novoa, B. RNA-Seq in *Mytilus galloprovincialis*: comparative transcriptomics and expression profiles among different tissues. *BMC Genomics* **2015**, *16* (1), 728.
- (32) Hijmans, R. J. raster: Geographic Data Analysis and Modeling. R package raster version 2.5-8. **2016**. <https://CRAN.R-project.org/package=raster>
- (33) Kolde, R. pheatmap: Pretty Heatmaps. R package version 1.0.8. **2015**. <https://CRAN.R-project.org/package=pheatmap>
- (34) Robin, X.; Turck, N.; Hainard, A.; Tiberti, N.; Lisacek, F.; Sanchez, J.-C. ; Muller, M. pROC : an open-source package for R and S+ to analyze and compare ROC curves. *BMC*

- Bioinformatics* **2011**, 12, 77.
- (35) Whiteaker, J.R.; Lin, C.; Kennedy, J.; Hou, L.; Trute, M.; Sokal, I.; Yan, P.; Schoenherr, R.M.; Zhao, L.; Voytovich, U.J.; Kelly-Spratt, K.S.; Krasnoselsky, A.; Gafken, P.R.; Hogan, J.M.; Jones, L.A.; Wang, P.; Amon, L.; Chodosh, L.A.; Nelson, P.S.; McIntosh, M.W.; Kemp, C.J.; Paulovich, A.G. A targeted proteomics-based pipeline for verification of biomarkers in plasma. *Nature Biotechnology* **2011**, 29, 625-634.
- (36) Matsumoto, T.; Nakamura, A. M.; Mori, K.; Kayano, T. Molecular characterization of a cDNA encoding putative vitellogenin from the Pacific oyster *Crassostrea gigas*. *Zoolog. Sci.* **2003**, 20 (1), 37-42.
- (37) Arcos, F. G.; Ibarra, A. M.; Rodríguez-Jaramillo, M. del C.; García-Latorre, E. A.; Vazquez-Boucard, C. Quantification of vitellin/vitellogenin-like proteins in the oyster *Crassostrea corteziensis* (Hertlein 1951) as a tool to predict the degree of gonad maturity. *Aquac. Res.* **2009**, 40 (6), 644-655.
- (38) Corporeau, C.; Vanderplancke, G.; Boulais, M.; Suquet, M.; Quéré, C.; Boudry, P.; Huvet, A.; Madec, S. Proteomic identification of quality factors for oocytes in the Pacific oyster *Crassostrea gigas*. *J. Proteomics* **2012**, 75 (18), 5554-5563.
- (39) Blaise, C.; Gagné, F.; Pellerin, J.; Hansen, P. D. Determination of vitellogenin-like properties in *Mya arenaria* hemolymph (Saguenay Fjord, Canada): A potential biomarker for endocrine disruption. *Environ. Toxicol.* **1999**, 14 (5), 455-465.
- (40) Won, S.-J.; Novillo, A.; Custodia, N.; Rie, M. T.; Fitzgerald, K.; Osada, M.; Callard, I. P. The Freshwater Mussel (*Elliptio complanata*) as a Sentinel Species: Vitellogenin and Steroid Receptors. *Integr. Comp. Biol.* **2005**, 45 (1), 72-80.
- (41) Fusi, F. M.; Vignali, M.; Busacca, M.; Bronson, R. A. Evidence for the presence of an

- integrin cell adhesion receptor on the oolemma of unfertilized human oocytes. *Mol. Reprod. Dev.* **1992**, *31* (3), 215–222.
- (42) Campbell, S.; Swann, H. R.; Seif, M. W.; Kimber, S. J.; Aplin, J. D. Integrins and adhesion molecules: Cell adhesion molecules on the oocyte and preimplantation human embryo. *Hum. Reprod.* **1995**, *10* (6), 1571–1578.
- (43) Suárez, M. P.; Alvarez, C.; Molist, P.; Juan, F. S. Particular Aspects of Gonadal Cycle and Seasonal Distribution of Gametogenic Stages of *Mytilus galloprovincialis* Cultures in the Estuary of Vigo. *J. Shellfish Res.* **2005**, *24* (2), 531–540.
- (44) Pedersen, S. K.; Harry, J. L.; Sebastian, L.; Baker, J.; Traini, M. D.; McCarthy, J. T.; Manoharan, A.; Wilkins, M. R.; Gooley, A. A.; Righetti, P. G.; et al. Unseen proteome: mining below the tip of the iceberg to find low abundance and membrane proteins. *J. Proteome Res.* **2003**, *2* (3), 303–311.
- (45) Groh, K. J.; Nesatyy, V. J.; Segner, H.; Eggen, R. I. L.; Suter, M. J.-F. Global proteomics analysis of testis and ovary in adult zebrafish (*Danio rerio*). *Fish Physiol. Biochem.* **2011**, *37* (3), 619–647.
- (46) Martin, G. G.; Jo Ellen Hose; Chong, C.; Hoodbhoy, T.; McKrell, N. Localization and roles of coagulogen and transglutaminase in hemolymph coagulation in decapod crustaceans. *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry* **1991**, *100* (3), 517–522.
- (47) Nozawa, H.; Mori, T.; Kimura, M.; Seki, N. Characterization of a transglutaminase from scallop hemocyte and identification of its intracellular substrates. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **2005**, *140* (3), 395–402.
- (48) Chaney, M. L.; Gracey, A. Y. Mass mortality in Pacific oysters is associated with a

- specific gene expression signature. *Mol. Ecol.* **2011**, *20* (14), 2942–2954.
- (49) Sathapondecha, P.; Treerattrakool, S.; Panyim, S.; Udomkit, A. Potential roles of transglutaminase and thioredoxin in the release of gonad-stimulating factor in *Penaeus monodon*: implication from differential expression in the brain during ovarian maturation cycle. *Mar. Genomics* **2011**, *4* (4), 279–285.
- (50) Swanson, W. J.; Vacquier, V. D. The rapid evolution of reproductive proteins. *Nat. Rev. Genet.* **2002**, *3* (2), 137–144.
- (51) Plumel, M. I.; Wasselin, T.; Plot, V.; Strub, J.-M.; Van Dorselaer, A.; Carapito, C.; Georges, J.-Y.; Bertile, F. Mass spectrometry-based sequencing and SRM-based quantitation of two novel vitellogenin isoforms in the leatherback sea turtle (*Dermochelys coriacea*). *J. Proteome Res.* **2013**, *12* (9), 4122–4135.
- (52) Klee, E. W.; Bondar, O. P.; Goodmanson, M. K.; Dyer, R. B.; Erdogan, S.; Bergstralh, E. J.; Bergen, H. R., 3rd; Sebo, T. J.; Klee, G. G. Candidate serum biomarkers for prostate adenocarcinoma identified by mRNA differences in prostate tissue and verified with protein measurements in tissue and blood. *Clin. Chem.* **2012**, *58* (3), 599–609.

