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

Emma B. Timmins-Schiffman[1], Grace A. Crandall[2], Brent Vadopalas[2], Michael E. Riffle[3], Brook L. Nunn[1], Steven B. Roberts[2,*]

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

[2] Department 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, Department of Aquatic and Fishery Sciences, University of Washington, Seattle WA 98105. Email: sr320@uw.edu

Abstract

Geoduck clams (*Panopea generosa*) are an increasingly important fishery and aquaculture product in the United States while also holding essential ecological roles. These long-lived clams are highly fecund, but 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. Assays of gonad physiology could indicate both sex and maturation stage. Proteomic profiles were determined for three reproductive maturation stages in both male and female clams using data dependent acquisition (DDA), or whole proteome profiling, 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 non-lethally determine geoduck sex and maturation stage pre-spawning. These results lay the foundation for the development of field-based tools to assess reproductive status in marine mollusks.

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 hatchery broodstock.

Geoducks are not sexually dimorphic, so these determinations require hatcheries to increase the temperature of seawater and introduce large rations of microalgae to induce the broodstock to spawn. Further, within a hatchery incubator only a few of the broodstock mature and spawn during a given spawning attempt. This thwarts the sustainability goal to breed large numbers of broodstock to maintain genetic diversity. Highly skewed sex ratios in some populations and the inability to determine when to induce spawning based on maturation state presents additional substantial economic challenges for shellfish hatcheries. Since sexual differentiation and maturation processes in marine mollusks cannot be well characterized with visual inspection, a low-cost assay for sex and maturation stage would be advantageous to these efforts.

To date, only a few studies have investigated molecular-level analyses to identify sex or reproductive stages in marine invertebrates. For example, contentious studies on vertebrate steroids explored these organic molecules as an analyte for tracking mollusk reproduction processes and stages.[4] Given the lack of functional receptors for androgens and estrogens in marine mollusks, vertebrate-specific steroids may not be critical analytes that play a role in reproduction; they have yet to yield viable assays for mollusk reproductive stages.[5] Further research on gonad lipids [6], specific genetic transcripts [7,8], and RNA/DNA ratios [9] have also not provided definitive, reproducible assays for the determination of mollusk sexual maturation. Arcos et al. [10] explored the use of enzyme-linked immunoassay assays (ELISA) to quantify proteins associated with vitellogenesis in order to determine oyster oocyte maturation. Although the two proteins targeted by Arcos et al.[10] allowed distinction of males and females, multiple mechanisms for maturation processing may be at play and, to date, unidentified.[4] Supporting this hypothesis, Li et al. [11] used an alternative protein-profile approach

69 to simultaneously examine 139 peptides in the Eastern oyster, *Crassostrea virginica*, and
70 revealed that numerous unidentified proteins change in abundance with relation to maturation
71 and the overall protein profile could predict gender differences.

72 Since male and female geoduck gonads are morphologically and functionally distinct at
73 reproductive maturity, molecular assays to track these early biochemical changes could help
74 identify sex and indicate when gametes are mature. Significant changes in both gene and protein
75 expression are expected as gonad tissue differentiates. Proteomic assays designed to detect these
76 expression changes could lead to non-lethal, informative, and efficient tools for selecting
77 broodstock. Here, we apply mass spectrometry (MS)-based proteomics technology to provide a
78 large-scale, unbiased approach for examining relative abundances of all proteins present in
79 gonad tissue across several maturation stages of both female and male geoducks. Whole
80 proteome characterization via data dependent acquisition (DDA) provided a framework for
81 understanding molecular processes behind reproductive differentiation and directed peptide
82 selection for the determination of sex and maturation stage in subsequent downstream MS
83 analyses. Selected peptides underwent targeted proteomic analyses via selected reaction
84 monitoring (SRM) to more accurately quantify geoduck gonad proteins. These analyses were
85 used to increase our fundamental understanding of sex and stage determination in marine
86 mollusks and to develop non-invasive peptide assays on the circulating hemolymph.
87
88 **Methods**
89 *Tissue Sampling*
90 Geoduck clams were collected in November 2014 from Nisqually Reach, Washington
91 (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 ~5-15µ), middle-stage (secondary oocytes ~50-70µ), and late-stage (secondary oocytes ~65-85µ). Male reproductive maturation stages were characterized as follows: early-stage (mostly somatic cells and ~5% spermatid composition per acinus), middle-stage (about equal parts somatic cells and reproductive tissue and ~50% spermatid composition per acinus), and late-stage (very little somatic cells and ~75-90% spermatid composition per acinus). Details of gonadal maturation classification and histological details have been previously described.[12] 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. (2015).[12] 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).

Protein Preparation

Gonad tissue from each of the eighteen geoduck clams was sonicated in lysis buffer (50 mM NH4HCO3, 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 ug - 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. [13] 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, 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.

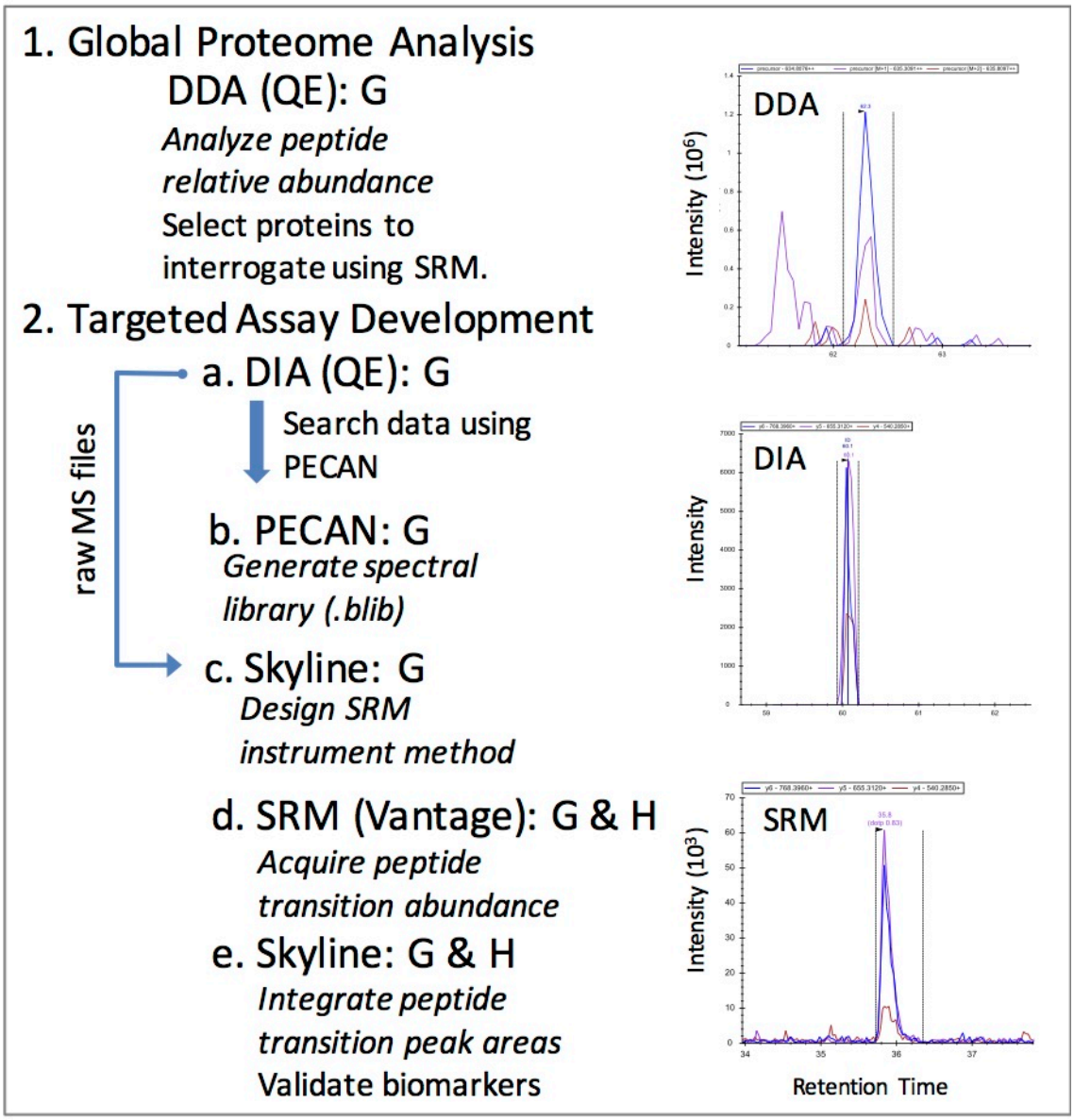
Global Proteome Analysis: Data-Dependent Acquisition (DDA) LC-MS/MS

Data-dependent acquisition (DDA), also referred to as shotgun proteomics, 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, 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) +

138 bovine serum albumin peptides (BSA)) were analyzed throughout the experiment to ensure

139 consistency of peptide detection and elution times.

140



141

142

143 *Figure 1. Illustration of experimental setup and workflow for mass spectrometry data*

144 *acquisition and analysis. MS experimental workflow: 1. Data dependent acquisition (DDA) was*

145 *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 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 [14] and deduced protein sequences determined using the Transdecoder algorithm within Trinity. Raw mass spectrometry data (PRIDE Accession #PXD003127) was searched against the protein database (Supporting Information 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) (Supporting information 2) to find peptide spectral matches.[15,16] Protein inference and match probability were found using the Trans-Proteomic

169 Pipeline with a probability cut-off of 0.9.[17,18] Protein identifications were considered true
170 matches when the probability of the match was at least 0.9 and at least 2 independent spectra
171 were associated with the protein across all samples.
172 Non-metric multidimensional scaling analysis (NMDS) was used to determine the
173 similarity of technical replicates (Supporting Information 3) using the vegan package [19] in R v.
174 3.2.3 [20]. NMDS was performed on log-transformed data using a Bray-Curtis dissimilarity
175 matrix. As technical replicates clustered closely together and showed less variability than
176 biological replicates (Supporting Information 4), spectral counts were averaged across each
177 sample (n=18). Normalized spectral abundance factors (NSAF) were calculated on the averaged
178 spectral counts for all high confidence, detected proteins as a proxy for protein expression.[21]
179 Proteomic differences between sexes and maturation stages were explored with two methods: 1)
180 NMDS and analysis of similarity (ANOSIM) were used to compare the entire proteomic profiles
181 in multivariate space and 2) QSpec [22] was used to determine significant differences at the
182 individual protein level. NMDS was performed on NSAF data followed by ANOSIM in the
183 vegan package in R to determine the differences between sexes and maturation stages.
184 Differentially abundant proteins among the six conditions were identified using QSpec.[22]
185 Spectral counts were summed across technical replicates to create input files for QSpec
186 (http://www.nesvilab.org/qspec.php/) with data normalized by abundance. Nine comparisons
187 were analyzed for differentially abundant proteins: Early-stage male (EM) vs. early-stage female
188 (EF); mid-stage male (MM) vs. mid-stage female (MF); late-stage male (LM) vs. late-stage
189 female (LF); EM vs. MM; EM vs. LM; MM vs. LM; EF vs. MF; EF vs. LF; MF vs. LF. Proteins
190 were considered significantly different if the absolute value of the log fold change was $\geq$ 0.5 and
191 the absolute value of the z-statistic was $\geq$ 2.

192 *Enrichment Analysis*

193 We used enrichment analysis to determine which biological processes (represented by
194 detected proteins) were predominant in each gonad sample. GO enrichment analysis was
195 performed for each maturation stage (for each sex) on 1) unique proteins and 2) all detected
196 proteins. In both cases, the entire detected proteome was used as the background. The uniquely
197 detected proteins were the basis for targeted proteomics assay development (see below).
198 Enrichment analysis was also performed on sets of differentially abundant proteins from QSpec
199 analyses. Code used to perform enrichment analysis is available in a corresponding GitHub
200 repository (https://github.com/yeastrc/compgo-geoduck-public) as is the underlying code for the
201 end user web-interface (http://yeastrc.org/compgo_geoduck/pages/goAnalysisForm.jsp).

202 Briefly, a p-value was calculated representing the likelihood that a GO term would be as
203 represented as it was (or more) by chance among the set of tested proteins, given the annotation
204 of the geoduck gonad proteome (3,627 proteins). A p-value cutoff of 0.01 was used to ascribe
205 statistical significance to GO terms representing biological process.

206 GO terms were first assigned directly to the geoduck gonad protein names resulting from
207 the *de novo*-assembled transcriptome. This was done by assigning GO terms associated with
208 Uniprot-KB Swiss-Prot BLASTp hits. A p-value representing the statistical significance of the
209 representation of a GO term in a set of proteins was calculated using the hypergeometric
210 distribution using the following formula:

211

212

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

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.

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 $\geqslant 0.01$. 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. Visualizations of DAGs for combined full and unique proteomes for each sex and stage are in Supporting Information 7 and proteins that contribute to specific enriched processes for all enrichment analyses are included in Supporting Information 6.

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 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 daily v. 3.5.1.9706 [23] 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 protein (1 ug) plus the QC (375 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 each [24] (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. MSConvert [25] 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[26]). 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 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.[27] 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 $\geq$ 3 good quality transitions and there were $\geq$ 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.[28]

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

278 described above for DIA (1 µg of protein per 3 µl injection). A new C18 trap (2 cm) and C18
279 analytical column (27.5 cm) were used and each sample was analyzed in triplicate across two
280 MS experiments to cover the entire peptide transition list (n=212). Raw data can be accessed in
281 the PeptideAtlas under accession PASS00943.

282 Hemolymph from early-, mid-, and late-stage males and females was also assayed for
283 sex- and stage-specific biomarkers. An augmented SRM assay was used on the hemolymph that
284 included all the peptide transitions analyzed in the gonad with an additional 40 transitions from
285 five new proteins. These proteins were selected based on gonad proteome annotation to include
286 1) proteins that would likely be circulating in the hemolymph and 2) gonad proteins that had
287 homology with the mussel hemolymph proteome [29]. The additional peptides for SRM analysis of
288 hemolymph were from vitellogenin (2 proteins), glycogen synthase (2 proteins), and glycogenin-
289 1. PECAN generated spectral libraries, as described above, and Skyline software selected
290 hemolymph protein transitions to target during SRM analyses. For the new hemolymph proteins,
291 the minimums for peptides and transitions could not be met for every protein, but were still
292 included in the analysis. These transitions, and the previous gonad transitions, were analyzed
293 across sixteen geoduck hemolymph samples in two technical replicates, as described above for
294 the gonad. Some of these peptides yielded no data in the hemolymph, resulting in a dataset of
295 171 peptide transitions that were reliably detected. Raw data can be accessed in PeptideAtlas
296 under accession PASS00942.

297 Acquired SRM data in gonad and hemolymph were analyzed in Skyline for peptide
298 transition quantification. Skyline documents that were used to analyze the monitored peptide
299 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 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 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, therefore technical replicate peak intensities were averaged for the rest of the analysis (Supporting Information 5). 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[30] in R, clustering rows (peptide transitions) and columns (samples) using euclidean distance and the average clustering method.

Results

DDA Proteomics

The mass spectrometry data (PRIDE Accession #PXD003127) interpreted with a species- and tissue-specific transcriptome yielded 3,627 proteins inferred with high confidence across all gonad tissue samples (total spectral count across all replicates > 1) (Supporting Information 6). 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 ($R = 0.4122$, $p = 0.002$) and maturation stages ($R = 0.3025$, $p = 0.004$).

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 unique proteomic profile across all stages contained 305 proteins whereas 522 proteins were unique to males. The number of proteins unique to a specific maturation stage increased with maturity from 24 and 20 in early-stage females and males, respectively, to 161 and 145 in late-stage females and males (Figure 3). These unique proteomes were used to develop the downstream SRM assays. Enriched GO biological processes for the unique and full proteomes 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 7).

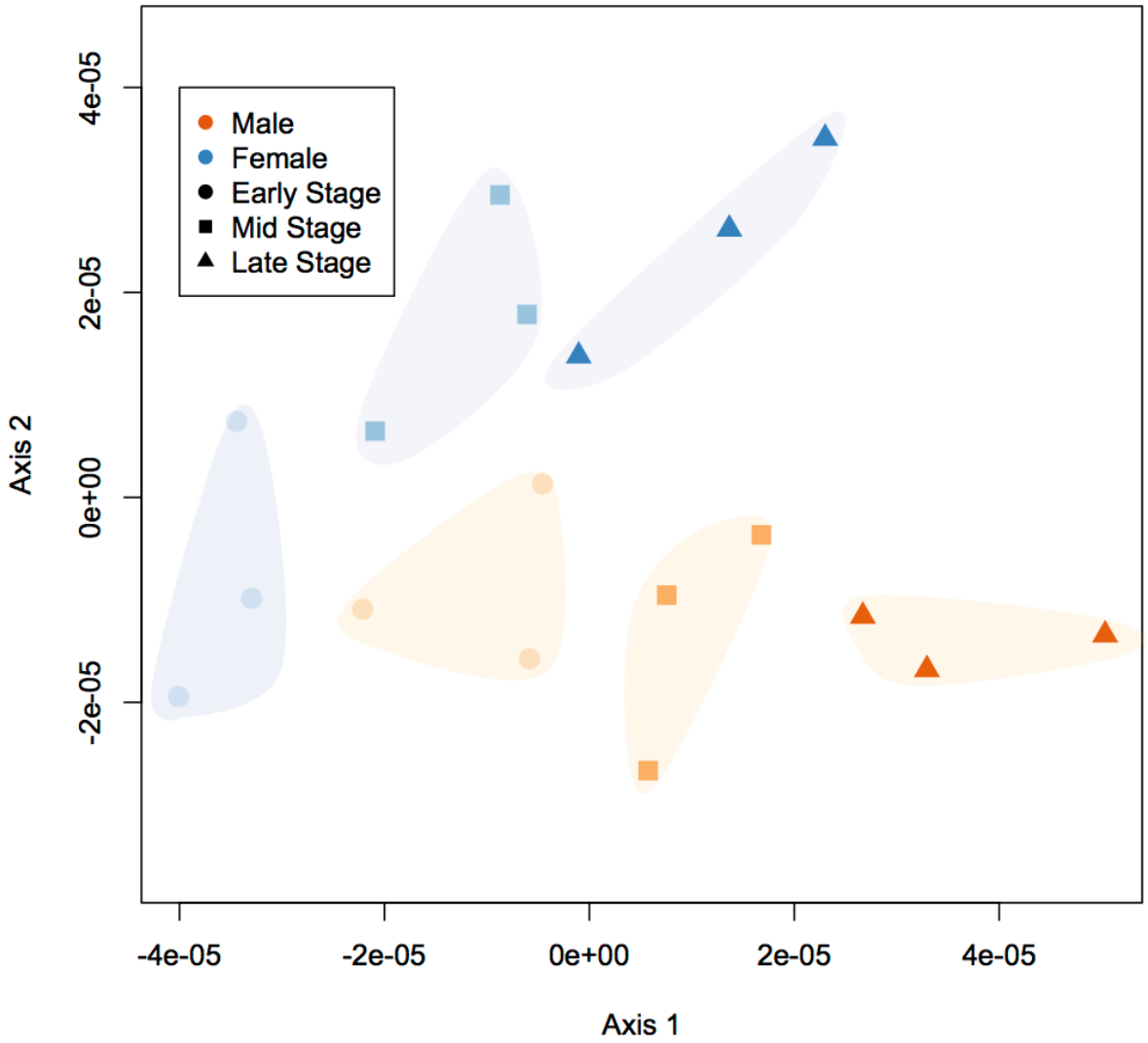
345 *Table 1. Enriched biological process GO terms for the detected and unique gonad proteomes. If*

346 *no GO terms are listed, no enriched processes were identified*

Gonad Stage	Detected Proteome GO terms	Unique Proteome GO terms
Early-stage Female	Actin filament organization; Translation	Protein phosphorylation
Mid-stage Female	Hexose metabolic process; Translation; Translational initiation	Regulation of hydrolase activity
Late-stage Female	DNA replication; Translation; Translational initiation	Protein glycosylation; Signal peptide processing; Homophilic cell adhesion via plasma membrane adhesion molecules
Early-stage Male	Translation; Oxidation-reduction process; Small molecule metabolic process	
Mid-stage Male	Translation; ATP metabolic process; Pyruvate metabolic process; Oxidation-reduction process; Nucleotide phosphorylation; Single organism catabolic process	
Late-stage Male	Translation; Cellular macromolecular complex assembly; Nucleotide metabolic process	Microtubule-based movement

347

348



349

350 *Figure 2. Non-metric multidimensional scaling plot (NMDS) of geoduck gonad whole proteomic*

351 *profiles generated by data dependent acquisition. Gonad proteomes differ among clams by both*

352 *sex (male = orange, female = blue) and stage (early-stage = circles, mid-stage = squares, late-*

353 *stage = triangles; $p<0.05$).*

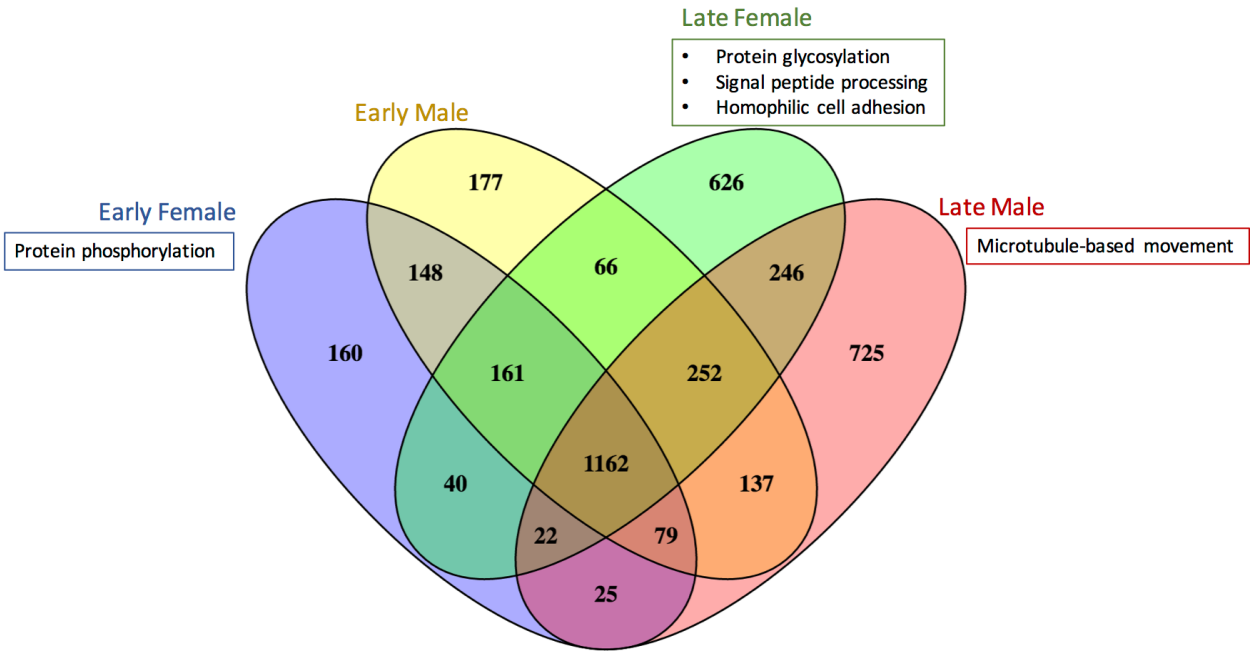
354

355 Numerous proteins were differentially abundant in comparisons between reproductive
356 stage and sex based on analysis with QSpec. The number of differentially abundant proteins
357 within a stage and between sexes increased from 387 between early-stage males and females to
358 1035 between late-stage males and females (Figure 4, Supporting Information 6). There were
359 more differentially abundant proteins between the stages within males than females (Supporting
360 Information 6). Enrichment analysis revealed which processes were over-represented in the sets
361 of differentially abundant proteins, when compared to the entire proteome (Figure 4 and Table
362 2).
363
364 *Table 2. GO biological processes enriched in proteins that were differentially abundant between*
365 *stages within a sex (e.g. early- vs. mid-stage female) or between sexes within a stage (e.g LF vs.*
366 *LM). Only the most specific GO terms in the GO hierarchy are listed.*
367

	Mid-stage Female	Late-stage Female	Early-stage Male	Mid-stage Male	Late-stage Male
Early-stage Female	Translation; DNA duplex unwinding	Translation; DNA duplex unwinding	Translation		
Mid-stage Female		DNA replication; DNA replication initiation; DNA duplex unwinding		Translation	
Late-stage Female	DNA replication; DNA				Translation; Microtubule-based process;

	replication initiation; DNA duplex unwinding				Translational initiation; Regulation of translational initiation
Early-stage Male				Ribonucleotide biosynthetic process; Purine ribonucleotide metabolic process; Purine ribonucleotide triphosphate biosynthetic process; Purine ribonucleoside metabolic process; Cytoskeletal anchoring at plasma membrane	Microtubule-based process; Oxidation-reduction process; Purine ribonucleoside triphosphate biosynthetic process; Purine ribonucleotide metabolic process
Mid-stage Male	Translation		Ribonucleotide biosynthetic process; Purine ribonucleotide metabolic process; Purine ribonucleotide triphosphate biosynthetic process; Purine ribonucleoside metabolic process; Cytoskeletal anchoring at plasma membrane		Monocarboxylic acid metabolic process; Microtubule-based movement

368



369

370 *Figure 3. Venn diagram with corresponding enriched biological processes for sex-stage full*

371 *proteomes of early-stage female (blue), early-stage male (yellow), late-stage female (green), and*

372 *late-stage male (red). Venn diagrams produced in Venny.* [31]

373

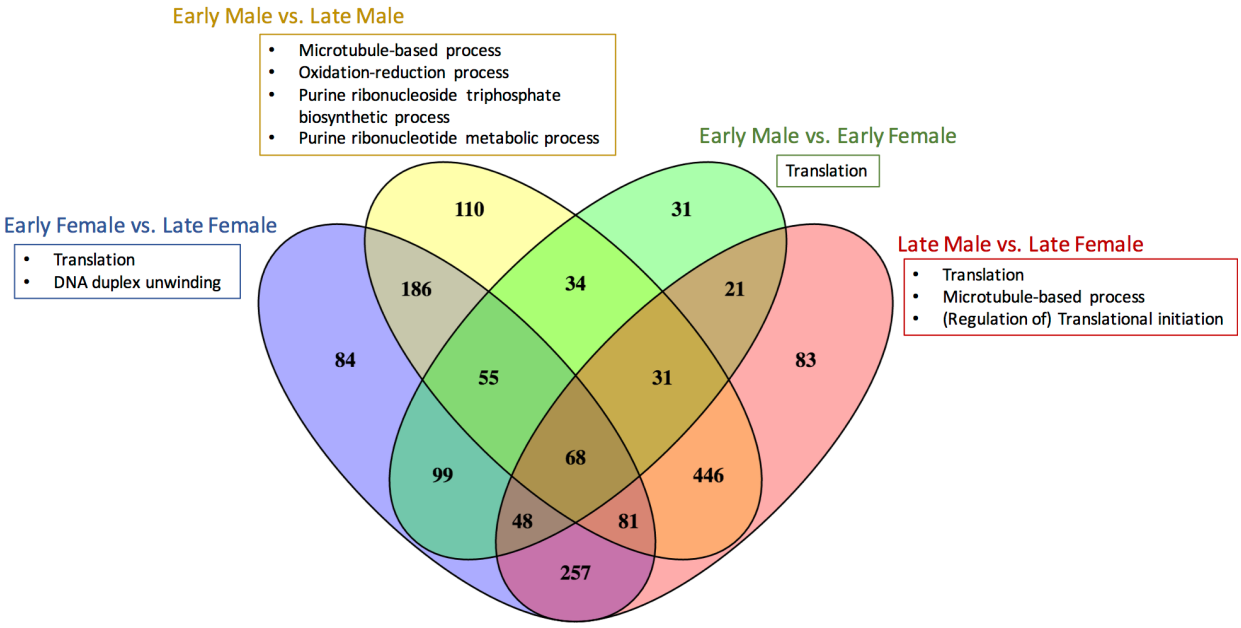


Figure 4. Venn diagram with corresponding enriched biological processes for differentially abundant proteins from QSpec comparisons between early-stage males (EM) and mid-stage males (MM); EM and late-stage males (LM); early-stage females (EF) and mid-stage females (MF); EF and late-stage females (LF). Venn diagrams produced in Venny.[31]

Targeted Assay Development: Selected Reaction Monitoring (SRM)

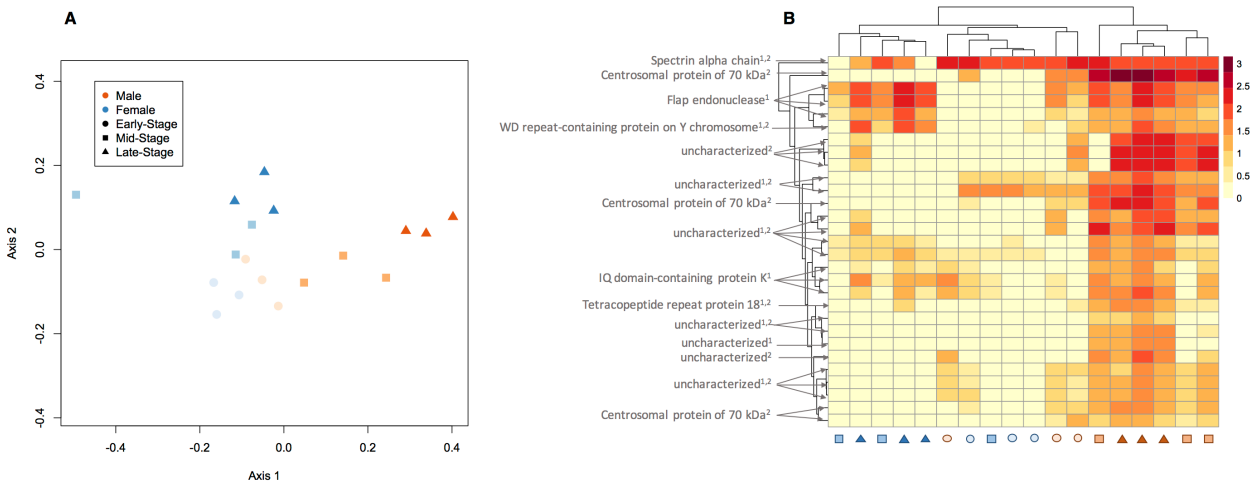
SRM was applied to create peptide assays that could resolve geoduck sex and maturation stages. Peptide transitions in the gonad (n=212) and hemolymph (n=171) derived from proteins detected in a single sex-maturation stage from one of the following; early-stage male (EM), early-stage female (EF), late-stage male (LM), and late-stage female (LF) proteomes were measured across 18 geoduck gonad samples and 16 hemolymph samples that included the 3 maturation stages and both sexes (Supporting Information 8). Many transitions had stable

coefficients of variation across all biological replicates, although more than 6% of the peptide transitions had CVs > 100, indicating a need for assay optimization (Supporting Information 9).

In an ordination plot, the gonad SRM data resolved males and females better in late-stage samples than the early- or mid-stage (Figure 5A). 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-stage female group from the rest of the geoduck (Figure 6A). There was significant separation of the 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 ($R=0.1435$, $p=0.065$).

Peptide transitions that directed the observed spread of the geoduck sex-stage data along Axis 1 and Axis 2 in the NMDS plots (determined by eigenvector analysis) were considered key analytes for differentiating groups. These peptide transitions represent a starting point for biomarker assay optimization. Peptides that drove the separation of the male maturation stages along Axis 1 in the gonad NMDS plot (Figure 5A) are from a set of 9 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 5B). Proteins that played the most significant role in the separation of the female stages along Axis 2 for the gonad data included WD repeat-containing protein on Y chromosome, tetratripcopeptide repeat protein 18, spectrin alpha chain, centrosomal protein of 70 kDa, and four uncharacterized proteins. Analysis of the hemolymph SRM NMDS (Figure 6A) reveals that the main peptide drivers along Axis 1 (separating late-stage females from all other geoduck) were from three proteins: flap endonuclease, vitellogenin, and one uncharacterized protein (Figure 6B). A cluster analysis of the 20 transitions that were primary drivers of sex

410 differentiation demonstrates that sex and stage groups from the gonad data clustered tightly, but
411 more specifically exposed a significant separation of males and females in the hemolymph data
412 (Figure 6B).
413



414

415 *Figure 5. A) NMDS of geoduck gonad peptide transition abundance generated by selected*
416 *reaction monitoring. Gonad proteomes differ among clams by both sex (male = orange, female*
417 *= blue) and stage (early = circles, mid = squares, late = triangles; $p<0.05$). B) Heatmap of log-*
418 *normalized peptide transition intensities for the top 20 most significant transitions (based on*
419 *NMDS analysis (A)) across all samples (early-stage/circles, mid-stage/squares, and late-*
420 *stage/triangles males/orange and females/blue) gonad. Each row (peptide transition) is labeled*
421 *by its corresponding protein annotation in gray. Superscripts indicate for which NMDS axis a*
422 *transition is significant.*
423

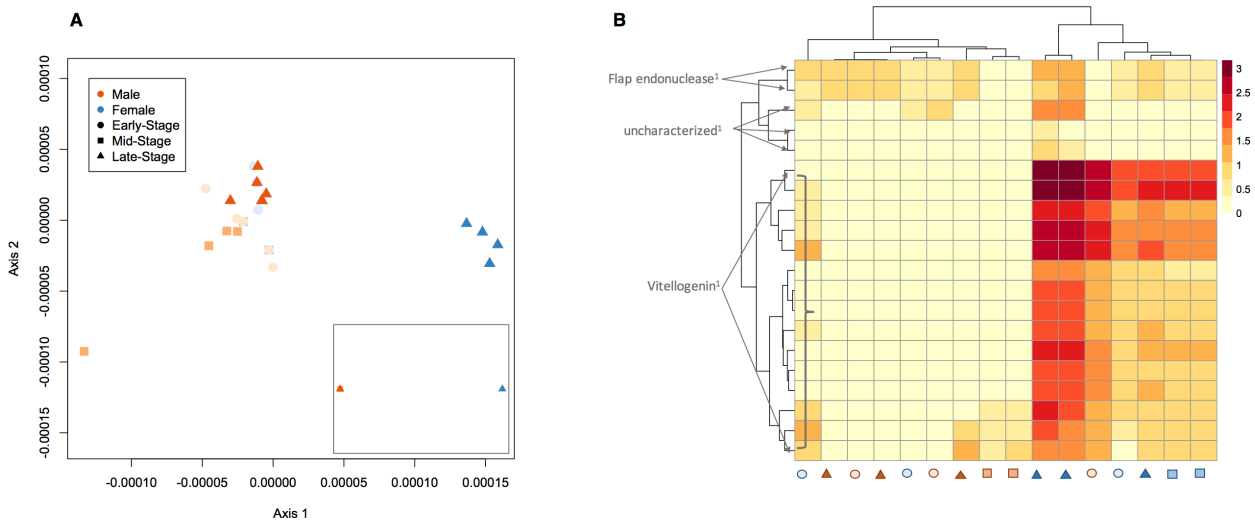


Figure 6. A) NMDS of geoduck hemolymph peptide transition abundance generated by selected reaction monitoring. The main plot contains a point for each technical and biological replicate; the inset (lower right-hand corner) shows the same data in a NMDS with the technical replicates averaged. Hemolymph proteomes differ among clams by sex (male = orange, female = blue; $p<0.05$). B) Heatmap of log-normalized peptide transition intensities for the top 20 most significant transitions (based on NMDS analysis (A)) 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 molecular physiological processes underlying functional phenotypic change, such as 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 3,627 detected proteins across both sexes and three maturation stages. This is a significant escalation in our

441 understanding of proteomic responses in maturation stages of marine mollusks compared to the
442 study by Li et al. [32] that focused on protein-chip assays of 139 peptides. Based on the DDA data,
443 27 proteins from early- and late-stage male and female clams were chosen for selected reaction
444 monitoring (SRM), or targeted proteomics, to initiate development of a biomarker assay to
445 identify geoduck reproductive stages based on 85 peptides. Peptide transitions (n=212) were
446 targeted from the selected peptides as candidates for biomarker development in gonad and 171
447 transitions were targeted in hemolymph.
448
449 *DDA*
450 There are clear proteomic profile differences in the geoduck gonad by both sex and
451 maturation stage and these differences reveal the biochemical pathways underlying tissue
452 reproductive specialization. More proteins are differentially abundant between early and late
453 stages of reproductive maturity, compared to early and mid-stages, for both males and females,
454 reflecting that, as maturation progresses, gonad tissue becomes highly specialized to male or
455 female reproduction. Throughout geoduck reproductive maturation, the gonad goes from
456 undifferentiated connective tissue to a highly specialized reproductive tissue that can release
457 fully mature sperm and eggs. These dramatic phenotypic changes required for reproductive
458 maturity are visible with histology [12] and are realized via changes at the molecular level that
459 prompt the biochemical shifts necessary to create mature gonad. For example, the zona pellucida
460 (ZP) containing protein (cds.comp134923_c0_seq3|m.23445) and vitellogenin
461 (cds.comp144315_c0_seq1|m.50928) became significantly more abundant in the female gonad as
462 geoduck matured. The zona pellucida surrounds vertebrate oocytes and its invertebrate homolog
463 is the vitelline envelope, which is likely the source of this protein annotation and is the site of

464 sperm-oocyte recognition and binding. Vitellogenin is an egg yolk protein that is essential in
465 reproduction and development across taxa.

466 The sex- and stage-specific proteomes were enriched for specific biological processes,
467 further demonstrating tissue specialization. In females, the early-stage gonad proteome was
468 enriched in proteins involved in protein phosphorylation, the mid-stage was enriched in
469 regulation of hydrolase activity and carbohydrate metabolism, and the late-stage was enriched in
470 protein glycosylation, signal peptide processing, DNA replication, and cell adhesion. These
471 shifts in biological pathways clarify the mechanisms behind what has already been observed in
472 bivalve tissue morphology during maturation. The enrichment in the process "protein
473 phosphorylation" likely represents the activation of translated proteins required to effect gonad
474 structural and functional changes. The enrichment of carbohydrate metabolism in mid-stage
475 females reflects the increased mobilization of carbohydrates to support the energy-intensive
476 process of gonad maturation. Previous biochemical work has highlighted the importance of
477 glycogen metabolism during bivalve reproductive maturation [33,34] and carbohydrate metabolism
478 during vertebrate oocyte maturation [35,36], however, a more complex overview of gonad
479 carbohydrate metabolism is revealed in the proteomic data. The proteins contributing to the
480 enrichment of carbohydrate metabolism processes in mid-stage females are involved in glycan
481 metabolism (2 enzymes), glycolysis (3 enzymes), gluconeogenesis (2 enzymes), and the pentose
482 phosphate pathway (2 enzymes). These enzymes were detected in the geoduck gonad, but were
483 not differentially abundant. It is likely that, similar to other marine invertebrates, there is a rapid
484 upregulation of glycolysis in geoduck eggs upon fertilization.[37]

485 As geoduck females entered the late stages of maturation, more proteins were detected
486 that are related to cell replication as well as cell adhesion, which is an essential biological

process in maturing oocytes in vertebrates.[38,39] The strong cell adhesion signal in late-stage females is driven in large part by the many cadherin and protocadherin proteins detected in a single female. However, seven protocadherins were detected at significantly higher abundance in more than a single late-stage female - compared to early- and mid-stages. Further, laminin, talin, and vinculin were significantly less abundant in late-stage females compared to other females. Evidence for the importance of DNA replication was more consistent: DNA helicase, DNA ligase, proliferating cell nuclear antigen, and ribonucleoside-diphosphate reductase were all more abundant in late-stage females than early- or mid-stages. This surge in DNA replication-associated proteins is likely a reflection of an up-regulation of meiosis and/or mitosis. These changes represent the diversity in function of female gonad tissue as it undergoes reproductive maturation.

The male tissue similarly undergoes increased complexity and specialization: the early-stage male gonad proteome was enriched in translation processes, oxidation-reduction processes, and small molecule metabolic process; the mid-stage male proteome was enriched in ATP metabolic process, pyruvate metabolism, oxidation-reduction process, and nucleotide phorphorylation; 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 reflected a greater metabolic demand in the tissue as it matures. This was particularly reflected in the increased abundance of ATP synthase proteins in mid-stage males relative to early-stage males and the decreased abundance of V-type proton ATPase relative to early-stage males (this trend continued, although not significantly, in late-stage

510 males). ATP synthase produces ATP from a proton gradient while V-type proton ATPase
511 consumes ATP to fuel a proton gradient. These pieces of evidence suggest a physiological need
512 for increased ATP production as male maturation proceeds. However, glycolysis pathway
513 enzymes trended towards decreased abundance in mid- versus early-stage males, suggesting a
514 decrease in ATP generated via this pathway, which is responsible for the proton gradient that
515 drives ATP synthase. There may be a trend of ATP consumption over production during male
516 geoduck gametogenesis. Adenosine phosphate molecule consumption was observed in mussel
517 gonad tissue, where amounts of AMP, ADP, and ATP decreased as gametogenesis
518 progressed.[40] In late-stage males, the enriched processes reflected an increase in cell replication
519 during the last steps of sperm formation. Several proteins implicated in mitosis and meiosis were
520 detected at significantly increased abundance in late-stage males compared to early- and mid-
521 stage: centrosomal proteins, dynein, intraflagellar transport protein, kinesin, and alpha and beta
522 tubulin. All of these proteins are instrumental in the cell division and replication and their
523 importance at this maturation stage suggests the presence of rapidly dividing cells. This could be
524 a sign of both meiosis and mitosis. In mussels off the Atlantic coast of Spain, meiotic division
525 begins approximately four months before the onset of spawning, while mitosis reconstructs the
526 gonad during and after spawning.[41] These pieces of proteomic evidence point towards increased
527 cellular energy metabolism and cell division during the development of mature, motile sperm.
528 Even with a highly specific protein identification database, the DDA method excludes
529 many important, relatively low abundance proteins from analysis.[42] The ten most abundant
530 proteins (as measured by total spectral count) for each sex-maturation stage accounted for 14-
531 26% of the total spectral counts across all proteins for a given sex-stage. Many of these proteins
532 are “housekeeping” proteins, such as actin and myosin, and their dominance in the peptide

533 mixture was likely masking many informative, low abundance peptides. Similarly, in zebrafish,
534 highly abundant vitellogenin (also highly abundant in the mid- and late-stage females in this
535 study) made it difficult to detect lower abundant proteins using DDA.[43] Vitellogenin can be an
536 informative biomarker, but by masking the detection of other proteins that may have a more
537 nuanced expression profile, it and other highly abundant proteins can obscure the finer-scale
538 changes that occur.
539
540 *SRM*
541 DDA analysis yields biologically relevant results, but for practical purposes of biomarker
542 development we do not need the abundance profiles of >3000 proteins. Selected reaction
543 monitoring allows for the selection and accurate detection of only the most informative proteins'
544 peptides to address a specific hypothesis. We leveraged the DDA dataset to create informative
545 SRM assays of geoduck sex and maturation stage in both gonad and hemolymph. Both of these
546 tissues have the potential for being the basis of non-lethal assays for broodstock assessment in
547 aquaculture. Reproductive maturation is a complex process and involves suites of interacting
548 proteins to achieve the end goal of mature gametes. By applying an unbiased method in the first
549 step of our biomarker development (DDA) we were able to detect and include these informative
550 peptides in our assay.
551 In the gonad tissue, we measured 212 peptide transitions that could differentiate males
552 and females as well as early and late maturation stages. Our assay is based on proteins that are
553 involved in calcium ion binding, phosphorylation, meiosis, DNA replication and repair, and
554 membrane structure. Additionally, seven of these informative proteins were unannotated, so we
555 cannot know their function based on homology. These unknown proteins may represent taxon-

556 specific reproductive proteins and warrant further characterization. Reproductive proteins are
557 some of the fastest-evolving, resulting in many species-specific proteins.[44] A subset of the
558 peptide transitions from these proteins (n=29) were deemed highly informative based on our
559 analysis and would be good candidates for a more streamlined SRM assay. These transitions
560 correspond to six annotated proteins and five unannotated proteins. The annotated proteins are
561 implicated in cytoskeleton structure (spectrin alpha chain), DNA replication and repair
562 (centrosomal protein of 70 kDa, flap endonuclease), and proteins that were identified by motifs
563 and thus yield little functional information (WD repeat-containing protein on Y chromosome, IQ
564 domain containing protein K, and tetracopeptide repeat protein 18). The 29 transitions were
565 almost uniformly at higher abundance in males than females, which is likely because we
566 included more proteins derived from the late-stage male proteome in our assay since these
567 yielded transitions with higher MS-based ion intensities.

568 Hemolymph is a more reliable non-lethal option for sampling and possesses differentially
569 abundant peptides based on sex and reproductive stage. The peptide-based assay developed from
570 the evidence found in the gonad tissue that revealed sex and stage was applied to circulating
571 hemolymph. To the gonad tissue SRM transition list we added 40 more transitions
572 that were related to reproductive maturation. Proteome profiles of bivalve hemocytes can reveal
573 patterns of reproductive status, such as differing oocyte quality in Pacific oysters, as measured by
574 larval production.[45] However, biomarkers of tissue-specific changes can be difficult to detect in
575 circulating fluid since hemolymph (or blood) is not the primary site of physiological change. The
576 sensitivity of targeted proteomics has a lower limit of detection than DDA and is sensitive
577 enough to detect some biomarkers of interest.[46] But even if the same proteins were circulating in
578 the hemolymph as were detected in the gonad, they may not be informative because 1) the

579 proteins could be fragmented in ways that make them undetectable with our assay or 2) the
580 protein abundance could be changing out of sync with reproductive changes. We were able to
581 collect data on 171 peptide transitions that were derived from gonad tissue but detectable in the
582 hemolymph. Peptide transitions from 16 proteins were reliably detected in both gonad and
583 hemolymph, however, only two proteins had transitions that were highly informative in both
584 tissues: flap endonuclease and an uncharacterized protein. The inclusion of vitellogenin in our
585 assay confirmed its presence in geoduck hemolymph during later stages of reproductive maturity
586 in females, but since it was such a strong physiological signal it dominated our statistical
587 analysis. It was similarly the strongest signal in hemocytes of maturing Pacific oysters [45] and the
588 basis of a SRM assay developed in sea turtles [47]. Since we were limited to the proteins from the
589 gonad DDA dataset for our assay development, there are likely additional circulating proteins
590 that would expand our hemolymph assay and make it informative beyond distinguishing late-
591 stage females.
592
593 *Conclusions*
594 Through characterization of the geoduck gonad proteome and development of proteomic
595 assays for reproductive maturity, we not only exponentially increased the molecular resources
596 available for this species, but we also developed effective tools for non-lethal detection of sex
597 and maturation stage in this species. The gonad peptide transitions could be further decreased to
598 a set of 29 transitions that would accurately predict a) geoduck sex and b) reproductive
599 maturation status. In the hemolymph, 20 peptide transitions can accurately differentiate late-
600 stage female geoduck from others. Together these molecular tools can directly address the

601 geoduck production problem of asynchronous spawning due to different maturation stages and
602 unknown sexes.
603
604 **Acknowledgements**
605 This work was supported by and funded by a grant from the University of Washington
606 Royalty Research Fund (to SR), a Training Grant from the National Institutes of Health (to ETS)
607 (T32 HG00035), and the University of Washington's Proteomics Resource (UWPR95794). We
608 would like to thank Jimmy Eng and Priska von Haller for their help in data acquisition and
609 analysis; Sonia Ting, Jarrett Egertson, Lindsay Pino, Yuval Boss, and Nick Shulman for
610 assistance with data analysis; and the Genome Sciences Information and Technology group for
611 their technical support. ETS and BLN would like to thank IJE and TAN for their ongoing
612 inspiration.
613
614
615 **Supporting Information.** The following files are available free of charge via the internet
616 goo.gl/uoi9H1
617
618 **S1-Pgenerosa-gonad-protein-v3.fasta** Deduced protein sequences derived from gonad
619 transcriptome sequencing
620 **S2-comet.params** Parameter file used in Comet searches for the DDA data
621 **S3-technical-reps.pdf** NMDS of the whole proteomic profiles (DDA data) for the three
622 biological replicates and corresponding three technical replicates for each sex and maturation
623 stage. Blue points represent female proteomes and orange represent male. Shapes and shade

624 (light to dark) represent the different stages: circles for early-stage, squares for mid-stage, and
625 triangles for late-stage
626 **S4-technical-reps-box-plots.pdf** Box plot of the coefficients of variation for protein spectral
627 counts from DDA data A) across technical replicates for each geoduck biological replicate and
628 B) across biological replicates for each sex-maturation stage. CVs were calculated in R using the
629 raster package. X-axis labels correspond to the geoduck sex and maturation stage, indicated by
630 e.g. "EF3" represents early-stage female sample 3. The boxes represent the upper and lower
631 quartiles of the data distribution; horizontal black line represents median value; "whiskers"
632 extend to the greatest and least values, excluding outliers; open circles represent outliers (more
633 or less than 3/2 times the upper or lower quartiles
634 **S5-NMDS-SRM-tech-reps.pdf** NMDS of gonad SRM data for the 3 biological replicates and
635 corresponding three technical replicates for each sex and maturation stage. Blue points represent
636 female proteomes and orange represent male. Shapes and shade (light to dark) represent the
637 different stages: circles for early-stage, squares for mid-stage, and triangles for late-stage
638 **S6-spc-nsaf-prot-names.txt** All identified proteins from the DDA experiment with Uniprot
639 annotations (e-value cut-off of 1E-10), total spectral counts for each technical replicate,
640 calculated normalized spectral abundance factor (NSAF) for combined technical replicates, and
641 indication of significantly differentially abundant proteins by pairwise comparison. Columns
642 containing spectral count data have headers "SpC" followed by the biological replicate number,
643 sex, maturation stage, and technical replicate (for example, "SpC 3FE" is technical replicate 1
644 from early-stage female 3). Notation for NSAF is similar to SpC. The 9 columns in the sheet
645 have headers such as "EFvLF" (comparison between early- and late-stage females) have
646 asterisks in the cells that correspond to proteins that were differentially abundant for each given

647 comparison. The last two columns ("NMDS Gonad" and "NMDS Hemolymph") have asterisks
648 in cells that correspond to the proteins with peptide transitions that contribute significantly to the
649 SRM NMDS plot distributions

650 **S7_DAGs.pdf** Enrichment plots for early- and mid-maturation stage geoduck (EM, EF, MM,
651 MF). Processes colored in blue were enriched from the set of all proteins detected in the given
652 sex-stage, while red processes are from the protein set that was uniquely detected in a sex-stage.
653 Darker colors represent higher significance of enrichment

654 **S8-SRM-data.xlsx** Raw Skyline output (in the tab "Skyline output) and peak intensities
655 normalized by QC peptide abundance ("Normalized Intensities") for the gonad and hemolymph
656 SRM data.

657 **S9-SRM-cv-boxplots.pdf** Integrated peak area coefficient of variation values across all gonad
658 peptide transitions (n=212) and hemolymph peptide transitions (n=171) A) across technical
659 replicates for each geoduck biological replicate in gonad tissue, B) across biological replicates
660 for each sex-maturation stage in gonad tissue, C) across technical replicates for each geoduck
661 biological replicate in hemolymph, and D) across biological replicates for each sex-maturation
662 stage in hemolymph. CVs were calculated in R using the raster package. X-axis labels
663 correspond to the geoduck sex and maturation stage, indicated by e.g. "EF3" represents early-
664 stage female sample 3. The boxes represent the upper and lower quartiles of the data distribution;
665 horizontal black line represents median value; "whiskers" extend to the greatest and least values,
666 excluding outliers; open circles represent outliers (more or less than 3/2 times the upper or lower
667 quartiles)

668

669 **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) Northern Economics, I. *The Economic Impact of Shellfish Aquaculture in Washington, Oregon and California. Prepared for Pacific Shellfish Institute*; 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) 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 oysterCrassostrea

corteziensis(Hertlein 1951) as a tool to predict the degree of gonad maturity. *Aquac. Res.* **2009**, *40* (6), 644–655.

(11) Li, Y.; Yaqin, L.; Ghazala, S.; 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.

(12) Crandall, G.; Roberts, S. Reproductive Maturation in Geoduck clams (Panopea generosa) https://figshare.com/articles/Reproductive_Maturation_in_Geoduck_clams_Panopea_generosa_/3205975/1. DOI: https://dx.doi.org/10.6084/m9.figshare.3205975.v1

(13) 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.

(14) Grabherr, M. G.; Haas, B. J.; Yassour, M.; Levin, J. Z.; Thompson, D. A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.; et al. *Full-length transcriptome assembly from RNA-seq data without a reference genome*; 2011.

(15) Eng, J. K.; Jahan, T. A.; Hoopmann, M. R. Comet: an open-source MS/MS sequence database search tool. *Proteomics* **2013**, *13* (1), 22–24.

(16) 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. *J. Am. Soc. Mass Spectrom.* **2015**, *26* (11), 1865–1874.

(17) Pedrioli, P. G. A. Trans-Proteomic Pipeline: A Pipeline for Proteomic Analysis. In *Methods in Molecular Biology*; 2009; pp 213–238.

(18) 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.

(19) 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.; et al. *vegan: Community Ecology Package*; 2016.

(20) R Core Team. *R: A language and environment for statistical computing*; R Foundation for Statistical Computing, Vienna, Austria, 2015.

(21) 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.

(22) Choi, H.; Fermin, D.; Nesvizhskii, A. I. Significance analysis of spectral count data in label-free shotgun proteomics. *Mol. Cell. Proteomics* **2008**, *7* (12), 2373–2385.

(23) MacLean, B.; Tomazela, D. M.; Shulman, N.; Chambers, M.; Finney, G. L.; Frewen, B.; Kern, R.; Tabb, D. L.; Liebler, D. C.; MacCoss, M. J. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **2010**, *26* (7), 966–968.

(24) 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.

(25) 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.

737 (26) Ting, Y. S.; Egertson, J. D.; Payne, S. H.; Kim, S.; MacLean, B.; Käll, L.; Aebersold, R.;
738 Smith, R. D.; Noble, W. S.; MacCoss, M. J. Peptide-Centric Proteome Analysis: An
739 Alternative Strategy for the Analysis of Tandem Mass Spectrometry Data. *Mol. Cell.*
740 *Proteomics* **2015**, *14* (9), 2301–2307.
741 (27) Wolf-Yadlin, A.; Hautaniemi, S.; Lauffenburger, D. A.; White, F. M. Multiple reaction
742 monitoring for robust quantitative proteomic analysis of cellular signaling networks. *Proc.*
743 *Natl. Acad. Sci. U. S. A.* **2007**, *104* (14), 5860–5865.
744 (28) Picotti, P.; Aebersold, R. Selected reaction monitoring-based proteomics: workflows,
745 potential, pitfalls and future directions. *Nat. Methods* **2012**, *9* (6), 555–566.
746 (29) Moreira, R.; Pereiro, P.; Canchaya, C.; Posada, D.; Figueras, A.; Novoa, B. RNA-Seq in
747 Mytilus galloprovincialis: comparative transcriptomics and expression profiles among
748 different tissues. *BMC Genomics* **2015**, *16* (1), 728.
749 (30) Kolde, R. *pheatmap: Pretty Heatmaps*; 2015.
750 (31) Oliveros, J. C. *Venny. An interactive tool for comparing lists with Venn's diagrams*; 2007-
751 2015.
752 (32) Li, Y.; Yaqin, L.; Ghazala, S.; Wikfors, G. H. Non-lethal determination of sex and
753 reproductive condition of Eastern oysters Crassostrea virginica Gmelin using protein
754 profiles of hemolymph by Proteinchip® and SELDI-TOF-MS technology. *Aquaculture*
755 **2010**, *309* (1-4), 258–264.
756 (33) Beukema, J. J.; De Bruin, W. Seasonal changes in dry weight and chemical composition of
757 the soft parts of the tellinid bivalve Macoma balthica in the Dutch Wadden Sea. *Neth. J. Sea*
758 *Res.* **1977**, *11* (1), 42–55.

759 (34) Li, Q.; Qi, L.; Makoto, O.; Katsuyoshi, M. Seasonal biochemical variations in Pacific oyster
760 gonadal tissue during sexual maturation. *Fish. Sci.* **2000**, *66* (3), 502–508.
761 (35) Dworkin, M. B.; Eva, D.-R. Metabolic regulation during early frog development:
762 Glycogenic flux in Xenopus oocytes, eggs, and embryos. *Dev. Biol.* **1989**, *132* (2), 512–
763 523.
764 (36) Harris, S. E.; Adriaens, I.; Leese, H. J.; Gosden, R. G.; Picton, H. M. Carbohydrate
765 metabolism by murine ovarian follicles and oocytes grown in vitro. *Reproduction* **2007**, *134*
766 (3), 415–424.
767 (37) Yasumasu, I.; Tazawa, E.; Fujiwara, A. Glycolysis in the eggs of the echiuroid, Urechis
768 unicinctus and the oyster, Crassostrea gigas. *Exp. Cell Res.* **1975**, *93* (1), 166–174.
769 (38) Fusi, F. M.; Vignali, M.; Busacca, M.; Bronson, R. A. Evidence for the presence of an
770 integrin cell adhesion receptor on the oolemma of unfertilized human oocytes. *Mol. Reprod.*
771 *Dev.* **1992**, *31* (3), 215–222.
772 (39) Campbell, S.; Swann, H. R.; Seif, M. W.; Kimber, S. J.; Aplin, J. D. Integrins and adhesion
773 mlecules: Cell adhesion molecules on the oocyte and preimplantation human embryo. *Hum.*
774 *Reprod.* **1995**, *10* (6), 1571–1578.
775 (40) Díaz Enrich, M. J.; Ibarguren, I. Implication of adenosine 3’,5'-cyclic monophosphate,
776 guanosine 3',5'-cyclic monophosphate, adenosine 5'-mono-, di-, and triphosphate and
777 fructose-2,6-bisphosphate in the regulation of the glycolytic pathway in relation to the
778 gametogenic cycle in the mussel mytilus galloprovincialis Lmk. *Mol. Cell. Biochem.* **2003**,
779 *252* (1-2), 339–345.

780 (41) Suárez, M. P.; Alvarez, C.; Molist, P.; Fs, A. J. Particular aspects of gonadal cycle and
781 seasonal distribution of gametogenic stages of Mytilus galloprovincialis cultured in the
782 estuary of Vigo. *J. Shellfish Res.* **2005**, *24* (2), 531–540.
783 (42) Pedersen, S. K.; Harry, J. L.; Sebastian, L.; Baker, J.; Traini, M. D.; McCarthy, J. T.;
784 Manoharan, A.; Wilkins, M. R.; Gooley, A. A.; Righetti, P. G.; et al. Unseen proteome:
785 mining below the tip of the iceberg to find low abundance and membrane proteins. *J.*
786 *Proteome Res.* **2003**, *2* (3), 303–311.
787 (43) Groh, K. J.; Nesatyy, V. J.; Segner, H.; Eggen, R. I. L.; Suter, M. J.-F. Global proteomics
788 analysis of testis and ovary in adult zebrafish (Danio rerio). *Fish Physiol. Biochem.* **2011**,
789 *37* (3), 619–647.
790 (44) Swanson, W. J.; Vacquier, V. D. THE RAPID EVOLUTION OF REPRODUCTIVE
791 PROTEINS. *Nat. Rev. Genet.* **2002**, *3* (2), 137–144.
792 (45) Corporeau, C.; Vanderplancke, G.; Boulais, M.; Suquet, M.; Quéré, C.; Boudry, P.; Huvet,
793 A.; Madec, S. Proteomic identification of quality factors for oocytes in the Pacific oyster
794 Crassostrea gigas. *J. Proteomics* **2012**, *75* (18), 5554–5563.
795 (46) Klee, E. W.; Bondar, O. P.; Goodmanson, M. K.; Dyer, R. B.; Erdogan, S.; Bergstralh, E. J.;
796 Bergen H., R.; Sebo, T. J.; Klee, G. G. Candidate Serum Biomarkers for Prostate
797 Adenocarcinoma Identified by mRNA Differences in Prostate Tissue and Verified with
798 Protein Measurements in Tissue and Blood. *Clin. Chem.* **2012**, *58* (3), 599–609.
799 (47) Plumel, M. I.; Wasselin, T.; Plot, V.; Strub, J.-M.; Van Dorsselaer, A.; Carapito, C.;
800 Georges, J.-Y.; Bertile, F. Mass spectrometry-based sequencing and SRM-based
801 quantitation of two novel vitellogenin isoforms in the leatherback sea turtle (Dermochelys
802 coriacea). *J. Proteome Res.* **2013**, *12* (9), 4122–4135.