1   **Integrating discovery-driven proteomics and selected reaction monitoring to develop a non-**

2   **invasive assay for geoduck reproductive maturation**

3

4   Emma B. Timmins-Schiffman[1], Grace A. Crandall[2], Brent Vadopalas[2], Michael E. Riffle[3], Brook

5   L. Nunn[1], Steven B. Roberts[2,]*

6

7   [1] Department of Genome Sciences, University of Washington, Seattle, WA 98105

8   [2] School of Aquatic and Fishery Sciences, University of Washington, Seattle, WA 98105

9   [3] Department of Biochemistry, University of Washington, Seattle, WA 98105

10   *Corresponding author:

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

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

13

14

15

16

17

18

19

20

21

22

23

24   **Abstract**

25   Geoduck clams (*Panopea generosa*) are an increasingly important fishery and

26   aquaculture product along the eastern Pacific coast from Baja California, Mexico to Alaska.

27   These long-lived clams are highly fecund, though sustainable hatchery production of genetically

28   diverse larvae is hindered by the lack of sexual dimorphism, resulting in asynchronous spawning

29   of broodstock, unequal sex ratios, and low numbers of breeders. Development of assays of gonad

30   physiology could indicate sex and maturation stage, as well as be used to assess the status of

31   natural populations. Proteomic profiles were determined for three reproductive maturation stages

32   in both male and female clams using data dependent acquisition (DDA) of gonad proteins.

33   Gonad proteomes became increasingly divergent between males and females as maturation

34   progressed. The DDA data was used to develop targets analyzed with selected reaction

35   monitoring (SRM) in gonad tissue as well as hemolymph. The SRM assay yielded a suite of

36   indicator peptides that can be used as an efficient assay to determine geoduck gonad maturation

37   status. Application of SRM in hemolymph samples demonstrates this procedure could effectively

38   be used to assess reproductive status in marine mollusks in a non-lethal manner.

39

40   **Key words: reproduction, gonad, geoduck, hemolymph, mollusk, proteomics, SRM,**

41   **targeted proteomics**

42

43

44

45

46

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 lipids [6], specific gene transcripts [7,8], and

70   RNA/DNA ratios [9] have clarified some of the physiology underlying molluscan sexual

71   maturation. A comprehensive study of the Pacific oyster gonadal gene expression across

72   reproductive stages further explained the molecular mechanisms at play [10]; however, none of

73   these studies translated physiological discovery into an applied assay.

74   Previous studies have demonstrated the promise of protein based assays for

75   characterization of reproductive status in bivalves. Arcos et al. [11] explored the use of enzyme-

76   linked immunoassay assays (ELISA) to quantify proteins associated with vitellogenesis to

77   determine oyster oocyte maturation. Although the two proteins targeted by Arcos et al. [11]

78   allowed distinction of males and females and differentiation of female reproductive stages, the

79   assay relies on a single physiological measure and cannot be adapted to a non-lethal assay. [4] The

80   complexity of molluscan reproductive maturation at the molecular level has been proven by

81   global profiling studies [10,12] that reveal possibilities for an informative assay that leverages the

82   involvement of diverse biochemical pathways.

83   Here, we apply mass spectrometry (MS)-based proteomics technology to provide a large-

84   scale, unbiased approach for examining relative abundances of all proteins present in gonad

85   tissue across several maturation stages of both female and male geoducks. Whole proteome

86   characterization via data dependent acquisition (DDA) provided basic information on

87   gametogenesis, gonad maturation, and informed peptide selection for the determination of sex

88   and maturation stage in subsequent downstream MS analyses. Selected peptides underwent

89   targeted proteomic analyses via selected reaction monitoring (SRM) to more accurately quantify

90   protein abundance. Together, these data provide fundamental insight into marine mollusk

91   reproductive biology and demonstrate effective non-invasive peptide quantification on

92   circulating hemolymph.

93

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 ~5-15µ), mid-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), mid-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.[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).

115

116   *Protein Preparation*

117   Gonad tissue from the eighteen geoduck clams was sonicated in lysis buffer (50 mM

118   NH_4HCO_3, 6 M urea). Gonad tissue protein homogenate content and hemolymph protein content

119   were quantified using Pierce BCA Protein Assay Kit (Thermo Fisher Scientific, Waltham, MA;

120   catalog #23225). Protein (100 µg - gonad samples, 50 µg - hemolymph) was evaporated to ~20

121   µl and resuspended in 100 µl of lysis buffer followed by sonication. Protein digestion for gonad

122   tissue samples and hemolymph followed the protocol outlined in Timmins-Schiffman et al.[14]

123   Briefly, each sample was incubated with 200 mM tris(2-carboxyethyl)phosphine buffered at pH

124   8.8 (1 hr, 37°C). Samples were alkylated with 200 mM iodoacetamide (IAM; 1hr, 20°C)

125   followed by an incubation (1 hr) with 200 mM dithiothreitol to absorb any remaining IAM. To

126   each sample, NH_4HCO_3 and HPLC grade methanol were added to dilute urea and to increase

127   solubilization of membrane proteins. Samples were digested overnight with trypsin at

128   37°C. Digested samples were evaporated and reconstituted in 5% acetonitrile (ACN) + 0.1%

129   trifluoroacetic acid (TFA) (100 µl) and pH was decreased to < 2. Desalting of the samples was

130   done using Macrospin columns (sample capacity 0.03-300 µg; The Nest Group, Southborough,

131   MA, USA) following the manufacturer's specifications. Dried peptides were reconstituted in

132   100 µl of 5% ACN + 0.1% formic acid.

133

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

135   Data-dependent acquisition (DDA) was performed to examine gonad proteomic profiles

136   across samples (Figure 1). Liquid chromatography coupled with tandem mass spectrometry (LC-

137   MS/MS) was accomplished on a Q-Exactive-HF (Thermo) on technical triplicates for each

138   sample. The analytical column (20 cm long) was packed in-house with C18 beads (Dr. Maisch

139   GmbH HPLC, Germany, 0.3 μm) with a flow rate of 0.3 μl/min. Chromatography was

140   accomplished with an increasing ratio of solvent A (ACN + 0.1% formic acid): solvent B (water

141   + 0.1% formic acid). The solvent gradient consisted of: 0-1 minutes of 2-5% solvent A; 1-60

142   minutes 5-35% solvent A; 60-61 minutes 35-80% solvent A; 61-70 minutes 80% solvent A; 71-

143   90 minutes 80-2% solvent A. Quality control standards (Pierce Peptide Retention Time

144   Calibration mixture (PRTC) + bovine serum albumin peptides (BSA)) were analyzed throughout

145   the experiment to ensure consistency of peptide detection and elution times.

146
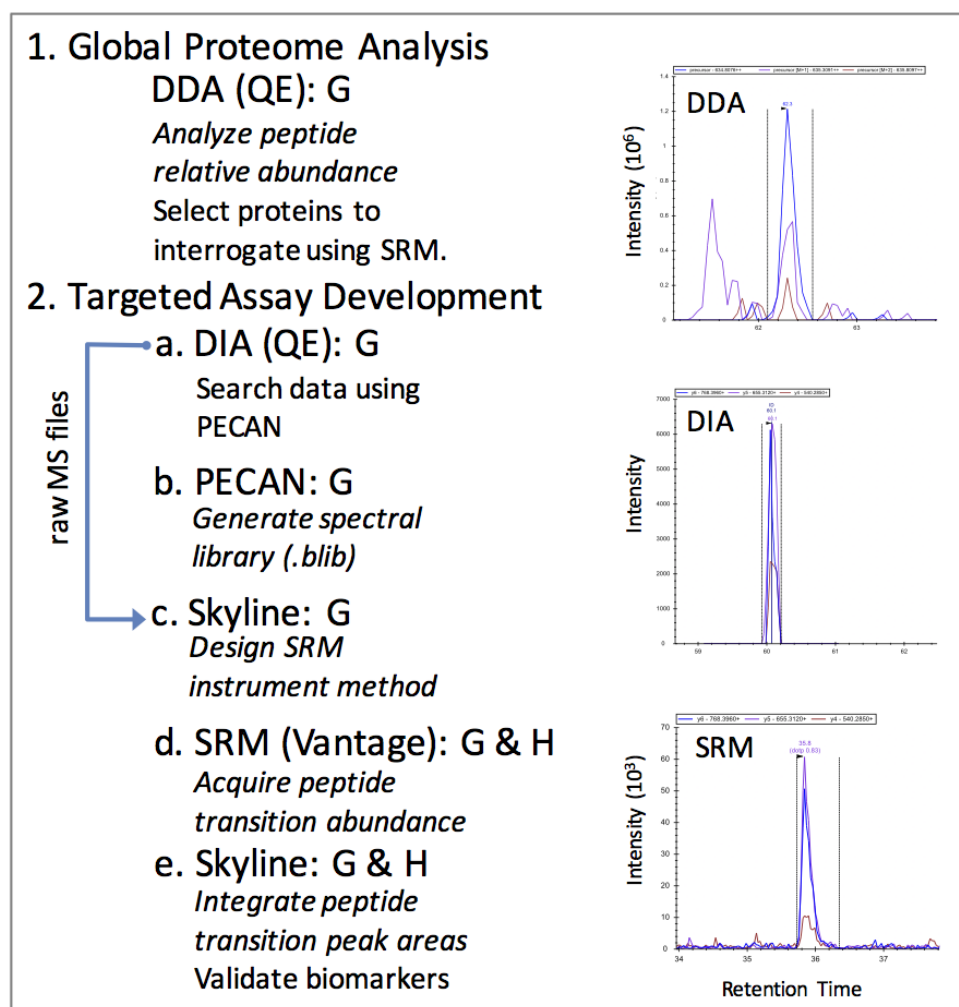
148   *Figure 1. Illustration of experimental setup and workflow for mass spectrometry data acquisition*

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

150   *on the Q-Exactive-HF (QE) to assess the global proteomic differences in gonad (G) tissue*

151   *between sexes and maturation stages. 2. Targeted assay development followed the steps of (a)*

152   *data independent acquisition (DIA) on gonad tissue was also completed on the QE to create*

153   *spectral libraries for selected reaction monitoring (SRM) method development; (b & c) spectral*

154   *libraries were analyzed in PECAN using Skyline to select optimal transitions and design an*

155   *instrument method for SRM analyses; (d) SRM was completed on the TSQ Vantage for geoduck*

156   *peptide transitions in gonad and hemolymph (G & H); (e) Peptide transition detection and*

157   *quantification was performed in Skyline. The chromatograms of peptide KEEELIDYMVKQ*

158   *(from protein 130261_c0_seq1|m.17926) were collected using the 3 different MS approaches*

159   *(DDA, DIA, and SRM) from the same late-stage female. Black vertical lines indicate peak*

160   *integration boundaries, and colored peaks represent the different transitions (i.e. peptide*

161   *fragments) collected.*

162

163   *DDA protein identification and quantification*

164   Gonad peptides were identified and proteins inferred using a proteome derived from a *de*

165   *novo* assembled transcriptome of a male and female geoduck clam gonad tissue libraries (NCBI

166   Bioproject Accession #PRJNA316216) (Del Rio-Portilla et al., unpublished). Briefly, reads were

167   assembled using Trinity [15] and deduced protein sequences determined using the Transdecoder

168   algorithm within Trinity. Raw mass spectrometry data (PRIDE Accession #PXD003127) was

169   searched against the protein sequences using Comet v 2016.01 rev.2.[15,16] Parameters used with

170   Comet included concatenated decoy search, specifying trypsin as the cleaving enzyme, two

171   missed cleavages allowed, peptide mass tolerance of 20 ppm, cysteine modification of 57 Da

172   (from the IAM) and methionine modification of 15.999 Da (from oxidation) to find peptide

173   spectral matches (all parameters can be found in Supporting information 2). Protein inference

174   and match probability were found using the Trans-Proteomic Pipeline.[16,17] Data files were

175   combined and normalized spectral abundance factor (NSAF) [18] was calculated in Abacus [19] to

176   determine consistent protein inferences across replicates. Protein identifications were considered

177   true matches when the probability of the match was at least 0.8611 (corresponding to error rate,

178   which approximates FDR, of 0.01) and at least 2 independent spectra were associated with the

179   protein across all samples.

180   Non-metric multidimensional scaling analysis (NMDS) was used to determine the

181   similarity of technical replicates (Supporting Information 3) using the vegan package [20] in R v.

182   3.2.3[21]. NMDS was performed on log-transformed data using a Bray-Curtis dissimilarity

183   matrix. As technical replicates clustered closely together and showed less variability than

184   biological replicates (Supporting Information 3), NSAF was averaged across each sample

185   (n=18). [18] Proteomic differences between sexes and maturation stages were explored with two

186   methods: 1) NMDS and analysis of similarity (ANOSIM) were used to compare the entire

187   proteomic profiles in multivariate space and 2) Fisher's exact test was used to determine

188   significant differences at the individual protein level. NMDS was performed on NSAF data

189   followed by ANOSIM in the vegan package in R to determine the differences between sexes and

190   maturation stages. Differentially abundant proteins among the six conditions were identified

191   using [22]Fisher's exact test in Excel. Spectral counts were summed across technical replicates to

192   compare biological replicates. Nine comparisons were analyzed for differentially abundant

193   proteins: Early-stage male (EM) vs. early-stage female (EF); mid-stage male (MM) vs. mid-stage

194   female (MF); late-stage male (LM) vs. late-stage female (LF); EM vs. MM; EM vs. LM; MM vs.

195   LM; EF vs. MF; EF vs. LF; MF vs. LF. All p-values resulting from the test were multiplied by 9

196   as a Bonferroni correction for multiple comparisons. Proteins with a corrected p-value ≤ 0.05

197   were considered differentially abundant.

198

199   *Enrichment Analysis*

200   We used enrichment analysis to determine biological processes (represented by detected

201   proteins) predominant in each gonad sample. GO enrichment analysis was performed for each

202   maturation stage (for each sex) on 1) unique proteins and 2) all detected proteins. In both cases,

203   the entire detected proteome was used as the background. Enrichment analysis was also

204   performed on sets of differentially abundant proteins. Code used to perform enrichment analysis

205   is available in a corresponding GitHub repository (https://github.com/yeastrc/compgo-geoduck-

206   public) as is the underlying code for the end user web-interface

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

208   Briefly, a p-value was calculated describing the enrichment of a GO term in a set of tested

209   proteins above what would be expected by chance, given the annotation of the geoduck gonad

210   proteome (3,627 proteins). A p-value cutoff of 0.01 was used to ascribe statistical significance to

211   GO terms representing biological process.

212   GO terms were first assigned directly to the geoduck gonad protein names resulting from

213   the *de novo*-assembled transcriptome. This was done by assigning GO terms associated with

214   Uniprot-KB Swiss-Prot BLASTp hits. A p-value representing the statistical significance of the

215   representation of a GO term in a set of proteins was calculated using the hypergeometric

216   distribution using the following formula:

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

217

218   Where A = total number of proteins submitted that have a GO annotation, B = the total number

219   of proteins in the background proteome annotated with the given GO term (or any of its

220   descendants), I (intersection of A and B) = the total number of submitted proteins annotated with

221   the given GO term (or any of its descendants), and T = the total number of annotated proteins in

222   the proteome background.

223   Then, the p-value describing the chance of having an intersection of size I or larger by

224   chance may be computed as:

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

226   The p-value is then corrected for multiple hypothesis testing using the Bonferroni method

227   by multiplying the p-value by the number of GO terms tested (setting resulting values over 1 to

228   1). The number of GO terms tested equals the number of GO terms found (and all ancestors) for

229   the submitted set of proteins.

230   Complete directed acyclic graphs (DAG) that represent a subset of the whole GO DAG

231   were generated to visualize enriched biological processes. These DAGs were then filtered by

232   removing all childless terms that had an associated p-value $\geqslant 0.01$. The resulting DAGs were

233   then filtered using the same method, and this process repeated until no childless terms remained

234   with a p-value above the cutoff. The final result is a filtered subset of the GO DAG that contains

235   no leaf nodes with a p-value greater than the cutoff, but where a given term is guaranteed to have

236   all of its ancestor terms, even if those terms have a p-value greater than the cutoff. Having the

237   ancestor terms present is critical to visualization as they provide context for interpreting the

238   results. Proteins that contribute to specific enriched processes for all enrichment analyses are

239   included in Supporting Information 4 and visualizations of enrichments for full and unique

240   proteomes for each sex and stage are visualized in a single DAG for each sex and stage

241   (Supporting Information 3). Venn diagrams of enriched processes were produced in Venny [22].

242

243   *Targeted Assay Development: Selected Reaction Monitoring (SRM)*

244   A subset of proteins was chosen for development of a suite of targeted assays using

245   selected reaction monitoring (SRM) on the mass spectrometer. Based on the data dependent

246   acquisition, proteins that were detected in only one gonadal stage (early-stage males (EM), early-

247   stage females (EF), late-stage males (LM), or late-stage females (LF)) or were at considerably

248   higher abundance in one of these stages were screened for usable peptide transitions in SRM in

249   Skyline daily v. 3.5.1.9706 [23].

250   Data independent acquisition (DIA) was used to generate spectral libraries for biomarker

251   development in the gonad tissue (Figure 1). Equal amounts of isolated peptides from the three

252   biological replicates for EM, EF, LM, and LF used in the DDA experiment (described above)

253   were pooled in equal quantities for DIA on the Q-Exactive HF (Thermo). Each sample included

254   a spiked-in internal quality control peptide standard (375 fmol PRTC; Pierce). Sample injections

255   for all DIA experiments included protein (1 µg) plus PRTC (50 fmol) in a 2 µl injection. An

256   analytical column (27cm) packed with C18 beads (3 µm; Dr. Maisch) and a trap (3 cm) with C12

257   beads (3 µm; Dr. Maisch) were used for chromatography. Technical replicate DIA spectra were

258   collected in 4 m/z isolation width windows spanning 125 m/z ranges each [24] (400-525, 525-650,

259   650-775, 775-900). For each method, a gradient of 5-80% ACN over 90 minutes was applied for

260   peptide spectra acquisition. Raw data can be accessed via ProteomeXchange under identifier

261   PXD004921. MSConvert [25] was used to generate mzML files from the raw DIA files.

262   Peptide Centric Analysis was completed with the software program PECAN to generate

263   spectral libraries for targeted method development. [26] Input files included the list of peptides

264   generated for SRM (n=217), as described above, and the mzML files generated from the raw

265   DIA files. PECAN correlates a list of peptide sequences with the acquired DIA spectra to locate

266   the peptide-specific spectra within the acquired DIA dataset. A background proteome of the *in*

267   *silico* digested geoduck gonad proteome was used.

268   The PECAN output file (.blib) was imported into Skyline to select peptide transitions and

269   create MS methods that would target specific peptides and transitions. Peptide transitions are the

270   reproducible fragments of peptides that are generated during the MS2 scan. Peptides reliably

271   fragment in the same way in the mass spectrometer, therefore transitions are a robust and

272   consistent signal of a peptide's presence.[27] Peptide transitions were selected if peak morphology

273   was uniform and consistent across the MS2 scans for technical replicates. Peptides were selected

274   for targeted analysis if they had $\geq$ 3 good quality transitions and there were $\geq$ 2 peptides per

275   protein. A maximum of 4 transitions per peptide were selected for targeted analysis and no more

276   than 3 peptides per protein were selected. This transition list was divided between two method

277   files for the final SRM analyses to provide adequate dwell time on individual transitions to

278   accurately detect and measure all peptides desired.[28]

279   Selected reaction monitoring (SRM) was carried out on a Thermo Vantage for all

280   eighteen geoduck gonad samples used in the original DDA analysis. Samples were prepared as

281   described above for DIA (1 µg of protein per 3 µl injection). A new C18 trap (2 cm) and C18

282   analytical column (27.5 cm) were used and each sample was analyzed in triplicate across two

283   MS experiments to cover the entire peptide transition list (n=228, 217 of which yielded

284   quantifiable data). Raw data can be accessed in the PeptideAtlas under accession PASS00943.

285   Hemolymph from early-, mid-, and late-stage males and females was also assayed for

286   sex- and stage-specific biomarkers. An augmented SRM assay was used on the hemolymph that

287   included all the peptide transitions analyzed in the gonad with an additional 40 transitions from

288   five new proteins. These proteins were selected based on gonad proteome annotation to include

289   1) proteins that would likely be circulating in the hemolymph and 2) gonad proteins that had

290   homology with the mussel hemolymph proteome.[29] The additional peptides for SRM analysis of

291   hemolymph were from vitellogenin (2 proteins: cds.comp100108_c2_seq1|m.5995,

292   cds.comp144315_c0_seq1|m.50928), glycogen synthase (2 proteins:

293   cds.comp140343_c0_seq2|m.34696, cds.comp141785_c0_seq1|39776), and glycogenin-1

294   (cds.comp140645_c0_seq3|m.35608). PECAN generated spectral libraries, as described above,

295   and Skyline software selected hemolymph protein transitions to target during SRM analyses. For

296   the new hemolymph proteins, the minimum targets for peptides ($\geq$3) and transitions ($\geq$2) as

297   described for gonad SRM could not be met for every protein, but were still included in the

298   analysis. These transitions, and the previous gonad transitions (total of n = 254), were analyzed

299   across sixteen geoduck hemolymph samples in two technical replicates, as described above for

300   the gonad. Some of these peptides yielded no data in the hemolymph, resulting in a dataset of

301   171 peptide transitions that were reliably detected. Raw data can be accessed in PeptideAtlas

302   under accession PASS00942.

303   Acquired SRM data in gonad and hemolymph were analyzed in Skyline for peptide

304   transition quantification. Skyline documents that were used to analyze the monitored peptide

305   transitions can be found on Panorama for the gonad dataset and the hemolymph datasets

306   (panoramaweb.org/labkey/geoduckrepro.url). Accurate Peak detection was determined based on

307   consistency of retention time (verified by spiked in PRTC peptides and correlation with DIA

308   peptide retention times) and peak morphology.

309   All peptide transition peak intensities were exported from Skyline for automated peak

310   selection and peak integration analysis. PRTC internal standard transitions were monitored for

311   consistency across runs by calculating the coefficient of variation (CV) of transition peak area

312   across injections. Peak intensities for the geoduck peptide transitions were normalized by

313   dividing by the averaged intensities for the 6 PRTC peptide transitions that had the lowest CV in

314   the gonad data (CV < 13) and the 6 with the lowest CV in the hemolymph data (CV < 10).

315   NMDS and ANOSIM were performed on the PRTC-normalized SRM dataset as

316   described above for the DDA dataset, except data were log(x+1) transformed for the gonad SRM

317   dataset. An initial NMDS showed that technical replicates clustered together well and that

318   variation was lower within biological replicates compared to between replicates, therefore

319   technical replicate peak intensities were averaged for the rest of the analysis (Supporting

320   Information 3). ANOSIM was performed using grouping by sex and reproductive stage alone, as

321   well as by a combined sex-stage factor. Coefficients of variation were calculated for combined

322   technical replicates using the raster [30] package in R.

323   Eigenvector loadings were calculated for the gonad and hemolymph data using the vegan

324   package in R. For each dataset, the top 20 transitions with the combination of lowest p-value and

325   highest loading value for each MDS axis were selected as biomarkers. Heatmaps of the log-

326   normalized transition intensities for these biomarkers were made using pheatmap[31] in R,

327   clustering rows (peptide transitions) and columns (samples) using euclidean distance and the

328   average clustering method.

329

330   **Results**

331   *DDA Proteomics*

332   The mass spectrometry data (PRIDE Accession #PXD003127) interpreted with a species-

333   and tissue-specific transcriptome yielded 3,651 proteins inferred with high confidence across all

334   gonad tissue samples (total spectral count across all replicates > 1) (Supporting Information

335   4). Female and male gonad proteomic profiles were more similar in early-stage maturation, with

336   proteomes diverging as reproductive maturity advanced (Figure 2). Proteomic profiles were

337   significantly different between sexes (R = 0.4167, p = 0.001) and maturation stages (R = 0.3494,

338   p = 0.002).

339   Unique proteomic profiles include the set of proteins that were detected in a specified sex

340   or specific stage within sex in the DDA analysis. The female proteomic profile across all stages

341   contained 156 unique proteins whereas 144 proteins were unique to males. The number of

342   proteins unique to a specific maturation stage increased with maturity from 16 and 18 in early-

343   stage females and males, respectively, to 132 and 104 in late-stage females and males (Figure 3).

344   Enriched GO biological processes for the unique proteins and all detected proteins for each sex

345   and stage are summarized in Table 1. Only the most specific GO terms (i.e., farthest down on the

346   DAG) are listed; parent terms were frequently enriched as well and can be seen in the DAGs

347   (Supporting Information 3).

348

349

350

351   *Table 1. Enriched biological process GO terms for the unique proteins and all detected proteins*

352   *in the gonad for each sex and stage. If no GO terms are listed, no enriched processes were*

353   *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

354

355
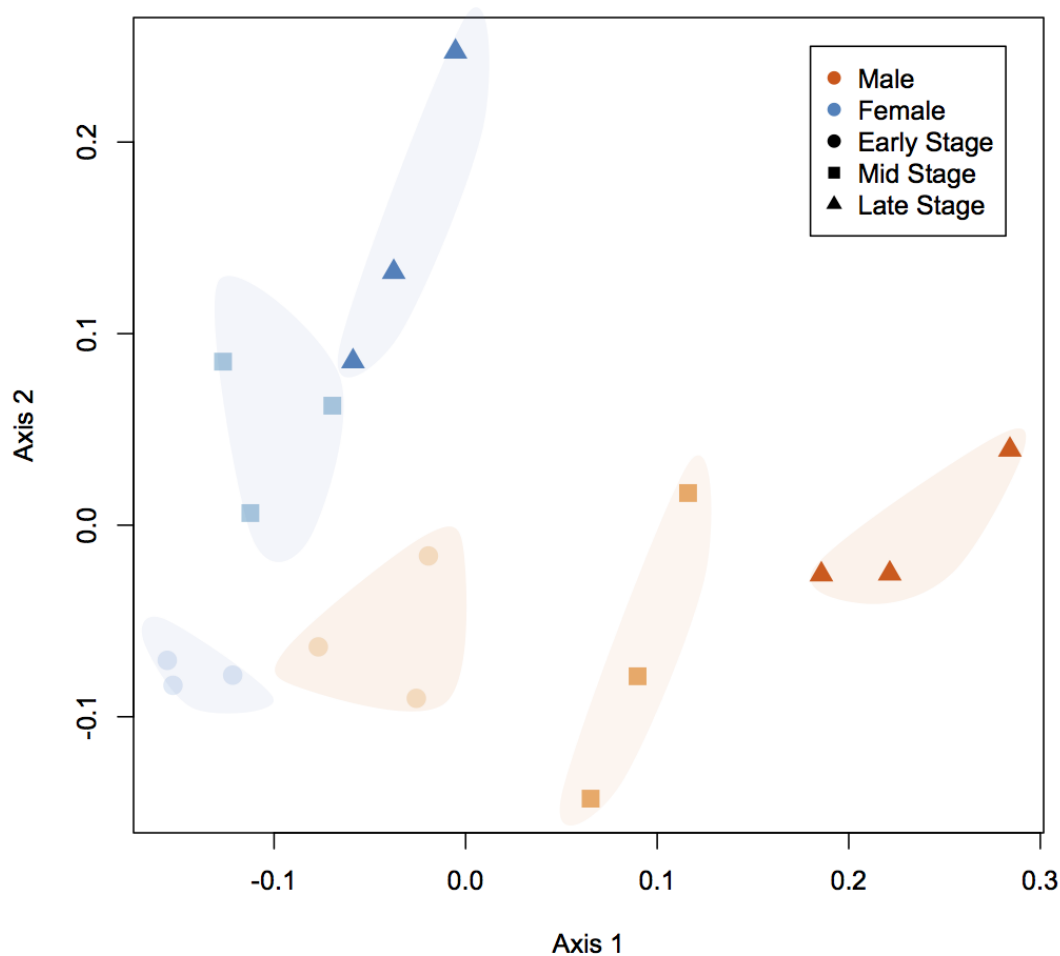
356   *Figure 2. Non-metric multidimensional scaling plot (NMDS) of geoduck gonad whole proteomic*

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

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

359   *stage = triangles; p<0.05).*

360

361   Numerous proteins were differentially abundant in comparisons between reproductive

362   stage and sex based on analysis with a Fisher's exact test with Bonferroni correction for multiple

363   comparisons. The number of differentially abundant proteins within a stage and between sexes

364   increased from 102 between early-stage males and females to 1109 between late-stage males and

365   females (Figure 4, Supporting Information 4). Enrichment analysis revealed which processes

366   were over-represented in the sets of differentially abundant proteins, when compared to the entire

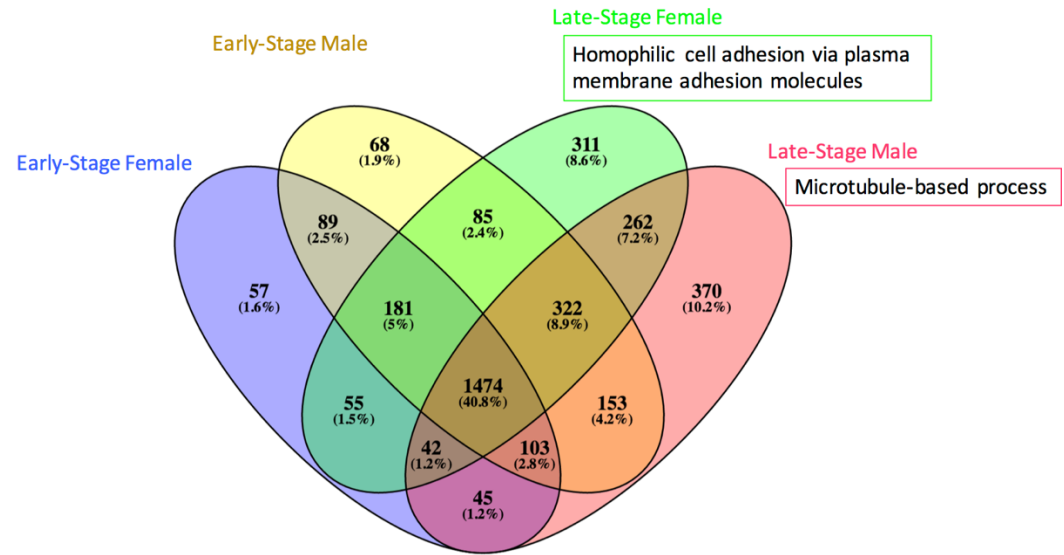367   proteome (Figure 4 and Supporting Information 4).

368

369

370   *Figure 3. Number of proteins detected in sex and stage specific proteomes for early-stage female*

371   *(blue), early-stage male (yellow), late-stage female (green), and late-stage male (red). Biological*

372   *processes enriched in the sex- and stage-specific proteomes compared to all proteins detected*

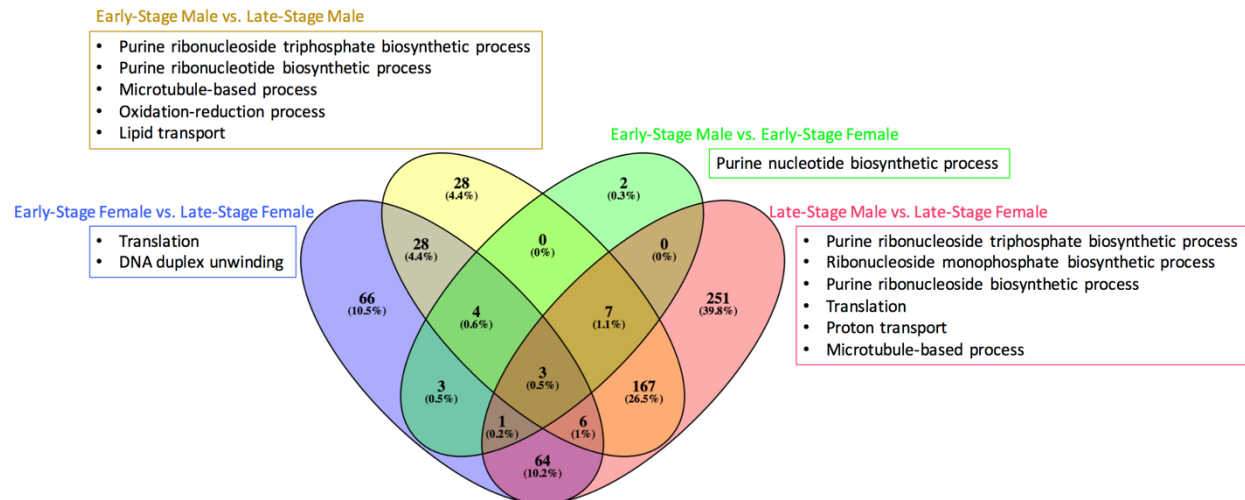373   *across sexes and stages are listed for each proteome.*

374

Figure 4. Number of differentially abundant proteins between geoduck sexes and stages based on comparisons between early-stage males and mid-stage males; early-stage males and late-stage males; early-stage females and mid-stage females; early-stage females and late-stage females. Biological processes enriched in the sets of differentially abundant proteins, compared to all proteins detected across sexes and stages, are listed for each proteome.

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=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 more 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).

392   In an ordination plot, the gonad SRM data resolved males and females better in late-stage

393   samples than the early- or mid-stage (Supporting Information 3). There was significant

394   separation based on gonad SRM data for sex (R = 0.2435, p=0.002), stage (R=0.1984, p=0.003),

395   and a combined sex-stage factor (R=0.3959, p=0.001). The hemolymph SRM data resolved the

396   late-stage female and mid-stage male groups from the rest of the geoduck (Supporting

397   Information 3). There was significant separation of the hemolymph proteomic profiles based on

398   stage (R=0.1772, p=0.048) and sex-stage (R=0.2749, p=0.028), but not by sex alone

399   (R=0.01786, p=0.37).

400   Peptide transitions that directed the observed spread of the geoduck sex-stage data along

401   Axis 1 and Axis 2 in the NMDS plots (determined by eigenvector analysis) were considered key

402   analytes for differentiating groups. These peptide transitions represent a starting point for

403   biomarker assay optimization. Peptides that drove the separation of the male maturation stages

404   along Axis 1 in the gonad NMDS plot (Supporting Information 4) are from a set of 7 proteins: IQ

405   domain-containing protein K (cds.comp133232_c0_seq1|m.21239), armadillo repeat-containing

406   protein (cds.comp140515_c2_deq1|m.35054), flap endonuclease

407   (cds.comp130260_c0_seq1|m.17926), dynein heavy chain (cds.comp144210_c0_seq2|m.50335),

408   spectrin alpha chain (cds.comp141647_c0_seq1|m.39026), and four uncharacterized proteins

409   (cds.comp135264_c2_seq1|m.24100, cds.comp133422_c0_seq1|m.21474,

410   cds.comp131654_c1_seq1|m.19394, cds.comp144596_c0_seq1|m.53297) (Figure 5). Proteins

411   that played a significant role in the separation of the female stages along Axis 2 for the gonad

412   data included centrosomal protein of 70 kDa (cds.comp133063_c0_seq1|m.21029), WD repeat-

413   containing protein on Y chromosome (cds.comp131715_c0_seq1|m.19472), tetratricopeptide

414   repeat protein 18 (cds.comp139151_c0_seq5|m.31757), and four uncharacterized proteins

415   (cds.comp131654_c1_seq1|m.19394, cds.comp139654_c0_seq1|m.32810,

416   cds.comp135264_c2_seq1|m.24100, cds.comp140866_c0_seq1|m.36276). Analysis of the

417   hemolymph SRM NMDS (Supporting Information 3) reveals that the main peptide drivers along

418   Axis 1 (separating late-stage females and mid-stage males) were from six proteins: flap

419   endonuclease (cds.comp130260_c0_seq1|m.17926), vitellogenin

420   (cds.comp144315_c0_seq1|m.50928), spectrin alpha chain (cds.comp141647_c0_seq1|m.39026),

421   vesicle-fusing ATPase (cds.comp144180_c0_seq1|m.39026), and two uncharacterized proteins

422   (cds.comp138018_c0_seq1|m.29238, cds.comp133422_c0_seq1|m.21474) (Figure 6). A cluster

423   analysis of the 40 transitions that were primary drivers of individual variation in the NMDS

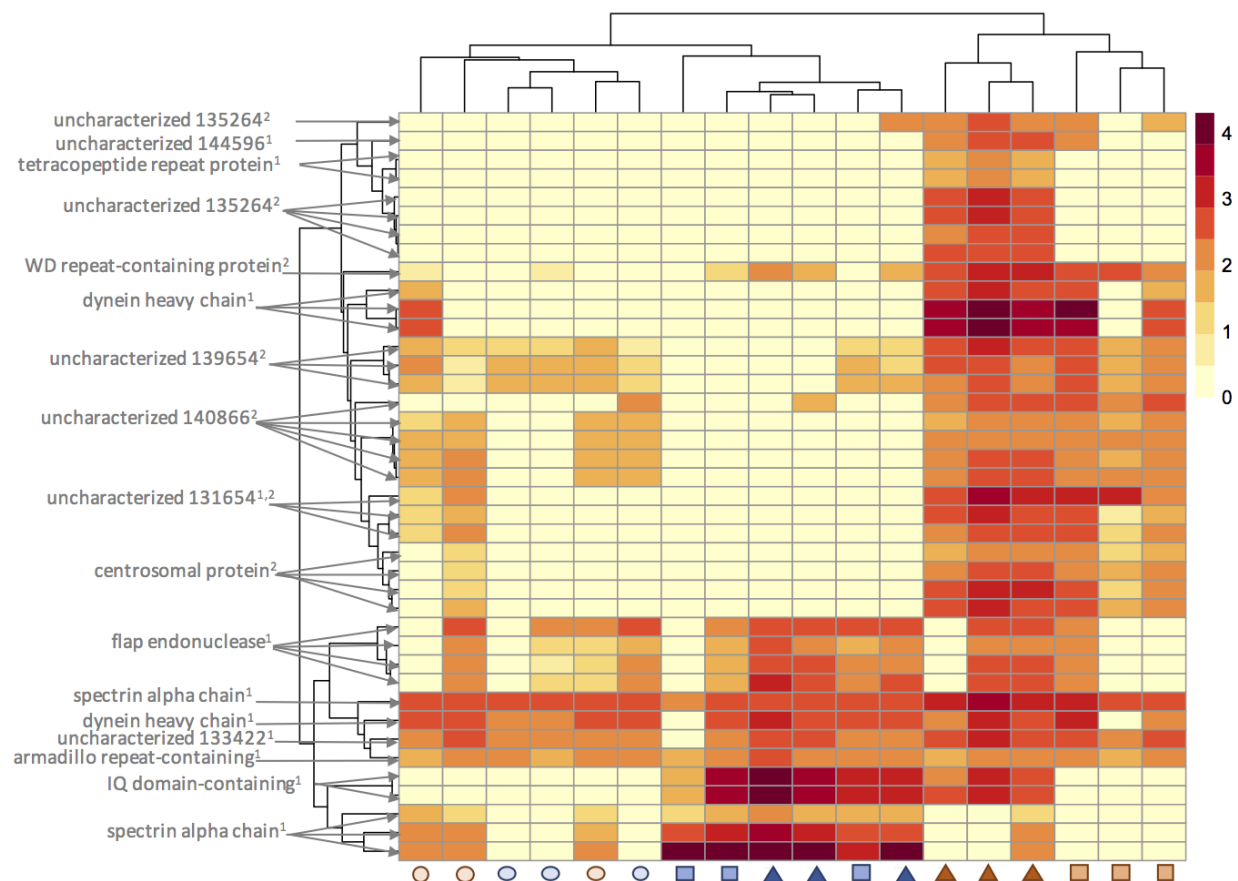424   exposed a strong cluster of the females in the hemolymph data (Figure 6).

425

426

427   *Figure 5. Heatmap of log-normalized peptide transition intensities for 40 significant transitions*

428   *(based on NMDS analysis) across all samples (early-stage/circles, mid-stage/squares, and late-*

429   *stage/triangles males/orange and females/blue) gonad. Each row (peptide transition) is labeled*

430   *by its corresponding protein annotation in gray. Superscripts indicate for which NMDS axis a*

431   *transition is significant.*

432

434   *Figure 6. Heatmap of log-normalized peptide transition intensities for 40 significant transitions*

435   *(based on NMDS analysis) across all samples (early-stage/circles, mid-stage/squares, and late-*

436   *stage/triangles males/orange and females/blue) hemolymph. Each row (peptide transition) is*

437   *labeled by its corresponding protein annotation in gray. Superscripts indicate for which NMDS*
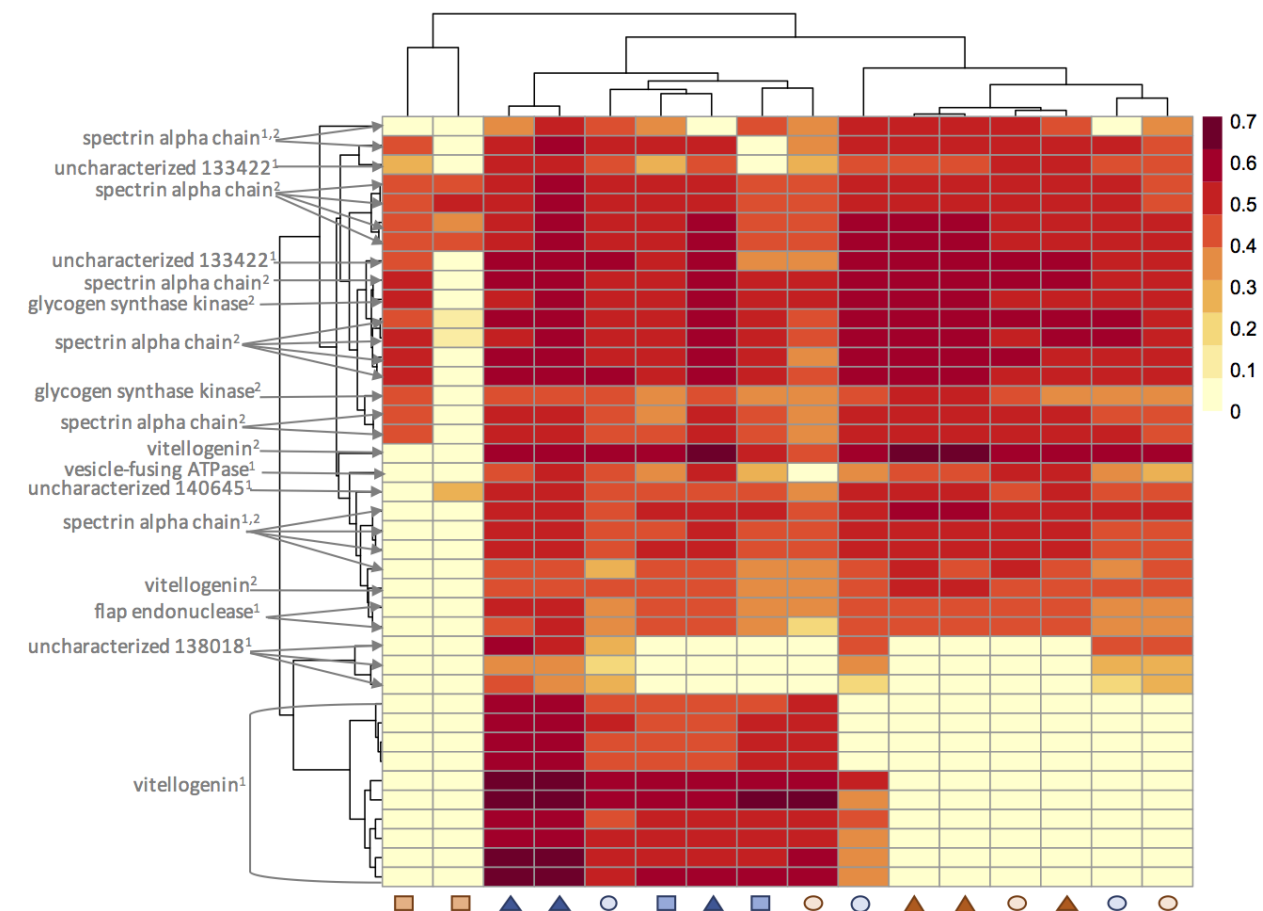
438   *axis a transition is significant.*

439

Discussion

441   Proteomics has the capacity to uncover physiological processes underlying functional

442   phenotypic change, including the maturation of reproductive tissue. We have characterized the

443   geoduck clam gonad proteome throughout reproductive maturation for both males and females.

444   Our data dependent analysis (DDA) approach yielded 3,651 detected proteins across both sexes

445   and three maturation stages. This is a significant escalation in the understanding of proteomic

446   responses in maturation stages of marine mollusks. Based on the DDA data, 27 proteins (with a

447   corresponding 85 peptides) from early- and late-stage male and female clams were chosen for

448   selected reaction monitoring (SRM). Peptide transitions were detected from the selected peptides

449   as candidates for biomarker development in gonad (n=217) and in hemolymph (n=171).

450

451   *DDA*

452   There are clear proteomic profile differences in the geoduck gonad by both sex and

453   maturation stage and these differences reveal the biochemical pathways underlying tissue

454   specialization. More proteins were differentially abundant between early and late stages of

455   reproductive maturity, compared to early and mid-stages, for both males and females, reflecting

456   that, as maturation progresses, gonad tissue becomes more specialized, expressing a more

457   diverse array of regulatory and structural proteins. Gonad goes from undifferentiated connective

458   tissue to specialized structures that regulate gametogenesis and release of gametes. These

459   phenotypic changes required for reproductive maturity are evident with histology[13] and are

460   realized via changes at the protein level. For example, the zona pellucida (ZP) containing protein

461   (cds.comp134923_c0_seq3|m.23445) and vitellogenin (cds.comp144315_c0_seq1|m.50928)

462   were more abundant in the female gonad as geoduck matured. The zona pellucida surrounds

463   vertebrate oocytes and its invertebrate homolog is the vitelline envelope; the vitelline envelope is

464   likely the origin of the zona pellucida (ZP) containing protein and the site of sperm-oocyte

465   recognition and binding. Vitellogenin is an egg yolk protein precursor that is closely linked to

466   gametogenesis in bivalves. Increasing levels of vitellogenin are correlated with advancement in

467   reproductive maturation in oysters [32,33,34], clams [35], and mussels [36].

468   In females, enrichment analysis of unique proteins and all detected proteins revealed that

469   the early-stage gonad proteome was enriched in proteins involved in translation and cytoskeleton

470   structure and proteins associated with purine metabolism and translation were predominant

471   during the mid-stage. These protein-level changes likely represent alterations to cell morphology

472   and shifts in protein translation required to effect gonad structural and functional changes.

473   The late-stage female gonad proteome was enriched in proteins associated with cell

474   adhesion and translation/phosphorylation. Cell adhesion is an essential biological process in

475   maturing oocytes in vertebrates.[37,38] The strong cell adhesion signal in late-stage females was

476   driven in large part by the many cadherin and protocadherin proteins detected in a single female.

477   However, twelve protocadherins were detected at significantly higher abundance in more than a

478   single late-stage female when compared to early- and mid-stages. Further, the cell adhesion

479   proteins laminin, talin, and vinculin were less abundant in late-stage females compared to other

480   females. The shift from laminin, talin, and vinculin to cadherins during female gonad

481   development may signal a change in cell-cell interactions as oocytes form. There was also storng

482   Evidence for the importance of protein translation and phosphorylation. Proteins with homology

483   to several protein kinases were all detected solely in late-stage females. This surge in

484   phosphorylating proteins is likely a reflection of an increase in protein activation in the later

485   stages of gonad maturation.

486   Enriched processes in late-stage male geoduck gonad proteome including "kinesin

487   complex" and many microtubule-related functions reflect an increase in cell replication during

488   the last steps of sperm formation. Several proteins implicated in mitosis and meiosis were

489   detected at significantly increased abundance in late-stage males compared to early- and mid-

490   stage: dynein, intraflagellar transport protein, kinesin, and alpha and beta tubulin. All of these

491   proteins are instrumental in the cell division and replication and their importance at this

492   maturation stage suggests the presence of rapidly dividing cells. In mussels off the Atlantic coast

493   of Spain, meiotic division begins approximately four months before the onset of spawning, while

494   mitosis reconstructs the gonad during and after spawning.[39] These pieces of proteomic evidence

495   point towards increased cellular energy metabolism and cell division during the development of

496   mature, motile sperm.

497   Even with a highly specific protein identification database, the DDA method excludes

498   many important, relatively low abundance proteins from analysis.[40] The ten most abundant

499   proteins (as measured by total spectral count) for each sex-maturation stage accounted for 6-28%

500   of the total spectral counts across all proteins for a given sex-stage. Many of these proteins are

501   "housekeeping" proteins, such as actin and myosin, and their dominance in the peptide mixture

502   was likely masking many informative, low abundance peptides. Similarly, in zebrafish, highly

503   abundant vitellogenin (also highly abundant in the mid- and late-stage females in this study)

504   made it difficult to detect lower abundance proteins using DDA.[41]

505   The most abundant protein across all replicates likely plays an essential role in

506   reproductive maturation. The protein is glutamine gamma-glutamyltransferase, or

507   transglutaminase, and it has total spectral counts ranging from 259 to 1878 per MS analysis. The

508   annotation to this protein was only recently updated in Uniprot, so it was originally annotated as

509   "uncharacterized" in our BLAST search. Transglutaminase is typically associated with protein

510   cross-linking during wound healing in crustaceans[42] and bivalves[43] and also serves as a

511   biomarker for summer mortality in bivalves.[44] However, gene expression of this protein

512   transcript has also been linked to vitellogenesis in the shrimp *Panaeus monodon* where it is

513   hypothesized to mediate ovarian maturation via a hormone-receptor interaction[45]. The high

514   abundance of transglutaminase in our geoduck dataset suggests a key role in invertebrate

515   reproductive maturation. It is significantly more abundant in mid-stage males and females

516   compared to late- and early-stage stages as well as more abundant in late-stage females than

517   males, which is consistent with the shrimp study where its gene transcript was at highest

518   abundance in early vitellogenesis[45].

519

520   *SRM*

521   Selected reaction monitoring allows for the detection and absolute quantification of only

522   the most informative proteins' peptides to address a specific hypothesis. We leveraged the DDA

523   dataset to create informative SRM assays of geoduck sex and maturation stage in both gonad and

524   hemolymph. In the gonad tissue, we measured 217 peptide transitions that could differentiate

525   males and females as well as early and late maturation stages. Our assay is based on proteins that

526   are involved in calcium ion binding, phosphorylation, meiosis, DNA replication and repair, and

527   membrane structure. Additionally, seven of these informative proteins were unannotated. These

528   unknown proteins may represent taxon-specific reproductive proteins and warrant further

529   characterization. Reproductive proteins are some of the fastest-evolving, resulting in many

530   species-specific proteins.[46]

531   A subset of the peptide transitions (n=40) from these proteins were deemed highly

532   informative based on our analysis in gonad samples and would be good candidates for a more

533   streamlined SRM assay. These transitions correspond to eight annotated proteins and six

534   unannotated proteins. The annotated proteins are implicated in cytoskeleton structure (spectrin

535   alpha chain, dynein heavy chain), DNA replication and repair (centrosomal protein of 70 kDa,

536   flap endonuclease), and proteins that were identified by motifs and thus yield little functional

537   information (WD repeat-containing protein on Y chromosome, IQ domain containing protein K,

538   armadillo repeat-containing protein, and tetracopeptide repeat protein 18). The 40 transitions

539   were almost uniformly at higher abundance in late stage males and females, compared to earlier

540   reproductive stages.

541   The peptide-based assay developed from the evidence found in the gonad tissue was

542   applied to circulating hemolymph. We were able to collect data on 171 peptide transitions that

543   were derived from gonad tissue but detectable in the hemolymph, including all the transitions

544   that were added specifically to the hemolymph assay. Peptide transitions from 16 proteins were

545   reliably detected in both gonad and hemolymph, however, only three proteins had transitions that

546   were highly informative in both tissues: spectrin alpha chain, flap endonuclease, and an

547   uncharacterized protein. The inclusion of vitellogenin in our assay confirmed its presence in

548   geoduck hemolymph during later stages of reproductive maturity in females. Vitellogenin is an

549   informative proteomic biomarker for oyster and sea turtle female reproductive stage. [34, 47] In the

550   oyster study, vitellogenin fragments were among the most abundant proteins detected and were

551   particularly accumulated at greater abundances in oocytes associated with low production of

552   larvae.[34] In this case, it was hypothesized that greater accumulation of vitellogenin fragments in

553   the poor versus high quality oyster oocytes was indicative of the enzymatic breakdown of the

554   protein during oocyte ageing.[34]

555   The full suite of peptide transitions was not detected in hemolymph, likely due to

556   physiological differences between the tissues in which the assay was designed (gonad) and

557   applied (hemolymph). Biomarkers of tissue-specific changes can be difficult to detect in

558   circulating fluid since hemolymph is not the primary site of physiological change. Proteins and

559   their peptide fragments can yield physiologically relevant assays, even for organ-specific

560   changes [48], but may not be detectable due to different fragmentation patterns or delay in

561   abundance changes between the site of physiological change and the hemolymph or blood. In

562   geoduck hemolymph, 98 transitions were detected in the gonad but not in the hemolymph, but

563   119 transitions were detected in both tissues. Flap endonuclease and an uncharacterized protein,

564   detected in both tissues, were informative for late-stage female identification in the hemolymph

565   assay with flap endonuclease also an informative biomarker in gonad (Figures 5 & 6). The

566   hemolymph results emphasize both the flexibility and limitations of a peptide-based SRM assay

567   when designed for measurement of physiological change in a specific organ or tissue.

568   Through characterization of the geoduck gonad proteome and development of targeted

569   peptide-based assays for reproductive maturity, we not only exponentially increased genomic

570   resources available for this species, but we also provide an effective approach for non-lethal

571   detection of sex and maturation stage in this species. The gonad peptide transitions could be

572   optimized to a set of 40 transitions that would accurately predict a) geoduck sex and b)

573   reproductive maturation status. In the hemolymph, 40 peptide transitions can accurately

574   differentiate late-stage female and mid-stage male geoduck from others. Together these

575   molecular tools can directly address the geoduck production problem of asynchronous spawning

576   due to different maturation stages and unknown sexes.

577

Acknowledgements

578

579   This work was supported and funded by a grant from the University of Washington

580   Royalty Research Fund (to SR), a Training Grant from the National Institutes of Health (to ETS)

581   (T32 HG00035), and the University of Washington's Proteomics Resource (UWPR95794). We

582   would like to thank Jimmy Eng and Priska von Haller for their help in data acquisition and

583   analysis; Sonia Ting, Jarrett Egertson, Lindsay Pino, Yuval Boss, Brian Searle, and Nick

584   Shulman for assistance with data analysis; the Genome Sciences Information and Technology

585   group for their technical support; and Michael MacCoss for providing support and space to

586   complete the work. ETS and BLN would like to thank IJE and TAN for their ongoing

587   inspiration.

588

Supporting Information.

589

590   **S-1** Cover page for Supporting Information.

591   **S-2** Parameter file used in Comet searches for the DDA data

592   **S-3** Plots demonstrating technical replication for DDA and SRM, and DAGs for enrichment

593   analysis.

594   **S-4** The first tab of the workbook contains all identified proteins from the DDA experiment with

595   Uniprot annotations (e-value cut-off of 1E-10), total spectral counts for each technical replicate,

596   calculated normalized spectral abundance factor (NSAF) for combined technical replicates, and

597   indication of significantly differentially abundant proteins by pairwise comparison. Columns

598   containing spectral count data have headers "SpC" preceded by the biological replicate number,

599   sex, maturation stage, and technical replicate (for example, "EF3.1 SpC" is technical replicate 1

600   from early-stage female 3). Notation for NSAF is similar to notation for SpC. The 9 columns in

601   the sheet have headers such as "EFvLF" (comparison between early- and late-stage females)

602   have asterisks in the cells that correspond to proteins that were differentially abundant for each

603   given comparison. The last two columns ("NMDS Gonad" and "NMDS Hemolymph") have

604   asterisks in cells that correspond to the proteins with peptide transitions that contribute

605   significantly to the SRM NMDS plot distributions. The second tab contains GO biological

606   processes enriched in proteins that were differentially abundant between stages within a sex (e.g.

607   early- vs. mid-stage female) or between sexes within a stage (e.g LF vs. LM). Only the most

608   specific GO terms in the GO hierarchy are listed. No terms are listed when a comparison was not

609   made. The third through sixth tabs contain raw Skyline output (in the tab "Skyline output") and

610   peak intensities normalized by PRTC peptide abundance ("Normalized Intensities") for the

611   gonad and hemolymph SRM data.

612

613   **References**

614   (1) Newell, R. I. Ecosystem influences of natural and cultivated populations of suspension-

615   feeding bivalve mollusks: A review. *J. Shellfish Res.* **2004**, *23* (1), 51–62.

616   (2) Hoffmann, A.; Bradbury, A.; Goodwin, C. L. Modeling geoduck, Panopea abrupta (Conrad,

617   1849) population dynamics. I. Growth. *J. Shellfish Res.* **2000**, *19* (1), 57–62.

618   (3) Northern Economics, I. *The Economic Impact of Shellfish Aquaculture in Washington,*

619   *Oregon and California. Prepared for Pacific Shellfish Institute*; 2013.

620   (4) Scott, A. P. Do mollusks use vertebrate sex steroids as reproductive hormones? II. Critical

621   review of the evidence that steroids have biological effects. *Steroids* **2013**, *78* (2), 268–281.

622   (5) Fernandes, D.; Denise, F.; Barbara, L.; Cinta, P. Biosynthesis and metabolism of steroids in

623   molluscs. *J. Steroid Biochem. Mol. Biol.* **2011**, *127* (3-5), 189–195.

624   (6) Palacios, E.; Racotta, I. S.; Arjona, O.; Marty, Y.; Le Coz, J. R.; Moal, J.; Samain, J. F.

625   Lipid composition of the pacific lion-paw scallop, Nodipecten subnodosus, in relation to

626   gametogenesis. *Aquaculture* **2007**, *266* (1-4), 266–273.

627   (7) Anantharaman, S.; Craft, J. A. Annual variation in the levels of transcripts of sex-specific

628   genes in the mantle of the common mussel, Mytilus edulis. *PLoS One* **2012**, *7* (11), e50861.

629   (8) Ciocan, C. M.; Cubero-Leon, E.; Minier, C.; Rotchell, J. M. Identification of reproduction-

630   specific genes associated with maturation and estrogen exposure in a marine bivalve Mytilus

631   edulis. *PLoS One* **2011**, *6* (7), e22326.

632   (9) Li, Q.; Qi, L.; Makoto, O.; Katsuyoshi, M. Seasonal biochemical variations in Pacific oyster

633   gonadal tissue during sexual maturation. *Fish. Sci.* **2000**, *66* (3), 502–508.

634   (10)   Dheilly, N. M.; Lelong, C.; Huvet, A.; Kellner, K.; Dubos, M.-P.; Riviere, G.; Boudry,

635   P.; Favrel, P. Gametogenesis in the Pacific oyster Crassostrea gigas: a microarrays-based

636   analysis identifies sex and stage specific genes. *PLoS One* **2012**, *7* (5), e36353.

637   (11)   Arcos, F. G.; Ibarra, A. M.; del Carmen Rodríguez-Jaramillo, M.; García-Latorre, E. A.;

638   Celia, V.-B. Quantification of vitellin/vitellogenin-like proteins in the oysterCrassostrea

639   corteziensis(Hertlein 1951) as a tool to predict the degree of gonad maturity. *Aquac. Res.*

640   **2009**, *40* (6), 644–655.

641   (12)   Li, Y.; Siddiqui, G.; Wikfors, G. H. Non-lethal determination of sex and reproductive

642   condition of Eastern oysters Crassostrea virginica Gmelin using protein profiles of

643   hemolymph by Proteinchip® and SELDI-TOF-MS technology. *Aquaculture* **2010**, *309* (1–

644   4), 258–264.

645   (13)   Crandall, G.; Roberts, S. Reproductive Maturation in Geoduck clams (Panopea generosa)

646   https://figshare.com/articles/Reproductive_Maturation_in_Geoduck_clams_Panopea_genero

647   sa_/3205975/1.

648   (14)   Timmins-Schiffman, E.; Nunn, B. L.; Goodlett, D. R.; Roberts, S. B. Shotgun proteomics

649   as a viable approach for biological discovery in the Pacific oyster. *Conserv Physiol* **2013**, *1*

650   (1), cot009.

651   (15)   Grabherr, M. G.; Haas, B. J.; Yassour, M.; Levin, J. Z.; Thompson, D. A.; Amit, I.;

652   Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.; et al. *Full-length transcriptome*

653   *assembly from RNA-seq data without a reference genome*; 2011.

654   (16)   Pedrioli, P. G. A. Trans-Proteomic Pipeline: A Pipeline for Proteomic Analysis. In

655   *Methods in Molecular Biology*; 2009; pp 213–238.

656   (17)   Deutsch, E. W.; Mendoza, L.; Shteynberg, D.; Slagel, J.; Sun, Z.; Moritz, R. L. Trans-

657   Proteomic Pipeline, a standardized data processing pipeline for large-scale reproducible

658   proteomics informatics. *Proteomics Clin. Appl.* **2015**, *9* (7-8), 745–754.

659   (18)   Florens, L.; Carozza, M. J.; Swanson, S. K.; Fournier, M.; Coleman, M. K.; Workman, J.

660   L.; Washburn, M. P. Analyzing chromatin remodeling complexes using shotgun proteomics

661   and normalized spectral abundance factors. *Methods* **2006**, *40* (4), 303–311.

662   (19)   Damien Fermin, Vekatesha Basrur, Anastasia K. Yocum, and Alexey I. Nesvizhskii.

663   Abacus: A computational tool for extracting and pre-processing spectral count data for

664   label-free quantitative proteomic analysis. *Proteomics* **2011**, *11* (7), 1340–1345.

665   (20)   Oksanen, J.; Guillaume Blanchet, F.; Friendly, M.; Kindt, R.; Legendre, P.; McGlinn, D.;

666   Minchin, P. R.; O'Hara, R. B.; Simpson, G. L.; Solymos, P.; et al. *vegan: Community*

667   *Ecology Package*; 2016.

668   (21)   R Core Team. *R: A language and environment for statistical computing*; R Foundation

669   for Statistical Computing, Vienna, Austria, 2015.

670 (22)   Choi, H.; Fermin, D.; Nesvizhskii, A. I. Significance analysis of spectral count data in

671   label-free shotgun proteomics. *Mol. Cell. Proteomics* **2008**, *7* (12), 2373–2385.

672 (22)   Oliveros, J. C. *Venny. An interactive tool for comparing lists with Venn's diagrams*;

673   2007-2015.

674 (23)   MacLean, B.; Tomazela, D. M.; Shulman, N.; Chambers, M.; Finney, G. L.; Frewen, B.;

675   Kern, R.; Tabb, D. L.; Liebler, D. C.; MacCoss, M. J. Skyline: an open source document

676   editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **2010**, *26*

677   (7), 966–968.

678 (24)   Panchaud, A.; Scherl, A.; Shaffer, S. A.; von Haller, P. D.; Kulasekara, H. D.; Miller, S.

679   I.; Goodlett, D. R. Precursor acquisition independent from ion count: how to dive deeper

680   into the proteomics ocean. *Anal. Chem.* **2009**, *81* (15), 6481–6488.

681 (25)   Chambers, M. C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D. L.; Neumann, S.;

682   Gatto, L.; Fischer, B.; Pratt, B.; Egertson, J.; et al. A cross-platform toolkit for mass

683   spectrometry and proteomics. *Nat. Biotechnol.* **2012**, *30* (10), 918–920.

684 (26)   Ting, Y. S.; Egertson, J. D.; Payne, S. H.; Kim, S.; MacLean, B.; Käll, L.; Aebersold, R.;

685   Smith, R. D.; Noble, W. S.; MacCoss, M. J. Peptide-Centric Proteome Analysis: An

686   Alternative Strategy for the Analysis of Tandem Mass Spectrometry Data. *Mol. Cell.*

687   *Proteomics* **2015**, *14* (9), 2301–2307.

688 (27)   Wolf-Yadlin, A.; Hautaniemi, S.; Lauffenburger, D. A.; White, F. M. Multiple reaction

689   monitoring for robust quantitative proteomic analysis of cellular signaling networks. *Proc.*

690   *Natl. Acad. Sci. U. S. A.* **2007**, *104* (14), 5860–5865.

691 (28)   Picotti, P.; Aebersold, R. Selected reaction monitoring-based proteomics: workflows,

692   potential, pitfalls and future directions. *Nat. Methods* **2012**, *9* (6), 555–566.

693   (29)   Moreira, R.; Pereiro, P.; Canchaya, C.; Posada, D.; Figueras, A.; Novoa, B. RNA-Seq in

694   Mytilus galloprovincialis: comparative transcriptomics and expression profiles among

695   different tissues. *BMC Genomics* **2015**, *16* (1), 728.

696   (30)   Hijmans, R. J. Geographic Data Analysis and Modeling [R package raster version 2.5-8].

697   (31)   Kolde, R. *pheatmap: Pretty Heatmaps*; 2015.

698   (32)   Matsumoto, T.; Nakamura, A. M.; Mori, K.; Kayano, T. Molecular characterization of a

699   cDNA encoding putative vitellogenin from the Pacific oyster Crassostrea gigas. *Zoolog. Sci.*

700   **2003**, *20* (1), 37–42.

701   (33)   Arcos, F. G.; Ibarra, A. M.; Rodríguez-Jaramillo, M. del C.; García-Latorre, E. A.;

702   Vazquez-Boucard, C. Quantification of vitellin/vitellogenin-like proteins in the oyster

703   Crassostrea corteziensis (Hertlein 1951) as a tool to predict the degree of gonad maturity.

704   *Aquac. Res.* **2009**, *40* (6), 644–655.

705   (34)   Corporeau, C.; Vanderplancke, G.; Boulais, M.; Suquet, M.; Quéré, C.; Boudry, P.;

706   Huvet, A.; Madec, S. Proteomic identification of quality factors for oocytes in the Pacific

707   oyster Crassostrea gigas. *J. Proteomics* **2012**, *75* (18), 5554–5563.

708   (35)   Blaise, C.; Gagné, F.; Pellerin, J.; Hansen, P. D. Determination of vitellogenin-like

709   properties inMya arenaria hemolymph (Saguenay Fjord, Canada): A potential biomarker for

710   endocrine disruption. *Environ. Toxicol.* **1999**, *14* (5), 455–465.

711   (36)   Won, S.-J.; Novillo, A.; Custodia, N.; Rie, M. T.; Fitzgerald, K.; Osada, M.; Callard, I. P.

712   The Freshwater Mussel (Elliptio complanata) as a Sentinel Species: Vitellogenin and Steroid

713   Receptors. *Integr. Comp. Biol.* **2005**, *45* (1), 72–80.

714   (37)   Fusi, F. M.; Vignali, M.; Busacca, M.; Bronson, R. A. Evidence for the presence of an

715   integrin cell adhesion receptor on the oolemma of unfertilized human oocytes. *Mol. Reprod.*

716   *Dev.* **1992**, *31* (3), 215–222.

717   (38)   Campbell, S.; Swann, H. R.; Seif, M. W.; Kimber, S. J.; Aplin, J. D. Integrins and

718   adhesion mlecules: Cell adhesion molecules on the oocyte and preimplantation human

719   embryo. *Hum. Reprod.* **1995**, *10* (6), 1571–1578.

720   (39)   Suárez, M. P.; Alvarez, C.; Molist, P.; Juan, F. S. PARTICULAR ASPECTS OF

721   GONADAL CYCLE AND SEASONAL DISTRIBUTION OF GAMETOGENIC STAGES

722   OF MYTILUS GALLOPROVINCIALIS CULTURED IN THE ESTUARY OF VIGO. *J.*

723   *Shellfish Res.* **2005**, *24* (2), 531–540.

724   (40)   Pedersen, S. K.; Harry, J. L.; Sebastian, L.; Baker, J.; Traini, M. D.; McCarthy, J. T.;

725   Manoharan, A.; Wilkins, M. R.; Gooley, A. A.; Righetti, P. G.; et al. Unseen proteome:

726   mining below the tip of the iceberg to find low abundance and membrane proteins. *J.*

727   *Proteome Res.* **2003**, *2* (3), 303–311.

728   (41)   Groh, K. J.; Nesatyy, V. J.; Segner, H.; Eggen, R. I. L.; Suter, M. J.-F. Global proteomics

729   analysis of testis and ovary in adult zebrafish (Danio rerio). *Fish Physiol. Biochem.* **2011**, *37*

730   (3), 619–647.

731   (42)   Martin, G. G.; Jo Ellen Hose; Chong, C.; Hoodbhoy, T.; McKrell, N. Localization and

732   roles of coagulogen and transglutaminase in hemolymph coagulation in decapod

733   crustaceans. *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry*

734   **1991**, *100* (3), 517–522.

735   (43)   Nozawa, H.; Mori, T.; Kimura, M.; Seki, N. Characterization of a transglutaminase from

736   scallop hemocyte and identification of its intracellular substrates. *Comp. Biochem. Physiol.*

737   *B Biochem. Mol. Biol.* **2005**, *140* (3), 395–402.

738   (44)   Chaney, M. L.; Gracey, A. Y. Mass mortality in Pacific oysters is associated with a

739   specific gene expression signature. *Mol. Ecol.* **2011**, *20* (14), 2942–2954.

740   (45)   Sathapondecha, P.; Treerattrakool, S.; Panyim, S.; Udomkit, A. Potential roles of

741   transglutaminase and thioredoxin in the release of gonad-stimulating factor in Penaeus

742   monodon: implication from differential expression in the brain during ovarian maturation

743   cycle. *Mar. Genomics* **2011**, *4* (4), 279–285.

744   (46)   Swanson, W. J.; Vacquier, V. D. The rapid evolution of reproductive proteins. *Nat. Rev.*

745   *Genet.* **2002**, *3* (2), 137–144.

746   (47)   Plumel, M. I.; Wasselin, T.; Plot, V.; Strub, J.-M.; Van Dorsselaer, A.; Carapito, C.;

747   Georges, J.-Y.; Bertile, F. Mass spectrometry-based sequencing and SRM-based

748   quantitation of two novel vitellogenin isoforms in the leatherback sea turtle (Dermochelys

749   coriacea). *J. Proteome Res.* **2013**, *12* (9), 4122–4135.

750   (48)   Klee, E. W.; Bondar, O. P.; Goodmanson, M. K.; Dyer, R. B.; Erdogan, S.; Bergstralh, E.

751   J.; Bergen, H. R., 3rd; Sebo, T. J.; Klee, G. G. Candidate serum biomarkers for prostate

752   adenocarcinoma identified by mRNA differences in prostate tissue and verified with protein

753   measurements in tissue and blood. *Clin. Chem.* **2012**, *58* (3), 599–609.

754