Shotgun proteomics as a viable approach for biological discovery in the Pacific oyster

Emma Timmins-Schiffman1, Brook L Nunn2, David R Goodlett2, Steven B Roberts1\*

1 University of Washington, School of Aquatic and Fishery Sciences

Box 355020

Seattle, WA 98195

2 University of Washington, Medicinal Chemistry

Box 357610

Seattle, WA 98195

\*Corresponding author

ABSTRACT

 Shotgun proteomics offers an efficient means to characterize proteins in a complex mixture, particularly when sufficient genomic resources are available. In order to assess the practical application of shotgun proteomics in the Pacific oyster, *Crassostrea gigas*, liquid chromatography coupled with tandem mass spectrometry was used to characterize the gill proteome. Using information from the recently published Pacific oyster proteome, 1,671 proteins were identified. Biological samples (n=4) and corresponding technical replicates (3) were similar in both specific proteins identified and expression as determined by total mass spectral counts. A majority of the proteins identified (1,249) were present in all biological samples. Functional analysis of the protein repertoire illustrates these proteins represent a wide range of biological processes, supporting the dynamic function of the gill. *In silico* analysis indicated this sequencing effort identified a large proportion of the complete gill proteome. Together these data demonstrate shotgun sequencing is a viable approach for biological discovery that will play an important role in future studies on oyster physiology.

Keywords: proteomics, Pacific oyster, *Crassostrea gigas*

INTRODUCTION

Fluctuation in gene and protein expression can be sensitive and specific indicators of biological processes. At the transcript level, several methodologies can be used to characterize expression from the gene-centric to systems level, including qPCR (e.g. Griffitt et al. 2006; Stumpp et al. 2011), microarrays (e.g. Todgham and Hofmann 2009; Lockwood et al. 2010), and high-throughput sequencing (e.g. Polato et al. 2011; Philipp et al. 2012). The use of high-throughput sequencing technology has exponentially increased available genome and transcript information for taxa of ecological interest in recent years. While these results provide an accurate portrayal of changes at the molecular level, it is common that proteins have a more direct role in regulating physiological processes and responding to environmental change.

Historically there have been several technical and analytical challenges in characterizing global protein expression. One challenge is the need to have sufficient genomic resources available to describe proteins of interest. Specifically, protein sequencing generally produces short amino acid fragments that require a known corresponding gene for identification and annotation purposes. However, the lack of genomic resources have not totally hampered proteomic studies. For example, researchers characterized the physiological response of *Gillichthys mirabilis* gill tissue exposed to osmotic and temperature stress using two dimensional gel electrophoresis without sequencing proteins (Kültz & Somero 1996). In another study, researchers used Surface Enhanced Laser Desorption/Ionisation and identified 11 differentially expressed proteins in the gill tissue of *Oncorhynchus mykiss* exposed to zinc stress (Hogstrand et al. 2002). Four proteins were identified based on a combination of their physical properties (*i.e.* mass and binding) coupled with sequence similarity comparisons with limited teleost protein sequences in the SwissProt database (Hogstrand et al. 2002).

Using predicted protein sequences in closely related species can assist in annotation, but species specific information will provide more accurate results. This is evident in a study on protein expression in pea (*Pisum sativum*) chloroplasts where concurrent cDNA sequencing facilitated a greater number of protein identifications compared to identifications through homology searches with closely related model species (Bräutigam et al. 2008). The reason that species specific information provides such an advantage is due to how modern day protein sequence identification is carried out. The vast majority of high-throughput mass spectrometry proteomics is accomplished by matching observed peptide fragmentation patterns (tandem mass spectra) to theoretical spectra. This is possible because peptides fragment in a predictable manner allowing for theoretical tandem mass spectra to be created *in silico* from a given protein sequence, stressing the importance of the database used. These correlation-based algorithms require the peptide mass (precursor mass) and peptide fragmentation (tandem mass spectrum). Even when employing databases of closely related species a large number of peptides viable tandem mass spectra might not accurately be assigned to a protein as a single amino acid mutation could significantly alter the peptide mass and resulting fragmentation pattern.

As technological advances have continued to increase accessibility of whole transcriptomes and genomes to researchers, there is increasing interest in leveraging these data to carry out proteomic studies. Recently, the Pacific oyster genome was sequenced (Zhang et al. 2012). Given the availability of this resource, our objective was to quantify the level of information (and respective variability) attainable in proteomic studies in oysters. There have been a several prior studies examining protein expression in oysters using liquid chromatography coupled with tandem mass spectrometry (MS) with samples separated by two-dimensional gel electrophoresis (2-DE) beforehand. These proteomic techniques have identified specific proteins that are responsible for early developmental changes in *C. gigas* (Huan et al. 2012) and larval *C. gigas* response to elevated *p*CO2 (ocean acidification) (Dineshram et al. 2012). Similarly, 2-DE methods have been used to identify and sequence proteins that are differentially regulated in a range of physiological situations in adult oyster species. For instance, antioxidant proteins are up-regulated in adult *Crassostrea virginica* mantle tissue upon exposure to ocean acidification (Tomanek et al. 2011), high quality oocytes in *Crassostrea gigas* have specific protein expression profiles (Corporeau et al. 2012), *C. hongkongensis* express both general and specific responses to metal contamination (Liu & Wang 2012), and differential proteome regulation in response to acid sulfate runoff has been described in *Saccostrea glomerata* (Amaral et al. 2012).

 An alternative to 2-DE approaches is to perform shotgun proteomics. Shotgun proteomics is the sequencing of a complex mixture of peptides using liquid chromatography and tandem MS without prior separation (*i.e.* 2-DE). One of the main advantages of using 2-DE methods is that information on the proteins’ physical properties (mass, isoelectric point) can be used in the protein identification, whereas these empirical data are lost in the strictly tandem MS approaches. However, tandem MS has significantly greater data efficiency than gel-based approaches. The use of shotgun proteomics allows for a greater number of proteins to be rapidly identified from a single sample providing a more complete metabolic picture of cellular function and physiology. With the exception of Muralidharan et al. (2012) who used shotgun proteomics to uncover *Saccostrea glomerata* hemocyte proteomic responses to metal contamination, there are limited examples in marine invertebrates where this approach has been implemented.

In this study, we used shotgun proteomics to sequence the gill proteome of the Pacific oyster, *Crassostrea gigas*. The goal of this study was to determine the effectiveness of using a shotgun proteomics approach and to functionally characterize proteins expressed in gill tissue.

MATERIALS AND METHODS

*Oysters*

 Pacific oysters (*Crassostrea gigas*, 18 months old) were collected in Shelton, Washington, United States. Oysters were transferred to Friday Harbor Laboratories (Friday Harbor, WA, USA) into a flow-through system at 13°C for 6 weeks. Eight 4-L vessels containing 6 oysters each were kept in a water bath with seawater flowing through at 57.5 mL/min. Vessels were cleaned every other day with fresh water and salt water rinses. Oysters were fed Shellfish Diet 1800 (Reed Mariculture, Campbell, CA, USA). At the end of 6 weeks, gill tissue was removed from four oysters and immediately flash frozen in liquid nitrogen for proteomic analysis.

*Protein Digestion and Desalting*

 Gill tissue samples (50-100 mg) were homogenized in 50 mM NH4HCO3 (100 ul) using RNAse-free plastic pestles. Each homogenized gill sample was sonicated four times with a probe sonicator and stored on dry ice between sonications. After sonication, protein concentrations were measured using the Bradford assay, following the manufacturer’s protocol (Pierce, Thermo Fisher Scientific, Rockford, IL, USA). Urea (36mg) was added to each sample (for a total concentration of 6M) to stabilize peptides. Next, 1.5 M Tris (pH 8.8) (6.6 µl) was added followed by 200 mM TCEP (2.5 µl). Samples were incubated for 1 hour at 37°C on a shaker. To alkylate the proteins, 200 mM iodoacetamide (IAM) (20 µl) was added. Samples were then vortexed, and incubated for 1 hour at room temperature in the dark. To absorb excess IAM, 200 mM dithiolthreitol (20 µl) was added, then samples were vortexed and incubated at room temperature for 1 hour. A volume equal to approximately 100 µg was removed and the remainder was discarded. NH4HCO3 (200 µl of 25 mM) was added to dilute the urea and then HPLC grade MeOH (50 µl) was added to each tube. Trypsin was solubilized in a trypsin dilution buffer (20 µl) (Promega, Madison, WI, USA) and 3 µl of this solution was added to each sample to enzymatically digest the proteins. The samples were incubated overnight at 37°C. The next day, the samples were evaporated on the speed vac to near dryness.

Samples were desalted by passage through a pre-prepared MacroSpin column following the manufacturer’s specifications (The Nest Group, Southborough, MA, USA). After desalting, remaining solvent was evaporated using a speed vac.

*Liquid Chromatography and Tandem Mass Spectrometry (LC-MS/MS)*

 Mass spectrometry was performed at the University of Washington Proteomics Resource (Seattle, WA, USA). Samples were resuspended in 2% acetonitrile and 0.1% fluoroacetic acid in water (100 µl). Samples were then vortexed to mix and spun down at 15,000 rpm for 10 minutes. The supernatant was aliquoted to autosampler vials. Nano LC separation was performed with a nanoACUITY system (Waters, Milford, MA, USA) interfaced to an LTQ Orbitrap XL mass spectrometer (Thermo Scientific, San Jose, CA, USA). Peptides were trapped on a 100 µm i.d. x 20 mm long pre-column packed with 200 angstrom (5 µm) Magic C18 particles (C18AQ; Michrom, Auburn, CA, USA). For separation, a 75 µm i.d. x 250 mm long analytical column with a laser pulled emitter tip packed with 100 angstrom (5 µm) Magic C18 particles (C18Q; Michrom) was used and analyzed in positive ion mode. For each LC-MS/MS analysis, an estimated amount of 0.5 µg of peptides was loaded onto the pre-column at 2 µL/min in water/acetonitrile (98%/2%) with 0.1% (v/v) formic acid. Peptides were eluted using an acetonitrile gradient flowing at 240 nL/min using mobile phase consisting of the following: Solvent C (water, 0.1% formic acid) and Solvent D (acetonitrile, 0.1% formic acid). The gradient program was as follows: 0 - 1 minute: Solvent C (98%) Solvent D (2%); 1 minute: Solvent C (90%) Solvent D (10%); 90 minutes: Solvent C (65%) Solvent D (35%); 91 - 101 minutes: Solvent C (20%) Solvent D (80%); 102 - 120 minutes, Solvent C (98%) Solvent D (2%). Peptide spectra were acquired by scans in the Orbitrap followed by the ion trap.

*Data acquisition*

 High resolution full precursor ion scans were acquired at 60,000 resolution in the Orbitrap over 400-2000 m/z while six consecutive tandem mass spectra (MS/MS) were acquired by collision induced dissociation (CID) in the Linear ion trap (LTQ). Data dependent ion threshold was set at 5,000 counts for MS/MS, and the maximum allowed ion accumulation times were 400 ms for full scans and 100 ms for MS/MS measurements. The number of ions accumulated was set to 1 million for Orbitrap scans and 10,000 for linear ion trap MS/MS scans. An angiotensin and neurotensin standard was run after every eight injections. Each sample was injected in triplicate in a novel randomized order.

*Protein Identification and Data Analysis*

 Peptide sequence and corresponding protein identification for all mass spectra were carried out using SEQUEST (Eng et al. 1994) and the *Crassostrea gigas* proteome version 9 (Zhang et al. 2012, <http://dx.doi.org/10.5524/100030>). A DECOY database was created by reversing the *C. gigas* proteome and adding it to the forward database. This was completed in order to determine false positive matches and yielded a false discovery rate of ~0.6%. Search parameters included trypsin as the assigned enzyme and a precursor mass accuracy of ± 3 Da. SEQUEST results were analyzed using PeptideProphet and ProteinProphet in order to statistically evaluate peptide matches and assign protein probabilities (Nesvizhskii et al. 2003). Only proteins with a probability of ≥0.9 (estimated false discovery rate of 0.6%) and a minimum of four total tandem mass spectral assignments in the combined technical replicates were used in further characterizations described below.

In order to annotate corresponding proteins, the *C. gigas* proteome (version 9) was compared to the UniProtKB/Swiss-Prot database (www.uniprot.org) using Blastp with an e-value limit of 1E-10. Associated Gene Ontology terms were used to classify sequences based on biological process as well as categorize genes into parent categories (GO Slim). Enrichment analysis was used to identify overrepresented biological processes in the gill proteome compared to the entire proteome (Database for Annotation, Visualization and Integrated Discovery (DAVID, v. 6.7; Huang et al. 2009a and 2009b, <http://david.abcc.ncifcrf.gov/>)). The results of the enrichment analysis were visualized in REViGO (Reduce and Visualize Gene Ontology; Supek et al. 2011, http://revigo.irb.hr/). In order to assess biological variation of expressed proteins, technical replicates were pooled by taking the sum of total independent tandem mass spectra for each protein to provide an expression value for a given oyster (*i.e.* biological sample). Comparisons of proteins identified across biological samples were visualized using Venny (Oliveros 2007).

The minimum number of peptides needed to be sequenced to optimize unique protein identifications was determined using an *in silico* approach. A list was constructed of all sequenced peptides and their matching protein identification. Redundancies were maintained in this list, so that if a certain peptide was sequenced multiple times it was included multiple times in the list. Randomized subsets of this list were generated using the sample function in R (R core Development Core Team 2009). The number of hypothetically sequenced peptides in these lists ranged from 500-70000. A plot was generated to visualize the relationship between each sample size of randomly chosen peptides and the number of unique proteins identified.

RESULTS

*LC-MS/MS*

A combined total of 175,818 tandem MS spectra were generated across all four biological and three technical replicates using the Orbitrap mass spectrometer. A total of 54,521 unique peptides contributed to the identification of 2,850 proteins with probability score threshold of 0.9 (Appendix 1). 1,671 of these proteins had at least four tandem mass spectral in the combined technical replicates. The mean amino acid coverage across all proteins was 13.3%. Protein identifications for each injection including protein probability scores, number of total and unique spectra, and peptide sequences are provided in Appendix 2.

 For all biological samples, the number of proteins identified in each technical replicate was consistent with minimal standard deviation (2.2-3.5%). Likewise, for each oyster sample 56-58% of the identified proteins were present in all three technical replicates. Using spectral counts as a proxy for relative expression, levels were consistent across technical replicates (Figure 1).

The number of proteins identified in each oyster (after pooling technical replicates, see Methods) was 1,558, 1,554, 1,519, and 1,500 (Table 1). Most proteins (n=1,249) were identified across all biological samples (Figure 2).

In order to evaluate general protein expression and assess sample variability, the ten most highly expressed proteins in all oysters were identified. These 12 proteins represent core cell structure and function, such as maintenance of membrane charge and cellular attachment, cytoskeleton structure, muscle components, turnover of intracellular proteins, and protection against oxidative stress. Seven out of twelve of these proteins (basement membrane-specific heparan sulfate proteoglycan core protein, rootletin, actin, probable cytosol aminopeptidase, myosin heavy chain, filamin-C, extracellular superoxide dismutase) were identified in the top ten most expressed proteins in all four oysters.

Of the 1,671 proteins expressed across all samples, 1,496 were annotated using the UnitProt-KB/SwissProt database. Of the annotated proteins, 1,368 were associated with Gene Ontology classifications. A majority of proteins were associated with the biological process of transport (n=175), followed by protein metabolism (n=124), and RNA metabolism (n=115).

Enrichment analysis was carried out to determine which biological processes were overrepresented in gill tissue as compared to the entire proteome (Figure 4). Several of the functional groups identified were associated with the abundant proteins involved in metabolism and transport as well as structural processes (*i.e.* actin-filament, microtubule). The most significantly enriched biological process was oxidation-reduction.

The number of unique proteins identified with different numbers of sequenced peptides created an exponential curve (Figure 5). The plateau began around 30,000-40,000 sequenced peptides with a total of 2,400-2,516 unique peptides identified. New unique peptides were still identified in larger sample sizes of peptides, but the return per sequenced peptide diminished.

DISCUSSION

 Technical and analytical challenges have resulted in limited focus on quantitative proteomics approaches in environmental physiology. Given the recent technological advances in the proteomics field (Yates et al. 2009) and the recent release of the Pacific oyster genome (Zhang et al. 2012), we set out to assess the practical use of quantitative proteomics in this model species. For all biological samples a majority of the proteins identified (56-58%) were present in all respective technical replicates. Relative expression across technical replicates was also consistent (Figure 1). However, there were proteins not identified in all technical replicates. Thus, proteins with limited expression might not be detected and/or expression levels might not be accurately reflected. It should be noted that the inclusion of proteins in our analysis is highly dependent on threshold selection. In the current study there needed to be 4 total spectra across the combined technical replicate data for identification purposes. Thus, if the threshold was adjusted to be more conservative (i.e. require 4 total spectra for each replicate) variability would be reduced.

The number of proteins identified and subsequently annotated can vary tremendously based on experimental design, target tissue, match thresholds, and genomic resources available. In the present study, the majority of the proteins (1,249) were identified in all biological samples. Based on *in silico* analysis (Figure 5) we have sequenced a relative complete proteome for oyster gill tissue. In a study of European whitefish, *Coregonus lavaretus*, proteomics on fish larvae yielded sequencing of peptides corresponding to 1,500 proteins (Papakostas et al. 2012). The similar number of protein identifications in whitefish compared to our study is likely associated with the tissue complexity. In the whitefish study whole body tissue was examined. In a metaproteomics study of marine microbes, 2,273 distinct proteins were identified across 10 samples (Morris et al. 2010). The large number of proteins identified by Morris et al. (2010) is evidence of the large number of organisms and ecological niches that were sampled in their study. Only 49 proteins were identified in the hemolymph of Sydney rock oysters, *Saccostrea glomerata*, likely due in part to the limited cell types in the biological material analyzed and limited genomic resources to compare mass spectrometry data (Muralidharan et al. 2012).

In addition to assessing the feasibility of shotgun proteomics in the Pacific oyster we also were able to provide a functional characterization of the gill proteome. Gene ontology characterization identified that a majority of the proteins were associated with transport, RNA metabolism and protein metabolism. These biological functions would be expected since gill tissue is the primary interface between the oyster and the environment (water) where the tissue’s major functions include ion regulation, respiration, and sorting of food particles. The high number of proteins involved in RNA and protein metabolism is not necessarily unique to gill tissue but does likely reflect that the multifunctional nature of the tissue that responds to variable environments. Enrichment analysis was also performed to identify which functional group of proteins expressed in gill tissue were overrepresented as compared to the complete protein repertoire. Several of the functional groups identified were associated with the abundant proteins involved in metabolism and transport as well as structural processes. The most significantly enriched biological process was oxidation reduction. Specific proteins that contributed to this are enzymes instrumental in the physiological response to oxidative stress including superoxide dismutases, peroxiredoxins, and catalase. Previous transcriptomic based studies on oysters support that oxidative stress response is an important role of the gill tissue (e.g. David et al. 2007; Fleury & Huvet 2012). The enrichment of these proteins in gill compared to the entire proteome reinforces the integral role that the gill plays in environmental response.

 The success of the shotgun sequencing effort was due in part to the recent publication of the *C. gigas* genome, emphasizing that the dissemination of genomic resources provides invaluable resources for the scientific community. Together these research results demonstrate shotgun sequencing is viable approach for biological discovery and will likely play an important role in future studies on oyster physiology.

ACKNOWLEDGEMENTS

 We would like to thank Taylor Shellfish for donating the oysters used for this study, Emily Carrington and Ken Sebens for providing lab space and Ronen Elad who assisted with oyster care. Priska von Haller helped with experimental design for the LC-MS/MS and provided technical assistance. Jimmy Eng and Tahmina “Eva” Jahan ran the database searches for protein identifications. This work was supported by National Oceanographic and Atmospheric Administration Saltonstall-Kennedy Program grant # NA09NMF4270093.

REFERENCES

Amaral V., Thompson E.L., Bishop M.J., Raftos D.A., 2012, The proteomes of Sydney rock oysters vary spatially according to exposure to acid sulfate runoff. Mar. Freshw. Res., 63, 361-369.

Bräutigam A., Shrestha R.P., Whitten D., Wilkerson C.G., Carr K.M., Froehlich J.E., Weber A.P.M., 2008, Low-coverage massively parallel pyrosequencing of cDNAs enables proteomics in non-model species: Comparison of a species-specific database generated by pyrosequencing with databases from related species for proteome analysis of pea chloroplast envelopes. J. Biotechnol., 136, 44-53.

Corporeau C., Vanderplancke G., Boulais M., Suquet M., Quéré C., Boudry P., Huvet A., Madec S., 2012, Proteomic identification of quality factors for oocytes in the Pacific oyster *Crassostrea gigas*. J. Proteomics, 75, 5554-5563.

David E., Tanguy A., Moraga D., 2007, Peroxiredoxin 6 gene: A new physiological and genetic indicator of multiple environmental stress response in Pacific oyster *Crassostrea gigas*. Aquat. Toxicol., 84, 389-398.

Dineshram R., Wong K.K.W., Xiao S., Yu Z., Qian P.Y., Thiyagarajan V., 2012, Analysis of Pacific oyster larval proteome and its response to high-CO2. Mar. Pollut. Bull., 64, 2160-2167.

Eng J.K., McCormack A.L., Yates J.R., 1994, An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. J. Am. Soc. Mass Spectrom., 5, 976-989.

Fleury E., Huvet A., 2012, Microarray analysis highlights immune response of Pacific oysters as a determinant of resistance to summer mortality. Mar. Biotechnol., 14, 203-217.

Griffitt R.J., Chandler G.T., Greig T.W., Quattro J.M., 2006, Cathepsin B and glutahione peroxidase show differing transcriptional responses in the grass shrimp, *Palaemonetes pugio* following exposure to three xenobiotics. Environ. Sci. Technol., 40, 3640-3645.

Hogstrand C., Balesaria S., Glover C.N., 2002, Application of genomics and proteomics for study of the integrated response to zinc exposure in a non-model fish species, the rainbow trout. Comp. Biochem. Physiol. B, Biochem. Mol. Biol., 133, 523-535.

Huan P., Wang H., Dong B., Liu B., 2012, Identification of differentially expressed proteins involved in the early larval development of the Pacific oyster *Crassostrea gigas*. J. Proteomics, 75, 3855-3865.

Huang D.W., Sherman B.T., Lampicki R.A., 2009a, Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. Nat. Protoc., 4, 44-57.

Huang D.W., Sherman B.T., Lempicki R.A., 2009b, Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res., 37, 1-13.

Kültz D., Somero G.N., 1996, Differences in protein patterns of gill epithelial cells of the fish *Gillichthys mirabilis* after osmotic and thermal acclimation. J. Comp. Physiol. B, Biochem. Syst. Environ. Physiol., 166, 88-100.

Liu F., Wang W., 2012, Proteome pattern in oysters as a diagnostic tool for metal pollution. J. Hazard. Mater., 239-240, 241-248.

Lockwood B.L., Sanders J.G., Somero G.N., 2010, Transcriptomic responses to heat stress in invasive and native blue mussels (genus *Mytilus*): molecular correlates of invasive success. J. Exp. Biol., 213, 3548-3558.

Morris R.M., Nunn B.L., Frazer C., Goodlett D.R., Ting Y.S., Rocap G., 2010, Comparative metaproteomics reveals ocean-scale shifts in microbial nutrient utilization and energy transduction. ISME J., 4, 673-685.

Muralidharan S., Thompson E., Raftos D., Birch G., Haynes P.A., 2012, Quantitative proteomics of heavy metal stress responses in Sydney rock oysters. Proteomics, 12, 906-921.

Nesvizhskii A.I., Keller A., Kolker E., Aebersold R., 2003, A Statistical Model for Identifying Proteins by Tandem Mass Spectrometry. Anal. Chem., 75, 4646-4658.

Oliveros J.C., 2007, VENNY. An interactive tool for comparing lists with Venn Diagrams.<http://bioinfogp.cnb.csic.es/tools/venny/index.html>

Papakostas S., Vasemăgi A., Văhă J., Himberg M., Peil L., Primmer C.R., 2012, A proteomics approach reveals divergent molecular responses to salinity in populations of European whitefish (*Coregonus lavaretus*). Mol. Ecol., 21, 3516-3530.

Philipp E.E.R., Kraemer L., Melzner F., Poustka A.J., Thieme S., Findeisen U., Schreiber S., Rosenstiel P., 2012, Massively Parallel RNA Sequencing Identifies a Complex Immune Gene Repertoire in the lophotrochozoan *Mytilus edulis*. PLoS One, 7, e33091.

Polato N.R., Vera J.C., Baums I.B., 2011, Gene Discovery in the Threatened Elkhorn Coral: 454 Sequencing of the *Acropora palmata* Transcriptome. PLoS One, 6, e28634.

R Development Core Team (2009) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL [http://www.R-project.org](http://www.r-project.org)

Stumpp M., Dupont S., Thorndyke M.C., Melzner F., 2011, CO2 induced seawater acidification impacts sea urchin larval development II: Gene expression patterns in pluteus larvae. Comp. Biochem. Physiol. A Mol. Integr. Physiol., 160, 320-330.

Supek F., Bošnjak M., Škunca N., Šmuc T., 2011, REVIGO summarizes and visualizes long lists of Gene Ontology terms. PLoS One. 6, e21800.

Todgham A.E., Hofmann G.E. (2009) Transcriptomic response of sea urchin larvae *Strongylocentrotus purpuratus* to CO2-driven seawater acidification. J. Exp. Biol., 212, 2579-2594.

Tomanek L., Zuzow M.J., Ivanina A.V., Beniash E., Sokolova I.M., 2011, Proteomic response to elevated PCO2 level in eastern oysters, *Crassostrea virginica*: evidence for oxidative stress. J. Exp. Biol., 214, 1836-1844.

Yates J.R., Ruse C.I., Nakorchevsky A., 2009, Proteomics by Mass Spectrometry: Approaches, Advances, and Applications. Annu. Rev. Biomed. Eng., 11, 49-79.

Zhang G., Fang X., Guo X., Li L., Luo R., Xu F., Yang P., Zhang L., Wang X., Qi H., Xiong Z., Que H., Xie Y., Holland P.W.H., Paps J., Zhu Y., Wu F., Chen Y., Wang J., Peng C., Meng J., Yang L., Liu J., Wen B., Zhang N., 2012, The oyster genome reveals stress adaptation and complexity of shell formation. Nature, 490, 49-54.

APPENDIX

**Appendix 1**: Proteins identified based on tandem mass spectra compared to the *Crassostrea gigas* proteome. The protein accession number is provided for each protein identified (n=2,850). When proteins could be annotated, SwissProt Accession Number (SPID), e-value for the BLASTp search, and gene description are provided. For all proteins, the total number of tandem mass spectra is provided.

**Appendix 2**: Spectra, peptide, and corresponding protein information for all 12 injections (4 biological samples x 3 technical replicates). Sheets are labeled as oyster ID and technical replicate (*i.e.* B\_03 is the third technical replicate for oyster B). Each protein search result includes an entry number, one or more protein accession numbers, the probability that the protein assignment is correct, the percent coverage of the protein by the sequenced peptides, the number of unique peptides (non-redundant) used to identify the protein, the total number of peptides used to identify the protein, and the peptide sequences. Some of the entries have multiple protein accession numbers because numbers correspond to proteins with the same or highly similar sequences.

FIGURE CAPTIONS

Figure 1. Total independent spectral counts for three technical replicates for oyster “D” plotted for each protein (n=1,500). Similar patterns were observed for the other 3 oysters (data not shown).

Figure 2. Venn diagram of proteins identified among biological samples. Proteins identified in oyster A are in the blue ellipse, B are in yellow, C are in green, and oyster D proteins are in red.

Figure 3. Representation of biological processes corresponding to the proteins identified from oyster gill tissue.

Figure 4. Biological processed found to be significantly (p<0.05) enriched in gill proteome compared to the full proteome. Processes are arranged in semantic space with color corresponding to significance (log10 p-value).

FIGURE 5. Predicted number of unique proteins that would be identified based on a sequential increase in peptides sequences.

TABLES

Table 1. Summary of the number of proteins identified for each oyster (labeled A-D).



 Table 2. Twelve most abundant proteins in gill proteome as determined by identifying the ten most abundant protein in each oysters. Protein ID is given, as well as protein description from UniProt-KB/SwissProt, SwissProt Accession Number, and the oysters where the protein was detected.



FIGURES

Figure 1



Figure 2



Figure 3



Figure 4



Figure 5

