Molecular techniques for resilient Pacific oyster (*Crassostrea gigas*) aquaculture

Yaamini R. Venkataraman

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2021

Reading Committee:

Steven B. Roberts, Chair

Jacqueline Padilla-Gamiño

Jonathan P. Davis

Program Authorized to Offer Degree:
School of Aquatic and Fishery Sciences

University of Washington

Abstract

Molecular techniques for resilient Pacific oyster (*Crassostrea gigas*) aquaculture

Yaamini R. Venkataraman

Chair of the Supervisory Committee:
Associate Professor Steven B. Roberts
School of Aquatic and Fishery Sciences

As ocean acidification continues to impact marine ecosystems at unprecedented rates, phenotypic plasticity may allow organisms to withstand more stressful conditions. Genomic methods can elucidate molecular mechanisms that contribute to phenotypic plasticity, allowing for a deeper understanding of how physiological processes will be impacted by low pH. My dissertation examines the effects of ocean acidification on the Pacific oyster (*Crassostrea gigas*) stress response and reproduction; elucidate how exposure history impacts phenotype; and explore the role of functional role DNA methylation in somatic and reproductive tissue. I investigated the effect of regional environmental variation on the molecular physiology of *C. gigas* outplanted at five different estuarine sites (four in Puget Sound, one in Willapa Bay) in Washington, USA using gel-free proteomic methods. While there was no difference in survival, or any protein abundances due to pH differences between sites, *C. gigas* outplanted at the site with the highest temperature had significantly higher abundances of antioxidant enzymes and molecular chaperones, elucidating the molecular underpinnings of thermotolerance. In a hatchery setting, I explored the impact of ocean acidification on reproductive maturity and output. A seven week low pH exposure did not affect sex ratio or maturation stage; however, it did significantly affect survival of larvae. Even though adult oysters

spent four months in ambient pH conditions between low pH exposure and strip spawning, larvae from females that experienced low pH conditions had significantly higher mortality. Finally, I conducted the first investigations examining the effect of ocean acidification in *C. gigas* methylomes. To investigate the role of environmentally-responsive methylation in reproductive tissue, I analyzed gonad methylomes of female *C. gigas* exposed to low pH. A total of 1,599 differentially methylated loci (DML) were found in gene bodies. The genic DML were associated with cilium movement, development, and cytoskeletal processes, implying a need to regulate cellular growth in the gonad in response to low pH. I then explored the influence of low pH on the somatic tissue methylome using diploid and triploid oyster ctenidia. Differences in ploidy status yielded 154 DML. These ploidy-DML were associated with cell-cell adhesion and dephosphorlylation processes, which are not commonly associated with methylome changes in organisms that undergo natural polyploidization. The 178 pH-DML were associated with processes commonly observed in oysters exposed to ocean acidification, including apoptosis, protein ubiquitination, zinc ion binding, and cytoskeletal processes. In both reproductive and somatic tissue, the enrichment of DML in genes with multiple transcripts could indicate a role for methylation to regulate gene expression via alternative splicing. Investigating the molecular underpinnings of responses to ocean acidification in *C. gigas* will provide a thorough understanding of this global aquaculture product's ability to withstand future ocean conditions.

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

DEDICATION

To my Appa, Dr. Shankar Venkataraman. Your Appa reminded you not to forget to complete your Ph.D, so that you could do the same for me.

I completed it.

INTRODUCTION

Pacific oysters (*Crassostrea gigas*; Thunberg 1793) are a commercially and ecologically relevant species, making them ideal models for studying physiological responses to ocean acidification. Research efforts have already identified consequences of ocean acidification for distinct life stages. Pacific oyster larvae experience developmental delays and reduced shell growth in response to experimental ocean acidification (Gazeau et al., 2011; Kurihara, Kato, & Ishimatsu, 2007; Timmins-Schiffman, O'Donnell, Friedman, & Roberts, 2013; Waldbusser et al., 2014). Key protein pathways are disrupted during metamorphosis (Dineshram et al., 2016) and adulthood (Timmins-Schiffman et al., 2014). As seawater pCO$_2$ increases, adult *C. gigas* calcification rates decrease (Gazeau et al., 2007), and their shells display significantly lower fracture toughness (Timmins-Schiffman et al., 2014). Exposure to experimental ocean acidification also negatively impacts growth, sperm motility, and egg viability in adult oysters (Omoregie, Mwatilifange, & Liswaniso, 2019). Although there is extensive research on ocean acidification's impact on *C. gigas*, key uncertainties remain surrounding the mechanisms behind observed physiological responses. Knowing how the environment shapes physiological phenotypes on a mechanistic level could help identify potential pathways for future-proofing *C. gigas* aquaculture operations.

Genomics are the next frontier for understanding how environmental variability affects an organism's physiological response and ability to acclimate to future ocean conditions. Epigenetic analysis can provide a direct link between changes in the environment and gene expression regulation. The epigenome consists of gene expression changes that do not arise from changes in the DNA sequence itself, with methylation of cytosine bases being the most studied mechanism (Bird, 2002; Deans & Maggert, 2015). Initial characterization of the *C. gigas* methylome has found that methylation occurs in a mosaic pattern and is concen-

trated in gene regions (Gavery & Roberts, 2013). Environmental response genes that are less methylated are likely prone to more spurious transcription and alternative splicing patterns, thereby increasing phenotypic plasticity (Roberts & Gavery, 2012). Recent studies have demonstrated that changes in the environment can affect DNA methylation patterns in other marine invertebrates (Eirin-Lopez & Putnam, 2018), so it is possible genomic regulation by DNA methylation may also be important for adaptation and acclimation in *C. gigas*. Additionally, manipulation of methylation patterns could be a method for "stress hardening" cultured species in hatchery settings (Gavery & Roberts, 2017). Similarly, examination of the proteome — all the proteins in a sample — can shed light on physiological changes on a molecular level, since proteins direct all cellular functions (Tomanek, 2014). Since the proteome is dynamic, proteomic analysis can capture organismal response to real-time environmental conditions, similar to those experienced by *C. gigas* outplanted at aquaculture grow-out sites.

My dissertation uses *C. gigas* as a model system to examine molecular mechanisms important for ocean acidification response. I first explored the impacts of natural pH variation on the *C. gigas* proteome. Next, I investigated the effect of ocean acidification on oyster physiology in a controlled hatchery setting. Finally, I elucidated the mechanistic role of DNA methylation in responding to ocean acidification in somatic and reproductive tissues. My work provides a foundation for using molecular tools to promote resilient aquaculture in the face of climate stressors like ocean acidification.

Chapter 1

CHARACTERIZATION OF PACIFIC OYSTER (*CRASSOSTREA GIGAS*) PROTEOMIC RESPONSE TO NATURAL ENVIRONMENTAL DIFFERENCES

1.1 Abstract

Global climate change is rapidly altering coastal marine ecosystems important for food production. A comprehensive understanding of how organisms will respond to these complex environmental changes can come only from observing and studying species within their natural environment. To this end, the effects of environmental drivers — pH, dissolved oxygen content, salinity, and temperature — on Pacific oyster (*Crassostrea gigas*) physiology were evaluated in an outplant experiment. Sibling juvenile oysters were outplanted to eelgrass and unvegetated habitat at five different estuarine sites within the Acidification Nearshore Monitoring Network in Washington State, USA to evaluate how regional environmental drivers influence molecular physiology. Within each site, we also determined if eelgrass presence that buffered pH conditions changed the oysters' expressed proteome. A novel, two-step, gel-free proteomic approach was used to identify differences in protein abundance in *C. gigas* ctenidia tissue after a 29 day outplant by 1) identifying proteins in a data independent acquisition survey step and 2) comparing relative quantities of targeted environmental response proteins using selected reaction monitoring. While there was no difference in protein abundance detected between habitats or among sites within Puget Sound, *C. gigas* outplanted at Willapa Bay had significantly higher abundances of antioxidant enzymes and molecular chaperones. Environmental factors at Willapa Bay, such as higher average temperature, may have driven this protein abundance pattern. These findings generate a suite of new hypotheses for lab

and field experiments to compare the effects of regional conditions on physiological responses of marine invertebrates.

1.2 Introduction

Global climate change will influence estuarine dynamics and impact the organisms that inhabit these environments. Estuaries are already variable across spatial and temporal scales in terms of phytoplankton production (Pennock & Sharp, 1986), nutrient availability (Paerl, Hall, Peierls, & Rossignol, 2014), heavy metal contamination (Liu et al., 2015), salinity (Banas, Hickey, MacCready, & Newton, 2004), and carbonate chemistry (Feely et al., 2010; Pelletier, Roberts, Keyzers, & Alin, 2018). Since climate change will affect these parameters, it is important to consider how estuarine organisms will respond.

Proteomics, or the study of protein abundance and expression, can be used to shed light on physiological changes on a molecular level. Proteins direct all major cellular functions, thus examining protein abundance provides direct evidence of an organism's physiological response to the estuarine environment (Tomanek, 2014). The proteome is dynamic, as it must rapidly respond to perturbation, providing mechanistic information that standard gene expression and mRNA quantification studies cannot (Flores-Nunes et al., 2015; Veldhoen, Ikonomou, & Helbing, 2012). As a result of the proteome's dynamic nature, proteins analyzed at the time of collection represent an organism's response to the environment in near real-time. Long-term exposure to environmental conditions, as well as natural organismal aging, are also reflected in the proteome (Hercus, Loeschcke, & Rattan, 2003). Discovery-based proteomic methods can elucidate responses to environmental drivers (Flores-Nunes et al., 2015). Several studies have connected protein abundances with changes in laboratory-simulated environmental conditions, identifying key proteins and mechanisms involved in specific environmental responses (Dineshram et al., 2016; Meng, Wang, Li, Yin, & Zhang, 2017; Timmins-Schiffman et al., 2014). While these studies provide insight into organismal adaptation and physiology, laboratory studies alone cannot fully encapsulate the effects of

multiple environmental drivers within an ecosystem context (**Riebesell2014?**).

Although challenging, in situ field studies provide a necessary biological realism when considering variable environments (Cornwall & Hurd, 2016; Slattery et al., 2012). Such experiments can be leveraged to study the effects of multiple environmental drivers on organismal physiology and to incorporate realistic variability, as opposed to examining the effect of a single stressor on an organism (**Riebesell2014?**). Through transcriptomics, Chapman et al. (2011) demonstrated the power of an in situ experimental design for examining the impacts of regional environmental conditions on Eastern oyster (*Crassostrea virginica*) physiology. Transcript signatures from *C. virginica* sampled from various locations in southeastern United States revealed that temperature, pH, salinity, dissolved oxygen and pollutant load at each location impacted gene expression. Furthermore, they were able to disentangle the interactions of these environmental factors on gene expression. RNA and protein abundances can be influenced by several environmental factors, and in situ studies can determine which drivers will be more important to consider for organismal physiology.

Marine invertebrates have proven to be informative bioindicators in proteomic studies to examine the effects of in situ conditions on organismal physiological responses to environmental change. When marine invertebrates have been exposed to varying environmental conditions, proteomics have demonstrated changes in cellular defense, immune responses, and genome function (Veldhoen, Ikonomou, & Helbing, 2012). Changes in protein abundance in bivalves like the Pacific oyster (*Crassostrea gigas*) and blue mussels (*Mytilus edulis spp.*) have been used to develop biomarkers for environmental contaminants (Beyer et al., 2017; Slattery et al., 2012). Proteomic responses to natural environmental drivers have also been evaluated in bivalves. For example, shotgun proteomic analysis of *M. edulis* ctenidia from Baltic Sea microcosms revealed that low salinity conditions lead to decreased abundance of cytoskeleton proteins, as well as calcium-binding messenger calmodulin, which plays an important role in signalling and intracellular membrane trafficking pathways (Campos et al., 2016). Using a growing wealth of genomic information to understand how these species fare under differen-

tial environmental conditions is critical for monitoring natural populations and commercial aquaculture.

Pacific oyster (*Crassostrea gigas*) rearing in estuarine environments in Washington State, USA (WA) provides an ideal system to examine the effect of in situ environmental conditions on the expressed proteome. *C. gigas* are extensively farmed in two different estuarine systems that show substantial regional variation: Puget Sound and Willapa Bay. Puget Sound is a complex estuarine system with interconnected sub-basins, each with different freshwater inputs, residence times, and stratification levels (Bianucci et al., 2018; Feely et al., 2010; Pelletier, Roberts, Keyzers, & Alin, 2018). Willapa Bay is a large shallow estuary on the Pacific coast that exchanges approximately half its water volume with the Pacific Ocean at each tide (Banas, Hickey, MacCready, & Newton, 2004; Banas, Hickey, Newton, & Ruesink, 2007). Seasonality and location within Puget Sound dictates temperature, dissolved oxygen, salinity, and pH conditions, while Willapa Bay conditions are influenced by diurnal fluctuations and proximity to either the ocean or rivers draining into the bay (Banas, Hickey, Newton, & Ruesink, 2007; Feely et al., 2010; Ruesink, Yang, & Trimble, 2015).

Both Puget Sound and Willapa Bay also host eelgrass beds (*Zostera spp.*) that affect environmental conditions, such as oxygen concentrations, on diurnal time scales. The 2012 Washington State Blue Ribbon Panel on Ocean Acidification outlines key early actions, which include the examination of "vegetation-based systems of remediation" (Action 6.1.1) to improve local pH through photosynthetic drawdown of carbon dioxide. This experiment set out to test whether protein abundance patterns reflect reduced stress within vegetation. For example, eelgrass beds may reduce emersion stress relative to unvegetated areas through shading, the retention of water, and increased evaporative cooling at low tide. They can also ameliorate effects of ocean acidification through photosynthetic activity. Reduced pathogen prevalence has also been documented in seagrass beds, but not specifically in eelgrass (Lamb et al., 2017). In contrast, eelgrass beds may drive more extreme carbonate chemistry conditions (Pacella, Brown, Waldbusser, Labiosa, & Hales, 2018). Lowe, Kobelt, Horwith, &

Ruesink (2018) also found that *C. gigas* shell strength and survival was significantly lower in eelgrass habitats in WA. Understanding how different aquaculture grow-out locations and habitats will affect the oyster's ability to persist through environmental change is crucial for the industry and the ecosystem.

The purpose of this study was to use proteomic techniques to uncover the impacts of environmental drivers on Pacific oyster physiological outcomes in estuarine environments in WA. Naturally existing environmental variation was harnessed by outplanting *C. gigas* in different locations within Puget Sound and Willapa Bay, and habitat effects were taken into consideration by placing oysters in eelgrass and unvegetated habitats. Gel-free proteomic methods were used to examine the effects of outplant conditions on relative quantities of all expressed proteins in a series of in situ experiments in order to identify differentially abundant proteins. We predicted that differences in environmental drivers at each outplant location and within outplant habitats would yield unique protein abundance patterns. Oysters at outplant locations with warmer water temperatures, more variable water temperatures, lower dissolved oxygen content, lower salinity, or lower pH may have higher abundances of proteins related to environmental response. Eelgrass beds were expected to ameliorate stressful conditions, resulting in lower abundances of environmental stress response proteins than oysters in unvegetated habitats.

1.3 Methods

1.3.1 Shellfish Deployment

Sibling juvenile *C. gigas* (average shell length = 27.2 mm, age = 2 months) were outplanted for 29 days starting June 19, 2016 at five locations: Case Inlet (CI), Fidalgo Bay (FB), Port Gamble Bay (PG), Skokomish River Delta (SK), and Willapa Bay (WB) in Washington State, USA (Table 1.1; Figure 1.1). These sites were selected for differences in environmental parameters, as well as for the presence of unvegetated areas and eelgrass beds within each

site. All sites were part of the Acidification Nearshore Monitoring Network, a network of sensors placed in various WA locations to monitor marine chemistry (ANeMoNe; Washington Department of Natural Resources). Prior to the outplant, oysters were reared in a controlled hatchery setting. At each site and habitat combination, custom-built Durafet-based sensors (Honeywell) were used to monitor pH. Commercially-available MiniDOT loggers (Precision Measurement Engineering) were used to measure dissolved oxygen, and Odyssey loggers were used to measure conductivity. All sensors recorded temperature measurements, and all sensors logged at 10-minute intervals across the outplant period, with the exception of SK, where sensors were installed two days into the outplant period. At each site, juvenile oysters were placed in bags of five oysters each directly onto the substrate at a tidal height of -1.5 MLLW, both inside and outside of eelgrass beds (n=15 per habitat type), for a total of thirty outplanted oysters per site. The animals were uniformly placed less than a lateral distance of 0.5 m from the sensors at the same tidal height as the instruments. Oysters were housed in exclusion cages to prevent predation. Juvenile oysters remained at each site for a 29-day exposure period. Because the ctenidia is the primary site where oysters interact with the environment, ctenidia samples were dissected at the end of the outplant and held on dry ice until storage at -80ºC [Beyer et al. (2017); Meng, Wang, Li, Yin, & Zhang (2017)).

Environmental data was treated as follows. Conductivity observations were removed when less than zero, which occurs when the instrument is dry at low tide. Remaining observations were converted to salinity measurements using the `swSCTp` function in the `oce` package in R (**Kelley2018?**), with temperature at 25ºC and pressure at 10 dbar. For dissolved oxygen, pH, and salinity datasets, data were removed when collected by probes 1) during low tide or 2) when tidal depth was less than one foot to remove readings where the probes may have been exposed. Values collected during low tide or a depths less than one foot were retained for temperature datasets. Outliers were screened using the Tukey method for temperature, dissolved oxygen, pH, and salinity datasets (Hoaglin, Iglewicz, & Tukey, 1986). Uniform outplant tidal heights were checked using `prop.test` in R.

A non-metric multidimensional scaling analysis (NMDS) was used to evaluate differences in environmental parameters. First, mean and variance were calculated for each day of the outplant. Values were log-transformed, and a separate Gower's distance matrix was calculated for daily mean and daily variances, accounting for missing data. The NMDS was conducted with the Gower's distance matrix to visually compare means or variances between sites and habitats. Significant differences between site and habitat were identified using a One-way Analysis of Similarities (ANOSIM) for each environmental parameter. Pairwise ANOSIM tests for significant one-way ANOSIM results and two-way ANOSIM tests by site and habitat were not conducted due to lack of replicates within each site-habitat combination. R Scripts are available in the associated Github repository (Appendix 1).

1.3.2 Protein Discovery

To identify select protein targets for characterization across locations and environmental conditions, a subset of tissue samples were analyzed with data independent acquisition (DIA) mass spectrometry analysis. Two tissue samples were used from each site to make a peptide library and maximize the amount of protein abundance data collected from each sample.

Protein Quantification

Tissue samples were homogenized in a solution of 50 mM NH4HCO3 with 6M Urea (500µl). Tissues were then sonicated 3 times (Fisher Scientific Sonic Dismembrator Model 100) for 10 seconds each and cooled between sonications in a bath of ethanol and dry ice. Protein quantities were measured with the Pierce BCA Protein Assay Kit microplate assay with a limited quantity of sonicated sample (11 µL). The protein concentration was measured via spectroscopy at 540 nm in a Labsystems (Waltham, MA) Multiskan MCC/340 and accompanying Ascent Software Version 2.6. Protein concentrations were calculated based on a standard curve with BSA (Pierce) per manufacturer's instructions.

Protein Digestion

Protein digestion followed the protocol outlined in Timmins-Schiffman, O'Donnell, Friedman, & Roberts (2013). To each sample of 30 µg protein,1.5 M Tris pH 8.8 buffer (6.6 µL) and 200 mM TCEP (2.5 µL) were added. After solvent additions, each sample's pH was verified to be basic (pH 8), and placed on a 37ºC heating block for one hour. Iodoacetamide (200 mM, 20 µL) was then added to each sample to maximize digestion enzyme access to protein cleavage sites. Samples were covered with aluminum foil to incubate in the dark for 1 hour at room temperature. Afterwards, diothiolthreitol (200 mM, 20 µL) was added and samples were incubated at room temperature for one hour. Lysyl Endopeptidase (Wako Chemicals) was then added to each sample in a 1 µg enzyme:30 µg oyster protein ratio, followed by one hour of incubation at room temperature. Urea was diluted with NH_4HCO_3 (25 mM, 800 µL) and HPLC grade methanol (200 µL). Trypsin (Promega) was added to each sample in a 1 µg trypsin: 30 µg oyster protein ratio for overnight digestion at room temperature.

Peptide Isolation

After overnight incubation, samples were evaporated to near dryness at 4ºC with a speedvac (CentriVap ® Refrigerated Centrifugal Concentrator Model 7310021). Samples were then reconstituted in 100 µL of a 5% Acetonitrile and 0.1% Trifluoroacetic Acid (Solvent A) to isolate peptides. If samples were not at pH 2, 10-20 µL aliquots of 10% Formic Acid were added until this pH was achieved.

Before desalting peptide samples, Macrospin C18 columns (The Nest Group) were prepared by adding 200 µL of a 60% Acetonitrile with 0.1% Trifluoroacetic Acid (Solvent B). The columns were spun for three minutes at 2000 rpm, and flow-through liquid from the column was discarded. The spinning and discarding process was completed a total of four times.To wash columns, 200 µL of Solvent A was added to each column. The columns were once again spun for three minutes at 2000 rpm and liquid was discarded afterwards; the solvent

addition, spinning, and discarding process was completed a total of three times.

To bind peptides to the columns, digested peptides were added to prepared columns, then the columns were spun at 3000 rpm for three minutes. The filtrate was pipetted back onto the column and spun again at 3000 rpm for three minutes. Solvent A (200 µL) was added to each column three separate times, then the column was spun for three minutes at 3000 rpm to wash salts off the column.

Peptides were eluted with two additions of 100 µL Solvent B to each column. Columns were spun at 3000 rpm for three minutes and the peptide fraction (filtrate) was reserved. Samples were placed in a speed vacuum at 4ºC until they were nearly dry (approximately two hours) to dry peptides. Peptides were reconstituted with 60 µL of 3% Acetonitrile + 0.1% Formic Acid, and stored at -80ºC.

Internal Standard Addition

Peptide Retention Time Calibration (PRTC; Pierce) is used as an internal standard to ensure consistency of peptides detected and measured throughout a mass spectrometry run. The stock solution of PRTC was diluted to 0.2 pmol/µL using 3% Acetonitrile with 0.1% Formic Acid. In a clean centrifuge tube, 6 µg of oyster protein and 0.376 pmol of PRTC were mixed together as per the PRTC user guide. Sample volume was brought up to 15 µL using a 3% acetonitrile and 0.1% formic acid solution. A quality control solution was also prepared (1 µL PRTC + BSA:3 µL 3% Acetonitrile and 0.1% Formic Acid solution).

1.3.3   Data Independent Acquisition Mass Spectrometry

Peptides were analyzed on an Orbitrap Fusion Lumos mass spectrometer (Thermo Scientific) using Data Independent Acquisition Mass Spectrometry (DIA). DIA analyses were completed as a comprehensive, non-random analytical method for detecting peptide ions present within a sample to create a peptide library. The peptide library was then leveraged to develop

a targeted proteomics assay for quantification (see *Selected Reaction Monitoring Assay*). A 30 cm analytical column and 3 cm pre-column were packed in-house with 3 µm C18 beads (Dr. Maisch). Samples were run in a randomized order. A blank injection followed each sample, with procedural blanks run at the very end. Every injection was 3 µL, which included 1 µg of oyster protein and 0.0752 pmol of PRTC. Peptides were analyzed in MS1 over the m/z range of 450-950 with 12 m/z wide windows with 5 m/z overlaps (Egertson et al., 2013). MS1 resolution was 60000 and AGC target was 400000 with a three second cycle time. The MS2 loop count was set to 20 and MS2 data was collected with a resolution of 15000 on charge state of 2 with an AGC target of 50000. No dynamic exclusion was used.

1.3.4 Peptide-Centric Proteomic Analyses

Unknown peptide spectra from mass spectrometry samples were matched with known peptides using Peptide-Centric Analysis in the PECAN software (Ting et al., 2015). Raw mass spectrometry files were converted to mzML files, then demultiplexed using MSConvert (Chambers et al., 2012). The C. gigas proteome was digested with in silico tryptic digest using Protein Digestion Simulator (Riviere et al., 2015). All known peptides from the mzML files were identified in comparison to the digested *C. gigas* proteome (Riviere et al., 2015).

The PECAN-generated spectral library (.blib) file was used to detect peptides of interest in raw DIA files in Skyline (MacLean et al., 2010). Skyline identified peptides using chromatogram peak picking, where ions that elute at the same time and mass are detected as a peptide (file available at Panorama Public). All PRTC peptides and approximately 100 different oyster proteins and their peptide transitions were manually checked for retention time and peak area ratio consistency to determine a Skyline auto peak picker error rate (24.3% $\pm$ 25%, range: 0% to 100%).

Proteins had to satisfy four criteria to be considered appropriate targets for the study. 1) After an extensive literature search, functions related to oxidative stress, hypoxia, heat shock,

immune resistance, shell formation, growth, and cellular maintenance were determined useful for evaluating environmental response. Proteins with annotations matching these functions were considered potential targets. 2) Protein data was then evaluated in Skyline to ensure there was no missing data for any peptide or sample. 3) Peaks with significant interference from other peptides were not considered. 4) Proteins needed at least two peptides with three transitions per peptide to quality as a potential target for downstream assays. The fifteen proteins (41 peptides and 123 transitions) that matched all of these criteria were selected as targets (Table 1.2).

1.3.5 Selected Reaction Monitoring Assay

Following the protein discovery phase (DIA), proteins were isolated as described above from an additional five randomly selected samples per site and habitat combination (for a total of 5 oysters per group) and analyzed with Selected Reaction Monitoring (SRM). Samples were prepared as described for DIA, except tissue samples were homogenized in 100 µL, and peptide samples were evaporated at 25ºC after peptide isolation. Proteins of interest identified from the DIA analysis were used as targets in a SRM assay following the workflow and informatic pipeline of (**Timmins-Schiffman2016?**). Target peptide transitions were monitored using SRM on a Thermo TSQ Vantage. SRM data were collected during a gradient of 2-60% acetonitrile over 40 minutes. All samples were run in technical duplicates in a randomized order with a 1 µg oyster peptide and 0.0752 pmol PRTC injection. A quality control injection and blank injection were run after every five sample injections, and PRTC peptides were monitored throughout the experiment.

Target Peptide Specificity

To ensure SRM assay specificity to oyster peptides of interest, oyster peptides were diluted in a background matrix of similar complexity (Pacific geoduck — *Panopea generosa* — peptides), then analyzed using the oyster SRM assay. An oyster-specific SRM target would

decrease in abundance with a decreasing abundance of oyster peptides in a mixture. Non-specific peptides — more likely to be found in background matrix of similar complexity — or peptides susceptible to interference would not correlate with oyster peptide abundance, and therefore, would be uninformative. Five *C. gigas* samples used for SRM were randomly selected and pooled in equal quantities. A ten-sample oyster:geoduck dilution series was prepared and run using the same methods as other SRM samples.

Target Analysis

Raw SRM files, a background *C. gigas* proteome, and the PECAN spectral library file from DIA were used to create a Skyline document (file available at Panorama Public). Correct transition peaks were selected based on predicted retention times from DIA results by comparing the relative retention times between identical PRTC peptides in the DIA and SRM datasets ($R^2 = 0.99431$). Based on peptide specificity analyses, heat shock protein 70 B2 and one constituent peptide of glucose-6-phosphate 1-dehydrogenase were removed from analyses.

Further filters were applied to the data to maintain only high quality peptides and transitions in the analysis. Coefficients of variation were calculated between technical replicates for each peptide transition. Peptides were removed from the dataset if CV > 20%. To maintain high sample quality, any sample missing data for more than 50% of peptide transitions was deemed poor quality for downstream analyses and excluded. Abundance data was normalized using total ion current (TIC) values from the mass spectrometer. Consistency between technical replicates was verified in remaining samples using a NMDS with TIC-normalized data and a euclidean dissimilarity matrix. Technical replicates were consistent if replicates lay closer together than to other samples in the NMDS. These replicates were then averaged for multivariate analytical methods.

Averaged technical replicate data was used to determine if peptides were differentially abundant between outplant sites and habitats. Before proceeding with analysis, peptide abundances were subjected to a Hellinger transformation to give low weights to any peptides with

low counts. A NMDS was used to visually compare relative peptide abundance. One-way ANOSIM tests by site, region (Puget Sound vs. Willapa Bay), and habitat, as well as a two-way ANOSIM test by site and habitat, were used to determine significant differences. Pairwise ANOSIM tests and post-hoc similarity percentage (SIMPER) analyses were conducted for each one- or two-way ANOSIM result significant at the 0.05 level. The first ten SIMPER entries were deemed influential peptides for each significant comparison.

The importance of environmental variables for explaining peptide abundance was evaluated with a redundancy analysis (RDA). For each site and habitat combination, mean and variance were calculated for pH, dissolved oxygen, salinity, and temperature over the course of the entire outplant. Environmental variables were then used as predictors to constrain peptide abundance. Predictors with missing values were not included. A triplot was used to visually assess differences in peptide abundance by site and habitat and the influence of individual peptides and environmental parameters. Analysis of Variance (ANOVA) was used to calculate significance of the RDA and environmental variables, with predictors deemed significant at $p < 0.05$. Since estuarine sites are highly variable, a second RDA was conducted constraining peptide abundance by environmental conditions on the day of collection to evaluate robustness of proteomic methods. R Scripts are available in the associated Github repository (Appendix 1).

1.4 Results

1.4.1 Conditions at Outplant Locations

Outplanted oysters experienced environmental variables representative of standard summer conditions in Puget Sound (PS) and WB. Wild or cultured oysters were present at the same tidal elevation as sensors, so outplant conditions represent experiences of non-experimental organisms. NMDS plots revealed mean environmental conditions were more similar among sites than variances. Both NMDS had stress values less than 0.20 and were considered ap-

16

propriate multivariate representations of environmental data (Mean NMDS: stress = 0.0170, p = 0.0100; Variance NMDS: stress = 0.0340, p = 0.0100). Mean dissolved oxygen and temperature were significantly different between sites (Table 1.4; ANOSIM; dissolved oxygen: R = 0.4063, p = 0.0530; temperature: R = 1.0000, p = 0.0020), but not between habitats. Variances of all environmental parameters were significantly different between sites (Table 5; ANOSIM; pH: R = 0.5313, p = 0.0180; dissolved oxygen: R = 0.6800, p = 0.0030; salinity: R = 0.8400, p = 0.0130; temperature: R = 0.9400, p = 0.0010). WB had the warmest, but least variable, temperature of 18.0ºC ± 1.3ºC compared to averages ranging from 14.8ºC ± 1.8ºC to 16.1ºC ± 1.7ºC in PS (Table 1.3; Figure 1.2). All variance values, mean pH, and mean salinity were not significantly different between habitats (Table 1.4; Table 1.5).

Although oyster position was standardized by tidal height, outplanted oysters experienced different amounts of exposure at low tide at each site due to differences in tidal amplitude (χ_4^2 = 25.29, p < 0.0001). Oysters at FB spent the highest percent of time emersed at tides less than 1 foot MLLW (13.99%; 188 hours), followed by PG (10.93%; 146 hours), SK (10.29%; 138 hours), CI (9.35%; 125 hours), and WB (8.48%; 113 hours). CI and FB outplants (p = 0.0021), FB and SK outplants (p = 0.0324), and FB and WB outplants (p < 0.0001) spent significantly different amounts of time emersed during low tide conditions.

1.4.2   Data Independent Acquisition Mass Spectrometry

Out of 39816 predicted proteins in the *C. gigas* FASTA proteome, 9047 proteins were detected in C. gigas across five sites and two habitats using DIA (Skyline auto peak picker error rate 24.3% ± 25%, range: 0% to 100%). Proteins detected included, but were not limited to, those annotated from processes such as responses to hypoxia and oxidative stress, removal of superoxide radicals, protein folding, muscle organ development, and negative regulation of apoptosis (Raw data: PeptideAtlas accession no. PASS01305).

1.4.3   Selected Reaction Monitoring Assay

Differential abundance of protein targets was evaluated at the peptide level after combining technical replicates. Proteins were considered differentially abundant if at least one monitored peptide was differentially abundant. There was no significant difference in SRM peptide abundance between unvegetated and eelgrass habitats across sites (One-way ANOSIM; R = 0.0440, p = 0.1220). Abundance data from both habitats were pooled for downstream analyses comparing protein abundances among sites (Raw data: PeptideAtlas accession no. PASS01304).

Rank distances between peptide abundances in multivariate space revealed differences in peptide abundance between WB versus the other four sites (NMDS: stress = 0.0750, p = 0.0099). One-way ANOSIM by site demonstrated no significant differences in peptide abundance (R = 0.0640, p = 0.065), but a one-way ANOSIM by region (PS vs. WB) revealed a trend towards peptide abundance differences (R = 0.2260, p = 0.0530). Peptide abundance was significantly different between FB and Willapa Bay (Table 1.6; Pairwise ANOSIM; R = 0.2568, p = 0.0350). Environmental variables explained 29% of variance in peptide abundance, but the proportion of variation explained was not significant (ANOVA for RDA; F6,19 = 1.3064; p = 0.1950). Temperature mean and variance were the two most influential environmental predictors of peptide abundance, but were not significant at the 0.05 level (ANOVA on RDA; Table 1.7; temperature mean: $F_{1,19}$ = 2.2375, p = 0.0650; temperature variance: $F_{1,19}$ = 2.1627, p = 0.0670). Peptides differentially abundant between FB and WB are primarily positively loaded onto temperature mean, with two negatively loaded on temperature variance. Robustness of analysis was evaluated by performing a second RDA to predict peptide abundance based on mean and variance of temperature and pH at time of collection. Due to missing values, dissolved oxygen and salinity could not be included in analysis. Temperature variance was the most influential predictor of peptide abundance at time of collection, but was not significant at the 0.05 level (ANOVA on RDA; $F_{1,13}$ = 2.2312, p = 0.0800).

Peroxiredoxin-5 (PRX), catalase (EC 1.11.1.6) (CAT), NAD(P) transhydrogenase (NADPt), glucose-6-phosphate 1-dehydrogenase (G6PD), carbonic anhydrase (CA), protein disulfide-isomerase 1 and 2 (PDI) had peptides that were identified as differentially abundant between Willapa Bay and Fidalgo Bay by SIMPER analysis (Figure 1.3). Fidalgo Bay peptide composition was differentiated by CA abundance, while all other significantly abundant proteins differentiated Willapa Bay. These proteins are involved in general cellular stress responses like reactive oxygen species neutralization or protein folding (Table 1.8). All differentially abundant proteins were detected at higher levels in the WB oysters than in oysters from the other four PS locations (CI, FB, PG, and SK), regardless of protein function (Figure 1.4). No differences in protein abundance were detected among the PS sites (Table 1.6).

1.5 Discussion

1.5.1 Antioxidant enzymes and Acid-Base Regulation

Higher antioxidant enzyme abundances can be a direct response to an increase in reactive oxygen species (ROS) (Y. Zhang et al., 2015; **Limón-Pacheco2009?**). During electron transport, oysters can produce ROS that induce oxidative stress if not neutralized (Abele, Philip, Gonzalez, & Puntarulo, 2007; **Limón-Pacheco2009?**). Peroxiredoxin-5 (PRX), and catalase (CAT) scavenge these ROS and degrade them before they cause cellular harm, while NAD(P) transhydrogenase (NADPt) maintains the cellular redox state (Flores-Nunes et al., 2015; Sussarellu et al., 2012; **Limón-Pacheco2009?**). Higher abundances of antioxidant enzymes in WB oysters suggests the need for ROS scavenging to acclimatize to local conditions.

Elevated antioxidant enzyme abundance at WB may be a response to higher levels of ROS brought on by heat stress. Mollusc species, like *C. gigas*, have been known to increase ROS production at higher temperatures (Abele, Philip, Gonzalez, & Puntarulo, 2007; Tomanek, 2014). Warmer and more variable temperature conditions at WB could be driving ROS

production and in turn, higher abundances of PRX, CAT, and NADPt to ameliorate ROS-associated stress (Y. Zhang et al., 2015). The shallow bathymetry of WB may have also contributed to elevated ROS presence and the need for antioxidant enzymes. At low tide, shallow waters would warm faster, leading to the higher temperatures observed at WB (Table 1.1). Warmer waters at low tide may prompt oysters to spend more time with their shells closed, inducing hypoxia and hypercapnia. Low oxygen concentrations within the shell could then augment ROS levels (Guzy & Schumaker, 2006). Oysters could respond through increased abundance of antioxidant enzymes (Sussarellu et al., 2012).

The need for WB outplants to regulate internal acid-base conditions is demonstrated by higher abundance of carbonic anhydrase (CA) in these oysters. *C. gigas* can regulate hemolymph pH by increasing HCO_3- concentration via the conversion of CO_2 to HCO_3-, catalyzed by CA (Michaelidis, Haas, & Grieshaber, 2005; Wei et al., 2015). If warmer water conditions at WB prompted shell closure, *C. gigas* would need to switch to anaerobic metabolism, inducing a significant reduction in hemolymph pH (Michaelidis, Haas, & Grieshaber, 2005). Buffering of hemolymph pH could be accomplished by elevated CA abundance. Although oysters at FB spent more time in low tide conditions that would also prompt shell closure and similar molecular responses, upregulated protein abundance at WB implies conditions at this site required a greater proteomic response in these specific biomarkers for acclimatization.

Higher abundance of glucose-6-phosphate 1-dehydrogenase (G6PD) in WB oysters indicates these oysters maintained metabolic activity in warmer temperature conditions. G6PD catalyzes the oxidative portion of the pentose phosphate pathway, and products from this pathway are often the precursors for nucleic and aromatic amino acids (Livingstone, n.d.). Additionally, G6PD activity generates NADPH and can indirectly regulate the redox environment and ameliorate effects of ROS (Livingstone, n.d.). For example, exposure to ROS-inducing pollutants led to increased G6PD activity in *C. gigas* ctenidia (Flores-Nunes et al., 2015). Increased abundance of G6PD in WB not only could have maintained transcrip-

tion and translation processes, but also levels of cellular and metabolic activity by indirectly dealing with ROS.

ROS are produced in response to many environmental changes, thus biomarkers of ROS scavenging are difficult to link to a single environmental difference in a variable and complex setting. For example, reduction of ROS was found to be a common response to both increased temperatures and aerial exposure in *C. gigas* (Y. Zhang et al., 2015). Since ROS mediation is a conserved response to several stressors, it is possible that environmental parameters we did not measure (e.g., contaminants, microbiota abundance, trace metals), or a combination of environmental parameters, could explain the observed variation in antioxidant enzyme abundance. Future work at these locations should take these variables into account.

1.5.2   Molecular Chaperones Involved in Protein Folding

Much like cellular response to ROS, increased levels of protein disulfide isomerase 1 and 2 (PDI) demonstrate a generalist response to conditions at WB. PDI is a general chaperone protein that forms disulfide bonds and assists with protein folding (Vargas-Albores, Martínez-Martínez, Aguilar-Campos, & Jiménez-Vega, 2009). Higher PDI abundance would reflect a need to refold misshapen proteins. In this regard, PDI would function similarly to molecular chaperones like heat shock proteins. Several invertebrate taxa demonstrate higher PDI abundance when faced with an immune challenge or metal contamination. When faced with an immune challenge, Pacific white shrimp (*Litopenaeus vannamei*) hemocytes synthesized higher abundances of immune response proteins, followed by elevated abundance of PDI to correct disulfide bonds in these proteins (Vargas-Albores, Martínez-Martínez, Aguilar-Campos, & Jiménez-Vega, 2009). An immune challenge could also lead to more misfolded proteins that PDI would need to refold to avoid cellular damage (P. Zhang et al., 2014). In flat oysters (*Ostrea edulis*), PDI expression increased in response to disseminated neoplasia (Silvestre et al., 2006; **Martín-Gómez2013?**). Metal contamination at WB could also elevate PDI abundance: Chinese mitten crabs (*Eriocheir sinensis*) chronically exposed to

cadmium over-expressed PDI (Silvestre et al., 2006; **Martín-Gómez2013?**). Since neither disease burden or environmental contamination was measured in this study, it is impossible to know if either triggered the PDI response. Examination of these factors in future studies may elucidate the specifics of elevated PDI abundance in WB.

1.5.3 Proteomic Responses in Puget Sound and Between Habitats

Due to the observed differences across environmental parameters between Willapa Bay and Puget Sound locations, similar abundances for proteins involved in various environmental responses may be evidence of physiological plasticity. One particular protein that we expected to be differentially abundant was heat shock 70 kDA protein 12A (HSP12A), since Willapa Bay had higher average temperatures (Table 1.3). However, this trend was not observed. Higher abundances of heat shock proteins (HSPs) are generally induced when organisms are exposed to thermal stress (Hamdoun, Cheney, & Cherr, 2003; Y. Zhang et al., 2015). In our experiment, average temperatures *C. gigas* experienced were lower than the 30°C threshold to induce elevated HSP expression in a controlled setting (Figure 1.2; Table 1.3), which would explain why we did not see elevated HSP12A expression in WB (Hamdoun, Cheney, & Cherr, 2003).

The lack of differential abundance for protein targets — both among Puget Sound sites and between unvegetated and eelgrass habitats — was unexpected. These similar proteomic profiles could be due to four factors. First, it is possible that a different suite of environmental response proteins in the SRM assay could have yielded a different view of acclimatization to the various outplant sites; however, the targets we chose have proven to yield insight into a range of environmental responses in previous studies (ex. Table 1.8). Second, outplant duration could have been too short to capture varied physiological response within Puget Sound, or oysters could have also acclimatized to conditions in their outplant locations. Third, post-translational modifications may have influenced how we detected proteins. Abundance between sites or habitats may have been similar, but addition of post-translational modifi-

cations may have varied. Finally, it is possible that the proteomic response was not different because the environmental conditions that would elicit up- or down-regulation of monitored proteins were similar across these five locations. For example, we found no differences in environmental conditions between eelgrass and unvegetated areas, nor any differences in protein responses between these habitats (Table 1.4; Table 1.5). Our results suggest that the potential vegetation-based systems of remediation could be limited under current conditions in the field. A recent comparison of stable isotopes and fatty acid composition in *C. gigas* at eelgrass and unvegetated habitats found no differences in 13C, 15N, total fatty acids, or proportional fatty composition in WB outplants, providing another line of evidence that suggests habitat may not affect *C. gigas* physiological performance (Lowe, Kobelt, Horwith, & Ruesink, 2018).

1.6 Conclusion

Differential abundance of target proteins observed between sibling oysters placed for 29 days in WB or PS indicates that environmental factors at WB tended to increase antioxidant enzyme and molecular chaperone protein abundance in Pacific oysters. This study is one the first to link regional environmental conditions to physiological responses in Puget Sound and Willapa Bay, as well as compare responses between Puget Sound and Willapa Bay. Understanding the difference between these two estuaries is important for the persistence of oyster reefs and aquaculture in the face of climate change. Of the environmental parameters measured, higher mean temperature, as well as less variable temperature conditions, at Willapa Bay may have influenced protein abundance. The lack of protein abundance differences between PS sites may imply that 2016 conditions were well-within the tolerances of outplanted *C. gigas*, so patterns of stress response witnessed under laboratory conditions were not apparent in the field. Together, the results generate a suite of new hypotheses for lab and field experiments comparing the effects of environmental conditions on physiological responses of marine invertebrates.

As global climate change continues to rapidly influence estuarine dynamics, this study illustrates the importance of investigating the effect of multiple modes of change on organismal physiology in situ. Our finding that oysters used generalist proteins to ameliorate stressors implies that assays for elucidating responses to multiple drivers in situ should include these conserved responses in addition to specific, stressor-related proteins identified in laboratory experiments. Pacific oysters are grown commercially throughout the Pacific Northwest region of the United States, so it is possible the population used for this study has a broad environmental tolerance. Proteomic assays allow for quantification of that tolerance, which is crucial for aquaculture industry and natural resource management. Future in situ explorations of environmental drivers on physiology should include a longer outplant duration, additional environmental variables, and multiple sampling points to provide a clearer connection between ecosystem dynamics and physiological performance.

1.7 Tables

Table 1.1: Latitude and longitude of *C. gigas* outplants, as well as time of day and tidal height at collection. Oysters were placed at five locations sites: Case Inlet (CI), Fidalgo Bay (FB), Port Gamble Bay (PG), Skokomish River Delta (SK), and Willapa Bay (WB).

Table 1.1: Outplant locations and time of day and tidal height at collection

Location	Latitude	Longitude	Time at Collection	Tidal Height MLLW feet at Collection
CI	47.35794	-122.7958	12:15	-1.8
FB	48.48169	-122.5835	12:12	-1.7
PG	47.84268	-122.5838	11:11	-1.6
SK	47.35523	-123.1572	11:51	-1.8
WB	46.49448	-124.0261	9:28	-1.7

Table 1.2: Proteins used as targets for a Selected Reaction Monitoring Assay (SRM). Targets were identified based on differential abundance and stress-related annotations. At least two peptides and six transitions were included in the assay for each protein. The protein Catalase had two isoforms under separate proteome IDs and target peptides were chosen from both IDs. A total of four peptides and twelve associated transitions were used as SRM targets.

Table 1.2: Target peptides used for Selected Reaction Monitoring Assay (SRM)

Protein	Proteome ID	Number of Peptides	Number of Transitions
3-ketoacyl-CoA thiolase	CHOYP_ACAA2.1.1|m.30666	3	9
Peroxiredoxin-5	CHOYP_BRAFLDRAFT_119799.1.1|m.23765	2	6
Thioredoxin reductase 3	CHOYP_BRAFLDRAFT_122807.1.1|m.3729	3	9
Protein phosphatase 1B	CHOYP_BRAFLDRAFT_275870.1.1|m.12895	3	9
Carbonic anyhdrase 2	CHOYP_CAH2.1.1|m.42306	3	9
Catalase 1	CHOYP_CATA.1.3|m.11120	3	9
Catalase 2	CHOYP_CATA.3.3|m.21642	1	3
Glucose-6-phosphate 1-dehydrogenase	CHOYP_G6PD.2.2|m.46923	3	9
Heat shock 70 kDa protein	CHOYP_HS12A.25.33|m.60352	2	6
Heat shock protein 70 B2	CHOYP_HSP74.1.1|m.13095	2	6
NAD(P) transhydrogenase	CHOYP_LOC100633041.1.1|m.35428	2	6
Glycogen phosphorylase	CHOYP_LOC100883864.1.1|m.41791	3	9
Multidrug resistance-associated protein	CHOYP_MRP1.5.10|m.34368	2	6
Protein disulfide-isomerase 1	CHOYP_PDIA1.1.1|m.5297	3	9
Protein disulfide-isomerase 2	CHOYP_PDIA3.1.1|m.60223	3	9
Puromysin-sensitive amirase	CHOYP_PSA.1.1|m.27259	3	9

Table 1.3: Environmental conditions (mean ± standard deviation) at outplant locations (Case Inlet (CI), Fidalgo Bay (FB), Port Gamble Bay (PG), Skokomish River Delta (SK), Willapa Bay (WB).

```
Warning in read.table(file = file, header = header, sep = sep, quote = quote, :
incomplete final line found by readTableHeader on 'data/Ch1/Table1.3.csv'
```

Table 1.3: Outplant conditions

Metric	CI	FB	PG	SK	WB
Mean Temperature (ºC)	16.1 ± 1.7	14.8 ± 1.8	15.1 ± 2.9	15.1± 2.2	18.0 ± 1.3
Mean Salinity (PSU)	24.47 ± 1.35	30.20 ± 0.38	23.25 ± 1.56	13.43 ± 1.04	27.28 ± 0.73
Mean Dissolved Oxygen Content (mg/L)	8.27 ± 1.91	9.63 ± 4.59	11.70 ± 3.70	9.99 ± 4.58	8.76 ± 1.78
Mean pH	7.63 ± 0.19	7.54 ± 0.23	7.33 ± 0.25	7.37 ± 0.24	7.56 ± 0.18

Table 1.4: Analysis of Similarities (ANOSIM) R statistic and p-value for mean values of environmental variables by site and habitat. Values significant at the 0.05 level are marked by a double asterisk, and values significant at the 0.10 level are marked by a single asterisk.

Table 1.4: ANOSIM for mean environmental conditions

Environmental Variable	Contrast	R statistic	p value
pH	Site	0.3281	0.135
Habitat	0.2125	0.151	
Dissolved oxygen	Site	0.4063 *	0.0530 *
Habitat	-0.14375	0.922	
Salinity	Site	-0.0133	0.482
Habitat	-0.1667	0.971	
Temperature	Site	1.0000 **	0.0020 **
Habitat	-0.192	0.905	

Table 1.5: Analysis of Similarities (ANOSIM) R statistic and p-value for environmental variable variances by site and habitat. Values significant at the 0.05 level are marked by a double asterisk, and values significant at the 0.10 level are marked by a single asterisk.

Table 1.5: ANOSIM for variance of environmental conditions

Environmental Variable	Contrast	R statistic	p value
pH	Site	0.5313 **	0.0180 **
Habitat	-0.1813	0.934	
Dissolved oxygen	Site	0.6800 **	0.0030 **
Habitat	-0.124	0.848	
Salinity	Site	0.8400 **	0.0130 **
Habitat	-0.125	0.708	
Temperature	Site	0.9400 **	0.0010 **
Habitat	-0.22	0.987	

Table 1.6: Results of pairwise ANOSIM tests between sites for protein abundance data. R values are listed above the diagonal, and p values are listed below the diagonal. Values significant at the 0.05 level are marked by a double asterisk, and values significant at the 0.10 level are marked by a single asterisk.

Table 1.6: Pairwise ANOSIM for protein abundance

X	CI	FB	PG	SK	FB 1
CI	-	0.0197	-0.0707	-0.0583	0.0926
FB	0.331	-	-0.0313	0.0949	0.2568
PG	0.78	0.5992	-	-0.0061	0.0727
SK	0.733	0.086	0.382	-	0.1547
WB	0.16	0.0350 **	0.192	0.0790 *	-

Table 1.7: Significance of each predictor in the RDA. Salinity mean and variance were not included due to missing values. Predictors significant at the 0.10 level are marked by a single asterisk.

Table 1.7: RDA Results

Environmental Variable	Df	Variance	F	Pr F
Temperature Mean	1	0.0014	2.3275 *	0.0650 *
Temperature Variance	1	0.0013	2.1627 *	0.0670 *
pH Mean	1	0.0003	0.5264	0.781
pH Variance	1	0.0007	1.2171	0.301
Dissolved Oxygen Mean	1	0.0006	1.0354	0.349
Dissolved Oxygen Variance	1	0.0003	0.5692	0.725
Residual	19	NA		

Table 1.8: Functions of differentially abundant proteins.

Table 1.8: Protein functions

Protein	Function
Peroxiredoxin-5 (PRX)	Scavenges ROS located in mitochondria using cysteine residuals to reduce substrates
Catalase (CAT)	Degrades hydrogen peroxide into water and oxygen
NAD(P) transhydrogenase (NADPt)	Involved in maintenance of cellular redox state
Carbonic anhydrase (CA)	Maintenance of acid-base balance by catalyzing CO_2 conversion to $HCO3-$
Glucose-6-phosphate 1-dehydrogenase (G6PD)	Catalyzes first step in the pentose phosphate pathway and controls rate of pathway's oxidative portion
Protein disulfide-isomerase 1 (PDI1) and Protein disulfide-isomerase 2 (PDI2)	General chaperone protein that aids in protein disulfide bond formation and isomerization

1.8 Figures

Figure 1.1: Map of outplant locations. Oysters were placed at five sites: Case Inlet (CI), Fidalgo Bay (FB), Port Gamble Bay (PG), Skokomish River Delta (SK), and Willapa Bay (WB). Sites in Puget Sound are outlined in a grey box.
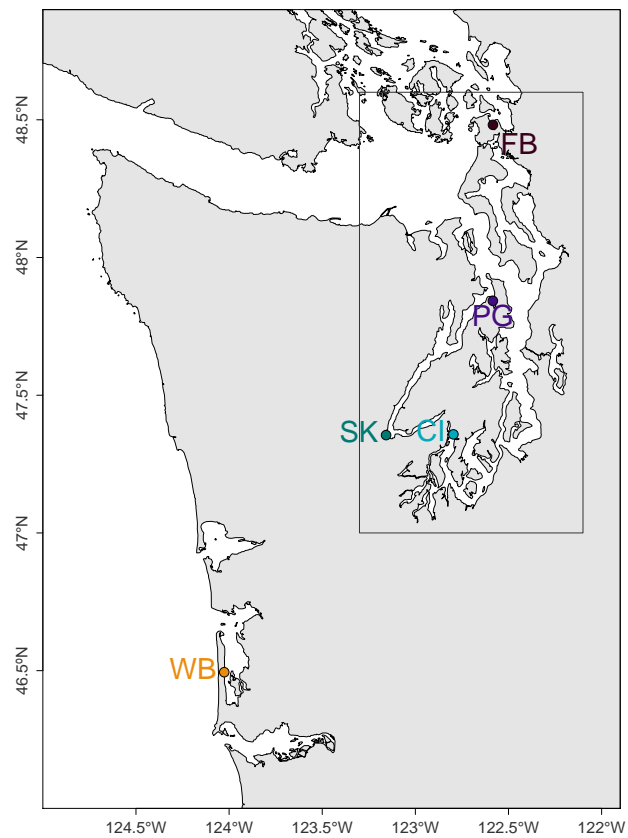
Figure 1.1: Map of outplant sites

Figure 1.2: Environmental variables for each site (Case Inlet, Fidalgo Bay, Port Gamble Bay, Skokomish River Delta, and Willapa Bay) and habitat (unvegetated and eelgrass) over the course of the 29 day outplant.
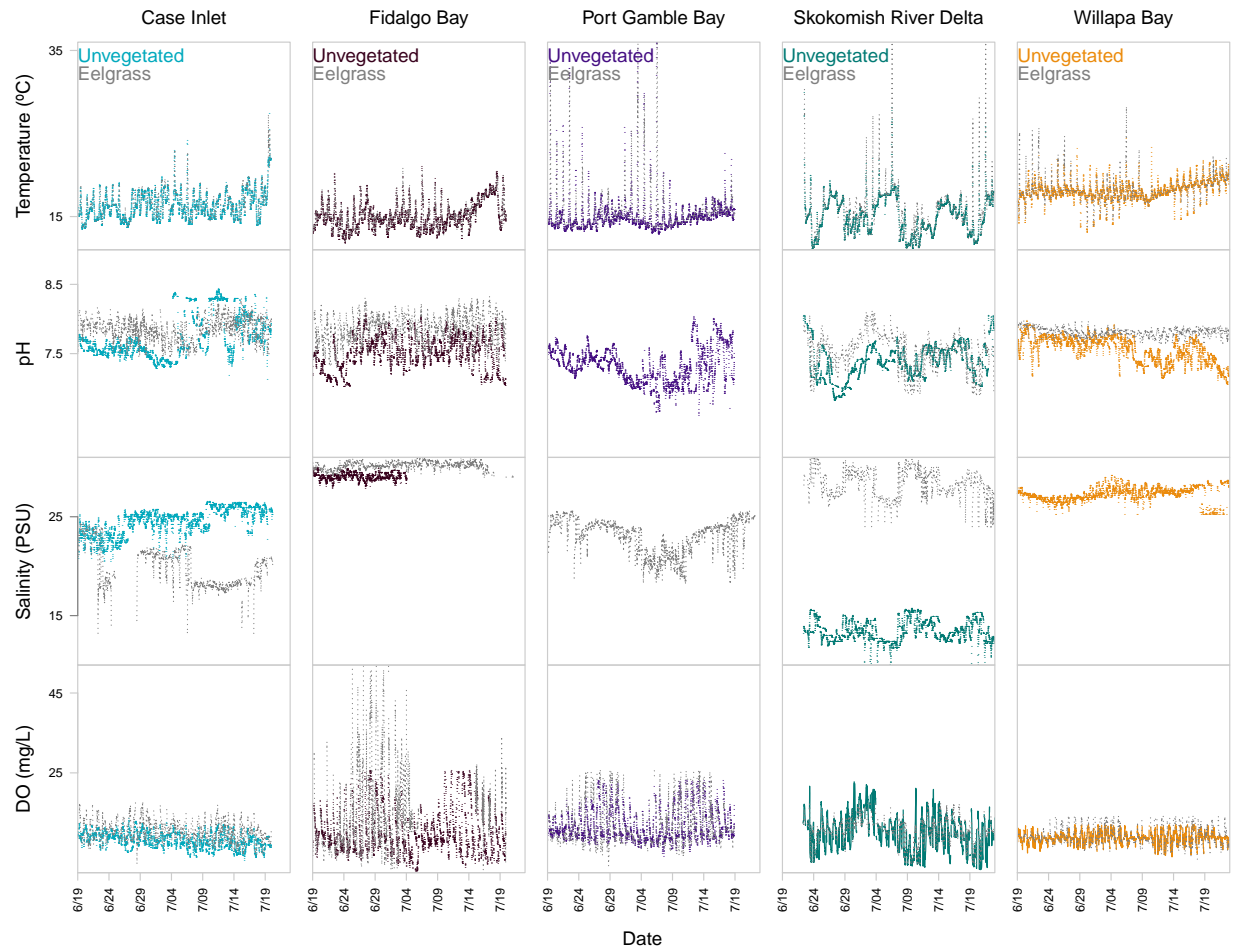
Figure 1.2: Environmental data at each site

Figure 1.3: Ordination results for peptide abundance and environmental data. a) Environmental variables (pH, dissolved oxygen, salinity, and temperature) explained 30% of variance in peptide abundance data; however, redundancy analysis (RDA) for peptide abundance constrained by environmental variable was not significant (ANOVA; $F_{6,19}$ = 1.306, p = 0.195). Temperature mean and variance were the most influential, yet nonsignificant, environmental predictors (ANOVA; temperature mean: $F_{1,25}$ = 2.3275, p = 0.065; temperature variance: $F_{1,25}$ = 2.1527, p = 0.067). Peptides differentially abundant between Fidalgo Bay (FB) and Willapa Bay (WB) are primarily positively loaded onto temperature mean, with two negatively loaded on temperature variance. b) Non-metric multidimensional scaling plot depicting peptide abundance by site and habitat, with 95% confidence ellipses around each site (stress = 0.075, p = 0.0099). c) Peptides that contributed to significant peptide abundance differences between FB and WB, as determined by post-hoc similarity percentage analysis (SIMPER). Peptides denoted 1 correspond to peroxiredoxin-5 (PRX), 2 for carbonic anhydrase (CA), 3 for catalase (CAT), 4 for glucose-6-phosphate 1-dehydrogenase (G6PD), 5 for NADP transhydrogenase (NADPt), and 6 for protein disulfide isomerases 1 and 2 (PDI). FB peptide composition was differentiated by CA abundance, while all other proteins influenced peptide abundances at WB.
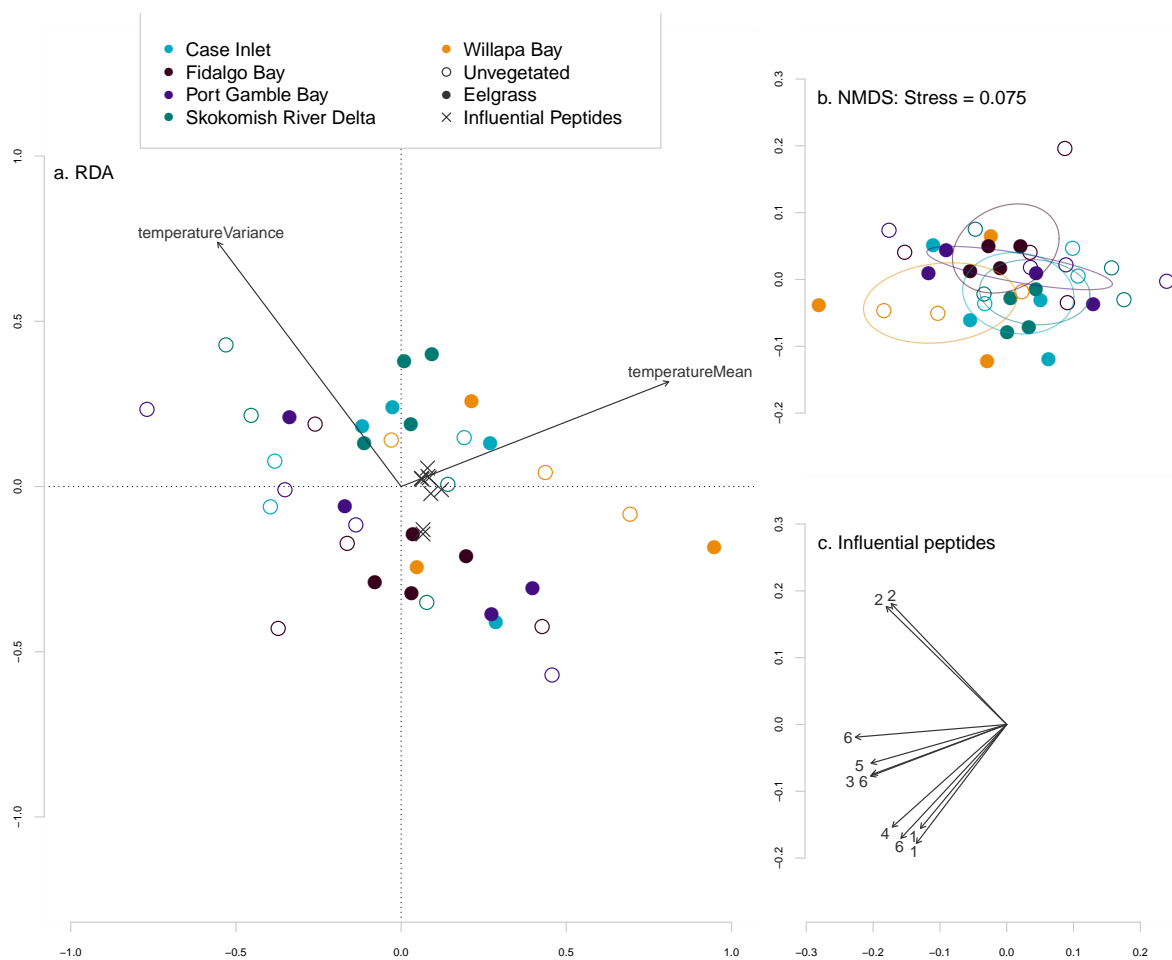
Figure 1.3: Peptide abundance ordination results

Figure 1.4: Average protein abundance by constituent peptides across experimental sites from SRM. Peptide abundance data at each site was averaged, then log transformed. Proteins were considered differentially abundant if at least one constituent peptide was significantly different (indicated by an asterisk). There were no significant differences in protein abundance among the Puget Sound locations, or between unvegetated and eelgrass habitats. Proteins were only significantly different between Fidalgo Bay and Willapa Bay or Skokomish River Delta and Willapa Bay.
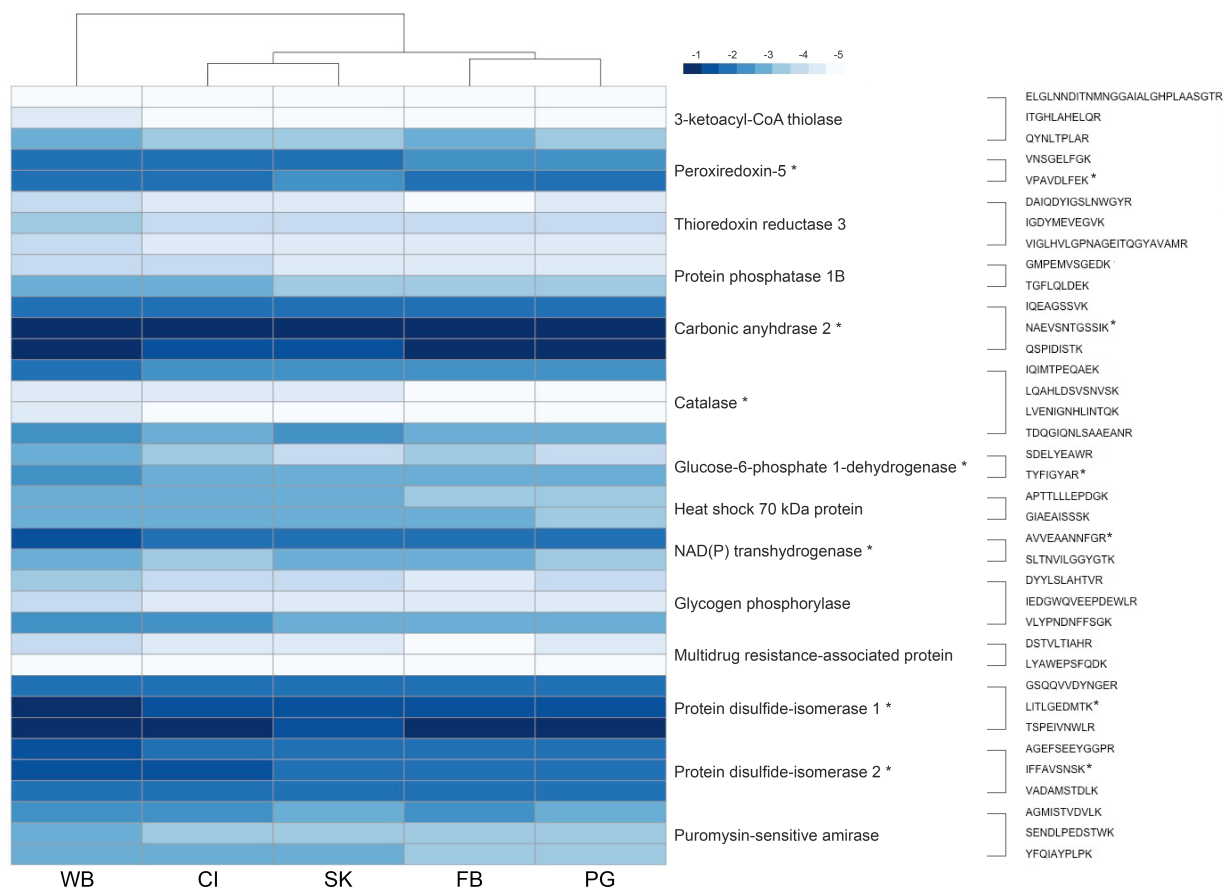
Figure 1.4: Average protein abundance by site

Chapter 2

LARVAL RESPONSE TO PARENTAL LOW PH EXPOSURE IN PACIFIC OYSTERS (*CRASSOSTREA GIGAS*)

2.1 Abstract

As negative effects of ocean acidification are experienced by coastal ecosystems, there is a growing trend to investigate the effect ocean acidification has on multiple generations. Parental exposure to ocean acidification has been shown to induce larval carryover effects, but whether or not acute exposure to a stressor as an adult can influence the larval generation long after the stress has been removed has yet to be tested. To assess how a temporary exposure to experimental ocean acidification affects the ecologically and commercially relevant Pacific oyster (*Crassostrea gigas*), adult oysters were exposed to either low pH (7.31 ± 0.02) or ambient pH (7.82 ± 0.02) conditions for seven weeks. Oysters were then held for eight weeks in ambient conditions, and subsequently reproductively conditioned for four weeks at ambient pH. After conditioning, oysters were strip-spawned to create four families based on maternal and paternal ocean acidification exposure. The number of D-hinge larvae were counted eighteen hours post-fertilization. A sex-specific broodstock response was observed, where female exposure to low pH conditions resulted in fewer D-hinge larvae. This study demonstrates that the effects of ocean acidification can last beyond the time from when the environmental perturbation is experienced. Broadening the understanding of environmental memory will be valuable when considering organismal ability to persist in the face of environmental change.

2.2 Introduction

Determining how parental exposure to ocean acidification carries over into early larval stages is important for understanding cumulative effects of climate-related environmental change. Gametogenesis is a key period during which parental exposure to ocean acidification can influence offspring (**Donelson2018?**). Several studies exposing Sydney rock oysters (*Saccostrea glomerata*) to high pCO₂ conditions (856 µatm, pHNBS 7.89-7.90) during reproductive conditioning identified positive larval carryover effects (**Parker2012?**; **Parker2015?**; **Parker2017?**). Specifically, larvae from parents exposed to low pH conditions were larger and developed faster in acidified conditions compared to those from parents reared in ambient pH conditions (**Parker2012?**; **Parker2015?**; **Parker2017?**). Conversely, similar experiments conducted with northern quahog (= hard clam; *Mercenaria mercenaria*) and bay scallops (*Argopecten irradians*) demonstrated negative larval carryover effects (**Griffith2017?**). Larvae from adult *A. irradians* and *M. mercenaria* exposed to low pH were more sensitive to acidified conditions than those spawned from parents exposed to ambient pH during reproductive conditioning (**Griffith2017?**). These studies demonstrate the importance of parental exposure during reproductive conditioning (late-stage gametogenesis) on offspring.

As the Pacific oyster (*Crassostrea gigas*; Thunberg, 1793) is a commercially and ecologically relevant species in much of the world, several research efforts have identified consequences of ocean acidification for distinct *C. gigas* life stages. While fertilization still occurs under near-future ocean acidification conditions (Kurihara, Kato, & Ishimatsu, 2007; **Havenhand2009?**; **Boulais2018?**), fertilization success in acidified conditions is variable between *C. gigas* populations (**Parker2010?**; **Barros2013?**). Researchers have found that larvae experience developmental delays and reduced shell growth when exposed to experimental ocean acidification conditions (Gazeau et al., 2011; Kurihara, Kato, & Ishimatsu, 2007; Timmins-Schiffman, O'Donnell, Friedman, & Roberts, 2013; Waldbusser et al., 2014). Natural upwelling-induced ocean acidification conditions also reduced larval production and

growth in a hatchery setting (**Barton2012?**). Ocean acidification hampers protein expression in larvae, especially for proteins related to calcification and cytoskeleton production (**Dineshram2012?**). During metamorphosis, oyster larvae experience down-regulation of proteins related to energy production, metabolism, and protein synthesis (Dineshram et al., 2016). Adult *C. gigas* calcification rates decrease as seawater pCO_2 increases (Gazeau et al., 2007), with oysters grown at 2800 µatm displaying significantly lower fracture toughness than oyster shells from ambient conditions (Timmins-Schiffman et al., 2014). Exposure to ocean acidification also affects adult antioxidant responses, carbohydrate metabolism, transcription, and translation protein pathways (Timmins-Schiffman et al., 2014). Predator-prey interactions can also change under experimental ocean acidification (**Wright2018?**). There is limited evidence, however, of how ocean acidification influences Pacific oysters across multiple generations.

The current study is the first to discern how exposure to experimental ocean acidification prior to reproductive conditioning affects larval abundance in *C. gigas*. This experiment not only describes how isolated exposure to low pH during early gametogenesis influences larvae, but also provides information on the effects of acute pH exposure on adult gonad morphology. Additionally, the study demonstrates how environmental perturbation experienced before reproductive maturity affects the subsequent generation, even if the stressor is long-removed.

2.3 Methods

2.3.1 Experimental overview

Experimental trials were conducted at the Kenneth K. Chew Center for Shellfish Research and Restoration at the National Oceanic and Atmospheric Administration (NOAA) Manchester Field Station (47°34'09.1"N 122°33'19.0"W, Manchester, Washington, USA) in 2017. Adult hatchery-raised *C. gigas* (average shell length $= 117.46 \pm 19.16$ cm) were acclimated in the facility for 10 days, then exposed to either low or ambient pH conditions for 48 days

(Figure 2.1). After pH exposure, oysters were held at ambient pH and water temperature conditions for 90 days. Oysters underwent reproductive conditioning for 22 days, then strip-spawned. D-hinge larvae were counted eighteen hours after fertilization occurred.

2.3.2 Experimental pH exposure

The experimental system consisted of a 1,610 liter storage tank that fed two 757 liter header tanks. Water from Clam Bay, WA was pumped through a sand filter, then UV-treated. The UV-treated water passed through a set of three sock filters (100 µm, 50 µm, and 25 µm) and a degassing column. Once degassed, water passed through three more sock filters (25 µm, then 10 µm, and 5 µm) before entering the storage tank. The storage tank was outfitted with an off-gas vent and pump to recirculate water such that CO_2 in the water could be equilibrated with atmospheric CO_2. Equilibrated water flowed into the two header tanks, each of which fed three 50L flow-through (1.2 L/min) experimental tanks (six experimental tanks total). For all header and experimental tanks, pH in header and experimental tanks was continuously monitored using Durafet pH probes (Honeywell Model 51453503-505) and an AVTECH system. The addition of CO_2 in the low pH header tank was controlled using a solenoid valve. A Dual Input Analytical Analyzer (Honeywell Model 50003691-501) automatically mediated solenoid injections. A CO_2 airline with a back pressure of 15 psi, controlled with a regulator, injected CO_2 into the low pH header tank every 180 seconds with an injection duration of 0.4 seconds. Injections only occurred if real-time pH from the Durafet was above pH 7.22. A venturi injector connected to the ambient water line mixed ambient pH water with CO_2-rich water to lower pH. There were no CO_2 injections in the ambient header tank.

Prior to the pH exposure trial, twenty randomly selected *C. gigas* were lethally sampled to assess gonadal status (see *Histological analysis*). Oysters were placed in each flow-through experimental tank in ambient water conditions and exposed to ambient or low pH conditions for seven weeks. Each treatment consisted of 3 tanks, each with 20 oysters. All experimental

tanks received algae from a common reservoir. The algal tank contained 300-500 mL of Shellfish Diet 1800® (Reed Mariculture) diluted in 200L of ambient pH seawater (**Helm2004?**). Algae was continuously dosed to oyster experimental tanks using an Iwaki Metering Pump. Algal lines were cleaned twice weekly, and experimental tanks were fully drained and cleaned once a week.

2.3.3   Seawater chemistry analysis

Twice a week, water samples (1L) were collected from each header and oyster experimental tank. For each sample, salinity was measured with a Bench/Portable Conductivity Meter (Model 23226-505, VWR), pH (mV) was measured with a Combination pH Electrode (Model 11278-220, Mettler Toledo), and temperature (ºC) was measured using a Traceable Digital Thermometer (Model 15-077, Fisher). To calibrate the pH probe, a Tris buffer (0.08 M, 28.0 salinity) was prepared using 0.3603 mol of NaCl (J.T. Baker), 0.0106 mol of KCl (Fisher Scientific), 0.0293 mol MgSO4-(H2O)7 (Fisher Scientific), 0.0107 mol of CaCl2-2(H2O) (MP Biomedicals), 0.0401 HCl (J.T. Baker), and 0.0799 mol of Tris base (Fisher Scientific). Deionized water was added for a final volume of 1L. Salinity, temperature, and pH measurements for the Tris buffer were obtained at five temperatures before measuring samples to generate a standard curve. This standard curve was used to calibrate the pH electrode and convert measured millivolts to pH units.

For total alkalinity measurements, duplicate seawater samples (250 mL) were collected from experimental tanks twice weekly and dosed with mercuric chloride (50 µL of 0.18 M solution) to preserve samples (**Bandstra2006?**). Samples from days 5, 33, and 48 were run on a T5 Excellence titrator (Mettler Toledo) to determine alkalinity. Salinity from discrete samples was used to calculate total alkalinity, using the `seacarb` library in R (**Gattuso2018?**). Calculated pH, total alkalinity, temperature, and salinity were also used in `seacarb` to calculate in situ pH, pCO_2, dissolved organic carbon (DIC), calcite saturation (Ω_{calcite}), and aragonite saturation ($\Omega_{\text{aragonite}}$) for days 5, 33, and 48. R code used to calculate water

chemistry parameters is available in the associated Github repository (Appendix 1).

2.3.4 Histological analysis

Twenty randomly selected *C. gigas* were lethally sampled before pH exposure for histological analyses. On the last day of low pH exposure, ten oysters from each treatment — randomly selected from each tank — were also lethally sampled to assess gonadal status. For each sampled oyster, a piece of gonad tissue was cut and placed in a histology cassette. Gonad tissue in cassettes was fixed for histology using PAXgene Tissue FIX and STABILIZER and sent to Diagnostic Pathology Medical Group, Inc. (Sacramento, CA) for staining with hematoxylin and eosin and slide preparation. Tissues exposed to ambient pH were confounded during processing, preventing any tank identification. Maturation state and organism sex was evaluated histologically at 40x magnification (**Fabioux2005?**; **Enríquez-Díaz2008?**).

2.3.5 Reproductive conditioning

Following seven weeks of low pH exposure, oysters were returned to a common garden and maintained at ambient pH conditions for eight weeks. Afterwards, oysters were reproductively conditioned. Water temperatures and food quantity are known to regulate the timing, speed, and intensity of gametogenesis in *C. gigas* (**Enríquez-Díaz2008?**). Conditioning protocol was modeled after standard hatchery practices (Molly Jackson, Broodstock Manager at Taylor Shellfish, pers. comm., June 2017). Water temperature was raised from ambient conditions (13ºC) to 23ºC over three weeks (1ºC/2 days), since optimal temperature for *C. gigas* gametogenesis is between 18ºC and 26ºC (**Parker2010?**). Conditions were maintained at 23ºC for one more week prior to spawning. During conditioning, *C. gigas* were fed 700-800 mL of Shellfish Diet 1800® daily (**Helm2004?**).

2.3.6   Strip spawning and larval rearing

After reproductive conditioning, all surviving oysters were prepared for strip spawning. A sample of gonad from each individual was assessed for presence of active sperm or eggs using a microscope at 10x magnification. Only *C. gigas* with active sperm or eggs were used for crosses (n~male, low~ = 6, n~female, low~ = 22, n~male, ambient~ = 6, n~female, ambient~ = 26). Presence of mature gametes and ripe oysters indicated that oysters were in good condition and not affected by use of Shellfish Diet 1800® instead of live algae during reproductive conditioning. For each treatment (low pH and ambient conditions), one gram of mature gonad from each ripe female was pooled. The number of eggs in both the ambient and low pH pools were counted to determine the number of eggs used for parental crosses. Parental crosses were created using 210,000 eggs from the female egg pools and sperm (200 µL) from individual males.

Four half-sibling families were created based on parental pH exposure: low pH female (pool) x low pH male, low pH female (pool) x ambient pH male, ambient pH female (pool) x low pH male, and ambient pH (pool) female x ambient pH male. These crosses were conducted using pooled eggs from either low pH or ambient pH females, and sperm from one of six males within each pH treatment (e.g. low pH female pool x low pH male-01, low pH female pool x low pH male-02, … low pH female pool x low pH male-06), totaling 24 crosses. All crosses were performed in duplicate, resulting in 48 separate fertilization events.

Fertilization was carried out in plastic beakers (1L) for 20 minutes with static 23ºC filtered seawater (1 µm) in ambient pH conditions. After confirming polar body formation, beaker contents were transferred to larger plastic tanks (19L) with aerated, static 23ºC filtered seawater (1 µm) for eighteen hours of incubation. Duplicate containers were combined eighteen hours post-fertilization, and D-hinge larvae were counted for each cross (n= 24).

2.3.7 Statistical analyses

Differences in in situ pH, total alkalinity, pCO$_2$, DIC, $\Omega_{\mathrm{calcite}}$, and $\Omega_{\mathrm{aragonite}}$ between pH treatments were evaluated with a one-way ANOVA. Because tissue samples were confounded during histological processing, a binomial GLM model was used to compare gonad maturation between pH treatments. Differences in sex ratios between pH treatments were evaluated using a chi-squared test of homogeneity. To identify differences in D-hinge larval counts, a linear mixed model was used, with sire and female egg pool as random effects. Differences in D-hinge larval counts by female treatment were assessed using a similar linear mixed model, with only sire as a random effect. Normality of data, as well as independence and homoscedasticity, were verified visually. All statistical analyses were carried out in R (Version 3.4.0). R Scripts are available in the associated Github repository (Appendix 1).

2.4 Results

2.4.1 Water chemistry

Pacific oysters exposed to low pH experienced different water chemistry parameters than those in the ambient pH treatment (Table 2.1). Using water samples from days 5, 33, and 48, pH (One-way ANOVA; F1, 16 = 5838.7810, p = 6.1165e-22), pCO$_2$ (One-way ANOVA; F~1, 16~ = 235.4018, p = 5.4421e-11), DIC (One-way ANOVA; F~1, 16~ = 7.1222, p = 0.0168), $\Omega_{\mathrm{calcite}}$ (One-way ANOVA; F~1, 16~ = 528.9468, p = 1.0989e-13), $\Omega_{\mathrm{aragonite}}$ (One-way ANOVA; F~1, 16~ = 526.5207, p =1.1389e-13) were significantly lower in the low pH treatment. Total alkalinity, however, was not significantly different between pH treatments (One-way ANOVA; F~1, 16~ = F = 1.382, p = 0.2570).

2.4.2   Gonad maturation

A binomial GLM was used to compare gonad maturation of individuals sampled before and immediately after pH exposure, but before reproductive conditioning. The most parsimonious model included only sampling time (before or after pH treatment). Gonad maturation status was not significantly different between *C. gigas* sampled before and after pH treatment (binomial GLM; F~2, 37~ = 0.7973, p = 0.3442). Additionally, maturation status was not different between pH treatments (binomial GLM; F~3, 36~ = 2.2675, p = 0.1408). No sampled oysters possessed fully mature gametes, but some males sampled appeared to be undergoing resorption (Appendix 1). Sex ratios were also similar between low and ambient pH treatments (Chi-squared test for homogeneity; X^2_2 = 3.2279; p = 0.1942).

2.4.3   Larval Survival

A linear mixed effect model, with female pool and sire as random effects, demonstrated no significant difference in the number of D-hinge larvae counted eighteen hours post-fertilization between all four parental families (Linear mixed effect model; X^2_3 = 3.1325; p = 0.1066). Sire and female egg pools accounted for 0.8530% and 3.1623% of total variance, respectively. Significantly fewer D-hinge larvae were present in half-sibling families where females were exposed to low pH conditions (Figure 2.2; Linear mixed effect model; X^2_1 = 8.1781; p = 0.0042), with sire accounting for 0.3116053% of total variance.

2.5   Discussion

The present study is the first to document the transgenerational influence of ocean acidification on Pacific oysters. Larval *C. gigas* was negatively impacted when maternal broodstock were exposed to low pH (pH = 7.31), suggesting a maternal carryover effect. The experimental design of this study is also unique — adult *C. gigas* experienced low pH conditions three months prior to reproductive conditioning, then were kept solely in ambient pH conditions

through strip spawning and larval rearing. Since environmental perturbation experienced before Pacific oysters were mature still affected larval oysters, the results indicate a role for environmental memory in *C. gigas* response to ocean acidification. Mechanisms for transgenerational environmental memory have been explored in response to acute stressors in other species. For example, *Daphnia magna* exposed to high salinity conditions had altered DNA methylation patterns, and these patterns were inherited by the following three non-exposed generations (**Jeremias2018?**). Significant carryover effects observed in *C. gigas* — solely exposed to low pH when immature — broaden the current understanding of stressor timing and its effect on organismal physiology.

While it is evident that acute exposure to low pH experienced by adult *C. gigas* resulted in detrimental effects for larvae, the fact that larvae were not reared in acidified conditions makes cross-study comparison difficult. If *C. gigas* larvae were also reared in acidified conditions, it is possible that larvae with a history of parental exposure to experimental ocean acidification may have exhibited a negative carryover effect on larval growth and performance. Negative carryover effects have been found in other marine invertebrate taxa, but all studies involved exposure to experimental ocean acidification during reproductive conditioning and larval rearing in acidified conditions. Tanner crabs (*Chionoecetes bairdi*) solely exposed to acidified water (pH 7.5 or 7.8) as larvae did not exhibit significant changes in morphology, size, Ca/Mg content, or metabolic rate, yet substantial effects on physiology was observed when larvae had a history of maternal exposure during oogenesis (**Long2016?**). Larvae from adult northern quahog (= hard clam; *M. mercenaria*) and bay scallops (*A. irradians*) developed slower when parents were reproductively conditioned in low pH conditions (pH_T = 7.4) (**Griffith2017?**). Additionally, larvae with a history of parental low pH exposure were more vulnerable to additional stressors like thermal stress, limited food, and harmful algae exposure (**Griffith2017?**). Although *C. gigas* were not reproductively conditioned in acidified water, and the present study cannot distinguish between hatching success and early mortality, identifying a similar negative larval carryover effect four months after an

acute environmental perturbation is arguably more surprising and significant, particularly in terms of efforts to understand the mechanism of environmental memory.

The severity of conditions experienced by organisms may also explain whether or not offspring demonstrate transgenerational acclimatization to stressors. For example, the negative carryover effect observed in *C. gigas* is different from the positive carryover effects observed in ocean acidification experiments conducted with Sydney rock oysters. When adult *S. glomerata* were exposed to acidified seawater (pCO_2 = 856 µatm; pH_{NBS} = 7.89-7.90) during reproductive conditioning, resultant larvae were larger and developed faster in acidified conditions when compared to larvae from parents exposed to ambient conditions (**Parker2012?**). This positive carryover effect was found to persist in the F2 generation. In acidified conditions, F2 offspring with a history of transgenerational (F0 and F1) pCO_2 exposure grew faster and demonstrated fewer shell abnormalities (**Parker2015?**). While species-specific responses can certainly explain the observed differences in larval phenotypes, it is also likely that inconsistencies in treatment conditions between experiments resulted in dose-dependent effects. (**Parker2012?**), (**Parker2015?**), and (**Parker2017?**) used a high pCO_2 treatment of 856 µatm (pH = 7.89-7.90), with a control of 380-385 µatm (pH = 8.19-8.20). Therefore, the elevated pCO_2 treatment used in those studies is similar to the ambient pH treatment (7.82; pCO_2 = 747.51-912.22) in the present study. Sydney rock oyster larvae with a history of transgenerational exposure exhibited faster development, but exhibited similar survival and were only 10% larger in acidified conditions when compared to larvae with no transgenerational exposure history (**Parker2012?**). With a relatively smaller effect size and a milder treatment than used in this study, it is possible these studies are not at odds but reflect dose-dependent effects on larval phenotypes. Negative carryover effects demonstrated in this study and in (**Griffith2017?**)] can also be attributed to similar treatment pH levels (**Griffith2017?** pHT = 7.4, this study: pH = 7.31). Both of these studies used treatment levels more extreme than International Panel on Climate Change projections for open ocean acidification, but consistent with coastal and estuarine acidification scenarios

experienced at study locations (Feely et al., 2010; Pelletier, Roberts, Keyzers, & Alin, 2018; **Griffith2017?**). More research is required to understand how location-specific conditions will affect multiple generations in a single species.

Although the effect of water chemistry on gametogenesis has been recorded in other taxa, it is unlikely that a low pH exposure occurring three months prior to reproductive conditioning could have affected gonad maturation. Studies in which reproductive conditioning and experimental ocean acidification occur concurrently have demonstrated negative effects on maturation and fecundity. Gametogenesis, especially oogenesis, was disrupted in eastern oysters (*Crassostrea virginica*) that experienced severe ocean acidification conditions during reproductive conditioning (pH = 7.71, 5584 µatm) (**Boulais2017?**). Green sea urchins (*Stronglyocentrotus droebachiensis*) exposed to high pCO_2 (1200 µatm) conditions for four months demonstrated low fecundity (**Dupont2013?**), and *S. glomerata* conditioned in high pCO_2 (856 µatm) conditions exhibited reduced rates of gametogenesis, smaller gonad area, and reduced fecundity (**Parker2018?**). Gonad histology from *C. gigas* taken immediately after low or ambient pH exposure did not indicate any differences in maturation state, or interaction between sex and maturation state, between treatments. Even if fecundity or rates of gametogenesis differed between treatments, a return to ambient conditions for three months may have reversed any detrimental effects.

Reduced *C. gigas* larval abundance could have been a result of altered maternal provisioning in female oysters exposed to low pH conditions. In the face of stressors, females can either increase maternal provisioning (**Allen2008?**; **Sunday2011?**) — diverting more resources, like lipids or proteins, into eggs — or decrease provisioning due to energetic constraints (**Liu2010?**; **Uthicke2013?**). For example, changes in fatty acid provisioning from maternal exposure to high pCO_2 conditions (2300 µatm) in Atlantic silverside (*Menidia menidia*) resulted in lower embryo survival when eggs lacked certain fatty acids (**Snyder2018?**). This phenomenon, however, was not documented in the Sydney rock oyster: while elevated pCO_2 conditions (856 µatm) reduced the amount of energy invested in maternal gonads, these

conditions did not impact *S. glomerata* egg size or total lipid content (**Parker2018?**). Since adult *C. gigas* did not experience environmental perturbation after low pH exposure, and received enough food to spawn well, any impact on maternal provisioning and subsequent larval abundance was likely a result of low pH three months prior to reproductive conditioning.

The documented effect on Pacific oyster larval abundance four months after low pH exposure indicates an important role for environmental memory in *C. gigas* response to ocean acidification. Low pH exposure may have induced epigenetic modifications (eg. changes in DNA methylation) in adult *C. gigas*. Studies of finfish and shellfish aquaculture species have demonstrated environmentally-induced epigenetic modifications that modify phenotypic responses in organisms (Gavery & Roberts, 2017). One notable study on *C. gigas* examined parental effects of adult pollutant exposure on offspring (**Rondon2017?**). Spat from parents exposed to the herbicide diuron had differential methylation in coding regions, with some changes leading to differential gene expression (**Rondon2017?**). This research indicates that a mechanism crucial for phenotypic plasticity and acclimation across generations exists, and this knowledge can be analyzed in the context of climate-related environmental stressors. Epigenetic modifications in response to ocean acidification have been documented in coral species (**Putnam2016?**), but not in molluscs. Several experimental ocean acidification studies, however, hint at the role of epigenetic memory. Olympia oysters (*Ostrea lurida*) exposed to high pCO_2 (1000 µatm) conditions still grew less in the juvenile life stage than counterparts reared in ambient pCO_2, even after the stressor had been removed (**Hettinger2013?**). Similarly, transgenerational acclimation of *S. glomerata* larvae with a history of exposure to acidified conditions could be explained by changes in epigenome that affect organismal performance (**Parker2012?**; **Parker2015?**; **Parker2017?**). Methylation levels are known to increase over the course of gametogenesis, with male and female *C. gigas* exhibiting significantly different methylation patterns (**Zhang2018?**). If epigenetic modifications were acquired by female oysters during low pH exposure, it could explain why a significant effect on larval abundance was detected four months after the exposure ended.

Epigenetic mechanisms and altered maternal provisioning are not necessarily mutually exclusive — changes in the methylome could influence maternal provisioning — and both could contribute to the results observed in this study.

The results of this study emphasize the need to broaden the scope of when environmental perturbation experienced by an organism is considered stressful, and when an effect can be detected. Although there was no observable effect on adult gonad maturation right after low pH exposure, significant differences in larval abundance were detected four months after the exposure ended. Stressor timing and duration can impact transgenerational responses between mature parents and offspring (**Donelson2018?**). While experimental ocean acidification (pH 7.7; $pCO_2 = 800$ µatm) increased female investment in amphipods (*Gammarus locusta*), the subsequent generation exhibited fewer eggs and lower fecundity in the same conditions (**Borges2018?**). Transgenerational benefits of maternal exposure to different temperatures (17ºC or 21ºC) in threespine stickleback (*Gasterosteus aculeatus*) differed based on exposure duration (**Shama2014?**). Grandparents (F0) were only exposed to treatment temperatures during reproductive conditioning, while parents (F1) experienced either temperature over the course of development. The F1 generation exhibited temperature tolerances similar to the F0 maternal rearing environment, but the F2 generation tolerance was more similar to the F0 generation than the F1 generation (**Shama2014?**). The present study demonstrates that length and timing of environmental perturbation experienced by immature individuals can still affect offspring. (**Massamba-N?**)'Siala2014 elucidated a similar phenomenon with marine polychaetes (*Ophryotocha labronica*): offspring experienced positive carryover effects of female exposure to temperature conditioning only when mothers were exposed to these conditions during late oogenesis; exposure during early oogenesis lead to negative carryover effects. More research should be conducted to understand how stressor timing, specifically before reproductive maturity, can impact carryover effects.

Most other experiments investigating stressor timing are conducted in a multiple stressor framework (**Gunderson2016?**). For example, elevated temperatures and low salinity had

synergistic effects on *O. lurida* when they were co-occurring stressors, but two to four weeks of recovery in between stressors negated these effects (**Bible2017?**). Incorporating recovery time in a single-stressor experimental design is also crucial for accurately understanding how environmental perturbation impacts organism physiology. Exposure at one point in time may elicit a response much later in time, in a different environmental setting, or in a different generation, as evidenced by the present study and (**Hettinger2013?**). The experimental design in the present study is unique, featuring a significant recovery time between low pH exposure and spawning. More single-stressor experiments should incorporate lag times between exposure to stress and measuring response variables to understand if these responses change over time. Adding a multigenerational component to such experiments can elucidate if acute exposures generate carryover effects.

Significant decreases in larval abundance four months after broodstock were exposed to acidified seawater has implications for both aquaculture and natural *C. gigas* populations. Parents and offspring — or even different offspring life stages — may not experience the same environmental chemistry. For example, upwelling conditions affecting adult *C. gigas* may subside once spawning occurs. Long-term monitoring of wild Pacific oyster populations, with detailed environmental chemistry reporting, will be crucial for understanding how brief exposures to adverse conditions affect reproductive success and larval abundance in the field. Responses to stressors should not only be documented during and after the perturbation occurs, but also for an extended time afterward. Hatchery-reared *C. gigas* larvae can also experience different conditions than broodstock. Facilities unable to control water chemistry conditions may be exposing immature individuals to environmental perturbations that could affect larvae once spawned. The success of "priming" — exposing *C. gigas* to stressful conditions to induce environmental memory and increase fitness — hinges on the identification of "programming windows" (Gavery & Roberts, 2017). The present study shows that the period of time before reproductive conditioning can be important for transferring environmental memory, although only negative carryover effects have been demonstrated in

C. gigas.

2.6   Conclusion

Four months after adult *C. gigas* experienced experimental ocean acidification, larval abundance of female oysters exposed to low pH was significantly lower than those exposed to ambient pH eighteen hours post-fertilization. Not only did this experiment elucidate intergenerational effects of ocean acidification on the Pacific oyster, but it also demonstrated a need to consider the timing of altered environmental conditions on organismal physiology. Although adult oysters experienced a low pH stressor prior to reproductive conditioning, larval abundance was still significantly affected. Therefore, conditions experienced by aquaculture broodstock before reproductive conditioning should be taken into consideration. Likewise, these results should be considered when modeling large-scale ecosystem responses to ocean change. Future work on multigenerational responses to ocean acidification should investigate how exposure to adverse conditions while an organism is immature can affect reproductive success and offspring fitness. The significant lag time between the end of the low pH exposure and spawning possibly indicates some form of epigenetic "memory." Additional research is needed to investigate the degree of environmental memory that can be maintained and the contributing epigenetic phenomenon.

2.7 Tables

Table 2.1: Average ($\pm$ SE) pH, total alkalinity (μmol/kg), pCO$_2$ (μatm), dissolved organic carbon (DIC; μmol/kg), calcite saturation state ($\Omega_{calcite}$), and aragonite saturation state ($\Omega_{aragonite}$) for three timepoints during low pH exposure (Day). The 'seacarb' library in R was used to calculate total alkalinity, and in situ pCO$_2$, Dissolved Inorganic Carbon (DIC), calcite saturation ($\Omega_{calcite}$), and aragonite saturation ($\Omega_{aragonite}$) for each oyster tank. Averages for both control (ambient pH) and experimental (low pH) values were calculated from three replicate tanks each. Between all three days, pH (One-way ANOVA; F$_{1, 16}$ = 5838.7810, p = 6.1165e-22), pCO$_2$ (One-way ANOVA; F$_{1, 16}$ = 235.4018, p = 5.4421e-11), DIC (One-way ANOVA; F$_{1, 16}$ = 7.1222, p = 0.0168), $\Omega_{calcite}$ (One-way ANOVA; F$_{1, 16}$ = 528.9468, p = 1.0989e-13), $\Omega_{aragonite}$ (One-way ANOVA; F$_{1, 16}$ = 526.5207, p =1.1389e-13) were significantly lower experimental treatment. Total alkalinity, however, was not significantly different between treatments (One-way ANOVA; F$_{1, 16}$ = 1.382, p = 0.2570).

Day	C pH	E pH	C TA	E TA	C pCO2	E pCO2	C DIC	E DIC	C Calcite	E Calcite	C Aragonite	E Aragonite
5	7.82 ± 0.004	7.33 ± 0.002	2307.41 ± 25.45	2332.36 ± 31.05	747.51 ± 13.94	2481.23 ± 29.83	2233.41 ± 25.29	2408.51 ± 31.76	1.86 ± 0.02	0.62 ± 0.01	1.16 ± 0.012	0.58 ± 0.007
33	7.81 ± 0.005	7.31 ± 0.004	2747.00 ± 21.13	2917.60 ± 18.36	912.22 ± 12.69	3309.52 ± 7.22	2664.57 ± 19.99	3020.99 ± 17.99	2.23 ± 0.03	0.77 ± 0.02	1.40 ± 0.020	0.48 ± 0.014
48	7.82 ± 0.015	7.29 ± 0.004	2611.40 ± 31.01	2808.39 ± 12.24	863.47 ± 42.42	3343.89 ± 49.49	2533.28 ± 35.45	2920.52 ± 15.11	2.13 ± 0.06	0.68 ± 0.01	1.32 ± 0.035	0.42 ± 0.004

Table 2.1: Carbonate chemistry parameters for control (C) and experimental (E) tanks

2.8 Figures

Figure 2.1: Pacific oysters (n = 140) were acclimated for 15 days, then twenty were randomly sampled for histological analyses. Remaining oysters were divided into ambient pH or low pH treatments for seven weeks. Three experimental tanks for each treatment were used with 20 oysters per tank for a total of 60 oysters per treatment. At the end of the pH exposure, a total of ten oysters were randomly selected from each treatment and sampled for histological analyses. All remaining oysters were then held in ambient pH conditions for 3 months. Finally, oysters were reproductively conditioned and strip spawned. Larvae were counted eighteen hours post-fertilization.
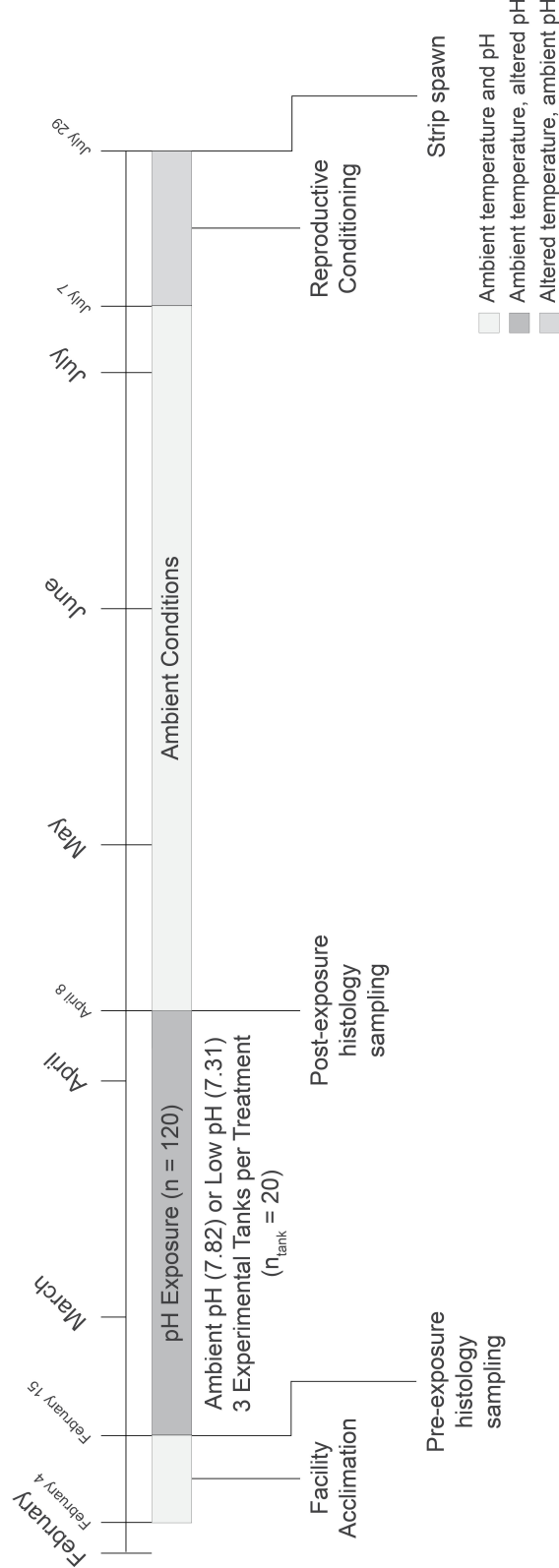
Figure 2.1: Experimental timeline

Figure 2.2: Proportion of live D-hinge larvae eighteen hours post-fertilization by female treatment. Each box represents proportions of live larvae between the first and third quartiles for half-sibling families where the female was exposed to either ambient or low pH conditions. Horizontal lines outside the box indicate the minimum value before the lower fence and the maximum value before the upper fence, with the solid line indicating the median. Circles represent outliers. A proportion of 1.0 indicates that all eggs in a cross were successfully fertilized and developed into D-hinge larvae. A linear mixed model, with sire as a random effect, indicated significantly fewer D-hinge larvae were present in half-sibling families where females were exposed to low pH conditions (t = -2.999; p = 0.0119). Significantly different proportions are indicated by a letter.
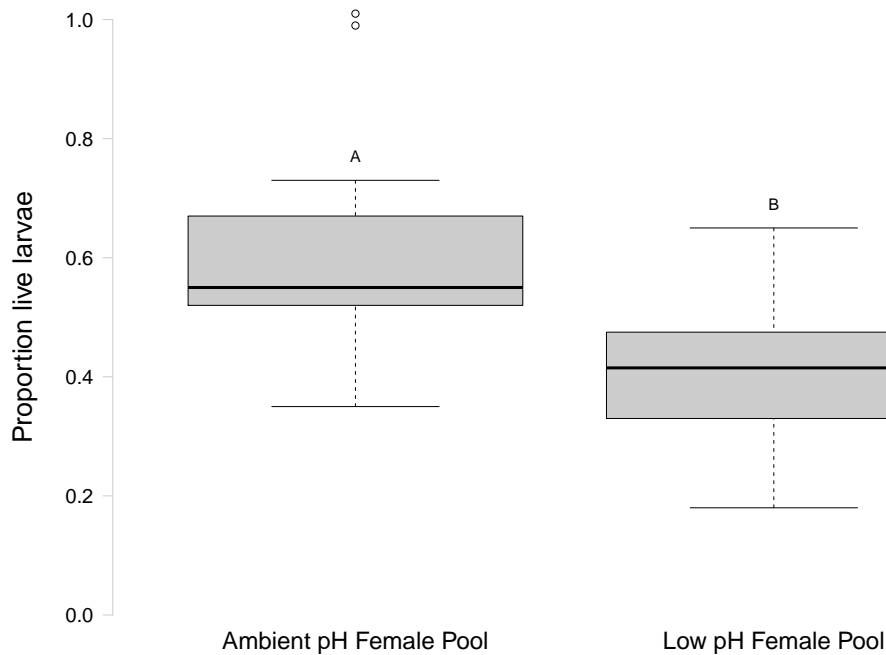
Figure 2.2: Larval survival 18 hours post-fertilization

Chapter 3

LOW PH INFLUENCES METHYLATION PATTERNS OF GONAD GROWTH GENES IN THE PACIFIC OYSTER (*CRASSOSTREA GIGAS*)

3.1 Abstract

There is a need to investigate mechanisms of phenotypic plasticity in marine invertebrates as negative effects of climate change are experienced by coastal ecosystems. Environmentally-induced changes to the methylome can confer plasticity through regulation of gene expression, but methylome responses may be tissue- and species-specific. To assess how ocean acidification affects the Pacific oyster (*Crassostrea gigas*) epigenome, we exposed oysters to either low pH (7.31 $\pm$ 0.02) or ambient pH (7.82 $\pm$ 0.02) conditions for seven weeks. Whole genome bisulfite sequencing was used to identify methylated regions in female oyster gonad samples. Analysis of gonad methylomes revealed a total of 1,599 differentially methylated loci (DML) were identified and found primarily in genes, with several genes containing multiple DML. Approximately 20% of DML overlapped with C/T SNPs. Overrepresented biological processes and cellular components in genes containing DML included cilium movement, development, and cytoskeletal components, suggesting that pH-responsive methylation may target gonad growth processes. Since enriched processes were associated with genic DML in multiple transcripts, it is possible DML in these genes could regulate alternative splicing. Our findings provide a foundation for understanding methylation-induced plasticity in *C. gigas*.

3.2 Introduction

There is great interest in elucidating how changes in the environment can impact marine invertebrate stress responses, and determining how plastic that phenotype is by examining DNA methylation in marine invertebrates (Eirin-Lopez & Putnam, 2018; **Hofmann2017?**). Environmental variation can alter the positions of methyl groups on CpG dinucleotides (Eirin-Lopez & Putnam, 2018). Methylation does not alter the DNA sequence itself, but can regulate gene expression by fine-tuning expression of housekeeping genes or providing additional transcriptional opportunities for environmental response genes (**Gavery2014?**). Methylation regulation of gene expression could then influence plasticity in response to environmental stressors (Eirin-Lopez & Putnam, 2018). Ocean acidification (**ChandraRajan2021?**; **Bogan2020?**; **Downey-Wall020?**; **Lim2020?**; **Venkataraman2020?**; **Liew2018?**), heat stress (**Arredondo-Espinoza2021?**), upwelling conditions (**Strader2019?**), disease (**Johnson2020?**), pesticide exposure (**Rondon2017?**), and temperature and salinity (**Liew2018?**) have all elicited changes in marine invertebrate methylomes, demonstrating a potential for methylation to regulate organismal responses to environmental conditions.

Given the impact of ocean acidification on calcifying species like oysters (**Melzner2019?**; **Parker2013?**; **Kroeker2010?**), several environmental epigenetic studies have examined the influence of ocean acidification on molluscan methylomes (Table 3.1). For example, eastern oyster (*Crassostrea virginica*) mantle (**Downey-Wall2020?**) and reproductive tissue (**Venkataraman2020?**), and Hong kong oyster (*Crassostrea hongkongensis*) mantle (**ChandraRajan2021?**), and larval methylomes (**Lim2020?**) all display changes in DNA methylation after experimental ocean acidification exposure. Between all four studies, only (**Venkataraman2020?**) reports the percent of CpGs methylated: the stated 22% is consistent with the baseline 15% methylation observed in initial Pacific oyster (*Crassostrea gigas*) methylation studies (Gavery & Roberts, 2013), as well as examination of methylome responses to heat stress (14.4%; (**Arredondo-Espinoza2021?**)) and diuron expo-

sure (16.6%; (**Rondon2017?**)). In *C. virginica*, the mantle methylome was hypomethylated after ocean acidification exposure (**Downey-Wall2020?**), while there were no differences in global gonad methylation patterns (**Venkataraman2020?**). Experimental ocean acidification exposure yielded differential methylation predominantly in gene bodies, but concentration in exons (**Lim2020?**; **Venkataraman2020?**) versus introns (**ChandraRajan2021?**) varies. Hyper- and hypomethylation occurs with equal frequency in response to ocean acidification. This is consistent with diuron exposure (**Rondon2017?**), but not with heat stress (**Arredondo-Espinoza2021?**). Gene functions enriched in differential methylation datasets were inconsistent between species and tissue types (Table 3.1). While aminotransferase complex and biosynthetic processes were found to be enriched in hypomethylated DML in *C. virginica* mantle tissue (**Downey-Wall2020?**), there were no enriched GOterms associated with genic DML in *C. virginica* gonad samples (**Venkataraman2020?**). In contrast, several processes were enriched in methylation datasets for *C. hongkongensis*, including acetoacetylco-A reductase activity, dehydrogenase activity, cellular response to pH, protein xylosyltransferase activity, translation factor activity, RNA binding, and diacylglycerol kinase activity in the mantle (**ChandraRajan2021?**) and cytoskeletal and signal transduction, oxidative stress, metabolic processes, and larval metamorphosis in larvae (**Lim2020?**). While considerable effort has been made to understand molluscan methylation responses to ocean acidification, it is clear that there may be species- and tissue-specific responses. Additionally, use of enrichment methods prior to bisulfite conversion and sequencing in previous studies may impact functional interpretation differential methylation data (**Trigg2021?**).

In addition to being an important global aquaculture species, *C. gigas* has been used to further our understanding of methylation in marine invertebrates (Gavery & Roberts, 2013, 2017; Roberts & Gavery, 2012; **Riviere2013?**; **Gavery2014?**; **Olson2014?**; **Olson2015?**; **Saint-Carlier2015?**; **Riviere2017?**; **Song2017?**). Examination of diuron exposure (**Rondon2017?**) and heat stress on different *C. gigas* phenotypes (**Arredondo-Espinoza2021?**) reveal that environmental conditions do influence the Pacific oyster methylome, and other

oyster methylomes are responsive to ocean acidification (Table 3.1). However, studies examining the influence of ocean acidification on the *C. gigas* methylome are currently absent. Additionally, a recent study found that female *C. gigas* exposure to low pH conditions, followed by four months in ambient pH conditions prior to spawning, can still negatively impact larval survival 18 hours post-fertilization (**Venkataraman2019?**). Since maturation stage was not different between female *C. gigas* after low pH exposure, this finding suggests environmental "memory" may be mediating the carryover effect.

We sought to understand impacts of ocean acidification on the Pacific oyster gonad methylome in order to not only understand DNA methylation as a potential mechanism for the observed negative maternal carryover effect previously reported in (**Venkataraman2019?**), but to also examine the functional role of methylation in regulating biological responses to ocean acidification. Female oyster gonad methylome sequencing revealed distinct methylation changes associated with low pH exposure. These results provide the foundation for examining phenotypic plasticity within a generation, as well as larval methylomes and potential intergenerational effects.

3.3   Methods

3.3.1   Experimental Design

The experimental design and seawater chemistry analysis with carbon chemistry parameters are described in (**Venkataraman2019?**). Briefly, adult hatchery-raised *C. gigas* (average shell length = 117.46 $\pm$ 19.16 cm) were exposed to either low pH (7.31 $\pm$ 0.02) or ambient pH (7.82 $\pm$ 0.02) conditions from February 15, 2017 to April 8, 2017 at the Kenneth K. Chew Center for Shellfish Research and Restoration at the National Oceanic and Atmospheric Administration Manchester Field Station (47º 34' 09.1" N 122º 33' 19.0" W, Manchester, WA). Shell length, width, and oyster wet weight were obtained before and after pH exposure, and gonad tissue was extracted for histological analysis at each sampling point. Prior to exposure,

twenty *C. gigas* were randomly selected for histology, while ten oysters from each treatment were randomly selected from tanks to assess maturation. Gonad tissue in cassettes was fixed for histology using PAXgene Tissue Fix and Stabilizer and sent to Diagnostic Pathology Medical Group, Inc. (Sacramento, CA) for hematoxylin and eosin staining and slide preparation. Tissues exposed to ambient pH were confounded during processing, preventing any tank identification. Maturation state and organism sex was evaluated histologically at 40x magnification (**Enríquez-Díaz2008?**; **Fabioux2005?**).

3.3.2 DNA Extraction and Library Preparation

To determine how pH exposure altered maternal gonad methylation, DNA was extracted from histology blocks for four female (or presumed female) oysters for each treatment using the PAXgene Tissue DNA Kit (Qiagen) with modifications specified below at the end of the seven-week exposure. For each sample, up to 0.02 g of histology tissue embedded in paraffin was cut from the histology block and placed in a 2 mL round-bottomed tube. The tissue was then further homogenized within the tube. After adding 1 mL of xylene, each sample was vortexed for 20 seconds, incubated at room temperature for three minutes, then centrifuged at maximum speed for three minutes. To evaporate ethanol from samples after ethanol addition and vortexing steps, samples were placed on a heat block at 37°C for 10-20 minutes to evaporate any residual ethanol.

The resuspended pellet was lysed using a TissueTearor at setting 1. Prior to lysis and between samples, the TissueTearor was run in a 10% bleach solution for 20 seconds, followed by two consecutive 20 second runs in DI water to clean the instrument. Lysates were transferred to clean, labelled 1.5 mL microcentrifuge-safelock tubes. Proteinase K (20 µL) was added to each sample, and pulse-vortexed for 15 seconds to mix. The sample-Proteinase K solution was incubated 56°C for 60 minutes. Every ten minutes, the samples were briefly removed from the heat block to vortex at maximum speed for one minute. After 60 minutes, RNase A (4 µL, 100 mg/mL) was added to each sample. Samples were vortexed at maximum speed for

20 seconds to mix and incubated for two minutes at room temperature to obtain RNA-free genomic DNA and reduce possible interference for downstream enzyme reactions. Samples were then kept at 80ºC for 60 minutes, and vortexed at maximum speed for one minute in ten-minute intervals. After the incubation, Buffer TD2 and 200 µL ethanol were added to each sample, vortexing thoroughly to mix.

DNA was isolated for each sample after lysis and incubation following manufacturer's instructions. To elute DNA, samples were loaded onto the spin column with Buffer TD5 (50 µL) and incubated at room temperature for five minutes to increase yield, then centrifuged at maximum speed for one minute. A Qubit™ 3.0 and dsDNA BR Assay (Invitrogen) was used to quantify sample yield using 1 µL of DNA for each sample. For two samples, DNA was further concentrated using an ethanol precipitation.

Samples were sent to ZymoResearch for bisulfite conversion and library preparation. Bisulfite conversion was performed with the Zymo-Seq WGBS Library Kit (Cat. #D5465) using 100 ng of genomic DNA following manufacturer's instructions. Per ZymoResearch protocol, samples were spiked with a mixture of six unique double-stranded synthetic amplicons (180-200 bp) from the *Escherichia coli* K12 genome. Since each amplicon represents a different percent methylation, the spike-in was used to determine bisulfite conversion efficiency. Libraries were barcoded and pooled into a single lane to generate 150 bp paired-end reads in a single Illumina NovaSeq flowcell.

3.3.3   Genome Information and Feature Tracks

Nuclear (**Peñaloza2021?**) (NCBI GenBank: GCA_902806645.1, RefSeq: GCF_902806645.1) and mitochondrial genome sequences (NCBI Reference Sequence: NC_001276.1) were used in downstream analysis. These sequences were combined for downstream alignment. The `fuzznuc` function within EMBOSS was used to identify CG motifs in the combined genomes.

The RefSeq annotation was used to obtain genome feature tracks, and create additional

tracks. The annotation included a combination of Gnomon, RefSeq, cmsearch, and tRNAscan-SE predictions. Gene, coding sequence, exon, and lncRNA tracks were extracted from the RefSeq annotation. These tracks were then used to obtain additional genome feature information. To create an intron track, the complement of the exon track was generated with BEDtools `complementBed` v.2.26.0 to create a non-coding sequence track (**Quinlan2010?**). The overlap between the genes and the non-coding sequences was found with `intersectBed` and defined as introns. Similarly, untranslated regions of exons were obtained by subtracting coding sequences from exon information with `subtractBed`. Flanking regions were defined as the 1000 bp upstream or downstream of genes. Upstream flanks were generated by adding 1000 bp upstream of genes, taking into account strand, with `flankBed`. Existing genes were removed from flanking sequences using `subtractBed`. A similar process was performed to generate downstream flanking regions. Intergenic regions were isolated by finding the complement of genes with `complementBed`, then using subtractBed to remove any flanks. Transposable element locations were obtained from RepeatMasker using the NCBI RefSeq annotation (RefSeq: GCF_902806645.1) (Smit et al., 2013; Tarailo-Graovac & Chen, 2009). The number of CG motifs in a given feature track was obtained using `intersectBed`.

3.3.4 Sequence Alignment

Reads were trimmed three times prior to alignment using TrimGalore! v.0.6.6 (**Martin2011?**). Known Illumina adapter sequences were removed (`--illumina`), along with 10 bp from both the 5' (`--clip_R1 10` and `--clip_R2 10`) and 3' ends of 150 bp paired-end reads (`--three_prime_clip_R1 10` and `--three_prime_clip_R2 10`). Any remaining adapters were removed in a second round of trimming. Finally, poly-G tails were trimmed out of samples by manually specifying adapter sequences (`--adapter GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG` and `--adapter2 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG`). Sequence quality was as-

sessed with FastQC v.0.11.9 (**Andrews2010?**) and MultiQC (**Ewels2016?**) after each round of trimming.

Trimmed samples were then aligned to the combined nuclear and mitochondrial genomes for *C. gigas* (**Peñaloza2021?**). The genome was bisulfite converted with Bowtie 2-2.3.4 (Linux x84_64 version (**Langmead2012?**)) using the bismark_genome_preparation function within Bismark v.0.22.3 (**Krueger2011?**). Reads were aligned to the bisulfite-converted genome. Non-directional input was specified (`--non_directional`), and alignment scores could be no lower than -90 (`-score_min L,0,-0.6`). Resultant BAM files were deduplicated (`--deduplicate_bismark`), and methylation calls were extracted from deduplicated BAM files (`bismark_methylation_extractor`). Deduplicated BAM files were also sorted and indexed for downstream applications using SAMtools v.1.10 (**Li2009?**). Using coverage files generated from methylation extraction, CpG information was consolidated between strands to report 1-based chromosomal coordinates (`coverage2cystosine`). Using coverage2cytosine coverage file output, genomic coordinates for CpGs with at least 5x coverage across all samples were written to BEDgraphs. Summary reports with alignment information for each sample were generated (`bismark2report` and `bismark2summary`) and concatenated with MultiQC. Code for all sequence alignment steps can be found in the associated Github repository.

3.3.5 *SNP Identification*

As genomic differences can play a role in methylation, single nucleotide polymorphisms (SNPs) were identified from WGBS data to determine if they overlapped with differentially methylated loci. SNP variants across all samples and in individuals were identified with `BS-Snper` (**Gao2015?**). As `BS-Snper` only takes one input file, deduplicated and sorted BAM alignments generated with Bismark v.0.22.3 and SAMtools v.1.10 (**Krueger2011?**; **Li2009?**) were merged (`samtools merge`). Variants were then identified in the merged BAM file or individual deduplicated and sorted BAM files. Default `BS-Snper` settings were

used, except for a minimum coverage of 5x to be consistent with methylation analyses. C/T SNPs were filtered by identifying overlaps between SNPs and CG motifs with intersectBed. The C/T SNPs were used in downstream analyses to identify differentially methylated loci that overlapped with SNPs.

3.3.6 Global Methylation Characterization

Global methylation patterns were characterized averaging DNA methylation across all samples (using a 5x coverage threshold). A CpG locus was considered highly methylated if average percent methylation was greater than or equal to 50%, moderately methylated if percent methylation was between 10-50%, and lowly methylated if percent methylation was less than or equal to 10%. The genomic location of highly, moderately, and lowly methylated CpGs in relation to UTR, CDS, introns, up- and downstream flanking regions, transposable elements, and intergenic regions were determined by using intersectBed. We tested the null hypothesis that there was no association between the genomic location of CpG loci and methylation status (all CpGs versus highly methylated CpGs) with a chi-squared contingency test (`chisq.test` in R Version 3.5.3).

3.3.7 Identification of Differentially Methylated Loci

Identification of differentially methylated loci, or DML, was conducted with `methylKit` v.1.17.4 (**Akalin2012?**) in R Version 3.5.3. Prior to DML identification, data were imported and processed. Data with at least 2x coverage were imported using `methRead` from merged CpG coverage files produced by `coverage2cytosine`. Imported data were filtered again (`filterByCoverage`) to require 5x coverage for each CpG locus (`lo.count = 5`) in each sample. Potential PCR duplicates in each sample were also removed by excluding CpG loci in the 99.9th percentile of coverage (`high.perc = 99.9`). Once filtered, data were normalized across samples (`normalizeCoverage`) to avoid over-sampling reads from one sample during downstream statistical analysis. Histograms of percent methylation (`getMethylationStats`)

and CpG coverage (`getCoverageStats`) were used to confirm normalization. After filtering and normalization, data at each CpG locus with 5x were combined (`unite`) to create a CpG background, ensuring that each locus had at least 5x coverage in each sample.

Methylation differences were then identified for *C. gigas* gonad samples. A logistic regression (`calculateDiffMeth`) modeled the log odds ratio based on the proportion of methylation at each CpG locus. A covariate matrix with gonad stage information (`covariates`) were applied to the model for the *C. gigas* data to account for varying maturation stages in samples. Additionally, an overdispersion correction with a chi-squared test (`overdispersion = "MN", test = "Chisq"`) was applied.

Differentially methylated loci, or DML, were defined as CpG dinucleotides with at least a 50% methylation difference between pH treatments, and a q-value < 0.01 based on correction for false discovery rate with the SLIM method (**Wang2011?**). Hypermethylated DML were defined as those with significantly higher percent methylation in low pH samples, and hypomethylated DML with significantly lower percent methylation in low pH samples. BEDfiles with DML locations were created and viewed with the Integrative Genomics Viewer (IGV) version 2.9.2 (**Thorvaldsdóttir2013?**).

3.3.8 DML Characterization

The location of DML was characterized in relation to various genome feature tracks. Presence of DML in UTR, CDS, introns, up- and downstream flanking regions, transposable elements, and intergenic regions was discerned using intersectBed. A chi-squared contingency test was used to test the null hypothesis of no association between genomic location and differential methylation by comparing the genomic location of all CpGs with 5x data and DML. Additionally, the number of DML present in each chromosome was quantified to see if DML were distributed uniformly across the genome. Overlaps between DML and SNPs were identified to determine if DML were solely attributed to treatment differences, or if genetic differences contributed to differential methylation.

3.3.9   Enrichment Analysis

To determine if genes containing DML were associated with overrepresented biological processes, cellular component, or molecular function gene ontologies (GO), we conducted functional enrichment analysis. Prior to this analysis, the *C. gigas* genome was annotated. A `blastx` was performed against the Uniprot-SwissProt database (accessed June 01, 2021) to get Uniprot Accession information for each transcript (**UniProt2019?**). The transcript nucleotide sequences were extracted from the genome (GCF_902806645.1_cgigas_uk_roslin_v1_rna_from_genomic.fna available in the NCBI Annotation Release 102). Transcript IDs from the `blastx` output were matched with GO terms from the Uniprot-Swissprot database using Uniprot Accession codes. These transcript IDs and corresponding GO terms were then used to create a gene ID-to-GO term database (`geneID2GO`) for manual GO term annotation. The transcript IDs were filtered for genes that contained CpGs with at least 5x coverage in all samples, as the same parameters were used to generate the CpG background for differential methylation analysis. Each line of the database contained transcript ID in one column, and all corresponding GO terms in another column.

Gene enrichment analysis was conducted with `topGO` v.2.34.0 in R (**Alexa2010?**). The `geneID2GO` database was used as the gene universe for enrichment, while transcript IDs for genes with DML were used as the genes of interest. These transcript IDs were filtered such that they did not include any transcripts associated with DML-SNP overlaps. A `topGO` object was generated for each DML list and GO category (biological process, cellular component, or molecular function), with GO term annotation performed using the `geneID2GO` database. A Fisher's exact test was used to identify GO terms in each DML list significantly enriched with respect to the gene background (P-value < 0.01). Keeping with `topGO` default settings, we did not correct for multiple comparisons (**Alexa2010?**). Visualization of enriched GO terms was conducted using `simplifyEnrichment` v.1.2.0 default settings (**Gu2021?**). Briefly, a semantic similarity matrix was created from enriched GO IDs using

the "Relevance" similarity measure and GO tree topology (`GO_similarity`). The semantic similarity matrix was clustered using the binary cut method, and visualized as a word cloud alongside a heatmap of semantic similarity (`simplifyGO`).

3.4 Results

3.4.1 Sequence Alignment

Whole genome bisulfite sequencing produced 1.38 x 10^9 total 150 bp paired-end reads. After quality trimming, 1.35 x 10^9 total paired reads remained (Appendix 1). Of the trimmed paired-end reads, 8.31 x 108 (61.7%) were uniquely aligned to the bisulfite-converted *C. gigas* genome with appended mitochondrial sequence (Appendix 1).

3.4.2 SNP Identification

C/T SNP variants were identified from WGBS data to determine if genomic differences contributed to differential methylation. A total of 13,234,183 unique SNPs were found in individual or merged BAM files, including 300,278 C/T SNPs. These C/T SNPs were used in downstream analyses (see *DML Characterization and Enrichment Analysis*).

3.4.3 Global Methylation Patterns

When data were combined, using a 5x coverage threshold 11,238,223 CpG loci (84.8% of 13,246,547 total CpGs in *C. gigas* genome) were characterized. The majority of CpGs, 9,201,922 loci (81.9%) were lowly methylated, followed by 1,029,894 highly methylated (9.2%) and 1,006,406 moderately methylated (9.0%) CpGs. Highly methylated CpGs were found primarily in genic regions (964,976 CpGs, 93.7%), with 57,896 (5.6%) in exon UTR, 355,575 (34.5%) in CDS, and 554,093 (53.8%) in introns (Figure 3.1). The distribution of highly methylated CpGs was significantly different than all 5x CpGs detected by WGBS (Appendix

1). A Principal Component Analysis (PCA) did not demonstrate any clear separation of samples by treatment, and consistently high Pairwise Pearson's correlation coefficients support the lack of global percent methylation differences between pH treatments (Appendix 1).

3.4.4   DML Characterization

A total of 1,599 DML were identified using a 50% methylation difference threshold (Figure 3.2), which were evenly split between 793 (49.6%) hypomethylated and 806 (50.4%) hypermethylated DML (Figure 3.2). Of these DML, 315 (19.7%) overlapped with C/T SNPs. The DML were distributed amongst the ten main chromosomes, as well as scaffolds not placed in any chromosome (Figure @ref(fig: chromdistr), Appendix 1).

The majority of DML (1,478; 92.4%) were located in genic regions (Figure @ref(fig: locDML)), consisting of 1080 unique genes. Most genes only contained one DML; however, several genes contained multiple DML, ranging from 2-104 DML in one gene (Appendix 1). The genic DML were primarily located in introns (941; 58.8%), followed by CDS (442; 27.6%) and exon UTR (95; 5.9%). Putative promoter regions upstream of transcription start sites contained DML (12; 0.8%), but more were located in downstream flanking regions (73; 4.6%). Additionally, 297 of the 1,478 genic DML overlapped with C/T SNPs in 283 unique genes. The DML were also found in transposable elements (496; 31.0%), lncRNA (9; 0.6%), and intergenic regions (42; 2.6%). The number of DML in all genome features differed significantly from all 5x CpGs (Appendix 1).

3.4.5   Enrichment Analysis

We identified 10 biological processes and 4 molecular function GOterms significantly enriched in genes containing DML that did not overlap with C/T SNPs (Figure @ref(fig: simplifyEnrichment), Appendix 1). A total of 339 genes with GOterm annotations containing 437 DML were used for this analysis. The enriched biological process GOterms were involved in motil-

ity (ex. cilium-dependent cell motility, cilium movement) and development (insta larval or pupal development, post-embryonic development), while the cellular component GOterms were involved in microtubule-associated and histone acetyltransferase complexes (Appendix 1). A total of 7 unique genes and 34 unique transcripts were associated with the biological process terms, and 2 unique genes and 15 unique transcripts were associated with cellular component terms (Appendix 1).

3.5 Discussion

Our work examined the potential role of environmentally-induced changes to the reproductive tissue methylome to mediate carryover effects in *C. gigas*. We identified a total of 1,599 differentially methylated loci, showing that the methylome is responsive to ocean acidification. We also found significantly enriched biological processes associated with genic DML, implying that low pH may impact regulation of distinct processes, such as gonad development. We also highlight the importance of understanding genomic contributions to differential methylation through the identification of SNPs that overlap with DML. Our findings contribute to the growing body of work examining epigenetic responses in molluscan species, and provide the foundation for future research examining intergenerational epigenetic inheritance and its potential to mediate plastic responses to environmental stressors.

The enriched GOterms associated with genes containing DML suggest environmentally-responsive methylation can regulate gonad growth processes. Enrichment of cilium movement and overrepresentation of microtubule-associated complexes could be related to changes in cytoskeletal structure. These changes may signify reorganization of female reproductive tissue to accommodate growing oocytes, or a need to regulate protein transport and cytoskeletal activity. Differentially methylated genes in *C. hongkongensis* larvae exposed to ocean acidification also were enriched in cytoskeletal processes (**Lim2020?**), suggesting that this particular process may be targeted by methylation across life stages. In female oysters, expression of microtubule-associated genes increases from stage 0 to 2, but decreases after

stage 3 (**Dheilly2012?**). As (**Venkataraman2019?**) did not identify any difference in maturation stage between oysters in low and ambient pH conditions, methylation regulation of these cytoskeletal genes may have allowed oysters in low pH to maintain gonad growth at the same rate as those in ambient pH conditions. It is also possible that the mixed cell types used in samples could be influencing enrichment results. Microtubule-associated genes are also present in somatic tissue (**Dheilly2012?**), so changes to the somatic tissue methylation may have also been captured in our analyses as we did not separate cell types. A similar study of *C. virginica* reproductive tissue did not find enrichment of growth processes, which may be due to confounding factors such as mixed cell types, as well as mixed sex and maturation stage in samples (**Venkataraman2020?**). By using females of known maturation stage, we were able to identify processes influenced by low pH in gonad tissue. Future work should separate different cell types to understand environmentally-responsive methylation specifically in gonad cells.

Additional functions associated with genes containing DML provide insight into potential mechanisms mediating the negative maternal carryover effect observed in (**Venkataraman2019?**)]. Enriched biological processes associated with development were involved in larval or post-embryonic development, suggesting methylation regulation of genes that may be important for early life-stages. The most common functions associated with genes containing DML were related to hormone signaling, secretion, and metabolism. Whether acidification alters invertebrate reproductive hormones is not yet known, but alterations to hormone pathways through DNA methylation may impact maturation. While there was no difference in female maturation status after the seven-week pH exposure (**Venkataraman2019?**), persistent changes to expression of genes related to hormonal processes could impact oyster gametogenesis. Several genes involved in lipid metabolism and biomineralization also contained DML. Reduction in lipid content has been observed in the Sydney rock oyster *S. glomerata* exposed to experimental acidification conditions, but this lipid content reduction did not yield negative effects on embryonic development (**Scanes2018?**), so the connection

between adult *C. gigas* low pH exposure, maternal provisioning, and poor larval survival may not be supported. Several studies note impacts of adult pH exposure on larval calcification process. For example, *S. glomerata* larvae exhibit faster shell growth in high pCO_2 conditions when parents mature in those same conditions (**Parker2012?**; **Parker2015?**). In contrast, larvae from other species found in the North Atlantic such as northern quahog (= hard clam; *Mercenaria mercenaria*) and bay scallops (*Argopecten irradians*) developed slower when parents were reproductively conditioned in low pH conditions (**Griffith2017?**). It is possible that *C. gigas* larvae experienced reductions in calcification which lead to lower survival after adult low pH exposure. Additional research is required to discern if methylation changes to hormone signaling, lipid metabolism, and biomineralization genes persist in larvae, and how those changes may influence larval gene expression or survival.

In *C. gigas*, low pH may elicit methylation changes in genes to improve transcriptional control and maintain homeostasis. Differential methylation was found primarily in introns and exons, which suggests that these loci may influence gene expression. This finding is consistent with the literature examining molluscan methylome responses to ocean acidification (**ChandraRajan2021?**; **Downey-Wall2020?**; **Lim2020?**; **Venkataraman2020?**), disease (**Johnson020?**), pesticide exposure (**Rondon2017?**), and heat stress (**Arredondo-Espinoza2021?**), as well as environmental responses in corals (**Liew2018a?**; **Liew2018b?**), urchins (**Strader2019?**), and pteropods (**Bogan2020?**). The relationship between methylation and gene expression has been investigated in *C. gigas*. High levels of methylation in genes were associated with increased gene expression and low expression variability, implying that methylation leads to transcriptional stability (Gavery & Roberts, 2013; **Olson2014?**). Methylation was also found to reduce transcriptional noise in response to ocean acidification in corals (**Liew2018?**). While pH-influenced differential methylation was not found to impact gene expression directly in *C. virginica* (**Downey-Wall2020?**) or *C. hongkongensis* (**ChandraRajan2021?**), neither study investigated methylation connections with alternative splicing. As enriched biological process GOterms were found in

various transcripts within genes, there may be a role for methylation regulation of expression regulation via alternative splicing. Additionally enrichment of GOterms related to histone acetyltransferases could be indicative of chromatin regulation, which would then impact methylation and gene expression (ex. (**Gatzmann2018?**)). Both elevated temperature and low pH conditions have been found to influence methylation of histone-related genes in *C. gigas* (**Arredondo-Espinoza2021?**) and *C. hongkongensis* (**Lim2020?**; **ChandraRajan2021?**). Our findings suggest that the connection between methylation and gene expression requires additional investigation when considering environmental response, with specific inclusion of mechanisms like alternative splicing and chromatin regulation.

Confounding elements may have influenced gonad methylome responses to ocean acidification, and our interpretation of those responses. Although specimens used were sibling oysters in an effort to reduce genomic variation, we identified several SNPs that overlapped with DML. It is possible that the DML algorithm used is biased towards identification of DML that overlap with SNP variants, as it removes individual variation (**Feng2014?**). With a limited sample size, SNP identification from WGBS is useful for evaluating if DML are solely environmentally-induced, or if genomic variation influences methylation differences. Inclusion of DML that overlap SNPs during enrichment analysis may also confound results, leading to a different biological interpretation. Additionally, the lack of annotated genes used in the enrichment analysis limits our interpretation of which processes are most impacted by pH-sensitive methylation.

Maturation stage is known to produce distinct baseline methylation patterns in *C. gigas* (**Zhang2018?**). While the overall maturation stage was used as a covariate when identifying DML, it is possible that each sample methylome contains information from cells with various maturation stages or cell types. Even though methylation differences in genic regions support our conclusion that low pH does impact the *C. gigas* gonad methylome, single-cell sequencing efforts or investigations within a single maturation stage would clarify how these sources of variation influenced our findings.

3.6 Conclusion

Our work — which is the first to explore methylation responses to ocean acidification in Pacific oysters — found that low pH treatment altered the *C. gigas* female gonad methylome after a seven-week exposure. We found that differential methylation occurs primarily in genic regions, which is consistent with other studies examining oyster methylome responses to ocean acidification. In conjunction with histology data from (**Venkataraman2019?**), enriched biological processes and cellular components in genes containing DML suggest that methylation may be used to maintain gonad growth in low pH conditions. Our work expands on previous molluscan methylation research by not only demonstrating a potential role for methylation in plastic responses to ocean acidification, but also incorporating removing genomic variation from analysis to strengthen biological interpretation of data.

As inheritance of environmentally-induced methylation — and any resulting phenotypic plasticity — is contingent on epigenetic marks being present in the germline, withstanding chromatin reorganization, and being packaged into gametes (Eirin-Lopez & Putnam, 2018), the next step in this work is correlating germline methylation with offspring methylation, and examining gene expression, protein abundance, and phenotype. Comparing parental and offspring phenotype will allow researchers to assess whether parental experience and resulting epigenetic responses are accurate, reliable cues for adaptive plasticity in offspring.

3.7 Tables

Table 3.1: Summary of published studies examining the impact of environmental stressors on oyster (*Crassostrea spp.*) methylation. CpGs methylated refers to the percentage of detected CpGs characterized as methylated by each study. Differentially methylated loci (DML), regions (DMR), and genes (DMG) are reported with the number of each differentially methylated feature hyper- or hypomethylated. Transcriptomics sequencing methods are included if the study investigated the connection between methylation and gene expression. Missing metadata is indicated by "NR" (not reported).

Table 3.1: Previous oyster methylation studies

Study	Species	Stressor	Exposure Time	Tissue	Method	CpGs Methylated	Differential methylation	Genome Features Impacted	Gene Functions Enriched	Connection to Transcription
Downey-Wall et al. 2020	C. virginica	Ocean acidification	80 days	Mantle	MBDBS	NR	85 DML (38 / 47)	Genes	Aminotransferase complex, biosynthetic process	Weak association (RNA-Seq)
Venkataraman et al. 2020	C. virginica	Ocean acidification	4 weeks	Gonad	MBDBS	22	598 DML (310 / 288)	Exons	None	NR
Chandra Rajan et al. 2021	C. hongkongensis	Ocean acidification	4.5 months	Mantle	MethylRAD	NR	377 DMG (214 / 163)	Introns	Acetoacetylco-A reductase activity, types of dehydrogenase activity, cellular response to pH, protein xylosyltransferase activity, translation factor activity, RNA binding, diacylglycerol kinase activity	None (RNA-Seq)
Lim et al. 2020	C. hongkongensis	Ocean acidification	21 days	Pediveliger	MethylRAD	NR	130 DMG (66 / 64)	Exons	Cytoskeletal and signal transduction, oxidative stress, metabolic processes, larval metamorphosis	NR

Arredondo-Espinoza et al. 2021	C. gigas	Heat stress	30 days	Gill	WGBS	14.4	161 DMR (147 / 14)	Introns	Binding processes, catalytic activity, transporter activity	NR
Johnson et al. 2020	C. virginica	Dermo	14 months	Gill	epiGBS/RRBS	NR	913 DMR (NR)	Genes	None	Weak association (TagSeq)
Rondon et al. 2017	C. gigas	Diuron	2 7-day periods	Spat	WGBS	16.6	236 DMR (121 / 115)	Genes	NR	Weak association (RNA-Seq)

3.8 Figures

Figure 3.1: Location of all CpGs with 5x coverage, highly methylated (50%), moderately methylated (10-50%), and lowly methylated (10%) CpGs in various genome features.
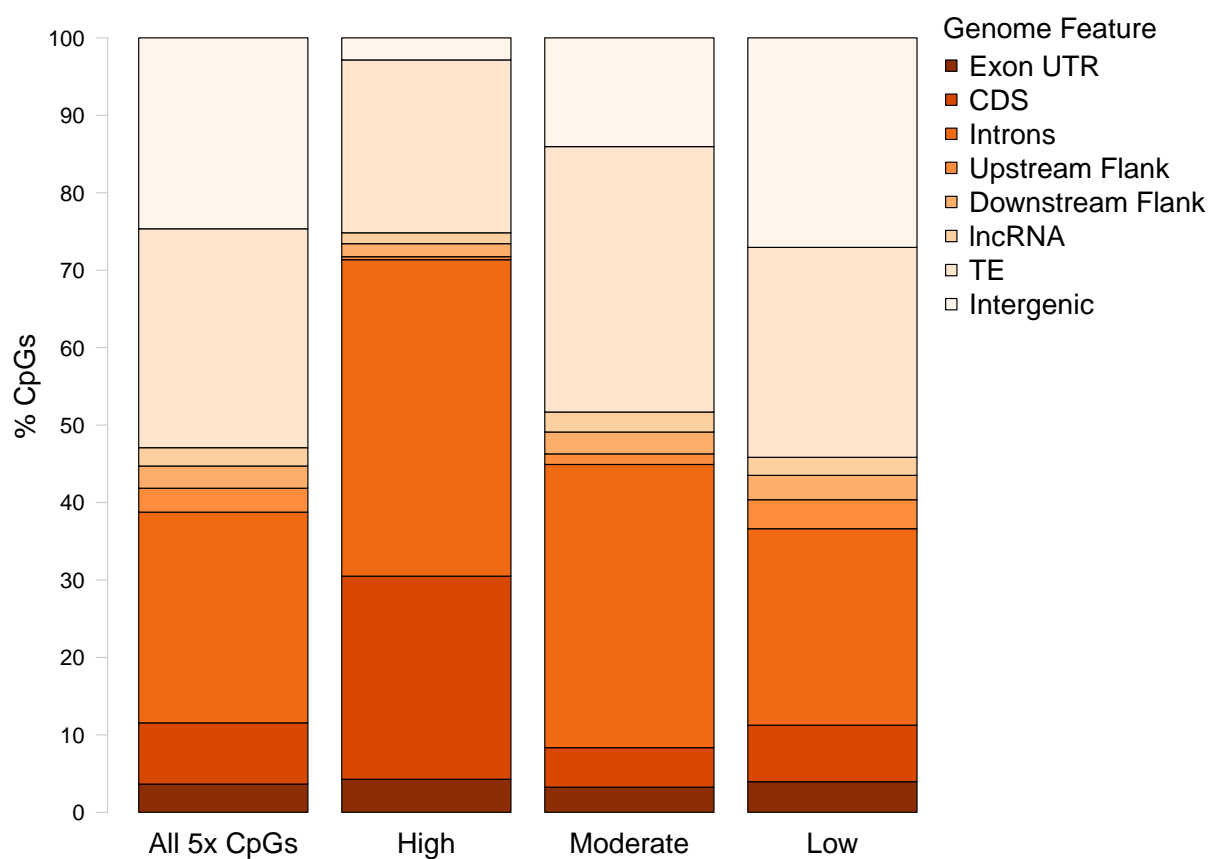
Figure 3.1: Genomic location of various CpG categories

Figure 3.2: Percent methylation values for DML created using an euclidean distance matrix. Samples in low pH conditions are represented by black, and samples in ambient pH conditions are represented by grey, with maturation stage along the bottom (0 = indeterminate, 3 = mature female). Darker colors indicate higher percent methylation, and a density plot depicts the distribution of percent methylation values for a panel. A total of 1599 DML were identified using a logistic regression, using a chi-squared test and 50% methylation difference cut-off.
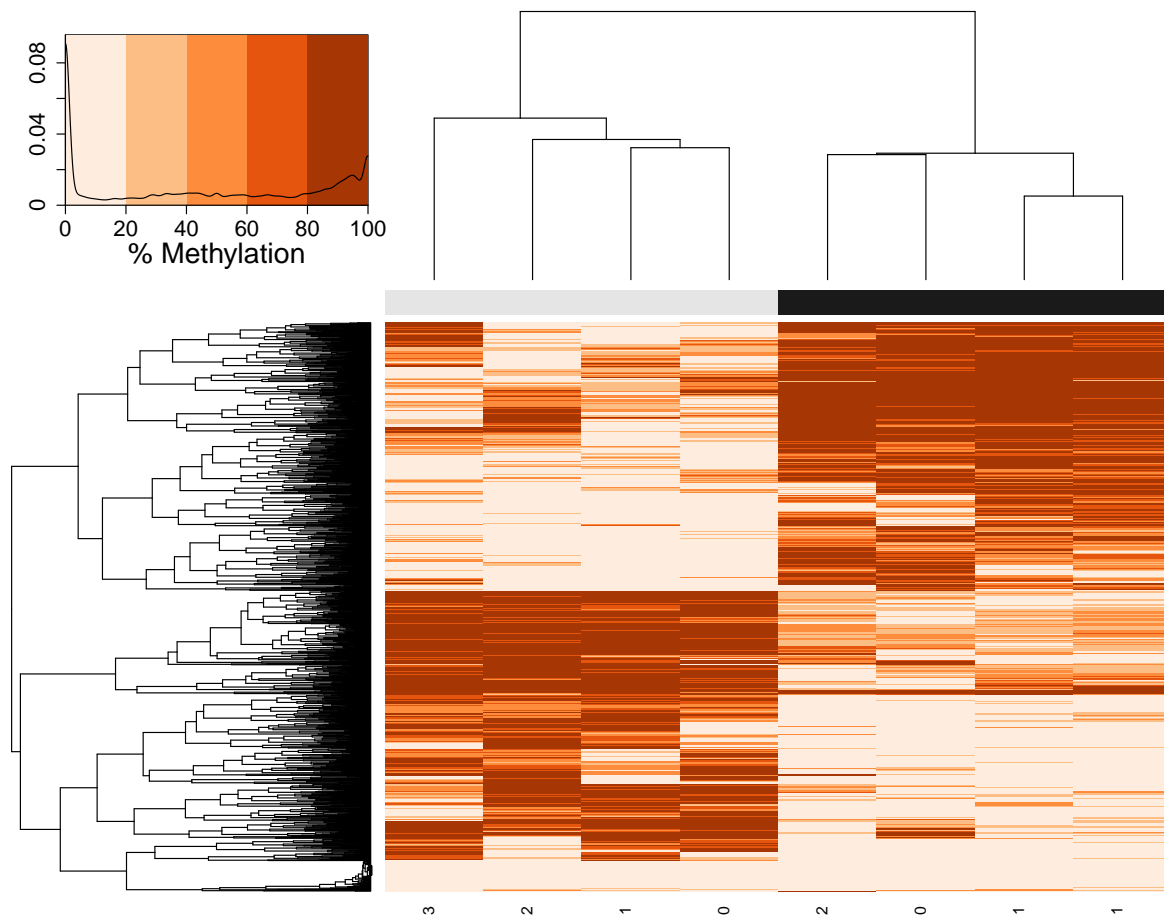
Figure 3.2: Heatmap of DML

Figure 3.3: Number of DML normalized by number of CpG in each chromosome (bars) and number of genes (line) in each chromosome. Additional DML were identified in scaffolds that were not mapped to any of the ten main linkage groups.
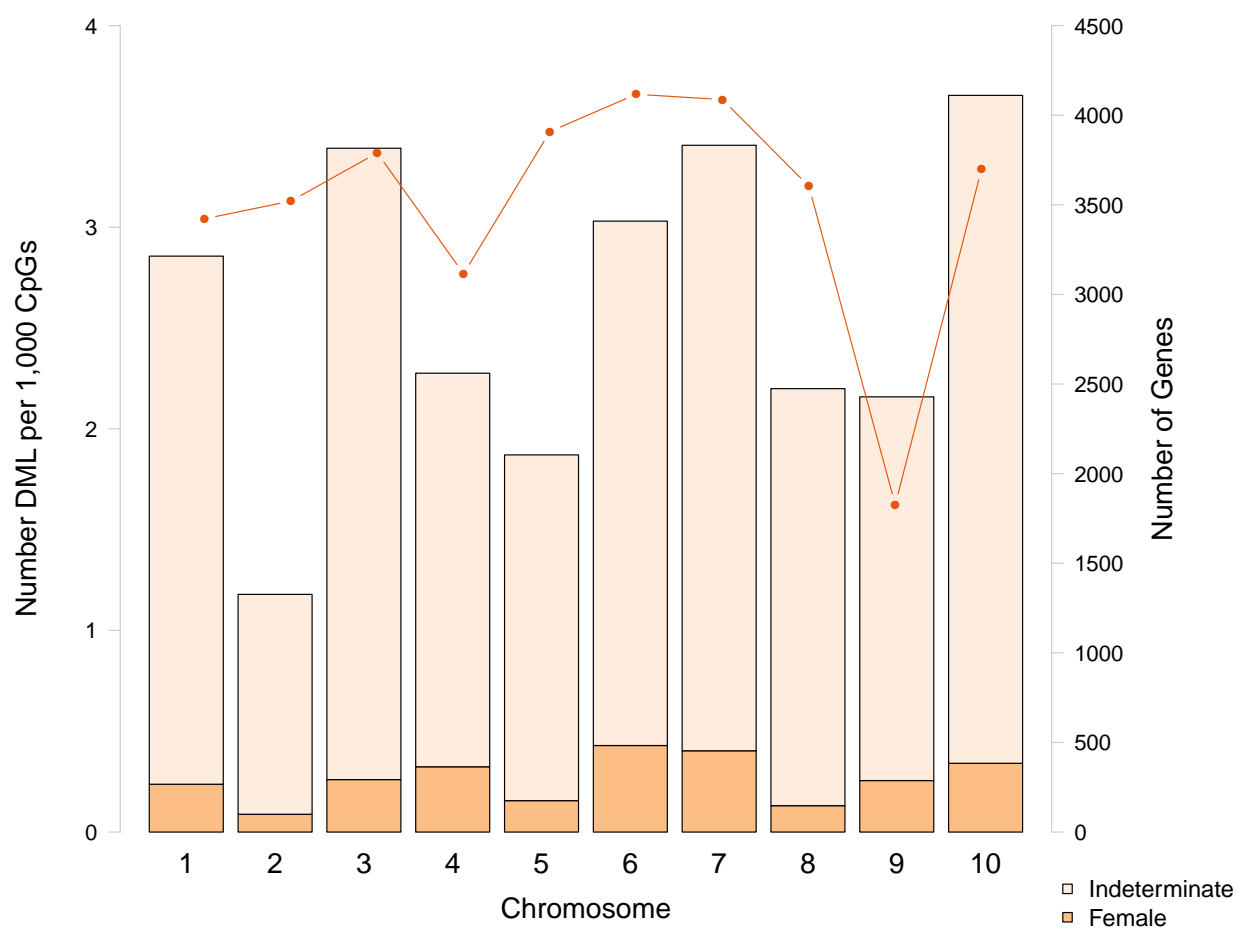
Figure 3.3: Distribution of DML in main chromosomes

Figure 3.4: Distribution of all CpGs with 5x coverage and DML in various genome features.
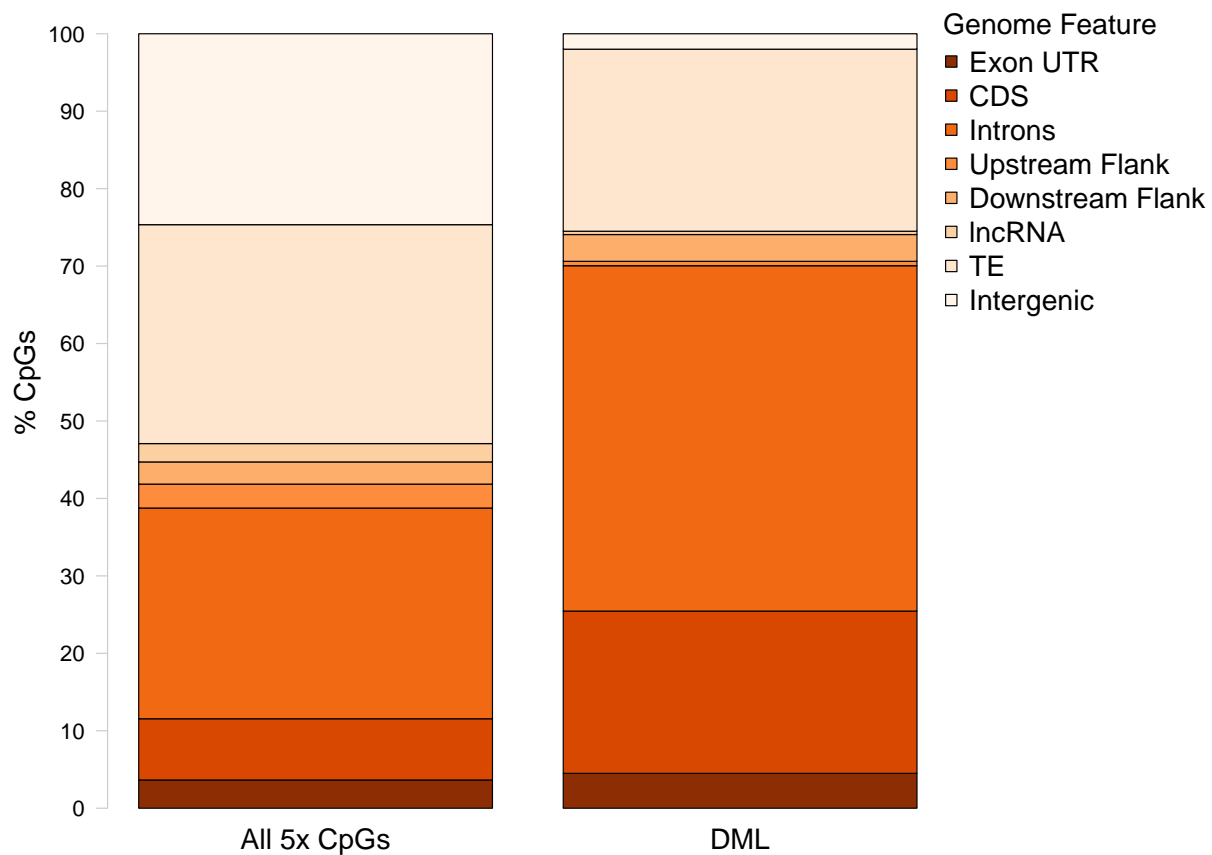
Figure 3.4: Genomic location of DML

Figure 3.5: Biological process GOterms enriched in genes containing DML. A semantic similarity matrix was generated using "Relevance" measures, then clustered using the binary cut method. The GOterms associated with each cluster are displayed in the word cloud, with the size of the word representing frequency in the dataset. Darker colors represent GOterms that have higher semantic similarity.

Figure 3.5: Significantly enriched GOterms

Chapter 4

POLYPLOIDY AND ENVIRONMENTAL STRESS HAVE DISTINCT IMPACTS ON PACIFIC OYSTER (*CRASSOSTREA GIGAS*) CTENIDIA METHYLOMES

4.1 Abstract

Production of polyploid organisms resilient to environmental change requires a molecular understanding of stress response mechanisms. This is especially important when considering the use of triploid oysters in commercial aquaculture settings. Examination of DNA methylation can elucidate gene expression regulation mechanisms employed by triploid oysters, and how they may differ from their diploid counterparts. In other cultured marine organisms, polyploid induction does not yield methylation variation similar to organisms that undergo natural genome duplication. We examined how artificial genome duplication and environmental variation contribute to stress response phenotypes in the Pacific oyster (*Crassostrea gigas*). We exposed diploid (n = 12) and triploid (n = 12) *C. gigas* to low or high pH stress for six weeks. Ctenidia methylomes were then obtained using whole genome bisulfite sequencing. While there were no global methylation differences between samples, distinct differentially methylated loci (DML) were associated with ploidy status (154 ploidy-DML) and pH condition (178 pH-DML). Functions of genes containing ploidy-DML were not similar to those observed in organisms that undergo natural polyploidization. However, genes with pH-DML were associated with apoptosis, protein ubiquitination, zinc ion binding, and cytoskeletal processes commonly observed in oyster response to ocean acidification. Given the enrichment of both ploidy- and pH-DML in genes with multiple transcripts, it is possible that these methylation changes can influence gene expression and physiological phe-

notype. Our study — which is the first to provide single base-pair resolution assessment of diploid and triploid methylomes in a marine invertebrate — suggests that while methylation may play different regulatory roles during artificial and natural genome duplication events, environmentally-sensitive methylome responses may be uniform between diploid and triploid *C. gigas.*

4.2 Introduction

As molluscan breeding programs focus on developing oyster lines resilient to environmental stressors associated with climate change, several programs are focusing on culturing triploid organisms. Triploid oysters, which may be generated through chemical induction or by mating diploid and tetraploid parents, contain three copies of the genome (**Nell2002?**). This third genome copy renders these oysters partially sterile, with triploids diverting resources that would be for reproduction towards growth (**Nell2002?**). As a result, triploids exhibit significantly faster growth rates, improved meat quality, and increased survival, with mated triploids having an additional advantage over chemically-induced oysters (**Wadsworth2019?**). The potential advantages of triploidy have been examined with respect to heat stress, salinity, hypoxia, and disease pressure in *Crassostrea spp.* While triploid Pacific oysters (*Crassostrea gigas*) may have an energetic advantage over their diploid counterparts at elevated temperatures due to reduced gonadal investment and corresponding increases in tissue weight, condition index, protein, and carbohydrate levels (**Shpigel1992?**), triploid eastern oysters (*C. virginica*) are more sensitive to desiccation stress associated with farm rearing, and display higher mortality than their diploid counterparts (**Bodenstein2021?**). Triploid hybrids of the edible oyster *Crassostrea hongkongensis* and Suminoe oyster *Crassostrea ariakensis* grew faster as larvae in the hatchery and as juveniles in environmentally-distinct outplant locations, but displayed less salinity tolerance than their diploid hybrid counterparts (**Qin2020?**). Low salinity conditions in the Chesapeake Bay removed growth advantages associated with *C. virginica*

triploids (**Callam2016?**). While triploid *C. virginica* juveniles did not display any survival differences in hypoxic conditions, time to mortality in *C. ariakensis* triploids was three times earlier than diploid oysters (**Lombardi2013?**). Active gametogenesis in *C. gigas* lead to increased *Vibrio* infection, irrespective of ploidy status, but triploid oysters had poorer survival after disease challenges in winter conditions (**DeDecker2011?**). In the Chesapeake Bay, triploid *C. virginica* had lower mortality and greater shell height, weight, yield, and a higher proportion of market-sized oysters than diploids, even when faced with similar *Perkinsus marinus* disease pressure (**Dégremont2012?**). However, triploid *C. virginica* were susceptible to mortality events that are not associated with disease or environmental conditions, suggesting that there are elements of triploid physiology and stress response still unknown (**Matt2020?**). To date, there is no documentation of triploid oyster responses to ocean acidification, and how this stress response differs from diploid organisms. Even though triploids are more sensitive to a variety of environmental stressors, without a mechanistic understanding of triploid sensitivity to ocean acidification, it is difficult to implement hatchery techniques such as environmental manipulation to produce more stress-tolerant triploid phenotypes (Gavery & Roberts, 2017).

Epigenetic mechanisms, like DNA methylation, can mediate phenotypic plasticity and responses to environmental stressors, which makes them useful for understanding triploid stress tolerance. Epigenetics refers to changes in gene expression that do not arise from changes in the DNA sequence, with methylation of cytosine bases being the most studied mechanism (Bird, 2002; Deans & Maggert, 2015). Baseline variation in methylation patterns, or standing epigenetic variation, can be associated with genomic structure. Currently, polyploidization impacts on DNA methylation is well-understood in organisms that demonstrate natural ploidy variation. Genome duplication events in plants are subsequently followed by sequence loss, gene duplication, transposon silencing, and changes to genome regulation mediated by epigenetic mechanisms (**Adams2005?**; **Chen2006?**; **Henry2005?**; **Matzke1999?**; **Salmon2005?**; **Verhoeven2010?**; **Wang2004?**). These methylation changes promote nor-

mal genomic function after polyploidization, and can persist for a few generations after the duplication event (**Wang2004?**). Social insects with natural genome duplication also display distinct methylation variation associated predominantly with ploidy status, which may be responsible for reducing transcriptional noise in haploid organisms (**Glastad2014?**). Organisms that experience induced triploidy may not display methylation variation associated with genome duplication. For example, examination of diploid and artificially-induced *C. gigas* and brown trout (*Salmo trutta* L.) did not find any global methylation differences (**Covelo-Soto2015?**; **Jiang2016?**). Even so, promoter sequences in triploid *Apostichopus japonicus* sea cucumbers exhibited correlation between methylation and expression (**Han2021?**). As standing methylation can still influence gene expression in triploid organisms, investigating these mechanisms can provide insight into triploid oyster gene expression regulation.

Ocean acidification, can also influence methylomes (Eirin-Lopez & Putnam, 2018), with pteropods (**Bogan2020?**), corals (**Liew2018?**; **Putnam2016?**), and urchins (**Strader2020?**; **Strader2019?**) demonstrating distinct methylation patterns in response to low pH. Diploid oysters have been well-studied in this respect. Eastern oyster (*C. virginica*) mantle (**Downey-Wall2020?**) and gonad (**Venkataraman2020?**), and *C. hongkongensis* mantle (**ChandraRajan2021?**), and larvae (**Lim2020?**) all display changes in DNA methylation after experimental ocean acidification exposure. However, no studies to date have examined environmentally-sensitive methylation in triploids. In plants, polyploidization can influence the standing methylation variation that evolves over time and interacts with environmental stressors (**Verhoeven2010?**). Transcriptomic investigations have started to unveil differences in diploid and triploid stress responses in molluscan species. For example, triploid abalone exposed to heat stress and hypoxia conditions exhibited significantly different transcriptional changes when compared to diploids (**Kim2021?**). Since DNA methylation could influence gene activity and impact stress responses, elucidating genomic and environmental variation contribute to phenotype is crucial for understanding potential acclimatization mechanisms and their application to

triploid aquaculture (Gavery & Roberts, 2017).

We present the first single-base pair resolution examination of diploid and triploid *C. gigas* methylomes in response to an environmental stressor. Our comparison of DNA methylation by ploidy status allows us to discern how epigenetic variation differs with genome duplication. By comparing oyster methylomes exposed to low and high pH conditions, we can elucidate how ocean acidification can exert an influence over ctenidia methylation. Finally, we determine if genetic and environmental variation interact to produce a unique methylation phenotype. Our work provides the foundation for understanding the mechanisms behind triploid *C. gigas* stress responses, and the application of knowledge for aquaculture practices that select for oysters resilient to climate stressors.

4.3 Methods

4.3.1 Experimental Design

Pacific oysters (n = 24) were obtained from the Hawaiian Shellfish LLC hatchery located in Kea'au, Hawai i Island and held at the University of Hawai'i, Hilo. The broodstock used to produce these oysters originated from the Molluscan Broodstock Program at Oregon State University. Triploids were produced by mating diploid female and tetraploid male oysters. The diploids used in the experiment were not used to produce the triploids. Diploid (n = 12) and triploid (n = 12) *C. gigas* were randomly assigned to each of two treatments (high pH and low pH).

The treatment tanks were 150 gallon conical bottom tanks with flow-through seawater from a deep well. Carbon dioxide was dispersed into both treatment tanks to achieve "non-acidification" conditions (pCO_2 ranged from 67.37 to 98.61 µatm) to "ocean acidification" conditions (pCO_2 ranged from 287.88 to 418.20 µatm). Water quality parameters were checked daily. pH probes were placed about 2.5 cm under the surface of the water and connected to pH controllers (American Marine). Dissolved oxygen, salinity, and temperature

were monitored daily with an YSI Pro 2030. Ammonia concentration was monitored daily using an aquarium test kit (API). Full water exchanges were performed daily. Since the ctenidia is the primary tissue that interacts with the environment, ctenidia tissue was obtained from each oyster after six weeks, preserved in 2 mL DNA/RNA Shield (ZymoResearch), then stored at -20ºC.

4.3.2 DNA Extraction and Library Preparation

DNA was extracted from each sample for Whole Genome Bisulfite Sequencing (WGBS). Samples were centrifuged for one minute at 10,000 RCF to pellet ctenidia tissue. After removing the DNA/RNA shield, ctenidia DNA was extracted using the E.Z.N.A. Mollusc Kit (Omega). Manufacturer's protocol was followed, with the exception of an overnight incubation at 37ºC after the addition of 350 µL ML1 Buffer and 25 µL Proteinase K Solution. After the MB1 lysis buffer was added to samples, a thick white precipitate formed. Samples were vortexed to emulsify the precipitate, but several samples did not lyse fully. Samples were eluted with 100 µL of Elution Buffer. A Qubit™ 3.0 and dsDNA BR Assay (Invitrogen) was used to confirm that samples contained at least 500 ng of DNA, using 2 µL of DNA for each sample.

Samples were sent to ZymoResearch for bisulfite conversion and library preparation. Bisulfite conversion was performed with the Pico Methyl-Seq Library Prep Kit (Cat. #D5455). Per ZymoResearch protocol, samples were spiked with a mixture of six unique double-stranded synthetic amplicons (180-200 bp) from the *Escherichia coli* K12 genome. Since each amplicon represents a different percent methylation, the spike-in was used to determine bisulfite conversion efficiency. Libraries were barcoded and pooled into a single lane to generate 150 bp paired-end reads in a single Illumina NovaSeq flowcell.

4.3.3   Genome Information and Feature Tracks

Nuclear (**Peñaloza2021?**) (NCBI GenBank: GCA_902806645.1, RefSeq: GCF_902806645.1) and mitochondrial genome sequences (NCBI Reference Sequence: NC_001276.1) were used in downstream analysis. These sequences were combined for downstream alignment. The `fuzznuc` function within EMBOSS was used to identify CG motifs in the combined nuclear and mitochondrial sequences.

The RefSeq annotation was used to obtain genome feature tracks, and create additional tracks. The annotation included a combination of Gnomon, RefSeq, cmsearch, and tRNAscan-SE predictions. Gene, coding sequence, exon, and lncRNA tracks were extracted from the RefSeq annotation. These tracks were then used to obtain additional genome feature information. To create an intron track, the complement of the exon track was generated with BEDtools `complementBed` v.2.26.0 to create a non-coding sequence track (**Quinlan2010?**). The overlap between the genes and the non-coding sequences was found with `intersectBed` and defined as introns. Similarly, untranslated regions of exons were obtained by subtracting coding sequences from exon information with `subtractBed`. Flanking regions were defined as the 1000 bp upstream or downstream of genes. Upstream flanks were generated by adding 1000 bp upstream of genes, taking into account strand, with `flankBed`. Existing genes were removed from flanking sequences using `subtractBed`. A similar process was performed to generate downstream flanking regions. Intergenic regions were isolated by finding the complement of genes with `complementBed`, then using `subtractBed` to remove any flanks. Transposable element locations were obtained from RepeatMasker using the NCBI RefSeq annotation (RefSeq: GCF_902806645.1) (**Smit2013?**; **Tarailo-Graovac2009?**). The number of CG motifs in a genome track was obtained using `intersectBed`.

4.3.4   Sequence Alignment

Prior to alignment, 150 bp paired-end reads generated for each sample underwent three rounds of trimming. First, Illumina adapter sequences were removed from samples using TrimGalore! v.0.6.6 (**Martin2011?**). Additionally, 10 bp were trimmed from both the 5' (`--clip_R1 10` and `--clip_R2 10`) and 3' ends of paired reads (`--three_prime_clip_R1 10` and `--three_prime_clip_R2 10`). Remaining adapters were then removed from sequences in the second round of TrimGalore! trimming. Poly-G tails present in data were removed by manually specifying a matching adapter sequence (`--adapter GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG` and `--adapter2 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG`). After each round of trimming, sequence quality was assessed using FastQC v.0.11.9 (**Andrews2010?**) and MultiQC reports (**Ewels2016?**).

Trimmed samples were then aligned to the *C. gigas* genome (**Peñaloza2021?**). The full genome was bisulfite converted with Bowtie 2-2.3.4 (Linux x84_64 version (**Langmead2012?**)) using the `bismark_genome_preparation` function within Bismark v.0.22.3 (Krueger & Andrews, 2011) prior to sequence alignment. Reads were then aligned to each bisulfite-converted genome, specifying non-directional input (`--non_directional`) and allowing for an alignment score no lower than -90 (`-score_min L,0,-0.6`). Aligned BAM files were deduplicated with the `--deduplicate_bismark` function, then sorted and indexed for downstream applications using SAMtools v.1.10 (**Li2009?**). Sequence methylation calls were extracted from deduplicated BAM files with `bismark_methylation_extractor`. Coverage files generated from methylation extraction contained CpG information from both strands. To consolidate CpG information between strands and report 1-based chromosomal coordinates for each CpG, `coverage2cytosine` was run on deduplicated coverage files. Using `coverage2cytosine` coverage file output, genomic coordinates for CpGs with at least 5x coverage across all samples were written to BEDgraphs. Finally, `bismark2report` and `bismark2summary` were used to create summary reports with alignment information for

each sample that were concatenated with MultiQC.

4.3.5 SNP Identification

As genomic differences can play a role in methylation, single nucleotide polymorphisms (SNPs) were identified from WGBS data. Two methods were used for SNP identification. First, SNP variants across all samples and in individuals were identified with BS-Snper (**Gao2015?**). As `BS-Snper` only takes one input file, deduplicated and sorted BAM alignments generated with Bismark v.0.22.3 and SAMtools v.1.10 (**Krueger2011?**; **Li2009?**) were merged (`samtools merge`). Variants were then identified in the merged BAM file or individual deduplicated and sorted BAM files. Default `BS-Snper` settings were used, except for a minimum coverage of 5x to be consistent with methylation analyses. C/T SNPs were filtered by identifying overlaps (`intersectBED`) between each set of SNPs and the CG motif track.

4.3.6 Global Methylation Characterization

Global methylation patterns were characterized using CpG loci with at least 5x coverage in all samples. BEDgraphs produced with Bismark v.0.22.3 were sorted (`sortBed`), then used to concatenate percent methylation information across samples (`unionBedGraphs`). Percent methylation values were averaged between all samples for each CpG locus. These loci were categorized as highly (50%), moderately (10-50%), or lowly methylated (10%) based on average percent methylation values. The genomic location of these methylation categories was determined with respect to UTR, CDS, introns, up- and downstream flanking regions, transposable elements, and intergenic regions (`intersectBed`). We tested the null hypothesis that there was no association between the genomic location of CpG loci and methylation status (all CpGs versus highly methylated CpGs) with a chi-squared contingency test (`chisq.test` in R Version 3.5.3). To quantify differences in percent methylation by ploidy status, percent

methylation values from the union BEDgraph were averaged for diploid and triploid samples separately.

As methylation detection relies on coverage, we also estimated average coverage for diploid and triploid oysters. Merged coverage files produced with Bismark v.0.22.3 were sorted (`sortBed`), then used to concatenate coverage information across samples (`unionBedGraphs`). Coverage was averaged for diploid and triploid samples separately to quantify if ploidy status influenced coverage.

Methylation patterns were compared between treatments using `methylKit` v.1.17.4 (**Akalin2012?**) with R Version 3.5.3. Data with at least 2x coverage from merged CpG coverage files produced by `coverage2cytosine` were imported using `methRead`. For each sample, data was filtered further (`filterByCoverage`) by requiring at least 5x coverage for each CpG (`lo.count = 5`), and removing potential PCR duplicates by excluding data in the 99.9th percentile of coverage (`high.perc = 99.9`). Filtered data were then normalized (`normalizeCoverage`) between samples to avoid over-sampling reads from one sample during downstream statistical analysis. Percent methylation (`getMethylationStats`) and CpG coverage (`getCoverageStats`) histograms were generated for each sample to confirm normalization. Filtered and normalized data were combined (unite) for every CpG position with at least 5x coverage with data in at least eight samples per treatment to generate a CpG background. Pairwise Pearson's correlation coefficients were calculated and a global Principal Component Analysis based on percent methylation profiles was conducted to understand similarities and differences in sample methylation.

4.3.7 Identification of Differentially Methylated Loci

Differentially methylated loci (DML) were identified using a generalized experimental design with `DSS` v.2.38.0 (Feng et al., 2014). The purpose of this generalized design was to test pH, ploidy, and their interaction simultaneously. For each sample, the corresponding merged coverage file from Bismark v.0.22.3 were modified to meet DSS input formatting requirements.

The coverage file contained information for CpG loci with at least 1x coverage in that sample. Only the start position in the coverage files was used, and the number of total reads was calculated by summing the number of methylated and unmethylated reads provided in the coverage files. The final input files included the following columns: chromosome, start position, number of total reads, and number of methylated reads. Data was not pre-filtered based on a coverage threshold, following DSS input specifications. After importing information into R Studio using `list2env`, a BSseq object (`makeBSseqData`, `bsseq` v.1.26.0) was generated with coverage data to use for model fitting. To define the experimental design, a data frame was created with ploidy status ("D" = diploid vs. "T" = triploid), pH treatment ("H" = high pH vs. "L" = low pH), and sample ID.

Methylation was calculated at each CpG locus using the Bayesian hierarchical model framework provided by `DSS`. Coverage data were modeled using a beta-binomial distribution with an arcsine link function and log-normal priors (**Feng2014?**). Unlike the Fisher's exact test used by `methylKit`, `DSS` assumes that data from each sample do not have identical distributions, retaining biological variation between samples in a treatment. The model was fitted onto transformed data using the generalized least squares method, defining pH, ploidy, and their interaction as coefficients (`DMLfit.multiFactor, formula = ~ ploidy + pH + ploidy:pH`). Smoothing, or using information from nearby CpG sites to inform methylation calculations, was not used, as invertebrate methylation data does not tend to be dense enough to provide multiple CpG dinucleotides in close proximity to each other.

A Wald test (`DMLtest.multiFactor`) was used at each CpG site for statistical testing of differential methylation for each coefficient separately (`ploidyT`, `pHL`, `ploidyT:pHL`). The test itself depends on sequencing depth, and a dispersion parameter defined by a shrinkage estimator from the Bayesian hierarchical model. The DML (ploidy-DML, pH-DML, and interaction-DML) were defined based on a P-value < 0.05 and Benjamini-Hochberg FDR < 0.01. The multiple regression framework used does not provide methylation values, so percent methylation differences were not used to define DML from `DSS`. BEDfiles with DML

information were created for downstream analysis and visualized with sample coverage information using IGV.

4.3.8   DML Characterization

Overlaps between all DML BEDfiles and UTR, CDS, introns, up- and downstream flanks, transposable elements, and intergenic regions (`intersectBed`) were used to describe where DML were located in the *C. gigas* genome. A chi-squared contingency test was used to test the null hypothesis of no association between methylation and specific genome features using all CpGs with 5x data and DML. The number of DML in each chromosome was quantified and divided by chromosome length to discern if DML were distributed uniformly across the genome.

To determine if ploidy and pH treatments elicited similar methylation responses, we identified loci common between pH-DML, ploidy-DML, and interaction-DML. Overlaps between these lists were obtained using intersectBed. Additionally, overlaps between C/T SNPs and DML were also identified to discern if genomic factors other than treatment contributed to methylation responses.

4.3.9   Functional Enrichment Analysis

A functional enrichment analysis was performed to identify gene ontology (GO) terms significantly overrepresented in genes with ploidy-, pH-, or interaction-DML. Prior to enrichment, a blastx was performed against the Uniprot-SwissProt database (accessed June 01, 2021) (**UniProt2019?**) to obtain Uniprot Accession information for each RNA nucleotide sequence from the genome (GCF_902806645.1_cgigas_uk_roslin_v1_rna_from_genomic.fna available in the NCBI Annotation Release 102). The Uniprot Accession codes were used to match transcript IDs from the blastx output with GO terms from the Uniprot-Swissprot database.

Gene enrichment analysis was conducted with `topGO` v.2.34.0 in R (**Alexa2010?**). First, a gene ID-to-GO term database (`geneID2GO`) was created for manual GO term annotation. The list of transcript IDs and corresponding GO terms generated during annotation was filtered to only include transcript IDs for genes with at least 1x coverage in the WGBS dataset. These transcript IDs were filtered such that they did not include any transcripts associated with DML-SNP overlaps. Each line of the database contained transcript ID in one column, and all corresponding GO terms in another. The list of transcript IDs from this database was used as the gene universe for enrichment. Genes of interest corresponded to the transcript IDs from individual DML lists. A `topGO` object was generated for each DML list and GO category (biological process, cellular component, or molecular function), with GO term annotation performed using the `geneID2GO` database. A Fisher's exact test was used to identify GO terms in each DML list significantly enriched with respect to the gene background (P-value < 0.05). Clustering of enriched GO terms was conducted using `simplifyEnrichment` v.1.2.0 default settings (**Gu2021?**). The "Relevance" similarity measure and GO tree topology (`GO_similarity`) were used to create a semantic similarity matrix from enriched GO IDs. The matrix was then clustered using the binary cut method to identify broad GO term categories impacted by environmentally-responsive methylation.

4.4 Results

4.4.1 Sequence Alignment

Whole genome bisulfite sequencing produced 2.5 x 10^9 total 150 bp paired-end reads. After trimming, 2.4 x 10^9 total reads remained. Poly-G tails were present in the second paired read for all samples, and constituted 1-2% of reads removed. Of the trimmed paired-end reads, 7.4 x 10^8 (61.0%) were uniquely aligned to the *C. gigas* genome and appended mitochondrial sequence. Across samples, 3.5 x 10^8 cytosines (11.6%) were methylated in a CpG context. Coverage between diploid and triploid samples were distributed similarly (Figure 4.1A-B). Of the 12,794,074 total CpG loci with 1x coverage in at least one sample, 5,395,386 had

higher average coverage in triploids, and 6,956,658 had equal or lower average coverage in triploids.

4.4.2 SNP Identification

Possible SNP variants were identified in WGBS data to determine if genomic differences contributed to differential methylation. A total of 14,857,859 unique SNPs, including 289,826 unique C/T SNPs, were found in WGBS data and used in downstream analyses.

4.4.3 Global Methylation Characterization

Global methylation patterns were first characterized using 10,497,320 CpG loci (79.2% of 13,246,547 total CpGs in *C. gigas* genome) with at least 5x coverage in each sample from the union BEDgraph. Of these loci, 989,453 (9.4%) were highly methylated (50%), 848,655 (8.1%) were moderately methylated (10-50%), and 8,659,212 (82.5%) were lowly methylated. Both diploid and triploid samples were predominantly lowly methylated (Figure 4.1C-D), with 2,735,188 loci having higher percent methylation in triploids and 6,605,916 loci with equal or lesser percent methylation in triploids. The highly methylated CpGs were found primarily in genes (932,568; 94.3%), with 526,656 (53.2%) genic CpGs in introns, followed by 353,333 (35.7%) in CDS, then 55,067 (5.6%) CpGs in exon UTR (Figure 4.2). More highly methylated CpGs were found in transposable elements (295,204; 29.8%) than intergenic regions (32,028; 3.2%), downstream flanking regions, (21,839; 2.2%), lncRNA (17,023; 1.7%), or upstream flanking regions (4,683; 0.5%). The proportion of highly methylated CpGs in various genome features was significantly different than the proportion of 5x CpGs in those same features (Appendix 1).

Methylation patterns were compared between ploidy and pH treatments using CpG loci with at least 5x coverage and eight samples per ploidy treatment in the CpG background. Pairwise Pearson's correlation coefficients were consistently between 0.87 and 0.93 for all

samples (Appendix 1). A global Principal Component Analysis did not separate samples based on pH treatment or ploidy status (Figure 4.1E).

4.4.4 Differential Methylation Characterization

Using a Bayesian hierarchical modeling approach, a total of 178 pH-DML, 154 ploidy-DML, and 53 interaction-DML were identified from 12,794,074 total CpG loci with 1x coverage in at least one sample (Figure @ref(fig: multiheatmap)A-C). Of these DML, 21 overlapped between pH and ploidy treatments, 11 between pH and ploidy:pH, 17 between ploidy and ploidy:pH. Additionally, 25 DML were common between all lists. A small number of DML — with five (2.8%) pH-DML, three (1.9%) ploidy-DML, and one (1.9%) interaction-DSS — were present in unique C/T SNPs. The DML were identified in the ten main chromosomes, as well as other unplaced scaffolds (Appendix 1). The ninth chromosome contained the most DML even though it has the fewest genes (Figure @ref(fig: multiheatmap)D).

The DML were primarily in genes (Figure @ref(fig: DMLgenomicloc)). For pH-DML, 123 (69.1%) DML were found in 94 unique genes, the majority of which were in intron sequences (104; 58.4%), followed by CDS (15; 8.4%) and exon UTR (4; 2.2%). Ploidy-DML and interaction-DML were similarly distributed between introns (ploidy: 114; 74.0% and interaction: 41; 77.4%), CDS (ploidy: 26; 16.9% and interaction: 6; 11.3%), and exon UTR (ploidy: 8; 5.2% and interaction: 3; 5.7%). Ploidy-DML and interaction-DML were also predominantly in 109 and 29 unique genes, respectively (ploidy: 145; 94.2% and interaction: 48; 90.1%). Most genes contained only one pH-DML, ploidy-DML, or interaction-DML. However, several genes contained multiple DML, with genes containing up to 8 pH-DML, 8 ploidy-DML, or 23 interaction-DML (Appendix 1).

Transposable elements contained the second highest amount of DML (pH: 86; 48.3%, ploidy: 66; 42.9%, interaction: 14; 26.4%), followed by intergenic regions (pH: 27; 15.2%, ploidy: 18; 11.7%, interaction: 5; 9.4%). Few DML were found in lncRNA (pH: 4; 2.2%, ploidy: 1; 0.6%, interaction: 2; 3.8%), upstream flanking regions (pH: 2; 1.1% and ploidy: 8; 5.2%), or

downstream flanks (pH: 2; 1.1% and ploidy: 8; 5.2%). No interaction-DML were found in either flanking region. The distribution of DML in the genome differed significantly from that of all CpGs with 5x data: specifically, the number of ploidy-DML in introns and intergenic regions, the number of pH-DML in introns, transposable elements, and intergenic regions, and the number of interaction-DML in introns and intergenic regions (Appendix 1).

4.4.5 Functional Enrichment Analysis

A series of Fisher's exact tests identified 21 significantly enriched GOterms total in genes that do not overlap with C/T SNPs (Table 4.1). A total of 84 genes annotated with GOterms containing 104 ploidy-DML were used to identify three enriched biological process terms associated with dephosphorylation and cell-cell binding (Appendix 1). The enriched GOterms were found in eight unique genes and 29 unique transcripts (Appendix 1). The remaining enriched GOterms were associated with pH-DML. For pH-DML, 69 genes with GOterm annotations and containing 87 DML were used for the analysis. Ten enriched biological process terms were identified, and were related to apoptotic processes (Appendix 1). An additional eight molecular function terms were related to myosin and cytoskeletal binding, ubiquitin activity, and zinc ion binding (Appendix 1).

4.5 Discussion

Understanding how ploidy and environmental variation interact to shape phenotype is important for aquaculture operations moving forward with developing polyploid lines resilient to climate stressors like ocean acidification. Our investigation of methylation in diploid and triploid *C. gigas* is the first single-base pair resolution study to examine the influence of ploidy on DNA methylation in a mollusc species. Additionally, we discern the influence of pH and its interaction with ploidy on methylation. Our results provide insight into stress responses in triploid oysters, which may be leveraged to improve triploid aquaculture production.

4.5.1   Small-scale methylation differences associated with polyploidization

The lack of global methylation differences observed between diploid and triploid *C. gigas* suggests that genome duplication does not induce broad changes in molluscan methylation. Conversely, organisms that naturally undergo polyploidization may exhibit global methylation changes associated with ploidy status. Methylation patterns in haplodiploid ants (*Solenopsis invicta*) are highly divergent between haploid and diploid individuals, irrespective of morphology or caste behavior (**Glastad2014?**). Haploid ants also exhibit consistently higher methylation at sequenced loci. Naturally polyploid loaches (*Misgurnus anguillicaudatus*) exhibit significant differences in methylation levels between diploids and triploids and diploids and tetraploids (**Zhou2016?**). In our experiment there were no significant differences between global methylation profiles in diploid and triploid *C. gigas*, which implies methylation regulation may function differently at different ploidy states for naturally polyploid organisms. Since triploidy is not a part of the natural reproductive process in oysters, it is possible that *C. gigas* has not evolved global methylation patterns associated with the triploid state. Methylation profiling with MSAP did not detect global methylation differences in between diploid and artificially-induced triploid *C. gigas* or *S. trutta* L. (**Covelo-Soto2015?**; **Jiang2016?**). While MSAP may only detect a subset of epigenetic variation, our WGBS results corroborate these results. The mating of diploid and tetraploid oysters to produce the triploid *C. gigas* used in our study itself may have introduced a higher level of standing epigenetic variation than what may be expected through natural polyploidization. Future studies should examine how different methods of triploid generation — mating versus chemical induction — influence methylation patterns in molluscs, and investigate genetic resistance to duplication processes.

Although we did not detect global methylation differences associated with polyploidization, we found individual CpG loci differentially methylated between diploid and triploid oysters. The identification of ploidy-DML suggests a role for methylation regulation of genome duplication, irrespective of pH condition. In plant species, DNA methylation regulates du-

plicated genes and transposable element silencing, and is generally involved in normal function and maintenance of polyploid genomes (**Adams2005?**; **Chen2006?**; **Henry2005?**; **Matzke1999?**; **Salmon2005?**; **Verhoeven2010?**; **Wang2004?**). Gene ontology enrichment revealed enrichment of cell-cell adhesion and protein dephosphorylation processes. These processes are not explicitly related to the transcriptional control processes observed during genome duplication in plants, and to our knowledge have not been documented in association with polyploidization in any species. However, DML in these genes may still modify gene expression, which could then influence tissue structure and function.

The concentration of ploidy-DML in genes, specifically introns, implies that these DML may control gene expression. Additionally, enriched GOterms associated with ploidy-DML were found in genes with multiple transcripts, suggesting gene expression regulation may occur through alternative splicing (Roberts & Gavery, 2012; **Glastad2014?**). We also found several ploidy-DML in transposable elements, which aligns with epigenetic regulatory pathways observed after genome duplication in plants. However, it is still unclear if methylation silencing of transposable elements occurs in invertebrate species (Eirin-Lopez & Putnam, 2018). Interestingly, three (1.9%) ploidy-DML overlapped with unique C/T SNPs identified in WGBS data. While SNPs may only have a small influence on methylation patterns associated with ploidy status, additional research is needed to examine how genome duplication and variation interact to produce methylation profiles.

4.5.2   Role of methylation in stress response elucidated by pH- and interaction-DML

The enriched gene ontologies associated with genes containing pH-DML reveal the potential for plastic responses to low pH stress. Enrichment of apoptotic processes imply a need to regulate the cell cycle in response to ocean acidification. Differential regulation of apoptotic processes in oysters exposed to ocean acidification conditions was also found in *C. virginica* mantle tissue (**Downey-Wall2020?**) and larval *C. hongkongensis* (**Lim2020?**), and there is a broader association between DNA methylation and apoptosis pathways (**Meng2011?**).

Enrichment of molecular functions related to protein ubiquitination complement enrichment of apoptosis functions. Protein ubiquitination is a post-translational protein modification that is involved in protein synthesis and degradation (**Komander2009?**; **Peng2003?**) and has been documented in molluscan response to ocean acidification. For example, shotgun proteomic characterization of posterior gill lamellae from adult *C. gigas* exposed to high pCO$_2$ revealed increased abundance of proteins involved in ubiquitination and decreased protein degradation (Timmins-Schiffman et al., 2014). Elevated pCO$_2$ levels were also found to up-regulate malate dehydrogenase in adult *C. virginica* mantle tissue (**Tomanek2011?**). Genes involved in protein ubiquitination also displayed differential methylation in *C. virginica* gonad tissue (**Venkataraman2020?**), and displayed both differential methylation and gene expression downregulation in *C. hongkongensis* mantle tissue (**ChandraRajan2021?**). Degradation processes are also prominent in coral methylome responses to temperature and salinity variation (**Liew2018?**), urchin responses to temperature and pH (**Strader2020?**), and oyster responses to disease (**Johnson2020?**). Together, these results imply that regulation of molecular and cellular degradation processes are conserved across taxa in response to environmental change, and concurrent gene expression and methylation studies are required to elucidate how changes in these processes impact physiology.

Other overrepresented molecular functions in genes with pH-DML are also consistent with existing research examining molluscan response to ocean acidification. Enrichment of myosin binding functions suggest low pH impacts the methylome associated with cytoskeletal processes. Low pH also induced changes to methylation in cytoskeletal genes in larval *C. hongkongensis* (**Lim2020?**). Proteomic analysis of *C. gigas* mantle, *C. virginica* mantle, *C. hongkongensis* larvae, and *C. gigas* larvae found that ocean acidification exposure altered the abundance of various cytoskeletal proteins (Dineshram et al., 2016; Wei et al., 2015; **Tomanek2011?**; **Dineshram2012?**; **Dineshram2013?**; **Dineshram2015?**). Additionally, selectively-bred Sydney rock oysters (*Saccostrea glomerata*) with a history of ocean acidification exposure had higher expression of cytoskeletal genes than wild type oysters

(**Goncalves2016?**). Changes to cytoskeletal proteins, gene expression, and methylation could be reflective of a conserved response to establish homeostasis in response to ocean acidification. The remaining enriched molecular function GOterms were related to zinc ion binding. Larval *C. hongkongensis* exposed to low pH also demonstrated methylation changes to genes associated with this function (**Lim2020?**). Since zinc ion binding is necessary for several enzymes and signaling pathways, methylation regulation of this function could represent a generalist response to ocean acidification.

Our finding that low pH impacted *C. gigas* ctenidia methylation at specific loci is consistent with previous studies investigating the influence of ocean acidification on invertebrate methylomes. However, there appear to be species- and tissue-specific methylome responses regarding the exact location of genic responses. The pH-DML and interaction-DML were identified primarily in introns, followed by CDS and untranslated regions of exons. Baseline assessment of *C. gigas* ctenidia methylome found higher levels of methylation in exons, but did note a larger proportion of intronic methylation in comparison to other invertebrate methylomes (Gavery & Roberts, 2013). Similarly, low pH-induced differential methylation in *C. hongkongensis* mantle tissue was more common in introns (**ChandraRajan2021?**). In contrast, adult *C. virginica* mantle and gonad methylomes (**Downey-Wall2020?**; **Venkataraman2020?**) and larval *C. honkongensis* demonstrate a higher concentration of DML associated with ocean acidification in exons over introns. If methylation does regulate gene expression through alternative splicing, then the presence of DML in introns versus exons may impact which components are retained in the alternatively-spliced transcript. To date, pH-responsive methylation in *Crassostrea spp.* have not corresponded with changes in gene expression (**ChandraRajan2021?**; **Downey-Wall2020?**). Additional research is needed in all three related species to determine how differential methylation in coding and non-coding sequences influence gene expression, alternative splicing, and protein abundance, and the influence of tissue type or developmental stage on those relationships.

The presence of interaction-DML suggests genomic and environmental variation can shape

the ctenidia methylome and subsequent phenotypes. We predicted that the triploid methylome would be more impacted by pH, as triploids can be more sensitive to environmental stressors. This sensitivity would have been associated with distinct gene ontologies involved in stress response. However, we found no significantly enriched gene ontologies associated with these DML. While ploidy and pH may interact, our results signify that this variation does not produce a clear physiological phenotype.

4.6 Conclusion

It is critical to understand the molecular mechanisms underpinning genomic regulation and environmental response in aquaculture species to improve production. Our single base-pair resolution sequencing of *C. gigas* ctenidia methylomes reveals distinct biological processes and molecular functions impacted by ploidy status and low pH exposure. In particular, pH influenced molecular and cellular degradation processes in both diploid and triploid oysters while ploidy had a smaller impact on cell binding. Interestingly, the lack of gene ontology enrichment in methylation responses associated with both ploidy and pH suggests that genomic and environmental variation have a lesser effect on physiology together than separately. Future studies should examine if the differential methylation observed in this study results in persistent changes in gene expression and stress tolerance phenotypes. If these DML can influence phenotype, environmental hardening in the hatchery may increase oyster resilience to low pH stressors and improve triploid production (Gavery & Roberts, 2017).

4.7 Tables

Table 4.1: Significantly enriched GOterms identified using a Fisher's exact test. There were 3 biological process terms associated with ploidy-DML, 10 biological process terms associated with pH-DML, and 8 molecular function terms associated with pH-DML. Biological process terms for ploidy-DML were broadly associated with cell-cell adhesion and dephosphorylation, while apopotis was enriched in pH-DML. Molecular functions overrepresented in genes with pH-DML were associated with protein ubiquitination, cytoskeletal proteins, and zinc ion binding.

Table 4.1: Enriched GO terms for ploidy- and pH-DML

DML	GO Category	GO ID	GO Term	Annotated Genes	Genes with DML	P value
ploidy	BP	GO:0006470	protein dephosphorylation	10	2	0.00870
ploidy	BP	GO:0098609	cell-cell adhesion	11	2	0.01050
ploidy	BP	GO:0016311	dephosphorylation	13	2	0.01460
pH	BP	GO:0043523	regulation of neuron apoptotic process	17	2	0.00062
pH	BP	GO:0051402	neuron apoptotic process	17	2	0.00062
pH	BP	GO:0070997	neuron death	20	2	0.00086
pH	BP	GO:1901214	regulation of neuron death	20	2	0.00086
pH	BP	GO:0042981	regulation of apoptotic process	37	2	0.00297
pH	BP	GO:0043067	regulation of programmed cell death	40	2	0.00346
pH	BP	GO:0006915	apoptotic process	41	2	0.00364
pH	BP	GO:0010941	regulation of cell death	42	2	0.00382
pH	BP	GO:0012501	programmed cell death	44	2	0.00418
pH	BP	GO:0008219	cell death	46	2	0.00456
pH	MF	GO:0017022	myosin binding	8	2	0.00021
pH	MF	GO:0061630	ubiquitin protein ligase activity	22	2	0.00173
pH	MF	GO:0061659	ubiquitin-like protein ligase activity	22	2	0.00173
pH	MF	GO:0008092	cytoskeletal protein binding	37	2	0.00488
pH	MF	GO:0004842	ubiquitin-protein transferase activity	41	2	0.00597
pH	MF	GO:0008270	zinc ion binding	41	2	0.00597
pH	MF	GO:0019787	ubiquitin-like protein transferase activ...	41	2	0.00597
pH	MF	GO:0046914	transition metal ion binding	41	2	0.00597

4.8 Figures

Figure 4.1: Coverage and global methylation information for diploid and triploid samples. Average coverage for A) diploid and B) triploids. Coverage was calculated using information for CpGs with at least 1x coverage. Percent methylation distribution for C) diploids and D) triploids. Percent methylation was calculated using information for CpGs with at least 5x coverage in all samples.
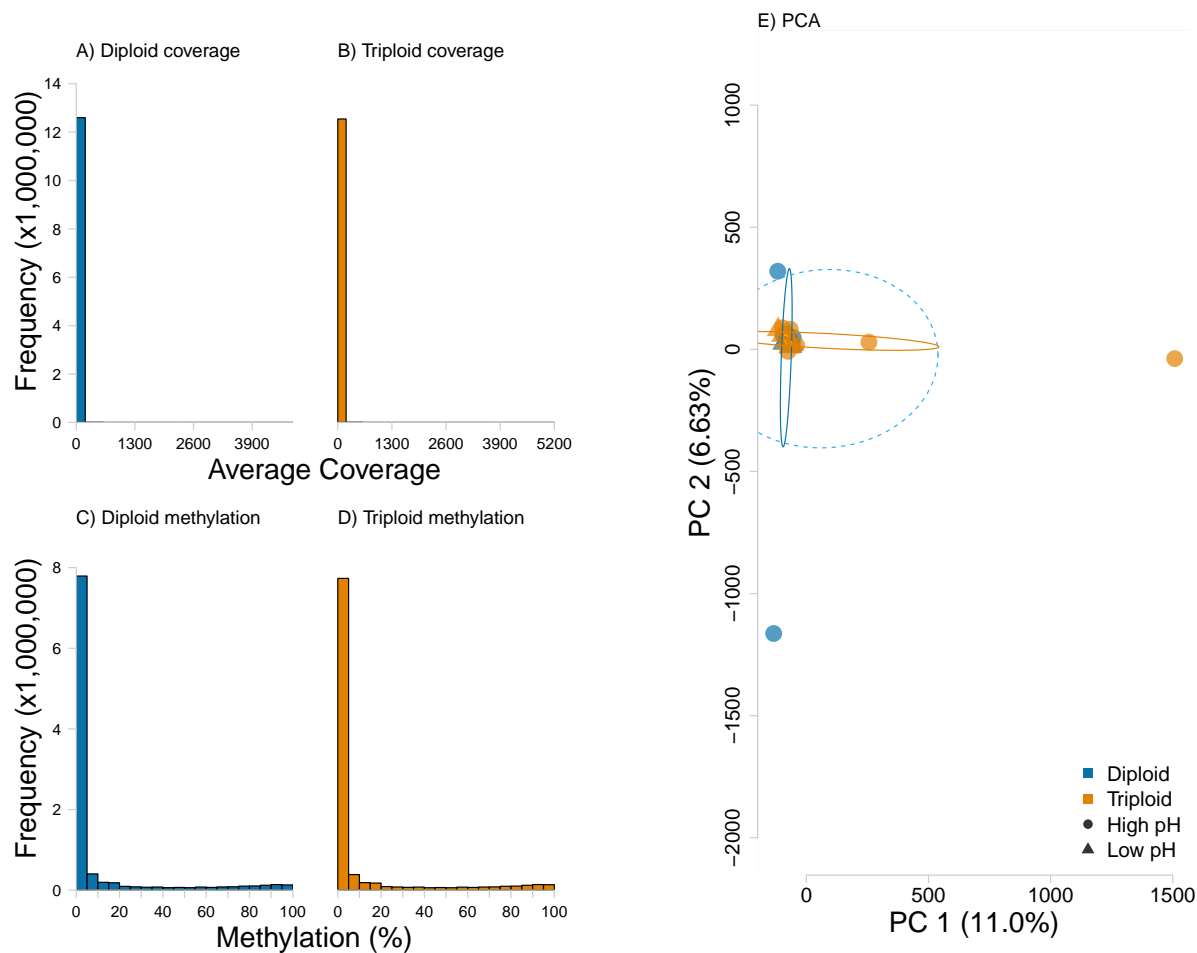

Figure 4.1: Diploid and triploid coverage and global methylation

Figure 4.2: Percent of total CpGs present in distinct *C. gigas* genome features for various CpG categories (from left to right: all CpGs with at least 5x coverage in all samples, highly methylated CpGs, moderately methylated CpGs, and lowly methylated CpGs).
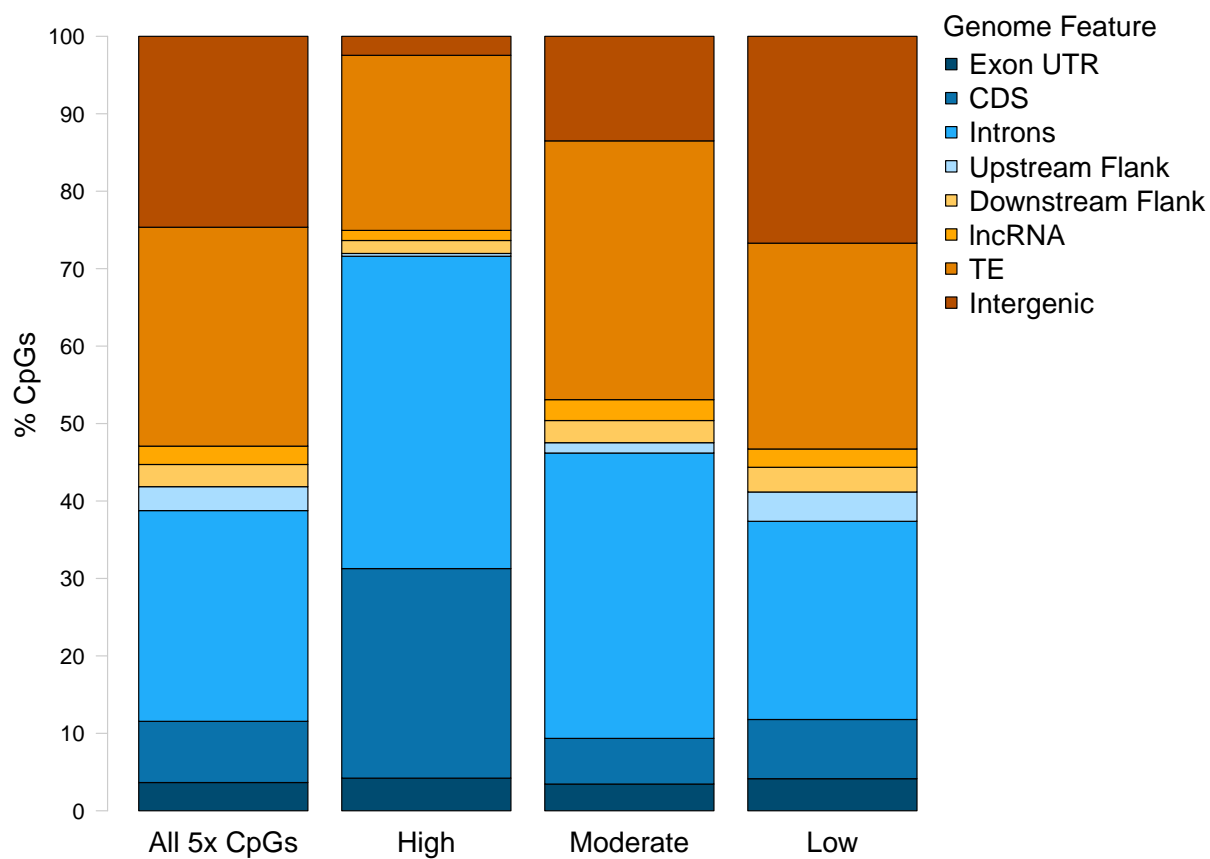
Figure 4.2: Genome distribution of various CpG categories

Figure 4.3: Heatmap of A) ploidy-, B) pH-, and C) interaction-DML identified, and D) distribution in *C. gigas* chromosomes. Density plots associated with heatmaps quantify the number of DML in various methylation bins.
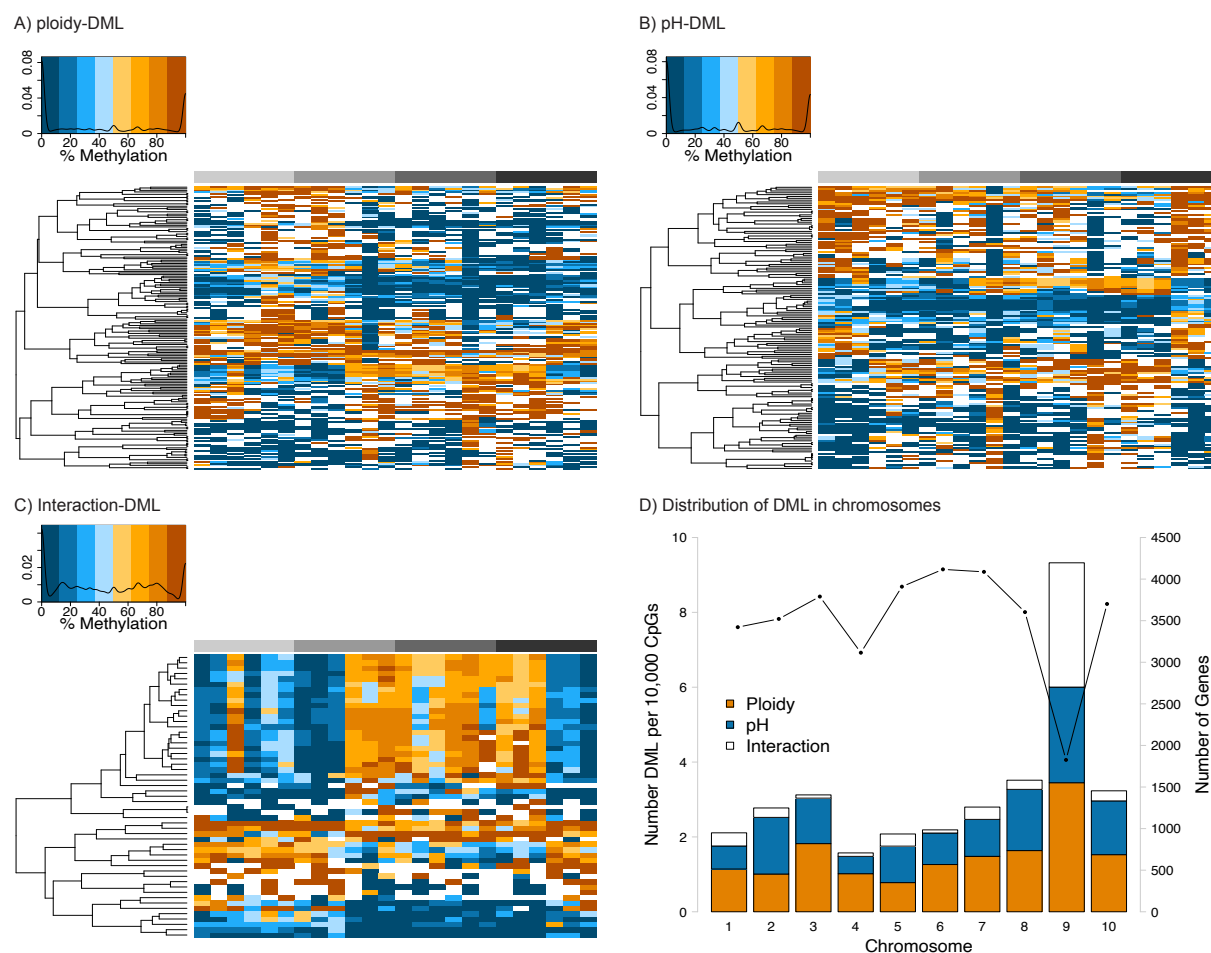
Figure 4.3: DML heatmaps and chromosomal distribution

Figure 4.4: Distribution of all 5x CpGs, ploidy-DML, pH-DML, and interaction-DML in various *C. gigas* genome features.
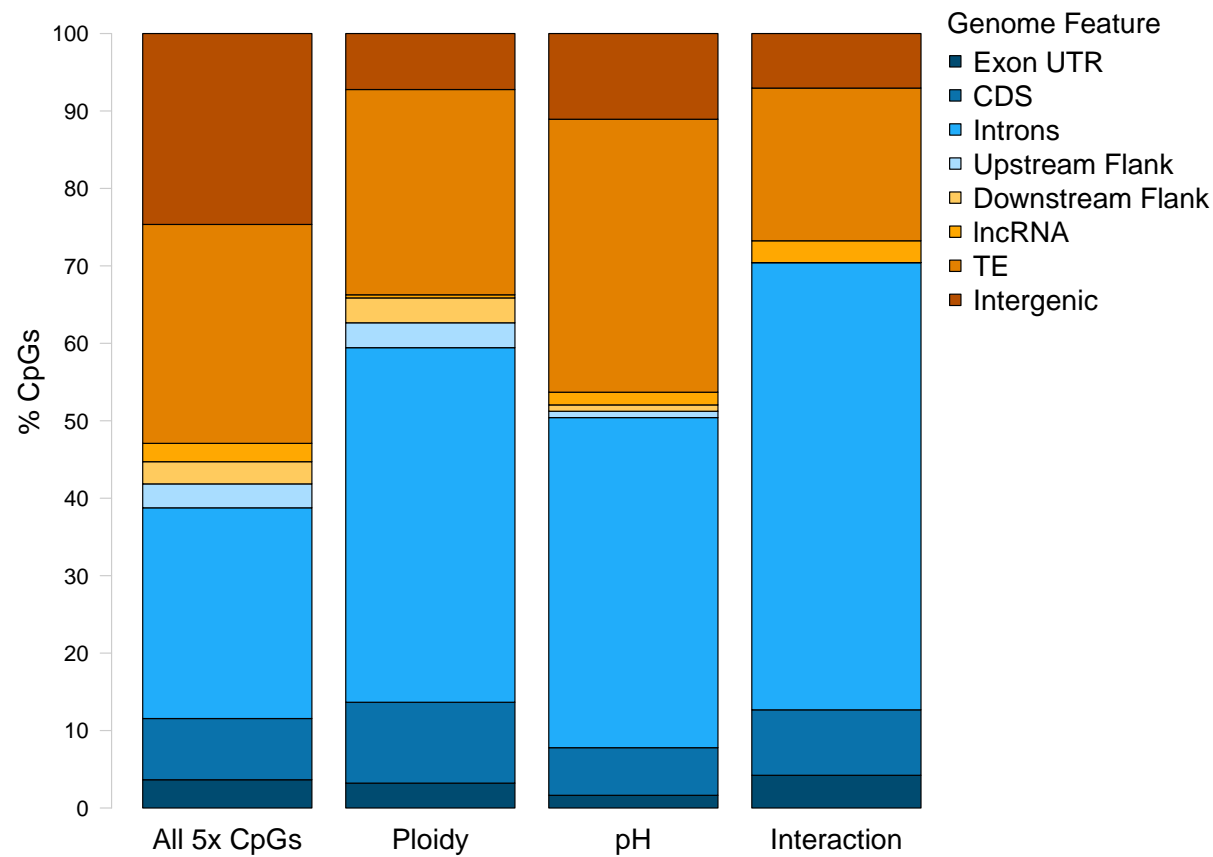
Figure 4.4: DML heatmaps and chromosomal distribution

CONCLUSION

While extensive research has been conducted on the impacts of ocean acidification on marine invertebrates, it is difficult to fully comprehend the influence of environmental stressors on organisms without molecular data. My dissertation demonstrates how proteomic and epigenetic analysis can elucidate sub- and nonlethal impacts of low pH on the Pacific oyster, Crassostrea gigas. My findings contribute to our understanding of stress response pathways in this species, which can in turn shape aquaculture practices. For example, my finding that eelgrass presence does not significantly impact stress protein abundance could signify that carbonate chemistry modulation by eelgrass may not create refugia from ocean acidification for oysters. Considering broodstock exposure history before spawning in hatcheries can impact larval yield and production.

Importantly, my dissertation helps elucidate the functional role of environmentally-responsive DNA methylation. Using C. gigas as a model organism, I found that somatic and reproductive methylomes have distinct responses to low pH stress. These responses may be linked to maintaining homeostasis under stressful conditions, which would depend on tissue function. If changes to methylation do impact gene expression, then DNA methylation would be critical for acclimation to ocean acidification conditions. With my dissertation as a foundation, future work can explore the role of DNA methylation in modulating carryover effects to environmental stressors, and determine how methylation can interact with gene expression or other epigenomic mechanisms to shape physiological phenotype.

Appendix A

APPENDIX 1

Data accessibility information and additional materials for each dissertation chapter can be found below.

A.1 Chapter 1: Characterization of Pacific oyster (Crassostrea gigas) proteomic response to natural environmental differences

Raw data can be accessed in the PeptideAtlas under accession PASS01304 and PASS01305. Skyline documents can be found on Panorama Public. All scripts and workflows can be found in the associated Github repository (https://github.com/RobertsLab/paper-gigas-DNR-proteomics).

A.2 Chapter 2: Larval response to parental low pH exposure in Pacific oysters (Crassostrea gigas)

All scripts and workflows can be found in the associated Github repository (https://github.com/RobertsLab/paper-gigas-early-gametogenic-exposure).

Additional histology can be found in this repository, including microscope images (https://github.com/RobertsLab/paper-gigas-early-gametogenic-exposure/tree/master/images/Gigas-gonad-histology) and specific sex information for each individual sampled (https://github.com/RobertsLab/paper-gigas-early-gametogenic-exposure/blob/master/data/2017-Adult-Gigas-Tissue-Sampling/2018-02-27-Gigas-Histology-Classification.csv).

A.3 Chapter 3: Low pH influences methylation patterns of gonad growth genes in the Pacific oyster (Crassostrea gigas)

All data, scripts, workflows, and outputs can be found in the associated Github repository (https://github.com/RobertsLab/project-gigas-oa-meth), including:

- A summary of sequencing and trimming information: https://github.com/RobertsLab/project-gigas-oa-meth/blob/master/output/12-functional-enrichment/all-CC-EnrichedGO-DML-withTranscript.csv

- Chi-squared test output comparing genomic locations of highly methylated CpGs against all CpGs with 5x data: https://github.com/RobertsLab/project-gigas-oa-meth/blob/master/output/09-methylation-landscape/CpG-location-statResults.txt

- Principal Components Analysis visualizing global methylation patterns: https://github.com/RobertsLab/project-gigas-oa-meth/blob/master/output/06-methylKit/figures/all-sample-PCA.pdf

- Plot of pairwise Pearson's Correlations for global methylation: https://github.com/RobertsLab/project-gigas-oa-meth/blob/master/output/06-methylKit/general-stats/Full-Sample-Pearson-Correlation-Plot-FilteredCov5Destrand.jpeg

- Chromosomal distribution of DML, including number of DML in scaffolds not placed in any chromosome: https://github.com/RobertsLab/project-gigas-oa-meth/blob/master/output/06-methylKit/DML/All-DML-by-chr.csv

- Number of DML in genes: https://github.com/RobertsLab/project-gigas-oa-meth/blob/master/output/10_DML-characterization/Number-of-DML-per-Gene.csv

- Chi-squared test output comparing genomic locations of DML against all CpGs with 5x data: https://github.com/RobertsLab/project-gigas-oa-meth/blob/master/output/10_DML-characterization/CpG-location-statResults-All.txt

- simplifyEnrichment figure for enriched cellular components: https://github.com/

```
RobertsLab/project-gigas-oa-meth/blob/master/output/12-functional-enrichment/
figures/simplifyEnrichment-CC.pdf
```

- Enriched biological process GO terms and associated gene and transcript information: `https://github.com/RobertsLab/project-gigas-oa-meth/blob/master/output/12-functional-enrichment/all-BP-EnrichedGO-DML-withTranscript.csv`

- Enriched cellular component GO terms and associated gene and transcript information: `https://github.com/RobertsLab/project-gigas-oa-meth/blob/master/output/12-functional-enrichment/all-CC-EnrichedGO-DML-withTranscript.csv`

A.4   Chapter 4: Polyploidy and environmental stress have distinct impacts on Pacific oyster (Crassostrea gigas) ctenidia methylomes

All data, scripts, workflows, and outputs can be found in the associated Github repository (`https://github.com/RobertsLab/project-oyster-oa`), including:

- Chi-squared test output comparing genomic locations of highly methylated CpGs against all CpGs with 5x data: `https://github.com/RobertsLab/project-oyster-oa/blob/master/analyses/Haws_06-methylation-landscape/CpG-location-statResults.txt`

- Plot of pairwise Pearson's Correlations for global methylation: `https://github.com/RobertsLab/project-oyster-oa/blob/master/analyses/Haws_04-methylKit/general-stats/Full-Sample-Pearson-Correlation-Plot-FilteredCov5Destrand.jpeg`

- Chromosomal distribution of ploidy-DML, including number of ploidy-DML in scaffolds not placed in any chromosome: `https://github.com/RobertsLab/project-oyster-oa/blob/master/analyses/Haws_04-DSS/DML/ploidy-DML-by-chr.csv`

- Chromosomal distribution of pH-DML, including number of pH-DML in scaffolds not placed in any chromosome: `https://github.com/RobertsLab/project-oyster-oa/`

`blob/master/analyses/Haws_04-DSS/DML/pH-DML-by-chr.csv`

- Chromosomal distribution of interaction-DML, including number of interaction-DML in scaffolds not placed in any chromosome: `https://github.com/RobertsLab/project-oyster-oa/blob/master/analyses/Haws_04-DSS/DML/ploidypH-DML-by-chr.csv`

- Number of ploidy-DML in genes: `https://github.com/RobertsLab/project-oyster-oa/blob/master/analyses/Haws_07-DML-characterization/Number-of-ploidy-DML-per-Gene.csv`

- Number of pH-DML in genes: `https://github.com/RobertsLab/project-oyster-oa/blob/master/analyses/Haws_07-DML-characterization/Number-of-pH-DML-per-Gene.csv`

- Number of interaction-DML in genes: `https://github.com/RobertsLab/project-oyster-oa/blob/master/analyses/Haws_07-DML-characterization/Number-of-ploidypH-DML-per-Gene.csv`

- Chi-squared test output comparing genomic locations of ploidy-DML against all CpGs with 5x data: `https://github.com/RobertsLab/project-oyster-oa/blob/master/analyses/Haws_07-DML-characterization/CpG-location-statResults-ploidy.txt`

- Chi-squared test output comparing genomic locations of pH-DML against all CpGs with 5x data: `https://github.com/RobertsLab/project-oyster-oa/blob/master/analyses/Haws_07-DML-characterization/CpG-location-statResults-pH.txt`

- Chi-squared test output comparing genomic locations of interaction-DML against all CpGs with 5x data: `https://github.com/RobertsLab/project-oyster-oa/blob/master/analyses/Haws_07-DML-characterization/CpG-location-statResults-ploidypH.txt`

- `simplifyEnrichment` figure for enriched biological process terms associated with ploidy-DML: `https://github.com/RobertsLab/project-oyster-oa/blob/`

`master/analyses/Haws_09-functional-enrichment/figures/simplifyEnrichment-ploidy-BP.`
`pdf`

- Enriched biological process GO terms and associated gene and transcript information for ploidy-DML: `https://github.com/RobertsLab/project-oyster-oa/blob/` `master/analyses/Haws_09-functional-enrichment/ploidy-BP-FisherTestResults.` `csv`

- `simplifyEnrichment` figure for enriched biological process terms associated with pH-DML: `https://github.com/RobertsLab/project-oyster-oa/blob/master/` `analyses/Haws_09-functional-enrichment/figures/simplifyEnrichment-pH-BP.` `pdf`

- Enriched biological process GO terms and associated gene and transcript information for pH-DML: `https://github.com/RobertsLab/project-oyster-oa/blob/` `master/analyses/Haws_09-functional-enrichment/pH-BP-FisherTestResults.` `csv`

- `simplifyEnrichment` figure for enriched molecular function terms associated with pH-DML: `https://github.com/RobertsLab/project-oyster-oa/blob/master/` `analyses/Haws_09-functional-enrichment/figures/simplifyEnrichment-pH-MF.` `pdf`

- Enriched molecular function GO terms and associated gene and transcript information for pH-DML: `https://github.com/RobertsLab/project-oyster-oa/blob/` `master/analyses/Haws_09-functional-enrichment/pH-MF-FisherTestResults.` `csv`

COLOPHON

This document is set in EB Garamond, Source Code Pro and Lato. The body text is set at 11pt with *lmr*.

It was written in R Markdown and LaTeX, and rendered into PDF using huskydown and bookdown.

This document was typeset using the XeTeX typesetting system, and the University of Washington Thesis class class created by Jim Fox. Under the hood, the University of Washington Thesis LaTeX template is used to ensure that documents conform precisely to submission standards. Other elements of the document formatting source code have been taken from the Latex, Knitr, and RMarkdown templates for UC Berkeley's graduate thesis, and Dissertate: a LaTeX dissertation template to support the production and typesetting of a PhD dissertation at Harvard, Princeton, and NYU

The source files for this thesis, along with all the data files, have been organised into an R package, xxx, which is available at `https://github.com/xxx/xxx`. A hard copy of the thesis can be found in the University of Washington library.

This version of the thesis was generated on 2021-08-20 01:09:34. The repository is currently at this commit:

The computational environment that was used to generate this version is as follows:

```
- Session info ----------------------------------------------------------------
 setting  value
 version  R version 4.1.0 (2021-05-18)
 os       macOS Big Sur 10.16
```

```
system   x86_64, darwin17.0
ui       X11
language (EN)
collate  en_US.UTF-8
ctype    en_US.UTF-8
tz       America/Los_Angeles
date     2021-08-20
```

```
- Packages -------------------------------------------------------------------
package      * version date       lib source
assertthat     0.2.1   2019-03-21 [1] CRAN (R 4.1.0)
bookdown       0.23.1  2021-08-18 [1] Github (rstudio/bookdown@6643bb9)
cachem         1.0.5   2021-05-15 [1] CRAN (R 4.1.0)
callr          3.7.0   2021-04-20 [1] CRAN (R 4.1.0)
cli            3.0.1   2021-07-17 [1] CRAN (R 4.1.0)
colorspace     2.0-2   2021-06-24 [1] CRAN (R 4.1.0)
crayon         1.4.1   2021-02-08 [1] CRAN (R 4.1.0)
DBI            1.1.1   2021-01-15 [1] CRAN (R 4.1.0)
desc           1.3.0   2021-03-05 [1] CRAN (R 4.1.0)
devtools     * 2.4.2   2021-06-07 [1] CRAN (R 4.1.0)
digest         0.6.27  2020-10-24 [1] CRAN (R 4.1.0)
dplyr        * 1.0.7   2021-06-18 [1] CRAN (R 4.1.0)
ellipsis       0.3.2   2021-04-29 [1] CRAN (R 4.1.0)
evaluate       0.14    2019-05-28 [1] CRAN (R 4.1.0)
fansi          0.5.0   2021-05-25 [1] CRAN (R 4.1.0)
fastmap        1.1.0   2021-01-25 [1] CRAN (R 4.1.0)
fs             1.5.0   2020-07-31 [1] CRAN (R 4.1.0)
generics       0.1.0   2020-10-31 [1] CRAN (R 4.1.0)
```

```
ggplot2       3.3.5   2021-06-25 [1] CRAN (R 4.1.0)
git2r         0.28.0  2021-01-10 [1] CRAN (R 4.1.0)
glue          1.4.2   2020-08-27 [1] CRAN (R 4.1.0)
gtable        0.3.0   2019-03-25 [1] CRAN (R 4.1.0)
htmltools     0.5.1.1 2021-01-22 [1] CRAN (R 4.1.0)
httr          1.4.2   2020-07-20 [1] CRAN (R 4.1.0)
huskydown   * 0.0.5   2021-08-18 [1] Github (benmarwick/huskydown@addb48e)
kableExtra    1.3.4   2021-02-20 [1] CRAN (R 4.1.0)
knitr         1.33    2021-04-24 [1] CRAN (R 4.1.0)
lifecycle     1.0.0   2021-02-15 [1] CRAN (R 4.1.0)
magrittr      2.0.1   2020-11-17 [1] CRAN (R 4.1.0)
memoise       2.0.0   2021-01-26 [1] CRAN (R 4.1.0)
munsell       0.5.0   2018-06-12 [1] CRAN (R 4.1.0)
pillar        1.6.2   2021-07-29 [1] CRAN (R 4.1.0)
pkgbuild      1.2.0   2020-12-15 [1] CRAN (R 4.1.0)
pkgconfig     2.0.3   2019-09-22 [1] CRAN (R 4.1.0)
pkgload       1.2.1   2021-04-06 [1] CRAN (R 4.1.0)
prettyunits   1.1.1   2020-01-24 [1] CRAN (R 4.1.0)
processx      3.5.2   2021-04-30 [1] CRAN (R 4.1.0)
ps            1.6.0   2021-02-28 [1] CRAN (R 4.1.0)
purrr         0.3.4   2020-04-17 [1] CRAN (R 4.1.0)
R6            2.5.0   2020-10-28 [1] CRAN (R 4.1.0)
remotes       2.4.0   2021-06-02 [1] CRAN (R 4.1.0)
rlang         0.4.11  2021-04-30 [1] CRAN (R 4.1.0)
rmarkdown     2.10    2021-08-06 [1] CRAN (R 4.1.0)
rprojroot     2.0.2   2020-11-15 [1] CRAN (R 4.1.0)
rstudioapi    0.13    2020-11-12 [1] CRAN (R 4.1.0)
rvest         1.0.0   2021-03-09 [1] CRAN (R 4.1.0)
```

```
scales       1.1.1   2020-05-11 [1] CRAN (R 4.1.0)
sessioninfo  1.1.1   2018-11-05 [1] CRAN (R 4.1.0)
stringi      1.7.3   2021-07-16 [1] CRAN (R 4.1.0)
stringr      1.4.0   2019-02-10 [1] CRAN (R 4.1.0)
svglite      2.0.0   2021-02-20 [1] CRAN (R 4.1.0)
systemfonts  1.0.2   2021-05-11 [1] CRAN (R 4.1.0)
testthat     3.0.4   2021-07-01 [1] CRAN (R 4.1.0)
tibble       3.1.3   2021-07-23 [1] CRAN (R 4.1.0)
tidyselect   1.1.1   2021-04-30 [1] CRAN (R 4.1.0)
usethis    * 2.0.1   2021-02-10 [1] CRAN (R 4.1.0)
utf8         1.2.2   2021-07-24 [1] CRAN (R 4.1.0)
vctrs        0.3.8   2021-04-29 [1] CRAN (R 4.1.0)
viridisLite  0.4.0   2021-04-13 [1] CRAN (R 4.1.0)
webshot      0.5.2   2019-11-22 [1] CRAN (R 4.1.0)
withr        2.4.2   2021-04-18 [1] CRAN (R 4.1.0)
xfun         0.25    2021-08-06 [1] CRAN (R 4.1.0)
xml2         1.3.2   2020-04-23 [1] CRAN (R 4.1.0)
yaml         2.2.1   2020-02-01 [1] CRAN (R 4.1.0)

[1] /Library/Frameworks/R.framework/Versions/4.1/Resources/library
```

REFERENCES

10 Abele, E., Philip, E., Gonzalez, P. M., & Puntarulo, S. (2007). Marine invertebrate mitochondria and oxidative stress. *Front. Biosci.*, *12*, 933–946.

Banas, N. S., Hickey, B. M., MacCready, P., & Newton, J. A. (2004). Dynamics of Willapa Bay, Washington: A Highly Unsteady, Partially Mixed Estuary. *J. Phys. Oceanogr.*, *34*(11), 2413–2427. http://doi.org/10.1175/JPO2637.1

Banas, N. S., Hickey, B. M., Newton, J. A., & Ruesink, J. L. (2007). Tidal exchange, bivalve grazing, and patterns of primary production in Willapa Bay, Washington, USA. *Mar. Ecol. Prog. Ser.*, *341*, 123–139.

Beyer, J., Green, N. W., Brooks, S., Allan, I. J., Ruus, A., Gomes, T., … Schøyen, M. (2017). Blue mussels (*Mytilus edulis spp.*) as sentinel organisms in coastal pollution monitoring: A review. *Mar. Environ. Res.*, *130*, 338–365. http://doi.org/10.1016/j.marenvres.2017.07.024

Bianucci, L., Long, W., Khangaonkar, T., Pelletier, G., Ahmed, A., Mohamedali, T., … Figueroa-Kaminsky, C. (2018). Sensitivity of the regional ocean acidification and carbonate system in Puget Sound to ocean and freshwater inputs. *Elem Sci Anth*, *6*(1), 22. http://doi.org/10.1525/elementa.151

Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes Dev.*, *16*(1), 6–21. http://doi.org/10.1101/gad.947102

Campos, A., Danielsson, G., Farinha, A. P., Kuruvilla, J., Warholm, P., & Cristobal, S. (2016). Shotgun proteomics to unravel marine mussel (*Mytilus edulis*) response

to long-term exposure to low salinity and propranolol in a Baltic Sea microcosm. *J. Proteomics*, *137*, 97–106. http://doi.org/10.1016/j.jprot.2016.01.010

Chambers, M. C., Maclean, B., Burke, R., Amodei, D., Ruderman, D. L., Neumann, S., … Mallick, P. (2012). A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol.*, *30*(10), 918–920. http://doi.org/10.1038/nbt.2377

Chapman, R. W., Mancia, A., Beal, M., Veloso, A., Rathburn, C., Blair, A., … Sanger, D. (2011). The transcriptomic responses of the eastern oyster, *Crassostrea virginica*, to environmental conditions. *Mol. Ecol.*, *20*(7), 1431–1449. http://doi.org/10.1111/j.1365-294X.2011.05018.x

Cornwall, C. E., & Hurd, C. L. (2016). Experimental design in ocean acidification research: problems and solutions. *ICES J. Mar. Sci.*, *73*(3), 572–581. http://doi.org/10.1093/icesjms/fsv118

Deans, C., & Maggert, K. A. (2015). What Do You Mean, "Epigenetic?" *Genetics*, *199*(4), 887–896. http://doi.org/10.1534/genetics.114.173492

Dineshram, R., Chandramouli, K., Ko, G. W. K., Zhang, H., Qian, P.-Y., Ravasi, T., & Thiyagarajan, V. (2016). Quantitative analysis of oyster larval proteome provides new insights into the effects of multiple climate change stressors. *Glob. Chang. Biol.*, *22*(6), 2054–2068. http://doi.org/10.1111/gcb.13249

Egertson, J. D., Kuehn, A., Merrihew, G. E., Bateman, N. W., MacLean, B. X., Ting, Y. S., … MacCoss, M. J. (2013). Multiplexed MS/MS for improved data-independent acquisition. *Nat. Methods*, *10*(8), 744–746. http://doi.org/10.1038/nmeth.2528

Eirin-Lopez, J. M., & Putnam, H. M. (2018). Marine Environmental Epigenetics. *Ann. Rev. Mar. Sci.* http://doi.org/10.1146/annurev-marine-010318-095114

Feely, R. A., Alin, S. R., Newton, J., Sabine, C. L., Warner, M., Devol, A., … Maloy, C. (2010). The combined effects of ocean acidification, mixing, and respiration on pH

and carbonate saturation in an urbanized estuary. *Estuar. Coast. Shelf Sci.*, *88*(4), 442–449. http://doi.org/10.1016/j.ecss.2010.05.004

Flores-Nunes, F., Gomes, T., Company, R., Moraes, R. R. M., Sasaki, S. T., Taniguchi, S., … Bebianno, M. J. (2015). Changes in protein expression of Pacific oyster *Crassostrea gigas* exposed in situ to urban sewage. *Environ. Sci. Pollut. Res. Int.*, *22*(22), 17267–17279. http://doi.org/10.1007/s11356-014-3821-8

Gavery, M. R., & Roberts, S. B. (2013). Predominant intragenic methylation is associated with gene expression characteristics in a bivalve mollusc. *PeerJ*, *1*, e215. http://doi.org/10.7717/peerj.215

Gavery, M. R., & Roberts, S. B. (2017). Epigenetic considerations in aquaculture. *PeerJ*, *5*, e4147. http://doi.org/10.7717/peerj.4147

Gazeau, F., Gattuso, J.-P., Greaves, M., Elderfield, H., Peene, J., Heip, C. H. R., & Middelburg, J. J. (2011). Effect of carbonate chemistry alteration on the early embryonic development of the Pacific oyster (*Crassostrea gigas*). *PLoS One*, *6*(8), e23010. http://doi.org/10.1371/journal.pone.0023010

Gazeau, F., Quiblier, C., Jansen, J. M., Gattuso, J.-P., Middelburg, J. J., & Heip, C. H. R. (2007). Impact of elevated CO_2 on shellfish calcification. *Geophys. Res. Lett.*, *34*(7), L07603. http://doi.org/10.1029/2006GL028554

Guzy, R. D., & Schumaker, P. T. (2006). Oxygen sensing by mitochondria at complex III: 695 the paradox of increased reactive oxygen species during hypoxia. Experimental 696. *Physiology*, *91*(807-819), 697.

Hamdoun, A. M., Cheney, D. P., & Cherr, G. N. (2003). Phenotypic plasticity of HSP70 and HSP70 gene expression in the Pacific oyster (*Crassostrea gigas*): implications for thermal limits and induction of thermal tolerance. *Biol. Bull.*, *205*(2), 160–169. http://doi.org/10.2307/1543236

Hercus, M. J., Loeschcke, V., & Rattan, S. I. S. (2003). Lifespan extension of *Drosophila melanogaster* through hormesis by repeated mild heat stress. *Biogerontology*, *4*(3), 149–156.

Hoaglin, D. C., Iglewicz, B., & Tukey, J. W. (1986). Performance of Some Resistant Rules for Outlier Labeling. *J. Am. Stat. Assoc.*, *81*(396), 991–999. http://doi.org/10.1080/01621459.1986.10478363

Kurihara, H., Kato, S., & Ishimatsu, A. (2007). Effects of increased seawater pCO$_2$ on early development of the oyster *Crassostrea gigas*. *Aquat. Biol.*, *1*, 91–98. http://doi.org/10.3354/ab00009

Lamb, J. B., Water, J. A. J. M. van de, Bourne, D. G., Altier, C., Hein, M. Y., Fiorenza, E. A., … Harvell, C. D. (2017). Seagrass ecosystems reduce exposure to bacterial pathogens of humans, fishes, and invertebrates. *Science*, *355*(6326), 731–733. http://doi.org/10.1126/science.aal1956

Liu, Z., Pan, S., Sun, Z., Ma, R., Chen, L., Wang, Y., & Wang, S. (2015). Heavy metal spatial variability and historical changes in the Yangtze River estuary and North Jiangsu tidal flat. *Mar. Pollut. Bull.*, *98*(1-2), 115–129. http://doi.org/10.1016/j.marpolbul.2015.07.006

Livingstone, D. R. (n.d.). Induction of enzymes as a mechanism for the seasonal control of metabolism in marine invertebrates: Glucose-6-phosphate dehydrogenases from the mantle and hepatopancreas of the common mussel *mytilus edulis* l. *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry*, *69*(2), 147–156.

Lowe, A. T., Kobelt, J., Horwith, M., & Ruesink, J. (2018). Ability of Eelgrass to Alter Oyster Growth and Physiology Is Spatially Limited and Offset by Increasing Predation Risk. *Estuaries Coasts*. http://doi.org/10.1007/s12237-018-00488-9

MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L., Frewen, B., … MacCoss, M. J. (2010). Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics*, *26*(7), 966–968. http://doi.org/10.1093/bioinformatics/btq054

Meng, J., Wang, W., Li, L., Yin, Q., & Zhang, G. (2017). Cadmium effects on DNA and protein metabolism in oyster (*Crassostrea gigas*) revealed by proteomic analyses. *Sci. Rep.*, *7*(1), 11716. http://doi.org/10.1038/s41598-017-11894-7

Michaelidis, B., Haas, D., & Grieshaber, M. K. (2005). Extracellular and Intracellular Acid-Base Status with Regard to the Energy Metabolism in the Oyster *Crassostrea gigas* during Exposure to Air. *Physiol. Biochem. Zool.*, *78*(3), 373–383. http://doi.org/10.1086/430223

Omoregie, E., Mwatilifange, N. S. I., & Liswaniso, G. (2019). Futuristic Ocean Acidification Levels Reduce Growth and Reproductive Viability in the Pacific Oyster (*Crassostrea gigas*). *J. Appl. Sci. Environ. Manage.*, *23*(9), 1747–1754.

Pacella, S. R., Brown, C. A., Waldbusser, G. G., Labiosa, R. G., & Hales, B. (2018). Seagrass habitat metabolism increases short-term extremes and long-term offset of CO_2 under future ocean acidification. *Proc. Natl. Acad. Sci. U. S. A.* http://doi.org/10.1073/pnas.1703445115

Paerl, H. W., Hall, N. S., Peierls, B. L., & Rossignol, K. L. (2014). Evolving Paradigms and Challenges in Estuarine and Coastal Eutrophication Dynamics in a Culturally and Climatically Stressed World. *Estuaries Coasts*, *37*(2), 243–258. http://doi.org/10.1007/s12237-014-9773-x

Pelletier, G., Roberts, M., Keyzers, M., & Alin, S. R. (2018, January). Seasonal variation in aragonite saturation in surface waters of Puget Sound – a pilot study. http://doi.org/10.1525/elementa.270

Pennock, J. R., & Sharp, J. H. (1986). Phytoplankton production in the Delaware Estuary: temporal and spatial variability. *Mar. Ecol. Prog. Ser.*, *34*(1/2), 143–155.

Riviere, G., Klopp, C., Ibouniyamine, N., Huvet, A., Boudry, P., & Favrel, P. (2015). GigaTON: an extensive publicly searchable database providing a new reference transcriptome in the Pacific oyster *Crassostrea gigas*. *BMC Bioinformatics*, *16*, 401. http://doi.org/10.1186/s12859-015-0833-4

Roberts, S. B., & Gavery, M. R. (2012). Is There a Relationship between DNA Methylation and Phenotypic Plasticity in Invertebrates? *Front. Physiol.*, *2*. http://doi.org/10.3389/fphys.2011.00116

Ruesink, J. L., Yang, S., & Trimble, A. C. (2015). Variability in Carbon Availability and Eelgrass (*Zostera marina*) Biometrics Along an Estuarine Gradient in Willapa Bay, WA, USA. *Estuaries Coasts*, *38*(6), 1908–1917. http://doi.org/10.1007/s12237-014-9933-z

Silvestre, F., Dierick, J.-F., Dumont, V., Dieu, M., Raes, M., & Devos, P. (2006). Differential protein expression profiles in anterior gills of *Eriocheir sinensis* during acclimation to cadmium. *Aquat. Toxicol.*, *76*(1), 46–58. http://doi.org/10.1016/j.aquatox.2005.09.006

Slattery, M., Ankisetty, S., Corrales, J., Marsh-Hunkin, K. E., Gochfeld, D. J., Willett, K. L., & Rimoldi, J. M. (2012). Marine proteomics: a critical assessment of an emerging technology. *J. Nat. Prod.*, *75*(10), 1833–1877. http://doi.org/10.1021/np300366a

Sussarellu, R., Fabioux, C., Camacho Sanchez, M., Le Göic, N., Lambert, C., Soudant, P., & Moraga, D. (2012). Molecular and cellular response to short-term oxygen variations in the Pacific oyster *Crassostrea gigas*. *J. Exp. Mar. Bio. Ecol.*, *412*, 87–95. http://doi.org/10.1016/j.jembe.2011.11.007

Timmins-Schiffman, E., Coffey, W. D., Hua, W., Nunn, B. L., Dickinson, G. H., & Roberts, S. B. (2014). Shotgun proteomics reveals physiological response to ocean acidification in *Crassostrea gigas*. *BMC Genomics*, *15*, 951. http://doi.org/10.1186/1471-2164-15-951

Timmins-Schiffman, E., O'Donnell, M. J., Friedman, C. S., & Roberts, S. B. (2013). Elevated pCO$_2$ causes developmental delay in early larval Pacific oysters, *Crassostrea gigas*. *Mar. Biol.*, *160*(8), 1973–1982. http://doi.org/10.1007/s00227-012-2055-x

Ting, Y. S., Egertson, J. D., Payne, S. H., Kim, S., MacLean, B., Käll, L., … MacCoss, M. J. (2015). Peptide-Centric Proteome Analysis: An Alternative Strategy for the Analysis of Tandem Mass Spectrometry Data. *Mol. Cell. Proteomics*, *14*(9), 2301–2307. http://doi.org/10.1074/mcp.O114.047035

Tomanek, L. (2014). Proteomics to study adaptations in marine organisms to environmental stress. *J. Proteomics*, *105*, 92–106. http://doi.org/10.1016/j.jprot.2014.04.009

Vargas-Albores, F., Martínez-Martínez, A., Aguilar-Campos, J., & Jiménez-Vega, F. (2009). The expression of protein disulfide isomerase from *Litopenaeus vannamei* hemocytes is regulated by bacterial inoculation. *Comp. Biochem. Physiol. Part D Genomics Proteomics*, *4*(3), 141–146. http://doi.org/10.1016/j.cbd.2009.01.001

Veldhoen, N., Ikonomou, M. G., & Helbing, C. C. (2012). Molecular profiling of marine fauna: integration of omics with environmental assessment of the world's oceans. *Ecotoxicol. Environ. Saf.*, *76*(2), 23–38. http://doi.org/10.1016/j.ecoenv.2011.10.005

Waldbusser, G. G., Hales, B., Langdon, C. J., Haley, B. A., Schrader, P., Brunner, E. L., … Gimenez, I. (2014). Saturation-state sensitivity of marine bivalve larvae to ocean acidification. *Nat. Clim. Chang.*, *5*, 273. http://doi.org/10.1038/nclimate2479

Wei, L., Wang, Q., Ning, X., Mu, C., Wang, C., Cao, R., … Zhao, J. (2015). Combined metabolome and proteome analysis of the mantle tissue from Pacific oyster Cras-

sostrea gigas exposed to elevated pCO$_2$. *Comp. Biochem. Physiol. Part D Genomics Proteomics*, *13*, 16–23. http://doi.org/10.1016/j.cbd.2014.12.001

Zhang, P., Li, C., Li, Y., Zhang, P., Shao, Y., Jin, C., & Li, T. (2014). Proteomic identification of differentially expressed proteins in sea cucumber *Apostichopus japonicus* coelomocytes after *Vibrio splendidus* infection. *Dev. Comp. Immunol.*, *44*(2), 370–377. http://doi.org/10.1016/j.dci.2014.01.013

Zhang, Y., Sun, J., Mu, H., Li, J., Zhang, Y., Xu, F., … Yu, Z. (2015). Proteomic basis of stress responses in the gills of the pacific oyster *Crassostrea gigas*. *J. Proteome Res.*, *14*(1), 304–317. http://doi.org/10.1021/pr500940s