Epigenomic and Transcriptomic Regulation of Environmental Responses in the Pacific oyster, *Crassostrea gigas*

Mackenzie R. Gavery

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

Doctor of Philosophy

University of Washington

2014

Reading Committee:

Steven Roberts, Chair

Graham Young

Carolyn Friedman

Program Authorized to Offer Degree:

Aquatic and Fishery Sciences

University of Washington

Abstract

Epigenomic and Transcriptomic Regulation of Environmental Responses in the Pacific oyster, *Crassostrea gigas*

Mackenzie R. Gavery

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

Intertidal invertebrates such as bivalve molluscs live in constantly changing and frequently stressful environments and must be equipped to both detect and quickly respond to environmental changes. Our understanding of the molecular and cellular response systems in these organisms are an important part of species conservation and well as our ability to predict their response and limits to environmental stress. This dissertation explores environmental responses in oysters using transcriptomic and epigenomic approaches. The first chapter examines the transcriptomic responses of oysters from two locations with varying anthropogenic input, using ultra short high-throughput sequencing reads. The work presented in Chapter 2 provides the first evidence that DNA methylation is present in the genome of a bivalve mollusc and

suggests a regulatory role in these species. Chapter 3 provides the first whole methylome analysis of a locotrophozoan and identifies relationships between DNA methylation and gene expression. Finally, Chapter 4 presents a review of the current DNA methylation data available for bivalves and proposes new hypotheses for how DNA methylation may be regulating the genome in oysters. By combining transcriptional and epigenetic datasets, this work provides the most complete picture of epigenomic regulation for any molluscan species and paves the way into future investigations of the role of epigenetics in environmental regulation and local adaptation and evolution in marine invertebrates.

Table of Contents

List of Figures

List of Tables

Acknowledgments and dedication

First and foremost, I would like to acknowledge and thank my advisor, Dr. Steven Roberts, for his mentorship. I will always appreciate Dr. Roberts' enthusiasm and encouragement of my research, and for both challenging and supporting me in equal amounts. I would also like to thank the other members of my supervisory committee: Dr. Graham Young, Dr. Joth Davis, Dr. Charles Laird and Dr. Carolyn Friedman, for their expertise, guidance and support.

This work would not have been completed without the support of the U.S. Environmental Protection Agency STAR Fellowship (Assistance Agreement no. FP917331) and the National Science Foundation (Grant Number 1158119). I am also grateful to have been a recipient of the William H. Pierre Sr. and Victor and Tamara Loosanoff Fellowships through the School of Aquatic and Fishery Sciences.

I would like to express my sincere thanks and gratitude to my colleagues and friends in the Roberts Lab: Sam White, Emma Timmins-Schiffman, Caroline Storer, Lisa Crosson, Claire Olson, Brent Vadopalas, Dave Metzger, and Jake Heare. They provided invaluable input and critiques of my experiments, manuscripts and presentations. They are also the best lab mates anyone could ever ask for.

I am constantly humbled by and grateful for the boundless support of my family and friends. I want to give special thanks to those who I convinced to join me for field work (in less than ideal conditions), for their enthusiasm and dedication. I am especially thankful for my husband, Brock Gavery, an unwavering supporter of my scientific aspirations, who provided much needed encouragement and humor through it all.

This work is dedicated with love and gratitude to my parents, who have always been my inspiration.

Introduction

Background

Intertidal invertebrates live in constantly changing and often stressful environments. From daily and seasonal changes in tides and temperature salinity, to disease presence and anthropogenic stressors, intertidal invertebrates must be equipped to both detect and quickly respond to environmental changes. Many organisms respond to environmental change by movement toward a more favorable environment, sessile organisms, such as many marine bivalves, must manage environmental change and stress in other ways. Understanding how these organisms cope with their ever-changing environments is an important step toward improving our ability to predict how populations may respond to stress, which would be particularly important for those species with key ecosystem roles and large economic importance. This understanding is also moves toward a greater understanding of broad molecular and evolutionary questions such as, how has the genome been shaped by living in these variable environments, and what genomic tools have individual species developed in order to persist in these harsh conditions?

There is a tremendous amount to be gained by extending our knowledge of genome regulation and gene expression in molluscan taxa, especially bivalves. Research surrounding bivalve physiology and stress responses has been on the rise in diverse disciplines such as aquatic and fishery science, toxicology, and climatology for a large part because bivalves, such as oysters, are ecologically, commercially and culturally important. As benthic filter-feeders, oysters play an important role in estuarine food webs and contribute to the removal of excess organics, nutrients, and particulates (Newell et al., 1999). Pacific oysters are also a key cultured shellfish species with Washington state

producing 87% of the US supply. Additionally, oysters are commonly used as estuarine and marine indicator species as they are easily obtained, likely to survive test conditions and accumulate the contaminants of interest to concentrations proportional to ambient waters (Phillips & Rainbow, 1993). There is also increasing interest in having a more integrated understanding of the impacts of climate change and ocean acidification on specific organisms in order to mitigate risks and anticipate potential impacts. To that end, this dissertation focuses on characterizing gene expression patterns and DNA methylation in the Pacific oyster.

The Studied Species, Crassostrea gigas

Crassostrea gigas (Thunberg, 1795) is a marine bivalve of the phylum Mollusca, family Ostreidae. The native range of *C. gigas* includes Japan and coastal regions of Asia, but this species has also been widely introduced in North America, Europe and Australia for cultivation purposes and is one of the most important food oysters in the world. *C. gigas* can be found on both hard and soft substrates in temperate tidal and sub-tidal zones in optimal salinities of 25 to 35 ppt (Quayle, 1969). As protandrous sequential hermaphrodites (Katkanksy & Sparks, 1966), *C. gigas* typically mature first as a male and then as a male or a female in subsequent reproductive seasons with only rare cases of simultaneous hermaphroditism. Spawning occurs annually in warm months and involves the release of gametes into the water column where fertilization occurs. *C. gigas* is a highly fecund species with a market sized female producing 50 – 100 million eggs (Quayle,1969). Larvae are planktonic for 20 – 30 days before settling and attaching to substrate (Shatkin et al., 1997).

Because their ecological importance as ecosystem engineers and commercial importance as a major aquaculture species throughout the world, the Pacific oyster has

emerged as a model species for marine bivalves. Consequently, substantial genomic resources have been produced in this species, including most recently the sequencing of the complete genome. The Pacific oyster has many unique characteristics that were argued as justification for the sequencing the genome including high genomic variability and genetic load, observations of heterosis and inbreeding depression, as well as the desire to understand complex traits such as growth and disease resistance (Hedgecock et al., 2005). The draft genome, released in 2012, revealed that oyster genome has an unusually high number of genes involved in responding to changes in the environment. Additionally, the oyster genome has a high rate of polymorphism and it has been suggested that transposable element activity may be involved in shaping this variation (Zhang et al., 2012).

Transcriptomic Tools to Understand Environmental Response

Understanding shellfish genome regulation and mechanisms of stress response in shellfish will facilitate discovery in diverse disciplines such as aquaculture, restoration and comparative biology. One approach that has allowed for discovery of many shellfish genes involved in various stress responses is through transcriptomics, the analysis of messenger ribonucleic acid (mRNA) expression. The transcriptome is dynamic and is influenced by many factors. Therefore, transcript expression profiles represent phenotypes shaped by the genotype of an organism and its environment. Many studies have used single gene approaches to characterize how bivalves respond to various stressors such as temperature (Tirapé et al., 2007; Farcy et al., 2008; Ivanina et al., 2008), heavy metals and hydrocarbon exposure (Jenny et al., 2006; Boutet et al., 2006; Manduzio et al., 2004), and pathogen exposure (Roberts et al., 2008, Tirapé et al., 2007).

Recent development of high-throughput DNA sequencing technologies, including whole transcriptome sequencing, has provided new opportunities to examine gene regulation and environmental responses in non-model organisms. Researchers have begun to examine the applicability of using high-throughput short-read technology for generating *de novo* transcriptomes in non-model species. In addition to assembling transcriptomes, this type of sequencing can be used directly to examine gene expression levels, a method referred to as RNA-Seq. In RNA-Seq, high-throughput sequencing reads generated from cDNA libraries are aligned to a common reference sequence or scaffold (e.g. whole genome) to produce a transcriptome map that includes transcript abundance for each gene. This approach has a number of benefits over previous methods including its utility in gene discovery, it is not reliant on a priori knowledge of genes involved in a particular biological process, and the ability to detect rare transcripts and alternative RNA splicing forms.

The primary goal of Chapter 1 was to evaluate the effectiveness of utilizing high-throughput, ultra short-read sequencing to both characterize the transcriptome and to analyze gene expression patterns in the Pacific oyster (*Crassostrea gigas*). As part of this study, gene expression patterns between oyster populations exposed to varying degrees of anthropogenic impact were compared. RNA-Seq was performed using only the *de novo* assembled transcripts as a scaffold. In order to evaluate the effectiveness of this approach, RNA-Seq was also performed using publicly available transcriptome data as a scaffold and the results were compared. This work not only evaluated the use of limited ultra-short read sequence data for characterizing transcriptomes in non-model organisms, but also

provides insight into the physiological responses of aquatic invertebrates in natural environments.

The Role of Epigenetics in Environmental Response

Epigenetics refers to heritable processes that alter gene activity without manipulating the underlying DNA sequence (Jablonka & Lamb, 2002). Common epigenetic mechanisms include DNA methylation, histone modifications and non-coding RNA activity. The best studied of these is DNA methylation, which refers to the enzymatic addition of a methyl group to a cytosine residue in DNA. DNA methylation is an important mechanism of gene regulation in both plants and animals. In vertebrates, DNA methylation plays important roles in providing genomic stability (Maloisel & Rossignol, 1998), genomic imprinting (Bell & Felsenfeld, 2000), and X-chromosome inactivation (Csankovszki et al., 2001).

Epigenetic mechanisms have received increased attention over the last 30 years in large part because of the association between aberrant epigenetic regulation and cancer (Feinberg & Tycko, 2004). Recently, research has become much broader in scope and it has been shown that epigenetic modifications, and the enzymes responsible for maintaining them, are susceptible to disruption by a range of environmental factors including diet (Wilson et al., 1984; Dolinoy et al., 2007), xenobiotic chemicals (Sutherland & Costa, 2003), and endocrine disrupting compounds (Anway et al., 2005) which can contribute to disease states. Susceptibility to environmental control and association with disease expression makes this an area rich with new data and exciting research questions.

The extent of cytosine methylation varies considerably among eukaryotes. In vertebrates, approximately 70-80% of CpG dinucleotides are methylated (Bird & Taggart,

1980) a pattern referred to as global methylation. Invertebrates display a wide range of DNA methylation, from very limited methylation in *Drosophilia melanogaster* (Gowher et al., 2000) and *Caenorhabditis elegans* (Simpson et al., 1986) to a mosaic-pattern of methylation in the sea urchin (*Strongylocentrotus purpuratus*) (Bird et al., 1979) and *Ciona intestinalis* (Simmen & Bird, 2000; Suzuki et al., 2007). In vertebrates, regulation of transcription by DNA methylation is accomplished by differential patterns of methylation in intergenic regions, namely gene promoters (Boyes & Bird, 1992; Kass et al., 1997; Hsieh, 1994). In contrast, there are no significant differences being reported in the methylation status of gene promoters in invertebrates where methylation appears to be targeted specifically to transcription units (Suzuki et al., 2007; Elango & Yi, 2008). Computational analyses of the methylation status of honey bee (*Apis mellifera*) provided some of the first evidence supporting a regulatory role of intragenic DNA methylation in invertebrates (Elango et al., 2009; Foret et al., 2009). In these studies, genes associated with general metabolic or 'housekeeping' functions were predicted to be hyper-methylated, whereas caste-specific genes were preferentially hypo-methylated. This functional clustering suggests DNA methylation serves to regulate gene transcription in *A. mellifera* but it is uncertain if this function is conserved across invertebrate taxa. Furthermore, it is unclear exactly how intragenic cytosine methylation directly affects transcription.

Studies in *A. mellifera* and others illustrate the diversity of DNA methylation patterns in invertebrate taxa and highlight gaps in our understanding of the evolutionary and functional significance of DNA methylation. One taxonomic group that has been notably absent from these investigations is the phylum Mollusca. Molluscs were first categorized as having 'echinoderm-type' DNA methylation patterns based on

experimental evidence using the common mussel (*Mytilus edulis*) (Bird & Taggart, 1980). Since then, there has been little investigation of DNA methylation in molluscs, with the exception of evidence suggesting the presence of CpG methylation in the clam, *Donux truculus* (Petrovic et al., 2009). In addition to increasing our understanding of the evolution of DNA methylation in invertebrate taxa, this study provides an opportunity to evaluate *C. gigas* as a model organism for analyzing DNA methylation in an aquatic species. Bivalve molluscs are important bioindicators (Markert et al., 2003) and elucidating the functional significance of DNA methylation in these organisms may prove valuable information for understanding the effects of environmental stress in aquatic organisms.

This dissertation explores environmental responses in oysters using transcriptomic and epigenomic approaches. The first chapter examines the transcriptomic responses of oysters from two locations with varying anthropogenic input, using ultra short high-throughput sequencing reads. The work presented in Chapter 2 provides the first evidence that DNA methylation is present in the genome of a bivalve mollusc and suggests a regulatory role in these species. Chapter 3 provides the first whole methylome analysis of a locotrophozoan and identifies relationships between DNA methylation and gene expression. Finally, Chapter 4 presents a review of the current DNA methylation data available for bivalves and proposes new hypotheses for how DNA methylation may be regulating the genome in oysters. The appendices of this dissertation cover additional research projects that broadly examined gene expression or genome regulation in oysters. Appendix A describes the characterization of the first full-length cyclooxygenase gene in oysters. Appendix B is a technical report generated as part of the development and

qualification of a novel methylation assay. Finally, Appendix C contains an archive list of the samples generated from an experiment aiming to disrupt normal methylation patterns in broodstock in order to examine the transferability of disrupted DNA methylation patterns to offspring. This work as a whole consists of novel investigations into the regulation of the genome in molluscan species. By combining transcriptional and epigenetic datasets, these analyses provides the most complete picture of epigenomic regulation for any molluscan species and pave the way for future investigations of the role of epigenetics in environmental response, local adaptation and evolution in marine invertebrates.

Works Cited

Anway MD, Skinner MK. 2006. Epigenetic transgenerational actions of endocrine disruptors. *Endocrinology* 147:43-49.

Bell AC, Felsenfeld G. 2000. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* 405:482-485.

Bird AP, Taggart MH, Smith BA. 1979. Methylated and unmethylated DNA compartments in the sea urchin genome. *Cell* 17:889–901.

Bird AP, Taggart MH. 1980. Variable patterns of total DNA and rDNA methylation in animals. *Nucleic Acid Res* 8:1485-1497.

Boutet I, Meistertzheim A, Tanguy A, Thebault M, Moraga D. 2005. Molecular characterization and expression of the gene encoding aspartate aminotransferase from the Pacific oyster *Crassostrea gigas* exposed to environmental stressors. *Comp Biochem Physiol C Toxicol Pharmacol* 140:69-78.

Boyes J, Bird A. 1992. Repression of genes by DNA methylation depends on CpG density and promoter strength. Evidence for involvement of a methyl-CpG binding protein. *EMBO J* 11:327–333.

Csankovszki G, Nagy A, Jaenisch R. 2001. Synergism of Xist RNA, DNA methylation, and histone hypoacetylation in maintaining X chromosome inactivation. *J Cell Biol* 153:773-784.

Dolinoy DC, Weidman JR, Waterland RA, Jirtle RL. 2006. Maternal genistein alters coat color and protects avy mouse offspring from obesity by modifying the fetal epigenome. *Environ Health Perspect* 114:567-572.

Elango N, Hunt BG, Goodisman MAD, Yi S. 2009. DNA methylation is widespread and associated with differential gene expression in castes of the honeybee, *Apis mellifera*. *Proc Natl Acad Sci USA* 106:11206-11211.

Elango N, Yi SV. 2008. DNA methylation and structural and functional bimodality of vertebrate promoters. *Mol Biol and Evol* 25:1602-1608.

Foret S, Kucharski R, Pittelkow Y, Lockett GA, Maleszka R. 2009. Epigenetic regulation of the honey bee transcriptome. unravelling the nature of methylated genes. *BMC Genomics* 10:472.

Gowher H, Leismann O, Jeltsch A. 2000. DNA of Drosophila melanogaster contains 5-methylcytosine. *EMBO J* 19:6918–6923.

Hedgecock D, Gaffney PM, Goulletquer P, Guo X, Reece K, Warr GW. 2005. The case for sequencing the Pacific oyster genome. *Journal of Shellfish Research* 24:429-441.

Hsieh CL. 1994. Dependence of transcriptional repression on CpG methylation density. *Mol Cell Biol* 14:5487–5494.

Jablonka E, Lamb M. 2002. The changing concept of epigenetics. *Ann N Y Acad Sci* 981:82-96.

Kass SU, Pruss D, Wolffe AP. 1997. How does DNA methylation repress transcription? *Trends Genet* 13:444–449.

Katkansky SC and Sparks AK. 1966. Seasonal sexual pattern in the Pacific oyster *Crassostrea gigas* in Washington State. *WA Dep Fish Fish Res Pap* 2: 80-89.

Maloisel L, Rossignol JL. 1998. Suppression of crossing-over by DNA methylation in Ascobolus. *Genes Dev* 12:1381-1389.

Manduzio H, Monsinjon T, Galap C, Leboulenger F, Rocher B. 2004. Seasonal variations in antioxidant defences in blue mussels *Mytilus edulis* collected from a polluted area. major contributions in gills of an inducible isoform of Cu/Zn-superoxide dismutase and of glutathione S-transferase. *Aquatic Toxicology* 70:83-93.

Markert BA, Breure AM, Zechmeister HG. 2003. Molluscs as bioindicators. In Bioindicators and Biomonitors. Elsevier Science 577–634. Amsterdam, Netherlands.

Newell RIE. 2004. Ecosystem influences of natural and cultivated populations of suspension-feeding bivalve molluscs: a review. *Journal of Shellfish Research* 23:51-61.

Petrović V, Pérez-García C, Pasantes JJ, Šatović E, Prats E, Plohl M. 2009. A GC-rich satellite DNA and karyology of the bivalve mollusk *Donax trunculus*: a dominance of GC-rich heterochromatin. *Cytogenet Genome Res* 124:63-71.

Phillips DJH, Rainbow PS. 1993. Biomonitoring of trace aquatic contaminants. Elsevier Applied Science. New York, NY.

Quayle DB. 1969. Pacific oyster culture in British Columbia. *Fish Res Board Can Bull* 169:192pp.

Roberts S, Goetz G, White S, Goetz F. 2008. Analysis of genes isolated from plated hemocytes of the Pacific oyster Crassostrea gigas. *Marine Biotechnol*11:24-44.

Simmen MW, Bird, AP. 2000. Sequence analysis of transposable elements in the sea squirt, Ciona intestinalis. *Mol Biol Evol* 17:1685–1693.

Simpson VJ, Johnson TE, Hammen RF. 1986. Caenorhabditis elegans DNA does not contain 5-methylcytosine at any time during development or aging. *Nucleic Acids Res* 14:6711–6719.

Sutherland JE, Costa M. 2003. Epigenetics and the environment. *Ann NY Acad Sci* 983:151–160.

Suzuki MM, Kerr ARW, De Sousa D, Bird A. 2007. CpG methylation is targeted to transcription units in an invertebrate genome. *Genome Res* 17:625–631.

Tirapé A, Bacque C, Brizard R, Vandenbulcke F, Boulo V. 2007. Expression of immune-related genes in the oyster *Crassostrea gigas* during ontogenesis. *Dev Compar Immunol* 31:859-873.

Wilson MJ, Shivapurkar N, Poirier LA. 1984. Hypomethylation of hepatic nuclear DNA in rats fed with a carcinogenic methyl-deficient diet. *Biochem J* 218:987–990.

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

Chapter I: Characterizing short read sequencing for gene discovery and RNA-Seq analysis in *Crassostrea gigas*

A version of this chapter was published in Comparative Biochemistry and Physiology Part D: Genomics and Proteomics: Gavery M and Roberts SB. 2012. Characterizing short read sequencing for gene discovery and RNA-Seq analysis in *Crassostrea gigas*. Comp Biochem and Phys Part D. 7:94-99.

Abstract

Advances in DNA sequencing technology have provided opportunities to produce new transcriptomic resources for species that lack completely sequenced genomes. However, there are limited examples that rely solely on ultra-short read sequencing technologies (e.g. Solexa, SOLiD) for transcript discovery and gene expression analysis (i.e. RNA-Seq). Here we use SOLiD sequencing to examine gene expression patterns in Pacific oyster (*Crassostrea gigas*) populations exposed to varying degrees of anthropogenic impact. Novel transcripts were identified and RNA- Seq analysis revealed several hundred differentially expressed genes. Gene enrichment analysis determined that in addition to biological processes predicted to be associated with anthropogenic influences (e.g. immune response), other processes play important roles including cell recognition and cell adhesion. To evaluate the effectiveness of restricting characterization solely to short-read sequences, mapping and RNA-Seq analysis were also performed using publicly available transcriptome sequence data as a scaffold. This study demonstrates that ultra-short read sequencing technologies can effectively generate novel transcriptome information, identify differentially expressed genes, and will be important for examining environmental physiology of non-model organisms.

Introduction

High-throughput DNA sequencing technologies are providing new opportunities to generate genomic resources for non-model organisms. A widely used approach is transcriptome sequencing, which has the benefit of providing increased coverage as a result of the reduced representation of the genome. A primary platform being used to generate transcriptomic resources in non-model species is the Roche 454 GS-FLX (454) followed by *de novo* assembly of sequence reads. This approach has been used to characterize transcriptomes of diverse taxa including plants (e.g. Novaes et al., 2008), insects (e.g. Vera et al., 2008), corals (e.g. Meyer et al., 2009), molluscs (e.g. Craft et al., 2010) and fish (e.g. Fraser et al., 2011). One benefit of using the 454 platform is that reads are longer compared to other common high-throughput sequencing systems, such as the Illumina Genome Analyzer IIx (Solexa) and Applied Biosystems SOLiD (SOLiD). Compared to the approximately 350 bp read length from the 454 platform, Solexa and SOLiD provide 'ultra-short reads' that are commonly less than 75 bp. The benefits of the ultra-short read platforms include increased number of reads and decreased cost. Sequencing on these platforms can be up to 30 times less expensive compared to 454 sequencing (Shendure and Ji 2008). Recently, researchers have begun to examine the applicability of using Solexa and SOLiD for generating *de novo* transcriptomes in non-model species. For example, a transcriptome was generated for the snail *(Radix balthica*) using Solexa (Feldmeyer et al., 2011). A study in sockeye salmon (*Oncorhynchus nerka)* used SOLiD to compare results of *de novo* assembly versus mapping to public expressed sequence tag (EST) databases (Everett et al., 2011).

Everett et al. (2011) determined that assemblies using public EST databases had a higher percentage of mapped reads and higher coverage than *de novo* assemblies. These studies demonstrate that current sequence assembler performance is sufficient for producing accurate and functionally informative transcriptomes generated from ultra-short read platforms.

In addition to assembling transcriptomes, high-throughput sequencing can also be used to directly examine gene expression levels, a method referred to as RNA-Seq. In RNA-Seq, high throughput sequencing reads generated from cDNA libraries are aligned to a common reference sequence or scaffold (*e.g.* whole genome) to produce a transcriptome map that includes transcript abundance for each gene. RNA-Seq provides similar information as hybridization based microarray analysis, however, RNA-Seq has an increased dynamic range compared to hybridization-based methods (Wang et al., 2009). Furthermore, RNA-Seq is not limited to analysis of known sequences like qPCR and microarray technology, which makes RNA-seq especially appropriate for non-model species. The RNA-Seq approach has been primarily used in organisms with sequenced genomes, but very recently RNA-Seq has been applied in non-model organisms. For example, RNA-Seq was used to investigate the basis of phenotypic variation between lake trout (*Salvelinus namaycush*) ecotypes using the 454 platform (Goetz et al., 2010). RNA-Seq was also used to identify genes expressed in guppies (*Poecilia reticulata*) in response to predator cues using Solexa sequencing (Fraser et al., 2011). SOLiD transcriptome sequence reads have been used to investigate genes involved in response to temperature and settlement cues in coral larvae (*Acropora millepora*) (Meyer et al.,

2011). In the latter two studies, Solexa or SOLiD short reads were mapped to a scaffold consisting of contigs generated from other sources (i.e. 454, ESTs). These studies conclude that this approach is effective in generating accurate and informative gene expression results. RNA-Seq analysis using one set of ultra-short read data as both the scaffold and individual reads for expression analysis would be the most cost efficient, especially for those organisms where genomic resources are limited. To date, a thorough evaluation of the effectiveness of this approach has not been performed.

The primary goal of this study was to evaluate the effectiveness of utilizing the SOLiD platform to both characterize the transcriptome and analyze gene expression patterns in the Pacific oyster (*Crassostrea gigas*). As part this study, gene expression patterns between oyster populations exposed to varying degrees of anthropogenic impact were compared. RNA-Seq was performed using only the ultra-short read consensus sequences generated from *de novo* assembly as a scaffold. In order to evaluate the effectiveness of using solely ultra-short read data, RNA-Seq was also performed using publicly available transcriptome data as a scaffold and the results were compared. This work not only evaluates the use of limited ultra-short read sequence data for characterizing transcriptomes in non-model organisms, but also offers insight into the physiological responses of aquatic invertebrates in natural environments.

Materials and Methods

Site selection

Oysters were collected from two locations in Puget Sound, Washington,

USA. The sites were selected based on a difference in perceived degree of anthropogenic impact. The mouth of Big Beef Creek (BBC) in Hood Canal is a low impact site, and Drayton Harbor (DH), located in North Puget Sound, is an elevated impact site. The level of impact refers to water quality as determined by the Washington State Department of Ecology and Puget Sound Assessment and Monitoring Program (Newton et al., 2002). BBC has a relatively low population density compared to DH and routine monitoring by Washington State Department of Health show low bacterial loads. DH is ranked as the number one shellfish growing area impacted by fecal coliform pollution (WSDOH, 2006). Additionally, the density of commercial dairies and animal keeping areas in the region surrounding DH is significantly higher than BBC (WSDOH, 2006), and a municipal wastewater treatment plant discharges in proximity to DH.

Sampling and library construction

Oysters were collected from both sites in April of 2009. At each site, gill tissue was immediately sampled from 16 oysters using sterile procedures and stored in RNAlater (Ambion). RNA was isolated from individual gill tissue samples (~50mg) using Tri-Reagent (Molecular Research Center). To eliminate possible DNA carryover, total RNA was DNase treated using the Turbo DNA-free Kit (Ambion) according to the manufacturer's "rigorous" protocol. RNA from all individuals at a site (n=16) was pooled in equal quantities (650ng) to provide template for SOLiD libraries. Pooled samples were enriched for mRNA using the Ribominus Eukaryote Kit for RNA- Seq (Invitrogen) and MicroPolyA Purist Kit

(Ambion). Libraries were prepared using the SOLiD Whole Transcriptome Analysis Kit (Applied Biosystems) and sequencing was performed using the SOLiD3 System (Applied Biosystems).

Sequence analysis

All sequence analysis was performed with CLC Genomics Workbench version 4.0 (CLC Bio). Initially, sequences were trimmed based on quality scores of 0.05 (Phred (Ewing and Green 1998; Ewing et al., 1998)) and the number of ambiguous nucleotides (>2 on ends). Sequences smaller than 20 bp were also removed. *De novo* assembly was carried out using the following parameters: limit=8, mismatch cost=2 and a minimum contig size of 200 bp. For comparison purposes, quality trimmed reads were also mapped to the 82,312 contigs in GigasDatabase (version 8) (Fleury et al., 2009). Parameters used for this reference mapping included: limit=8 and mismatch cost=2. Sequences and corresponding annotations from GigasDatabase were downloaded from the *Crassostrea gigas* Public Sigenae Contig Browser (http://public-contigbrowser.sigenae.org:9090/Crassostrea_gigas). Reference mapping, using the same parameters, was used to distinguish mitochondrial transcripts using the *C. gigas* mitochondrian genome (GenBank: AF177226).

Consensus sequences from the *de novo* assembly were compared to the UniProtKB/Swiss-Prot database (http://uniprot.org) in order to determine putative descriptions. Comparisons were made using the BLAST algorithm (Altschul et al., 1990). A cutoff E-value of 1E-05 was used for annotations. Associated GO terms (Gene Ontology database: http://www.geneontology.org) were used to categorize

genes into parent categories and were assigned a functional group based on the MGI GO Slim database (URL: http://www.informatics.jax.org). The MGI GO Slim terms for 'other biological processes' and 'other metabolic processes' are not included in this analysis. For RNA-Seq analysis, expression values were measured as RPKM (reads per kilobase of exon model per million mapped reads) (Mortazavi et al., 2008) with an unspecific match limit of 10 and maximum number of 2 mismatches. Statistical comparison of RPKM values between the BBC and DH libraries was carried out using Baggerly's test (Baggerly et al., 2003), and multiple comparison correction was performed using a false discovery rate. Genes were considered differentially expressed in a given library when 1) the p-value was less than or equal to 0.05 and 2) a greater- than-or-equal-to two-fold change in expression across libraries was observed. Galaxy was used for analysis (*i.e.* table joins) during annotation and RNA-Seq analysis (Goecks et al., 2010; Blankenberg et al., 2010). RNA-Seq analysis was performed using two different scaffolds including 1) the consensus sequences from *de novo* assembly of SOLiD reads and 2) contigs in GigasDatabase.

In order to identify enriched biological themes and GO terms, the Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.7 was used (Huang et al., 2009a; Huang et al., 2009b). Specifically, corresponding UniProt accession numbers for differentially expressed genes were used as the gene list, and compared to a complete list of the corresponding UniProt accession numbers of the respective transcriptome (i.e. results of *de novo* assembly or reference mapping) for the background. Biological Process terms (DAVID 'BP Level 2' categories) were

considered significantly enriched when the p-value was less than 0.05.

Results

Crassostrea gigas SOLiD sequencing

After quality trimming, 20.7 and 24.6 million reads (average length: 40.6 bp) remained from the BBC and DH cDNA libraries, respectively. A majority of the reads (98%) corresponded to nuclear transcripts with the other 2% mapping to mitochondria protein coding genes. The quality trimmed reads from each library were combined for *de novo* assembly and reference mapping. All sequence data has been submitted to the NCBI Short Read Archive under accession number SRP007621.

De novo assembly

De novo assembly of reads from the combined libraries resulted in 18,510 consensus sequences with an average length of 276 bp. Twenty three percent of the reads assembled using this approach. The average number of assembled reads per consensus sequence was 454 and the mean coverage was 61.7x (Figure 1).

Reference mapping (GigasDatabase)

SOLiD reads were also mapped to publicly available *C. gigas* transcriptomic resources (GigasDatabase v8). Reads from the combined libraries mapped to 64,645 of the 82,314 contigs in the database. The average number of reads per contig was 376 and the mean coverage was 15.8x (Figure 1). See Table 1 for a full comparison

of results of the *de novo* assembly compared to reference mapping.

De novo assembly: annotation

A total of 7,724 consensus sequences could be annotated, 3,931 of which could be classified using GO Slim terms. The most highly represented biological process was transport, followed by protein metabolism (data not shown). Of those consensus sequences associated with transport a majority were involved in protein and ion transport. Comparatively, 7,296 of the GigasDatabase contigs with mapped reads were annotated with biological process GO terms. When the associated GO terms were evaluated, two of the most highly represented biological processes identified after binning into broader GO Slim terms included protein and RNA metabolism (data not shown).

De novo assembly: identification of novel transcripts

Short read consensus sequences generated from *de novo* assembly were compared to GigasDatabase v8 to identify novel sequences. Approximately 10% of the sequences (1,776) did not have a significant match (E-value > 1.0E-01). Of these, 742 could be annotated (see Supplementary Table 1) and 690 could be classified using GO Slim. The 4 most highly represented biological processes included: transport, developmental processes, cell organization and biogenesis, and cell adhesion (Figure 2).

RNA-Seq analysis

De novo-based RNA-Seq

RNA-Seq analysis using the *de novo* assembled short read consensus sequences as the scaffold identified 2,991 differentially regulated features. Most of these features represented moderately expressed transcripts (100 – 10,000 total reads), but 20% were rare transcripts (<100 total reads). Six consensus sequences were expressed uniquely in the BBC library and 5 were expressed only in the DH library. None of the uniquely expressed features could be annotated. Of differentially expressed features with reads in both libraries, 1200 were expressed higher in the BBC library and 1791 were expressed higher in the DH library. A subset of the differentially expressed features (751 in BBC and 313 in DH, respectively) could be annotated (see Supplementary Table 2). A majority of these annotated features represented a 2 fold difference, but overall differences ranged between 2 – 409 fold.

Functional enrichment analysis identified 15 biological processes that were overrepresented in the differentially expressed gene set (Figure 3). The most significantly enriched process was cell adhesion (p-value = 8E-15), followed by cell recognition (p-value = 5E-5).

Reference-based RNA-Seq

For comparison, RNA-Seq was also performed using GigasDatabase v8 as the scaffold. In total, 427 differentially expressed features were identified. Of those, 239 were expressed higher in the BBC library and 189 were expressed higher in the DH library. Of these, 216 contigs could be annotated. Table 1 provides a comparison of

data from both RNA-Seq procedures.

Functional enrichment analysis identified three biological processes that were enriched in the differentially expressed gene set. The most significantly enriched process was microtubule-based processes followed by oxidation reduction and cell recognition. One term, cell recognition (p value = 6E-3), overlapped between the *de novo* based and reference based RNA-Seq analysis. The other terms were unique to each analysis.

Discussion

This study evaluates the effectiveness of using high-throughput, short-read sequencing technology to characterize the transcriptome of taxa with limited genomic resources. Specifically, SOLiD sequencing was carried out on cDNA libraries from Pacific oysters from two locations with differing anthropogenic influence. Sequence assembly and RNA-Seq analyses were carried out using resources generated solely as part of this study and compared to respective analysis using a publicly available transcriptome database. We found that limited ultra short read sequence data can provide valuable information about transcriptome activity. Furthermore, we provide new genomic resources for *C. gigas* and have identified differences in oysters from areas that have experienced different degrees of human impact. These combined data significantly contribute to what we know about oyster biology but also offer a framework for efficiently characterizing transcriptomic differences in species lacking sequenced genomes. Advantages and limitations of using short-read sequencing technology for gene discovery and RNA-Seq analysis are discussed.

Gene Discovery

The number of Pacific oyster consensus sequences generated from *de novo* assembly is comparable to similar studies in sockeye salmon (Everett et al., 2011) and *Radix balthica* (Feldmeyer et al., 2011). However, as expected, mean contig length (276bp) was shorter than transcriptome characterizations that use 454 pyrosequencing. Recent studies in guppies (Fraser et al., 2011) and chum salmon (Seeb et al., 2010) produced mean contig lengths of 464bp and 412bp, respectively. In the current study our average coverage was 62x compared to 5x reported by Seeb et al. (2010). Dohm et al. (2008) have indicated greater than 20x coverage is sufficient to minimize effects of sequencing errors. We were able to annotate 42% of the consensus sequences generated from the *de novo* assembly. This included a large number of transcripts (742 contigs) not present in public databases. The number of novel sequences identified is slightly higher than reported in studies using Sanger sequencing for gene discovery in *C. gigas* (Gueguen et al., 2003; Roberts et al., 2009). The functional classification of the novels transcripts identified using SOLiD sequencing were highly diverse with a large proportion being involved in transport, developmental processes, stress response, and cell adhesion.

Several genes of interest were identified in the novel contigs, many of which are associated with response to stress. A number of these transcripts have been shown to be involved specifically in the immune response. For instance, a sequence with similarity to dual oxidase 2 was identified. In *Drosophila melanogaster* this protein regulates the production of reactive oxygen species in response to infectious and commensal microbes (Ha et al., 2009). The mitogen-activated protein kinase (MAPK)

signaling pathway is involved in phagocytosis and the prophenoloxidase cascade in invertebrates (Lamprou et al. 2007). A subset of genes involved in this pathway has been previously identified in a *C. gigas* (Roberts et al., 2008). Here we identified a novel sequence in this pathway, mitogen-activated protein kinase kinase kinase 7 (M3K7). Another important component of the invertebrate immune system are bactericidal enzymes. A transcript similar to myeloperoxidase (MPO), which functions as a bactericide by generating hypochlorous acid (Harrison and Schultz, 1976), was present in the *de novo* consensus sequences. While this protein has been identified in molluscs based on its catalytic activity (Shenk et al., 1990), this is the first time the nucleotide sequence has been reported in oysters. An additional sequence of interest possesses homology to a SAM domain and HD domain-containing protein, which has been shown to be involved in anti-viral responses in humans (Rice et al., 2009).

Oysters and other coastal invertebrates are frequently exposed to xenobiotics. One of the first steps involved in the metabolism and subsequent exclusion of xenobiotics is binding of a ligand (i.e. aromatic hydrocarbon) to the aryl hydrocarbon receptor. As part of this sequencing effort we identified a transcript similar to aryl hydrocarbon receptor nuclear translocator (ARNT). ARNT encodes a protein that forms a complex with the ligand-bound aryl hydrocarbon receptor, and is required for receptor function (Hoffman et al., 1991). Activation of the aryl hydrocarbon receptor initiates transcription of cytochrome p450 oxidases. Several genes in this family have been previously reported in *C. gigas* (Roberts et al., 2009). Xenobiotic conjugates and metabolites are eventually excreted from the cell by membrane transporters in the multidrug resistance

protein family. A contig generated as part of the *de novo* sequencing effort identified a transcript similar to multidrug resistance protein 1. Together the new sequences identified here demonstrate that limited ultra-short read sequencing provides an important resource for gene discovery.

When reference mapping was carried out, the proportion of reads that could be putatively annotated increased. While we have demonstrated that the sole use of a limited short-read sequencing data set can provide cost-effective, valuable, novel genomic information, an available scaffold (i.e. EST contigs, genome) can provide benefits with respect to number of mapped reads and subsequent ability to annotate.

RNA-Seq

Using limited short-read data we were able to effectively perform RNA-Seq analysis in the Pacific oyster. This is one of the first studies describing RNA-Seq analysis using solely ultra-short read data, along with other very recent publications in the crustacean *Pandalus latirostris* (Kawahara- Miki et al., 2011) and insect *Plutella xylostella* (Etebari et al., 2011). A similar approach is Tag-Seq, which utilizes short (<30 bp) tags, generally from the 3' ends of transcripts to characterize differentially expressed genes. A recent study by de Logeril et al. (2011) utilized Tag-Seq to identify approximately 4,000 unique, immune responsive genes in *C. gigas*. In the current study, we were able to identify and annotate 1,064 differentially expressed transcripts in *C. gigas* populations exposed to varying degrees of anthropogenic impact. Tag-Seq can be relatively less expensive than RNA-Seq with respect to coverage, however a reference scaffold is required. In addition, because tags are

usually generated from a single end of a transcript, RNA-Seq analysis, as described here, has the advantage of identifying and quantifying novel transcripts (Cullum et al., 2011). In our RNA-Seq study, 18% of the differentially expressed transcripts were novel, representing a significant contribution to genomic resources. Together these studies demonstrate how advances in sequencing technology will continue improve our ability to characterize physiological responses in non-model organisms.

When comparing differentially expressed genes in oysters from the two sites, there was a large difference in the number of differentially expressed genes depending on whether the RNA-Seq was based on *de novo* assembly or reference based. Specifically, RNA-Seq performed using the *de novo* assembled consensus sequences reported seven-times as many differentially expressed genes as the RNA-Seq analysis using GigasDatabase v8. One possible explanation for this discrepancy is that using the *de novo* assembly as a scaffold may result in multiple sequences representing the same gene. In other words, the consensus sequences are relatively short and fragments representing different regions of the same gene may not overlap. As a majority of these differentially expressed genes could not be annotated, it is difficult to determine the precise impact of this possibility. However, 889 of the 1064 annotated, differentially expressed genes were deemed unique based on the protein identification code of the UniProt ID, suggesting there may be other factors contributing to this difference. As would be expected, based on the proportion of differentially expressed genes, the number of enriched GO biological processes identified was also different between the two analyses. It is likely that this difference is related to the scaffold itself, as all genes making up the scaffold are used as the

"background" for the enrichment analysis. Therefore, it is possible that the *de novo* based enrichment analysis is more biologically relevant, as the background is a better representation of the genes expressed under similar conditions.

RNA-Seq analysis revealed that the set of transcripts differentially expressed between BBC and DH were most significantly enriched in genes associated with cell adhesion. In general, cell adhesion can be divided into to two general types. The first is a stable cell-cell adhesion that is critical for the organization of tissues. The second is a transient cell adhesion involved in processes such as cell adhesion between hemocytes and cell adhesion to pathogens. This transient type of cell adhesion is a critical part of invertebrate innate immunity by way of recognition of non-self particles, as well as chemotaxis and aggregation of hemocytes (reviewed by Johansson 1999). The specific genes that are contributing to the difference between the two libraries include integrins, laminins and cadherins, which are expressed approximately 2 – 4 times higher in the DH library. While the precise biological role for this increased expression cannot be determined from this study, it could indicate the presence of specific contaminants in the environment. For instance, integrin expression increased in response to pathogen exposure in white shrimp (*Litopenaeus vannemai*) (Lin et al., 2010). In addition, estrogen exposure stimulates hemocyte binding to laminin 1 and collagen IV in mussels (*Mytilus galloprovincialis*) (Koutsogiannaki and Kaloyianni, 2011). While we can only speculate on the functional role, it is interesting to note that it is consistent with the environmental data from this locale, as DH is a site close to urban wastewater discharge and intensive agriculture exposure. However, additional research is required to determine

the role of genes associated with cell adhesion and environmental exposures in oysters.

Conclusions

Ultra-short read sequencing technology, such as SOLiD, provides a powerful and effective means for gene discovery and expression analysis in organisms with limited genomic resources available. We have shown that it is technically possible and efficient to use this approach to 1) generate transcriptomic resources, 2) identify novel genes, and 3) perform RNA-Seq analysis. In terms of gene expression, *de novo* based RNA-Seq analysis does not rely on previous transcriptome information and results can be annotated at the biological process level. As high-throughput sequencing platforms continue to improve, they will serve as important tools for examining environmental physiology of non-model organisms.

Acknowledgments

This work was partially funded by University of Washington, School of Aquatic and Fishery Sciences. MRG's graduate appointment is funded in part by the Victor and Tamara Loosanoff Endowed Fellowship and the John G. Peterson Scholarship. The authors would also like to thank Caroline G Storer for comments on earlier versions of this manuscript.

Works cited

Altschul, S.F., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool, J. Mol. Biol. 215, 403-410.

Baggerly, K., Deng, L., Morris, J., Aldaz, C., 2003. Differential expression in SAGE: accounting for normal between-library variation, Bioinformatics 19, 1477-1483.

Blankenberg, D., Von Kuster, G., Coraor, N., Ananda, G., Lazarus, R., Mangan, M., Nekrutenko, A., Taylor, J., 2010. Galaxy: a web-based genome analysis tool for experimentalists, Curr. Protoc. Mol. Biol. 19, Unit 19.10.1-21.

Craft, J.A., Gilbert, J.A., Temperton, B., Dempsey, K. E., Ashelford, K., Tiwari, B., Hutchinson, T.H., Chipman, J.K., 2010. Pyrosequencing of Mytilus galloprovincialis cDNAs: Tissue-specific expression patterns, PlosOne 5, e8875.

Cullum, R., Alder, O., Hoodless, P.A., 2011. The next generation: Using new sequencing technologies to analyse gene regulation, Respirology 16, 210–22.

de Lorgeril, J., Zenagui, R., Rosa, R.D., Piquemal, D., Bachère, E., 2011. Whole transcriptome profiling of successful immune response to Vibrio infections in the oyster Crassostrea gigas by Digital Gene Expression analysis, PLoS One 6(8), e23142.

Dohm, J.C., Lottaz, C., Borodina, T., Himmelbauer, H., 2008. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing, Nucleic Acids Res. 36, e105.

Etebari, K., Palfreyman, R., Schlipalius, D., Nielsen, L., Glatz R., Asgari, S., 2011. Deep sequencing-based transcriptome analysis of Plutella xylostella larvae parasitized by Diadegma semiclausum, BMC Genomics 12, 446.

Everett, M., Grau, E., Seeb, J., 2011. Short reads and non-model species: exploring the complexities of next generation sequence assembly and SNP discovery in the absence of a reference genome, Mol. Ecol. Res. 11, 93–108.

Ewing, B., Green, P., 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities, Genome Res. 8, 186–194.

Ewing, B., Hillier, L., Wendl, M.C., Green, P., 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment, Genome Res. 8, 175–185.

Feldmeyer, B., Wheat, C.H., Krezdorn, N., Rotter, B., Pfenninger, M., 2011. Short read Illumina data for the de novo assembly of a non-model snail species transcriptome (Radix balthica, Basommatophora, Pulmonata), and a comparison of assembler performance, BMC Genomics 12, 317.

Fleury, E., Huvet, A., Lelong, C., de Lorgeril, J., Boulo, V., Gueguen, Y., Bachère, E., Tanguy, A., Moraga, D., Fabioux, C., Lindeque, P., Shaw, J., Reinhardt, R., Prunet, R., Davey, G., Lapègue, S., Sauvage, C., Corporeau, C., Moal, J., Gavory, F., Wincker, P., Moreews, F., Klopp, C., Mathieu, M., Boudry P., Favrel, B., 2009. Generation and analysis of a 29,745 unique Expressed Sequence Tags from the Pacific oyster (Crassostrea gigas) assembled into a publicly accessible database: the GigasDatabase, BMC Genomics 10, 341.

Fraser, B.A., Weadick, C.J., Janowitz, I., Rodd, H., Hughes, K.A., 2011. Sequencing and characterization of the guppy (Poecilia reticulata) transcriptome, BMC Genomics 12, 202.

Goecks, J., Nekrutenko, A., Taylor, J., 2010. The Galaxy Team. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences, Genome Biol. 1, R86.

Goetz, F. Rosauer, D., Sitar, S., Goetz, G., Simchick, C., Roberts, S., Johnson, R., Murphy, C., Bronte C., MacKenzie, S., 2010. A genetic basis for the phenotypic differentiation between siscowet and lean lake trout (Salvelinus namaycush), Mol. Ecol. 19, 176–196.

Gueguen, Y., Cadoret, J.-P., Flament, D., Barreau-Roumiguière, C., Girardot, A.-L., Garnier, J., Horeau, A., Bachère, E., Escoubas, J.-M., 2003. Immune gene discovery by expressed sequence tags generated from hemocytes of the bacteria-challenged oyster, Crassostrea gigas, Gene 303, 139–145.

Ha, E., Lee, K., Seo, Y.Y., Kim, S., Lim, J., Oh, B., Kim J., Lee, W., 2009. Coordination of multiple dual oxidase–regulatory pathways in responses to commensal and infectious microbes in Drosophila gut, Nat Immunol 10, 949 – 957.

Harrison, J.E., Schultz, J., 1976. Studies on the chlorinating activity of myeloperoxidase, J Biol Chem 251, 1371-137.

Hoffman, E.C., Reyes, H., Chu, F.F., Sander, F., Conley, L.H., Brooks, B.A., Hankinson, O., 1991. Cloning of a factor required for activity of the Ah (dioxin) receptor, Science 252, 954–958.

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

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

Johansson, M.W., 1999. Cell adhesion molecules in invertebrate immunity, Dev. Comp. Immunol. 23, 303– 315.

Kawahara-Miki, R., Wada, K., Azuma, N., Chiba, S., 2011. Expression profiling without genome sequence information in a non-model species, Pandalid shrimp (*Pandalus latirostris*), by next-generation sequencing, PLoS One 6(10), e26043.

Koutsogiannaki, S., Kaloyianni, M., 2011. Effect of 17β-estradiol on adhesion of Mytilus galloprovincialis hemocytes to selected substrates. Role of alpha2 integrin subunit, Fish Shellfish Immunol. 31, 73-80. Lamprou, I., Mamali, I., Dallas, K., Fertakis, V., Lampropoulou, M., Marmaras VJ (2007) Distinct signalling pathways promote

phagocytosis of bacteria, latex beads and lipopolysaccharide in medfly haemocytes. Immunology 121:314–327

Lin, Y.C., Tayag, C. M., Huang, C.L., Tsui, W.C., Chen, J.C., 2010. White shrimp Litopenaeus vannamei that had received the hot-water extract of Spirulina platensis showed earlier recovery in immunity and up- regulation of gene expression after pH stress, Fish Shellfish Immunol. 29, 1092–1098.

Meyer, E., Aglyamova, G., Wang, S., Buchanan-Carter, J., Abrego, D., Colbourne, J.K., Willis, B.L., Matz, M.V., 2009. Sequencing and de novo analysis of a coral larval transcriptome using 454 GS-Flx, BMC Genomics 10, Meyer, G.V. Aglyamova and M.V. Matz, Profiling gene expression responses of coral larvae (Acropora millepora) to elevated temperature and settlement inducers using a novel RNA-Seq procedure. Mol. Ecol. 20, 3599–3616.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., Wold, B., 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq, Nat. Methods 5, 585-587.

Newton, J. A., Albertson, S.L., Voorhis, K.V., Maloy, C., Siegel, E., 2002. Washington State Marine Water Quality, 1998 through 2000, Publication #02-03-056, 111pp.

Novaes, E., Drost, D.R., Farmerie, W.G., Pappas, G.J., Grattapaglia, D., Sederoff, R.R., Kirst, M., 2008. High-throughput gene and SNP discovery in Eucalyptus grandis, an uncharacterized genome, BMC Genomics 9, 312.

Rice, G.I., Bond, J., Asipu, A., Brunette, R.L., Manfield, I.W., Carr, I.M., Fuller, J.C., Jackson, R.M., Lamb, T., Briggs, T.A., Ali, M., Gornall, H., Couthard, L.R., Aeby, A., Attard-Montalto, S.P., Bertini, E., Bodemer, C., Brockmann, K., Brueton, L.A., Corry, P.C., Desguerre, I., Fazzi, E., Cazorla, A.G., Gener, B., Hamel, B.C.J., Heiberg, A., Hunter, M., van der Knaap, M.S., Kumar, R., Lagae, L., Landrieu, P.G., Lourenco, C.M., Marom, D., McDermott, M.F., van der Merwe, W., Orcesi, S., Prendiville, J.S., Rasmussen, M., Shalev, S.A., Soler, D.M., Shinawi, M., Spiegel, R., Tan, T.Y., Vanderver, A., Wakeling, E.L., Wassmer, E., Whittaker, E., Lebon, P., Stetson, D.B., Bonthron D.T., Crow, Y.J., 2009. Mutations involved in aicardi-goutieres syndrome implicate samhd1 as regulator of the innate immune response, Nat Genet 41, 829–832.

Roberts, S.B., Goetz, G., White, S., Goetz, F., 2008. Analysis of genes isolated from plated hemocytes of the Pacific oysters Crassostrea gigas, Mar. Biotechnol. 11, 24-44.

Seeb, J.E., Pascal, C.E., Grau, E.D., Seeb, L.W., Templin, W.D., Roberts, S.B., Harkins, T., 2010. Transcriptome sequencing and high-resolution melt analysis advance SNP discovery in duplicated salmonids, Mol. Ecol. Resour. 11, .

Shendure, J., Ji, H.L., 2008. Next-generation DNA sequencing, Nature Biotechnol. 26, 1135–1145.

Vera, J.C., Wheat, C.W., Fescemyer, H.W., Frilander, M.J., Crawford, D.L., Hanski, I., Marden, J.H., 2008. Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing, Mol. Ecol. 17, 1636-1647.

Wang, Z., Gerstein, M., Snyder, M., 2009. RNA-Seq: A revolutionary tool for transcriptomics, Nat. Rev. Genet. 10, 57–63.

Washington State Department of Health: Office of Shellfish and Water Protection, A Report for the Puget Sound Assessment and Monitoring Program: Atlas of Fecal Coliform Pollution in Puget Sound: Year 2005, 2006. Publication #332-061, 80pp.

Figures

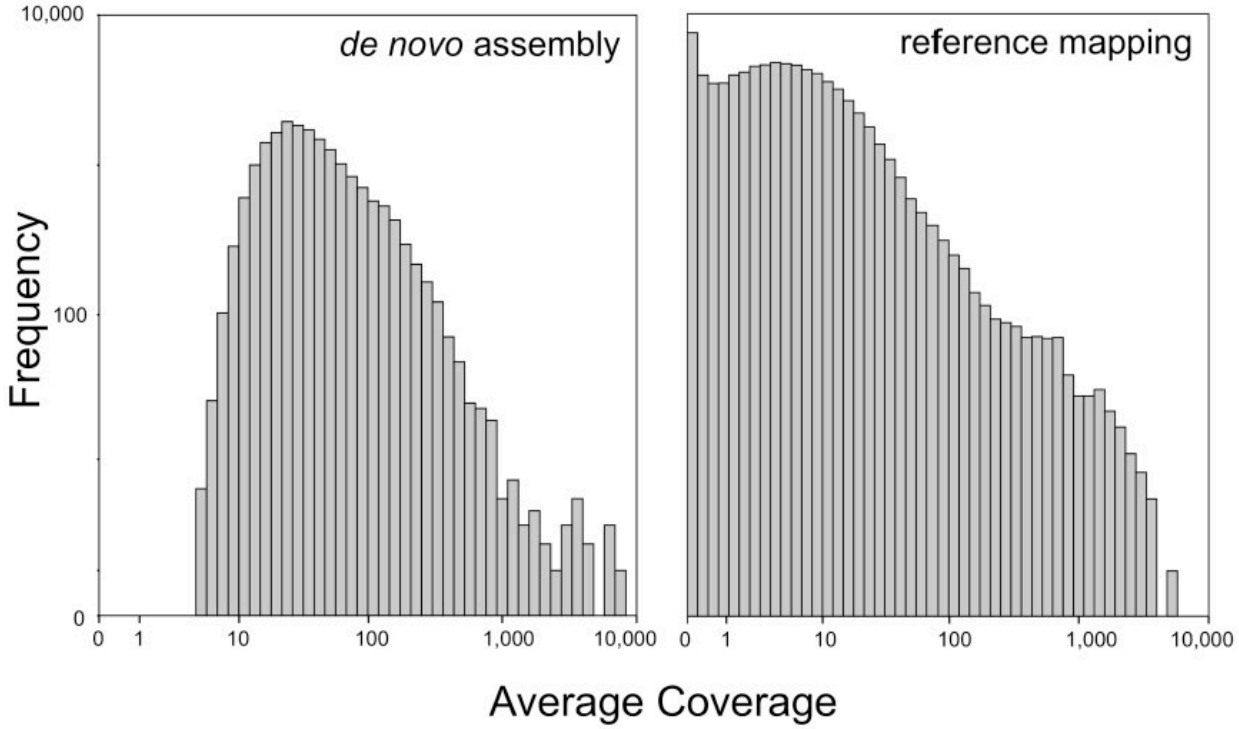


Figure 1. *Coverage distribution for de novo assembly and reference mapping.* Histograms showing average read coverage for de novo assembly and reference mapping to GigasDatabase v8 for the combined C. gigas SOLiD transcriptome libraries.

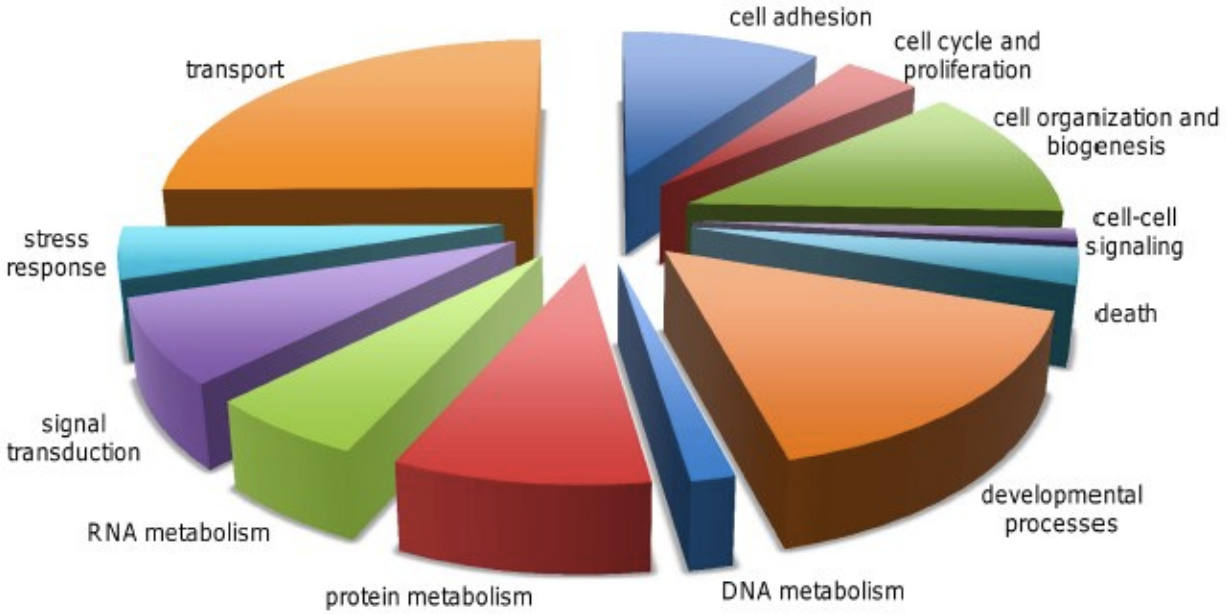


Figure 2. *Functional classification of novel transcripts identified by de novo assembly of the combined SOLiD transcriptome libraries.* Categories represent 'GO Slim' terms associated with Biological Processes.

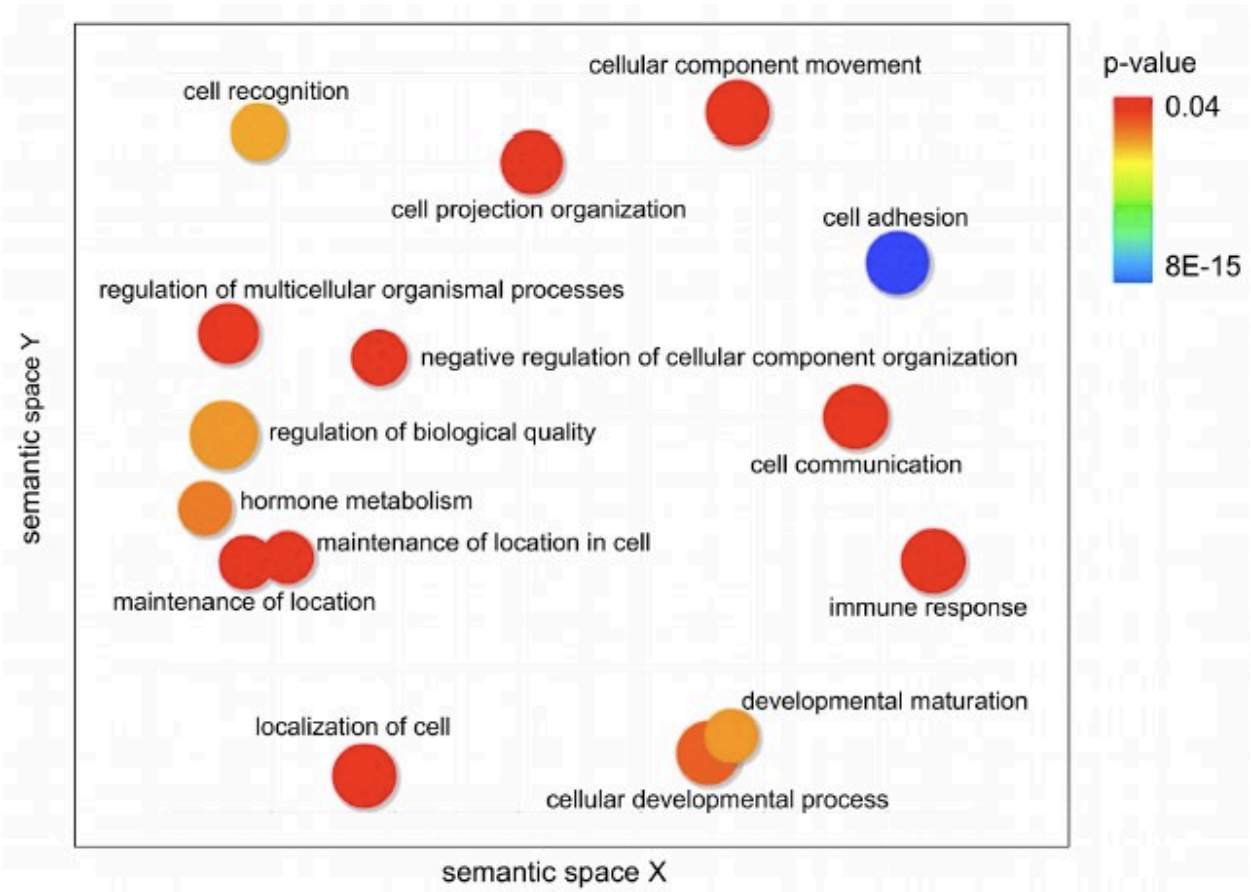


Figure 3. *Gene ontology categories overrepresented in the differentially expressed gene set*. Color indicates p-value of the enrichment and size is proportional to the number of genes in the category. Spatial arrangement reflects a grouping of categories by semantic similarity.

Supplementary Information

Supplementary information is available at the online version of Gavery & Roberts, 2012.

Tables

Table 1. *Summary of assembly and RNA-Seq statistics for de novo assembly and reference mapping (GigasDatabase v8).*

		De novo assembly	**Reference mapping**
Assembly	mapped reads	8,407,963	29,107,760
	unmapped reads	36,944,698	16,244,901
	contigs	18,510	77,433
	average contig length	276	554
	average contig coverage	62	16
	contigs annotated to GO Slim	3,931	7,296
RNA-Seq	differentially expressed genes	2991	427
	enriched GO biological process	15	3

Chapter II: DNA methylation patterns provide insight into epigenetic regulation in the Pacific oyster (*Crassostrea gigas*)

A version of this chapter was published in BMC Genomics as: Gavery M and Roberts SB. 2010. DNA methylation patterns provide insight into epigenetic regulation in the Pacific oyster (*Crassostrea gigas)*. BMC Genomics 11:483. doi:10.1186/1471-2164-11-483

Abstract

Background
DNA methylation is an epigenetic mechanism with important regulatory functions in animals. While the mechanism itself is evolutionarily ancient, the distribution and function of DNA methylation is diverse both within and among phylogenetic groups. Although DNA methylation has been well studied in mammals, there is limited data on invertebrates, particularly molluscs. Here we characterize the distribution and investigate potential functions of DNA methylation in the Pacific oyster (*Crassostrea gigas*).

Results
Methylation sensitive PCR and bisulfite sequencing PCR approaches were used to identify CpG methylation in *C.gigas* genes and demonstrated that this species possesses intragenic methylation. *In silico* analysis of CpGo/e ratios in publically available sequence data suggest that DNA methylation is a common feature of the *C.gigas* genome, and specific functional categories of genes have significantly different levels of methylation.

Conclusions
The Pacific oyster genome displays intragenic DNA methylation and contains genes necessary for DNA methylation in animals. Results of this investigation suggest that DNA methylation has regulatory functions in *Crassostrea gigas*, particularly in gene families that have inducible expression, including those involved in stress and environmental responses.

Background

Epigenetic mechanisms induce changes in gene activity without alteration to the underlying DNA sequence [1]. Common epigenetic mechanisms include DNA methylation, histone modifications and non-coding RNA activity. The most well-studied of these is DNA methylation, which refers to the addition of a methyl group to position 5 of cytosines. In animals, this reaction is catalyzed by a family of enzymes called DNA (cytosine-5) methyltransferases (DNMTs) and occurs almost exclusively in CpG dinucleotides. DNA methylation is typically associated with transcriptional repression, and is primarily achieved by methylation in gene promoters [2, 3, 4]. The functional significance of DNA methylation in vertebrates includes providing genomic stability [5], regulation of imprinted genes [6] and X-chromosome inactivation [7]. In mammals, DNA methylation is essential for development and cell differentiation [8] and defects or unintended changes in DNA methylation can have deleterious consequences such as embryonic lethality [9] and tumorgenesis [10]. DNA methylation, like many epigenetic marks, may be heritable, therefore unintended changes as a result of environmental exposures or other processes can be passed on for multiple generations [11].

The extent of cytosine methylation varies considerably among eukaryotes. In vertebrates, approximately 70-80% of cytosines in CpG dinucleotides are methylated [12], a pattern referred to as global methylation. Invertebrates display a wide range of DNA methylation, from very limited methylation in *Drosophilia melanogaster* [13] and *Caenorhabditis elegans* [14] to a mosaic-pattern of methylation in the sea urchin (*Strongylocentrotus purpuratus*) [15] and *Ciona intestinalis* [16 and 17]. Bird and Taggart [12] concluded that

there were three general types of methylation patterns: the 'insect-type' which shows little to no methylation, the 'echinoderm-type', whose genomes contained both methylated and non-methylated fractions, and the heavily methylated 'vertebrate-type'. Recent studies in the honey bee (*Apis mellifera*) indicate these patterns may be more complex [18 and 19]. In contrast to *D. melanogaster,* which lacks most of the classical DNMTs [20] and shows limited cytosine methylation [21], *A.mellifera* has a fully functional set of DNA methylation enzymes and shows substantial methylation across its genome [18].

In vertebrates, regulation of transcription by DNA methylation is accomplished by differential patterns of methylation in intergenic regions, namely gene promoters [2, 3, 4]. In contrast, there are no significant differences being reported in the methylation status of gene promoters in invertebrates where methylation appears to be targeted specifically to transcription units [17 and 22]. Computational analyses of the methylation status of *A. mellifera* genes have provided some of the first evidence supporting a regulatory role of intragenic DNA methylation in invertebrates [19 and 23]. In these studies, genes associated with general metabolic or 'housekeeping' functions were predicted to be hyper-methylated, whereas caste-specific genes were preferentially hypo-methylated. This functional clustering suggests DNA methylation serves to regulate gene transcription in *A. mellifera*, however, it is uncertain if this function is conserved across invertebrate taxa. Furthermore, it is unclear exactly how intragenic cytosine methylation directly affects transcription.

Studies in *A. mellifera* and others illustrate the diversity of DNA methylation patterns in invertebrate taxa and highlight gaps in our understanding of the evolutionary and functional significance of DNA methylation. One taxonomic group that has been notably absent from these investigations is the phylum Mollusca. Molluscs were first categorized as having 'echinoderm-type' DNA methylation patterns based on experimental evidence using the common mussel (*Mytilus edulis*) [12]. Since then, there has been little investigation of DNA methylation in molluscs with the exception of evidence suggesting the presence CpG methylation in the clam, *Donux truculus* [24]. In addition to increasing our understanding of the evolution of DNA methylation in invertebrate taxa, this study provides an opportunity to evaluate the Pacific oyster (*Crassostrea gigas*) as a model organism for analyzing DNA methylation in an aquatic species. Bivalve molluscs are important bioindicators [25] and elucidating the functional significance of DNA methylation in these organisms may prove valuable for understanding the effects of environmental stress in aquatic organisms. Here, we report the first investigation into DNA methylation profiles in the genome of the Pacific oyster. We confirm the presence of intragenic CpG methylation in *C.gigas*. We also demonstrate a relationship between predicted methylation status and gene function, suggesting that DNA methylation performs important regulatory functions in *C.gigas*. Implications of these findings are discussed in both an evolutionary and ecological context.

Methods

Animal collection & DNA isolation

Oysters used in this study were collected from naturalized *C.gigas* populations in Puget Sound, Washington. To isolate genomic DNA, 25 mg of gill tissue was processed according to manufacturer's protocol using the DNeasy Blood & Tissue Kit (Qiagen, CA).

Methylation Sensitive PCR

Oyster genomic DNA was enzyme digested with either HpaII or MspI. Five immune related genes containing one or more CCGG recognition sites and covering a broad range of predicted methylation status (based on CpGo/e) were selected from a set of ESTs generated from a cDNA library of plated hemocytes [52]. PCR primers were designed to flank one or more restriction sites. Primer sequences are provided in Table 3. Quantitative PCR was performed using digested (HpaII or MspI) and undigested samples using 1x Immomix Master Mix (Bioline USA Inc., Boston, MA), 2uM SYTO-13 (Invitrogen, Carlsbad, CA) and 0.2uM forward and reverse primers in an Opticon 2 System (Bio-Rad, Hercules, CA) with the following cycling conditions: 10 min at 95C , followed by 37 cycles of 15 sec at 95C, 30 sec at 55C, and 1 min at 72C and a final extension at 72C for 10 min. Results were scored qualitatively based on the presence or absence of amplification as determined by florescence.

Bisulfite Conversion and Sequencing

Genomic DNA was bisulfite converted using the Epitect Bisulfite conversion kit (Qiagen, Carlsbad, CA). Briefly, 1.75ug of DNA was subjected to treatment with sodium bisulfite at increased temperature to deaminate unmethylated cytosine residues to uracil. Following treatment, the solution was desulfonated on a column, washed and eluted.

To identify methylated cytosines in expressed regions of the oyster genome, Meth Primer [53] was used to design primers that flank multiple CpG sites, but do not contain CpGs. Primer sequences are provided in Table 3. The mean expected amplicon length for bisulfite primers was ~180bp. PCR of bisulfite treated samples (54ng/ PCR reaction) was carried out using Taq DNA Polymerase Master Mix (Apex BioResearch Products, Research Triangle Park, NC) for 10 min at 95C, followed by 40 cycles of 15 sec at 95C, 30 sec at 55C, and 1 min at 72C and a final extension at 72C for 10min.

PCR products were separated using gel electrophoresis. Single bands were excised from the gel, purified using Ultra-DA purification columns (Ambion, Foster City, CA) and cloned using TOPO TA Cloning kit (Invitrogen). Four clones were sequenced for each primer pair. Methylation status was determined by comparing the sequence of bisulfite treated DNA to sequence of untreated DNA using Geneious 4.5.4 software (Biomatters Ltd., Aukland, NZ) and annotated using BLAST [54].

in silico analysis: predicted DNA methylation status

For *in silico* analysis, the non-redundant *C.gigas* expressed sequence tag (EST) contig database, 'GigasDatabase' version 6 (http://public-contigbrowser.sigenae.org:9090/Crassostrea_gigas /index.html, [29]), was utilized. Analysis was limited to annotated sequences (n=12,210) in order to be confident that sequences were reported in the 5' to 3'direction. It should be noted that this transcriptomic dataset is appropriate for predicting methylation status of the *C. gigas* genome as investigation into other invertebrate species show that DNA methylation is specifically targeted to transcribed regions of the genome [17 and 12]. CpG observed/expected ratio (CpGo/e) was calculated using the following equation where l is the number of nucleotides in the contig:

$$\text{CpGo/e} = \frac{\text{number of CpG}}{\text{number of C} \times \text{number of G}} \times \frac{l^2}{l-1}$$

To evaluate the distribution of Pacific oyster EST contigs a mixture model was fit to the CpGo/e ratios using the mixtools package in R (Benaglia, 2009) yielding a two component mixture where $p_1 + p_2 = 1$. Hence the data C_i, are distributed as

$$Ci \sim p_1 N(\mu_1,\sigma_1) + p_2 N(\mu_2,\sigma_2).$$

The log likelihood statistic of the bimodal mixture model was compared to the

The log likelihood statistic of the bimodal mixture model was compared to the normal null model to test for a significant improvement in fit.

In order evaluate the variation of CpGo/e within and among functional classes of genes, contigs from the GigasDatabase annotated with a biological process GO term (n = 10,699 contigs) were assigned a functional group based on the MGI GO Slim database (URL: http://www.informatics.jax.org) [57]. Since each contig may have multiple GO terms,

contigs were allowed to fall into multiple GO Slim bins. However, to avoid weighting within a single category, an individual contig was not allowed to be included more than once in a single GO Slim bin. The mean CpGo/e and standard errors were calculated for each GO Slim term. A one-way ANOVA followed by Tukey's test for multiple comparisons was performed using SPSS software (SPSS Inc., Chicago, IL). A significance level of $p < 0.05$ was accepted.

Results

Methylation Sensitive PCR

A Methylation Sensitive PCR (MSP) approach was used to identify specific methylated sites. Five genes associated with immune function were analyzed and methylation status determined (Table 1). Methylation status can be concluded based on the presence or absence of a PCR product in the methylation sensitive HpaII digest. Of the five genes analyzed, CpG methylation was confirmed for *heat shock protein 70* (*hsp70*), whereas no methylation was detectable at restriction site(s) for the other sequences examined. The CpG observed to expected ratios (see Methods for calculation) are included in Table 1 for each gene. It should be noted that *hsp70* has the lowest ratio of all the genes analyzed (0.57). This low ratio is predictive of a hyper-methylated status, which is confirmed here by MSP.

Bisulfite Sequencing PCR

In order to describe methylated cytosines outside of CCGG sites, Bisulfite Sequencing PCR (BSP) was used. Five genes predicted to be hyper-methylated, and five predicted to

be hypo-methylated (based on CpG observed to expected ratio) were randomly selected for analysis. Valid PCR products were produced for two of the genes. This is a typical result as the conversion of unmethylated cytosines results in challenges for primer specificity. Four individual clones were sequenced for each of the two products. There was a 100% conversion rate for non-CpG cytosines for each of the clones sequenced. In the first fragment, a 136 bp fragment with homology to the amino terminal fragment of the human neuromedin-u receptor [Swiss-Prot: Q9GZQ4], one of seven CpGs sites displayed methylation in 25% of the clones sequenced (Figure 1(a)). In a second fragment, one of two CpGs sites was determined to be methylated in 50% of the clones sequenced in a 93 bp region (Figure 1(b)). The latter sequence has significant homology to human bromodomain adjacent to zinc finger domain, 1A [Swiss-Prot: Q9NRL2].

in silico analysis of C.gigas transcriptome

The ratio of observed to expected CpG dinucleotides (CpGo/e) was used to predict methylation status in the *C.gigas* transcriptome. This approach is based on the known hyper-mutability of methylated cytosines, which readily deaminate to thymine residues [26]. This CpG mutation is not easily corrected by DNA repair machinery, and as a result consistently methylated regions of DNA are depleted of CpG dinucleotides over evolutionary time [27]. Consequently, regions of DNA with a low CpGo/e are predicted to be methylated, whereas regions with a high CpGo/e (approaching 1.0) are predicted to be unmethylated. This approach has been used to reliably predict methylation status across many taxonomic groups [17, 19, 22, 28].

A non-redundant *C.gigas* contig database, 'GigasDatabase' version 6 [29] was utilized for this analysis. To ensure only CpG (and not GpC) dinucleotides were being evaluated, analysis was limited to annotated sequences. The probability density function of the CpGo/e for 12,210 annotated *C.gigas* expressed sequence tag (EST) contigs is illustrated in Figure 2. We find that the data fit a bimodal mixture model (blue curve) significantly better than a unimodal distribution. The red curves represent the scaled, normal mixture components, which have means of 0.40 (±0.12 SD) and 0.7 (±0.2 SD) respectively (Figure 2). A majority of the contigs have a CpGo/e less than 1.0.

The ratio of observed to expected GpC dinucleotide frequencies (GpCo/e) was calculated in order to be assured that the bimodal distribution of CpGo/e was not biased toward G+C content of specific genes as there are no known mechanisms for preferential depletion of the GpC dinucleotide. As predicted, the ratio of observed to expected GpC's approaches 1.0 following a unimodal Gaussian distribution (Figure 2 inset). In addition, there is a significant negative correlation between CpGo/e and TpGo/e ($p=0.00$) indicating that the depletion of CpG dinucleotides is associated with the conversion of methylated CpG sites to TpG dinucleotides.

In order to determine any functional difference that might exist among those genes with lower than expected CpGo/e ratios, data were analyzed in the context of each gene's biological process GO Slim term (Figure 3). Several biological processes have CpGo/e ratios that are significantly different from each other (see Table 2). Specifically, genes with lower CpGo/e ratios (predicted to be hyper-methylated) were associated with DNA

metabolism, RNA metabolism, and cell cycle and proliferation. Biological processes with higher CpGo/e ratios (predicted to be hypo-methylated) include cell adhesion, cell-to-cell signalling and signal transduction. This analysis confirms that the normal mixture components described previously in Figure 2 are enriched with genes from particular functional categories.

Discussion

Results of methylation specific PCR and bisulfite sequencing PCR indicate that the Pacific oyster (*Crassostrea gigas*) genome is methylated. Further evidence supporting the presence and importance of methylation in *C.gigas* is the identification of genes that encode DNA methyltransferases (DNMT), the family of proteins responsible for the enzymatic conversion of cytosines to 5-methylcytosine. Animals that lack DNA methylation such as *C.elegans* also lack essentials DNMTs, while invertebrates with an intermediate level of DNA methylation such as honey bees, sea urchins and urochordates have the full set of DNMT genes [30]. Sequences with high homology to DNMT3 (responsible for *de novo* methylation), DNMT1 (associated with maintenance methylation), and methyl-CpG-binding domain protein 2 (mediates the effects of DNA methylation) are present in a publically available *C.gigas* contig database, GigasDatabase version 6 [29]. These annotated sequences can be found in GigasDatabase with accession numbers CU684371.p.cg.6 (e-value 1e-61), CU994437.p.cg.6 (e-value 2e-26), and AM861084.p.cg.6 (e-value 1e-11), respectively. While a DNMT2 homolog has not been identified, it may not be required for DNA methylation in *C.gigas* as it functions primarily as a tRNA methyltransferase and shows only weak DNA methyltransferase activity *in*

vitro [31]. DNMTs are an evolutionary conserved group of proteins, but show structural diversity both within and among taxa [32]. The evolutionary diversity of DNA methylation within and among phylogenetic groups provides justification for further evaluation of the functions of this epigenetic mark.

The presence of intragenic methylation in *C.gigas* is similar to that of other invertebrates that primarily exhibit intragenic DNA methylation patterns [33 and 17], the role of which has been largely unexplored. Studies of DNA methylation in mammals have generally focused on promoter regions, where hyper-methylation of promoters inhibits initiation of transcription [2]. In contrast, invertebrate genomes do not show differentially methylated gene promoters [22]. One of the long-standing hypotheses is that intragenic DNA methylation prevents inappropriate initiation of transcription outside of promoter regions [34], however new studies have begun to investigate a more active role for intragenic DNA methylation, namely in regulation of expression. For example, exonic DNA methylation has been shown to regulate transcription of the *phytochrome A* gene in *Arabidopsis thaliana* [35]. In humans, investigation of intragenic CpG islands (≥200 bp regions with G+C content of at least 50% and CpGo/e close to expected) has revealed that CpG islands in terminal exons may regulate transcription of non-coding RNAs [36]. Here, using BSP, we observed methylation variability in two CpG sites that may indicate cell-specific methylation. The function of intragenic DNA methylation in *C.gigas* cannot be conclusively determined from this study, but results of studies in other organisms suggest that it could be involved in either repression of transcription outside of transcription start sites, and/or regulation of expression.

Within the transcriptome of the Pacific oyster, a significant difference in methylation pattern was observed across gene families. A majority of *C.gigas* genes analyzed were depleted in CpG dinucleotides (i.e. CpGo/e <1.0) and show a significantly bimodal distribution, suggesting that DNA methylation is a common feature of the *C.gigas* transcriptome, and that certain groups of genes have significantly different levels of methylation. The bimodal distribution of CpGo/e is similar to the pattern observed in the honey bee *A.mellifera*, where authors reported a hyper-methylated fraction that was enriched in genes involved with general metabolic functions and a hypo-methylated fraction enriched with genes that are associated with caste-specific functions [19]. Similarly when *C.gigas* transcripts were clustered according to their functional annotations using GO Slim terms, we see that the two distributions are comprised of functionally distinct classes of genes with varying regulatory requirements. Specifically, genes predicted to be hyper-methylated are associated with housekeeping functions and those predicted to be hypo-methylated are associated with general immune functions. Hyper-methylation of intragenic regions of housekeeping genes is consistent between *C.gigas* and honey bees [19], but stands in contrast to observations in vertebrates, where distinct hypo-methylation of housekeeping gene promoters is associated with global expression [37]. Constitutive DNA methylation in housekeeping genes in *C.gigas* could be important for repressing transcription outside of promoter regions as previously discussed. It has also been proposed that hyper-methylation of housekeeping genes in *A. mellifera* indicates epigenetic control of gene activity in housekeeping genes [23]. Further experiments will be required to determine whether hyper-methylation of housekeeping

genes plays a passive role in preventing inappropriate transcription or a more active role in maintaining expression in *C.gigas*.

Highest CpGo/e ratios were observed in genes involved in the oyster's innate immune response, including categories of cell adhesion, cell-cell signaling, and signal transduction. Our experimental data using MSP supports the predicted hypo-methylation of this class of genes as only 1 of the 5 immune related genes were methylated. Our results do not indicate that DNA methylation is entirely absent from genes in the hypo-methylated group as CpG depletion is still observed (CpGo/e 0.7) which stands in contrast to the hypo-methylated genes in *A.mellifera* (CpGo/e >1.0). One explanation as to why it would be advantageous for this class to of genes to be hypo-methylated is that it allows for greater epigenetic flexibility and higher regulatory control. Oysters have been shown to have high phenotypic plasticity in response to environmental changes and stress [38 and 39] and it could be postulated that an epigenetic mark, such as DNA methylation, could provide this level of control. DNA methylation has been generally considered to be a less dynamic epigenetic mark, however, it has been reported in plants that DNA methylation levels are involved in regulating gene expression in response to stress and show active methylation and demethylation in response to various stressors [40, 41, 42]. It has been hypothesized from these studies that DNA methylation is a possible mechanism to impart protection against local stresses in future generations [43]. The identification of genes involved in demethylation in *C.gigas* would be an important step toward uncovering the nature of these epigenetic marks.

DNA methylation patterns have been shown to be heritable in mammalian taxa [44], and changes in DNA methylation patterns can persist for multiple generations [45]. Little work has been done to investigate heritability of DNA methylation in invertebrates, although a recent study of the crustacean, *Daphnia magna*, has shown transgenerational heritability of DNA methylation patterns after exposures to 5-azacytidine [46]. If DNA methylation does play a role in regulation of transcription in *C.gigas* it may provide a mechanism not only for regulating responses to stress, but also for adapting to local stressors through heritability of DNA methylation patterns. Investigating the potential of epigenetic control in mechanisms of local adaptation may prove useful in understanding of impacts of anthropogenic inputs in aquatic ecosystems and populations. Likewise, it is possible that epigenetic mechanisms may provide explanation for other phenomena associated with heritability such as inbreeding depression and hybrid vigour.

Elucidating functional significance of DNA methylation in aquatic invertebrates may change the way we study impacts of environmental change in aquatic organisms. A range of factors such as diet [47 and 48], xenobiotic chemicals [49], and endocrine disruptors [11] have been shown to disrupt DNA methylation patterns. These epigenetic disruptions are increasingly being associated with disease susceptibility, which in some cases can be passed on for multiple generations [50]. Although these investigations have been performed almost exclusively in mammalian species, recent studies have reported a dose dependent relationship between concentration of mercury and cadmium and levels and total DNA methylation in *D. magna* [46 and 51]. Understanding which environmental factors can affect DNA methylation and elucidating the functional significance of DNA

methylation in these important bioindicator species will be major steps toward clarifying the complex interactions between the environment, gene expression, and organismal responses.

Conclusions

The Pacific oyster genome displays methylation. *In silico* analysis reveals intragenic regions are targeted for methylation consistent with reports of methylation in other invertebrate species. Results of this investigation suggest that DNA methylation has regulatory functions in *Crassostrea gigas*, particularly in gene families involved in stress and environmental response. Experiments are underway in our lab to investigate relationships between the environment, DNA methylation, and control of gene expression to better characterize this process. In-depth analysis of methylation patterns in *Crassostrea gigas*, will help to advance the field of evolutionary epigenetics and will serve to illuminate functions of DNA methylation in invertebrates.

Acknowledgements

We thank Dustin Lennon for statistical support. This research was partially supported by the School of Aquatic and Fishery Sciences at the University of Washington.

Works Cited

1. Jablonka E, Lamb M: The changing concept of epigenetics. *Ann N Y Acad Sci* 2002, 981:82-96.

2. Boyes J, Bird A: Repression of genes by DNA methylation depends on CpG density and promoter strength: Evidence for involvement of a methyl-CpG binding protein. *EMBO J* 1992, 11:327–333.

3. Kass SU, Pruss D, Wolffe AP: How does DNA methylation repress transcription? *Trends Genet* 1997, 13:444–449.

4. Hsieh CL: Dependence of transcriptional repression on CpG methylation density. *Mol Cell Biol* 1994*,* 14:5487–5494.

5. Maloisel L, Rossignol JL: Suppression of crossing-over by DNA methylation in *Ascobolus Genes Dev* 1998, 12:1381-1389.

6. Bell AC, Felsenfeld G: Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* 2000, 405:482-5.

7. Csankovszki G, Nagy A, Jaenisch R: Synergism of Xist RNA, DNA Methylation, and Histone Hypoacetylation in Maintaining X Chromosome Inactivation. *J Cell Biol* 2001, 153:773-784.

8. Okano M, Bell DW, Haber DA, Li E: DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* 1999, 99:247–257.

9. Li E, Bestor TH, Jaenisch R: Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* 1992, 69:915–926.

10. Jones PA, Baylin SB: The epigenomics of cancer. *Cell* 2007, 128:683–92.

11. Anway MD, Skinner MK: Epigenetic transgenerational actions of endocrine disruptors. *Endocrinology* 2006, 147:43-49

12. Bird AP, Taggart MH: Variable patterns of total DNA and rDNA methylation in animals. *Nucleic Acid Res* 1980 8:1485-1497.

13. Gowher H, Leismann O, Jeltsch A: DNA of Drosophila melanogaster contains 5-methylcytosine. *EMBO J.* 2000, 19:6918–6923.

14. Simpson VJ, Johnson TE, Hammen RF. *Caenorhabditis elegans* DNA does not contain 5-methylcytosine at any time during development or aging. *Nucleic Acids Res* 1986, 14:6711–6719.

15. Bird AP, Taggart MH, Smith BA: Methylated and unmethylated DNA compartments in the sea urchin genome. *Cell* 1979, 17:889–901.

16. Simmen MW, Bird, AP: Sequence analysis of transposable elements in the sea squirt, *Ciona intestinalis*. *Mol Biol Evol* 2000, 17:1685–1693.

17. Suzuki MM, Kerr ARW, De Sousa D, Bird A: CpG methylation is targeted to transcription units in an invertebrate genome. *Genome Res* 2007, 17:625–631.

18. Wang Y, Jorda M, Jones PL, Maleszka R, Ling X, Robertson HM, Mizzen CA, Peinado MA and Robinson GE: Functional CpG methylation system in a social insect. *Science* 2006, 314:645–647.

19. Elango N, Hunt BG, Goodisman MAD, Yi S: DNA methylation is widespread and associated with differential gene expression in castes of the honeybee, *Apis mellifera*. *Proc Natl Acad Sci USA* 2009, 106:11206-11211.

20. Hung MS, Karthikeyan N, Huang B, Koo HC, Kiger J, Shen C J: Drosophila proteins related to vertebrate DNA (5-cytosine) methyltransferases. *Proc Natl Acad Sci USA* 1999, 96:11940 -11945.

21. Lyko F, Ramsahoye BH, Jaenisch R: DNA methylation in Drosophila melanogaster. *Nature* 2000*,* 408:538-540.

22. Elango N, Yi SV: DNA methylation and structural and functional bimodality of vertebrate promoters. *Mol Biol and Evol* 2008, 25:1602-1608.

23. Foret S, Kucharski R, Pittelkow Y, Lockett GA, Maleszka R: Epigenetic regulation of the honey bee transcriptome: unravelling the nature of methylated genes. *BMC Genomics* 2009, 10:472.

24. Petrović V, Pérez-García C, Pasantes JJ, Šatović E, Prats E, Plohl M: A GC-rich satellite DNA and karyology of the bivalve mollusk *Donax trunculus:* a dominance of GC-rich heterochromatin. *Cytogenet Genome Res* 2009, 124:63-71.

25. Markert BA, Breure AM, Zechmeister HG: Molluscs as Bioindicators. In *Bioindicators and Biomonitors*. Edited by: B.A Markert, A.M Breure and H.G Zechmeister. Amsterdam: Elsevier Science; 2003, 577–634.

26. Coulondre C, Miller JH, Farabaugh PJ, Gilbert W: Molecular basis of base substitution hotspots in *Escherichia coli*. *Nature* 1978, 274:775–780.

27. Schorderet DF, Gartler SM: Analysis of CpG suppression in methylated and nonmethylated species. In: *Proc Natl Acad Sci USA* 1992, 89:957–961.

28. Shimizu T.S., Takahashi K., Tomita M: CpG distribution patterns in methylated and non-methylated species *Gene* 1997, 205:103-107.

29. Fleury E, Huvet A, Lelong C, de Lorgeril J, Boulo V , Gueguen Y, Bachère E, Tanguy A, Moraga D, Fabioux C, Lindeque P, Shaw J, Reinhardt R, Prunet R, Davey G, Lapègue S, Sauvage C, Corporeau C, Moal J, Gavory F, Wincker P, Moreews F, Klopp C, Mathieu M, Boudry P, Favrel B: Generation and analysis of a 29,745 unique Expressed Sequence Tags from the Pacific oyster (Crassostrea gigas) assembled into a publicly accessible database: the GigasDatabase. *BMC Genomics* 2009, 10:341.

30. Cañestro C,Yokoi H,Postlethwait JH: Evolutionary developmental biology and genomics. *Nat Rev Genet* 2007, 8:932-942.

31. Schaefer M, Lyko F: Solving the Dnmt2 enigma. *Chromosoma* 2010, 119: 35-40.

32. Colot V, Rossignol J: Eukaryotic DNA methylation as an evolutionary device. *BioEssays* 1999, 21: 402-411.

33. Tweedie S, Charlton J, Clark V, Bird A: Methylation of genomes and genes at the invertebrate-vertebrate boundary. *Mol Cell Biol* 1997 17:1469–1475.

34. Bird AP: Gene number, noise reduction and biological complexity. *Trends Genet* 1995, 11:94-100.

35. Chawla R, Nicholson SJ, Folta KM, Srivastav V: Transgene-induced silencing of *Arabidopsis* phytochrome A gene via exonic methylation. *Plant J* 2007, 52:1105–1118.

36. Medvedeva Y, Fridman MV, Oparina NJ, Malko DB, Ermakova EO, Kulakovskiy IV, Heinzel A, Makeev V: Intergenic, gene terminal, and intragenic CpG islands in the human genome. *BMC Genomics* 2010, 11: 48.

37. Weber M, SchubelerD: Genomic patterns of DNA methylation: targets and function of an epigenetic mark. *Current Opin Cell Biol* 2007, 19:273-280.

38. Hamdoun AM, Cheney DP, Cherr GN: Phenotypic plasticity of HSP70 and HSP70 gene expression in the Pacific oyster (*Crassostrea gigas*): Implications for thermal timits and induction of thermal tolerance. *Biol Bull* 2003, 205:160-169.

39. Honkoop PJC, Bayne BL, Drent J: Flexibility of size of gills and palps in the Sydney rock oyster *Saccostrea glomerata* (Gould, 1850) and the Pacific oyster *Crassostrea gigas* (Thunberg, 1793). *J. Exp. Mar. Biol. Ecol.* 2003, 282:113–133.

40. Kovarik A, Koukalova B, Bezdek M, Opatrn Z: Hypermethylation of tobacco heterochromatic loci in response to osmotic stress. *Theor Appl Genet* 1997, 95: 301–306.

41. Steward N, Ito M, Yamaguchi Y, Koizumi N, Sano H: Periodic DNA Methylation in Maize Nucleosomes and Demethylation by Environmental Stress. *J Biol Chem* 2002, 277: 37741-37746.

42. Choi CS, Sano H: Abiotic-stress induces demethylation and transcriptional activation of a gene encoding a glycerophosphodiesterase-like protein in tobacco plants. *Mol Genet Genome* 2007, 277:589–600.

43. Chinnusamy V, Zhu J: Epigenetic regulation of stress responses in plants. *Current Opin Plant Biol* 2009, 12:133-139.

44. Anway MD, Cupp AS, Uzumcu M, Skinner MK: Epigenetic transgenerational actions of endocrine disruptors and male fertility. *Science* 2005, 308:1466–1469.

45. Cropley JE, Suter CM, Beckman KB, Martin DIK: Germ-line epigenetic modification of the murine A(vy) allele by nutritional supplementation. *Proc Natl Acad Sci USA* 2006, 103:17308–1731.

46. Vandegehuchte MB, Lemière F, Vanhaecke L, Berghe VW, Janssen CR: Direct and transgenerational impact on Daphnia magna of chemicals with a known effect on DNA methylation. *Comp Biochem Physiol C Toxicol Pharmacol*, in press.

47. Wilson MJ, Shivapurkar N, Poirier LA: Hypomethylation of hepatic nuclear DNA in rats fed with a carcinogenic methyl-deficient diet. *Biochem J* 1984 , 218: 987–990.

48. Dolinoy DC, Weidman JR, Waterland RA, Jirtle RL: Maternal genistein alters coat color and protects A^{vy} mouse offspring from obesity by modifying the fetal epigenome. *Environ Health Perspect* 2006, 114: 567-572.

49. Sutherland JE, Costa M: Epigenetics and the environment. *Ann NY Acad Sci* 2003, 983:151–160.

50. Jirtle RL, Skinner MK: Environmental epigenomics and disease susceptibility. *Nat Rev Genet* 2007, 8: 253-262.

51. Vandegehuchte MB, Vanholme B, Haegeman A, Gheysen G, Janssen CR: Occurrence of DNA methylation in Daphnia magna and influence of multigeneration Cd exposure. *Environ Int* 2009, 35:700-706.

52. Roberts S, Goetz G, White S, Goetz F: Analysis of genes isolated from plated hemocytes of the Pacific oyster, *Crassostrea gigas*. *Mar Biotechnol* 2009, 11:24–44.

53. Li LC, Dahiya R. MethPrimer: designing primers for methylation PCRs. *Bioinformatics* 2002, 18:1427-31.

54. Altschul SF, Gish W, Miller W, Myers E.W, Lipman DJ: Basic local alignment search tool *J Mol Biol* 1990, 215:403-410.

55. Young DS, Hunter DR, Elmore RT, Xuan F, Hettmansperger TP, Thomas H: The mixtools package Vvrsion 0.2.0: Tools for mixture models. R Package Version 0.2.0. 2007: [http://www.r-project.org].

56. R Development Core Team. R: A language and environment for statistical computing: [http://www.R-project.org].

57. Mouse Genome Database (MGD) at the Mouse Genome Informatics [http://www.informatics.jax.org].

Figures

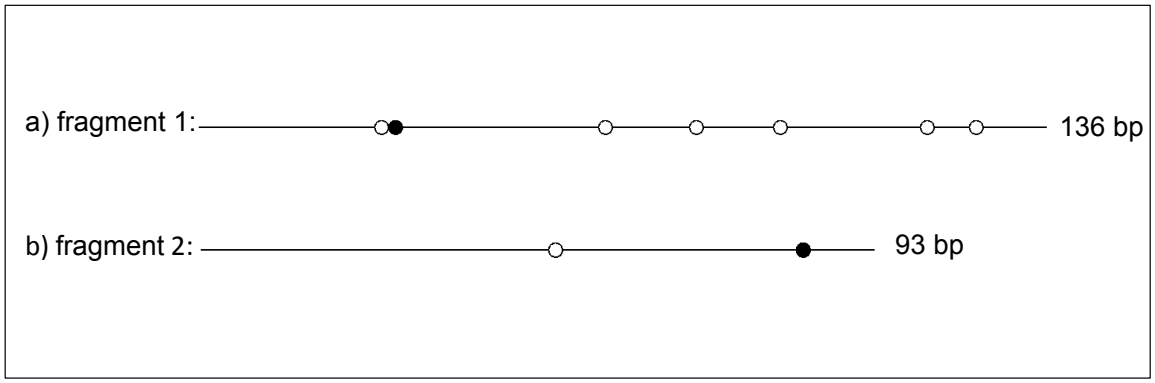


Figure 1. Methylation status of two C.gigas DNA fragments by bisulfite sequencing. Methylation status of a 136 bp (a) and 93 bp (b) fragment of *C.gigas* DNA as determined by bisulfite sequencing. Solid and open circles represent methylated and non-methylated CpG dinucleotides, respectively. One of four clones was determined to be methylated at the CpG indicated by the solid circle in (a) and 2 of 4 clones were determined to be methylated at the CpG dinucleotide indicated in (b).

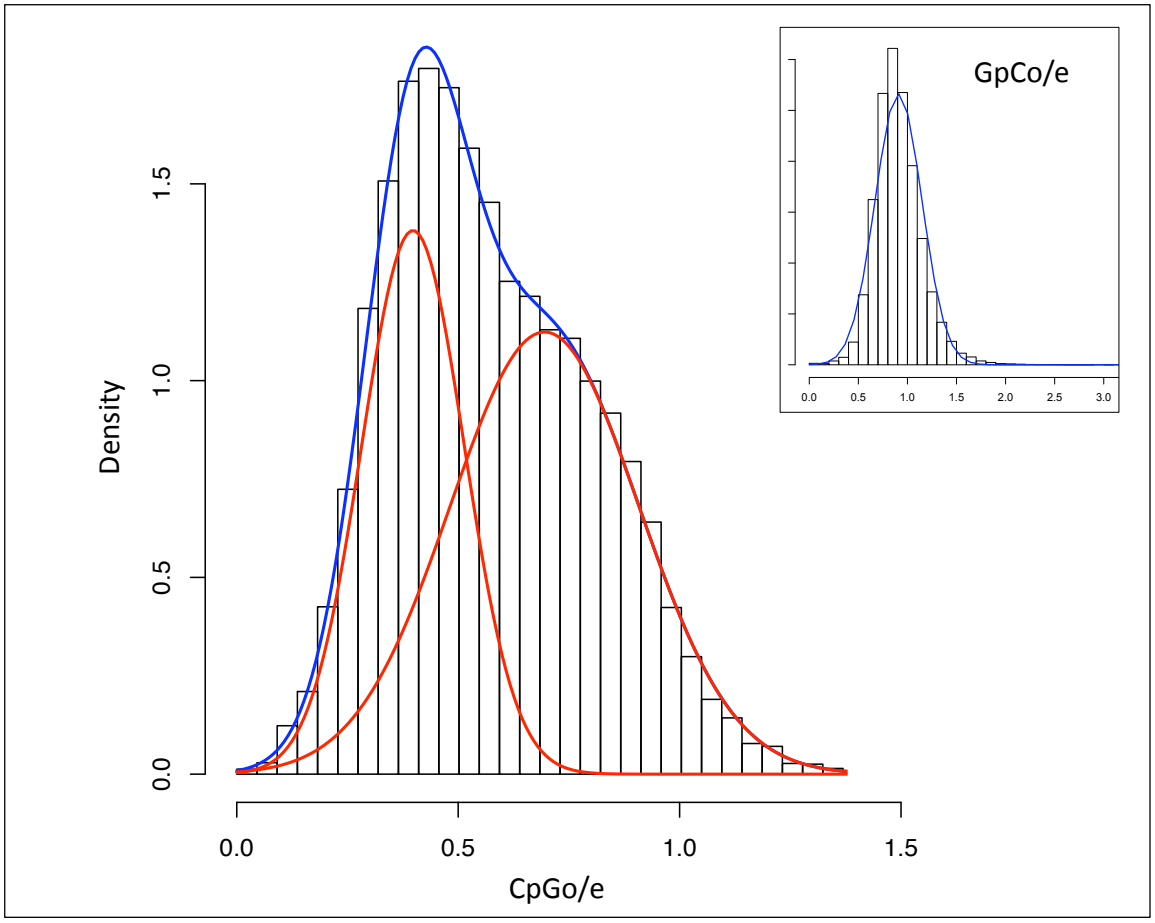


Figure 2. **Distribution of predicted methylation status of 12,210 annotated C.gigas transcripts measured computationally by CpGo/e ratio.**
Probability density function of CpGo/e for 12,210 C.gigas contigs. Blue curve is fitted mixture model; red curves are scaled, normal mixture components with means of 0.40 and 0.70 respectively. For contrast, a control dinucleotide (GpCo/e) is also shown with the blue curve representing a normal, unimodal distribution (inset).

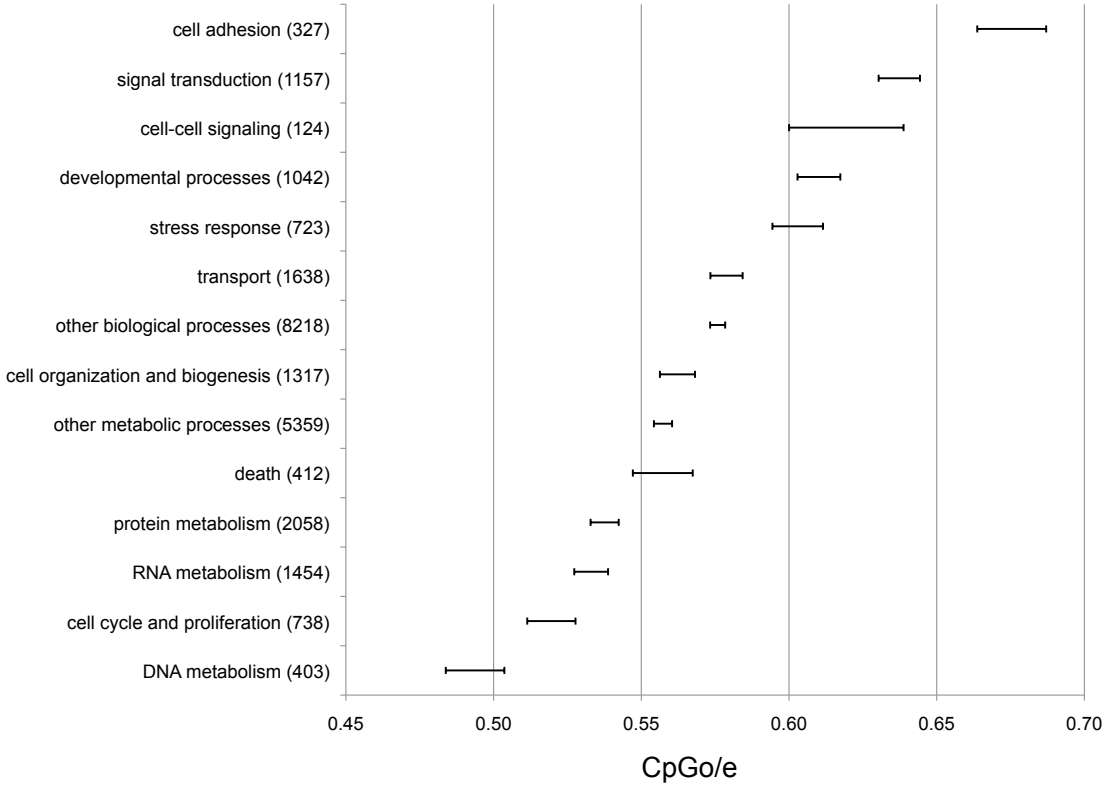


Figure 3. Differential methylation between categories of genes involved in discrete biological processes as measured by CpGo/e.
Mean CpGo/e for 10,699 *C.gigas* contigs categorized by GO Slim term. Bars represent mean ± 1 standard error. The number of contigs in each category is listed in parenthesis.

Tables

Table 1. Results of Methylation Specific PCR analysis for five C.gigas genes
Results of methylation status of five genes associated with immune response by MSP. PCR was carried out on undigested, HpaII digested, and MspI digested DNA. Results of methylation status of five genes associated with immune response by MSP. PCR was carried out on undigested, HpaII digested, and MspI digested DNA.

Accession #	Best blast hit [Organism]	Undi-gested	HpaII	MspI	Number of restriction sites	CpG o/e
EW778441	heat shock protein 70 [*Crassostrea gigas*]	+	+	–	2	0.57
EW777519	heat shock protein 25 [*Danio rerio*]	+	–	–	3	0.81
EW778166	cytochrome P450 [*Haliotis diversicolor*]	+	–	–	1	0.85
EW778608	macrophage expressed protein 1-like protein [*Crassostrea gigas*]	+	–	–	6	1.08
EW778905	14-3-3 protein gamma (Protein kinase C inhibitor protein 1 [*Bos taurus*]	+	–	–	2	0.92

Table 2. Matrix of p-values for comparisons between GO Slim categories based on CpGo/e
This table contains the p-values of each comparison. Significant differences ($p < 0.05$) are highlighted.

	cell adhesion	signal transduction	cell-cell signaling	developmental processes	stress response	transport	other biological processes	cell organization and biogenesis	other metabolic processes	death	protein metabolism	RNA metabolism	cell cycle and proliferation	DNA metabolism
cell adhesion		4.11E-2	5.13E-1	1.36E-5	2.81E-6	6.89E-12	6.43E-12	6.43E-12	6.43E-12	7.79E-12	6.43E-12	6.43E-12	6.43E-12	6.43E-12
signal transduction			1.00	0.32E-1	6.48E-2	1.15E-9	6.43E-12	6.43E-12	6.43E-12	5.18E-8	6.43E-12	6.43E-12	6.43E-12	6.43E-12
cell-cell signaling				1.00	9.94E-1	4.63E-1	3.15E-1	7.56E-2	2.79E-2	9.55E-2	7.55E-4	3.82E-4	4.23E-5	2.56E-7
developmental processes					1.00	2.49E-2	2.32E-4	1.34E-5	2.71E-10	3.74E-3	6.43E-12	6.43E-12	6.43E-12	6.43E-12
stress response						5.12E-1	1.21E-1	6.62E-3	3.76E-5	6.83E-2	1.79E-9	1.08E-9	1.89E-10	6.74E12
transport							1.00	7.24E-1	4.73E-2	9.04E-1	2.12E-6	1.46E-6	2.73E-7	4.38E-10
other biological processes								6.74E-1	2.50E-4	9.38E-1	2.96E-10	1.98E-9	7.19E-9	3.72E-11
cell organization and biogenesis									1.00	1.00	1.14E-1	5.22E-2	3.94E-3	5.65E-6
other metabolic processes										1.00	3.55E-2	1.76E-2	1.58E-3	1.92E-6
death											9.31E-1	8.07E-1	2.65E-1	2.96E-3
protein metabolism												1.00	8.67E-1	1.76E-2
RNA metabolism													9.90E-1	7.62E-2
cell cycle and proliferation														7.91E-1
DNA metabolism														

Table 3. Primer Sequences: This table contains primer sequences used for methylation sensitive PCR and bisulfite sequencing analysis.

Methylation Sensitive PCR:

Accession # [NCBI]	Best BLAST hit [Organism]	5' primer sequence	3'primer sequence
EW778441	heat shock protein 70 [*C. gigas*]	AGGGTATCGATTTCTACACAAG	GTTTCTCTTGATGAGATTGGTC
EW777519	heat shock protein 25 [*Danio rerio*]	AAATGAGCAAAATATTTAACGAGGA	TGGGATGGTAAGGATCAAGG
EW778166	cytochrome P450 [*Haliotis diversicolor*]	ATATTGGAGCCCTCGTTGTG	TGAGCGCAGAGAAACTTCAA
EW778608	macrophage expressed protein 1-like protein [*C. gigas*]	CGGAACCGAAGTAGATGGAA	ACGCTCATAAACGAGGCACT
EW778905	14-3-3 protein gamma (Protein kinase C inhibitor protein 1) [*Bos taurus*]	ATAGAGCGGAAGCCGTGATA	TAACTCGCAAGCAGTGTTGG

Bisulfite Sequencing PCR:

Accession # [GigasDatabase[1]]	Best BLAST hit [Organism]	5' primer sequence	3'primer sequence
AM858698.p.cg.6	neuromedin-u receptor [*H. sapiens*]	AGTTTTATATTGATTTTTTGGAGAG	AAATTCTTCTCAAAATACATTCTTC
AM860932.p.cg.6	bromodomain adj. to zinc finger protein [*H. sapiens*]	TTGTAAATTGATAAATGAAATATTT	AAAATTCTTAAAAACCTTCTCCTC

[1] http://public-contigbrowser.sigenae.org:9090/Crassostrea_gigas/index.html

Chapter III: Predominant intragenic methylation is associated with gene expression characteristics in a bivalve mollusc

A version of this chapter was published in PeerJ as: Gavery M and Roberts SB. 2013. Predominant intragenic methylation is associated with gene expression characteristics in a bivalve mollusk. PeerJ 1:e215. doi:10.7717/peerj.215

Abstract

Characterization of DNA methylation patterns in the Pacific oyster, Crassostrea gigas, indicates that this epigenetic mechanism plays an important functional role in gene regulation and may be involved in the regulation of developmental processes and environmental responses. However, previous studies have been limited to in silico analyses or characterization of DNA methylation at the single gene level. Here, we have employed a genome-wide approach to gain insight into how DNA methylation supports the regulation of the genome in C. gigas. Using a combination of methylation enrichment and high-throughput bisulfite sequencing, we have been able to map methylation at over 2.5 million individual CpG loci. This is the first high-resolution methylome generated for a molluscan species. Results indicate that methylation varies spatially across the genome with a majority of the methylated sites mapping to intra-genic regions. The bisulfite sequencing data was combined with RNA-seq data to examine genome-wide relationships between gene body methylation and gene expression, where it was shown that methylated genes are associated with high transcript abundance and low variation in expression between tissue types. The combined data suggest DNA methylation plays a complex role in regulating genome activity in bivalves.

Introduction

Epigenetic marks such as DNA methylation are important for genome regulation (Bell & Felsenfeld 1992; Li et al., 1992; Hsieh 1994). DNA methylation has been well-studied in mammals and plants where it has been shown to play important roles in temporal and spatial regulation of gene expression (Okano et al.,1999, Zhang et al., 2006), reduction of transcriptional noise (Bird 1995), and genome stabilization (Wolffe & Matzke 1999). However, the distribution and context of DNA methylation varies greatly among phylogenetic groups indicating that these functional roles may not be conserved (Colot & Rossignol 1999).

In contrast to the heavily methylated vertebrate genomes, invertebrate genomes exhibit a relatively low level of methylation that is primarily confined to gene bodies (Zemach et al., 2010). It is perhaps because of these differences that DNA methylation has remained largely understudied in invertebrates. Recently, however, there has been a renewed interest in invertebrate DNA methylation patterns as it is now being recognized that they are exceptional models to study function and evolutionary roles of gene body methylation. Furthermore, it has been shown that DNA methylation mediates phenotypes in response to environmental cues in the invertebrate *Apis mellifera* (Kucharski et al., 2008, Lyko et al., 2010), indicating an important role integrating environmental signals. Thus, understanding both the distribution and function of DNA methylation in diverse invertebrate lineages will contribute significantly to our understanding of the evolution of genome regulation and environmental physiology.

The focus of the work presented here is to explore the role of DNA methylation in bivalve molluscs. The presence of DNA methylation has been confirmed in several bivalve species (Wang et al., 2008, Petrovic et al., 2009, Gavery & Roberts 2010). A majority of the research on DNA methylation in molluscs has focused on the Pacific oyster (*Crassostrea gigas*),

an economically and ecologically important species. Previous studies in the Pacific oyster identified a relationship between gene function and methylation pattern. Specifically, it was shown that genes with housekeeping functions are more methylated than genes involved in inducible functions (i.e. genes involved in response to environment, embryonic development or tissue-specific functions) (Gavery & Roberts 2010, Roberts & Gavery 2012). More recently, Riviere et al. (2013) determined that DNA methylation plays a critical role in development as indicated by differential methylation patterns throughout embryogenesis. This was further supported by their observation that 5-aza-cytidine, a potent demethylating agent, significantly disrupts embryonic development (Riviere et al., 2013).

These recent studies on DNA methylation in oysters provides important foundational information on DNA methylation in bivalves. However, previous studies were not able to provide fine scale resolution of DNA methylation patterns, nor examine the relationship with gene expression at the genome-wide level. Here, we provide the first high resolution methylome of a mollusc and examine this in relationship to gene expression data to get a better understanding of the role of DNA methylation in invertebrates.

Methods

Bisulfite treated DNA (BS-Seq) Analysis

The cohort of adult oysters used in this study was from Samish Bay, WA, USA. Briefly, genomic DNA was isolated using DNAzol (Molecular Research Center) from gill tissue of 8 oysters, pooled, and methylation enrichment performed using the MethylMiner Kit (Invitrogen) following the manufacturer's instructions. Specifically, pooled DNA was sheared by sonication on a Covaris S2 (Covaris) (parameters: 10 cycles at 60 seconds each, duty cycle of 10%,

intensity of 5, 100 cycles/burst). Approximately 13ug of sheared DNA was used as input DNA and incubated with MBD-Biotin Protein coupled to M-280 Streptavidin Dynabeads following the manufacturer's instructions (MethylMiner (Invitrogen)). Enriched, methylated DNA was eluted from the bead complex with 1M NaCl and purified by ethanol precipitation. This enriched fraction represented approximately 15% of the total DNA recovered from the enrichment procedure. The DNA library was prepared using the Illumina Tru-Seq system with methylated TruSeq adapters (mean fragment size of library: 350 bp). Bisulfite treatment was then performed using the EpiTect Bisulfite Kit (Qiagen) following manufacturer instructions. Library preparation and sequencing was performed on the Illumina HiSeq 2000 platform at the University of Washington high throughput sequencing facility (Seattle, WA). High-throughput reads (36bp single end) were mapped back to the oyster genome (Fang et al., 2012) using BSMAP software version 2.73 (Xi & Li 2009). Methylation ratios (i.e. number of unconverted cytosines/the number of converted and unconverted cytosines at each locus) were extracted from BSMAP output (SAM) using a Python script (*methratio.py)* that is distributed with the BSMAP package. Only cytosines in a CpG context with sufficient sequencing depth (defined here as greater than or equal to 5x coverage) were retained for further analysis.

DNA Methylation Landscape of Genomic Features

In order to examine the relationship of DNA methylation and other genomic features, data from BSMAP (i.e. *methratio*) was converted to genomic feature tracks (i.e. generic feature format [GFF] files). Conversion was done using SQLShare (Howe et al., 2011), with the files and corresponding query language published (Gavery & Roberts 2013).

The distribution of methylated CpGs with respect to specific genomic features was determined using BEDtools (i.e. *intersectBED*) (Quinlan & Hall 2010). For this analysis, a CpG locus was considered methylated if at least half of the reads remained unconverted after bisulfite treatment. Genomic features that were examined include: exons and introns (Fang et al., 2012), promoters (defined as 1kb upstream of open reading frames), and transposable elements. Putative transposable elements were identified using RepeatMasker (Smit et al., 1996-2010), based on protein similarities to the Transposable Element Protein Database. At the time of analysis the database contained 5411 predicted proteins. For comparative purposes, total CpG across the entire *C. gigas* genome was also examined. Locations of all CpGs were identified using the EMBOSS tool *fuzznuc* (Rice et al., 2000), and the proportion of total CpG in each of the genomic features listed above was determined using *intersectBED*. A Chi-squared test was performed to determine if the distribution of methylated CpG was different from what would be expected by a random distribution of the total CpG in the genome (p-value <0.05 was considered significant).

Average methylation ratios were determined for full length genes and also the cumulative exons and cumulative introns comprising a gene. Average methylation was determined by the number of methylated cytosines divided by the total number of CpG per region. Correlations between the methylation status of exons and introns of individual genes were performed using Pearson's correlation coefficient in SPSS (SPSS Inc.).

The relationship between predicted methylation status, using the CpG observed to expected ratio (CpGo/e), and the average methylation ratio for each gene was examined to assess the effectiveness of the CpGo/e method for predicting methylation in bivalve species. For this analysis, the CpGo/e ratio was calculated for each gene using the method described in Gavery &

Roberts (2010). Correlation between CpGo/e and the methylation ratio was performed using Spearman rank correlation in SPSS (SPSS Inc.).

Gene Expression Analysis

RNA was isolated from gill tissue of the same 8 individuals used for DNA isolation and bisulfite sequencing using Tri-Reagent (Molecular Research Center). RNA was pooled in equal quantities and enriched for mRNA using Sera-Mag oligo dT beads (Thermo Scientific). First strand synthesis was performed using SuperScript III (Invitrogen) and the second strand of cDNA was synthesized using dUTP instead of dTTP, making the library strand-specific. A shotgun library was constructed from double stranded cDNA for paired end sequencing by end-polishing, A-tailing and ligation of sequencing adaptors. Sequencing was performed on the Illumina HiSeq 2000 platform at the Northwest Genomics Center at the University of Washington (Seattle, WA). High-throughput reads (50bp paired end) were mapped back to the oyster genome (Fang et al., 2012) using CLC Genomics Workbench version 6.5 (CLC Bio). Initially, sequences were trimmed based on quality scores of 0.05 (Phred, Ewing and Green, 1998; Ewing et al., 1998), and the number of ambiguous nucleotides (>2 on ends). Sequences smaller than 20 bp were also removed. For RNA-Seq analysis, expression values were measured as RPKM (reads per kilobase of exon model per million mapped reads) (Mortazavi et al., 2008) with an unspecific match limit of 10 and maximum number of 2 mismatches.

The RPKM values were used to examine the relationship between gene expression and DNA methylation in gill tissue. All genes containing at least 1 CpG locus (n=28,105) were grouped into deciles according to transcriptomic representation in gill tissue (RPKM) and the average methylation ratios for each decile were compared. A one-way ANOVA followed by

Tukey's test for multiple comparisons was performed using R (R Core Development Team 2012) and a significance level of $p < 0.05$ was accepted.

A principal component analysis (PCA) was used as an exploratory tool to identify relationships between DNA methylation, gene expression profiles and gene attributes such as length. To explore variables related to gene expression, publicly available RNA-seq data from a variety of adult *C. gigas* tissues were leveraged from Zhang et al. (2012). Specifically, mean transcript abundance and variation in transcript abundance across tissues were calculated using RPKM values for 7 tissues adult tissues (digestive gland, female and male gonad, gill, anterior muscle, hemocytes and labial palps). Mean transcript abundance was calculated using the mean RPKM across all tissues for each gene. Variation in transcript abundance across tissues was calculated as the coefficient of variation (%CV) of the RPKM across all 7 tissues for each gene. Other gene attributes that were examined as they may associate with DNA methylation include gene length, number of exons per gene and number of CpG per gene. In summary, the following attributes were included as variables in the PCA performed in R (R Core Development Team 2012): average methylation ratio of the full length gene (as described above), gene length in base pairs (bp), number of exons, average transcriptomic representation (average RPKM across 7 adult tissues), coefficient of variation (%CV) of transcript abundance (RPKM) among tissues. All variables were log transformed, with the exception of the methylation ratio which was arcsine transformed prior to analysis. The significance of each principal component was calculated using Monte-Carlo randomization tests. Principal components were considered significant at $p \leq 0.05$. Correlation loadings of 0.6 were considered significant.

Results

DNA Methylation Mapping

Bisulfite treated DNA sequence reads (139,728,554 total reads; 36bp) are available in the NCBI Short Read Archive under the accession number SRX32737. A total of 120,734,949 reads (86%) mapped to the *C. gigas* genome. Fifty-six percent of the 164,873,219 cytosines in the *C. gigas* genome, had at minimum of 1x coverage. Of the 9,978,551 CpG dinucleotides in the genome, 2.6 million (26%) had ≥ 5x coverage. The distribution of methylation ratios found at CpG dinucleotides ranged between 0.0 - 1.0, but a majority of the loci were either heavily methylated or unmethylated. Specifically, 55% (1,453,752) were methylated (i.e. ≥ 0.50) and another 28% were unmethylated (i.e. = 0.0) (Figure 1). Genome feature track files (i.e. GFF) representing 1) all CpG dinucleotides and 2) methylated CpG dinucleotides (>50%) for this dataset were developed and are available (Gavery & Roberts 2013).

Methylation Landscape of Genomic Features

Methylated CpG dinucleotides, defined as having a methylation ratio of 0.5 or greater, were located predominantly in intragenic regions (exons and introns), but were also present in putative promoters (defined as 1 kb upstream of TSS), transposable elements and unannotated intergenic regions. The distribution of methylated CpGs across various genomic regions is significantly different than what would be expected if the methylation were distributed randomly throughout the CpG dinucleotides in the genome (X^2 = 513,194.1, df = 4, p <0.0001). Specifically, DNA methylation appears to be overrepresented in intragenic regions (64% of methylated CpG in combined exons and introns) when compared to the proportion of all CpG in the genome (38%) (Figure 2). When methylation was examined on a per gene basis a strong

positive correlation (R^2 : 0.86) was observed between exonic and intronic methylation within a gene. Additionally, a strong correlation was observed between the gene methylation measured via high-throughput bisulfite sequencing and the predicted methylation ratio based on the CpG observed to expected ratio (CpGo/e) (Spearman *rho*: -0.616, p-value: $<1x10^{-4}$).

Gene Expression & DNA methylation

After quality trimming, 45,751,574 RNA-seq reads mapped to the genome. Raw reads are available in the NCBI Short Read Archive under the accession number TBD.

The relationship between the proportion of methylation in a gene and overall transcript abundance in gill tissue was characterized (Figure 3). In general, transcription abundance increases significantly with increasing DNA methylation until the 40^{th} percentile after which it remains relatively stable until the 100^{th} percentile when methylation significantly decreases.

The first two principal components (PC) of the PCA of gene attributes were significant and explained 76.4% of the variation among the genes. This variation was being driven by multiple factors, including DNA methylation (Figure 4 and Table 1). The only variable that did not contribute significantly to the first two principal components was mean transcript abundance (correlation loading 0.2). The first PC, which explained 50.2% of the variation was loaded heavily by number of CpG dinucleotides, the length of the mRNA and the number of exons. The second PC, which explained 26.1% of the variation was loaded heavily by the %CV of gene expression among tissues and the methylation ratio. DNA methylation is negatively correlated with transcript variance between tissues (%CV) and relatively uncorrelated with attributes such as gene length.

Discussion

Here we have used methylation enriched high-throughput bisulfite sequencing in conjunction with genomic feature annotation and transcriptomic data in an attempt to gain a better understanding of the role of DNA methylation in oysters. This work not only provides new information on DNA methylation in invertebrates but also provides a framework for characterizing DNA methylation in other taxa.

The reduced representation approach was selected to obtain a higher coverage of methylated regions. In addition, since methylation was likely to occur in gene bodies (Zemach 2010), and because transcriptomic data was the primary genomic resource for *C. gigas* at the time of sequencing (the genome was released soon after), it was expected that the methylation enrichment would significantly limit the proportion of unmappable reads. Quantitative methylation data were obtained for both methylated CpG as well as unmethylated CpG that were either interspersed with or flanking these more heavily methylated regions. Therefore, methylation enriched bisulfite-sequencing was effective in generating a comprehensive invertebrate methylome.

One of our primary findings was the overall level of genome methylation in the oyster. Here we found that 15% of CpG dinucleotides (2% of total cytosines) are methylated in gill tissue. This degree of methylation is much lower than the global methylation patterns seen in mammals where 70-80% of CpGs are methylated (Bird & Taggart 1980), but still higher than what has been reported in other invertebrates. For instance, only 0.8% of the CpGs are methylated in the brain of *A. mellifera* (Foret et al., 2010) and between 0 – 8% of CpGs are methylated in the nematode, *Trichinella spiralis*, depending on the life stage (Gao et al., 2012). Although methylation in *C. gigas* is relatively high for an invertebrate, it is not outside

the range of what has been reported in other species by liquid chromatography-mass spectrometry analysis. For example, similar to the oyster, 2% of total cytosines are methylated in the mollusc *Biomphalaria glabrata* (Fneich et al.,2013). It should be noted that methylation in oysters does likely vary in both a temporal and possibly tissue specific manner, as clearly indicated by Riviere et al. (2013) by characterizing differences in total methylation during development. In addition, because the sample represents a pool of multiple individuals, it cannot be determined whether the variation in methylation at a particular locus represents hemimethylation or differential methylation between individuals. In general, the bimodal pattern observed (Figure 1) indicates that a CpG locus is either heavily methylated or unmethylated, but future work sequencing individual oysters would provide valuable information regarding individual epigenetic variation in oysters.

This work also provided the first direct evidence in oysters that DNA methylation is prominent in gene bodies (see Figure 2) and these data are well correlated with previous investigations using an *in silico* approach (i.e. CpGo/e) to predict methylation in *C. gigas* (Gavery & Roberts 2010).The predominance of gene body methylation is consistent with what has been described in other invertebrates (e.g. Suzuki et al., 2007, Zemach et al., 2010, Lyko et al., 2010) and there is increasing evidence that gene body methylation is the ancestral pattern (Lechner et al., 2013). The function of gene body methylation remains unclear, but studies indicate possible active roles in preventing spurious transcription (Bird et al.,1995, Huh et al.,2013) and regulating alternative splicing (Manukea et al., 2010, Shukla et al., 2011, Foret et al., 2012), as well as a more passive role for methylation as a byproduct of an open chromatin state (Jjingo et al., 2012). Given the nature of the study design, we are not able to directly test the hypothesis that DNA methylation contributes to spurious transcription or the regulation of

alternative isoforms in *C.gigas*. However, genomic feature tracks have been developed and published (Gavery & Roberts 2013) so that genome wide methylation can be easily visualized with respect to gene expression patterns (exon-specific RPKM).

Exons are the preferential target for gene body methylation for most species (Feng et al., 2012), and methylation was enriched in exons in the oyster. However, there is also a relatively large amount of intronic methylation in oysters when compared to other invertebrate species. For example, DNA methylation occurs almost exclusively in exons in the honey bee *A. mellifera* (Lyko et al., 2010). Genome-wide methylation studies in other invertebrate species also report very low levels of intronic methylation relative to other genomic regions (e.g. Gao et al, 2012, Bonasio et al., 2012). Similarly, in plants, methylation is preferably targeted to exons; however it has been reported that in globally methylated mammalian genomes gene body methylation is not biased toward exons (Feng et al., 2012), although exon/intron boundaries can be marked by differences in DNA methylation (Sati et al., 2012). It appears that bivalves may be unique among the invertebrates examined in terms of the degree of methylation in introns. Intronic methylation has been implicated to be involved in gene regulation through the expression of alternative isoforms of genes in other species (e.g. Manukea et al., 2010, Foret et al., 2012). Variation in methylation patterns between taxa may indicate that additional model invertebrates are needed to study the function of these epigenetic marks.

This distribution of DNA methylation in the *C. gigas* genome is consistent with the fractionated or 'mosaic' pattern of methylation previously described in invertebrates (Tweedie et al., 1999, Simmen et al 1999). In oysters, as in other invertebrates, the methylated fraction tends to consist of gene bodies, while other genomic regions exhibit less methylation (Figure 2). Interestingly, transposable elements (TE) show little methylation in oyster gill tissue. This is in

contrast to vertebrate genomes where TE are heavily methylated and function to suppress their activity (Yoder et al 1997). While there is no general consensus regarding the extent of TE methylation across invertebrate taxa, the pattern of sparse TE methylation observed in oysters is similar to what has been described in other invertebrate species (Simmen et al 1999, Feng et al 2010, Zemach et al 2010).

Intra-genic DNA methylation is positively correlated with gene expression in *C. gigas* with the moderately and highly expressed genes showing the highest degree of methylation (Figure 3). This relationship is similar to what has been reported for other invertebrate species (Zemach et al., 2010). Interestingly, Riviere et al (2013) reported a negative relationship between DNA methylation and expression of certain *homeobox (hox)* genes during embryonic development in *C. gigas*. The authors hypothesized that the apparent suppression of *hox* expression by DNA methylation may be due to repression by DNA methylation proximal to the transcription start site in these genes. Although the results reported here and those of Riviere et al. may appear contradictory, it is possible that depending on the context of the methylation (i.e. whether gene body or promoter methylation) it may play either a repressive or expressive role. This is referred to as the DNA methylation paradox (Jones 1999) and is observed in a wide range of taxa.

We used an ordination approach to explore genomic attributes or groups of attributes that predictably co-occur with methylated genes in the *C. gigas* genome. Because multiple factors may be linked with methylation (either through causative or correlative associations), this approach allowed us to identify relationships between multiple variables. The most interesting finding from the PCA analysis is that amount of methylation in a gene is related to the variance in expression between tissues. Genes that show the least variation in expression between tissues

have higher DNA methylation levels than those exhibiting a tissue-specific expression profile (i.e. high %CV between tissues). This observation provides corroboration for previous reports based on *in silico* analyses in oysters showing that housekeeping genes have the highest amount of methylation in *C. gigas* (Gavery & Roberts 2010). Housekeeping genes perform functions required by all cell types, therefore it's expected that their expression patterns would show low variation across tissues. The results of this study are consistent with that expectation that genes with low expression variation across tissues show a high degree of methylation relative to genes with a more tissue-specific expression pattern. Again, this study supports previous findings (Gavery & Roberts 2010, Roberts & Gavery 2012) that heavily methylated genes are enriched in housekeeping functions, which are essential for cellular function. One theory is that the lack of methylation in these genes with more tissue-specific expression is that it can contribute to phenotypic plasticity by allowing more transcriptional opportunities through process such as allowing access to alternative TSS, facilitate exon skipping or other alternative splicing mechanisms and allow for increased sequencing variation (Roberts & Gavery 2012).

Conclusions

Through the current effort, quantitative methylation data were obtained for over 2.5 million CpG dinucleotides throughout the genome of *Crassostrea gigas*. These data represent the first high resolution methylome in any mollusc and the analytical approaches provide a framework for DNA methylation characterization in other species. In addition, this dataset developed here will be beneficial for phylogenetic analysis of DNA methylation in invertebrates, which will be more robust with the addition of a lophotrochozoan species. The results of this study highlight similarities in epigenetic profiles of invertebrates such as a predominance of gene body methylation and a positive relationship between intragenic methylation and gene

expression. In addition, they highlight interesting differences between invertebrate epigenomes including a higher level of intronic methylation in bivalves than what has been reported, for example, in insects. Although the functional role of DNA methylation in bivalves remains elusive, two scenarios could explain our findings. One possibility is DNA methylation in gene bodies is a byproduct of transcription resulting from an open chromatin state, as proposed by Jjingo et al (2013). Thus the methylation patterns are influenced by transcriptional activity. The second scenario is DNA methylation is involved in regulating the gene activity in bivalves. If in fact DNA methylation does influence transcription, the regulatory role is likely very complex. For instance, DNA methylation could have both a have direct regulatory effect on certain genes as proposed by Riviere et al. (2013), as well as facilitating expanded transcriptional opportunities in other cases. Future studies will certainly be challenging given the dynamic nature of DNA methylation, but will hopefully help better delineate if DNA methylation plays a functional role in regulating genome activity in bivalves and what that role might be.

Acknowledgements
The authors would like to thank Bill Howe and Daniel Halperin for their assistance with SQLShare. They would also like to acknowledge Samuel J White for comments that improved the manuscript.

Works Cited

Bell AC, Felsenfeld G. 2000. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* 405:482-5.

Bird A. 1995. Gene number, noise reduction and biological complexity. *Trends Genet* 11:94–100.

Bird AP, Taggart MH.1980. Variable patterns of total DNA and rDNA methylation in animals. *Nucleic Acid Res* 8:1485-1497.

Bonasio R, Li Q, Lian J, MuttiNS, Jin L, Zhao H, Zhang, P, Wen P, Xiang H, DingY, Jin Z, Shen S, Wang Z, Wang W, Wang J, Berger S, Liebig J, Zhang G. 2012. Genome-wide and caste-specific DNA methylomes of the ants Camponotus floridanus and Harpegnathos saltator. *Curr Biol* 22:1755-1764.

Colot V, Rossignol JL.1999. Eukaryotic DNA methylation as an evolutionary device. *BioEssays* 21:402–411.

Fang X, Li L, Luo R, Xu F, Wang X, Zhu Y, Yang L, Huang Z. 2012. Genomic data from the Pacific oyster (*Crassostrea gigas)*. GigaScience. http://dx.doi.org/10.5524/100030.

Feng S, Cokus SJ, Zhang X, Chen PY, Bostick M, Goll MG, Hetzel J, Jain J,Strauss SH, Halpern ME, Ukomadu C, Sadler KC, Pradhan S, Pellegrini M, Jacobsen SE. 2010. Conservation and divergence of methylation patterning in plants and animals. *Proc Natl Acad Sci* 107: 8689–8694.

Fneich S, Dheilly N, Adema C, Rognon A, Reichelt M, Bulla J, Grunau C, Cosseau C. 2013. 5-methyl-cytosine and 5-hydroxy-methyl-cytosine in the genome of Biomphalaria glabrata, a snail intermediate host of Schistosoma mansoni. *Parasites & Vectors* 6:167.

Foret S, Kucharski R, Pellegrini M, Feng S, Jacobsen SE, Robinson GE, Maleszka R. 2012. DNA methylation dynamics, metabolic fluxes, gene splicing, and alternative phenotypes in honey bees. *Proceedings of the National Academy of Sciences* 109:4968–4973.

Gao F, Liu XS, Wu X-P, Wang X-L, Gong D, Lu H, Song Y, Wang J, Du J, Liu S, Han X, Tang Y, Yang H, Jin Q, Zhang X, Liu M. 2012. Differential methylation in discrete developmental stages of the parasitic nematode Trichinella spiralis. *Genome Biol* 13:R100

Gavery M, Roberts SB. 2010. DNA methylation patterns provide insight into epigenetic regulation in the Pacific oyster (*Crassostrea gigas*). *BMC Genomics* 11:483.

Gavery M, Roberts S. 2013. Crassostrea gigas high-throughput bisulfite sequencing (gill tissue). figshare. [http://dx.doi.org/10.6084/m9.figshare.749728] Retrieved 18:44, Oct 21, 2013 (GMT).

Howe B, Cole G, Souroush E, Koutris P, Key A, Khoussainova N, Battle L. 2011. Database-as-a-service for long-tail science. In Proceedings of the 23rd international conference on Scientific and statistical database management (SSDBM'11), Judith Bayard Cushing, James French, and Shawn Bowers (Eds.). Springer-Verlag, Berlin, Heidelberg, 480-489.

Hsieh CL. 1994. Dependence of transcriptional repression on CpG methylation density. *Mol Cell Biol* 14:5487-5494.

Huh I, Zeng J, Park T, Yi S. 2013. DNA methylation and transcriptional noise. *Epigenetics & Chromatin* 6:9.

Jjingo D, Conley AB, Yi SV, Lunyak VV, Jordan IK. 2012. On the presence and role of human gene-body DNA methylation. *Oncotarget* 3:462-474.

Jones PA. 1999. The DNA methylation paradox. *Trends Genet* 15:34–37.

Kucharski R, Maleszka J, Foret S, Maleszka R. 2008. Nutritional control of reproductive status in honeybees via DNA methylation. *Science* 319:1827–1830.

Lechner M, Marz M, Ihling C, Sinz A, Stadler PF, Krauss V. 2013. The correlation of genome size and DNA methylation rate in metazoans. *Theory in Biosciences* 132:47–60.

Li E, Bestor TH, Jaenisch R. 1992. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* 69:915-926.

Lyko F, Foret S, Kucharski R, Wolf S, Falckenhayn C, Maleszka R. 2010. The honey bee epigenomes: differential methylation of brain DNA in queens and workers. *PLoS Biology* 8:e1000506.

Maunakea AK, Nagarajan RP, Bilenky M, Ballinger TJ, D'Souza C, Fouse SD, Johnson BE, Hang C, Nielsen C, Zhao Y, Turecki G, Delaney A, Varhol R, Thiessen N, Shchors K, Heine VM, Rowitch DH, Xing X, Fiore C, Schillebeeckx M, Jones SJM, Haussler D, Marra MA, Hirst M, Wang T, Costello JF. 2010. Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* 466:253–257.

Okano M, Bell DW, Haber DA, Li E. 1999. DNA methyltransferases Dnmt3a and Dnmt3b are essential for *de novo* methylation and mammalian development. *Cell* 99:247–257.

Petrovic V, Perez-Garcia C, Pasantes JJ, Satovic E, Prats E, Plohl M. 2009. A GC-rich satellite DNA and karyology of the bivalve mollusk *Donax Trunculus*: a dominance of GC-rich heterochromatin. *Cytogenet Genome Res* 124: 63–71.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841–842.

R Development Core Team: R: A language and environment for statistical computing. [http://www.R-project.org].

Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* 16:276-7.

Riviere G, Wu GC, Fellous A, Goux D, Sourdaine P, Favrel P. 2013. DNA methylation is crucial for the early development in the oyster *C. gigas*. *Marine Biotechnology* 1–15. doi:10.1007/s10126-013-9523-2.

Roberts SB, Gavery MR. 2012. Is there a relationship between DNA methylation and phenotypic plasticity in invertebrates? *Frontiers in Physiology* 2:116.

Sati S, Tanwar VS, Kumar KA, Patowary A, Jain V, Ghosh S, Ahmad S, Singh M, Reddy SU, Chandak GR, Raghunath M, Sivasubbu S, Chakraborty K, Scaria V, Sengupta S. 2012. High resolution methylome map of rat indicates role of intragenic DNA methylation in identification of coding region. *PloS One* 7:e31621.

Shukla S, Kavak E, Gregory M, Imashimizu M, Shutinoski B, Kashlev M, Oberdoerffer P, Sandberg R, Oberdoerffer S. 2011. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* 479:74–79.

Simmen MW Leitgeb S, Charlton J, Jones SJM, Harris BR, Clark VH, Bird A. 1999. Nonmethylated transposable elements and methylated genes in a chordate genome. *Science* 283:1164–1167.

Smit AFA, Hubley R, Green P. 1996-2010. RepeatMasker Open-3.0. [http://www.repeatmasker.org].

Suzuki MM, Kerr ARW, De Sousa D, Bird A. 2007. CpG methylation is targeted to transcription units in an invertebrate genome. *Genome Res* 17:625-631.

Tweedie S, Charlton J, Clark V, Bird A. 1999. Methylation of genomes and genes at the invertebrate–vertebrate boundary. *Mol Cell Biol* 17:1469–1475.

Wang S, Bao Z, Hu X, Shao M, Zhang L, Hu J. 2008. Two novel elements (CFG1 and PYG1) of mag lineage of Ty3/Gypsy retrotransposons from Zhikong scallop (Chlamys Farreri) and Japanese scallop (Patinopecten Yessoensis). *Genetica* 133:37–46.

Wolffe AP, Matzke MA. 1999. Epigenetics: regulation through repression. *Science* 286:481-486.
Xi Y, Li W. 2009. BSMAP: whole genome Bisulfite Sequence MAPping program. *BMC Bioinformatics* 10:232.

Yoder JA, Walsh CP, Bestor TH. 1997. Cytosine methylation and the ecology of intragenomic parasites. *Trends in Genet* 8:335–340.

Zemach A, McDaniel IE, Silva P, Zilberman D. 2010. Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* 328:916–919.

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

Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SW, Chen H, Henderson IR, Shinn P, Pellegrini M, Jacobsen SE, Ecker JR. 2006. Genome-wide high-resolution mapping and functional analysis of DNA methylation in Arabidopsis. *Cell* 126:1189–1201.

Zilberman D, Gehring M, Tran RK, Ballinger T, Henikoff S. 2007. Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. *Nat Genet* 39: 61–69.

Figures

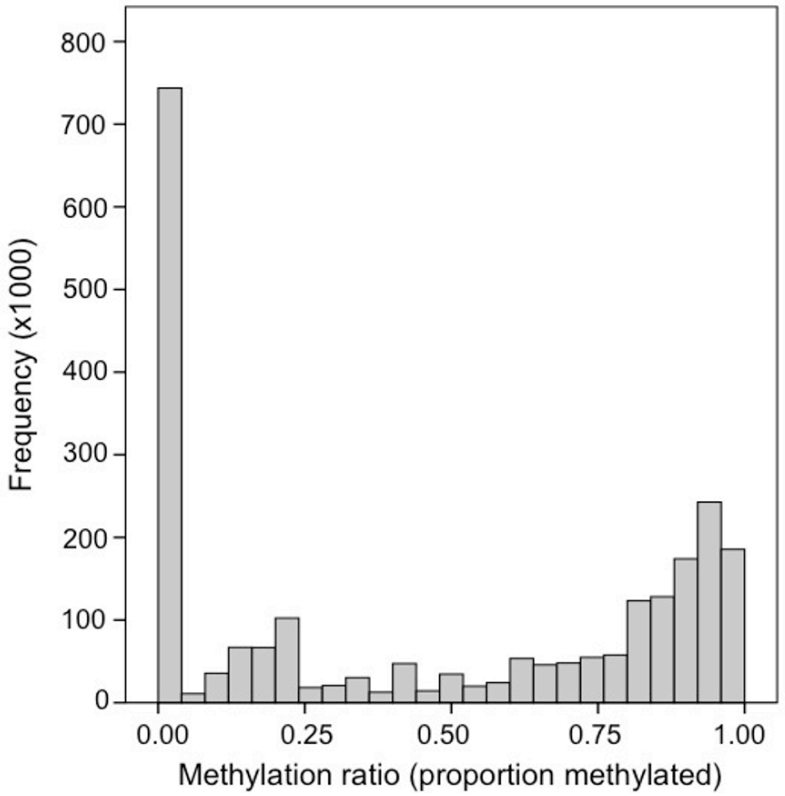


Figure 1.Frequency distribution of methylation ratios for CpG dinucleotides in oyster gill tissue.
A total of 2,625,745 CpG dinucleotides with ≥ 5x coverage are represented.

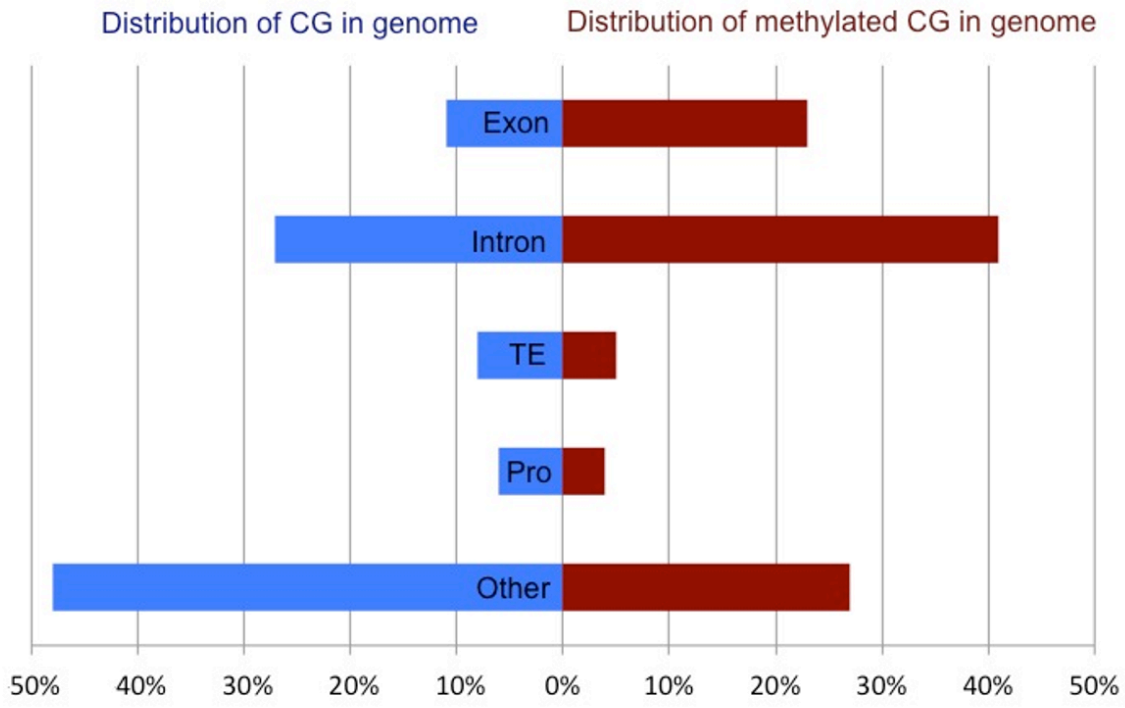


Figure 2. Comparison of the total CpG versus methylated CpG in oyster gill tissue by genomic feature.
Proportion of all CpG (blue) and methylated CpG (red) in gill tissue across genomic features of C. gigas. Percent of CpG dinucleotides in Exons, Introns, Transposable Elements (TE), promoters (Pro) and unannotated intergenic regions (Other) are reported.

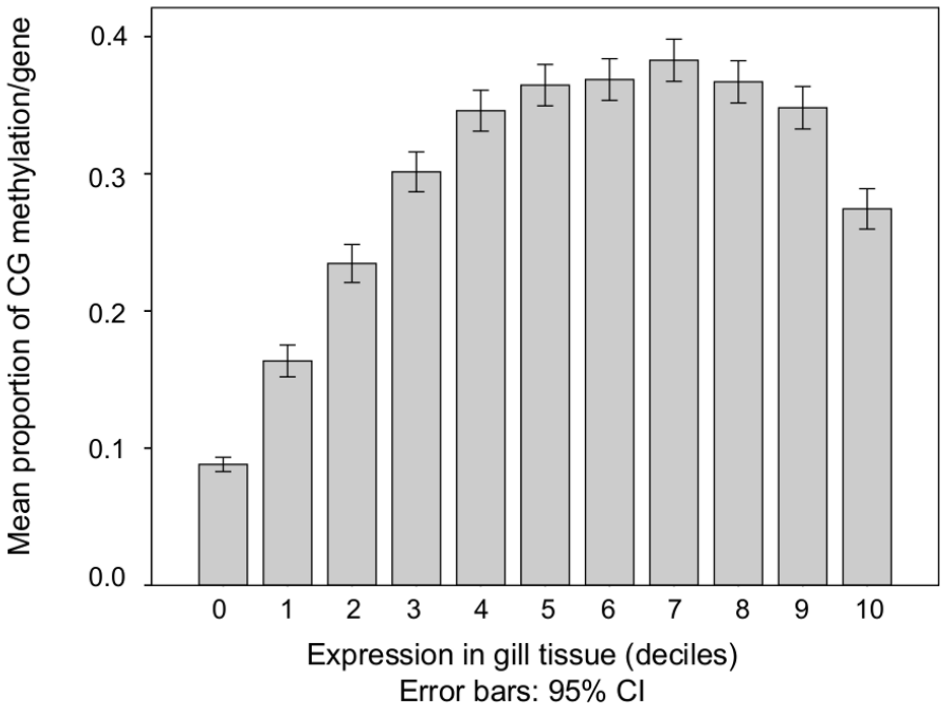


Figure 3. DNA methylation among genes with increasing transcript abundances.
Expressed genes were grouped into deciles by transcription abundance. Genes not expressed in gill (i.e. RPKM=0) are also shown (leftmost column).. Error bars represent 95% confidence intervals.

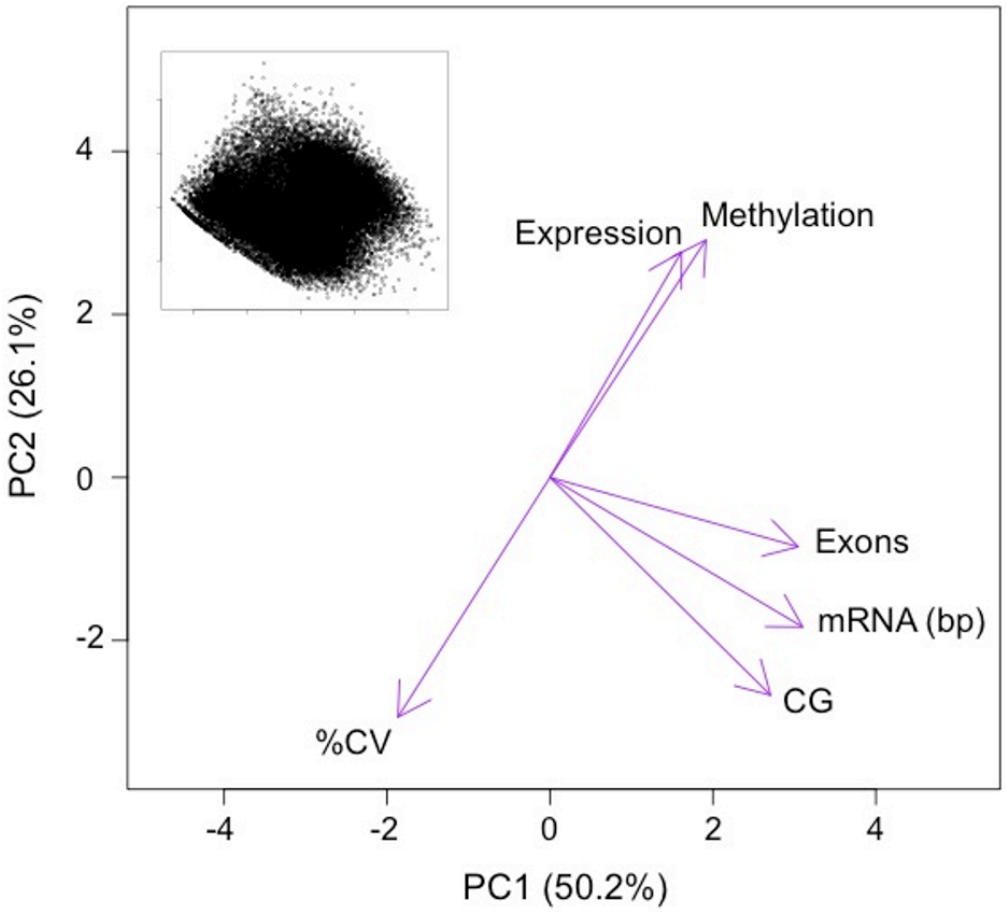


Figure 4 PCA ordination of oyster genes by gene attributes.
Variables loadings shown by purple arrows. Variables significantly contributing to PC1 and PC2 include: methylation ratio (Methylation), the coefficient of variance of expression between tissues (%CV), the number of exons (Exons), the length of the mRNA in base pairs (mRNA) and the number of CpG dinucleotides in the gene (CpG). Variables that did not significantly contributes to PC1 and PC2 include the mean transcript abundance (Expression). Inset depicts ordination of the genes analyzed on PC1 and PC2 (n= 27,181).

Tables

Table 1. Summary of PCA for gene attributes.

Principal Component	% variance explained	Significance value	Significant variable loadings	
PC1	50.2	<0.001	Number of CpG	0.9
			Length mRNA	0.9
			Number of exons	0.8
PC2	26.1	<0.0001	Expression %CV	-0.6
			Methylation ratio	0.6

Chapter IV: A Context Dependent Role for DNA Methylation in Bivalves

A version of this chapter was published in Briefings in Functional Genomics as: Gavery M and Roberts SB. 2014. A context dependent role for DNA methylation in bivalves. Briefings in Functional Genomics (e-published ahead of print January 7th 2014) doi:10.1093/bfgp/elt054

Abstract

The function of DNA methylation in species such as bivalves where the limited amount of DNA methylation is predominantly found in gene bodies remains unclear. An emerging possible explanation is that the role of gene body DNA methylation is dependent on gene function, a potential phenomenon that has arisen from selective pressure on lineage-specific life history traits. In genes contributing to phenotypes that benefit from increased plasticity, the absence of DNA methylation could contribute to stochastic transcriptional opportunities and increased transposable element activity. In genes where regulated control of activity is essential, DNA methylation may also play a role in targeted, predictable genome regulation. Here we review the current knowledge concerning DNA methylation in bivalves and explore the putative role of DNA methylation in both an evolutionary and ecological context.

The variability observed in DNA methylation landscapes and functionality in invertebrates is fascinating from both a molecular and evolutionary perspective. At the molecular level we are still uncovering the many nuances associated with the functional mechanism of methylation, which in turn should eventually provide insight into the evolution of this prevalent epigenetic mark. Although we continue to understand more about DNA methylation in invertebrates, relatively limited information exists concerning the role of DNA methylation in molluscs. The phylum Mollusca is incredibly diverse and consists of over 100,000 species. The class Bivalvia is a particularly relevant group as it includes species of significant ecological (i.e. sentinel species, ecosystem engineers) and commercial (i.e. fisheries, aquaculture) importance. Here we review the current knowledge concerning DNA methylation in bivalves and explore the putative role of DNA methylation in both an evolutionary and ecological context.

The presence of DNA methylation has been confirmed in several bivalves including the Japanese scallop, *Chlamys farreri* [1], the salt water clam, *Donax trunculus* [2], and the Pacific oyster *Crassostrea gigas* [3]. Using high-throughput sequencing of bisulfite treated DNA (BS-Seq) it was recently determined 15% of CpG dinucleotides (1.8% of total cytosines) in the *C. gigas* genome are methylated [4], similar to the 2% total methylation for a gastropod (snail) as measured by LC-MS [5]. Methylation levels reported for the Pacific oyster were characterized in adult gill tissue but it is important to note that methylation levels are likely to vary among life history stages and among tissue types. This point is clearly indicated by Riviere et al [6] where they used an ELISA to quantify relative DNA methylation in developing oysters. Although the ELISA approach does not provide comparable values with respect to the extent of absolute DNA methylation levels, methylation almost doubled during the morula and blastula stage as compared to oocyte and then decreased again during later developmental stages [6].

DNA methylation in bivalves appears to be predominantly found in gene bodies [4]. The observation that gene bodies are the primary methylated genomic feature is consistent with what has been described in other invertebrates (e.g. [7-9]. There is increasing evidence that this form of methylation is the ancestral pattern [10] as gene body methylation is observed not only in invertebrates and vertebrate species [11] but also in plants [12]. The function of DNA methylation in species such as bivalves where the limited amount of DNA methylation is predominantly found in gene bodies remains unclear. One possible explanation that is emerging is that the role of gene body DNA methylation is dependent on the gene function, a potential phenomenon that has arisen from selective pressure on lineage-specific life history traits. In genes whose function may benefit from increased variability (e.g. immune response), the absence of DNA methylation contributes to stochastic 'transcriptional opportunities', whereas genes considered core to survival (e.g. housekeeping genes) are protected from this type of transcriptional variation by the presence of DNA methylation [13]. This theory of beneficial stochastic variation as a result of hypomethylation could also be extended to other regions of the genome such as transposable elements. Further, and not mutually exclusively, DNA methylation may also play a role in a directed and targeted genome regulation. It should also be noted that an alternative explanation for intra-genic DNA methylation patterns is that it is solely a byproduct of having an open and accessible chromatin state [14]. Here we will explore studies of both stochastic and targeted methylation functions that are emerging as potential roles for DNA methylation in bivalves.

Stochastic Variation

A classical explanation of gene body methylation is that it reduces transcriptional noise by preventing initiation of transcription outside of traditional transcription start sites (TSS) [15].

There are data to support this explanation in mammals [16], though to our knowledge, this idea has not been tested directly in an invertebrate. The implication of this explanation is that unmethylated regions would be inherently 'noisier'. In other words, a variety of transcriptional products are produced. It has been proposed that this type of transcriptional noise could result in more diverse transcriptional opportunities [13], which may be beneficial for organisms such as marine bivalves that live in unpredictable and variable nearshore habitats, and as a result, have unpredictable and variable reproduction and recruitment success. As such, oysters may use epigenetic systems to maintain the genomic and phenotypic diversity necessary for a species that undergoes this type of 'sweepstakes reproduction' [17] where chance events dictate which individuals are successful each spawning season. The lack of methylation, by allowing more transcriptional opportunities in genes functionally associated with environmental response, may contribute to phenotypic plasticity by providing access for transcription factors to bind to alternative transcription start sites, facilitating exon skipping or other alternative splicing mechanisms, and/or through unknown mechanisms supporting increased sequencing variation [13] (Figure 1). Although direct evidence is currently lacking to support the idea that hypo-methylation is correlated with increased transcriptional opportunities in bivalves, recent evidence is concordant with this possibility in insects. Specifically, in the honeybee *A. mellifera*, knockdown of global methylation was associated with increased transcriptional opportunities in the form of the generation of splice variants [18] .

Consistent with the theory that limited methylation allows for a variety of transcriptional opportunities is the possibility that transposable element (TE) mobilization may be facilitated by the lack of repressive DNA methylation in bivalves (Figure 1). In vertebrates and some plants, extensive methylation of TEs suppresses their activity in the genome [19]. In invertebrates,

species such as *A. mellifera* show very little methylation in TEs [9]. Likewise, in oysters there appears to be no preponderance of TE methylation [4]. The finding that TEs are not methylated in oysters is consistent with the theory outlined above regarding the ability of a population to present a variety of phenotypes in response to environment change (i.e. phenotypic plasticity). Thus the absence of TE methylation may indicate an evolutionary pressure to retain the variation generated by TE mobilization to maintain genetic diversity in a species inhabiting heterogeneous environments [20].

It is worth considering the relationships between DNA methylation, TEs and transcriptional / genomic variation in light of recent evidence coming from studies of DNA methylation and stress response in plants. For example, it has recently been reported that DNA methylation is involved in regulating the defense response of *Arabidposis thaliana* to the pathogen *Pseudomonas syringae* [21]. Using mutant strains of *A. thaliana* defective in the various types of DNA methylation, Dowen and colleagues were able to show that genome-wide hypomethylation increased plant resistance to the pathogen and was associated with mobilization of TEs and disregulaton of several immune response genes. This was further examined by Yu and colleagues [22] where *Arabidposis* subject to bacterial challenge exhibited globally reduced DNA methylation. This resulting hypomethylation was associated with the reactivation of previously silent TEs. The authors conclude that the defense response in *A. thaliana* is negatively regulated by DNA methylation, and propose that hypomethylation is a part of the plant immune response that acts by priming transcriptional activation of defense genes that are linked to TEs. Considering these studies as a whole, it is interesting that oysters, like plants, which are immobile and face intense selection at early life stages, may benefit from these 'noisy' or 'unstable' genomes. It is important to note that the lack of DNA methylation does not preclude

TE silencing, which can be repressed by a variety of epigenetic mechanisms (reviewed [23]). Future investigation in bivalves should focus on characterizing these additional epigenetic marks (e.g. histone modification, noncoding RNAs) to determine what roles they might play in stabilizing bivalve genomes, and examining the relationship between TE activity and environmental stress.

Targeted Regulation

A second explanation regarding a role for DNA methylation as it pertains to gene body methylation in bivalves is that the epigenetic mark regulates transcriptional activity in a targeted, predictable manner (Figure 2). Evidence is emerging linking gene body methylation to a potential function in regulating alternative splicing [24,25]. The production of both constitutive and alternative isoforms by alternative splicing is important for developmental processes and tissue-specific functions. In oysters, alternative splicing regulates the generation of both tissue-specific [26] and stress activated [27] isoforms of genes. The relationship between methylation and splicing has been examined in a number of studies performed in *Apis mellifera* [9,18,28]. Mechanistically, it has been proposed that exon specific DNA methylation may impact exon-skipping through interactions with DNA binding proteins (CTCF) and subsequent effects on RNA polymerase II pausing [25]. Interestingly, although intronic methylation is rare in *A. mellifera*, Foret et al. [28] identified a relationship between differential methylation in an intron upstream of a differentially expressed cassette exon of the ALK gene. Specifically, they reported that low methylation was correlated with high inclusion of the upstream exon [28].

Additional support for a targeted role in regulating transcription in bivalves is the recent work of Riviere et al [6]. Investigators examined the relationship between expression and methylation in homeobox (hox) genes, a family of genes that are critical developmental genes.

Riviere et al. [6] observed an inverse relationship with gene body methylation and expression, and hypothesized that the apparent suppression of hox expression by DNA methylation was due to a "CpG island-like" repression by DNA methylation proximal to the transcription start site (TSS) in these genes. Results were obtained using methylated DNA immunoprecipitation (MeDIP) qPCR, so the context of the region investigated was known. When possible (6 out of 10 genes) the region examined was in the 1st exon or 5'UTR. The trend is similar to repression in proximal promoter/1st exon repression as seen in mammals. Riviere et al [6] not only provide evidence of active regulation of transcription via gene body methylation, but their work also suggests mechanism similar to the conventional repressive nature of promoter methylation in vertebrates. While little research exists on the relationship of promoter methylation and expression in invertebrates, there is at least one report in molluscs. In *Aplysia*, Rajasethupathy et al [29] found that serotonin exposure induced an increase in methylation in the promoter of the CREB2 gene, which is also associated with the downregulation of CREB2 mRNA in neurons. In general, CpG island containing promoter methylation is not typical in invertebrates [8]; however it is possible that depending on the context of the methylation (i.e. whether gene body or promoter methylation) it may play either a repressive or expressive role. This is known as the DNA methylation paradox [30] and has been observed in a wide range of taxa.

Future Directions

Continued endeavors exploring the role of DNA methylation in invertebrates will certainly shed light on general similarities and lineage specific nuances. There remains a multitude of research questions and phenomena that need attention; among them are some of the ideas presented here. The only direct evidence available relating DNA methylation and expression in bivalves focuses either on a single family of genes (i.e. hox) [6] or genome-wide

analysis of pooled individuals [4]. To ultimately gain a better understanding of this, future studies are needed to characterize genome wide methylation and gene expression on individuals with consideration towards cell-type, developmental stage, and environmental condition. A draft genome of *C. gigas* is now available [31] and new bivalve genomic resources are increasingly available to the scientific community, allowing us to characterize stochastic versus targeted roles for DNA methylation in bivalves. Future investigations into other epigenetic phenomena, including histone modifications and non-coding RNAs, will provide a fuller picture regarding genome regulation in bivalves. Another important question is the extent that the environment influences DNA methylation in bivalves. In other species it has been clearly shown that DNA methylation can be influenced by the environment [32–34]. Interestingly, one of the best examples of this phenomenon comes from findings in an invertebrate. In honeybees, larvae fed royal jelly become queens, which are phenotypically distinct from workers. It has been shown that DNA methylation serves as an intermediary between this environmental signal (nutrition) and the developmental outcome into a queen or a worker [35]. It is a likely generality that the environment influences DNA methylation in bivalves, though possibly in a different fashion, in light of the ideas presented here with respect to the stochastic nature of new transcriptional opportunities and local adaptation.

It remains to be determined to what extent transgenerational epigenetic inheritance occurs in invertebrate taxa. In mammals, evidence exists of transgenerational inheritance of DNA methylation patterns and phenotypes in response to certain xenobiotics (e.g. [36,37]). Transgenerational inheritance of DNA methylation patterns associated with phenotypes (epialleles) have also been observed in plants [38,39], including evidence that environmental stress induces heritable changes [40]. Transgenerational epigenetic inheritance has not been

investigated in bivalves, but one particularly intriguing possibility to explore is the role of DNA methylation in protecting future generations through an acquired stress response. Bivalves are generally sessile and do not directly interact with their offspring. One way a bivalve could 'inform' their offspring about recent environmental conditions is through the transmission of epigenetic marks such as DNA methylation. If epigenetic marks are heritable, they may play a role in evolutionary processes. To address the question of heritability, we need to compare levels of existing epigenetic variation in natural populations with genetic variation. This indeed could be a game changer, as epigenetic variation may offer a new platform for selection. There has been some work done in vertebrates and plants [41–43], though information in invertebrates is limited. Researchers have started to address this fact in oyster aquaculture settings in response to mass selection protocols. Jiang et al [44] used a methylation-sensitive amplified polymorphism (MSAP) methodology to identify epigenetic variation between a base population and a fourth generation mass selection population. They also used AFLP to look at genetic variation. Jiang et al [44] found genetic variation with no epigenetic variation over all, though specific differences were observed. The authors observed a correlation between epigenetic and genetic variation. Despite the limitations of this study in using a relatively small number of random markers, it is the first study comparing epigenetic and genetic variation in bivalves and illuminates an interesting direction for future work.

The relationship of heritable transmission of genome patterns and epigenetic resetting is another research avenue that should be explored. In mammalian systems epigenetic resetting, a clearing and re-establishment of DNA methylation with each generation, is thought to be necessary to induce pluripotency of cells (reviewed [45]). Nevertheless, there are certain loci (e.g. imprinted genes) where the clearing of epigenetic marks is incomplete resulting in meiotic

inheritance of DNA methylation patterns. This type of transgenerational inheritance has been studied in plants and mammals, but to our knowledge has yet to be addressed in invertebrates. As mentioned above, oysters show temporal changes in the total amount of DNA methylation during embryonic development, with lower methylation in the 2-4 cell stages and increasing in morula and blastula [6]. This observation may be indicative of an epigenetic resetting event. However, characterizing epigenetic changes at finer temporal time scales are needed.

Exploring these questions of epigenetic flexibility to environmental cues, natural variation, heritability, as well as the possibility of epigenetic resetting in bivalves will inform the direction of much larger research questions. While we are gaining a better understanding of invertebrate epigenetics, we certainly have a lot more to learn, which could considerably change our comprehension of organismal and ecosystem responses to environmental change.

Works Cited

1. Wang S, Bao Z, Hu X, et al. Two novel elements (CFG1 and PYG1) of Mag lineage of Ty3/Gypsy retrotransposons from Zhikong scallop (Chlamys farreri) and Japanese scallop (Patinopecten yessoensis). Genetica 2008; 133:37–46

2. Petrovic V, Perez-Garcia C, Pasantes JJ, et al. A GC-rich satellite DNA and karyology of the bivalve mollusk Donax trunculus: a dominance of GC-rich heterochromatin. Cytogenet Genome Res 2009; 124:63–71

3. Gavery MR, Roberts SB. DNA methylation patterns provide insight into epigenetic regulation in the Pacific oyster (Crassostrea gigas). BMC genomics 2010; 11:483

4. Gavery M, Roberts S. Predominant intragenic methylation is associated with gene expression characteristics in a bivalve mollusc. PeerJ 2013; 1:e215; DOI 10.7717/peerj.215

5. Fneich S, Dheilly N, Adema C, et al. 5-methyl-cytosine and 5-hydroxy-methyl-cytosine in the genome of Biomphalaria glabrata, a snail intermediate host of Schistosoma mansoni. Parasites & Vectors 2013; 6:167

6. Riviere G, Wu G-C, Fellous A, et al. DNA methylation is crucial for the early development in the oyster C. gigas. Marine Biotechnology 2013; 1–15

7. Simmen MW, Bird AP. Sequence analysis of transposable elements in the sea squirt, Ciona intestinalis. Mol Biol Evol 2000; 17:1685–1693

8. Zemach A, McDaniel IE, Silva P, et al. Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science 2010; 328 :916–919

9. Lyko F, Foret S, Kucharski R, et al. The honey bee epigenomes: differential methylation of brain DNA in queens and workers. PLoS Biology 2010; 8:e1000506

10. Lechner M, Marz M, Ihling C, et al. The correlation of genome size and DNA methylation rate in metazoans. Theory in Biosciences 2013; 132:47–60

11. Aran D, Toperoff G, Rosenberg M, et al. Replication timing-related and gene body-specific methylation of active human genes. Human Molecular Genetics 2011; 20 :670–680

12. Zilberman D, Gehring M, Tran RK, et al. Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. Nat Genet 2007; 39:61–69

13. Roberts SB, Gavery MR. Is there a relationship between DNA methylation and phenotypic plasticity in invertebrates? Frontiers in physiology 2012; 2:116

14. Jjingo D, Conley AB, Yi S V, et al. On the presence and role of human gene-body DNA methylation ABSTRACT. Oncotarget 2012; 3:462–474

15. Bird AP. Gene number, noise reduction and biological complexity. Trends Genet 1995; 11:94–100

16. Huh I, Zeng J, Park T, et al. DNA methylation and transcriptional noise. Epigenetics & Chromatin 2013; 6:9

17. Hedgecock D. Does variance in reproductive success limit effective population sizes of marine organisms? Genetics and Evolution of Aquatic Organisms. 1994; 122–134

18. Li-Byarlay H, Li Y, Stroud H, et al. RNA interference knockdown of DNA methyl-transferase 3 affects gene alternative splicing in the honey bee. Proceedings of the National Academy of Sciences 2013; 110 :12750–12755

19. Yoder JA, Walsh CP, Bestor TH. Cytosine methylation and the ecology of intragenomic parasites. Trends in Genetics 1997; 13:335–340

20. Casacuberta E, González J. The impact of transposable elements in environmental adaptation. Molecular Ecology 2013; 22:1503–1517

21. Dowen RH, Pelizzola M, Schmitz RJ, et al. Widespread dynamic DNA methylation in response to biotic stress. Proceedings of the National Academy of Sciences 2012; 109 :E2183–E2191

22. Yu A, Lepère G, Jay F, et al. Dynamics and biological relevance of DNA demethylation in Arabidopsis antibacterial defense. Proceedings of the National Academy of Sciences 2013; 110 :2389–2394

23. Slotkin RK, Martienssen R. Transposable elements and the epigenetic regulation of the genome. Nat Rev Genet 2007; 8:272–285

24. Maunakea AK, Nagarajan RP, Bilenky M, et al. Conserved role of intragenic DNA methylation in regulating alternative promoters. Nature 2010; 466:253–257

25. Shukla S, Kavak E, Gregory M, et al. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. Nature 2011; 479:74–79

26. Rodet F, Lelong C, Dubos M-P, et al. Alternative splicing of a single precursor mRNA generates two subtypes of gonadotropin-releasing hormone receptor orthologues and their variants in the bivalve mollusc Crassostrea gigas. Gene 2008; 414:1–9

27. Guévélou E, Huvet A, Sussarellu R, et al. Regulation of a truncated isoform of AMP-activated protein kinase α (AMPKα) in response to hypoxia in the muscle of Pacific oyster Crassostrea gigas. Journal of Comparative Physiology B 2013; 183:597–611

28. Foret S, Kucharski R, Pellegrini M, et al. DNA methylation dynamics, metabolic fluxes, gene splicing, and alternative phenotypes in honey bees. Proceedings of the National Academy of Sciences 2012; 109 :4968–4973

29. Rajasethupathy P, Antonov I, Sheridan R, et al. A role for neuronal piRNAs in the epigenetic control of memory-related synaptic plasticity. Cell 2012; 149:693–707

30. Jones PA. The DNA methylation paradox. Trends in Genetics 1999; 15:34–37

31. Fang X, Li L, Lou R, et al. Genomic data from the Pacific oyster (Crassostrea gigas). GigaScience 2012;

32. Chinnusamy V, Zhu J. Epigenetic regulation of stress responses in plants. Current Opin Plant Biol 2009; 12:133–139

33. Navarro-Martín L, Viñas J, Ribas L, et al. DNA methylation of the gonadal aromatase (*cyp19a*) promoter is involved in Temperature-Dependent Sex Ratio Shifts in the European Sea Bass. PLoS Genet 2011; 7:e1002447

34. Yu Y, Yang X, Wang H, et al. Cytosine methylation alteration in natural populations of *Leymus chinensis* induced by multiple abiotic stresses. PLoS ONE 2013; 8:e55772

35. Kucharski R, Maleszka J, Foret S, et al. Nutritional control of reproductive status in honeybees via DNA methylation. Science 2008; 319 :1827–1830

36. Manikkam M, Guerrero-Bosagna C, Tracey R, et al. Transgenerational actions of environmental compounds on reproductive disease and identification of epigenetic biomarkers of ancestral exposures. PLoS ONE 2012; 7:e31901

37. Guerrero-Bosagna C, Settles M, Lucker B, et al. Epigenetic transgenerational actions of vinclozolin on promoter regions of the sperm epigenome. PLoS ONE 2010; 5:e13100

38. Cubas P, Vincent C, Coen E. An epigenetic mutation responsible for natural variation in floral symmetry. Nature 1999; 401:157–161

39. Manning K, Tor M, Poole M, et al. A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. Nat Genet 2006; 38:948–952

40. Verhoeven KJF, Jansen JJ, van Dijk PJ, et al. Stress-induced DNA methylation changes and their heritability in asexual dandelions. New Phytologist 2010; 185:1108–1118

41. Liu S, Sun K, Jiang T, et al. Natural epigenetic variation in the female great roundleaf bat (Hipposideros armiger) populations. Molecular Genetics and Genomics 2012; 287:643–650

42. Schrey AW, Coon CAC, Grispo MT, et al. Epigenetic variation may compensate for decreased genetic variation with introductions: a case study using house sparrows (Passer domesticus) on two continents. Genetics research international 2012; 2012:

43. Herrera CM, Bazaga P. Epigenetic differentiation and relationship to adaptive genetic divergence in discrete populations of the violet Viola cazorlensis. New Phytologist 2010; 187:867–876

44. Jiang Q, Li Q, Yu H, et al. Genetic and epigenetic variation in mass selection populations of Pacific oyster Crassostrea gigas. Genes & Genomics 2013; 1–7

45. Santos F, Dean W. Epigenetic reprogramming during early development in mammals. Reproduction 2004; 127 :643–651

Figures

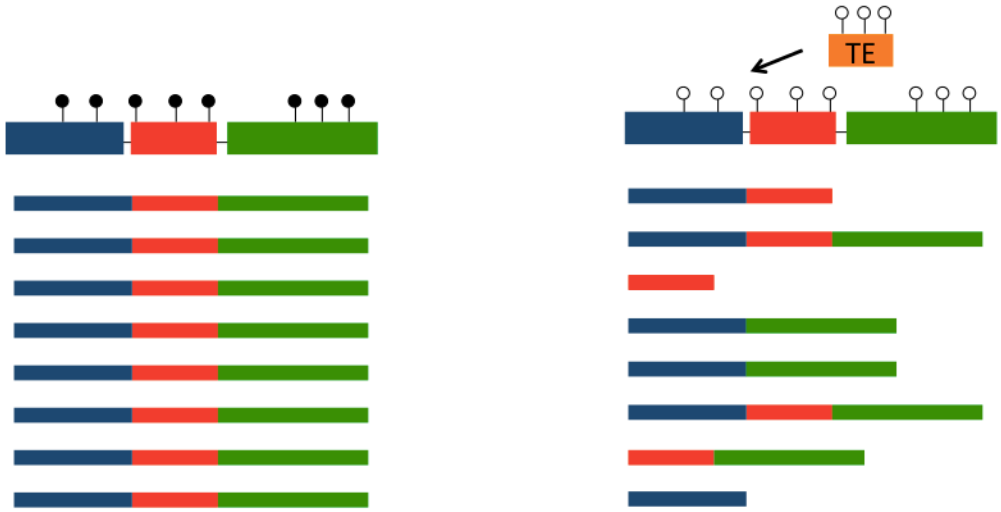


Figure 1. Stochastic Regulation

A simplified model of stochastic transcriptional opportunities based on limited DNA methylation. Open and closed circles represent unmethylated and methylated CpG, respectively. Thick boxes represent individual exons of a single gene, thin boxes below represent transcriptional outcomes. In this theoretical model predominantly methylated genes (left) produce consistent transcriptional outcomes, whereas unmethylated genes (right) generate transcriptional 'noise' in the form of splice variants. In addition, unmethylated TEs may actively insert into the genome where they could produce transcriptional variation in the form of truncated transcripts or splice variants.

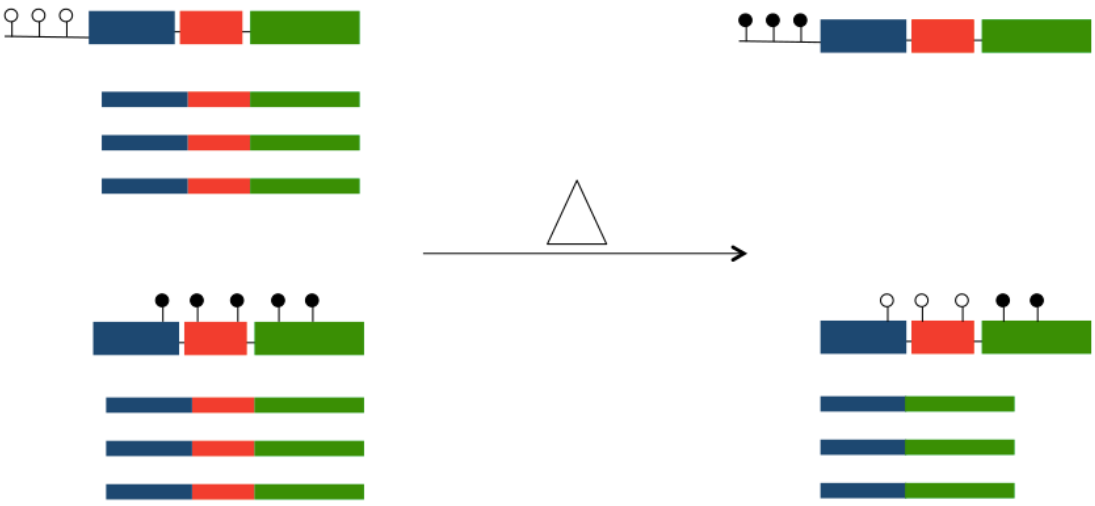


Figure 2. Targeted Regulation
A simplified model of targeted regulation of gene products via dynamic methylation/demethylation in response to extrinsic or intrinsic signals Open and closed circles represent unmethylated and methylated CpG, respectively. Thick boxes represent individual exons of a single gene, thin boxes below represent transcriptional outcomes. In the top example, changes in methylation status in proximity to transcription initiation site may inhibit or promote transcription. In the lower example a change in methylation status in the gene body produces a transcriptional variant. Potential initiators of these methylation changes could be cues from the environment or associated with developmental processes.

Synthesis and Conclusion

By combining transcriptomic and epigenetic datasets, this work provides the most complete picture of epigenomic regulation for a molluscan species and paves the way for future investigations of the role of epigenetics in environmental response, local adaptation, and evolution in marine invertebrates. Data in Chapters 1 and 3 also highlight the advantages of applying high-throughput sequencing technology to study gene regulation and DNA methylation to examine the interactions between genotype, phenotype, and environment in non-model species. The work presented here on DNA methylation in oysters is the first in any bivalve species and highlights both conserved and novel properties of the DNA methylation system in molluscs.

A significant finding from this research is that diversity of DNA methylation patterns across species cannot be underestimated. It is well known that vertebrates and invertebrates display large differences in the extent and context of methylation in their genomes (Bird & Taggart 1980, Elango & Yi, 2008). However, this dissertation also highlights there are differences in the extent and context of DNA methylation within invertebrate taxa. Among invertebrates, insects are the most well-studied group in terms of DNA methylation patterns, and the honey bee, *A. mellifera,* is emerging as a model species for studies of DNA methylation in invertebrates (Hunt et al., 2013). It is now clear from the results of Chapter 3 that there are substantial differences between DNA methylation patterns in social insects and bivalve molluscs. For example, while DNA methylation is limited almost exclusively to exons in *A. mellifera,* there is substantial methylation present in intronic and intergenic regions in *C. gigas*. These data suggest that functions of DNA methylation may not be conserved even among invertebrate taxa. In 1999, Colot and Rossignol hypothesized that the presence of a functioning DNA methylation

system confers a selective advantage not because it performs a single essential function, but instead, it provides a means for setting up a variety of functions that could be used to different ends depending on the need of the species (Colot & Rossignol 1999). Comparative analysis of oyster and honey bee DNA methylation data supports this hypothesis. An oyster's unique life history as a broadcast spawner inhabiting estuarine intertidal zones poses unique challenges that other emerging model species do not face. Thus, it will be important to continue this line research and develop *C. gigas* as a model of DNA methylation for marine invertebrates.

The hypothesis proposed in Chapter 4 regarding stochastic and targeted gene regulation via DNA methylation provides a strong basis to experimentally address the functional role of DNA methylation in oysters. The DNA methylation system in oysters may have significant ecological ramifications in terms of promoting phenotypic plasticity within populations. For example, if DNA methylation patterns are facilitating stochastic regulation in the oyster, it is possible that increased transcriptional variation in environmental response genes would result in the production of various alternatively spliced isoforms. One benefit of this type of promiscuous regulation would be that some individuals would produce isoforms that function optimally in a particular environment. On the other hand, if DNA methylation is being utilized to promote targeted regulation of genes in the oyster, it is possible that a specific environmental cue, such as exposure to high temperature, could induce transient methylation events that result in the production of a specific protein or protein isoform that is best suited to respond to that stress. In both of these scenarios, whether the transcriptional response is stochastic or targeted, the methylation patterns are increasing the robustness of a population under conditions of environmental change by contributing to phenotypic plasticity.

Furthermore, it is possible that DNA methylation may function to enhance genetic variation in oysters. This is based on the observation that transposable elements lack methylation in the oyster genome. In other species it has been shown that transposable elements are heavily methylated, contributing to stability. Transposable element mobility can cause rearrangements in the genome creating new genetic variation. In terms of what this may mean for oyster populations, it is possible that even in the case of a severe environment event, such as a disease outbreak that can exert a strong selection pressure for a specific phenotype, populations may still have the ability to generate new genetic variation through the activity of transposable elements. This would increase diversity and possibly increases odds of survival upon facing future stressors. One important research direction going forward will be to understand and predict how organisms like oysters will respond to climate change and ocean acidification. If the DNA methylation system is indeed facilitating variation, it may indicate that oysters are epigenetically equipped, in terms of transcriptional and genomic plasticity, to demonstrate increase resilience in the face of environment change beyond what is currently predicted considering genetic variation alone.

In addition to advancing our ecological and physiological understanding of aquatic invertebrates, this research also has important implications for the aquaculture industry. Phenotypic plasticity is generally not a desired trait in an aquaculture setting. If an oyster is selected for a particular phenotype (e.g. fast growth), the oyster would ideally exhibit that phenotype under a wide variety of environments and locations. If it holds true that the DNA methylation system in oysters is predisposed toward plasticity, it may follow that that the production of selected oyster lines that produce the same phenotype in multiple environments will be a major challenge for the industry.

In conclusion, these combined results, proposed functional theories, and technical approaches provide substantial footholds for future research investigating molecular mechanisms of phenotypic variation, plasticity, and genotype by environment interaction in oysters. At the same time this work begins to challenge current paradigms in genetics and physiology that will undoubtedly garner significant attention as genome sequencing technology improves and epigenetic processes are more readily incorporated into organismal and ecological studies.

Works Cited

Bird AP, Taggart MH. 1980. Variable patterns of total DNA and rDNA methylation in animals. *Nucleic Acid Res* 8:1485-1497

Colot V, Rossignol J. 1999. Eukaryotic DNA methylation as an evolutionary device. *BioEssays* 21: 402-411.

Elango N, Yi SV. 2008. DNA methylation and structural and functional bimodality of vertebrate promoters. *Mol Biol and Evol* 25:1602-1608.

Hunt BG, Glastad KM, Yi SV, Goodisman MAD. 2013. The function of intragenic DNA methylation: insights from insect epigenomes. *Integrative and Comparative Biology* 53: 319-328.

Appendix A: Characterization of prostaglandin endoperoxide synthase from the Pacific oyster, *Crassostrea gigas*.

Work accomplished in collaboration with Samuel White

Background

Prostaglandins are a group of lipid compounds derived from fatty acids that play important local (e.g., paracrine) roles in numerous physiological processes including muscle contraction, inflammation and endocrine signalling (Riccioti & Fitzgerald, 2011). Synthesis of prostaglandins is mediated by prostaglandin H2 synthase (PGS), which converts the unsaturated fatty acid arachidonic acid, to prostaglandin H2 (PGH2) which is then be converted to the primary bioactive prostaglandins.

Prostaglandin synthases have been well-studied in mammals and have been shown to be important molecules in the innate immune response, particularly in regulating inflammation (Riccioti & Fitzgerald, 2011). However, discovery and descriptions of prostaglandins in lower organisms has lagged behind mammals. In invertebrates, only two full-length coding sequences for prostaglandin synthases have been described in the corals *G. fruiticosa* and *P. homomalla* (Koljak et al., 2001, Valmsen et al., 2001). In the Pacific oyster, *Crassostrea gigas*, a gene expression study revealed increased expression in genes involved in prostaglandin biosynthesis and receptors (Roberts et al., 2008). In order to facilitate further research into the role of prostaglandins in oyster immunity we present the first description and characterization of a full-length prostaglandin synthase gene in shellfish.

Methods & Materials

Oyster Sampling

Oysters (mean length = 114mm; Taylor Shellfish Farms) were held in 18^{o} C seawater at the University of Washington. To examine overall gene expression patterns, five oysters were

randomly selected for tissue collection. Tissues from gill, mantle, muscle, and visceral mass were aseptically collected, immediately frozen and then stored at -80 ^{o} C.

Oysters were exposed to *Vibrio vulnificus* via bath immersion. A bacterial culture was grown overnight in 400 mL 1x Luria Bertani broth (LB), plus an additional 1% NaCl, in a shaking incubator at 37^{o} C, 150 RPM. The culture was pelleted (10 min, 4,300 RPM, 25 ^{o} C, Sorvall ST-H750 Rotor) and supernatant discarded. Bacteria (3.64 x 10^{19} colony forming units [CFU]) were resuspended in 50 mL of sterile seawater and added to a container of oysters (n=16) in 8 L of seawater. A second 8 L container of oysters (n=16) contained only seawater. Both containers were aerated. At 1 and 3 hours post-exposure, gill tissue from 8 oysters under both conditions was collected using sterile techniques. Tissue was immediately frozen and stored at -80^{o} C.

Bacterial counts in the seawater containing *V. vulnificus* were 3.84 x 10^{12} CFUs and 1.84 x 10^{12} CFUs, after 1 and 3 hours, respectively. Bacterial counts in the control seawater after 1 and 3 hours were 1.17 x 10^{6} CFUs and 8.48 x 10^{5} CFUs, respectively.

Sample Preparation and Quantitative PCR

Total RNA was isolated from tissues using TriReagent (Molecular Research Center, Inc.) according to the manufacturer's protocol. To eliminate possible DNA carryover, total RNA was DNase treated using the Turbo DNA-free Kit (Ambion) according to the manufacturer's "rigorous" protocol. Removal of genomic DNA was confirmed via real-time PCR on DNAse-treated total RNA. RNA (1 mg was reverse transcribed using M-MLV Reverse Transcriptase (Promega) and oligo dT primers (Promega) according to the manufacturer's protocol. Resulting cDNA was diluted four-fold with molecular biology grade water and stored at -20^{o} C.

Quantitative PCR (qPCR) was performed using 1 mL of cDNA in a 20 mL reaction containing 1x Sso Fast EvaGreen Supermix (BioRad) and 0.25 mM each of forward and reverse primers (Table 1). Thermal cycling and fluorescence detection were performed on a CFX96 Real-Time Detection System (BioRad) using the following cycling parameters: one cycle of 98^{o}C for 2 min; 40 cycles of: 98^{o}C for 5 sec, then 60^{o}C for 5 sec. Immediately after cycling, a melting curve protocol was run to verify that a single product was generated in each reaction. All qPCR reactions were performed in duplicate.

Analysis of qPCR data was carried out using raw fluorescence data (no baseline subtraction) based upon the kinetics of individual qPCR reactions using Real-time PCR Miner v2.1 (Zhao & Fernald, 2005). All data were normalized to corresponding GAPDH (NCBI Accession: AJ544886) values and represented as fold increases over the minimum. Statistical analysis was performed using one-way ANOVA followed by Tukey's multiple-comparison posttest using SPSS software.

Cloning and Sequencing of Oyster PGS1

To obtain the full coding sequence of PGS, 5'/3' rapid amplification of cDNA ends (RACE) was performed using the SMART RACE cDNA Amplification Kit (Clontech), according to the manufacturer's protocol. Primers for RACE (CgPGS1 5' RACE, CgPGS1 3' RACE; Table 1) were designed with Primer3 (Rozen & Skaletsky, 2000) based on a *C. gigas* EST (AM856036) in NCBI's GenBank. The full length coding sequence obtained through RACE was confirmed by producing a single amplicon with PCR. PCR reactions were performed using 1 mL of equally pooled cDNA from all samples (see above) in a 25 mL reaction containing 1x Apex Red TAQ DNA Polymerase Master Mix (1.5 mM $MgCl_2$; Genesee Scientific) and 0.2 mM each of forward and reverse primers (Table 1). RACE and PCR products were separated on

low TE (10 mM Tris-HCl, 1 mM EDTA) agarose gels and bands were excised. DNA was purified using the Ultra-free DA Kit (Millipore) according to the manufacturer's protocol, cloned into vector pCR 2.1-TOPO, and transformed into Top10 chemically competent cells using the TOPO TA Cloning Kit (Invitrogen) according to the manufacturer's protocol. Plasmid DNA was isolated from bacterial cultures using the Qiagen MiniPrep Kit (Qiagen) and sequenced at the High Throughput Genomics Unit (University of Washington). Electropherograms were visually inspected using Geneious 5.3.6 (Drummond et al., 2010). Sequence analysis and characterization were performed using Geneious 5.3.6. Raw nucleotide sequencing data was trimmed to remove low quality and vector sequence. Sequence alignments were performed in Geneious 5.3.6 using ClustalW. A phylogenetic tree was generated using the Jukes-Cantor Genetic Distance Model with a Neighbor-Joining tree building model and bootstrapped 10,000 times using Geneious 5.3.6.

Results

Sequence Analysis

Trimmed sequences were de novo assembled. The assemblies yielded a consensus sequence of 2,416 base pairs. An open reading frame was identified from bases 416 - 2,179. Translation of this open reading frame resulted in a sequence of 587 amino acids. BLASTp of the deduced amino acid sequence revealed high homology with PGS1 of other species. Analysis of the CgPGS1 protein sequence using GenBank's Conserved Domains tool (Marchler-Bauer A et al., 2011) revealed a match (e value = 0) to animal prostaglandin endoperoxide synthase from amino acids 77 - 563, providing additional support for characterization of this sequence as a prostaglandin synthase. The CgPGS1 nucleotide sequence was submitted to GenBank and assigned the following accession number: FJ375303.

CgPGS1 was aligned with 15 other species of PGS1 using ClustalW (Figure 1). The alignment revealed 30.7% identical sites and pairwise identity of 65.8% across all species. Phylogenetic mapping (Figure 2) of all 16 PGS1 sequences revealed that CgPGS1 was most similar to the marine soft coral, *Gersemia fruticosa* (55.8% identical sites, 55.6% pairwise identities), and the marine Black Sea rod coral, *Plexaura homomalla* (53.8% identical sites, 53.6% pairwise identities).

Gene Expression

The highest expression of CgPGS1 was observed in the gill followed by the mantle. The Dg/gonad and muscle showed the lowest expression of all tissues examined (Fig. 3). Vibrio vulnificus had no effect on CgPGS1 expression at either 1 or 3 hours after exposure (Fig. 4).

Discussion

Here, we have present the first description and characterization of a full-length prostaglandin synthase sequence in shellfish. The *Crassostrea gigas* PGS sequence has high homology to other animal PGS1 sequences and is most similar to coral species. Gene expression analysis indicates the highest expression in gill tissue consistent with a role in the immune system as gill tissue contains a large number of circulating hemocytes, the oyster's immune cells. Interestingly, there was no change in expression PGS1 after a short exposure to *Vibrio vulnificus*, indicating that mRNA synthesis PGS is not involved in short exposures to bacteria, however it is possible that transcription may occur after longer term exposures. Characterization of this sequence will facilitate additional research into the role of prostaglandins in invertebrate species.

Tables and Figures

Table 1. Primer sequences used for RACE PCR and qPCR

Primer Name	Accession Number	Sequence	Product Size(s)
CgPGS1 5' RACE	AM856036	CCCTTCACAGAATACGGGGCACCAA	~1700bp ~1900bp
CgPGS1 3' RACE	AM856036	CCCCGAAATCATGGCCCTGTGACAA	~1600bp
CgPGS1 qPCR F	FJ375303	GGCCGCACCTTAAGGCTGCAGT	117bp
CgPGS1 qPCR R	FJ375303	GCACCAATTTGTACCAAAGAGTCACCAAAC	
CgPGS1 PCR F	FJ375303	TCTTCACAGGTTGTATTTCACGTGGG	~2300bp
CgPGS1 PCR R	FJ375303	CACACTCACAGTACAATATCAGTCAAATTTC	
CgGAPDH qPCR F	AJ544886	GCGAACGGGATCCAGCCAACA	151bp
CgGAPDH qPCR R	AJ544886	GCGTCAGCTGAGGGGGCAGA	

Figure 1. Sequence alignment of PGS1

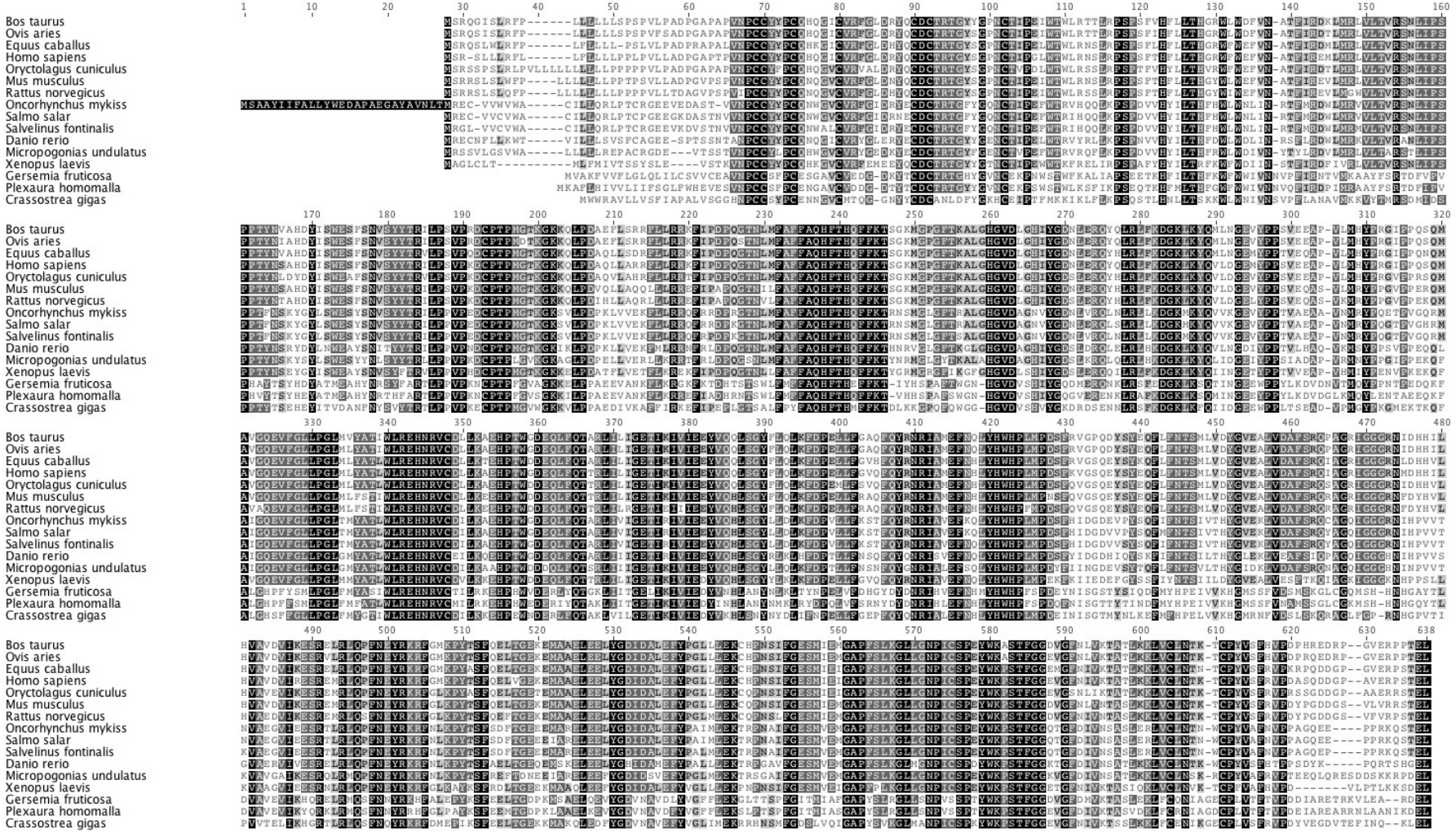


Figure 2. Phylogenetic map of 16 PGS1 sequences

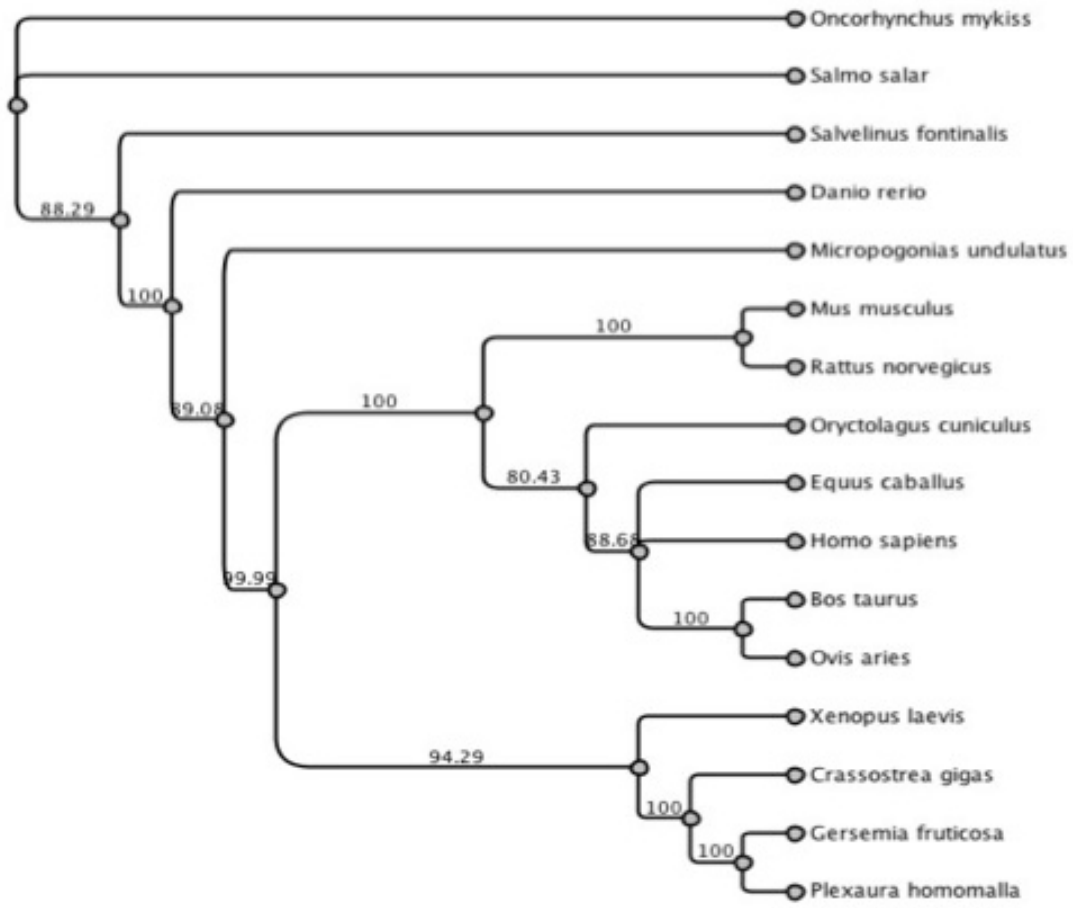


Figure 3. Messenger RNA expression levels of PGS1 across *C. gigas* tissues

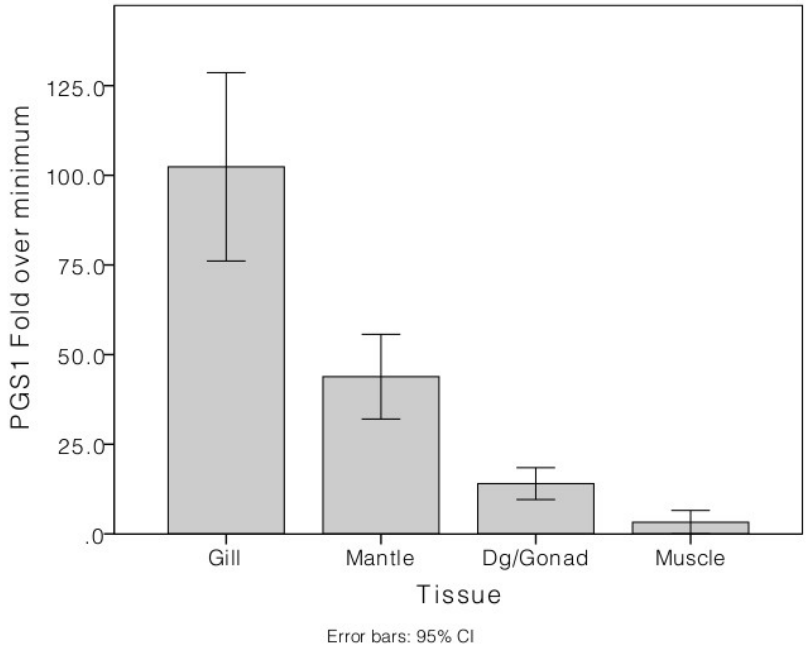
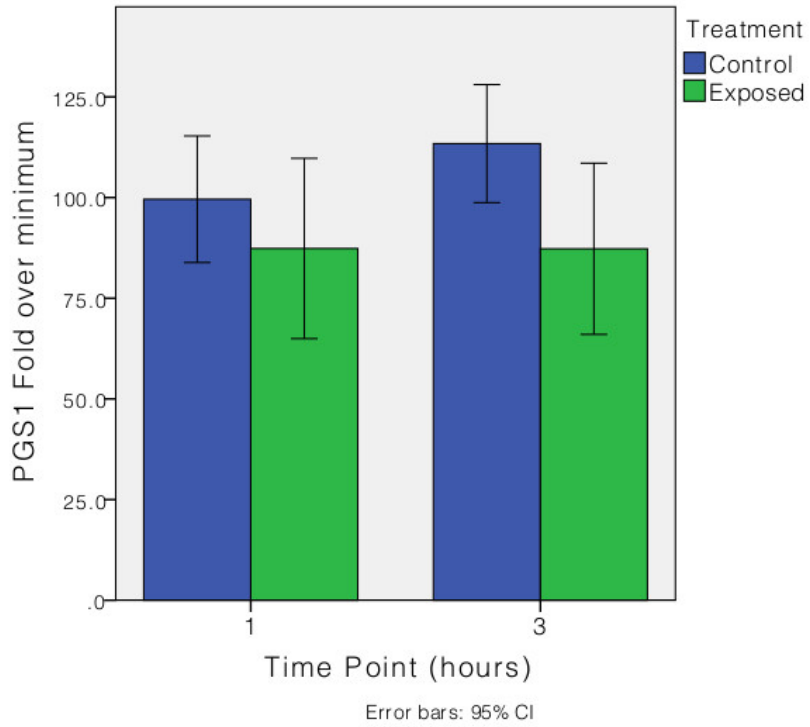


Figure 4. Messenger RNA expression levels of PGA1 in gill tissue of Pacific oysters challenged with Vibrio vulnificus (green bars) compared to controls (blue bars), one and three hours after exposure. Error bars represent 95% confidence intervals.



Works Cited

Drummond AJ, Ashton B, Buxton S, Cheung M, Cooper A, Heled J, Kearse M, Moir R, Stones-Havas S, Sturrock S, Thierer T, Wilson A. 2010 Geneious v5.3, Available from http://www.geneious.com

Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Lu F, Marchler GH, Mullokandov M, Omelchenko MV, Robertson CL, Song JS, Thanki N, Yamashita RA, Zhang D, Zhang N, Zheng C,Bryant SH. 2011. CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res* 39:D225-9.

Ricciotti E, FitzGerald GA. 2011. Prostaglandins and inflammation. *Arterioscler Thromb Vasc Biol* 31:986–1000.

Roberts S, Goetz G, White S, Goetz F. 2009. Analysis of genes isolated from plated hemocytes of the Pacific oyster, Crassostrea gigas. *Mar Biotechnol* 11:24-44.

Rozen S and Skaletsky HJ. 2000. Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) *Bioinformatics Methods and Protocols: Methods in Molecular Biology.* Humana Press, Totowa, NJ, pp 365-386
Source code available at http://fokker.wi.mit.edu/primer3/.

Appendix B: DNA methylation analysis in *Crassostrea gigas* using Nanostring nCounter Technology

Work accomplished in collaboration with Jeanette Nussbaum, Philippa Webster and Robin Lynn White from Nanostring Technologies.

Background

In collaboration with Nanostring, assays were developed to detect DNA methylation in oyster samples. This technology would allow for methylation detection in multiple genomic regions at a time in order to assess if DNA methylation patterns were dynamic between different developmental stages, tissue types or in response to different environmental exposures (see Table 1. for a description of samples). The assays combined methylation specific restriction enzymes MspI (methylation insensitive) and HpaII (methylation sensitive) with Nanostring's nCounter technology (http://www.nanostring.com). Fifty-six regions were targeted for analysis based on, 1) the presence of a restriction site (CCGG), 2) suitability of sequence surrounding the site for probe design (performed by Nanostring), and 3) the biological interest in the probe he following report summarizes the results from this collaboration. Supporting documents, including the collaboration document and a summary of the codeset design, can be found in Gavery & Roberts 2014.

Assay Date	Description of Samples
11/05/2011	Gill A (sample prepped for Phase I repeatability assay)
	Gill C (sample prepped for Phase I repeatability assay)
	Gill F (sample prepped for Phase I repeatability assay)
	Gill H (sample prepped for Phase I repeatability assay)
11/14/11	OA gill (no additional description)
	EE2 gill (96hr exposure to ethinylestradiol 10/27/11)
	Larvae (no additional description)
	Male gamete (no additional description)
12/19/11	Female gonad (vinclozolin exposed parent ID: 17)
	Male gonad (control parent ID: 19)
	5-azacytidine treated larvae (DNA isolated by SJW 6/8/10)
	Larvae (V. tubiashii exposed (ETS 8/23/10)
3/21/12	Female gonad (control parent ID: 28)
	Male gonad (control parent ID: 10)
	Larvae 155 (ETS 4/19/11)
	Larvae 159 (ETC 4/19/11)
4/23/13	Male gonad of control offspring (vinclozolin experiment) ID: 44
	Male gonad of control offspring (vinclozolin experiment) ID: 45
	Male gonad of control offspring (vinclozolin experiment) ID: 57
	Male gonad of control offspring (vinclozolin experiment) ID: 58
6/5/13	Male gonad of vinclozolin exposed offspring ID: 3
	Male gonad of vinclozolin exposed offspring ID: 4
	Male gonad of vinclozolin exposed offspring ID: 20
	Male gonad of vinclozolin exposed offspring ID: 21

Table 1. Description of samples assayed using the Nanostring codeset.

Methods

DNA extraction, enzyme digestion, nCounter analysis

DNA was isolated from genomic DNA and ~1 µg was submitted to Nanostring. Nanostring performed the enzyme digests. All digests include the AluI rescriction enzyme, which fragments DNA to sizes appropriate for the hybridization (determined *in silico* during probe design). There were three digests assayed for each sample: AluI alone, MspI + AluI and HpaII + AluI. The fragmented DNA was then used as input into the nCounter assay using the codeset designed for this project (see Gavery & Roberts, 2014 for a summary of the codeset). A

dilution series of positive controls (spike-in RNA hybridization controls (not oyster specific)) and negative controls (no template) were run in each assay. Nanostring provided raw and normalized count data. To generate the normalization factor (lane-specific), the sum of positive control counts were calculated for each lane and divided by the average of these sums (n=12 lanes per assay). This lane-specific scaling factor was then applied to all of the counts generated for that lane. Overall, the positive control scaling factors were all within the expected range (0.3 – 3). The exception is the 06/05/13 assay, where the positive controls failed to produce counts. Nanostring felt the data for this assay were still usable but it should be noted that the counts were not normalized.

Percent methylation of a particular probe was calculated as the HpaII +AluI counts/AluI counts x 100%. Percent methylation was only reported when probe validity criteria were met (see below).

Probe validity criteria

In order to report percent methylation for a probe, two criteria were required to be met. First, the AluI digest had to be >100 counts (or >200 for the 06/05/13 assay). Second, the percent background had to be less than 25%. Percent background is calculated as the ratio of the MspI counts/AluI counts. This is designed to control for the presence of the CCGG restriction site. If the counts of the MspI digested sample are not significantly lower than the undigested sample, it indicates that there was no restriction site present (or alternatively the outer cytosine is methylated and therefore MspI was unable cut at the restriction site). Figure 1 shows the proportion of assays that were valid for a given probe (total samples = 23) Total number of probes on the x-axis is 41. Figure 1 highlights the results that not all probes performed well across all samples. This is likely due to polymorphisms (e.g. no restriction site present or poor

hybridization of probes). There were a number probes that failed to meet validity criteria across all assays. Of the 56 targets, 15 probes consistently had very low counts. Therefore, these probes were removed from further analysis and a total of 41 target loci were evaluated for methylation status.

Results

A majority of the regions assayed for methylation were determined to be unmethylated across all samples tested regardless of tissue type, developmental stage or treatment (Figure 2).

Methylation was revealed by two probes (EU342886_1129 and EU342886_3306) in a gene annotated as hexokinase (NCBI: EU342886). The genomic region assayed by probe EU342886_1129 was unmethylated in gill tissue while heavy methylation is observed in both gonad and larvae samples (Figure 3). Initial verification of these results using targeted bisulfite sequencing in the same region as the Nanostring probe corroborated these findings (Figure 4). This initial verification was performed using a limited number of clones and would be strengthened by sequencing additional clones. The genomic region assayed by a second probe in the hexokinase gene (EU342886_3306) is also methylated (between 15 – 50% methylation) in samples where the probe met validity criteria. This hemimethylated state could be a result of differential methylation between individuals (larvae) or cell types (gonad). In other words, some cell types could be 100% methylated while others are 0% methylated. Alternatively, it may represent hemi-methylation of all the genomes present in the sample. A region of the phosducin gene (NCBI: EW779247), assayed by probe: EW779247_393, is constitutively methylated regardless of tissue type or developmental stage (Figure 6). A region of the tubulin tyrosine ligase-like family gene (GU207415), assayed by probe GU207415_8453 is methylated between 20 – 80% in samples where probe validity criteria were met (Figure 7).

Conclusions

This collaboration was an effort to apply the Nanostring nCounter technology to assess DNA methylation at targeted loci. The success of this effort was limited by the lack of a genome at the time of assay development. Probe selection was challenging, as very little gDNA sequence for *C. gigas* was publicly available. Instead, a majority of the sequence available was in the form of expressed sequence tag (EST) resources, meaning some probes may have crossed exon/intron boundaries. The lack of a genome also made it impossible to test for specificity of the probes. As a result, a number of the probes did not provide reliable results.

However, this collaboration did generate some interesting results. First, we identified the first differentially methylated CpG site between oyster tissues/life stages in the hexokinase gene, which is involved in glucose metabolism. We also identified a number of consistently unmethylated genomic regions as well as one constitutively methylated genomic region in the phosducin gene. Although this codeset would not be useful in the future due to the lack of informative sites (i.e. only 4 of 41 regions assayed were methylated), in the future, this technology could be used to develop a new codeset if probes were designed based on predefined regions of interest and tested for specificity using the *C. gigas* genome.

Works Cited

Gavery M, Roberts S. 2014. Nanostring. figshare. http://dx.doi.org/10.6084/m9.figshare.1109994. Retrieved 19:04, Jul 18, 2014 (GMT)

Figures

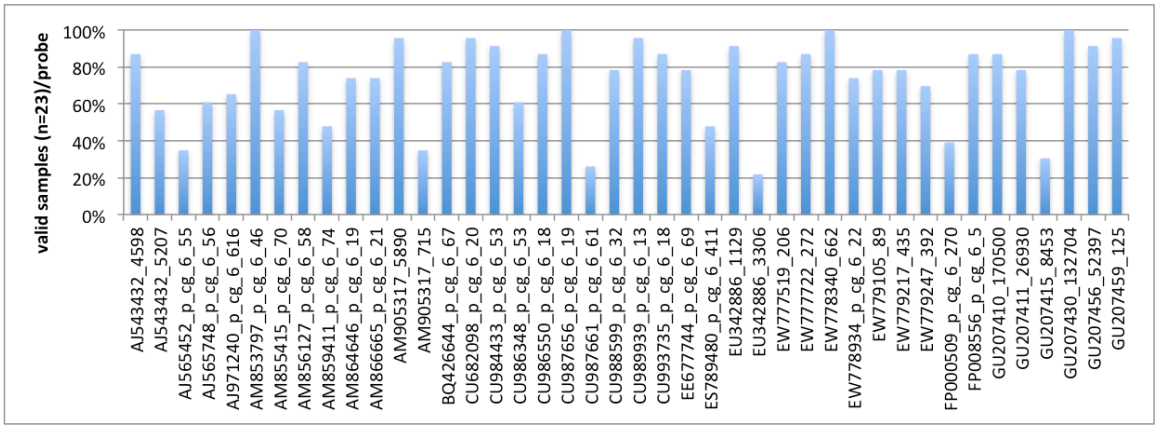


Figure 1. Probe validity summary (as percent valid samples). Probes showed variable performance across all 23 samples.

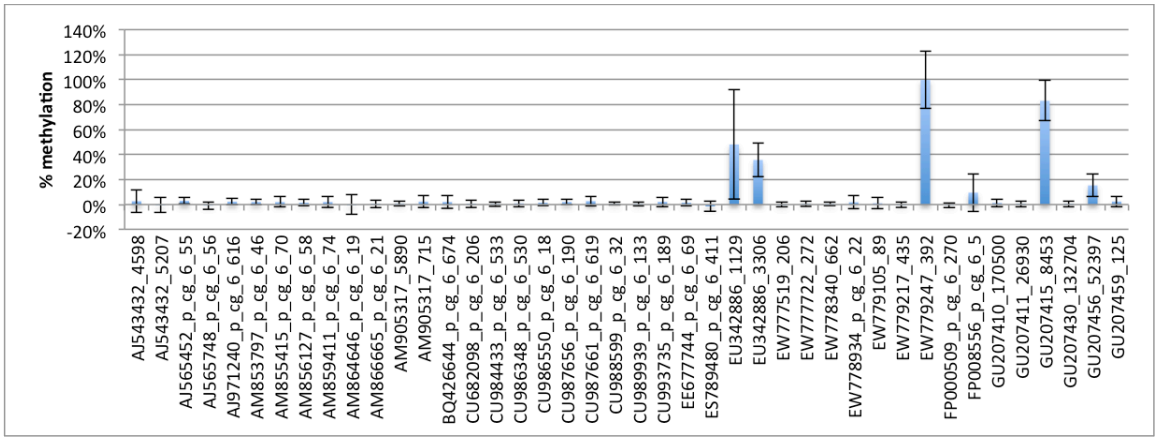


Figure 2. Percent methylation (+/- 1SE) for each probe across all samples (n=23). A majority of the probes are unmethylated regardless of sample type. DNA methylation (>20%) was detected in one or more of the samples for 4 of the 41 probes.

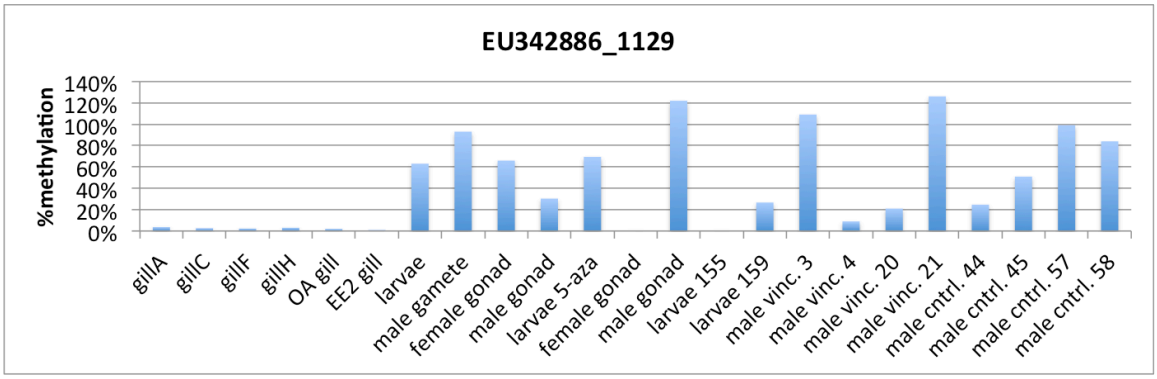
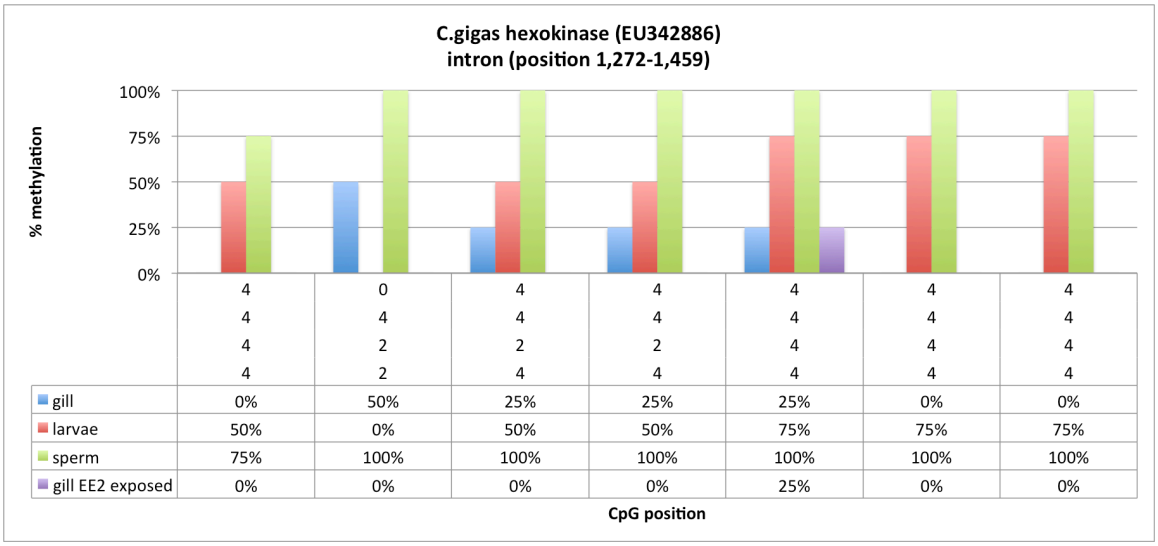


Figure 3. Percent methylation for the genomic region covered by probe EU342886_1129. Percent methylation is not reported for three samples: EE2 gill, female gonad and larvae 155, as probe validity criteria were not met.



	4	0	4	4	4	4	4
	4	4	4	4	4	4	4
	4	2	2	2	4	4	4
	4	2	4	4	4	4	4
gill	0%	50%	25%	25%	25%	0%	0%
larvae	50%	0%	50%	50%	75%	75%	75%
sperm	75%	100%	100%	100%	100%	100%	100%
gill EE2 exposed	0%	0%	0%	0%	25%	0%	0%

Figure 4. Bisulfite sequencing results for 7 CpGs in C. gigas hexokinase gene. The uppermost numbers in the legend represent the number of clones sequenced for each sample. The lower numbers represent the average percent methylation at each CpG position.

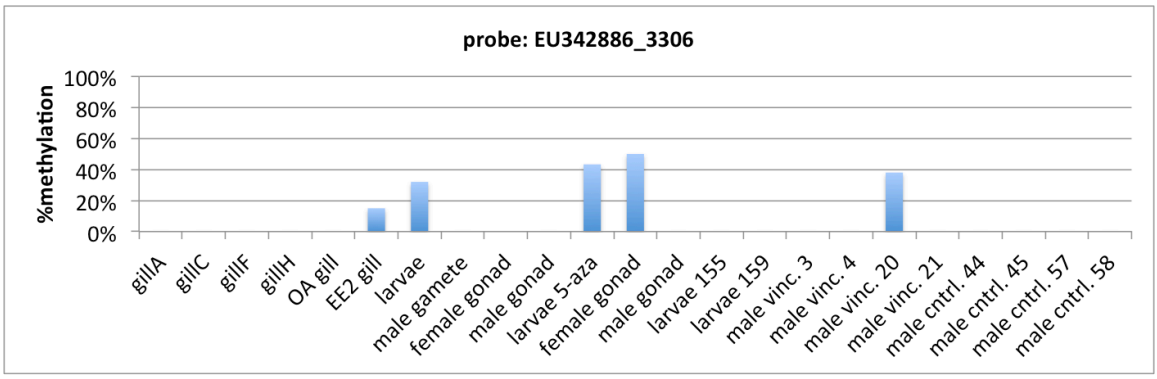


Figure 5. Percent methylation for the genomic region covered by probe EU342886_3306. Percent methylation was not reported for samples where blue bars are not visible, as probe validity criteria were not met.

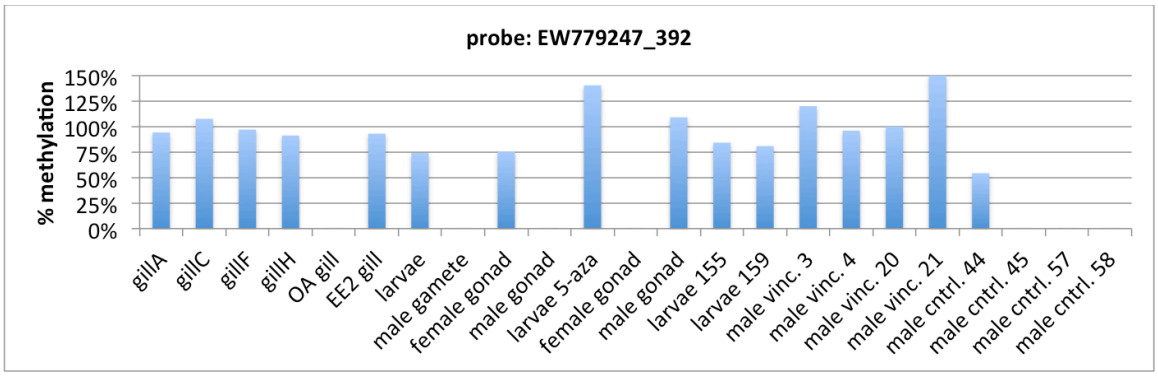


Figure 6. Percent methylation for the genomic region covered by probe EW779247_392. Percent methylation was not reported for samples where blue bars are not visible, as probe validity criteria were not met.

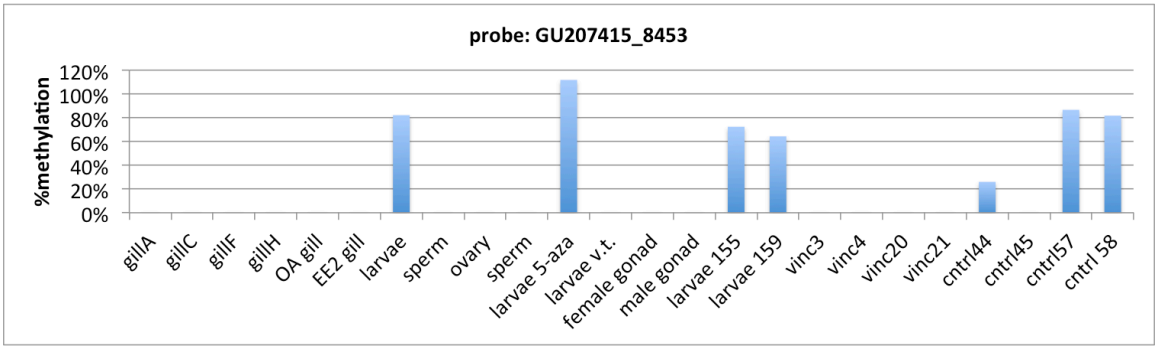


Figure 7. Percent methylation for the genomic region covered by probe GU207415_8453. Percent methylation was not reported for samples where blue bars are not visible, as probe validity criteria were not met.

Appendix C: Experiment to investigate transgenerational epigenetic inheritance in *Crassostrea gigas*

Background

Epigenetic patterns have been shown to be heritable in plants and mammals and in certain cases environmentally induced changes in these patterns can persist for multiple generations. Less work has been done to investigate transmissibility of epigenetic marks in invertebrates. Here, oysters were treated with 5-azacytidine, a chemical known to significantly decrease methylation levels, and the pesticide vinclozolin, which has been shown to transgenerationally alter DNA methylation patterns in mammals, in order to address the question of transmissibility in bivalves. An experiment was performed to assess the potential, stable transfer of DNA methylation patterns between exposed broodstock (F_0) and non-exposed F_1 generation offspring. Limited DNA methylation analysis was performed, but samples have been archived for future analysis.

Methods

Conditioning

Crassostrea gigas broodstock were obtained from Taylor Shellfish and conditioned in the laboratory. Broodstock were acclimated at 18° C then divided into three groups: 5-azacytidine treated (10 mg/L), vinclozolin treated (0.3 mg/L) and control (n=20 each group). Each group was placed into a static tank with 5 L seawater and aerated with air stones and fed twice daily. Temperature was raised by approximately 1° C every 4 days during treatment to induce gamete maturation. Treatment and control tanks were refreshed every third day.

Spawning and Sampling F_0

Oysters reached maturity and were strip spawned 16 days after treatment began. Single pair matings were performed (n=2 for both treatment and control groups). For the treatment replicates, sperm from the same male was crossed independently with eggs from two different females. For the 5-azacytidine group, a treated male was mated with control females as no mature females from the 5-azacytidine group were available. For the vinclozolin treated group, both the males and the female had been exposed to the pesticide. Tissues were sampled from parents (gill and gonad) for DNA isolation. See Table 1 for the identification numbers of the F_0 oyster from each cross.

On the day of spawning (7/1/10), egg counts were performed for each of the females. A total of 250,000 eggs were added to a beaker in a total of 250mL seawater. Sperm viability was estimated and between 2 - 6 mL of sperm was added to the eggs based on estimated motility (e.g. added 4 mL of sperm at approximately 50% motility). Fertilized gametes were added to 3.5 gallons of seawater. Larvae were filtered and fed every other day. F_1 Larvae were reared in static tanks at Taylor Shellfish Hatchery, set at day 21 (7/21/10) in an upwelling system, then placed in grow out bags as spat in Dabob Bay, WA.

Spawning and Sampling F_1

On 12/19/12, 1.5 years post spawn, F_1 oysters were collected and sampled. There were no surviving progeny from the 5-azacytidine crosses. Fifteen individuals each from the control and vinclozolin treated crosses were measured, sex determination was performed using light microscopy of the gonad and tissue samples were taken from gill gonad. All samples were taken

aseptically and stored immediately on dry ice prior to long-term storage at -80C. See Table 2 for a list of samples.

DNA methylation analysis

A limited number of individuals from F_0 and F_1 generation were analyzed for DNA methylation using the Nanostring assay (see Appendix B). No differences in methylation were observed. All remaining tissues are stored at -80C.

Discussion and Conclusions

This experiment was performed to determine if DNA methylation patterns could be transgenerationally inherited in oysters. The purpose of using 5-azacytidine and vinclozolin treatments was to disrupt normal DNA methylation patterns in oysters so that the changes could be tracked over multiple generations. One of the issues with working with the demethylating agent 5-azacytidine is that it can be cytotoxic. Only a very small proportion of the offspring of the 5-azacytidine treated oysters were viable indicating either that demethylation inhibits the production of quality gametes or that the dose of 5-azacytidine was toxic to the gametes. A very limited DNA methylation analysis was performed on the vinclozolin and control offspring and revealed no differences in the methylation patterns between these treatments. It is possible that a more thorough investigation of methylation in the parents and the offspring, perhaps using MBD-bisulfite sequencing, may reveal differences in methylation. Future efforts should be directed at performing a more through methylation analysis and also repeating the experiment with a lower dose of 5-azacytidine and following the crosses through to a second generation of offspring (F_2).

Tables

Table 1. Identification of parents used in crosses

Treatment	Male ID (family 35)	Female ID (family 51)
Control	19	30
		31
5-azacytidine	2	28*
		32*
Vinclozolin	10	11
		17

*females were from the control group not treated with 5-azacytidine.

Table 2. Sizes and sex determination of sampled F_1 individuals at 18 months (12/19/12)

ID	treatment	tag	gill sample ID	gonad sample ID	length	width	whole mass	sex (M/F)
1	vinclozolin (F1)	10 x 11	1.gi	1.go	74.8	52.2	51.23	M
2	vinclozolin (F1)	10 x 11	2.gi	2.go	74.8	46.5	37.78	?
3	vinclozolin (F1)	10 x 11	3.gi	3.go	78.5	46.5	45.93	M
4	vinclozolin (F1)	10 x 11	4.gi	4.go	76.8	47.9	50.74	M
5	vinclozolin (F1)	10 x 11	5.gi	5.go	78.7	48.7	43.49	F
6	vinclozolin (F1)	10 x 11	6.gi	6.go	68.4	49	37.36	M
7	vinclozolin (F1)	10 x 11	7.gi	7.go	61.7	39.9	29.96	M
8	vinclozolin (F1)	10 x 11	8.gi	8.go	64.8	43.9	32.25	F
9	vinclozolin (F1)	10 x 11	9.gi	9.go	50.3	39.4	23.99	F
10	vinclozolin (F1)	10 x 11	10.gi	10.go	49.6	36.3	18.25	M
11	vinclozolin (F1)	10 x 11	11.gi	11.go	58.6	43.1	22.22	M
12	vinclozolin (F1)	10 x 11	12.gi	12.go	57.9	41.8	25.13	F
13	vinclozolin (F1)	10 x 11	13.gi	13.go	60.1	43.7	42.38	?
14	vinclozolin (F1)	10 x 11	14.gi	14.go	65.5	53.5	38.61	F
15	vinclozolin (F1)	10 x 11	15.gi	15.go	67.7	44.4	33.53	M
16	vinclozolin (F1)	10 x 17	16.gi	16.go	49	35	20.18	F
17	vinclozolin (F1)	10 x 17	17.gi	17.go	70.3	36.9	29.68	F
18	vinclozolin (F1)	10 x 17	18.gi	18.go	62.7	47.2	32.77	M
19	vinclozolin (F1)	10 x 17	19.gi	19.go	56.9	33.4	21.27	?
20	vinclozolin (F1)	10 x 17	20.gi	20.go	68.5	37.7	29.2	M
21	vinclozolin (F1)	10 x 17	21.gi	21.go	63	38.8	24.71	M
22	vinclozolin (F1)	10 x 17	22.gi	22.go	55.6	44.7	27.07	M
23	vinclozolin (F1)	10 x 17	23.gi	23.go	58.2	42.3	38.97	M
24	vinclozolin (F1)	10 x 17	24.gi	24.go	56.9	36.8	22.65	M
25	vinclozolin (F1)	10 x 17	25.gi	25.go	54.8	31	14.61	M

26	vinclozolin (F1)	10 x 17	26.gi	26.go	40	29.6	10.81	M
27	vinclozolin (F1)	10 x 17	27.gi	27.go	63.1	40.9	23.62	M
28	vinclozolin (F1)	10 x 17	28.gi	28.go	78.6	48.9	51.98	M
29	vinclozolin (F1)	10 x 17	29.gi	29.go	48.9	30	15.91	M
30	vinclozolin (F1)	10 x 17	30.gi	30.go	52.8	37.5	23.1	M
31	control (F1)	19 x 31	31.gi	31.go	69.8	48.5	41.62	M
32	control (F1)	19 x 31	32.gi	32.go	58.5	47	37.61	?
33	control (F1)	19 x 31	33.gi	33.go	72	50.8	52.98	?
34	control (F1)	19 x 31	34.gi	34.go	66	52.1	36.19	?
35	control (F1)	19 x 31	35.gi	35.go	63.6	48.5	39.39	M
36	control (F1)	19 x 31	36.gi	36.go	80.5	54.2	54.59	F
37	control (F1)	19 x 31	37.gi	37.go	60.7	50.5	50.44	M
38	control (F1)	19 x 31	38.gi	38.go	70.5	50.2	57.73	F
39	control (F1)	19 x 31	39.gi	39.go	76.5	50.1	65.66	F
40	control (F1)	19 x 31	40.gi	40.go	77.5	65.4	57.18	F
41	control (F1)	19 x 31	41.gi	41.go	73	60	51.04	F
42	control (F1)	19 x 31	42.gi	42.go	62.2	46.1	35.98	F
43	control (F1)	19 x 31	43.gi	43.go	70.5	52.5	42.71	F
44	control (F1)	19 x 31	44.gi	44.go	63	47	32.79	M
45	control (F1)	19 x 31	45.gi	45.go	62.5	39	33.76	M
46	control (F1)	19 x 30	46.gi	46.go	67.9	51	45.26	F
47	control (F1)	19 x 30	47.gi	47.go	71.1	41.3	37.73	F
48	control (F1)	19 x 30	48.gi	48.go	85.1	47.2	56.89	?
49	control (F1)	19 x 30	49.gi	49.go	89	60.3	70.28	F
50	control (F1)	19 x 30	50.gi	50.go	69.5	43.1	42.24	M
51	control (F1)	19 x 30	51.gi	51.go	90	50.2	64.79	F
52	control (F1)	19 x 30	52.gi	52.go	57.8	50.9	29.56	M
53	control (F1)	19 x 30	53.gi	53.go	73.6	48.6	36.5	F
54	control (F1)	19 x 30	54.gi	54.go	72.7	42.4	39.35	?
55	control (F1)	19 x 30	55.gi	55.go	85.4	51	62.58	?
56	control (F1)	19 x 30	56.gi	56.go	77.8	43.8	39.31	?
57	control (F1)	19 x 30	57.gi	57.go	67	45.9	38.71	M
58	control (F1)	19 x 30	58.gi	58.go	69.2	56.2	51.09	M
59	control (F1)	19 x 30	59.gi	59.go	73.4	46.4	40.38	M
60	control (F1)	19 x 30	60.gi	60.go	79.5	56.3	50.27	M

M=male, F=female, ?=unable to determine sex