1. INTRODUCTION

Animal production plays a central role in the sustainability of the US agriculture industry. Maintaining and improving animal production efficiency requires an understanding of the genetic and physiological mechanisms that control desired traits. Already, an understanding of these mechanisms has lead to the development of pioneering biotechnological methods that have important applications in the food industry. For example, molecular markers have been adopted for use in broodstock selection and transcriptomic studies have been used to improve environmental conditions to decrease physiological stress in animals.

Recently, there has been an expanded interest in understanding the role of epigenomic variation in commercially important traits. This interest stems at least partly from human based genome-wide association studies and ecological investigations where genetic differences do not completely explain phenotypic diversity. Similarly, the completion of the human genome project has failed to meet expectations to successfully identify causes and cures to major diseases. There is also evidence that epigenetic mechanisms are directly responsible for commercially important traits. In sheep, a mutation in the *callipyge* gene results in muscle hypertrophy only in heterozygous animals where the mutated allele is of paternal origin, indicating an epigenetically controlled imprinting mechanism (Charlier et al. 2001). Another example has been described in Texel sheep, where a mutation in the *myostatin* 3' untranslated region interferes with miRNA binding (an important epigenetic process) that ultimately results in the decreased expression of this potent negative regulator of muscle growth (Clop et al. 2006).

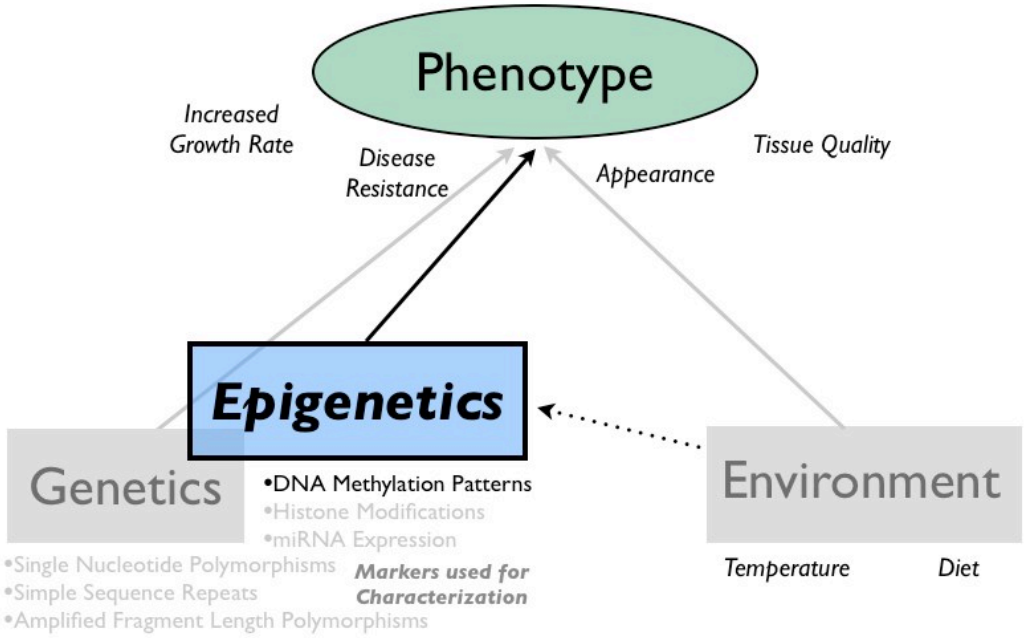

Figure 1. Schematic representation of epigenetic influence on phenotype. Markers used for characterizing organismal variation are listed.

A second reason for the increased interest in epigenomic variation as it relates to commercially important traits, is a result of advances in technology that have improved researchers' ability to accurately identify and describe epigenetic features. This includes high-throughput sequencing technologies (e.g. Applied Biosystems SOLiD) that have facilitated genome wide characterization of epigenetic marks (histone modification, DNA methylation) when coupled with other approaches such as chromatin immunoprecipation and bisulfite treatment (e.g. Johnson et al 2007; Zeschnigk et al 2009). Specialized sample preparation techniques have also facilitated the use of hybridization array based epigenetic characterizations (e.g. Guerro-Bosagna et al 2010).

A primary goal of the proposed project is to ***characterize DNA methylation patterns*** in a commercially important aquaculture species, the Pacific oyster (*Crassostrea gigas*). In addition, we will ***investigate how DNA methylation contributes to phenotypes associated with inbreeding depression and hybrid vigor***. In the remainder of this section, background information concerning DNA methylation is provided, including results from preliminary research on the oyster. As there is limited knowledge concerning DNA methylation in taxa closely related to oysters, much of the following information comes from mammalian studies. Inbreeding depression and hybrid vigor will also be discussed as it relates to the Pacific oyster.

1.1 Epigenetics and DNA methylation

The term epigenetics was coined by Waddington in 1942 and referred to how genes might interact with other factors to contribute towards phenotype (Waddington 1942). Scientific discoveries, along with advances in technology in the past half century, have allowed for a better understanding of what these factors are. While there is some debate as to what the term "epigenetics" refers to, for this proposal we will refer to epigenetics as mechanisms capable of inducing changes in genetic activity without altering the underlying DNA sequence (Jablonka & Lamb, 2002). DNA methylation, histone modifications and non-coding RNA activity (e.g. miRNA) are the most commonly described epigenetic mechanisms. DNA methylation is one of the most studied mechanisms of epigenetic regulation and refers to the addition of a methyl group to position 5 of cytosines. In animals, the addition of methyl groups to cystosine-5 occurs almost exclusively in CpG dinucleotides. DNA methylation is typically associated with transcriptional repression. This methylation, and subsequent repression of transcription, frequently occurs in gene promoters (Boyes & Bird, 1992, Kass et al. 1997, Hsieh 1994). The functional significance of DNA methylation in vertebrates is wide in scope and includes regulation of imprinted genes (Bell & Felsenfeld 2000) and X-chromosome inactivation (Csankovszki et al. 2001). In mammals, DNA methylation is essential for development (Okano et al. 1999) and defects or unintended changes can have deleterious consequences such as embryonic lethality (Li et al. 1992) and tumorgenesis (Jones & Baylin 2007). In addition, DNA methylation is one mechanism that contributes to cellular differentiation

across cell-types despite the fact that all cells in an organism possess the same DNA. 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 (Anway & Skinner 2006).

1.2 DNA methylation patterns across taxa

The extent of cytosine methylation varies considerably among eukaryotes. In vertebrates, approximately 70-80% of cytosines in 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). One taxonomic group that has been significantly 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 CpG methylation in the clam, *Donux truculus* (Petrović et al. 2009).

In addition to differences in the extent of methylation among different phyla, there are also differences in the distribution of methylation. Methylation appears to be confined to intragenic regions in invertebrates (Suzuki et al. 2007, Elango & Yi 2008), whereas methylation is found globally across vertebrate genomes (Ehrlich et al. 1982). It has been suggested that this may impart some functional significance for differences in regulatory processes.

1.3 Transcriptional Regulation

Regulation of transcription by DNA methylation is accomplished primarily at gene promoters in vertebrates (Boyes & Bird 1992, Kass et al. 1997, Hsieh 1994). In contrast, methylation appears to be targeted specifically to transcription units in invertebrates (Suzuki et al. 2007). Some of the first evidence supporting a regulatory role of intragenic DNA methylation in invertebrates comes from computational analyses of the methylation status of *A. mellifera* genes (Elango et al. 2009, Petrović 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*, however, it is uncertain if this function is conserved across invertebrate taxa. Furthermore, it is unclear exactly how intragenic cytosine methylation directly affects transcription. One hypothesis is that it prevents inappropriate initiation of transcription outside of promoter regions (Bird 1995).

1.4 Imprinted genes and development

Classically imprinted genes (also referred to as epialleles) are expressed from either the paternal or maternal allele. DNA methylation is the primary mechanism regulating the expression of imprinted genes. In mammals, there are approximately 80 genes that are imprinted (Morison et al. 2005, Wood & Oakey 2006). One of the most well-studied examples of this is the *insulin growth factor 2* gene where the maternal allele is silenced by promoter methylation (DeChiara et al. 1991, Feil et al. 1994). In addition to imprinted genes that are expressed strictly on the basis of parental origin, there is also another class of imprinted genes referred to as metastable alleles that are not strictly dependent on the parent of origin. One example includes the Avy allelle that has an intracisternal A particle (IAP) insertion upstream of the *agouti* gene (Duhl et al. 1994). A cryptic promoter in the IAP can induce expression of Agouti, resulting in mice with yellow coats (Duhl et al. 1994). While characteristics of imprinted genes in invertebrates are not completely understood, a characterization of DNA methylation patterns in oysters would provide important tools to study the phenomenon.

In mammals, most imprinted alleles are reset during development. This reprogramming (i.e. demethylation) occurs at two specific times during development. One reprogramming occurs in primordial germ cells (PGCs) once they have reached the embryonic gonad (E10.5 to E13.5) (Sasaki & Matsui 2008) and the second occurs just following fertilization, when paternal DNA is demethylated (Oswald et al. 2000, Mayer et al. 2000). This process continues during cell division passively as DNMT1 is excluded from the nucleus (Howell et al. 2001). However, differentially methylated regions in imprinted genes are protected from demethylation in the zygote.

During both demethylation events in developing mammals, recent evidence suggests that active demethylation is occurring through processes such as base excision repair and DNA deaminase activity (Popp et al. 2010, Hajkova et al. 2010). It is not known if oysters and similar species undergo a reprogramming. It is beyond the scope of the current proposal to examine this process in detail, in oysters. However it should be noted this phenomenon has significant implications regarding the degree of inheritance of DNA methylation patterns.

1.5 DNA methylation in the Pacific Oyster – (Preliminary Research Results)

In order to better understand the role of DNA methylation the Pacific oyster, we recently characterized DNA methylation patterns, primarily using an *in silico* approach (Gavery & Roberts 2010). This work suggests that intragenic methylation is important for regulating gene expression in oysters. Specifically, the predicted methylation status of *C. gigas* genes were characterized based on the known hyper-mutability of methylated cytosines, which readily deaminate to thymine residues (Coulondre et al. 1978). 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 (Schorderet & Gartler 1992). Consequently,

regions of DNA with a low CpG observed:expected ratio (o/e) are predicted to be methylated, whereas regions with a high CpGo/e (approaching 1.0) are predicted to be unmethylated. The results of this analysis indicate that *C. gigas* genes predicted to be hyper-methylated are generally associated with housekeeping functions and those predicted to be hypo-methylated are associated with immune related processes (Figure 2). One hypothesis is that genes predicted to be hypo-methylated have greater epigenetic flexibility, which allows for higher regulatory control of these inducible classes of genes. Oysters have been shown to have high phenotypic plasticity in response to environmental changes and stress (Hamdoun et al. 2003, Honkoop et al. 2003) and it is possible that an epigenetic mark, such as DNA methylation, could provide this level of control. This finding is similar to research on *A. mellifera* where genes associated with general housekeeping functions were predicted to be hyper-methylated, whereas caste-specific genes were preferentially hypo-methylated (Elango et al. 2009, Foret et al., 2009).

In addition to the *in silico* based approaches, we have begun to characterize DNA methylation in the Pacific oyster using gene-specific bisulfite sequencing to identify DNA methylation in a select number of genes. Initially, five genes predicted to be hyper-methylated and five predicted to be hypo-methylated (based on CpGo/e) were selected for analysis. Valid PCR products were produced for only two of the

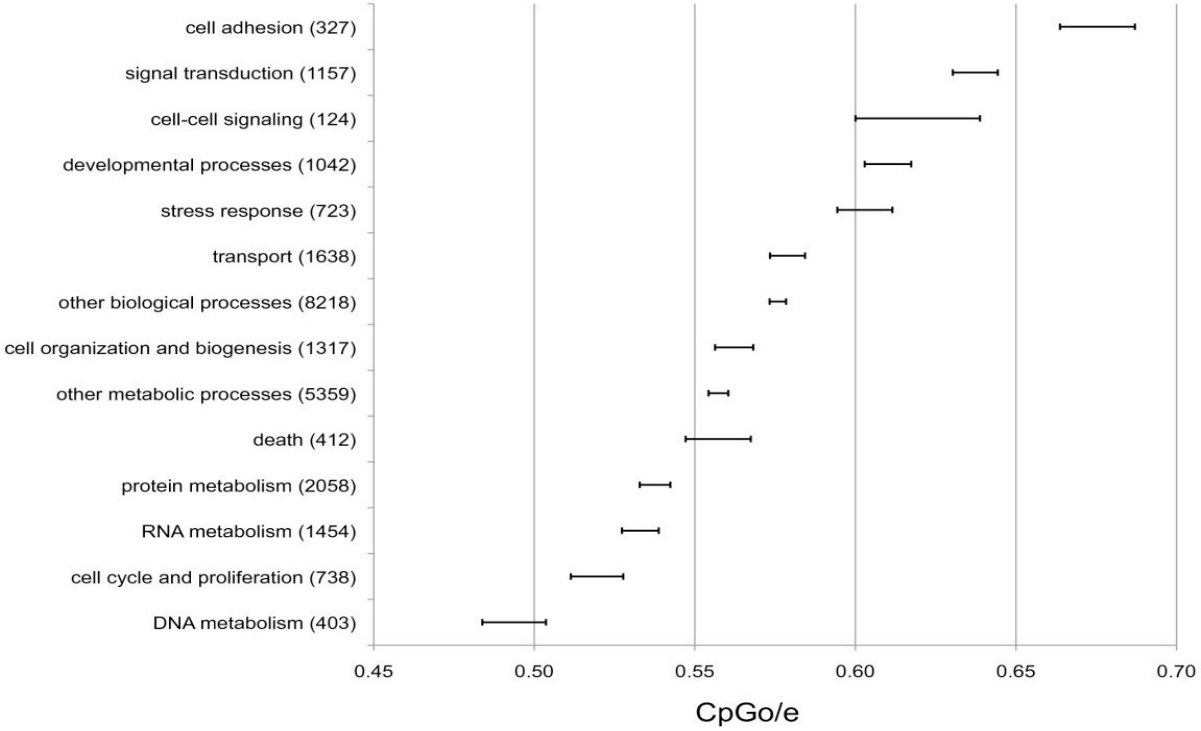

Figure 2. Differential methylation between categories of genes involved in discrete biological processes as measured by CpGo/e. Mean CpGo/e for 10,699 *C. gigas* genes (*i.e.* contigs from ESTs) categorized by GO Slim term. Bars represent mean ± 1 standard error. The number of contigs in each category is listed in parenthesis. Modified from Gavery and Roberts 2010.

genes. This is a typical result as the conversion of unmethylated cytosines results in challenges for primer specificity. In the first target, one of seven CpGs sites displayed methylation in 25% of the clones sequenced and in the second target, one of two CpGs sites was determined to be methylated in 50% of the clones (Gavery & Roberts 2010).

More recently we have been attempting to characterize the specific DNA methylation pattern in heat shock cognate 70 (hsc70) (GenBank Accession number AJ305315) (unpublished data) using nested primers to increase specificity. Figure 3 illustrates a portion of the gene and the respective DNA methylation pattern based on PCR amplifying, cloning, and sequencing bisulfite-treated DNA from different regions of the gene. Chromatogram files are then assembled onto the respective reference sequence to determine methylation status.

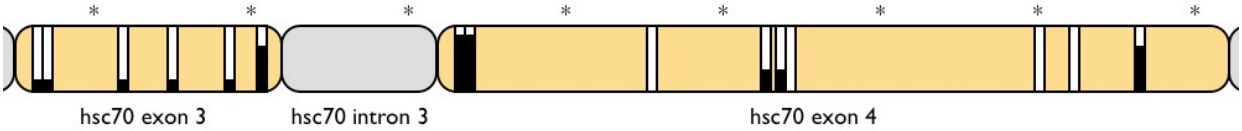

Figure 3. DNA methylation pattern in a portion of the Pacific oyster hsc70 gene. Vertical bars indicate CpG residues and shading represents percent of clones where cytosine was determined to be methylated, based on bisulfite treatment. Asterisks are located every 100bp. Width of bars representing CpG residues are not to scale.

Currently, we are in the process of sequencing two methylation enriched DNA libraries using the SOLiD 4 high-throughput sequencing platform (Applied Biosystems). Here, genomic DNA is randomly sheared, and methylated fragments enriched by preferential binding to methyl-CpG binding domain of human MBD2 protein (MBD). Based on DNA quantification following the MBD enrichment, approximately 22% of the Pacific oyster genome is methylated. With the sequencing of these libraries we expect to corroborate our *in silico* analysis (Figure 2). This approach cannot, however, provide the base pair resolution that can be obtained with bisulfite treatment. As part of this research proposal we are planning to couple MBD enrichment and bisulfite treatment to obtain increased resolution of DNA methylation patterns (Described in detail below- Section 3). This approach (which will be referred to as MBD-BS-Seq) will be used to accomplish our research objective to ***characterize DNA methylation patterns in larval and adult oysters***. The second research objective is to ***investigate how DNA methylation contributes to phenotypes associated with inbreeding depression and hybrid vigor***.

1.6 Hybrid Vigor

Hybrid vigor, or heterosis, is a well-known phenomenon where hybrid progeny display superiority compared with their parents (Shull 1908). Heterosis is agriculturally important and advantages in yields as a result of hybrid crops can range between 15% and 50% (Duvick 1999). Therefore, understanding the mechanisms behind hybrid vigor, and its inverse, inbreeding depression, are important. The major genetic models proposed to explain heterosis include dominance, overdominance, and epistasis (Crow 2000, Lippman & Zamir 2007). However, these genetic models have not been completely satisfactory in describing these phenotypes and studies in the same species can support varying results regarding the mechanism of heterosis (reviewed by Lippman and Zamir, 2007). Recent studies into the physiological basis of heterosis, including transcriptomic analyses, indicate regulatory mechanisms may be responsible. In hybrids, novel patterns of gene expression resulting from the combination of allelic variants are thought to be involved in heterosis (Birchler et al. 2003, Swanson-Wagner et al. 2006, Springer & Stupar 2007). One hypothesis that is beginning to gain support is that ***epigenetics may regulate genes associated with the increased growth observed in hybrids of agriculturally important species.*** For example, a recent study showed that differential expression of a small number of epigenetically regulated genes is associated with growth in hybrid *Arabidopsis thaliana* (Ni et al. 2009). In pigs, differential DNA methylation patterns were observed in hybrid offspring compared to parents (Cao-de et al. 2005). There have been similar findings in the potato (Nakamura & Hosaka 2009, corn (Zhao et al., 2007), and rice (Xiong et al.,1999). With increased understanding, it is possible that desired methylation patterns associated with hybrid vigor could be selected for.

Hybrid vigor has been examined in oysters for more than a decade. Researchers initially observed non-additive components of genetic variance for larval survival (Lannan 1980) and body size at harvest (Hedgecock et al. 1991) which led to interest in crossbreeding inbred lines in an effort to increase performance. As a result of these efforts, heterosis was directly demonstrated for growth and survival in Pacific oysters by Hedgecock et al. (1995). Subsequently, investigation into the mechanism behind this observed heterosis provided evidence of a high genetic load in Pacific oysters, supporting the dominance theory (Launey & Hedgecock 2001). Dominance theory proposes that independent sets of deleterious alleles accumulate throughout the genome during inbreeding of parental lines and that hybridization causes complementation in F1 progeny by the superior alleles, resulting in phenotypic superiority over both parents (Bruce 1910).

More recently, studies are beginning to look at physiological mechanisms for heterosis. For example, hybrid oysters show increased feeding rates and efficiencies relative to inbred equivalents (Bayne et al. 1999, Pace et al. 2006). A study looking at transcriptomic differences between hybrid and inbred crosses identified approximately 350 genes that showed non-additive expression in the hybrids (Hedgecock et al. 2007). The regulatory mechanisms underlying these differences in gene expression are not understood, but based on research mentioned previously, it

is possible that epigenetic effects, including changes in DNA methylation patterns, could be involved. In the current study we will use hybrid and inbred lines of Pacific oysters to investigate the role of DNA methylation in hybrid vigor. This will be carried out using a combination of methylated DNA enrichment and a tiling array approach described in Section 3.

2. RATIONALE AND SIGNIFICANCE

The research objectives proposed here will increase our understanding of the animal at the molecular and cellular level to provide the foundation to improve production in shellfish. Epigenetic processes are critical mechanisms responsible for regulating all aspects of animal physiology, including those that contribute to commercially important traits. Improving our fundamental knowledge will provide a framework to improve production by assisting selection practices and improving management strategies. Epigenetic marker assessment could be incorporated into ongoing marker-assisted selection programs. More importantly, epigenetics will likely provide explanations of desired phenotypes including disease resistance, reproductive control, and hybrid vigor. As described, a better understanding of hybrid vigor will significantly impact the larger agricultural community, including plant species.

While the project will have impacts on other species, it directly targets oysters, which are a key cultured shellfish species in the US with production estimated at over 40,000 tonnes (WRAC, 2004; FAO 2006). In 2010, over 3500 tonnes of oysters were exported from the United States at a value of over $22 million (USDA Economic Research Services). There are substantial farming operations along the East Coast, West Coast and Gulf of Mexico, which contribute significantly to the economies across the US. It is estimated that close to $100 million is contributed to the economies of these regions annually (WRAC, 2004).

Given the nature of crop outplanting (adult oysters are usually grown in cages in coastal waters), oysters serve as an important ecosystem service and are an excellent example of an environmentally sustainable product. 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; Rice 1999, Officer et al. 1982). Nutrients such as nitrogen and phosphorus that are not incorporated into oyster tissue are excreted and can then be utilized by keystone plants such as eelgrass (Newell 2004). Oysters commonly grow, and are farmed, in aggregations that have been shown to enhance biodiversity (Ferraro and Cole 2007). Such structures are integral not only for larval oysters, but for other organisms such as worms, snails, crabs, fish and birds that utilize the structures as habitat. Like other shellfish, oysters are also carbon fixers, incorporating carbon into their shell thereby helping reduce atmospheric carbon dioxide levels (Newell 2004).

The proposed project directly addresses the program priority to target Animal Breeding, Genetics, and Genomics. This project will *generate new genome sequences as well as refine the relatively limited amount of existing genome sequence* for this species by increasing coverage. In sum, it is estimated that we will generate 80 gigabases of sequence data. Furthermore, this project will *describe epigenomic variation, including its relevance to a phenotype that directly contributes to improved animal production, hybrid vigor*.

3. APPROACH

DNA methylation is a fundamental biological process that is increasingly recognized as having an important role in controlling phenotypic traits. In the Pacific oyster, we have provided evidence that genes with differing regulatory requirements have different levels of DNA methylation (Gavery and Roberts 2010). It still remains to be determined precisely what portion of the genome is methylated, how the methylation patterns vary between cell types, if DNA methylation patterns change over the lifetime of the organism and if epigenetic processes influence phenotypes associated with increased efficiency of production. To begin to address these points, the specific research objectives of the current proposal are to:

1) ***Characterize DNA methylation patterns across the oyster genome***

2) ***Investigate how DNA methylation contributes to phenotypes associated with inbreeding depression and hybrid vigor***

The first research objective will provide a map of DNA methylation patterns in larval and adult oysters that will characterize the epigenomic variation between these two developmental time points as well as at the individual gene level. This will allow us to validate our *in silico* analysis (Figure 2) and provide direct insight into molecular process by which DNA methylation acts in invertebrate species.

As described, DNA methylation in invertebrates is significantly different from vertebrates. Hypotheses regarding how intragenic methylation may regulate transcription include passive processes such as the prevention of inappropriate initiation of transcription outside of promoter regions (Bird, 1995) and also more active roles in transcriptional regulation through associations with non-coding RNAs. While our investigations to date are very limited, it is intriguing that both exons examined in the *HSC70* gene have a pair of methylated CpGs at the start of the exon. Large-scale interrogation of DNA methylation across the genome will certainly elucidate patterns that will provide insight into the basic role of these epigenetic marks.

Completion of the second research objective will provide a greater understanding of how DNA methylation varies between inbred lines (slow growth rates) and crosses of two different inbred lines which have been shown to have a significantly higher growth rate than either of the parental lines (Hedgecock et al.

1995). Researchers examining heterosis in other agriculturally important species have found correlations with DNA methylation patterns as previously described. Given the importance of hybrids in agriculture, the results of this project will certainly contribute to increased production efficiency for other species. The remainder of the section provides a description of the activities associated with each research objective, including the expected outcomes and potential limitations. This is followed by detailed description of the methods and a project timeline.

3.1 *Objective 1* - Characterize DNA methylation patterns the oyster genome

In order to carry out the first research objective, a MBD-BS-Seq approach will be used. DNA will first be subjected to methylation enrichment by preferential binding of methylated fragments to the methyl-CpG binding domain of human MBD2 protein. Samples will then be treated with sodium bisulfite to deaminate cytosine, but not methylcytosine, to uracil. Libraries will then be sequenced using the ABI SOLiD 4 sequencing platform. The key advantage to this procedure is that the methylation enrichment allows us to increase sequence coverage necessary for sequence read assembly and mapping. The SOLiD libraries that we have already constructed (sequencing in progress; data available April 2011) will also assist in the analysis (see below and Figure 4). Furthermore, irrespective of the bisulfite treated libraries, this methylation enrichment approach will allow us to empirically determine the relative degree of methylation in different functional categories of genes.

An alternative approach to methylation enrichment would be to immunoprecipitate with a monoclonal antibody that specifically binds 5-methylcytidine (MeDIP). However, MBD approaches have advantages over the MeDIP procedures with respect to downstream sequencing. First, the end product in MeDIP is single stranded DNA, which requires increased sample manipulation to make double-stranded DNA required for library preparation. MBD enrichment does not require this step. Second, MBD enrichment allows for fractionation based on methylation density to generate pools of densely methylated, moderately methylated and unmethylated fractions. This can be beneficial by increasing coverage of the targeted fraction.

Hemocytes from adult oysters will be analyzed in the fashion described above. Hemocytes were selected for analyses based on their function in the immune system, as evidence suggests immune related genes may be regulated by DNA methylation (Gavery and Roberts 2010). Examining larval samples in addition to adult hemocyte samples will provide insight into the degree of differences in DNA methylation across life history.

3.1.1 Experimental Design - *Objective 1*

For adult Pacific oyster samples, a hemolymph sample will be taken from one individual, hemocytes isolated, and DNA extracted. For larval samples, DNA will be extracted from a pool of whole body veliger larvae (Figure 4). Methylation enrichment will be performed for each sample using the Methyl-Miner Kit (Invitrogen). Two fractions will be obtained: one analogous to the highly methylated fraction and a

second, a moderately methylated fraction. For each fraction, both a bisulfite treated and non-treated portion will be used to generate two separate SOLiD libraries. In total, 8 libraries will be sequenced (Figure 4). Sequencing reads will be quality trimmed and a genomic scaffold will be developed using the combined non-bisulfite treated libraries, as well as publically available sequence data. Once this scaffold is constructed, the quality-trimmed sequence reads from the bisulfite treated libraries will be mapped back to the scaffold with remaining cystosines representing methylcytosines and non-methylated cytosines are converted to thymines. The lab of the PI has extensive experience analyzing short-read sequencing data as well as working with bisulfite treated sequence products. See Section 3.3 for detailed methods.

3.1.2 Expected Outcomes - *Objective 1*

Based on our preliminary research results, and what has been observed in other species, our expected results from the MBD-BS-Seq component of the project are: a) methylated cytosines will be exclusively located in exonic portions the genome, b) overall, approximately 30-60% of the genomic sequence will be methylated, c) methylation will be correlated with the functional role of the corresponding transcript (see Figure 2), d) varying levels of methylation will be observed at a given CpG residue, and e) we will observe significantly different methylation patterns between hemocytes and whole body larval samples. The rationale for expecting varying levels of methylation within hemocyte samples and between adult (hemocyte) and larval samples is based on the understanding that all cells in an organism have the same DNA and epigenetic processes regulate transcriptional activity essential for developmental and diverse functional processes. Therefore, when a sample is taken for DNA isolation it will possess different cell types that will likely have different methylation patterns. It is expected that there will be much less variation in the hemocyte libraries as a limited number of cell types are found in hemolymph (i.e. granulocytes, hyalinocytes).

3.1.3 Potential Pitfalls - *Objective 1*

One of the greatest challenges in the proposed research is related to the fact that the Pacific oyster genome has not been completely sequenced. A completed genome greatly improves the ability to assemble and analyze data. For this study we are taking steps to ensure we are able to effectively analyze the sequencing reads generated as part of the MBD-BS-Seq approach. Specifically, MBD enriched DNA samples will serve as a form of genome reduction to allow us to obtain enough sequence coverage. The oyster genome is 824 megabases (Hedgecock et al. 2005) and based on some preliminary MBD procedures in the lab it is estimated that approximately 30% of the genome is methylated. We plan to sequence an octet on the SOLiD 4 system (UW HTGU) that is expected to generate close to 12 gigabases (ABI Specification Sheet). The estimated coverage would therefore be over 50x. This an exaggerated estimation as it does not take into consideration low quality data,

sequencing errors, and sub-optimal recoveries. However, based on handling similar types of data in RNA-Seq analysis, we are confident in the ability to analyze this data. Furthermore, it is likely that the oyster genome will be released in the near future. Recently, the genomics institute BGI in Shenzhen, China stated in a press release that they had sequenced the Pacific oyster genome (Nature News, 2010). This data would assist in the analysis of the data, however, it is not necessary for completion of this project. *Furthermore, the sequence information generated from the non-bisulfite treated libraries will provide direct empirical evidence of what portions of the oyster genome are methylated irrespective of the bisulfite treated libraries.*

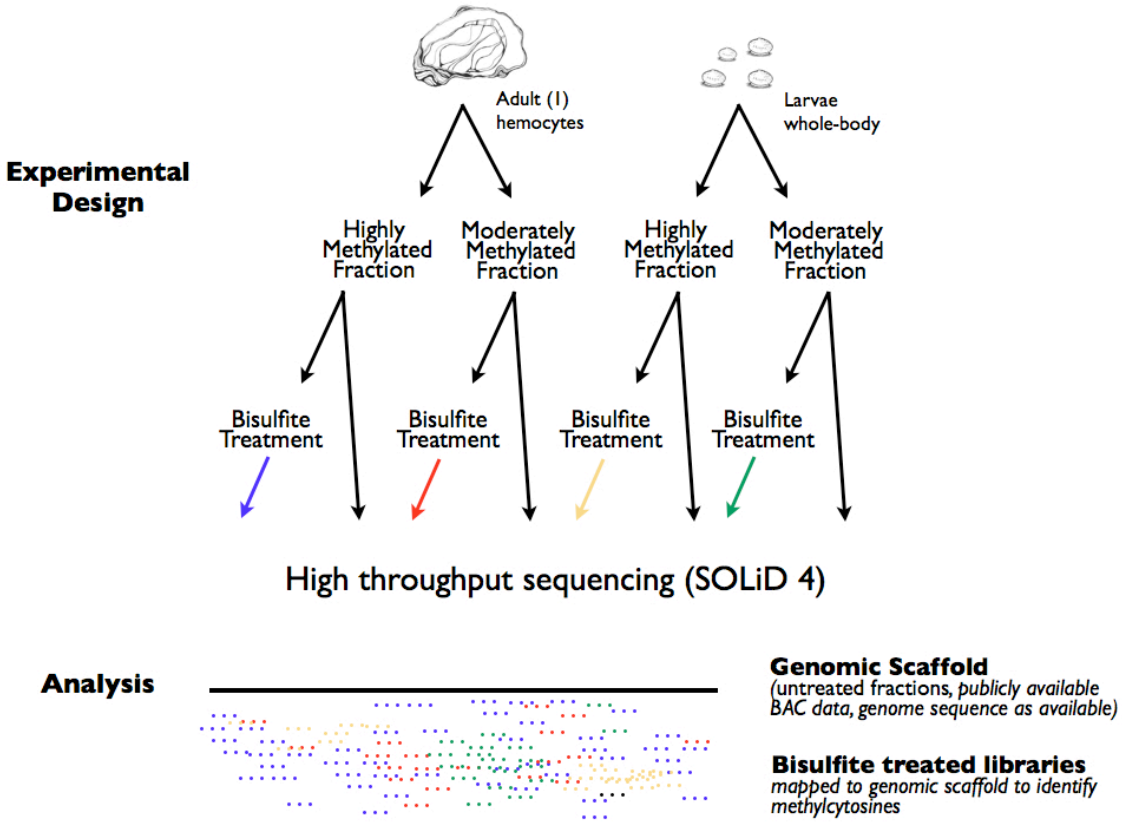

Figure 4. Schematic illustrating experimental design and analysis related to research objective 1. DNA from adult oyster hemocytes and whole bodied larvae will be fractionated based on degree of methylation. Half of each sample will be subjected to bisulfite treatment and eight libraries will be generated and sequenced on the SOLiD 4 System. For analysis, bisulfite treated reads will be mapped back to a genomic scaffold comprised of reads from the 4 untreated libraries, publicly available genomic sequence, and the reads from the MBD libraries sequenced in our lab in April 2011.

3.2 *Objective 2* - Investigate how DNA methylation contributes to phenotypes associated with inbreeding depression and hybrid vigor

The second objective of the proposed project is to identify how DNA methylation contributes to phenotypes associated with hybrid vigor. As described, it has been demonstrated in other systems that DNA methylation and other epigenetic mechanisms play a role in this phenomenon. For this component of the project, the MBD methylation enrichment procedure will be used in conjunction with tilling array chip hybridizations to identify differences in DNA methylation between hybrid and inbred oysters (Figure 5). Selectively bred oyster lines shown to display heterosis (Hedgecock and Davis 1995), will be utilized to investigate differentially methylated regions between inbred (oyster lines 35 & 51) and hybrid crosses associated with growth phenotypes (The designation “35” and “51” are arbitrary names of inbred oyster lines used in this study). Hybrid individuals of these lines have shown increased yield and growth in previous generations. It also is worth noting that these lines have been used by other research groups for sequencing efforts, including ESTs, BAC sequences, and whole genome sequencing efforts (BGI).

The MBD fractionation is the same procedure the PI has used in construction of the SOLiD libraries (Preliminary Research Results) and that is proposed for research objective one. The chip based approach, referred to as MBD-chip, was selected as compared to the direct sequencing based method because the array platform offers an efficient and economical means to compare genome-wide methylation changes associated with different treatments or conditions. For instance, this technology has been used to analyze methylation patterns in cancerous cells compared to controls. In addition, MBD-chip has been used to evaluate whole organism exposures to environmental compounds such as cocaine, BPA, and pesticides (Nvikova et al 2008, Yaoi et al 2008, Guerrero-Bosagna et al. 2010).

While the MBD-chip approach is ideal for comparing samples it does not provide high resolution CpG methylation pattern information. Therefore, select targets that are differentially methylated will be selected for bisulfite treatment and gene-specific pyrosequencing. In addition, quantitative PCR will be carried out on a limited number of genes to begin to examine relationships of DNA methylation and transcriptional regulation. Methods are available to follow up MBD-chip with sequencing all hybridized products and examining whole transcriptome differences (i.e. gene expression array) however, this would be beyond the scope of the current proposal.

3.2.1 Experimental Design - *Objective 2*

For this experiment, tissue samples have already been harvested and preserved by Taylor Resources, Inc. (see letter of support from Dr. Jonathon Davis) Inbred parent lines (35 and 51) were used to generate inbred offspring from both parental lines (i.e. 35 x 35 (A) and 51 x 51 (B)), as well as hybrid offspring (35 x 51 (C)). Two separate crossing events (denoted by subscript) will be examined from

each of the three crosses for a total of six families (Figure 5). DNA will be isolated from eight individuals from each of the six families and pooled in equal concentrations. Individual samples will be maintained in order to perform single template pyrosquencing on targets that are differentially methylated. MBD enrichment will be performed, to fractionate methylated DNA.

A tiling array will be developed to assess differential methylation in oysters. Given the current available sequence data for the Pacific oyster a 3 x 720k chip will be designed by NimbleGen Roche. The probe sizes will range from 50 – 75 bp in length with a median probe spacing of 100 bp. The tiling array approach will be the most appropriate tool to use for this analysis as it allows for a direct comparison of samples and provides good resolution (~50 – 70bp).

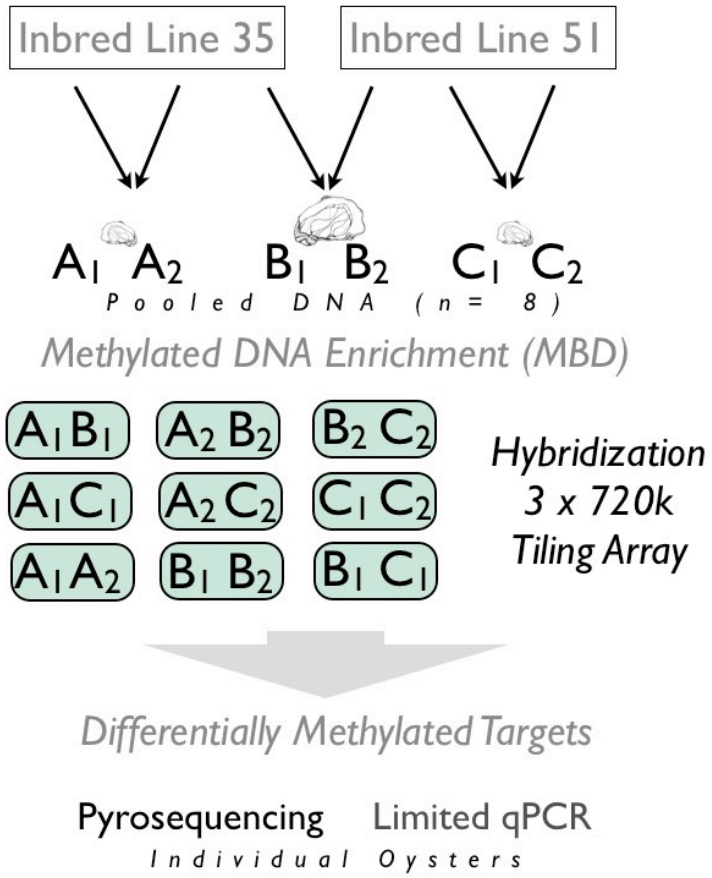

Figure 5. Schematic illustrating experimental design related to research objective two. Differential methylation patterns will be examined in a hybrid cross of two inbred lines (B) and the two parental inbred lines (A & C). Two separate crossing events (denoted by subscript) will be examined from each of the three crosses for a total of six families. Nine direct comparisons will be made (green circles) with each direct comparison performed in duplicate with a dye swap. Differentially methylated targets will be further interrogated by pyrosequencing and quantitative PCR.

The array analysis will be able to identify small regions of DNA that are differentially methylated between inbred and hybrid families as well different cross of the same parental lines (denoted with subscript in Figure 5). Briefly, labeled MBD fractions will be co-hybridized to the array, washed, and scanned. Each comparison will be performed in duplicate in order to include a dye swap for quality control of the data. Methylation enrichment and probe labeling will be done in the lab of PI with hybridization and scanning performed by the Fred Hutchinson Cancer Research Center Genomics Core Facility at the University of Washington. This core facility has extensive experience in all aspects ofarray technology and analysis. Working with this facility will ensure the procedures are carried out in a

manner that is MIAME compliant. In addition, personnel will work with the PI to ensure proper data analysis methodologies are used.

Once differentially methylated regions have been identified, these differences will be interrogated directly by pyrosequencing bisulfite treated DNA from individual oysters. This is a quantitative sequencing-by-synthesis method that monitors real-time incorporation of nucleotides through the enzymatic conversion of released pyrophosphate into a proportional light signal (Tost & Gut 2007). In addition, a limited number of targets will be selected and quantitative PCR performed as described in Roberts et al. (2009) to begin to examine the relationship between DNA methylation and transcriptional regulation.

3.2.2 Expected Outcomes - *Objective 2*

Upon the completion of this objective we will have gained significant information regarding the role of DNA methylation in contributing to phenotypes associated with hybrid vigor. Based on physiological observations suggesting the efficiency of protein turnover likely contributes to heterosis in oysters (Bayne et al 1999, Pace et al 2006, Hedgecock et al 2007), we expect that genes associated with metabolism and protein degradation will be differentially methylated between inbred and hybrid oysters. Genes identified as being differentially regulated by the tiling array will be further validated by qPCR. With regard to the replicate families for a given cross, it is expected that there will be limited differences in methylation observed for families A and B with increased variation in family C (hybrids).

3.2.3 Potential Pitfalls - *Objective 2*

In order to carry out this research objective we have chosen to use a tiling array platform that limits the interrogation to known sequence information. A substantial amount of the transcriptome has been assembled (GigasBase website), however, it is not complete. Furthermore, we are excluding non-coding sequence information (e.g. introns). Most of the research to date has indicated that DNA methylation is limited primarily to exonic sequences in invertebrates. We feel confident that differences in methylation will be observed using this method, particularly given the differences in gene expression patterns observed with heterosis and the role of DNA methylation in transcriptional regulation in other species. Furthermore, if the sequence of the oyster genome is released in the near future, we could revise our array design.

3.3 Detailed Methods

Animal Collection & DNA Isolation

All Pacific oyster samples will be provided from a local commercial shellfish hatchery (Taylor Resources, Inc., Shelton, Washington - See attached letter of support). DNAzol (MRC) will be used to isolate DNA from all samples according to the manufacturer's protocol

Methylation Enrichment

Methylation enrichment will be performed using the Methyl-Miner Kit (Invitrogen) which binds fragmented double-stranded genomic DNA using biotin labeled Methyl Binding Domain 2 protein (MBD2), and can be eluted using a salt gradient. DNA will first be fragmented to ~150 bp, then incubated with methyl-CpG binding domain of human MBD2 protein, coupled to paramagnetic bead via a biotin linker. The bound fraction (methylated fraction) will be eluted with a high salt concentration buffer either in multiple steps of increasing salt concentration (objective 1) or in a single high concentration (objective 2). The unbound fraction will also be retained.

Library Construction

DNA libraries will be prepared using the SOLiD DNA Fragment Library Kit (Applied Biosystems). Briefly, this involves end-repairing the DNA fragments (fragmentation performed during methylation enrichment – see above), ligating adaptors (P1 and P2), size selecting the DNA, then performing nick-translation and amplification via PCR with primers specific to the adaptor sequences. DNA will be analyzed using a Bioanalyzer (Agilent) prior to emulsion PCR. Initially, an octect (1/8 of a slide) will be sequenced from each library. Sequencing will be carried out at the UW High-Throughput Sequencing Facility (UW HTGU).

Bisulfite Treatment

Methylation enriched DNA fragments will be bisulfite converted using the Epitect Bisulfite conversion kit (Qiagen, Carlsbad, CA). DNA will be subjected to treatment with sodium bisulfite at increased temperature to deaminate unmethylated cytosine residues to uracil. Following treatment, the solution will be desulfonated on a column, washed and eluted prior to library construction.

Library Construction – Bisulfite Sequencing

Library construction for sodium bisulfite treated samples will be similar to the above library construction procedure with the following exceptions: the top strand of the P1 adaptor will be synthesized with 5-methyl cytosine instead of cytosine to prevent modification during bisulfite conversion and during nick-translation 2′-deoxycytidine-5′-triphosphate (dCTP) will be replaced with 5-methyl-2′-deoxycytidine-5′-triphosphate (5mC-dNTP) in the original dNTP mixture (standard dNTPs for A, G, and T will be used). Bisulfite conversion will be performed in solution according to Renade *et al*, 2009. After bisulfite conversion, PCR amplification will be performed

as described above. Note that only one strand will amplify during PCR (the bottom strand is not protected during bisulfite conversion), which will simplify downstream analysis.

Analysis
Sequencing data will be analyzed using CLC Genomics WorkBench (CLC Bio) along with publicly available databases (NCBI, SWISS-PROT, GigasBase) and our own unpublished *C. gigas* RNA-Seq libraries. The PI has significant experience with this data format, characterizing RNA-Seq libraries from trout (Goetz et al. 2010), chum salmon (Seeb et al. 2010), and several shellfish species (unpublished). Analysis will include quality trimming, *de novo* assembly, and BLAST. Comparisons among libraries will be made within CLC Genomics Workbench (ChIP-Seq, RNA-Seq) and Galaxy tools (http://main.g2.bx.psu.edu/).

Array Design and Analysis
A custom tiling array will be designed utilizing NimbleGen technology. A 3 x 720k design will be used with probes derived to cover the oyster transcriptome as assembled by GigasDatabase (Fleury et al.2009). Two different comparative hybridizations will be performed for each sample. For each comparison outlined in Figure 5, labeled samples will be co-hybridized to the array, washed, and scanned. Array images will be used for data extraction as pair files. Genomic feature format (GFF) files will be produced for visualization of scaled log2-ratio data. P-value files (.gff) files will be generated from the scaled log2-ratio data, where each probe is tested for positive enrichment of DNA methylation against all probes on the array. All array processing and analysis will be conducted at the Fred Hutchinson Cancer Institute Core Facility at the University of Washington. This core facility has significant expertise with this platform.

Pyrosequencing
Differentially methylated targets identified using MBD-chip will be subjected to gene specific pyrosequencing to determine which CpGs are methylated. DNA from individual oysters will be treated with sodium bisulfite (see procedure above). Primers will be designed for the bisulfite treated sample. For this analysis, PCR will be performed with one biotinylated PCR primers. This is required for the conversion of the PCR product to single-stranded DNA templates. A sequencing primer will then be added, which anneals to the single stranded template. The limitation of this method is that all differentially methylated regions (DMRs) will not be easily interrogated due to the reduced complexity of the sequence as a result of the C to T conversion. However, for regions where suitable primers can be designed, PCR will be performed and products will be sent to EpigenDx (Worchester MA) to perform the pyrosequencing analysis. Paired t-tests will be performed to evaluate significant differences at $p < 0.05$ level.

Data Release
All high-throughput sequencing data will be released in a timely manner through NCBI Short Read Archive and local servers at the University of Washington. MBD-chip data will be deposited in the Gene Expression Omnibus (GEO) at NCBI.

3.4 Timeline

Activities to be carried out during Year 1 of the proposed project will focus on research objective 1 and the direct sequencing of hemocyte samples from adult oysters and whole body larvae to characterize differences in DNA methylation patterns across the oyster genome. At the end of Year 1 we expect this milestone to be completed and research activities required to accomplish research objective 2 will begin. A tiling array will be made by NimbleGen from the Pacific oyster transcriptome. Pooled, methylated enriched DNA from hybrid and inbred oyster lines will be directly compared using the MBD-chip approach early in Year 2. Data analysis is projected to be completed by February of Year 2. Pyrosequencing and quantitative PCR will be finished in the Spring. The schematic below provides timing for completion of research activities with primary milestones italicized. Research results will be presented at national meeting at the end of Year 1 and Year 2. Throughout the duration of the project, research progress will be made available to the general public through a dedicated project website as well as via open access electronic laboratory notebooks maintained by all personnel in the lab of the PI available at http://genefish.wikispaces.com.