**Characterization of DNA methylation as a source of epigenetic regulation in *Crassostrea gigas***

*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. DNA methylation has been well studied in mammals, however the same level of research has not been extended to invertebrates and surprisingly little is known about this mechanism in these taxa. Previously, we have applied *in silico* approaches to characterize DNA methylation in Pacific oysters (*Crassostrea gigas*). Our results suggest that DNA methylation has regulatory functions in *C. gigas*, particularly in gene families involved in stress and environmental responses. In light of these findings, we have performed high throughput bisulfite sequencing using C. gigas genomic DNA. This approach will allow us to quantitatively assess the methylation status of hundreds of thousands of cytosine loci in the genome. The recent release of the *C.gigas* genome allows us the opportunity to not only map and report the methylation status of these sites, but also to evaluate the structure of the DNA methylation landscape using multivariate techniques.

*Questions/Objectives*: The purpose of this analysis is to determine if there are any particular genomic attributes or groups of attributes that predictably co-occur with methylated cytosines in the *C. gigas* genome. We expect that the results from this analysis will shed light on the structure of the DNA landscape in this genome and allow us to make informed hypothesis about the regulatory role of DNA methylation in molluscs (e.g. repression of transposable elements, regulation of gene expression, control of alternative splicing)

*Source of Data*: The data for this analysis come from two sources. The DNA methylation data (given as % methylation at a particular genomic locus) comes from mapping the high-throughput sequencing reads to the *C. gigas* genome using BSMAP software. The genomic attribute data for each locus was compiled from recently released C.gigas genome (Zhang et al., 2012).

*Structure of Data*: The sample unit for this analysis is a genomic locus. Specifically, it is a cytosine locus with a known percent methylation status (0 – 100%). The high-throughput sequencing effort provided quantitative methylation status for approximately 250,000 loci. The data matrix will include the loci as objects and percent methylation as one variable. The additional variables include categorical (binary) genomic attributes (e.g. intra- or inter-genic) with some hierarchical structure within the variables. For example, if the locus is intragenic is it in an exon or intron, and if exon what is the biological function of the gene (up to 16 non-exclusive descriptors). Conversely, is an intergenic locus within a repetitive element such as a transposon, or is it in a promoter region or neither. Regarding aggregation, it may be more appropriate to aggregate the methylation data to a level above an individual cytosine, such as the methylation status of a 100bp window. I believe that type of aggregation is possible using available software and I will look into this option if it appears to be more appropriate for the analysis.

Zhang, G et al. (2012) Genomic data from the Pacific oyster (Crassostrea gigas).GigaScience. <http://dx.doi.org/10.5524/100030>