Professor Xiang-Jiao Yang

Academic Editor for Peer J

Dear Professor Yang,

## Please find attached our revised manuscript "Predominant intragenic methylation is associated with gene expression and tissue specificity in a bivalve mollusc" (#2013:09:844:0:0:REVIEW). All revisions are indicated using “track changes”. We would like to thank you and the reviewers for their comments; we believe the revised version is considerably improved in response to the comments provided.

## Our responses to the questions and suggestions *(in italics)* are given below:

### *Reviewer 1*

### *Basic reporting*

*The manuscript entitled “Predominant intragenic methylation is associated with gene expression and tissue specificity in a bivalve mollusc” by Gavery and Roberts describes their analyses of genome-wide relationships between gene body methylation and gene expression. While for the most part I find their analyses to be interesting, it is mostly a descriptive study with no examples of specific genes that provide insight or corroborate their finding.*

#### ***(1) Experimental design***

*The authors of this study state that “methylation in oysters does likely vary in both a temporal and possibly tissue specific manner”, while the Zhang’s study observed that “The oyster genome is highly polymorphic”. Given that “The cohort of adult oysters used in this study was from Samish Bay, WA, USA”, while the pacific oyster used in Zhang’s study is an inbred female produced by four generations of brother–sister mating. Question is the compatibility of the methylome data from this study with the RNA-seq data from Zhang’s study. RT-PCR validation of the expression levels of selected example genes and some in-depth discussion would help to address this issue.*

#### REPLY: The reviewer makes a good point regarding the polymorphic nature of the genome. In order to ensure there were no confounding effects of specimen origin we are no longer using the Zhang study data for comparison but rather a RNA-seq library that corresponds to the same individuals used for DNA methylation analysis. We did consider this before but opted to use the Zhang data as a commonality through other analyses. The new results are the same with respect to overall methylation and expression (FIGURE 3), as might be expected at the gene level. More importantly, the inclusion of this data provides the ability to derive specific examples, a point raised by the reviewer. As we are now using complementary data we see no need to ‘validate’ with RT-PCR. The comment regarding specific examples is addressed below.

As we are now including new data, the following section has been added to the Methods section:

**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.

#### ***(2) Validity of the findings***

*The author noted that “Specifically, DNA methylation appears to be overrepresented in intragenic regions” and in intron and Intronic methylation has been implicated to be involved in gene regulation through the expression of alternative isoforms of genes. Beside just global analyse of the methylome data, some detailed analyse on example genes would provide some insight.*

REPLY: Given the experimental design (or rather lack of experiment here) and considering pooled nature of the samples (that does offer benefits for our purpose) we do not want to offer up specific examples as proof and there is not an appropriate genome wide analytic approach given the design. For instance we do not have the ability to identify alternative isoforms, and even if we could, we could not test for whether DNA methylation patterns contributed to different isoforms.

We did however come up with what we feel is an appropriate action that addresses (and goes beyond) the reviewers request without overstepping our ability to draw accurate conclusions. The solution we came up with is to co-publish all data in a manner that anyone can easily examine and explore the relationship with gene expression and DNA methylation. Specifically, as part of the figshare fileset (cited in the original manuscript: http://dx.doi.org/10.6084/m9.figshare.749728) we have now added an exon-specific gene expression (RPKM) track, CpG methylation ratio feature track, an Integrative Genomics Browser (IGV) session file (xml), and instructions for visualizing. IGV is free and the steps are as simple as:

* Go to **File** > **Load from url**
* paste http://eagle.fish.washington.edu/trilobite/Crassostrea\_gigas\_v9\_tracks/BiGill\_igv\_charlie.xml

This will result in a browser window as shown below:



####

#### The following text was added to the Discussion to clarify this point and direct readers to the available data: “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).”

#### ***(3) Comments for the author***

*Missing ")" on line 120.*

### REPLY: This correction has been made.

###

### *Reviewer 2*

### *Basic reporting*

*The authors report a high-resolution methylome analysis in Crassotrea gigas. They used a bisulfite conversion strategy that allows identifying DNA methylation at single nucleotide resolution. They show that methylation is essentially intragenic and associated with gene expression as this was previously demonstrated in other invertebrate species.*

#### ***Experimental design***

*They used commercially available kits to enrich their starting biological material in methylated DNA and to do the bisulfite treatment. They used the illumina strategy for the sequencing of their fragmented DNA. Their bioinformatic pipeline is robust and based on software that was previously successfully used.*

#### ***Validity of the findings***

*This is the first report of a genome wide methylome analysis in Crassotrea gigas and also in a molluscan species. This work is certainly of great value to help at deciphering methylated region in this organism. Therefore this publication will certainly be a reference for further analysis in oyster and other bivalves. It comes with other papers that were recently published in the field with the aim at describing the structure of methylated DNA in invertebrate species. This is so far a unique reference for the molluscan phyla. I have however some suggestion to improve the quality of the manuscript (please see next section).*

#### ***Comments for the author***

*Comments on the method section:*

*(1) About the MethylMiner Kit (Invitrogen) :*

*- This kit requires DNA to be fragmented before the methylation enrichment. There is no details about the fragmentation of the DNA, starting amount of DNA? Was sonication used, which machine, what condition? what was the size of the fragmented DNA?*

*- Enriched methylated DNA may be eluted either as a single fraction or doing fractionate steps with increasing NaCl concentration, which condition was used?*

*- Several procedure exists depending on the starting amount of DNA, salt concentration etc.. Which one was used? What was the starting amount of DNA for this procedure?*

*In this sense, an accurate summary of each of these two procedures would be appreciated for the readers to be able to repeat the experiment. Please, could you also provide a brief summary of the library construction and sequencing experiment?*

REPLY: We agree with the reviewer that these details should be provided to ensure reproducibility of the experiments. The following details were provided in Methods:

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 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).

*(2) Comments on the result section:*

*The authors published in a previous paper the distribution of predicted methylation status of 12,210 annotated C. gigas transcripts measured computationally by CpGo/e ratio. This distribution is bimodal as observed in other invertebrate organisms. The results presented in the figure 1 of this paper seem to confirm this bimodality. I would find interesting to perform a correlation analysis between in silico predicted methylation status and methylation ratio in genes that encodes these 12,210 transcripts. This kind of correlation has been validated in insect species previously in the work of Sarda et al (2012) and it would be interesting to validate it in mollusk. It would be helpful to predict methylation profil based on in silico prediction in other bivalves.*

REPLY: We agree with the reviewer that this correlation analysis would be interesting to validate previous findings and beneficial for future work in other molluscs using the CpGo/e. To increase the robustness of this analysis we did not limit the dataset to the 12,210 transcripts previously analyzed, but instead extended it to the whole genome. The following sections were included in the manuscript.

Methods:

The relationship between in silico prediction of methylation status of genes using the CpG observed to the expected ratio (CpGo/e) and the average methylation ratio for each gene was examined in order to assess the usefulness of the CpGo/e method for predicting methylation in bivalve species. For this analysis 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.).

Results:

Additionally, a strong negative 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).

Discussion:

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).

*(3) The principal component analysis is not clear; the authors should revise this section and clearly explain the aim of their analysis and their interpretation.*

REPLY: The principal component analysis was utilized as an exploratory tool to identify relationships between DNA methylation, gene expression characteristics, and gene attributes. The following additions were made to the manuscript to clarify the exploratory nature of this analysis:

Methods:

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.

Discussion:

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.

*(4) Comments on the discussion section:*

*The authors highlight similarities in epigenetic profiles with other invertebrates such as a predominance of gene body methylation and a positive relationship between intragenic methylation and gene expression. Other point could be discussed: transposon methylation and mosaic like feature. Indeed, genomes of invertebrates are characterized by interspaced regions of methylated and unmethylated DNA. Have the authors any evidence for such a mosaic like feature in oyster?*

REPLY: We agree with reviewer 2 that this additional information would improve the manuscript. The following paragraph was added to the Discussion:

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).

*(5) Comments on the title:*

*I agree with a title mentioning that predominant intragenic methylation is associated with gene expression in a bivalve mollusk, but I don’t see where the authors discussed the association with tissue specificity. The authors should revise their title or go deeper into an analysis that indeed demonstrates a link with tissue specificity.*

REPLY: We agree with the comments of the reviewer, particularly given the exploratory nature of the ordination analysis, and have revised the title accordingly:

Previous Title: Predominant intragenic methylation is associated with gene expression and tissue specificity in a bivalve mollusc.

Revised Title: Predominant intragenic methylation is associated with gene expression characteristics in a bivalve mollusc.

*(6) Minor revisions:*

*L90: Supplementary instead of supplemental*

*L213: Riviere instead of Riverie*

*Legend of figure 3, y axis: of is repeated twice*

REPLY: The 3 suggested minor revisions were corrected in the revised manuscript.