Elucidating the effects of the Deepwater Horizon oil spill on the Atlantic oyster using global transcriptome analysis

Natàlia G. Reyero[1], Nalini Raghavachari[2], Kurt Showmaker[1], Eric Olson[3], Poching Liu[2], Nadereh Jafari[4], Natalie Barker[5], Kristine Willett[6], Jone Corrales[6], Heather Patterson[7], Ruth Carmichael[7], Don Baldwin[8]

[1]Institute for Genomics, Biocomputing and Biotechnology, Mississippi State University, Starkville, MS, USA; [1]National Heart, Lung, and Blood Institute, NIH, Bethesda, MD 20892, USA; [3]Geospiza/Perkin Elmer, Seattle, WA, USA; [4]Center for genetic medicine, Northwestern University, Chicago, IL 60611, USA; [5]Environmental Laboratory, US Army Engineer Research and Development Centre, Vicksburg, MS, USA ; [6]School of Pharmacy, University of Mississippi, Oxford MS 38677, USA; [7]Dauphin Island Sea Lab, Dauphin Island, AL 36528, USA; [8]Pathonomics LLC, Philadelphia PA 19104, USA

The problem

Global transcriptome analysis is of growing importance in understanding the effect of environmental pollutants to living organisms. The Genomics Research Group applied next generation sequencing technologies to study the effects of deepwater horizon oil spill on the transcriptome of Atlantic oysters.

The Deepwater Horizon oil spill resulted in the release of over 200 million gallons of crude oil into the waters of the Gulf of Mexico. Over two million gallons of chemical were used to emulsify and disperse oil plumes posing further risks to the environment in addition to the direct impacts of crude oil. Biota such as the commercially important oyster *Crassostrea virginica*, were inevitably exposed to spill-related contaminants in the Gulf.

The potential effects of oiled water and sediments on oysters range from non-detectable to reduced settlement to impaired immune function, acute intoxication, and death due to bioaccumulation of contaminants. Oil also may affect oxygen diffusion through the water column, and in some cases lead to hypoxic conditions that prompt avoidance migration by mobile species. Sedentary organisms such as oysters are even more susceptible to these negative effects of oil contamination.

The mechanisms of toxicity of the oil and spill-related compounds and their toxicity as secondary stressor in hypoxic conditions are not well understood. In order to understand these mechanisms, we used RNA-sequencing of oyster samples from before and after the spill.

Our approach

Atlantic oysters were transplanted (Sand Reef, AL) at two different depths in either normoxic or hypoxic conditions. Oysters were collected before the spill (06/09/10) and after the spill (06/28/10) and RNA was extracted from the gill. Total PAHs were measured from water and oyster. Dissolved oxygen was measured in the water to confirm the normoxic/hypoxic conditions.

Fig 1: oil reaching Mobile Bay on June 29th 2010. Site of exposure is shown in red.

RNA-sequencing was performed on the oyster samples (3 replicates per treatment). Libraries were constructed using the Nugen RNAseq Complete System. The libraries were sequenced using the Illumina HiSeq (single end, 50 bp length reads). In order to help with the de novo assembly, another run was performed on an Illumina MiSeq (single end 150 bp).

As the Atlantic oyster genome is not yet available, we followed three different approaches to assemble the transcriptome. First, the sequences were mapped to the Pacific oyster genome using GeneSifter (Perkin Elmer). Secondly, a reference transcriptome was assembled *de novo* using Trinity (http://trinityrnaseq.sourceforge.net/). Finally, the HiSeq and MiSeq sequences were assembled using the Oases module of Velvet (European Bioinformatics Institute). All transcripts were annotated using BLAST and gene expression analyses were performed.

Results and Discussion

• Chemistry

Total polycyclic aromatic hydrocarbons (PAHs) were measured in the water and in oyster homogenates. The total amount of PAHs is significantly higher on June 28th, when the oil reached the area. The PAH composition was also different at that time (Fig 2).

Table 1: Total polycyclic aromatic hydrocarbons (PAHs) measured in the water (ng/L).

Date	Top	Bottom
5/28/10	16.04	12.82
6/9/10	9.66	10.47
6/28/10	268.14	30.99
7/20/10	91.5	48.39

Table 2: Total PAHs measured in the oyster (ng/g dry weight).

Date	Top	Bottom
6/9/10	84.7	47.05
6/28/10	116.47	205.88

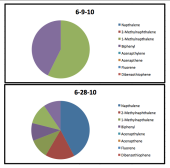

Fig 2: Percentage of different PAHs measured in oysters.

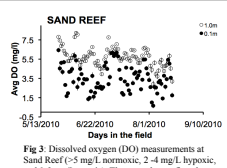

Fig 3: Dissolved oxygen (DO) measurements at Sand Reef (>5 mg/L normoxic, 2 -4 mg/L hypoxic, and 0-2 mg/L anoxic). These results confirm the hypoxic conditions in our hypoxic treatments.

• Method 1

A total of 4,818,117 assembled sequences were mapped to the recently published Pacific oyster genome (http://gigadb.org/pacific_oyster/). About 16% of the sequences matched the Pacific oyster. Differentially expressed genes were calculated using a 2-way ANOVA. A total of 4,848 genes were found differentially expressed, with 543 genes differentially expressed by *time* and 629 were differentially expressed by *depth*.

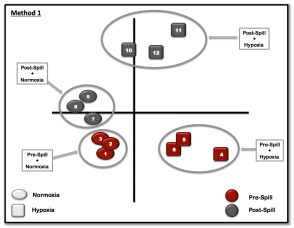

Fig 4: Principal Component Analysis (PCA) of the 12 different samples using all differentially expressed genes. The samples cluster clearly by treatment.

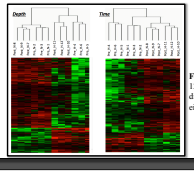

Fig 5: Hierarchical clustering of the 12 different samples using the differentially expressed genes by either *depth* or *time*.

• Method 2

A total of 4,818,117 assembled sequences were used to create *de novo* a reference transcriptome using Trinity. The assembled sequences were then mapped to the transcriptome in order to create a transcript list for each sample. A 66% of the sequences mapped to the reference transcriptome. Differentially expressed genes were calculated using a two-way ANOVA (5% FDR). A total of 1,002 genes were differentially expressed by *time* and 1,606 were differentially expressed by *depth*.

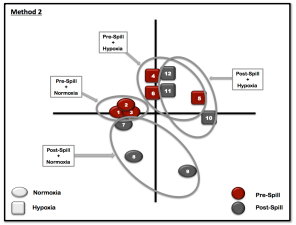

Fig 6: PCA of the 12 different samples. Samples cluster mostly by treatment, although two of the treatments (pre- and post- spill hypoxia) appear to be more similar.

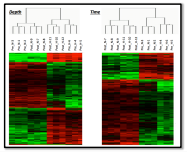

Fig 7: Hierarchical clustering of the 12 different samples using the differentially expressed genes by either *depth* or *time*.

• Method 3

The 1,204,744 Miseq and 100,959,733 Hiseq sequences were used in a *de novo* transcriptome assembly using the oases module of the velvet short read assembler. Hiseq reads were mapped to the assembled transcriptome using Bowtie (Johns Hopkins University). Read counts and RPKM values for each transcript were generated using RPKMforgenes.py. Differential gene expression was determined by a negative binomial model with the DESeq library in the statistical programming language R. The 9,446 homologies with the Pacific Oyster were determined by reciprocal BLAST alignments between the Atlantic Oyster transcriptome and the proteome of *Crassostrea gigas*.

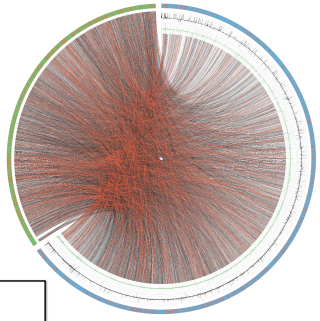

Fig 8: Outer track represents Atlantic Oyster transcriptome (blue) and Pacific Oyster proteome (green), with red stroked and lines (n = 689) linking them representing differentially expressed genes ($p<0.05$). Black lines represent the 9,446 homologies identified by the reciprocal BLAST. Inner tracks show the mean difference in RPKM values in test 1 (black, treatments 1 and 3) and test 2 (green, treatments 2 and 4).

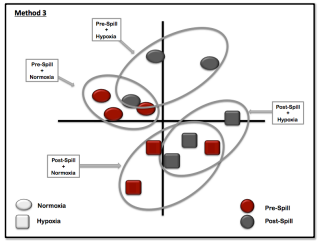

Fig 9: PCA of the 12 different samples. Samples cluster mostly by treatment, some of the samples share some similarity.

GO	Biological Process
GO:0048856	anatomical structure development
GO:0030154	cell differentiation
GO:0034641	cellular nitrogen compound metabolic process
GO:0009058	biosynthetic process
GO:0006950	response to stress
GO:0000003	reproduction
GO:0006810	transport
GO:0006464	cellular protein modification process
GO:0044281	small molecule metabolic process
GO:0002376	immune system process
GO:0009790	embryo development
GO:0016192	vesicle-mediated transport
GO:0040007	growth
GO:0006629	lipid metabolic process
GO:0022607	cellular component assembly
GO:0007010	cytoskeleton organization
GO:0050877	neurological system process
GO:0040011	locomotion

Table 1: enriched biological processes from differentially expressed genes. Some of these processes are known to be related to PAH exposure.

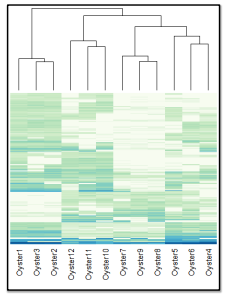

Fig 10: Hierarchical clustering of the 12 different samples using the top differentially expressed genes.

Conclusions

Global transcriptome analysis is a promising tool to identify the effects of environmental stressors on wildlife. A major challenge when using this type of analysis is the availability of a reference genome. Interestingly, our results show that RNA-sequencing is extremely useful for the study of environmental samples. Nevertheless, the approach used to assemble the transcriptome seems to have a big impact on the final results. Methods 2 and 3 use a *de novo* transcriptome assembly and give a higher number of mapped sequences than method 1, which maps the sequences to a reference genome. That potentially translates into better biological information due to a higher number of annotated transcripts.

Acknowledgements

This work was funded by the US Army Environmental Quality Research Program, IGBB, NIH, Pathonomics, NSF, and the Gulf of Mexico Research Initiative.

Contact: natalia@icnanotox.org