FINAL REPORT

INSTITUTION: Purdue University
TITLE: Utilization of genomic signatures from *Hyalella azteca* as a way to quickly evaluate toxicity and need of sediment remediation in the Great Lakes basin.
PRINCIPAL INVESTIGATOR: Sepúlveda, Maria S.
AFFILIATION: Forestry and Natural Resources and School of Civil Engineering, Purdue University
CO-PRINCIPAL INVESTIGATOR: Colbourne, John K.
AFFILIATION: The Center for Genomics and Bioinformatics, Indiana University

OBJECTIVE:
The main objective of our study was to apply state-of-the-art technologies for a quick and accurate evaluation of the hazards posed by contaminated sediments across the Great Lakes basin. More specifically, we utilized microarrays for the evaluation of gene expression signatures after exposure to different sediment pollutants. We achieved this goal by utilizing a recently developed microarray for the evaluation of gene expression signatures after exposure of the amphipod *Hyalella azteca* to different pollutants commonly found in sediments across the Great Lakes basin.

PROBLEM:
Decades of point and non-point pollution sources to the Great lakes and its tributaries have resulted in major sediment contamination. Indeed, sediments from this basin are known to contain many different types of contaminants, such as heavy metals (mercury, lead, cadmium), herbicides (atrazine), PCBs, and PAHs. More recently, emerging contaminants, such as polybrominated flame retardants, have also been detected in Great Lakes sediments. Sediments contaminated with organic and inorganic contaminants have the potential to negatively affect ecosystem health through impacts on aquatic organisms, wildlife, and humans. The problem is so widespread, that over 2,000 miles (20%) of the Great Lakes shoreline is considered impaired because of sediment contamination and fish consumption advisories. Altogether, these areas fall into the category of Areas of Concern (AOC) and are defined as "places where beneficial uses of water resources such as drinking, swimming, fishing, and navigation are impaired by anthropogenic pollution or perturbation" (Adriaens et al., 2002, "Great Lakes Sediments: Contamination, Toxicity and Beneficial Re-Use", White paper commissioned by Michigan Sea Grant and the School for Natural Resources and the Environment, 37 pp.). Since it is impossible to remediate this vast geographical area, it is imperative that contaminated areas get ranked in terms of sediment toxicity and ecological impacts.

RATIONALE (IMPACT OF PROBLEM):
The expense involved with removing or treating contaminated sediments requires an accurate characterization of risks. Currently, one of the standard approaches is to evaluate sediments through the sediment triad approach, which includes sub-chronic and chronic toxicity testing using the benthic invertebrate *H. azteca*. The high cost of these tests results in either very high evaluation costs, or remedial decisions based on very limited data. A gene microarray chip specific to *H. azteca* can eventually be used as a surrogate for chronic tests, resulting in

significant money and time savings.

Gene arrays (also called microarrays), are a relatively new research tool where thousands of genes specific to an organism are spotted onto a solid support matrix and queried with RNA samples from animals exposed to stressors, which are compared to controls. Our own work and that of others has shown that classes of chemicals, and perhaps even individual chemicals, are likely to display their own “chemical signatures" in terms of the specific genes that are differentially activated upon exposure. A major advantage of this approach is that these “chemical signatures” can be elicited rapidly (within hours of exposure), even though the physiological responses to which they are correlated may require much prolonged exposures. Moreover, such biomarkers can provide early warnings of ecological duress, even when there is limited knowledge about the mode of action of individual toxicants or of chemical mixtures. The application of such arrays for screening is already proven in the diagnosis of several human diseases. However, their application in ecotoxicological studies have yet to be rigorously tested, largely because the production of microarrays for such purpose requires substantial resources, including characterized and relevant cDNA databases to source the DNA probes on the arrays.

Our research has the potential to revolutionize toxicogenomics using model species whose genomes are not yet fully sequenced. If we succeed in producing a proof of principle in extracting genomic signatures that are diagnostic of the toxicological effects using this model aquatic crustacean for toxicological evaluations, then other species will follow for cross-species extrapolations of environmental toxic effects from contaminants. This project is similar in design to the recent endeavors by the *Daphnia* Genomics Consortium – based within Indiana University’s Center for Genomics and Bioinformatics (CGB) – to develop a comprehensive database of genomic signatures of environmental compounds for this distantly related limnetic crustacean. Moreover, given the paucity of functional genomic data for Arthropods other than *Drosophila* and its insect allies, the comparative value of microarray experiments in crustaceans that root the insect phylogenetic tree will provide important insights into the evolution of toxicological responses and their gene regulatory networks.

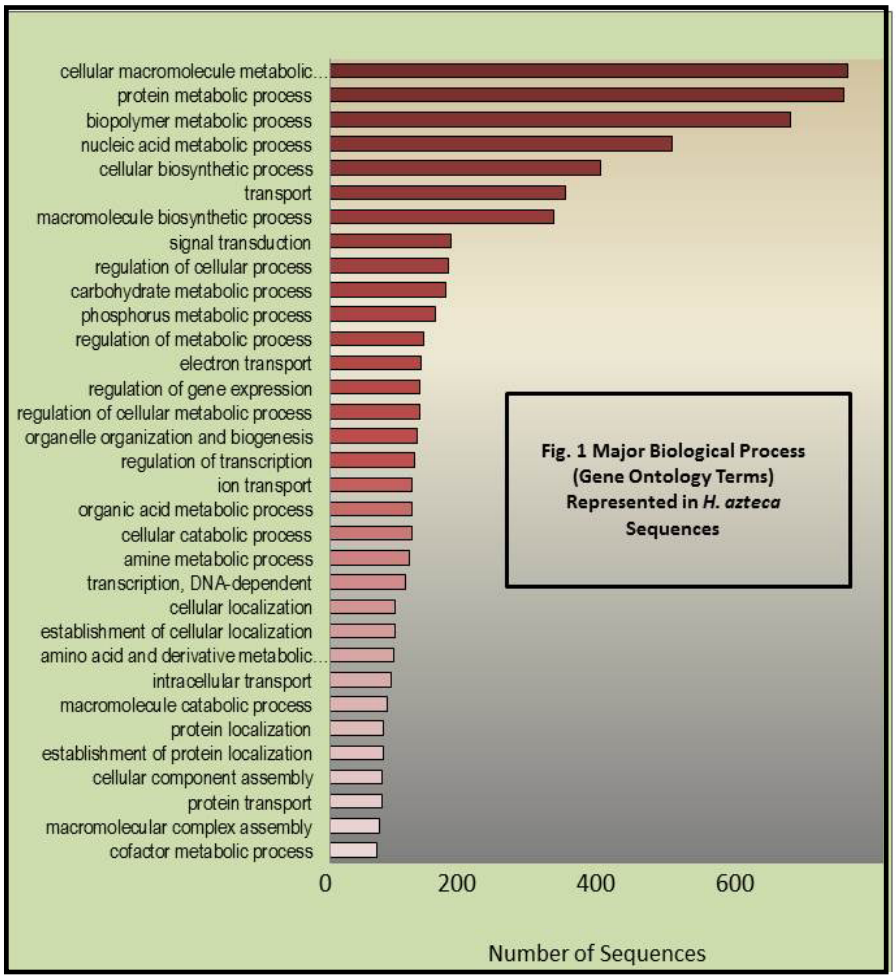
Since all organisms respond to environmental stressors by regulating the expression of genes, we propose to utilize recently developed microarrays for measuring these fundamental changes. Over the last year, we have produced over one million cDNA sequences for *H. azteca* using a recently published procedure from one of our labs that characterizes the full transcriptome from single runs of next-generation sequencers. This constitutes the largest catalog of transcribed genes for an amphipod species. In collaboration with Roche NimbleGen (Madison, WI) we have also produced a multiplex microarray containing three probes for every predicted gene for genome wide expression analysis. Our main goal now is to evaluate gene expression signatures after exposure to different sediment pollutants and to correlate these changes with whole animal toxicity. The use of microarrays will allow replacing long, expensive toxicity tests with short exposures using microarrays for the identification of DNA signatures that predict chronic toxicity.

METHODOLOGY AND RESULTS:

We conducted laboratory experiments exposing *H. azteca* to sediments spiked with

contaminants singly and in mixtures and evaluate changes on gene expression using microarrays.

Microarrays: For the modest investment of a single run of the Roche 454 genome sequencer, we assembled over one million Expressed Sequence Tags (ESTs) (1,039,832) into 65,961 contigs (overlapping sequence fragments) of which 13,113 have average lengths of 1,000 bases. Our bioinformatics pipeline then generated 147,877 long oligonucleotide (60 bp) sequences that match uniquely to the gene transcripts, which were then synthesized onto 12 sub-arrays of a single glass slide. Therefore, by processing up to four glass slides in a single day, we can generate transcription profiles from 96 independent RNA samples (4 chips x 12 sub arrays x 2 co-hybridizing samples using different fluors). **Fig. 1** summarizes the major biological processes that are represented in our array.



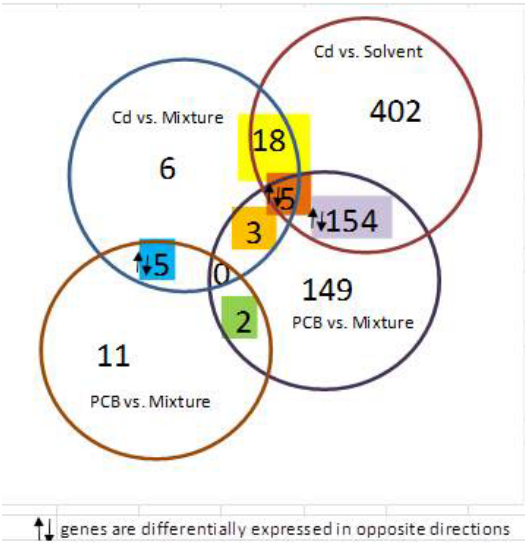
Controlled contaminant exposures: In order to validate the array, *H. azteca* were exposed to contaminants commonly associated with Great Lakes sediments. These included a heavy metal (cadmium, Cd 5.5 µg/L), a PCB (PCB 126, 7µg/L) and their mixture (at same concentrations). The concentration chosen for PCB 126 is equivalent to the invertebrate final acute value determined for Aroclor 1254 by Fuchsman et al. (2006, "An evaluation of cause-effect relationships between polychlorinated biphenyl concentrations and sediment toxicity to benthic invertebrates", Environ Toxicol Chem. 25:2601-12). The concentration chosen for Cd is equal to the experimentally determined 48-h LC_{10} for Cd in our lab. Exposures lasted for 48 hr with 6 replicates per condition and 20 juvenile *Hyalella* per replicate. **Table 1** and **Fig. 2** summarize the microarray results obtained from each of these exposures. As expected, exposure of *Hyalella* to each chemical class yielded a unique set of genes. This was particularly

true for Cd and the mixture. However, we were expecting more genes to change in expression after the PCB exposure.

Table 1. Representative differentially expressed genes for each condition with putative function.

	Log2 ratio			closest protein homolog	Putative Function	GO terms
Sequence ID	Cd	Mixture	PCB			
contig00092		2.11		A1Z8U0_DROME: CG8877-PA	pre-mRNA splicing factor	nuclear mRNA splicing, via spliceosome
contig02659		2.27		Q25BN1_CHICK: Dicer protein	Dicer protein	double-stranded RNA binding; helicase activity; ribonuclease III activity; ATP binding
contig05129	2.09	2.12		A7S3C5_9CNID: Predicted protein	AN1-type zinc finger protein	zinc ion binding
contig08539		2.11		Q7Q2Y5_ANOGA: ENSANGP00000011363	Short stop/Kakapo long isoform	calcium ion binding; cell cycle arrest
contig09877		2.12		Q17GY9_AEDAE: Kakapo	Short stop/Kakapo long isoform	calcium ion binding; cell cycle arrest
contig08305			1.39	none found	unknown	none found
contig08319			1.21	none found	unknown	none found
contig18554	3.11	2.37		Q17CE9_AEDAE: Histone deacetylase	Histone deacetylase	histone deacetylase activity; regulation of transcription, DNA-dependent
contig23551		2.32		Q16KQ8_AEDAE: Putative uncharacterized protein	Notch-like protein	serine-type endopeptidase inhibitor activity
contig25551	2.12	2.07		Q16EB4_AEDAE: Sodium/chloride dependent amino acid transporter	Sodium-dependent nutrient amino acid transporter	neurotransmitter:sodium symporter activity neurotransmitter transport
contig30217		2.33		PUM_DROME: Maternal protein pumilio	Pumilio	RNA binding
contig31498		2.06		Q291U7_DROPS: GA21822-PA	glycogen biosynthesis	4-alpha-glucanotransferase activity; glycogen biosynthetic process
contig33751	2.26	2.19		DNK_DROME: Deoxynucleoside kinase	Deoxynucleoside kinase	ATP binding; nucleic acid metabolic process; phosphotransferase activity, alcohol group as acceptor
contig47432		2.08		Q86GD6_PROCL: Projectin	Projectin	nucleotide binding; protein serine/threonine kinase activity; protein phosphorylation
contig48146	3.82	3.86		Q8ITL4_9DIPT: Heat shock cognate 70	Heat shock protein 70	ATP binding; response to stress
contig48593	2.17	2.17		A5HL62_ORYLA: Heat shock protein 70 isoform 5	Heat shock protein 70	ATP binding; response to stress
contig49974	3.18	3.00		Q6QR01_9EUCA: Hsp-90	Heat shock protein 90	ATP binding; protein folding; unfolded protein binding
contig50642	3.56	3.52		A5PMG2_DANRE: Novel protein	Heat shock protein 70	ATP binding; response to stress
contig51807		2.10		A7T022_9CNID: Predicted protein	NFX1-type zinc finger-containing protein 1	sequence-specific DNA binding transcription factor activity; regulation of transcription, DNA-dependent; zinc ion binding
contig55979	3.07	2.91		Q6SXP5_LOCMI: Heat shock protein 90	Heat shock protein 90	ATP binding; protein folding; response to stress; unfolded protein binding
contig57132	2.28	2.23		O73922_OREMO: Heat shock protein 70	Heat shock protein 70	ATP binding; response to stress
contig63196	2.57	2.54		A0ZT12_COTJA: Heat shock protein 70kDa	Heat shock protein 70	ATP binding; response to stress
contig65761	2.96	3.05		A5A8V7_PIG: Heat shock 10kDa protein 1-like	Heat shock 10kDa protein 1-like	ATP binding; response to stress

Fig. 2. Number of differentially expressed genes by each exposure.



Overlap between Cd vs. Sol and PCB vs. Mix (opposite directions) = differentially expressed by Cd and differentially expressed in the mixture. These are Cd response genes whose signature is preserved in the mixture (154 genes)

Overlap between PCB vs. Sol and Cd vs. Mix (opposite directions) = differentially expressed by PCB and differentially expressed in the mixture. These are PCB responsive genes whose signature is preserved in the mixture (5 genes)

Overlap between PCB vs. Sol and PCB vs. Mixture = uniquely differentially expressed by PCB. These are PCB responsive genes which lose responsiveness in the mixture. (2 genes)

Overlap between Cd vs. Sol and Cd vs. Mixture = uniquely differentially expressed by Cd. These are Cd responsive genes which lose responsiveness in the mixture (18 genes).

Overlap between Cd vs. Mixture and PCB vs. Mixture (and not in Cd vs. Sol) = uniquely differentially expressed (down-regulated) by the mixture. These are genes that are responsive only in the presence of both Cd and PCB.

Overlap between Cd vs. solvent and Cd vs. Mixture and (in opposite directions) PCB vs. mixture = differentially expressed by Cd and differentially expressed in the mixture (but at about 1/2 the level). These are Cd response genes whose signature is lessened in the mixture, but still significant (5 genes)

EXPECTED IMPACT:

We have developed a tool that could be applied to determine whether *Hyalella*, a commonly used benthic model organism, respond to contaminants commonly present in AOCs across the Great Lakes. Our approach will help address the goal of improving Great Lakes ecosystem health and water quality by providing information on sediment toxicity and informing the risk assessment process.

There are three main areas where this project is of major significance. First, the proposed work will greatly facilitate the development of a streamlined sediment evaluation approach. The expense involved with removing or treating contaminated sediments requires an accurate characterization of risks. Currently, the standard approach is to evaluate sediment toxicity by measuring effects on *H. azteca* reproduction, which can be cost-prohibitive (> $5,000/sample) and very time consuming (each test lasts 42 days). A gene microarray chip based on *H. azteca* can eventually be used as a surrogate for chronic 42-d toxicity tests. Second, the development of gene arrays for a distinctive invertebrate species will enhance efforts to better extrapolate results from toxicity studies on test species to other organisms. The extrapolation of results observed in one test species is often problematic when there is no information as to the mechanism by which a chemical induces a certain response. Gene expression profiles may be able to define global biochemical pathways that are affected and thus clarify the mechanisms by which compounds act. Finally, this proposal has provided the first sequence survey of genes responding to environmental stressors in this sentinel species. To date, there are no sequence data for nuclear genes for this amphipod within public databases.