Technical Report 0660

Application of Gene Expression Analysis for Sediment Toxicity Stressor Identification

Steven M. Bay[1], Chris D. Vulpe[2], Alexander V. Loguinov[2] and Audrey Arai[2]

[1] *Southern California Coastal Water Research Project, Costa Mesa, CA*

[2] *University of California, Department of Nutritional Sciences and Toxicology, Berkeley, CA*

EXECUTIVE SUMMARY

Sediment toxicity tests are frequently used to assess sediment quality along the west coast of North America. However, the use of sediment toxicity information to assist in the development of management actions (e.g., source control, sediment remediation) is limited by the difficulty of determining which contaminants are responsible for the toxic effects. An investigative process known as a Toxicity Identification Evaluation (TIE) is one approach commonly used to determine the toxicants present in environmental samples. However, the TIE process can be expensive, time consuming, and may yield a nonspecific result due to low toxicity or a lack of analytical methods. These constraints limit the application of TIEs, resulting in uncertainty regarding the cause of sediment toxicity in coastal water bodies such as San Francisco Bay.

This report describes the results of a project to investigate the feasibility of a new sediment TIE approach, known as a molecular TIE. This approach uses changes in gene expression of the toxicity test organism, rather than chemical manipulation, as a method for identifying the cause of sediment toxicity. The research utilized a newly-developed gene microarray for the estuarine amphipod *Eohaustorius estuarius,* a widely used sediment toxicity test organism. The project was designed to accomplish three objectives:

- **Determine gene expression in amphipods exposed to diverse contaminant types.**
- **Develop tools for toxicant identification.**
- **Evaluate toxicant identification ability.**

Samples of *E. estuarius* from a variety of sediment and water toxicity tests were analyzed using the microarray. A training data set consisting of a total of 16 different chemical treatments, each with matching controls, was analyzed. The treatments represented a diversity of contaminants of concern, such as trace metals, chlorinated pesticides, petroleum hydrocarbons, and current use pesticides. The RNA from each replicate was extracted and analyzed using the microarray. The gene expression results for the exposed samples were compared to matching controls in order to calculate differential gene expression (i.e., ratio of expression in test sample relative to control).

Two preliminary approaches for classifying unknown samples were developed using the training data: cluster analysis and class prediction models based on random forest analysis. The effectiveness of the classification methods was tested by analyzing independent evaluation samples having different chemical characteristics.

Each training sample analyzed showed evidence of differential gene expression. Both upregulation and down-regulation of genes were evident in all samples. A total of 3182 microarray probes with statistically significant differential gene expression (i.e., candidate genes) were identified in the training samples. Between 12% and 54% of the candidate genes in a treatment group were uniquely expressed (i.e., had significant differential expression in only one chemical treatment group). The magnitude of differential

expression for individual gene probes also varied widely among the samples, with extremes ranging from 40-fold down-regulation for fipronil to 100-fold up-regulation for DDT.

Cluster analysis identified seven clusters of the 16 training sample types, indicating that characteristic patterns of gene expression were present among the samples. Three clusters were composed of single chemicals (copper 750, DDT, and fipronil), suggesting that these chemicals produced very distinctive patterns of expression. The remaining clusters each contained 2-4 chemical treatments of variable composition. For example, one cluster was comprised of samples exposed to pyrene, chlordane, and cadmium. Samples representing different exposure concentrations of the same chemical usually grouped in the same cluster, suggesting consistency of gene expression patterns within chemical type.

Three independent evaluation samples (not part of the training data set) were used to test the ability of the candidate molecular TIE methods to identify the cause of toxicity. One sample (T1) was from a spiked sediment exposure to cyfluthrin, a pyrethroid pesticide. A second sample (T2) was from a field site in southern California where TIEs had determined the cause of toxicity to be pyrethroid pesticides. The third evaluation sample (T3) was from RMP monitoring in San Francisco Bay and the cause of toxicity was not known. Both of the field sediment samples were contaminated by a complex mixture of trace metals and trace organics.

Cluster analysis yielded limited success in classifying the evaluation samples. Sample T1 (cyfluthrin exposure) was grouped into the first cluster, which also contained samples exposed to bifenthrin, ammonia, and pyrene. It was expected that sample TI would cluster with the other cyfluthrin samples, which were grouped nearby in a separate cluster. However, bifenthrin is a pyrethroid pesticide that is chemically similar to cyfluthrin and has a similar mode of toxic action. Sample T2 (pyrethroid contaminated) was grouped into a cluster with two other chemicals: chlordane and pyrene. Neither of these chemicals has a chemical similarity to pyrethroids, which were identified by TIE as the cause of toxicity at this site.

Results for the third evaluation sample (T3) were similar to those for T1; this samples was grouped into the same cluster associated with bifenthrin, ammonia, and pyrene. There was no a priori expectation regarding the cause of toxicity in this sample. However, both PAHs (such as pyrene) and ammonia have been suggested as contributing to sediment toxicity in portions of San Francisco Bay.

More accurate classification results were obtained using a class prediction model developed using random forest analysis. This analysis developed a prediction model based on three stressor classes: pyrethroid pesticide, trace organic (other than pyrethroids), and other (e.g., metals, ammonia). Class predictions using the evaluation samples gave results consistent with expectations. Replicate samples of T1 and T2 were correctly classified into the Pyrethroid category 100% of the time when average response values were used. Variable results were obtained for sample T3 using random forest, with classification into both the Pyrethroid and Trace Organic categories.

This project has achieved several important milestones in the effort to develop a molecular TIE method for sediment toxicity. First, a gene microarray for *E. estuarius* was successfully developed based on the first ever sequencing of RNA from this amphipod. Second, this microarray was used to identify a subset of candidate genes having differential expression in response to 11 different types of chemical exposure. Third, many of the chemical treatments were shown to produce distinctive patterns of differential expression, confirming a key assumption of the approach. Finally, we developed and applied multiple approaches for evaluating unknown samples. Some of these approaches were shown to correctly identify the cause of toxicity in independent evaluation samples, providing a demonstration that a molecular TIE approach is feasible and has the potential to provide an effective and powerful tool for sediment TIEs.

Several data gaps and areas of uncertainty need to be addressed before this molecular TIE approach can be used with confidence in monitoring programs. A primary data gap is the lack of gene expression information for many contaminants of concern. Contaminant groups particularly underrepresented in the current training set include PAHs, trace metals, and PCBs. It is possible that greater discrimination between the evaluation sample from San Francisco Bay (T3) and the other evaluation samples would have been obtained had the training data set included these additional constituents.

Standardization of sample preparation and analysis methods is needed to develop a reliable tool for use in multiple laboratories. There are few standard protocols for sample preservation, extraction, analysis, and quality assurance, especially for marine invertebrates. Consequently, data for the same sample analyzed by different laboratories may vary as a result of method variations, possibly leading to incorrect conclusions regarding the gene expression patterns obtained. Interlaboratory studies are needed to investigate these issues and develop methods that yield comparable results among laboratories. This issue is of importance to all TIE methods, as the methods must be accurate and reliable in order for end users to have confidence in their use for guiding environmental management actions that may have high cost.

Finally, further development and evaluation of classification models based on gene expression is needed. This study investigated only two types of statistical approaches for stressor identification: clustering and random forest analysis. Alternative data analysis methods should be investigated to help in determining the most accurate approach for TIE applications. A molecular TIE method, like other stressor identification methods, must be accurate and reliable in order for the results to be accepted by management agencies and used to guide potentially costly clean up and control actions.

The research conducted in this study has laid the foundation for development of a new and potentially effective TIE approach with wide application. While further development and validation is needed, this project has demonstrated that rapid progress can be made in these areas through collaboration and partnership.

Full Text

ftp://ftp.sccwrp.org/pub/download/DOCUMENTS/TechnicalReports/660_GenomicStressorID.pdf