Integration of 'omics technologies for characterization of complex microbial ecosystems - 8119

A.L. Miracle, V.L. Bailey, S.E. Baker, A.L. Bunn, J. Magnuson, B.J. Webb-Robertson
Pacific Northwest National Laboratory
P. O. Box 999, Richland, WA 99352

ABSTRACT

Environmental remediation sites possess complexity at both biotic and abiotic levels, with temporal shifts that are associated with the interaction of both systems. Single scientific disciplines no longer serve to address and understand the complex nature of these sites. An integration of physical, chemical, and biological characterizations or a more inclusive environmental systems approach is needed. As a proof of concept, an integrated approach was developed to identify biosignatures from a complex environmental community to provide sensitive, early indicators and predictors of response to radionuclide and chemical exposures of interest for environmental management. Specifically, the integrated approach used a combination of genetics, transcriptomics, proteomics, and computational statistics to characterize a complex periphyton community following uranium exposure. This additional, specific information has promise to further reduce uncertainties in environmental remediation and monitoring in realizing the potential impacts of exposure in advance to reduce potential mitigation efforts. Results from this study establish a basis for biosignature characterization of any ecosystem for comparison or monitoring of biotic exposure and effects in response to a specific contaminant.

INTRODUCTION

Environmental Monitoring

Through various activities, mankind has altered the environment to a degree where these sites now require remediative and monitoring programs**.** Current environmental monitoring methods include point source measurements for contaminants or chemicals of interest. However, sensitivity of detection methods, dilution of contaminants in various (aquatic) media, and shifts in chemical speciation all make accurate and timely risk decision capability difficult. This issue is compounded further by the heterogeneity of the environment as the chemicals of concern are found within a complex mixture of other chemicals, and cross different substrates.

The purpose of environmental monitoring from a regulatory perspective is to provide opportunities to determine if a contaminant is present and bioavailable in the environment in advance of irreparable damage or harm to human and environmental health and safety. This requires tools and guidance to assess relative impacts and an understanding of allowable levels of change. Traditionally, risk endpoints such as growth, survival, and reproduction are the metrics used to determine maximum exposure limits, and thus these levels for individual chemicals are used as threshold values.

Unfortunately, mixtures of contaminants with unknown synergistic or antagonistic interaction do not lend themselves to a single chemical threshold level, and may fail to determine biological availability, accumulation, and adverse effects (Relyea 2003; Borgert *et al*. 2004; Gust *et al.* 2005; Relyea 2005; Gust *et al.* 2006). Environmental conditions such as season, precipitation, and population densities are also factors that can affect interpretation. These compounding influences indicate that additional, ecosystem-inclusive approaches to monitor exposure and effects are needed. Rather than stopping at data that only indicates an impacted environment, gathering information that also provides a mechanistic understanding of how the environment is being impacted will afford better protection of human and environmental safety.

Biomarkers

By definition, a biomarker is any physiological measure used to indicate a condition or disease. The degree of accuracy in determining the diagnostic or predictive value of biomarkers lies with having the understanding of what mechanistic processes are indicated by that particular biomarker. For instance, blood pressure is used as a gross biomarker for heart disease. The diagnostic resolution of using arterial pressure is less than other biomarkers for heart disease, and is a weak predictor of heart condition in and of itself. Biological endpoints are currently used in environmental monitoring and include measurements at the single cell, individual tissue and the whole animal level (van der Oost *et al.* 2003; Venturino *et al.* 2003; Galloway *et al.* 2004; Broeg *et al.* 2005). Therefore, the use of "biomarkers" in environmental monitoring is not a unique concept. However, current biomarkers used for environmental monitoring have significant limitations. Like the blood pressure example, some currently used biomarkers used for classes of compounds (like metals), and are not individual chemical-specific (Galloway *et. al.* 2004). These markers also lack the discernment of temporal effects and may not be able to distinguish adaptation or be dose-responsive (Marques et al 2007).

The inclusion of ecosystem-level molecular approaches will add value to the current approaches used for environmental monitoring. Molecular tools (described in the next section) are not intended to replace existing monitoring approaches, but will provide complementary information that will best inform a weight of evidence approach for environmental monitoring. Molecular measures can aid monitoring programs at different levels. Discovery of novel biomarkers from ecosystem-level exposure studies provides a near-term capability to assess exposure, bioavailability, and population effects in the context of changing environmental conditions, and may serve as a rapid screen for assessing relevant contaminant exposure in sentinel species. This discovery process is a robust approach in that all metabolic pathways, genes, and proteins are screened to provide the most informative sets of biomarkers that can be cross-validated between methods and traditional endpoints such as bioaccumulation and population structure.

MOLECULAR TOOLS

Population Characteristics

Genome analysis (genomics) is the functional study of genetic composition or variation within a population, leading to identification of sequence variability (i.e., polymorphisms) associated with a phenotype/response of interest. Genomics provides a functional link between genotype and individual variability within a population. Given that the genetic code defines the heritable attributes of individuals, sequencing of genomes offers insights into individual response differences to contaminants and associations between genotype and phenotype. From an environmental perspective, complex communities such as microbial communities, can be fingerprinted and observed through time for changes in population frequency by tracking the unique genetic fragments that are species specific from a number of different approaches.

Terminal restriction fragment length polymorphism analysis relies on the high resolution of automated sequencing to discern ribosomal fragments that are unique to individual species and is thus capable of providing information on community structure and composition (Marsh 1999; Kitts 2001). This method has been used extensively as a microbial population characterization technique for community and ecotoxicological assessments. A similar approach is denaturing gradient gel electrophoresis (DGGE) allows assessment of changes in microbial (prokaryotic and eukaryotic) community structure between conditions or treatments and between time points (Muyzer et al, 1993). Both methods allow for sequencing and identification of predominant species or species of interest within samples. Briefly, DNA extracts from all samples undergo a polymerase chain reaction (PCR) amplification of the 16S (prokaryotic) or 18S (eukaryotic) genes using the appropriate PCR primers. The amplified product is separated by electrophoresis and the resulting pattern of bands down the length of the gel lane is representative of species within the community. Changes in banding pattern or the number of bands between samples reflects changes in complexity of community membership.

Phospholipid fatty acid (PLFA) analysis examines lipid components of membranes that differ between phylogenetically or metabolically defined groups (Hedrick et al. 2000). PLFAs are essential components of cell membranes. Within minutes of cell death, enzymes start to degrade the phospholipids (White 1994), so PLFA profiles are primarily the product of cells present in the living, functional biomass. The quantity of PLFA in a sample relative to other samples provides a viable biomass comparison. Using standard equations, total cell counts can be derived from PLFA data. Differences in PLFA structure provide broad phylogenetic information and functional group membership (sulfate reducers, anaerobic metal reducers, methane oxidizers, etc.). PLFA data can also provide information on physiological status of some community members, specifically gram negative bacteria, as cells with high metabolic activity contain different fatty acids than do those with low metabolic activity or those that are stressed.

While the use of single nucleotide polymorphisms (SNPs) has been predominately used in human genetics to understand the etiology and pathology of human susceptibility to disease, the use of genetic variation as subpopulation indicators can be tied to functional changes in response to environmental stress and historic exposure (Schmid et

al, 2006). For environmental use, this capability will rely on well characterized genome information, or loci unique to metabolic processes adapted for extreme environments.

Metagenomics is a fairly new approach that has the potential to elucidate the vast amount of genome information present in uncultured, or environmental, microorganisms. Because of DNA sequencing advances, these samples can be rapidly sequenced, resulting in a heterogeneous collection of genome fragments representative of the mixture from which it was derived (Tringe and Rubin, 2006). While this is a great advantage over methods that require pure cultures, there are complicating factors such as annotation for identification and attribution of information, and the lack of discernment between living and dead microbes.

One technique to circumvent the annotation issue is the construction of phylogenetic arrays. Phylogenetic microarrays can be targeted to identify community membership within a given sample (Flanagan et al, 2007). Such a tool with information on a high density microarray is widely applicable for assessing the diversity of any ecosystem with regard to the taxonomy of the prokaryotic or eukaryotic microbes present. For a taxonomic microarray, sets of probes are developed for the widely used ribosomal ribonucleic acid (RNA) genes in order to detect and classify microbes present in a sample. The principal challenge in building a phylogenetic microarray is retrieving, aligning and analyzing thousands of gene sequences from public databases to develop the oligonucleotide probes for the chip. For assessing the eukaryotic taxa present, we are currently using the ITS1/5.8S/ITS2 region of the ribosomal RNA genes to represent a broad taxonomic spectrum of fungi, green algae and diatoms. The relatively high conservation of these ribosomal RNA sequences through evolutionary time makes them the preferred target for molecular classification of organisms.

Gene Expression

Gene activation and transcription are assessed through multiple techniques that can be referred to collectively as transcriptomics methods. Measuring changes in gene expression over time and across environmental conditions provides critical information regarding the kinetics and coordination of gene expression that contributes to the dynamic processes of cellular homeostasis and toxicity. Analyzing gene expression data across multiple samples or responsive species also reveals underlying similarities among different conditions, thus producing correlates of gene behavior that can be used to predict and diagnose cellular responses to contaminants of concern.

Microarrays are a well-established technology that provides the greatest breadth of global gene expression discovery. For select species, extensive annotation of the genes represented on these arrays continues to increase and is another significant advantage that facilitates the conversion of microarray data into mechanistic and pathway knowledge to support risk assessment. However, full utility of microarray technology to risk assessment is limited by a number of factors including the number of different platforms (e.g. cDNA arrays and oligonucleotide arrays) and multi-center cross validation studies that indicate significant inconsistencies in data between laboratories. Use of

microarray technology to investigate environmental microbial communities is limited by the level of known sequences for community members. However, scientific advances continue to improve on detection limits, increase specificity and quantification of signal, and provide a high throughput platform to assess environmental samples (reviewed in Wagner et al, 2007).

In the absence of annotated sequences, cDNA libraries and their variants are still a viable option for obtaining information regarding patterns of expression change within a complex community. With DNA sequencing technology growing more rapid and affordable, the ability to assess a community or select organisms by sequencing the expressed genome is now a reality (Raes et al, 2007). Further manipulation of libraries containing "global expression" content can be achieved with suppression subtraction to isolate the unique genetic responses of different microbial community samples. This approach narrows the pool of sequences to relevant genes that are differentially expressed. Comparisons of raw sequences still rely on the growing databases of annotated sequences for probable function by homology. However, pattern expression is still a useful tool in understanding relative changes between environmental conditions.

Protein Expression

Proteomics refers to the global analysis of expressed proteins within an organism. The protein component has been estimated to be at least one order of magnitude larger than the genome and includes proteins that result from alternative splicing and post-translational modifications. The identification of proteins within the proteome depends first and foremost upon the availability of annotated sequences. While this may appear to be a limitation for assessing protein expression in environmental samples, genome sequencing and annotation technologies continue to provide new data that help to improve the identification of differentially expressed proteins. Both gel-based and non gel-based methods have been developed and each requires high-resolution separations in one or more dimensions.

Current global proteomic strategies are critically dependent on the existence of a genome annotation database for each organism under study. This currently limits uncovering changes in ecosystem proteins by high-throughput protein identifications and potential biomarker discovery to systems where the organisms have compiled genome annotations. However, by estimating peptide presence and abundance profiles without regard to initial peptide identification, patterns of expression can be assessed. The ecosystem-level putative peptide profiles would be obtained using current proteomic identification algorithms using a comprehensive non-redundant protein database to match experimental and virtual/database generated spectra. A suite of internal standards will be utilized for sample- to-sample and run-to-run normalization. Thus, a genome-specific annotation that is unattainable for almost all ecosystem communities would not be needed to compare changes in the ecosystem protein profile across conditions.

PERIPHYTON CASE STUDY

Experimental Design

As a proof of concept, a focused uranium dose-response in periphyton experiment was performed to demonstrate the usefulness of the integrated molecular approach. The relevant hypothesis will address if the periphyton community responds differently to the presence of uranium in terms of diversity and metabolic function. Periphyton communities can be used as monitors of ecosystem health and as indicators of contamination in lotic systems. In a previous study, measures of biomass, community structure, and genetic diversity were used to investigate impacts of uranium (U) exposure on periphyton (Small et al, in press) up to 100 $\mu g\ L^{-1}$ over 5 days of exposure. Results from this study indicated that the dose and duration of U exposure did not alter population structure or cause cessation of growth of a natural, complex community despite reports to the contrary for single algal cultures that had a minimum detectable effect at 1.7 $\mu g\ L^{-1}$ U (Charles et al. 2002). Water hardness and pH are variable factors that can significantly affect U exposure growth metrics which may account for the differences between single species lab cultures and natural river water complex communities. However, the significance of these exposures in setting criteria for minimum detectable effects should reflect the natural environment, and not a contrived, sterile setting that has little bearing on environmental exposure scenarios. To that effect, an examination of U exposures to a natural complex community was performed for interrogation at multiple levels of molecular organization, across phyla, to assess the relevant impact of low, medium, and high U concentration exposures to a valuable environmental sentinel.

The experimental design for the periphyton uranium uptake study will include the following:

- The exposure phase of the study consisted of four replicate blocks of tanks at four uranium treatment concentrations of 0 (un-amended river water), 10, 100, and 500 µg/L uranium
- Samples taken from each treatment block (Figure 1):
    - At times 0, 24, and 120 hours for molecular project analyses as described below. Samples are homogenized, split into four aliquots, and preserved for analyses.
    - At times 0, 24, 48, 72, and 120 hours for biomass by gravimetric analysis, concentration of the uranium in the tissue and water, and water quality (pH, temperature, dissolved oxygen, hardness, and alkalinity).

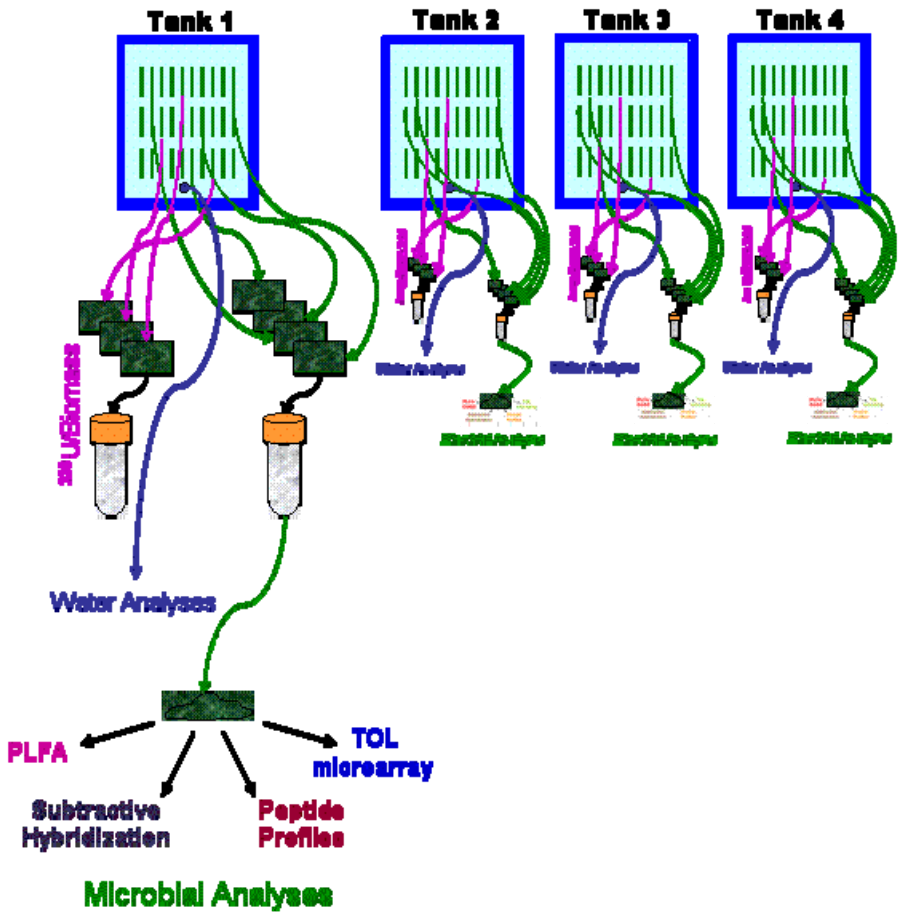

Figure 1. Diagram of periphyton sampling plan following uranium exposure.

Molecular Projects

Uranium in the water and periphyton were analyzed using ICP-MS. Periphyton samples were also analyzed for gravimetric biomass and phospholipid fatty acids (PLFA). Water was analyzed for standard water quality parameters, including pH, DO, hardness, and alkalinity. Uranium accumulated in rapidly in the periphyton exposures. Uptake of uranium by the periphyton biofilm reached an apparent equilibrium in 48 hrs, and the uptake was concentration dependent. There was significant accumulation of uranium in the biofilm compared to the water concentration. The bioconcentration factor ranged with the exposure concentration from 6 to 17. However, there was no significant difference in any biomass analysis (gravimetric or PLFA) based on treatment or time. PLFA indicated that the periphyton community was dominated by prokaryotes, and diatoms were the most common of the eukaryotic community. However, there was no significant change in PLFA with time and uranium concentration to indicate community structure changes. These results were significant because uranium has been shown to inhibit growth of *Chlorella* sp. at concentrations < 5 µg L-1 at similar pH and water hardness in comparison to this study (Franklin et al. 2000; Charles et al. 2002).

Peptide patterns are being assessed using LC-MS/MS, current proteomic identification algorithms (e.g., Sequest) and a non-redundant protein database to match experimental and database generated spectra. Computational statistics and software has been generated to automate comparisons of MS/MS spectra from multiple samples, enabling spectra not identified by Sequest to be utilized in a fingerprint-type analysis. The proteomic peptide profiling is being used to characterize changes in the protein profiles of the uranium-exposed periphyton samples versus controls.

Subtraction hybridization-PCR is being used to isolate unique genetic responses of periphyton community samples exposed to uranium. The resulting biosignatures of uranium influence will be sequenced, and the data will be used to identify patterns (unsupervised and supervised learning) leading to robust characterization of uranium influence and response by the periphyton community.

A fungal and algal taxonomic microarray is being used to assess the complex periphyton community itself for relative population level changes between uranium exposure conditions. The resulting biosignatures will be indicative of uranium exposures to the fungal/algal component of the periphyton community.

Data Integration

Once the samples are abstracted to the level of ecosystem profiles, a wide variety of statistical methods may be applied. The estimated peptide profiles closely resemble (from the perspective of statistical analysis) the data (fingerprints, profiles) accumulated by other biological fingerprinting techniques. Consequently, advances in statistical fingerprinting developed in these problems may be applied in a straightforward fashion to this problem. Each of the molecular data sources described above captures an important component of the periphyton system – each a snapshot of a complex and highly dynamic system. To integrate these heterogeneous data sources we employ a Bayesian statistical approach. Each data source is described as a probability model. Based on the data from the molecular results, the statistical model describing the probability of observing a specific biosignature pattern given a specific Uranium dose (U_i) and time after exposure (T_k) can be derived; $P_{Method}(S \mid U_i, T_k)$. Since this transformation places all the data sources on the same scale, Bayes Rule can be used to derive the posterior probability of interest, the probability of a specific Uranium dosage and time after exposureas given in equation 1.

$$P(U_i, T_k \mid S) \approx P_{PLFA}(S \mid U_i, T_k) * P_{TOL}(S \mid U_i, T_k) * P_{PEP}(S \mid U_i, T_k) * P_{SH}(S \mid U_i, T_k) \quad \text{(Eq. 1)}$$

Functional characterization of changes in the periphyton community in response to uranium exposure will require use of orthogonal and evolutionary information. The community proteomics, fungal/algal array, and the subtractive hybridization approaches will generate sequence-associated data. These sequences can be associated with functional pathways or general functional categories using a number of different comparative genomics techniques. Like the compositional analysis the Bayesian

approach will allow a detailed analysis of the contribution of a particular identified function to the overall response of the community.

CONCLUSIONS

Given that environmental samples are a mixture of communities of unknown concentration with unknown levels of interactions, the ability to interrogate a single sample with multiple technologies will allow us to perform an ecological evaluation at a "systems" level. How these communities change in species, concentration and function is of specific interest, but no single technology can answer all these questions. Having an integrated assessment of relative change allows for comparisons at temporal and spatial scales. These factors are increasingly more important in determining impacts of anthropogenic remediation or monitoring strategies in a timely and relevant manner.

REFERENCES

Borgert C.J., T.F. Quill, L.S. McCarty, and A.M. Mason. 2004. Can mode of action predict mixture toxicity for risk assessment? *Toxicol. Appl. Pharmacol.* 201: 85-96.

Broeg K, H.V. Westernhagen, S. Zander, W. Korting, and A. Koehler. 2005. The ''bioeffect assessment index'' (BAI) A concept for the quantification of effects of marine pollution by an integrated biomarker approach. *Mar Pollut Bull* 50:495–503.

Flanagan, J. L., E. L. Brodie, L. Weng, S. V. Lynch, O. Garcia, R. Brown, P. Hugenholtz, T. Z. DeSantis, G. L. Andersen, J. P. Wiener-Kronish, and J. Bristow. 2007. Loss of Bacterial Diversity during Antibiotic Treatment of Intubated Patients Colonized with *Pseudomonas aeruginosa*. *J Clin Microbiol.* 45(6):1954–1962.

Galloway T, R. Brown, M.A. Browne, A. Dissanayake, D. Lowe, M.B. Jones, and M.H. DePledge. 2004 A Multibiomarker Approach To Environmental Assessment. *Environ. Sci. Technol.*, 38:1723-1731.

Gust K.A., and J.W. Fleeger. 2005. Exposure-related effects on Cd bioaccumulation explain toxicity of Cd-phenanthrene mixtures in *Hyalella azteca*. *Environ Tox Chem* 24:2918-2926.

Gust K.A., and J.W. Fleeger. 2006. Exposure to Cadmium-Phenanthrene Mixtures Elicits Complex Toxic Responses in the Freshwater Tubificid Oligochaete, *Ilyodrilus templetoni*. *Arch Environ Contam Toxicol. 51: 54-60.*

Kitts, C.L. 2001. Terminal restriction fragment patterns: a tool for comparing microbial communities and assessing community dynamics. *Curr Iss Intest Microbiol* 2(1):17-25.

Marques, C.C., A. Sánchez-Chardi , S.I. Gabriel , J. Nadal, A.M. Viegas-Crespo, and M. da Luz Mathias. 2007. How does the greater white-toothed shrew, *Crocidura russula*, responds to long-term heavy metal contamination? -- A case study. *Sci Total Environ.* 376(1-3):128-33.

Marsh, T.L. 1999. Terminal restriction fragment length polymorphism (T-RFLP): an emerging method for characterizing diversity among homologous populations of amplification products. *Curr Opin Microbiol* 2:323-327.

Raes, J., K.U. Foerstner, and P. Bork. 2007. Get the most out of your metagenome: Computational analysis of environmental sequence data. *Curr Opin Microbiol* Oct 11 Epub: http://www.sciencedirect.com/science/journal/13695274

Relyea, R.A.. 2003. Predator cues and pesticides: a double dose of danger for amphibians. *Ecolog Appl* 13:1515–1521.

Relyea, R.A. 2005. The Lethal Impacts of Roundup and Predatory Stress on Six Species of North American Tadpoles. *Arch. Environ. Contam. Toxicol*. 48: 351–357.

Schmid, K.J., O. Törjék, R. Meyer, H. Schmuths, M.H. Hoffmann, T. Altmann 2006. Evidence for a large-scale population structure of *Arabidopsis thaliana* from genome-wide single nucleotide polymorphism markers. *Theor Appl Genet*. 112(6):1104-1114.

Small, J.A., A.L. Bunn, C. McKinstry, A. Peacock, and A.L. Miracle. 2007. Investigating freshwater periphyton community response to uranium with phospholipid fatty acid and denaturing gradient gel electrophoresis analyses. *J Environ Radioact* in press.

Tringe, S.G., and E.M. Rubin. 2006. Metagenomics: DNA sequencing of environmental s amples. *Nat Rev Genet* 6(11):805-814.

van der Oost, R., J. Beyer, and N. Vermeulen. 2003. Fish bioaccumulation and biomarkers in environmental risk assessment: a review. *Environ Toxicol Pharmacol* 13:57-149.

Venturino, A., E. Rosenbaum, A. Caballero de Castro, O.L. Anguiano, L. Gauna, T. Fonovich de Schroeder, and A.M. Pechen de D'Angelo. 2003. Biomarkers of effect in toads and frogs. *Biomarkers*. 8:167-186.

Wagner, M., H. Smidt, A. Loy, and J. Zhou. 2007. Unravelling microbial communities with DNA-Microarrays: Challenges and future directions. *Microb Ecol* 53(3):498-506.