

## Omics: How Not to Make Your Site a Science Project

*Dora Taggart* (dtaggart@microbe.com), Sam Rosolina, and Katherine Clark (Microbial Insights, Knoxville, TN, USA)

**Background/Objectives.** As high-throughput molecular technologies, including metagenomics and proteomics, are becoming more affordable, the costs associated with storing and analyzing the large amounts of data produced may soon overtake the investment in generating raw data. While these more novel analyses can be exciting to utilize, it is important to recognize that the resulting data will need to be distilled, with a plan in place for filtering the important pieces of information from the noise. Laboratories offering these services generally will use an established pipeline to translate the raw data into microorganisms, genes, or protein identifications, but additional analysis and expertise is still needed to extract meaning from the data set and interpret it within the context of the study. Without this additional effort, the analysis loses much of its utility and the data can be easily overinterpreted. Through case studies and examples of data output, this presentation aims to highlight the power of metagenomic and proteomic analyses and the expected impact that they will have on the environmental remediation field.

**Approach/Activities.** Although brief comprehensive reviews of metagenomics and proteomics are discussed within this study, the focus is given to the resulting data and approaches to data analysis. Next generation sequencing (NGS) results of 16S rRNA genes from a petroleum pipeline release currently undergoing monitored natural attenuation will be used to illustrate the importance of proper study design, sampling strategy, and statistical analyses. Groundwater samples were periodically obtained from background and monitoring wells within the dissolved plume for a period of two years. Principal component analysis (PCA), hierarchical clustering dendrograms, and diversity indices were used to elucidate differences in microbial community composition and any changes over time.

With advances in mass spectrometry, the detection or quantification of specific proteins (targeted environmental proteomics) will become more widely commercially available. Because the enzyme is the catalyst that actually mediates the biological reactions responsible for contaminant biodegradation, quantification of the target protein is believed to be the most direct indicator of activity. Current and future applications for proteomics in environmental remediation will be discussed.

**Results/Lessons Learned.** While the state of metagenomic and proteomic technologies is impressive, the importance of data analysis cannot be overlooked. With traditional molecular biological tools, such as quantitative polymerase chain reaction (qPCR), assays have been developed and validated to specifically target microorganisms or genes of interest. However, metagenomics necessarily includes all sequence data regardless of their potential relevance to site remediation objectives, so more time and expertise may be needed to interpret the results. Proteomics analysis is a powerful tool that can detect thousands of proteins being expressed in a single sample, but, similar to metagenomics, additional statistical methods and research into environmentally relevant proteins may be required before arriving at an interpretation that can inform site management decisions.