Enabling ENIGMA collaborative research in the DOE Systems Biology Knowledgebase

Pavel S. Novichkov[1]*(psnovichkov@lbl.gov), Alexey E. Kazakov[1], Adam P. Arkin[1] and **Paul D. Adams**[1]

[1]*Lawrence Berkeley National Laboratory, Berkeley CA*

http://enigma.lbl.gov

Project Goals: Sharing the data, computational analysis, and hypotheses between members of a large-scale collaborative project like ENIGMA is critical to facilitate collaboration and accelerate the pace of scientific discovery. The goal of this project was to support the ENGIMA collaborative research by enabling necessary data models and tools to start working in the DOE Systems Biology Knowldegebase.

The Ecosystems and Networks Integrated with Genes and Molecular Assemblies (ENIGMA) is a large-scale, multi-institutional collaborative project that involves many teams with different expertise working together to address challenging biological problems. A project like this generates a wide range of experimental and computational types of data that needs to be properly stored and shared between members of the project to facilitate collaboration and accelerate the pace of scientific discovery. It is critical to enable the sharing of not only the raw data, but also the major steps in the data analysis, and ultimately, the generated hypotheses. The DOE Systems Biology Knowledgebase (KBase) is a powerful computational platform that was originally designed to specifically address these needs of the scientific community. Collaboration and sharing, data provenance and data integration, extension by new data types and new data analysis are the key features of the KBase computational infrastructure.

The goal of this project was to enable collaborative research of the ENIGMA Metal Metabolism campaign in the DOE Systems Biology Knowledgebase by implementing necessary data models and computational tools in the KBase infrastructure. The overarching goal of the ENIGMA Metal Metabolism campaign is to elucidate the fundamental mechanisms that drive metal assimilation and investigate metallobiochemistry of microbial communities in the ORNL wells. The primary data types generated by this campaign toward the goal include: (i) survey of the ORNL wells with analysis of different geochemical parameters such as concentration of various metals; (ii) growth assays of the selected isolates under different metal conditions; (iii) HPLC-ICPMS screens to identify metalloproteins involved in metal toxicity and metal reduction. To support these primary data types, we designed and implemented KBase data models representing environmental sample sets, growth data, and chromatography data. The deposition of the data into KBase Narrative is facilitated by the well-documented file formats and developed KBase uploaders for all three data types. A collection of KBase widgets and methods was developed to enable visualization and analysis of the uploaded data. “View Well Samples 2D Plot” and “View Well Sample Histogram” allows for studying the correlation of a particular geochemical parameter across all wells and comparison of different geochemical parameters across a selected number of wells. “View Growth Curves” and “View Growth Parameters Plot” generates plots visualizing growth curves for either all or selected conditions, calculates various growth parameters, such as growth rate or maximum OD, and allows to study its dependence on the growth conditions. “View Growth Parameters 2D Plot” was designed specifically to support

large-scale growth assays to investigate several media parameters, e.g. concentrations of several metals or a combination of concentration of a particular metal and knockout strains. The "View Chromatograms" method allows for basic visualization of chromatography data with zoom in/out features to study particular peaks.

This material by ENIGMA- Ecosystems and Networks Integrated with Genes and Molecular Assemblies (http://enigma.lbl.gov), a Scientific Focus Area Program at Lawrence Berkeley National Laboratory is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research under contract number DE-AC02-05CH11231