ENIGMA Computation: An integrated framework of databases and software for exploring genes, genomes, proteomes, and networks in microbes and communities.

J. Christopher Bare[1]* (Christopher.Bare@systemsbiology.org), Paramvir S. Dehal[2], Karen Foley[1], Keith Keller[2], Morgan N. Price[2], Wei-Ju Wu[1], Serdar Turkarslan[1], Jason Baumohl[2], Marcin P. Joachimiak[2], John T. Bates[2], David J. Reiss[1], Pavel S. Novichkov[2], **Adam P. Arkin[2]**, **Nitin S. Baliga[1]** and **Paul D. Adams[2]**.

http://microbesonline.org

Project Goals: Currently, one of the grand challenges in biological discovery is not the production of data but it is the data processing and analysis. Often, algorithms or computational tools exist tailored to the analysis of specific datasets yet they lack modularity to handle new or non-standard data formats, automation for non-expert usage or connectivity with downstream applications. Therefore it is critical to develop biology driven accessible computational frameworks that will promote, support and validate biological discovery. In order to tackle this grand challenge, the goals of this project is to i) integrate existing tools and ongoing efforts for algorithm, software and database development within ENIGMA, ii) develop an automated and modular framework for storing, analyzing and visualizing biological data for microbes or communities iii) promote knowledge transfer within ENIGMA and in the scientific community by providing a suite of computational resources.

[1]Institute for Systems Biology, Seattle, WA; [2]Lawrence Berkeley National Lab, Berkeley CA

While current technological advances enabled production of massive amounts of biological data, the real challenge is to connect these data sources to biological discovery. Here we describe our efforts to develop a modular and integrated framework for storage, automated analysis and visualization of biological datasets. Using Microbes Online as a source of high-quality curated data, we infer regulatory networks for several organisms. Feeding gene expression data and experimental metadata through the cMonkey and Inferelator, network inference algorithms produce, for a given organism, a set of co-expressed and putatively co-regulated modules. These networks are archived in the Network Portal, which provides a web interface for searching, visualizing and annotating the networks. The Network Portal will be tightly integrated with Microbes Online, which already provides a rich set of features for comparative microbial genomics as well as RegPrecise for curation of regulatory inferences. Additional experimental validation of network inferences can be performed using the MicrobesOnline Fitness Experiment data. Automation of the network inference pipeline will enable us to capture phylogenetic diversity to drive comparative studies and availability of high quality fitness data will help us to asses and curate network inferences. In addition, the Gaggle integration framework provides interoperability with several popular bioinformatics tools including Cytoscape, MeV, and R. Firegoose provides Gaggle connectivity to web resources such as KEGG, DAVID and EMBL STRING. This suite of integrated analysis and visualization tools provides a powerful and easily extended environment for the study of the regulatory systems of microbial biology.

This work conducted by ENIGMA was supported by the Office of Science, Office of Biological and Environmental Research, of the U. S. Department of Energy under Contract No. DE-AC02-05CH11231.