ENIGMA: MAGI: A Method For Metabolite, Annotation, And Gene Integration

Onur Erbilgin[1], **Benjamin P. Bowen[1,3,*] (bpbowen@lbl.gov),** Oliver Rübel[2], Katherine B. Louie[3], Matthew Trinh[1], Markus de Raad[1], Tony Wildish[3,4], Daniel W. Udwary[3,4], Cindi A. Hoover[3], Samuel Deutsch[1,3], **Trent R. Northen[1,3,*],** A.P. Arkin[1,5] and P.D. Adams[1,5]

[1]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory; [2]Computational Research Division, Lawrence Berkeley National Laboratory; [3]Joint Genome Institute, Lawrence Berkeley National Laboratory; [4]National Energy Research Scientific Computing Center, Lawrence Berkeley National Laboratory; [5]University of California, Berkeley, CA.

http://enigma.lbl.gov

Project Goals: ENIGMA -Ecosystems and Networks Integrated with Genes and Molecular Assemblies use a systems biology approach to understand the interaction between microbial communities and the ecosystems that they inhabit. To link genetic, ecological, and environmental factors to the structure and function of microbial communities, ENIGMA integrates and develops laboratory, field, and computational methods.

Abstract: Metabolomics is a widely used technology for obtaining direct measures of metabolic activities from diverse biological systems. However, it is limited by ambiguous metabolite identifications. Furthermore, interpretation is limited by incomplete and inaccurate genome-based predictions of enzyme activities (*i.e.* gene annotations). Metabolite, Annotation, and Gene Integration (MAGI) attempts to addresses these challenges by generating metabolite-gene associations via biochemical reactions based on a score between probable metabolite identifications and probable gene annotations. This is calculated by a method that emphasizes consensus between metabolites and genes via biochemical reactions. To demonstrate the potential of this method, we applied MAGI to integrate sequence data and metabolomics data collected from *Streptomyces coelicolor* A3(2), an extensively characterized bacterium that produces diverse secondary metabolites. Our findings suggest that coupling metabolomics and genomics data by scoring consensus between the two increases the quality of both metabolite identifications and gene annotations in this organism. MAGI also made biochemical predictions for poorly annotated genes that were consistent with the literature. This limited analysis suggests the potential using metabolomics data to improve annotations in sequenced organisms and also providing testable hypotheses for specific biochemical functions. MAGI is freely available for academic use both as an online tool at https://magi.nersc.gov and with source code available at https://github.com/biorack/magi.

This material by ENIGMA- Ecosystems and Networks Integrated with Genes and Molecular Assemblies 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