

160. Analysis capabilities for microbial communities in KBase

Folker Meyer*¹(folker@anl.gov), Dylan Chivian², Jason Baumohl², Aaron Best³, Jared Bischof¹, Ben Bowen², Thomas Brettin¹, Tom Brown¹, Shane Canon¹, Stephen Chan², John-Marc Chandonia², Ric Colasanti¹, Neal Conrad¹, Brian Davison⁴, Matt DeJongh³, Paramvir Dehal², Narayan Desai¹, Scott Devoid¹, Terry Disz¹, Meghan Drake⁴, Janaka Edirisinghe¹, Gang Fang⁷, José Pedro Lopes Faria¹, Mark Gerstein⁷, Elizabeth M. Glass¹, Annette Greiner², Dan Gunter², James Gurtowski⁶, Nomi Harris², Travis Harrison¹, Fei He⁵, Matt Henderson², Chris Henry¹, Adina Howe¹, Marcin Joachimiak², Kevin Keegan¹, Keith Keller², Guruprasad Kora⁴, Sunita Kumari⁶, Miriam Land⁴, Steve Moulton⁴, Pavel Novichkov², Taeyun Oh⁸, Gary Olsen⁹, Bob Olson¹, Dan Olson¹, Ross Overbeek¹, Tobias Paczian¹, Bruce Parrello¹, Shiran Pasternak⁶, Sarah Poon², Gavin Price², Srividya Ramakrishnan⁶, Priya Ranjan⁴, Bill Riehl², Pamela Ronald⁸, Michael Schatz⁶, Lynn Schriml¹⁰, Sam Seaver¹, Michael W. Sneddon², Roman Sutormin², Mustafa Syed⁴, James Thomason⁶, Nathan Tintle³, Will Trimble¹, Daifeng Wang⁷, Doreen Ware^{5,6}, David Weston⁴, Andreas Wilke¹, Fangfang Xia¹, Shinjae Yoo⁵, Dantong Yu⁵, **Robert Cottingham⁴**, **Sergei Maslov⁵**, **Rick Stevens¹**, **Adam P. Arkin²**

¹Argonne National Laboratory, Argonne, IL, ²Lawrence Berkeley National Laboratory, Berkeley, CA, ³Hope College, Holland, MI, ⁴Oak Ridge National Laboratory, Oak Ridge, TN, ⁵Brookhaven National Laboratory, Upton, NY, ⁶Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, ⁷Yale University, New Haven, CT, ⁸University of California, Davis, CA, ⁹University of Illinois at Champaign-Urbana, Champaign, IL, ¹⁰University of Maryland, College Park, MD

<http://kbase.us>

Project Goals: The KBase project aims to provide the capabilities needed to address the grand challenge of systems biology: to predict and ultimately design biological function. KBase enables users to collaboratively integrate the array of heterogeneous datasets, analysis tools and workflows needed to achieve a predictive understanding of biological systems. It incorporates functional genomic and metagenomic data for thousands of organisms, and diverse tools for (meta)genomic assembly, annotation, network inference and modeling, allowing researchers to combine diverse lines of evidence to create increasingly accurate models of the physiology and community dynamics of microbes and plants. KBase will soon allow models to be compared to observations and dynamically revised. A new prototype Narrative interface lets users create a reproducible record of the data, computational steps and thought process leading from hypothesis to result in the form of interactive publications.

KBase provides the infrastructure and tooling for in-depth metagenome analysis, facilitating the annotation of microbial communities and the quest for identification of key players in a microbial community or the identification of trends. By automatically transforming microbial communities into abundance profiles KBase is enabling users to drill down so that trends, specific taxa or functions can be identified. Combining metagenomic and environmental data makes it possible to correlate information about organism or function abundance to metadata that describe a variety of biologically intriguing characteristics of the samples, such as the biome the samples were collected from, the pH of the samples, etc.

Metabolic modeling can help elucidate the roles played by individual taxa in microbial communities by providing a detailed characterization of their functional repertoire. Once obtained this knowledge can be used for a number of purposes, such as the prediction of cultivation conditions for functionally important taxa. While the current state of modeling and our ability to annotate microorganisms has advanced greatly in recent years, the emerging models should be viewed as a first approximation rather than the final answer to these questions.

With KBase the functionality is in place to perform comparisons of multiple strategies for deriving metabolic models from microbial community data. Those strategies include the use of PCR primer amplified ribosomal genes as a reporter for the organisms informing PICRUSt predictions, the use of shotgun metagenomics to obtain functional information, the use of EMIRGE to extract complete ribosomal sequences from shotgun metagenomics sequences and the use of taxonomic information obtained from metagenomic sequences to inform PiCRUST predictions.

KBase is funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research.