

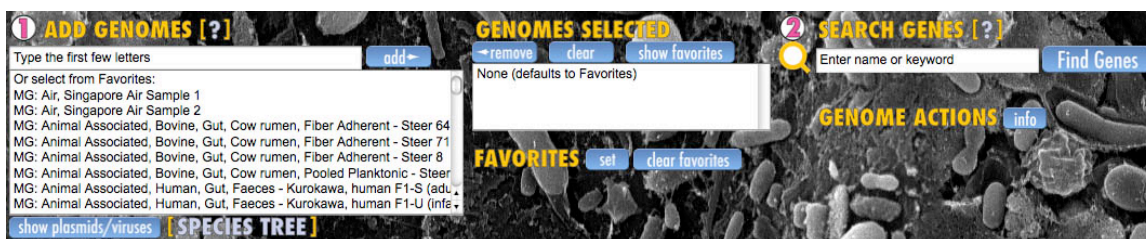
*For Immediate Release. Oct. 8<sup>th</sup>, 2010*

MicrobesOnline.org Enhanced for Metagenomics and Metabolism

# metaMicrobesOnline & GLAMM

The Arkin lab ([genomics.lbl.gov](http://genomics.lbl.gov)) computational biology and bioinformatics team, as part of the Joint BioEnergy Institute (JBEI, [www.jbei.org](http://www.jbei.org)) and the Virtual Institute for Microbial Stress and Survival (VIMSS, [vimss.lbl.gov](http://vimss.lbl.gov)), has added significant new functionality to the MicrobesOnline.org web resource for comparative and functional analysis of microbial genomes. This work, led by **Dylan Chivian** with **John Bates**, **Keith Keller**, **Morgan Price**, **Paramvir Dehal**, under the guidance of **Adam Arkin**, offers the scientific community new capabilities for metagenomic and metabolic analyses.

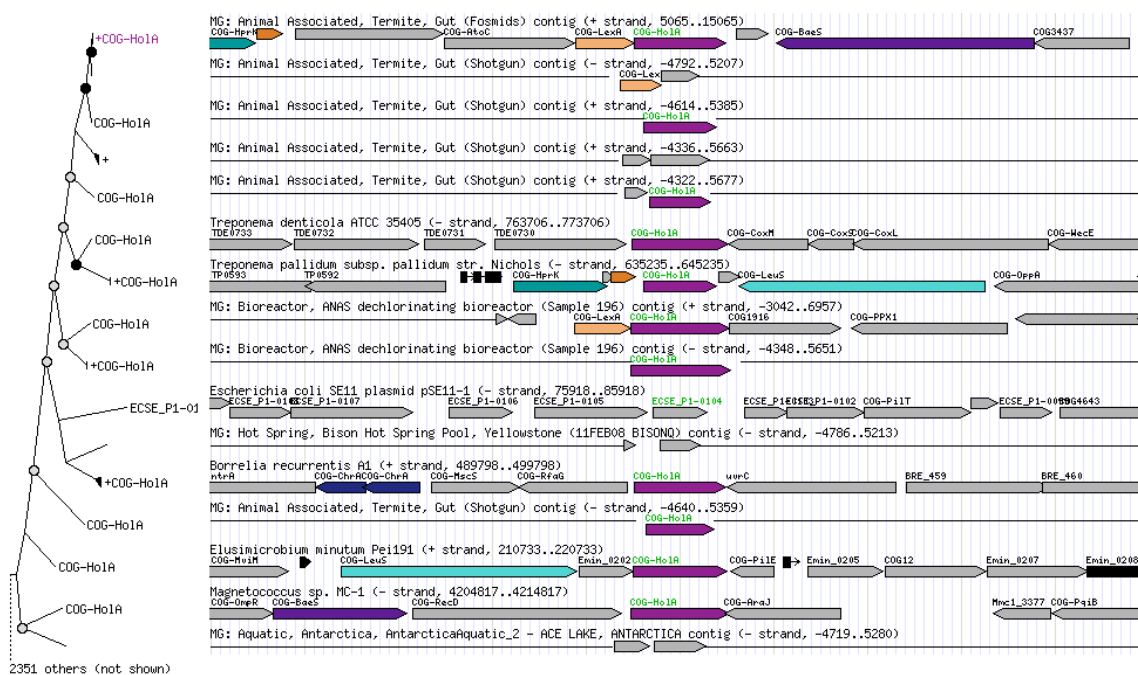
## metaMicrobesOnline



**Figure 1.** metaMicrobesOnline permits the user to select metagenomes and isolates they want to study and then perform easy searches for keywords or gene families (e.g. “amylase” or “PFAM00128”). Scanning electron microscopy image of microbial compost community courtesy Bernhard Knierim and Manfred Auer.

The new [meta.MicrobesOnline.org](http://meta.MicrobesOnline.org) web resource, which currently contains over 1600 genomes from bacterial, archaeal, and microeukaryotic isolates, offers combined phylogenetic gene tree analysis of millions of genes from over 150 ecological and organismal metagenomes. These trees are built using our FastTree [1] program, which offers rapid highly accurate tree building, even for very large trees. Such combined analysis is superior to BLAST-based homology approaches in that trees offers the ability to **place genes from environmental samples into an evolutionary context** and permits more precise functional grouping within a gene family, yielding information about the key genes for a given environment. Additionally,

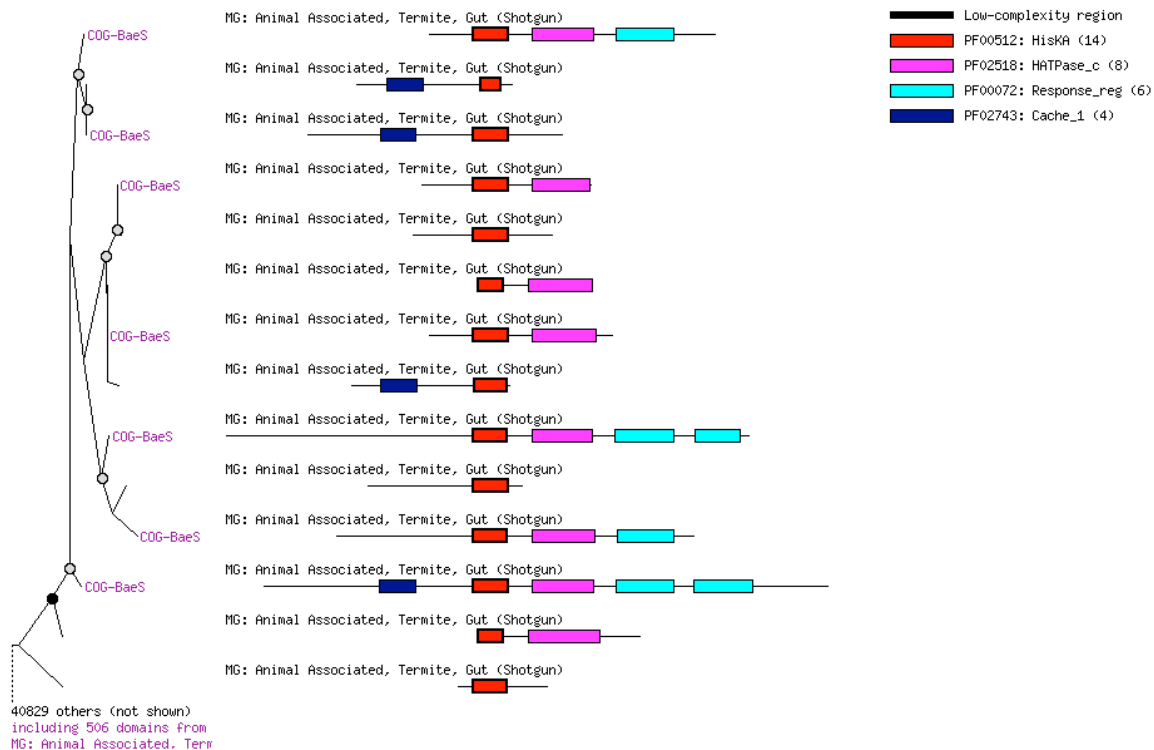
comparison with isolate genomes gives researchers clues for which additional genes to look for to complete the components of systems, or may possess phylogenetic markers to aid in assigning the species for environmental sequence fragments, permitting the determination of which community members are responsible for which roles. Lastly, when there is a reference genome that is sufficiently closely related to a member of the community, comparison of the environmental sequences to the reference genome is very valuable for grouping the fragmentary sequences and overcoming the challenge of putting “humpty-dumpty” back together. Grouping fragmentary sequences increases the ability of the researcher to perform whole-genome analyses from environmental sequence, and thus furthers the detailed analysis of the microbial community.



**Figure 2. TreeBrowser** shows phylogenetic gene trees in the context of the neighboring region of the genome sequence. Members of the same gene family are given the same color. Fragmentary sequences are indicated with a line over the truncated region.

With the goal of simplifying access to the desired genes, metaMicrobesOnline allows the researcher to select metagenomes and isolate genomes and then perform easy keyword searches on information associated with the genes in the database in a typical intuitive fashion (Figure 1). Searches include the locus id (if known), descriptive terms in annotations (such as “amylase”), or canonical gene or domain families (such as “PFAM00128”). Metagenomes and isolate genomes have had their genes assigned to all canonical gene and domain families from COG, Pfam, and TIGRFAM. These gene and domain families are then used to build phylogenetic trees with both the environmental and isolate genes. Trees may include hundreds of thousands of genes. These trees provide an evolutionarily meaningful grouping of the

genomic information, captured in the TreeBrowser (Figure 2). Additionally, the researcher may investigate the domain combinations present in the homologs of a gene of interest using our DomainBrowser, again using the evolutionary context provided by metaMicrobesOnline's trees (Figure 3).



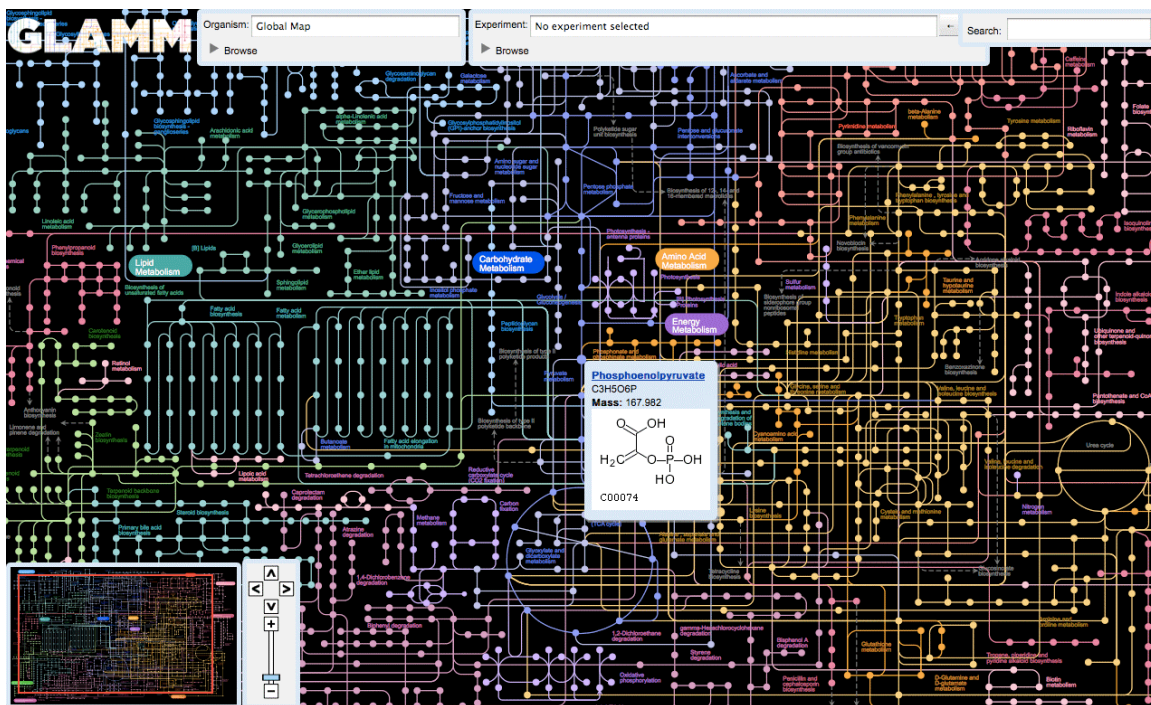
**Figure 3. DomainBrowser** tree with domain combinations found in histidine kinase (HK) signaling genes. For metagenomic sequences genes may be missing domains as a result of sequencing truncation.

There are many new environmental sequencing projects underway, and metaMicrobesOnline will endeavor to include those that are amenable to our computationally expensive tree-building approach. Additionally, we offer private access to any user who would like to annotate their metagenomic sequence in metaMicrobesOnline. Please contact [help@microbesonline.org](mailto:help@microbesonline.org) for more information.

## GLAMM

The Google-Like Application for Metabolic Maps (GLAMM) has been developed to provide an interactive browsing experience with metabolic networks based on that used with web mapping technology. GLAMM is integrated with the MicrobesOnline.org web resource, allowing a researcher to access the many powerful comparative genomic and functional analysis tools present in MicrobesOnline [2].

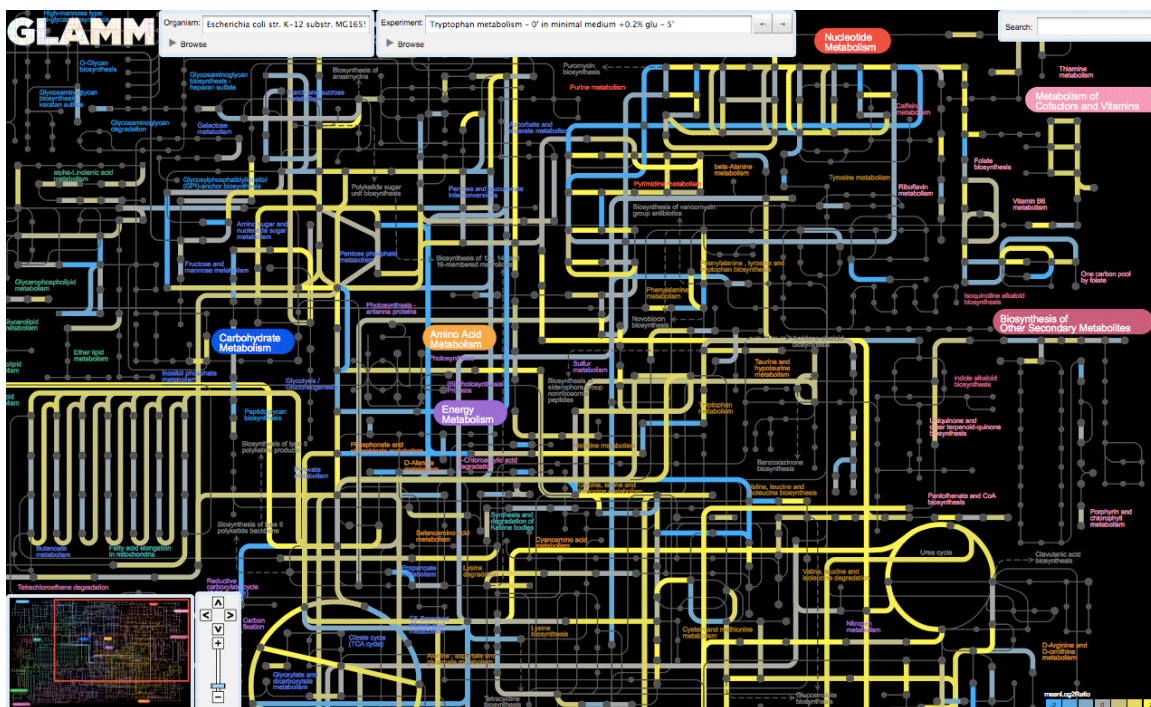
GLAMM allows a researcher to browse the metabolic network as though it were a web map, including the ability to search for and obtain information about metabolites and reactions (Figure 4). These “dialog boxes” include links back to MicrobesOnline, where and genes and metabolites that take part in given reactions are available for further study. All of the organisms present in MicrobesOnline have automatic metabolic reconstructions computed using assignments to enzyme families and homology to genes in KEGG [3], which can be selected for browsing.



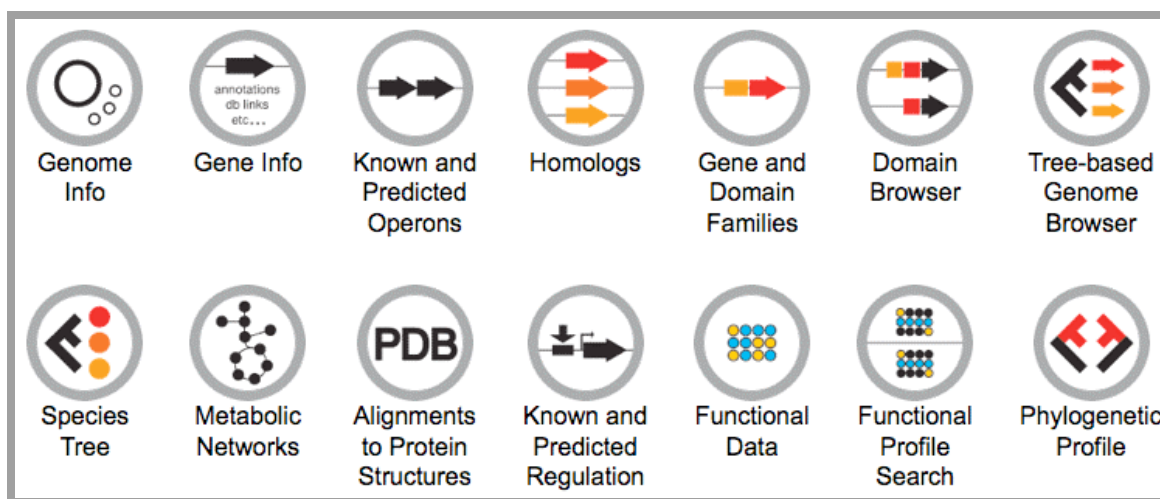
**Figure 4. GLAMM global map** allows browsing of reactions from automated metabolic reconstructions, search, and information about reactions and compounds, integrated with MicrobesOnline’s comparative tools.

MicrobesOnline contains functional systems biology data for many organisms. This data can be displayed on the metabolic map (Figure 5) by selecting the desired organism and then selecting experiments to add to the set. These data can then be studied in series, revealing which reactions are up or down-regulated under particular conditions, and the differences between the experiments. GLAMM then makes it simple to find the genes that carry out the reactions that have been indicated by the experiments and investigate with the comparative and functional tools in MicrobesOnline. These tools (Figure 6) include known and predicted operons, homolog assignments, gene and domain family assignments, the DomainBrowser domain tree and domain combination browser, the TreeBrowser gene tree and genome context browser, a species tree based on conserved protein coding genes, alignments to protein structures, information on gene regulation, analysis tools for functional data, expression pattern correlation search tools,





**Figure 5. GLAMM functional systems biology data overlay** on the map reveals which reactions are up (yellow) or down-regulated (blue).



**Figure 6. MicrobesOnline features** include summary information about genomes and genes, links to other databases, and tools for comparative genomic and functional systems biology analyses.

regulatory binding site motif finding tools, phylogenetic profile search tools, gene carts that allow users to build multiple sequence alignments and gene trees using sets of genes of their choosing, and so on. We continue to expand the capabilities of MicrobesOnline as a powerful platform for comparative and functional analysis of environmental and isolate microbes.



U.S. DEPARTMENT OF  
**ENERGY**

Office of  
Science

## Funding

This work was part of the DOE Joint BioEnergy Institute ([www.jbei.org](http://www.jbei.org)), the ENIGMA Scientific Focus Area Program, the Genomics:GTL Foundational Science supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory (LBNL) and the U.S. Department of Energy. Additionally, the GLAMM project received support from the "Knowledgebase R&D" project, sponsored by the Office of Biological and Environmental Research in the DOE Office of Science with American Recovery and Reinvestment Act funding to Oak Ridge National Laboratory (ORNL) and performed at LBNL. ORNL is managed by UT-Battelle, LLC, for the U.S. Department of Energy under contract DE-AC05-00OR22725.

## References

- [1] FastTree2 – approximately maximum-likelihood trees for large alignments. Price MN, Dehal PS, Arkin AP. *PLoS One*. 2010; 5(3):e9490
- [2] MicrobesOnline: an integrated portal for comparative and functional genomics. Dehal PS, *et al.* *Nucleic Acids Res.* 2010; 38(Database issue):D396-400
- [3] KEGG for linking genomes to life and the environment. Kanehisa M, *et al.* *Nucleic Acids Res.* 2008; 36(Database issue):D480-4

## For more information, please contact

Dylan Chivian  
[DCChivian@lbl.gov](mailto:DCChivian@lbl.gov)

Adam Arkin  
[APArkin@lbl.gov](mailto:APArkin@lbl.gov)

