Title

MicrobesOnline: An Integrated Portal For Comparative Functional Genomics (R-018)

Author

Huang, Y. Wayne

Publication Date

2008-06-02

07-GM-A-4111-ASM

MicrobesOnline: An Integrated Portal For Comparative Functional Genomics (R-018)

Y. Wayne Huang[1,2], Katherine H. Huang[1,2], Eric J. Alm[1,3], Dylan Chivian[1,2], Paramvir S. Dehal[1,2],
Marcin P. Joachimiak[1,2], Keith Keller[1,4], Morgan N. Price[1,2], Adam P. Arkin[1,2,4,5,6]

[1] Virtual Institute for Microbial Stress and Survival, http://vimss.lbl.gov; [2] Lawrence Berkeley National Laboratory, Berkeley, CA;
[3] Department of Biological Engineering, MIT, Cambridge, MA; [4] University of California, Berkeley, CA;
[5] Howard Hughes Medical Institute, Chevy Chase, MD; [6] Department of Bioengineering, University of California, Berkeley, CA

gtlweb@vimss.lbl.gov

http://vimss.lbl.gov/

Introduction

The MicrobesOnline database (http://microbesonline.org) currently contains 430 microbial genomes as of our March 2007 release and will soon be expanded to over 600 by Fall 2007. Together with a suite of visualization, analysis, and annotation tools, MicrobesOnline provides a comprehensive framework for studying important comparative and evolutionary genomic questions. MicrobesOnline offers many useful tools including a multi-species genome browser, operon and regulon prediction methods and results, gene ontology browser, capabilities for community annotation of genomes, and a user workbench, which allows users to store arbitrary sets of sequences for further analysis including the ability to generate multiple sequence alignments and custom gene trees.

In our most recent software release, we've added a combined gene and species phylogeny browser (see *Tree Browser*), sequence motif detection and motif search tools in the user workbench, and tools for viewing and mining gene expression microarray, proteomic, and phenotype microarray data (see *Expression Profile Searches*).

MicrobesOnline has facilitated comparative studies on prokaryotic genomes since 2003, and currently contains 431 genomes.

- To describe the potential functions of genes, MicrobesOnline includes protein family analyses (from InterPro and COG), metabolic maps (from KEGG), links to research papers (from UniProt and PubMed), and operon predictions for every genome.
- To examine each gene's evolutionary history, MicrobesOnline includes precomputed phylogenetic trees for all the gene families. It displays gene trees with genomic context or it compares the gene tree to the species tree.
  - You can also compute your own motifs, sequence alignments, and phylogenetic trees by saving genes to "My Gene Carts" (registration or login required).
- MicrobesOnline includes gene expression data from microarrays: you can view up-regulated or down-regulated genes or operons, and you can view changes in expression patterns in a metabolic map.
- You can annotate a gene by changing the gene's name, description, enzyme classification (EC), or Gene Ontology (GO) assignment, or by adding a comment.

Sponsored by the U.S. Department Of Energy Genomics: GTL program.
To cite this website, please use this paper.

News

- Over 400 genomes now available (March 2007)
- Public microarray data now available on MicrobesOnline (December 2005)

Recent Publications

Gene Cart and Workbench Tools

Gene carts allow users to create arbitrary sets of genes for further study. Each cart can be saved and recalled at a later date, allowing the user to download the associated sequences or to perform further analysis. Users may create multiple sequence alignments using *MUSCLE*, view the alignment graphically using *JalView*, and subsequently generate a phylogenetic tree using a number of different methods. MicrobesOnline also offers a unique multi-species browser with color-coding based on orthology. Results generated using the workbench tools are associated with the user's profile and can be viewed at a later date, with the option to save the results permanently.

Recently, we've added support for motif discovery using *AlignACE, MEME,* and *Weeder*, and motif searches using *ScanACE, MAST,* and *Patser*, respectively.

VIMSS Id	Name	Organism
208612	PerR	Desulfovibrio vulgaris Hildenborough
208611	rbr	Desulfovibrio vulgaris Hildenborough
208610	rdl	Desulfovibrio vulgaris Hildenborough
207805	rbr2	Desulfovibrio vulgaris Hildenborough
207792	ahpC	Desulfovibrio vulgaris Hildenborough
206199	DVU0772	Desulfovibrio vulgaris Hildenborough

Motif Search Tool: MEME
Sequence Region: ● Search around start of gene: offset -150 to -50 bp
○ Search around end of gene: offset to bp
○ Search entire gene
Truncate?: ☐ Truncate regions overlapping other genes
Motif Distribution:
Max Motifs to Find:
E-value Threshold:
#Sites: Min: Max:
Motif Width: Min: Max:
Gap Costs: Open: Extend:
Options: ☐ Include - strand in search
☑ Force palindromes
☐ Do not adjust motif width using multiple alignment
☐ Do not count end gaps in multiple alignment
Name of Motif Search:

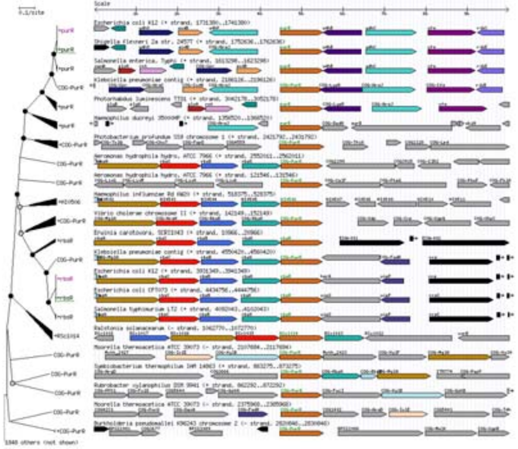
Scan for Hits with Selected Motif
Genome to Scan: Desulfovibrio vulgaris Hildenborough
Min Score (in bits):
Max Number of Hits:
Options: ☑ Treat motif as palindromic
Max Distance from Gene (bp):
Name of Motif Scan:

Tree Browser

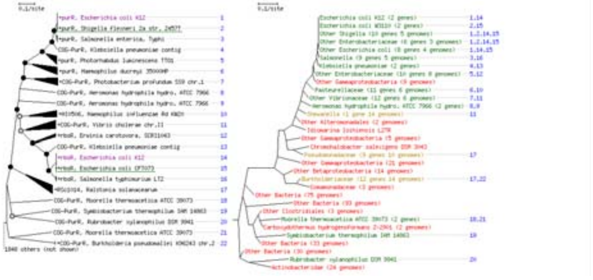
Sequence similarity searches are the first steps in computational gene characterization. This process, however, currently generates thousands of significant matches per gene and identifying the useful matches within this large dataset, of which many of the matches are to closely related strains or gene families, can be quite tedious and time consuming. Without knowing the species relationships, the similar matches alone will not reveal the gene's distribution or evolutionary history.

Our solution is the MicrobesOnline *Tree Browser*. A phylogenetic tree is constructed for each gene family on the basis of domains, COGs, or ad-hoc BLAST families. The *Tree Browser* can then display one's gene in the context of its phylogenetic tree, highlighting the characterized homologs, and allowing one to collapse branches of close relatives to allow more distant relationships to be shown and to reconcile the gene tree with the species tree to elucidate one's gene's evolutionary history.

E. coli purR closest homologs



E. coli purR gene tree vs. Species tree



Future Work

- Import more microarray expression data (GEO, ArrayExpress); as of 2/9/07, we had 354 experiments (1,213 hybridizations) and growing
- Microarray tools for comparing multiple experiments across multiple genomes
- Extend motif search to incorporate co-expression data to support predictions
- Protein-protein interaction data
- Metabolomic and proteomic data

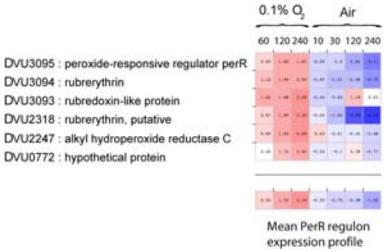
Expression Profile Searches

MicrobesOnline integrates functional genomic data and provides novel web-based viewing and mining tools for gene expression microarray, proteomic, and phenotype microarray data. Currently these data are project-specific and limited to *Desulfovibrio vulgaris* and *Shewanella oneidensis* exposed to stress conditions found at DOE field sites, but we are planning to incorporate data from public microarray databases such as GEO and ArrayExpress, as well as data for other organisms. In addition to being able to view associated microarray data for a specific gene of interest, one can now search any subset of experiments in the microarray data compendium for similar gene matches to a mean expression profile derived from an a priori set of genes such as a known or predicted regulon. These new compendium-wide searches allow one to observe patterns in gene expression changes across multiple conditions and to be able to search for similarities to these patterns.

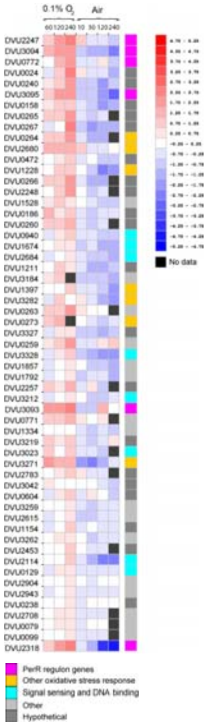
Rodionov et al (Genome Biology 2003) identified a putative peroxide response (PerR) regulon in the *Desulfovibrio vulgaris Hildenborough* genome using comparative genomics and sequence motif detection. Subsequently, a mild environmental stress microarray experiment, micro-aerobic stress (0.1 % O_2), identified members of this regulon amongst the only 12 significantly changing genes (total of genes with $\log_2$ ratio ≥ 2 or ≤ -2 and z-score ≥ 2 and ≤ -2, respectively).

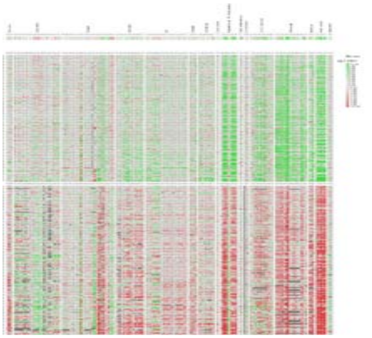
The previously identified PerR sequence motif

PerR regulon gene expression and mean expression profile in two types of oxygen stress experiments



Profile search results using an empirical score cutoff





Compendium gene-gene expression profile searches

ACKNOWLEDGEMENT

ESPP is part of the Virtual Institute for Microbial Stress and Survival supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomics:GTL Program through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy.