

Lawrence Berkeley National Laboratory

Recent Work

Title

MicrobesOnline: A Community Resource for Data-Integrated Comparative Functional Genomics

Permalink

<https://escholarship.org/uc/item/1w027679>

Author

Arkin, A.P.

Publication Date

2006-05-22



INTRODUCTION

Since 2003, the MicrobesOnline.org web site of the Environmental Stress Pathway Project (ESPP) has provided novel web-based tools for comparative genomics of prokaryotes.

- To describe the potential functions of genes, MicrobesOnline includes protein family analyses (from InterPro and COG), metabolic maps (from KEGG), and links to characterized genes (from UniProt).
- To examine each gene's evolutionary history, MicrobesOnline includes precomputed phylogenetic trees for all the gene families, and displays them with the genes' genomic context, or it compares the gene tree to the species tree.
- Users can also compute their own sequence alignments and phylogenetic trees.
- To examine the potential regulation of genes of interest, MicrobesOnline also includes tools for finding sequence motifs and for finding regulatory sites.
- MicrobesOnline includes microarray data: it shows up-regulated or down-regulated genes or operons, shows changes in expression patterns in a metabolic map, or gives an overview of the expression pattern for specific genes of interest.
- Finally, users can use our gene annotation interface to add or revise gene names, descriptions, EC or GO assignments, or comments.

Currently MicrobesOnline hosts over 350 public genomes and 20 private genomes, and ~300 public and private microarrays. Our web site is unique in that, in addition to providing our own automated analyses, it allows users to perform their own analyses of genes of their choice.

All of these tools are freely available at our web site:

<http://www.microbesonline.org>

Our examples concern perR and related genes in *Desulfovibrio vulgaris* Hildenborough. Searching for perR (above) returns search results (below) where there are links to view the various pages relating to each gene.

Every gene page has links for viewing basic gene annotations, operon predictions, domain alignments from InterPro, BLASTp homologs, and the public annotation interface.

Future Work

- More links to external databases/web sites, such as RegTransDB, EcoCyc, IMG, STRING, PhIC
- More microarray data, and a new interface including heat maps
- Protein-protein interaction data
- Metabolic and proteomic data
- Operon-wise significance analysis of microarray data

Gene Information and Annotation Pages

The Gene Info page (left) gives an overview of perR. A user-contributed annotation notes that perR may be a part of the rbr operon, so perhaps rbr is also a gene of interest. The Add Annotation page (below) allows users to add additional annotations.

The operon view of rbr (left) shows that perR is not predicted by our methods to be part of the rbr operon. User contributed annotations allow for manually-entered information to supplement automated annotations.

Microarray Data Display and Analysis

In ESPP experiments, perR was significantly up-regulated, as well as rbr2 and rbr (above). The operon view of this microarray experiment (upper right) shows that rbr was up-regulated, as was the other gene in the operon.

Phylogenetic Tree Browser

Pre-computed trees are available for each PFAM, for each COG, and for additional ad-hoc families from BLASTp. Trees are built with MUSCLE and quickjoin. There are 41,000 trees covering 88% of all protein coding genes.

Phylogenetic trees allow the user to view the gene in context with genomes which have closely-related genes. These example trees are evidence of the history of that gene: three proteobacteria contain perR, but no others do. Perhaps these organisms obtained perR through horizontal transfer. The gene context (above) shows perR in relation to the related COG in other genomes; the species context (right) shows that the gene is not present in most proteobacteria.

Gene Cart Workbench Tools

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 Program:GTL through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy.