Title

Integrated Microbial Genomes: From Genomes to Metagenomes

Author

Kyrpides, N.C.

Publication Date

2006-05-01

Integrated Microbial Genomes : From genomes to metagenomes

Markowitz, V.M.[1,2], Palaniappan, K. [1,2], Szeto, E. [1,2], Korzeniewski, F. [1,2], Dubchak, I. [2], Hugenholtz, P. [2], Mavromatis, K. [2], Anderson, I. [2], Lykidis, A. [2], Ivanova, N. [2], Kyrpides, N.C. [2]

1. Biological Data Management Center, Lawrence Berkeley National Lab
2. DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA

Abstract

Integrated Microbial Genomes (IMG) is a microbial genome analysis system developed at the DOE Joint Genome Institute (JGI) to promote individual and comparative analysis of genomes sequenced at the JGI and elsewhere. Version 1.3, released on December 1, 2005, contains 674 genomes.

Following release will also include data from metagenomes (IMG/M) with the same functionality as IMG v1.3 and additional features that assist metagenomic analysis.

New features include the implementation of analysis carts that allow the comparative study of functions against selected genomes and abundance profiles that show the relative presence of function and homologs in genomes and environmental samples.

The study of metagenomic data is further facilitated by providing information for individual bins and the phylogenetic mapping of metagenomic fragments.

IMG continues to be updated on a quarterly basis with new public and JGI genomes. The next update is scheduled for March 1st, 2006.

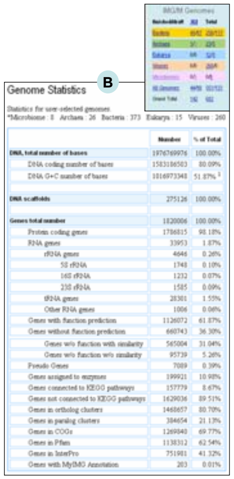
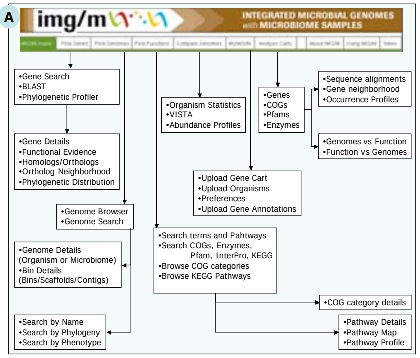


Figure 1. Summary of genomes (A) and genome statistics (B) in IMG/M. Statistics can be displayed for any individual organism or any group of organisms. A table for comparison of individual organism statistics can be generated with any subset of categories.



- •Gene Search
- •BLAST
- •Phylogenetic Profiler

- •Gene Details
- •Functional Evidence
- •Homologs/Orthologs
- •Ortholog Neighborhood
- •Phylogenetic Distribution

- •Genome Browser
- •Genome Search

- •Genome Details (Organism or Microbiome)
- •Bin Details (Bins/Scaffolds/Contigs)

- •Search by Name
- •Search by Phylogeny
- •Search by Phenotype

- •Organism Statistics
- •VISTA
- •Abundance Profiles

- •Genes
- •COGs
- •Pfams
- •Enzymes

- •Sequence alignments
- •Gene neighborhood
- •Occurrence Profiles

- •Genomes vs Function
- •Function vs Genomes

- •Upload Gene Cart
- •Upload Organisms
- •Preferences
- •Upload Gene Annotations

- •Search terms and Pahtways
- •Search COGs, Enzymes, Pfam, InterPro, KEGG
- •Browse COG categories
- •Browse KEGG Pathways

- •COG category details

- •Pathway Details
- •Pathway Map
- •Pathway Profile

Figure 2. User interface map for IMG 1.3 and IMG/M. Users can limit their analysis to a specific group of organisms through the *Genome Browser* or *Genome Search* functions and get information for specific genomes and bins (Figures 5 and 6). Under *Compare Genomes* the user can compare *Genome Statistics*, or genome sequence, for selected groups of organisms, using *VISTA, or Abundance Profiles* (Figure 3). Sets of genes or organisms can be exported for storage and then uploaded through *MyIMG. Genes*, or functions of interest, can be added to the *Analysis Carts* and then analyzed using the *Compare Genes* tools, or *Function Profiles.* The distributions of COGs, Pfams or enzymes in KEGG pathways can be compared in up to 20 organisms (Figure 4).

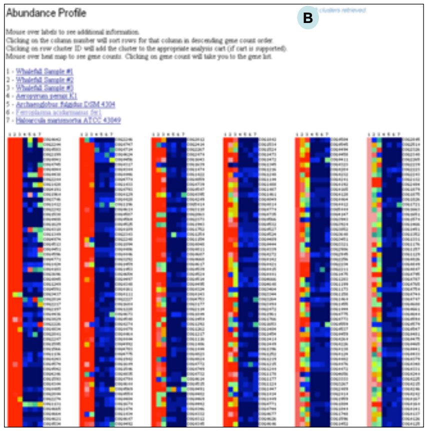


Figure 3. Genomes may be viewed against functional clusters (COG, Pfam, Enzymes, and native clusters)(A). A high density heat map facilitates visualization of relative abundance (B). Columns may be sorted by descending gene count (most to least abundant) for a given genome, to be compared against other genomes. Normalization selection affects heat map colors.

IMG (v1.3 December 2005) http://img.jgi.doe.gov
IMG/M (March 2006) http://img.jgi.doe.gov/m/

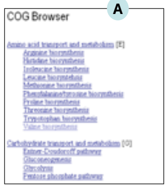
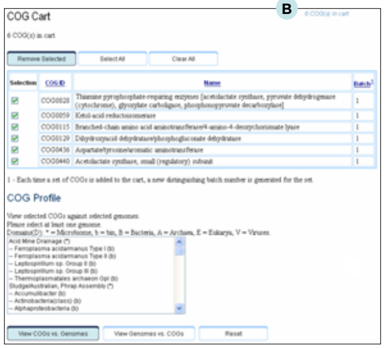
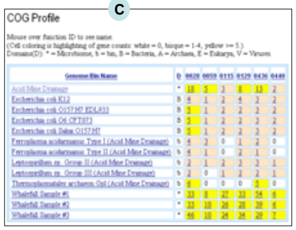
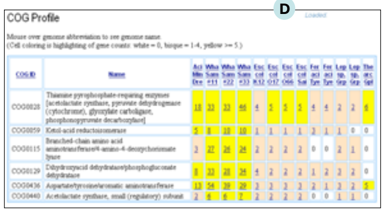


Figure 4. COG Analysis Cart. This page lists all the COG categories (A) and individual COGs within a category (B). User can add any of these COGs to the *Analysis Cart* and compare the profiles of genes belonging to the COG in a user-defined set of organisms (C and D). Clicking on the gene number takes you to a list of genes (for 2 or more genes) or to the *Gene Details* page (for a single gene). The same type of profile can be generated for Pfam and enzymes in a KEGG pathway.







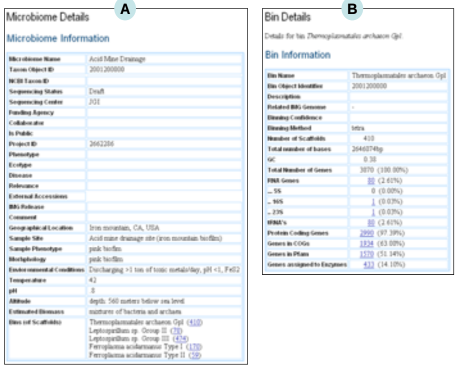


Figure 5. Microbiome details. Information about the sequencing project and the bins is shown (A). User can select a bin and see detailed information about it (B).

Figure 6. Phylogenetic map of metagenomic fragments (in IMG/M). The histogram is a count of best hits of genes within the phylum or class at 30%, 60% and 90% BLAST identities. *Unassigned* are the remainder of genes less than the percent identity cutoff, or that are not best hits at the cutoff, or have not hits

This work was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48, Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231 and Los Alamos National Laboratory under contract No. W-7405-ENG-36.