Integrated Microbial Genomes (IMG): New Data and Functionality in Version 1.2

Victor M. Markowitz, Frank Korzeniewski, Krishna Palaniappan, Ernest Szeto, Greg Werner, Anuradha Padki, Xueling Zhao, Philip Hugenholtz, Konstantinos Mavromatis, Iain Anderson*, Athanasios Lykidis, Natalia Ivanova and Nikos C. Kyrpides

DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

Abstract

IMG (http://img.jgi.doe.gov) 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 1.2, released on September 1, 2005, contains 618 genomes including the addition of 260 phage genomes. New functionality in version 1.2 includes the ability to compare profiles of all the members of a COG category or KEGG pathway map in up to 20 user-selected genomes. The occurrence of enzymes and COG members can now be viewed phylogenetically. A link to VISTA allows detailed genome alignment of closely related groups of organisms. Organism search capabilities have been enhanced, and users can now upload their own annotations through MyIMG.

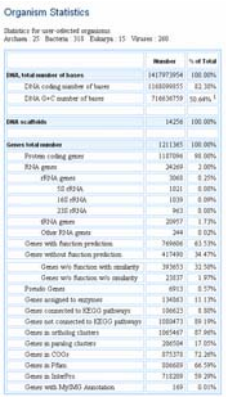

Figure 1. Summary of organism statistics for all genomes in IMG 1.2. Statistics can be displayed for any individual organism or any group of organisms. A table for comparison of individual organism statistics can be generated with any subset of categories.

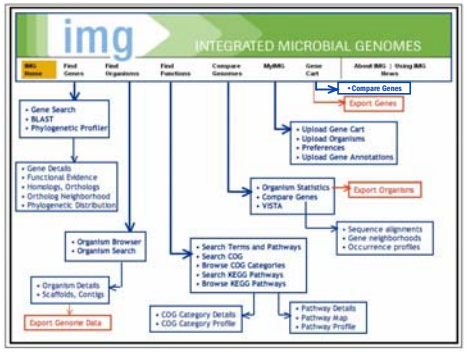

Figure 2. User interface map for IMG 1.2. Users can limit their analysis to a specific group of organisms through the Organism Browser or Organism Search functions. Genes of interest can be added to the Gene Cart and then analyzed using the Compare Genes tools. The distributions of COGs or enzymes in KEGG pathways can be compared in up to 20 organisms (see Figure 4), or the distribution of a single COG or enzyme can be viewed phylogenetically (see Figure 5). Sets of genes or organisms can be exported for storage and then uploaded through MyIMG.

Figure 3. COG category details page. This page lists all the COGs within a category and the number of organisms in which each COG is found. Clicking on the number of organisms for a COG takes you to a table showing the number of genes in each organism that belong to the COG (see Figure 5 left panel). At the bottom of the COG category details page, up to 20 organisms can be selected for a comparative profile (shown in Figure 4).

Figure 4. COG category profile page. This page lists all the COGs within a category and the number of genes belonging to the COG in a user-defined set of organisms (see Figure 3). 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). Clicking on an organism abbreviation causes the COGs to be displayed in order of prevalence in that organism. The same type of profile can be generated for enzymes in a KEGG pathway.

Figure 5. COG phylogenetic distribution. Selecting a COG category from the COG Browser leads to the COG Category Details page (see Figure 3). Clicking on the number of organisms for a COG in that page brings you to the table shown on the left in this figure. A button above the table (indicated by the red arrow) leads to a phylogenetic distribution of the COG (right side of this figure). If the COG is present in an organism, the organism name and all of its higher taxonomic categories are displayed in red, and the number of genes belonging to the COG in the organism or higher category are shown in parentheses. The distribution of KEGG pathway enzymes can also be viewed in this way.

Figure 6. Detailed genome comparison with VISTA. One of the alternatives available under the Compare Genomes tab (see Figure 2) is the analysis of a group of closely related genomes with the VISTA viewer. Currently organism groups are available for *Bacillus* (shown), *E. coli*/*Salmonella*, *Mycoplasma*, *Ehrlichia*/*Anaplasma*, *Pseudomonas*, *Pyrococcus*/*Thermococcus*, and *Vibrio*. Each graph shows the level of DNA sequence identity between an organism and a reference genome (in this case *B. subtilis*). Purple regions correspond to genes called in the reference genome. The toolbar allows users to zoom in and out and scroll along the genome. DNA sequence alignments can be viewed between a genome and the reference genome.

Figure 7. Adding your own annotations, part 1. MTH943 has been shown to be a new type of phosphoenolpyruvate carboxylase, but it is annotated as a hypothetical protein in IMG. The proteins in the ortholog table are likely to have the same function. To annotate these genes, select them and add them to the Gene Cart.

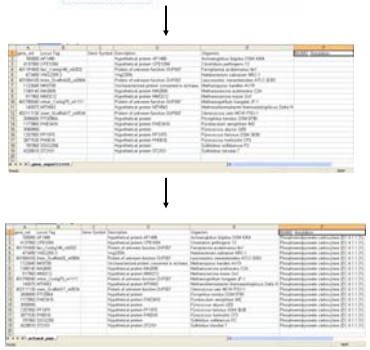

Figure 8. Adding your own annotations, part 2. Go to the Export Genes function on the Gene Cart page, and select the tab-delimited format. An Excel file will be generated with information about the genes and a column labeled "MyIMG Annotation" (middle panel). Add your annotations in this column.

Figure 9. Adding your own annotations, part 3. Go to the "Upload Gene Annotations" function on the MyIMG page. First you will have to log in. Instructions for obtaining a username and password are on the MyIMG page. After you upload your file, the annotations will be displayed in a table underneath the Gene Information table on the Gene Details page. You can delete your annotations through the "View My Annotations" function in MyIMG.

This work was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program and the by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48, Lawrence Berkeley National Laboratory under contract No. DE-AC03-76SF00098 and Los Alamos National Laboratory under contract No. W-7405-ENG-36.