

What's New in IMG 3.5



DOE JOINT GENOME INSTITUTE
U.S. DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

Quick Genome Search:



INTEGRATED MICROBIAL GENOMES

IMG Home Find Genomes Find Genes Find Functions Compare Genomes Analysis Cart My IMG Companion Systems Using IMG

IMG Genomes

	Total
Bacteria	3062
Archaea	121
Eukarya	124
Plasmids	1199
Viruses	2809
Genome Fragments	674
All Genomes	7989
GEBA	112

[Genome by Metadata](#)
[Project Map](#)
[Content History](#)
[What's New](#)
[Using IMG](#)
[About IMG](#)
[FAQ](#)

 Hands on training available at the [Microbial Genomics & Metagenomics Workshop](#)

Database updated: 2011-12-01
Next IMG release: June 2012

The **Integrated Microbial Genomes (IMG)** system ([Nucleic Acids Research, Vol 40, 2012](#)) serves as a community resource for comparative analysis and annotation of all publicly available genomes from three domains of life in a uniquely integrated context. Plasmids that are not part of a specific microbial genome sequencing project and phage genomes are also included into IMG in order to increase its genomic context for comparative analysis.

IMG 3.5: What's New

Count

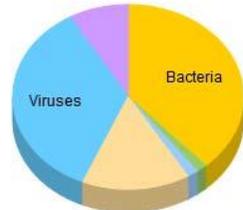
DNA, number of bases

Total Genes

Total Genomes

[IMG Statistics](#) [IMG Family Systems](#)

All Genomes



Category	Count
Bacteria	3062
Archaea	121
Eukarya	124
Plasmids	1199
Viruses	2809
Genome Fragments	674



System Requirements

IMG recommends the use of the following browsers: Firefox Version 9+, IE 9 and Chrome version 16+, because IMG uses the latest in JavaScript 1.8.5

(<http://en.wikipedia.org/wiki/JavaScript#Versions>) and CSS3

(http://en.wikipedia.org/wiki/Cascading_Style_Sheets#Browser_support).

IMG uses third party JavaScript libraries from Yahoo and Google. For users using JavaScript blockers, please allow JavaScript from: yahoo.com, yimg.com, yahooapis.com, google.com and google-analytics.com to view IMG correctly.

IMG uses Java Applets and Java Web Start for some of its viewers. Currently all applets have been tested using Java SE 6 update 30

(<http://www.oracle.com/technetwork/java/javase/downloads/index-jsp-138363.html>).

IMG 3.5 Content Genomes

The content of **IMG 3.5** has been updated with new microbial genomes available in **RefSeq version 48** (July 26, 2011). **IMG 3.5** contains a total of **7,989** genomes consisting of **3,062** bacterial, **121** archaeal, **124** eukaryotic genomes, **2,809** viruses (including bacterial phages), and **1,199** plasmids that did not come from a specific microbial genome sequencing project. In addition, **674** Genomic Fragments (partial genomic fragments of biosynthetic gene clusters) have been integrated in **IMG 3.5**. Among these genomes, **6,329** are **finished** genomes, and **1,652** are **draft** genomes, and **8** are **permanent draft** (i.e., will never be finished) genomes.

Note that **17** microbial genomes from **IMG 3.4** were **replaced** in **IMG 3.5** because (1) a **draft** genome has been replaced by its **finished** version, or (2) the composition of the genome has changed through the addition of new replicons (plasmids, chromosomes). For replaced genomes, the gene object identifiers (gene OIDs) for the protein-coding genes (CDS) were mapped to their new version in **IMG 3.5**. See IMG Data Evolution History for details.

IMG 3.5 also contains **proteomic** data from *Arthrobacter chlorophenicus*1, *Cryptobacterium curtum*, and *Brachybacterium faecium* studies.

IMG Statistics

Various statistics are provided via the **IMG Statistics** link on the home page of IMG, as shown below, including **IMG Total Gene Count** which consists of counting all the genes (protein coding genes, RNA genes) in IMG, except obsolete genes. Compared to **IMG 3.4**, **IMG 3.5** contains **12,607,998** genes, an increase of **973,261** genes.

What's New in IMG 3.5

The following (3) eukaryotic genomes have been included into IMG in order to increase its genomic context for comparative analysis. Currently there are **124 Eukaryotic genomes in IMG 3.5**.

3 Genomes Protists

1. Cryptosporidium muris RN66
2. Entamoeba dispar SAW760
3. Hydra magnipapillata

IMG 3.5 New UI Features

New Home Page

In an attempt to reduce clutter and improve readability, the IMG Home Page (see Figure 1) has been reorganized to display only the total number of the genomes of each domain. Links to finished, draft and permanent draft genome linked have been removed from the Home Page and moved to the IMG Statistics page (see description in next section).

Main section of the Home Page has been reduced down to a mission statement, table of DNA base count, gene count and genome count. Followed by a pie chart of genome break down, colors and counts are the same as the Genome summary table in the “IMG Genomes Column” on the left hand side. Pie slice are clickable and will go to the genome list page.

Home Page includes several quick links below the genome summary.

- Genome by Metadata
- Project Map
- Content History
- What's New
- Using IMG
- About IMG
- FAQ
- Hands on training.

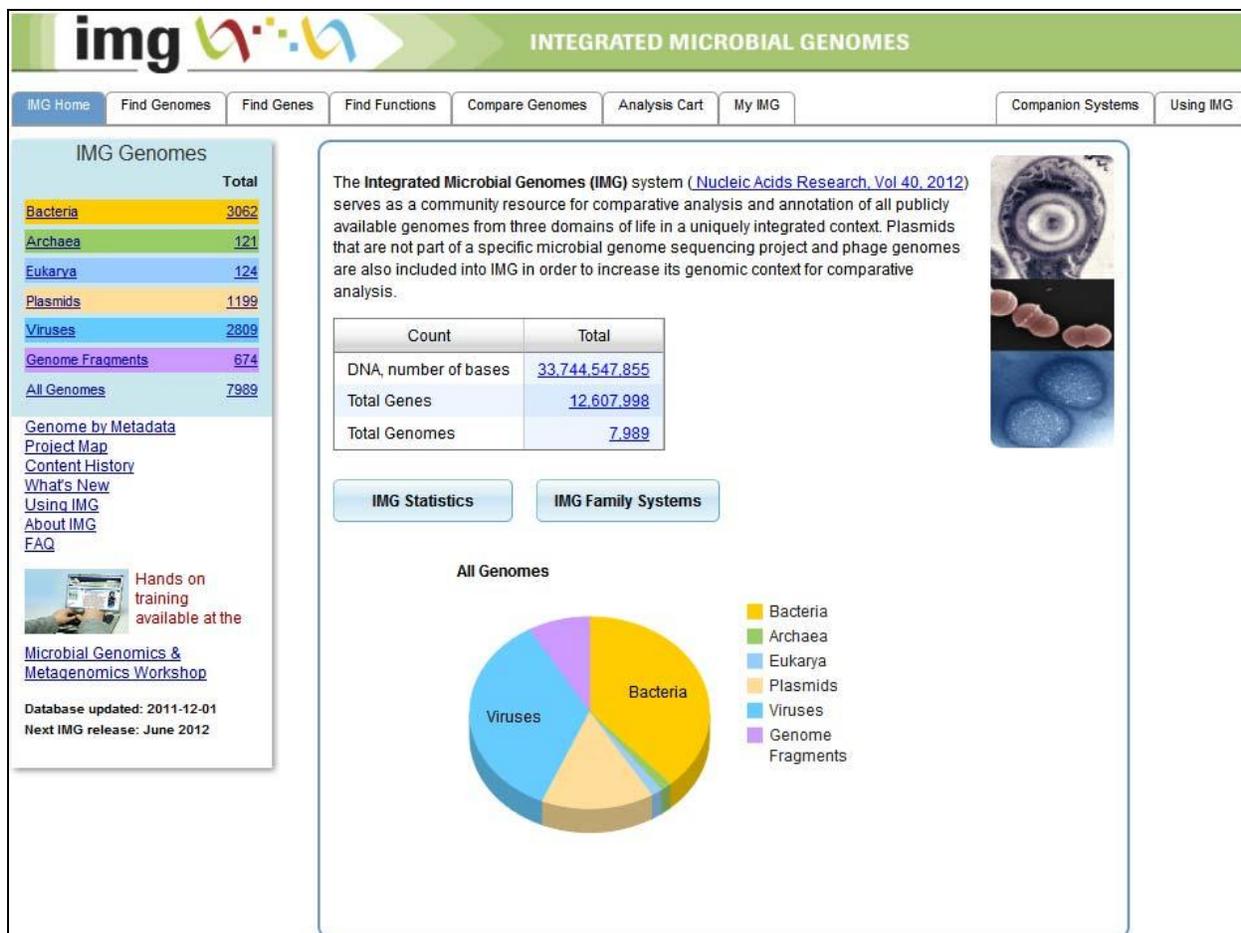


Figure 1: IMG 3.5 Home Page

New IMG Statistics Page

IMG Statistics is a detailed page outlining all statistics available in IMG. This page has undergone a major overhaul in this version of IMG. Not only has it been reorganized, new statistics information has been added. The noticeable difference from previous versions is the tabulation of statistics by domain. IMG Statistics can be reached by clicking on the *IMG Statistics* button on the IMG home page. Please refer to **Error! Reference source not found.** for the location of this button.

IMG Statistics is organized into four separate tabs: *Genome Statistics*, *Gene/Cluster Statistics*, *Function/Pathway Statistics*, and *Experiments*.

1. Genome Statistics: The Genome Statistics tab (see Figure 2) is subdivided into two sections: (a) *Genome Count* and (b) *DNA Statistics*.
 - a. Genome Count table shows the genome counts per domain for genome completion status of *Finished*, *Draft*, and *Permanent Draft*. Clicking on any count will show a new Genome Browser page with detailed information.

- b. The DNA Statistics table shows DNA totals by domain of number of bases, number of coding bases, G+C number of bases, and scaffolds. For convenience, both tables have grand totals for each row and column.

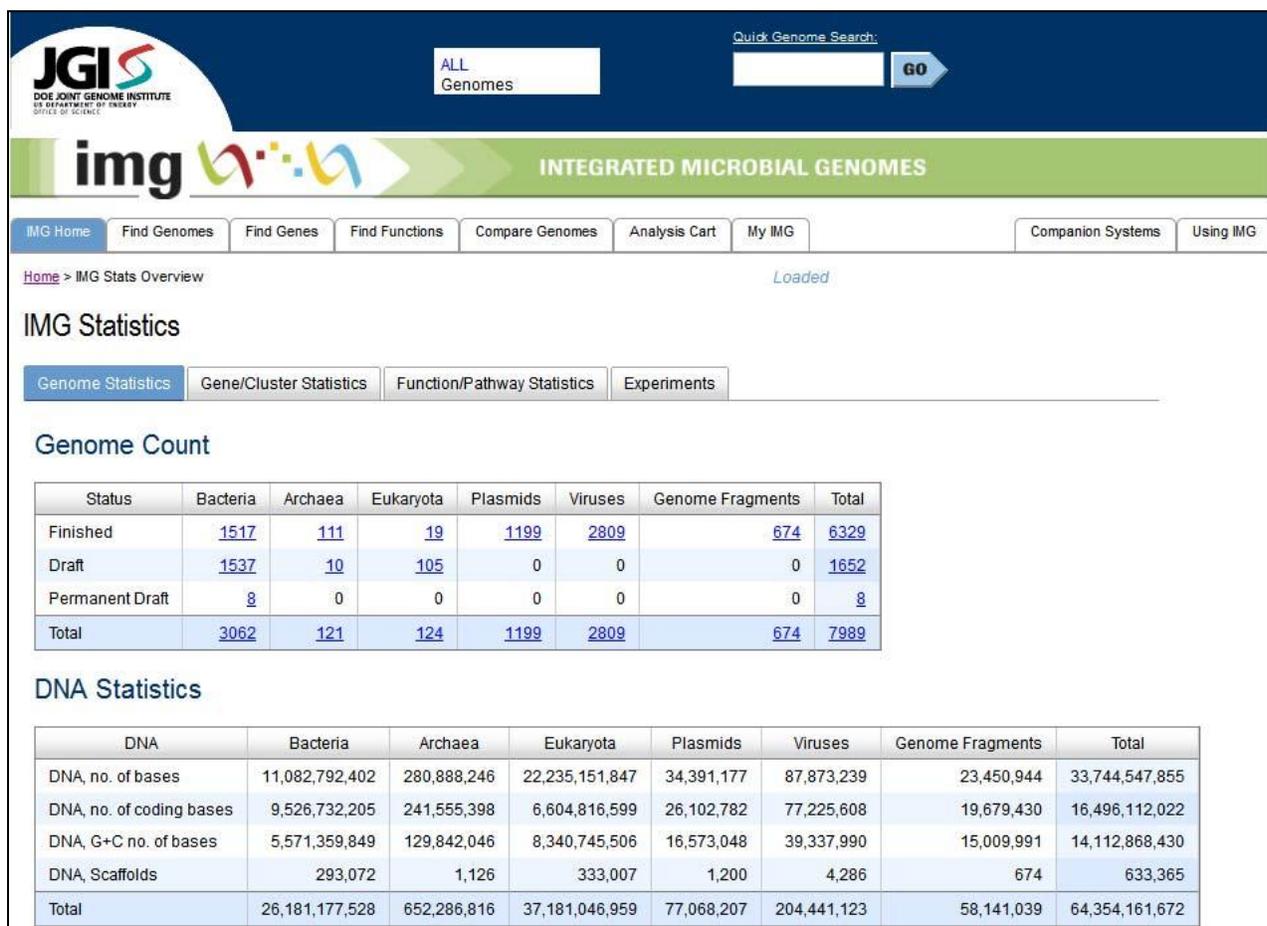


Figure 2: IMG Statistics Page: Genome Statistics

2. Gene/Cluster Statistics: The Gene/Cluster Statistics tab (see Figure 3) is subdivided into two sections: (a) *Gene Statistics* and (b) *Cluster Statistics*.
 - a. Gene Statistics table shows gene counts per domain. The Gene column lists the counts for the available types of genes. For convenience, a grand total column is provided.
 - b. Cluster Statistics is a table that shows the available cluster counts per domain. In addition to the cluster totals, each cluster has two subordinate rows. One row lists the number of protein coding genes with each cluster. The second row lists the percentage of protein coding genes of the total genes per domain. Items with no statistics are marked with “N/A” (Not Available).

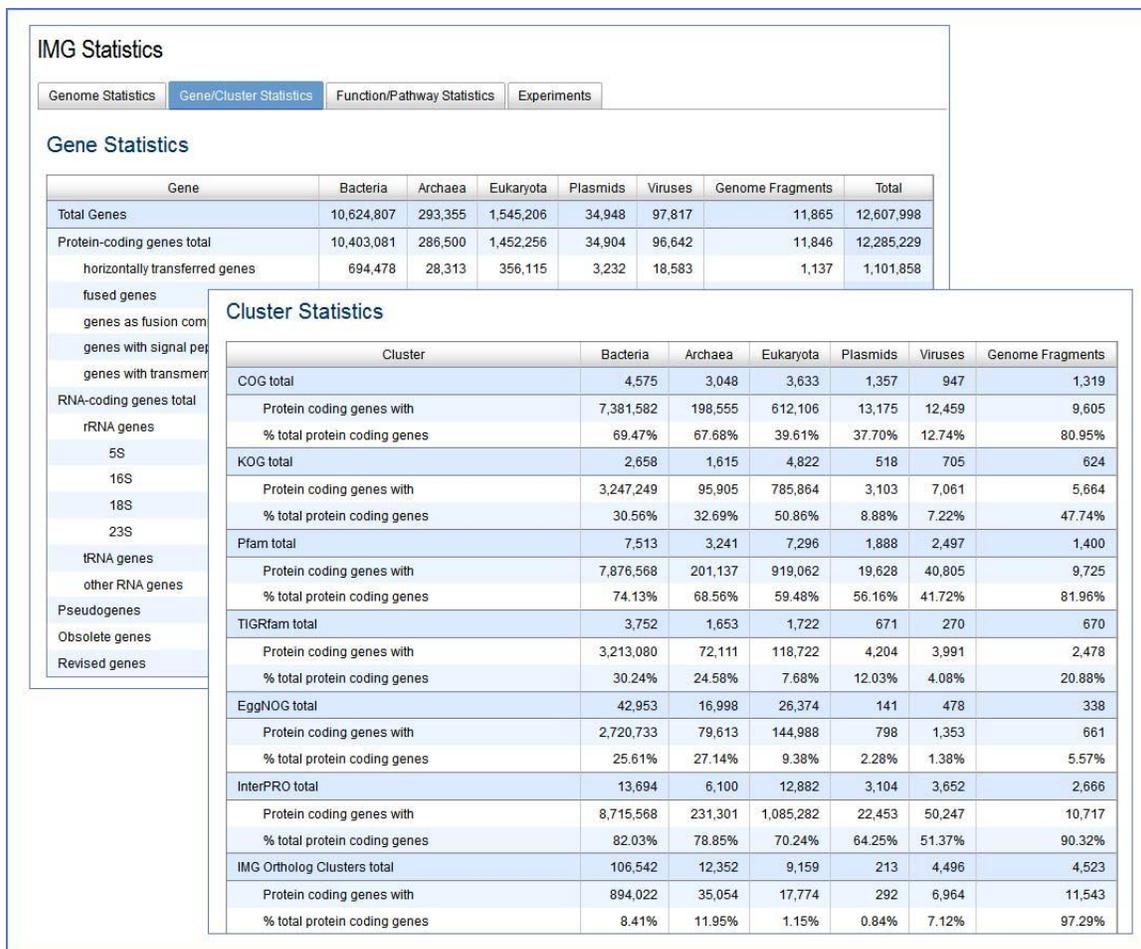


Figure 3: IMG Statistics Page: Gene/Cluster Statistics

3. Function/Pathway Statistics: The Function/Pathway Statistics tab (see Figure 4) is subdivided into two sections: (a) *Function Statistics* and (b) *Pathway Statistics*.
 - a. Function Statistics is a table that shows the counts for available functions per domain. In addition to the function totals, each function has two subordinate rows. One row lists the number of protein coding genes with each function. The second row lists the percentage of protein coding genes with that function out of the total genes per domain. Items with no statistics are marked with “N/A” (Not Available).
 - b. Pathway Statistics is a table that shows the counts for available pathways per domain. In addition to the pathway totals, each pathway has two subordinate rows. One row lists the number of protein coding genes with each pathway. The second row lists the percentage of protein coding genes with that pathway out of the total genes per domain. Items with no statistics are marked with “N/A” (Not Available).

What's New in IMG 3.5

IMG Statistics

Genome Statistics | Gene/Cluster Statistics | **Function/Pathway Statistics** | Experiments

Function Statistics

Function	Bacteria	Archaea	Eukaryota	Plasmids	Viruses	Genome Fragments
Product Names total	459,127	41,768	412,701	9,867	21,211	6,585
Protein coding genes with	6,960,135	171,498	646,300	19,854	56,842	9,994
% total protein coding genes	65.93%	59.07%	41.83%	56.81%	58.11%	84.23%
No Product Names total	21,262	4,752	15,012	360	1,492	46
Protein coding genes with	3,431,513	115,000				
% total protein coding genes	32.50%	39.61%				
Enzyme total	2,764	1,195				
Protein coding genes with	1,318,772	29,670				
% total protein coding genes	12.49%	10.22%				
IMG Terms total	3,276	1,550				
Protein coding genes with	2,009,747	65,270				
% total protein coding genes	19.04%	22.48%				
GO-Molecular Functions total	3,079	1,460				
Protein coding genes with	2,420,170	107,280				
% total protein coding genes	22.92%	36.96%				
KO Terms total	7,241	2,970				
Protein coding genes with	5,220,569	134,630				
% total protein coding genes	49.45%	46.38%				
SEED total	105,285	12,900				
Protein coding genes with	4,000,954	106,210				

Pathway Statistics

Pathway	Bacteria	Archaea	Eukaryota	Plasmids	Viruses	Genome Fragments
COG Pathways total	77	77	77	77	77	77
Protein coding genes with	4,952,474	126,497	321,924	8,563	7,432	5,933
% total protein coding genes	46.91%	43.57%	20.83%	24.50%	7.60%	50.00%
KOG Pathways total	25	25	25	25	25	25
Protein coding genes with	N/A	N/A	N/A	N/A	N/A	N/A
% total protein coding genes	N/A	N/A	N/A	N/A	N/A	N/A
KEGG Pathways total	314	237	330	176	248	206
Protein coding genes with	2,974,567	79,553	291,707	3,077	2,739	3,321
% total protein coding genes	28.18%	27.40%	18.88%	8.80%	2.80%	27.99%
KO Modules total	350	263	318	142	91	172
Protein coding genes with	1,710,573	53,719	119,283	1,718	1,223	1,506
% total protein coding genes	16.20%	18.50%	7.72%	4.92%	1.25%	12.69%
TIGRfam Roles total	100	93	93	77	53	84
Protein coding genes with	2,861,756	58,918	90,110	2,621	3,557	1,655
% total protein coding genes	27.11%	20.29%	5.83%	7.50%	3.64%	13.95%
IMG Pathways total	633	463	435	167	77	176
Protein coding genes with	745,880	23,058	22,420	421	416	609

Figure 4: IMG Statistics Page: Function/Pathway Statistics

- Experiments: The Experiments tab (see Figure 5) lists statistics of the current experiments available in IMG. Each number is hyperlinked to a detailed page of that experiment.

IMG Statistics

Genome Statistics | Gene/Cluster Statistics | Function/Pathway Statistics | **Experiments**

Experiment statistics

Column Selector

Experiment Type	Bacteria	Archaea	Eukaryota	Plasmids	Viruses	Genome Fragments	Total
Protein Experiments	3	0	0	0	0	0	3
Essential Gene Experiments	1	0	0	0	0	0	1
Total	4	0	0	0	0	0	4

Figure 5: IMG Statistics Page: Experiments tab

InterPro Function

A new InterPro function browser has been added in IMG 3.5 under Find Function -> InterPro Browsers (see Figure 6). The UI will display a list of InterPro functions with gene and genome counts. InterPro function can be added to the function cart for additional analysis. See Functional Closure section for more information regarding how to compare InterPro functions with other functional annotations such as COG or Pfam.

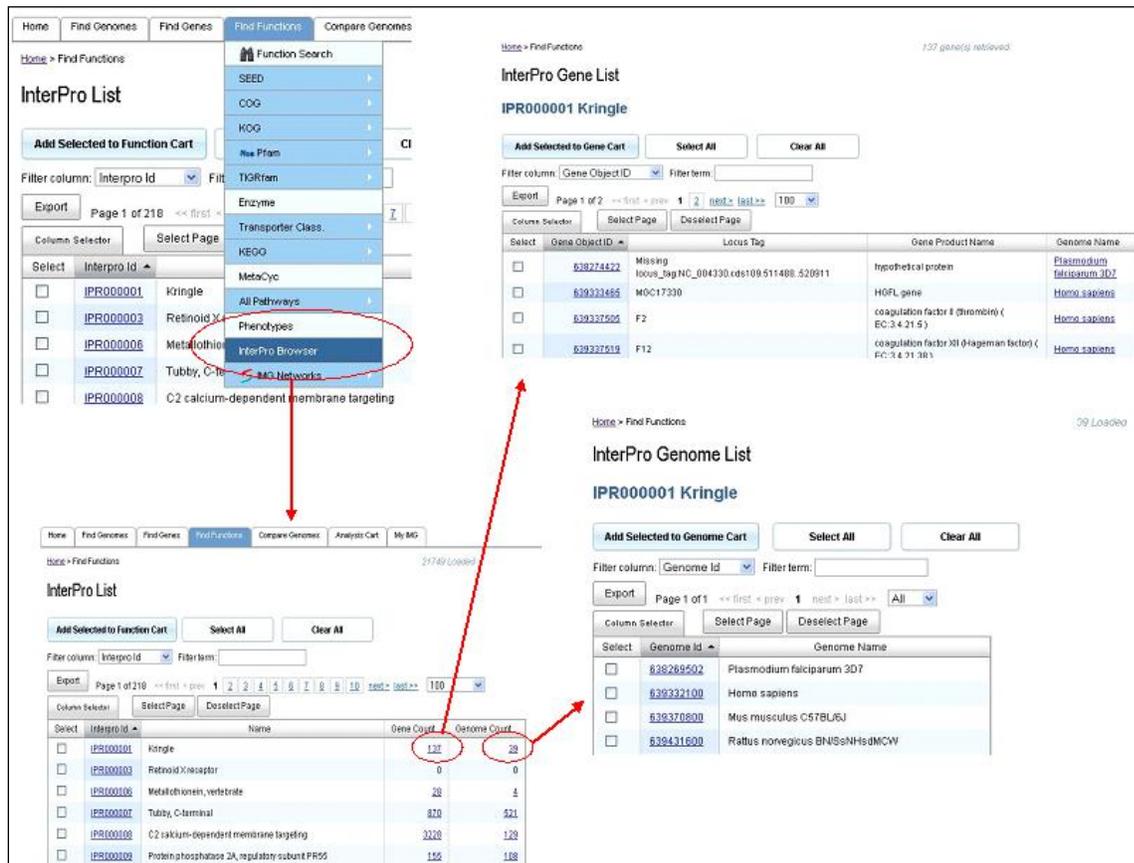


Figure 6: InterPro Function

Genome Gene Ortholog Tool

Genome Gene Ortholog tool under Compare Genomes (see Figure 7) allows user to find gene orthologs in a list of selected reference genomes. The result is a list of genes from the subject genome and the best blast hit of each individual gene to genes of the selected reference genomes. The data is similar to phylogenetic distribution of genes. However, the phylogenetics distribution of genes is the best blast hit across all reference isolate genomes instead of a subset of selected genomes.

What's New in IMG 3.5

Genome Gene Orthology

Find a Genome's Gene Orthologs in a selected sets of Genomes, using Blast best top hits. Please select only **ONE** Find In Genome and maximum **50** Ortholog In Genome

Percent Identity: 60+

Find Genes In	Ortholog In	Ignoring	
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Archaea
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Crenarchaeota
<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	Candidatus Korarchaeum
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Candidatus Korarchaeum cryptofilum OPF8 [F]
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Cenarchaeum
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Cenarchaeum symbiosum A [F]
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Candidatus Nitrosoarchaeum
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Candidatus Nitrosoarchaeum limnia SFB1 [D]
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Nitrosopumilus
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Nitrosopumilus maritimus SCM1 [F]
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Candidatus Nitrosoarchaeum koreensis MY1 [D]

Genome Gene Orthologs

Candidatus Korarchaeum cryptofilum OPF8

[Candidatus Nitrosoarchaeum limnia SFB1](#)
[Nitrosopumilus maritimus SCM1](#)
[Cenarchaeum symbiosum A](#)
[Candidatus Nitrosoarchaeum limnia SFB1](#)

Filter column: Query Gene ID | Filter term:

Export | Page 1 of 1 | << first | prev | 1 | next | last >> | 10

Query Gene ID	Query Gene Name	Query Locus Tag	Query Coords	Query Genome	Subject Gene ID	Subject Gene Name	Subject Locus Tag	Subject Coords
641612314	AAA family ATPase, CDC48 subfamily	Kor_0229	370,646	Candidatus Korarchaeum cryptofilum OPF8	651664710	AAA family ATPase, CDC48 subfamily	MY1_0044	429,704
641613314	AAA family ATPase, CDC48 subfamily	Kor_0229	1,368	Candidatus Korarchaeum cryptofilum OPF8	651664710	AAA family ATPase, CDC48 subfamily	MY1_0044	1,365
641612314	AAA family ATPase, CDC48 subfamily	Kor_0229	1,709	Candidatus Korarchaeum cryptofilum OPF8	641725984	AAA ATPase	CEN8Va_1969	1,712
641612848	Vitamin B6 biosynthesis protein	Kor_0774	44,338	Candidatus Korarchaeum cryptofilum OPF8	651662627	Vitamin B6 biosynthesis protein	MY1_1591	27,319
641612848	Vitamin B6 biosynthesis protein	Kor_0774	44,338	Candidatus Korarchaeum cryptofilum OPF8	651419790	pyridoxine biosynthesis protein	Mim_1057	27,319

Figure 7: Genome Gene Ontology Tool

Artemis – ACT

Artemis – ACT scaffold selection (Compare Genomes -> Synteny Viewers -> Artemis ACT) form has been update via user request to include a select all and clear all ignore check boxes, such that user can easily ignore a majority of scaffolds (see Figure 8).

What's New in IMG 3.5

Contig Reorder - Artemis - ACT

Select *ignore* to remove contig.
Press *Update* button to update order and remove any sel
Press *Next* button to create files for ACT.

Acidilobus saccharovorans 345-15

Ignore	Order	Ext Accession	
<input type="checkbox"/>	<input type="text" value="1"/>	<input type="text" value="NC_014374"/>	Acid

Acidulprofundum boonei T469

Ignore	Order	Ext Accession	
<input type="checkbox"/>	<input type="text" value="1"/>	<input type="text" value="NZ_DS990515"/>	Acid NZ_
<input type="checkbox"/>	<input type="text" value="2"/>	<input type="text" value="NZ_DS990516"/>	Acid NZ_
<input type="checkbox"/>	<input type="text" value="3"/>	<input type="text" value="NZ_DS990517"/>	Acid NZ_
<input type="checkbox"/>	<input type="text" value="4"/>	<input type="text" value="NZ_DS990518"/>	Acid NZ_

Figure 8: Artemis - ACT with Select All and Clear All ignore check boxes

Create Files for Genbank Submission

New Genbank files compatible for NCBI submission are now available. In order to use this feature, first go to the genome detail page of your selected genome, and then scroll down to find the "Generate Genbank File" button. Click this button will bring you to the Generate Genbank File page, and the new function is in the "NEW Create NCBI Genbank Files for submission" section toward the end of the page (see Figure 9). The script will create five files (fsa, gbf, sqn, tbl and val). Link to these files will be emailed to the user and the files will only be available for 24 hours. The email will be sent of either rt-img@cuba.jgi-psf.org or imgsupp@quagmire.jgi-psf.org. Please do reply to email because it was automatically generated.

What's New in IMG 3.5

NEW Create NCBI Genbank Files for submission

My Email (Results will be mailed to you)

Go

Home > Find Genomes loaded.

Your request to process has been successfully submitted. You will be notified via email from **rt-img@cuba.jgi-psf.org** or **imgsupp@quagmire.jgi-psf.org** with a URL to the result. The URL will be valid for **only 24 hours**.

from: imgsupp@quagmire.jgi-psf.org reply forward

subject: **NCBI submission file for Genome 641522611 done**

to:

DO NOT reply to this email. It was automatically generated.
Use links below to download your results. (The URL will be valid for only 24 hours)

Files:

- <http://img-stage.jgi-psf.org/w/tmp/public/public/4550/641522611/641522611.fsa>
- <http://img-stage.jgi-psf.org/w/tmp/public/public/4550/641522611/641522611.gb>
- <http://img-stage.jgi-psf.org/w/tmp/public/public/4550/641522611/641522611.sgn>
- <http://img-stage.jgi-psf.org/w/tmp/public/public/4550/641522611/641522611.tbl>
- <http://img-stage.jgi-psf.org/w/tmp/public/public/4550/641522611/641522611.val>

It is best to have the IMG web page open first.
<http://img-stage.jgi-psf.org/w>
before downloading files.

Figure 9: Create Genbank Files for NCBI Submission

Content History

Content history is now a category bar chart instead of a stacked bar chart (see Figure 10).

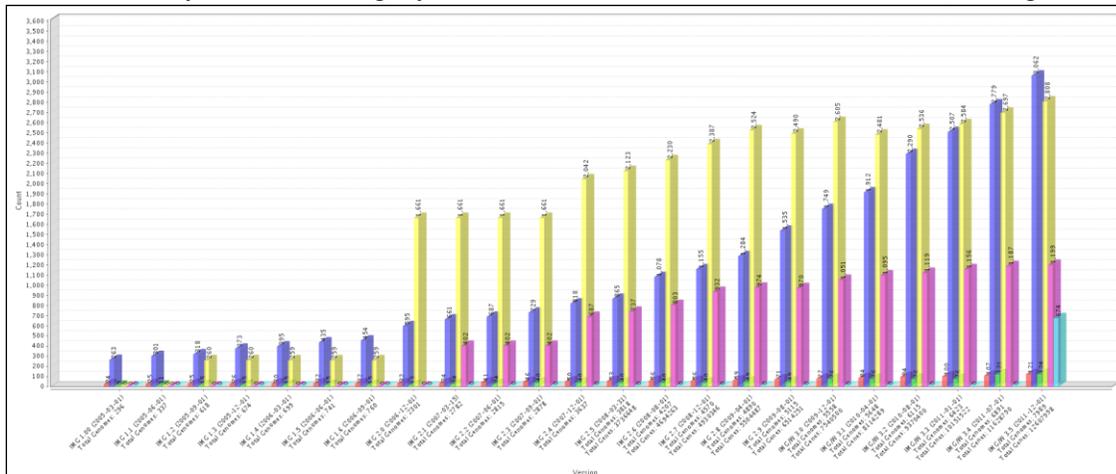


Figure 10: Content History

Tabbed Pages

All analysis carts now use tabs to separate different functionalities. For example, Figure 11 shows the new Gene Cart page with 7 tabs. Function Cart, Genome Cart and Scaffold Cart also use tabbed pages.

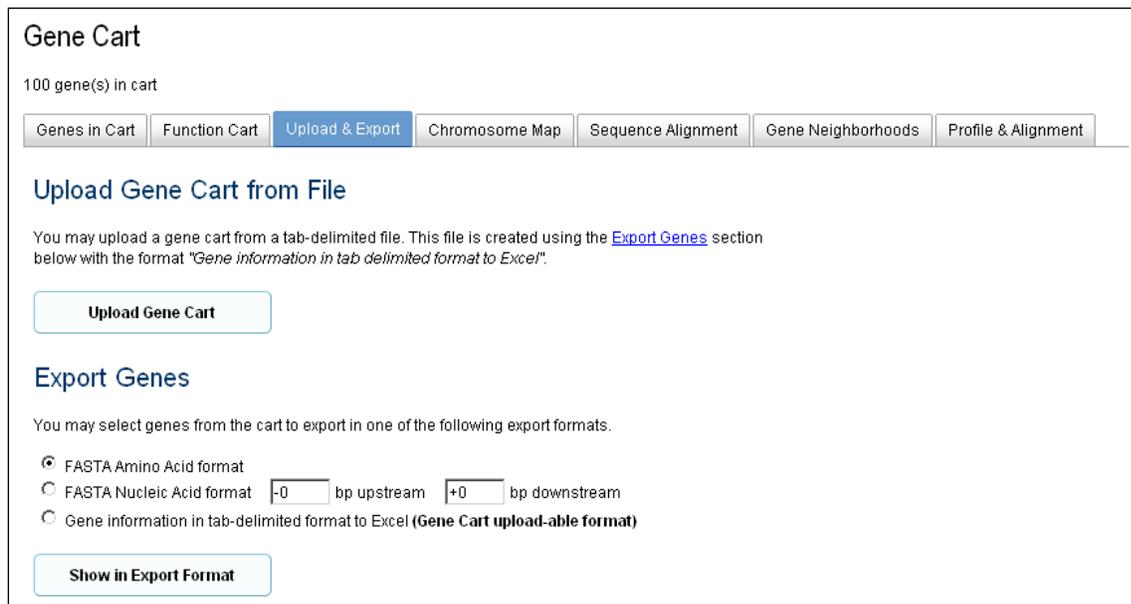


Figure 11: Gene Cart Display with Tabbed Pages

Phylogenetic Profiler for Single Genes

The Phylogenetic Profiles for Single Genes is a new tool for finding genes in a genome of interest qualified by similarity to sequences in other genomes. The function can be found under Find Genes -> Phylogenetic Profilers -> Single Genes.

After going to the Phylogenetic Profiler for Single Genes page, a user can:

- select exactly one genome of interest in the "Find Genes In" column;
- specify genomes that should have homologs in the "With Homologs In" column;
- specify genomes that should not have homologs in the "Without Homologs In" column;
- specify similar cutoffs criteria at the end of the page; and
- click the "Go" button.

After making the desired selection condition (see Figure 12), the tool displays the results in a page with a Summary Statistics table and a Detailed table.

What's New in IMG 3.5

Phylogenetic Profiler for Single Genes

Find genes in genome (bin) of interest qualified by similarity to sequences in other genomes (based on BLASTP alignments). Only use in the profiler.

Genome Completion: [F]inished, [P]ermanent Draft, [D]raft.

Profile

Find Genes In*	With Homologs In	Without Homologs In	Ignoring	Taxon Name
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Archaea
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Crenarchaeota
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Candidatus Korarchaeum
<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Candidatus Korarchaeum cryptofilum OPE8 [F]
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Cenarchaeum
<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	Cenarchaeum symbiosum A [F]
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Candidatus Nitrososarc
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	Candidatus Nitrososarc
<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	Nitrosopumilus
<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	Nitrosopumilus maritimus
<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	Candidatus Nitrososarc
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Acidilobus
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	Acidilobus saccharovorans

Similarity Cutoffs

Max. E-value	1e-5
Min. Amino Acid Percent Identity	30
Exclude Pseudo Genes	No
Algorithm	By Present/Absent Homologs
Min. Taxon Percent With Homologs	100
Min. Taxon Percent Without Homologs	100

Figure 12: Phylogenetic Profiler for Single Genes

The Summary Statistics table list the counts of genes found for the genome of interest. Previous versions of IMG had the counts broken down into the “features” *COG*, *Enzyme*, *Pfam*, *InterPro*, *KO Term*, *Tigrfam*, *No Functional Hit*, and *Unique In IMG*. In this version of IMG, two additional statistics elements *Cassette* and *KEGG Map* have been added to the Summary Statistics table. Furthermore, an additional column that lists the unique number in each “feature” has been included in this IMG version (see Figure 13).

What's New in IMG 3.5

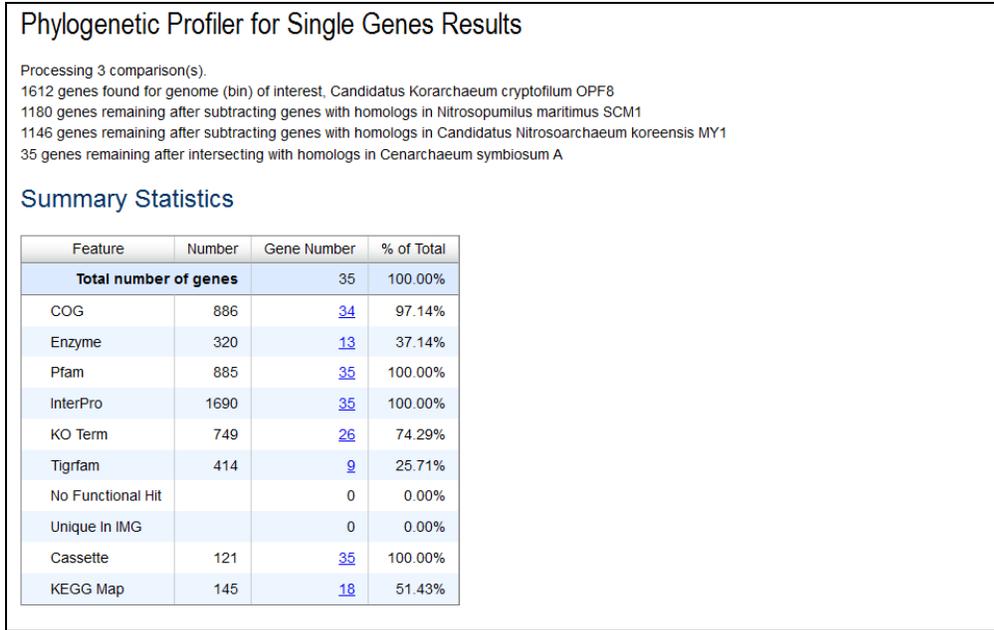


Figure 13: Phylogenetic Profilers for Single Genes - Summary Statistics

Below the Summary Statistics table is a detailed table listing the genes and their information in each row. Rows can be selected and added to the Gene Cart for further analysis. Three additional columns: *Gene Cassette ID*, *KEGG Map Name*, and *KEGG Module Name* have been added to this table (see Figure 14). Users can select a gene from this table and click the "Missing Gene?" button. This function will allow to discover missing genes using TBlastn of the selected gene against selected genomes in the "Without Homologs In" column.

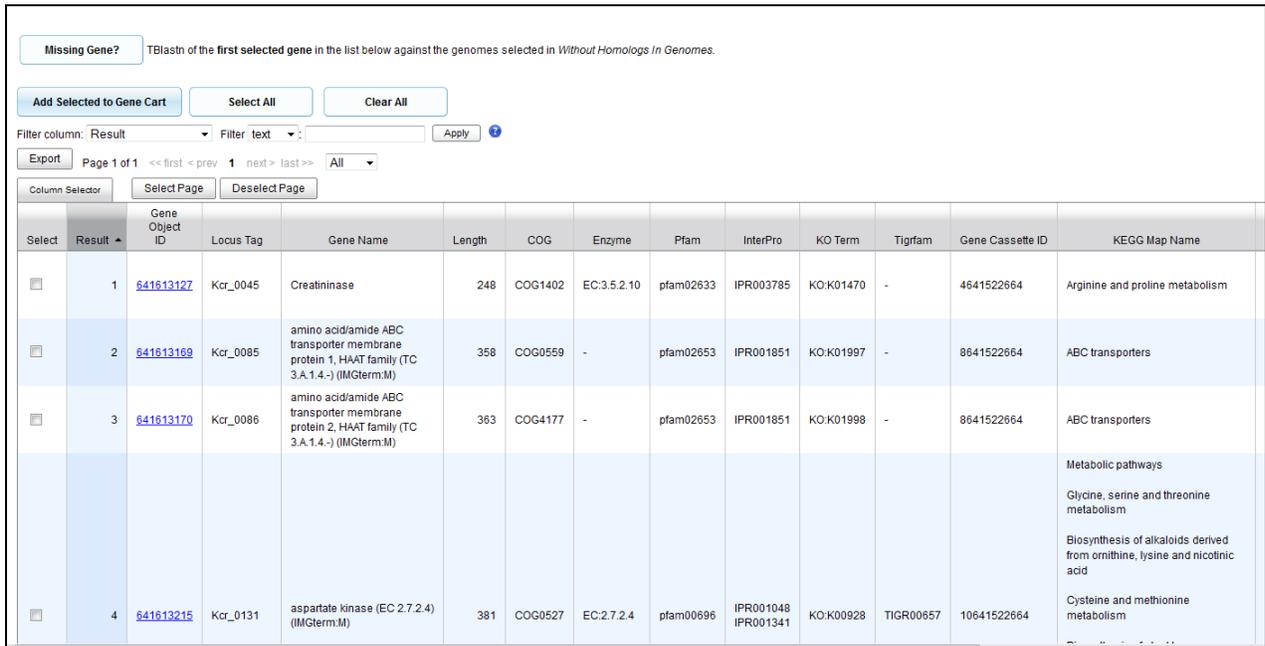


Figure 14: Phylogenetic Profilers for Single Genes - Detail Page

Dynamic table filter enhancements

A common component in all dynamic IMG tables is the filter section. A dynamic table, if it is so configured when developed, can be filtered by a search keyword. The search keyword can be applied to any table column that is enabled for search. Once the keyword is applied to the table, it is said to be “filtered” by that keyword. The table is immediately refreshed to display only rows that satisfy the search criterion. In previous IMG versions, only plain text searches were possible. The current version of IMG has two new enhancements to the filter experience. They are *Regular Expression Search* and *Arithmetic Comparison*.

Regular expressions (or “regex”) offer a concise method for matching strings of text or patterns of characters. To filter a dynamic IMG table by a regular expression, select *regex* from the *Filter* dropdown menu and enter the regex match pattern into the adjoining text field (see Figure 15). To filter the table to the pattern, either click on the *Apply* button OR presser the Enter/Return key from the keyboard. For regular expression syntax, please refer to resources online. Clicking on the ? icon brings up a brief help panel.



Figure 15: Dynamic table filter enhancement

Numeric columns in dynamic IMG tables can be filtered by prefixing or infixing the search keyword with specific comparison operators. Arithmetic comparisons can be applied either in the “text” or the “regex” mode of the *Filter* dropdown. The following comparisons are allowed:

Operator	Operation	Examples
>	Greater than	>400: filter by values greater than 400
<	Less than	<100: filter by values less than 100
>=	Greater than or equal to	>=250: filter by values greater than or equal to 250
<=	Less than or equal to	<=400: filter by values less than or equal to 400
..	Between	400..700: filter by values between 400 and 700 both inclusive

IMG Pathway Profile

IMG users can view assertion profile of a set of selected IMG pathways vs. a set of selected genomes using the new IMG Pathway Profiler in the Function Cart. To use this function, first select and add some IMG pathways into the Function Cart. In the "Profile & Alignment" tab of Function Cart, select at least one genome in the Genome Filter and then go to the new IMG Pathway Profile section to find the "View IMG Pathways vs. Genomes" button.

A new IMG Pathways vs. Genomes Profile page will show up with a table indicate whether each selected IMG pathway is *absent* (or not asserted) or *present* (or asserted) in each selected genome (see Figure 16).

What's New in IMG 3.5

IMG Pathways vs. Genomes Profile

Assertion:
a - absent or not asserted
p - present or asserted
u - unknown
N/A - no data available

Evidence Level (g/R):
g - number of reactions with associated genes.
R - total number of reactions in pathway.

hint: *Mouse over genome abbreviation to see full name.*

Export Page 1 of 1 << first < prev 1 next > last >> All ▾

Column Selector

IMG Pathway	Met mar C775	Met mar S226	Met van SBB6	Met vol A339	Met lab ZZ7	Met mar 1488	Met eve 3719	Met mah 5290
00335 - Acetyl-CoA synthesis by reverse TCA cycle	g(5/8)	g(5/8)	g(5/8)	g(0/8 (1/8 with orthologs))	g(5/8)	g(5/8)	g(4/8)	g(4/8)
00339 - 6-phosphogluconate synthesis via gluconate	g(0/3)	g(0/3)	g(0/3)	g(0/3)	g(0/3)	g(1/3)	g(0/3)	g(0/3)
00433 - Archaeal nucleoid proteins	g(2/2)	g(2/2)	g(2/2)	g(2/2)	g(2/2)	g(2/2)	g(2/2)	g(2/2)

Export Page 1 of 1 << first < prev 1 next > last >> All ▾

Figure 16: IMG Pathways vs. Genomes Profile

The associated evidence level (g/R) shows the number of reactions with associated genes (g) over total number of reactions in pathway (R). Users can click on "a" or "p" to check the pathway assertion detail (see Figure 17).

IMG Pathway Assertion Details

Pathway OID	00433
Pathway Name	Archaeal nucleoid proteins
Genome	Methanococcus vannieli SB
Modify Date	2011-10-28
Modified By	IMG_PIPELINE
Assertion	<input checked="" type="radio"/> Asserted <input type="radio"/> Not Asserted <input type="radio"/> Unknown
Evidence	2/2
Comments	

Evidence

Reaction Order	IMG Terms	Reaction Definition	Genes
1.	<input type="checkbox"/> 00139 archaeal histone HmvA	Archaeal histones	640786007 640786326 640787070
	or		
	<input type="checkbox"/> 00140 archaeal histone		
2.	<input type="checkbox"/> 00141 archaeal histone-like protein	Archaeal non-histone nucleoid proteins	640786594
	or		
	<input type="checkbox"/> 00142 nucleoid protein Sul7d		
	or		
	<input type="checkbox"/> 00143 nucleoid protein Alba		
	<input type="checkbox"/> 00144 nucleoid protein MC1		
	or		
	<input type="checkbox"/> 00145 nucleoid protein HTa		

Add selected terms to function cart.

Figure 17: IMG Pathway Assertion Detail

Functional Closure

The new Functional Closure function allows users to check relationships among closely related COG, Pfam, TIGRfam, Enzyme, KO, InterPro functions and IMG terms in IMG. A user will be able to dynamically construct a functional closure by first selecting a function and then adding closely related functions of different category. The iteration can continue until there is a set (or closure) of related functions of different category. A closure with small amount of functions usually indicates clean functional annotation, while a closure with too many functions often indicates messed up annotations that may require some manual cleanup.

To find a functional closure, a user starts with selecting a function from the Function Cart. If there are multiple selections in the Function Cart, then the first selected function will be used. Figure 18 shows that COG0081 is selected for analysis by a user (Figure 18 (i)). The user then go to the Functional Closure function section in the Profile & Alignment tab to select Pfam as a new functional category and click the Find Closure button (Figure 18 (ii)).

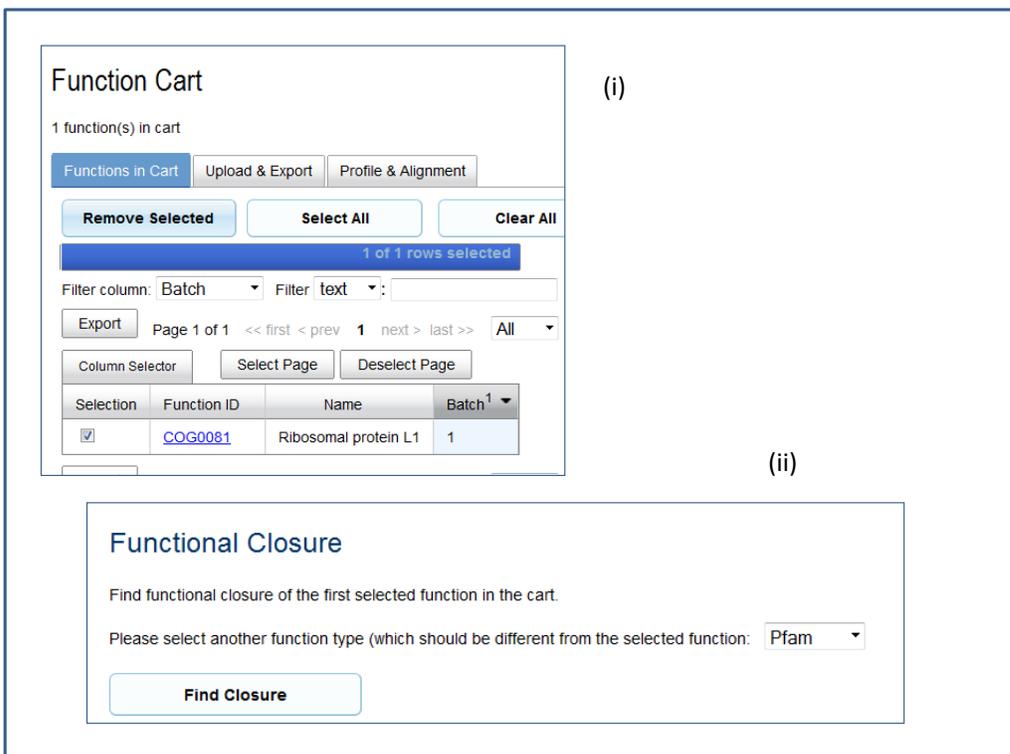


Figure 18: Select a function in Function Cart, and then select a different function category to find closure.

A new Find Functional Closure window will show up. This window consists of 2 parts. The first part shows the gene and genome distribution of the selected function (COG0081 in this case) (Figure 19 (i)). Most genomes only have a single gene annotated with COG0081; however, there are genomes with multiple genes associated with COG0081. Users can select any row in the gene/genome count table, and then click the Show Detail button to see the gene and genome list. Figure 19 (ii) shows the 9 genes annotated with COG0081 in *Populus balsamifera trichocarpa*.

What's New in IMG 3.5

(i)

Find Functional Closure

Selected function: COG0081

Selected function gene count: 3437

The following table lists number of genes in genomes that are annotated with the selected function(s).

Select	Gene Count	No. of Genomes
<input type="radio"/>	1	3077
<input type="radio"/>	2	85
<input type="radio"/>	3	32
<input type="radio"/>	4	3
<input type="radio"/>	5	6
<input type="radio"/>	6	1
<input type="radio"/>	7	3
<input type="radio"/>	8	2
<input type="radio"/>	9	1

Show Detail

Genes with Selected Functions

Selected functions

COG: COG0081

Per genome gene count: 9

Filter column: Genome Filter text: Apply

Export Page 1 of 1 << first < prev 1 next > last >> All

Column Selector Select Page Deselect Page

Selection	Gene OID	Gene Name	Genome
<input type="checkbox"/>	649479691	hypothetical protein	Populus_balsamifera_trichocarpa
<input type="checkbox"/>	649479690	hypothetical protein	Populus_balsamifera_trichocarpa
<input type="checkbox"/>	649435833	hypothetical protein	Populus_balsamifera_trichocarpa
<input type="checkbox"/>	649434177	hypothetical protein	Populus_balsamifera_trichocarpa
<input type="checkbox"/>	649432048	hypothetical protein	Populus_balsamifera_trichocarpa
<input type="checkbox"/>	649427157	hypothetical protein	Populus_balsamifera_trichocarpa
<input type="checkbox"/>	649423927	hypothetical protein	Populus_balsamifera_trichocarpa
<input type="checkbox"/>	649423918	hypothetical protein	Populus_balsamifera_trichocarpa
<input type="checkbox"/>	649412605	hypothetical protein	Populus_balsamifera_trichocarpa

Export Page 1 of 1 << first < prev 1 next > last >> All

Add Selected to Gene Cart Select All Clear All

(ii)

Figure 19: Gene and genome distribution of COG0081.

The second part of the Find Functional Closure window shows Pfam distribution of genes associated with this COG function (Figure 20).

What's New in IMG 3.5

Top 3 Pfam functions displayed.

1 of 3 rows selected

Filter column: Count Filter text: Apply

Export Page 1 of 1 << first < prev 1 next > last >> All

Column Selector Select Page Deselect Page

Selection	Function ID	Function Name	Count	Percent
<input checked="" type="checkbox"/>	pfam00687	Ribosomal protein L1p/L10e family	3431	99.83
<input type="checkbox"/>	pfam00400	WD domain, G-beta repeat	2	0.06
<input type="checkbox"/>	pfam00583	Acetyltransferase (GNAT) family	1	0.03

Export Page 1 of 1 << first < prev 1 next > last >> All

1 of 3 rows selected

Select any functions to add to the Function Cart, to view gene list, or to find closure with a new function type.

Any of the selected functions (union) All of the selected functions (intersection)

Select All Clear All Add Selected to Function Cart List Genes

Select a different new function type to continue: TIGRfam Find Closure

Figure 20: Pfam distribution of genes with COG0081 annotation

What's New in IMG 3.5

The Pfam distribution shows that most of the genes are annotated with pfam00687, while there are some genes associated with pfam00400 or pfam00583. The user can select pfam00687 to be added to the functional closure and then continue to investigate TIGRfam annotation.

Figure 21 shows that most genes annotated with COG0081 and pfam00583 are also associated with TIGRfam function TIGR01169, KO term K02863, InterPro functions IPR002143 or IPR016094 or IPR016095, and IMG term 163: LSU ribosomal protein L1P.

Since this functional closure contains a relatively small number of functions, it is considered as well annotated.

(i) Top 2 TIGRfam functions displayed.

Selection	Function ID	Function Name
<input type="checkbox"/>	TIGR01169	ribosomal protein L1, bacterial-type
<input type="checkbox"/>	TIGR01170	ribosomal protein L1, mitochondrial-type

(ii) Top 4 InterPro functions displayed.

Selection	Function ID	Function Name	Count	Percent
<input checked="" type="checkbox"/>	IPR002143	Ribosomal protein L1	3029	100.00
<input checked="" type="checkbox"/>	IPR016094	Ribosomal protein L1, 2-layer alpha/beta-sandwich	3028	99.97
<input checked="" type="checkbox"/>	IPR016095	Ribosomal protein L1, 3-layer alpha/beta-sandwich	3006	99.24
<input type="checkbox"/>	IPR005878	Ribosomal protein L1, bacterial-type	1269	41.90

(iii) Top 1 KO functions displayed.

Selection	Function ID	Function Name
<input checked="" type="checkbox"/>	KO:K02863	large subunit ribosomal protein L1

(iv) Top 4 IMG Term functions displayed.

Selection	Function ID	Function Name	Count	Percent
<input type="checkbox"/>	ITERM:00163	LSU ribosomal protein L1P	2104	69.90
<input type="checkbox"/>	ITERM:02780	LSU ribosomal protein CL1P	1	0.03
<input type="checkbox"/>	ITERM:02282	transcriptional regulator, GntR family	1	0.03
<input type="checkbox"/>	ITERM:02617	LSU ribosomal protein MRPL1P	1	0.03

Figure 21: Add TIGRfam, Interpro, KO and IMG terms to the Functional Closure.

Protein Family Comparison

The new Protein Family Comparison page (under Find Functions) provides 6 types of statistics and comparisons:

- KO Term Protein Families
- KO Term Genomes & Paralog Clusters
- IMG Term Paralog

What's New in IMG 3.5

- IMG Term Combinations
- IMG Ortholog Clusters
- IMG Term KO

KO Term Protein Families shows KO term distribution across protein families in IMG. Each row in the table list an individual KO term, the number of associated COG count, Pfam count, TIGRfam count, and the total number of unique combinations (see Figure 22).

KO Term Distribution across Protein Families in IMG

Filter column: KO Term ID Filter text:

Page 1 of 152 << first < prev 1 2 3 4 5 6 7 8 9 10 next > last >> 100

Select	KO Term ID	KO Term	KO Definition	COG Count	Pfam Count	TIGRfam Count	Num of Unique Combos
<input type="checkbox"/>	KO:K00001	E1.1.1.1, adh	alcohol dehydrogenase [EC:1.1.1.1]	11	7	9	50
<input type="checkbox"/>	KO:K00002	E1.1.1.2, adh	alcohol dehydrogenase (NADP+) [EC:1.1.1.2]	5	5	2	14
<input type="checkbox"/>	KO:K00003	E1.1.1.3	homoserine dehydrogenase [EC:1.1.1.3]	3	6	1	14
<input type="checkbox"/>	KO:K00004	BDH, butB	(R,R)-butanediol dehydrogenase / diacetyl reductase [EC:1.1.1.4 1.1.1.303]	2	4	2	6
<input type="checkbox"/>	KO:K00005	E1.1.1.6, glcA	glycerol dehydrogenase [EC:1.1.1.6]	1	2	0	3
<input type="checkbox"/>	KO:K00006	E1.1.1.8, GPD1	glycerol-3-phosphate dehydrogenase (NAD+) [EC:1.1.1.8]	1	2	1	5
<input type="checkbox"/>	KO:K00007	dalD	D-arabinitol 4-dehydrogenase [EC:1.1.1.11]	1	2	0	2
<input type="checkbox"/>	KO:K00008	E1.1.1.14, gutB	L-iditol 2-dehydrogenase [EC:1.1.1.14]	6	3	2	14
<input type="checkbox"/>	KO:K00009	mttD	mannitol-1-phosphate 5-dehydrogenase [EC:1.1.1.17]	2	3	0	7
<input type="checkbox"/>	KO:K00010	E1.1.1.18,	myo-inositol 2-dehydrogenase	1	2	0	6

Figure 22: KO Term Distribution across Protein Families in IMG

KO Term Genomes & Paralog Clusters shows KO term distribution across genomes and paralog clusters in IMG. Each row list a KO term with number of genes associated with this KO term, number of genomes, average number of genes per genome, number of genes in paralog clusters, number of genes whose paralog has same term, and average percentage identity (see Figure 23).

What's New in IMG 3.5

KO Term Distribution across Genomes and Paralog Clusters in IMG

Filter column: KO Term ID Filter text:

Export Page 1 of 152 << first < prev 1 2 3 4 5 6 7 8 9 10 next > last >> 100

Column Selector Select Page Deselect Page

Select	KO Term ID	KO Term Name	KO Term Definition	Num of Genes	Num of Genomes	Avg # Genes per Genome	Num of Genes in Paralog Clusters	Num of Genes whose Paralog has Same Term	Avg % identity
<input type="checkbox"/>	KO:K00001	E1.1.1.1, adh	alcohol dehydrogenase [EC:1.1.1.1]	3368	1599	2.11	2421	1426	30.52
<input type="checkbox"/>	KO:K00002	E1.1.1.2, adh	alcohol dehydrogenase (NADP+) [EC:1.1.1.2]	454	295	1.54	263	90	34.2
<input type="checkbox"/>	KO:K00003	E1.1.1.3	homoserine dehydrogenase [EC:1.1.1.3]	2607	2178	1.2	645	591	38.8
<input type="checkbox"/>	KO:K00004	BDH, butB	(R,R)-butanediol dehydrogenase / diacetyl reductase [EC:1.1.1.4, 1.1.1.303]	186	171	1.09	147	26	28
<input type="checkbox"/>	KO:K00005	E1.1.1.6, gidA	glycerol dehydrogenase [EC:1.1.1.6]	963	772	1.25	577	258	30.23
<input type="checkbox"/>	KO:K00006	E1.1.1.8, GPD1	glycerol-3-phosphate dehydrogenase (NAD+) [EC:1.1.1.8]	199	109	1.83	90	89	87.48
<input type="checkbox"/>	KO:K00007	daID	D-arabinitol 4-dehydrogenase [EC:1.1.1.11]	151	147	1.03	105	2	29.94

Figure 23: KO Term Distribution across Genomes and Paralog Clusters in IMG

IMG Term Paralog shows IMG term distribution across genomes and paralog clusters in IMG. Each row list an IMG term with number of genes associated with this IMG term, number of genomes, average number of genes per genome, number of genes in paralog clusters, number of genes whose paralog has same term, and average percentage identity (see Figure 24).

IMG Terms Paralog

Filter column: IMG Term ID Filter text:

Export Page 1 of 43 << first < prev 1 2 3 4 5 6 7 8 9 10 next > last >> 100

Column Selector Select Page Deselect Page

Select	IMG Term ID	IMG Term Name	Num of Genes	Num of Genomes	Avg # Genes per Genome	Num of Genes in Paralog	Num of Genes whose Paralog has Same Term	Avg % identity
<input type="checkbox"/>	00001	chromosomal replication initiator protein DnaA	2365	2338	1.01	816	41	29.89
<input type="checkbox"/>	00002	DNA gyrase subunit A (EC 5.99.1.3)	2273	2265	1.00	1581	8	37.74
<input type="checkbox"/>	00003	DNA gyrase subunit B (EC 5.99.1.3)	2315	2301	1.01	1526	18	35.98
<input type="checkbox"/>	00004	replicative DNA helicase loader DnaC	130	130	1.00	74	0	37.84
<input type="checkbox"/>	00005	replicative DNA helicase loader DnaI	586	581	1.01	225	8	32.79
<input type="checkbox"/>	00006	replicative DNA helicase loader DnaB	744	701	1.06	205	38	35.71

Figure 24: IMG Terms Paralog

IMG Term Combinations shows IMG term distribution across protein families in IMG. Each row in the table list an individual IMG term, the number of associated COG count, Pfam count, TIGRFam count, KO count, and the total number of unique combinations (see Figure 25).

What's New in IMG 3.5

IMG Term Combinations

Filter column: Term ID Filter text:

Page 1 of 40 << first < prev 1 [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) next > last >> 100

Select	Term ID	Name	COG Count	Pfam Count	TIGRfam Count	KO Count	Num of Unique Combos
<input type="checkbox"/>	00001	chromosomal replication initiator protein DnaA	1	6	2	1	13
<input type="checkbox"/>	00002	DNA gyrase subunit A (EC 5.99.1.3)	1	3	5	2	12
<input type="checkbox"/>	00003	DNA gyrase subunit B (EC 5.99.1.3)	1	4	4	2	18
<input type="checkbox"/>	00004	replicative DNA helicase loader DnaC	1	1	0	2	2
<input type="checkbox"/>	00005	replicative DNA helicase loader DnaI	1	3	0	2	9
<input type="checkbox"/>	00006	replicative DNA helicase loader DnaB	4	9	5	3	17
<input type="checkbox"/>	00007	primary replicative DNA helicase (EC 3.6.1.-)	2	4	4	1	15
<input type="checkbox"/>	00008	DNA primase (EC 2.7.7.-)	1	7	1	1	14
<input type="checkbox"/>	00024	bacterial translation initiation factor 1 (bIF-1)	1	1	3	1	3
<input type="checkbox"/>	00025	bacterial translation initiation factor 2 (bIF-2)	3	5	2	1	21

Figure 25: IMG Term Combinations

IMG Terms IMG Ortholog Clusters shows IMG terms and its associated IMG ortholog cluster count (see Figure 26). Users can click on counts to see more details.

IMG Terms IMG Ortholog Clusters

Filter column: IMG Term ID Filter text:

Page 1 of 43 << first < prev 1 [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) next > last >> 100

Select	IMG Term ID	Name	IMG Ortholog Cluster Count
<input type="checkbox"/>	00001	chromosomal replication initiator protein DnaA	1210
<input type="checkbox"/>	00002	DNA gyrase subunit A (EC 5.99.1.3)	1283
<input type="checkbox"/>	00003	DNA gyrase subunit B (EC 5.99.1.3)	1291
<input type="checkbox"/>	00004	replicative DNA helicase loader DnaC	31
<input type="checkbox"/>	00005	replicative DNA helicase loader DnaI	284
<input type="checkbox"/>	00006	replicative DNA helicase loader DnaB	373
<input type="checkbox"/>	00007	primary replicative DNA helicase (EC 3.6.1.-)	1120
<input type="checkbox"/>	00008	DNA primase (EC 2.7.7.-)	1077
<input type="checkbox"/>	00024	bacterial translation initiation factor 1 (bIF-1)	892
<input type="checkbox"/>	00025	bacterial translation initiation factor 2 (bIF-2)	1178
<input type="checkbox"/>	00026	bacterial translation initiation factor 3 (bIF-3)	1150

Figure 26: IMG Terms IMG Ortholog Clusters

IMG Term KO shows the relationship between an IMG term with KO, paralog and IMG ortholog clusters. Each row in the table lists an IMG term, number of genes associated with this IMG term that also has KO terms, numbers of genes without KO terms, numbers of genes in paralog and IMG ortholog cluster count (see Figure 27).

What's New in IMG 3.5

IMG Term KO

Filter column: IMG Term ID Filter: text

Page 1 of 43 << first < prev 1 2 3 4 5 6 7 8 9 10 next > last >> 100

Select	IMG Term ID	Name	Num of Genes with KO	Num of Genes without KO	Num of KO terms	Num of Genes in Paralog	IMG Ortholog Cluster Count
<input type="checkbox"/>	00001	chromosomal replication initiator protein DnaA	2362	3	1	816	1210
<input type="checkbox"/>	00002	DNA gyrase subunit A (EC 5.99.1.3)	2273	0	2	1561	1283
<input type="checkbox"/>	00003	DNA gyrase subunit B (EC 5.99.1.3)	2314	1	2	1526	1291
<input type="checkbox"/>	00004	replicative DNA helicase loader DnaC	130	0	2	74	31
<input type="checkbox"/>	00005	replicative DNA helicase loader DnaI	575	11	2	225	284
<input type="checkbox"/>	00006	replicative DNA helicase loader DnaB	689	55	3	205	373

Figure 27: IMG Term KO