IMG 3.2 is the **20th** release of the **Integrated Microbial Genomes** (IMG) genomic data management and analysis system. **IMG 3.2** was released on **September 7th, 2010.**

IMG 3.2 Content

Genomes

The content of **IMG 3.2** has been updated with new microbial genomes available in **RefSeq version 41** (June 14, 2010). **IMG 3.2** contains a total of **6,115** genomes consisting of **2,290** bacterial, **94** archaeal, **76** eukaryotic genomes, **2,536** viruses (including bacterial phages), and **1,119** plasmids that did not come from a specific microbial genome sequencing project. Among these genomes, **4,850** are **finished** genomes, and **1,261** are **draft** genomes, and **4** are **permanent draft** (i.e. will never be finished) genomes.

Note that **31** microbial genomes from **IMG 3.1** were **replaced** in **IMG 3.2** because (1) a **draft** genome has been replaced by its **finished** version, or (2) the composition of the genome has changed through the addition of new replicons (plasmids, chromosomes).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. 3.2**. See IMG Data Evolution History for details.

IMG 3.2 also contains **proteomic** data from *Arthrobacter chlorophenolicus*[1], *Cryptobacterium curtum*, and *Brachybacterium faecium* studies.

[1] Unell, M., Abraham, P.E., Shah, M., Zhang, B. et al. (2009) Impact of Phenolic Substrate and Growth Temperature on the *Arthrobacter chlorophenolicus* Proteome. *J. Proteome Res.* **8** (4): 1953-1964.

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 3.1**, **IMG 3.2** contains **9,370,400 genes**, an increase of **1,256,224** genes.

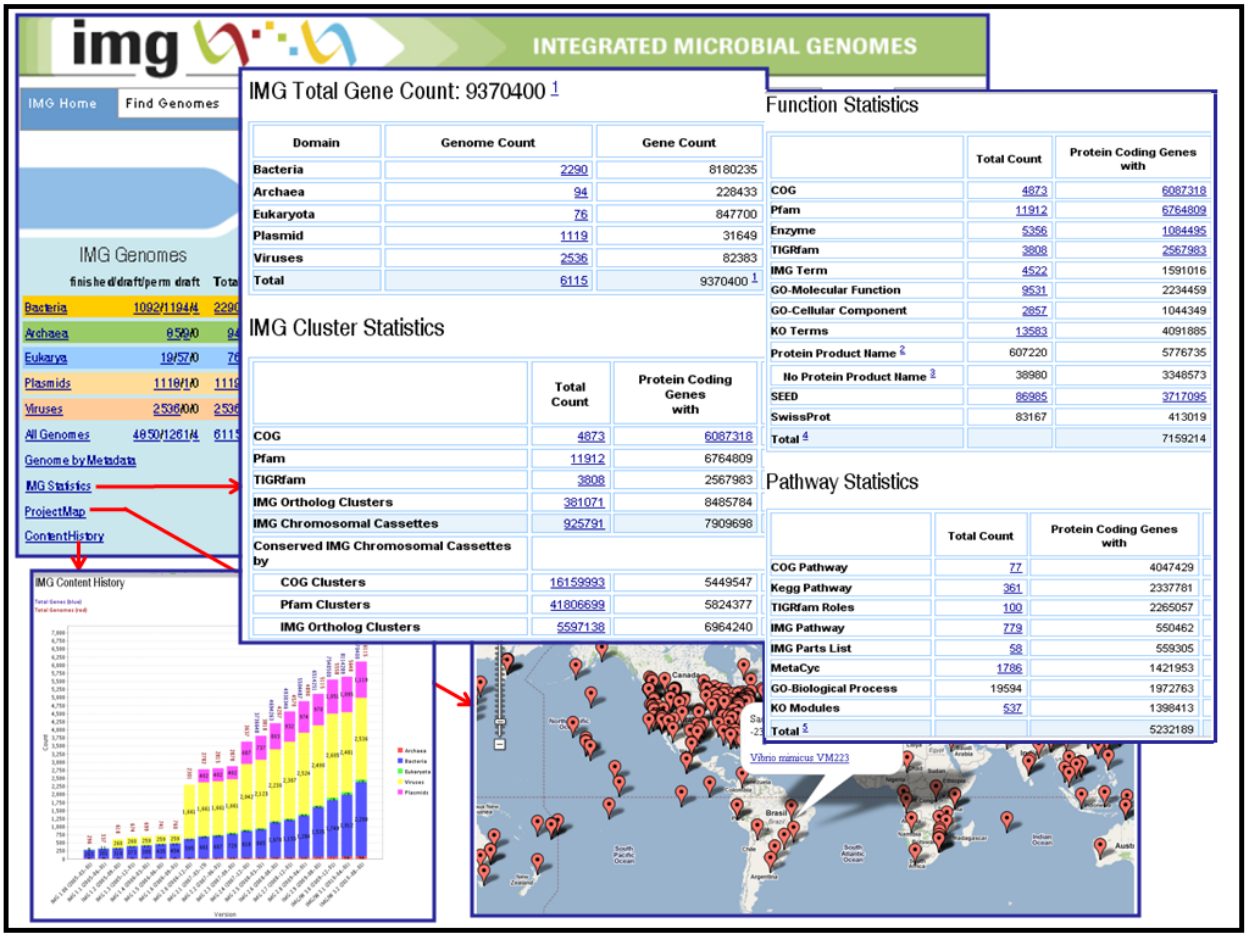


Figure 1. IMG Statistics.

The **Content History** link on the home page of IMG leads to a bar chart, as illustrated above, representing the growth of IMG in terms of number of genomes and genes since the release of its first version in March 2005. Note that the number of genes is recorded only for the last five versions of the system.

The **Project Map** link on the home page of IMG leads to a Google Map, as illustrated above, displaying the location of isolation sites for genomes that are associated with longitude/latitude coordinates in GOLD (http://www.genomesonline.org/).

IMG 3.2 User Interface

The IMG User Interface (UI) has been extended in order to improve its overall functionality and usability. The main IMG UI changes include:

1. **New features**
    a. Site Map
    b. Find Alignment
    c. Distance Tree
2. **Extended features**
    a. Genome Search
        i. by Fields
        ii. by Metadata
        iii. Genome Filter
    b. Find Function – SEED , SwissProt and KEGG added
    c. Generate GenBank File for large genomes
    d. Table Display
        i. Search / Filter rows
        ii. Select / Deselect page
        iii. Show / Hide Column
    e. Gene/Function/Scaffold Carts
        i. Gene Cart
            - add to Function Cart
            - Gene Cart Neighborhoods red underlines genes in Gene Cart.
            - Toggle select.
        ii. Function Cart
            - Add to Gene Cart
            - Toggle select
        iii. Scaffold Cart
            - Toggle select
    f. Protein Expression Studies Comparative Analysis Tools

New Features

Site Map

A new **Site Map** link is available in the **User Guide** section of **Using IMG**, as shown in Figure 2, leading to a list of IMG's menu and analysis tools.

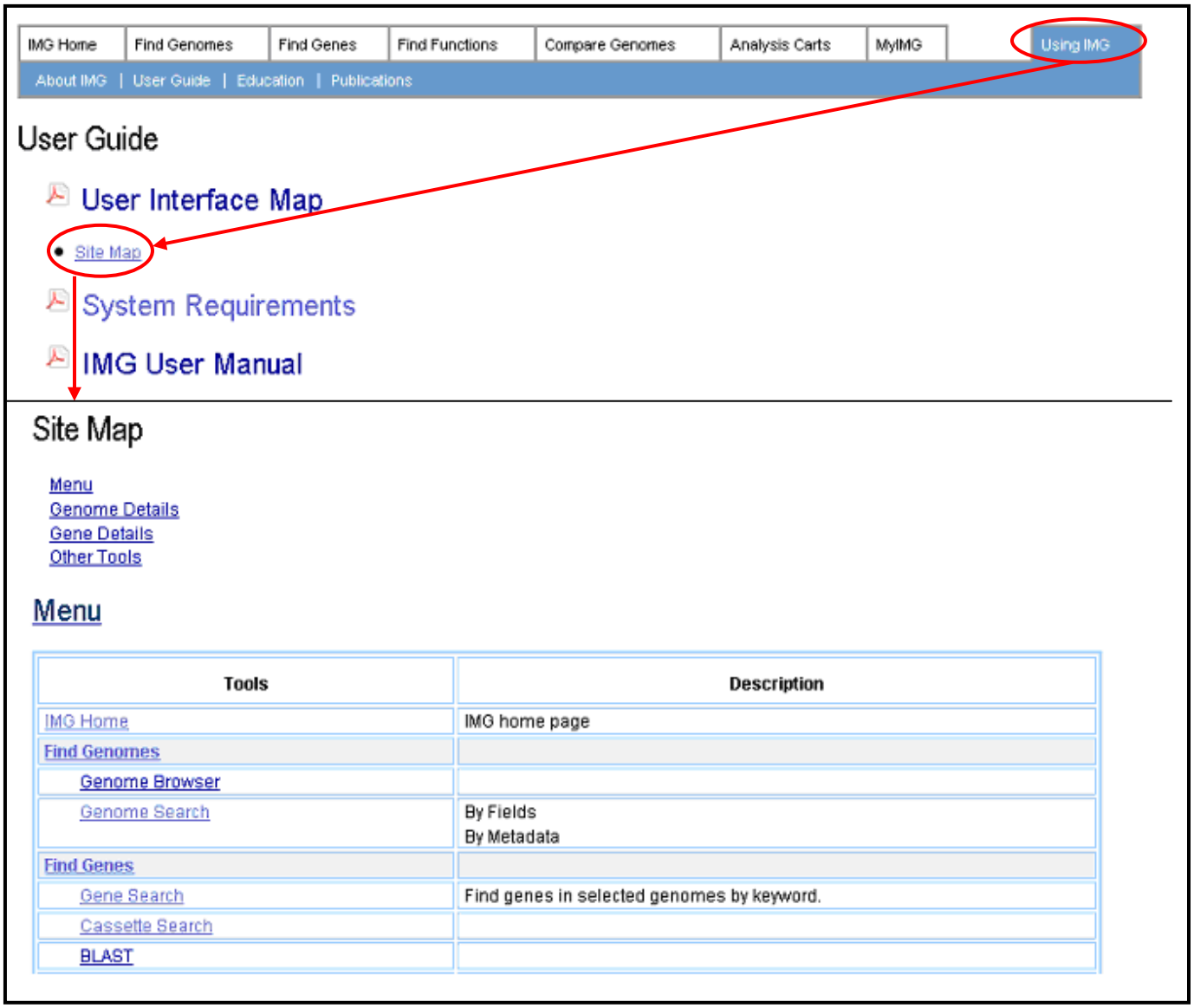


Figure 2. IMG Site Map.

Find Alignment

Function Alignment is a new tool that lists alignments of functional prediction for genes. User first enters keyword, next selects COG or Pfam or both as filter to look for specific functions and genes. The **Genome Filter** is provided to limit the scope of search. The result page is split into two parts through tab view: COG and Pfam. If the **Function Alignment** search involves only one type of function, then either COG or Pfam tab view is displayed (Figure).

Function Alignment can be also accessed from **Gene Cart** and **Function Cart**.

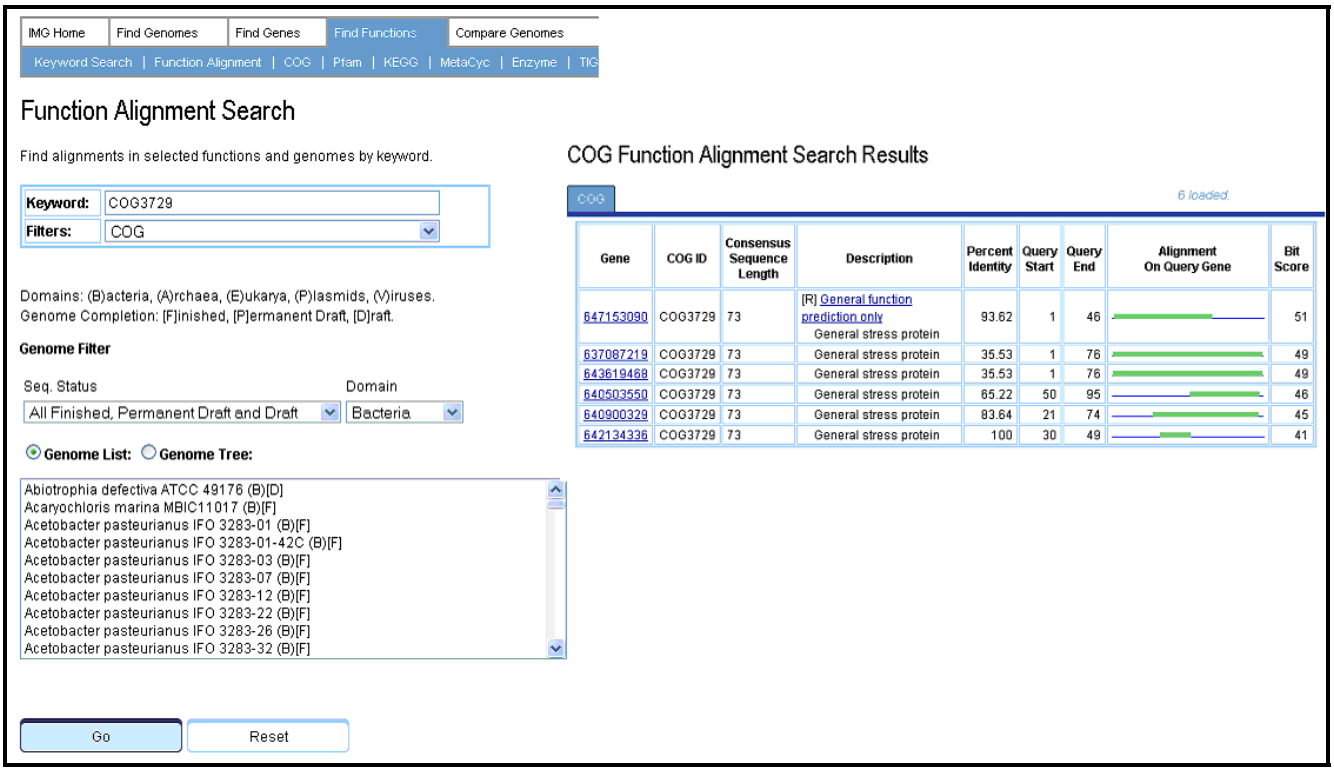


Figure 3. Function Alignment applied on COG3729 for all bacterial genomes.

Phylogenetic Distance Tree

Distance Tree is a new tool available under the **Compare Genomes** tag on the IMG main menu, as shown in Figure 4.

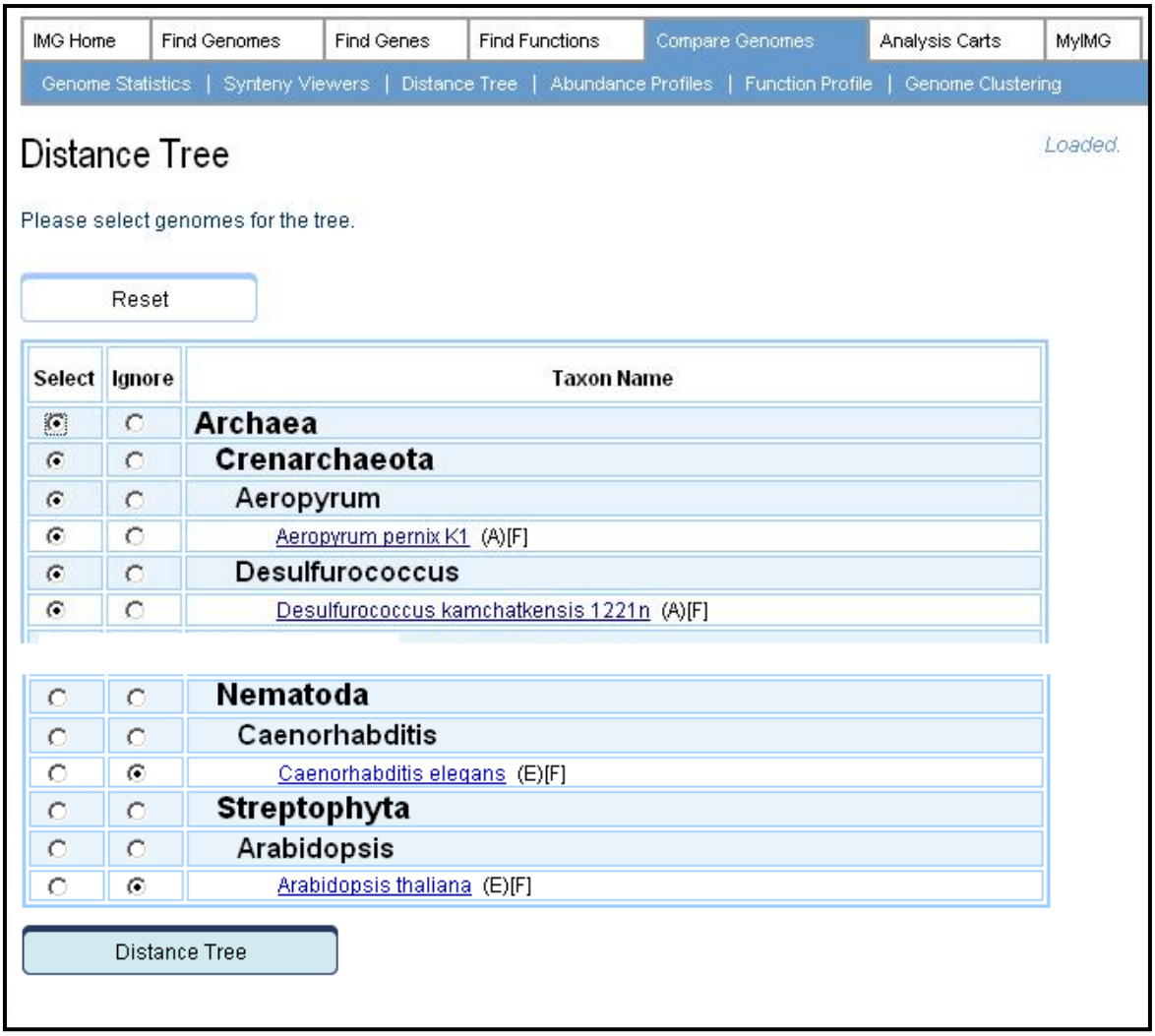


Figure 4. Distance Tree.

Genomes are first selected for distance tree calculation, as shown in Figure . The tool then computes the phylogenetic distance between genomes and displays the tree in the Archaeopterix applet, as shown in Figure . This applet allows users to save the analyzed tree as a pdf file. Each node in the tree links to the gene page for that node.

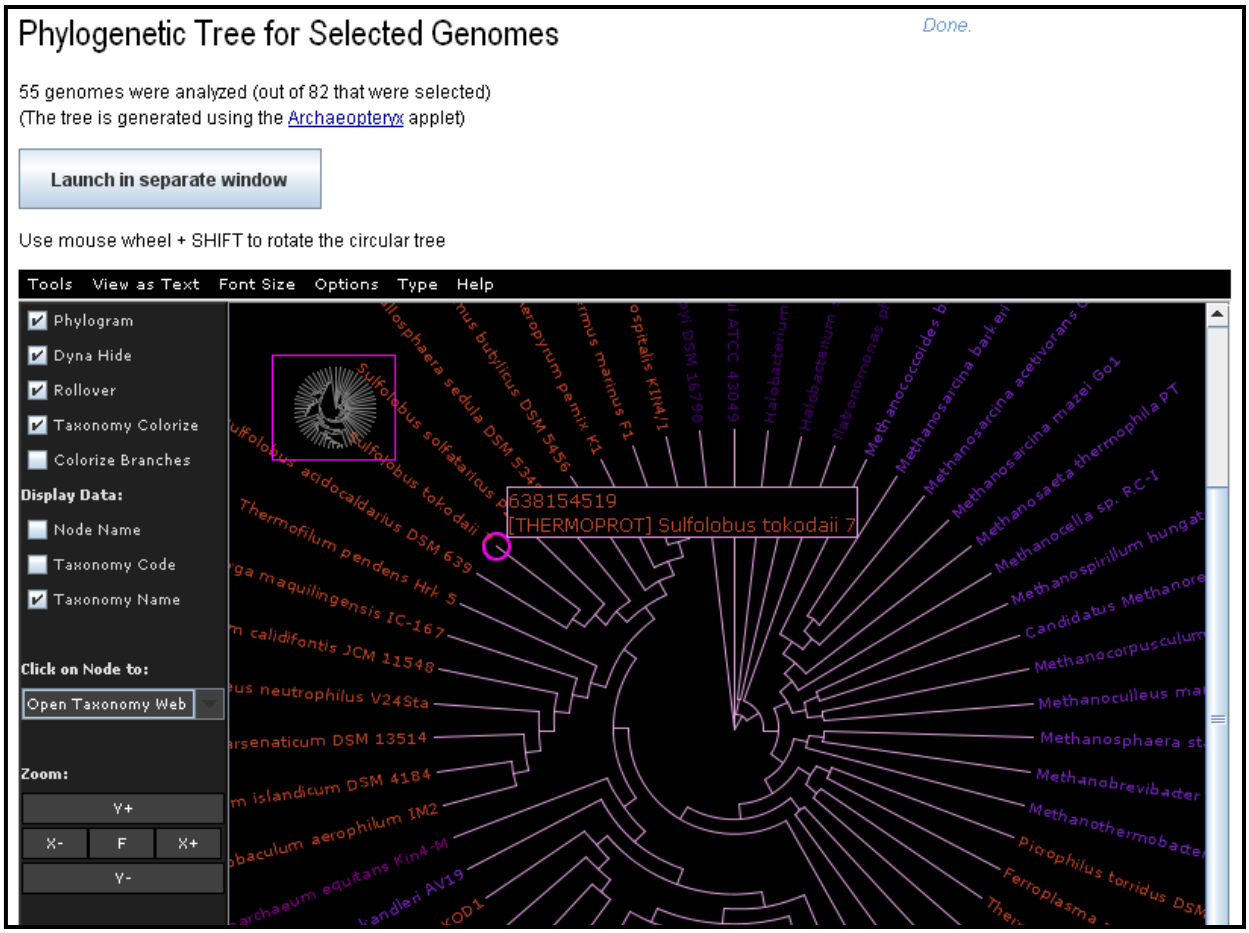


Figure 5. Distance tree viewed with **Archaeopterix** applet**.**

Extended User Interface Features

Genome Search

The Genome Search page in Find Genomes is split into two parts through the application of tab view: "by Fields" and "by Metadata", as illustrated in Figure .

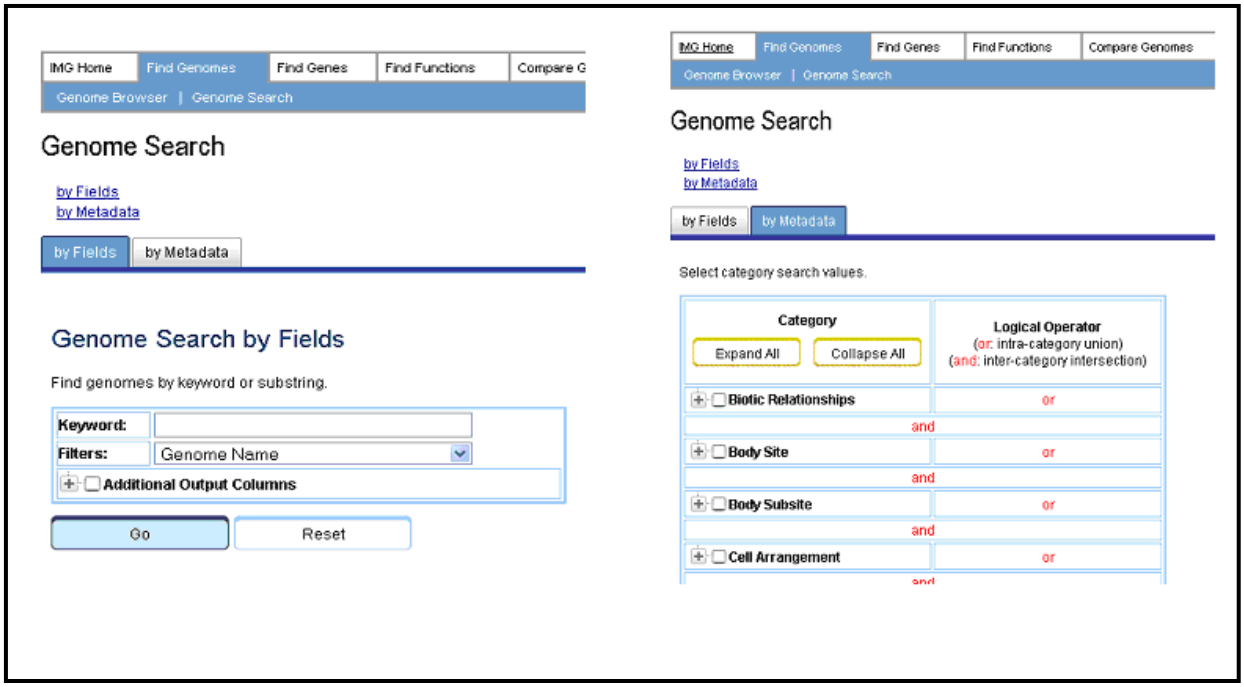


Figure 6. Genome Search: search by fields or metadata.

Genome Search by Fields

The search based on "NCBI Project ID" is replaced with "Refseq Project ID" and "GenBank Project ID". The new "**All field names***" is aka 'Quick Genome Search' on the top. In addition to the traditional search on genome name, taxonomy ranks, certain attributes and various IDs, this search broadened its coverage to include area such as "RefSeq Project ID", "GenBank Project ID", "Is Public", "IMG Release", "IMG Product Assignment", "Gold ID in IMG Database" and "Add Date" (in format: yyyy-mm-dd), as illustrated in Figure 7.

The result page now displays "Domain" and "Genome Completion" as default. In addition, through the selection in search page, the result can display a variety of "Additional Output Columns":

1. Taxon Object ID; NCBI Taxon ID; RefSeq Project ID; GenBank Project ID
2. Phylum; Class; Order; Family; Genus; Species; Strain
3. Sequencing Center; Funding Agency
4. Add Date; Is Public
5. IMG Submission ID; IMG Release; IMG Product Assignment; GOLD ID in IMG Database

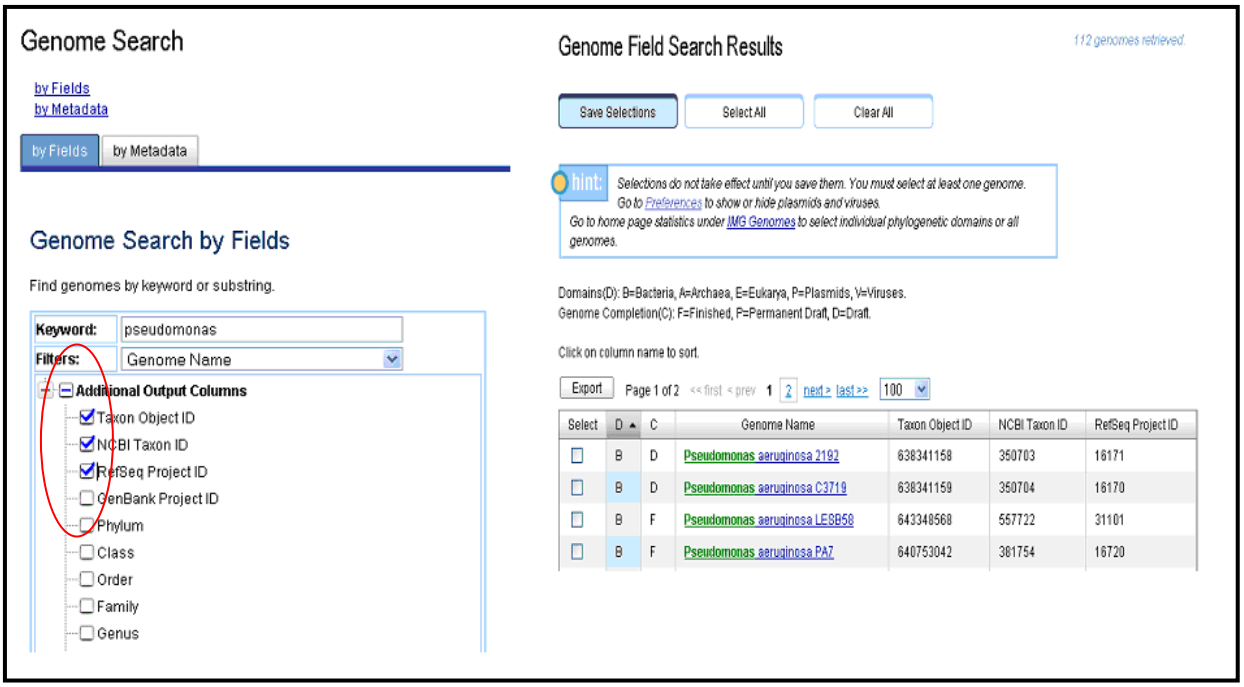


Figure 7. Example of **Genome Search** by **Fields** and its results.

Genome Search by Metadata

Each metadata category is now displayed as a tree. User can open the tree, or expand all trees. The logical operator for intra-category is union, and for inter-category is intersection.

The Habitat category is replaced with five other ecosystem classifications: Ecosystem, Ecosystem Category, Ecosystem Type, Ecosystem Subtype and Specific Ecosystem (**Error! Reference source not found.**).

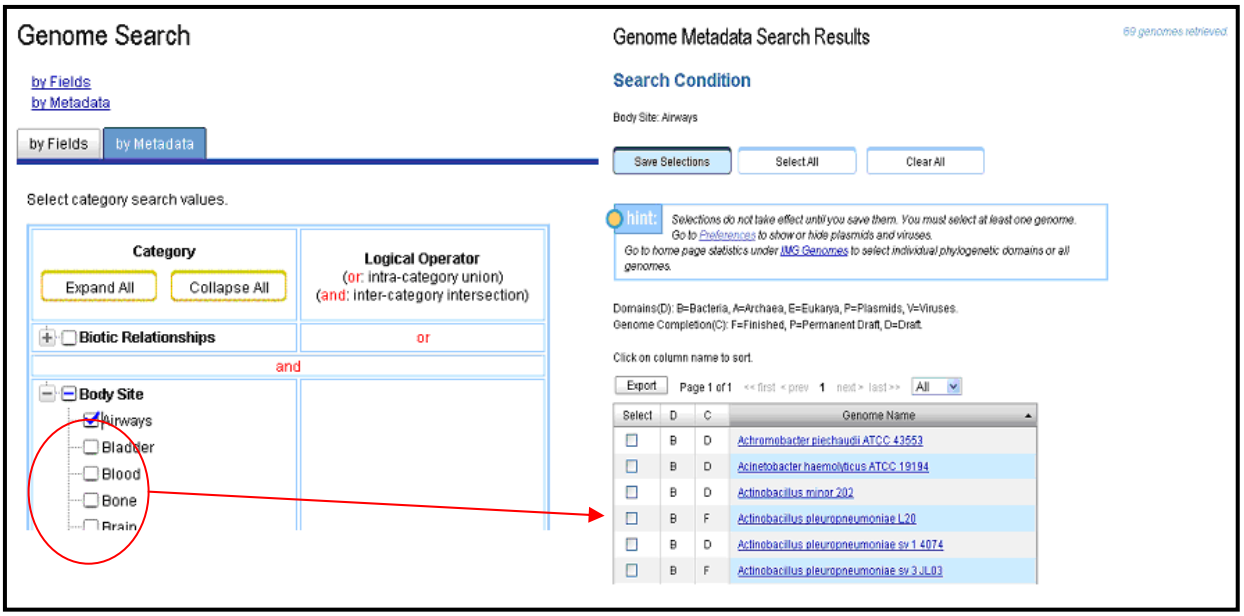


Figure 8. Example of **Genome Search** by **metadata** and results.

Genome Filter

The Genome List Filter is renamed into Genome Filter, since it added a "Genome Tree" feature in addition to the default "Genome List". This tree feature is accessible through selection. User has three options to open a tree (**Error! Reference source not found.**):

1. Single click to open the node itself only.
2. Double click to open the node and its subsequent children.
3. Use button to expand all nodes recursively.

In pages that require multiple selection of genomes, such as "**Gene Search**", "**Function Search**", "**Function Alignment**", "**Cassette Search**" and "**Artemis ACT**", a checkbox is applied to each node, and the highlight selection is propagated both up and down. In the page of "**A Genome vs Metagenomes**", only the end node hosts a radio since it requires single selection of genome. A radio selection excludes the others.

The Genome Filter is also introduced into pages of "**Gene Cart**", **"Function Cart**" and "**Gene Product Name Search Results**". In pages of "Gene Cart" and "Function Cart", the Genome Filter is share-used by several profile and alignment tools.

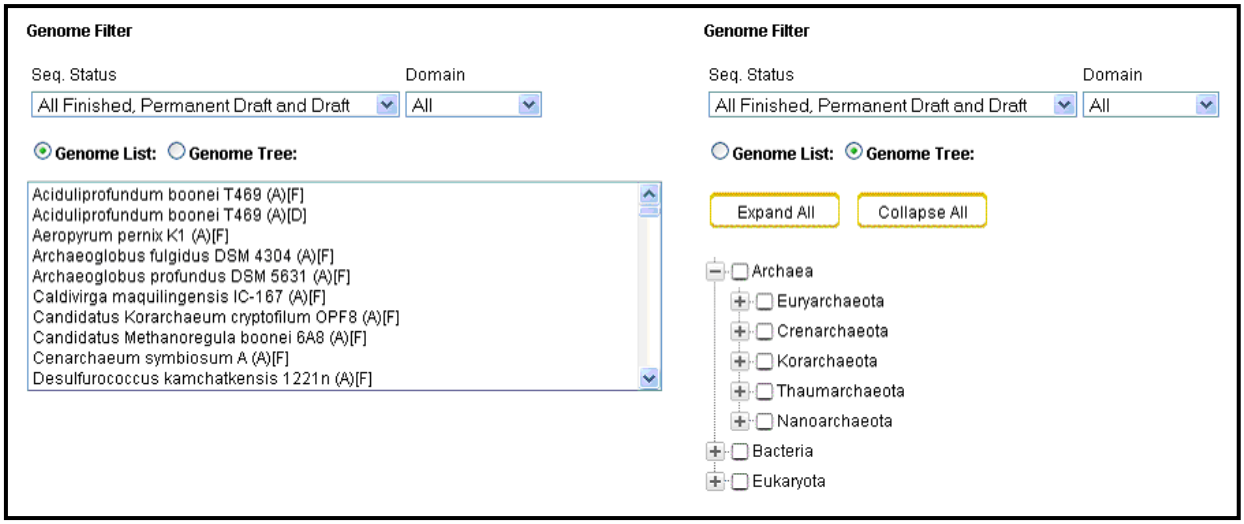


Figure 9. Genome Filter.

Find Function

Find Function has two new search entries: **SEED** Product Name/Subsytem Name and **SwissProt** Product Name. Both are included into "**All Function names**" search. In addition, **KEGG Orthology ID**, **KEGG Orthology Name** and **KEGG Orthology Definition** also made way into the list of "All Function name" search.

Many queries in Find Function have been tuned for performance. In general, "All Function name" search now takes 40% less time than before. The most dramatic improvement comes from the search of "IMG Term and Synonyms". An "ase" search with all genome can be accomplished within 1 min, before it took 8 min.

Generate GenBank File

The **Generate GenBank File** tool resides in the pages of Taxon Details, metagenome Bin Details and Scaffold Cart. Now when user selects over 100 scaffolds, the program will ask user to input his/her email address. Therefore, the program can generate the GenBank file at the background and sent the results back to the user through an email. The email contains a link, which allows the user to download the generated .zip file (**Error! Reference source not found.**). User needs to unzip the file to view its .gbk or .embl content.

By taking this strategy, not only user can continue his/her work without the waiting time, but also the program can perform works that cost more than 5 min connection time.

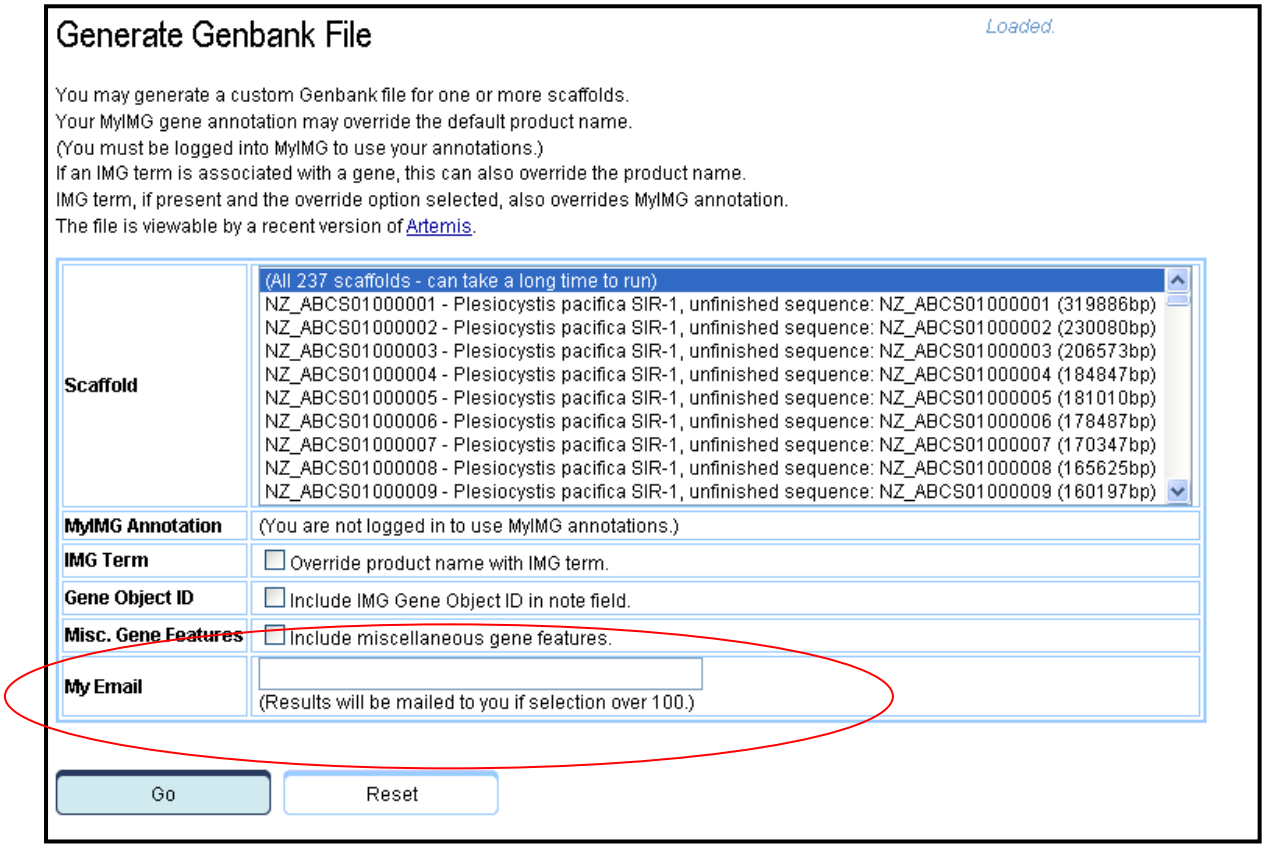


Figure 10. Generate Genbank File with submit request for large genomes.

Table Displays in the User Interface

Table Display: Search / Filter Rows

Rows in tables can now be filtered by a search term. The search term can be a part of a word, a word, a phrase, multiple words or a combination of these.

1. Select Search column from the dropdown list.
2. Enter Search term (see Figure).

Note: Using the search feature when rows are selected (checkboxes checked) is more time intensive.

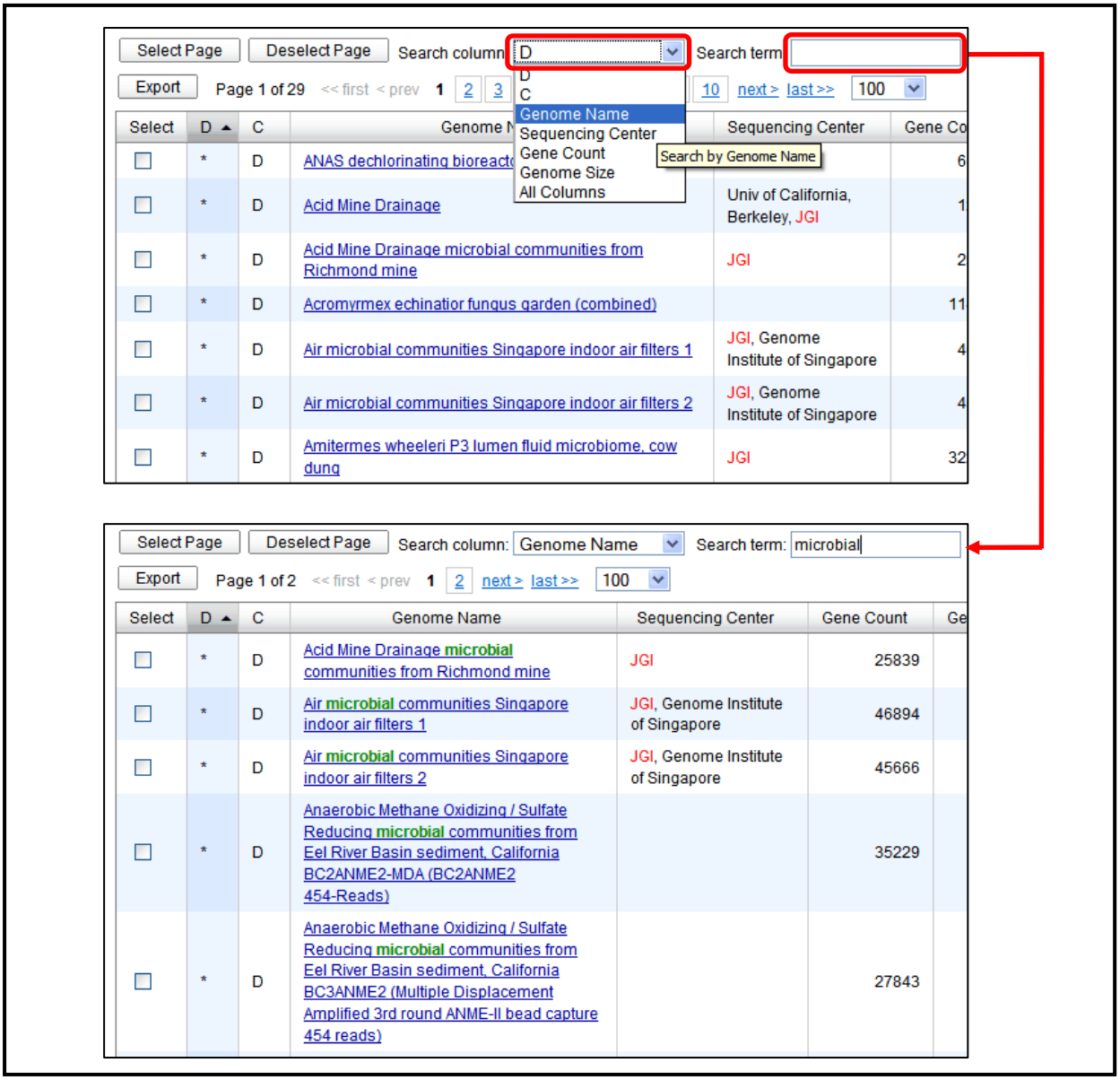


Figure 11. Filter table rows: before filter & after filter.

Table Display: Select / Deselect Page

Rows in tables can now be selected or deselected one page at a time.

- Click the "Select Page" button to select all rows on the current page (Figure).
- Click the "Deselect Page" button to deselect all rows on the current page (Figure).

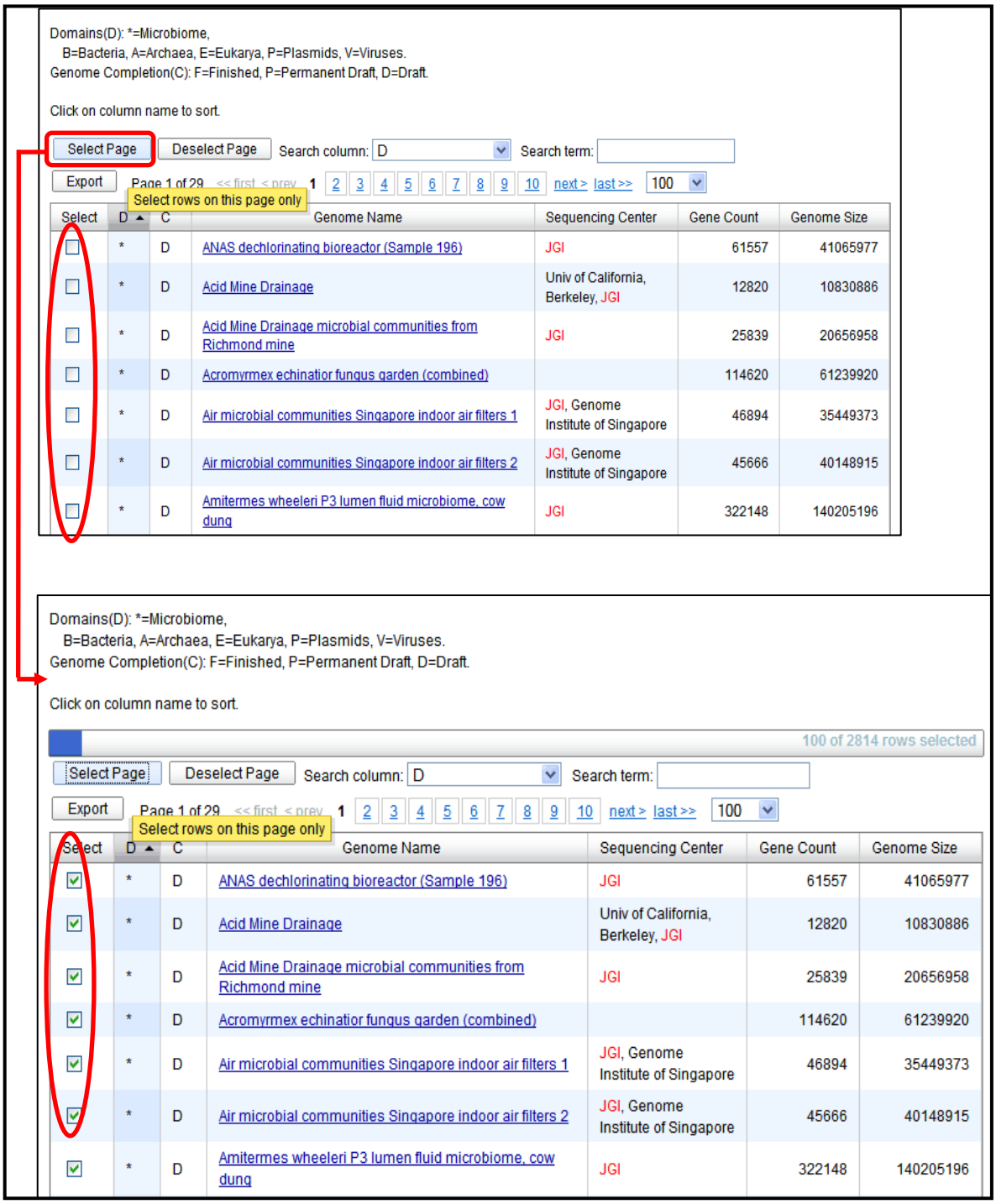


Figure 12. Select page: before selecting page & after selecting page.

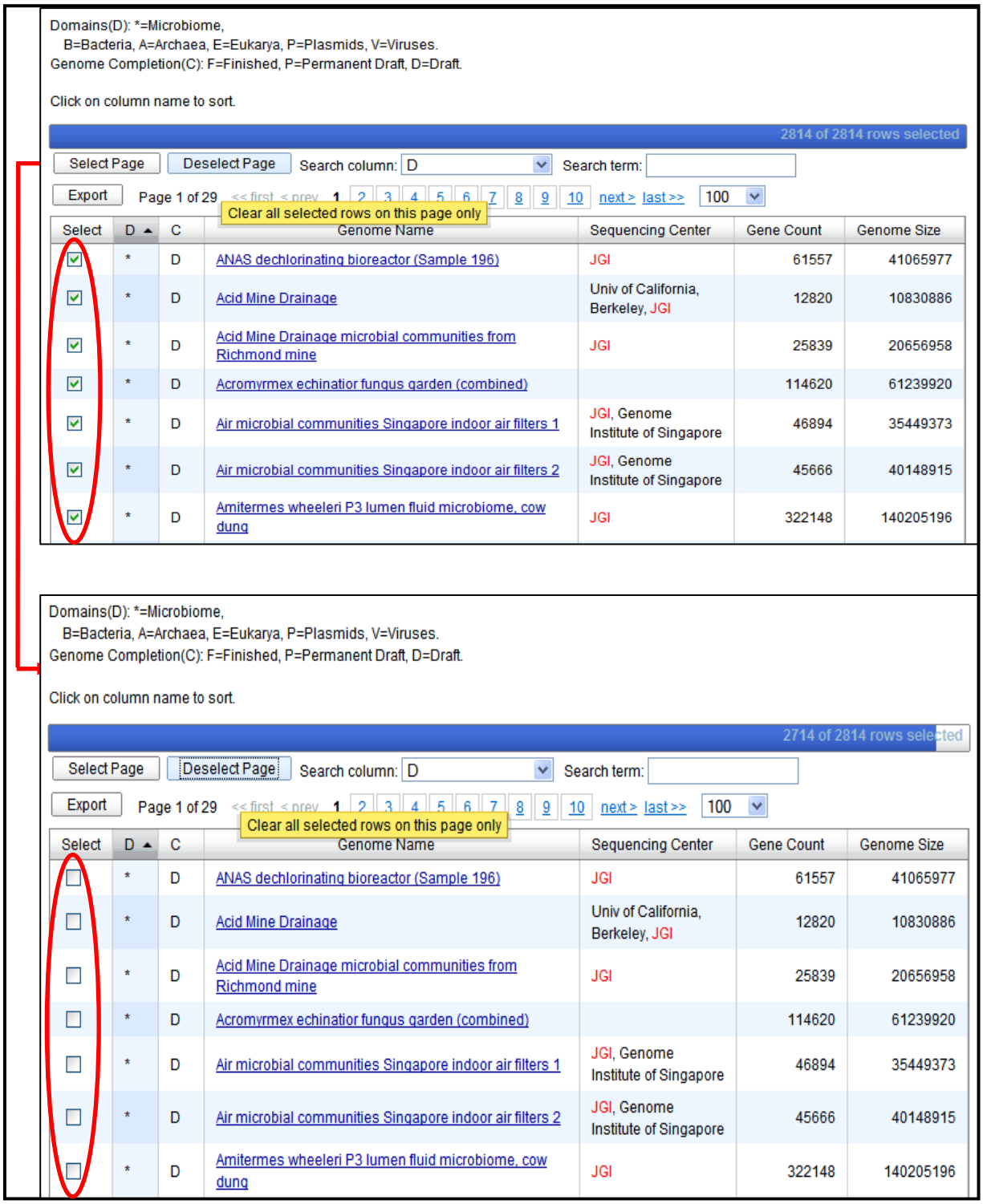


Figure 13. Deselect page: before deselecting page & after deselecting page.

Table Display: Show/Hide Column

Columns in tables can now be shown or hidden.

1. Click on the Column Selector tab at the top left hand corner of the table. (Figure)
2. Show or Hide columns by clicking the corresponding button in the dialog box.
3. Click on Close to dismiss the dialog box and return to the table.

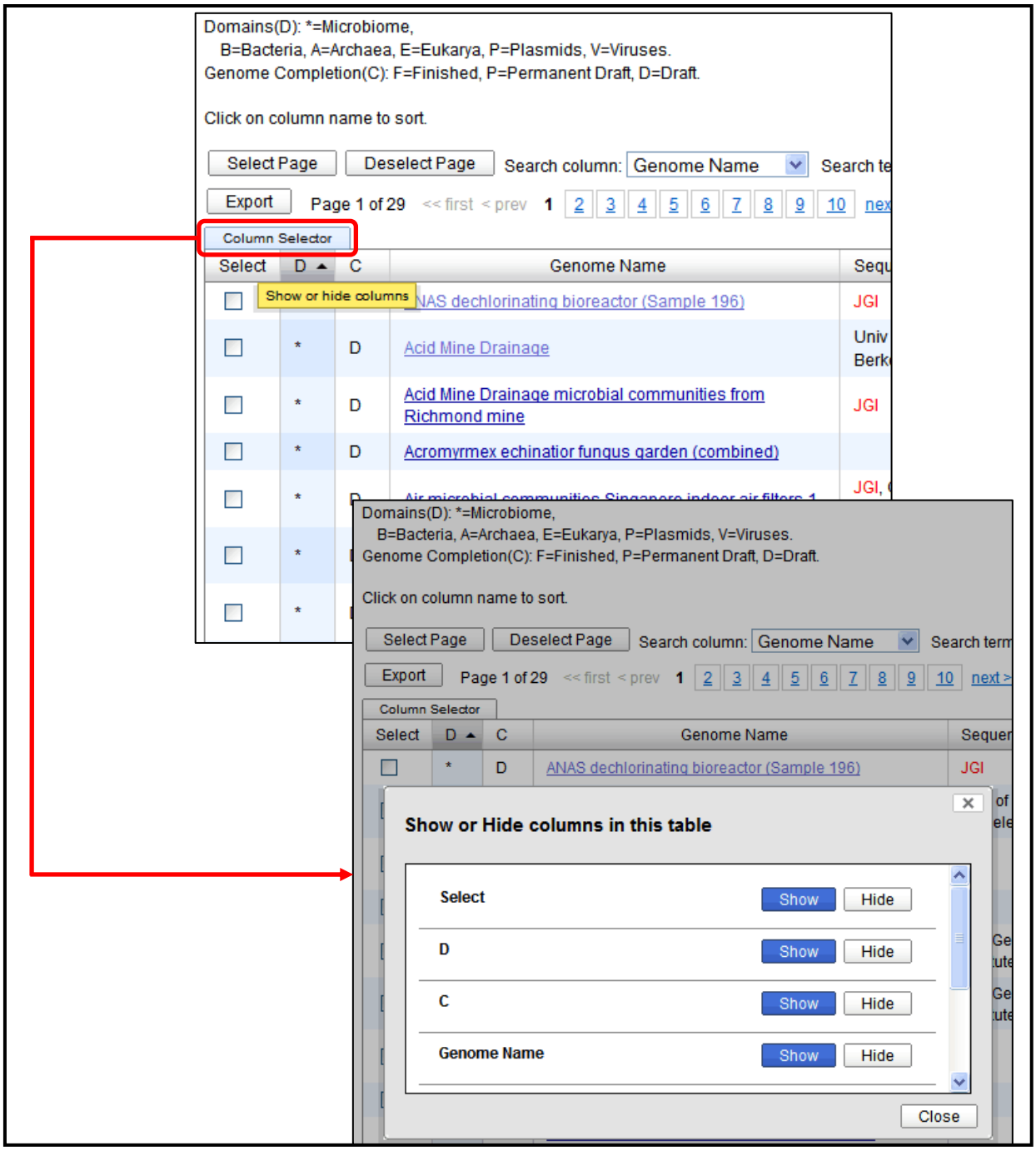


Figure 14. Using the Column Selector to show or hide columns.

Gene/Function/Scaffold Carts

Gene Cart

Genes selected in **Gene Cart** can be **added** to **Function Cart**, as illustrated in Figure 11. First, select genes from the list of genes in **Gene Cart**. Then, select a function type (COG, Pfam TIGRfam, EC Numbers, MetaCyc, KEGG KO, IMG Pathways, IMG Parts List, and IMG Terms), and add the genes to Function Cart where they will be associated with the functions corresponding to the selected function type.

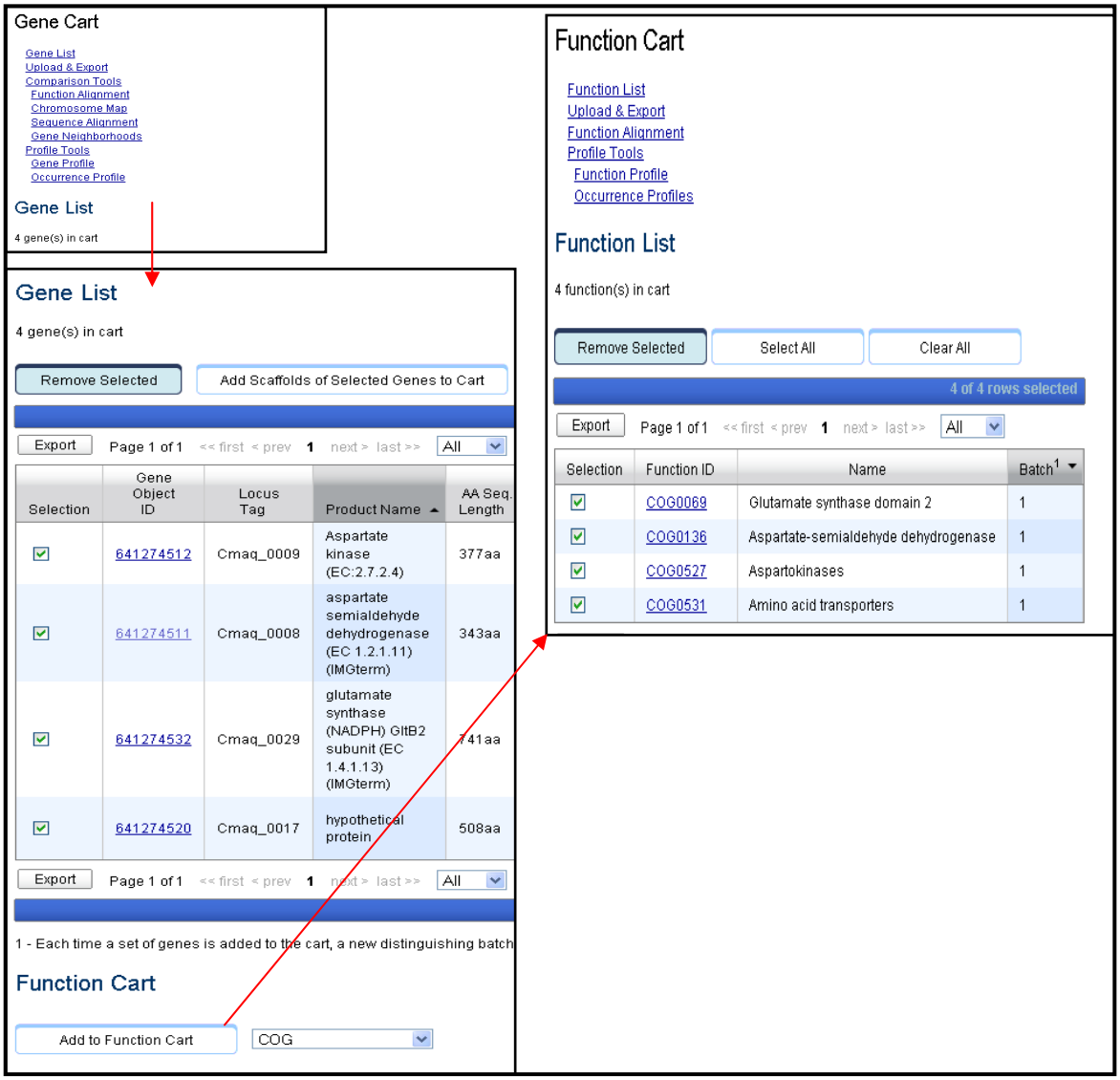


Figure 11. Add **Gene Cart** genes to **Function Cart** associated with functions of a selected type.

A new **Gene Neighborhoods** tool in **Gene Cart** allows displaying gene neighborhood plots with the genes in the Gene Cart underlined with red lines, as illustrated in Figure 12.

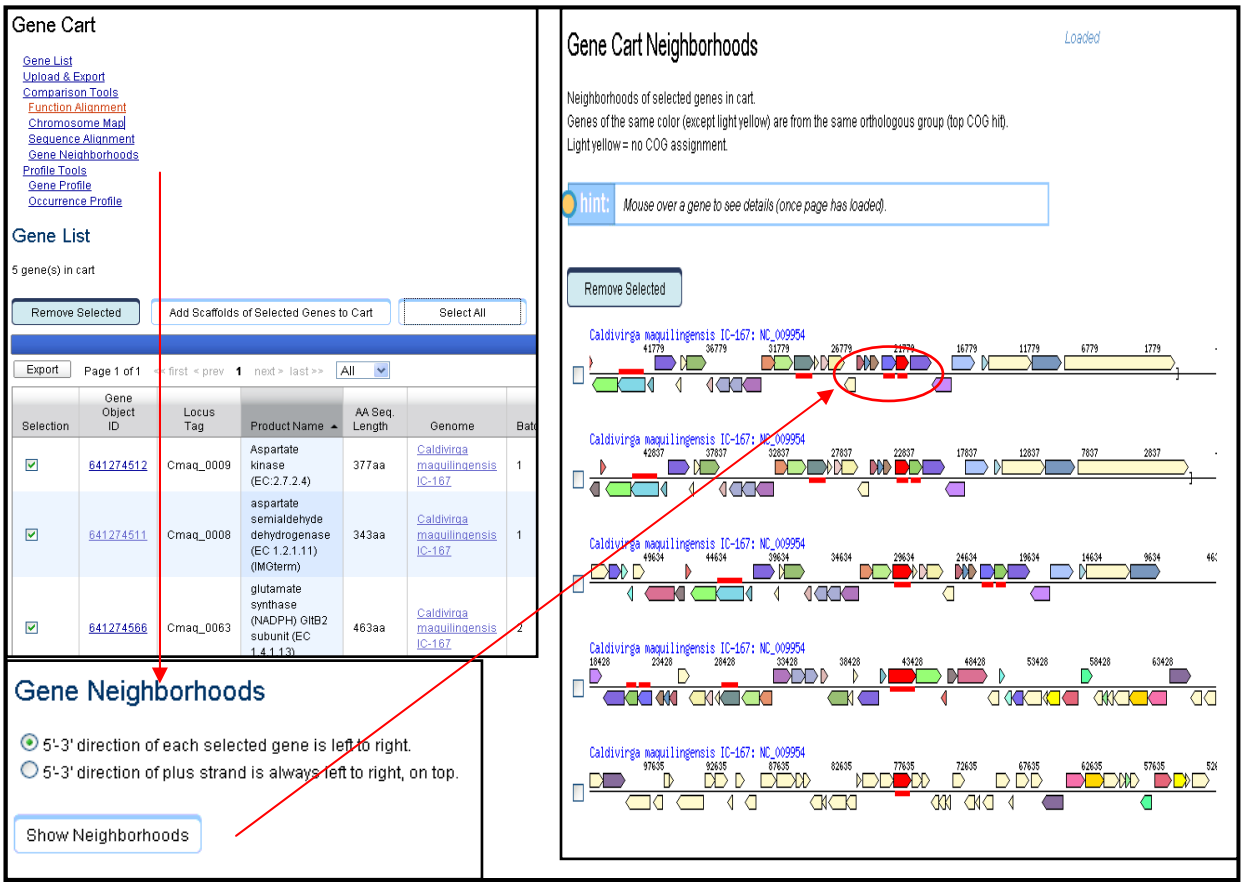


Figure 12. Gene Cart Neighborhood underlines, in red, genes found in **Gene Cart.**

Genes selected in **Gene Cart** can be **toggled**, that is, their selection states can be reverted, as illustrated in

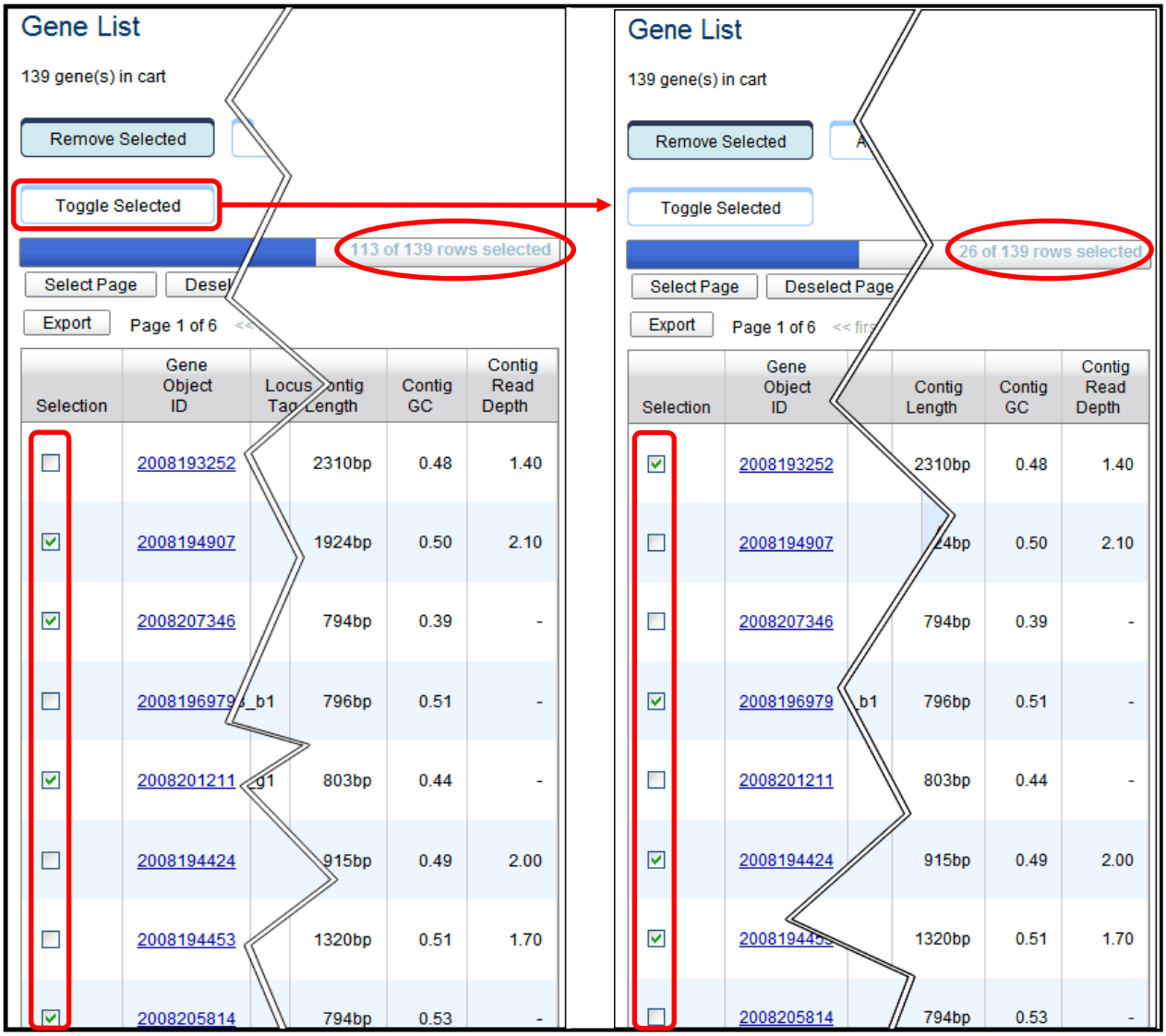


Figure 13.

Figure 13. **Toggle** genes selected in **Gene Cart**: before toggle & after toggle.

Function Cart

Genes that are associated with functions selected from the **Function List** in **Function Cart**, can be added to **Gene Cart**, as illustrated in Figure 18.

First, select functions from the **Function List**. Next, select genomes from the list of genomes provided in the **Function Profile** section of **Function Cart**. Click **Add to Gene Cart** to add all the genes associated with selected functions and genomes to **Gene Cart**.

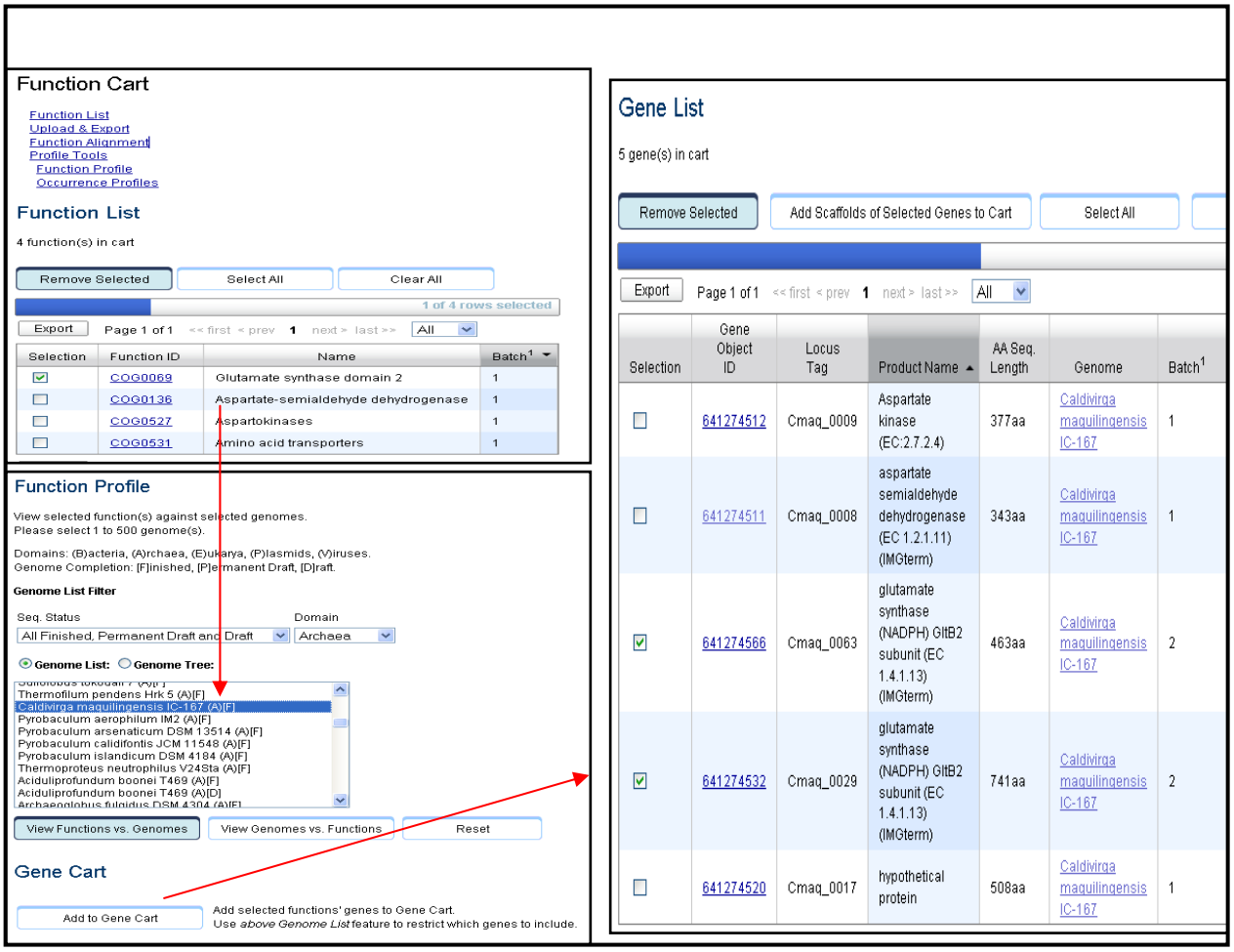


Figure 14. Genes with selected functions & genomes in **Function Cart** are added to **Gene Cart.**

Functions selected in **Function Cart** can be toggled, similar to the genes in Gene Cart, as illustrated in Figure 17 above.

Scaffold Cart

Scaffolds selected in **Scaffold Cart** can be toggled, that is, their selection states can be reverted, as illustrated in Figure 19.

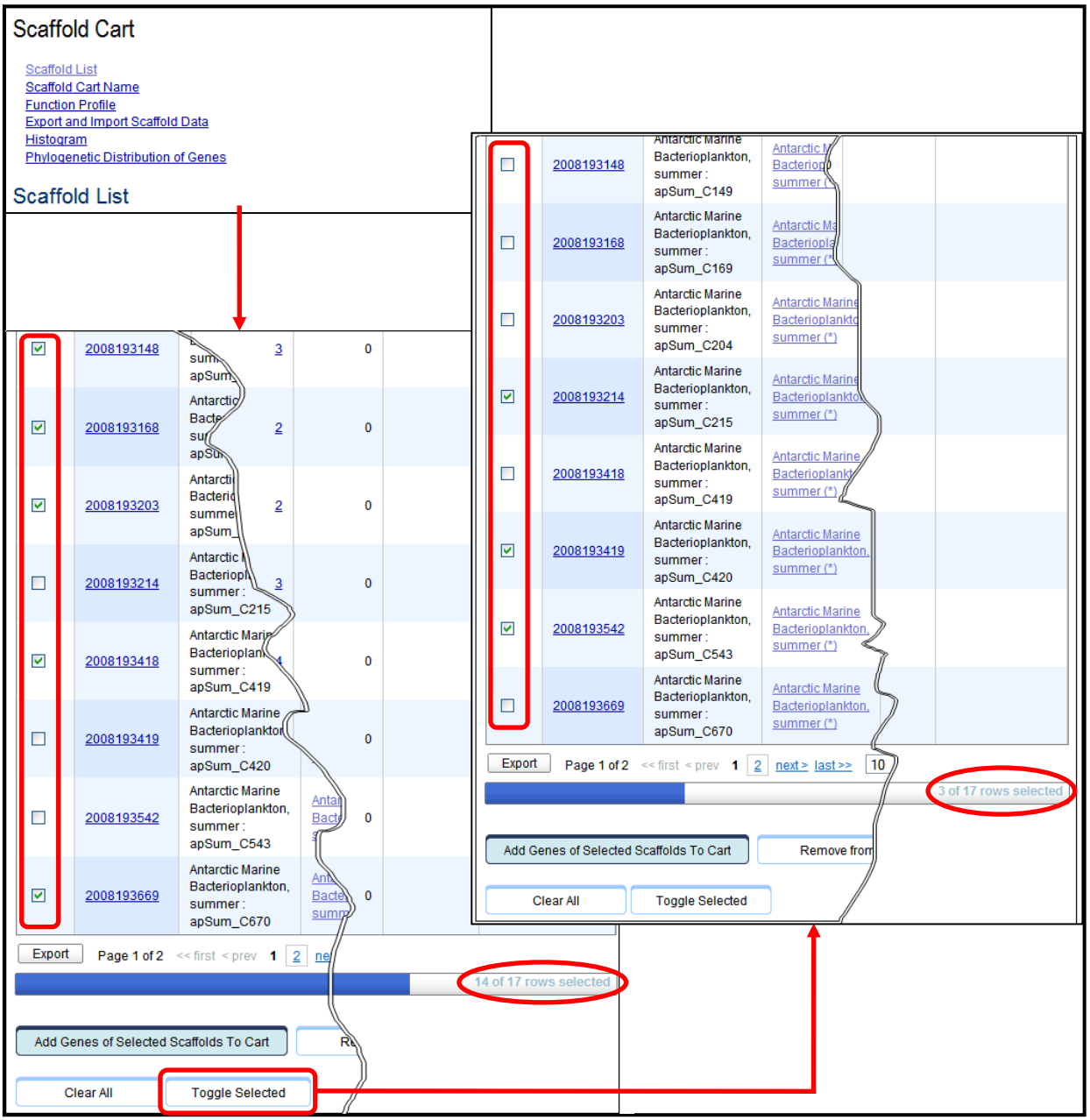


Figure 19. Toggle selected scaffolds in **Scaffold Cart:** before toggle & after toggle.

Chromosome Viewer

The **Chromosome Viewer** has been extended with additional coloring by KEGG category, Pfam category, and TIGRfam roles, as illustrated in Figure .

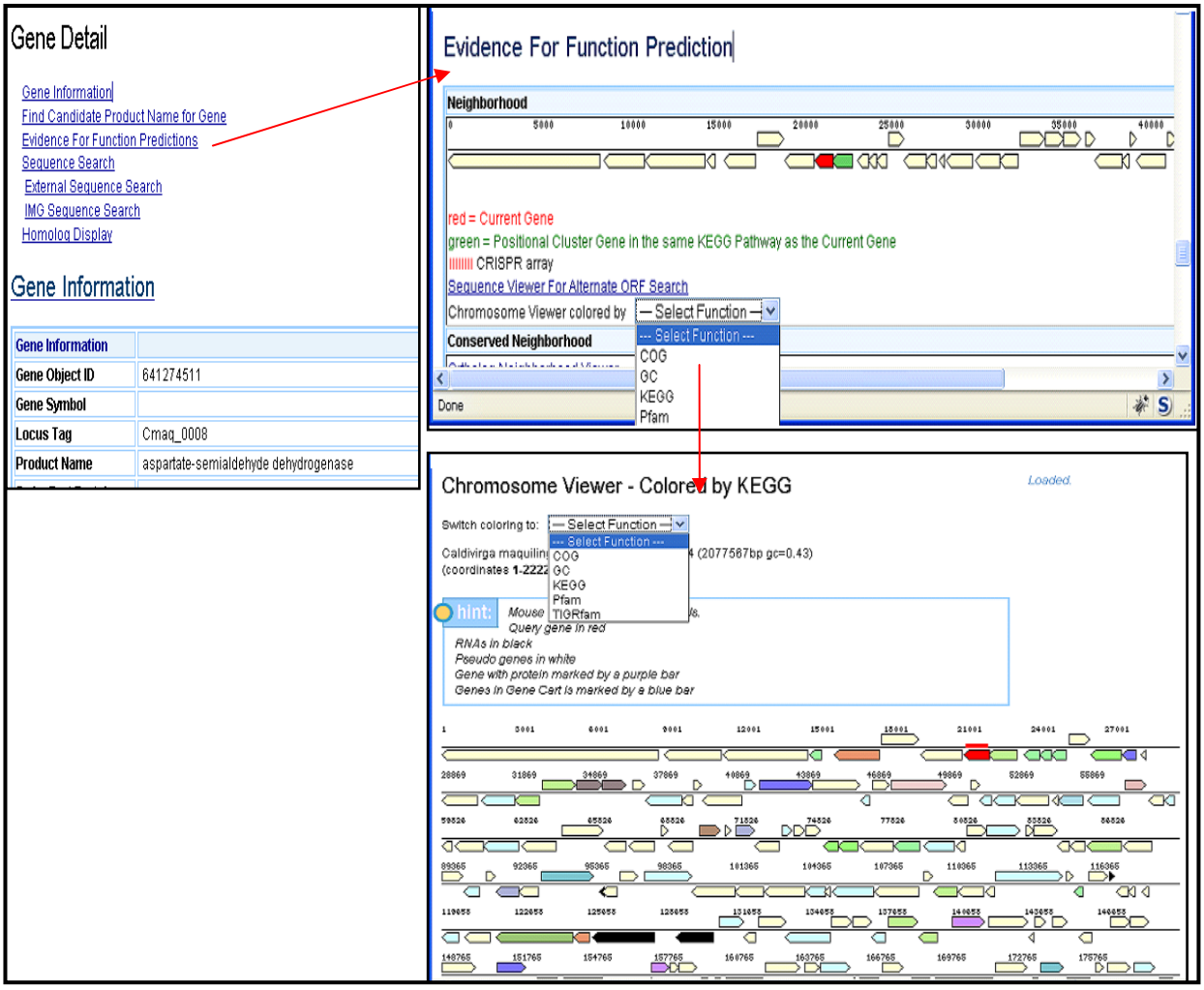


Figure 20. Chromosome Viewer coloring by KEGG category, Pfam category, TIGRfam roles.

Protein Expression Studies

Protein Expression Sample "Describe"

For each study, the experiments (samples) can be selected not just for **clustering**, but also for a description of functionality, as illustrated in Figure 5. The user first selects some samples to compare and then clicks on the "describe" button.

Describe is used to compare samples with respect to gene coverage and to display information and links to cassettes, associated cog functions, and KEGG pathways (Figure 6).

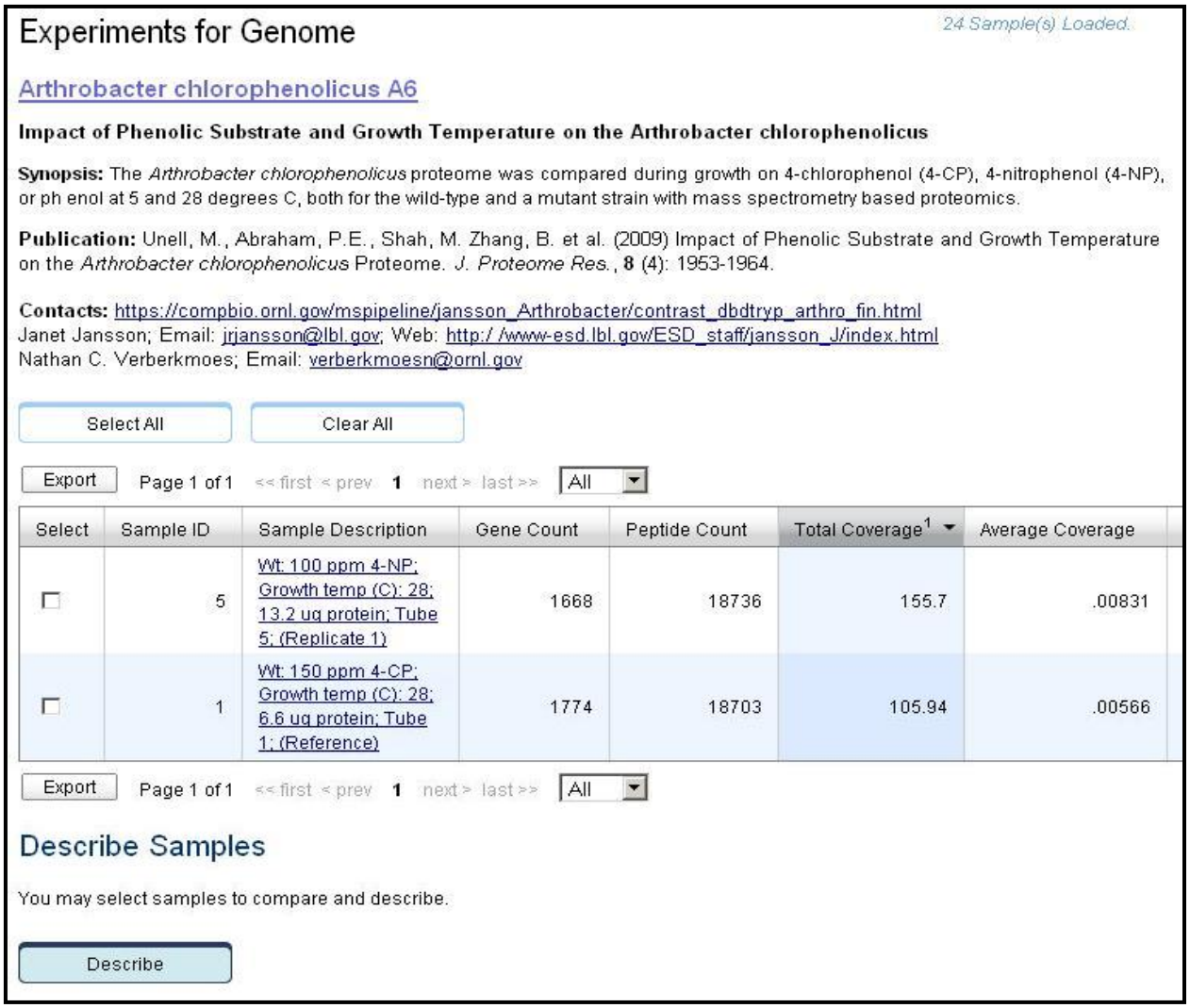


Figure 5 Describe Samples

Expression Data for Selected Samples

2182 Genes Loaded.

3 samples selected (5,6,15)

Add Selections To Gene Cart | Select All | Clear All

Export | Page 1 of 22 << first < prev 1 2 3 4 5 6 7 8 9 10 next > last >> | 100

Select	Gene ID	Locus Tag	Product Name	5	6	15	Cassette ID	COG function	KEGG pathway
☐	643589402	Achl_0997	adenosine deaminase	0.0538244	0.0566572	0.0254958	57643348670	F - Nucleotide transport and metabolism	Primary immunodeficiency Purine metabolism
☐	643589623	Achl_1218	adenosine deaminase	0.1031746	0.0714286	0.0687831	74643348670	F - Nucleotide transport and metabolism	Primary immunodeficiency Purine metabolism
☐	643589620	Achl_1215	pyrimidine-nucleoside phosphorylase	0.0767494	0.0902935	0.1128668	74643348670	F - Nucleotide transport and metabolism	Bladder cancer Drug metabolism - other enzymes Pyrimidine metabolism
☐	643590232	Achl_1827	Superoxide dismutase	0.1971154	0.3173077	0.1538462	11064334867(	P - Inorganic ion transport and metabolism	Huntington's disease

Figure 6. Describe functionality.

The KEGG pathway link from the sample "describe" page, displays the pathway either for multiple samples or for a single sample, depending on what the user selected. If a single sample was selected, the pathway will be colored by gene expression (coverage) for that sample (Figure). If multiple samples were selected, the pathway will be colored by percentage of samples in which a given gene is expressed (Figure).

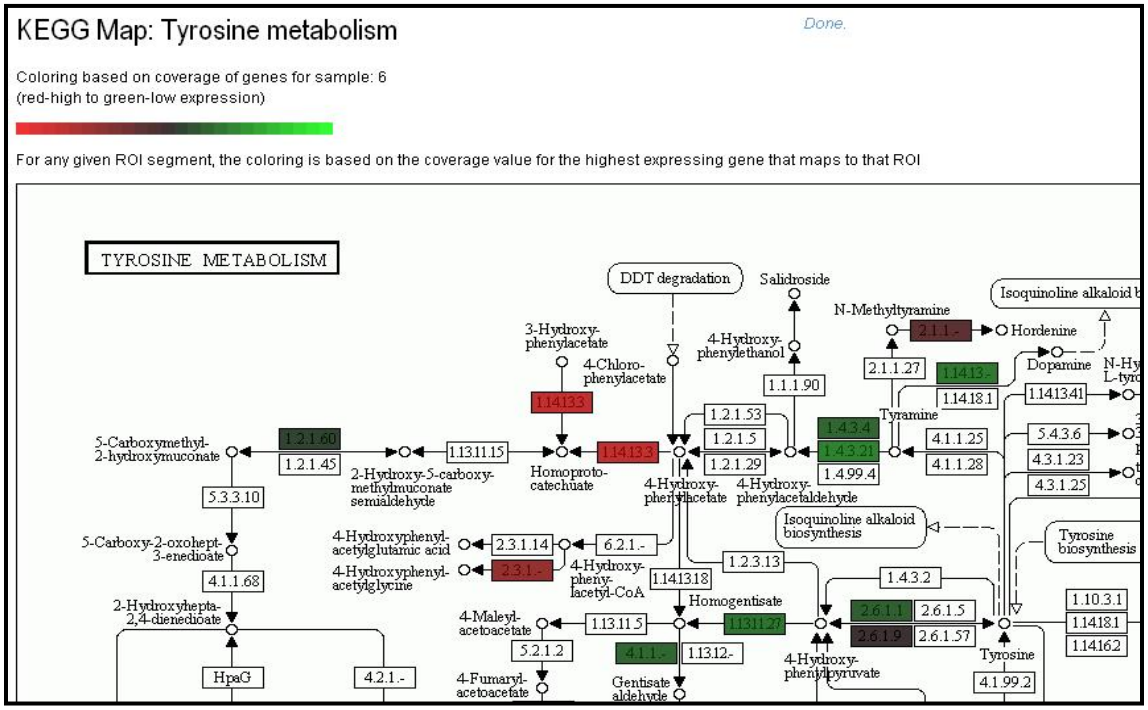


Figure 23. KEGG Map colored by gene expression in single sample.

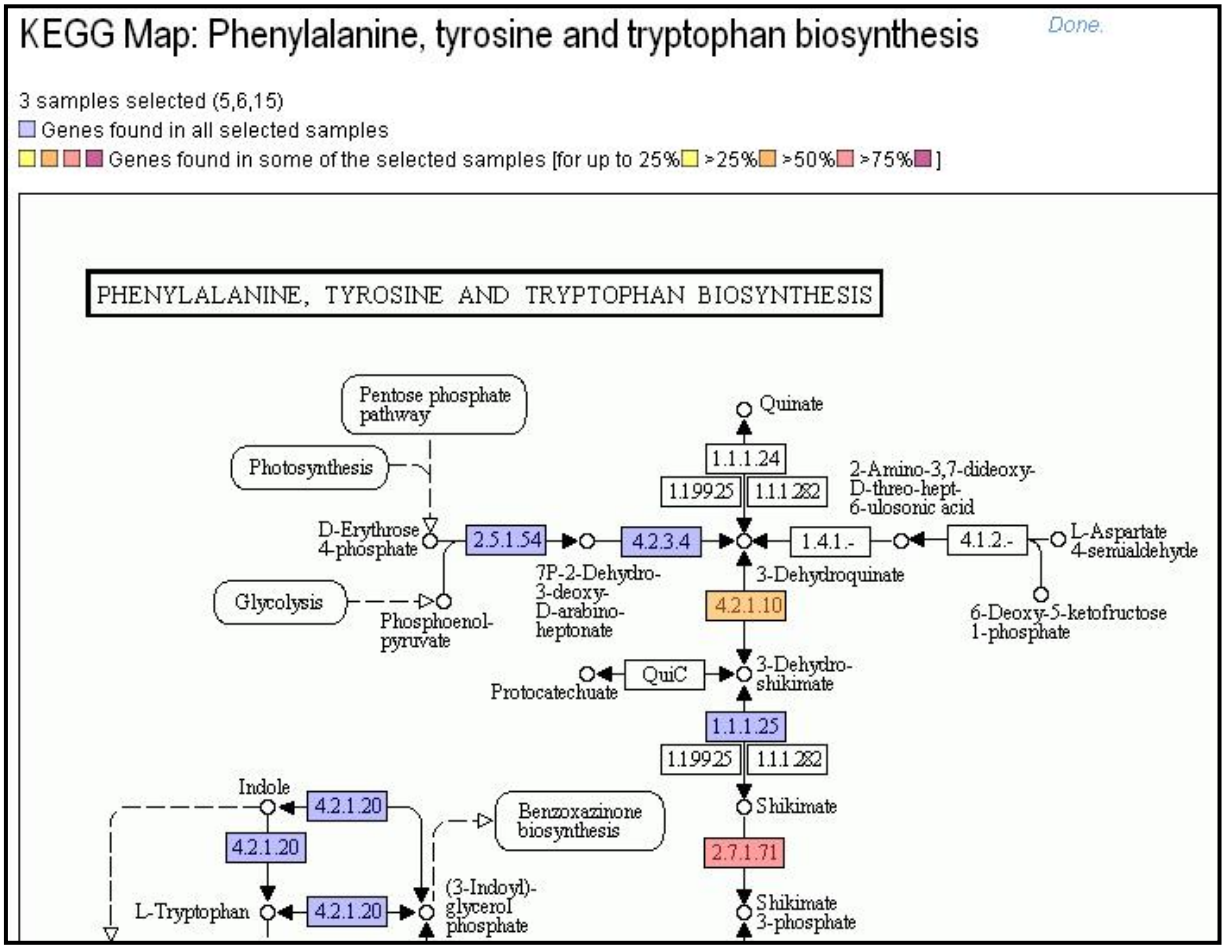


Figure 24. KEGG Map colored by percentage of samples in which a given gene is expressed.

Each region of interest on the kegg map links to a list of genes for the selected samples that fall in that region (Figure 7).

Genes with KO: K01695,K01696 found among selected samples

Add Selected to Gene Cart | Select All | Clear All

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

Select	Gene ID	Locus Tag	Gene Name	Sample ID	Sample Description	KO
☐	643590098	Achl_1683	tryptophan synthase, beta subunit	5	Wt: 100 ppm 4-NP; Growth temp (C): 28; 13.2 ug protein; Tube 5; (Replicate 1)	K01696
☐	643590098	Achl_1683	tryptophan synthase, beta subunit	6	Wt: 100 ppm 4-NP; Growth temp (C): 28; 0.7 ug protein; Tube 6; (Replicate 2)	K01696
☐	643590098	Achl_1683	tryptophan synthase, beta subunit	15	Wt: 100 ppm 4-NP; Growth temp (C): 5; 5.0 ug protein; Tube 15; (Replicate 2)	K01696
☐	643590099	Achl_1684	tryptophan synthase, alpha subunit	5	Wt: 100 ppm 4-NP; Growth temp (C): 28; 13.2 ug protein; Tube 5; (Replicate 1)	K01695
☐	643590099	Achl_1684	tryptophan synthase, alpha subunit	6	Wt: 100 ppm 4-NP; Growth temp (C): 28; 0.7 ug protein; Tube 6; (Replicate 2)	K01695
☐	643590099	Achl_1684	tryptophan synthase, alpha subunit	15	Wt: 100 ppm 4-NP; Growth temp (C): 5; 5.0 ug protein; Tube 15; (Replicate 2)	K01695

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

Add Selected to Gene Cart | Select All | Clear All

Figure 7. List of genes for selected samples that fall on a given region of interest.

Protein Samples: Compare Samples using COG statistics

For each genome that has associated protein experiment samples, the COG statistics that are usually displayed per individual genome (see Figure 8), can now be displayed per samples in that genomes, as illustrated in Figure .

Statistics by COG Categories

Show Cogs in Protein Samples

D	C	Genome Name	Total COG Genes	Amino acid transport and metabolism	Amino acid transport and metabolism percentage	Carbohydrate transport and metabolism	Carbohydrate transport and metabolism percentage	RNA processing and modification	RNA processing and modification percentage
B	F	Arthrobacter chlorophenolicus A6	3124	357	11.43%	394	12.61%	2	0.06%

Figure 8. COG statistics displayed per individual genome.

Statistics by COG Categories

Hide Cogs in Protein Samples

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

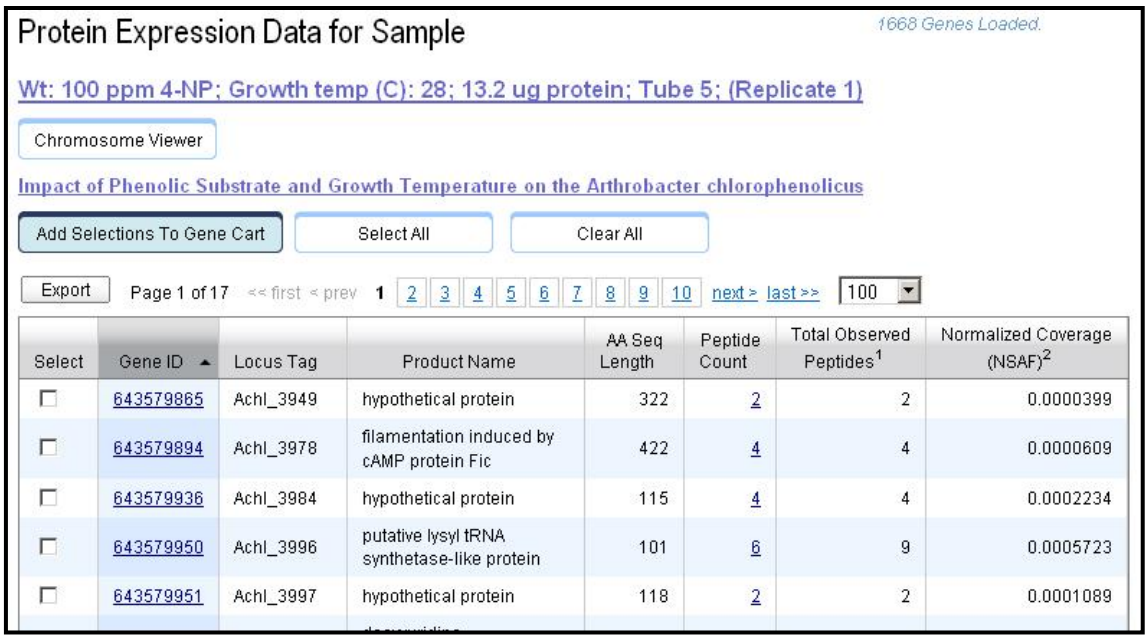
Sample ID	Sample Name		%		%		%
1	Wt: 150 ppm 4-CP; Growth temp (C): 28; 6.6 ug protein; Tube 1; (Reference)	201	6.16%				
2	Wt: 150 ppm 4-CP; Growth temp (C): 28; 6.9 ug protein; Tube 2; (Replicate 1)	212	6.50%	142	4.35%	2	0.06%
3	Wt: 150 ppm 4-CP; Growth temp (C): 28; 4.9 ug protein; Tube 3; (Replicate 2)	188	5.76%	146	4.47%	1	0.03%
4	Wt: 100 ppm 4-NP; Growth temp (C): 28; 11.9 ug protein; Tube 4; (Reference)	199	6.10%	137	4.20%	2	0.06%

Carbohydrate transport and metabolism

Figure 27. COG statistics displayed per samples in a genom.

Protein Sample: Chromosome Viewer

For a protein sample, the user can select **Chromosome Viewer**, as illustrated in Figure , which will then display genes colored by their expression level, as illustrated in Figure 29.



Select	Gene ID	Locus Tag	Product Name	AA Seq Length	Peptide Count	Total Observed Peptides[1]	Normalized Coverage (NSAF)[2]
☐	643579865	Achl_3949	hypothetical protein	322	2	2	0.0000399
☐	643579894	Achl_3978	filamentation induced by cAMP protein Fic	422	4	4	0.0000609
☐	643579936	Achl_3984	hypothetical protein	115	4	4	0.0002234
☐	643579950	Achl_3996	putative lysyl tRNA synthetase-like protein	101	6	9	0.0005723
☐	643579951	Achl_3997	hypothetical protein	118	2	2	0.0001089

Figure 28. Chromosome Viewer link on Protein Sample page.

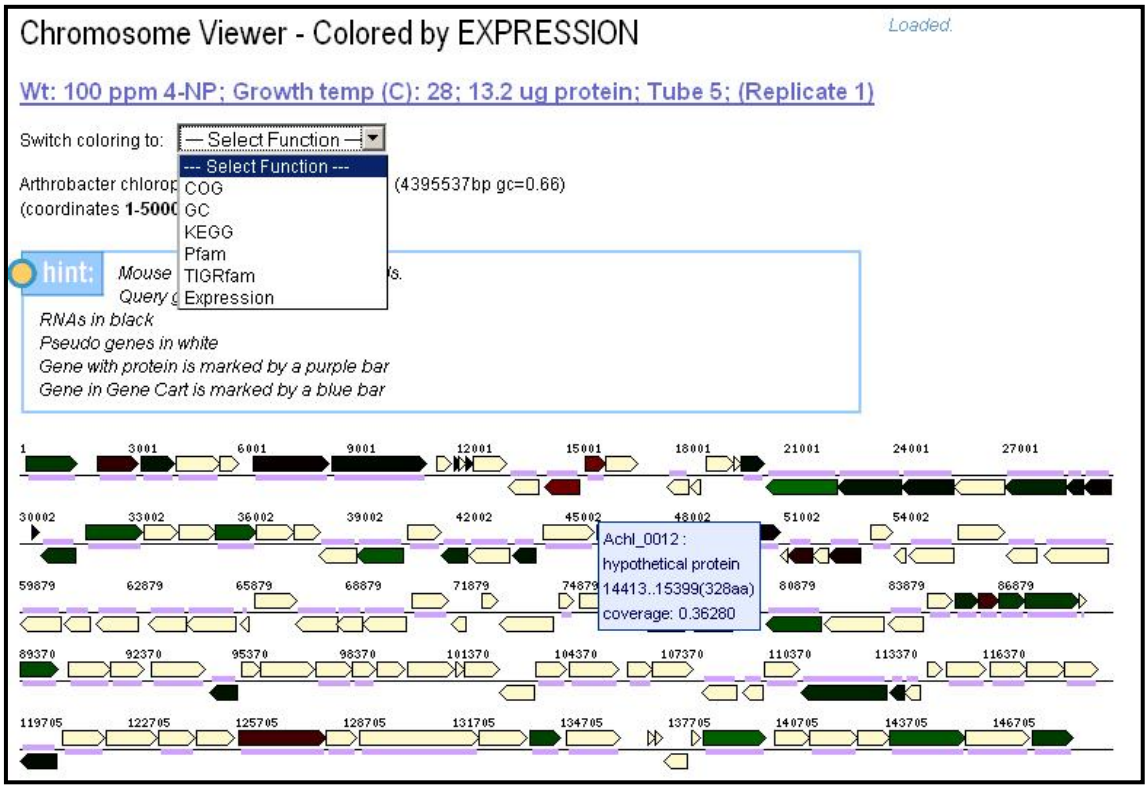


Figure 9 Chromosome Viewer colored by gene expression in a given sample

Gene Detail - Chromosome Viewer

In the **Gene Detail** page for a gene, genes associated with protein expression data are displayed by the **Chromosome Viewer** underlined by purple lines, as illustrated in Figure 30.

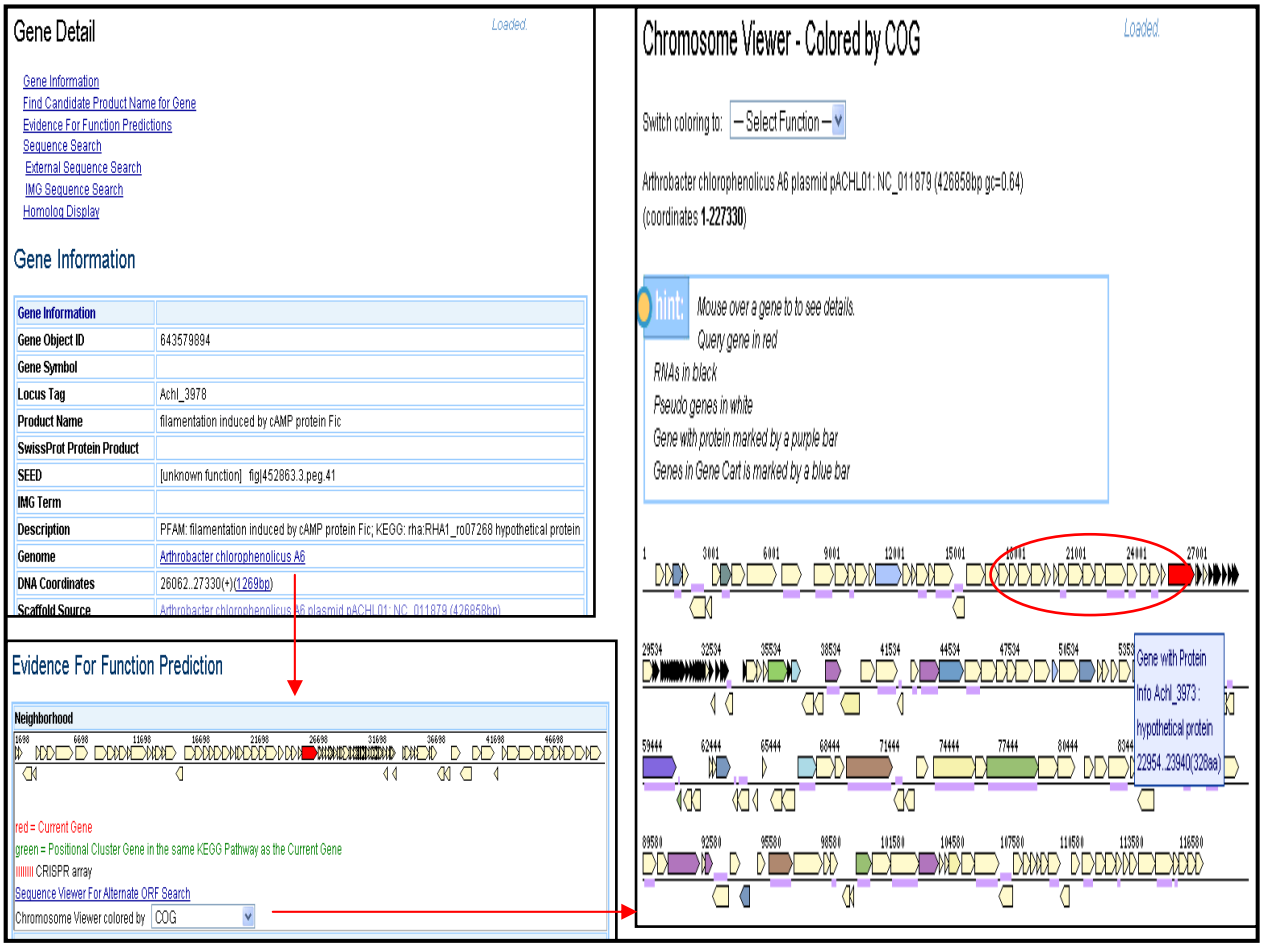


Figure 30. Chromosome Viewer: genes with protein information underlined in purple.