

Protein Expression Studies IMG 2.9

Introduction

IMG 2.9 contained **protein expression** data from an *Arthrobacter chlorophenolicus* study. IMG 3.0 contains data from two additional protein expression studies. These data can be explored using the tools illustrated in Figures 1 and 2.

The screenshot illustrates a web-based interface for exploring protein expression data. It is divided into five main sections:

- Organism Details (i):** Shows 'Organism Information' with 'Arthrobacter chlorophenolicus A6' and a 'List of Studies' link.
- Studies for Genome (ii):** A table listing studies. The study 'Impact of Phenolic Substrate and Growth Temperature on the Arthrobacter chlorophenolicus' is highlighted.
- Experiments for Genome (iii):** A table listing samples. Sample 6, 'Wt: 100 ppm 4-NP; Growth temp (C): 28; 0.7 ug protein; Tube 6; (Replicate 2)', is highlighted.
- Protein Expression Data for Sample (iv):** A table showing data for sample 6. The gene '643579969' (Achl_4015) is highlighted.
- Peptides for Gene and Sample (v):** Shows the protein sequence and a list of peptides. Peptide 79913 (IAQFVIQR) is highlighted.

Red arrows indicate the navigation path: from 'List of Studies' to the study table, then to the sample table, then to the protein expression data table, and finally to the peptide details page.

Figure 1. Organism Details - Protein Expression Studies.

The **Organism Details** for a genome associated with protein expression studies provides two starting points for exploring protein expression data. First, a **Protein Expression Studies** link on this page, illustrated in Figure 1(i), leads to the list of studies for the genome of interest, as illustrated in Figure 1(ii). Each protein expression study is displayed together with the number of expressed genes and observed peptides. The **Study Name** provides a link to the list of experiments/samples for the study, as illustrated in Figure 1(iii). The description for each sample consists of the experimental conditions and provides a link to the protein expression data for the sample organized per expressed gene, as illustrated in Figure 1(iv). For each expressed gene, the number of observed peptides and **normalized coverage** (NSAF) are provided. **NSAF** is the coverage for an expressed gene in an experiment divided by the total coverage of all the genes in that experiment, whereby the **coverage** for a gene is defined as the sum of the sizes of all observed peptides for the gene divided by the size of the gene¹. The number of observed peptides leads to the peptide details page, illustrated in Figure 1(v), where the peptide sequences are displayed aligned on the gene's protein sequence.

¹ Florens, L., Carozza, M.J., Swanson, S.K. et al. (2006) Analyzing chromatin remodeling complexes using shotgun proteomics and normalized spectral abundance factors, *Methods* **40**(4): 303-311.

The **Organism Details** provides another starting point for exploring protein expression data through the number of expressed genes in the Genome Statistics section, as illustrated in Figure 2(i). This number leads to the list of expressed genes, with each gene linked to its **Gene Detail** page, as illustrated in Figure 2(ii).

Organism Information (i)

Organism Name	Arthrobacter chlorophenolicus A6
Taxon Object ID	643348509

Genome Statistics

Genes with Proteomic data: 2853

Gene Detail (ii)

Gene Information

Gene Object ID	643579856
Gene Symbol	
Locus Tag	Achl_3940
Product Name	GatB/YqeY do

Protein Information

Amino Acid Sequence Length	153aa
COG	COG1610 - U
Families	- IPR003789 - IPR019004
Transmembrane Helices	No
Signal Peptide	No
Statistics	peptide
Expression	protein
Pfam	pfam09424 - Y

Proteomic Data for Gene (iii)

>Gene: 643579856 YP_002477708 GatB/YqeY domain protein
[Arthrobacter chlorophenolicus A6 plasmid pACHL01: NC_011879]

MTLKERLKED VVAHMKAGHK TALTTVNRNL GEISTKEKAG KTFIELDVQ
VTSLQKEAA KRRDTARIYT EAGQGDRAAA EITEAEIIEA YLPKALIRDE
VEVIVDEAIN ALKADGQELS MRSIGAVMKP VTAKVAGRFD GKTVSEIVRG
RLS

Sample ID	Description	Peptide Count	Study
10	Wt: 150 ppm 4-CP; Growth temp (C): 5; 8.4 ug protein; Tube 10; (Reference)	2	Impact of Phenolic Substrate and Growth Temperature on the Arthrobacter chlorophenolicus
13	Wt: 100 ppm 4-NP; Growth temp (C): 5; 3.6 ug protein; Tube 13; (Reference)	2	Impact of Phenolic Substrate and Growth Temperature on the Arthrobacter chlorophenolicus
14	Wt: 100 ppm 4-NP; Growth temp (C): 5; 4.8 ug protein; Tube 14; (Replicate 1)	2	Impact of Phenolic Substrate and Growth Temperature on the Arthrobacter chlorophenolicus
24	Mutant T99- 400 ppm phenol; Growth temp (C): 5; 6.5 ug protein; Tube 24; (Replicate 2)	2	Impact of Phenolic Substrate and Growth Temperature on the Arthrobacter chlorophenolicus

Peptides for Gene and Sample (iv)

Wt: 150 ppm 4-CP; Growth temp (C): 5; 8.4 ug protein; Tube 10; (Reference)

>Gene: 643579856 YP_002477708 GatB/YqeY domain protein
[Arthrobacter chlorophenolicus A6 plasmid pACHL01: NC_011879]

MTLKERLKED VVAHMKAGHK TALTTVNRNL GEISTKEKAG KTFIELDVQ
VTSLQKEAA KRRDTARIYT EAGQGDRAAA EITEAEIIEA YLPKALIRDE
VEVIVDEAIN ALKADGQELS MRSIGAVMKP VTAKVAGRFD GKTVSEIVRG
RLS

Peptide ID	Peptide Seq
138475	DEVEVIVDEAINALKADGQELSMR
138476	NVLGEISTKEKAGKTFIELDVQVTSLQK

Figure 2. Gene Details - Protein Expression Data.

The **Protein Information** section of the **Gene Detail** for an expressed gene provides a link to a **Proteomic Data**, as illustrated in Figure 2(iii), which displays the list of experiments/samples involving the expressed gene. Each sample is associated with a description consisting of the experimental conditions and the number of peptides observed for the expressed gene as part of this experiment. The number of peptides leads to a page displaying peptide details page, illustrated in Figure 2(iv), where the peptide sequences are displayed aligned on the gene's protein sequence.

Sample Clustering *IMG 3.1*

For each study, the experiments (samples) can be selected for **clustering** based on coverage values for the genes expressed in each sample (Figure 3). The user can specify the clustering method and the distance measure to use, as well as the minimum number of samples in which a gene should appear in order to be included in the calculation.

Sample Clustering

You may select samples and cluster them based on the coverage of the expressed genes.
Proximity of grouping indicates the relative degree of similarity of samples to each other.

Select All

Select	Sample ID	Sample Description	Gene Count	Peptide Count	Total Coverage ¹
<input type="checkbox"/>	1	Wt: 150 ppm 4-CP; Growth temp (C): 28; 6.6 ug protein; Tube 1; (Reference)	1773	18703	105.94
<input type="checkbox"/>	2	Wt: 150 ppm 4-CP; Growth temp (C): 28; 6.9 ug protein; Tube 2; (Replicate 1)	1898	22709	133.17
<input type="checkbox"/>	3	Wt: 150 ppm 4-CP; Growth temp (C): 28; 4.9 ug protein; Tube 3; (Replicate 2)	1753	21017	120.75
<input type="checkbox"/>	4	Wt: 100 ppm 4-NP; Growth temp (C): 28; 11.9 ug protein; Tube 4; (Reference)	1905	21016	140.96
<input type="checkbox"/>	5	Wt: 100 ppm 4-NP; Growth temp (C): 28; 13.2 ug protein; Tube 5; (Replicate 1)	1666	18736	155.69
<input type="checkbox"/>	6	Wt: 100 ppm 4-NP; Growth temp (C): 28; 0.7 ug protein; Tube 6; (Replicate 2)	1523	16178	120.12
<input type="checkbox"/>	23	Mutant T99: 400 ppm phenol; Growth temp (C): 5; 12.2 ug protein; Tube 23; (Replicate 1)	1901	19050	129.03
<input type="checkbox"/>	24	Mutant T99: 400 ppm phenol; Growth temp (C): 5; 6.5 ug protein; Tube 24; (Replicate 2)	1889	20708	137.15

Clustering Method:

- Pairwise complete-linkage (default)
- Pairwise single-linkage
- Pairwise centroid-linkage
- Pairwise average-linkage

Distance Measure:

- No gene clustering
- Uncentered correlation
- Pearson correlation (default)
- Uncentered correlation, absolute value
- Pearson correlation, absolute value
- Spearman's rank correlation
- Kendall's tau
- Euclidean distance
- City-block distance

Minimum number of samples in which a gene should appear in order to be included: (default=2)

Go

Reset

FIGURE 3. Clustering of protein samples

Both samples and genes are clustered. What will be displayed in IMG are a hierarchical cluster tree of samples and a normalized heat map of coverage values for each gene for each sample (Figure 4). On the heat map, the samples link to sample pages and the genes link to gene pages. Clicking on a sample in the cluster tree **sorts** the heat map on descending coverage values for that sample.



FIGURE 4. Clustering results

In addition, the results can be viewed in **Java TreeView** applet (Figure 5). The applet is loaded with the cluster files by clicking on the “View with Java TreeView” button. Within the Java TreeView, the genes link to IMG gene pages and the samples link to IMG sample pages.

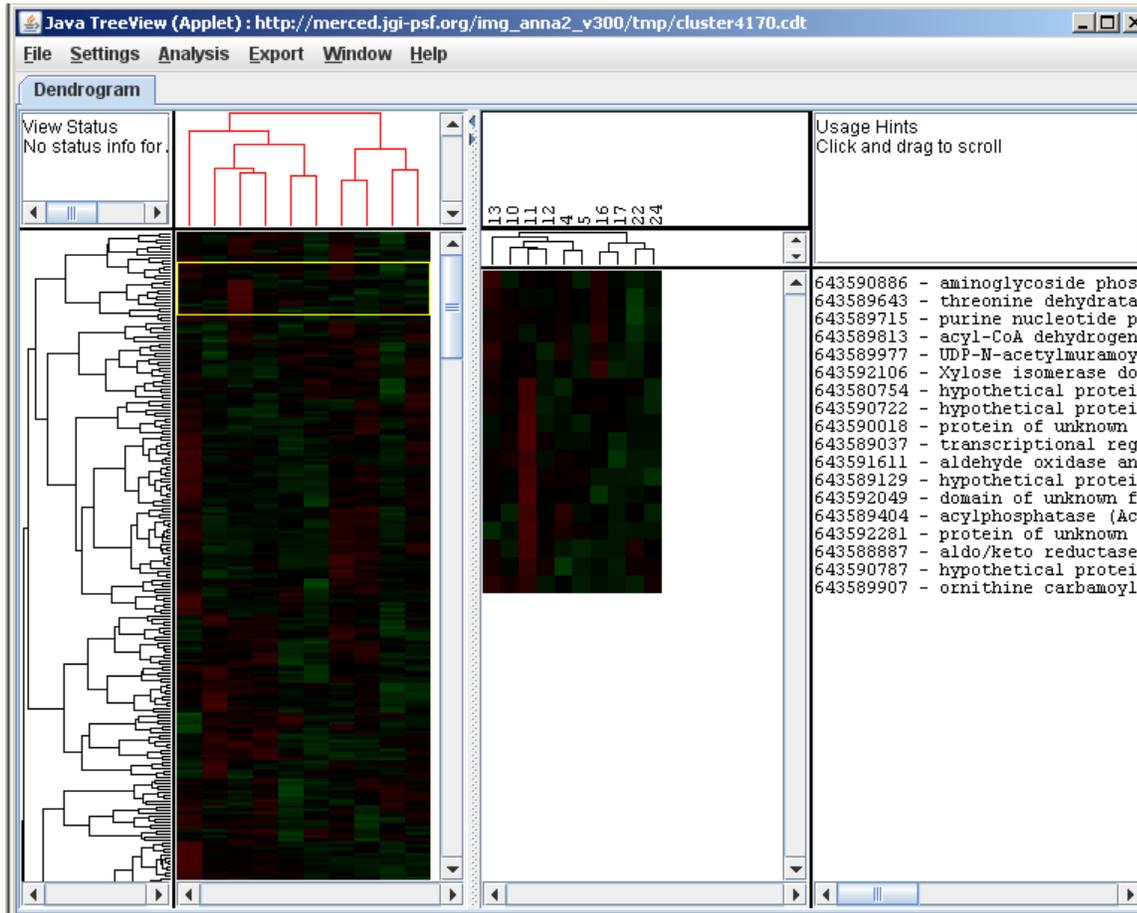


FIGURE 5. Cluster results displayed in Java TreeView application

Sample “Describe” IMG 3.2

For each study, the experiments (samples) can be selected not just for **clustering**, but also for **describe** functionality (Figure 6). The user first selects some samples to compare and then clicks on the “describe” button.

Experiments for Genome 24 Sample(s) Loaded.

Arthrobacter chlorophenolicus A6

Impact of Phenolic Substrate and Growth Temperature on the Arthrobacter chlorophenolicus

Synopsis: The *Arthrobacter chlorophenolicus* proteome was compared during growth on 4-chlorophenol (4-CP), 4-nitrophenol (4-NP), or ph enol at 5 and 28 degrees C, both for the wild-type and a mutant strain with mass spectrometry based proteomics.

Publication: 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.

Contacts: https://compbio.ornl.gov/mspipeline/jansson_Arthrobacter/contrast_dbdtryp_arthro_fin.html
Janet Jansson; Email: jjansson@lbl.gov; Web: http://www-esd.lbl.gov/ESD_staff/jansson_J/index.html
Nathan C. Verberkmoes; Email: verberkmoesn@ornl.gov

Select All Clear All

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

Select	Sample ID	Sample Description	Gene Count	Peptide Count	Total Coverage ¹ ▾	Average Coverage
<input type="checkbox"/>	5	Wt: 100 ppm 4-NP; Growth temp (C): 28; 13.2 ug protein; Tube 5; (Replicate 1)	1668	18736	155.7	.00831
<input type="checkbox"/>	1	Wt: 150 ppm 4-CP; Growth temp (C): 28; 6.6 ug protein; Tube 1; (Reference)	1774	18703	105.94	.00566

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

Describe Samples

You may select samples to compare and describe.

Describe

Figure 6. Describe Samples

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 7).

Expression Data for Selected Samples 2182 Genes Loaded.

3 samples selected (5,6,15)

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
<input type="checkbox"/>	643589402	Achl_0997	adenosine deaminase	0.0538244	0.0566572	0.0254958	57643348670	E - Nucleotide transport and metabolism	Primary immunodeficiency Purine metabolism
<input type="checkbox"/>	643589623	Achl_1218	adenosine deaminase	0.1031746	0.0714286	0.0687831	74643348670	E - Nucleotide transport and metabolism	Primary immunodeficiency Purine metabolism
<input type="checkbox"/>	643589620	Achl_1215	pyrimidine-nucleoside phosphorylase	0.0767494	0.0902935	0.1128668	74643348670	E - Nucleotide transport and metabolism	Bladder cancer Drug metabolism - other enzymes Pyrimidine metabolism
<input type="checkbox"/>	643590232	Achl_1827	Superoxide dismutase	0.1971154	0.3173077	0.1538462	110643348670	E - Inorganic ion transport and metabolism	Huntington's disease

Figure 7. 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 8). If multiple samples were selected, the pathway will be colored by percentage of samples in which a given gene is expressed (Figure 9).

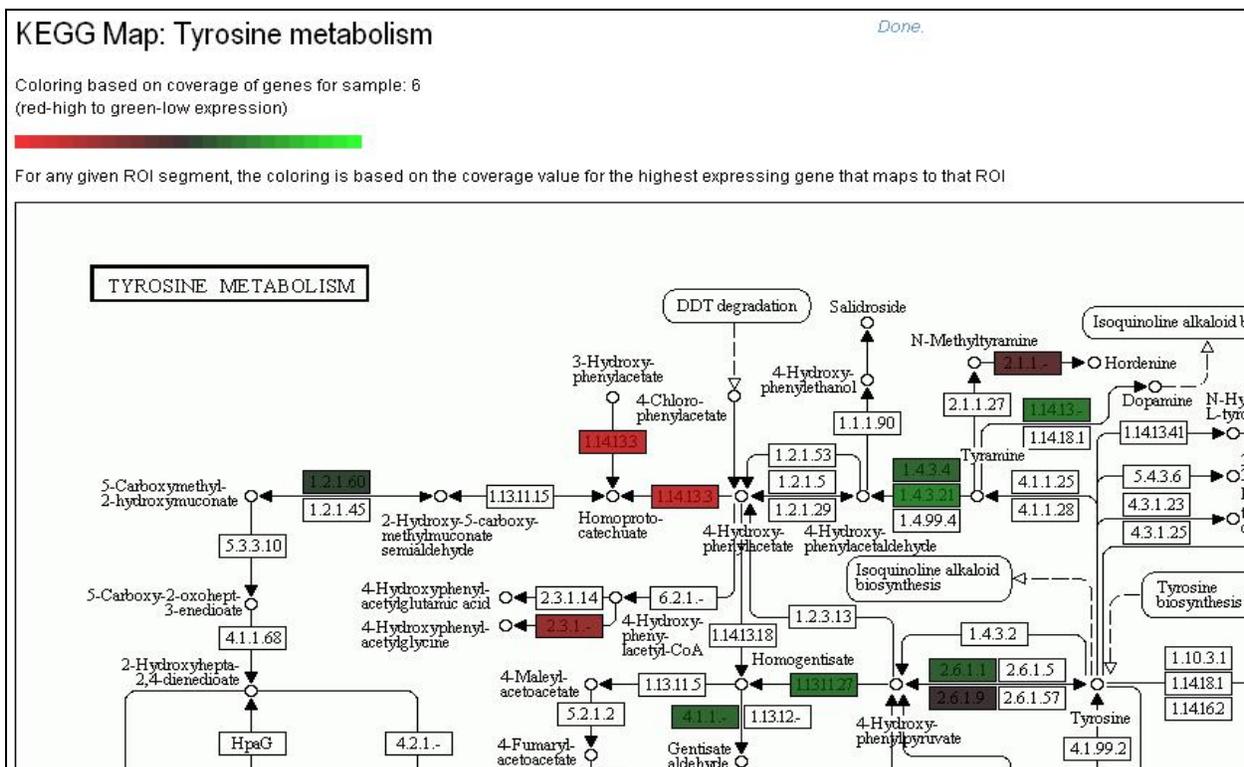


Figure 8. Kegg Map colored by gene expression in single sample

KEGG Map: Phenylalanine, tyrosine and tryptophan biosynthesis

Done.

3 samples selected (5,6,15)

■ Genes found in all selected samples

■ Genes found in some of the selected samples [for up to 25% ■ >25% ■ >50% ■ >75% ■]

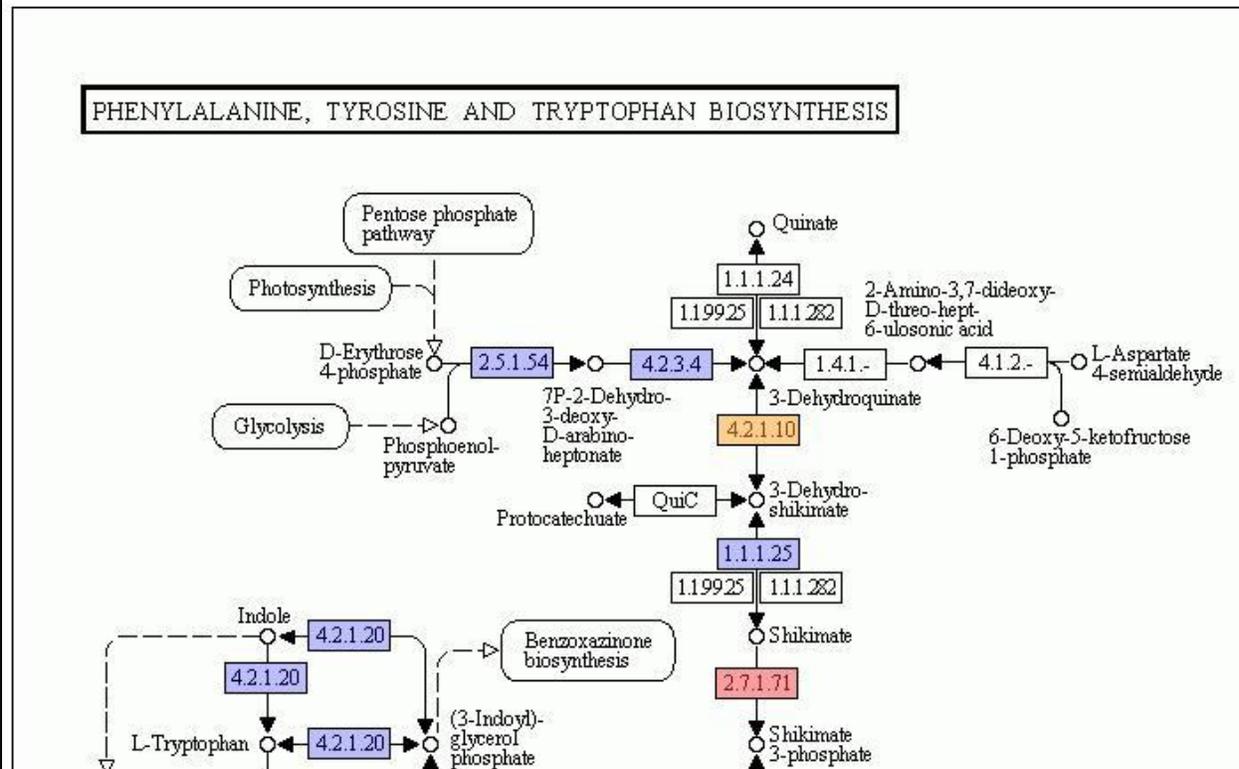


Figure 9. Kegg Map colored by percentage of samples in which a given gene is expressed

Each region of interest on the kegg map links to a list of genes for the selected samples that fall in that region (Figure 10).

Genes with KO: [K01695](#), [K01696](#) found among selected samples

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

Select	Gene ID ▲	Locus Tag	Gene Name	Sample ID	Sample Description	KO
<input type="checkbox"/>	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
<input type="checkbox"/>	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
<input type="checkbox"/>	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
<input type="checkbox"/>	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
<input type="checkbox"/>	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
<input type="checkbox"/>	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

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

Figure 10. List of genes for selected samples that fall on a given region of interest

Chromosome Viewer - Color by “Expression” IMG 3.2

There are 2 ways to color the chromosome by expression.

From the [sample page](#) - from each experimental sample page, the user can select to go to the Chromosome Viewer page (Figure 11).

Protein Expression Data for Sample *1668 Genes Loaded.*

[Wt: 100 ppm 4-NP; Growth temp \(C\): 28; 13.2 ug protein; Tube 5; \(Replicate 1\)](#)

[Chromosome Viewer](#)

Impact of Phenolic Substrate and Growth Temperature on the *Arthrobacter chlorophenicus*

Page 1 of 17 << first < prev **1** [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [next](#) > last >>

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

Figure 11. Chromosome Viewer link from Sample page

From the [genome page](#) - if a genome has samples associated with it, the “Scaffolds and Contigs” page allows the user to display the list of all available samples (by clicking on the “Select Protein Samples” button – Figure 12) and to choose one sample to color the chromosome by the expression of genes for that sample.

Chromosome Viewer *Loaded.*

Scaffolds and contigs for [Arthrobacter chlorophenicus A6](#)

User Selectable Coordinates

Select	Scaffold ▲	Length (bp)	GC	Type	Topology	No. Genes	Coordinate Range
<input type="checkbox"/>	Arthrobacter chlorophenicus A6 plasmid pACHL01: NC_011879	426858	0.64	plasmid	circular	593	1..426858

Figure 12. User can choose samples to use for coloring the chromosome viewer

The Chromosome Viewer that opens up displays coloring by “expression”, that is, coverage values for each gene in the given sample (Figure 13).

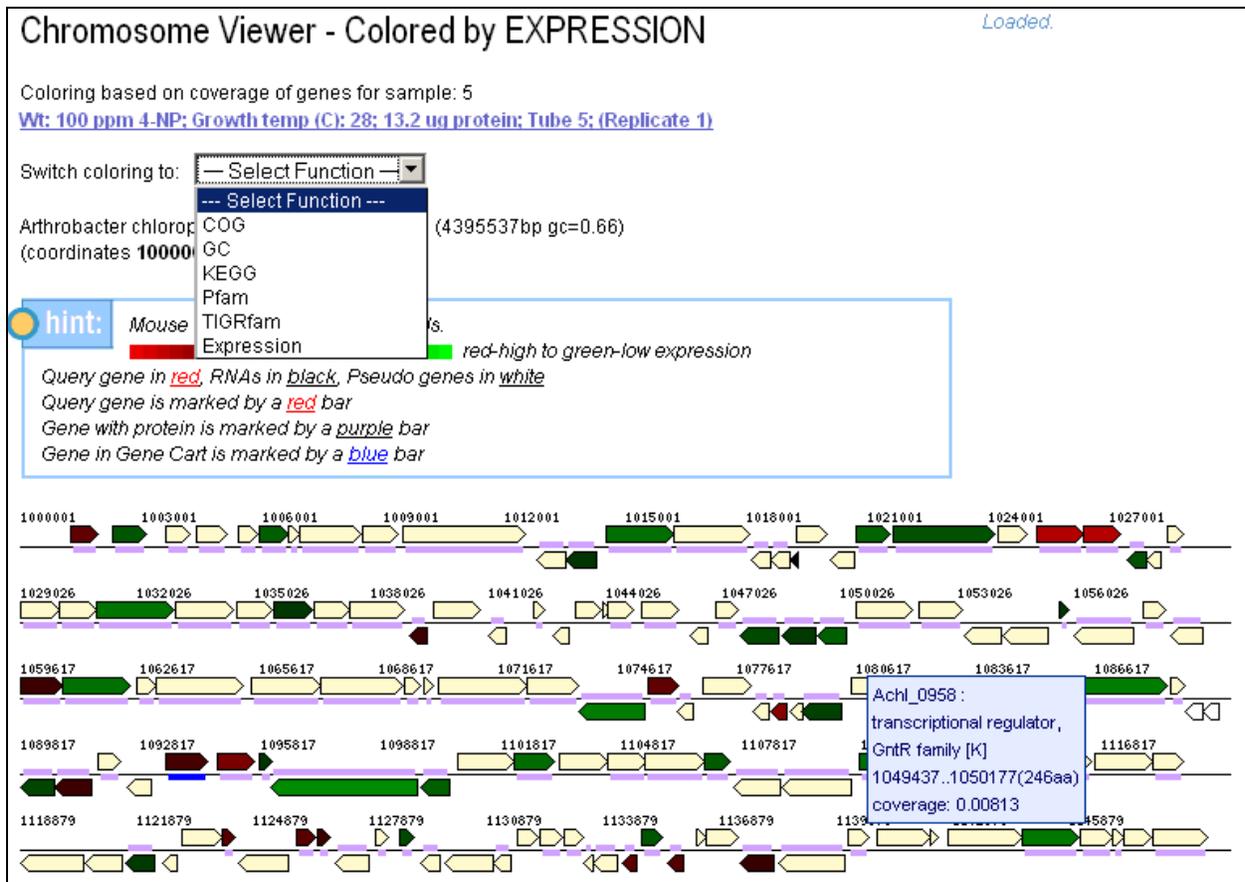


Figure 13. Chromosome Viewer colored by gene expression in a given sample

Find Up/Down Regulated Genes IMG 3.3

The user can select 2 samples of interest to see which genes are up-regulated and which are down-regulated between the two experimental conditions.

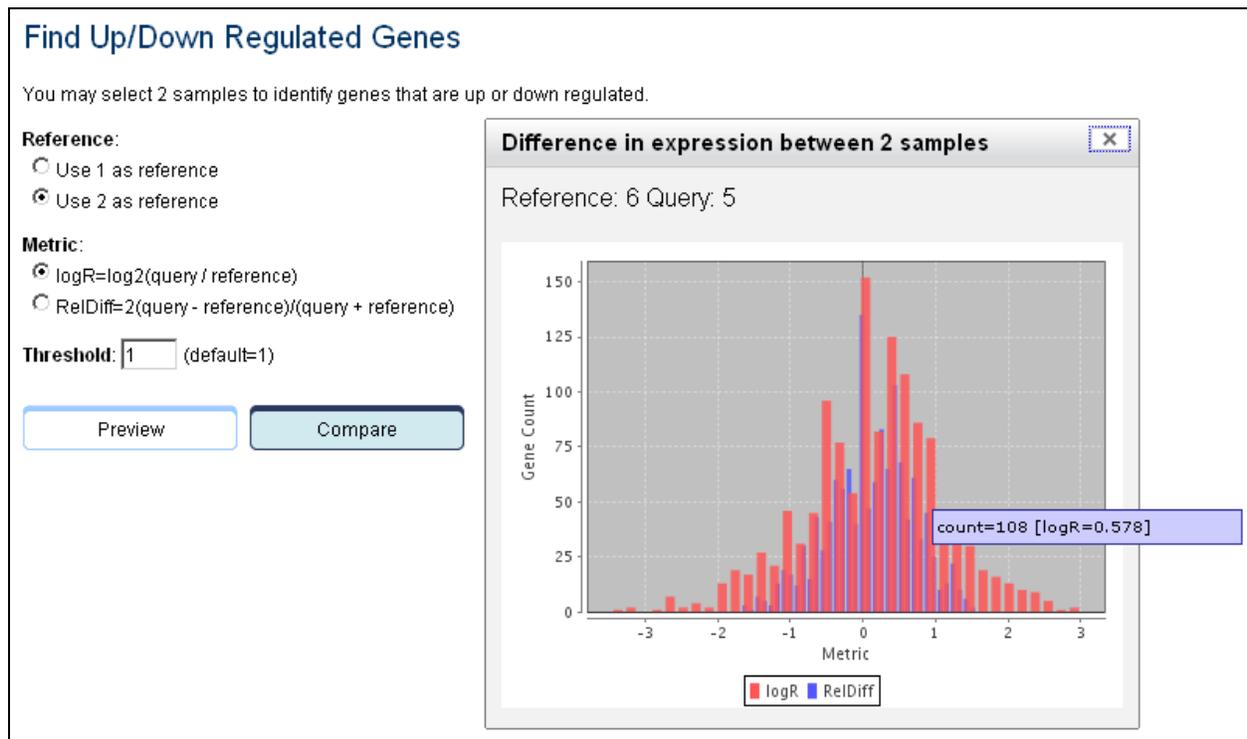


Figure 14. Find Up/Down Regulated Genes - Preview

The user can find all the genes whose expression levels differ by a specified threshold. The difference in expression is computed using either the $\log R = \log_2(\text{query} / \text{reference})$ or the $\text{RelDiff} = 2(\text{query} - \text{reference}) / (\text{query} + \text{reference})$ metric. The user can preview the comparison to see the value spread in a histogram (Figure 14). This helps in choosing which metric and what threshold value to use.

Up/Down Regulation

[Arthrobacter chlorophenolicus A6](#)

Impact of Phenolic Substrate and Growth Temperature on the *Arthrobacter chlorophenolicus*

Reference sample: [Wt. 100 ppm 4-NP; Growth temp \(C\): 28; 0.7 ug protein; Tube 6; \(Replicate 2\)](#)

Query sample: [Wt. 100 ppm 4-NP; Growth temp \(C\): 28; 13.2 ug protein; Tube 5; \(Replicate 1\)](#)

hint: Click on a tab to select up- or down- regulated genes to add to gene cart
Difference in expression levels is computed using the *logR* metric
Expression levels differ by a *threshold*: 1

Up-regulated Genes

Down-regulated Genes

Add Selections To Gene Cart

Select All

Clear All

Search column:

Search term:

Export

Page 1 of 3

<< first < prev

1

2

3

next > last >>

100

Column Selector

Select Page

Deselect Page

Select	Gene ID	Locus Tag	Product Name	6	5	logR
<input type="checkbox"/>	643590150	Achl_1740	transcriptional regulator, GntR family	0.0069444	0.0555556	3.00001
<input type="checkbox"/>	643591875	Achl_3469	FAD dependent oxidoreductase	0.0023952	0.0179641	2.90690
<input type="checkbox"/>	643590353	Achl_1949	Cytochrome b/b6 domain protein	0.0053957	0.0341727	2.66296
<input type="checkbox"/>	643589091	Achl_0694	DNA polymerase III, subunits gamma and tau	0.0016639	0.0099834	2.58496
<input type="checkbox"/>	643589253	Achl_0852	transcriptional regulator, CarD family	0.0125000	0.0750000	2.58496
<input type="checkbox"/>	643590398	Achl_1994	phosphopantothenoylcysteine decarboxylase/phosphopantothenate/cysteine ligase	0.0048193	0.0289157	2.58496

Figure 15. Find Up/Down Regulated Genes

When the user clicks “Compare”, the up- and down- regulated genes are presented in separate tabs (Figure 15). The user can then select which genes to add to the Gene Cart.

Cluster Samples

You may select samples and cluster them based on the coverage of the expressed genes. Proximity of grouping indicates the relative degree of similarity of samples to each other.

Clustering Method:

- Pairwise complete-linkage (default)
- Pairwise single-linkage
- Pairwise centroid-linkage
- Pairwise average-linkage

Distance Measure:

- No gene clustering
- Uncentered correlation
- Pearson correlation (default)
- Uncentered correlation, absolute value
- Pearson correlation, absolute value
- Spearman's rank correlation
- Kendall's tau
- Euclidean distance
- City-block distance

Minimum number of samples in which a gene should appear in order to be included: (default=2)

Use only genes from gene cart

Cluster

Reset

Figure 16a. Cluster using only genes from gene cart

The user can now go back, select all samples, and cluster them using only the genes in the Gene Cart (Figure 16a). Alternatively, the user can use “Describe” to see the Gene Cart genes in the context of the KEGG maps, chromosomal neighborhood, and cassettes that they belong to.

Describe Samples

You may select samples to compare and describe.

Use only genes from gene cart

Describe

Figure 16b. Cluster using only genes from gene cart