



Pathway Analysis

Niels W. Hanson

Ph.D. Candidate Bioinformatics

Tuesday, February 11 2014

Hallam Laboratory

Hydrocarbon MetaPathways Workshop

The University of British Columbia, Vancouver

https://github.com/nielshanson/mp_tutorial/wiki/Pathway-Analysis



Prerequisites

- Install Pathway Tools
<http://biocyc.org/download-bundle.shtml>
- Perl v5.0 <http://www.perl.org/>

Downloads

- Presentation Slides: [MetaPathways_Tutorial_Pathway_Analysis.pdf](#)
- HOT Fosmid-end ePGDBs: [HOT_Sanger_ePGDBs.zip](#)
- Perl Pathway Extractor Script: [extract_pathway_table_from_pgdb.pl](#)
- ORF Abundance Tables:
[1_upper_euphotic_rxn.wide.txt](#), [HOT_Sanger_rxn.wide.txt](#)

Goals of Tutorial

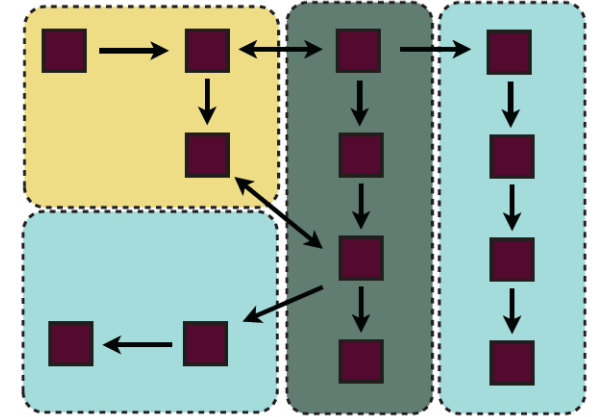
1. Give a brief overview of the Pathway Tools and the Pathologic Algorithm
2. Load processed Environmental Pathway Genome Databases (ePGDBs) into Pathway Tools
3. Explore predicted Pathways in Pathway Tools in the Cellular Overview, Pathway, and Reaction pages.
4. Highlight predicted pathways on the cellular overview and compare samples
5. Extract pathway and associated ORFs into long and wide formats.
6. Understand how the 'Omics Tools' feature to overlay quantitative metadata about pathways, e.g., ORF abundance.

1. Pathway Tools & PathoLogic



Pathway Tools

- Genes operate within structure of metabolism
- Pathway tools a software framework to integrate genomic annotations with MetaCyc pathways
- Data structure of Genes + Pathways: Pathway/Genome Database (PGDB)



Pathologic

- rule-based model on pathways class and completeness



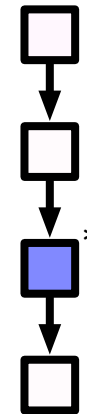
Abundance



Degradation



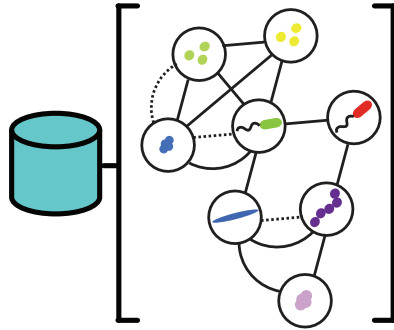
Biosynthesis



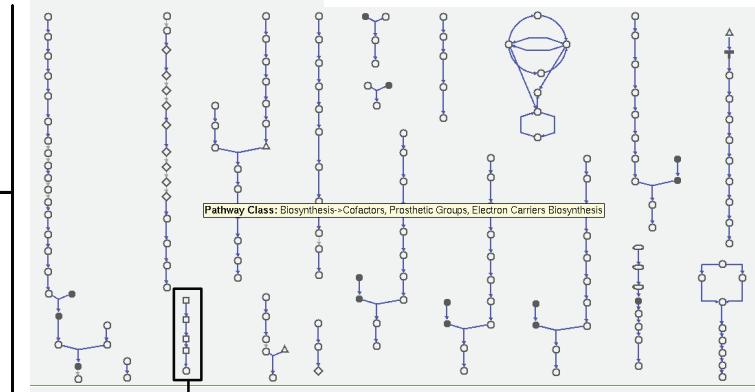
Key Enzyme

1. Pathway Tools & PathoLogic

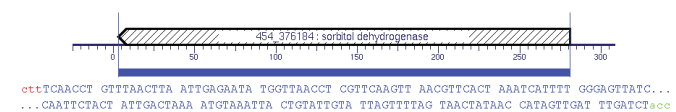
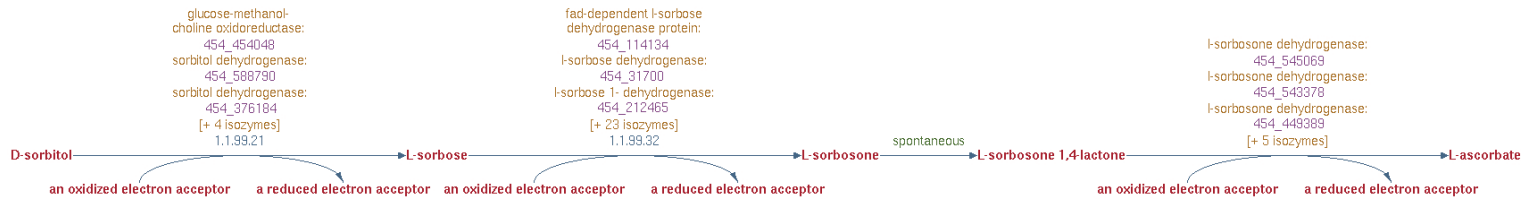
Pathway Tools



ePGDB



Cellular Overview





2. Loading ePGDBs into Pathway Tools

Two Operations:

1. Place completed ePGDB <sample>cyc/ folder into ptools-local/pgdbs/user/
2. Update organism.dat in each ePGDB <sample>cyc/1.0/input/organism.dat

```
28  -
29  ID> 1_UPPER_EUPHOTIC-
30  STORAGE>FILE-
31  NAME> unclassified sequences-
32  ABBREV-NAME>u. sequences-
33
```

```
28  -
29  ID> 1_UPPER_EUPHOTIC-
30  STORAGE>FILE-
31  NAME> 1_UPPER_EUPHOTIC-
32  ABBREV-NAME>1_UPPER_EUPHOTIC-
33
```



3. Cellular Overview, Pathway, and Reaction Pages

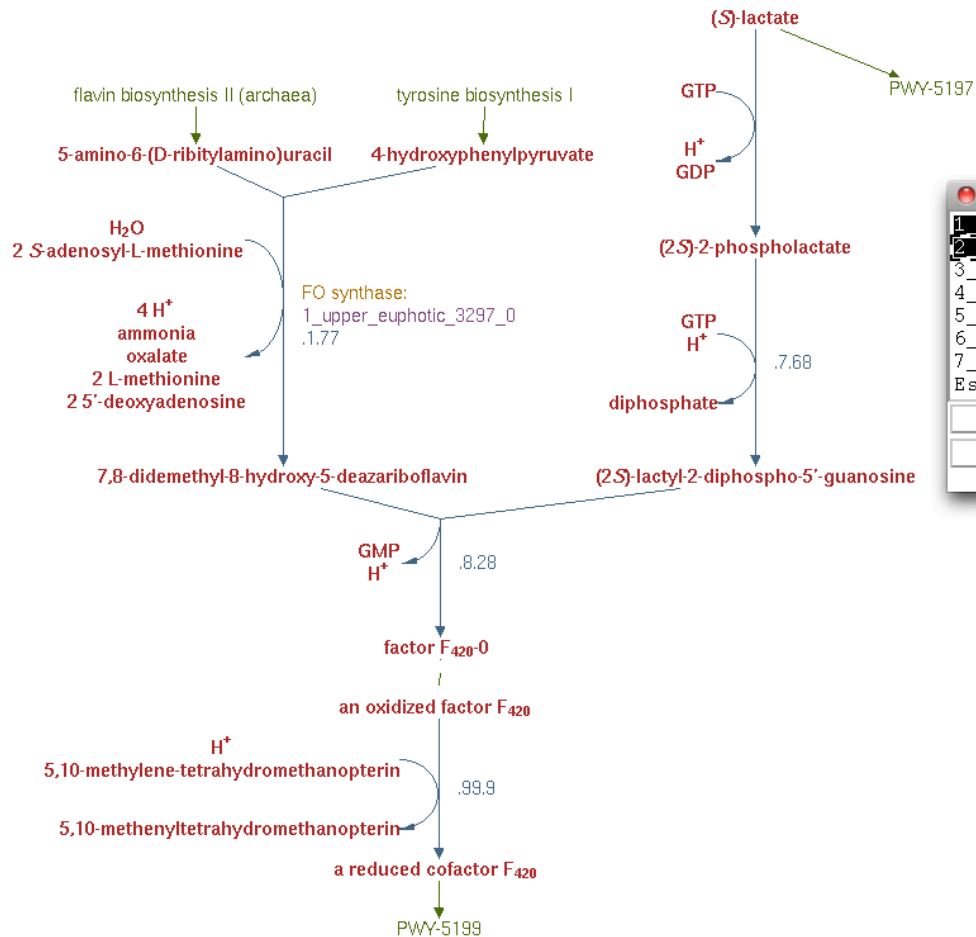
1_UPPER_EUPHOTIC Pathway: fac

Navigate to a page with a table comparing this pathway (its enzymes, genes, operon organization) across select

More Detail

Less Detail

Species Comparison



Select Organisms for Comparison

1_UPPER_EUPHOTIC
2_CHLOROPHYLLMAX
3_BELOW_EUPHOTIC
4_DEEPABYSS
5_UPPERMESOPELAGIC
6_UPPER_EUPHOTIC
7_OMZ
Escherichia coli K-12 substr. MG1655

Select All Deselect All

Cancel Use these values

3. Cellular Overview, Pathway, and Reaction Pages

Organism	Evidence Glyph	Enzymes and Genes for factor 420 biosynthesis																		
1_UPPER_EUPHOTIC		<table><tr><td>RXN-8076</td><td>None</td></tr><tr><td>EC .1.77</td><td>FO synthase: 1_upper_euphotic_3297_0</td></tr><tr><td>EC .7.68</td><td>None</td></tr><tr><td>EC .8.28</td><td>None</td></tr><tr><td>EC .99.9</td><td>None</td></tr></table>	RXN-8076	None	EC .1.77	FO synthase: 1_upper_euphotic_3297_0	EC .7.68	None	EC .8.28	None	EC .99.9	None								
RXN-8076	None																			
EC .1.77	FO synthase: 1_upper_euphotic_3297_0																			
EC .7.68	None																			
EC .8.28	None																			
EC .99.9	None																			
2_CHLOROPHYLLMAX		<table><tr><td>RXN-8076</td><td>None</td></tr><tr><td>EC .1.77</td><td>FO synthase: 2_chlorophyllmax_3549_0</td></tr><tr><td>EC .7.68</td><td>None</td></tr><tr><td>EC .8.28</td><td>LPPG:FO 2-phospho-L-lactate transferase: 2_chlorophyllmax_594_0</td></tr><tr><td></td><td>2-phospho-L-lactate transferase: 2_chlorophyllmax_5766_1</td></tr><tr><td>EC .99.9</td><td>None</td></tr></table>	RXN-8076	None	EC .1.77	FO synthase: 2_chlorophyllmax_3549_0	EC .7.68	None	EC .8.28	LPPG:FO 2-phospho-L-lactate transferase: 2_chlorophyllmax_594_0		2-phospho-L-lactate transferase: 2_chlorophyllmax_5766_1	EC .99.9	None						
RXN-8076	None																			
EC .1.77	FO synthase: 2_chlorophyllmax_3549_0																			
EC .7.68	None																			
EC .8.28	LPPG:FO 2-phospho-L-lactate transferase: 2_chlorophyllmax_594_0																			
	2-phospho-L-lactate transferase: 2_chlorophyllmax_5766_1																			
EC .99.9	None																			
3_BELOW_EUPHOTIC		<table><tr><td>RXN-8076</td><td>None</td></tr><tr><td>EC .1.77</td><td>None</td></tr><tr><td>EC .7.68</td><td>None</td></tr><tr><td>EC .8.28</td><td>2-phospho-L-lactate transferase: 3_below_euphotic_1432_1</td></tr><tr><td>EC .99.9</td><td>None</td></tr></table>	RXN-8076	None	EC .1.77	None	EC .7.68	None	EC .8.28	2-phospho-L-lactate transferase: 3_below_euphotic_1432_1	EC .99.9	None								
RXN-8076	None																			
EC .1.77	None																			
EC .7.68	None																			
EC .8.28	2-phospho-L-lactate transferase: 3_below_euphotic_1432_1																			
EC .99.9	None																			
4_DEEPABYSS		<table><tr><td>RXN-8076</td><td>None</td></tr><tr><td>EC .1.77</td><td>FO synthase subunit 1: 4_deepabyss_3687_1</td></tr><tr><td></td><td>FO synthase subunit 1: 4_deepabyss_4601_1</td></tr><tr><td>EC .7.68</td><td>None</td></tr><tr><td>EC .8.28</td><td>2-phospho-L-lactate transferase: 4_deepabyss_4262_1</td></tr><tr><td></td><td>2-phospho-L-lactate transferase: 4_deepabyss_3652_0</td></tr><tr><td></td><td>LPPG:FO 2-phospho-L-lactate transferase: 4_deepabyss_3332_1</td></tr><tr><td></td><td>2-phospho-L-lactate transferase: 4_deepabyss_10739_0</td></tr><tr><td>EC .99.9</td><td>None</td></tr></table>	RXN-8076	None	EC .1.77	FO synthase subunit 1: 4_deepabyss_3687_1		FO synthase subunit 1: 4_deepabyss_4601_1	EC .7.68	None	EC .8.28	2-phospho-L-lactate transferase: 4_deepabyss_4262_1		2-phospho-L-lactate transferase: 4_deepabyss_3652_0		LPPG:FO 2-phospho-L-lactate transferase: 4_deepabyss_3332_1		2-phospho-L-lactate transferase: 4_deepabyss_10739_0	EC .99.9	None
RXN-8076	None																			
EC .1.77	FO synthase subunit 1: 4_deepabyss_3687_1																			
	FO synthase subunit 1: 4_deepabyss_4601_1																			
EC .7.68	None																			
EC .8.28	2-phospho-L-lactate transferase: 4_deepabyss_4262_1																			
	2-phospho-L-lactate transferase: 4_deepabyss_3652_0																			
	LPPG:FO 2-phospho-L-lactate transferase: 4_deepabyss_3332_1																			
	2-phospho-L-lactate transferase: 4_deepabyss_10739_0																			
EC .99.9	None																			
5_UPPERMESOPELAGIC		<p>This pathway is not marked as present in this organism. No enzymes or genes have been identified for this pathway</p>																		

3. Cellular Overview, Pathway, and Reaction Pages

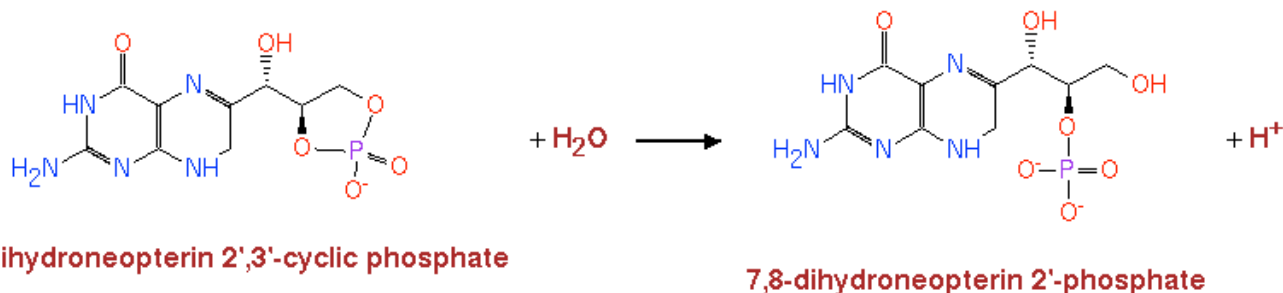
1_UPPER_EUPHOTIC Reaction: [no EC number assigned]

Species Comparison

Superclasses: **Reactions-Classified-By-Conversion-Type -> Simple-Reactions -> Chemical-Reactions**
Reactions-Classified-By-Substrate -> Small-Molecule-Reactions

In Pathway: **6-hydroxymethyl-dihydropterin diphosphate biosynthesis II (archaea)**

Atom Mapping: None found for this reaction.



- The reaction direction shown, that is, $\text{A} + \text{B} \leftrightarrow \text{C} + \text{D}$ versus $\text{C} + \text{D} \leftrightarrow \text{A} + \text{B}$, is in accordance with the direction in which it was curated.
- Most BioCyc compounds have been protonated to a reference pH value of 7.3, and some reactions have been computationally balanced for hydrogen by adding free protons. Please see the PGDB Concepts Guide for more information.
- Mass balance status: Balanced.

Credits: Created in MetaCyc 26-Apr-2011 by **Caspi R, SRI International**
 Imported from **MetaCyc** 13-Jan-2014 by

4. Highlight predicted pathways to compare

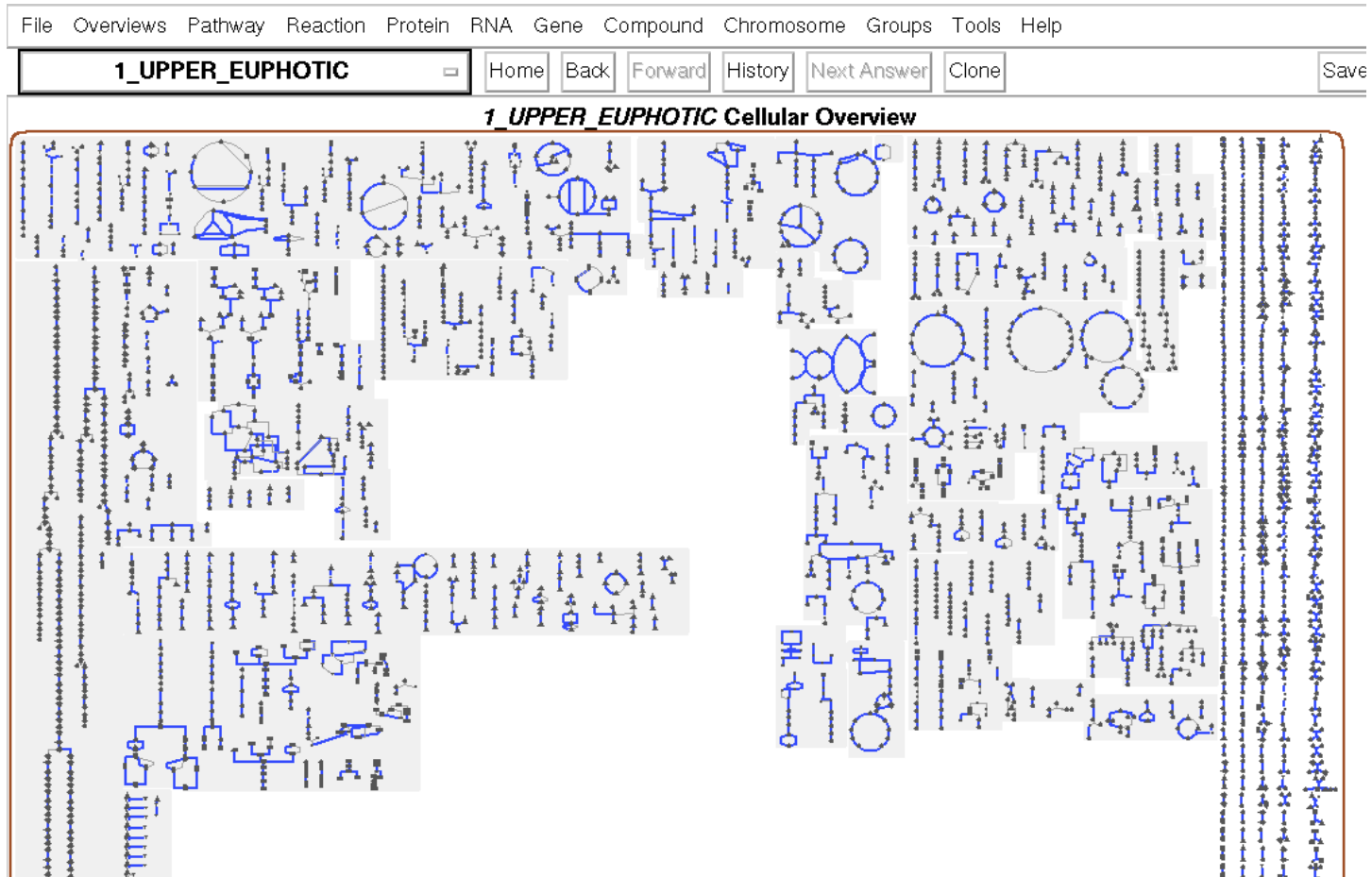
[Overviews](#)
[Pathway](#)
[Reaction](#)
[Protein](#)
[RNA](#)
[Gene](#)

Omics Viewer: Overlay Experimental Data from [▶](#)

Clear All Highlighting

Show Cellular Overview [Ctrl+O](#)

Show Legend



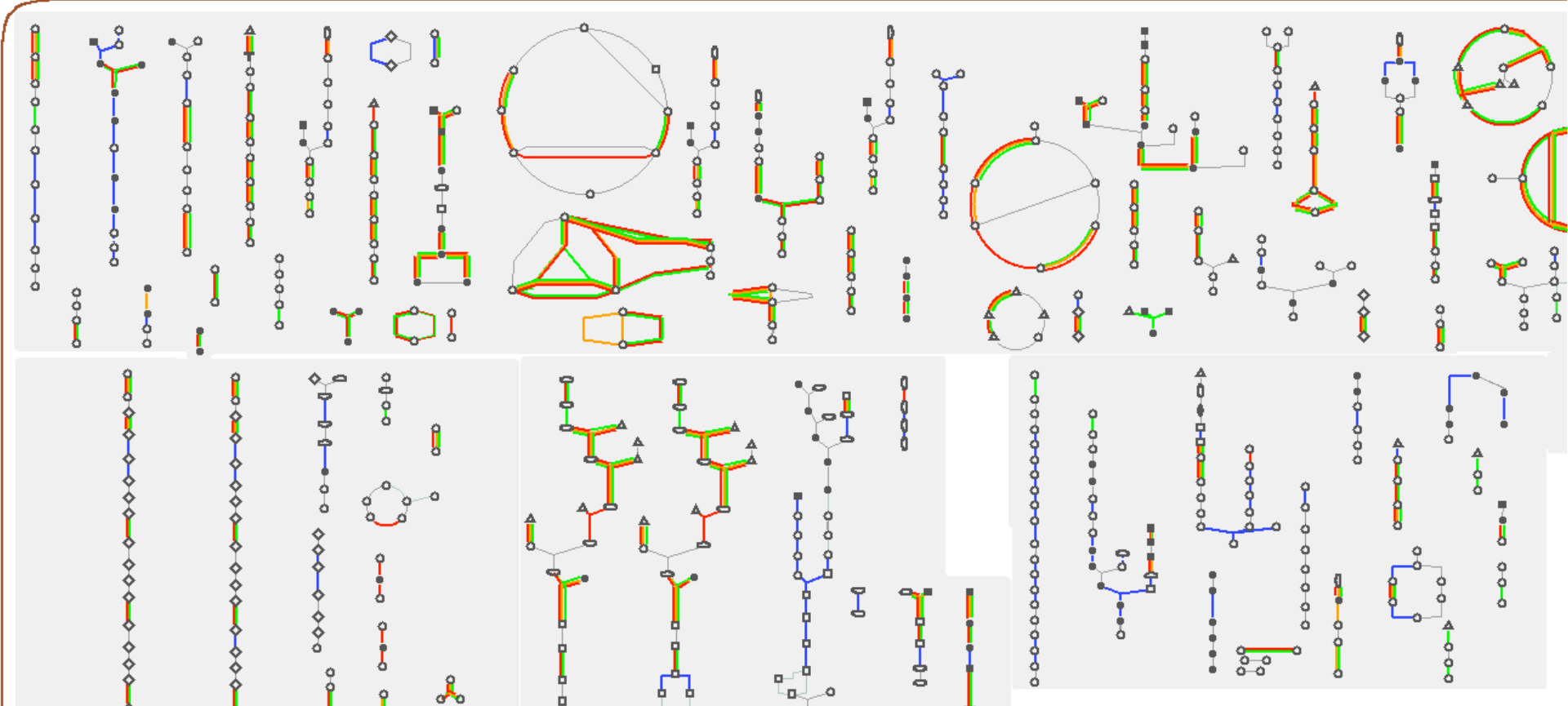
4. Highlight predicted pathways to compare

The screenshot shows the 'Omics Viewer' application with the 'Highlight' menu open. The menu options include: 'Clear All Highlighting', 'Show Cellular Overview' (Ctrl+O), 'Show Legend', 'Show/Hide Transport Links', 'Highlight' (which has a submenu), 'Metabolite Tracing', 'Print as Poster', 'Update...', 'Show Complete Regulatory Overview' (Ctrl+R), 'Highlight Genes', 'Redisplay Highlighted Genes Only', 'Zoom Regulatory Overview', 'Preferences for Regulatory Overview', 'Show Regulatory Key Code', 'Save Current Regulatory Overview to File', 'Load Regulatory Overview from File', 'Show Genome Overview' (Ctrl+G), and 'Highlight Genes by Substring'. The 'Highlight' submenu is open, showing options: 'Species Comparison...', 'Pathway', 'Reaction(s)', 'Gene', 'Compound(s)', 'Answer List', 'Undo', 'Redo', 'Save to File...', and 'Load from File...'. The background shows a metabolic pathway diagram titled 'EUPHOTIC Cellular C'.

The 'Species Comparison Dialog' window is shown. It contains the text: 'Highlight all 1_UPPER_EUPHOTIC reactions that are'. Below this, there are two dropdown menus: 'Shared' and 'Any', followed by the text 'with' and 'of the following species:'. A list of species is provided: 2_CHLOROPHYLLMAX, 3_BELOW_EUPHOTIC, 4_DEEPABYSS, 5_UPPERMESOPELAGIC, 6_UPPER_EUPHOTIC, 7_OMZ, and E. coli. To the right of the list are two buttons: 'Select All' and 'Deselect All'. At the bottom of the dialog are 'OK' and 'Cancel' buttons.

4. Highlight predicted pathways to compare

1_UPPER_EUPHOTIC Cellular Overview





5. Extract pathways and ORFs

extract_pathway_table_from_pgdb.pl

1. Start Pathway Tools in `–api` mode:
`pathway-tools/pathway-tools –api`
2. In **another shell** run `extract_pathway_table_from_pgdb.pl` to extract pathways:

```
# List Available ePGDBs in Pathway Tools
perl extract_pathway_table_from_pgdb.pl -l
```

```
# Extract pathways from pathway tools
perl extract_pathway_table_from_pgdb.pl
    -f [list of samples]
    -out [output file]
    -c [pathway coverage] <0.0-1.0>
    -s [pathway support] <1-100>
    -t [type of output table] <lookup, long, or wide>
    -orfs [output orfs in lookup mode]
    -rxn [output reactions in wide mode]
```





5. Extract pathways and ORFs

```
perl extract_pathway_table_from_pgdb.pl -f 1_upper_euphotic -out  
1_upper_euphotic.lookup.txt -t lookup
```

```
head 1_upper_euphotic.lookup.txt
```

SAMPLE	PWY_NAME	PWY_COMMON_NAME	NUM_REACTIONS	NUM_COVERED_REACTIONS	ORF_COUNT
1_upper_euphotic	PWY-5913	TCA cycle VI (obligate autotrophs)	11	10	65
1_upper_euphotic	REDCITCYC	TCA cycle III (helicobacter)	9	8	33
1_upper_euphotic	PWY-5690	TCA cycle II (eukaryotic)	9	8	42
1_upper_euphotic	TCA	TCA cycle I (prokaryotic)	10	9	43
1_upper_euphotic	ANARESP1-PWY	respiration (anaerobic)	13	10	44

Only output pathways with more than 50% reactions covered and at least 7 ORFs:

```
perl extract_pathway_table_from_pgdb.pl -f 1_upper_euphotic -out  
1_upper_euphotic.lookup.c05.s7.txt -t lookup -c 0.5 -s 7
```





5. Extract pathways and ORFs

the **-t long** table format displays each each ORF in each pathway:

```
perl extract_pathway_table_from_pgdb.pl -f 1_upper_euphotic -out
1_upper_euphotic.long.txt -t long
```

```
head 1_upper_euphotic.long.txt
```

SAMPLE	PWY_NAME	PWY_COMMON_NAME	NUM_REACTIONS	NUM_COVERED_REACTIONS	ORF_COUNT	ORF
1_upper_euphotic	PWY-5913	TCA cycle VI (obligate autotrophs)	11	10	65	1_upper_euphotic_3417_1
1_upper_euphotic	PWY-5913	TCA cycle VI (obligate autotrophs)	11	10	65	1_upper_euphotic_5953_0
1_upper_euphotic	PWY-5913	TCA cycle VI (obligate autotrophs)	11	10	65	1_upper_euphotic_7270_0
1_upper_euphotic	PWY-5913	TCA cycle VI (obligate autotrophs)	11	10	65	1_upper_euphotic_643_1
1_upper_euphotic	PWY-5913	TCA cycle VI (obligate autotrophs)	11	10	65	1_upper_euphotic_14_0
1_upper_euphotic	PWY-5913	TCA cycle VI (obligate autotrophs)	11	10	65	1_upper_euphotic_670_0

the **-t wide** with multiple samples creates a “master” table:

```
# - wide table format of pathways from multiple samples
perl extract_pathway_table_from_pgdb.pl -f 1_upper_euphotic 6_upper_euphotic
2_chlorophyllmax 3_below_euphotic 5_uppermesopelagic 7_omz 4_deepabyss -out
HOT_Sanger_pwy.wide.txt -t wide
```

```
head 1_upper_euphotic.long.txt
```

PWY	1_upper_euphotic	6_upper_euphotic	2_chlorophyllmax	3_below_euphotic	5_uppermesopelagic	7_omz	4_deepabyss
SUCSYN-PWY	0	16	10	8	17	13	16
PWY-6733	0	1	0	0	0	0	0
PWY-5274	0	0	0	1	0	0	0
PWY-6728	0	0	0	46	0	59	0
PWY-241	12	6	8	6	7	12	7





5. Extract pathways and ORFs

the **-t wide -rxn** options produce a list of reactions and abundance:

```
perl extract_pathway_table_from_pgdb.pl -f 1_upper_euphotic -out
1_upper_euphotic_rxn.wide.txt -t wide -rxn
```

```
head 1_upper_euphotic_rxn.wide.txt
```

```
RXN    1_upper_euphotic
RXN1G-617    1
RXN-6641    1
RXN0-2381    2
DTPDGLUCOSEPP-RXN 2
RXN0-6479    3
DADPKIN-RXN 4
```

the **-t wide -rxn** with multiple samples creates a “master” rxn table:

```
perl extract_pathway_table_from_pgdb.pl -f 1_upper_euphotic 6_upper_euphotic
2_chlorophyllmax 3_below_euphotic 5_uppermesopelagic 7_omz 4_deepabyss -out
HOT_Sanger_rxn.wide.txt -t wide -rxn
```

```
head HOT_Sanger_rxn.wide.txt
```

PWY	1_upper_euphotic	6_upper_euphotic	2_chlorophyllmax	3_below_euphotic	5_uppermesopelagic	7_omz	4_deepabyss
SUCSYN-PWY	0	16	10	8	17	13	16
PWY-6733	0	1	0	0	0	0	0
PWY-5274	0	0	0	1	0	0	0
PWY-6728	0	0	0	46	0	59	0
PWY-241	12	6	8	6	7	12	7





5. Extract pathways and ORFs

the **-t wide -rxn** options produce a list of reactions and abundance:

```
perl extract_pathway_table_from_pgdb.pl -f 1_upper_euphotic -out
1_upper_euphotic_rxn.wide.txt -t wide -rxn
```

```
head 1_upper_euphotic_rxn.wide.txt
```

```
RXN    1_upper_euphotic
RXN1G-617    1
RXN-6641    1
RXN0-2381    2
DTDPGLUCOSEPP-RXN  2
RXN0-6479    3
DADPKIN-RXN  4
```

****Use these two to highlight ORF counts on the Cellular Overview****

the **-t wide -rxn** with multiple samples creates a “master” rxn table:

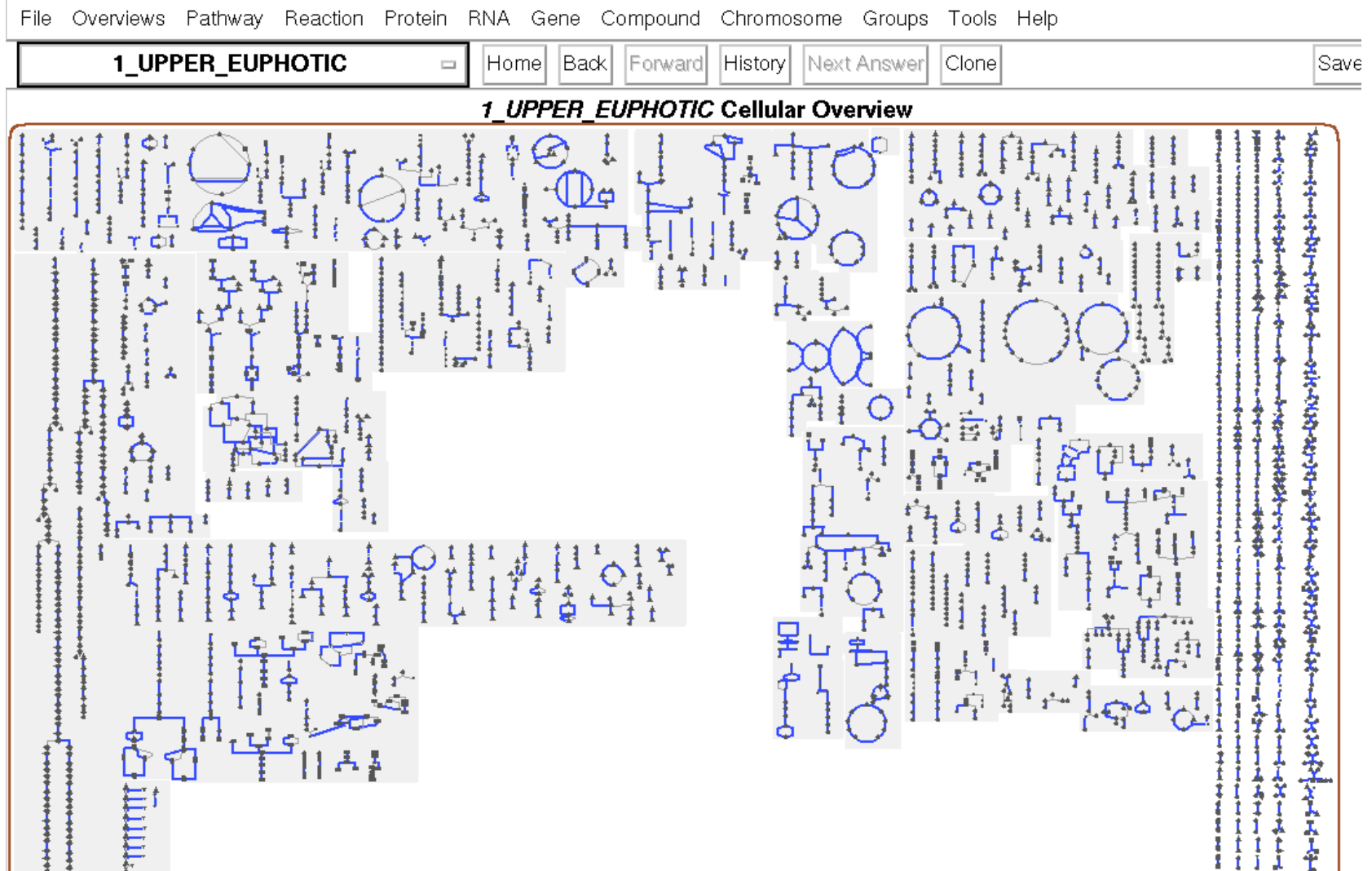
```
perl extract_pathway_table_from_pgdb.pl -f 1_upper_euphotic 6_upper_euphotic
2_chlorophyllmax 3_below_euphotic 5_uppermesopelagic 7_omz 4_deepabyss -out
HOT_Sanger_rxn.wide.txt -t wide -rxn
```

```
head HOT_Sanger_rxn.wide.txt
```

```
RXN    1_upper_euphotic  6_upper_euphotic  2_chlorophyllmax  3_below_euphotic  5_uppermesopelagic
      7_omz  4_deepabyss
1.7.7.2-RXN  0      0      2      0      0      1      1
PPGPPSYN-RXN      0      1      0      0      0      0      0
RXN-6641      1      0      1      0      1      3      2
RXN3DJ-170    1      0      0      0      0      0      2
```

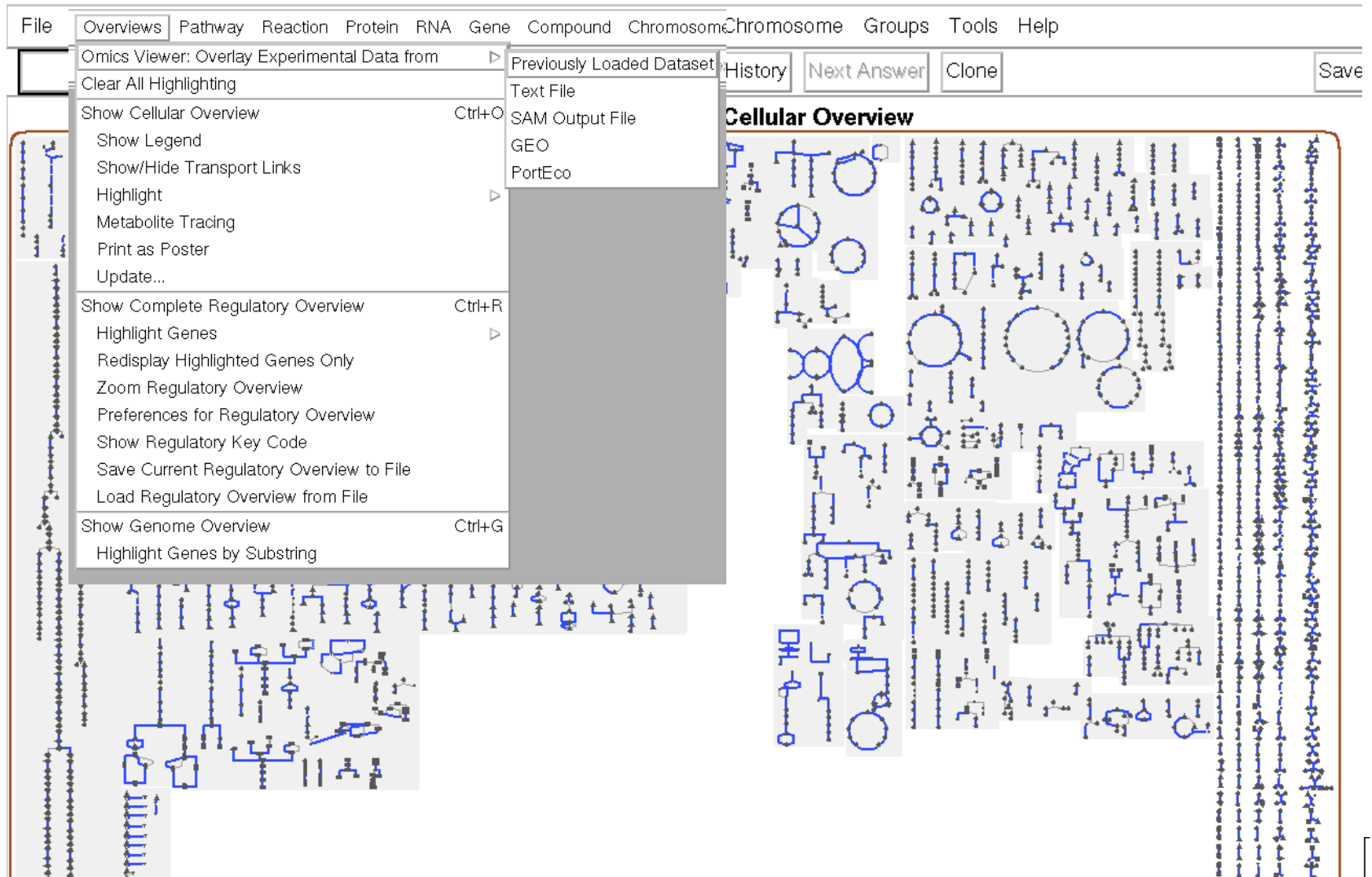


6. 'Omics Tools' feature for Metadata



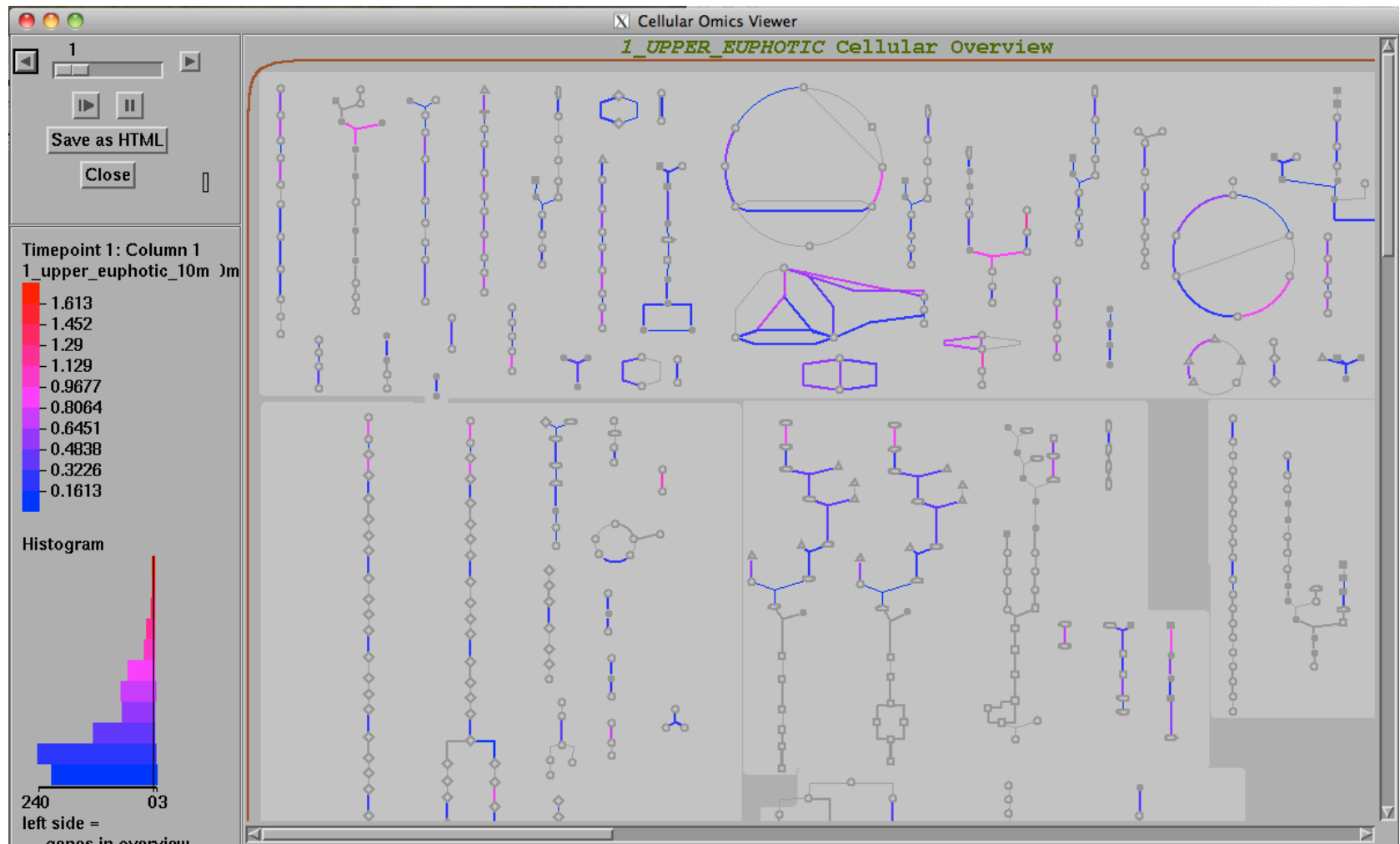
6. 'Omics Tools' feature for Metadata

Load 1_upper_euphotic_rxn.wide.txt

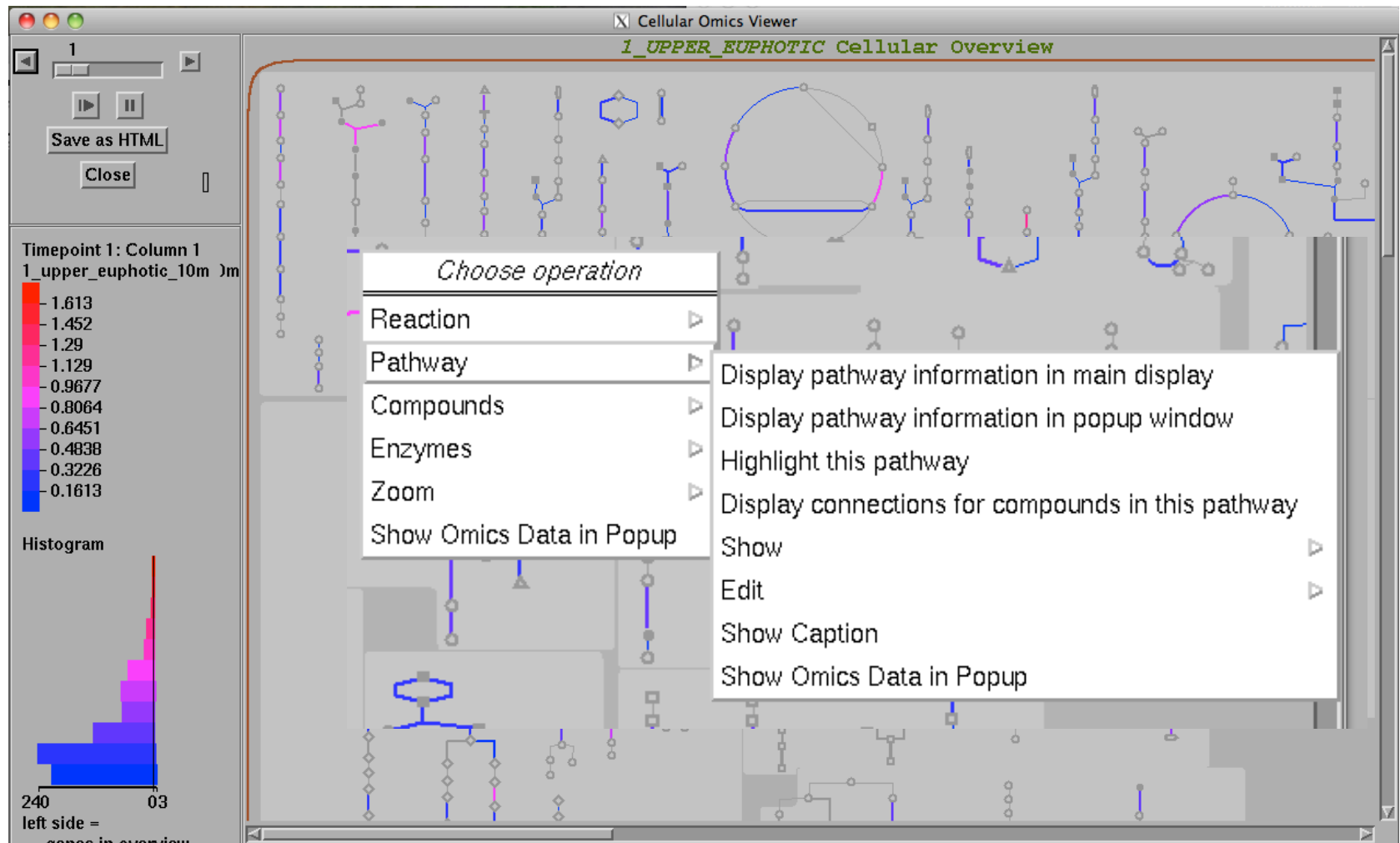


The screenshot displays the 'Omics Tools' software interface. The top menu bar includes 'File', 'Overviews', 'Pathway', 'Reaction', 'Protein', 'RNA', 'Gene', 'Compound', 'Chromosome', 'Chromosome', 'Groups', 'Tools', and 'Help'. The 'Omics Viewer: Overlay Experimental Data from' menu is open, showing options like 'Previously Loaded Dataset', 'Text File', 'SAM Output File', 'GEO', and 'PortEco'. Other menu items include 'Clear All Highlighting', 'Show Cellular Overview' (Ctrl+O), 'Show Legend', 'Show/Hide Transport Links', 'Highlight', 'Metabolite Tracing', 'Print as Poster', 'Update...', 'Show Complete Regulatory Overview' (Ctrl+R), 'Highlight Genes', 'Redisplay Highlighted Genes Only', 'Zoom Regulatory Overview', 'Preferences for Regulatory Overview', 'Show Regulatory Key Code', 'Save Current Regulatory Overview to File', 'Load Regulatory Overview from File', 'Show Genome Overview' (Ctrl+G), and 'Highlight Genes by Substring'. The 'Cellular Overview' panel on the right shows a complex network diagram with various nodes and connections. The bottom left corner features the 'Genome Alberta' logo.

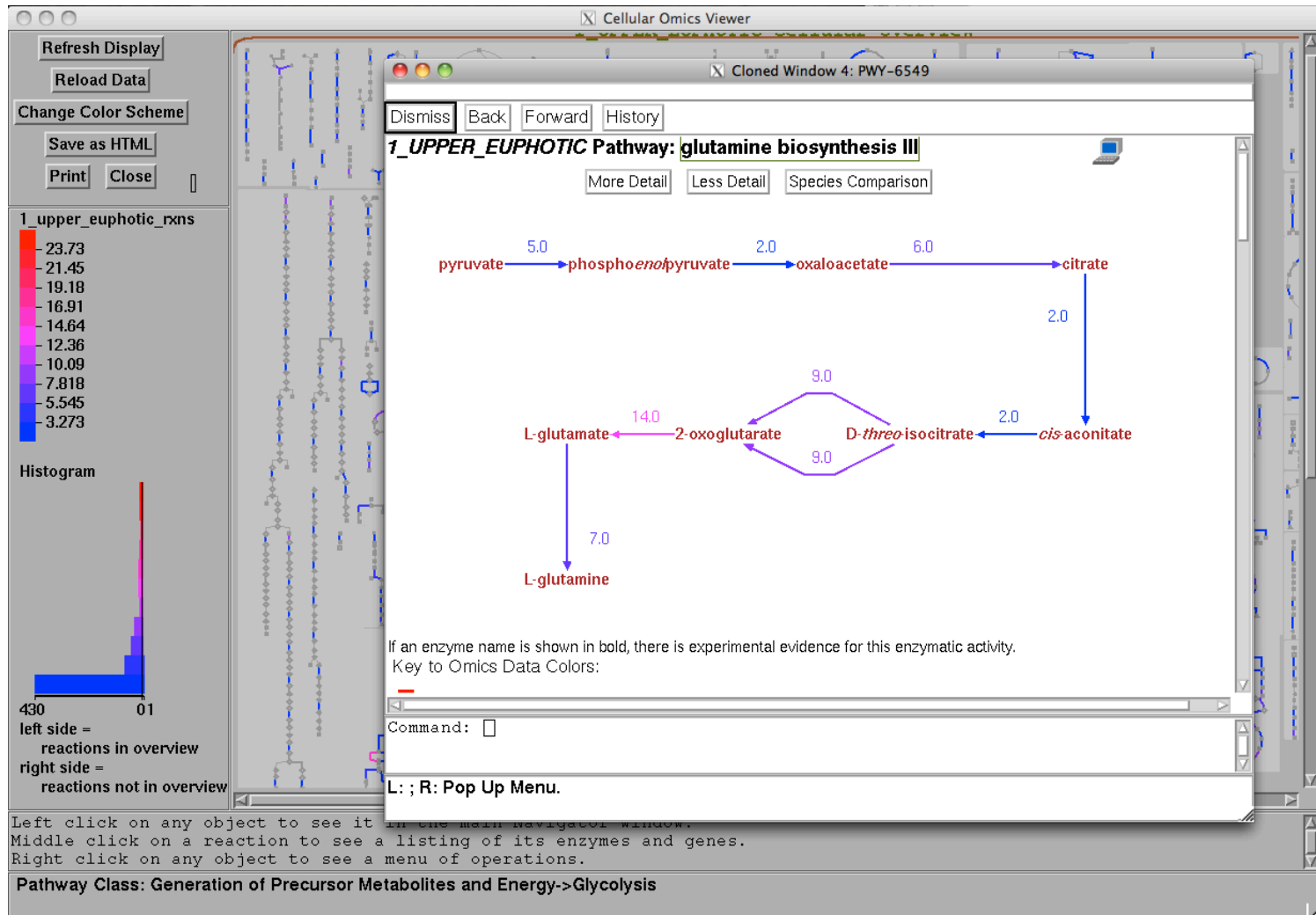
6. 'Omics Tools' feature for Metadata



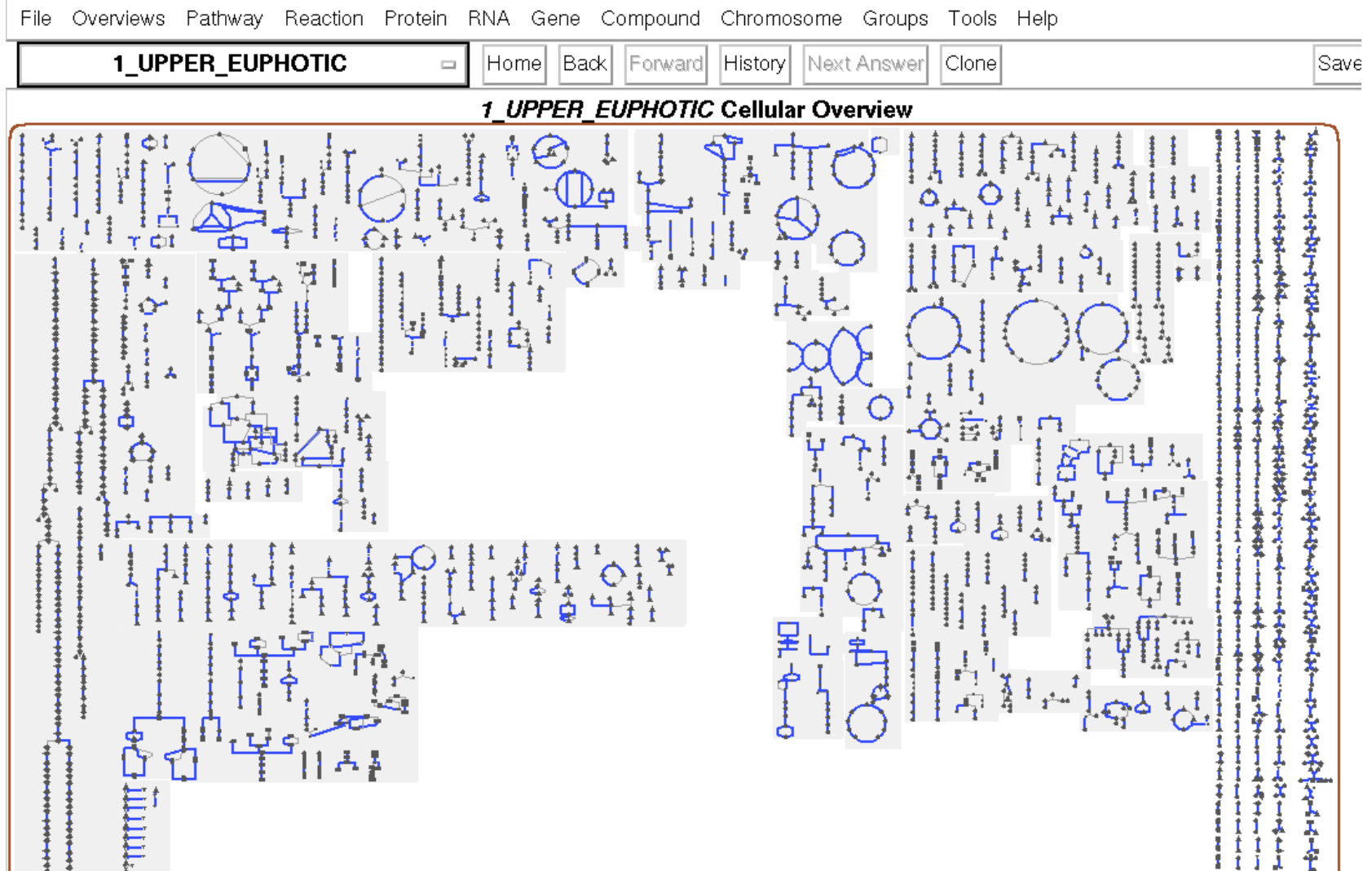
6. 'Omics Tools' feature for Metadata



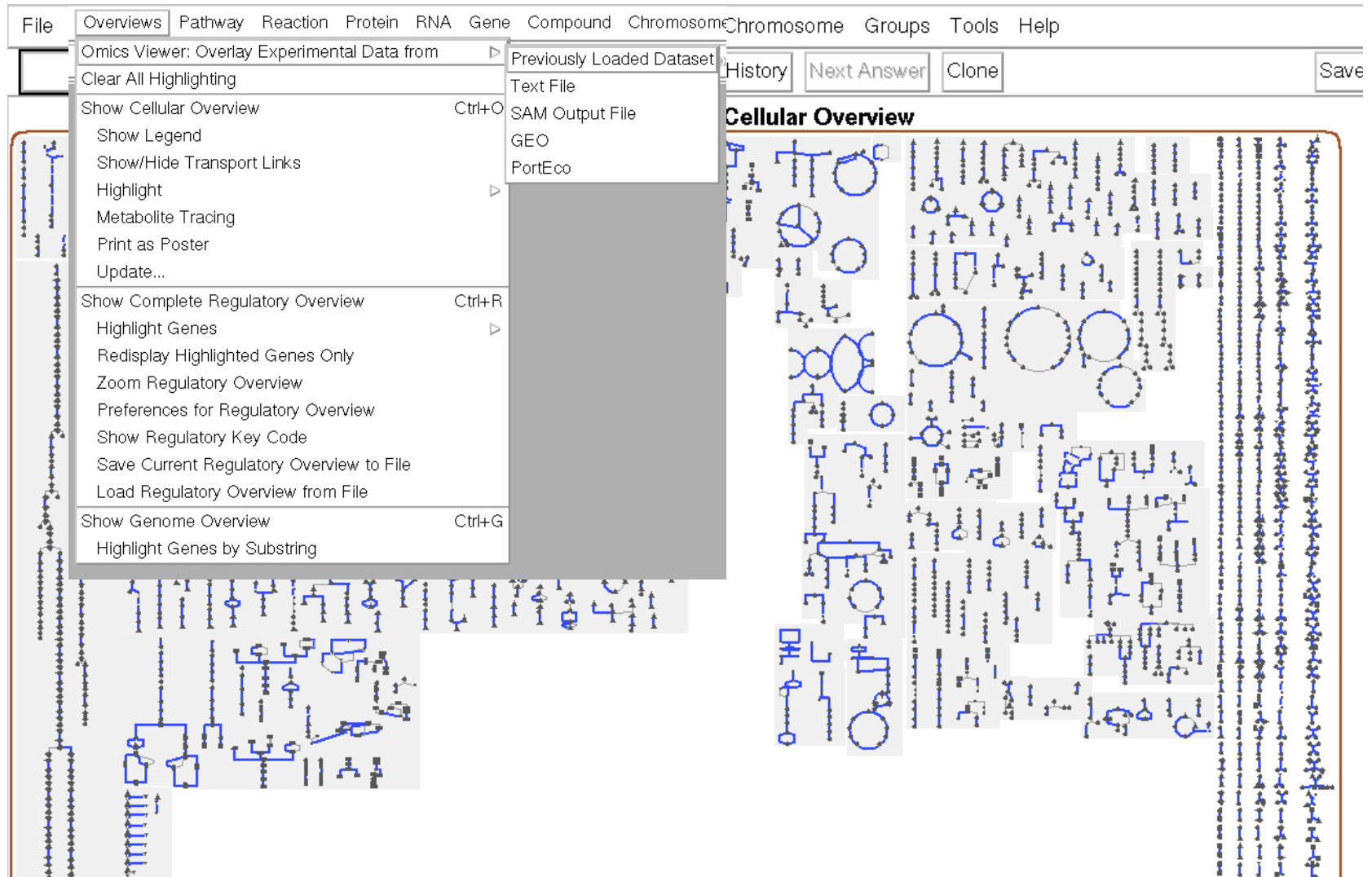
6. 'Omics Tools' feature for Metadata



6. 'Omics Tools' feature for Metadata



6. 'Omics Tools' feature for Metadata



The screenshot displays a software interface with a menu bar at the top: File, Overview, Pathway, Reaction, Protein, RNA, Gene, Compound, Chromosome, Chromosome, Groups, Tools, and Help. The 'Tools' menu is open, showing a list of options:

- Omics Viewer: Overlay Experimental Data from
 - Previously Loaded Dataset
 - Text File
 - SAM Output File
 - GEO
 - PortEco
- Clear All Highlighting
- Show Cellular Overview (Ctrl+O)
 - Show Legend
 - Show/Hide Transport Links
 - Highlight
 - Metabolite Tracing
 - Print as Poster
 - Update...
- Show Complete Regulatory Overview (Ctrl+R)
 - Highlight Genes
 - Redisplay Highlighted Genes Only
 - Zoom Regulatory Overview
 - Preferences for Regulatory Overview
 - Show Regulatory Key Code
 - Save Current Regulatory Overview to File
 - Load Regulatory Overview from File
- Show Genome Overview (Ctrl+G)
 - Highlight Genes by Substring

The background of the interface shows a 'Cellular Overview' diagram, which is a complex network of blue lines and nodes representing cellular pathways and interactions. The diagram is divided into several sections, with some areas highlighted in orange. The 'Cellular Overview' title is visible at the top of the diagram area.

6. 'Omics Tools' feature for Metadata

File Overviews Pathway Reaction Protein RNA Gene Compound Chromosome Chromosome Groups Tools Help

Omics Viewer: Overlay Experimental Data from Previously Loaded Dataset History Next Answer Clone Save

Clear All Highlighting Text File

Show Cellular Overview Ctrl+O SAM Output File Cellular Overview

Omics Viewer

Experiment Title:

File:

Paint data on: ☒ Cellular Overview Diagram ☐ Regulatory Overview Diagram ☐ Genome Overview Diagram

Type of display: ☒ Single Experiment ☐ Animation

Column containing identifiers: Identifier types: ?

Data columns to use: ?

Select type of values: ?

Assign a label to each timepoint (optional)

Choose color scheme: ?

Specify Timepoint Labels

Timepoint labels can be specified either here or in the data file. To include them in the data file, make sure the first line (excluding comment lines or blank lines) starts with a "\$" character. The values in each column of this first line will then be treated as labels. If labels are specified both in the data file and here, the labels entered here will take precedence.

Enter label for each timepoint column:

1:

2:

3:

4:

5:

6:

7:

6. 'Omics Tools' feature for Metadata

File Overviews Pathway Reaction Protein RNA Gene Compound Chromosome Chromosome Groups Tools Help

Omics Viewer: Overlay Experimental Data from Previously Loaded Dataset History Next Answer Clone Save

Clear All Highlighting Text File

Show Cellular Overview Ctrl+O SAM Output File Cellular Overview

Experiment Title: HOT_Sanger_Reaction_Series

File: /Users/nielsh/mp_tutorial/pathway_analysis/code/HOT_Sanger_rxn_log.wide.txt

Paint data on: ☐ Cellular Overview Diagram ☐ Regulatory Overview Diagram ☐ Genome Overview Diagram

Type of display: ☐ Single Experiment ☒ Animation

Column containing identifiers: Identifier types: Reaction IDs/EC#s

Data columns to use: 1,2,3,4,5,6,7

Select type of values: Absolute

Assign a label to each timepoint (optional)

Choose color scheme: Color range from data (blue-red)

Save Color Scheme Parameters Retrieve Saved Color Scheme Parameters

OK Cancel

Specify Timepoint Labels

Timepoint labels can be specified either here or in the data file. To include them in the data file, make sure the first line (excluding comment lines or blank lines) starts with a "\$" character. The values in each column of this first line will then be treated as labels. If labels are specified both in the data file and here, the labels entered here will take precedence.

Enter label for each timepoint column:

1: 1_upper_euphotic_10m

2: 6_upper_euphotic_70m

3: 2_chlorophyllmax_130m

4: 3_below_euphotic_200m

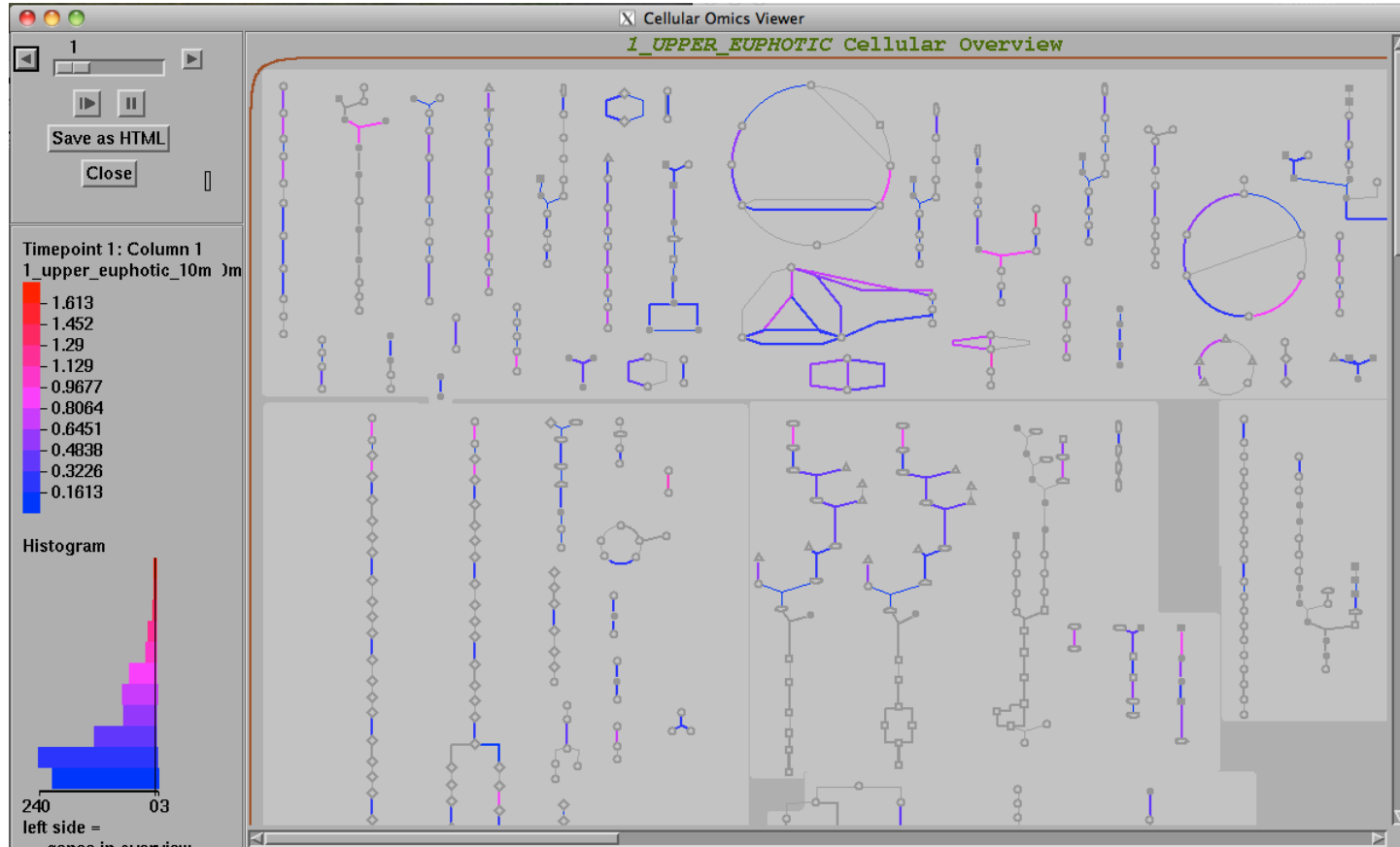
5: 5_uppermesopelagic_500m

6: 7_omz_770m

7: 4_deepabyss_4000m

OK Cancel

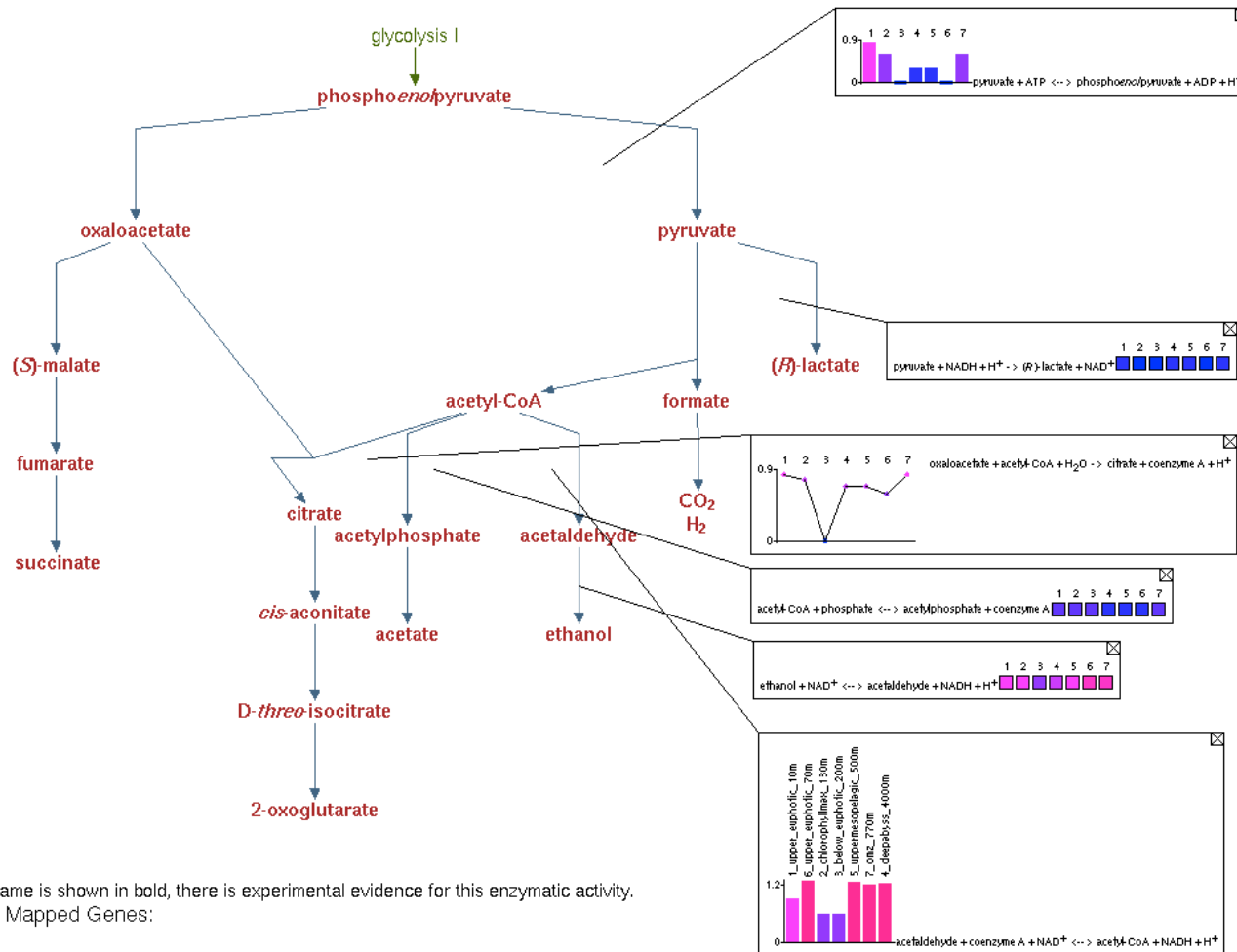
6. 'Omics Tools' feature for Metadata



6. 'Omics Tools' feature for Metadata

1_UPPER_EUPHOTIC Pathway: mixed acid fermentation

More Detail Less Detail Species Comparison



If an enzyme name is shown in bold, there is experimental evidence for this enzymatic activity.
Locations of Mapped Genes:



Questions?

