Constructing a Yeast Interactome

The yeast interactome used in the current work was constructed in April 2008 using protein-protein and genetic (gene-gene) interaction data downloaded from the *Saccharomyces* Genome Database (SGD; http://www.yeastgenome.org/), and protein-DNA interaction data downloaded from the YEASTRACT transcription factor database (http://yeastract.com/). These data were formatted as a tab-delimited file for use in the data analysis and visualization platform, Cytoscape (http://www.cytoscape.org/).[1,2] These steps are summarized here, then described in detail.

A. Format the SGD interaction data file:

1. Delete unnecessary text (e.g. "Bait" and "Hit") and columns

2. Simplify the interaction types as physical or genetic interactions

3. Count and delete duplicate interactions; assign the numbers of duplicates as a column of interaction weights

B. Format the YEASTRACT (transcription factors) interaction file:

1. Convert common yeast gene names (YEASTRACT) to systematic names (SGD), a requirement for Cytoscape

2. Change the case of all letters from lowercase to uppercase (also required by Cytoscape)

3. Process the list of genes from YEASTRACT through the "Batch Download" tool at SGD, to identify "troublesome" genes (e.g., MAL63 is not in the systematic sequence of SGD reference strain S288c; common names assigned to two or more genes)

4. Add a column of weights (all equal to 1), for compatibility with the SGD interaction data

C. Concatenate the SGD and YEASTRACT interactions files

D. Import the interactome into Cytoscape; import the annotation files; assign visual display properties for the nodes (genes; proteins) and edges (interaction types), using the Cytoscape VizMapper tool

The computer used for this work employed an Intel Pentium 4 CPU operating at 3.0 GHz, 1.5 GB of RAM, and the Microsoft Windows XP Professional Version 2002 Service Pack 2 operating system. To facilitate these manipulation in Steps A-C, above, simple perl and awk

---

[1] Shannon, P. *et al.* (2003) "Cytoscape: A software environment for integrated models of biomolecular interaction networks." Genome Res. **13:** 2498-2504.

[2] See also "Introduction to Cytoscape" (http://www.cytoscape.org/features2.pphp).

scripts were employed using Cygwin (http://www.cygwin.com/), a Linux-like environment for Windows [GNU bash shell, version 3.2.33(18)-release (i686-pc-cygwin]. These commands can be entered directly at a command prompt ($) in the Macintosh and Linux operating systems.

Detailed Construction Notes

A. SGD Interaction Data

An interaction data file (interactions.tab[3]; 23,338 kb; Mar-31-2008) was downloaded from SGD, the first three lines of which are shown here:

Synthetic Lethality YFL039C (Bait)|YBR243C (Hit) inviable Davierwala AP, et al. (2005) The synthetic genetic interaction spectrum of essential genes. Nat Genet 37(10):1147-52 16155567 BioGRID
Synthetic Lethality YFL039C (Bait)|YKL052C (Hit) inviable Davierwala AP, et al. (2005) The synthetic genetic interaction spectrum of essential genes. Nat Genet 37(10):1147-52 16155567 BioGRID
Synthetic Lethality YFL039C (Bait)|YPR105C (Hit) inviable Davierwala AP, et al. (2005) The synthetic genetic interaction spectrum of essential genes. Nat Genet 37(10):1147-52 16155567 BioGRID

This file contains interaction data for various genes and proteins distributed among a number of sub-categories that are described in a glossary file[4] available at the SGD website and below, in this supplementary file**.**

The total number of lines contained in the interactions.tab file was determined using the awk programming language command:

Command:

$ awk 'END{print NR}' interactions.tab

Output:

96347

Unnecessary text (e.g. "(Bait)|" and " (Hit)") was deleted using a perl-based search and replace command:

Command:

[3] ftp://genome-ftp.stanford.edu/pub/yeast/data_download/literature_curation/interactions.tab
[4] http://www.yeastgenome.org/help/glossary.html

```
$ perl -p -i.bak -e 's/\Q (Bait)|\E/\t/g' interactions.tab && perl -p -i.bak2 -e 's/\Q (Hit)\E//g'
  interactions.tab && rm -f *.bak2
```

Output (first three lines shown):

```
Synthetic Lethality     YFL039C  YBR243C     inviable        Davierwala AP, et al. (2005)
    The synthetic genetic interaction spectrum of essential genes. Nat Genet 37(10):1147-52
    16155567  BioGRID
Synthetic Lethality     YFL039C  YKL052C     inviable        Davierwala AP, et al. (2005)
    The synthetic genetic interaction spectrum of essential genes. Nat Genet 37(10):1147-52
    16155567  BioGRID
Synthetic Lethality     YFL039C  YPR105C     inviable        Davierwala AP, et al. (2005)
    The synthetic genetic interaction spectrum of essential genes. Nat Genet 37(10):1147-52
    16155567  BioGRID
```

The numerous interaction types provided by SGD results in nodes being connected by multiple edges, reflecting multiple interaction associations between two genes or proteins. To simplify the display of the interactome in Cytoscape, using the glossary provided by SGD, these interactions were simplified as being either “pp” (physical) or “gi” (genetic) interactions, using the following “SGD_interactions_search_replace.awk” script (file):

```
# START OF SCRIPT
# This awk (gawk) script searches and replaces the SGD interaction labels
# with Cytoscape interaction types (pp; gi).  For the SGD glossary of terms, see
# http://www.yeastgenome.org/help/glossary.html#DosageLethality
# The following script recodes these as follows:
#
# PHYSICAL INTERACTIONS (pp):
#
# Affinity Capture-MS
# Affinity Capture-RNA
# Affinity Capture-Western
# Biochemical Activity
# Co-crystal Structure
# Co-fractionation
# Co-localization
# Co-purification
# Far Western
# FRET
# Protein-peptide
# Protein-RNA
# Reconstituted Complex
# Two-hybrid
```

```
#
# GENETIC INTERACTIONS (gi):
#
# Dosage Growth Defect
# Dosage Lethality
# Dosage Rescue
# Epistatic MiniArray Profile
# Phenotypic Enhancement
# Phenotypic Suppression
# Synthetic Growth Defect
# Synthetic Lethality
# Synthetic Rescue
#
# This script is implemented as follows:
# $ awk -f <awk program file name> <source file>
# e.g.,
# $ awk -f SGD_interactions_search_replace.awk interactions.tab
#
# SCRIPT:
#
# PHYSICAL INTERACTIONS (pp):
#
{
sub(/Affinity Capture-MS/, "pp");
sub(/Affinity Capture-RNA/, "gi");
sub(/Affinity Capture-Western/, "gi");
sub(/Affinity Capture-RNA/, "pp");
sub(/Affinity Capture-Western/, "pp");
sub(/Biochemical Activity/, "pp");
sub(/Co-crystal Structure/, "pp");
sub(/Co-fractionation/, "pp");
sub(/Co-localization/, "pp");
sub(/Co-purification/, "pp");
sub(/Far Western/, "pp");
sub(/FRET/, "pp");
sub(/Protein-peptide/, "pp");
sub(/Protein-RNA/, "pp");
sub(/Reconstituted Complex/, "pp");
sub(/Two-hybrid/, "pp");
#
# GENETIC INTERACTIONS (gi):
#
sub(/Dosage Growth Defect/, "gi");
sub(/Dosage Lethality/, "gi");
```

```
sub(/Dosage Rescue/, "gi");
sub(/Epistatic MiniArray Profile/, "gi");
sub(/Phenotypic Enhancement/, "gi");
sub(/Phenotypic Suppression/, "gi");
sub(/Synthetic Growth Defect/, "gi");
sub(/Synthetic Lethality/, "gi");
sub(/Synthetic Rescue/, "gi");
print > "SGD_interactions_Mar-31-2008_for_Cytoscape.tab"
}
#
# END OF SCRIPT
```

The print command at the end of this script saves the output as a tab-delimited text (.tab) file, "SGD_interactions_Mar-31-2008_for_Cytoscape.tab." This awk script was implemented as follows:

Command:

```
$ awk -f SGD_interactions_search_replace.awk interactions.tab
```

Output (first three lines shown):

```
gi   YFL039C  YBR243C      inviable        Davierwala AP, et al. (2005)  The synthetic genetic
  interaction spectrum of essential genes. Nat Genet 37(10):1147-52    16155567  BioGRID
gi   YFL039C  YKL052C      inviable        Davierwala AP, et al. (2005)   The synthetic genetic
  interaction spectrum of essential genes. Nat Genet 37(10):1147-52    16155567  BioGRID
gi   YFL039C  YPR105C      inviable        Davierwala AP, et al. (2005)  The synthetic genetic
  interaction spectrum of essential genes. Nat Genet 37(10):1147-52    16155567  BioGRID
```

The following script:

Command:

```
$ awk 'BEGIN {OFS=FS="\t"} $3 < $1 { t = $1; $1 = $3; $3 = t } { a[$1 "\t" $2 "\t" $3]++ }
  END { for (k in a) { print a[k], k } } ' SGD_interactions_Mar-31-2008_basic.tab | sort >
  pp_gi.tab
```

- Extracts the second ("bait" i.e. source node), first (interaction type) and third ("hit" i.e. target node) columns
- Looks for duplicates - e.g. "Node_A pp Node_B" and "Node_B pp Node_A" results in two edges of the same type (pp) being displayed between Node_A and Node_B in Cytoscape
- Counts the duplicates (for use as interaction weights)

- Prints the data to an output file containing the weight, source node (gene / protein), interaction type, and target node (gene / protein)
- Sorts the data
- Generates the output file pp_gi.tab

The following command allows you to quickly view (verify) the contents of this file:

Command:

$ cat pp_gi.tab

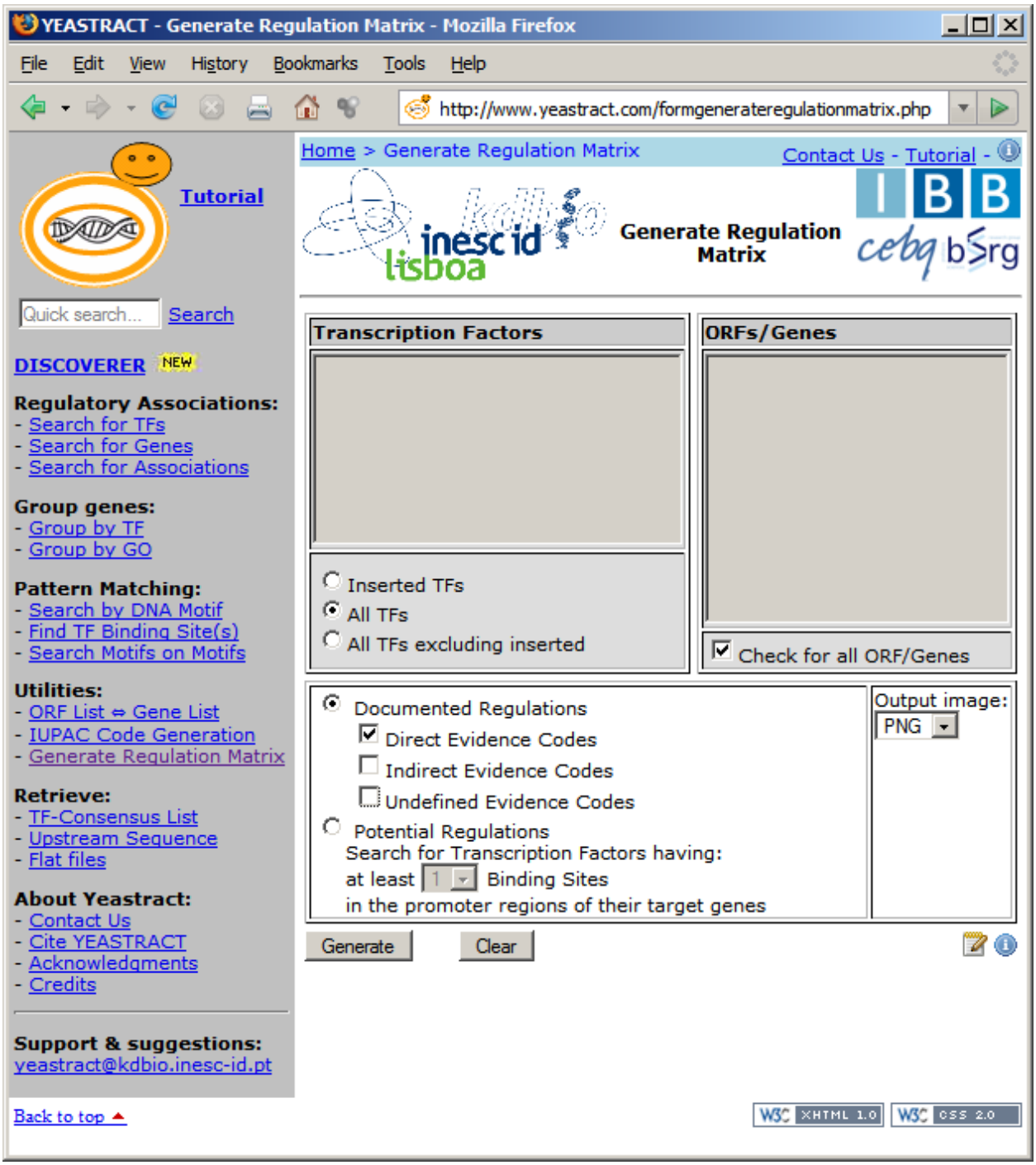
Output (first three and last 13 lines shown):

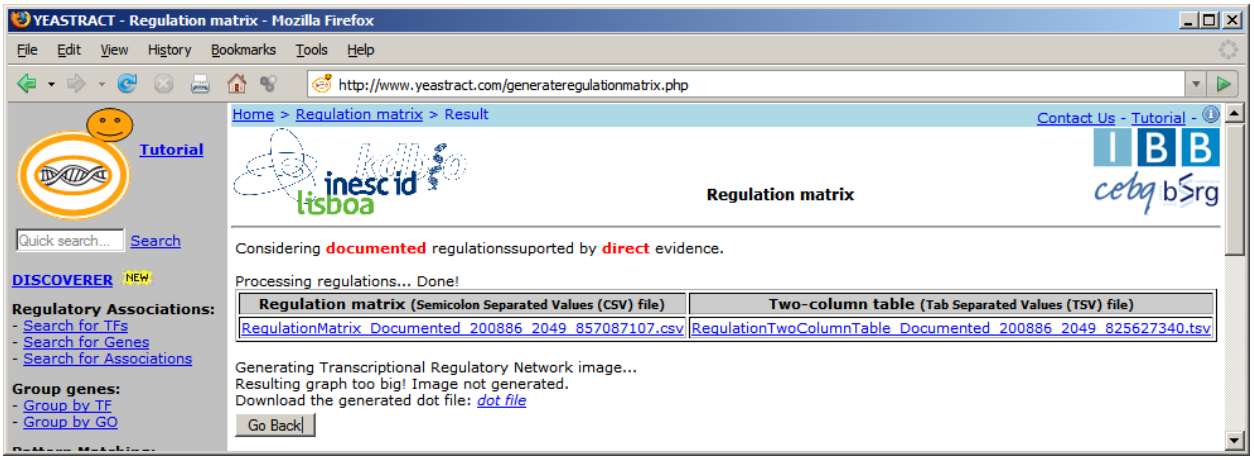
1	Q0045	gi	YDL044C
1	Q0045	gi	YER154W
1	Q0045	gi	YGL064C
etc.			
7	YLR234W	gi	YMR190C
7	YMR167W	pp	YNL082W
8	YDR129C	gi	YFL039C
8	YDR201W	pp	YKR037C
8	YDR217C	pp	YPL153C
8	YFL008W	pp	YJL074C
8	YKR081C	pp	YOR294W
8	YMR146C	pp	YOR361C
8	YMR224C	pp	YNL250W
8	YNL098C	gi	YOR101W
8	YNL262W	pp	YPR175W
9	YDR092W	pp	YGL087C
9	YJL092W	gi	YMR190C

Each line, above, shows the interaction weight, the source node, the interaction type, and the target node.

B. YEASTRACT Interaction Data

Documented interactions between 168 *S. cerevisiae* transcription factors (March 2008) and their target genes were downloaded from the YEASTRACT transcription factor database, as shown in the following screen captures.


YEASTRACT - Generate Regulation Matrix - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://www.yeastract.com/formgenerateregulationmatrix.php
Tutorial
Quick search... Search
DISCOVERER NEW
Regulatory Associations:
- Search for TFs
- Search for Genes
- Search for Associations
Group genes:
- Group by TF
- Group by GO
Pattern Matching:
- Search by DNA Motif
- Find TF Binding Site(s)
- Search Motifs on Motifs
Utilities:
- ORF List ⇔ Gene List
- IUPAC Code Generation
- Generate Regulation Matrix
Retrieve:
- TF-Consensus List
- Upstream Sequence
- Flat files
About Yeastract:
- Contact Us
- Cite YEASTRACT
- Acknowledgments
- Credits
Support & suggestions:
yeastract@kdbio.inesc-id.pt
Back to top
Home > Generate Regulation Matrix
Contact Us - Tutorial -
Generate Regulation Matrix
Transcription Factors
ORFs/Genes
Inserted TFs
All TFs
All TFs excluding inserted
Check for all ORF/Genes
Documented Regulations
Direct Evidence Codes
Indirect Evidence Codes
Undefined Evidence Codes
Potential Regulations
Search for Transcription Factors having:
at least 1 Binding Sites
in the promoter regions of their target genes
Output image:
PNG
Generate
Clear
W3C XHTML 1.0
W3C CSS 2.0



This file, "RegulationTwoColumnTable_Documented_2008410_1839_1043605408.tsv" consisted of rows of transcription factors and their target genes, illustrated here using selected rows:

Abf1 YKL112w
Abf1 YAL054c
Abf1 YGL234w
etc.
Ace2 YKL150w
Ace2 YNL328c
etc.
Cup9 YDR441c
Cup9 YDR442w
Cup9 YEL040w
etc.

Cytoscape requires uppercase systematic gene names, e.g. YKL112W rather than abf1, and YKL112W rather than YKL112w. Prior to formatting this list of transcription factors and documented target genes for use in Cytoscape, each of the columns of genes from YEASTRACT was passed through the "Batch Download" tool at SGD[5] to identify ORFs that are no longer identified as genes, genes not present in the standard sequence of *S. cerevisiae* strain S288c (the reference, sequenced strain curated at SGD), and nomenclature conflicts due to common gene names being associated with two or more systematic gene names.

These genes were corrected as needed (renamed or deleted). For example, the common gene names provided by YEASTRACT for the the transcription factors were converted to their systematic names using a Microsoft Excel "lookup table" (common_to_systematic.tab) created by one of the authors (GRS). As well, genes not present on the Agilent G4140A Yeast Oligo Microarray were deleted. Columns representing the interaction type ("tf," for transcription factor) and the interaction weights (all set to a weight of 1) were added for compatibility with the pp-gi file constructed from the SGD interaction data. Each of the tf weights were set to "1" since

[5] http://db.yeastgenome.org/cgi-bin/batchDownload

there is a 1:1 relationship between a transcription factor and any single target gene. Lastly, as for the pp-gi data, these tf data were tab-delimited, in the final output file.

The operations described in the two preceding paragraphs were accomplished using the following script, generating the tab-delimited output text file, tf.tab:

Command:

```
$ awk 'FNR==NR{a[tolower($1)]=$2;next}tolower($1) in a{print "1\t" a[tolower($1)] "\ttf\t"
  toupper($2)}' "common_to_systematic.tab"
  "RegulationTwoColumnTable_Documented_2008410_1839_1043605408_parsed.tsv" | sort
  > tf.tab
```

The first three lines of tab.tf are shown here:

Command:

```
$ cat tf.tab
```

Output:

```
1       YAL051W     tf      YAL016W
1       YAL051W     tf      YAL034W-A
1       YAL051W     tf      YAL035W
```

C. Combined (SGD + YEASTRACT: pp-gi-tf) Yeast Interactome

To generate the yeast interactome used in this study (Stuart et al. 2009), the interaction files described above in Sections A (SGD) and B (YEASTRACT) were combined and saved as a single tab-delimited file (interactome), containing four columns, as follows:

Command:

```
$ cat pp_gi.tab tf.tab > pp_gi_tf.tab
```

Selected lines from the resulting output file, pp_gi_tf.tab, are shown here:

Command:

```
$ cat pp_gi_tf.tab
```

Output (selected lines shown, including the junction between the pp_gi and tf source files):

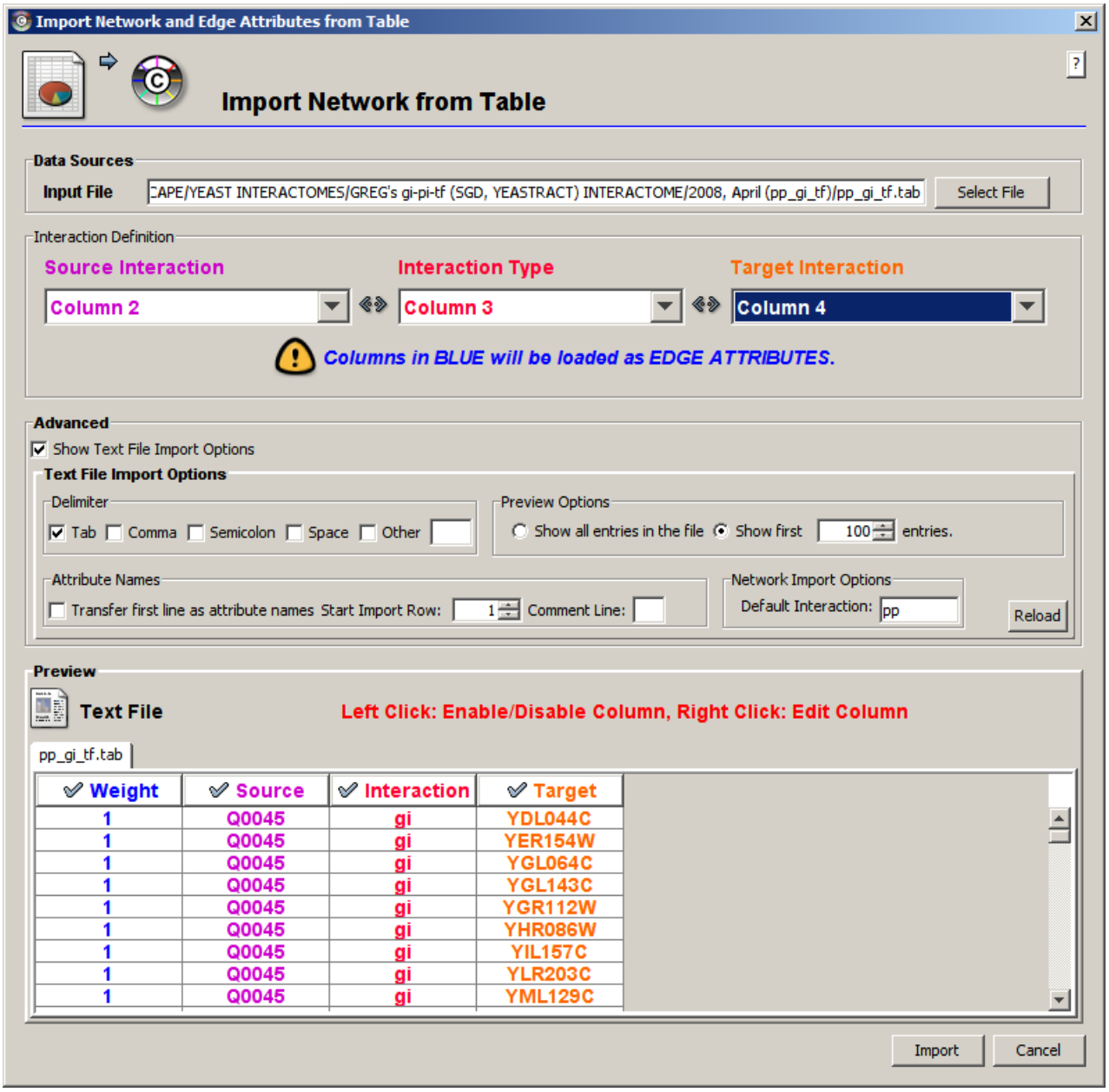
1	Q0045	gi	YDL044C
1	Q0045	gi	YER154W
1	Q0045	gi	YGL064C
etc.			
8	YNL098C	gi	YOR101W
8	YNL262W	pp	YPR175W
9	YDR092W	pp	YGL087C
9	YJL092W	gi	YMR190C
1	YAL051W	tf	YAL016W
1	YAL051W	tf	YAL034W-A
1	YAL051W	tf	YAL035W
etc.			
1	YPR199C	tf	YPR200C
1	YPR199C	tf	YPR200C
1	YPR199C	tf	YPR201W
1	YPR199C	tf	YPR201W

D. Importing the Yeast Interactome, Data and Annotation Files Into Cytoscape; Visual Display Properties

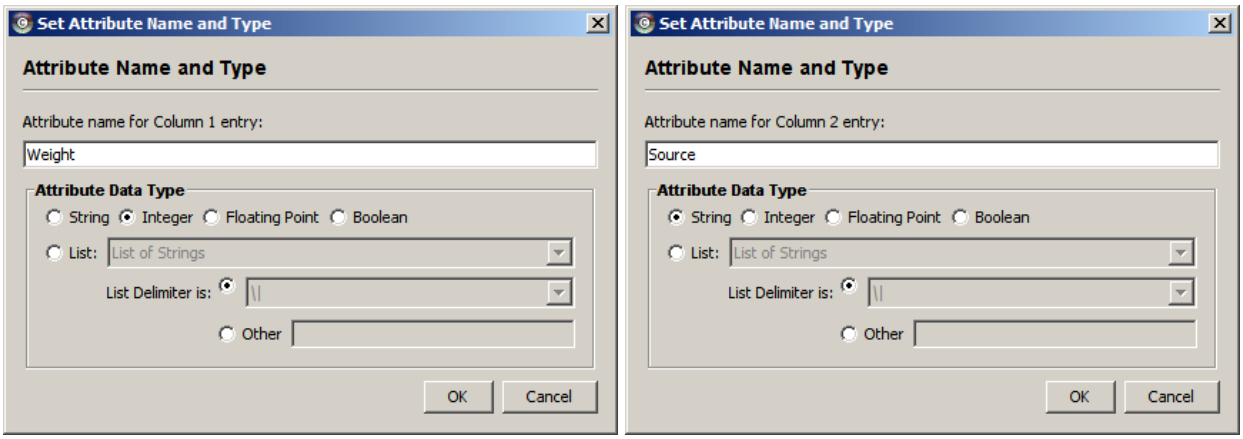
The following steps were done using Cytoscape version 2.6.0. From the ***File*** menu, select the pp_gi_tf.tab interaction file, created above:

File > Import > Network from Table (Text/MS Excel)... > Select File > pp_gi_tf.tab

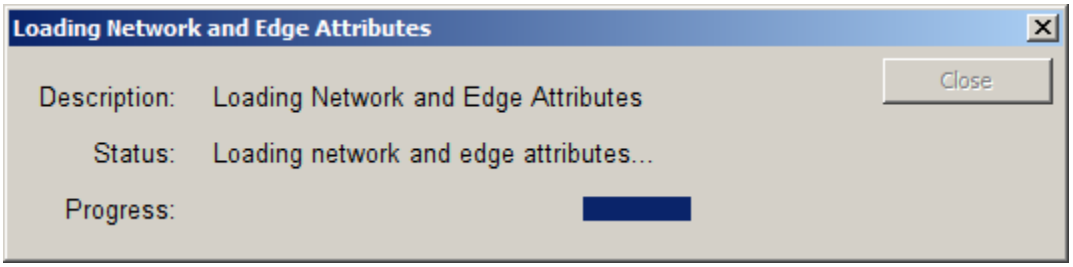
Next, select the options shown below:



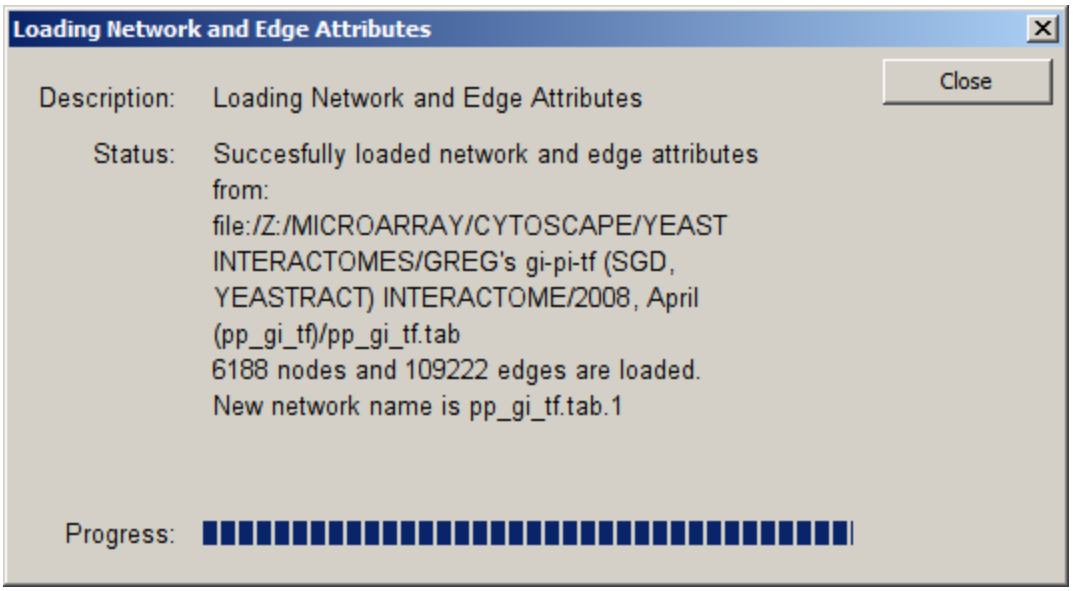
Edit the column headings (Weight; Source; Interaction; Target) by right-clicking each of the column headings:



Returning to the "Import Network and Edge Attributes from Table" window, click ***Import*** to import the interaction data file into Cytoscape.

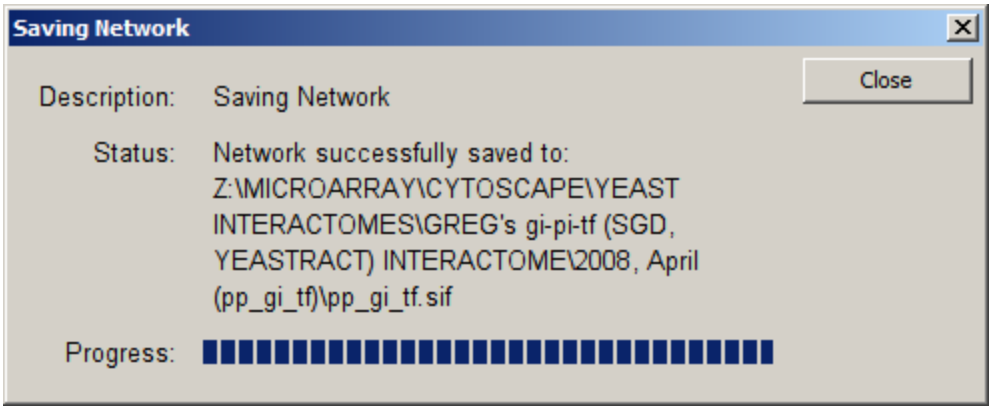


After a few seconds, the following window should appear:



This new network can now be saved as a SIF (Simple Interaction Format) file, using the ***File*** menu in Cytoscape:

File > Export > Network as SIF File... > pp_gi_tf.sif > ***Save***



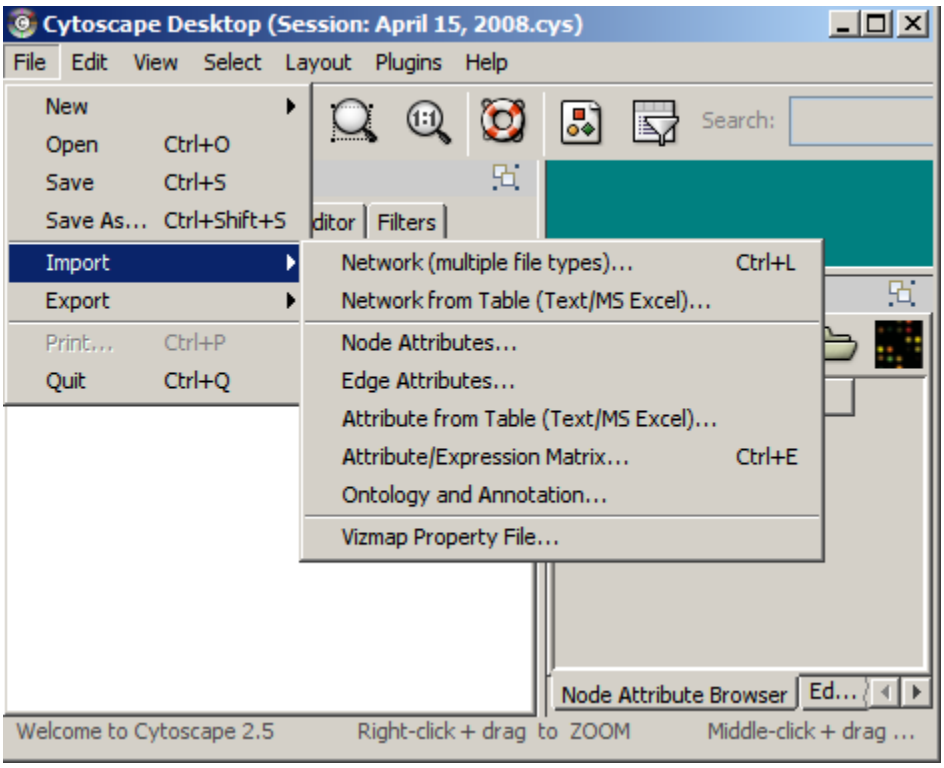
Next, the microarray expression data can be imported, as well as various data attribute (annotation) files, as needed. For example, to load our yeast diauxic shift expression data into Cytoscape, we went to the File menu,

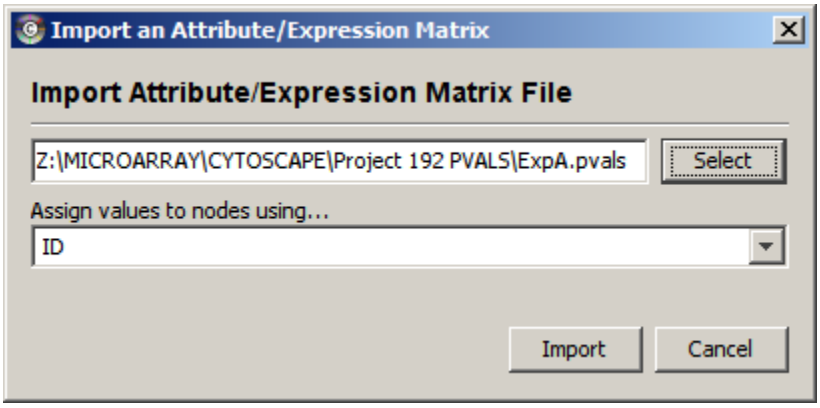
File > Import > Attribute/Expression Matrix...

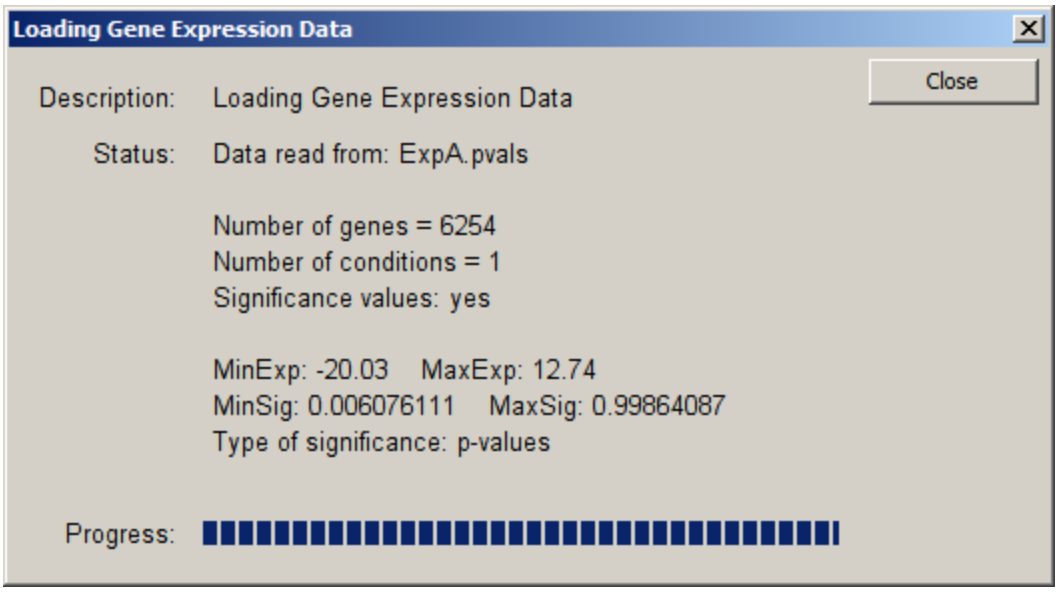
and selected our "Experiment_F_WT_YPG_Shift.pvals" file, a tab-delimited text file containing our microarray data (tabulated as the systematic gene name, the common gene name(s), the fold-change and the p-value, respectively, on each line). The first five lines of this file are shown, here:

GENE	COMMON	EXP_F	EXP_F
YPR121W	THI22	-1.23	0.80547
YLR281C	YLR281C	-1.27	0.73793
YCR020W-B	HTL1	-1.18	0.90121
YIL002C	INP51,SJL1	-1.05	0.91376

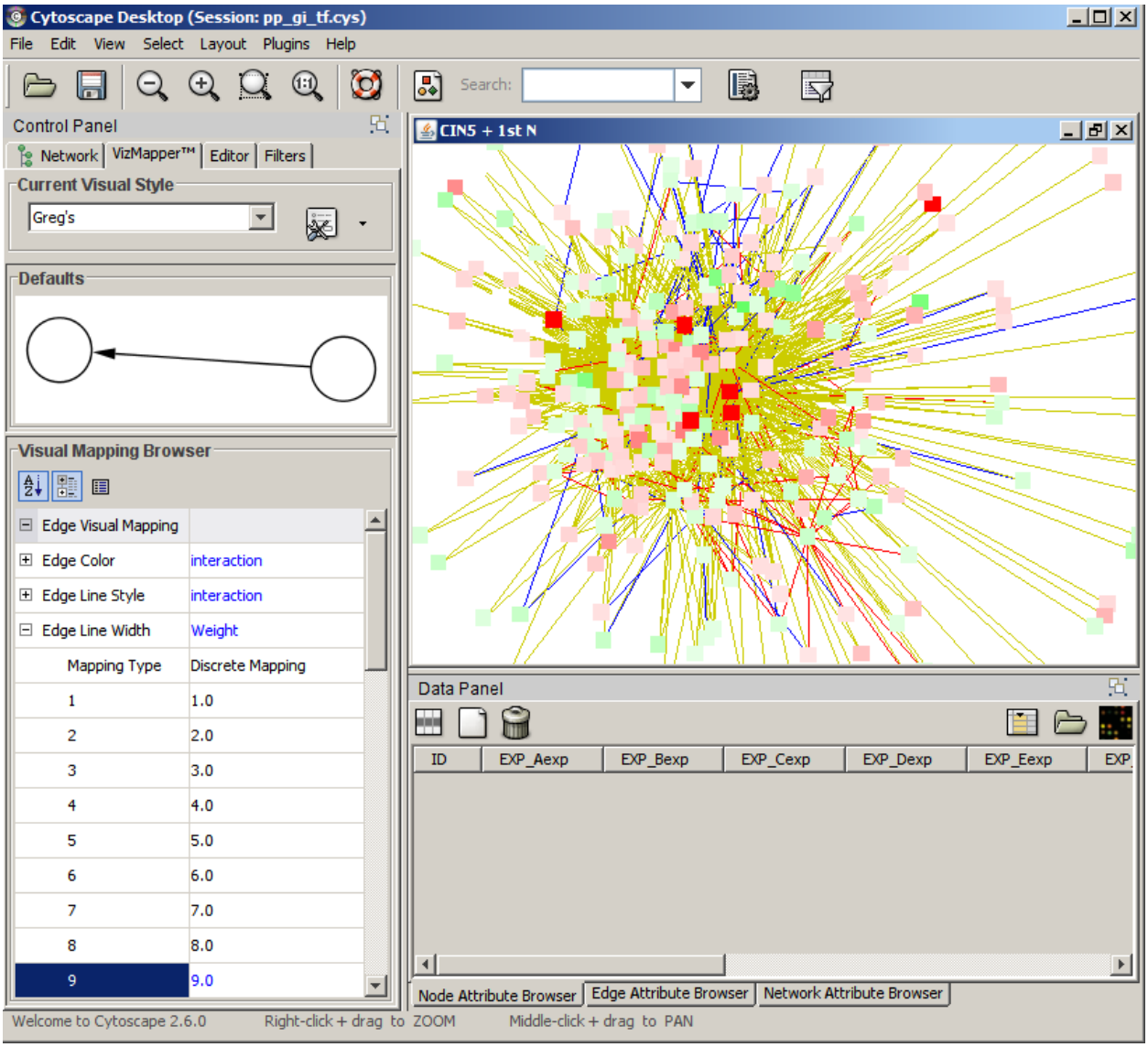
Node attributes, properties associated with the nodes displayed in Cytoscape, are user-defined. Annotations embedded in the interactome included in our companion paper (Stuart *et al.* 2009) include lists of genes associated with the response to stress, mitochondrial-associated genes, the common gene names associated with each of the systematic gene names, and descriptions of each gene as provided by SGD. These attributes are loaded as shown in the following screen captures:

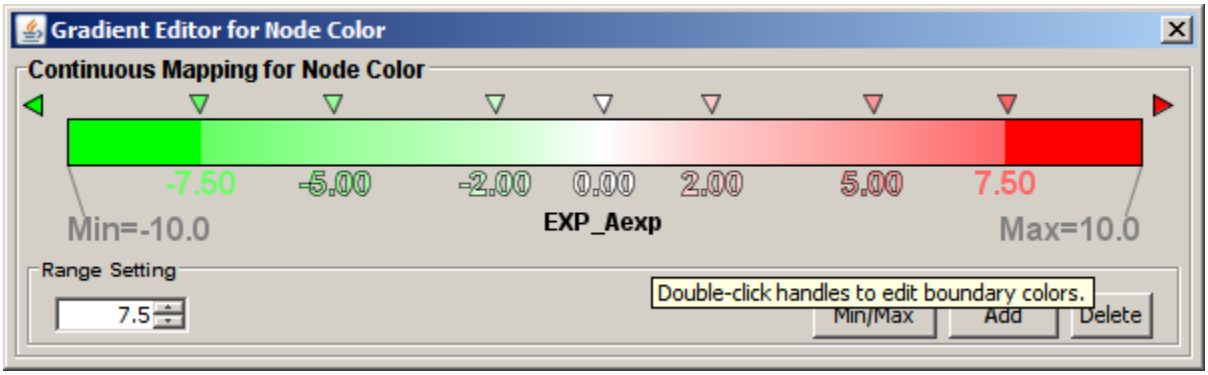

Cytoscape Desktop (Session: April 15, 2008.cys)
File
Edit
View
Select
Layout
Plugins
Help
New
Open Ctrl+O
Save Ctrl+S
Save As... Ctrl+Shift+S
Import
Export
Print... Ctrl+P
Quit Ctrl+Q
Search:
ditor
Filters
Network (multiple file types)... Ctrl+L
Network from Table (Text/MS Excel)...
Node Attributes...
Edge Attributes...
Attribute from Table (Text/MS Excel)...
Attribute/Expression Matrix... Ctrl+E
Ontology and Annotation...
Vizmap Property File...
Node Attribute Browser
Ed...
Welcome to Cytoscape 2.5
Right-click + drag to ZOOM
Middle-click + drag ...


Import an Attribute/Expression Matrix
Import Attribute/Expression Matrix File
Z:\MICROARRAY\CYTOSCAPE\Project 192 PVALS\ExpA.pvals
Select
Assign values to nodes using...
ID
Import
Cancel


Loading Gene Expression Data
Close
Description: Loading Gene Expression Data
Status: Data read from: ExpA.pvals
Number of genes = 6254
Number of conditions = 1
Significance values: yes
MinExp: -20.03 MaxExp: 12.74
MinSig: 0.006076111 MaxSig: 0.99864087
Type of significance: p-values
Progress:

Recent versions of Cytoscape (starting with version 2.5) include a visual display properties editor, “VizMapper,” that allows the display properties to be customized by the user:

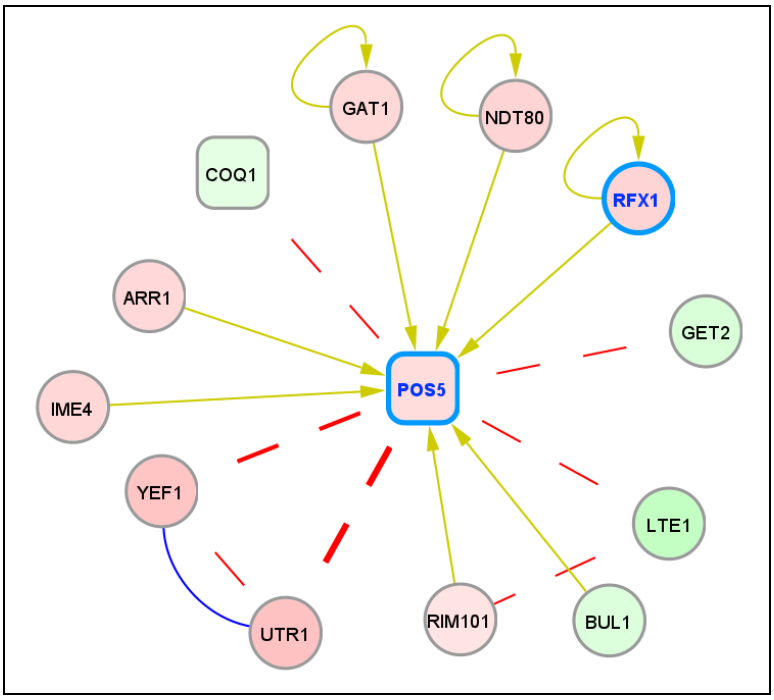




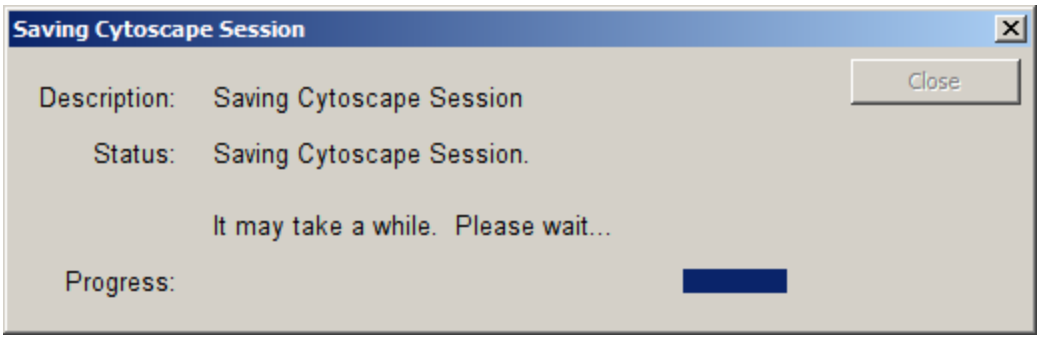
For the current interactome / diauxic shift experiment, the following display properties were define, using VizMapper:

- Edges were displayed according to their interaction type:
  - gold arrows for transcription factors (tf), pointing from the transcription factor to the regulated gene
  - blue lines for physical (pp: protein-protein) interactions
  - broken red lines for genetic interactions (gi; e.g. synthetic lethality)
- Mitochondrial nodes (genes) were displayed as rectangles
- Stress-response genes were displayed as nodes with exaggerated (emphasized) blue borders

These display properties are illustrated in the following screen capture from Cytoscape showing the *POS5* gene plus it's directly-associated, "first-neighbor" nodes:



For convenience, the interactome, imported expression data, attributes, display properties and various analyses can be saved as a Cytoscape session (.cys) file from the ***File*** menu > ***Save*** or ***Save As...*** :


Saving Cytoscape Session
Close
Description: Saving Cytoscape Session
Status: Saving Cytoscape Session.
It may take a while. Please wait...
Progress: