

Gene Database Construction and GenMAPP Analysis Conducted on *Sinorhizobium meliloti* (Strain 1021) Microarray Data



(Photo Credit: Wishart Lab - BacMap Genome Atlas)

Stephen Louie, Lauren Magee, Miles Malefyt, and Mitchell Petredis

BIOL/CMSI 367

Biological Databases

Decemeber 13, 2013

## Introduction

*Sinorhizobium meliloti* is a gram-negative rhizobacteria that is the symbiont of the legume alfalfa. As a plant growth promoting rhizobacteria, *S. meliloti* is able to induce nodule growth within the roots of its symbiont. It is within these nodules that the process of nitrogen fixation takes place. This process is particularly vital for plants that happen to be situated within Nitrogen-deficient soils. Given the relative importance of this process, *S. meliloti* acts an ideal model organism in which to study the mechanisms of this process. The DNA microarray experiment performed by Ana Domínguez-Ferreras and her team specifically focused studying the transcriptional responses of *S. meliloti* when placed under high salinity and sucrose stress. This serves to help identify the gene that is responsible for hyperosmotic stress.

We'd like to analyze the data from the DNA microarray experiment performed by Ana Domínguez-Ferreras and her team, but it was not possible to do so since no gene database exists for *S. meliloti*. Until now: thanks to the free resources provided by XMLPipeDB, we are able to apply their GenMAPP Builder (gmbuilder) program to create a dense, custom database for our species of interest in conjunction with pgAdmin III. Our end goal is to create a fully functional database and custom version of gmbuilder for *S. meliloti* and release it freely to the public for all to utilize and make contributions to; additionally, we want to discover new information about the microarray data by the Domínguez-Ferreras team via the GenMAPP program. The microarray data took the txt files that were offered up in the experiment page and showed a lot of evidence for their findings in the scientific paper. In addition to supporting their scientific data, it also shed light on some unmentioned cellular processes that were not included in their paper but may still be relevant to how the cell responds to the sucrose or salt stimuli.

## Methods

1. Download the following files (All files can be found on the [[Team Name]] wiki page at this url under the “Important Files” sections. The only different files per section are the .gdb files:

[https://xmlpipedb.cs.lmu.edu/biodb/fall2013/index.php/Team\\_Name](https://xmlpipedb.cs.lmu.edu/biodb/fall2013/index.php/Team_Name)).

- a. UniProt XML proteome set for *S. meliloti*
  - i. Go to [UniProt Complete Proteomes](#)
  - ii. Find the link for "[List all Bacteria](#)"
  - iii. Do an advanced search, with filters “AND,” “Field: ALL,” and term “Sinorhizobium meliloti 1021”
    1. You should only see one result (at the time of this writing)
    2. If there’s more than one result, then use this link: [Rhizobium meliloti \(strain 1021\) \(Ensifer meliloti\) \(Sinorhizobium meliloti\)](#)
  - iv. At the top right corner of the page, an orange “Download” box should appear. Click on it.
  - v. Download the XML file. There may be newer versions available, but the one we used for this project is listed below.
    1. [Sinorhizobium meliloti 1021 mpetredi 2013115 UniProt XML.xml](#)
    2. Version: UniProt release 2013\_10
- b. GO OBO-XML
  - i. Go to <http://beta.geneontology.org/page/download-ontology>.
  - ii. Look under “Legacy Downloads” for OBO-XML, and download [obo-xml.gz](#). You’ll need a program like 7-zip to properly extract the .gz file.

- iii. After extracting the file, keep the go\_daily-termdb.obo-xml.
- iv. Here are the files we used at the time of the project.

- 1. [Go daily-termdb.obo-xml mpetredi 2013116.gz](#)
- 2. Version: From 11/4/2013

c. GOA

- i. Going to <ftp://ftp.ebi.ac.uk/pub/databases/GO/goa/> in Google Chrome results in an error message.
- ii. Changed url from "ftp" to "http" at beginning, like this:  
<http://ftp.ebi.ac.uk/pub/databases/GO/goa/>
- iii. You'll be taken to "Index of/pub/database/GO/goa." Click on "proteomes" folder.
- iv. You'll be directed to "Index of/pub/databases/GO/goa/proteomes." Download 58.R\_meliloti.goa.
  - 1. Note: R. meliloti is an alternative name to S. Melitoti.
- v. Here are the files we used at the time of the project.

- 1. [R meliloti.goa](#)
- 2. Version: From 11/7/2013

2. Create a new database using pgAdmin III and gmbuilder

Importing data into the pgAdmin III database and exporting gene database files (.gdb) involves the free, open source program called GenMAPP Builder (gmbuilder). First, we had to import the .sql file within the .zip package of gmbuilder into our PostgreSQL database to create the data tables needed for the gmbuilder to work properly, and do not close pgAdmin III; this .sql file remained unchanged throughout the entire project and with every new release of gmbuilder. Once the data tables were loaded, we launched the gmbuilder-32bit.bat program and began importing the UniProt XML, GO OBO-XML, and GOA files one at a time. When all three files imported, we started the first .gdb export

for analysis by our quality assurance coordinator and GenMAPP users. The speed of the whole process depends on your computer hardware, so ideally you'll want to use the most upgraded computer available. If any errors occurred, the quality assurance coordinator would identify the errors and forward them to the team coder, who would make the necessary adjustments to gmbuilder after every export.

After our database was established, we began looking at microarray data from the following article: "Transcriptome Profiling Reveals the Importance of Plasmid pSymB for Osmoadaptation of *Sinorhizobium meliloti*", which we found online in the Journal of Bacteriology. The article used DNA microarrays to examine gene expression in response to various abiotic stresses. These stresses included adding .3 M NaCl, .4 M NaCl, .5 M sucrose, and .7 M sucrose to cultures of *Sinorhizobium meliloti*. After each was added their separate cultures, the researchers observed the number of genes still present on the microarray chip at four different time intervals. The times were 15 minutes, 30 minutes, 60 minutes, and 240 minutes. The researchers thought it beneficial to use these times, because they were observing whether the bacterium would adapt to the new stressor over the course of its exposure. In addition, since there are three replicants included in the *Sinorhizobium meliloti* genome, each replicant genome was cultured and tested individually with all four of the stressors, at the four time intervals. This produces a lot of data, so we decided to only focus on the .7 M sucrose stressor and the .3 M NaCl stressor.

We downloaded the raw data produced from our microarray article's .3M NaCl and .7M sucrose experiments on the ArrayExpress website. Then we extracted the Cy5 Signal, Cy5 Background, Cy3 Signal, and Cy3 Background median values and placed them into a new worksheet. Then we calculated the  $\log\left(\frac{\text{Cy5 Signal} - \text{Cy5 background}}{\text{Cy3 Signal} - \text{Cy3 background}}\right)$  for every replicate (1-3) corresponding with the four different times offered in the experiment (15, 30, 60, 240).

The log values were carried over to a new spreadsheet with the Gene IDs as well, and we our began scaling and centering the data. Each column of this spreadsheet represented one of the three replicants at one of the four time intervals. Each column was averaged, and the standard deviation was also calculated. Each value in the column was subtracted by this column average, and divided by the column standard deviation. This was done to all of the column.

Statistical Analysis was then performed on this data and four average log fold change values were calculated. These values, were the average of all three replicants at the same time interval. Therefore, there was a average log full change value for all the data at the 15 minute mark, the 30 minute mark, the 60 minute mark, and the 240 minute mark. Based off of these averages, a T stat was calculated and then a P-value for each of the column values (see tables in the results section).

A file was formatted based off the specific GenMAPP qualifications and then the file was loaded into GenMAPP using the “Expression Dataset” function with the gene database previously created. Color sets were created for those with log fold change values less than -0.25 or more than 0.25, both with a p-value less than 0.05. Those greater than .25 were marked by the color green and labeled “increased”. Those less than -0.25 were marked by the color red. Then the GenMAPP data was run in the MAPPFinder program and gene ontology results were produced and studied. From these results, a biological pathway was studied and noted in the discussion section of this paper. Excel files created from MAPPFinder were also examined and filters were created to determine how many results had a Z score greater than 2, a PermuteP less than 0.05, a Number Change greater than or equal to 5 and less than 100, and a Percent Change greater than or equal to 25.

## **Results**

Our initial import/export cycle for *S. meliloti* involved [gmbuilder-2.0b71](#), which was originally

designed for the species *Vibrio cholerae*. On one of our slower HP lab computers, the times were as follows: UniProt XML: 19.17 minutes, GO OBO-XML: 17.81 minutes to import and 15.54 minutes to process, GOA: 0.21 minutes, .gdb: around 8 hours. According to Microsoft Access and Tally Engine (part of gmbuilder), the number of OrderedLocusNames matched with 4712 records.

Since `gmbuilder-2.0b71` was not designed for *S. meliloti*, we had to make changes within its code so that it may give us better readings for our species. For this, we used Eclipse and SourceForge to access and checkout the code for gmbuilder for further modification and creation of custom gmbuilder versions. The first set of changes to the code involved creating and customizing the species profile for *S. meliloti* as well as adding the species profile to the catalog of known species profiles; the instructions to do these things are available at this link:

<https://xmlpipedb.cs.lmu.edu/biodb/fall2013/index.php/Coder>. For our species, the code looked like this prior to building, testing, and committing the program to SourceForge:

```
SinorhizobiumMelilotiUniProtSpeciesProfile.java
package edu.lmu.xmlpipedb.gmbuilder.databasetoolkit.profiles;

import java.sql.SQLException;

public class SinorhizobiumMelilotiUniProtSpeciesProfile extends UniProtSpeciesProfile {

    public SinorhizobiumMelilotiUniProtSpeciesProfile() {
        super("Sinorhizobium meliloti",
            266834,
            "This profile customizes the GenMAPP Builder export for " +
            "Sinorhizobium meliloti" +
            " data loaded from a UniProt XML file.");
    }

    @Override
    public TableManager getSystemsTableManagerCustomizations(TableManager tableManager, DatabaseProfile dbProfile) {
        super.getSystemsTableManagerCustomizations(tableManager, dbProfile);
        tableManager.submit("Systems", QueryType.update, new String[][] {
            { "SystemCode", "N" },
            { "Species", "|" + getSpeciesName() + "|" }
        });

        tableManager.submit("Systems", QueryType.update, new String[][] {
            { "SystemCode", "N" },
            { "Link", "http://cmr.jcvi.org/tigr-scripts/CMR/shared/GenePage.cgi?locus=~" }
        });

        return tableManager;
    }
}
```

For our second export using [GenMAPP\\_Builder\\_2.0b72](#) *S. meliloti*, we decided to use a different, more powerful Dell lab computer to help speed up the process, which involved creating another database and importing all the important files again. Times to import and export were as follows: UniProt XML: 7.34 minutes, GO OBO-XML: 6.31 minutes to import and 4.60 minutes to process, GOA: 0.07 minutes, .gdb: about three hours. As before, Microsoft Access and Tally Engine confirm the same OrderedLocusNames with 4712 records.

In order to account for the missing genes within the database, Microsoft Access was used to screen for any tags that may have been potentially missing from the code. Upon observation, it was shown that all of the Gene IDs shared the tags “ORF” and “ordered locus names”.

To compensate for gmbuilder’s inability to locate the Gene IDs, another modification to the code was done. This time, Dr. David Dionisio assisted us with the coding to include Gene IDs found within “ORF” and “ordered locus”; the code is stated below and highlighted to show its placement amongst the other strings of code within gmbuilder:

```
public TableManager getSystemTableManagerCustomizations(TableManager tableManager,
TableManager primarySystemTableManager, Date version) throws SQLException,
InvalidParameterException {
    List<String> comparisonList = new ArrayList<String>();
    comparisonList.add("ordered locus");
    comparisonList.add("ORF");

    return systemTableManagerCustomizationsHelper(tableManager, primarySystemTableManager,
version, "OrderedLocusNames", comparisonList);
}
```



```

SinorhizobiumMellilotiUniProtSpeciesProfile.java
package edu.lmu.xmlpipeedb.gmbuilder.databasetoolkit.profiles;

import java.sql.SQLException;

public class SinorhizobiumMellilotiUniProtSpeciesProfile extends UniProtSpeciesProfile {

    public SinorhizobiumMellilotiUniProtSpeciesProfile() {
        super("Sinorhizobium melliloti",
            266834,
            "This profile customizes the GenMAPP Builder export for " +
            "Sinorhizobium melliloti" +
            " data loaded from a UniProt XML file.");
    }

    @Override
    public TableManager getSystemsTableManagerCustomizations(TableManager tableManager, DatabaseProfile dbProfile) {
        super.getSystemsTableManagerCustomizations(tableManager, dbProfile);
        tableManager.submit("Systems", QueryType.update, new String[][] {
            { "SystemCode", "N" },
            { "Species", "|" + getSpeciesName() + "|" }
        });

        tableManager.submit("Systems", QueryType.update, new String[][] {
            { "SystemCode", "N" },
            { "Link", "http://cmr.jcvi.org/tigr-scripts/CMR/shared/GenePage.cgi?locus=" }
        });

        return tableManager;
    }

    @Override
    public TableManager getSystemTableManagerCustomizations(TableManager tableManager, TableManager primarySystemTableManager, Date version) throws SQLException,
        List<String> comparisonList = new ArrayList<String>();
        comparisonList.add("ordered locus");
        comparisonList.add("ORF");

        return systemTableManagerCustomizationsHelper(tableManager, primarySystemTableManager, version, "OrderedLocusNames", comparisonList);
    }
}

```

We mistakenly imported the same set of data into the database on our Dell computer and started an export without realizing that it would create duplicate data; therefore, we discarded this export and created another database on a different Dell computer for the final import/export cycle. Times to import and export were as follows: UniProt XML: 6.25 minutes, GO OBO-XML: 6.21 minutes to import and 4.50 minutes to process, GOA: 0.07 minutes, .gdb: about three hours. Again, Microsoft Access and Tally Engine agree that the total number of records for OrderedLocusNames is 4712.

The results from the third export had yielded fewer exceptions with the modification to the code. These exceptions were compiled and pasted into an excel spreadsheet along with the compiled Gene ID list given from the original XML file. A MATCH test was run in order to ensure that the presence of the exceptions was due to the fact that the Gene IDs from the microarray dataset were never in the XML file to start with. In all columns, the results of the comparisons showed up as #N/A. This meant that the

Gene IDs were not present in the XML file to begin with and that all of the exceptions were not due to an error within the code.

The following table represents the results of the statistical analysis performed on the raw microarray data:

.7 M Sucrose (20160 total microarray dots)

| P-value           | Time Interval | Number of Microarray Dots |
|-------------------|---------------|---------------------------|
| <b>&lt; .05</b>   | 15 minutes    | 3613 (17.9%)              |
|                   | 30 minutes    | 5225 (25.9%)              |
|                   | 60 minutes    | 5207 (25.8%)              |
|                   | 240 minutes   | 6790 (33.7%)              |
| <b>&lt; .01</b>   | 15 minutes    | 907 (4.5%)                |
|                   | 30 minutes    | 1518 (7.5%)               |
|                   | 60 minutes    | 1553 (7.7%)               |
|                   | 240 minutes   | 2437 (12.1%)              |
| <b>&lt; .001</b>  | 15 minutes    | 92 (0.5%)                 |
|                   | 30 minutes    | 179 (0.9%)                |
|                   | 60 minutes    | 172 (0.9%)                |
|                   | 240 minutes   | 347 (1.7%)                |
| <b>&lt; .0001</b> | 15 minutes    | 7 (0.03%)                 |
|                   | 30 minutes    | 15 (0.07%)                |
|                   | 60 minutes    | 13 (0.06%)                |
|                   | 240 minutes   | 36 (0.18%)                |

The table above shows that the time of 240 minutes consistently had the most significantly changed microarray dots. However, the time of 15 minutes was chosen in the case of the .7 M sucrose

for further study of gene ontology and biological pathways. The following image is the top twenty gene ontology results for the “increased” data found using GenMAPP and MAPPFinder:

| Gene Ontology Results  |                     |               |  |
|--|---------------------|---------------|--|
| regulation of RNA biosynthetic process                         | NESTED 68/377 18%   | 377/514 73.3% | z score = 5 permute p = 0 adjusted p = 0.192     |
| regulation of RNA metabolic process                            | NESTED 68/377 18%   | 377/514 73.3% | z score = 5 permute p = 0 adjusted p = 0.192     |
| regulation of transcription, DNA-dependent                     | NESTED 68/377 18%   | 377/514 73.3% | z score = 5 permute p = 0 adjusted p = 0.192     |
| regulation of nucleobase-containing compound metabolic process | NESTED 68/379 17.9% | 379/516 73.4% | z score = 4.952 permute p = 0 adjusted p = 0.192 |
| regulation of nitrogen compound metabolic process              | NESTED 68/379 17.9% | 379/516 73.4% | z score = 4.952 permute p = 0 adjusted p = 0.192 |
| transcription, DNA-dependent                                   | NESTED 68/381 17.8% | 381/519 73.4% | z score = 4.903 permute p = 0 adjusted p = 0.193 |
| RNA biosynthetic process                                       | NESTED 68/385 17.7% | 385/524 73.5% | z score = 4.806 permute p = 0 adjusted p = 0.196 |
| DNA binding  | NESTED 80/478 16.7% | 478/627 76.2% | z score = 4.728 permute p = 0 adjusted p = 0.21  |
| nucleic acid metabolic process                                 | NESTED 98/622 15.8% | 622/789 78.8% | z score = 4.649 permute p = 0 adjusted p = 0.212 |
| regulation of macromolecule biosynthetic process               | NESTED 68/393 17.3% | 393/530 74.2% | z score = 4.616 permute p = 0 adjusted p = 0.214 |
| regulation of biosynthetic process                             | NESTED 68/393 17.3% | 393/530 74.2% | z score = 4.616 permute p = 0 adjusted p = 0.214 |
| regulation of cellular biosynthetic process                    | NESTED 68/393 17.3% | 393/530 74.2% | z score = 4.616 permute p = 0 adjusted p = 0.214 |
| regulation of cellular macromolecule biosynthetic process      | NESTED 68/393 17.3% | 393/530 74.2% | z score = 4.616 permute p = 0 adjusted p = 0.214 |
| regulation of cellular metabolic process                       | NESTED 68/394 17.3% | 394/532 74.1% | z score = 4.592 permute p = 0 adjusted p = 0.214 |
| regulation of primary metabolic process                        | NESTED 68/394 17.3% | 394/532 74.1% | z score = 4.592 permute p = 0 adjusted p = 0.214 |
| regulation of gene expression                                  | NESTED 68/395 17.2% | 395/532 74.2% | z score = 4.569 permute p = 0 adjusted p = 0.215 |
| regulation of macromolecule metabolic process                  | NESTED 68/397 17.1% | 397/535 74.2% | z score = 4.522 permute p = 0 adjusted p = 0.223 |
| regulation of metabolic process                                | NESTED 68/402 16.9% | 402/541 74.3% | z score = 4.407 permute p = 0 adjusted p = 0.279 |
| regulation of cellular process                                 | NESTED 78/502 15.5% | 502/667 75.3% | z score = 3.921 permute p = 0 adjusted p = 0.968 |
| RNA metabolic process  | NESTED 78/466 15.7% | 466/609 76.5% | z score = 3.851 permute p = 0 adjusted p = 0.968 |

The following image is the top twenty gene ontology results for the “decreased” data found using GenMAPP and MAPPFinder:

| Gene Ontology Results  |                      |               |  |
|--|----------------------|---------------|--|
| cytoplasmic part   | NESTED 40/78 51.3%   | 78/81 96.3%   | z score = 5.746 permute p = 0 adjusted p = 0     |
| structural constituent of ribosome   | NESTED 31/55 56.4%   | 55/55 100%    | z score = 5.7 permute p = 0 adjusted p = 0       |
| intracellular part   | NESTED 157/456 34.4% | 456/474 96.2% | z score = 5.678 permute p = 0 adjusted p = 0     |
| ribosome   | NESTED 31/56 55.4%   | 56/56 100%    | z score = 5.574 permute p = 0 adjusted p = 0     |
| structural molecule activity   | NESTED 36/70 51.4%   | 70/70 100%    | z score = 5.466 permute p = 0 adjusted p = 0     |
| ribonucleoprotein complex  | NESTED 31/57 54.4%   | 57/57 100%    | z score = 5.431 permute p = 0 adjusted p = 0     |
| non-membrane-bounded organelle   | NESTED 42/87 48.3%   | 87/87 100%    | z score = 5.41 permute p = 0 adjusted p = 0      |
| intracellular non-membrane-bounded organelle                                     | NESTED 42/87 48.3%   | 87/87 100%    | z score = 5.41 permute p = 0 adjusted p = 0      |
| intracellular organelle  | NESTED 42/88 47.7%   | 88/88 100%    | z score = 5.32 permute p = 0 adjusted p = 0.001  |
| organelle  | NESTED 42/88 47.7%   | 88/88 100%    | z score = 5.32 permute p = 0 adjusted p = 0.001  |
| translation  | NESTED 48/106 45.3%  | 106/108 98.1% | z score = 5.255 permute p = 0 adjusted p = 0.001 |
| nucleoside monophosphate biosynthetic process                                    | NESTED 25/44 56.8%   | 44/44 100%    | z score = 5.161 permute p = 0 adjusted p = 0.001 |
| ribonucleoside monophosphate biosynthetic process                                | NESTED 24/42 57.1%   | 42/42 100%    | z score = 5.091 permute p = 0 adjusted p = 0.001 |
| cytoplasm  | NESTED 138/407 33.9% | 407/423 96.2% | z score = 5.058 permute p = 0 adjusted p = 0.001 |
| ATPase activity, coupled to transmembrane movement of ions, rotational mechanism | NESTED 8/8 100%      | 8/8 100%      | z score = 5.058 permute p = 0 adjusted p = 0.002 |
| plasma membrane ATP synthesis coupled proton transport                           | NESTED 8/8 100%      | 8/8 100%      | z score = 5.058 permute p = 0 adjusted p = 0.002 |
| proton-transporting ATP synthase activity, rotational mechanism                  | NESTED 8/8 100%      | 8/8 100%      | z score = 5.058 permute p = 0 adjusted p = 0.002 |
| ribosomal subunit  | NESTED 11/14 78.6%   | 14/14 100%    | z score = 4.812 permute p = 0 adjusted p = 0.003 |
| intracellular  | NESTED 162/504 32.1% | 504/528 95.5% | z score = 4.715 permute p = 0 adjusted p = 0.005 |
| macromolecular complex   | NESTED 56/139 40.3%  | 139/161 86.3% | z score = 4.637 permute p = 0 adjusted p = 0.005 |

Based off of these gene ontology results, the following biological pathway was created using MAPPFinder:



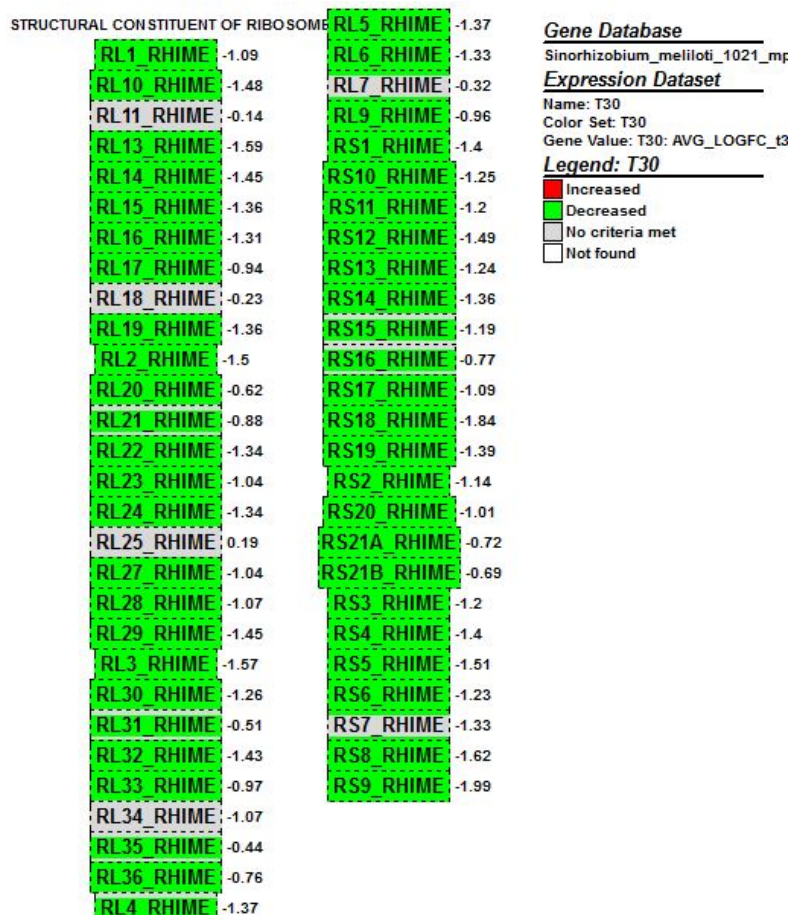
## Gene Ontology Results

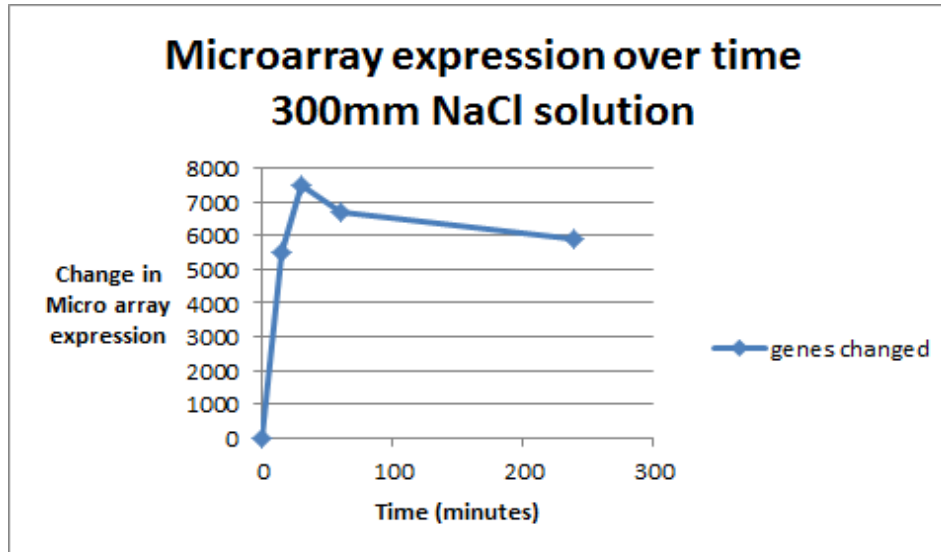
```

intracellular part NESTED 269/456 59%, 456/474 96.2% z score = 10.501 permute p = 0 adjusted p = 0
cytoplasm NESTED 239/407 58.7%, 407/423 96.2% z score = 9.722 permute p = 0 adjusted p = 0
translation NESTED 85/106 80.2%, 106/108 98.1% z score = 9.388 permute p = 0 adjusted p = 0
structural molecule activity NESTED 62/70 88.6%, 70/70 100% z score = 9.057 permute p = 0 adjusted p = 0
intracellular NESTED 276/504 54.8%, 504/528 95.5% z score = 9.001 permute p = 0 adjusted p = 0
structural constituent of ribosome NESTED 51/55 92.7%, 55/55 100% z score = 8.655 permute p = 0 adjusted p = 0
intracellular organelle NESTED 71/88 80.7%, 88/88 100% z score = 8.629 permute p = 0 adjusted p = 0
organelle NESTED 71/88 80.7%, 88/88 100% z score = 8.629 permute p = 0 adjusted p = 0
non-membrane-bounded organelle NESTED 70/87 80.5%, 87/87 100% z score = 8.535 permute p = 0 adjusted p = 0
intracellular non-membrane-bounded organelle NESTED 70/87 80.5%, 87/87 100% z score = 8.535 permute p = 0 adjusted p = 0
ribosome NESTED 51/56 91.1%, 56/56 100% z score = 8.476 permute p = 0 adjusted p = 0
cytoplasmic part NESTED 64/78 82.1%, 78/81 96.3% z score = 8.365 permute p = 0 adjusted p = 0
ribonucleoprotein complex NESTED 51/57 89.5%, 57/57 100% z score = 8.3 permute p = 0 adjusted p = 0
cell NESTED 458/964 47.5%, 964/1136 84.9% z score = 8.048 permute p = 0 adjusted p = 0
cell part NESTED 458/964 47.5%, 964/1136 84.9% z score = 8.048 permute p = 0 adjusted p = 0
macromolecular complex NESTED 96/139 69.1%, 139/161 86.3% z score = 8.03 permute p = 0 adjusted p = 0
cellular_component NESTED 656/1504 43.6%, 1504/1782 84.4% z score = 7.186 permute p = 0 adjusted p = 0
cellular protein metabolic process NESTED 115/191 60.2%, 191/205 93.2% z score = 6.878 permute p = 0 adjusted p = 0
rRNA binding NESTED 33/38 86.8%, 38/38 100% z score = 6.42 permute p = 0 adjusted p = 0
RNA binding NESTED 66/99 66.7%, 99/105 94.3% z score = 6.235 permute p = 0 adjusted p = 0

```

## Ribosome constituents MAPP





The graph shows microarray data points and not the actual genes that were changed.

## Discussion

Initial results of the first export were troublesome but expected, since the version of gmbuilder we used was not designed for *S. meliloti*. However, as we continued to make changes to the code, we were able to significantly reduce the total amount of errors by 87%. The results of the gene ontology showed a significant increase in the number of genes activated as opposed to the genes that were being repressed in the .3 M NaCl solution. In the NaCl experiment the genes that were shown to be most heavily induced were those that coded for the ribosomal subunits as well as a number of genes that acted in different parts of the metabolic pathway and 1 family of genes that coded for the cell wall structure of the bacteria. The one gene that was found to be significantly repressed was a gene that functioned in the nitrogen fixation elements of the bacteria symbiote. These results were interpreted to mean that the bacteria, under salt stress, acted in a way to bolster its metabolism for structural integrity as well as to create the metabolic pathways needed to harvest more energy so that the salt could be pumped out of the cell. During this stress the energy metabolism was diverted away from nitrogen

fixation and concentrated more on preserving the cell against salt stress.

When dealing with the data of the .7 M experiment, there was an error encountered in loading the GenMAPP dataset in GenMAPP, because the gene IDs in the dataset did not match those in the gene database. However, this issue could be easily changed for most of the microarray dots. because they had IDs similar to those within the database, but they has been packed between sets of extraneous information. However, this was not easy for about 25% of the microarray dot points, because they were IDed differently than was seen in the database. This caused GenMAPP to recognize 5,538 errors of “Gene not found in OrderedLocusNames or any related system”.

Upon studying the Gene Ontology of the “increased” GenMAPP data, I chose to look at the top Gene Ontology term, “Regulation of RNA, DNA dependent”. However, there were many other top terms also related to RNA, and quite a few relating to nitrogen fixing metabolic processes. These changes, show that the addition of .7 M to a culture of *Sinorhizobium meliloti*, did significantly change the the expression of numerous genes. Not much can be concluded inrelation to the addition of .7 M specifically, but many of the same genes that were effected in this experiment, were also seen to be effected in the other three experiments as well. Therefore, it is expected that there is a group of genes responsible for responding to this type of salinity and hyperosmotic stress. Many of the genes involved in these processes were also seen to have unidentified functions in the cell. Therefore, it is interesting to note that there is still much to be studied about the *Sinorhizobium* genome.

## References

1. Domínguez-Ferreras, A., Pérez-Arnedo, R., Becker, A., Olivares, J., Soto, M.J., Sanjuán, J. (2006) Transcriptome Profiling Reveals the Importance of Plasmid pSymB for Osmoadaptation of *Sinorhizobium meliloti* *Journal of Bacteriology* 188:7617-7625

2. Galibert, F., Finan, T.M., Long, S., Puhler, A., et al. (2001) The composite genome of the legume symbiont *Sinorhizobium meliloti* *Science* 293:668-672

### **Acknowledgments**

We would like to acknowledge our two professors who offered us immense help throughout this entire project. Specifically thank you Dr. Dionisio for your contribution to the coding portions of this project, and thank you Dr. Dahlquist for your aid in all things GenMAPP related. We would also like to thank Dr. Lum for her expertise in *Sinorhizobium meliloti*, a rhizobacterium that she conducts independent research on.