



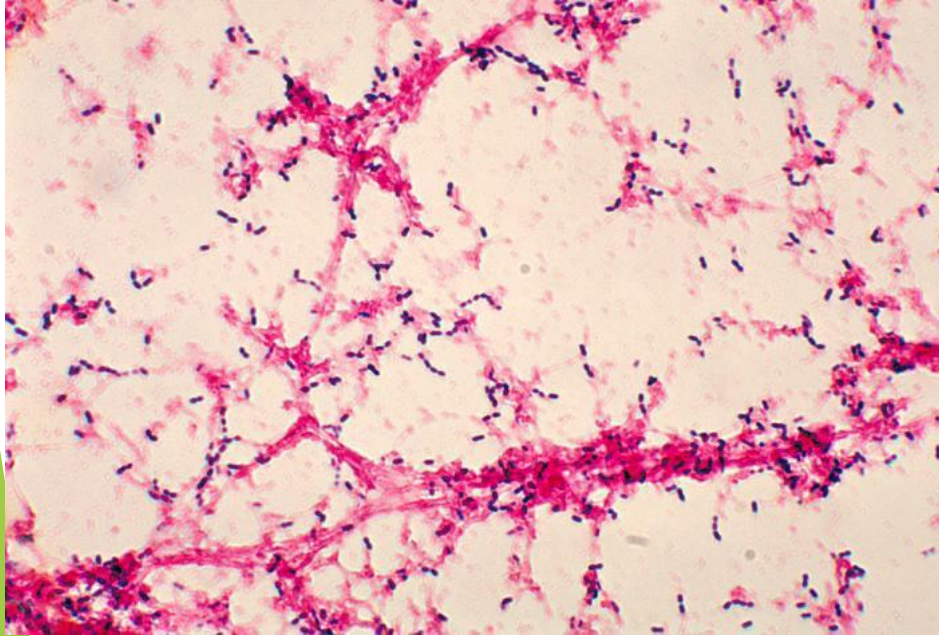
# Changes in *Streptococcus pneumoniae* Carbohydrate Transporter Expression During Biofilm Formation

Kevin Meilak, Tauras Vilgalys, Alina Vreeland  
BIOL/CMSI 367: Biological Databases  
Loyola Marymount University  
December 12, 2013

# Outline

- It is currently unclear what genomic changes *Streptococcus pneumoniae* undergoes during biofilm formation
- We created a gene database for *S. pneumoniae*
- Using microarray data from Sanchez et al. (2011), we examined changes in gene expression using GenMAPP builder
- The pathways most significantly changed after 12 hours of growth were related to carbohydrate transformation

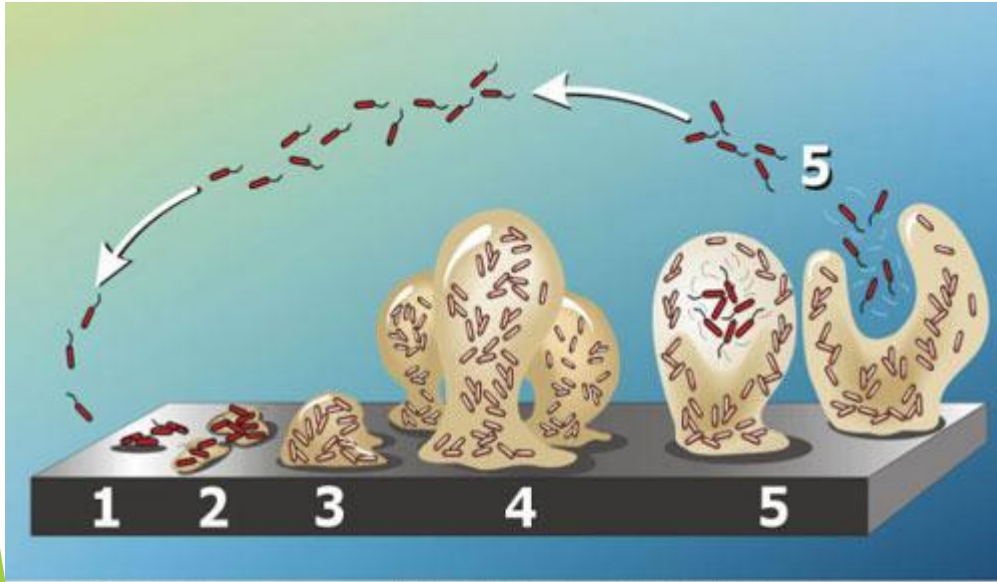
# *Streptococcus pneumoniae*: The Infectious Agent of Pneumonia



<http://textbookofbacteriology.net/PHILspbloodGram.jpg>

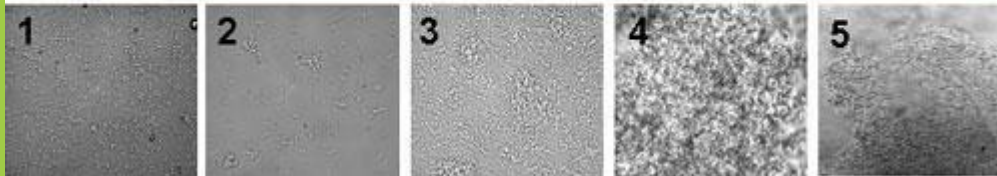
- Infectious agent of pneumococcal diseases such as pneumonia, meningitis, conjunctivitis
- One of top 10 causes of death in the US
- Commonly researched because it transitions between virulent and avirulent phases
- Gram positive coccus bacteria

# Biofilms are a transitional stage of the *S. pneumoniae* life cycle



Adherent cells surrounded by self-produced 'slime'

- 1) Initial Attachment
- 2) Final Attachment
- 3) Maturation I
- 4) Maturation II
- 5) Dispersion



# Species customization in GenMAPP Builder

- GenMAPP builder version 2.0b73 was downloaded from [sourceforge.net/projects/smlpipedb](http://sourceforge.net/projects/smlpipedb)
  - Included a species profile for *S. pneumoniae* TIGR4 and link to the Ensembl Bacteria gene page
  - Had VC-like code to adjust Gene IDs from SP##### to SP\_####

```
@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://bacteria.ensembl.org/streptococcus_pneumoniae_tigr4/Gene/Summary?g=~" }
    });

    return tableManager;
}
```

## Assembling data files for the gene database

- Downloaded the UniProt complete proteome for *Streptococcus pneumoniae* serotype 4 (strain ATCC BAA-334 / TIGR4), last updated November 13, 2013
- Downloaded the GOA file for 57.S\_pneumoniae\_TIGR4, last updated November 12, 2013 14:49:00.
- Downloaded the November 20, 2013 OBO-XML file from [gene ontology.org](http://geneontology.org)

# GenMAPP Builder was used to create a gene database

## In PostgreSQL:

- Created new database
- Ran `gmbuilder.sql` to import GenMAPP tables

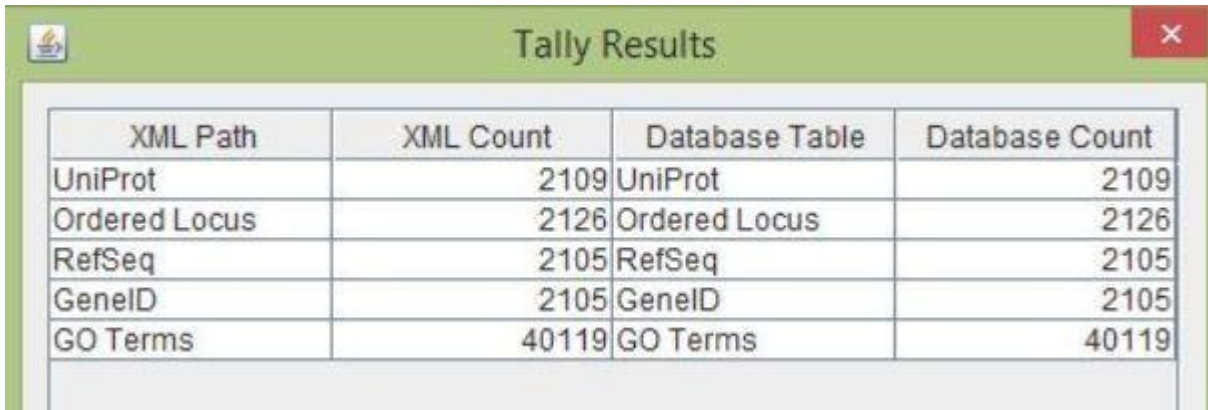
## In GenMAPP Builder

- Connected to the PostgreSQL database
- Imported OBO-XML, GOA, and UniProt XML files
- Exported database titled `Streptococcus_pneumoniae_TIGR4_20131125.gdb`

# Different counting programs were used to validate gene ID totals

## Tally Engine:

- Ordered Locus totals for XML and Database counts matched with 2126 results

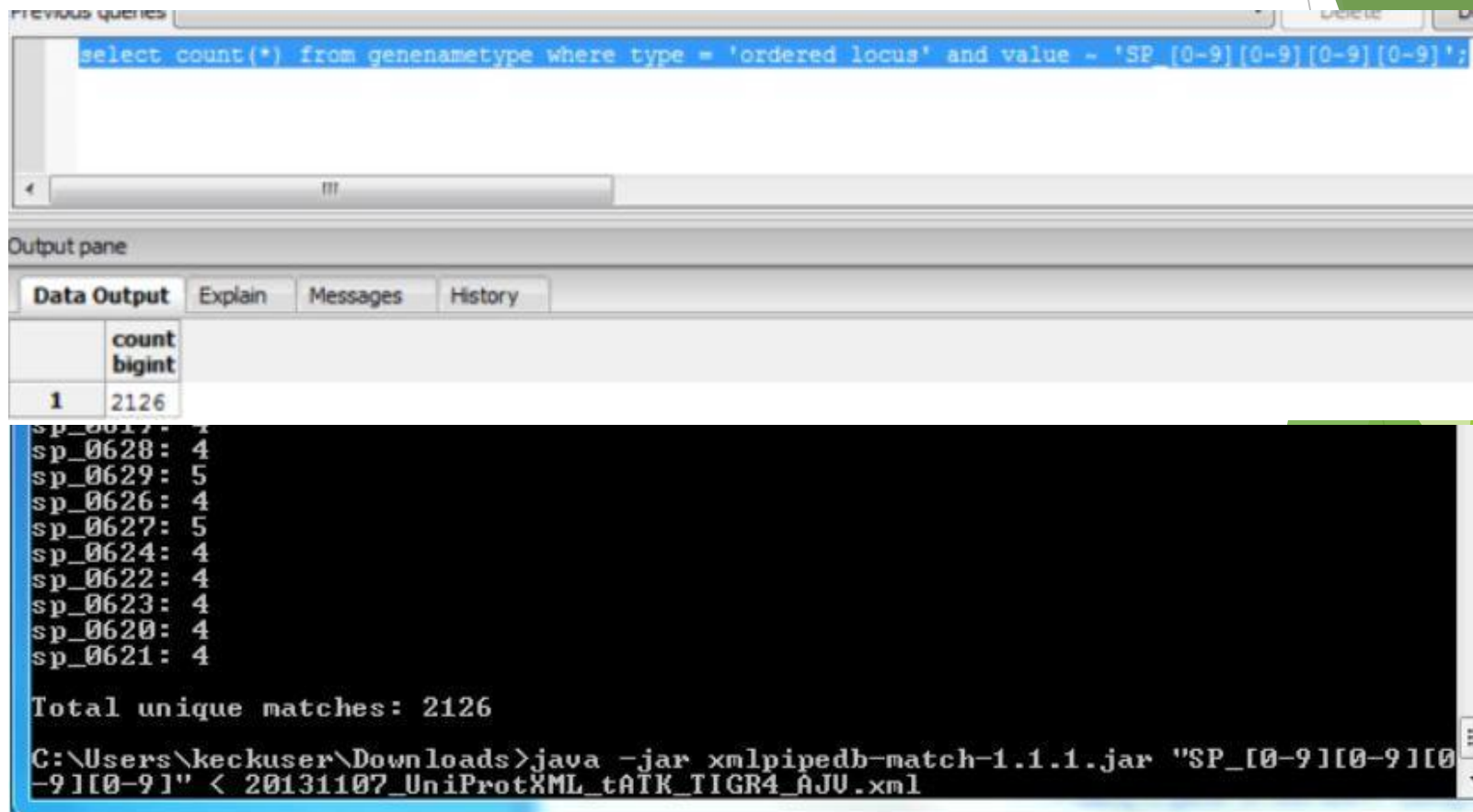


The screenshot shows a window titled "Tally Results" with a table comparing XML and Database counts for five categories: UniProt, Ordered Locus, RefSeq, GeneID, and GO Terms. The counts for each category are identical in both columns.

XML Path	XML Count	Database Table	Database Count
UniProt	2109	UniProt	2109
Ordered Locus	2126	Ordered Locus	2126
RefSeq	2105	RefSeq	2105
GeneID	2105	GeneID	2105
GO Terms	40119	GO Terms	40119



# Postgres SQL and XMLPipeDB Match validated TallyEngine results for the XML file



The screenshot displays a database query interface. The query window contains the following SQL statement:

```
select count(*) from genenametype where type = 'ordered locus' and value ~ 'SP_[0-9][0-9][0-9][0-9]';
```

The output pane shows a table with the following data:

	count bigint
1	2126

Below the table, a list of unique matches is shown:

```
sp_0617: 4  
sp_0628: 4  
sp_0629: 5  
sp_0626: 4  
sp_0627: 5  
sp_0624: 4  
sp_0622: 4  
sp_0623: 4  
sp_0620: 4  
sp_0621: 4
```

Total unique matches: 2126

The command prompt at the bottom shows the execution of the XMLPipeDB match tool:

```
C:\Users\keckuser\Downloads>java -jar xmlpipedb-match-1.1.1.jar "SP_[0-9][0-9][0-9][0-9]" < 20131107_UniProtXML_tATK_TIGR4_AJU.xml
```

# Row Counts performed on the Gene Database confirmed prior results

ID System	ID Count
EMBL	201
GeneOntology	3648
InterPro	2641
OrderedLocusNames	4252*
PDB	225
Pfam	1277
RefSeq	2105
UniProt	2109

- There are 2126 unique genes/proteins in the current version of the Gene Database
- The 4252 count includes duplicate IDs of the form SP##### and SP\_#####.

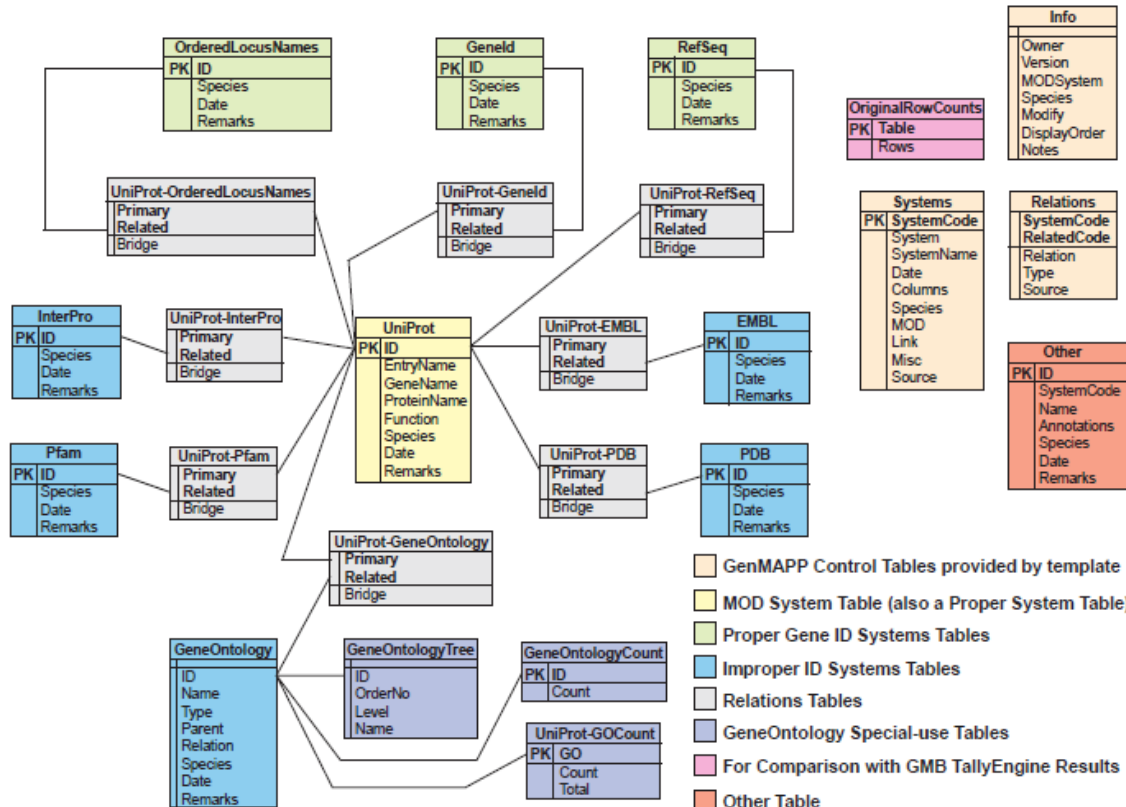
# Visual inspection of tables within the gdb determined accuracy of gene ID's found

- **UniProt table:**

- Gene ID was in the format SP\_####, as expected
- Species matched the Species Profile name incorporated into GenMAPP builder

ID	Species	Date
SP_0002	StreptococcusPneumoniaeTIGR4	11/25/2013
SP_0274	StreptococcusPneumoniaeTIGR4	11/25/2013
SP_0291	StreptococcusPneumoniaeTIGR4	11/25/2013
SP_0458	StreptococcusPneumoniaeTIGR4	11/25/2013
SP_1644	StreptococcusPneumoniaeTIGR4	11/25/2013

# The *S. pneumoniae* Schema had all the same linkages and information as the *V. Cholera* file

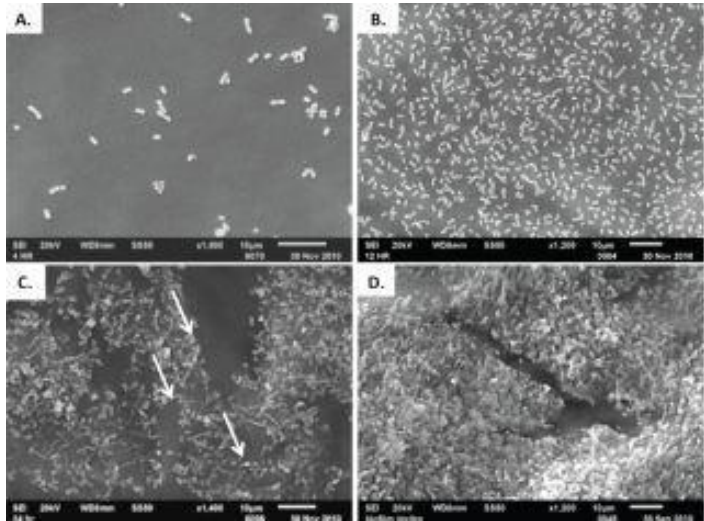


NOTE: Some Relations tables are not shown. All possible pairwise Relations tables exist between Proper ID systems and between Proper and Improper ID systems, but not between Improper ID systems (i.e., Proper-Proper, Proper-Improper, but NOT Improper-Improper).

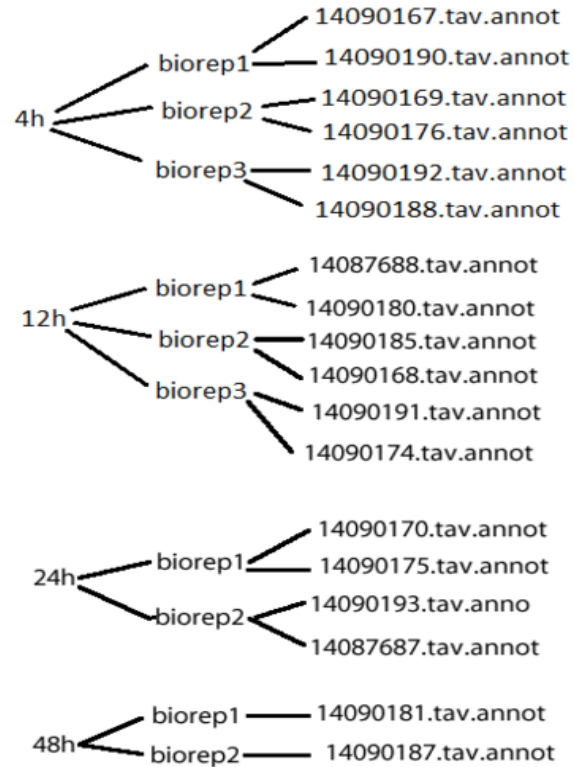
# Sanchez et al. 2011 found pneumococci biofilms were avirulent

- Tested three strains (TIGR4, R6, G54)
- Injected mice with planktonic, biofilm, and biofilm derived *S. pneumoniae*
- Biofilm and biofilm derived bacteria infected the lung, but did not enter into the bloodstream
- Planktonic bacteria resulted in mouse death
- Compared gene expression between biofilm to planktonic bacteria
- Concluded that pneumococci from biofilms were non-virulent, but hyper adhesive

# Sanchez et al. contained technical and biological replicates for four stages of biofilm growth



Images of biofilm growth are shown at 4 (A), 12 (B), 24, and 48 (D) hours. In each picture, there was an increase in total biofilm. (Sanchez et al. 2012)



Technical replicates were not available due to limited amount of RNA extracted

# Microarray data revealed the greatest differences between samples at 12 hours

- Scaled & centered, averaged replicates, calculated T-stat and P-value for each gene at each time
  - Generally followed *V. cholerae* instructions
  - Replaced error messages (50,571 in total) with space character
- Averages were by biological replicate then time point
- Issues with raw data: no column headers

# 12 Hour Time Focused on due to Highest Percent of Significance

Time	Significant P-values	Percent of all genes
4hr	1218	26%
12hr	1902	41%
24hr	918	20%
48hr	129	2.8%



# Exceptions found after filtering for TIGR4 strain were compared to original XML file IDs

- List of 333 exceptions was compared with the original UniProt gene IDs compiled by XMLpipeDB Match
- All genes found as #N/A
  - Means that they were absent from the in original UniProt XML file

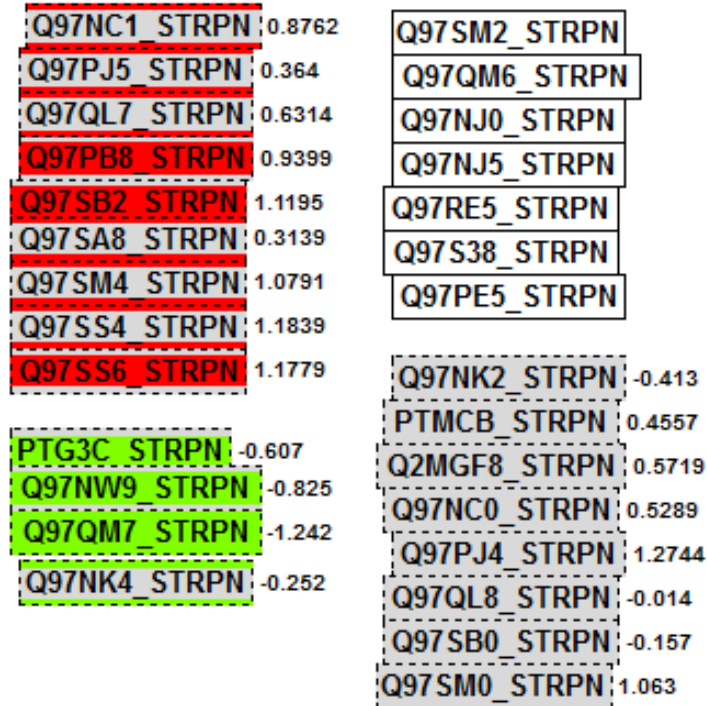
	A	B	C
1	Gene ID	Match Gene IDs	1 to 2 <input type="text"/>
2	SP0365	SP0220	#N/A
3	SP0388	SP0221	#N/A
4	SP0131	SP0222	#N/A
5	SP0644	SP0223	#N/A
6	SP0432	SP0224	#N/A
7	SP0697	SP0225	#N/A
8	SP0365	SP0226	#N/A
9	SP0131	SP0227	#N/A
10	SP0432	SP0228	#N/A
11	SP0697	SP0229	#N/A
12	SP1194	SP1861	#N/A
13	SP1613	SP1862	#N/A
14	SP0131	SP1863	#N/A
15	SP1194	SP1864	#N/A
16	SP0388	SP1860	#N/A
17	SP0644	SP1869	#N/A

# Significant differences were found mostly within pathways for carbohydrate transport

Gene Ontology Results	pvalue	zscore
phosphoenolpyruvate-dependent sugar phosphotransferase system	0.004	6.665
transporter activity	0.004	6.335
protein-N(PI)-phosphohistidine-sugar phosphotransferase activity	0.004	6.213
carbohydrate transport	0.004	6.153
carbohydrate transporter activity	0.033	5.744
carbohydrate transmembrane transporter activity	0.033	5.744
transport	0.033	5.674
establishment of localization	0.033	5.674
localization	0.033	5.674
sugar transmembrane transporter activity	0.039	5.447
solute:cation symporter activity	0.039	5.273

# The created MAPP revealed differences in genes functional in sugar transmembrane transport activity at the 12 hour time point

## SUGAR:HYDROGEN SYMPORTER ACTIVITY



## MALTOSE TRANSMEMBRANE TRANSPORTER ACTIVITY

MALX\_STRPN -0.307

## TREHALOSE TRANSMEMBRANE TRANSPORTER ACTIVITY

Q97NW9\_STRPN -0.825

## SUGAR EFFLUX TRANSMEMBRANE TRANSPORTER ACTIVITY

CPSC\_STRPN -1.207

# Sugar transmembrane transporter activity did not change in a predictable fashion

- 6 genes decreased
  - All transmembrane transporter proteins or involved in sugar phosphorylation
- 9 genes increased
  - Phosphoenol-pyruvate transport system, transmembrane transporter protein, involved in sugar phosphorylation
- 9 genes no criteria met
  - Some sugar transport and involved in sugar phosphorylation, though many predicted proteins with no known function
- 7 genes not found
  - catalytic activity or no known function

# Conclusions

- Biofilm formation is a transitional phase in the life cycle of *S. pneumoniae* during which it is avirulent
- We created a gene database for *S. pneumoniae* str. TIGR4 using GenMAPP Builder
- Analysis in GenMAPP revealed gene expression in the carbohydrate transport pathway was influenced by biofilm formation
- However, it is unclear what changes are actually occurring and more research must be performed to determine the function of specific genes

# Reference

Sanchez, C.J., Kumar, N., Lizcano, A., Shivshankar, P., Dunning Hotopp, J.C., Jorgensen, J.H., Tettelin, H., and Orihuela, C.J. (2011) *Streptococcus pneumoniae* in Biofilms Are Unable to Cause Invasive Disease Due to Altered Virulence Determinant Production. *PLoS ONE* 6(12): e28738.  
doi:10.1371/journal.pone.0028738

# Acknowledgements

We would like to thank Drs. Dahlquist and Dionisio for all their help and expertise while preparing and planning this project.

Special thanks to Dr. Dahlquist for assistance in using GenMAPP and to Dr. Dionisio for helping modify GenMAPP builder code.