Examining Virulence in *Streptococcus pneumoniae* by Sequencing the R6 Laboratory Strain

Kevin Meilak, Tauras Vilgalys, Alina Vreeland

BIOL/CMSI 367: Biological Databases

Loyola Marymount University

November 12, 2013

Outline

- R6 is an unencapsulated, avirulent strain of *S. pneumoniae*, one of the leading causes of death in the US

- Genome sequencing reveals R6 possesses some virulence genes, but a large deletion of capsule biosynthesis genes

- SpR6 is a model organism database for *S. pneumoniae* which provides nucleotide, protein, and organism information

Streptococcus pneumoniae: The Infectious Agent of Pneumonia

- One of the top 10 causes of death in the United States

- One third of patient isolates in the US are penicillin resistant

- Research focuses on determining putative virulence genes and antibiotic targets

http://textbookofbacteriology.net/PHILspbloodGram.jpg

- Gram-positive coccus bacteria whose primary metabolic product is lactic acid
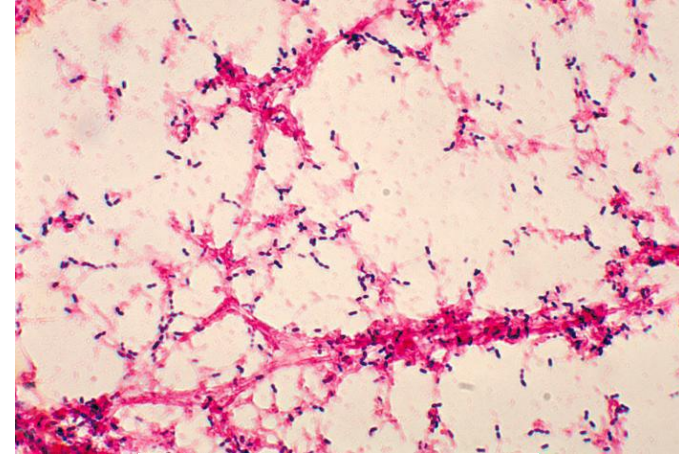
The R6 Strain: An Avirulent Version for Labratory Testing

- Descendant of the type 2 capsule S clinical isolate (Avery et al., 1944)

- No polysaccharide capsule
    - Avirulent and safe to work with
    - Genetically malleable

The Genome of *S. pneumoniae*

- Shot-gun sequencing with sequence spanning PCR and custom primers end-contig sequencing to fill in gaps

- Single circular chromosome of 2.04 bbp (40% G+C)
- 2,043 protein coding regions and 73 noncoding RNA genes

- Differentiated from virulence strains by a 7,504 bp deletion from capsule biosynthesis genes

19 R6-encoded proteins that have been studied for a role in *S. pneumoniae* virulence or as protective antigens to *S. pneumoniae*

Gene no.	Gene name	Description
spr0121	*pspA*	Surface protein (choline binding), functions in inhibition of complement activation and may involve lactoferrin binding (30, 48)
spr0286	*hysA*	Hyaluronidase (30)
spr0674	*sodA*	Manganese cofactored superoxide dismutase (30)
spr0867	*lytB*	Endo-beta-*N*-acetylglucosaminidase (15)
spr0884	*ppmA*	Surface-located proteinase maturation protein (33)
spr1042	*iga*	Immunoglobulin A1 protease (49)
spr1323	*nox (nadH)*	NADH oxidase (1)

14 Possible/probable drug efflux pumps encoded by *S. pneumoniae* R6
Orthologs play a role in drug resistance that does not appear in the R6 strain

Gene no.	Gene name	Description
spr0137	ABC-NBD&MSP	This is similar to the *E. coli msbA* gene, which is proposed by Milton Saier, Jr., and Ian Paulsen to function in drug efflux
spr0144	Conserved hypothetical	Possible macrolide efflux permease of the major facilitator superfamily
spr0610	ABC-NBD/truncation	Probably part of a drug resistance ABC transporter pump
spr0875	*pmrA*	Major facilitator superfamily multidrug resistance efflux pump

2% of R6 genes are truncated compared to orthologous genes in other species

Gene no.	Gene name	Description
spr0105	ABC-NBP&MSP-truncation	ABC transporter similar to *comA*
spr0106	ABC-NBP&MSP-truncation	ABC transporter similar to*comA*
spr0223	ABC-SBP-truncation	ABC transporter substrate-binding protein (iron transport)
spr0224	ABC-SBP-truncation	ABC transporter substrate-binding protein (iron transport)
spr0299	Conserved hypothetical	
spr0300	Conserved hypothetical	
spr0324	Transposase G-truncation	Uncharacterized transposase
spr0346	*alkD*-truncation	DNA alkylation repair enzyme
spr0347	*alkD*-truncation	DNA alkylation repair enzyme

# Transcriptional orientation of genes flanking BOX and RUP repetitive elements

Element	No. (%) of elements[a]		
	→BOX or RUP→	→BOX or RUP←	←BOX or RUP→
BOX	80 (4)	26 (13)	9 (4)[b]
RUP	49 (3)	21 (11)	14 (7)[b]

a The arrows indicate the transcriptional orientation of the genes flanking the BOX and RUP elements. The numbers of BOX and RUP elements in each orientation are listed. The parenthetic percentage values are the fractions of all intergenic regions in each transcriptional orientation that contain BOX or RUP elements.

b These numbers may be artificially high. Three of these 9 BOX elements and 6 of these 14 RUP elements are adjacent either to small hypothetical genes predicted with low confidence or to insertion elements. Were those genes falsely predicted or the IS elements not present, these BOX and RUP elements would be in the →BOX/RUP→ orientation relative to their flanking genes instead of the ←BOX/RUP→ orientation.
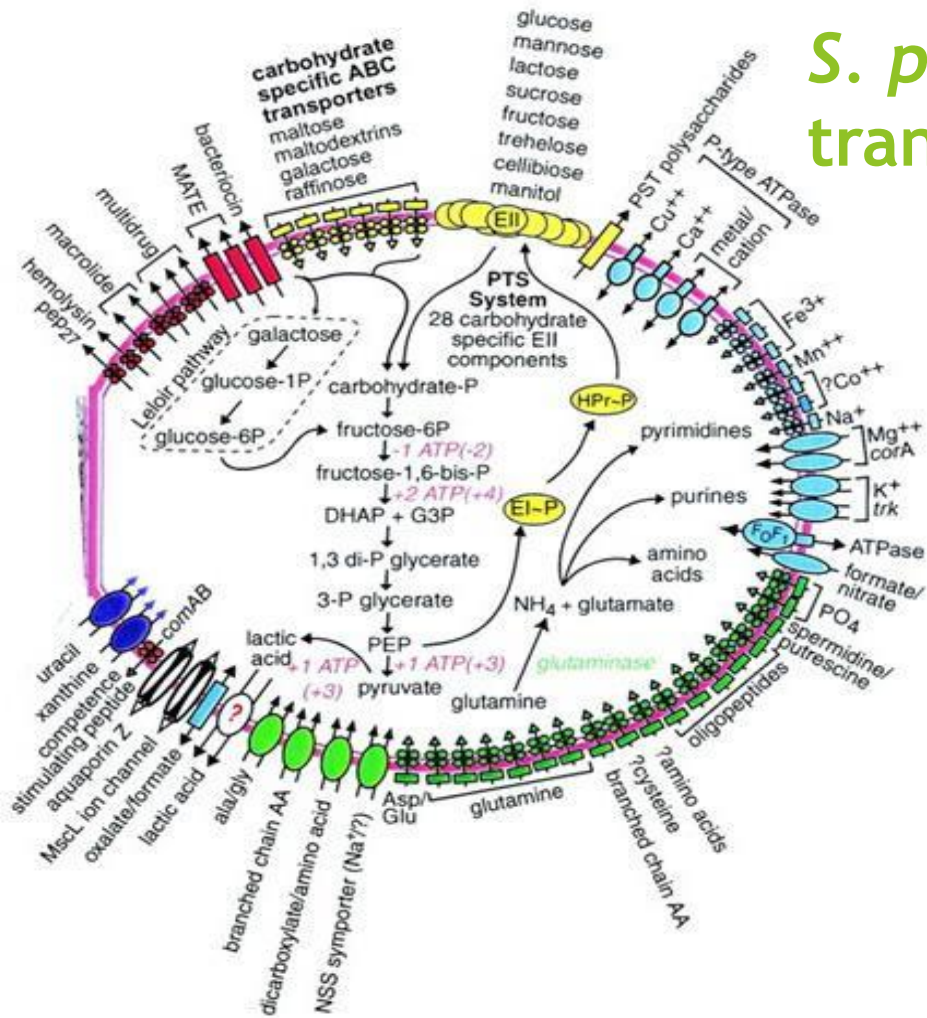
S. pneumoniae substrate transport, carbohydrate and glutamine metabolism, and selected categories of cell surface proteins

The choline-binding proteins
- autolysin (LytA)
- endo-beta-N-acetylglucosaminidase (LytB)
- surface protein involved in adherence and immunoglobulin A inactivation (CbpA)
- putative lactoferrin-binding protein (PspA)
- several surface proteins of unknown function (CbpD, PcpA, PcpC, PcpC1).
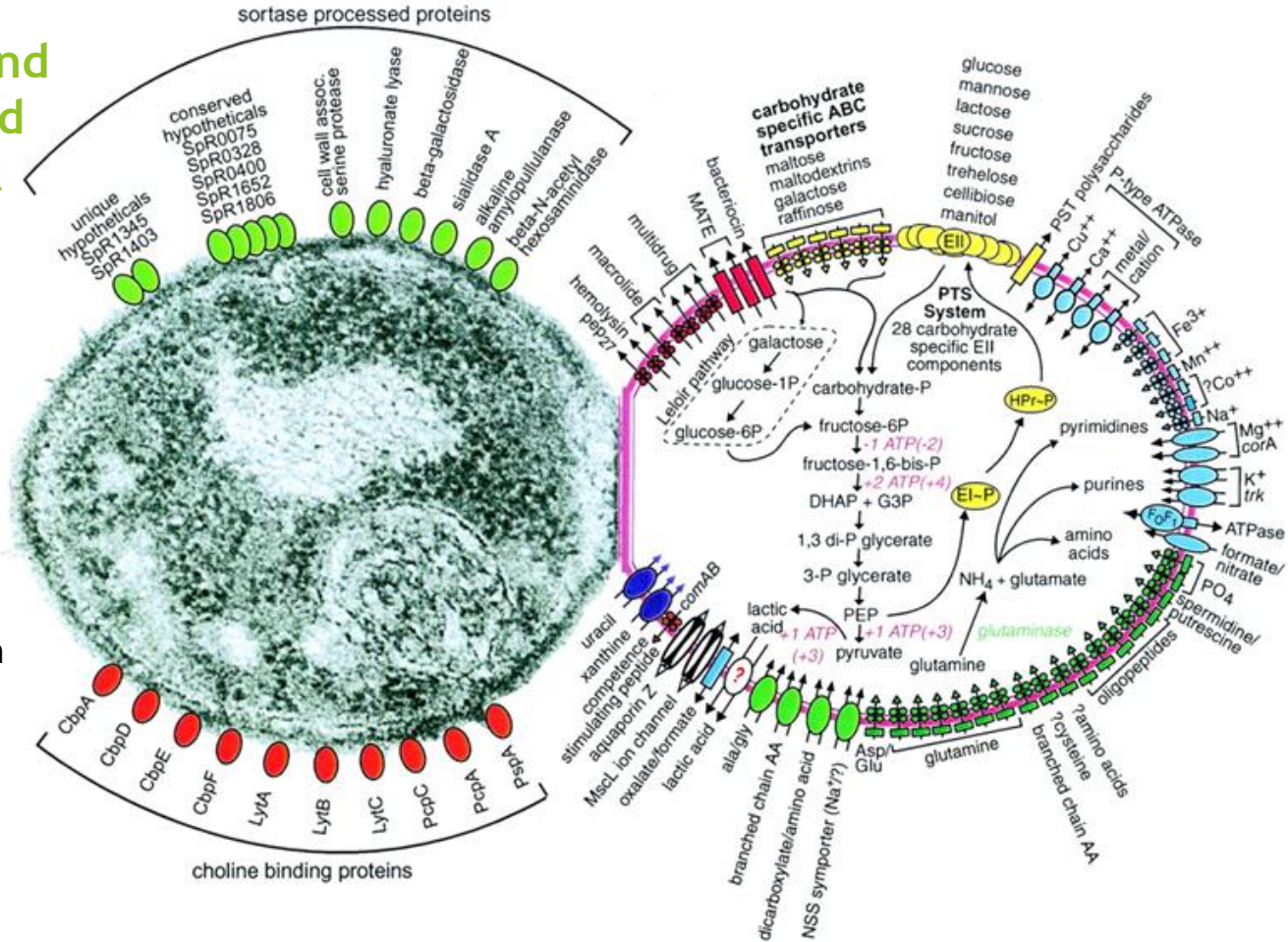
S. pneumoniae transporter proteins

| Color	Substrate Transported
Red	Multidrug and peptides
Yellow	Carbohydrates
Blue	Cations
Green	Anions and Amino Acids
Purple	Nucleosides, purines, pyrimidines
White	Other Substrates

S. *pneumoniae* has two main metabolic pathways

The glycolytic pathway leading to lactate
Pathway for glutamine to nitrogen-containing compounds

- ATP consumption and production values listed are for monosaccharides
- ATP values for disaccharide catabolism are in parentheses
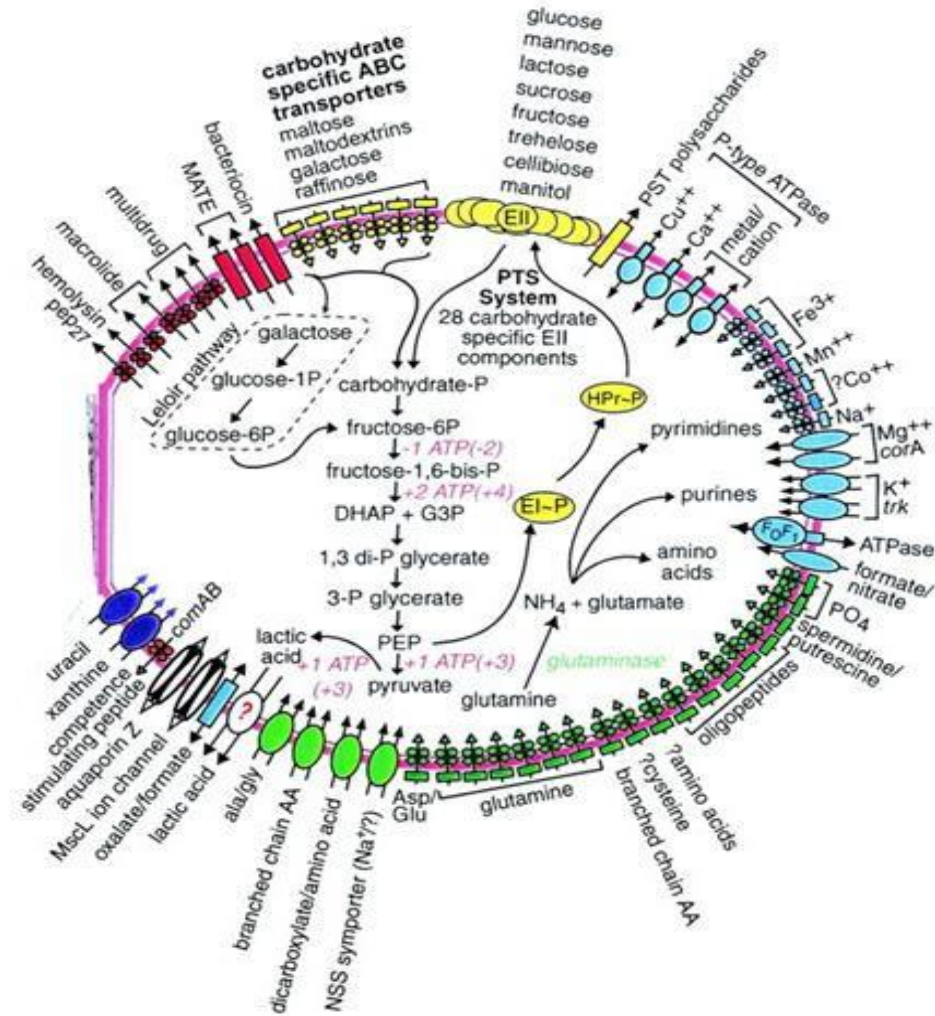
Comparison of the predicted *S. pneumoniae* R6 ORFs with those of other completely sequenced genomes

All S. pneumoniae R6 ORFs were searched against other genomes with BLAST2. The percentage of S. pneumoniae R6 ORFs with significant similarity (E ≤ 10−12) to predicted ORFs in the other genomes is shown.
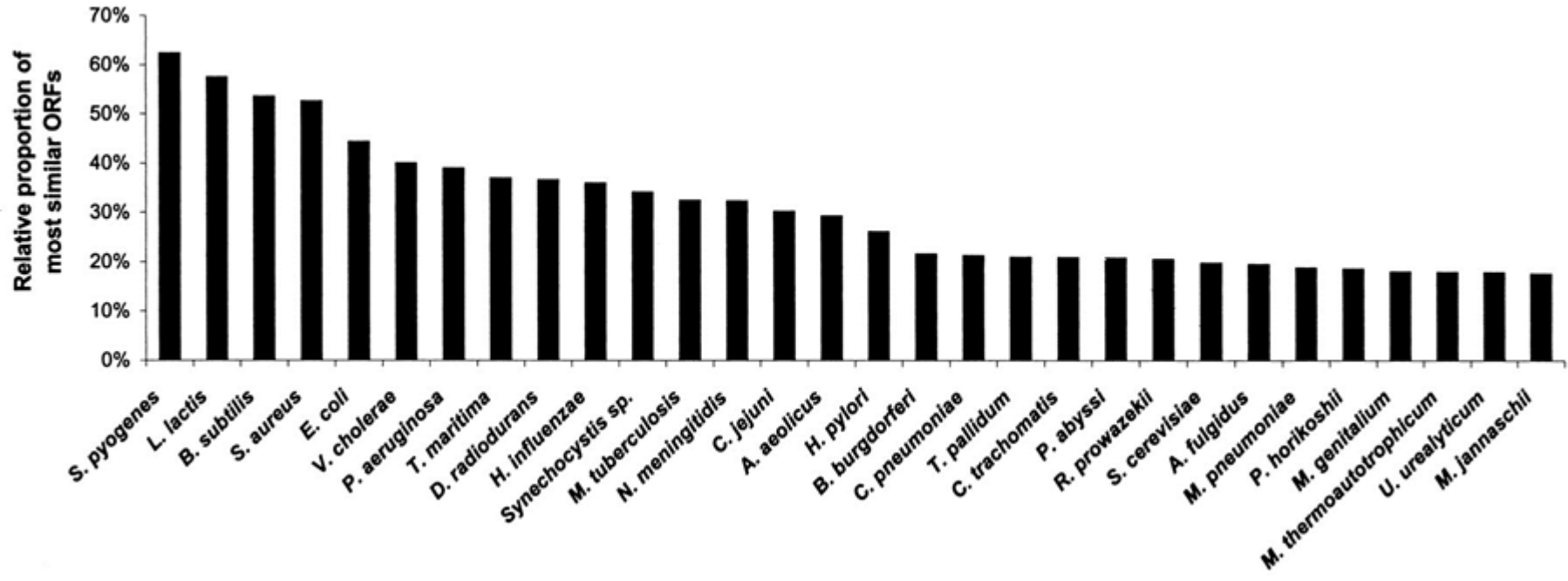
SpR6 (www.streppneumoniae.com) is a model organism database for *S. pneumoniae*

SpR6

UAB

analysis   contact us

welcome

Streptococcus pneumoniae
**The Complete Genomic Sequence
Strain SpR6**

- Journal of Bacteriology 183:5709-5717, 2001 (1016KB)
- ATCC culture number BAA-255
- GenBank accession number AE007317

Important Legal Information The diagnosis and treatment of *Streptococcus pneumoniae* can be performed only by a physician or qualified health care professional. The information on this Web site is intended for informational purposes only. The tools are not designed to provide diagnosis or substitute for treatment. If you have questions or concerns about your overall health, you should consult your doctor.

As a convenience to our visitors, this Web site contains links to a number of sites that we believe may offer useful information. The policies and procedures described in the Privacy Statement do not apply to those sites. We suggest contacting those sites directly for information on their privacy, security, data collection and distribution policies.

The electron micrograph is provided by Dr. Alexander Thomaz, from "*Streptococcus pneumoniae* Molecular Biology and Mechanisms of Disease" Mary Ann Liebert Inc. Publishers, NY, 2000. This figure used with permission from Hoskins, et al. 2001 The genome of the bacterium *Streptococcus pneumoniae* strain R6. J. Bacteriol. 183:5509-5517. Copyright © American Society for Microbiology. All Rights Reserved.

Information Available

- Nucleotide sequences (DNA/RNA)

- Specific gene info

- Strain info

- Medical relevance of strain

- Strain Metabolism

SpR6 database for genes is searchable based on name, protein characteristics, and cellular function
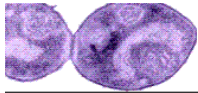

SpR6

UAB
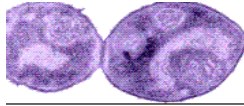
analysis   home

gene search

Gene Search | BLAST Search | Sequence Download | Strain Information | Medical Relevance | Carbohydrate Utilization

Check the **Search Limit** box of the fields you wish to search (check multiple boxes to combine search parameters)
Then enter appropriate parameters and press the **Submit** button to get a list of matching genes

Submit   Reset

Search Limits	Search Field	Search Parameters	
☑	spr name:	%	(e.g. spr0001)
☐	Gene name:		(e.g. dnaA)
☐	Molecule:	⦿ Protein ○ Nucleic Acid ○ Repeat unit	
☐	Genome position:	Start: 1	End: 2038615
☐	Description:		
☐	Notes:		
☐	Molecular Weight:	Greater Than or Equal To: 1000	Less Than or Equal To: 500000
☐	pI:	Greater Than or Equal To: 1	Less Than or Equal To: 13
☐	Prosite motif:	2-oxo acid dehydrogenases acyltransferase component lipoyl binding site	
☐	Role:	Amino acid Biosynthesis	
☐	Subrole:	Amino acid Biosynthesis:Aromatic amino acid	

Search output made sense and was easily interpreted

SpR6

analysis   home

search results

Gene Search | BLAST Search | Sequence Download | Strain Information | Medical Relevance | Carbohydrate Utilization

Number of sequences matching your search parameters: 2315

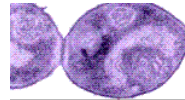| Name	Gene	Molecule	Frame	Start	Finish	Protein Length	Nt Length	Molecular Weight	pI	Description
spr0001	dnaA	p	+	1	1359	453	1359	51,782	5.95	DNA biosynthesis, initiation, binding protein
spr0002	dnaN	p	+	1521	2654	378	1134	42,070	4.89	DNA biosynthesis; sliding clamp subunit, required
spr0003	hypothetical	p	+	2722	2913	64	192	7,480	10.17	Hypothetical protein
spr0004	conserved hypothetical	p	+	3000	4112	371	1113	41,171	4.61	Conserved hypothetical protein
spr0005	pth	p	+	4186	4752	189	567	21,419	9.74	Peptidyl-tRNA hydrolase
spr0006	mfd	p	+	4756	8262	1169	3507	134,801	5.43	Transcription-repair coupling factor
spr0007	conserved hypothetical	p	+	8323	8586	88	264	10,065	10.50	Conserved hypothetical protein
spr0008	hypothetical	p	+	8582	8947	122	366	14,805	10.73	Hypothetical protein
spr0009	conserved hypothetical	p	+	9001	10335	445	1335	51,084	9.28	Conserved hypothetical protein
spr0010	conserved hypothetical	p	+	10335	11609	425	1275	49,966	9.25	Conserved hypothetical protein
spr0011	hgt	p	+	11616	12155	180	540	20,212	5.19	Hypoxanthine guanine phosphoribosyltransferase
spr0012	ftsH	p	+	12174	14129	652	1956	71,254	5.34	Cell-division protein / general stress protein (cl
spr0013	comX1	p	+	14254	14730	159	477	19,873	7.56	Competence-specific global transcription modulator
tRNA-Glu1	tRNA-Glu1	n	+	14827	14898	0	72			tRNA-Glu
rRNA_16S-1	rRNA_16S-1	n	+	15161	16674	0	1514			16S ribosomal RNA
tRNA-Ala1	tRNA-Ala1	n	+	16748	16820	0	73			tRNA-Ala
rRNA_23S-1	rRNA_23S-1	n	+	16945	19846	0	2902			23S ribosomal RNA

- Well organized database, but no help section/tutorials
- Gene id format: spr####

SpR6 is a metadatabase containing nucleotide, gene, and organism data

- Established by John Glass (University of Alabama-Birmingham) in 1991 out of Elliot Lefkowitz's lab
- Has been maintained as part of the UAB Center for AIDS Research and UAB Health Services Foundation

- Was funded by the NIH 1991-2008 (NIH P30 AI027767) and the UAB Health Services Foundation General Endowment Fund (2000-2004).
- Last update was 17 Feb. 2005

- The database is open to the public and contains no restrictions or licensing information
- Initial electronic curation followed by manual quality assurance

SpR6 information can be downloaded as FastA files

- Each gene links to GenBank and the CDC database website

- FastA (text) files available for:
  - Complete nucleotide sequence
  - Protein sequences of predicted ORFs
  - Nucleotide sequences of predicted ORFs
  - RNA sequences
  - Repeat elements

Conclusions

- R6 is an avirulent strain of *S. pneumoniae*, one of the leading causes of death in the US

- Genome sequencing discovered 2,043 protein coding regions, 19 of which may be virulence genes.

- 2% of genes are truncated or non-functional including the capsule synthesis genes

- Nucleotide and protein sequences of *S. pneumoniae* can be found at its model organism database
  - However this information has been poorly maintained and not updated

Reference

Avery, O.T., MacLeod, C.M., & McCarty, M. (1944) Studies on the chemical nature of the substance inducing transformation of pneumococcal types induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. *The Journal of experimental medicine*, 79(2): 137-158.

Hoskins, J., Alborn, W.E. Jr., Arnold, J., Blaszczak, L.C., Burgett, S., Bradley, S., DeHoff, S.T., Estrem, L.F., Fu, D.J., Fuller, W., Geringer, C., Gilmour, R., Glass, J.S., Khoja1, H., Kraft, A.R., Lagace, R.E., LeBlanc, D.J., Lee, L.N., Lefkowitz, E.J., Lu, J., Matsushima, P., McAhren, S.M., McHenney, M., McLeaster, K., Mundy, C.W., Nicas, T.I., Norris, F.H., O'Gara1, M.J., Peery, R.B., Robertson, G.T., Rockey, P., Sun, P.M., Winkler, M.E., Yang, Y., Young-Bellido, M., Zhao, G., Zook, C.A., Baltz, R.H., Jaskunas, S.R., Rosteck Jr., P.R., Skatrud, P.L., and Glass, J.I. (2001) Genome of the bacterium Streptococcus pneumoniae strain R6. *Journal of Bacteriology* 183(19): 5709-5717. doi: 10.1128/JB.183.19.5709-5717.2001