

Using NMPDR Subsystems-based Annotation and Genome Comparison Tools to Understand *Streptococcus pyogenes* Virulome

Details in
poster(s)

Ramy K. Aziz, PhD

**SDSU, San Diego, CA & U. Chicago, IL, USA
Cairo University, Cairo, EGYPT**

Rocky 2008 Meeting: Sun Dec 7 2008

Acknowledgment:

The FIG + MCS Argonne National Lab + NIH (\$\$)

L. Mcneil, R. Overbeek, R. Stevens, V. Vonstein, M. Kotb, Robert A. Edwards

Main Question: Can Pathogenesis Be Computed?

What's a pathogen?

Koch's postulates (1890)

What's a virulence factor?

Molecular Koch's postulates (Stanley Falkow, 1988)/ expression in vivo

Why then is it not trivial to “reconstruct virulence?”

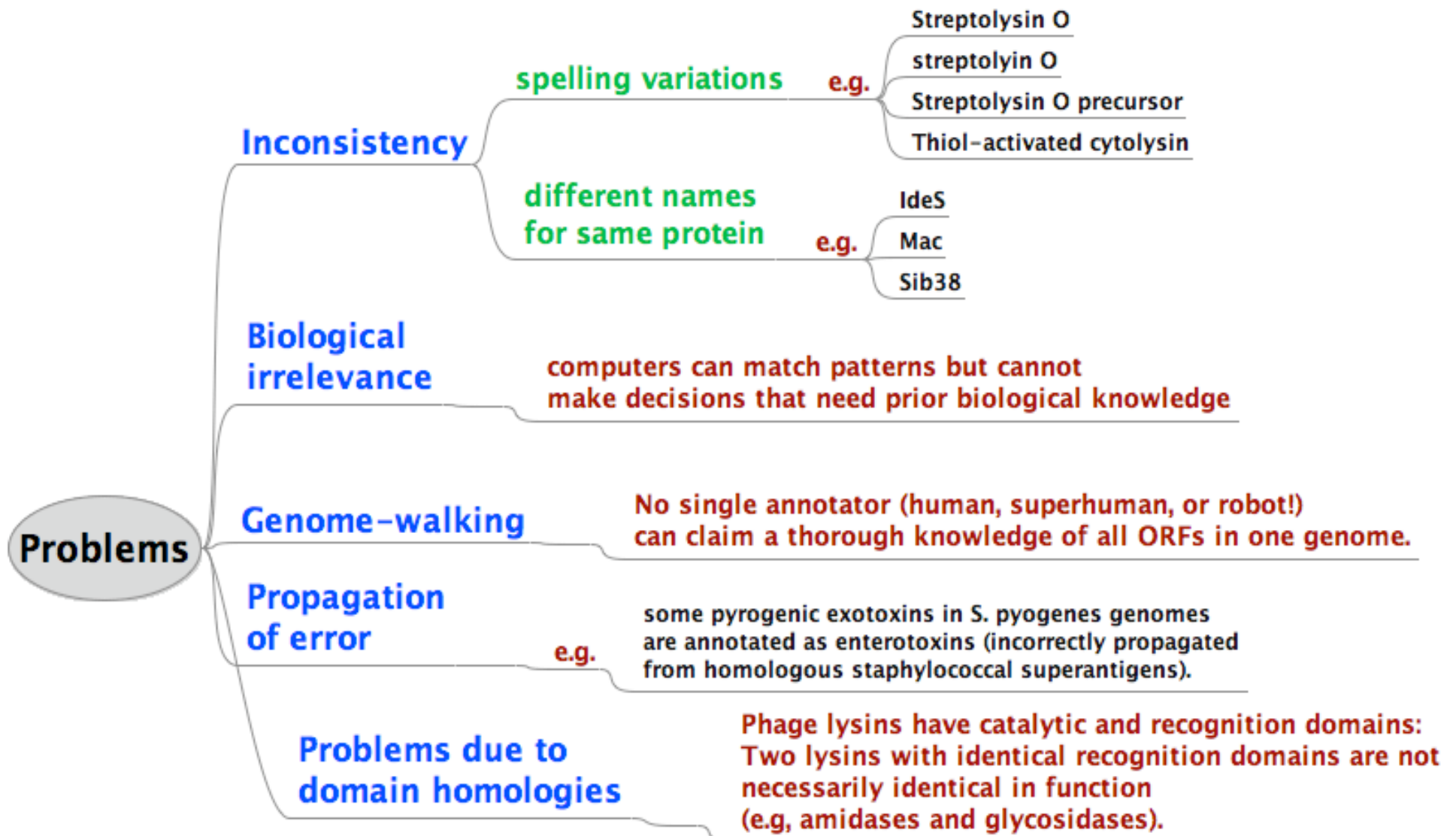
- bi/multifunctional proteins
- opportunistic pathogens/ commensals in secondary niches
- host factors controlling infectious diseases
- complex traits
- are virulence regulators and regulatory RNA virulence factors?
- are essential genes virulence factors?

The key (if any)?

more genomes + better annotations + expression and functional data
in a “pathogen-friendly” database

Platform: NMPDR/ SEED- Another Biological Database!! Why?

Known(?) Problems with current databases:



Platform: NMPDR/ SEED- Another Biological Database!! Why?

Need a platform that...

consistency

biological relevance

**automation/
rapid propagation**

compatibility

offers



<http://www.nmpdr.org>

Using Subsystems to Understand Virulence?!

What's a subsystem?

“A subset of functional roles studied across genomes”

“ A spreadsheet where:

- each row represents a genome
- each column represents a functional role/ feature/ protein”

	Protein 1	Protein 2	...	Protein n
Genome a				
Genome b				
...				
Genome z				

Using Subsystems to Understand Virulence?!

What can fit in a subsystem?

- 1) Chromosomal clusters or virulence operons.
- 2) Pathogenicity islands/ Integrons.
Genome context will reveal them; they may be unique to very few strains.
- 3) Indirectly coupled proteins.
Proteins known to be related or to contribute to **common phenotypes** based on mutational analysis, functional studies, extensive literature mining, etc...(Need wet-lab experts)
- 4) Expression coupling/ Co-expressed proteins.
Members of regulatory networks,
- 5) Scattered proteins sharing motifs/ domains.
e.g., proteins with LPXTG motif (cell wall-anchored), proteins with antibiotic resistance domain.

Using Subsystems to Understand Virulence?!

Examples 1) *S. pyogenes* virulome

An attempt to collect all *S. pyogenes* virulence proteins in comparison to other Streptococci

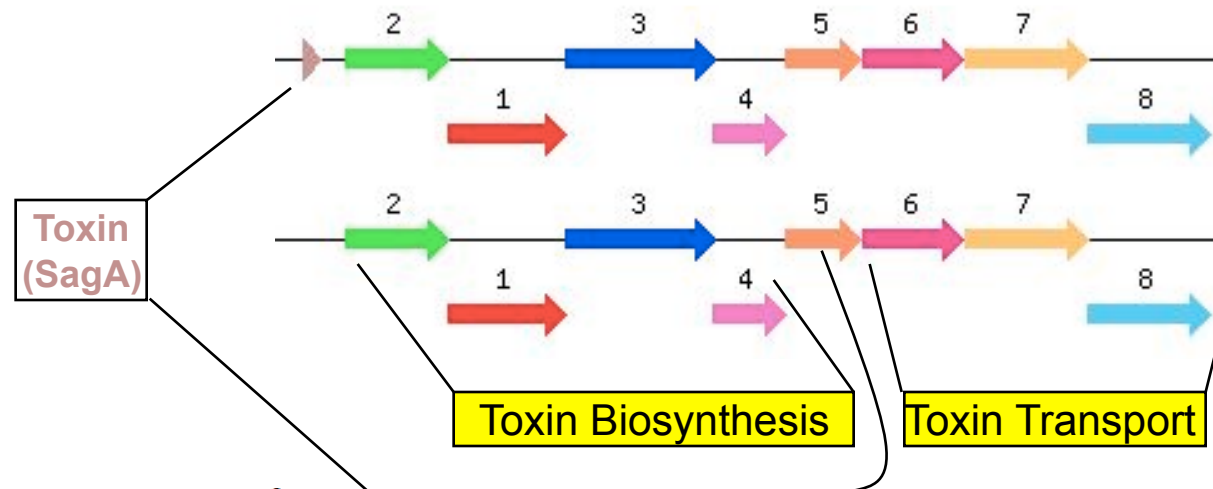
Looking at the subset “Exotoxins: Superantigens”: Strain-specific patterns

Organism	Variant Code	SpeA	SpeB	SpeC	SpeG	SpeH	SpeI	Spd1	SSA	SpeK	SpeL	SpeJ	SmeZ	psSpeG	psSpeJ
Streptococcus agalactiae NEM316 [B]	0001_B														
Streptococcus equi subsp. zooepidemicus [B]	0110						911	1758		476	475				
Streptococcus pyogenes M1 GAS [B]	1111_A01 M1	1562	520	159	763	1719	1566					312	1536		
Streptococcus pyogenes MGAS5005 [B]	1111_A01 M1T1	1308	382		274			385				635	349		979
Streptococcus pyogenes MGAS315 [B]	1111_A03 M3	1301	1742		155			1745	920	1205			1715, 1716		
Streptococcus pyogenes SSI-1 [B]	1111_A03 M3	560	1739		161			1743	1119	2041					
Streptococcus pyogenes M5 [B]	1111_A05 M5		568	1484	91	936	934, 1888	566					595		792
Streptococcus pyogenes MGAS10394 [B]	1111_A06 M6	2061	1735	1196			740	2088		985			2108	217	
Streptococcus pyogenes MGAS8232 [B]	1111_A18 M18	305	1723	609	159			1727			1012		1699		
Streptococcus pyogenes MGAS6180 [B]	1111_A28 M28		536	1353	273			540		1604		583	502		1032

Using Subsystems to Understand Virulence?!

Examples 2) Streptolysin-S (SagA) operon in *S. pyogenes*

- SagA and SagF are linked.
- In M5 pyogenes, the ORF was missed.
- In *Clostridium*, the toxin is not there. What else could be transported?
- In *Listeria*, only 3 genes are conserved.



Organism	Variant Code	SagA	SagB	SagC	SagD	SagE	SagF	SagG	*ABC_permease
<i>Clostridium botulinum</i> ATCC 3502 [B]	0102		67	68	69	70		72	73-10 , 74-10
<i>Listeria monocytogenes</i> str. 4b F2365 [B]	0100		1103	1104	1105				
<i>Listeria monocytogenes</i> str. 4b H7858 [B]	0100		613	612	611				
<i>Streptococcus equi</i> subsp. zooepidemicus [B]	0111		837	838	839	840	841	842	843-8 , 844-9
<i>Streptococcus pyogenes</i> M1 GAS [B]	1111	540	541	542	543	544	545	546	547-8 , 548-9
<i>Streptococcus pyogenes</i> M5 [B]	0111		686	687	688	689	690	691	692-8 , 693-9
<i>Streptococcus pyogenes</i> MGAS10394 [B]	1111	578	579	580	581	582	583	584	585-8 , 586-9
<i>Streptococcus pyogenes</i> MGAS315 [B]	1111	480	481	482	483	484	485	486	487-8 , 488-9
<i>Streptococcus pyogenes</i> MGAS5005 [B]	1111	874	875	876	877	878	879	880	881-8 , 882-9
<i>Streptococcus pyogenes</i> MGAS6180 [B]	1111	925	926	927	928	929	930	931	932-8 , 933-9

Future: a Project to Integrate 1 million transcriptomes/ proteomes/ phenomes

Basic Spreadsheet

	Genome ID	Organism	Variant Code	RR	HK	ISP	MAP	Mga	Emm	Scp
<input type="checkbox"/>	211110.1	Streptococcus agalactiae NEM316 [B]	0			1692				1971 , 449
<input type="checkbox"/>	1336.1	Streptococcus equi subsp. zooepidemicus [B]	0			2147			5	
<input checked="" type="checkbox"/>	160490.1	Streptococcus pyogenes M1 GAS [B]	0	1555	1554	1553	1552	1551	1550	1547
<input type="checkbox"/>	1314.1	Streptococcus pyogenes M5 [B]	0	578	579	580 , 237	581	583	585	586
<input type="checkbox"/>	286636.1	Streptococcus pyogenes MGAS10394 [B]	0	1725	1724	1523 , 1723		1720	1719	1718
<input type="checkbox"/>	198466.1	Streptococcus pyogenes MGAS315 [B]	0	1733	1732	1731	1730	1728	1727	1726
<input type="checkbox"/>	186103.1	Streptococcus pyogenes MGAS8232 [B]	0	1715	1714	1713	1712	1711	1710	1709
<input type="checkbox"/>	193567.1	Streptococcus pyogenes SSI-1 [B]	0	1730	1929	1728	1727	1726	1725	1724
<input type="checkbox"/>	1349.1	Streptococcus uberis [B]	0					1771	1770	

Sort spreadsheet genomes by ▾

Pick transcriptomes for checked genome(s)

Transcriptome for selected Genome:	Variant	RR	HK	ISP	MAP	Mga	Emm	Scp
Streptococcus pyogenes M1 GAS [B]		1555	1554	1553	1552	1551	1550	1547
Growth at 37C	100	10	50	0	-30	-20	-10	0
Growth at 42C	110	11	44	2	-12	-23	-14	0
Growth in THB medium	-1		0		-1			
Growth in Iron-deficient medium	101	49	99		100	0	100	100
In vivo BALB/c mice	200	-20	-30	100	34	55	66	100
Taken from TSS patients	200	-21	-35	100	33	88	90	90
Taken from pharyngitis patients	300	-10	-20	-5	0	0	10	-10