



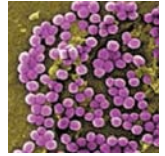
Steps toward broad-spectrum therapeutics: discovering virulence-associated genes present in diverse pathogens

Chris Stubben¹, Melanie Dufflied² and Murray Wolinksy¹

¹ Bioscience Division, LANL, Los Alamos, NM

² Biomedical Sciences, Dstl, Porton Down, UK

New strategies for antibiotic resistance



- New drugs and therapeutics are needed to treat the emergence of bacterial strains that are resistant to available antibiotics.
- An alternative to killing bacteria is to “disarm” bacteria by interfering with their capacity to be virulent.
- Targeting virulence factors may enable a pathogen to survive and evoke an appropriate immune response, and also avoids disruption of the host flora and selection for antimicrobial resistance.

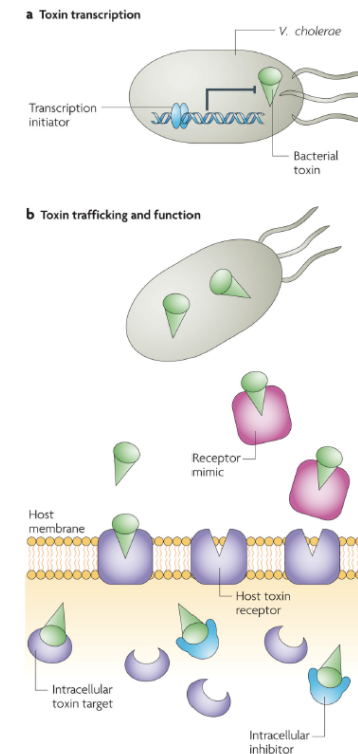
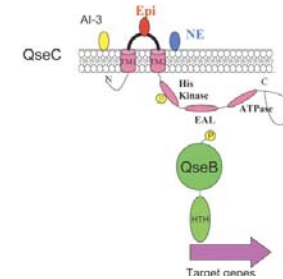


Figure 3. Targeting toxin-powered pathogens
a | The inhibition of toxin transcription, as described for *Vibrio cholerae*, is one way to inhibit the consequences of toxin-mediated virulence. b | Neutralizing toxins, or preventing their trafficking and/or enzymatic activity, at cellular targets is an alternative strategy to inhibit toxin damage to the host.

From Cegelski et al 2008

Problem 1: how to identify virulence genes

- **Homology based** searches to “known” genes involved in pathogenic processes, e.g., regulators of virulence gene expression (QseC signaling in *Science* 22 August 2008).
- Predictions based on **genomic context**, e.g., genes that are present in pathogens and absent in non-pathogens (*phylogenetic profiles*)

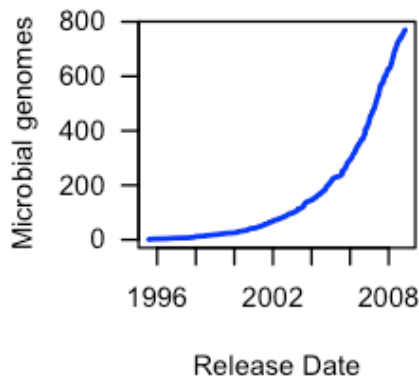


Gene	Al-3	Epi	NE	His Kinase	EAL	ATPase	QseB	PTH	Target nucleic acid
1 pflon0004	1	1	1	1	1	1	1	1	1
2 pflon0005	1	1	1	1	1	1	1	1	1
3 pflon0006	1	1	1	1	1	1	1	1	1
4 pflon0009	1	1	1	1	1	1	1	1	1
5 pflon0010	1	1	1	1	1	1	1	1	1
6 pflon0011	1	1	1	1	1	1	1	1	1
7 pflon0012	1	1	1	1	1	1	1	1	1
8 pflon0013	1	1	1	1	1	1	1	1	1
9 pflon0015	1	1	1	1	1	1	1	1	1
10 pflon0016	1	1	1	1	1	1	1	1	1
11 pflon0018	1	1	1	1	1	1	1	1	1
12 pflon0023	1	1	1	1	1	1	1	1	1
13 pflon0024	1	1	1	1	1	1	1	1	1
14 pflon0027	1	1	1	1	1	1	1	1	1
15 pflon0028	1	1	1	1	1	1	1	1	1
16 pflon0032	1	1	1	1	1	1	1	1	1
17 pflon0033	1	1	1	1	1	1	1	1	1
18 pflon0034	1	1	1	1	1	1	1	1	1
19 pflon0035	1	1	1	1	1	1	1	1	1
20 pflon0036	1	1	1	1	1	1	1	1	1
21 pflon0037	1	1	1	1	1	1	1	1	1
22 pflon0041	1	1	1	1	1	1	1	1	1
23 pflon0042	1	1	1	1	1	1	1	1	1
24 pflon0043	1	1	1	1	1	1	1	1	1

In particular, we were interested in known or novel virulence factors that are conserved across a broad range of biothreat pathogens

Problem 2: how to search profiles

- Almost 800 complete microbial genomes have been sequenced (plus 600 WGS sequences).
 - **Abundance profiles** at Integrated Microbial Genomes (IMG) from JGI allow users to select columns from a drop-down list of 900 genomes (good luck!)
 - We also created profiles tables using **Database cross-References** in UniProt files from 800 complete genomes.
- Genomes should be linked to accurate and detailed information on phenotypes, host-pathogen interactions, etc., for easy selection of target and background genomes



JGI
DOE JOINT GENOME INSTITUTE
A DIVISION OF LANL

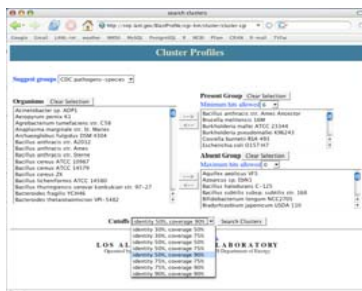
Quick Genome Search:

img/m INTEGRATED MICROBIAL GENOMES
with MICROBIOME SAMPLES

Select	Row No.	ID	Name	Bac ant Ams	Bac hal C-5	Bra jap 119	Bru met 16M	Bur pse K93	Clo ace 924	Cox bur 493	Eri lul 9-44	Geo PCA	Man MBE	Mes MA9	Nit C71	Pho pro 559	Shi file 301	Vib sho N11	Yer pen CO2	Rif len DJA	Pro mar 925
<input type="checkbox"/>	1	pfam00004	AAA	11	11	13	8	16	16	8	7	13	8	10	11	8	8	9	11	8	12
<input type="checkbox"/>	2	pfam00005	ABC_tran	105	95	200	102	111	94	20	21	52	46	219	36	86	73	64	105	54	27
<input type="checkbox"/>	3	pfam00006	ATP-synt_ab	4	4	6	4	9	4	3	3	4	3	5	6	7	5	4	7	3	2
<input type="checkbox"/>	4	pfam00009	GTP_EFTU	6	5	8	8	9	9	7	7	11	9	9	8	10	9	9	8	6	6
<input type="checkbox"/>	5	pfam00010	HLH	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
<input type="checkbox"/>	6	pfam00011	HSP20	2	1	11	2	2	1	1	1	6	0	8	1	2	2	1	2	1	0
<input type="checkbox"/>	7	pfam00012	HSP70	2	1	1	1	2	3	1	1	1	2	1	1	1	3	3	2	2	1
<input type="checkbox"/>	8	pfam00013	KH_1	2	2	3	2	1	2	1	1	3	1	3	1	1	1	1	1	1	1
<input type="checkbox"/>	9	pfam00015	MCPsignal	13	12	36	0	21	28	0	0	24	0	1	3	39	5	45	9	0	0
<input type="checkbox"/>	10	pfam00016	RuBisCO_large	1	0	1	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1

Discovering virulence-associated genes

- Step 1. Cluster into gene families
 - clustered all against all Blast results from 617,000 proteins at 50% identity and 90% coverage (77,000 total clusters/rows and 214 genomes/columns)
- Step 2. Search for virulence genes (www.tvfac.lanl.gov)
 - Selected 12 representative strains of CDC category A and B biothreat agents as targets and 52 non-pathogens as background
 - Allowed minimum of 5 hits to target and maximum of 3 hits to background



Genomes linked to detailed information on host-pathogen interactions

Search Clusters			
Absent Hits	Present Hits	Clusters	Select
0	9	1	<input type="checkbox"/>
0	8	1	<input checked="" type="checkbox"/>
0	7	9	<input checked="" type="checkbox"/>
0	6	20	<input type="checkbox"/>
1	8	4	<input type="checkbox"/>
1	7	17	<input type="checkbox"/>
1	6	119	<input type="checkbox"/>

Display Selected Rows

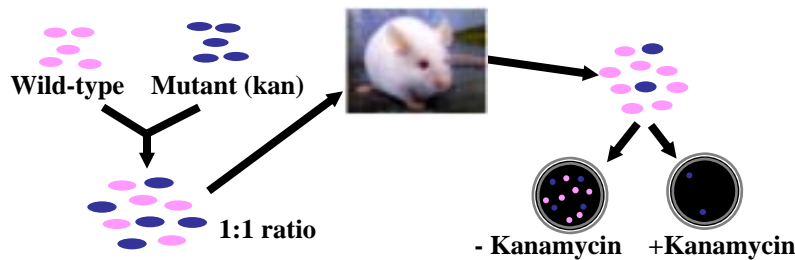
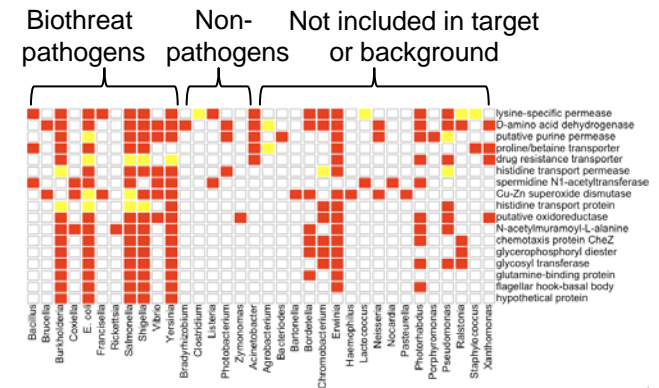
Results sorted based on hits to target genomes and missing in background genomes

		Non-pathogens			
		0	1	2	3
Pathogens	5	368	125	44	33
	6	20	119	145	82
	7	9	17	13	20
	8	1	4	4	14
	9	1	-	2	3

=1024 candidates

Discovering virulence-associated genes

- **Step 3. Narrow down candidates**
 - Manual down-selection of profile candidates to a manageable number for experimental testing (13 total)
- **Step 4. Experimental screening**
 - Disrupt genes and assess virulence in mutant strains of *Y. pseudotuberculosis* (with DstI)



- **Results are promising...**
 - Six (out of 13 tested) genes with reduced virulence in YPTB (42%)

YPTB Locus	CI	Cluster
2410	0.01	Magnesium transport ATPase mgtB
2913	0.03	Deoxyribopyrimidine photolyase phrB
2705	0.06	Putative manganese transport protein mntH
3827	0.12	Outer membrane biogenesis protein
0181	0.13	Putative protohaeme IX biogenesis protein hemY
0756	0.16	Superoxide dismutase C sodC
1340	0.39	Lysine specific permease cadR
0242	0.42	Glycerophosphoryl diester phosphodiesterase glpQ
2699	0.45	Hypothetical protein
1424	0.76	Hypothetical protein
3166	0.87	Thiol:disulphide interchange protein dscB
0188	2.38	Fratxin-like protein cyaY
3505	3.15	Stringent starvation protein sspB