

Genome perspectives of *Mycoplasma* pathobiology

Michael Calcutt
University of Missouri-Columbia

Acknowledgements

University of Missouri

Kim Wise	Mark McIntosh	Thomas Reilly
Kerstin Roeske	Tendai Mutangadura	
Mark Foecking	Manijeh Heidari	
Tara Martin	Miranda Hvinden	Bethany Kent

J. Craig Venter Institute

John Glass	Shibu Yooseph	Barbara Methe
Ramana Madupu	Robert DeBoy	

Funding: USDA-ARS and NIH

Overview

Genomic analysis of *M. bovis* PG45

Deduced features

Mobile Elements

Surface Lipoprotein genes of *M. bovis*

Antigenic/phase variable proteins

Surface protein genes

Phase variable antigens of *M. sp. bovine* group 7

M. bovis PG45 genome

Univ. of MO and TIGR/JCVI

- Recognised by USDA as an emerging pathogen
- Calf pneumonia, otitis, arthritis and mastitis
- Chronic disease, poor response to therapy

Genome status- Annotation complete

Images courtesy of Dr. Mary Brown, Univ. Florida



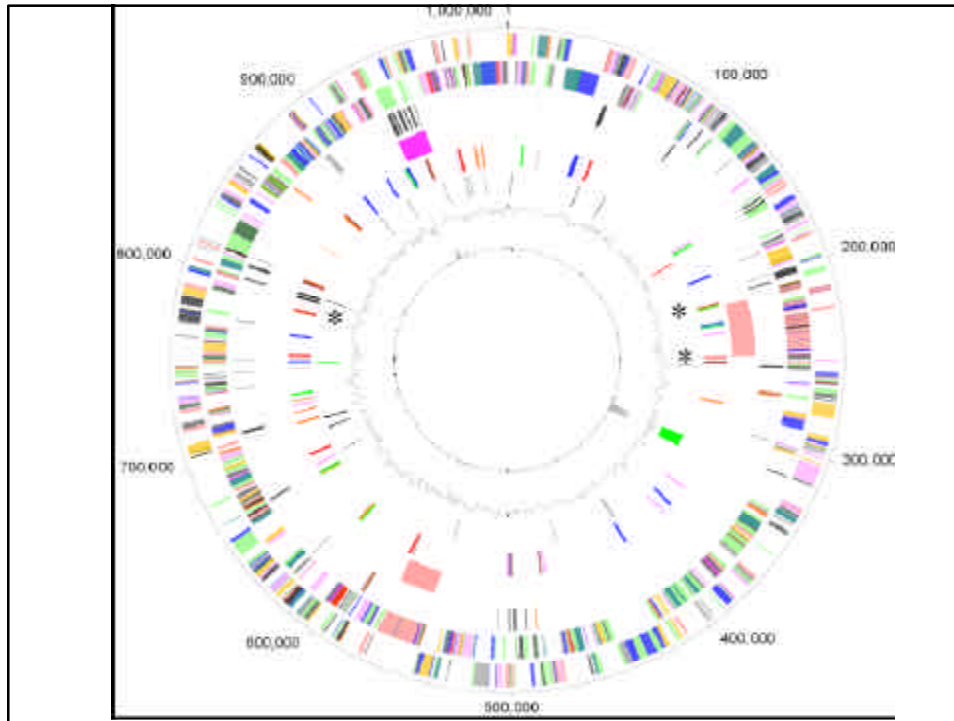
M. bovis PG45 genome summary

Genome Size	1,003,404 base pairs
G+C content	29.3%
Genes	868
ORFs	827
Gene density	89%
Coverage	8X
IS units	56 (~100 kb)
Lipoproteins	96
Hypothetical ORFs	89*
Regulatory proteins	3?

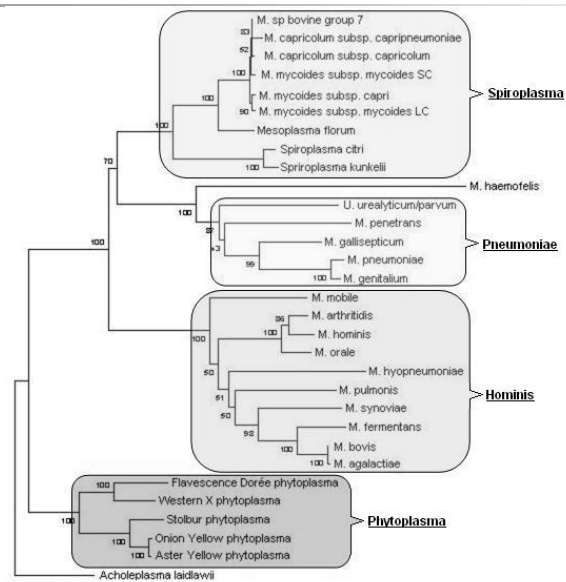
Insertion Sequence Inventory of *M. bovis* PG45

IS	Copies ⁺	Size (bp)	Total	homology
<i>ISMbov1</i>	12	1,521	16.2 kb	<i>ISMagal1</i> (92%)
<i>ISMbov2</i>	8	1,671	12.5 kb	<i>ISMmy1</i> (97%)
<i>ISMbov3</i>	7	1,873	12.9 kb	<i>IS1634</i> (97%)
<i>ISMbov4</i>	8	1,642	13.1 kb	<i>ISMmy1</i> (34%)
<i>ISMbov5</i>	3	1,388	3.2 kb	<i>ISMmy1</i> (32%)
<i>ISMbov6</i>	6	1,395	8.4 kb	<i>ISMpen</i> (42%)
<i>ISMbov7</i>	10	1,343	10.7 kb	UU372 (60%)
Total:	54 (43 complete)		~ 77 kb	(7.7% of genome)

⁺ Includes complete and partial IS units



Phylogeny of *Mollicutes* (16S rRNA)



Genome comparisons reveal gene synteny and inversions

M. agalactiae
100 kb

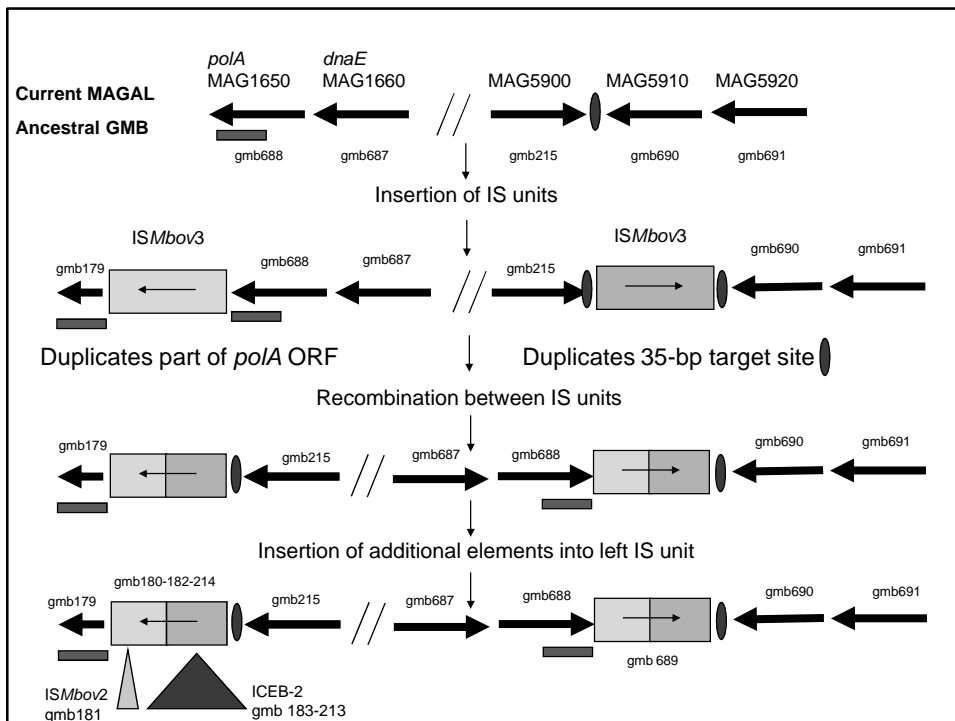
M. agalactiae
700 kb



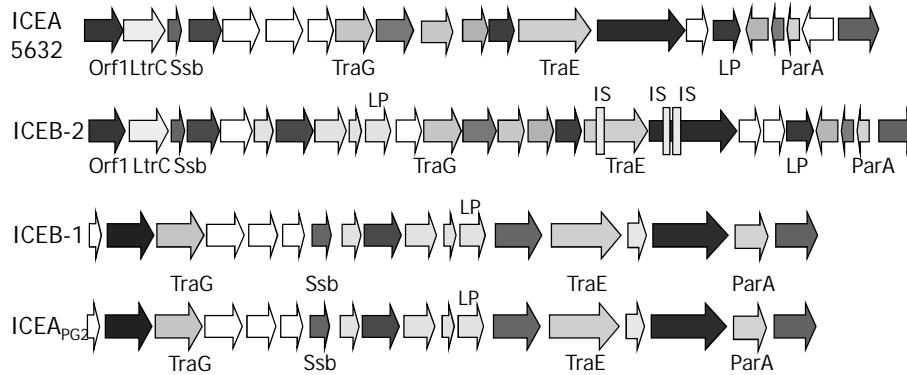
M. bovis
150 kb

M. bovis
800 kb

ACT analysis—regions of strong DNA homology joined by red lines
- inverted regions are shown in blue

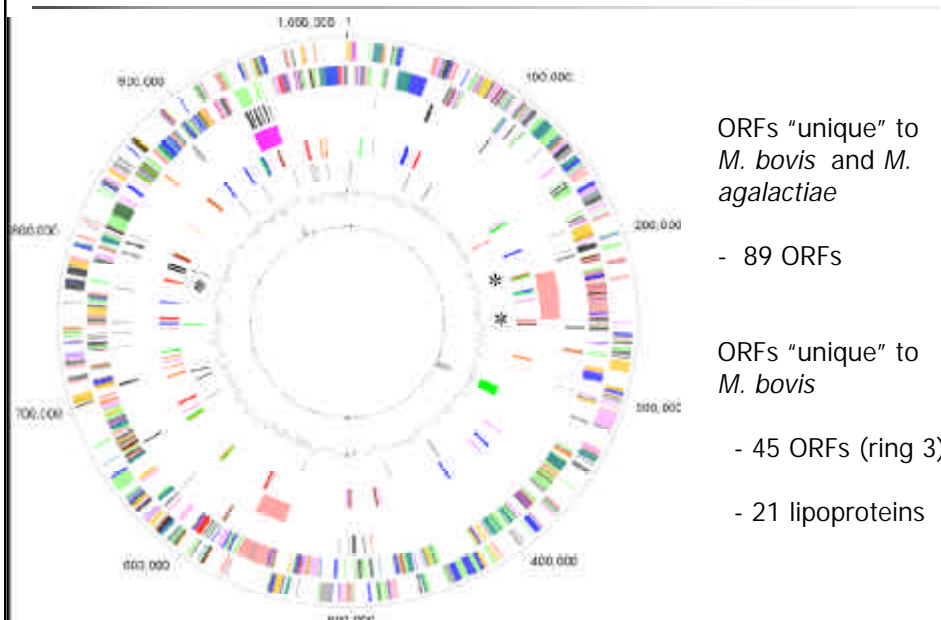


M. agalactiae and *M. bovis* contain Integrative Conjugal Elements



- ICEA and ICEB-2 are very closely related (75-94% aa identity)
- Each ICEB unit is flanked by 8-bp direct repeats
- ICEA_{PG2} and ICEB-2 exhibit strong homology-identical insertion site
- extrachromosomal forms of each unit can be detected

How many genes are “unique” to *M. bovis*



M. bovis contains a paucity of identifiable regulatory functions

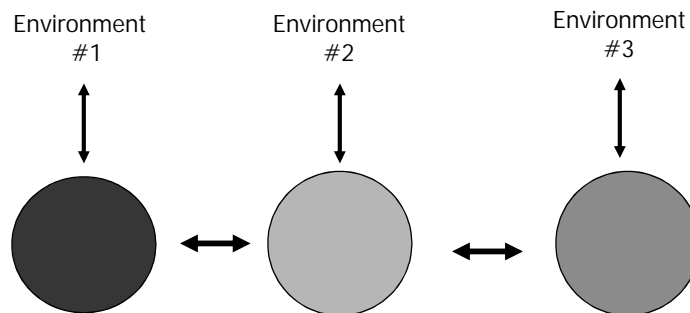
Transcriptional Regulation

- Single sigma factor for promoter recognition
- Heat shock repressor, HrcA
- LacI family repressor

- Absence of two component phosphorelay systems
- No homologs for quorum sensing regulation

- Post-transcriptional regulation?
- Translational regulation? Riboswitches?

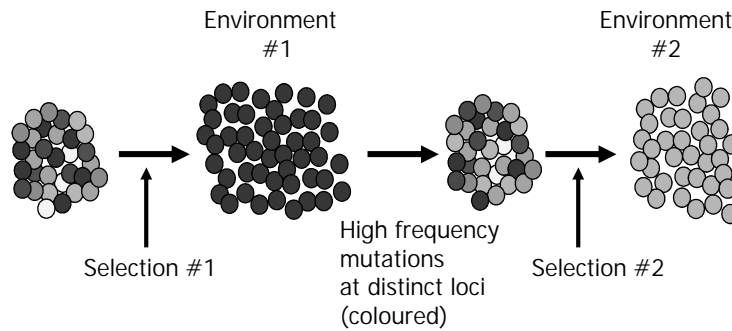
Adaptation Plan A: A single organism adapts to alternative niches through responsive networks of regulated genes



Gene regulation in mycoplasmas: Do they use plan A?

- Classical regulatory gene sets and networks are sparse
- Regulation may/may not be reflected in transcript levels
- Role of regulation in adaptive strategies is barely explored

Adaptation Plan B: A population adapts by generating diversity through phase variable mutations in contingency genes



Features of phase variation:

- Heritable genotypic changes
- Reversible genotypic changes
- High frequency

The *Mycoplasma* Surface and Antigenic Variation

- Membrane is unique interface between mycoplasma and host
- Components of the "exoproteome" perform diverse functions
 - adherence
 - immunomodulation
 - nutrient acquisition
 - immune evasion
- Mycoplasmas deploy systems of antigenic variation to generate diversity
- Proposed functions:
 - escape immune surveillance; host adaptation (tropism)

M. bovis encodes an extensive repertoire of surface lipoproteins

- Modified LP signal peptide (“lipobox”) search pattern used (based on unusual residues at the -1 position of expressed LPs)

[LIVMFWSTAG](2)-[LIVMFYSTAGCQ]-[AGSKRQT]-C

- Modified search pattern identified 38 additional LP genes
- Total putative lipoproteins = 96 (>10% of predicted proteome)
- Modified search pattern may be clade specific and applies to *M. agalactiae*

M. bovis encodes an extensive repertoire of surface lipoproteins

Putative lipoproteins = 96 (> 10% of the predicted proteome)

Orthologs of immunodominant surface proteins

LppA	<i>M. mycoides</i> cluster
LppB (3 paralogs)	<i>M. mycoides</i> cluster
LppD (2 paralogs)	<i>M. mycoides</i> cluster
P48 [MALP] (2 paralogs)	diverse species
P80	<i>M. agalactiae</i>

Potential antigenic targets and virulence factors

- Vsp family of surface lipoproteins
 - adhesins
 - biofilms ?
 - immunological shielding? modulation?
- Glycerol dependent hydrogen peroxide production?

M. bovis is the only species to have 2 *gts-lppB* clusters
- Surface hydrolases: nucleases, proteases, lipases, phosphatase
- Immunomodulins?

M. bovis expresses diverse lipopeptide structures

M. bovis encodes a large repertoire of potentially variable lipoproteins

Putative lipoproteins = 96 (> 10% of the predicted proteome)

Variable due to high frequency mutations:

13 Vsp phase variable LPs (transcriptional; site-specific recombination)

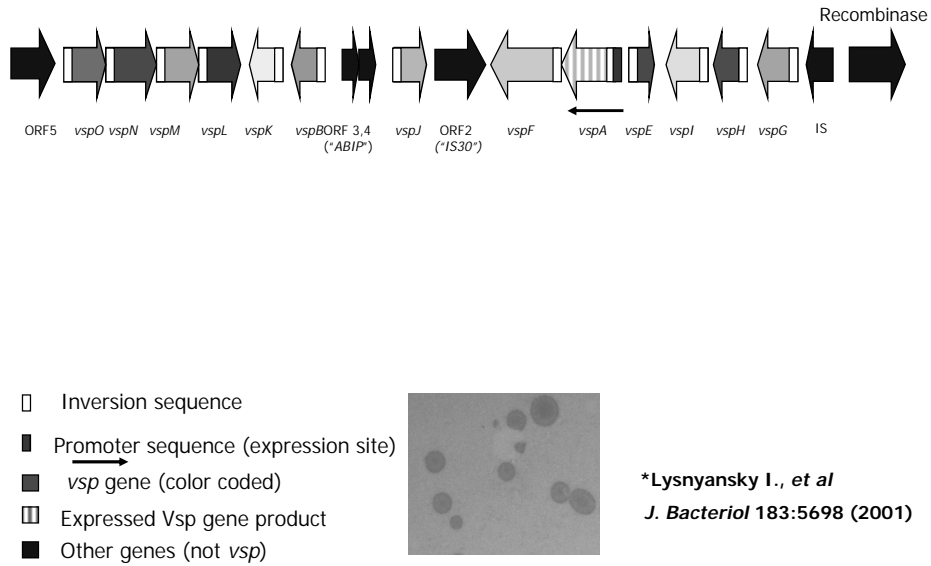
PLUS:

3 putative phase variable LPs (transcriptional; 5' poly [CT] tracts)

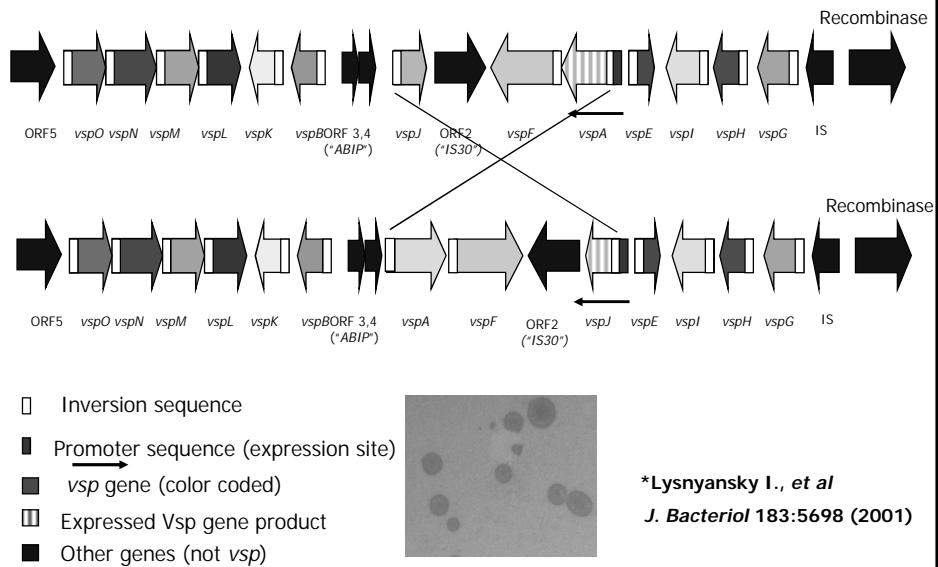
21 putative phase-variable LPs (translational; poly [G or C] frame-shifts)

10 putative LPs have tandem repeats in coding region

The *vsp* locus: a hotspot for DNA inversions

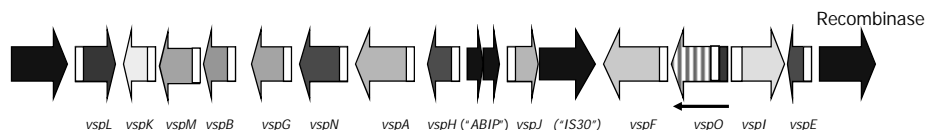


The *vsp* locus: a hotspot for DNA inversions

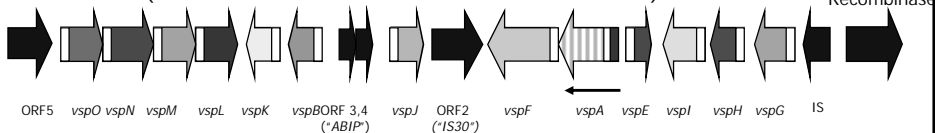


The *vsp* locus: extensive intra-strain complexity

STRAIN PG45 ATCC 25523 CLONE A2 (GENOMIC SHOTGUN SEQUENCE)



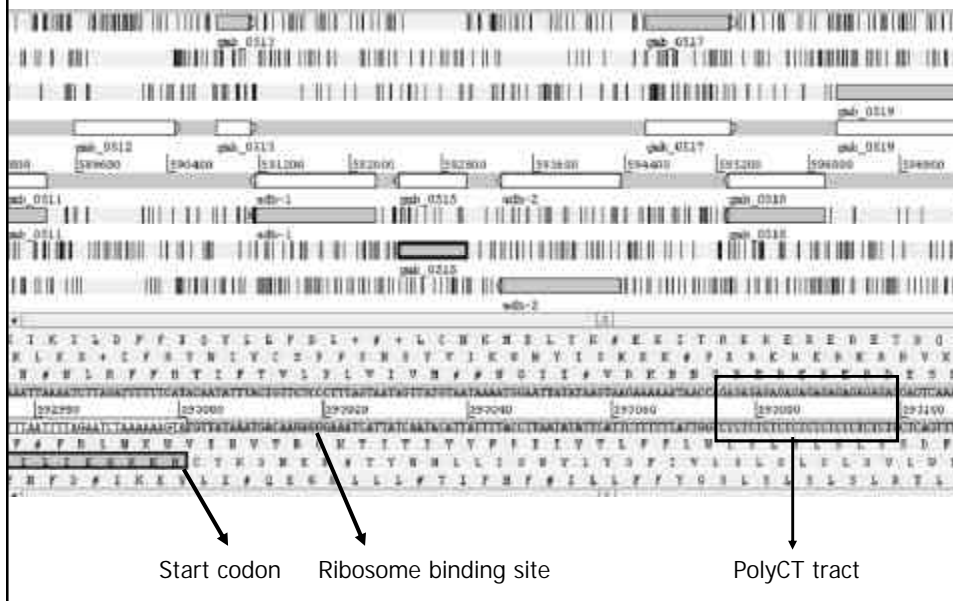
STRAIN PG45 (PREVIOUSLY SEQUENCED BY MOLECULAR CLONING) *



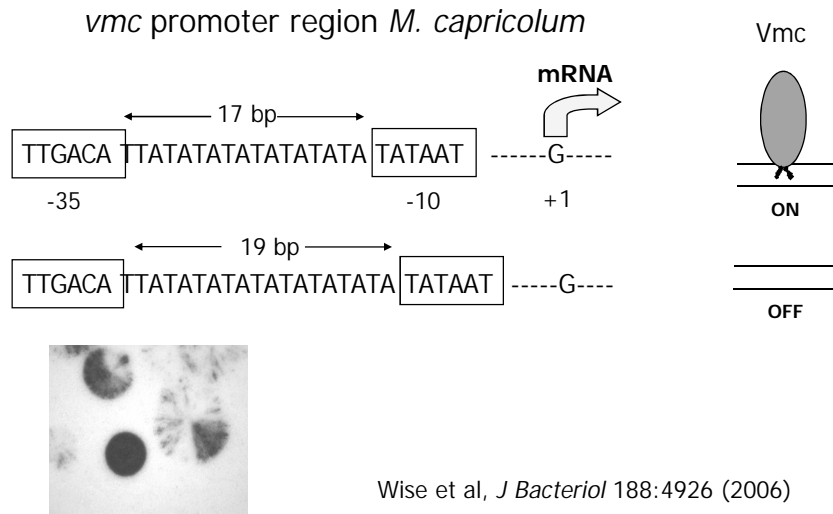
- Inversion sequence
- Promoter sequence (expression site)
- *vsp* gene (color coded)
- ▨ Expressed Vsp gene product
- Other genes (not *vsp*)

*Lysnyansky I., et al
J. Bacteriol 183:5698 (2001)

PolyCT tracts are indicative of phase variation

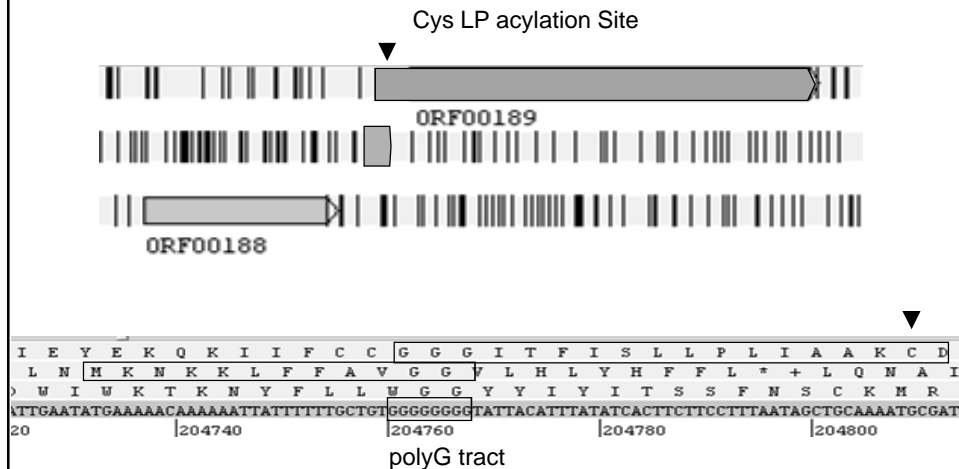


High frequency indels mediate phase variation (transcription)



PolyG tracts are indicative of phase variation

e.g. ORF189 Lipoprotein



M. bovis genome has 19 putative LP genes with polyG or polyC tracts

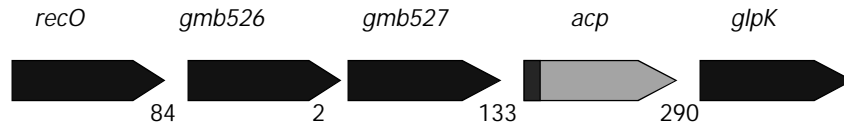
Lessons for vaccine development?

- The organism is an antigenic moving target
- Genome annotation provides a database to map the expressed proteome and immunoproteome
- Genome sequence provides a platform for comparative analysis of field isolates
- Genetic manipulation may provide some control of surface variation
- Post-genomic studies are complicated by inherent instability and mutability by error-prone mechanisms

“From Sequence to Consequence”



Identification of a putative surface phosphatase



- *gmb528* (*acp*) : a putative lipoprotein (463 amino acids)
- BLAST analysis: ORF has a limited region of homology to acid phosphatases
- The signal peptide is atypical with T at the -1 position

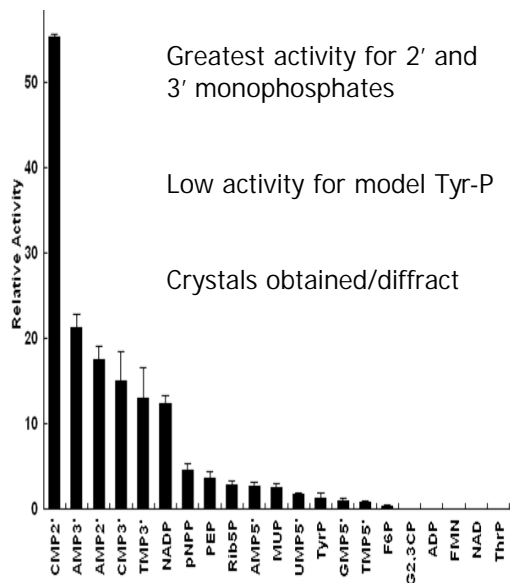
MTNRKFFKGMLIGSAPVILLPTI AAT C

Purification of recombinant phosphatase

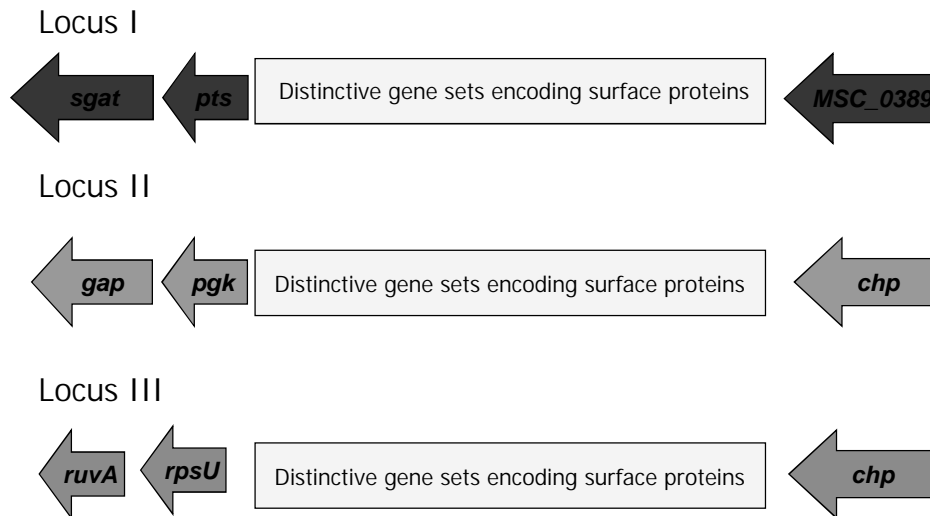


Synthetic gene in pET20b (T7)

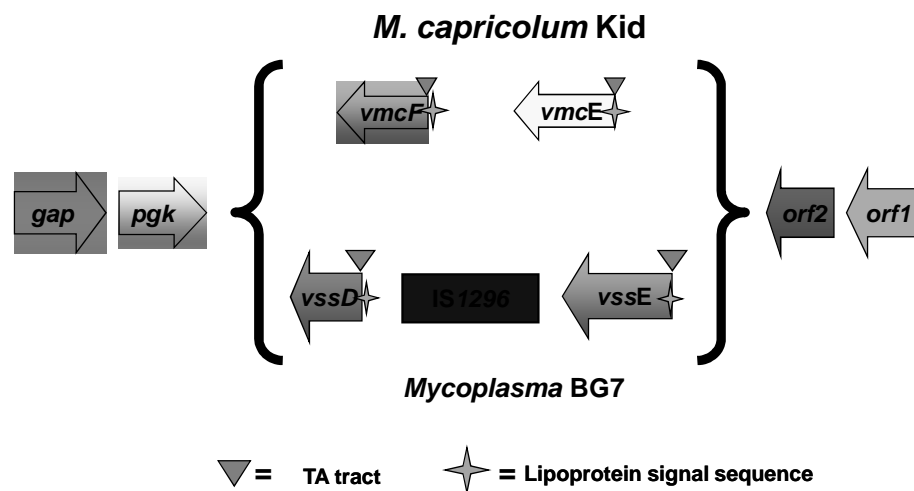
1 NiTA column 2 SP-Sepharose



Contingency Loci-variable genes in a conserved genomic location



Locus II



Miranda Hvinden

