

## 1. Comparative Genomics:

- ACT exercise

Orthologue of  
*E. coli* b0169

Orthologue of  
*E. coli* b0170

Orthologue of  
*E. coli* b0171

Artemis Entry Edit: S\_typhi.dna

File Entries Select View Goto Edit Create Write Run Graph Display

Selected feature: bases 54 misc\_feature (/colour=8 /note="PS01066 Undecaprenyl pyrophosphate synthetase family signature")

Entry: ☐ S\_typhi.dna ☒ S\_typhi.tab

0 247200 248000 248800 249600 250400 251200 252000 252800 253600 254400 255200 256000 256800 257600

misc\_feature

STY0237

STY0238

STY0239

STY0240

STY0241

STY0242 dxr

STY0244

STY0245

I C A V \* C R A D Q I F R G F + N G \* N Y R I P G K R A A \* A Y R R R H D G L Q K S A D \* S  
S V R F D A G Q I K S F E D F R M A E I T A S L V K E L R E R T G A G M M D C K K A L T E  
L C G L M P G R S N L S R I L E W L K L P H P W # K S C V S V P A Q A \* W I A K K R \* L K  
TCTGTGCGGTTTGATGCCGGGCAGATCAAACTCTTCGAGGATTTAGAAATGGCTGAAATACCGCATCCCTGGTAAAGAGCTGCGTGAGCGTACCGGCGCAGGCATGATGGATTGCAAAAAAGCGCTGACTGAAC  
252120 252130 252140 252150 252160 252170 252180 252190 252200 252210 252220 252230 252240 252250  
AGACACGCCAAACTACGGCCCGTCTAGTTTAAAGCTCCTAAATCTTACCGGACTTTAATGGCGTAGGGACCATTTTCTCGACGCACTCGCATGGCCGCGTCCGTACTACCTAACGTTTTTTCGCGACTGACTTC  
Q A T Q H R A S \* I K R P N # F P Q F # R M G P L L A A H A Y R R L C S P N C F L A S Q L  
D T R N S A P C I L D K S S K L I A S I V A D R T F S S R S R V P A P M I S Q L F A S V S A  
R H P K I G P L D F R E L I K S H S F N G C G Q Y F L Q T L T G A C A H H I A F F R Q S F

CDS	254032	254589	Orthologue of <i>E. coli</i> frf (RRF_ECOLI); Fasta hit to RRF_ECOLI (185 aa), 94% identity in 185 aa overlap
CDS	254731	255927	Similar to Escherichia coli 1-deoxy-D-xylulose 5-phosphate reductoisomerase dxr SW:DXR_ECOLI (P4556)
misc_feature	255064	255087	PS00017 ATP/GTP-binding site motif A (P-loop)
CDS	256240	256998	Orthologue of <i>E. coli</i> UPPS ECOLI; Fasta hit to UPPS ECOLI (253 aa), 94% identity in 252 aa overlap
misc feature	256804	256857	PS01066 Undecaprenyl pyrophosphate synthetase family signature
CDS	257011	257868	Fasta hit to YNBB ECOLI (298 aa), 33% identity in 298 aa overlap
misc_feature	257725	257805	PS01315 Phosphatidate cytidylyltransferase signature
CDS	257880	259232	Similar to Escherichia coli hypothetical 49.1 kDa protein in cdsa-hlpa intergenic region yaeL SW:YAE
misc_feature	257934	257963	PS00142 Neutral zinc metalloproteinases, zinc-binding region signature

# Representing sequence similarity/identity

```
fasta results for STY0122 from /nfs/disk222/yeastpub3/Salmonella_typhi/whole_genome/./old_whole_genome/fasta/St.tal
10      20      30      40      50      60
STY012 MQALLEHFITQSTLYSLIAVLLVAFLESALVGLILPGTVLMAGLGALIGSGELNFWHTW
BAA013      .X:.....
      10      20      30      40
      70      80      90      100     110     120
STY012 LVGIIGCLMGDWISFWLGRFKKPLHRWSFMKKNKSLLDKTEHALHQHSMFTILVGRFVG
BAA013      .....
      50      60      70      80      90      100
      130     140     150     160     170     180
STY012 PTRPLVPMVAGMLDLPVAKFIGPNLIGCLLWPPFYFLPGILAGAAIDIPSDMQSGDFKWL
BAA013      .....
      110     120     130     140     150     160
      190     200     210     220     230     240
STY012 LLATAALLWVGWLCWRLWRSKAAVDRLTAYLPRSRLLYLAPLTIGVVALVVRHP
BAA013      .....
      170     180     190     200     210     220
      250
STY012 LMPVYIDILRKVVGY
BAA013      .....
      230
Close Send to browser
```

MQALL...

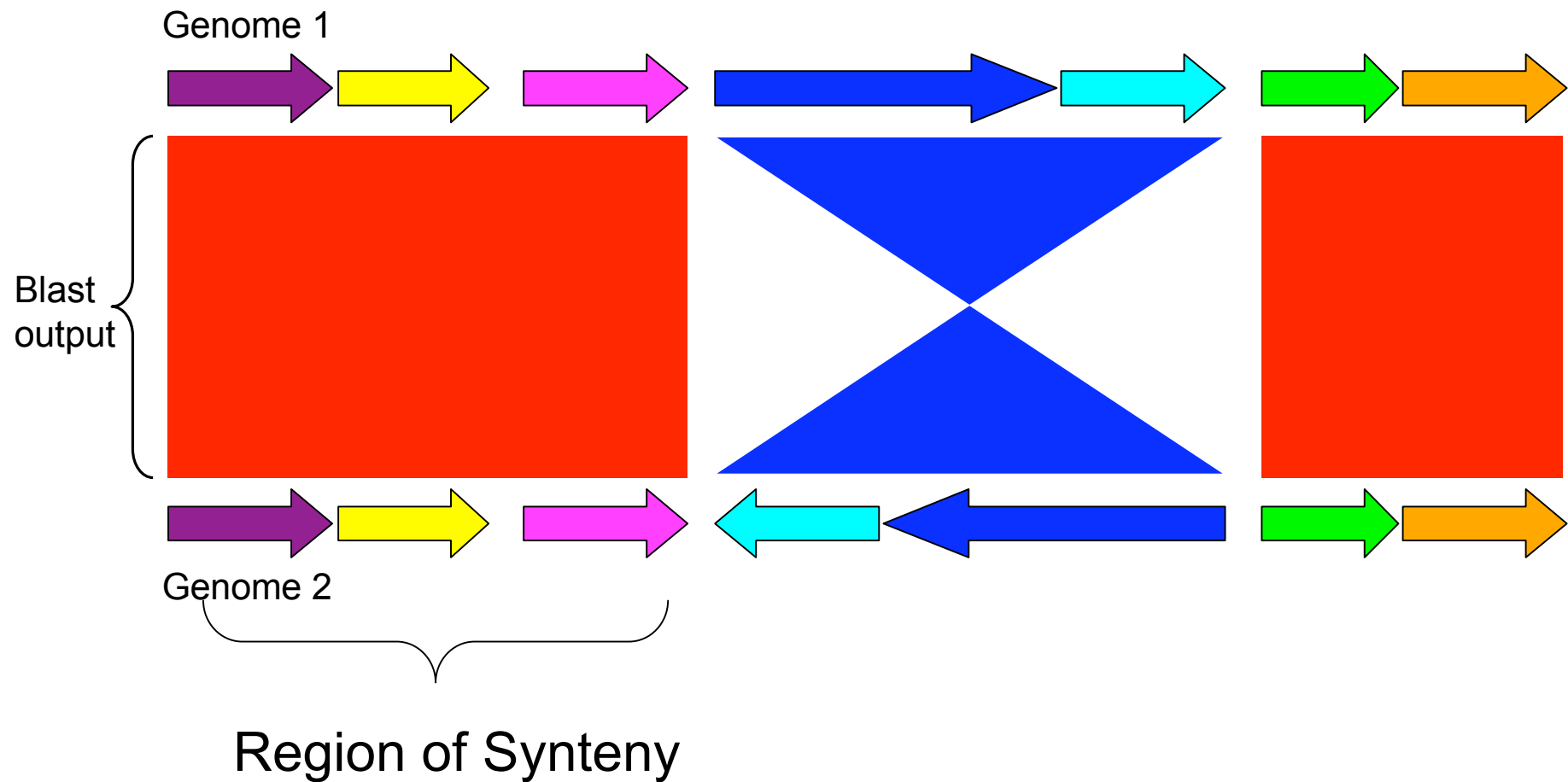
...RHP

Region sharing similarity  
(BlastP)

MAVV...

...RHP

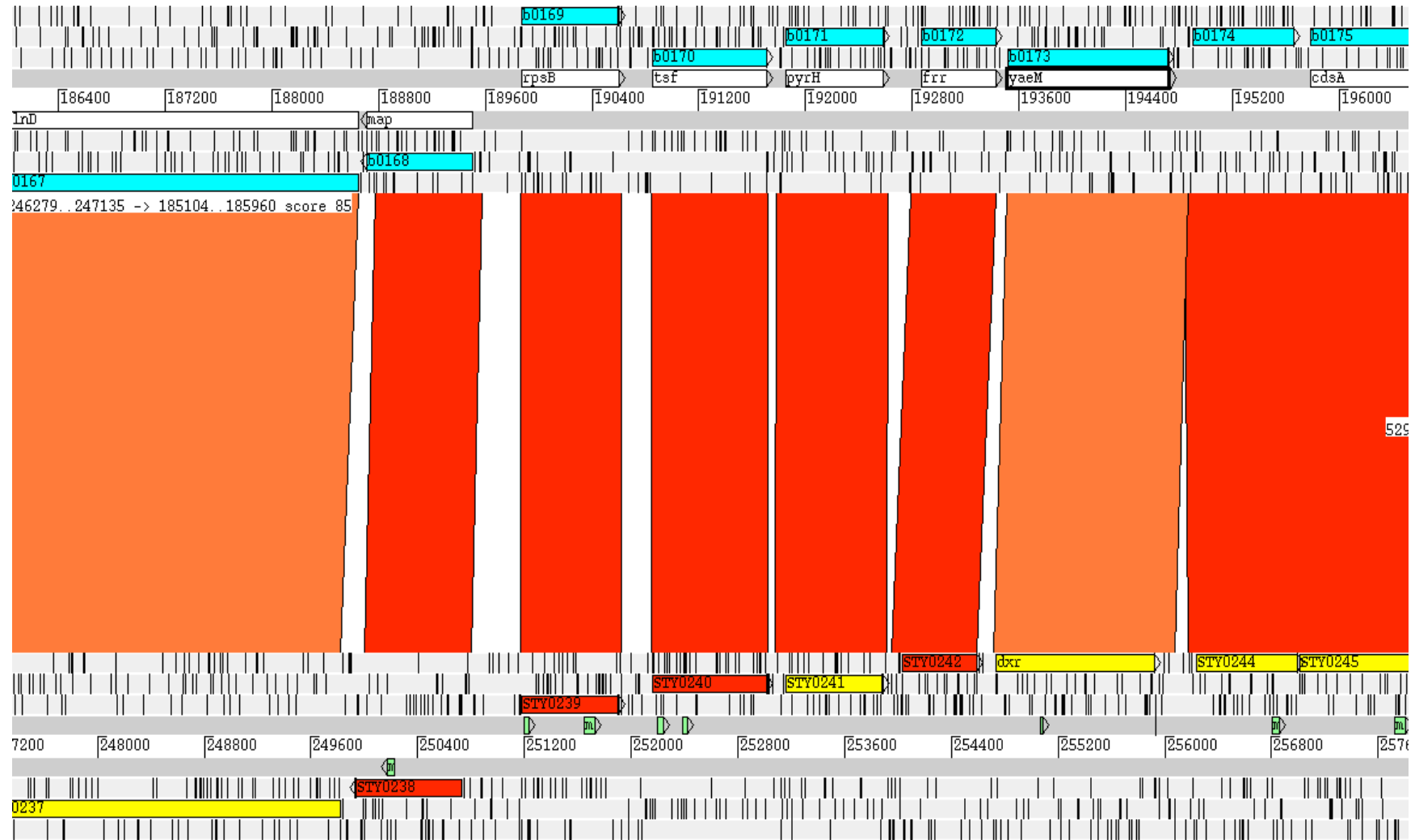
# ACT - Artemis Comparison Tool





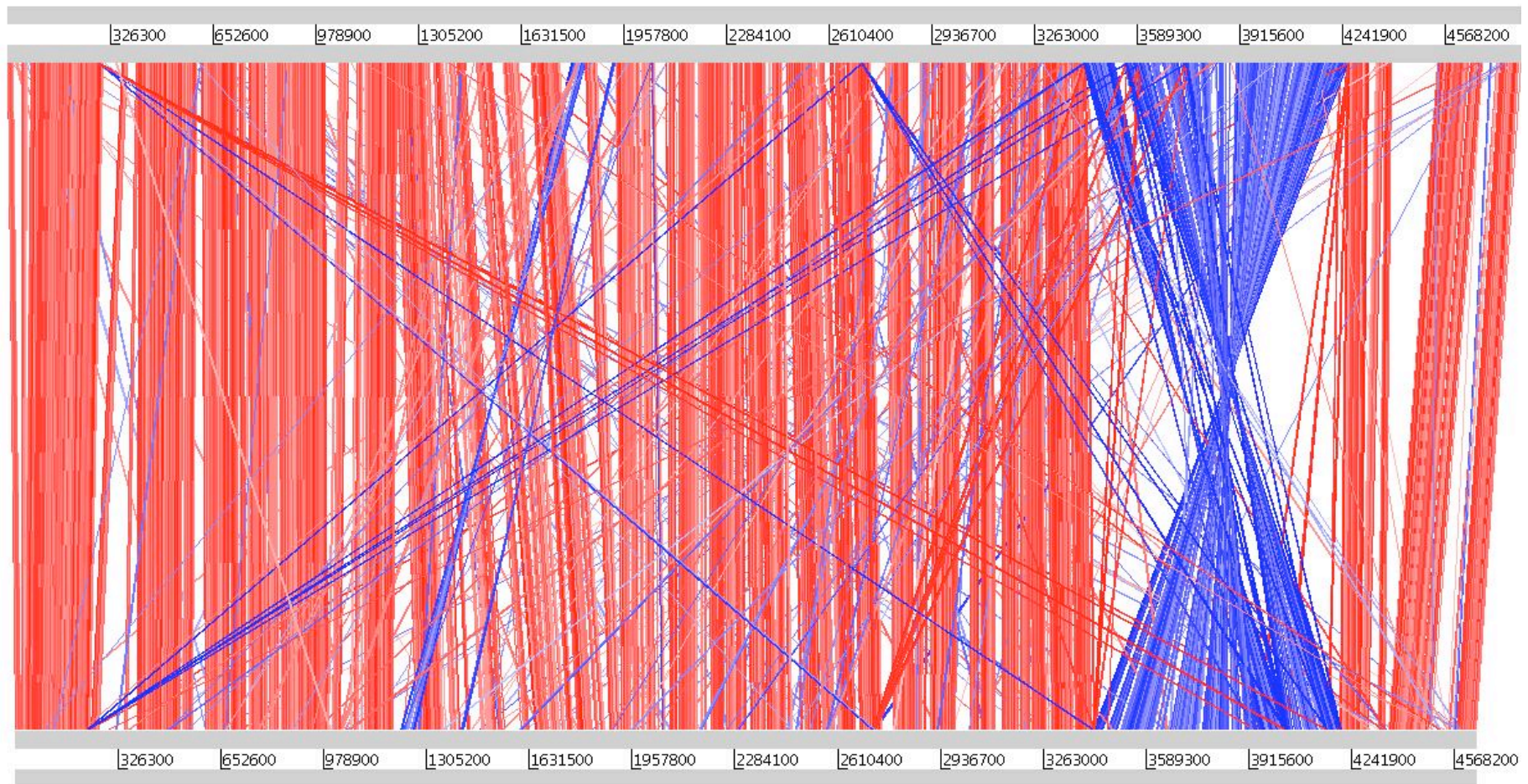


## ACT- Artemis Comparison Tool

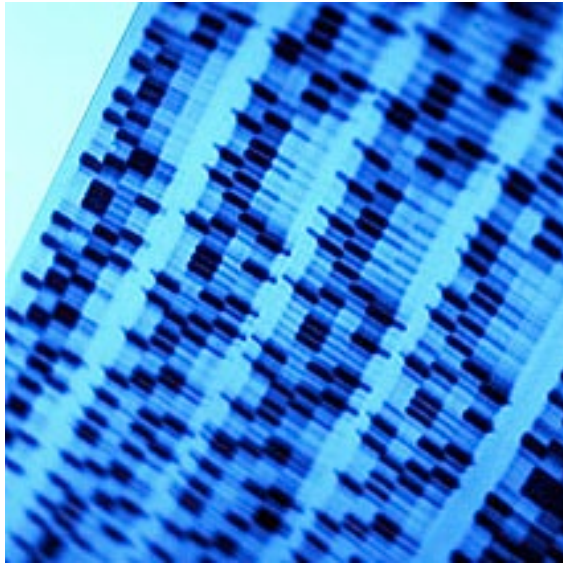


## *Salmonella Typhi* vs. *E. coli* – ACT

**S. Typhi**



***E. coli***



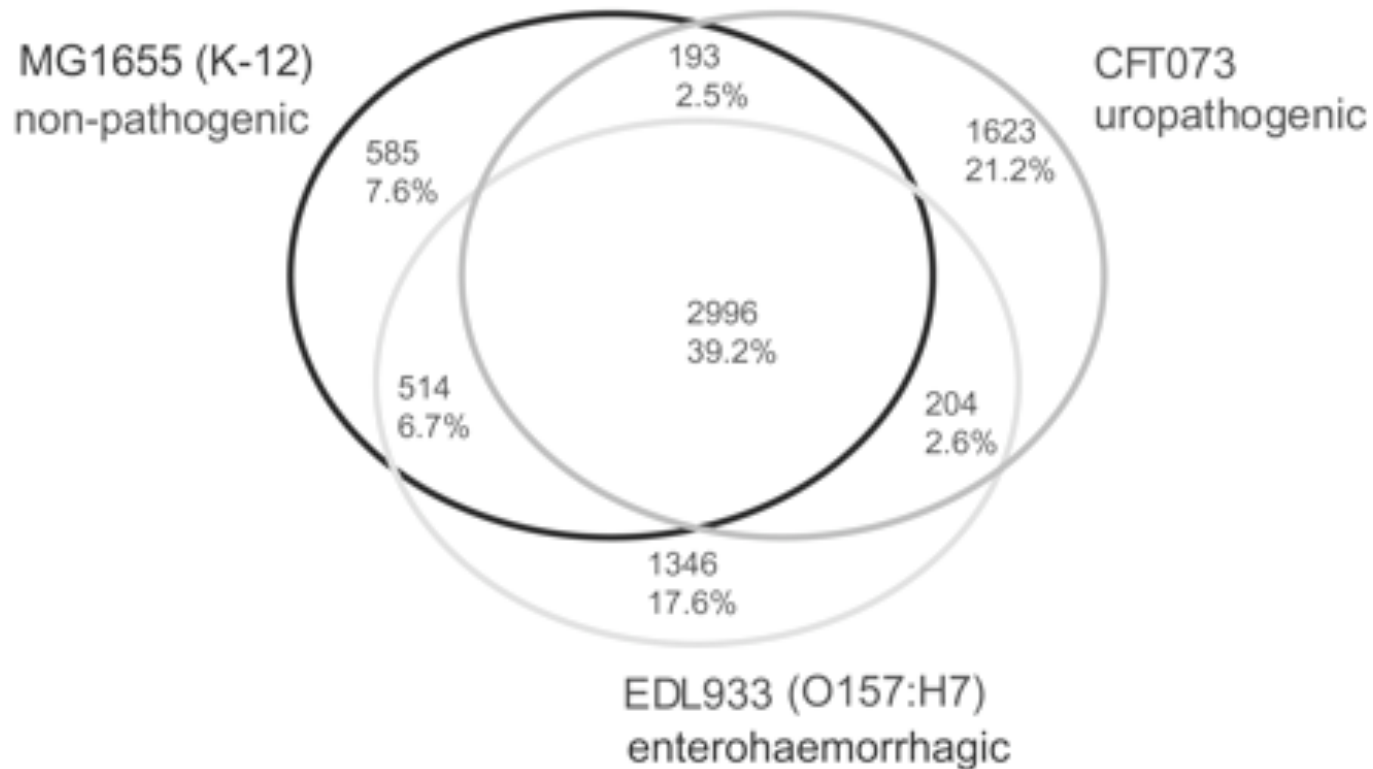
## 1. Comparative Genomics:

- Core and accessory genome
- Pathogenomics



Back to the conserved and unique functions

Where do they fit into pathogen evolution?





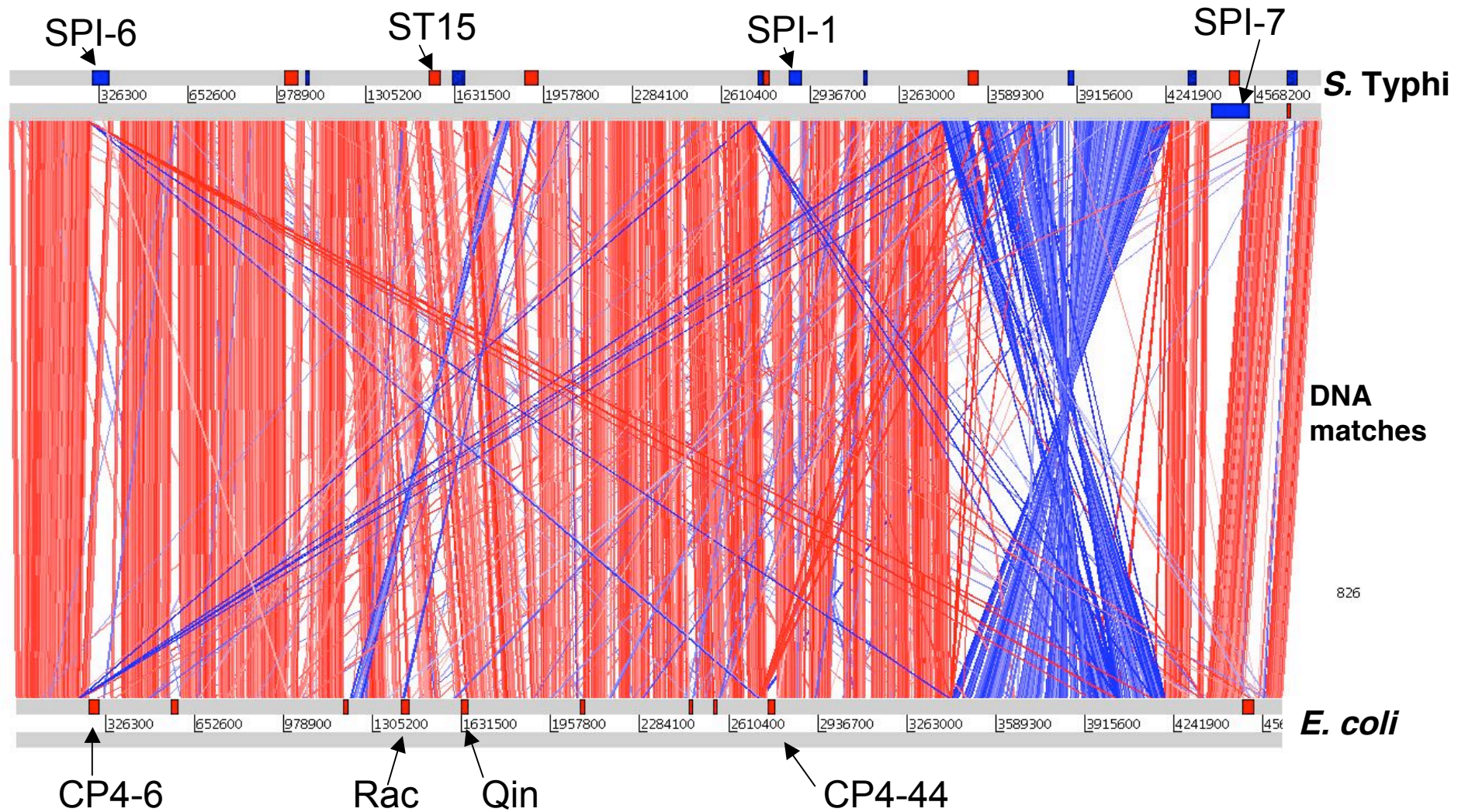
# Genome Content

- Core - Stable
  - House keeping genes
    - Central metabolism
    - Transporters
    - Cell division
    - Replication and information transfer
  - Regulators
  - Surface proteins
  - Virulence factors
- Accessory - Variable
  - Miscellaneous metabolism
  - Virulence factors
  - Drug resistance
  - Phage
  - PAI's or Genomic islands
  - Plasmids

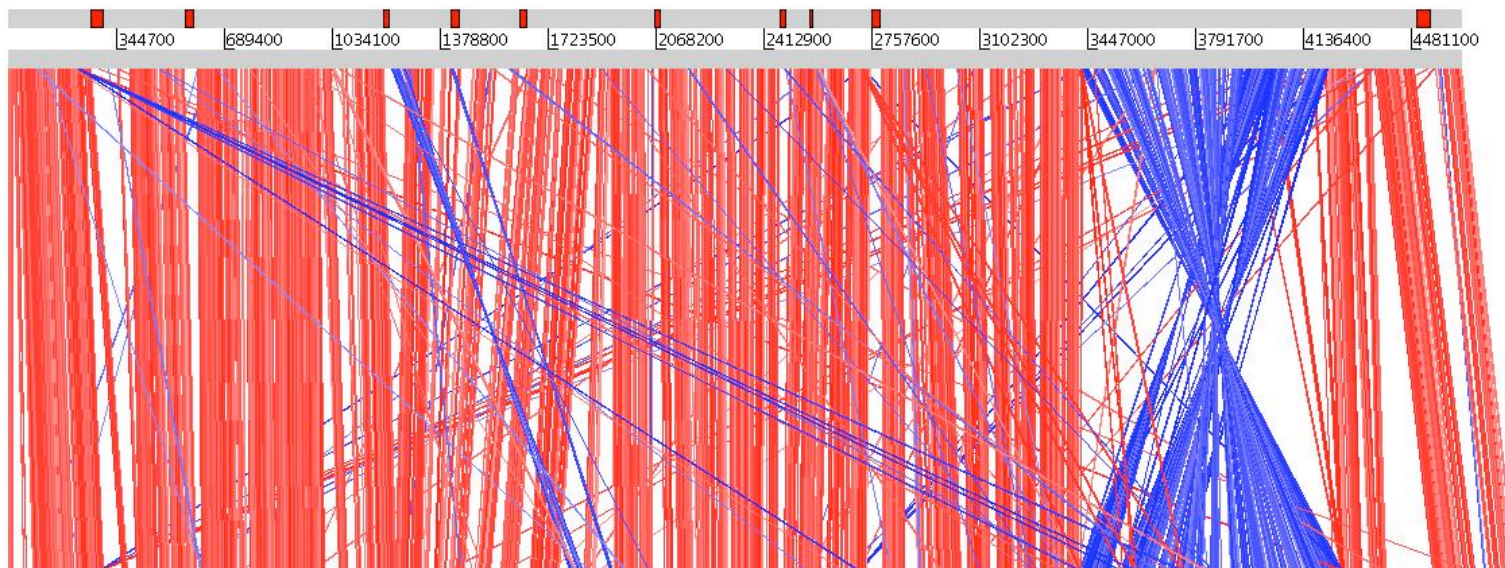


# *Salmonella* vs. *E. coli* - core and accessory genomes

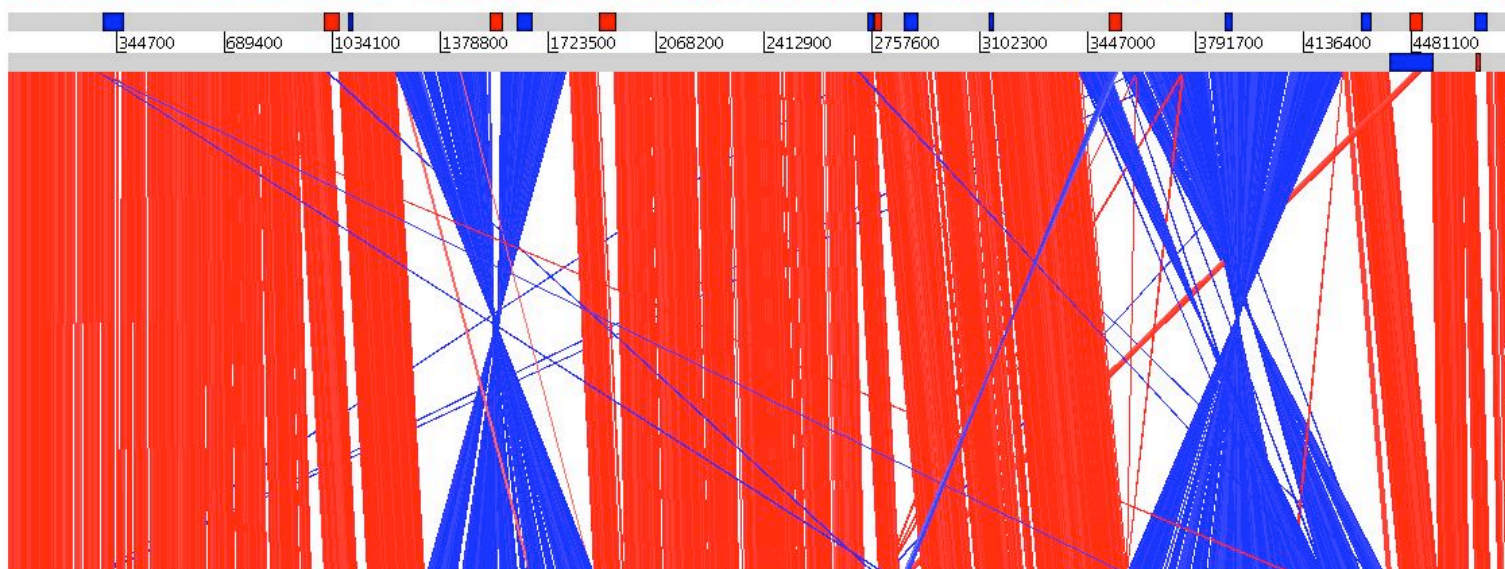
 prophage     PAI



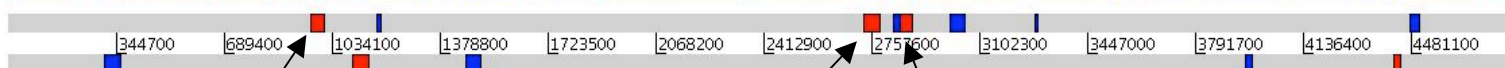




*E. coli*



*S. Typhi*



*S. Typhimurium*

Fels-1

Gifsy-2

Gifsy-1

Fels-2



# The Accessory Genome

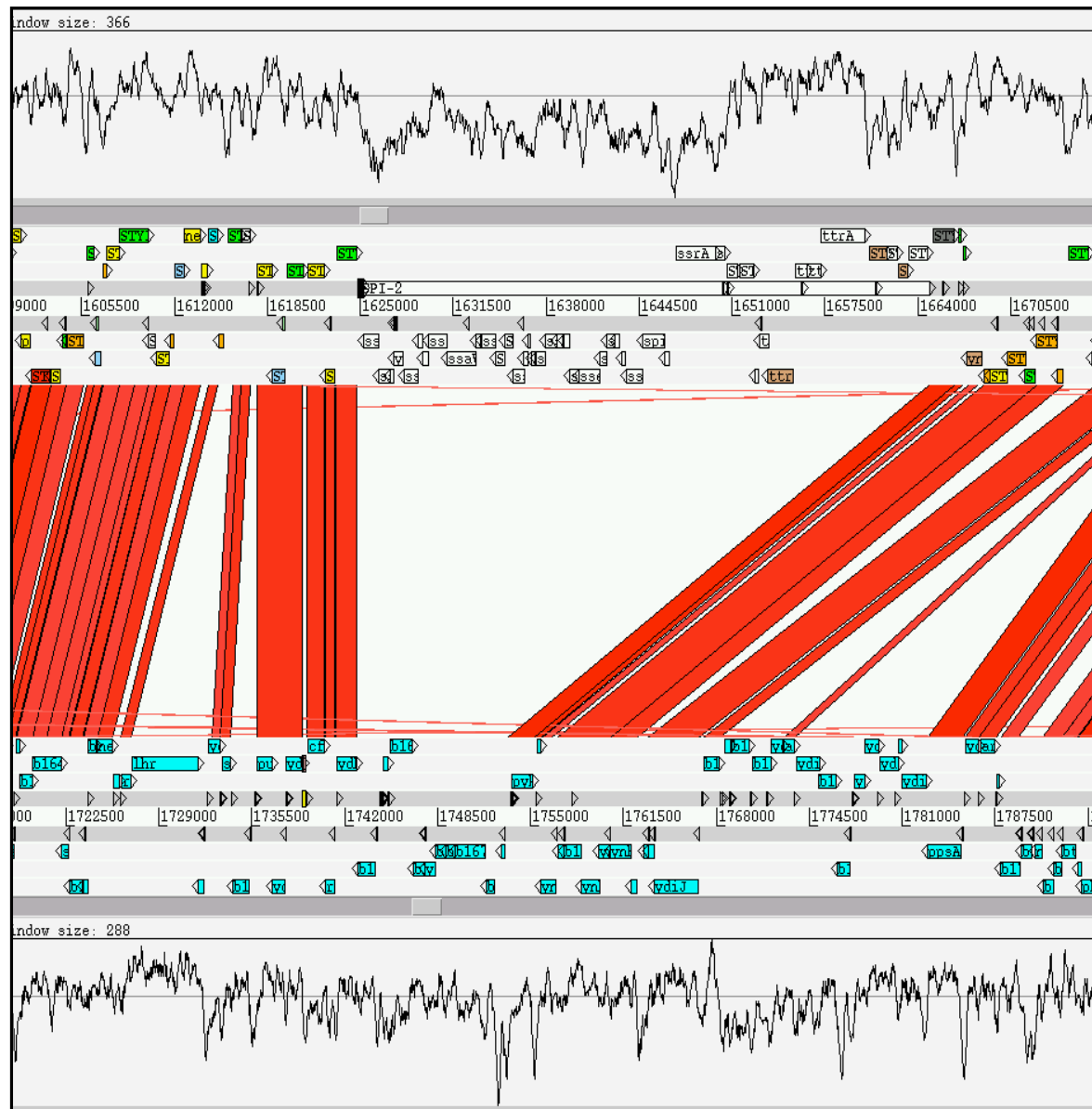
## *Pathogenicity Islands*

- 1) DNA loops
- 2) 'Jorg Hacker' Definition
  - Mobile?
  - Inserted next to tRNA
  - Anomolous G+C
  - Carry virulence genes
  - Pseudogenes
  - In pathogens absent from non-pathogens
- 3) Genomic Islands/loops, Islets, Metabolic islands
- 4) Size 10- >100kb





## *Salmonella typhi* vs. *E. coli* – SPI-2

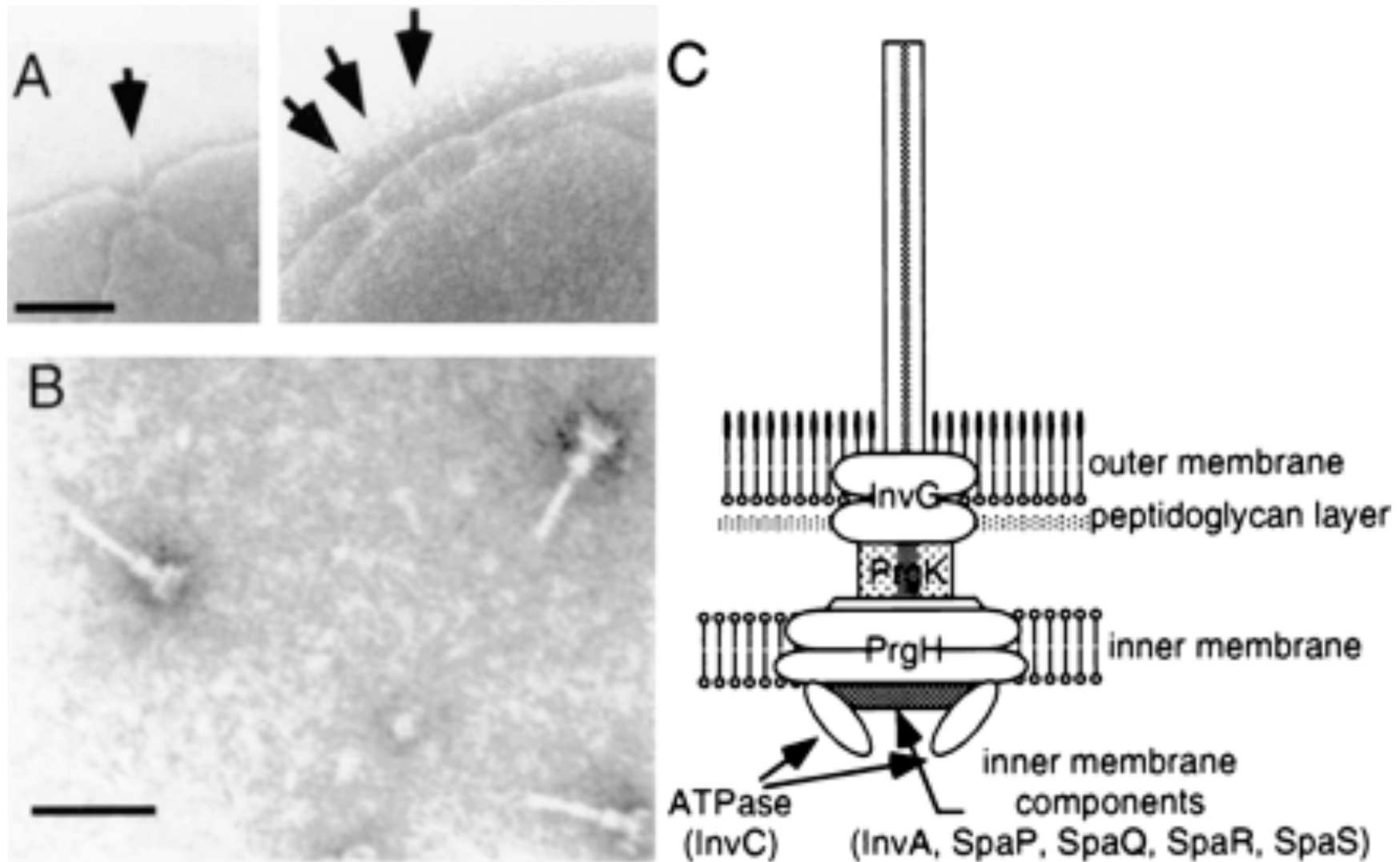


G+C  
tRNA  
phage/IS genes  
Pseudogenes  
virulence





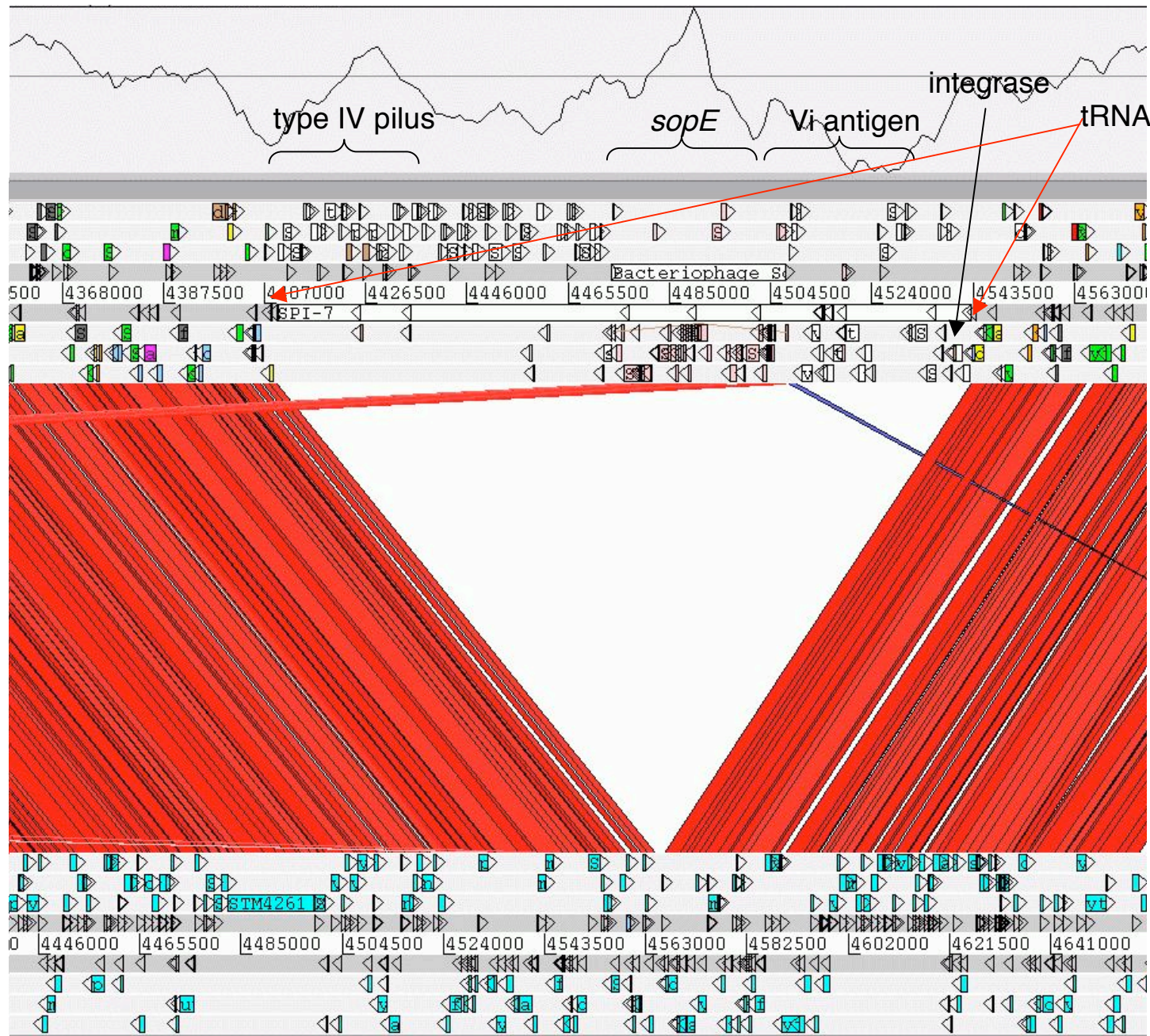
## *Salmonella* Typhi type III secretion system SPI-1



## *Salmonella typhi* and *Yersinia pestis* type III secretion systems



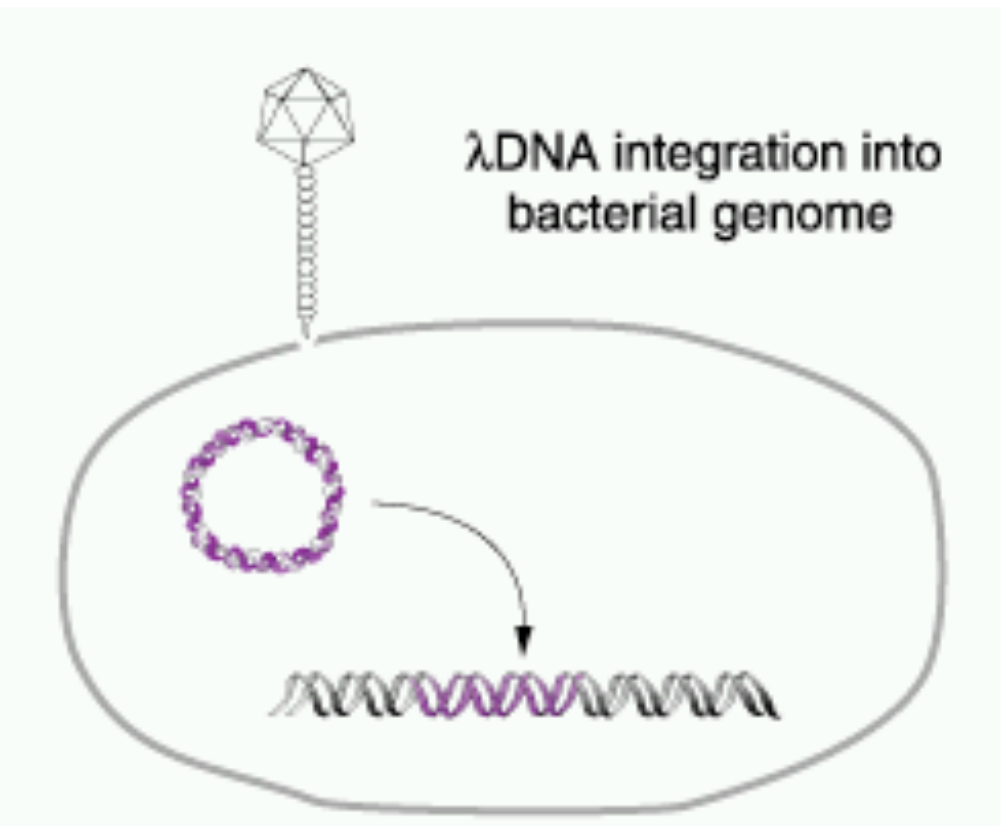
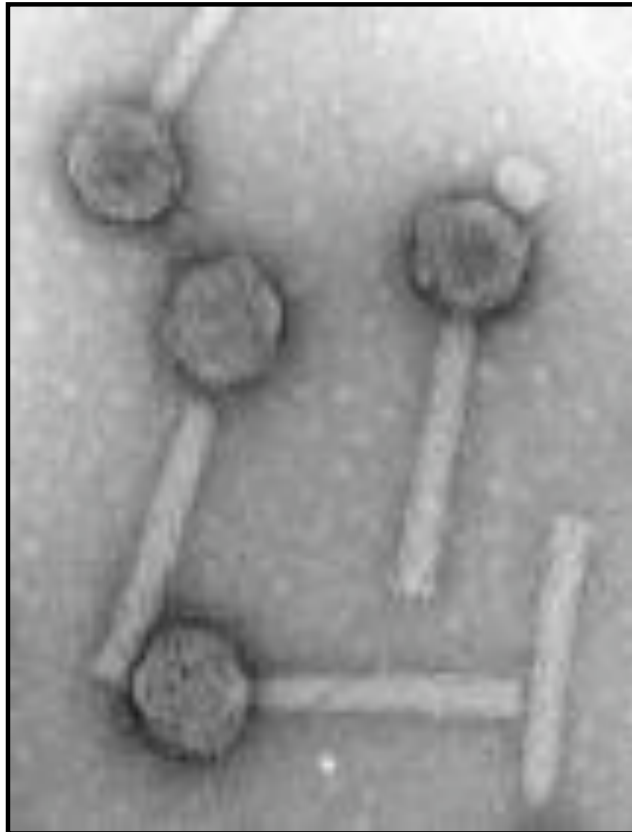


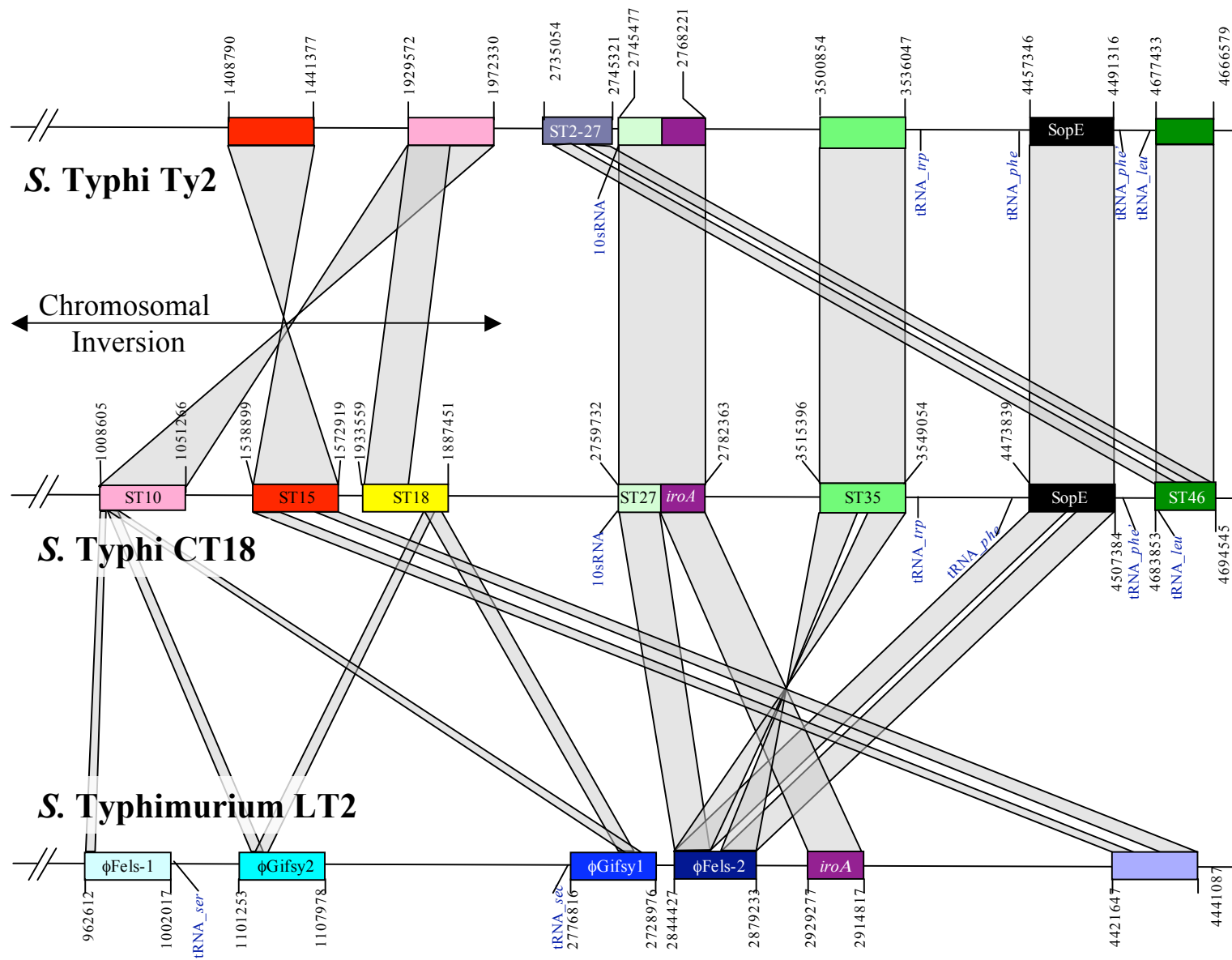


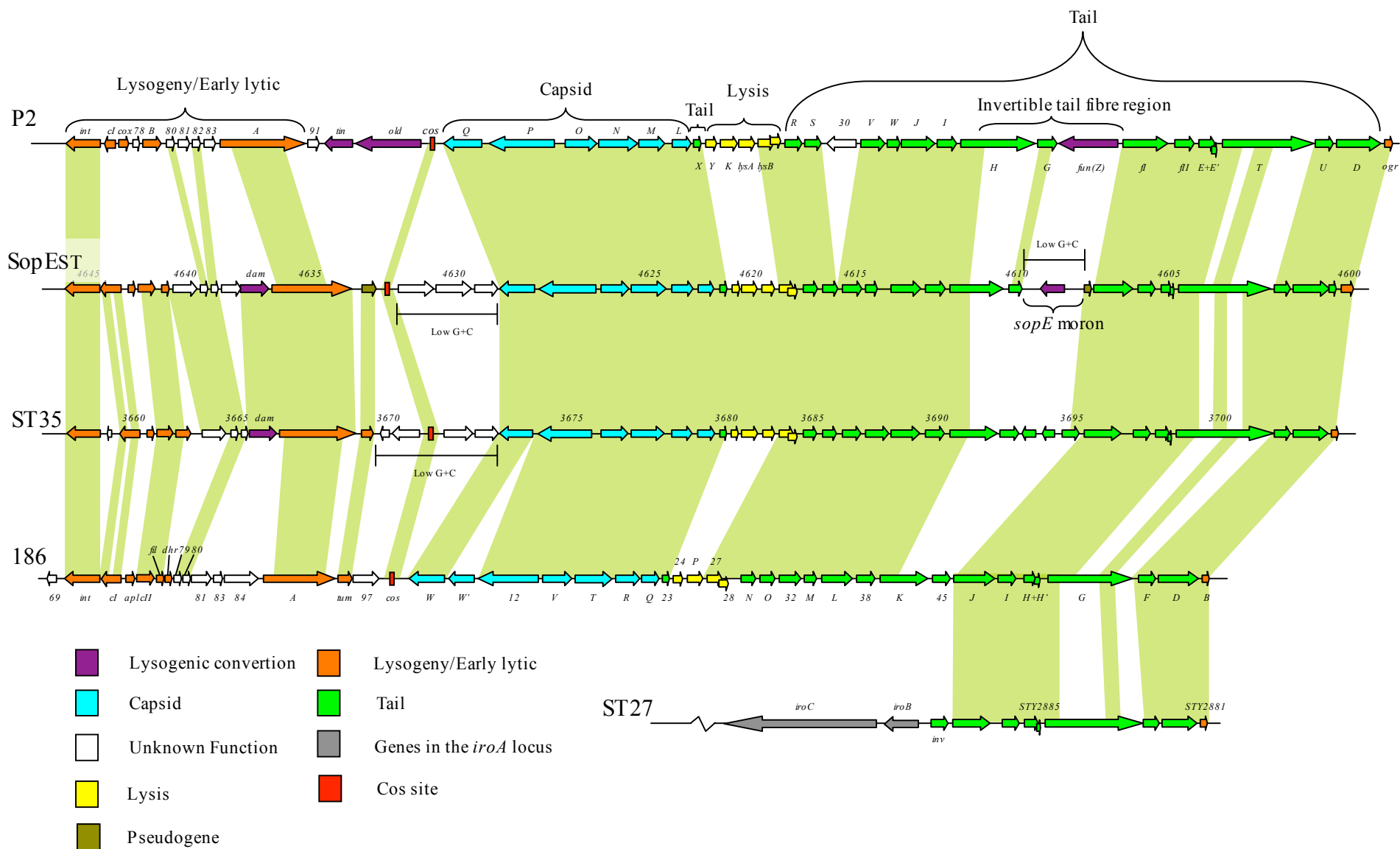
**S. Typhi**

**S. Typhimurium**

# The Accessory Genome Bacteriophage



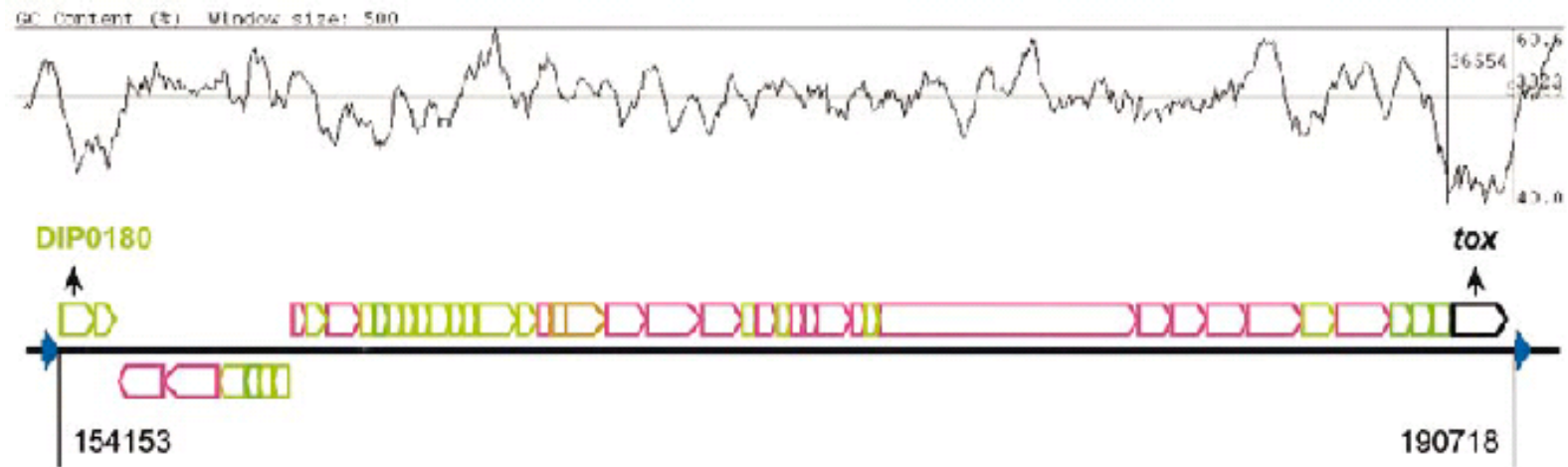






# Lysogenic conversion

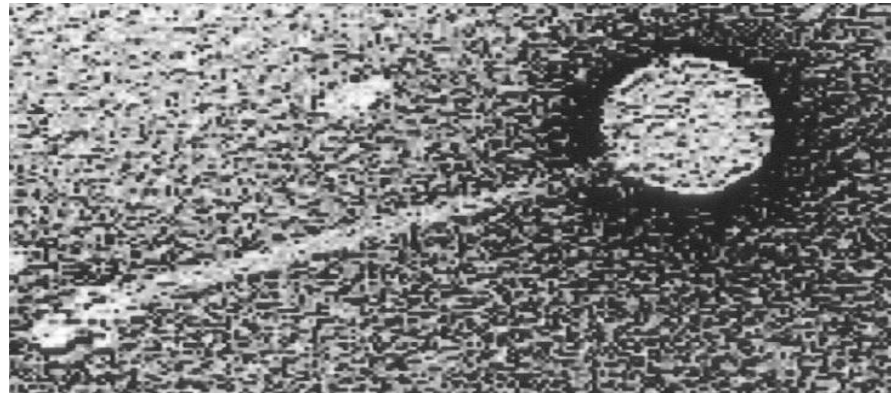
## *Corynebacterium diphtheriae*



Regulation and replication

Head tail and lysis genes

Diphtheria toxin





# Phage carried virulence factors

Protein	Gene	Phage	Organism
<b>Extracellular toxins</b>			
Diphtheria toxin	<i>tox</i>	β-Phage	<i>C. diphtheriae</i>
Cholera toxin	<i>ctxAB</i>	CTXφ	<i>V. cholerae</i>
<b>Proteins altering antigenicity</b>			
Membrane proteins	Mu-like	Pnm1	<i>N. meningitidis</i>
O-antigen acetylase	<i>oac</i>	Sf6	<i>S. flexneri</i>
<b>Effector proteins involved in invasion</b>			
Type III effector	<i>sopE</i>	SopEf	<i>S. enterica</i>
Type III effector	<i>ssel (gtgB)</i>	GIFSY-2	<i>S. enterica</i>
<b>Enzymes</b>			
Superoxide dismutase	<i>sodC-I</i>	GIFSY-2	<i>S. enterica</i>
Neuraminidase	<i>nanH</i>	Fels-1	<i>S. enterica</i>
<b>Serum resistance</b>			
OMPb	<i>bor</i>	λ	<i>E. coli</i>
OMP	<i>eib</i>	λ-like	<i>E. coli</i>
<b>Adhesions for bacterial host attachment</b>			
Vir	<i>vir</i>	MAV1	<i>M. arthritis</i>
Phage coat proteins	<i>pbIA , pbIB</i>	SM1	<i>S. mitis</i>
<b>Others</b>			
Mitogenic factor	<i>toxA</i>	Unnamed	<i>P. multocida</i>
Antivirulence	<i>grvA</i>	GIFSY-2	<i>S. enterica</i>

Adapted from Boyd & Brussow (2002) *Trends Microbiol.*

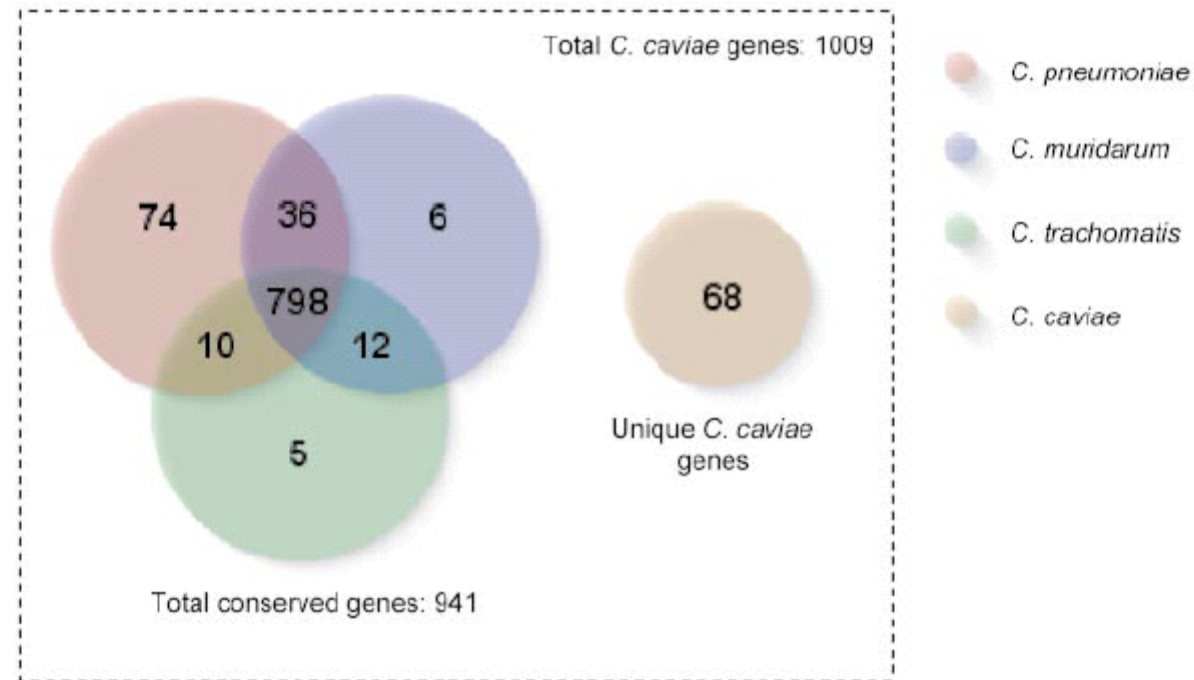
*The story outside of the gut*

*Same or different?*

*Chlamydia: Opposite end of the spectrum*



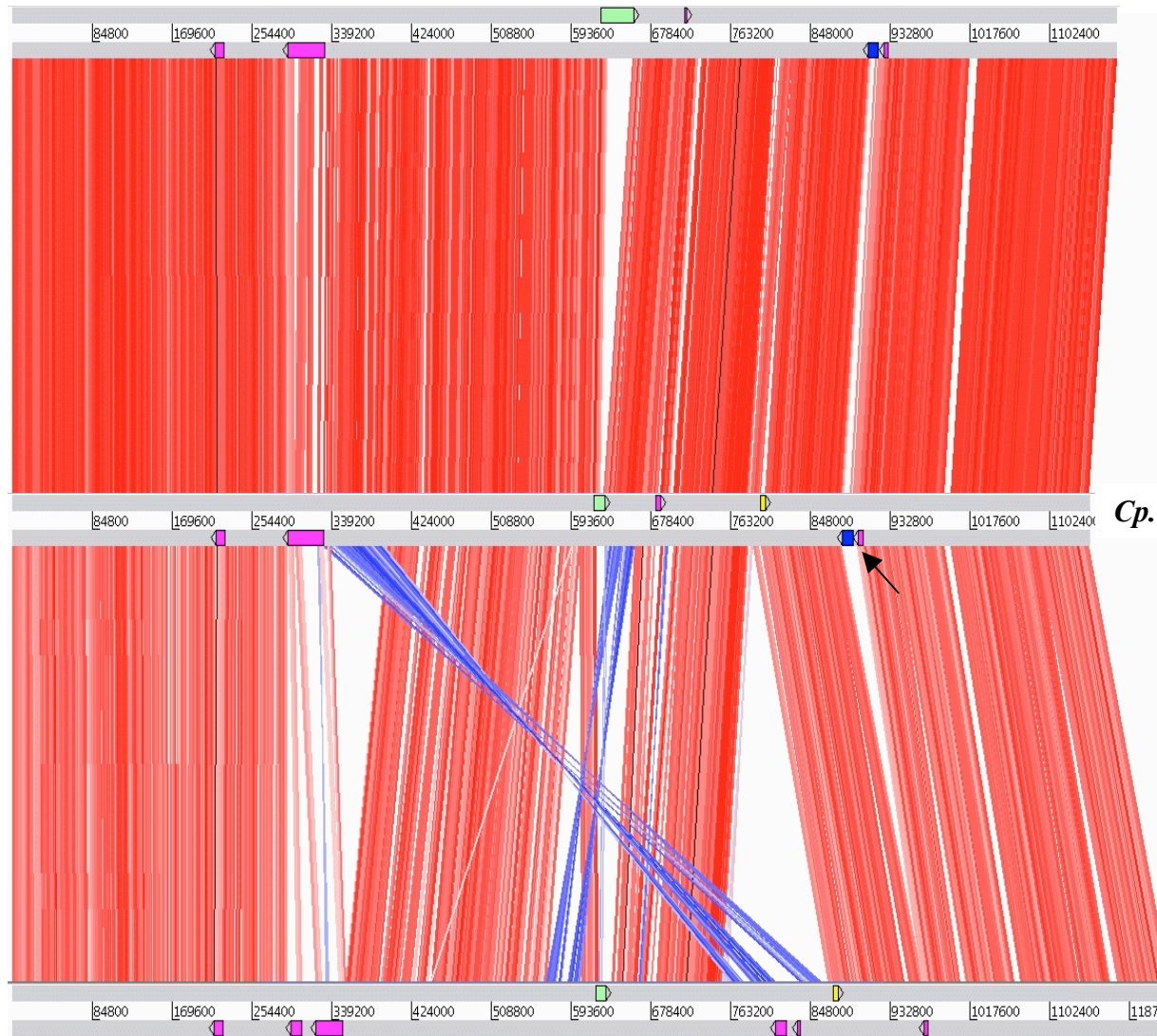
## Intracellular microbial parasites: *Chlamydophila*



**Figure 2.** Breakdown of orthologous *C. caviae* genes in other Chlamydiaceae genomes. Numbers represent a count of the most similar genes in particular genomes to *C. caviae*.

T. Read et al., 2003 Nucleic Acid Research

*Cp. caviae* (GPIC)



Infxn point  
PZ-tryp:  
STD  
ocular

*Cp. abortus* (S26/3)

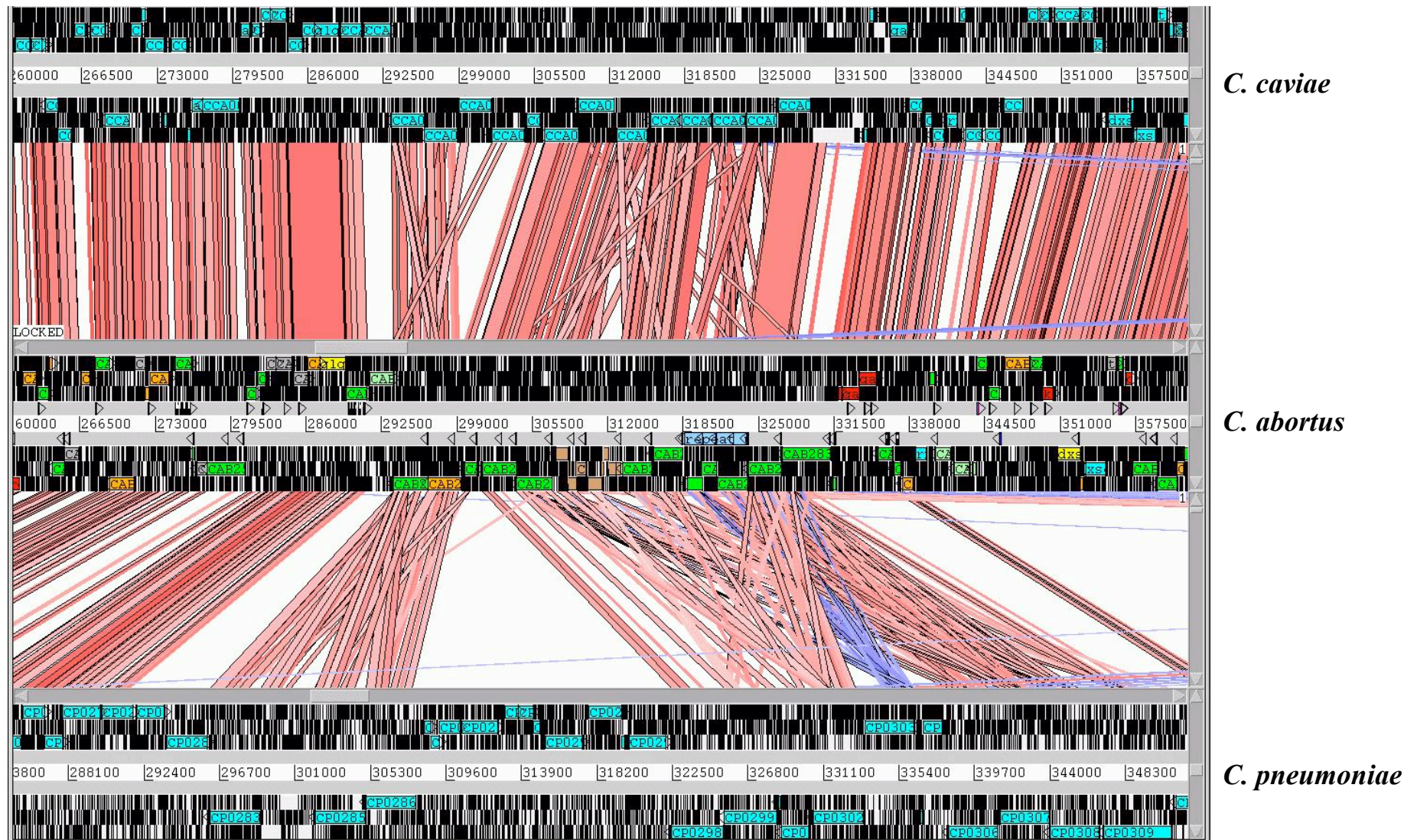
*Cp. pneumoniae* (AR39)







# Polymorphic membrane proteins

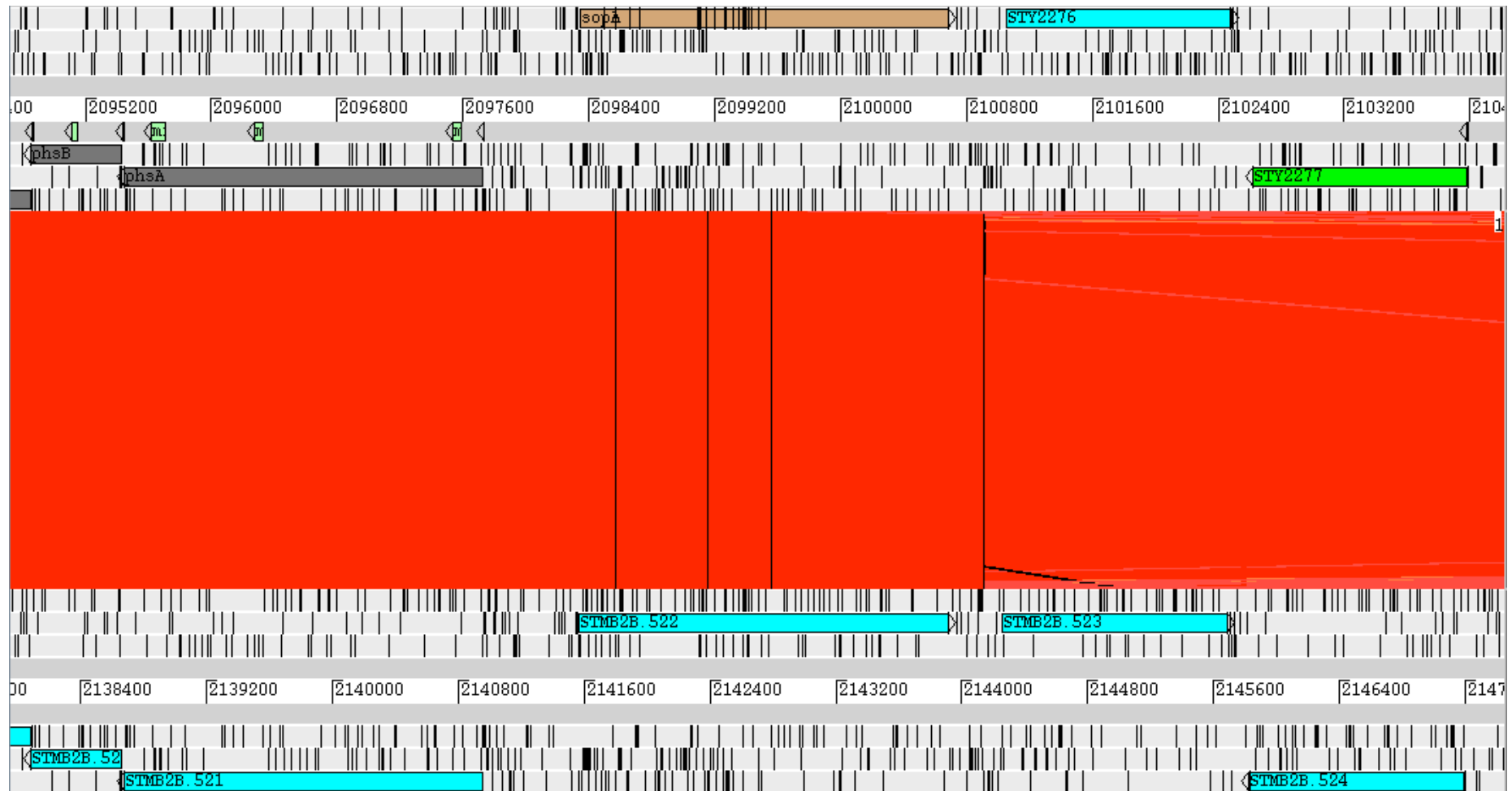


*Fine Detail*



## *Salmonella typhi* pseudogenes

secreted effector protein (type-III)



## *Salmonella typhi* pseudogenes

61% due to single mutations

10% in IS / phage

37% “metabolic” genes

59% lie in unique regions w.r.t. *E. coli*, compared to 33% of all genes

22% involved in virulence / host interaction, including:

components of 7/12 chaperone/usher fimbrial systems

flagella methylation                      *fliB*

type-III secreted effectors              *sseJ*  
   *sopE2*  
   *sopA*

host-range determinants                *slrP*  
   *shdA/ratA/sivH*

SPI-associated                              *ttrS* (SPI-2)  
   *cigR, marT, misL* (SPI-3)



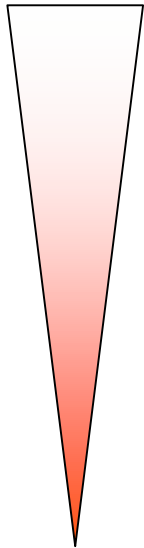




# *Summary to finish*

## Prokaryotic genomes are far from static

- Genome sequencing set to explode: 454
- Sources of genome variation:



Global

Bacteriophage

$10^{31}$  tailed phage particles

$10^{25}$  phage infections per second

Pathogenicity islands

remain controversial, diversity and source unknown

pseudogenes

- Access to diversity: Environment
- With all this DNA exchange how does that affect the Species concept

## Project Management

### Bart Barrell

Julian Parkhill  
Stephen Bentley  
Matthew Berriman  
Al Ivens  
Marie-Adèle  
Rajandream  
Nick Thomson

## Programming

Ellen Adlem  
Tim Carver  
David Harper  
Rob Davies  
Arnaud Kerhornou  
Paul Mooney  
Adrian Tivey  
Ed Zuiderwijk

## Administration

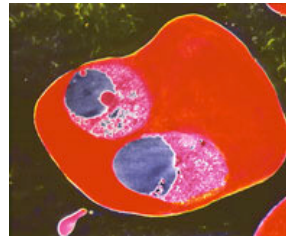
Sarah Locke  
Yvonne Shaw

## Visitors

Peter  
Ashton  
Sendu Bala  
Ed Burke

## Microarrays

Gareth Bloomfield  
Celine Carret  
Theresa Feltwell  
Maria Fookes  
Nefeli Nikolaidou-  
Katsaridou  
Matloob Qureshi  
Jason Skelton



## Analysis & Databases

Martin Aslett  
Andrew Berry  
Ulrike Böhme  
Ana Cerdeño-Tárraga  
Lisa Crossman  
Christiane Hertz-Fowler  
Matthew Holden  
Arnab Pain  
Chris Peacock  
Hubert Renault  
Mohammed Sebahia  
Valerie Wood

## Sequencing

### Carol Churcher

Becky Atkin  
Karen Brooks  
Inna Cherevach  
Tracey Chillingworth  
Nancy Hamlin  
Kay Jagels  
Sharon Moule  
Brian White  
Sally Whitehead

### David Harris

Nigel Fosker  
Lee Murphy  
Simon Rutter  
David Saunders  
Kathy Seeger  
Robert Squares

## Subcloning

Mike Quail  
Angela Lord  
Halina Norbertczak  
Claire Price  
Ester Rabinowitsch  
John Woodward

### Karen Mungall

Ian Goodhead  
Zahra Hance  
Heidi Hauser  
Mandy Sanders  
Mark Simmonds  
Danielle Walker

### Barbara Harris

Andrew Barron  
Louise Clark  
Craig Corton  
Nicola Lennard  
Alexandra Line  
Doug Ormond

## Comparative Genomics

Alison Dennis  
Emily Kay  
Helena Seth-Smith

