



# Structure and evolution of bacterial genomes

Olga Bochkareva

postdoc, Kondrashov group ([olga.bochkaryova@gmail.com](mailto:olga.bochkaryova@gmail.com))

# Data is coming ...

## Bacteria

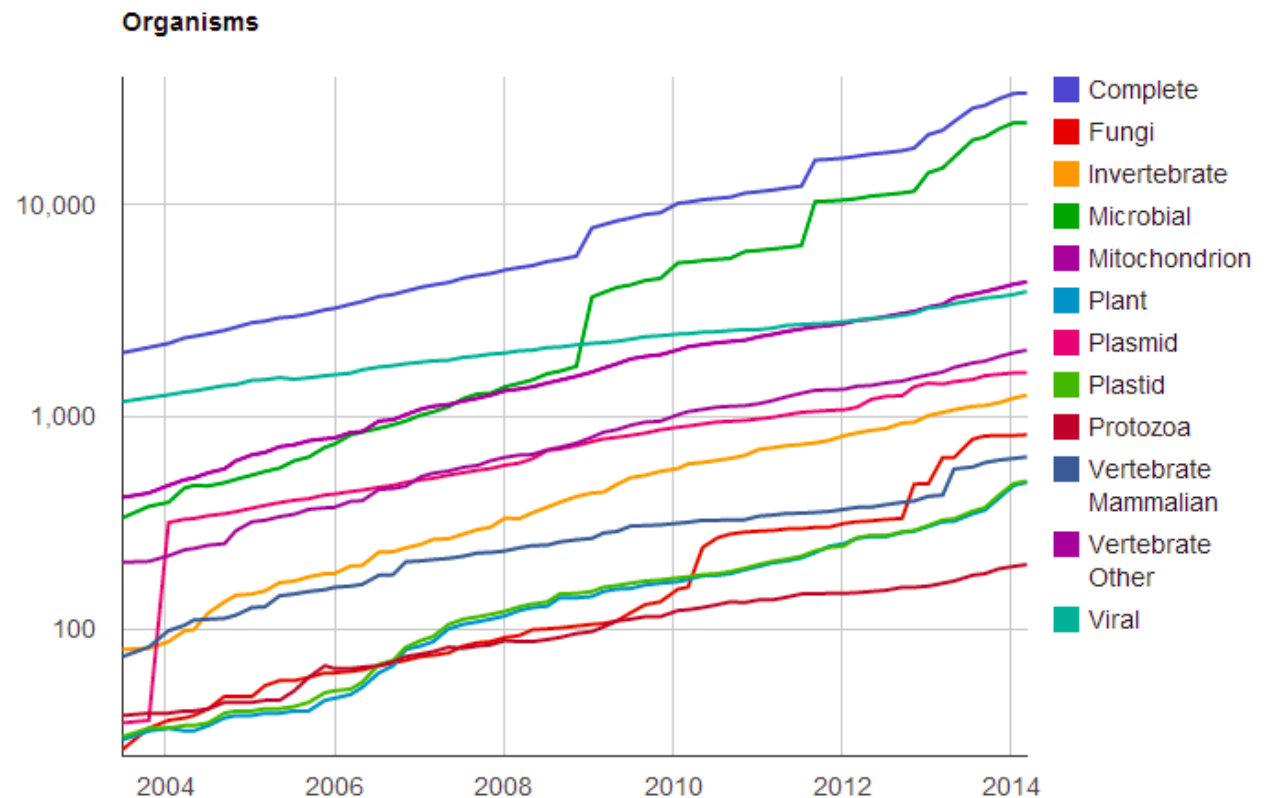
479 320 assemblies



16 527 complete genome  
assembly level



1 631 representative  
genomes



<https://www.ncbi.nlm.nih.gov/refseq/statistics/>

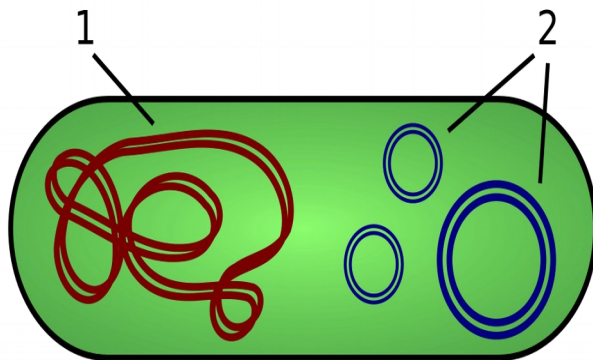
# What is bacterial genome?

## 1. Chromosome(-s)

- circular, double-stranded DNA molecule
- 1Mb – 6 Mb
- one origin of replication
- carry house-keeping genes

## 2. Plasmids

- circular, double-stranded DNA molecules
- 15 kb – 200 kb
- replicate independently
- carry genes that benefit the survival of the organism.



*wikipedia.org*

# How does it evolve?

## What we can observe in genomes

- Mutations – genes sequences

```

A5A9C3.1 14 SIKLPPSQITRLIVERRMNNIST..PSIFSRK..YGSISKEERARENKQIEVACSTAND....HYEKEPDCGGSAVILYAKESKILILEVLK 101
B4F917.1 13 SIKLPPSQESTRIMLVDRMTNINLST..ESIFSRK..YRLIGKDEAHENAKTIEELCFALADE.....HFREEPDCGGSSAVILYAKETSMMLEVLK 100
A9S1V2.1 23 VEKLPSPSQICREAVKDMHLKSS..ALCFEESG...FRIELHAGDENHAKHLEEVAFGAKGE.....ADSGDKTDSAVMVYKHAHSLNLETLR 109
B9G9W7.1 18 SVKLPSPQSTRLIVERRITQITII..PSIFSRK..YGLSKEEKEQNKIEVAFRANND....RYEKDPDCGGSSAVIYAKESKILILEVLK 100
Q8H056.1 30 SFESLPPPTQRTKDAVVRRLVDTEGG..DITLCKR..YGVVPRADAEPRARGIEAEAFDARAA..SGEAAATASVEEIKALQLYSEVSRRLDFFVK 120
Q0Q423.2 44 SLSLPPSPQRTKDAVVRRLVDIVR..PSILSKR..YGVVPRADAEPRARGIAAEVAFVAVTES..SRAHAPASVEEIGLEVLQYSEVSRRLLELAK 135
B9W44B.1 56 SFESLPPPTQRTKDAVVRRLVETLST..PSVLSKR..YGTIPKEEESAEKRIEERFGRST.....YVSEKEDQLEVLQYSEVSRRLDFFVK 141
Q0TYC5.1 29 SFAMVPPPTQRTKDAVVRRLVAVLSGOTTALRKRKYRYGVVPRADAEPRARGIAAEVAFVAVTES...SSSSSSVEEIGLETLQLYSEVSRRLDFFVK 121
P9M446.1 13 SIKLPPSQESTRLIVERRITDNLSS..YVFSRKR..YGLSKEEERHAKRIEETAFALAND....HEKKEPDLQSSVDFYKAEARLRHLEHLK 100
Q9C500.1 57 SIKLPPPTQRTKDAVVRRLVETLST..ESILSKR..YTLKSDHATTVNLIEEERVGVASH.....AVSDDQDQILELQYSEVSRRLDFFVK 142
Q2HR17.1 26 YSILPPKQRTKDAVVRRLVETLST..PSVLTQR..YGTMSADEASAAITQIEEERFVAVNA.....SSSTSNQNTILEVYSEVSRRLDFFVK 110
Q9M7N3.1 28 SFKLPPTQRTKDAVVRRLVETLST..DSVLSKR..YGVPEEDATSHARILEEERFVAVSV..ASARSTGGREDBILEVHLYSEIDVQVVESSH 119
Q9M7N6.1 20 SFESLPPPTQRTKDAVVRRLVETLST..PSILSKR..YTLPDEASETARLIEEERFARAS.....TASDADQILELQYSEVSRRLDFFVK 110
Q9LEB2.1 14 SVFMPPSQESTRLIVERRIKNIIT..PSIFSRK..YGLLSVEEEREDDKPTEDLRFATANK.....HFQNEPDCGDSAVIYAKESKILILEVLK 101
Q9M651.2 13 SIKLPPSPFTRKALERTTAVPSS..KILITET..YSLTKDQRTKRIEETIDHFTAND....DFEKEPDCGGSAVILYAKESKILILEVLK 100
B9R749.1 48 SLSLPPPTQRTKDAVVRRLVETLST..PSVLSKR..YGTISHDEASARRIEEERFQVANT.....RTSREDDQLELQYSEVSRRLDFFVK 133
  
```

# How does it evolve?

## What we can observe in genomes

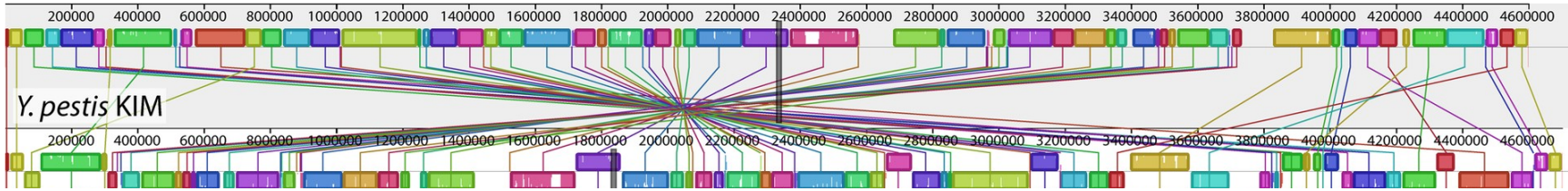
- Mutations – genes sequences

```

A59C3.1 14 SIKLPPSQTTRLLVERMANNIST..PSIFSRK..YGSISKEERARENKQIEEVCSTAND....HYKEPDCGGSAVILYAKESKILILEVLK 101
B4F917.1 13 SIKLPPSSESTRIMLVDMTNNLST..ESIFSRK..YRLIGKDEAHENAKTIEELCFALADE....HFREPPDCGSSAVILYAKETSQMLEVLK 100
A9S1V2.1 23 VEKLPSPSQTREAVKDMHLKLS..ALFESUS..FRIELHAGDENHAKIEEVAFGAKDE....ADSGDKTDSAVMVYAKHAKMLLETLR 109
B9GSH7.1 18 SVKLPFDOSTRLMVERMTTQIT..PSIFSRK..YGLSKEEKEENKIEEVAFGAKND....HYKQPDGCGSSAVIYAKESKILILEVLK 100
O8H056.1 30 SFESIMPTQTRDQAVVRKLVDTLGG..DITLCKR..YGVAPRADEPARAGIEAERFQBARA..SGEAAATASVEEIGIKALQYKESRRLLDVFK 120
B0Q423.2 44 SLSIMPPSQTREAVKDMHLKLS..PSILSKR..YGVAPRADEPARAGIEAERFQBARA..SGEAAATASVEEIGIKALQYKESRRLLDVFK 120
B9W44B.1 56 SFESIMPTQTRDQAVVRKLVDTLGG..DITLCKR..YGVAPRADEPARAGIEAERFQBARA..SGEAAATASVEEIGIKALQYKESRRLLDVFK 120
Q0TYC5.1 29 SFAMPPPTQTRDQAVVRKLVAVLSCOTTALKRKYRYGVAPRADEPARAGIEAERFQBARA..SGEAAATASVEEIGIKALQYKESRRLLDVFK 121
P9M446.1 13 SIKLPPSSESTRIMLVDMTNNLST..PSIFSRK..YGLSKEEKEENKIEEVAFGAKND....HYKEPDCGGSAVILYAKESKILILEVLK 100
Q9C500.1 57 SLSIMPPSQTREAVKDMHLKLS..PSILSKR..YGLSKEEKEENKIEEVAFGAKND....HYKEPDCGGSAVILYAKESKILILEVLK 100
Q2HR17.1 26 YSIMPQKQTRDQAVKRLIETLST..PSVLTQR..YVITMSDEKSAARITQIEERFQBARA..SGEAAATASVEEIGIKALQYKESRRLLDVFK 110
Q9M7N3.1 28 SFKLPPTQTRDQAVVRKLVDTLGG..DITLCKR..YGVAPRADEPARAGIEAERFQBARA..SGEAAATASVEEIGIKALQYKESRRLLDVFK 120
Q9M7N6.1 26 SFESIMPTQTRDQAVKRLIETLST..PSVLTQR..YVITMSDEKSAARITQIEERFQBARA..SGEAAATASVEEIGIKALQYKESRRLLDVFK 110
Q9LEB2.1 14 SVYMPSPSSTRIMLVDMTNNLST..PSIFSRK..YGLSVEEKEEDKPTIEDLRFATANK....HFONEPDCGDSAVIYAKESKILMLDVIK 101
Q9M651.2 13 SIKLPPSPTKRLIETRETNVPS..PILITET..YGLSKEEKEENKIEEVAFGAKND....HFONEPDCGDSAVIYAKESKILILEVLK 100
B9R748.1 48 SLSIMPTQTRDQAVITRLIETLST..PSVLSKR..YVITSHDEKSAARIEERFQBARA..SGEAAATASVEEIGIKALQYKESRRLLDVFK 133
    
```



- Gene gain/loss – genes content
- Rearrangements - genes order



# How does it evolve?

## What we can observe in genomes

- Mutations – genes sequences

```

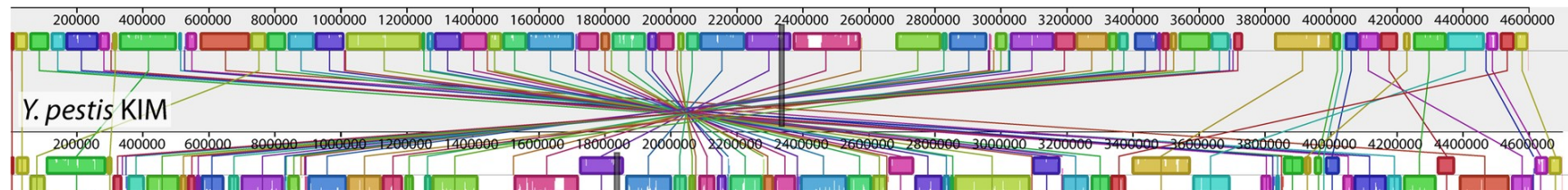
A59C3.1 14 SIKLPPSPTTRLLVERMANNIST..PSIFSRK..YGSISKEERARENKQIEEVCSTAND....HYEKEPDCGGSAVQLYAKESKLLILEVLK 101
B4F917.1 13 SIKLPPSPTTRMLVDMTNNLST..ESIFSRK..YRLGKDEAHENAKTIEELCFALADE....HFREEPDCGSSAVQLYAKETSQMLLEVLK 100
A9S1V2.1 23 VEKLPSPSPTTRKRWKDMHKLKSS..ALFESUS..FRIELHAGDENHAKHLEEVAFGAKDE....ADSGDKTDSAVMYYVHAKSLMLLETLR 109
B9S3W7.1 18 SVKLPSPDSTRMLVERMTTQFII..PSIFSRK..YGLSKEEKEENKIEEVAFRANNO....RYEKQPDGSSAVQIYKESKLLILEVLK 100
08H056.1 30 SFESIMPTQTRDQAVRRLVDTGG..DITLCKR..YGVAFADAEPAARCTEAEAFDARA..SGEAAATASVEEIKALQLYKESVRRLLDFVK 120
00Q423.2 44 SLSIMPPSPTQTRDQAVRRLVDTVA..PSILSKR..YGVAFADAEPAARCTEAEAFDARA..SGEAAATASVEEIKALQLYKESVRRLLDFVK 135
B9W44B.1 56 SFESIMPTQTRDQAVRRLVDTVA..PSILSKR..YGVAFADAEPAARCTEAEAFDARA..SGEAAATASVEEIKALQLYKESVRRLLDFVK 141
Q0IYCS.1 29 SFAMPPPTQTRDQAVRRLVAVLSCOTTALKRKYRYGVAFADAEPAARCTEAEAFDARA..SSSSSSVEEDIEITDLYSREVSRLLAFFVR 121
P9M446.1 13 SIKLPPSPTTRMLVERMTNNLSS..VYFYSRK..YGLSKEEKEENKIEEAFRANNO....HEKEEPLDSSVQVYKESRRLMLLEVLK 100
Q9S500.1 57 SIKLPPSPTQTRDQAVRRLVDTVA..PSILSKR..YGLSKEEKEENKIEEAFRANNO....RVSDDQKILELYKESVRRLLDFVK 142
Q2HR17.1 25 SFESIMPPKQTRDQAVRRLVDTVA..PSVLTQR..YGVAFADAEPAARCTEAEAFDARA..SSSSSSVEEDIEITDLYSREVSRLLAFFVR 110
Q9M7N3.1 28 SFKLPPTQTRDQAVRRLVDTVA..PSVLTQR..YGVAFADAEPAARCTEAEAFDARA..SSSSSSVEEDIEITDLYSREVSRLLAFFVR 119
Q9M7N6.1 20 SFESIMPTQTRDQAVRRLVDTVA..PSILSKR..YGVAFADAEPAARCTEAEAFDARA..SSSSSSVEEDIEITDLYSREVSRLLAFFVR 110
Q9LEB2.1 14 SVFMPPSPTTRMLVERMTNNLST..PSIFSRK..YGLSKEEKEENKIEEAFRANNO....HFONEPDCGGSAVQLYAKESKLLILEVLK 101
Q9M651.2 13 SIKLPPSPTTRKRLERETWVPS..PILITET..YGLSKEEKEENKIEEAFRANNO....DFREEPDCGGSAVQLYAKESKLLILEVLK 100
B9R749.1 48 SLSLPPPTQTRDQAVRRLVDTVA..PSVLSKR..YGVAFADAEPAARCTEAEAFDARA..SSSSSSVEEDIEITDLYSREVSRLLAFFVR 133
    
```



- Gene gain/loss – genes content
- Rearrangements - genes order

## What are the mechanisms?

1. Homologous recombination
2. Intra-genomic recombination
3. Horizontal gene transfer:
  - transformation
  - transduction
  - bacterial conjugation



# From Biology to Maths

## Rearrangement

> Deletion

## Math illustration

**1 2 -3 4 -5 6**  $\Rightarrow$  **1 2 -3 -5 6**

> Insertion

**1 2 -3 4 -5 6**  $\Rightarrow$  **1 2 -3 4 7 -5 6**

> Translocation

**1 2 -3 4 -5 6**  $\Rightarrow$  **1 -3 4 -5 2 6**

> Inversion

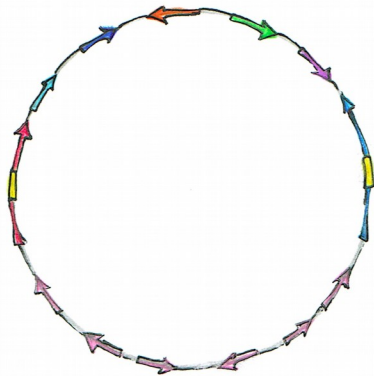
**1 2 -3 4 -5 6**  $\Rightarrow$  **1 2 5 -4 3 6**

> Duplication

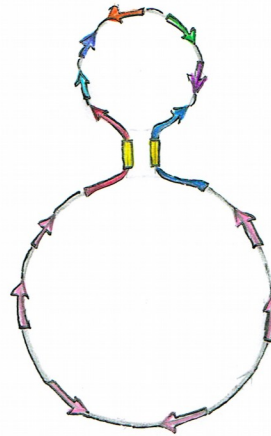
**1 2 -3 4 -5 6**  $\Rightarrow$  **1 2 -3 -3 4 -5 6**

# Intra-genomic recombination

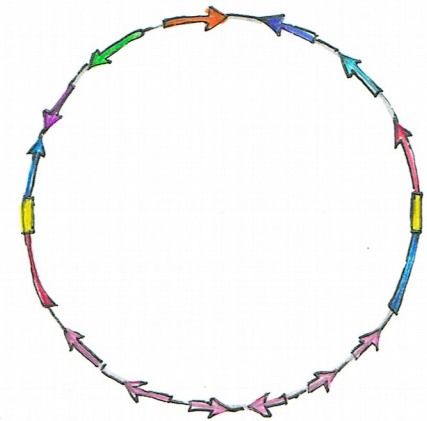
**Circular chromosome with repeats**



**Intra-chromosome recombination**



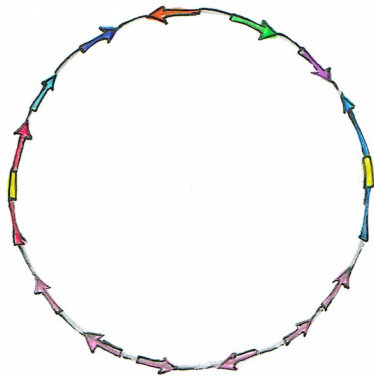
**Inversion of the segment between repeats**



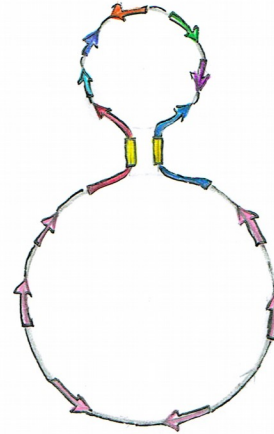


# Intra-genomic recombination

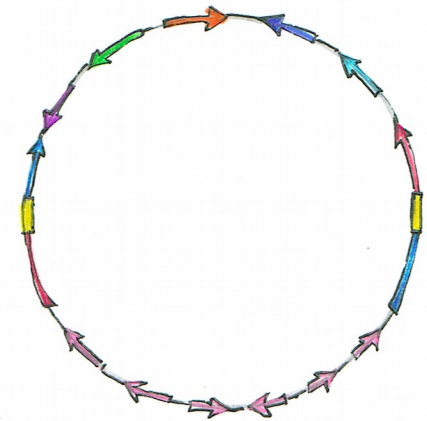
**Circular chromosome with repeats**



**Intra-chromosome recombination**



**Inversion of the segment between repeats**

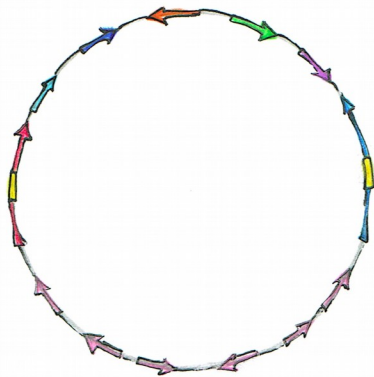


## Genomic repeats:

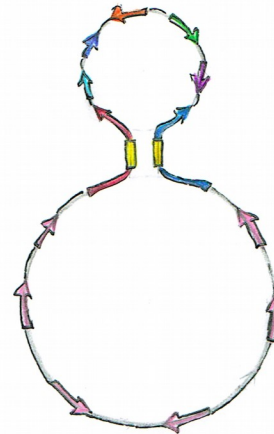
- mobile elements (transposase) – up to 100 copies per genomes
- rRNA gene operons – up to 15 copies per genomes
- gene paralogs

# Intra-genomic recombination

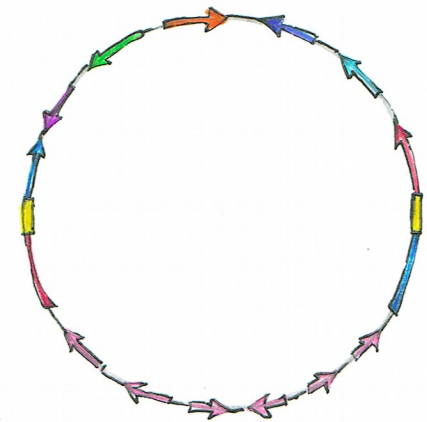
**Circular chromosome with repeats**



**Intra-chromosome recombination**



**Inversion of the segment between repeats**



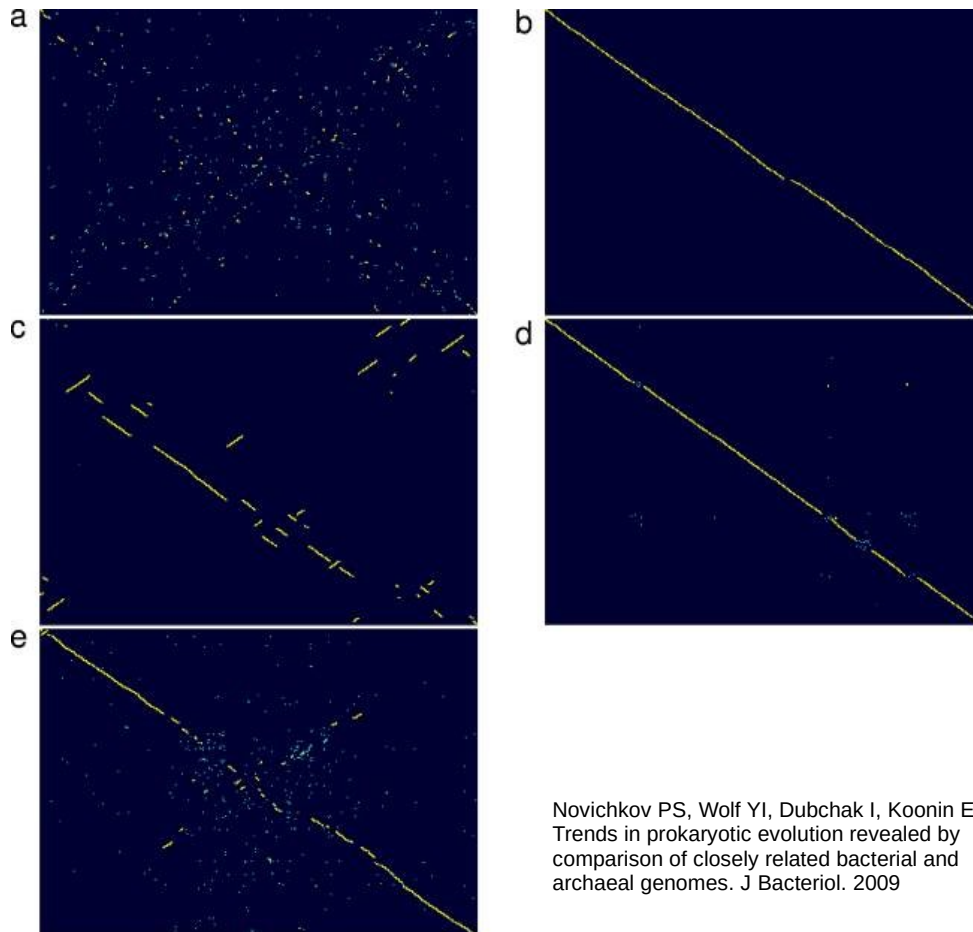
## Genomic repeats:

- mobile elements (transposase) – up to 100 copies per genomes
- rRNA gene operons – up to 15 copies per genomes
- gene paralogs

- short fragments

Site-specific inversion

# Is gene order conserved?



(a) Nearly complete decay of synteny; *Streptococcus sanguinis* SK36 and *Streptococcus pneumoniae* R6.

(b) Virtual absence of rearrangements; *Chlamydomophila caviae* GPIC and *Chlamydomophila abortus* S26/3.

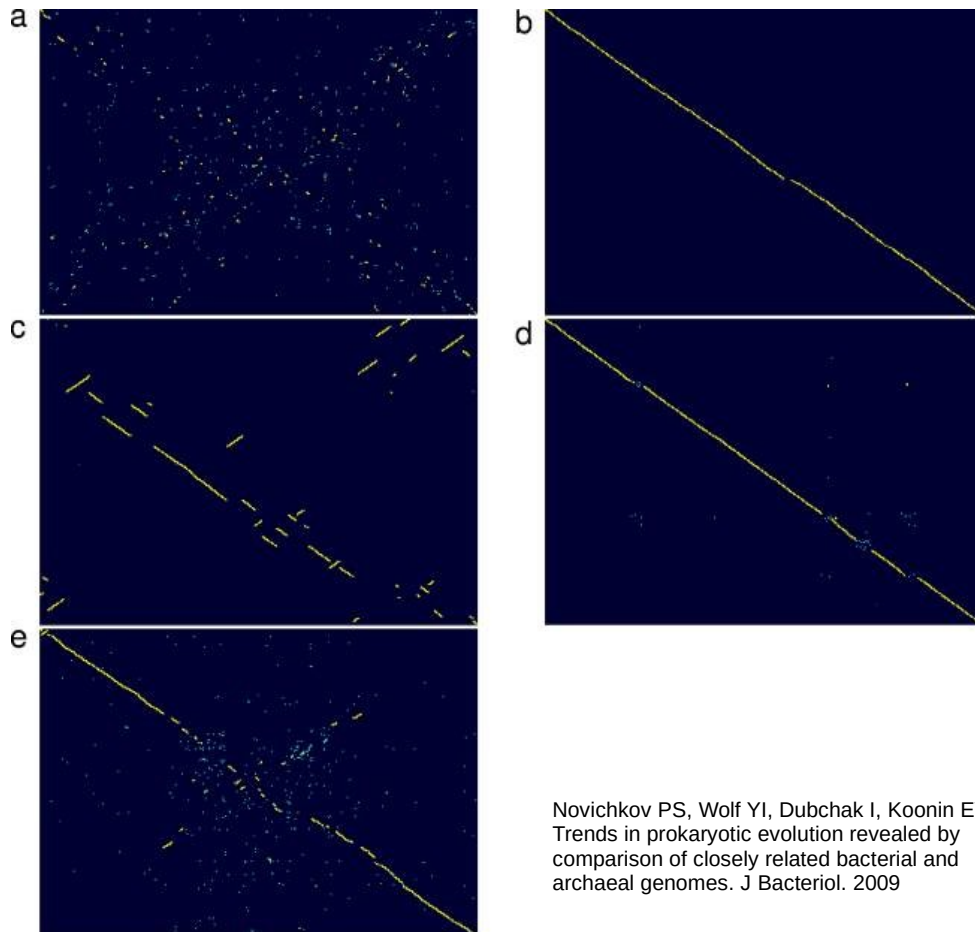
(c) Multiple inversions with limited transposition of individual genes; *Yersinia pestis Antiqua* and *Y. pestis CO92*.

(d) No inversion; hot spots of transposition of individual genes; *P. marinus* AS9601 and *P. marinus* MIT 9215.

(e) Multiple inversions and transposition of individual genes; *Pseudomonas fluorescens* PfO-1 and *P. fluorescens* Pf-5.

Novichkov PS, Wolf YI, Dubchak I, Koonin EV.  
Trends in prokaryotic evolution revealed by  
comparison of closely related bacterial and  
archaeal genomes. J Bacteriol. 2009

# Is gene order conserved?



(a) Nearly complete decay of synteny; *Streptococcus sanguinis* SK36 and *Streptococcus pneumoniae* R6.

(b) Virtual absence of rearrangements; *Chlamydomophila caviae* GPIC and *Chlamydomophila abortus* S26/3.

(c) Multiple inversions with limited transposition of individual genes; *Yersinia pestis Antiqua* and *Y. pestis CO92*.

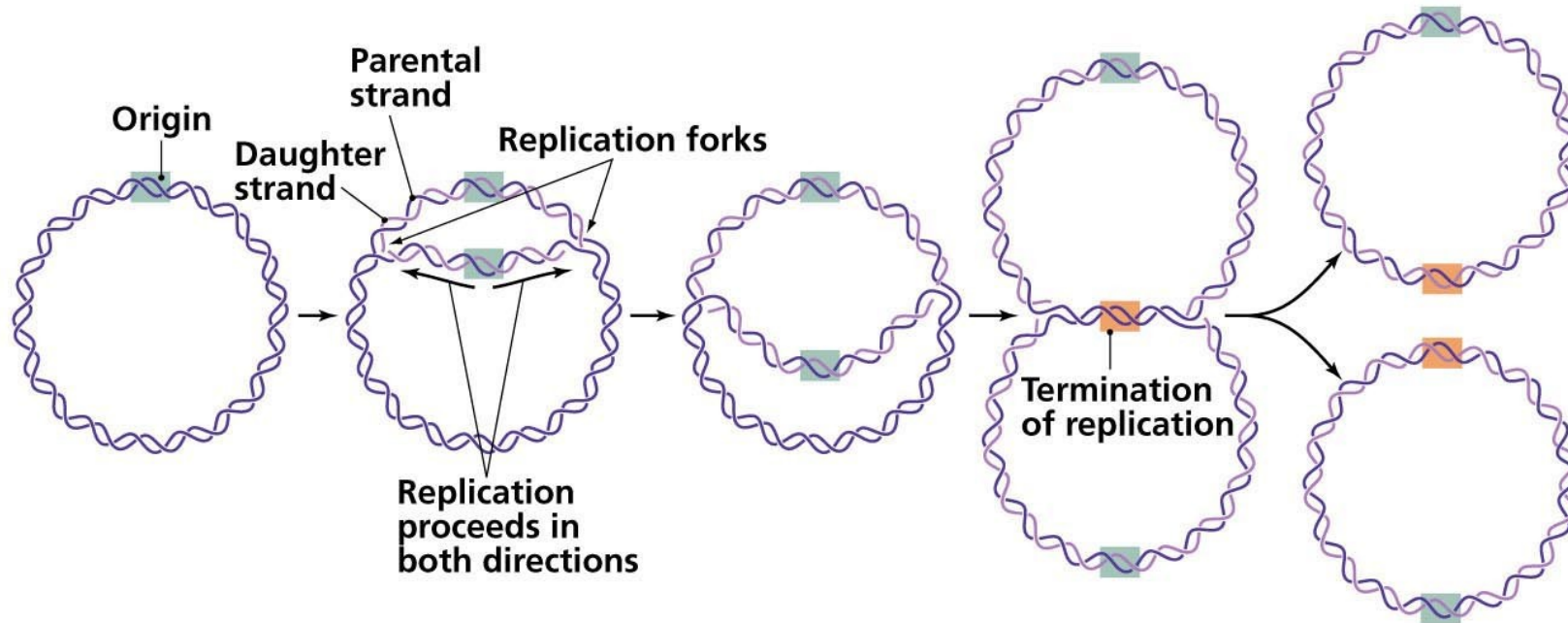
(d) No inversion; hot spots of transposition of individual genes; *P. marinus* AS9601 and *P. marinus* MIT 9215.

(e) Multiple inversions and transposition of individual genes; *Pseudomonas fluorescens* PfO-1 and *P. fluorescens* Pf-5.

Novichkov PS, Wolf YI, Dubchak I, Koonin EV. Trends in prokaryotic evolution revealed by comparison of closely related bacterial and archaeal genomes. J Bacteriol. 2009

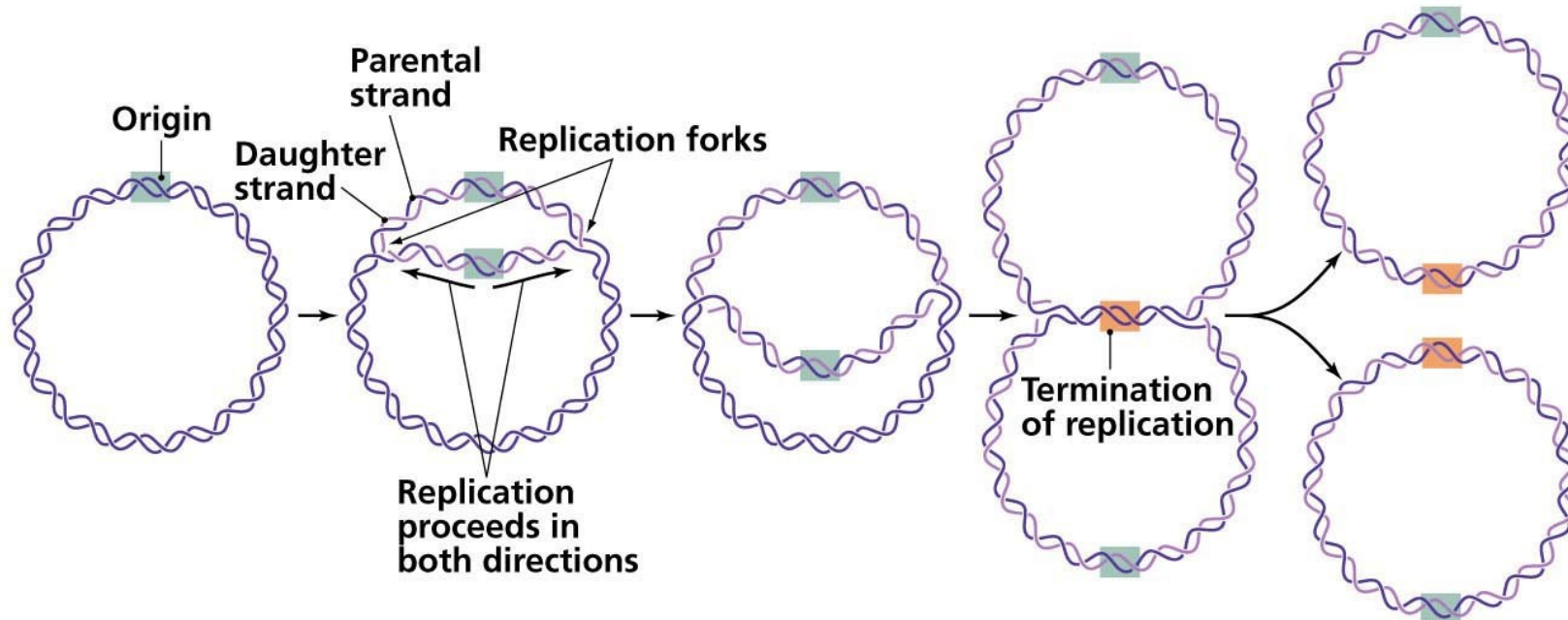
symmetric inversions  
around the replication origin

# Chromosome structure



Copyright © 2006 Pearson Education, Inc., publishing as Benjamin Cummings.

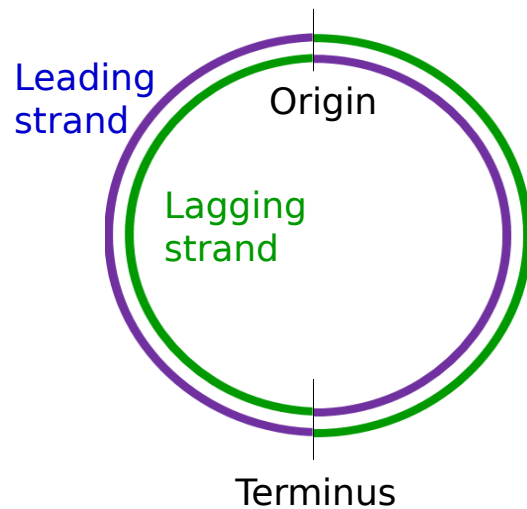
# Chromosome structure



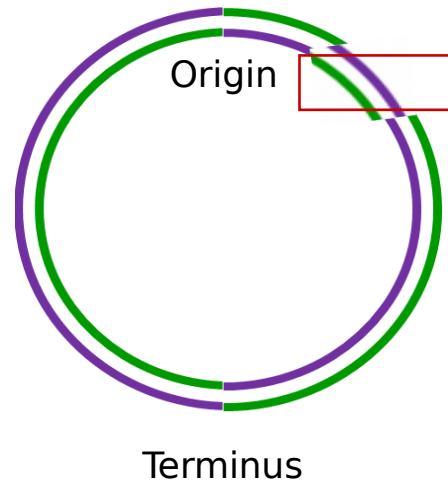
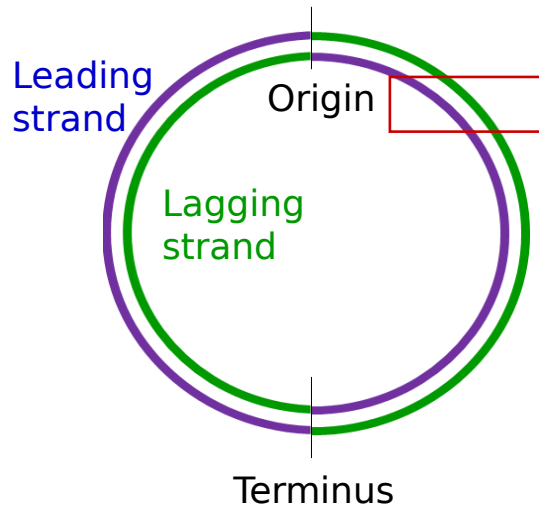
## Selection on genome rearrangements / gene transfer:

1. Highly-expressed genes placed closed to origin of replication
2. Highly-expressed genes placed on the leading strand (co-direction of replication and transcription processes)
3. Origin and terminus of replication placed symmetrically (equal replicore lengths)

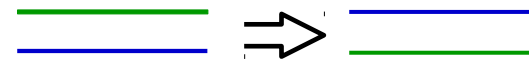
# If the inversion occurred



# If the inversion occurred

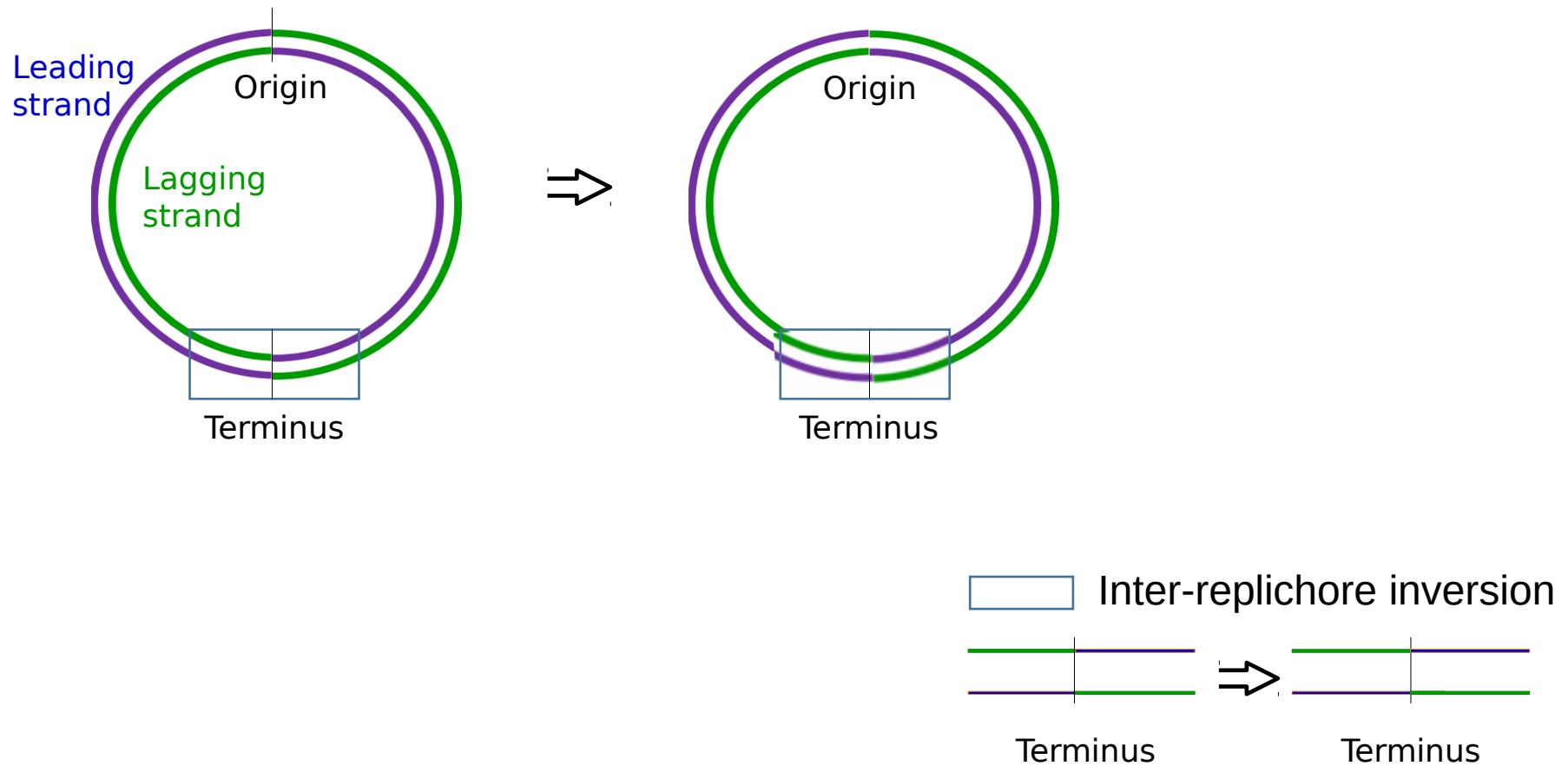


 Intra-replichore inversion

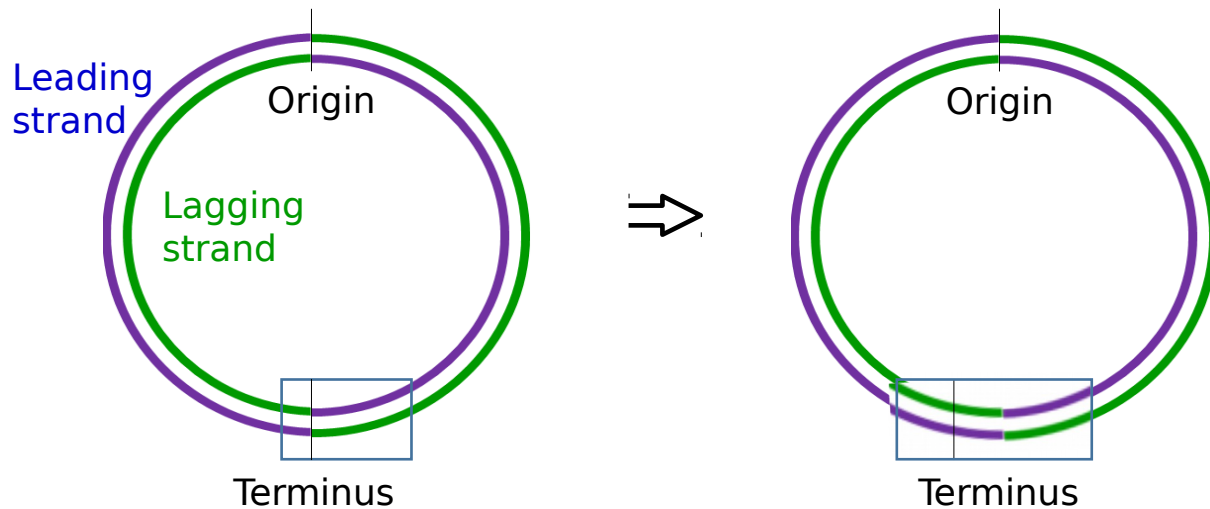




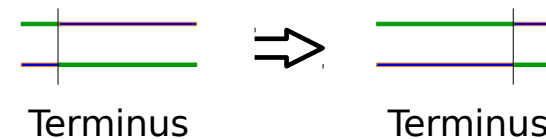
# If the inversion occurred



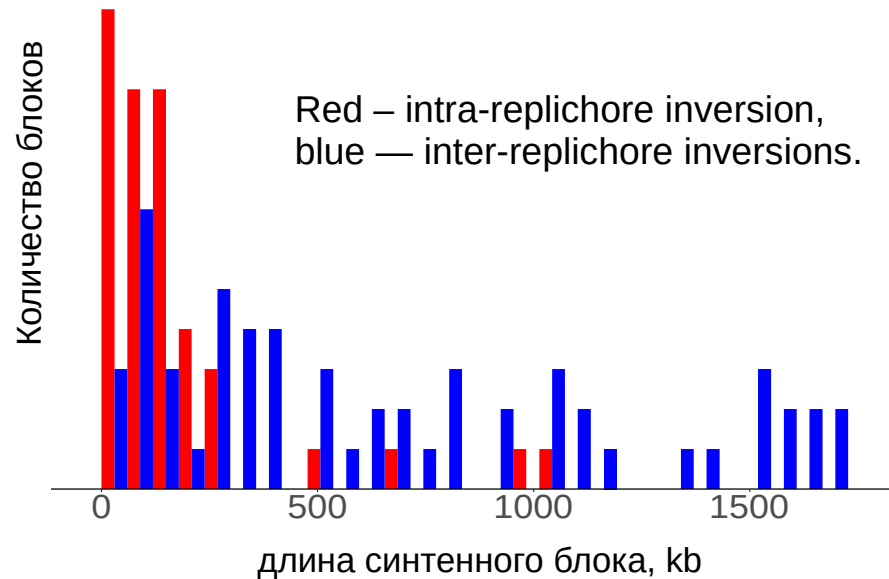
# If the inversion occurred



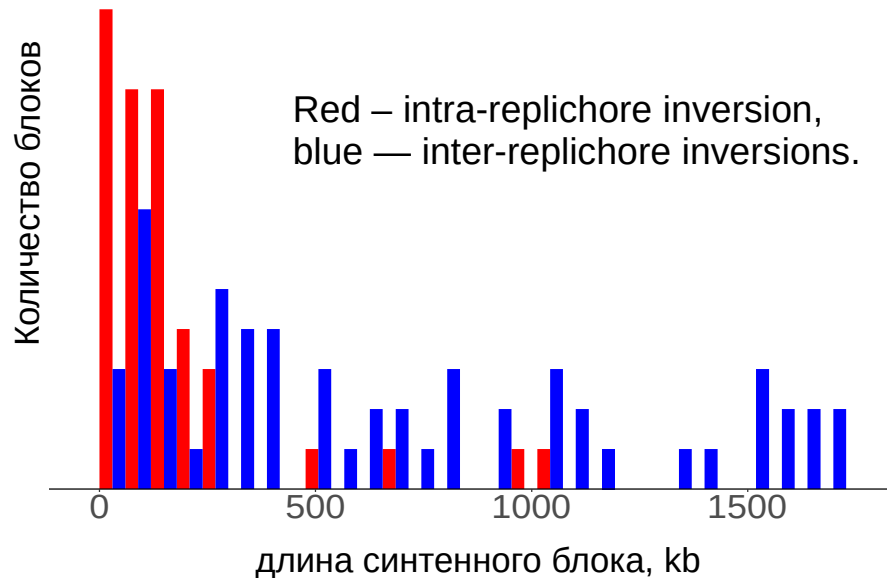
Inter-replichore inversion



# What we see on the data



# What we see on the data



Вероятность инверсии длиной  $m$  пройти через точку начала (конца) репликации:

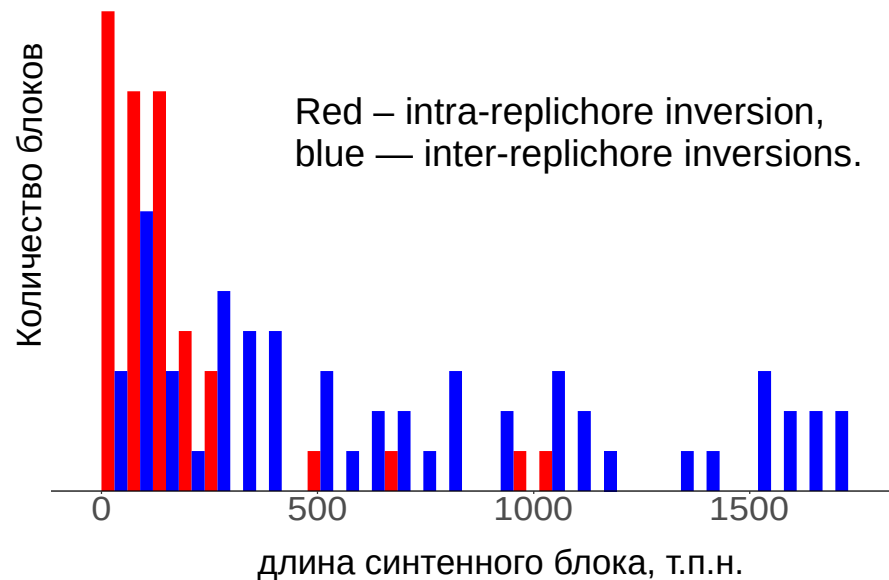
$$p(l) = m/L,$$

где  $L$  – размер реплихоры.

Доля инверсий с концами на разных реплихорах:  
58 из 101

Вероятность получить такой набор инверсий:  
 $p \sim 10^{-33}$

# What we see on the data



Вероятность инверсии длиной  $m$  пройти через точку начала (конца) репликации:

$$p(l) = m/L,$$

где  $L$  – размер реплихоры.

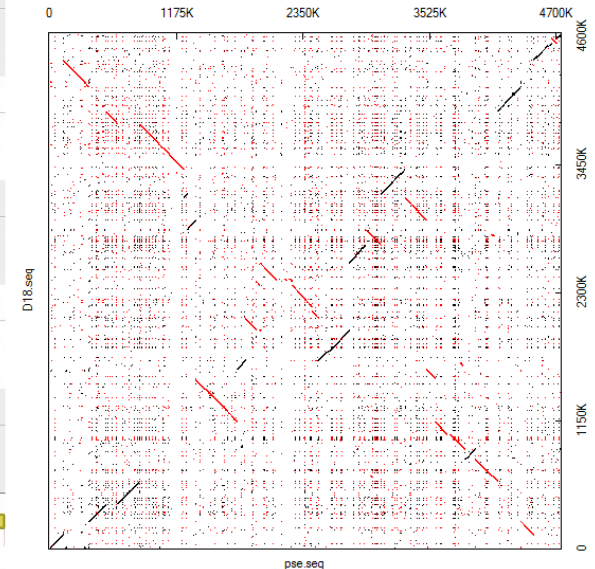
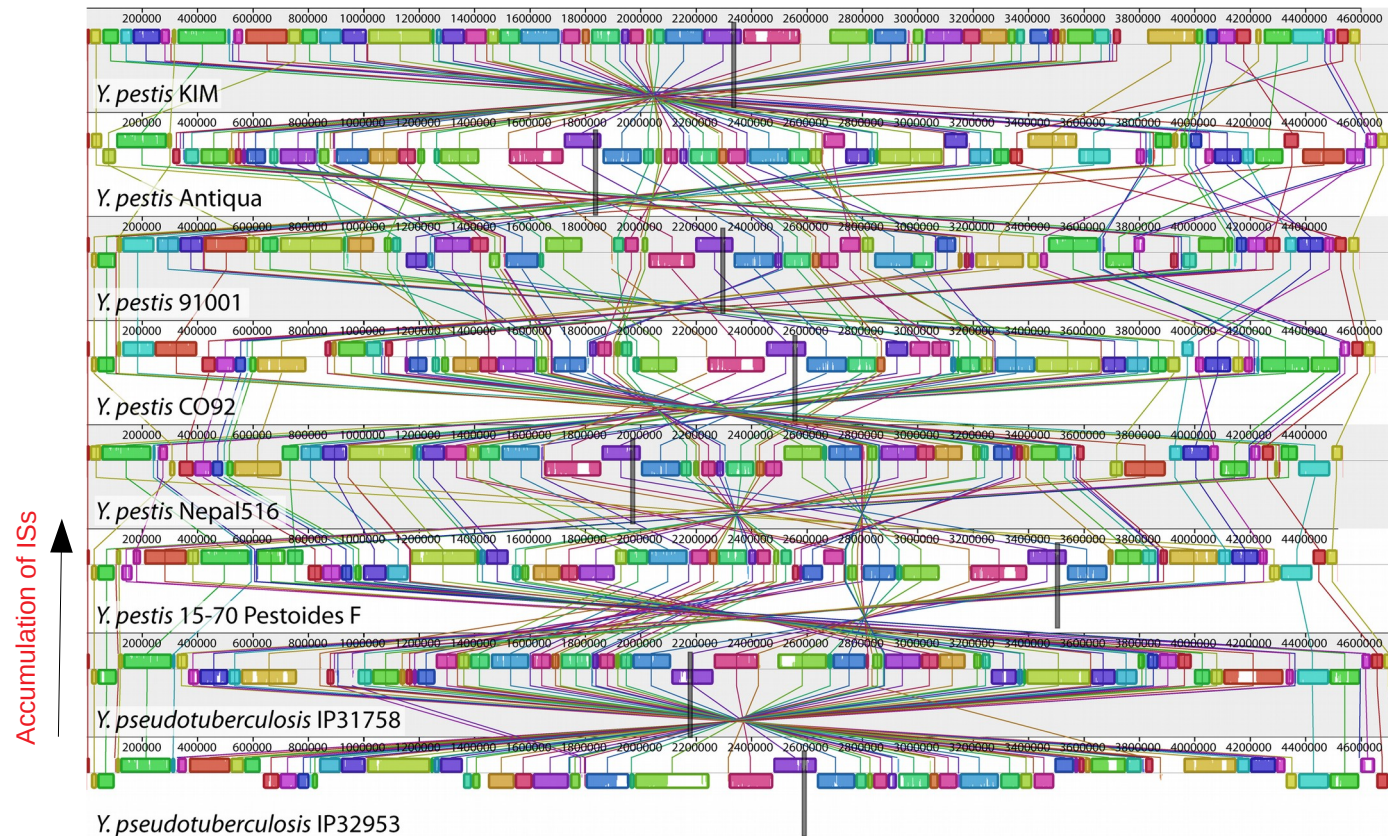
Доля инверсий с концами на разных реплихорах:  
58 из 101

Вероятность получить такой набор инверсий:  
 $p \sim 10^{-33}$

Darling AE, Miklós I, Ragan MA. Dynamics of genome rearrangement in bacterial populations. PLoS Genet. 2008

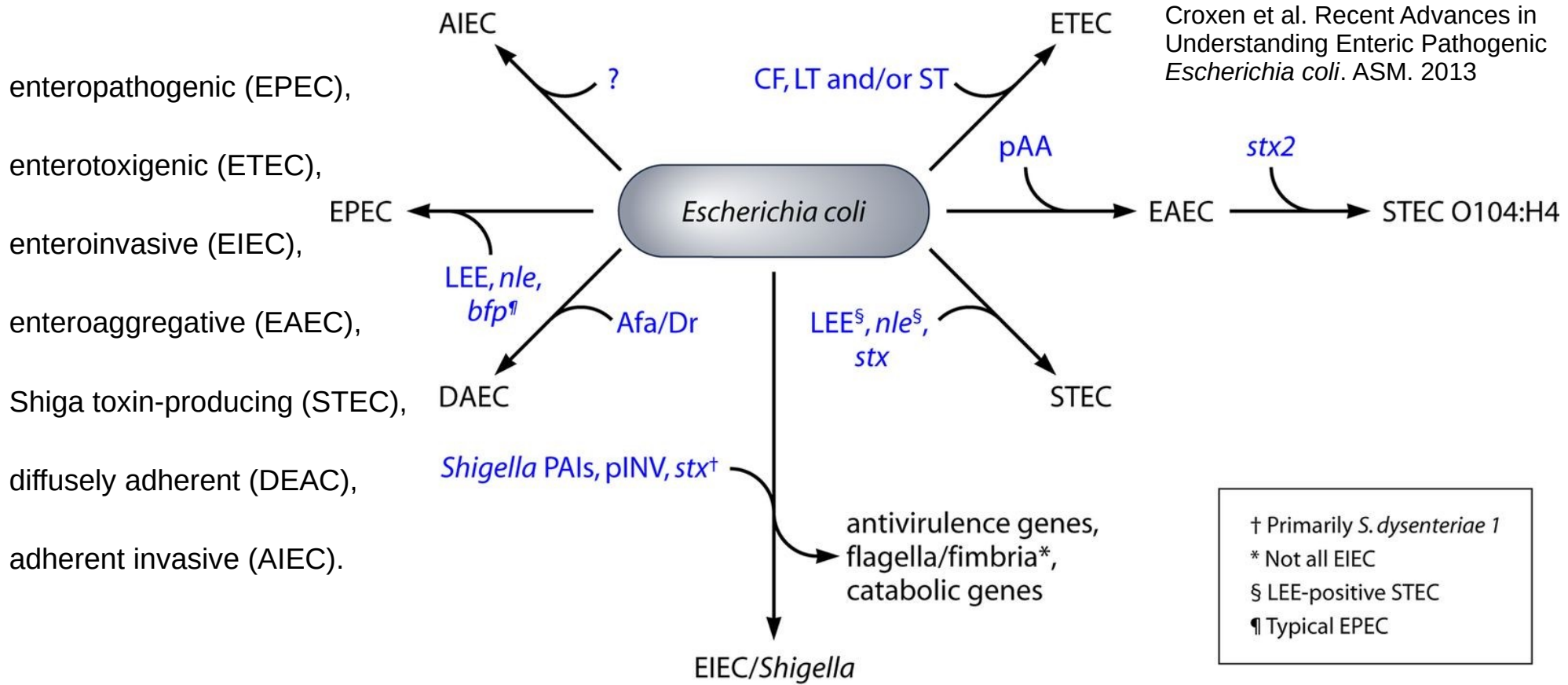
# ISs expansion drives rearrangements

*Yersinia pseudotuberculosis* → *Y. pestis*



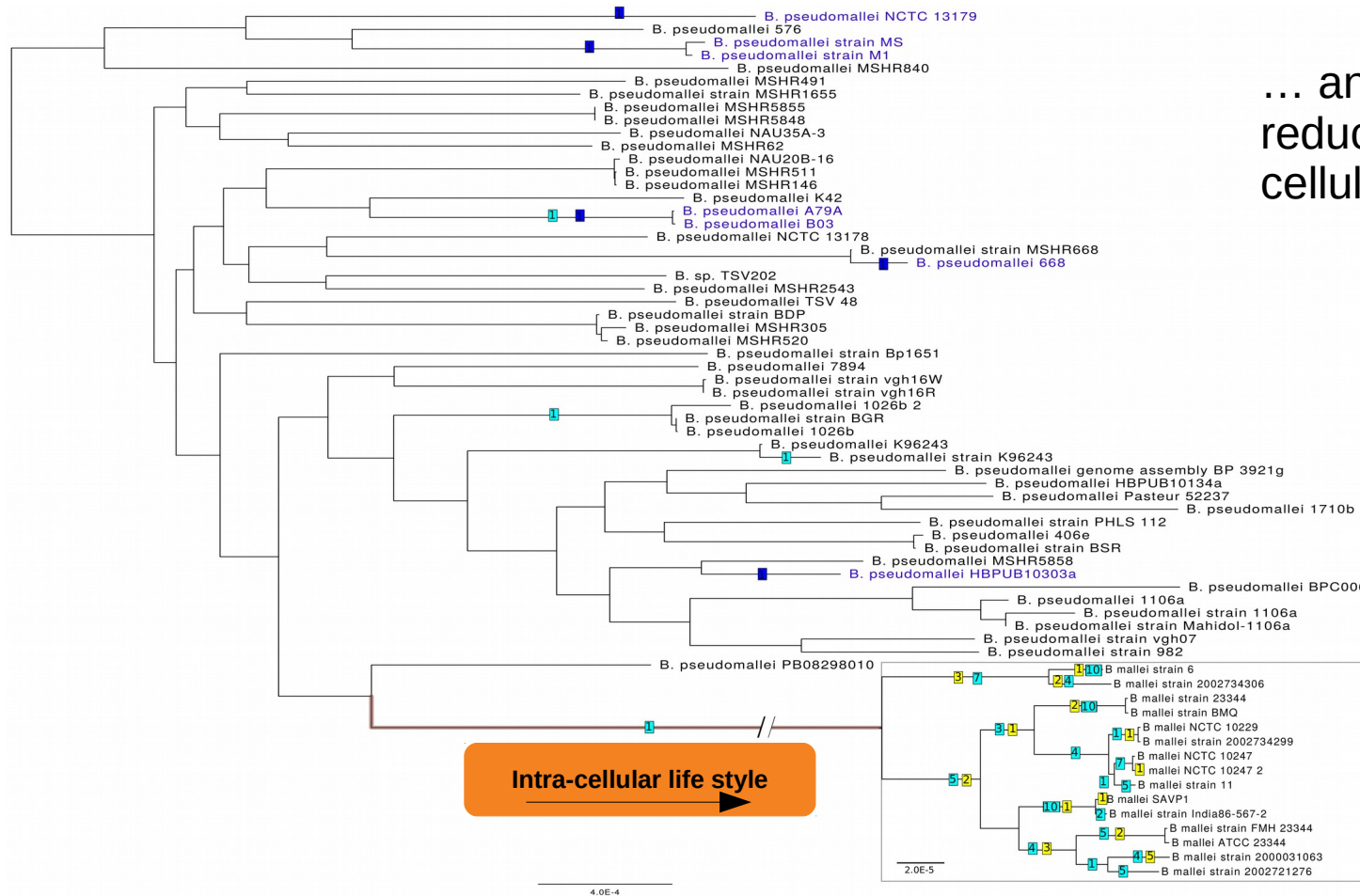
AE, Miklós I, Ragan MA. Dynamics of genome rearrangement in bacterial populations. PLoS Genet. 2008

# Pathogenic E.coli



*Shigella* strains are distinguished from *E. coli* by their nonmotility, absence of decarboxylated lysine and several biochemical characteristics, such as inability to ferment lactose and mucate.

# ISs expansion drives rearrangements

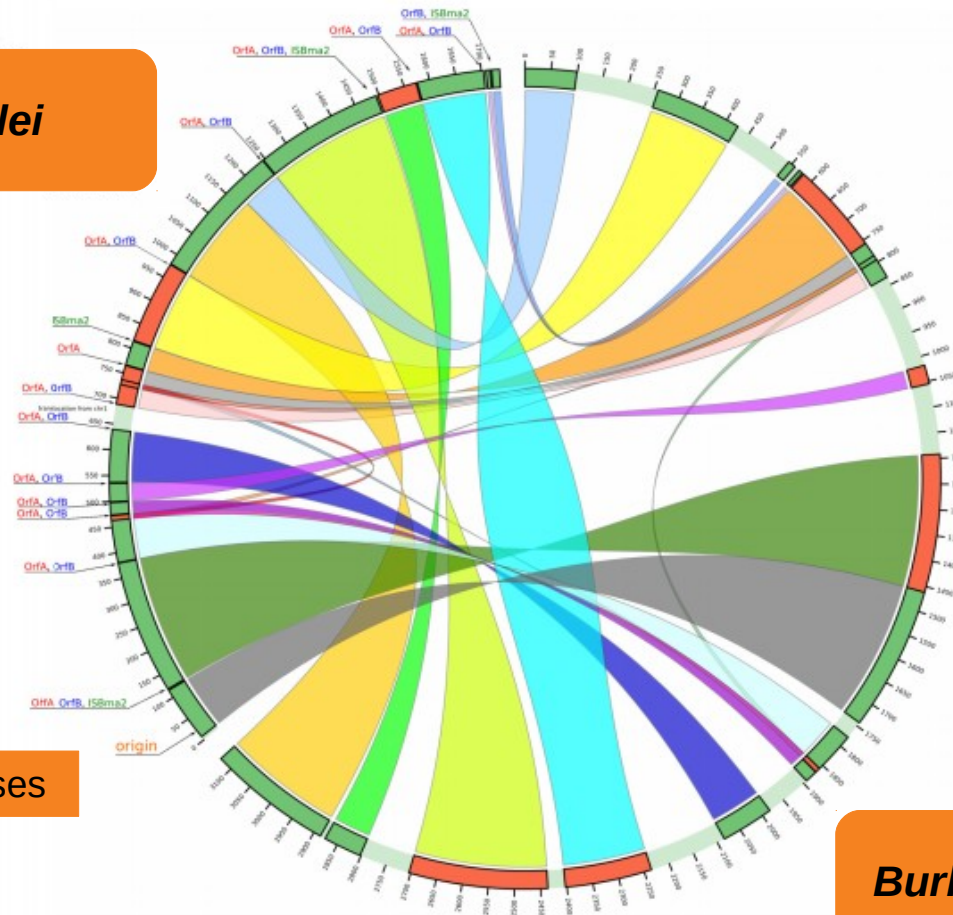


... and results in genome reduction in young intra-cellular pathogens



# Genome reduction by deletions

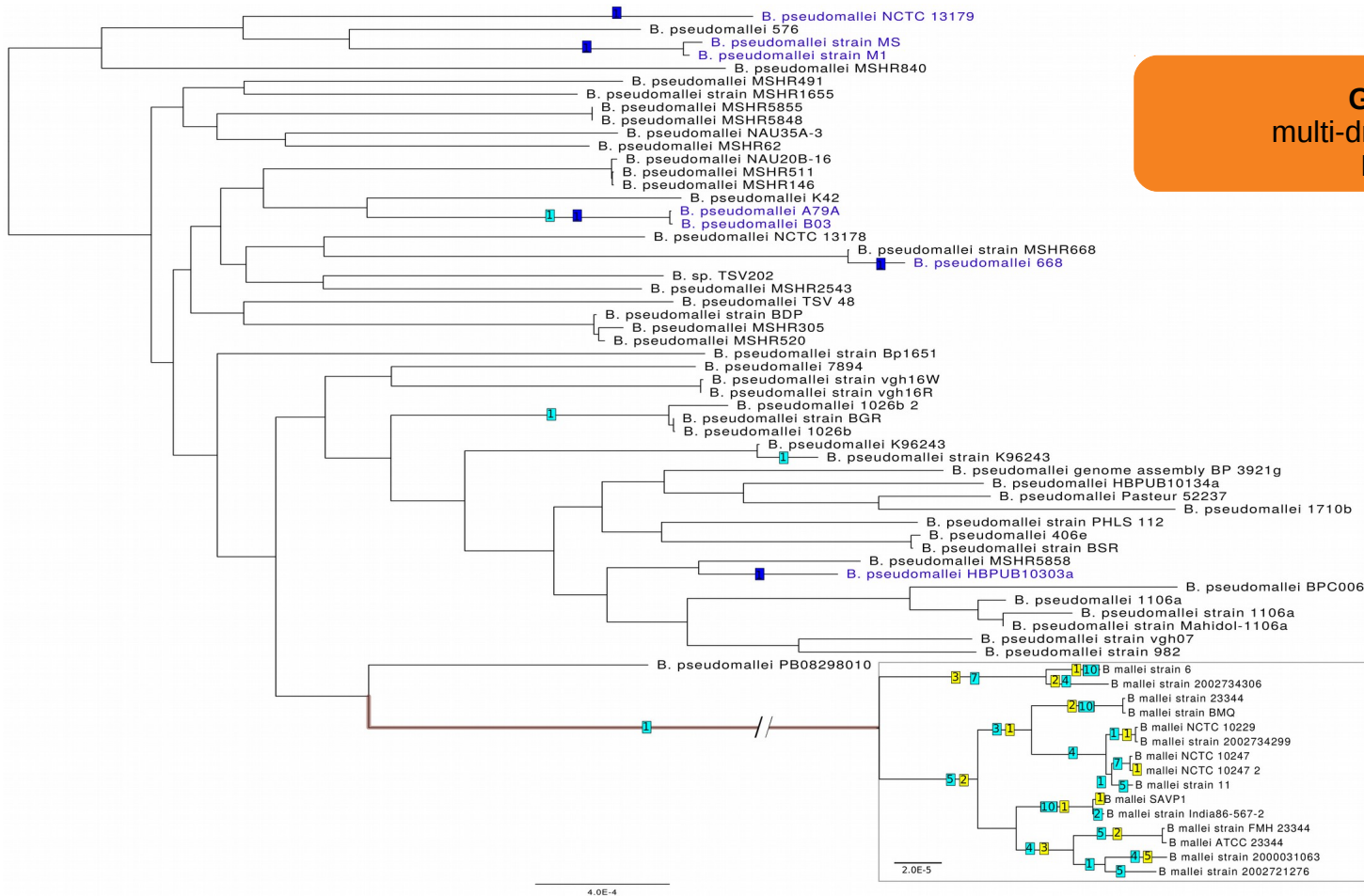
*Burkholderia mallei*



transposases

*Burkholderia pseudomallei*

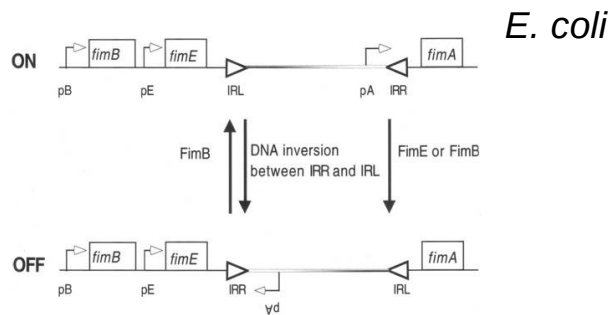
# Parallel inversions – genomic switch?



**Genomic repeats:**  
multi-drug resistance complex,  
RND efflux system

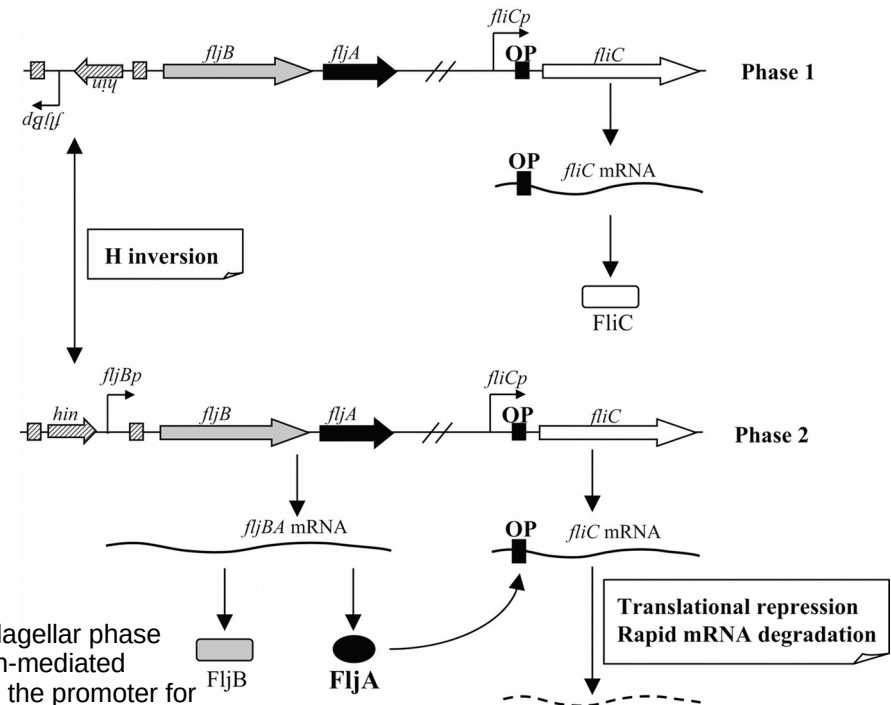
# Site-specific micro-inversions

## Phase variation – Site specific inversion



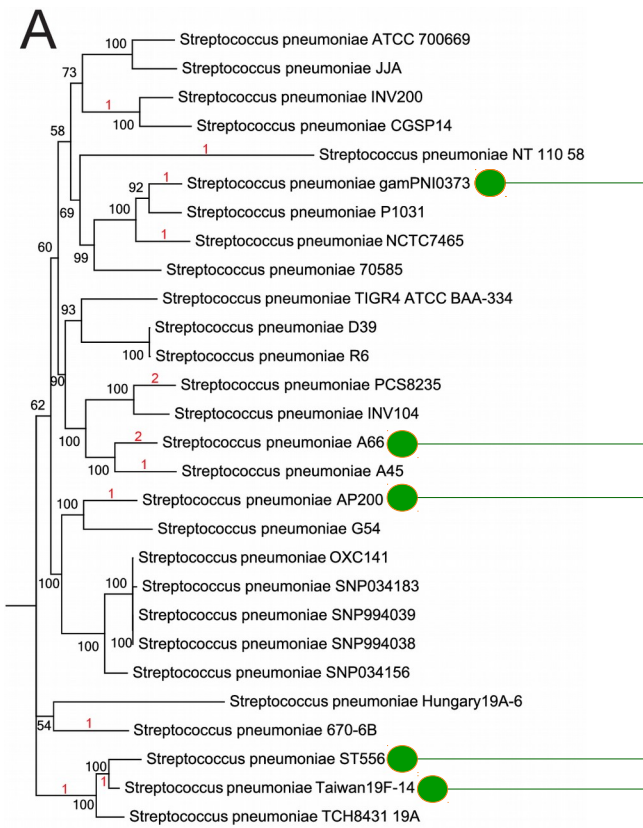
Phase variation – site specific inversion. pB, pE and pA are promoters for the genes *fimB*, *fimE* and *fimA* respectively. IRL and IRR are inverted repeats. FimB and FimE are recombinases that bind to the open triangles IRL and IRR. The result is an inversion of the DNA sequence (shaded bar) that turns ON or OFF the transcription of *fimA*.

## Expression of multiple types of flagellin by *S. typhimurium*



The dual controlling system governs flagellar phase variation; one part of the system is Hin-mediated inversion of the H segment containing the promoter for the *fljBA* operon, and the other is FljA-mediated inhibition of *fliC* expression. FljA binds to the 5'-UTR of *fliC* mRNA, which inhibits its translation and facilitates its degradation.

# Parallel inversions – genomic switch?

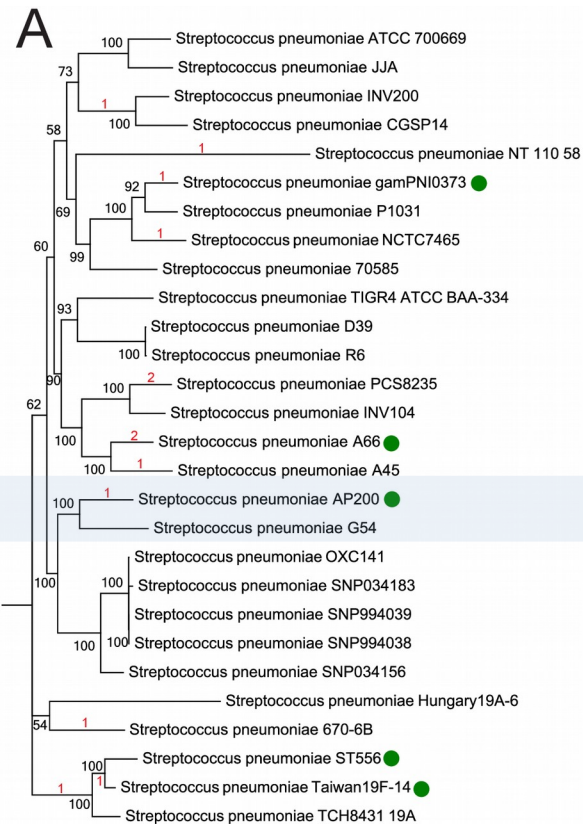


The same inverted sequence length 15kB  
Breakpoints are formed by genes encoding  
proteins PhtB and PhtD

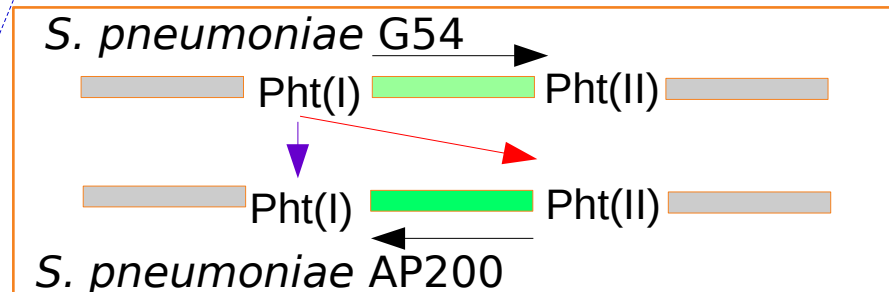
## Белки гистидиновой триады (Pht)

- белки внешней мембраны, антигены
- характеризуются наличием мотива  
*HxxHxH*

# Parallel inversions – genomic switch?



**B** Nucleotide alignments

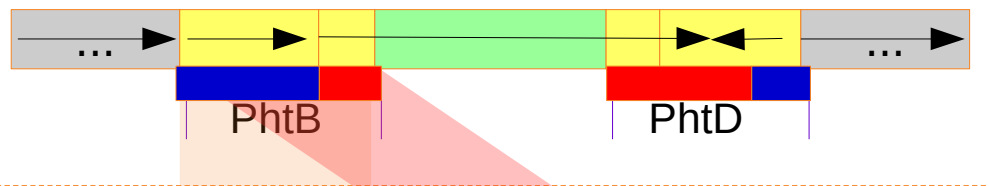


● The same inverted sequence length 15kB  
Breakpoints are formed by genes encoding  
proteins PhtB and PhtD

# Parallel inversions – genomic switch?

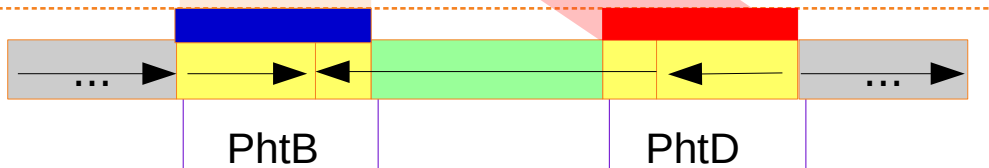
## Phts' alignment

*S. pneumoniae* AP200



Alignment  
t 1

Alignment  
t 2



*S. pneumoniae* G54

WP\_0018462

WP\_0018475

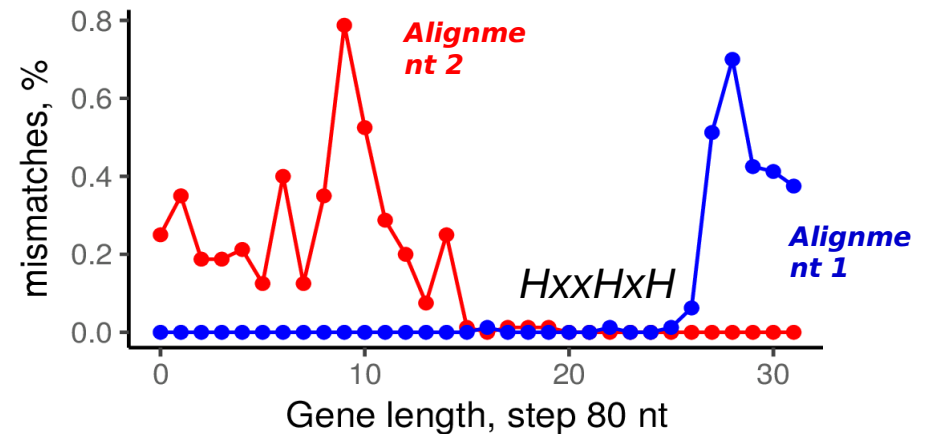
cons

1 MKINKKYVAGSVAVLALSVCSEYELGRYQAGQDKKESNRVAYIDGDQAG

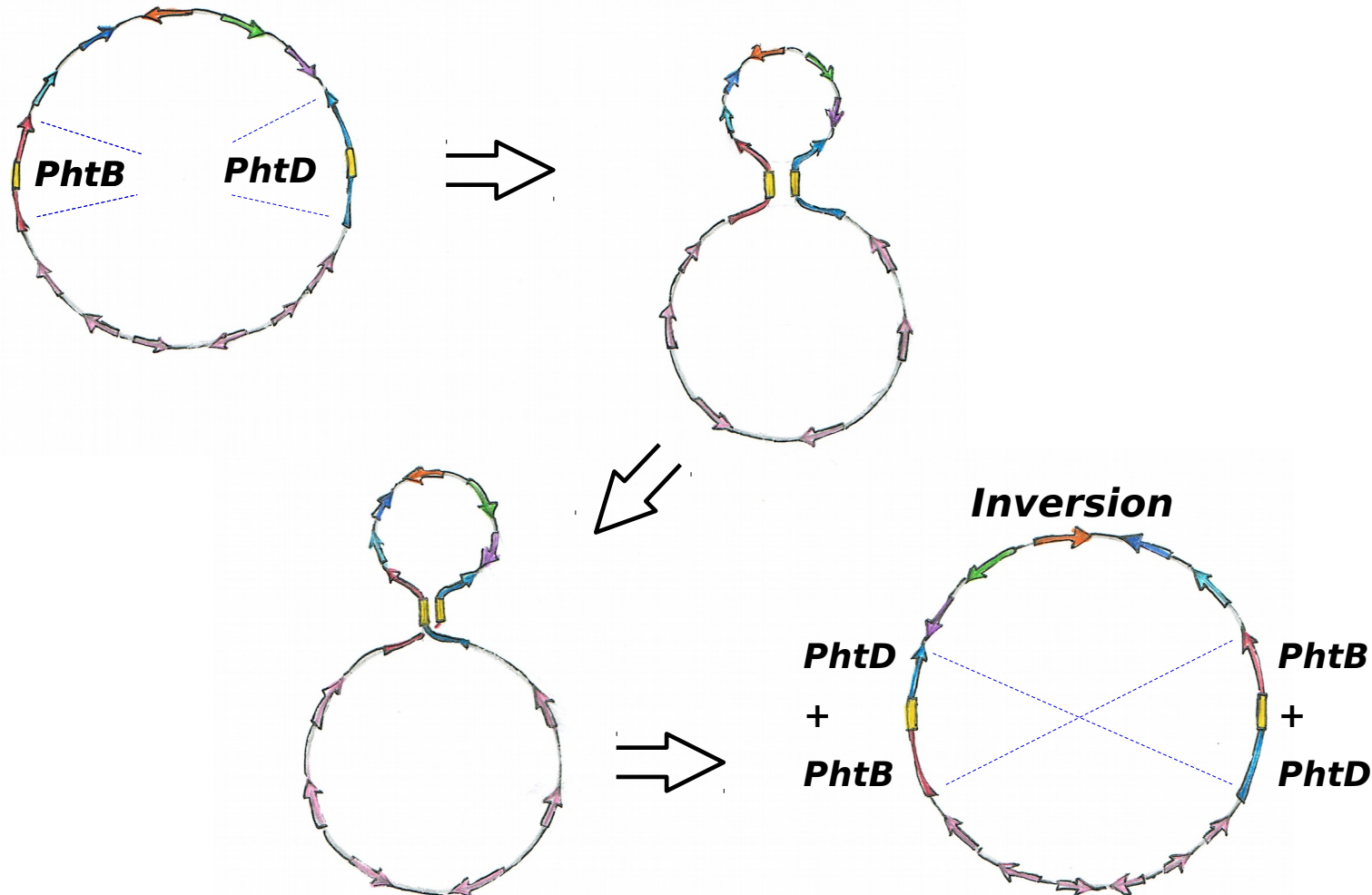
1 MKINKKYLAGSVATLVLSVCAYELGLHQA-QTVKENNRVSYIDGKQAT

1 \*\*\*\*\*:\*\*\*\*\*.\*.\*\*\*\*\*:\*\*\*\*\*:\*\*\* \*\* .\*\*\*\*\*:\*\*\*\*\*.\*

## Mismatches





# Parallel inversions – genomic switch?



# Parallel inversions – genomic switch?

## Биоинформатическое предсказание

### Micro-evolution of three *Streptococcus* species: selection, antigenic variation, and horizontal gene inflow

Pavel V. Shelyakin <sup>†</sup>  , Olga O. Bochkareva <sup>†</sup>, Anna A. Karan and Mikhail S. Gelfand

<sup>†</sup>Contributed equally

*BMC Evolutionary Biology* 2019 19:83

<https://doi.org/10.1186/s12862-019-1403-6> | © The Author(s) 2019

Received: 22 December 2017 | Accepted:  | Published: 





# Parallel inversions – genomic switch!!!



## Биоинформатическое предсказание

### Micro-evolution of three *Streptococcus* species: selection, antigenic variation, and horizontal gene inflow

Pavel V. Shelyakin <sup>†</sup>  , Olga O. Bochkareva <sup>†</sup>, Anna A. Karan and Mikhail S. Gelfand

<sup>†</sup>Contributed equally

*BMC Evolutionary Biology* 2019 19:83

<https://doi.org/10.1186/s12862-019-1403-6> | © The Author(s) 2019

Received: 22 December 2017 | Accepted: 25 February 2019 | Published: 27 March 2019

## Экспериментальное подтверждение

### Deep genome annotation of the opportunistic human pathogen *Streptococcus pneumoniae* D39

Jelle Slager, Rieza Aprianto, Jan-Willem Veening 

*Nucleic Acids Research*, Volume 46, Issue 19, 02 November 2018, Pages 9971–9989,

<https://doi.org/10.1093/nar/gky725>

Published: 13 August 2018 | [Article history](#) ▼

# Parallel inversions – genomic switch!!!

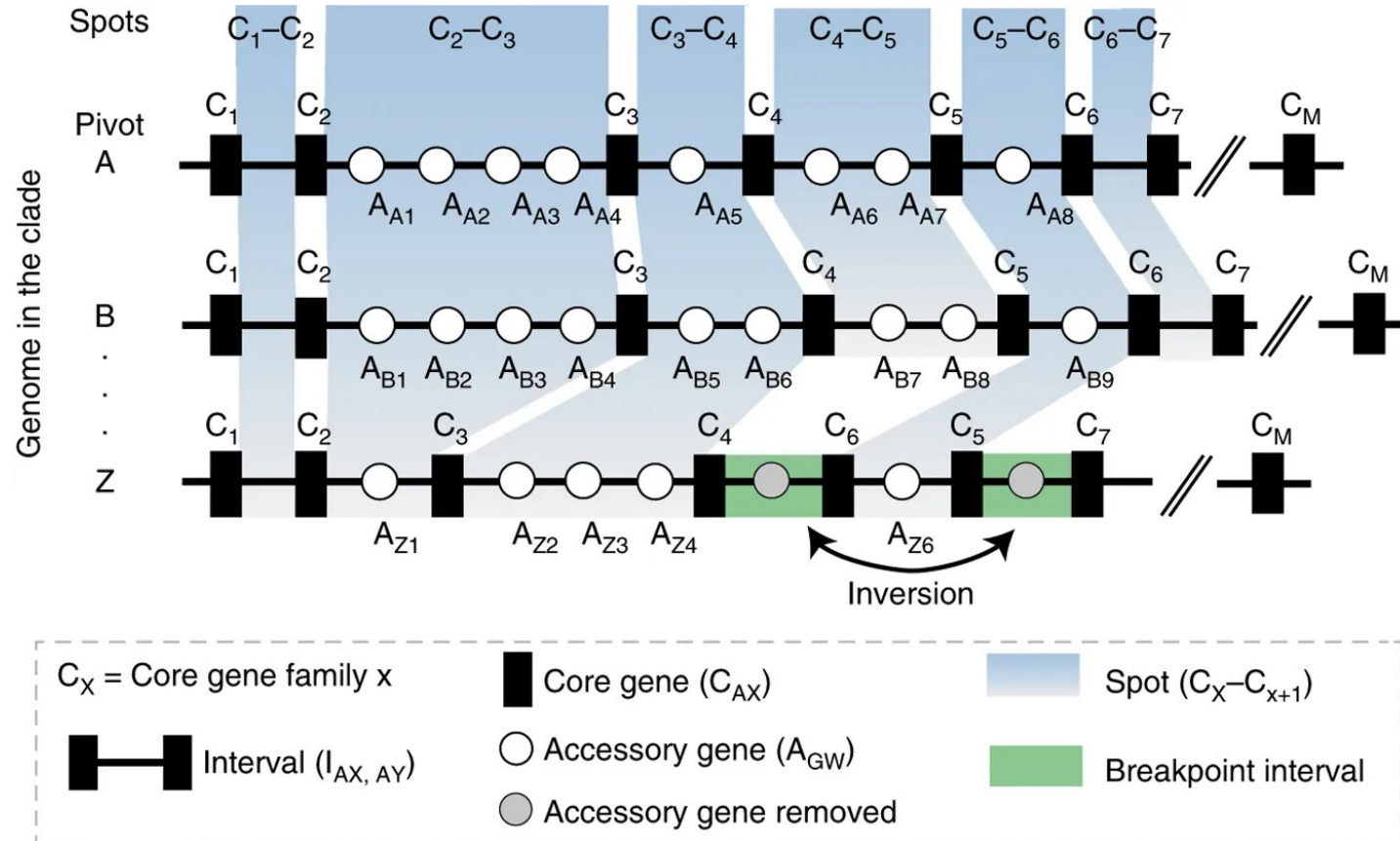


These two out of four pneumococcal histidine triad (Pht) proteins **are considered to be good vaccine candidates.**

**PhtD was already used in several phase I/II clinical trials.**

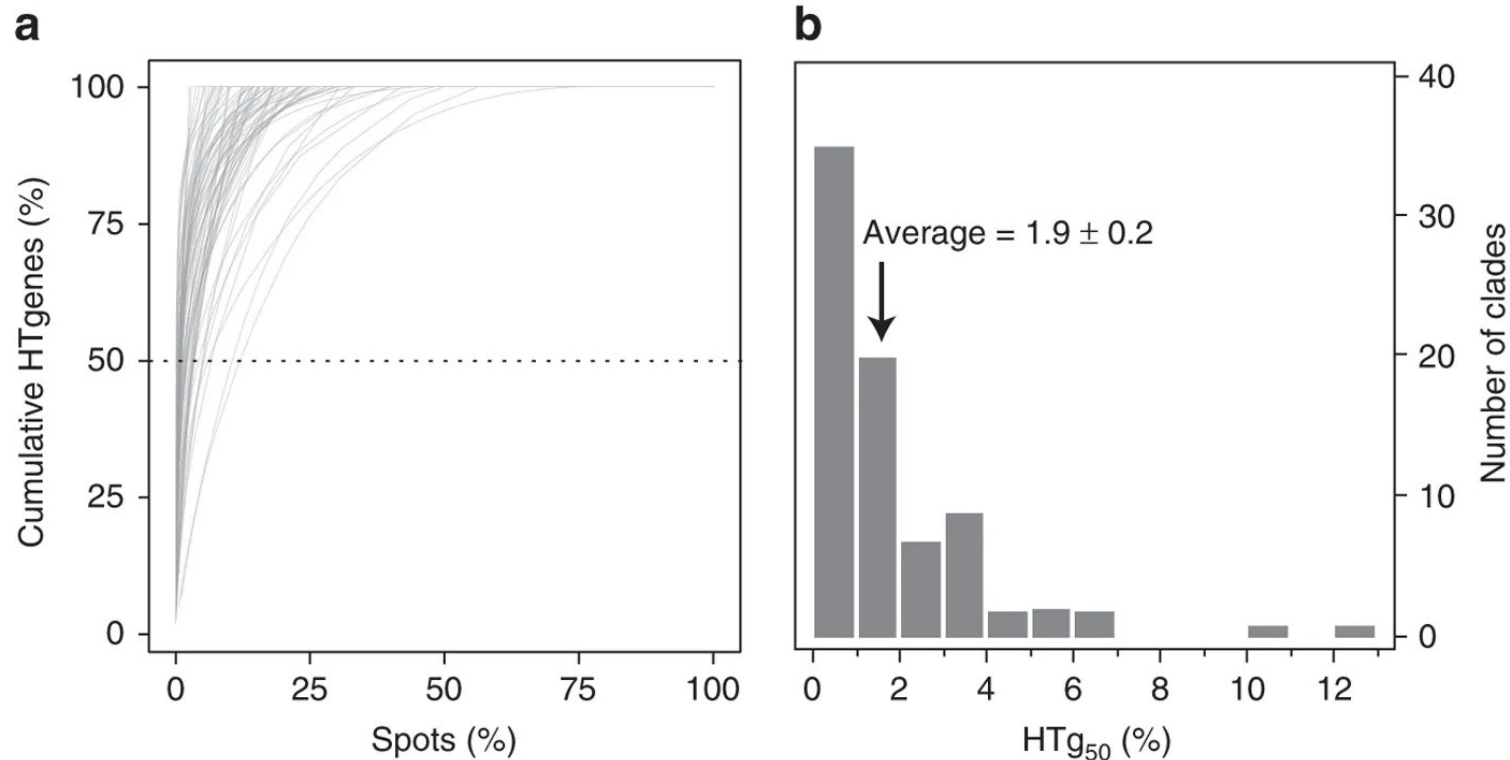
Yun et al.\* (PLoS One. 2015) analyzed the diversity of phtD alleles from 172 clinical isolates and concluded that **the sequence variation was minimal.**

# Evolution of gene content



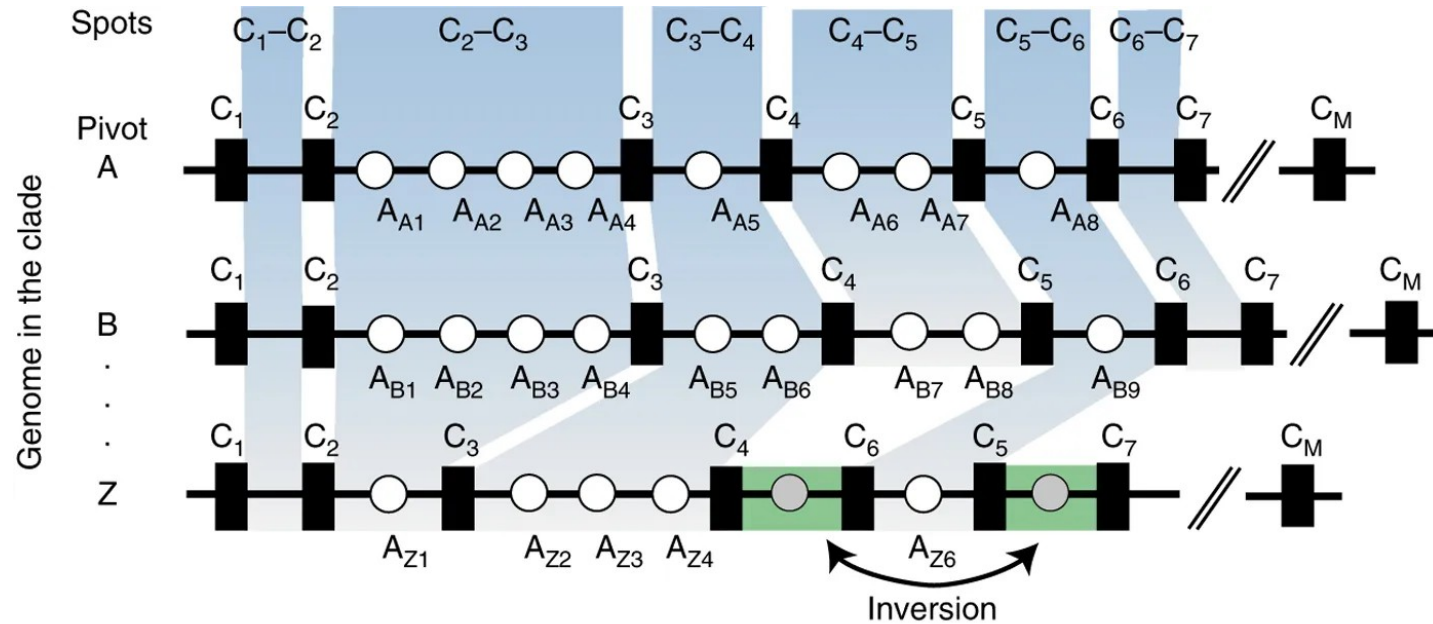
Oliveira, P.H., Touchon, M., Cury, J. et al. The chromosomal organization of horizontal gene transfer in bacteria. Nat Commun. 2017

# Evolution of gene content



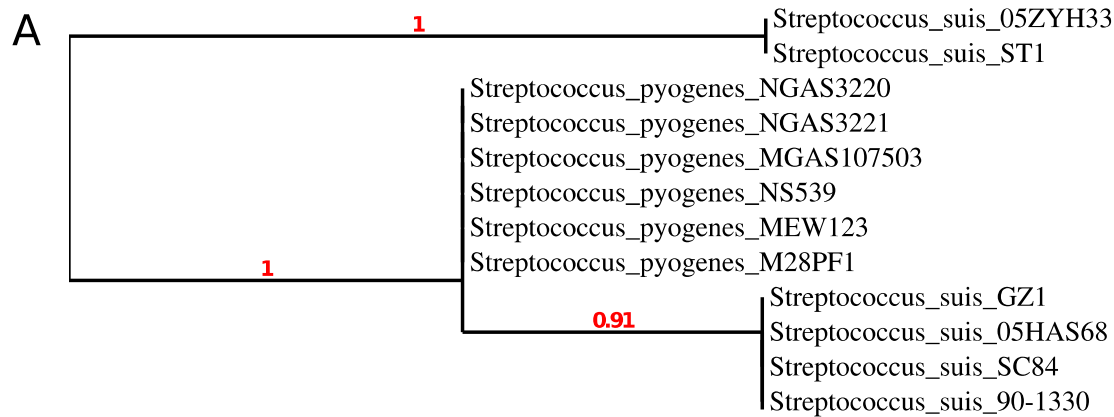
“A total of 1841 hotspots in the 80 clades. They represent only 1.2% of the spots, but they concentrate 47% of the accessory gene families and 60% of the HTgenes.”

# Evolution of gene content

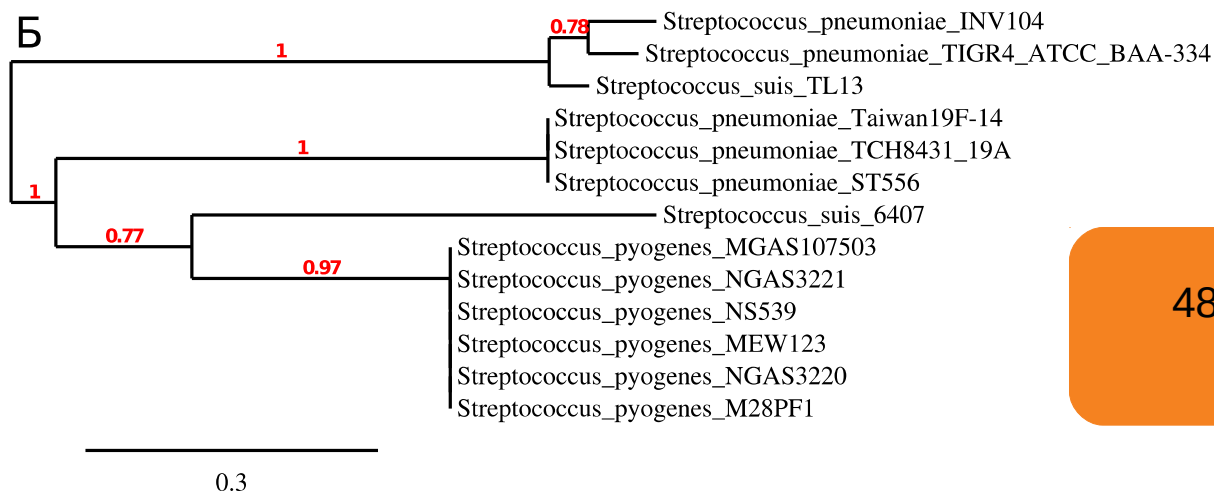


Orthologous groups in different breakpoint intervals are HT candidates

# Evolution of gene content

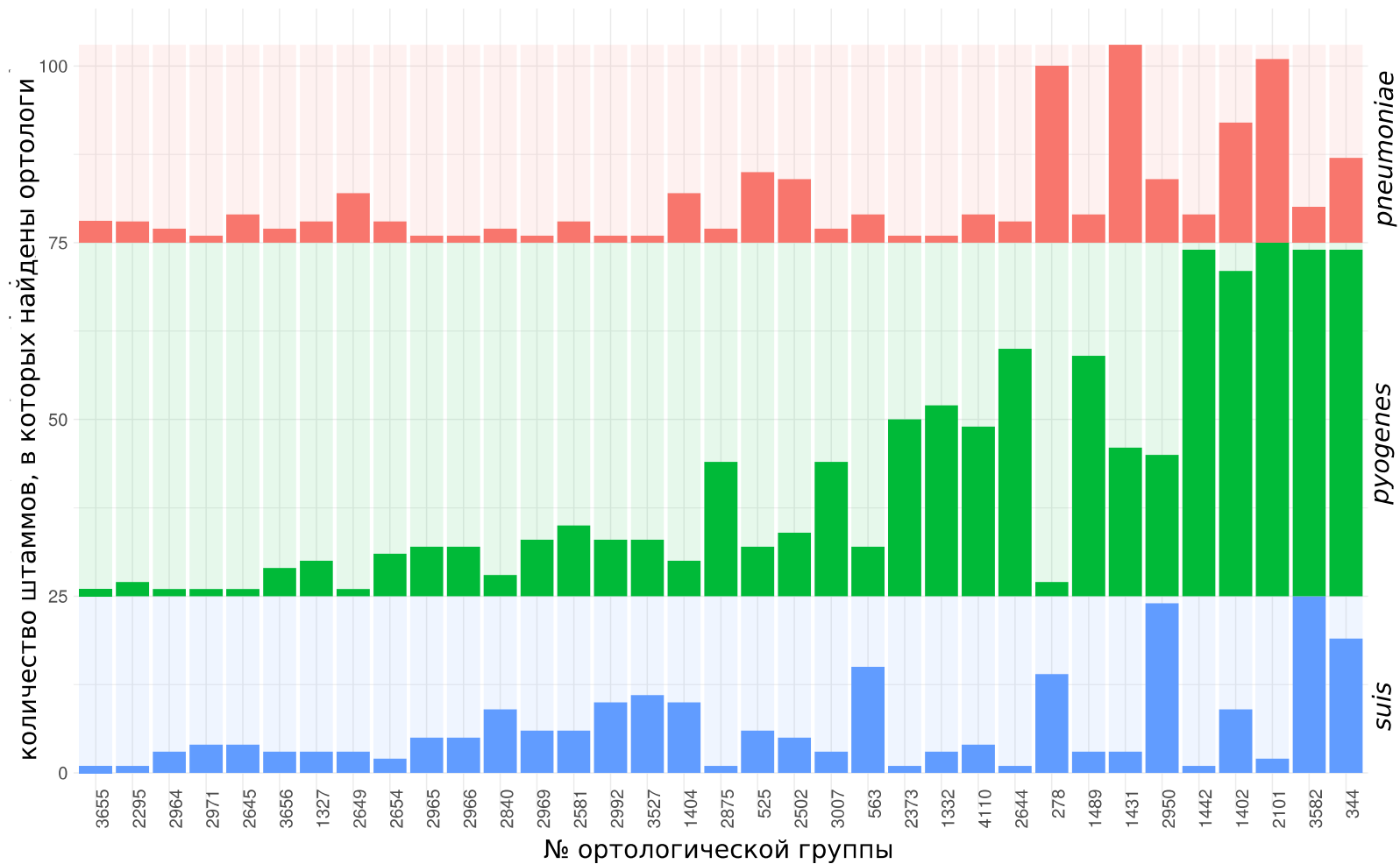


176 ortologous groups



48 of them have inconsistent phylogenetic trees

# Evolution of gene content



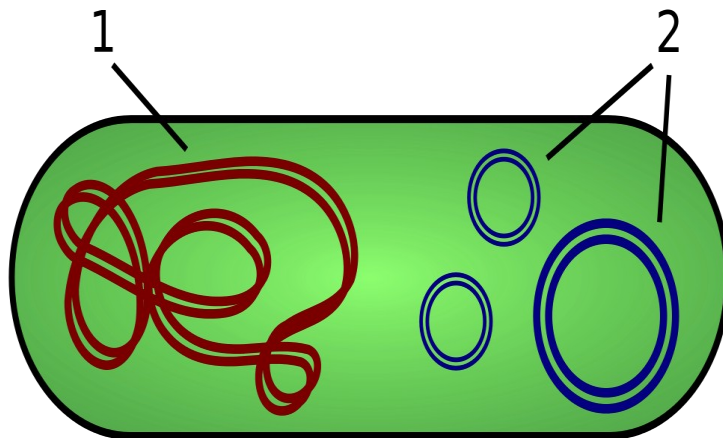
# Bacterial genome structure

## 1. Chromosome(-s)

- circular, double-stranded DNA molecule
- 1Mb – 6 Mb
- one origin of replication
- carry house-keeping genes and a lot of others

## 2. Plasmids

- circular, double-stranded DNA molecules
- 15 kb – 200 kb
- replicate independently
- carry genes that benefit the survival of the organism.

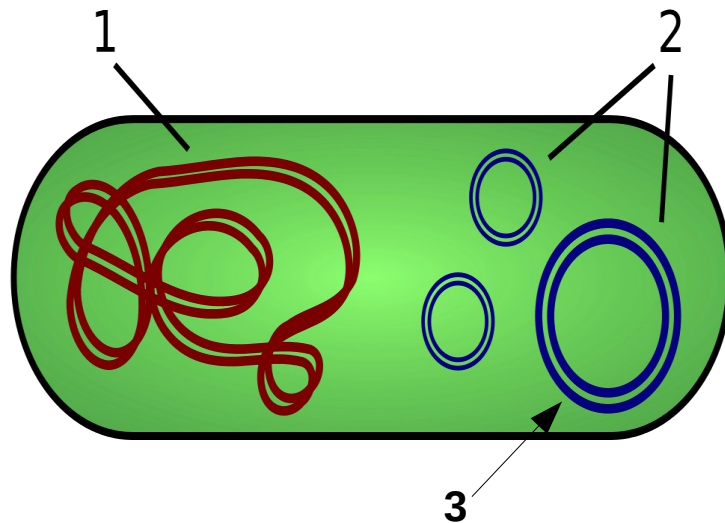




# Bacterial genome structure

## 1. Chromosome

- circular, double-stranded DNA molecule
- 1Mb – 6 Mb
- one origin of replication
- carry house-keeping genes and a lot of others



## 2. Plasmids

- circular, double-stranded DNA molecules
- 15 kb – 200 kb
- replicate independently
- carry genes that benefit the survival of the organism.

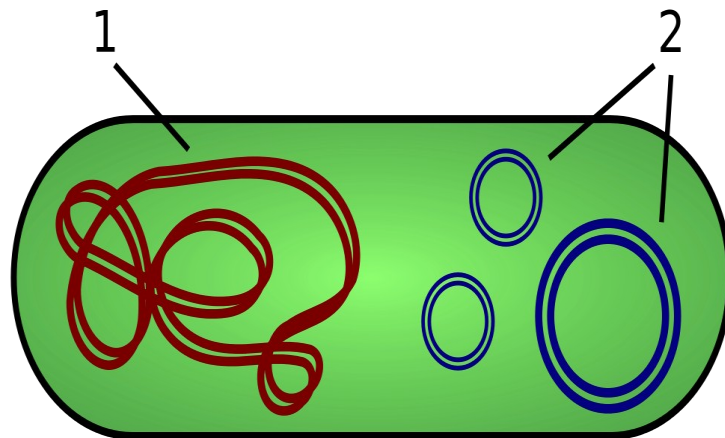
## 3. Other secondary replicons

- circular, double-stranded DNA molecule
  - 300 Mb – 3 Mb
- 
- plasmid type origin of replication
  - have small part of core genes
  - is subjected to regulation by chromosomally encoded mechanisms

# Bacterial genome structure

## 1. Chromosome

- circular, double-stranded DNA molecule
- 1Mb – 6 Mb
- one origin of replication
- carry house-keeping genes and a lot of others



**Chromosome + plasmid = chromid**

## 2. Plasmids

- circular, double-stranded DNA molecules
- 15 kb – 200 kb
- replicate independently
- carry genes that benefit the survival of the organism.

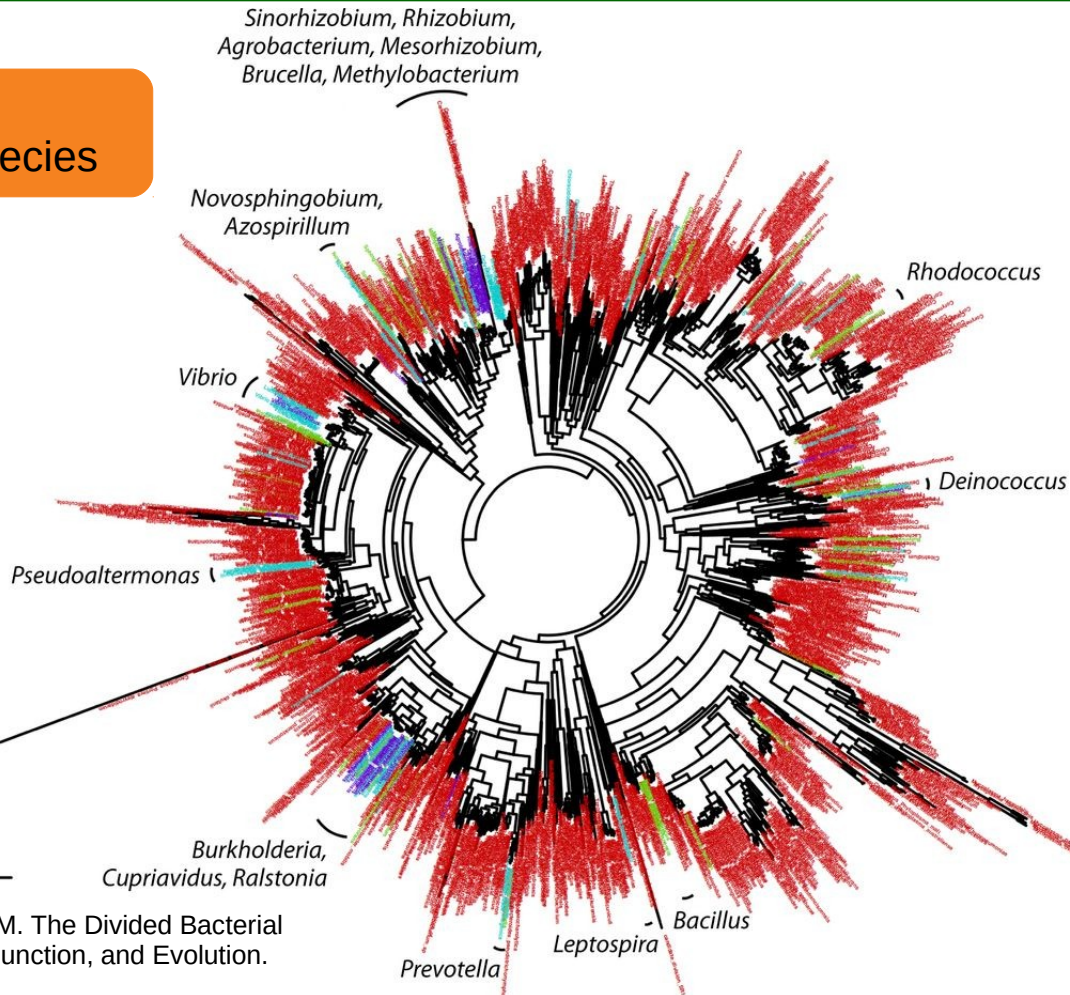
## 3. Other secondary replicons

- circular, double-stranded DNA molecule
- 300 Mb – 3 Mb

- 
- plasmid type origin of replication
  - have small part of core genes
  - is subjected to regulation by chromosomally encoded mechanisms

# Bacterial tree of life

**1,708**  
bacterial species



## Colors:

red for species with no megaplasmid or chromid

green for species with a megaplasmid(s) but no chromid

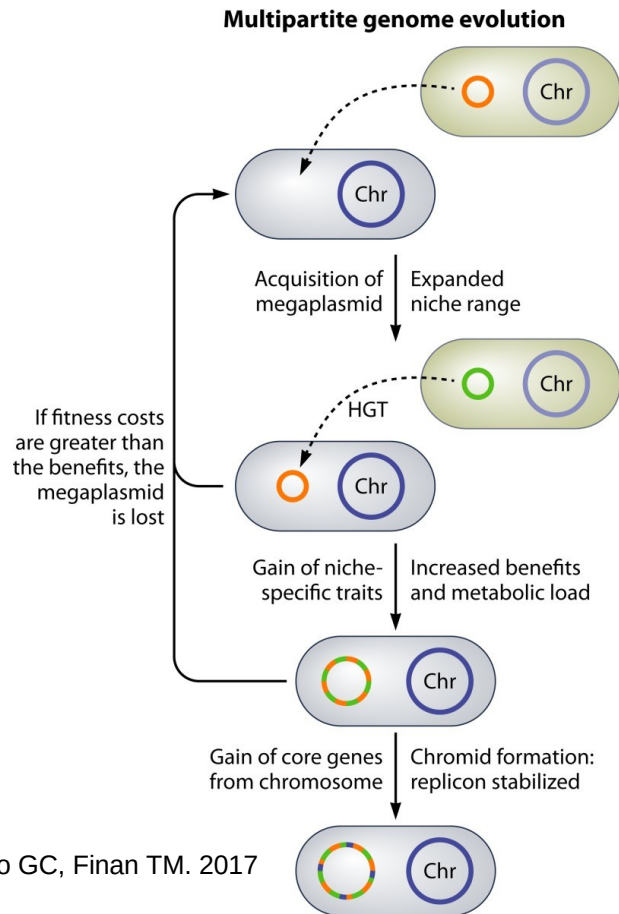
blue for species with a chromid(s) but no megaplasmid

purple for species with both a megaplasmid(s) and a chromid(s).

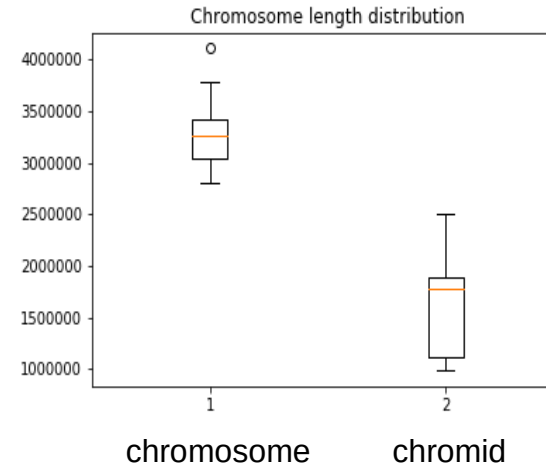
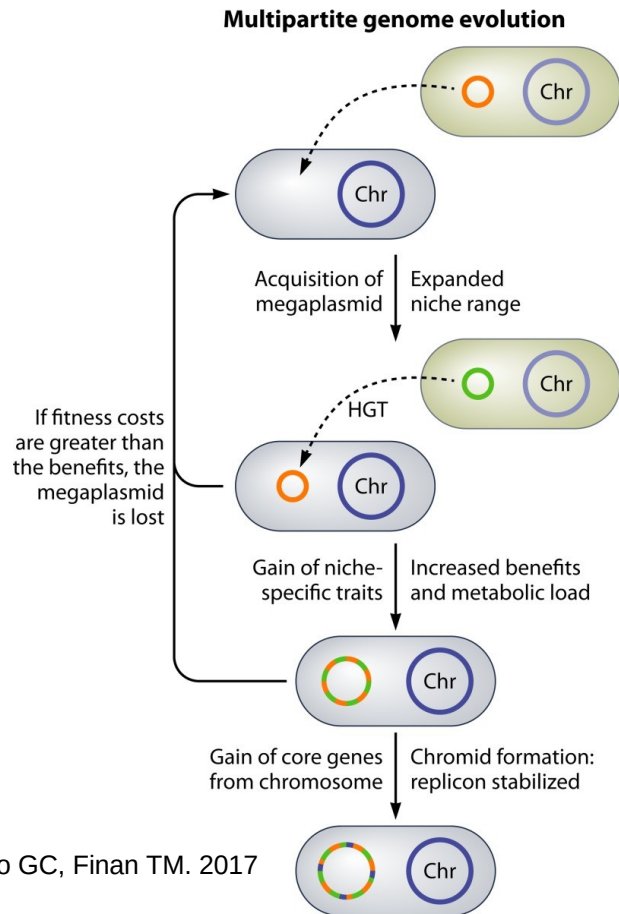
**10%** of bacterial genomes are split between two or more large DNA fragments

diCenzo GC, Finan TM. The Divided Bacterial Genome: Structure, Function, and Evolution. 2017

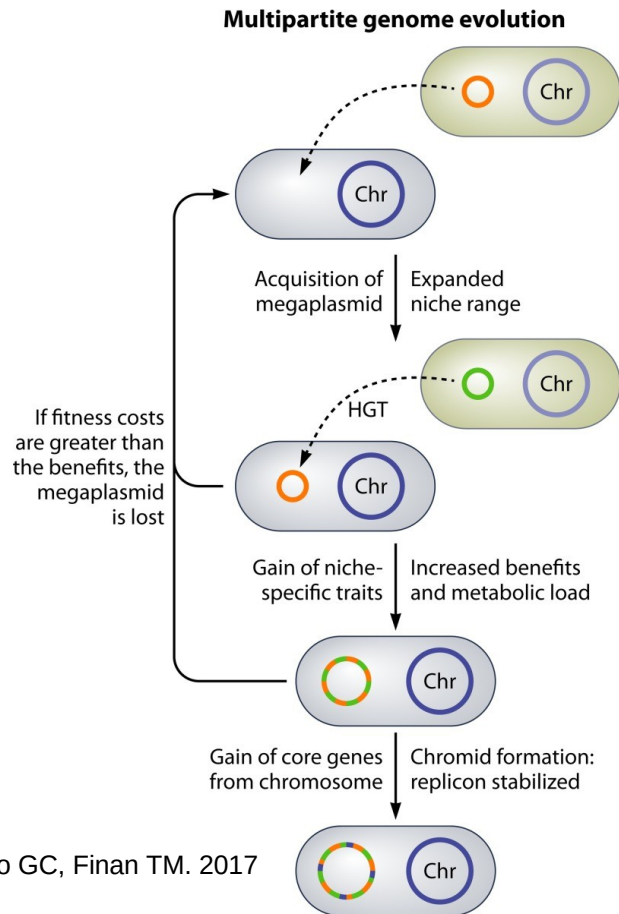
# Chromid formation hypothesis



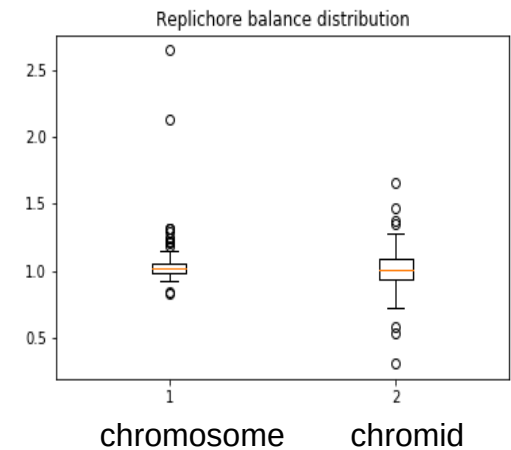
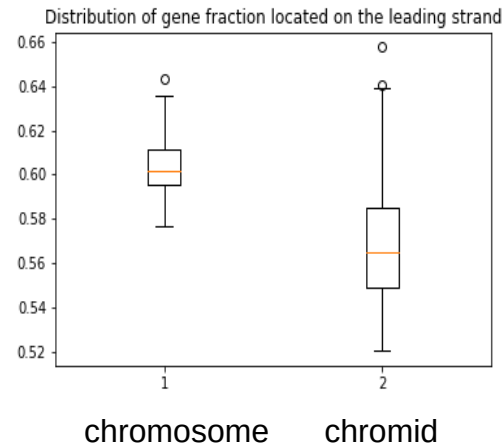
# Chromid formation hypothesis



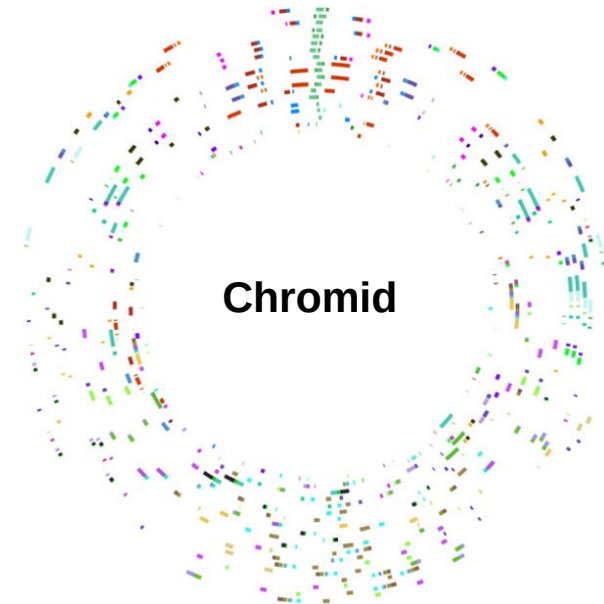
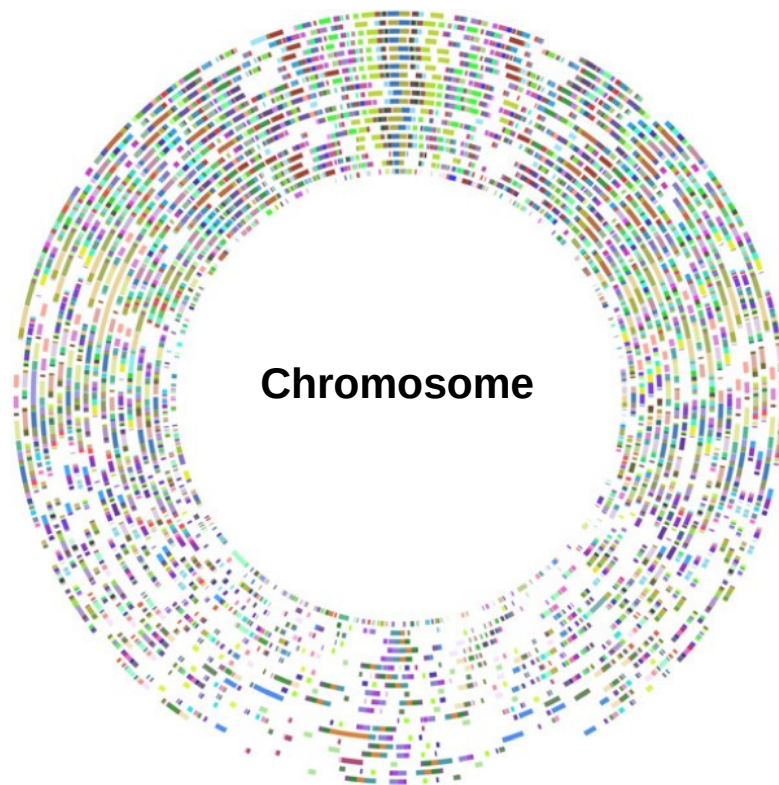
# Chromid formation hypothesis



diCenzo GC, Finan TM. 2017



# Chromosome and chromid structure



Dikow, Rebecca B, and William Leo Smith. "Genome-level homology and phylogeny of Vibrionaceae (Gammaproteobacteria: Vibrionales) with three new complete genome sequences." *BMC Microbiology* vol. 13 80. 11 Apr. 2013, doi:10.1186/1471-2180-13-80

# Acknowledgments

ИППИ РАН  
Михаил Гельфанд,  
Павел Шелякин, Елена Мороз

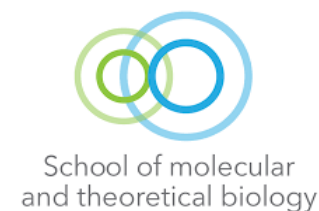
IST Austria  
Федор Кондрашов

ФББ МГУ  
Анна Карань  
Кристина Перевощикова  
Заира Сефербекова

ШМТБ  
Анна Тоидзе



This program has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 754411.





# IST Austria

## **Для студентов:**

internship: the ISTernship summer program and year-round scientific internships

## **Для поступающих в аспирантуру:**

interdisciplinary PhD program

## **Для постдоков:**

ISTplus postdoctoral program

<https://ist.ac.at/en/home/>

