



Structure and evolution of bacterial genomes

Olga Bochkareva

postdoc, Kondrashov group (olga.bochkaryova@gmail.com)

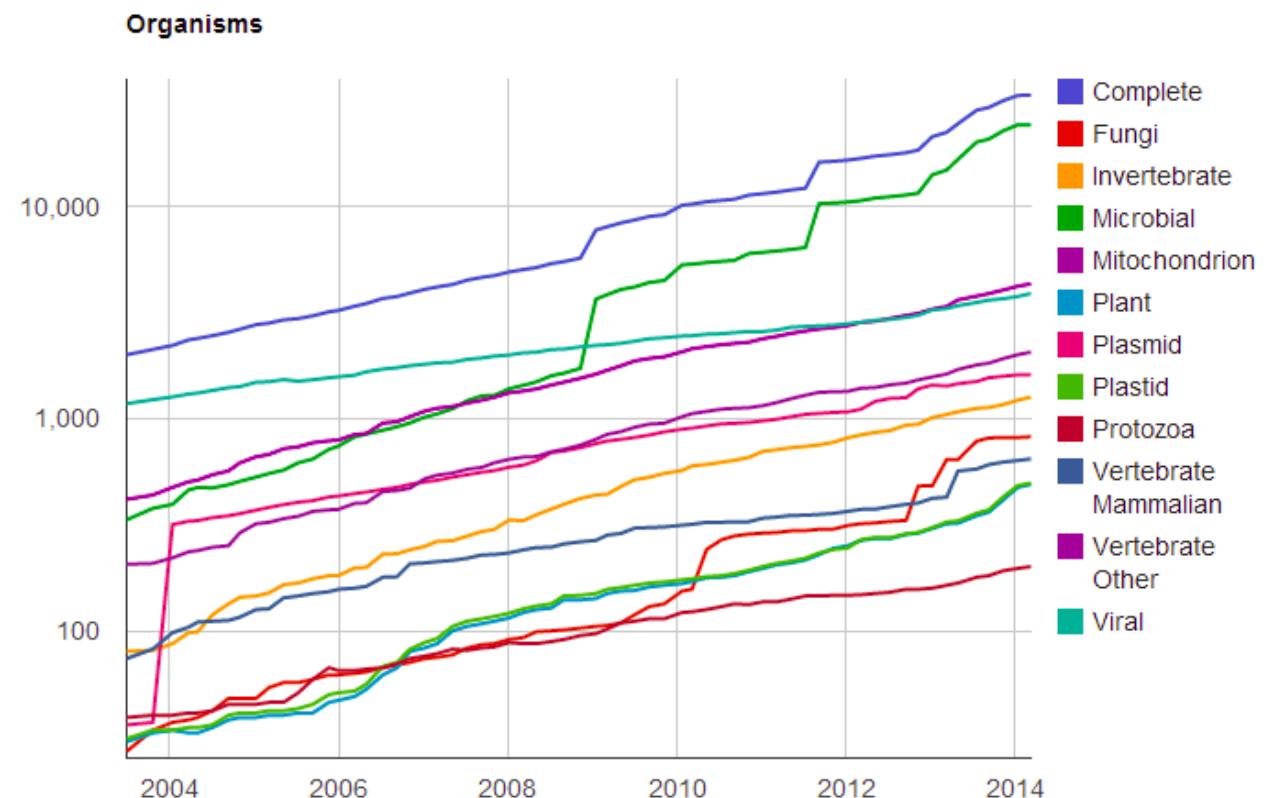
Data is coming ...

Bacteria

479 320 assemblies

↓
16 527 complete genome assembly level

↓
1 631 representative genomes



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

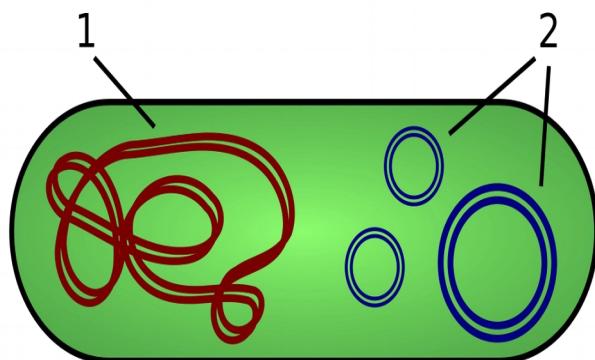
What is bacterial genome?

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

```

ASASC3_1 14 SIKLUMPPSQTTRLLVERMANNLST..PSIFTRK..YGLSLKEERPENRKOTIEEVACSTRNQ....HYEKEPODGGSAVOLYAKESKRLTILEVVK 101
B4F917_1 13 SIKLUMPPSESTRIMIVDMHTNNLST..ESIFSRK..YRLLGKQEARHENRKTTIELFLALADE....HFRREEPODGGSAVOLYAKESKRLTILEVVK 100
A9S1V2_1 23 VFKLUMPPSQTTRDVMVNRMLKLS...ACFESOS..PARTELDPRDREMRRIEVEFGRACE....ADSGCQDTGSAVMVYAHASKRLTILEVVK 109
B9H9V1_1 17 SIKLUMPPSQTTRDVMVNRMLKLS...ACFESOS..PARTELDPRDREMRRIEVEFGRACE....ADSGCQDTGSAVMVYAHASKRLTILEVVK 100
Q8H056_1 30 SFSIUMPPORTTDWVVRILVOTVGC..DTILSKR..YGHVPADECPARCTIRENDRDRR..SGEMARTSVEETZIRALOLYSKEVSRRLLFVK 120
000423_2 44 SLSIUMPPORTTDWVVRILVOTVLR..PSILSKR..YGAPEAEAGRAAAGVEAEVARYARVTES..SSARAAAPPSVEDGIEVLDAVSKYEVSRLLFVK 135
B9MVW8_1 56 SFSIUMPPORTTDWVVRILVOTLST..TSVLSKR..YTIPIKEESEASERKRIEFEERPSGAST....VASEXEDGLEVLVOLYSKEISKRMLETVK 141
000423_1 25 SIKLUMPPSQTTRDVMVNRMLKLS...ACFESOS..PARTELDPRDREMRRIEVEFGRACE....ADSGCQDTGSAVMVYAHASKRLTILEVVK 100
R9NA46_1 12 SIKLUMPPSQTTRLMVERMTDNSS..VSFFSRK..YGLLSKEERPENRKTTETAFBLAND...HEKERNLDOSSVVFYAREGRKLMLFALK 100
09C500_1 57 SIKLUMPPSQTTRDWVNLNRLIETLST..ESILSKR..YGTLSKDDATTVAKLIEEAYGVASN....AVBSDODGKILELYSKETISKRMLETVK 142
02HR17_1 28 NYSIUMPPORTTDWVNRMLKLS...PSVLTKA..YOTMSAEDERSAHAIQIEDERFSVVHN....SSSTSDDNVTILEVYSKEISKRMLETVK 110
09C500_2 20 SIKLUMPPSQTTRDWVNLNRLIETLST..ESILSKR..YGTLSKDDATTVAKLIEEAYGVASN....AVBSDODGKILELYSKETISKRMLETVK 100
Q9M7N6_1 25 SFSIUMPPORTTDWVNLNRLIETLST..PSILSKR..YGTLPQDAASETBRLTIEEERFRRGGS...TASDHQDNVTEILLOVYSKEISKRMLETVK 110
09LE82_1 14 SVKUMPPPSQTTRLMVERMTNTT..PSIFSRK..YGLLSVVEEHDOKRPIEOLFATANK....HFDNEPODGDTSAVHVYAKESKRLMDLVVK 101
Q9M651_2 13 SIKLUMPPPLPTRKNLIERITINPSS..KTIFTRK..YGLSLTKDQHTEENKRRIEIDIFRSTANG....OFREREPODGGSAVOLYAKESKRLTILEVVK 100
B9R748_1 48 SLSIUMPPORTTDWVNLNRLIETLSS..PSVLSKR..YGTISHDEAESARRIEDEHFGVHN....RTSHEDQALEIQLQYSKEISKRMLETVK 133

```

How does it evolve?

What we can observe in genomes

- Mutations – genes sequences

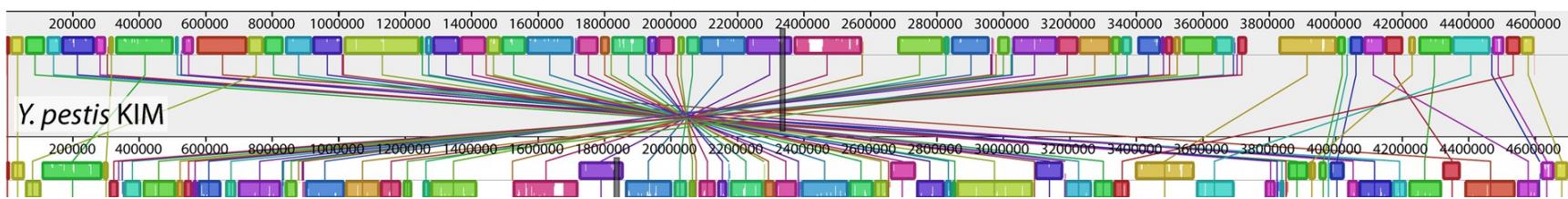
```

ASASC3_1 14 SIKLUPPSQTTRLLVERMANNLST..PSIFTRK..YGLSLKEERPENRKOTIEEVACSTRNQ..,....HYEKEPODGGSAAVOLYAKESKRLTILEVK 101
B4F917_1 13 SIKLUPPSESTRIMLVDHMNLLST..ESIFSRK..YRLLGKQEARHENKTTIELFLALADE..,....HFEKEPODGGSAAVOLYAKESKRLTILEVK
A9S1V2_1 23 VFKLUPPSSQTRDVMVDMHKLSS..ACFESOS..FARIELDGRHEMRRIEVEVGRACE..,....ADSGQDKTSGVVMVYAHASHKLMLETL 109
B9S1V3_1 17 SIKLUPPSESTRIMLVDHMNLLST..ESIFSRK..YRLLGKQEARHENKTTIELFLALADE..,....HFEKEPODGGSAAVOLYAKESKRLTILEVK
Q8H056_1 30 BSIUIMPPPTQRTTDWVWRILVOTGG..DTILSKR..YGHVPADKEPARGTIRENDRDRR..,....SGEAKARTSVEETZIRALOLSYKEVRRLLLEVK 120
000423_2 44 SLSIUMPPPTQRTTDWVWRILVOTGV..PSILSKR..YGAPEAEAGRAAGAEAEVAVARYAVTES..SSARAPRPSVEDGIEVELDAYSKEVSRLLEVK 135
B9MVW8_1 56 BSIUIMPPPTQRTTDWVWRILVOTLST..TSVLSKR..YTIPIKEESEASERKIEEERPSGAST..,....VASSEDQGLEVLVOLYSKEISKRMILTVK 141
000424_2 25 SIKLUPPSESTRIMLVDHMNLLST..ESIFSRK..YRLLGKQEARHENKTTIELFLALADE..,....HFEKEPODGGSAAVOLYAKESKRLTILEVK 142
R9N446_1 12 SIKLUPPSESTRIMLVDHMNLLST..VSFFSRK..YGLLSKEERAEENKTTIELFLALADE..,....HEKPNLDDSVVDFYAREGRKLMLFALK 100
09C500_1 57 SIKLUPPSESTRIMLVDHMNLLST..ESILSKR..YGTLSKDDATTVAKLIEEEAYGVASN..,....AVBSDODGIKILELYSKETISKRMILSVK 142
02HR17_1 25 NYSIUMPPPTQRTTDWVWRILVOTLST..PSVLTK..YOTMSDERSAHAIQIEDERFSVHN4..,....SSSTSDDNVTILEVYSKEISKRMILETVK 110
02HR18_1 20 SIKLUPPSESTRIMLVDHMNLLST..ESILSKR..YGTLSKDDATTVAKLIEEEAYGVASN..,....AVBSDODGIKILELYSKETISKRMILSVK 110
Q9M7N6_1 25 BSIUIMPPPTQRTTDWVWRILVOTLST..PSILSKR..YGTLPDQE4SETBRLTIEEEAFRRGGS..,....TASDHQDNVTEILLOVYSKEISKRMILTVK 110
09LE82_1 14 SVKMPUPPSKSTRLMVERMKNNTT..PSIFSRK..YGLLSVVEEHDOKRPIEDLIFRATANK..,....HFEKEPODGGSAAVOLYAKESKRLLMDLVK 101
Q9M651_2 13 SIKLUPPPLPTRKNLIERLITNPF..KTIFTEK..YGLSLTKDQHTENKRRIEDLIFRSTANG..,....OFEKEPODGGSAAVOLYAKESKRLLMDLVK 100
B9R748_1 48 SLSIUMPPPTQRTTDWVWRILVOTLSS..PSVLSKR..YGTISHDEESEASRRIEDLIFRVGNT..,....RTHEDQALEIOLQYSKETISKRMILTVK 133

```



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



How does it evolve?

What we can observe in genomes

- Mutations – genes sequences

```

A5ASC3_1 14 SIKLUPPSQTTRLLVERMANNLST..PSIFTRK..YGLSLKEERPENRKOTIEEVACSTRNQ..,....HYEKEPDGIGGSAVOLYAKESKRLTILEVK 101
B4F917_1 13 SIKLUPPSQTTRIMVDMTMNLLST..ESIFSRK..YRLLGKQEARHENKTTIELCFALADE..,....HFEKEPDGIGGSAVOLYAKESKRLTILEVK 100
A9S1V2_1 23 VFKLUPPSQTTRDVMVOMHLKLS..ACFESOS..FARTIELDRCRMRKRIEVEPRGRCE..,....ADSGQDKTSGVMMVYAHASKLMLETL 109
B9V9W_1 17 SIKLUPPSQTTRDVMVOMHLKLS..ACFESOS..FARTIELDRCRMRKRIEVEPRGRCE..,....ADSGQDKTSGVMMVYAHASKLMLETL 109
Q8H056_1 30 SFISIUPPTQORTDWMVVRILVOTGG..DTILSKR..YGHVPARDEPARCIEERFDRA00A..SGEMARATSVETZIRKALOLSYKESVRLLVLK 120
000423_2 44 SLSIUPPSQTTRDWMVVRILVOTGR..PSILSKR..YGAPEAEAGRAAAGREAYARVATES..SSARAAAPPSVEDGIEVELD04SKESVRLLVLK 135
B9MVW8_1 56 SFISIUPPTQORTDWMVVRILVOTLVR..PSILSKR..YTIPIKEESEASERKRIEERPSGAST..,....VASSKEDQLEVLVOLYSKESVRLLVLK 141
B9V9W_1 57 SFISIUPPTQORTDWMVVRILVOTLVR..PSILSKR..YTIPIKEESEASERKRIEERPSGAST..,....VASSKEDQLEVLVOLYSKESVRLLVLK 141
R9NA46_1 12 SIKLUPPSQTTRLMVERMTNLLSS..VSFFSRK..YGLLSKEERPENRKOTIEEVACSTRNQ..,....HEKPNLDOSSVIVYAREGRKLMLFLK 100
Q9C500_1 57 SIKLUPPSQTTRDWLVNLKLIETLST..ESILSKR..YGTLSKDDRTTVKULIEEEAYGVASH..,....AVBSDODGKILELYSKESVRLLVLK 142
02HR17_1 25 NYSIUPPSQTTRDWVNKLKLIETLST..PSVLTKR..YOTMSDERSAHAI01DERPSVVAH..,....SSSTSDDNHTILEVYSKESVRLLVLK 110
02HR17_2 26 NYSIUPPSQTTRDWVNKLKLIETLST..PSVLTKR..YOTMSDERSAHAI01DERPSVVAH..,....SSSTSDDNHTILEVYSKESVRLLVLK 110
Q9M7N6_1 25 SFISIUPPTQORTDWMVNLKLIETLST..PSILSKR..YSTLPD0E4SETBRLIEEEAFRAGGS..,....TASDHQDNHTEILDVYSKESVRLLVLK 109
09LE82_1 14 SVKMPUPPSQTTRLMVERMTKNNTI..PSIFSRK..YGLLSVEEHEODDKRPIEIDLAFITANK..,....HFNEPDGODGTSVAVHVYAKESKMLLVLK 101
B9M651_2 13 SIKLUPPSQTTRKLKLIETINPFS..KTIFTEK..YGLSLTKDOUTENRKRIEIDIFRSTANG..,....OFEPEPDQGIGGSAVOLYAKESKRLLVLK 100
B9R748_1 48 SLSIUPPSQTTRDWVNLKLIETLSS..PSVLSKR..YGTISHDEAESARKRIEIDHRFGVHN..,....RTHSDQALEILQLYSKESVRLLVLK 133

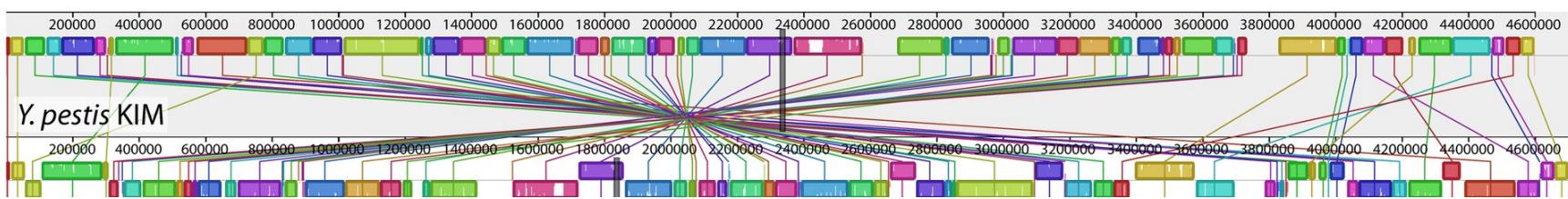
```



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

What are the mechanisms?

- Homologous recombination
- Intra-genomic recombination
- Horizontal gene transfer:
 - transformation
 - transduction
 - bacterial conjugation



From Biology to Maths

Rearrangement

> Deletion

Math illustration

$$1 \ 2 \ -3 \ \textcolor{red}{4} \ -5 \ 6 \rightarrow 1 \ 2 \ -3 \ -5 \ 6$$

> Insertion

$$1 \ 2 \ -3 \ 4 \ -5 \ 6 \rightarrow 1 \ 2 \ -3 \ 4 \ \textcolor{red}{7} \ -5 \ 6$$

> Translocation

$$1 \ \textcolor{red}{2} \ -3 \ 4 \ -5 \ 6 \rightarrow 1 \ -3 \ 4 \ -5 \ \textcolor{red}{2} \ 6$$

> Inversion

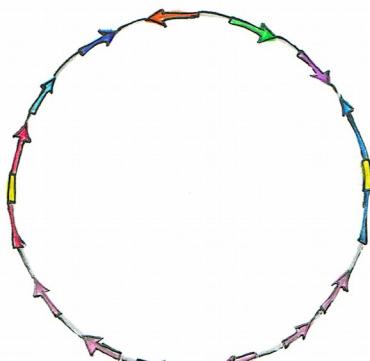
$$1 \ 2 \ -3 \ 4 \ -5 \ 6 \rightarrow 1 \ 2 \ \textcolor{red}{5} \ -4 \ 3 \ 6$$

> Duplication

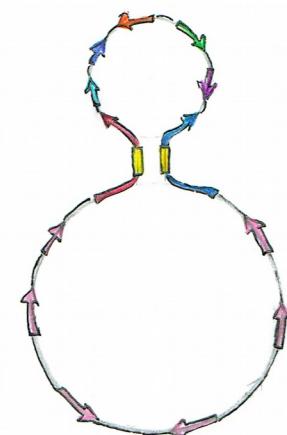
$$1 \ 2 \ -3 \ 4 \ -5 \ 6 \rightarrow 1 \ 2 \ \textcolor{red}{-3} \ -3 \ 4 \ -5 \ 6$$

Intra-genomic recombination

**Circular chromosome
with repeats**



**Intra-chromosome
recombination**

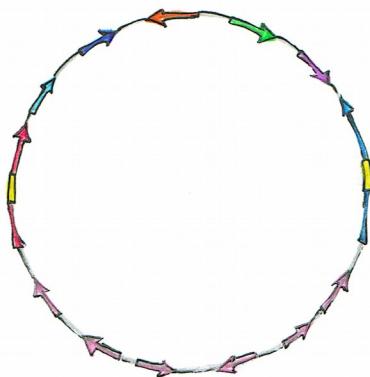


**Inversion of the segment
between repeats**

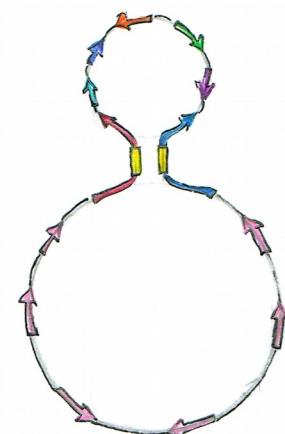


Intra-genomic recombination

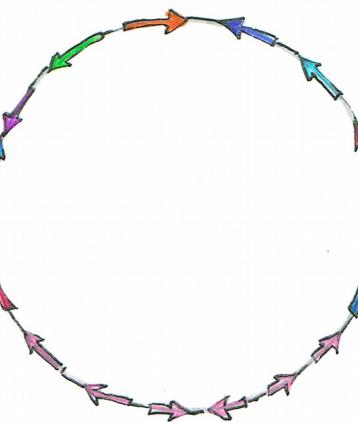
Circular chromosome with repeats



Intra-chromosome recombination



Inversion of the segment between repeats

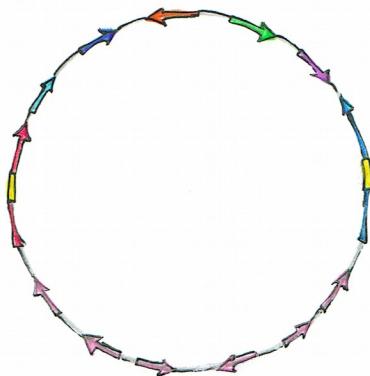


Genomic repeats:

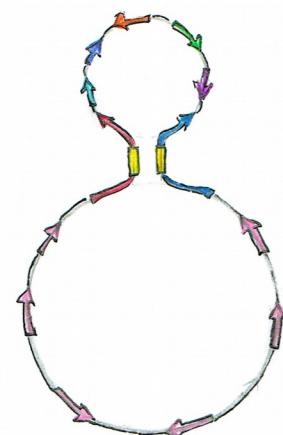
- mobile elements (transposase) – up to 100 copies per genomes
- rRNA gene operons – up to 15 copies per genomes
- gene paralogs

Intra-genomic recombination

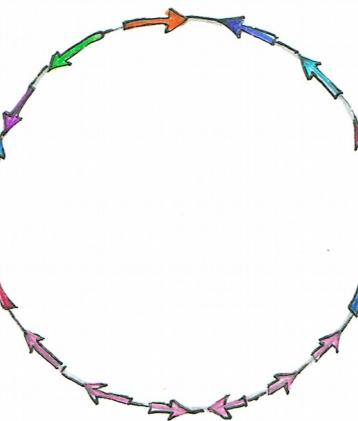
Circular chromosome with repeats



Intra-chromosome recombination



Inversion of the segment between repeats

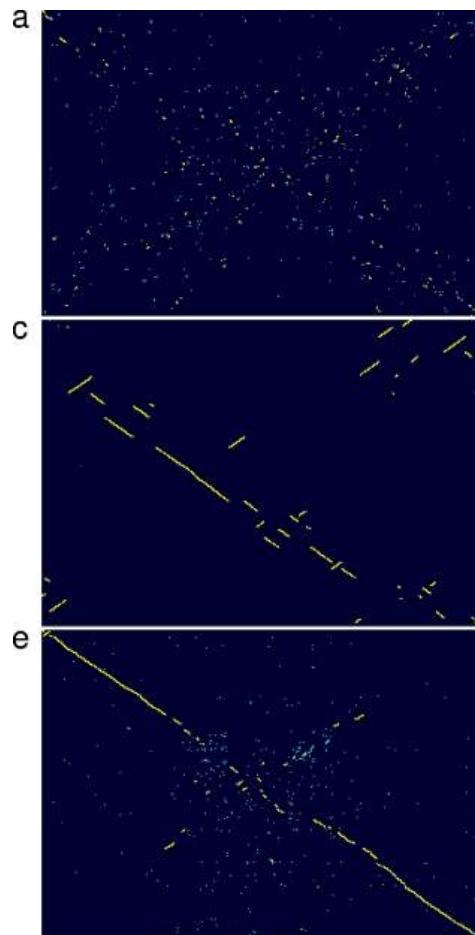


Genomic repeats:

- 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; *Chlamydophila caviae* GPIC and *Chlamydophila abortus* S26/3.

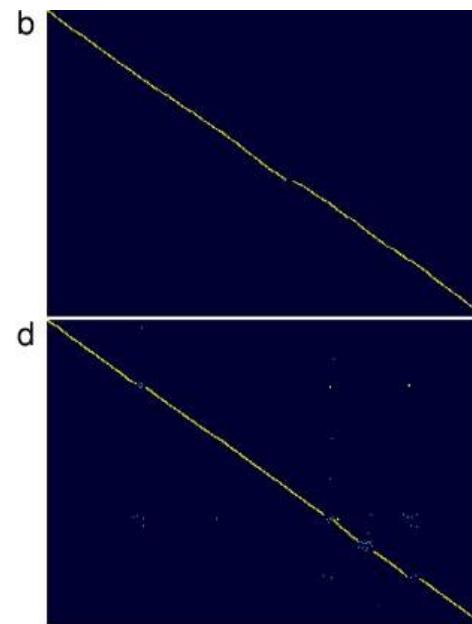
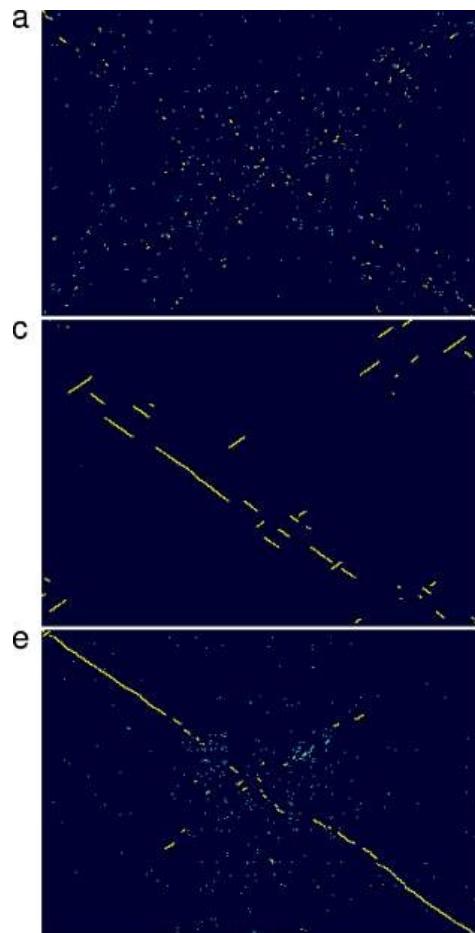
(c) Multiple inversions with limited transposition of individual genes; *Yersinia pestis* Antiqua and *Y. pestis* CO92.

(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; *Chlamydophila caviae* GPIC and *Chlamydophila abortus* S26/3.

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

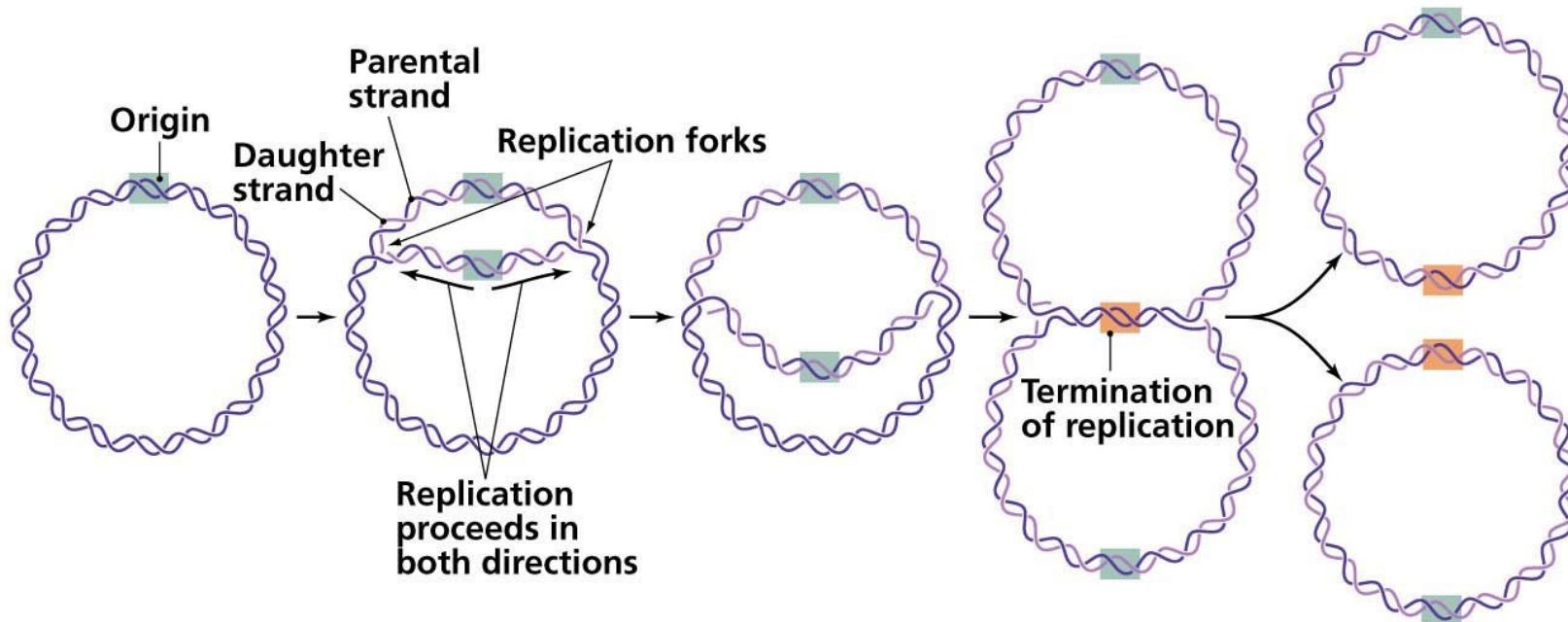
(d) No inversion; hot spots of transposition of individual genes; *P. marinus* AS9601 and *P. marinus* MIT 9215.

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

symmetric inversions
around the replication origin

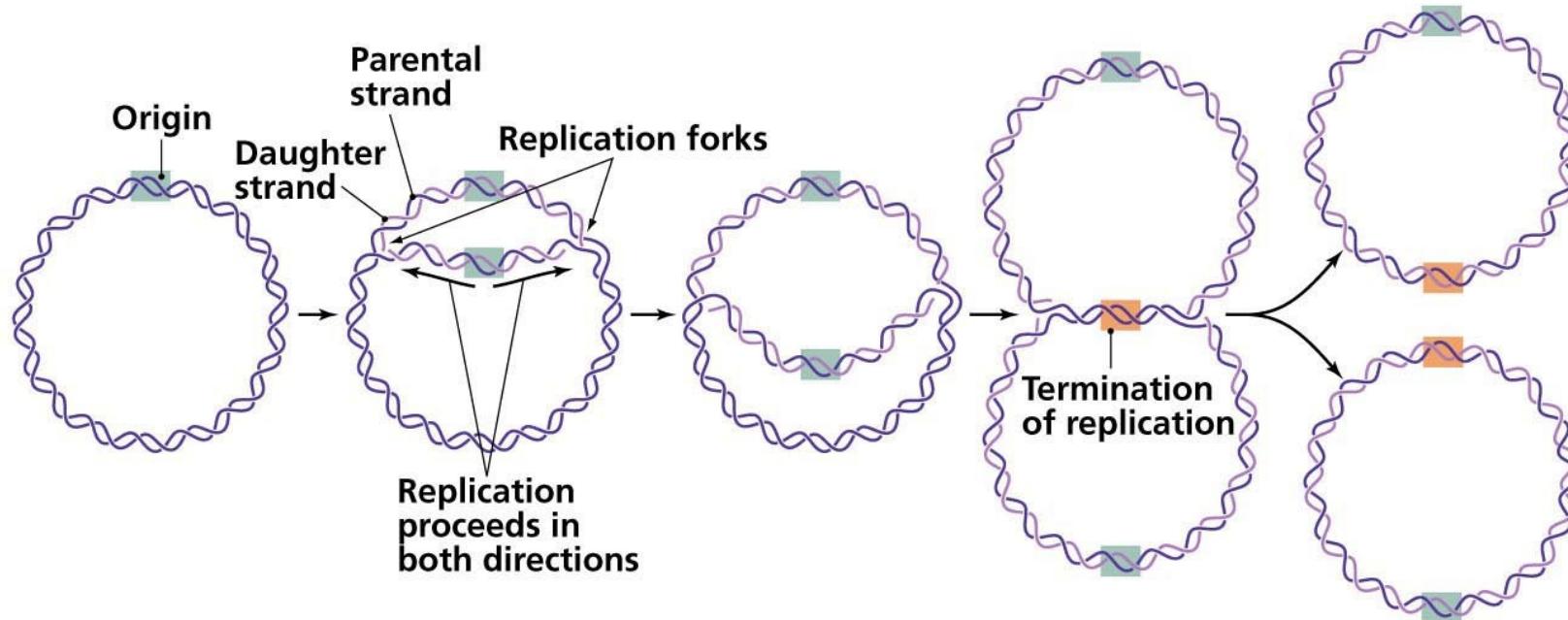
Novichkov PS, Wolf YI, Dubchak I, Koonin EV.
Trends in prokaryotic evolution revealed by
comparison of closely related bacterial and
archaeal genomes. *J Bacteriol.* 2009

Chromosome structure



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

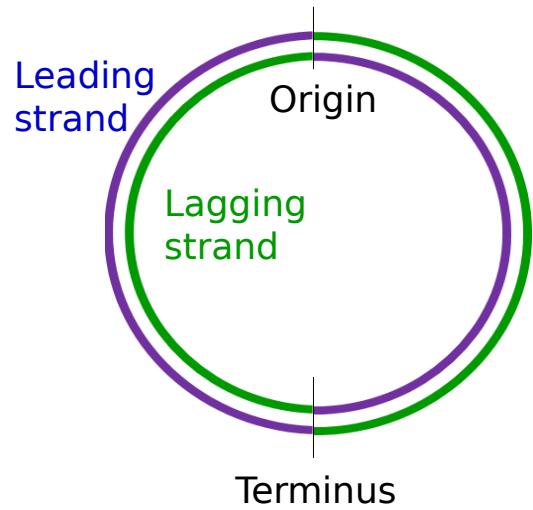
Chromosome structure



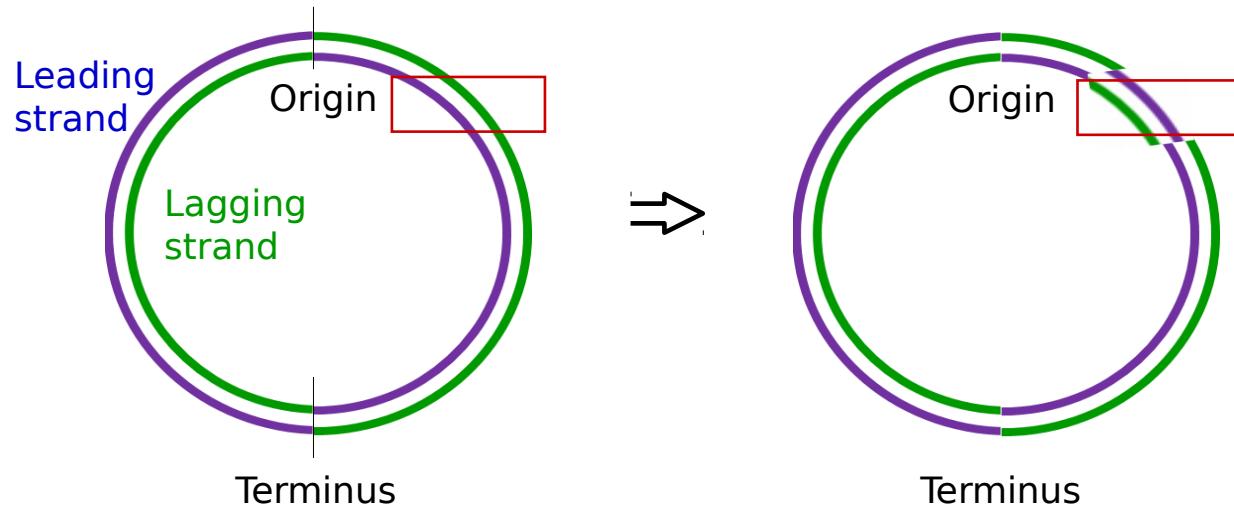
Selection on genome rearrangements / gene transfer:

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)

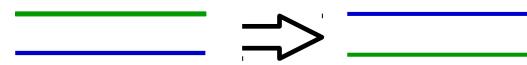
If the inversion occurred



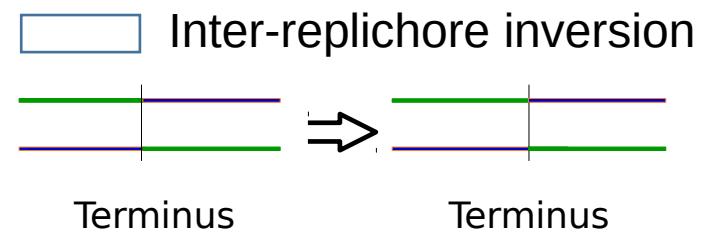
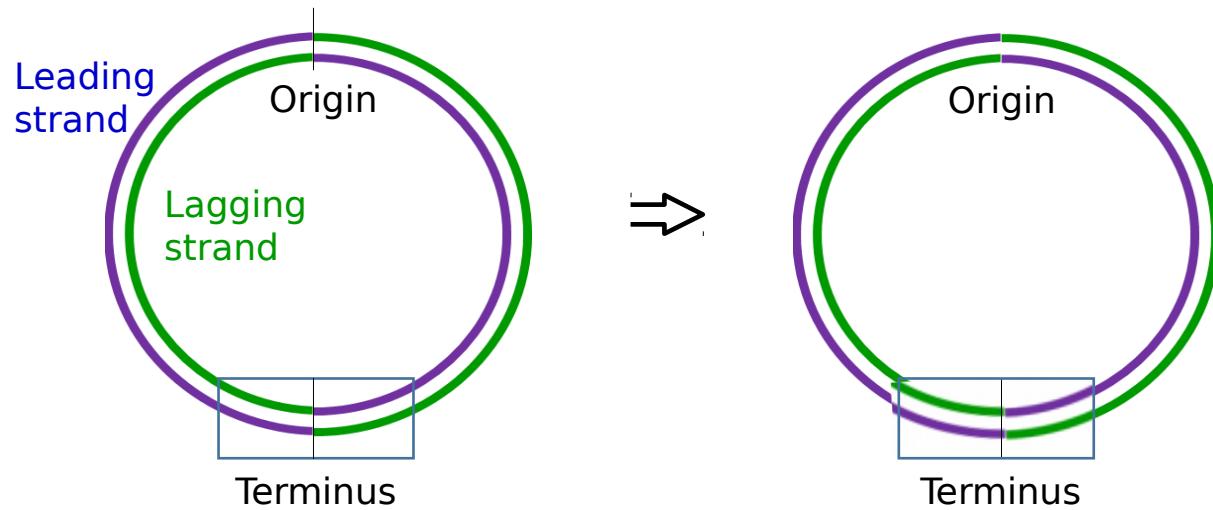
If the inversion occurred



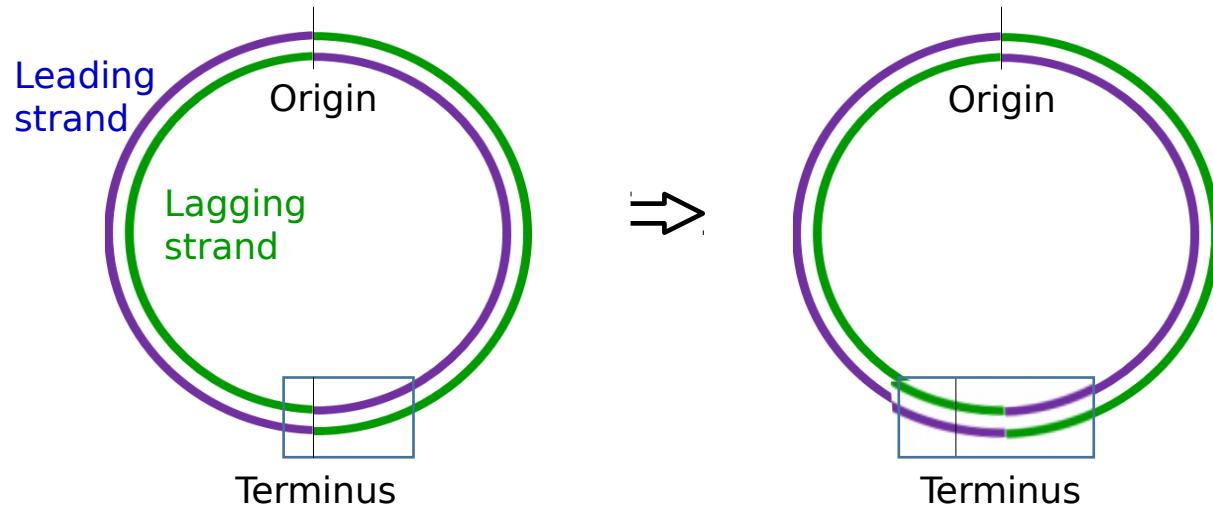
▀ Intra-replichore inversion



If the inversion occurred



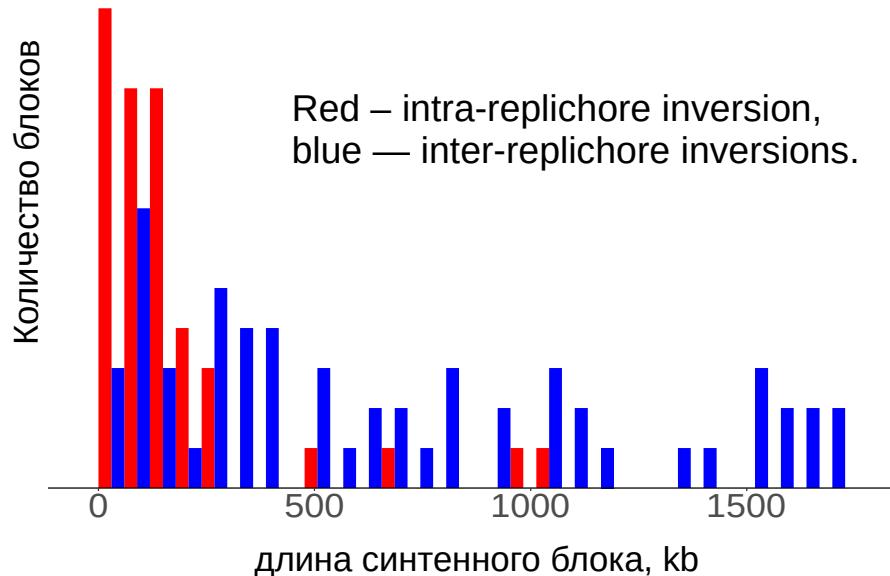
If the inversion occurred



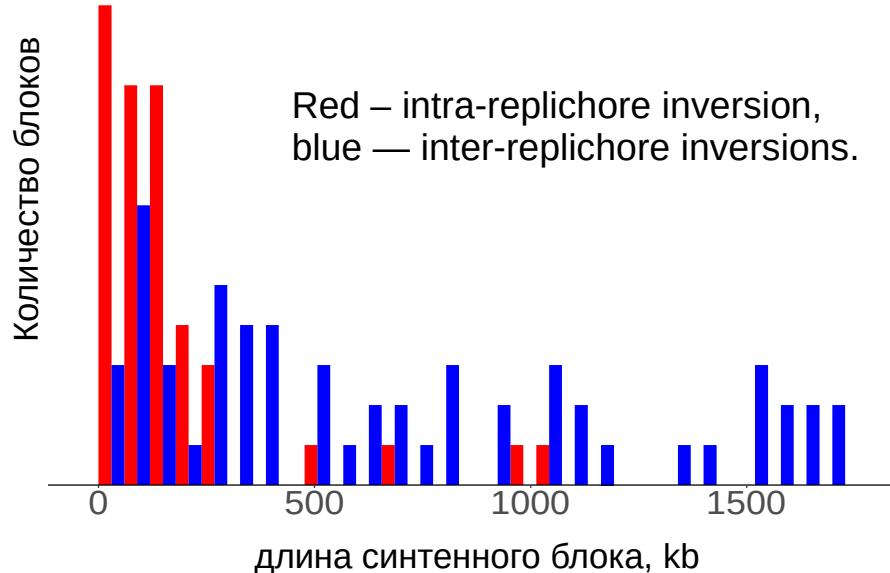
■ Inter-replichore inversion



What we see on the data



What we see on the data



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

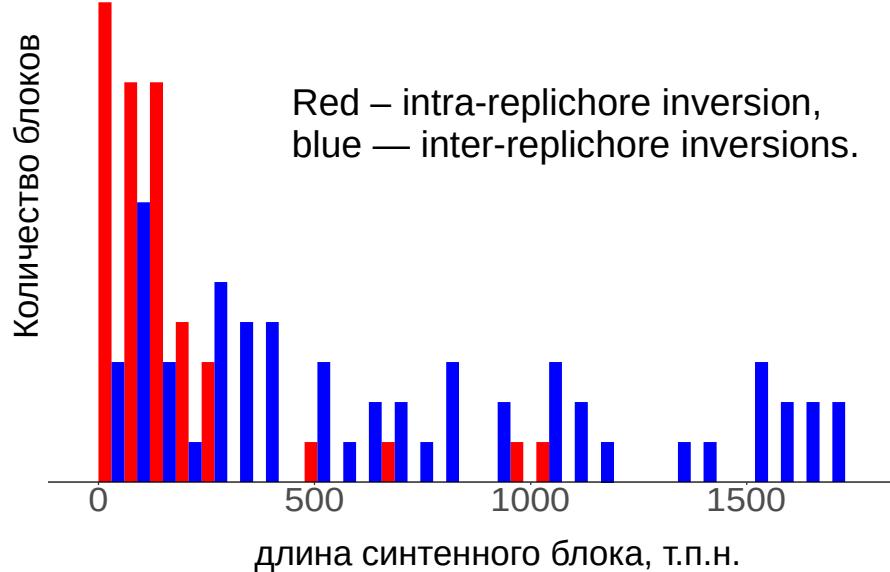
$$p(l) = m/L,$$

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:

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

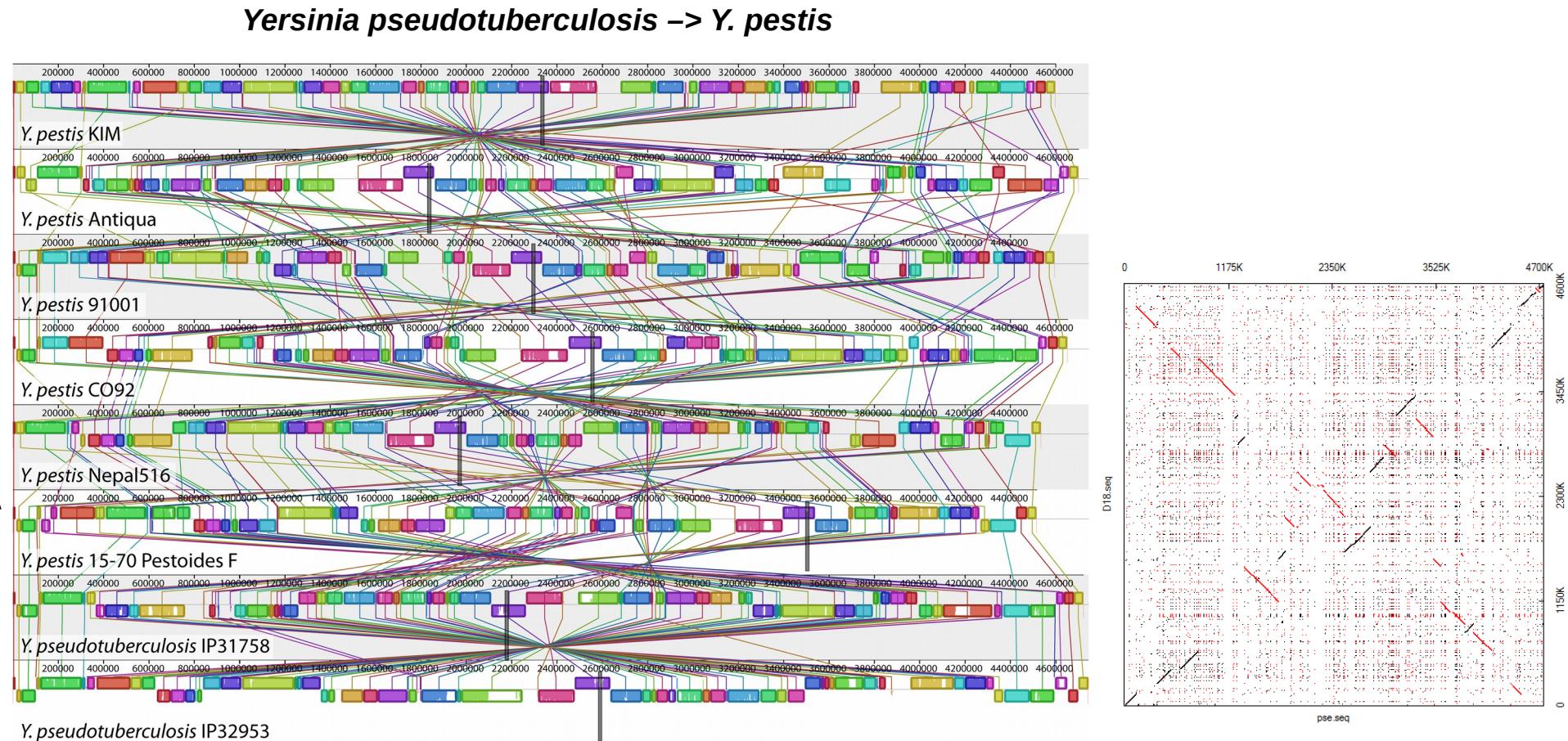
где L – размер реплихоры.

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

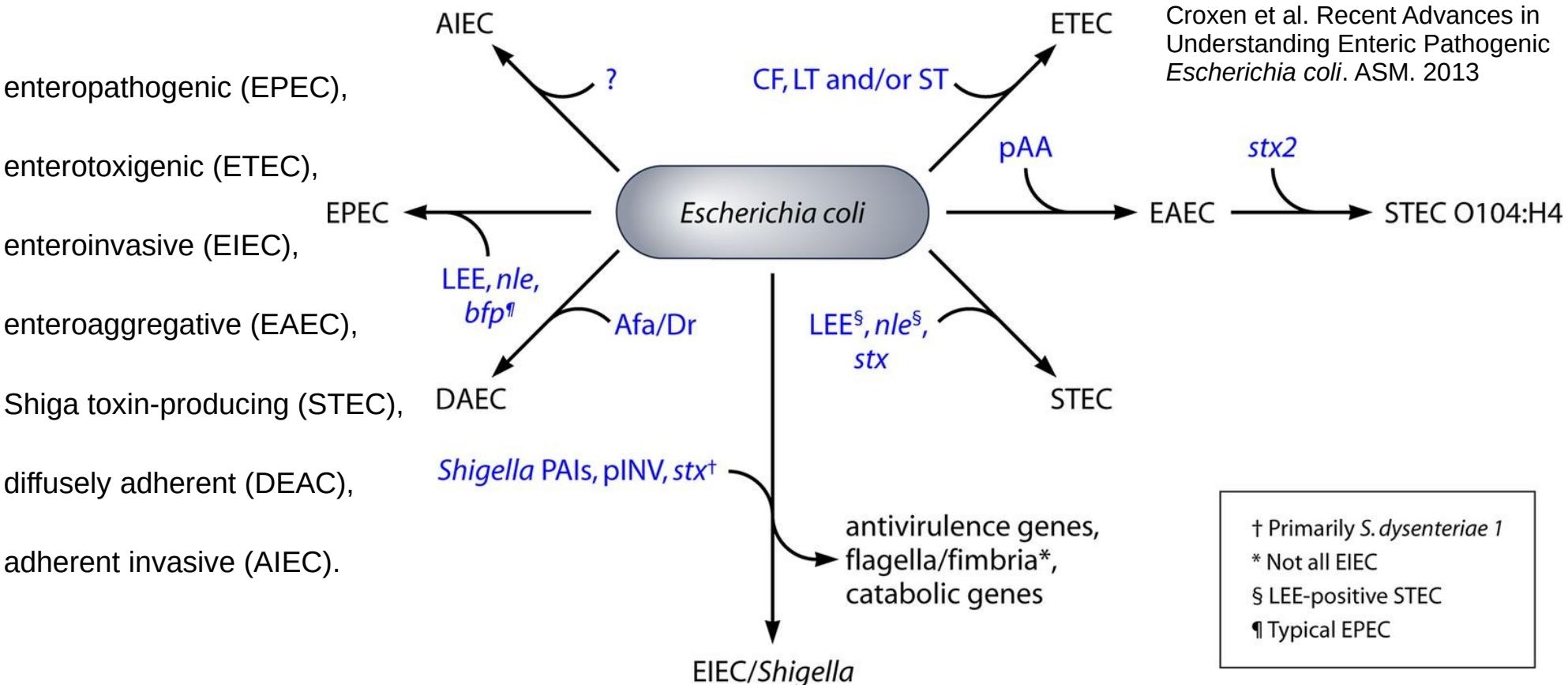
ISs expansion drives rearrangements



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

Pathogenic E.coli

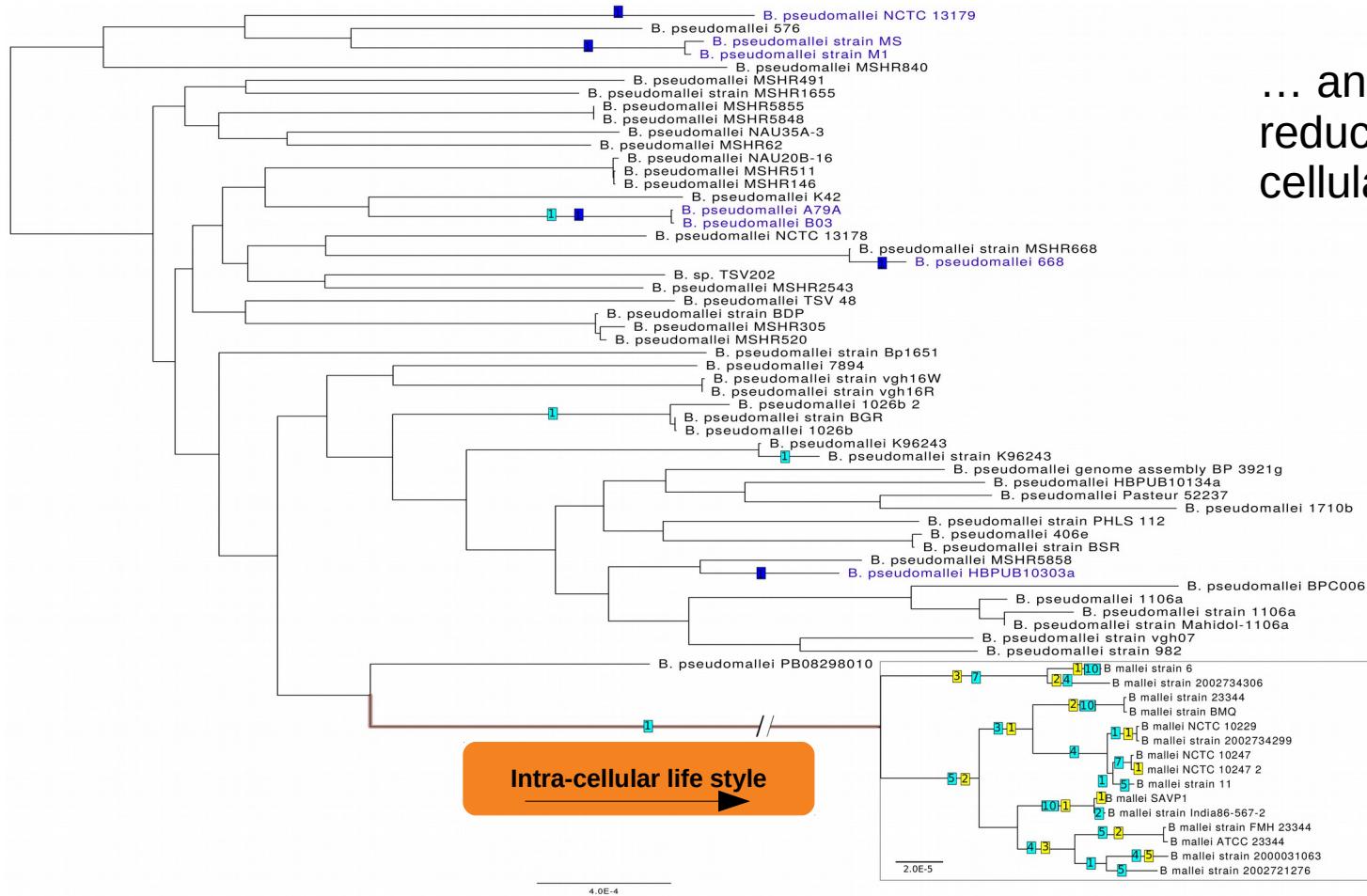
enteropathogenic (EPEC),
enterotoxigenic (ETEC),
enteroinvasive (EIEC),
enteroaggregative (EAEC),
Shiga toxin-producing (STEC),
diffusely adherent (DEAC),
adherent invasive (AIEC).



[†] Primarily *S. dysenteriae* 1
^{*} Not all EIEC
[§] LEE-positive STEC
[¶] Typical EPEC

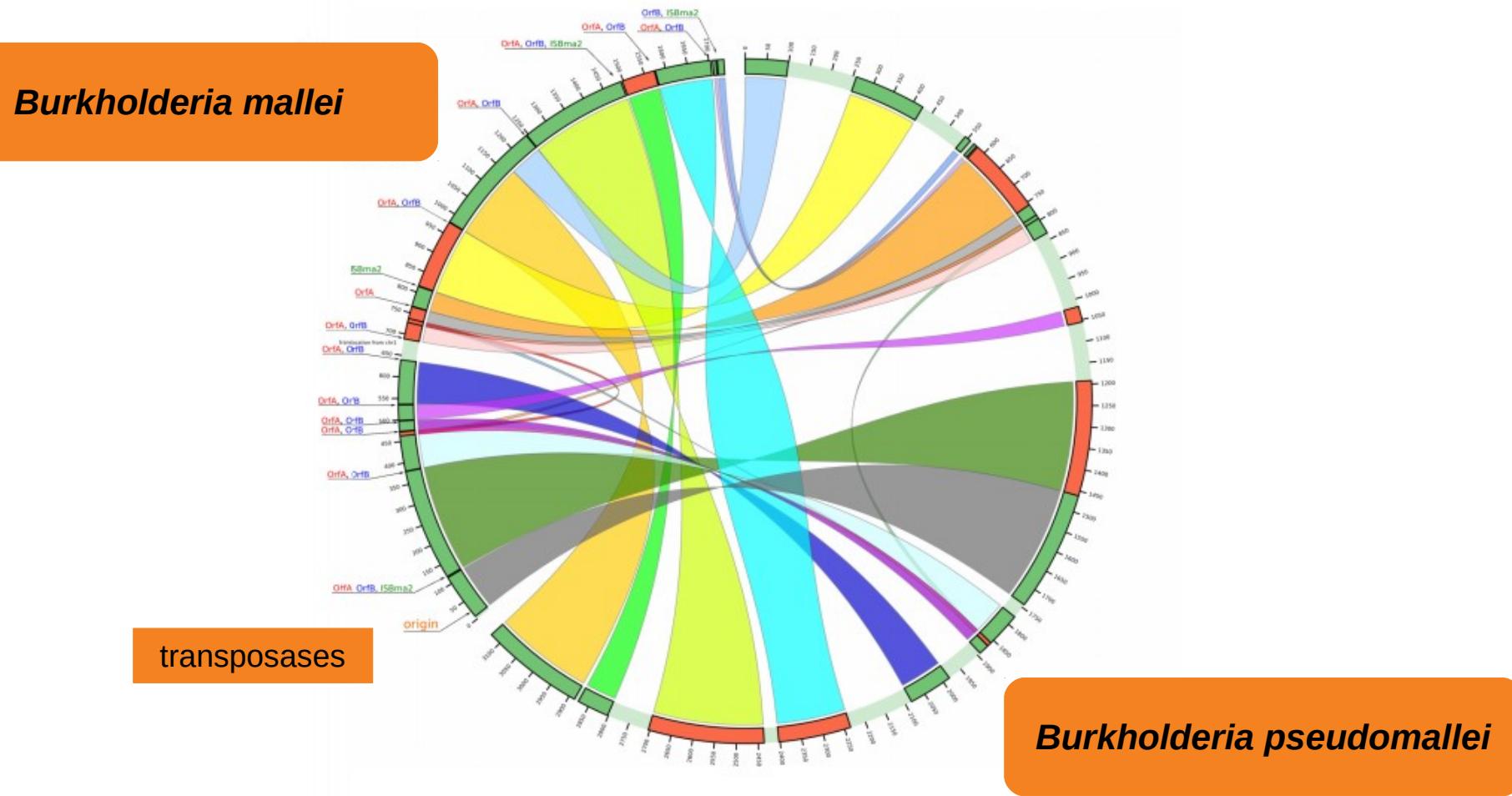
Shigella strains are distinguished from *E. coli* by their nonmotility, absence of decarboxylated lysine and several biochemical characteristics, such as inability to ferment lactose and mucate.

ISs expansion drives rearrangements

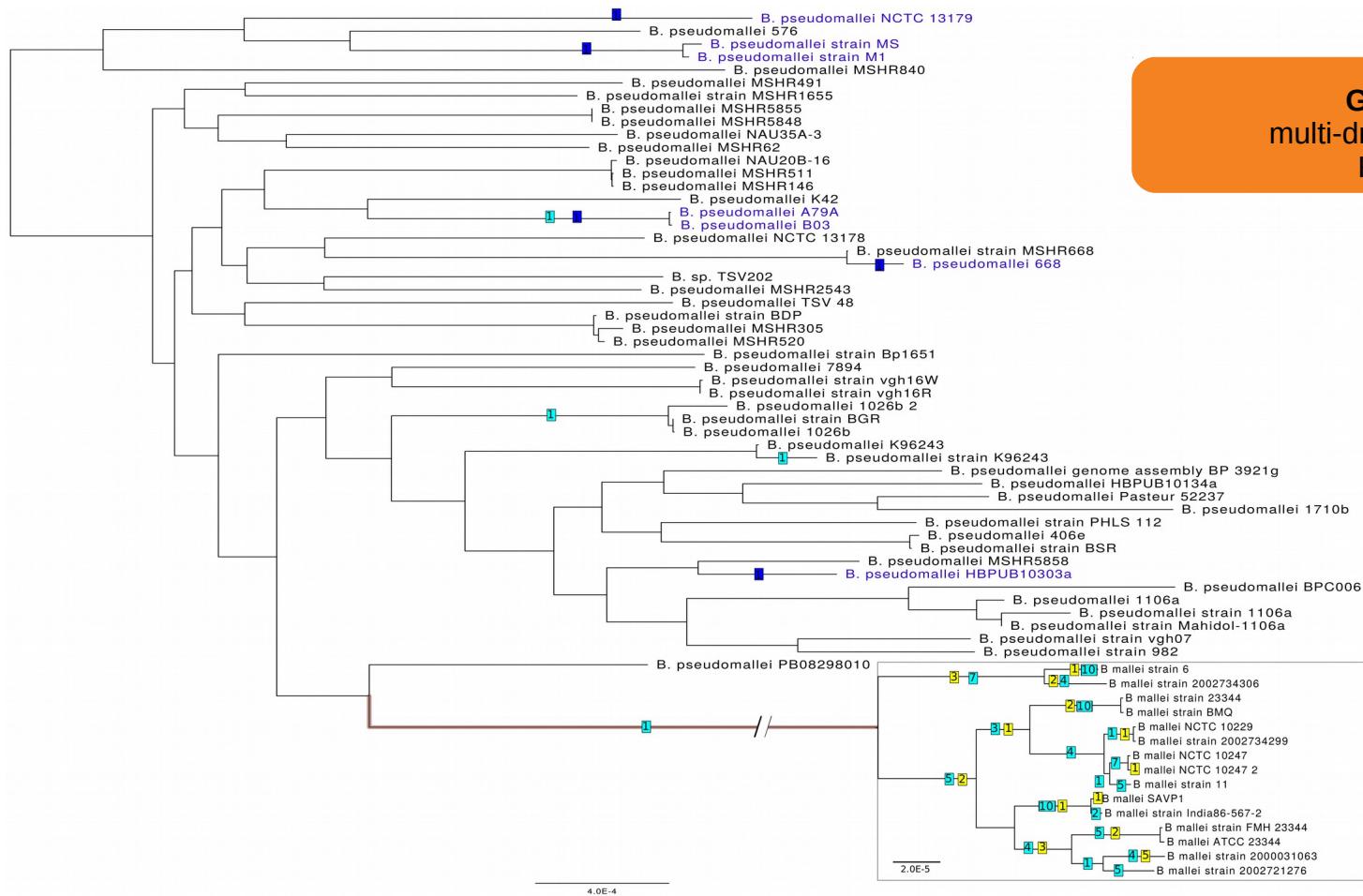


... and results in genome reduction in young intra-cellular pathogens

Genome reduction by deletions



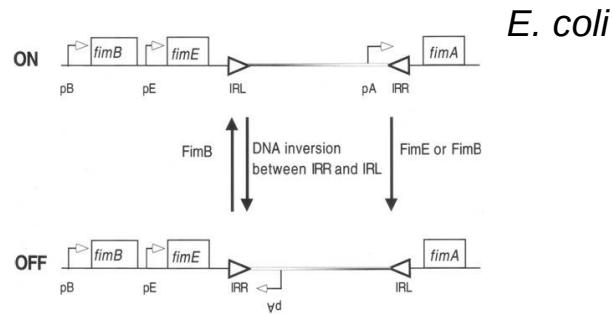
Parallel inversions – genomic switch?



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

Site-specific micro-inversions

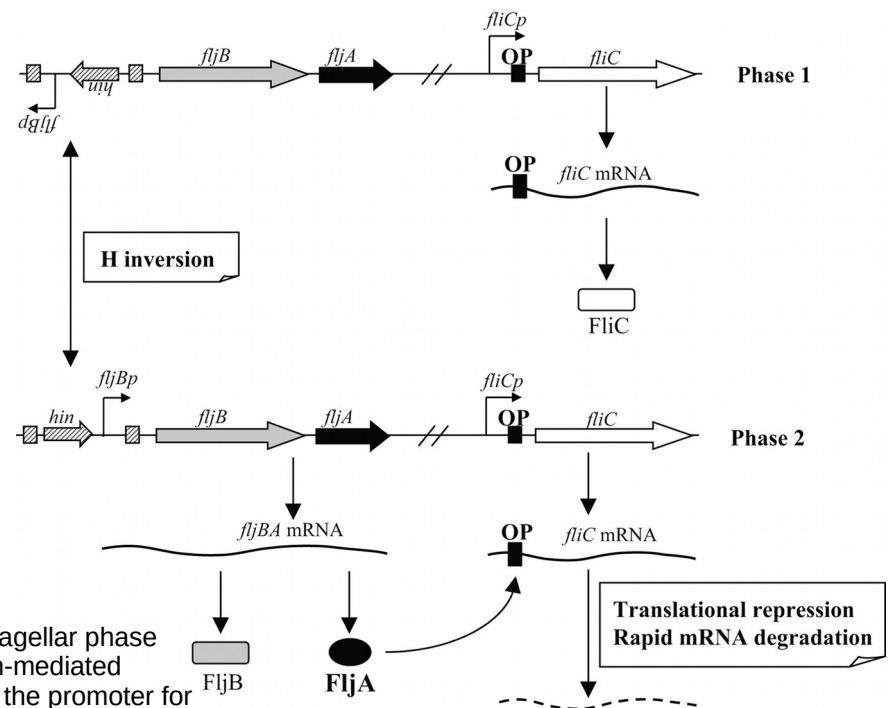
Phase variation – Site specific inversion



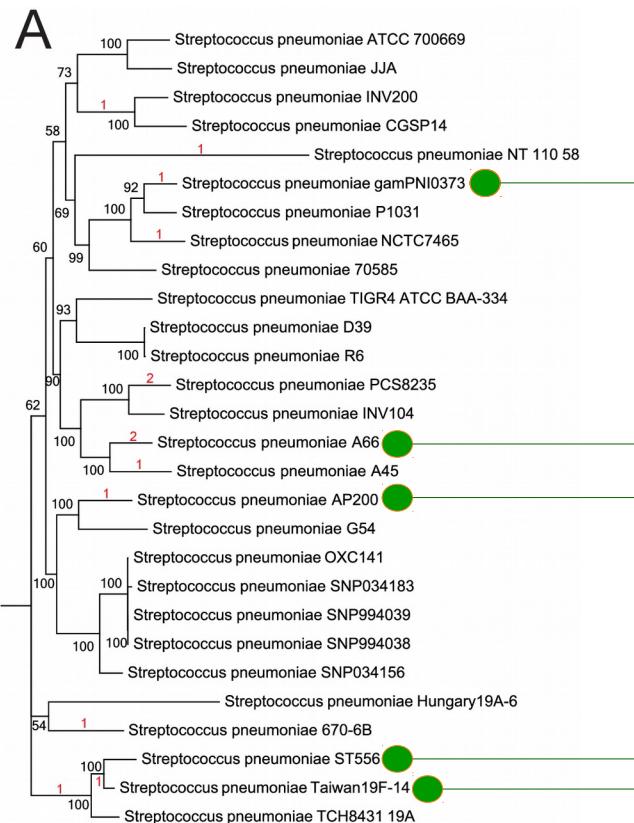
Phase variation – site specific inversion. pB, pE and pA are promoters for the genes *fimB*, *fimE* and *fimA* respectively. IRL and IRR are inverted repeats. FimB and FimE are recombinases that bind to the open triangles IRL an IRR. The result is an inversion of the DNA sequence (shaded bar) that turns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

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.

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



Parallel inversions – genomic switch?

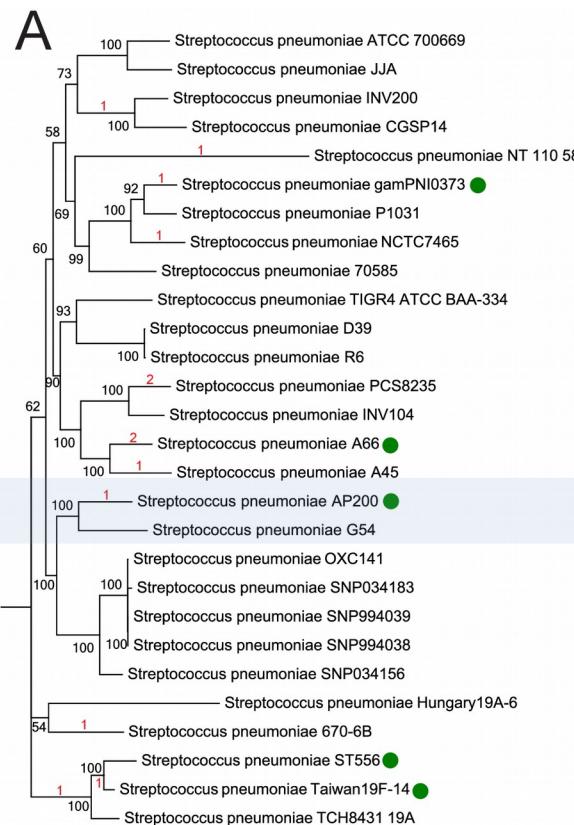


The same inverted sequence length 15kB
Breakpoints are formed by genes encoding
proteins PhtB and PhtD

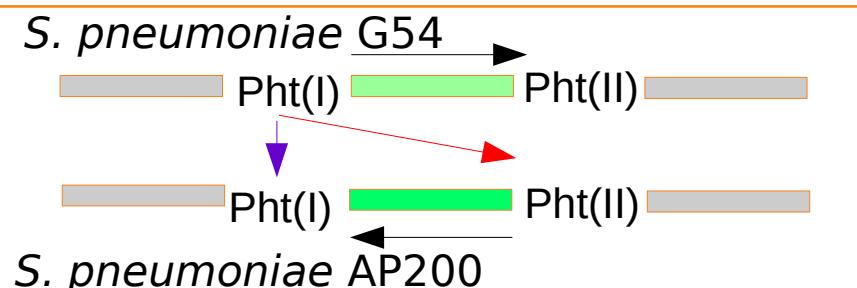
Белки гистидиновой триады (Pht)

- белки внешней мембрани, антигены
- характеризуются наличием мотива
HxxHxH

Parallel inversions – genomic switch?



B *Nucleotide alignments*

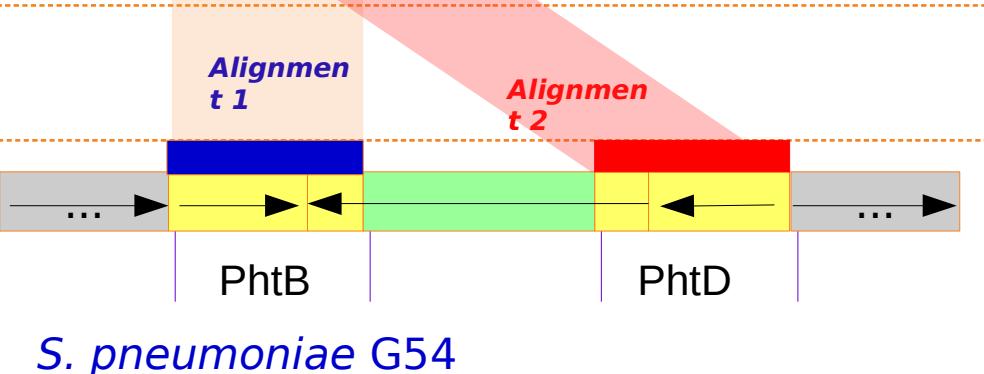
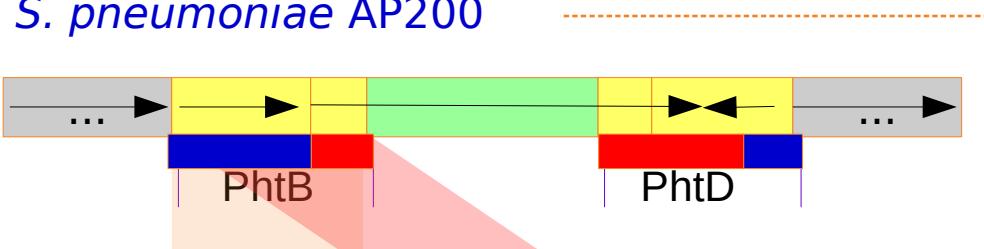


- The same inverted sequence length 15kB
- Breakpoints are formed by genes encoding proteins PhtB and PhtD

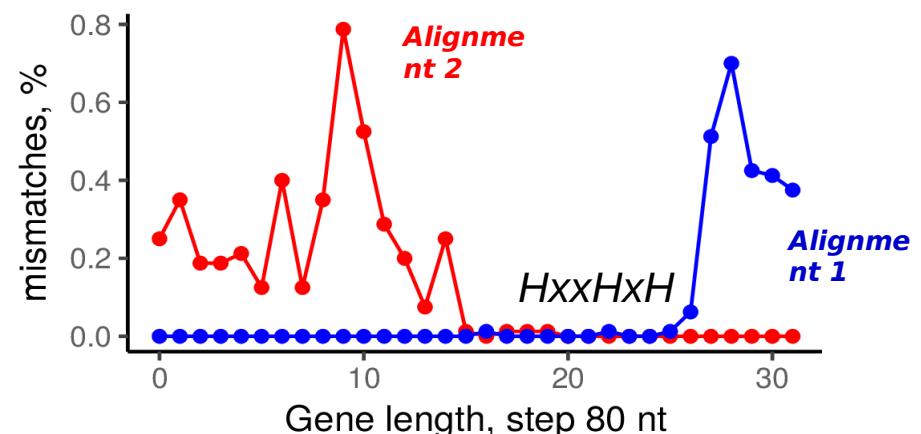
Parallel inversions – genomic switch?

Phts' alignment

S. pneumoniae AP200



Mismatches



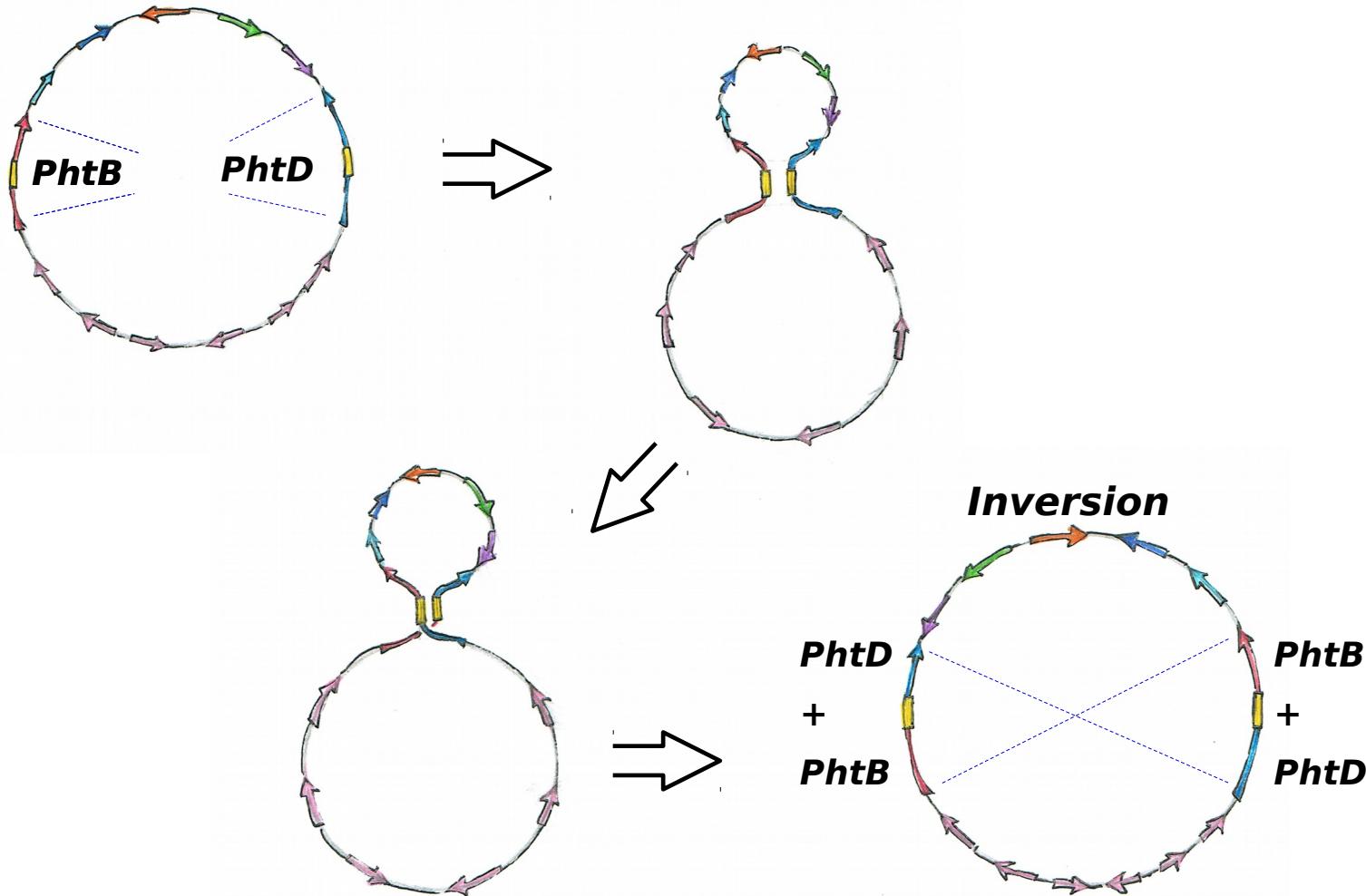
WP_0018462
WP_0018475

1 MKINKKYVAGSVAVLALSVC SYELGRYQAGQDKKESNRVAYIDGDQAG
1 MKINKKYL AGSVATLVLSVCAYE LGLHQA-QTVKENNRVSYIDGKQAT

cons

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

Parallel inversions – genomic switch?



Parallel inversions – genomic switch?

Биоинформационическое предсказание

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

Pavel V. Shelyakin [†]✉  , Olga O. Bochkareva [†], Anna A. Karan and Mikhail S. Gelfand

[†]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!!!



Биоинформационическое предсказание

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

Pavel V. Shelyakin [†]✉ [✉], Olga O. Bochkareva [†], Anna A. Karan and Mikhail S. Gelfand

[†]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!!!

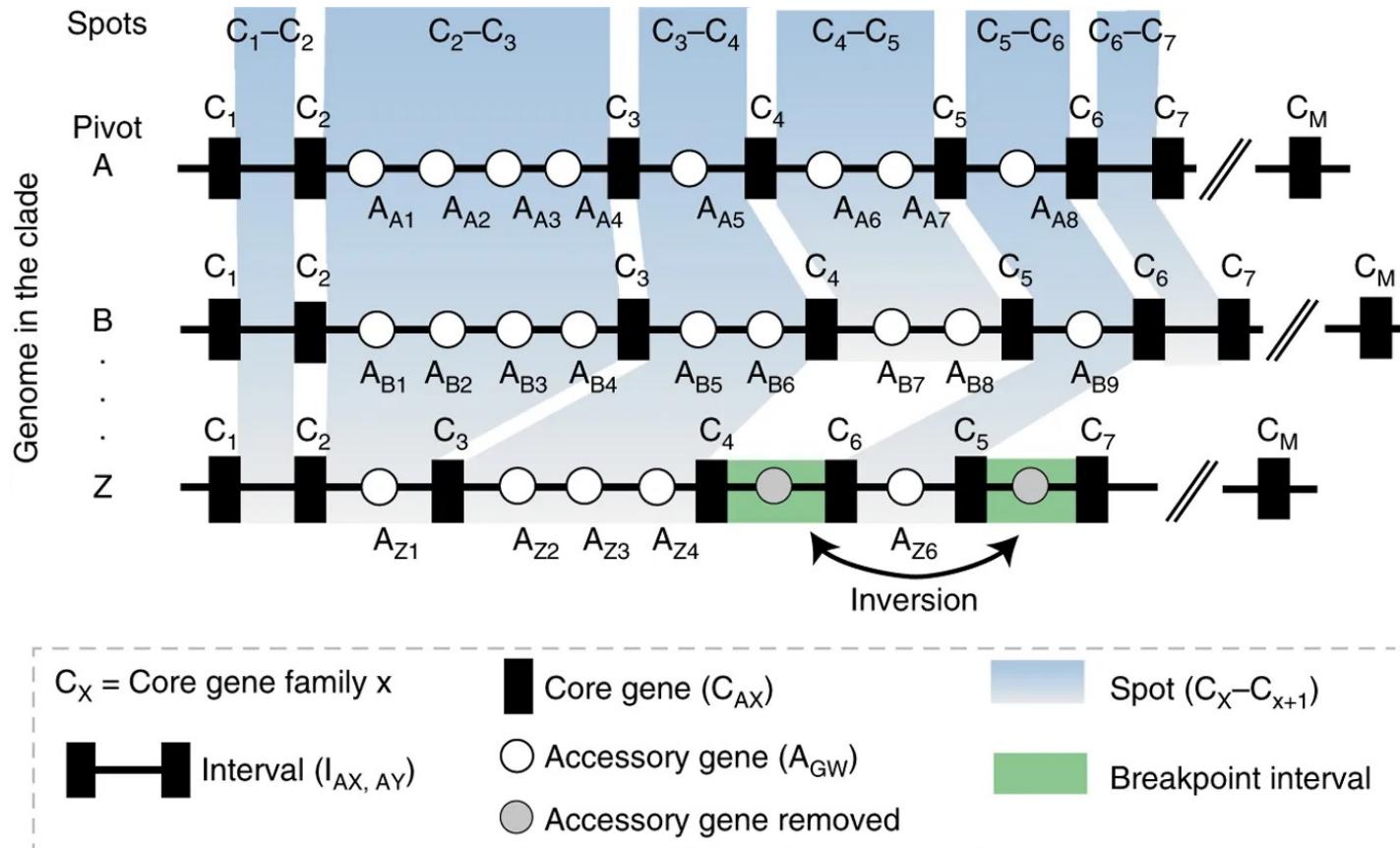


These two out of four pneumococcal histidine triad (Pht) proteins **are considered to be good vaccine candidates.**

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

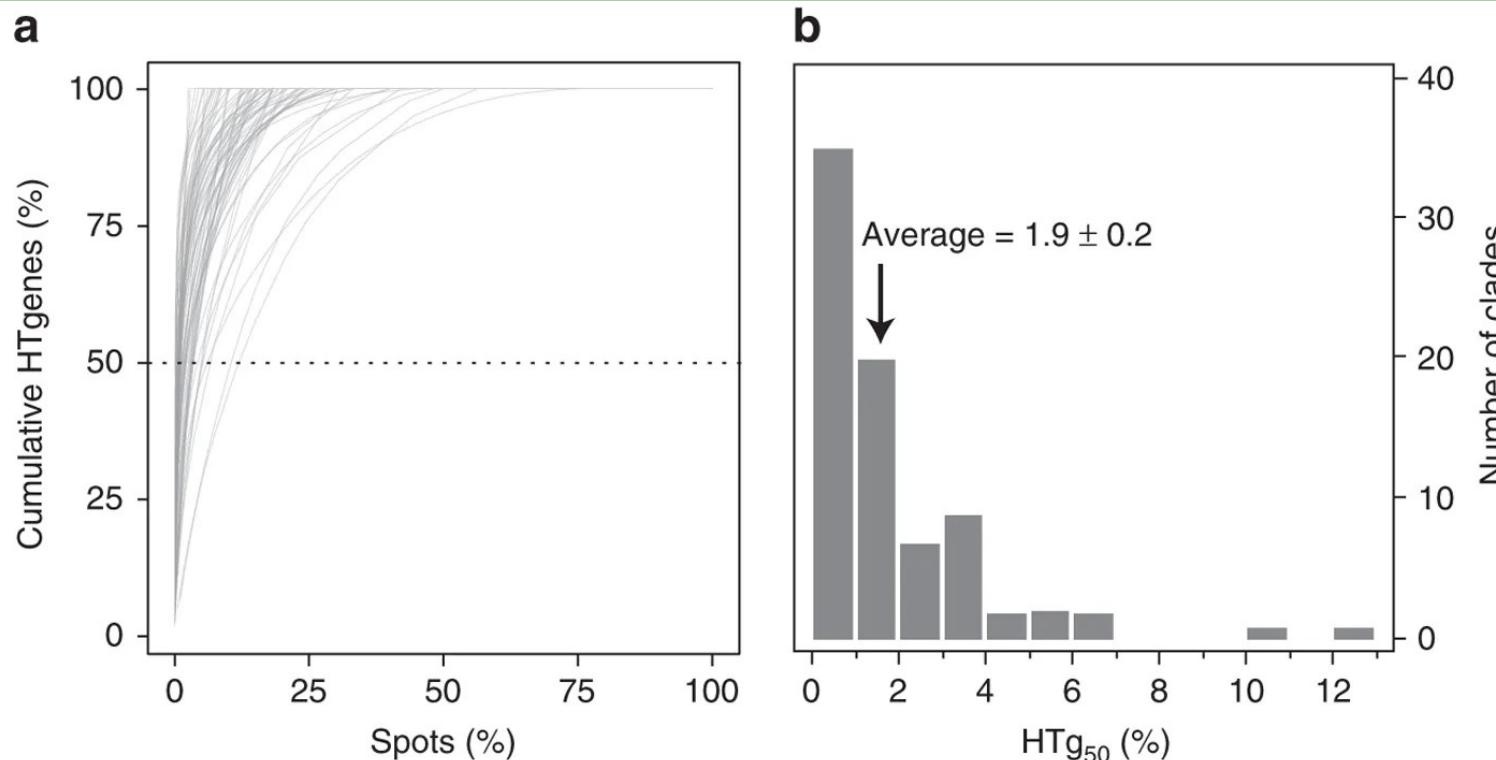
Yun et al.* (PLoS One. 2015) analyzed the diversity of phtD alleles from 172 clinical isolates and concluded that **the sequence variation was minimal.**

Evolution of gene content



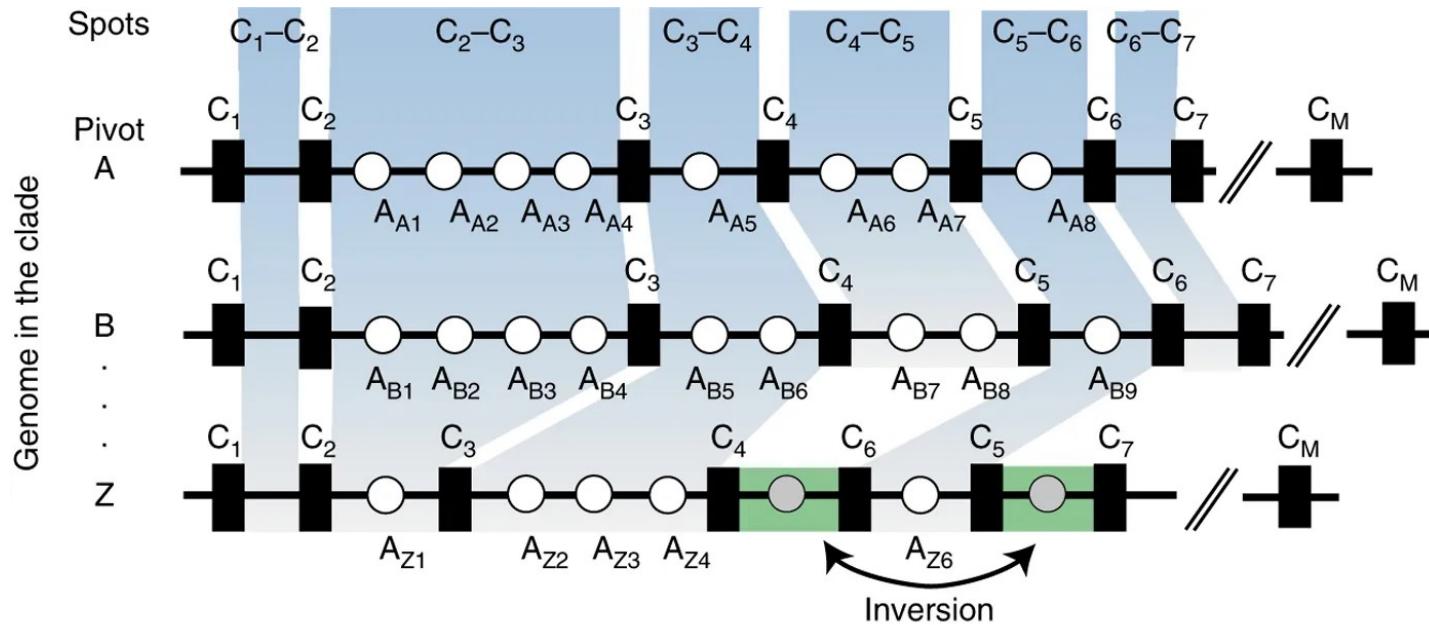
Oliveira, P.H., Touchon, M., Cury, J. et al. The chromosomal organization of horizontal gene transfer in bacteria. *Nat Commun.* 2017

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.”

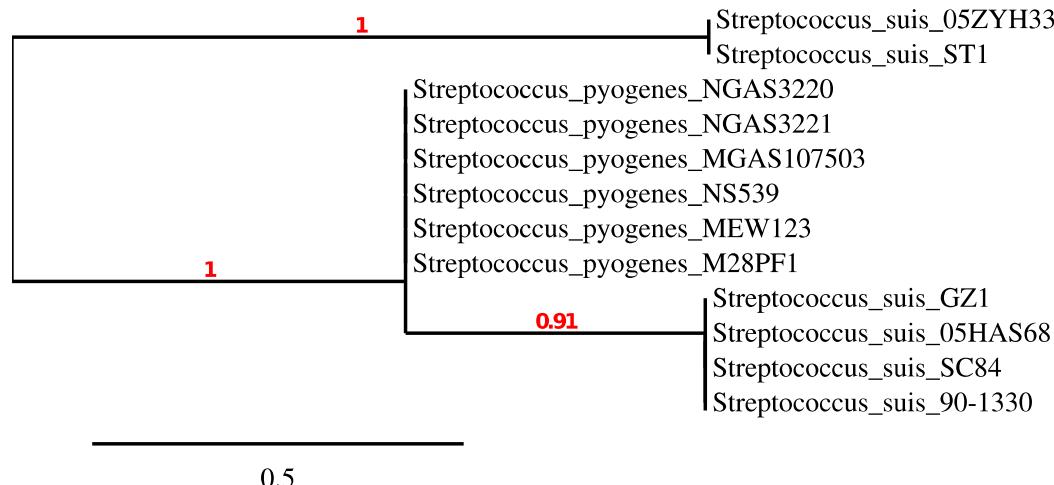
Evolution of gene content



Orthologous groups in different breakpoint intervals are
HT candidates

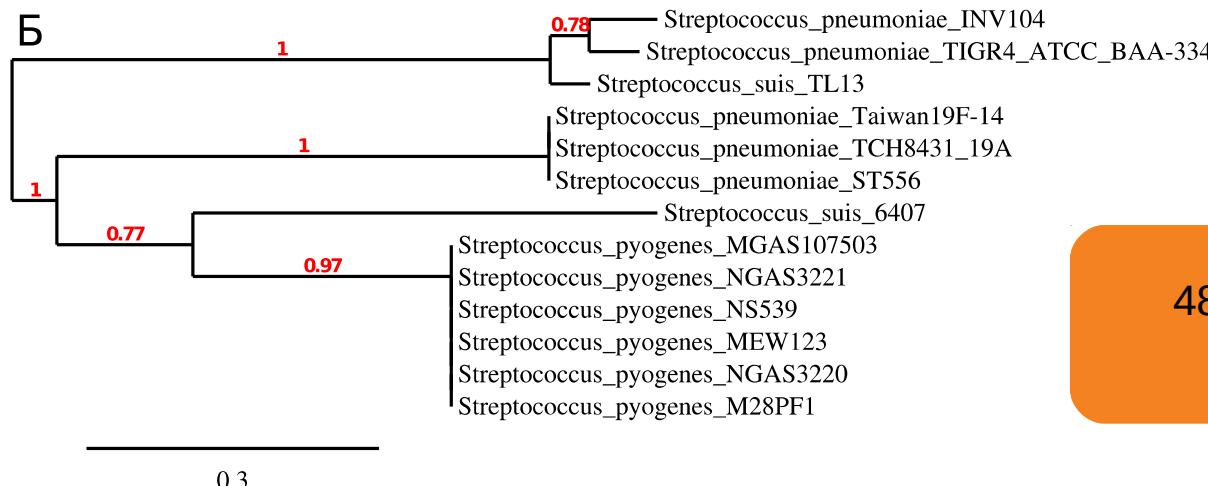
Evolution of gene content

A



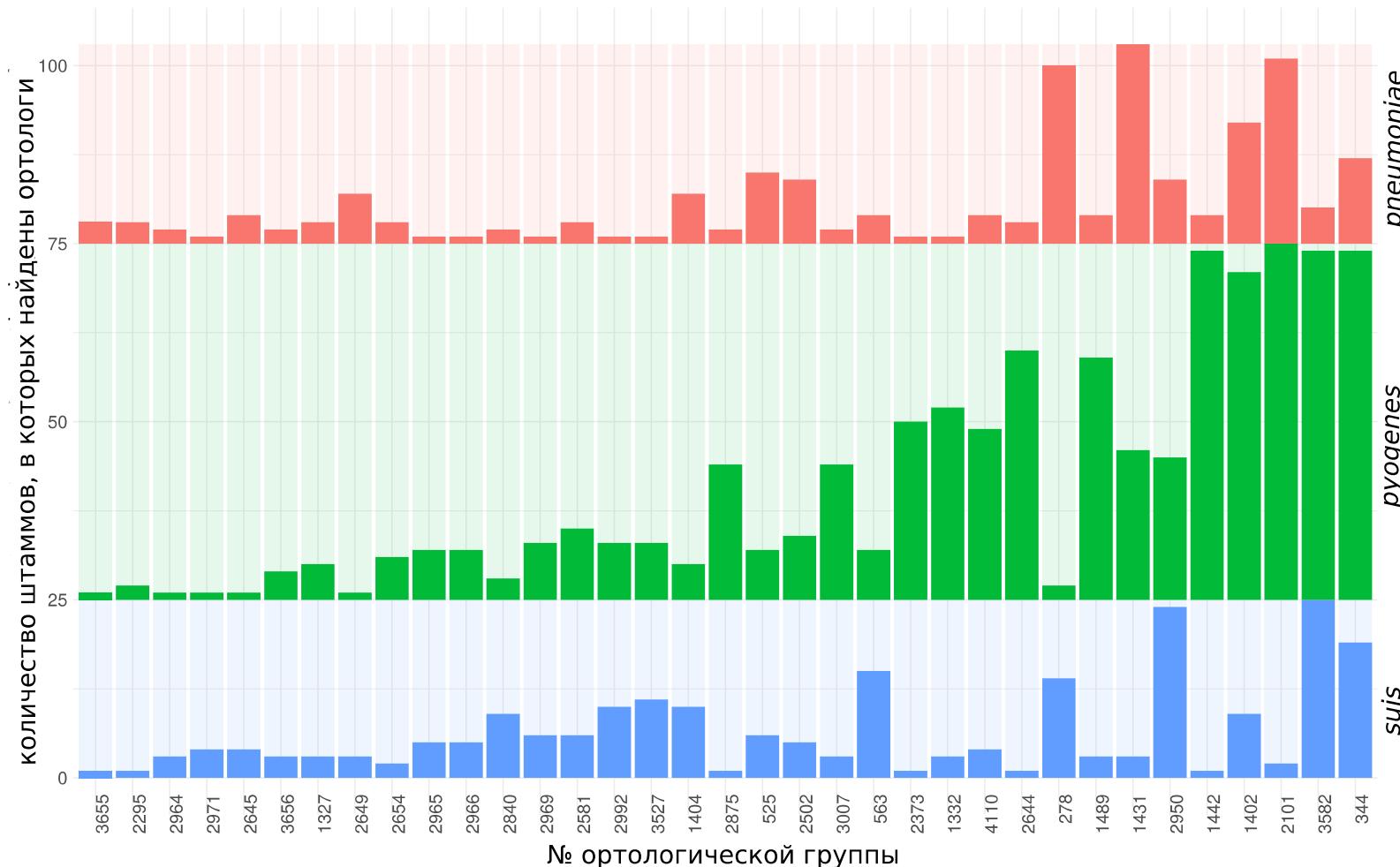
176 orthologous groups

B



48 of them have inconsistent phylogenetic trees

Evolution of gene content



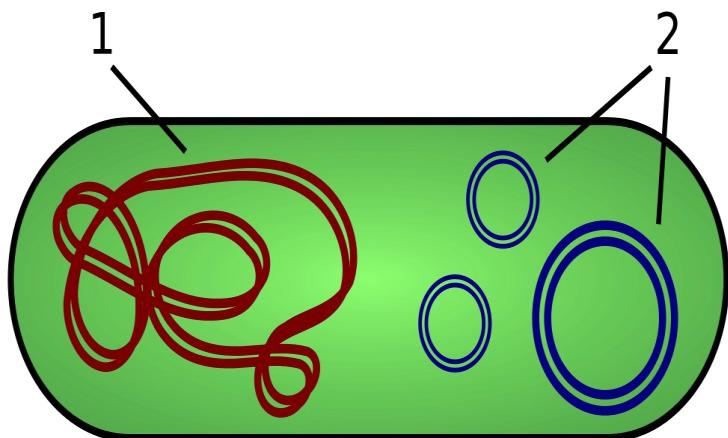
Bacterial genome structure

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

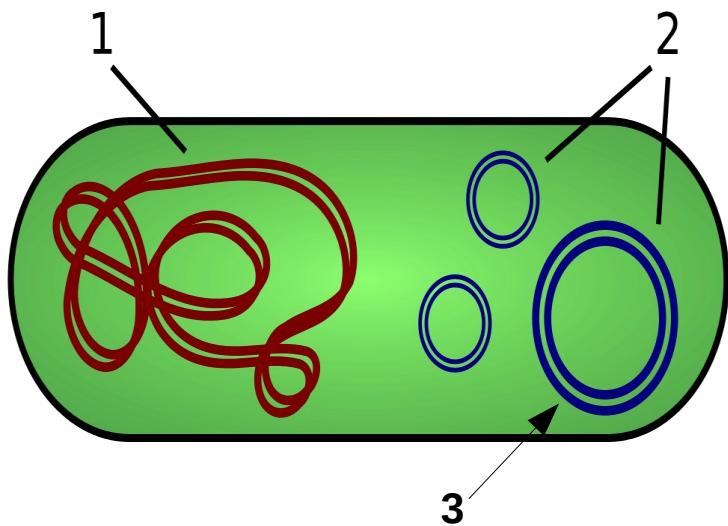
- circular, double-stranded DNA molecules
- 15 kb – 200 kb
- replicate independently
- carry genes that benefit the survival of the organism.



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

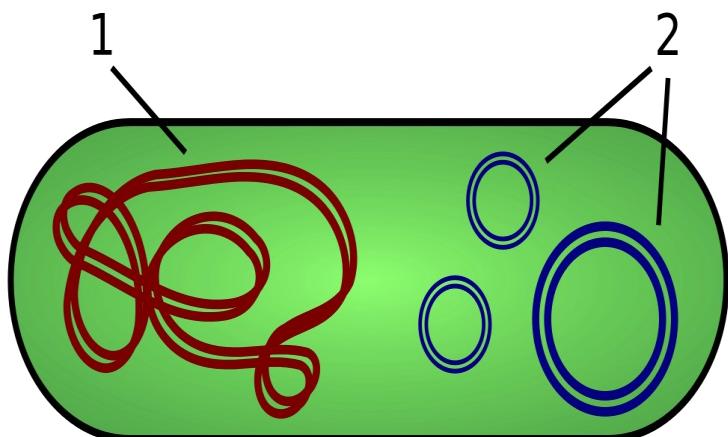
- circular, double-stranded DNA molecule
- 300 Mb – 3 Mb

-
- 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.

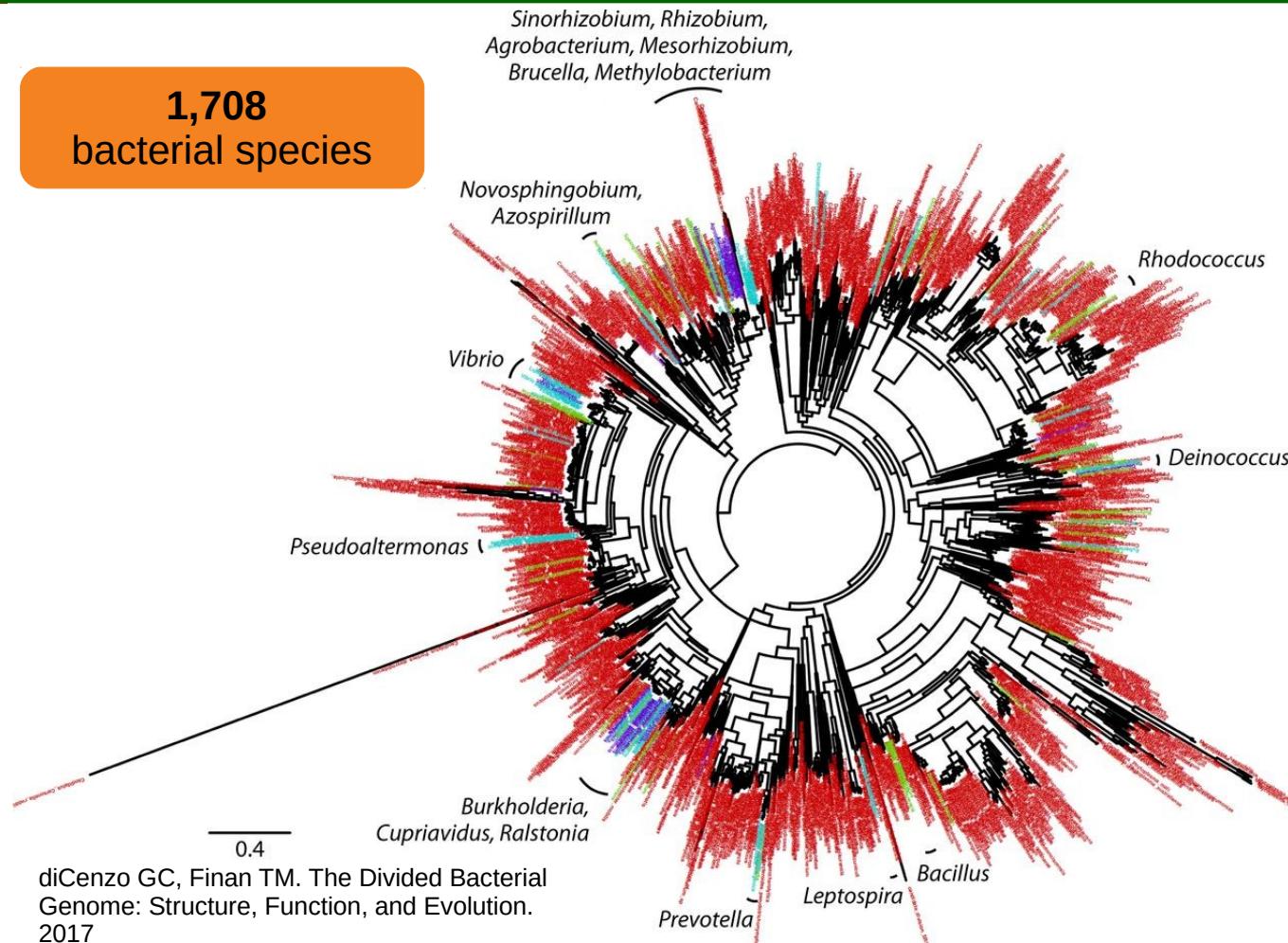
3. Other secondary replicons

- 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

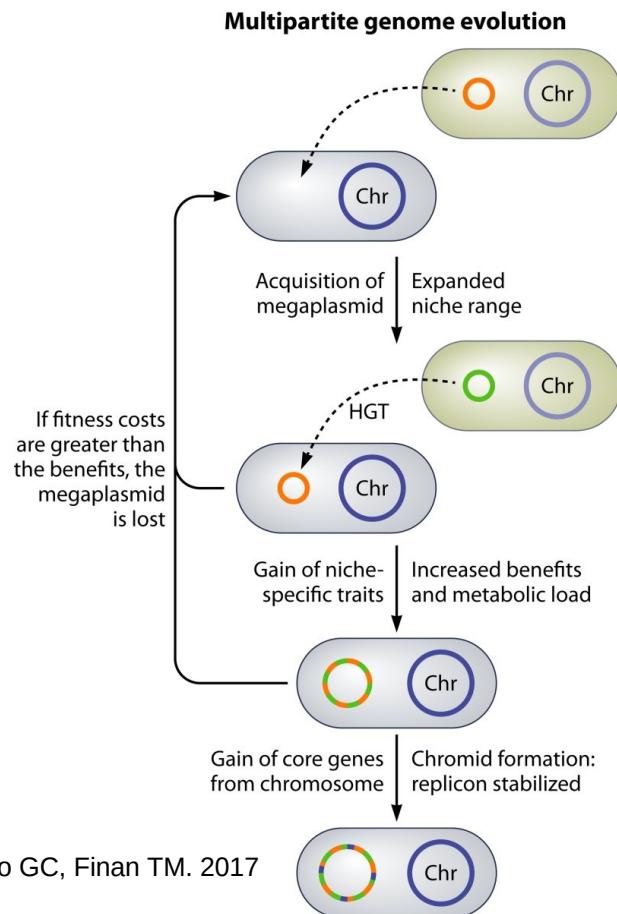
green for species with a megaplasmid(s) but no chromid

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

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

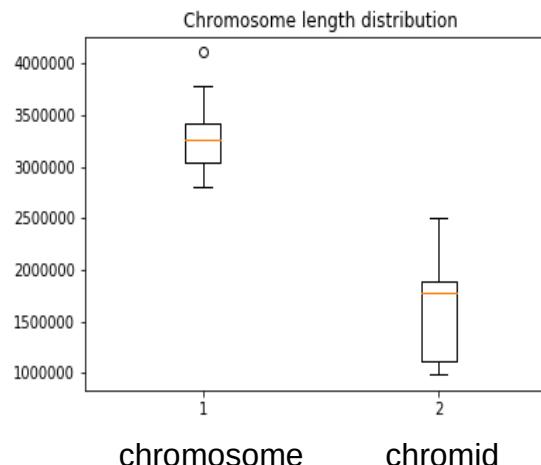
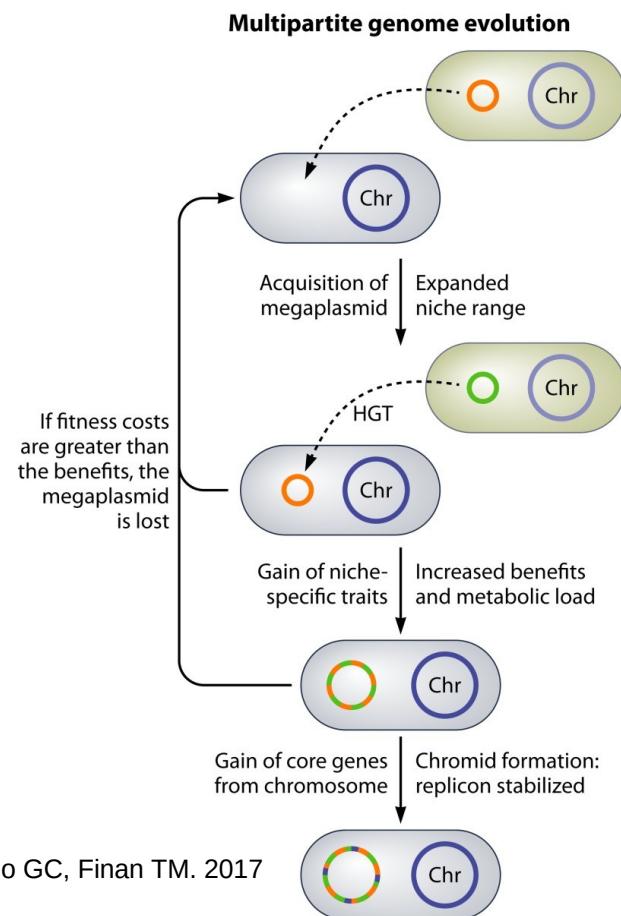
10% of bacterial genomes are split between two or more large DNA fragments

Chromid formation hypothesis

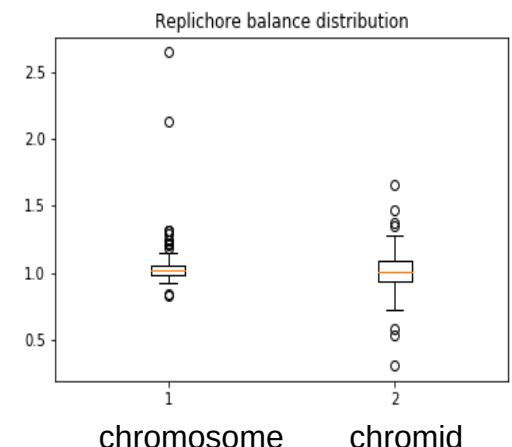
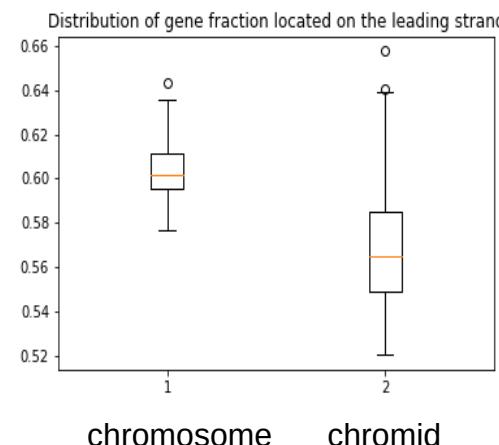
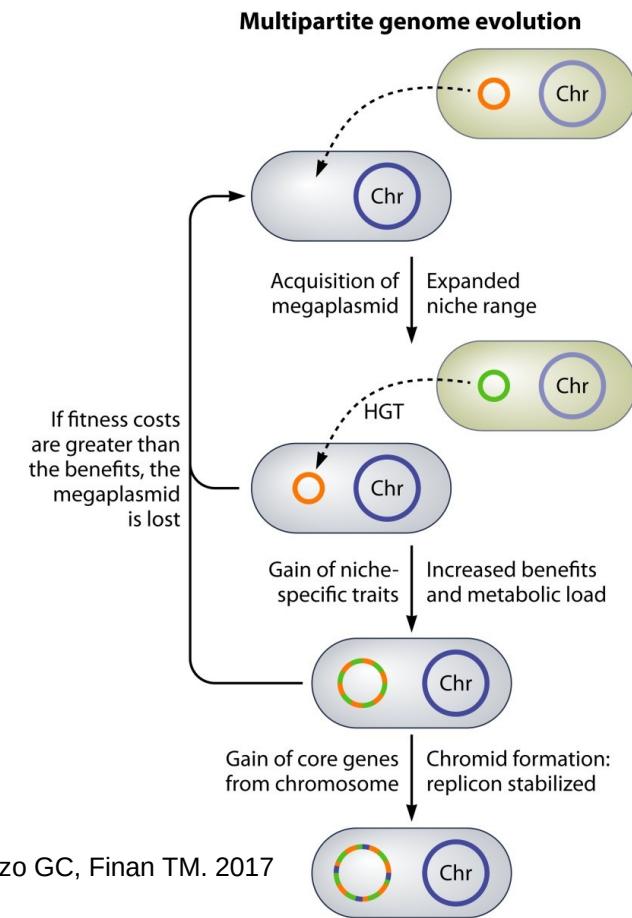


diCenzo GC, Finan TM. 2017

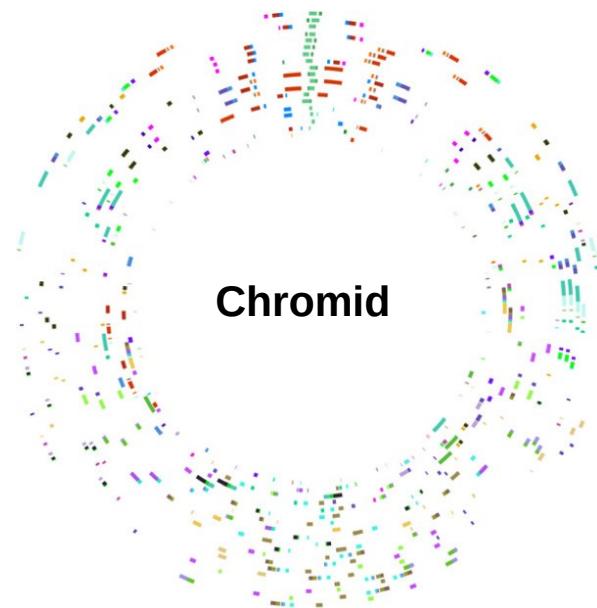
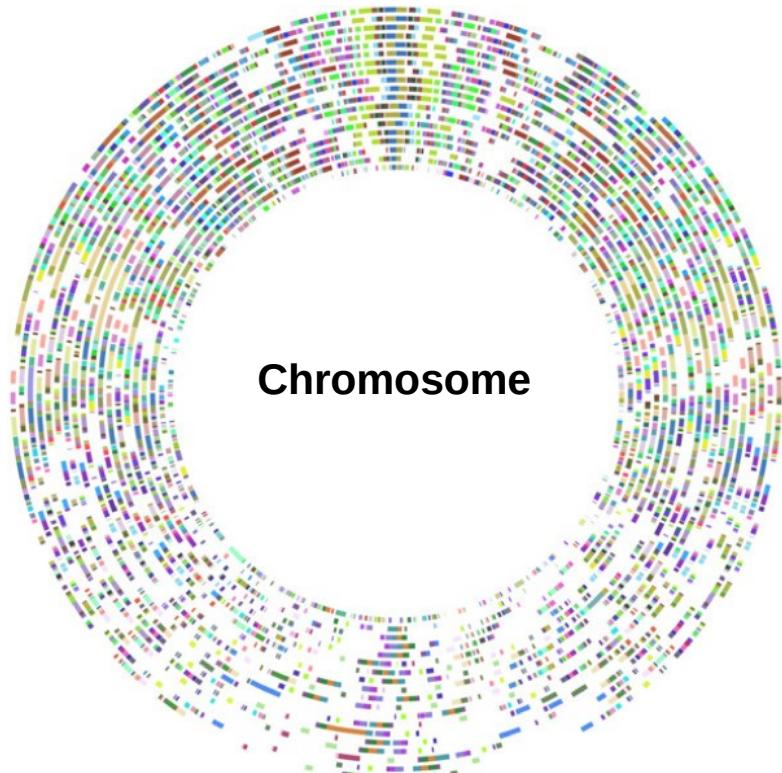
Chromid formation hypothesis



Chromid formation hypothesis



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 Карапь
Кристина Перевощикова
Заира Сефербекова

ШМТБ

Анна Тойдзе



Institute of Science and Technology



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.



School of molecular
and theoretical biology

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/>

