



# How *Ecoli* switch to dark side



Olga Bochkareva, PhD

# My career track

**BSc in Computational Physics,**  
Moscow Institute of Physics and  
Technology, Russia

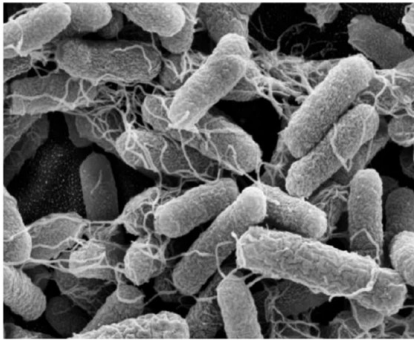
**PhD in Bioinformatics,**  
Institute for Information Transmission Problems, Russia  
*“Reconstruction of evolutionary history of genome  
rearrangements”*



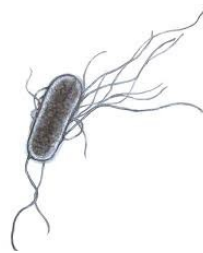
**MSc in Bioinformatics,**  
Moscow Institute of Physics  
and Technology, Russia

**IST plus (Marie Curie) postdoc fellowship**  
IST Austria  
*“Origin and maintenance of the multipartite  
bacterial genome”*

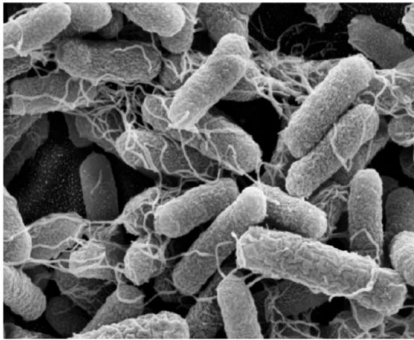
# Let's start with Wikipedia



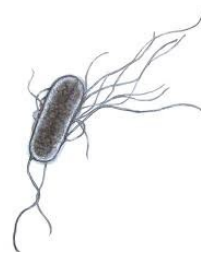
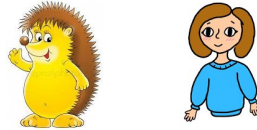
In 1885, the German-Austrian pediatrician [Theodor Escherich](#) discovered this organism in the feces of healthy individuals. He called it *Bacterium coli* commune because it is found in the colon. Later the organism was reclassified in the newly created genus *Escherichia*, named after its original discoverer.



# Let's start with Wikipedia

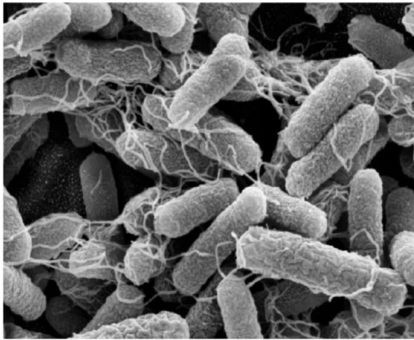


In 1885, the German-Austrian pediatrician [Theodor Escherich](#) discovered this organism in the feces of healthy individuals. He called it *Bacterium coli* commune because it is found in the colon. Later the organism was reclassified in the newly created genus *Escherichia*, named after its original discoverer.

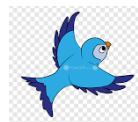
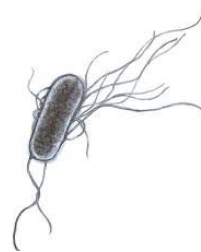
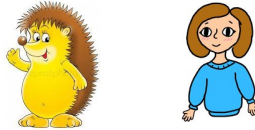




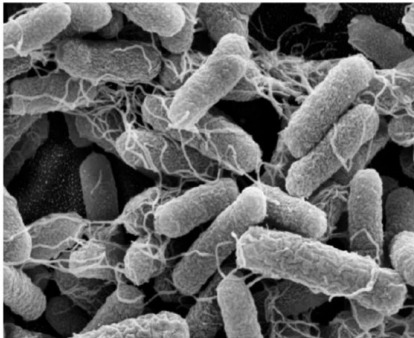
# Let's start with Wikipedia



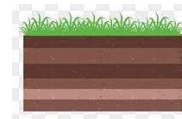
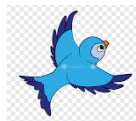
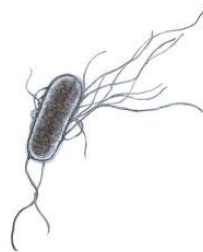
In 1885, the German-Austrian pediatrician [Theodor Escherich](#) discovered this organism in the feces of healthy individuals. He called it *Bacterium coli* commune because it is found in the colon. Later the organism was reclassified in the newly created genus *Escherichia*, named after its original discoverer.



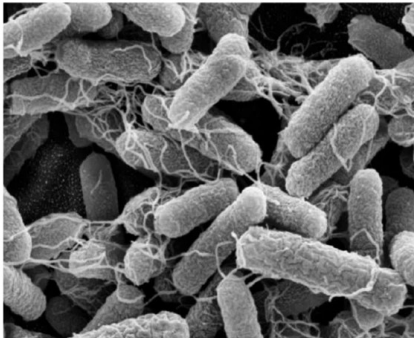
# Let's start with Wikipedia



In 1885, the German-Austrian pediatrician [Theodor Escherich](#) discovered this organism in the feces of healthy individuals. He called it *Bacterium coli* commune because it is found in the colon. Later the organism was reclassified in the newly created genus *Escherichia*, named after its original discoverer.



# Let's start with Wikipedia

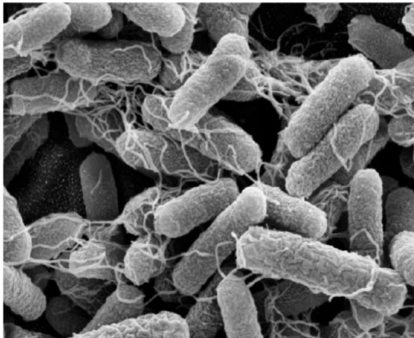


In 1885, the German-Austrian pediatrician [Theodor Escherich](#) discovered this organism in the feces of healthy individuals. He called it *Bacterium coli* commune because it is found in the colon. Later the organism was reclassified in the newly created genus *Escherichia*, named after its original discoverer.

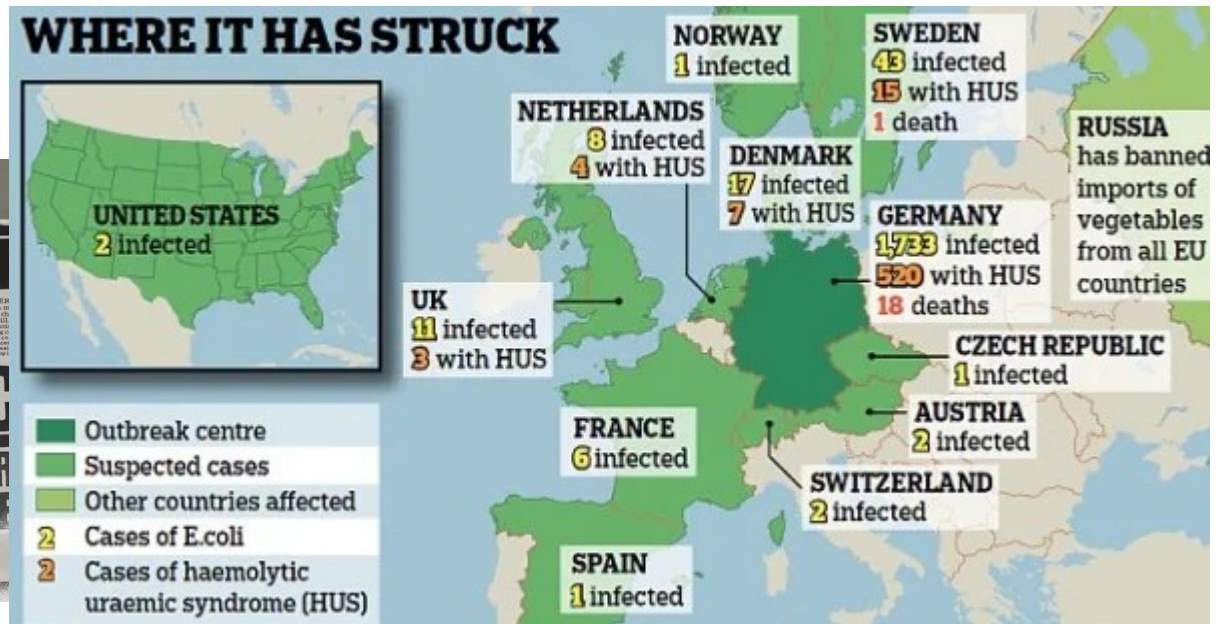
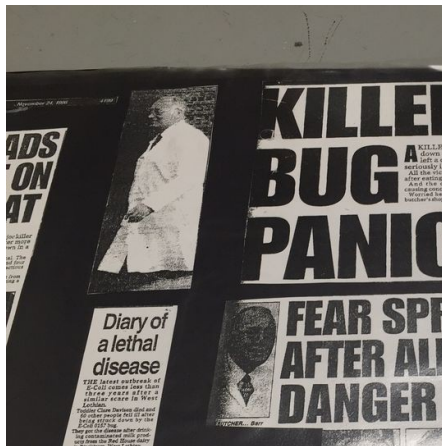


1996 *E. coli* food poisoning occurred in Wishaw, Scotland, killing 21 people

# Let's start with Wikipedia



In 1885, the German-Austrian pediatrician [Theodor Escherich](#) discovered this organism in the feces of healthy individuals. He called it *Bacterium coli* commune because it is found in the colon. Later the organism was reclassified in the newly created genus *Escherichia*, named after its original discoverer ...

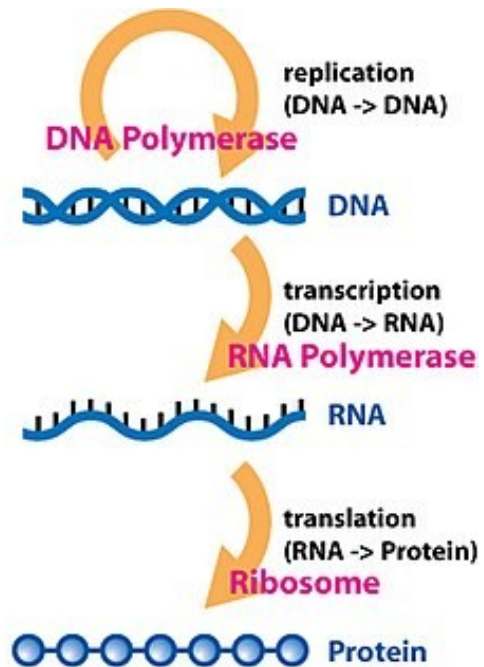


2011 Germany, Ecoli outbreak linked to organic fenugreek sprouts, killed 53 people.



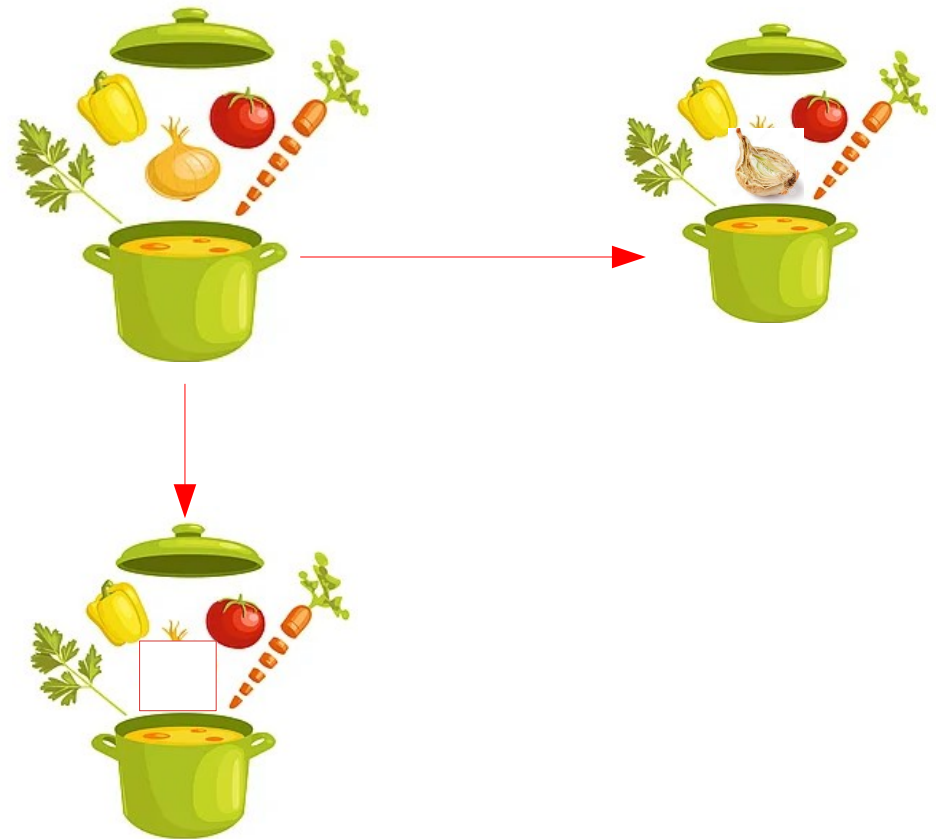
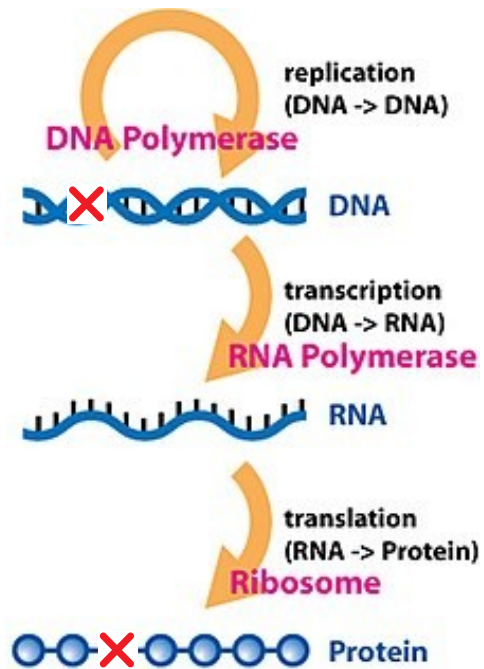
# How bacteria evolve

Central dogma of molecular biology



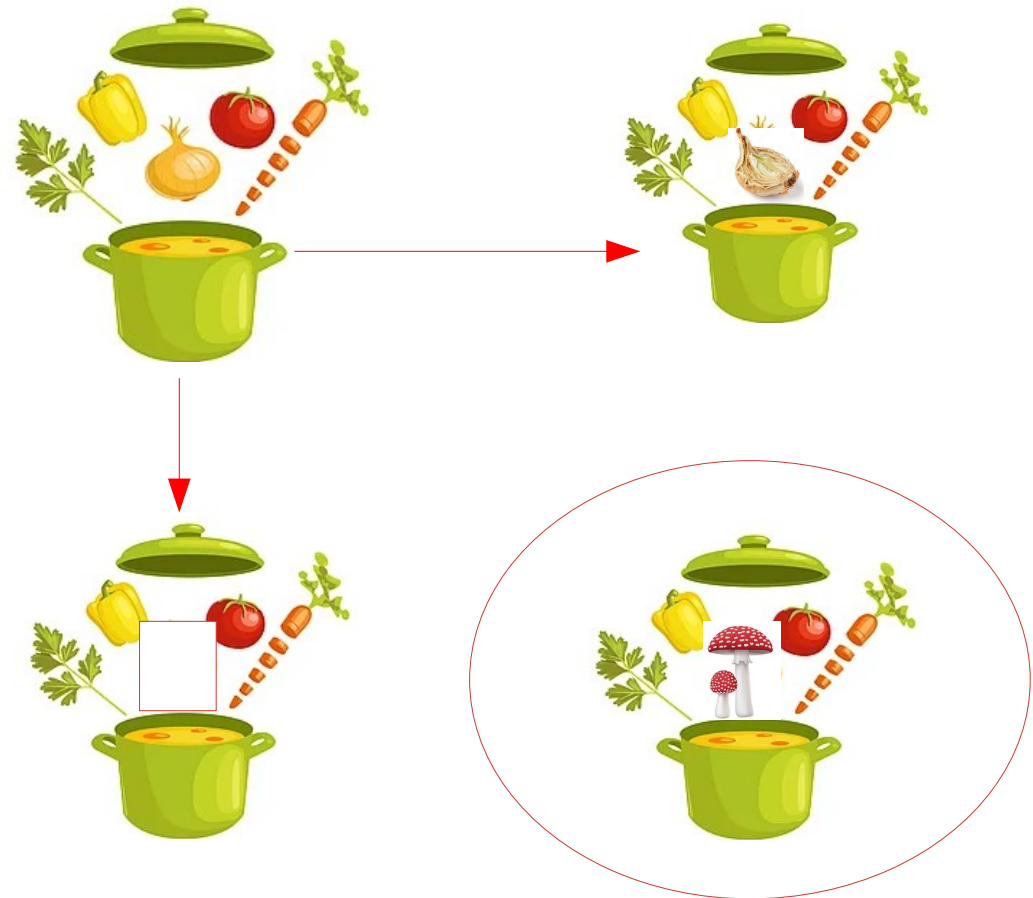
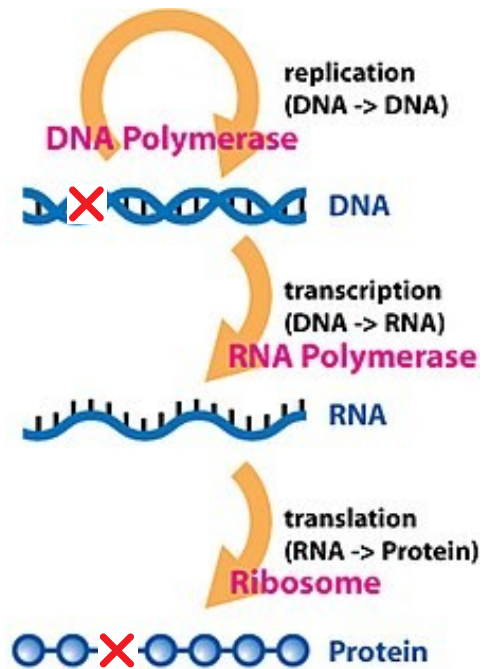
# How bacteria evolve

Central dogma of molecular biology



# How bacteria evolve

Central dogma of molecular biology



# How bacteria evolve

## Transformation



## Conjugation



## Transduction



1) Bacteria take up DNA from their environment

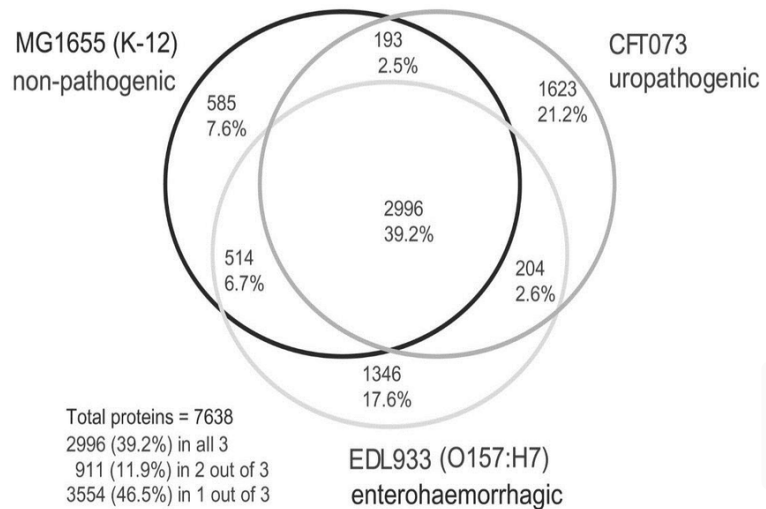
2) Bacteria directly transfer genes to another cell

3) Bacteriophages move genes from one cell to another

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4536854/>



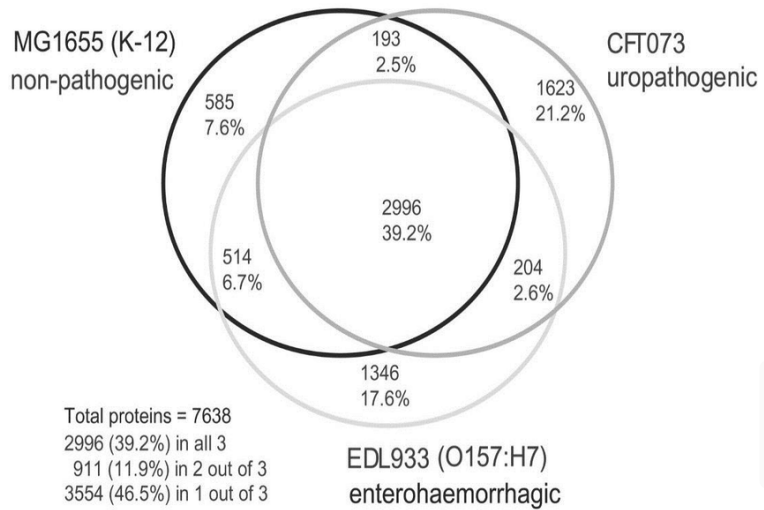
# What makes Ecoli pathogenic



Variations in gene content  
Venn diagram for 3 *E. coli*

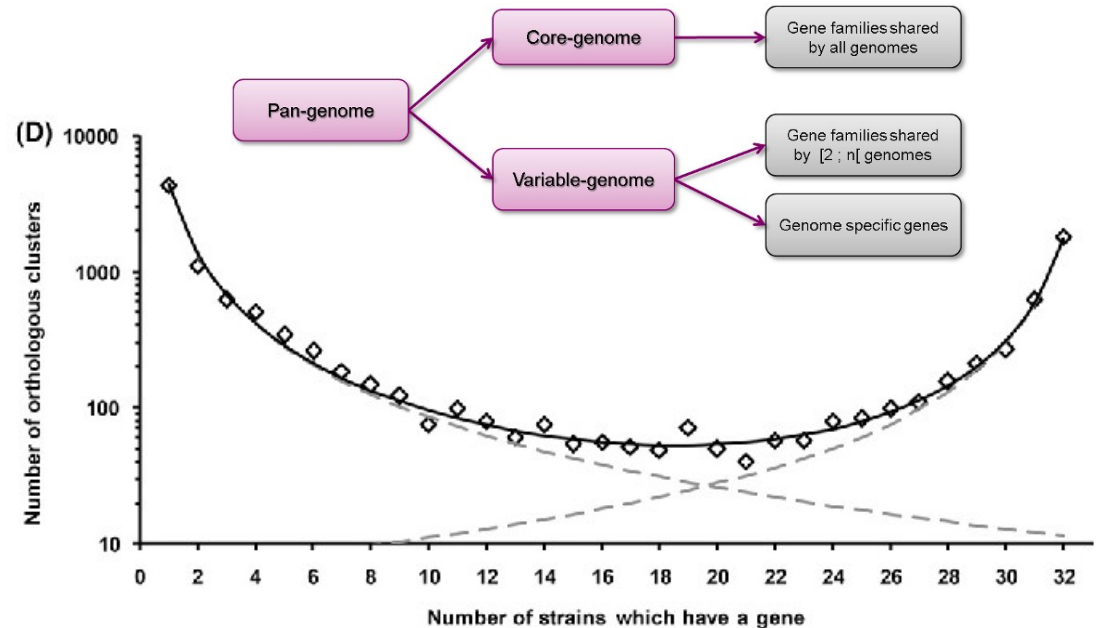


# What makes Ecoli pathogenic

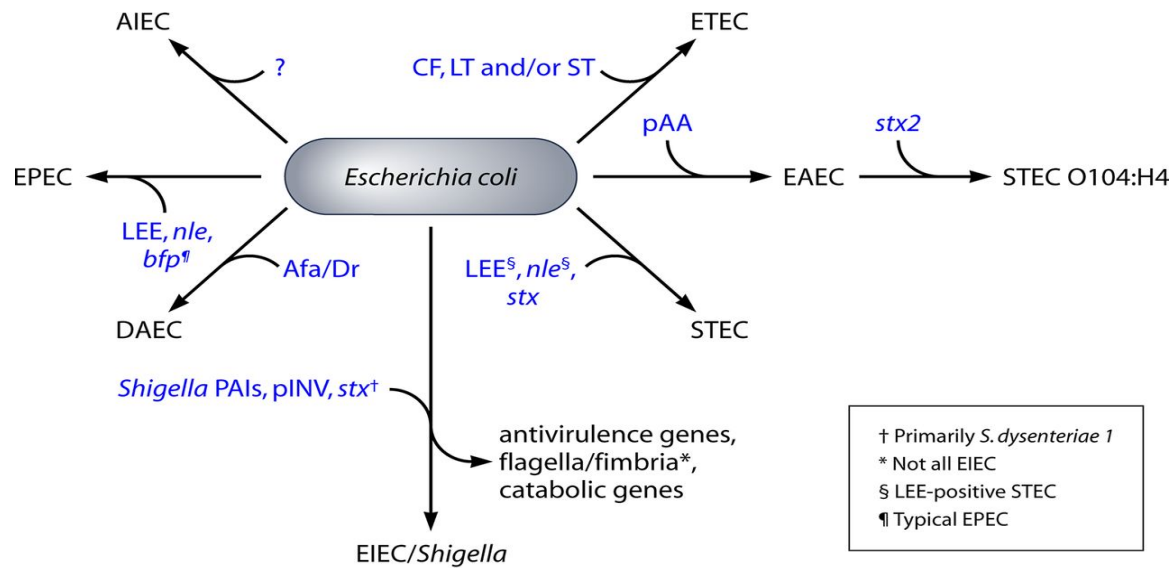


Variations in gene content  
 Venn diagram for 3 *E. coli*

Pan-genome approach  
 the entire set of genes from all strains within a clade



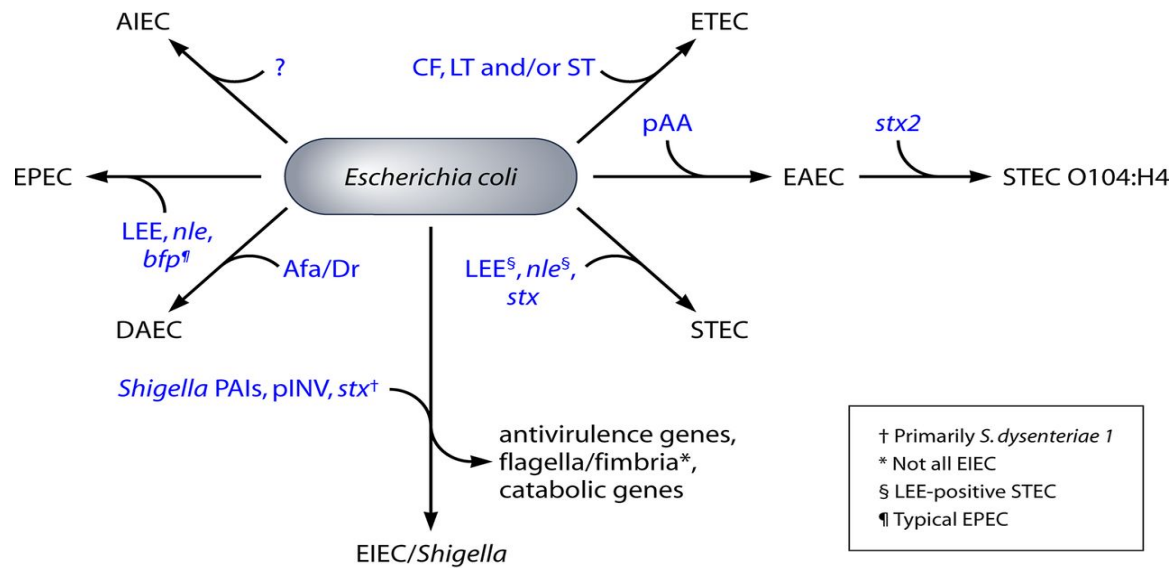
# What makes Ecoli pathogenic



## Pathotypes:

- enteropathogenic (EPEC) - Attaching and effacing lesions on intestinal epithelial cells
- enterotoxigenic (ETEC) - Heat-stable and heat-labile enterotoxins
- enteroinvasive (EIEC) - Colonocyte invasion
- enteroaggregative (EAEC) - Aggregative adhesion on enterocytes
- Shiga toxin-producing (STEC) - Shiga toxin
- diffusely adherent (DEAC) - Diffuse adhesion on enterocytes
- adherent invasive (AIEC) - Adhesion and invasion of intestinal epithelial cells

# What makes Ecoli pathogenic



The genus *Shigella* is named after Japanese physician Kiyoshi Shiga, who researched the cause of dysentery. In 1897 Shiga studied 32 dysentery patients and identify the bacterium causing the disease. He continued to study and characterize the bacterium, identifying its methods of toxin production i.e Shiga toxin.

## Pathotypes:

- enteropathogenic (EPEC) - Attaching and effacing lesions on intestinal epithelial cells
- enterotoxigenic (ETEC) - Heat-stable and heat-labile enterotoxins
- enteroinvasive (EIEC) - Colonocyte invasion
- enteroaggregative (EAEC) - Aggregative adhesion on enterocytes
- Shiga toxin-producing (STEC) - Shiga toxin
- diffusely adherent (DEAC) - Diffuse adhesion on enterocytes
- adherent invasive (AIEC) - Adhesion and invasion of intestinal epithelial cells



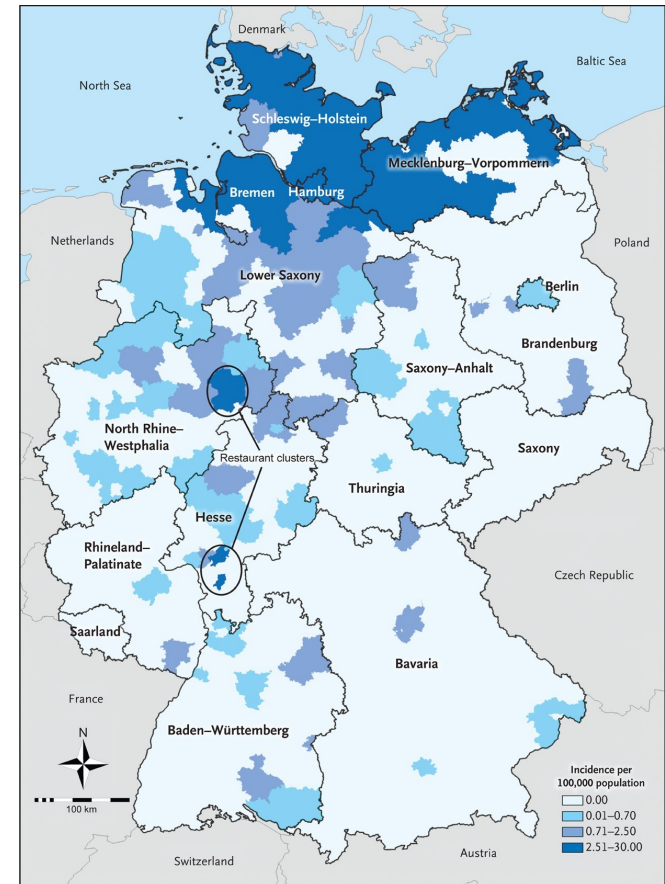
# What makes Ecoli pathogenic

## Strain *E.coli* O104:H4

~4000 people infected; 900 of those affected developed HUS, which was fatal for 54 individuals.

A combination of properties characteristic of **EHEC**, Stx2a production (although not harbouring the LEE of **EPEC**) and **EAEC**-type adherence:

- genes typical of enteroaggregative *E. coli*, such as *attA*, *aggR*, *aap*, *aggA*, and *aggC*, located on a virulence plasmid.
- gene for a Shiga-toxin 2 variant (*stx2a*).
- resistant to beta-lactam antibiotics (e.g., ampicillin) and third-generation cephalosporins and are partially resistant to fluoroquinolones (nalidixic acid).

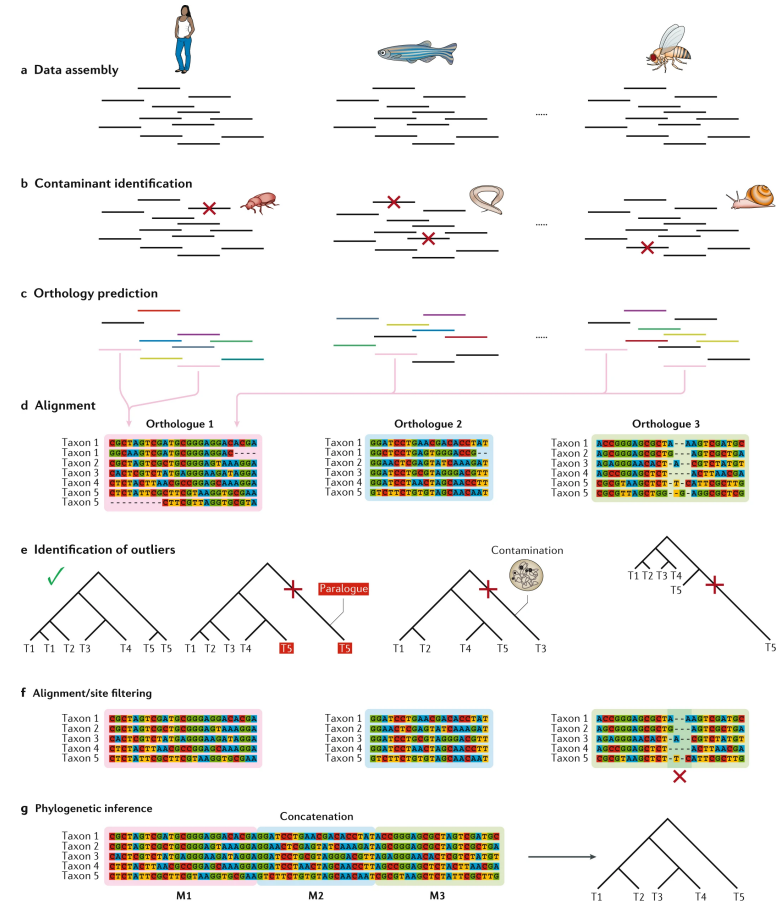
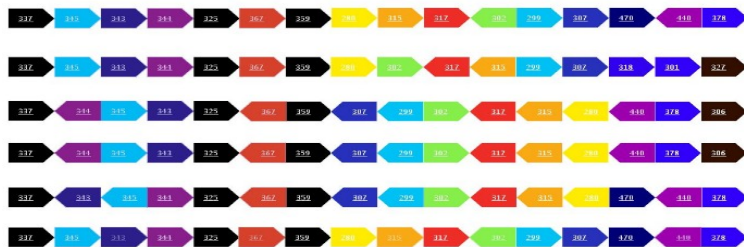


[https://en.wikipedia.org/wiki/2011\\_Germany\\_E.\\_coli\\_O104:H4\\_outbreak](https://en.wikipedia.org/wiki/2011_Germany_E._coli_O104:H4_outbreak)

# How can we analyze it?



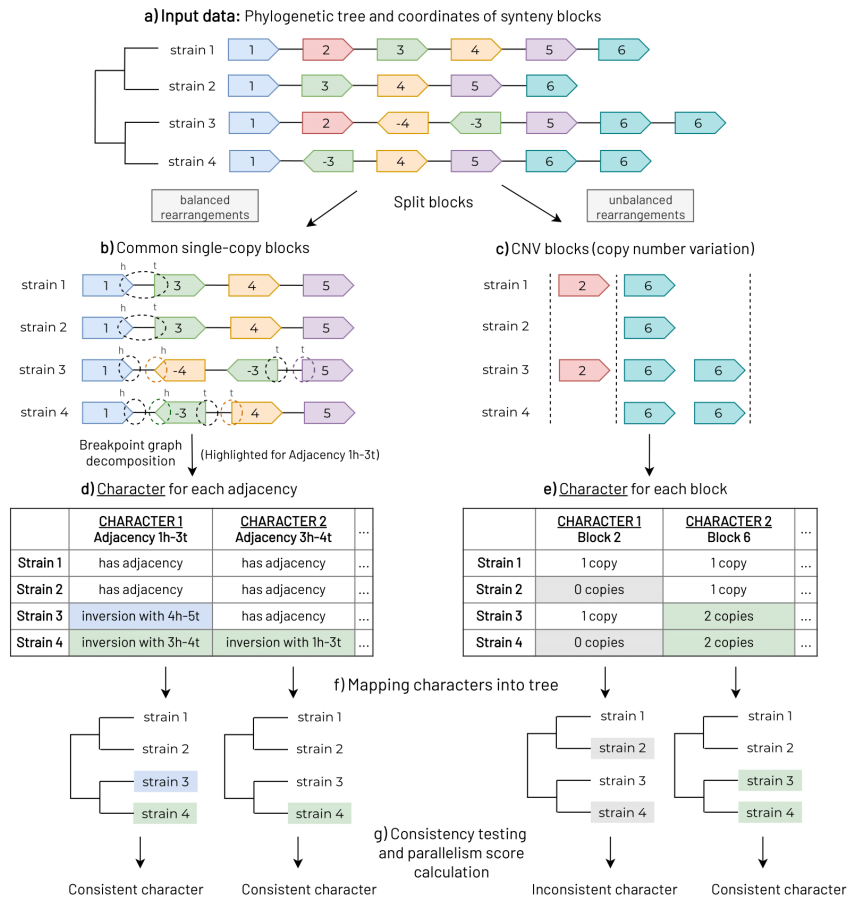
genome as an ordered set of genes and regulatory network



Kapli, P., Yang, Z. & Telford, M.J. **Phylogenetic tree building in the genomic age.** *Nat Rev Genet* 21, 428–444 (2020). <https://doi.org/10.1038/s41576-020-0233-0>

# How can we analyze it?

## Pipeline: Parallel Rearrangements and Breaks identification

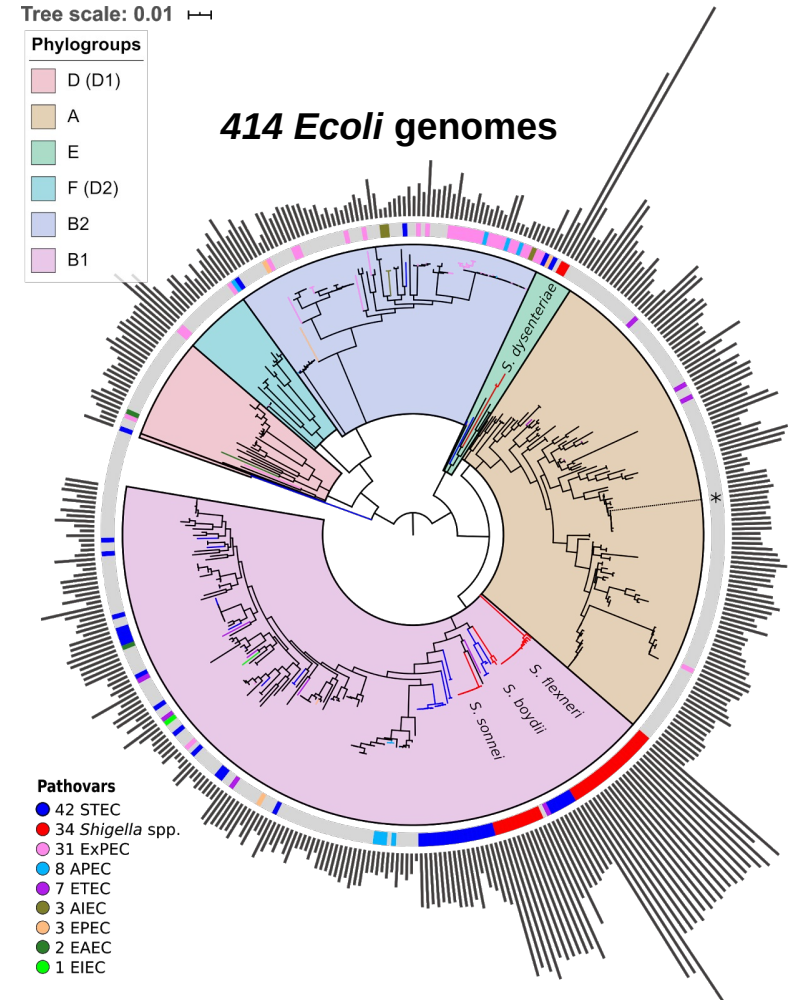


Tree scale: 0.01

### Phylogroups

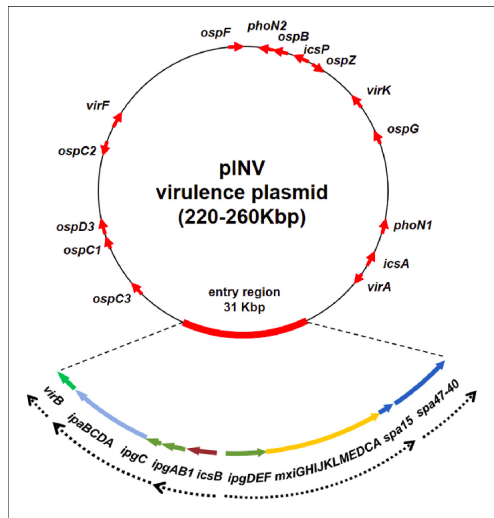
- D (D1)
- A
- E
- F (D2)
- B2
- B1

414 Ecoli genomes



# Acquisition of invasion

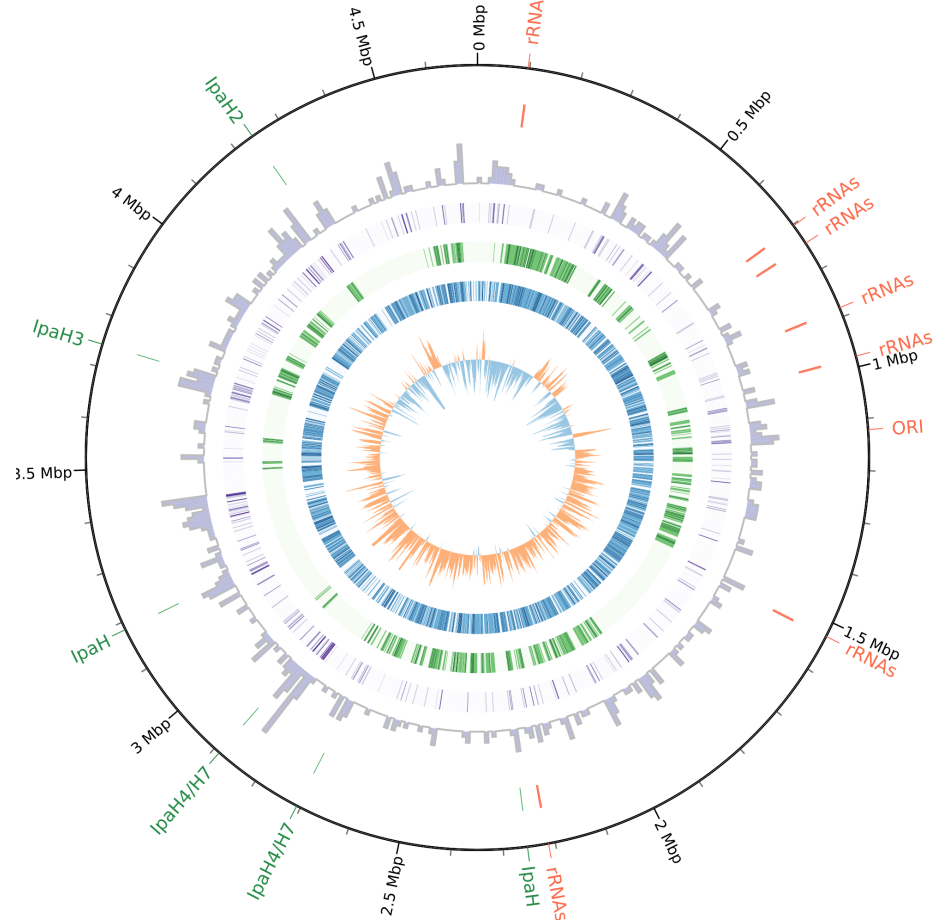
## Plasmid



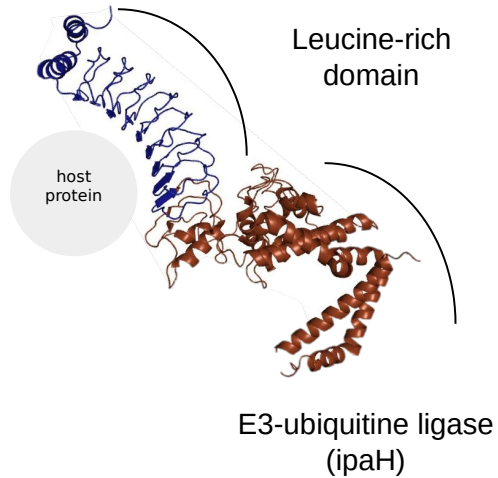
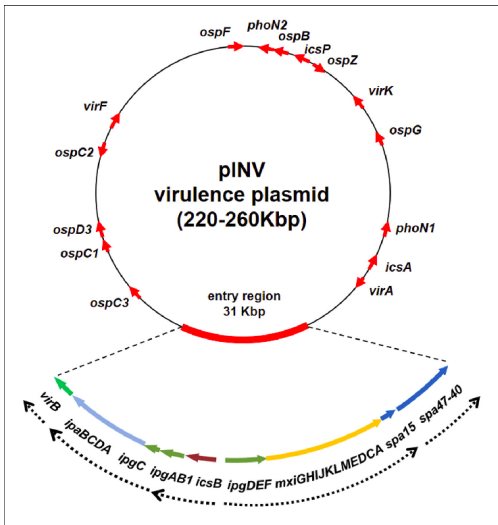
Genetic map of the pINV. The red arrows indicate major virulence determinants. Due to the variability in position and number, the *ipaH* genes are not shown.

Source: "The Intriguing Evolutionary Journey of Enteroinvasive *E. coli* (EIEC) toward Pathogenicity"

## Chromosome



# Acquisition of invasion



WP_134797310.1/1-547	1 MLVFNPLSTHVFYR.....TIDN.AANNYLEWVWTNSISGENTAFTRQLQGLSEGLDLCQRFL.RLFDNDIVNVDLHM.....LPELFRALKEINRQYV.LSALPELIVLEVINVSQHHFALPELFALEVINVSQHNLVLP	169
WP_000936833.1/1-568	1 MLVINNH.FLSQNSFWN.....ISG..TTADYFAWKWKEKALGGENNEAVSLKKEQLINFELQLNRLNLSLQDNLPFOITVLEIRQALISLPELPALEIVLDQDPLLPELFAKHLNDVNHQ.LLMLPELALLEYINADHQLTMLPELRTSLEVLSVRRHL	174
WP_134805189.1/1-583	1 MLTINNH.NRLISNFR.S.....YIDTSRAVENYLIHWTKNNRIQEQSDIAPORLVSCQDQNETINDLELGLITLP.EIIPQKSIINIKNNLISLPLASLTOLNVSYRLLIEPALPOGKLLNANHQ.LITLPTLRILKELKLVSNQCL	176
WP_128860970.1/1-587	1 MKRAHNE.....FFRFGLGLGICSRLLVEE.QNITDYRRIWNNWAKGGAATDITAVRLLKICLAFEPALNLSLLRRLP.YLPPHQELNINSNELRSLELPPSLTVLKASDRLRPLPALPHEVALDVSLNRVLTCLYLPSELSSALLNSLE	180
CDX08027.1/1-603	1 M.....LRNISSCLFRH.....LITITSNHYLSEWDDWEGKGLREBQTEAVRRLRAQLTSKHKLDLRALASGLR.VLPACIKKLDVCKKLLITLDRLENIKELIARDPLRHISALPHYLITLDVENQ.LENLPLLRDIKLSAEYRRLS	189
WP_134797310.1/1-547	157 LFMLELELDAAANALEVI.PDFPERDDH.....II..RIFWLNRIRITAIPEILQSSDQVNVNRENLQSPRIMQTLTQTAQPDYHGPRIFYFMSD.....GQNTLHRPLADAVTAWFPENKQSDVSIWHAFEHEEHANTFSAFLDR	302
WP_000936833.1/1-568	175 LPESLEALDVSTLLESIPAVPVR.....NHHHEETEI.....FFRCRENRITHIPEINLSLDRCTIILEDNPLSRIRELSQTAQPDYHGPRIFYFMSD.....GQNTLHRPLADAVTAWFPENKQSDVSIWHAFEHEEHANTFSAFLDR	323
WP_134805189.1/1-583	177 LPLFLEIISAIQNLSLEPLPHNIHSIWAIDNMLDIPYLPENL.....RNGYFDINDIHIPEISILNLRNECSIDISDNPLSRHALQSLQRLTSSPDYHGPRIFYFMSD.....GQNTLHRPLADAVTAWFPENKQSDVSIWHAFEHEEHANTFSAFLDR	338
WP_128860970.1/1-587	181 LCELELSAFDRLLEPLLPQNRLLVGEGQLHRLPELPRL.....QSLYIRNQNQLTLPDSIMNLHIYADNVIYNNPLSRTLALQRLTSSPDYHGPRIFYFMSD.....GQNTLHRPLADAVTAWFPENKQSDVSIWHAFEHEEHANTFSAFLDR	342
CDX08027.1/1-603	170 LSGNKILKVAHMLTELPLRRLQLFAYSRLNENIENIMRRFYFENGKHTITPTNLRDRHITIEANLPLSDTLFLIQTGTVENNGQERILEDDNRLFRQMLRQNLNHSRHRIVITEGQNFQIPPLPELVAWFPENKQSDVSIWHAFEHEEHANTFSAFLDR	358
WP_134797310.1/1-547	303 NTSGPREQVAAWLEKLSAELRQDSFAVAADATECEDRVALTWNLRKTLVHDAEGLFDNDTGALLSGREMPLEILEDIARDQVRTLHFVDEIEVYLAFDTMLAEKLLSLTAVNEMRFYGVSGVTANDLRTAEAMVRSRENEFDWFLWG	461
WP_000936833.1/1-568	324 NTSGPREQVAAWLEKLSAELRQDSFAVAADATECEDRVALTWNLRKTLVHDAEGLFDNDTGALLSGREMPLEILEDIARDQVRTLHFVDEIEVYLAFDTMLAEKLLSLTAVNEMRFYGVSGVTANDLRTAEAMVRSRENEFDWFLWG	512
WP_134805189.1/1-583	330 NTSGPREQVAAWLEKLSAELRQDSFAVAADATECEDRVALTWNLRKTLVHDAEGLFDNDTGALLSGREMPLEILEDIARDQVRTLHFVDEIEVYLAFDTMLAEKLLSLTAVNEMRFYGVSGVTANDLRTAEAMVRSRENEFDWFLWG	527
WP_128860970.1/1-587	343 NTSGPREQVAAWLEKLSAELRQDSFAVAADATECEDRVALTWNLRKTLVHDAEGLFDNDTGALLSGREMPLEILEDIARDQVRTLHFVDEIEVYLAFDTMLAEKLLSLTAVNEMRFYGVSGVTANDLRTAEAMVRSRENEFDWFLWG	531
CDX08027.1/1-603	350 NTSGPREQVAAWLEKLSAELRQDSFVAADATECEDRVALTWNLRKTLVHDAEGLFDNDTGALLSGREMPLEILEDIARDQVRTLHFVDEIEVYLAFDTMLAEKLLSLTAVNEMRFYGVSGVTANDLRTAEAMVRSRENEFDWFLWG	547
WP_134797310.1/1-547	402 VADRLLKASGLSDADAEREAGQVMRETEQQIYRQLTDEVLALRLPENGSQLHHS	547
WP_000936833.1/1-568	513 VADRLLKASGLSDADAEREAGQVMRETEQQIYRQLTDEVLALRLPENGSQLHHS	568
WP_134805189.1/1-583	528 VADRLLKASGLSDADAEREAGQVMRETEQQIYRQLTDEVLALRLPENGSQLHHS	583
WP_128860970.1/1-587	532 VADRLLKASGLSDADAEREAGQVMRETEQQIYRQLTDEVLALRLPENGSQLHHS	587
CDX08027.1/1-603	548 VADRLLKASGLSDADAEREAGQVMRETEQQIYRQLTDEVLALRLPENGSQLHHS	603

T3SS effector E3 ubiquitin-protein ligase IpaH

# Mess with gene names

Gene	Size (bp)	<i>S. flexneri</i>		<i>S. flexneri</i>	<i>S. dysent.</i>	<i>S. sonnei</i>	<i>S. boydii</i>
		8401	301	2457T	197	046	227
<i>ipaHa1</i>	1764	<i>ipaH7</i>	<i>ipaH1</i>	<i>ipaH1</i> (fs)	<i>ipaH1</i>	<i>ipaH1</i>	<i>ipaH6</i>
<i>ipaHa2</i>	1764	-	-	-	<i>ipaH2</i> (IS)	-	-
<i>ipaHb1</i>	1827	<i>ipaH1</i>	<i>ipaH2</i> (fs)	<i>ipaH2</i>	<i>ipaH6</i>	<i>ipaH5</i>	<i>ipaH1</i>
<i>ipaHb2</i>	1827	<i>ipaH6</i>	<i>ipaH7</i>	<i>ipaH7</i>	-	-	-
<i>ipaHc</i>	1716	<i>ipaH2</i>	<i>ipaH3</i>	-	<i>ipaH4</i>	<i>ipaH3</i>	<i>ipaH4</i>
<i>ipaHd</i>	1752	<i>ipaH3</i>	<i>ipaH4</i>	<i>ipaH4</i>	<i>ipaH5</i>	<i>ipaH2</i>	<i>ipaH5</i>
<i>ipaHe1</i>	1644	<i>ipaH4</i>	<i>ipaH5</i> (IS)	<i>ipaH5</i>	<i>ipaH3</i> (IS)	<i>ipaH4</i>	<i>ipaH2</i>
<i>ipaHe2</i>	1644	<i>ipaH5</i>	<i>ipaH6</i> (IS)	<i>ipaH6</i> (fs)	-	-	<i>ipaH3</i>

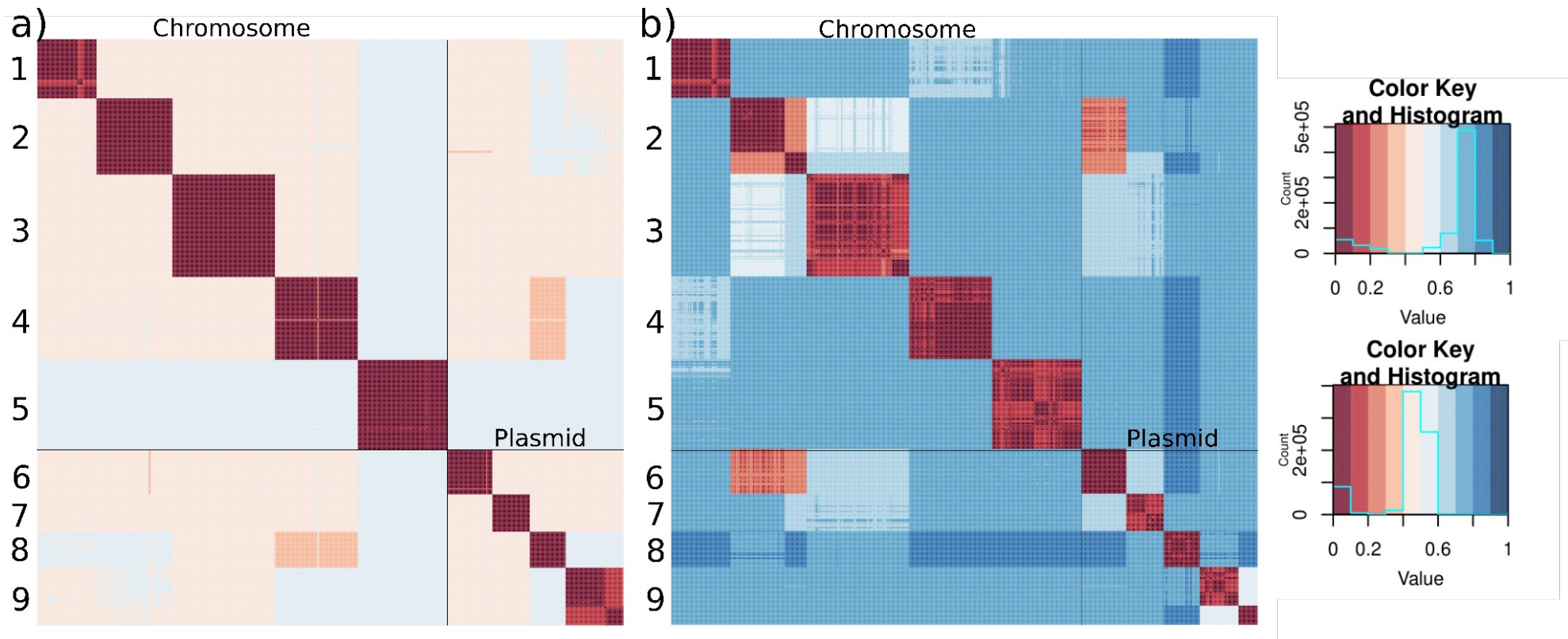
<sup>a</sup>Names used to annotate the chromosomal *ipaH* genes in different genomes are indicated; in this study, we used the letter-based nomenclature indicated in the left column. Genes inactivated by a frameshift mutation or an insertion sequence are indicated (fs) and (IS), respectively, and genes absent from a genome are indicated by a hyphen.

doi:10.1371/journal.pone.0032862.t001



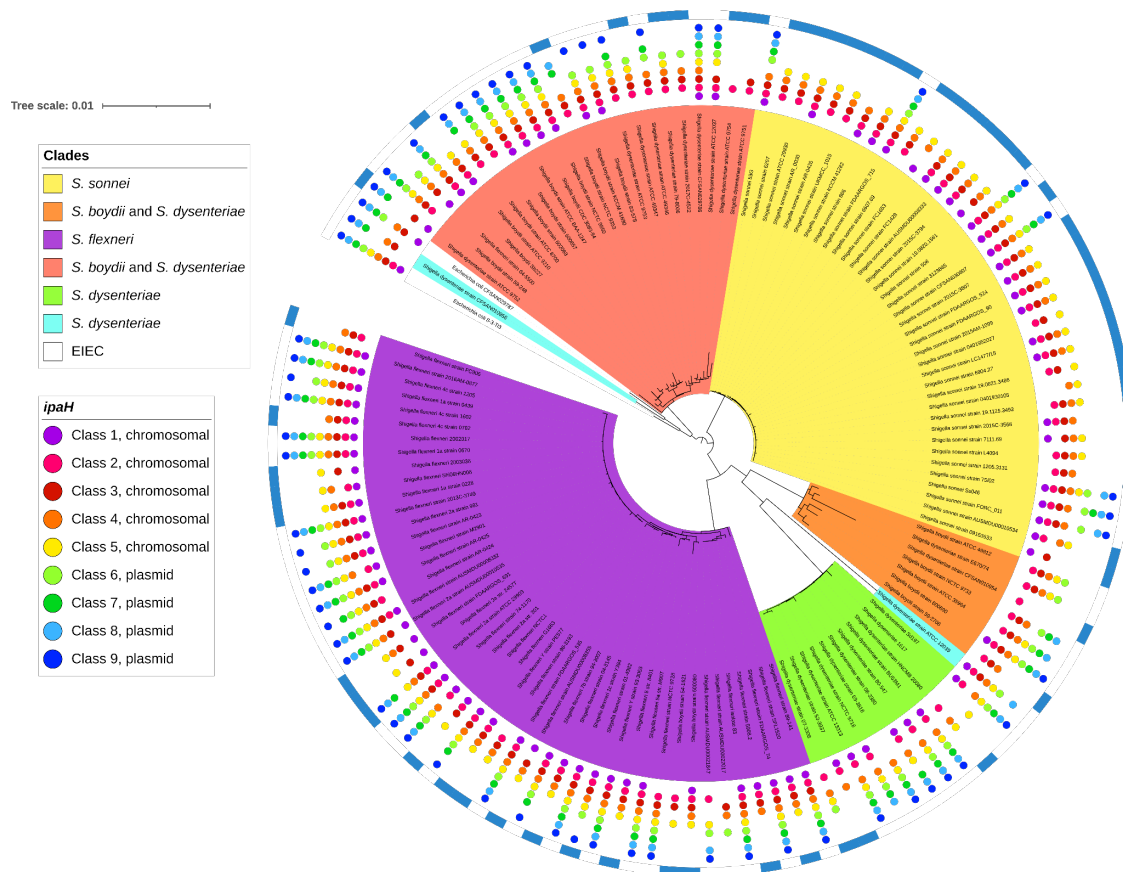
# Now it's solved

Heatmap of identity level of (a) the *ipaH* genes; (b) their upstream sequences in *Shigella*.



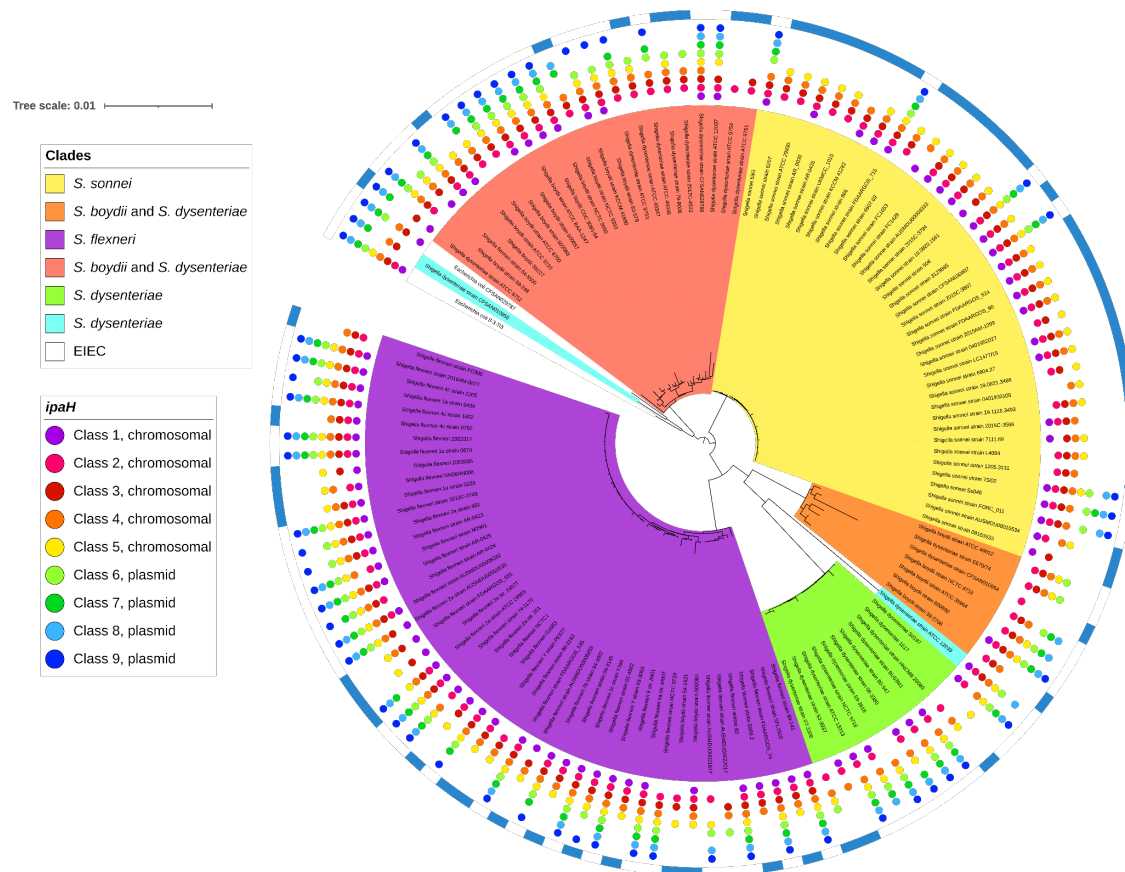
# And classified!

## Composition of *virulence* genes in human-host pathogenic *E.coli*

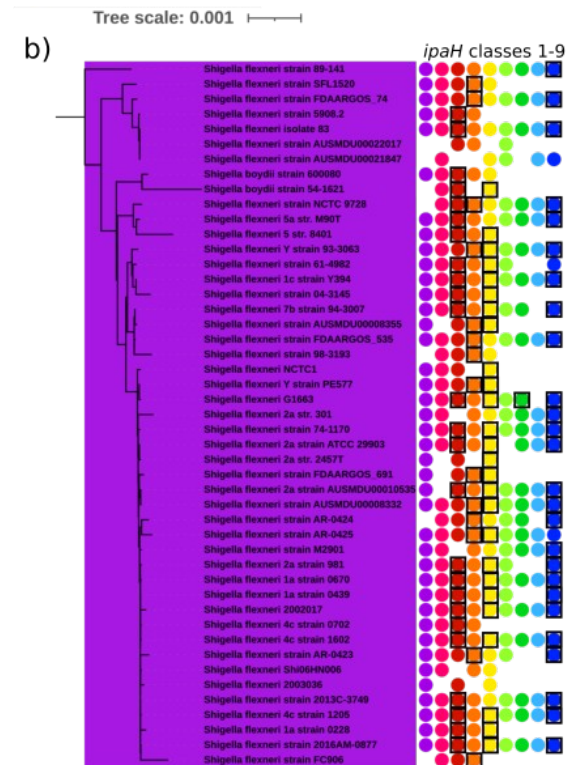


# And classified!

## Composition of virulence genes in human-host pathogenic *E.coli*

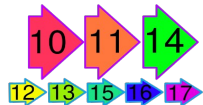


## And their copies

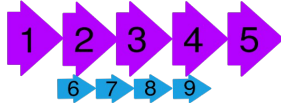


# But wait... what's about animals?

**ipaH class**  
Non-human-host *ipaH*

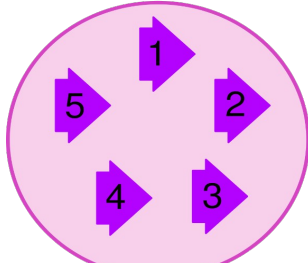


Human-host *ipaH*

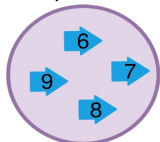


**Shigella/EIEC**  
(human)

Chromosome

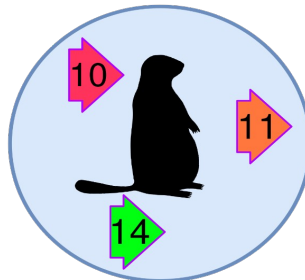


*pINV*



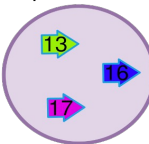
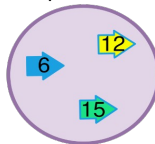
**Escherichia marmotae**  
(marmot)

Chromosome



*pEM76*

*pEM148*



Genomic and molecular characterisation of *Escherichia marmotae* from wild rodents in Qinghai-Tibet plateau as a potential pathogen. *Sci Rep*, 2019

## Alignment of C-terminal domains

**ShigellaCDomain** GQQNTLHRPLADAVTAWFPENKQSDVVSQIWHAFEEHEEHANTFSAFLDRLSDTVSARNISG  
**NonHumanCDomain1** GQQHTSVRPLPEAVAAWFPQNKQSDVVSQIWLAFEREEHANTFSAFLDRLADTVSARNAQG  
**NonHumanCDomain2** GQQHTPVRSLPEAVAAWFPESLRSEVSQRWGAFTDEENAAATFSAFLDRLADTVTARNAPG

**ShigellaCDomain** FREQVAWLEKLSASAE LRQQSFAVAADATESCEDRVALTWNLRKTL LVHQASEGLFDN  
**NonHumanCDomain1** FRQQVSFALEKLSTSAELRQQSFAVAADATESCEDRVALTWNLRKTL LVHQASEGLFDN  
**NonHumanCDomain2** FAQQVSEFLEKLSRSEALRQQCFVAVAADATRSCEDRVALTWNLRKTYRVHQASEGEFDS

**ShigellaCDomain** DTGALLSLGREMERLEILEDIARQKVRTLHFVDEIEVYLAFTQMLAEKQLSTAVKEMRF  
**NonHumanCDomain1** DTGALLSLGREMERLEILEDIARQKVRTLRFVDEIEVYLAFTQMLAEKQLSTAVKEMRF  
**NonHumanCDomain2** DLTGALLSLGREMYRLEVLEEIAREKVRTLHFVDEIEVYLAFTQMLAEKQLSTAVKEMRF

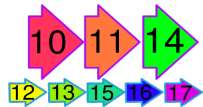
**ShigellaCDomain** YGVSGVTANDLRTAEAMVRSREENEFDTWFSLWGPWHAVLKRTEADRWAAEECKYEMLE  
**NonHumanCDomain1** YGVSGVTENDLRTAEAMVRSREENEFDTWFALWGPWHAVLKRTEADRWAAEECKYEMLE  
**NonHumanCDomain2** YGVSGVTEDDLSSALVRVLSREEREFAEWFAWSPWHAVLKRTEAERWARAEECKYEMLE

**ShigellaCDomain** NEYPQRVADRLKASGLSGDAEREAGQVMRETEQQIYRQLTDEVIALRLSENGSQLHH  
**NonHumanCDomain1** NEYPQRVADRLKASGLSDDADAEREAGQVMRETEQQIYRQLTDEVIALRLPENGSRLHH  
**NonHumanCDomain2** REYPORLAERLSALGLSGDGAEREAGVRVMEETETETIYRQLTDEVISGERLAENRARRxxx

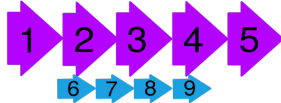


# Oh... there are many others!

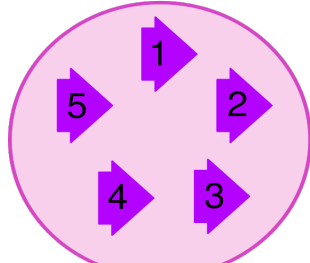
**ipaH class**  
Non-human-host *ipaH*



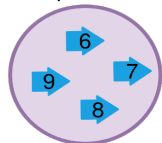
Human-host *ipaH*



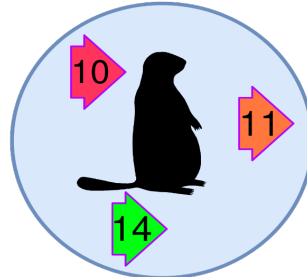
**Shigella/EIEC**  
(human)  
Chromosome



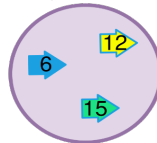
*pINV*



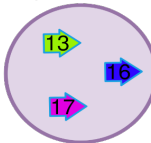
**Escherichia marmotae**  
(marmot)  
Chromosome



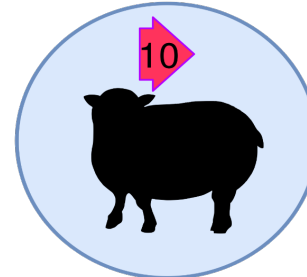
*pEM76*



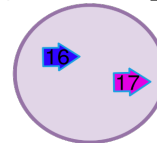
*pEM148*



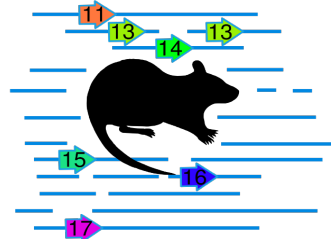
**Escherichia coli**  
(sheep)  
Chromosome



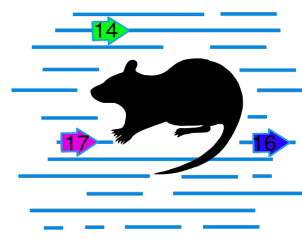
*pRHB04-C17\_2*



**Escherichia coli**  
(rat 1)



**Escherichia coli**  
(rat 2)

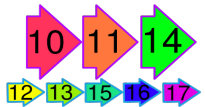


Composition of *ipaH* genes in *Escherichia* genomes from different hosts.

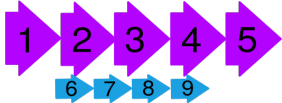


# But how many?..

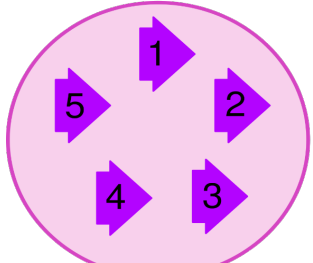
**ipaH class**  
Non-human-host ipaH



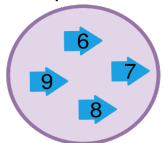
Human-host ipaH



**Shigella/EIEC**  
(human)  
Chromosome

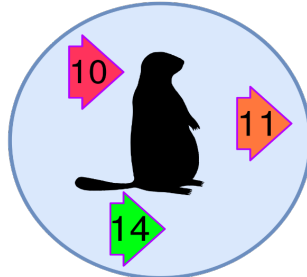


pINV

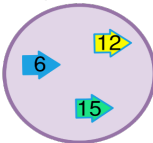


**Escherichia marmotae**  
(marmot)

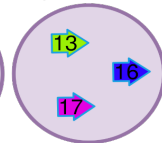
Chromosome



pEM76

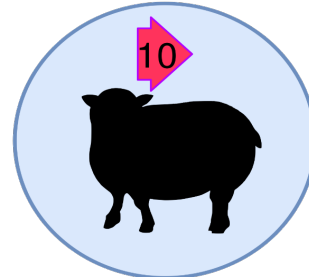


pEM148

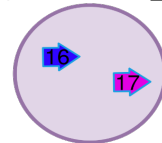


**Escherichia coli**  
(sheep)

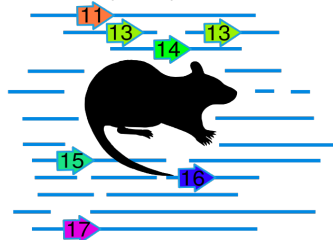
Chromosome



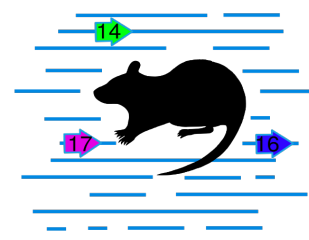
pRHB04-C17\_2



**Escherichia coli**  
(rat 1)



**Escherichia coli**  
(rat 2)



Composition of ipaH genes in *Escherichia* genomes from different hosts.



# Take-home messages



Wash your  
hands

# Take-home messages



Wash your  
hands



Sequence  
microbes

# Take-home messages



Wash your  
hands



Sequence  
microbes



Do  
bioinformatics

