



How *Ecoli* switch to dark side

Olga Bochkareva, PhD



My career track

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

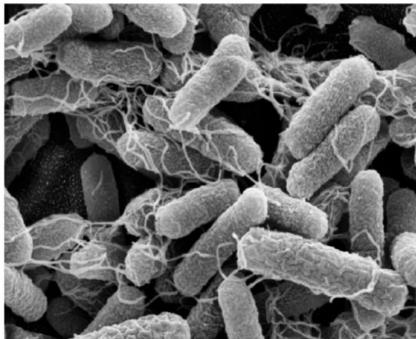
PhD in Bioinformatics,
Institute for Information Transmission Problems, Russia
“Reconstruction of evolutionary history of genome rearrangements”



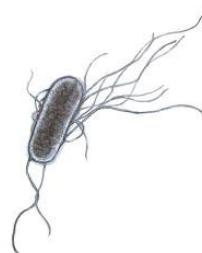
MSc in Bioinformatics,
Moscow Institute of Physics
and Technology, Russia

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

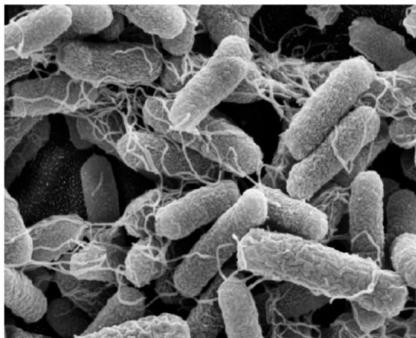
Let's start with Wikipedia



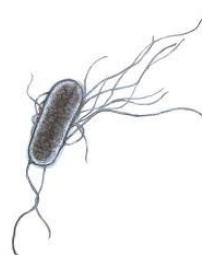
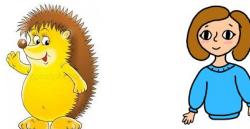
In 1885, the German-Austrian pediatrician [Theodor Escherich](#) discovered this organism in the feces of healthy individuals. He called it *Bacterium coli* commune because it is found in the colon. Later the organism was reclassified in the newly created genus *Escherichia*, named after its original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



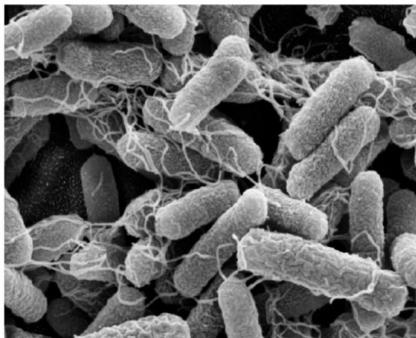
Let's start with Wikipedia



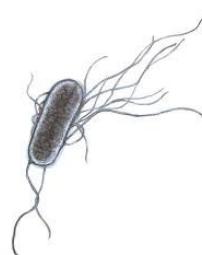
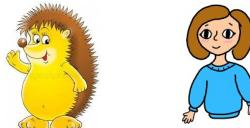
In 1885, the German-Austrian pediatrician [Theodor Escherich](#) discovered this organism in the feces of healthy individuals. He called it *Bacterium coli* commune because it is found in the colon. Later the organism was reclassified in the newly created genus *Escherichia*, named after its original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



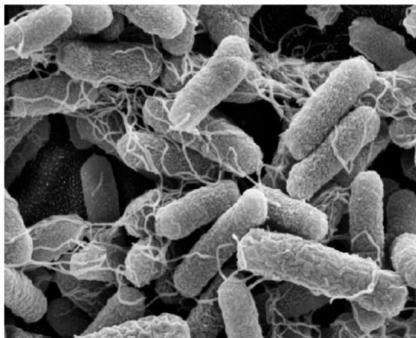
Let's start with Wikipedia



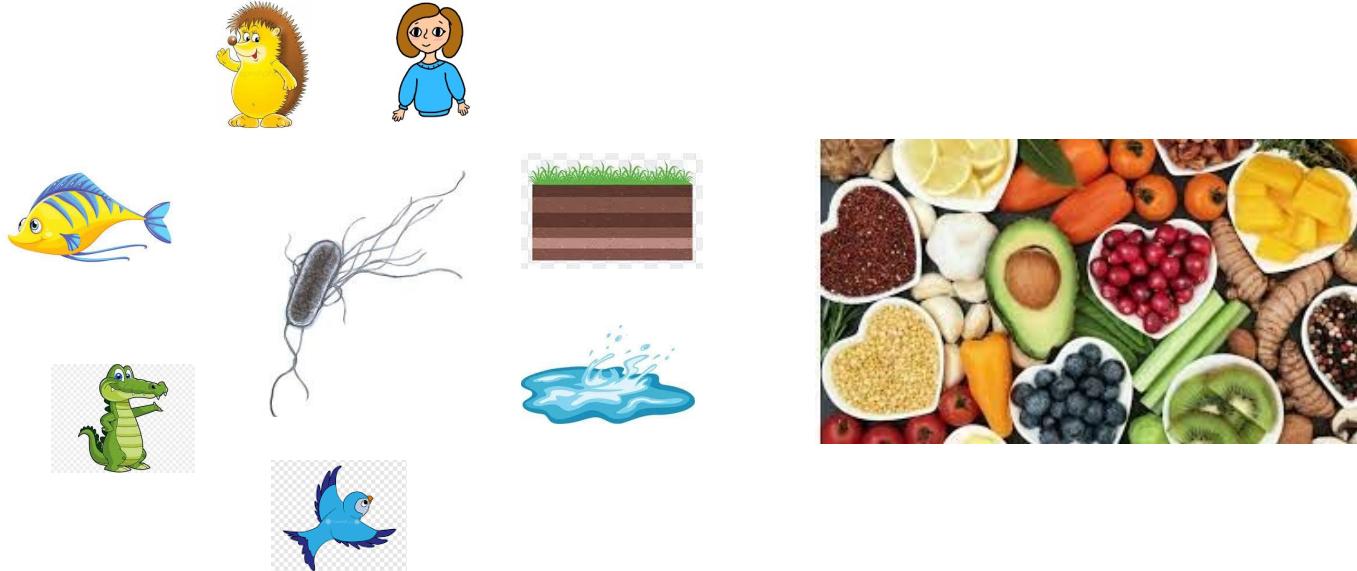
In 1885, the German-Austrian pediatrician [Theodor Escherich](#) discovered this organism in the feces of healthy individuals. He called it *Bacterium coli* commune because it is found in the colon. Later the organism was reclassified in the newly created genus *Escherichia*, named after its original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



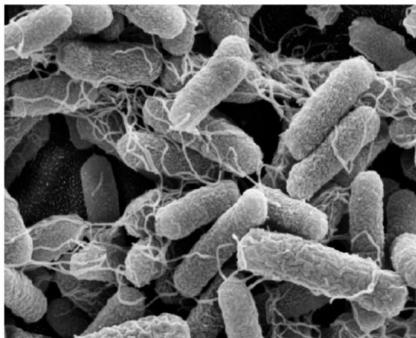
Let's start with Wikipedia



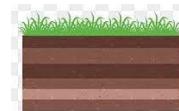
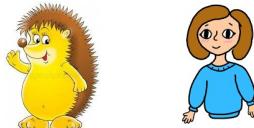
In 1885, the German-Austrian pediatrician [Theodor Escherich](#) discovered this organism in the feces of healthy individuals. He called it *Bacterium coli* commune because it is found in the colon. Later the organism was reclassified in the newly created genus *Escherichia*, named after its original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



Let's start with Wikipedia

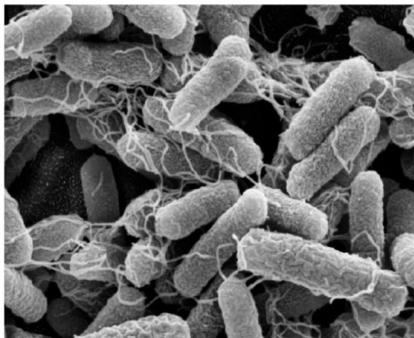


In 1885, the German-Austrian pediatrician [Theodor Escherich](#) discovered this organism in the feces of healthy individuals. He called it *Bacterium coli* commune because it is found in the colon. Later the organism was reclassified in the newly created genus *Escherichia*, named after its original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

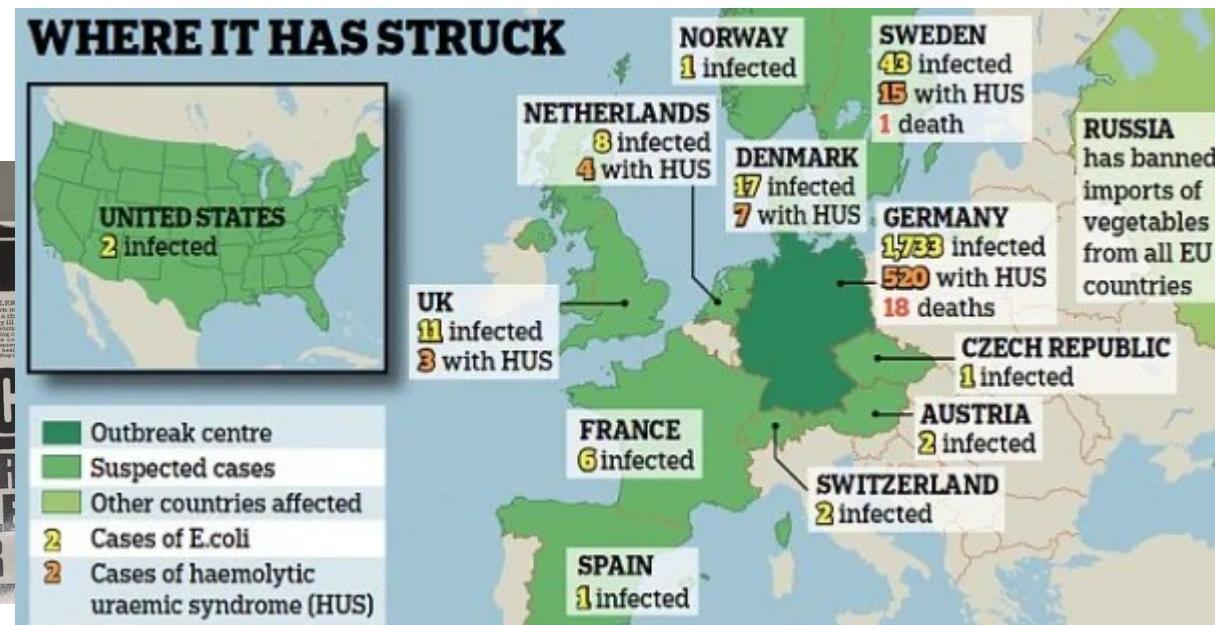


1996 E. coli food poisoning occurred in Wishaw, Scotland, killing 21 people

Let's start with Wikipedia



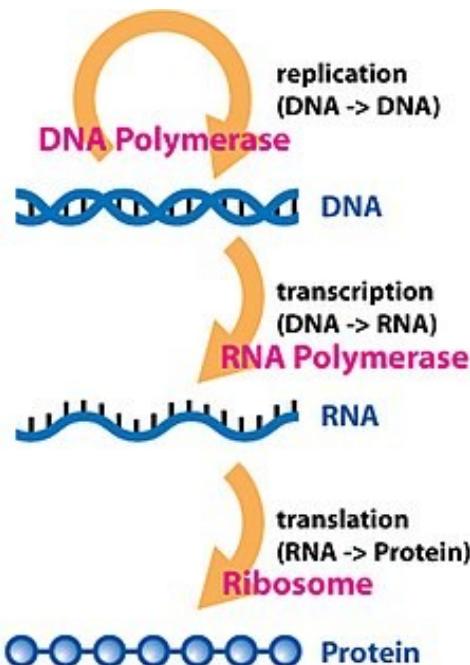
In 1885, the German-Austrian pediatrician [Theodor Escherich](#) discovered this organism in the feces of healthy individuals. He called it *Bacterium coli* commune because it is found in the colon. Later the organism was reclassified in the newly created genus *Escherichia*, named after its original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



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

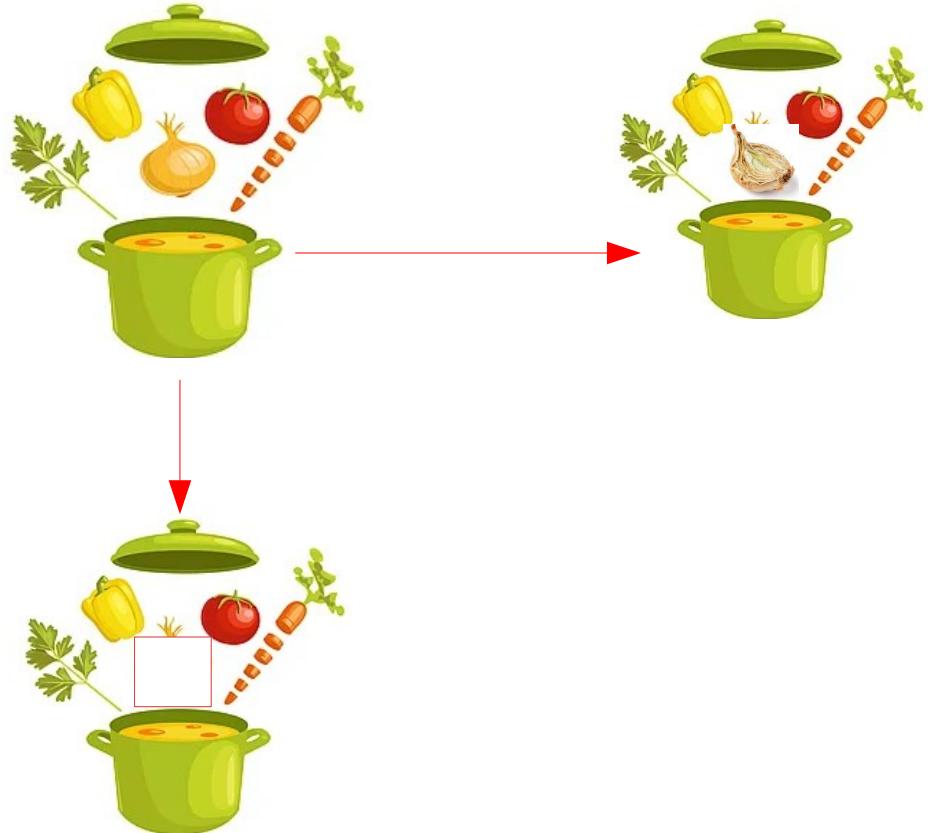
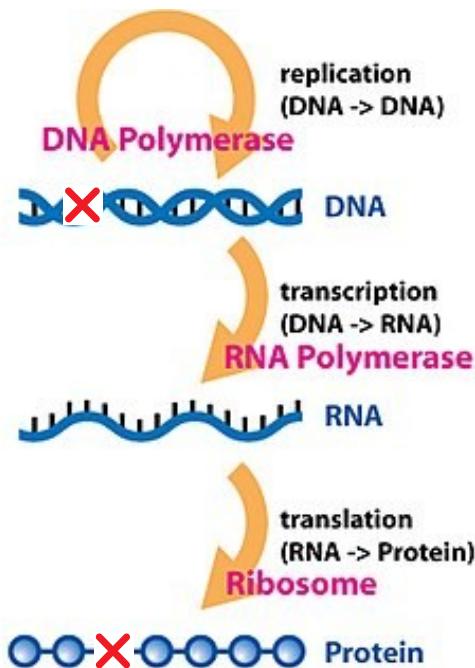
How bacteria evolve

Central dogma of molecular biology



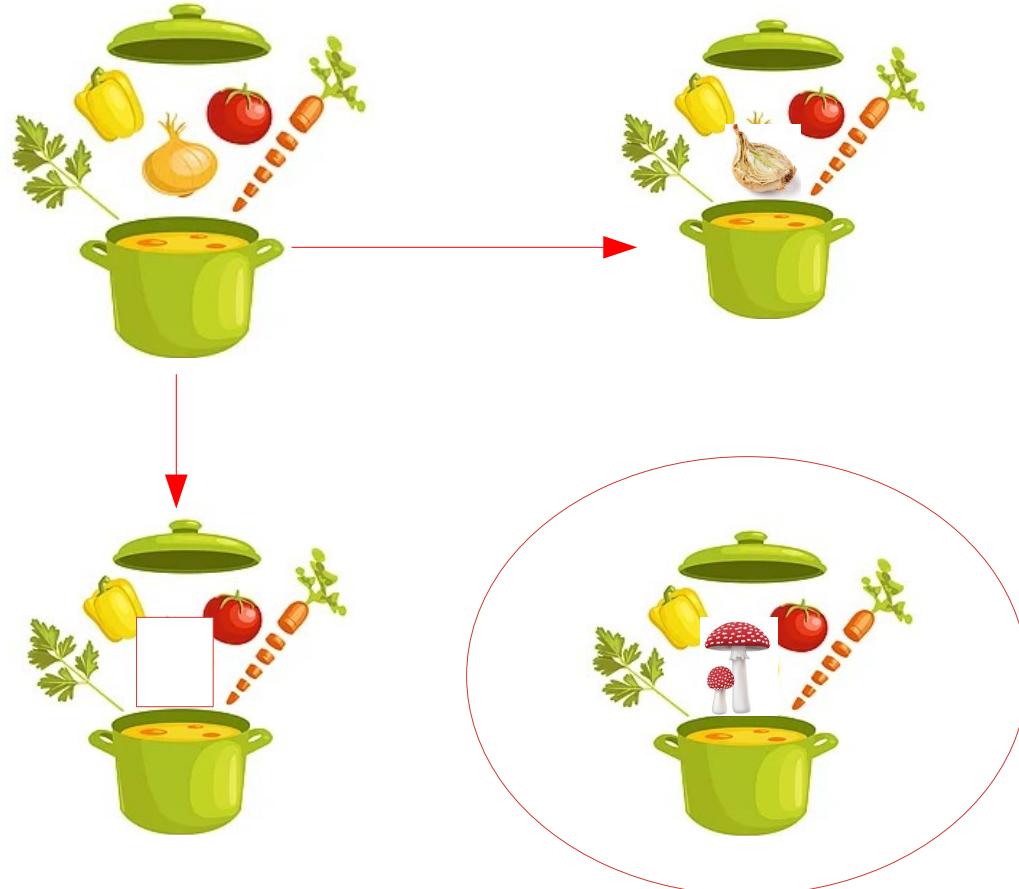
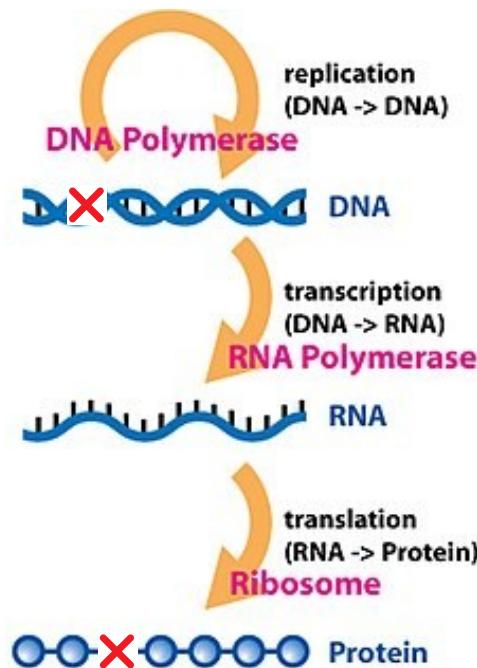
How bacteria evolve

Central dogma of molecular biology



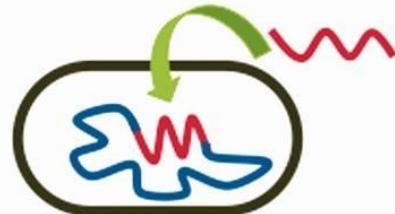
How bacteria evolve

Central dogma of molecular biology

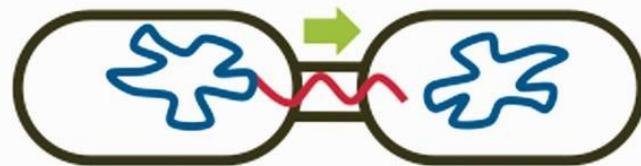


How bacteria evolve

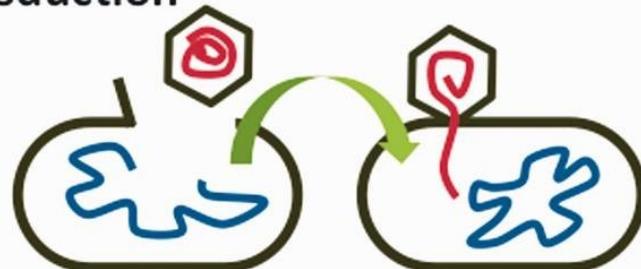
Transformation



Conjugation



Transduction



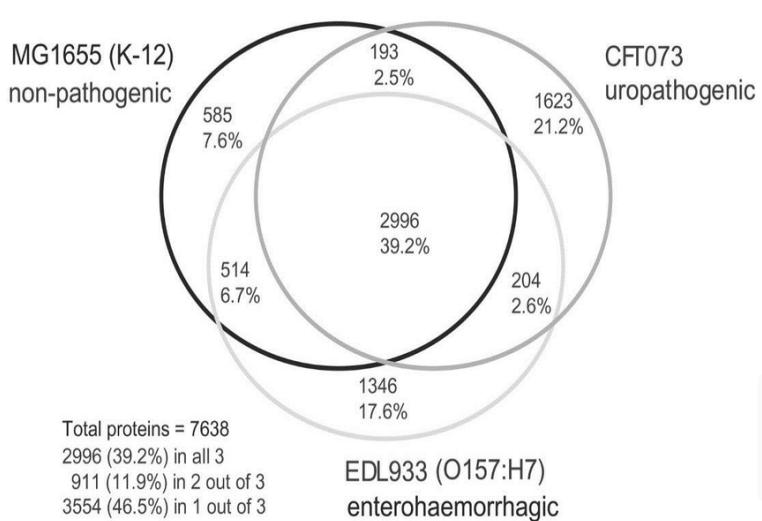
1) Bacteria take up DNA from their environment

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/>

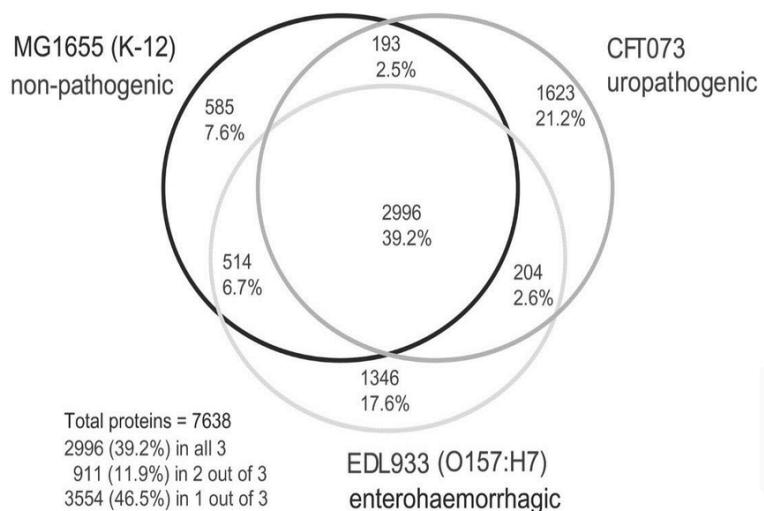
What makes Ecoli pathogenic



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



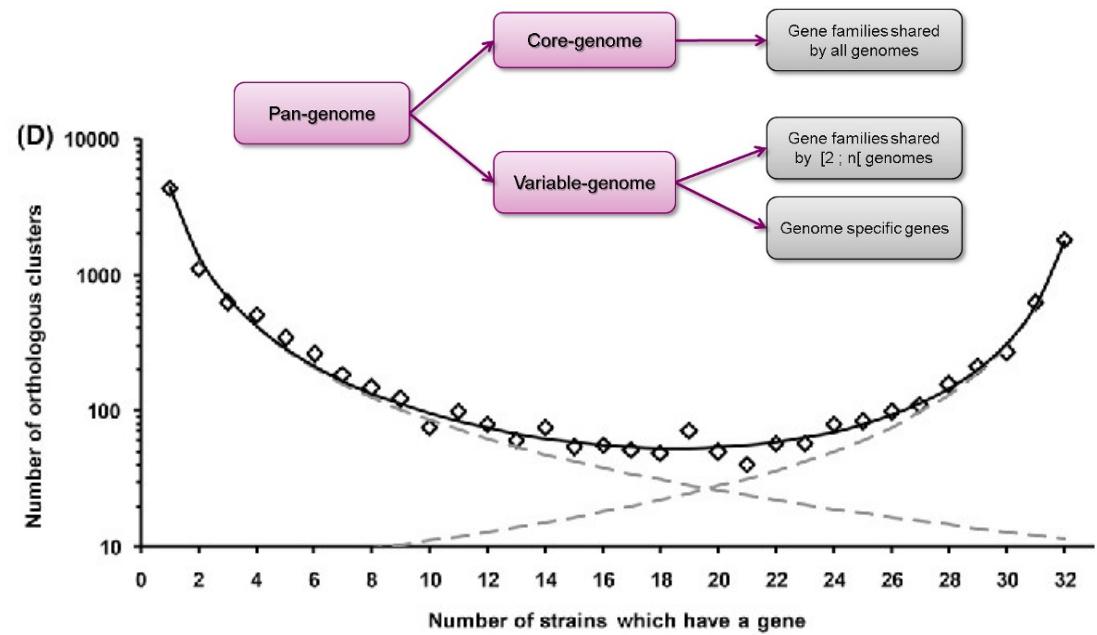
What makes Ecoli pathogenic



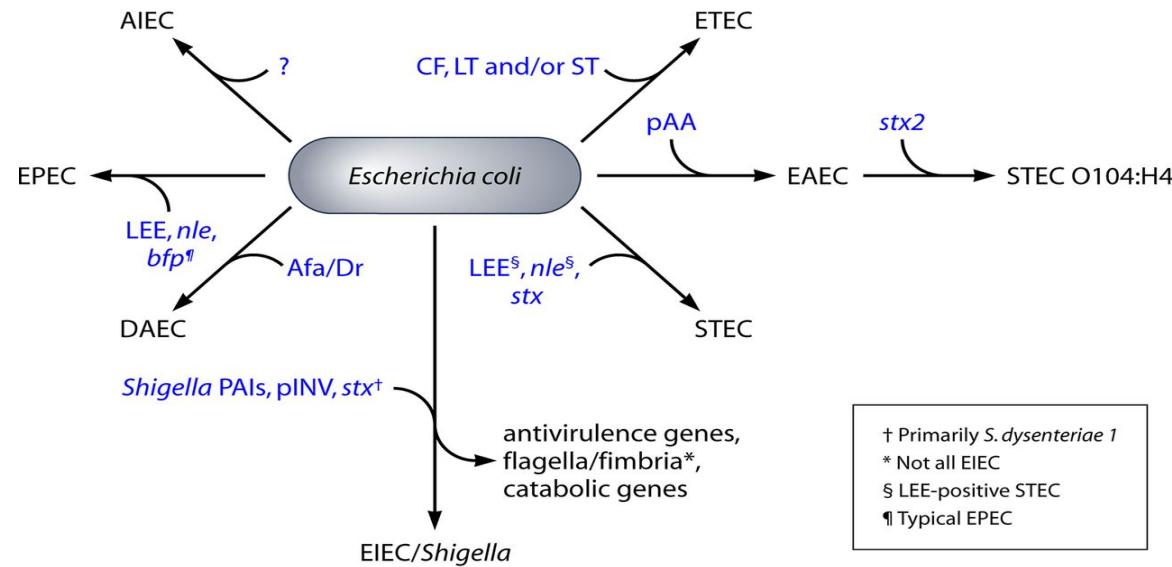
Variations in gene content
Venn diagram for 3 *E.coli*

Pan-genome approach

the entire set of genes from all strains within a clade



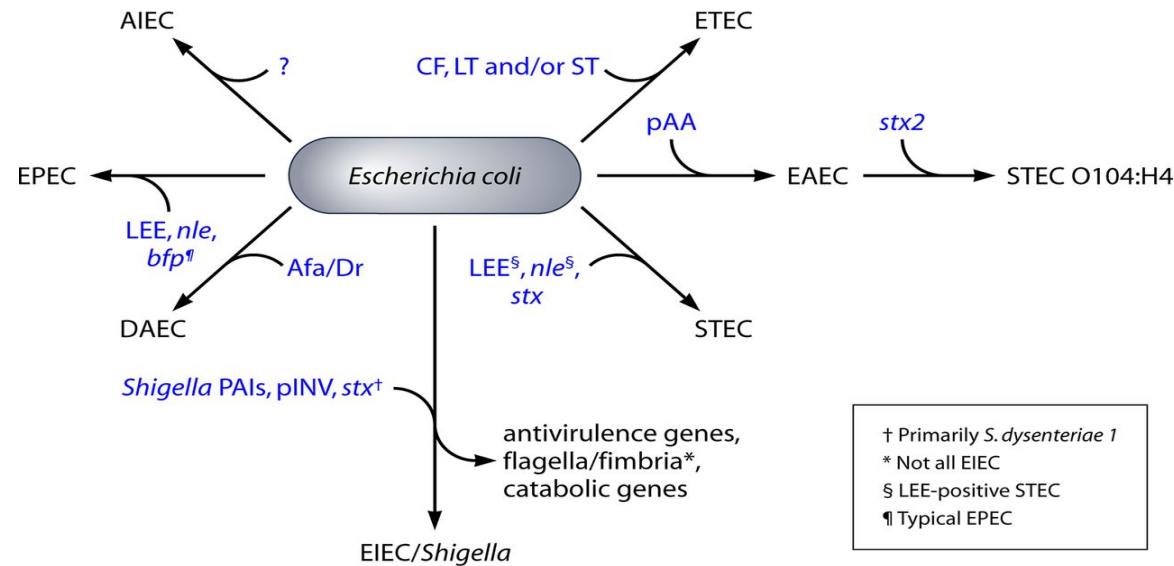
What makes Ecoli pathogenic



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



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



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.

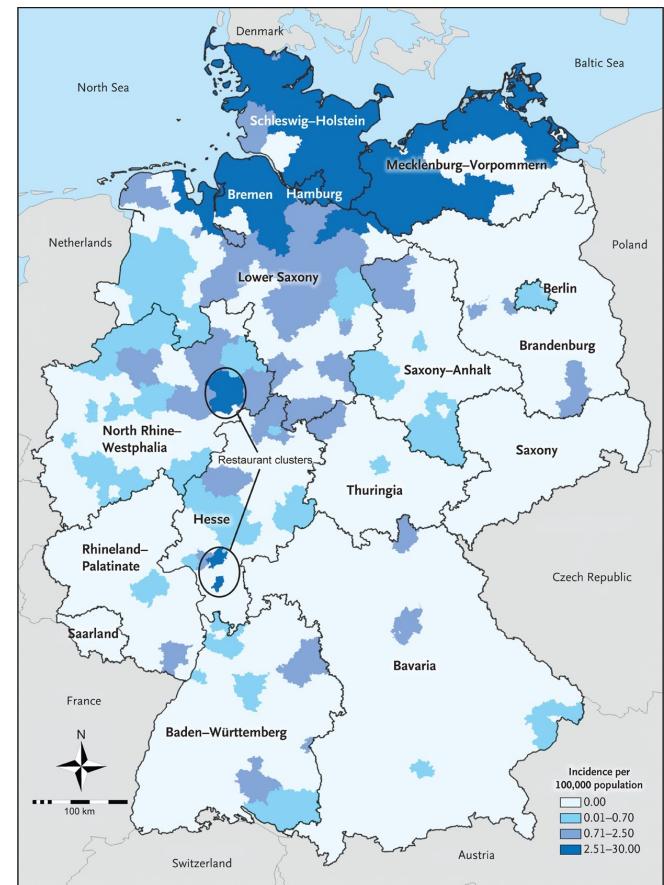
What makes Ecoli pathogenic

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).

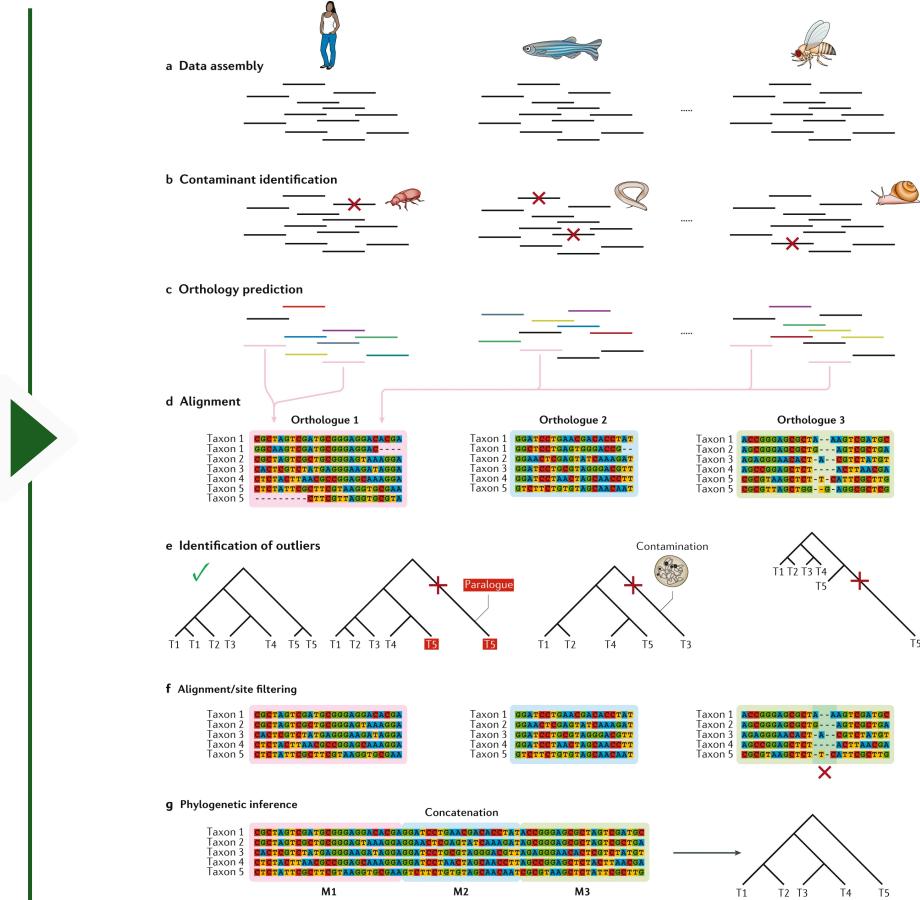
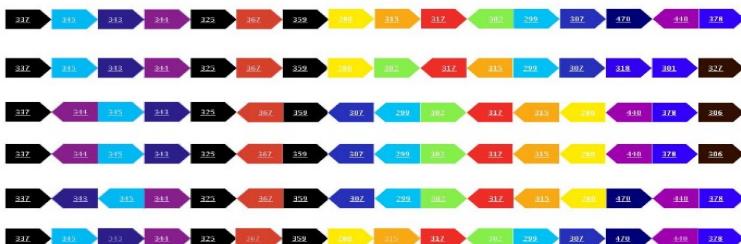


https://en.wikipedia.org/wiki/2011_Germany_E._coli_O104:H4_outbreak

How can we analyze it?



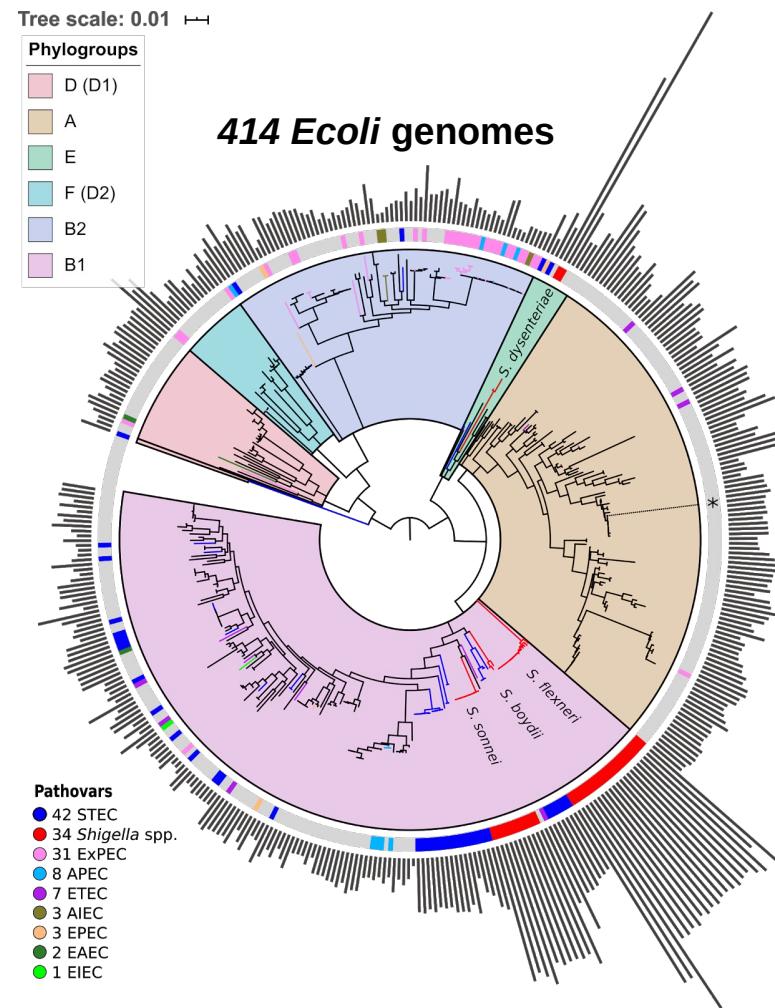
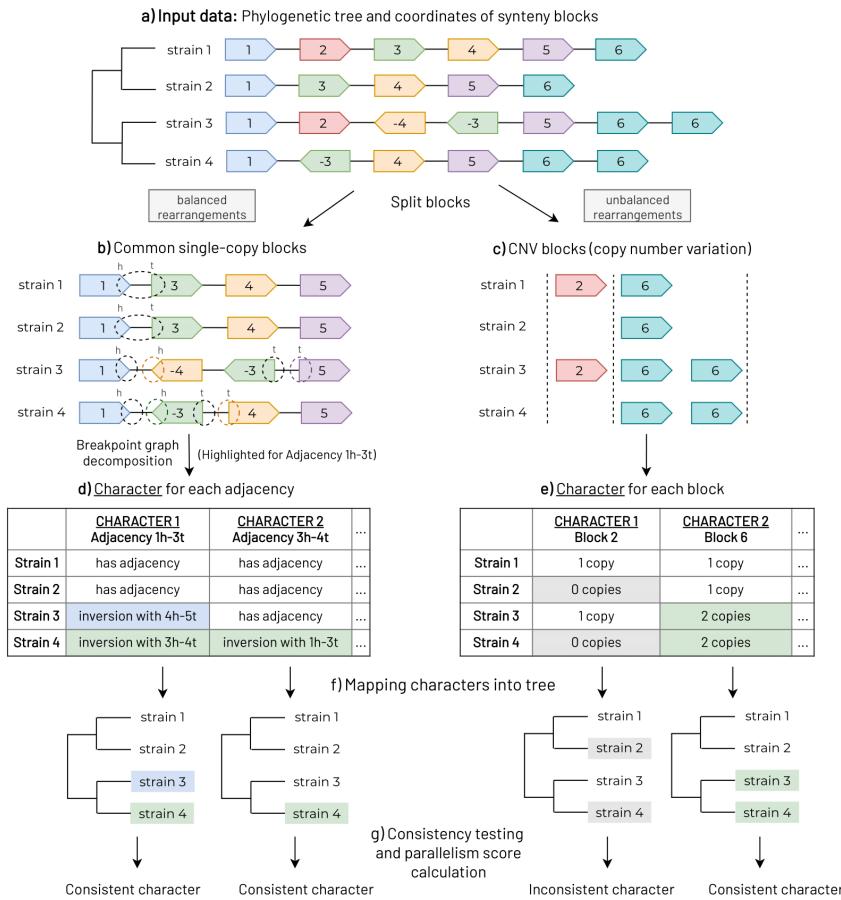
genome as an ordered set of genes and regulatory network



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>

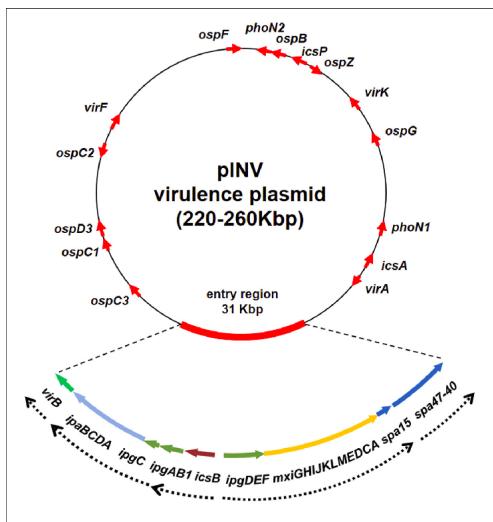
How can we analyze it?

Pipeline: Parallel Rearrangements and Breaks identification



Acquisition of invasion

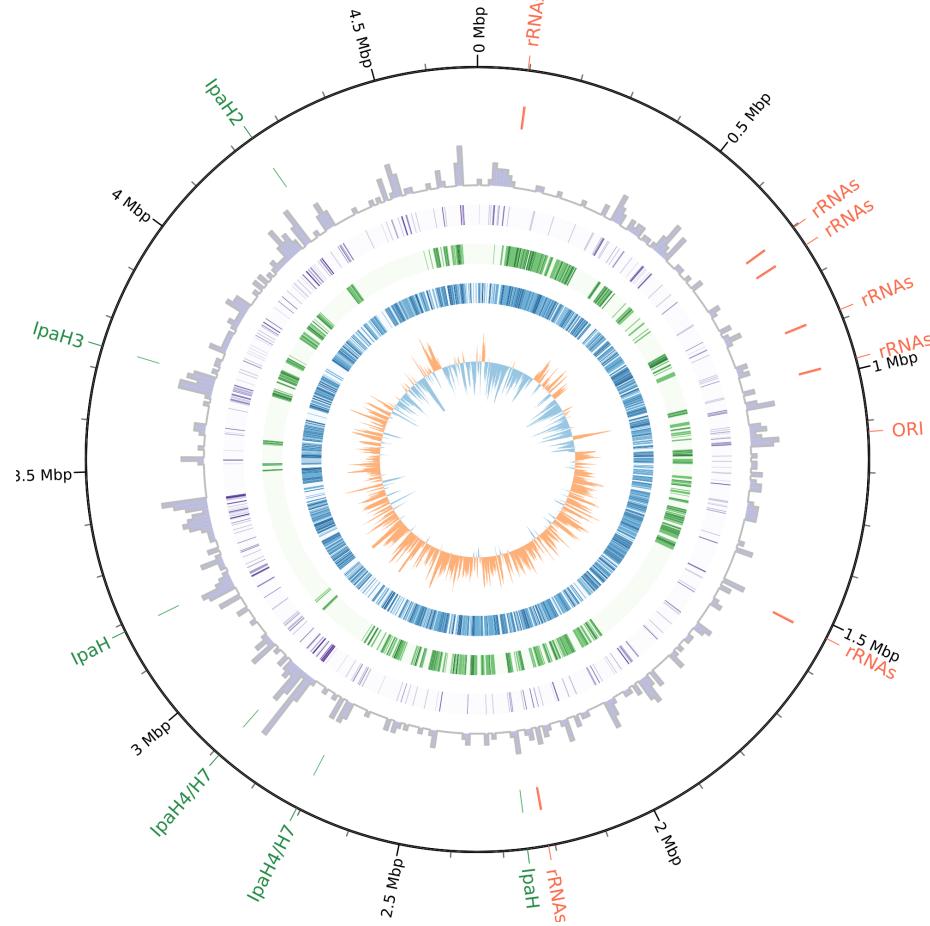
Plasmid



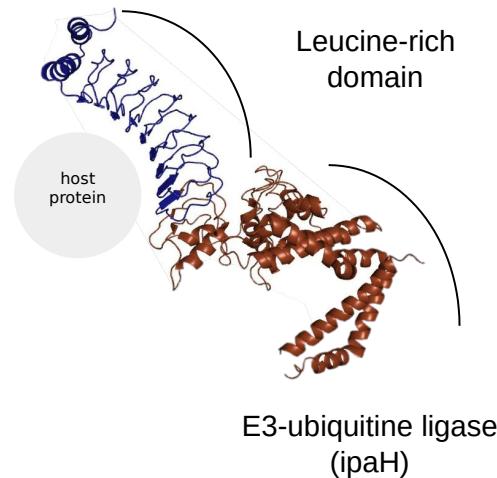
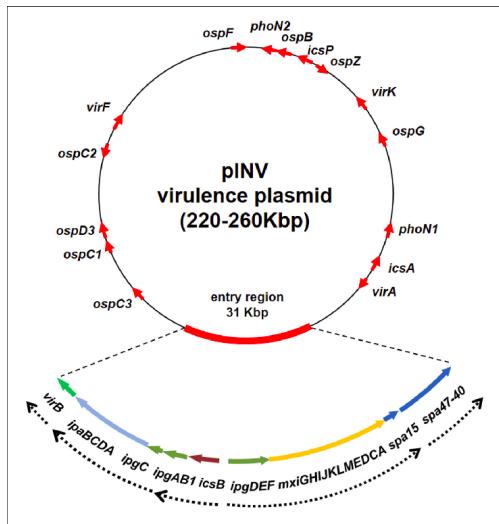
Genetic map of the pINV. The red arrows indicate major virulence determinants. Due to the variability in position and number, the *ipaH* genes are not shown.

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

Chromosome



Acquisition of invasion



WP_134797310_1/1-547 1 MLPVNPPLSTSNVFYRT.....TIDN-AHNIVLDEVVENTNSISGENRETAFTRQLQCLENSETSLDL-CLGRLRSLP-RLFDNLDEIINVNIOLM.....LPELPRALKELNASHW-LALPELFVLEYINVDSDHLISVLFR 156
 WP_000936833_1/1-568 1 MLPVNKN-FSLSONFVNNT.....ISQ-TYADVFTAWDKWEAKALPGENRNBAVSLNKLKECLINPFEQLNRENLLSSLPDNLPPQITVLEIQNALISLPELAQEYLDACDRLSTLPELPASLKHLDVDNHD-LMLPELALLEYINADNQLTMLPELPTSLEVLSVRHNLIFLFE 174
 WP_13405189_1/1-583 1 MLPPTNNN-MLLSNF-FST.....YIDTSRAENVLYHWTWKNNRIOEGEODIAQGRUVSCDLSQDPTINLDELOGLTLP-EIPPOIKSNTIKNLISLSPFAUTOLNNSYRRELPLAPQSLKLLNASHHD-LITLTLFIBLUKELVHVSQNLCPVPELLETLDVSCNOLAVLP 176
 WP_128860970_1/1-587 1 MKPAHNF-FCGFGCFCISRLVEE-QNITOYHRIWDNWAHEGAATEDRTAVRLLKICLAFP-SPAULSLERLRS-L-YLPFHQELNHSMEHLRSLELPFLUTVLKASDRLRPLALPFLVALDVSLHVRVLCYLTFLPSL-SLSALLSELTLPDUPALAKLSVGNGNLTALE 180
 CDX08027_1/1-603 1 M.....LENISSCLFFH.....STITSNPHYL-EWDDEWEKOLREECSTEAVRRLERAGLTSKCHKLDRALAESL-VLPACIKKDVCKKL-LTDLBENIKELIARDFLTHISALPHYLTEDVH-LENLPLIEDIJKLSAHEYRESLPLNLLKKLEVRINELQLES 169

WP_134797310_1/1-547 157 LPMILELLDAARVAVLAEVIFPDFPERDDHI.....I.....I.....RIFWLNQPRITATPEELIOLSSDVVNLRPEYQPRIMOTILLBOTAQPDYHOPRIPYFSMSD.....DQNTLHRPLADAVFWFPEKCDSDVSDQIWHAFFEHHHANTTSAFLDRLDTVSA 302
 WP_000936833_1/1-568 158 LPEELALDWTVTLLESLIPAVPVVR.....VHSEETI.....FRRGRERITHTPENILSLDPTCTIILEDNPFLSSRIRFELSBOTAOPDYHGPRIYFSMSD.....DQNTLHRPLADAVFWFPEKCDSDVSDQIWHAFFEHHHANTTSAFLDRLDTVSA 323
 WP_13405189_1/1-583 177 LPFELDEIISAIQISLSELSPFLPHNIHSIWADNMLDIDIPYVPEL.....RNYYDFDQNSHPEVILNLRNECSIDISDNPLSSHALDZLORLTSSPDQYHGPRIYFSMSD.....DQNTLHRPLADAVFWFPEKCDSDVSDQIWHAFFEHHHANTTSAFLDRLDTVSA 338
 WP_128860970_1/1-587 181 LPCEDDELSAFDWRLQELPPLPONLRLNIVBENOLHRLPELPDRL-..QSLYIPNNQNLNTLP2IMMLH1YADVNLYNNPLS3RTLQALORLTSSPDQYHGPRIYFSMSD.....DQNTLHRPLADAVFWFPEKCDSDVSDQIWHAFFEHHHANTTSAFLDRLDTVSA 342
 CDX08027_1/1-603 170 LPSNLKILKVAHHTLTELPPPLPRLQLLFAYSNRQLNPNQDENTIMRRFFYRENQQTITETNLFRLDPHTIEIANPLDQTLFLIBOTSVPNENGPOFRLSLEQNLHSRHIVRITEDQOFOIPLPPLFTEAVAWFPEADRRREVSTOWTSFESTEENRAFCAFLLDRLDTVSA 358

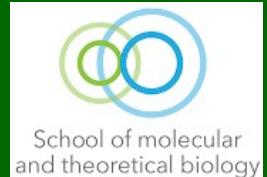
WP_134797310_1/1-547 303 NTSGFREQVAAWLRLBASAEIQQSFAVAADATEESEDVRVALTNLRLTLLVHQA-BGLFD-DTGALLSLGREMPFRLIELEDIARDKVPTLHFVDEIEVYLAFTOMLAELCQLSTAVKEMRFYGVSGVVTANDLRTAEAMVRSEEEEFIDWFSLWGPRWHAVLKRTEDRWAAEEDQYEMLENEYPS 491
 WP_000936833_1/1-568 324 NTSGFREQVAAWLRLBASAEIQQSFAVAADATEESEDVRVALTNLRLTLLVHQA-BGLFD-DTGALLSLGREMPFRLIELEDIARDKVPTLHFVDEIEVYLAFTOMLAELCQLSTAVKEMRFYGVSGVVTANDLRTAEAMVRSEEEEFIDWFSLWGPRWHAVLKRTEDRWAAEEDQYEMLENEYPS 512
 WP_13405189_1/1-583 339 NTSGFREQVAAWLRLBASAEIQQSFAVAADATEESEDVRVALTNLRLTLLVHQA-BGLFD-DTGALLSLGREMPFRLIELEDIARDKVPTLHFVDEIEVYLAFTOMLAELCQLSTAVKEMRFYGVSGVVTANDLRTAEAMVRSEEEEFIDWFSLWGPRWHAVLKRTEDRWAAEEDQYEMLENEYPS 527
 WP_128860970_1/1-587 343 NTSGFREQVAAWLRLBASAEIQQSFAVAADATEESEDVRVALTNLRLTLLVHQA-BGLFD-DTGALLSLGREMPFRLIELEDIARDKVPTLHFVDEIEVYLAFTOMLAELCQLSTAVKEMRFYGVSGVVTANDLRTAEAMVRSEEEEFIDWFSLWGPRWHAVLKRTEDRWAAEEDQYEMLENEYPS 531
 CDX08027_1/1-603 359 NTSGFREQVAAWLRLBASAEIQQSFAVAADATEESEDVRVALTNLRLTLLVHQA-BGLFD-DTGALLSLGREMPFRLIELEDIARDKVPTLHFVDEIEVYLAFTOMLAELCQLSTAVKEMRFYGVSGVVTANDLRTAEAMVRSEEEEFIDWFSLWGPRWHAVLKRTEDRWAAEEDQYEMLENEYPS 547

WP_134797310_1/1-547 402 RVADPRLKAQGLGDAADEAREAGAQVMRETEQDQYRQLTDEVLALFLPENGSOLHHH
 WP_000936833_1/1-568 513 RVADPRLKAQGLGDAADEAREAGAQVMRETEQDQYRQLTDEVLALFLPENGSOLHHH
 WP_13405189_1/1-583 528 RVADPRLKAQGLGDAADEAREAGAQVMRETEQDQYRQLTDEVLALFLPENGSOLHHH
 WP_128860970_1/1-587 532 RVADPRLKAQGLGDAADEAREAGAQVMRETEQDQYRQLTDEVLALFLPENGSOLHHH
 CDX08027_1/1-603 548 RVADPRLKAQGLGDAADEAREAGAQVMRETEQDQYRQLTDEVLALFLPENGSOLHHH

T3SS effector E3 ubiquitin-protein ligase IpaH

547
568
583
587
603

Mess with gene names



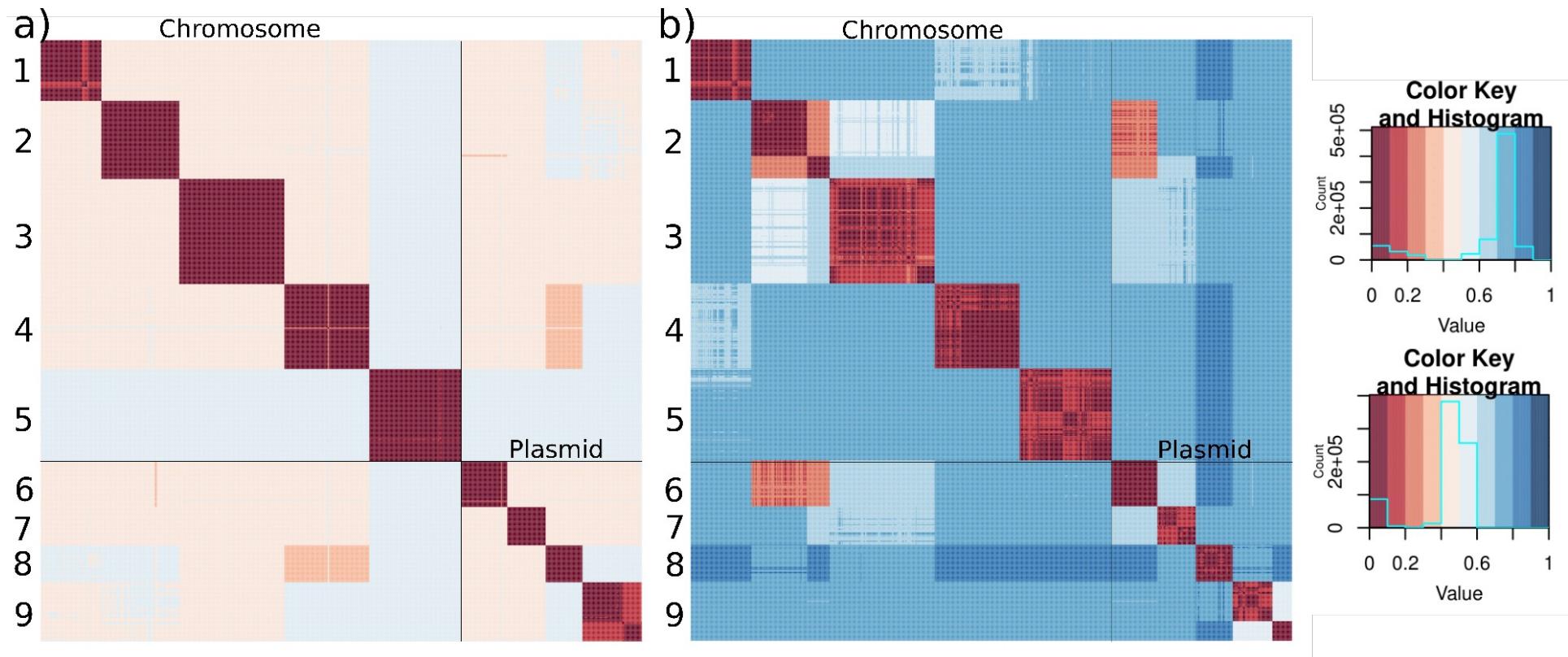
Gene	Size (bp)	<i>S. flexneri</i>	<i>S. flexneri</i>	<i>S. flexneri</i>	<i>S. dysent.</i>	<i>S. sonnei</i>	<i>S. boydii</i>
		8401	301	2457T	197	046	227
<i>ipaH</i> a1	1764	<i>ipaH</i> 7	<i>ipaH</i> 1	<i>ipaH</i> 1 (fs)	<i>ipaH</i> 1	<i>ipaH</i> 1	<i>ipaH</i> 6
<i>ipaH</i> a2	1764	-	-	-	<i>ipaH</i> 2 (IS)	-	-
<i>ipaH</i> b1	1827	<i>ipaH</i> 1	<i>ipaH</i> 2 (fs)	<i>ipaH</i> 2	<i>ipaH</i> 6	<i>ipaH</i> 5	<i>ipaH</i> 1
<i>ipaH</i> b2	1827	<i>ipaH</i> 6	<i>ipaH</i> 7	<i>ipaH</i> 7	-	-	-
<i>ipaH</i> c	1716	<i>ipaH</i> 2	<i>ipaH</i> 3	-	<i>ipaH</i> 4	<i>ipaH</i> 3	<i>ipaH</i> 4
<i>ipaH</i> d	1752	<i>ipaH</i> 3	<i>ipaH</i> 4	<i>ipaH</i> 4	<i>ipaH</i> 5	<i>ipaH</i> 2	<i>ipaH</i> 5
<i>ipaH</i> e1	1644	<i>ipaH</i> 4	<i>ipaH</i> 5 (IS)	<i>ipaH</i> 5	<i>ipaH</i> 3 (IS)	<i>ipaH</i> 4	<i>ipaH</i> 2
<i>ipaH</i> e2	1644	<i>ipaH</i> 5	<i>ipaH</i> 6 (IS)	<i>ipaH</i> 6 (fs)	-	-	<i>ipaH</i> 3

^aNames 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

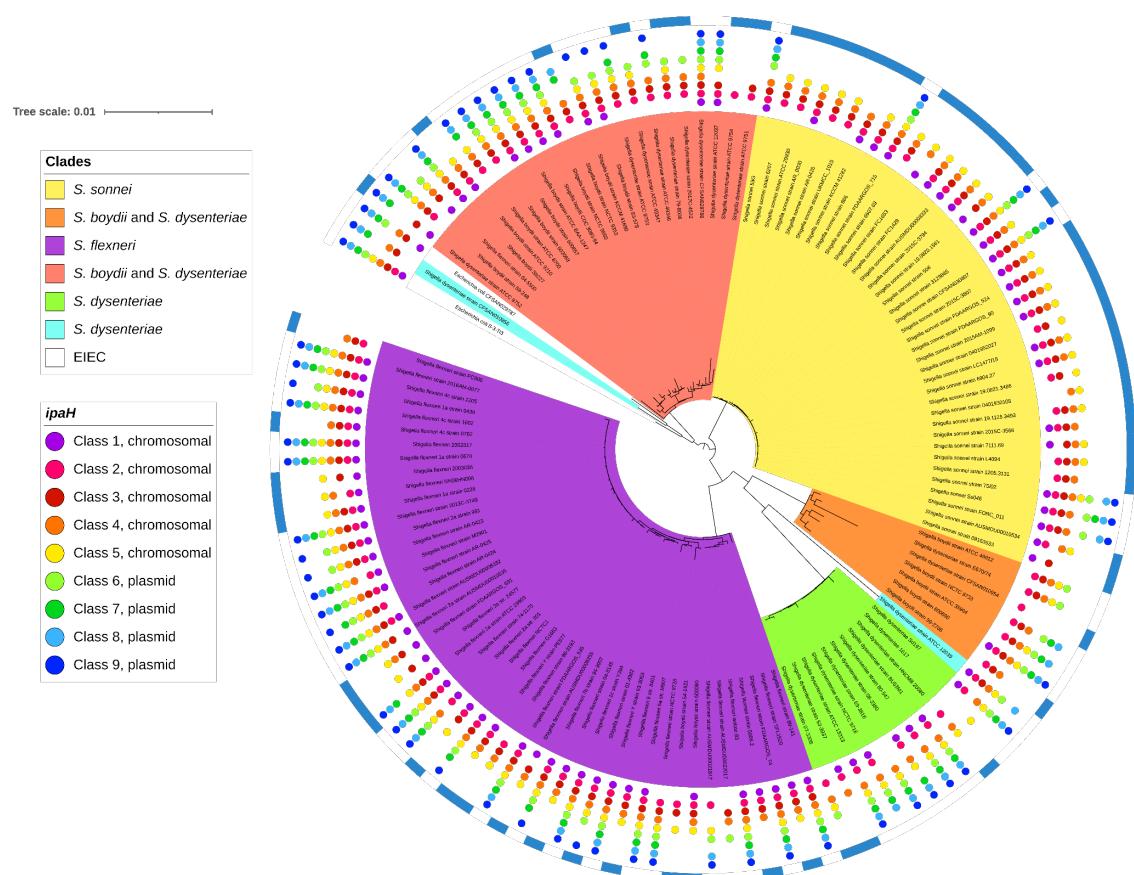
Now it's solved

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



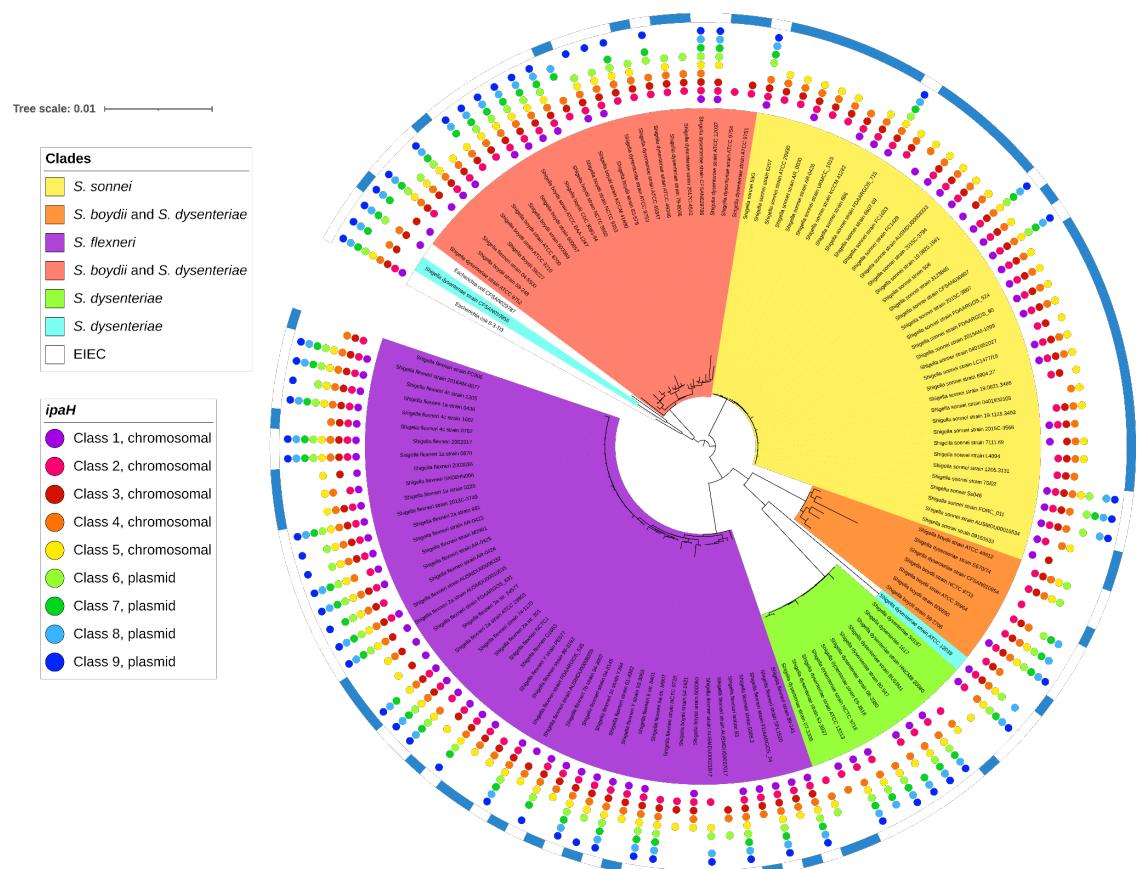
And classified!

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

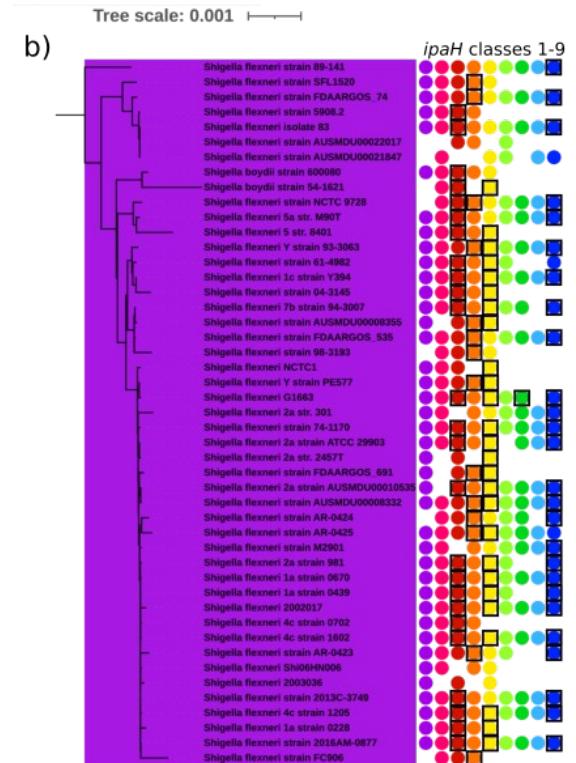


And classified!

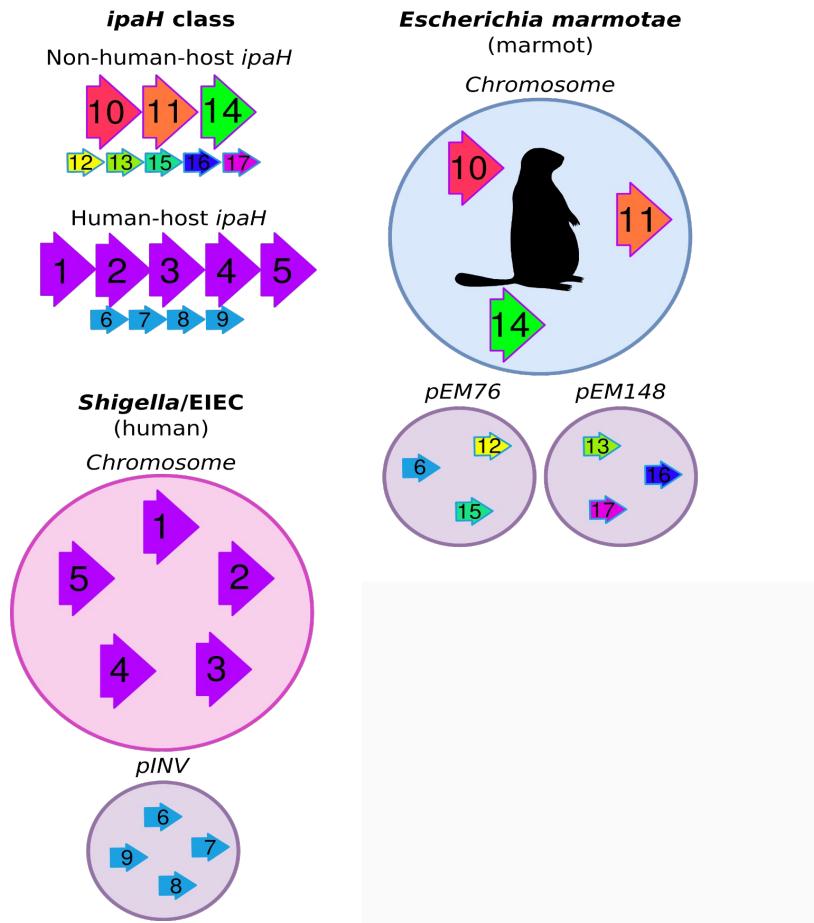
Composition of *virulence genes* in human-host pathogenic *E.coli*



And their copies



But wait... what's about animals?

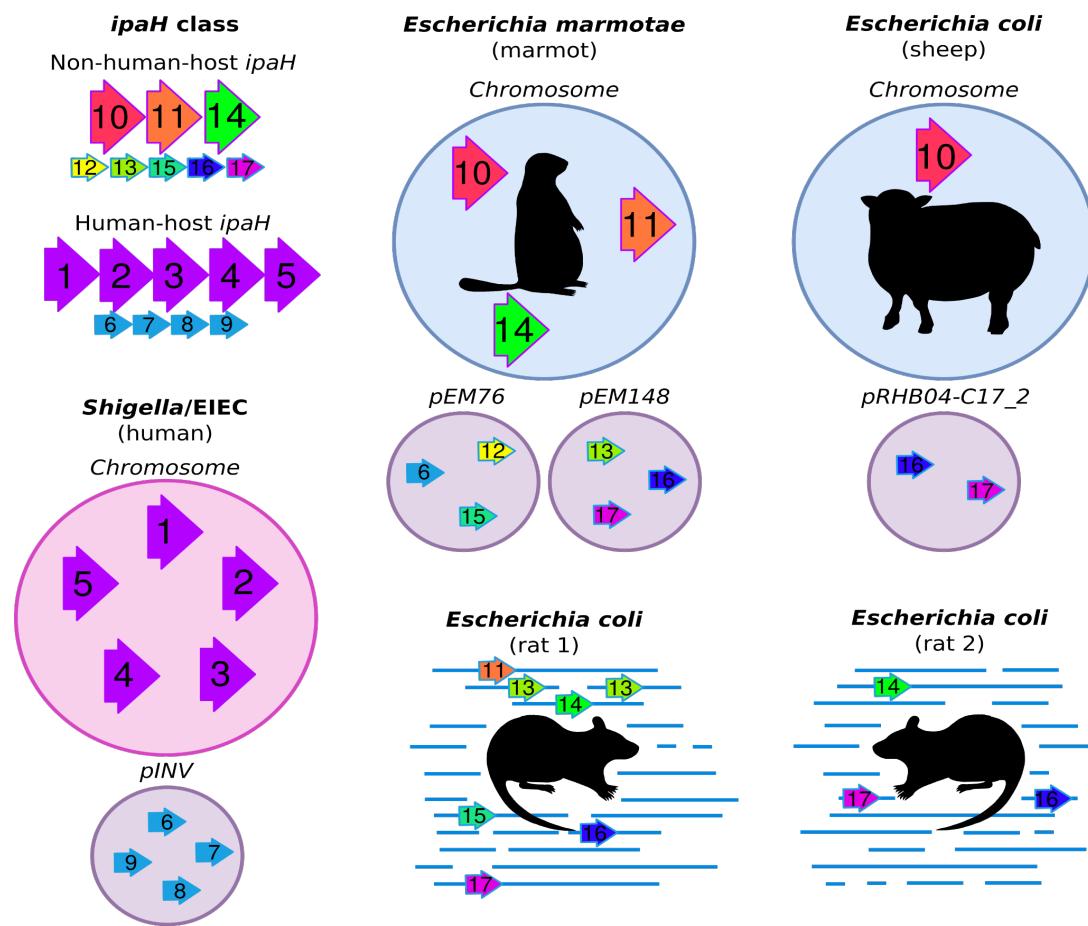


Genomic and molecular characterisation of *Escherichia marmotae* from wild rodents in Qinghai-Tibet plateau as a potential pathogen. Sci Rep, 2019

Alignment of C-terminal domains

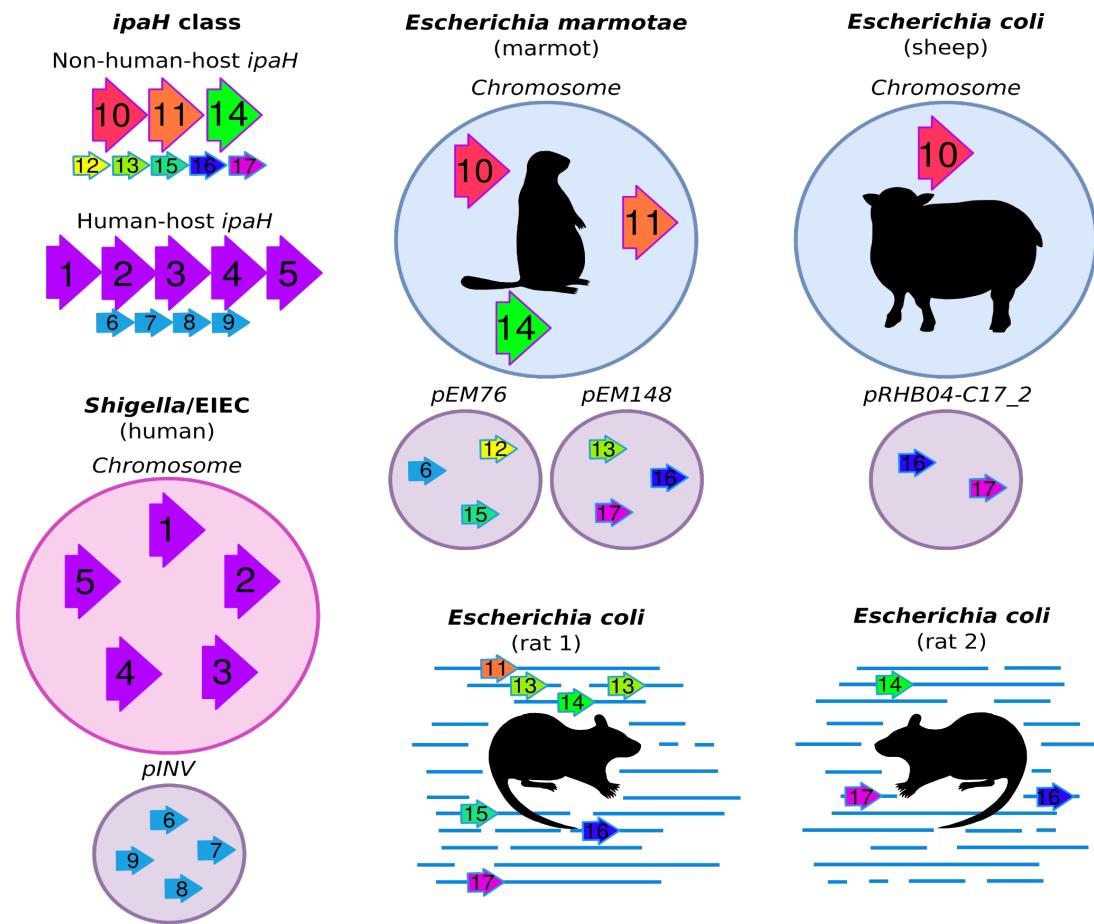
ShigellaCDomain	GQQ T L H RPLADAVTAWFPE N KQSDVSQIWHAF E HEEHANT T SAFLDRL S D T V A R N TS G
NonHumanCDomain1	GQQHTSVRPLPEAVAAWF P Q N KQSDVSQIWLAFREEHANT T SAFLDRLADTV S ARNAQ G
NonHumanCDomain2	GQQHTPVRSLP E AVAAWF P ESLRSEVSQRWG A FTDEENAA T SAFLDRLADTV T ARNAPG
ShigellaCDomain	F R QVAAWLEKLSASAE L RQQSFAVAADATE E SCEDRV V ALT T NNL R KTLLVHQASEGLFDN
NonHumanCDomain1	F R QVASF L EKLSTS A E L RQQSFAVAADATE E SCEDRV V ALT T NNL R KTLLVHQASEGLFDN
NonHumanCDomain2	FAQQVSEF L EKLRS E ALRQQCF A VAADAT R SCEDRV V ALT T NNL Q KTYRVHQASEGFDS
ShigellaCDomain	D T GALLSLGREM R LE I LD I ARDKV V RTLHF V DEIEVYLA F QTMLAEK L Q L STAV E MRF
NonHumanCDomain1	D T GALLSLGREM R LE I LD I ARDKV V RTLRF V DEIEVYLA F QTMLAEK L Q L STAV E MRF
NonHumanCDomain2	DLTG L LSLGREM Y RL V LE E IAREK V RTLHF V DEIEVYLA F QTMLAEK L Q L STAV E MRF
ShigellaCDomain	YGVSGV T AN D LRTA E AMV S RE E NE F T D WF S WGPWHAVLKRTEAD R W A Q E E O KYEM L
NonHumanCDomain1	YGVSGV T EN D LRTA E AMV S RE E NE F T D WF A WGPWHAVLKRTEAD R W A Q E E O KYEM L
NonHumanCDomain2	YGVSGV T ED D LSS S ALVRVL S RE E FAE W FW R SPW H AVLKRTEA R W A RE E EKKYEM L
ShigellaCDomain	NEYPQRVADRLKA S GLSGD A DA E RE A GAQVM R E T E Q QIYRQLT D EV L ALRL S ENG C QLHH
NonHumanCDomain1	NEYPQRVADRLKA S GLSD D DA E RE A GAQVM R E T E Q QIYRQLT D EV L ALRLP G EN R LLHH
NonHumanCDomain2	REYPORLAERLSA G LGSGD G DA E RE A GV R ME E ET E YIYR L TEEVSGERLAENRAR X X

Oh... there are many others!



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

But how many?..



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

