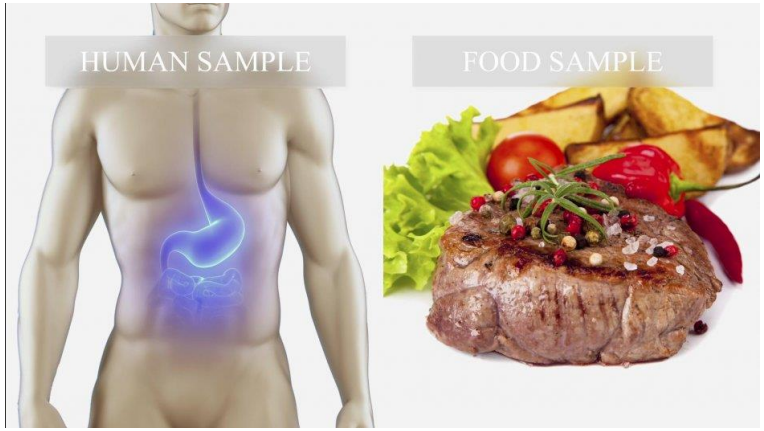
Bacterial strain typing:
characterizing a number of strains in detail and ascertaining whether they are derived from a single parental organism, is **a way to identify bacteria at the strain level and to uncover the genetic diversity underlying important phenotypic characteristics**



https://multimedia.efsa.europa.eu/moleculartyping/index.htm

https://multimedia.efsa.europa.eu/moleculartyping/index.htm

Uses of Typing methods

1.Epidemiologic usefulness:

1. To investigate the source of different strains in outbreak situations, to check the possibility of laboratory cross-contamination etc.
2. To determine whether the second episode of disease is due to a previously isolated strain or to a newly infecting strain, and
3. To determine whether an infection is caused by more than one strain of the organism

2.As an important infection control tool: to monitor the prevalence of certain strains within a healthcare institution or to investigate if a cluster of infections are unrelated or part of an outbreak

A: Phenotypic Typing Methods: Phenotyping techniques detect characteristics expressed by the microorganism. They are based on biochemical, antigenic or susceptibility (to phages or antimicrobial agents) properties of the organism.

1. **Biotyping:** Based on metabolic characteristics expressed by an isolate; referred to as 'biotypes'
2. **Serotyping:** Based on antigenic determinants expressed by the microorganism; referred to as 'serotypes'. O-antigen, K-antigen H-antigen
3. **Phage typing:** Based on the pattern of resistance or susceptibility to a standard set of phages; referred to as 'phage types'.
4. **Resistotyping**: Based on the resistance or susceptibility of the isolates against a set of arbitrarily chosen chemical agents
5. **Bacteriocin typing:** Based on the susceptibility to a set of bacterial peptides (bacteriocin) produced by certain bacteria.

https://microbeonline.com/bacterial-typing-methods-aim-attributes-and-types/

B: Molecular Typing Methods: Molecular techniques are based on the analysis of chromosomal or extrachromosomal genetic elements (such as plasmid) of the organism.
In recent years a plethora of molecular-typing methods have appeared based on the analysis of fragments of DNA split by specific restriction enzymes. Their discriminatory powers and complexity vary widely. With the advancement of molecular epidemiology, a single machine is now able to generate a wealth of information needed to detect, monitor and control new threats such as drug resistance and the emergence of new pathogens. With the widespread use of molecular typing methods, phenotyping typing methods are now being obsolete.

- **Commonly used molecular tying methods are as follows:**

1. Amplified fragment length polymorphism (AFLP)
2. Enterobacterial repetitive intergenic consensus (ERIC)-PCR
3. Multilocus sequence typing (MLST)
4. Multilocus variable-number tandem repeat analysis (MLVA)
5. Pulsed-field gel electrophoresis (PFGE)
6. PCR Ribotyping (agarose based or sequence-based)
7. Repetitive element PCR typing
8. Restriction endonuclease analysis (REA)
9. Surface layer protein A gene sequence typing (slpAST)
10. Whole-genome sequencing (WGS)

Bacterial typing methods

TYPING

FINGERPRINTING

The **discriminating power** of a method defines the level of correlation of strains in an epidemiological investigation

Typing methods: pattern of bands or DNA sequences

Typing methods: bands or sequences

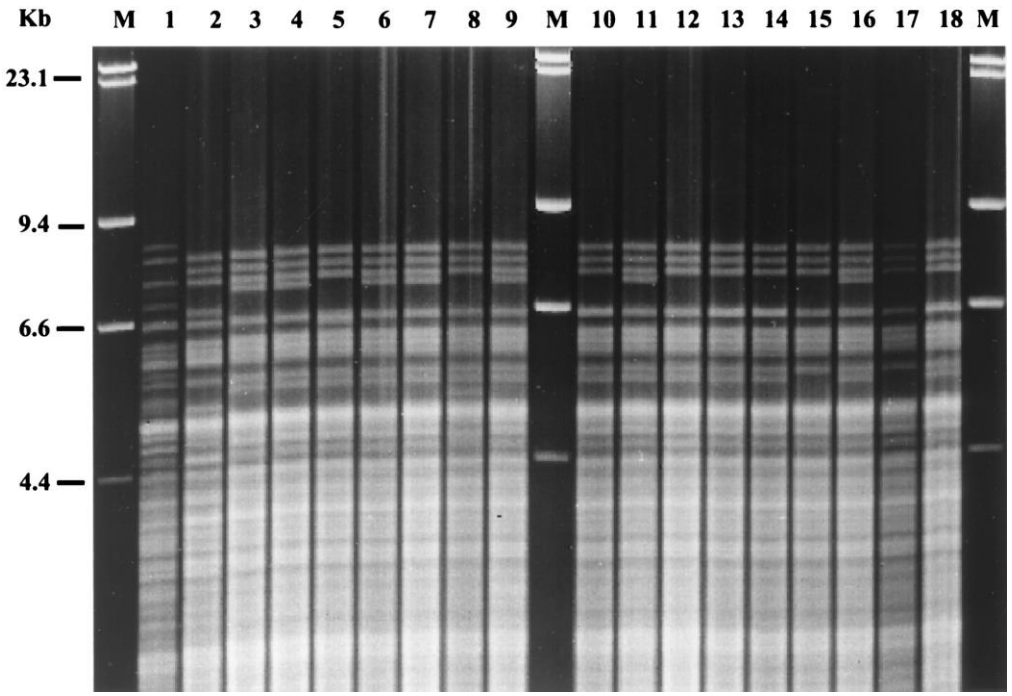
REA: Restriction endonuclease analysis
Rep-PCR: Repetitive element PCR typing
ERIC-PCR: Enterobacterial repetitive intergenic consensus
AFLP: Amplified fragment length polymorphism
PFGE: Pulsed-Field Gel Electrophoresis
MLVA :Multiple-Locus Variable-number tandem-repeat Analysis
PCR Ribotyping (agarose based or sequence-based)

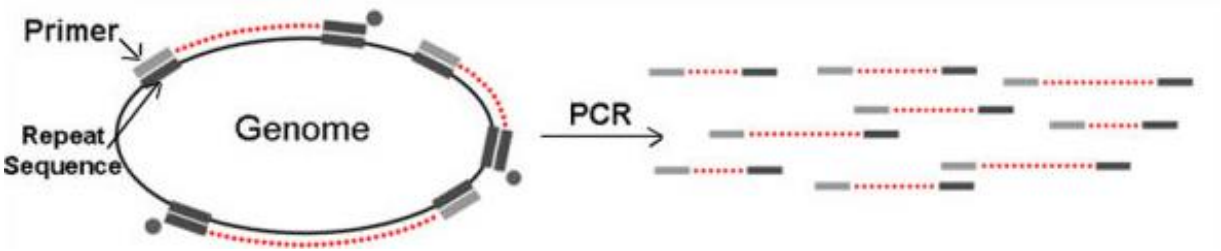
MLST: MultiLocus Sequence Typing
slpAST:Surface layer protein A gene sequence typing
WGS: Whole Genome Sequencing

REA: Restriction endonuclease analysis

Restriction endonuclease analyses (REAs) **constitute the only inexpensive molecular approach capable of typing and characterizing all strains based on their entire genome**. However, the application of this method is limited by the need for time-consuming and labor-intensive procedures.

Digestion of a preparation of genomic DNA and gel electrophoresis.

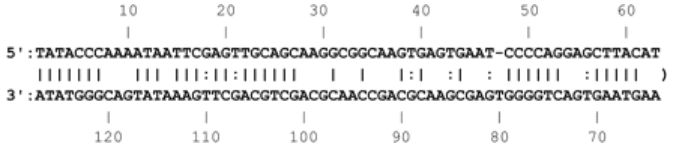


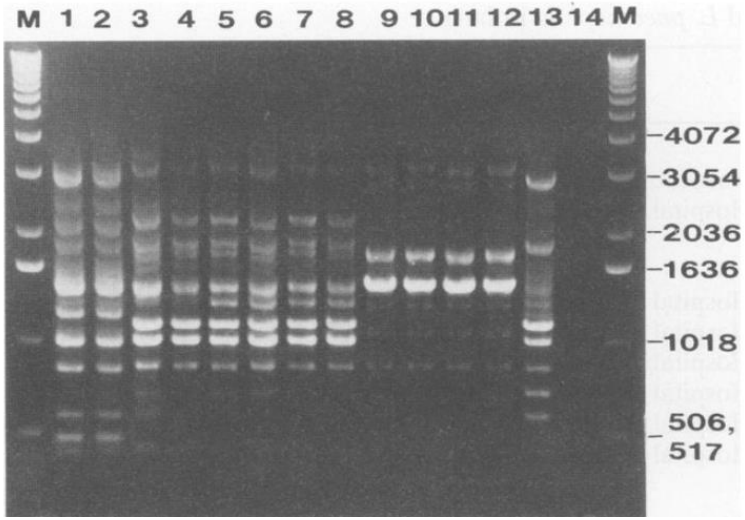


Rep-PCR: **R**epetitive ***e***xtragenic **p**alindromic PCR

It is based on the presence of regions of highly conserved and randomly interdispersed repetitive DNA in the genome of a bacterium. The number and location of these regions varies within strains that show differences at the genomic level. The primers are found within the regions and are then generated a series of fragments that will have for each strain length and a defined number

Enterobacterial **r**epetitive **i**ntergenic **c**onsensus (ERIC) sequences are 127-bp imperfect palindromes that occur in multiple copies in the genomes of enteric bacteria and vibrios



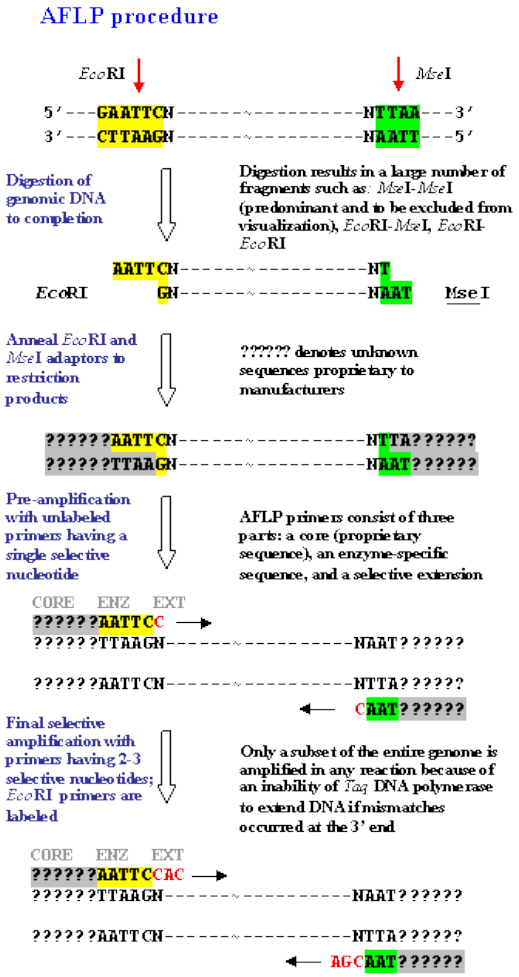


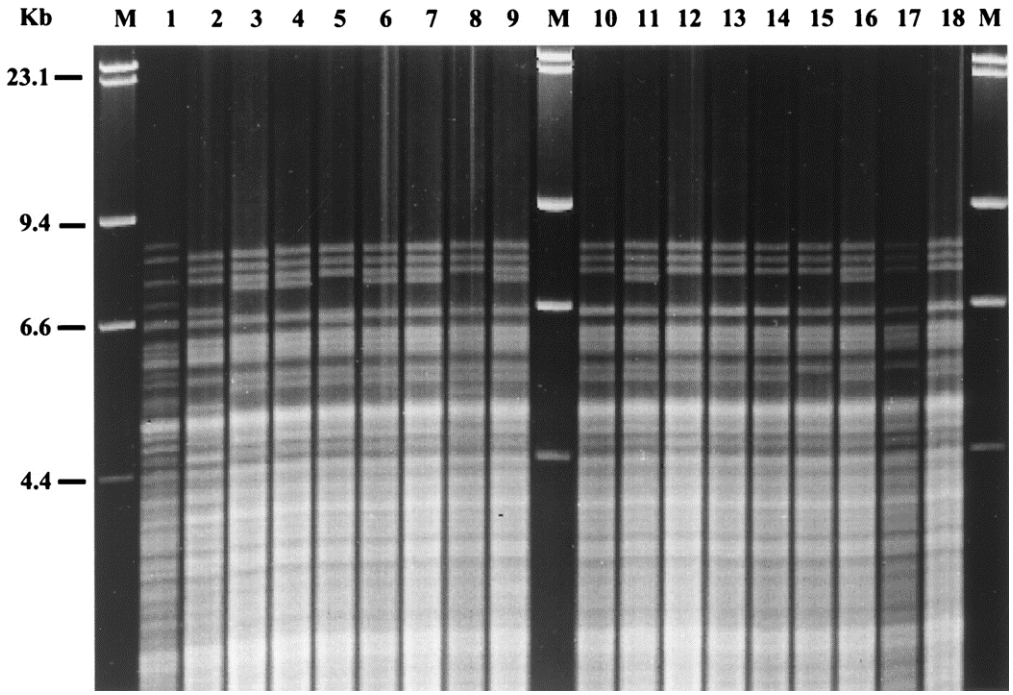
Genomic DNA of different strains were amplified by PCR using specific primers based on repeated rep sequences.

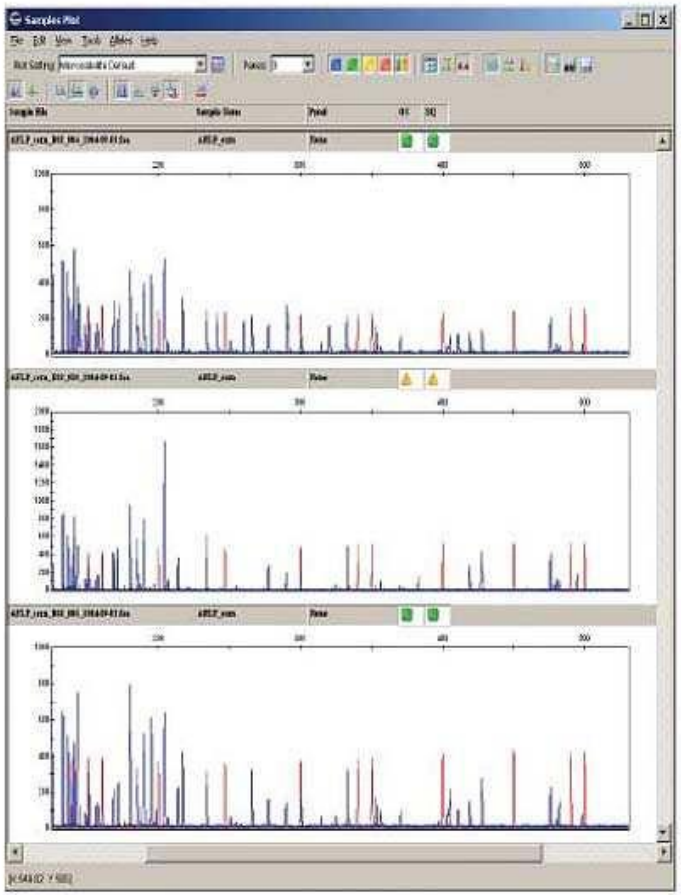
The products obtained by PCR are then separated on agarose gel and allow to identify similar / different strains
e.g. (1,2)(3,4)(5-8)(9-12)13.

Amplified fragment length polymorphism (AFLP) is a PCR-based technique that uses selective amplification of a subset of digested DNA fragments to generate and compare unique fingerprints for genomes of interest. The power of this method relies mainly in that it does not require prior information regarding the targeted genome, as well as in its high reproducibility and sensitivity for detecting polymorphism at the level of DNA sequence.

(a) restriction of genomic DNA and ligation of adaptors (most often performed together) to restricted fragments; (i.e. MseI and EcoRI digestions and MseI-adaptor pair and EcoRI-adaptor pair)
(b) preselective PCR amplification of a subset of the restricted fragments;
(c) selective PCR amplification, reducing further fragment number; (i.e EcoRI primers: 5- GACTGCGTACCAATTCXXX where X stands for selective nucleotides)
(d) electrophoretic separation of amplified DNA fragments (capillary electrophoresis in a Sanger sequencer);
(e) scoring and interpretation of the data




Kb
M 1 2 3 4 5 6 7 8 9 M 10 11 12 13 14 15 16 17 18 M
23.1
9.4
6.6
4.4

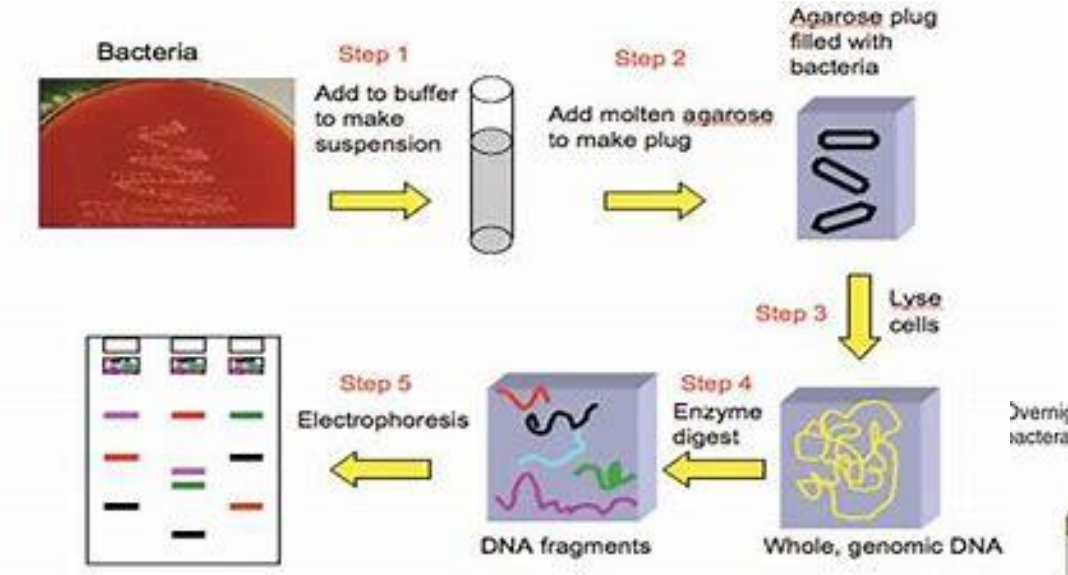


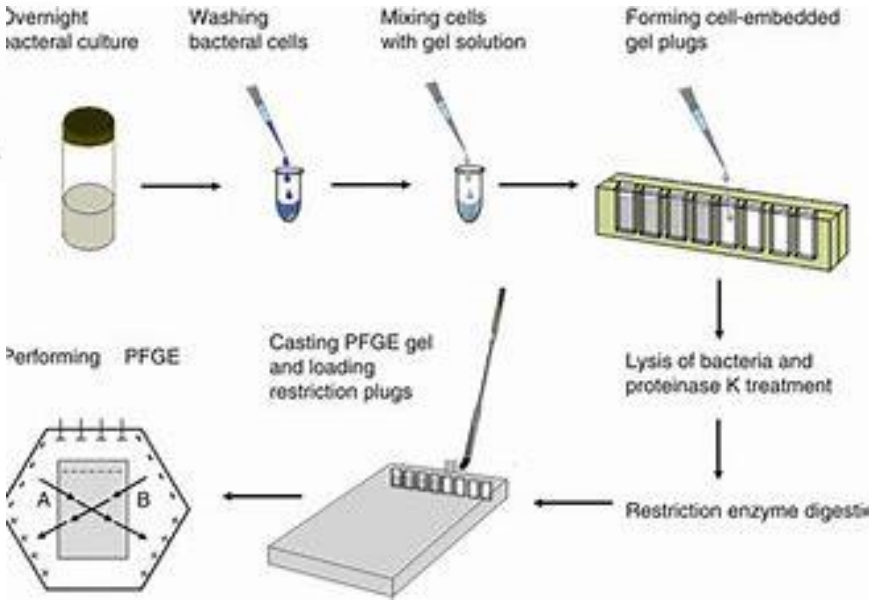
Pulsed-Field Gel Electrophoresis:
This is a widely used technique for analyzing a large amount of chromosomal DNA found in large bacterial chromosomal fragments generated by endonuclease digestion.

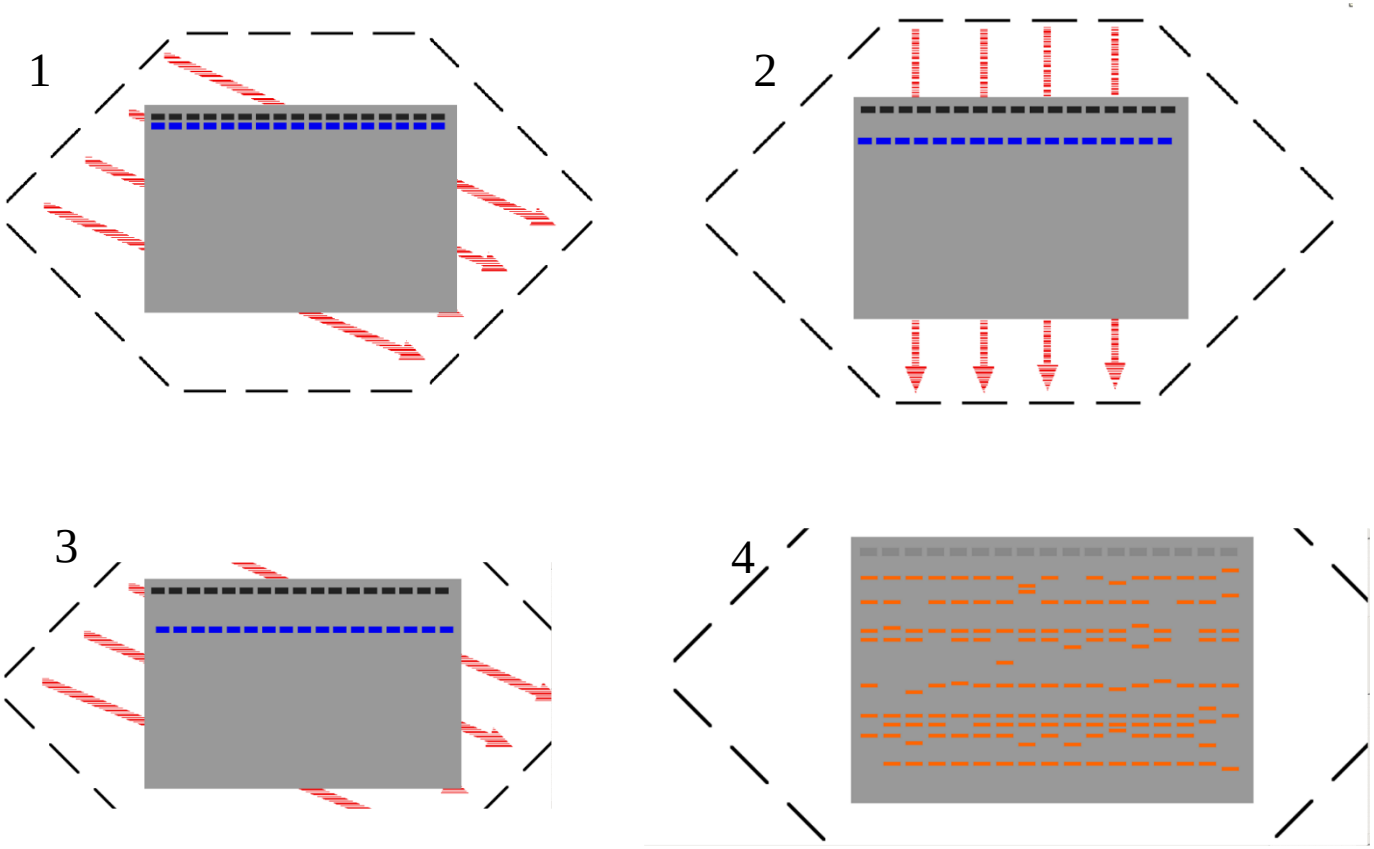
PULSE FIELD GEL ELECTROPHORESIS

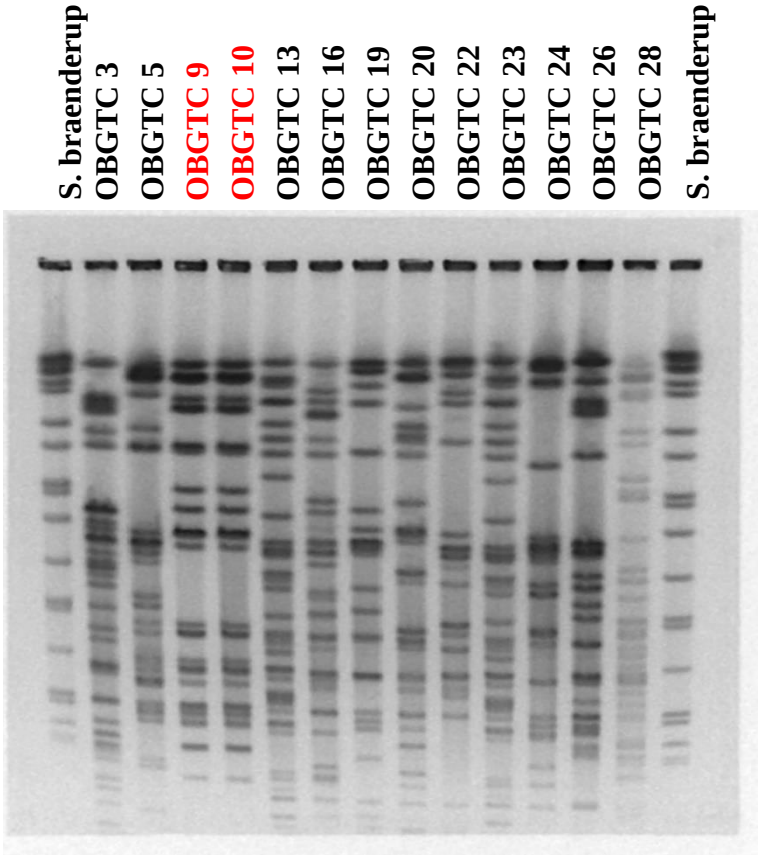
The method consists of an electrophoresis on agarose gel in which two electrical fields with different angles are applied alternately for defined periods of time (e.g. 60s).
The action of the first electric field induces the movement of DNA fragments along the direction of the field. The interruption of this field and the application of the second causes the molecules to move in the new direction.

Since for a linear long-chain molecule there is a relationship between the conformational change induced by an electric field and the length of the molecule itself, the smaller molecules will realign faster in the new electric field than the larger ones.
In this way, not only the smaller molecules are separated from the larger ones but, thanks to the different re-orientation times typical of larger fragments, also large molecules between them.


Bacteria
Step 1
Add to buffer to make suspension
Step 2
Add molten agarose to make plug
Agarose plug filled with bacteria
Step 3
Lyse cells
Step 4
Enzyme digest
Whole, genomic DNA
DNA fragments
Step 5
Electrophoresis


Overnight bacteral culture
Washing bacteral cells
Mixing cells with gel solution
Forming cell-embedded gel plugs
Lysis of bacteria and proteinase K treatment
Restriction enzyme digesti
Casting PFGE gel and loading restriction plugs
Performing PFGE
A
B


1
2
3
4



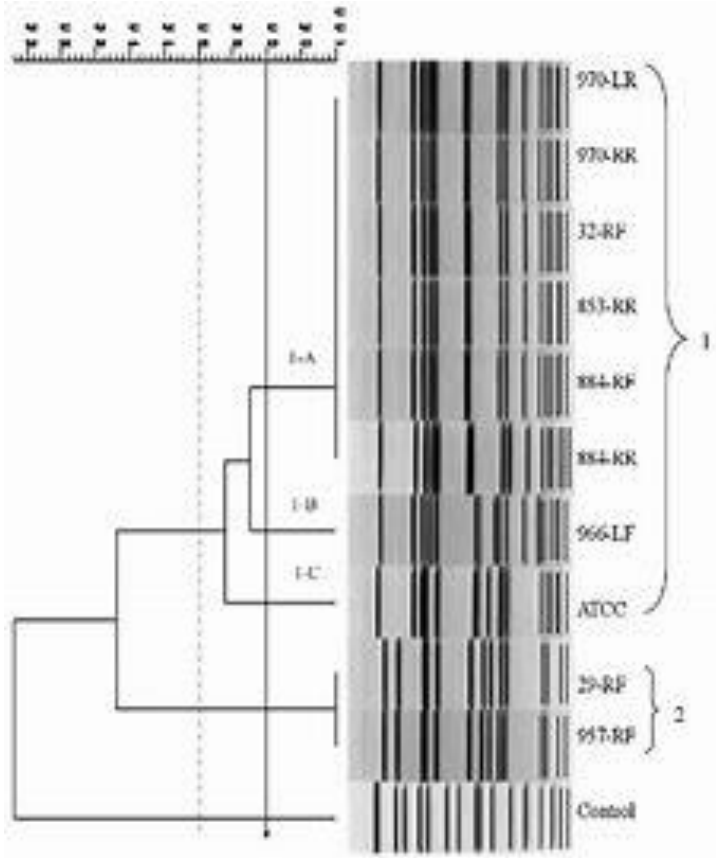
PFGE

XbaI digestion

Band pattern

Pulsotypes

Genomic DNA macrorestriction profiles of *S. maltophilia* produced by PFGE



PulseNet International

Search...

Home | International | Networks | Molecular Typing | Publications

Home / Home

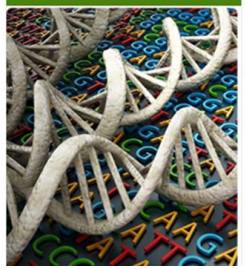
PulseNet International

- PulseNet
- About PulseNet
- What We Do
- PulseNet Methods & Protocols
  - WGS
  - PFGE
  - Protocols
  - Protocol Images
- Next Generation Technology
- PulseNet Participants

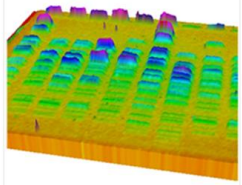
PulseNet Methods

PulseNet uses a variety of methods to subtype: *E. coli* (O157 and other Shiga toxin-producing *E. coli*), *Campylobacter*, *Listeria monocytogenes*, *Salmonella*, *Shigella*, *Vibrio cholerae*, *Vibrio parahaemolyticus*, and *Cronobacter* isolates. Pulsed-field gel electrophoresis (PFGE), multiple locus variable number tandem repeat analysis (MLVA),and whole genome sequencing (WGS) are PulseNet's main subtyping (or fingerprinting) tools.

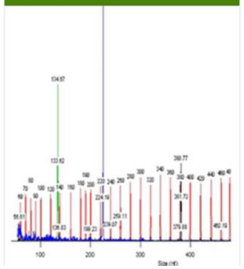
WGS



PFGE



MLVA

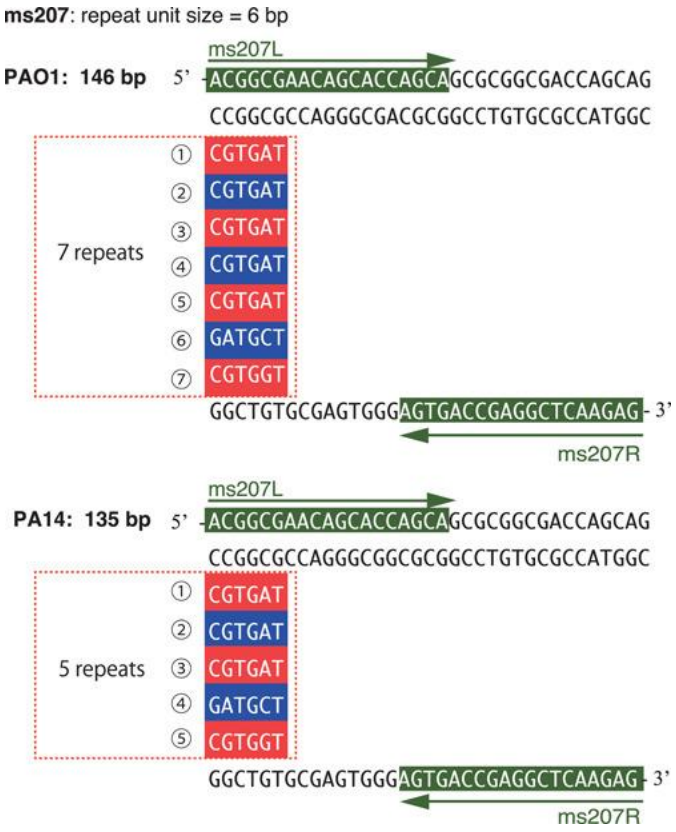


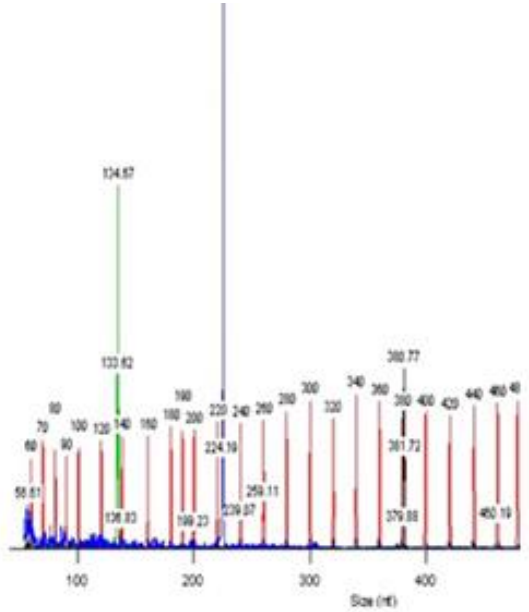
MLVA: Multiple-Locus Variable-number tandem-repeat Analysis

"The MLVA method indexes genetic variation at well defined genomic loci and produces reproducible allelic profiles that can be coded in a simple digital format.
Hence, they represent an attractive alternative to banding profile-based methods such as pulsed-field gel electrophoresis (PFGE), which requires dedicated efforts (e.g. http://www.cdc.gov/pulsenet) in order to produce fingerprinting data that are comparable across laboratories.
Indeed, to be useful to surveillance networks and for global epidemiology, a genotyping method has to be technically accessible, reproducible and to yield easily portable data. In addition, electronic databases that are made accessible through the Internet can render exchange and comparison of data among laboratories very effective for local, national, and international surveillance."

Guigon G, Cheval J, Cahuzac R, Brisse S. MLVA-NET – a standardised web database for bacterial genotyping and surveillance. Euro Surveill. 2008;13(19):pii=18863.

MLVA



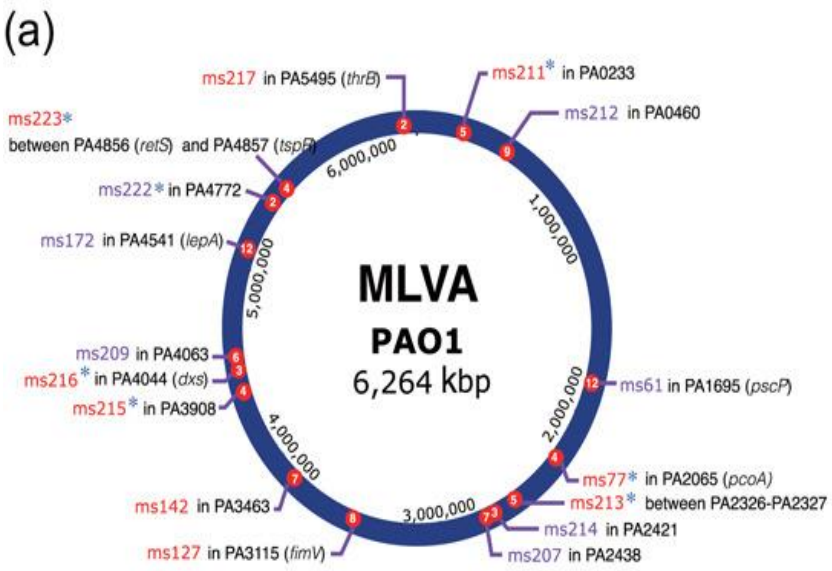


No of repeats

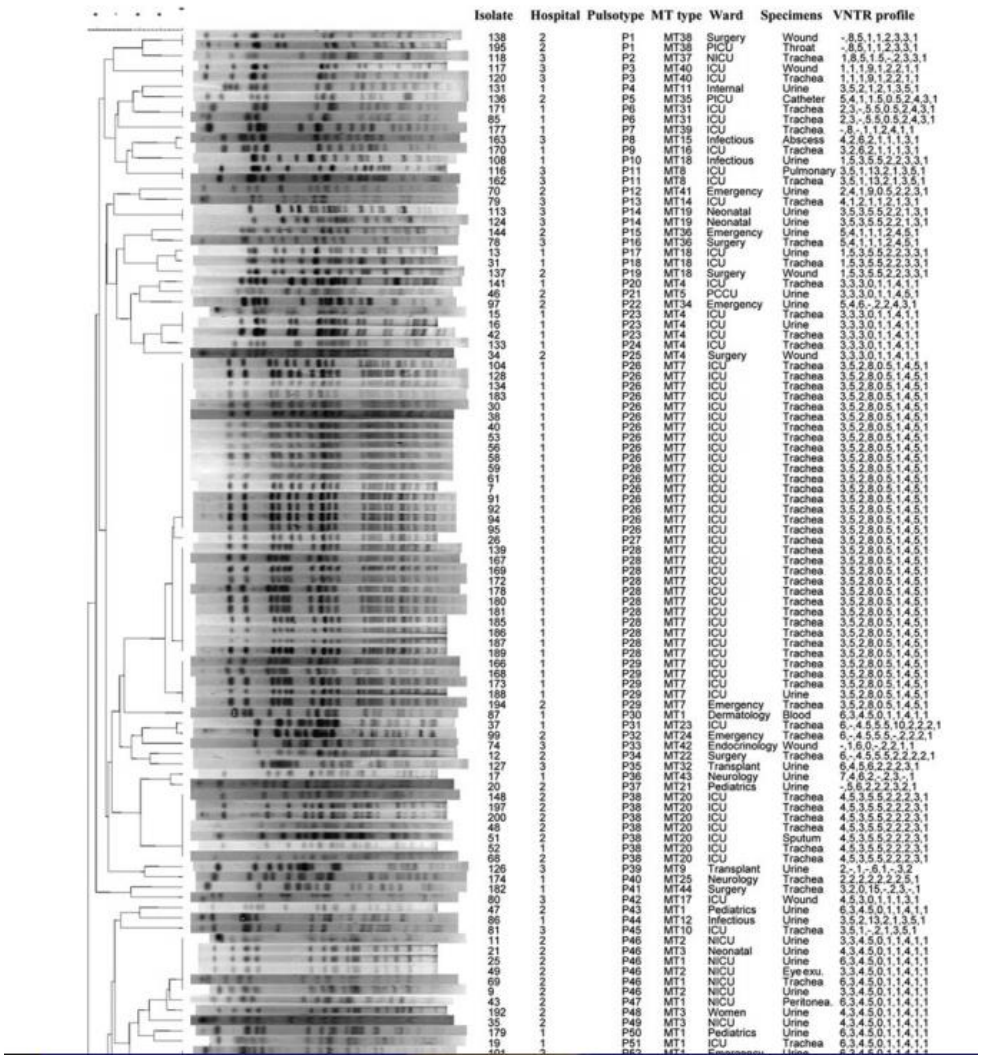
MLVA TYPES

Microbiology and Immunology, Volume: 64, Issue: 5, Pages: 331-344, First published: 22 January 2020, DOI: (10.1111/1348-0421.12776)

Molecular epidemiology of clinically high-risk Pseudomonas aeruginosa strains: Practical overview



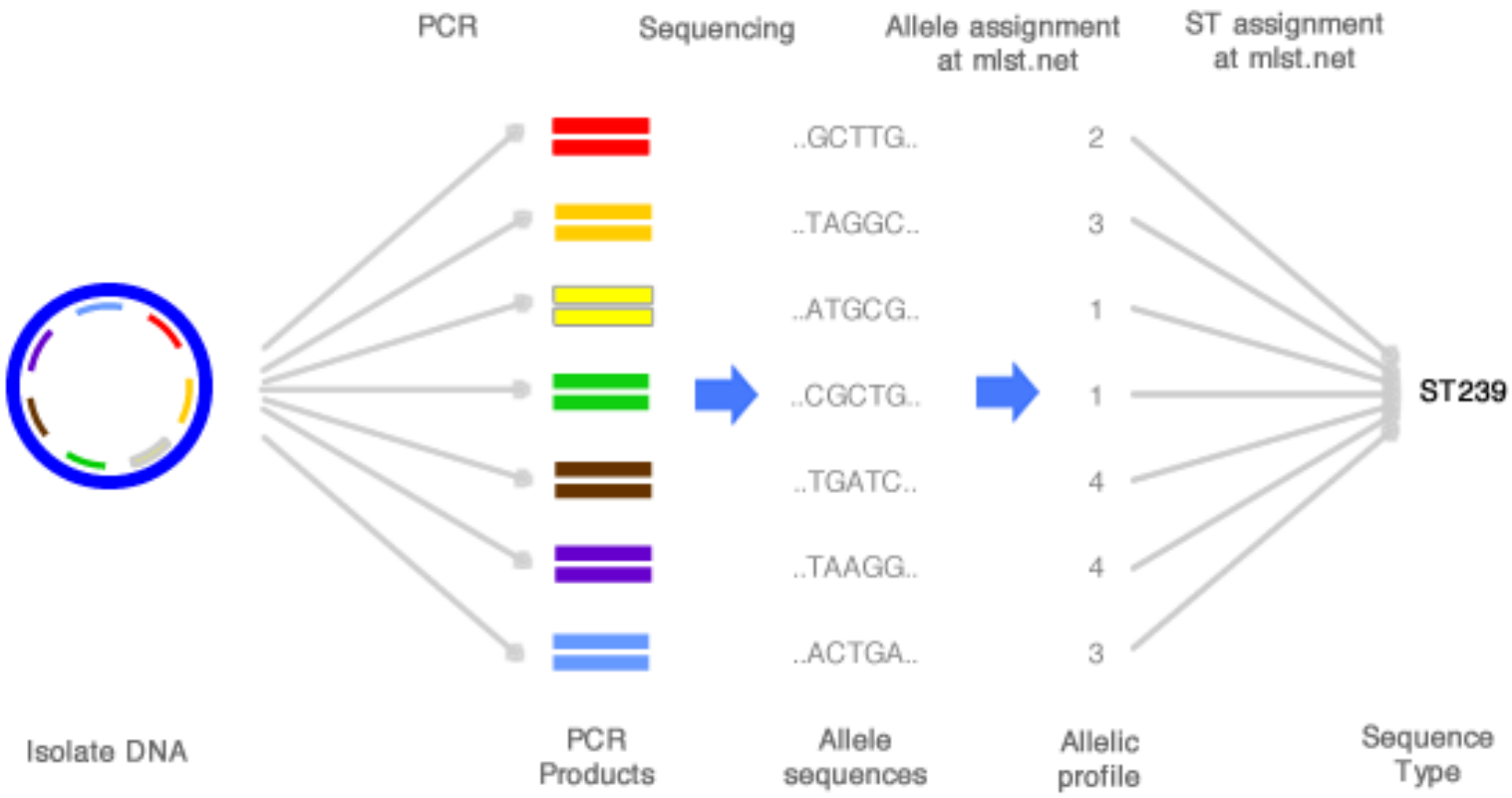
Microbiology and Immunology, Volume: 64, Issue: 5, Pages: 331-344, First published: 22 January 2020, DOI: (10.1111/1348-0421.12776)



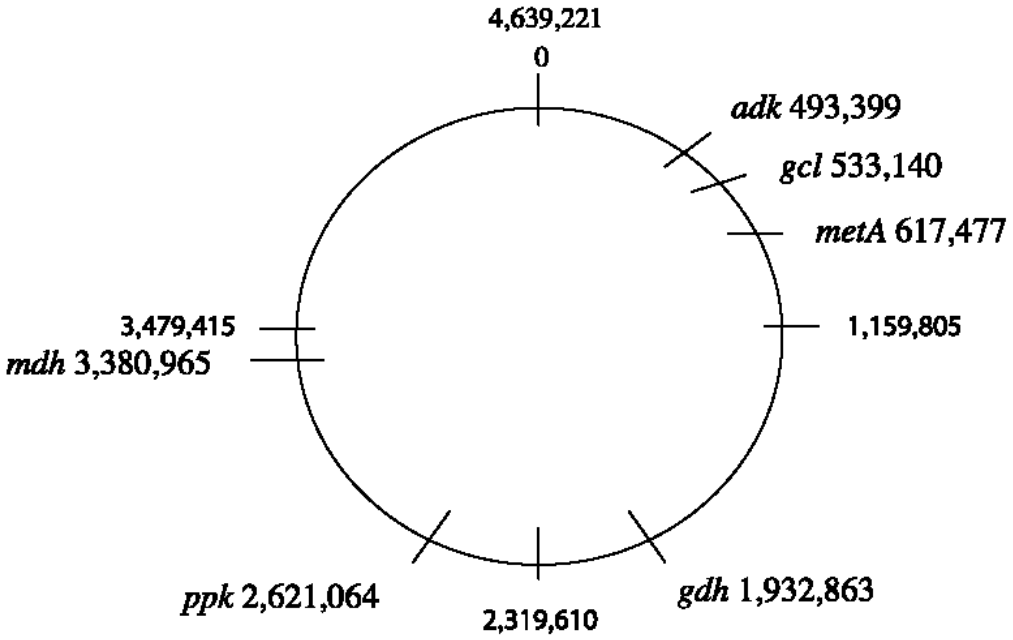
Isolate	Hospital	Pulsotype	MT type	Ward	Specimens	VNTR profile
138	2	P1	MT38	Surgery	Wound	-,8,5,1,1,2,3,3,1
195	2	P1	MT38	PICU	Throat	-,8,5,1,1,2,3,3,1
118	3	P2	MT37	NICU	Trachea	1,8,5,1,5,-,2,3,3,1
117	3	P3	MT40	ICU	Wound	1,1,1,9,1,2,2,1,1
120	3	P3	MT40	ICU	Trachea	1,1,1,9,1,2,2,1,1
131	1	P4	MT11	Internal	Urine	3,5,2,1,2,1,3,5,1
136	2	P5	MT35	PICU	Catheter	5,4,1,1,5,0,5,2,4,3,1
171	1	P6	MT31	ICU	Trachea	2,3,-,5,5,0,5,2,4,3,1
85	1	P6	MT31	ICU	Trachea	2,3,-,5,5,0,5,2,4,3,1
177	1	P7	MT39	ICU	Trachea	-,8,-,1,1,2,4,1,1
163	3	P8	MT15	Infectious	Abscess	4,2,6,2,1,1,1,3,1
170	1	P9	MT16	ICU	Trachea	3,2,6,2,1,1,1,3,1
108	1	P10	MT18	Infectious	Urine	1,5,3,5,5,2,2,3,3,1
116	3	P11	MT8	ICU	Pulmonary	3,5,1,13,2,1,3,5,1
162	3	P11	MT8	ICU	Trachea	3,5,1,13,2,1,3,5,1
70	2	P12	MT41	Emergency	Urine	2,4,1,9,0,5,2,2,3,1
79	3	P13	MT14	ICU	Trachea	4,1,2,1,1,2,1,3,1
113	3	P14	MT19	Neonatal	Urine	3,5,3,5,5,2,2,1,3,1
124	3	P14	MT19	Neonatal	Urine	3,5,3,5,5,2,2,1,3,1
144	2	P15	MT36	Emergency	Urine	5,4,1,1,1,2,4,5,1
78	3	P16	MT36	Surgery	Trachea	5,4,1,1,1,2,4,5,1
13	1	P17	MT18	ICU	Urine	1,5,3,5,5,2,2,3,3,1
31	1	P18	MT18	ICU	Trachea	1,5,3,5,5,2,2,3,3,1
137	2	P19	MT18	Surgery	Wound	1,5,3,5,5,2,2,3,3,1
141	1	P20	MT4	ICU	Trachea	3,3,3,0,1,1,4,1,1
46	2	P21	MT5	PCCU	Urine	3,3,3,0,1,1,4,5,1
97	2	P22	MT34	Emergency	Urine	5,4,6,-,2,2,4,3,1
15	1	P23	MT4	ICU	Trachea	3,3,3,0,1,1,4,1,1
16	1	P23	MT4	ICU	Urine	3,3,3,0,1,1,4,1,1
42	1	P23	MT4	ICU	Trachea	3,3,3,0,1,1,4,1,1
133	1	P24	MT4	ICU	Trachea	3,3,3,0,1,1,4,1,1
34	2	P25	MT4	Surgery	Wound	3,3,3,0,1,1,4,1,1
104	1	P26	MT7	ICU	Trachea	3,5,2,8,0,5,1,4,5,1
128	1	P26	MT7	ICU	Trachea	3,5,2,8,0,5,1,4,5,1
134	1	P26	MT7	ICU	Trachea	3,5,2,8,0,5,1,4,5,1
183	1	P26	MT7	ICU	Trachea	3,5,2,8,0,5,1,4,5,1
30	1	P26	MT7	ICU	Trachea	3,5,2,8,0,5,1,4,5,1
38	1	P26	MT7	ICU	Trachea	3,5,2,8,0,5,1,4,5,1
40	1	P26	MT7	ICU	Trachea	3,5,2,8,0,5,1,4,5,1
53	1	P26	MT7	ICU	Trachea	3,5,2,8,0,5,1,4,5,1
56	1	P26	MT7	ICU	Trachea	3,5,2,8,0,5,1,4,5,1
58	1	P26	MT7	ICU	Trachea	3,5,2,8,0,5,1,4,5,1
59	1	P26	MT7	ICU	Trachea	3,5,2,8,0,5,1,4,5,1
61	1	P26	MT7	ICU	Trachea	3,5,2,8,0,5,1,4,5,1
7	1	P26	MT7	ICU	Trachea	3,5,2,8,0,5,1,4,5,1
91	1	P26	MT7	ICU	Trachea	3,5,2,8,0,5,1,4,5,1
92	1	P26	MT7	ICU	Trachea	3,5,2,8,0,5,1,4,5,1
94	1	P26	MT7	ICU	Trachea	3,5,2,8,0,5,1,4,5,1
95	1	P26	MT7	ICU	Trachea	3,5,2,8,0,5,1,4,5,1
26	1	P27	MT7	ICU	Trachea	3,5,2,8,0,5,1,4,5,1
139	1	P28	MT7	ICU	Trachea	3,5,2,8,0,5,1,4,5,1
167	1	P28	MT7	ICU	Trachea	3,5,2,8,0,5,1,4,5,1
169	1	P28	MT7	ICU	Trachea	3,5,2,8,0,5,1,4,5,1
172	1	P28	MT7	ICU	Trachea	3,5,2,8,0,5,1,4,5,1
178	1	P28	MT7	ICU	Trachea	3,5,2,8,0,5,1,4,5,1
180	1	P28	MT7	ICU	Trachea	3,5,2,8,0,5,1,4,5,1
181	1	P28	MT7	ICU	Trachea	3,5,2,8,0,5,1,4,5,1
185	1	P28	MT7	ICU	Trachea	3,5,2,8,0,5,1,4,5,1
186	1	P28	MT7	ICU	Trachea	3,5,2,8,0,5,1,4,5,1
187	1	P28	MT7	ICU	Trachea	3,5,2,8,0,5,1,4,5,1
189	1	P28	MT7	ICU	Trachea	3,5,2,8,0,5,1,4,5,1
166	1	P29	MT7	ICU	Trachea	3,5,2,8,0,5,1,4,5,1
168	1	P29	MT7	ICU	Trachea	3,5,2,8,0,5,1,4,5,1
173	1	P29	MT7	ICU	Trachea	3,5,2,8,0,5,1,4,5,1
188	1	P29	MT7	ICU	Urine	3,5,2,8,0,5,1,4,5,1
194	2	P29	MT7	Emergency	Trachea	3,5,2,8,0,5,1,4,5,1
87	1	P30	MT1	Dermatology	Blood	6,3,4,5,0,1,1,4,1,1
37	1	P31	MT23	ICU	Trachea	6,-,4,5,5,5,10,2,2,2,1
99	2	P32	MT24	Emergency	Trachea	6,-,4,5,5,5,-,2,2,2,1
74	3	P33	MT42	Endocrinology	Wound	-,1,6,0,-,2,2,1,1
12	2	P34	MT22	Surgery	Trachea	6,-,4,5,5,5,2,2,2,2,1
127	3	P35	MT32	Transplant	Urine	6,4,5,6,6,2,2,3,1
17	1	P36	MT43	Neurology	Urine	7,4,6,2,-,2,3,-,1
20	2	P37	MT21	Pediatrics	Urine	-,5,6,2,2,2,3,2,1
148	2	P38	MT20	ICU	Trachea	4,5,3,5,5,2,2,2,3,1
197	2	P38	MT20	ICU	Trachea	4,5,3,5,5,2,2,2,3,1
200	2	P38	MT20	ICU	Trachea	4,5,3,5,5,2,2,2,3,1
48	2	P38	MT20	ICU	Trachea	4,5,3,5,5,2,2,2,3,1
51	2	P38	MT20	ICU	Sputum	4,5,3,5,5,2,2,2,3,1
52	1	P38	MT20	ICU	Trachea	4,5,3,5,5,2,2,2,3,1
65	2	P38	MT20	ICU	Trachea	4,5,3,5,5,2,2,2,3,1
126	3	P39	MT9	Transplant	Urine	2,-,1,-,6,1,-,3,2
174	1	P40	MT25	Neurology	Trachea	2,2,2,2,2,2,2,5,1
182	1	P41	MT44	Surgery	Trachea	3,2,0,15,-,2,3,-,1
80	3	P42	MT17	ICU	Wound	4,5,3,0,1,1,1,3,1
47	2	P43	MT1	Pediatrics	Urine	6,3,4,5,0,1,1,4,1,1
86	1	P44	MT12	Infectious	Urine	3,5,2,13,2,1,3,5,1
81	3	P45	MT10	ICU	Trachea	3,5,1,-,2,1,3,5,1
11	2	P46	MT2	NICU	Urine	3,3,4,5,0,1,1,4,1,1
21	2	P46	MT3	Neonatal	Urine	4,3,4,5,0,1,1,4,1,1
25	2	P46	MT1	NICU	Urine	6,3,4,5,0,1,1,4,1,1
49	2	P46	MT2	NICU	Eye exu.	3,3,4,5,0,1,1,4,1,1
69	2	P46	MT1	NICU	Trachea	6,3,4,5,0,1,1,4,1,1
9	2	P46	MT2	NICU	Urine	3,3,4,5,0,1,1,4,1,1
43	2	P47	MT1	NICU	Peritonea.	6,3,4,5,0,1,1,4,1,1
192	2	P48	MT3	Women	Urine	4,3,4,5,0,1,1,4,1,1
35	2	P49	MT3	NICU	Urine	4,3,4,5,0,1,1,4,1,1
179	1	P50	MT1	Pediatrics	Urine	6,3,4,5,0,1,1,4,1,1
19	1	P51	MT1	ICU	Trachea	6,3,4,5,0,1,1,4,1,1

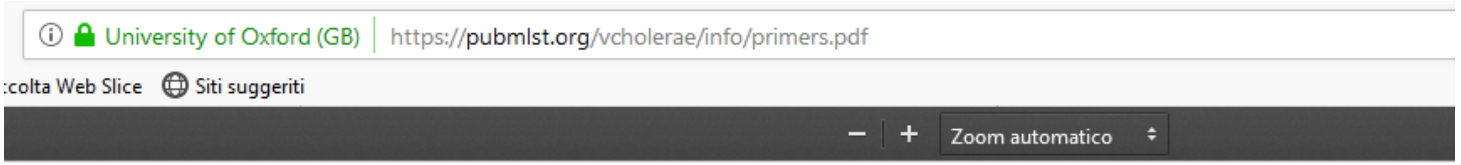
Derakhshan et al., 2016. Multiple-Locus Variable Number Tandem Repeat Analysis of *Klebsiella pneumoniae* : Comparison with Pulsed-Field Gel Electrophoresis
Microbial Drug Resistance 23(5)

Multilocus Sequence Typing:
This is a genome-based version of the conventional method of multilocus enzyme electrophoresis. It helps in the typing of various bacterial species by identifying DNA alleles from various organisms.
This method involves PCR amplification and the nucleic acid sequencing of multiple internal fragments of housekeeping genes. The advantages of this method are that the culturing of pathogenic micro-organisms is avoided and that the sequencing data are unambiguous, easy to standardize, and electronically portable.


PCR
Sequencing
Allele assignment at mlst.net
ST assignment at mlst.net
..GCTTG..
..TAGGC..
..ATGCG..
..CGCTG..
..TGATC..
..TAAGG..
..ACTGA..
2
3
1
1
4
4
3
ST239
Isolate DNA
PCR Products
Allele sequences
Allelic profile
Sequence Type

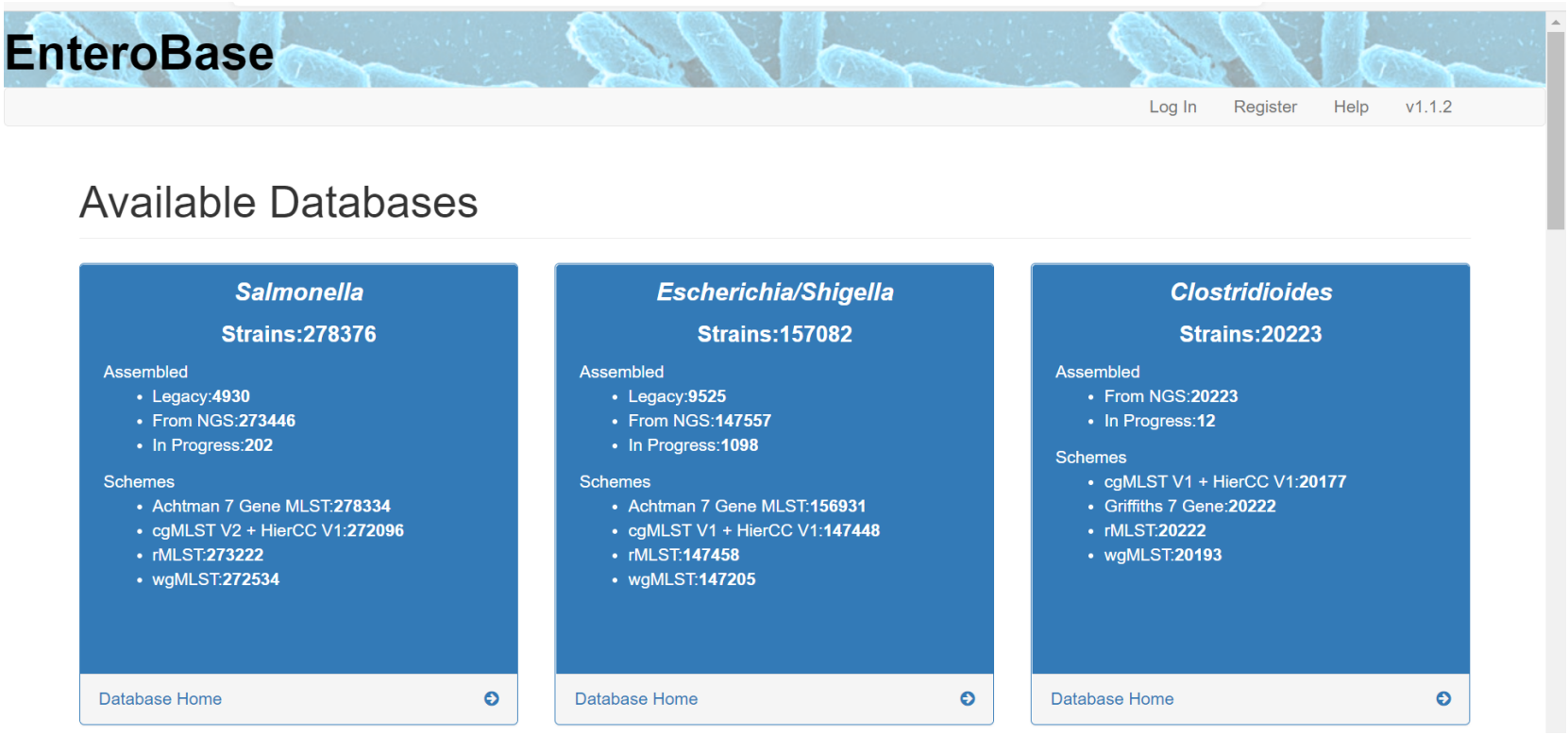
MultiLocus Sequence Typing: genes along the chromosome





***Vibrio cholerae* primers**

Gene	Gene Product	Direction	Oligonucleotide Sequence (5' à 3')
adk	adenylate kinase	F	CATCATTCTTCTCGGTGCTC
		R	AGTGCCGTCAAACTTCAGGTA
gyrB	DNA gyrase subunit B	F	GTACGTTTCTGGCCTAGTGC
		R	GGGTCTTTTTCCTGACAATC
metE	methionine synthase	F	CGGGTGACTTTGCTTGGT
		R	CAGATCGACTGGGCTGTG
mdh	malate dehydrogenase	F	ATGAAAGTCGCTGTTATTGG
		R	GCCGCTTGGCCCATAGAAAG
		R	TAGCTTGATAGGTTGGG
pntA	pyridine nucleotide transhydrogenase	F	CTTTGATGGAAAAACTCTCA
		R	GATATTGCCGTCTTTTTCTT
		F	GGCCAGCCCAAAATCCT
purM	phosphoribosyl-formylglycinamide cyclo-ligase	F	GGTGTCGATATTGATGCAGG
		R	GGAATGTTTTCCCAGAAGCC
pyrC	dihydroorotase	F	ATCATGCCTAACACGGTTCC
		R	TTCAAACACTTCGGCATA


EnteroBase
Log In
Register
Help
v1.1.2
Available Databases
Salmonella
Strains:278376
Assembled
Legacy:4930
From NGS:273446
In Progress:202
Schemes
Achtman 7 Gene MLST:278334
cgMLST V2 + HierCC V1:272096
rMLST:273222
wgMLST:272534
Database Home
Escherichia/Shigella
Strains:157082
Assembled
Legacy:9525
From NGS:147557
In Progress:1098
Schemes
Achtman 7 Gene MLST:156931
cgMLST V1 + HierCC V1:147448
rMLST:147458
wgMLST:147205
Database Home
Clostridioides
Strains:20223
Assembled
From NGS:20223
In Progress:12
Schemes
cgMLST V1 + HierCC V1:20177
Griffiths 7 Gene:20222
rMLST:20222
wgMLST:20193
Database Home

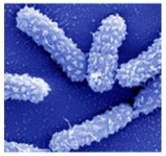
Institut Pasteur

INSTITUT PASTEUR MLST
and whole genome MLST databases

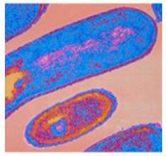
This site hosts databases of multilocus sequence typing (MLST) and whole-genome based typing schemes, which are used for genotyping of bacterial isolates. They provide reference nomenclatures of microbial strains and are mainly intended for molecular epidemiology of pathogens of public health importance, detection of virulence and antimicrobial resistance genes, and for population biology research. This site is powered by the BIGSdb software.

Databases hosted on this site

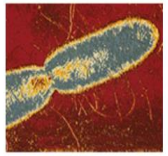
Klebsiella pneumoniae



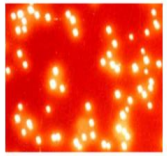
Listeria monocytogenes



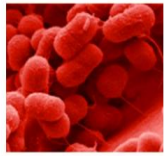
Escherichia coli



Bordetella



Acinetobacter baumannii



INSTITUT PASTEUR | RESEARCH | EDUCATION | PUBLIC HEALTH | INTERNATIONAL | CAREERS | DONATE

INSTITUT PASTEUR

PUBLICATION : JOURNAL OF CLINICAL MICROBIOLOGY

Multilocus sequence typing of Klebsiella pneumoniae nosocomial isolates

Member: Sylvain Brisse

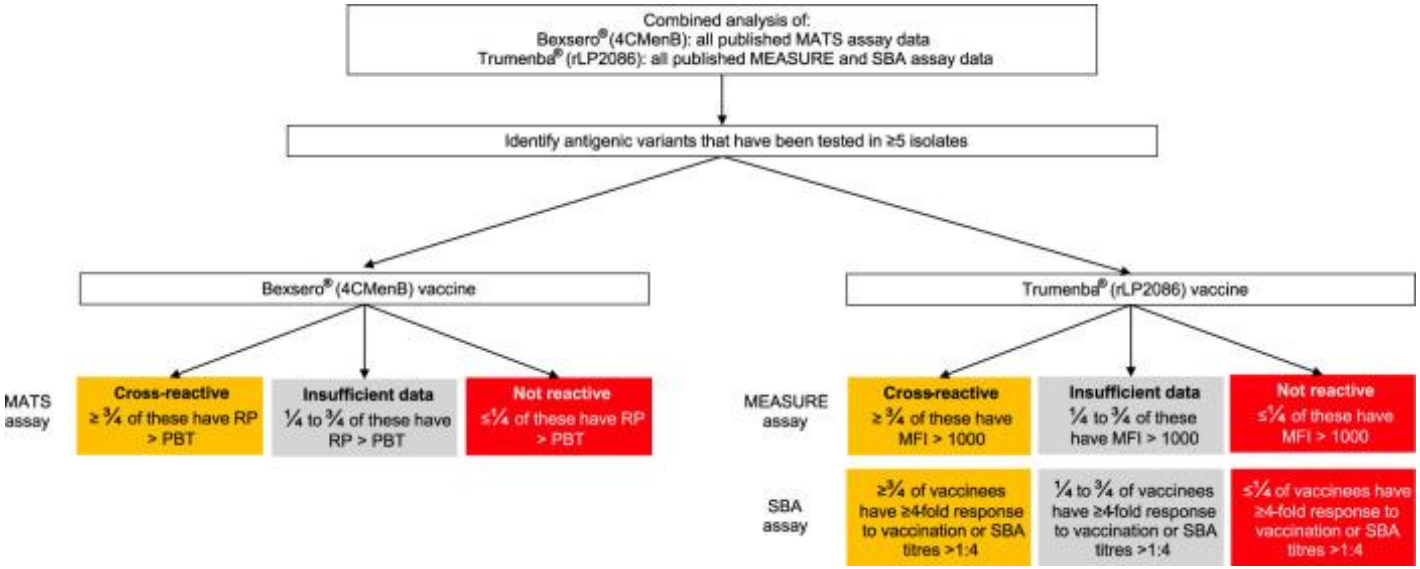
© RESEARCH

The Deduced Vaccine Antigen Reactivity (MenDeVAR) Index was developed to combine multiple, complex data that inform the reactivity of the Bexsero® and Trumenba® vaccines against specific antigenic variants.

•Bexsero® (4CMenB) is a multicomponent vaccine. Peptide sequence diversity can be analysed using the Bexsero Antigen Sequence Typing (BAST) scheme (Brehony *et al.* 2016). The vaccine contains antigen variants: fHbp peptide 1; NHBA peptide 2; NadA peptide 8; PorA VR2 4.

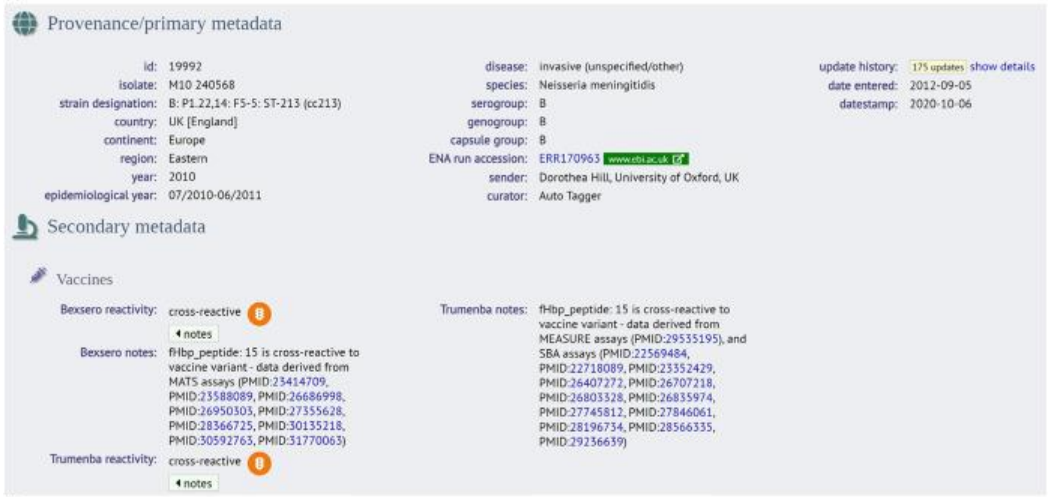
•Trumenba® (rLP2086) is a bivalent fHbp-containing vaccine. Peptide sequence diversity can be analysed using the fHbp peptide locus. The vaccine contains fHbp peptide variants 45 and 55.

Meningococcal Deduced Vaccine Antigen Reactivity (MenDeVAR) Index: a Rapid and Accessible Tool That Exploits Genomic Data in Public Health and Clinical Microbiology Applications. Charlene M. C. Rodrigues, Keith A. Jolley, Andrew Smith, J. Claire Cameron, Ian M. Feavers, Martin C. J. Maiden. J Clin Microbiol. 2021 Jan; 59(1): e02161-20.

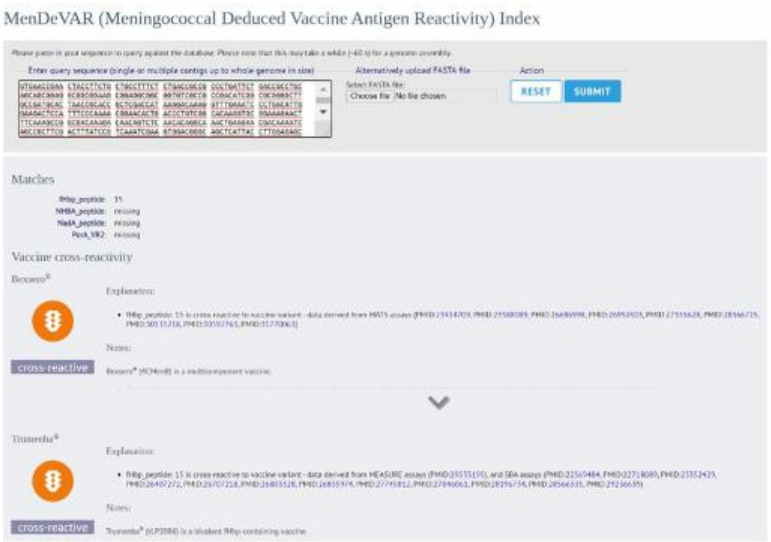


The Meningococcal Deduced Vaccine Antigen Reactivity (MenDeVAR) Index algorithm used to identify which antigens are included as cross-reactive in the combined analysis of published experimental data from the meningococcal antigen typing system (MATS) , the meningococcal antigen surface expression (MEASURE) assay, and the serum bactericidal activity (SBA) assay. RP, relative potency; PBT, positive bactericidal threshold; MFI, mean fluorescence intensity.

(a)



(b)



Upload a meningococcal genome assembly by pasting in to the web form or uploading a FASTA file. On submission, the protein variants found in the genome are determined and the vaccine reactivity results given as below for each vaccine:

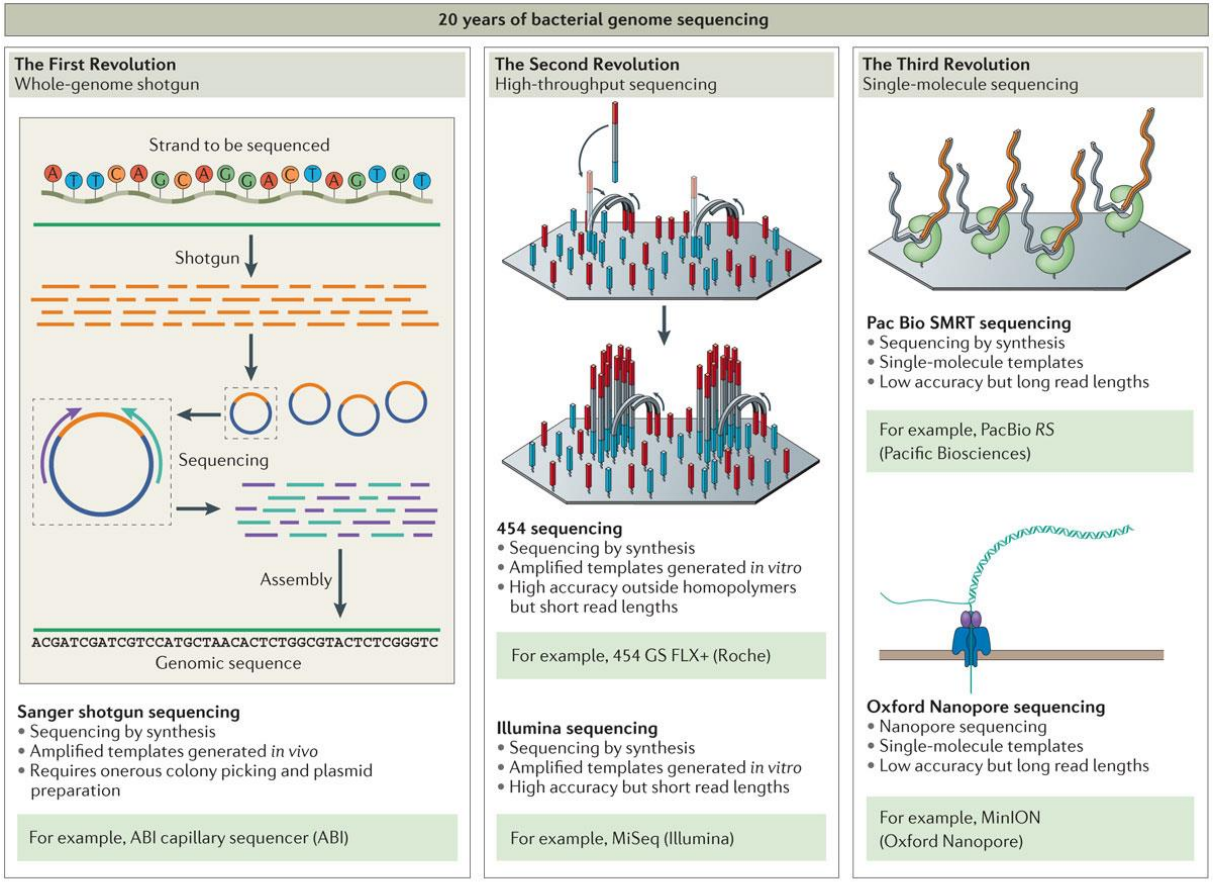
- **exact match:** isolate contains ≥1 exact sequence match to antigenic variants found in the vaccine.
- **cross-reactive:** isolate contains ≥1 antigenic variant deemed cross-reactive to vaccine variants through experimental studies.
- **none:** all the isolate's antigenic variants have been deemed not cross-reactive to vaccine variants through experimental studies.
- **insufficient data:** isolate contains antigens for which there is insufficient data from or are yet to be tested in experimental studies.

The output is supplemented with a detailed description of the evidence supporting the result. *Only evidence published in the scientific literature is used*.

In conclusion, we present a generalizable multilocus gene-by-gene framework for interpreting complex genomic data sets that can be used by practitioners to address clinical questions in a timely manner. Specifically, the MenDeVAR Index combines genomic and experimental data to provide a rational, evidence-based estimate of the likelihood that either of the meningococcal protein-based vaccines offers protection against a given meningococcus. To ensure broad accessibility, the MenDeVAR Index is implemented with a “red,” “amber,” and “green” interpretive interface that is easy to use and informative for practitioners without expertise in genomic analysis. In the light of new published evidence, the MenDeVAR Index can be regularly reevaluated using the criteria described here, adjusting antigenic variant designations accordingly, to ensure that public health and clinical microbiologists globally benefit from the latest research findings

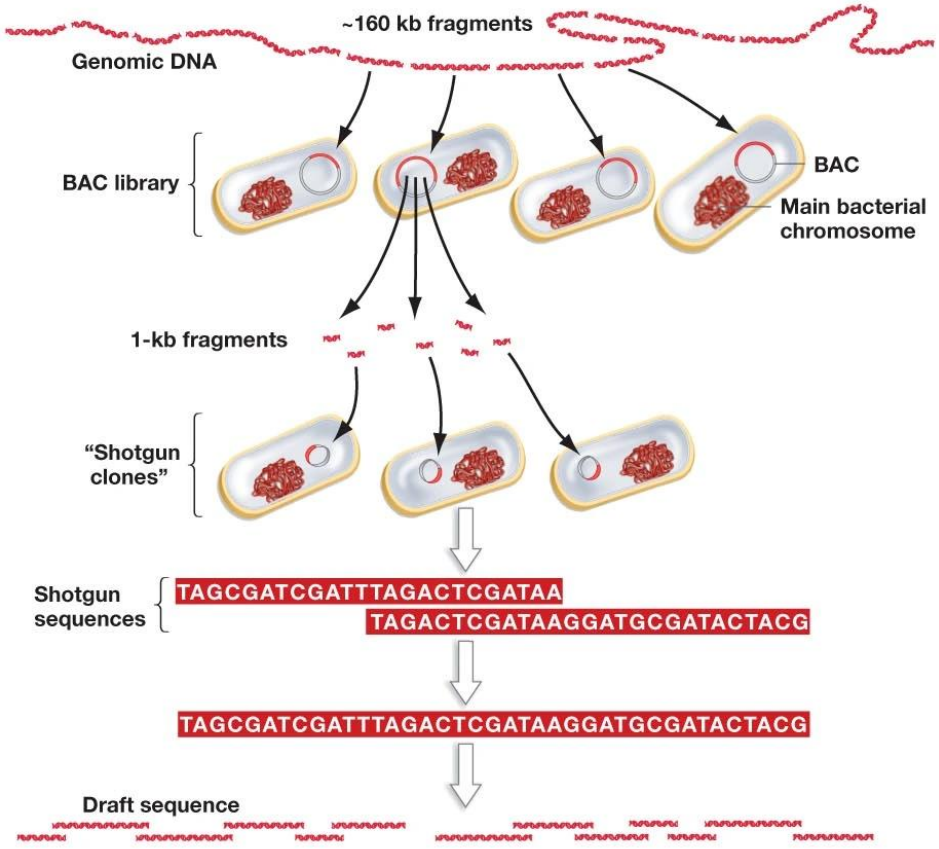
20 years of Whole Genome Sequencing (WGS) of bacteria

- Robust data: one method for all bacterial species
- Data storage for later analysis
- Monitoring of epidemic cases in hospital
- Monitoring emergency prevalent and emerging clones
- Identification of all genes of interest
- International comparison of prevalent and emerging clones



Loman, N., Pallen, M. Twenty years of bacterial genome sequencing. *Nat Rev Microbiol* **13,** 787–794 (2015). https://doi.org/10.1038/nrmicro3565

SHOTGUN SEQUENCING A GENOME



1. Cut DNA into fragments of ~160 kb, using sonication.

2. Insert fragments into bacterial artificial chromosomes; grow in *E. coli* cells to obtain large numbers of each fragment.

3. Purify each 160-kb fragment, then cut each into a set of 1-kb fragments, using sonication, so that 1-kb fragments overlap.

4. Insert 1-kb fragments into plasmids; grow in *E. coli* cells. Obtain many copies of each fragment.

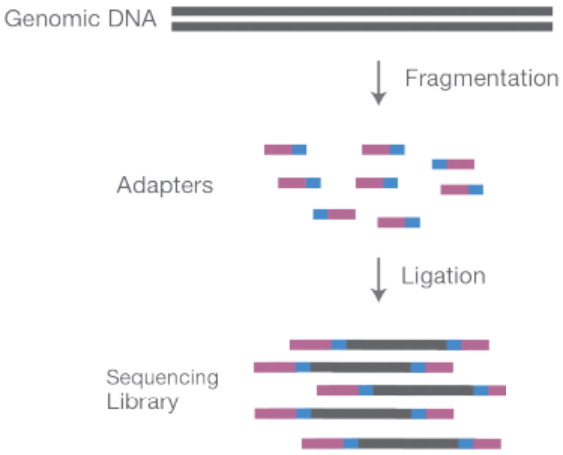
5. Sequence each fragment. Find regions where different fragments overlap.

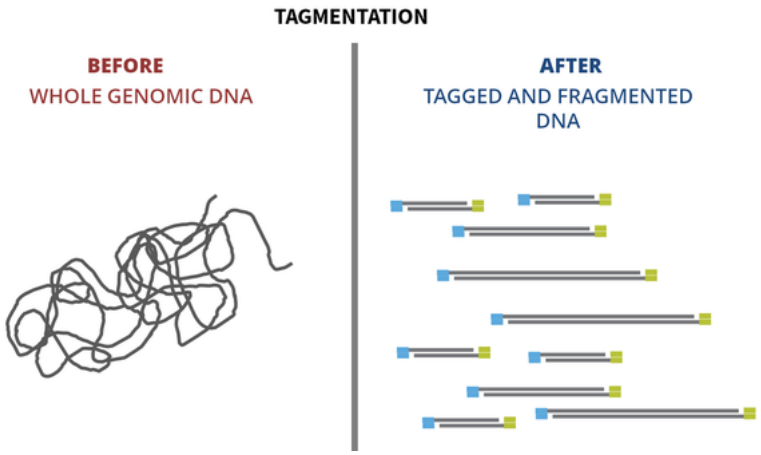
6. Assemble all the 1-kb fragments from each original 160-kb fragment by matching overlapping ends.

7. Assemble sequences from different BACs (160-kb fragments) by matching overlapping ends.

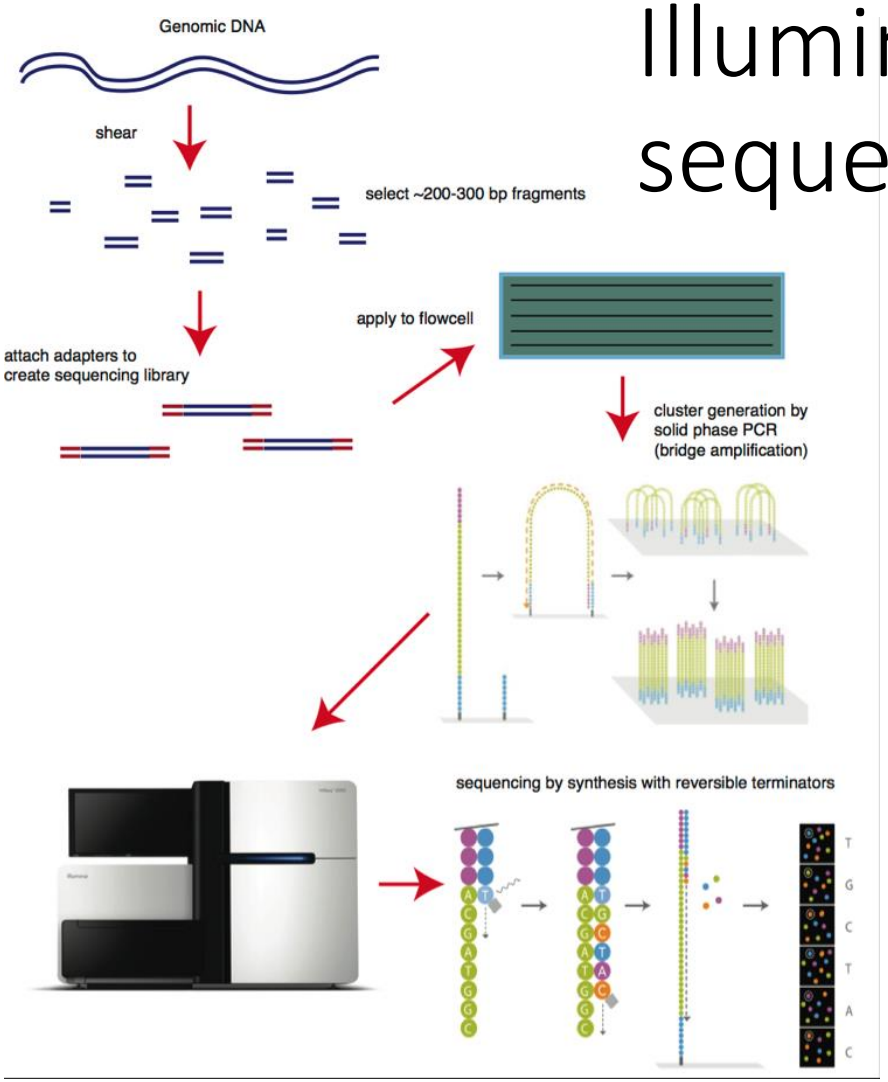
Whole-genome sequencing

2nd generation

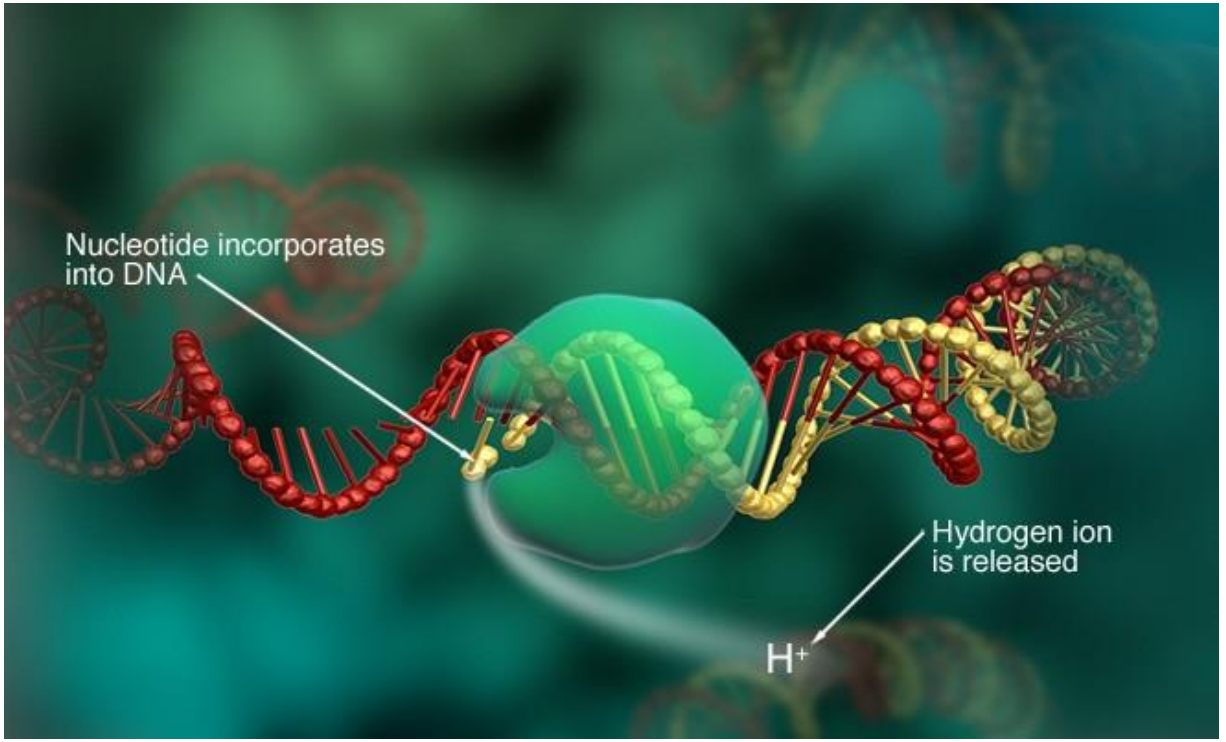




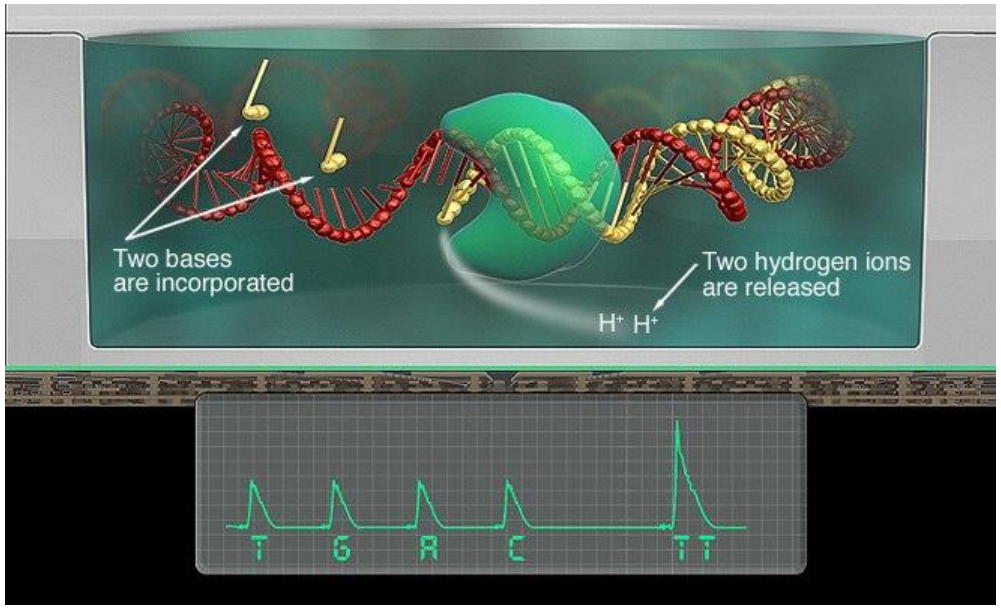
Illumina sequencing



ION TORRENT Technology



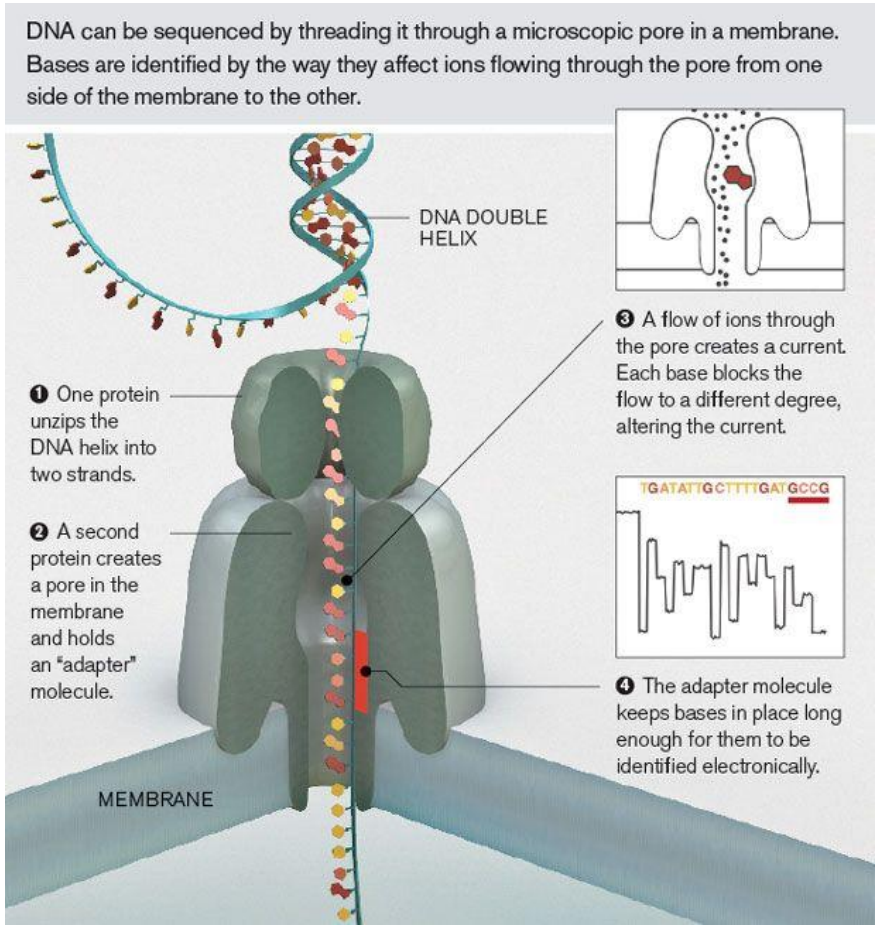
Ion Torrent™ technology directly translates chemically encoded information (A, C, G, T) into digital information (0, 1) on a semiconductor chip. This approach marries simple chemistry to proprietary semiconductor technology

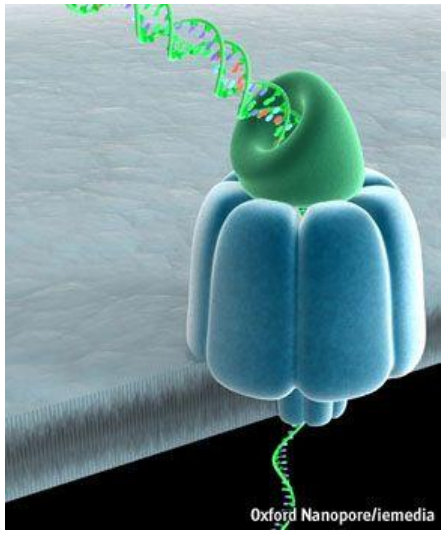


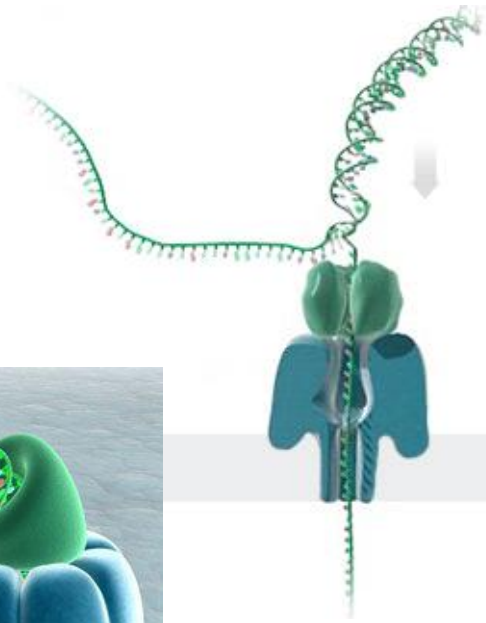
https://www.thermofisher.com/it/en/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-technology.html

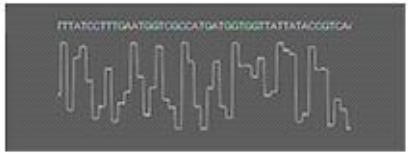
Nanopore sequencing

Nanopore sequencing

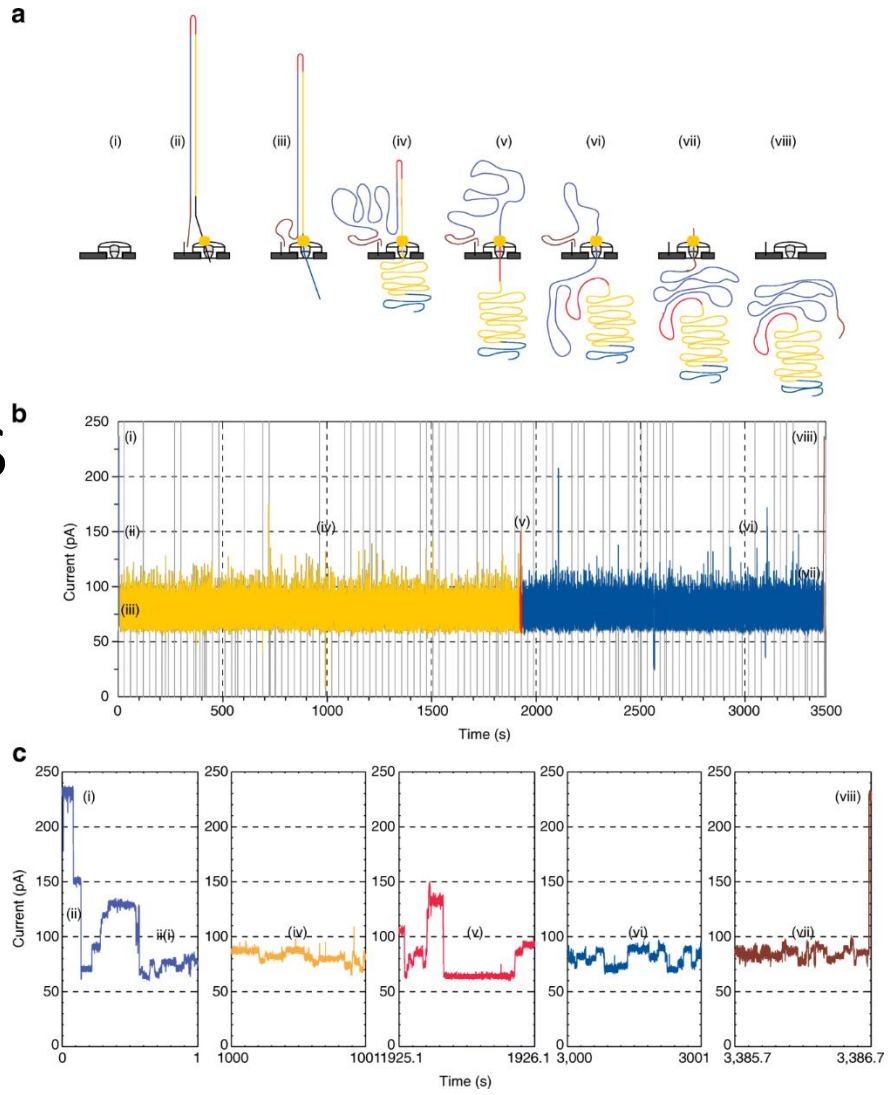



Oxford Nanopore/iemedia

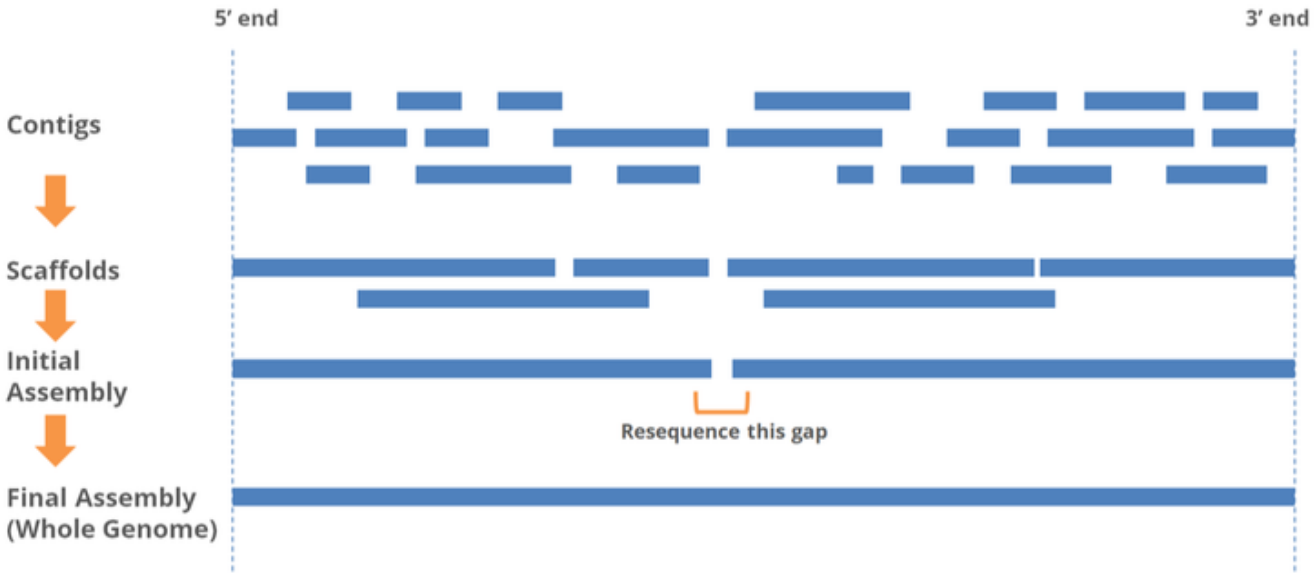




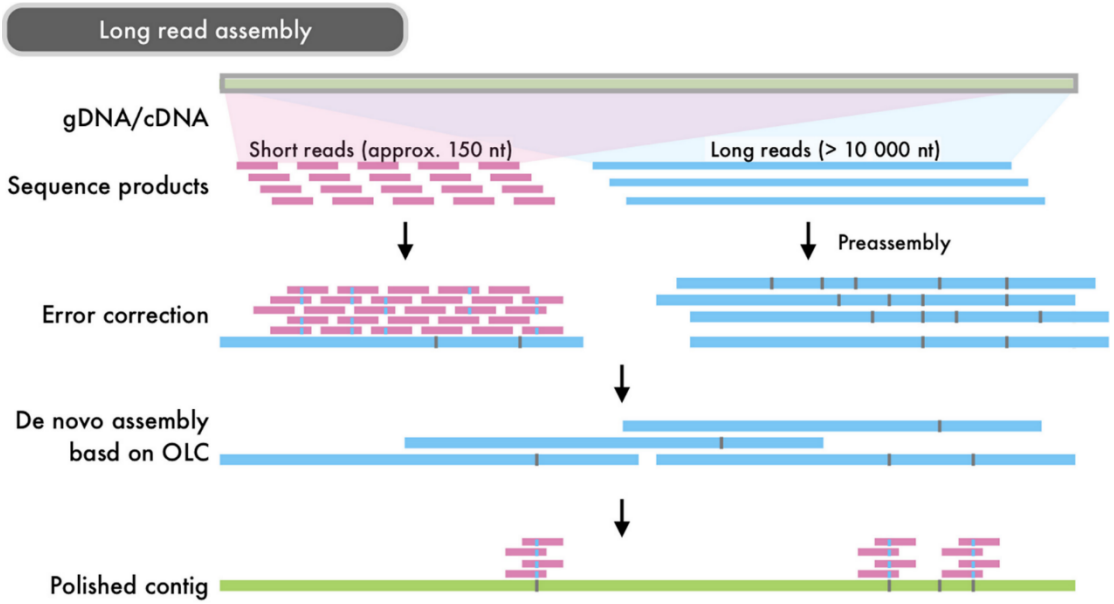
Nanopore sequencing:
long reads but mistakes!!!



Assembler


5′ end
3′ end
Contigs
Scaffolds
Initial Assembly
Resequence this gap
Final Assembly (Whole Genome)

short reads and long reads



Studying genome content

- De novo sequencing
- Comparative analysis with reference genomes
- Identification of peculiar genes

DATABASES

Published in final edited form as:
Euro Surveill. ; 18(4): 20379.

Automated extraction of typing information for bacterial pathogens from whole genome sequence data: *Neisseria meningitidis* as an exemplar

Keith A. Jolley and **Martin C.J. Maiden**
Department of Zoology, University of Oxford, South Parks Road, Oxford, OX1 3PS, United Kingdom

Abstract

Whole genome sequence (WGS) data are becoming a major means of characterising samples of bacterial pathogens. These data have the advantage of providing detailed information on the genotypes and likely phenotypes of aetiological agents, enabling the relationships of samples from potential disease outbreaks to be established precisely. However, the generation of increasing quantities of sequence data does not, in itself, resolve the problems that a wide variety of microbiological typing methods have addressed over the last 100 years or so; indeed, the provision of very high volumes of unstructured data can confuse rather than resolve these issues. Here we review the nascent field of the storage of WGS data for clinical application and show how curated sequence-based typing schemes on websites such as PubMLST.org, accumulated over the past 14 years or so, has generated an infrastructure that can be used to exploit WGS for bacterial typing efficiently. We review the tools that have been implemented within the PubMLST.org website to extract clinically useful, strain characterisation information which can be provided to physicians and public health scientists and officials in a timely, concise and understandable way. These data can be used to inform medical decisions such as how to treat a patient, whether to institute public health action, and what action might be appropriate. The information is compatible both with previous sequence-based typing data and also with data that can be obtained in the absence of WGS data, for example by real-time PCR tests, providing a flexible infrastructure for WGS-based clinical microbiology.

The first MLST scheme was designed to identify major clones within populations of Neisseria meningitidis with just seven gene fragments, totalling only 3,284 bp, or about 0.15% of the whole genome.

For meningococcus, the extent of genetic diversity present even in this small number of genes under stabilising selection is extensive: as of November 2012 each of the gene fragments used as meningococcal MLST loci had between 424 to 675 distinct alleles recorded on the PubMLST Neisseria website with 54-94% (71% mean) sites that were polymorphic.

Furthermore, in the representative abcZ locus all four bases were present at a given site over the known population in 54/433 (12%) of the nucleotide positions (Figure 1). Much of this variation is at low frequency and transitory, but the precise variants that this is the case for cannot be known without exhaustive, or at least extensive, sampling over time

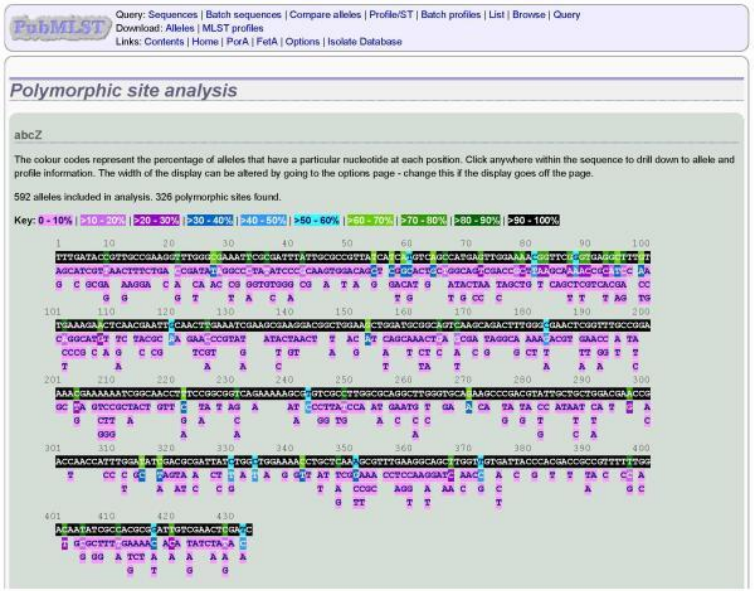


Figure 1. A schematic of one of the MLST loci showing the number and positions of known polymorphic sites within the gene fragment (unmodified PubMLST.org screenshot).

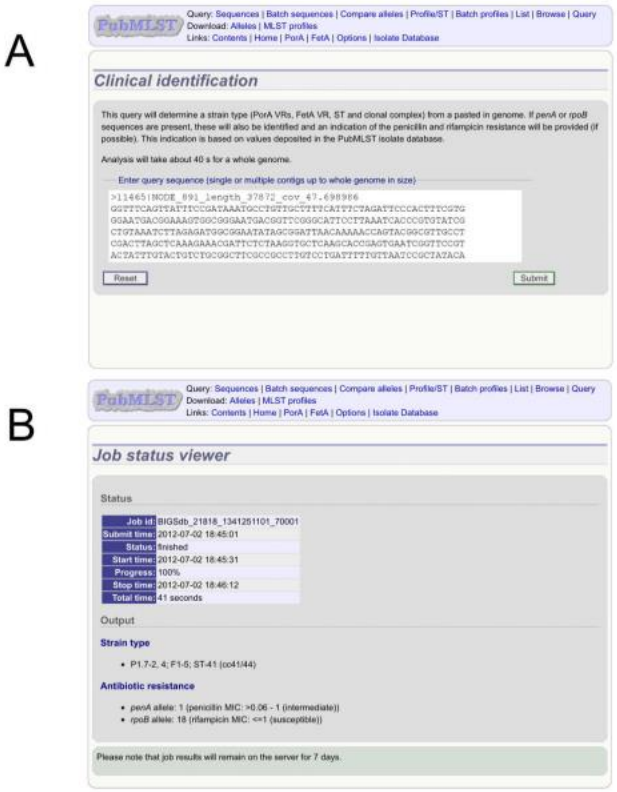


Figure 2.
Extracting antigen and antibiotic resistance data from whole genome sequences. A whole genome sequence, which may consist of multiple contigs, can be pasted in to the *Neisseria* PubMLST website (A) with typing and antibiotic resistance data for penicillin and rifampicin rapidly extracted (B) (unmodified PubMLST.org screenshots).

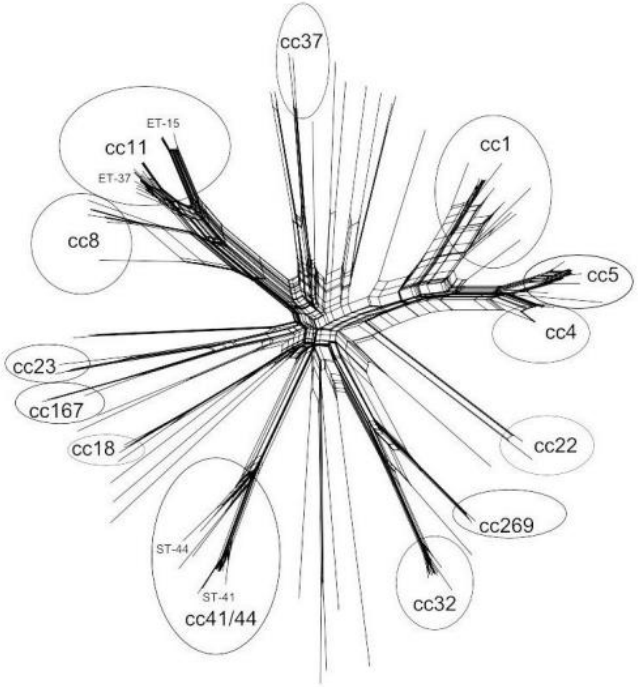


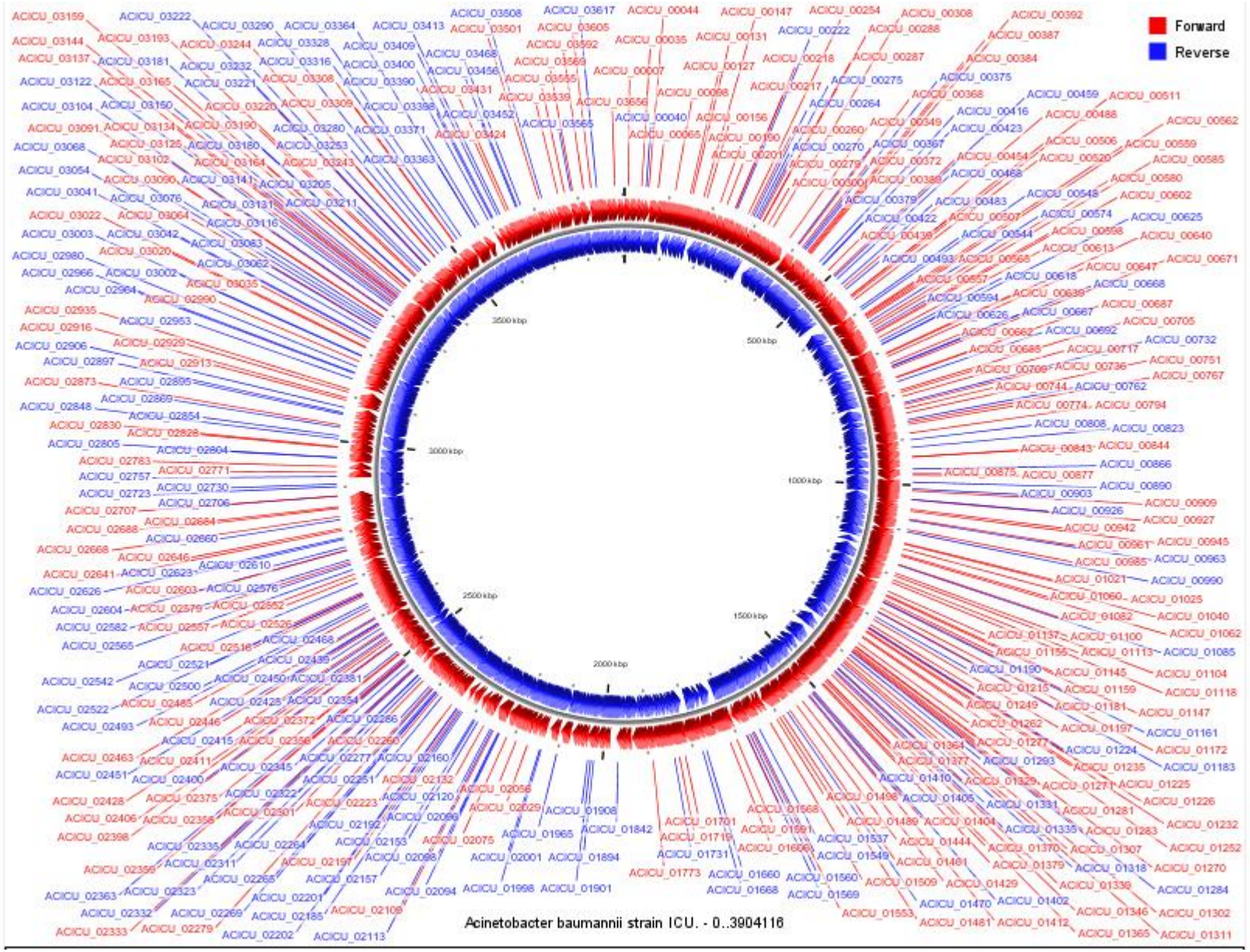
Figure 3.
The relationships of 139 *Neisseria meningitidis* genomes stored in the PubMLST.org *Neisseria* database, generated with Genome Comparator and NeighborNet from allelic profiles data for rMLST loci. The locations of isolates belonging to major clonal complexes identified by conventional MLST are indicated (cc1, etc). The figure illustrates relationships not apparent from sevenlocus MLST, including the diversity of some clonal complexes (e.g. cc1) and the interrelationships of others, e.g. cc8 and cc11 clonal complexes, and the relationships of the 'ET-15 and 'ET-37' variants within cc11.

Studying genome content

- De novo sequencing
- Comparative analysis with reference genomes
- Identification of peculiar genes

DATABASES

Genome annotation



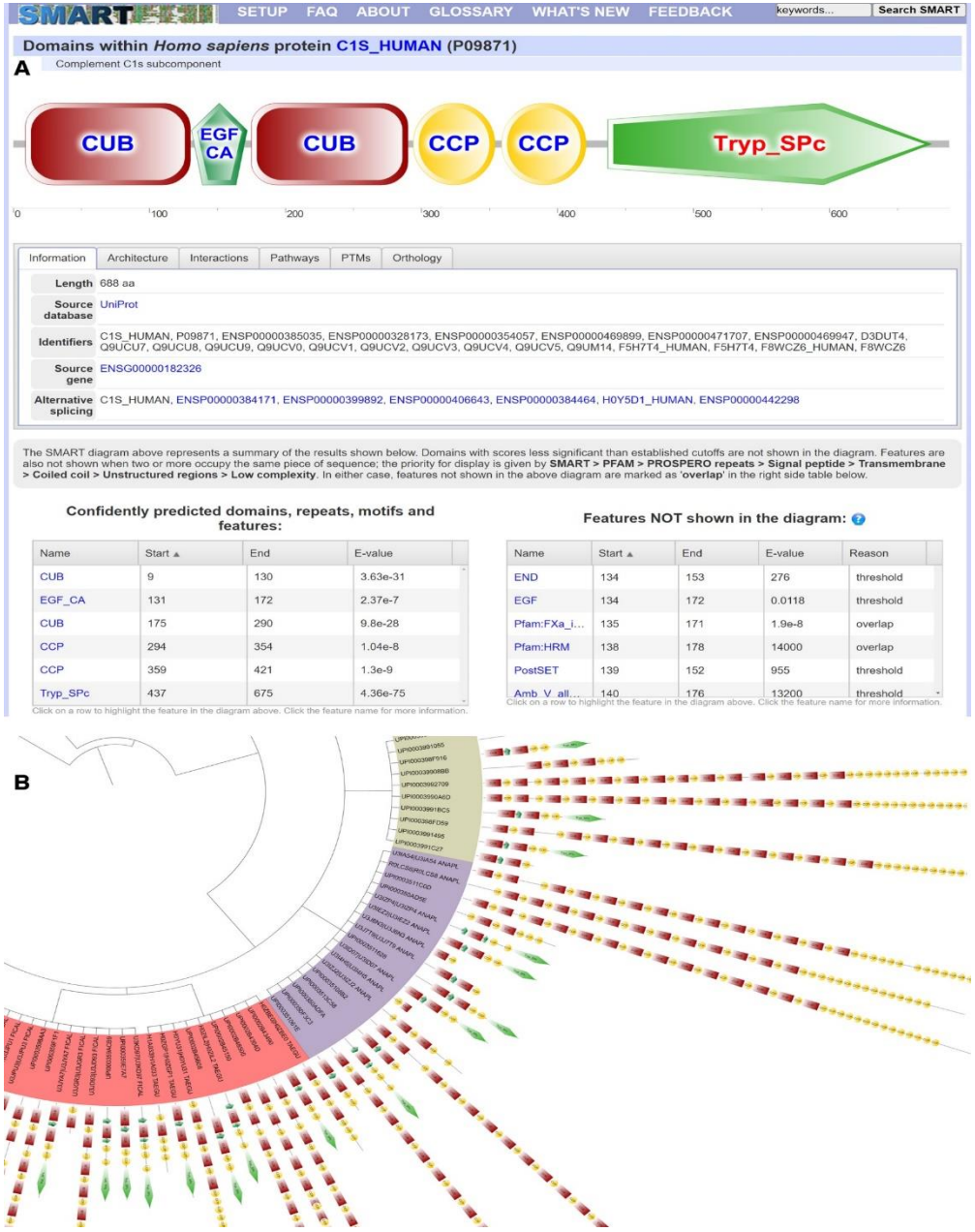
Functional classes of the proteins

- Transporters
- Energy metabolism
- Biosynthesis
  - Amino acids, lipids, nucleotides
- Cell cycle
- Virulence
- Phages

Fundamental protein domains

www.ncbi.nlm.nih.gov/COG/

Prediction of the function of a protein deduced from a DNA sequence on the basis of its functional domains



Nucleic Acids Res, Volume 46, Issue D1, 4 January 2018, Pages D493–D496 **20 years of the SMART protein domain annotation resource**

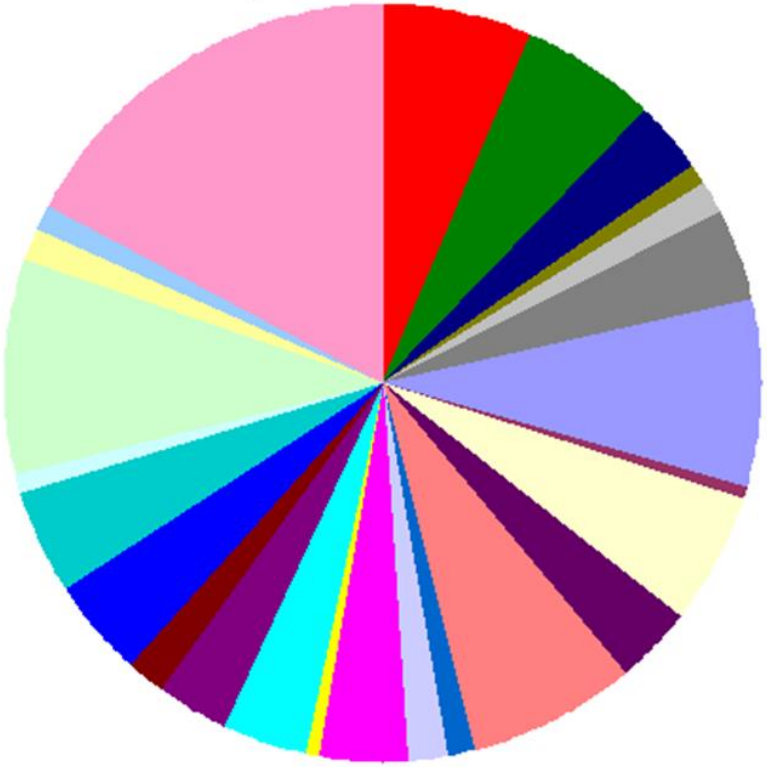
https://doi.org/10.1093/nar/gkx922

Subsystem Information

Subsystem Statistics | Features in Subsystems

Subsystem Coverage

Subsystem Category Distribution



Subsystem Feature Counts

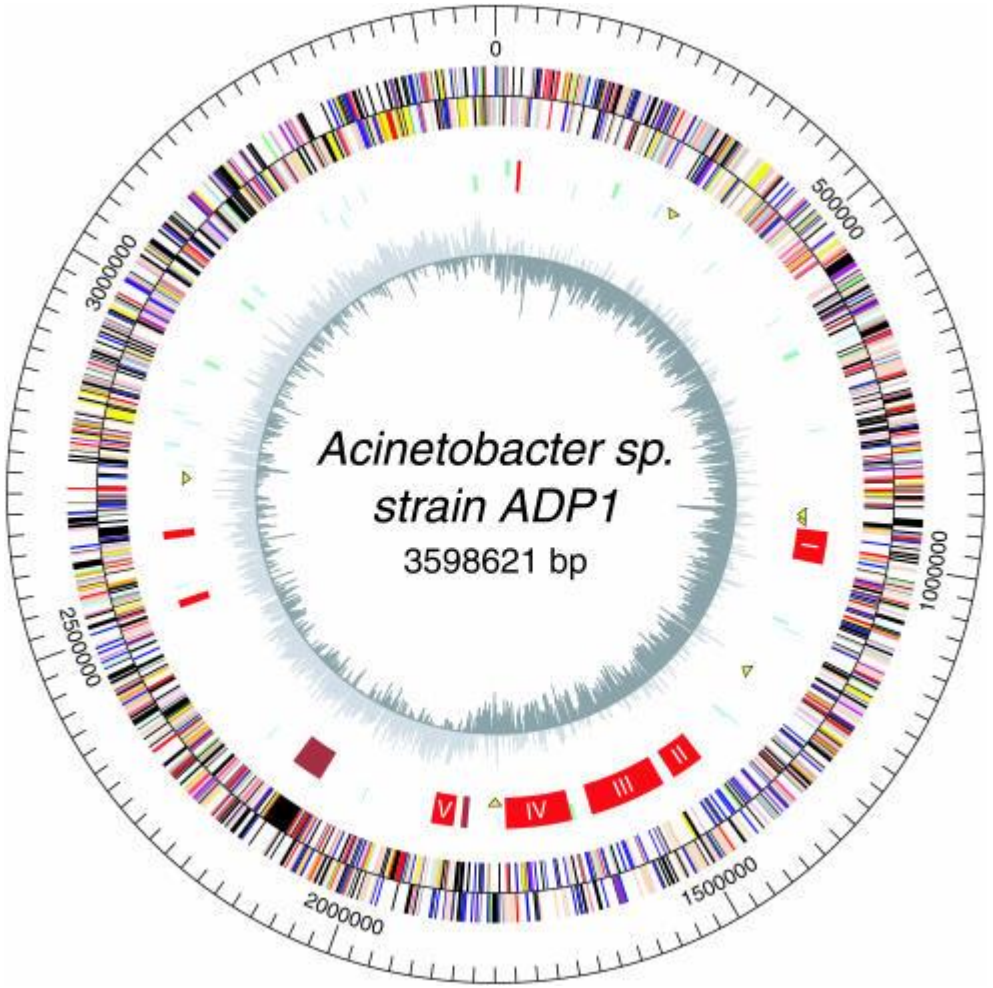
- Cofactors, Vitamins, Prosthetic Groups, Pigments (290)
- Cell Wall and Capsule (262)
- Virulence, Disease and Defense (138)
- Potassium metabolism (28)
- Photosynthesis (0)
- Miscellaneous (65)
- Phages, Prophages, Transposable elements, Plasmids (175)
- Membrane Transport (356)
- Iron acquisition and metabolism (25)
- RNA Metabolism (245)
- Nucleosides and Nucleotides (152)
- Protein Metabolism (322)
- Cell Division and Cell Cycle (45)
- Motility and Chemotaxis (80)
- Regulation and Cell signaling (175)
- Secondary Metabolism (26)
- DNA Metabolism (156)
- Fatty Acids, Lipids, and Isoprenoids (135)
- Nitrogen Metabolism (76)
- Dormancy and Sporulation (7)
- Respiration (192)
- Stress Response (196)
- Metabolism of Aromatic Compounds (27)
- Amino Acids and Derivatives (417)
- Sulfur Metabolism (58)
- Phosphorus Metabolism (54)
- Carbohydrates (750)

Example of genomics

Acinetobacter

Barbe V, Vallenet D, Fonknechten N, Kreimeyer A, Oztas S, Labarre L, Cruveiller S, Robert C, Duprat S, Wincker P, Ornston LN, Weissenbach J, Marlière P, Cohen GN, Médigue C.
Unique features revealed by the genome sequence of Acinetobacter sp. ADP1, a versatile and naturally transformation competent bacterium. Nucleic Acids Res. **2004 Oct** 28;32(19):5766-79. doi: 10.1093/nar/gkh910.

Acinetobacter baylyi



Acinetobacter baumannii

Smith MG, Gianoulis TA, Pukatzki S, Mekalanos JJ, Ornston LN, Gerstein M, Snyder M. New insights into Acinetobacter baumannii pathogenesis revealed by high-density pyrosequencing and transposon mutagenesis. Genes & development. 2007 Mar 1;21(5):601-14. **ATCC strain**

Other *Acinetobacter baumannii*

Iacono M, Villa L, Fortini D, Bordoni R, Imperi F, Bonnal RJ, Sicheritz-Ponten T, De Bellis G, Visca P, Cassone A, Carattoli A. Whole-genome pyrosequencing of an epidemic multidrug-resistant Acinetobacter baumannii strain belonging to the European clone II group. Antimicrobial agents and chemotherapy. 2008 Jul;52(7):2616-25. **Clinical carbapenem resistant ACICU strain**
Received 21 December 2007 **Revision received** 28 February 2008 **Accepted** 8 April 2008

Vallenet D, Nordmann P, Barbe V, Poirel L, Mangenot S, Bataille E, et al. (2008) Comparative Analysis of Acinetobacters: Three Genomes for Three Lifestyles. PLoS ONE 3(3): e1805, March 19 2008. **AYE and SDF**
Received: September 20, 2007; **Accepted:** February 9, 2008; **Published:** March 19, 2008

NEWS & ANALYSIS

GENOME WATCH

Opportunity knocks

Helena Seth-Smith & Alan Walker

This month's Genome Watch examines recent genome papers that provide insight into opportunistic pathogenesis.

Acinetobacter baumannii is emerging as an opportunistic pathogen that primarily infects immunocompromised patients in hospitals and particularly those in intensive-care units. The main clinical outcomes of infection (pneumonia, meningitis, bacteraemia and urinary-tract infections) are compounded by the problem of multidrug resistance. The natural reservoir of *A. baumannii* is unknown, but it can persist in hospital environments and is commonly found on the skin. *A. baumannii* has also been isolated from body lice, which suggests that it might use these insects as vectors[1].

Four strains of *A. baumannii* were recently sequenced: *A. baumannii* ATCC 17978, an historic strain from 1951 that was implicated in fatal meningitis in a 4-month-old baby; *A. baumannii* SDF, which was isolated from a human-body louse in France; *A. baumannii* AYE, which was isolated in 2001 during a nationwide outbreak in France; and *A. baumannii* ACICU, which was isolated from the cerebrospinal fluid of a patient during an outbreak in Italy in 2005 (REFS 2–4).

One of the most striking observations is the amount of apparently horizontally acquired DNA that is present in *A. baumannii* genomes. Members of the *Acinetobacter* genus can take up foreign DNA and incorporate it into their own genomes. *A. baumannii* ATCC 17978 carries 28 putative alien islands, which account for more than 17% of the predicted coding sequences (CDSs). The more recent strain *A. baumannii* ACICU possesses an additional 8 putative alien islands. The louse-associated strain *A. baumannii* SDF does not contain intact copies of all the genes that are necessary for natural transformation and, perhaps as a result, it contains fewer strain-specific CDSs than *A. baumannii* AYE. However, it contains 428 copies of insertion sequences, a massive expansion compared with *A. baumannii* AYE (33) and *A. baumannii* ACICU (14). This has resulted in a greater proportion of pseudogenes (more than 9%) and associated deletions, reducing the overall genome size (3.4 Mb compared with 3.9 Mb in the other strains) and perhaps restricting its host range.

Both *A. baumannii* ACICU and *A. baumannii* ATCC 17978 contain two plasmids, whereas *A. baumannii* SDF has three plasmids and *A. baumannii* AYE has 4 plasmids. Plasmid pACICU1 from *A. baumannii* ACICU might encode carbapenem resistance, but none of the other plasmids contains obvious resistance or virulence markers.

Generally considered a low-virulence species, candidate virulence factors have proved hard to identify. Many of the putative alien islands carry potential virulence genes, including type IV secretion systems, siderophores and haemolysins/haemagglutinins. Screens of transposon mutants of *A. baumannii* ATCC 17978 in both *Caenorhabditis elegans* and *Dictyostelium discoideum* identified several genes that are involved in virulence. However, some of these were strain specific. Surface structures may be important in the ability of *A. baumannii* to form biofilms, which could aid the survival of this organism in hospital environments.

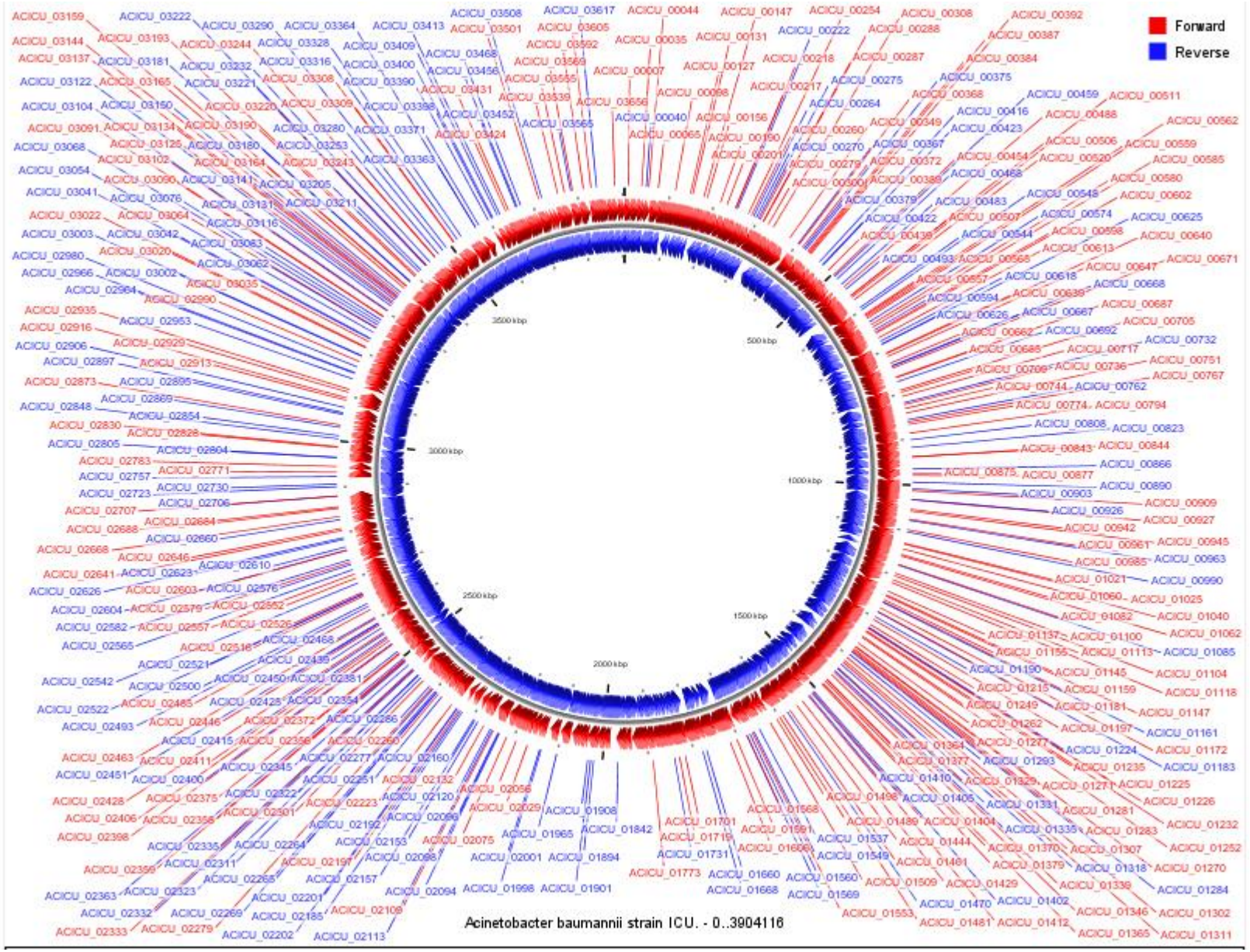
Glucokinase is absent from the sequenced genomes, which means that the strains cannot perform the first steps of glycolysis: an inability to grow on glucose as a sole carbon source has long been used to identify *Acinetobacter* species[3]. However, *A. baumannii* can catabolize a wide range of alternative carbon sources: *A. baumannii* ACICU seems to have the ability to use benzoate, citrate and glycerol, among other sources, and *A. baumannii* AYE has a substantial catabolic repertoire, including several uncharacterized oxygenases.

A. baumannii is intrinsically resistant to many antibiotics, putatively owing to the presence of many outer-membrane proteins and efflux pumps. Even *A. baumannii* ATCC 17978, which was isolated in 1951 and therefore had not been exposed to many antibiotics, possesses several efflux pumps, including 19 resistance–nodulation–division (RND) transporters, 3 major facilitator superfamily (MFS)

Four strains of *A. baumannii* were recently sequenced: *A. baumannii* ATCC 17978, an historic strain from 1951 that was implicated in fatal meningitis in a 4-month-old baby; *A. baumannii* SDF, which was isolated from a human-body louse in France; *A. baumannii* AYE, which was isolated in 2001 during a nationwide outbreak in France; and *A. baumannii* ACICU, which was isolated from the cerebrospinal fluid of a patient during an outbreak in Italy in 2005 (Refs 2,3,4).

Nature Reviews Microbiology **6**, pages 652–653 (2008)

Genome annotation

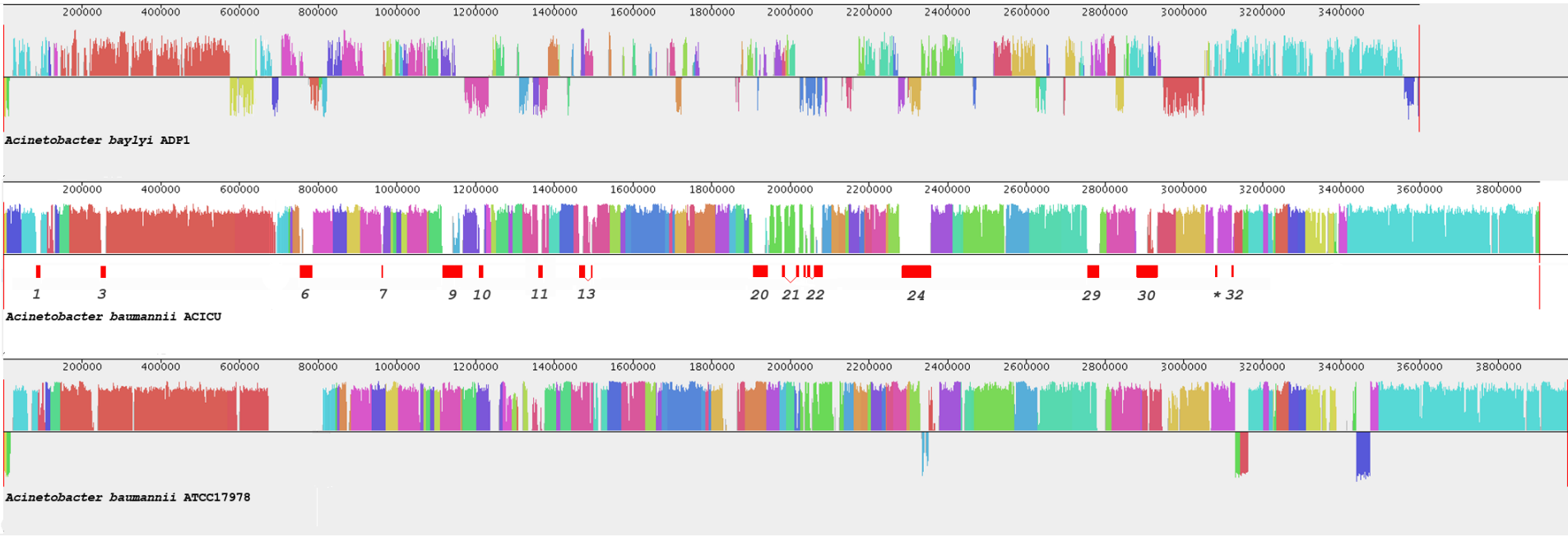


- *A. baumannii* ACICU contains a single circular chromosome of 3,904,116 bp and two plasmids (pACICU1 and pACICU2) of 28,279 and 64,366 bp, respectively; 3,758 genes were annotated in the ACICU chromosome, including 3,670 predicted protein-encoding CDSs, 64 tRNA genes, and 8 rRNA operons.

- Nearly 70% of the CDSs (n = 2,670) were assigned to a COG functional category; several genes belonged to more than one COG class.

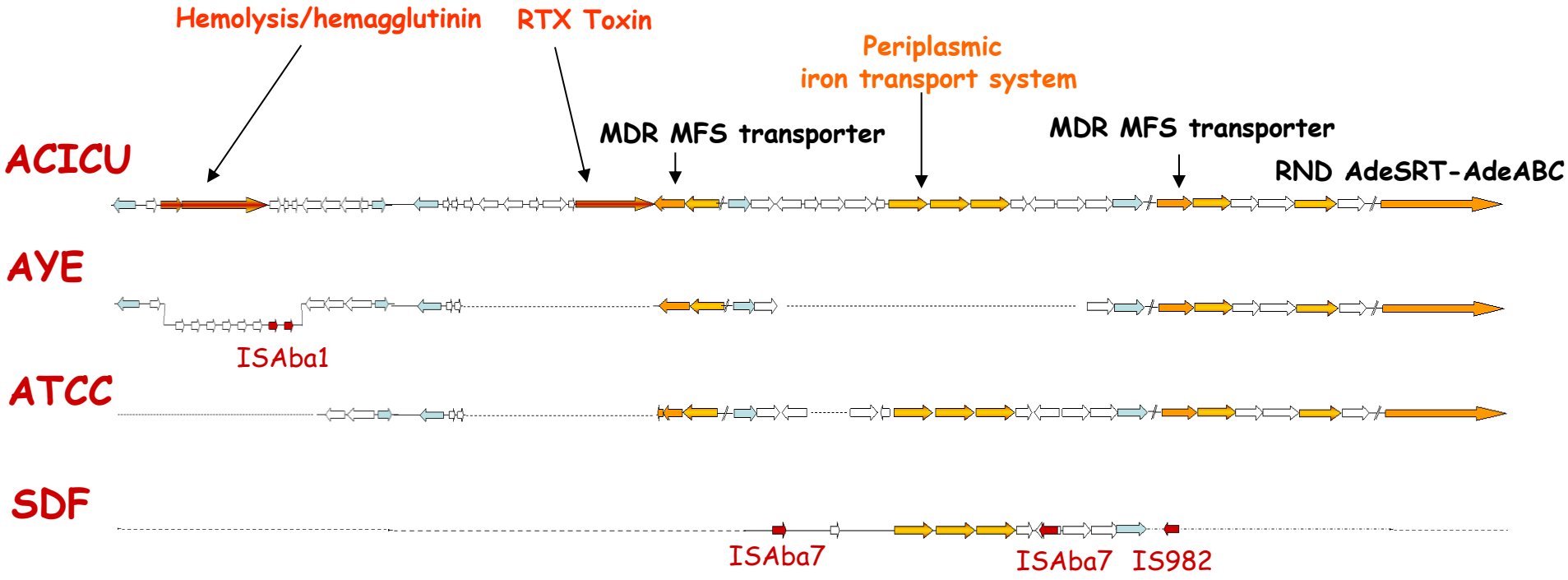
- The *A. baumannii* ACICU genome was initially compared with the unique genomes of *Acinetobacter* available*, A. baumannii* ATCC 17978 and *Acinetobacter baylyi* ADP1, with the aim of identifying novel genes related to virulence and drug resistance.
- Genome comparison showed 86.4% synteny with *A. baumannii* ATCC 17978 and 14.8% synteny with *A. baylyi* ADP1
- For many COG classes, the number of CDSs identified in ACICU largely exceeds the number identified in ATCC 17978, since in the latter strain only 60.1% of the genes were assigned to a COG class

Acinetobacter spp. synthenia



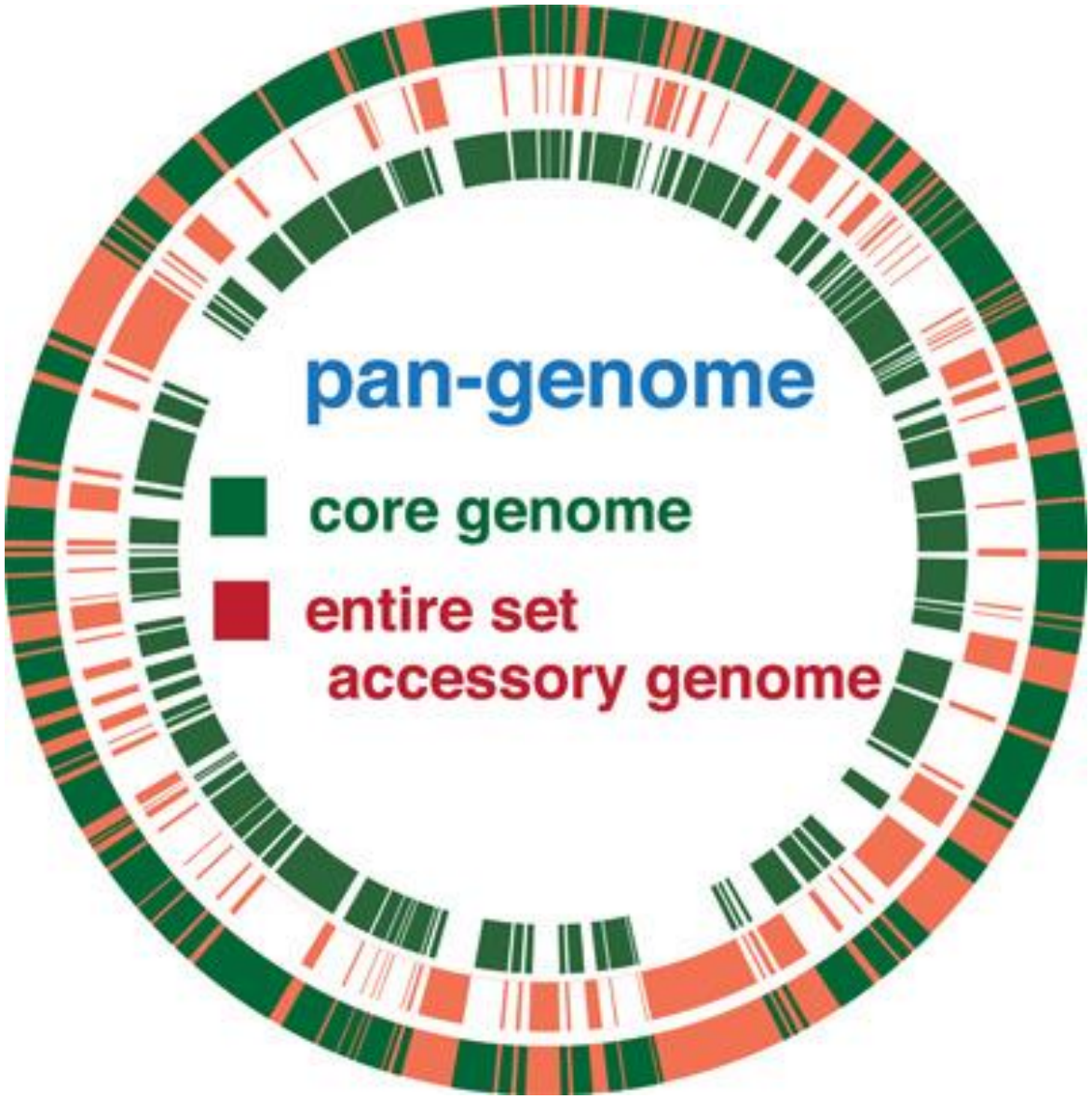
36 putative alien islands (pAs) were detected in the ACICU genome; 24 of these had previously been described in the ATCC 17978 genome, 4 are proposed here for the first time and are present in both ATCC 17978 and ACICU, and 8 are unique to the ACICU genome.

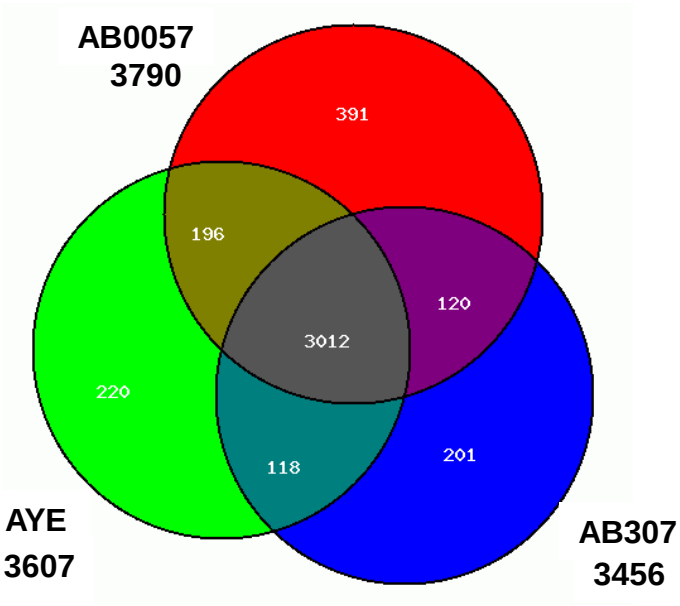
- ACICU also contains 14 ISs in the chromosome, including 7 ISAba125 elements, 4 ISAba2 elements, 2 IS26 elements, and 1 ISPu12 element, and 11 on plasmids, including 3 ISAba3 elements, 3 IS26 elements, 4 ISAba2 elements, and 1 ISAba125 element.
- The chromosome is composed of 0.38% short repetitive mini- and microsatellite DNA sequences

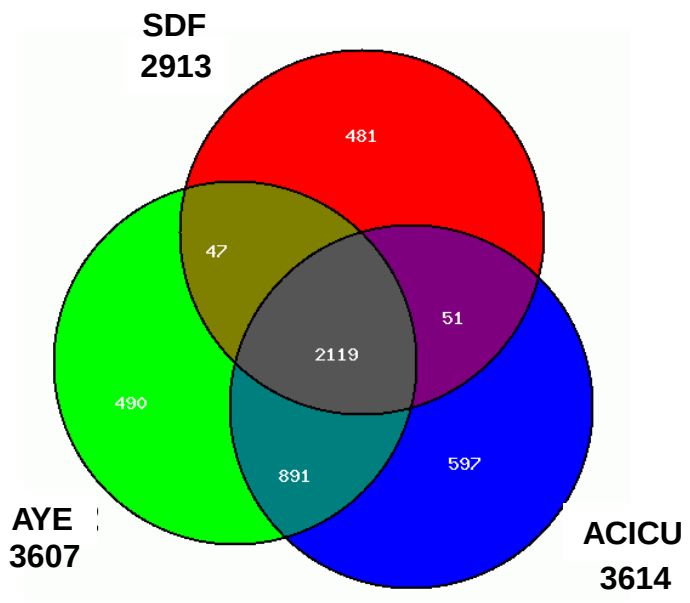

Hemolysis/hemagglutinin
RTX Toxin
Periplasmic
iron transport system
MDR MFS transporter
MDR MFS transporter
RND AdeSRT-AdeABC
ACICU
AYE
ISAba1
ATCC
SDF
ISAba7
ISAba7 IS982

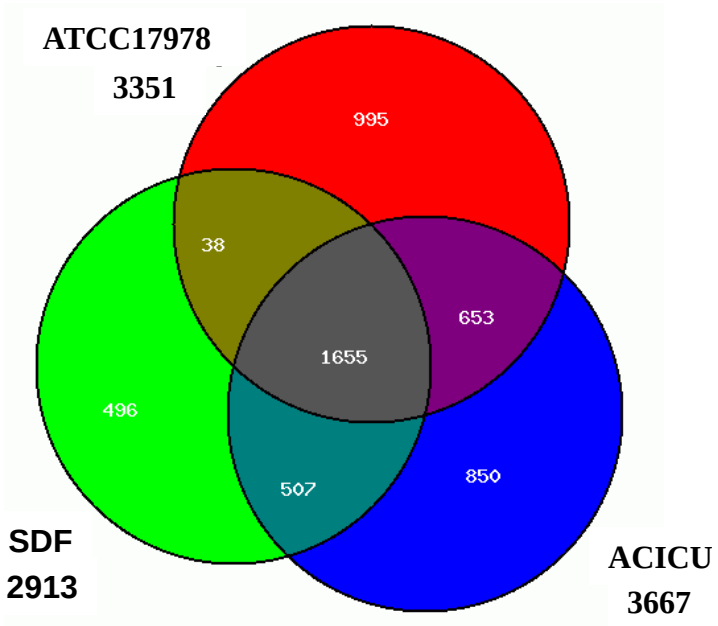
Coregenome Pangenome

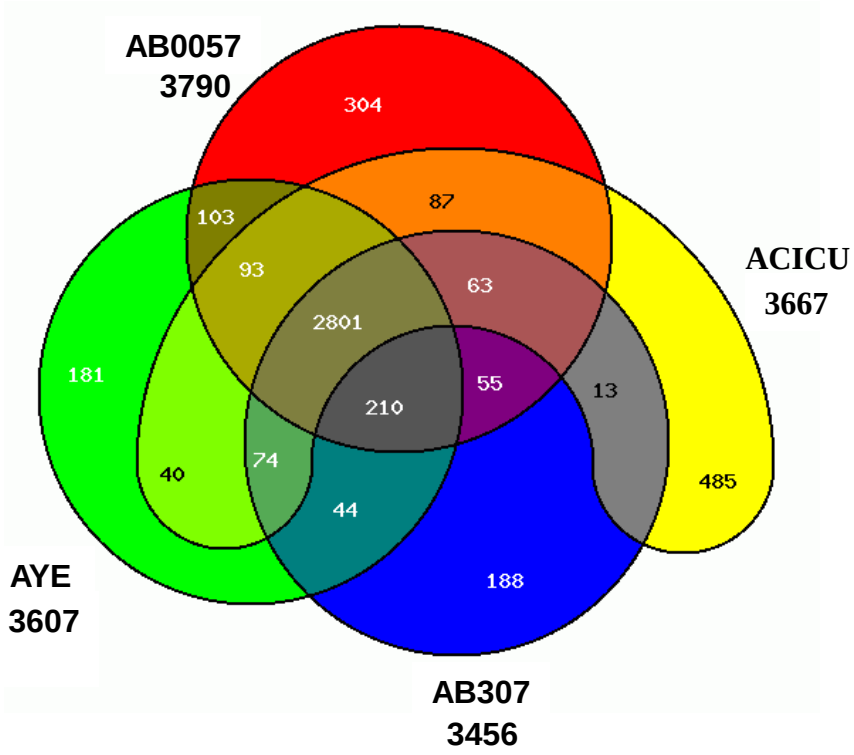
- Coregenome represents the genes present in all strains of a species = indispensable genome
- The accessory or flexible genome or dispensable genome: represents the genes that are present in some strains but not in the whole species
- The pangenome is the pool of all genes accessible to a species, both those of the coregenome and the accessory genes
- The ultimate goal is to understand the phenotype of a species, but also the phenotypic differences between isolates of the same species


pan-genome
core genome
entire set
accessory genome

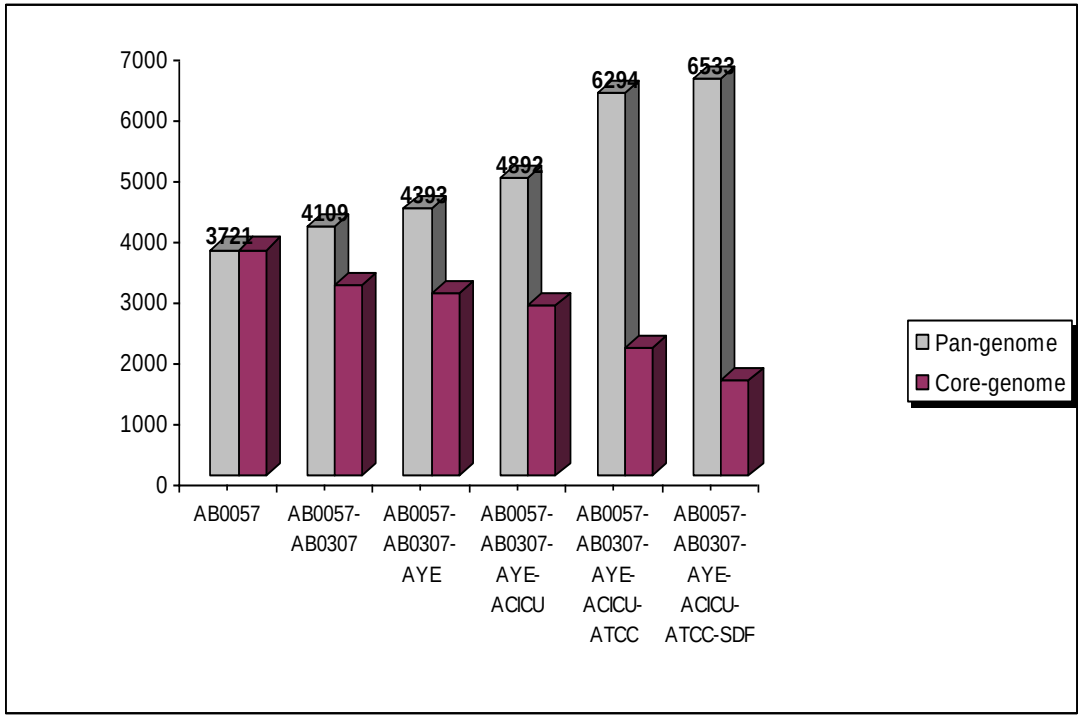


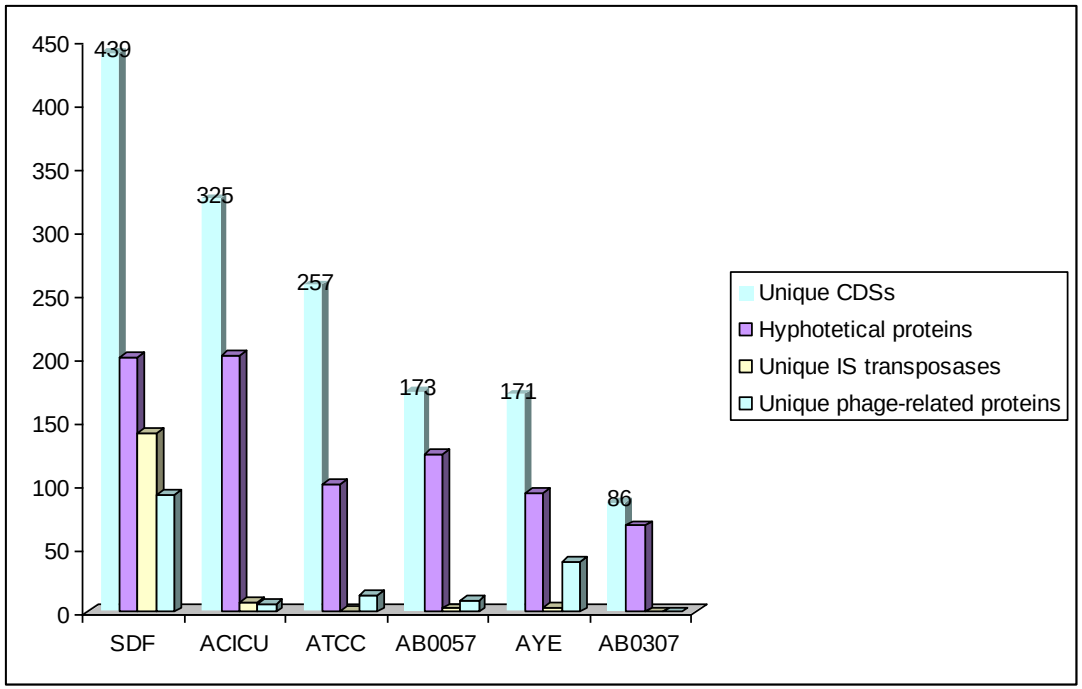






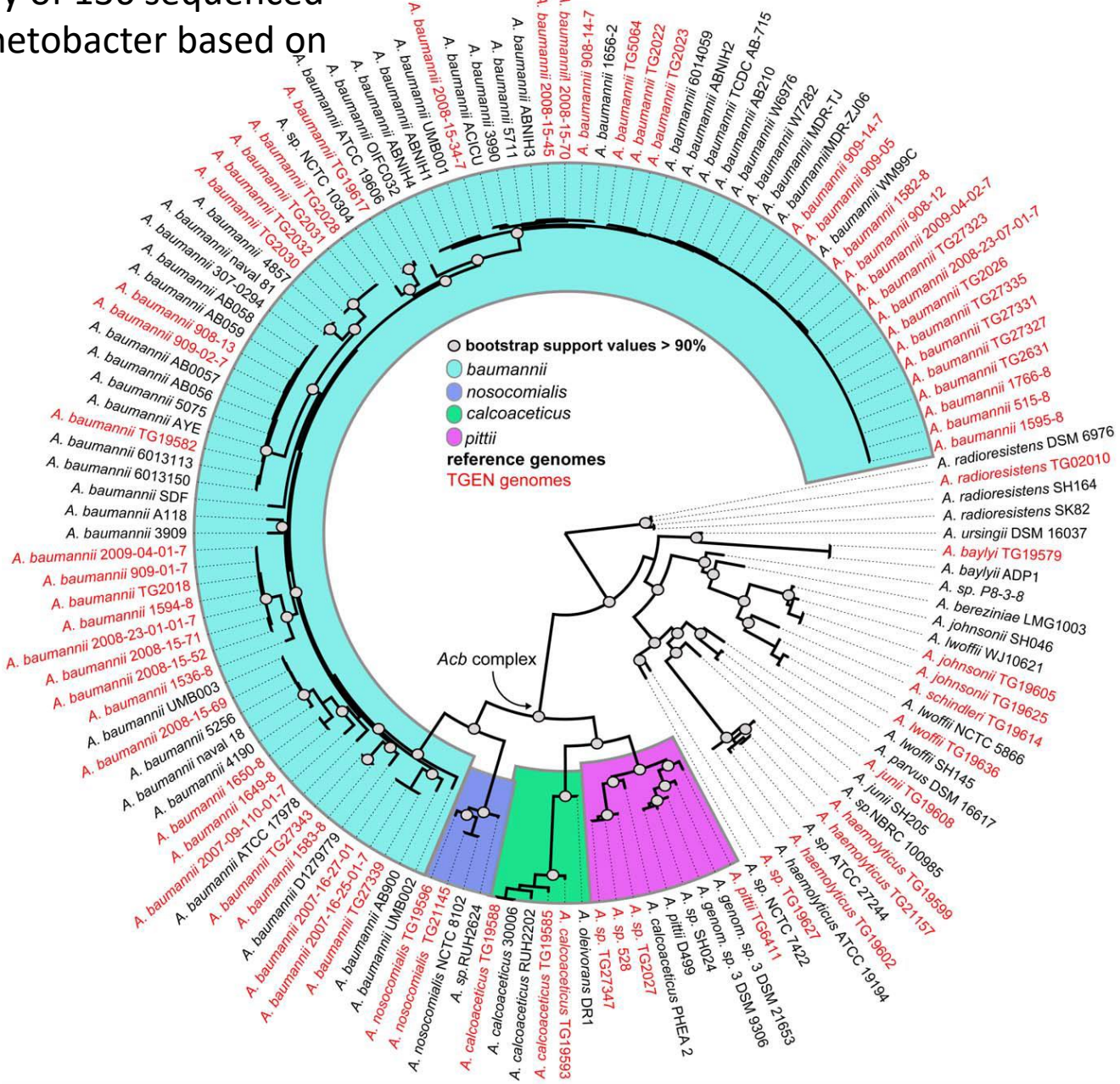
The genomics of Acinetobacter baumannii: insights into genome plasticity, antimicrobial resistance and pathogenicity. **Imperi** et al IUBMB Life. 2011 Dec;63(12):1068-74





The genomics of Acinetobacter baumannii: insights into genome plasticity, antimicrobial resistance and pathogenicity. **Imperi** et al IUBMB Life. 2011 Dec;63(12):1068-74

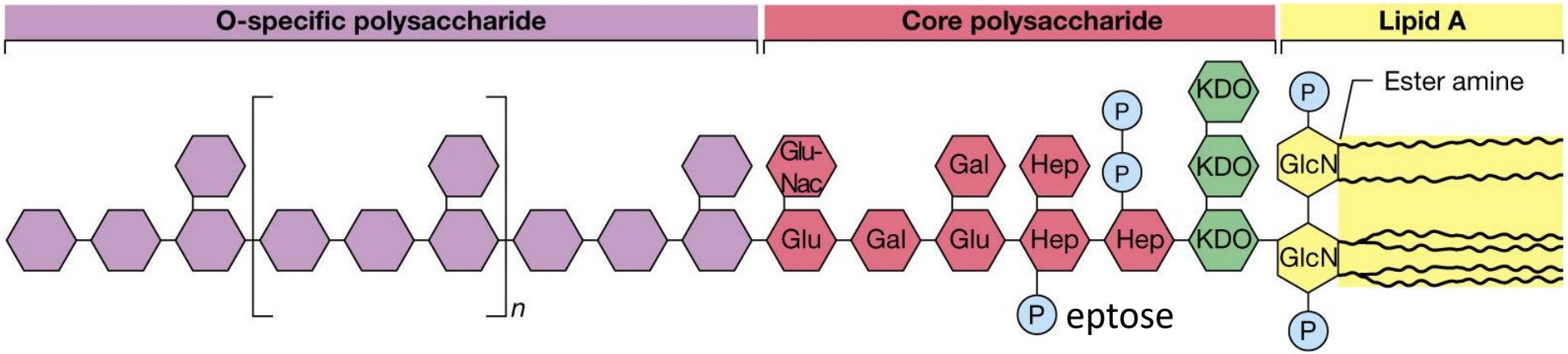
A whole genome phylogeny of 136 sequenced genomes in the genus Acinetobacter based on SNPs



Sahl et al. (2013) Evolution of a Pathogen: A Comparative Genomics Analysis Identifies a Genetic Pathway to Pathogenesis in Acinetobacter. PLoS ONE 8(1): e54287.

identification of genes in the genome

LPS



Chain of tetra or penta-saccharides

N-Acetylgluc

2-keto-3deoptonatec

dTDP-glucose 4,6-dehydratase
Glucose-1-phosphate thymidylyltransferase
dTDP-6-deoxy-3,4-keto-hexulose isomerase
dTDP-6-deoxy-3,4-keto-hexulose isomerase
Aminotransferase
O-antigen flippase
O-antigen polymerase
Glycosyltransferase
Serine transferase
Amylovoran biosynthesis protein
E. coli O5
JUMPstart
rmlB
rmlA
fdtA
fdtA
fdtB
wzx
wzy
wbuM
ORF9
wbuO
amsE
gnd
E. coli O76
JUMPstart
ORF1
wzx
ORF3
wzy
wzy
wcaN
rfaG
galE
galE
gnd
bp
500
1000
1500
2000
2500
3000
3500
4000
4500
5000
5500
6000
6500
7000
7500
8000
8500
9000
9500

Serotyping


SALMONELLA
SG 6
SALMONELLA
SG 2
d + i + e,h
SALMONELLA
O:9
SALMONELLA
H:a
SALMONELLA
O:2
SALMONELLA
H:2
SALMONELLA
SG 1
a + b + c + z10
SALMONELLA
H:m
SALMONELLA
O:4
SALMONELLA
H:i
SALMONELLA
O:46

PLOS ONE

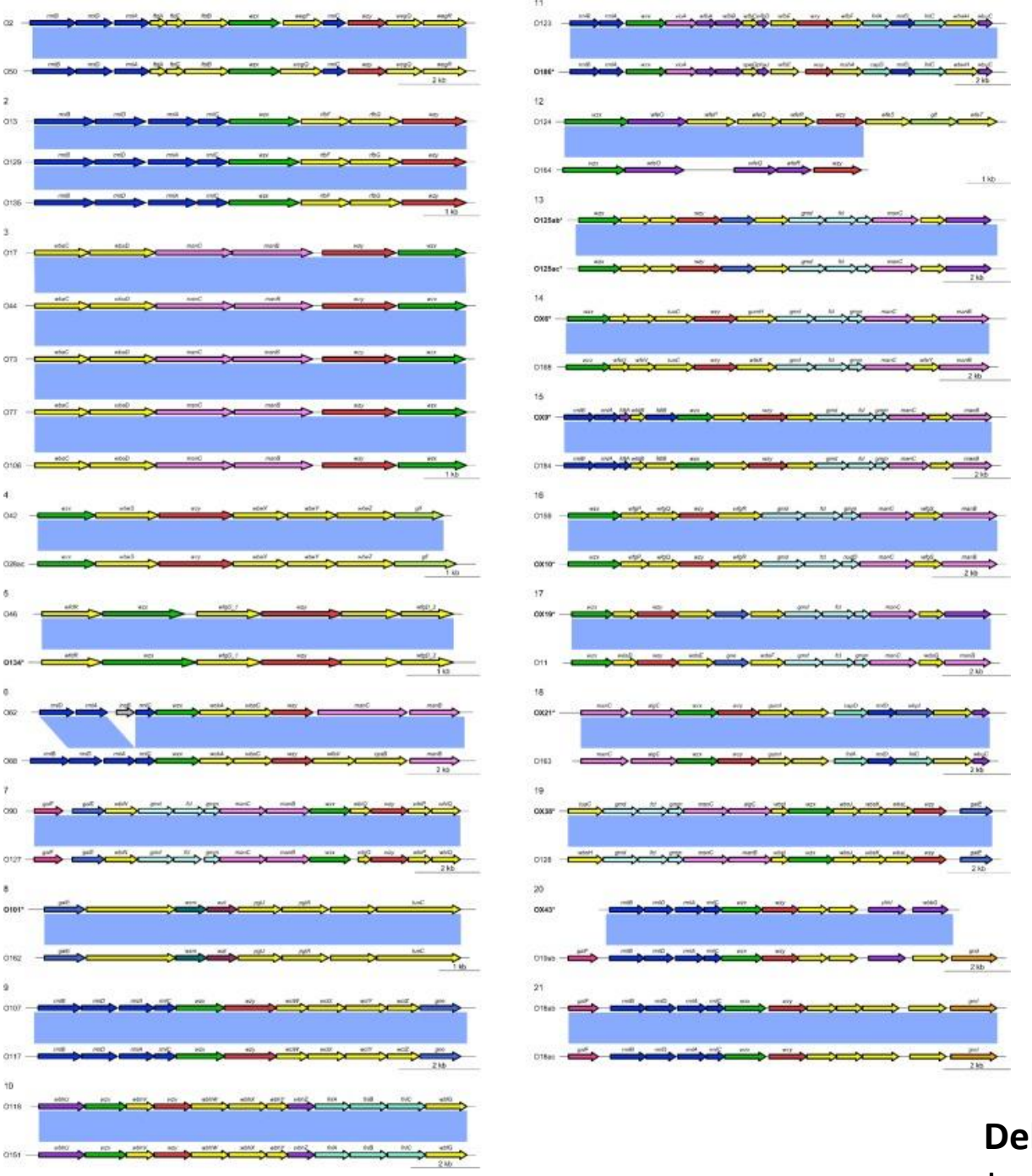
RESEARCH ARTICLE

Comparison of O-Antigen Gene Clusters of All O-Serogroups of *Escherichia coli* and Proposal for Adopting a New Nomenclature for O-Typing

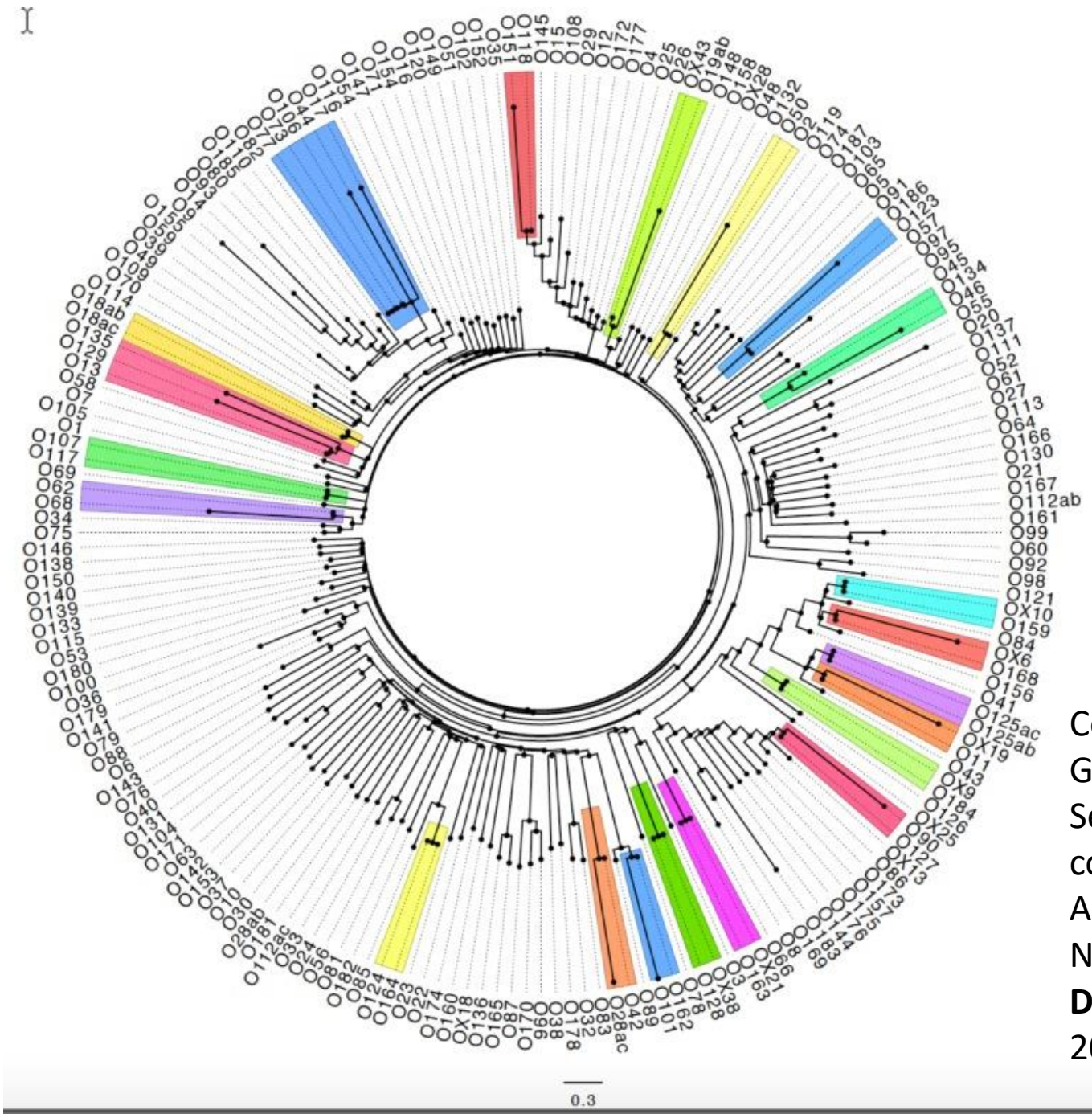
Chitrita DebRoy[1]*, Pina M. Fratamico[2], Xianghe Yan[2], GianMarco Baranzoni[2], Yanhong Liu[2], David S. Needleman[2], Robert Tebbs[3], Catherine D. O'Connell[3], Adam Allred[3], Michelle Swimley[3], Michael Mwangi[1], Vivek Kapur[1], Juan A. Raygoza Garay[1], Elisabeth L. Roberts[1], Robab Katani[1]

1 *E. coli* Reference Center, Department of Veterinary and Biomedical Sciences, The Pennsylvania State University, University Park, Pennsylvania, United States of America, **2** Eastern Regional Research Center, Agricultural Research Service, U.S. Department of Agriculture, Wyndmoor, Pennsylvania, United States of America, **3** Animal Health & Food Safety, Life Sciences Solutions, Thermo Fisher Scientific, Austin, Texas, United States of America

* rcd3@psu.edu

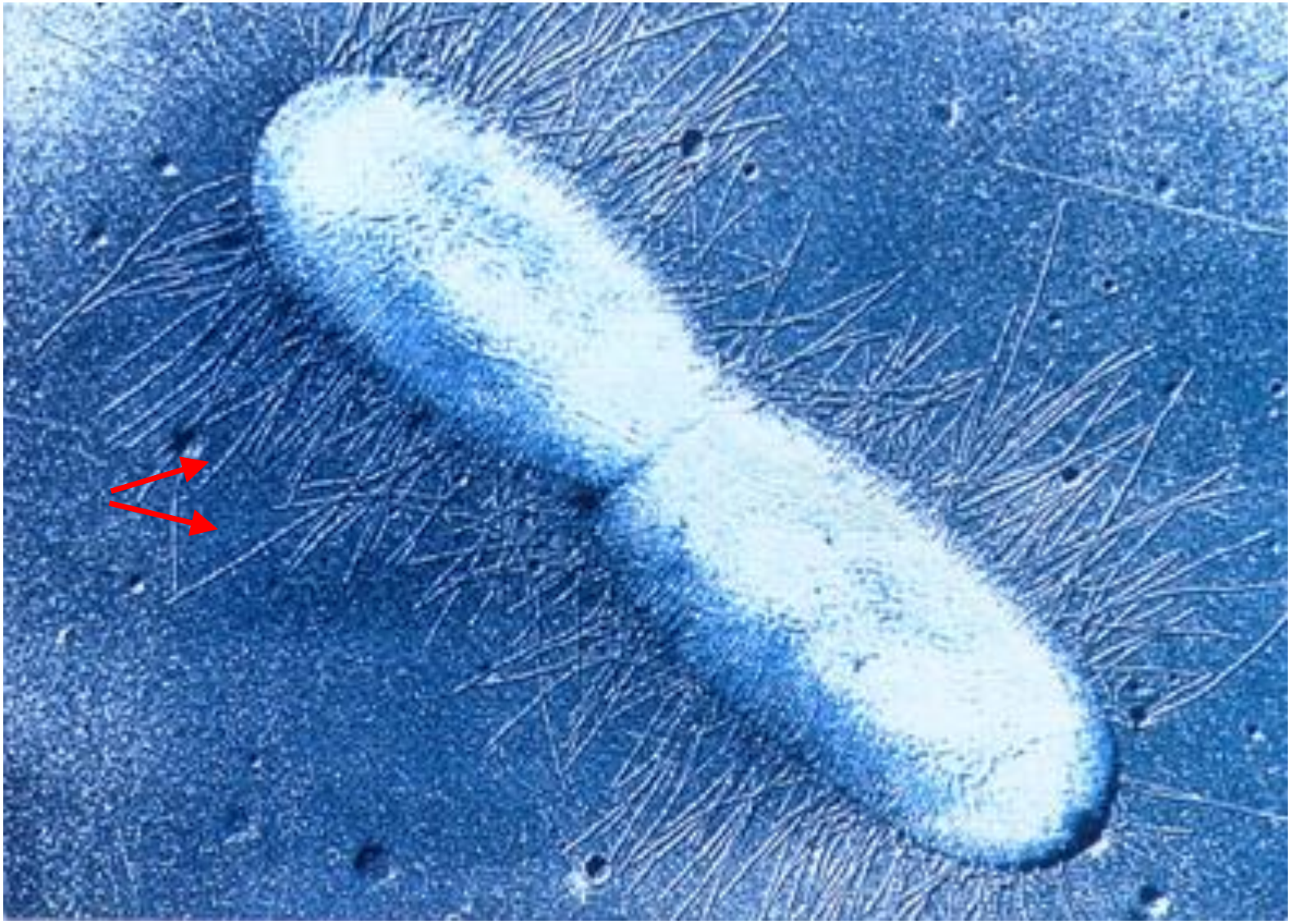


DebRoy C et al. PLoS One. 2016 Jan 29;11(1):e0147434

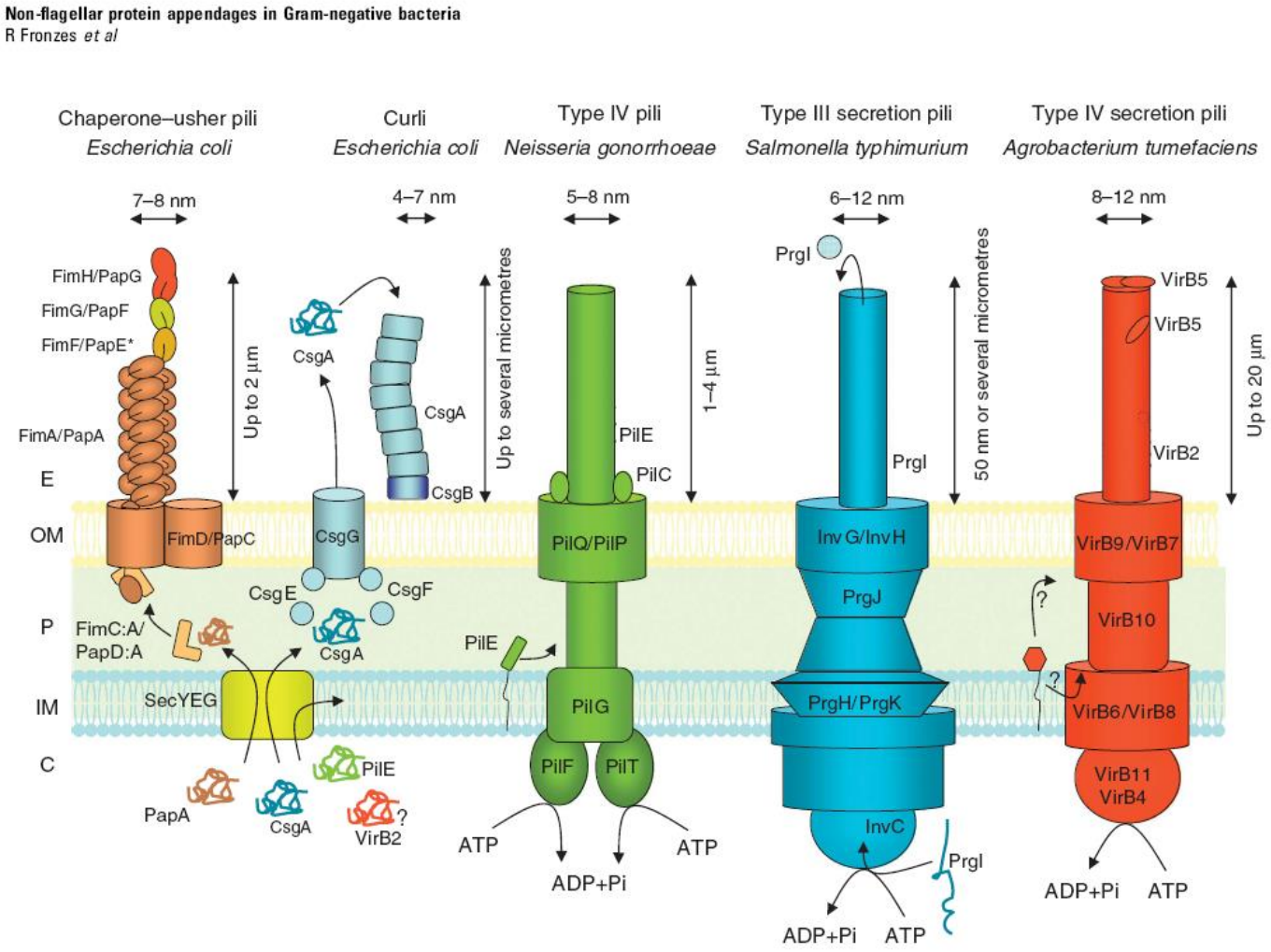


Comparison of O-Antigen Gene Clusters of All O-Serogroups of Escherichia coli and Proposal for Adopting a New Nomenclature for O-Typing. **DebRoy** C et al. PLoS One. 2016 Jan 29;11(1):e0147434

Pili



Pili and fimbriae



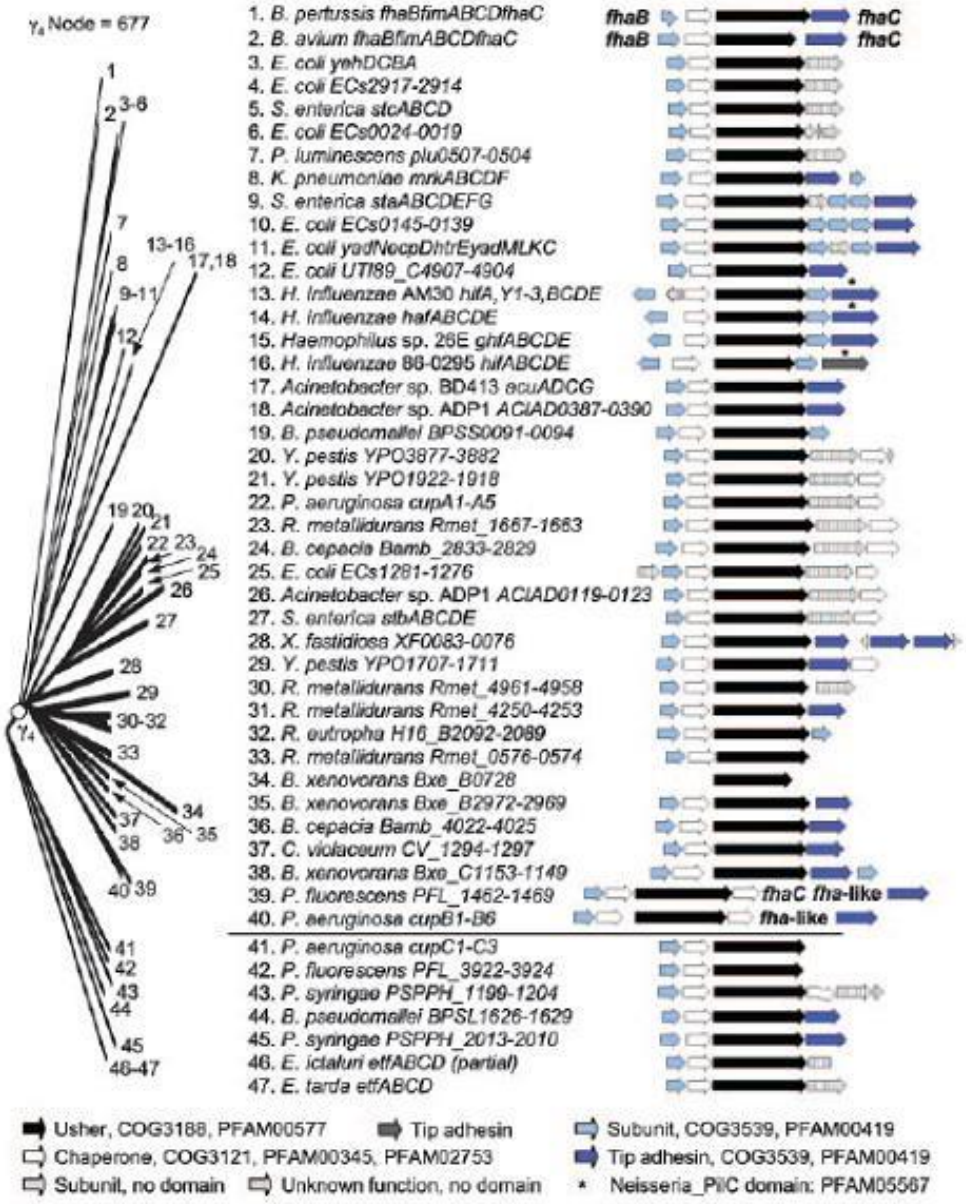
fimbriae

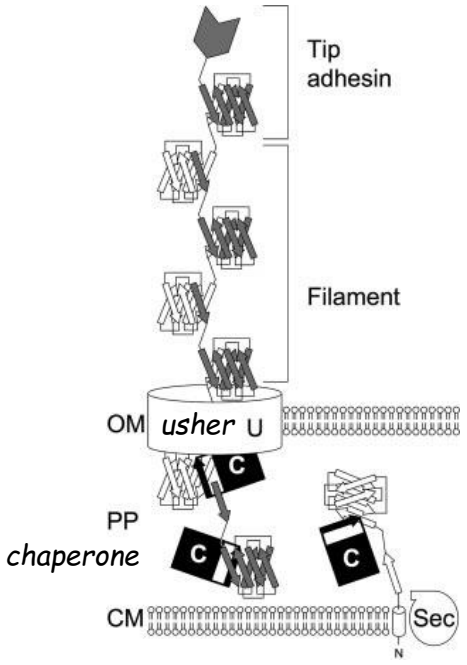
Microbiology and Molecular Biology Reviews, Dec. 2007, p. 551–575
1092-2172/07/$08.00+0 doi:10.1128/MMBR.00014-07

Vol. 71, No. 4

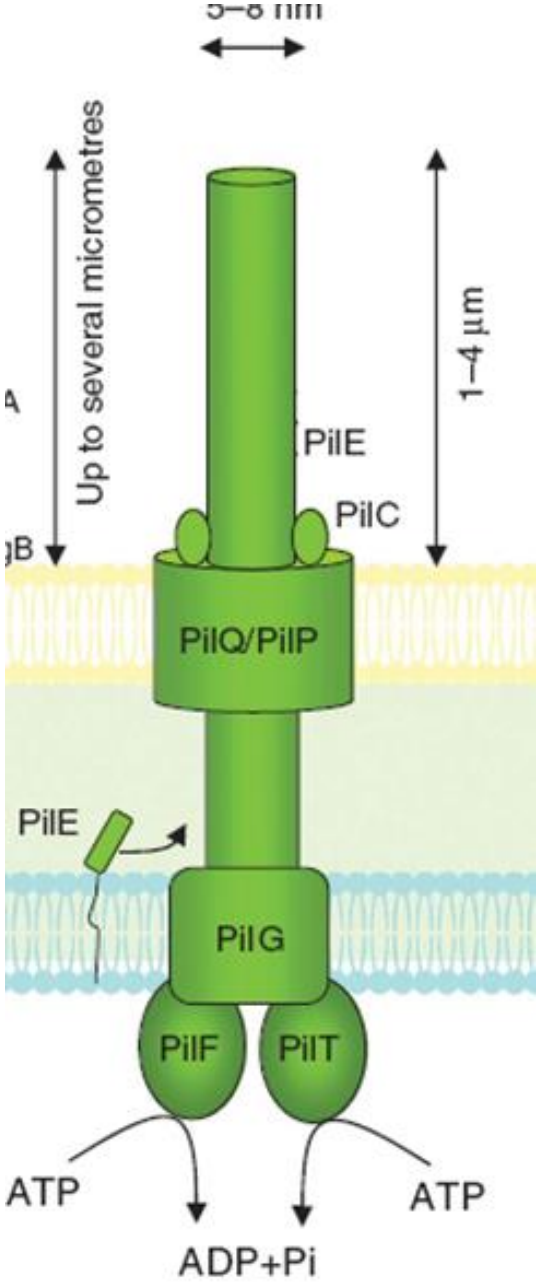
'olution of the Chaperone/Usher Assembly Pathway: Fimbrial Classification Goes Greek†

Sean-Paul Nuccio and Andreas J. Bäumler*



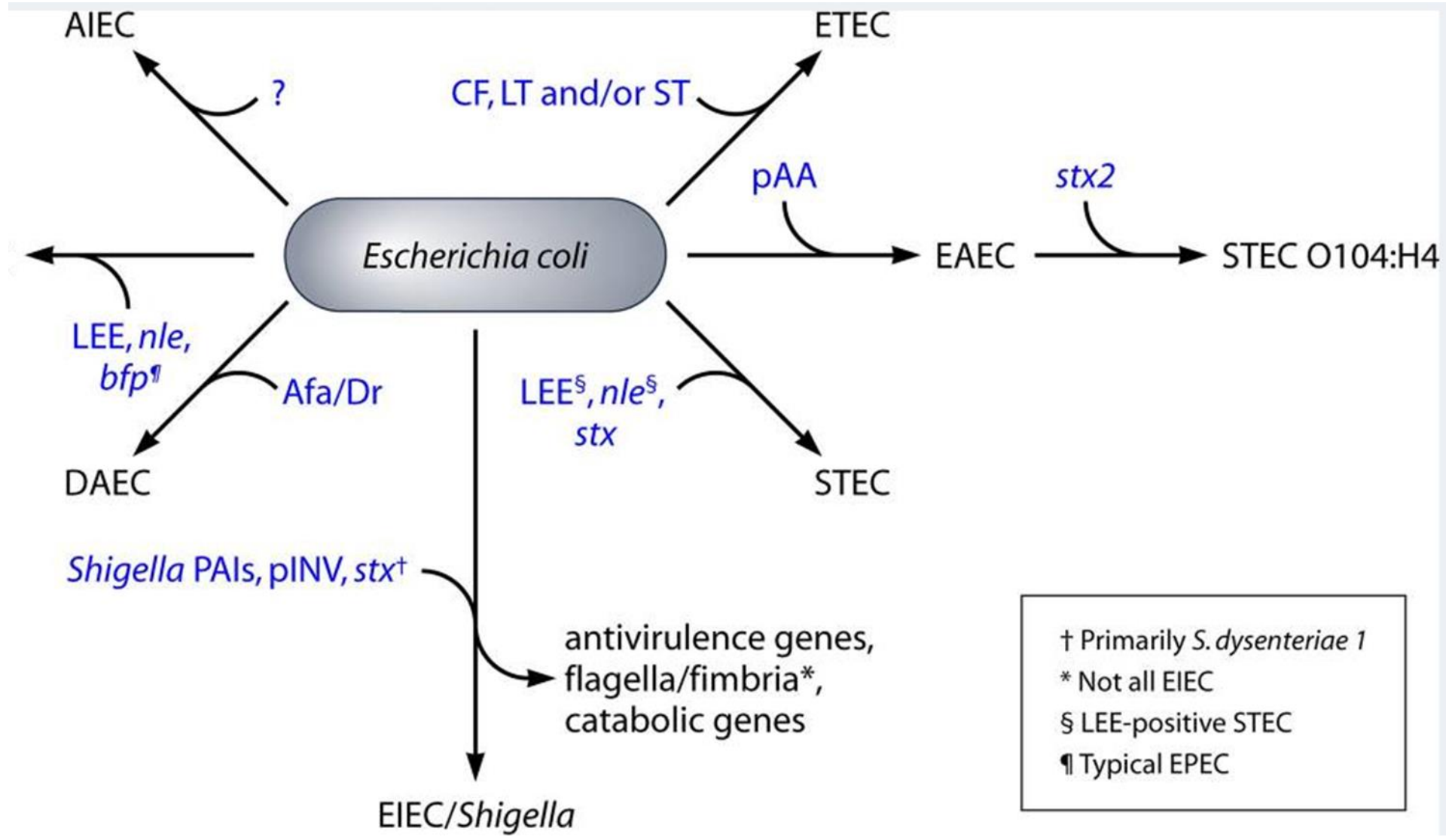


Type IV pilus



Escherichia coli

K12-MG1655 (no pathogenic)
Enterohemorrhagic EHEC (O157:H7, STX)
Uropathogenic UPEC (pili P)
Enteropathogenic EPEC (T3SS)
Enterotoxigenic ETEC (LT e adesine)
Enteroaggregative EAEC (fimbriae)



35 Kb locus of enterocyte effacement (LEE); bundle-forming pilus gene (*bfp*); Shiga toxin genes (stx_1, stx_2,); Heat-Labile toxin (LT); Heat-Stable toxin (ST); colonization factors (CFs); acquired fimbriae that enhance adherence (Afa/Dr); pAA plasmid; pINV plasmid; chromosomal pathogenicity islands (PAIs) Croxen et al. CMR 2013

nature communications

Article

https://doi.org/10.1038/s41467-023-43854-3

Global emergence of a hypervirulent carbapenem-resistant *Escherichia coli* ST410 clone

Received: 10 May 2023

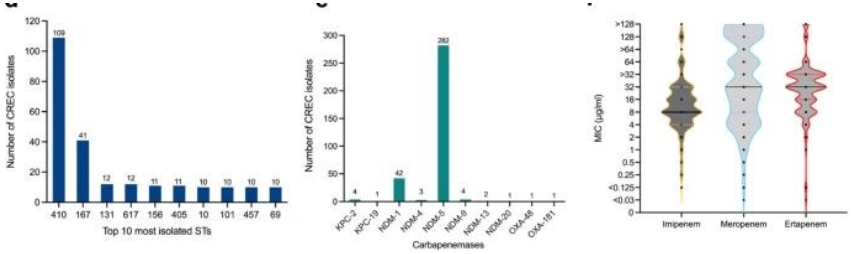
Accepted: 22 November 2023

Published online: 12 January 2024

Check for updates

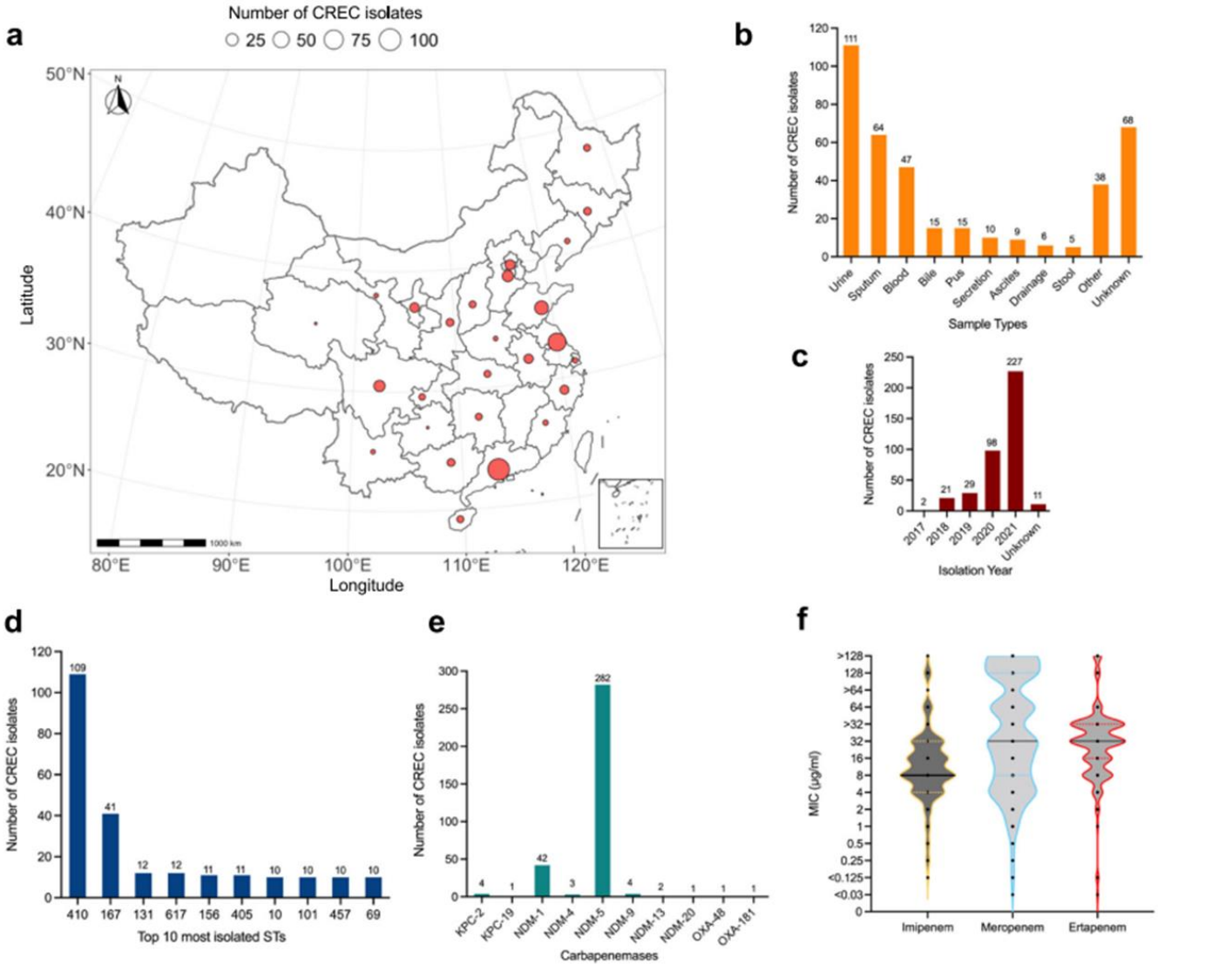
Xiaoliang Ba[1,8], Yingyi Guo[2,8], Robert A. Moran[3], Emma L. Doughty[3], Baomo Liu[4], Likang Yao[2], Jiahui Li[2], Nanhao He[2], Siquan Shen[5,6], Yang Li[7], Willem van Schaik[3], Alan McNally[3], Mark A. Holmes[1,9] ✉ & Chao Zhuo[2,9] ✉

1Department of Veterinary Medicine, University of Cambridge, Cambridge, United Kingdom. 2State Key Laboratory of Respiratory Disease, First Affiliated Hospital of Guangzhou Medical University, Guangzhou, China. 3Institute of Microbiology and Infection, College of Medical and Dental Sciences, University of Birmingham, Birmingham B15 2TT, United Kingdom. 4Department of Pulmonary and Critical Care Medicine, The First Affiliated Hospital of Sun Yat-sen University, Guangzhou, China. 5Institute of Antibiotics, Huashan Hospital, Fudan University, Shanghai, China. 6Key Laboratory of Clinical Pharmacology of Antibiotics,MinistryofHealth,Shanghai,China.7DepartmentofClinicalLaboratory,Children'sHospitalofSoochowUniversity,Suzhou,Jiangsu,China.8These authors contributed equally: Xiaoliang Ba, Yingyi Guo. 9These authors jointly supervised this work: Mark A. Holmes, Chao Zhuo. e-mail:mah1@cam.ac.uk; chaosheep@sina.com



Carbapenem-resistant *Escherichia coli* (CREC) ST410

- CREC prevalence increased in Chinese hospitals between 2017 and 2021.
- ST410 is the most frequent. Four groups of *E. coli* ST410 were identified in this children's hospital,
- **Genomic analysis identifies a hypervirulent CREC ST410 clone, B5/H24RxC**
- It may have emerged from the previously characterized B4/H24RxC in 2006
- Compared with B4/H24RxC, **B5/H24RxC lacks the $bla_{OXA-181}$-bearing X3 plasmid, but carries a F-type plasmid containing bla_{NDM-5}**
- **Most of B5/H24RxC also carry a high pathogenicity island YBT and a novel O-antigen gene cluster**
- B5/H24RxC grew faster *in vitro* and is more virulent *in vivo*
- Globally disseminated hypervirulent CREC clone, highlights the ongoing evolution of ST410 towards increased resistance and virulence.



Carbapenem-resistant *Escherichia coli* (CREC) ST410

- CREC prevalence increased in Chinese hospitals between 2017 and 2021.
- ST410 is the most frequent. Four groups of *E. coli* ST410 were identified in this children's hospital,

Genomic analysis identifies a hypervirulent CREC ST410 clone, B5/H24RxC

It may have emerged from the previously characterized B4/H24RxC in 2006

Article https://doi.org/10.1038/s41467-023-43854-3

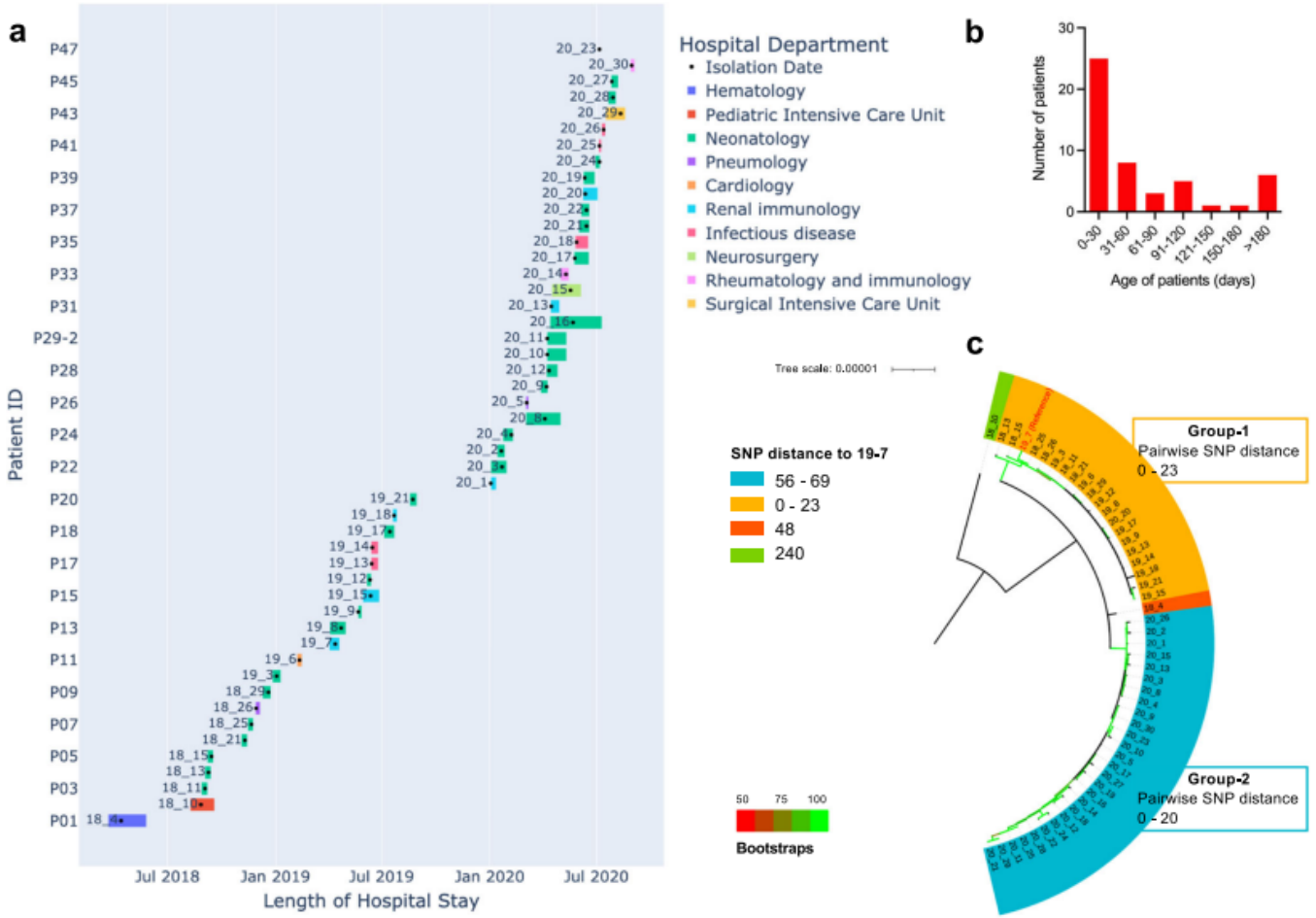
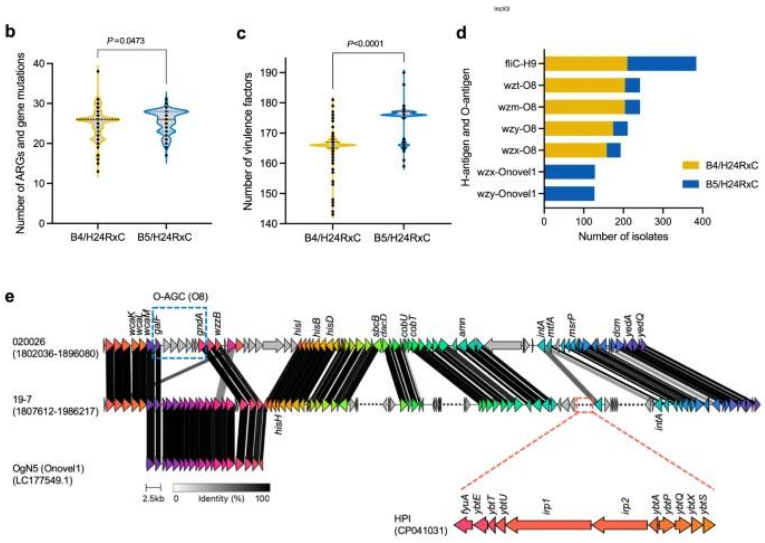
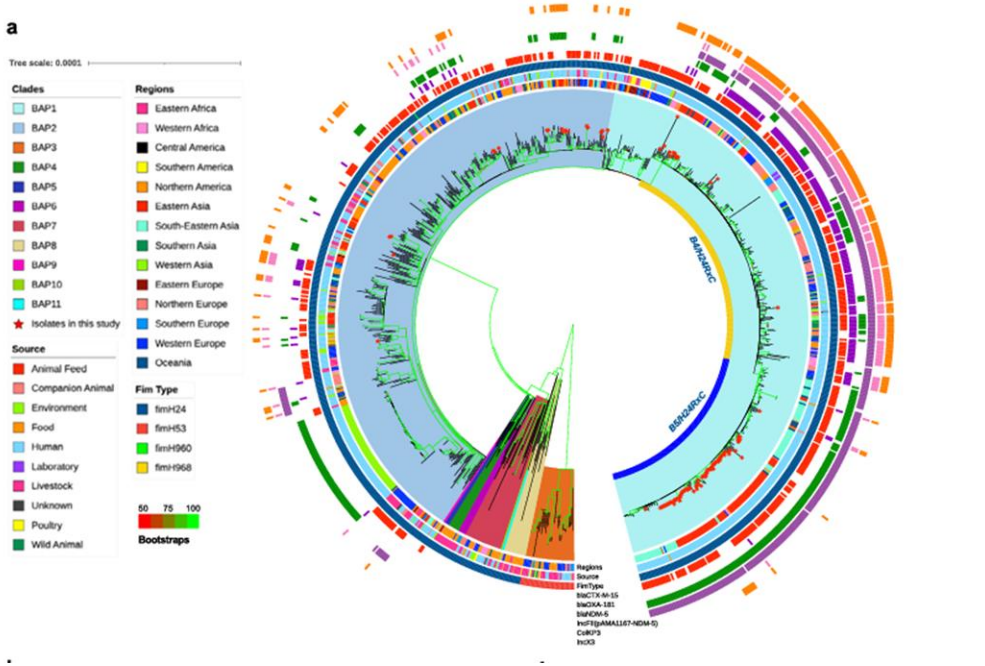
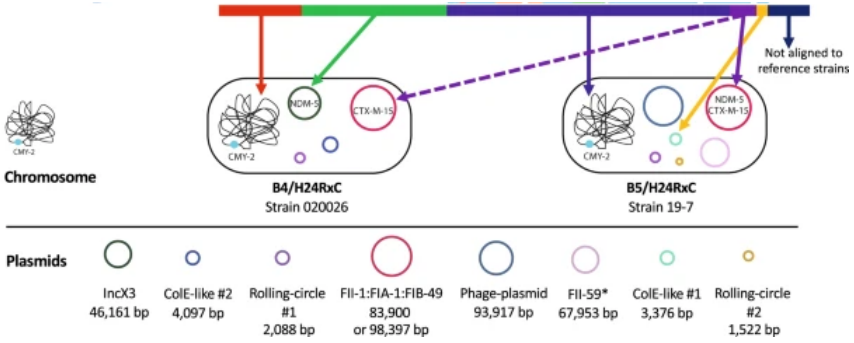


Fig. 2 | Outbreaks of a ST410 lineage in a children's hospital. a Gantt plot showing the length of hospital stay of the patients in the children's hospital in eastern China. Patient ID are presented on the *y* axis and the length of stay of each patient is represented with coloured bars. A black dot within the coloured bars indicates the time of the isolation of the isolates. **b** Bar chart showing the age distribution of the patients. **c** Maximum-likelihood core-genome SNP phylogeny of the 49 ST410 CREC isolates in the children's hospital. Colours indicates the SNP distance to the reference genome 19-7. Bootstrap values are represented by gradient colours. Source data are provided as a Source Data file.





B5/H24RxC had 176 putative virulence genes, B4/H24RxC had 166
The difference: the presence of the **high pathogenicity island (Yersiniabactin)** originated from in Yersinia enterocolitica (*fyuA, irp1, irp2, ybtA, ybtE, ybtP, ybtQ, ybtS, ybtT, ybtU,* and *ybtX)*
And the O-antigen genes were associated with B4/H24RxC (*wzm*/*wzt* O8) **changed with *wzx*/*wzy* Onovel1 in B5/H24RxC**
Compared with B4/H24RxC, **B5/H24RxC lacks the *bla*$_{OXA-181}$-bearing X3 plasmid, but carries a F-type plasmid containing *bla*$_{NDM-5}$**



- B5/H24RxC grew faster *in vitro* and is more virulent *in vivo*

Time of the most recent common ancestor (TMRCA) of different phylogenetic groups

A mutation rate of 6.42E-7 SNPs per site per was estimated. The analysis estimated the age of the ST410 lineage to be approximately 205 years, with a TMRCA of around 1816. The B4/H24RxC ancestor was estimated to have originated in 2003. The TMRCA of B5/H24RxC was estimated at around May 2006

https://doi.org/10.1038/s41467-023-43854-3

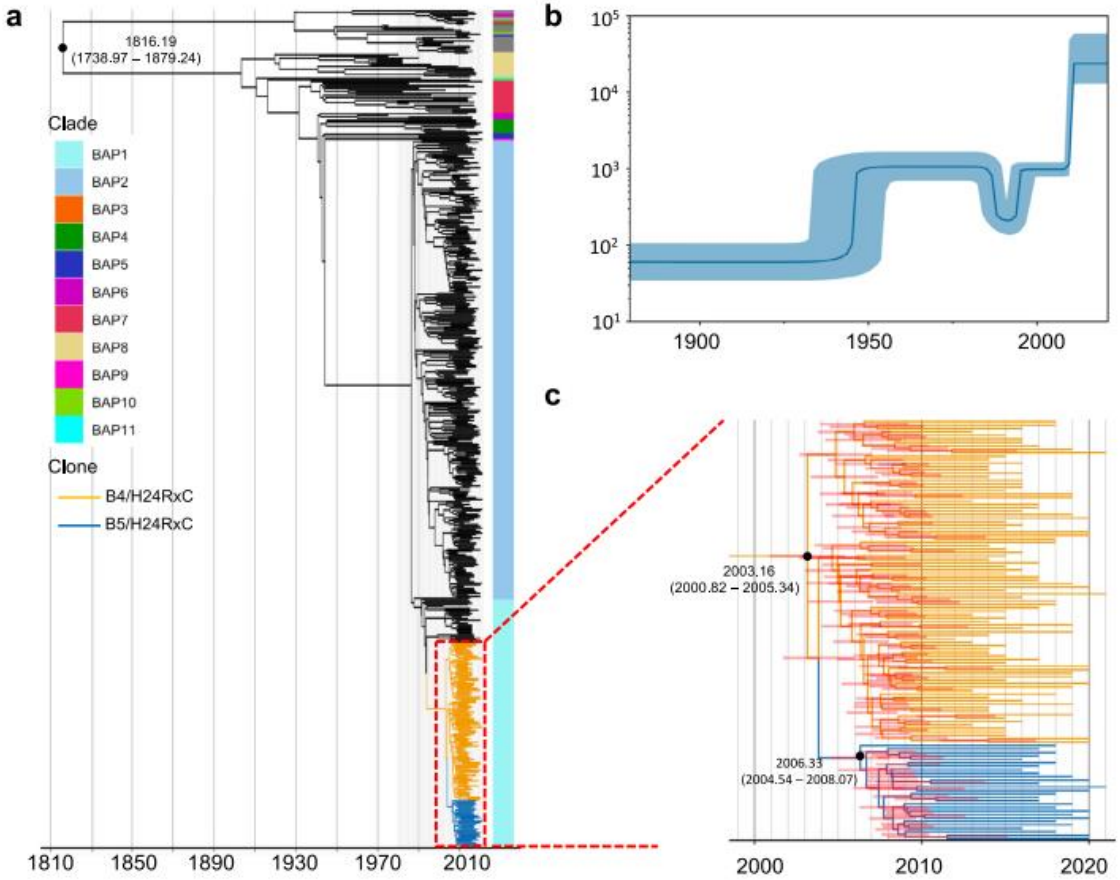


Fig. 6 | Coalescence-based analysis of *E. coli* ST410. a A time-calibrated phylogeny was reconstructed using BEAST2.0 based on the nonrecombinant SNPs for the 500 selected *E. coli* ST410. The MDR clone B4/H24RxC and B5/H24RxC are coloured in orange and blue, respectively. **b** The Bayesian skyline plot illustrates the predicted demographic changes of the ST410 clades. The thick solid line represents the median estimate of the effective population size, with 95% confidence interval shown in lighter blue area. **c** An enlarged phylogenetic tree showing the B4/H24RxC and B5/H24/RxC clones.