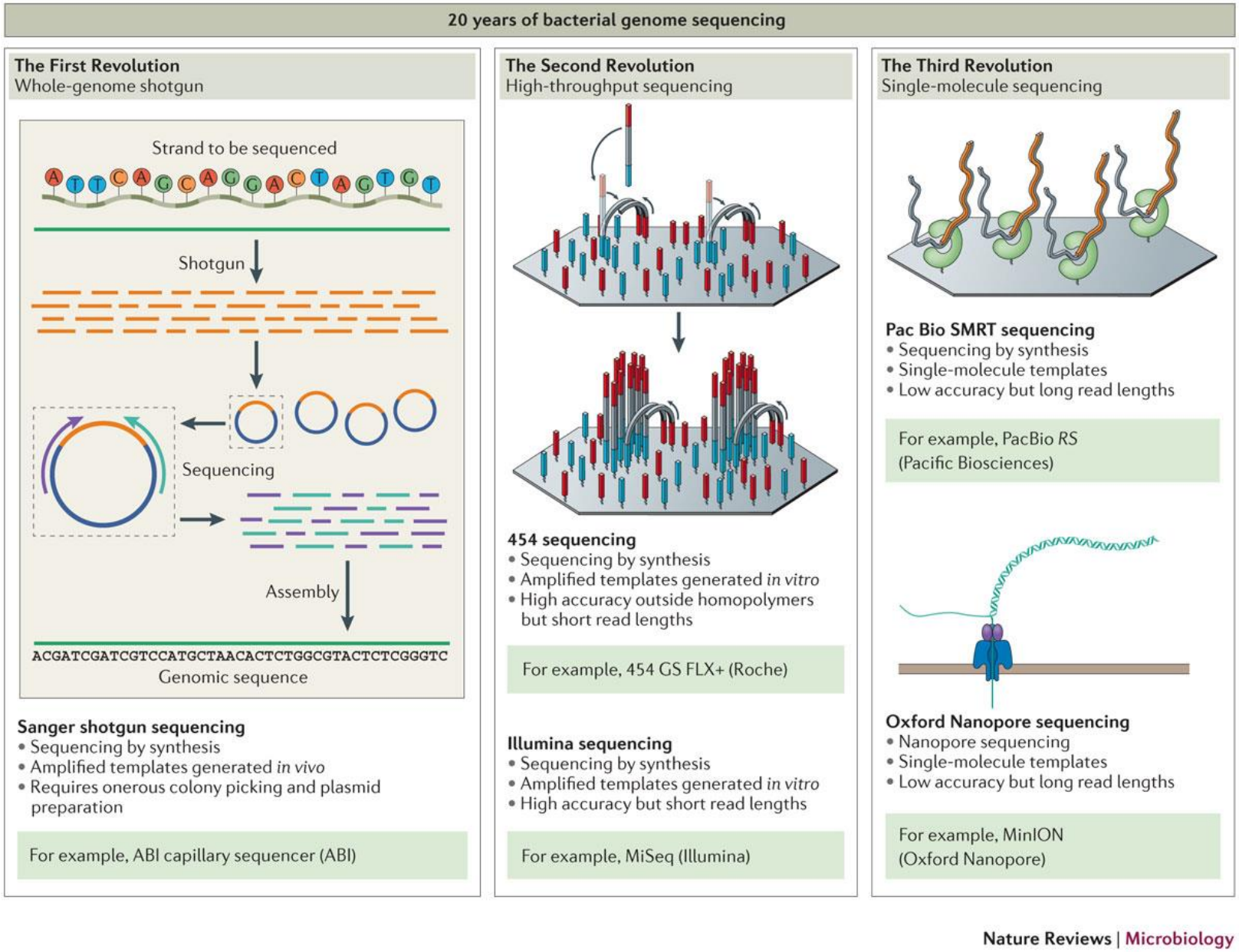
20 years of Whole Genome Sequencing (WGS) of bacteria

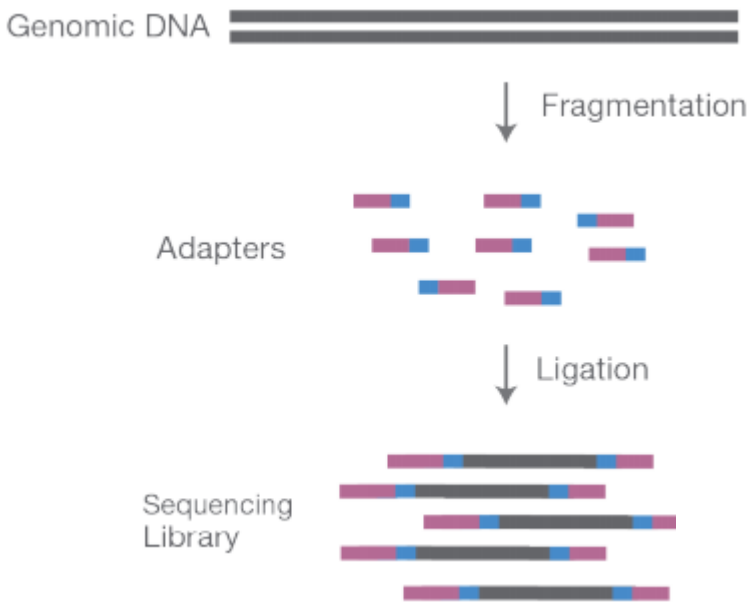
- Robust data: one method for all bacterial species
- Data storage for later analysis
- Monitoring of epidemic cases in hospital
- Monitoring emergency prevalent and emerging clones
- Identification of all genes of interest
- International comparison of prevalent and emerging clones

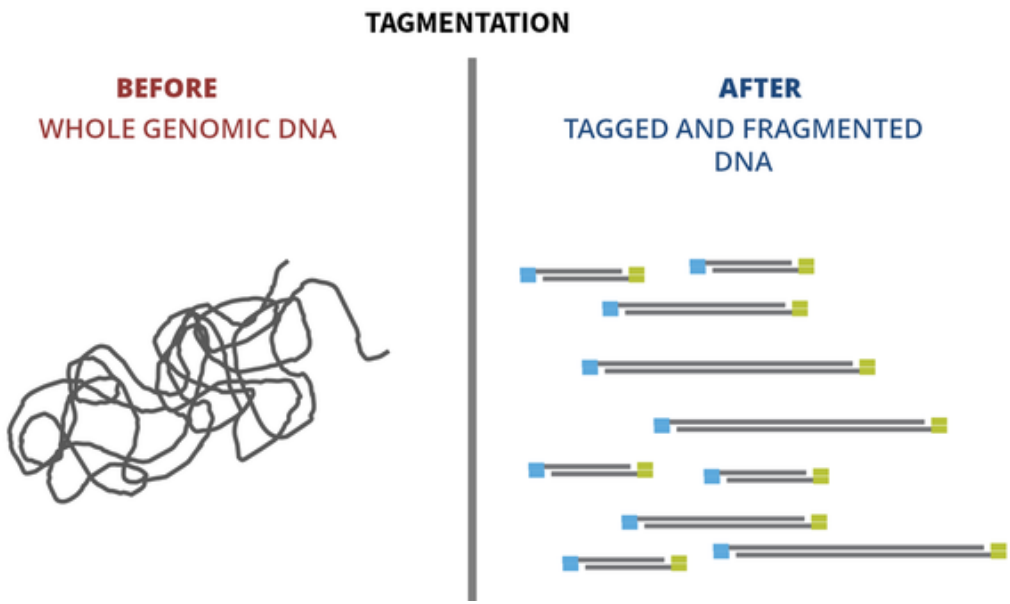


Loman, N., Pallen, M. Twenty years of bacterial genome sequencing. *Nat Rev Microbiol* **13,** 787–794 (2015). https://doi.org/10.1038/nrmicro3565

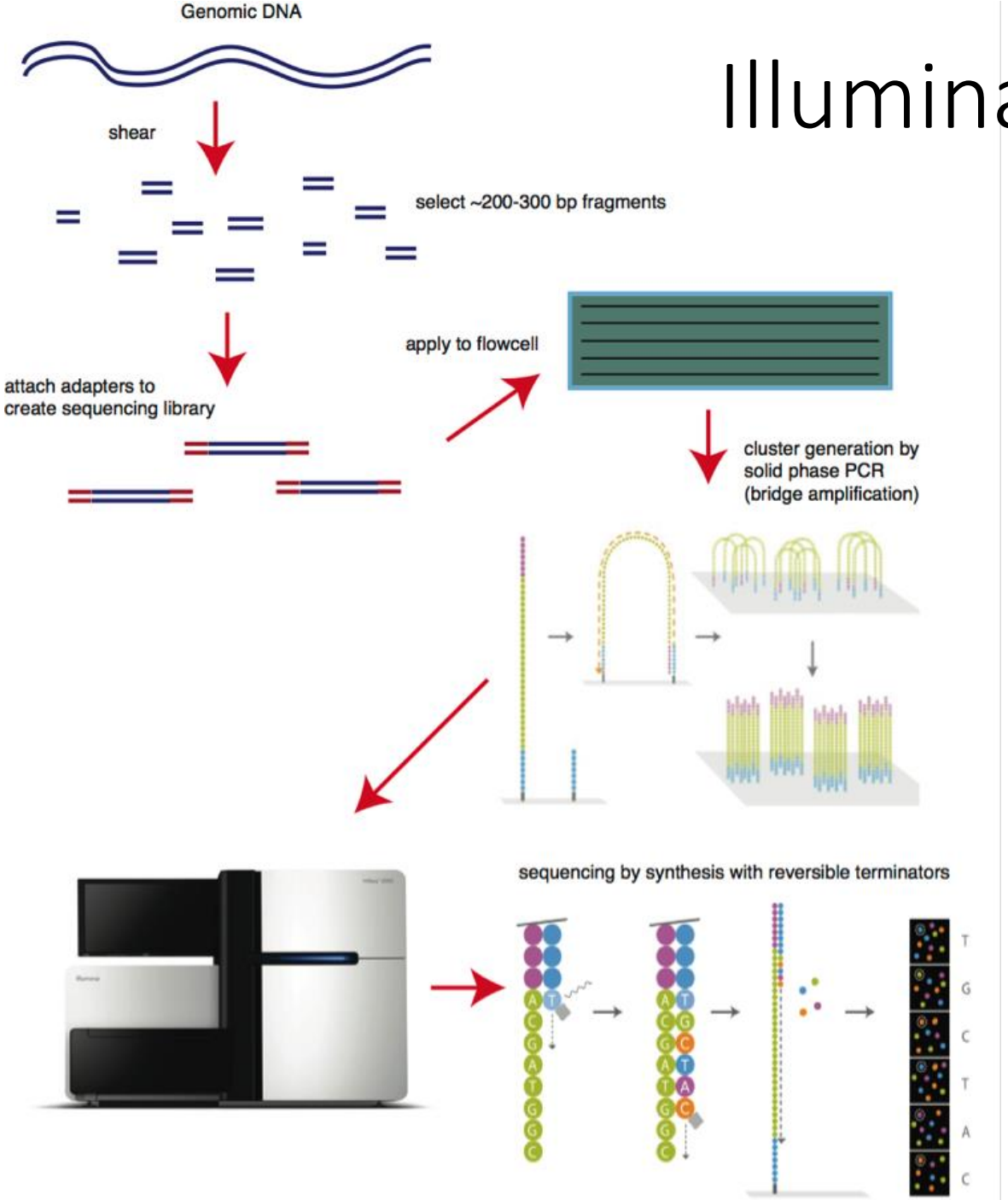
Whole-genome sequencing

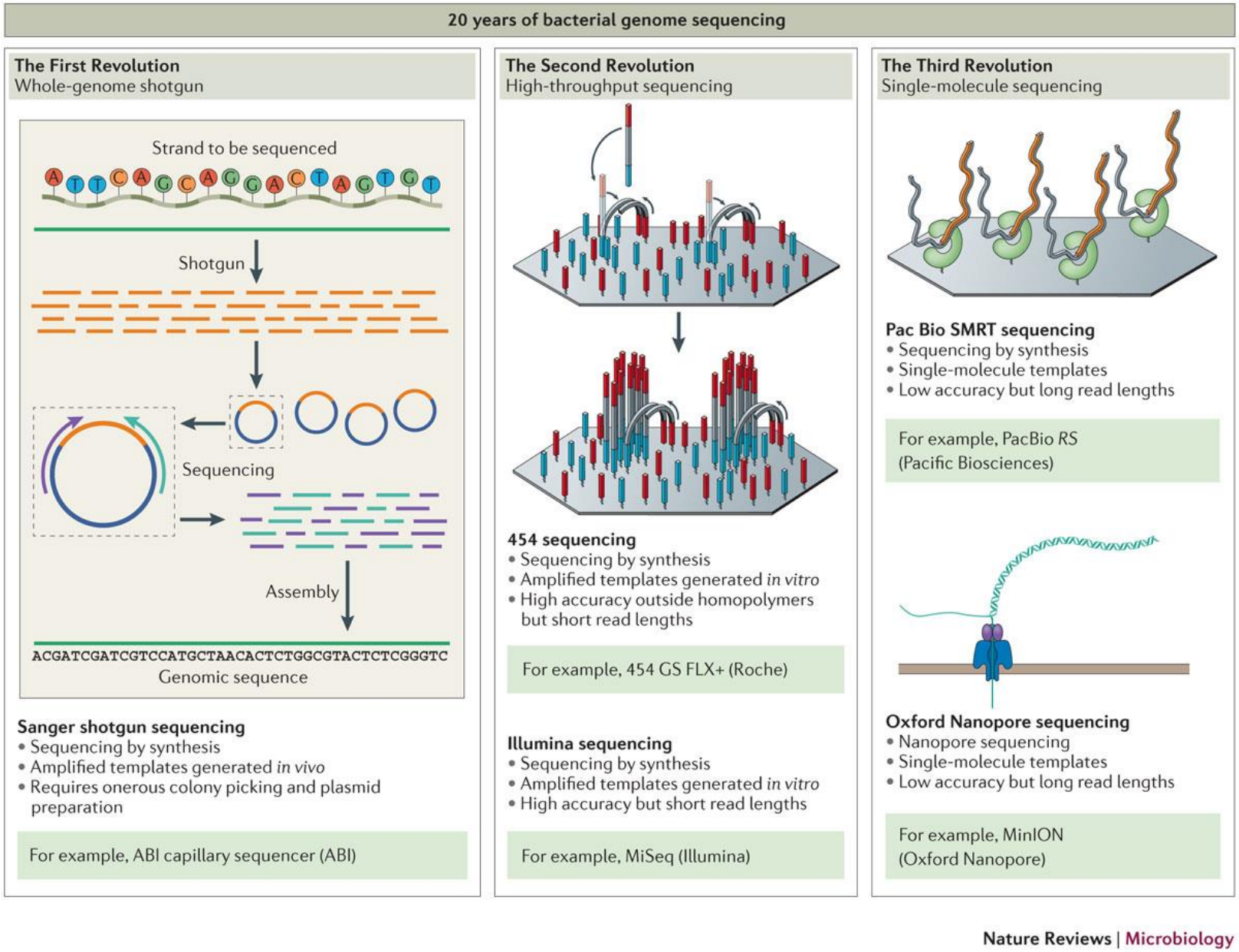
2nd generation





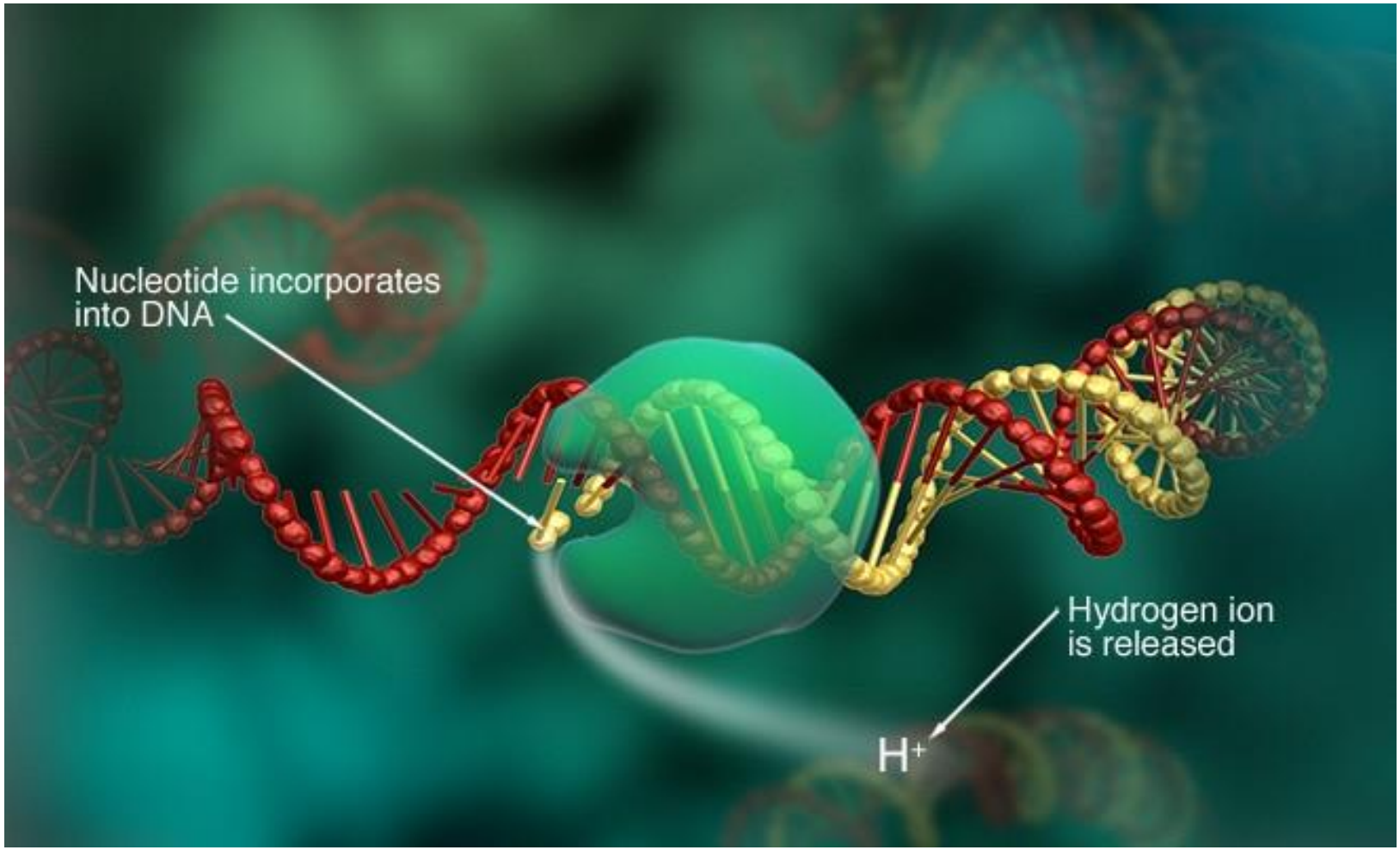
Illumina sequencing



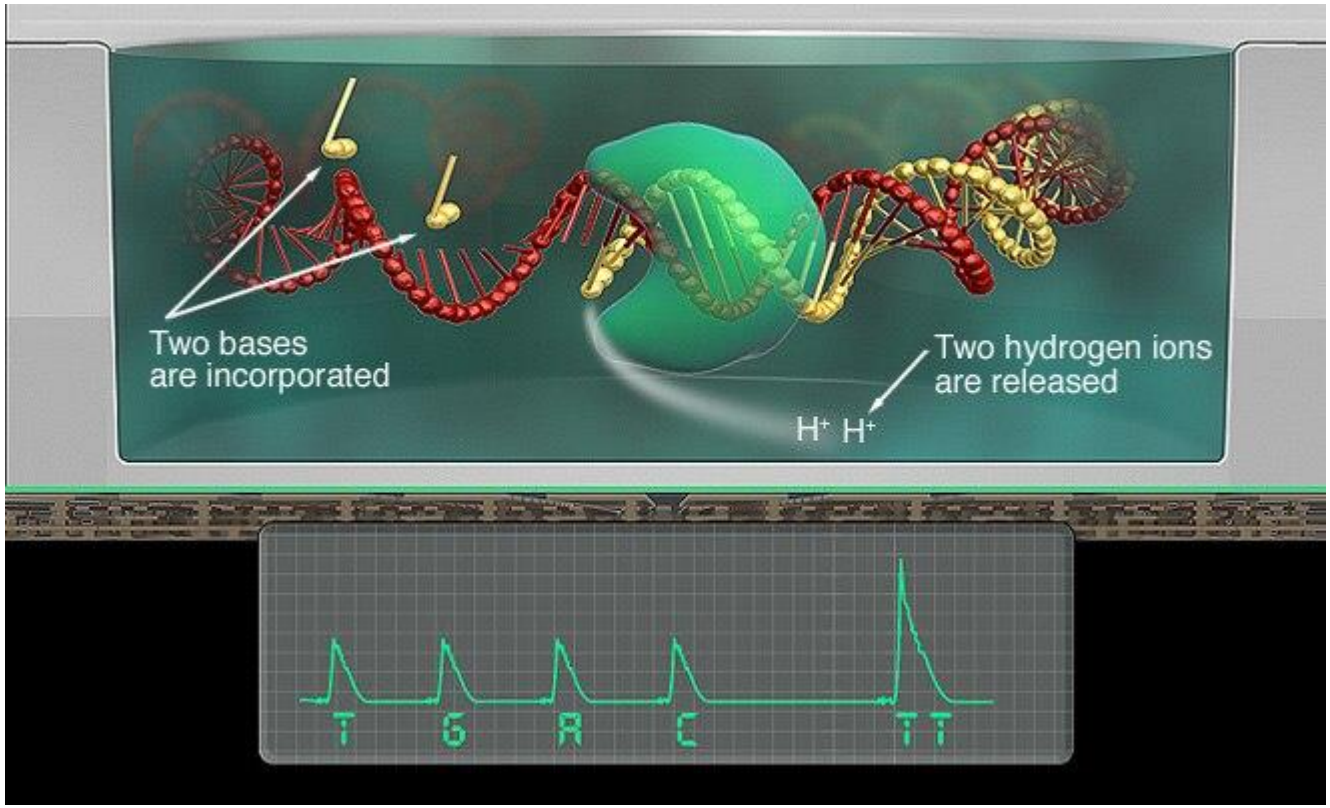


Loman, N., Pallen, M. Twenty years of bacterial genome sequencing. *Nat Rev Microbiol* **13,** 787–794 (2015). https://doi.org/10.1038/nrmicro3565

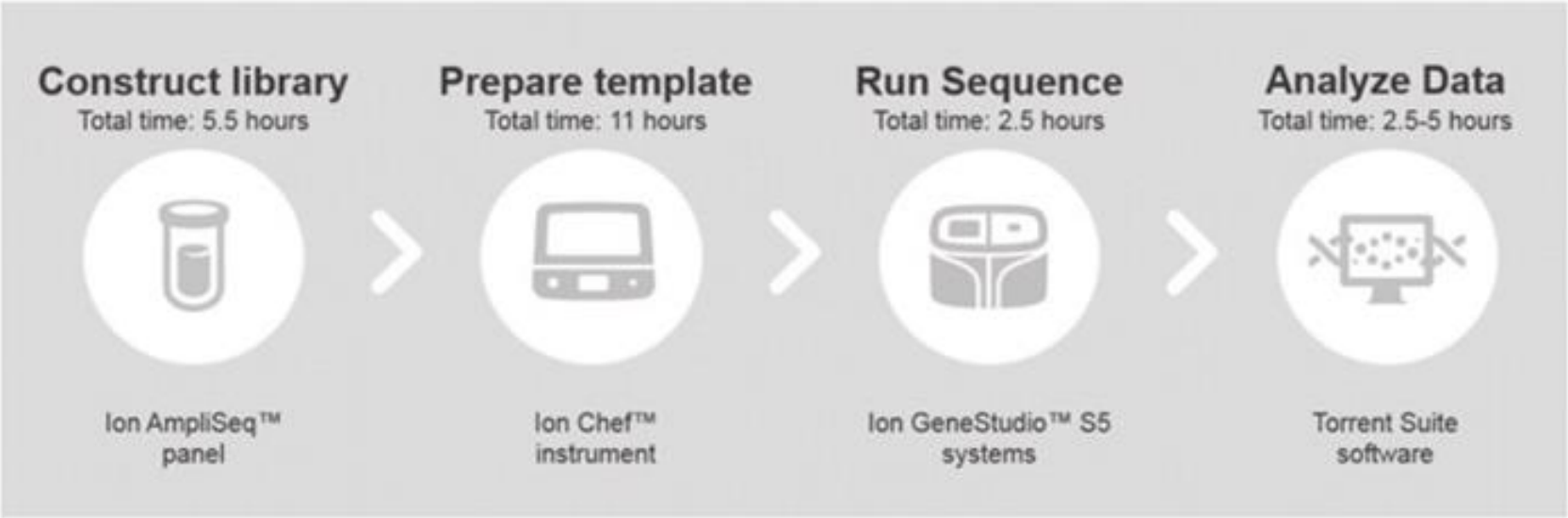
ION TORRENT Technology



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



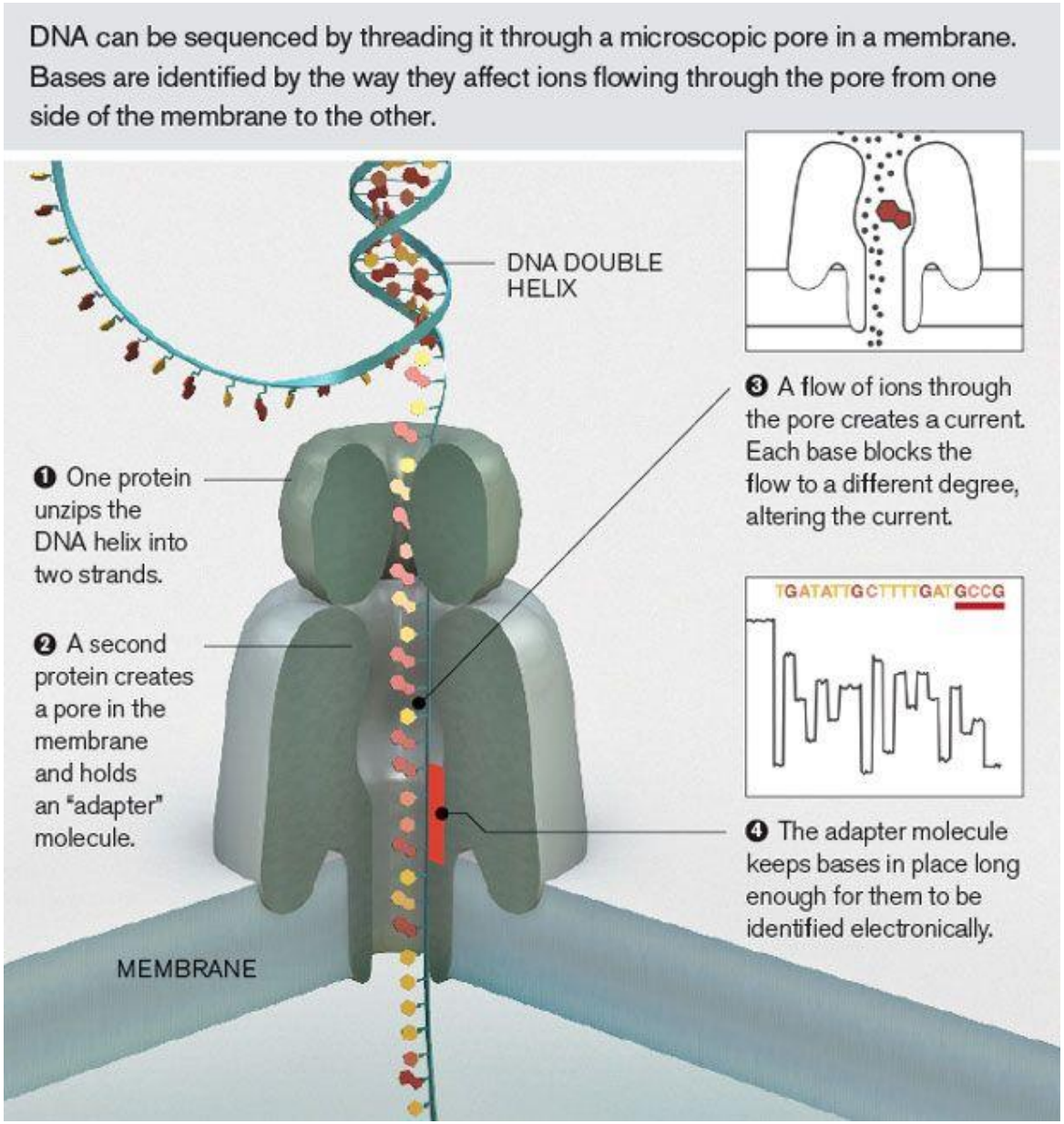
https://www.thermofisher.com/it/en/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-technology.html

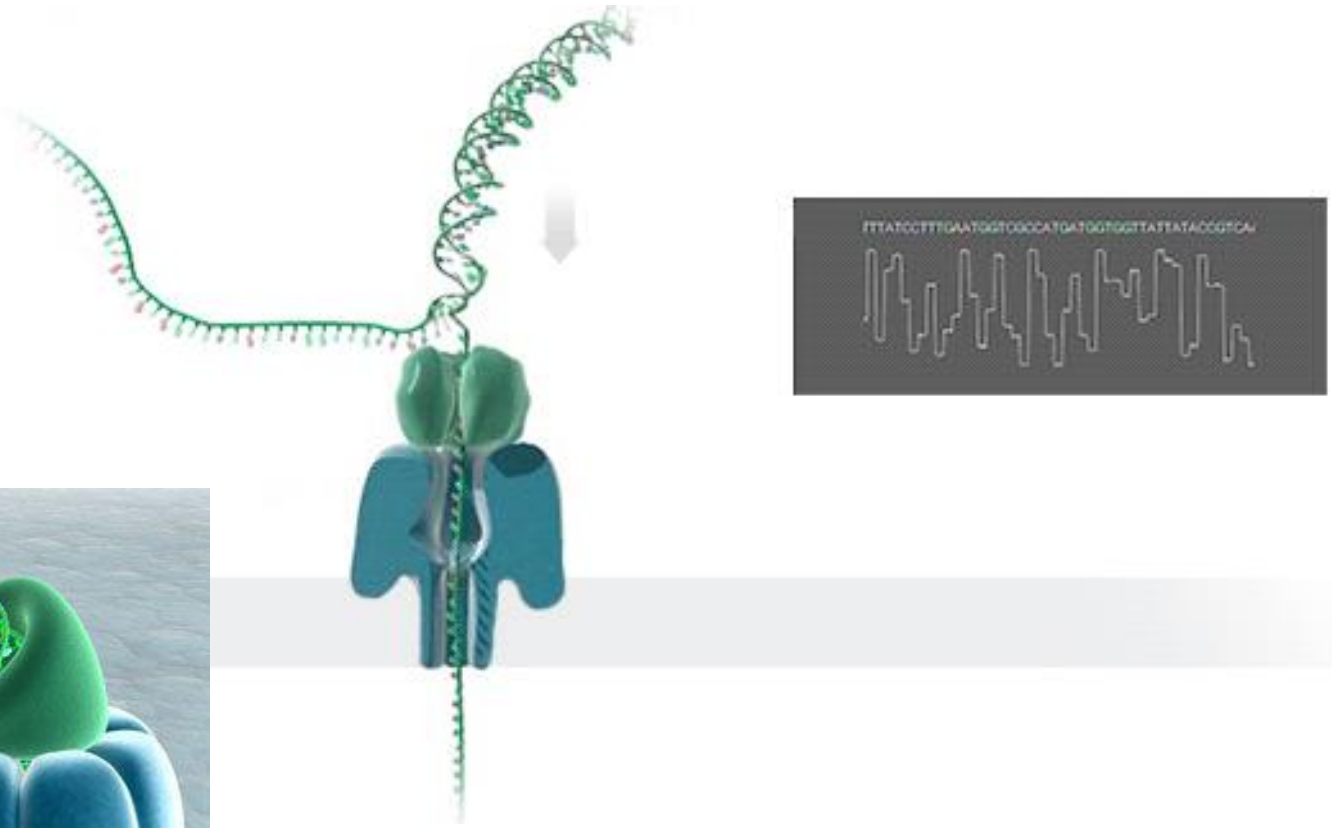

Construct library
Total time: 5.5 hours
Prepare template
Total time: 11 hours
Run Sequence
Total time: 2.5 hours
Analyze Data
Total time: 2.5-5 hours
Ion AmpliSeq™ panel
Ion Chef™ instrument
Ion GeneStudio™ S5 systems
Torrent Suite software

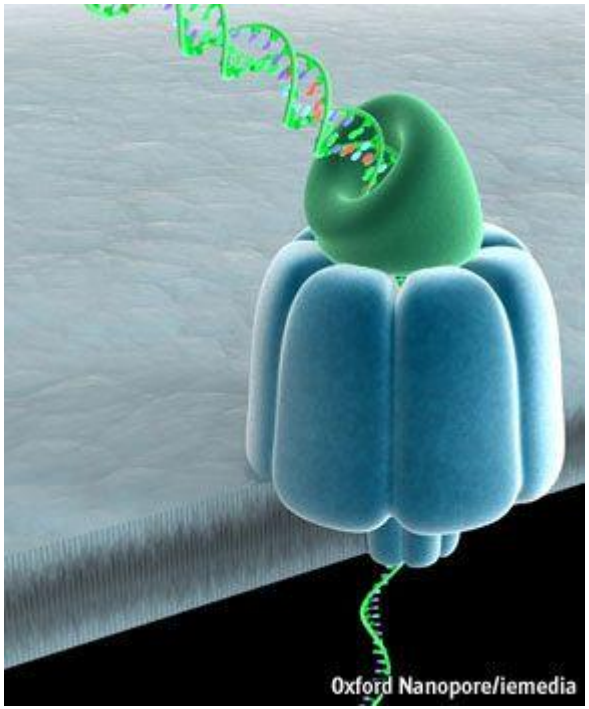

Step 1
Step 2
Step 3
Step 4
Step 5
Select targets
Construct library
Prepare template
Run sequence
Analyze data
Select from pre-designed panels or configure your own with the Ion AmpliSeq Designer.
Featured research solutions:
• Coronavirus (SARS-CoV-2)
• Ebola (EBOV)
• Zika virus (ZIKV)
• Human immunodeficiency virus (HIV)
• Human papillomaviruses (HPV)

Nanopore sequencing

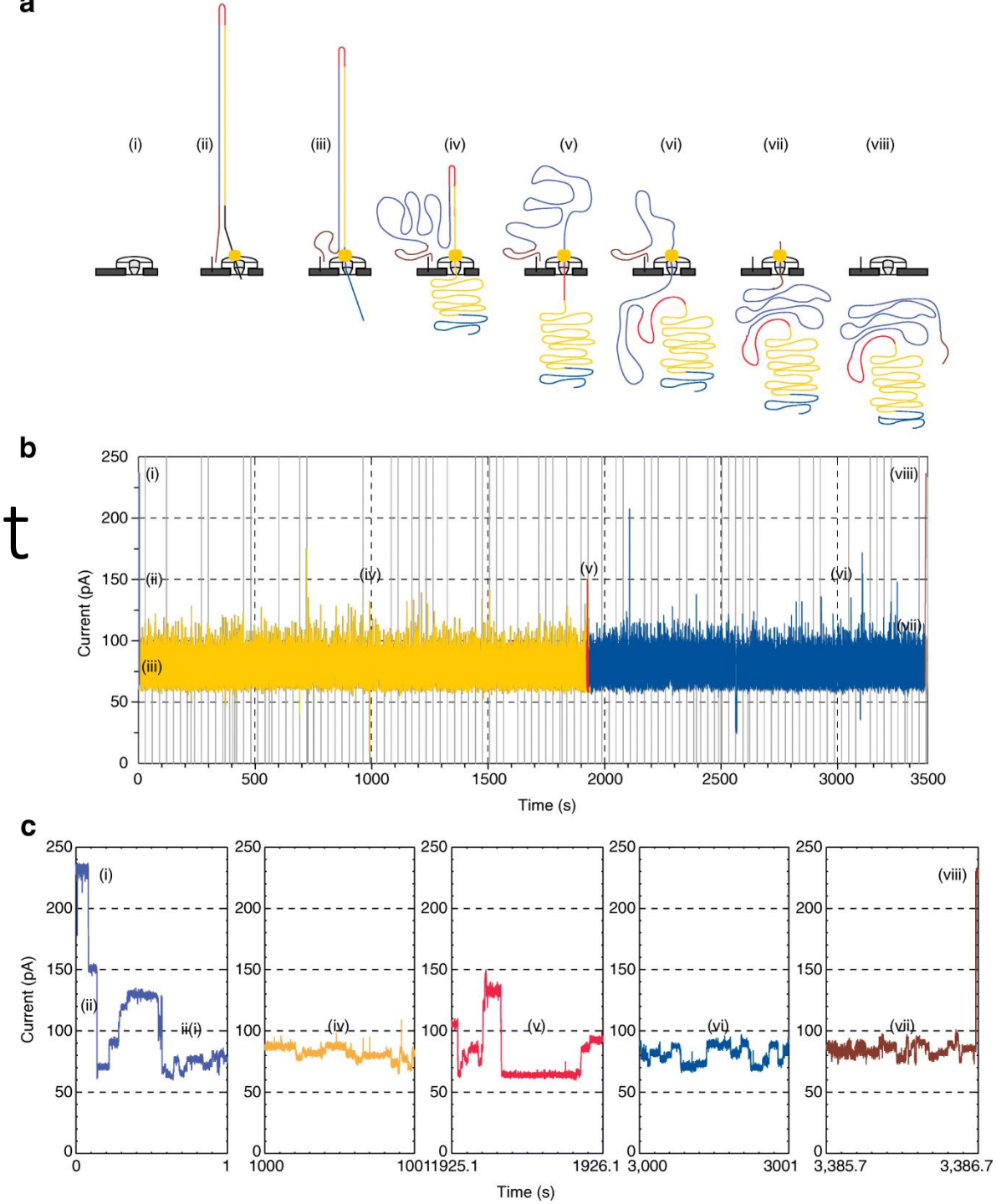
Nanopore sequencing

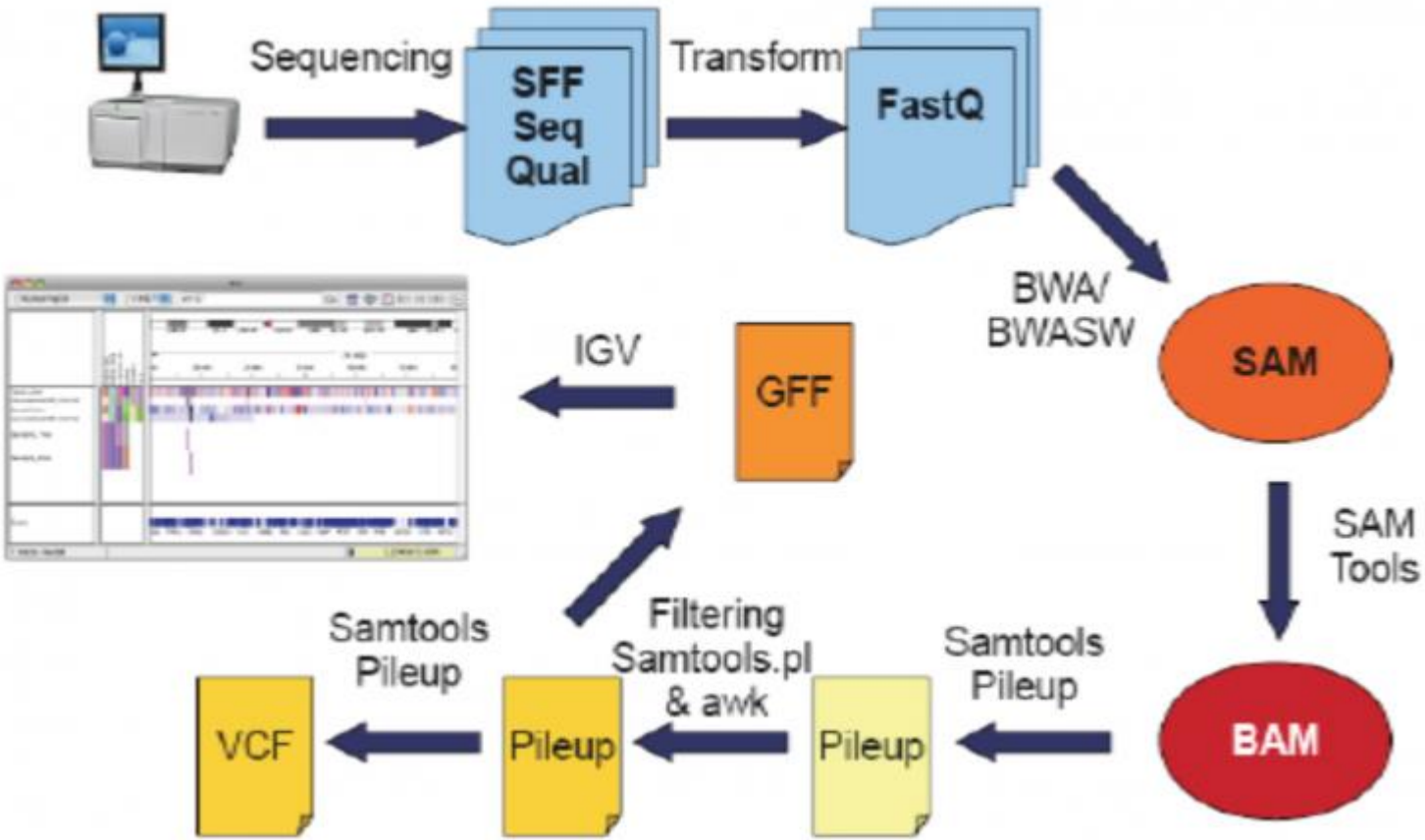



TTTATCCTTTGAATGGTCGCCATGATGGTGGTTATTATACCGTCAA


Oxford Nanopore/iemedia

Nanopore sequencing: long reads but mistakes!!!




Sequencing
SFF
Seq
Qual
Transform
FastQ
BWA/
BWASW
SAM
SAM
Tools
BAM
Samtools
Pileup
Pileup
Filtering
Samtools.pl
& awk
Pileup
Samtools
Pileup
VCF
GFF
IGV

The average coverage for a whole genome can be calculated from the length of the original genome (G), the number of reads (N), and the average read length (L) as

$$N \times L/G.$$

For example, a hypothetical genome with 2,000 base pairs reconstructed from 8 reads with an average length of 500 nucleotides will have
8 x 500 : 2000 = 2× redundancy.
This parameter also enables one to estimate other quantities, such as the percentage of the genome covered by reads (sometimes also called breadth of coverage).
A high coverage in shotgun sequencing is desired because it can overcome errors in base calling and assembly. The subject of DNA sequencing theory addresses the relationships of such quantities.

FastQ: Each sequence requires at least 4 lines:

1. The first line is the sequence header which starts with an '@'
2. The second line is the sequence.
3. The third line starts with '+'
4. The fourth line are the quality scores

```
@HWI-ST911:111:C0N4WACXX:5:1101:2249:2216 1:N:0:TTAGGC CGATC:@@@FF
NATGGCACCATTAAAAAGAATGTTTTATATGGTGTTGAGAAGGACAAAGCTGAAGAAGAAATTTAGTCTGCACTTGATGTTGCAAATGCAAAGAAA
+
#2A2<CCFHIIIIIIIIIIIIIGCCHIIIGIIIFFHIIDGHIGIIIIIIICHGIIIGGCECEGICFHCECDEFFFFDEEEEEDDDDDCDDCDDDDBC
@HWI-ST911:111:C0N4WACXX:5:1101:2509:2197 1:N:0:TTAGGC CGATC:1+4=B
NATGAGATAAATCAATTGTCTTTAATGAAGTACAGTCTTTGAATAATGAGTTTTGAACTCTTCTGCAACTTTTTGGAAACTTTAAAGTTTGTAATG
+
#4A2<AADHIIIIIIIIIHHIIIIIIIIIIFGIII@GIIFIIIGIIIIEIDHEHIIIHIIIIIIIIIIICHIIIHHEEDFFFFFEEECEEEADDFC
@HWI-ST911:111:C0N4WACXX:5:1101:3746:2179 1:N:0:TTAGGC CCATC:+11+A
NATGTCATCCATCTTTTCCTATCTAAAAAAGAATCAAAAAAGGGATAGTACAGAGGGAAAGTTCAATCCAGAGGACGATGAAACACTGATTGATGG
+
```

HW-ST911	the unique instrument name
111	the run id
C0N4WACXX	the flowcell id
5	flowcell lane
1101	tile number within the flowcell lane
2249	'x'-coordinate of the cluster within the tile
2216	'y'-coordinate of the cluster within the tile
1	the member of a pair, 1 or 2 (paired-end or mate-pair reads only)
Y	Y if the read is filtered, N otherwise
18	0 when none of the control bits are on
TTAGGC, CGATC	index sequence

FastA: Each sequence consists of at least two lines:

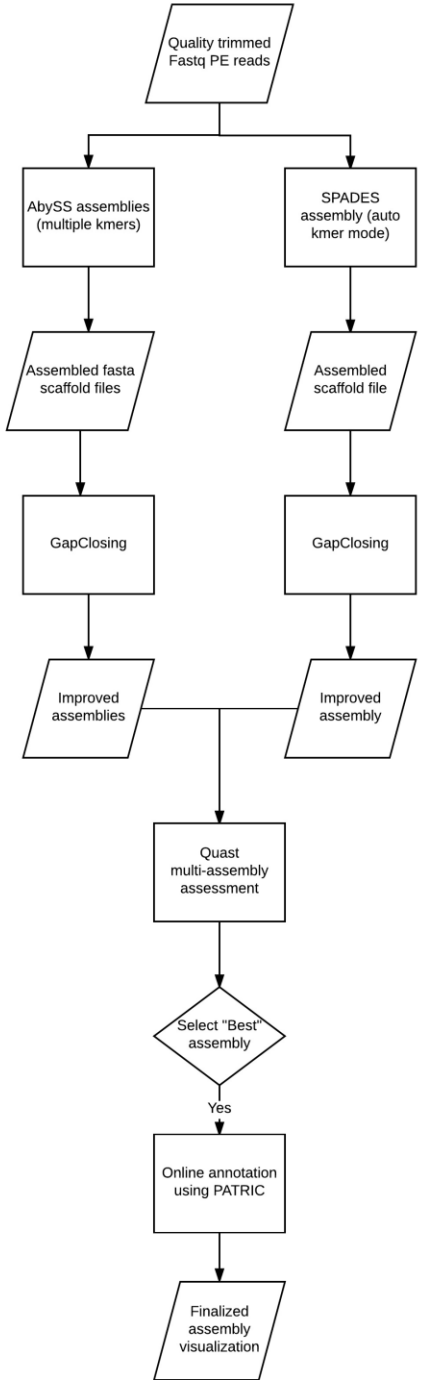
1. The first is the sequence header, which always starts with a ‘>’
    1. Everything from the beginning ‘>’ to the first whitespace is considered the sequence identifier. Everything after that is considered the sequence description (this can be metadata, machine serial number, read orientation, etc.)
2. The sequence itself

    Note that the sequence can span multiple lines, depending on the length of the sequence.

```
>Chr1 CHROMOSOME dumped from ADB: Jun/20/09 14:53; last updated: 2009-02-02
CCCTAAACCCTAAACCCTAAACCCTAAACCTCTGAATCCTTAATCCCTAAATCCCTAAATCTTTAAATCCTACATCCAT
GAATCCCTAAATACCTAATTCCCTAAACCCGAAACCGGTTTCTCTGGTTGAAAATCATTGTGTATATAATGATAATTTT
ATCGTTTTTATGTAATTGCTTATTGTTGTGTGTAGATTTTTTAAAAATATCATTTGAGGTCAATACAAATCCTATTTCT
TGTGGTTTTCTTTCCTTCACTTAGCTATGGATGGTTTATCTTCATTTGTTATATTGGATACAAGCTTTGCTACGATCTA
CATTTGGGAATGTGAGTCTCTTATTGTAACCTTAGGGTTGGTTTATCTCAAGAATCTTATTAATTGTTTGGACTGTTTA
TGTTTGGACATTTATTGTCATTCTTACTCCTTTGTGGAAATGTTTGTTCTATCAATTTATCTTTTGTGGGAAAATTATT
TAGTTGTAGGGATGAAGTCTTTCTTCGTTGTTGTTACGCTTGTCATCTCATCTCTCAATGATATGGGATGGTCCTTTAG
```

Studying a genomic sequence

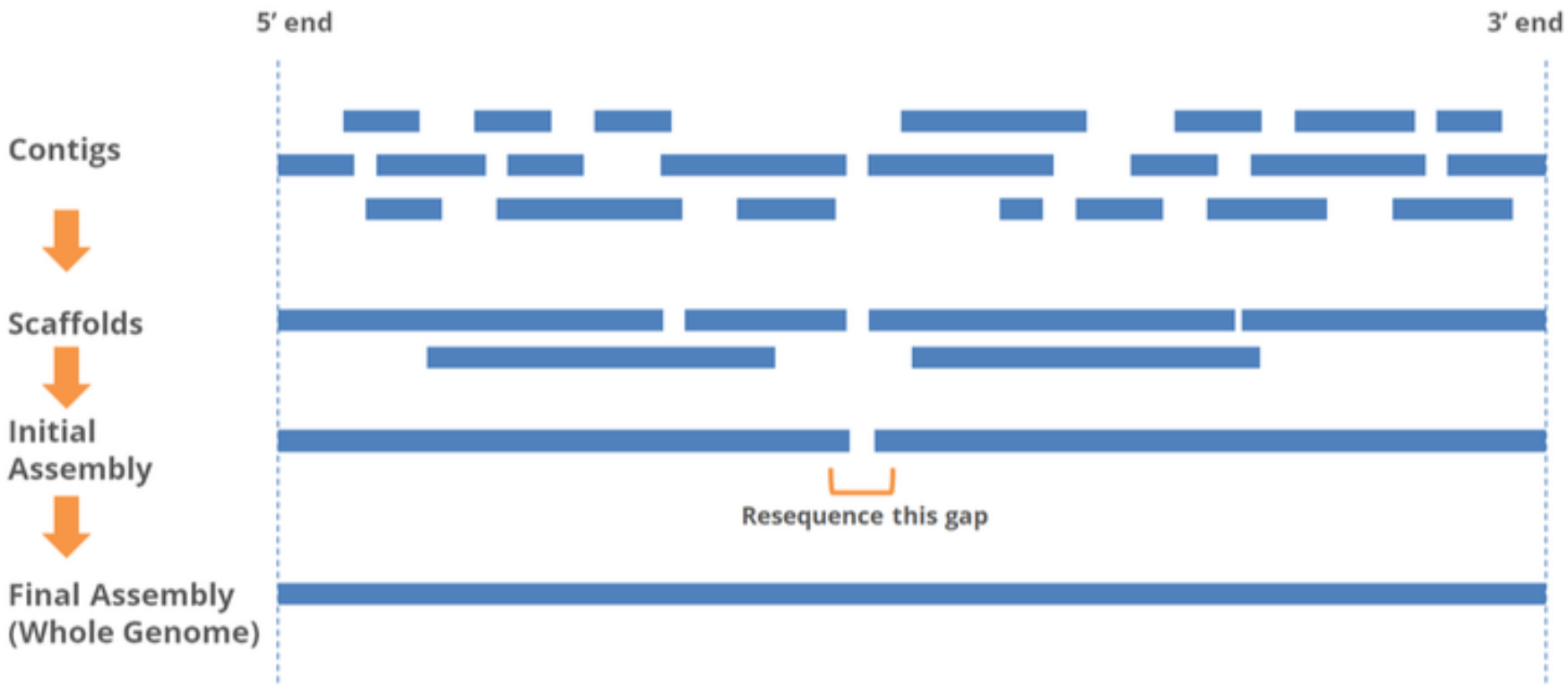
- Bacterial genome lenght approx. 3-5 x 10^6 base pairs
- Coverage (or depth) in DNA sequencing is the number of unique reads that include a given nucleotide in the reconstructed sequence. Deep sequencing refers to the general concept of aiming for high number of unique reads of each region of a sequence.
- Even though the sequencing accuracy for each individual nucleotide is very high, the very large number of nucleotides in the genome means that if an individual genome is only sequenced once, there will be a significant number of sequencing errors. Furthermore, many positions in a genome contain rare single-nucleotide polymorphisms (SNPs). Hence to distinguish between sequencing errors and true SNPs, it is necessary to increase the sequencing accuracy even further by sequencing individual genomes a large number of times
- The term "ultra-deep" can refer to higher coverage (>100-fold), which allows for detection of sequence variants in mixed populations



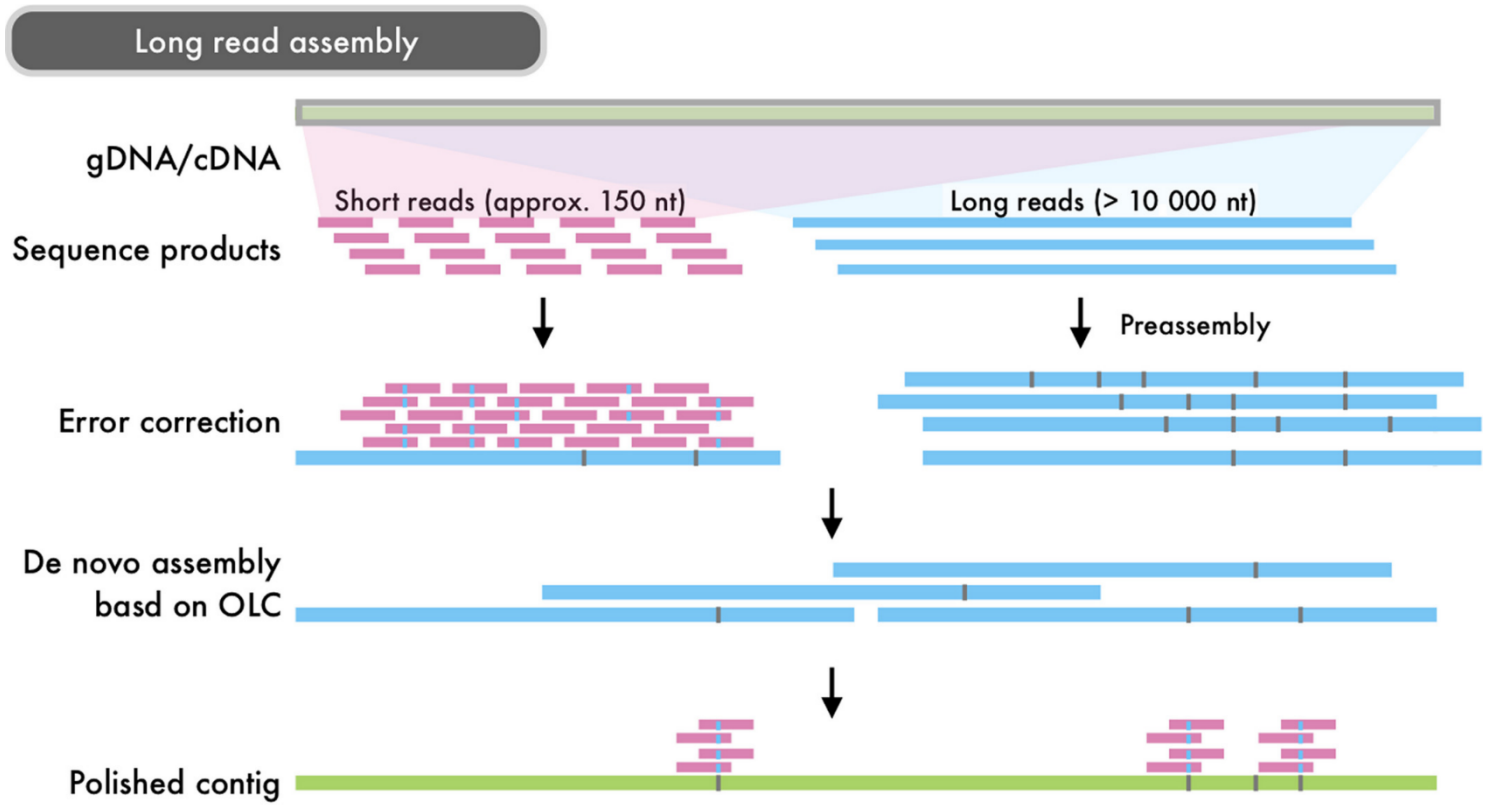
De novo genome assembly

1-Informatic tools normally use FastQ files
to produce
FastA files
2-GapClosing
to produce
Longer contigs or scaffolds
(long reads)
PCR-based closure
3-on line annotation

Assembler


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

short reads and long reads

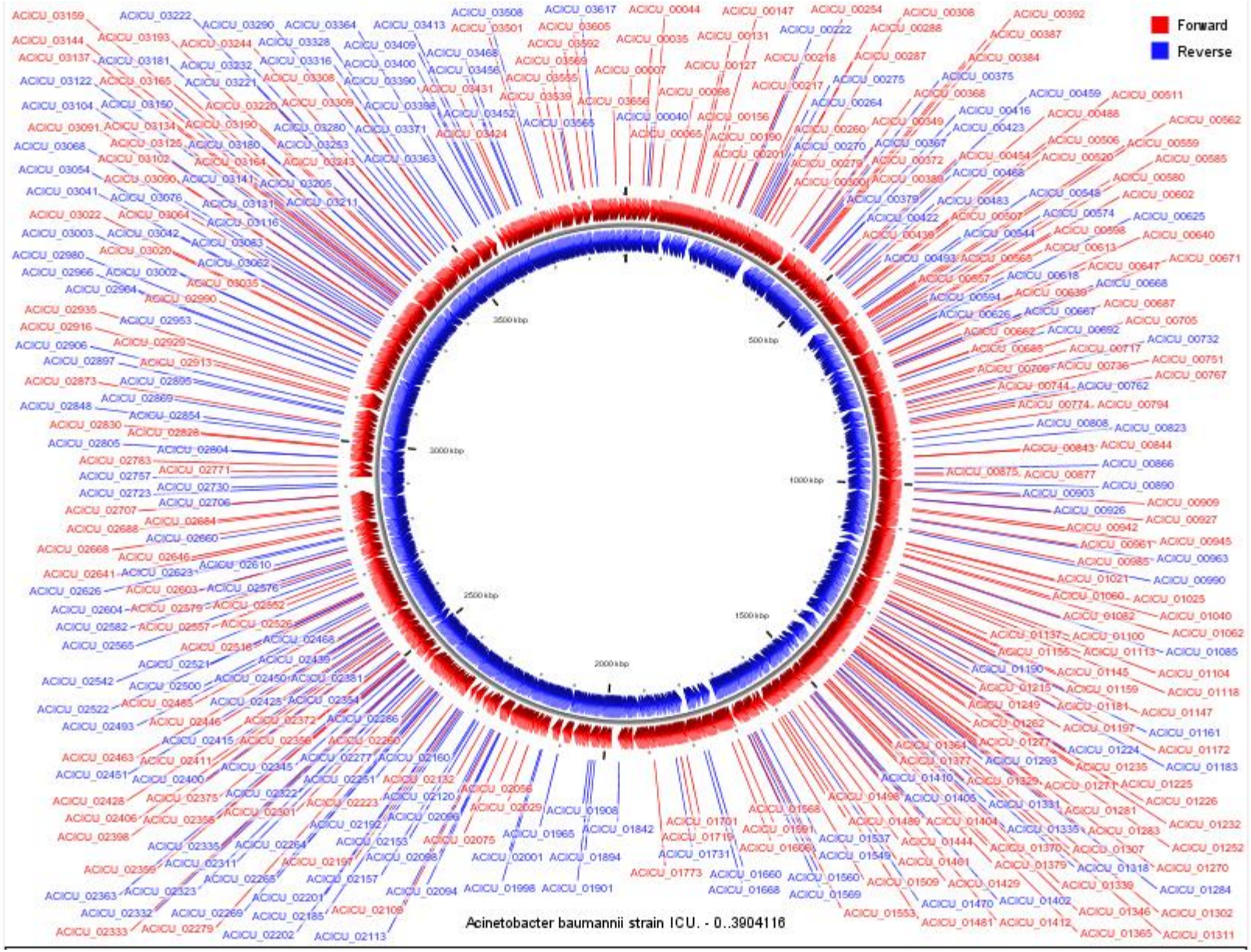


Studying genome content

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

DATABASES

Genome annotation



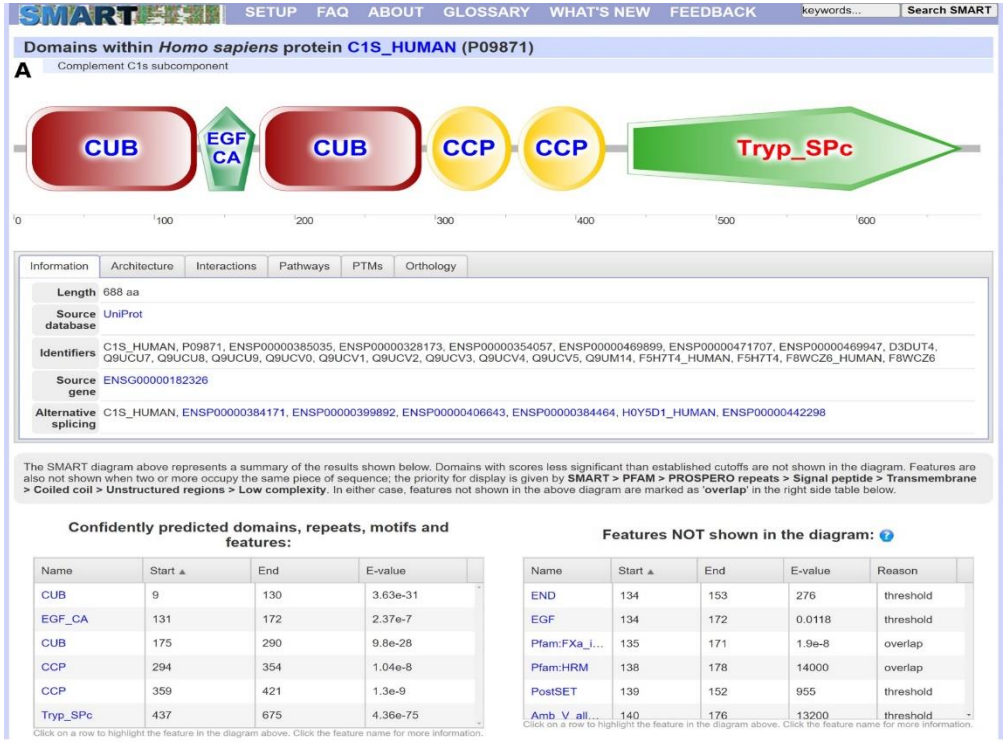
Functional classes of the proteins

- 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**

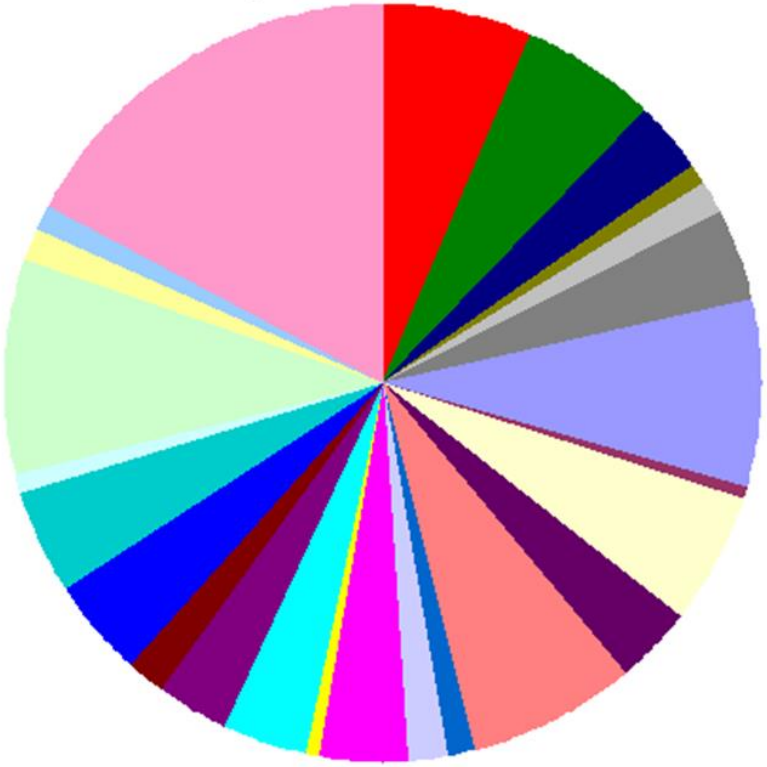
https://doi.org/10.1093/nar/gkx922

Subsystem Information

Subsystem Statistics | Features in Subsystems

Subsystem Coverage

Subsystem Category Distribution



Subsystem Feature Counts

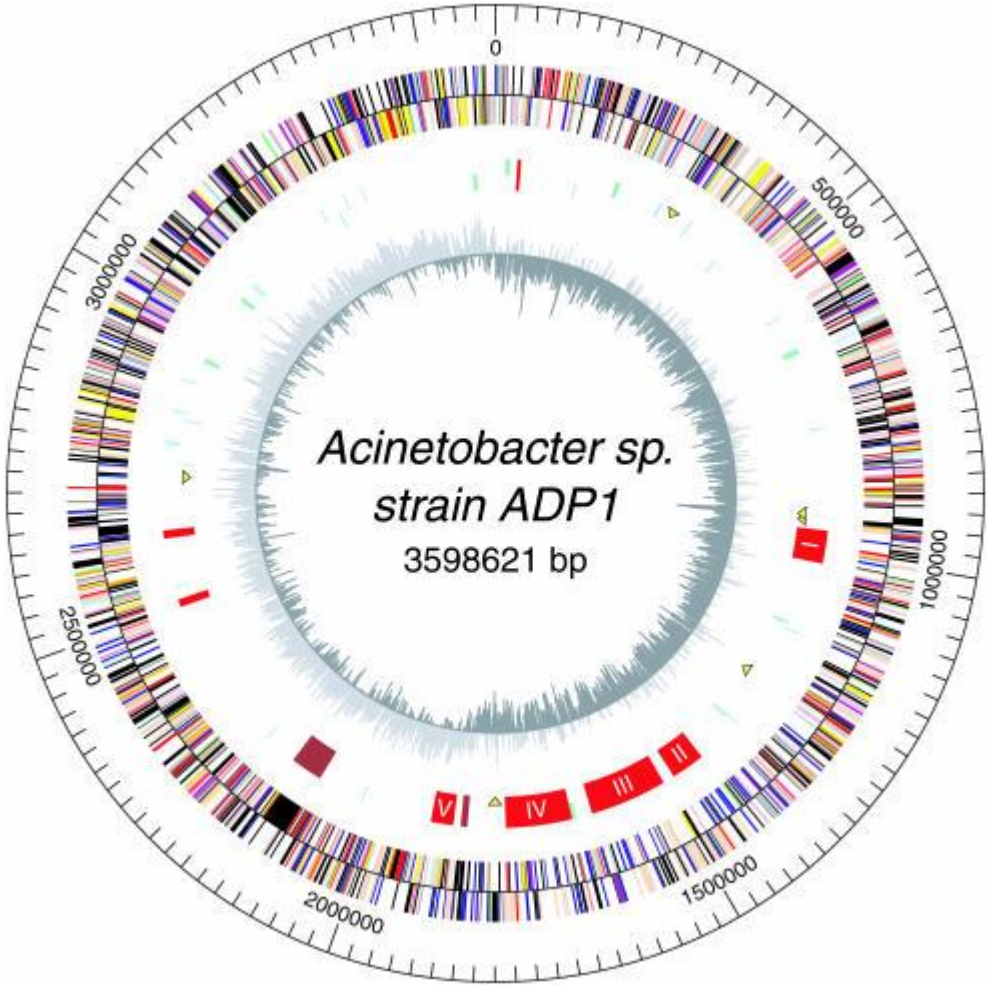
- Cofactors, Vitamins, Prosthetic Groups, Pigments (290)
- Cell Wall and Capsule (262)
- Virulence, Disease and Defense (138)
- Potassium metabolism (28)
- Photosynthesis (0)
- Miscellaneous (65)
- Phages, Prophages, Transposable elements, Plasmids (175)
- Membrane Transport (356)
- Iron acquisition 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



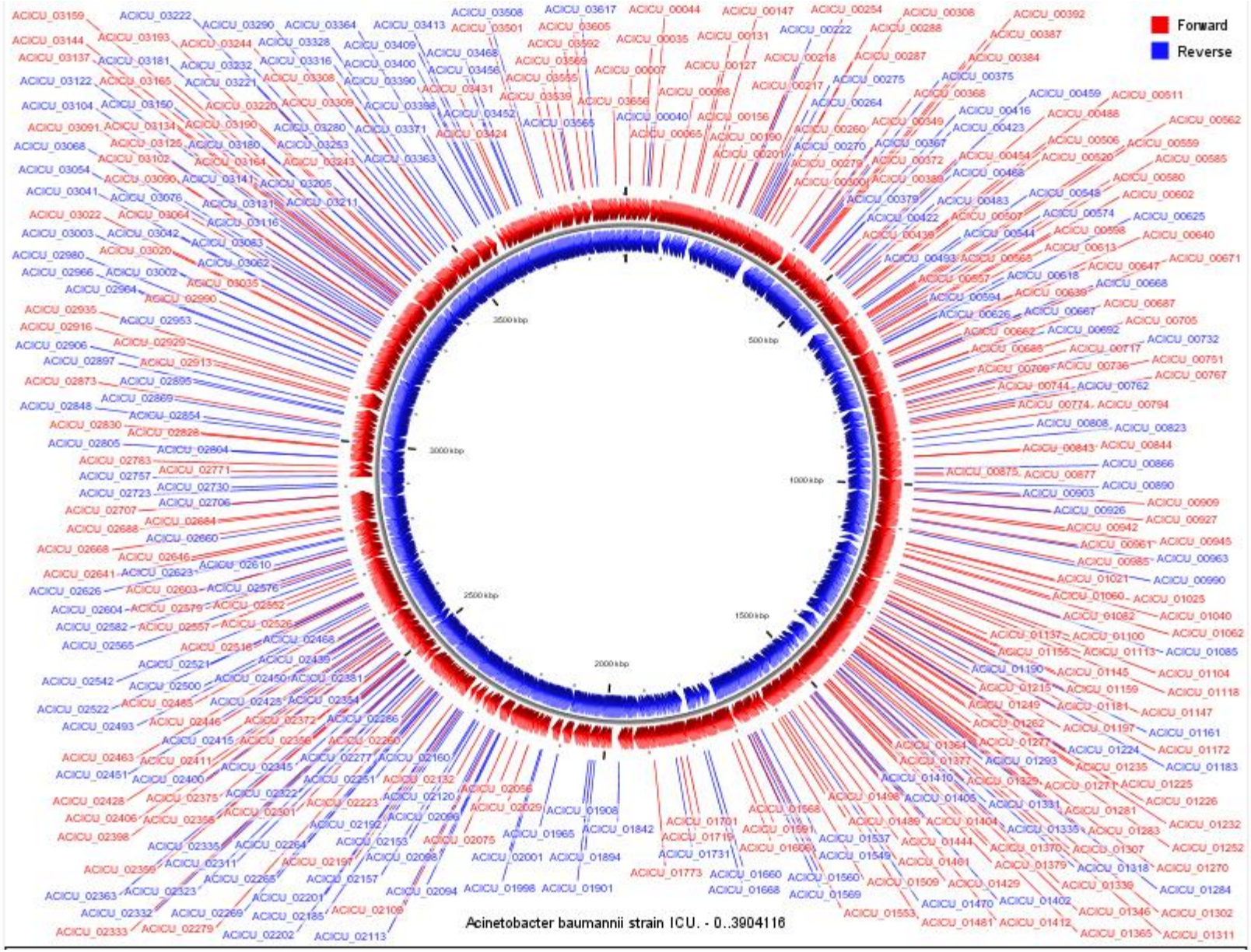
Acinetobacter baumannii

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**

- *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.

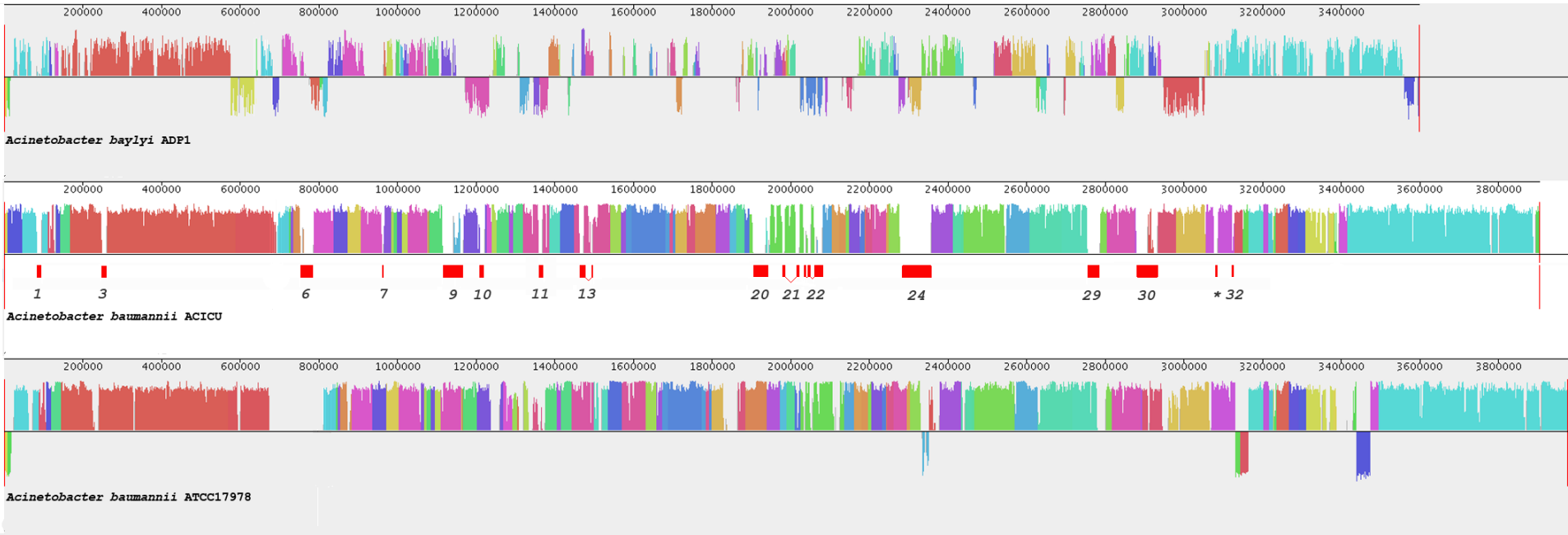
Genome annotation



- 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

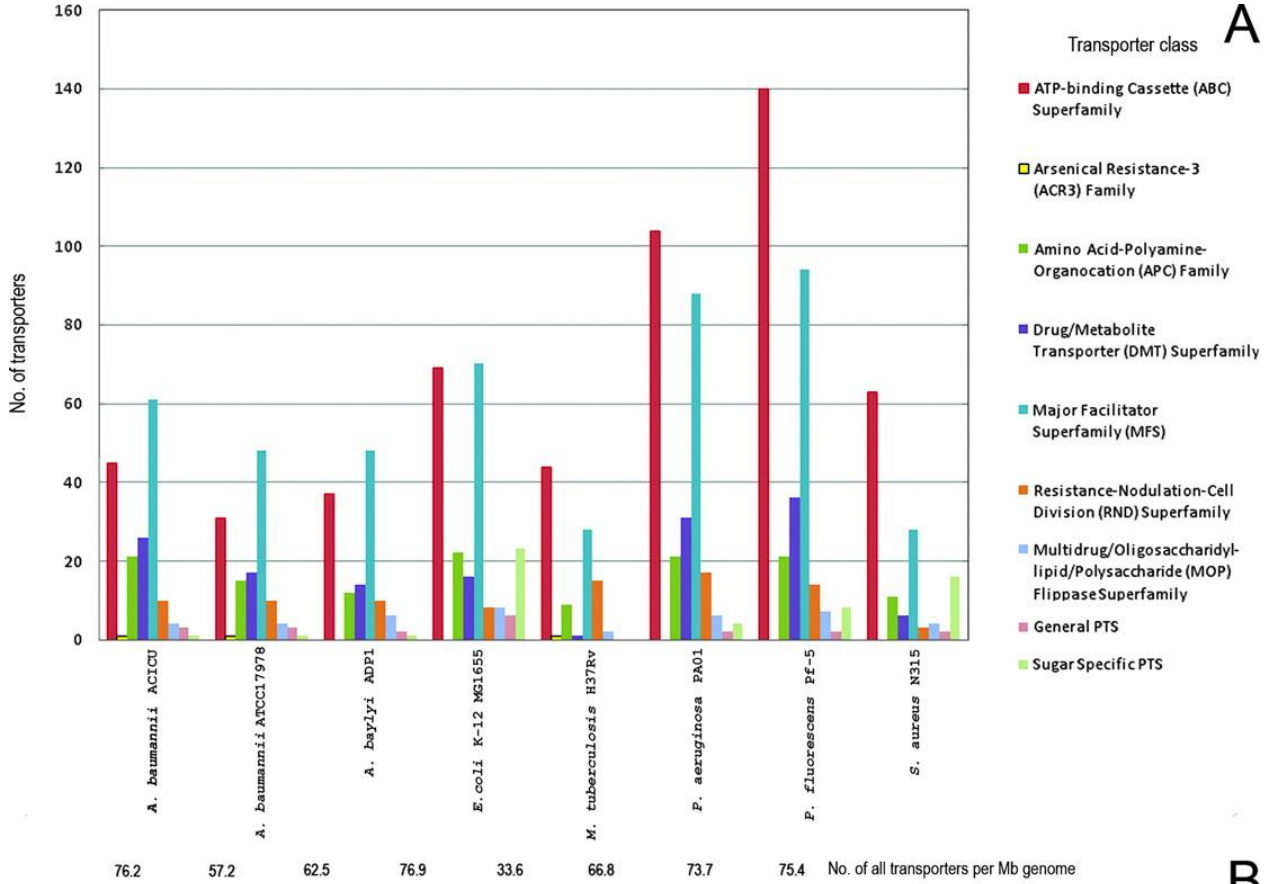
36 putative alien islands (pAs) were detected in the ACICU genome; 24 of these had previously been described in the ATCC 17978 genome, 4 are proposed here for the first time and are present in both ATCC 17978 and ACICU, and 8 are unique to the ACICU genome.

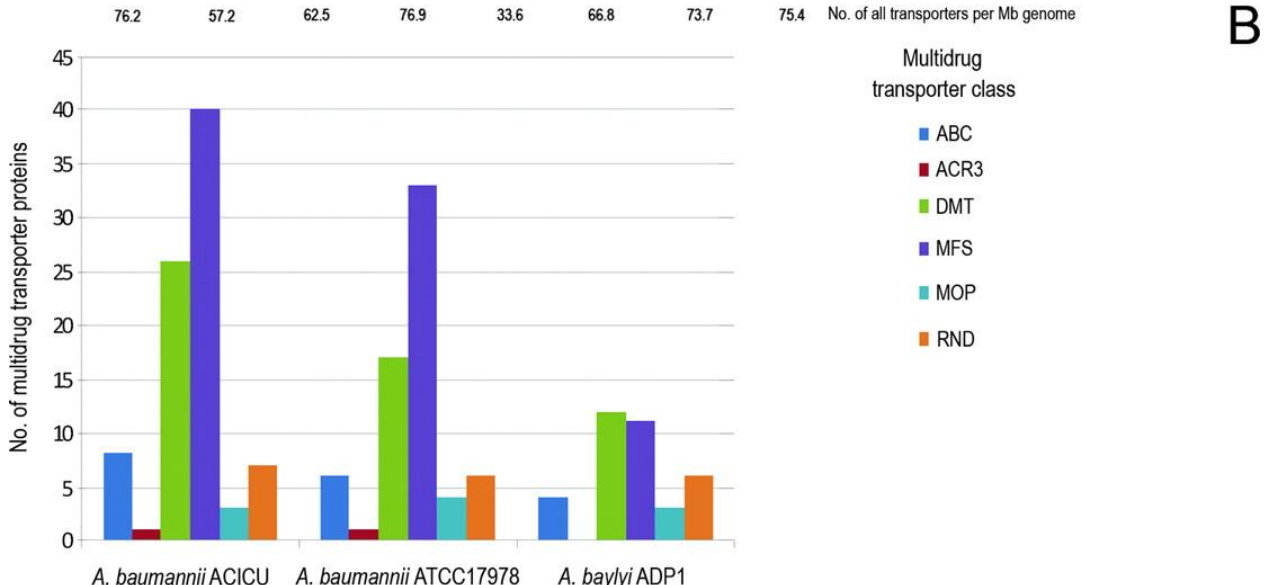
Acinetobacter spp. synthenia



- 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

A conspicuous number of transporters belonging to different superfamilies was predicted for *A. baumannii* ACICU. The relative number of transporters was much higher in ACICU than in ATCC 17978 and ADP1 (76.2, 57.2, and 62.5 transporters per Mb of genome, respectively).





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 **Published** 1 July 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.

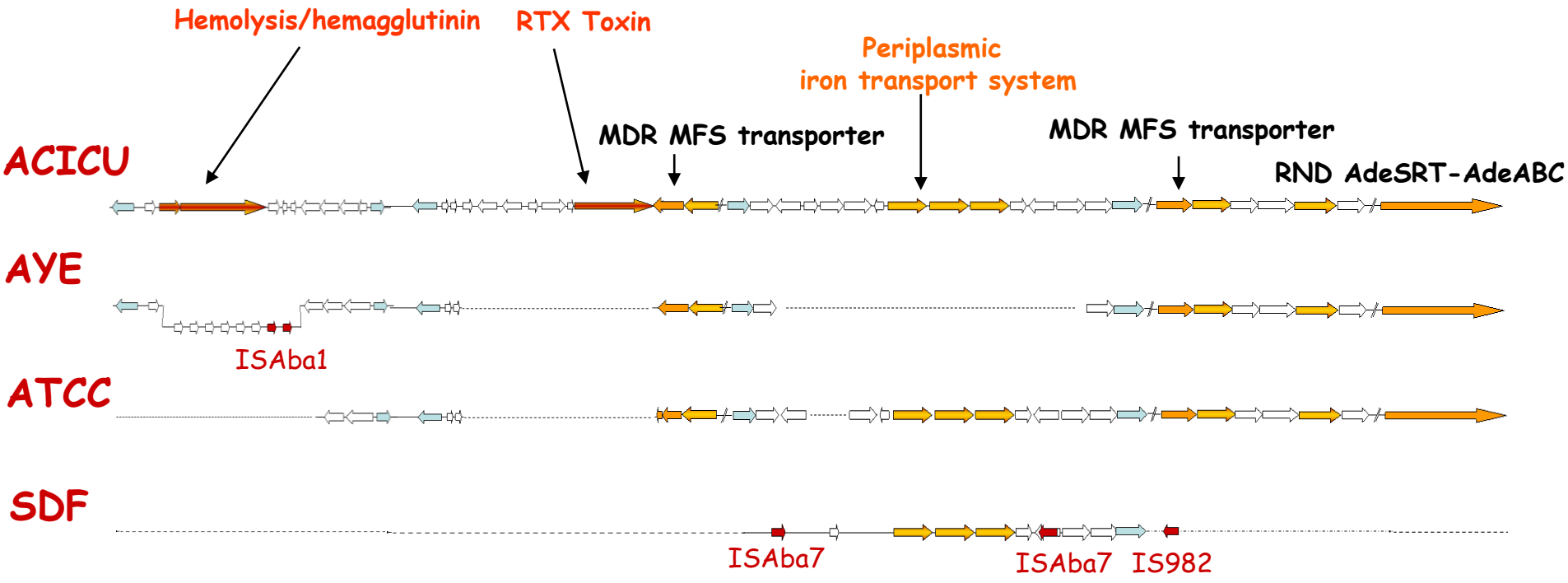
Generally considered a low-virulence species, candidate virulence factors have proved hard to identify. Many of the putative alien islands carry potential virulence genes, including type IV secretion systems, siderophores and haemolysins/haemagglutinins. Screens of transposon mutants of *A. baumannii* 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)

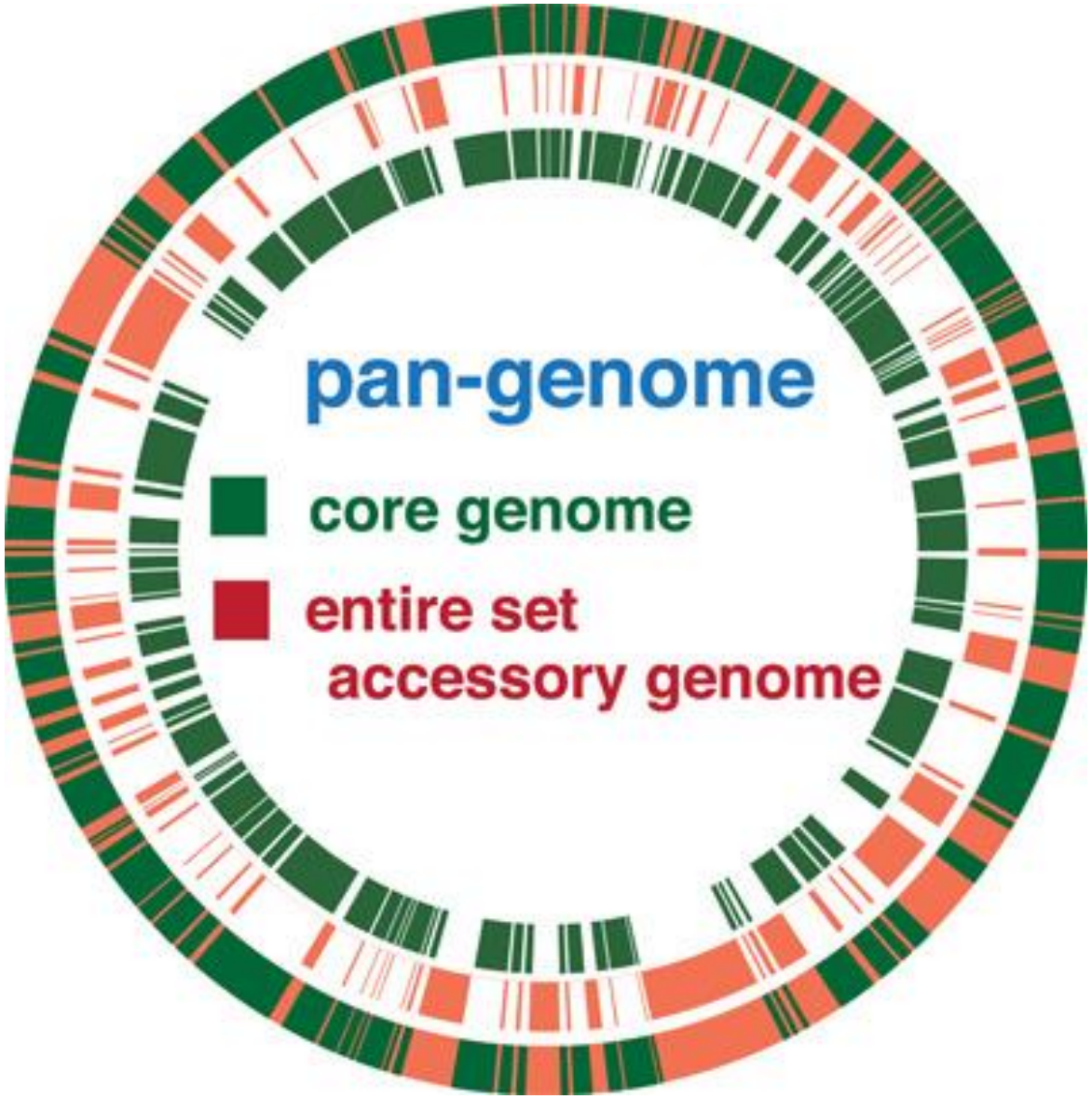
Four strains of *A. baumannii* were recently sequenced: *A. baumannii* ATCC 17978, an historic strain from 1951 that was implicated in fatal meningitis in a 4-month-old baby; *A. baumannii* SDF, which was isolated from a human-body louse in France; *A. baumannii* AYE, which was isolated in 2001 during 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)


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

Core genome pan genome calculation

https://www.uni-giessen.de/fbz/fb08/Inst/bioinformatik/software/EDGAR/Features/core_pan

Bioinformatics and Systems Biology | News | Research | Education | Software | BiGi | Core Facility | Publications | People | Website durchsuchen

- ASA³P
- Bakta
- Conveyor
- EDGAR
  - Features
    - Score Ratio Value Plots
    - Core/Pan genome calculation
    - Singleton gene calculation
    - Venn diagrams
    - Calculate genesets
    - Core/Singleton development plots
    - Synteny plots
    - Comparative Viewer
    - Define Metacontig / Replicon group
    - CoreHMM scanner
    - Phylogenetic trees
  - Documentation
  - Access
- Galaxy
- GenDB
- GenDBE
- MGX
- Platon
- ReadXplorer
- WASP
- ECF Hub

Core/Pan genome calculation

One main feature of EDGAR is the fast calculation of the genomic subsets of the core and pan genome. The interfaces and results for this features are identical.

The core genome as well as the pan genome calculation require the selection of one reference genome in the left selection list and the selection of comparison genomes or sets in the right selection list. Results are presented in tabular form and the reference genome will appear in the first column of the result table.

EDGAR

Choose a reference contig from the list on the left to calculate the pan genome of EDGAR_Xanthomonas for the set of contigs chosen on the right!

Start Analyses

- Synteny Plots
- Score Ratio Value Plots
- Core Genome
- Core development plot
- Venn Diagrams
- Pan Genome
- Singletons
- Singleton development plot
- Calculate genesets
- Comparative Viewer
- Phylogenetic tree
- CoreHMM scan
- Define replicon group

Logout.

Pan Genome

Pan Genome consists of 6172 CDS.

save pan genome (.csv)

save dna fas sequence

save aa fas sequence

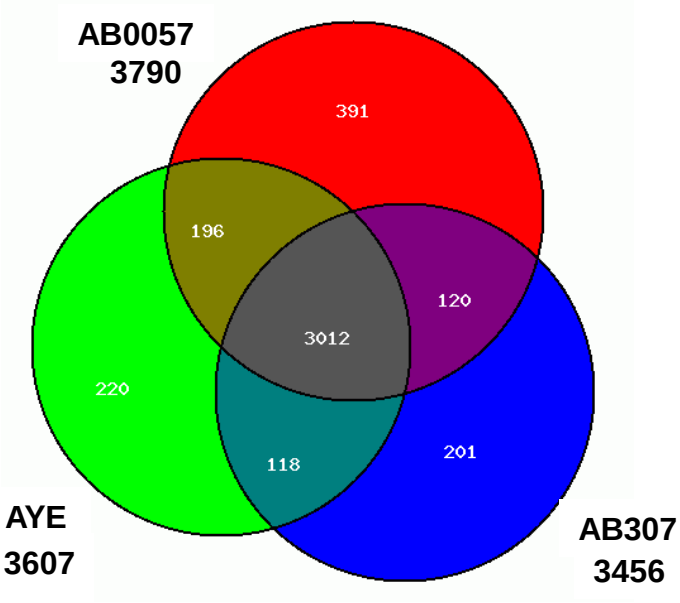
X_campestris_pv_campestris_str_B100	ALL_X_campestris_pv_vesicatoria_str_85-10	X_campestris_pv_campestris_str_8004	X
xccb100_0001 chromosomal replication initiation protein	XCV0001 chromosomal replication initiation protein	XC_0001 chromosome replication initiator DnaA	
xccb100_0002 DNA polymerase III subunit beta	XCV0002 DNA polymerase III subunit beta	XC_0002 DNA polymerase III subunit beta	
⋮			
xccb100_0010 macromolecule import protein	XCV0010 biopolymer transport ExbD protein	XC_0010 biopolymer transport ExbD1 protein	
xccb100_0011 macromolecule import protein	XCV0011 biopolymer transport ExbD protein	XC_0011 biopolymer transport ExbD2 protein	
xccb100_0012 hypothetical protein	- -	- -	
xccb100_0013 macromolecule import protein	XCV0012 biopolymer transport ExbD protein	- -	
- -	XCV0013 pyridoxine 5'-phosphate synthase	XC_0012 pyridoxine 5'-phosphate synthase	

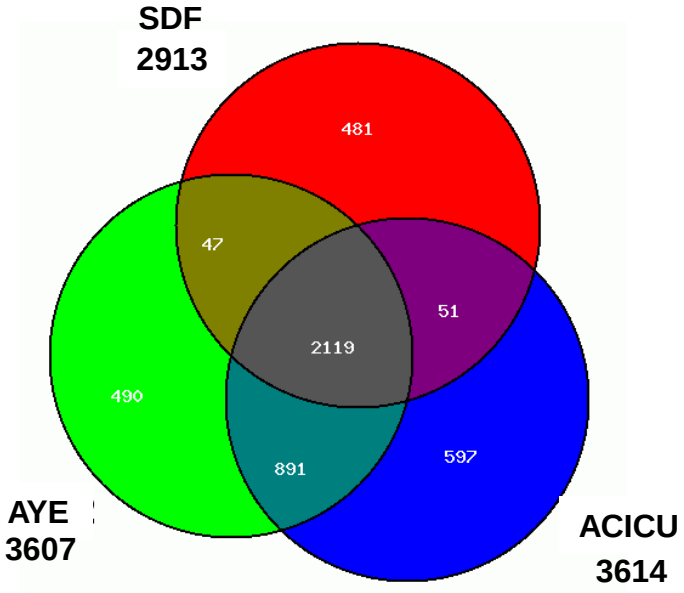
Contact

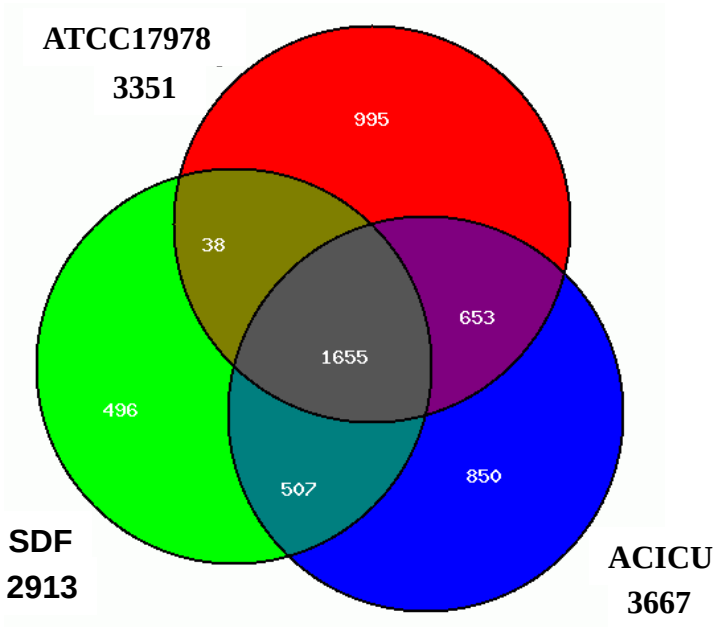
Heinrich-Buff-Ring 58
Justus-Liebig-Universität
35392 Gießen
Sekretariat Goesmann
Erdgeschoss, Raum 0003.A
Tel.: +49 (0) 641 99 35801
Fax: +49 (0) 641 99 35809
sekretariat@computational.bio

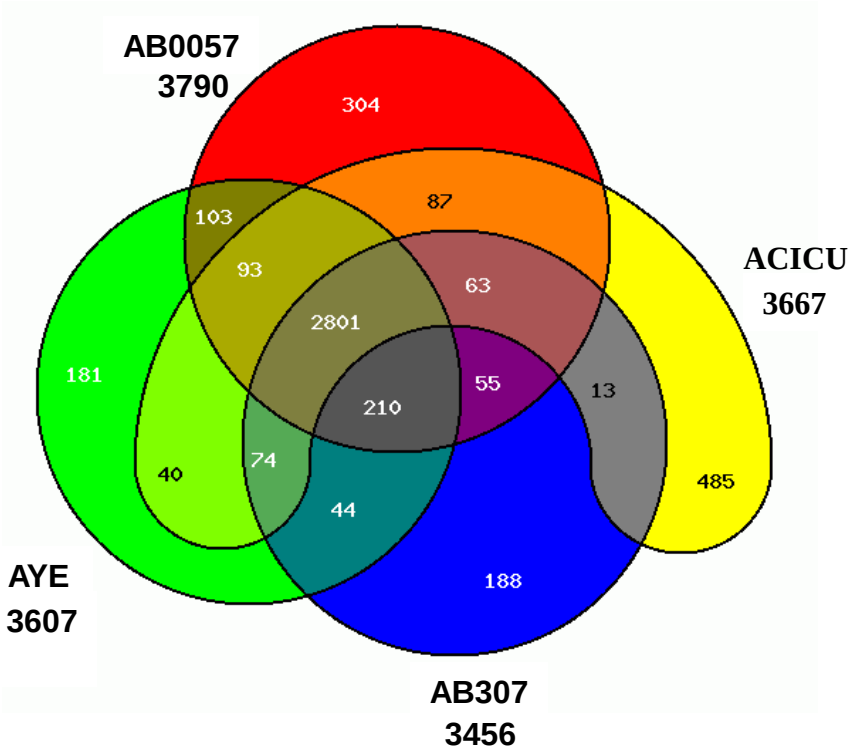
Reguläre Sprechzeiten:
Montag bis Freitag 09:00 - 12:00 Uhr

How to find us ...

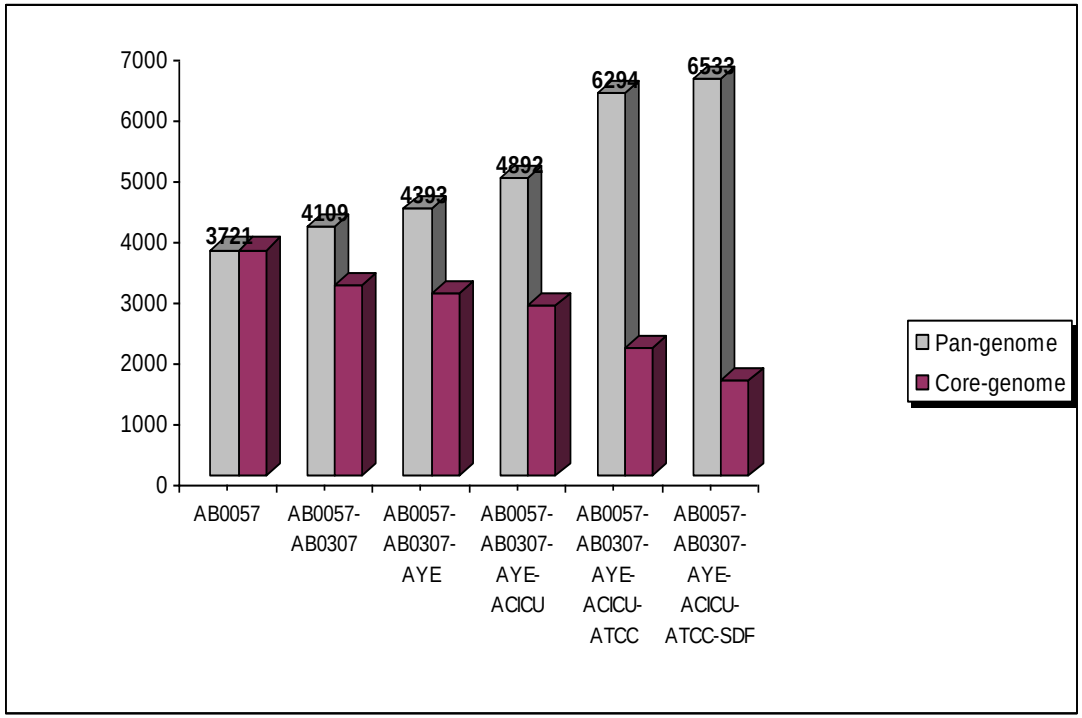


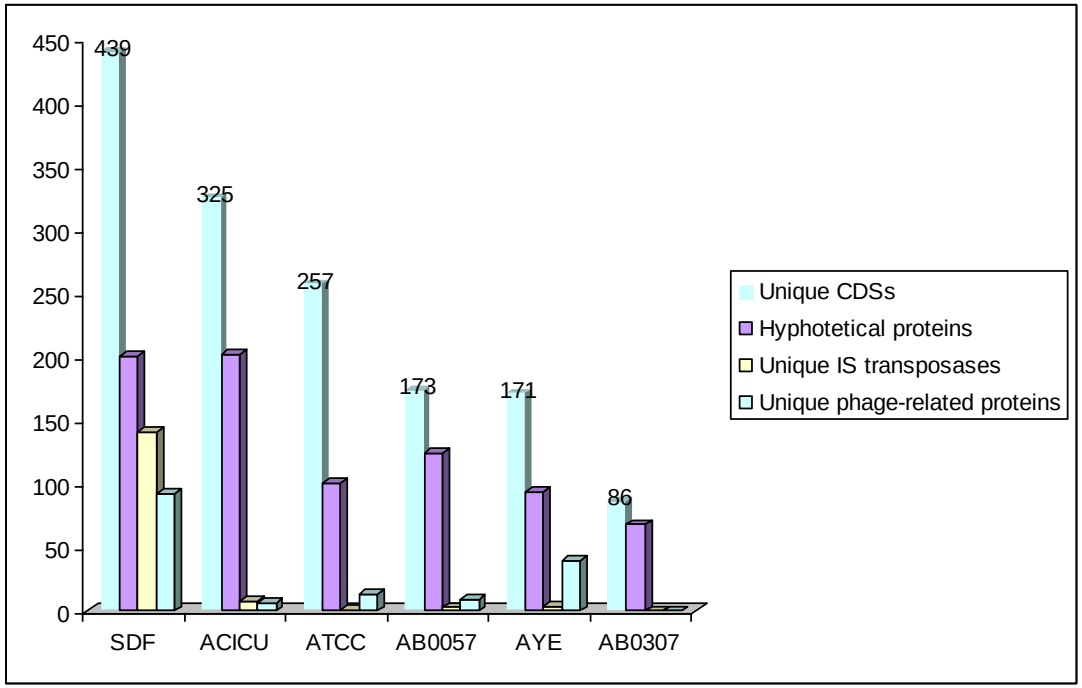






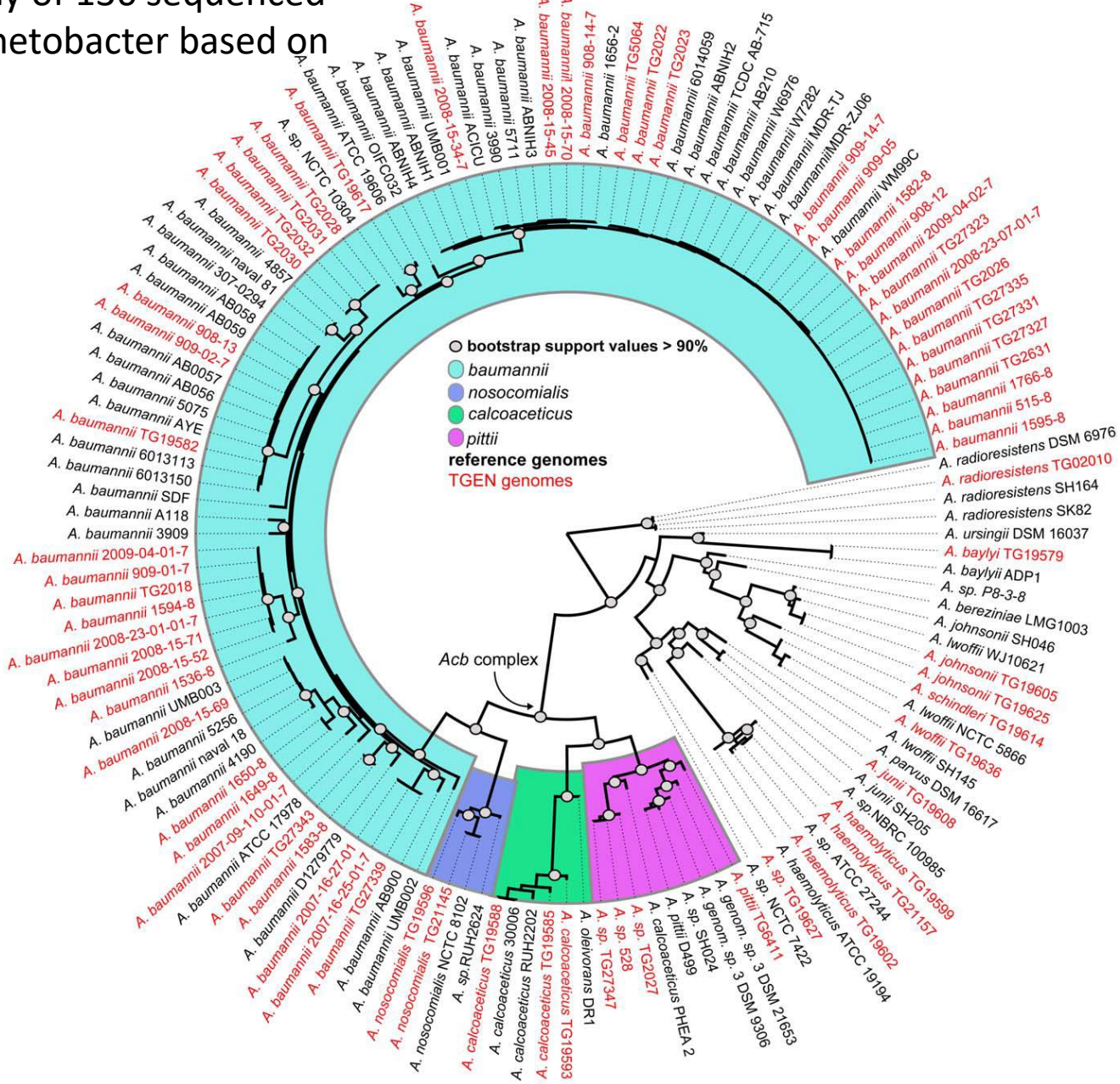
The genomics of Acinetobacter baumannii: insights into genome plasticity, antimicrobial resistance and pathogenicity. **Imperi** et al IUBMB Life. 2011 Dec;63(12):1068-74





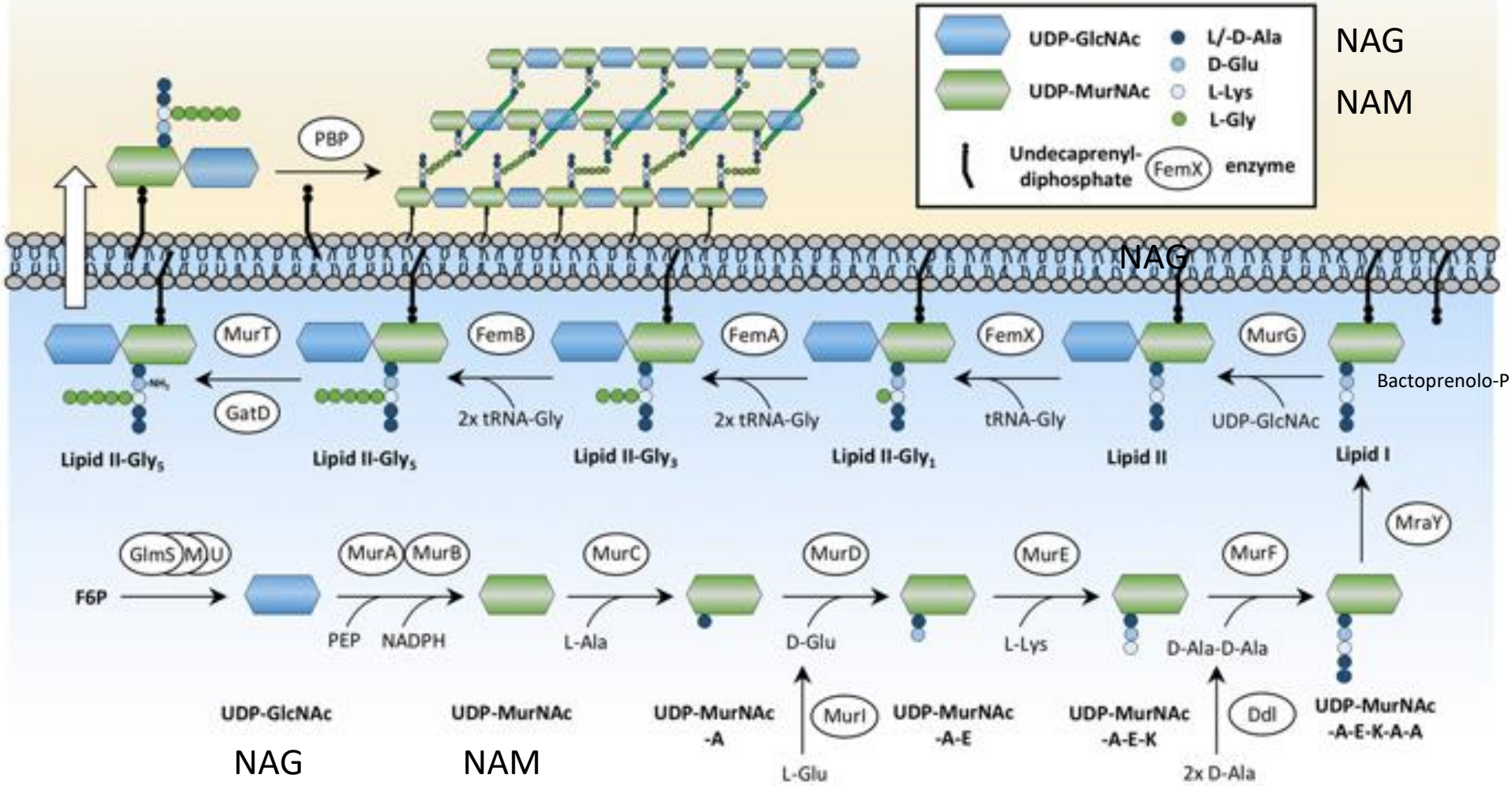
The genomics of Acinetobacter baumannii: insights into genome plasticity, antimicrobial resistance and pathogenicity. **Imperi** et al IUBMB Life. 2011 Dec;63(12):1068-74

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



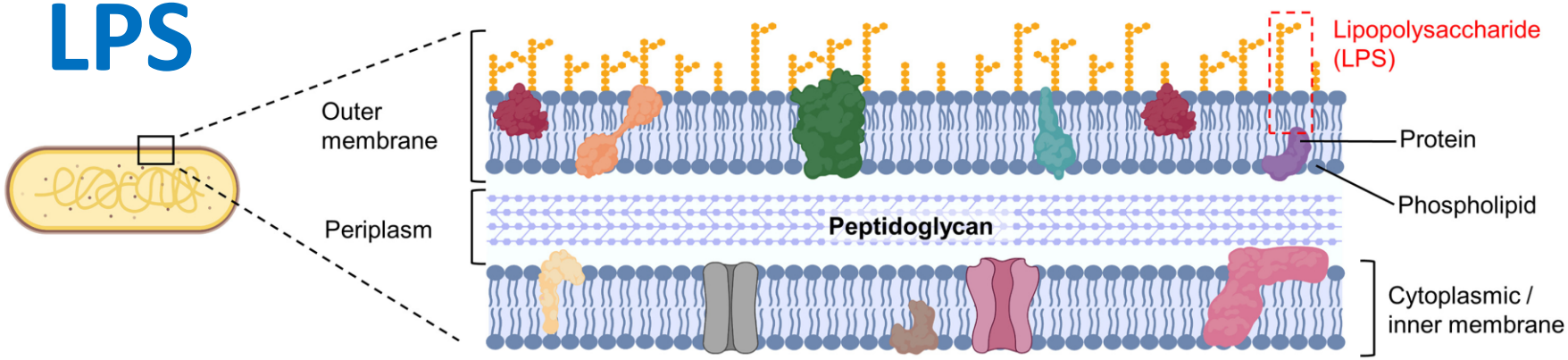
Acido mesodiaminopimelico
DAP nei Gram-

Lipopolysaccharide (LPS)

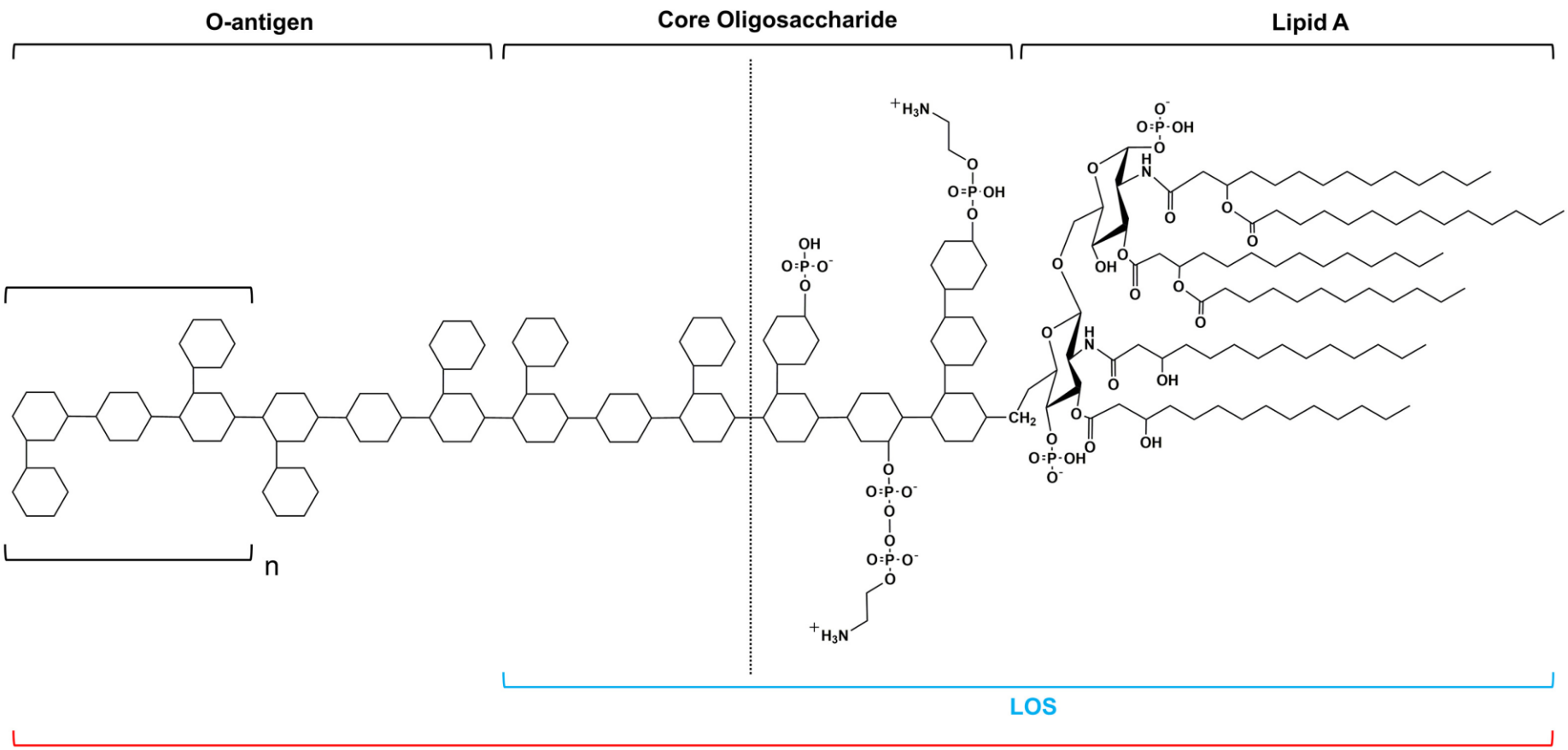
- Endotoxin or Pyrogen
  - Fever causing
  - Toxin nomenclature
    - Endo- part of bacteria
    - Exo- excreted into environment
- Structure
  - Lipid A
  - Polysaccharide
    - O Antigen of E. coli, Salmonella
- G- bacteria only
  - Alcohol/Acetone removes

(a)

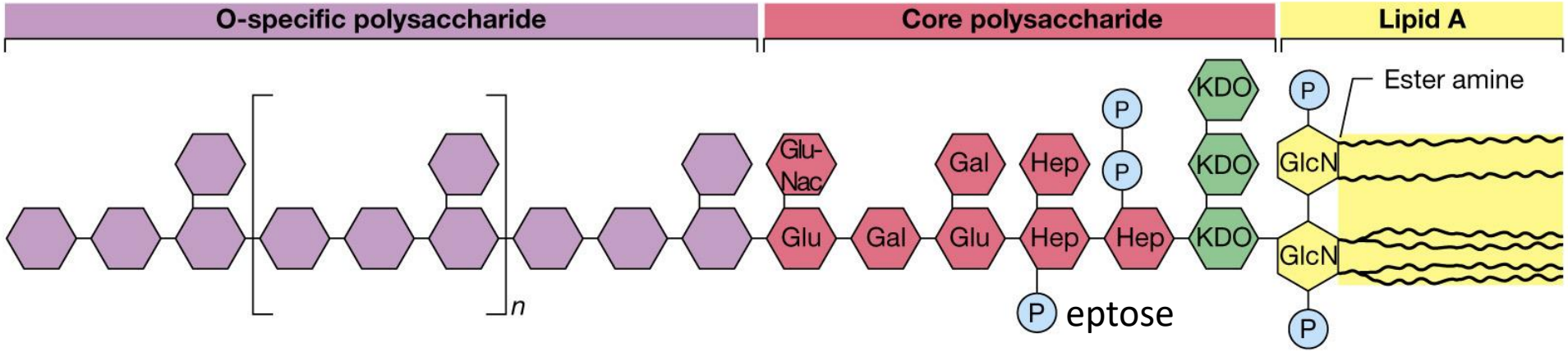
LPS


Outer membrane
Periplasm
Lipopolysaccharide (LPS)
Protein
Phospholipid
Peptidoglycan
Cytoplasmic / inner membrane

(b)


O-antigen
Core Oligosaccharide
Lipid A
n
LOS

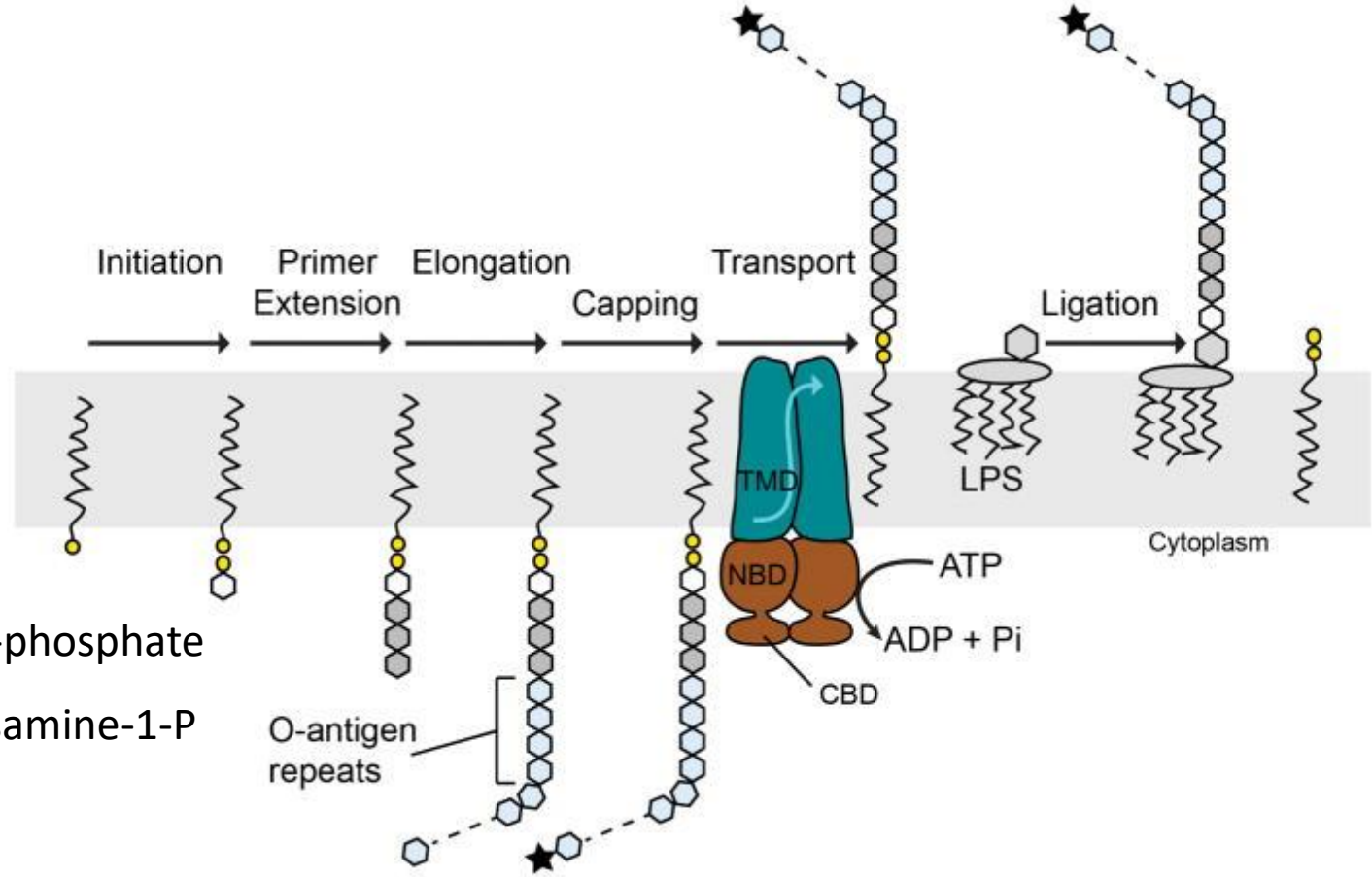
LPS



Chain of tetra or penta-saccharides

N-Acetylgluc

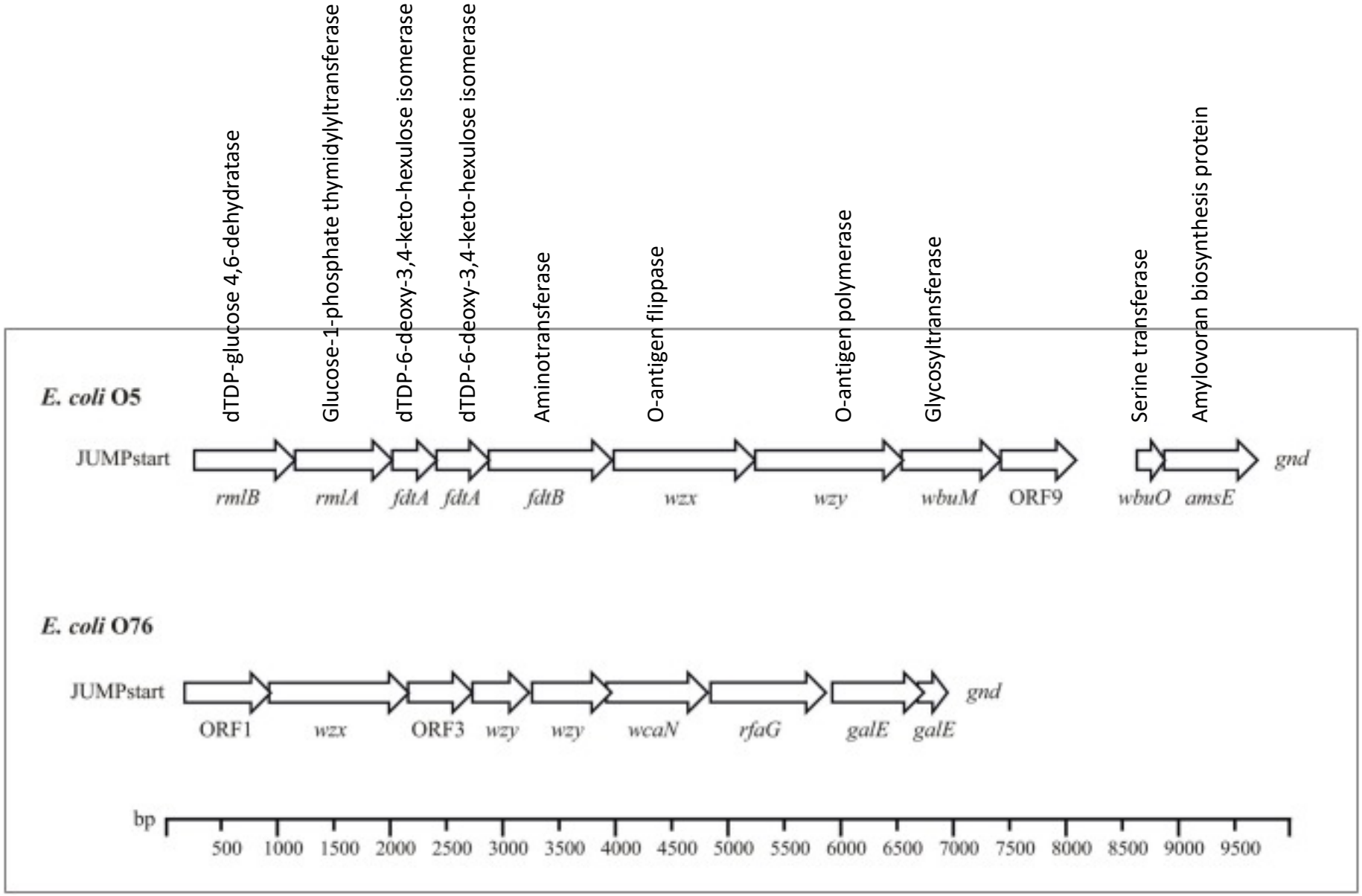
2-keto-3deoptonatec



Undecaprenyl-phosphate

N-acetylglucosamine-1-P

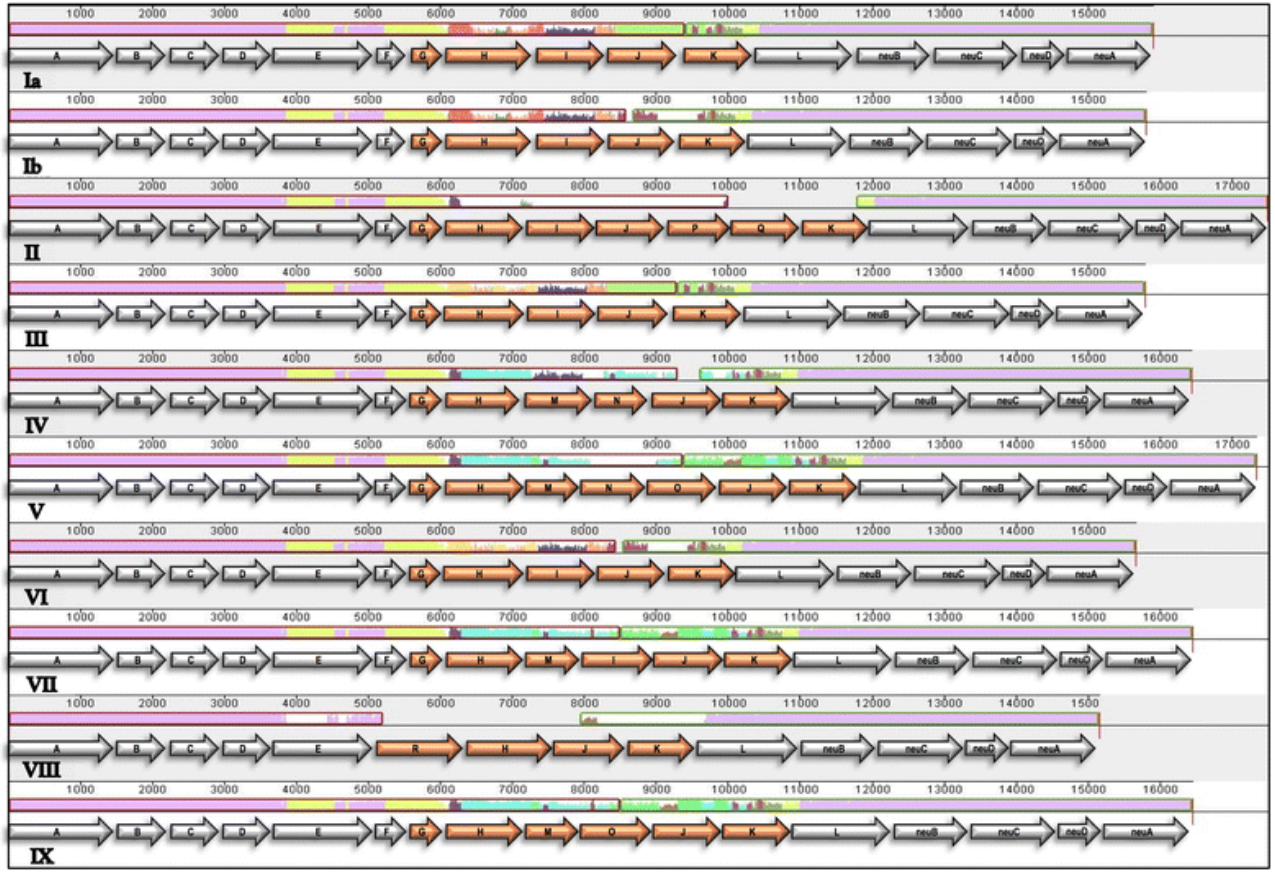
Architecture of a channel-forming O-antigen polysaccharide ABC transporter
Bi et al. Nature. 2018 Jan 18; 553(7688): 361–365.

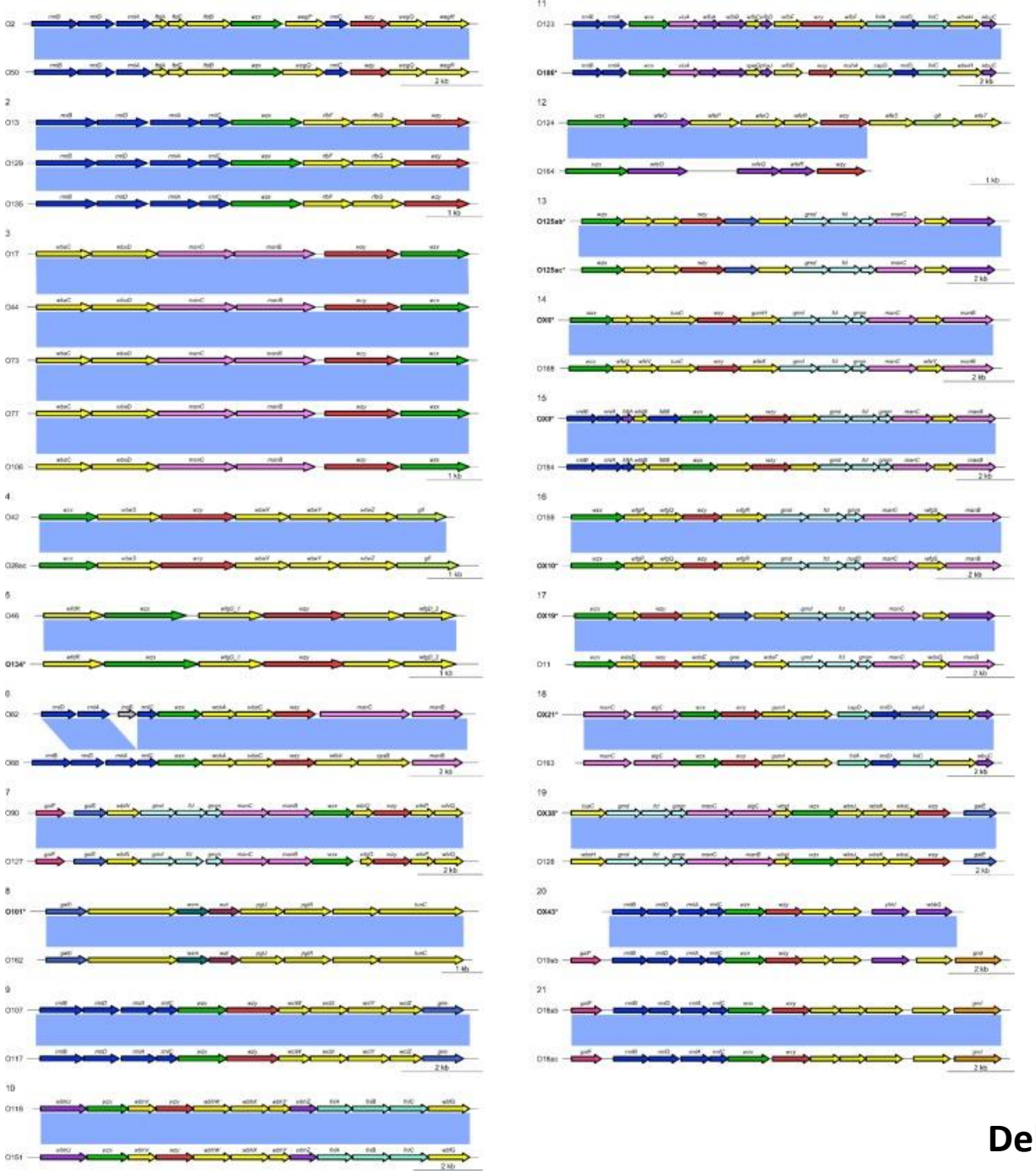

dTDP-glucose 4,6-dehydratase
Glucose-1-phosphate thymidylyltransferase
dTDP-6-deoxy-3,4-keto-hexulose isomerase
dTDP-6-deoxy-3,4-keto-hexulose isomerase
Aminotransferase
O-antigen flippase
O-antigen polymerase
Glycosyltransferase
Serine transferase
Amylovoran biosynthesis protein
E. coli O5
JUMPstart
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

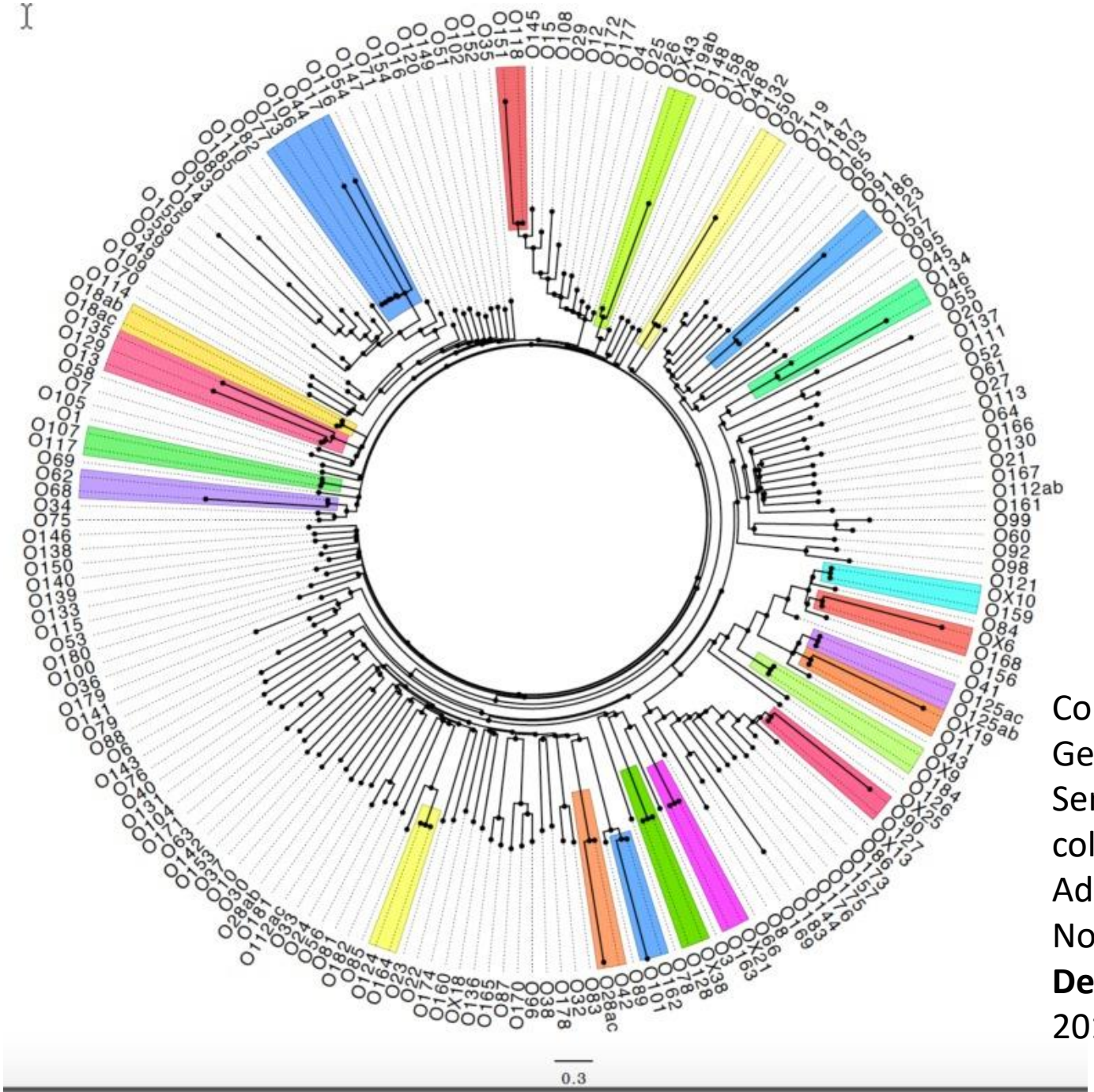

SALMONELLA
SG 6
SG 2
d + i + e,h
H:q,s,t,p,u
O:9
H:m
H:a
O:2
SG 1
a + b + c + z10
H:2
H:b
H:i
O:46
O:4
Store at 2-8°C
Exp. January 2008

In silico Serotyping



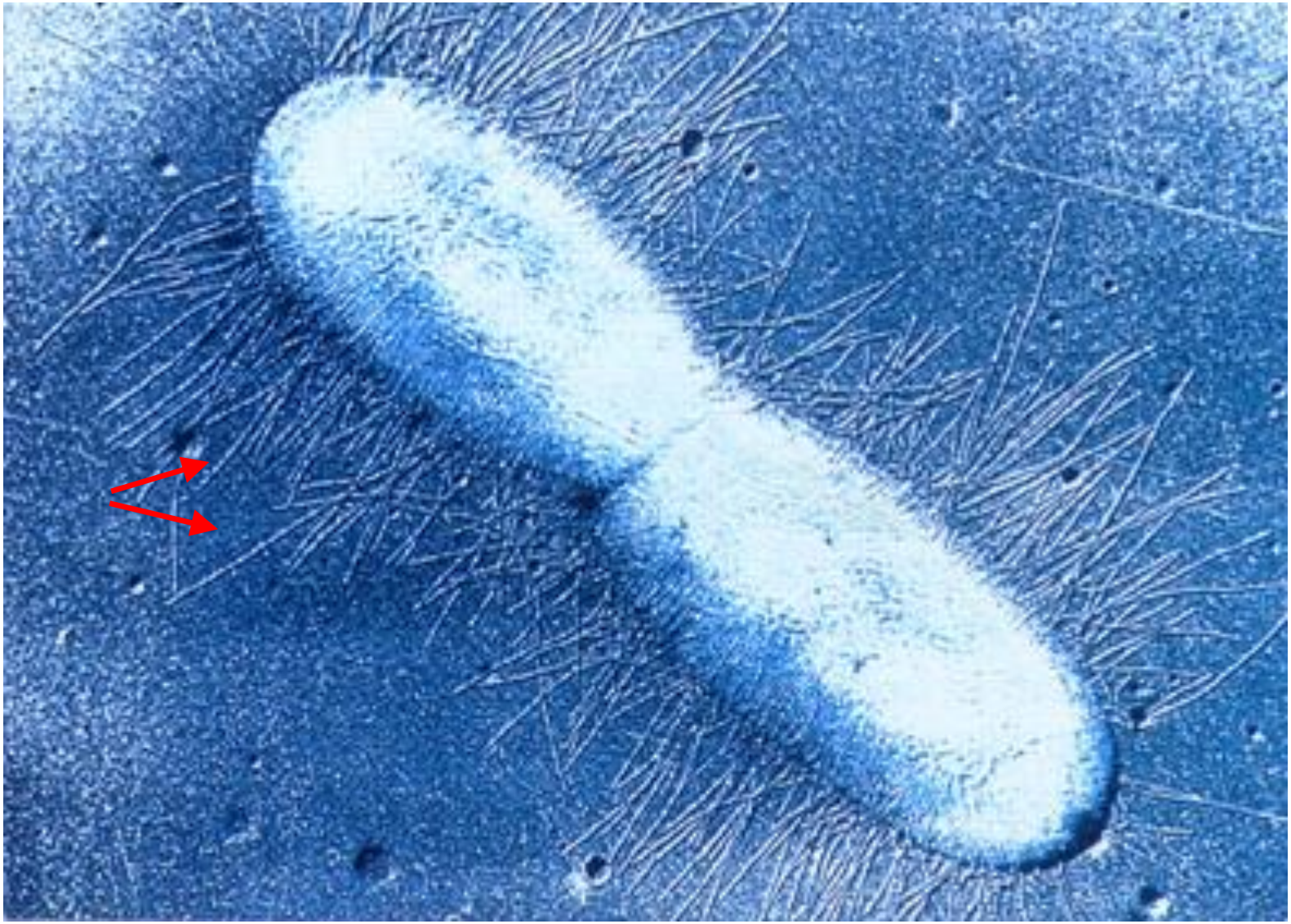


DebRoy C et al. PLoS One. 2016 Jan 29;11(1):e0147434



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

The EMBO Journal (2008) 27, 2271–2280 | © 2008 European Molecular Biology Organization | Some Rights Reserved 0261-4189/08
www.embojournal.org

THE EMBO JOURNAL
EMBO open

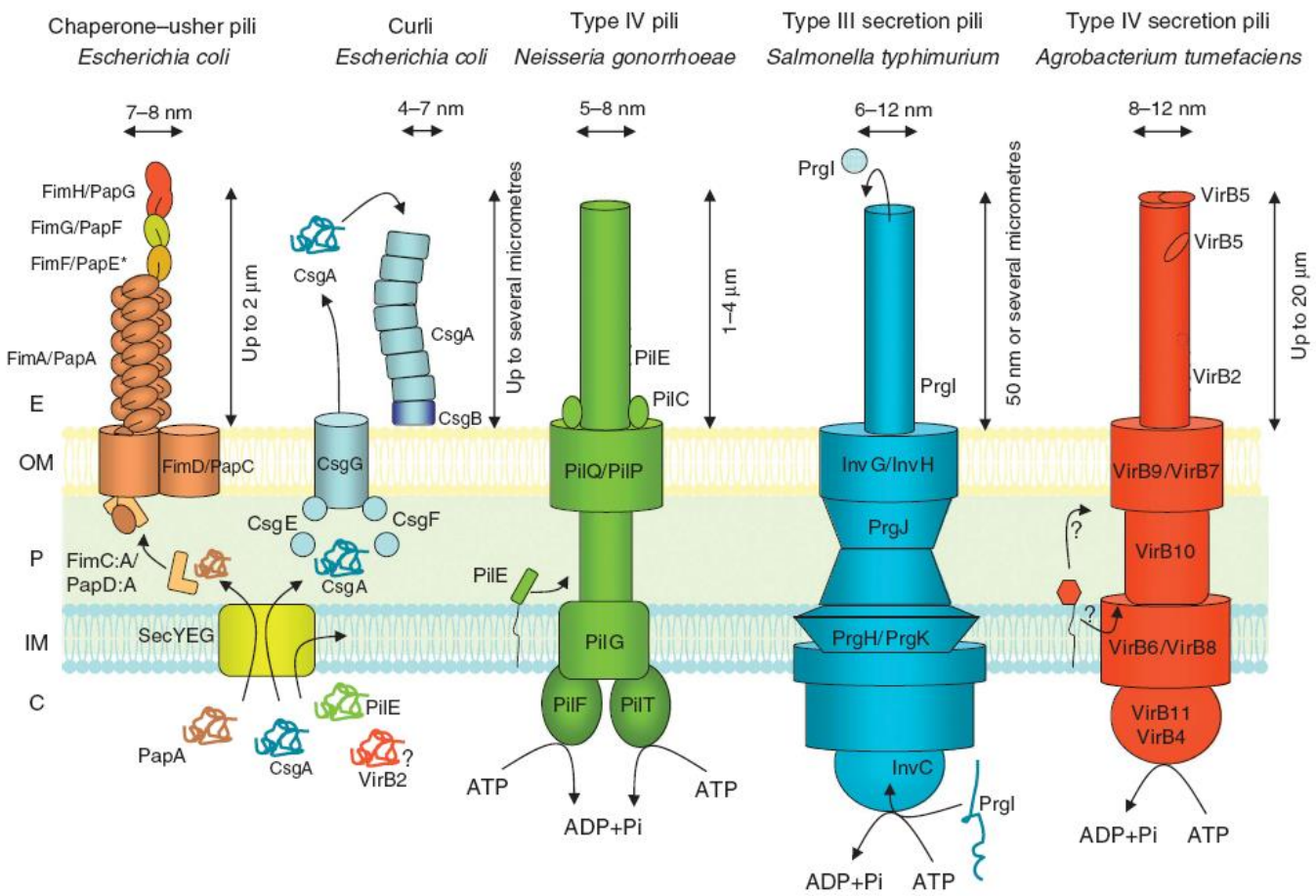
New EMBO Member's Review

Architectures and biogenesis of non-flagellar protein appendages in Gram-negative bacteria

Remi Fronzes, Han Remaut and Gabriel Waksman*

way: chaperone–usher (CU) pili, curli, type IV pili, type III secretion needle and type IV secretion pili (Figure 1).

Non-flagellar protein appendages in Gram-negative bacteria
R Fronzes *et al*

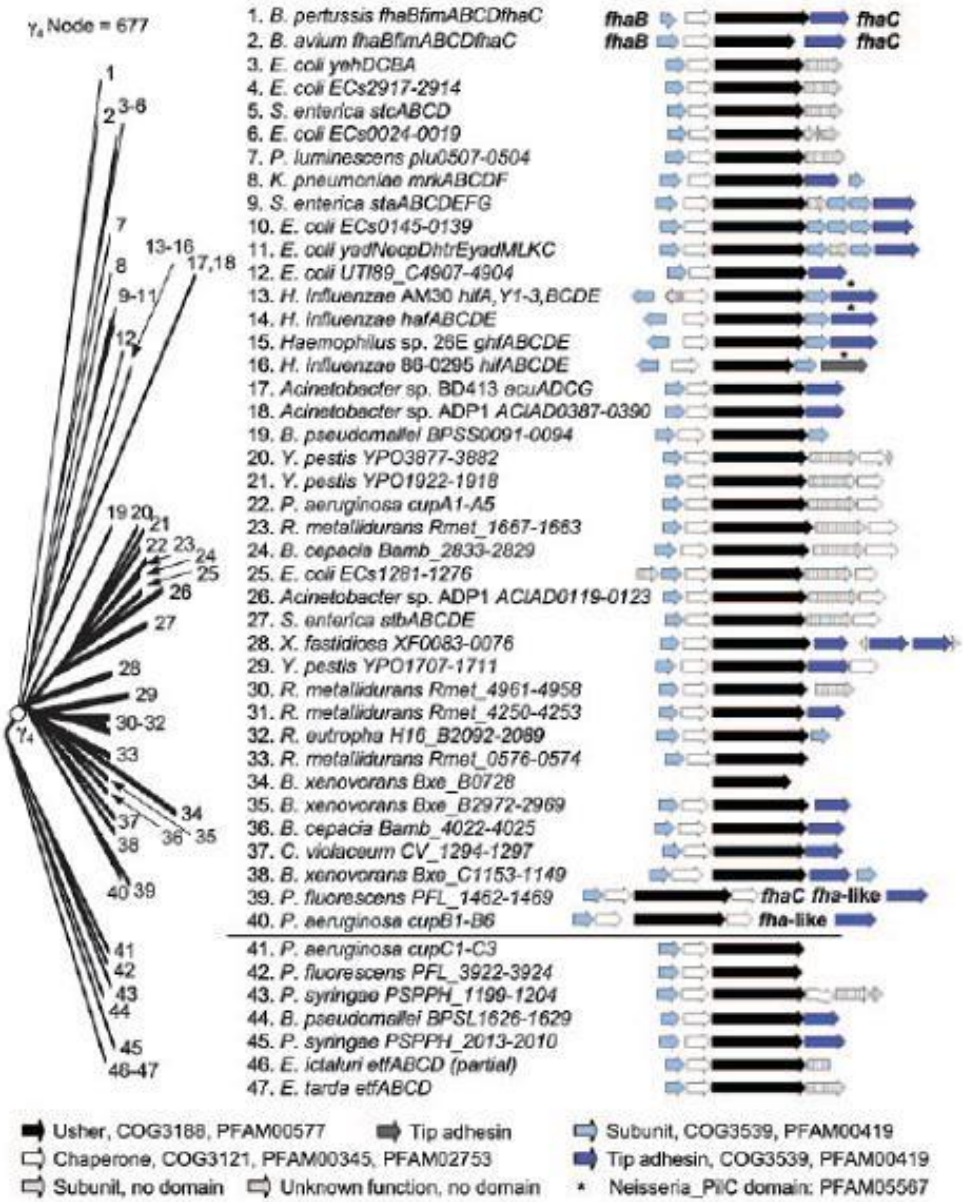


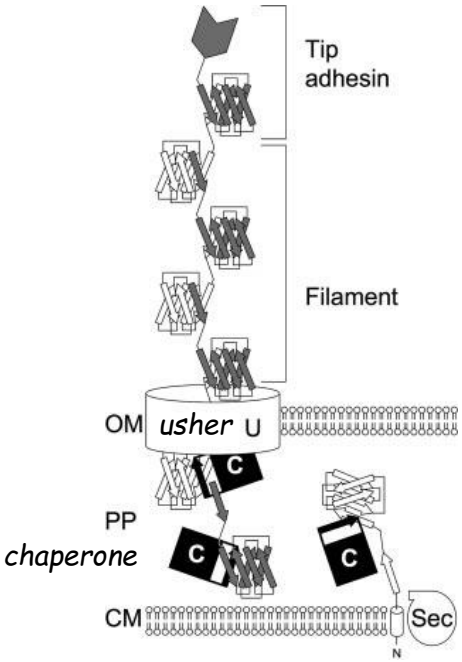
fimbriae

Microbiology and Molecular Biology Reviews, Dec. 2007, p. 551–575 Vol. 71, No. 4
1092-2172/07/$08.00+0 doi:10.1128/MMBR.00014-07

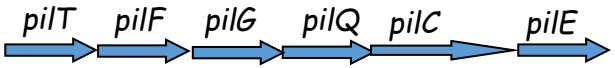
'olution of the Chaperone/Usher Assembly Pathway: Fimbrial Classification Goes Greek†

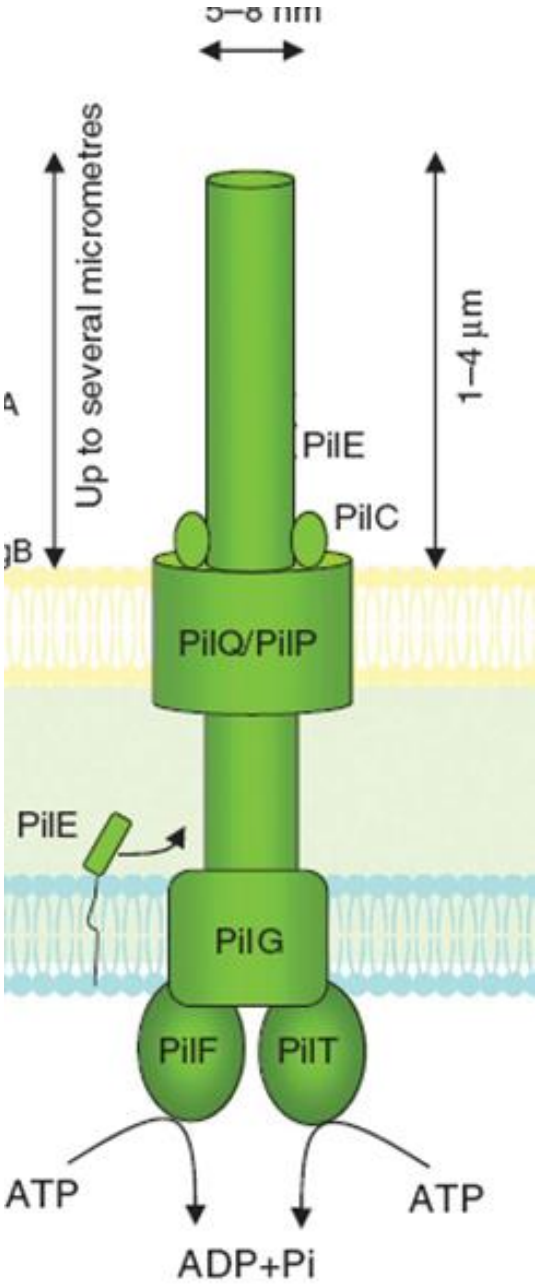
Sean-Paul Nuccio and Andreas J. Bäumler*





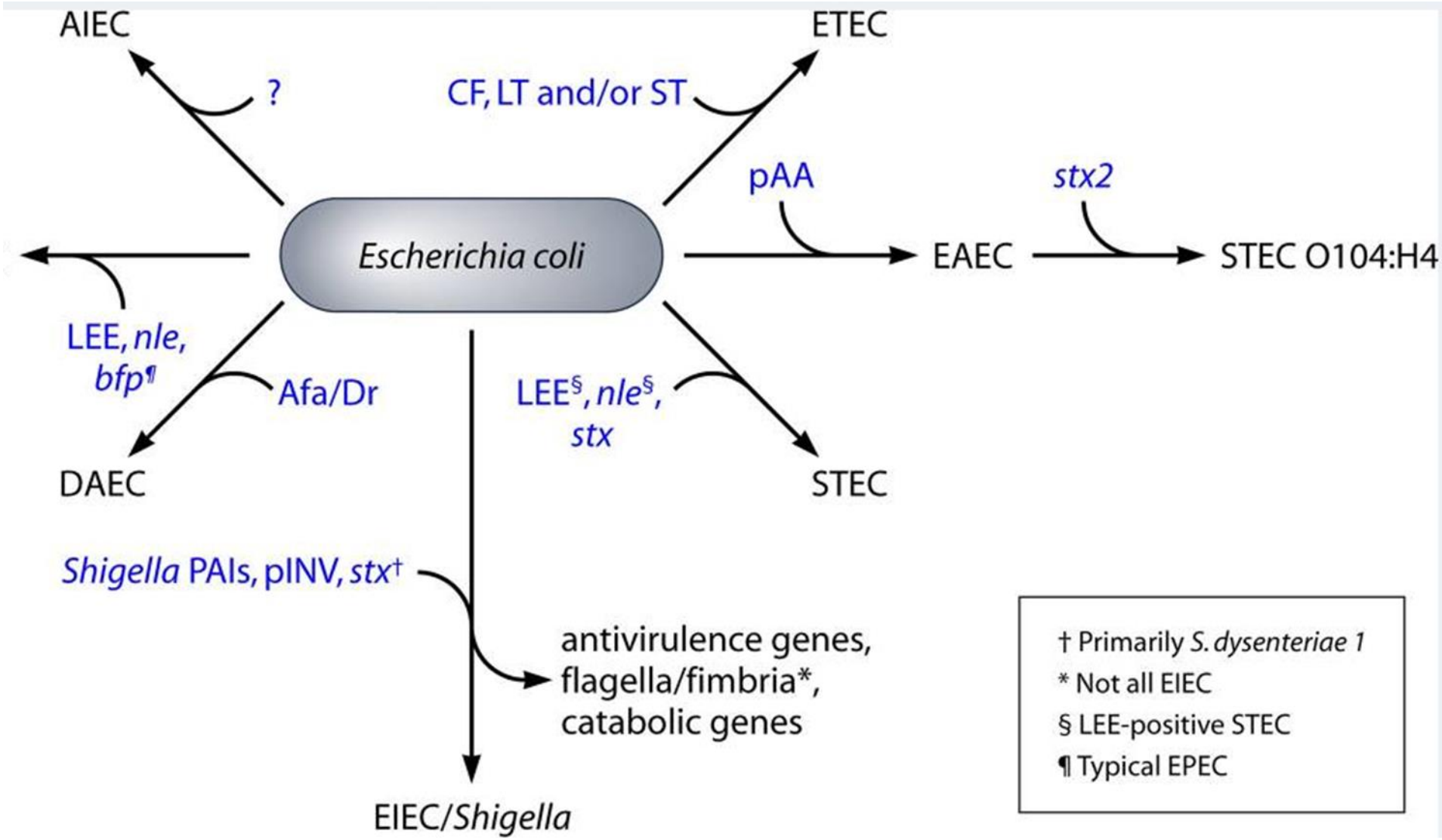
Type IV pilus





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

Google

microbial genomes ncbi

Tutti | Immagini | Notizie | Libri | Video | Altro | Strumenti

Circa 3.600.000 risultati (0,40 secondi)

nih.gov
https://www.ncbi.nlm.nih.gov › ...

Microbial Genomes - NCBI

Microbial Genomes **resource presents public data from prokaryotic genome sequencing projects**. Prokaryotes are the earliest forms of life, appearing on earth ...

Articoli accademici per **microbial genomes ncbi**

Update on RefSeq **microbial genomes** resources - Tatusova - Citato da 144

The integrated **microbial genomes** (IMG) system - Markowitz - Citato da 459

... **microbial genomes** database: new representation and ... - Tatusova - Citato da 487

nih.gov
https://www.ncbi.nlm.nih.gov › ...

Microbial Genome Resources - NCBI

Microbial Genomes Resources **presents public data from prokaryotic genome sequencing projects**. The sequence collection contains data from finished genomes as ...

Genome

Genome

Search

Microbial Genomes

Microbial Genomes resource presents public data from prokaryotic genome sequencing projects. Prokaryotes are the earliest forms of life, appearing on earth 4 billion years ago. The Prokaryotes include the Archaea, which include inhabitants of some of the most extreme environments on the planet, and the Bacteria, which include both important pathogens and producers of fermented food, antibiotics, and vitamins.

Using Microbial Genomes

- Browse microbial genomes
- Download/FTP Refseq Archaea genomes
- Download/FTP Refseq Bacteria genomes

Annotation Tools

- Prokaryotic Annotation Pipeline
- GeneMark
- Glimmer
- ORF finder

Analysis Tools

- Microbial Genomes BLAST

Genome Submission

- RegisterBioproject
- Submit SRA data
- Submit a genome
- Submission Guide

Related Resources

- Assembly
- Genome
- BioProject
- BioSample

Contact and Outreach

- NCBI Handbook

New Genome Table

Try our new **Genome page** and use the feedback button to let us know what you think

Genome > **Genome Information by Organism**

Escherichia coli | Search | Download Reports from FTP site

Overview (72067); Eukaryotes (24955); **Prokaryotes (437192)**; Viruses (51007); Plasmids (41645); Organelles (25346)

Filters | Download

Choose Columns | Page 1 of 8,744 | 50

#	Organism Name	Organism Groups	Strain	BioSample	BioProjec	Assembly	Leve	Size	GC%	F
1	'Brassica napus' phytoplasma	Bacteria;Terrabacteria group;Tenericutes	TW1	SAMN09083457	PRJNA46439	GCA_003181115.1		0.743598	27.20	
2	'Candidatus Kapabacteria' thiocyanatum	Bacteria;FCB group;Bacteroidetes/Chloro group	SCN18_14_9_16_R1_B_59	SAMN18061348	PRJNA629336	GCA_017307255.1		3.25	59.50	
3	'Candidatus Kapabacteria' thiocyanatum	Bacteria;FCB group;Bacteroidetes/Chloro group	59-99	SAMN05660602	PRJNA279279	GCA_001899175.1		3.27	59.40	
4	'Catharanthus roseus' aster yellows phytoplasma	Bacteria;Terrabacteria group;Tenericutes	De Villa	SAMN10923938	PRJNA522055	GCA_004214875.1		0.603949	28.38	chromosome: NZ_CP035 plasmid unnamed1: NZ_
5	'Chrysanthemum coronarium' phytoplasma	Bacteria;Terrabacteria group;Tenericutes	OY-V	SAMD00018609	PRJDB2922	GCA_000744065.1		0.739592	27.50	
6	'Cynodon dactylon' phytoplasma	Bacteria;Terrabacteria group;Tenericutes	LW01	SAMN12727363	PRJNA564951	GCA_009268075.1		0.483935	20.50	
7	'Echinacea purpurea' witches'-broom phytoplasma	Bacteria;Terrabacteria group;Tenericutes	NCHU2014	SAMN04017316	PRJNA294131	GCA_001307505.2		0.639808	24.52	chromosome: CP040925 plasmid pEpWB: CP040

Genome > **Genome Information by Organism**

Escherichia coli | Search | Download Reports from FTP site

Overview (1); **Prokaryotes (31230)**; Plasmids (6529)

Filters | Download

Choose Columns | Page 1 of 625 | 50

#	Organism Name	Organism Groups	Strain	BioSample	BioProjec	Assembly	Leve	Size	GC%	F
1	Escherichia coli str. K-12 substr. MG1655	Bacteria;Proteobacteria;Ga	K-12 substr. MG1655	SAMN02604091	PRJNA225	GCA_000005845.2	●	4.64	50.80	chromosome: NC_00091
2	Escherichia coli O157:H7 str. Sakai	Bacteria;Proteobacteria;Ga	Sakai substr. RIMD 0509952	SAMN01911278	PRJNA226	GCA_000008865.2	●	5.59	50.45	chromosome: NC_00269 plasmid pOS231: NC_00 plasmid pO157: NC_002
3	Escherichia coli	Bacteria;Proteobacteria;Ga	136	SAMN08773043	PRJNA445267	GCA_005221645.1	●	5.56	50.55	chromosome: NZ_CP028 plasmid pTA136: NZ_CP plasmid pTA136-2: NZ_C
4	Escherichia coli	Bacteria;Proteobacteria;Ga	140	SAMN08773047	PRJNA445267	GCA_005221925.1	●	5.56	50.55	chromosome: NZ_CP028 plasmid pTA140: NZ_CP plasmid pTA140-2: NZ_C
5	Escherichia coli	Bacteria;Proteobacteria;Ga	117	SAMN08773029	PRJNA445267	GCA_005221825.1	●	5.56	50.55	chromosome: NZ_CP028 plasmid pTA117: NZ_CP plasmid pTA117-2: NZ_C
6	Escherichia coli	Bacteria;Proteobacteria;Ga	120	SAMN08773032	PRJNA445267	GCA_005221805.1	●	5.56	50.55	chromosome: NZ_CP028 plasmid pTA120: NZ_CP plasmid pTA120-2: NZ_C
7	Escherichia coli	Bacteria;Proteobacteria;Ga	138	SAMN08773045	PRJNA445267	GCA_005221965.1	●	5.56	50.55	chromosome: NZ_CP028 plasmid pTA138: NZ_CP plasmid pTA138-2: NZ_C
8	Escherichia coli	Bacteria;Proteobacteria;Ga	121	SAMN08773033	PRJNA445267	GCA_005222025.1	●	5.56	50.55	chromosome: NZ_CP028 plasmid pTA121: NZ_CP plasmid pTA121-2: NZ_C

GenBank Send to:

Due to the large size of this record, sequence and annotated features are not shown. Use the "Customize view" panel to change the display.

Escherichia coli str. K-12 substr. MG1655, complete genome

NCBI Reference Sequence: NC_000913.3

FASTA Graphics

Go to:

```
LOCUS       NC_000913            4641652 bp    DNA     circular CON 09-MAR-2022
DEFINITION  Escherichia coli str. K-12 substr. MG1655, complete genome.
ACCESSION   NC_000913
VERSION     NC_000913.3
DBLINK      BioProject: PRJNA57779
            BioSample: SAMN02604091
KEYWORDS    RefSeq.
SOURCE      Escherichia coli str. K-12 substr. MG1655
  ORGANISM  Escherichia coli str. K-12 substr. MG1655
            Bacteria; Pseudomonadota; Gammaproteobacteria; Enterobacterales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 4641652)
  AUTHORS   Riley,M., Abe,T., Arnaud,M.B., Berlyn,M.K., Blattner,F.R.,
            Chaudhuri,R.R., Glasner,J.D., Horiuchi,T., Keseler,I.M., Kosuge,T.,
            Mori,H., Perna,N.T., Plunkett,G. III, Rudd,K.E., Serres,M.H.,
            Thomas,G.H., Thomson,N.R., Wishart,D. and Wanner,B.L.
  TITLE     Escherichia coli K-12: a cooperatively developed annotation
            snapshot--2005
  JOURNAL   Nucleic Acids Res. 34 (1), 1-9 (2006)
   PUBMED   16397293
  REMARK    Publication Status: Online-Only
REFERENCE   2  (bases 1 to 4641652)
  AUTHORS   Hayashi,K., Morooka,N., Yamamoto,Y., Fujita,K., Isono,K., Choi,S.,
            Ohtsubo,E., Baba,T., Wanner,B.L., Mori,H. and Horiuchi,T.
  TITLE     Highly accurate genome sequences of Escherichia coli K-12 strains
            MG1655 and W3110
  JOURNAL   Mol. Syst. Biol. 2, 2006 (2006)
   PUBMED   16738553
REFERENCE   3  (bases 1 to 4641652)
```

GCF_000005845.2_ASM584v2_genomic - Blocco note

File Modifica Visualizza

```
>NC_000913.3 Escherichia coli str. K-12 substr. MG1655, complete genome
AGCTTTTCATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGATTAAAAAAAGAGTGTCTGATAGCAGCTTCTGAACTG
GTTACCTGCCGTGAGTAAATTAAAATTTTATTGACTTAGGTCACTAAATACTTTAACCAATATAGGCATAGCGCACAGAC
AGATAAAAATTACAGAGTACACAACATCCATGAAACGCATTAGCACCACCATTACCACCACCATCACCATTACCACAGGT
AACGGTGCGGGCTGACGCGTACAGGAAACACAGAAAAAAGCCCGCACCTGACAGTGCGGGCTTTTTTTTTCGACCAAAGG
TAACGAGGTAACAACCATGCGAGTGTTGAAGTTCGGCGGTACATCAGTGGCAAATGCAGAACGTTTTCTGCGTGTTGCCG
ATATTCTGGAAAGCAATGCCAGGCAGGGGCAGGTGGCCACCGTCCTCTCTGCCCCCGCCAAAATCACCAACCACCTGGTG
GCGATGATTGAAAAAACCATTAGCGGCCAGGATGCTTTACCCAATATCAGCGATGCCGAACGTATTTTTGCCGAACTTTT
GACGGGACTCGCCGCCGCCCAGCCGGGGTTCCCGCTGGCGCAATTGAAAACTTTCGTCGATCAGGAATTTGCCCAAATAA
AACATGTCCTGCATGGCATTAGTTTGTTGGGGCAGTGCCCGGATAGCATCAACGCTGCGCTGATTTGCCGTGGCGAGAAA
ATGTCGATCGCCATTATGGCCGGCGTATTAGAAGCGCGCGGTCACAACGTTACTGTTATCGATCCGGTCGAAAAACTGCT
GGCAGTGGGGCATTACCTCGAATCTACCGTCGATATTGCTGAGTCCACCCGCCGTATTGCGGCAAGCCGCATTCCGGCTG
ATCACATGGTGCTGATGGCAGGTTTCACCGCCGGTAATGAAAAAGGCGAACTGGTGGTGCTTGGACGCAACGGTTCCGAC
TACTCTGCTGCGGTGCTGGCTGCCTGTTTACGCGCCGATTGTTGCGAGATTTGGACGGACGTTGACGGGGTCTATACCTG
CGACCCGCGTCAGGTGCCCGATGCGAGGTTGTTGAAGTCGATGTCCTACCAGGAAGCGATGGAGCTTTCCTACTTCGGCG
CTAAAGTTCTTCACCCCCGCACCATTACCCCCATCGCCCAGTTCCAGATCCCTTGCCTGATTAAAAATACCGGAAATCCT
CAAGCACCAGGTACGCTCATTGGTGCCAGCCGTGATGAAGACGAATTACCGGTCAAGGGCATTTCCAATCTGAATAACAT
GGCAATGTTCAGCGTTTCTGGTCCGGGGATGAAAGGGATGGTCGGCATGGCGGCGCGCGTCTTTGCAGCGATGTCACGCG
CCCGTATTTCCGTGGTGCTGATTACGCAATCATCTTCCGAATACAGCATCAGTTTCTGCGTTCCACAAAGCGACTGTGTG
CGAGCTGAACGGGCAATGCAGGAAGAGTTCTACCTGGAACTGAAAGAAGGCTTACTGGAGCCGCTGGCAGTGACGGAACG
GCTGGCCATTATCTCGGTGGTAGGTGATGGTATGCGCACCTTGCGTGGGATCTCGGCGAAATTCTTTGCCGCACTGGCCC
GCGCCAATATCAACATTGTCGCCATTGCTCAGGGATCTTCTGAACGCTCAATCTCTGTCGTGGTAAATAACGATGATGCG
ACCACTGGCGTGCGCGTTACTCATCAGATGCTGTTCAATACCGATCAGGTTATCGAAGTGTTTGTGATTGGCGTCGGTGG
CGTTGGCGGTGCGCTGCTGGAGCAACTGAAGCGTCAGCAAAGCTGGCTGAAGAATAAACATATCGACTTACGTGTCTGCG
GTGTTGCCAACTCGAAGGCTCTGCTCACCAATGTACATGGCCTTAATCTGGAAAACTGGCAGGAAGAACTGGCGCAAGCC
AAAGAGCCGTTTAATCTCGGGCGCTTAATTCGCCTCGTGAAAGAATATCATCTGCTGAACCCGGTCATTGTTGACTGCAC
TTCCAGCCAGGCAGTGGCGGATCAATATGCCGACTTCCTGCGCGAAGGTTTCCACGTTGTCACGCCGAACAAAAAGGCCA
ACACCTCGTCGATGGATTACTACCATCAGTTGCGTTATGCGGCGGAAAAATCGCGGCGTAAATTCCTCTATGACACCAAC
GTTGGGGCTGGATTACCGGTTATTGAGAACCTGCAAAATCTGCTCAATGCAGGTGATGAATTGATGAAGTTCTCCGGCAT
TCTTTCTGGTTCGCTTTCTTATATCTTCGGCAAGTTAGACGAAGGCATGAGTTTCTCCGAGGCGACCACGCTGGCGCGGG
AAATGGGTTATACCGAACCGGACCCGCGAGATGATCTTTCTGGTATGGATGTGGCGCGTAAACTATTGATTCTCGCTCGT
GAAACGGGACGTGAACTGGAGCTGGCGGATATTGAAATTGAACCTGTGCTGCCCGCAGAGTTTAACGCCGAGGGTGATGT
TGCCGCTTTTATGGCGAATCTGTCACAACTCGACGATCTCTTTGCCGCGCGCGTGGCGAAGGCCCGTGATGAAGGAAAAG
TTTTGCGCTATGTTGGCAATATTGATGAAGATGGCGTCTGCCGCGTGAAGATTGCCGAAGTGGATGGTAATGATCCGCTG
TTCAAAGTGAAAAATGGCGAAAACGCCCTGGCCTTCTATAGCCACTATTATCAGCCGCTGCCGTTGGTACTGCGCGGATA
TGGTGCGGGCAATGACGTTACAGCTGCCGGTGTCTTTGCTGATCTGCTACGTACCCTCTCATGGAAGTTAGGAGTCTGAC
ATGGTTAAAGTTTATGCCCCGGCTTCCAGTGCCAATATGAGCGTCGGGTTTGATGTGCTCGGGGCGGCGGTGACACCTGT
TGATGGTGCATTGCTCGGAGATGTAGTCACGGTTGAGGCGGCAGAGACATTCAGTCTCAACAACCTCGGACGCTTTGCCG
```

15:28
25/09/2022

GCF_000008865.2_ASM886v2_genomic - Blocco note

File Modifica Visualizza

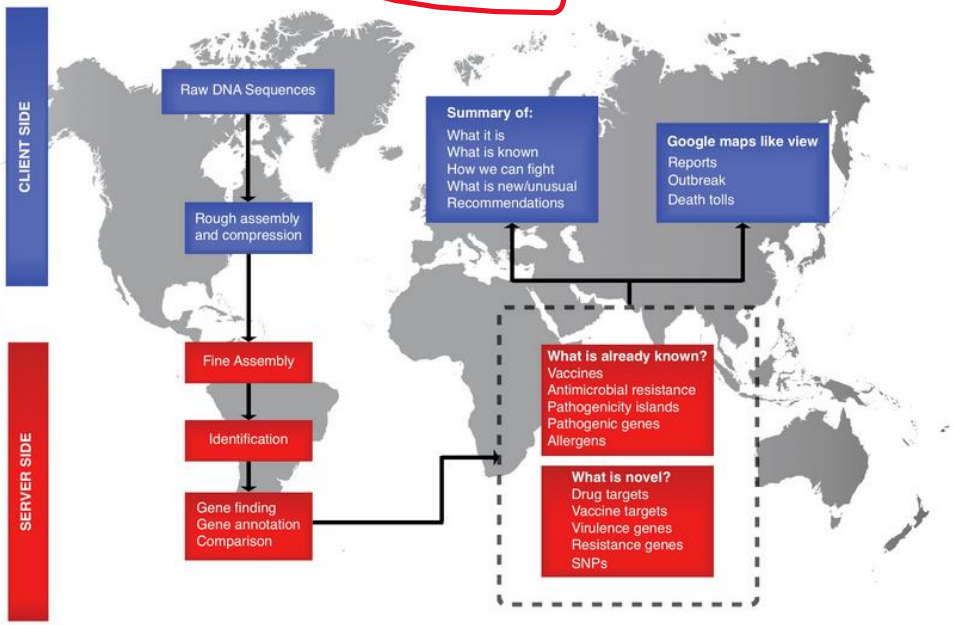
```
>NC_002695.2 Escherichia coli O157:H7 str. Sakai DNA, complete genome
AGCTTTTCATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGATTAAAAAAAGAGTCTCTGACAGCAGCTTCTGAACTG
GTTACCTGCCGTGAGTAAATTAAAATTTTATTGACTTAGGTCACTAAATACTTTAACCAATATAGGCATAGCGCACAGAC
AGATAAAAATTACAGAGTACACAACATCCATGAAACGCATTAGCACCACCATTACCACCACCATCACCACCACCATCACC
ATTACCATTACCACAGGTAACGGTGCGGGCTGACGCGTACAGGAAACACAGAAAAAAGCCCGCACCTGACAGTGCGGGCT
TTTTTTTCGACCAAAGGTAACGAGGTAACAACCATGCGAGTGTTGAAGTTCGGCGGTACATCAGTGGCAAATGCAGAACG
TTTTCTGCGGGTTGCCGATATTCTGGAAAGCAATGCCAGGCAGGGGCAGGTGGCCACCGTCCTCTCTGCCCCCGCCAAAA
TCACCAACCACCTGGTGGCGATGATTGAAAAAACCATTAGCGGCCAGGATGCTTTACCCAATATCAGCGATGCCGAACGT
ATTTTTGCCGAACTTCTGACGGGACTCGCCGCCGCCCAGCCGGGATTCCCGCTGGCGCAATTGAAAACTTTCGTCGACCA
GGAATTTGCCCAAATAAAACATGTCCTGCATGGCATTAGTTTGTTAGGGCAGTGCCCGGATAGCATTAACGCTGCGCTGA
TTTGCCGTGGCGAGAAAATGTCGATCGCCATTATGGCCGGCGTATTAGAAGCGCGCGGTCACAACGTTACCGTTATCGAT
CCGGTCGAAAAACTGCTGGCAGTGGGGCATTACCTCGAATCTACTGTCGATATTGCAGAGTCCACCCGCCGTATTGCGGC
AAGTCGTATTCCGGCTGATCACATGGTGCTGATGGCAGGTTTCACCGCCGGTAATGAAAAAGGCGAACTGGTGGTACTTG
GACGCAACGGTTCCGACTACTCCGCGGCGGTGCTGGCTGCCTGTTTACGCGCCGATTGTTGCGAGATTTGGACGGACGTT
GACGGGGTATATACCTGCGACCCGCGTCAGGTGCCCGATGCGAGGTTGTTGAAATCGATGTCCTACCAGGAAGCGATGGA
GCTTTCCTACTTCGGCGCTAAAGTTCTTCACCCCCGCACCATTACCCCCATCGCCCAGTTCCAGATCCCTTGCCTGATTA
AAAATACCGGAAATCCTCAAGCTCCAGGTACGCTCATTGGTGCCAGTCGTGATGAAGACGAATTACCGGTCAAGGGCATT
TCCAATCTGAATAATATGGCAATGTTCAGCGTTTCCGGCCCGGGGATGAAAGGAATGGTCGGCATGGCGGCGCGCGTCTT
TGCTGCAATGTCACGCGCCCGTATTTCCGTGGTGCTGATTACGCAATCATCTTCCGAATACAGTATCAGTTTCTGCGTTC
CGCAAAGCGACTGTGTGCGAGCTGAACGGGCAATGCAGGAAGAGTTCTACCTGGAACTGAAAGAAGGCTTACTGGAGCCG
CTGGCGGTGACGGAACGGCTGGCCATTATCTCGGTGGTAGGTGATGGTATGCGCACCTTGCGTGGGATCTCGGCGAAATT
CTTTGCCGCGCTGGCCCGCGCCAATATCAACATTGTCGCTATTGCTCAGGGATCTTCTGAACGCTCAATCTCTGTCGTGG
TAAATAACGATGATGCGACCACTGGCGTGCGCGTTACTCATCAGATGCTGTTCAATACCGATCAGGTTATCGAAGTGTTT
GTGATTGGCGTCGGTGGCGTTGGCGGTGCGCTGCTGGAGCAACTGAAGCGTCAGCAAAGCTGGTTGAAGAATAAACATAT
CGACTTACGTGTCTGCGGTGTTGCTAACTCGAAGGCTCTGCTCACCAATGTGCATGGCCTAAATCTGGAAAACTGGCAGG
AAGAACTGGCGCAAGCCAAAGAGCCGTTTAATCTCGGGCGCTTAATTCGCCTCGTGAAAGAATATCATCTGCTGAACCCG
GTCATTGTTGACTGCACCTCCAGCCAGGCAGTGGCGGATCAATATGCCGACTTCCTGCGCGAAGGTTTCCACGTTGTCAC
GCCGAACAAAAAGGCCAACACCTCGTCGATGGATTACTACCATCTGTTGCGTCATGCGGCTGAAAAATCGCGGCGTAAAT
TCCTCTATGACACCAACGTTGGGGCTGGATTACCGGTTATTGAGAACCTGCAAAATCTGCTCAATGCTGGTGATGAATTG
ATGAAGTTCTCCGGCATTCTTTCAGGTTCGCTTTCTTATATCTTCGGCAAGTTAGACGAAGGCATGAGTTTCTCCGAGGC
GACTACGCTGGCGCGGGAAATGGGTTATACCGAACCGGATCCGCGAGATGATCTTTCTGGTATGGATGTAGCGCGTAAAC
TATTAATTCTCGCTCGTGAAACGGGACGTGAACTGGAGCTGGCGGATATTGAAATTGAACCTGTGCTGCCCGCAGAGTTT
AACGCTGAGGGTGATGTTGCCGCTTTTATGGCGAATCTGTCACAGCTCGACGATCTCTTTGCCGCGCGCGTGGCGAAGGC
CCGTGATGAAGGAAAAGTTTTGCGCTATGTTGGCAATATTGATGAAGATGGCGTCTGCCGCGTGAAGATTGCCGAAGTGG
ATGGTAATGATCCGCTGTTCAAAGTGAAAAATGGCGAAAACGCCCTGGCCTTTTATAGCCACTATTATCAGCCGCTGCCG
```

center for genomic epidemiology
https://www.genomicepidemiology.org/

Center for Genomic Epidemiology

Home | Services | Publications | Contact

Due to scheduled maintenance, the webpage will experience down time during next week (Week 42)



News

MINTyper: an outbreak-detection method for accurate and rapid SNP typing of clonal clusters with noisy long reads
April 2021
Link to article....

Automated download and clean-up of family specific databases for kmer-based virus identification
October 2020
Link to article....

CRHP Finder, a webtool for the detection of clarithromycin resistance in Helicobacter pylori from whole-genome sequencing data
September 2020
Link to article....

ResFinder 4.0 for predictions of phenotypes from genotypes
August 2020
Link to article....

CCMetagen: comprehensive and accurate identification of eukaryotes and prokaryotes in metagenomic data
April 2020
Link to article....

Large scale automated phylogenomic analysis of bacterial isolates and the Evergreen Online platform
March 2020
Link to article....

Improved Resistance Prediction in

Welcome to the Center for Genomic Epidemiology

The use of sequencing technologies is currently transforming almost every aspect of biological science. In relation to infectious diseases, the advances are rapidly changing our scientific discoveries, as well as diagnostic and outbreak

genome ncbi - Cerca | center for genomic epidemiolog | VirulenceFinder 2.0

https://cge.food.dtu.dk/services/VirulenceFinder/

VirulenceFinder 2.0

Service | Instructions | Output | Article abstract | Citations | Version history

Software version: 2.0.3 (2020-05-21)
Database version: (2022-09-06)

The database is curated by:
Flemming Scheutz, SSI
(click to contact)

Select species

- Listeria
- S. aureus
- Escherichia coli
- Enterococcus

Select threshold for %ID

90 %

Select minimum length

60 %

15:38

genome ncbi - Cerca | center for genomic epidemiolog | VirulenceFinder 2.0

https://cge.food.dtu.dk/services/VirulenceFinder/

Select type of your reads

Only data from one single isolate should be uploaded. If raw sequencing reads are uploaded KMA will be used for mapping. KMA supports the following sequencing platforms: Illumina, Ion Torrent, Roche 454, SOLiD, Oxford Nanopore, and PacBio.

Assembled or Draft Genome/Contigs* (

Choose File(s)

Name	Size	Progress	Status

Upload Remove

* Please note also that "Assembled Genomes/Contigs" should be selected, if you have already assembled your short sequencing reads into one continuos genome or into several contigs. "Assembled Genomes/Contigs" is defined as one or several contigs in one FASTA file (one entry per contig). It is indifferent which type of short sequence reads were used to

15:38

genome ncbi - Cerca | center for genomic epidemiolog | VirulenceFinder 2.0

https://cge.food.dtu.dk/services/VirulenceFinder/

Select type of your reads

Only data from one single isolate should be uploaded. If raw sequencing reads are uploaded KMA will be used for mapping. KMA supports the following sequencing platforms: Illumina, Ion Torrent, Roche 454, SOLiD, Oxford Nanopore, and PacBio.

Assembled or Draft Genome/Contigs* (

Choose File(s)

Name	Size	Progress	Stat
GCF_000008865.2_ASM886v2_genomic.fna	5.40 MB		

Upload Remove

* Please note also that "Assembled Genomes/Contigs" should be selected, if you have already assembled your short sequencing reads into one continuos genome or into several contigs. "Assembled Genomes/Contigs" is defined as one or several contigs in one FASTA file (one entry per contig). It is indifferent which type of short sequence reads were used to

15:39

genome ncbi - Cerca | center for genomic epidemiolog | Webservices : Job queue

https://cge.food.dtu.dk/cgi-bin/webface.fcgi?jobid=6330526400004467575B7DDB;wait=

Center for Genomic Epidemiology

Your job is being processed

Wait here to watch the progress of your job, or fill in the form below to get an email message upon completion.

To get notified by email: [ ] Notify me via email

This page will update itself automatically.

15:40

			genome			
espB	100	939 / 939	NC_002695.2 Escherichia coli O157:H7 str. Sakai DNA, complete genome	4591168..4592106	Secreted protein B	AE005174
espF	100	747 / 747	NC_002695.2 Escherichia coli O157:H7 str. Sakai DNA, complete genome	4589382..4590128	Type III secretion system	AE005174
espJ	100	654 / 654	NC_002695.2 Escherichia coli O157:H7 str. Sakai DNA, complete genome	2668710..2669363	Prophage-encoded type III secretion system effector	AE005174
espP	100	3903 / 3903	NC_002128.1 Escherichia coli O157:H7 str. Sakai plasmid pO157, complete sequence	80757..84659	Extracellular serine protease plasmid-encoded	AB011549
etpD	100	1758 / 1758	NC_002128.1 Escherichia coli O157:H7 str. Sakai plasmid pO157, complete sequence	3675..5432	Type II secretion protein	AB011549
gad	100	1401 / 1401	NC_002695.2 Escherichia coli O157:H7 str. Sakai DNA, complete genome	2092027..2093427	Glutamate decarboxylase	BA000007
gad	100	1401 / 1401	NC_002695.2 Escherichia coli O157:H7 str. Sakai	4408441..4409841	Glutamate decarboxylase	BA000007

			genome			
ompT	100	954 / 954	NC_002695.2 Escherichia coli O157:H7 str. Sakai DNA, complete genome	1661315..1662268	Outer membrane protease (protein protease 7)	AKKX01000238
stx1A	100	948 / 948	NC_002695.2 Escherichia coli O157:H7 str. Sakai DNA, complete genome	2924904..2925851	Shiga toxin 1, subunit A, variant a	EF079675
stx1B	100	270 / 270	NC_002695.2 Escherichia coli O157:H7 str. Sakai DNA, complete genome	2924625..2924894	Shiga toxin 1, subunit B, variant a	AM230663
stx2A	100	960 / 960	NC_002695.2 Escherichia coli O157:H7 str. Sakai DNA, complete genome	1267107..1268066	Shiga toxin 2, subunit A, variant a	AB048837
stx2B	100	270 / 270	NC_002695.2 Escherichia coli O157:H7 str. Sakai DNA, complete genome	1268078..1268347	Shiga toxin 2, subunit B, variant a	AE005174
tccP	100	1014 / 1014	NC_002695.2 Escherichia coli O157:H7 str. Sakai DNA, complete genome	2669688..2670701	Tir-cytoskeleton coupling protein	AB253537
		1014 /	NC_002695.2 Escherichia coli		Tir-cytoskeleton	

genome ncbi - Cerca | center for genomic epidemiolog | VirulenceFinder 2.0

https://cge.food.dtu.dk/services/VirulenceFinder/

Select type of your reads

Only data from one single isolate should be uploaded. If raw sequencing reads are uploaded KMA will be used for mapping. KMA supports the following sequencing platforms: Illumina, Ion Torrent, Roche 454, SOLiD, Oxford Nanopore, and PacBio.

Assembled or Draft Genome/Contigs* (

Choose File(s)

Name	Size	Progress	Status

Upload Remove

* Please note also that "Assembled Genomes/Contigs" should be selected, if you have already assembled your short sequencing reads into one continuos genome or into several contigs. "Assembled Genomes/Contigs" is defined as one or several contigs in one FASTA file (one entry per contig). It is indifferent which type of short sequence reads were used to

15:38

genome ncbi - Cerca | center for genomic epidemiolog | VirulenceFinder 2.0

https://cge.food.dtu.dk/services/VirulenceFinder/

Select minimum length

60 %

Select type of your reads

Only data from one single isolate should be uploaded. If raw sequencing reads are uploaded KMA will be used for mapping. KMA supports the following sequencing platforms: Illumina, Ion Torrent, Roche 454, SOLiD, Oxford Nanopore, and PacBio.

Assembled or Draft Genome/Contigs* (

Choose File(s)

Name	Size	Progress	Stat
GCF_000005845.2_ASM584v2_genomic.fna	4.48 MB		

Upload Remove

15:42

Virulence factor	Identity	Template length	Contig	Position in contig	Protein function	Accession number
gad	100	1401 / 1401	NC_000913.3 Escherichia coli str. K-12 substr. MG1655, complete genome	3666180..3667580	Glutamate decarboxylase	U00096
gad	100	1401 / 1401	NC_000913.3 Escherichia coli str. K-12 substr. MG1655, complete genome	1570645..1572045	Glutamate decarboxylase	U00096
hlyE	100	918 / 918	NC_000913.3 Escherichia coli str. K-12 substr. MG1655, complete genome	1229483..1230400	Avian E.coli haemolysin	ECU57430
iss	98.98	294 / 294	NC_000913.3 Escherichia coli str. K-12 substr. MG1655, complete genome	578600..578893	Increased serum survival	CP001509
ompT	100	954 / 954	NC_000913.3 Escherichia coli str. K-12 substr. MG1655, complete genome	584680..585633	Outer membrane protease (protein protease 7)	AP009048
terC	100	714 / 714	NC_000913.3 Escherichia coli str. K-12 substr. MG1655, complete genome	2970420..2971133	Tellurium ion resistance protein	CP007491
terC	99.9	966 / 966	NC_000913.3 Escherichia coli str. K-12 substr. MG1655, complete genome	3238580..3239545	Tellurium ion resistance protein	MG591698

pMLST

Multi Locus Sequence Typing (MLST) from an assembled plasmid or from a set of reads.

cgMLSTFinder

Core genome Multi Locus Sequence Typing (cgMLST) from a set of reads.

KmerFinder

Prediction of bacterial species using a fast K-mer algorithm.

MGE

Identification of mobile genetic elements and their relation to antimicrobial resistance genes and virulence factors.

SpeciesFinder

Prediction of bacterial species using the S16 ribosomal DNA sequence.

SeroTypeFinder

Prediction of serotypes in total or partial sequenced isolates of E. coli.

SeqSero

SeqSero predicts the Salmonella serotype of either the pre-assembled or raw read sequence data provided to the service.

spaTyper

spaTyper predicts the S. aureus spa type.

FimTyper

FimTyper predicts the E. coli Fim type.

CHTyper

CHTyper predicts the E. coli FimH type and FumC type.

Other

MyKMAfinder

MyKMAfinder performs typing or pheno typing using KMA based on a user defined database.

MyDbFinder

MyDbFinder performs typing or pheno typing using blast based on a user defined database.

MyKmerFinder

MyKmerFinder performs typing or pheno typing using Kmers based on a user defined database.

DeHumanizer

The DeHumanizer web-server is a tool for human filtering based on the method described by Zhang et al.

HostPhinder

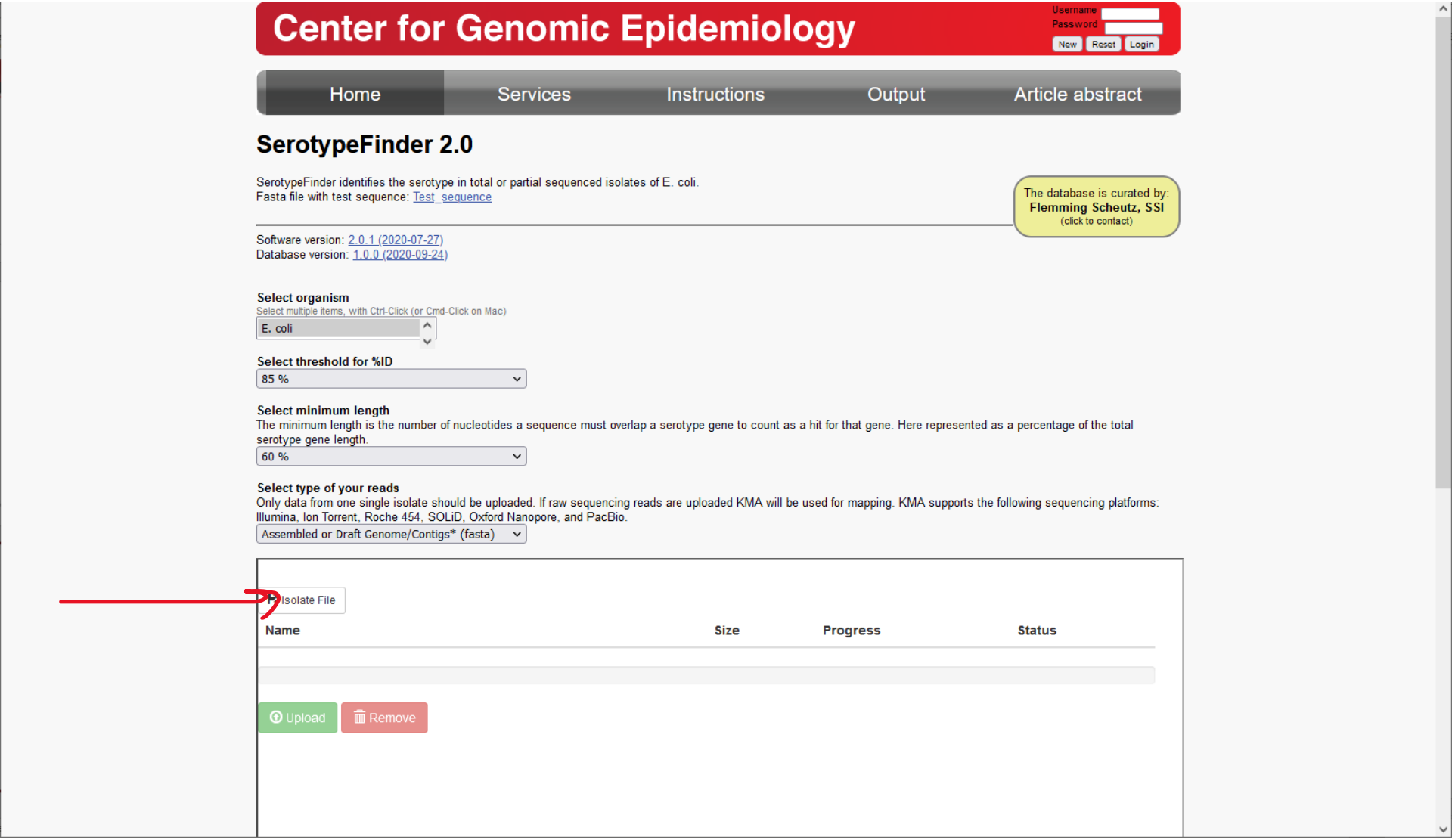
HostPhinder identifies the bacterial host of a query phage genome based on its genomic similarity to a database of phage genomes with known host.

MetaPhinder

MetaPhinder: Identifying Bacteriophage Sequences in Metagenomic Data Sets.

Plain text

- sequence.fasta



Center for Genomic Epidemiology

Home | Services | Instructions | Output | Overview of genes | Article abstract

SerotypeFinder-2.0 Server - Results

Database(s): *H_type,O_type*

Database for H type genes						
Gene	Serotype	Identity	Template / HSP length	Contig	Position in contig	Accession number
fliC	H7	100	1758 / 1758	NC_002695.2 Escherichia coli O157:H7 str. Sakai DNA, complete genome	2624516..2626273	AF228487

Database for O type genes						
Gene	Serotype	Identity	Template / HSP length	Contig	Position in contig	Accession number
wzy	O157	100	1185 / 1185	NC_002695.2 Escherichia coli O157:H7 str. Sakai DNA, complete genome	2785447..2786631	JH953200
wzx	O157	100	1392 / 1392	NC_002695.2 Escherichia coli O157:H7 str. Sakai DNA, complete genome	2783354..2784745	JH959508

extended output

Results as text | Results tsv | Hits in genome seqs | Serotype gene sequences

Selected %ID threshold: *85 %*

Selected minimum length: *60 %*

Input Files: *sequence.fasta*

Support | Scientific problems | Technical problems

Center for Genomic Epidemiology, DTU, Kemitorvet, Building 204, 2800 Kgs. Lyngby, Denmark
Contact: Vibeke Dybdahl Hammer, Telephone: +45 3588 6420, E-mail: vdha@food.dtu.dk
Funded by: The Danish Council for Strategic Research
Last modified May 22, 2012 11:08:01 GMT

https://usegalaxy.eu/

Galaxy Europe

Workflow Visualize Shared Data Help Login or Register

Using 0 bytes

Tools

search tools

Upload Data

Get Data
Send Data
Collection Operations
GENERAL TEXT TOOLS
Text Manipulation
Filter and Sort
Join, Subtract and Group
GENOMIC FILE MANIPULATION
Convert Formats
FASTA/FASTQ
Quality Control
SAM/BAM
BED
VCF/BCF
Nanopore
COMMON GENOMICS TOOLS
Operate on Genomic Intervals
Fetch Sequences / Alignments
GENOMICS ANALYSIS
Annotation
Multiple Alignments
Assembly

COVID-19 Research!

Want to learn the best practices for the analysis of SARS-CoV-2 data using Galaxy? Visit the Galaxy SARS-CoV-2 portal. We mirror **all public SARS-CoV-2 data** from ENA in a Galaxy data library for your convenience. The Galaxy community has created COVID-19 dedicated training materials. Please check our recent activities for more details.

If you need help submitting your data to public archives, like ENA, please get in touch. We will support you in sharing your data.

"Anyone, anywhere in the world should have free, unhindered access to not just my research, but to the research of every great and enquiring mind across the spectrum of human understanding." – Prof. Stephen Hawking

News

Oct 13, 2021
BY-COVID: A new EU project for pandemic preparedness

Oct 12, 2021
UseGalaxy.eu Use Case: cellular specification, differentiation and morphogenesis of the mucociliary epithelium

Oct 11, 2021
UseGalaxy.eu Use Case: microRNAs in heart disease

Oct 8, 2021
A proteomics sample metadata representation for multiomics integration and big data analysis

Oct 6, 2021
New brochure from ELIXIR Germany

Oct 5, 2021
UseGalaxy.eu FTP Server Update

Events

Oct 19, 2021
Analyse avancée de séquences

Oct 20, 2021
SARS-CoV-2 Data Analysis and Monitoring with Galaxy

Oct 20, 2021
International Galaxy Proteomics Meeting Series

Oct 21, 2021
Galaxy Paper Cuts

Oct 22, 2021
How Galaxy Imaging makes cloud-based image analysis possible

Oct 22, 2021
The evolution of a Galaxy project for running hybrid events

Currently Running and Queued Jobs

2 K
1.50 K
1 K
500

OPEN CHAT

History

search datasets

Unnamed history

(empty)

This history is empty. You can **load your own data** or **get data from an external source**

https://usegalaxy-eu.github.io/posts/2021/10/13/bycovid-kickoff/plain.html

Prokka prokaryotic genome annotation in Galaxi Europe

Galaxy Europe

Workflow Visualize Shared Data Help User

Using 13%

Tools

prokka

Upload Data

Show Sections

Prokka Prokaryotic genome annotation

Roary the pangenome pipeline - Quickly generate a core gene alignment from gff3 files

WORKFLOWS

All workflows

Prokka Prokaryotic genome annotation (Galaxy Version 1.14.6+galaxy0)

Favorite Versions Options

Contigs to annotate

No fasta dataset available.

FASTA format

Locus tag prefix

(--locustag)

Locus tag counter increment

1

(--increment)

GFF version

3

(--gffver)

Force GenBank/ENA/DDJB compliance

No

Equivalent to --addgenes --mincontiglen 200 --centre Prokka (or other centre specified below) (--compliant)

Add 'gene' features for each 'CDS' feature (--addgenes)

No

Minimum contig size (--mincontiglen)

200

NCBI needs 200

Sequencing centre ID

(--centre)

Genus name

History

search datasets

0323

(empty)

This history is empty. You can **load your own data** or **get data from an external source**

https://rast.nmpdr.org/

RAST Rapid Annotation using Subsystem Technology version 2.0

The NMPDR, SEED-based, prokaryotic genome annotation service.
For more information about The SEED please visit theSEED.org.

»Tutorials »Help carattoli ••••••••••• login

Info: RAST Access Problems

Click here for instructions on how to resolve several of the most common problems accessing RAST or your RAST data.

Comand-Line API "301 Permanently Moved" Errors

Click here for instructions on how to resolve "301 Permanently Moved" errors when using the RAST batch command-line interface.

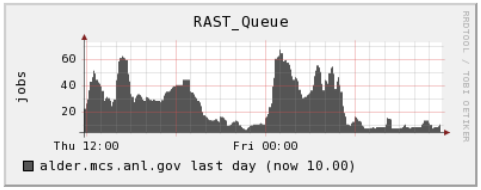
To monitor RAST's load and view other news and statistics for RAST and the SEED, please visit "The Daily SEED."

Welcome to RAST

» Register for a new account, service, or user-group
» Forgot your password?

Login carattoli
Password ••••••••••• Login

RAST Job Load, last 24 hours



What is RAST?

RAST (Rapid Annotation using Subsystem Technology) is a fully-automated service for annotating complete or nearly complete bacterial and archaeal genomes. It provides high quality genome annotations for these genomes across the whole phylogenetic tree.

We have a number of presentations and tutorials available:

- Registering for RAST
- The IRIS/Automated-Assembly/RASTtk Workshop Presentations and Tutorials
- The SEED/"Classic-RAST" Workshop presentations and Tutorials
- Downloading and installing the RASTtk Toolkit
- Downloading and installing the myRAST Toolkit
- The RAST batch submission interface (a part of myRAST)
- Making manual improvements to RAST-annotated genomes (first tutorial). This is a powerpoint presentation; bring it up in slide-show mode and click through to see the animations and movies.
- Making manual improvements to RAST-annotated genomes (second tutorial). This is a second tutorial on the topic of manually improving RAST annotations; it is also a powerpoint presentation with animations.

As the number of more or less complete bacterial and archaeal genome sequences is constantly rising, the need for high quality automated initial annotations is rising with it. In response to numerous requests for a SEED-quality automated annotation service, we provide RAST as a free service to the community. It leverages the data and procedures established within the SEED framework to provide automated high quality gene calling and functional annotation. RAST supports both the automated annotation of high quality genome sequences AND the analysis of draft genomes. The service normally makes the annotated genome available within 12-24 hours of submission.

Please note that while the SEED environment and SEED data structures (most prominently FIGfams) are used to compute the automatic annotations, the data is NOT added into the SEED automatically. Users can however request inclusion of a their genome in the SEED. Once annotation is completed, genomes can be downloaded in a variety of formats or viewed online. The genome annotation provided does include a mapping of genes to subsystems and a

https://youtu.be/4H-L1DVD3z8 MLST video in Galaxy

https://youtu.be/MhIZ6llaAac

NCBI Minute: Use Web RAPT to Assemble and Annotate Prokaryotic Genomes