

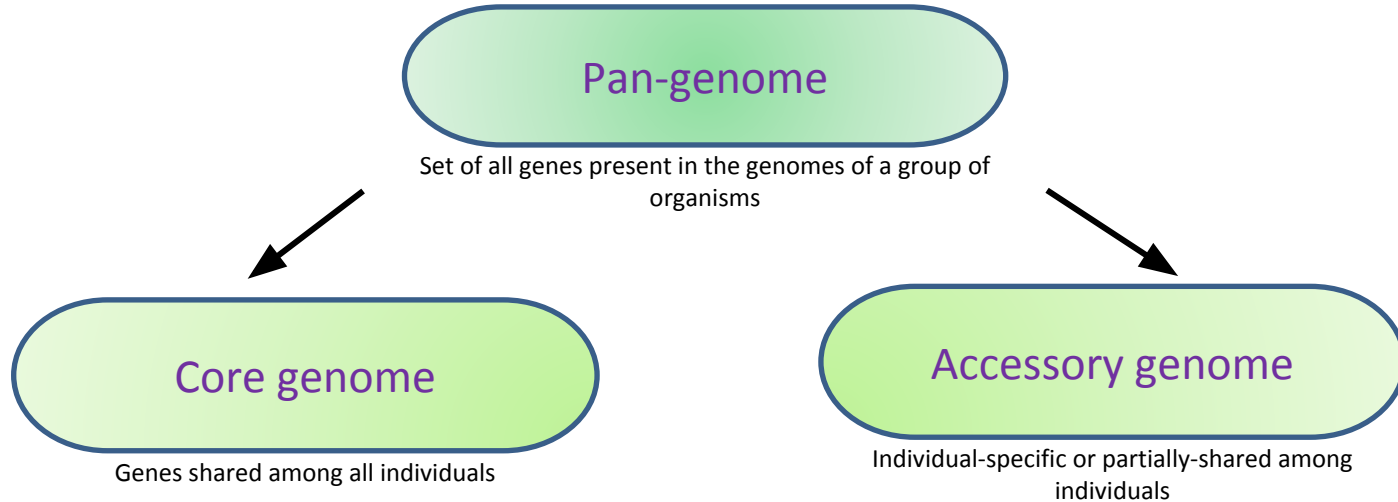
*Precourse: Basics in bioinformatics and
Klebsiella pneumoniae genomics*

Marit Hetland & Kelly Wyres

Session outline

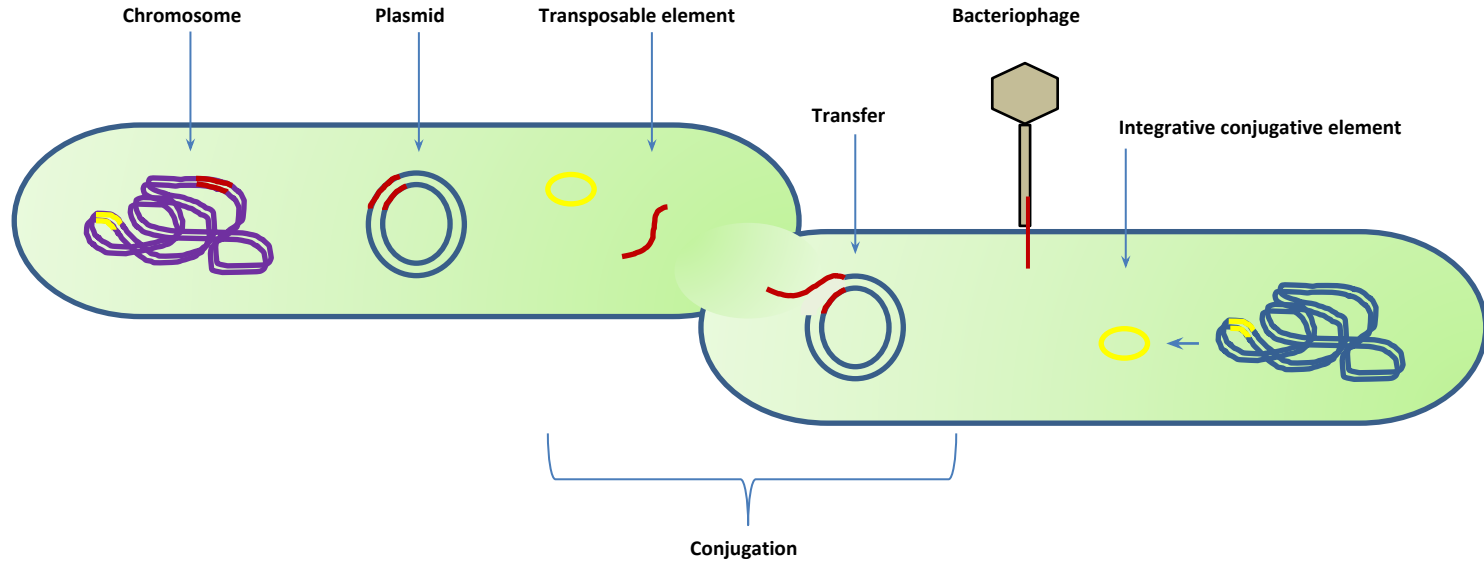
- Bacterial genetics
- Next-Generation Sequencing (NGS)
- Handling of raw data
 - Read mapping/alignment
 - *De novo* assembly
 - Assembly graphs
- General bacterial analyses
- *Klebsiella pneumoniae* genomics
 - What do we want to know?
 - What tools can help us?
- Interactive use of *Klebsiella pneumoniae* genomics tools

Bacterial genetics: The bacterial genome

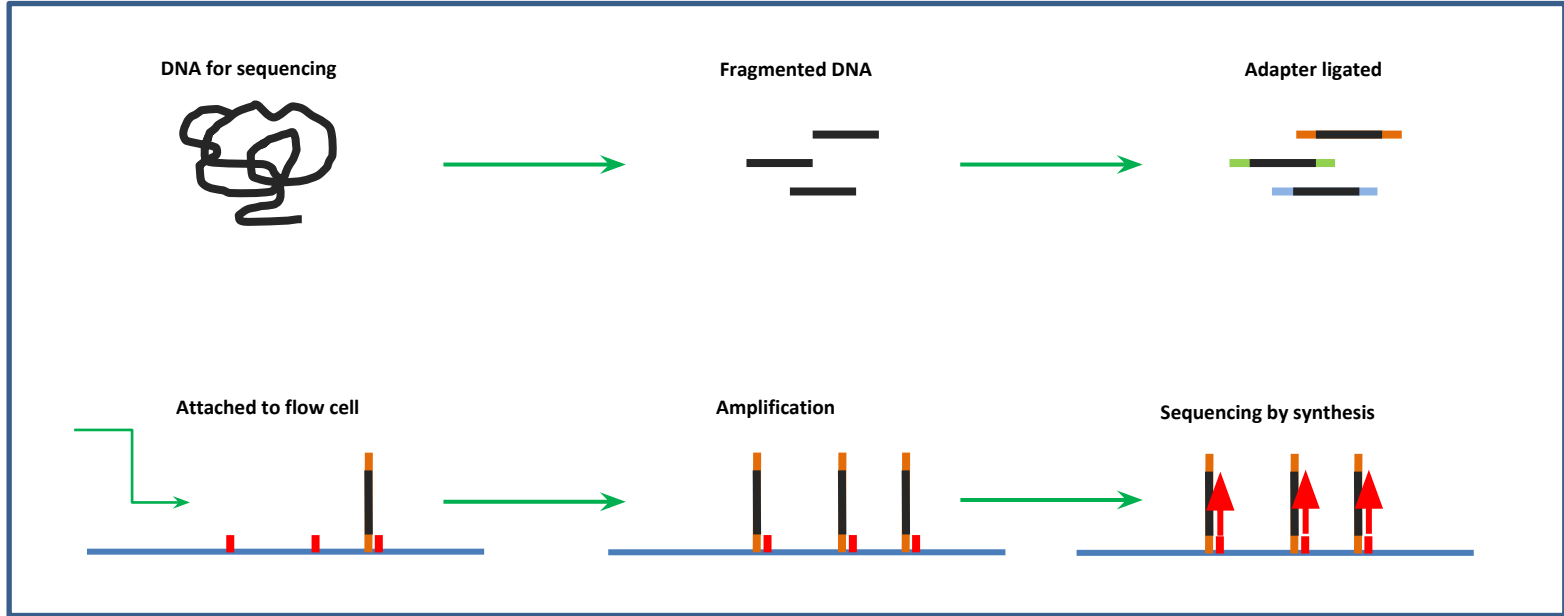


Note: The specific genes in these categories vary based on how many individual isolates are included in your sample

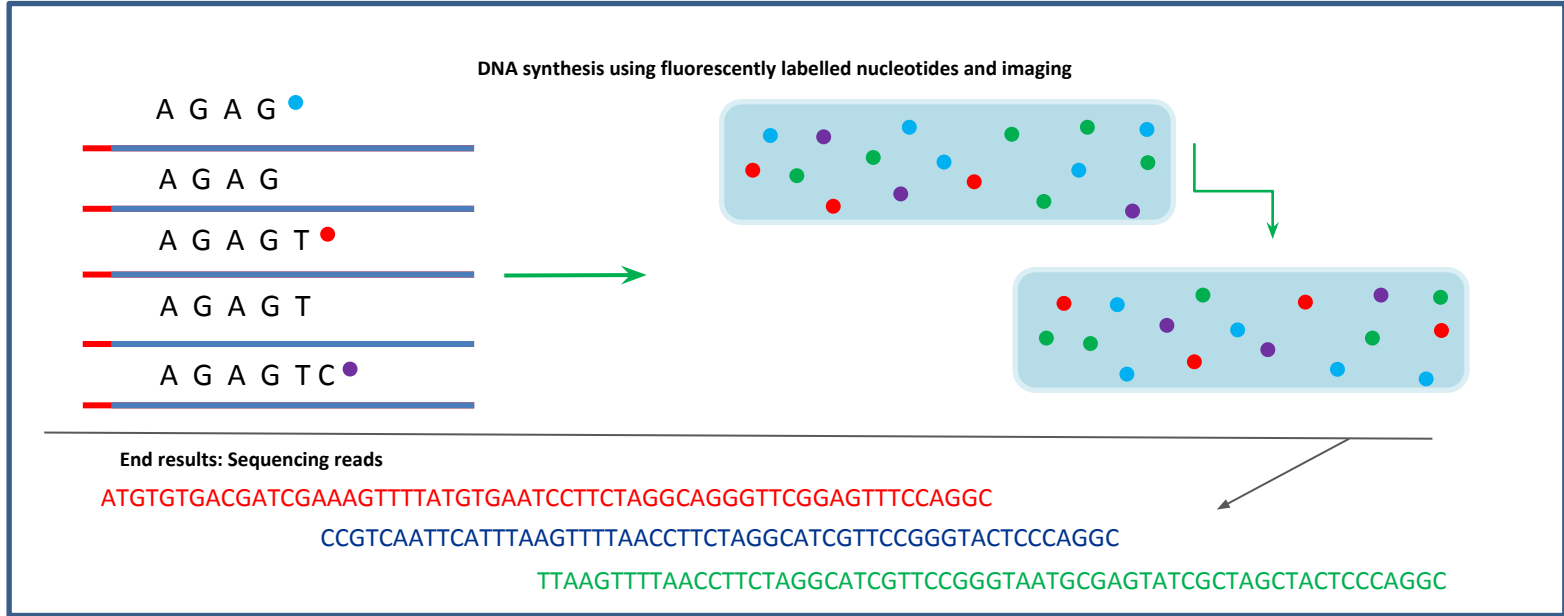
Bacterial genetics: Mobile genetic elements



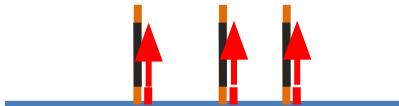
NGS: Short-read Illumina sequencing



NGS: Short-read Illumina sequencing by synthesis



Sequencing by synthesis

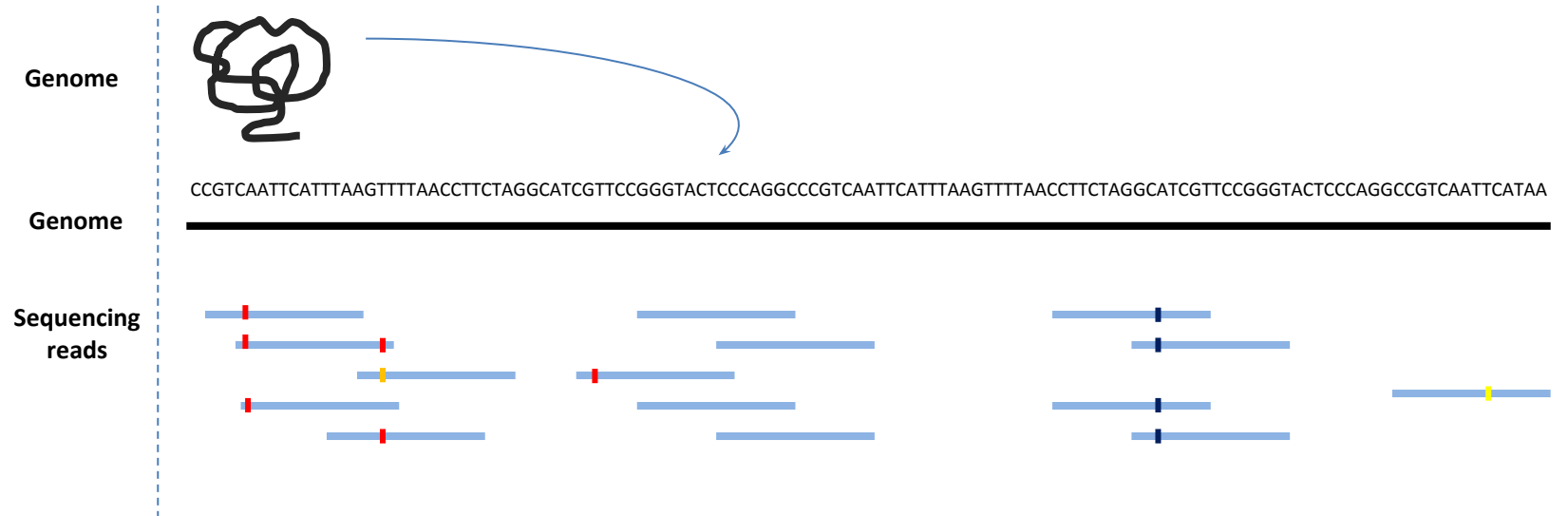


Post-NGS: Raw data

Label	Sequence
@ROFUSJP01AJDW2	CCGTCAATTCATTTAAGTTTTAACCTTCTAGGCATCGTTCCGGGTACTCCCAGGC
+	AAAAAAAAAAAAA::99@::A::??@::FAAAAACAACAAA::BB@?@A?:FA?
Q Scores (ASCII)	Base = T, Q=A=25

?

Post-NGS: Read mapping alignment



Post-NGS: Handling raw data

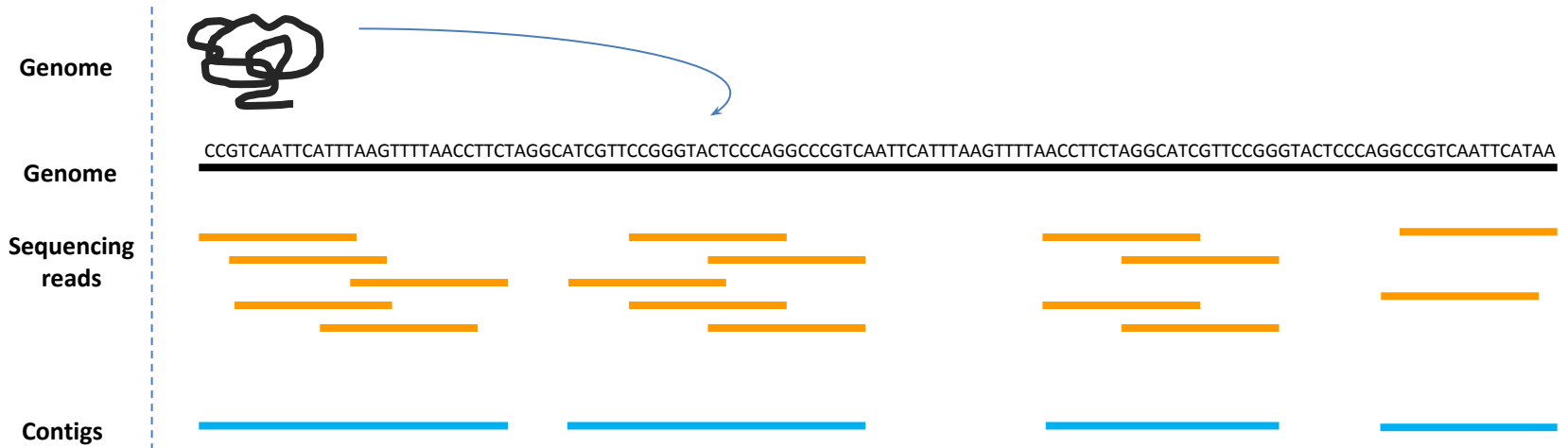
@ROFUSIP01AJDW2

CCGTCAATTCATTTAAGTTTTAACCTTCTAGGCATCGTTCCGGGTACTCCCAGG
+
AAAAAAAAAAAAA::99@::A::??@::FAAAAAACAACAAA::BB@?@A?:FA

Read mapping alignment

```
##bcftools_callCommand=call -c -v -o /home/marit/Tmp_0
#CHROM POS ID REF ALT QUAL FILTER
1 27497 . A C 3.0136 .
1 43031 . T ATT 29.0978 .
1 49784 . G A 6.97853 .
1 53432 . A G 3.5427 .
1 80664 . A C 25.7758 .
1 81800 . C T 4.76877 .
1 85261 . G T 3.5427 .
GT:PL 0/1:31,0,197
1 86102 . A G 60.0087 .
GT:PL 0/1:90,0,32
1 89336 . G T 4.76875 .
```

Post-NGS: *De novo* assembly



Post-NGS: Handling raw data

```
@ROFUSJP01AJDW2
```

```
CCGTCAATTCATTTAAGTTTTAACCTTCTAGGCATCGTTCCGGGTACTCCCAGG
```

```
¢
```

```
AAAAAAAAAAAAAAAA::99@::A::??@::FAAAAAACAACAAA::BB@?@A?:FA  
A?
```

Read mapping alignment

```
##bcftools_callCommand=call -c -v -o /home/marit/Tmp_0  
#CHROM POS ID REF ALT QUAL FILTER  
1 27497 . A C 3.0136 .  
1 43031 . T ATT 29.0978 .  
1 49784 . G A 6.97853 .  
1 53432 . A G 3.5427 .  
1 80664 . A C 25.7758 .  
1 81800 . C T 4.76877 .  
1 85261 . G T 3.5427 .  
GT:PL 0/1:31,0,197  
1 86102 . A G 60.0087 .  
GT:PL 0/1:90,0,32  
1 89336 . G T 4.76875 .
```

De novo assembly

```
>NODE_182_length_251_cov_57.532258  
TTTCATGCTGTACCCAGTTGTAATAACCGCTCCTGGATACACCAAATACCTGACACATCGC  
TTCAATGGGAAATTGTTGTCGCCATTGTCGATTAACGCGTATTTTTAGCGACTCCTGT  
GCAAATCTACAGTTGCTTTTTTAATATATCTCGCTCAAGGCGAGCTTCATTTAACGCC  
TTACGCAGTTGCAGAATTCAGATTCCAGTTCAGCCACCGTGCGGGAACCAGGAGTACCG  
AGCCCTTTTCT  
>NODE_183_length_245_cov_19.364407  
TGGCGCTCGCCAGCGCATGTTTCGTACCTGCTGAAAGAAGAACAACCGGACGAATCGATTC  
GTGCTGAGGTGTTCAAGTACATCCCAAGGCAAAAGCTGGCTGAAATCATCACGTTGTCC  
GTGAAATTGCCCGCCCTCAGACGATAATTTTCATGAAGAAATGGTGGAGCAGTACGGGC  
GCGTTCGTGTTTCTGCCCATCTGCTGAATACCGTTAAATTTTCATCCGACCTGCCG  
GGGTT  
>NODE_184_length_245_cov_16.788136  
GTCAGAATAGTGGACAGCGGCCAAGAACTTCGTTTCATGATAGTCTCCGAACCCGTTTCG  
AGTCGTTTTCCGCCCGTGCTTTCATATCAATTGTCGGGGTTGATCGCAACGTACAACA  
CCTGTGGTACGTATGCCAACACCATCCAACGACACCGCAAAGCCGGCAGTGGGGCAAAA  
TTGCCTCCGCTGGTTACGGGCACAACAACAGGCGAGGCGGGTACGCGATTAAAGGCCGCC  
GGTGT|
```

Post-NGS: *De novo* assembly graphs

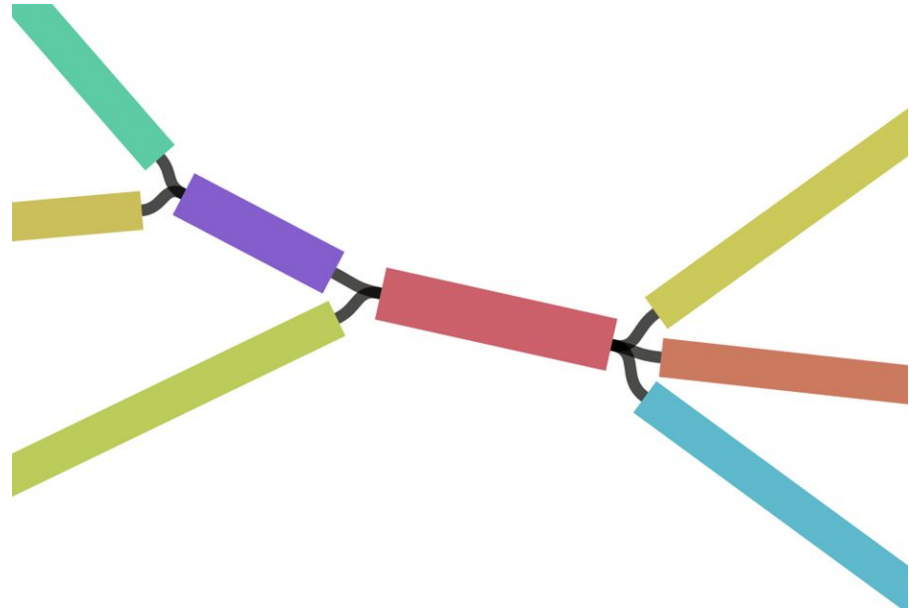
- 1 GACTTGTATTAGATCA
- 2 CATAATTGTCCCCGGTATTATTG
- 3 TAGCTAAAGG
- 4 CGTTAGCTGCTGGTACATCAGCTCC
- 5 CCCGGGA
- 6 GTTACTTCGCTATTCAT
- 7 TTTTCAGAATTTAGC
- 8 CTGAAGATTCA



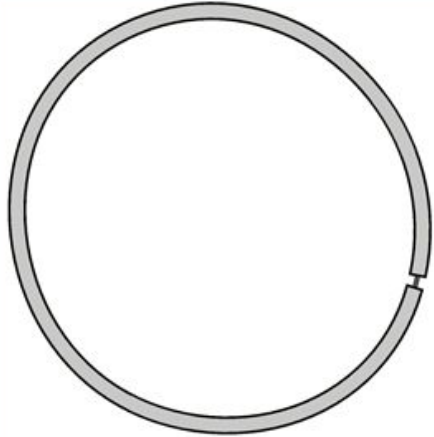
CCGTCAATTCATTTAAGTTTTAACCTTCTAGGCATCGTCCGGGTACTCCCAGGCCGTCAATTCATTTAAGTTTTAACCTTCTAGGCATCGTCCGGGTACTCCCAGGCCGTCAATTCATAA

Post-NGS: De novo assembly graphs

- 1 GACTTGTATTAGATCA
- 2 CATAATTGTCCCCGGTATTATTG
- 3 TAGCTAAAGG
- 4 CGTTAGCTGCTGGTACATCAGCTCC
- 5 CCCGGGA
- 6 GTTACTTCGCTATTCAT
- 7 TTTTCAGAATTTAGC
- 8 CTGAAGATTCA



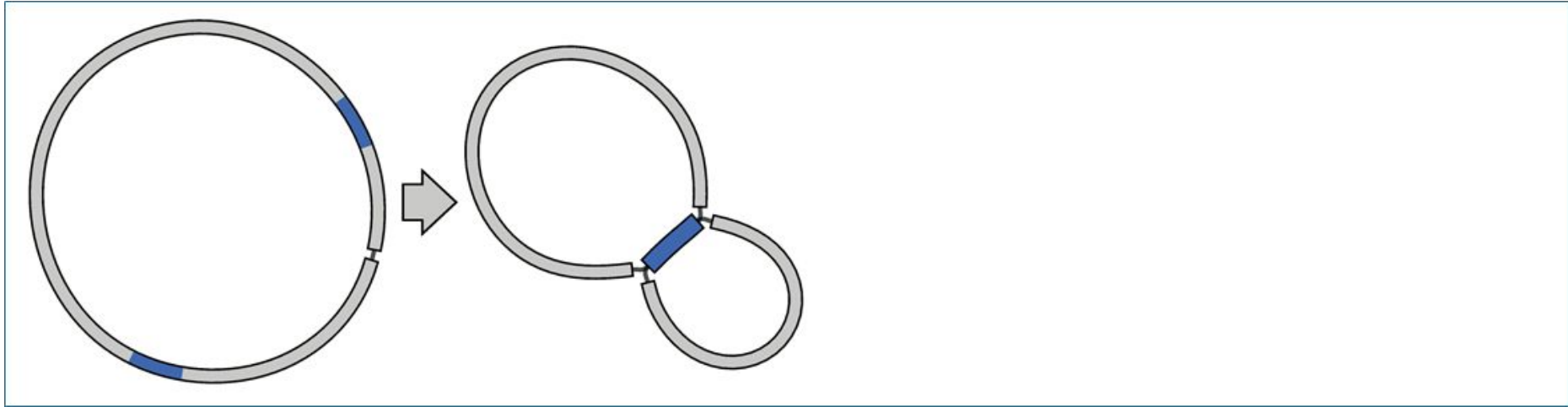
Post-NGS: De novo assembly graphs



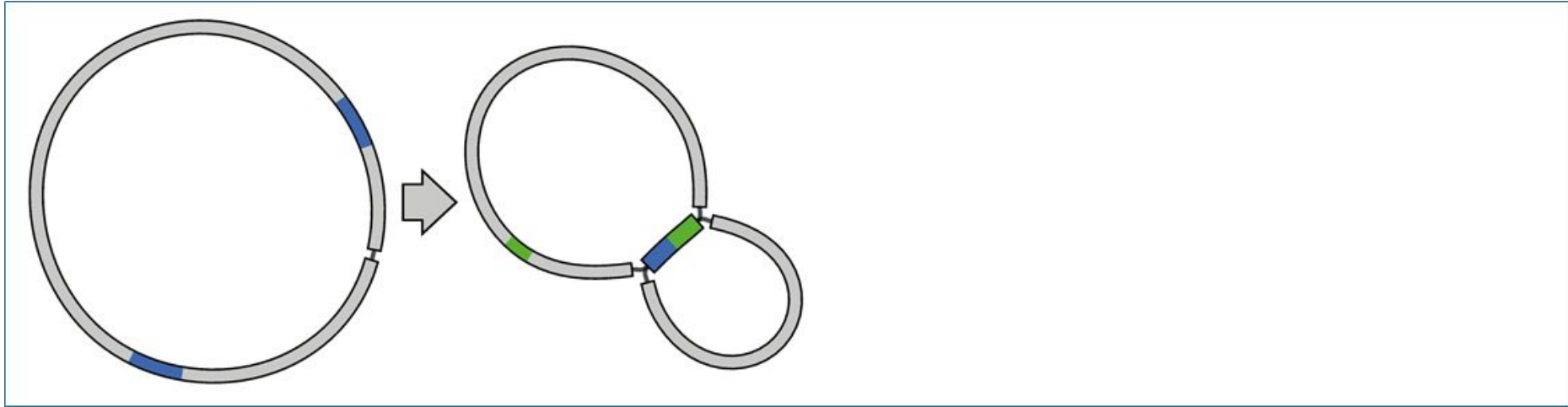
Post-NGS: De novo assembly graphs



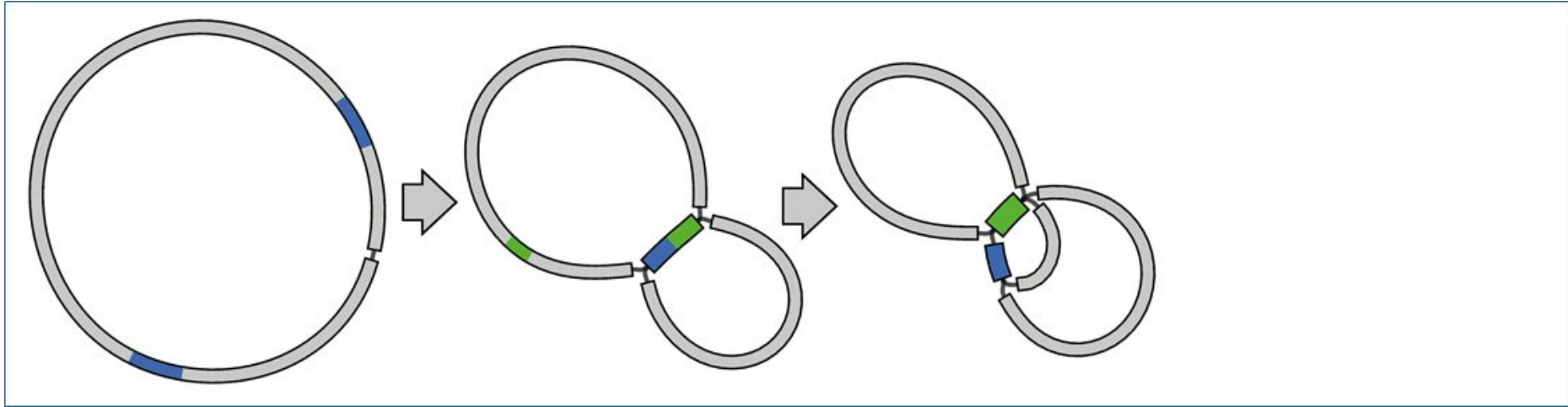
Post-NGS: De novo assembly graphs



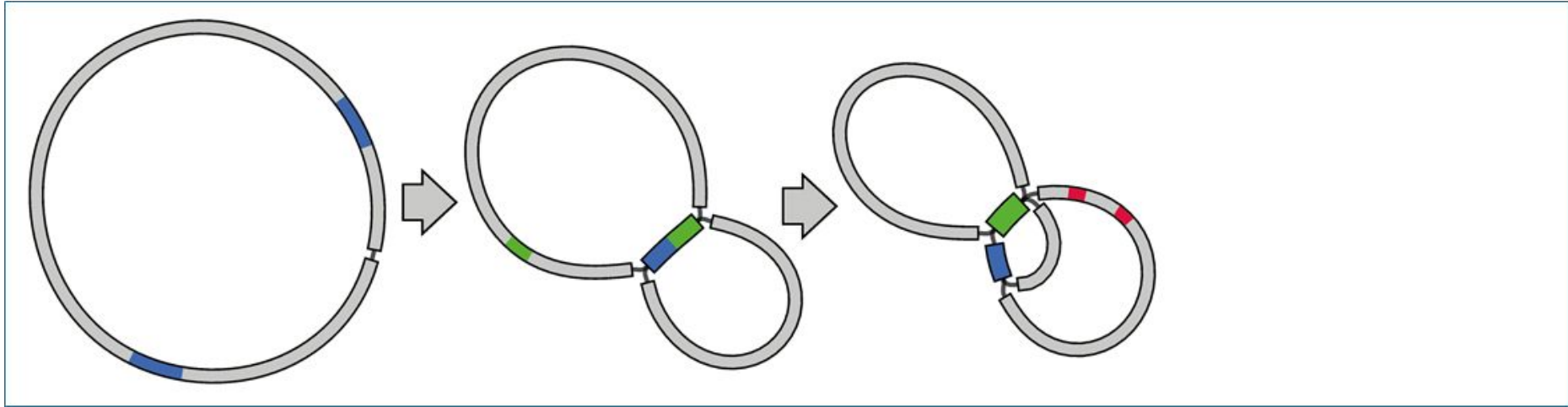
Post-NGS: De novo assembly graphs



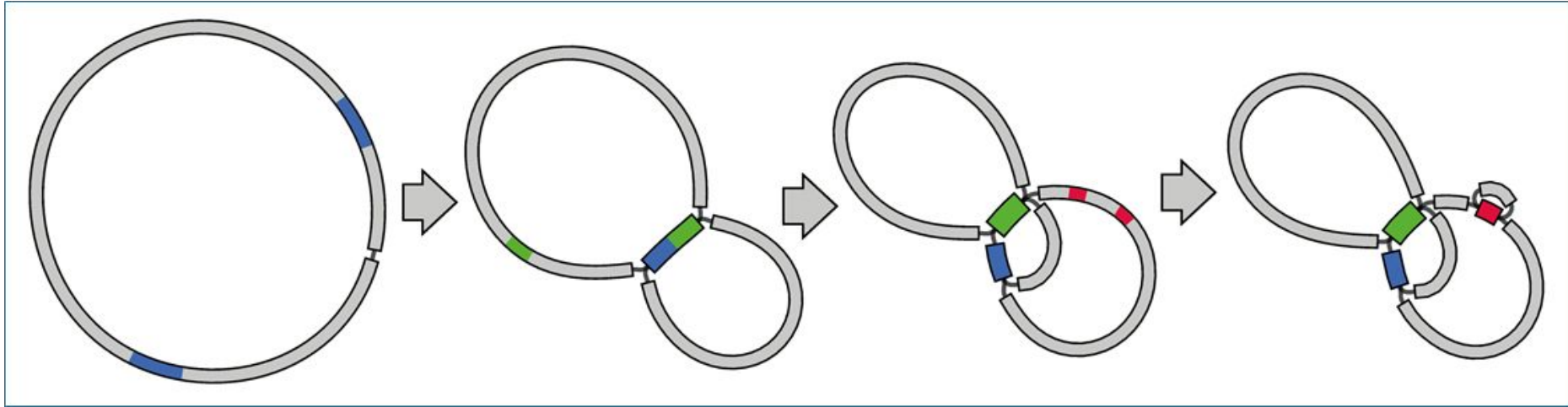
Post-NGS: De novo assembly graphs



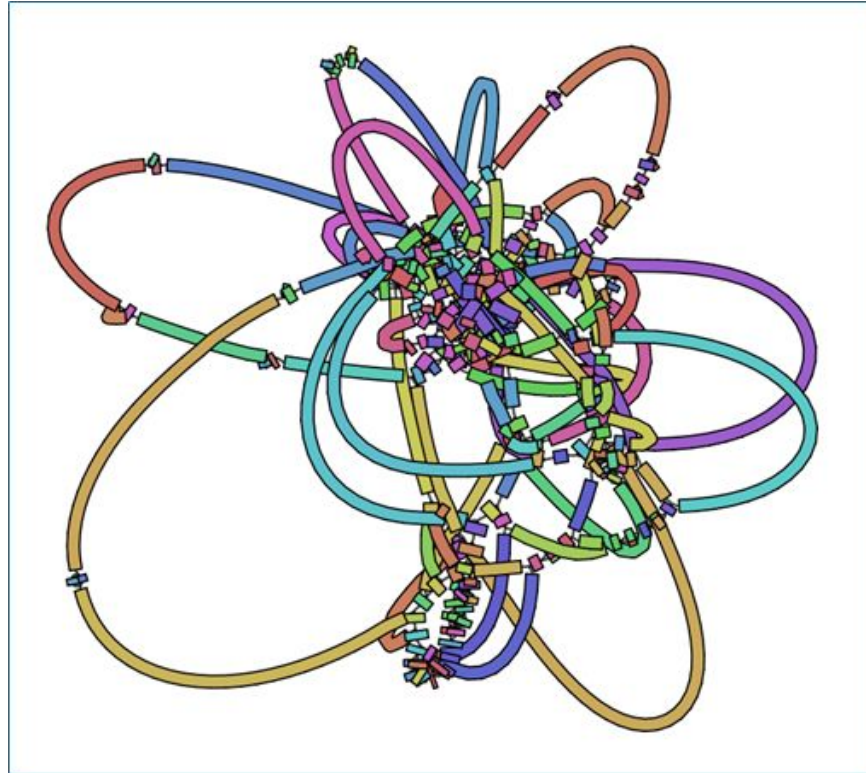
Post-NGS: De novo assembly graphs



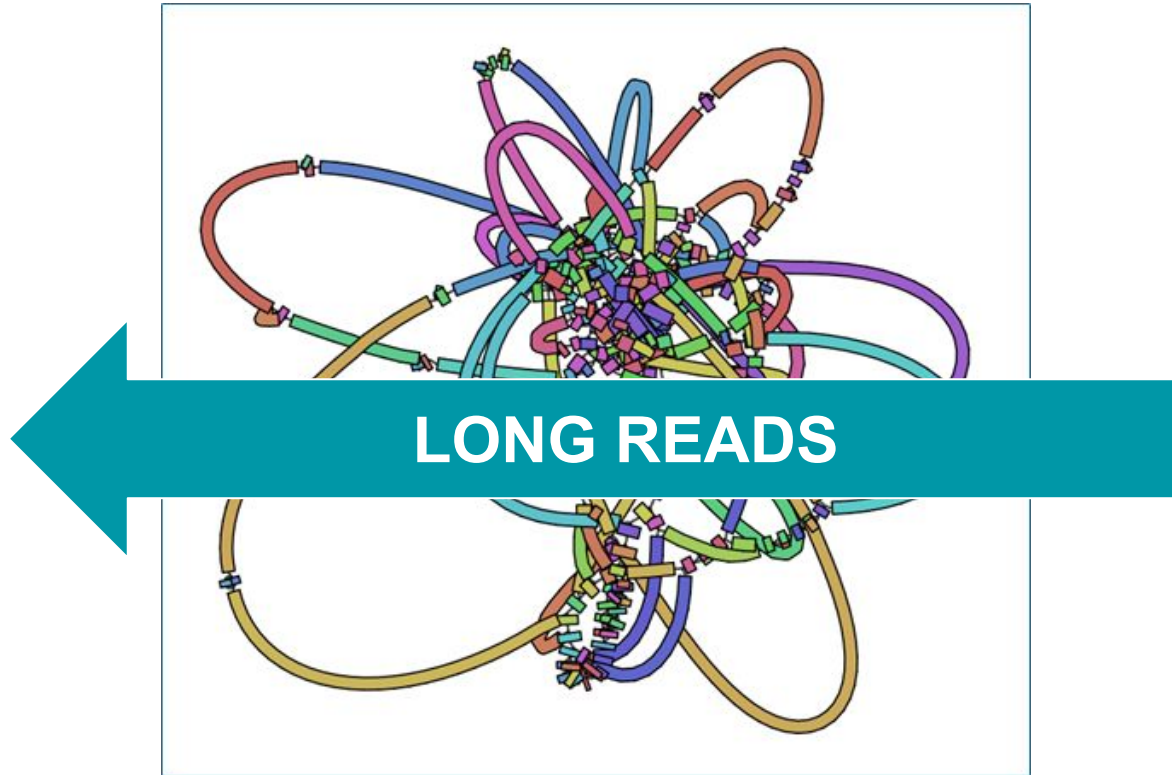
Post-NGS: De novo assembly graphs



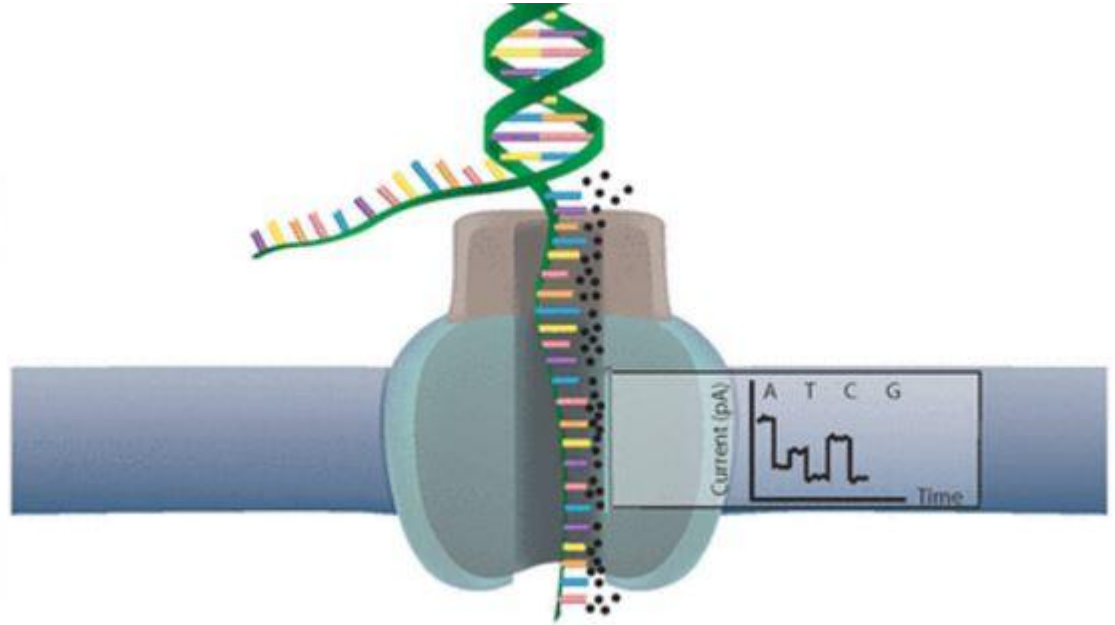
Post-NGS: De novo assembly graphs



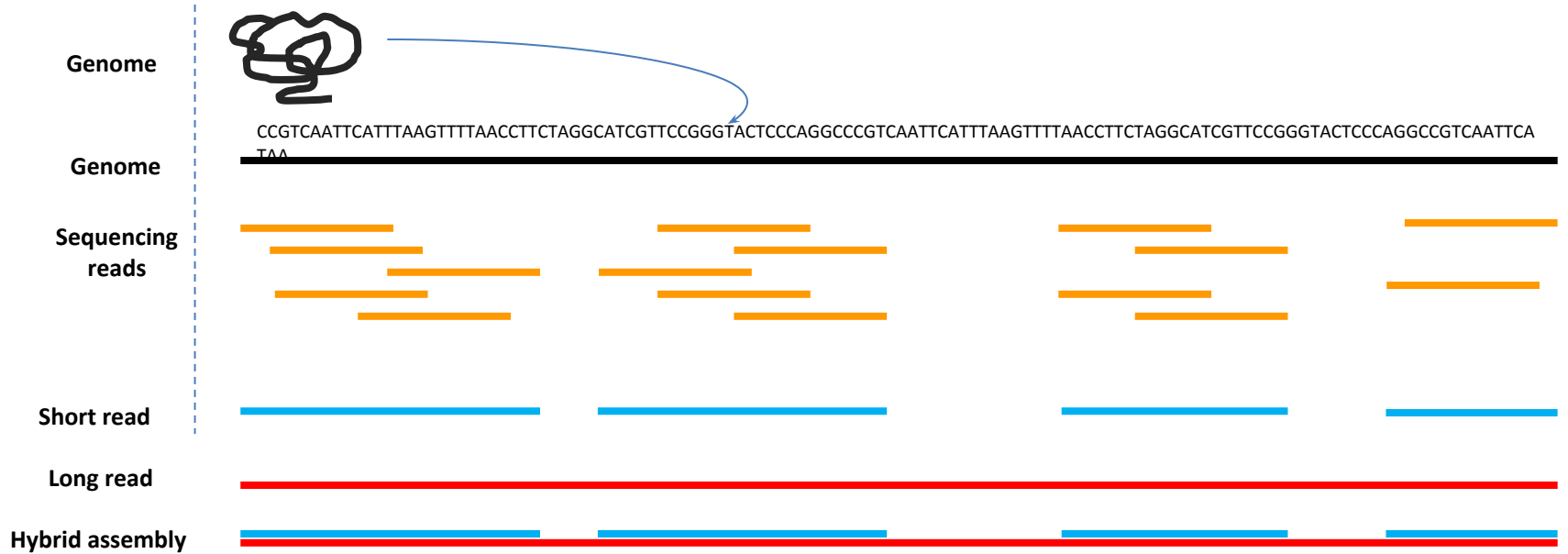
Post-NGS: De novo assembly graphs



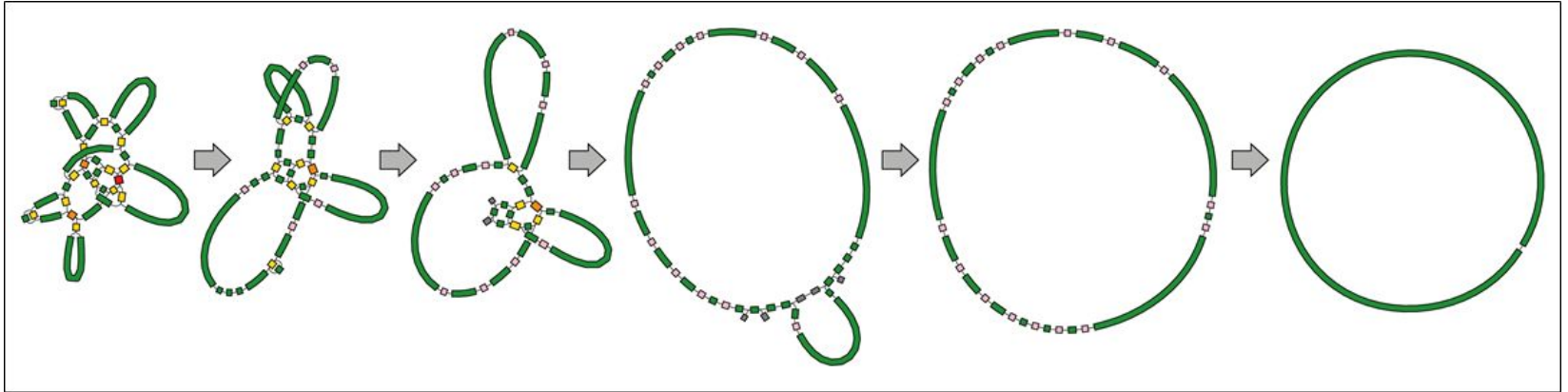
Oxford Nanopore sequencing



Post-NGS: *De novo* assembly - long and short reads



Post-NGS: De novo assembly graphs



Identify genes of interest

MLST

- 7 'housekeeping' genes
- Identify ST from schemes

Several databases for identification of known genes

- AMR: ResFinder, ARG-ANNOT, CARD, NCBI AMR
- Plasmids: PlasmidFinder, Plasmid Atlas
- Virulence factors: Virulence Finder, Virulence Factor

BLAST® is useful for

- Searching for one (or more) particular gene(s) of interest
- Querying a contig sequence from your assembly to identify gene content
- Querying flanking regions of identified gene(s) to determine location

Things to be aware of

- Sequence coverage and identity
- Only partial match - on contig ends?

The Comprehensive Antibiotic Resistance Database

A bioinformatic database of resistance genes, their products and associated phenotypes.

Center for Genomic Epidemiology

ARG-ANNOT.

Antibiotic Resistance Gene-ANNOTation

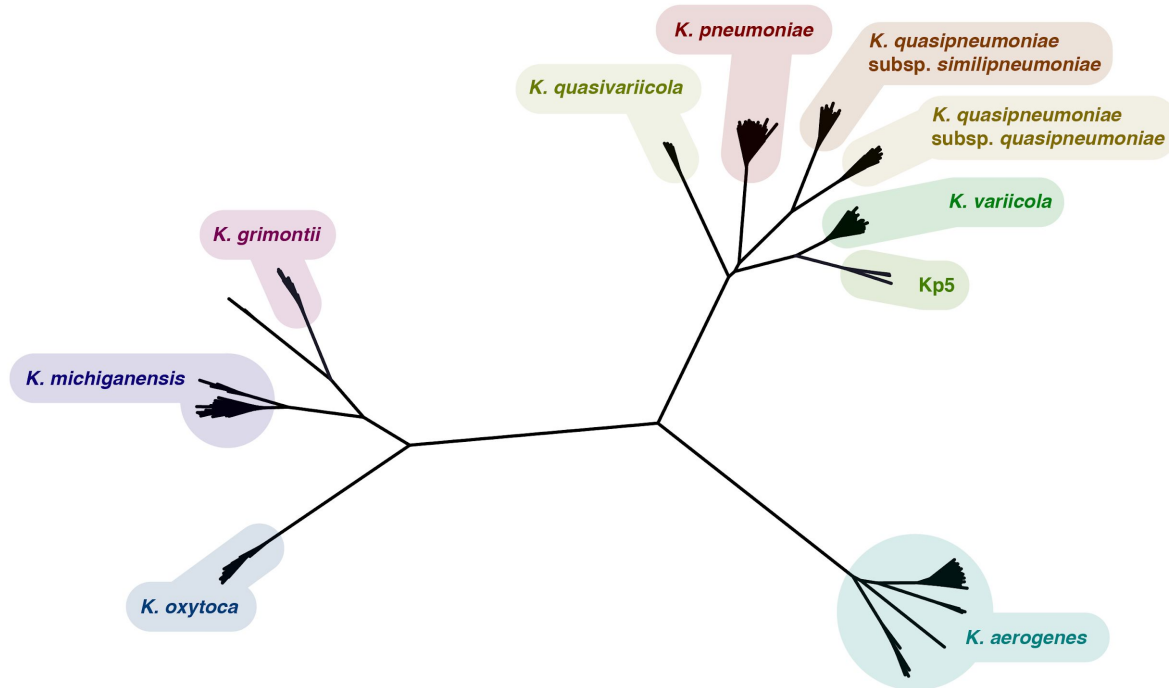


Klebsiella pneumoniae genomics: what do we want to know?

- Species confirmation

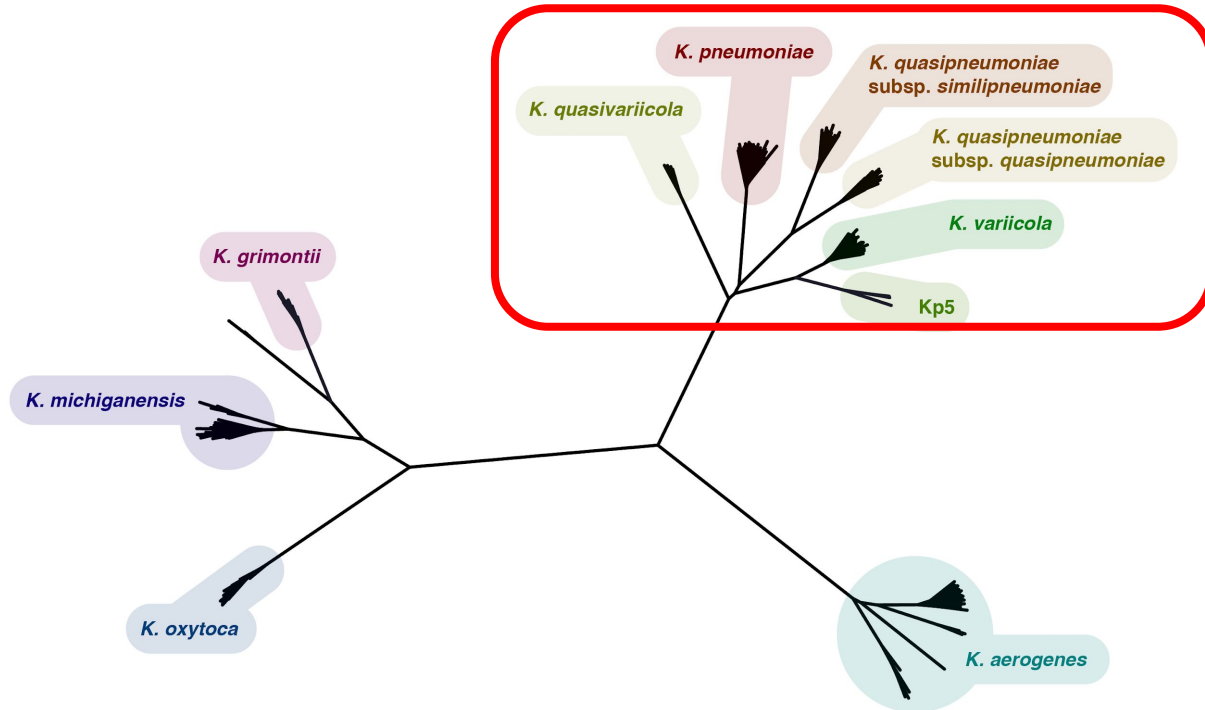
Klebsiella pneumoniae genomics: what do we want to know?

- Species confirmation



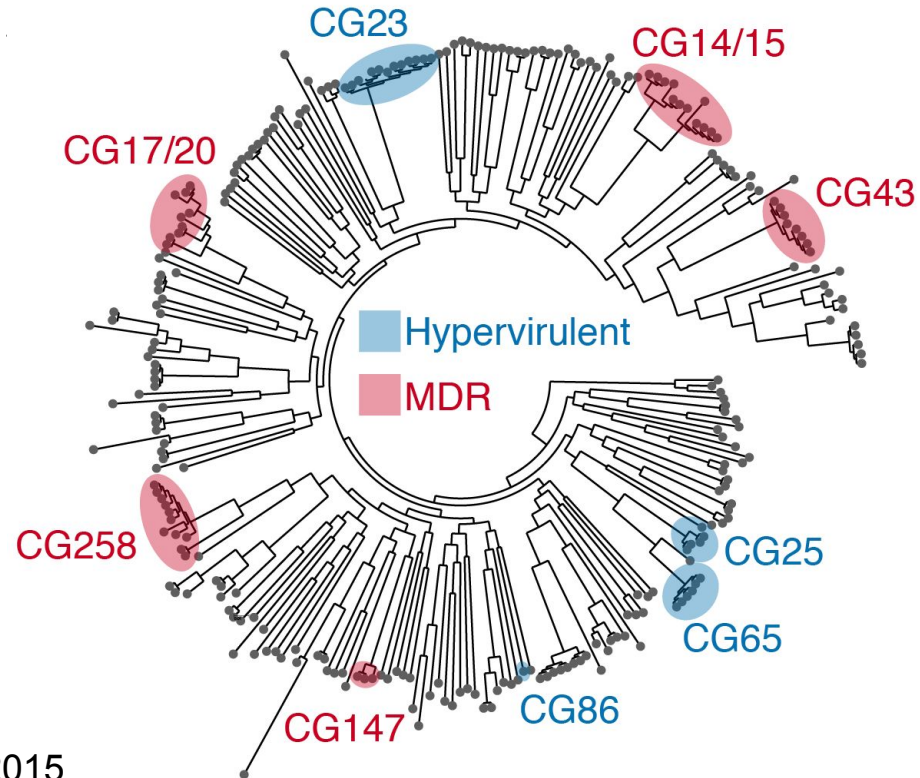
Klebsiella pneumoniae genomics: what do we want to know?

- Species confirmation



Klebsiella pneumoniae genomics: what do we want to know?

- Species confirmation
- Lineages / clonal groups (multi-locus sequence



Klebsiella pneumoniae genomics: what do we want to know?

- Species confirmation
- Lineages / clonal groups (multi-locus sequence types)
- Resistance genes / mutations

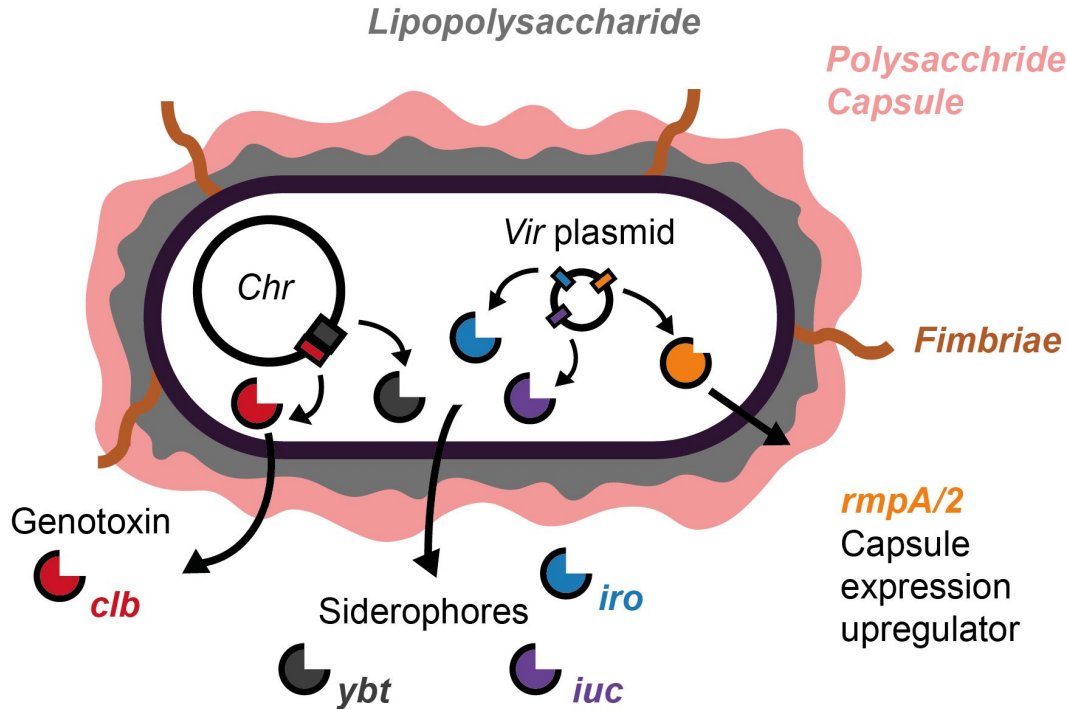
Klebsiella pneumoniae genomics: what do we want to know?

Beta-lactamases		<i>bla</i> Genes conferring resistance (*intrinsic)
Class A		CARB-3, PSE-1, SCO-1, SHV-1*, TEM-1
- ESBL		CTX-M, SHV-5, TEM-10, VEB
- Carbapenemase		KPC, GES-5
Class B (Metallo-beta-lactamase)		CphA, IMP, NDM, SIM, VIM
Class C (Cephalosporinase)		AmpC, CMY, DHA, FOX, MIR
Class D		OXA-1, OXA-2, OXA-7, OXA-9, OXA-10, OXA-12
- ESBL		OXA-11, OXA-15
- Carbapenemase		OXA-48, OXA-51, OXA-181, OXA-237
Other AMR	Genes conferring resistance (*intrinsic)	Mutations
Aminoglycosides	<i>aac, aadA, aadB, aph, armA, rmt, strAB</i>	-
Carbapenems	(see carbapenemase <i>bla</i> genes, class A & D above)	Mutations in <i>ompK35, ompK36</i>
Colistin	<i>mcr-1, mcr1.2</i>	Inactivation of <i>pmrB, mgrB</i> ; mutations in <i>crrB</i>
Fluoroquinolones	<i>qepA, qnrA, qnrB, qnrD, qnrS</i>	SNPs in <i>gyrA, parC</i> ; Upregulation of <i>acrAB</i> or <i>oqxAB</i> efflux
Macrolides	<i>ereA, ereB, ermB, mef, mph, msrE</i>	-
Phenicol	<i>catA, catB, cml, floR</i>	-
Rifampin	<i>arr</i>	-
Sulfonamides	<i>folP, sul1, sul2, sul3</i>	-
Tetracycline	<i>tet</i> genes	-
Tigecycline	-	Upregulation of <i>acrAB</i> or <i>oqxAB</i> efflux; mutation in <i>rpsJ</i>
Trimethoprim	<i>dhfr</i> genes	-

Klebsiella pneumoniae genomics: what do we want to know?

- Species confirmation
- Lineages / clonal groups (multi-locus sequence types)
- Resistance genes / mutations
- Virulence loci

Klebsiella pneumoniae virulence determinants



NOTE:

Core virulence determinants
e.g. *mrk* (type 3 fimbriae), *ent*
(siderophore)

Finding the genes: *Klebsiella pneumoniae* BIGSdb

<http://bigsdb.pasteur.fr/klebsiella/klebsiella.html>

← → ↻ 🏠 bigsdb.pasteur.fr/perl/bigsdb/bigsdb.pl?db=pubmlst_klebsiella_seqdef_public 🔍 ☆ 🇫🇷 🇮🇹 🇵🇸 ⋮

PasteurMLST

Query: [Search](#) | [Browse](#) | [List](#)
Options: [General](#) | [Schemes](#) | [Loci](#) | [Scheme fields](#)
Breakdown: [Isolate fields](#) | [Scheme/alleles](#) | [Publications](#)
Links: [Contents](#) | [Home](#) | [Curate](#)

🔑 Log in

Toggle: ⓘ ☰

Klebsiella locus/sequence definitions database

The *Klebsiella* PasteurMLST sequence definition database contains allele and profile data representing the total known diversity of *Klebsiella* species. Every new ST deposited in this database should have a corresponding record in the isolate database.



Query database

- [Sequence query](#) - query an allele sequence.
- [Batch sequence query](#) - query multiple sequences in FASTA format.
- [Sequence attribute search](#) - find alleles by matching criteria (all loci together)
- [Locus-specific sequence attribute search](#) - select, analyse and download specific alleles.
- [Search, browse or enter list of profiles](#)
- [Search by combinations of alleles](#) - including partial matching.
- [Batch profile query](#) - lookup profiles copied from a spreadsheet.
- [Determine 7-gene MLST sequence type from genome sequence](#) **NEW**



Downloads

- [Allele sequences](#)
-



Option settings

- [Set general options](#)
- [Scheme options](#)



General information

- Number of sequences: 316747
- Number of profiles: [Show](#)
- Last updated: 2018-04-25
- [Profile update history](#)
- [About BIGSdb](#)



Export

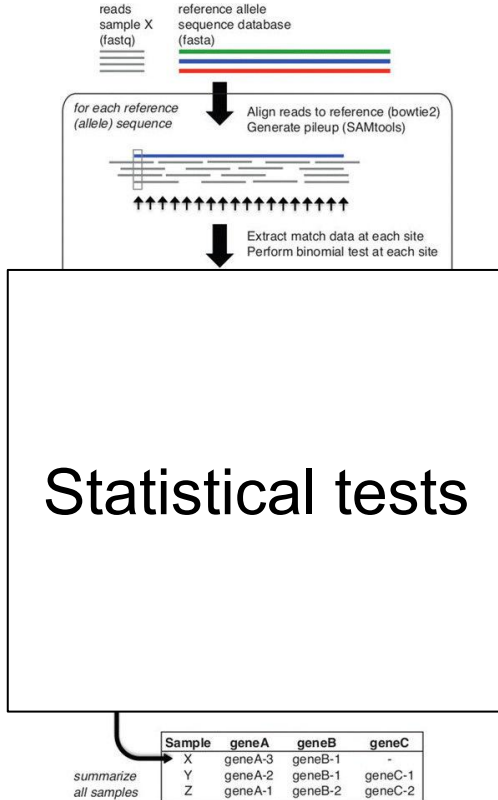
- [Sequences](#) - XMFA / concatenated FASTA formats



Analysis

- [Sequence similarity](#) - find sequences most similar to selected allele.
- [Sequence comparison](#) - display a comparison between two sequences.
- [Locus Explorer](#) - tool for analysing allele sequences stored for particular locus.

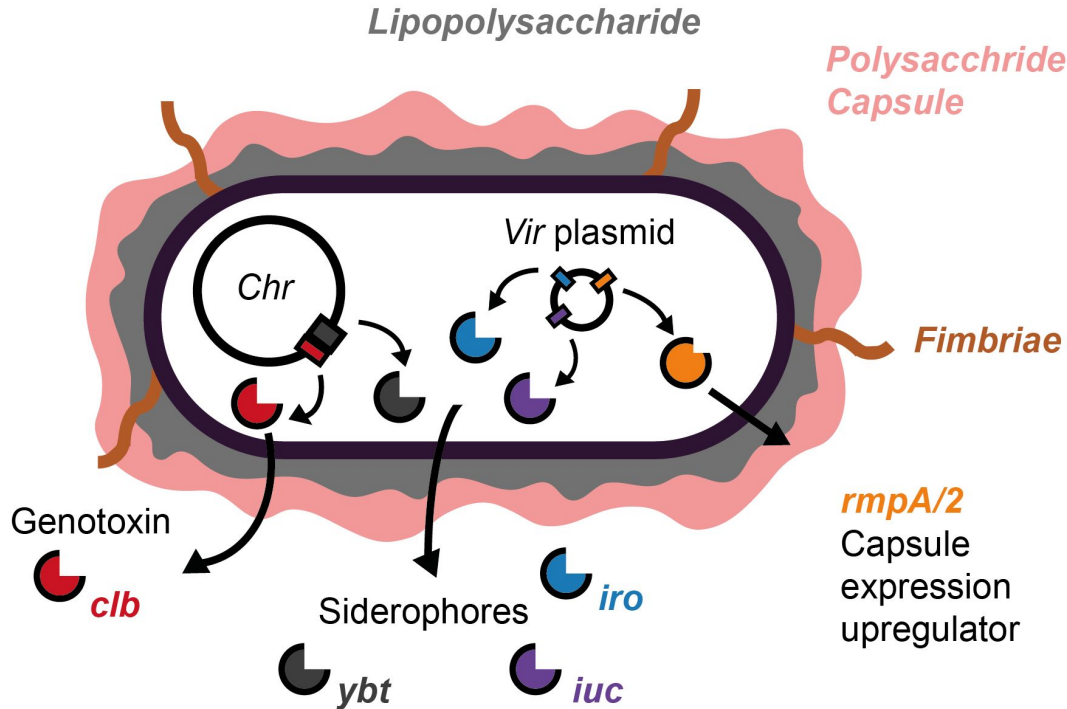
Finding the genes: SRST2



Inouye *et al* Genome Med 2014

<https://github.com/katholt/srst2>

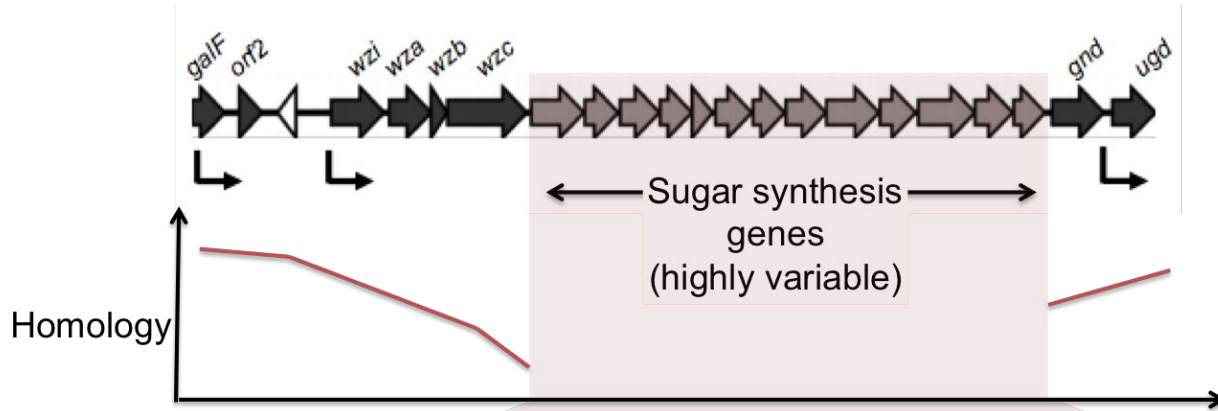
Klebsiella pneumoniae virulence determinants



NOTE:

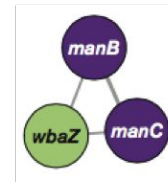
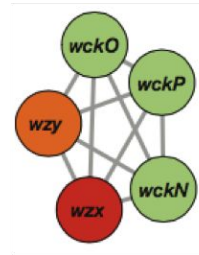
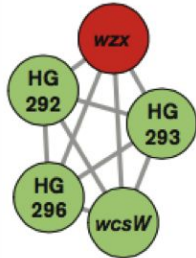
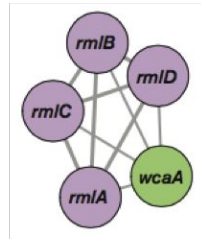
Core virulence determinants
e.g. *mrk* (type 3 fimbriae), *ent*
(siderophore)

Klebsiella pneumoniae capsule (K) loci



Drawn from pool of ~500 proteins

**134 distinct
K loci**

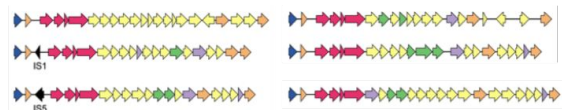


Typing *Klebsiella pneumoniae* capsule (K) loci

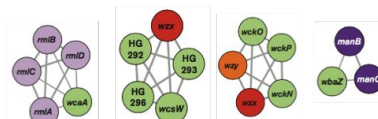


Bioinf tool for K locus typing

<http://kaptive.holtlab.net/>



1. Nucleotide BLAST for all full length loci



2. Protein BLAST for all AA seqs

- Extract best match
- Report any missing or extra genes



Klebsiella pneumoniae genomics: what do we want to know?

- Species confirmation
- Lineages / clonal groups (multi-locus sequence types)
- Resistance genes / mutations
- Virulence loci
- Where are the resistance and virulence genes?

Why do we care about gene location?

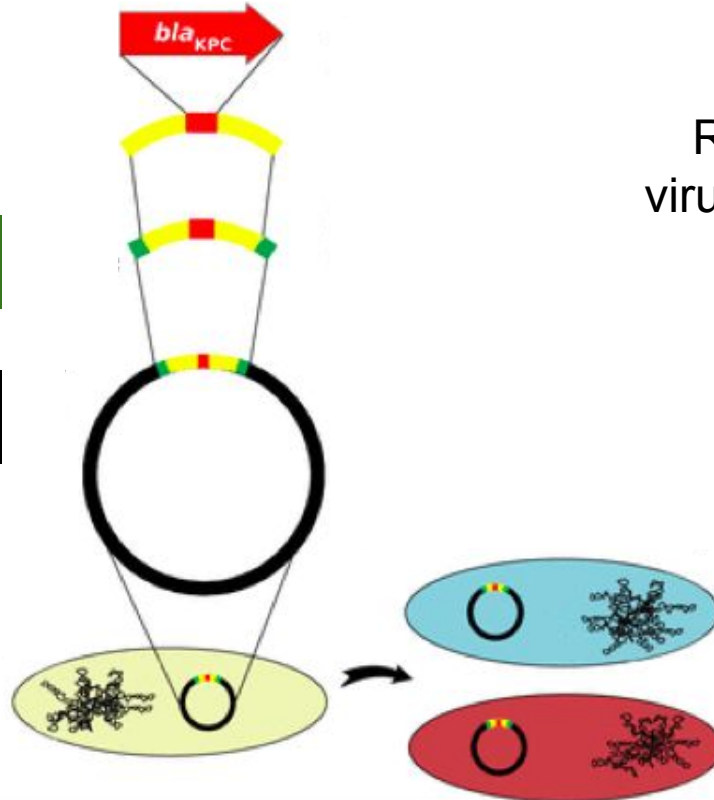
Gene

Transposon

Within a transposon

On a plasmid

Plasmid moves
between cells

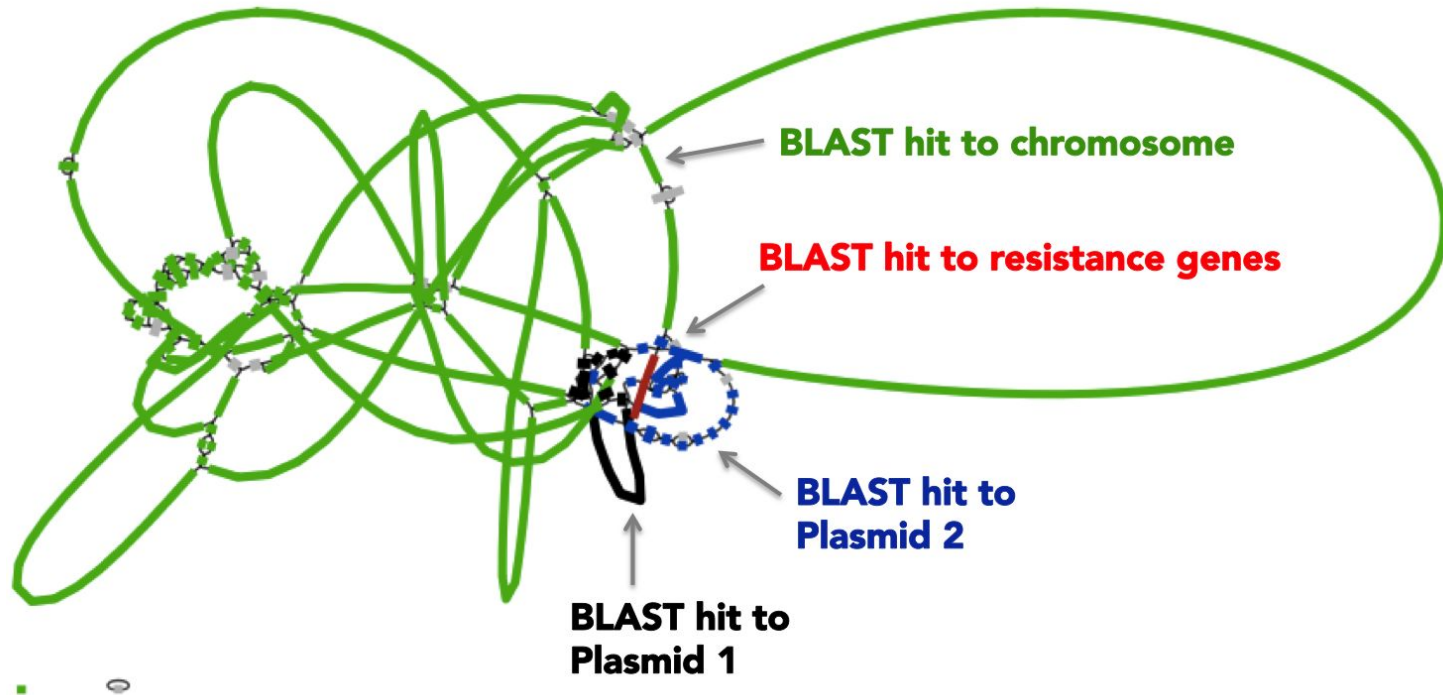


Resistance and
virulence genes can
spread

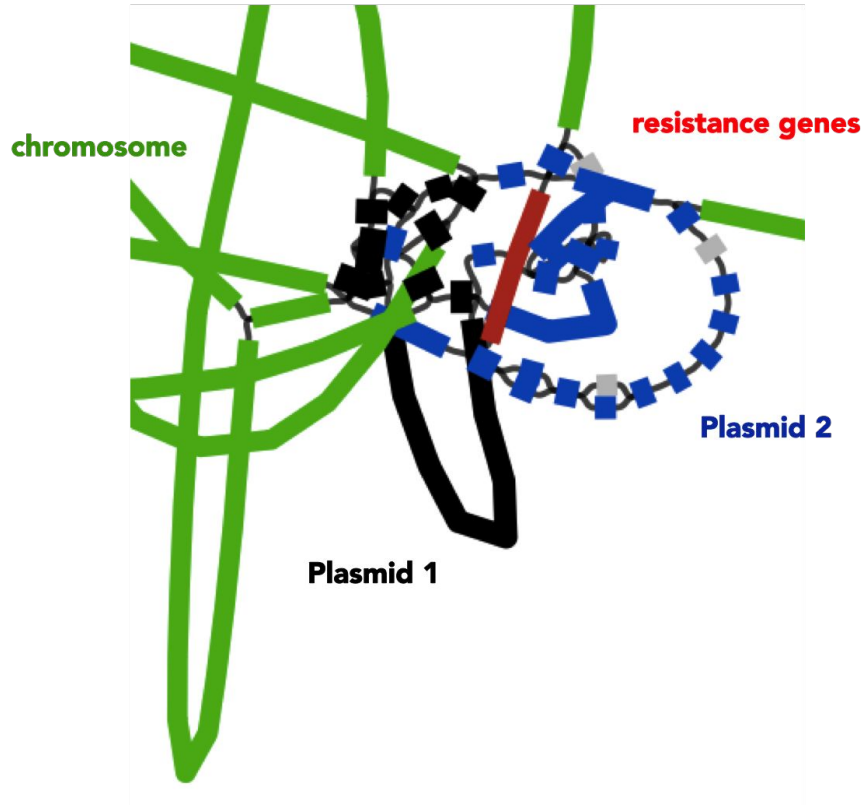
Finding resistance plasmids



Finding resistance plasmids

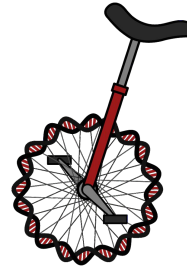


Finding resistance plasmids



Long reads and hybrid assembly can help!

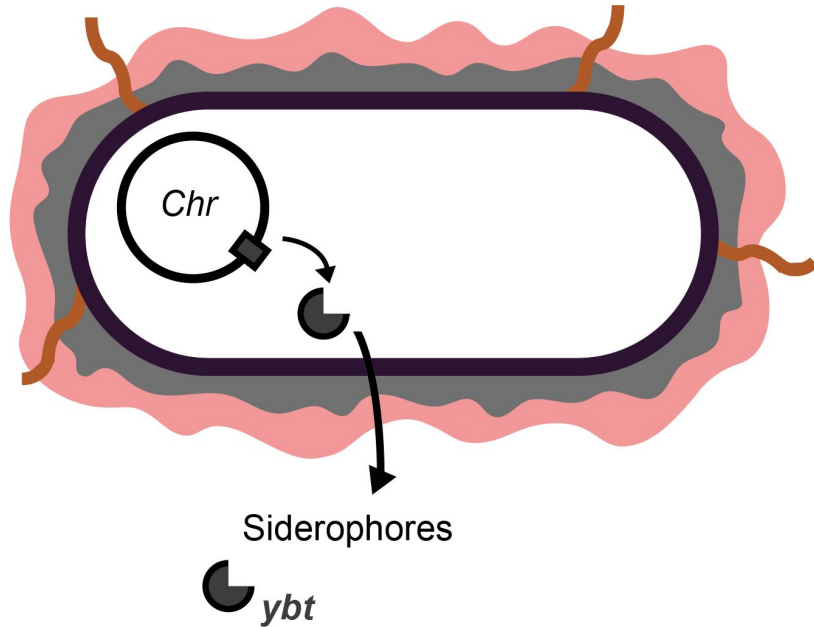
E.g. Wick *et al* MGen 2017 (protocol)
E.g. Wick *et al* PLoS Comp Biol 2017
(Unicycler hybrid assembly)



Unicycler

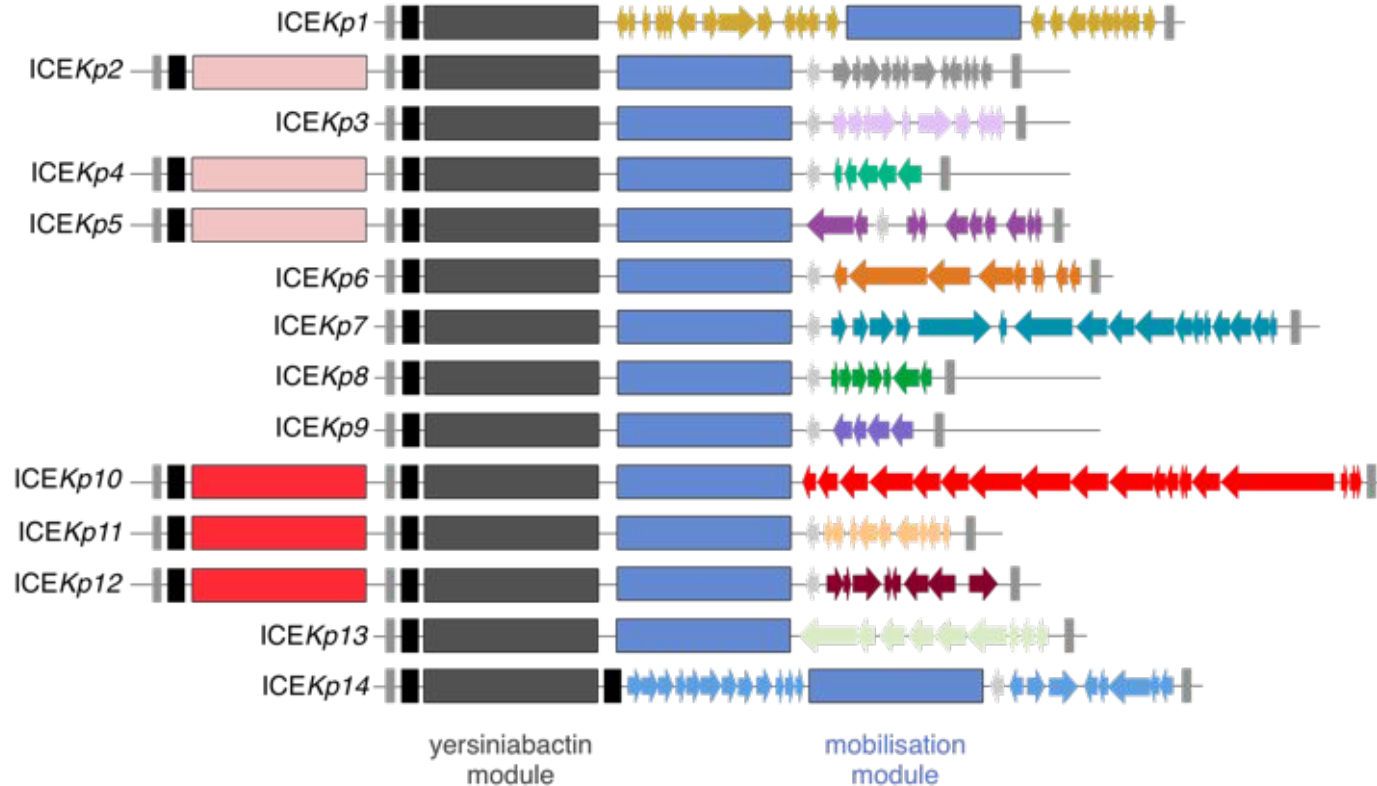
<https://github.com/rswick/Unicycler/>

Klebsiella pneumoniae virulence locus context: yersiniabactin (*ybt*)



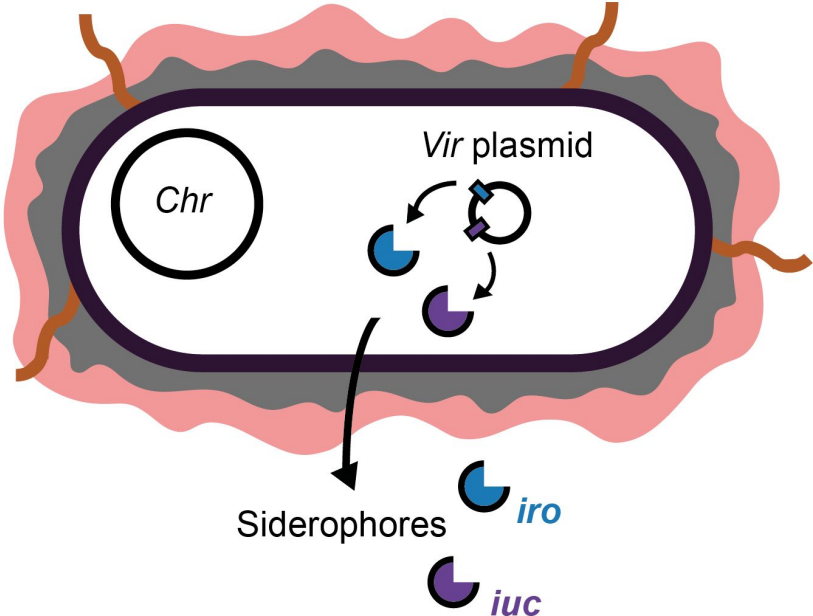
Chromosomally located within an integrative conjugative element (ICE)

Klebsiella pneumoniae virulence locus context: yersiniabactin (*ybt*)



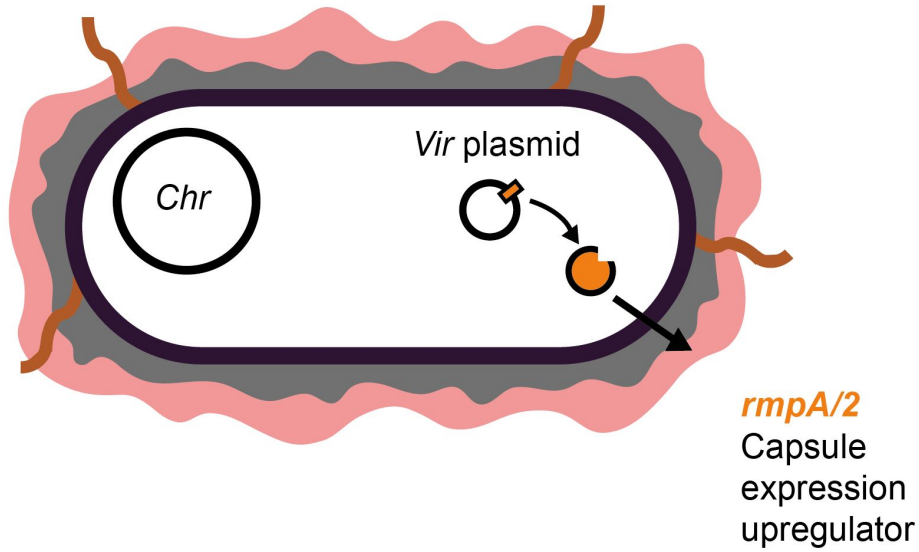
Klebsiella pneumoniae virulence locus context: salmochelin and aerobactin

Co-located on virulence plasmid



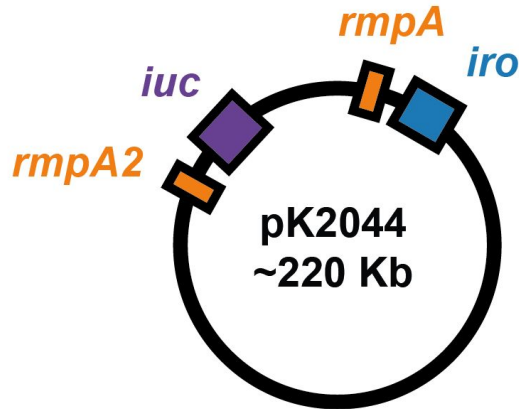
Klebsiella pneumoniae virulence locus context: *rmpA/rmpA2*

Co-located with *iro* and *iuc*

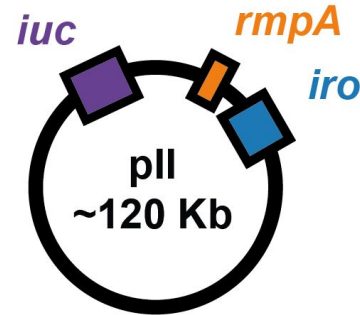


Klebsiella pneumoniae has multiple virulence plasmids

- Two plasmids published to-date; pK2044 most common
- At least 15 virulence plasmids in total (Lam *et al* unpublished)



Strain: NTUH-K2044
(K1, ST23)

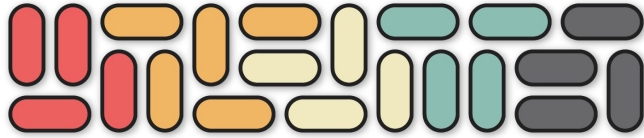


Strain: Kp52.145
(K2, ST66)

Klebsiella pneumoniae genomics: what do we want to know?

- Species confirmation
- Lineages / clonal groups (multi-locus sequence types)
- Resistance genes / mutations
- Virulence loci
- Where are the resistance and virulence genes?

KLEBORATE



- Species confirmation
- Chromosomal MLST
- Resistance genes
- Virulence genes + context

Demos and interactive examples

www.nor-kleb.net/precourse.php