Salmonella serotype identification: SeqSero tool description and application

Rolf Sommer Kaas

Method: SeqSero

- Direct targets or indirect targets
  - MLST: indirect targets
  - SeqSero: direct targets

- Direct targets:
  - *rfb* gene cluster (O)
  - *fliC* and *fliB* (flagellar)

- Finding the targets
  - Raw data: BWA
  - Assembled data: BLAST

Method: SerotypeFinder

- Direct targets or indirect targets
  - MLST: indirect targets
  - SeroTypeFinder: direct targets

- Direct targets:
  - *wzx*, *wzy*, *wzm*, and *wzt* genes (O)
  - *fliC* and a for a few types, non-*fliC* genes (flagellar)
    - Non-*fliC* genes: *flkA*, *fllA*, *flmA*, and *flnA*

- Finding the targets
  - Raw data: Assemble, then BLAST
  - Assembled data: BLAST

Method: SerotypeFinder

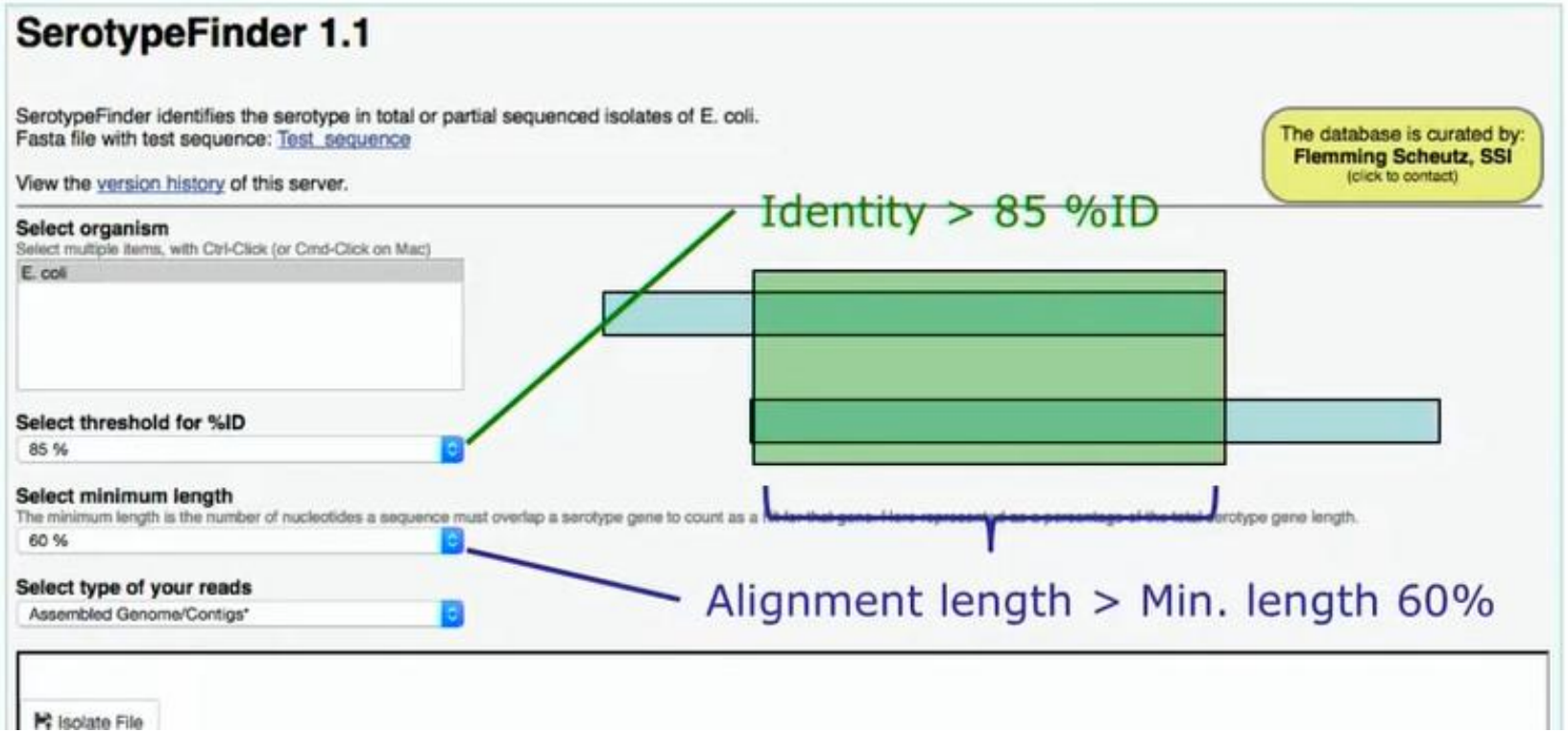
- Curated database
  - Covers all O-types (1-187) except O14 and O57
  - Covers all 53 H-types
- Genotype vs. Phenotype
  - Rough, non-motile, non-typable

Method: SerotypeFinder

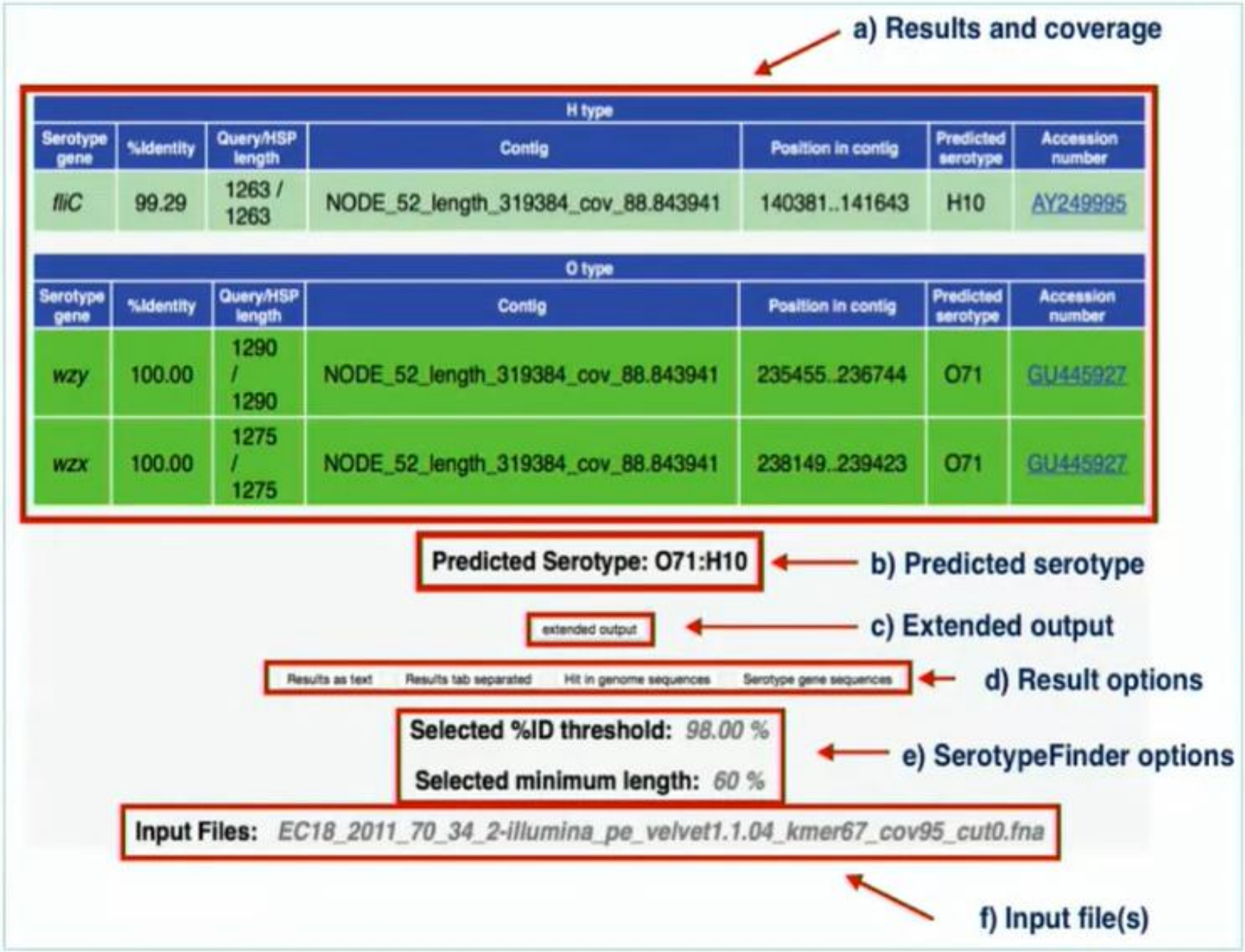
- Curated database
  - Covers all O-types (1-187) except O14 and O57
  - Covers all 53 H-types
- Genotype vs. Phenotype
  - Rough, non-motile, non-typable
- Similar types
  - O90/O127
  - O107/O117
  - O20/O137
  - O13/O135/O129

CGEs SerotypeFinder

https://cge.cbs.dtu.dk/services/SerotypeFinder/



Output


a) Results and coverage
H type
Serotype gene | %Identity | Query/HSP length | Contig | Position in contig | Predicted serotype | Accession number
fliC | 99.29 | 1263 / 1263 | NODE_52_length_319384_cov_88.843941 | 140381..141643 | H10 | AY249995
O type
Serotype gene | %Identity | Query/HSP length | Contig | Position in contig | Predicted serotype | Accession number
wzy | 100.00 | 1290 / 1290 | NODE_52_length_319384_cov_88.843941 | 235455..236744 | O71 | GU445927
wzx | 100.00 | 1275 / 1275 | NODE_52_length_319384_cov_88.843941 | 238149..239423 | O71 | GU445927
Predicted Serotype: O71:H10
b) Predicted serotype
extended output
c) Extended output
Results as text
Results tab separated
Hit in genome sequences
Serotype gene sequences
d) Result options
Selected %ID threshold: 98.00 %
Selected minimum length: 60 %
e) SerotypeFinder options
Input Files: EC18_2011_70_34_2-illumina_pe_velvet1.1.04_kmer67_cov95_cut0.fna
f) Input file(s)

Background

- Plasmids are double-stranded circular or linear DNA molecules. They can replicate and transfer between different bacterial species and clones

- Most of the known plasmids have been identified because they confer phenotypes that are subject to positive selection on bacterial host such as the presence of antimicrobial resistance or virulence genes

- It is important not only to study the molecular epidemiology of different bacterial clones but also to study and understand the molecular epidemiology of transferable plasmids

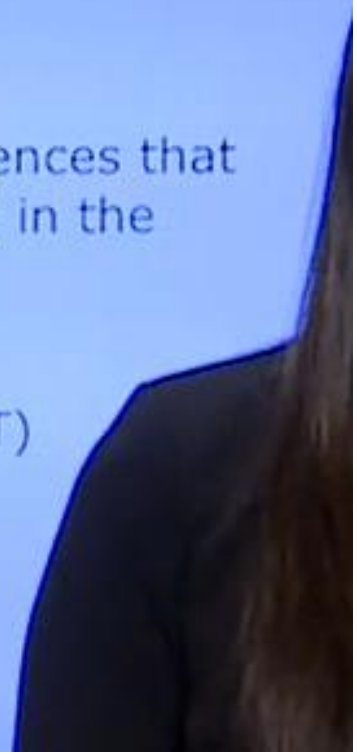
- For this specific purpose, plasmid typing systems are needed

Background

- **PlasmidFinder** is an easy-to-use web tool for in *silico* detection and characterisation of WGS and whole-plasmid sequence data

- The PlasmidFinder database currently consists of 116 replicon sequences that match with at least at 80% identity all replicon sequences identified in the 559 fully sequenced plasmids

- **pMLST** is a web tool for plasmid multilocus sequence typing (pMLST) analysis, a database that is updated weekly was generated from www.pubmlst.org



Background

- Both tools were evaluated using draft genomes from a collection of *Salmonella* Typhimurium isolates

- PlasmidFinder was able to detect a broad variety of plasmids that are often associated with antimicrobial resistance in clinically relevant bacteria pathogens

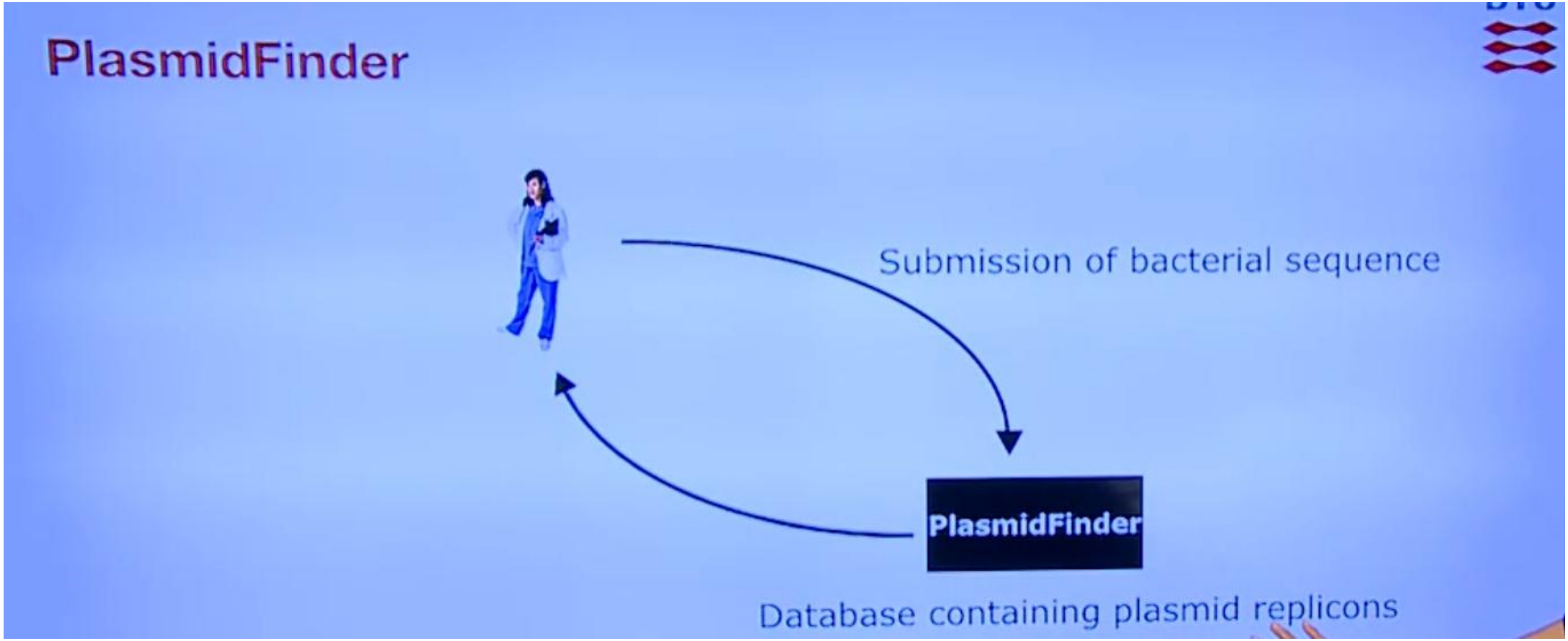
- pMLST tool was able to subtype genomic sequencing data of plasmids, revealing both known plasmid sequence types (STs) and new alleles and ST variants

What does PlasmidFinder do ?

- ○ Identify ST type
- ○ Identity plasmid ST type
- ◉ Identify plasmid replicons
- ○ Identify whole plasmid sequences

✓ **Correct**

PlasmidFinder


PlasmidFinder
Submission of bacterial sequence
PlasmidFinder
Database containing plasmid replicons

PlasmidFinder-1.3 Server - Results

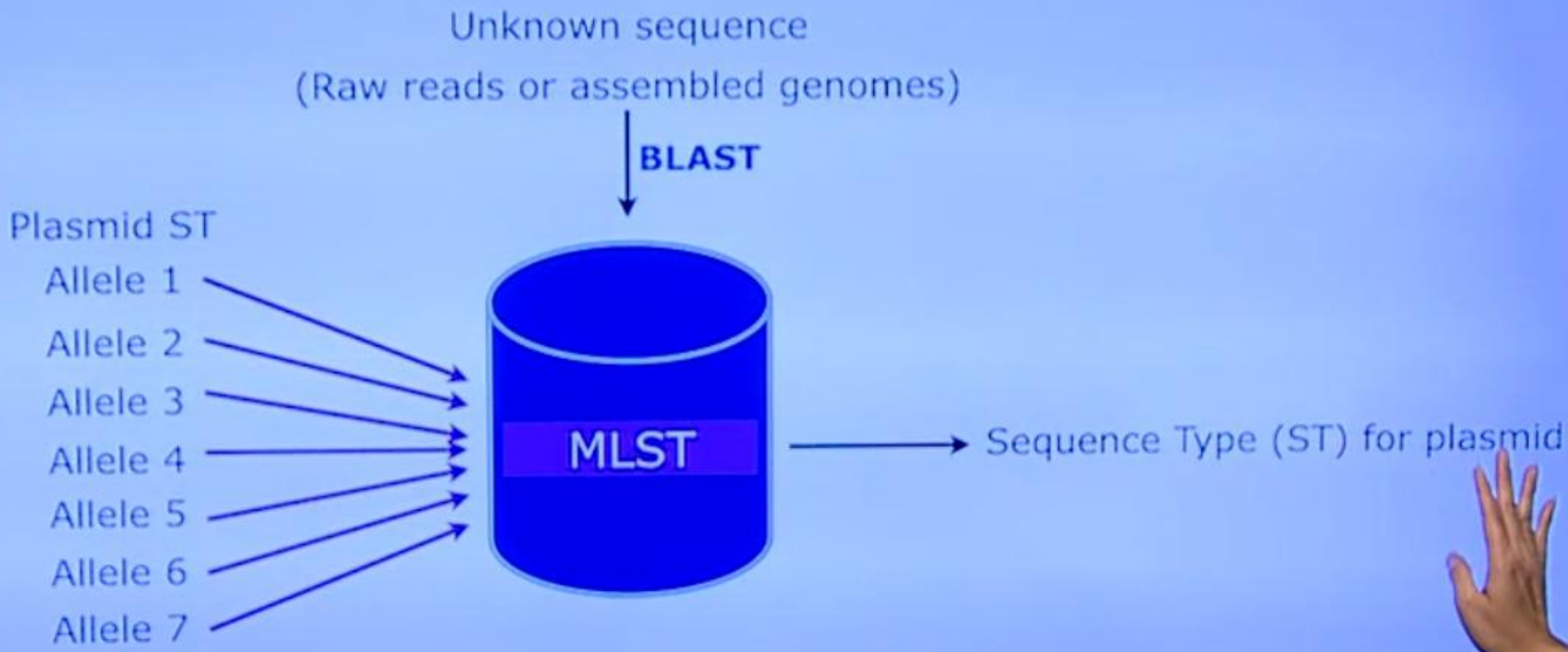
PlasmidFinder Results

SETTINGS:
Selected %ID threshold: *95.00*

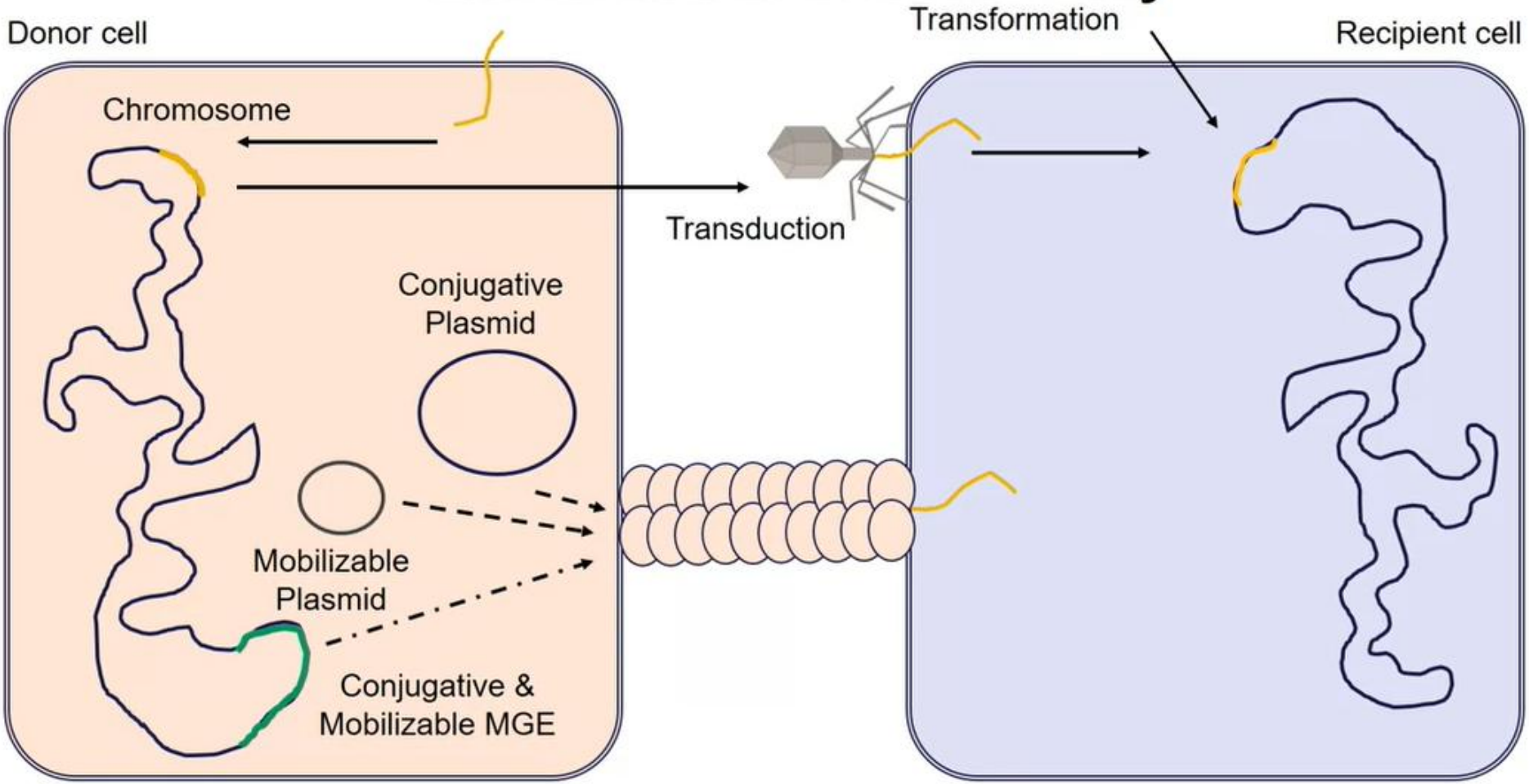
HSP is the length of the alignment between the best matching allele and the corresponding sequence in the genome

Plasmid	%Identity	Query/HSP length	Contig	Position in contig	Note	Accession number
IncP	99.44	535 / 534	strain_1_contig_11	13583..14117	alpha	L27758

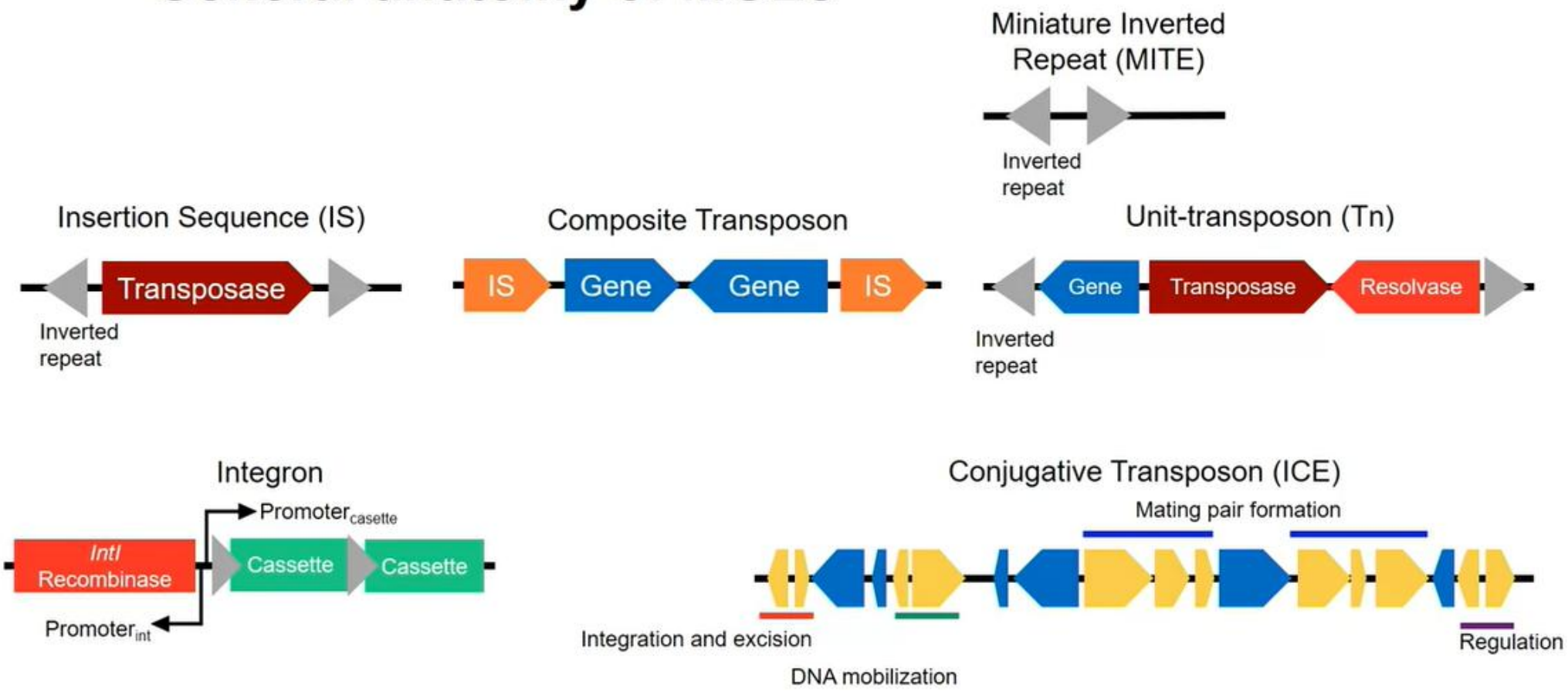
pMLST tool


Unknown sequence
(Raw reads or assembled genomes)
BLAST
Plasmid ST
Allele 1
Allele 2
Allele 3
Allele 4
Allele 5
Allele 6
Allele 7
MLST
Sequence Type (ST) for plasmid

Extracellular DNA mobility



General anatomy of MGEs



Overview of Mobile Genetic Elements

Transposition

Within cell	**Modulating** • MITEs • Insertion Sequences	Gene inactivation Up-regulation & repression of gene expression
	Recruiting • Integrons	
	Gene transporter • Unit transposons • Composite transposons	Can carry passenger genes
Between cell	**Conjugating** • Cis-Mobilizable Elements (CIME) • Integrative Mobilizable Elements (IME) • Integrative Conjugative Elements (ICE) • Plasmids	Transpose genes between bacteria

Which mobile elements are capable of transporting within a bacterial cell AND carrying accessory genes?

- ○ Insertion sequences and Integrons
- ◉ Unit transposons and Composite transposons
- ○ Unit transposons and Integrative Conjugative Elements (ICE)

The function of MGEs

- Recruits and mobilizes DNA through interplay of MGEs and plasmids
- Disseminate genes across bacterial populations, eg antimicrobial resistance genes
- Important for bacterial evolution
  - Rearrangements and deletions of the bacterial genome
  - Recruit and disseminate new genes
  - Harbors genes at low fitness cost

With great power comes great needs for regulation

- MGEs are tightly regulated but modulated by several factors
- Regulation generally coupled to the SOS response systems
- Repression is lifted during stress and/ or presence of ssDNA, result of conjugation.
- Transcription of MGE associated genes can be modulated by host factors, e.g. increased transcription of virulence factors during infection.

How are MGEs often regulated by the bacteria?

- ○ They are not regulated
- ◉ Regulation is coupled to the SOS response systems
- ○ Regulation is coupled to the citric acid cycle

Question

How are MGEs being predicted?

- (●) By sequence similarity to known MGEs
- ( ) By using deep learning neural networks
- ( ) By looking for SNPs in core genes