



Public Health
England

The implementation and application of Whole Genome Sequencing in the *Campylobacter* Reference Laboratory at Public Health England

Craig Swift

Campylobacter EURL workshop (2018)



The Gastrointestinal Bacteria Reference Unit

- The Gastrointestinal Bacteria Reference Unit (GBRU) at Public Health England routinely detects and characterizes foodborne pathogens involved in outbreaks of gastroenteritis including:
 - *Campylobacter*
 - *Salmonella*
 - *Listeria monocytogenes*
 - pathogenic *E. coli*
- GBRU are currently using Whole Genome Sequencing (WGS) to characterize these pathogens and are replacing many of the conventional microbiological techniques used to identify and type bacterial pathogens.
 - High throughput, highly automatable and cost effective.
 - Validation: parallel testing was undertaken with existing methodology.



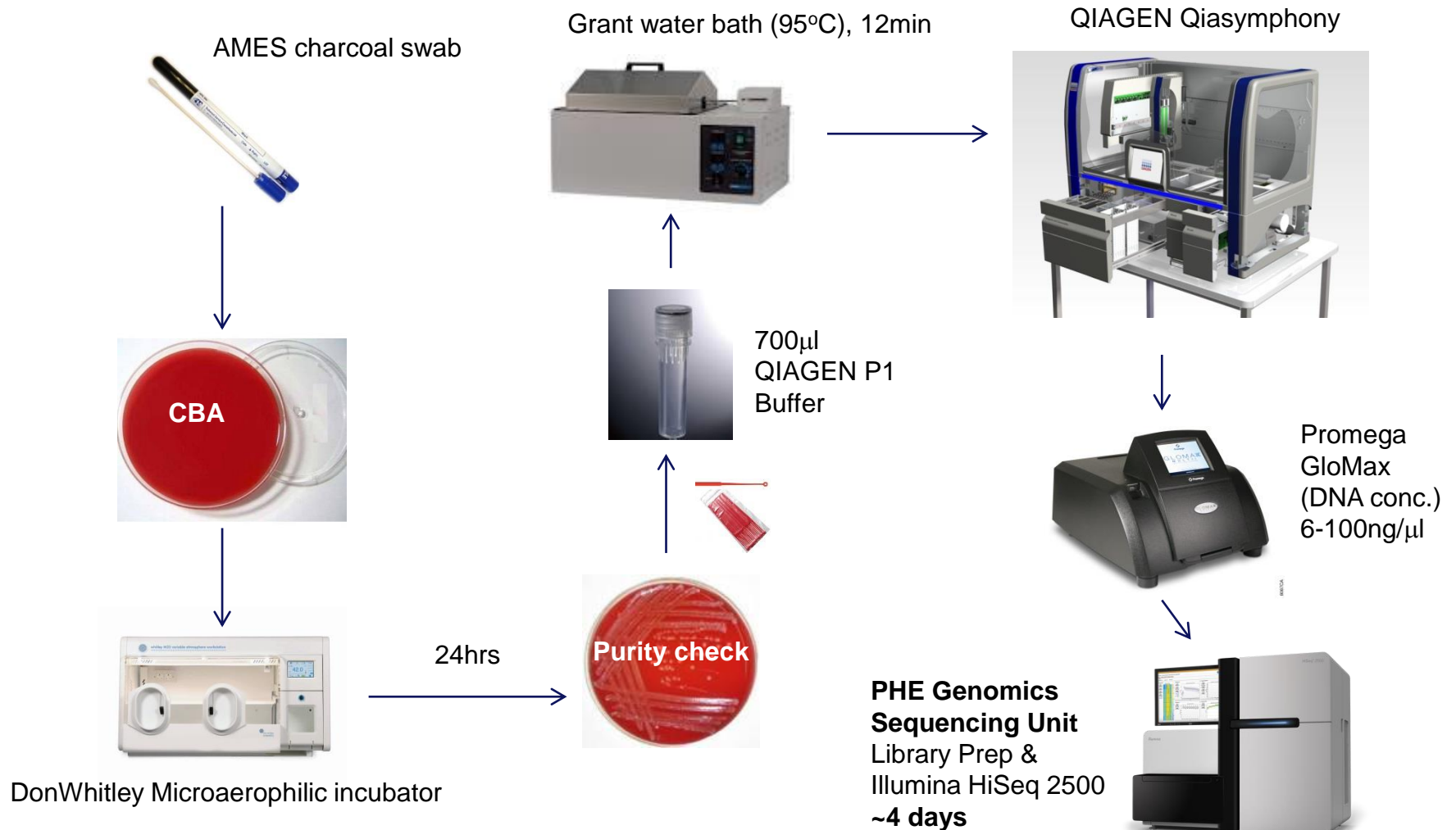
Campylobacter reference service over the years

Typing and identification methods

- Classical phenotypic methods used at the start of the *Campylobacter* reference laboratory:
 - Hippurate hydrolysis - differentiation of *C. jejuni* and *C. coli*.
 - Serotyping (heat stable antigens) (*Frost et. al. 1998*).
 - Phage typing (*Frost et. al. 1999*).
 - Antimicrobial resistance by agar Minimum Inhibitory Concentration (*Thwaites et. al. 1999*).
- **(2003)** Real Time TaqMan PCR for species identification for *C. jejuni* (*mapA* gene) and *C. coli* (*ceuE* gene) (*Best et. al. 2003*).
- **(2012)** Serotyping was replaced with Multi Locus Sequence Typing (MLST) (*PubMLST scheme*).
- **(2015)** Whole Genome Sequencing (WGS) was launched for **routine** identification and typing of every isolate of *Campylobacter* received into the reference laboratory.

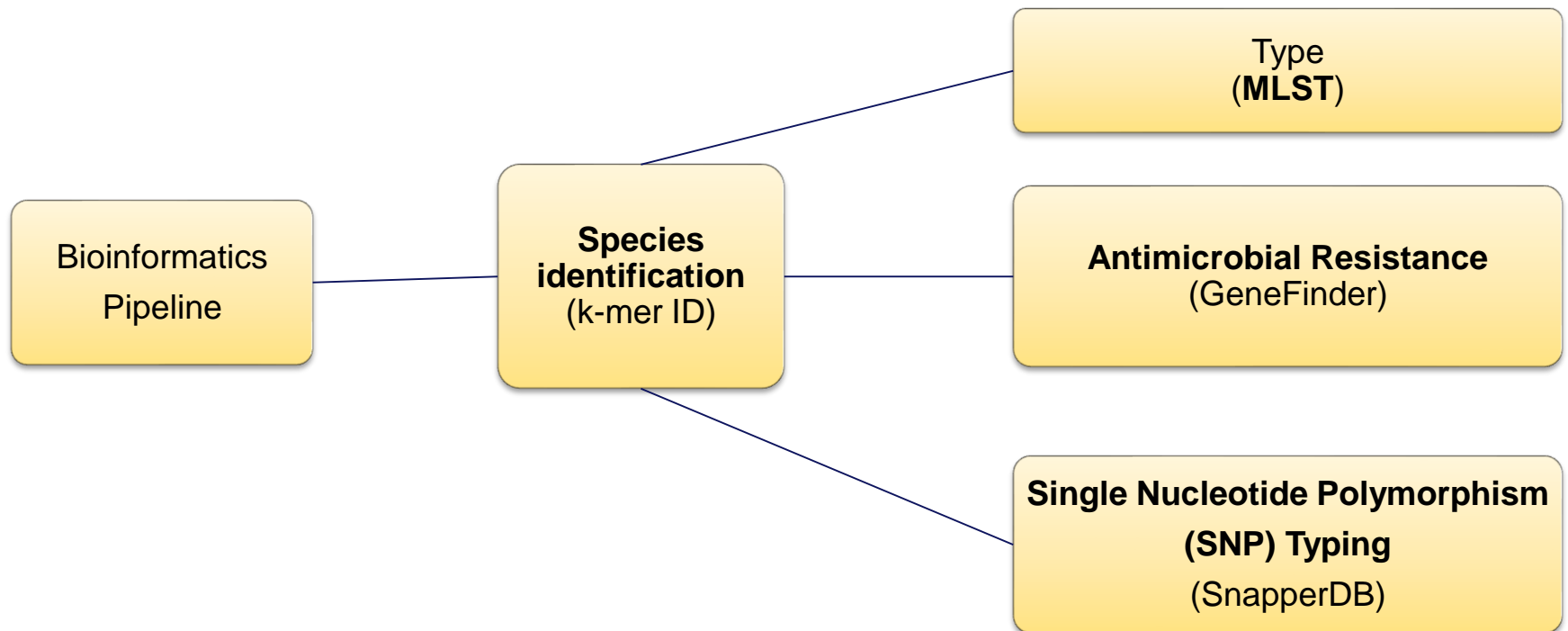


Whole Genome Sequencing workflow overview





Overview of the bioinformatics workflow





Species identification by k-mer

- 40 reference genomes from *Campylobacter* are present in our k-mer database alongside numerous other genomes from 136 other bacterial genus.
- Use k-mer of length 18
- Check what bacteria is present in the sample.
- Check if there is a mixture:
 - *C. jejuni* / *C. coli*
 - *C. jejuni* / *C. jejuni*
 - *C. coli* / *C. coli*



Quality metric: post-trim-yield

Quality metric PASS threshold: 150

Sample	Metric: post-trim-yield	Metric Status	Identification	meanQ	Percent clusters
1	484	PASS	Campylobacter coli LMG 23336 (89.38%)	37.0	1.5



Multi Locus Sequence Typing (MLST)

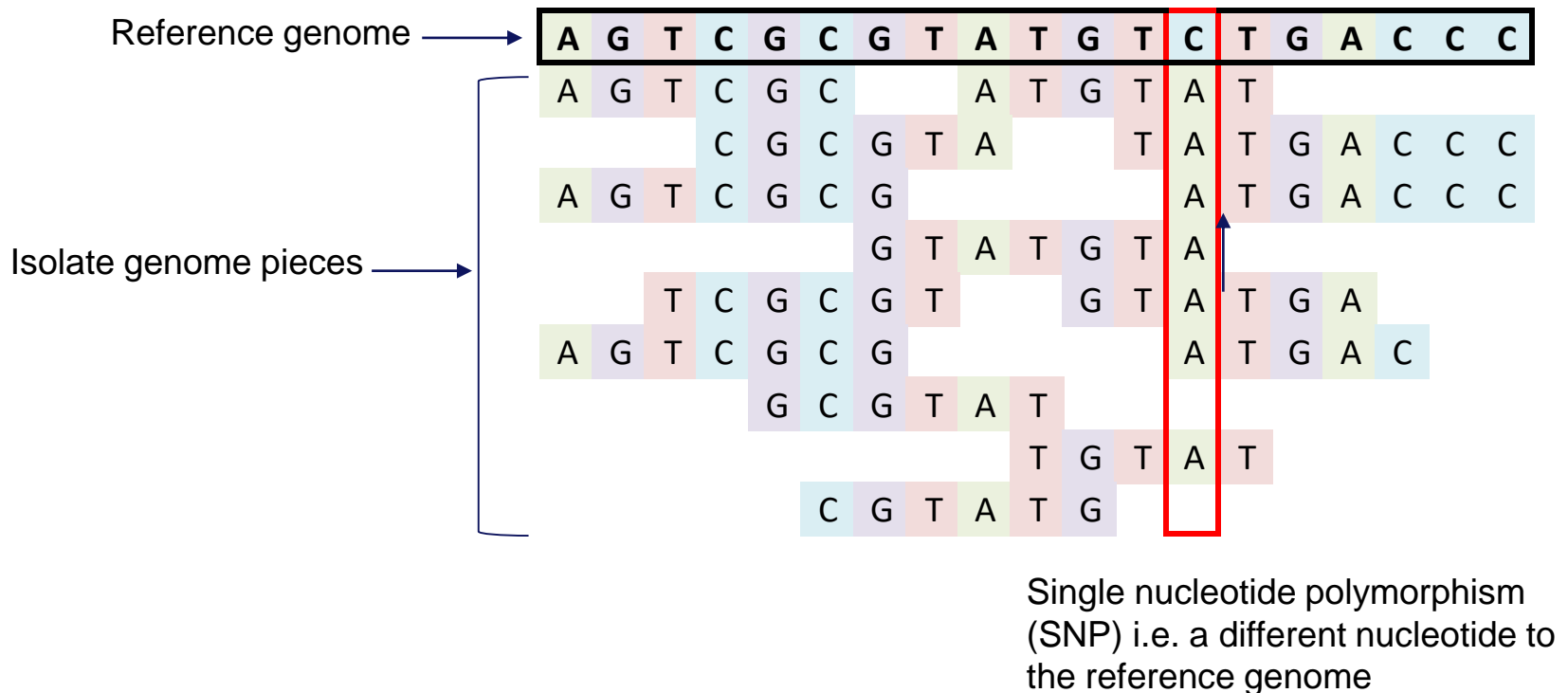
- PubMLST scheme <https://pubmlst.org/campylobacter/>
- 7 housekeeping genes
- More than 8000 individual ST have been defined
- Each ST is associated to an ST-complex (if possible)
- The 7-loci MLST is derived from WGS data using a bioinformatics tool “MOST”, a modified MLST typing tool based on short read sequencing (*Tewolde et al. 2016*)

Sample	MLST ST	MLST profile	QC mean cons depth	QC max % non cons base	QC % coverage	QC min cons depth
1	827	33,39,30,82,104,56,17	390.54 422,300,515 401,399,426,498	1.47 1,0,0 1,1,0,0	100	97 308,252,97 296,101,289,326



Single Nucleotide Polymorphism (SNP) Typing

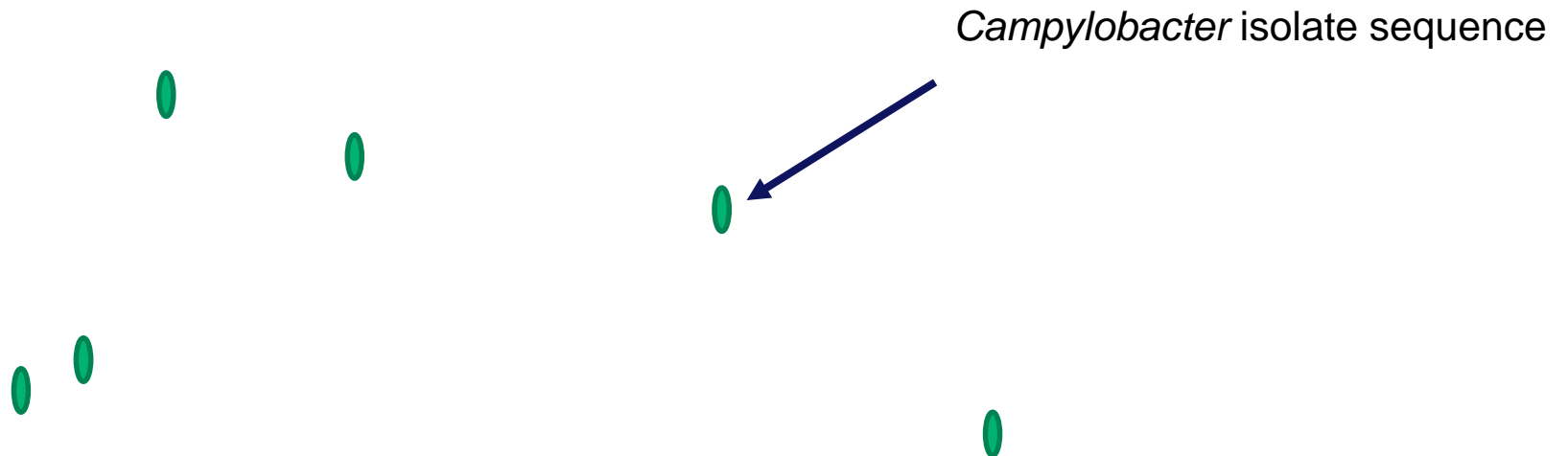
- Compare sequences to a known reference genomes for each major ST-complex.
- Reference genomes obtained from public databases.
- SNP typing used for detecting genetic diversity, surveillance & outbreak investigations.





Generating the SNP address

Imagine a space where all *Campylobacter* isolate sequences can be laid out

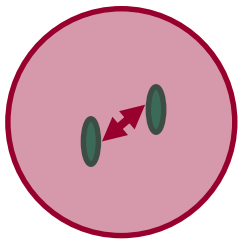




Generating the SNP address

In this space, similar sequences are next to each other

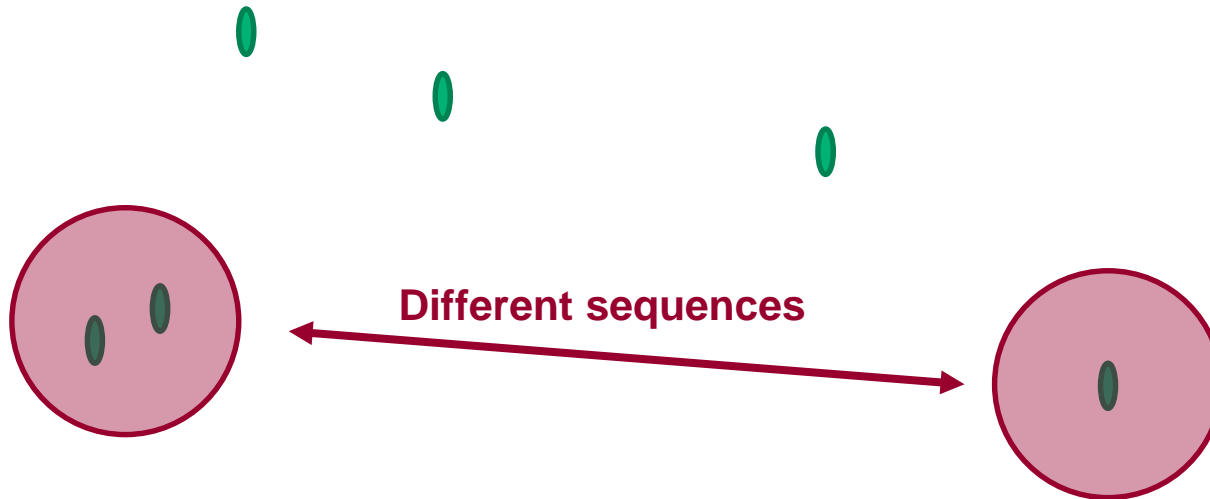
**Similar
sequences**





Generating the SNP address

The further apart the isolates are, the more different

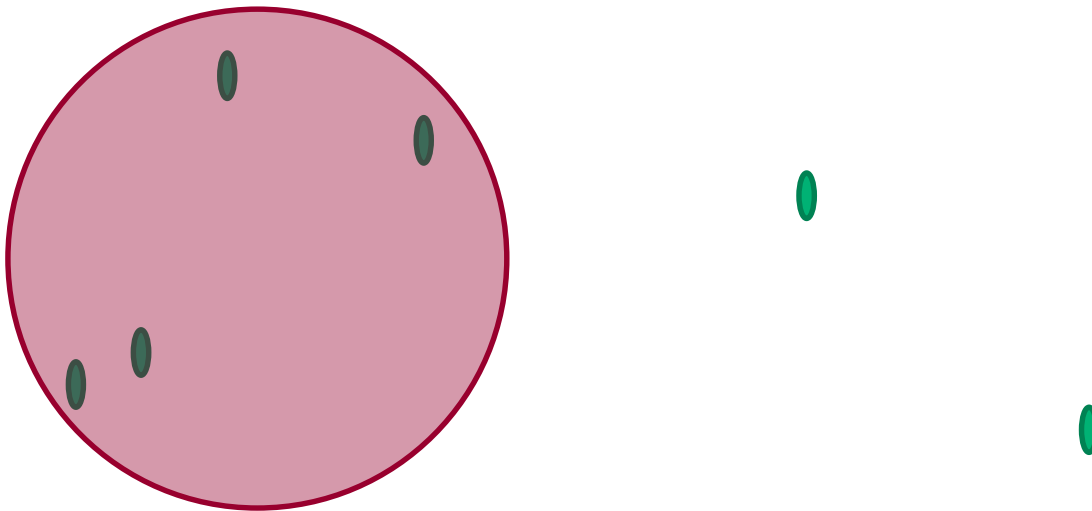




Generating the SNP address

The isolates are then grouped according to the number of SNP differences between them

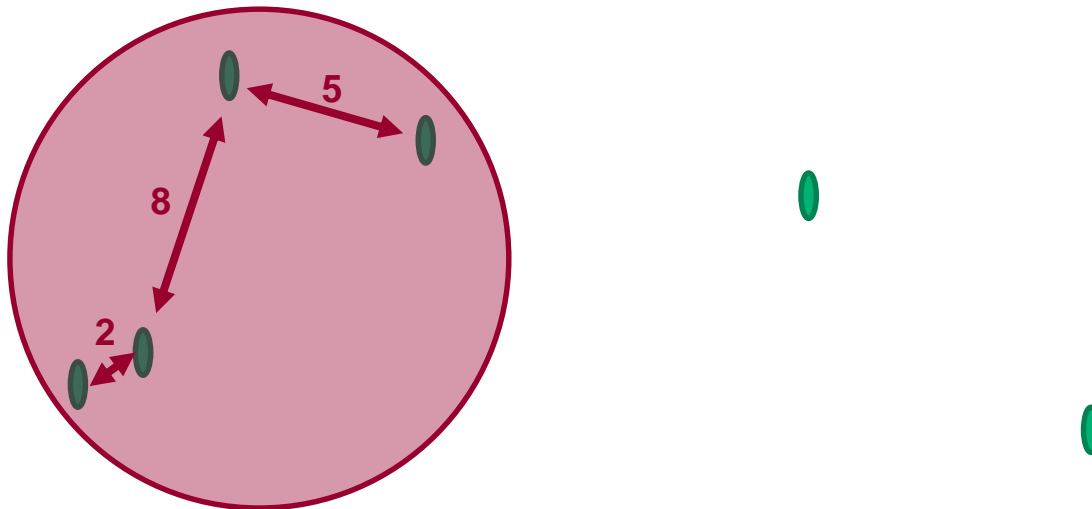
All of the isolates within the red circle have up to **10 SNP differences** to another isolate in the group





Generating the SNP address

The numbers indicate the number of SNP differences between the isolates

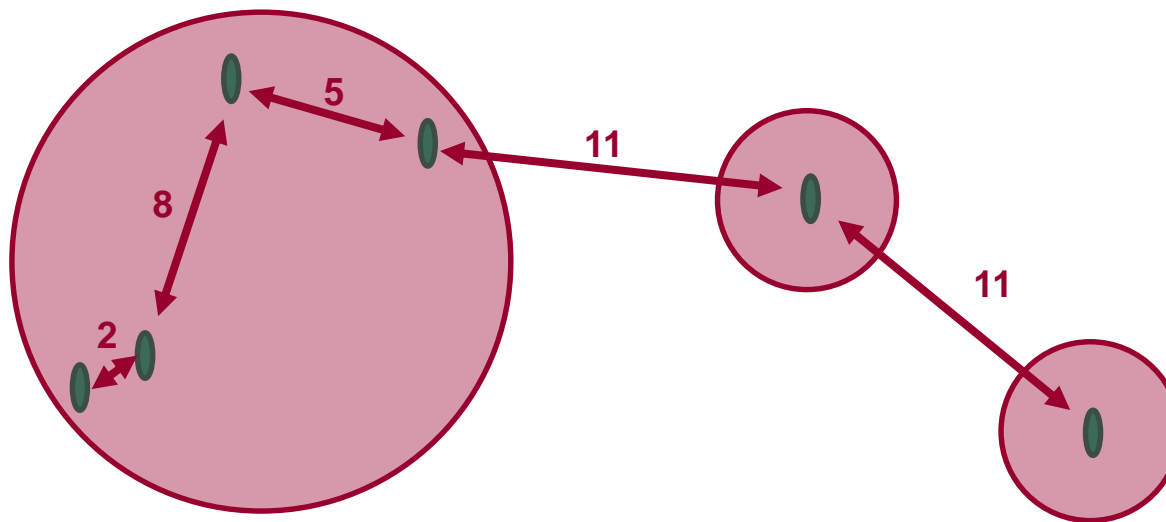




Generating the SNP address

The next closest isolate is 11 SNP differences away so is not included in the first group

The other two isolates in this example are in their own 10 SNP group

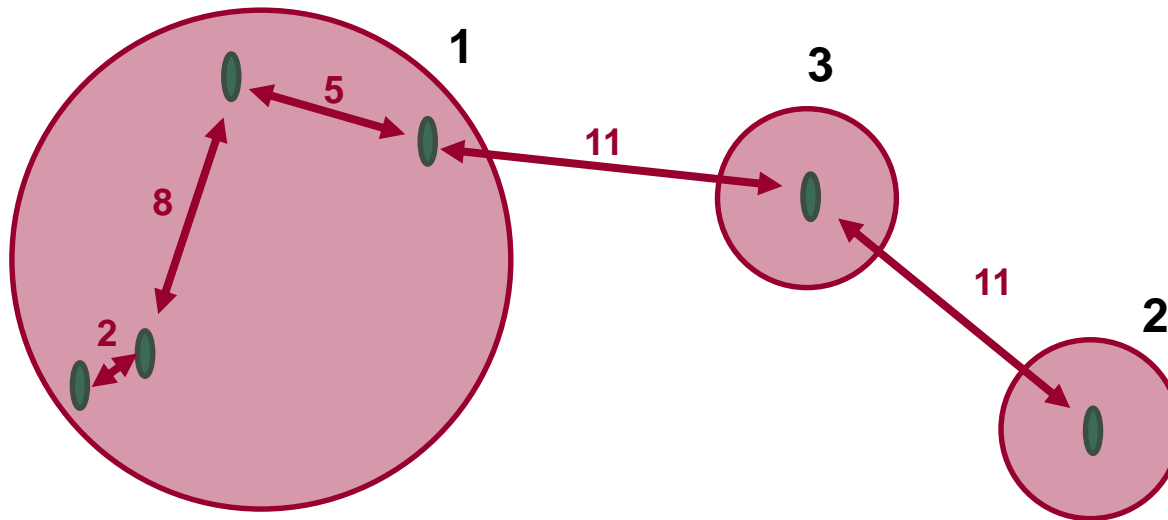




Generating the SNP address

The groups are given numbers to identify between them (numbers in black)

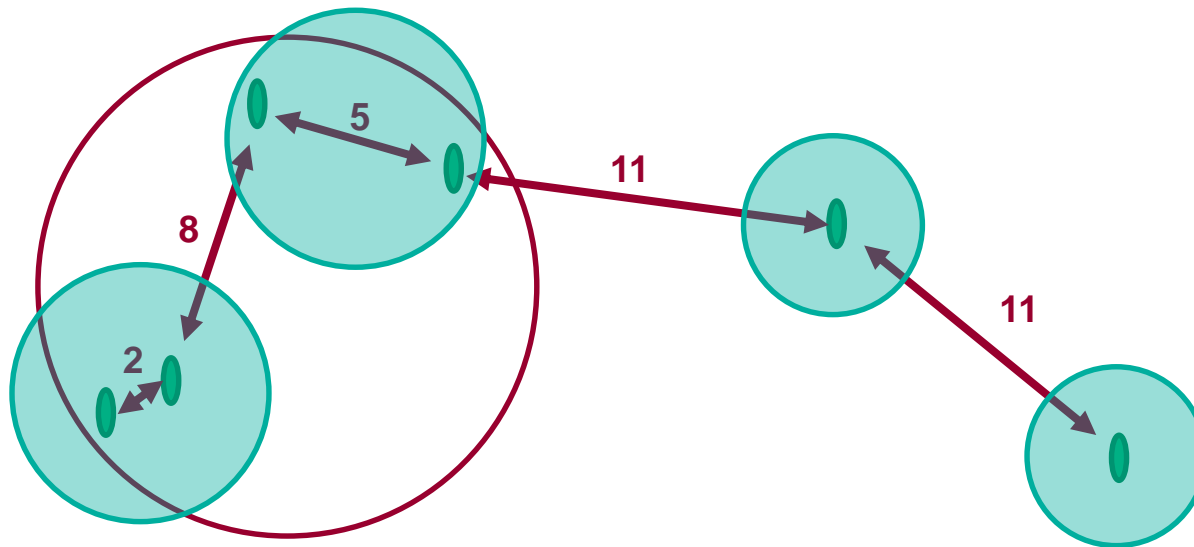
The numbers are associated only with the order in which they were discovered





Generating the SNP address

We also want to look for clusters of isolates that are more genetically similar so create groups with up to 5 SNP differences (green circles)

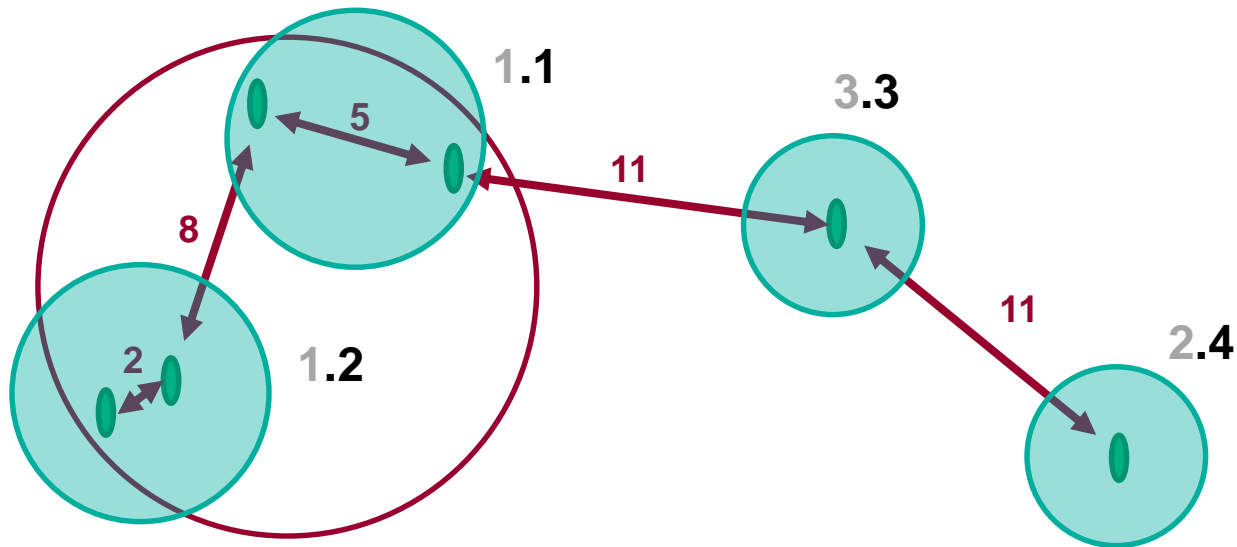




Generating the SNP address

We can again number these groups according to what was discovered first (black numbers)
– (grey numbers 10 SNP designation)

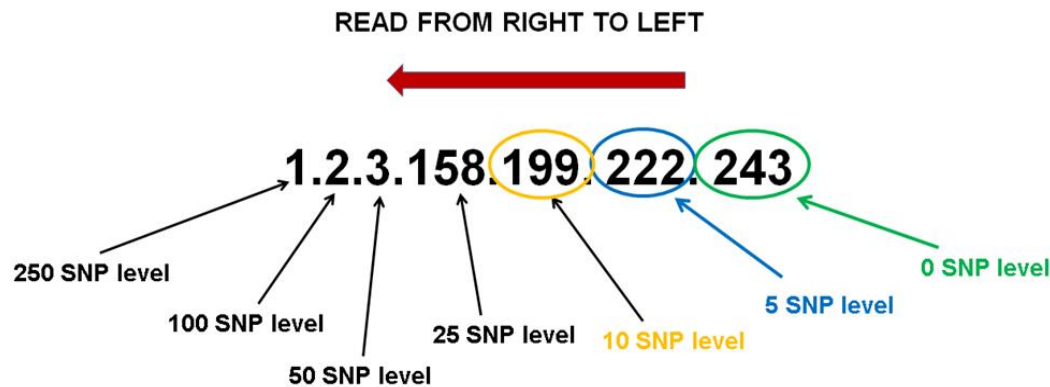
The two lone isolates are again in their own 5 SNP group





Generating the SNP address

- We can look at many different thresholds of SNP differences
- At present in PHE, we look at seven different levels: **250, 100, 50, 25, 10, 5** and **0**
- Each isolate gets a number for each of these different thresholds according to which cluster it falls into.



SNP address (Isolate 1)	1	1	1	158	199	222	243
SNP address (Isolate 2)	1	1	1	158	199	222	256
SNP address (Isolate 3)	1	2	2	35	60	125	160
SNP threshold	250	100	50	25	10	5	0



Antimicrobial resistance (AMR)

- GBRU classical method for determining AMR resistance for surveillance purposes is an **in agar dilution method** (labour intensive, prone to operator errors).
- Apply EUCAST interpretative threshold as outline in the “ECDC protocol for harmonised monitoring of antimicrobial resistance in human *Salmonella* and *Campylobacter* isolates (2016)”
- From WGS data we can use an in-house bioinformatics tool “**Genefinder**” to detect known antimicrobial resistance mechanisms to:
 - **Macrolides** – mutations in the 23S gene
 - **Fluroquinolones** – mutation in *gyrA* gene
 - **Tetracycline** – presence of *tetO* gene
 - **Aminoglycosides** – genes controlling reduced uptake or decreased cell permeability; alterations at the ribosomal binding sites; production of aminoglycoside modifying enzymes.



Summary

- GBRU have implemented routine WGS for *Campylobacter*:
 - Confirm species identification.
 - Derive the 7-loci MLST
 - Perform high discriminatory typing using SNP analysis
 - Infer phenotypic sensitivity to several classes of antimicrobial agents.
- GBRU report on implementing genomics produced with the Public Health Genomics Foundation (2018)

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/731057/implementing_pathogen_genomics_a_case_study.pdf





Questions

