

Multiple choice questions day 3 (for recap of day 2, IFAT form #E025)

1. Which of the below CGE methods does NOT use an assembly+BLAST approach to make its predictions?

- A: MLST
- B: ResFinder
- C: KmerFinder
- D: VirulenceFinder
- E: SerotypeFinder

2. Which of the below statements is NOT true, when it comes to WGS-based serotyping?

A: The H-type of *E. coli* is determined from WGS data by analysing the flagellin genes (*fliC*, *flkA*, *flhA*, *flmA*, *flnA*)

B: If no *E. coli* H-type is determined by SerotypeFinder it can only be explained by poor quality of the sequence data

C: The O-type of *E. coli* is determined from WGS data by analysing the O-processing genes (*wzx*, *wzy*, *wzm*, and *wzt*, and *wzx/wzy* or *wzm/wzt* pairs)

D: The tool SerotypeFinder is for serotyping of *E. coli*

E: The tool SeqSero is for serotyping of *Salmonella*

3. Which of the below statements is NOT true, when it comes to VirulenceFinder?

A: VirulenceFinder can identify virulence factors in *Vibrio cholera*

B: VirulenceFinder can identify virulence factors in *Escherichia coli*

C: VirulenceFinder can identify virulence factors in *Staphylococcus aureus*

D: VirulenceFinder can identify virulence factors in *Enterococcus*

E: VirulenceFinder is an example of an assembly+BLAST-based method

4. Below, the sequence of four strains (strain1-4) is shown. Which pair of strains have the highest number of SNPs?

Strain 1: CGTTACGTT

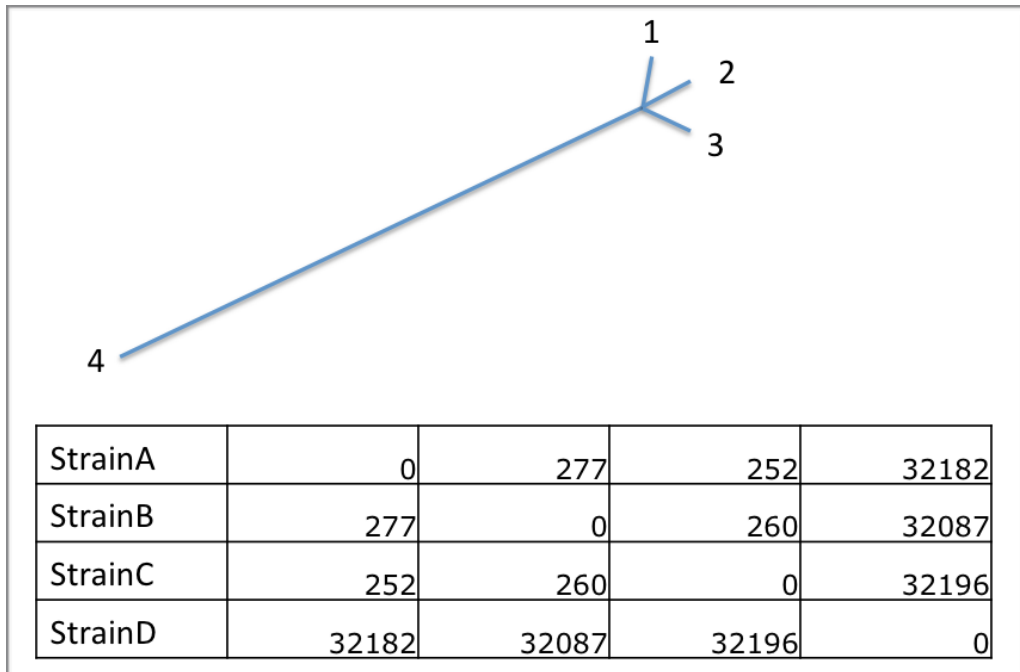
Strain 2: CGATACGTT

Strain 3: CCATACGTT

Strain 4: TCATACGAT

- A: Strains 3 and 4
- B: Strains 1 and 2
- C: Strains 2 and 3
- D: Strains 2 and 4
- E: Strains 1 and 4

5. The phylogenetic tree below illustrates the relationship of four isolates. Below the tree, the distance matrix upon which the tree is based, is shown. Which strain does “4” correspond to?



- A: It is not possible to determine
- B: StrainC
- C: StrainB
- D: StrainD
- E: StrainA

6. Which of the below statements related to phylogenetic trees is NOT true

- A: A rule of thumb states that outbreak strains differ by less than 10 SNPs
- B: The Neighbour Joining algorithm works well when samples have been taken at different time points
- C: If possible, always use raw reads as the input to NDtree and CSIPhylogeny - not assembled genomes
- D: Two strains that differ by 200 SNPs are likely to be part of the same outbreak
- E: CSIPhylogeny is more strict than NDtree and hence considers fewer positions in the same set of input genomes

7. Which of the below statements is true, when it comes to assembly+BLAST-based methods versus read mapping methods?

- A: The read mapping methods are more sensitive, and hence have a tendency to result in more false positives
- B: Read mapping methods are likely to miss entire genes, if the sequence coverage in the area is a bit low
- C: The CGE MLST tool currently uses read mapping to identify the Sequence Type of the input sequence

- D: The assembly+BLAST-based methods are generally faster than the read mapping methods
- E: KmerResistance predicts the resistance of bacterial strains towards kmers

8. Which method is considered to have the highest discriminatory power?

- A: core genome MLST (cgMLST)
- B: A SNP-based method
- C: A method sampling only the 16S rRNA gene
- D: MLST
- E: rMLST, which samples app. 53 ribosomal genes

9. Which CGE method allows you to upload and use your own database with any genes of interest?

- A: KmerFinder
- B: NDtree
- C: ResFinder
- D: MyDBFinder
- E: CSIPhylogeny

10. This questions can probably only be answered after the first lecture of today. Which of the below methods is NOT a part of the Bacterial Analysis Pipeline?

- A: KmerFinder
- B: MGMapper
- C: ResFinder
- D: MLST
- E: pMLST