**EURGen RefLabCap Multidisciplinary workshop on CCRE outbreak detection**

Hereby some questions for consideration when working with the data set for the exercise.

Outbreak detection:

Do you find one or more clusters within the 12 E. coli in the dataset?

What is the number of SNPs within the cluster(s)?

Did it do any difference if you used draft genomes or raw data in CSIphylogeny? [Analysis 2 vs 4]?

Did exclusion of less related isolates from the SNP analysis change the number of SNPs? [Analysis 4 vs 5]?

How well did the draft assemblies of ONT MinION data perform compared to draft assemblies from Illumina [Analysis 3 vs 7]?

Did the number of SNPs change when using the optimal reference compared to the kmer-reference? [Analysis 2 vs 3] and [Analysis 5 vs 6].

Did you obtain the same clustering using Illumina data with CSIphyogeny and MinTyper? [Analysis 6 vs 8]?

Did you get the same clustering with Illumina raw data in CSIphylogeny as you do with ONT MinION data in MinTyper using the best reference on the 9 closest isolates [Analysis 6 vs 10]

Was it possible to combine Illumina data and ONT MinION data (fast or Super accuracy\*) in one analysis in a way that was able to group individual isolates sequenced with different methods (e.g. Ec001 from Illumina grouping with Ec0001 from ONT MinION)? [Analysis 12 and 13]?

\*note: only the fast accuracy fastq files are shared

Additional subtyping

Which of the subtyping method you tried has the highest discriminatory power?

Which subtyping method adds most value in relation to the cluster analysis – MLST, serotyping, cgMLST?

Antimicrobial resistance and plasmids

Which OXA- and NDM- genes were found in the bacteria?

Did clustering strains contain the same type(s) of carbapenemase genes?

Did any of the carbapenemase genes seem to be on a plasmid?

If yes, are there other resistance genes on the same plasmid?

-Hint: Running MobileElementFinder on Illumina fasta files shows the location of plasmid replicons and resistance genes on contigs

<https://cge.food.dtu.dk/services/MobileElementFinder/>

Based on the sum of all your analysis, which isolates do you think belongs to the same hospital outbreak?

Do this molecular clustering find support in the epidemiological data?

Is there any of the molecular data (SNPs, Resistance genes, Plasmids ect) which is not supporting what would be expected from a clonal outbreak? What can be the explanation(s)?