First EURGen-RefLabCap virtual multidisciplinary training workshop

Exercise overview

Objective:

The objective of this virtual multidisciplinary training workshop is to equip the participants with background information about bacterial subtyping and cluster analysis, as well as suggestions for online available analytical tools, to get started on bacterial comparison and outbreak detection.

There will be opportunity to work with both Illumina short read sequences and Oxford Nanopore (ONT MinION) long read sequences, but it is not a requirement to work with both types of files to participate in the exercise.

Scenario:

A recent rise in cases of carbapenemase producing *E. coli* in several regional hospitals indicate one or more ongoing outbreaks, and it has been suggested that the NRL could give assistance by performing outbreak investigation by WGS. Patients involve both domestic and travel-related cases and a batch of samples has already been sequenced using Illumina sequencing platform (NextSeq). From these sequences, subtyping by MLST was performed and a selection (12 *E. coli* isolates) of the most predominant MLST (ST410) isolates has been transported to your laboratory for further analysis. Your laboratory has just finalized setting up MinION (Oxford Nanopore; ONT) sequencing, and you wish to use this occasion to work with both types of sequences. The currently available metadata for the samples can be seen in Table 1.

Table 1 Metadata for the 12 carbapenemase producing *E. coli* isolates

Species	Date	Region of isolation	Travel	MLST	Sequence	Carba genotype (PCR)
E. coli	2015	Copenhagen	Pakistan	ST410	Ec001	OXA-48-like
E. coli	2015	Copenhagen	Thailand	ST410	Ec002	OXA-48-like
E. coli	2015	Jutland - M	India	ST410	Ec003	NDM
E. coli	2015	Copenhagen	Lebanon	ST410	Ec004	OXA-48-like
E. coli	2016	Zealand	No	ST410	Ec005	NDM, OXA-48-like
E. coli	2016	Zealand	No	ST410	Ec006	NDM, OXA-48-like
E. coli	2017	Copenhagen	Pakistan	ST410	Ec007	OXA-48-like
E. coli	2018	Jutland - N	Thailand	ST410	Ec008	NDM
E. coli	2018	Zealand	No	ST410	Ec009	NDM, OXA-48-like
E. coli	2018	Zealand	No	ST410	Ec010	NDM, OXA-48-like
E. coli	2018	Zealand	No	ST410	Ec011	NDM
E. coli	2018	Zealand	No	ST410	Ec012	OXA-48-like

Materials for exercise:

Two sets of sequences of 12 bacterial isolates (*Escherichia coli*) can be included in the exercise, and the sequences are available in both fasta and fastq format. You can find a short description of the file formats and size of data in the end of this document (Table 3). The data files are available for download from the website sciencedata.dk through the link below:

https://sciencedata.dk/shared/007e3242ab05e33a01de62cf24ce8eda

There will be 4 zipped folders covering:

1) Illumina fastq sequences (trimmed raw reads)

2) Illumina assemblies (contigs; fasta format)

3) ONT MinION fastq sequences (basecalled)

4) ONT MinION assemblies (contigs; fasta format)

Tasks:

The main task in this exercise is to perform cluster analysis of either Illumina data alone or together with MinION data for comparison, and use the results hereof together with the metadata in table 1 to elucidate possible outbreak isolates in the dataset. There is no need to perform the analysis on both fasta and fastq files from the same sequencing technology, so chose the format that is most suitable for your analysis setup and connection. During the first day of the exercise, we will give an introduction to the tools, how to select a reference and how to work with and interpret the results.

Suggested tools, which will be presented during the first day (Sept 26th):

CSIphylogeny (Illumina data only): https://cge.food.dtu.dk/services/CSIPhylogeny/

MinTyper (Illumina AND MinION data): https://cge.food.dtu.dk/services/MINTyper/

Evaluation:

There will be no hand-in of analysis results from the participants. We will give a walk-through of the cluster analysis results on the second day of the exercise (Oct 10th), including considerations on additional analyses like identification of resistance genes and plasmids, and subtyping by cgMLST. There will be time for questions, and participants are welcome to show their results from e.g. their own pipelines etc.

Additional analyses:

Participants are encouraged to perform additional analyses related to subtyping (e.g. cgMLST, serotyping) or identification of resistance determinants and plasmids (ResFinder, KmerResistance, PlasmidFinder etc.)

The main website with an overview of CGE tools (described in brief in the webinar Sept 14th):

http://www.genomicepidemiology.org/services/

Here you can find a range of additional tools that can be explored. Some of the relevant tools are listed below:

Table 2: Examples of tools that can be used for additional analyses. Some tools can be used for both MinION and Illumina data

Tool	Illumina	Both Illumina and MinION
cgMLST		https://cge.food.dtu.dk/services/cgMLSTFinder/
ResFinder	https://cge.food.dtu.dk/services/ResFinder/	
KmerResistance		https://cge.food.dtu.dk/services/KmerResistance/
PlasmidFinder		https://cge.food.dtu.dk/services/plasmidfinder/
MobileElementFinder	https://cge.food.dtu.dk/services/MobileElementFinder/	

File formats and data details:

The assemblies/fasta files are smaller files to work with (as they contain less information about e.g. sequence quality) than the raw reads, and are especially suitable for low speed internet connections or low computer data storage. The fastq files are available if you prefer to work with these or want to work with your own analysis pipeline.
The Illumina data are available as trimmed reads and in the trimming stage, sequencing adaptors were removed and reads were end-trimmed to quality Q_20 using Trimmomatic. The Illumina fasta files were generated using the pipeline Bifrost.
For the ONT MinION data base-calling, demultiplexing, and conversion to fastq format from the raw fast5 reads were done using Albacore v2.3.4. Sequencing adapters were removed with Porechop v0.2.3.

Table 3: Overview of available files and sequence ID for each of the 12 *E. coli* strains 001-012. The data amount in the zipped download folders is indicated in brackets.

	Illumina	**ONT MinION**
Fastq – raw reads	Ec0xx.illumina_R1.trimmed.fastq (4 GB)	Ec0xx_fast.fastq (6 GB)
Fasta - assemblies	Ec0xx_assembly.fa (15 MB)	Ec0xx_ONT_assembly (15 MB)