

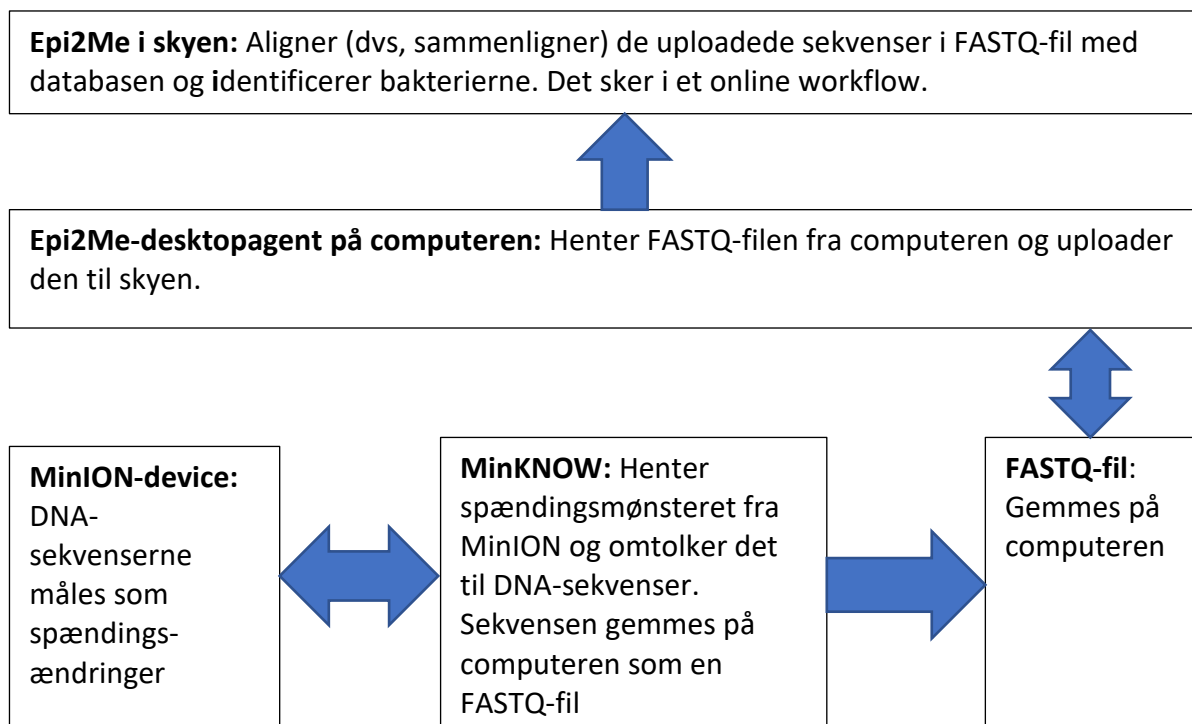
Databehandling fra sekventering

Indhold

Databehandling fra sekventering.....	1
Tolkning af Workflow-rapporten i Epi2Me	2
Efterbehandling af data	4
Hvad har I fundet?	4
Videre muligheder for databehandling	5
FASTQ-filer	5

Data fra MinION-devicen oversættes i programmet minKNOW til en FASTQ-fil. FASTQ-filen gemmes på computeren i den mappe (directory), som man har angivet i minKNOW.

Data kan derefter bearbejdet i programmet Epi2Me:

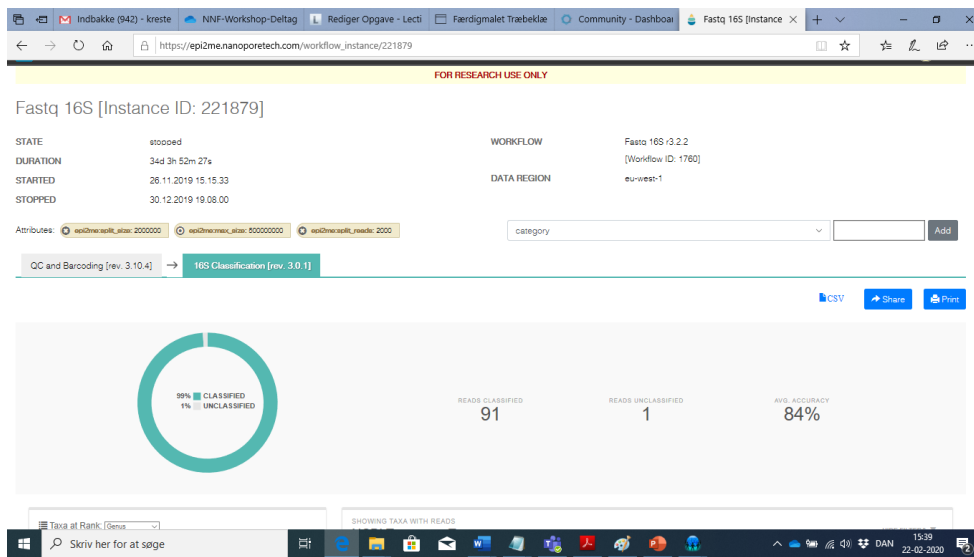


Figur 1. Programmer og databehandling.

Vi laver databehandlingen i Epi2Me fælles for klassen. Derefter tolker I på resultaterne i grupper.

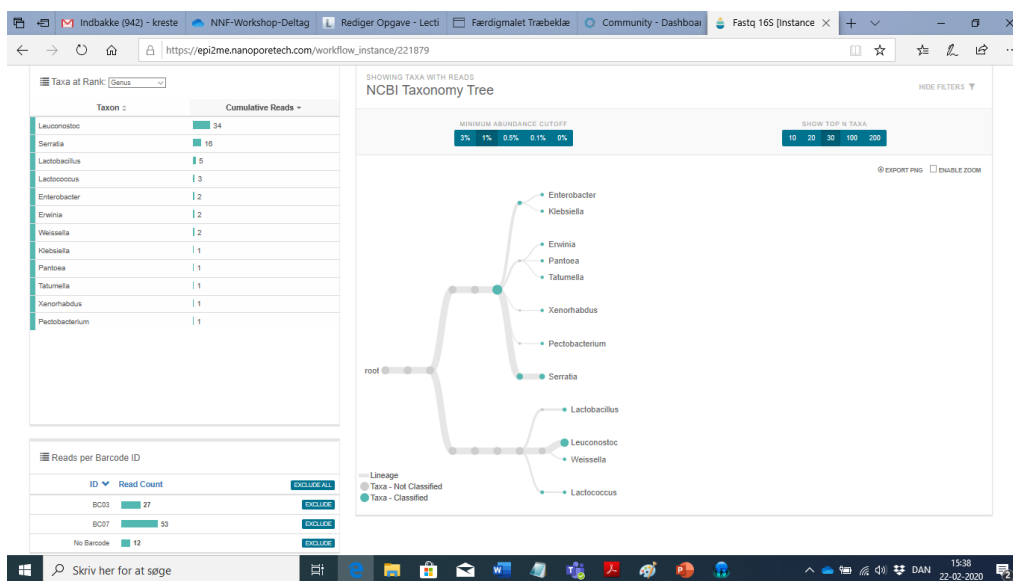
Tolkning af Workflow-rapporten i Epi2Me

Første side indeholder oplysninger om sekventeringens forløb, antal succesfulde og mislykkede reads og deres gennemsnitlige nøjagtighed i forhold til databasens sekvenser.



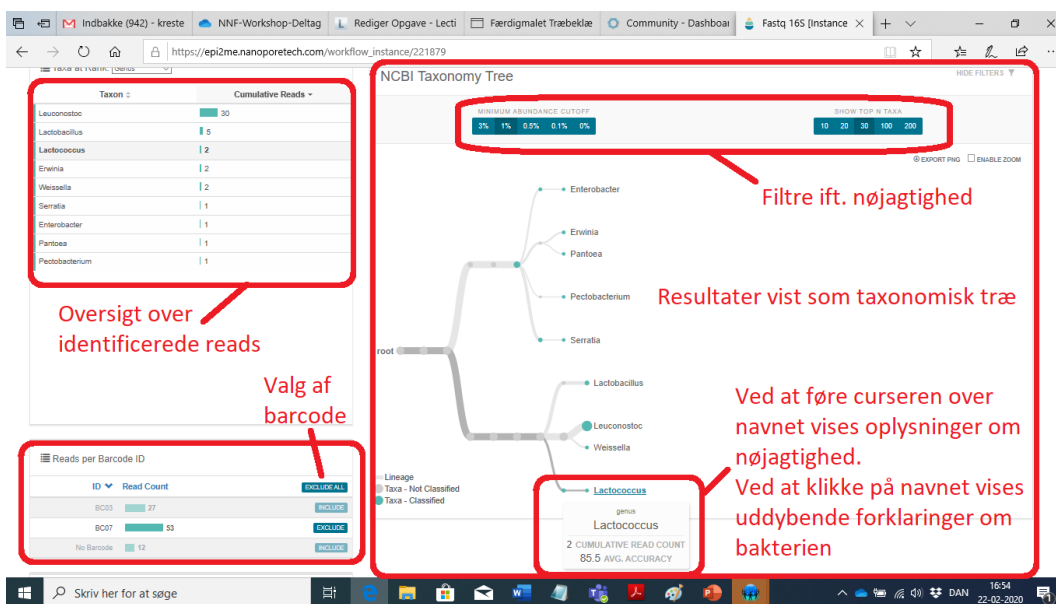
Figur 8.

Længere nede på siden vises de samlede resultater fra alignment med databasen:



Figur 9.

Figur 10 viser hvad de forskellige elementer på siden angiver:





Figur 10.

Efterbehandling af data

Hvad har I fundet?

Del de identificerede bakterier imellem jer og undersøg dem nærmere. Vær opmærksom på om de oplysninger I finder er på arts-, slægts- eller familie-niveau. Indenfor samme slægt kan der være stor forskel på enkelte bakterietypers biologi og eventuelle patologi (evne til at fremkalde sygdomme).

Navn på bakterie					
Hvordan ser den ud?					
Form: kok, stav, skrueformet? 					
Lejring: -enkelt (monokok), -dobbelt (diplokok, diplobacillus), -kæde (streptokok /streptobacillus), -stak (staphylokok)? 					
Gram+ eller gram-?					
Hvilke egenskaber og stofskifteprocesser er karakteristiske for bakterien?					
Kan bakterien være patogen for planter, dyr eller mennesker?					
Hvor forekommer den normalt?					
Hvilke abiotiske forhold karakteriserer dens niche?					
Lever den associeret til andre organismer, fx planter eller dyr?					
Ville du forvente at finde den, der hvor I har jeres prøve fra?					

Fremlæg jeres bakterier for klassen.

Sammenlign de bakterier der forekommer på de fire lokaliteter. Hvad viser de om lokaliteterne?

Videre muligheder for databehandling

De enkelte sekvenser i FASTQ-filerne kan kopieres og sammenlignes med indrapporterede sekvenser i genbank. Det gøres ved Nucleotid BLAST.

1. Åben hjemmesiden:
<https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE=BLASTHome>
2. Vælg Nucleotide BLAST
3. Kopier en sekvens fra FASTQ-filen og indsæt den i søgefeltet 'Enter... FASTA sequence'.
4. Tryk på BLAST.

Nu fremkommer en side med de indrapporterede sekvenser, som matcher sekvensen bedst.

5. Hvilke oplysninger kan ses på hjemmesiden?
6. Hvilke bakterier matcher bedst?
7. Hvor godt matcher sekvensen med de indrapporterede?

FASTQ-filer

Resultaterne fra sekventeringen vil ligge på computeren som en FASTQ-fil. I FASTQ-filen er hver DNAsekvens repræsenteret af 4 linjer.

- Unikt ID starter med et '@'. Linien kan fx indeholde information om prøven og flowcellen.
- Selve DNA-sekvensen (ATGC....)
- Et '+', som markerer at sekvensen slutter
- En linje med en kvalitetskode for hver af baserne i linie 2. Kvalitetskoden er angivet med ASCII-symboler¹.

1. Åben en af FASTQ-filerne i en text-editor.
2. Identificer de fire linjer for hver DNA-sekvens. Hvilke oplysninger kan du finde i linierne?

¹ Værdien er angivet med ASCII-symboler, hvis betydning kan findes ved søgning på nettet. Værdien angiver $Q = -\log_{10} P$, hvor P er sandsynligheden for at en base er fejlbestemt. Dvs. at jo højere Q -værdi, jo mere sikker er bestemmelsen.