

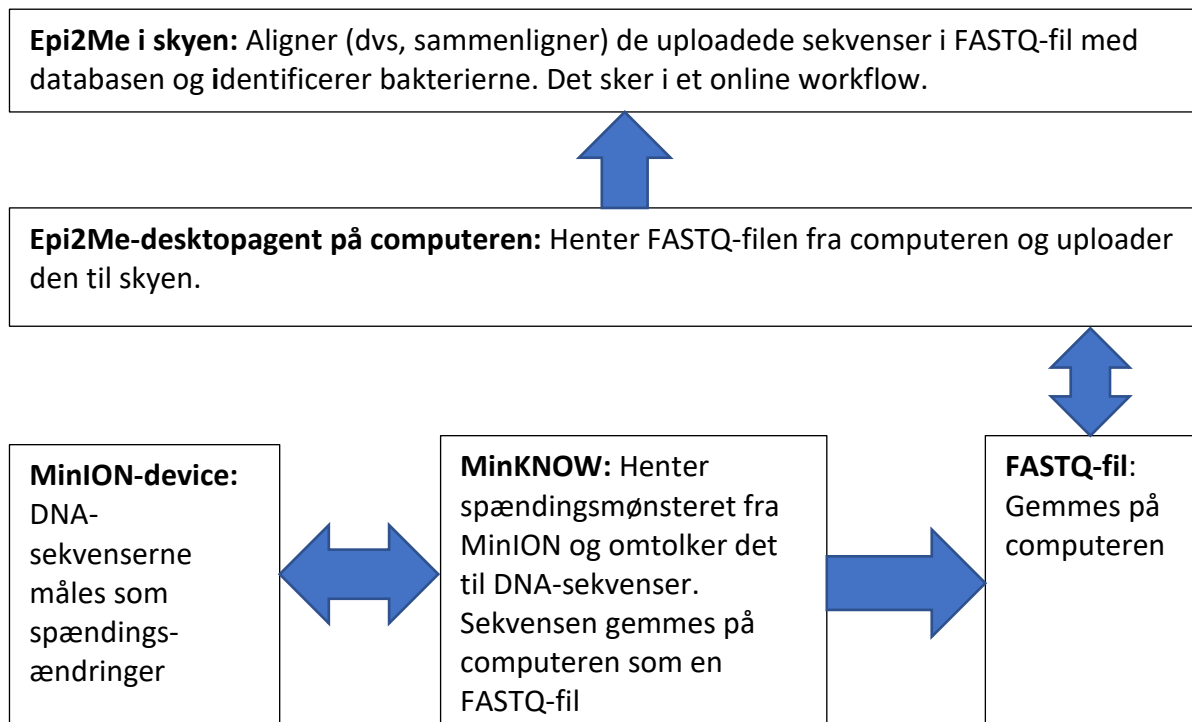
Databehandling fra sekventering

Indhold

Databehandling fra sekventering.....	1
FASTQ-filer	2
Databehandling i Epi2Me.....	2
Starte et workflow i Epi2Me	2
Tolkning af resultater	6
Efterbehandling.....	7
Hvad har I fundet?	Fejl! Bogmærke er ikke defineret.
Fylogenetiske træer: Hvordan er bakterier beslægtede?.....	Fejl! Bogmærke er ikke defineret.
Videre muligheder for databehandling	Fejl! Bogmærke er ikke defineret.
Installation af programmet Epi2Me.....	2

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.

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?

Installation af programmet Epi2Me

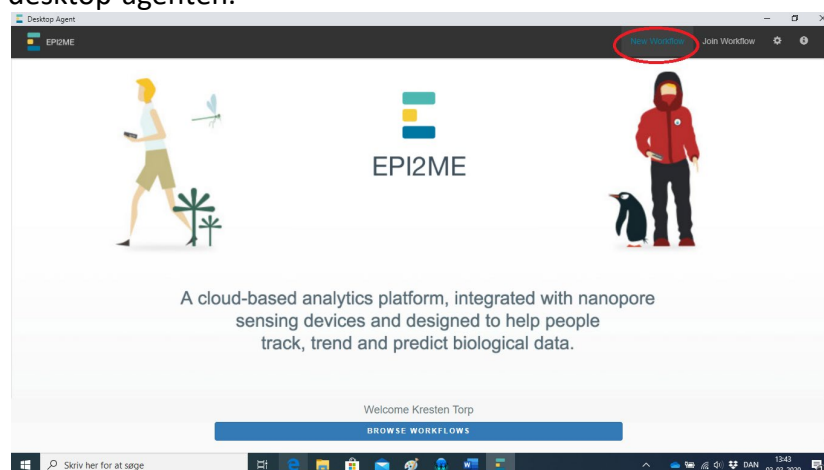
Programmet Epi2Me downloades som en desktop-version fra Nanopore Community og installeres på computeren, hvorefter man åbner sine FASTQ-filer i workflows online.

1. Log ind på Nanopore Community: <https://community.nanoporetech.com/>
2. Find software downloads til højre.
3. Vælg Epi2Me desktop agent og 'download'.
4. Vælg "kør".
5. Du skal muligvis godkende at computeren installerer programmet.
6. Start programmet, når installationen er slut. Du bliver nu bedt om at registrere dig. Klik på linket og kopier registreringskoden. Indsæt den, og gå videre.

Databehandling i Epi2Me

Start et workflow i Epi2Me

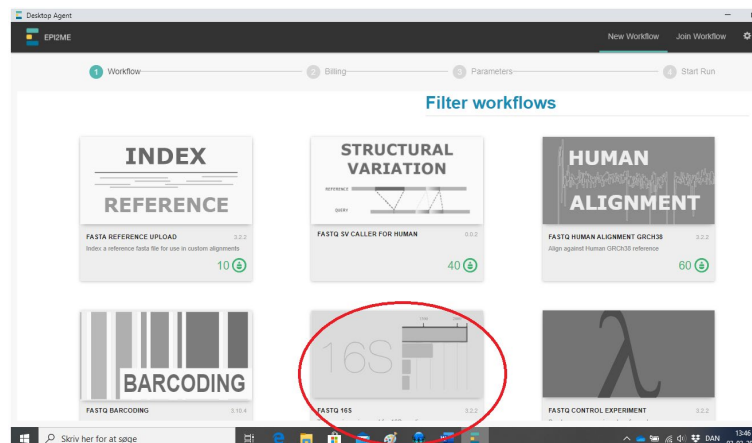
1. Log på Nanopore Community og Epi2Me
2. Start Epi2Me desktop-agenten.



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.

Figur 2

3. Vælg 'New workflow'

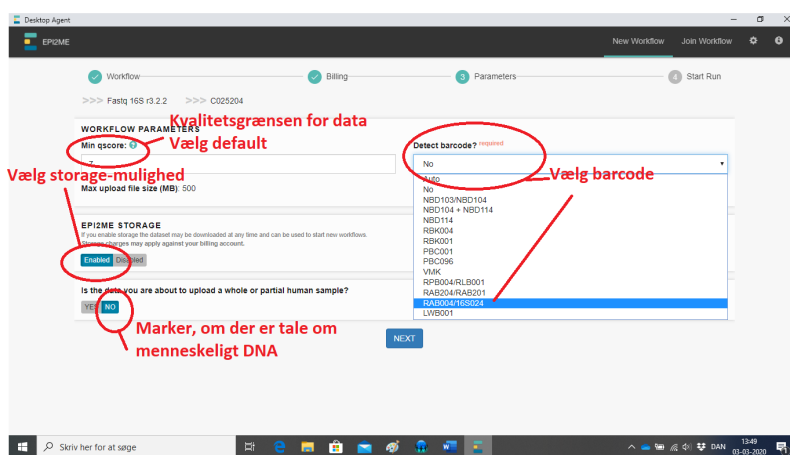


Figur 3.

4. Vælg 16S

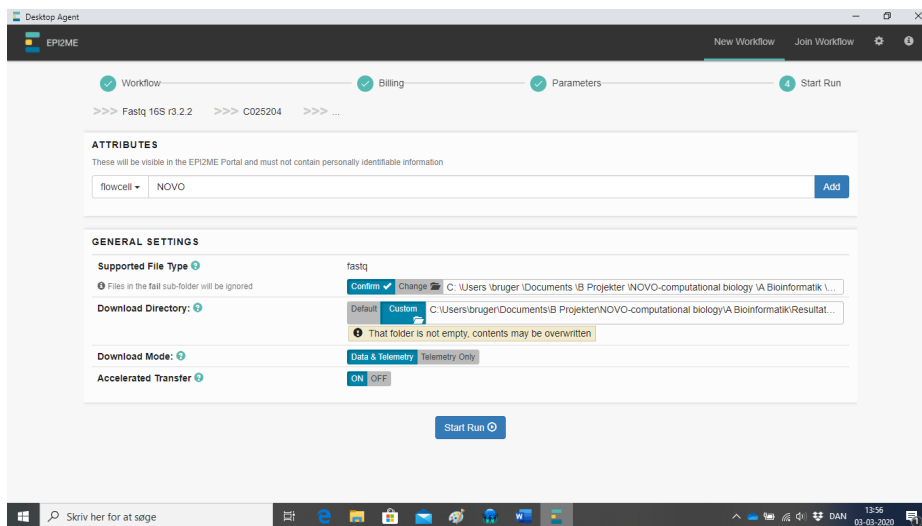
5. Herefter kommer nogle registreringer, som skal foretages:

- Min. qscore: Kvalitetsgrænsen for hvilke resultater, der skal medtages.
- Detect barcodes: Vælg 16S024 i dette tilfælde.
- Human sample: Skal angives. Hvis der sekventeres bakterier: 'No'
- Epi2Me storage er en mulighed for at gemme på hjemmesiden. Der kan i fremtiden komme betaling på dette, men har vi ikke oplevet indtil nu.



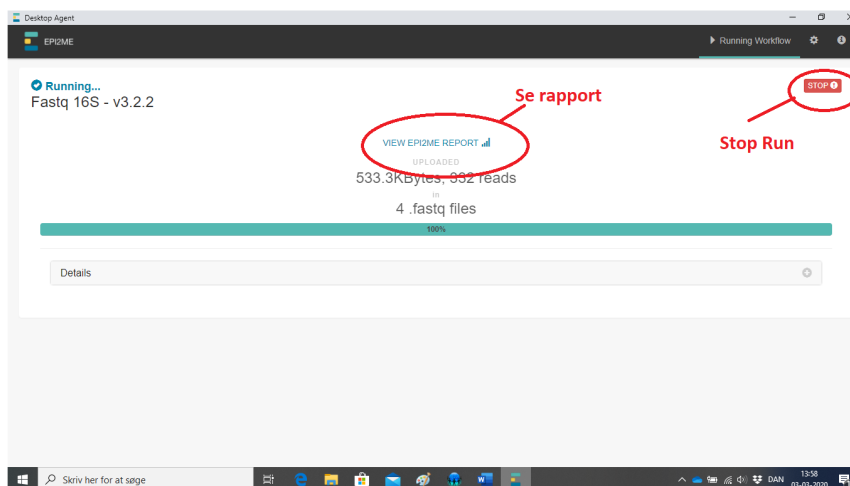
Figur 4.

6. Tryk 'Next'.



Figur 5.

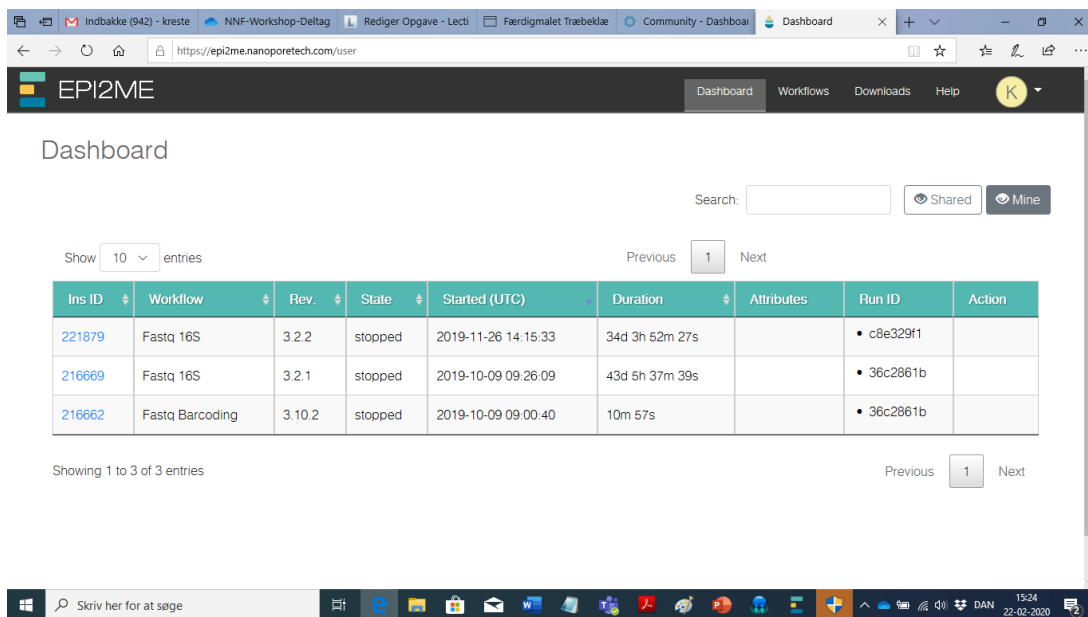
7. Herefter angives et projekt-navn, der vælges mappe, hvorfra FASTQ-filerne skal hentes, og hvor resultaterne skal gemmes.
8. Vælg 'Accelerated transfer'
9. Vælg 'Start Run'. Nu uploades datafilen til skyen.



Figur 6.

10. Når filen er hentet, trykkes 'Stop Run'
11. Vælg 'View Epi2Me report'. Nu åbnes hjemmesiden, og man kan se resultatet af workflowet.

På 'dashboard' i online-programmet kan du altid vende tilbage til workflows, du har kørt:

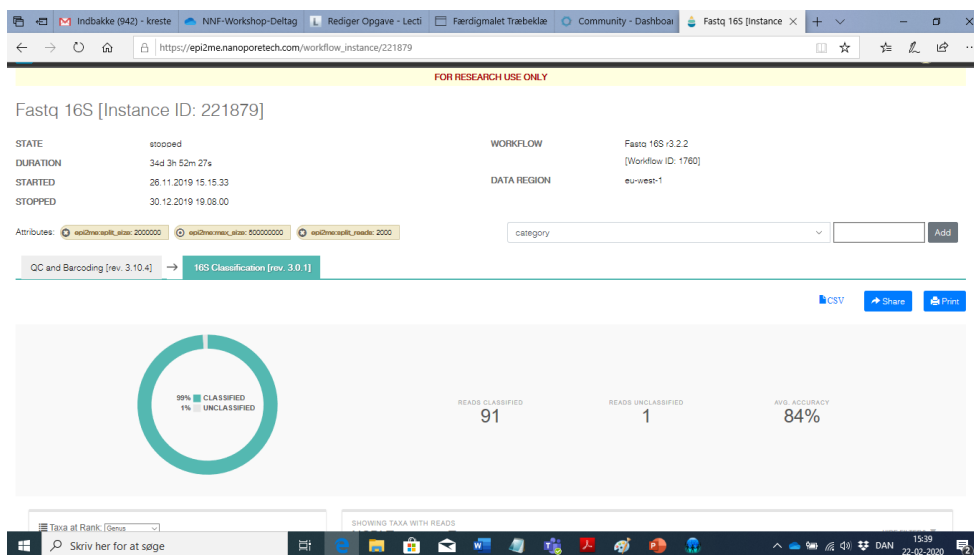


Figur 7.

Åben det workflow du vil arbejde med.

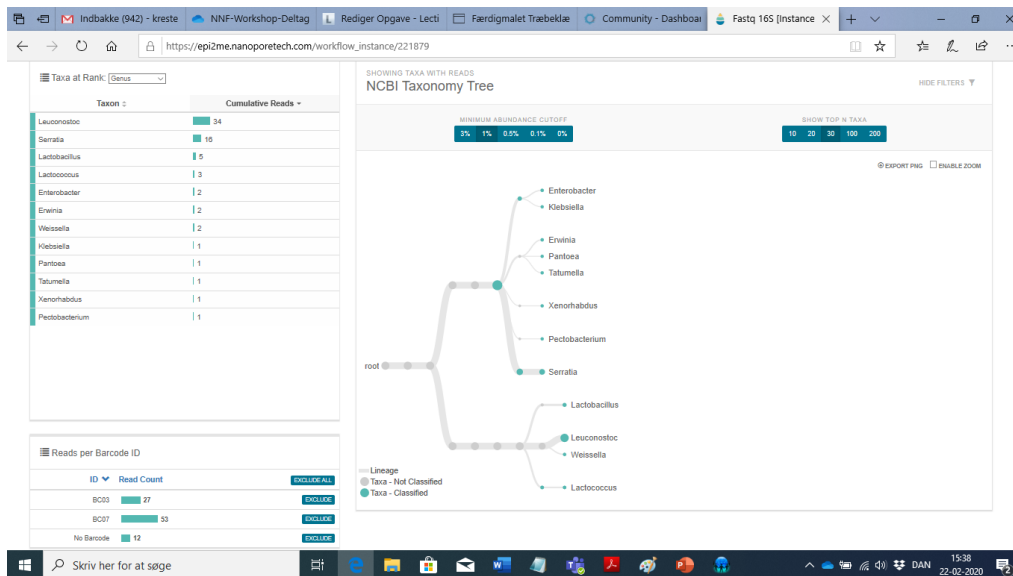
Tolkning af Workflow-rapporten i Epi2Me

Første side indeholder oplysninger om sekventeringens forløb, antal succesfulde og mislykkede reads og deres gennemsnitlige nøjagtighed i forhold til databasens sekvenser. Det er vigtigt at huske på at bakterier pga. deres ukønnede formering vil optræde med meget stor diversitet, og et match på 80-85 % med databaserne er derfor normalt.



Figur 8.

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



Figur 9.

Over artslisten kan man skifte mellem art, slægt osv. Matchet er forbundet med usikkerheder, og databaserne har deres begrænsning. Derfor vil der være usikkerhed ved en bestemmelse til artsniveau.

Tolkning af resultater

I venstre side udvælges den barcode, man vil arbejde med:



Figur 10.

7. Før cursoren hen over de identificerede bakterier og find det gennemsnitlige procentuelle match mellem sekvensen og databasen.
8. Angiv på baggrund af længden af DNA-sekvensen fra PCR-processen, hvor mange nucleotider der matcher databasens sekvenser med den viste procentuelle nøjagtighed.
9. Forklar hvad årsagen kan være til at der ikke er et nøjagtigt match med databasens sekvenser.

Opgaven 'Matematikken bag alignment' forklarer, hvordan man matcher sekvenser ved alignment.

Efterbehandling af data

- I elevvejledningen er der opgaver, hvor eleverne læser mere om de bakterier de har fundet. De vil ikke kunne finde fyldige oplysninger om alle de fundne bakterier. Det har imidlertid den vigtige pædagogiske pointe i sig, at der er et stort rum for forskning for den nysgerrige. Vi ved ufatteligt lidt om Jordens bakterieflora.
- Der er desuden supplerende opgaver, hvor de kan afprøve BLAST i en database. Det kan anbefales at lade eleverne klikke rundt, når de har foretaget en søgning her. Særligt kan man se på hvordan sekvenserne matcher, og identificere forskellige typer af mutationer.
- I de øvrige teoriopgaver (filerne 2a-) er der opgaver i bakteriernes formering og betydningen for identifikation, opgaver i hvorfor man identificerer ud fra 16S rRNA og opgaver som giver en matematik-vinkel på alignment og fylogenetiske træer.
- De fylogenetiske træer i Epi2Me kan suppleres med opgaver i hvordan disse konstrueres med andre eksempler, og har derfor tråde til et evolutionstema.