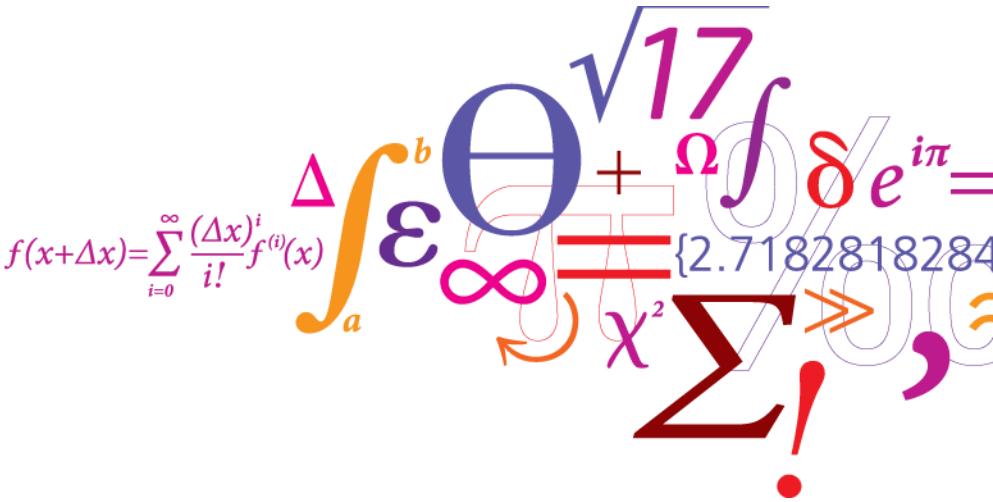
From Organisms to Genomics: From Alive to Live (cont'd)



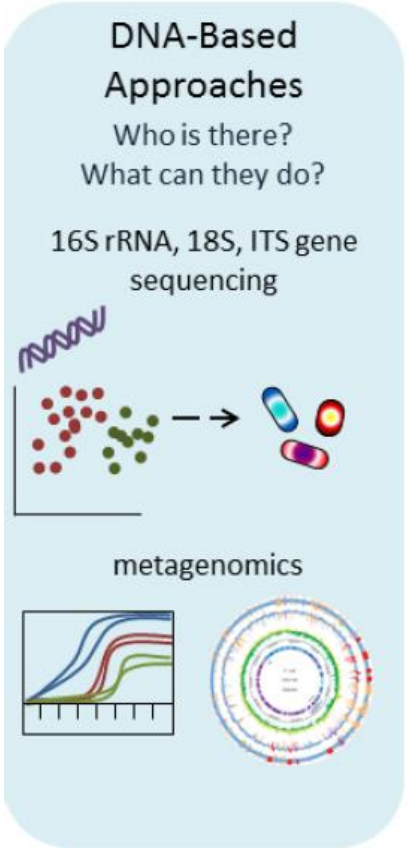
Saria Otani, MRes., MSc., PhD

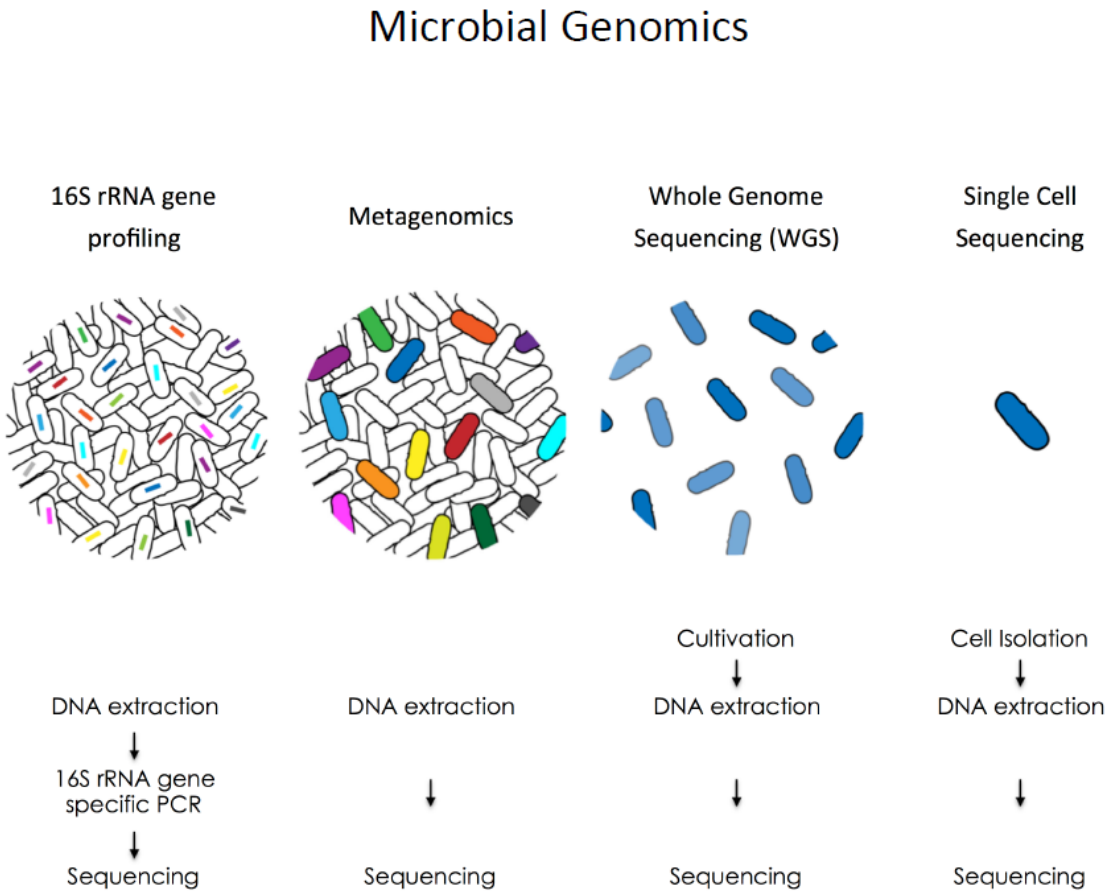
Applied Methods in Metagenomics - 23260
Denmark, 6/09/2022

DTU Food
National Food Institute

How do we detect/study a microbiome

Current approach



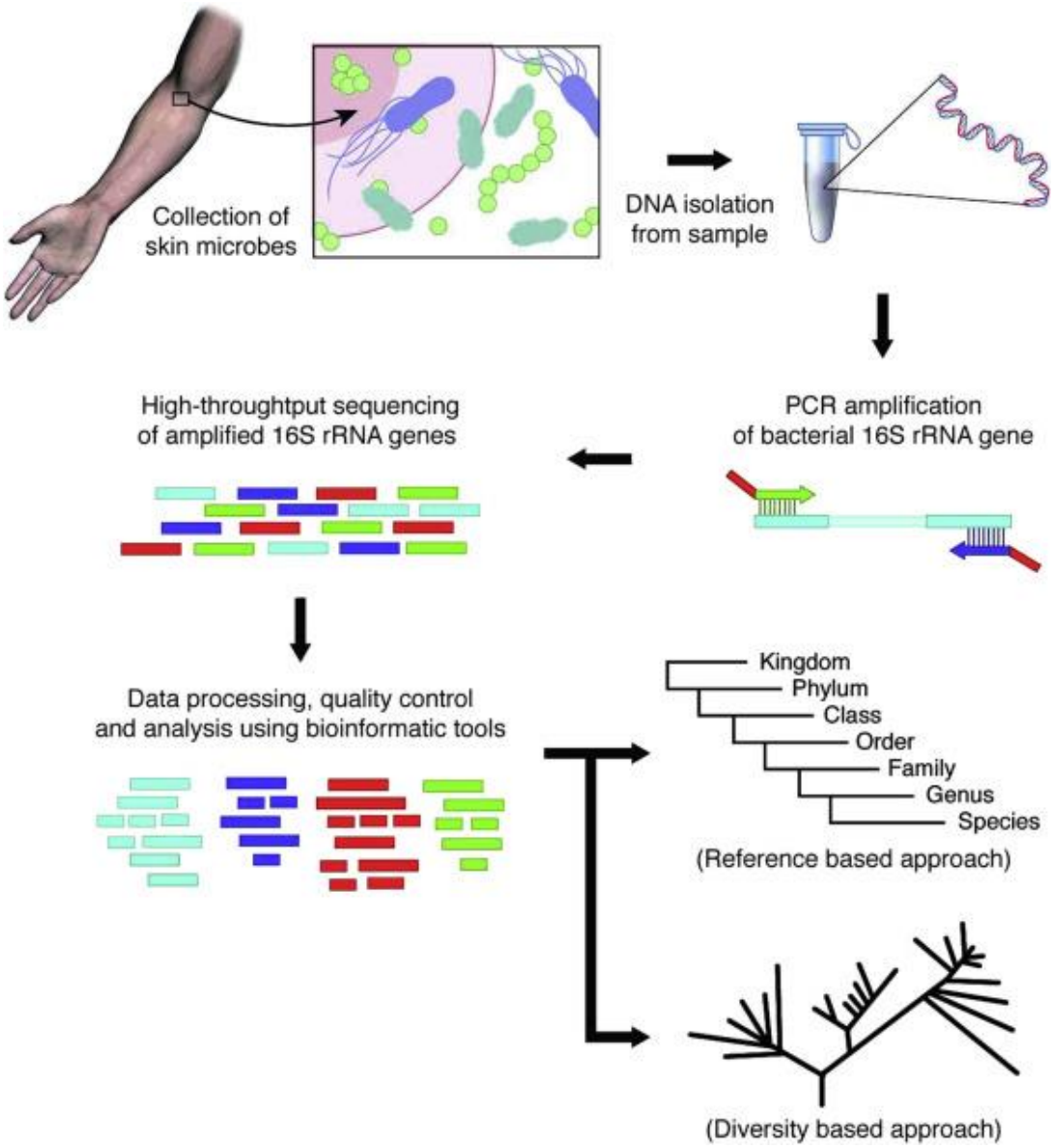


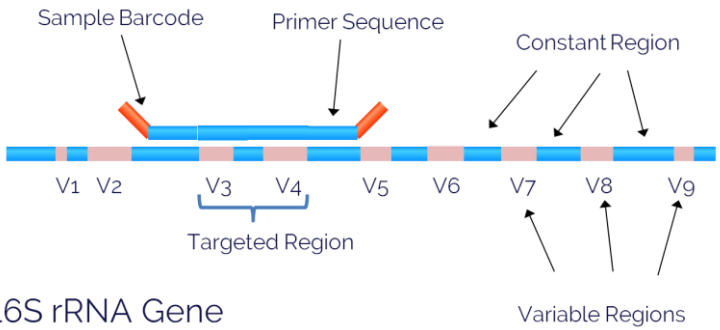
Environmental Chemicals, the Human Microbiome, and Health Risk: A Research Strategy, 2018; Pamp, S. 2017

How do we detect/study a microbiome

Previously on …

16S rRNA profiling





End product:

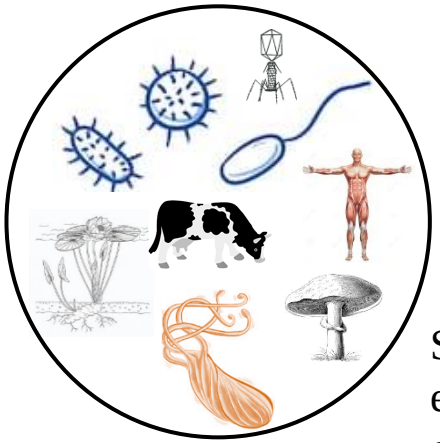
A profile of all bacterial taxa in the studied niche.

Composition, who is there!

How do we detect/study a microbiome

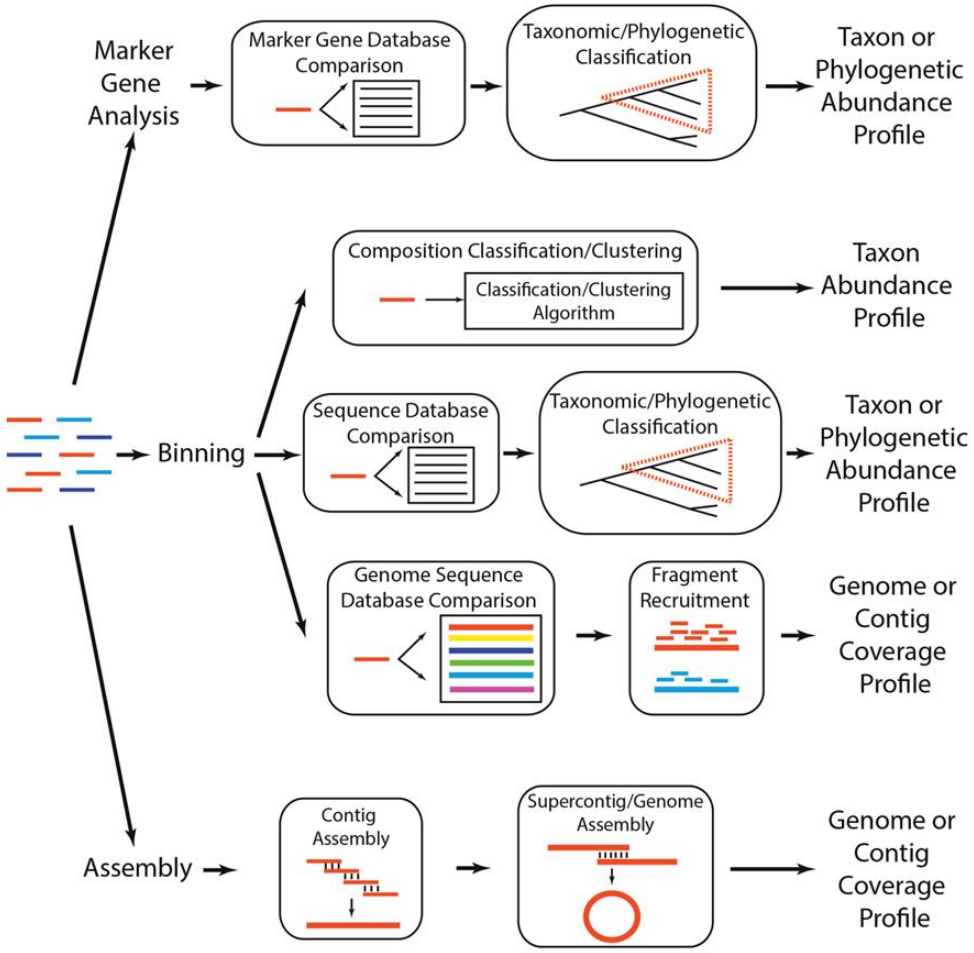
Metagenomics

sequencing-based analysis of the entire genomic content in an environmental sample.



Extract DNA

Sequence the entire DNA content



End product:

Information on the diversity and function of all organisms in the sample

Composition, who is there!

Roles, what do they do!

Analyse a metagenome:

Previously on …

Table 1 from https://learn.inside.dtu.dk/d2l/le/content/126041/Home

These are real 5 microbiomes.

Can you tell from this if the host of the microbiome is a chimpanzee?

Is table 1 a 16S rRNA pyrosequencing output?

What questions can you answer (what are you interested in)? (5-8 minutes)

What is your answer?

What is the most abundant bacterium (with assigned taxon)? Is it pathogenic?
Can you calculate the taxonomic assignment success? At each level.

Table 2 from https://learn.inside.dtu.dk/d2l/le/content/126041/Home

Is Table 2 metagenomics output? Why?
Can you tell from this if there is a chimpanzee host amongst the samples?

Table 3 from https://learn.inside.dtu.dk/d2l/le/content/126041/Home

What is it? Is it metagenomic output?

A project workflow in metagenomics

Plan ahead: what exactly are you looking for? design the experiment accordingly!

Sample.

DNA extraction.

Library preparation.

Sequencing platforms.

Analyses: *e.g.*, diagnostics

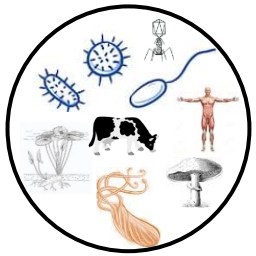
4

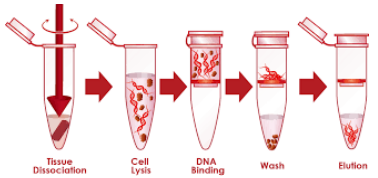
3

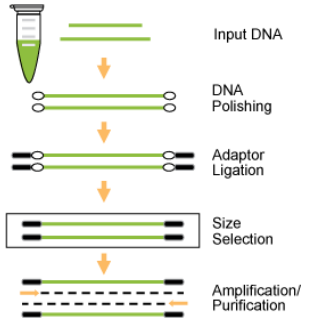
2

1

Plan it this way



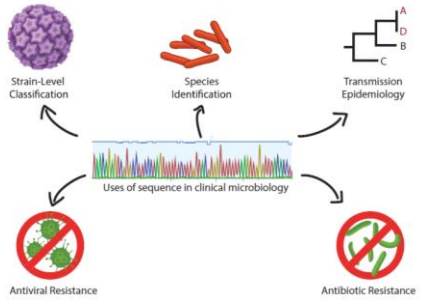




illumina®

PACBIO®

Oxford NANOPORE Technologies



Genomic DNA extraction – kit example

Starting material

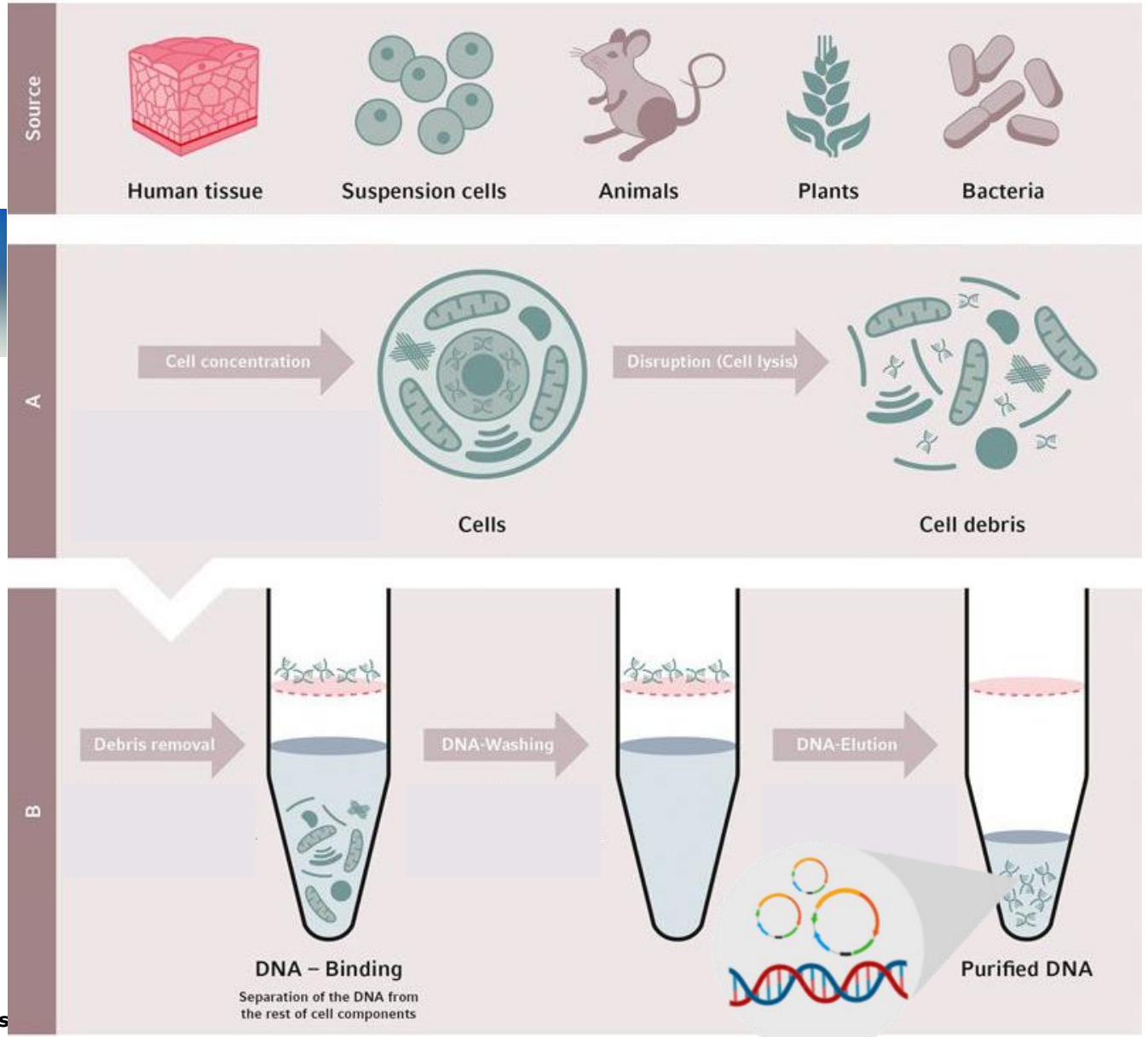
Cell lysis

Left after the lysis:

DNA, RNA, protein

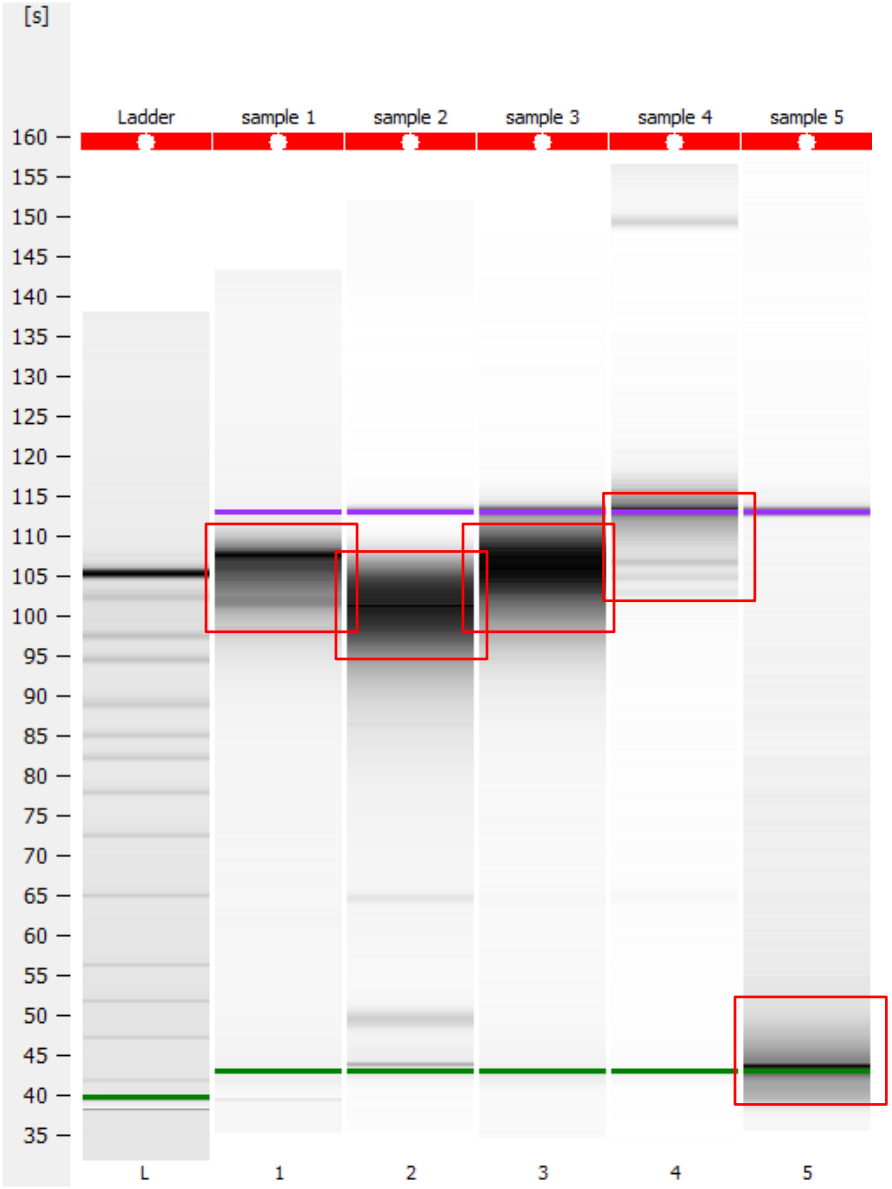
Precipitation

Clean up



Modified from Printerest.com

Previously on …



High molecular weight
(large size DNA)

low molecular weight
(small size DNA)

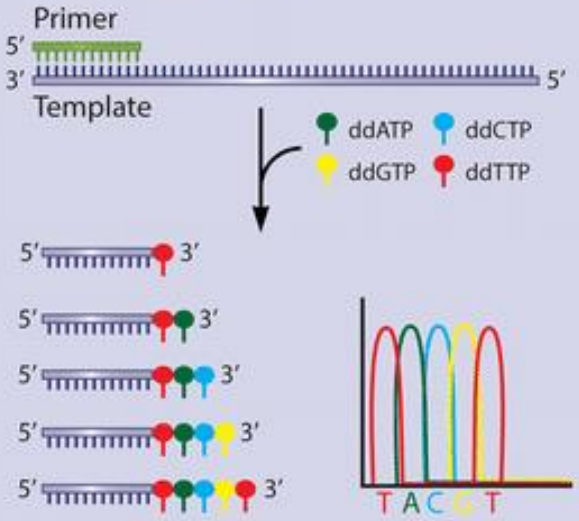
Illumina Sequencing

This is commonly known as NGS

Previously on …

DTU

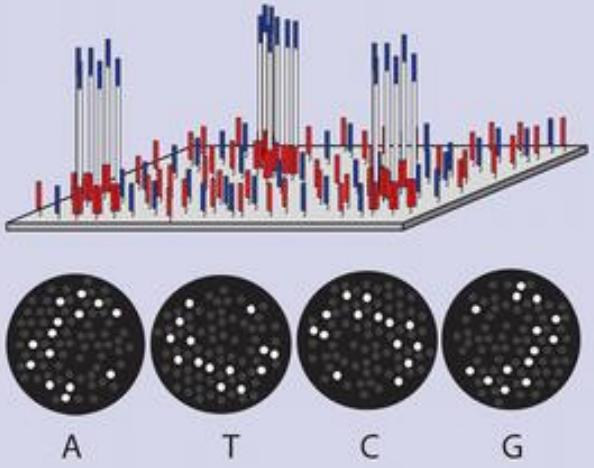
First Generation
Shotgun Sequencing



- Sequencing by synthesis
- High accuracy
- Long read lengths
- Relatively small amount of data generated

e.g., ABI capillary sequencer (ABI)

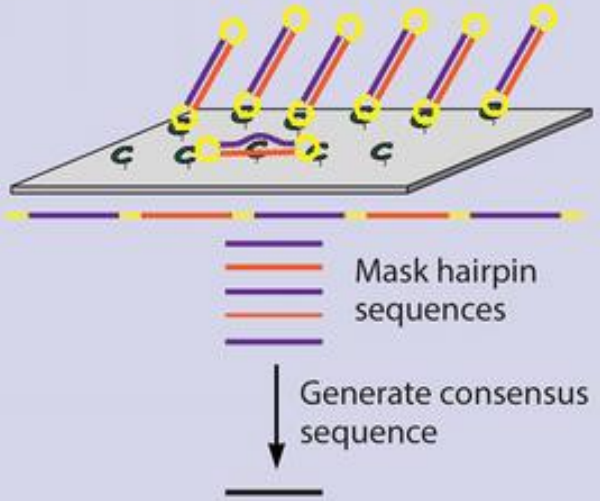
Second Generation
Massively Parallel Sequencing



- Sequencing by synthesis
- Amplified templates are generated during sequencing, reducing the requirements for starting material
- High accuracy
- Short read lengths

e.g., MiSeq (Illumina), Ion Torrent (Thermo Fisher Scientific)

Third Generation
Single-molecule Sequencing



- Single-molecule templates
- Low accuracy
- Long read lengths

e.g., Single-Molecule Real-Time (SMRT) — Sequencing (Pacific Biosciences), MinION (Oxford Nanopore Technologies)

Sequencing technology selection

- **Short read technologies**

– Illumina (MiSeq, NextSeq, HiSeq, NovaSeq…)

– Ion Torrent

- **Long read technologies**

– Pacific Biosciences (PacBio)

– Oxford Nanopore Technologies (MinION)

Previously on …

DNA library preparation MinIon:

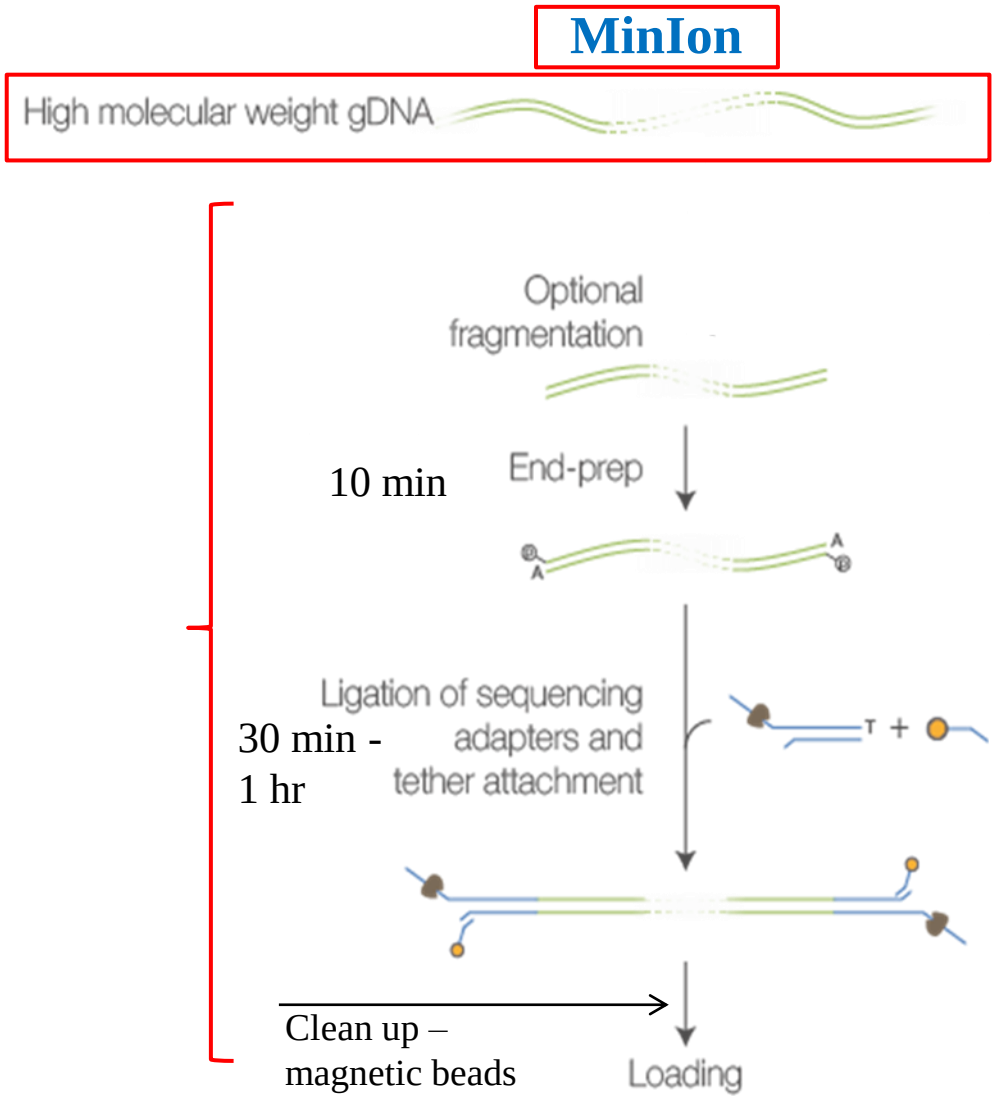
Same steps, yet different ingredients and handling!

Keep in the fridge.

Wide bore tips whenever handling the DNA.

Flick tubes, no vortex.

Increase all incubation times in lib. prep. protocol.



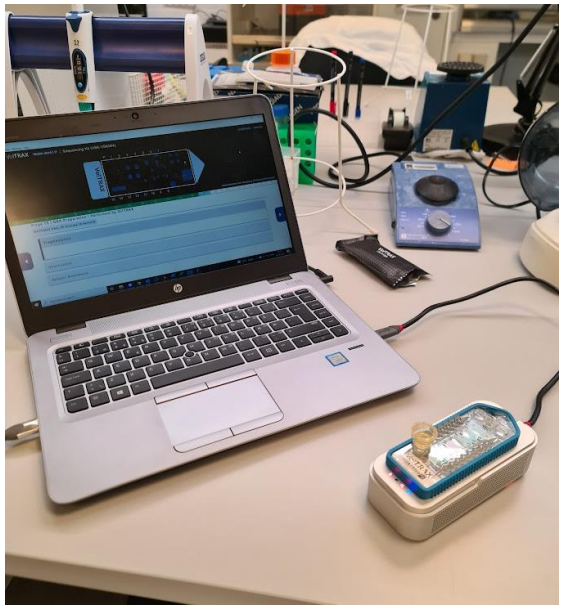
Previously on …

DNA library preparation head ups - Illumina:

Low sequencing output:
low DNA input - library prep failure.

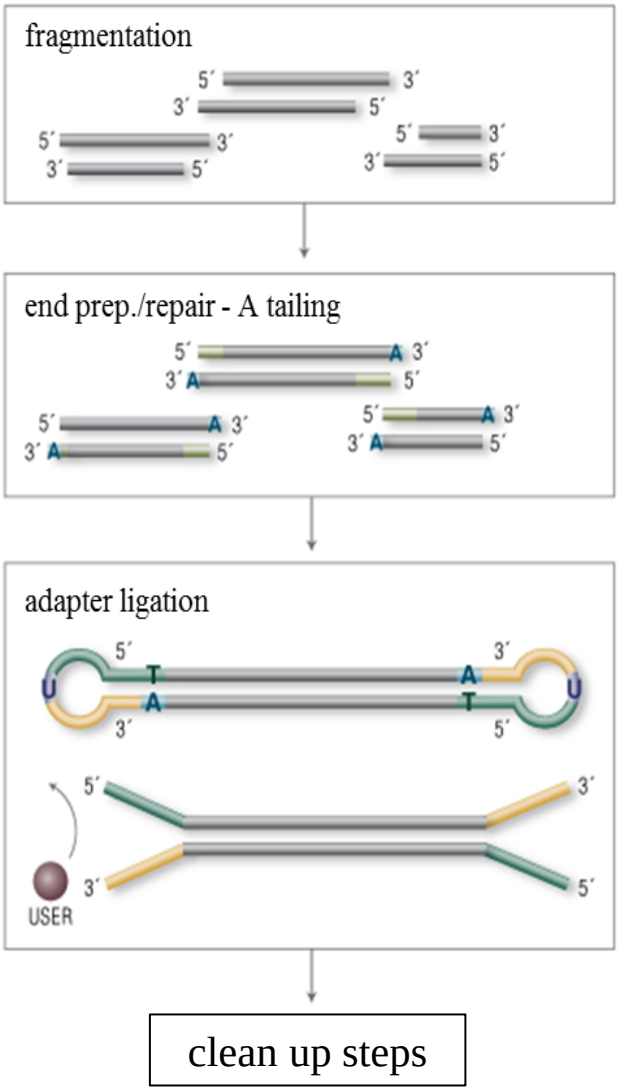
Failed clustering:

Failed library prep.



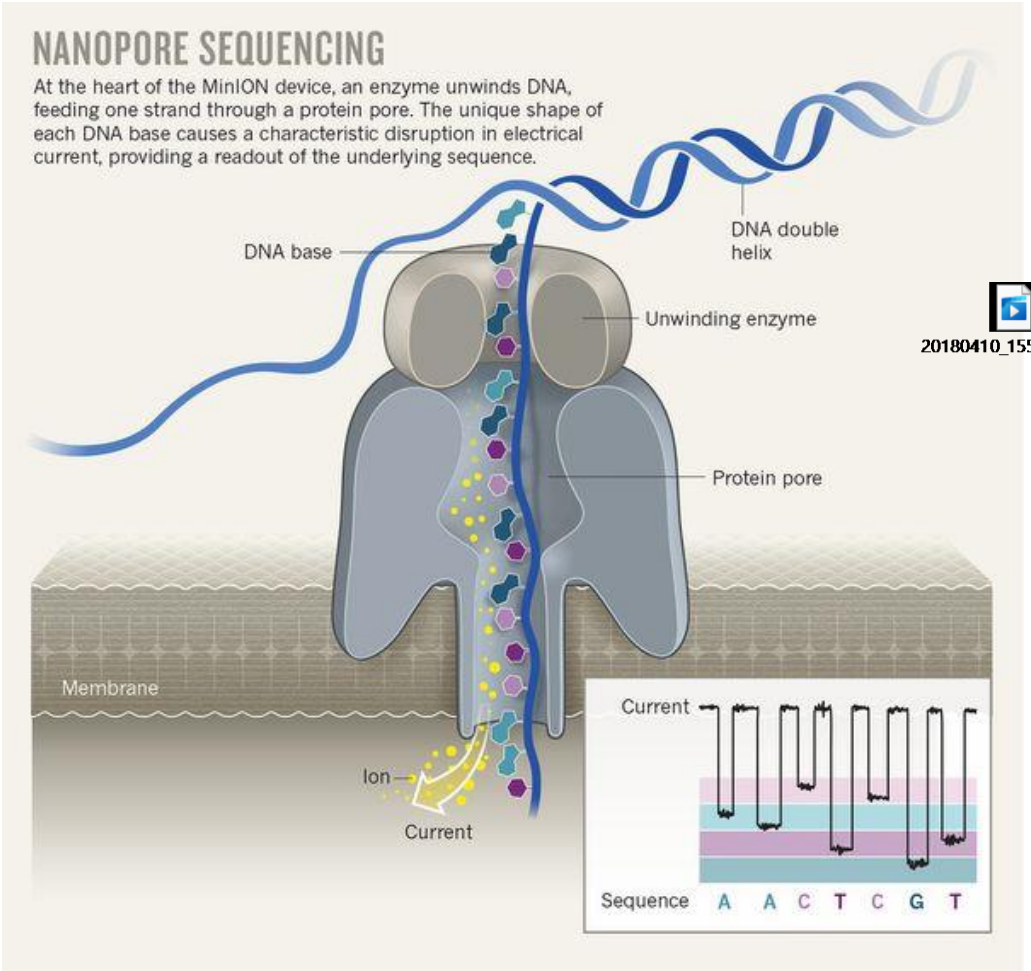
Library prep machines

VolTRAX

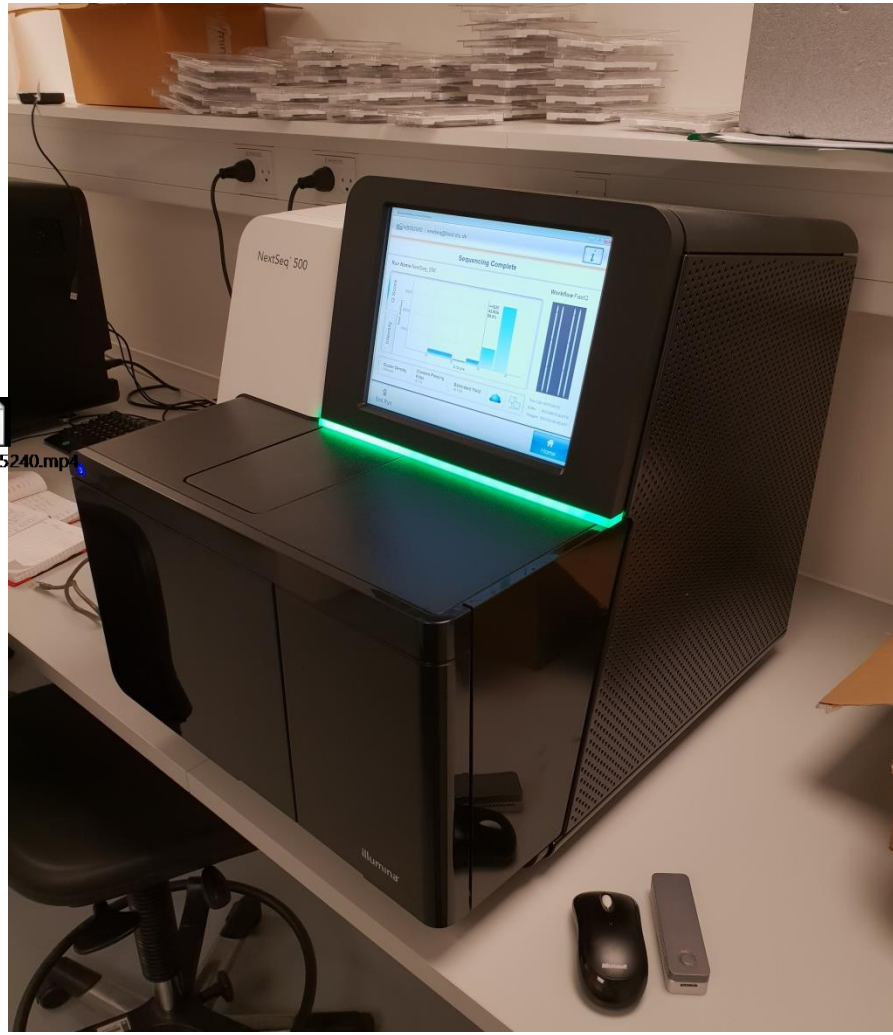


Oxford Nanopore Technologies (MinION):

Long and ultra-long reads – Best high throughput available (100 Gbp) – Portable, fast and real-time sequencing.



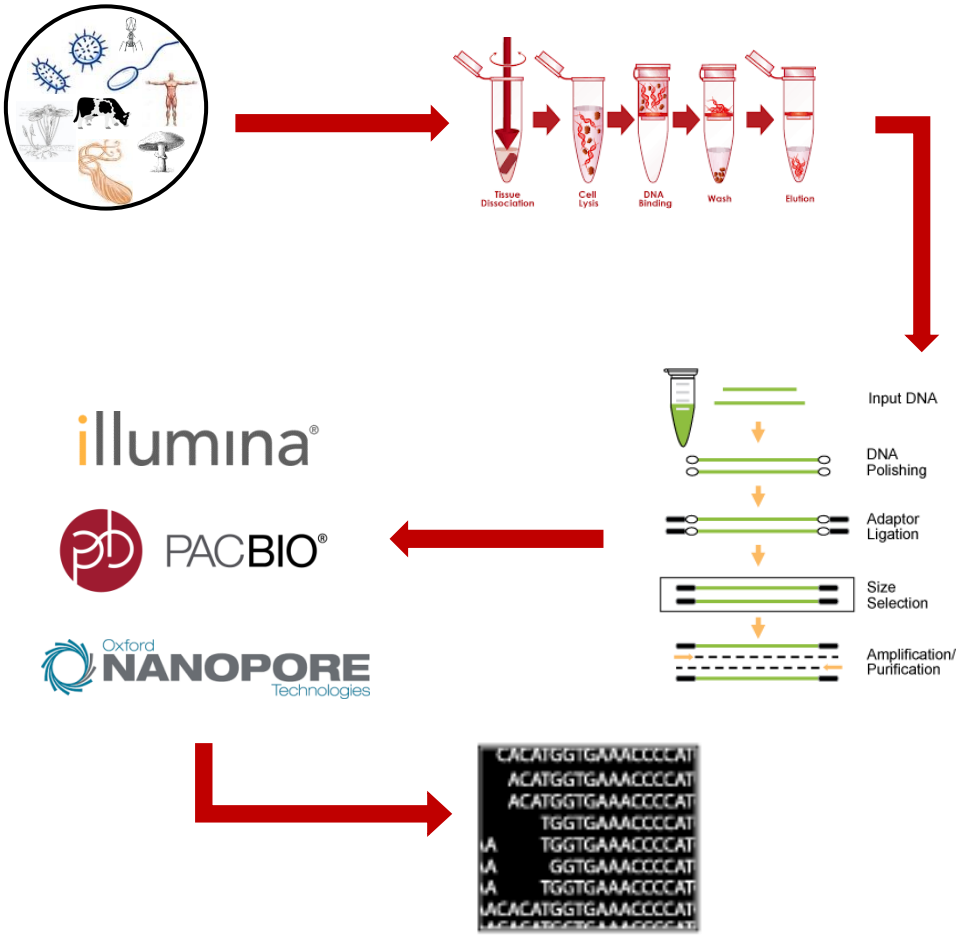
20180410_155240.mp4



1. The workflow from alive material to DNA:

-Starting materials and sample selection

-DNA extraction and challenges



2. DNA preparation for sequencing:

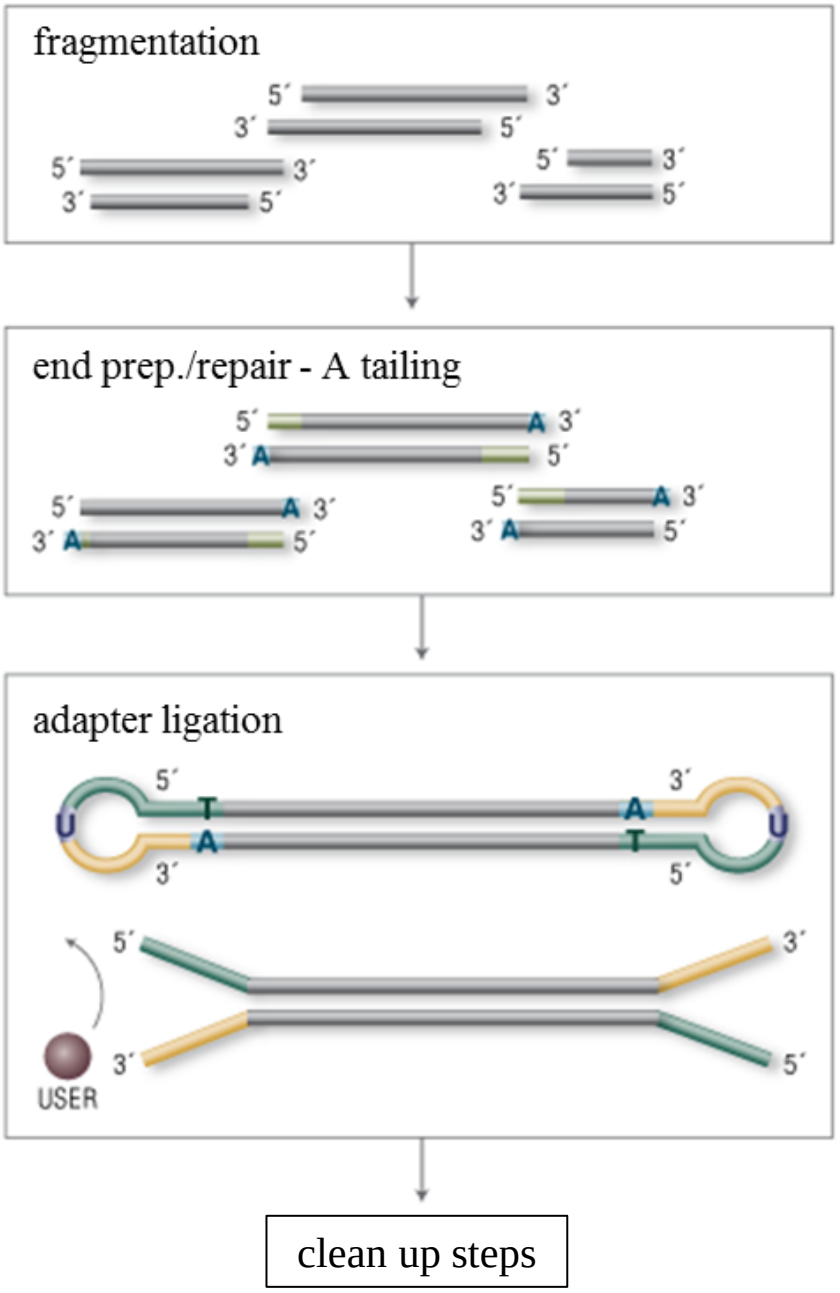
-DNA library preparation

-Sequencing technology selection

DNA library preparation:

Collection of steps to prepare DNA to be read by a sequencing machine.

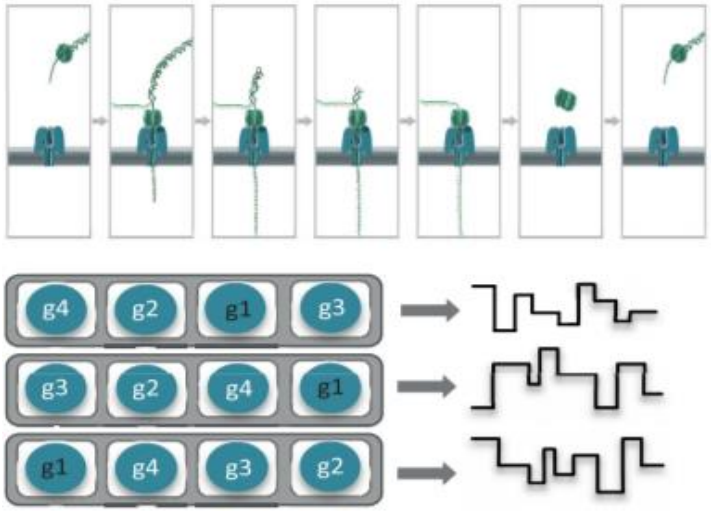
It is the same for all platforms.

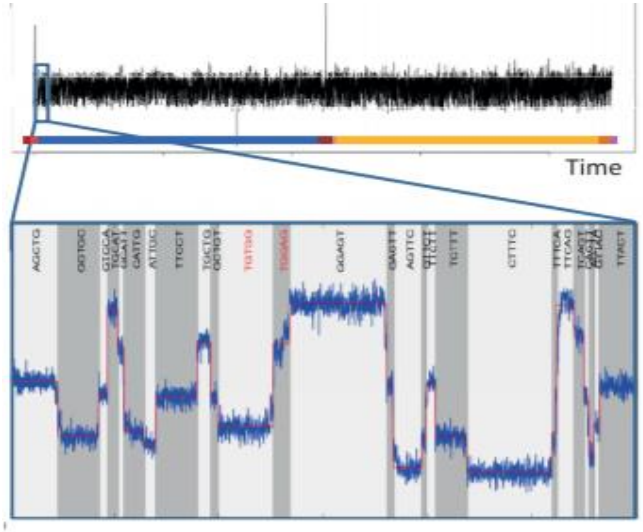


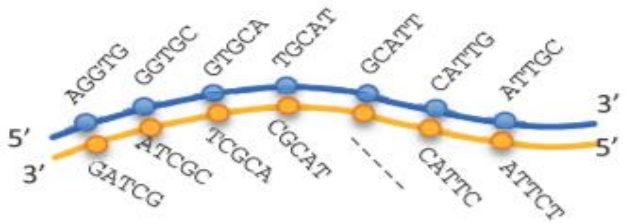
Modified from BioLine and NEB

Oxford Nanopore - MinKNOW

Generates FAST5 files.



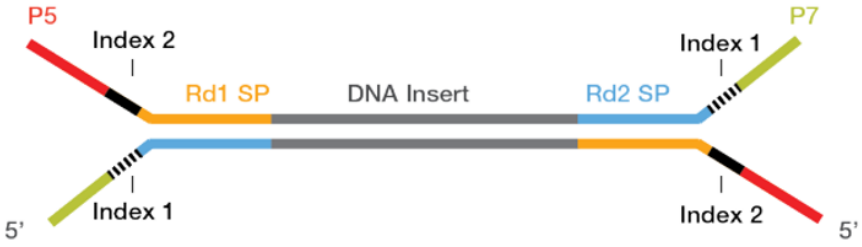


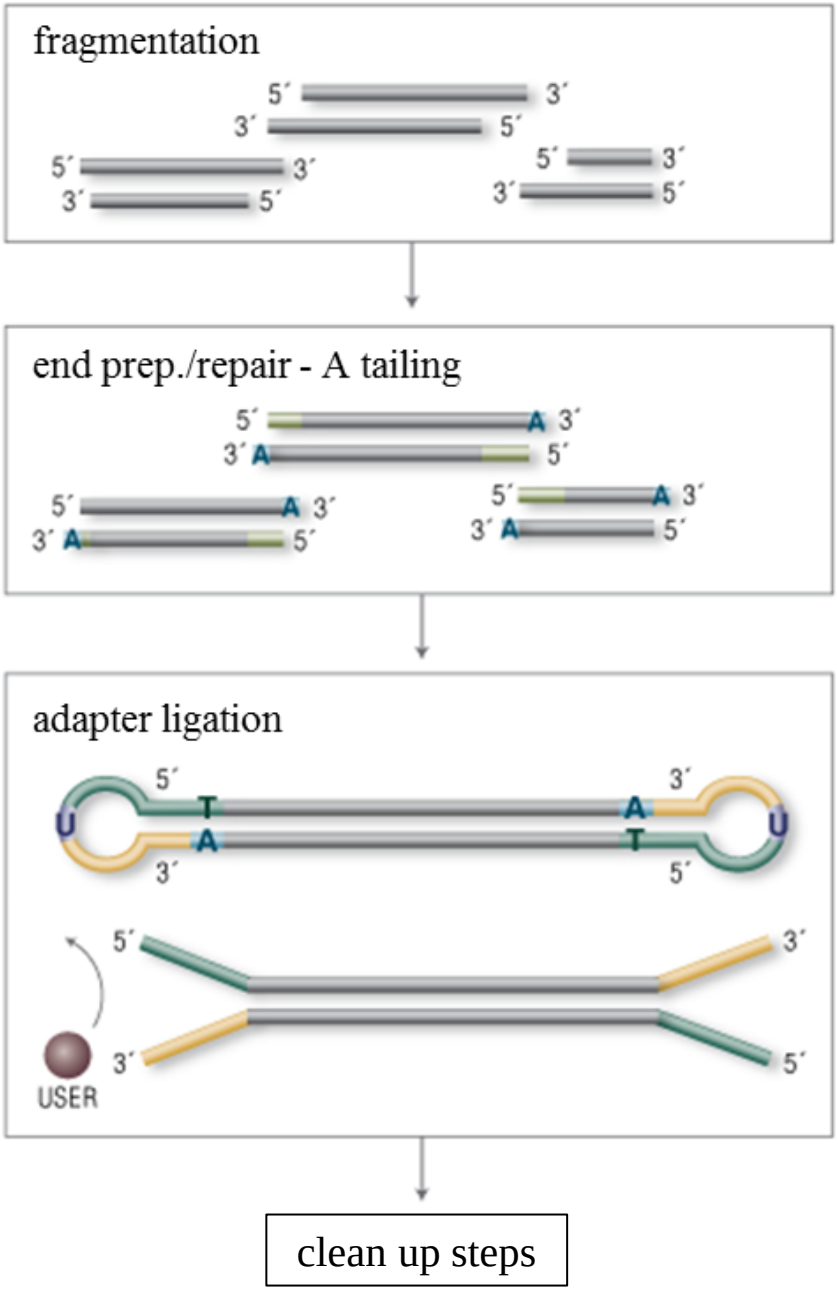



```
5′-GTGCATT-G-CTTGAGTTCTTT-CAGTT-A-3′
   |||| || | |||||||||||| ||||| |
5′-CTGCATTCTCCGTGAATTCTTTCGAATTCT-3′
   |||| || | |||||||||||| ||||| |
5′-TCGC-TT-TCCGCGAATTCCTT-GAATT-C-3′
```

DNA library preparation:

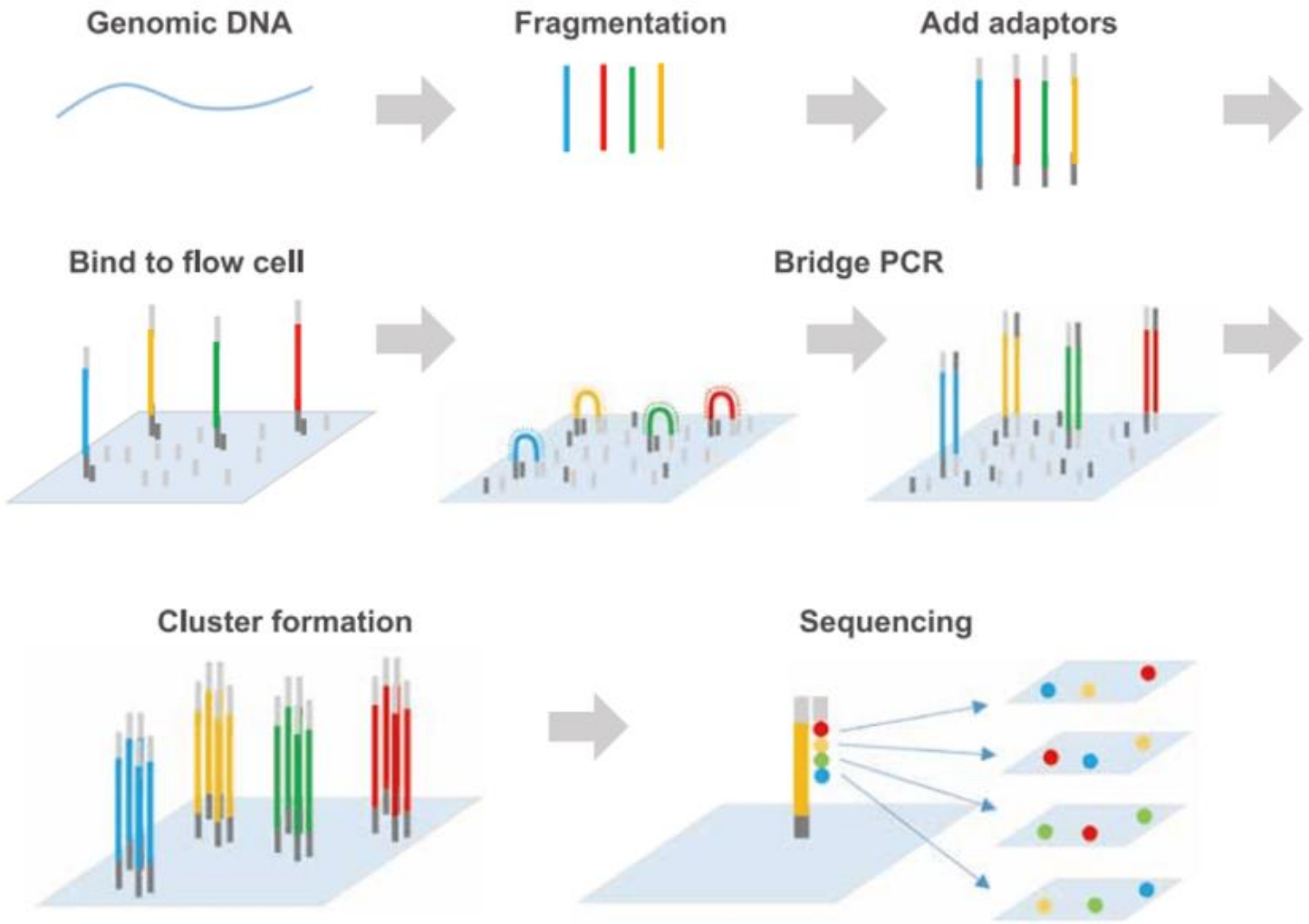
End library product



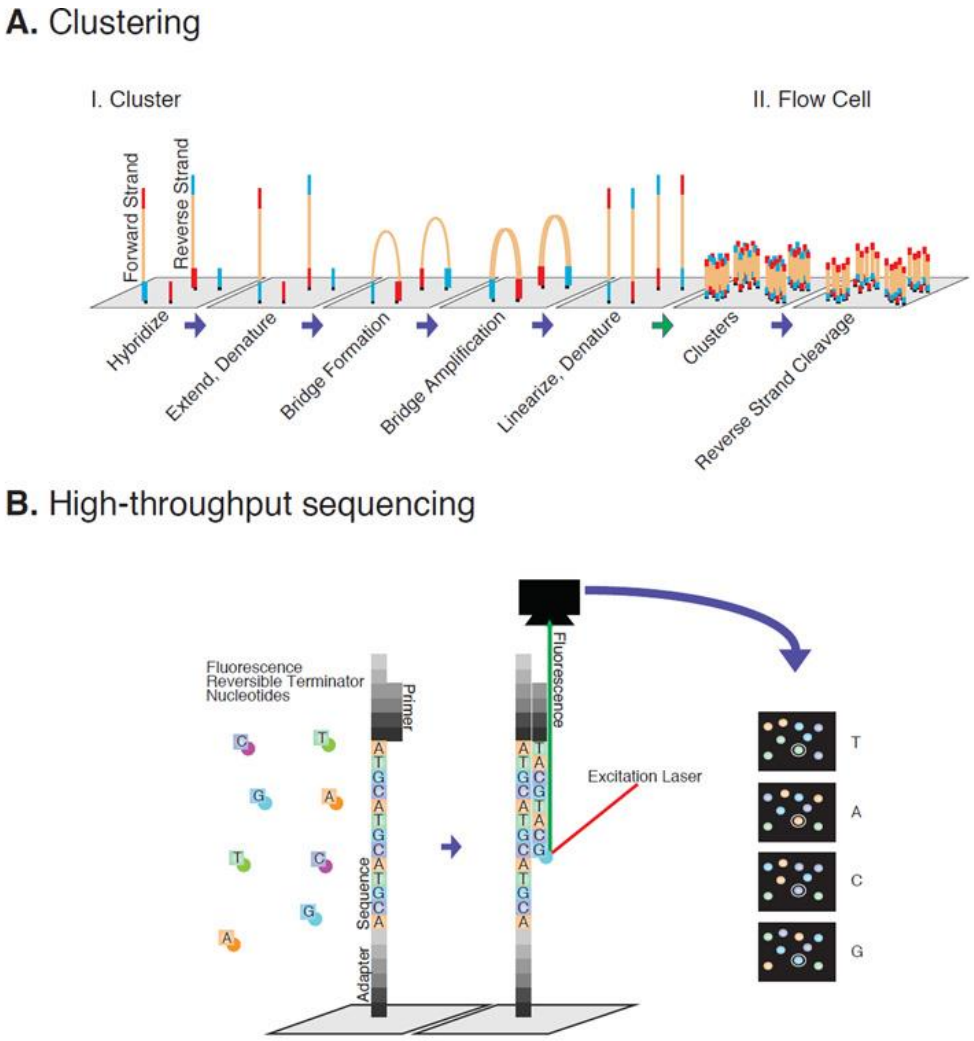


Modified from BioLine and NEB

Illumina Sequencing

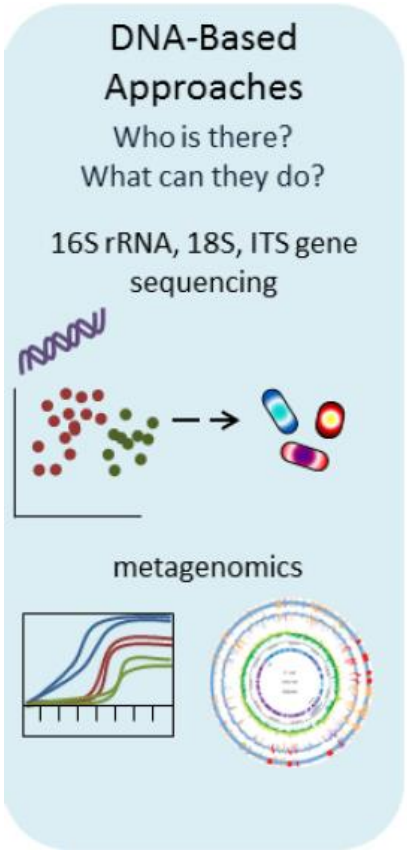


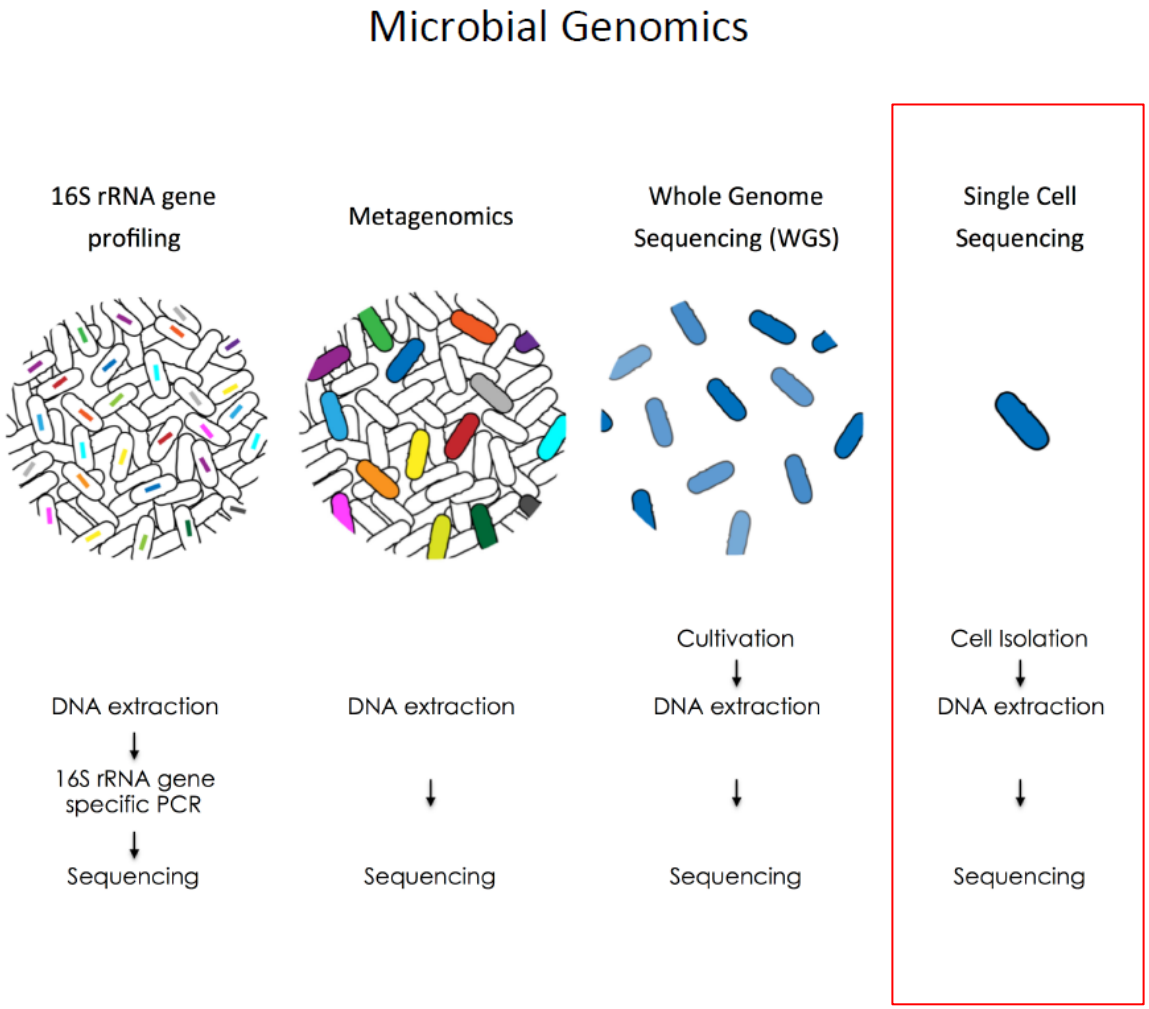
Illumina Sequencing



How do we detect/study a microbiome

Current approach





Environmental Chemicals, the Human Microbiome, and Health Risk: A Research Strategy, 2018; Pamp, S. 2017

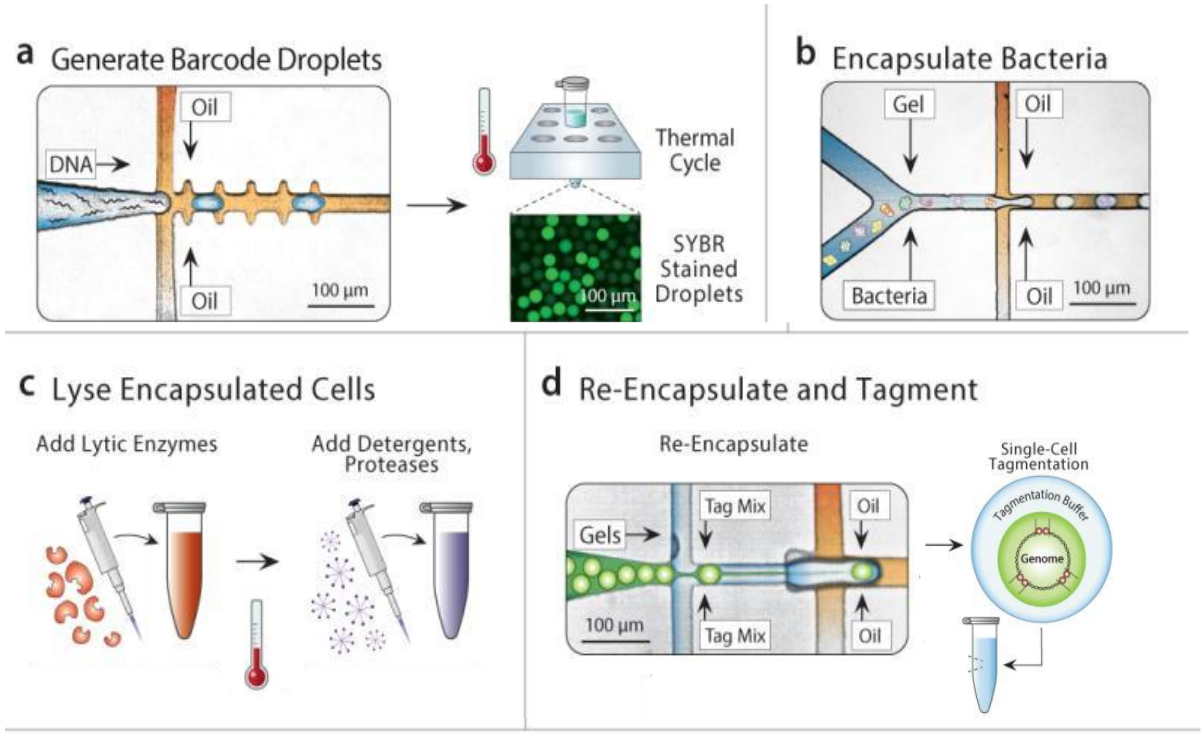
Single cell sequencing:

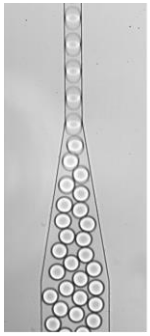
Microfluidic.

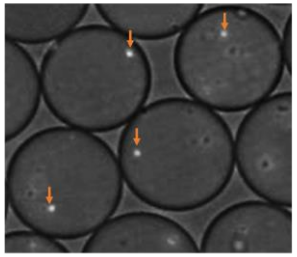
Established in Illumina.

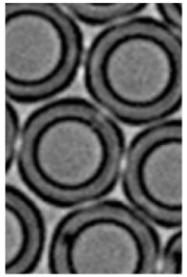
Enables targeted sequencing of genetic content within one cells.

Traceability!





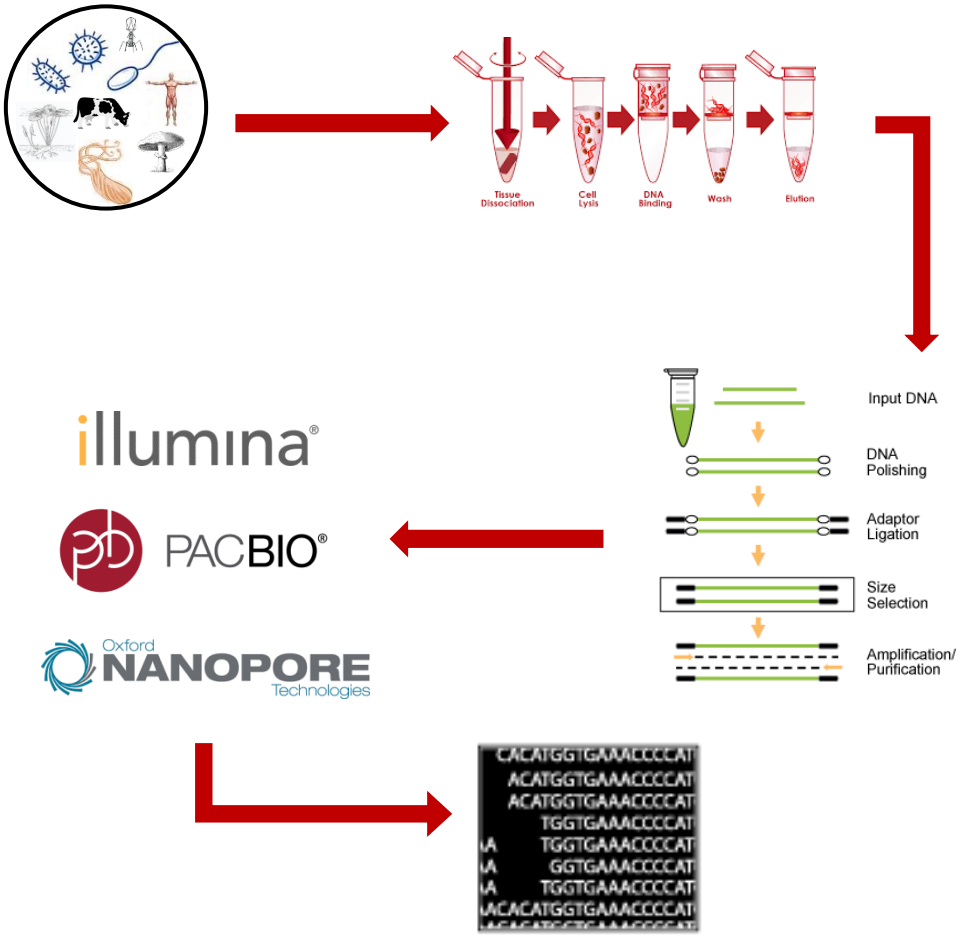




1. The workflow from alive material to DNA:

-Starting materials and sample selection

-DNA extraction and challenges



2. DNA preparation for sequencing:

-DNA library preparation

-Sequencing technology selection

Illumina *vs.* Oxford Nanopore

– Illumina (MiSeq, NextSeq, HiSeq, NovaSeq…):

Average 300 bp reads - short

Good accuracy

Error rate ~0.1%

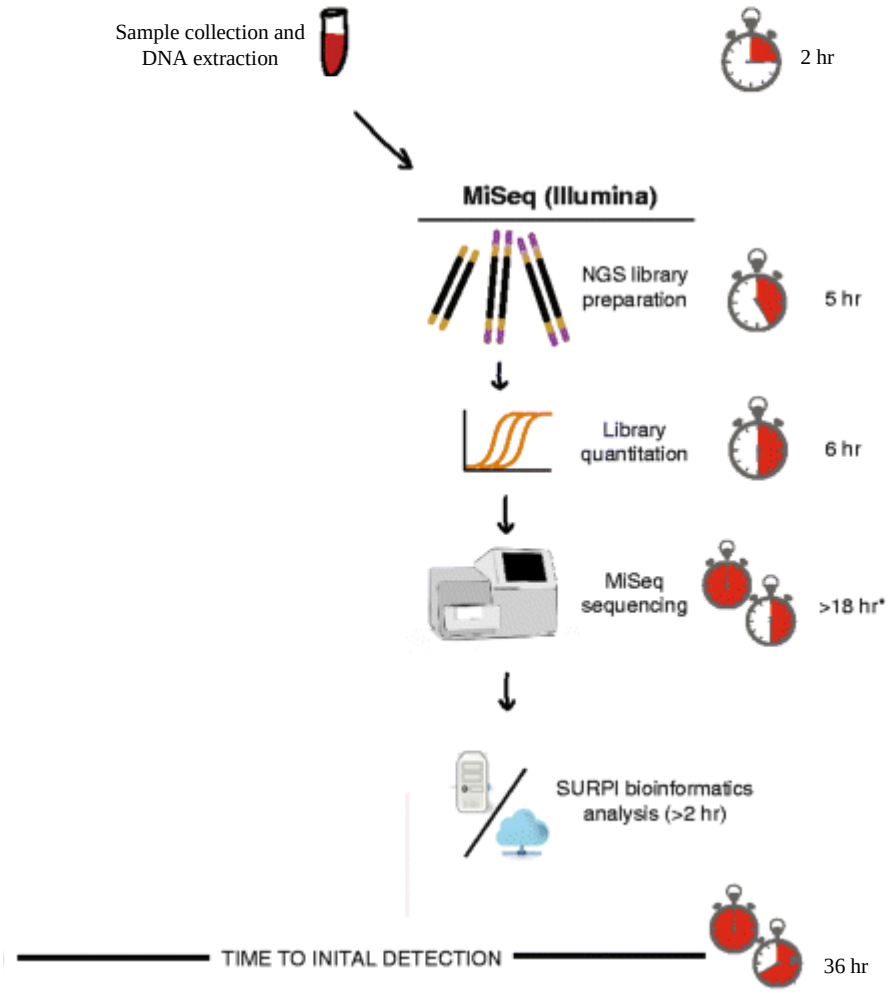
– Oxford Nanopore Technologies (MinION):

Very long reads (up to 900 kb.)

Fast turnaround time (down to 2 hrs)

Portable and real-time sequencing

Large error rates (<1% - 2022)

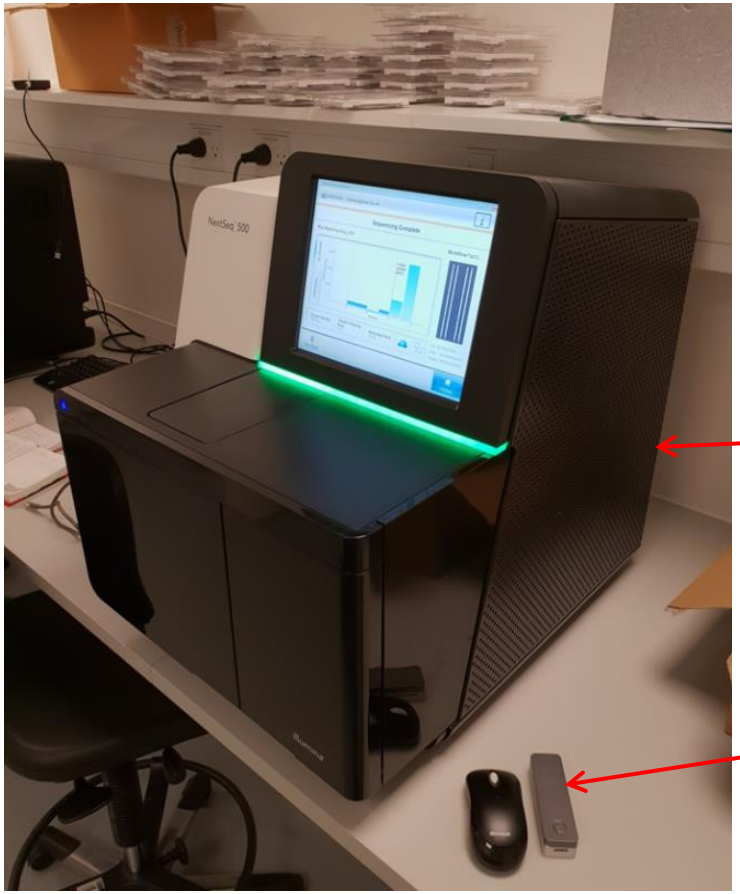


(Greninger., *et al*., 2015)

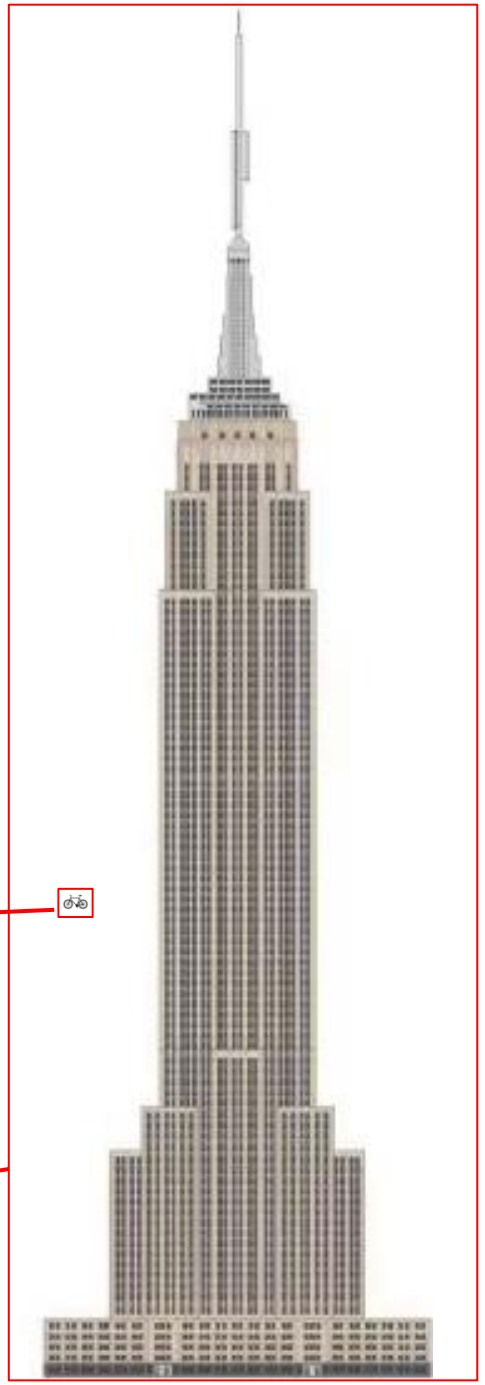
Illumina *vs.* Oxford Nanopore: DNA input

Illumina: DNA concentration ≃ 1 ng

Nanopore: DNA concentration > 1000 ng

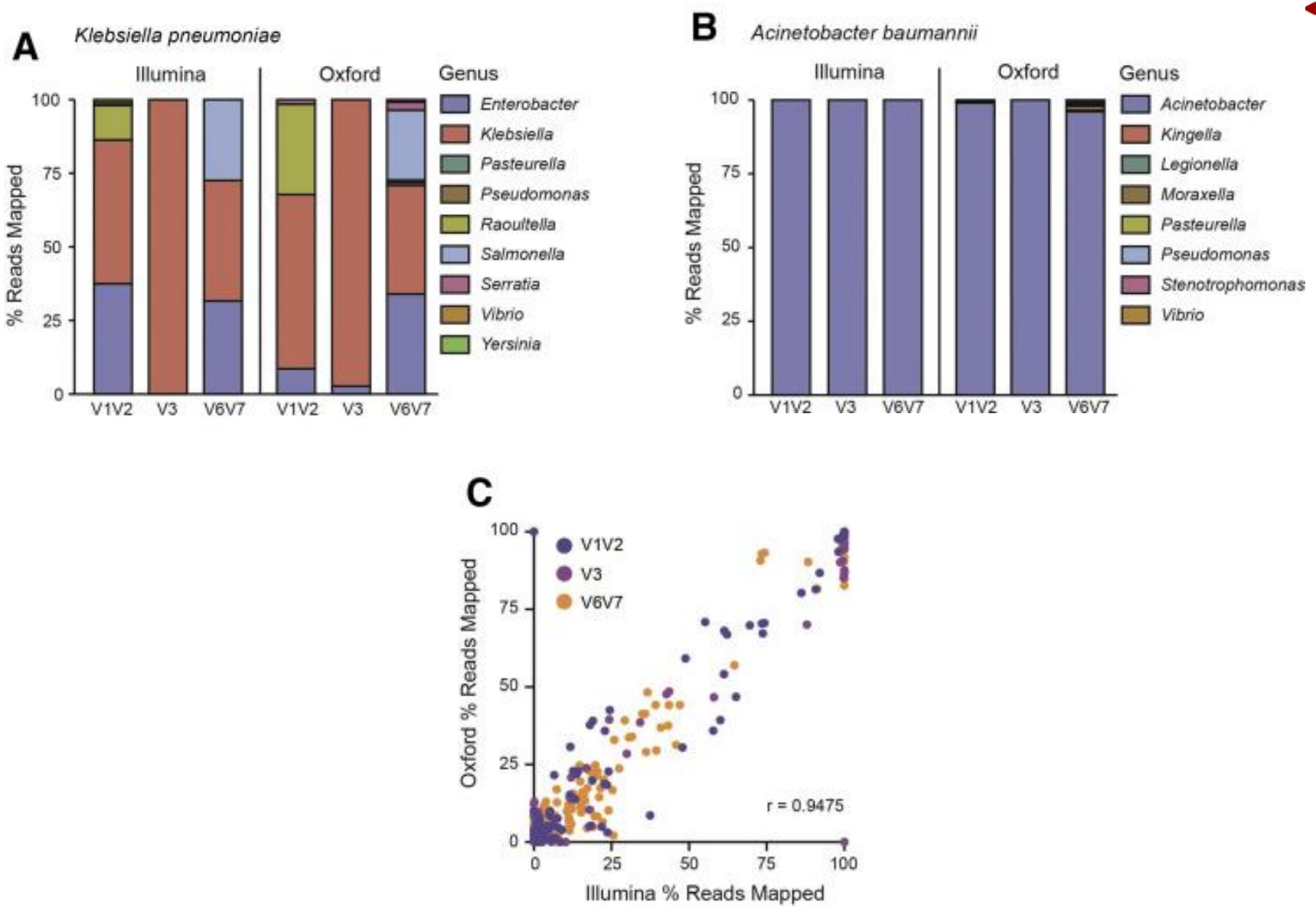


Library preparation



Illumina vs. Oxford Nanopore

Blah blah blah



(Stefan., *et al*., 2022)

Example

What are the plasmidome dynamics in environmental samples like soil and sewage water?

a. What is the first thing to do?

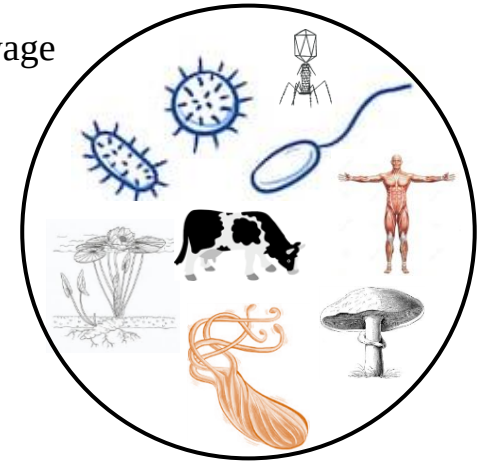
b. Which DNA should be extracted?

c. Which sequencing technology do you recommend?

Global sewage plasmidome characterisation (Nanopore)

Plasmidome dynamics in 24 sewage samples from around the globe

Sewage



High quality DNA from sewage is problematic due to sample nature:

-Low DNA content

-Mixed microbial community with extra amount of **contaminants** (block DNA extractions)

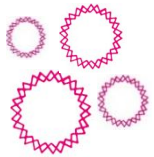
Global sewage plasmidome characterisation (Nanopore)

Pre-sequencing treatment:

Sewage

Several plasmid extraction protocols were tested

All were sequenced with Nanopore



Plasmid DNA

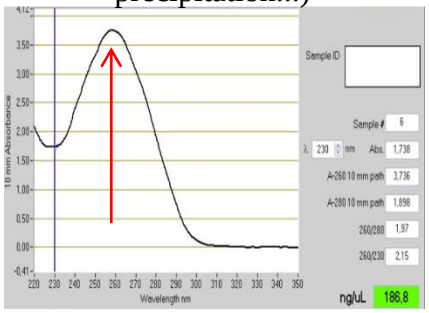
Global sewage plasmidome characterisation

Input QC: DNA quality, purity and length

Ensure good quality DNA input

Goes to sequencing

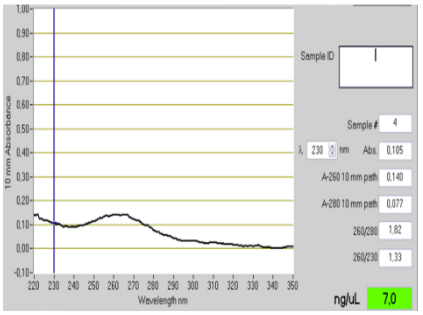
DNA modified protocol (beading, longer lysis, longer protein precipitation...)



✓ concentration

✓ DNA purity

DNA standard protocol



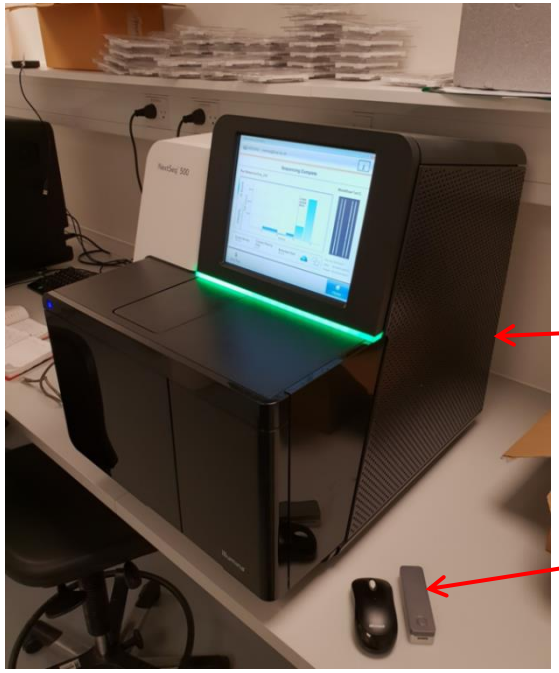
Global sewage plasmidome characterisation

Comparison example, differences in DNA input between 2 sequencing technologies:

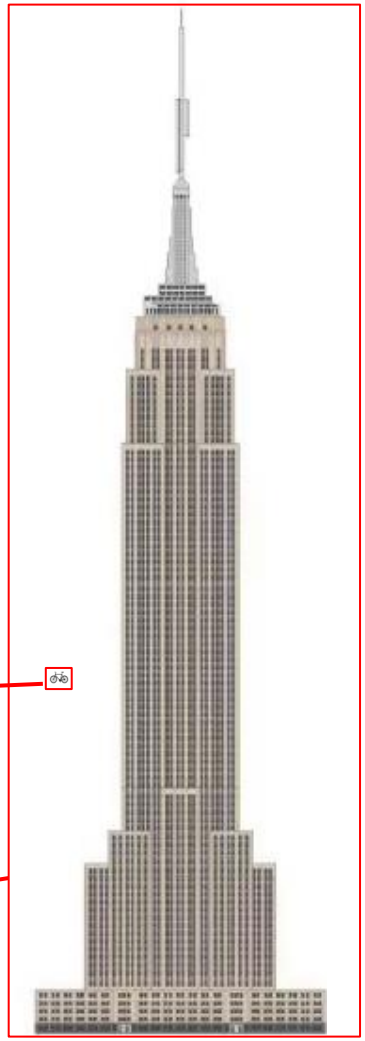
Illumina DNA concentration ≃ 1 ng

:

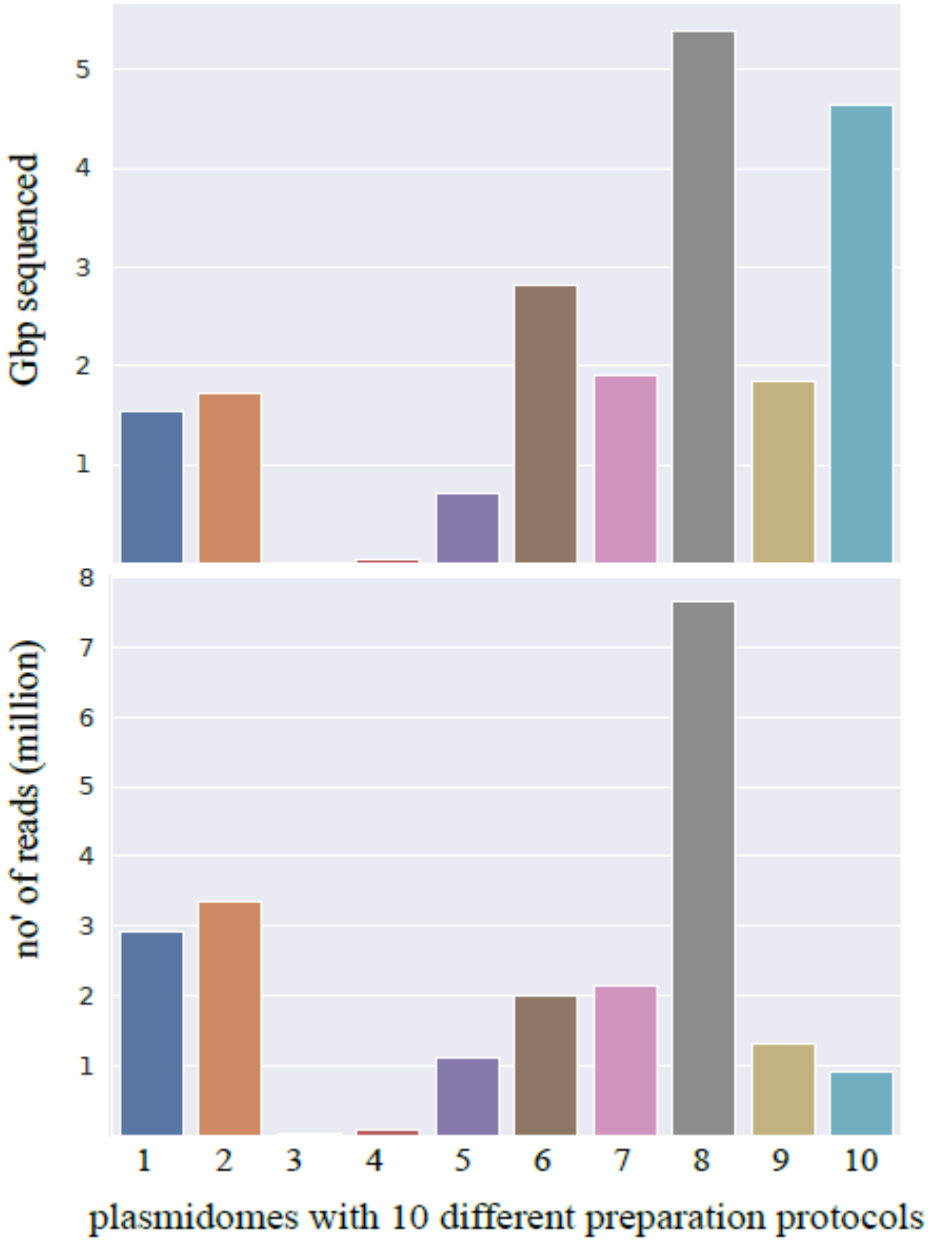
Nanopore: DNA concentration > 1000 ng



Library preparation



Plasmidome characterisation



Comparing plasmidome

Nanopore throughputs

Plasmidome dynamics in 24 sewage samples from around the globe

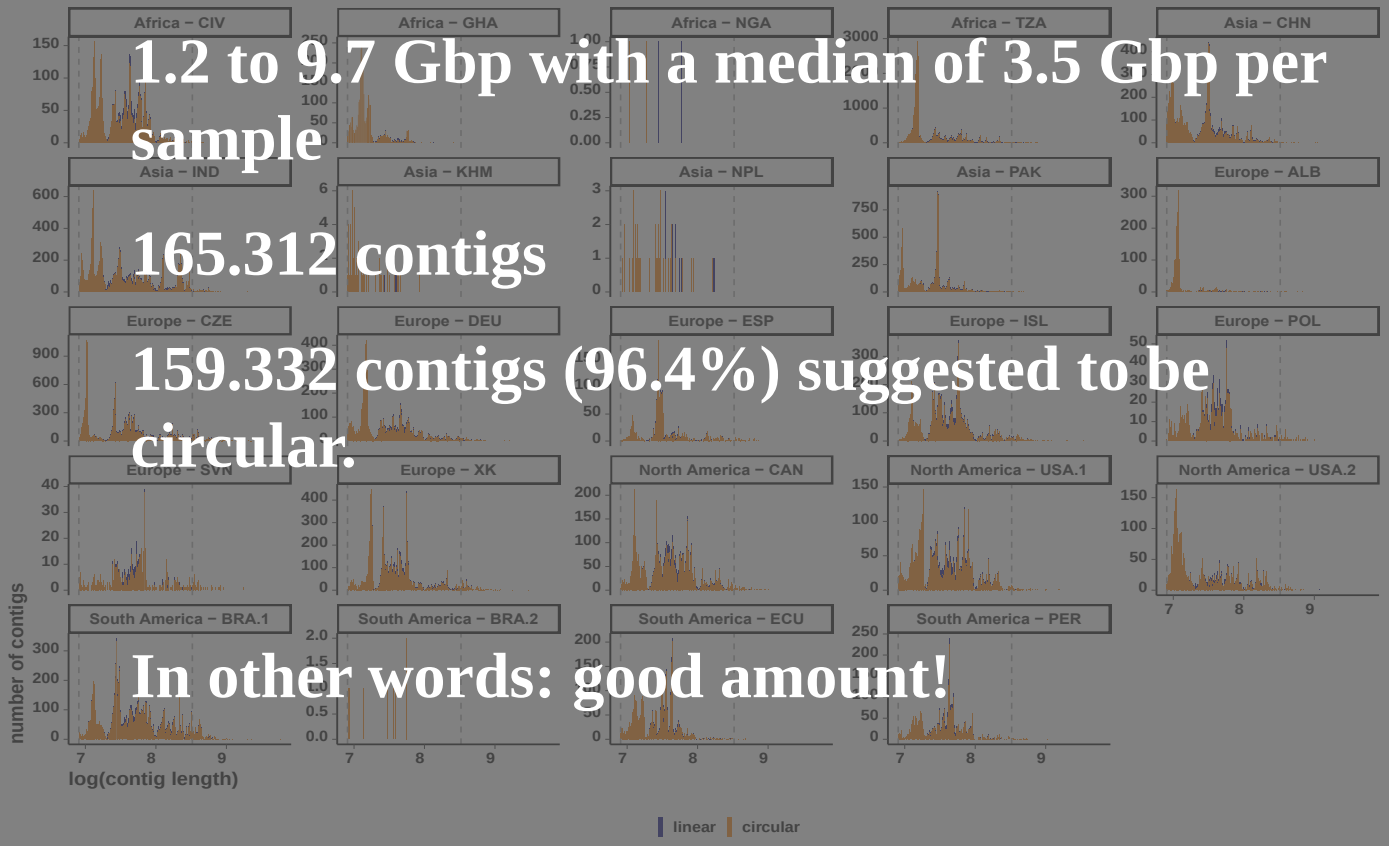
Which protocol is the best at exploiting long read seq.?

slido

Which one is the best output to exploit long read sequencing techn. the best way

ⓘ Start presenting to display the poll results on this slide.

Global sewage plasmidome characterisation



1.2 to 9.7 Gbp with a median of 3.5 Gbp per sample

165.312 contigs

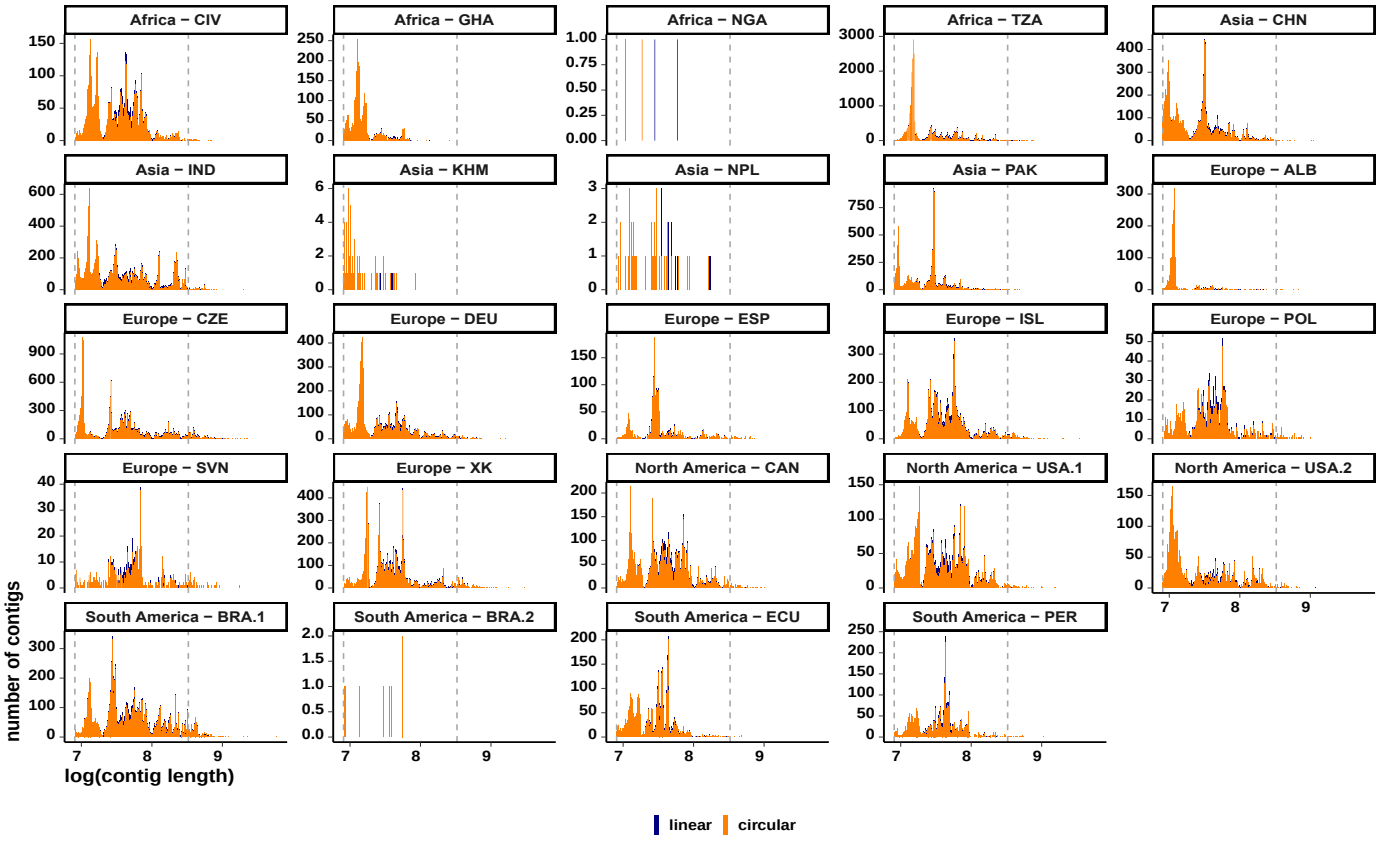
159.332 contigs (96.4%) suggested to be circular.

In other words: good amount!

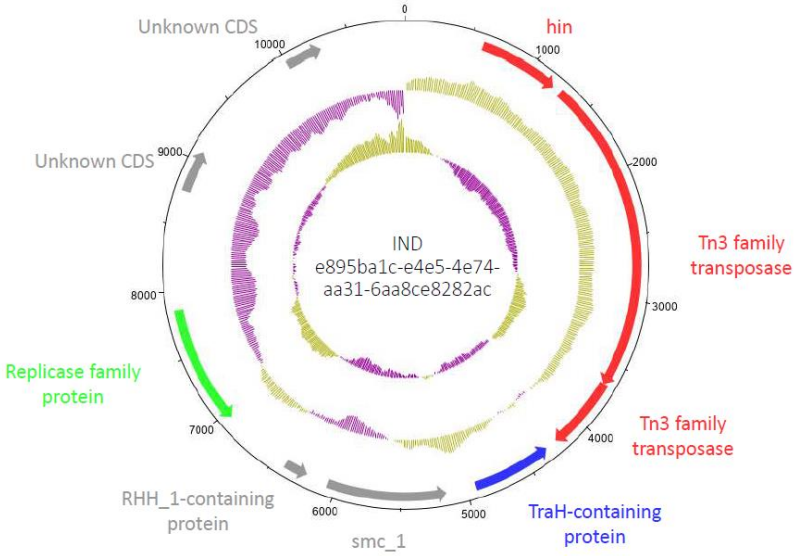
Global sewage plasmidome characterisation

Plasmid content in 24

samples from 22 countries show diltued signal of geograpgical differences.



Global sewage plasmidome characterisation (Nanopore)



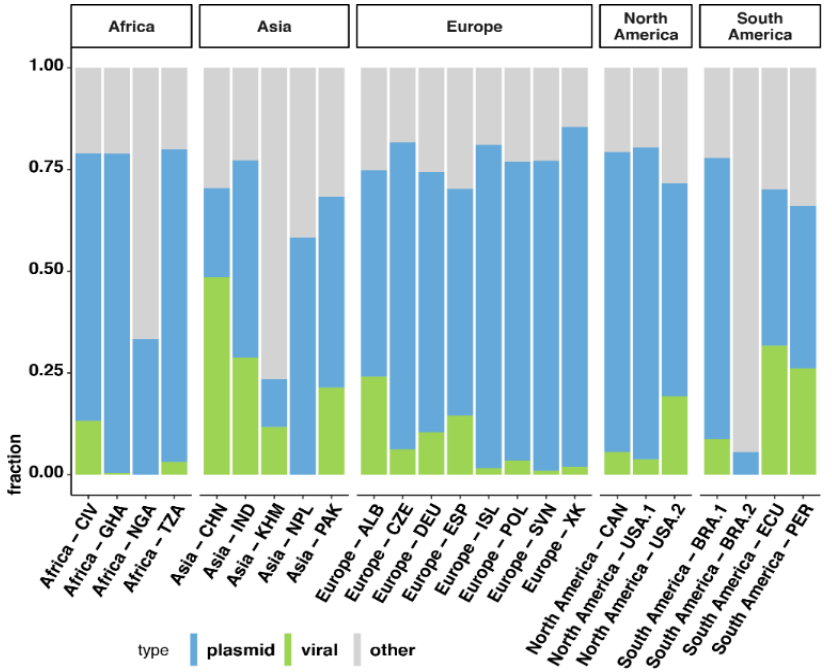
Unknown/other

Transposition

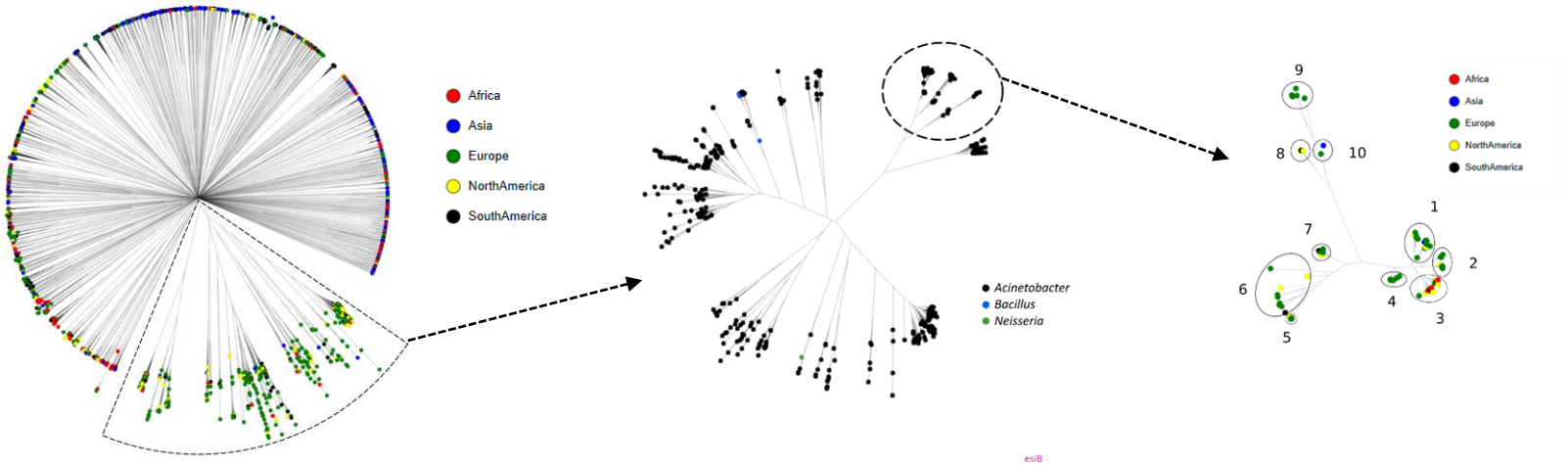
Conjugation/Mobilisation

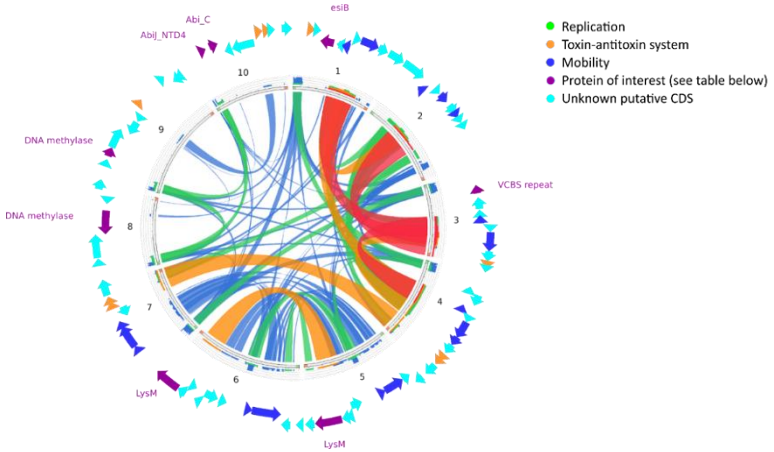
Replication

Resistance



Global sewage plasmidome functional characterisation





Global sewage plasmidome characterisation – more analyses

MGEfinder

- **Input**: Assembled sequence data (contigs or scaffolds)*
- Detects MGEs based on sequence similarity to known elements
- Includes a curated database of ~4,400 MGEs
- **Read more:** https://doi.org/10.1093/jac/dkaa390

Webserver,

- Graphical user interface
- Annotates AMR genes, Virulence factors & Plasmid re
- User friendly but less flexible

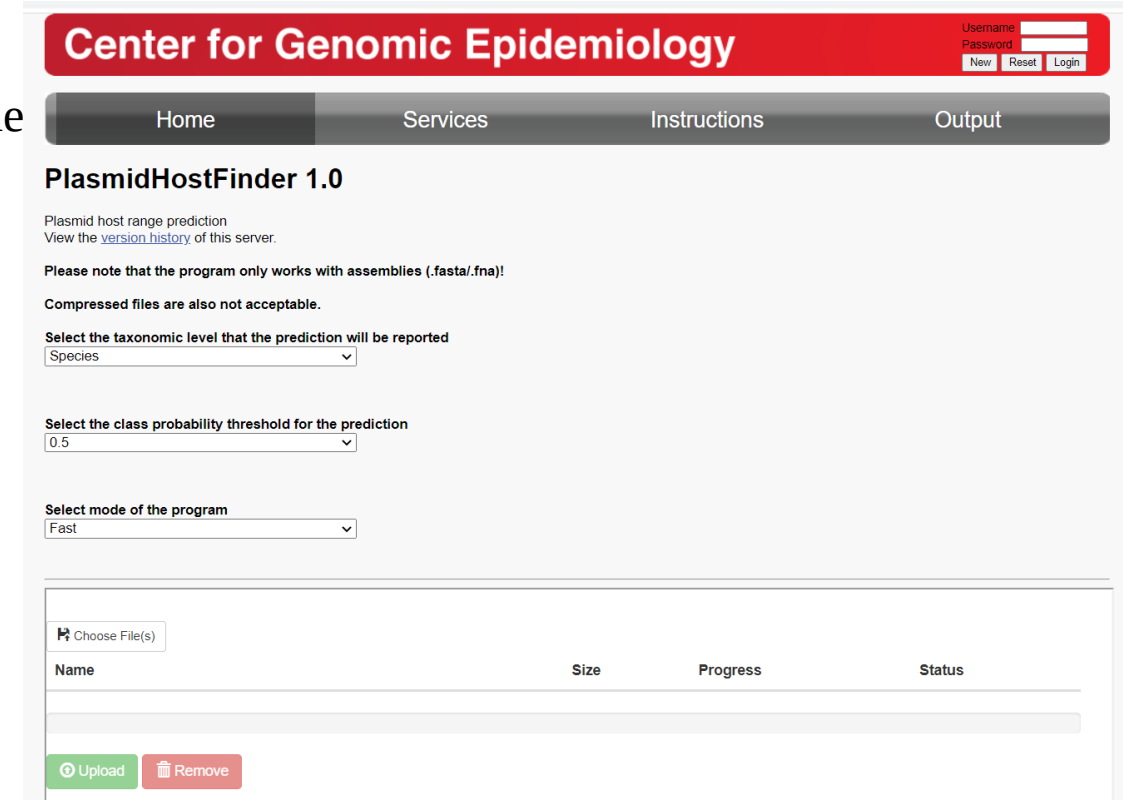
Local installation

- Python package, hosted on PyPI, installed with pip
- Flexible with user customizable thresholds

Plasmid-host prediction from sewage plasmidomes

Plasmid-host prediction using machine learning from plasmidome

"PlasmidHostFinder: Prediction of plasmid hosts using random forest"



Case 3:

Patient with gastrointestinal infection and diarrhoea: Sample: Faeces. Went to the hospital and was admitted to ICU immediately.

2: If we are to do metagenomics sequencing on the faecal sample of the patient, which technology would you use if the results are needed within one day with real-time result observation:

A: Illumina

B: 454

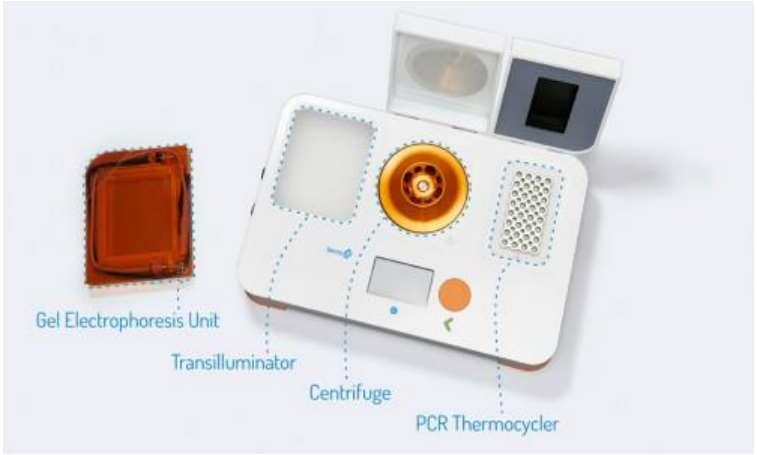
C: Oxford Nanopore

D: PacBio

Antimicrobial resistance and bacterial pathogen detection on-site

DNA extraction and library preparation (wet lab):

Can it be done in the field?



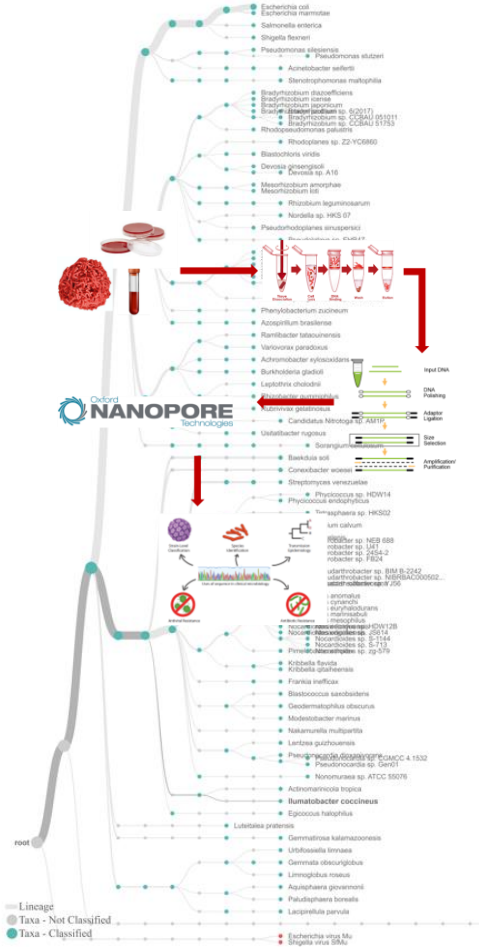
Antimicrobial resistance and bacterial pathogen detection on-site

FARMED:

Fast detection of pathogens and AMR genes on-site

November - May

DNA extraction and sequencing in field




```
#!/bin/bash

# Download software, reference genome, and FASTQ files
wget http://downloads.sourceforge.net/project/bio-bwa/bwakit/bwakit-0.7.12_x64-
linux.tar.bz2
wget -O refGenomeFiles.zip https://ndownloader.figshare.com/files/4841809
wget -O FASTQ.zip https://ndownloader.figshare.com/articles/3114454/versions/1

# Extract software and reference genome files
tar -jxvf bwakit-0.7.12_x64-linux.tar.bz2
unzip refGenomeFiles.zip
unzip FASTQ.zip

# Align FASTQ data to reference genome and save to SAM file
bwa.kit/bwa mem  KJ660346.fa SRR1972917_1.fastq SRR1972917_1.fastq >
SRR1972917_aligned.sam

# Convert SAM file to BAM file
bwa.kit/samtools view -bS SRR1972917_aligned.sam > SRR1972917_aligned.bam

# Sort BAM file
bwa.kit/samtools sort SRR1972917_aligned.bam SRR1972917_aligned_sorted

# Index BAM file
bwa.kit/samtools index SRR1972917_aligned_sorted.bam
```

Sequence analyses in field

Not only in Europe

In Tanzania (different project)

To remember ☺

Your sequence quality is of your DNA.

The better your DNA quality, the better your sequencing.

DNA extraction is material-dependent.

Know your starting material and downstream application.
Why are you sequencing.

To remember ☺

- **Short read technologies**

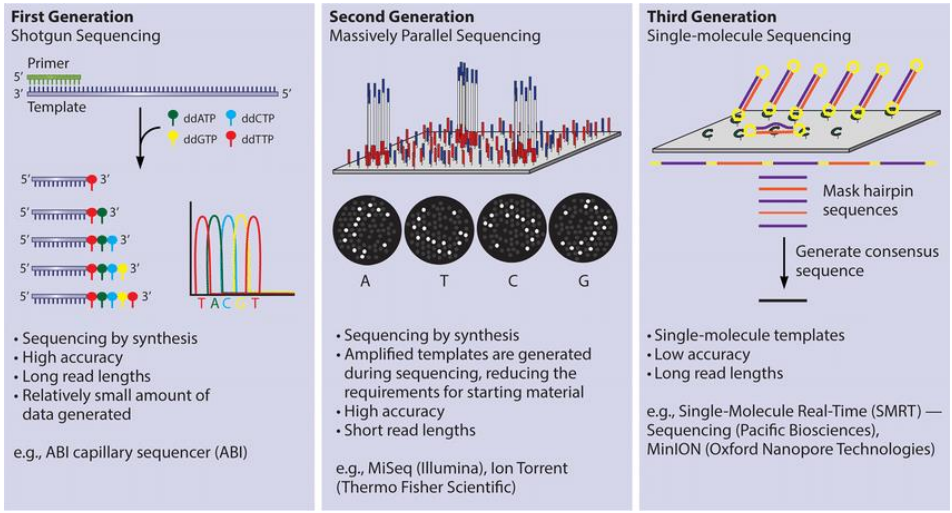
– Illumina (MiSeq, NextSeq, HiSeq, NovaSeq…)

– Ion Torrent

- **Long read technologies**

– Pacific Biosciences (PacBio)

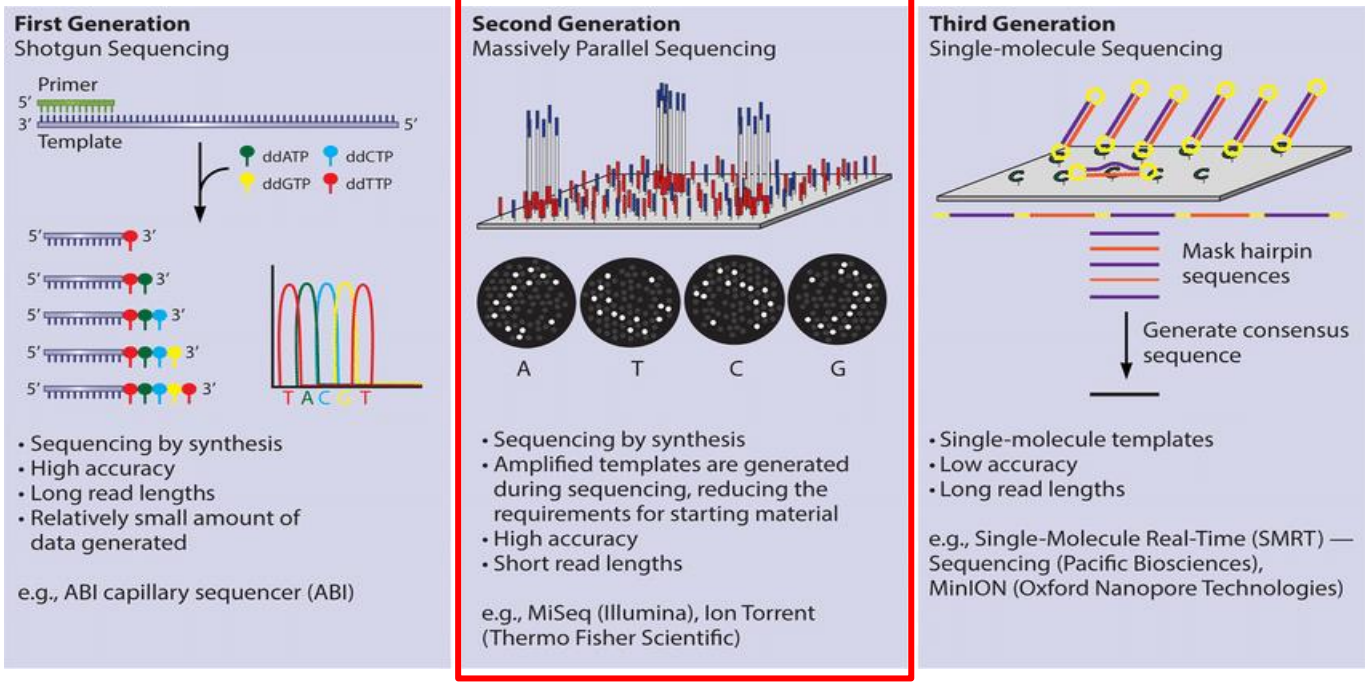
– Oxford Nanopore Technologies (MinION)



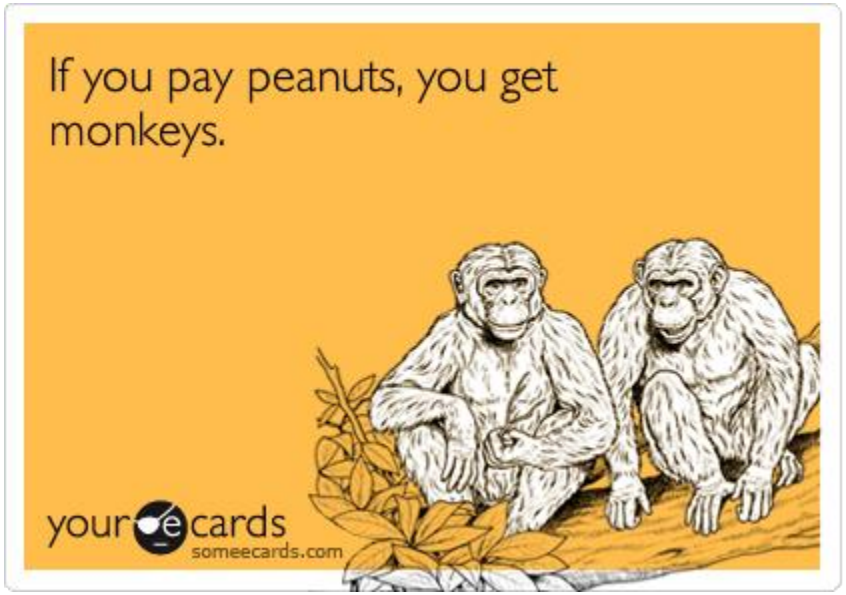
To remember ☺

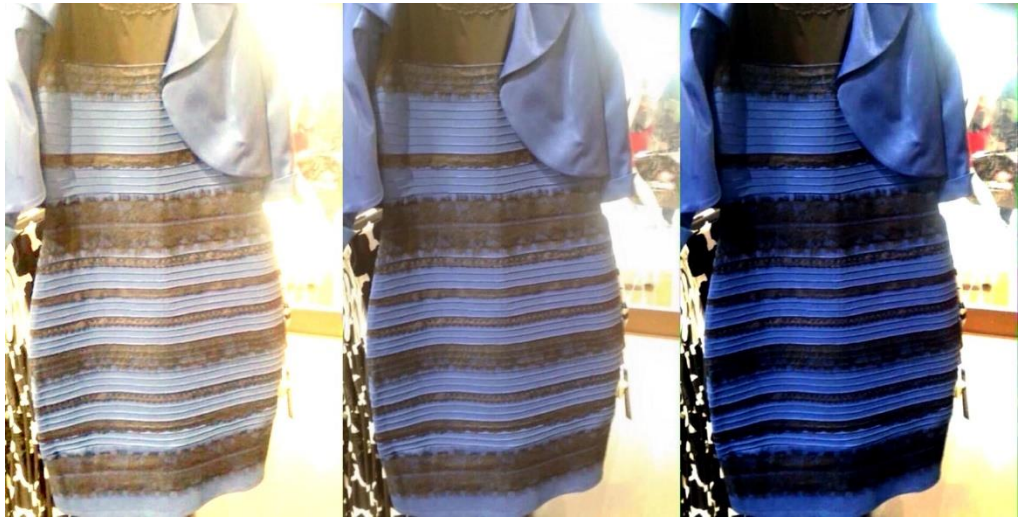
Choose your sequencing technology that serves your analysis best (*e.g.*, gene detection, identification).

This is commonly known as NGS



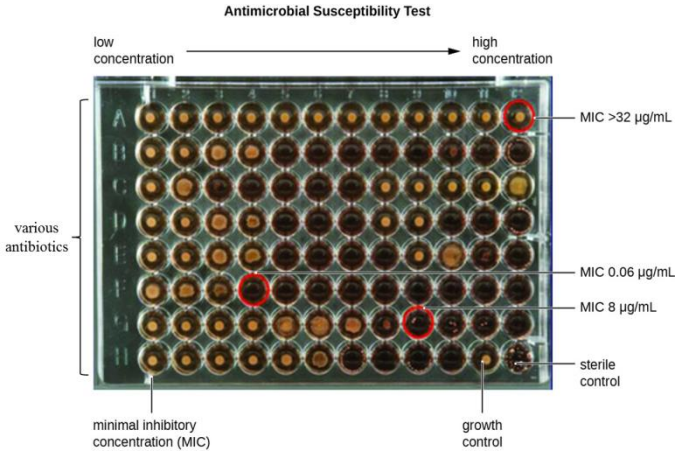
To remember ☺

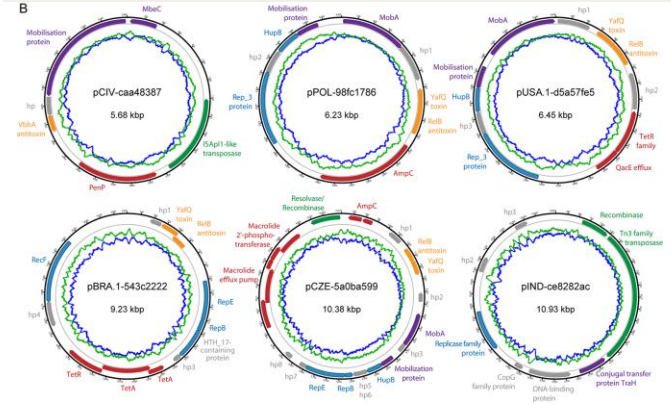

If you pay peanuts, you get monkeys.
yourecards
someecards.com



Next-Generation sequencing as a diagnostic tool (Microbiology)

We have come a long way





(Kristahler., Otani., *et al*., 2021)

"Now, here, you see, it takes all the running you can do just to keep in the same place. If you want to get somewhere else, you must run at least twice as fast!"

- Red Queen from *Through the Looking Glass*

Thank you ☺

Twitter: @SariaOtani

LinkedIn: Saria Otani

Email: saot@food.dtu.dk

Questions ☺