Next-Generation Sequencing for detection of emerging viruses

Aurelia Kwasiborski

Research engineer

Institut Pasteur Paris

Laboratory for Urgent Response to Biological Threats

Environment and Infectious Risks Research and Expertise Unit

aurelia.kwasiborski@pasteur.fr

Agenda

What will be covered in this presentation?

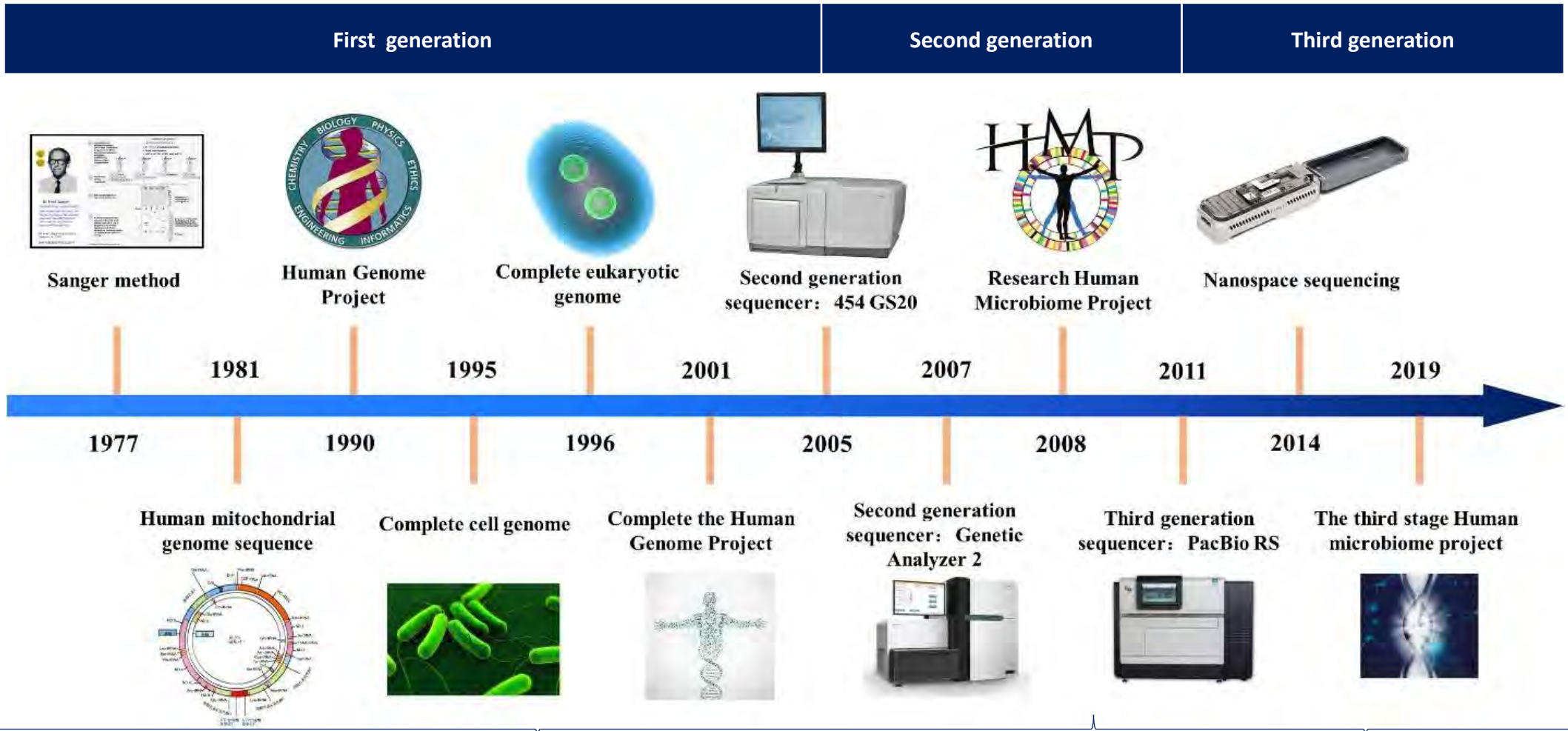
1. Overview of sequencing technology platforms

2. NGS applications (Illumina, Oxford Nanopore Technologies)

3. Metagenomics Next-Generation Sequencing for detection of viruses

4. Outbreak investigations : Metagenomics sequencing of Rift Valley Fever Virus, West Nile Virus…

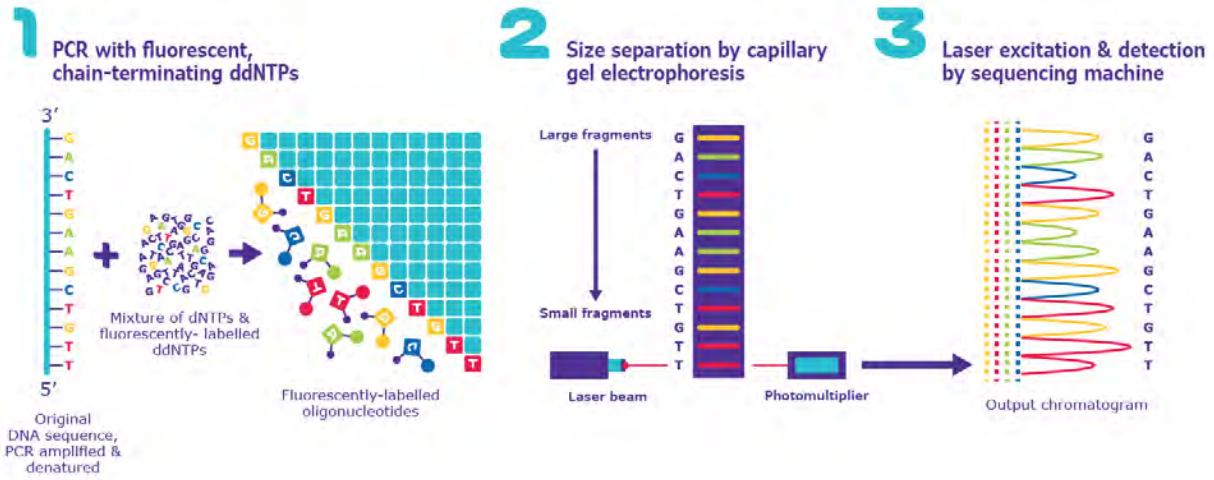
A brief history of sequencing technology



From Review on the Application of Machine Learning Algorithms in the Sequence Data Mining of DNA (https://doi.org/10.3389/fbioe.2020.01032)

First Generation Sequencing

Sanger sequencing workflow



Applied Biosystems

3700xl 96 capillary

3500xl 24 capillary

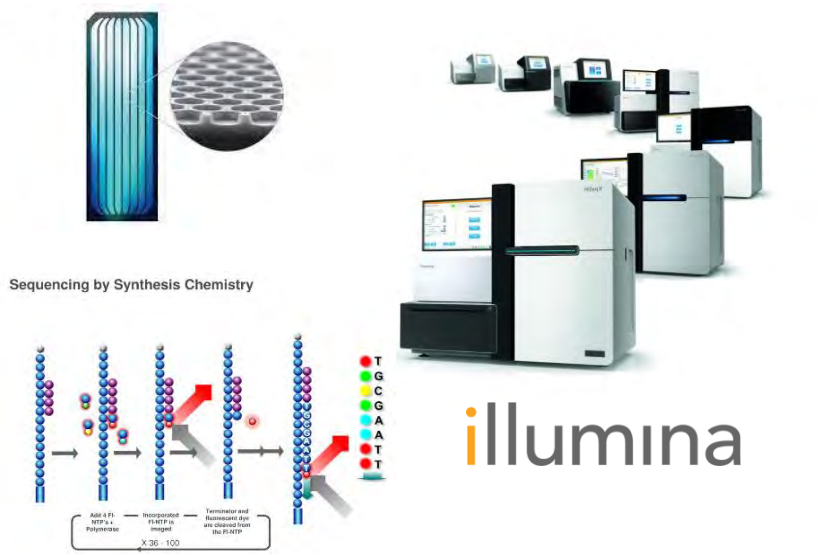
Advantages:

- Simple
- 500-1000pb fragments
- High accuracy

Limitations:

- Low throughput
- Requires primers that may not work
- Limited to very small portion of genome
- No possibility of scaling up
- No detection of low frequency variants
- Relatively cost per base still high

Second Generation Sequencing

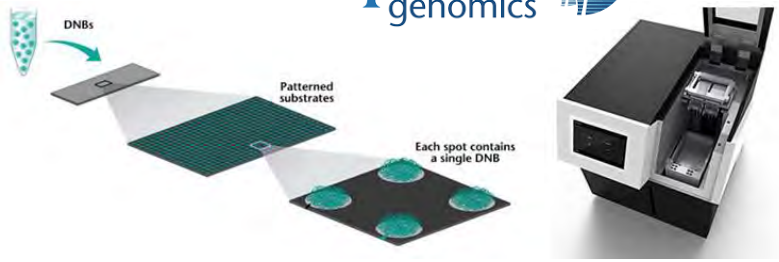


illumina

ThermoFisher SCIENTIFIC

Ion Torrent

Complete genomics

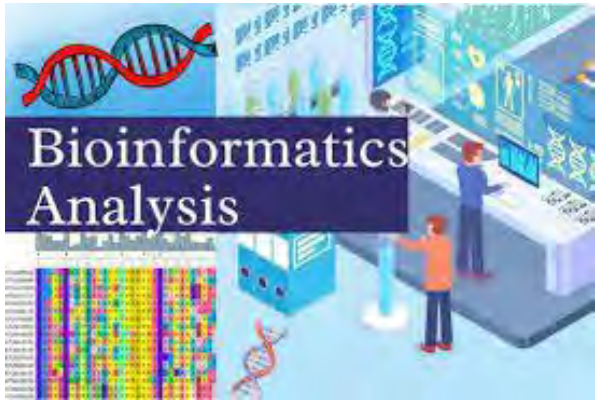


BGISEQ

QIAGEN

General characteristics and weaknesses

- High throughput from the parallelization of sequencing reactions
- Short reads (~50 - 600 bases)
- Moderate accuracy
- Clonal amplification phase (with PCR) is a source of bias (duplicates, PCR errors)
- Expensive sequencers and consumables
- Long time runs



Second Generation Sequencing

Illumina sequencers

	iSeq 100*	MiniSeq	MiSeq*†	NextSeq*†	HiSeq 4000*	NovaSeq 6000*††
Output Range	1.2 Gb	1.8–7.5 Gb	0.3–15 Gb	20–120 Gb	125–1500 Gb	134–6000 Gb
Run Time	9–17.5 hr	24 hr	5–55 hr	11–29 hr	< 1–3.5 days	13–44 hr
Reads per Run	4 million	8–25 million	1–25 million	130–400 million	2.5–5 billion	Up to 20 billion
Maximum Read Length	2 × 150 bp	2 × 150 bp	2 × 300 bp	2 × 150 bp	2 × 150 bp	2 × 150 bp

30 000 €

125 000 €

€€€€

Reagents

Fluorometer Qubit 4

Bioanalyzer 2100

Applications

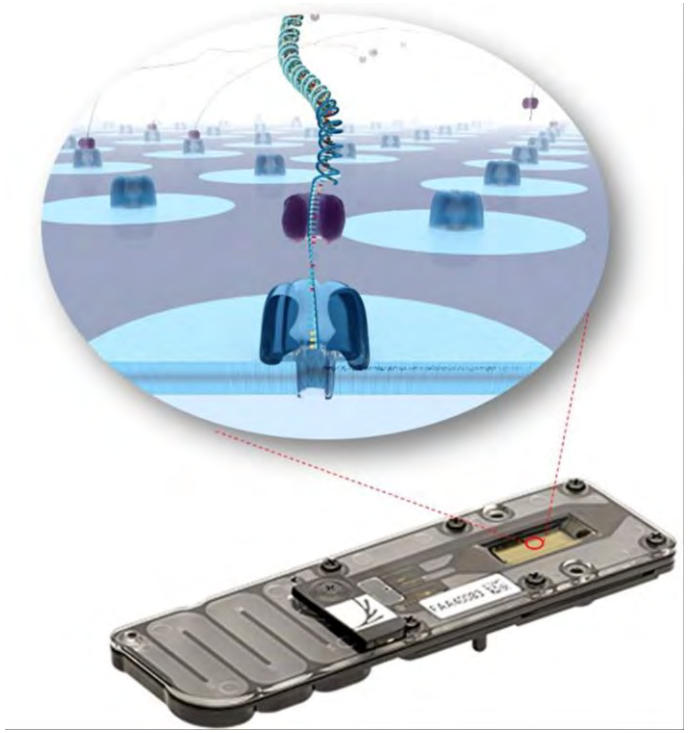
Popular Applications & Methods	Key Application	Key Application	Key Application	Key Application	Key Application
Large Whole-Genome Sequencing (human, plant, animal)					
Small Whole-Genome Sequencing (microbe, virus)	●	●	●	●	●
Exome & Large Panel Sequencing (enrichment-based)				●	●
Targeted Gene Sequencing (amplicon-based, gene panel)	●	●	●	●	●
Single-Cell Profiling (scRNA-Seq, scDNA-Seq, oligo tagging assays)				●	●
Transcriptome Sequencing (total RNA-Seq, mRNA-Seq, gene expression profiling)				●	●
Targeted Gene Expression Profiling	●	●	●	●	●
miRNA & Small RNA Analysis	●	●	●	●	●
DNA-Protein Interaction Analysis (ChIP-Seq)			●	●	●
Methylation Sequencing				●	●
16S Metagenomic Sequencing		●	●	●	●
Metagenomic Profiling (shotgun metagenomics, metatranscriptomics)		✓	✓	●	●
Cell-Free Sequencing & Liquid Biopsy Analysis				●	●

https://www.illumina.com/systems/sequencing-platforms.html

Third Generation Sequencing

Nanopore Sequencing (Oxford Nanopore Technology)

Sequencing technologies by detecting successive bases of a DNA/RNA molecule through a pore (protein)



Advantages:

- Sequence native DNA in real-time with single molecule resolution
- Long sequencing reads (20bases - 4 million bases…)
- Without loss of information (methylated bases, …)
- Faster (turnaround time, rapid access to result, …)
- Able to detect a very low signal
- Portable and versatile technology …etc

Mk1C

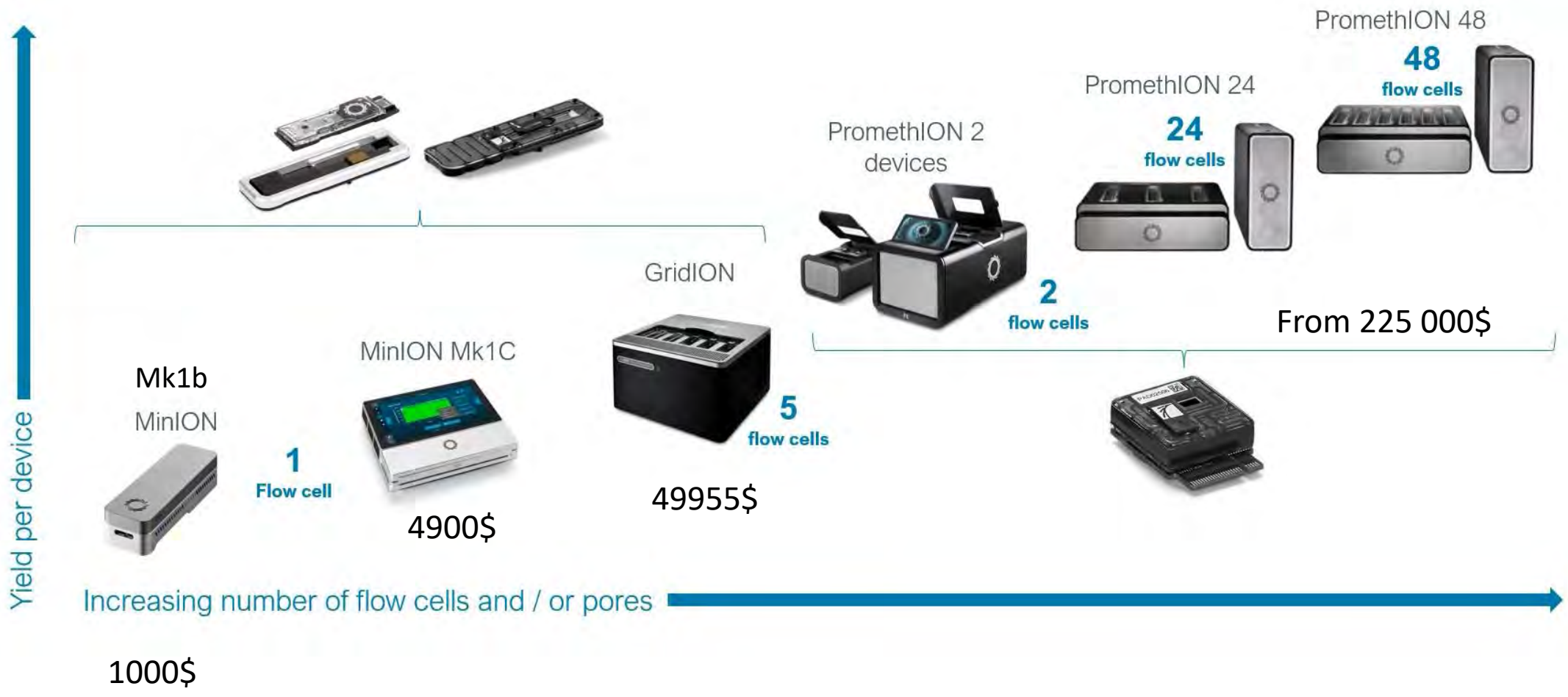
Mk1b

Limitations:

- Lower accuracy
- Cost of consumables

https://nanoporetech.com/community

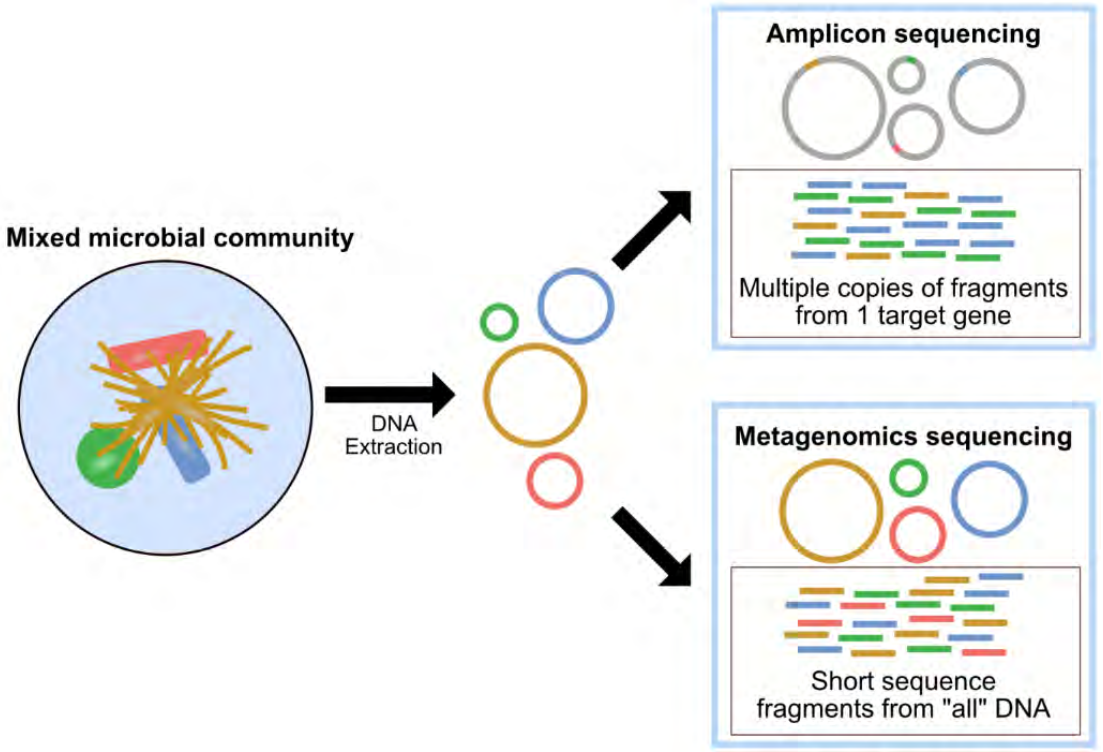
Third Generation Sequencing



NCM2022: https://nanoporetech.com/community

Metagenomics NGS

Sequencing approaches



TARGETED AMPLICON

- PCR-based rRNA 16S, 18S and ITS gene etc
- Hypothesis required
- Requires primers that may not work
- Lower sequencing depth required
- Less complex computational analysis required (phylogeny)
- PCR amplification biases

SHOTGUN METAGENOMICS

- Sequence the entire genetic content present in a sample
- Without a priori hypothesis
- Beyond pathogen identification (virulence, resistance…)
- Detection of novel pathogens
- Must also sequence host background
- Expensive (higher sequencing depth required)
- Prone to contaminations with environmental species

Biological question?
Hypothesis?
Sample type?
Experimental design?

Metagenomics NGS workflow

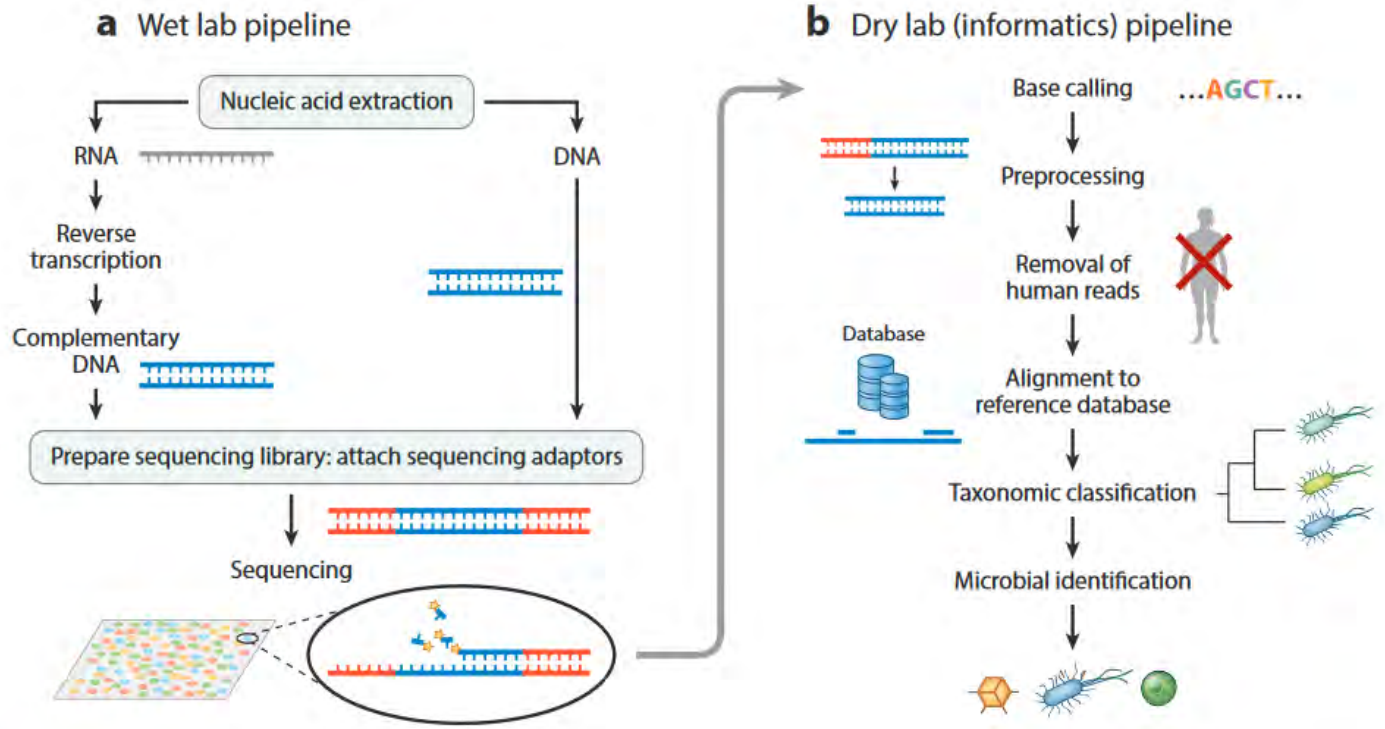


Figure 1

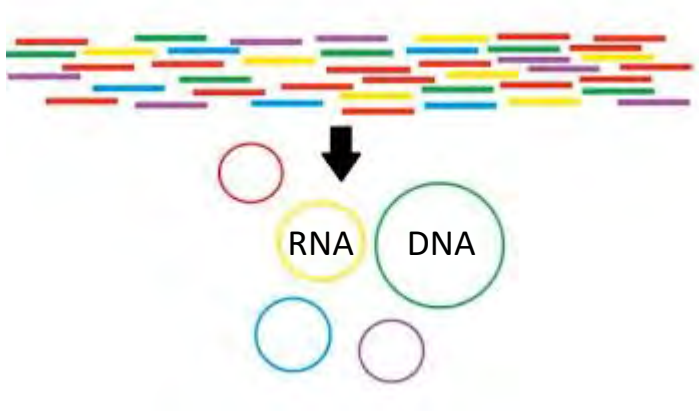
Schematic of the generalized workflow of metagenomic next-generation sequencing for diagnostic clinical use. The workflow has two components: (*a*) a wet lab protocol in which samples are collected, processed, extracted for nucleic acids, prepared into a sequencing library, and sequenced, and (*b*) a dry lab computational pipeline that includes microbial identification, statistical analysis, and interpretation. The sequencing library may be targeted, undergo DNA amplification, or both.

Gu W, Miller S, Chiu CY. Clinical Metagenomic Next-Generation Sequencing for Pathogen Detection. Annu Rev Pathol. 2019 Jan 24;14:319-338. doi: 10.1146/annurev-pathmechdis-012418-012751. Epub 2018 Oct 24. PMID: 30355154; PMCID: PMC6345613.

Preparing samples for metagenomics sequencing

Sample collection and transport

- Potential samples for mNGS: body fluids, swabs, environmental samples
- Input DNA/RNA <100pg
- Minimize the possibility of nucleic acid degradation during sampling and storage
- Avoid multiple freeze-thaw steps



Nucleic acid extraction

- Extracting DNA: ZymoBIOMICS Microbial Community Standard, Qiagen Purgene etc
- Extracting RNA: Direct-zol RNA Kits, Quick-RNA Viral Kits, High Pure Viral Nucleic Acid Kit Viral RNA/DNA (Roche)
- Isolation of viral RNA and DNA and bacterial DNA from a variety of animal sample types : QIAamp cador Pathogen Mini Kit (QIAGEN)
- Extraction controls: negative and positive (contamination control)

Preparing samples for metagenomics sequencing

Host NA depletion

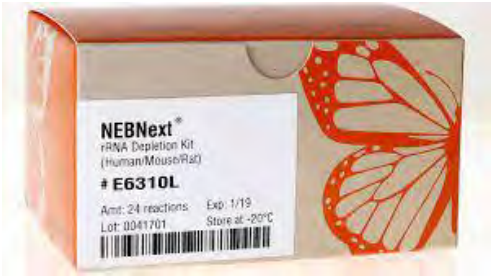
Experimental design

Pre-extraction:

- centrifugation to pellet human cells
- filtration to remove human and bacterial cells
- nuclease treatment

Post-extraction rRNA depletion :

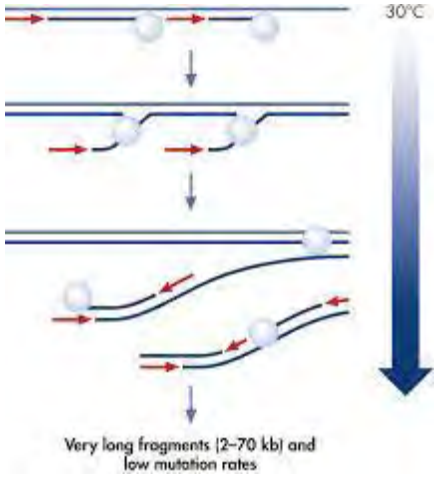
- hybridization [Ribo-Zero (Illumina)]
- exonuclease digestion [MICROBExpress (Ambion)]
- endonuclease digestion (RNase H), etc
- NEBNext® rRNA Depletion Kit (Human/Mouse/Rat)



QIAGEN (MDA)

Viral Enrichment

- Sequence Independent, Single Primer Amplification (SISPA)
- Single Primer Isothermal Amplificaiton (SPIA) (Ovation RNA Amplification System, NuGen)
- Multiple displacement amplification Phi29 (MDA) (QIAGEN)
- Multiple Annealing and Looping Based Amplification Cycles (MALBAC) etc



Lopez-Labrador, 2021, Journal of Clinical Virology https://doi.org/10.1016/j.jcv.2020.104691

Outbreak investigation : Mayotte, RVFV

- **Rift Valley Fever Virus (tri-segmented single-stranded RNA)**

(Collaboration with Louis Collet from Mayotte hospital and Catherine Cêtre-Sossah and Eric Cardinale, CIRAD, UMR 117 ASTRE)

- Nov 2018 – Aug 2019
- 143 confirmed human cases (2 severe)
- 126 animal foci (100 bovines, 26 caprines/ovines)

Viral load testing :

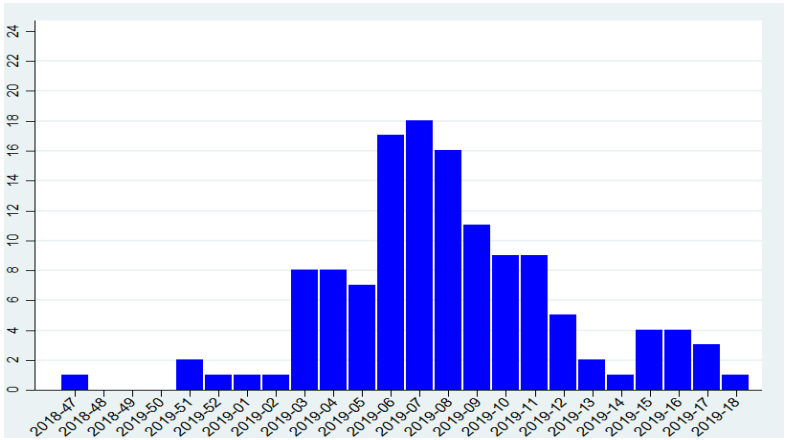
- Quantitative real-time RT-PCR

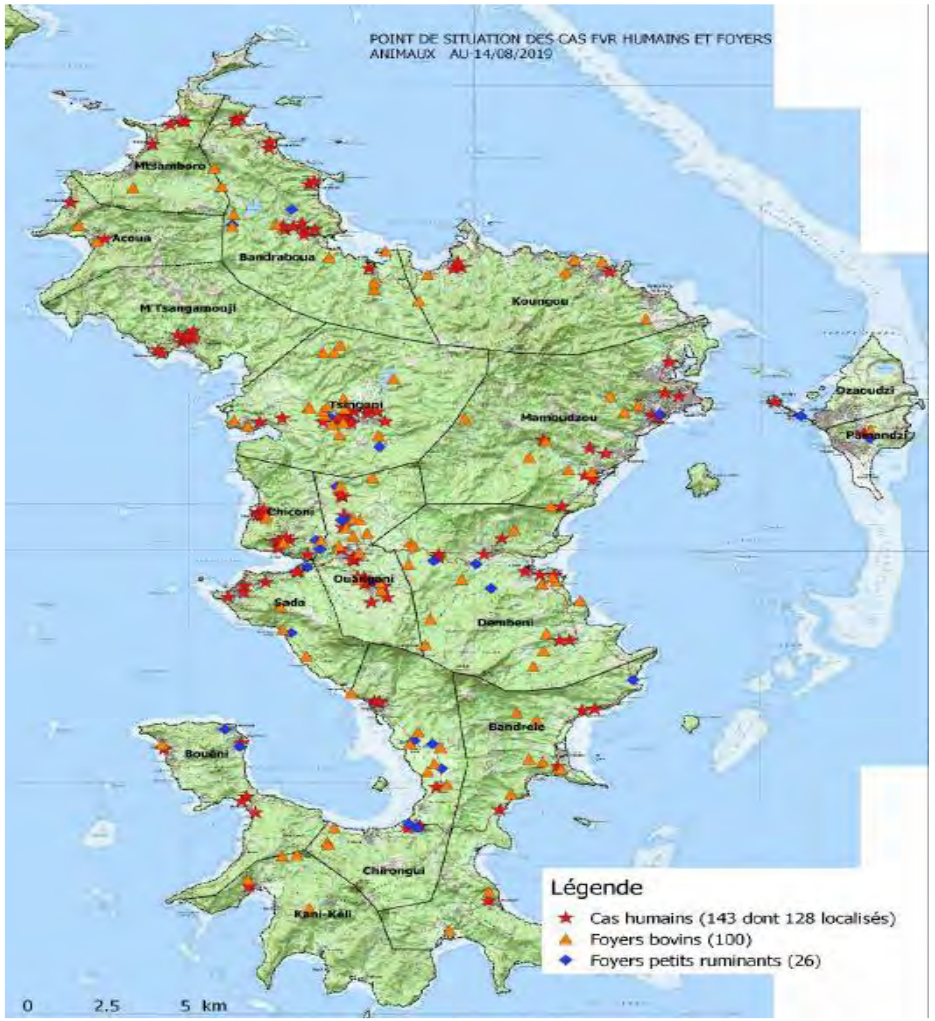
Viral isolation

- Vero E6 cell line

Next-generation sequencing

- Clinical sample
- Illumina sequencing
- MinION sequencing





Outbreak investigation : Mayotte, RVFV

Next-generation sequencing : Illumina

Sample preparation

- Host depletion
  - DNAse treatment
  - Ribosomal RNA removal kit (NEB)
- Untargeted amplification
  - Whole Transcriptome Amplification (MDA) (plasma)
- Sequencing library preparation
  - Nextera XT
  - Multiplex libraries (x5)
  - MiSeq sequencer (Ilumina)

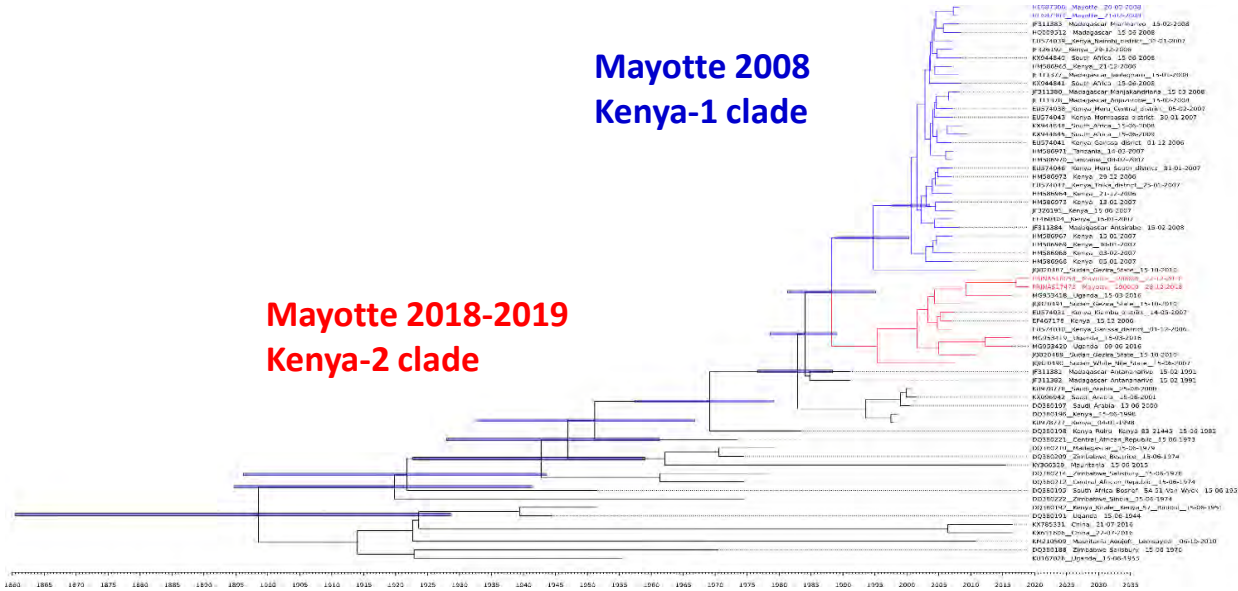
Case ID	#Reads	Segment S (% coverage)	Segment M (% coverage)	Segment L (% coverage)	Ct value
1	1 210 424	38%	92%	87%	27.4
2	**1 345 032**	**100%**	**96%**	**100%**	**21.8**
3	1 447 561	46%	57%	67%	25.7
4	**1 617 180**	**100%**	**96%**	**100%**	**22.6**
5	1 233 073	Nd	Nd	Nd	31.95

Results:

- Mayotte 2018 strains (in red) belong to the lineage H, and Kenya-2 clade
- Re-introduction rather than reemergence from East Africa

Possible drivers:

- importation of infected livestock from the African mainland
- presence of competent vector species on the island

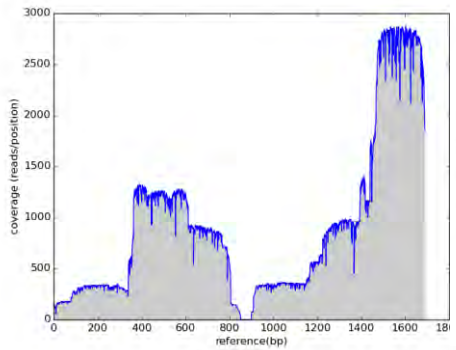


Outbreak investigation : Mayotte, RVFV

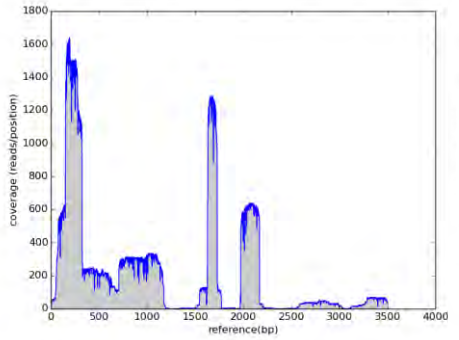
Next-generation sequencing : MinION

- Metagenomics sequencing parameters:
  - Rapid PCR Barcoding Kit (SQK-RPB004)
  - Input 1ng
  - Multiplexing x5
  - Flow-cell FLO-MIN 106 (R9.4)
- Data processing:
  - MinKNOW
  - Run time 24h
- Data analysis
  - What is my pot? (EPI2Me)
  - Mapping : Minimap2
  - Visualization Samtools

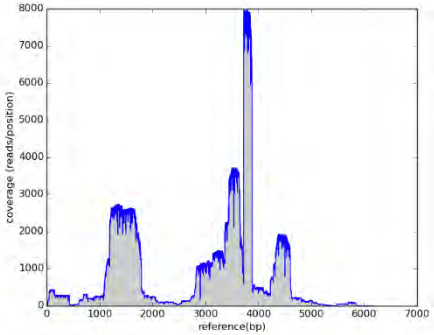
S (EU574072)



M (EU574044)



L (EU574017)



Case ID	#Reads	Segment S	Segment M	Segment L	Ct value
1	7770	24%	9%	30%	27.4
2	**16178**	**97%**	**56%**	**85%**	**21.8**
3	5058	30%	0%	35%	25.7
4	**4785**	**90%**	**47%**	**71%**	**22.6**
5	8698	**Nd**	**Nd**	**Nd**	31.95

Outbreak investigation: Tunisia, WNV

WNV outbreak in Tunisia 2018-2019

(collaboration with IP Tunis - Wasfi Fares and Youmna Mghirbi)

- WNV outbreak occurred between August and November 2018
- Total of 96 human cases were confirmed among 399 suspicion cases with neurological symptoms
- Additionally, medical entomologists at IP Tunis captured mosquitoes in the region -> positive mosquitoes for WNV
- Molecular characterization of WNV strains was performed based on 23 partial sequences grouped in a distinct monophyletic group and were clearly distinct from the Tunisia 1997 WNV strain
- Aim of this project is to obtain whole genome of WNV strains to better understand the WNV epidemic emergence

Outbreak investigation: Tunisia, WNV

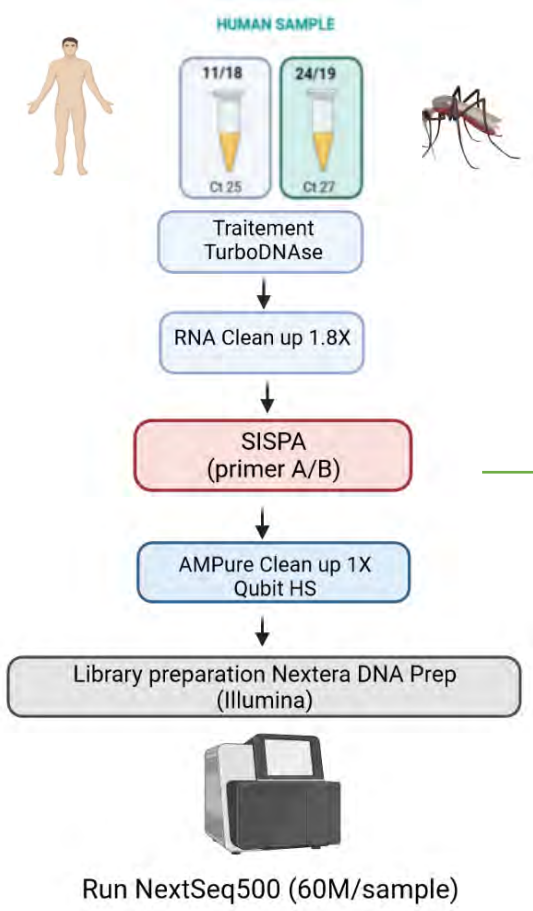
Metagenomics NGS characterization of the novel West Nile Virus strain related to the 2018 Tunisian outbreak

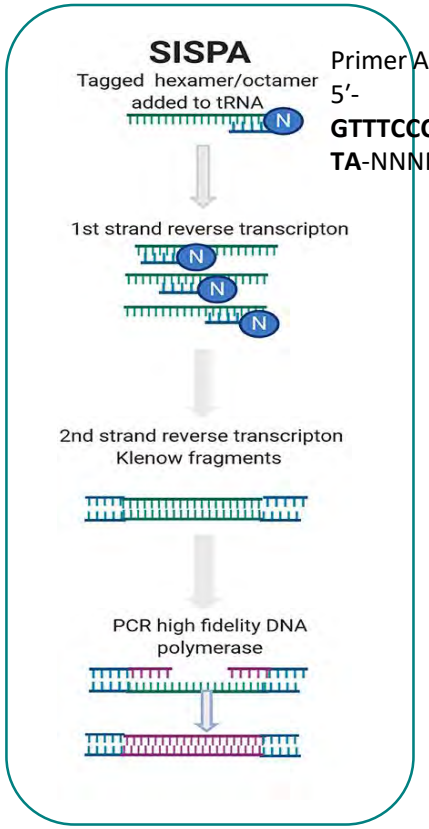
mNGS

- Host depletion (DNAse treatment)
- Viral Enrichment (SISPA amplification)
- NGS library preparation (Nextera DNA prep, Illumina)

Next steps:

- Comparison of human and mosquitos sequences of WNV
- Development of amplicon-based sequencing approach using MinION to better understand this epidemic emergence of WNV
- Phylogenetic analysis





Primer A
5'-**GTTTCCCACTGGAGGA TA**-NNNNNNNNN-3'

Amplification with Primer B
5'-**GTTTCCCACTGGAGGATA**-3'

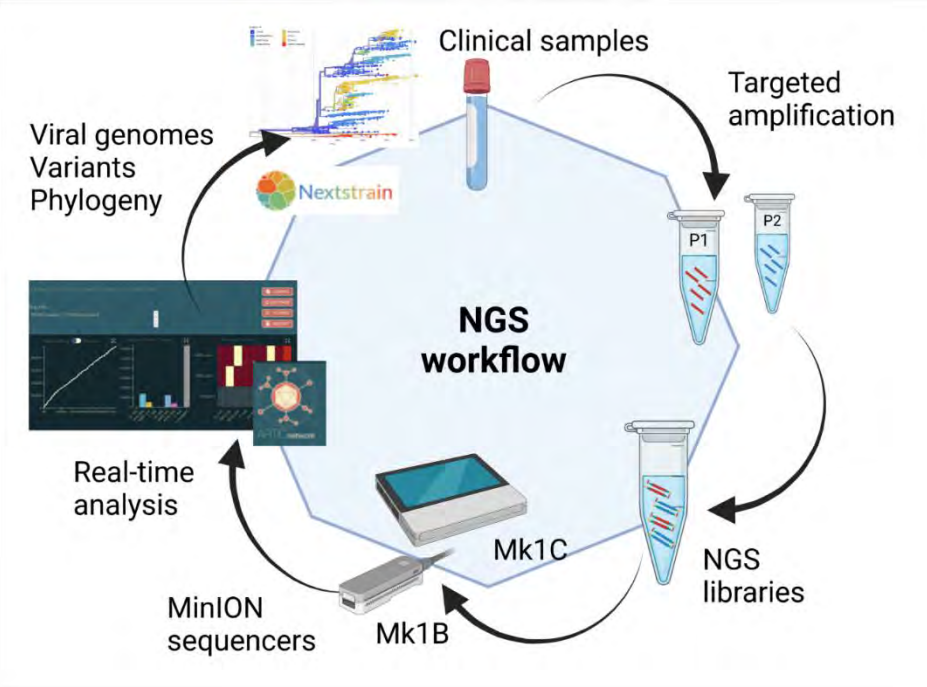
Outbreak investigation: SARS-CoV-2

French-COVID-19, SARS-CoV-2

- **Amplicon-based approach** (cDNA synthesis and multiplex PCR)

  NGS library preparation (SQK-LSK-109 + Native barcoding (multiplex X24)

  MinIOn Mk1C (ONT) **MinION**, (FLO-MIN106)
- **Real time** sequencing using RAMPART (**R**ead **A**ssignment **M**apping and **P**hylogenetic **A**nalysis in **R**eal **T**ime).
- Artic-nCov pipeline from reads to consensus sequences



RESULTS

- Using this amplicon-based approach, we found that viral genomic sequencing could be performed with high confidence and in just a few hours, **directly from clinical specimens**, in particular with low viral loads

Hourdel et al, 2020, Rapid Genomic Characterization of SARS-CoV-2 by Direct Amplicon-Based Sequencing Through Comparison of MinION and Illumina iSeq100™ System

https://doi.org/10.3389/fmicb.2020.571328

Thank you for your attention

Laboratory for Urgent Response to Biological Threats
Environment and Infectious Risks Research and Expertise Unit

Valérie Caro
Aurelia Kwasiborski
Véronique Hourdel
Charlotte Balière

cibu-pgp@pasteur.fr