



---

# Workshop

## Molecular Diagnostic Techniques (Genomics and Bioinformatics – Infectious diseases)

### Module-1

Dr. Srinivas Reddy Pallerla, Mr. Tung Tran and Ms. Linh Le  
Prof. Dr. Thirumalaisamy P. Velavan, University Hospital, Tübingen, Germany  
Prof. Dr. Daniel Olusola Ojurongbe, Nigeria  
Prof. Dr. Mohamed Osman, University of Khartoum, Sudan

# Training schedule

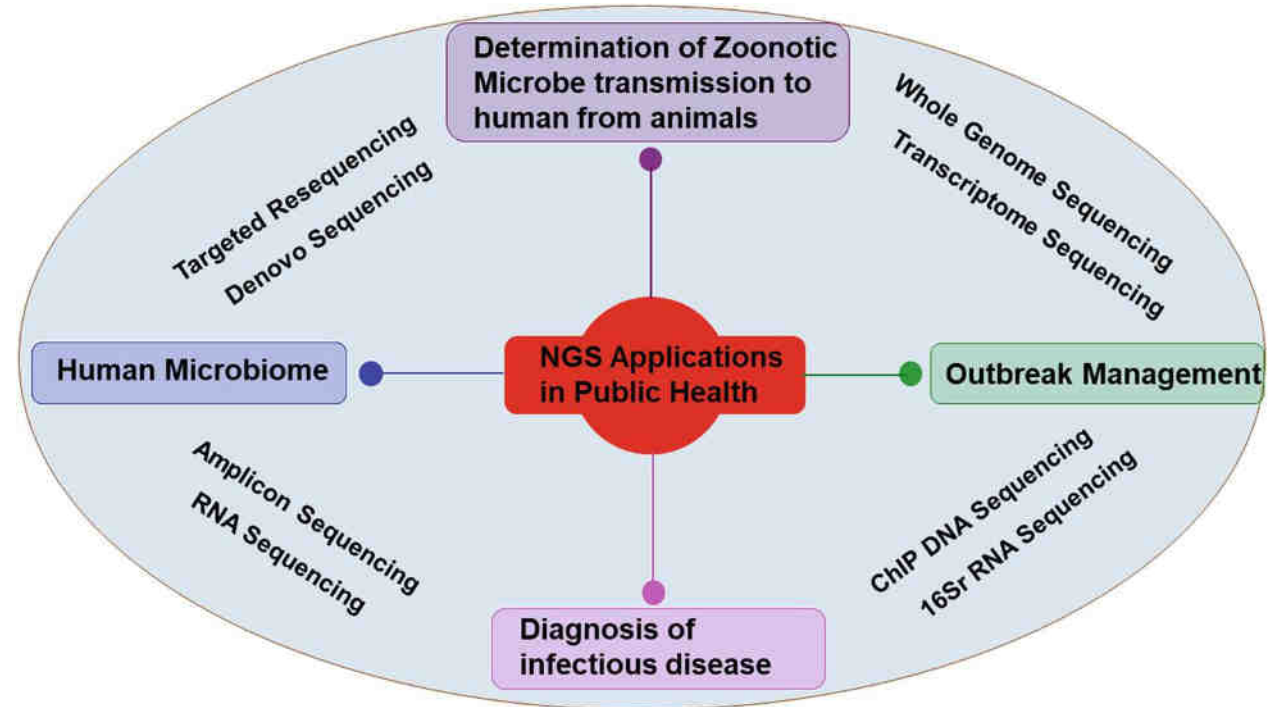
Time/Day	Monday 25 <sup>th</sup>	Tuesday 26 <sup>th</sup>	Wednesday 27 <sup>th</sup>	Thursday 28 <sup>th</sup>	Friday 29 <sup>th</sup>
9-10 am	Overview of CERID Workshop rules – Prof Ojurongbe  Objectives and expectations; Workshop Pre-test Prof Olowe	<b>Wet Lab</b>  Intro. to Laboratory equipment Pipetting skills DNA extraction DBS – Plasmodium Crude - Bacteria isolate – Dr. Ayoola	RNA Extraction  Nasal Swab for Sars-Cov-2 detection  Tung/Dolapo/Salma	<b>MODULE -3</b>  Oxford Nanopore Seq Demonstration of Nanopore Seq Barcoding, Flow cell loading, Sequencing Flow cell washing  Tung/Dolapo/Salma	<b>MODULE -4</b>  Bioinformatics:  Nanopore data analysis Linux (Ubuntu) OS. Linux commands Open terminal and practice commands  Dr. Pallerla/Tung
10-11 am	Principle and Types of PCR - Prof Opaleye				
11-11.15 am	<b>Coffee/Tea Break</b>				
11.15-11 am	Real-Time PCR	<b>Hands on PCR</b>  Master Mix prep PCR amplification Gel electrophoresis  -Drs. Funwei / Ayoola	■ RT-cDNA Synthesis Real-time PCR Sars-Cov-2 and Monkey Pox detection  Tung/Dolapo/Salma	Oxford Nanopore Seq  Demonstration of Nanopore Seq Barcoding, Flow cell loading, Sequencing Flow cell washing  Tung/Dolapo/Salma	Bioinformatics:  Nanopore data analysis Linux (Ubuntu) OS. Linux commands Open terminal and practice commands  Dr. Pallerla/Tung
11-12 pm	Principles and Types – Prof Ojurongbe				
12-1 pm	Application of RT-PCR  COVID-19 and other pathogen detection  Dr. Pallerla				
1-2 pm	<b>LUNCH BREAK</b>				
2-3 pm	Introduction to NGS and Bioinformatics – Dr. Pallerla	<b>Hands on PCR</b>  Master Mix prep PCR amplification Gel electrophoresis  - Drs. Funwei / Ayoola	■ RT-cDNA Synthesis Real-time PCR Sars-Cov-2 and Monkey Pox detection  Tung/Dolapo/Salma	Oxford Nanopore Seq  Demonstration of Nanopore Seq Barcoding, Flow cell loading, Sequencing Flow cell washing  Tung/Dolapo/Salma	Other Bioinformatic tools  <b>MODULE - 5</b>  MEGA Phylogenetic Analysis  Dr. Ayoola/Tung
3-4 pm	<b>MODULE -1</b>				
4-4.15 pm	<b>Coffee/Tea Break</b>				
4.15-5 pm	Introduction to Nanopore Sequencing - Dr. Pallerla	<b>Hands on PCR</b>  Master Mix prep PCR amplification Gel electrophoresis  Drs. Funwei / Ayoola	■ RT-cDNA Synthesis Real-time PCR Sars-Cov-2 and Monkey Pox detection  Tung/Dolapo/Salma	Oxford Nanopore Seq  Demonstration of Nanopore Seq Barcoding, Flow cell loading, Sequencing Flow cell washing  Tung/Dolapo/Salma	Other Bioinformatic tools  MEGA Phylogenetic Analysis  Dr. Ayoola/Tung
5-5.45pm	<b>MODULE -2</b>				

# Learning objectives of this modul

- Introduction NGS and bioinformatics
- First, second and third generation sequencing technologies
- Genomics and bioinformatics
- Genomics in Infectious disease research and public health
- File systems, terminology and softwares/repositories
- Oxford Nanopore Sequencing and data analysis
- Programming languages used commonly (Python, R and Bash)
- Linux (Ubuntu) operating system
- Linux commands usage: Open terminal and practice to use these commands

# Genomics in Infectious disease research and public health

Diagnostics,  
monitoring for viral evolution,  
transmissibility and virulence,  
transmission dynamics,  
antibiotic resistance,  
nosocomial outbreaks,  
vaccine development, and  
the overall assessment of infection prevention and control measures



COVID-19 pandemic has highlighted the need to incorporate pathogen genomics for enhanced disease surveillance and outbreak management.

# First, second and third generation sequencing technologies

## First Generation



Sanger Sequencing  
Maxam and Gilbert  
Sanger Chain-termination

- Infer nucleotide identity using dNTPs then visualize with electrophoresis
- 500-1000 bp fragments

## Second Generation (Next Generation Sequencing)



454, Solexa,  
Ion Torrent  
Illumina

- High throughput from the parallelization of sequencing reactions
- ~50-500 bp fragments

## Third Generation



PacBio  
Oxford Nanopore

- Sequence native DNA in real time with single-molecule resolution
- Tens of kb fragments, on average

**Short-read sequencing**

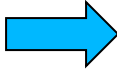
**Long-read sequencing**

# Advances in Sequencing technologies

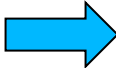


2001 (Whitehead)

13 Years & 3.8 billion \$

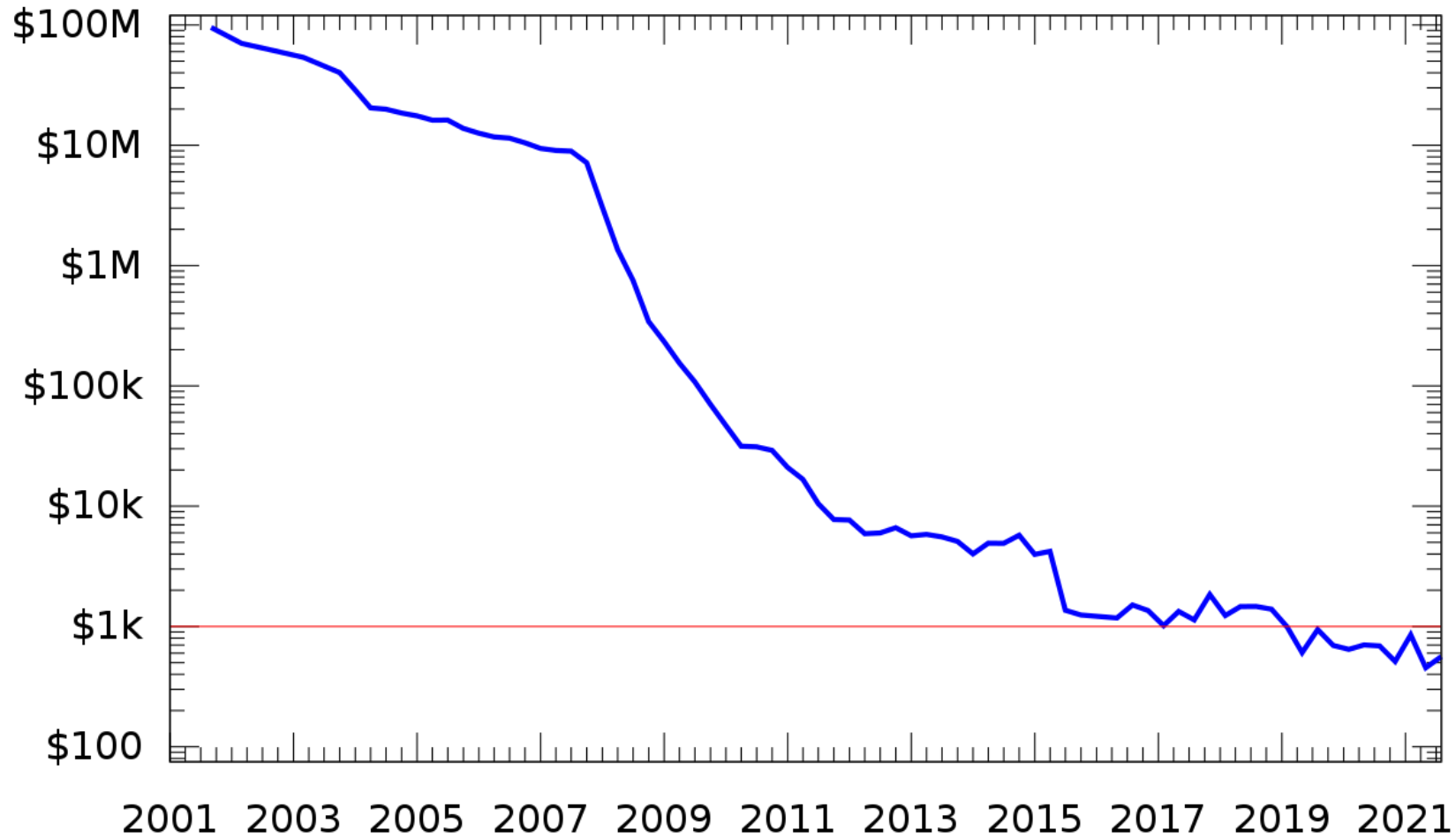


2018 (Illumina)

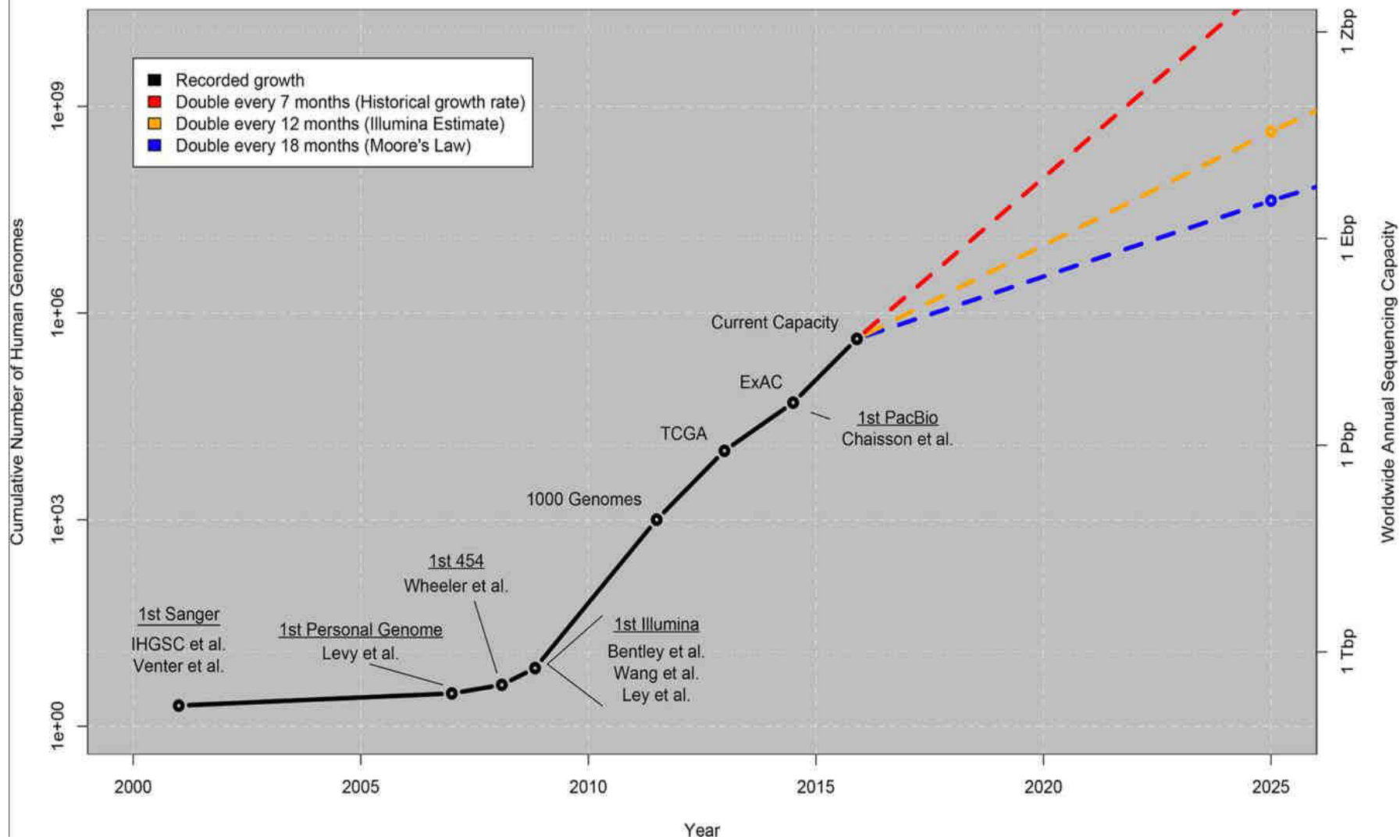


<12 hours & 1000\$

Cost to sequence a human genome (USD)



### Growth of DNA Sequencing



<https://goo.gl/3RhFkH>





## The Largest Current Bottleneck in Genomics...



# Advantages and disadvantages for short vs long read sequencing

## Short-read sequencing

- Higher sequence fidelity
- Cheap
- Can sequence fragmented DNA
- Not able to resolve structural variants or distinguish highly homologous genomic regions
- Unable to provide coverage of some repetitive regions

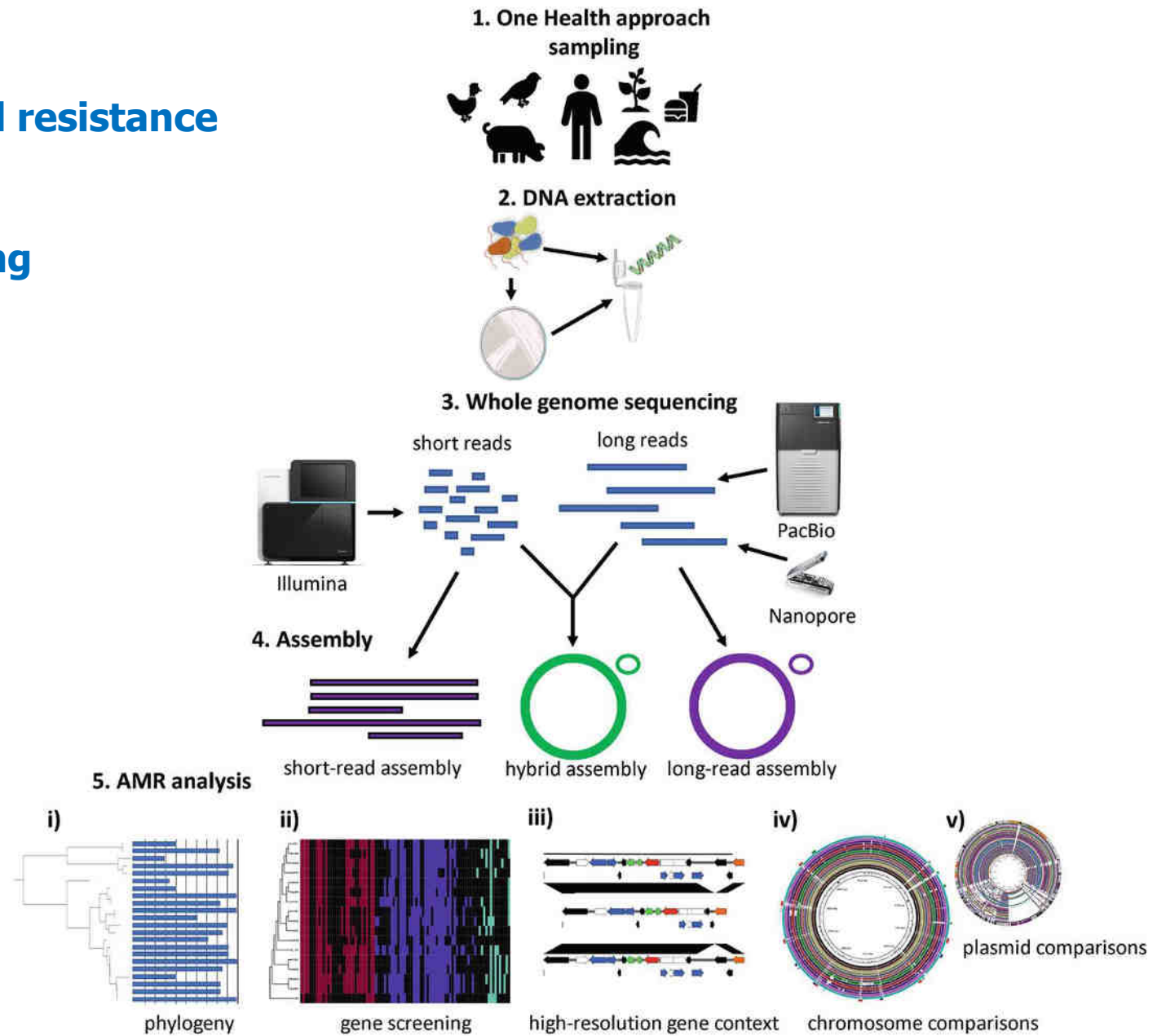
## Long-read sequencing

- Able to sequence genetic regions that are difficult to characterize with short-read seq due to repeat sequences
- Able to resolve structural rearrangements or homologous regions
- Able to read through an entire RNA transcript to determine the specific isoform
- Assists de novo genome assembly
- Lower per read accuracy
- Bioinformatic challenges, caused by coverage biases, high error rates in base allocation, scalability and limited availability of appropriate pipelines

**Hybrid Bioinformatic assembly:** Data generated from the same sample using long and short read sequencers

# Whole genome sequencing – Bacterial genomes - Antimicrobial resistance (AMR)

Both short vs long read sequencing



# Various steps showing SARS-CoV-2 sequencing using NGS platforms



# Oxford Nanopore sequencing technology



## Research Areas:

- Microbiology
- Microbiome
- Environmental
- Animal
- Infectious disease
- Human genomics
- Clinical research
- Cancer
- Transcriptome
- Populations genomics

## Applications:

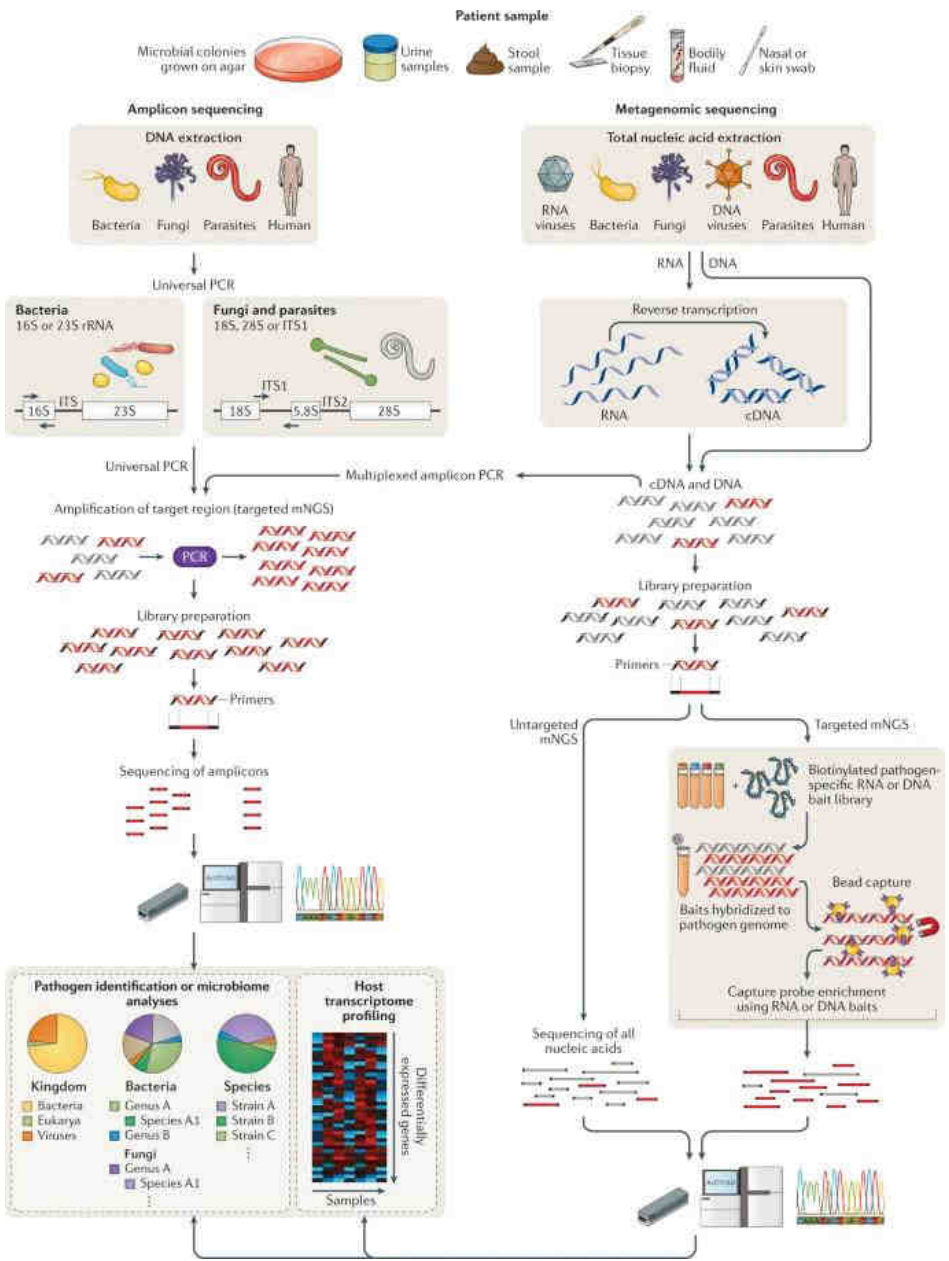
Nanopore sequencing offers advantages in all areas of research; includes **DNA**, **RNA** sequencing and gene expression.

<https://www.youtube.com/c/OxfordNanoporeTechnologies/playlists>

## Key considerations - library preparation and sequencing platform

- (a) Research question ? being asked
- (b) Sample type
- (c) Short-read or long-read sequencing
- (d) DNA or RNA sequencing – do you need to look at the genome or transcriptome?
- (e) Is the whole genome required or only specific regions?
- (f) Read depth (coverage) needed – experiment-specific
- (g) Extraction method
- (h) Sample concentration
- (i) Single end, paired end or mate pair reads
- (j) Specific read length required
- (K) Could samples be multiplexed ?
- (l) Bioinformatic tools – experiment dependent. Depending on the sample and the biological question, the entire process of sequence analysis can be adapted

# Infectious Disease Diagnostics

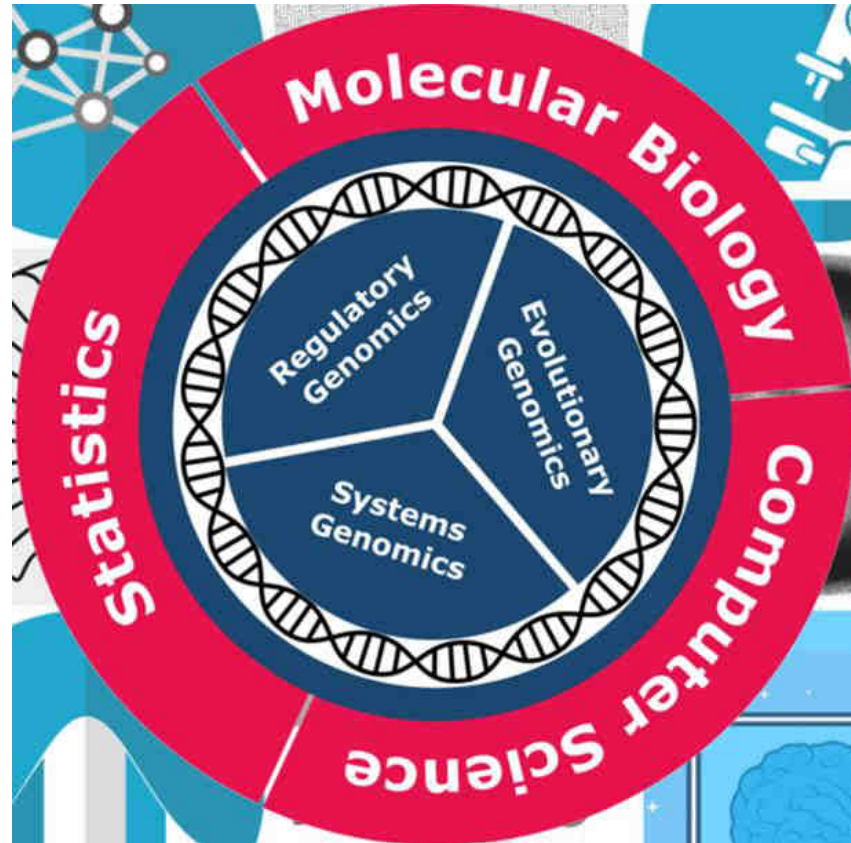


## Sequencing Methods

- Whole genome sequencing
- Targeted sequencing
- Metagenomics
- Rna sequencing
- Methylation sequencing

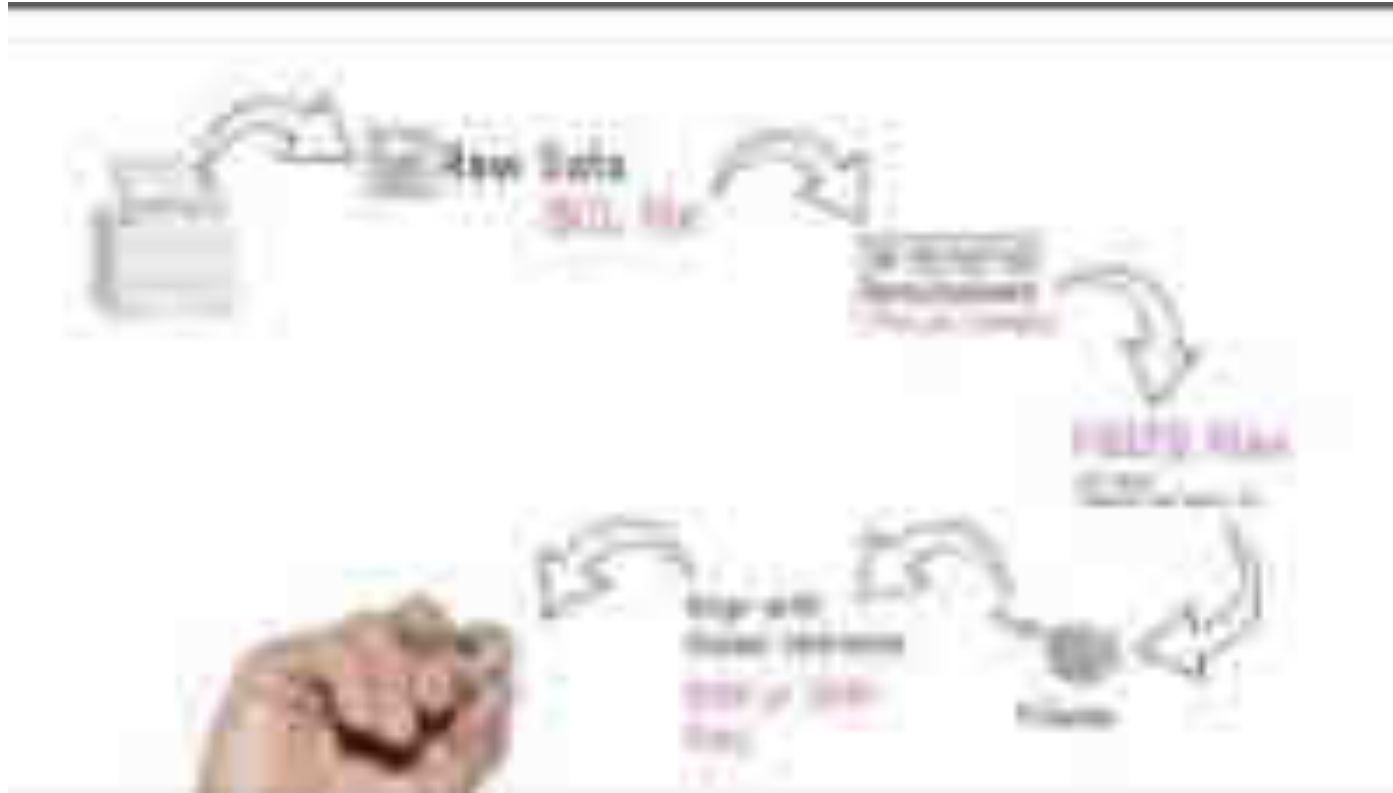
# Bioinformatics

- Bioinformatics and sequencing technologies have flourished in recent years.
- Bioinformatic deals with, how to mine, interpret, as well as utilize sequencing and experimental data, and convert them into useful information.





## NGS data analysis, Pipelines and Bioinformatics



<https://www.youtube.com/watch?v=RkttaYc8hfw>

[2021 Canadian Bioinformatics Workshop \(CBW\) series: Bioinformatics DotCa](https://www.youtube.com/channel/UCKbkfKk65PZyRCzUwXOJung/featured)  
<https://www.youtube.com/channel/UCKbkfKk65PZyRCzUwXOJung/featured>

## Any one can become a Bioinformatician

**Category 1:** "Biologists that use tools"- Biologists view the details of quantitative tools as largely unimportant. Rather the only thing that is important is 'biological questions'. Bioinformatics is about using computational approaches to study biological questions that would be impossible or infeasible to approach with other methods.

**Category 2:** Professions who develops methods and software tools for understanding biological data, in particular when the data sets are large and complex.

<https://www.youtube.com/c/SimonCockell/videos>

## Any one can become a Bioinformatician



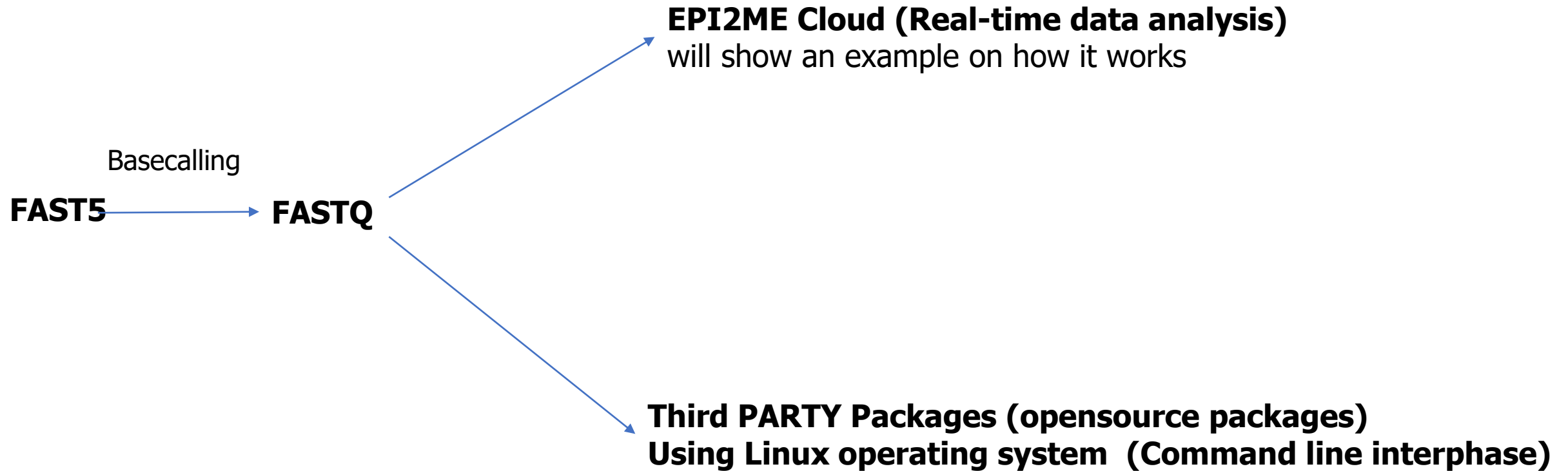
- Most bioinformatics (and most experts) are self-taught, and you can do the same – learn from online resources (Google and Youtube) (you cannot believe where you will reach....)
- Require curiosity and an interest in bioinformatics

# Oxford Nanopore Sequencing and data analysis



- Oxford Nanopore produces FAST5 (HDF5) files and/or FASTQ files.
- FAST5 files contain raw signal data that can be used for basecalling.
- FASTQ format that contain nucleotide sequence data (obtained from FAST5 Raw data)

# Data Analysis



# Common file systems, terminology and softwares/repositories

## File Systems

FAST5 (Nanopore)  
FASTQ  
FASTA  
BAM/SAM  
GTF/GFF/BED  
VCF

## Terminology

Pipeline  
MultiQC  
Reference sequence  
Consensus sequence  
De novo assembly  
Alignment to reference

## Softwares/repositories

Github  
Conda (Anaconda, Miniconda, Bioconda)  
Medaka  
Minimap2



# File Systems

FASTQ: Sequence (Fasta) with quality score

```
Identifier | @HWI-EAS209_0006_FC706VJ:5:58:5894:21141#ATCACG/1
Sequence  | TTAATTGGTAAATAAATCTCCTAATAGCTTAGATNTTACCTTNNNNNNNNNTAGTTTCTTGAGA
+ sign & identifier | +HWI-EAS209_0006_FC706VJ:5:58:5894:21141#ATCACG/1
Quality scores | efcfffffcfeefffcfffffddf`feed]` ]_Ba_^__ [YBBBBBBBBBBRTT\]] [ ] dddd`
```

Base T  
phred Quality ] = 29

FASTA: Text-based format for representing nucleotide sequences

```
>NG_008679.1:5001-38170 Homo sapiens paired box 6 (PAX6)
ACCCTCTTTTCTTATCATTGACATTTAAACTCTGGGGCAGGTCCTCGCGTAGAACGCGGCTGTCAGATCT
GCCACTTCCCCTGCCGAGCGGCGGTGAGAAGTGTGGGAACCGGCGCTGCCAGGCTCACCTGCCTCCCCGC
CCTCCGCTCCCAGGTAACCGCCCGGGCTCCGGCCCCGGCCCGGCTCGGGGCCCGCGGGGCCTCTCCGCTG
CCAGCGACTGCTGTCCCCAAATCAAAGCCCGCCCCAAGTGGCCCCGGGGCTTGATTTTTGCTTTTAAAAG
GAGGCATACAAAGATGGAAGCGAGTTACTGAGGGAGGGATAGGAAGGGGGGTGGAGGAGGGACTTGTCTT
TGCCGAGTGTGCTCTTCTGCAAAAGTAGCAAAATGTTCCACTCCTAAGAGTGGACTTCCAGTCCGGCCCT
GAGCTGGGAGTAGGGGGCGGGAGTCTGCTGCTGCTGTCTGCTAAAGCCACTCGCGACCGCGAAAAATGCA
GGAGGTGGGGACGCACTTTGCATCCAGACCTCCTCTGCATCGCAGTTCACGACATCCACGCTTGGGAAAG
TCCGTACCCGCGCCTGGAGCGCTTAAAGACACCCTGCCGCGGGTCCGGCGAGGTGCAGCAGAAGTTTCCC
GCGGTTGCAAAGTGCAGATGGCTGGACCGCAACAAAGTCTAGAGATGGGGTTCGTTTCTCAGAAAGACGC
```

# File Systems

**FAST5 (Nanopore)** : Raw reads from Oxford Nanopore sequencer (only for Nanopore)

**SAM/BAM** : Sequence Alignment Map (SAM) is a text-based format originally for storing biological sequences aligned to a reference sequence; The binary equivalent of a SAM file is a Binary Alignment Map (BAM) file

**GTF/GFF** : General Feature Format (GFF) file is a simple tab-delimited text file for describing genomic features.

**VCF** : Variant Call Format is text file used in storing gene sequence variations



## Terminology

**Pipeline :** Composed of a wide array of software algorithms to process raw sequencing data and generate a consensus and list of annotated sequence variants

**Reference sequences:** The Reference Sequence (RefSeq) database is an open access, annotated and curated collection of publicly available nucleotide sequences

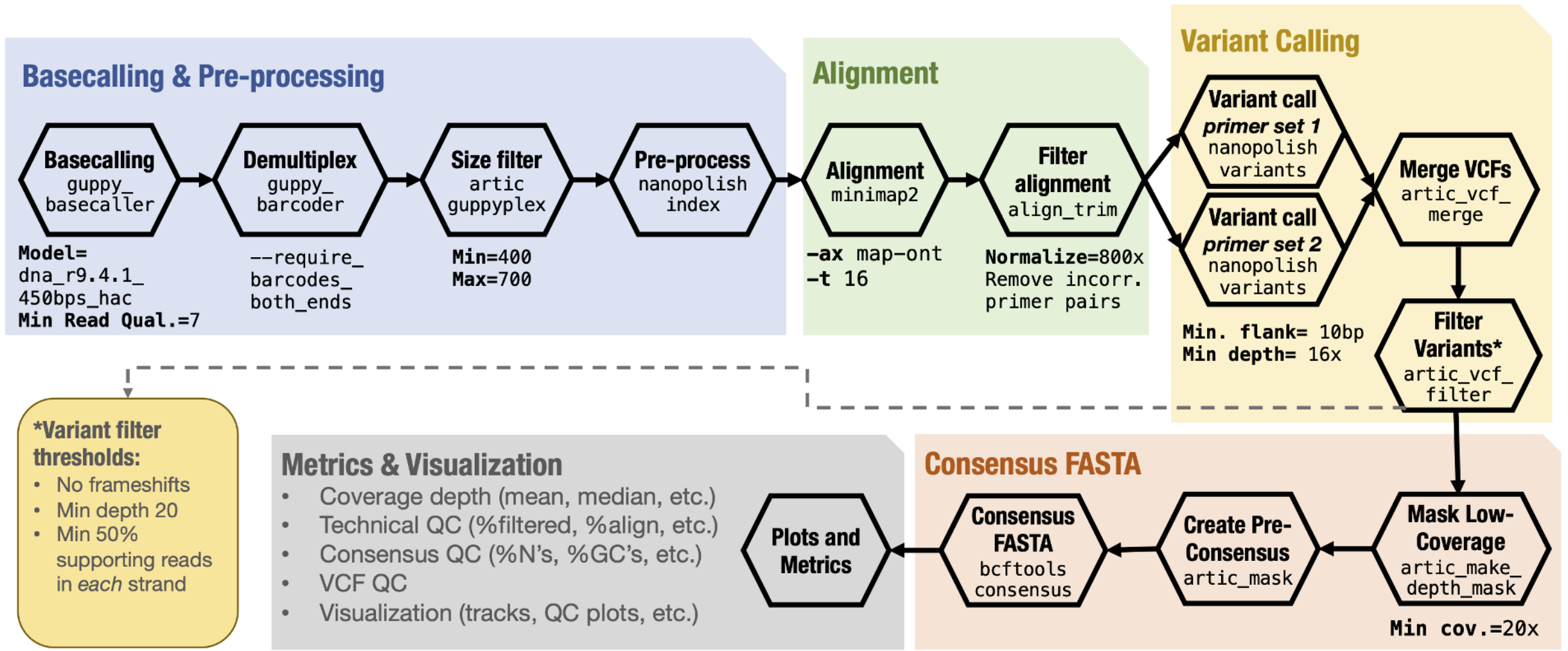
**MultiQC :** Aggregate results from bioinformatics analyses across many samples into a single report

**Sequence alignment :** Aligns the reads/sequences based in the similarities

**Consensus sequence:** A consensus sequence is a sequence of DNA represents aligned, related sequences.

**De novo assembly:** De novo sequence assemblers are a type of program that assembles short nucleotide sequences into longer ones without the use of a reference genome.

# SARS-CoV2 ARTIC Analysis Pipeline (Nanopolish)

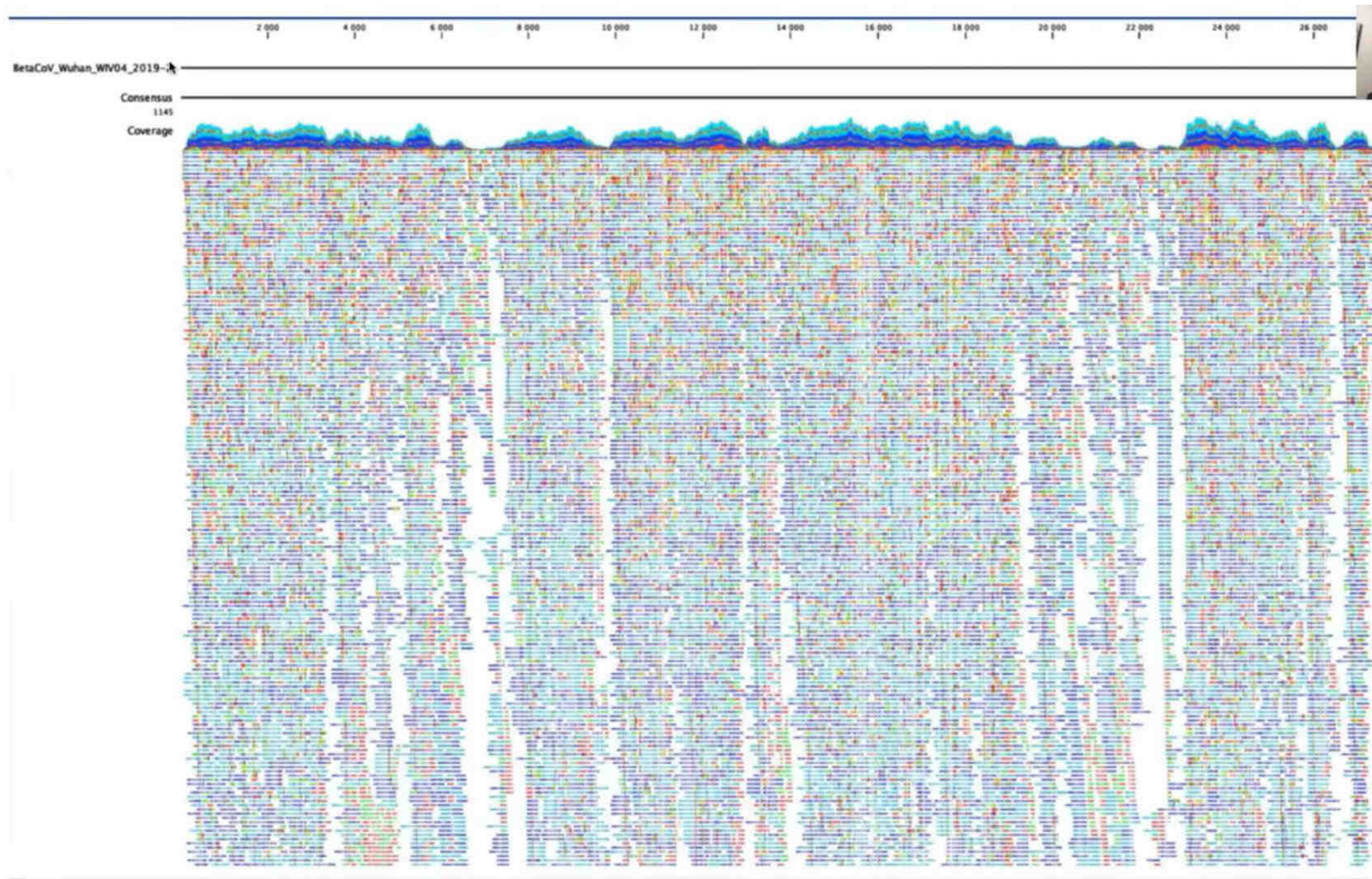


**Reference Genome:** Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1 (GenBank MN908947.3)

**Software versions:** guppy-GPU v3.4.4; minimap2 v2.17; samtools v1.9; bcftools v1.9; bedtools v2.27.0; python v3.6; nanopolish v0.13.1; muscle v3.8.31

**ARTIC pipeline cloned from:** <https://github.com/artic-network/fieldbioinformatics>

# Read alignment (Bowtie 2) : Wuhan reference sequence to the raw reads to get the final consensus seq.



## Softwares and repositories

**Github:** Millions of developers and companies build, ship, and maintain their software on GitHub—the largest and most advanced development platform in the world.

**Conda (Anaconda, Miniconda, Bioconda):** Conda is an open-source package management system and environment management system that runs on Windows, macOS, and Linux. Conda quickly installs, runs, and updates packages and their dependencies. Conda easily creates, saves, loads, and switches between environments on your local computer. It was created for Python programs but it can package and distribute software for any language.

**Minimap2:** is a versatile sequence alignment program

**Medaka:** is a tool to create consensus sequences and variant calls from nanopore sequencing data

# EPI2ME Cloud (Real-time data analysis)- No need bioinformatics

Bacterial sepsis/meningitis detection using real-time data analysis on EPI2ME cloud (16S amplicon sequencing)

The screenshot displays the EPI2ME Desktop Agent interface. The top navigation bar includes the EPI2ME logo, a cloud icon, and a 'PROFILE [LINH]' link. A progress indicator at the top shows three steps: 'SELECT SOURCE' (completed), 'CHOOSE ANALYSIS' (current step), and 'CONFIRM & START'. Below the progress bar, the 'FOLDERS' tab is active, showing a large dashed box with the text 'Drop files and folders here, or click to select.' A file selection box below this contains the path 'D:\Nanopore\_Microbiota\_Sepsis\_Projects\CSF\_Blood\_Sepsis\Flongle\_Sepsis\_Hanoi\_23.04.2021\Flongle\_Sepsis\_fastq\_pass\_all 200'. The left sidebar lists several 'Fastq 16S' files with their respective dates and sample IDs. The Windows taskbar at the bottom shows the search bar, application icons, and system tray information including the date and time (05:03 20/09/2021).

File Name	Date	Sample ID
Fastq 16S	Mon Aug 23 2021 15:34:51	285968
Fastq 16S	Fri May 07 2021 12:14:30	271510
Fastq 16S	Mon May 03 2021 15:03:08	271043
Fastq 16S	Mon May 03 2021 15:02:28	271042
Fastq 16S	Mon May 03 2021 14:57:32	271041

**Epi2ME Cloud example live demo**

# EPI2ME Cloud (Real-time data analysis)

EPI2ME Desktop Agent  
EPI2ME Agent Edit View Help

EPI2ME

PROFILE [LINH]

START AGAIN

Mon Aug 23 2021 15:34:51  
Fastq 16S  
v2021.03.05  
285968

Fri May 07 2021 12:14:30  
Fastq 16S  
v2021.03.05  
271510

Mon May 03 2021 15:03:08  
Fastq 16S  
v2021.03.05  
271043

Mon May 03 2021 15:02:28  
Fastq 16S  
v2021.03.05  
271042

Mon May 03 2021 14:57:32  
Fastq 16S  
v2021.03.05  
271041

1 SELECT SOURCE

2 CHOOSE ANALYSIS

3 CONFIRM & START

FAVOURITES ALL

ANALYSIS FASTA REFERENCE UPLOAD 2021.07.15

ANALYSIS FASTQ SV CALLER FOR HUMAN 2021.06.19

ANALYSIS FASTQ HUMAN ALIGNMENT GRCH38 2021.06.08

ANALYSIS FASTQ ANTIMICROBIAL RESISTANCE 2021.05.17

ANALYSIS FASTQ HUMAN EXOME 2021.05.15

ANALYSIS FASTQ WIMP ( HUMAN + VIRAL ) 2021.04.05

ANALYSIS FASTQ RNA CONTROL EXPERIMENT 2021.03.30

ANALYSIS FASTQ CONTROL EXPERIMENT 2021.03.30

ANALYSIS FASTQ CUSTOM ALIGNMENT 2021.03.25

ANALYSIS FASTQ QC + ARTIC + NEXTCLADE 1.0.4

ANALYSIS FASTQ 16S 2021.03.05

ANALYSIS FASTQ WIMP 2021.03.05

ANALYSIS FASTQ BARCODING 2021.03.05

← BACK NEXT →

Type here to search

Desktop 12°C 05:03 20/09/2021

# EPI2ME Cloud (Real-time data analysis)

The screenshot displays the EPI2ME Cloud interface. The browser address bar shows the URL <https://epi2me.nanoporetech.com/report-285968>. The page header includes the EPI2ME logo and navigation links for Dashboard, Workflows, and Software. A user profile icon is visible in the top right corner.

The main content area is divided into two sections. On the left, there is a table with a search filter and a list of taxa with their cumulative read counts. On the right, there is a phylogenetic tree visualization with a minimum abundance cutoff and a 'SHOW TOP N TAXA' dropdown.

**Table Data:**

Taxon	Cumulative Reads
Streptococcus pneumoniae	174
<b>Streptococcus mitis</b>	5
Streptococcus himalayensis	3
Streptococcus azizii	3
Streptococcus sanguinis	2
Streptococcus gordonii	2
Streptococcus marmotae	2
Neisseria gonorrhoeae	1
Neisseria cinerea	1
Streptococcus cristatus	1
Streptococcus equi	1
Streptococcus mutans	1

**Phylogenetic Tree Parameters:**

- MINIMUM ABUNDANCE CUTOFF: 3%, 1%, 0.5%, 0.1%, 0%
- SHOW TOP N TAXA: 10, 20, 30, 100, 200

**Phylogenetic Tree Labels:**

- Streptococcus pneumoniae
- Streptococcus azizii
- Streptococcus gordonii
- Streptococcus himalayensis
- Streptococcus marmotae
- Streptococcus mitis

The tree shows a root node on the left, with a horizontal line leading to a vertical line that branches into the taxa listed on the right. The taxa are ordered by their relative abundance in the sample.

At the bottom of the screen, the Windows taskbar is visible, showing the search bar, taskbar icons, and system tray information including the time 13:27 and date 20/09/2021.

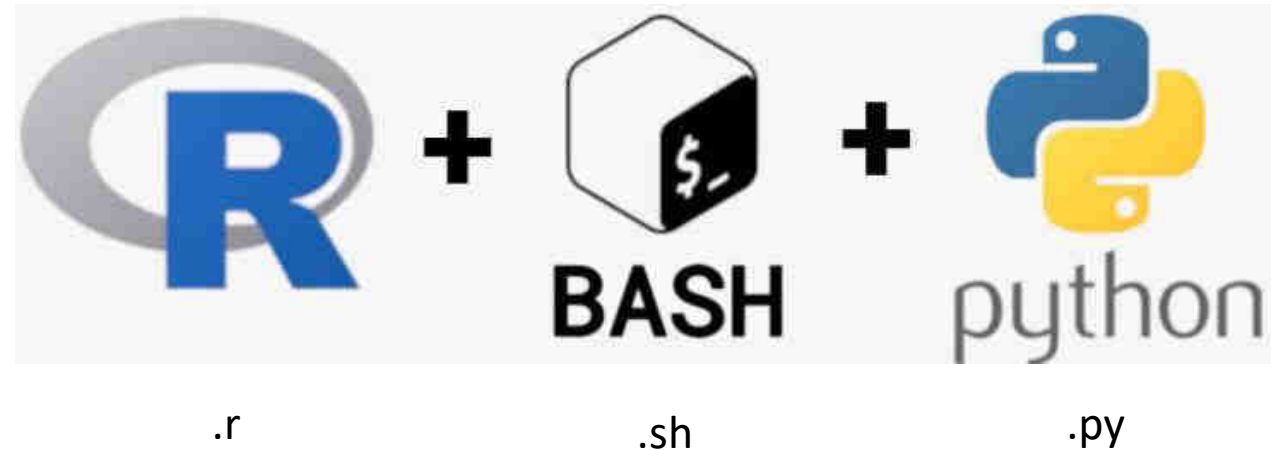
# Third Party packages (open source packages) on Linux operating system (Command line interphase)



Linux Operating System (Ubuntu)

```
srinivas@DESKTOP-H9B3T2K: ~  
(base) srinivas@DESKTOP-H9B3T2K:~$ ls  
Corona          bbc              index.html      test-nfcore     test.sh.save.2  
Miniconda2-latest-Linux-x86_64.sh  corona          miniconda2     test-runner.sh.save  test.sh.save.3  
bashrc         fieldbioinformatics  req.save       test.sh         vgcare  
bashrc.save   folder            test.sh.save   test.sh.save    windows  
bashrc.save.1 gabun_aligned.fasta.log  shared        test.sh.save.1  work  
(base) srinivas@DESKTOP-H9B3T2K:~$ x_
```

Most common languages used in the Third PARTY Packages (opensource packages)





# Fieldbioinformatics package on Github (SARS-CoV-2 data analysis)

The screenshot displays the GitHub repository page for `artic-network/fieldbioinformatics`. The repository is on the `master` branch, which is 401 commits ahead and 2 commits behind `zibraprojectmaster`. A recent merge pull request #68 is visible, merged on Jan 14 with 518 commits.

The repository structure includes the following files and folders:

File/Folder	Description	Time Ago
<code>artic</code>	patching version	8 months ago
<code>docs</code>	updating docs	8 months ago
<code>test-data</code>	forgot to add medaka consensus for tests...	17 months ago
<code>.gitignore</code>	updating gitignore with more test datasets	2 years ago
<code>.travis.yml</code>	updating pytest decorator and adding CL flag for specifying workflow	2 years ago
<code>CHANGELOG</code>	merging CHANGELOG and HISTORY	2 years ago
<code>LICENSE</code>	update to python3 - lose docker implementatio	3 years ago
<code>README.md</code>	Update README.md	12 months ago
<code>ansible</code>	add lab-on-an-ssd -> ansible link	2 years ago
<code>environment.yml</code>	replacing artic multiqc fork with the vanilla one in the dep list	8 months ago
<code>mkdocs.yml</code>	updating docs	13 months ago
<code>requirements.txt</code>	deprecating <code>plot_amplicon_depth</code> script in favour of multiqc report	8 months ago
<code>setup.py</code>	adding code to produce stats from <code>align_trim</code> and <code>vcf_check</code> reports; u...	8 months ago
<code>test-runner.sh</code>	make <code>--medaka-model</code> a required flag if <code>--medaka set</code>	8 months ago

The 'Languages' section shows the following distribution:

Language	Percentage
Python	95.7%
Shell	2.6%
Makefile	1.1%
R	0.6%

The Windows taskbar at the bottom shows the system clock as 05:30 on 20/09/2021, with a temperature of 12°C.

<https://github.com/artic-network/fieldbioinformatics>

# Difference between Linux and Windows

S.NO	Linux	Windows
1.	Linux is a open source operating system.	While windows are the not the open source operating system.
2.	Linux is free of cost.	While it is costly.
3.	It's file name case-sensitive.	While it's file name is case-insensitive.
4.	In linux, monolithic kernel is used.	While in this, micro kernel is used.
5.	Linux is more efficient in comparison of windows.	While windows are less efficient.
6.	There is forward slash is used for Separating the directories.	While there is back slash is used for Separating the directories.
7.	Linux provides more security than windows.	While it provides less security than linux.
8.	Linux is widely used in hacking purpose based systems.	While windows does not provide much efficiency in hacking.

## What after performing the Bioinformatic analysis

- Deposit the RAW data and the final sequences to the public repositories.
- If metagenomics is there any pathogen detected – is it novel ?
- Combine the phenotype and genotype data
- Perform statistical (in case there are groups)/phylogenetic analysis (Viruses)
- Check the results
- Write a manuscript.... edit...edit.....Final draft
- Submit and publish

## Next-generation sequencing data repositories (to upload and download raw data)

<https://www.gisaid.org/> (Repository)

<https://www.ncbi.nlm.nih.gov/sra> (NCBI SRA Toolkit to download the data directly)

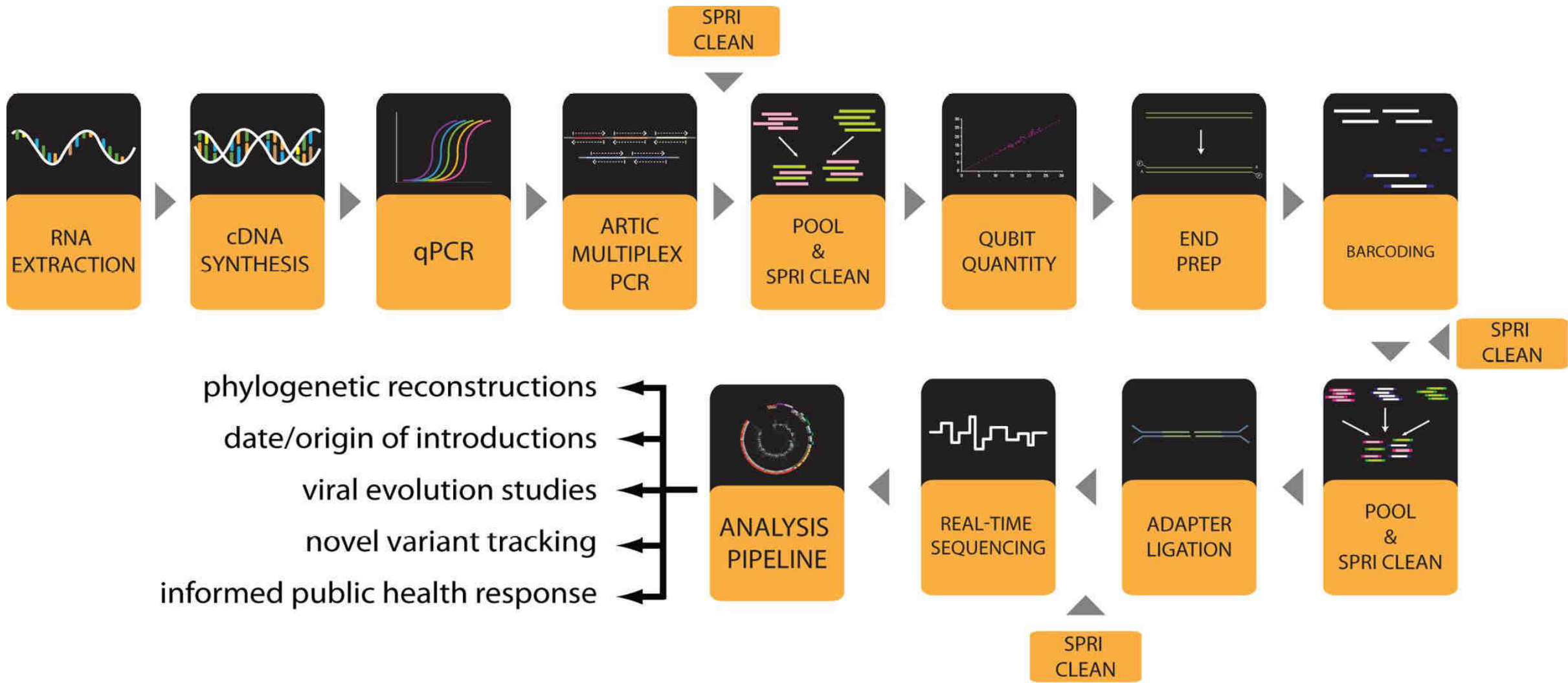
<https://www.ncbi.nlm.nih.gov/genbank>

<https://www.insdc.org/>

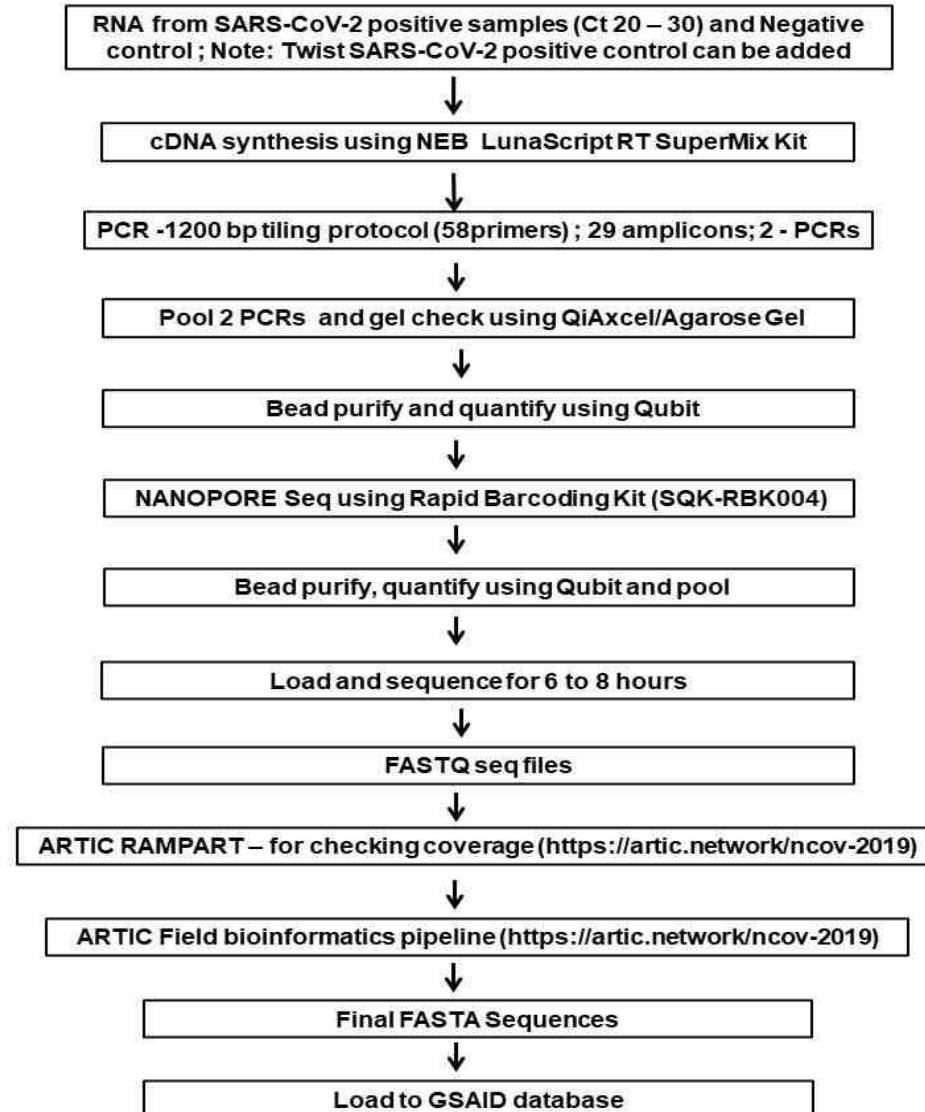
<https://www.ebi.ac.uk/ena/browser/home>

<https://www.ebi.ac.uk/submission/>

If you don't have the your own raw data and want to work on the pipelines: Data can be downloaded from NCBI or other databases .....



# Flow diagram of the SARS-CoV-2 Whole-genome sequencing using Oxford Nanopore





THANK YOU

## Linux : important links for using linux operating system

<https://www.youtube.com/c/SimonCockell/videos>

[https://www.youtube.com/watch?v=v\\_1zB2WNN14](https://www.youtube.com/watch?v=v_1zB2WNN14)

<https://www.youtube.com/watch?v=V1y-mbWM3B8>

<https://www.youtube.com/watch?v=x5MhydijWmc>

<https://www.youtube.com/watch?v=G36l1iqDZig>

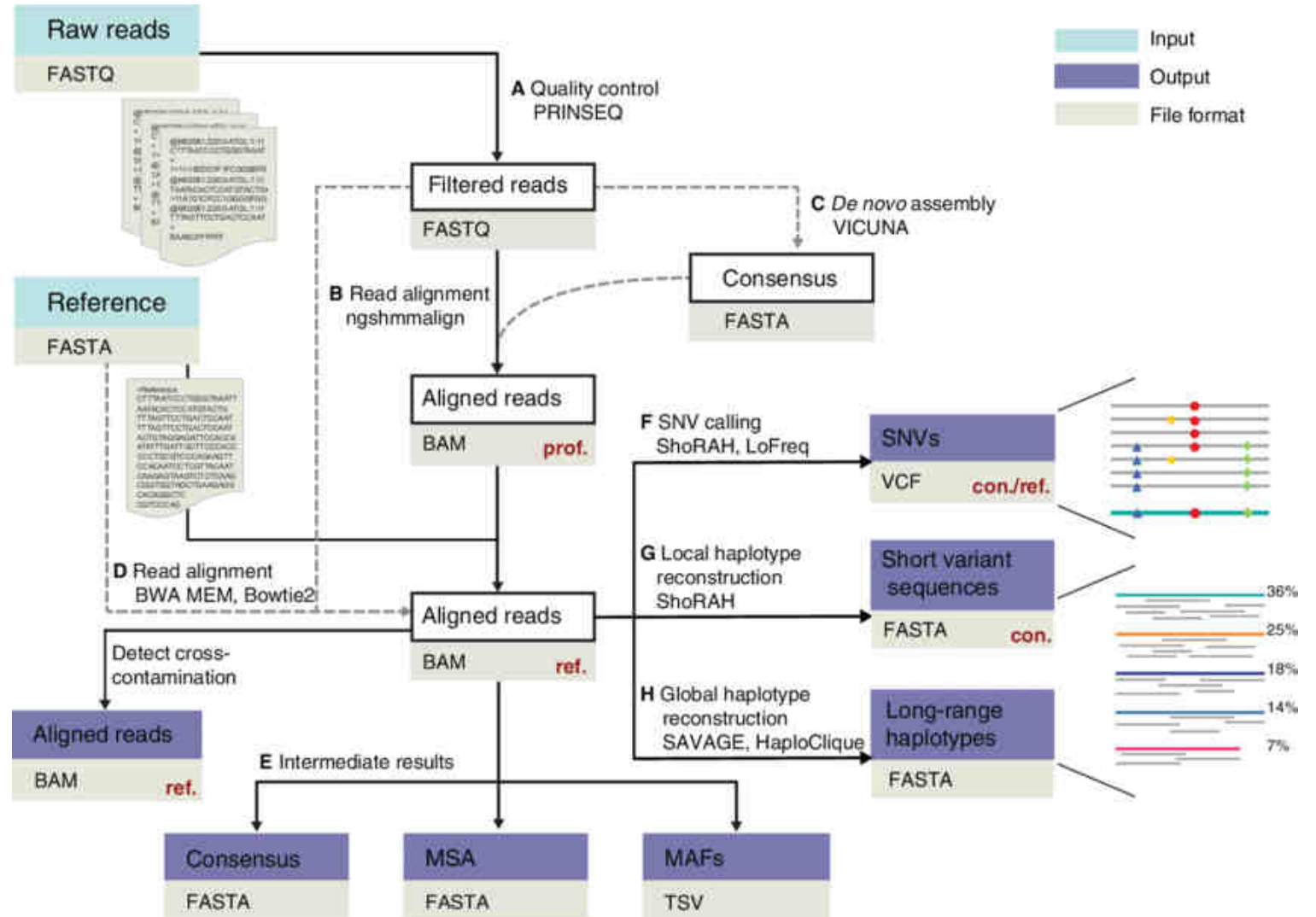


# V-Pipe: Viral Quasiespecies analysis pipeline (example)

Illumina  
Nanopore

FAST5 reads  
Nanopore

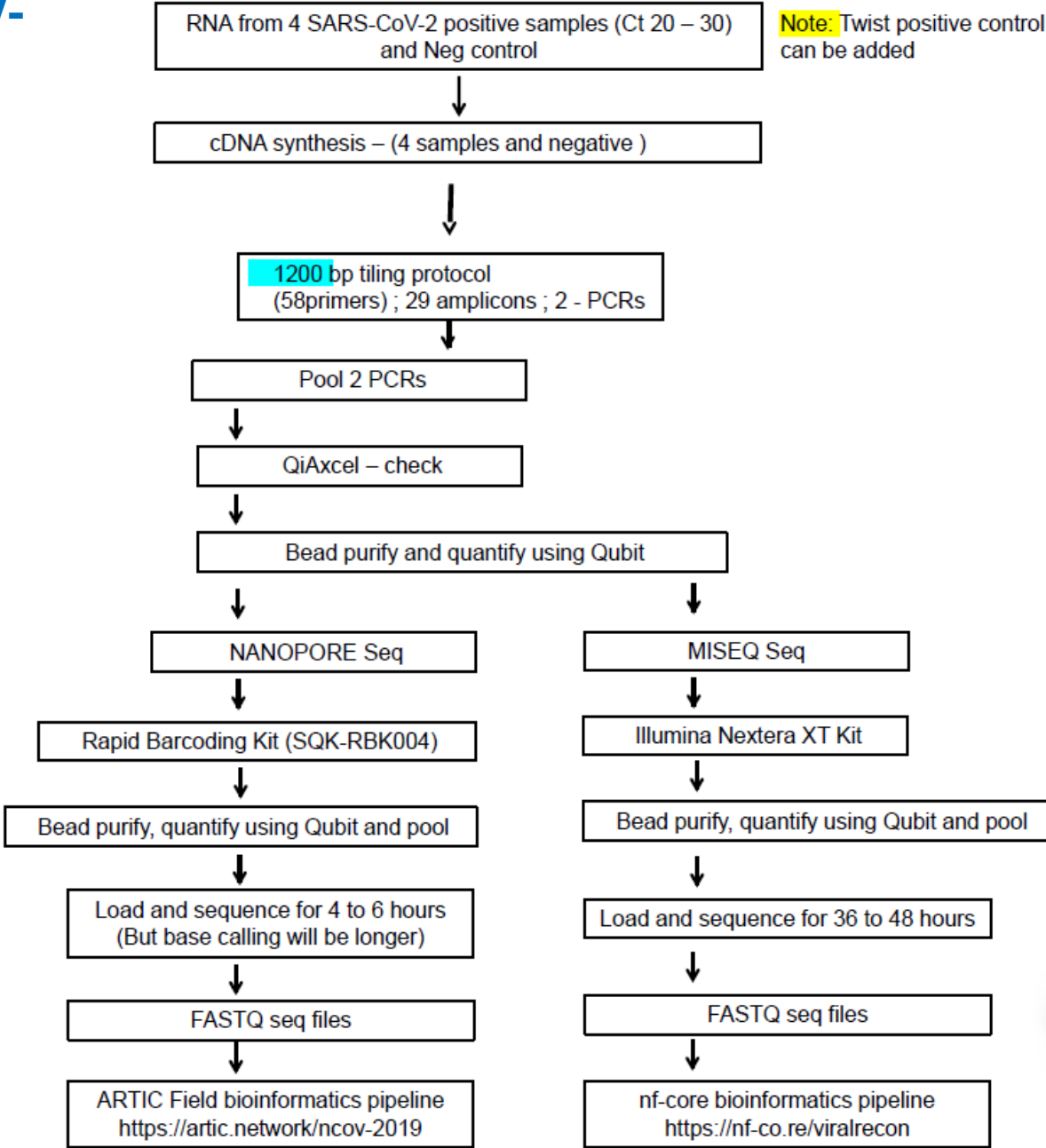
→

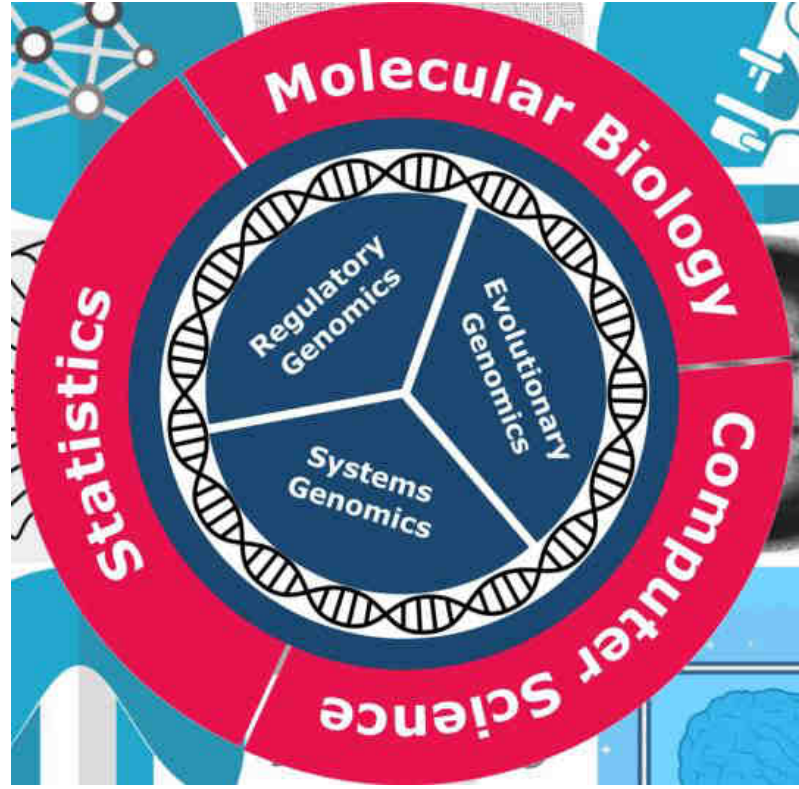


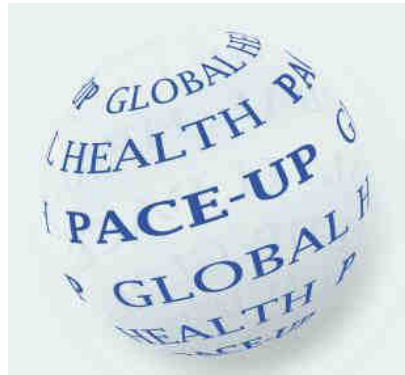
- Pipeline
- Softwares used
- File formats

<https://github.com/cbg-ethz/V-pipe>

# Flow chart showing Sars-CoV-2 sequencing using both illumina and Nanopore







---

# Workshop

## Molecular Diagnostic Techniques (Genomics and Bioinformatics – Infectious diseases) module-2

Dr. Srinivas Reddy Pallerla, Mr. Tung Tran and Ms. Linh Le  
Prof. Dr. Thirumalaisamy P. Velavan, University Hospital, Tübingen, Germany  
Prof. Dr. Daniel Olusola Ojurongbe, Nigeria  
Prof. Dr. Mohamed Osman, University of Khartoum, Sudan

# Overview

- Introduction
  - Who is who? Who knows what?
- How will the training be organized?
  - What is in the “additional information”?
- Overview of NGS-nanopore work-flow as described in the protocol
- Theory of nanopore-sequencing
  - Material + devices

“Our goal is to enable the analysis of anything, by anyone, anywhere”.  
From Nanopore

<https://www.youtube.com/c/OxfordNanoporeTechnologies/featured>



<https://nanoporetech.com/>

# Nanopore technique – Devices available



Flongle

← Adapting MinION and GridION for smaller, routine tests and analyses. Low plex targeted sequencing, RNA isoform analysis, and quality control applications.

[View Flongle](#)



MinION Mk1B

Access the benefits of nanopore technology from just \$1,000 – suitable for targeted sequencing and gene expression studies.

[View MinION](#)



MinION Mk1C

Integrated sequencing and analysis in a powerful handheld device – suitable for targeted sequencing and gene expression studies.

[View MinION](#)



GridION

From genome assembly to gene expression, run multiple experiments on-demand using 5 independent MinION flow cells.

[View GridION](#)



PromethION

Flexible, population-scale sequencing using up to 48 independent, high-capacity flow cells – complete genomic and transcriptomic characterisation of large sample numbers.

[View PromethION](#)



VoITRAX

→ Automated sample extraction and library preparation.

[View VoITRAX](#)

<https://nanoporetech.com/how-nanopore-sequencing-works>

Nanopore sequencing, the only technology that offers scientific researchers:

- Sequence any DNA/RNA fragment length from short to ultra-long  
Characterise more genetic variation, versatile to broad applications
- Direct sequencing of native DNA/RNA  
Generate content-rich data, including methylation
- Data available in real time  
Rapid insights, and analyses that can respond to results in real time
- Scalable from portable devices to ultra-high throughput desktop devices  
Sequence anything, anywhere
- No capital investment required  
Accessible and cost effective
- Simple & rapid, or automated, library prep  
Easy to use and versatile



# Overview

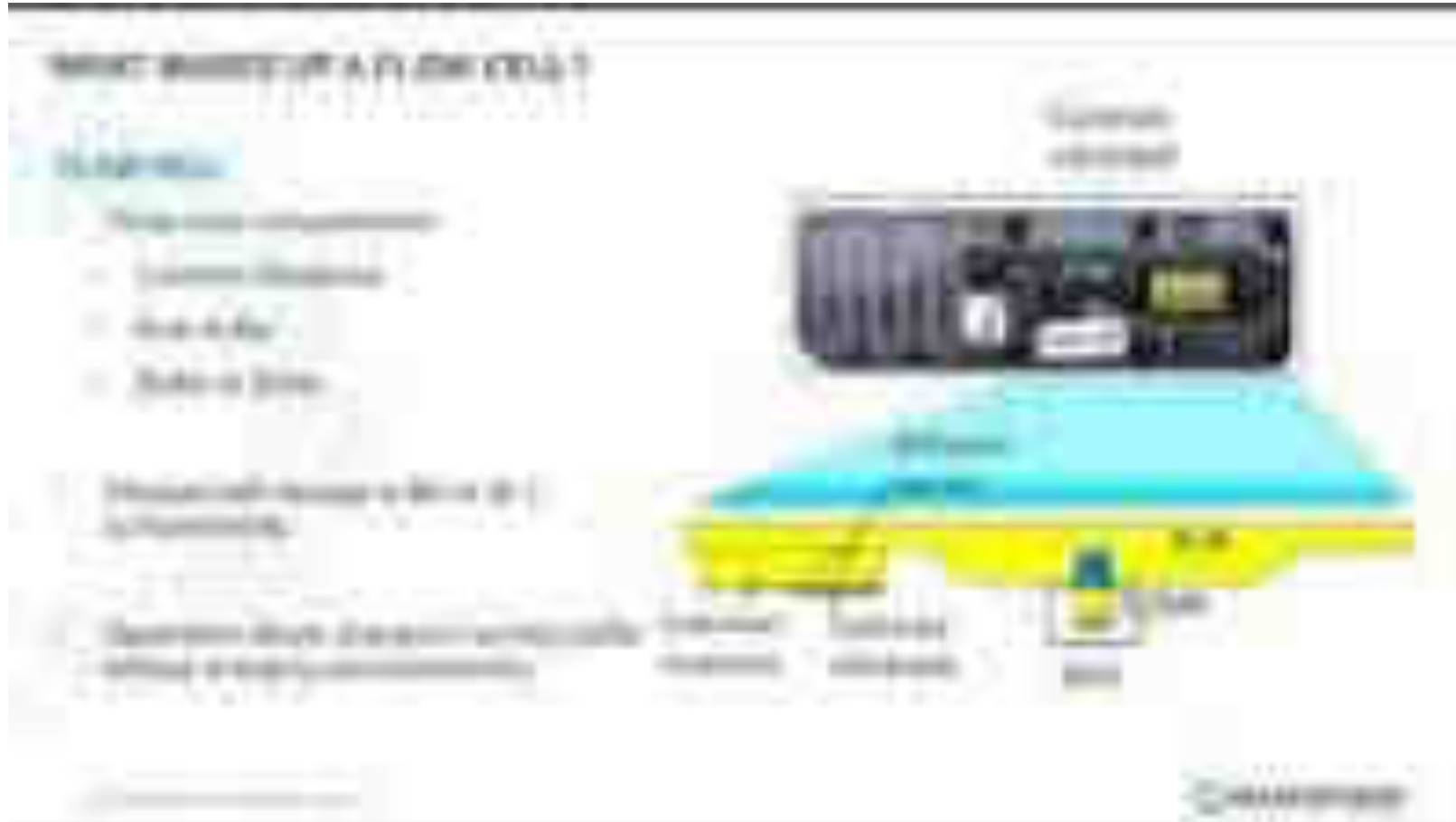
<https://www.youtube.com/c/OxfordNanoporeTechnologies/featured>

All the updates on the Nanopore sequencing advancements, kits, improvements, ..... please follow to adapt to the updates..... that will reduce the costs, new kits, new products.....



**VoTRAX – a small device designed to perform library preparation automatically**





<https://nanoporetech.com/how-nanopore-sequencing-works>

# Nanopore technique – MinION Mk1B



**Read length**

Nanopores read the length of DNA or RNA presented to them – from short to ultra-long (longest >4 Mb)

**Suitable applications include**

- Whole genomes/exomes
- Metagenomics
- Targeted sequencing
- Whole transcriptome (cDNA)
- Smaller transcriptomes (direct RNA)
- Multiplexing for smaller samples

**Dimensions**


- Size: W 105 mm, H 23 mm, D 33 mm
- Weight: 87 g

**High yields**

Up to 50 Gb per MinION Flow Cell /  
2.8 Gb per Flongle Flow Cell\*

<https://nanoporetech.com/products/minion#tabs-0=mk1b>

# Nanopore technique – MinION Mk1C



**Read length**

Nanopores read the length of DNA or RNA presented to them – from short to ultra-long (longest >4 Mb)

**Suitable applications include**

- Whole genomes/exomes
- Metagenomics
- Targeted sequencing
- Whole transcriptome (cDNA)
- Smaller transcriptomes (direct RNA)
- Multiplexing for smaller samples

**Dimensions**

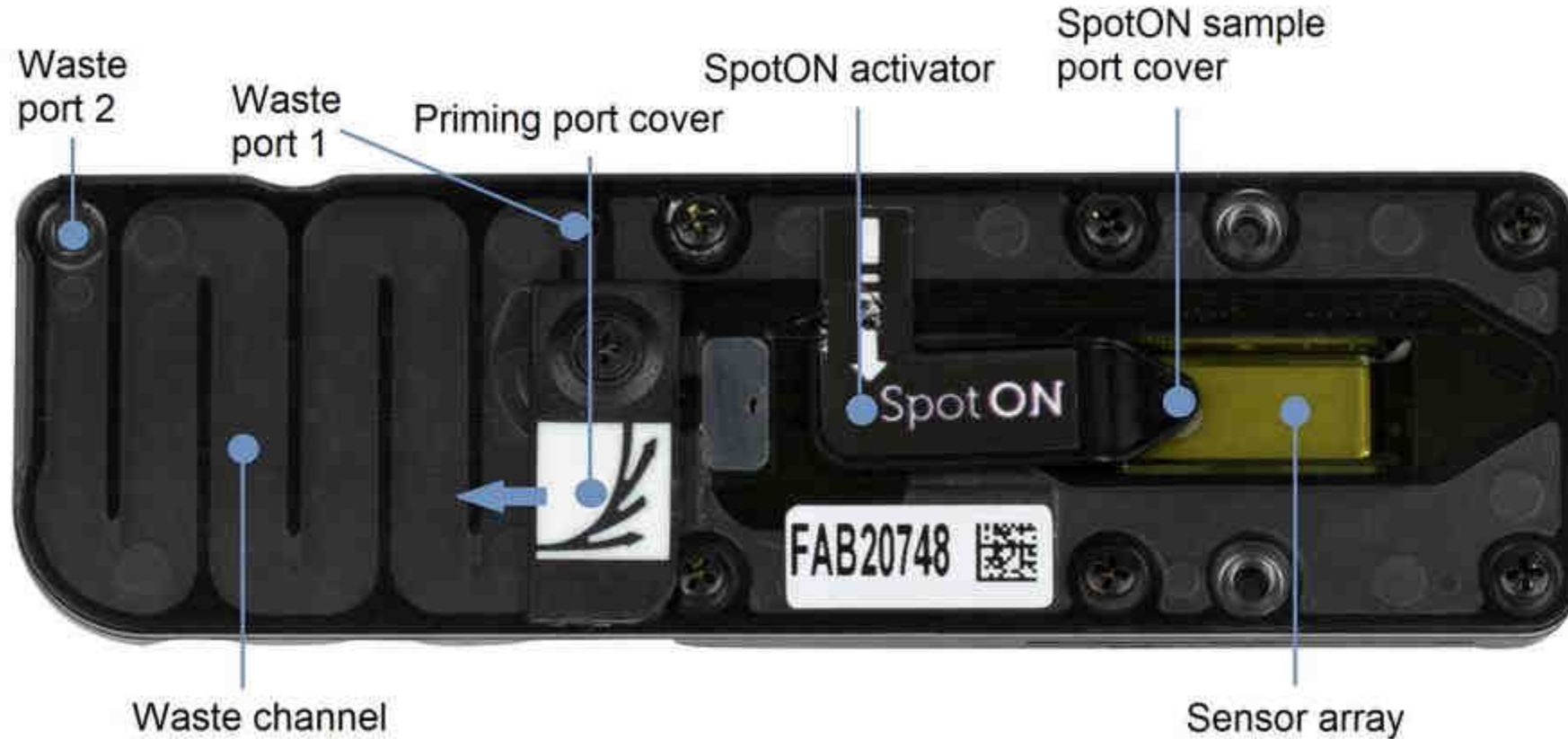
- Size: W 140 mm, H 30 mm, D114
- Weight: 450 g

**High yields**

Up to 50 Gb per MinION Flow Cell / 2.8 Gb per Flongle Flow Cell\*

\* Theoretical max output when system is run for 72 hours (or 16 hours for Flongle) at 420 bases / second. Outputs may vary according to library type, run conditions, etc.

# Nanopore technique – Flowcell



[https://community.nanoporetech.com/protocols/flow-cell-wash-kit-exp-wsh004/v/wfc\\_9120\\_v1\\_revb\\_08dec2020/flushing-a-minion-gridion-flow-cell?devices=minion](https://community.nanoporetech.com/protocols/flow-cell-wash-kit-exp-wsh004/v/wfc_9120_v1_revb_08dec2020/flushing-a-minion-gridion-flow-cell?devices=minion)

# Nanopore technique – Flongle (adapter for MinION for smaller, single-use flowcells)

## Dimensions

- Size: W 105, H 23, D 8 mm
- Weight: 20 g

## Read length

Nanopores read the length of DNA or RNA presented to them – from short to ultra-long (longest >4 Mb)



Up to 2.8 Gb yield\*

- Suitable for smaller assays
- Multiplexing options available
- \* Theoretical max output when system is run for 16 hours at 420 bases / second. Outputs may vary according to library type, run conditions, etc.

# Nanopore technique – Flowcell



The Flongle Flow Cell can generate up to 2.8 Gb of data enabling direct, real-time DNA & cDNA sequencing on smaller, single-use flow cells.



The MinION Flow Cell can generate up to 50 Gb of data for sequencing DNA, cDNA or native RNA in real-time.



The PromethION Flow Cell can generate up to 290 Gb for sequencing DNA, cDNA or native RNA in real-time.

<https://nanoporetech.com/flow-cells-and-nanopores>



# What is in the additional information ?

## 1. Theory

- 1.5 Links to get familiar with technology and experiments that you may plan in future using NANOPORE
  - > <https://nanoporetech.com/>
  - > <https://nanoporetech.com/nanopore-sequencing-data-analysis>
  - > <https://www.youtube.com/c/OxfordNanoporeTechnologies/playlists>

# What is in the additional information ?

## 1. Theory

- 1.5 Links to get familiar with technology and experiments that you may plan in future using NANOPORE
  - > <https://nanoporetech.com/>
  - > <https://nanoporetech.com/nanopore-sequencing-data-analysis>
  - > <https://www.youtube.com/c/OxfordNanoporeTechnologies/playlists>

# Nanopore technique

- Video : Introduction to nanopore sequencing

<https://www.youtube.com/watch?v=qzusVw4Dp8w&list=PLxpxXZj-IgXyR0it4QjoBETC5JeEanW-o&index=2>

# Nanopore technique

- Oxford Nanopore Website :

<https://nanoporetech.com/>

# Nanopore technique – Kits

Different kits used in the workflow:

- cDNA-synthesis – LunaScript RT SuperMix Kit by NEB
- Multiplex PCR – Q5 Hot Start HF Polymerase by NEB  
dNTP mix, 10mM by Thermo Fisher Scientific  
(+ in-house Primer Pool)
- Purification – Agencourt Ampure XP Beads by Beckman Coulter
- Quantification – Qubit dsDNA BR Kit by Thermo Fisher Scientific
- Library preparation – Rapid Barcoding Kit by Oxford Nanopore  
Flow Cell Priming Kit by Oxford Nanopore
- Flowcell washing – Flowcell Wash Kit by Oxford Nanopore

# Nanopore technique – Oxford Nanopore Kits

- Many different kits available, for an overview check: <https://store.nanoporetech.com/sample-prep.html>
- ONT-Kits used in the workflow:
  - Library preparation – Rapid Barcoding Kit by Oxford Nanopore  
Flow Cell Priming Kit by Oxford Nanopore
  - Flowcell washing – Flowcell Wash Kit by Oxford Nanopore

# Nanopore technique – Rapid Barcoding Kit

## Rapid Barcoding Kit

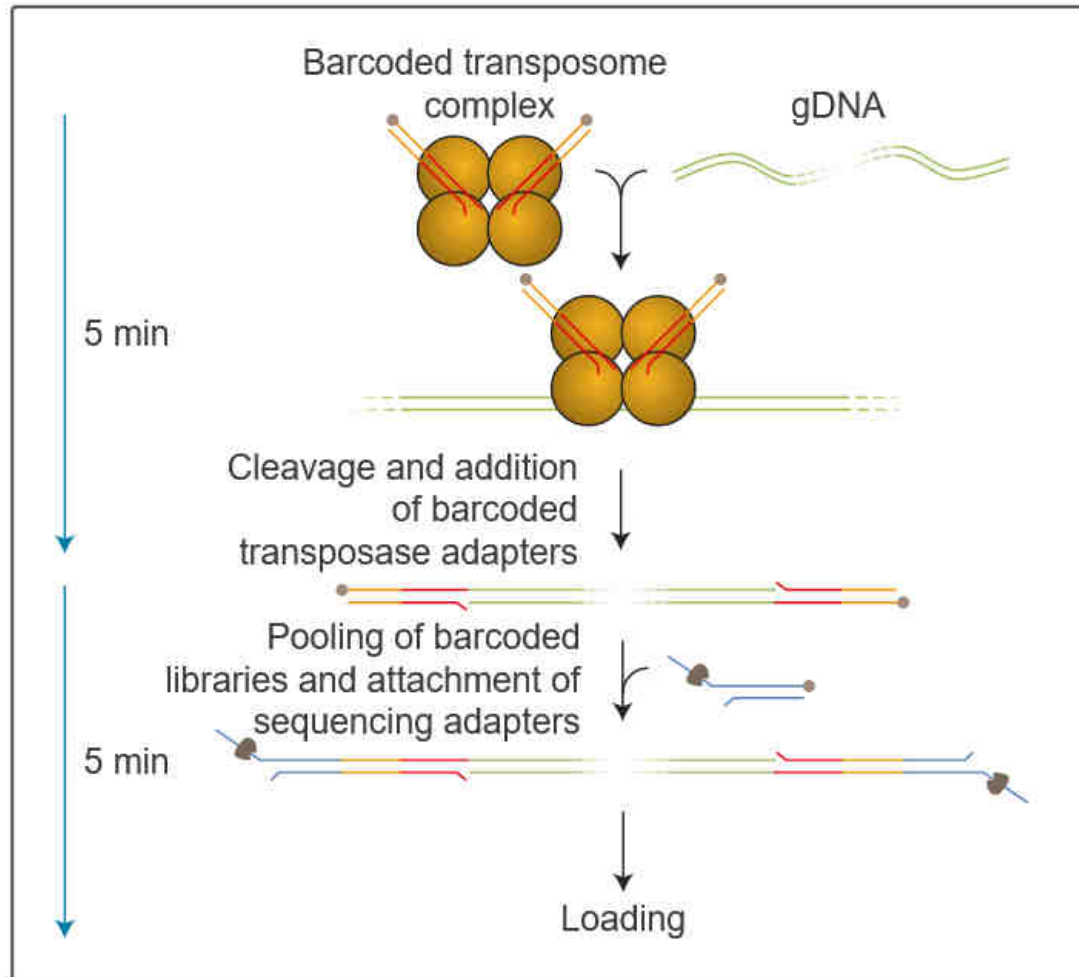
SQK-RBK004

- Simple and rapid library preparation, with barcoding for up to 12 gDNA samples
- Shipped at 2–8°C
- Long-term storage -20°C



<https://store.nanoporetech.com/rapid-barcoding-kit.html>

# Nanopore technique – Rapid Barcoding Kit



<https://store.nanoporetech.com/rapid-barcoding-kit.html>

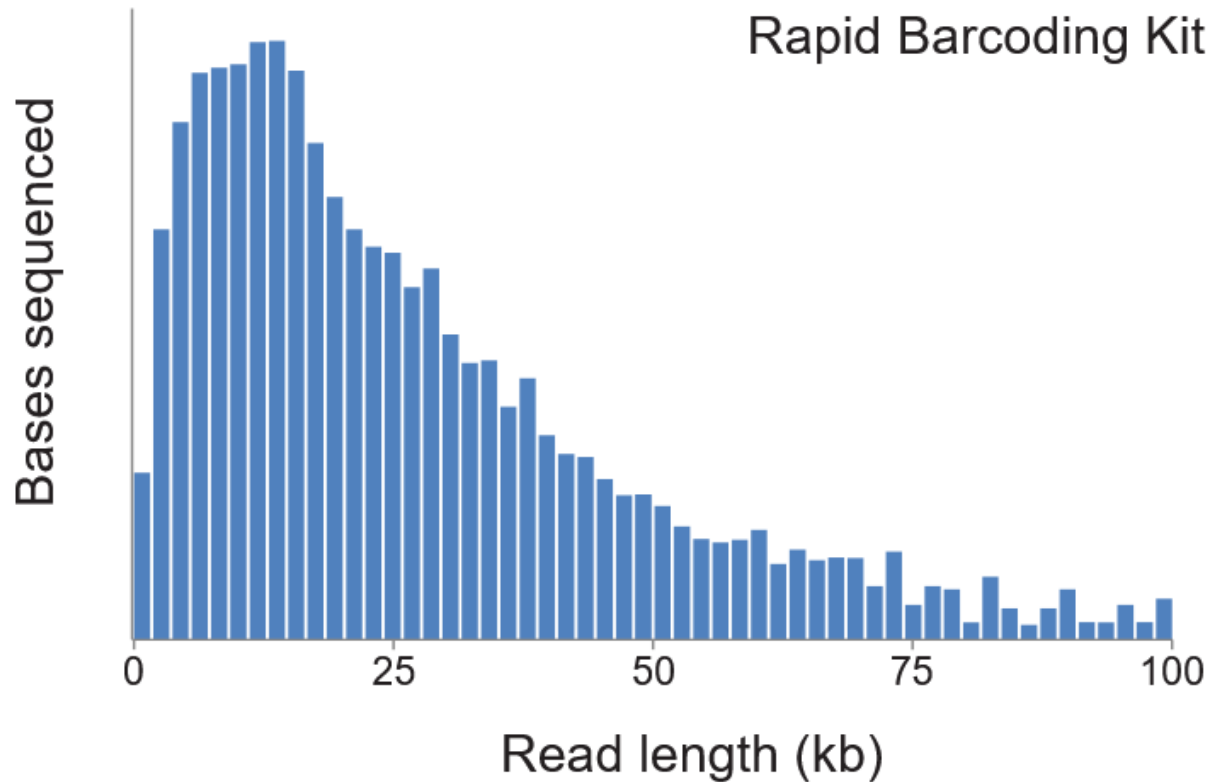
The Rapid Barcoding Kit generates barcoded sequencing libraries from extracted gDNA in 10 minutes using a simple 2-step protocol.

At the heart of the kit is a transposase which simultaneously cleaves template molecules and attaches barcoded tags to the cleaved ends: there are 12 unique barcoded tags in the kit.

Barcoded samples are pooled and Rapid Sequencing Adapters are then added to the tagged ends.



# Nanopore technique – Rapid Barcoding Kit



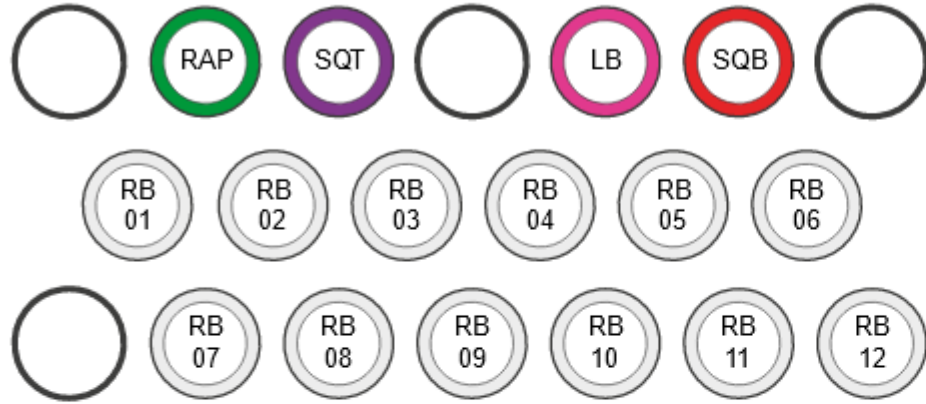
*Typical read length histogram when preparing an E.coli library with the Rapid Barcoding Kit.*

<https://store.nanoporetech.com/rapid-barcoding-kit.html>

Due to the simple nature of the workflow and the fact that little sample manipulation is required (e.g. minimal pipetting steps and no clean-ups), some very long reads can be achieved with this kit, despite the required transposase fragmentation. However, in order for long reads to be observed in sequencing, long fragments need to be present in the sample in the first place.

# Nanopore technique – Rapid Barcoding Kit

The Rapid Barcoding Kit contains twelve unique barcodes and sufficient reagents to generate six sequencing libraries.



RAP : Rapid adapter

SQT : Sequencing tether

LB : Loading beads

SQB : Sequencing buffer

RB01 : Fragmentation Mix RB 1

RB02 : Fragmentation Mix RB 2

RB03 : Fragmentation Mix RB 3

RB04 : Fragmentation Mix RB 4

RB05 : Fragmentation Mix RB 5

RB06 : Fragmentation Mix RB 6

RB07 : Fragmentation Mix RB 7

RB08 : Fragmentation Mix RB 8

RB09 : Fragmentation Mix RB 9

RB10 : Fragmentation Mix RB 10

RB11 : Fragmentation Mix RB 11

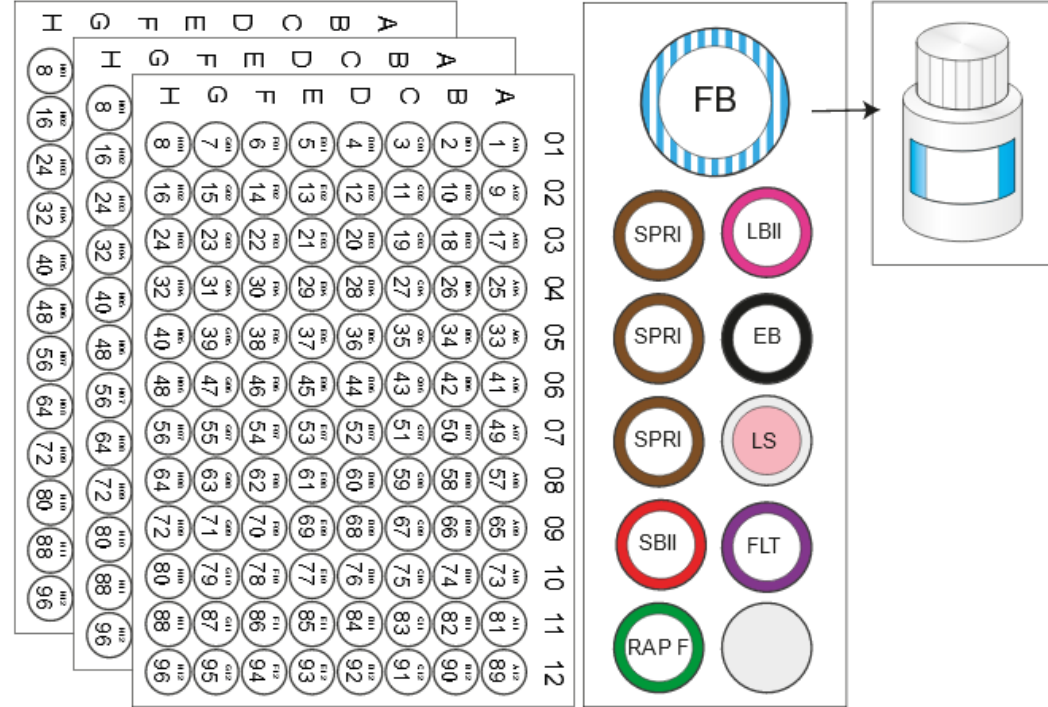
RB12 : Fragmentation Mix RB 12

<https://store.nanoporetech.com/rapid-barcoding-kit.html>

# Nanopore technique – Rapid Barcoding Kit 96

## ← Rapid Barcoding Kit 96

SQK-RBK110.96



FB : Flush Buffer  
 SPRI : SPRI Beads  
 SBII : Sequencing Buffer II  
 RAP F : Rapid Adapter F

LBII : Loading Beads II  
 EB : Elution Buffer  
 LS : Loading Solution  
 FLT : Flush Tether

<https://store.nanoporetech.com/rapid-barcoding-kit-1.html>

# Nanopore technique – Flow Cell Wash Kit

## ◀ Flow Cell Wash Kit

EXP-WSH004



- Shipped at 2–8°C
- Long-term storage -20°C

<https://store.nanoporetech.com/flow-cell-wash-kit-r9.html>

# Nanopore technique – Flowcell



- Never freeze!!
- Room temperature: flow cells can be stored, unopened, at room temperature for one month.
- 2-8°C: flow cells can be stored, unopened, at 2-8°C for 12 weeks.

<https://store.nanoporetech.com/flow-cell-r9-4-1.html>

# Nanopore technique – Flowcell

How to load the flowcell:

<https://www.youtube.com/watch?v=Pt-iaemrM88>

# Nanopore technique – MinION Mk1B

Presentation of MinION Video:

[https://www.youtube.com/watch?v=1\\_mER5qmaVk&list=PLxpxXZj-IgXyR0it4QjoBETC5JeEanW-o&index=12](https://www.youtube.com/watch?v=1_mER5qmaVk&list=PLxpxXZj-IgXyR0it4QjoBETC5JeEanW-o&index=12)

# Nanopore technique – MinION Mk1C

Presentation of MinION-Mk1C Video:

<https://www.youtube.com/watch?v=A1NLE0Jbvo8&list=PLxpxXZj-IgXyR0it4QjoBETC5JeEanW-o&index=15>



# Nanopore technique – Flongle

Presentation of Flongle Video:

<https://www.youtube.com/watch?v=Ov889BEkh5I&list=PLxpxXZj-IgXyR0it4QjoBETC5JeEanW-o&index=16>

# SARS-CoV-2: Nanopore MinION sequencing

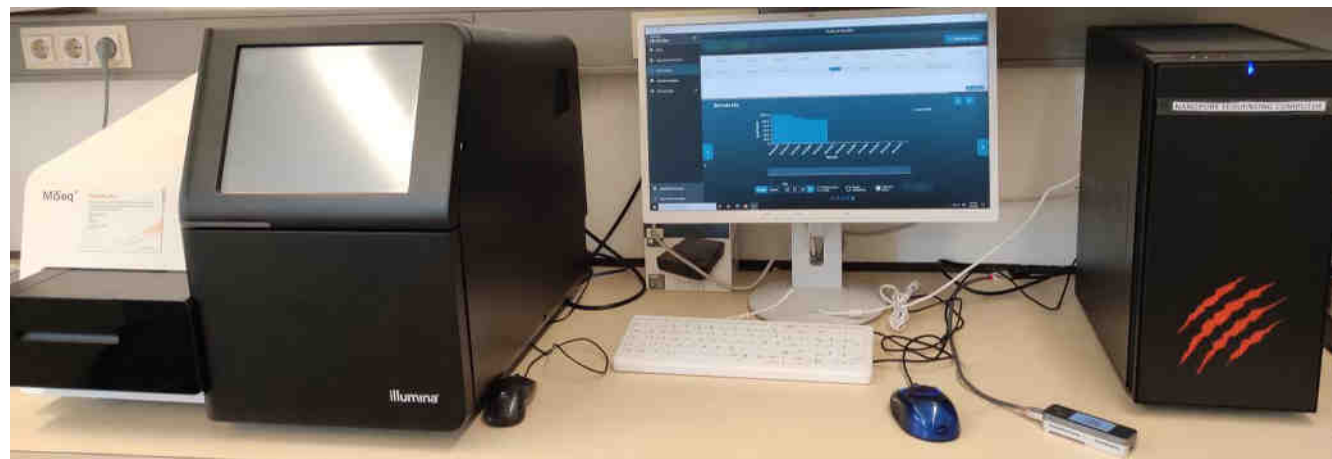
**Aim:** Sustainable capacities to be built place at CERID

## **Training module:**

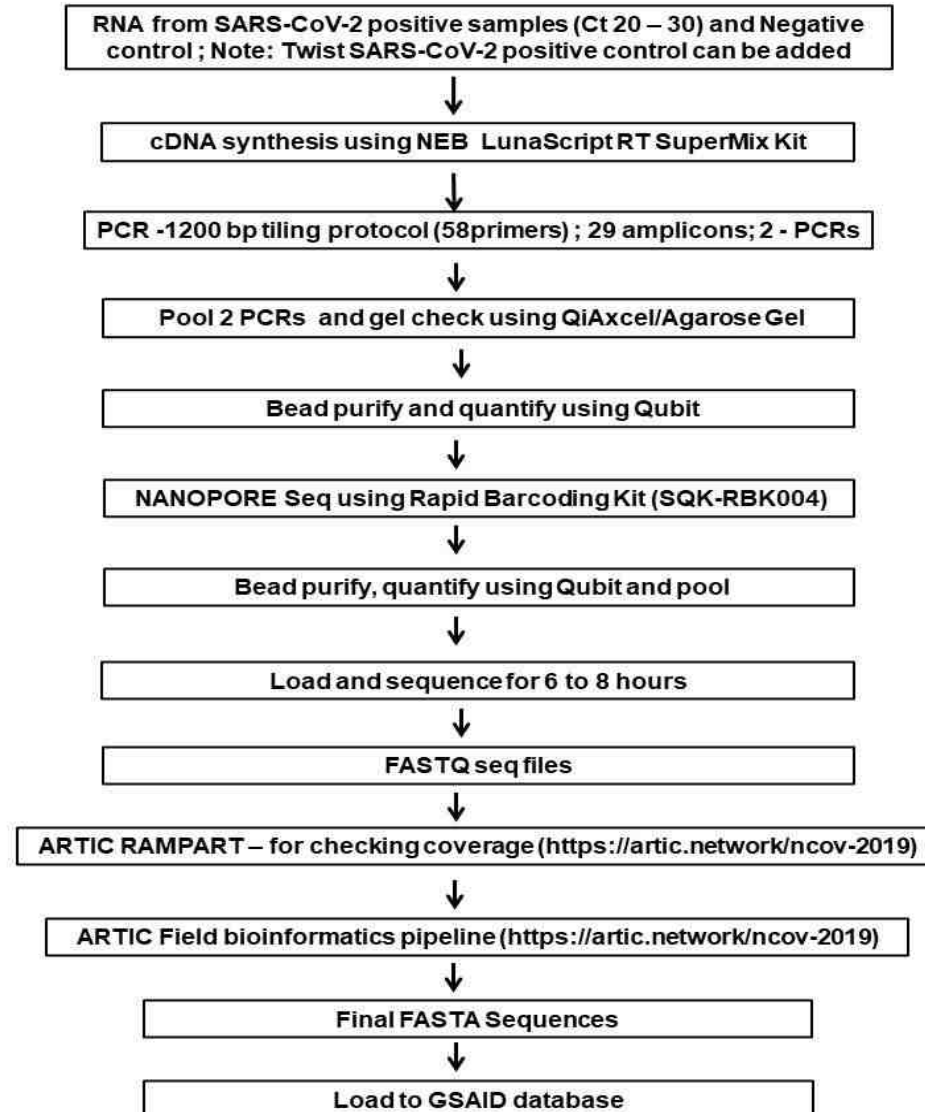
Theory : NGS and Oxford nanopore methodology

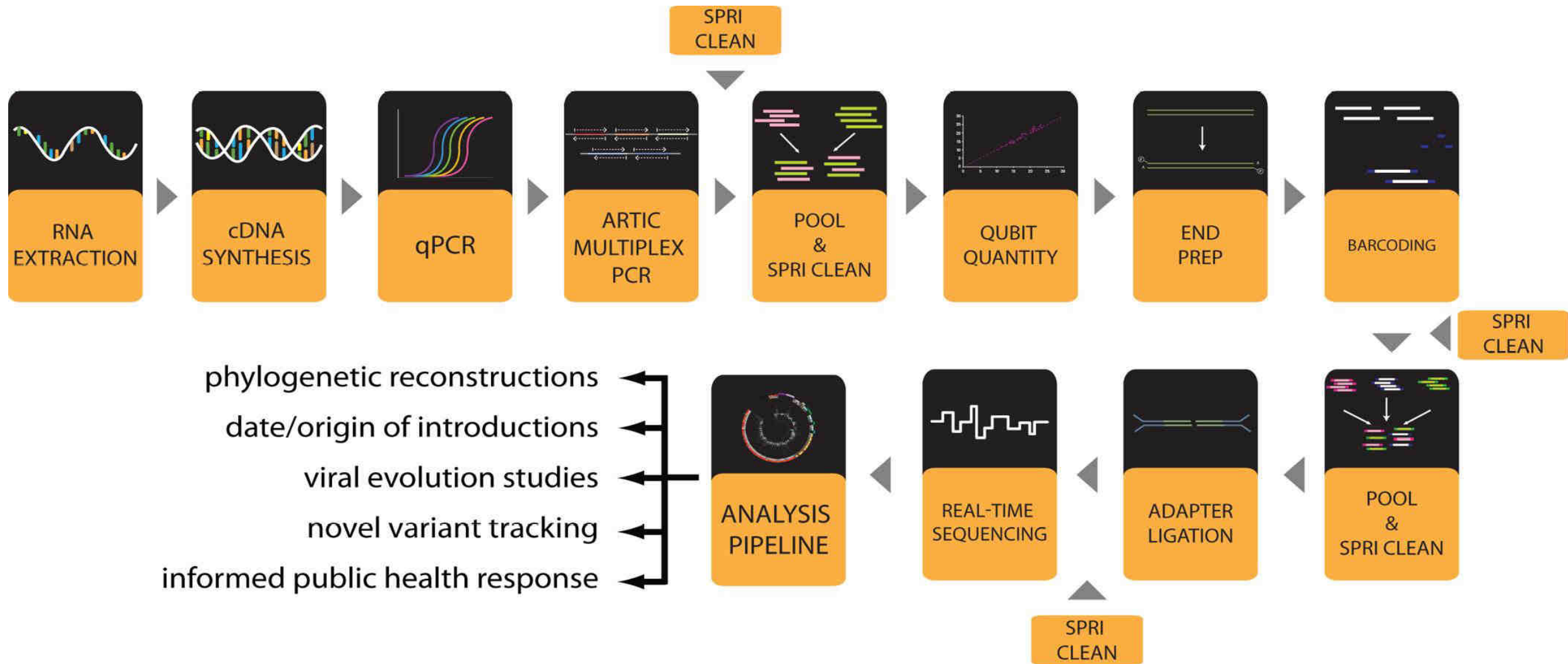
Wet-lab : SARS-CoV-2 sequencing

Dry-lab : LINUX and Bioinformatics pipelines

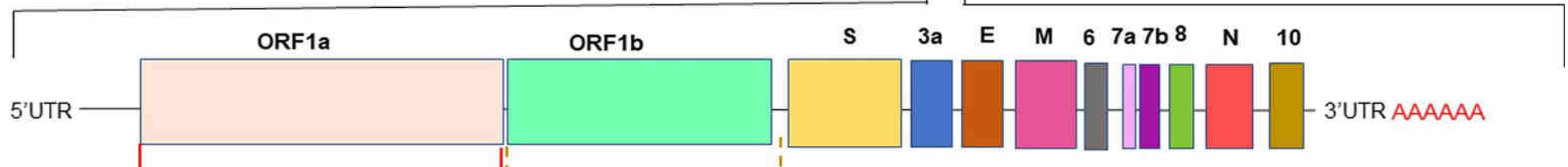
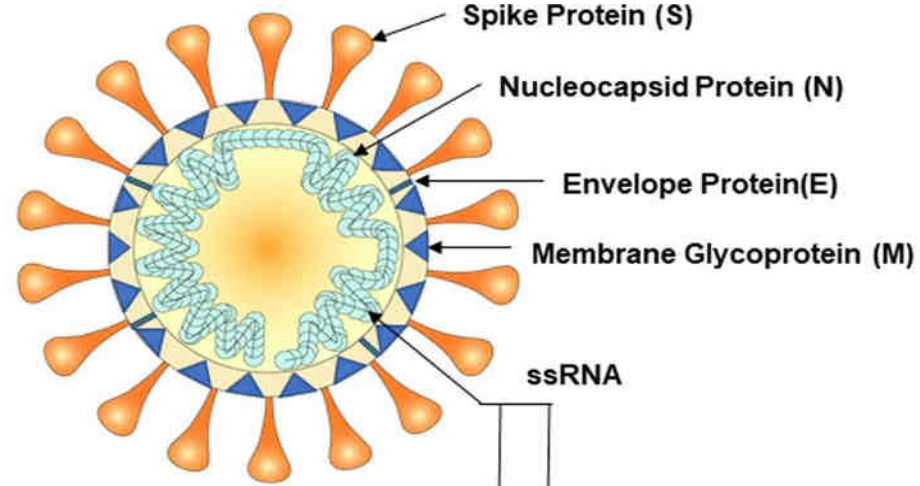


# Flow diagram of the SARS-CoV-2 Whole-genome sequencing using Oxford Nanopore

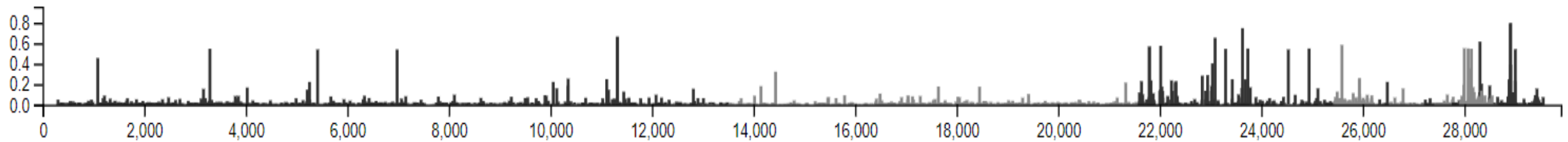




# SARS-CoV-2 genome – 29.9 kb



Diversity



# How concerning is the SARS-CoV-2 variants of concern ?

- Increased transmissibility;
- Increased morbidity and mortality;
- Evade detection by diagnostic tests;
- Decreased neutralizing to antibodies (e.g., convalescent plasma or monoclonal antibodies);
- Evade natural immunity (e.g., causing reinfections);
- Infect vaccinated individuals;
- Increased risk - multisystem inflammatory syndrome or long COVID;
- Increased affinity for particular demographic or clinical groups, such as children or immunocompromised individuals.

# Global - genomic surveillance of SARS-CoV-2

GISAID (Global Initiative on Sharing All Influenza Data)

hCoV-19 data sharing via GISAID

**3,545,029**  
submissions

- <https://outbreak.info/>
- <https://cov-lineages.org/index.html>
- <https://nextstrain.org/ncov/global>

- Global/Country/State/City to local
- Circulating variants
- New viral lineages
- Variants of interest, under investigation, of concern
- Distribution over time and space
- No of mutation accumulating
- Much more.....

# What is in the additional information ?

## 1. Theory

- 1.1 SARS-CoV-2
  - > Presentation given on 15.09.2021 (pdf)
- 1.2 COVID-19 Genomic Epidemiology Toolkit by CDC
  - > <https://www.cdc.gov/amd/training/covid-19-gen-epi-toolkit.html>
- 1.3 Details about Nanopore Technology -> workshop-presentations (pdf)
  - > <https://nanoporetech.com/how-nanopore-sequencing-works>
  - > <https://nanoporetech.com/nanopore-sequencing-data-analysis>
- 1.4 Links SARS-CoV-2 sequencing principles videos Oxford Nanopore:
  - > <https://nanoporetech.com/resource-centre/bioinformatics-workflows-sars-cov-2-raw-nanopore-reads-consensus-genomes-using>
  - > <https://www.youtube.com/watch?v=rYaFcDE-Ewg>



# What is in the additional information ?

## 2. Wet-lab

- 2.1 Protocol! (pdf)
- 2.2 Video links on Oxford Nanopore Flow cell introduction, Priming and loading flow cell
  - > [https://www.youtube.com/watch?v=zC6lAtzqi\\_k](https://www.youtube.com/watch?v=zC6lAtzqi_k)
  - > <https://www.youtube.com/watch?v=Pt-iaemrM88>
  - > <https://www.youtube.com/watch?v=IknVaEnuDz0&t=604s>
- 2.3. Flow cell wash protocol (pdf)
- 2.4 Methods manuscript, literature protocol (pdf)

# What is in the additional information ?

## 3. Dry-lab

### 3.1 Protocol! (pdf)

Protocol NGS\_WGS\_SARS-CoV-2\_1200bp amplicons using Nanopore\_V1.2.pdf

### 3.2 Introduction to bioinformatics

-> <https://www.youtube.com/c/SimonCockell/videos>

### 3.3 Linux

-> linuxcommands (pdf)

-> Links on Linux operating system

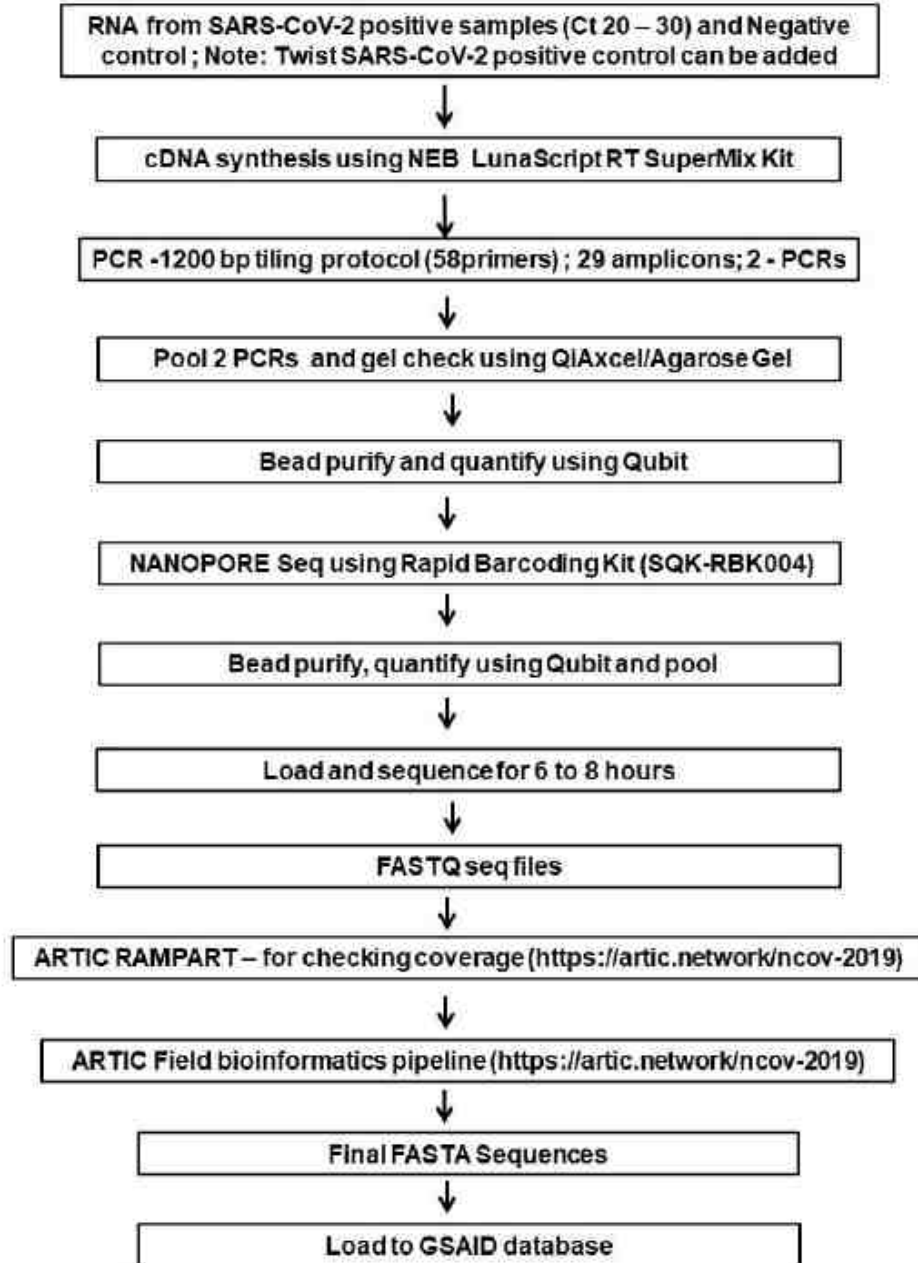
[https://www.youtube.com/watch?v=v\\_1zB2WNN14](https://www.youtube.com/watch?v=v_1zB2WNN14)

<https://www.youtube.com/watch?v=V1y-mbWM3B8>Ubuntu

<https://www.youtube.com/watch?v=x5MhydiWmc>

<https://www.youtube.com/watch?v=G36I1iqDZig>

# NGS-nanopore WORK-FLOW Overview





Alexander von Humboldt  
Stiftung/Foundation



EBERHARD KARLS  
UNIVERSITÄT  
TÜBINGEN



---

# Workshop

## Molecular Diagnostic Techniques (Genomics and Bioinformatics – Infectious diseases)

### Module-3

Dr. Srinivas Reddy Pallerla, Mr. Tung Tran and Ms. Linh Le  
Prof. Dr. Thirumalaisamy P. Velavan, University Hospital, Tübingen, Germany  
Prof. Dr. Daniel Olusola Ojurongbe, Nigeria  
Prof. Dr. Mohamed Osman, University of Khartoum, Sudan

# SARS-CoV-2 - Nanopore sequencing training module

## 1. Theory module

- Introduction into NGS technology and the library preparation
- Experimental design
- Standard operating procedures
- Introduction to Oxford Nanopore technology and its application

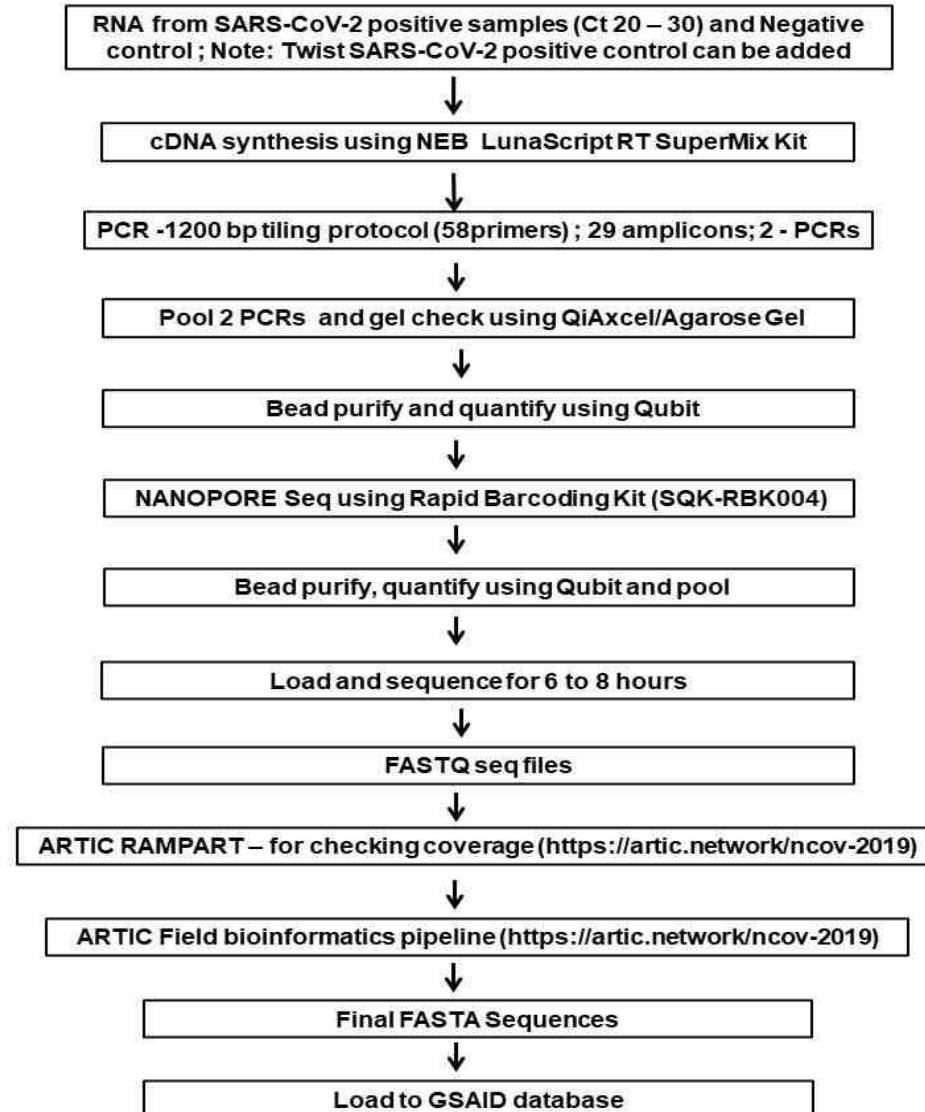
## 2. Wet-lab module specific for SARS-CoV-2

- NGS library preparation and validation.
- Multiplexing and barcoding
- Quantify and quality control of the prepared libraries

## 3. Dry-lab module

- Introduction to the bioinformatics pipelines
- NGS data analysis using Linux operating system
- Basic Linux commands for beginners
- SARS-CoV-2 data analysis using ARTIC bioinformatic pipelines (<https://artic.network/ncov-2019>)

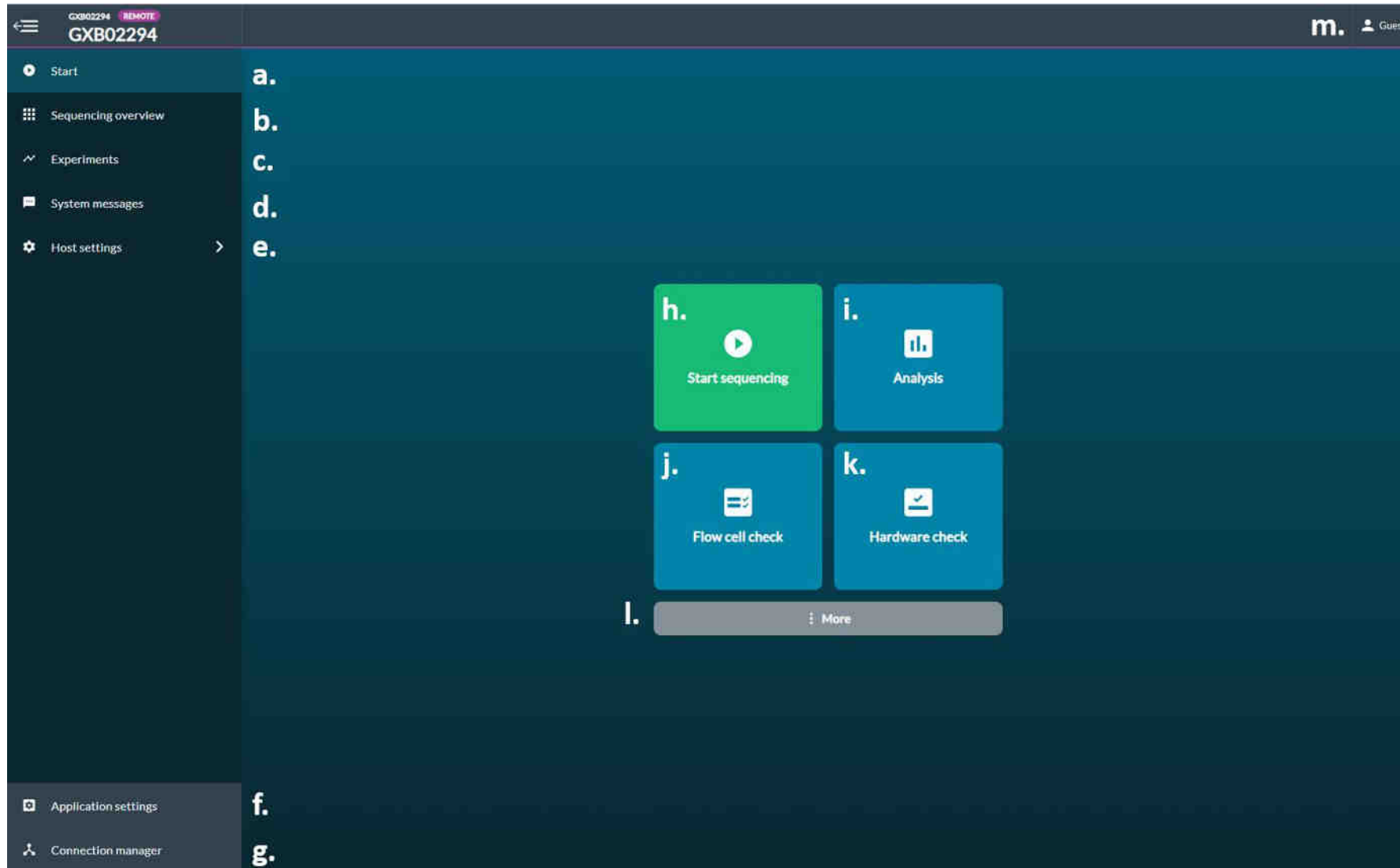
# Flow diagram of the SARS-CoV-2 Whole-genome sequencing using Oxford Nanopore



# Overview

- MinKNOW-Software
- Data formats
- Programs / work-flow used for data procession and analysis
  - Installation
  - Execution

# MinKNOW-Software

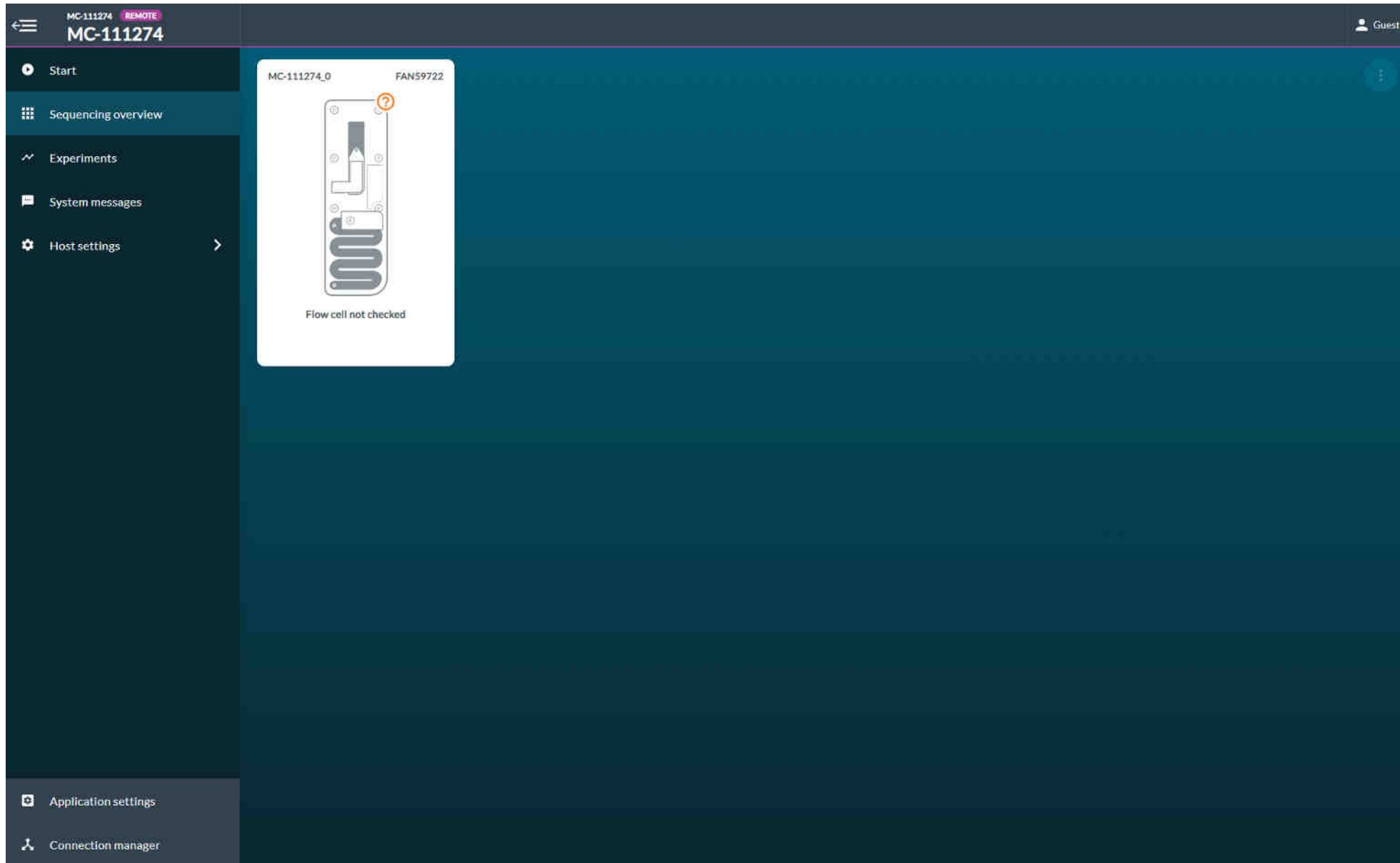


- a. **Start** homepage
- b. **Sequencing Overview** of connected flow cells
- c. Recent and current **Experiments**
- d. **System Messages**
- e. **Host Settings**
- f. **Application Settings** contain tutorials, account/login settings and MinKNOW GUI information
- g. **Connection Manager** to connect with other available devices
- h. **Start Sequencing** experiment
- i. Post-run **Analysis**
- j. **Flow Cell Check**
- k. **Hardware Check**
- l. **More** includes option to generate .mmi from .fasta file or to import a sample sheet
- m. **Guest/initials** to logout

[https://community.nanoporetech.com/protocols/experiment-companion-minknow/v/mke\\_1013\\_v1\\_revbp\\_11apr2016/homepage](https://community.nanoporetech.com/protocols/experiment-companion-minknow/v/mke_1013_v1_revbp_11apr2016/homepage)

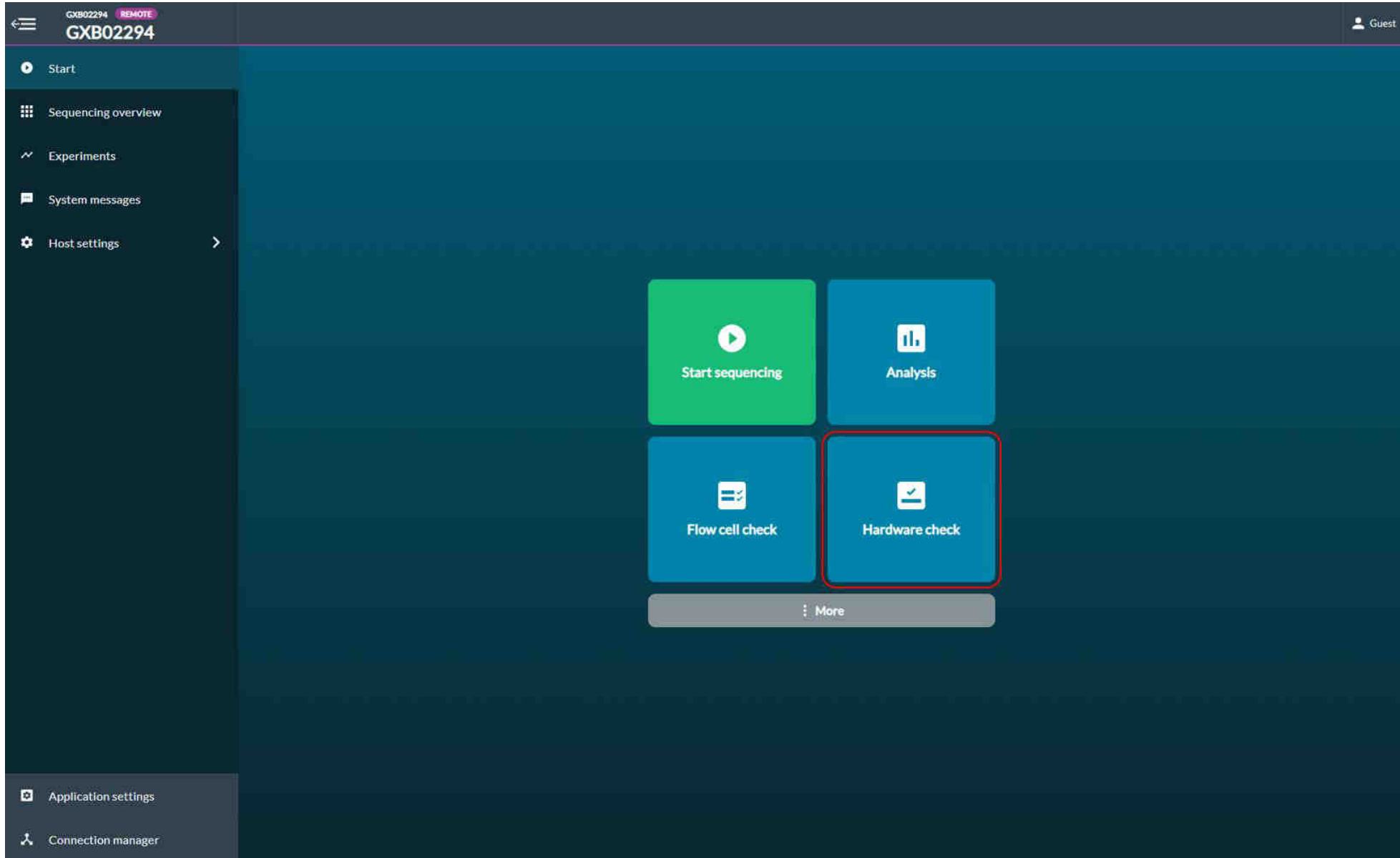


# MinKNOW-Software



[https://community.nanoporetech.com/protocols/experiment-companion-minknow/v/mke\\_1013\\_v1\\_revbp\\_11apr2016/homepage](https://community.nanoporetech.com/protocols/experiment-companion-minknow/v/mke_1013_v1_revbp_11apr2016/homepage)

# MinKNOW-Software



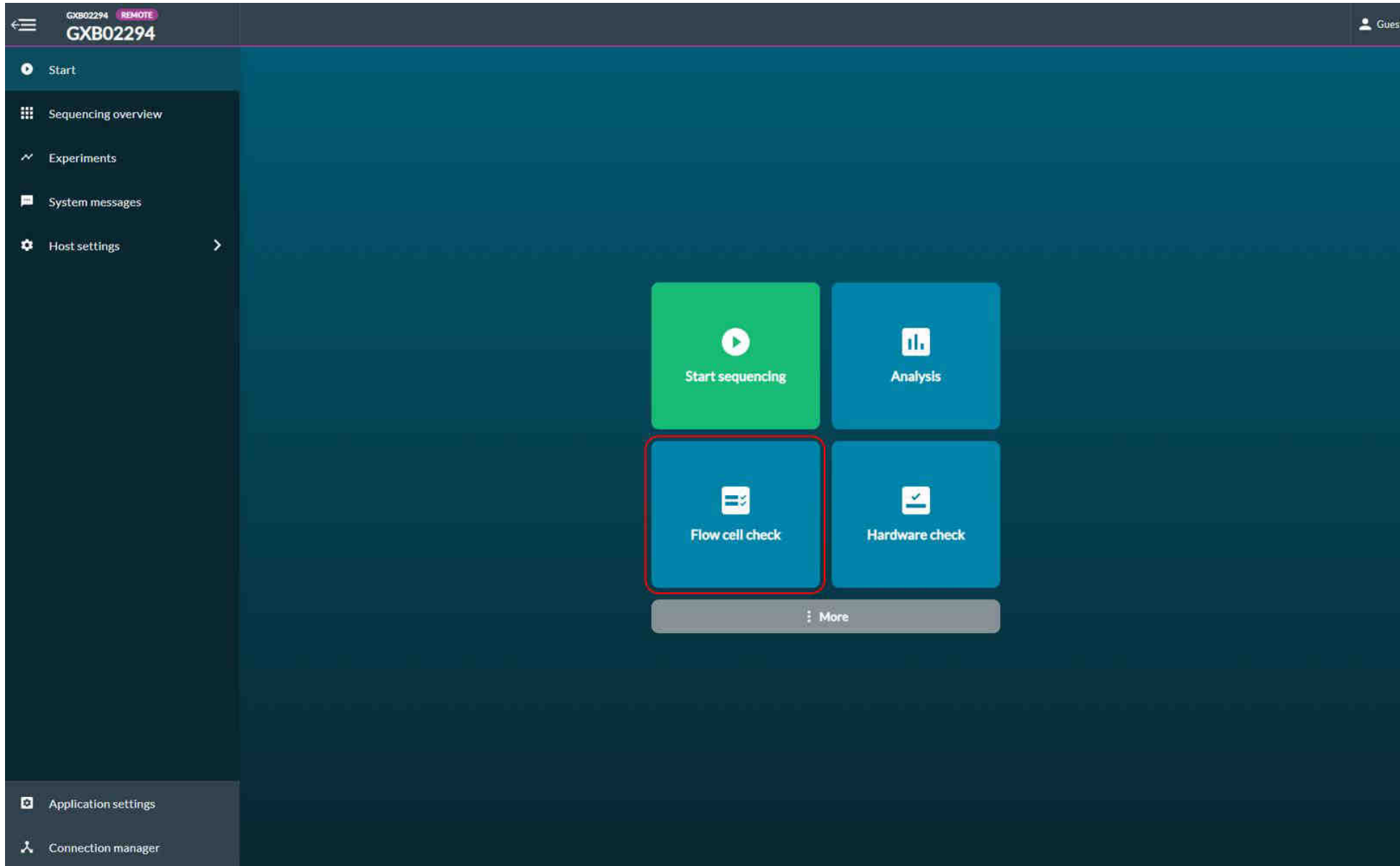
[https://community.nanoporetech.com/protocols/experiment-companion-minknow/v/mke\\_1013\\_v1\\_revbp\\_11apr2016/homepage](https://community.nanoporetech.com/protocols/experiment-companion-minknow/v/mke_1013_v1_revbp_11apr2016/homepage)

# MinKNOW-Software

The screenshot displays the MinKNOW software interface. At the top left, the device ID 'MC-111274' is shown with a 'REMOTE' indicator. The top right corner shows the user 'Guest'. A dark sidebar on the left contains navigation options: 'Start', 'Sequencing overview', 'Experiments', 'System messages', 'Host settings', 'Application settings', and 'Connection manager'. The main content area is titled 'Hardware check' and shows '1 of 1 positions selected' and a green checkmark indicating 'Ready to start'. A large button labeled 'CONFIGURATION TEST CELL' is present, with a note below it: 'Please ensure that the configuration test cell has been inserted correctly.' At the bottom, there are 'Back to start' and 'Start' buttons.

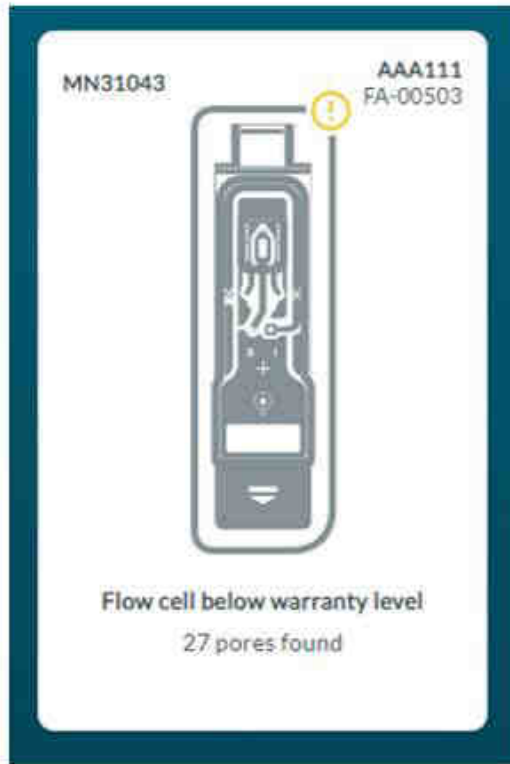
[https://community.nanoporetech.com/protocols/experiment-companion-minknow/v/mke\\_1013\\_v1\\_revbp\\_11apr2016/homepage](https://community.nanoporetech.com/protocols/experiment-companion-minknow/v/mke_1013_v1_revbp_11apr2016/homepage)

# MinKNOW-Software

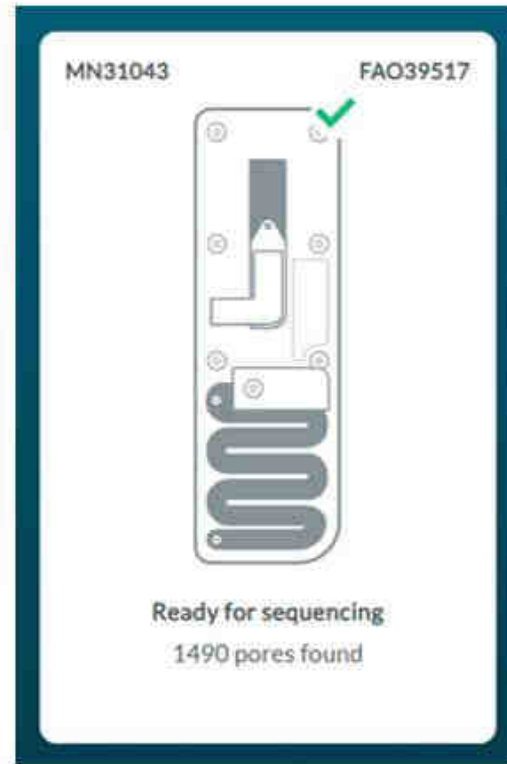


[https://community.nanoporetech.com/protocols/experiment-companion-minknow/v/mke\\_1013\\_v1\\_revbp\\_11apr2016/homepage](https://community.nanoporetech.com/protocols/experiment-companion-minknow/v/mke_1013_v1_revbp_11apr2016/homepage)

# MinKNOW-Software



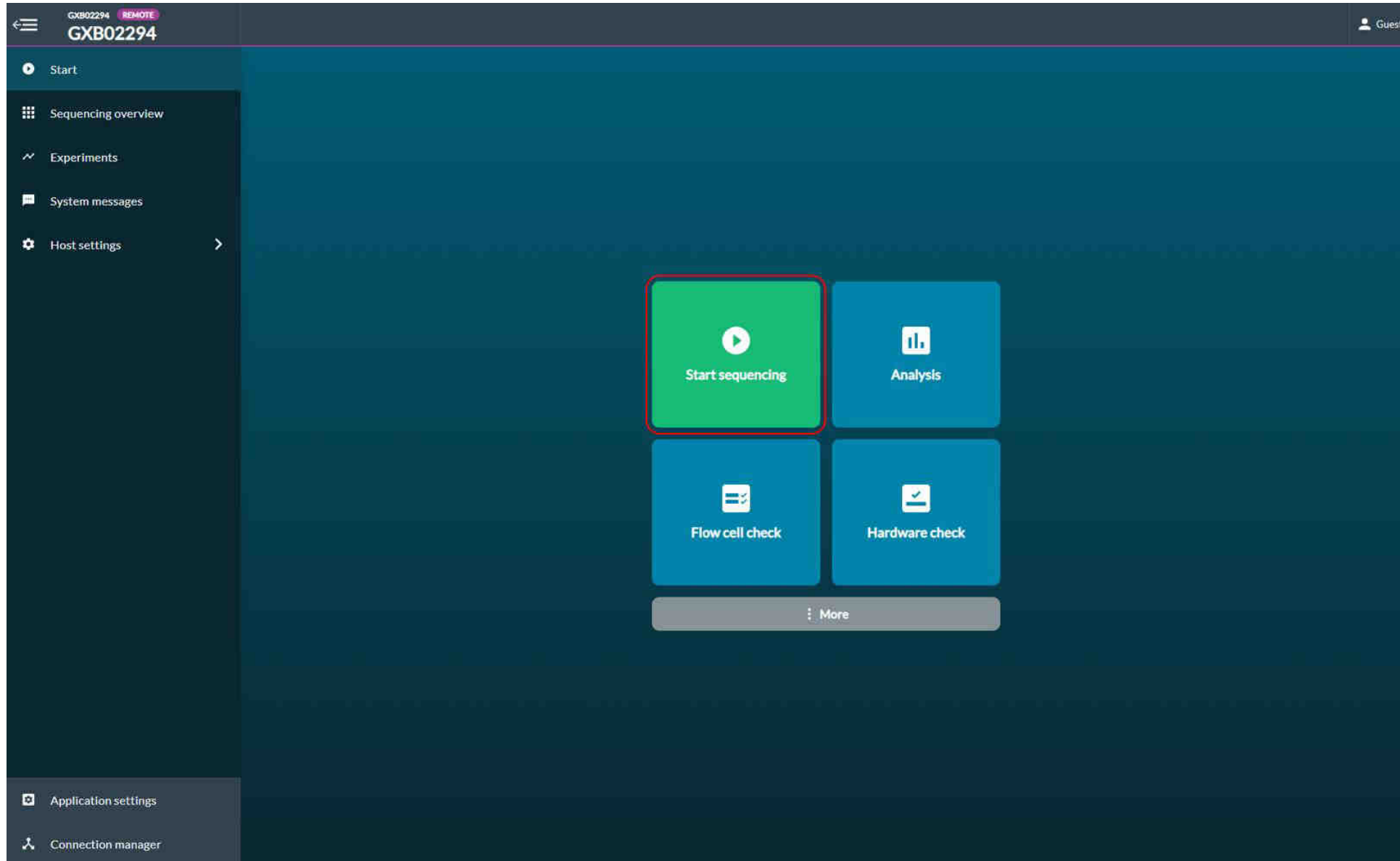
Yellow exclamation mark: The number of sequencing pores is below warranty.



Green tick: The number of sequencing pores is above warranty and ready for sequencing.

- warranty for flow cells -> 800 nanopores or above checked within 5 days of receipt
- Sequencing can also be run with less than 800 nanopores.
- The lower the number of pores, the longer will it need to sequence.

# MinKNOW-Software



[https://community.nanoporetech.com/protocols/experiment-companion-minknow/v/mke\\_1013\\_v1\\_revbp\\_11apr2016/starting-a-sequencing-run-on-minion](https://community.nanoporetech.com/protocols/experiment-companion-minknow/v/mke_1013_v1_revbp_11apr2016/starting-a-sequencing-run-on-minion)

# MinKNOW-Software

The screenshot displays the MinKNOW software interface for a remote session (MC-110168). The top navigation bar shows a sequence of steps: 1. Positions, 2. Kit, 3. Run options, 4. Basecalling, 5. Output, and 6. Review. The left sidebar contains menu items: Start, Sequencing overview, Experiments, System messages, Host settings, Application settings, and Connection manager. The main content area is titled 'Select positions' and features a search bar with the text 'Test'. Below the search bar are two buttons: 'Join existing' and 'Load settings from template'. A table is displayed with the following columns: Position, Flow cell ID, Flow cell type, and Sample ID. The table contains one row with the following values: Position: MC-110168\_0, Flow cell ID: FAK23143, Flow cell type: FLO-MIN106, and Sample ID: Sample\_ID. At the bottom of the interface, there are three buttons: 'Back to start', 'Continue to kit selection', and 'Skip to final review'.

Position	Flow cell ID	Flow cell type	Sample ID
MC-110168_0	FAK23143	FLO-MIN106	Sample_ID

[https://community.nanoporetech.com/protocols/experiment-companion-minknow/v/mke\\_1013\\_v1\\_revbp\\_11apr2016/starting-a-sequencing-run-on-minion](https://community.nanoporetech.com/protocols/experiment-companion-minknow/v/mke_1013_v1_revbp_11apr2016/starting-a-sequencing-run-on-minion)

# MinKNOW-Software

The screenshot shows the 'Kit selection' step in the MinKNOW software. The interface includes a top navigation bar with steps 1-6, a left sidebar with menu items, and a main content area with filter buttons and a grid of sequencing kits. The 'Rapid Barcoding Kit' (SQK-RBK004) is selected.

**Kit selection**

Sample type:   PCR-free:   Multiplexing:    Control [Reset filters](#)

Ligation Sequencing Kit SQK-LSK109	Rapid Barcoding Kit SQK-RBK004	Rapid Sequencing Kit SQK-RAD004	Direct RNA Sequencing Kit SQK-RNA002
PCR Barcoding Kit SQK-PBK004	16S Barcoding Kit (BC1-24) SQK-16S024	CAS109 Sequencing Protocol SQK-CAS109	CAS109 Sequencing Kit SQK-CS9109
Direct cDNA Sequencing Kit SQK-DCS109	Field Sequencing Kit SQK-LRF001	Ligation Sequencing Kit (48 reactions) SQK-LSK109-XL	Ligation Sequencing Kit SQK-LSK110
Ligation Sequencing Kit XL SQK-LSK110-XL	Native Barcoding Sequencing Kit 24 SQK-NBD110-24	Native Barcoding Sequencing Kit 96 SQK-NBD110-96	PCR cDNA Barcoding Kits SQK-PCB109
PCR cDNA Barcoding Kit SQK-PCB110	PCR-cDNA Sequencing Kit SQK-PCS109	PCR cDNA Sequencing Kit SQK-PCS110	Pore-C Sequencing Protocol SQK-PRC109
PCR Sequencing Kit SQK-PSK004	SQK-RBK110-96 SQK-RBK110-96	Direct RNA Sequencing Kit SQK-RNA003	Rapid PCR Barcoding Kit SQK-RPB004
SQK-ULK001 SQK-ULR001	VoITRAX Multiplex Sequencing Kit VSK-VMK002	VSK-VMK003 VSK-VMK003	VoITRAX Sequencing Kit VSK-VSK002
VSK-VSK003 VSK-VSK003	Direct cDNA Sequencing Kit SQK-DCS108	Ligation Sequencing Kit SQK-LSK108	PCR-cDNA Sequencing Kit SQK-PCS108

**Select barcode expansion pack**

Native Barcoding Expansion 1-12 (PCR-free) EXP-NBD194	Native Barcoding Expansion 13-24 (PCR-free) EXP-NBD114	EXP-NBD196 EXP-NBD196	PCR Barcoding Expansion 1-12 EXP-PBC001
PCR Barcoding Expansion 1-96 EXP-PBC096			

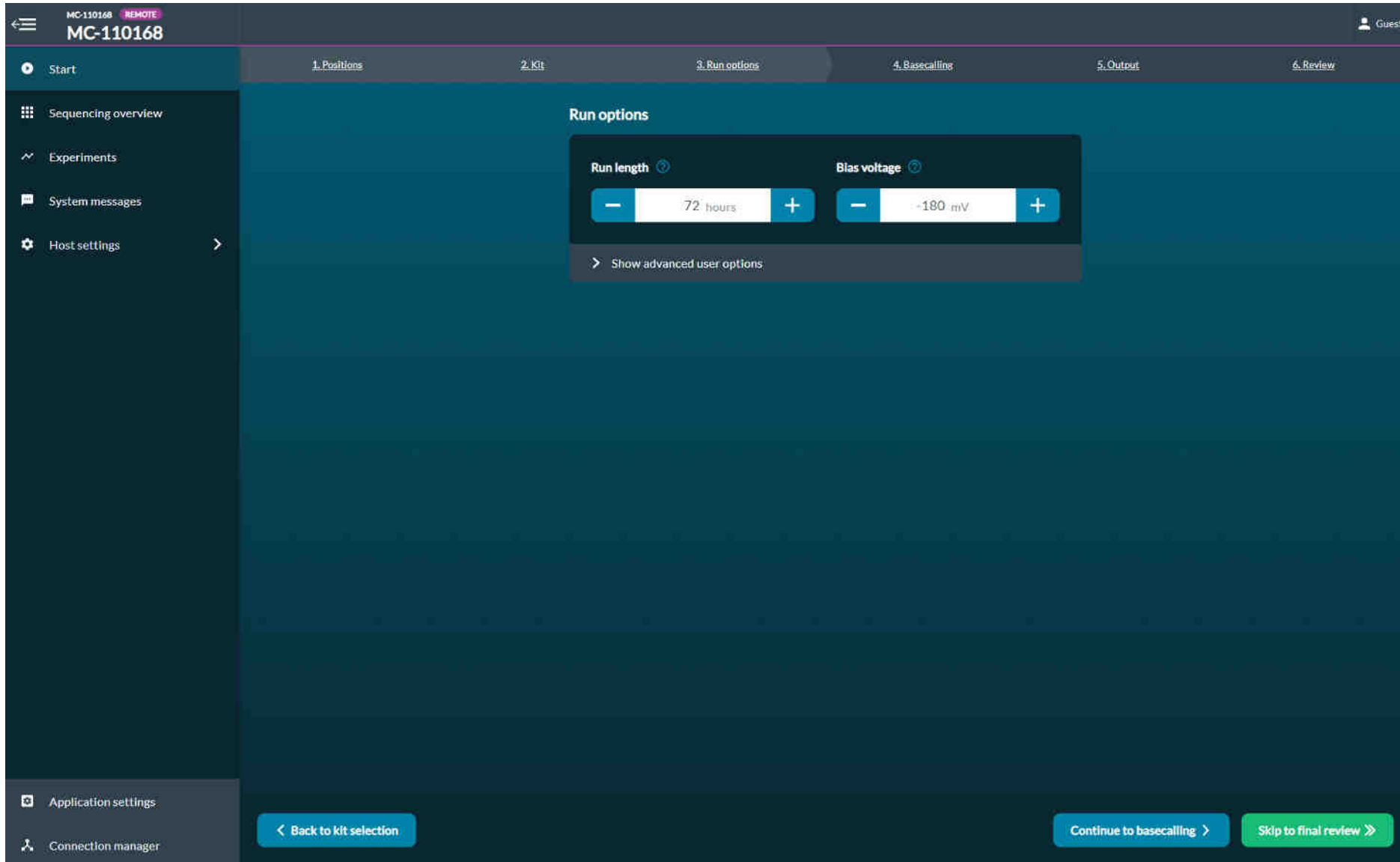
Navigation buttons: [Back to position selection](#) [Continue to run options](#) [Skip to final review](#)

Select the kit!

For the training we use:  
„Rapid Barcoding-Kit“  
SQK-RB004



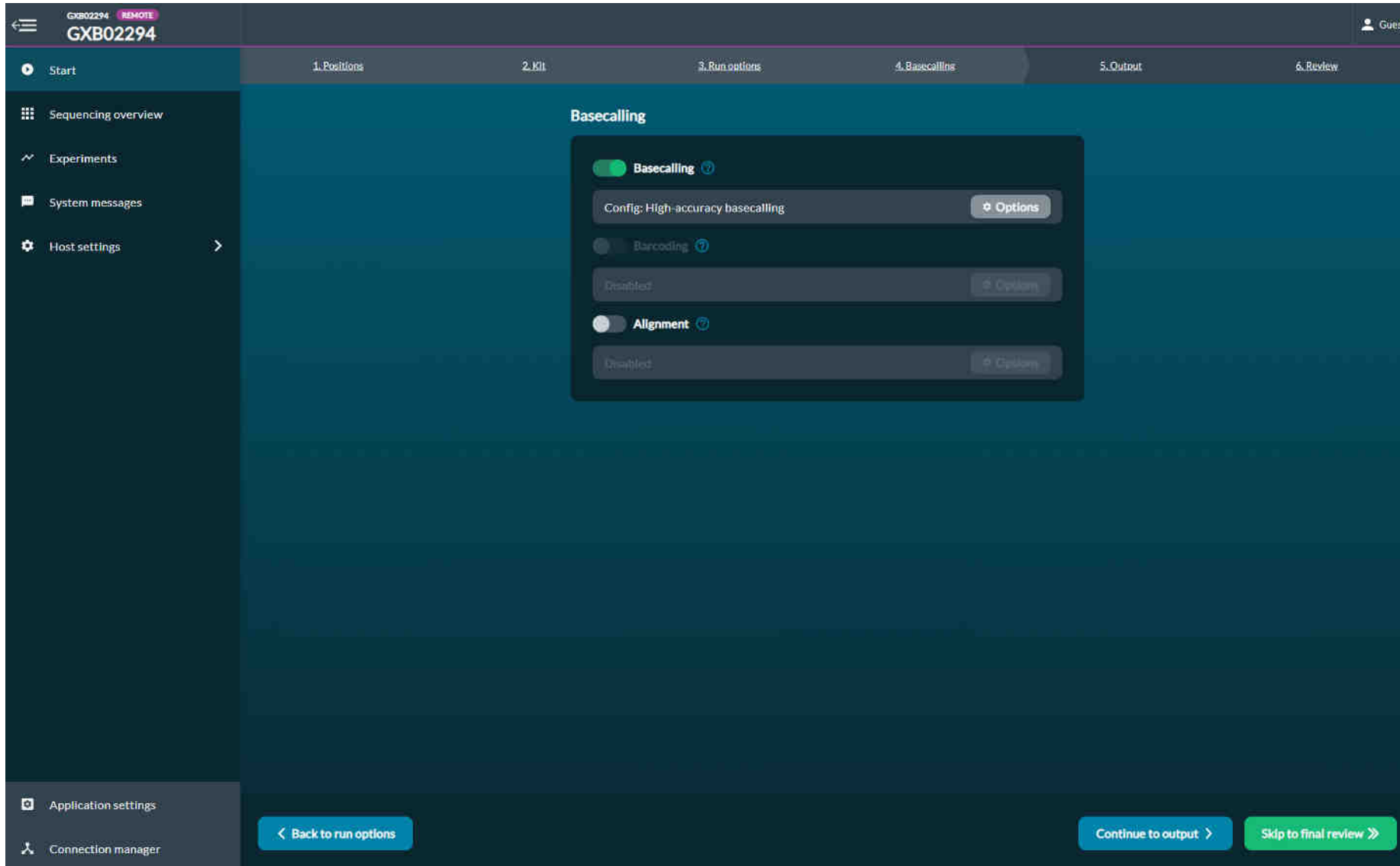
# MinKNOW-Software



Set run length to  
6 hours

(it can be stopped  
earlier)

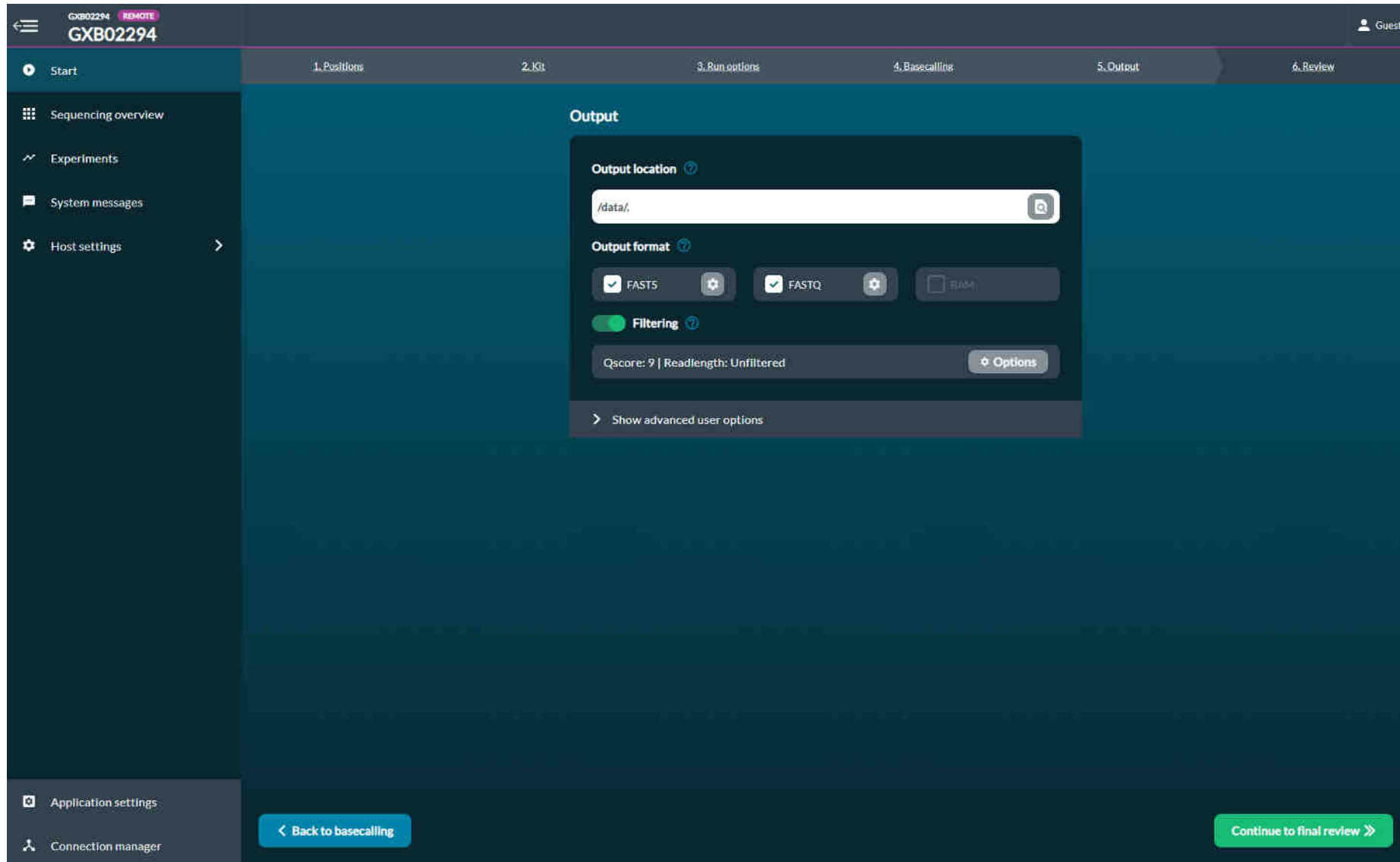
# MinKNOW-Software



Select basecalling  
and barcoding

[https://community.nanoporetech.com/protocols/experiment-companion-minknow/v/mke\\_1013\\_v1\\_revbp\\_11apr2016/starting-a-sequencing-run-on-minion](https://community.nanoporetech.com/protocols/experiment-companion-minknow/v/mke_1013_v1_revbp_11apr2016/starting-a-sequencing-run-on-minion)

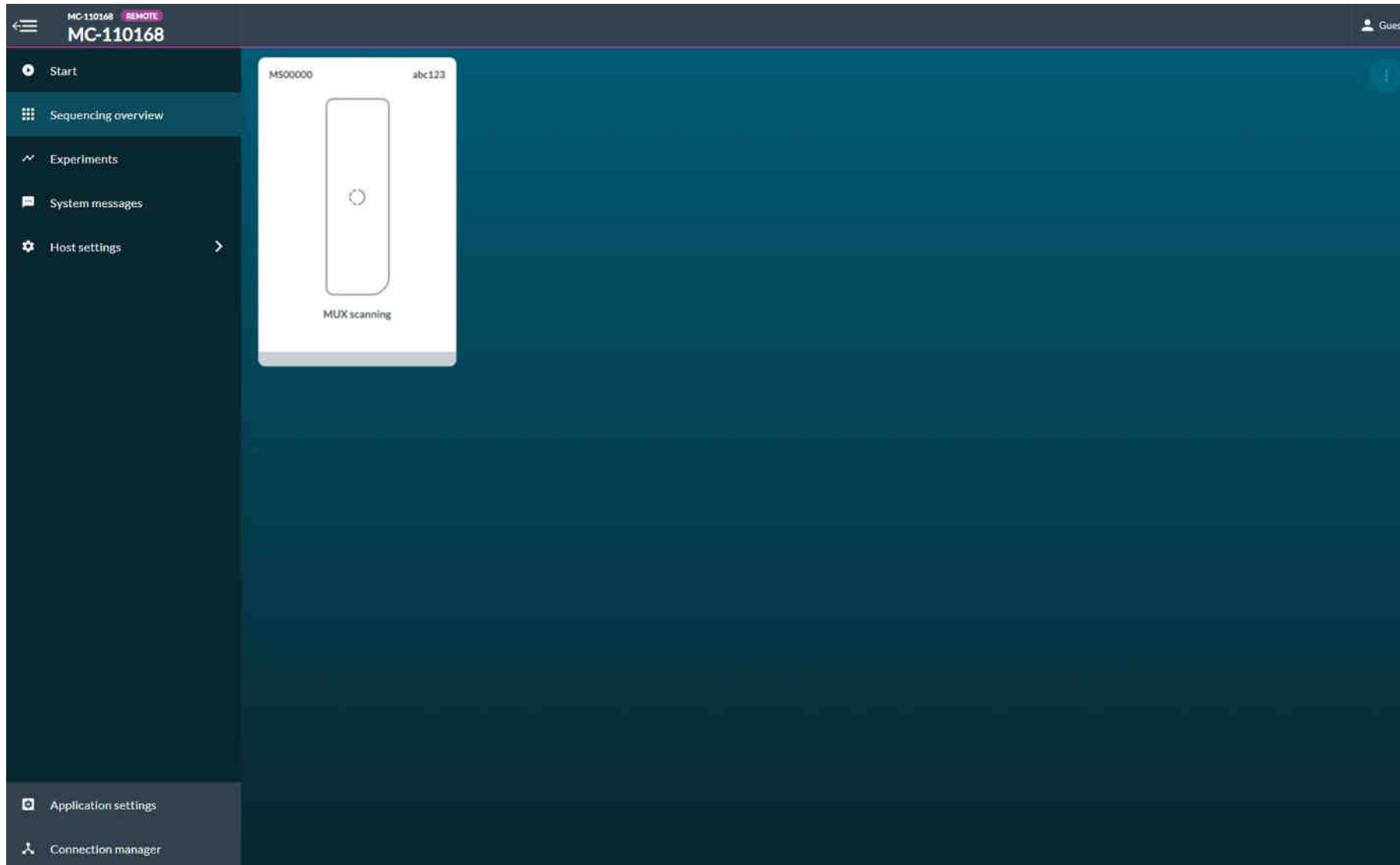
# MinKNOW-Software



Switch off  
fastq-compression

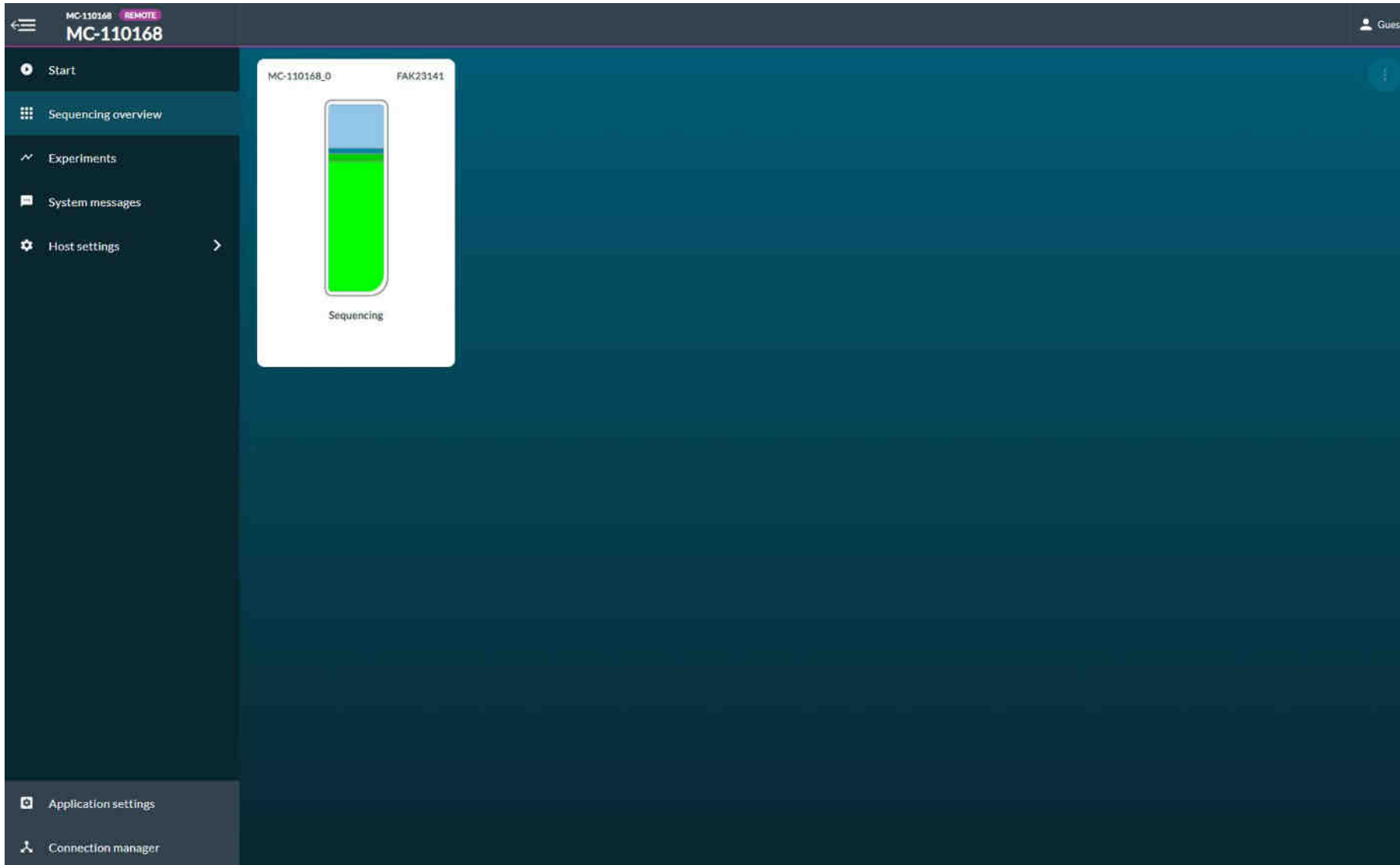
[https://community.nanoporetech.com/protocols/experiment-companion-minknow/v/mke\\_1013\\_v1\\_revbp\\_11apr2016/starting-a-sequencing-run-on-minion](https://community.nanoporetech.com/protocols/experiment-companion-minknow/v/mke_1013_v1_revbp_11apr2016/starting-a-sequencing-run-on-minion)

# MinKNOW-Software



MUX-Scan:  
multiplex scan,  
allows MinKNOW  
to prioritise the  
order in which the  
nanopores are  
used, maximising  
the data output in  
the initial stages of  
the run.

# MinKNOW-Software



[https://community.nanoporetech.com/protocols/experiment-companion-minknow/v/mke\\_1013\\_v1\\_revbp\\_11apr2016/homepage](https://community.nanoporetech.com/protocols/experiment-companion-minknow/v/mke_1013_v1_revbp_11apr2016/homepage)

# MinKNOW-Software

The screenshot displays the MinKNOW software interface for a remote host (MC-110168). The interface includes a sidebar with navigation options: Start, Sequencing overview, Experiments, System messages, Host settings, Application settings, and Connection manager. The main area shows a summary for the experiment '4\_3\_basecalling\_test' with the following statistics:

- Reads: 236.22 k
- Estimated bases: 1.07 Gb
- Basecalled bases: 1.11 Gb
- Active runs: 2
- Total runs: 4

Below the summary are control buttons: Resume, Pause, Stop, Start MUX scan, Export PDF, and Experiment group. A table below shows the current run details:

Position	Flow cell ID	Sample ID	Health	Run time	Run state	Reads	Bases	Basecalled %
MC-110168_0	FAK23141	exp_14_04_2021_runtime		16 m / 72 h	Active	19.65 k	86.77 Mb basecalled 84.64 Mb estimated	100%

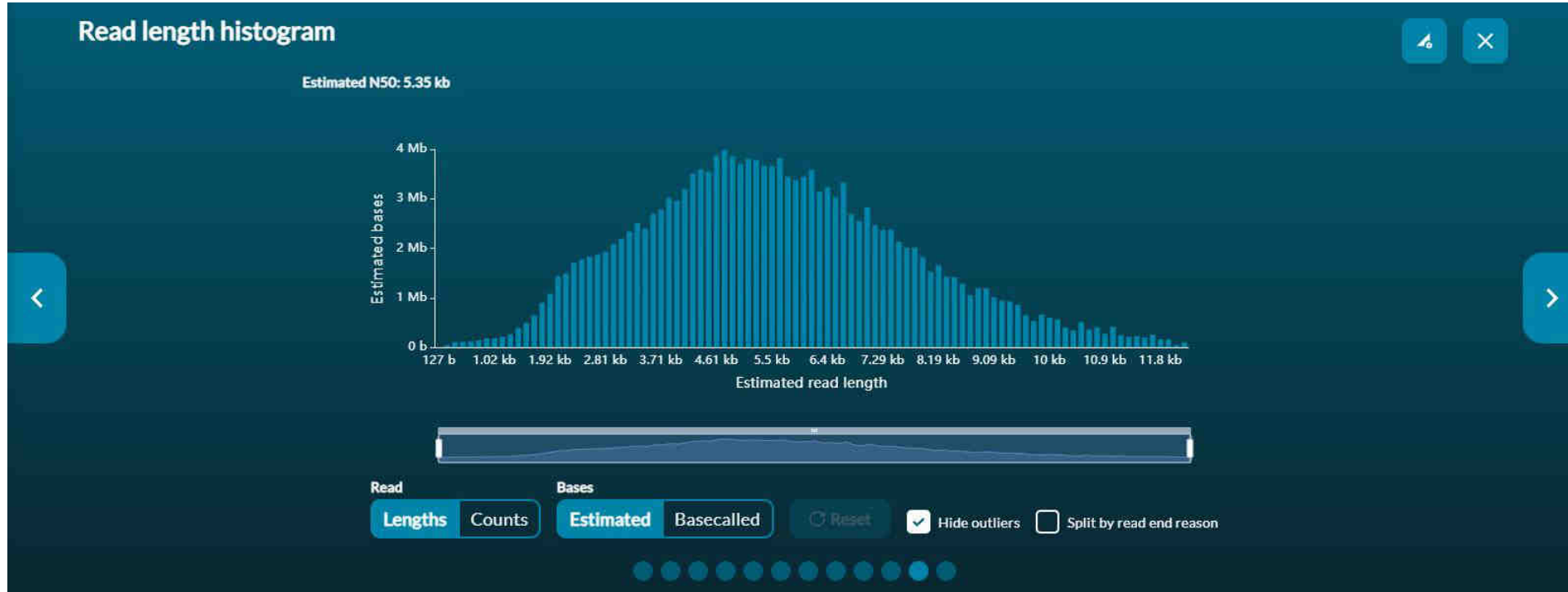
At the bottom right, there is a 'Scroll right >' button.

[https://community.nanoporetech.com/protocols/experiment-companion-minknow/v/mke\\_1013\\_v1\\_revbp\\_11apr2016/homepage](https://community.nanoporetech.com/protocols/experiment-companion-minknow/v/mke_1013_v1_revbp_11apr2016/homepage)

# MinKNOW-Software



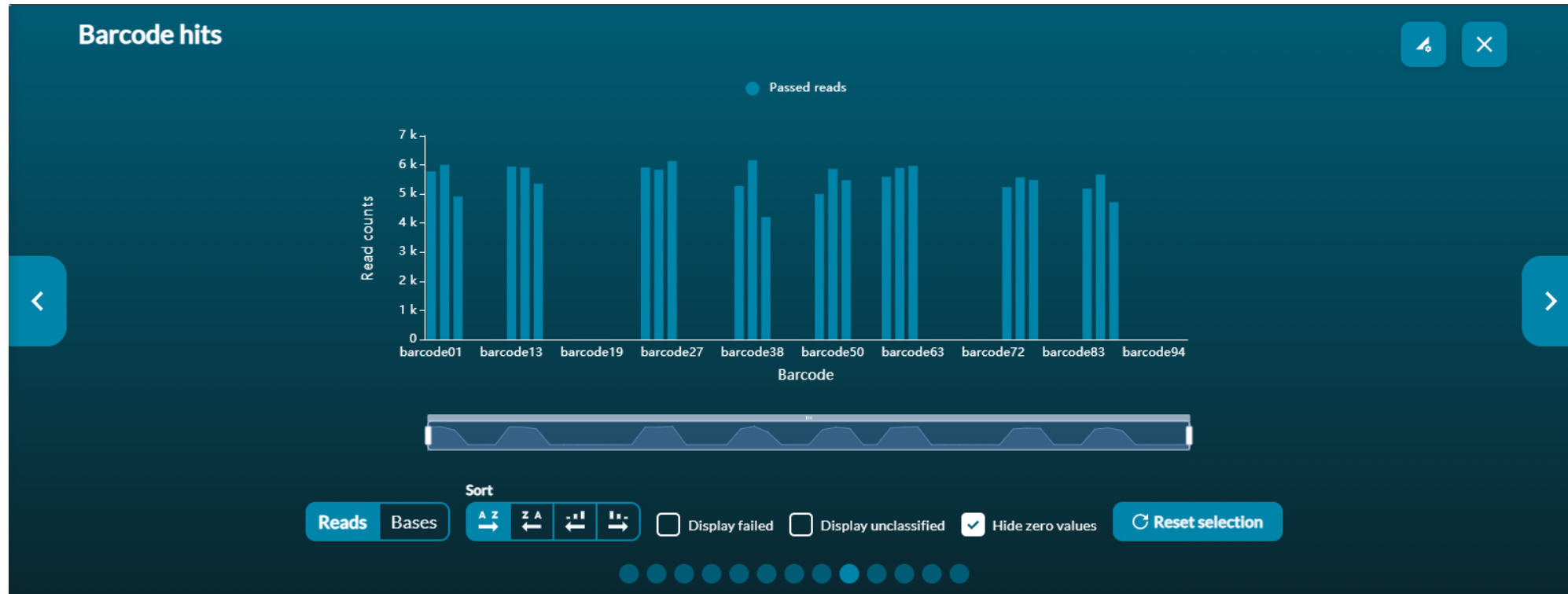
# MinKNOW-Software



[https://community.nanoporetech.com/protocols/experiment-companion-minknow/v/mke\\_1013\\_v1\\_revbp\\_11apr2016/progression-of-minknow-pro](https://community.nanoporetech.com/protocols/experiment-companion-minknow/v/mke_1013_v1_revbp_11apr2016/progression-of-minknow-pro)



# MinKNOW-Software



[https://community.nanoporetech.com/protocols/experiment-companion-minknow/v/mke\\_1013\\_v1\\_revbp\\_11apr2016/progression-of-minknow-pro](https://community.nanoporetech.com/protocols/experiment-companion-minknow/v/mke_1013_v1_revbp_11apr2016/progression-of-minknow-pro)

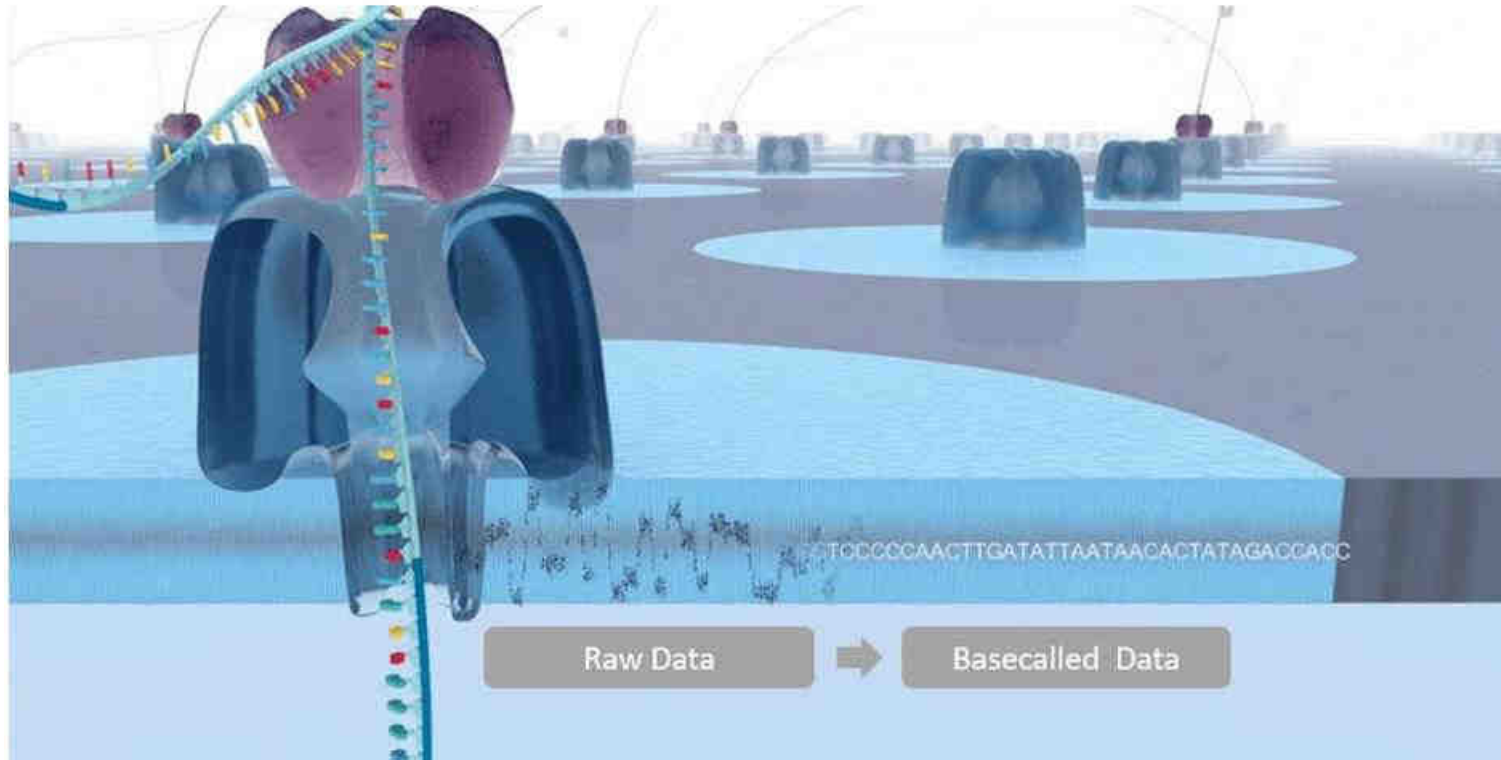
## Data Formats – fast5

- raw-data
- contain all information that was generated while sequencing
- big size
- not human-readable

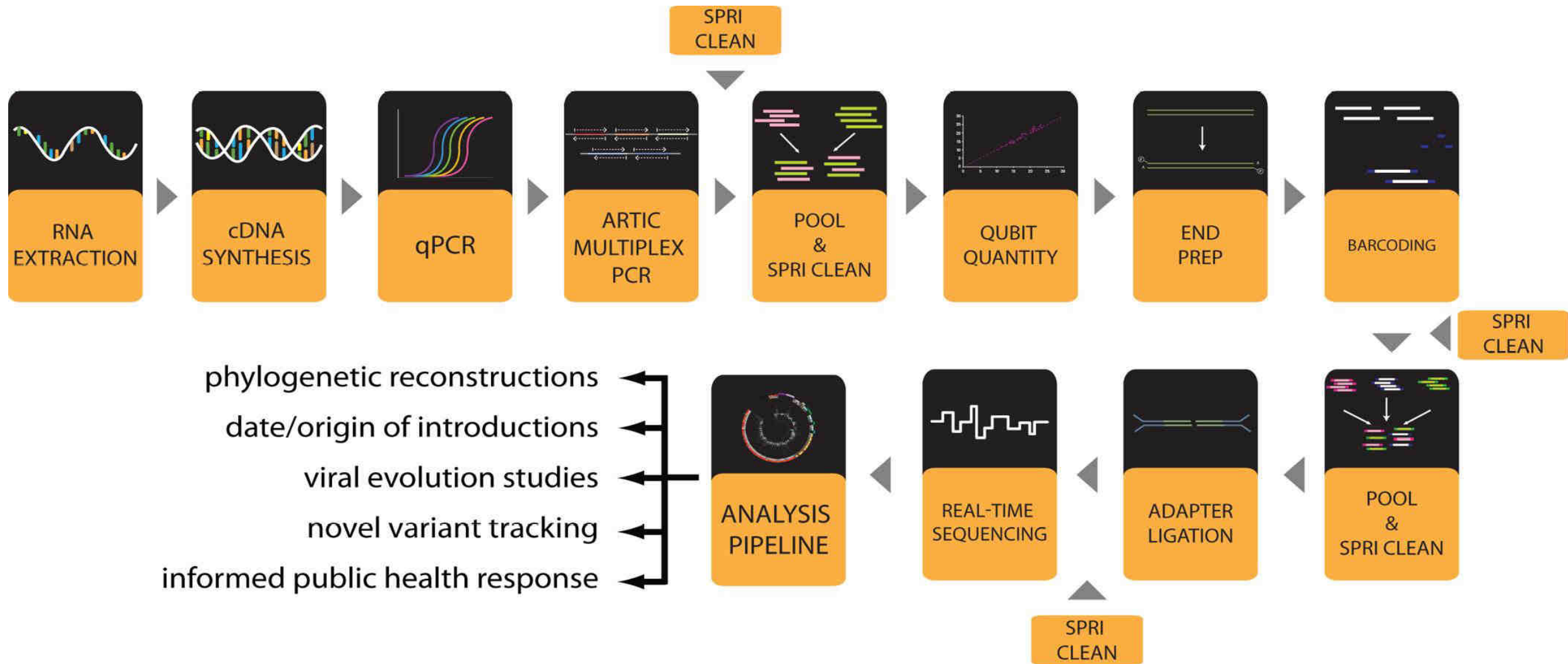
# High accuracy basecalling with Guppy

- Toolkit by Oxford Nanopore, integrated in MinKNOW
- Basecalling algorithm based on a Recurrent Neural Network (RNN)

Basecalling is the process of converting the electrical signals generated by a DNA or RNA strand passing through the nanopore into the corresponding base sequence of the strand. The general data flow in a nanopore sequencing experiment is shown below.



[https://community.nanoporetech.com/protocols/Guppy-protocol/v/gpb\\_2003\\_v1\\_revz\\_14\\_dec2018/guppy-software-overview](https://community.nanoporetech.com/protocols/Guppy-protocol/v/gpb_2003_v1_revz_14_dec2018/guppy-software-overview)



# Overview

- Introduction
  - Who is who? Who knows what?
- How will the training be organized?
  - Detailed schedule
  - What is in the “additional information”?
- Overview of NGS-nanopore work-flow as described in the protocol
- Theory of nanopore-sequencing
  - Material + devices

# What is in the additional information ?

## 1. Theory

- 1.1 SARS-CoV-2
  - > Presentation given on 15.09.2021 (pdf)
- 1.2 COVID-19 Genomic Epidemiology Toolkit by CDC
  - > <https://www.cdc.gov/amd/training/covid-19-gen-epi-toolkit.html>
- 1.3 Details about Nanopore Technology -> workshop-presentations (pdf)
  - > <https://nanoporetech.com/how-nanopore-sequencing-works>
  - > <https://nanoporetech.com/nanopore-sequencing-data-analysis>
- 1.4 Links SARS-CoV-2 sequencing principles videos Oxford Nanopore:
  - > <https://nanoporetech.com/resource-centre/bioinformatics-workflows-sars-cov-2-raw-nanopore-reads-consensus-genomes-using>
  - > <https://www.youtube.com/watch?v=rYaFcDE-Ewg>

# What is in the additional information ?

## 1. Theory

- 1.5 Links to get familiar with technology and experiments that you may plan in future using NANOPORE
  - > <https://nanoporetech.com/>
  - > <https://nanoporetech.com/nanopore-sequencing-data-analysis>
  - > <https://www.youtube.com/c/OxfordNanoporeTechnologies/playlists>

# What is in the additional information ?

## 2. Wet-lab

- 2.1 Protocol! (pdf)
- 2.2 Video links on Oxford Nanopore Flow cell introduction, Priming and loading flow cell
  - > [https://www.youtube.com/watch?v=zC6lAtzqi\\_k](https://www.youtube.com/watch?v=zC6lAtzqi_k)
  - > <https://www.youtube.com/watch?v=Pt-iaemrM88>
  - > <https://www.youtube.com/watch?v=IknVaEnuDz0&t=604s>
- 2.3. Flow cell wash protocol (pdf)
- 2.4 Methods manuscript, literatur protocol (pdf)



# What is in the additional information ?

## 3. Dry-lab

### 3.1 Protocol! (pdf)

Protocol NGS\_WGS\_SARS-CoV-2\_1200bp amplicons using Nanopore\_V1.2.pdf

### 3.2 Introduction to bioinformatics

-> <https://www.youtube.com/c/SimonCockell/videos>

### 3.3 Linux

-> linuxcommands (pdf)

-> Links on Linux operating system

[https://www.youtube.com/watch?v=v\\_1zB2WNN14](https://www.youtube.com/watch?v=v_1zB2WNN14)

<https://www.youtube.com/watch?v=V1y-mbWM3B8>Ubuntu

<https://www.youtube.com/watch?v=x5MhydiWmc>

<https://www.youtube.com/watch?v=G36I1iqDZig>



Alexander von Humboldt  
Stiftung/Foundation



EBERHARD KARLS  
UNIVERSITÄT  
TÜBINGEN



---

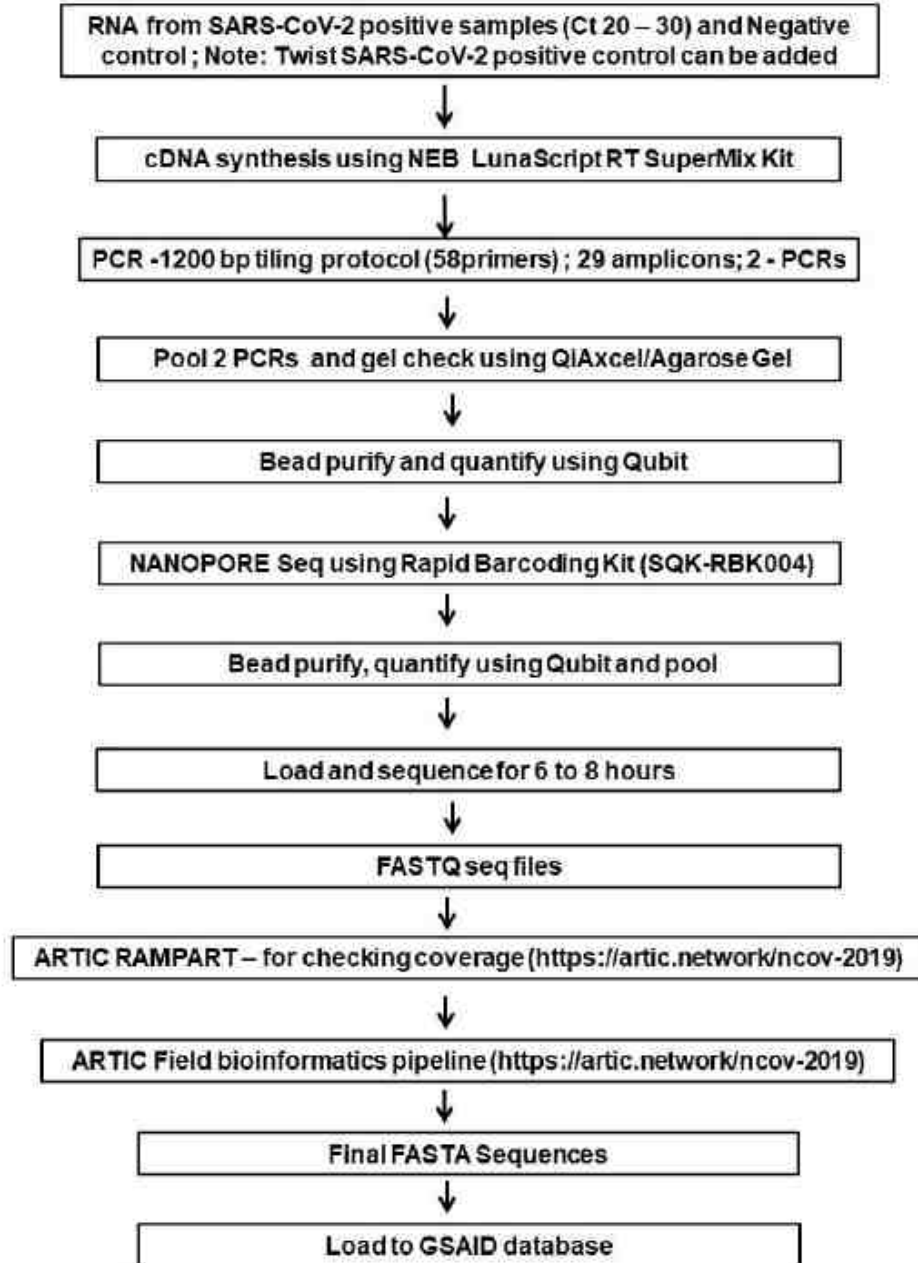
# Workshop

## Molecular Diagnostic Techniques (Genomics and Bioinformatics – Infectious diseases)

### Module-4

Dr. Srinivas Reddy Pallerla, Mr. Tung Tran and Ms. Linh Le  
Prof. Dr. Thirumalaisamy P. Velavan, University Hospital, Tübingen, Germany  
Prof. Dr. Daniel Olusola Ojurongbe, Nigeria  
Prof. Dr. Mohamed Osman, University of Khartoum, Sudan

# NGS-nanopore WORK-FLOW Overview



# What is in the additional information ?

## 3. Dry-lab

### 3.1 Protocol! (pdf)

Protocol NGS\_WGS\_SARS-CoV-2\_1200bp amplicons using Nanopore\_V1.2.pdf

### 3.2 Introduction to bioinformatics

-> <https://www.youtube.com/c/SimonCockell/videos>

### 3.3 Linux

-> linuxcommands (pdf)

-> Links on Linux operating system

[https://www.youtube.com/watch?v=v\\_1zB2WNN14](https://www.youtube.com/watch?v=v_1zB2WNN14)

<https://www.youtube.com/watch?v=V1y-mbWM3B8>Ubuntu

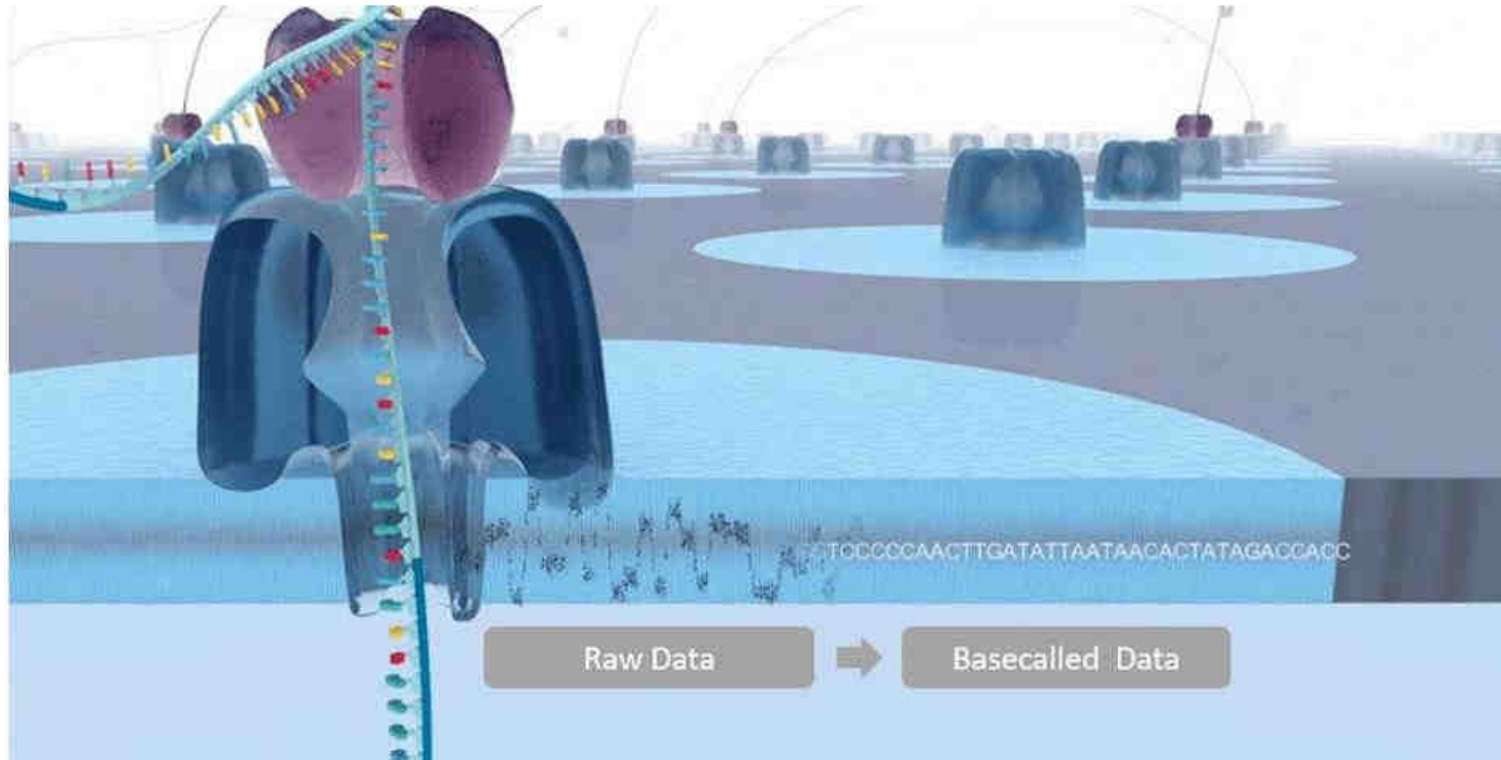
<https://www.youtube.com/watch?v=x5MhydiWmc>

<https://www.youtube.com/watch?v=G36I1iqDZig>

# High accuracy basecalling with Guppy

- Toolkit by Oxford Nanopore, integrated in MinKNOW
- Basecalling algorithm based on a Recurrent Neural Network (RNN)

Basecalling is the process of converting the electrical signals generated by a DNA or RNA strand passing through the nanopore into the corresponding base sequence of the strand. The general data flow in a nanopore sequencing experiment is shown below.



[https://community.nanoporetech.com/protocols/Guppy-protocol/v/gpb\\_2003\\_v1\\_revz\\_14\\_dec2018/guppy-software-overview](https://community.nanoporetech.com/protocols/Guppy-protocol/v/gpb_2003_v1_revz_14_dec2018/guppy-software-overview)

# Programs / work-flow – fieldbioinformatics

For further analysis, you have to merge all fasta-files into one, this can either be done manually or with the script below.

(the script is also in the dropbox, you have to edit in input-path)

```
GNU nano 4.8 merge_consensus_files.sh
#!/bin/bash

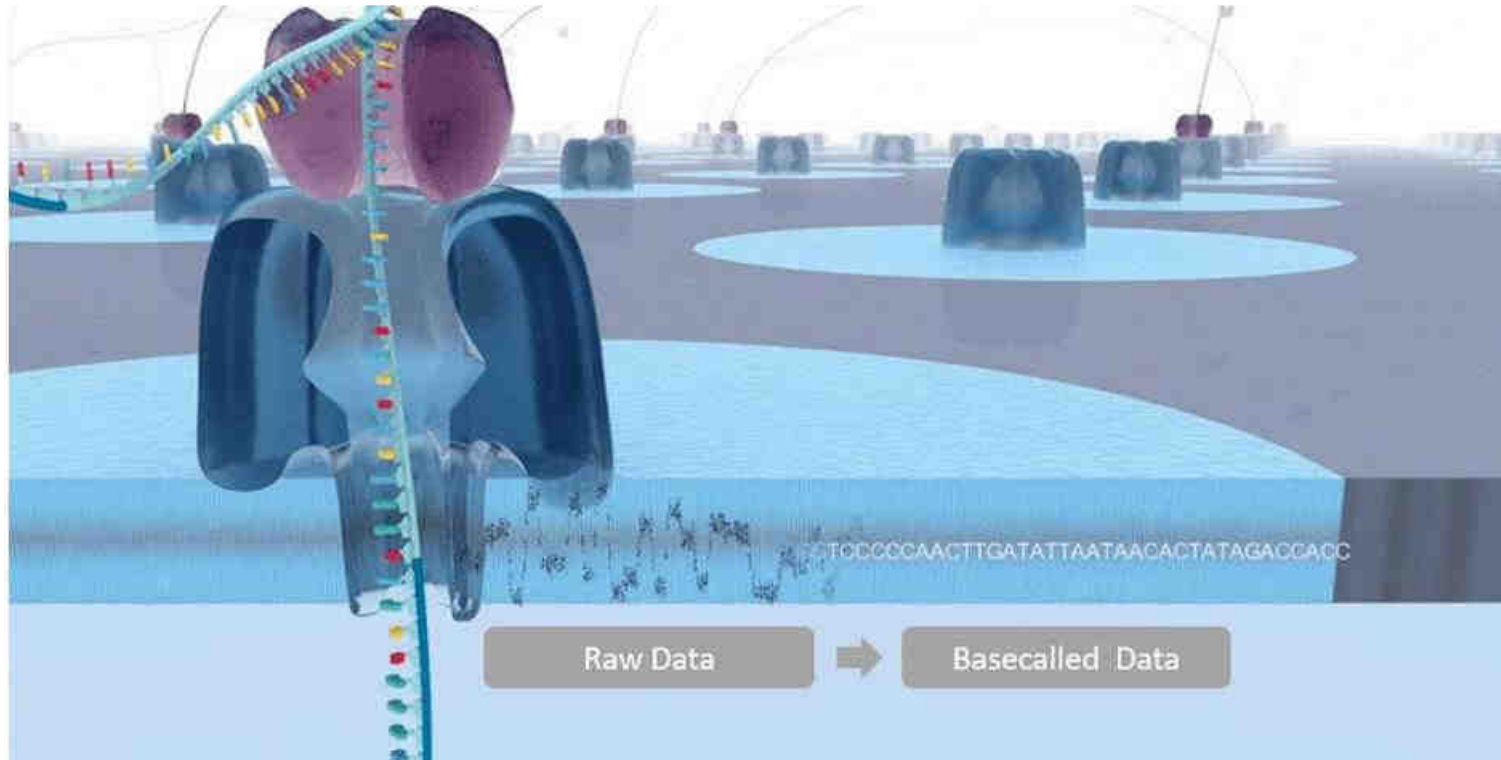
# create file
touch all_barcodes_merged.fasta
echo "file created"

# add consensus files
for N in {07..11};
do
    cat ~/data/data_210820/fastq_pass_high_accuracy/barcode$N/SARS-CoV_high_accuracy_Barcode$N/SARS-CoV_high_accuracy_Barcode$N.consensus.fasta \
    >> all_barcodes_merged.fasta
    echo "done adding Barcode$N"
done
```

# High accuracy basecalling with Guppy

- Toolkit by Oxford Nanopore, integrated in MinKNOW
- Basecalling algorithm based on a Recurrent Neural Network (RNN)

Basecalling is the process of converting the electrical signals generated by a DNA or RNA strand passing through the nanopore into the corresponding base sequence of the strand. The general data flow in a nanopore sequencing experiment is shown below.



[https://community.nanoporetech.com/protocols/Guppy-protocol/v/gpb\\_2003\\_v1\\_revz\\_14\\_dec2018/guppy-software-overview](https://community.nanoporetech.com/protocols/Guppy-protocol/v/gpb_2003_v1_revz_14_dec2018/guppy-software-overview)

# High accuracy basecalling with Guppy

## Toolkit:

- Basecaller
  - Calibration strand detection
  - Adapter trimming
- Barcoding/demultiplexing
- Alignment
- Modified basecalling

[https://community.nanoporetech.com/protocols/Guppy-protocol/v/gpb\\_2003\\_v1\\_revz\\_14dec2018/guppy-software-overview](https://community.nanoporetech.com/protocols/Guppy-protocol/v/gpb_2003_v1_revz_14dec2018/guppy-software-overview)



# High accuracy basecalling with Guppy

## System-requirements:

- 4 GB RAM plus 1 GB per thread for 1D basecalling
- 4 GB RAM plus 2 GB per thread for 1D2 basecalling
- Administrator access for .deb or .msi installers
- 2 GB of drive space for installation, minimum 512 GB storage space for basecalled read files (1 TB recommended)
  
- CUDA 10 for Linux 4 Tegra running Ubuntu 18

More details can be checked at the Oxford Nanopore website, GPU is recommended

[https://community.nanoporetech.com/protocols/Guppy-protocol/v/gpb\\_2003\\_v1\\_revz\\_14dec2018/guppy-software-overview](https://community.nanoporetech.com/protocols/Guppy-protocol/v/gpb_2003_v1_revz_14dec2018/guppy-software-overview)

# High accuracy basecalling with Guppy

## Different models:

“The Fast model is designed to keep up with data generation on Oxford Nanopore devices (MinION Mk1C, GridION, PromethION). The HAC model provides a higher raw read accuracy than the Fast model and is currently 5-8 times more computationally-intensive. The Super accurate model has an even higher raw read accuracy, and is ~3 times more intensive than the HAC model.”

Model	R9.4.1 modal accuracy	R10.3 modal accuracy
Fast	95.8	91.4
HAC	97.8	95.7
sup	98.3	97.5

[https://community.nanoporetech.com/protocols/Guppy-protocol/v/gpb\\_2003\\_v1\\_revz\\_14dec2018/guppy-software-overview](https://community.nanoporetech.com/protocols/Guppy-protocol/v/gpb_2003_v1_revz_14dec2018/guppy-software-overview)

# High accuracy basecalling with Guppy

Different models:

Device	Basecalling speed in Gbases per hour			Keep-up, number of flow cells
	Fast model	HAC model	sup model	Fast model
PromethION P24	82	19	5.2	23
GridION	34	5	2	51
MinION Mk1C	0.8	0.07	-	1.25
MinION CPU, high-spec laptop	0.13	0.014	0.003	0.21
MinION GPU, RTX2070 laptop	14	1.6	0.06	22

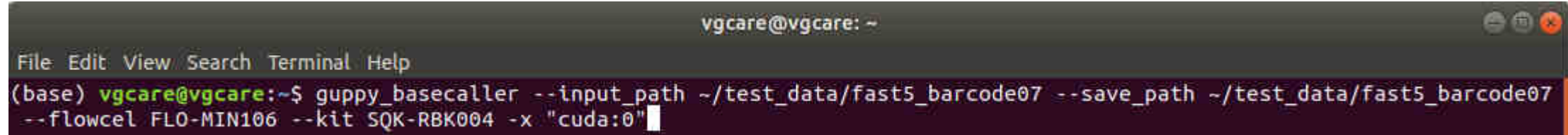
“Keep up” is defined as 80% of the theoretical flow cell output.

e.g. MinION = 4000 kHz x 512 channels x 0.8 = 1.6 M samples/s = 160 kbases/s at 400 b/s

[https://community.nanoporetech.com/protocols/Guppy-protocol/v/gpb\\_2003\\_v1\\_revz\\_14dec2018/guppy-software-overview](https://community.nanoporetech.com/protocols/Guppy-protocol/v/gpb_2003_v1_revz_14dec2018/guppy-software-overview)

# High accuracy basecalling with Guppy

Command line – program

A terminal window with a dark background and light text. The title bar reads "vgcare@vgcare: ~". The menu bar includes "File", "Edit", "View", "Search", "Terminal", and "Help". The command prompt shows the user is in a virtual environment "(base)". The command entered is "guppy\_basecaller --input\_path ~/test\_data/fast5\_barcode07 --save\_path ~/test\_data/fast5\_barcode07 --flowcel FLO-MIN106 --kit SQK-RBK004 -x "cuda:0"". A cursor is visible at the end of the command.

```
vgcare@vgcare: ~  
File Edit View Search Terminal Help  
(base) vgcare@vgcare:~$ guppy_basecaller --input_path ~/test_data/fast5_barcode07 --save_path ~/test_data/fast5_barcode07  
--flowcel FLO-MIN106 --kit SQK-RBK004 -x "cuda:0"
```

“guppy\_basecaller” – program command

“- - input\_path” – argument, path to input data

“- - save\_path” – argument, where to store output data

“- - flowcel” – argument, flowcell model

“- - kit” – argument, barcoding kit used

“- x” – argument, use GPU

# High accuracy basecalling with Guppy

## Command line – program

```
vgcare@vgcare: ~  
File Edit View Search Terminal Help  
(base) vgcare@vgcare:~$ guppy_basecaller --input_path ~/test_data/fast5_barcode07  
--save_path ~/test_data/fast5_barcode07 --flowcel FLO-MIN106 --kit SQK-RBK004 -x  
"cuda:0"  
ONT Guppy basecalling software version 5.0.14+8f53ee955  
config file: /opt/ont/guppy/data/dna_r9.4.1_450bps_hac.cfg  
model file: /opt/ont/guppy/data/template_r9.4.1_450bps_hac.json  
input path: /home/vgcare/test_data/fast5_barcode07  
save path: /home/vgcare/test_data/fast5_barcode07  
chunk size: 2000  
chunks per runner: 256  
minimum qscore: 9  
records per file: 4000  
num basecallers: 4  
gpu device: cuda:0  
kernel path:  
runners per device: 4  
  
Found 8 fast5 files to process.  
Init time: 584 ms  
  
0% 10 20 30 40 50 60 70 80 90 100%  
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
```

## Data Formats – fast5

- raw-data
- contain all information that was generated while sequencing
- big size
- not human-readable

# Data Formats – fastq

- q = quality
- Information about sequencing run and quality score
- Human-readable but complicated to read

```
@42b099dc-b20b-4f3d-af60-d859ebc8d976 runid=a55ad8caa1c516614cc04cc21c9dbdb356b66c41 read=101 ch=58 start_time=2021-08-20T13:47:46Z flow_cell_id=FAP81457 protocol_group_id=1234 sample_id=no_s
GTTGTACTTCGTTTCAGTTTGAATTTGGGTGTTCTACGTGTTTGGTGCCTGGGAATGAATCCTTGTTCGCGATTTATCGTGAACGCTTCGCGTTTTGGTGCGCCGCTTCATCTTGTGCAGAATGAATTCGCCGTAACCTACGTAGCACAAGTAGATGTAGTTAACTTTAATCTCCTGGCAATCTTTAA1
+
&'(.0,, -2.&).4''',+&$&$&+**3.]D8,%&&(,/-* '&&&$%(*//39/+&&%.22>;<>FC@+,,, ,65582/- -8<;44564617336810(')(*)*+)&+004; ;9//155(787.+ ,+, /+3'&+232..04+++*+*,&&0599:;81/2442422100>=], -/, ...669>?=85
@129269c0-d564-4052-8359-9c2a91f80a43 runid=a55ad8caa1c516614cc04cc21c9dbdb356b66c41 read=65 ch=316 start_time=2021-08-20T13:47:46Z flow_cell_id=FAP81457 protocol_group_id=1234 sample_id=no_s
GTACTTCACGTTCAATCAGTTGGGTGTTTAAACATTTACCGTGGGAATGGAATCCTTGATGGGGTGAAGCAACTTTCGCGTTTTTCGTGCGCCCTTCAGCATTGTCAGTACAAAAGACATACTGTTCTAATGTTGAATTCACCTTGAATTTATCAAACACTCTACACGAGCAGTGCAGGTATAATTCTACT
+
$'&&&0,&%'+00+)))+'' (+)]63, (&&$%0'''''+1864<0:427('&&&''))+, .03**((( -,+&)*, /9:961368-, -&-+ -00<=9'&&+, .., -?=@;9<5,+**;>A+./:><=@:=<7=:9;:65232348<?4>?=?;+, .86//*)'+-+.045-. -044:]]+297624€
@aa9e903f-69cf-415b-8d4e-8af728fb8cc9 runid=a55ad8caa1c516614cc04cc21c9dbdb356b66c41 read=52 ch=390 start_time=2021-08-20T13:47:46Z flow_cell_id=FAP81457 protocol_group_id=1234 sample_id=no_s
GTTGTACTTCGTTTCGATTTCAATTTGGGTGTTTATGTGTTACCGTGGGAATGGATGTTTCATGTTATCGTGAACGCTTTCGCGTTTTTCGTGCGCCAATTCACGAACGCTGTATTACAAGTGGGAGCTTCGCAGCACGTGTAGCAGAGTACTCAGGTTTTACTGCATACAGTCGCACAGGATTAGCAAC
+
$,, //..010+.33%$$%/'$$&&&&+ -];64*&'(.95***)%&'&2:10((+&' +--(*%$%%%).)&$(( '%),334))*58AB?. -*+**%(*30-, ./8843/(((, &&'0,%%')+')),,04346&&05]70110-, ,)*4304)).,9?+?++14466921128:8542074.-'%%&&'
```

# Data Formats – fastq

A single read sequence in a FASTQ file is described in four lines:

1. Line 1 begins with a '@' and is followed by a header containing information about the sequencing run.
2. Line 2 is the basecalled sequence (using A, C, T, G and N).
3. Line 3 contains a '+'.
4. Line 4 encodes the per-base quality scores for the sequence in Line 2.

An example of a FASTQ file generated by MinKNOW or Guppy is shown below:

```
@75be78f7-bd62-4972-92d2-aba16f465b0d runid=ff83cfafb0cb3bfc28ac370b841f59798ab3d63a
sampleid=RB02_lambda_ovn1 read=19343 ch=53 start_time=2019-12-23T13:44:31Z
CGGTATTACTTCGTTTCAGTTTCGGACAGGTGTTTTAACC[...]TCGTACCTAT
+
'%' - ($&&&&' (:+7) -% (&$$.%##) 868;;87/9; [...] 68 (* (2) /%$
```

[https://community.nanoporetech.com/technical\\_documents/data-analysis/v/datd\\_5000\\_v1\\_revo\\_22aug2016/fastq-files](https://community.nanoporetech.com/technical_documents/data-analysis/v/datd_5000_v1_revo_22aug2016/fastq-files)

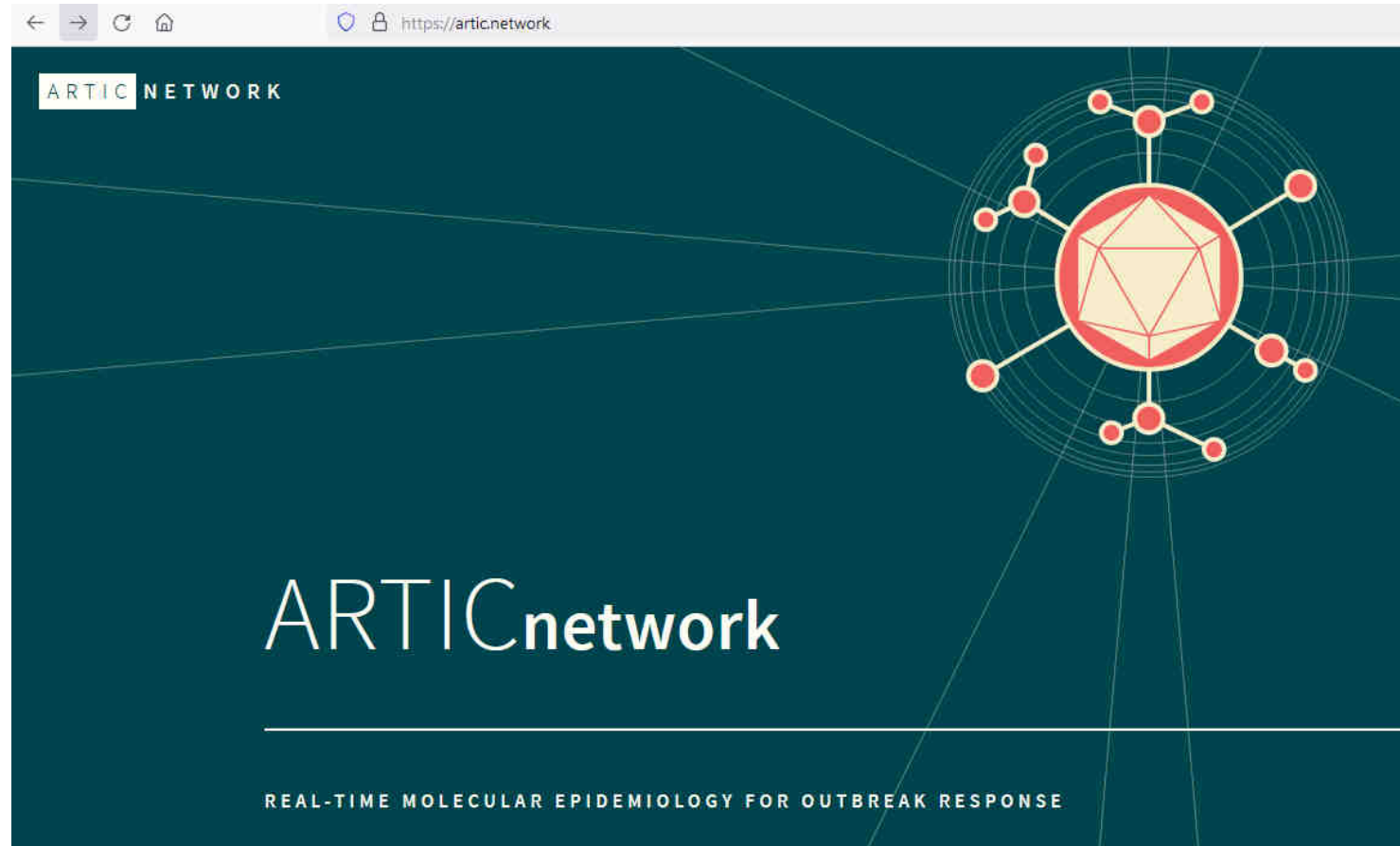






# Programs / work-flow – fieldbioinformatics

- Bioinformatics pipeline for virus sequencing, developed by the ARTIC-Network



# Programs / work-flow – fieldbioinformatics

Features include:

- read filtering
- primer trimming
- amplicon coverage normalization
- variant calling
- consensus building

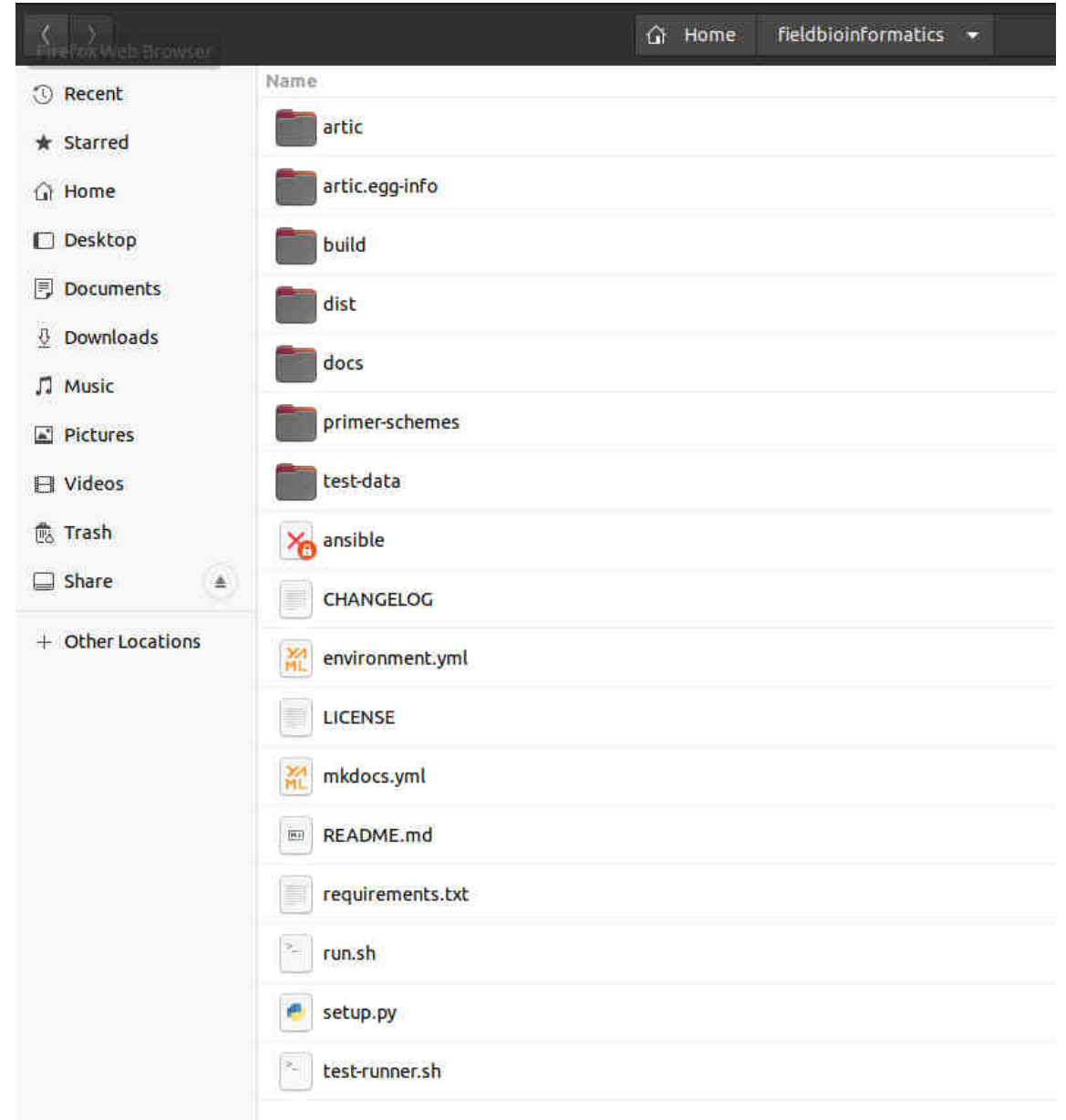
Information, download and installation-instructions:

<https://github.com/artic-network/fieldbioinformatics>

To install the fieldbioinformatics-pipeline you have to have conda installed!

# Programs / work-flow – fieldbioinformatics

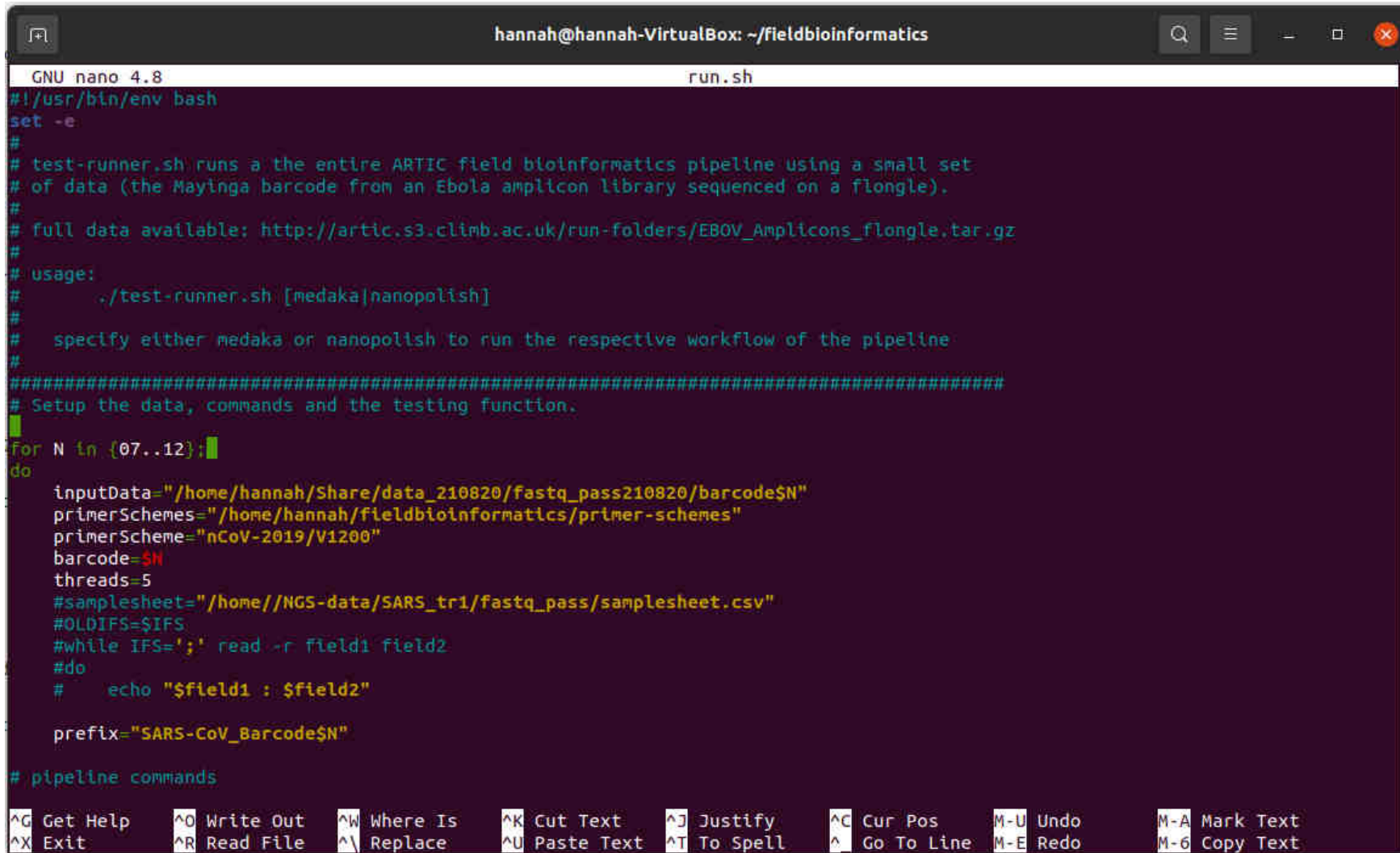
- After downloading and installing fieldbioinformatics:
    - > exchange run.sh file
    - > add our primer schemes into the fieldbioinformatics folder
- (both are available in the dropbox-folder shared with you)



# Programs / work-flow – fieldbioinformatics

```
hannah@hannah-VirtualBox: ~/fieldbioinformatics
(base) hannah@hannah-VirtualBox:~$ cd fieldbioinformatics
(base) hannah@hannah-VirtualBox:~/fieldbioinformatics$ ls -lh
total 72K
lrwxrwxrwx 1 hannah hannah 13 Aug 10 00:08 ansible -> lab-on-an-ssd
drwxrwxr-x 3 hannah hannah 4,0K Aug 10 00:08 artic
drwxrwxr-x 2 hannah hannah 4,0K Aug 10 11:52 artic.egg-info
drwxrwxr-x 4 hannah hannah 4,0K Aug 10 11:52 build
-rw-rw-r-- 1 hannah hannah 2,7K Aug 10 00:08 CHANGELOG
drwxrwxr-x 2 hannah hannah 4,0K Aug 10 11:52 dist
drwxrwxr-x 2 hannah hannah 4,0K Aug 10 00:08 docs
-rw-rw-r-- 1 hannah hannah 445 Aug 10 00:08 environment.yml
-rw-rw-r-- 1 hannah hannah 1,1K Aug 10 00:08 LICENSE
-rw-rw-r-- 1 hannah hannah 541 Aug 10 00:08 mkdocs.yml
drwxrwx--- 3 hannah hannah 4,0K Aug 10 12:02 primer-schemes
-rw-rw-r-- 1 hannah hannah 2,8K Aug 10 00:08 README.md
-rw-rw-r-- 1 hannah hannah 55 Aug 10 00:08 requirements.txt
-rwxrwx--- 1 hannah hannah 5,1K Sep 19 20:32 run.sh
-rw-rw-r-- 1 hannah hannah 1,9K Aug 10 00:08 setup.py
drwxrwxr-x 5 hannah hannah 4,0K Aug 10 00:08 test-data
-rwxrwxr-x 1 hannah hannah 4,6K Aug 10 00:08 test-runner.sh
(base) hannah@hannah-VirtualBox:~/fieldbioinformatics$ nano run.sh
```

# Programs / work-flow – fieldbioinformatics



```
hannah@hannah-VirtualBox: ~/fieldbioinformatics
GNU nano 4.8 run.sh
#!/usr/bin/env bash
set -e
#
# test-runner.sh runs a the entire ARTIC field bioinformatics pipeline using a small set
# of data (the Mayinga barcode from an Ebola amplicon library sequenced on a flongle).
#
# full data available: http://artic.s3.climb.ac.uk/run-folders/EBOV_Amplicons_flongle.tar.gz
#
# usage:
#   ./test-runner.sh [medaka|nanopolish]
#
# specify either medaka or nanopolish to run the respective workflow of the pipeline
#
#####
# Setup the data, commands and the testing function.
for N in {07..12};
do
  inputData="/home/hannah/Share/data_210820/fastq_pass210820/barcode$N"
  primerSchemes="/home/hannah/fieldbioinformatics/primer-schemes"
  primerScheme="nCoV-2019/V1200"
  barcode=$N
  threads=5
  #samplesheet="/home//NGS-data/SARS_tr1/fastq_pass/samplesheet.csv"
  #OLDIFS=$IFS
  #while IFS=';' read -r field1 field2
  #do
  #   echo "$field1 : $field2"

  prefix="SARS-CoV_Barcode$N"

# pipeline commands

^G Get Help      ^O Write Out    ^W Where Is     ^K Cut Text     ^J Justify      ^C Cur Pos      M-U Undo        M-A Mark Text
^X Exit          ^R Read File    ^_ Replace      ^U Paste Text   ^T To Spell     ^ Go To Line   M-E Redo        M-6 Copy Text
```

Modify the script for each analysis (input-path, barcodes, prefix..).

# Programs / work-flow – fieldbioinformatics

```
hannah@hannah-VirtualBox: ~/fieldbioinformatics
(base) hannah@hannah-VirtualBox:~$ cd fieldbioinformatics
(base) hannah@hannah-VirtualBox:~/fieldbioinformatics$ ls -lh
total 72K
lrwxrwxrwx 1 hannah hannah 13 Aug 10 00:08 ansible -> Lab-on-an-ssd
drwxrwxr-x 3 hannah hannah 4,0K Aug 10 00:08 artic
drwxrwxr-x 2 hannah hannah 4,0K Aug 10 11:52 artic.egg-info
drwxrwxr-x 4 hannah hannah 4,0K Aug 10 11:52 build
-rw-rw-r-- 1 hannah hannah 2,7K Aug 10 00:08 CHANGELOG
drwxrwxr-x 2 hannah hannah 4,0K Aug 10 11:52 dist
drwxrwxr-x 2 hannah hannah 4,0K Aug 10 00:08 docs
-rw-rw-r-- 1 hannah hannah 445 Aug 10 00:08 environment.yml
-rw-rw-r-- 1 hannah hannah 1,1K Aug 10 00:08 LICENSE
-rw-rw-r-- 1 hannah hannah 541 Aug 10 00:08 mkdocs.yml
drwxrwx--- 3 hannah hannah 4,0K Aug 10 12:02 primer-schemes
-rw-rw-r-- 1 hannah hannah 2,8K Aug 10 00:08 README.md
-rw-rw-r-- 1 hannah hannah 55 Aug 10 00:08 requirements.txt
-rwxrwx--- 1 hannah hannah 5,1K Sep 19 20:32 run.sh
-rw-rw-r-- 1 hannah hannah 1,9K Aug 10 00:08 setup.py
drwxrwxr-x 5 hannah hannah 4,0K Aug 10 00:08 test-data
-rwxrwxr-x 1 hannah hannah 4,6K Aug 10 00:08 test-runner.sh
(base) hannah@hannah-VirtualBox:~/fieldbioinformatics$ nano run.sh
(base) hannah@hannah-VirtualBox:~/fieldbioinformatics$ chmod u+x run.sh
(base) hannah@hannah-VirtualBox:~/fieldbioinformatics$
```

When running the first time, you may have to add execution-permissions to the script.



# Programs / work-flow – fieldbioinformatics

```
hannah@hannah-VirtualBox: ~/fieldbioinformatics
(base) hannah@hannah-VirtualBox:~$ cd fieldbioinformatics/
(base) hannah@hannah-VirtualBox:~/fieldbioinformatics$ ls -lh
total 72K
lrwxrwxrwx 1 hannah hannah 13 Aug 10 00:08 ansible -> lab-on-an-ssd
drwxrwxr-x 3 hannah hannah 4,0K Aug 10 00:08 artic
drwxrwxr-x 2 hannah hannah 4,0K Aug 10 11:52 artic.egg-info
drwxrwxr-x 4 hannah hannah 4,0K Aug 10 11:52 build
-rw-rw-r-- 1 hannah hannah 2,7K Aug 10 00:08 CHANGELOG
drwxrwxr-x 2 hannah hannah 4,0K Aug 10 11:52 dist
drwxrwxr-x 2 hannah hannah 4,0K Aug 10 00:08 docs
-rw-rw-r-- 1 hannah hannah 445 Aug 10 00:08 environment.yml
-rw-rw-r-- 1 hannah hannah 1,1K Aug 10 00:08 LICENSE
-rw-rw-r-- 1 hannah hannah 541 Aug 10 00:08 mkdocs.yml
drwxrwx--- 3 hannah hannah 4,0K Aug 10 12:02 primer-schemes
-rw-rw-r-- 1 hannah hannah 2,8K Aug 10 00:08 README.md
-rw-rw-r-- 1 hannah hannah 55 Aug 10 00:08 requirements.txt
-rwxrwx--- 1 hannah hannah 5,1K Sep 19 20:32 run.sh
-rw-rw-r-- 1 hannah hannah 1,9K Aug 10 00:08 setup.py
drwxrwxr-x 5 hannah hannah 4,0K Aug 10 00:08 test-data
-rwxrwxr-x 1 hannah hannah 4,6K Aug 10 00:08 test-runner.sh
(base) hannah@hannah-VirtualBox:~/fieldbioinformatics$ nano run.sh
(base) hannah@hannah-VirtualBox:~/fieldbioinformatics$ chmod u+x run.sh
(base) hannah@hannah-VirtualBox:~/fieldbioinformatics$ conda activate artic
(artic) hannah@hannah-VirtualBox:~/fieldbioinformatics$
```

Activate the correct conda environment.

# Programs / work-flow – fieldbioinformatics

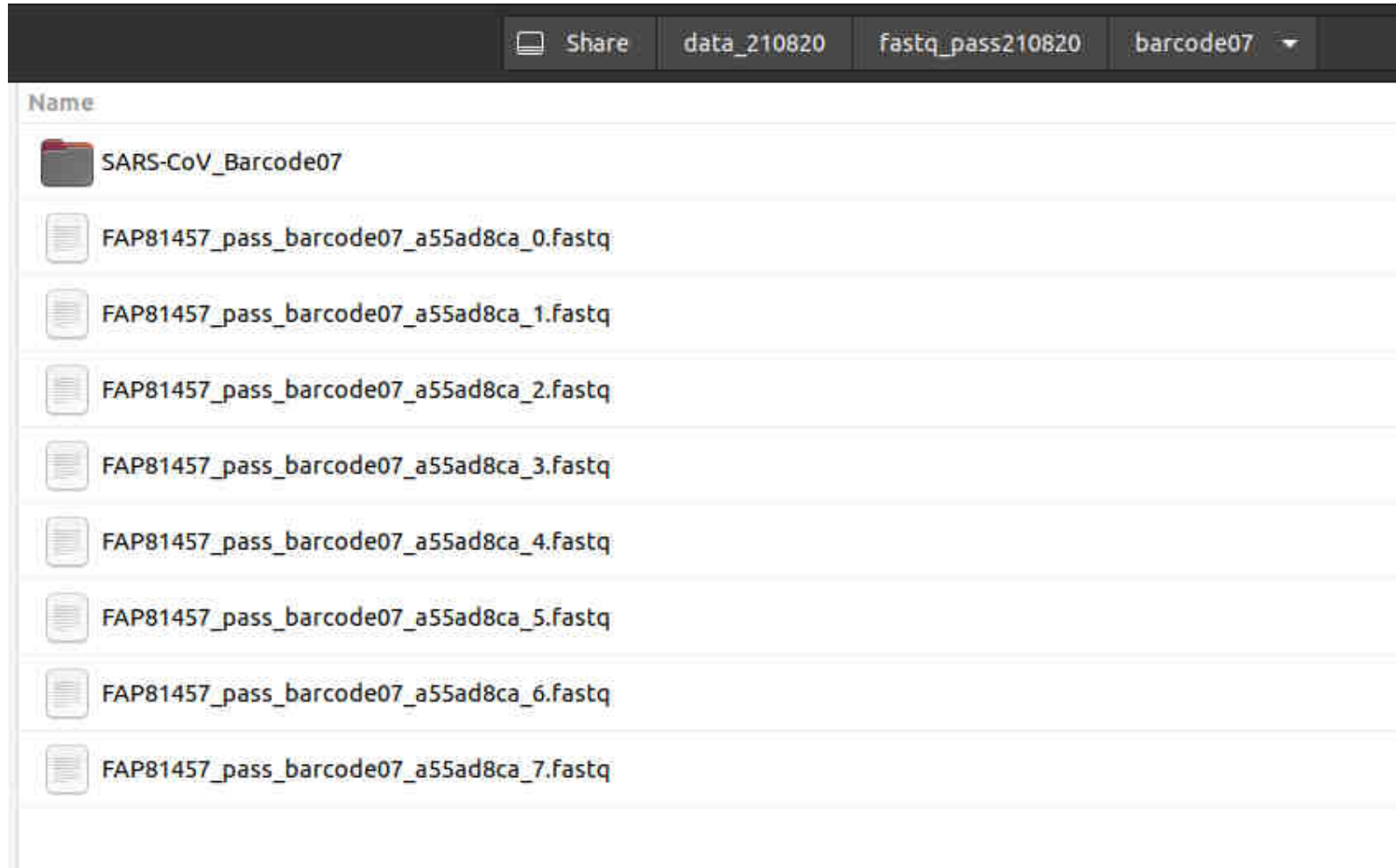
```
hannah@hannah-VirtualBox: ~/fieldbioinformatics
total 72K
lrwxrwxrwx 1 hannah hannah 13 Aug 10 00:08 ansible -> lab-on-an-ssd
drwxrwxr-x 3 hannah hannah 4,0K Aug 10 00:08 artic
drwxrwxr-x 2 hannah hannah 4,0K Aug 10 11:52 artic.egg-info
drwxrwxr-x 4 hannah hannah 4,0K Aug 10 11:52 build
-rw-rw-r-- 1 hannah hannah 2,7K Aug 10 00:08 CHANGELOG
drwxrwxr-x 2 hannah hannah 4,0K Aug 10 11:52 dist
drwxrwxr-x 2 hannah hannah 4,0K Aug 10 00:08 docs
-rw-rw-r-- 1 hannah hannah 445 Aug 10 00:08 environment.yml
-rw-rw-r-- 1 hannah hannah 1,1K Aug 10 00:08 LICENSE
-rw-rw-r-- 1 hannah hannah 541 Aug 10 00:08 mkdocs.yml
drwxrwx--- 3 hannah hannah 4,0K Aug 10 12:02 primer-schemes
-rw-rw-r-- 1 hannah hannah 2,8K Aug 10 00:08 README.md
-rw-rw-r-- 1 hannah hannah 55 Aug 10 00:08 requirements.txt
-rwxrwx--- 1 hannah hannah 5,1K Sep 19 20:32 run.sh
-rw-rw-r-- 1 hannah hannah 1,9K Aug 10 00:08 setup.py
drwxrwxr-x 5 hannah hannah 4,0K Aug 10 00:08 test-data
-rwxrwxr-x 1 hannah hannah 4,6K Aug 10 00:08 test-runner.sh
(base) hannah@hannah-VirtualBox:~/fieldbioinformatics$ nano run.sh
(base) hannah@hannah-VirtualBox:~/fieldbioinformatics$ chmod u+x run.sh
(base) hannah@hannah-VirtualBox:~/fieldbioinformatics$ conda activate artic
(artic) hannah@hannah-VirtualBox:~/fieldbioinformatics$ ./run.sh medaka
Starting tests...
- using the medaka workflow

running the pipeline...
#####
Running: artic gather --min-length 400 --max-length 1200 --prefix SARS-CoV_Barcode07 --directory /home/hannah/Share/data_21
0820/fastq_pass210820/barcode07 --no-fast5s

Processing 8 files in barcode07
SARS-CoV_Barcode07_barcode07.fastq      23426    23426
Processing 1 files in SARS-CoV_Barcode07
```

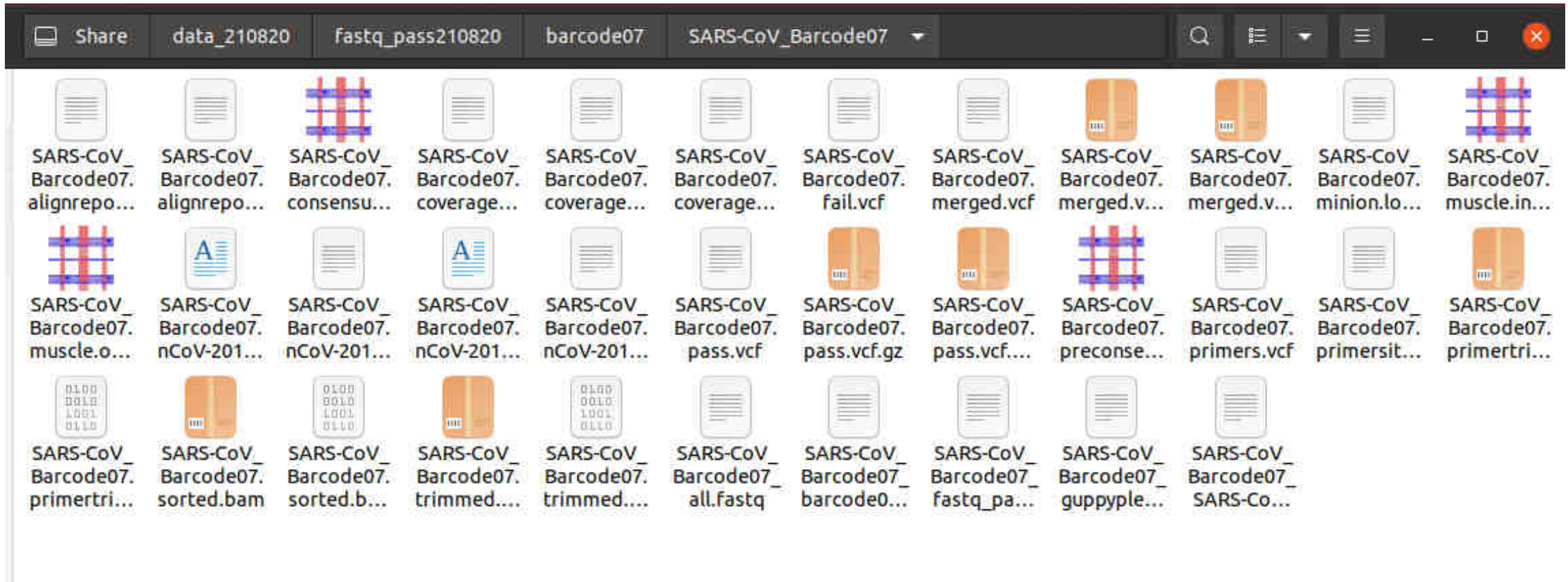
Run the pipeline.

# Programs / work-flow – fieldbioinformatics



The output-folder will be in the same place as the input-data.

# Programs / work-flow – fieldbioinformatics



The content of the output-folder.

# Programs / work-flow – fieldbioinformatics

```
Open  SARS-CoV_Barcode07.consensus.fasta  Save  ≡
Share ~/Share/data_210820/fastq_...0820/barcode07/SARS...

1 >SARS-CoV_Barcode07/ARTIC/medaka MN908947.3
2 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNAGATCT
3 GTTCTCTAAACGAACCTTTAAAATCTGTGTGGCTGTCACCTCGGCTGCATGCTTAGTGCACT
4 CACGCAGTATAATTAATAACTAATTACTGTCGTTGACAGGACACGAGTAACTCGTCTATC
5 TTCTGCAGGCTGCTTACGGTTTCGTCCGTGTTGCAGCCGATCATCAGCACATCTAGGTTT
6 TGTCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCTGGTTTCAACGAGAAAAC
7 ACACGTCCAACCTCAGTTTGCCTGTTTTACAGGTTGCGGACGTGCTCGTACGTGGCTTTGG
8 AGACTCCGTGGAGGAGGTCTTATCAGAGGCACGTCAACATCTTAAAGATGGCACTTGTGG
9 CTTAGTAGAAGTTGAAAAAGGCGTTTTGCCTCAACTTGAACAGCCCTATGTGTTCACTAA
10 ACGTTTCGGATGCTCGAACTGCACCTCATGGTCATGTTATGGTTGAGCTGGTAGCAGAACT
11 CGAAGGCATTCAGTACGGTCGTAGTGGTGAGACACTTGGTGTCTTGTCCCTCATGTGGG
12 CGAAATACCAGTGGCTTACCGCAAGGTTCTTCTTCGTAAGAACGGTAATAAAGGAGCTGG
13 TGGCCATAGTTACGGCGCCGATCTAAAGTCATTTGACTTAGGCGACGAGCTTGGCACTGA
14 TCCTTATGAAGATTTTCAAGAAAACCTGGAACACTAAACATAGCAGTGGTGTACCCGTA
15 ACTCATGCGTGAGCTTAACGGAGGGGCATACACTCGCTATGTGATAACAACCTTCTGTGG
16 CCCTGATGGCTACCCTCTTGAGTGCATTAAGACCTTCTAGCACGTGCTGGTAAAGCTTC
17 ATGCACTTTGTCCGAACAACCTGGACTTTATTGACACTAAGAGGGGTGTATACTGCTGCCG
18 TGAACATGAGCATGAAATTGCTTGGTACACGGAACGTTCTGAAAAGAGCTATGAATTGCA
19 GACACCTTTTGAAATTAATTTGGCAAAGAAATTTGACACCTTCAATGGGGAATGTCCAAA
20 TTTTGTATTTCCCTTAAATTCCATAATCAAGACTATTCAACCAAGGGTTNNAAGAAAAA
```

The “.consensus.fasta” file is the one, which interests us most.

# Programs / work-flow – fieldbioinformatics

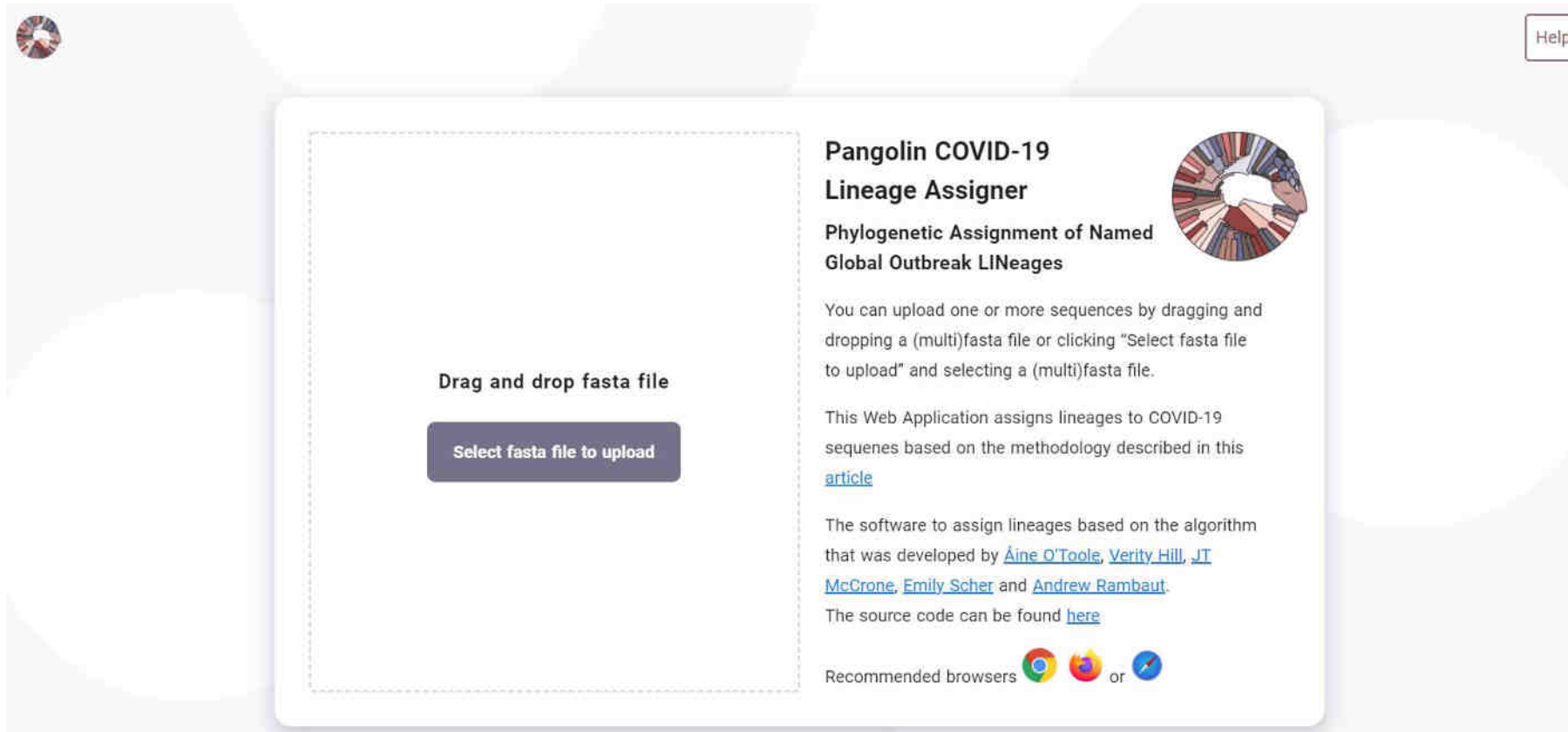
```
*SARS-CoV_Barcode07.nCoV-2019_1.vcf
Share ~/Share/data_210820/fastq_...0820/barcode07/SARS-CoV_Barcode07

1 ##fileformat=VCFv4.1
2 ##medaka_version=1.0.3
3 ##contig=<ID=MN908947.3>
4 ##FORMAT=<ID=GT,Number=1,Type=String,Description="Medaka genotype.">
5 ##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Medaka genotype quality score">
6 #CHROM      POS      ID      REF      ALT      QUAL      FILTER  INFO      FORMAT  SAMPLE
7 MN908947.3  40       .       C        T        36.972    PASS    .         GT:GQ   1:37
8 MN908947.3  241      .       C        T        55.435    PASS    .         GT:GQ   1:55
9 MN908947.3  1130     .       GA       G        0.817     PASS    .         GT:GQ   1:1
10 MN908947.3 3037     .       C        T        23.787    PASS    .         GT:GQ   1:24
11 MN908947.3 4457     .       A        G        42.077    PASS    .         GT:GQ   1:42
12 MN908947.3 4891     .       C        T        42.832    PASS    .         GT:GQ   1:43
13 MN908947.3 5184     .       C        T        35.099    PASS    .         GT:GQ   1:35
14 MN908947.3 9204     .       A        G        32.587    PASS    .         GT:GQ   1:33
15 MN908947.3 11074    .       CT       C        8.197     PASS    .         GT:GQ   1:8
16 MN908947.3 11430    .       A        G        31.292    PASS    .         GT:GQ   1:31
17 MN908947.3 15187    .       A        G        40.072    PASS    .         GT:GQ   1:40
18 MN908947.3 15324    .       C        T        48.63     PASS    .         GT:GQ   1:49
19 MN908947.3 22802    .       C        A        40.773    PASS    .         GT:GQ   1:41
20 MN908947.3 22912    .       T        G        51.175    PASS    .         GT:GQ   1:51
21 MN908947.3 23403    .       A        G        35.788    PASS    .         GT:GQ   1:36
22 MN908947.3 25445    .       GTGAAATCAAGGATGCTACTCCTTCAGATTT  G        1029.872
23 MN908947.3 27576    .       T        C        36.103    PASS    .         GT:GQ   1:36
24 MN908947.3 28846    .       C        T        46.757    PASS    .         GT:GQ   1:47
25 MN908947.3 28887    .       C        T        31.681    PASS    .         GT:GQ   1:32
26 MN908947.3 29730    .       C        T        32.8      PASS    .         GT:GQ   1:33
```

The “.vcf” file lists all variations from the original WUHAN-sequence.

# Analysis of the sequences for variants/clades/lineages/mutations

<https://pangolin.cog-uk.io/>



The screenshot shows the Pangolin COVID-19 Lineage Assigner web application. On the left, there is a large dashed box for file upload with the text "Drag and drop fasta file" and a button labeled "Select fasta file to upload". On the right, the application title "Pangolin COVID-19 Lineage Assigner" is displayed, along with a circular logo. Below the title, the subtitle "Phylogenetic Assignment of Named Global Outbreak LINEages" is shown. The main text explains that users can upload sequences by dragging and dropping a (multi)fasta file or clicking "Select fasta file to upload". It also states that the application assigns lineages based on the methodology described in a linked article. At the bottom, it lists the developers: Aine O'Toole, Verity Hill, JT McCrone, Emily Scher, and Andrew Rambaut, and provides a link to the source code. Recommended browsers (Chrome, Firefox, or Safari) are also indicated.




**Pangolin COVID-19 Lineage Assigner**

Phylogenetic Assignment of Named Global Outbreak LINEages

You can upload one or more sequences by dragging and dropping a (multi)fasta file or clicking "Select fasta file to upload" and selecting a (multi)fasta file.

This Web Application assigns lineages to COVID-19 sequences based on the methodology described in this [article](#)

The software to assign lineages based on the algorithm that was developed by [Aine O'Toole](#), [Verity Hill](#), [JT McCrone](#), [Emily Scher](#) and [Andrew Rambaut](#). The source code can be found [here](#)

Recommended browsers   or 

# Analysis of the sequences for variants/clades/lineages/mutations

<https://clades.nextstrain.org/>



## Nextclade <sup>beta</sup> v0.14.2

Clade assignment, mutation calling, and sequence quality checks

<b>Simple</b> No installation or setup - drop a file and see the results	<b>Private</b> No remote processing - sequence data never leaves your computer		
<b>Mutation Calling</b> Find differences of your sequences relative to the reference in standard numbering	<b>Clade Assignment</b> Find out which Nextstrain clades your samples are from	<b>Phylogenetic Placement</b> See where on the SARS-CoV-2 tree your sequences fall	<b>Quality Control</b> Check your data against multiple QC metrics

SARS-CoV-2  Simple mode  Advanced mode

\* Sequences **required**

Drag & Drop a file here  
or

[Show me an Example](#)



# High accuracy basecalling with Guppy



◀ Back

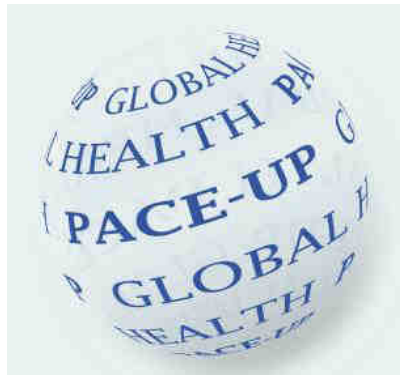
Done. Total sequences: 5. Succeeded: 5

ID	Sequence name	QC	Clade	Mut.	non-ACGTN	Ns	Gaps	Ins.
0	✔ SARS-CoV_high_accuracy_Barcode07/ARTIC/mex	N M P C F S	20A	25	0	220	2	9
1	✔ SARS-CoV_high_accuracy_Barcode08/ARTIC/mex	N M P C F S	20A	26	0	220	2	9
2	✔ SARS-CoV_high_accuracy_Barcode09/ARTIC/mex	N M P C F S	20A	30	0	220	11	9
3	✔ SARS-CoV_high_accuracy_Barcode10/ARTIC/mex	N M P C F S	20A	29	0	220	2	9
4	✔ SARS-CoV_high_accuracy_Barcode11/ARTIC/mex	N M P C F S	19B	3	0	189	0	0

# THANK YOU

---





  
**Alexander von Humboldt**  
Stiftung/Foundation

EBERHARD KARLS  
**UNIVERSITÄT**  
**TÜBINGEN**



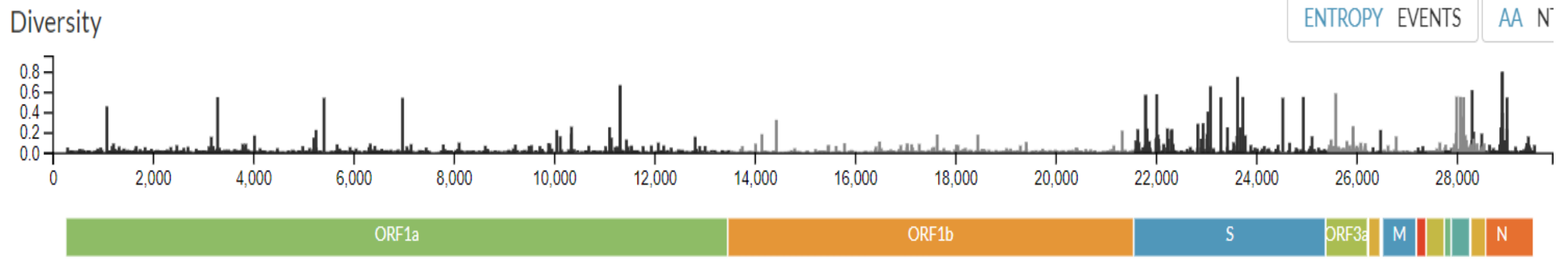
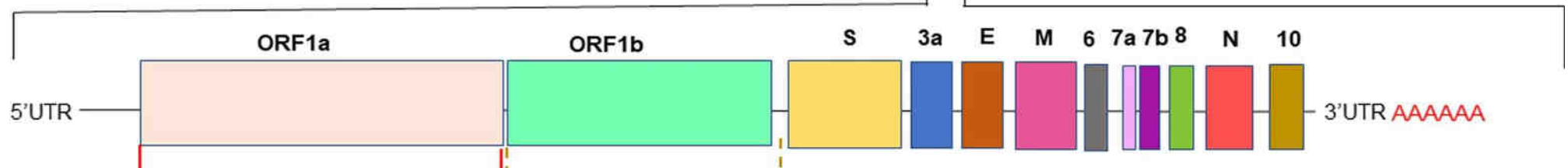
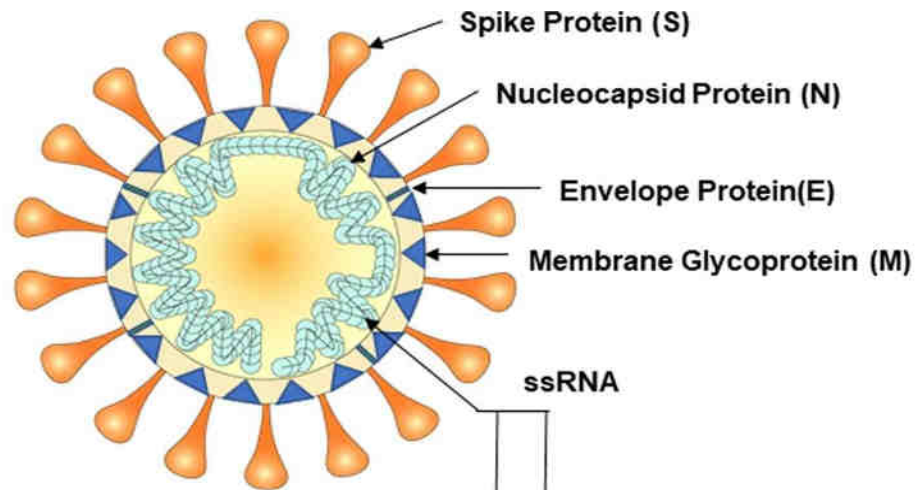
---

# Workshop

## Molecular Diagnostic Techniques (Genomics and Bioinformatics – Infectious diseases) module-5

Dr. Srinivas Reddy Pallerla, Mr. Tran Thanh Tung and Ms. Linh Le  
Prof. Dr. Thirumalaisamy P. Velavan, University Hospital, Tübingen, Germany  
Prof. Dr. Daniel Olusola Ojurongbe, Nigeria  
Prof. Dr. Mohamed Osman, University of Khartoum, Sudan

# SARS-CoV-2 genome – 29.9 kb



## Rationale for sequencing of SARS-CoV-2

- Evolution and epidemiology of the virus
- Understand transmissibility, morbidity and mortality
- Evade detection by diagnostic tests
- Decreased neutralizing to antibodies (e.g., convalescent plasma or monoclonal antibodies)
- Evade natural immunity (e.g., causing reinfections)
- Infect vaccinated individuals (outbreak infections)
- Monitor variants of concern (VOC) and variants of interest (VOI)
- Monitor trends at the national level
- Monitor emergence of important new strains
- Monitor trends after interventions such as vaccination
- Better understand epidemiology at the local level

# Global - genomic surveillance of SARS-CoV-2

GISAID (Global Initiative on Sharing All Influenza Data)

hCoV-19 data sharing via GISAID

**3,545,029**  
submissions



- <https://outbreak.info/>
- <https://cov-lineages.org/index.html>
- <https://nextstrain.org/ncov/global>



- Global/Country/State/City to local
- Circulating variants
- New viral lineages
- Variants of interest, under investigation, of concern
- Distribution over time and space
- No of mutation accumulating
- Much more.....

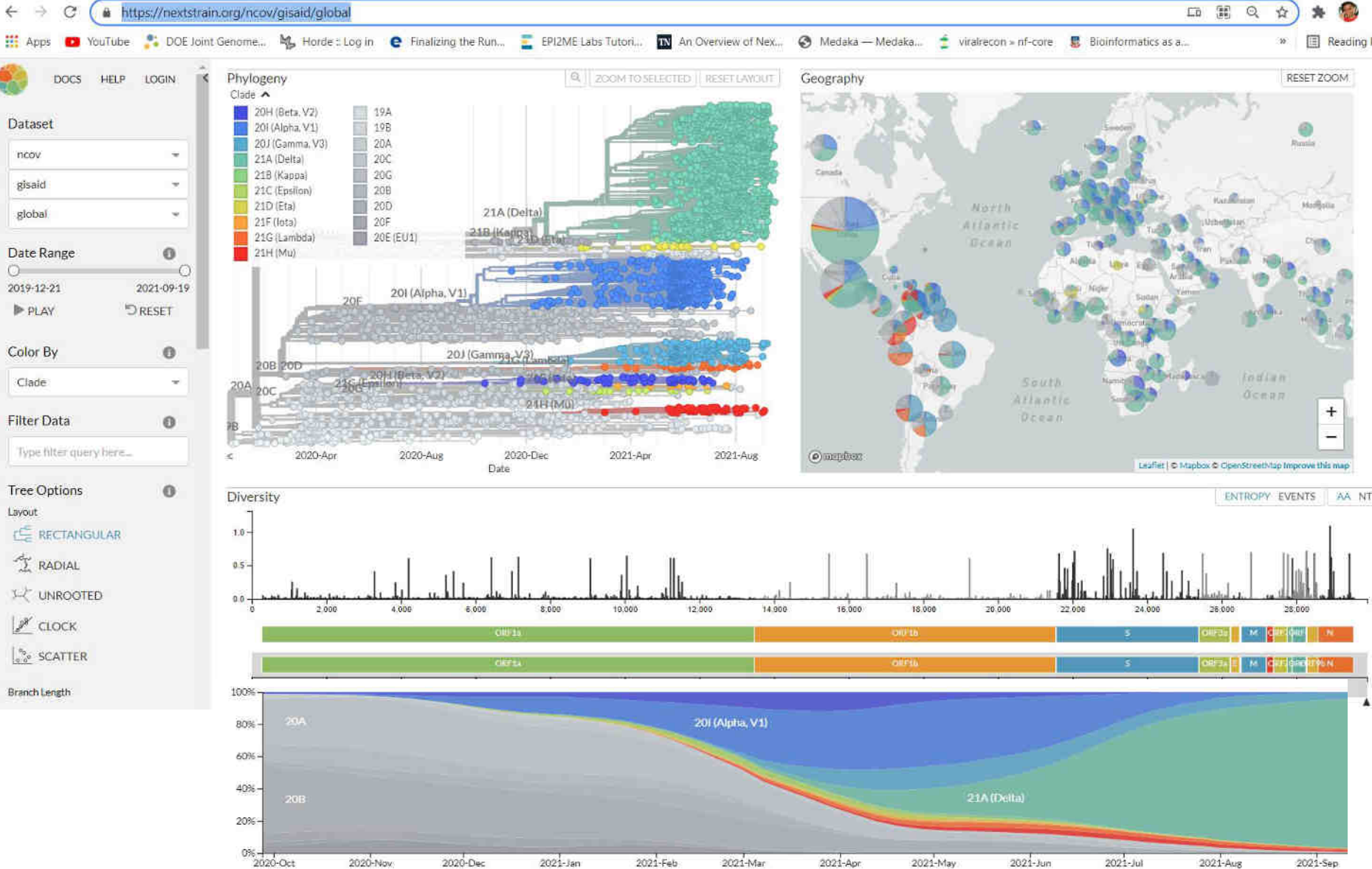
## Nomenclatures for SARS-CoV-2 variants

**GISAID Clades:** (e.g. L, V, S, O, G, GH, GR, GV, GRY) ; <https://www.gisaid.org/>

**NeXTstrain Clades:** Major clades: 19A, 19B, and 20A–20I ; <https://nextstrain.org/ncov/gisaid/global>

**PANGOLIN Lineages:** Phylogenetic Assignment of Named Global Outbreak Lineages ; <https://cov-lineages.org/> Predominant - B lineage: e.g. B.1.1.7 (Alpha), B.1.351 (Beta), B.1.617.2 (Delta) etc.

# Genomic epidemiology of novel coronavirus - Global subsampling (nextstrain.org)





# Genomic epidemiology of novel coronavirus (outbreak.info)

The screenshot shows the outbreak.info website in a browser. The browser's address bar displays 'https://outbreak.info'. The website's navigation menu includes 'Cases & Deaths', 'Variants', 'Research Library', 'API', and 'About'. The main content area is divided into three sections: 'COVID-19 Cases & Deaths' with a search bar for location (suggestions: Brazil, New York City, Kansas City metro area); 'Variants' with search bars for PANGO lineage (suggestions: Delta, B.1.617.2, B.1.617, Alpha / B.1.17) and location (suggestions: USA, UK, New York, San Diego); and 'Research Library' with a search bar for resources (suggestions: E484K, Moderna, remdesivir, NIAID-funded). Below the navigation is a descriptive paragraph: 'Outbreak.info is a project from the [Su](#), [Wu](#), and [Andersen](#) labs at Scripps Research to unify COVID-19 and SARS-CoV-2 epidemiology and genomic data, published research, and other resources. Researchers, health officials, and the public can track the pandemic using data on cases, deaths, and genomic variants, and stay updated on related research through interactive visualizations, a searchable library, and downloadable raw data.' To the right of the text is a video player showing a large red circle on a dark blue background, with a progress bar at 0:00 / 1:05.

<https://outbreak.info/>

Details will be showed in the browser

# Phylogenetic analysis of SARS-CoV-2 genomes

## Short video on the phylogenetics

<https://www.cdc.gov/amd/training/covid-toolkit/module1-3.html>

<https://www.youtube.com/watch?v=Ny38Aa2edk8>

<https://www.cdc.gov/amd/training/covid-19-gen-epi-toolkit.html>

# Load MAFFT and IQTREE programmes using following commands in Linux terminal

## 12.5 Phylogeny

For phylogeny you can use a program called **MegaX** <https://megasoftware.net/> which comes with a graphical user interface. There is a free version which allows all important steps, for more features you can buy a full version. But MegaX is rather slow, especially for large numbers of samples.

Alternatively there are some command line tools.

**MAFFT** <https://mafft.cbrc.jp/alignment/software/> is a multiple sequence alignment program. It can be installed via the command line

```
$ sudo apt install mafft
```

To run mafft type:

```
$ mafft all_barcodes_merged.fasta > all_barcodes_alinged.fasta (input > output)
```

**IQTREE** <http://www.iqtree.org/> is a phylogeny program. It can be installed via the command line.

```
$ sudo apt install iqtree
```

To run type:

```
$ iqtree -s all_barcodes_alinged.fasta -bb 1000
```

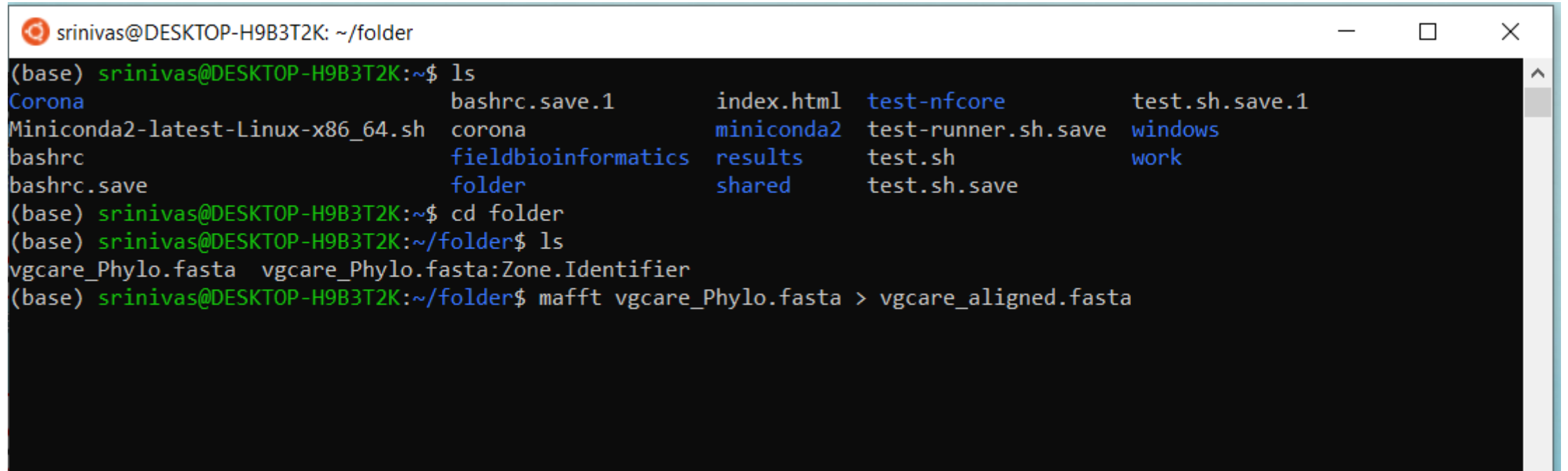
“-bb 100” is to set the ultrafast bootstrap (UFBoot) feature to 1000 replicates which will give more unbiased values.

To visualize your phylogenetic analysis you can load the treefile of the iqtree-output to the **ITOL-website**: <https://itol.embl.de/upload.cgi>

1. Create a folder (using command mkdir xxx) and copy vgcare\_phylo.fasta file from dropbox

```
srinivas@DESKTOP-H9B3T2K: ~/folder
(base) srinivas@DESKTOP-H9B3T2K:~$ ls
Corona          bashrc.save.1      index.html         test-nfcore        test.sh.save.1
Miniconda2-latest-Linux-x86_64.sh  corona            miniconda2        test-runner.sh.save  windows
bashrc          fieldbioinformatics  results           test.sh            work
bashrc.save    folder              shared            test.sh.save
(base) srinivas@DESKTOP-H9B3T2K:~$ cd folder
(base) srinivas@DESKTOP-H9B3T2K:~/folder$ ls
vgcare_Phylo.fasta  vgcare_Phylo.fasta:Zone.Identifier
(base) srinivas@DESKTOP-H9B3T2K:~/folder$
```

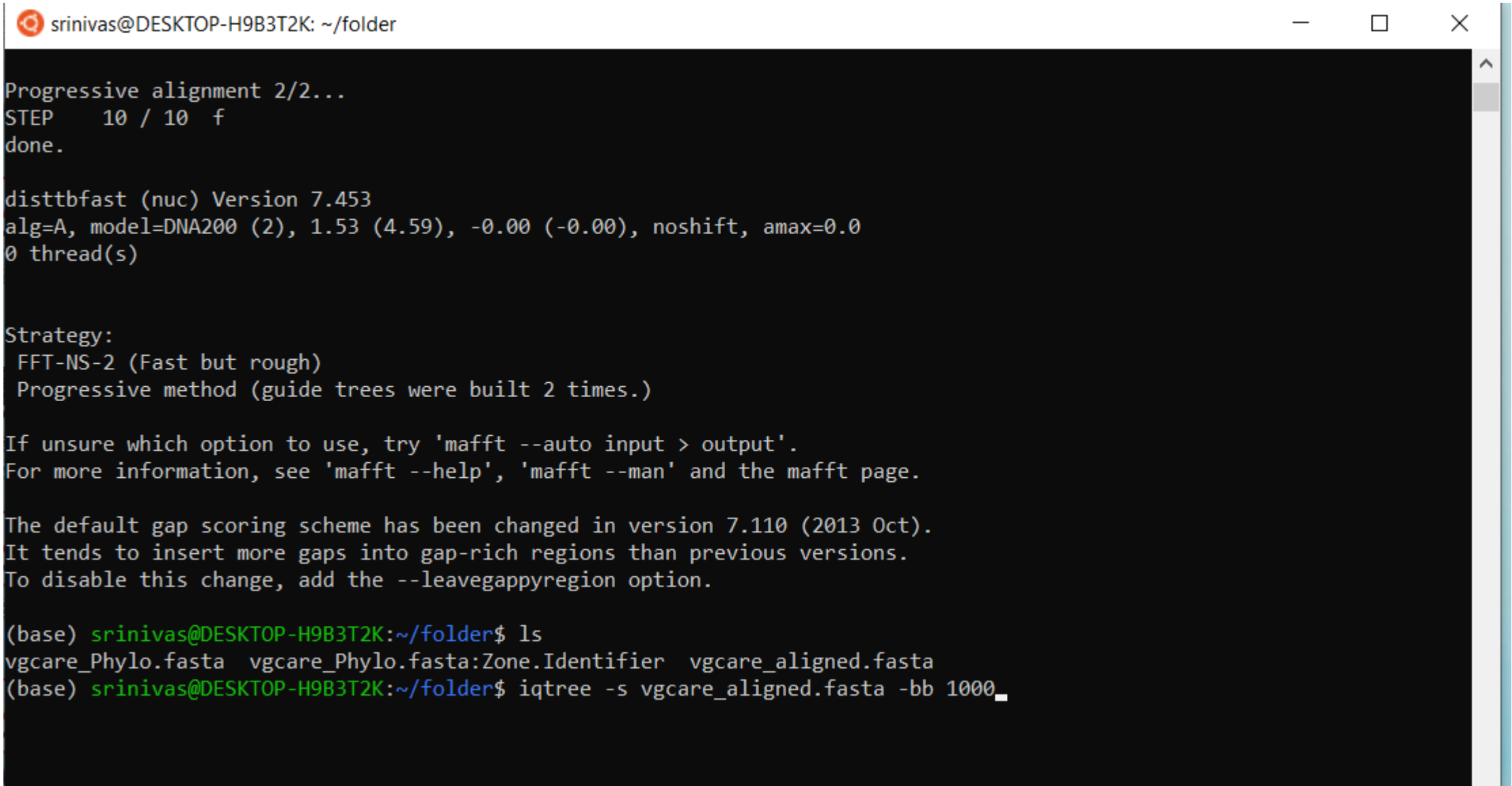
## 2. \$ mafft input > output files as shown below



```
srinivas@DESKTOP-H9B3T2K: ~/folder
(base) srinivas@DESKTOP-H9B3T2K:~$ ls
Corona          bashrc.save.1    index.html      test-nfcore     test.sh.save.1
Miniconda2-latest-Linux-x86_64.sh  corona          miniconda2     test-runner.sh.save  windows
bashrc          fieldbioinformatics  results        test.sh         work
bashrc.save    folder            shared         test.sh.save
(base) srinivas@DESKTOP-H9B3T2K:~$ cd folder
(base) srinivas@DESKTOP-H9B3T2K:~/folder$ ls
vgcare_Phylo.fasta  vgcare_Phylo.fasta:Zone.Identifier
(base) srinivas@DESKTOP-H9B3T2K:~/folder$ mafft vgcare_Phylo.fasta > vgcare_aligned.fasta
```

## Load MAFFT and IQTREE programmes using following commands in Linux terminal

3. \$ "iqtree -s vgcare\_aligned.fasta -bb 1000" command



```
srinivas@DESKTOP-H9B3T2K: ~/folder
Progressive alignment 2/2...
STEP 10 / 10 f
done.

disttbfast (nuc) Version 7.453
alg=A, model=DNA200 (2), 1.53 (4.59), -0.00 (-0.00), noshift, amax=0.0
0 thread(s)

Strategy:
FFT-NS-2 (Fast but rough)
Progressive method (guide trees were built 2 times.)

If unsure which option to use, try 'mafft --auto input > output'.
For more information, see 'mafft --help', 'mafft --man' and the mafft page.

The default gap scoring scheme has been changed in version 7.110 (2013 Oct).
It tends to insert more gaps into gap-rich regions than previous versions.
To disable this change, add the --leavegappyregion option.

(base) srinivas@DESKTOP-H9B3T2K:~/folder$ ls
vgcare_Phylo.fasta  vgcare_Phylo.fasta:Zone.Identifier  vgcare_aligned.fasta
(base) srinivas@DESKTOP-H9B3T2K:~/folder$ iqtree -s vgcare_aligned.fasta -bb 1000_
```

#### 4. \$ several files generated and use "ls" command to check and copy all the files to the desktop

```
Select srinivas@DESKTOP-H9B3T2K: ~/folder
-----
FINALIZING TREE SEARCH
-----
Performs final model parameters optimization
Estimate model parameters (epsilon = 0.010)
1. Initial log-likelihood: -41106.764
Optimal log-likelihood: -41106.762
Rate parameters: A-C: 1.00000 A-G: 2.27693 A-T: 1.00000 C-G: 1.00000 C-T: 8.16370 G-T: 1.00000
Base frequencies: A: 0.299 C: 0.184 G: 0.196 T: 0.322
Parameters optimization took 1 rounds (0.006 sec)
BEST SCORE FOUND : -41106.762
Creating bootstrap support values...
Split supports printed to NEXUS file vgcare_aligned.fasta.splits.nex
Total tree length: 0.002

Total number of iterations: 200
CPU time used for tree search: 1.170 sec (0h:0m:1s)
Wall-clock time used for tree search: 1.180 sec (0h:0m:1s)
Total CPU time used: 1.212 sec (0h:0m:1s)
Total wall-clock time used: 1.223 sec (0h:0m:1s)

Computing bootstrap consensus tree...
Reading input file vgcare_aligned.fasta.splits.nex...
11 taxa and 44 splits.
Consensus tree written to vgcare_aligned.fasta.contree
Reading input trees file vgcare_aligned.fasta.contree
Log-likelihood of consensus tree: -41107.272

Analysis results written to:
IQ-TREE report:          vgcare_aligned.fasta.iqtree
Maximum-likelihood tree: vgcare_aligned.fasta.treefile
Likelihood distances:    vgcare_aligned.fasta.mldist













Ultrafast bootstrap approximation results written to:
Split support values:    vgcare_aligned.fasta.splits.nex
Consensus tree:         vgcare_aligned.fasta.contree
Screen log file:        vgcare_aligned.fasta.log

Date and Time: Tue Sep 21 16:14:52 2021
(base) srinivas@DESKTOP-H9B3T2K:~/folder$
```

## 4. Copied files on the desktop

his PC > ubuntu-20.04 (\\wsl\$) (Z:) > home > srinivas > folder

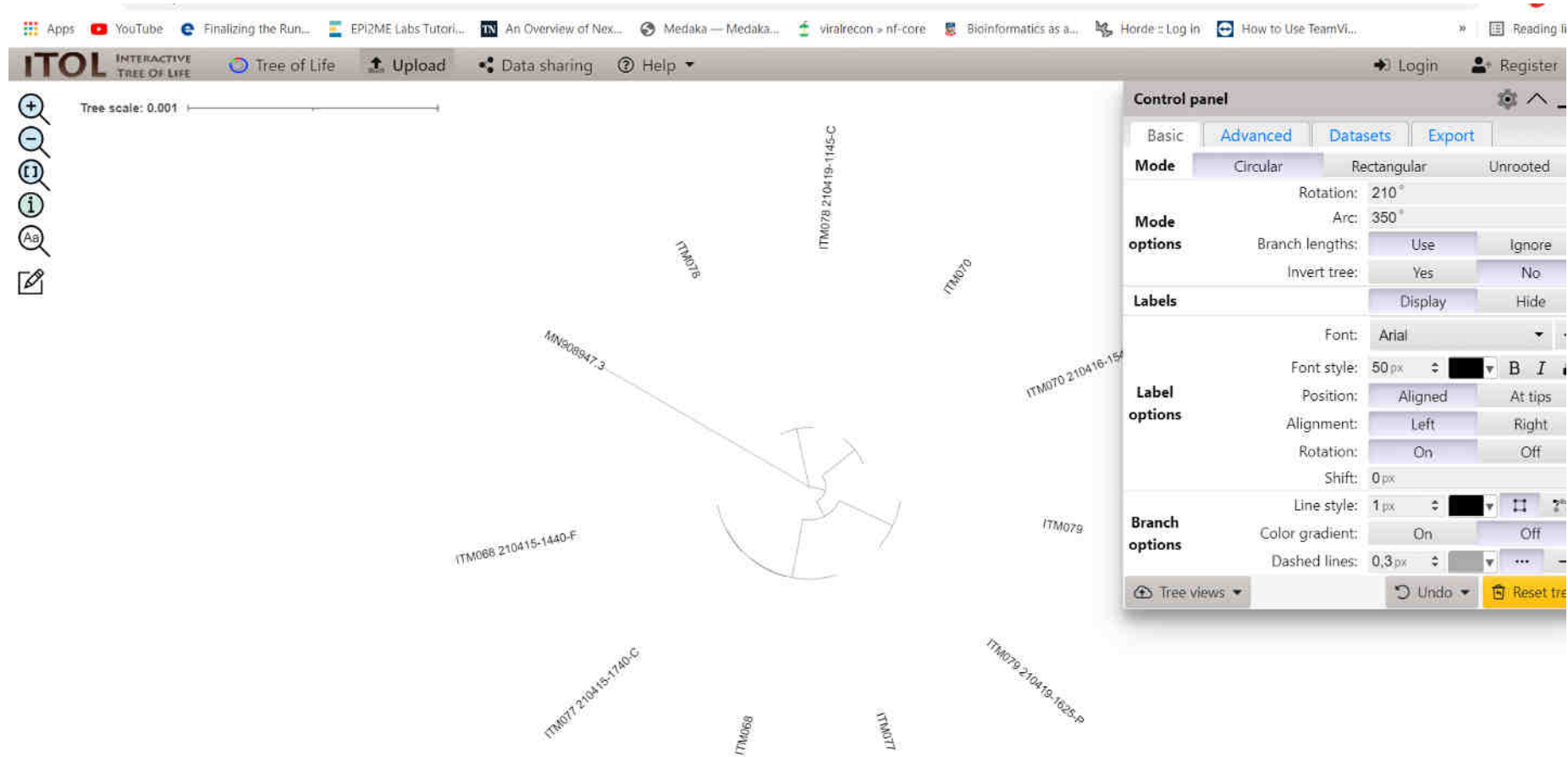
ler

Name	Date modified	Type	Size
 vgcare_aligned.fasta	21/09/2021 16:12	FASTA File	327 KB
 vgcare_aligned.fasta.bionj	21/09/2021 16:14	BIONJ File	1 KB
 vgcare_aligned.fasta.ckp.gz	21/09/2021 16:14	GZ File	15 KB
 vgcare_aligned.fasta.contree	21/09/2021 16:14	CONTREE File	1 KB
 vgcare_aligned.fasta.iqtree	21/09/2021 16:14	IQTREE File	18 KB
 vgcare_aligned.fasta.log	21/09/2021 16:14	Text Document	17 KB
 vgcare_aligned.fasta.mldist	21/09/2021 16:14	MLDIST File	2 KB
 vgcare_aligned.fasta.model.gz	21/09/2021 16:14	GZ File	3 KB
 vgcare_aligned.fasta.splits.nex	21/09/2021 16:14	NEX File	1 KB
 vgcare_aligned.fasta.treefile	21/09/2021 16:14	TREEFILE File	1 KB
 vgcare_Phylo.fasta	23/04/2021 11:32	FASTA File	333 KB
 vgcare_Phylo.fasta· Zone.Identifier	23/04/2021 11:32	IDENTIFIER File	1 KB



5. Go to <https://itol.embl.de/upload.cgi> and upload file "vgcare\_aligned.fasta.treefile"

It can be edited and exported



# Questions and discussion