

Method of the Year 2022:
Long-read sequencing



ONT- Long-read sequencing for virus and applications

Huyha
Phan Tấn Phát <https://doi.org/10.1038/s41592-022-01730-w>

Contents

I. Overview about ONT for Virus

1. ONT basecalling in nutshell
2. Bioinformatic software for long-read
3. Structure variants
4. WGS for Virus
5. Application for ONT

II. Practice

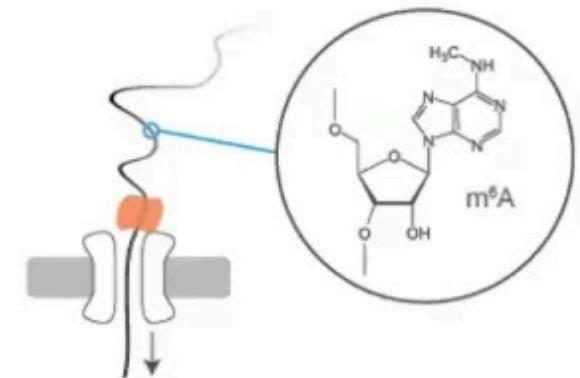
1. Mapping strategy
2. De-novo assembly strategy

Review previous week

- Short read, long read
 - SNV, SNP
- Structure variant
- Contig, scaffold
- Hybird assembly
- Summary

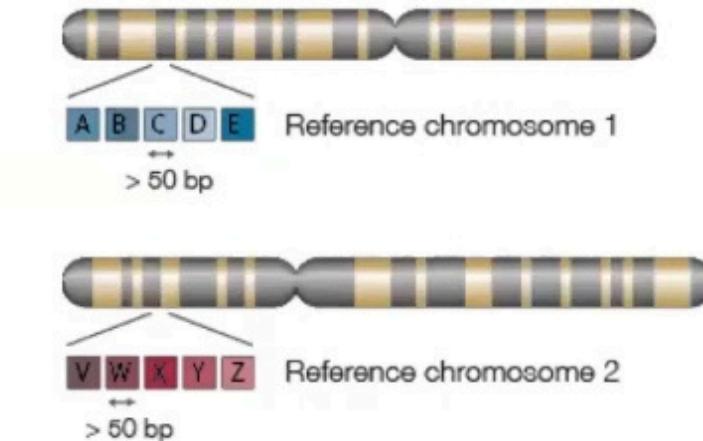
Why nanopore sequencing?

Access a wide range of unique benefits



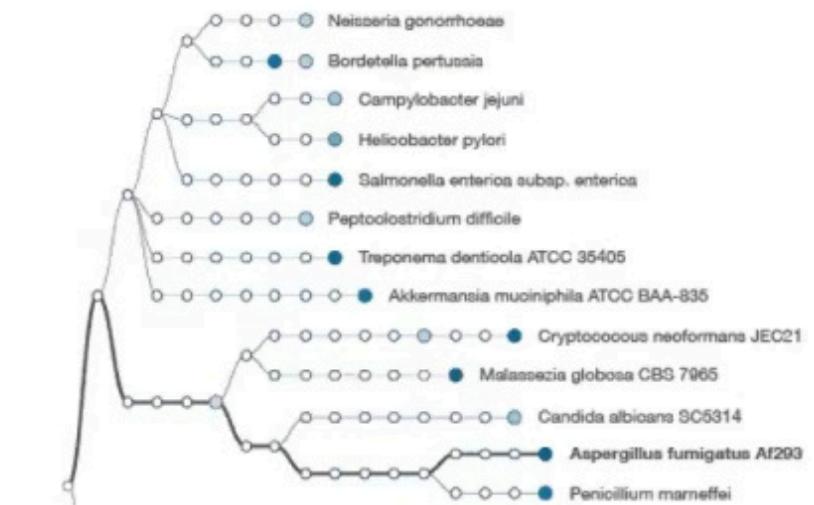
Direct sequencing of DNA or RNA

- Native PCR-free sequencing: processes read lengths presented rather than generating read lengths
- No light-based sequencing steps or surrogate markers required
- Base modification e.g. methylation information is also collected



Read short to ultra-long sequences

- Widest application range possible
- Long reads enable quick and easy elucidation of genomic variations: SVs; repeat regions; phasing; and transcript isoform resolution
- Longest read length: $>4 \text{ Mb}$

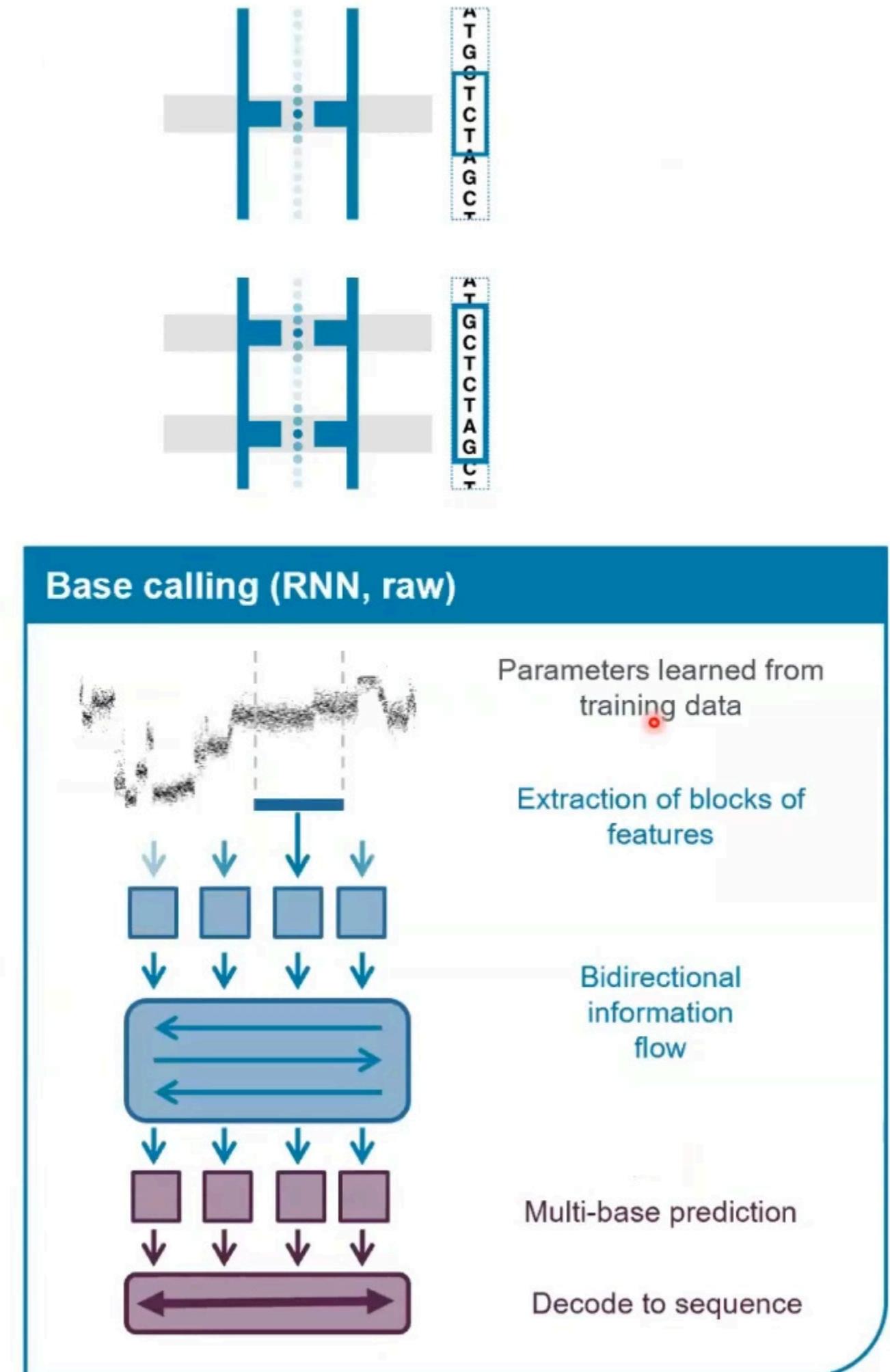
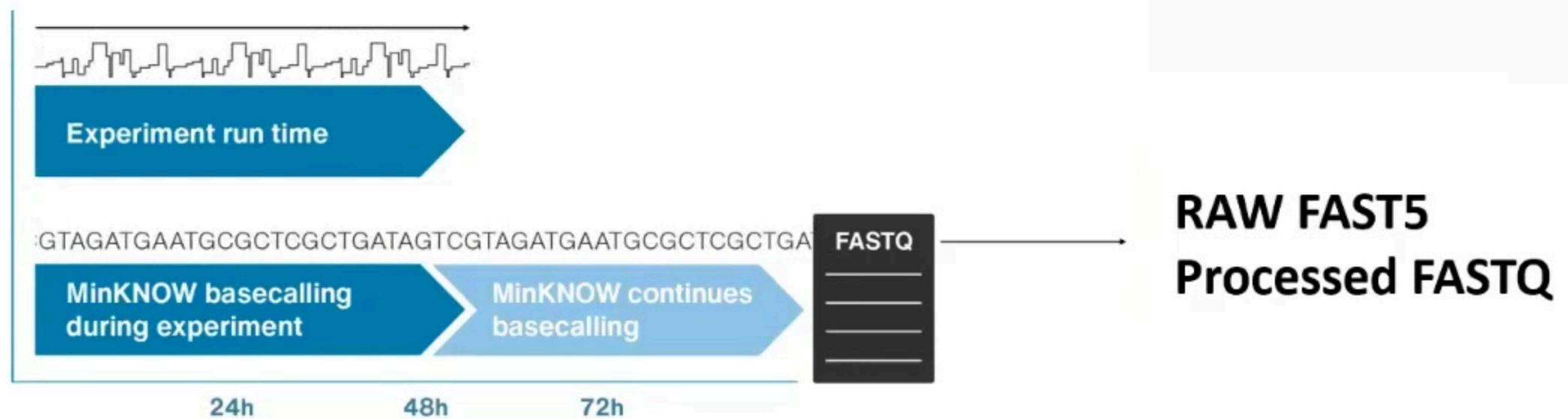


REAL real-time analysis!

- For each flow cell, start and stop as required; no fixed run time
- Additional flow cells can be added whilst a run is in progress (no batching required)
- Adaptive sampling: on-sequencer target enrichment; reject or accept reads on a strand by strand basis

Basecalling in a nutshell

- When sequencing DNA or RNA with nanopores, the changes in current caused by the strand of DNA or RNA as it passes through the pore are recorded by the MinKNOW™ software which runs all of Oxford Nanopore's sequencing devices.
- The processive movement of bases through the pore leads to a continual change in current, known as the “squiggle”.

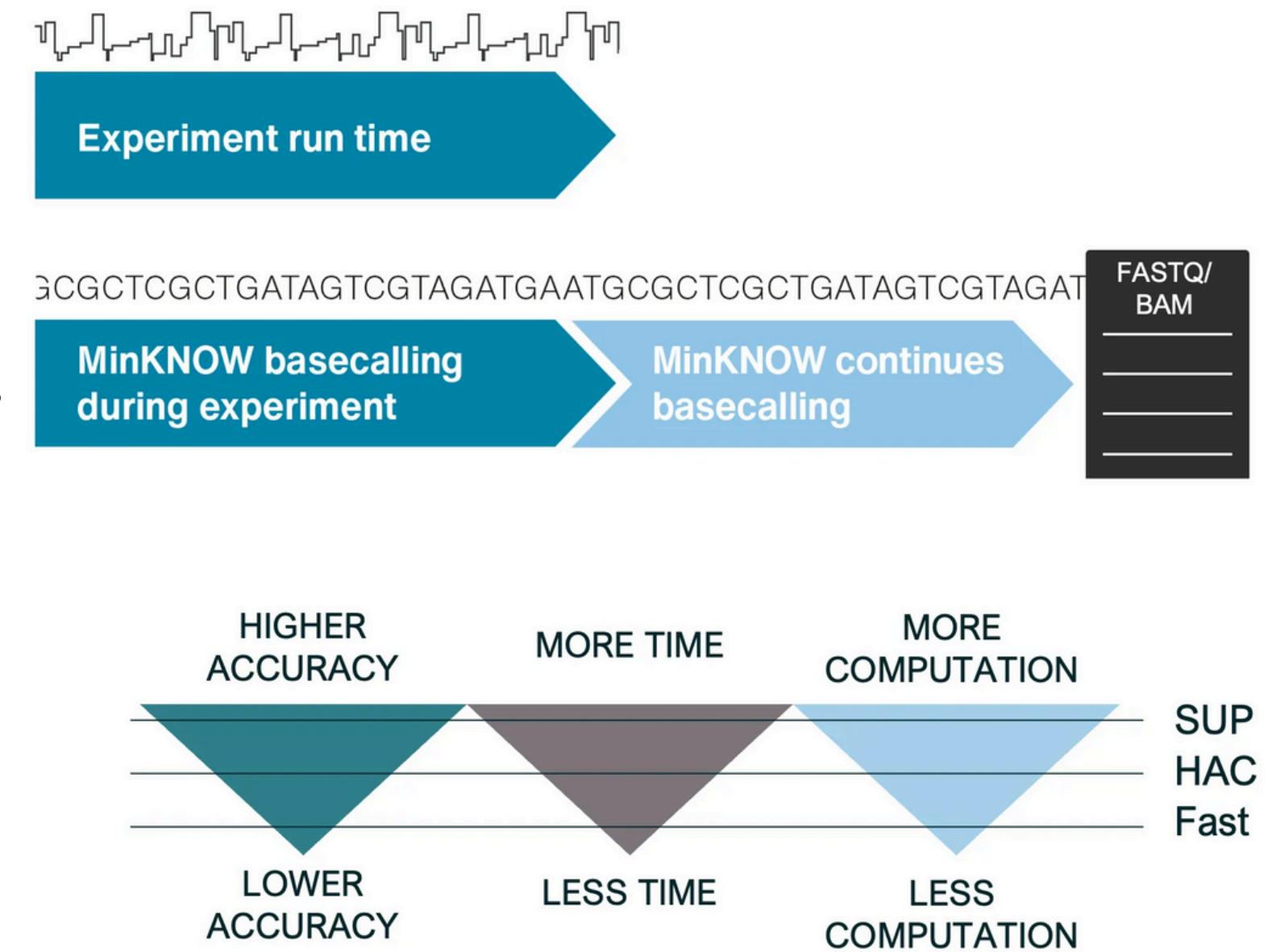


Basecalling mode

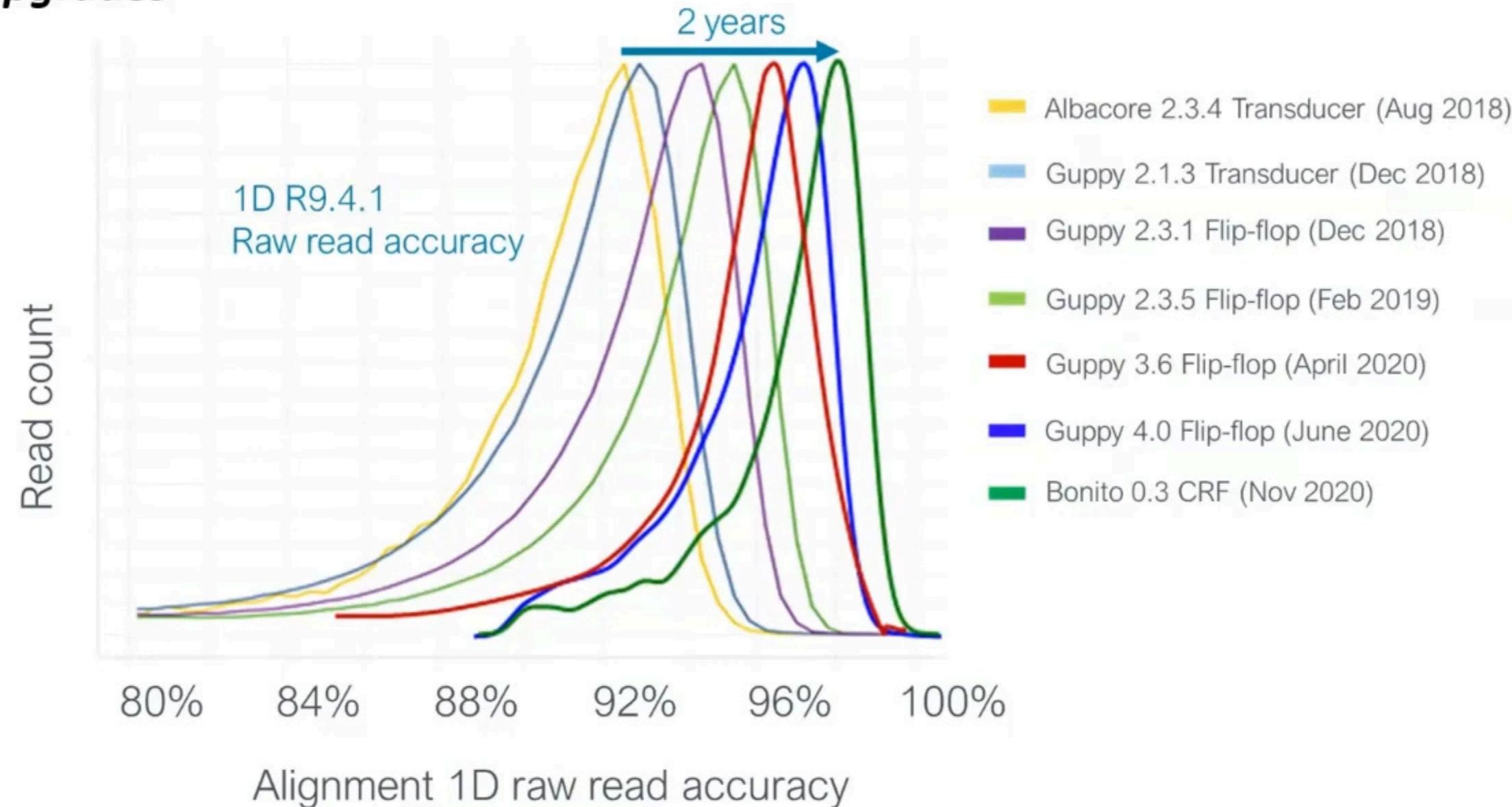
Optimise accuracy according to your requirements by selecting the most suitable basecalling model.

- **Fast basecalling:** fastest, least computationally intense. Compatible with real-time basecalling on all nanopore devices with compute. Recommended for quick, real-time insights on sequencing data when compute resources are limited.
- **High accuracy basecalling (HAC):** highly accurate, intermediate speed and computational requirement. Compatible with real-time basecalling on GridION and PromethION devices with compute. Recommended for high-throughput projects focusing on variant analysis.
- **Super accuracy basecalling (SUP):** the most accurate and computationally intense. Recommended for de novo assembly projects and low-frequency variant analysis (e.g. somatic variation and single-cell applications).
- **Duplex basecalling:** is recommended for hemi-methylation investigation, enabling the methylation signature of each DNA strand to be distinguished.

Nanopore raw reads now achieve 99.75% (Q26) accuracy with the latest Dorado basecalling models (v5) available in GitHub.



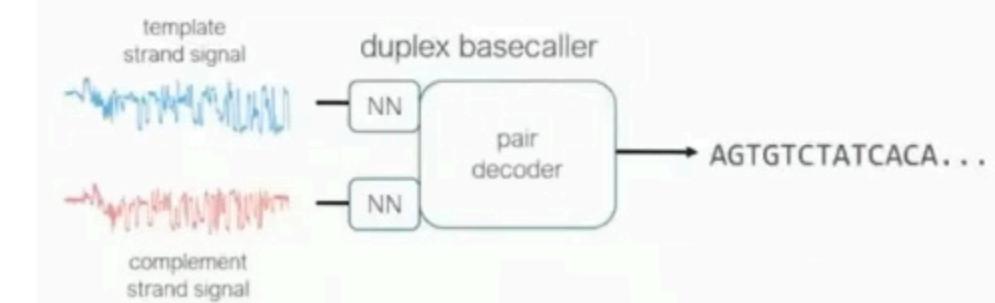
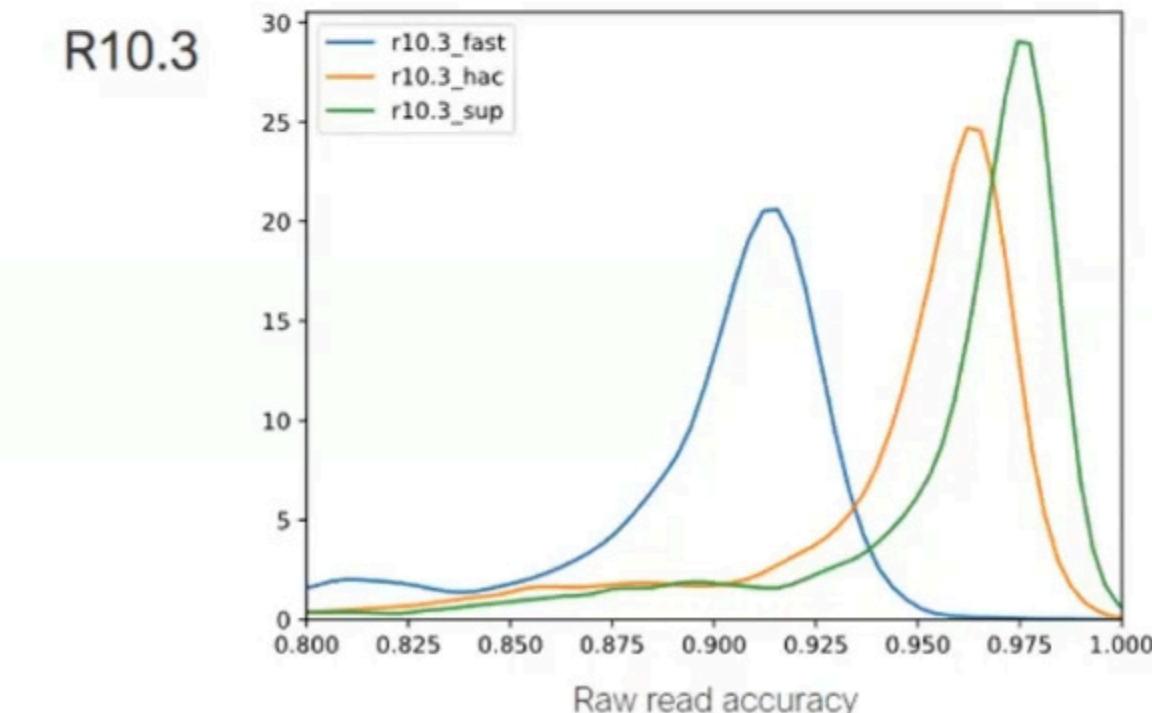
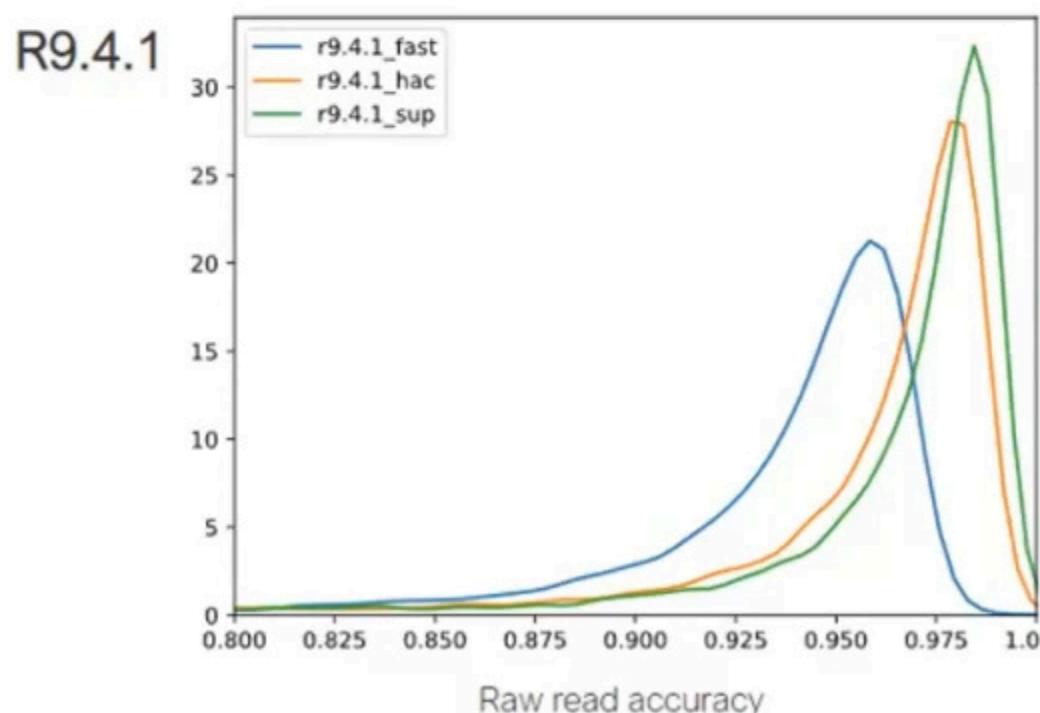
Software upgrades



At the moment :

- Guppy 6.1.7
- Dorado 0.0.3

Flowcell upgrades



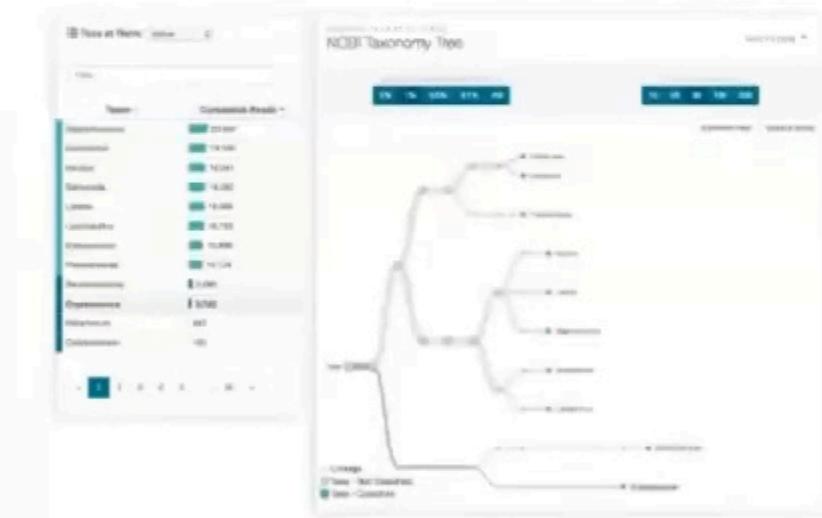
Nanopore sequencing data analysis solutions

A range of bioinformatics options available

Point and click

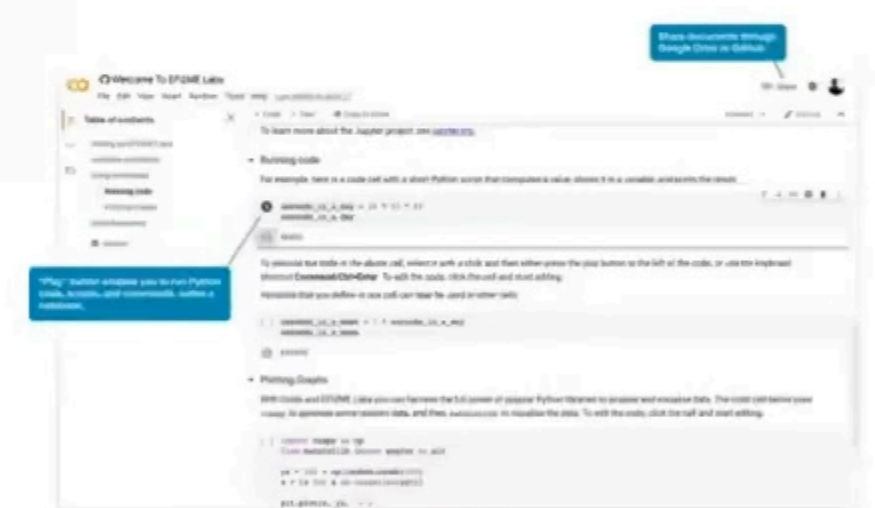


EPI2ME



- Real-time end-to-end analysis
- Wide range of workflows
- Intuitive GUI
- Cloud-based compute

Tutorial



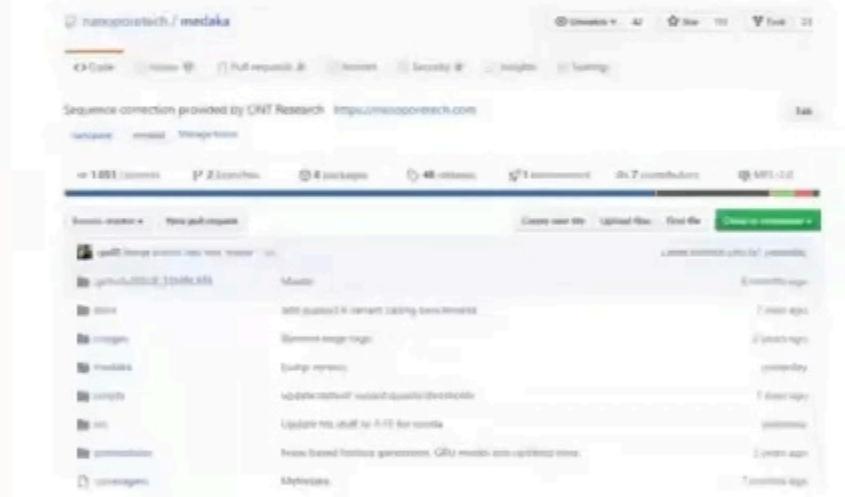
- Tutorial format
- Best-practice analysis methods
- Emphasis on visualisation
- Run on device

Custom



Oxford Nanopore Technologies

Oxford, UK <https://nanoporetech.com/> support@nanoporetech.com Verified



- Open source software
- Customisable pipelines
- State-of-the-art research
- API to ONT instruments
- Integrate with own systems

Popular long-read sequencing software

<https://long-read-tools.org/>

- *De novo* assembly : Flye, Canu, Shasta
- Reads QC : PoreChop, FiltLong, NanoFilt
- Read Aligner : NGLMR, Minimap2, Winnowmap2
- SNP Caller : Clair3, PEPPER, DeepVariant
- CNV/SV Caller: Sniffles2, CuteSV2, NanomonSV, Spectre
- Methylation caller: Remora, F5C, Nanopolish, Tombo...
-

The screenshot shows the homepage of the long-read-tools.org website. At the top right is a search bar with the placeholder "Search tools by full or partial name". Below the search bar is a navigation menu with links: HOME, TABLE, TOOLS, TUTORIALS, BENCHMARKS, STATISTICS, SUBMIT, UPDATES, FAQS, and CONTACT. To the right of the menu are social media icons for Twitter and GitHub. The main content area has a yellow header with the text "We currently track 763+ Tools in 35+ Categories". Below the header are three large buttons: "Browse the table" (with a magnifying glass icon), "View the stats" (with an eye icon), and "Submit a Tool" (with a plus sign icon). The bottom half of the page features a grid of icons representing different tool categories: Oxford Nanopore, PacBio, Alignment, Error Correction and Polishing, and Base Modification Detection. Each category has a corresponding icon of a wrench and screwdriver.

HOME TABLE TOOLS TUTORIALS BENCHMARKS STATISTICS SUBMIT UPDATES FAQS CONTACT

Search tools by full or partial name

We currently track 763+ Tools in 35+ Categories

Browse the table

View the stats

Submit a Tool

Oxford Nanopore

PacBio

Alignment

Error Correction and Polishing

Base Modification Detection

prev next

Structural variants

- Structural variants (SVs) are typically defined as genomic variants larger than 50bps (e.g. deletions, duplications, inversions).
- Largely unexplored feature in most genomes

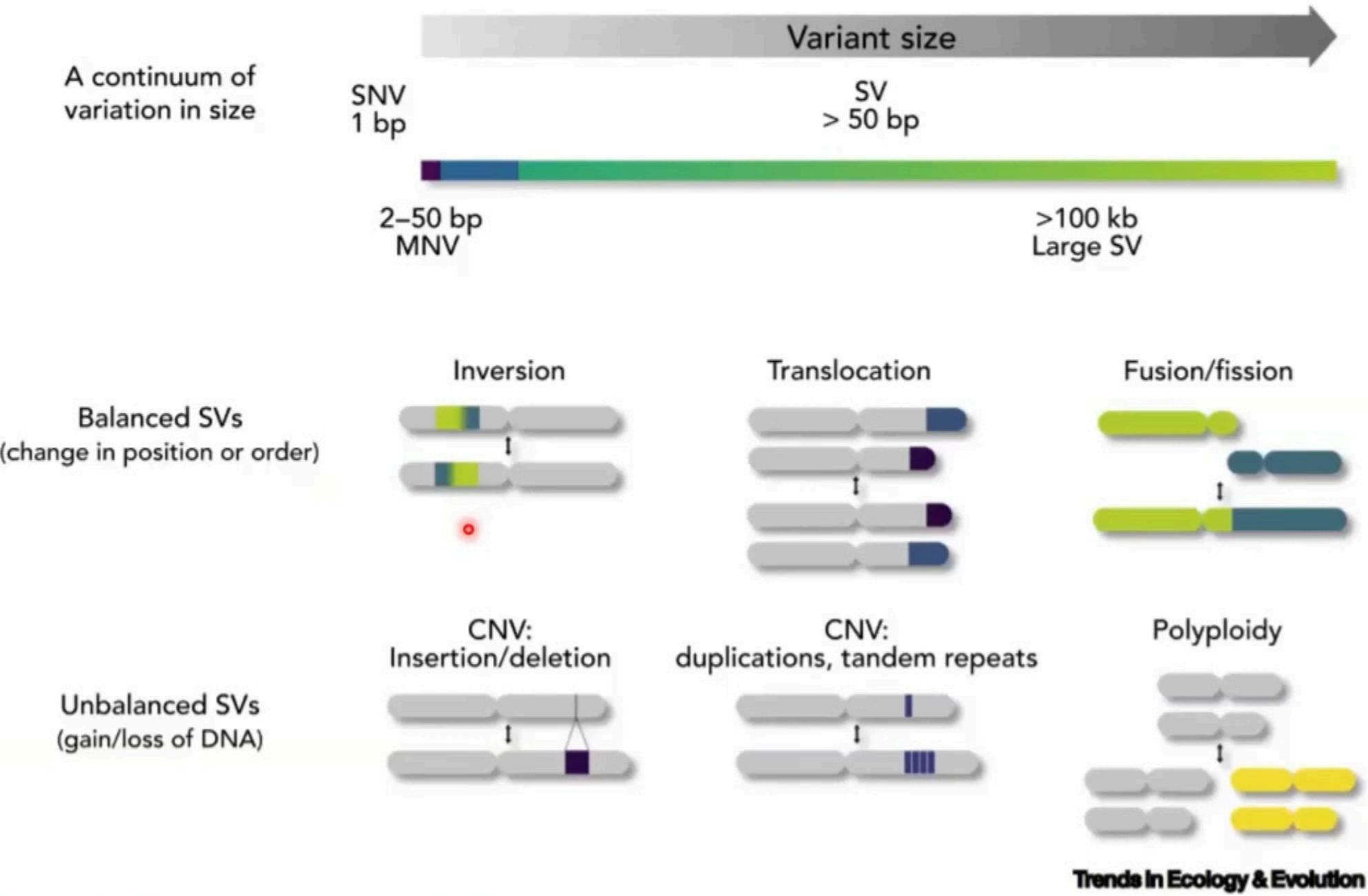
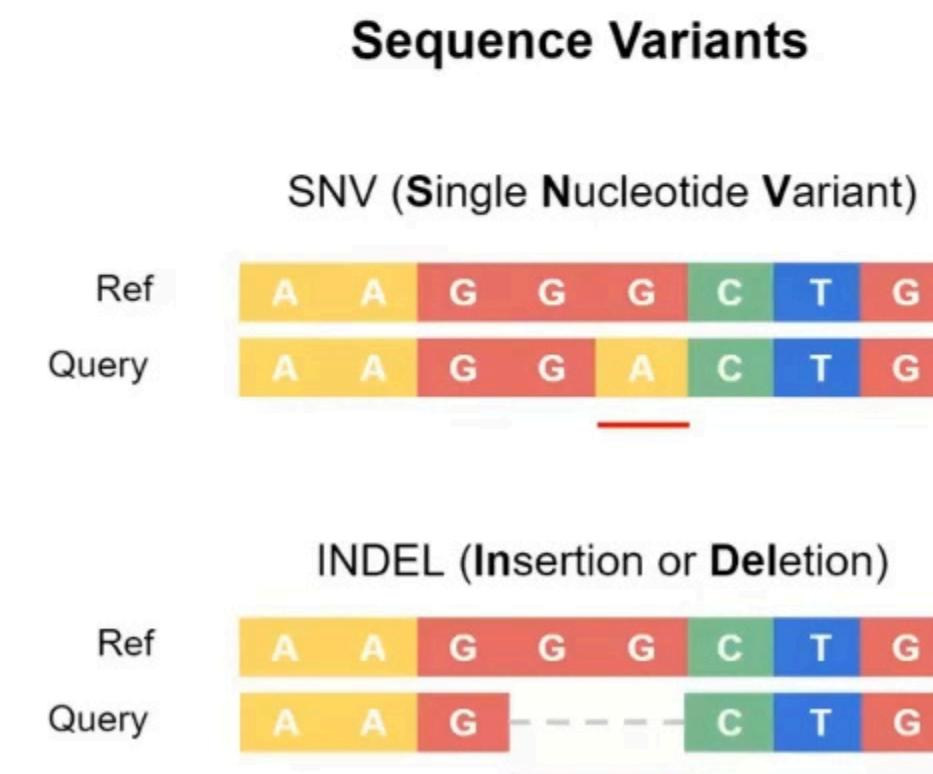


Figure 1. Diversity of Structural Variants. Genetic variants vary in size from a single nucleotide to hundreds-of-Mb-long structural variants (SVs). SVs are classified according to how they change the genome sequence. Balanced SVs change the position and/or order of genomic areas. Unbalanced SVs involve a gain or loss of sequence. Note that transposable elements can cause translocations, indels, and/or duplications. Abbreviations: CNV, copy number variant; MNV, multiple nucleotide variant; SNV, single nucleotide variant.

1000 Genomes project

New Results

Follow this preprint

Previous

Next

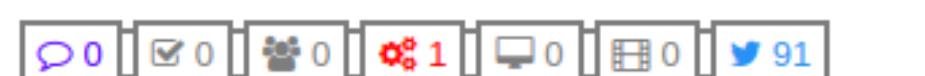
Long-read sequencing and structural variant characterization in 1,019 samples from the 1000 Genomes Project

Posted April 20, 2024.

Siegfried Schloissnig, Samarendra Pani, Bernardo Rodriguez-Martin, Jana Ebler, Carsten Hain, Vasiliki Tsapalou, Arda Söylev, Patrick Hüther, Hufsah Ashraf, Timofey Prodanov, Mila Asparuhova, Sarah I Jan O. Korbel

doi: <https://doi.org/10.1101/2024.04.18.590093>

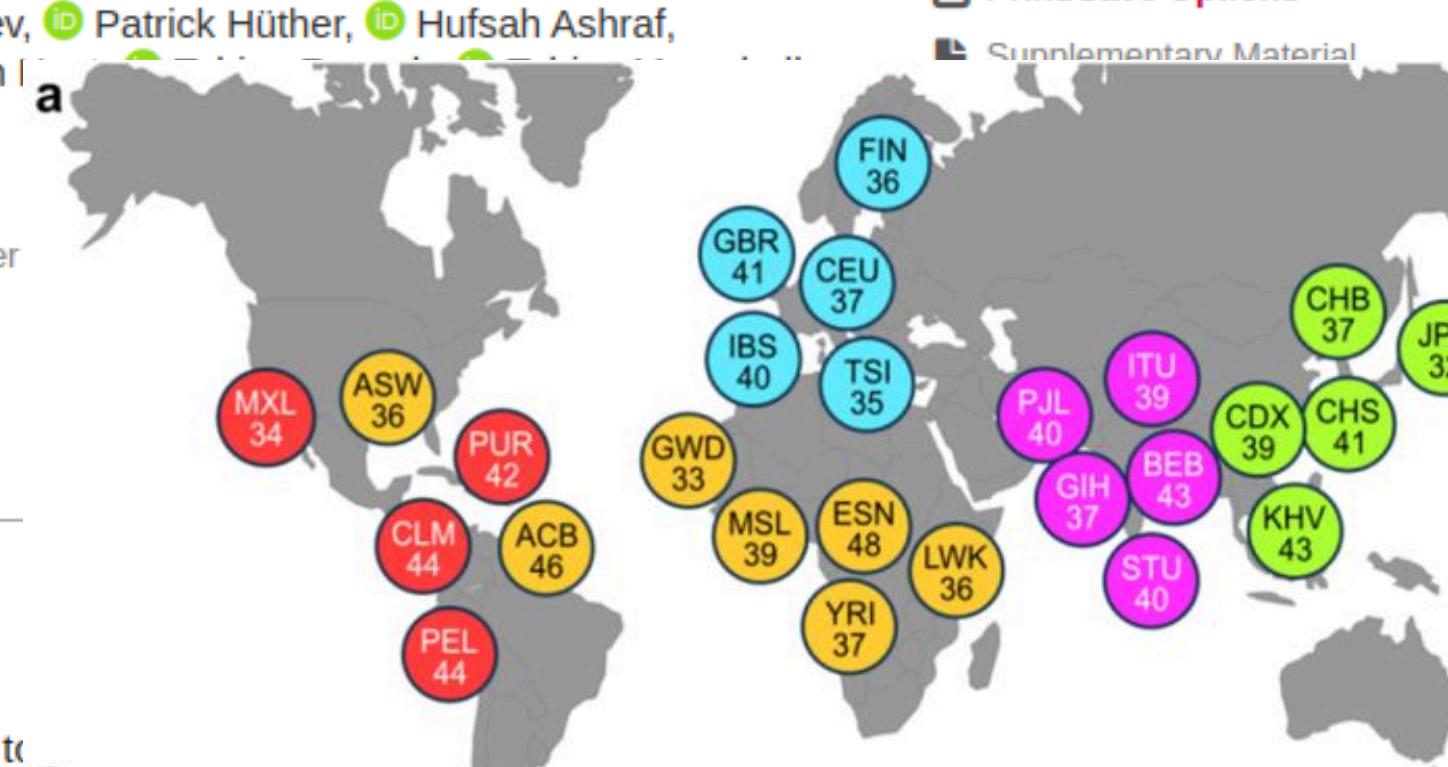
This article is a preprint and has not been certified by peer review.



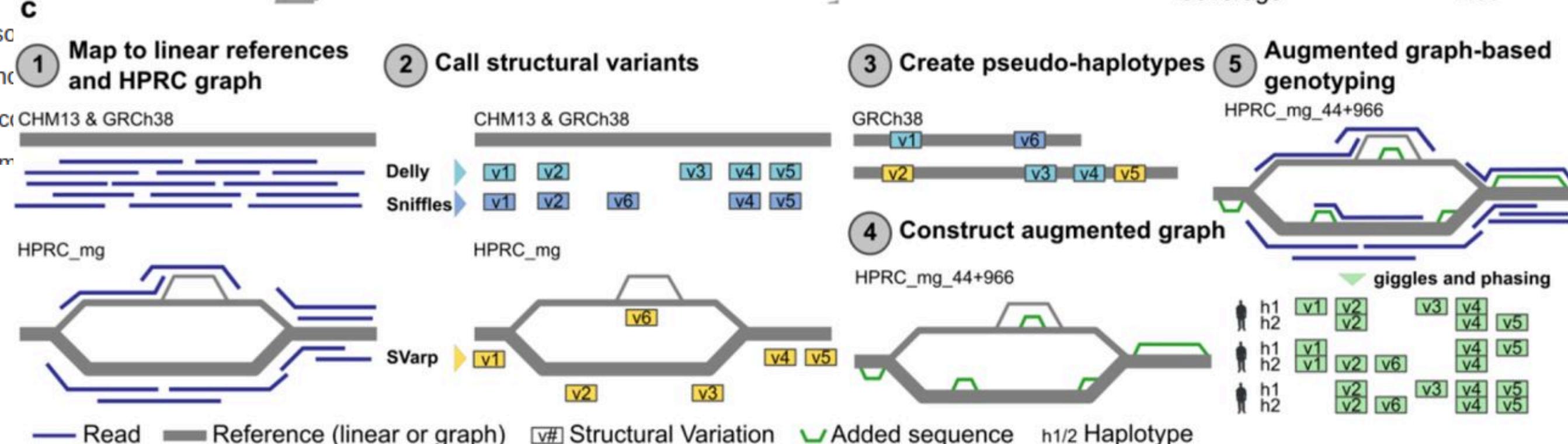
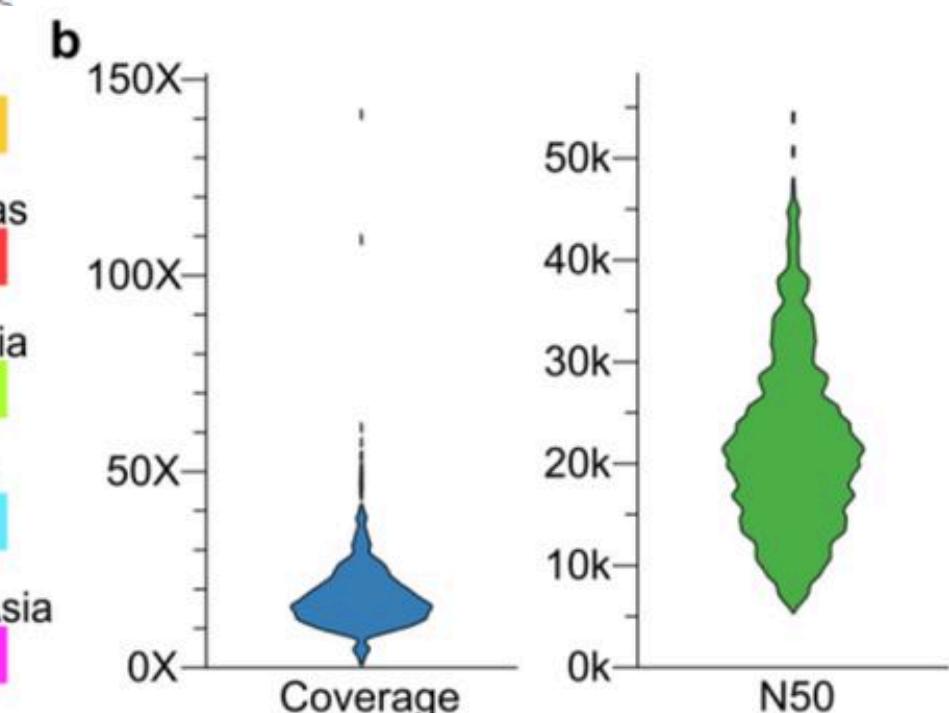
Abstract Full Text Info/History Metrics

Abstract

Structural variants (SVs) contribute significantly to



Download PDF Email
 Print/Save Options Share
 Supplementary Material Citation Tools



Icelandic structure variation profiling

Long-read sequencing of 3,622 Icelanders provides insight into the role of structural variants in human diseases and other traits

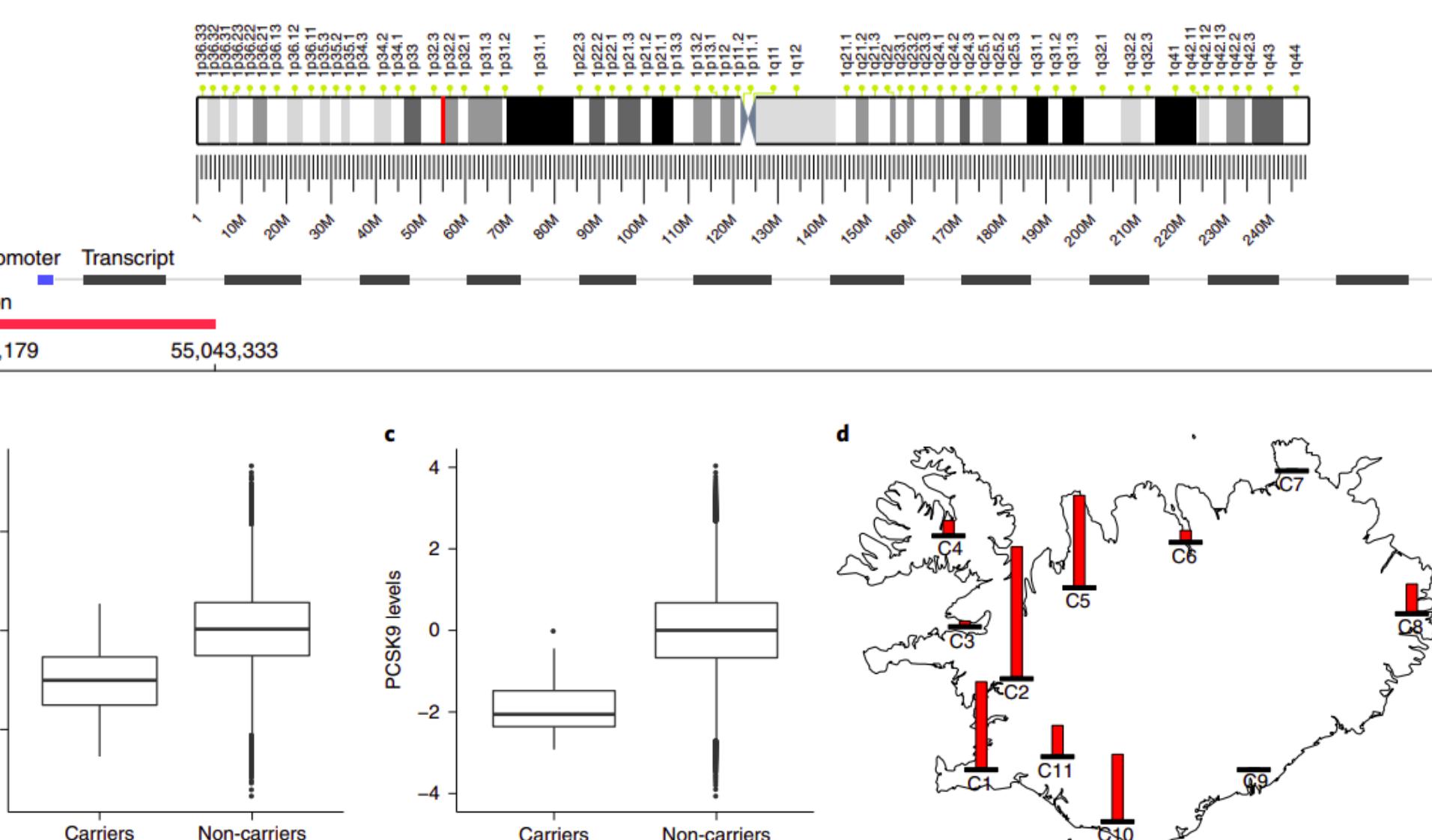
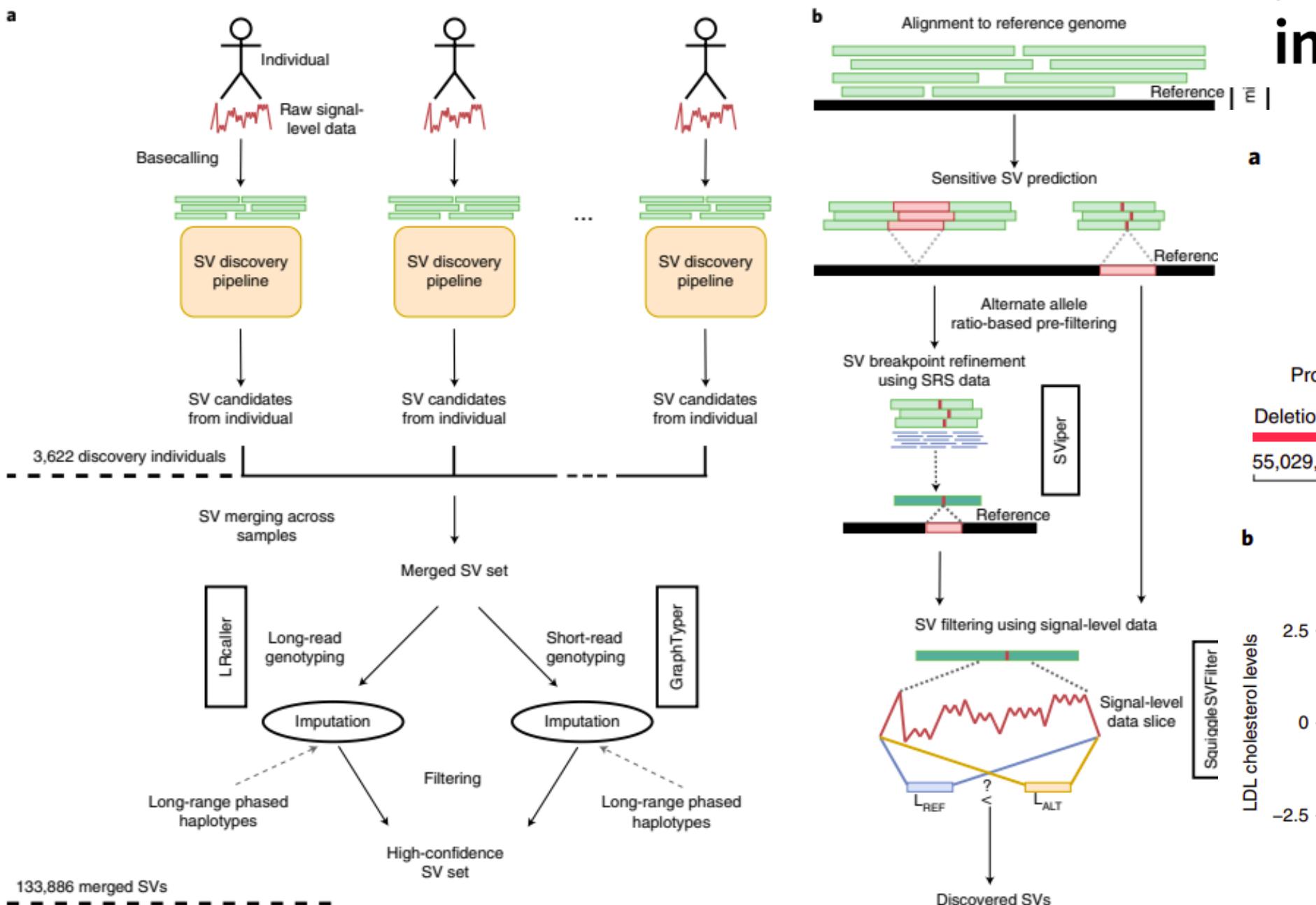
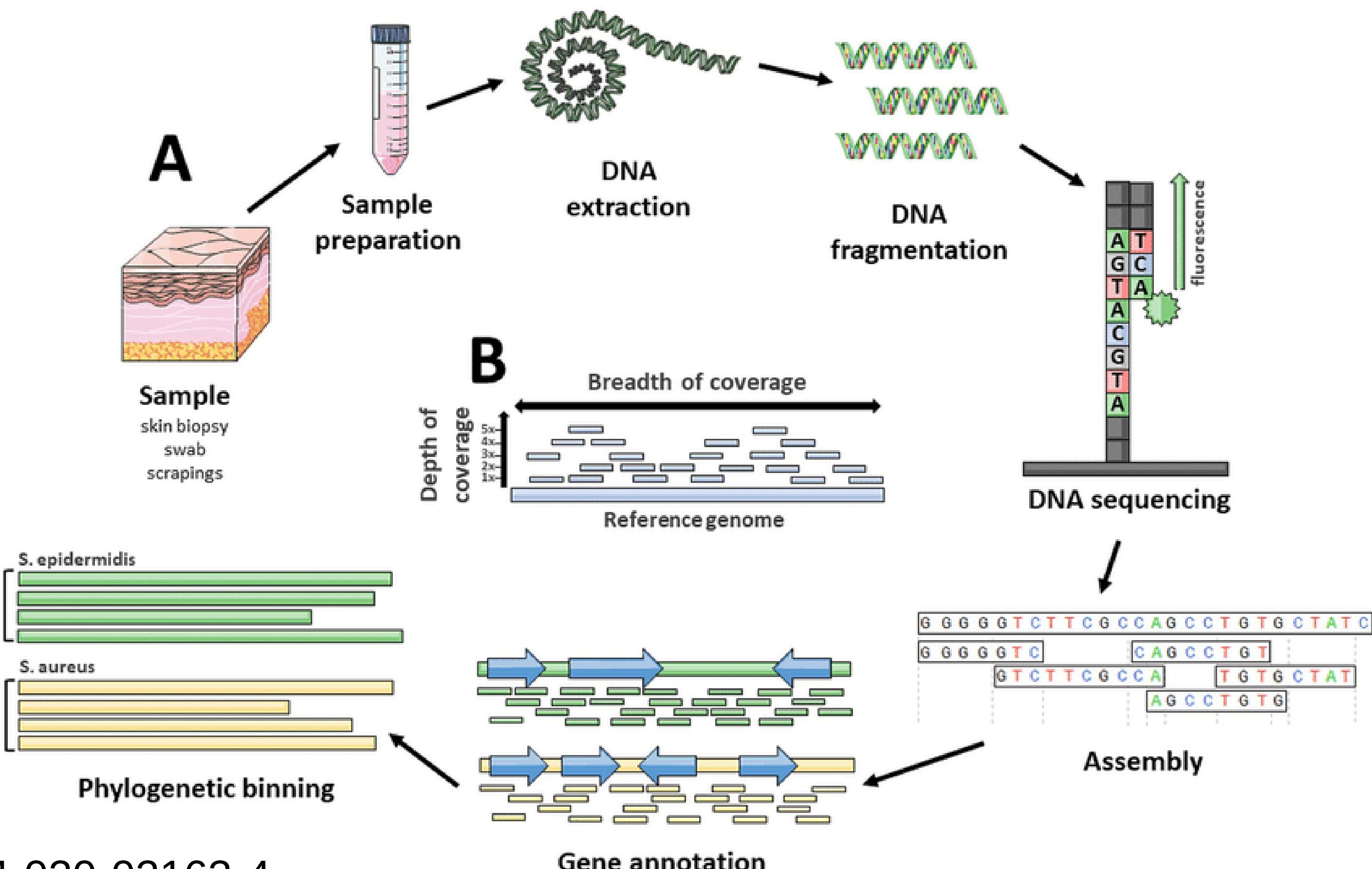


Fig. 3 | Large deletion in PCSK9 associated with lower LDL cholesterol levels. a, Chromosome 1 ideogram with cytobands, highlighting the SV site in red. A 14,154-bp (allele frequency, 0.037%) deletion removes the promoter and the first coding exon of PCSK9 at 55,029,179–55,043,333 bp (GRCh38). **b**, LDL cholesterol levels in carriers and non-carriers ($n=72$ carriers, $n=96,840$ non-carriers, effect = -1.31 s.d., $P=7.0\times 10^{-20}$, two-sided linear regression). **c**, PCSK9 protein levels in carriers and non-carriers using SOMAscan ($n=20$ carriers, $n=38,385$ non-carriers, effect = -1.99 s.d., $P=3.1\times 10^{-13}$, two-sided linear regression). **d**, Geographical distribution of the PCSK9 deletion in 166,281 chip-typed Icelanders. Each bar shows the allele frequency of the variant.

Whole genome sequencing and shotgun sequencing?



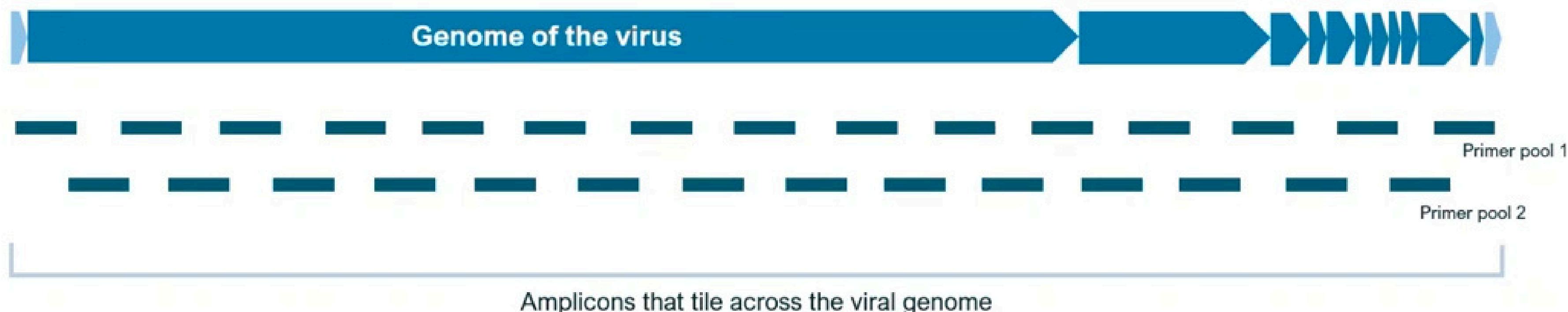
PCR tiling of virus genomes

Viral whole-genome sequencing (WGS) has become a critical tool used to guide public health

PCR can provide target enrichment and amplification in a single step. Compared to other methods, this approach is cost-effective, readily accessible and efficient

To generate whole viral genome coverage, a tiling amplicon scheme is commonly used¹

Both short and long amplicons can be sequenced using nanopore technology, further simplifying whole viral genome recovery²

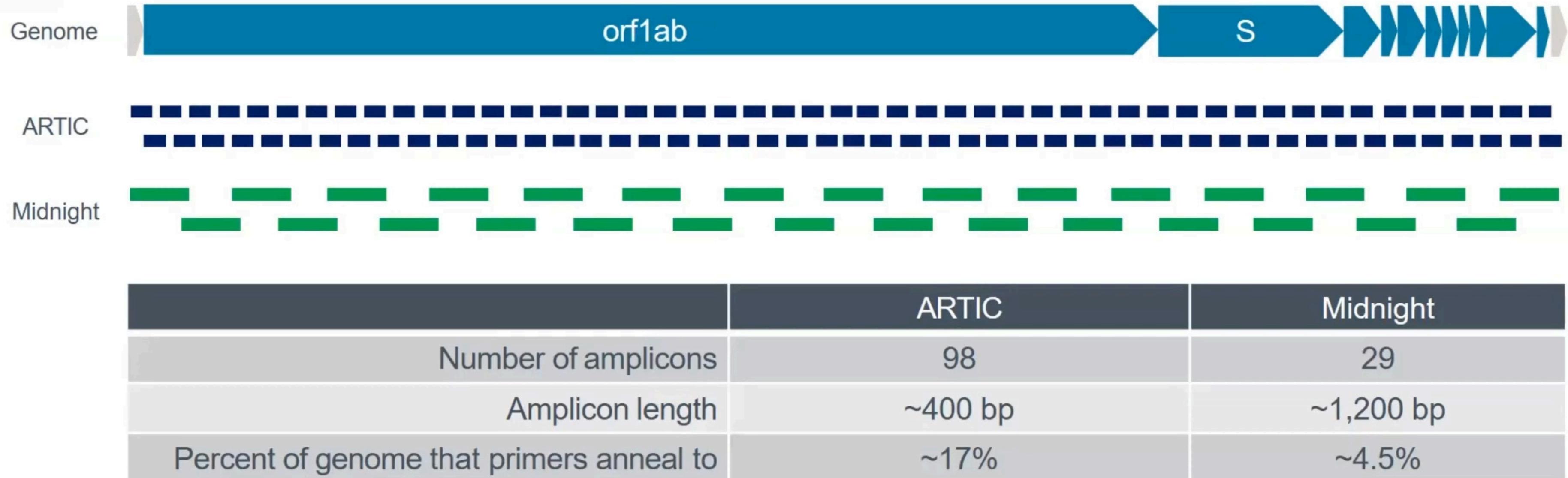


1. Quick, Joshua, et al. "Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples." *Nature protocols* 12.6 (2017): 1261-1276.

2. Freed, Nikki E., et al. Rapid and inexpensive whole-genome sequencing of SARS-CoV-2 using 1200 bp tiled amplicons and Oxford Nanopore Rapid Barcoding. *Biology Methods and Protocols* 5.1 (2020): bpea014.

SARS-CoV-2 genome sequencing method

PCR-based target enrichment — producing tiled amplicons that span the viral genome



Quick, Josh. NCoV-2019 Sequencing Protocol. Protocols.io. (2020) <https://www.protocols.io/view/nkov-2019-sequencing-protocol-bbmuik6w>

Freed, Nikki E., et al. Rapid and inexpensive whole-genome sequencing of SARS-CoV-2 using 1200 bp tiled amplicons and Oxford Nanopore Rapid Barcoding. Biology Methods and Protocols 5.1 (2020): bpaa014

Article | Open access | Published: 26 June 2023

Universal whole-genome Oxford nanopore sequencing of SARS-CoV-2 using tiled amplicons

Ruslan Kalendar , Ulykbek Kairov, Daniyar Karabayev, Akbota Aitkulova, Nuray Tynyshtykbayeva, Asset Daniyarov, Zhenis Otarbay, Saule Rakhimova, Ainur Akilzhanova & Dos Sarbassov

[Scientific Reports](#) 13, Article number: 10334 (2023) | [Cite this article](#)

3029 Accesses | 1 Citations | 4 Altmetric | [Metrics](#)

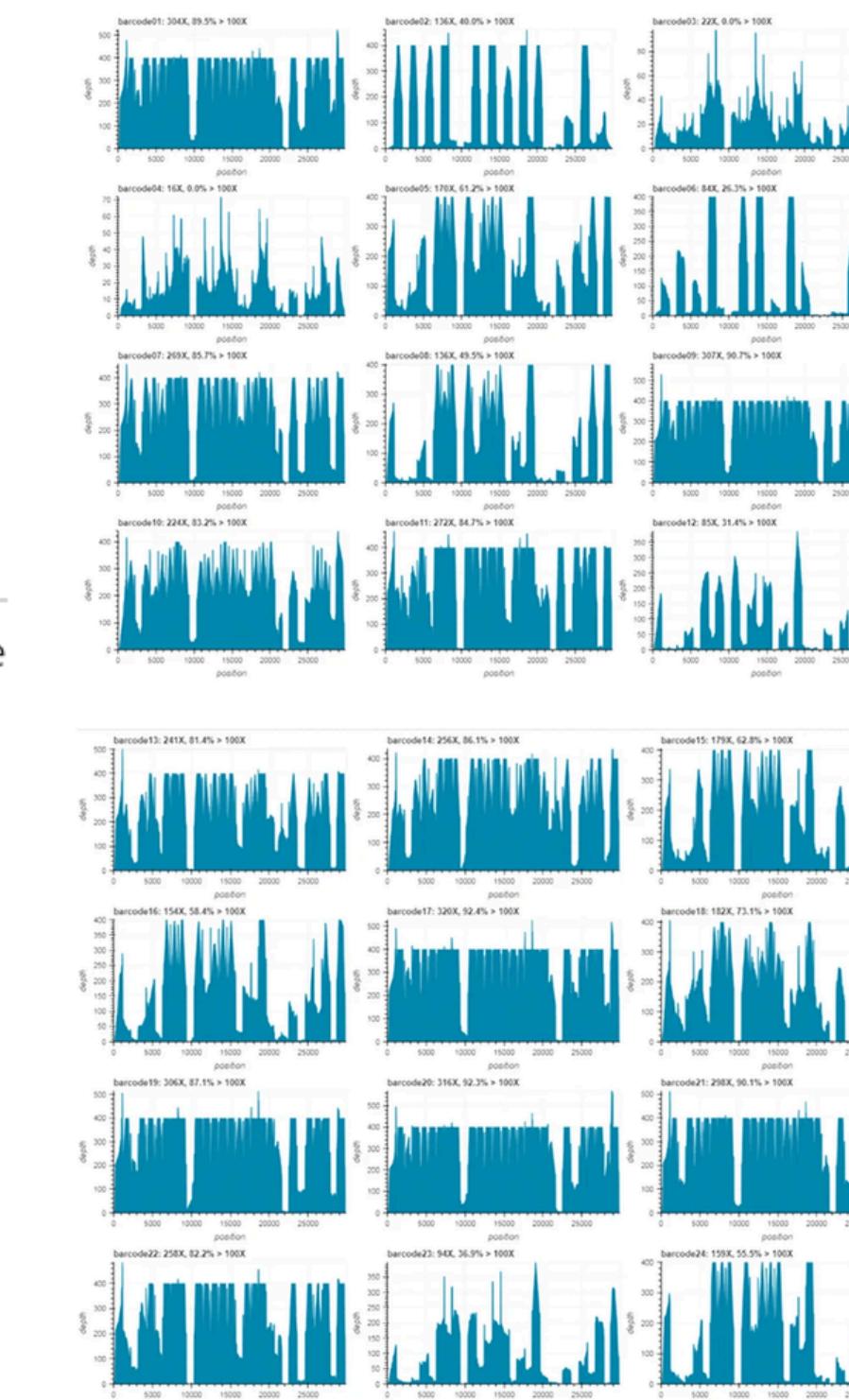
Abstract

We developed a comprehensive multiplexed set of primers adapted for the Oxford Nanopore Rapid Barcoding library kit that allows universal SARS-CoV-2 genome sequencing. This primer set is designed to set up any variants of the primers pool for whole-genome sequencing of SARS-CoV-2 using single- or double-tiled amplicons from 1.2 to 4.8 kb with the Oxford Nanopore. This multiplexed set of primers is also applicable for tasks like targeted SARS-CoV-2 genome sequencing. We proposed here an optimized protocol to synthesize cDNA using Maxima H Minus Reverse Transcriptase with a set of SARS-CoV-2 specific primers, which has high yields of cDNA template for RNA and is capable of long-length cDNA synthesis from a wide range of RNA amounts and quality. The proposed protocol allows whole-genome sequencing of the SARS-CoV-2 virus with tiled amplicons up to 4.8 kb on low-titer virus samples and even where RNA degradation has occurred. This protocol reduces the time and cost from RNA to genome sequence compared to the Midnight multiplex PCR method for SARS-CoV-2 genome sequencing using the Oxford Nanopore.

Midnight workflow • ~7 hrs.



LunaScript RT
(random hexamer and oligo-dT primers)

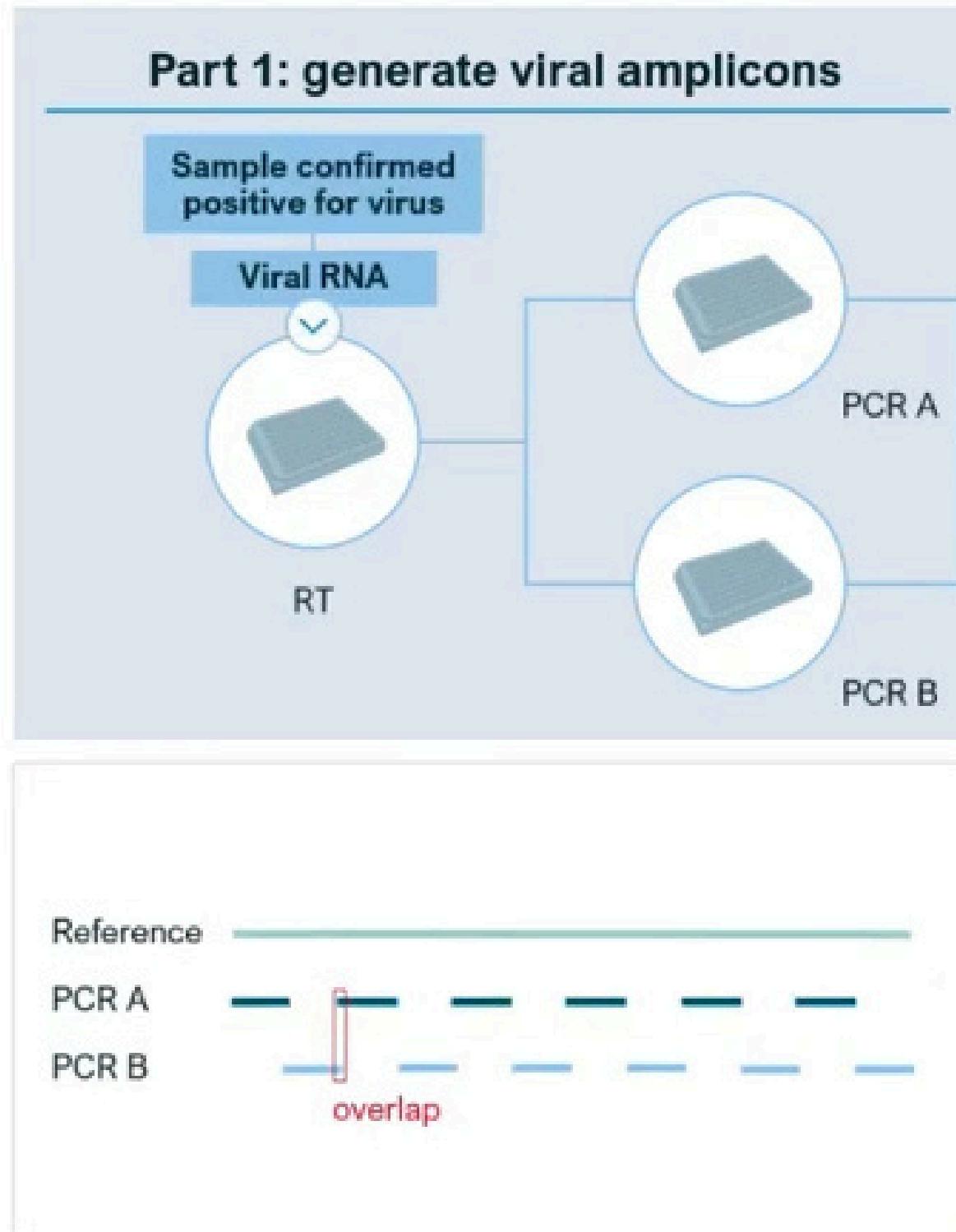


Maxima H Minus RT
(SARS-CoV-2 specific primers set)



Overview of nanopore sequencing for viral whole genome sequencing

From RNA input to viral genome characterization

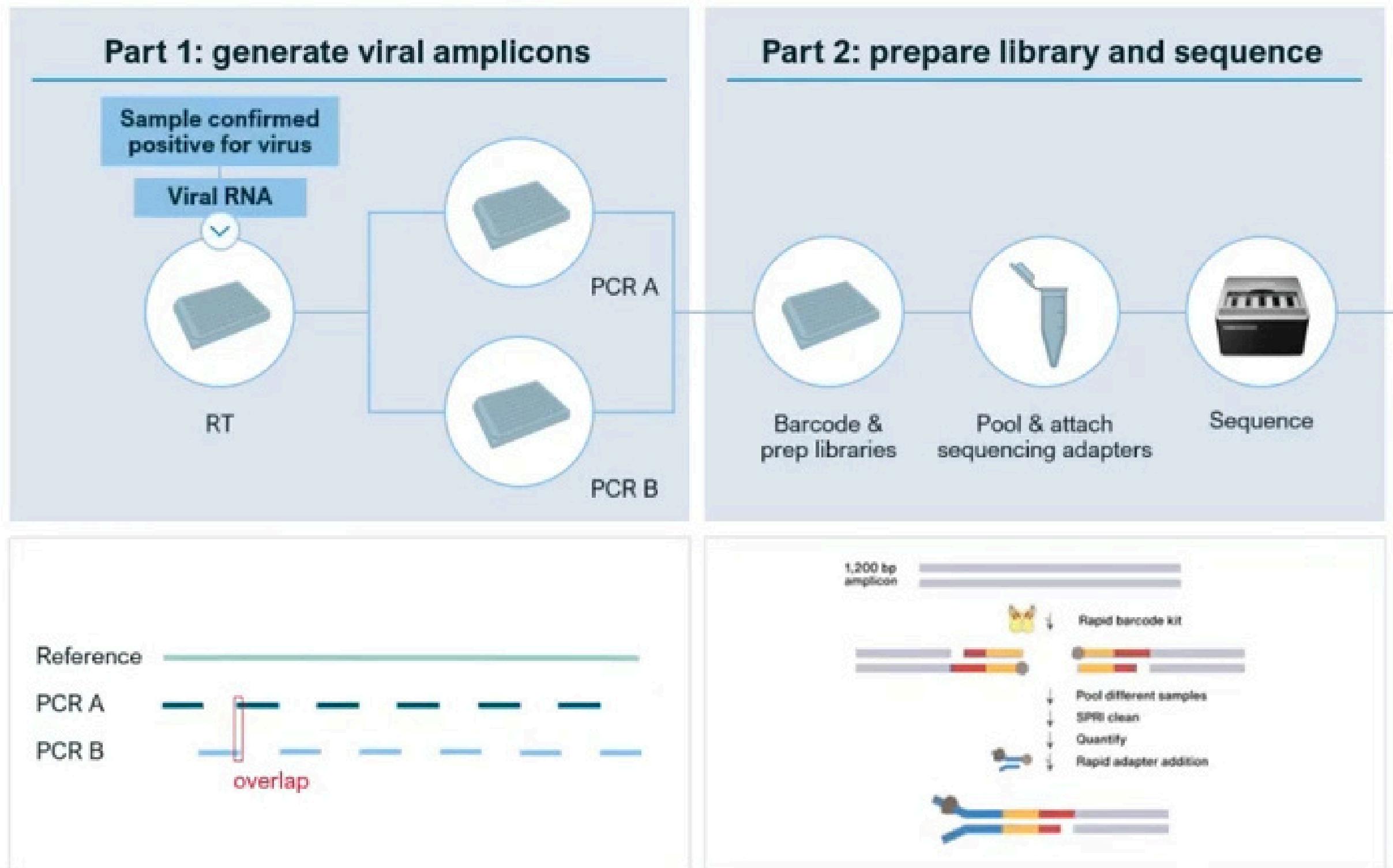


© 2024 Oxford Nanopore Technologies plc.
Oxford Nanopore Technologies products are not intended for use for health assessment or
to diagnose, treat, mitigate, cure, or prevent any disease or condition.

<https://www.youtube.com/watch?v=TT8ZnzKJka8>

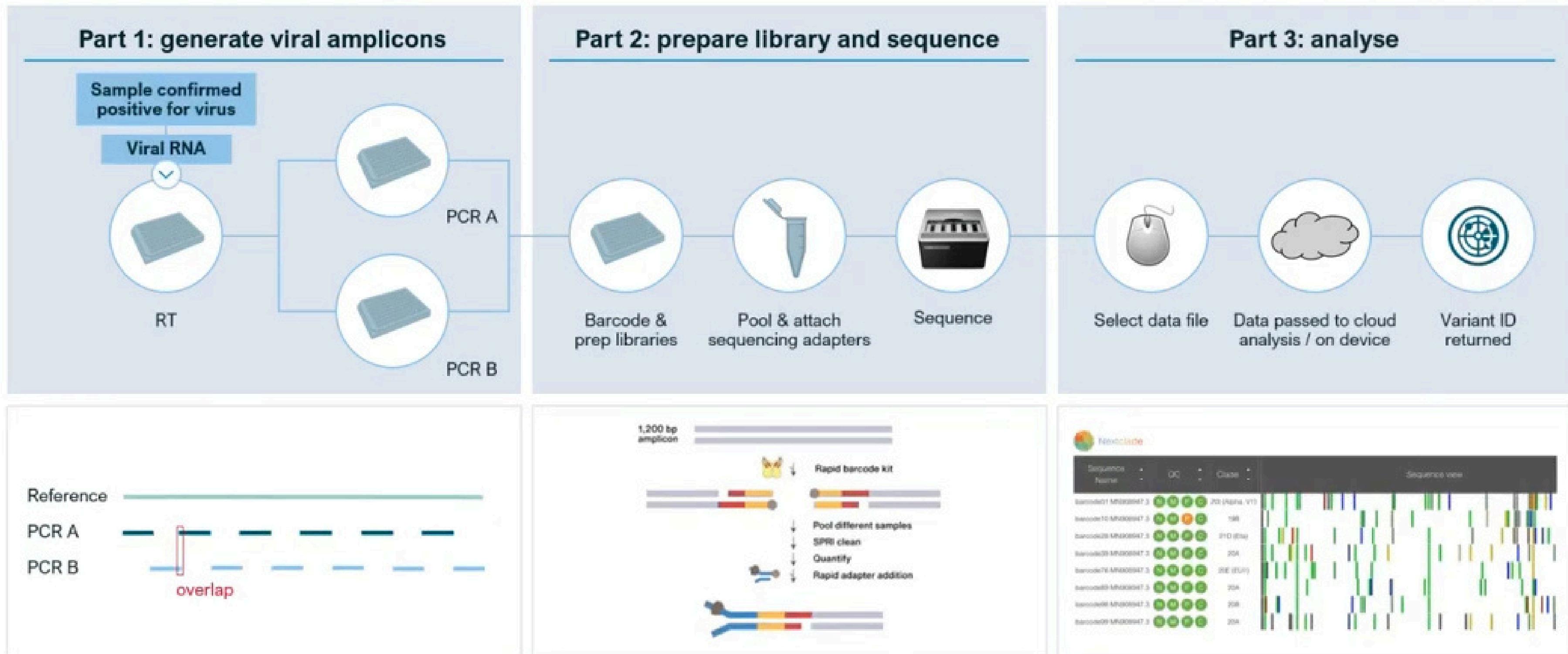
Overview of nanopore sequencing for viral whole genome sequencing

From RNA input to viral genome characterization

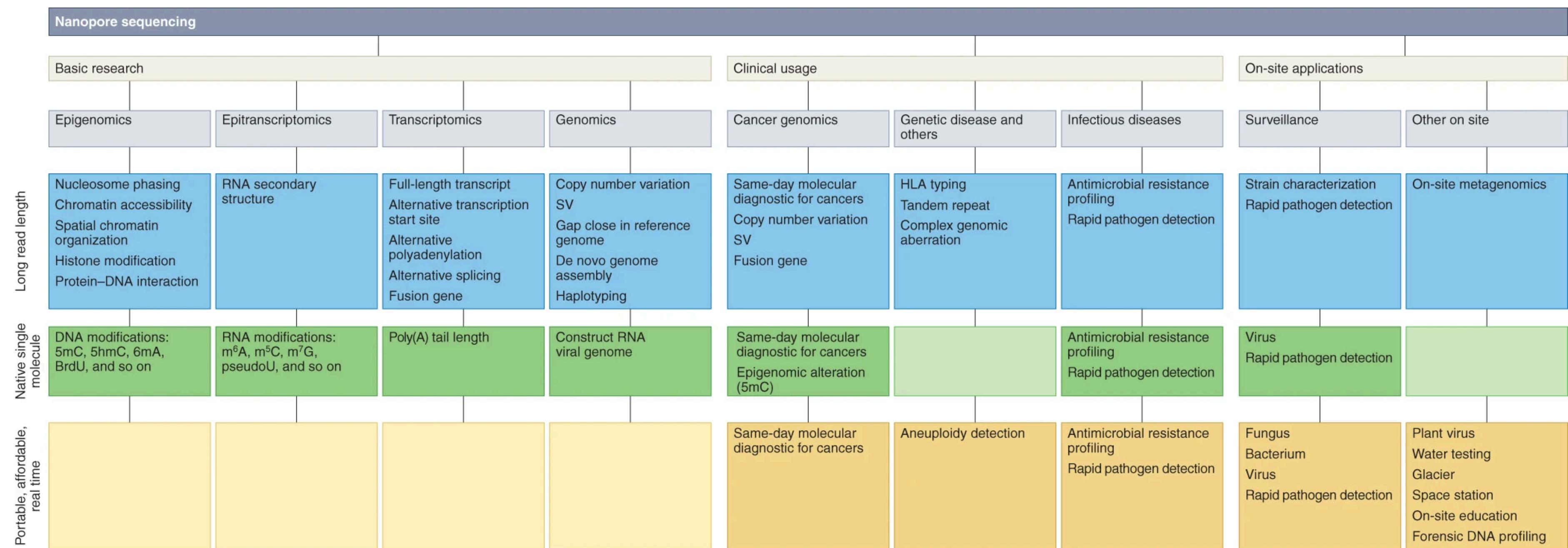


Overview of nanopore sequencing for viral whole genome sequencing

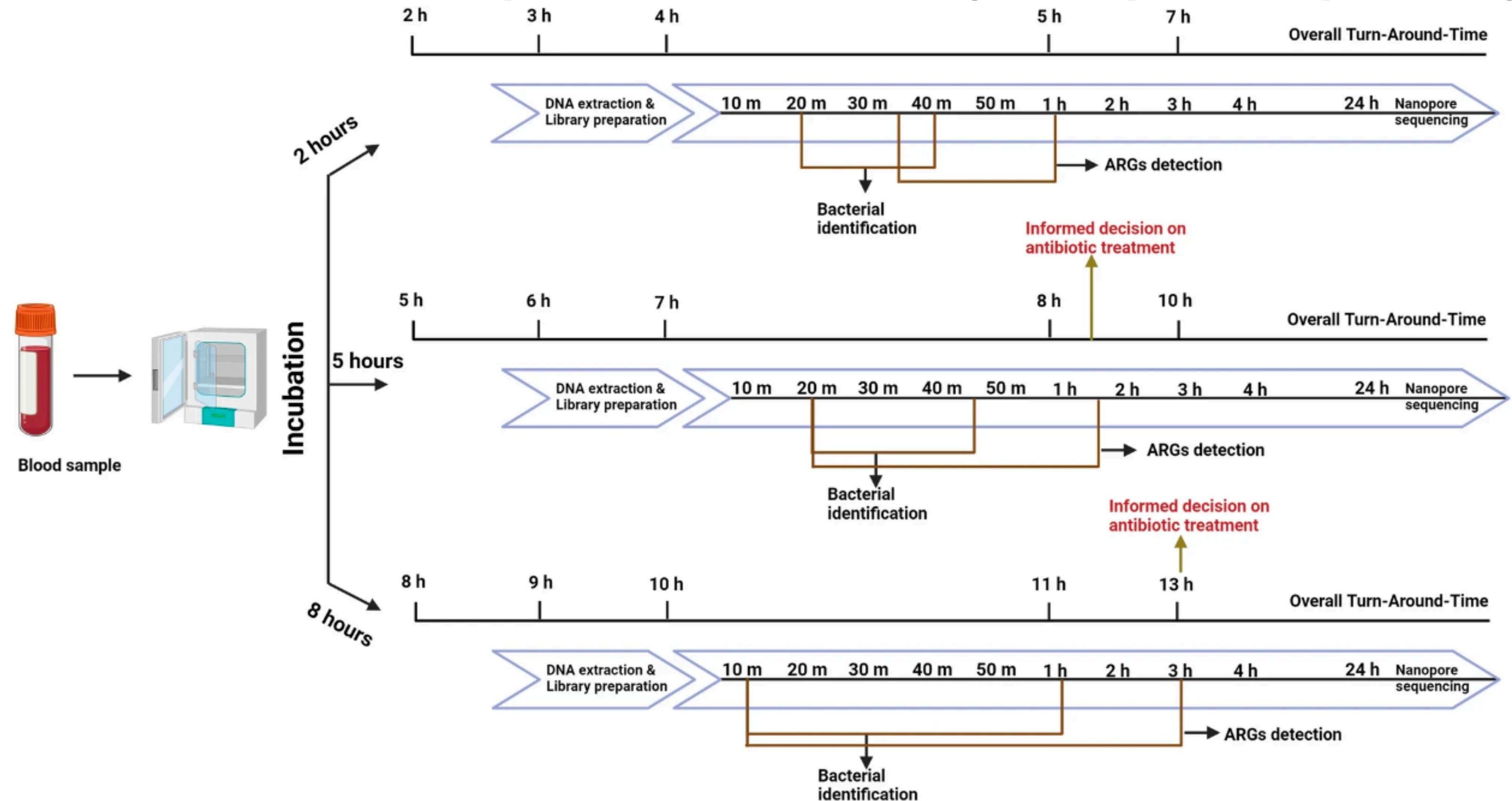
From RNA input to viral genome characterization



Application

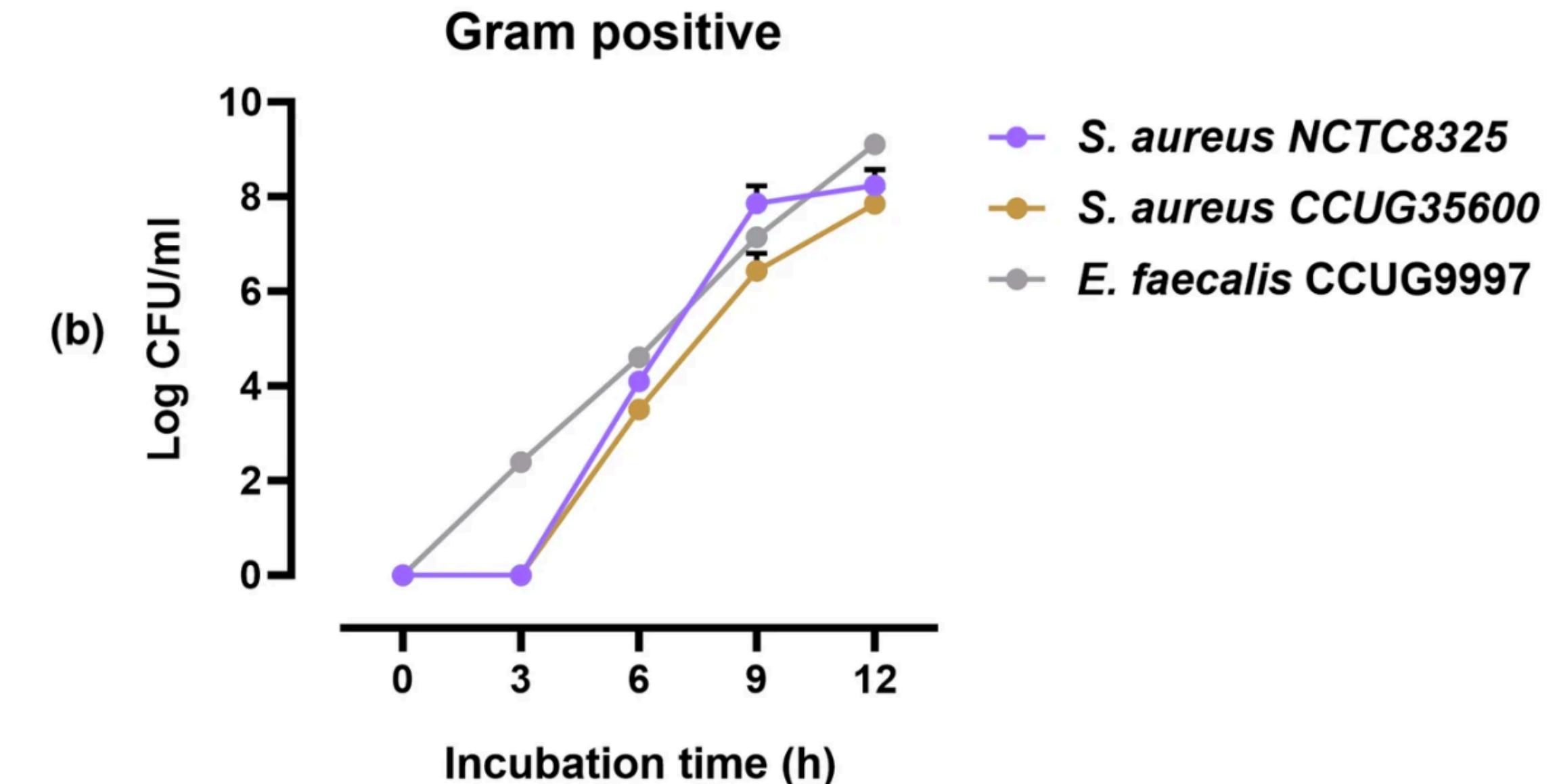
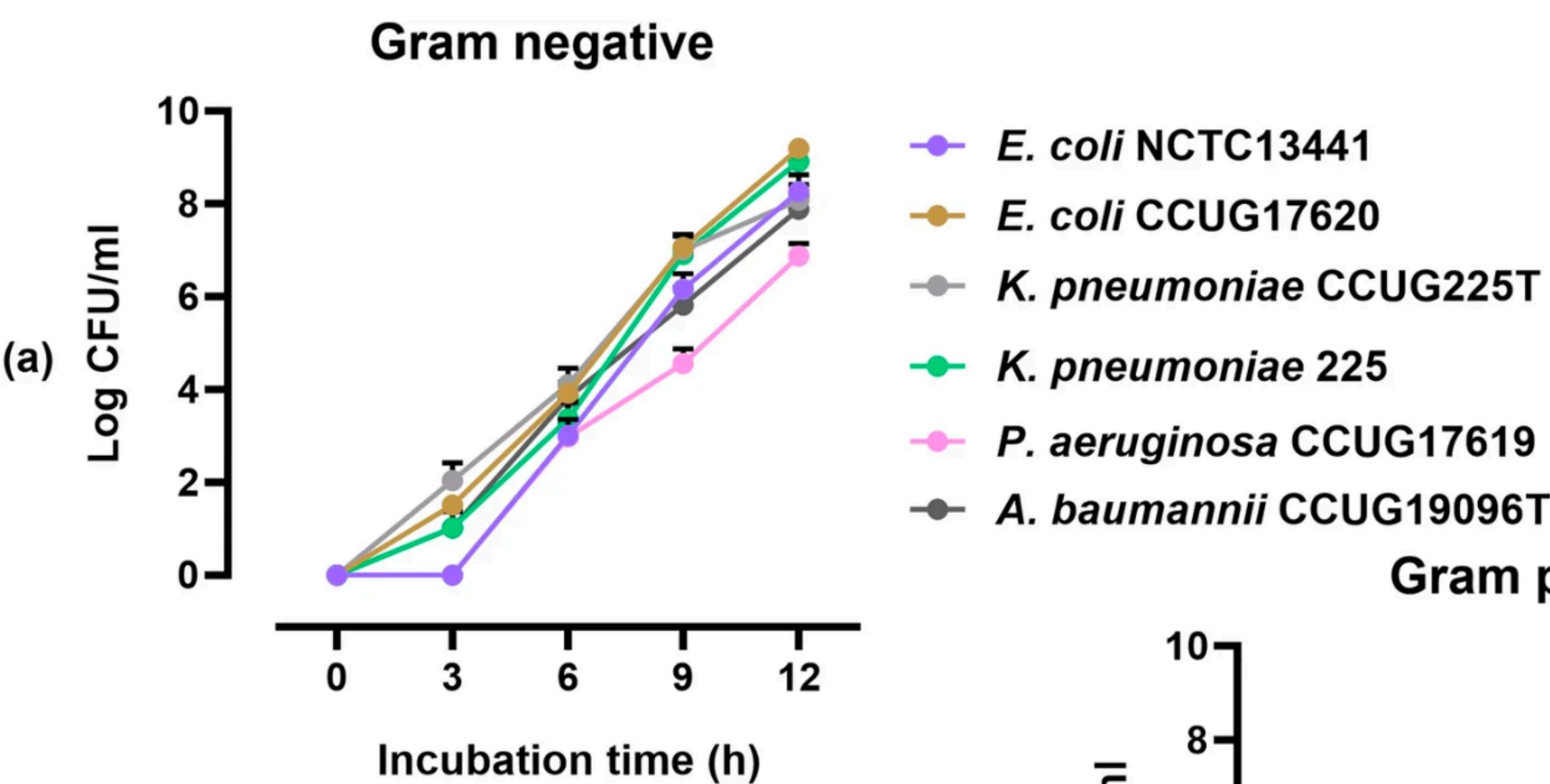


Short turnaround time of seven to nine hours from sample collection until informed decision for sepsis treatment using nanopore sequencing



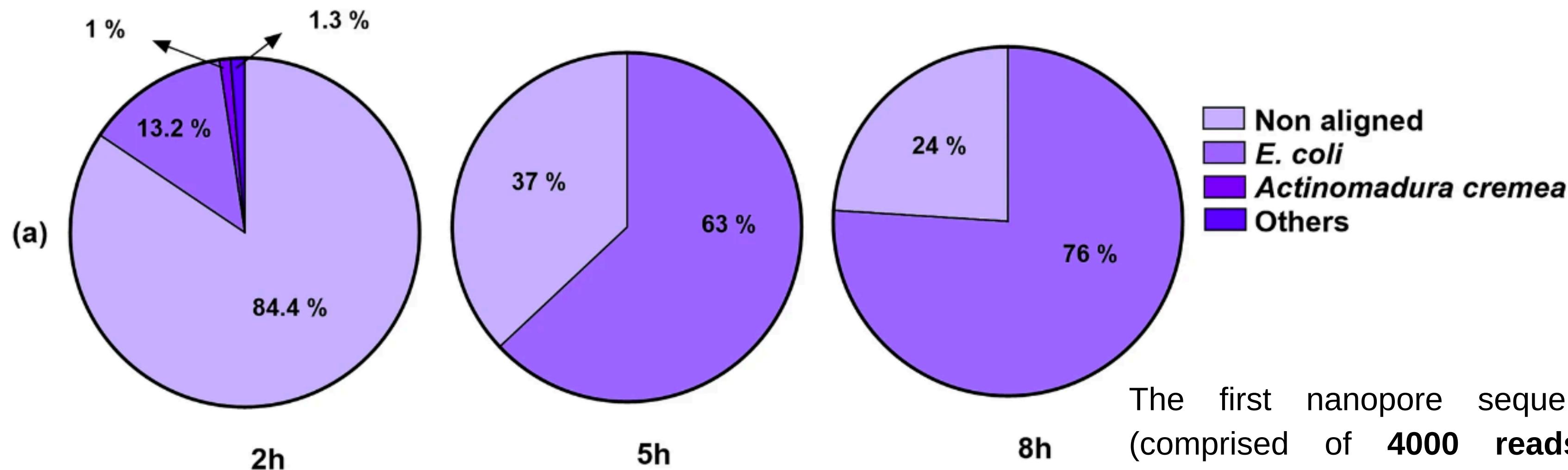
Timeline for the detection of pathogens and ARGs from blood culture samples. The timeframe for identifying pathogens and ARGs is calculated using t_0 .
DOI <https://doi.org/10.1038/s41598-024-55635-z>

Growth characteristics of sepsis-relevant bacteria in BD BACTEC blood culture medium grown in a standard incubator



The log CFU/mL of (a) gram-negative and (b) gram-positive bacterial strains following 3, 6, 9, and 12 h of incubation. The log phase of the bacteria starts after 3 h of incubation, and the growth curve increases sharply up to 12 h.

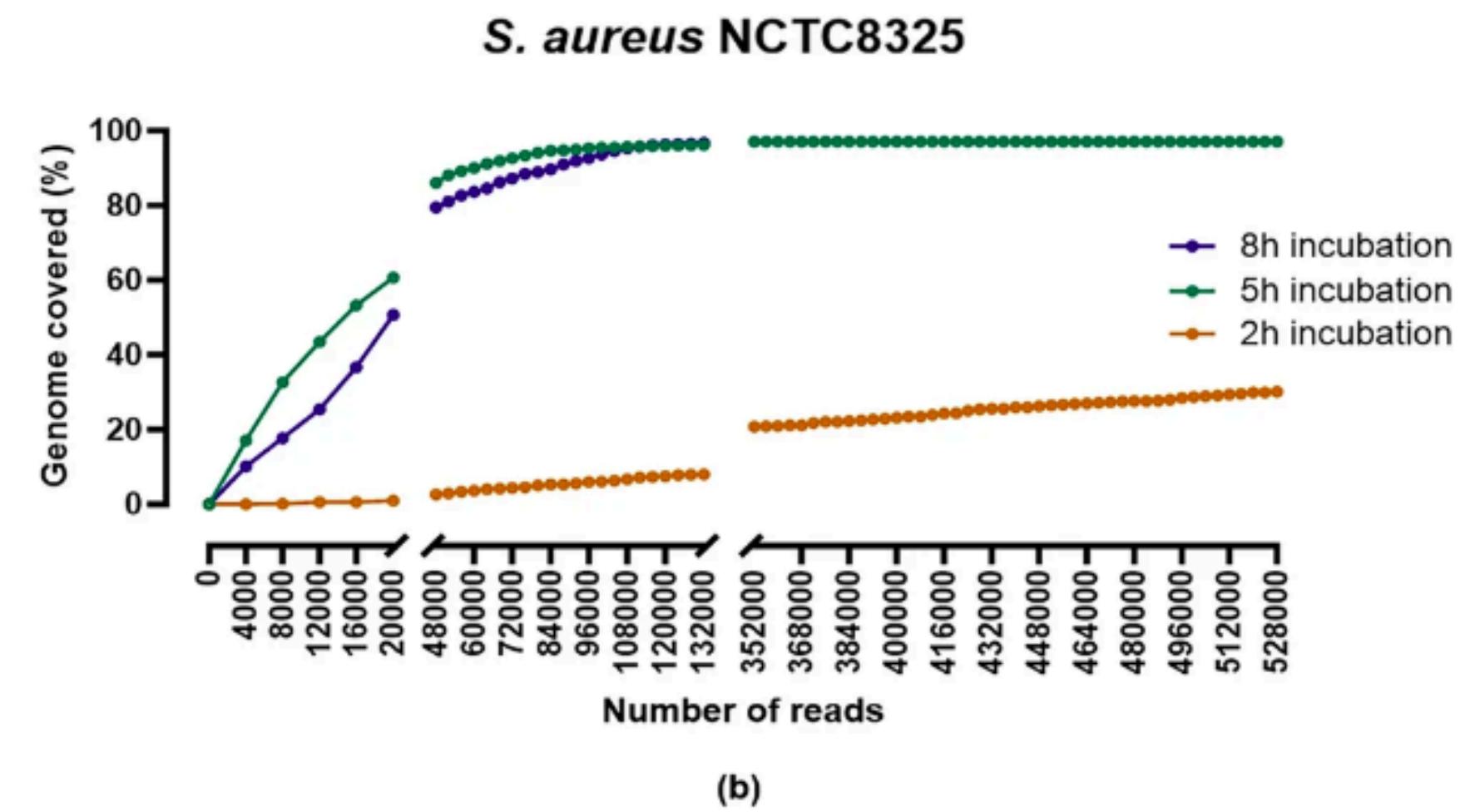
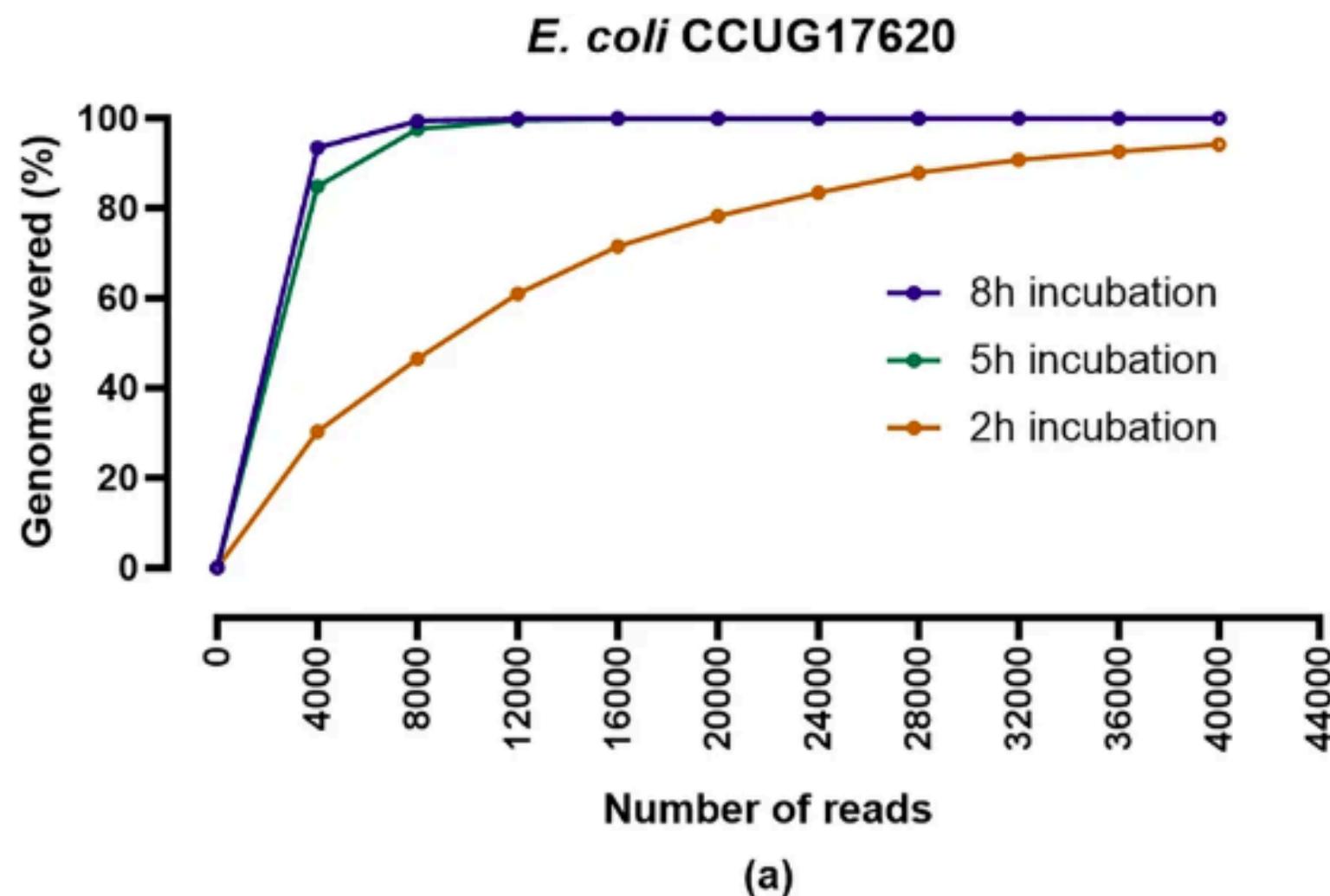
10^2 - 10^4 CFU/mL obtained after **2h of incubation** was sufficient for pathogen detection using nanopore sequencing



Relative distribution of sequencing reads generated by nanopore sequencing from blood culture

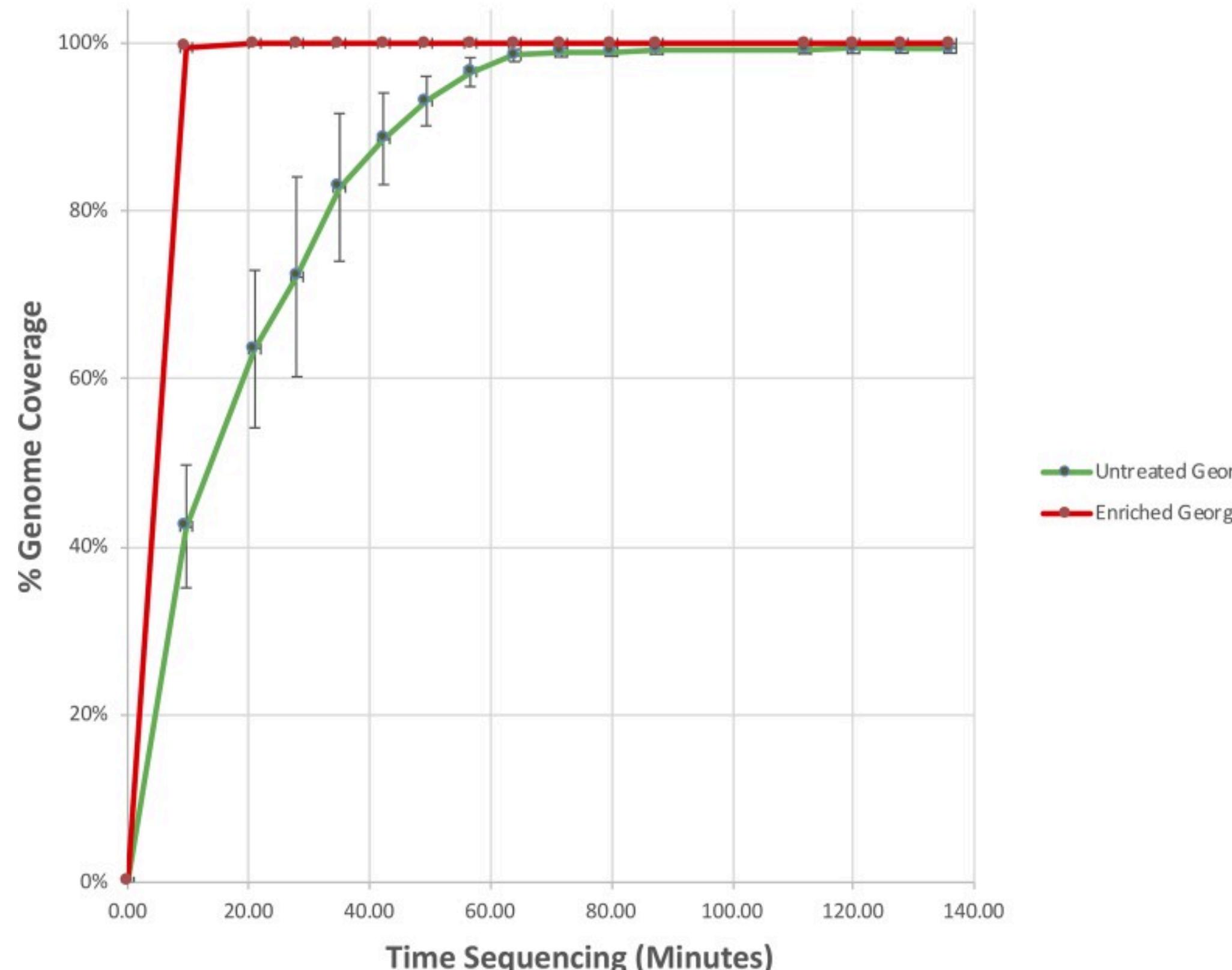
The first nanopore sequencing file (comprised of **4000 reads**), which became available **after 10–63 min** of sequencing, was enough for detecting the target bacteria at all time points and for all cultures.

90% of the bacterial genome was sequenced in < 10 h of nanopore sequencing

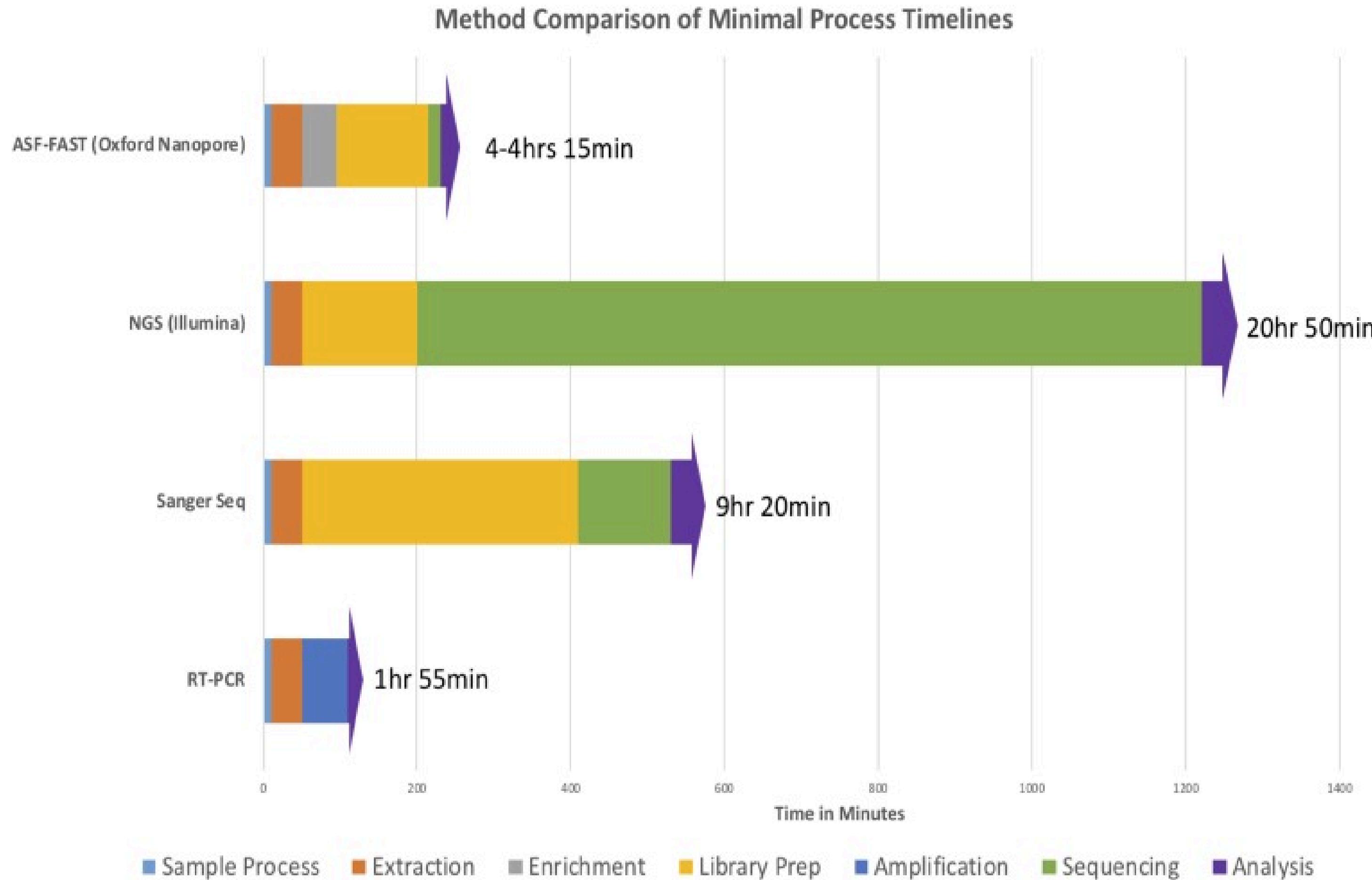


Rapid Sequence-Based Characterization of African Swine Fever Virus by Use of the Oxford Nanopore MinION

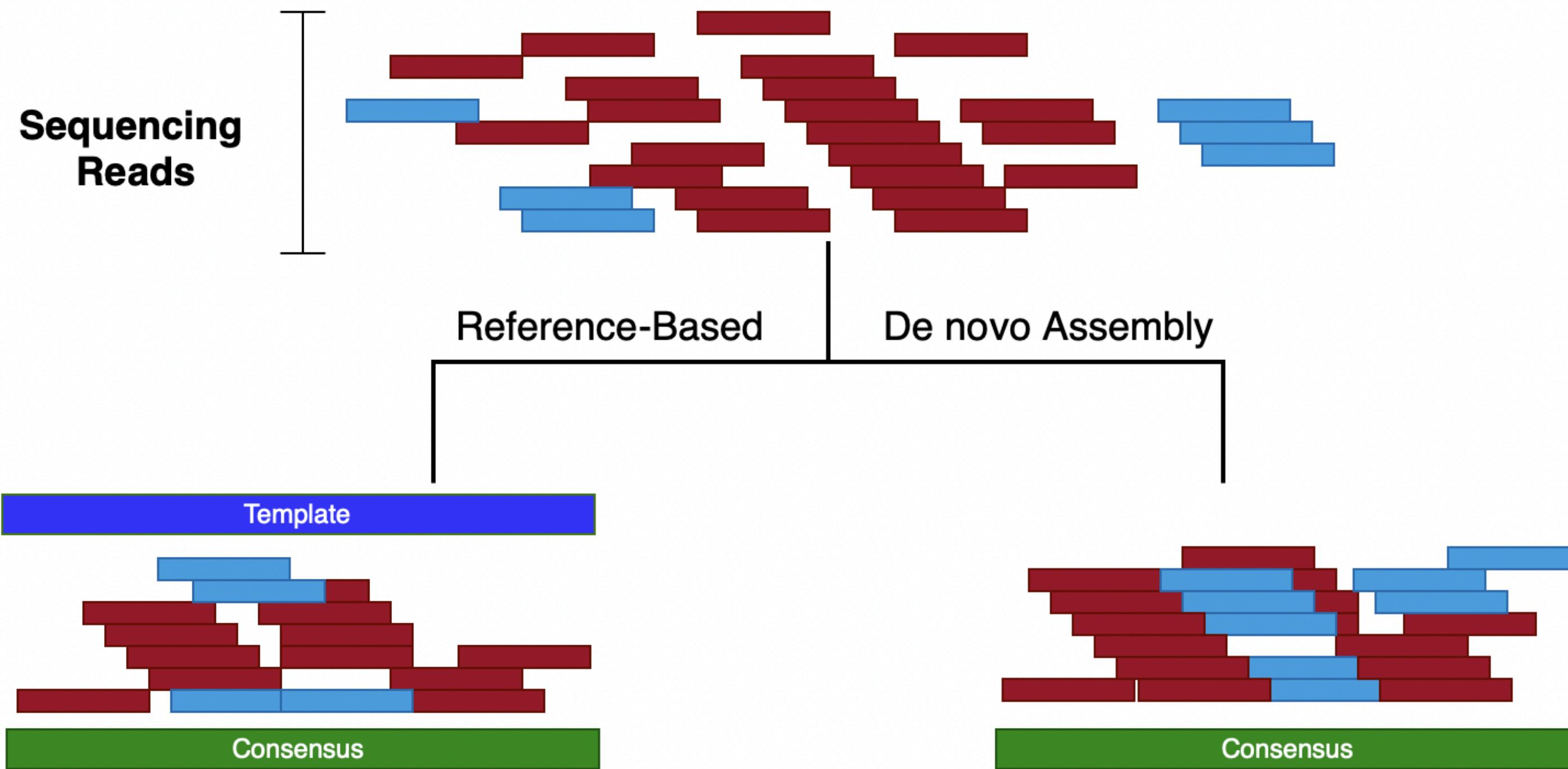
ASF Georgia Time-Course Sequencing Untreated vs.
Enriched Sample (Cell Culture Grown Virus)



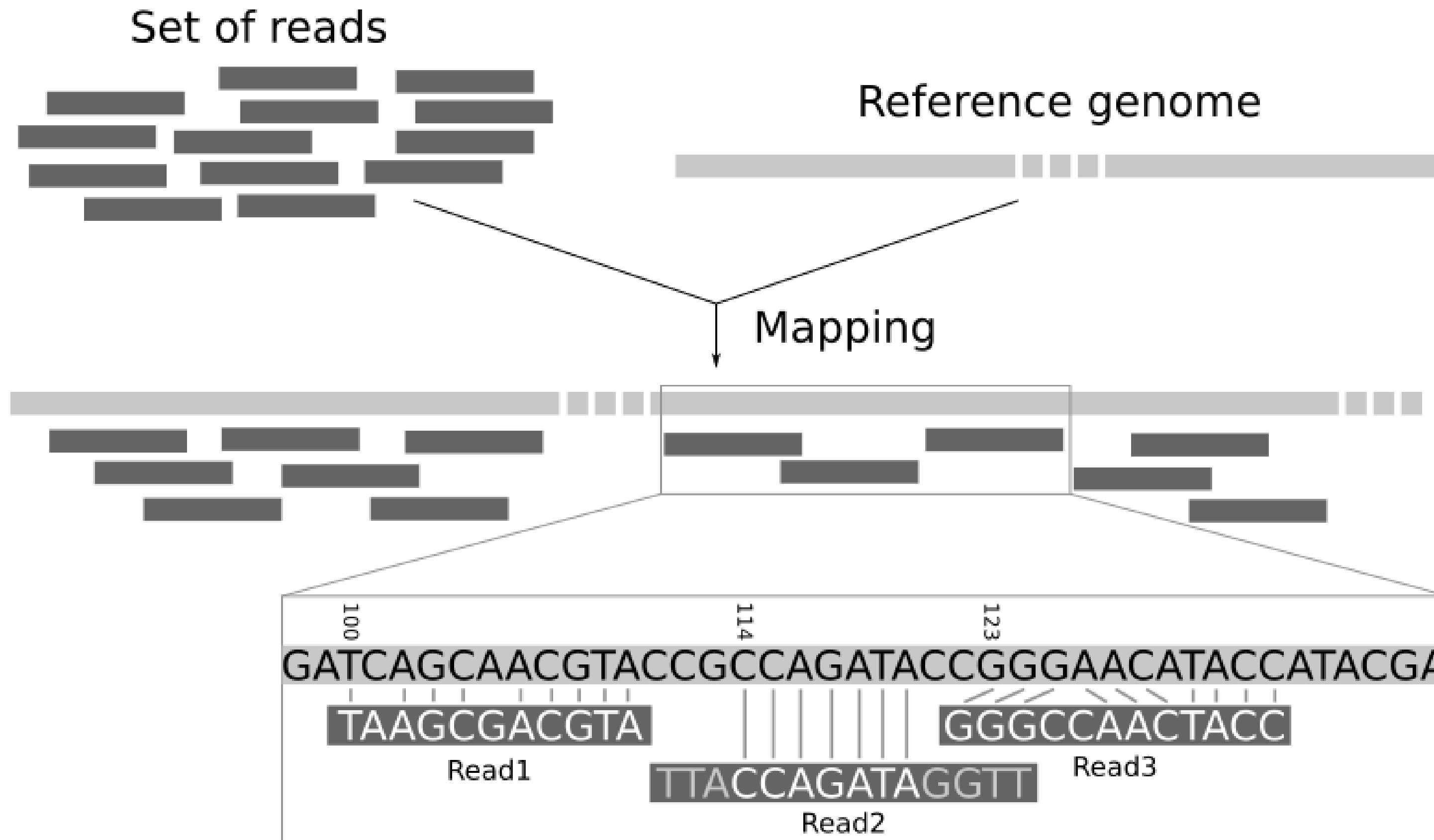
Comparison of different ASFV characterization/diagnostic method process timelines



Stratery

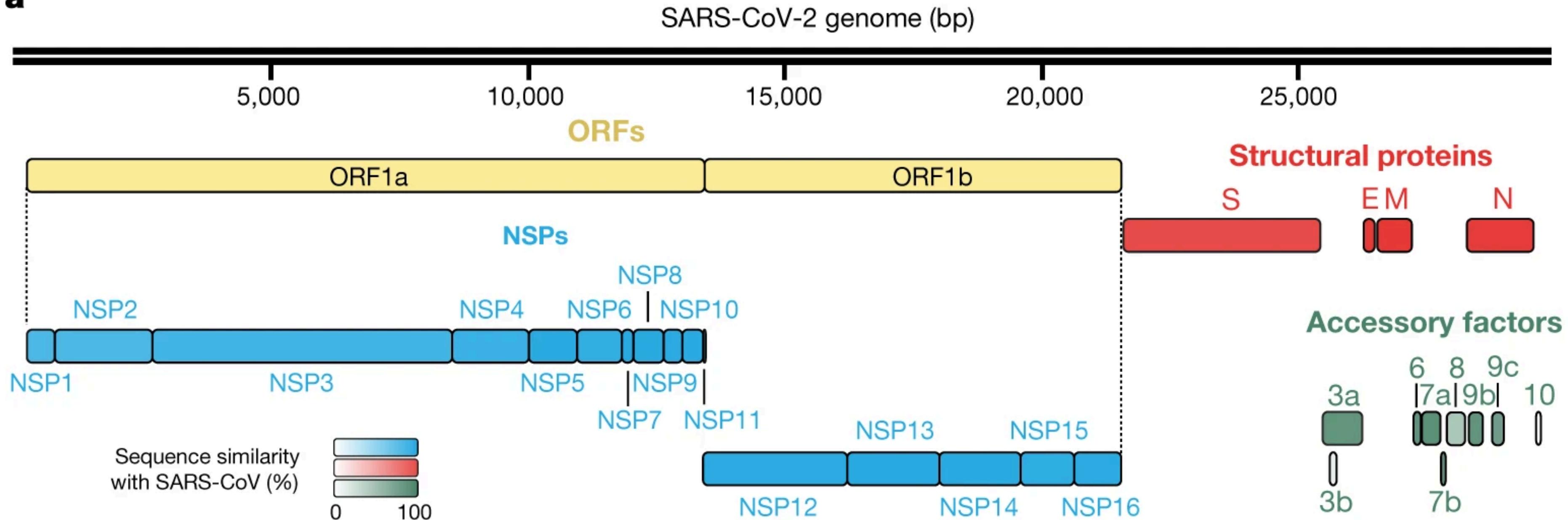


What is mapping?



You need to know the reference genome

a



SARS-CoV-2 genome annotation. The colour intensity is proportional to the protein sequence similarity with SARS-CoV homologues (when homologues exist). n = 4 structural proteins; n = 16 NSPs; n = 9 accessory factors

Reference based (mapping)



EPI2ME -ARTIC

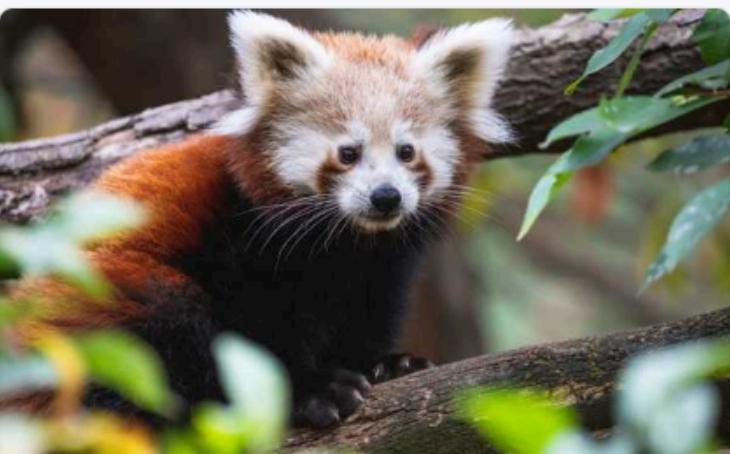
 **Dashboard**

 **Update application**
An update for this application is available.

 **View workflows**
Run and install analysis pipelines.

 **Track analyses**
View current and previous runs.

Latest updates



EPI2ME 24.08-01 Release
With the ever-growing popularity of EPI2ME workflows, we present a consolidated maintenance release of a selection of our workflows.
Published: 15 August 2024



Unexpected results, so now what?
Essentials of a Bioinformatician's toolkit.
Published: 2 July 2024



IGV for EPI2ME workflows
IGV integration for EPI2ME workflows.
Published: 10 June 2024



EPI2ME 24.06-01 Release
With all the excitement of London Calling we skipped our previous release window, so this one is a cracker!
Published: 5 June 2024

<input type="checkbox"/> Name	Workflow	Created	Status
<input type="checkbox"/> clever_fermat	wf-artic	18 Sept 2024	 Completed
<input type="checkbox"/> affectionate_yallow	wf-artic	11 Sept 2024	 Completed
<input type="checkbox"/> mystifying_knuth	wf-artic	9 Sept 2024	 Completed

← Q Search workflows and analyses ⌂ X

Workflows

Installed Available Workflows Pending Updates

 EPI2ME-LABS wf-alignment
Align Nanopore reads and visualize mapping statistics.

 EPI2ME-LABS wf-artic
Run the ARTIC SARS-CoV-2 methodology on multiplexed MinION, GridION, and PromethION data.

 EPI2ME-LABS wf-metagenomics
Identification of the origin of single reads from both amplicon-targeted and shotgun metagenomics s...

 EPI2ME-LABS wf-bacterial-genomes
Assembly, variant calling, and annotation of bacterial genomes.

 EPI2ME-LABS wf-cas9
Summarise the results of Cas9 enrichment sequencing.

 EPI2ME-LABS wf-clone-validation
De-novo reconstruction of synthetic plasmid sequences.

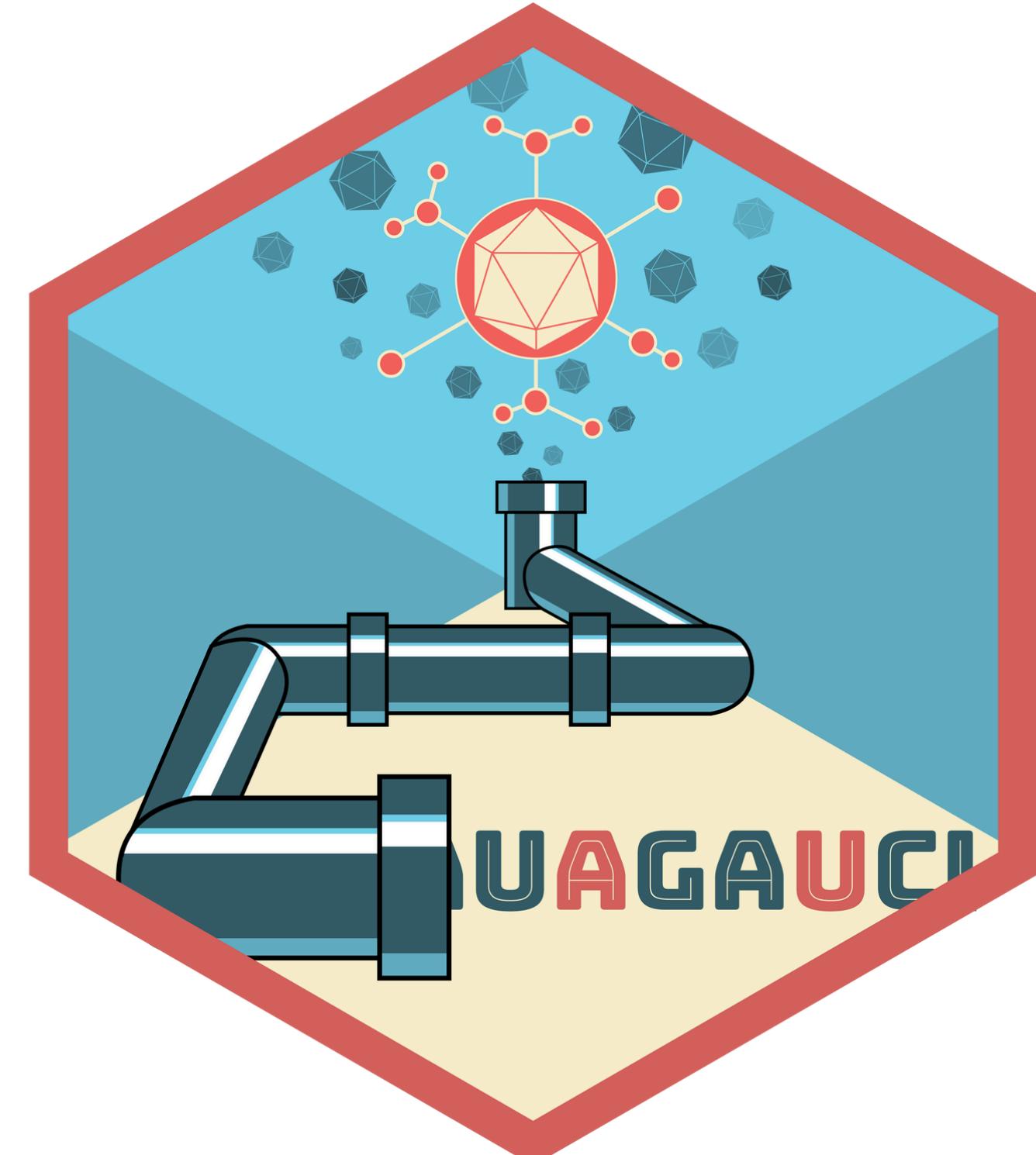
 EPI2ME-LABS wf-human-variation
Basecalling, SNV calling, SV calling, modified base calling, CNV calling, and STR genotyping of huma...

 EPI2ME-LABS wf-single-cell
Identification of cell- and UMI barcodes from single-cell sequencing.

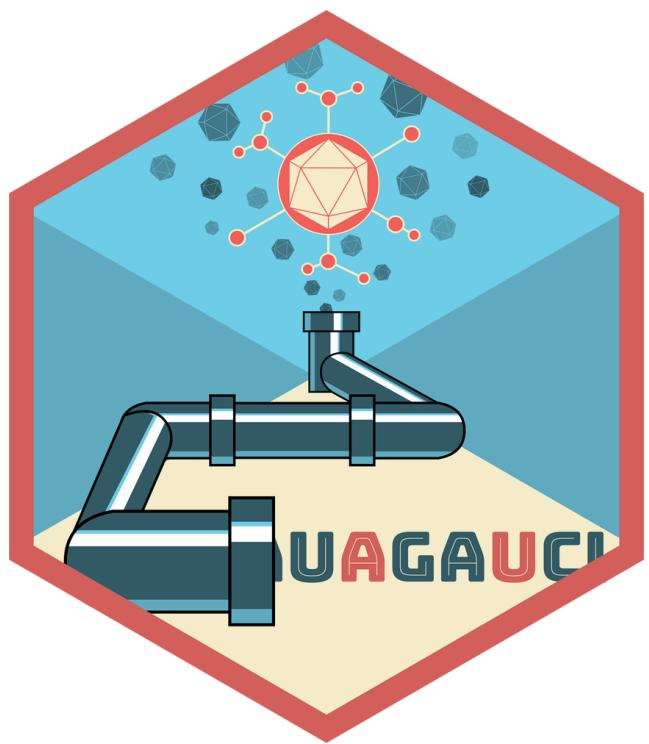
 EPI2ME-LABS wf-tb-amr

 EPI2ME-LABS wf-transcriptomes

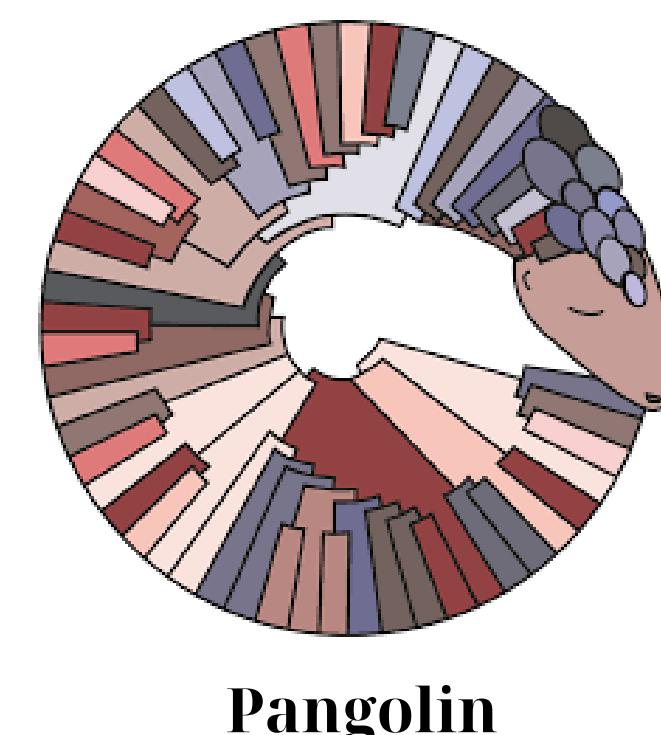
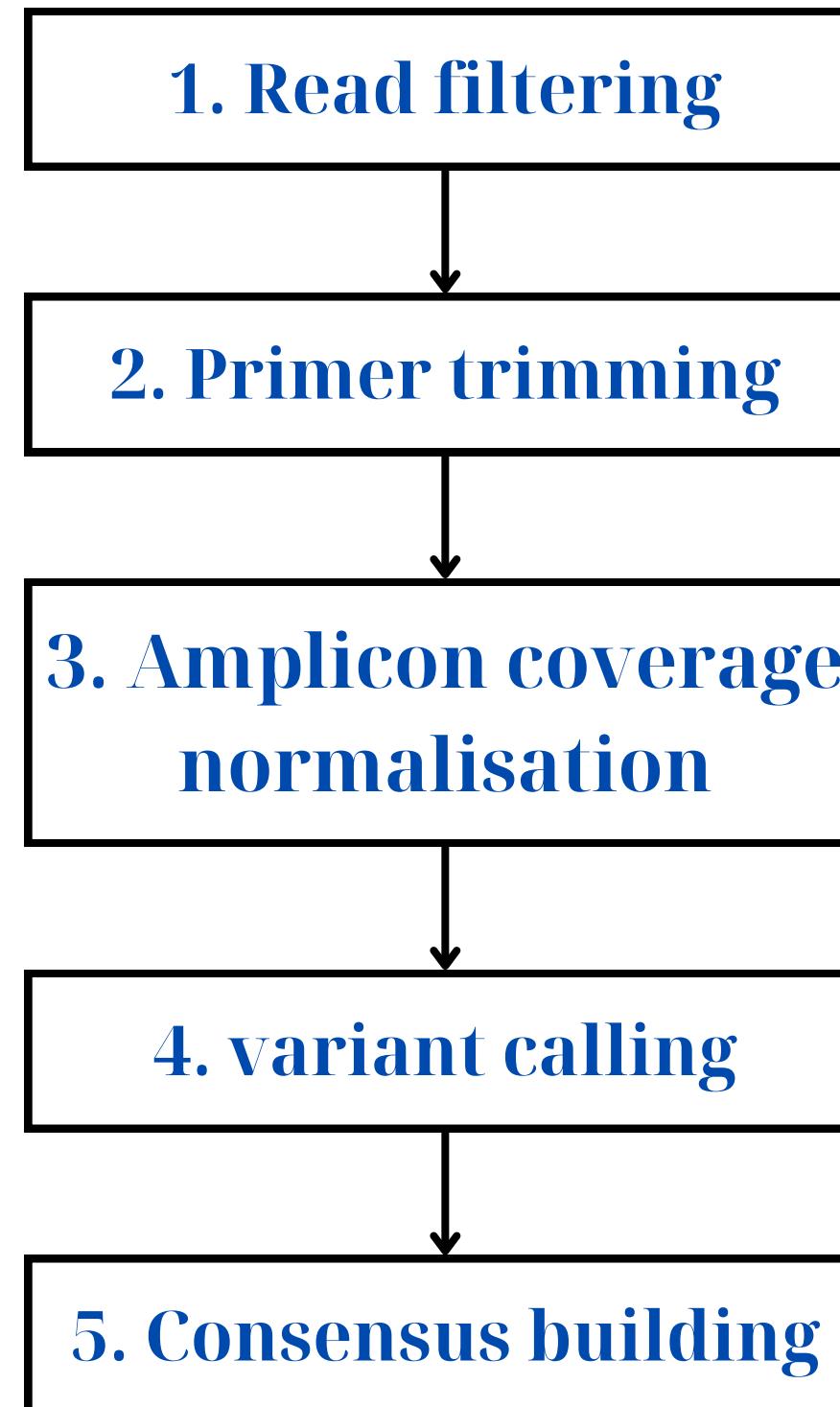
Reference based (mapping)

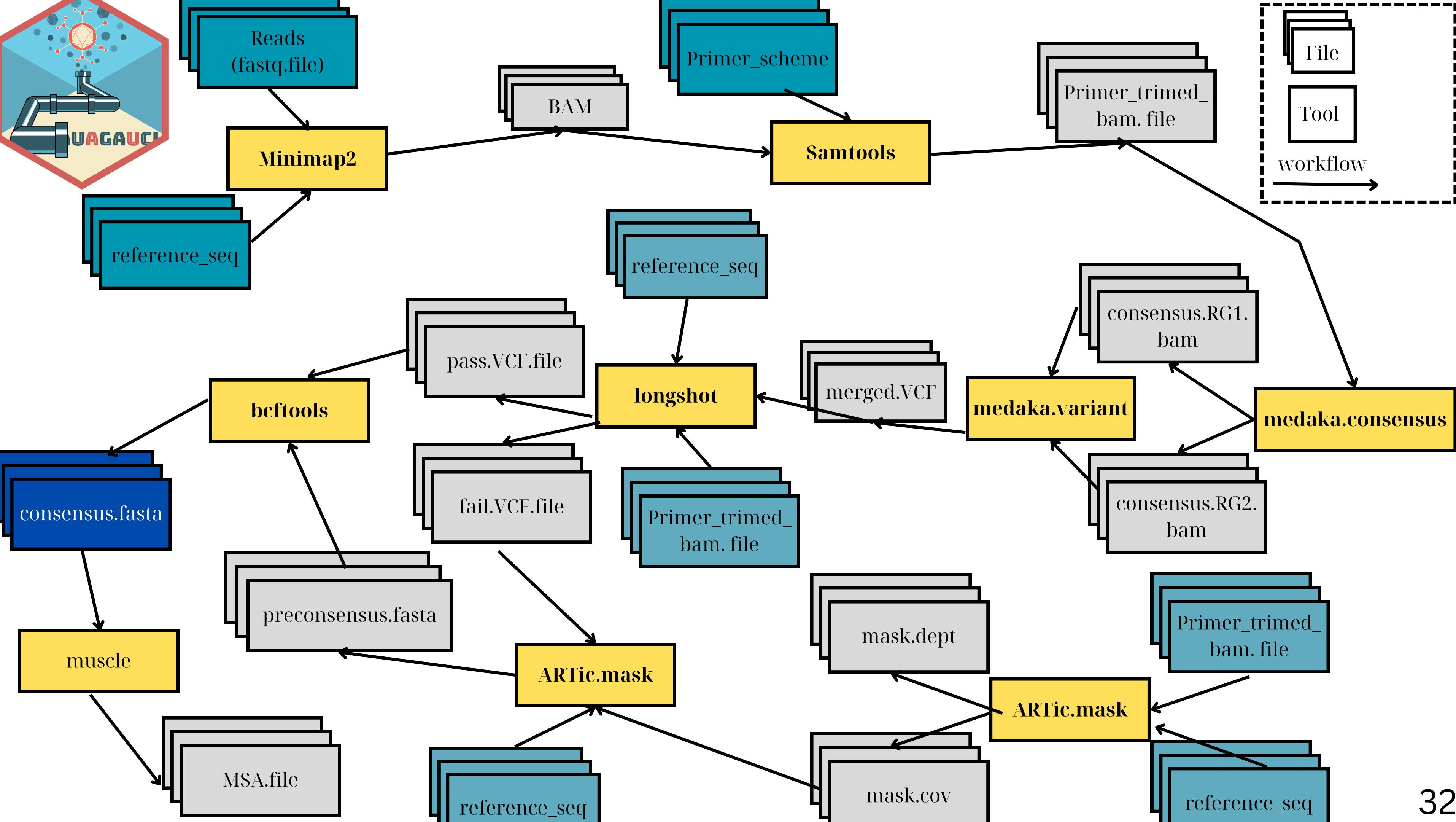


ARTIC-pipeline



ARTIC-pipeline





[nature](#) > [scientific reports](#) > [articles](#) > [article](#)

Article | [Open access](#) | Published: 26 June 2023

Universal whole-genome Oxford nanopore sequencing of SARS-CoV-2 using tiled amplicons

[Ruslan Kalendar](#)✉, [Ulykbek Kairov](#), [Daniyar Karabayev](#), [Akbota Aitkulova](#), [Nuray Tynyshtykbayeva](#), [Asset Daniyarov](#), [Zhenis Otarbay](#), [Saule Rakhimova](#), [Ainur Akilzhanova](#) & [Dos Sarbassov](#)

[Scientific Reports](#) **13**, Article number: 10334 (2023) | [Cite this article](#)

3235 Accesses | 1 Citations | 4 Altmetric | [Metrics](#)

<https://doi.org/10.1038/s41598-023-37588-x>

multiqc report

The screenshot shows the multiqc reporting interface. On the left, a sidebar lists various analysis modules: General Stats, FastQC, Sequence Counts, Sequence Quality Histograms, Per Sequence Quality Scores, Per Base Sequence Content, Per Sequence GC Content, Per Base N Content, Sequence Length Distribution, Sequence Duplication Levels, Overrepresented sequences by sample, Top overrepresented sequences, Adapter Content, Status Checks, and Software Versions. The main content area displays the 'General Statistics' section, which includes a table with columns for Sample Name, Dups, GC, and Seqs. The table shows data for samples 65, 66, and 662. A 'FastQC' section is also visible at the bottom. The right side features a vertical 'Toolbox' with icons for various tools.

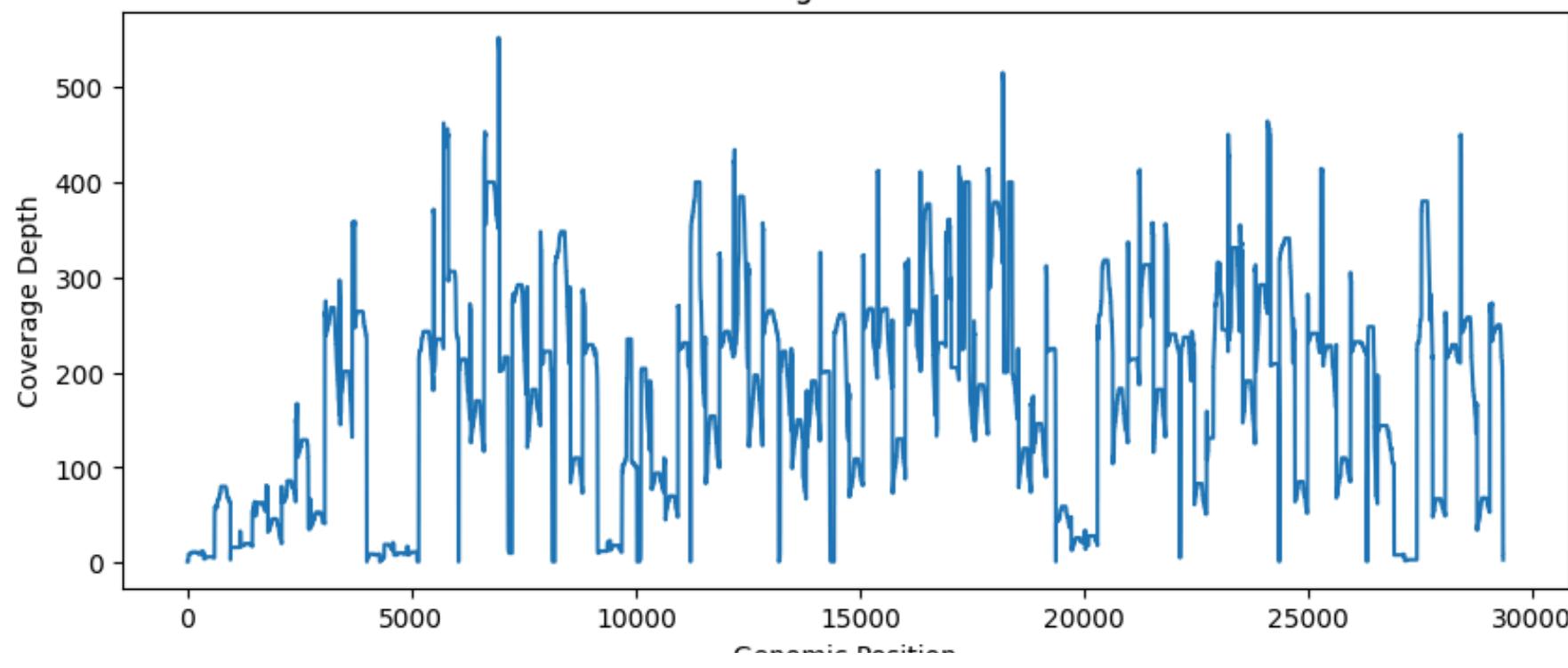
General Statistics

Sample Name	Dups	GC	Seqs
65	9.4 %	39.0 %	0.3 M
66	13.5 %	40.0 %	0.2 M
662	10.1 %	40.0 %	0.2 M

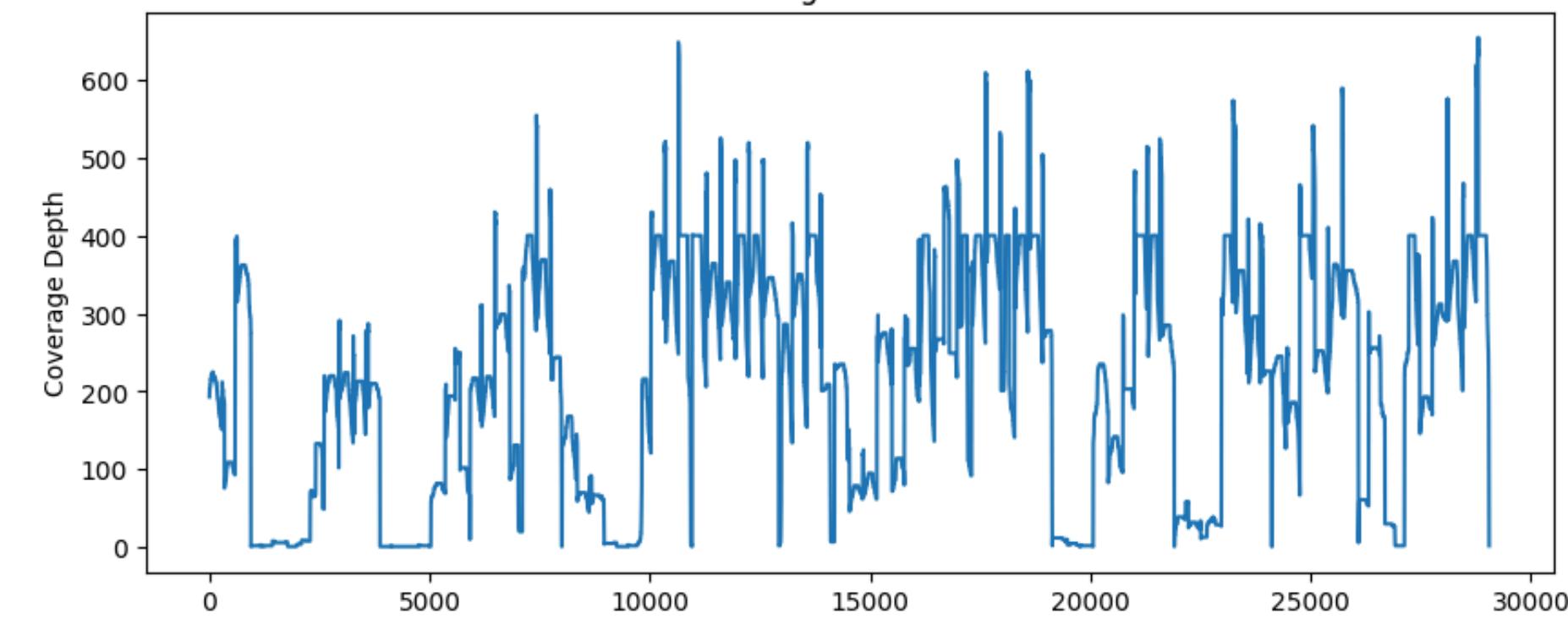
FastQC Version: 0.12.1

Read coverage

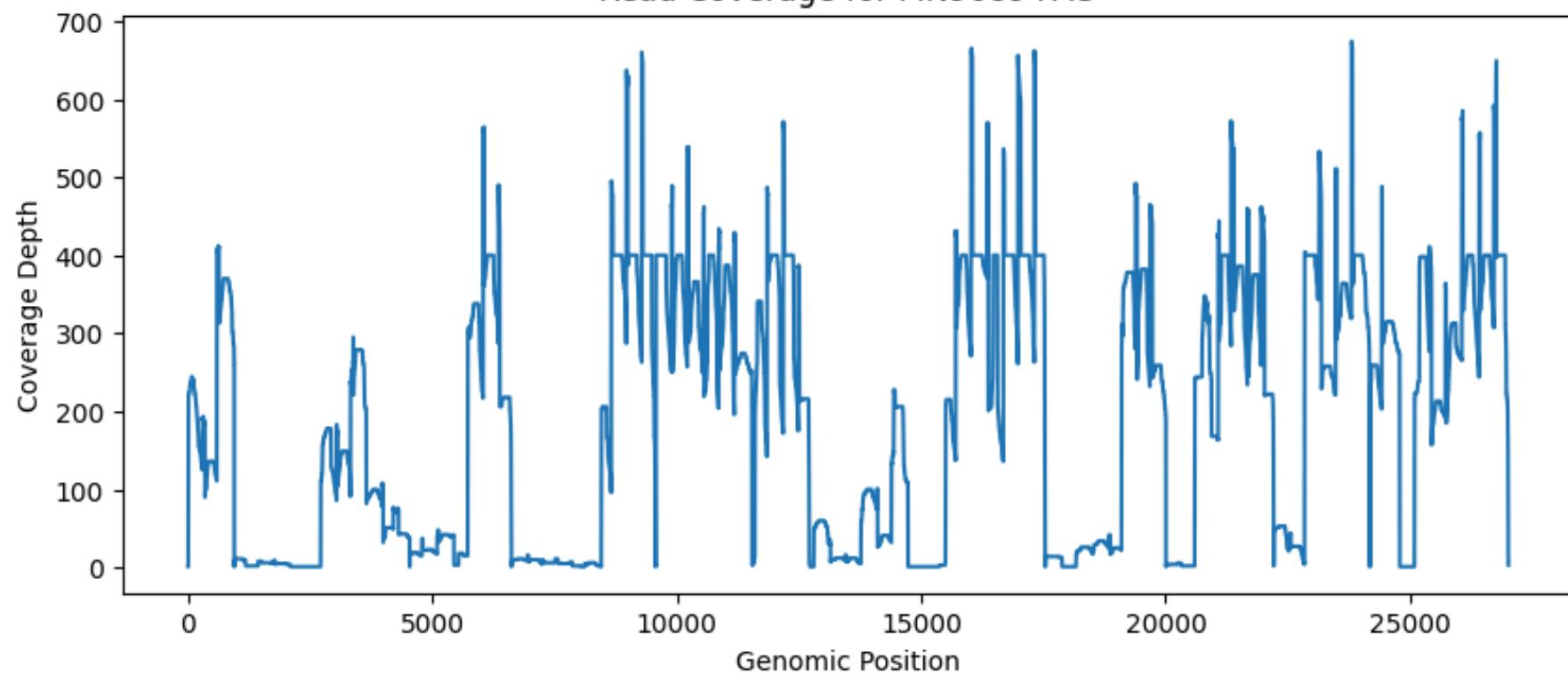
Read Coverage for MN908947.3



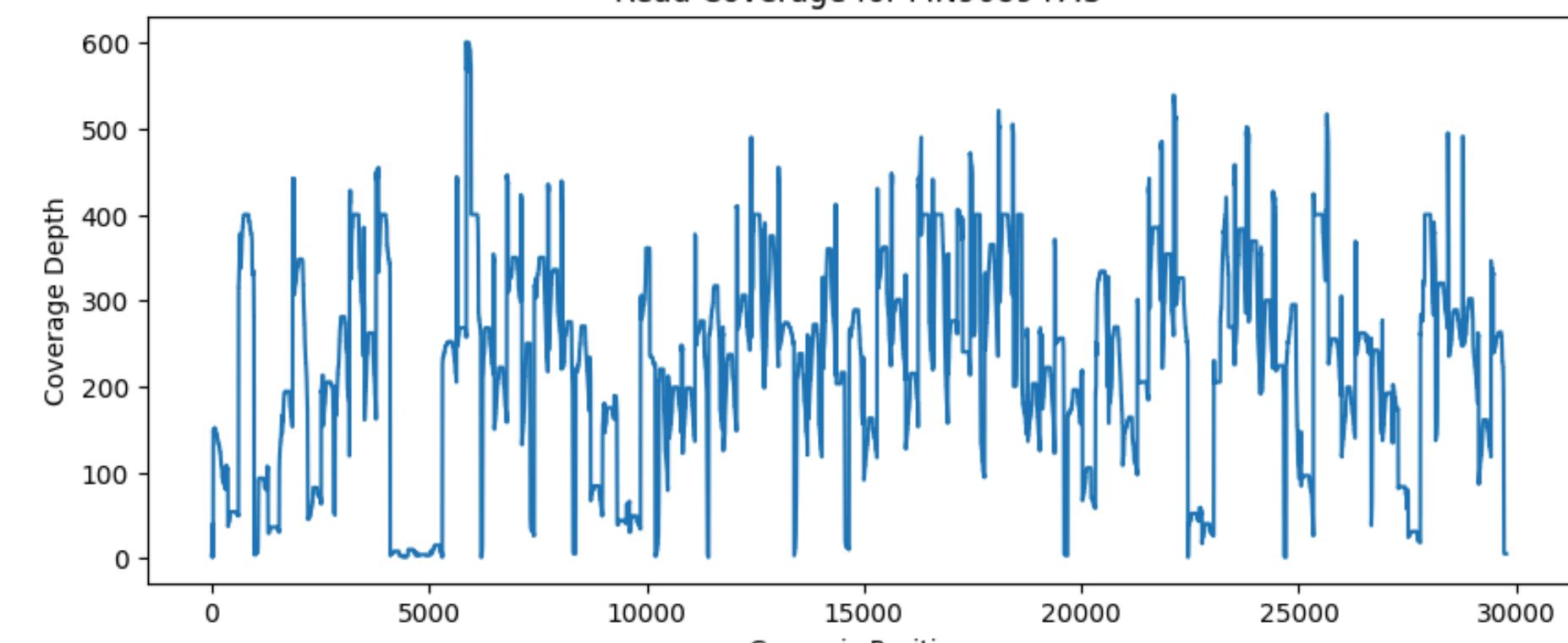
Read Coverage for MN908947.3



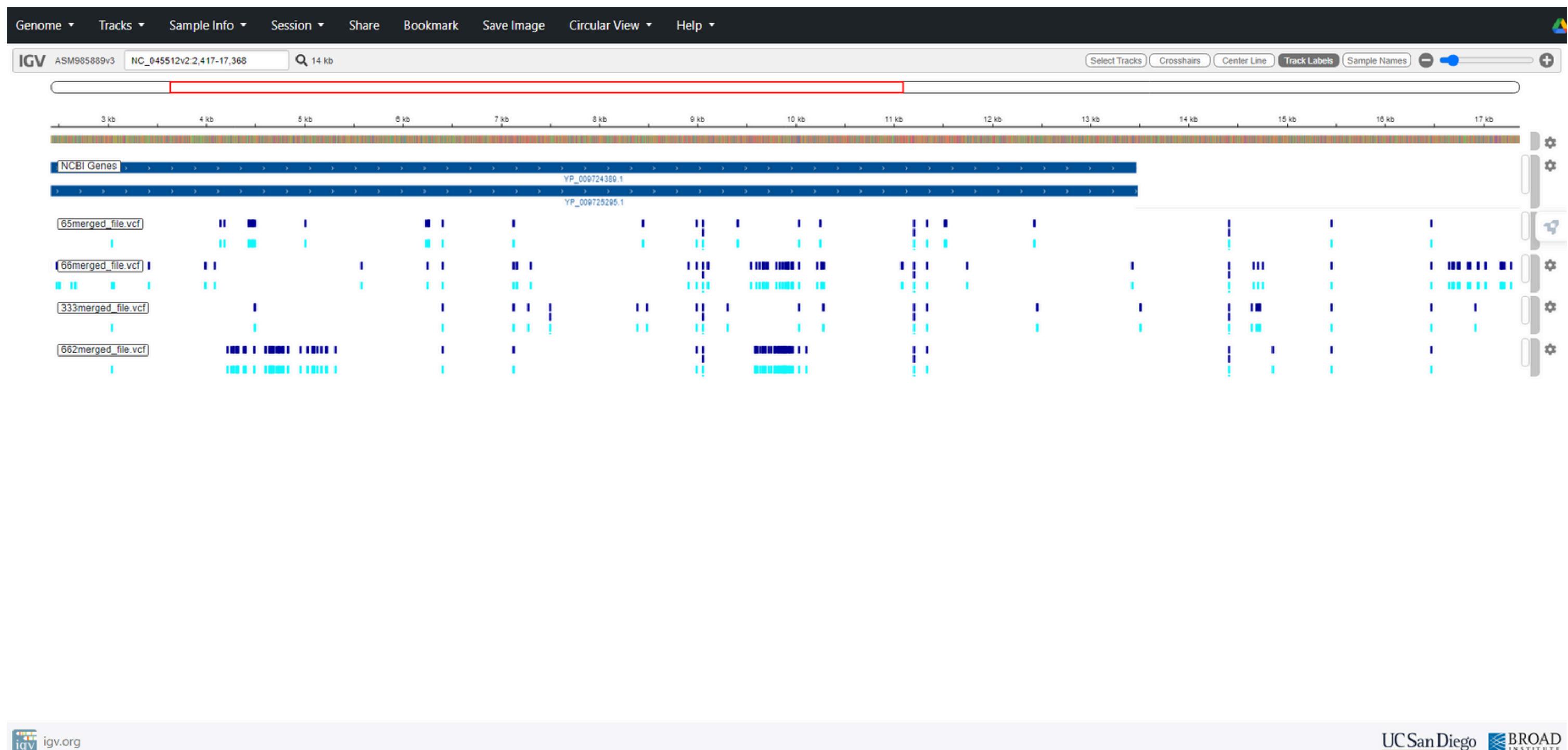
Read Coverage for MN908947.3



Read Coverage for MN908947.3

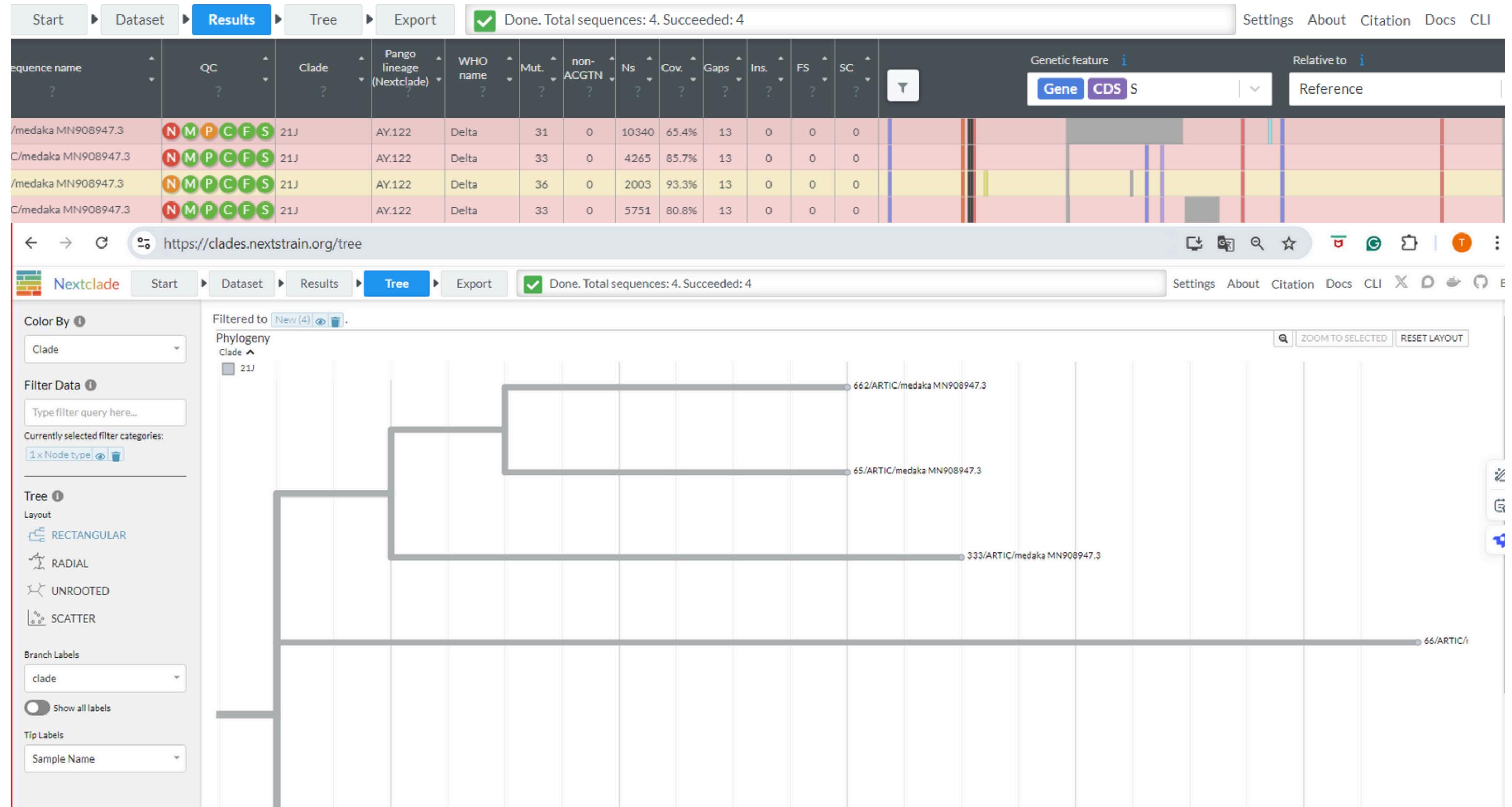


Variant calling



<https://igv.org/app/>

Nextclade result



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

De novo assembly



Genome size estimation



Raw-read statistics



De novo genome assembly



Quality assessment of genome assembly

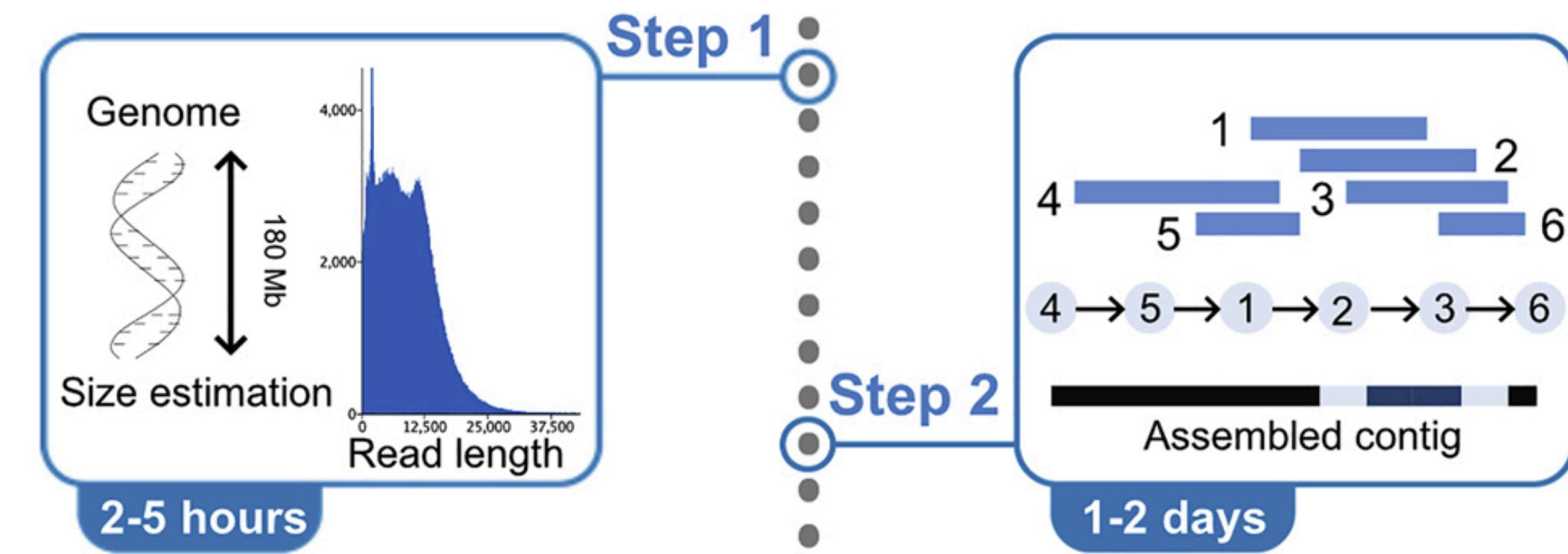


Scaffolding using a reference genome



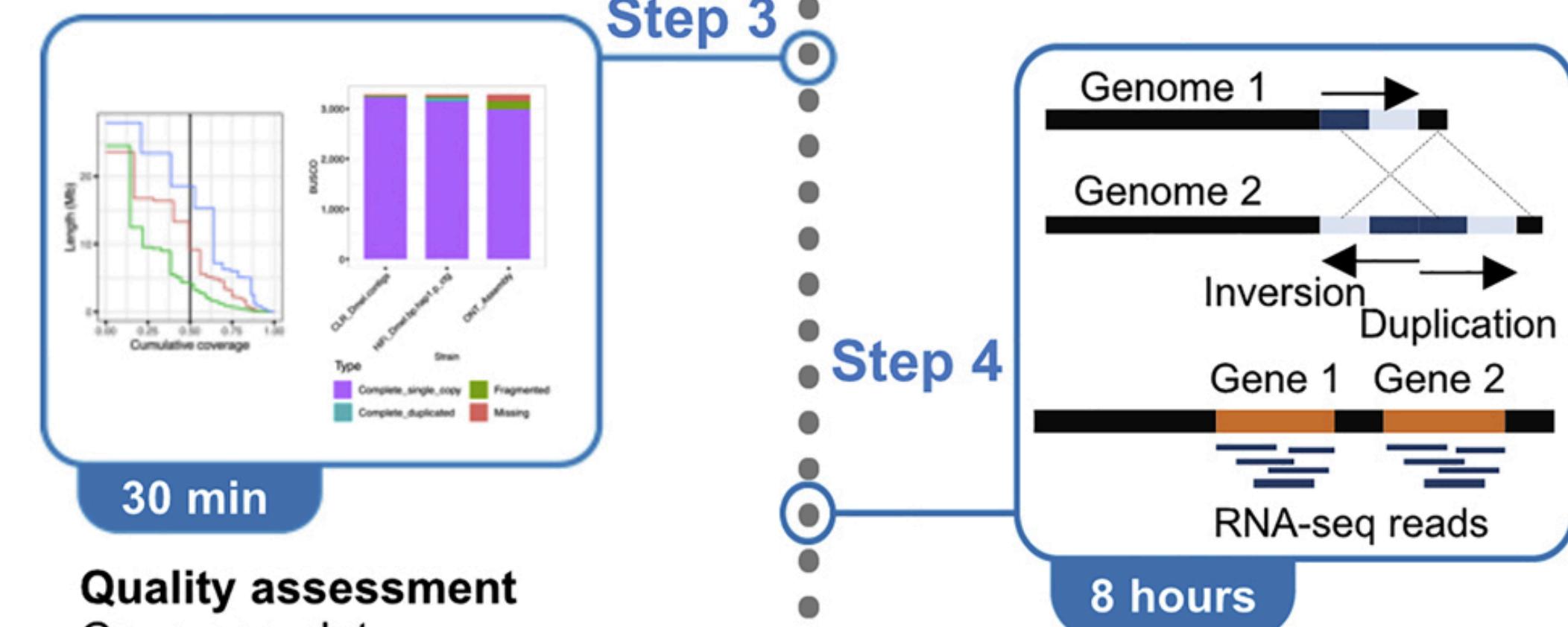
Structural variant calling and gene annotation

Detail Workflow



Prepare genome assembly
Genome size estimation
Long-read quality assessment

Genome assembly
PacBio CLR – Canu
PacBio HiFi – Hifiasm
ONT Long - Shasta

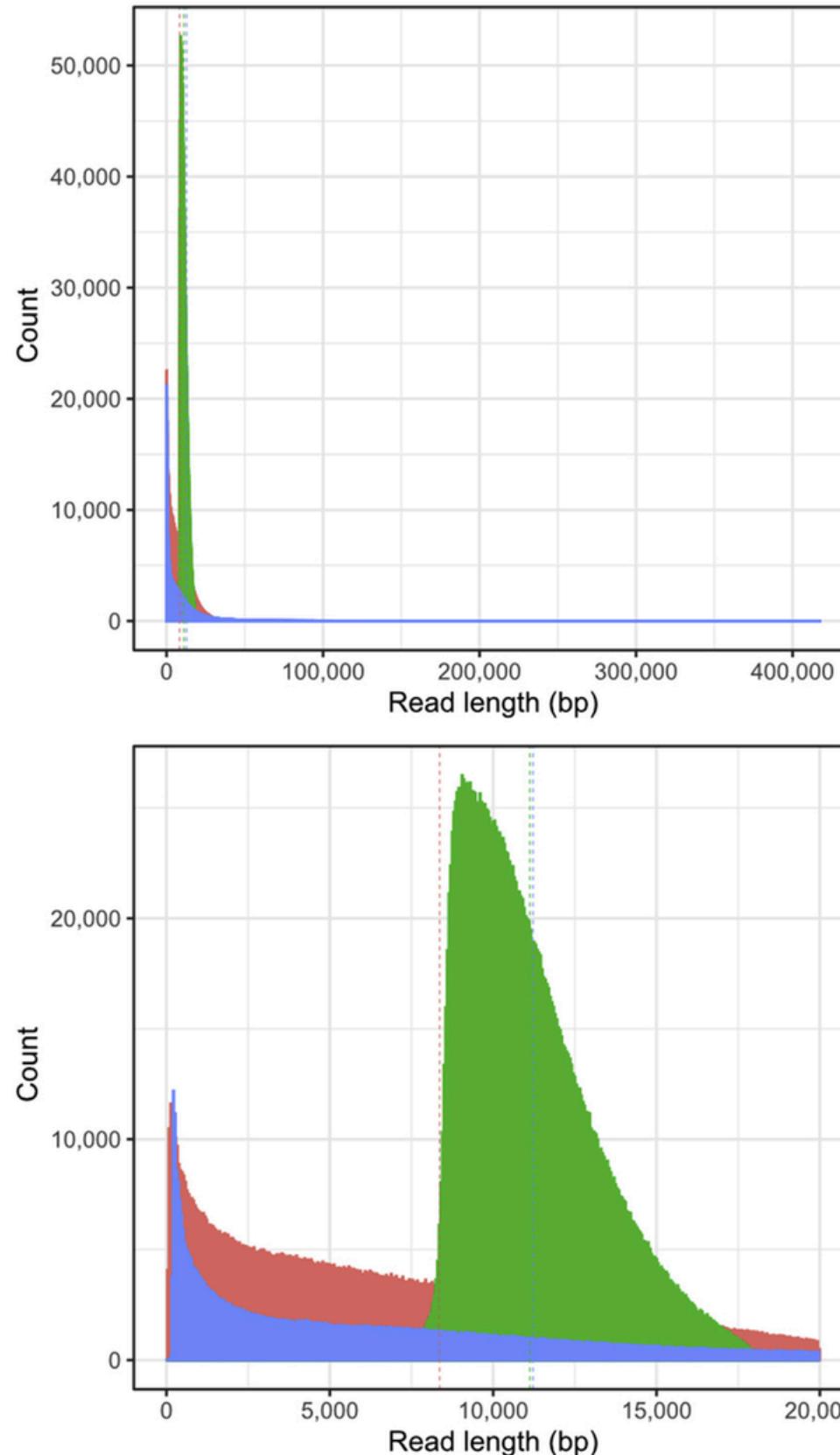


Quality assessment
Coverage plot
BUSCO analysis

Genomic feature
Structural variant calling
Gene annotation

A

Expected result



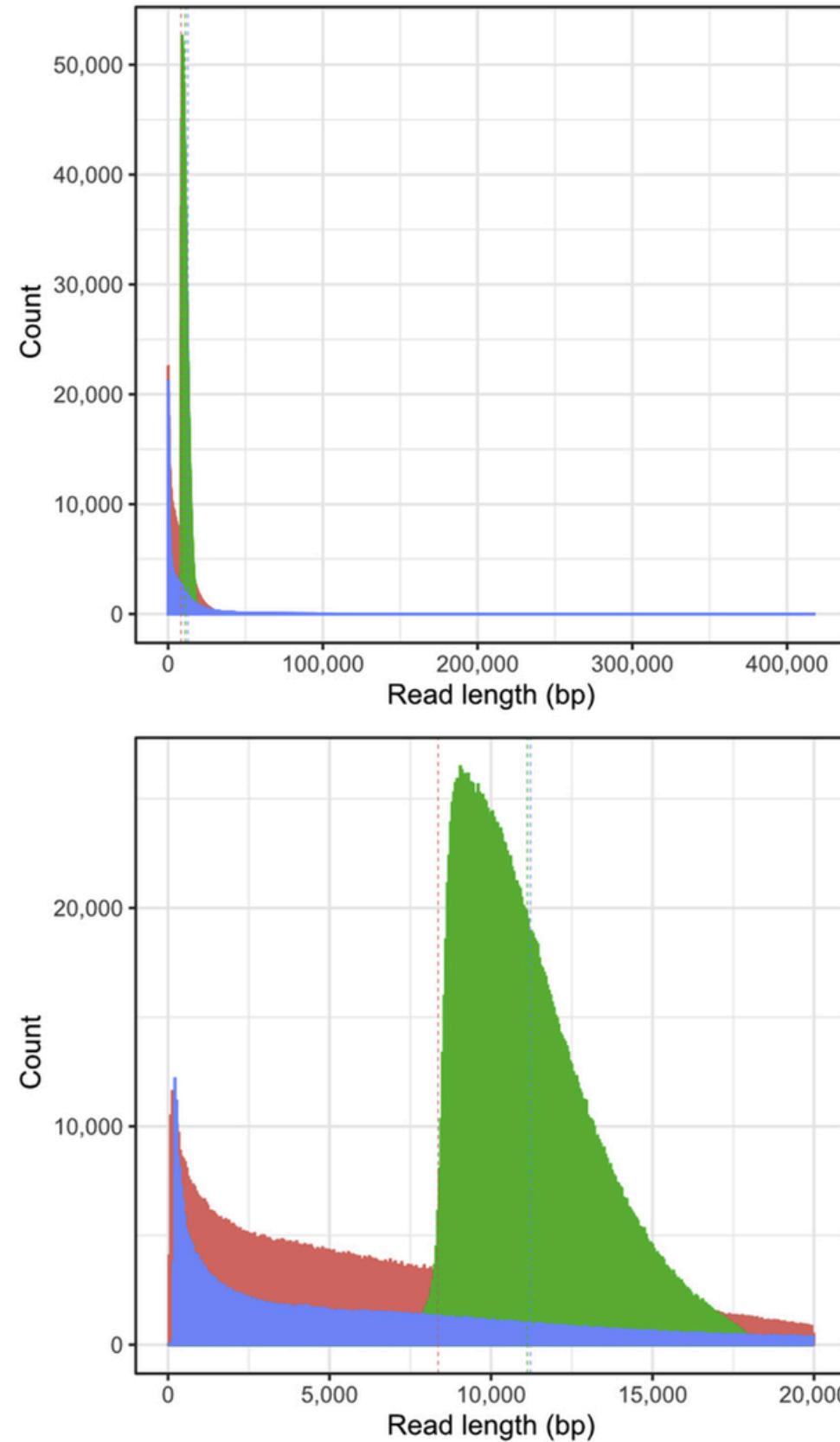
De novo genome assembly read-length distribution and quality assessment

(A) Read-length distributions of the three publicly available datasets used in this study. Each vertical dotted line represents the mean value of each dataset.

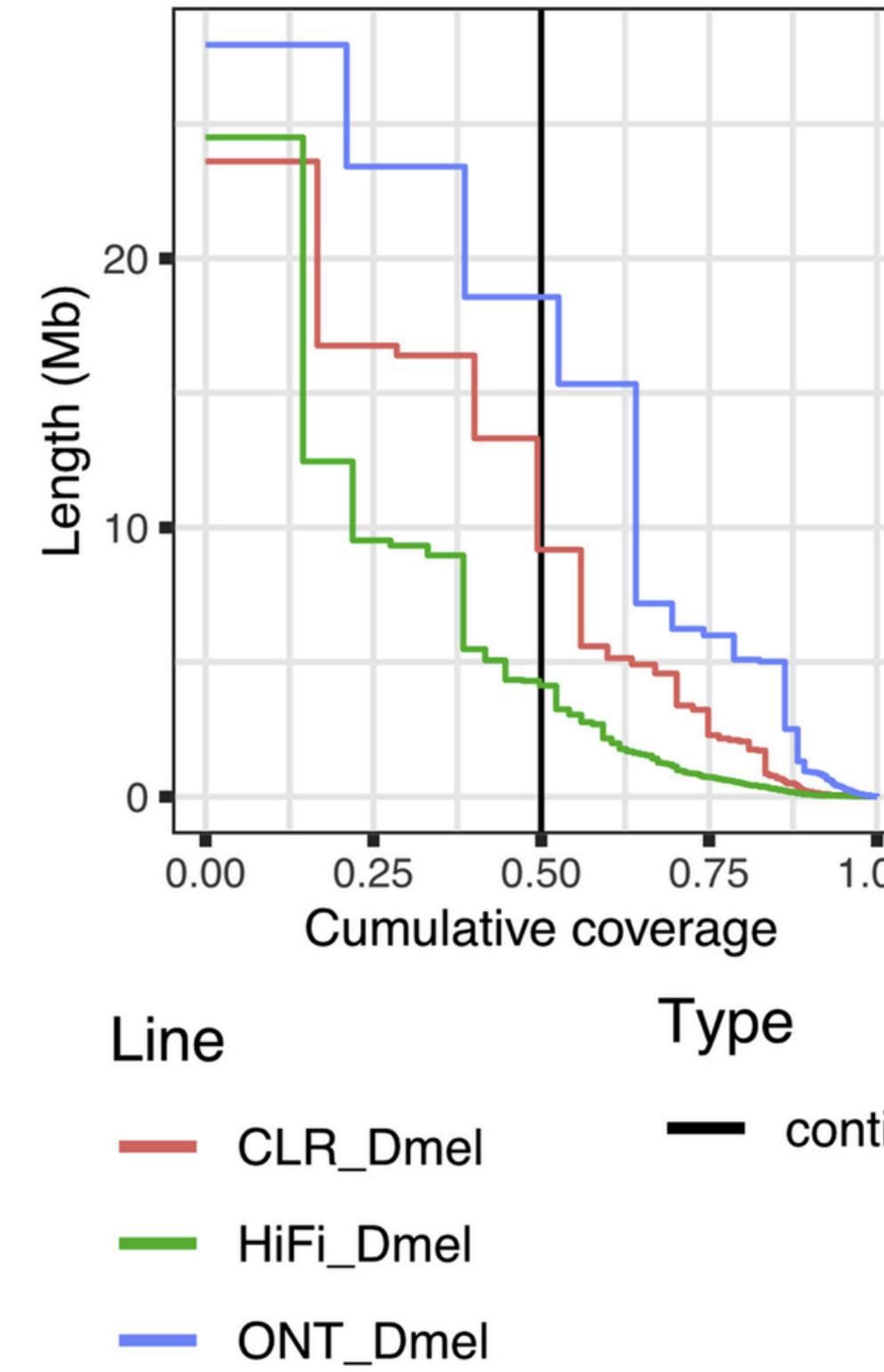
(B) Cumulative coverage plot for the contig/scaffold length.

(C) BUSCO analysis used to determine the number of single-copy orthologs known in a lineage.

A Expected result



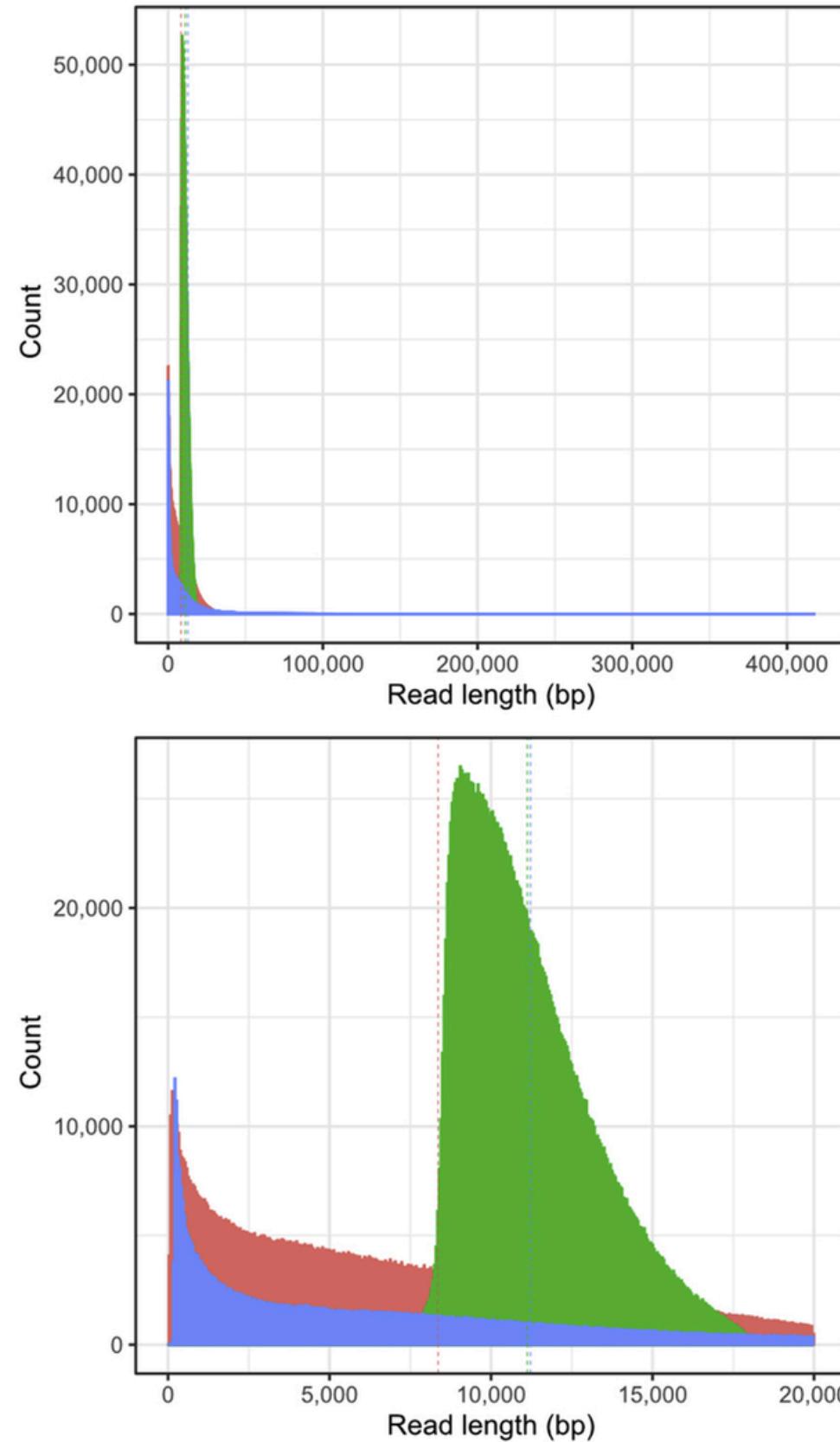
B



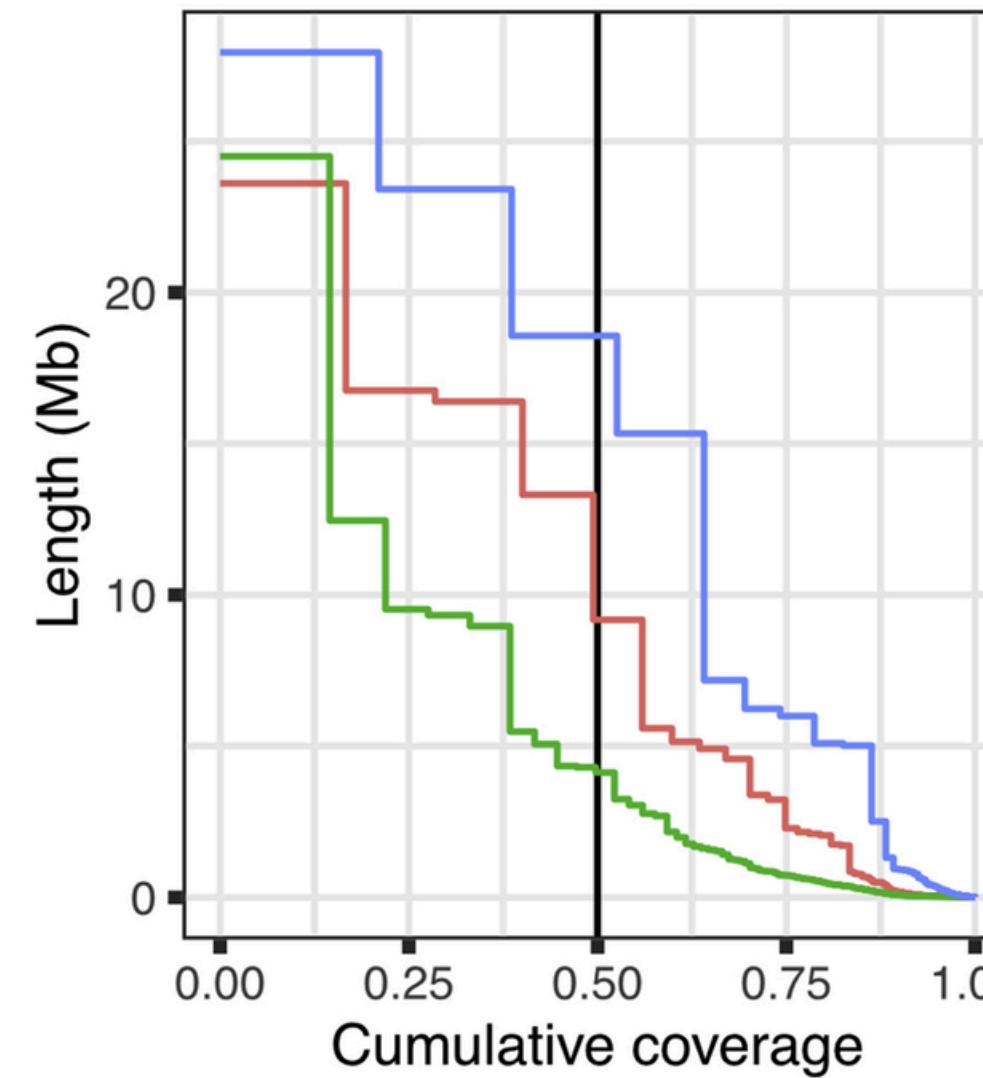
De novo genome assembly read-length distribution and quality assessment

- (A) Read-length distributions of the three publicly available datasets used in this study. Each vertical dotted line represents the mean value of each dataset.
(B) Cumulative coverage plot for the contig/scaffold length.
(C) BUSCO analysis used to determine the number of single-copy orthologs known in a lineage.

A Expected result



B



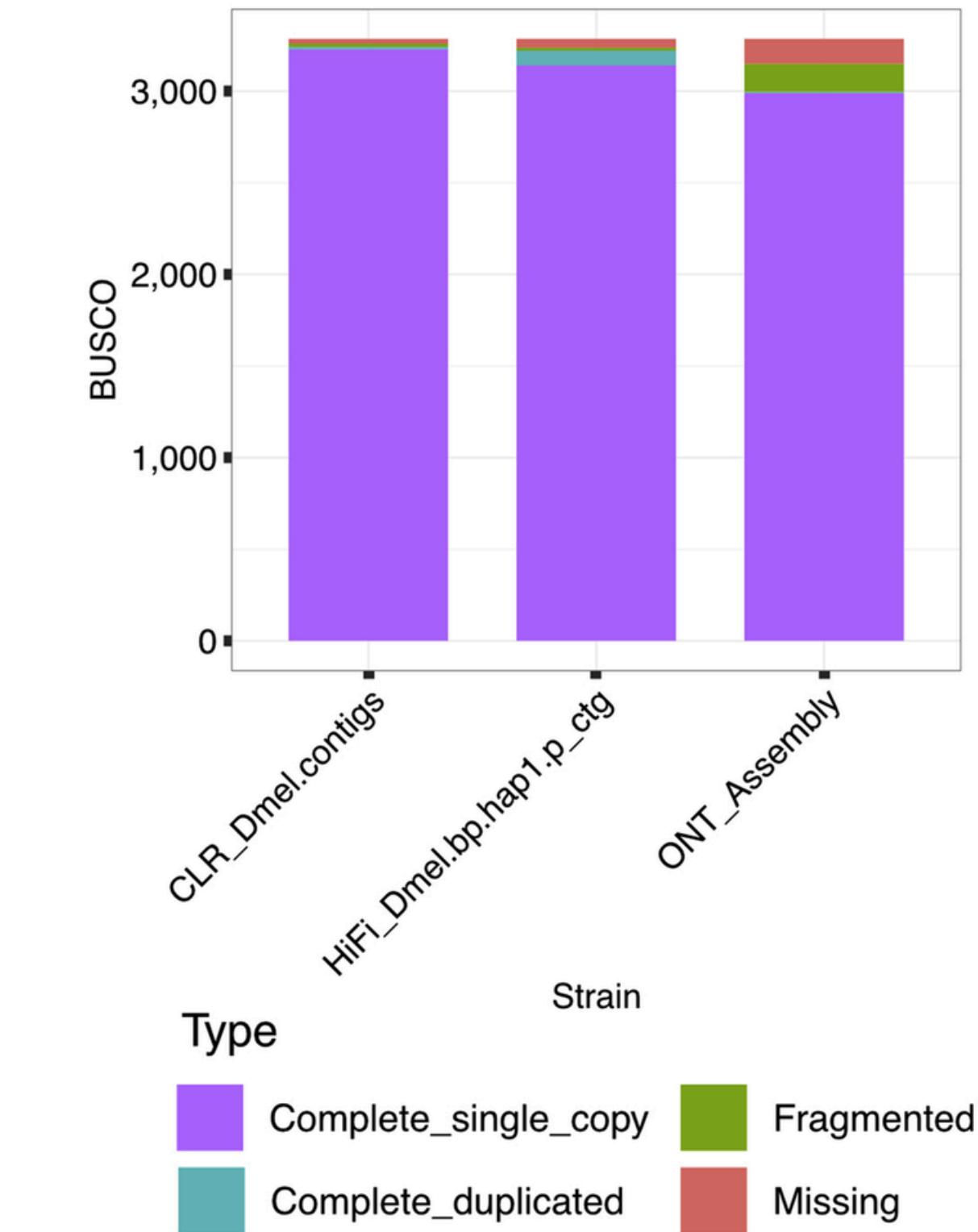
Line

- CLR_Dmel
- HiFi_Dmel
- ONT_Dmel

Type

— contig

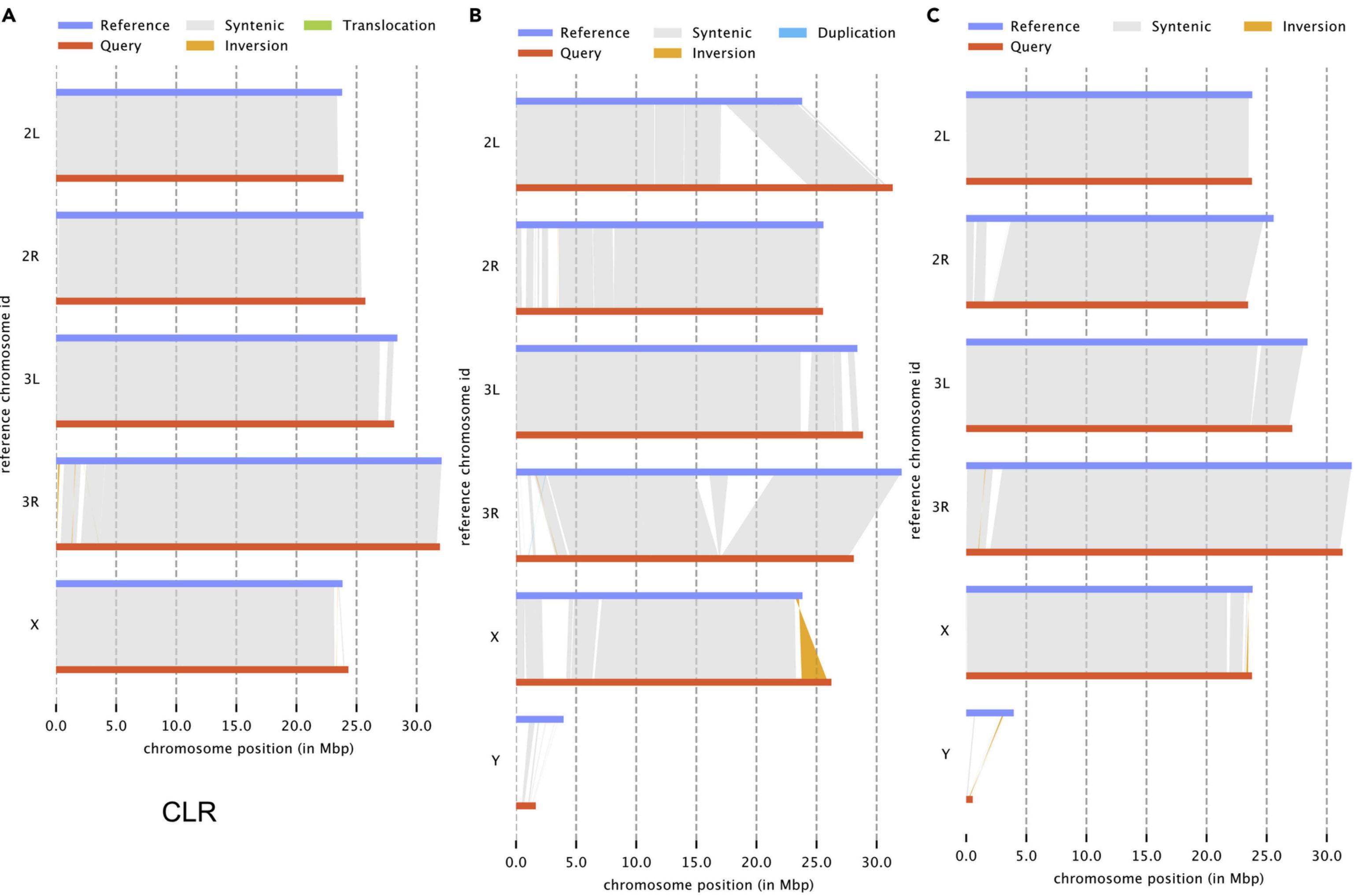
C



De novo genome assembly read-length distribution and quality assessment

- Read-length distributions of the three publicly available datasets used in this study. Each vertical dotted line represents the mean value of each dataset.
- Cumulative coverage plot for the contig/scaffold length.
- BUSCO analysis used to determine the number of single-copy orthologs known in a lineage.

Long reads Structure variant



Output summary statistic

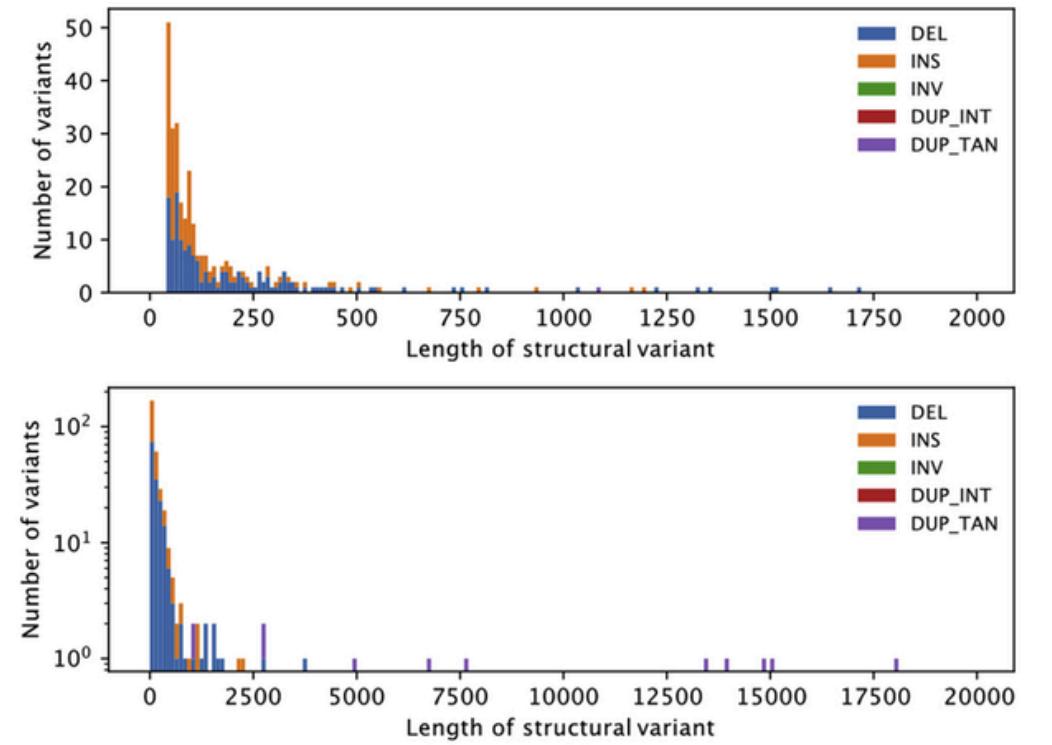
Table 1. Output summary statistics for three different long-read sequencing platforms used in this paper

	PacBio CLR	PacBioHiFi	ONT
Number of reads	1,437,524	2,301,518	640,215
Read max (bp)	99,345	26,462	417,450
Read N50 (bp)	13,094	11,151	21,491
Read min (bp)	50	369	1
Total read length (bp)	12,016,661,679	25,600,110,705	7,133,020,037
Number of contigs	452	654	208
Contig max (bp)	23,607,911	24,502,687	27,938,801
Contig N50 (bp)	9,177,974	4,127,200	18,567,724
Contig min (bp)	1,381	9,867	21
Total contig length (bp)	141,740,149	168,692,738	133,002,022
Number of placed contigs	169	175	99
Length of placed contigs (bp)	134,417,363	142,931,941	130,673,188
Number of unplaced contigs	283	479	109
Length of unplaced contigs (bp)	7,322,786	25,760,797	2,328,834

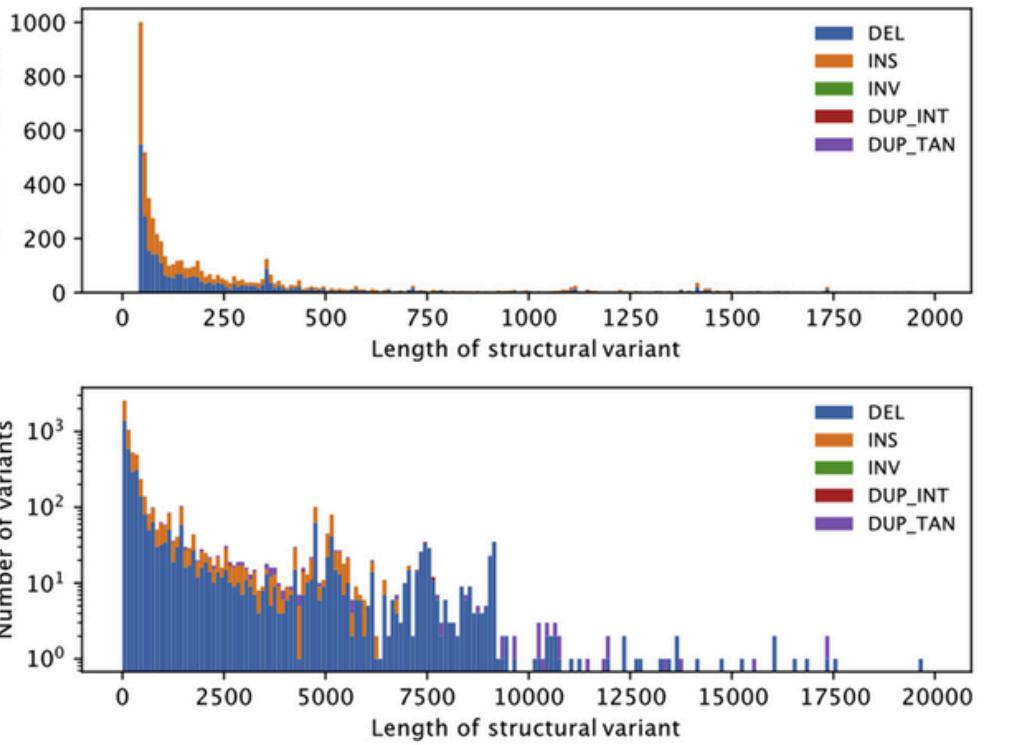
Long reads Structure variant

A

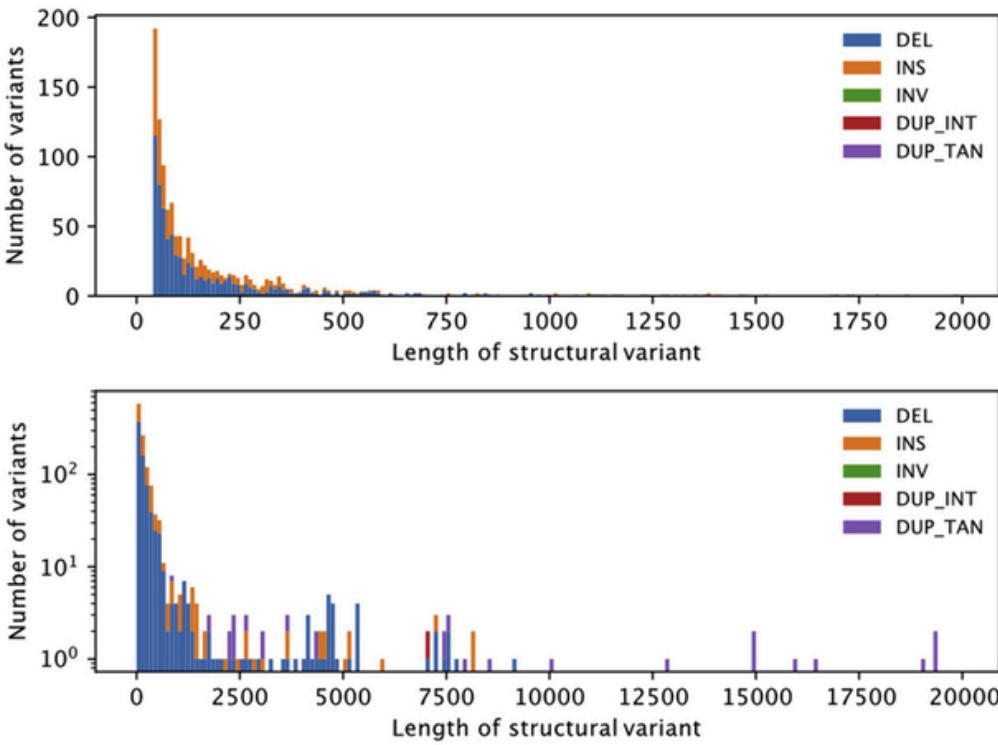
CLR



HiFi

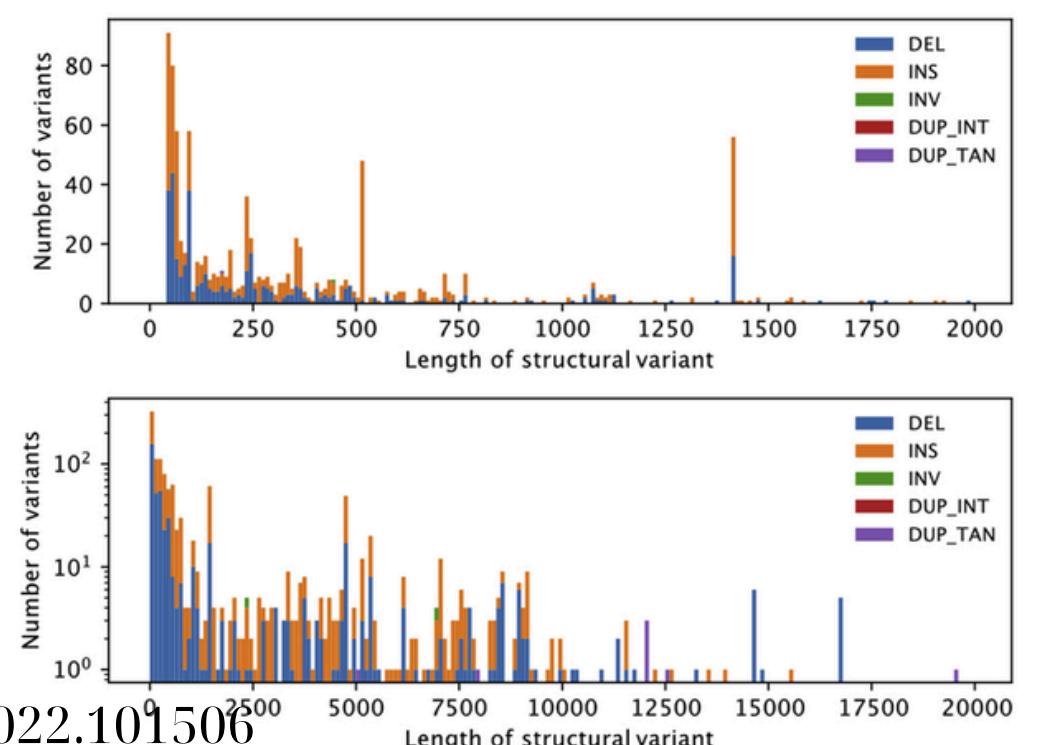


ONT

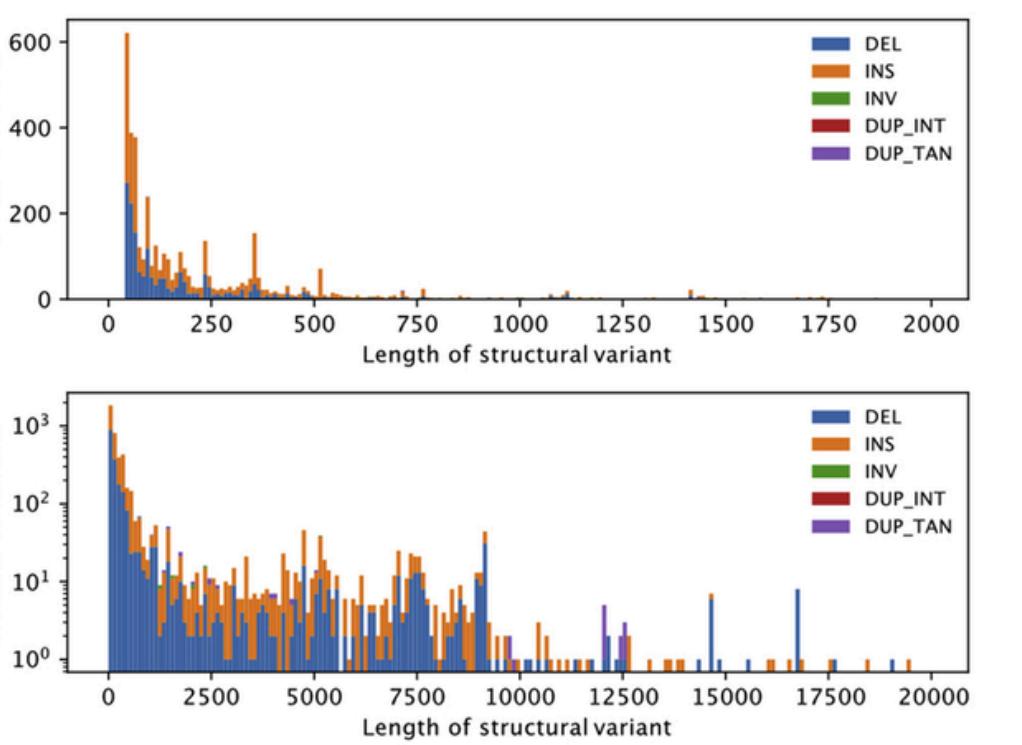


B

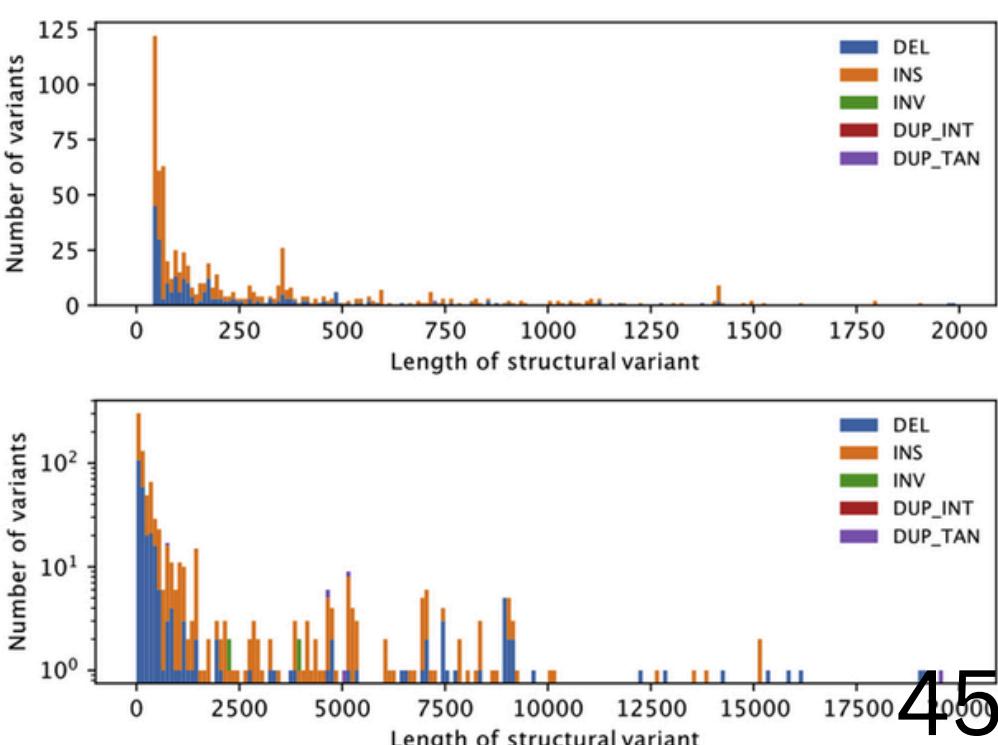
CLR



HiFi



ONT



SV calling output summary

(A) Read-based SV calling with SVIM.

(B) Assembly-based SV calling with SVIM-asm.

Summary

I. Overview about ONT for Virus

1. ONT basecalling
2. Bioinformatic software for long read sequencing
3. 1000 genomes project - population genomic
4. PCR tilling for Covid 19
5. Real-time and onsite characteristic application

II. Practice

1. Mapping strategy
2. De novo assembly strategy

THANK FOR WATCHING