Optimized protocol for sequencing the SARS-CoV-2 genome with PCR enrichment

Introduction

The novel coronavirus SARS-CoV-2 continues to be one of the biggest global public health concerns of the modern era. The virus can cause severe respiratory illness with a wide range of symptoms. The onset of disease can be severe in older individuals, particularly those with underlying medical issues [1]. The scientific community is researching this threat to better understand the virus and its evolution and spread.

Faster and more sensitive analysis of mutations in the SARS-CoV-2 genome can enable tracking of global transmission and provide a better understanding of infection mechanisms. Sequencing libraries enriched for SARS-CoV-2 genetic material allow deeper sequence coverage and are valuable when single-nucleotide variants need to be identified. Targeted amplification is an enrichment approach in which PCR is used to generate amplicons to tile the entire SARS-CoV-2 genome prior to preparing the sequencing library [1]. This application note describes a workflow for preparing a next-generation sequencing (NGS) library from samples that contain SARS-CoV-2—from total RNA purification to library quantification and amplicon sequencing—using the Invitrogen™ Collibri™ ES DNA Library Prep Kit for Illumina™ Systems (Figure 1).

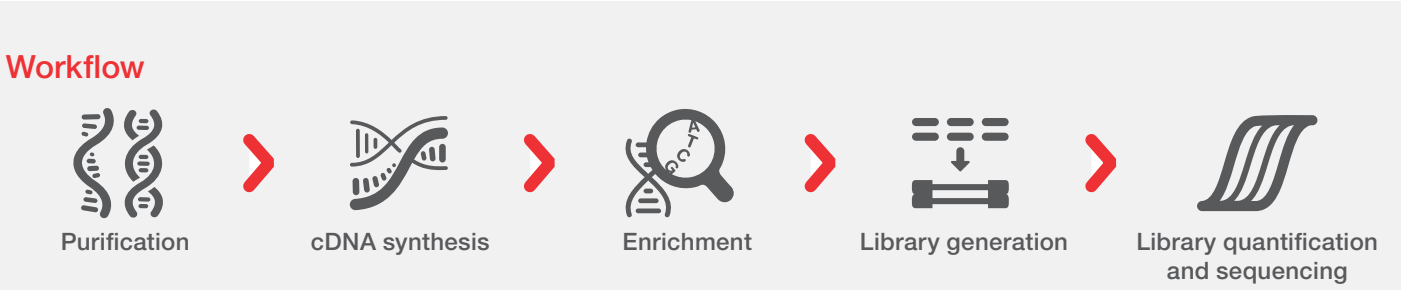


Figure 1. NGS workflow for the analysis of SARS-CoV-2 samples using a targeted amplicon-based approach.

Materials and methods

Samples

Samples were obtained with consent from the Santara Clinics Biobank in Lithuania. Each sample represents a unique donor. C_t values confirmed the presence of SARS-CoV-2 in the biobank samples. Sequencing was performed on six separate clinical sample collections. The quality and C_t values of the six collections were within the acceptable ranges for further analysis.

Purification of total RNA

Total RNA was extracted from clinical bronchoalveolar lavage (BAL) specimens. The BAL samples were lysed, and the RNA was purified using the Applied Biosystems™ MagMAX™ Viral/Pathogen Nucleic Acid Isolation Kit according to the standard protocol.

cDNA synthesis

The RNA in each 10 µL sample was reverse-transcribed into cDNA using the Invitrogen™ SuperScript™ IV cDNA Synthesis Kit along with Invitrogen™ RNaseOUT™ Recombinant Ribonuclease Inhibitor, Invitrogen™ 10 mM dNTP Mix, and Thermo Scientific™ Random Hexamer Primer according to the standard protocols for the kits.

PCR enrichment

The following workflow was optimized for studying coronaviruses based on the amplicon sequencing protocol initially created by the ARTIC Network [2]. PCR enrichment required Invitrogen™ Platinum™ SuperFi™ U Multiplex Master Mix, which was developed for highly multiplexed PCR, and ARTIC V.3 primer pools (Table 1). The optimized protocol for SARS-CoV-2 sequencing with ARTIC V.3 primers and an Applied Biosystems™ ProFlex™ PCR System or an equivalent PCR system is outlined in Table 2.

Table 1. PCR setup for enrichment.

Component	Final concentration (amount)	Volume
4X Platinum™ SuperFi™ U Multiplex PCR Master Mix	1X	6.25 µL
Forward primers	0.1–0.3 µM total	X µL
Reverse primers	0.1–0.3 µM total	X µL
Template DNA or RT reaction	0.1–10 ng or 2.5 µL (10% volume)	X µL
Water, nuclease free	—	Up to 25 µL
Total volume		**25 µL**

Table 2. PCR enrichment protocol for SARS-CoV-2 sequencing.

Cycle step	Temp.	Time	Cycles
Initial denaturation	98°C	30 sec	1
Denaturation	98°C	15 sec	30–35
Annealing/extension	63–65°C*	5 min	30–35
Hold	4°C	hold	1

* We strongly recommend performing a gradient PCR with your thermal cycler to determine the optimal annealing temperature. Subtle differences in thermal cycler calibration can cause specific amplicons to drop out.

Library generation

The cDNA samples were converted to sequencing-ready libraries through enzymatic fragmentation using the Collibri ES DNA Library Prep Kit for Illumina Systems. Using combinatorial dual (CD) and unique dual (UD) indexes together enabled higher sequencing throughput. See Table 2 for additional recommendations.

Table 3. Recommendations for high- and medium-throughput and alternative combinations.

Throughput	Comment
192 preps	For the highest throughput, use CD (Cat. No. A38607096) and UD (Cat. No. A38607196) indexes together
120 preps	Medium throughput can be achieved using CD (Cat. No. A38605024) and UD (Cat. No. A38607196), or any UD set of 24 preps with A38607096 (CD)

Note: Do not pool together differently sized kits that contain the same types of indexed adaptors (CDs or UDs only).

Note: Kits A38606024, A43605024, A43606024, and A43607024 (24 preps) together are equivalent to kit A38607196 (96 preps). They can be used interchangeably during periods of high demand.

Materials and methods (continued)

The standard protocol for the Collibri ES DNA Library Prep Kit for Illumina Systems was used with the modifications summarized in Table 3.

Table 3. Recommended protocol optimization for library generation with the Collibri ES DNA Library Prep Kit for Illumina Systems.

<table>
<tr><th>Step</th><th colspan="2">Standard recommendation</th><th colspan="2">Recommended adaptations for SARS-CoV-2 samples</th></tr>
<tr><td>1. Remove EDTA from DNA samples</td><td colspan="2">(If needed)
Input: 1–500 ng</td><td colspan="2">Input: amplified pools after PCR enrichment</td></tr>
<tr><td>2. Fragment the DNA and add dA tails</td><td colspan="2">Assemble the fragmentation and dA-tailing reaction for each DNA sample in a sterile 0.2 mL thin-wall PCR tube on ice or a cooling rack. Add the reagents in the order shown.
Component | Volume
10 mM Tris-HCl, pH 7.5–8.5 | To 40 µL
Double-stranded DNA (1–500 ng) | X µL
10X fragmentation and dA-tailing buffer (blue) | 5 µL
Total volume (light blue mixture) | 40 µL
Recommended fragmentation time and optimization range to obtain the desired fragment size.
Fragment size | Fragmentation time at 37°C — Recommendation | Optimization range
150–300 bp | 20 min | 20–30 min
300–500 bp | 10 min | 10–20 min
500–700 bp | 5 min | 5–10 min</td><td colspan="2">Component | Volume
Equally combined amplicon pools (total: 50 ng) | 40 µL
10 mM Tris-HCl, pH 7.5–8.5 | X µL
10X fragmentation and dA-tailing buffer (blue) | 5 µL
Total volume (light blue mixture) | 45 µL
Fragment for 15 minutes at 37°C. Seal the plate and incubate the mixture in a thermal cycler programmed as outlined below with the heated lid set to 80–85°C and the block pre-cooled to 4°C.
Step | Temperature | Time
Pre-cool block | 4°C | As required
Fragmentation | 37°C | 15 min
dA-tailing | 65°C | 10 min
Hold | 4°C | Hold</td></tr>
</table>

Library quantification

The final concentrations of the sequencing libraries were determined using the Invitrogen™ Collibri™ Library Quantification Kit.

Sequencing and data analysis

The libraries were sequenced on an Illumina™ MiSeq™ System via 2 x 150 bp paired-end sequencing using the Illumina™ PhiX Control v3 Library (10% v/v) as a sequencing control. Alignment was performed using the BWA-MEM algorithm included in the Burrows-Wheeler alignment tool (bwa mem -t 32 -r 1.0 -k 19 -M -B 6 -v 1). The mean coverage, or the average number of reads that aligned with the reference NCBI ASM985889v3 genome, was calculated using the QualiMap Bam QC package.

Results

A total of 1,001 clinical samples were positive for SARS-CoV-2 and sequenced using the amplicon-based workflow solution. As expected, the quality of individual samples in each collection (Table 4) corresponded to the percentage of aligned reads, which provided sufficient coverage for alignment with the reference genome.

The heatmap of sequencing depth for all 1,001 positive samples is shown in Figure 2. The mean percentage of bases not called in raw reads across the length of the 30 kb virus genome was less than 1%. This meant the amplicon-based workflow for the Collibri ES DNA Library Prep Kit for Illumina Systems nearly eliminated the potential for amplicon failure. The mean coverage at different depths across the six clinical sample collections was of high quality, and the robust results could be used to identify SARS-CoV-2 strains in individual samples.

Table 4. C_t values of the six collections analyzed in this study.

Value	Collection 1	Collection 2	Collection 3	Collection 4	Collection 5	Collection 6
Maximum	31.86	38.42	37.25	37.78	32.49	39.61
Minimum	14.37	12.54	12.11	10.13	12.91	10.31
Mean	23.94	23.57	23.84	23.07	23.16	23.18
Median	24.95	23.71	23.50	23.53	22.95	23.61

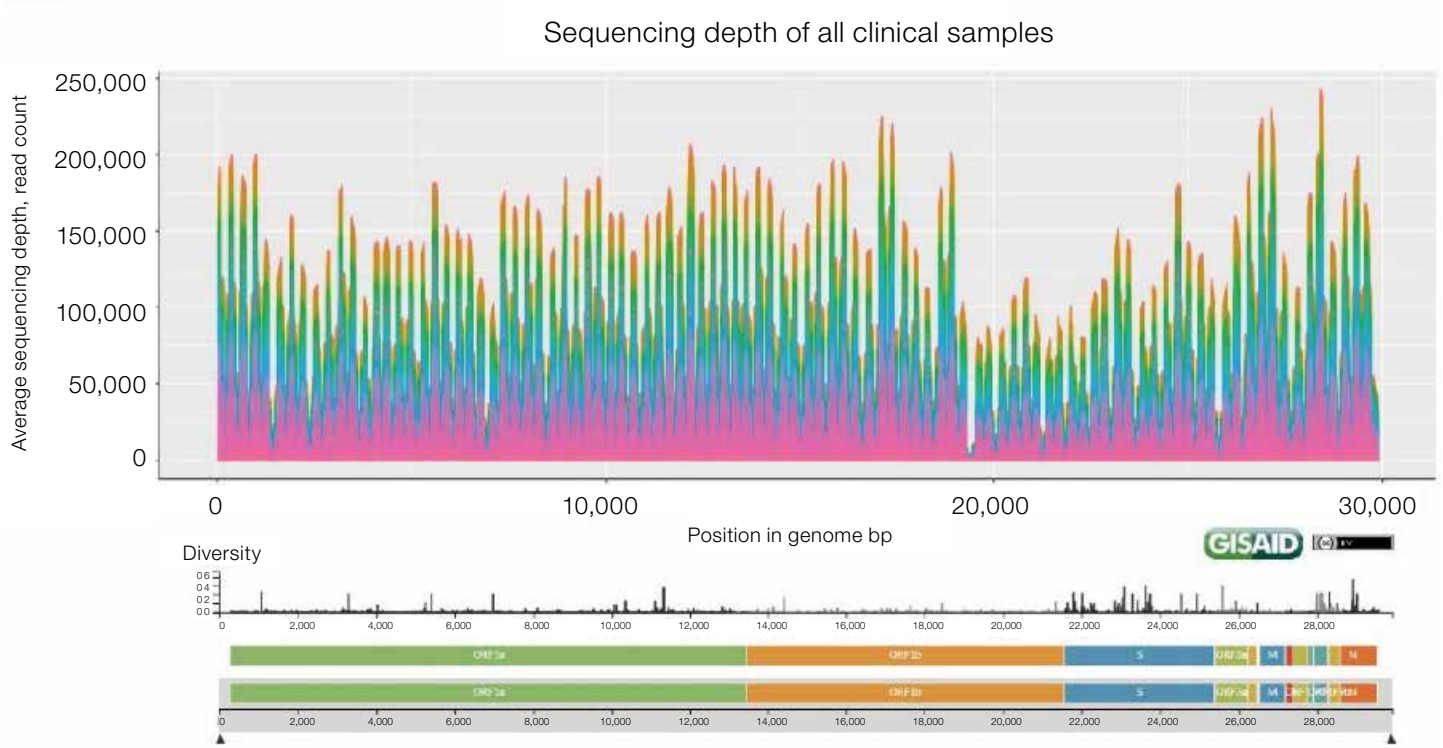
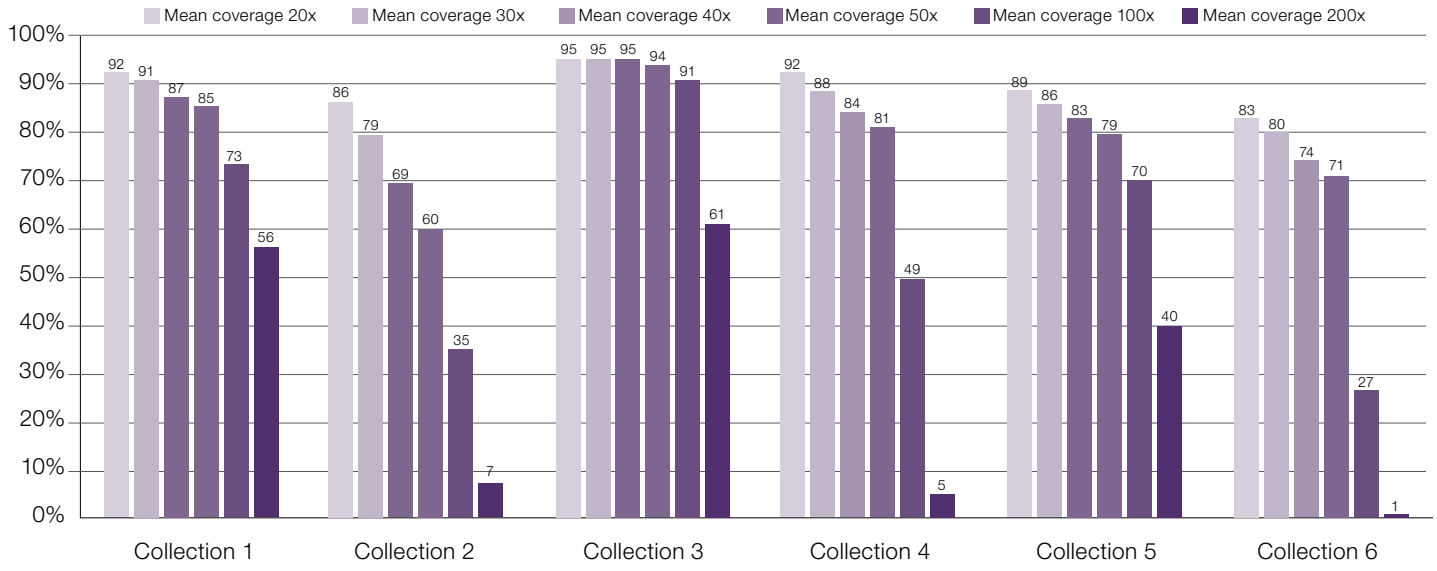


Figure 2. Heatmap of sequencing depth of 1,001 SARS-CoV-2 positive clinical samples across the NCBI ASM985889v3 genome. The heatmap is compared to a snapshot of the positions and frequencies of known mutations from the GISAID database [3].

For example, 57% of the reads across all six collections had a mean coverage of 100x with less than 100,000 reads (Figure 3). The mean alignment rate for a collection corresponded to the quality of samples in that collection.

The results also indicated that the clinical C_t values correlated with expected sequencing coverage. The association between the C_t values and genome coverage is illustrated in Figure 4. There was a clear cutoff near C_t of 30, above which coverage was more likely to drop.

Although C_t is an excellent indicator of sequencing coverage, it is not absolute. Other factors influence sequencing coverage heavily, such as sample storage conditions and subsequent degradation during handling. This means it is possible to get good coverage of a viral genome with a C_t of 35 and inadequate coverage with a C_t of 15. However, consistently good coverage is more likely to be obtained using samples with a C_t below 30.



	Collection 1	Collection 2	Collection 3	Collection 4	Collection 5	Collection 6
Number of samples	93	192	192	192	140	192
Mean aligned reads (%)	97.3	70.2	85.9	85.6	85.6	86.4
Mean aligned reads (counts)	134,623	26,335	91,653	60,920	60,920	91,653

Figure 3. Percentage of samples with different depths of coverage within each of six collections. A total of 1,001 SARS-CoV-2 positive clinical samples were sequenced and analyzed.

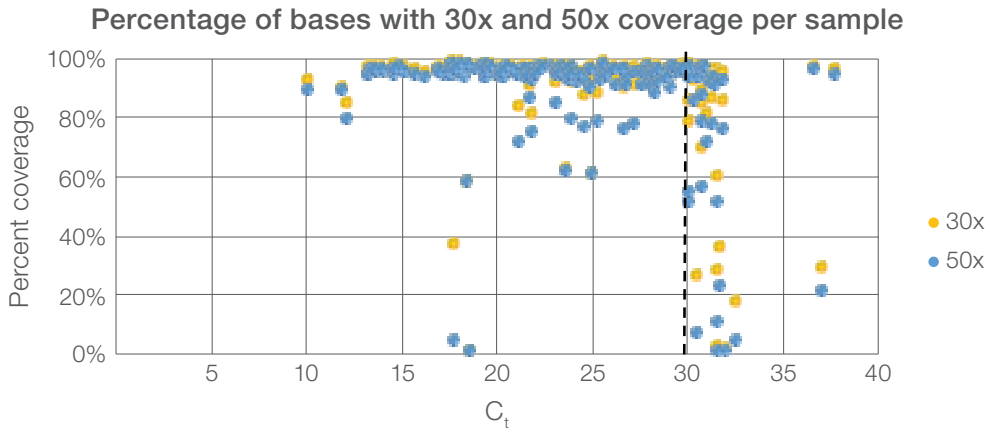


Figure 4. Decrease in genome coverage for samples (Collection 4) with C_t > 30.

invitrogen

Conclusion

The amplicon-based approach to NGS allows robust SARS-CoV-2 sequencing and analysis to identify new mutations and viral strains. The workflow used with the Collibri ES DNA Library Prep Kit for Illumina Systems minimizes the potential occurrence of amplicon failure.

References

1. World Health Organization, Coronavirus, https://www.who.int/health-topics/coronavirus#tab=tab_1 (accessed 4/15/2021)
2. ARTIC Network, https://artic.network/ncov-2019 (accessed 4/15/2021)
3. GISAID, https://www.gisaid.org/phylodynamics/global/nextstrain/ (accessed 4/15/2021)

Ordering information

Step	Kit and link to user guide	Quantity	Cat. No.
Key component			
Purification of total RNA*	MagMAX Viral/Pathogen Nucleic Acid Isolation Kit	Up to 2,000 preps	A48310
		Up to 200 preps	A42352
cDNA synthesis	SuperScript IV Reverse Transcriptase	10,000 units	18090050
		4 x 10,000 units	18090200
		2,000 units	18090010
	RNaseOUT Recombinant Ribonuclease Inhibitor	5,000 units	10777019
	dNTP Mix (10 mM each)	5 x 1 mL	18427089
		1 mL	18427088
		100 µL	18427013
	Random Hexamer Primer	120 µL	SO142
PCR enrichment	ARTIC v3 Primer Pools	Choose a provider	Choose a provider
	Platinum SuperFi U Multiplex Master Mix	24 preps	A5140024
		96 preps	A5140096
Library generation	Collibri ES DNA Library Prep Kit for Illumina Systems with Combinatorial Dual (CD) Indexes	24 preps	A38605024
		96 preps	A38607096
	Collibri ES DNA Library Prep Kit for Illumina Systems with Unique Dual (UD) Indexes	24 preps	A38606024 (Set A, 1–24)
			A43605024 (Set B, 25–48)
			A43606024 (Set C, 49–72)
			A43607024 (Set D, 73–96)
		96 preps	A38607196 (Set A–D, 1–96)
Library quantification**	Collibri Library Quantification Kit (preferred)	100 reactions	A38524100
		500 reactions	A38524500
	Qubit dsDNA HS Assay Kit	100 assays	Q32851
		500 assays	Q32854

* RNA can be purified using either the MagMAX Viral/Pathogen Nucleic Acid Isolation Kit or the MagMAX Viral/Pathogen II Nucleic Acid Isolation Kit. Optional protocol modifications to reduce sample and reagent use are described in the application note "MagMAX Viral/Pathogen kit protocol changes to enable increased SARS-CoV-2 testing throughput".

** Library quantification can be performed using the Collibri Library Quantification Kit (qPCR assay) or the Qubit dsDNA BR Assay Kit (fluorometric assay).

Find out more at **thermofisher.com/sarscov2-collibri**

ThermoFisher SCIENTIFIC