Rapid Library Preparation for RNA Virus Sequencing Utilizing Long cDNA Synthesis and Long-Range HiFi PCR

Keywords: RNA virus sequencing, long cDNA synthesis, long range HiFi amplification, qScript® Ultra Flex Kit, repliQa® HiFi ToughMix®, sparQ PureMag Beads, sparQ DNA Frag & Library Prep Kit

ABSTRACT

RNA viruses are often pathogenic and can cause disease in humans, animals and other host organisms. Sequencing of the viral genome is critical to characterize viruses, survey strain circulation, and detect variants. cDNA synthesis and subsequent amplification of the viral targets are essential steps for selectively enriching the virus genome for downstream sequencing applications. There are several methods available in the market that utilize small amplicon panels limited to specific viral targets. In this application note, we demonstrate a method for viral whole genome sequencing utilizing rapid long cDNA synthesis using qScript® Ultra Flex Kit, followed by long-range PCR using repliQa® HiFi ToughMix® to generate long PCR products that can be a substrate for a short read, long read, and other sequencing methods.

INTRODUCTION

Viruses can have a significant impact in humans and other living organisms by causing illness. RNA viruses are more likely to cause human disease compared to DNA viruses, which is attributed to their high mutation rates and their low-fidelity RNA polymerase. RNA viruses are highly diverse, making them responsible for 44% of all emerging infectious diseases.[1] Viruses such as H1N1 influenza, Ebola, Zika and SARS-nCoV-2 are responsible for recent pandemics. Rapid and accurate RNA sequencing is crucial to better understand the pathogenicity, epidemiology, and infection control of these emergent viruses.

Next-generation sequencing is a powerful tool for the identification and characterization of viruses.[2] One of the main methods for viral genome sequencing is PCR amplicon sequencing. This well-established method is highly specific, sensitive, and cost-effective offering deep coverage even at low viral load.[3] However, PCR-based sequencing tools often target specific genomic regions, which can limit their ability to sequence large genomes, potentially missing insertions, deletions, and novel regions.

The first step in many RNA sequencing methods is conversion into cDNA using a reverse transcriptase (RT). Despite its pivotal role, many commercially available RTs have a limited ability to generate long cDNA products because of low processivity and poor activity at temperatures sufficient to relax complex secondary structures often adopted by RNA viruses. The recently developed qScript Ultra RT overcomes these challenges with higher processivity and a higher cDNA synthesis temperature, allowing for long cDNA synthesis from challenging templates. Using this novel enzyme, we developed a simple and robust methodology for whole viral genome sequencing based on long cDNA synthesis and long-range PCR.

In this study, we used RNA from the large OC43 coronavirus and SARS-CoV-2 as a model system. Gene-specific primers were used for long cDNA synthesis using qScript Ultra Flex Kit (Quantabio) followed by long amplicon (~10 kb) generation using repliQa HiFi ToughMix (Quantabio). Finally, NGS libraries were generated from long amplicons and sequenced on the Illumina platform. Overall, longer and faster cDNA synthesis followed by rapid, long-range, high-fidelity PCR, provides a reliable method for monitoring viral evolution, mutation surveillance, and pathogen detection.

MATERIALS AND METHODS

Samples

Human coronavirus OC43 (VR-1558D™, ATCC) and SARS-CoV2 novel coronavirus strain 2019-nCoV/USA-WA1/2020 RNA (2019-nCoV/USA-WA1/2020, ATCC) were used as the template for cDNA synthesis with a final input of 1 µL per reaction.

Long cDNA synthesis using qScript Ultra Flex Kit

1. Gene-specific primers (GSP) were designed for the first-strand cDNA synthesis from the genomes of human coronavirus OC43 and SARS CoV-2 strains (Table 1).

Target	Primer name	Sequence 5' – 3'
2019-nCoV/ USA-WA1/2020	WA1-8349 REV	CTACAGTCAATACAAGCAC-CAAGGTCA
	WA1-16401 REV	ACAACCTGGAGCATTGCAAA-CATACGG
	WA1-24221 REV	TCCAACCAGAAGTGATTGTAC-CCGCTA
	WA1-29801 REV	GCTCTTCCATATAGGCAGCT-CTCCC
HCoV-OC43	OC43-12174 REV	GAGAGCCAAATCAGCCA-TACGCTCCA
	OC43-21745 REV	AATTTAATCTGAAAACCATCA-CCCCAA
	OC43-30683 REV	TGGTAACTTAACATGCTGGCT-CTTCCC

Table 1 Gene specific primers used for first-strand cDNA synthesis using qScript Ultra Flex Kit.

2. The qScript Ultra Flex master mix was prepared on ice according to Table 2.

Component	Volume for 1 sample (µL)
Nuclease-free Water	12
Template RNA	1
10 µM Gene-specific primer mix	1
10x GSP Enhancer	2
Final volume	16

Table 2 Reaction setup for first-strand cDNA synthesis using qScript Ultra Flex Kit.

3. The mix was incubated for 5 min at 65°C and cooled on an ice block.

4. Then 4 µL of 5x qScript Ultra Reaction Mix was added.

5. The first-strand cDNA reaction was run following the cycling program detailed in Table 3.

Step	Temperature	Incubation time
1	55°C	10 min
2	85°C	5 min

Table 3 Cycling program for first-strand cDNA synthesis using qScript Ultra Flex Kit.

6. After the completion of cDNA synthesis, 2 µl of the reaction was taken for PCR amplification.

Long amplicon generation using repliQa HiFi ToughMix

7. Primer sets were picked to amplify specific regions of the human coronavirus OC43 and 2019-nCoV/USA-WA1/2020 genomes. These region-specific targets were between 8.1 kb and 11.9 kb (Table 4).

Target	Primer name	Sequence 5' – 3'	Amplicon length (bp)
2019-nCoV/USA-WA1/2020	WA1-75 FWD	CTTTAAAATCTGTGT-GGCTGTCACTCG	8275
	WA1-8349 REV	CTACAGTCAATACA-AGCACCAAGGTCA	
	WA1-8159 FWD	ACTTTTATTTCAGCA-GCTCGGCAAGGG	8243
	WA1-16401 REV	ACAACCTGGAGCAT-TGCAAACATACGG	
	WA1-16191 FWD	ACCTGAGTTTTATGA-GGCTATGTACACACCG	8031
	WA1-24221 REV	TCCAACCAGAAGTGA-TTGTACCCGCTA	
	WA1-22771 FWD	AGGTGATGAAGTCAG-ACAAATCGCTCC	7031
	WA1-29801 REV	GCTCTTCCATATAGGC-AGCTCTCCC	
HCoV-OC43	OC43-179 FWD	CCCGCCCATAGGTCA-CAATGTCGAAG	11888
	OC43-12066 REV	ATTAGCAGAACCACT-AAAACGCGCCTC	
	OC43-10201 FWD	ACAGTGACCCTGCA-AAATTCTCGTACGC	10881
	OC43-21081 REV	CACTGCCACCCAGA-GCTAACTTGTCTCG	
	OC43-19624 FWD	AATTATTTGCCAAGC-GCAGTGTTCGAC	10901
	OC43-30524 REV	CCATTCTGATAGAGA-GTGCCTTATCTCGACT	

Table 4 Primer sequences used to amplify HCoV-OC43 and 2019-nCoV/ USAWA1/2020 targets using repliQa HiFi ToughMix.

8. The repliQa HiFi master mix was prepared on ice according to Table 5.

Component	Volume for 1 sample (µL)	Final concentration
repliQa HiFi ToughMix (2X)	10	1X
10 µM Primer Mix	1	500 nM
Nuclease-free water	7	–
Long cDNA template	2	–
Final volume	20	

Table 5 Reaction setup for PCR amplification using repliQa HiFi ToughMix.

9. The PCR was run according to the cycling programs detailed in Table 6.

Step	Temperature	Incubation time	Cycles
Initial Denaturation	98°C	30 sec	1
Denaturation	98°C	10 sec	35
Extension	68°C	45 sec for CoV-2 targets 2 min for OC43 targets	
Hold	4°C	Hold	1

Table 6 2-Step cycling program for PCR amplification using repliQa HiFi ToughMix.

10. Length and purity of the amplicons were checked using Genomic DNA ScreenTape® on the 4200 TapeStation® System (Agilent).

11. Long amplicons were bead purified using 0.45X sparQ PureMag Beads according to manufacturer recommendations and resuspended in 20 µL of 10 mM Tris-HCl, pH 8.0.

DNA library preparation

12. Libraries were prepared using 100 ng PCR products from step 10 and the sparQ DNA Frag & Library Prep Kit following the product manual (IFU-122.1_REV_08_95194_sparQ_DNA_Frag_Library_Prep_Kit_0624).

13. DNA samples were fragmented for 9 min at 32°C.

14. Libraries were prepared by ligating 250 nM sparQ UDI adapters and amplified for 5 PCR cycles.

15. Libraries were then purified using sparQ PureMag Beads and eluted in 30 µL elution buffer (10 mM Tris-HCl, pH 8.0).

Size Verification

16. The size of the libraries was verified using a D5000 ScreenTape on the 4200 TapeStation System. The expected average size of the library was around ~450 bp.

Quantification

17. Libraries were quantified using Qubit® dsDNA BR Assay Kit (Thermo Fisher Scientific) using manufacturer's recommendations. The tubes were incubated at room temperature for 2 min before loading the tubes into a Qubit 3.0 Fluorometer.

Sequencing and data analysis

18. Libraries were normalized to 4 nM stock. For sequencing, a 10 pM library pool was loaded into an Illumina® MiSeq® V3 kit, and libraries were sequenced using 2x150 bp reads on the MiSeq V3 platform.

19. FastQ files generated by MiSeq were analyzed using CLC Genomics Workbench (QIAGEN).

RESULTS

Long cDNA and long-range PCR using viral RNA targets

Next-generation sequencing is a powerful tool for characterizing novel viral infections, conducting strain surveillance, and detecting variants. The workflow for RNA virus sequencing begins with cDNA generation from RNA templates, followed by amplicon generation using HiFi PCR and subsequent library preparation. Most well-known RNA virus genomes are approximately 10 kb in size, such as HIV (9.2 kb), HPV (8 kb), HCV (9.6 kb), ▶

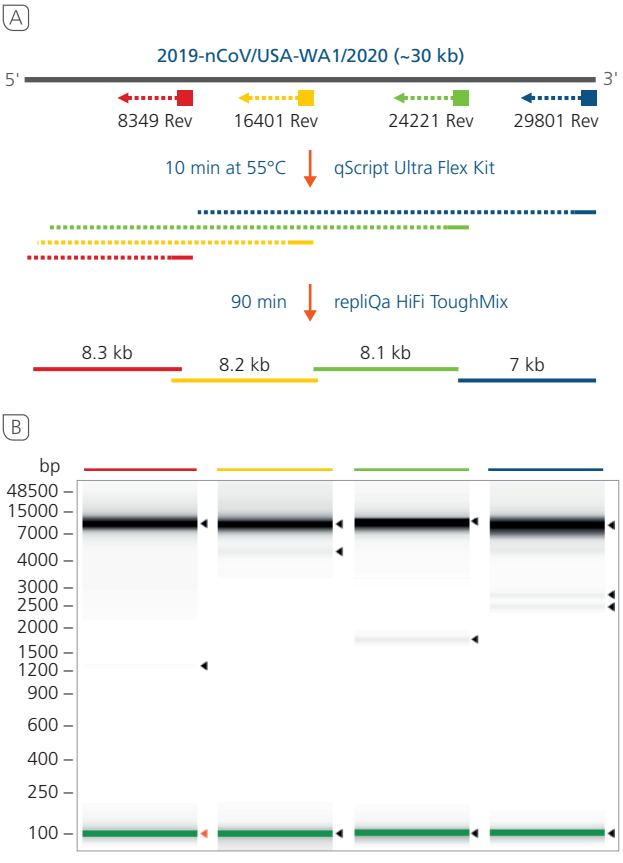


Figure 1 Long amplicon generation from novel coronavirus 2019-nCoV/USA-WA1/2020 using qScript Ultra Flex Kit and repliQa HiFi ToughMix.
A Schematic workflow showing individual cDNA synthesis and PCR reactions.
B Genomic TapeStation analysis of the long amplicon.

H1N1 (13.5 kb), and Norovirus (7.5 kb). In contrast, Nidoviruses which includes coronaviruses, have larger genomes ranging from 26 to 32 kb. Most commercially available reverse transcriptases can only generate cDNA of lengths between 10 and 20 kb. Additionally, the success of long cDNA generation is often influenced by the integrity and quality of the input RNA.

We used novel coronavirus 2019-nCoV/USA-WA1/2020 RNA (~30 kb) as a template to generate cDNA using qScript Ultra Flex Kit. Four gene-specific primers (WA1-8349 REV, WA1-16401 REV, WA1-24221 REV, and WA1-29801 REV) were used for four independent cDNA reactions by incubating at 55°C for 10 min. Then, repliQa HiFi ToughMix was used for the PCR amplification using the respective primer sets for each specific first-strand reaction following a 2-step cycling program (Figure 1 A). Genomic TapeStation analysis confirmed the sizes of the four DNA fragments as 8.3 kb (WA1-75-8349), 8.2 kb (WA1-8159-16401), 8.1 kb (WA1-16191-24221), and 7 kb (WA1-22771-29801) (Figure 1 B). This data clearly showed that qScript Ultra Flex Kit generated long cDNA in just 10 min and repliQa HiFi ToughMix generated 7– 8 kb amplicons in just 90 minutes. This workflow provides a fast and robust solution for amplicon generation for RNA viruses where most viral genomes can be covered by a single gene-specific primer and a single long amplicon.

Multiplex first-strand and PCR reaction

To reduce the number of cDNA reactions needed for each sample, we utilized two gene-specific primers for each cDNA reaction. As illustrated in Figure 2A, primers WA1-16401 REV and WA1-29801 REV were employed in the first cDNA reaction, while primers WA1- 8349 REV and WA1-24221 REV were used in the second cDNA reaction. For amplicon generation, we amplified WA1-8159-16401 and WA1-22771-29801 amplicons individually from the first cDNA products. Similarly, the second cDNA products were used to amplify WA1-75-8349 and WA1-16191-24221 amplicons separately. The presence of clear, single bands in the first four lanes of the TapeStation analysis indicates that the dual primer strategy was successful for long cDNA synthesis (Figure 2 B).

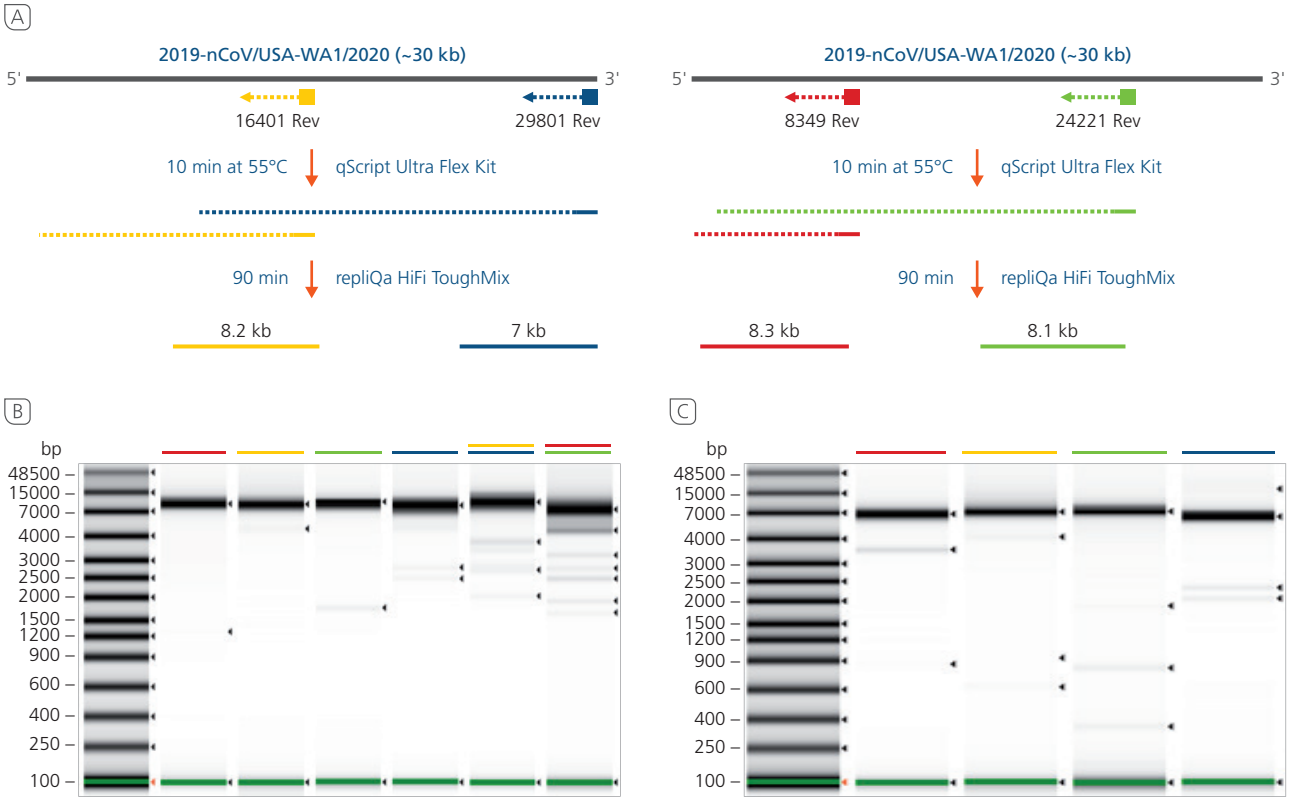


Figure 2 **A** A schematic representation illustrating the process of cDNA synthesis using two gene-specific primers per reaction, followed by either singleplex or duplex PCR to generate long amplicons. **B** TapeStation analysis using the Genomic DNA ScreenTape on a 4200 TapeStation. The first four lanes were produced using four singleplex PCR reactions, while the last two lanes were generated using duplex PCR reactions. **C** cDNA synthesis was conducted using dual gene-specific primers per reaction without the initial 65°C step. TapeStation analysis indicated that omitting this step had no effect on cDNA generation for this template.

Two different primer sets were also used for multiplex long-range PCR amplification. Both WA1-8159-16401 and WA1-22771-29801 amplicons were generated in a single reaction using the first cDNA products, and WA1-75-8349 and WA1-16191-24221 amplicons were generated in a single reaction using the second cDNA products, respectively. Multiplex PCR also generated clear lanes of the expected size (the last two lanes in Figure 2 B), suggesting multiplex PCR was successful. Overall, this multiplex cDNA and multiplex long-range PCR approach reduced the number of cDNA and PCR reactions by half, saving cost and improving the throughput of the reaction.

Effect of initial denaturation on the first strand reaction

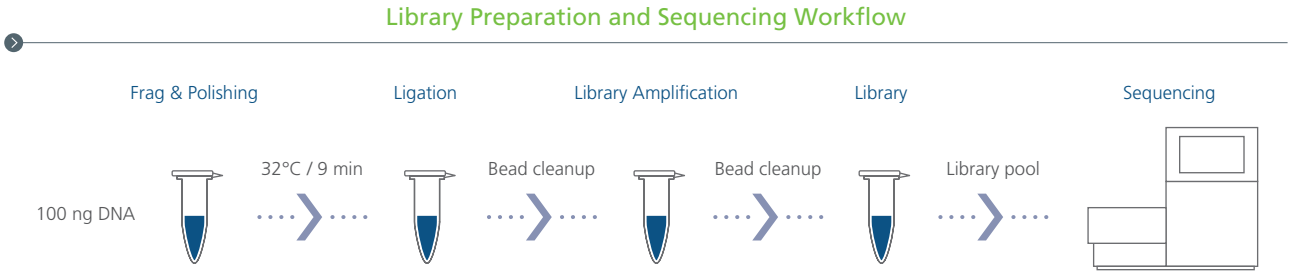
To improve the binding of gene-specific primers to the RNA template, we implemented a 65°C step prior to the cDNA synthesis reaction. Typically, this procedure involved mixing template RNA, nuclease-free water, gene-specific primers, and a GSP Enhancer, then incubating the mixture for 5 minutes at 65°C before cooling on ice. Next, we added the 5x qScript Ultra Reaction Mix to the mixture, and cDNA synthesis was conducted by incubating the reaction at 55°C for 10 minutes. To determine whether the 65°C step was necessary, we conducted cDNA synthesis using a dual primer strategy without this step.

In this case, all components for the cDNA reaction were mixed and incubated directly at 55°C for 10 minutes. As shown in Figure 2 C, we observed single bands from individual PCR reactions with similar DNA yields compared to the first four bands in Figure 2 B. This indicates that, for the 2019-nCoV/USA-WA1/2020 template, the 65°C step is not required. However, this additional step may still be beneficial for RNA templates with high GC content and stable secondary structures.

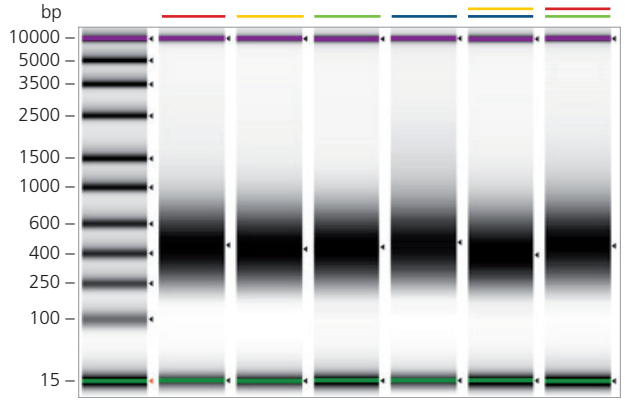
Library preparation from long amplicons

To sequence the long amplicons on the Illumina platform, 100 ng purified amplicons (as shown in Figure 2B: WA1-75-8349, WA1-8159-16401, WA1-16191-24221, WA1-22771-29801, WA1-75-8349 + WA1-16191-24221, WA1-8159-16401 + WA1-22771-29801) were used as input for DNA library preparation. Libraries were prepared using sparQ DNA Frag & Library Prep Kit by enzymatically fragmenting the DNA for nine minutes (Figure 3 A) to achieve a DNA fragment size larger than 300 bp to be sequenced in Illumina MiSeq platform using 2x150 bp reads. Amplified libraries were analyzed using a D5000 ScreenTape on a 4200 TapeStation, showing the median library size ranges from 420 bp to 460 bp (Figure 3 B). The Qubit quantification of the libraries revealed a total library yield of more than 500 ng (Table 3 C).

A



B



C

Library Yield (ng)	1020	984	828	738	744	690
Mapped reads (%)	99	98.9	98.3	97.8	98.7	97.4
Duplication rate (%)	3.6	3.6	3.3	4.6	3	3.5

Figure 3 **Library preparation from novel coronavirus 2019-nCoV/USA-WA1/2020 for Illumina sequencer.** A A schematic representation of the library preparation and sequencing workflow. B Analysis of the DNA libraries using D5000 ScreenTape on the 4200 TapeStation System. C Sequencing QC matrices showed a high percentage of mapped reads and low duplication rates after mapping.

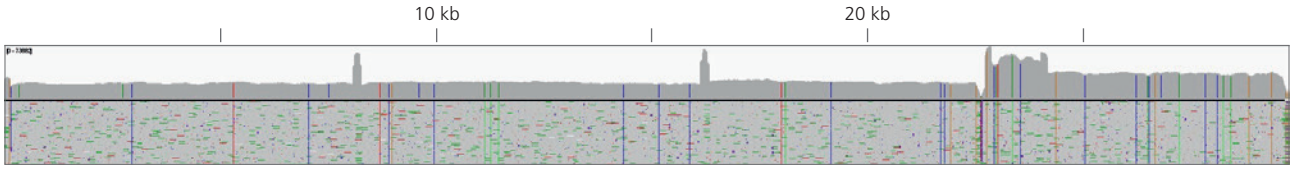


Figure 4 IGV view of mapped sequencing reads generated by Illumina MiSeq platform.

High-quality sequencing data and genomic coverage

FastQ files generated from Illumina MiSeq were analyzed using CLC Genomics Workbench. Library QC analysis showed a high percentage of mapped reads (>98%) and a low duplication rate (<5%) (Table 3 C), suggesting the libraries were high quality and predominantly represented 2019-nCoV/USA-WA1/2020 targets.

Sequencing reads from four targets (WA1-75-8349, WA1-8159-16401, WA1-16191-24221, WA1-22771-29801) were combined and visualized in IGV. An IGV image represented in Figure 4 showed uniform coverage from 5' to 3' on the target genome. In three locations, the coverage was double that of other regions, due to overlapping regions from two neighboring targets. This is essential to identify any mutation present at the primer binding sites.

cDNA synthesis of the entire 30 kb target using a single GSP

Most commercially available reverse transcriptases can generate cDNA up to 20 kb. The qScript Ultra Flex Kit has been developed to synthesize cDNA targets up to 20 kb in just 10 minutes. While most RNA targets fall below this length, some RNA targets such as coronaviruses can be around 30 kb. To address this challenge, we utilized multiple gene-specific primers (GSPs) for cDNA synthesis. In this experiment, we specifically tested the capability of Ultra RT to synthesize cDNA beyond 20 kb.

To ensure a high-quality integrated RNA template, we used human coronavirus OC43 RNA instead of SARS-CoV-2 RNA. For cDNA synthesis, we employed a single GSP (OC43-30683 Rev) and incubated the reaction at 55 °C for either 10 minutes or 30 minutes. To assess the length of the cDNA, we used three different primer sets to amplify targets OC43-179-12066, OC43-10201-21081, and OC43-19624-30524 with repliQa HiFi ToughMix. TapeStation analysis (Figure 5 B) clearly indicated that only the targets OC43-10201-21081 and OC43-19624-30524 were amplified when cDNA synthesis was performed for 10 minutes. However, when the synthesis time was extended to 30 minutes, all three targets were amplified. This data suggests that it is possible to synthesize cDNA beyond 20 kb using the qScript Ultra Flex Kit by extending the cDNA synthesis time.

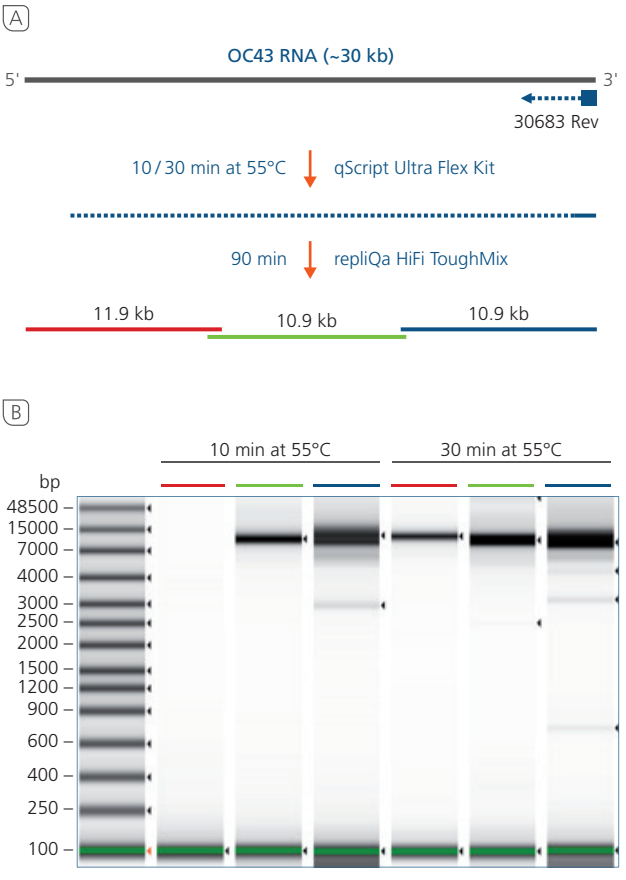


Figure 5 **cDNA synthesis beyond 20 kb using a single GSP for the OC43 RNA target.** **A** Schematic workflow showing cDNA synthesis and PCR reactions. **B** Genomic TapeStation analysis of the long amplicons using the cDNA generated by either 10-minute or 30-minute reactions.

CONCLUSIONS

In this application note, we have detailed a protocol for whole genome viral sequencing based on long range PCR amplification using qScript Ultra Flex Kit, repliQa HiFi ToughMix and sparQ DNA Frag and Library Prep Kit. We showed that the combination of these three products generate high quality libraries from whole viral genomes.

References

1. Woolhouse M. and Gowtage-Sequeria S. Host Range and Emerging and Reemerging Pathogens. Emerging infectious diseases. 2005 Dec; 11(12): 1842–1847.
2. Houldcroft C. Beale M and Breuer J. Clinical and biological insights from viral genome sequencing. Nature Review Microbiology. 2017 Mar;15(3):183-192
3. Misu, M Yoshikawa T. Sugimoto, S. et al. Rapid whole genome sequencing methods for RNA viruses. Frontiers in microbiology. 2023; 14: 1137086.

Trademarks: qScript®, repliQa® and ToughMix® are registered trademarks of Quantabio. ScreenTape® and TapeStation® are registered trademarks of Agilent Technologies, Inc. VR-1558D™ is a trademark of American Type Culture Collection. Qubit® is a registered trademark of Thermo Fisher Scientific, Inc. LLC. Illumina® and MiSeq® are trademarks of Illumina, Inc.

Quantabio products are intended for molecular biology applications. The products are not intended for the diagnosis, prevention or treatment of a disease.

MK-AN-0022 REV 01 Rapid LP RNA Virus Seq 0525