Streamlined whole-genome and targeted sequencing workflows for prevalent RNA viruses: Sars-CoV-2, Respiratory Syncytial Virus and Influenza A

Kaylinnette Pinet, Matthew Angel, Bradley W. Langhorst, V K Chaithanya Ponnaluri, and Keerthana Krishnan

New England Biolabs, Inc.

Introduction

The public health benefits of monitoring respiratory viruses through whole genome and targeted sequencing are evident in the wake of the recent global pandemic. We have developed robust sequencing approaches for respiratory RNA viruses, such as SARS-CoV-2, Human Respiratory Syncytial Virus (RSV), and Influenza to support scientists and public health experts in the monitoring of these viral pathogens.

We have further streamlined the RNA virus sequencing approach by combining reverse transcription and targeted amplification steps into a single step through the application of a reverse transcriptase and high-fidelity DNA polymerase enzyme mix. Combining the reverse transcription and targeted amplification reactions has the added benefits of reducing plastic consumption and probability of sample cross-contamination.

Following cDNA synthesis and targeted amplification, libraries can be prepared for sequencing on various sequencing platforms. For example, we have developed robust library prep reagents and protocols that provide complete genome coverage from amplicon-based sequencing on either Oxford Nanopore Technologies® or Illumina® sequencing platforms. Broadening established multiplex targeted amplification sequencing techniques to include additional viral targets, in conjunction with a single-step cDNA synthesis protocol, can support public health and research efforts by providing fast and resilient sequencing.

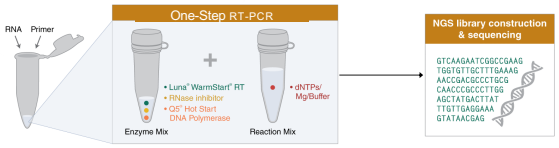
Methods

1. Conscientious design or selection of primer sets for targeting respiratory RNA viruses of interest.

- NEBNext VarSkip Short v2 primers - Variant skipping primers for SARS-CoV-2[1,2]
- iiMS Influenza A primers - Integrated indexing primers that target the homologous regions of all Influenza A genome segments for whole genome segment sequencing[3]
- H5N1 primers - Select tiled-amplicon primers for targeting Highly Pathogenic Avian Influenza (H5N1) HA and NA segments[4]
- NEBNext RSV primers - Primers with degenerate nucleotides for targeted RSVA and RSVB sequencing

2. One-Step cDNA synthesis and targeted amplification (sequencing platform agnostic).

3. Library preparation with optimized reagents and protocols.



Results

SARS-Co-V2

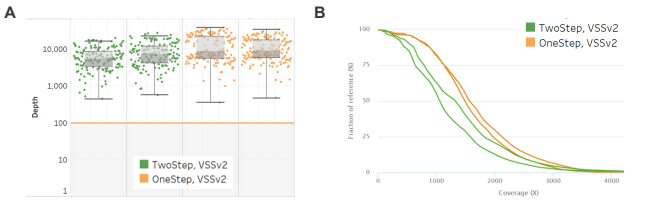
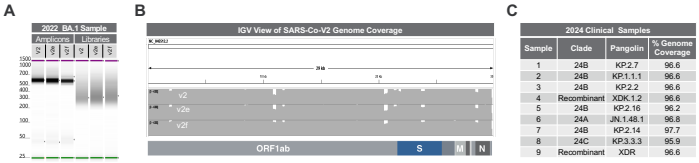


Figure 1. One-Step cDNA synthesis and amplification with SARS-CoV2 VSSv2 primers produces high amplicon and genome coverage. Amplicons were generated from 1,000 copies of SARS-CoV-2 viral gRNA (ATCC VR-1986) in 100 ng of Universal Human Reference RNA (Thermo Fisher Scientific®, QS0639) using VSSv2 primers. Either a two-Step or one-step cDNA synthesis and amplification reaction was applied. Libraries were constructed using the NEBNext ARTIC SARS-CoV-2 FS Library Prep Kit (Illumina®) and sequenced on a NextSeq® instrument (2X75 bp). 500,000 read pairs were sampled with seqtk[5] and aligned to the SARS-CoV-2 reference genome (NCBI, NC_045512) with Bowtie 2[6]. A) Sequencing reads per amplicon for each library. B) Genome fraction coverage by X reads (Mosdepth[7], MultiQC[8] tool plot).



Sample	Clade	Pangolin	% Genome Coverage
1	24B	KP.2.7	96.6
2	24B	KP.1.1.1	96.6
3	24B	KP.2.2	96.6
4	Recombinant	XDK.1.2	96.6
5	24B	KP.2.16	96.2
6	24A	JN.1.48.1	96.8
7	24B	KP.2.14	97.7
8	24C	KP.3.3.3	95.9
9	Recombinant	XDR	96.6

Figure 2. Spike-ins for NEB's Variant-skipping SARS-CoV-2 primers (VSSv2+) provide high genome coverage of previous and present-day variants. RNA extracted from SARS-CoV-2 positive clinical samples were sequenced with the NEBNext® ARTIC SARS-CoV-2 Library Prep Kit for Illumina or companion kit for Oxford Nanopore Technologies sequencing. VSSv2+ primer pools were utilized for targeted cDNA amplification. A) Diluted amplicons and final NEBNext UltraExpress® FS DNA library traces from a 2022 BA.1 clinical sample. B) Integrative Genome Viewer (IGV) visualization of read coverage across the same 2022 BA.1 genomes, respectively (0-1000 log scale). C) Strain and genome coverage for nine variants from 2024 clinical samples.

References

1. VarSkip multiplex PCR designs for SARS-CoV-2 sequencing. https://github.com/nebiolabs/VarSkip
2. Bei Y, Pinet K, Vrtis KB, Borgaro JG, Sun L, Campbell M, Apone L, Langhorst BW and Nichols NM (2022). Overcoming variant mutation-related impacts on viral sequencing and detection methodologies. *Front. Med.* 9:969813. doi: 10.3389/fmed.2022.969813
3. Kaylinnette Pinet, Abhinaya Srikanth, Peter Thielen, New England Biolabs, Juliet Bonnevie 2024. NEBNext iiMS Influenza A DNA Library Prep for Oxford Nanopore. protocols.io https://dx.doi.org/10.17504/protocols.io.n92ldd8mnl5b/v1
4. William C. Vuyk, Andrew Lail, Isla Emmen, Nura Hassan, Patrick Barros Tiburcio, Christina Newman, Nicholas R. Minor, Nancy Wilson, Thomas Friedrich, David O'Connor 2024. Whole genome sequencing of H5N1 from dairy products with tiled 250bp amplicons . protocols.io https://dx.doi.org/10.17504/protocols.io.kqdg3226pv25/v1
5. H. Li, Seqtk: a fast and lightweight tool for processing FASTA or FASTQ sequences, 2013. https://github.com/lh3/seqtk/
6. Langmead, B., Salzberg, S. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–359 (2012). https://doi.org/10.1038/nmeth.1923.
7. Pedersen, Brent S. and Aaron Quinlan. "Mosdepth: Quick Coverage Calculation for Genomes and Exomes." BioRxiv, no. November (2017): 10–11. https://doi.org/10.1101/185843.
8. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: Summarize analysis results for multiple tools and samples in a single report. Bioinformatics, 32(19), 2016, 3047–3048. doi: 10.1093/bioinformatics/btw354. PMID: 27312411.

Results

Influenza A (H3N2 and H5N1)

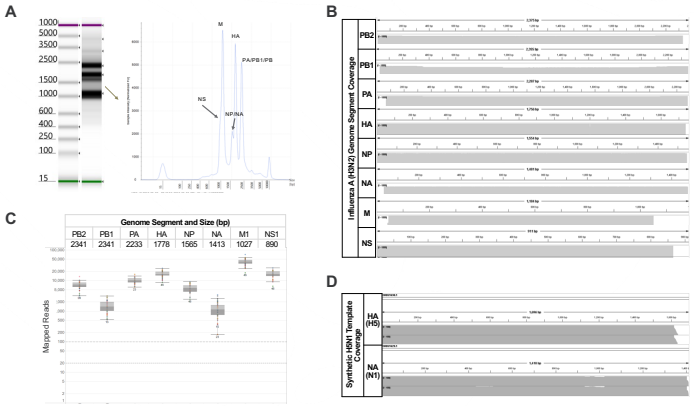


Figure 3. LunaScript® Multiplex One-Step cDNA synthesis and amplification of Influenza A targets provides robust amplicon yields for whole genome segment sequencing. For Fig. 3A-C, non-overlapping indexed amplicons were generated from 10,000 copies of Influenza A H3N2 gRNA (ATCC VR-1680) using NEBNext iiMS Flu A Primers. A one-step cDNA synthesis and amplification reaction was applied and indexed amplicons were subsequently pooled and cleaned-up prior to ONT library preparation (SQK-LSK114) and sequencing on a R10 GridION flow cell. A) Clean pooled amplicon traces. B) IGV visualization of read coverage across Influenza A genome segments (0-1000 log scale). C) Mapped read counts per genome segment for each of 48 barcoded samples. For Fig. 3D overlapping 250 bp amplicons were generated from 1,000 copies of Influenza A H5N1 Twist synthetic control genome segments in a one-step cDNA synthesis and amplification reaction. Then, the overlapping amplicons were used as input for NEBNext® UltraExpress FS DNA Library Prep for Illumina sequencing on a NextSeq® instrument (2X75 bp). D) IGV visualization of read coverage across the HA and NA genome segments of H5N1 Twist synthetic control genome segment templates (0-1000 log scale).

Respiratory Syncytial Virus (RSV A and RSV B)

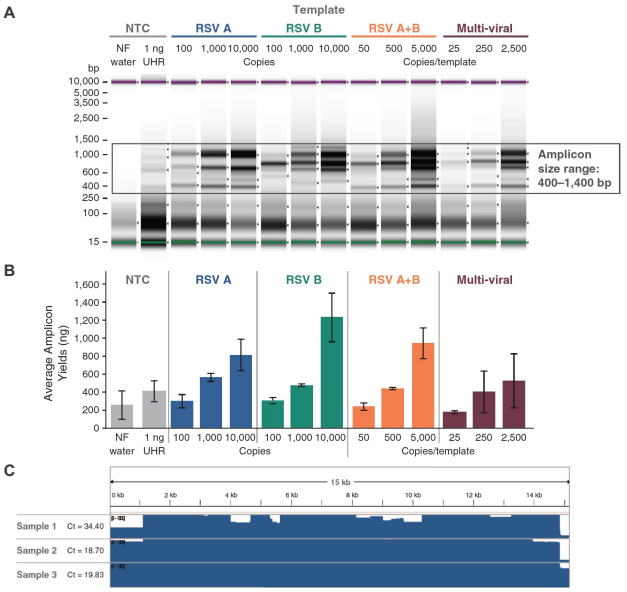


Figure 4. LunaScript® Multiplex One-Step cDNA synthesis and amplification of RSV genome targets of varying sizes provides robust amplicon yields and high genome coverage. Overlapping amplicons were generated from 100–10,000 total copies of control viral gRNA templates, with 1 ng Universal Human Reference RNA (UHR) background, or clinical samples, using dual multiplexed RSV-targeting primer pools. The control gRNA templates were RSV A gRNA (ATCC® VR1540), RSV B gRNA (ATCC VR-1580), a 1:1 [RSV A : RSV B] gRNA mix, a 1:1:1:1 [RSV A : RSV B : Influenza A : SARS-CoV-2] mix. The clinical samples with Ct values ranging from 18-35. A) 1/10th diluted amplicons from control templates run on TapeStation® using DS 5000 HS reagents. TapeStation traces show expected amplicon peaks within the 400 to 1,400 bp size range. B) Average amplicon yields (n=3) within 400 to 1400 bp window determined via TapeStation Analysis of diluted post-RT-PCR amplicons. C) IGV visualization of coverage across RSV A clinical sample genomes (0-1000 log scale).

Conclusions

- NEBNext is focused on developing variant-tolerant primer schemes and streamlined protocols for the targeted sequencing of viral pathogens
- One-step cDNA synthesis and amplification further streamlines the process of sequencing respiratory RNA viruses
- SARS-CoV-2 VSSv2b primers continue to provide high genome coverage across variant strains
- iiMS Influenza A sequencing shows high and balanced coverage of all eight gene segments
- Our dual multiplexed primer pool amplicon-based targeted sequencing method for RSV A and RSV B provides high genome coverage