



Short communication

Open access

## Assessment of Ion S5 NGS protocol for SARS-CoV-2 genome sequencing

Dino Pećar, Ivana Čeko, Lana Salihefendić, Rijad Konjhodžić

ALEA Genetic Center, Sarajevo, Bosnia and Herzegovina

DOI: 10.31383/ga.vol5iss2pp24-30

### Abstract

#### \*Correspondence

E-mail:  
 dinopecar@yahoo.com;  
 info@agc.ba

#### Received

November, 2021

#### Accepted

December, 2021

#### Published

December, 2021

**Copyright:** ©2021 Genetics & Applications, The Official Publication of the Institute for Genetic Engineering and Biotechnology, University of Sarajev

#### Keywords

*SARS-CoV-2, whole-genome sequencing, NGS*

Monitoring of the lineages SARS-CoV-2 is equally important in a fight against COVID-19 epidemics, as is regular RT - PCR testing. Ion AmpliSeq Library kit plus is a robust and validated protocol for library preparation, but certain optimizations for better sequencing results were required. Clinical SARS-CoV-2 samples were transported in three different viral transport mediums (VTM), on arrival at the testing lab, samples were stored on -20°C. Viral RNA isolation was done on an automatic extractor using a magnetic beads-based protocol. Screening for positive SARS-CoV-2 samples was performed on RT-PCR with IVD certified detection kit. This study aims to present results as follows: impact of first PCR cycle variation on library quantity, comparison of VTMs with a quantified library, maximum storage time of virus and correlation between used cDNA synthesis kit with generated target base coverage. Our results confirmed the adequacy of the three tested VTMs for SARS-CoV-2 whole-genome sequencing. Tested cDNA synthesis kits are valid for NGS library preparation and all kits give good quality cDNA uniformed in viral sequence coverage. Results of this report are useful for applicative scientists who work on SARS-CoV-2 whole-genome sequencing to compare and apply good laboratory practice for optimal preparation of the NGS library.

## Introduction

SARS-CoV-2 is an RNA virus first isolated in December of 2019. It is responsible for a severe acute respiratory syndrome that caused the COVID-19 pandemic (Wu F et al, 2020). The virus rapidly spread globally and on 11 March 2020 WHO declared a pandemic (Cucinotta & Vanelli, 2021).

There are many arguments for whole-genome SARS-CoV-2 sequencing: lineage shift monitoring, development of the diagnostic procedure, understanding the mutation rates, virulence changes, planning of public health strategies, support to therapeutic procedures, development of vaccines, etc. This study aims to test optimizations of the procedure as a practical guide for better genome reads and uniformity in target base coverage. The results obtained in this study are a collection of data acquired during screening for SARS-CoV-2 variants present in Bosnia and Herzegovina. Sequencing with Ion Torrent S5 platform by Thermo Fischer Scientific NGS instrument is deep coverage multi parallel and it is scalable for high throughput testing.

High coverage is necessary for the screening of variants in clinical samples with different viral loads. Optimization of protocol arises as necessary for library preparation being more challenging for RNA than DNA, because of the presence of hydroxyl group at 2' position in ribose and it is susceptible to base-catalyzed hydrolysis (Elliot & Lodomery, 2011). Therefore, we examined the effect of the three VTMs, customarily used in a clinical setting, as well as three reverse transcription protocols on the quality of viral genome sequence obtained by NGS.

## Material and methods

In this study, 37 clinical SARS-CoV-2 samples were analyzed from which complete genomes were sequenced using Ion S5 System. The samples for SARS-CoV-2 virus detection were transported and stored in three different viral transport mediums

(VTM). Nasopharyngeal/oropharyngeal swabs were taken from patients by different clinics and transported either in CITOSWAB® (8), Bio-Speedy® (16) or saline solution (13). RNA was isolated with automated extractor Tianlong Technology GeneRotex 96, which is a magnet beads-based method. After screening for the presence of SARS-CoV-2 virus using Bio-Rad CFX96 Touch Real-Time PCR Detection System, reverse transcription (cDNA synthesis) was done on PCR. In this study cDNA was synthesized with Ampliseq cDNA Synthesis for Illumina (8), Applied Biosystems™ High-Capacity cDNA Reverse Transcription Kit (13) and Invitrogen™ SuperScript™ VILO™ cDNA Synthesis Kit (16). For the cDNA synthesis protocol maximum amount of RNA was added and Ct of RT-PCR was not considered as maximum quantities of libraries were targeted.

Afterward, cDNA was quantified by Invitrogen™ Qubit™ 3 Fluorometer by Fisher Scientific; the concentration was expressed in ng/μl. PCR instrument used for cDNA synthesis and Ion AmpliSeq™ Library Kit Plus protocol was SimpliAmp™ Thermal Cycler by Applied Biosystems. For NGS library preparation Ion AmpliSeq™ Library Kit Plus and Ion AmpliSeq SARS-CoV-2 Research Panel were used.

The noted panel consists of two primer pools which include 247 primer pairs; 237 targets the SARS-CoV-2 sequence and 10 target human sequences that serve as the control. The first primer pool consists of 125 primer pairs (120 SARS-CoV-2 specific and 5 targeting human DNA), the second primer pool consists of 122 primer pairs (117 SARS-CoV-2 specific and 5 targeting human DNA). Ion AmpliSeq SARS-CoV-2 Research Panel is designed to cover 99% of the SARS-CoV-2 genome, starting from reference genome position 43 up to 29842 bp; the missing regions are due to primer placement. The panel is designed in a fashion that amplicons overlap, thus allowing single reads to obtain 2x coverage of 62,78%. This design allows greater accuracy in base calling protocol as deeper base coverage ensures that

base call is correct and not sequencing artifact or false call. The first PCR is crucial for NGS sequencing and its optimization results in the improvement of target base coverage.

Digestion, barcode/adaptor ligation and purification were done according to the manufacturer's guidelines. Libraries were purified with Beckman Coulter™ Agencourt AMPure XP magnetic beads technology. NGS libraries were quantified on QuantStudio™ 7 Flex Real-Time PCR System with Ion Library TaqMan™ Quantitation Kit and the concentration of library was expressed in pM. Pooled libraries were loaded on the Ion 530™ Chip using Ion Chef™ Instrument and then sequenced on Ion S5 System. Torrent Suite Software 5.12 was utilized for monitoring, tracking and analyzing runs and for the coverage statistics built-in Coverage Analysis Plugin. In the Coverage Analysis application there are four tiers of coverage: 1x, 20x, 100x and 500x. Ion Torrent software filters nucleotide calls with only Q30 Phred quality score which means the chance of incorrect base call 1 in 1000 or 99.9% base call accuracy (Ewing et al, 1998). Coverage of 1x means each base in reference was read at least once with the highest sequencing quality Q30, while 20x means the same base was read at least 20 times. The samples sequenced with coverage above 90% gave us full genomic alignment with 99% similarity with the reference genome.

During our study different SARS-CoV-2 lineages were detected using the same method which proves its applicability to sequence different strains of the virus. Assembly of sequence from BAM (Binary Alignment Map) files was performed with the Iterative Refinement Meta-Assembler (IRMA). IRMA was designed for the robust assembly, variant calling, and phasing of highly variable RNA viruses. IRMA is free to use and parallelizes computations for both cluster computing and single computer multi-core setups. Acquired sequences were uploaded to The Basic Local Alignment Search Tool (BLAST) on the NCBI site, the goal being to compare the acquired sequences with known SARS-CoV-2 isolate Wuhan-Hu-1

(NCBI Reference Sequence: NC\_045512.2). To position variant of sequenced sample on SARS-CoV-2 phylogenetic tree open source web site, Pangolin was applied. In this study, changes in the standard protocol were noted as well as practical tips for researchers who work on SARS-CoV-2 sequencing. Individual GISAID accession numbers for analyzed samples are:

EPI\_ISL\_463893, EPI\_ISL\_722201, EPI\_ISL\_722209,  
EPI\_ISL\_922076, EPI\_ISL\_910329, EPI\_ISL\_910334,  
EPI\_ISL\_959650, EPI\_ISL\_2029122, EPI\_ISL\_2029260,  
EPI\_ISL\_2709213, EPI\_ISL\_2800339, EPI\_ISL\_2800338,  
EPI\_ISL\_2708955, EPI\_ISL\_2708954, EPI\_ISL\_2820513,  
EPI\_ISL\_2820510, EPI\_ISL\_2810027, EPI\_ISL\_2854258,  
EPI\_ISL\_2800346, EPI\_ISL\_3230255, EPI\_ISL\_3184312,  
EPI\_ISL\_3184314, EPI\_ISL\_3230471, EPI\_ISL\_3230739,  
EPI\_ISL\_3184309 and EPI\_ISL\_3184316.

## Results and Discussion

### *Storage of SARS-CoV-2 samples*

The quality of RNA was validated by NGS sequencing, only 1x 98-99% coverage libraries were used for analysis. After screening, positive samples were stored at -20°C for a period ranging from 1 to 28 days. Table 1. presents maximum, minimum, and average storage time expressed in a number of days for different VTMs. The full genome coverage sequence was acquired from CITOSWAB® VTM stored at -20°C for 28 days.

### *RNA isolation*

The automated RNA extraction protocol was our method of choice because of: high yield of RNA, reduced chance of sample cross-contamination and human error also good and validated cleanup of RNA. RNA is an unstable molecule and storage of isolated and purified RNA is risky and expensive. RNAs from samples chosen for sequencing were re-isolated from VTMs freshly before cDNA synthesis. All of the RNA was isolated with Tianlong Technology GeneRotex 96 deep well automated extractor which utilizes magnetic beads-based isolation protocol. This protocol requires 200 µl of VTM and results in 50 µl of purified RNA elute.

Table 1. Maximum, minimum and average storage time on - 20° C, expressed in days, comparing three tested VTMs

Medium	Maximum storage time (days)	Minimum storage time (days)	Average storage time (days)
CITOSWAB®	28	1	11
Bio-Speedy vNAT®	15	1	4
Saline solution	12	1	5

*cDNA synthesis*

Purified viral RNA is used as a template for cDNA synthesis. Good innovative practice is to quantify cDNA using the standard absorbance DNA quantification method to verify if cDNA synthesis was completed. The results of this study indicate that cDNA concentration, from which NGS libraries can be prepared, ranges from 0,12 ng/ µl to 3,97 ng/ µl. Reverse transcription PCR from each clinical sample was done in duplicate, to apply cDNA with higher concentration for library preparation. For cDNA quantification Qubit 3.0 fluorometer was used and results of quantification are presented in Table 2. The results of our study indicate that all tested protocols produce good quality cDNA from clinical samples and obtained cDNA can be used for NGS library preparation.

*Ion AmpliSeq™ Library Kit Plus*

After successful cDNA synthesis, the next step was PCR amplification of viral target regions. In Ion AmpliSeq™ SARS-CoV-2 Research Panel Quick Reference recommended number of PCR cycles for low viral load is 24-27 and for high viral load is 14-21. Alessandrini et al, 2020 stated they used 12 and 20 PCR cycles, but they cultivated *in vitro* SARS-CoV-2 virus and the RNA isolation method was different from ours. Plitnick et al, 2021 used 17 PCR cycles but they prepared NGS libraries on Ion Chef Instrument using fully automated protocol; also they used the same cDNA synthesis kit as in this study which gave good results. The main obstacle we were facing during sequencing runs was to get a sufficient concentration of NGS libraries. Good quantity libraries were not achieved by using 14 to 27

Table 2. Maximum, minimum and average cDNA concentration obtained using three different kits

cDNA synthesis kit	Maximum concentration (ng/ µl)	Minimum concentration (ng/ µl)	Average concentration (ng/ µl)
Invitrogen™ SuperScript™ VILO™ cDNA Synthesis Kit	3,97	0,12	1,61
Ampliseq cDNA Synthesis for Illumina	3,9	0,54	1,81
Applied Biosystems™ High-Capacity cDNA Reverse Transcription Kit	2,47	0,4	1,19

first PCR cycles, therefore optimization was required. The number of first PCR cycles was increased from 35 to 40, giving superior results of NGS library quantification and target sequence coverage. From our experience, libraries with yields below 100 pM cannot give good quality reads. Interestingly, three libraries with quantification of 68, 71 and 86 pM gave good coverage of 1x 99% and they were added to the pool in the same volume as other libraries diluted to 100 pM. In Ion AmpliSeq™ SARS-CoV-2 Research Panel Quick Reference recommended loading concentration for Ion 530 Chip is 30pM. Diluting libraries to such recommended concentration resulted in very low and no reads at all. NGS libraries with a high enough concentration were diluted to a loading concentration of 100pM. Libraries with concentrations varying from 90 to 110 pM were pooled directly without diluting. The results of this

study show that it is possible to get 500x coverage of the viral genome utilizing all three tested mediums with 35 to 40 cycles of first PCR. Comparing average coverage of samples delivered in VTMs tested in this study shows that CITOSWAB® ensures slightly better results than Bio-Speedy vNAT® VTM and saline solution. As presented in Figure 1, the trend can be noticed in coverage fraction comparing different VTMs and all the tested mediums are valid for SARS-CoV-2 whole-genome sequencing. Figure 2 represents maximum, minimum and average target base coverage of all samples plotted with three cDNA synthesis kits. On the first three stacks in the Figure 2 maximum coverage can be noticed, Invitrogen™ SuperScript™ VILO™ cDNA Synthesis Kit and Applied Biosystems™ High-Capacity cDNA Reverse Transcription Kit can produce nearly 100% of 500x coverage. Comparing the results of average coverage, a trend can be identified that interpret the possibility of all three kits to give good cDNA.

**Table 3.** Maximum, minimum and average library concentration acquired from three VTMs

<b>Medium</b>	<b>Maximum library concentration (pM)</b>	<b>Minimum library concentration (pM)</b>	<b>Average library concentration (pM)</b>
<b>CITOSWAB®</b>	2701	110	1001
<b>Bio-Speedy vNAT®</b>	6986	68	1082
<b>Saline solution</b>	5323	71	1145

Results of this study reveal the slightly better performance of SuperScript™ VILO™ cDNA Synthesis Kit by Invitrogen and we recommend it also because hands-on time and PCR run is shorter than High-Capacity cDNA Reverse Transcription Kit by Applied Biosystems.

#### *NGS library quantification*

NGS libraries were quantified according to manufacturer protocol for Ion Library TaqMan™ Quantitation Kit

Libraries with concentrations above 2000 pM should be diluted then quantified again, for a more precise dilution factor and better overall sequencing results. Comparing the average library concentration of different VTMs tested, 1076 pM libraries were obtained and results are presented in Table 3. The results show that all VTMs tested in this study can be used for high coverage SARS-CoV-2 whole-genome sequencing.

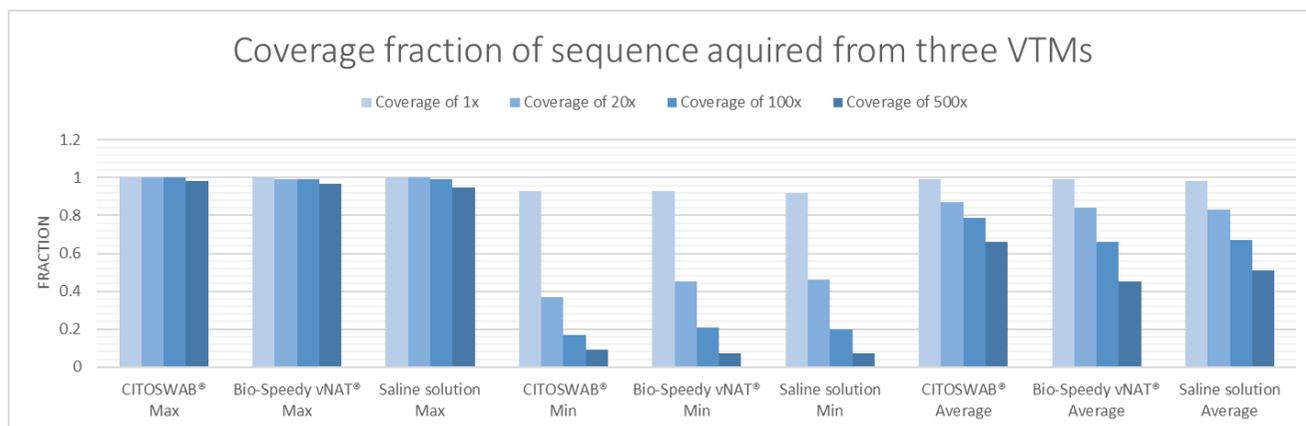


Figure 1. Maximum, minimum and average target base coverage acquired from different VTM's

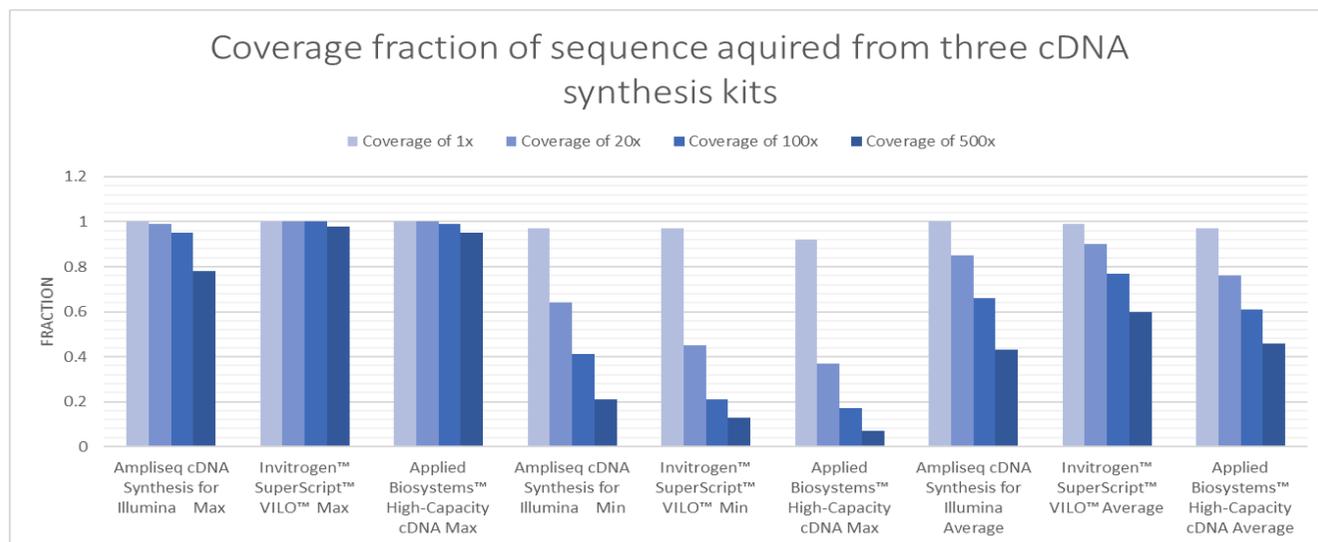


Figure 2. Maximum, minimum and average target base coverage comparing three cDNA synthesis kits

## Conclusion

Good quality RNA for NGS sequencing can be isolated from saline solution, CITOSWAB® and Bio-Speedy vNAT® stored on -20°C from 1 to 28 days. For longer storage we recommend CITOSWAB®. RNA should be isolated freshly before cDNA synthesis and we recommend using a maximum volume of RNA for the cDNA synthesis protocol. A good practice is to quantify cDNA before library preparation to test if cDNA synthesis was successful. SARS-CoV-2 whole-genome sequencing NGS libraries can be prepared from cDNA quantity ranging from 0,12 ng/ µl to 3,97 ng/ µl. Three of the tested cDNA synthesis kits provide a good template for NGS library preparation as all three gave sequences with maximum target base coverage. For whole-genome sequencing of SARS-CoV-2 from clinical samples with Ion AmpliSeq™ Library Kit Plus manual preparation protocol, the number of 1st PCR cycles should be 35 to 40. Libraries with concentrations above 2000 pM should be diluted and quantified again for a more precise dilution factor. Adequate loading concentration of each library for Ion AmpliSeq SARS-CoV-2 Research Panel is 100pM for 530 Ion Chip and should be pooled in the same ratio up to 24 samples.

## Acknowledgment

All funding for this study was provided by ALEA Genetic Center from Sarajevo, Bosnia and Herzegovina.

## Conflict of interest

The authors of this study declare that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or

beliefs) in the subject matter or materials discussed in this manuscript.

## References

- Alessandrini F, Caucci S, Onofri V, Melchionda F, Tagliabracci A, Bagnarelli P, Di Sante L, Turchi C, Menzo S, (2020) Evaluation of the Ion AmpliSeq SARS-CoV-2 Research Panel by Massive Parallel Sequencing. *Genes* 11, 929. <https://doi.org/10.3390/genes11080929>
- Cucinotta D, Vanelli M (2021) WHO Declares COVID-19 a Pandemic. *Acta Biomed.* 91(1):157-160.
- Elliot D, Ladomery M, (2011) *Molecular Biology of RNA* (1 ed.). Oxford University Press, New York, pp. 34–64.
- Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* 8(3):175–185.
- Plitnick J, Griesemer S, Lasek-Nesselquist E, Singh N, Lamson DM, St. George K. (2021) Whole-genome sequencing of SARS-CoV-2: assessment of the Ion Torrent AmpliSeq panel and comparison with the Illumina MiSeq ARTIC protocol. *J Clin Microbiol* 59:e00649-21. <https://doi.org/10.1128/JCM.00649-21>.
- Wu F, Zhao S, Yu B et al. (2020) A new coronavirus associated with human respiratory disease in China. *Nature* 579:265–269