COVseq is a cost-effective workflow for mass-scale SARS-CoV-2 genomic surveillance

Michele Simonetti
Karolinska Institutet https://orcid.org/0000-0003-3322-1697
Ning Zhang
Karolinska Institutet and Qilu Hospital of Shandong University https://orcid.org/0000-0002-6430-4236
Luuk Harbers
Karolinska Instituet https://orcid.org/0000-0003-3910-6497
Maria Grazia Milia
Ospedale Amedeo di Savoia
Thi Thu Huong Nguyen
Karolinska Instituet
Silvia Brossa
Instituto di Candiolo FPO - IRCCS
Magda Bienko
Karolinska Institutet https://orcid.org/0000-0002-6499-9082
Anna Sapino
Instituto di Candiolo FPO - IRCCS
Antonino Sottile
Instituto di Candiolo FPO - IRCCS
Valeria Ghisetti (✉ valeria.ghisetti@gmail.com)
Ospedale Amedeo di Savoia
Nicola Crosetto (✉ nicola.crosetto@ki.se)
Karolinska Institutet https://orcid.org/0000-0002-3019-6978

Method Article

Keywords: SARS-CoV-2, Next-generation sequencing, genomic surveillance, COVseq, virology

Posted Date: February 5th, 2021

DOI: https://doi.org/10.21203/rs.3.pex-1338/v1

Abstract

Genomic surveillance of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is critical to monitor the spread and evolution of the virus across different populations, geographical regions and species. Here, we present a streamlined workflow—COVseq—based on the CUTseq method that we previously described, which can be used to generate highly multiplexed sequencing libraries compatible with Illumina platforms, from hundreds of SARS-CoV-2 samples in parallel, in a rapid and cost-effective manner. We validated COVseq on RNA extracted from the supernatant of a SARS-CoV-2 culture as well as from 85 left-over samples from nasopharyngeal swabs, demonstrating the ability of COVseq to achieve almost complete genome coverage, including the S region encoding the spike protein. A cost analysis showed that COVseq could be used to sequence thousands of samples per week at less than 20 USD per sample. COVseq is a versatile and scalable method that can be readily applied for genomic surveillance of the ongoing pandemic and easily adapted to other pathogens such as influenza viruses.

Introduction

Since the identification of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) as the causative agent of coronavirus disease 2019 (COVID-19)[1], thousands of SARS-CoV-2 genomes have been sequenced worldwide and the sequences have been made publicly available (https://www.gisaid.org/). This has enabled a phylogenetic reconstruction of the viral spread and evolution across different countries and continents at an unprecedented scale[2], allowing the rapid identification of genomic variants of potential epidemiological concern, such as the A23403G mutation in the S region causing the D614G amino acid substitution in the spike protein S, which enhances infectivity[3]. SARS-CoV-2 whole genome sequencing (WGS) is being increasingly applied in epidemiological surveillance to track infections in hospitals and community settings, thus informing public health decisions[4,5]. SARS-CoV-2 WGS was also recently deployed to monitor outbreaks of the virus in mink farms[6], which represent a potential source of viral variants with increased pathogenicity. In genomic surveillance, the availability of rapid and cost-effective methods for sequencing hundreds or even thousands of samples per week would be greatly beneficial.

Various approaches have been adopted for SARS-CoV-2 WGS on different sequencing platforms. These include standard RNA sequencing[7–9], amplicon-based approaches[8–15], oligonucleotide capture-based methods[8,10–12], and direct RNA sequencing[13]. In addition, direct tagmentation of retro-transcribed RNA extracted from patient samples was shown to enable high SARS-CoV-2 genome coverage and simultaneous metagenomic analysis of other viruses, bacteria and yeast present in the same sample[14]. Recently, RNA-mediated oligonucleotide annealing selection and ligation coupled with next-generation sequencing (RASL-seq) was applied to detect SARS-CoV-2 in clinical samples, without the need for nucleic acid extraction and reverse transcription[15]. However, this method can only cover a small fraction of the SARS-CoV-2 genome. More generally, an important limitation of existing SARS-CoV-2 WGS methods is that they cannot be scaled up in a cost-effective manner. This is due to the fact that, typically,

one sequencing library must be prepared for each individual sample and multiple indexed libraries must be carefully quantified and balanced before pooling them together prior to sequencing. In the case of transposase-based approaches, such as Illumina's Nextera, the preparation of indexed libraries from multiple samples can be performed rapidly, but the cost per sample is very high. Here, we describe a versatile workflow—COVseq—based on the CUTseq method that we previously described[16], which allows preparing multiplexed sequencing libraries from a large number of SARS-CoV-2 samples in parallel, in a streamlined and cost-effective manner. We technically validate COVseq on RNA extracted from a SARS-CoV-2 viral culture and on 85 SARS-COV-2 positive left-over RNA samples from two hospitals in Italy. Lastly, we perform a cost analysis to demonstrate that COVseq is a highly cost-effective method that can be adopted for mass-scale genome surveillance of the ongoing pandemic.

Reagents

Reagents:

- Absolute Ethanol (VWR, cat. no. 20816.367)

- Nuclease-Free Water (Thermo Fisher Scientific, cat. no. AM9932)

- Mineral oil (Sigma, cat. no. M5904)

- Random hexamers (50 µM) (Thermo Fisher Scientific, cat. no. N80800127)

- dNTPs (10 mM) (Thermo Fisher Scientific, cat.no. R0191)

- NEBNext Q5 Hot Start HiFi PCR Master Mix (NEB, cat.no. M0543L)

- Primer pools 1,2,3,4,5 and 6 (IDT custom @ 50 µM)

- MseI (NEB, cat.no. R0525L)

- NlaIII (NEB, cat.no. R0125L)

- CutSmart® buffer (NEB, cat. no. B7204S)

- T4 DNA Ligase (Thermo Fisher Scientific, cat. no. EL0011)

- COVseq oligonucleotide adapters

- UltraPure™ BSA (50 mg/ml) (Thermo Fisher Scientific, cat. no. AM2616)

- ATP Solution (100 mM) (Thermo Fisher Scientific, cat. no. R0441)

- dNTPs (25mM) (Thermo Fisher Scientific, cat.no. R1121)

- MEGAscript® T7 Transcription Kit (Thermo Fisher Scientific, cat. no. AM1334)

- DNase I, RNase-free (Thermo Fisher Scientific, cat. no. AM2222)

- RA3 adaptor and RTP, RP1 and RPI primers (custom-made by Integrated DNA Technologies Inc. based on the sequences in the TruSeq Small RNA Library Preparation kit, Illumina)

- RNaseOUT™ Recombinant Ribonuclease Inhibitor (Invitrogen, cat. no. 10777-019)

- T4 RNA ligase 2, truncated (NEB, cat. no. M0242L)

- SuperScript® IV Reverse Transcriptase (Thermo Fisher Scientific, cat. no. 18090050)

- NEBNext UltraII Q5 PCR MasterMix (NEB, cat. no. M0544S)

- Agencourt RNAClean XP (Beckman Coulter, cat. no. A63987)

- Agencourt AMPure XP (Beckman Coulter, cat. no. A63881)

- Qubit™ RNA BR Assay Kit (Thermo Fisher Scientific, cat. no. Q10211)

- Qubit™ dsDNA BR Assay Kit (Thermo Fisher Scientific, cat. no. Q32850)

- Qubit™ dsDNA HS Assay Kit (Thermo Fisher Scientific, cat. no. Q32851)

- Bioanalyzer High Sensitivity DNA Kit (Agilent, cat. no. 5067-4627)

Consumables:

- Eppendorf® DNA LoBind microcentrifuge tubes 0.5 ml (Sigma, cat. no. EP0030108035-250EA)

- Eppendorf® DNA LoBind microcentrifuge tubes 1.5 ml (Sigma, cat. no. EP0030108051-250EA)

- Sapphire Filter tips, low retention (Greiner Bio-One, cat. no. 771265, 773265, 738265, 750265)

- microTUBE-50 AFA Fiber Screw-Cap (25) (Covaris, cat. no. 520166)

- # 96-well plates (Thermo Fisher Scientific, cat. no. 4316813)

- # 384-well plates (Thermo Fisher Scientific, cat. no. 4483320)

- Qubit™ Assay Tubes (Thermo Fisher Scientific, cat. no. Q32856)

- Bioanalyzer High-sensitivity DNA kit (Chips) (Agilent, cat. no. 5067-4626)

Equipment

- Incubator (for example, Binder incubator, Model KB 53 or Boekel Scientific InSlide Out, cat. no. 05-450-50)

- Tabletop centrifuge (for example, Eppendorf® Microcentrifuge 5424)

- I-DOT One (Dispendix GmbH, Stuttgart, Germany)

- Thermoshaker (for example, Eppendorf® Thermomixer Compact)

- PCR thermocycler (for example, Biometra TRIO)

- Sonication device (for example, ME220 Focused-ultrasonicator, Covaris)

- DynaMag™-2 Magnet (Thermo Fisher Scientific, cat. no. 12321D)

- Qubit® 2.0 Fluorometer (Thermo Fisher Scientific, cat. no. Q32866)

- Bioanalyzer 2100 (Agilent, cat. no. G2943CA)

Procedure

DAY 1

First strand synthesis

1. Mix the following components:

RNA 5 µl

50 µM random hexamers 1 µl

10 mM dNTPs 1 µl

Nuclease-Free Water 6 µl

2. Incubate for 5 min @ 65 °C

3. Place the tube immediately on ice for 5 min

4. Add the following components:

5x SSIV buffer 4 µl

0.1 M DTT 1 µl

RNaseOUT 1 µl

SSIV Reverse Transcriptase enzyme 1 µl

5. Perform the following steps in a PCR thermocycler with the lid set @ 85 °C:

1. 25 °C 10 min

2. 50 °C 10 min

3. 85 °C 10 min

4. 4 °C Hold

6. Add 1 µl RNase H to the tube and incubate for 20 min at 37 °C

Multiplex PCR

Note: Prepare primers as 50 µM primer stocks. Add an equal amount of each 50 µM primer stock to six different Eppendorf tubes labeled as pool 1,2,3,4,5 and 6. Prepare 10 µM working concentration by diluting each pool 1:5 with Nuclease-Free Water.

7. Mix the following components:

NEBNext Q5 Hot Start HiFi PCR Master Mix 15 µl

Nuclease-Free Water 9.2 µl

Primer pool 1,2,3,4,5 or 6 (10 µM) 1.8 µl

4x SYBR Green 1 µl

8. Add 3 µl of cDNA to each tube

9. Perform the following steps in PCR thermocycler with the lid set @ 105 °C:

1. 98 °C 30 sec

2. 98 °C 15 sec

3. 65 °C 5 min

GOTO step 2, 40 times

4. 4 °C Hold

DNA purification

Note: This is an optional step. For the continuation even unpurified material can be used

10. Pool 20 µl from each of the six amplicon pools into a 1.5 ml LoBind tube
11. Add 1x vol/vol ratio of AMPure XP beads pre-warmed at room temperature
12. Mix thoroughly and incubate for 10 min at room temperature
13. Place the sample on a magnetic stand
14. Incubate for at least 5 min until the liquid appears clear
15. Remove and discard the supernatant
16. Wash the beads twice with 200 µl of freshly prepared 80% ethanol
17. Air-dry the beads at room temperature

Note: do not dry the beads for more than 5–8 min, since this may result in low DNA yield

18. Remove the sample from the magnetic stand
19. Resuspend the beads in 80 µl of nuclease-free water
20. Incubate for 2 min at room temperature
21. Place the sample back on the magnetic stand

22. Incubate for at least 5 min until the liquid appears clear

23. Transfer the supernatant to a new 1.5 µl DNA LoBind tube

24. Check the library concentration using Qubit dsDNA BR kit

Note: Samples can be stored for long time @ -20 °C

DAY 2

Note: To process multiple samples in parallel, we performed all reactions until IVT in 384-well plates. We used the I-DOT One nanodispensing device (Dispendix GmBH) to reduce volumes of each reagent. For this step one can use unpurified material coming from the multiplex PCR after pooling an equal volume from each of the six pools. Other dispensing systems may also be used; however, volumes might have to be adjusted depending on the technical specifications of each instrument.

DNA digestion

25. Dispense manually 5 µl of mineral oil per well in the targeted region of 384-well plates

26. Dispense 50 nL of purified or 100 nL of non-purified PCR amplicons

27. Dispense Nuclease-Free Water to a total volume of 350 nL

Note: From now, after dispensing for each step, we shake the plate in a ThermoMixer at 1,000 rpm for 1 min and centrifuge at 3,220 g for 5 min before each incubation

28. Mix the following components:

NlaIII enzyme 50 nL

MseI enzyme 50 nL

10x CutSmart Buffer 50 nL

29. Dispense 150 nL per well

30. Perform the following incubation steps:

1. 37 °C 1 h

2. 65 °C 20 min

3. 4 °C Hold

Ligation of COVseq adapters

31. Dispense 150 nL of 33 nM COVseq adaptor of NlaIII per well

32. Dispense 150 nL of 33 nM COVseq adaptor of MseI per well

33. Dispense 700 nL of ligation mix per well:

Nuclease-free water 250 nL

10x T4 ligase buffer 150 nL

ATP 10 mM 120 nL

BSA 50 mg/ml 30 nL

T4 standard ligase 150 nL

34. Perform incubation at 22° C for 1 h followed by inactivation at 70° C for 5 min

35. Dispense manually 5 µl of Nuclease-Free Water per well

36. Pool the content of multiple wells manually in a 1.5 mL or 5 mL eppendorf tube

37. Spin down the tube and carefully remove the upper phase containing mineral oil

Note: Pooling step can be performed by centrifuging the plate upside down into a collection plate placed at the bottom at 800 rpm for 1min.

DNA cleanup

38. Add 1.2x vol/vol ratio of AMPure XP beads pre-warmed at room temperature

39. Mix thoroughly and incubate for 10 min at room temperature

40. Place the sample on a magnetic stand

41. Incubate for at least 5 min until the liquid appears clear

42. Remove and discard the supernatant

43. Wash the beads twice with freshly prepared 80% ethanol (the ethanol should be enough to cover the beads)

44. Air-dry the beads at room temperature

Note: do not dry the beads for more than 5–8 min, since this may result in low DNA yield

45. Remove the sample from the magnetic stand

46. Resuspend the beads in 10 µl of nuclease-free water

47. Incubate for 2 min at room temperature

48. Place the sample back on the magnetic stand

49. Incubate for at least 5 min until the liquid appears clear

50. Transfer the supernatant to a new 1.5 µl DNA LoBind tube

51. Check the library concentration using Qubit dsDNA HS kit

Note: Samples can be stored for long time @ -20 °C

In vitro transcription

52. Start with 8 µl from the previous step

53. Add the following reagents on ice:

rATP+rUTP+rGTP+rCTP* 8 µl

10x T7 polymerase buffer 2 µl

T7 polymerase 1.5 µl

RNaseOUT™ Recombinant Ribonuclease Inhibitor 0.5 µl

*Prepared from separate rNTP solutions provided with the MEGAscript® T7 Transcription Kit

54. Incubate for 14 hours at 37 °C in a PCR thermocycler with the lid set @ 70 °C

Note: IVT can also be performed at 37°C for 2 hours to save time.

DAY 3

RNA cleanup

55. Add 1 µl of DNase I (RNase-free) to the IVT product

56. Incubate for 15 min @ 37 °C

57. Bring up the volume to 50 µl by adding 29 µl Nuclease-Free Water, then mix with 90 µl (1.8x vol/vol) of RNAClean XP beads pre-warmed at room temperature

58. Mix thoroughly and incubate for 10 min at room temperature

59. Place the sample on a magnetic stand

60. Incubate for at least 5 min until the liquid appears clear

61. Remove and discard the supernatant

62. Wash the beads twice with 200 µl of freshly prepared 70% ethanol

63. Air-dry the beads at room temperature

Note: do not dry the beads for more than 5–8 min, since this may result in low DNA yield

64. Remove the sample from the magnetic stand

65. Resuspend the beads in 9.5 µl of nuclease-free water

66. Incubate for 2 min at room temperature

67. Place the sample back on the magnetic stand

68. Incubate for at least 5 min until the liquid appears clear

69. Transfer 8.8 µl of supernatant to a new 0.5 µl DNA LoBind tube

70. Check the library concentration with 1 µl using Qubit dsDNA BR kit

RA3 adapter ligation

71. Add 1 µl of 10 µM RA3 adapter to 7.8 µl obtained after RNA cleanup

72. Incubate for 2 min @ 70 °C in a PCR thermocycler, then immediately place sample on ice

73. Add 3.2 µl of the following mix:

RNA ligase buffer 1.2 µl

RNaseOUT™ Recombinant Ribonuclease Inhibitor 1 µl

T4 RNA ligase truncated 1 µl

74. Incubate for 2 hours @ 25 °C in a PCR thermocycler with the lid set @ 30 °C

Reverse transcription

75. Add 2 µl per sample of RTP primer

76. In a PCR thermocycler, incubate for 2 min @ 70 °C

77. Quickly transfer the sample to ice

78. Add 11 µl of the following mix:

5x SSIV buffer 5 µl

12.5mM dNTPs 1 µl

0.1M DTT 2 µl

RNaseOUT™ Recombinant Ribonuclease Inhibitor 1 µl

SuperScript IV reverse transcriptase 2 µl

79. Incubate for 20 min @ 50°C followed by inactivation for 10 min @ 80 °C in a PCR thermocycler with the lid set @ 80 °C

Library indexing and amplification

80. Add 16 µl per sample of the desired indexed Illumina primer

81. Add 359 µl of the following mix:

Nuclease-free water 143 µl

NEBNext® Ultra™ II PCR Master Mix 200 µl

RP1 primer 16 µl

82. Divide the final mix in 8 PCR tubes with each containing 50 µl

83. In a PCR thermocycler perform the following cycles:

1. 98 °C 30 sec

2. 98 °C 10 sec

3. 60 °C 30 sec

4. 65 °C 45 sec

GOTO step 2, 10 times

5. 65 °C 5 min

6. 4 °C Hold

Note: 10 PCR cycles are used for an input to the IVT of ≈ 200 ng. Please, adjust PCR cycles accordingly to the input to the IVT step.

Final library cleanup

84. Pool the 8 tubes for each sample and add 0.8x vol/vol ratio of AMPure XP beads pre-warmed at room temperature

85. Mix thoroughly and incubate for 10 min at room temperature

86. Place the sample on a magnetic stand

87. Incubate for at least 5 min until the liquid appears clear

88. Remove and discard the supernatant

89. Wash the beads twice with 1 ml of freshly prepared 80% ethanol

90. Air-dry the beads at room temperature

Note: do not dry the beads for more than 5–8 min, since this may result in low DNA yield

91. Remove the sample from the magnetic stand

92. Resuspend the beads in 50 µl of nuclease-free water

93. Incubate for 2 min at room temperature

94. Place the sample back on the magnetic stand

95. Incubate for at least 5 min until the liquid appears clear

96. Transfer the supernatant to a new 1.5 µl DNA LoBind tube

97. Check the library concentration using Qubit dsDNA HS kit

98. Check the fragment distribution on a Bioanalyzer 2100 using DNA HS chip

Note: Libraries can be stored for long time @ -20 °C

Troubleshooting

Time Taken

Anticipated Results

References

1. Zhu, N. *et al.* A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N. Engl. J. Med.* **382**, 727–733 (2020).

2. Mercatelli, D. & Giorgi, F. M. Geographic and Genomic Distribution of SARS-CoV-2 Mutations. *Front. Microbiol.* **11**, 1800 (2020).

3. Plante, J. A. *et al.* Spike mutation D614G alters SARS-CoV-2 fitness. *Nature* (2020) doi:10.1038/s41586-020-2895-3.

4. Harilal, D. *et al.* SARS-CoV-2 Whole Genome Amplification and Sequencing for Effective Population-Based Surveillance and Control of Viral Transmission. *Clin. Chem.* **66**, 1450–1458 (2020).

5. Meredith, L. W. *et al.* Rapid implementation of SARS-CoV-2 sequencing to investigate cases of health-care associated COVID-19: a prospective genomic surveillance study. *Lancet Infect. Dis.* **20**, 1263–1272 (2020).

6. Oude Munnink, B. B. *et al.* Transmission of SARS-CoV-2 on mink farms between humans and mink and back to humans. *Science* (2020) doi:10.1126/science.abe5901.

7. Pillay, S. *et al.* Whole Genome Sequencing of SARS-CoV-2: Adapting Illumina Protocols for Quick and Accurate Outbreak Investigation during a Pandemic. *Genes* **11**, (2020).

8. Nasir, J. A. *et al.* A Comparison of Whole Genome Sequencing of SARS-CoV-2 Using Amplicon-Based Sequencing, Random Hexamers, and Bait Capture. *Viruses* **12**, (2020).

9. Tyson, J. R. *et al.* Improvements to the ARTIC multiplex PCR method for SARS-CoV-2 genome sequencing using nanopore. *bioRxiv* (2020) doi:10.1101/2020.09.04.283077.

10. Xiao, M. *et al.* Multiple approaches for massively parallel sequencing of SARS-CoV-2 genomes directly from clinical samples. *Genome Med.* **12**, 57 (2020).

11. Paden, C. R. *et al.* Rapid, Sensitive, Full-Genome Sequencing of Severe Acute Respiratory Syndrome Coronavirus 2. *Emerg. Infect. Dis.* **26**, 2401–2405 (2020).

12. Doddapaneni, H. *et al.* Oligonucleotide capture sequencing of the SARS-CoV-2 genome and subgenomic fragments from COVID-19 individuals. *bioRxiv* (2020) doi:10.1101/2020.07.27.223495.

13. Kim, D. *et al.* The Architecture of SARS-CoV-2 Transcriptome. *Cell* **181**, 914-921.e10 (2020).

14. Chen, C. *et al.* MINERVA: A facile strategy for SARS-CoV-2 whole genome deep sequencing of clinical samples. *bioRxiv* 2020.04.25.060947 (2020) doi:10.1101/2020.04.25.060947.

15. Credle, J. J. *et al.* Highly multiplexed oligonucleotide probe-ligation testing enables efficient extraction-free SARS-CoV-2 detection and viral genotyping. *BioRxiv Prepr. Serv. Biol.* (2020) doi:10.1101/2020.06.03.130591.

16. Zhang, X. *et al.* CUTseq is a versatile method for preparing multiplexed DNA sequencing libraries from low-input samples. *Nat. Commun.* **10**, 4732 (2019).

Acknowledgements

We would like to acknowledge the thousands of research groups from all over the world who publicly shared their SARS-CoV-2 genome sequences through the GISAID initiative platform. This work was supported by funds from the National Natural Science Foundation of China (no. 81972475) and Chinese Postdoctoral Science Foundation (2019T120593, 2018M630787) to N.Z.; by funds from the Fondazione Piemontese per la Ricerca sul Cancro (INTEGRAZIONI DIAGNOSTICA IN ONCOLOGIA – INTERONC FPRC 5x1000 MIUR 2017) to A.S. and A.SO.; by a SciLifeLab/KAW National COVID-19 Research Program project grant, Research Area Viral Sequence Evolution to N.C.; and by a private donation for COVID-19 research from Chiesi Pharma AB, also to N.C.