Alternate primers for whole-genome SARS-CoV-2 sequencing

Matthew Cotten*, Dan Lule Bugembe, Pontiano Kaleebu, My V.T. Phan

Author affiliations: UK Medical Research Council–Uganda Virus Research Institute and London School of Hygiene and Tropical Medicine Uganda Research Unit, Entebbe, Uganda (D. Bugembe, P. Kaleebu, M. Cotten); Uganda Virus Research Institute, Entebbe (P. Kaleebu); Erasmus Medical Center, Rotterdam, the Netherlands (M.V.T. Phan); UK Medical Research Council–University of Glasgow Centre for Virus Research, Glasgow, Scotland, UK (M. Cotten)

*Correspondence to Matthew Cotten, MRC/UVRI & LSHTM Uganda Research Unit, Entebbe, Uganda email: matthew.cotten@lshtm.ac.uk

Abstract

As the world is struggling to control the novel Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), there is an urgency to develop effective control measures. Essential information is encoded in the virus genome sequence with accurate and complete SARS-CoV-2 sequences essential for tracking the movement and evolution of the virus and for guiding efforts to develop vaccines and antiviral drugs. While there is unprecedented SARS-CoV-2 sequencing efforts globally, approximately 19 to 43% of the genomes generated monthly are gapped, reducing their information content. The current study documents the genome gap frequencies and their positions in the currently available data and provides an alternative primer set and a sequencing scheme to helps improve the quality and coverage of the genomes.

Keywords: SARS-CoV-2; COVID-19, primers; next generation sequencing;

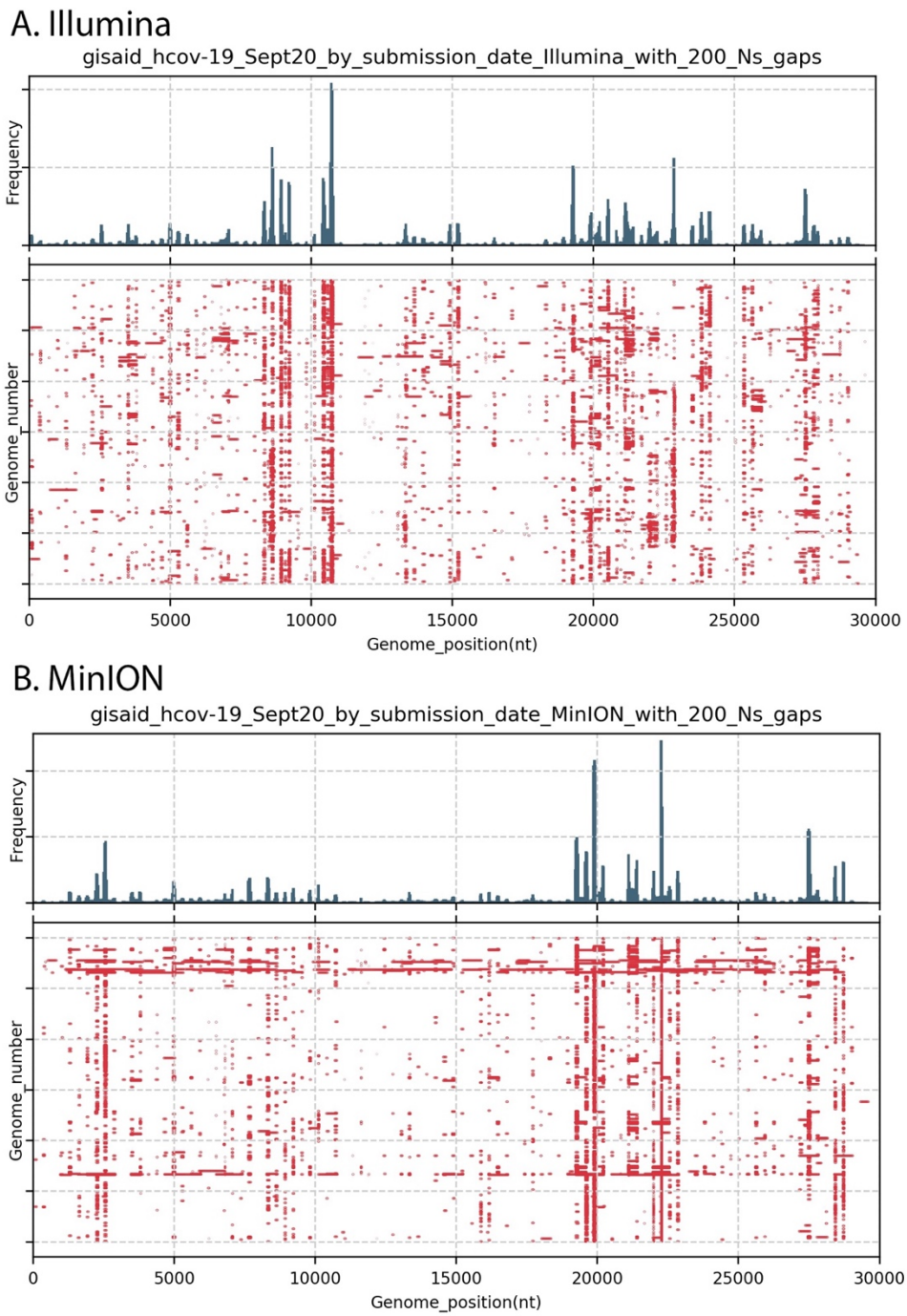
Introduction

Since the first report on 30th December 2019 in Wuhan China and the WHO declaration of the pandemic on 12th March 2020, the novel Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) (1) and the associated disease Coronavirus Disease 2019 (COVID-19) (2)(3) have continued to spread throughout the world, causing >35 million infections and >1,000,000 death globally (4). The virus genome sequences carry important information, which can be used to interpret the virus transmission, evolution patterns and origin tracing. Furthermore, accurate and complete genomic sequences are essential for monitoring diagnostics and developing novel therapeutics and vaccines. We have seen an unprecedented amount of virus sequencing with over 130,000 complete or nearly complete genome sequences of SARS-CoV-2 now available in the GISAID database by the end of September 2020 (5). Most of the sequences have been generated by next generation sequencing using targeted amplicon methods. A scan through SARS-CoV-2 genomes from GISAID with the filter "complete genome" revealed a high frequency of gaps occurring across the genome, influencing the overall genomics quality and interpretation. Here we describe an alternate primer scheme for whole-genome sequencing to improve the genome sequence quality and coverage.

Documenting the problem

We retrieved genomes deposited to GISAID in September 2020 (9 months into the pandemic), using the “complete genome” filter and sorting the genomes by sequencing platforms information included in the metadata. Figure 1 illustrated the positions across the 30kb genome of every stretch of 200 Ns (N200; nearly the size of an amplicon) in the first genomes deposited in September 2020 using the Illumina platform (Panel A; N = 3000) versus MinION platform (Panel B; N = 3000). Histograms of gap frequencies across the genomes are shown for each platform. The gaps are not randomly distributed but occur with higher frequency in a subset of positions across the genome. Although genomes generated by the

52 two platforms (Illumina and MinION) show similar problem regions (nt 8000-11000, and nt
53 19,000-24,000 relative to the reference genome NC_045512), the patterns are not completely
54 identical. Given the use of several primer amplification schemes, we suspect the gaps in
55 coverage may be due to unexpected primer interactions, complicated sequence regions (odd
56 composition or secondary structure), issues with primer trimming during quality control of read
57 data, or some combinations of these factors.



58

Figure 1. Positions of 200nt gaps across SARS-CoV-2 genomes listed as complete in GISAID. Genomes deposited in September 2020 (n= 38,228) were retrieved from GISAID, sorted by sequencing platform and the presence of 200N motifs was monitored. Panel A present gaps in the first 3000 Illumina-generated genome sequences deposited, the start positions of each N200 in each genome are indicated with red markers. The upper panel histogram shows the frequency (in 30 nt bins) of N200 motifs across the genomes. Panel B is the same analysis of the first 3000 MinION-generated genome sequences in September 2020.

The phenomenon is unlikely to be due to an isolated set of genomes as we observed similar N200 frequencies in genomes submitted from each month of the pandemic (Table 1), suggesting that gaps in coverage is a more general phenomenon. Of note, genomes generated using Ion Torrent show much lower levels of N200 (Table 1). The very low frequency of large gaps in the Ion Torrent data may be due to the use of a dedicated alternative primer set (6). There have been discussions and reports on the SARS-CoV-2 genome changes due to sequencing errors as well as long gaps in the genomes due to missing amplicons from the amplicon-approach sequencing (7) (8) (9). Updates of the ARTIC primers have been presented in late March 2020 to address these issues (10). However, the percentage of reported complete genomes in GISAID with 1 or more N200s continues, with 10,611 (28%) of the 38,228 genomes deposited in September 2020 having 1 or more N200 gaps (Table 1), indicating the challenges remain largely unsolved.

Table 1. Frequency of SARS-CoV-2 genomes with 1 or more 200 nt gaps (N200) by month and by sequencing platform.

Deposition period	Complete genomes	Genomes with 1 or more N200[2]	%. with 1 or more N200[3]	Illumina total[4]	Illumina % with N200[5]	MinION total[6]	MinION % with N200[7]	Ion Torrent total[8]	Ion Torrent % with N200[9]	Method unclear[10]	Method unclear % with N200[11]
1Jan20-31Jan20	54	1	1.85	9	0	2	0	0	0	43	2
1Feb20-29Feb20	126	2	1.59	44	0	9	0	5	0	65	4
1Mar20-31Mar20	2872	559	19.46	1518	16	548	32	35	6	771	18
1Apr20-30Apr20	12411	3745	30.17	4970	38	1286	27	264	0	5424	26
1May20-31May20	19787	8606	43.49	8634	52	2634	30	529	0	7990	42
1Jun20-30Jun20	21665	8723	40.26	7043	36	3844	35	629	2	10149	47
1Jul20-31Jul20	17986	4834	26.88	4965	23	1585	33	471	2	10965	29
1Aug20-31Aug20	17276	4005	23.18	11074	22	2270	26	486	0	3446	28
1Sep20-30Sep20	38227	10611	27.76	22740	23	7973	44	580	1	6934	28

1. Number of genomes with the annotation "complete" retrieved from GISAID (https://www.gisaid.org/).

2. Genomes were sorted by the presence or absence of the sequence N200.

3. ((The number of genomes with at least 1 N200)/total number of genomes)*100.

4. Number of genomes in GISAID for this period generated using any of the Illumina methods as noted in the GISAID "Sequencing technology metadata" .

5. ((The number of Illumina genomes with at least 1 N200)/total number of genomes)*100.

6.,7. Number of genomes in GISAID for this period generated using any of the MinION methods as noted in the GISAID "Sequencing technology metadata" and their percentage.

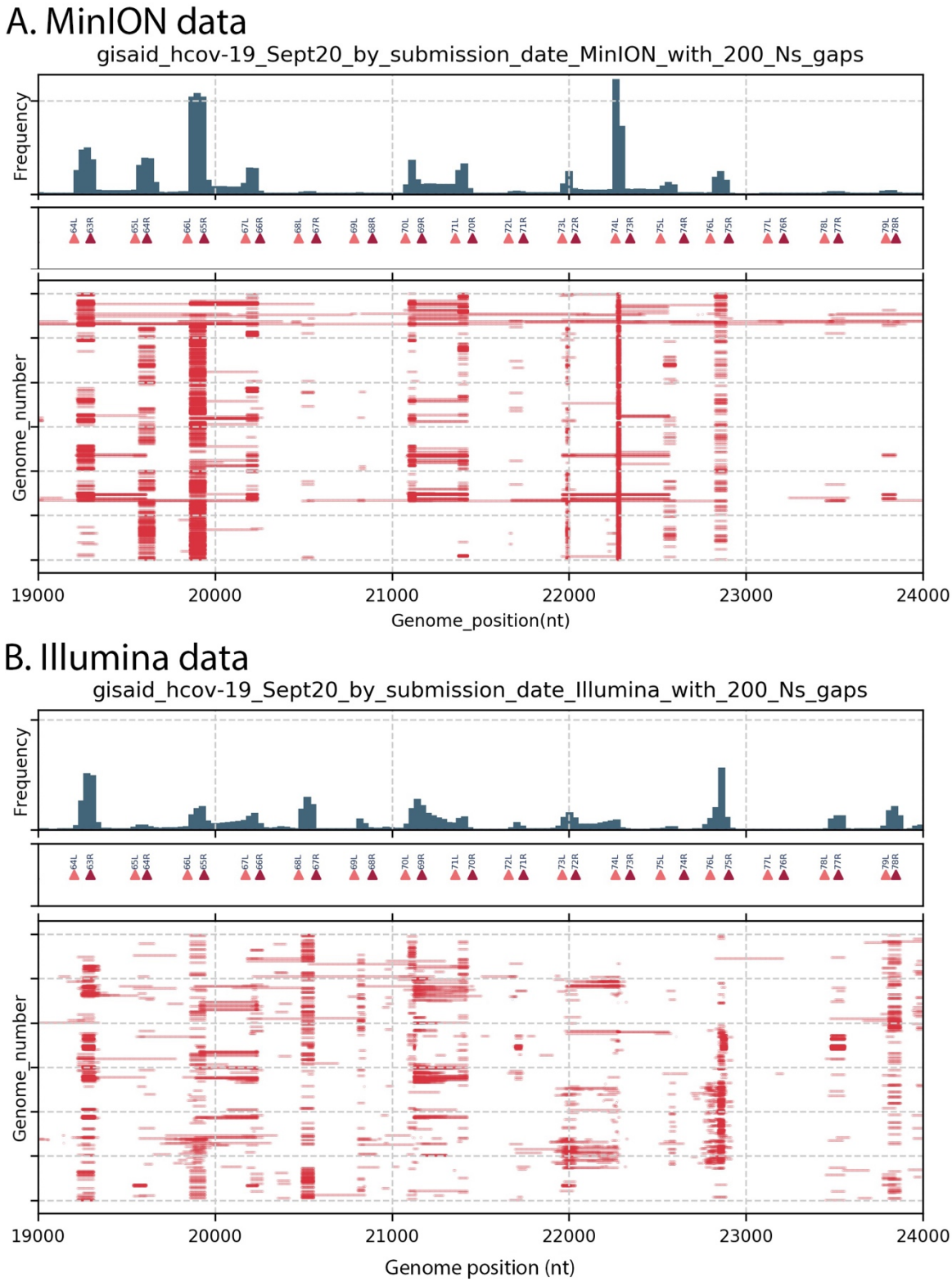
8.,9. Number of genomes in GISAID for this period generated using any of the Ion Torrent methods as noted in the GISAID "Sequencing technology metadata" and their percentage.

10.,11. Number of genomes in GISAID for this period generated using unclear methods as noted in the GISAID "Sequencing technology metadata" and their percentage.

Detailed analysis of gaps.

A more focused analysis of the frequent gaps is provided in Figure 2. The gap pattern between nt 19,000 and 24,000 (relative to the reference genome NC_045512) is shown for both MinION and Illumina sequences (first 3000 genomes of each deposited in September 2020). For reference the positions of the ARTIC primers (v.1) in the region are indicated (middle panel). A histogram of gap start positions (top panel) and the individual genome gaps (bottom panel) are also shown. The peaks are all between Forward from Amplicon X and Reverse from Amplicon X-1, for both MinION and Illumina data. Some of the gaps span the entire amplicons (the region between primerL and primerR, but the peaks of gap regions frequently span the region between Forward primer from amplicon X and Reverse primer from Amplicon X-1. Possible explanation for these gaps could be: (1) The primer pair fails at the amplification and no amplicon is available to sequence; or (2) The amplicons sequence is in

the raw data but a step in the data analysis, perhaps the primer trimming process, results in a gap in the data.



112 **Figure 2. Positions of 200nt gaps across SARS-CoV-2 genomes stratified by MinION or**
113 **Illumina, in region 19000 to 24000.** Genomes deposited in September 2020 as "complete
114 were retrieved from GISAID, sorted by sequencing platform and the presence of stretches of
115 N200 were monitored. For clarity only the first 3000 genomes in each set were plotted. The
116 upper histogram shows the frequency (in 30 nt bins) of N200 motifs across the genome region.
117 The middle panel shows the positions and names of ARTIC primers colored by forward or
118 reverse, the bottom panel shows start positions of 200 nt gap by genome, in the first 3000
119 genomes of each set. Panel A: MinION derived genome sequences, Panel B: Illumina-derived
120 genome sequences.

121

122 **Alternate primers as a potential solution to avoid gapped genomes.**

123 We explored an alternate set of amplification primers (termed the **Entebbe primers**)
124 designed using methods we had previously used for MERS-CoV (11), Norovirus (12), RSV
125 (13) and Yellow Fever virus (14). Important for the design were the amplicon size and the
126 primer placement. For their implementation, the use of primers for the reverse transcription
127 step, and the multiplexing of the amplicons in two staggered sets was important for the PCR.
128 Our experience had suggested an optimum amplicon size of around 1500 bp. The larger
129 amplicons reduced the total primers content of the reactions but still allowed high reverse
130 transcription efficiency (which, in our hands, declined beyond 1500 nt). Here we describe
131 primers designed for whole-genome sequencing of SARS-CoV-2, as well as sharing the
132 detailed laboratory methods that we used for reverse transcription, PCR amplification and
133 MinION library preparation to successfully sequence the SARS-CoV-2 genome.

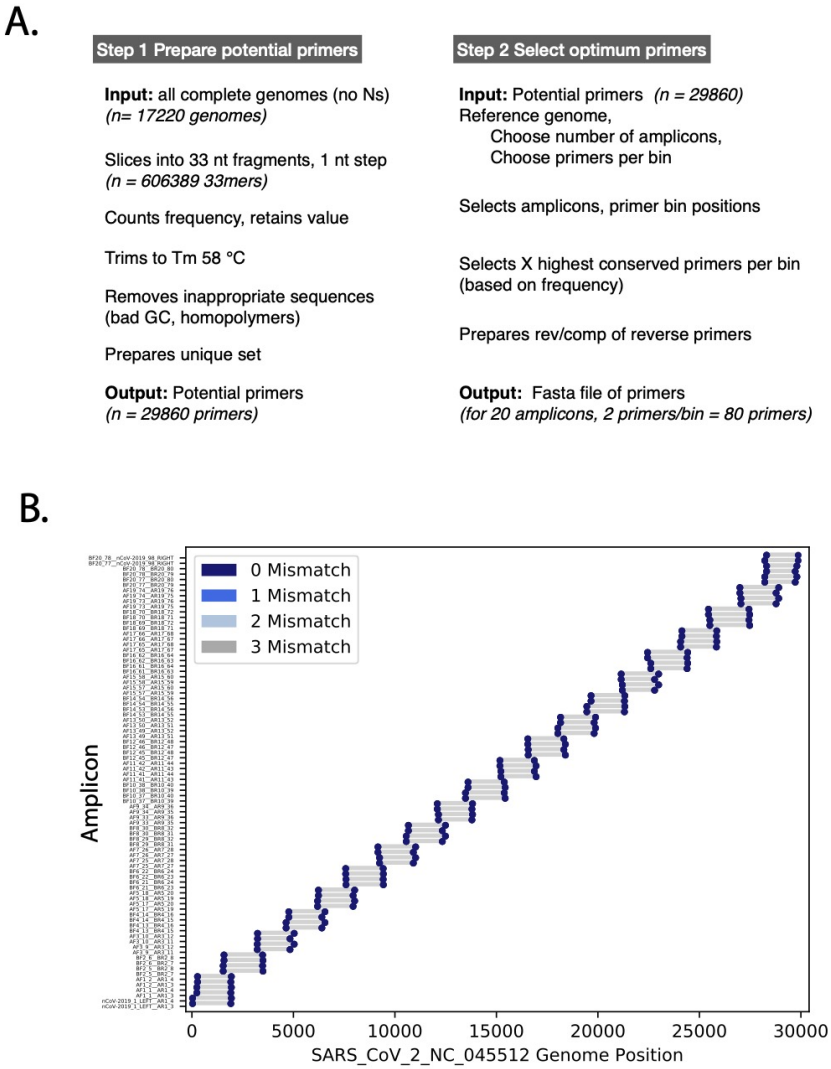


Figure 3. Primer design and amplicon layout. Panel A. The two main steps involving in primers generation and selection are shown. Panel B. The layout of the 20 amplicons across the SARS-CoV-2 genome is shown in lower panel. The blue markers indicate target positions in the SARS-CoV-2 genome (NC_045512 used here), the grey bars indicate the resulting amplicon.

Briefly, the primer design (Figure 3A) started with the set of complete SARS-CoV-2 genome sequences available in the GISAID database on 22 June 2020 (N = 21,687). Spaces and disruptive characters were removed from the sequence IDs and the sequences were further screened to remove genomes containing gaps of 6Ns or more, resulting in 17,220

145 clean genome sequences. Next, all sequences were sliced into 33nt strings (33mers), with a
146 1nt step and 606,389 unique 33mers were generated. The frequency of each 33mer was
147 counted to identify highly conserved 33mers. This counting method avoids the multiple
148 sequence alignment step commonly used in primer design and becomes prohibitive with large
149 and or diverse genome sets. This alignment-free approach allowed us to use all suitable
150 genome sequences of interest rather than a set that could be conveniently aligned. Finally,
151 primer-like 33mers sequences were generated by trimming the sequences to a calculated
152 desired melting temperature and removing any primers greater than 26nt.

153 In the second step we defined forward and reverse primer target regions (bins)
154 for the amplicons. For SARS-CoV-2, we selected 20 amplicons with an overlap of
155 300nt, regularly spaced across the SARS-CoV-2 genome sequence (Figure 3B). We
156 then selected the top conserved primer sequences (the highest frequency primers)
157 mapping in the 5' or 3' 185nt of each amplicon. For security, the two highest frequency
158 primers per bin were selected for the SARS-CoV-2 sequence, this provided some
159 insurance against primer failure either due to target evolution or unexpected
160 secondary structure. The binning and primer target locations for the final set of primers
161 are shown in Figure 2 and the final calculated amplicon lengths were 1495-2093nt. In
162 the second step we defined forward and reverse primer target regions (bins) for the
163 amplicons. For SARS-CoV-2, we selected 20 amplicons with an overlap of 300nt,
164 regularly spaced across the SARS-CoV-2 genome sequence (Figure 3B). We then
165 selected the top conserved primer sequences (the highest frequency primers)
166 mapping in the 5' or 3' 185nt of each amplicon. For security, the two highest frequency
167 primers per bin were selected for the SARS-CoV-2 sequence, this provided some
168 insurance against primer failure either due to target evolution or unexpected

169 secondary structure. The binning and primer target locations for the final set of primers
170 are shown in Figure 2 and the final calculated amplicon lengths were 1495-2093nt.
171 The reverse transcription, PCR amplification and library protocols were modified to
172 accommodate the new primers. Important changes to note are the following. Reverse
173 transcription was performed using the reverse primers and reverse transcription at 42°C. The
174 PCR cycling conditions (using Phusion enzyme) were adjusted for the new T_ms and an
175 increased elongation time required for the longer PCR products. Finally, the library purification
176 steps were adjusted to recover longer PCR and library products. A detailed step-by-step
177 protocol is provided in the supplementary material.
178
179 **Testing the performance of primers to sequence SARS-CoV-2 using MinION**
180 We tested the Entebbe primers performance for sequencing SARS-CoV-2 from
181 nucleic acid extracted from positive samples. The amplicon sizes and genome coverage are
182 summarised in Figure 4. In particular, panel A and B (Figure 4) illustrate the amplicon
183 products after the reverse transcription and PCR amplification with the expected sizes of
184 1400-2000 bp. These amplicons were then pooled and used for library preparation using the
185 MinION sequencing kits SQK-LSK109. Final libraries were quantified and sequenced using
186 a MinION Flow Cell (R9.4.1). The resulting read data, after quality and primer and adapter
187 trimming, were then mapped to the SARS-CoV-2 Wuhan1 reference genome NC_045512
188 (Figure 4, panels C and D) to document sequence coverage across the genome. The twenty
189 individual amplicons are detected in the coverage pattern with small peaks appearing where
190 amplicons overlap. The coverage is consistent across the genome with no missing amplicons
191 and the data were readily assembled into good coverage full genomes. Initial experiments
192 showed that amplicons 2 (spanning nt 2,400) and 16 (spanning nt 23,500) had reduced
193 yields (Figure 4, panel C). The primer mixes were subsequently adjusted to increase
194 concentrations of amplicon 1 and 16 primers for reverse transcription and PCR (see detailed

protocol in Supplementary materials), this improved yields relative to the other amplicons (Figure 4 panel D).

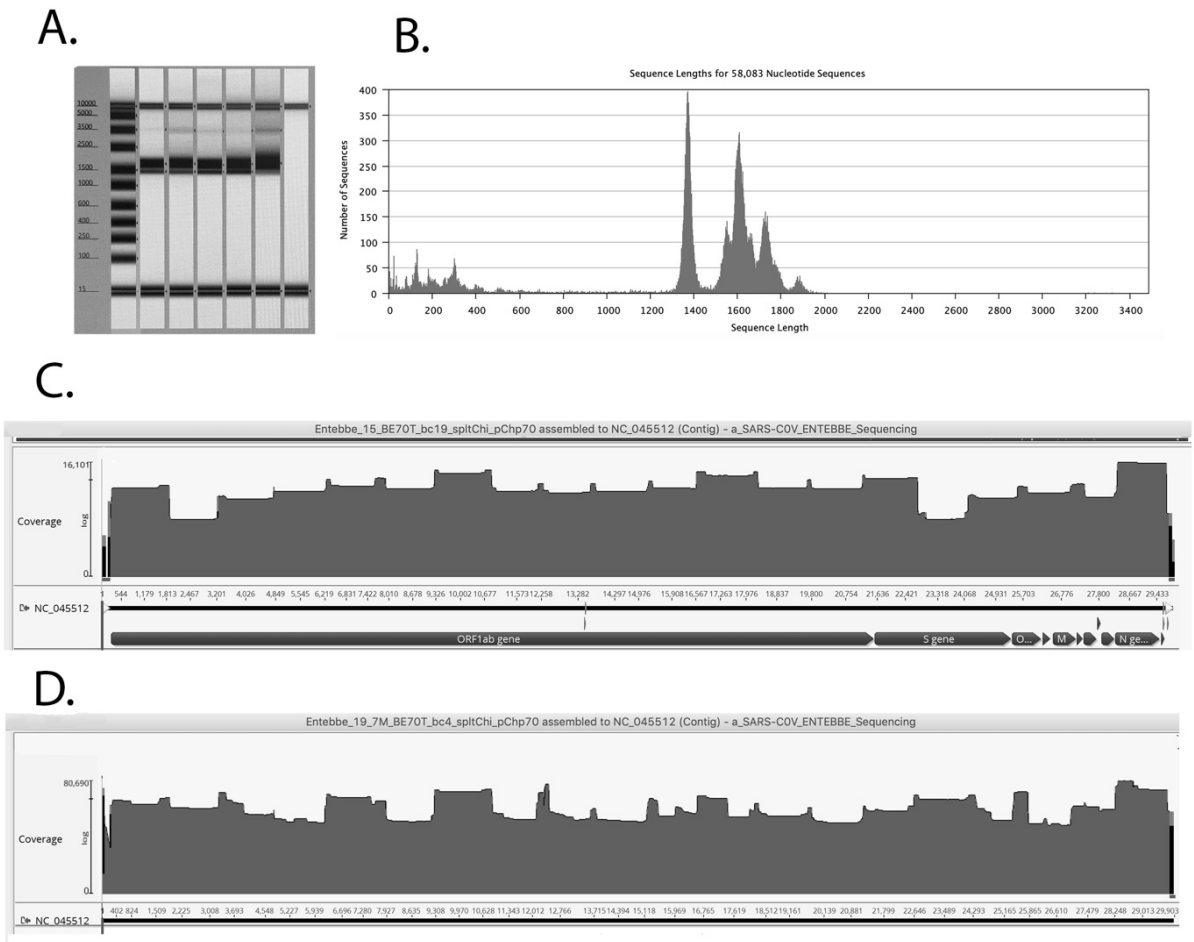


Figure 4. Testing the primer performance. Panel A, PCR product size after pooling of reaction A and B. Expected sizes of amplicons are from 1500 bp to 2093 bp before primer trimming. Panel B, MinION reads after quality control, primer, adapter trimming. Panel C, Reads mapped to SARS-CoV-2 reference genome, before amplicon 2 and 16 primer boosting. Panel D, Reads mapped to SARS-CoV-2 reference genome, after amplicon 2 and 16 primer boosting.

Conclusions

Given the urgency of controlling the SARS-CoV-2 pandemic and the importance of having good quality SARS-CoV-2 genomes, we are providing these alternative primers (the

208 Entebbe primers) with detailed step-by-step laboratory protocols to the community with the
209 hope that they benefit from the new design. The costs and efforts of sequencing SARS-CoV-
210 2 in the large numbers that are currently being seen is substantial and if these new primers
211 result in a higher proportion of gap-less genomes, this will provide added value and will
212 increase the utility of the resulting data.
213
214 **Acknowledgements.**

215 We thank all global SARS-CoV-2 sequencing groups for their open and rapid sharing
216 of sequence data and GISAID for providing an effective platform for making these data
217 available. We are grateful to the Oxford Nanopore Technologies and the ARTIC Network for
218 their support with protocols and analysis software.

219 We acknowledge the support of the Uganda Ministry of Health and its COVID-19
220 Scientific Advisory Committee, the National COVID-19 Task Force, and the staff of the
221 Emerging and Remerging Infections Department of the Uganda Virus Research Institute, and
222 the US Centers for Disease Control and Prevention. The SARS-CoV-2 diagnostic and
223 sequencing award is jointly funded by the UK MRC and the UK DFID under the MRC–DFID
224 Concordat agreement (grant agreement no. NC_PC_19060) and is also part of the European
225 and Developing Countries Clinical Trials Partnership 2 program supported by the European
226 Union. The diagnostics were also supported by the UK Research and Innovation/MRC, the
227 Global Fund, the Government of Uganda, the Islamic Development Bank, the World Health
228 Organization, GAVI, the US Centers for Disease Control and Prevention, and the Jack Ma
229 Foundation, among others. M.V.T.P. was supported by a Marie Sklodowska-Curie Individual
230 Fellowship, funded by European Union's Horizon 2020 research and innovation programme
231 (grant agreement no. 799417). The Uganda Medical Informatics Centre high performance
232 computer was supported by the UK MRC (grant no. MC_EX_MR/L016273/1) to P.K. The study

is supported by a Wellcome Epidemic Preparedness–Coronavirus grant, jointly funded by the Wellcome Trust and UK DFID (grant agreement no. 220977/Z/20/Z) awarded to M.C.

References

1. Edward C. Holmes, Yong-Zhen Zhang EC. Initial genome release of novel coronavirus http://virological.org/t/initial-genome-release-of-novel-coronavirus/319. Virological.org [Internet]. 2020; Available from: http://virological.org/t/initial-genome-release-of-novel-coronavirus/319

2. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. N Engl J Med. 2020 Jan 29;NEJMoa2001316.

3. Yang X, Yu Y, Xu J, Shu H, Xia J, Liu H, et al. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. Lancet Respir Med. 2020 Feb;S2213260020300795.

4. Lauren Gardner et al. Coronavirus COVID-19 Global Cases by Johns Hopkins CSSE. 2020; Available from: https://www.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6

5. Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data – from vision to reality. Eurosurveillance. 2017 Mar 30;22(13):30494.

6. Alessandrini F, Caucci S, Onofri V, Melchionda F, Tagliabracci A, Bagnarelli P, et al. Evaluation of the Ion AmpliSeq SARS-CoV-2 Research Panel by Massive Parallel Sequencing. Genes. 2020 Aug 12;11(8):929.

7. Nicola De Maio, Conor Walker, Rui Borges, Lukas Weilguny, Greg Slodkowicz, Nick Goldman. Issues with SARS-CoV-2 sequencing data. Virological.org [Internet]. 2020; Available from: https://virological.org/t/issues-with-sars-cov-2-sequencing-data/473

8. Page AJ, Mather AE, Le Viet T, Meader EJ, Alikhan N-FJ, Kay GL, et al. Large scale sequencing of SARS-CoV-2 genomes from one region allows detailed epidemiology and enables local outbreak management [Internet]. Epidemiology; 2020 Sep [cited 2020 Oct 9]. Available from: http://medrxiv.org/lookup/doi/10.1101/2020.09.28.20201475

9. Itokawa K, Sekizuka T, Hashino M, Tanaka R, Kuroda M. Disentangling primer interactions improves SARS-CoV-2 genome sequencing by the ARTIC Network's multiplex PCR [Internet]. Microbiology; 2020 Mar [cited 2020 Oct 9]. Available from: http://biorxiv.org/lookup/doi/10.1101/2020.03.10.985150

10. Josh Quick and Nick Loman. nCoV-2019 Version 3 Amplicon Releas. 2020; Available from: https://community.artic.network/t/ncov-2019-version-3-amplicon-release/19

11. Cotten M, Lam TT, Watson SJ, Palser AL, Petrova V, Grant P, et al. Full-Genome Deep Sequencing and Phylogenetic Analysis of Novel Human Betacoronavirus. Emerg Infect Dis [Internet]. 2013 May [cited 2020 Aug 29];19(5). Available from: http://wwwnc.cdc.gov/eid/article/19/5/13-0057_article.htm

12. Cotten M, Petrova V, Phan MVT, Rabaa MA, Watson SJ, Ong SH, et al. Deep Sequencing of Norovirus Genomes Defines Evolutionary Patterns in an Urban Tropical Setting. J Virol. 2014 Oct 1;88(19):11056–69.

13. Agoti CN, Otieno JR, Munywoki PK, Mwihuri AG, Cane PA, Nokes DJ, et al. Local Evolutionary Patterns of Human Respiratory Syncytial Virus Derived from Whole-Genome Sequencing. Perlman S, editor. J Virol. 2015 Apr 1;89(7):3444–54.

14. Phan MV, Murad SD, van der Eijk AA, Metselaar HJ, Hartog H, Harinck F, et al. Genomic sequence of yellow fever virus from a Dutch traveller returning from the Gambia-Senegal region, the Netherlands, November 2018. Eurosurveillance [Internet]. 2019 Jan 24 [cited 2020 Aug 29];24(4). Available from: https://www.eurosurveillance.org/content/10.2807/1560-7917.ES.2019.24.4.1800684

15. Oxford Nanopore Technologies. Nanopore Protocol PCR tiling of COVID-19 virus. 2020; Available from: https://community.nanoporetech.com/protocols/pcr-tiling-ncov/v/PTC_9096_v109_revH_06Feb2020

16. Josh Quick. nCoV-2019 sequencing protocol. 2020; Available from: https://www.protocols.io/view/ncov-2019-sequencing-protocol-bbmuik6w

17. Hsin-Ru Chang, Jay Crespo, Kathy Lee, Kathleen McGall, Susan MacWhorter, Paul Russell, Jeff Schatz, Donna Seid, Criss Walworth, Joe Yu. TaqMan® Assays Shipped at Ambient Temperature Reduce Environmental Impact and Retain Their Quality and Stability. Available from: https://assets.thermofisher.com/TFS-Assets/LSG/brochures/cms_081489.pdfhttps://assets.thermofisher.com/TFS-Assets/LSG/brochures/cms_081489.pdf

298 **Supplementary Materials: 1. Detailed protocol. 2. Primer table.**

299 We acknowledge the Oxford Nanopore Technologies and the ARTIC Network for their
300 extensive3 protocols which provided important background details for the method presented
301 here (15)(16). Potential users are strongly advised to study the original protocos carefully.
302 The protocol presented here assumes the users has all the ARTIC and ONT background
303 experience.

304

305 **Detailed protocol for reverse transcription, PCR and sequencing library preparation.**

306 **Practical comments.** Primers were ordered and delivered dry and were dissolved at
307 100 µM (= 100 pmol/µl) in PCR quality water. Dry shipped primers are reported to have better
308 stability at room temperature (17). Ideal the primers were allowed to dissolve in PCR grade
309 water for at least 3 hours at room temperature or overnight at 4 °C before diluting for use (see
310 below).

311

312 **Prepare the following primer mixes:**

313 **RPM_A**: 21 primers. 3 µl of 100 µM stock of the following Reverse primers plus PCR
314 grade water to 200 µl (= 300 pmol x 21 = 6300 pmol/200 µl = 32 pmol/µl. 2 µl per 20 µl RT
315 reaction = 3.2 pmol/µl = 3.2 µM concentration in reaction) .

316 **Primers in RPM_A**: nCoV-2019_1_LEFT, AR1_3, AR1_4, AR3_11, AR3_12,
317 AR5_19, AR5_20, AR7_27, AR7_28, AR9_35, AR9_36, AR11_43, AR11_44, AR13_51,
318 AR13_52, AR15_59, AR15_60, AR17_67, AR17_68, AR19_75, AR19_76.

319 **RPM_B**: 21 primers. 3 µl of 100 µM stock of the following Reverse primers plus PCR
320 grade water to 200 µl. (= 300 pmol x 21 = 6300 pmol/200 µl = 32 pmol/µl. 2 µl per 20 µl RT
321 reaction = 3.2 pmol/µl = 3.2 µM concentration in reaction). **Primer Boost:** For Primers BR2_7,
322 BR2_8, BR16_63, BR16_64 use 9 µl of 100 µM stock.

323 **Primers in RPM_B:** BR2_7, BR2_8, BR4_15, BR4_16, BR6_23, BR6_24, BR8_31,
324 BR8_32, BR10_39, BR10_40, BR12_47, BR12_48, BR14_55, BR14_56, BR16_63,
325 BR16_64, BR18_71, BR18_72, BR20_79, BR20_80, nCoV-2019_98_RIGHT.

326

327 **PCR primer mix A (PPM_A):** 1.5 µl of each primer (forward and reverse primer) plus
328 439 µl PCR grade water. = 41 primers (100 pmol/µl) = 4100 pmol in 500 µl water = 8.2 pmol/µl
329 in the PPM. When 2 µl PPM used per 25 µl PCR reaction = 0.66 pmol/µl in reaction (= 660
330 nM).

331 **Primers in PPM_A:** nCoV-2019_1_LEFT, AF1_1, AF1_2, AR1_3, AR1_4, AF3_9,
332 AF3_10, AR3_11, AR3_12, AF5_17, AF5_18, AR5_19, AR5_20, AF7_25, AF7_26, AR7_27,
333 AR7_28, AF9_33, AF9_34, AR9_35, AR9_36, AF11_41, AF11_42, AR11_43, AR11_44,
334 AF13_49, AF13_50, AR13_51, AR13_52, AF15_57, AF15_58, AR15_59, AR15_60,
335 AF17_65, AF17_66, AR17_67, AR17_68, AF19_73, AF19_74, AR19_75, AR19_76 .
336 **PCR primer mix B (PPM_B):** 1.5 µl of each primer (forward and reverse primer) plus
337 439 µl PCR grade water = 41 primers (100 pmol/µl) = 4100 pmol in 500 µl water = 8.2 pmol/µl
338 in the PPM. When 2 µl PPM used per 25 µl PCR reaction = 0.66 pmol/µl in reaction (= 660
339 nM). **Primer Boost:** for primers BF2_5, BF2_6, BR2_7, BR2_8, BF16_62, BF16_61,
340 BR16_63, BR16_64, use 4.5 µl of 100 µM stock
341
342 **Primers in PPM_B:** BF2_5, BF2_6, BR2_7, BR2_8, BF4_13, BF4_14, BR4_15,
343 BR4_16, BF6_21, BF6_22, BR6_23, BR6_24, BF8_29, BF8_30, BR8_31, BR8_32, BF10_37,
344 BF10_38, BR10_39, BR10_40, BF12_45, BF12_46, BR12_47, BR12_48, BF14_53,
345 BF14_54, BR14_55, BR14_56, BF16_61, BF16_62, BR16_63, BR16_64, BF18_69,
346 BF18_70, BR18_71, BR18_72, BF20_77, BF20_78, BR20_79, BR20_80, nCoV-
347 2019_98_RIGHT.
348
349 **1. Set up Reverse Transcription reactions (two reactions per sample).**
350
351 Mix the following components in a 0.2mL 8-strip tube, 1 RT_A and 1 RT-B reaction per
352 sample.
353 Component Volume (per sample)
354 **RT_A_Reaction**
355 **Component** **Volume**
356 RPM_A 1 µl
357 Template RNA 11µl (if using less than 11 µl viral RNA adjust to 11µl
358 with water)
359 Total Volume 12µl
360
361 **RT_B_Reaction**
362 **Component** **Volume**
363 RPM_B 1 µl

Template RNA	11µl (if using less than 11 µl viral RNA adjust to 11µl with water)
Total Volume	12µl

Heat at 65 °C for 5:00 mins, and quickly chill in ice/water bath for at least 1:00 min.

Viral RNA input from a clinical sample should be between Ct 18-35. If Ct is between 12-15, then dilute the sample 100- fold in water, if between 15-18 then dilute 10-fold in water. This will reduce the likelihood of PCR-inhibition.

Reverse Transcription

Add the following to the 12 µl annealed template RNA sample:

Component	**Volume (1 reaction)**
SuperScript IV 5X Buffer	4 µl
100mM DTT	1 µl
RNaseOUT RNase Inhibitor	1 µl
10mM dNTPs mix (10mM each)	1µl
SuperScript IV	1 µl
(add 8 µl mix per reaction)	
Total Volume	20 µl

2. Incubate the RT reaction as follows:

42 °C 50 minutes

70 °C 10 minutes

Hold at 4 °C

3. Set up Amplicon PCR

In the Mastermix hood set up the multiplex PCR reactions as follows in 0.2mL 8-strip PCR tubes. Prepare one PPM_A and one PPM_B mix per sample.

Component	**1 reaction**	**8rxns_ PPM_A_Mix**	**8rxns_ PPM_B_Mix**
5 x Phusion HF buffer	5µl	40 µl	40 µl
10 mM dNTPs	0.5µl	4 µl	4 µl
Phusion DNA Pol	0.5µl	4 µl	4 µl
Primer PPM_A or PPM_B (10µM)	2 µl	16 µl	16 µl

Nuclease-free water	12 µl	96 µl	96 µl
Total	20 µl	per rxn	

Aliquot 20 µl per tube PPM_A_Mix or PPM_B_Mix

In the extraction and sample addition cabinet add **5µl cDNA** to each tube and mix well by pipetting.

4. Run PCR

Step	**Cycles**	**Temp**	**Time**
Heat Activation	1	98 °C	30sec
Denaturation	35	98 °C	15sec
Annealing	35	63 °C	30sec
Extension	35	72 °C	3 min
Polish	1	72 °C	10 min
Hold		4 °C	infinite

In post-PCR cabinet: After PCR, for each sample, pool the two reactions in a 1.5 ml microcentrifuge tube.

5. Clean up using AMPure XP beads, Agencourt.

Add an equal volume (1:1) of AMPure XP beads to the sample tube and mix gently by either flicking or pipetting. E.g. add 50 µl AMPure XP bead suspension to a 50 µl reaction.

Pulse centrifuge to collect all liquid at the bottom of the tube.

Incubate for 5 min at room temperature.

Place on magnetic rack and for 2 min

Carefully remove and discard the supernatant, being careful not to touch the bead pellet.

Ethanol wash 1: Add 200µl of room-temperature 70%volume ethanol to the pellet.
Carefully remove and discard ethanol, being careful not to touch the bead pellet.
Repeat Ethanol wash
Pulse centrifuge to collect all liquid at the bottom of the tube and carefully remove as much residual ethanol as possible using a P10 pipette.

436 Open lid, allow to dry briefly (2 min)

437

438 Resuspend pellet in 30µl Elution Buffer (EB, QIAGEN)

439 Mix gently by flicking,

440 Incubate 5 min

441

442 Place on magnet and transfer sample (supernatant) to a clean 1.5mL Eppendorf tube

443 ensuring no beads are transferred into this tube.

444

445 **6. Quantitate** (A Tapestation 4200 analyzer with the D5000 assay kit as well as the Qubit

446 fluorometer were used.)

447

448 **7. Add Barcodes (Steps 7 onward derived from the Oxford Nanopore Technologies**

449 **protocol (15)).**

450 Set up the following reaction for each sample:

451 **Component**	**Volume**
452 DNA amplicons	5 µl*
453 Nuclease-free water	7.5 µl
454 Ultra II End Prep Reaction Buffer	1.75 µl
455 Ultra II End Prep Enzyme Mix	0.75 µl
456 Total Volume	15 µl

457

458 *Volumes usually varies based on the DNA concentration. For the 1500bp amplicons and
459 pool size of appx 8 we have input 50ng and for a pool size of 12 we have input 20ng

460

461 Incubate at room temp (20°C) for 20 min

462 Incubate at 65 °C for 5:00 min

463 Incubate on ice for 1 min

464

465 E7645S/L NEBNext Quick Ligation Module

466 **Component**	**Volume**
467 End repaired sample	15µl
468 Barcode vol	2.5 µl
469 Ultra II Ligation Master Mix	17.5 µl
470 Ligation Enhancer	0.5 µl
471 Water	1.5 µl
472 final volume	35.5 µl

473

474 Incubate in a thermocycler for 15 minutes at 20°C with heated lid set to 30°C (or off).
475 Incubate at 70 °C for 10 min
476 Incubate on ice for 1 min
477
478 The 70°C incubation is to inactivate the DNA ligase to prevent barcode cross-ligation when
479 reactions are pooled in the next step.
480
481 Pool all barcoded fragments together into a new 1.5 ml Eppendorf tube.
482
483 **8. Clean up on AMPure XP beads, Agencourt.**
484
485 Add an equal volume (1:1) of SPRI beads to the sample tube and mix gently by either flicking
486 or pipetting. E.g. add 50 µl SPRI beads to a 50 µl reaction.
487
488 Pulse centrifuge to collect all liquid at the bottom of the tube.
489
490 Incubate for 5 min at room temperature.
491
492 Place on magnetic rack and for 5 min
493
494 Carefully remove and discard the supernatant, being careful not to touch the bead pellet.
495
496 **Ethanol wash 1:** Add 200µl of room-temperature 70%volume ethanol to the pellet.
497 Carefully remove and discard ethanol, being careful not to touch the bead pellet.
498 **Repeat Ethanol wash**
499
500 Pulse centrifuge to collect all liquid at the bottom of the tube and carefully remove as much
501 residual ethanol as possible using a P10 pipette.
502 Open lid, allow to dry briefly (2 min)
503
504 Resuspend pellet in 30µl Elution Buffer (EB, QIAGEN)
505 Mix gently by flicking
506
507 Incubate 5 min
508

509 Place on magnet and transfer sample (supernatant) to a clean 1.5mL Eppendorf tube
510 ensuring no beads are transferred into this tube.
511
512 **9. Quantitate as above.**
513
514 **10. Add AMII adapters.**
515
516 AMII adapters ligation reaction:

517 **Component**	**Volume**
518 Barcoded amplicon pools	30 µl
519 NEBNext (E60562)	
520 Quick Ligation Reaction Buffer (5X)	10 µl
521 AMII adapter mix	5 µl
522 Quick T4 DNA Ligase	5 µl
523 Total	50 µl

524
525 The input of barcoded amplicon pools will depend on the number of barcoded pools and
526 should be between 40 ng (8 barcodes) and 160 ng (24 barcodes).
527
528 Incubate at room temperature for 15:00 min
529
530 **11. Clean up on AMPure XP beads, Agencourt.**
531
532 **USE SFB instead of Ethanol for two washes ***CRUCIAL,** Ethanol strips off motor protein
533 from adapter. You don't want to do this!
534
535 Clean up on AMPure XP beads, Agencourt
536
537 Add an equal volume (1:1) of SPRI beads to the sample tube and mix gently by either flicking
538 or pipetting. E.g. add 50 µl SPRI beads to a 50 µl reaction.
539
540 Pulse centrifuge to collect all liquid at the bottom of the tube.
541
542 Incubate for 5 min at room temperature.
543

544 Place on magnetic rack and for 2 min
545
546 Carefully remove and discard the supernatant, being careful not to touch the bead pellet.
547
548 **SFB washes:** Add 200µl of room-temperature **SFB** to the pellet.
549 resuspend
550 collect on magnet
551 Carefully remove and discard **SFB** being careful not to touch the bead pellet.
552
553 **Repeat SFB wash**
554
555 Pulse centrifuge to collect all liquid at the bottom of the tube and carefully remove as much
556 residual **SFB** as possible using a P10 pipette.
557
558 **Resuspend pellet in 15 µl Elution Buffer (EB, QIAGEN)**
559 Mix gently by flicking,
560 **Incubate 5 min**
561 Place on magnet and transfer sample (supernatant) to a clean 1.5mL Eppendorf tube
562 ensuring no beads are transferred into this tube.
563
564 **12. Quantitate.**
565
566 **13. Proceed to MinION sequencing.**
567

568 **Supplementary material Table 1. Entebbe Primers**

Amplicon_no	Primer_id	Sequence[1]	Position[2]
1	AF1_1	GTCCGGGTGTGACCGAAAG	242
1	AF1_2	CCTTGTCCCTGGTTTCAACGA	274
1	AR1_3	TGCGGGAGAAAATTGATCGTACA	1894
1	AR1_4	GCACAGAATTTTGAGCAGTTTCA	1921
2	BF2_5	ACGTGCTAGCGCTAACATAGG	1540
2	BF2_6	GTTGGAGAAGGTTCCGAAGGT	1580
2	BR2_7	TAGCCTTATTTAAGGCTCCTGCAA	3477
2	BR2_8	TGCAACACCTCCTCCATGTTTAA	3459
3	AF3_9	ACAAACTGTTGGTCAACAAGACG	3220
3	AF3_10	CAACAAGACGGCAGTGAGGA	3233
3	AR3_11	TTGTGTAGATTGTCCAGAATAGGACC	4806
3	AR3_12	TCCATATGTCATTGACATGTCCACA	5014
4	BF4_13	TTTGGAAGAAGCTGCTCGGTATAT	4639
4	BF4_14	GAAACCATCTCACTTGCTGGTTC	4772
4	BR4_15	GGTTTTAGATCTTCGCAGGCAAG	6380
4	BR4_16	CCATTAGATCTGTGTGGCCAA	6534
5	AF5_17	CACACCCTCTTTTAAGAAAGGAGCTA	6190
5	AF5_18	TTGTTTGGCATGTTAACAATGCAAC	6234
5	AR5_19	CAGACGCTGATTTTGCAGATGA	7919
5	AR5_20	GCACTATCACCAACATCAGACAC	7994
6	BF6_21	AGGCTTTTGCAAACTACACAATTG	7591
6	BF6_22	TATGCTAATGGAGGTAAAGGCTTTTG	7574
6	BR6_23	CGATAGCTACAATACCACCAGCT	9409
6	BR6_24	CAATACCACCAGCTACTATAGATGCT	9397
7	AF7_25	AAGCTGGTGTTTGTGTATCTACTAGT	9243
7	AF7_26	ACCTACCTTGAAGGTTCTGTTAGAG	9164
7	AR7_27	CACTACCCAATATGGTACGTCCA	10885
7	AR7_28	TGAGTAACAACCAGTGGTGTGT	11000
8	BF8_29	CTGGAGTTCATGCTGGCACA	10560
8	BF8_30	GTTTTAGCTTGGTTGTACGCTG	10664
8	BR8_31	TTTGCCCTCTTGTCCTCAGATCTA	12313
8	BR8_32	TGGTATGACAACCATTAGTTTGGCT	12466
9	AF9_33	CTCAAGAAGCTTATGAGCAGGCT	12144
9	AF9_34	CAAGCTATAGCCTCAGAGTTTAGTTC	12089
9	AR9_35	ACGTTGACGTGATATATGTGGTACC	13770
9	AR9_36	ACGAGGTCTGCCATTGTGTATT	13802
10	BF10_37	GCGGTGTAAGTGCAGCC	13475
10	BF10_38	TCGCTTCCAAGAAAAGGACGAA	13602
10	BR10_39	CTCAATACTTGAGCACACTCATTAGC	15406
10	BR10_40	GTGACAAGCTACAACACGTTGT	15367
11	AF11_41	ATTCTATGGTGGTTGGCACAACA	15219
11	AF11_42	CAATAGCCGCCACTAGAGGAG	15173
11	AR11_43	TAGTGTAGGTGCACTTAATGGCATT	16932
11	AR11_44	GGTAAACAACAGCATCACCATAGTC	16846
12	BF12_45	AACATGTGACTGGACAAATGCTG	16566
12	BF12_46	ACTGACTTTAATGCAATTGCAACATG	16546
12	BR12_47	TACAGCAACTAGGTTAACACCTGTAG	18374
12	BR12_48	CACCCCTCGACATCGAAGC	18302

13	AF13_49	GTGGCAACTTTACAAGCTGAAAATG	18025
13	AF13_50	GACATACCTGGCATACCTAAGGAC	18160
13	AR13_51	GTTTAATGTTGCGCTTAGCCCAA	19791
13	AR13_52	GTAGTCCCAGATCACAGTATTAGCAG	19859
14	BF14_53	CGTGTATAACACGTTGCAATTTAGGT	19454
14	BF14_54	ACTTTGATGGACAACAGGGTGAA	19661
14	BR14_55	CGCGTGGTTTGCCAAGATAATTA	21285
14	BR14_56	CATAACCATCTATTTGTTCGCGTGG	21301
15	AF15_57	AGCTCATGGGACACTTCGC	21203
15	AF15_58	TTGGAGGTTCCGTGGCTATAAAG	21146
15	AR15_59	CGATTTGTCTGACTTCATCACCTC	22770
15	AR15_60	CTACCGGCCTGATAGATTTCAG	22971
16	BF16_61	TGCATCTGTTTATGCTTGGAACAG	22603
16	BF16_62	CTCTCTCAGAAACAAAGTGTACGTTG	22446
16	BR16_63	CACTTGCTGTGGAAGAAAGTGA	24371
16	BR16_64	GTTGACCACATCTTGAAGTTTTCCAA	24396
17	AF17_65	GGTGATTGCCTTGGTGATATTGC	24074
17	AF17_66	CTGTTTTGCCACCTTTGCTCA	24138
17	AR17_67	CCAGCAAAGAAAATAGTTGGCATC	25816
17	AR17_68	CGTAACAATTAGTATGCCAGCAAAGA	25830
18	BF18_69	TCCCTTTCGGATGGCTTATTGTT	25514
18	BF18_70	TGAAATCAAGGATGCTACTCCTTCAG	25446
18	BR18_71	GTTGTACCTCTAACACACTCTTGGTA	27451
18	BR18_72	CTCACAAGTAGCGAGTGTTATCAG	27418
19	AF19_73	ACGCTTTCTTATTACAAATTGGGAGC	27045
19	AF19_74	CGCTGTGACATCAAGGACCT	26994
19	AR19_75	TGGCAATGTTGTTCCTTGAGGAA	28755
19	AR19_76	CAGCCATTCTAGCAGGAGAAGTT	28885
20	BF20_77	TAGAGTATCATGACGTTCGTGTTGTT	28219
20	BF20_78	ACCCCGCATTACGTTTGGT	28309
20	BR20_79	TGGCTCTTTCAAGTCCTCCCT	29700
20	BR20_80	GCTCTTCCATATAGGCAGCTCTC	29779
	nCoV-2019_1_LEFT[3]	ACCAACCAACTTTCGATCTCTTGT	31
	nCoV-2019_98_RIGHT[3]	TTCTCCTAAGAAGCTATTAAAATCACATGG	29866

Footnotes 1. Sequence listed 5' to 3'.

2. Position in SARS-CoV-2 reference genome GenBank NC_045512

3. From original ARTIC primer set V.1 (16).