

1 A rapid, cost efficient and simple method to identify current SARS-CoV-2
2 variants of concern by Sanger sequencing part of the spike protein gene

3

4 Tue Sparholt Jørgensen¹, Kai Blin¹, Franziska Kuntke², Henrik K. Salling³, Thomas Y
5 Michaelsen⁴, Mads Albertsen⁴, Helene Larsen².

6 Affiliation

7 1 The Novo Nordisk Foundation Center for Biosustainability, Technical University of
8 Denmark - DTU, 2800 Kgs Lyngby

9 2 Centre for Diagnostics, Department of Health Technology, Technical University of
10 Denmark - DTU, 2800 Kgs. Lyngby

11 3 Novo Nordisk A/S, Dept. 609.04 Virology, Hagedornsvej 1, 2820 Gentofte, Denmark

12 4 Department of Chemistry and Bioscience, Aalborg University, 9220 Aalborg, Denmark

13

14 Fig. 1: Workflow

15 Fig. 2: Sensitivity estimation graph

16 Table 1: Primers used

17 Table 2: Theoretical profiling of VoC and non-VoC mutations in our amplicon window

18 Suppl. Fig. 1: PCR primer annealing temperature test

19 Suppl. Table 1: Proof of concept DNA concentration of 24 samples, Ct values and mutation

20 **NOTE:** This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

21 Suppl. Table. 2: Ct values of 195 samples and mutation calls

22 Suppl. Table. 3: GISAID references

23

24 Abstract

25 In 2020, the novel coronavirus, SARS-CoV-2, caused a pandemic, which is still raging at the
26 time of writing this. Many countries have set up high throughput RT-qPCR based diagnostics
27 for people with COVID-19 symptoms and for the wider population. In addition, with the use of
28 whole genome sequencing (WGS) new lineages of SARS-CoV-2 have been identified that
29 have been associated with increased transmissibility or altered vaccine efficacy, so-called
30 Variants of Concern (VoC). WGS is generally too labor intensive and expensive to be
31 applied to all positive samples from the diagnostic tests, and often has a turnaround time too
32 long to enable VoC focused contact tracing. Here, we propose to use Sanger sequencing for
33 the detection of common variants of concern and key mutations in early 2021, using a single
34 set of the recognized ARTIC Network primers. The proposed setup relies entirely on
35 materials and methods already in use in diagnostic RT-qPCR labs and on existing
36 infrastructure from companies that have specialized in cheap and rapid turnaround Sanger
37 sequencing. In addition, we provide an automated mutation calling software
38 (<https://github.com/kblin/covid-spike-classification>). We have validated the setup on 195
39 SARS-CoV-2 positive samples, and we were able to profile >85% of RT-qPCR positive
40 samples, where the last 15% largely stem from samples with low viral count. At
41 approximately 4€ per sample in material cost, with minimal hands-on time, little data
42 handling, and a turnaround time of less than 30 hours, the setup is simple enough to be
43 implemented in any SARS-CoV-2 RT-qPCR diagnostic lab. Our protocol provides results
44 that can be used to focus contact-tracing efforts and it is cheap enough for the tracking and
45 surveillance of all positive samples for emerging variants such as B.1.1.7, B.1.351 and P.1
46 as of January 2021.

47 Keywords

48 SARS-CoV-2, COVID-19, CoV, spike protein, profiling, coronavirus, sequencing

49 Introduction

50 The pandemic caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-
51 2) is a crisis faced by nearly every country in the world at the time of writing this article.

52 Throughout 2020, the spread of SARS-CoV-2 caused millions of deaths, and more than 100
53 million people have been diagnosed with COVID-19, primarily by the use of Real-Time

54 quantitative Polymerase Chain Reaction (RT-qPCR) diagnostic tests. In the RT-qPCR tests,
55 the relative number of viral copies in a sample is estimated by the threshold cycle (Ct) value,

56 where a higher value indicates fewer viral copies. Variants of the virus are identified by

57 whole genome sequencing (WGS) based mutation calls and compared to the originally

58 published genome¹ or by variant specific qPCR tests^{2,3}. Numerous protocols for SARS-CoV-

59 2 WGS exist, with different versions of the ARTIC Network protocols being the most

60 commonly used⁴. WGS derived genomes are deposited in the GISAID database⁵, where

61 researchers can then use them to analyze outbreaks. Tracking new mutations throughout

62 the population has led to the identification of several VoCs of SARS-CoV-2, such as the

63 lineages B.1.1.7⁶, B.1.351⁷, and P.1⁸. The lineage B.1.1.7 carries the spike protein N501Y

64 mutation, and is thought to be more transmissible⁹, and this mutation is also found in the

65 B.1.351 and P.1 variants. The lineages B.1.351 and P.1 in addition are characterized by a

66 E484K mutation, which has been reported to significantly reduce antibody binding¹⁰. This

67 could potentially lower the efficacy of current vaccines. Monitoring VoCs makes it possible to

68 tailor the societal response for maximal containment of SARS-CoV-2 and at the lowest cost.

69 While WGS is crucial for surveillance work, it is not optimally suited for near real-time contact

70 tracing of VoCs, as it is more expensive, requires a centralized setup and a high degree of

71 technical and bioinformatics expertise. Instead, we have developed a cost efficient, rapid,

72 and simple method for SARS-CoV-2 variant mutation typing using no additional equipment

73 to that used for the diagnostic RT-qPCR SARS-CoV-2 test. (Figure 1). At a cost of approx.
74 4€ per reaction in materials, and approximately one and a half hours of hands on time per 96
75 samples (Figure 1), this system can readily be implemented for large scale typing of all
76 positive samples, without the use of specialized equipment. The material cost is several
77 times cheaper than the cost of WGS, regardless of the sequencing platform used⁴. Our
78 method consists of an RT-PCR with a single ARTIC Network primer set, using the same
79 enzyme mix used in diagnostic RT-qPCR tests, and using Sanger sequencing to sequence a
80 1001bp amplicon in the spike protein. Sanger sequencing as a technology is more than 40
81 years old¹¹, and a large existing commercial infrastructure offers Sanger sequencing
82 services. The resulting sequence information from our proposed amplicon area of the spike
83 gene can be used to establish whether a patient has an infection with the current SARS-
84 CoV-2 VoCs or any of the biologically important mutations in the spike gene. Finally, we
85 provide a bioinformatics tool that enables automated basecalling, mapping and calling of
86 mutations from Sanger trace files (.ab1). The results are reported in a table as well as in
87 fastq format, available as a command line interface from [https://github.com/kblin/covid-spike-](https://github.com/kblin/covid-spike-classification)
88 [classification](https://github.com/kblin/covid-spike-classification) or in a browser-based app running locally from [https://ssi.biolib.com/app/covid-](https://ssi.biolib.com/app/covid-spike-classification/run)
89 [spike-classification/run](https://ssi.biolib.com/app/covid-spike-classification/run).

90

91 Methods

92 Purified RNA from the diagnostic RT-qPCR SARS-CoV-2 test facility at the Technical
93 University of Denmark (DTU) was set up with two primers from the ARTIC Network SARS-
94 CoV-2 v3 amplicon set for WGS v3 ([https://github.com/artic-network/artic-](https://github.com/artic-network/artic-ncov2019/blob/master/primer_schemes/nCoV-2019/V3/nCoV-2019.tsv)
95 [ncov2019/blob/master/primer_schemes/nCoV-2019/V3/nCoV-2019.tsv](https://github.com/artic-network/artic-ncov2019/blob/master/primer_schemes/nCoV-2019/V3/nCoV-2019.tsv)) (Table 1). All assays
96 were tested with RPC purified primers from TAG Copenhagen (Copenhagen, Denmark,
97 <http://tagc.dk>). RNAAdvance Viral Reagent Kit - Built on SPRI (Solid Phase Reversible
98 Immobilization) bead-based technology (#C63510 Bechman Coulter, IN, USA), was used for

99 the purification of input RNA. RT-PCR were set up in 20 μ l RT-PCR reactions using 10 μ l One
100 Step PrimeScript III (Takara, Shimogyō-ku, Kyoto) used in the diagnostic SARS-CoV-2 test
101 facility at DTU, 0.4 μ M forward primer, 0.4 μ M reverse primer, and 2-5 μ l purified template
102 RNA per reaction (Table 1). The RT-PCR reaction was performed (after the initial test of
103 annealing temperature) using the following program: reverse transcription for 5 min at 52 °C,
104 then hot start polymerase activation at 95 °C for 10s, followed by 45 cycles of 95 °C for 5s,
105 58 °C for 30 s, and 72 °C for 1 min, followed by 5 min at 72 °C. This protocol is in line with
106 the manufacturer's instruction for a PCR product length of approx. 1kb. The resulting PCR
107 products were analyzed using gel electrophoresis on a 1% agarose gel to confirm
108 amplification (Suppl. Fig. 1). Then 10 μ l PCR product was purified with 0.8x volume of SPRI
109 beads (KAPA Pure Beads, Roche, Basel, Switzerland), and both purified and unpurified
110 PCR product was shipped to Eurofins Genomics (Eurofins, Luxembourg, Luxembourg) for
111 Sanger sequencing using their Europe wide Mix2Seq Overnight Service. Samples were
112 shipped at 15:00 one day and the Sanger sequencing results were received before noon the
113 next day. We expect that any commercial Sanger sequencing provider can be used with the
114 proposed protocol. For the proof of concept test with six samples, the recommended 5 ng/ μ l
115 in 15 μ l was shipped when possible, along with a final concentration of 1.2 μ M left primer
116 diluted in nuclease free water. For the larger scale test of 195 samples, 1.5 μ l of unpurified
117 PCR products were shipped. For RT-PCR product DNA concentration measurement, a Qubit
118 4 fluorometer was used with the dsDNA BR kit (Thermo Fisher Scientific, Waltham,
119 Massachusetts, U.S.). Primer melting temperature was calculated using the Oligo Analyzer
120 from IDT (<https://eu.idtdna.com/calc/analyzer>, Integrated DNA Technologies, Coralville,
121 Iowa, USA). For analyzing the Sanger sequencing data, we set up a workflow consisting of:
122 1) basecalling ab1 files using Tracy¹², followed by 2) mapping of fastq data to the
123 NC_045512.2 SARS-CoV-2 sequence using Bowtie2¹³, and finally 3) mutation calling with
124 Samtools¹⁴. Subsequently, the codons with known mutations are extracted and translated,
125 and the amino acid calls are compared to a list of known mutations (currently N439K,
126 L452R, Y453F, S477N, E484K, N501Y, A570D, Q613H, D614G, A626S, H655Y, Q677H,

127 P681H, P681R, I692V, A701V, T716I). Information is primarily derived from covariant.org¹⁵,
128 and reported as a csv table file, along with the basecalls in fastq format. The program is
129 available from <https://github.com/kblin/covid-spike-classification>. Alternatively, Statens
130 Serum Institut (SSI), the Danish government body for SARS-CoV-2 surveillance, has set up
131 a self-contained web-app running the program without data upload at
132 <https://ssi.biolib.com/app/covid-spike-classification/run>.

133 The protocol for this method has been published at protocols.io at
134 [dx.doi.org/10.17504/protocols.io.bsbdnai6](https://doi.org/10.17504/protocols.io.bsbdnai6)¹⁶.

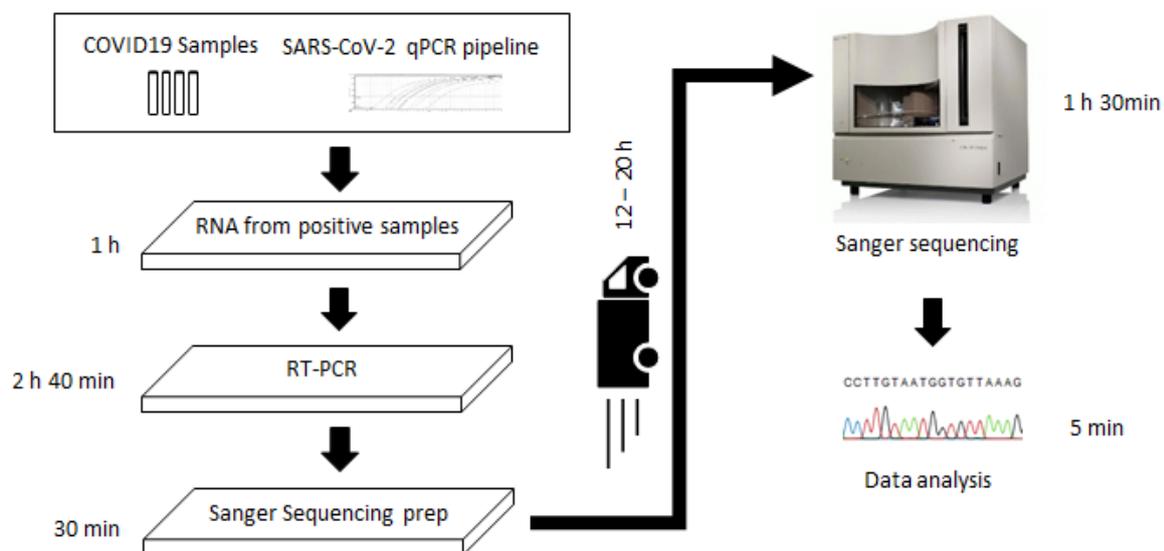
135 The input RNA from the diagnostic SARS-CoV-2 RT-qPCR facility was obtained from
136 oropharyngeal swabs with specimens placed immediately into a tube with 3M Guanidine
137 Thiocyanate for instant lysis of viral particles. This ensures stability of RNA during
138 transportation and during the process of transferring patient material from tube to plate. The
139 RT-qPCR kit is demonstrated on PCR applications and has a Limit of Detection of 1 copy/ μ L
140 for viral RNA (Centre for Diagnostics, DTU). RNA was purified in cooled racks and placed
141 directly into the fridge prior to RT-qPCR reaction set-up. Following the reaction set up, the
142 RNA was frozen at -20 C° prior to being thawed for sequencing within 1-2 days. For the
143 diagnostic SARS-CoV-2 test, two separate primer sets and probes target the SARS-CoV-2
144 Nucleoprotein gene (N1 and N2), and one internal positive control primer set and probe
145 which target the human RNase P. A three-color multiplex probe system from Pentabase
146 (Odense, Denmark) was used.

147 To estimate the profiling potential of the spike gene amplicon, SARS-CoV-2 strain data was
148 collected from GISAID on 2021-01-29. Data was first pre-filtered by host "human" and
149 minimum length > 25000. The data were then further filtered by excluding cases older than
150 the first date at which all three VoCs appear (2020-12-04). This resulted in a dataset of
151 46,822 sequences. The VoCs B.1.1.7, B.1.351 and P.1 were assigned using Pangolin^{17,18}
152 v2.1.7. To evaluate potential false-positive calls, a reference dataset of 5000 non-VoC

153 samples was randomly selected from this dataset. To generate VoC reference data, the
154 prefiltered data was randomly subsampled to a maximum of 1000 samples of each VoC.
155 This led to downsampling of B.1.1.7 to 1000 and all P.1 (n = 40) and B.1.351 (n = 708)
156 cases included. Sequences were assigned amino acid substitutions using Nextclade
157 v0.11.2. To ensure amplicon dropout did not interfere with the evaluation, we removed
158 samples which had ≥ 50 consecutive bases missing within the amplicon region. This
159 resulted in a final dataset of 878 B.1.1.7, 34 P.1, 516 B.1.351, and 4418 non-VoCs. We used
160 the set of hallmark spike amino acid mutations that fall within the range of amino acid
161 positions (position 419 to 759 on the spike protein), which are N439K, E484K, N501Y,
162 A570D, H655Y, P681H, A701V, and T716I, along with the universal D614G mutation. Using
163 the set of amino acid substitutions defined, we were able to compute all possible mutation
164 combinations for each VoC within the amplicon window.

165

166



167

168 Figure 1. Workflow of the proposed Sanger sequence typing of SARS-CoV-2 variant
169 mutations. Time estimates are based on a single 96 well PCR plate and no automation. The
170 total hands on time is one hour 30 minutes, and only requires setting up an RT-PCR and

171 transferring 1.5µl of the product to a new 96 well PCR plate, which is sent for Sanger
 172 sequencing at a commercial provider. The total cost per sample including plastware,
 173 enzymes, transport, and sequencing, is approx. 4 €. After the raw data is received, data
 174 analysis takes less than five minutes for 96 samples. If Sanger sequencing is available in-
 175 house, the time to result could be less than six hours.

176

primer name	primer sequence	Primer melting temperature	Primer end position in NC_045512	Covered positions in the spike protein
nCoV-2019_76_LEFT_al t3	GGGCAAACCTGGAAAGA TTGCTGA	65.4 °C	22822	N439..T716
nCoV-2019_78_RIGHT	TGTGTACAAAACTGCC ATATTGCA	63.9 °C	23823	

177 Table 1. Primers suggested for typing SARS-CoV-2 VoCs. The Primer melting temperature
 178 is based on the use of Primescript III master mix, which has 2mM MgCl₂.

179

	Pangolin lineage			
Substitution	B.1.1.7 (n = 878)	B.1.351 (n=516)	P.1 (n=34)	non-VoC (n=4418)

N439K	0% (0)	0% (0)	0% (0)	7.4% (329)
E484K	0% (0)	99.2% (512)	100% (34)	0.1% (4)
N501Y	99.9% (877)	99.6% (514)	100% (34)	0.4% (16)
A570D	100% (878)	0% (0)	0% (0)	0% (0)
D614G	99.9% (877)	100% (516)	100% (34)	99.8% (4408)
H655Y	0% (0)	0% (0)	100% (34)	0.2% (11)
P681H	99.8% (876)	0.2% (1)	0% (0)	0.7% (31)
A701V	0.1% (1)	100% (516)	0% (0)	0.1% (4)
T716I	99.8% (876)	0% (0)	0% (0)	0.1% (6)
None of the above substitutions	0% (0)	0% (0)	0% (0)	0.2% (9)

180 Table 2. Key mutations in four SARS-CoV-2 variants and the potential to profile them based
181 on a single amplicon in the spike protein. Note that even without the near universal D614G
182 mutation, 100% of the VoC strains (n=1428) have one or more of the mutations, and less
183 than 1% of non VoC strains would be wrongly interpreted as belonging to a VoC. The non-
184 VoC lineage strains with e.g. E484K (four strains out of 4418) are important in their own
185 right, as this mutation is thought to lower the current vaccine efficacy and this warrants extra
186 contact tracing focus.

187

188 Results & Discussion

189 Classification of SARS-CoV-2 genomes is often performed using Nextclade¹⁹ or Pangolin¹⁸
190 which uses several lineage defining mutations, but only a subset of mutations is needed as
191 this allows profiling of incomplete genomes. Using 9 key amino-acid mutations included in a
192 single 1001 bp amplicon in the spike protein gene (Table 2) from a total of 1428 VoC strains
193 and 4418 non-VoC strains, we get an overview of the ratio of individual mutations for each of
194 the B.1.1.7, B.1.351, P.1, and non-VoC strains (Accession numbers and references are
195 stated in Suppl. Table 3). While nearly all strains shared the D614G mutation (99.8% of all
196 strains), several mutations were exclusively or predominantly found in the VoC strains. The
197 mutation N501Y was found in 99.9% of B.1.1.7, in 99.6% of B.1.351, and in 100% of P.1 of
198 VoC classified strains (Table 2), while only present in 0.4% of non-VoC classified strains.
199 Similarly, the mutation E484K was found in 99.2% of B.1.351 variants and in 100% of P.1
200 variants, but not in the included set of B.1.1.7 strains and in less than 0.1% of non-VoC
201 strains. Several additional positions predominantly occur in VoC strains as shown in Table 2.
202 We propose to use the Sanger sequence information in this amplicon to report and confirm
203 individual mutations, which is sufficient to focus potentially limited contact tracing resources
204 on patients who harbor VoCs, and it will provide enough information for surveillance of
205 emerging variants.

206 The proposed workflow for Sanger sequencing of SARS-CoV-2 positive samples uses a
207 minimum of plastware and includes very few hands on steps (Figure 1). First, purified RNA
208 from RT-qPCR positive samples are transferred to a 96 well plate. Then, the RT-PCR
209 master mix is prepared in another 96 well plate and template is added. After RT-PCR, 1.5µl
210 from each well is transferred to a PCR plate containing the Sanger sequencing primer and
211 the plate is then ready for sequencing. The whole process requires only 3 pipette tips per
212 sample, and approx. an hour and a half of hands on time (1 hour for RT-PCR preparation

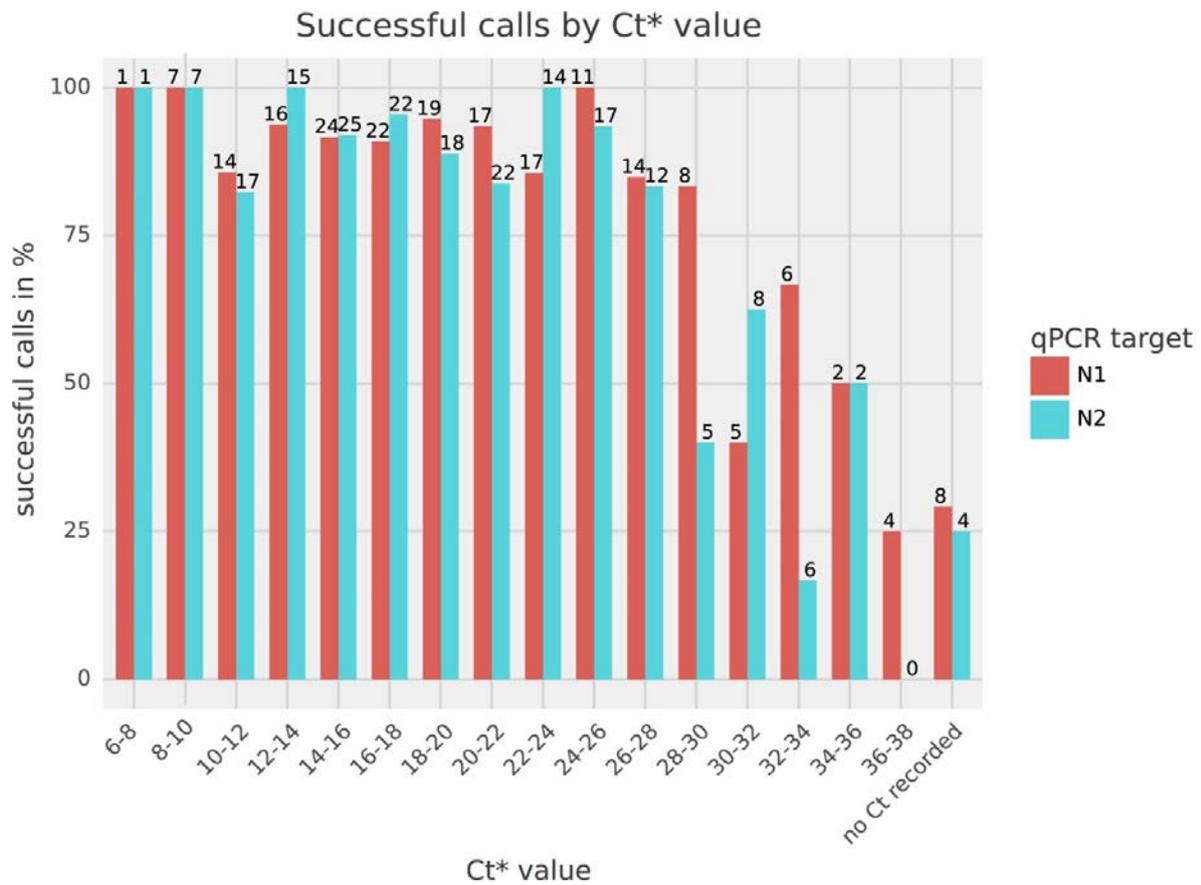
213 and 30 minutes for preparation of samples for sequencing), along with the two hour and 40
214 minutes run-time for the RT-PCR per 96 well PCR plate. The workflows proposed here are
215 easy to automate. We estimate that the total price of these simple steps is 4€ per sample in
216 material cost including plastware, enzymes, buffers, shipping, and sequencing, making the
217 workflow cost efficient compared to WGS or multi-target RT-qPCR typing approaches. If
218 Sanger sequencing can be performed in-house the time to result can be less than six hours
219 from positive COVID-19 sample to variant call, which is faster than any high throughput
220 WGS based analysis, with less computational workload and data storage.

221 To evaluate the suitability of the Sanger sequencing method for mutation calling in the RT-
222 PCR product, we initially used six SARS-CoV-2 positive RNA samples from the large scale
223 testing facility at DTU. We shipped both SPRI purified and unpurified RT-PCR products for
224 Sanger sequencing (Table 1, Suppl Figure 1). In this way, we got two Sanger reads from
225 each of the 6 samples, which we used to verify that the mutations are called identically
226 regardless of purification status. Suppl. Table 1 shows an overview of the 12 Sanger reads.
227 We tested a custom made primer set with similar results (Suppl. Table 1). A single sample
228 had mutation calls identical to B.1.1.7 variants in all four reads, as it had the key mutations
229 N501Y, A570D, and P681H in all four Sanger reads. This successful proof of concept study
230 shows that the single ARTIC primer set can be used for typing current SARS-CoV-2 VoC
231 key defining mutations and that unpurified RT-PCR product can be used, significantly
232 reducing the workload involved in mutation typing. Following RT-PCR and shipping of a
233 larger batch of 195 samples for Sanger sequencing, we received the results within 20 hours,
234 which is routine delivery time for most large commercial providers. In Figure 2, the 195
235 samples have been binned according to the diagnostic RT-qPCR Ct value for both the N1
236 and N2. Note that the initial seven cycles are not fluorescence recorded and therefore do not
237 count towards the final Ct value in the RT-qPCR set up used at the DTU testing facility.
238 Overall, the success rate of Sanger sequencing of the RT-qPCR positive samples was
239 85.1% (166/195 Sanger reads mapped to the correct position of the SARS-CoV-2 spike

240 gene). If failed samples were rerun or duplicate PCR reactions implemented, we expect that
241 a slightly higher fraction of samples would yield mutation calls. The fraction of samples with
242 a mappable result from Sanger sequencing is stable at approx. 90% below Ct 30 (Figure 2,
243 note the seven initial unrecorded cycles). For samples where the Sanger read successfully
244 mapped, a total of 15 positions out of 1,494 (166 mapped samples and 9 investigated
245 targets) were not assigned a value (1.0% of positions, see suppl. Table 2), mainly from the
246 ends of the sequence. As expected, the “No Template Control” did not yield a result. Five of
247 the 166 sequences had the combination of the N501Y, A570D, P681H, and T716I mutations,
248 which strongly suggests infection with the B.1.1.7 lineage (Table 2, Suppl. Table 2).

249 The enzymes we used in the proposed RT-PCR set up are identical to those used in the
250 COVID-19 diagnostic test facility at DTU, where up to 10,000 patient laryngeal swab
251 samples are processed per day. Because the enzymes are identical, the supply chain is
252 simplified. It is likely that any RT-PCR enzyme mix will perform well with our setup.
253 Identifying emerging VoCs requires large-scale tracking of the complete SARS-CoV-2
254 genome, and we do not recommend substituting the WGS efforts with Sanger sequencing of
255 a single area of the spike gene. However, spike-gene Sanger sequencing of all SARS-CoV-2
256 positive samples can enable rapid mutation typing and association to current variants and
257 mutations of concern, and potentially the proposed amplicon can cover mutations arising in
258 the future. In February 2021, there are currently two additional lineages of note, B.1.525 and
259 A.23.1, that contain key defining mutations within the amplicon sequenced. RT-qPCR with
260 primers and probes specific for VoC mutations is being used to monitor outbreaks, and while
261 this method is arguably faster than Sanger sequencing, there are several drawbacks. For
262 instance, RT-qPCR assays usually use several primers and probes, rendering it more
263 expensive than the proposed Sanger sequencing based method. qPCR assays can only
264 investigate a few mutations in a sample, where Sanger sequencing gives the sequence over
265 a 1 kb part of the spike gene of SARS-CoV-2. In this way, new VoCs are likely to be
266 detected without change to the Sanger sequencing setup, whereas a VoC specific RT-qPCR

267 setup would likely require extensive testing of a new primer probe combination each time a
268 new VoC is classified.



269

270 Figure 2. Sensitivity of Sanger sequencing of the proposed amplicon compared to the
271 diagnostic RT-qPCR test (Ct values, n=195). The numbers above each bar refer to the
272 number of samples in that bin. Overall, 85.1% of samples were of sufficient quality for
273 mapping. The difference between N1 and N2 reflects that the Ct for the two RT-qPCR
274 targets can be different, and sometimes one target is not recorded (n=8 for N1 and n=4 for
275 N2). The NTC is not depicted, as it did not produce a Sanger read that mapped to the
276 reference genome. *: Ct values stem from a 2-step PCR where the first 7 cycles are not
277 fluorescence registered.

278

279 Conclusion

280 We show that typing of current SARS-CoV-2 VoC specific mutations can be performed by
281 Sanger sequencing of an RT-PCR product from the spike gene using a single primer set and
282 existing commercial sequencing infrastructure in a cheap and fast way. This requires no
283 other enzymes or equipment besides those already in daily use at any SARS-CoV-2 RT-
284 qPCR testing facility. Our protocol has the potential to change how SARS-CoV-2 VoC
285 mutations are typed globally, as it is the cheapest and simplest way proposed yet to yield
286 sequencing information from SARS-CoV-2 positive samples. We achieved mutation calls
287 from 85.1% of all samples, some of which only recorded one of two targets in the RT-qPCR
288 diagnostic SARS-CoV-2 test. We do not suggest abandoning the use of WGS, as it is
289 important for tracking emerging variants. The proposed workflow is currently being used
290 routinely to type all positive samples at the High Throughput SARS-CoV-2 RT-qPCR testing
291 facility at DTU.

292

293 Statement of data use

294 All samples are pseudoanonymized before arrival at the DTU facility, and once more before
295 shipment of samples to Eurofins genomics. For this study, Sanger reads were anonymized.

296

297

298 Acknowledgements

299 We thank Jeppe Hallgren and Jørn Emborg from biolib for setting up the self-contained web-
300 app and Kim Ng and Thor Bech Johannesen for coordination assistance from Statens
301 Serum Institut. We thank Benni Winding Hansen for equipment access. We thank the NGS
302 lab at The Novo Nordisk Center for Biosustainability for allowing parts of the work to be
303 performed in their facilities. We thank Søren Karst for valuable discussions.

304 This work was supported by the Poul Due Jensen Fonden (Grundfos Fonden, Corona-
305 Danica). T.S.J. and K.B. received support by The Novo Nordisk Foundation through the
306 grants NNF-IIMENA (NNF16OC0021746) and NNF CFB (NNF20CC0035580). We are

307 gratefully to researchers, clinicians, and public health authorities for making SARS-CoV-2
308 sequence data available through GISAID (supplementary table 3).

309

310 References

311

312 1. Wu, F. *et al.* A new coronavirus associated with human respiratory disease in China.

313 *Nature* **579**, 265–269 (2020).

314 2. Vogels, C. Multiplexed RT-qPCR to screen for SARS-COV-2 B.1.1.7, B.1.351, and P.1

315 variants of concern. (2021) doi:10.17504/protocols.io.brrhm536.

316 3. Vogels, C. B. F. *et al.* Analytical sensitivity and efficiency comparisons of SARS-CoV-2

317 RT-qPCR primer-probe sets. *Nat. Microbiol.* **5**, 1299–1305 (2020).

318 4. Quick, J. nCoV-2019 sequencing protocol v3 (LoCost). (2020).

319 5. Elbe, S. & Buckland-Merrett, G. Data, disease and diplomacy: GISAID's innovative

320 contribution to global health. *Glob. Chall.* **1**, 33–46 (2017).

321 6. Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK

322 defined by a novel set of spike mutations - SARS-CoV-2 coronavirus / nCoV-2019

323 Genomic Epidemiology. *Virological* <https://virological.org/t/preliminary-genomic->

324 characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-

325 spike-mutations/563 (2020).

326 7. Tegally, H. *et al.* Emergence and rapid spread of a new severe acute respiratory

327 syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in

328 South Africa. *medRxiv* 2020.12.21.20248640 (2020) doi:10.1101/2020.12.21.20248640.

329 8. Genomic characterisation of an emergent SARS-CoV-2 lineage in Manaus: preliminary

330 findings - SARS-CoV-2 coronavirus / nCoV-2019 Genomic Epidemiology. *Virological*

331 <https://virological.org/t/genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in->

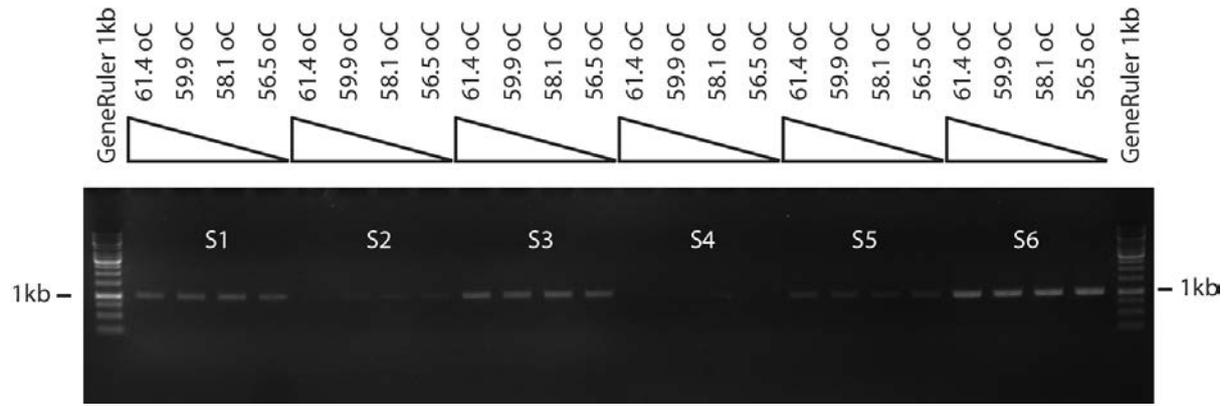
332 manaus-preliminary-findings/586 (2021).

- 333 9. Graham, M. S. *et al.* The effect of SARS-CoV-2 variant B.1.1.7 on symptomatology, re-
334 infection and transmissibility. *medRxiv* 2021.01.28.21250680 (2021)
335 doi:10.1101/2021.01.28.21250680.
- 336 10. Greaney, A. J. *et al.* Comprehensive mapping of mutations to the SARS-CoV-2
337 receptor-binding domain that affect recognition by polyclonal human serum antibodies.
338 *bioRxiv* 2020.12.31.425021 (2021) doi:10.1101/2020.12.31.425021.
- 339 11. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating
340 inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463–5467 (1977).
- 341 12. Rausch, T., Fritz, M. H.-Y., Untergasser, A. & Benes, V. Tracy: basecalling,
342 alignment, assembly and deconvolution of sanger chromatogram trace files. *BMC*
343 *Genomics* **21**, 230 (2020).
- 344 13. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat.*
345 *Methods* **9**, 357–359 (2012).
- 346 14. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinforma. Oxf.*
347 *Engl.* **25**, 2078–2079 (2009).
- 348 15. Hodcroft, E. B. CoVariants: SARS-CoV-2 Mutations and Variants of Interest.
349 <https://covariants.org/> (2021).
- 350 16. Jørgensen, T. S. Sanger sequencing of a part of the SARS-CoV-2 spike protein.
351 (2021) doi:10.17504/protocols.io.bsbdnai6.
- 352 17. Rambaut, A. *et al.* A dynamic nomenclature proposal for SARS-CoV-2 lineages to
353 assist genomic epidemiology. *Nat. Microbiol.* **5**, 1403–1407 (2020).
- 354 18. O'Toole, A. *cov-lineages/pangolin*. (CoV-lineages, 2021).
- 355 19. Hadfield, J. *et al.* Nextstrain: real-time tracking of pathogen evolution. *Bioinforma.*
356 *Oxf. Engl.* **34**, 4121–4123 (2018).

357

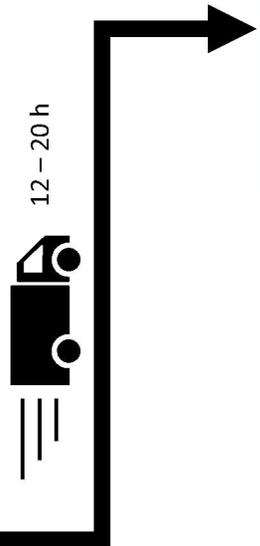
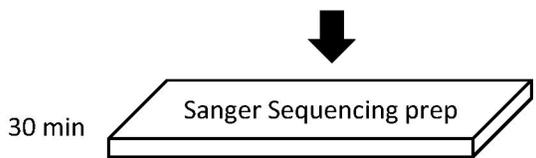
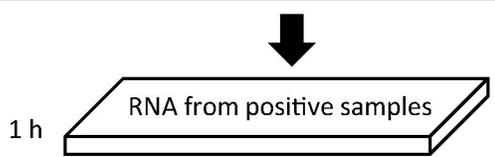
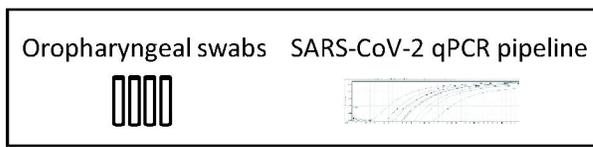
358

359



360

361 Suppl. Fig. 1. 1% Agarose gel electrophoresis of primer test. Each well has 1.5µl RT-PCR
362 product regardless of DNA concentration.



Sanger sequencing

1 h 30min



Data analysis

5 min

Successful calls by Ct* value

