medRxiv preprint doi: https://doi.org/10.1101/2021.03.20.21253956; this version posted March 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

1 **A Sanger-based approach for scaling up screening of SARS-CoV-2 variants of interest and**
2 **concern**

3 Matheus Filgueira Bezerra[1], Lais Ceschini Machado[2], Viviane do Carmo Vasconcelos de
4 Carvalho[3], Cássia Docena[3], Sinval Pinto Brandão-Filho[4], Constância Flávia Junqueira Ayres[2],
5 Marcelo Henrique Santos Paiva[2,5]*, Gabriel Luz Wallau[2]*.

6 [1]Departamento de Microbiologia – Instituto Aggeu Magalhães; Fiocruz. Recife, PE, Brazil;

7 [2]Departamento de Entomologia e Núcleo de Bioinformática – Instituto Aggeu Magalhães;
8 Fiocruz. Recife, PE, Brazil;

9 [3]Núcleo de Plataforma Técnológica – Instituto Aggeu Magalhães; Fiocruz. Recife, PE, Brazil;

10 [4]Departamento de Parasitologia, Instituto Aggeu Magalhães; Fiocruz. Recife, PE, Brazil;

11 [5]Centro Acadêmico do Agreste, Núcleo de Ciências da Vida. Universidade Federal de
12 Pernambuco, Caruaru, PE, Brazil.

13

14 **Running title:** Identification of SARS CoV-2 variants with Sanger sequencing

15

16 * Corresponding authors: Marcelo Paiva & Gabriel Wallau

17 Av. Professor Moraes Rego, s/n, Recife/PE 50740-465, Brazil. Tel: +55-81-21012634.

18 Contact: marcelo.paiva@fiocruz.br / gabriel.wallau@fiocruz.br

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2021.03.20.21253956; this version posted March 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

19 **ABSTRACT**

20 The global spread of new SARS-CoV-2 variants of concern underscore an urgent need of
21 simple deployed molecular tools that can differentiate these lineages. Several tools and protocols
22 have been shared since the beginning of the COVID-19 pandemic, but they need to be timely
23 adapted to cope with SARS-CoV-2 evolution. Although whole-genome sequencing (WGS) of
24 the virus genetic material have been widely used, it still presents practical difficulties such as
25 high cost, shortage of available reagents in the global market, need of a specialized laboratorial
26 infrastructure and well-trained staff. These limitations result in genomic surveillance blackouts
27 across several countries. Here we propose a rapid and accessible protocol based on Sanger
28 sequencing of a single PCR fragment that is able to identify and discriminate all SARS-CoV-2
29 variants of concern (VOCs) identified so far, according to each characteristic mutational profile
30 at the Spike-RBD region (K417N/T, E484K, N501Y, A570D). Twelve COVID-19 samples from
31 Brazilian patients were evaluated for both WGS and Sanger sequencing: three from P.2, two
32 from P.1 and seven from B.1.1 lineage. All results from the Sanger sequencing method perfectly
33 matched the mutational profile of VOCs and non-VOCs described by WGS. In summary, this
34 approach allows a much broader network of laboratories to perform molecular surveillance of
35 SARS-CoV-2 VOCs and report results within a shorter time frame, which is of utmost
36 importance in the context of rapid public health decisions in a fast evolving worldwide
37 pandemic.

38 **Keywords:** SARS-CoV-2 variants of concern; Sanger sequencing; molecular surveillance.

39

40

medRxiv preprint doi: https://doi.org/10.1101/2021.03.20.21253956; this version posted March 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

41 As of December 2020, the United Kingdom reported a new SARS-CoV-2 variant, the
42 B.1.1.7 lineage, which presented a higher transmissibility rate, bringing deep concerns about the
43 prospects of the COVID-19 pandemic (1). Shortly after, other so-called "Variants Of Concern"
44 (VOCs) were reported in South Africa (B.1.3.51), Brazil (P.1) and more recently, in the U.S.A
45 (B.1.526) (2-4). Specific mutations, such as the N501Y and the E484K, in the residue binding
46 domain (RBD) of the Spike protein are recurrent across the VOCs. These mutations play an
47 important role on the lineage phenotype, allowing higher affinity to the human ACE2 receptor
48 and/or immune evasion from previously elicited antibodies (5,6). It is likely that continuous
49 circulation of SARS-CoV-2 in previously exposed and vaccinee populations will drive SARS-
50 CoV-2 evolution towards lineages with increased transmissibility and escape from immune
51 responses, allowing these variants to spread quickly throughout the world (6,7). In this scenario,
52 the development of large-scale molecular surveillance strategies to monitor SARS-CoV-2 VOCs
53 is crucial to provide timely information for proper public health control and adaptation of
54 vaccination measures.

55 Since the release of the first SARS-CoV-2 genome, many molecular tools have been
56 adapted to detect and monitor this virus in parallel with its emerging genomic changes (8). One
57 of the most employed tools, capable of yielding unprecedented results is the whole genome
58 sequencing (WGS) of SARS-CoV-2 from clinical samples. However, WGS is still very
59 expensive to be applied as a front-line method for massive testing, particularly in
60 underdeveloped and developing countries. Additionally, other PCR-based methodologies have
61 been developed as well, focusing mainly on lineage-specific deletions of emerging VOCs and/or
62 Spike mutation differentiation based on amplification dropouts and specific probes in RT-PCR
63 assays (9,10). However, worldwide shortage of imported reagents, limited laboratorial

medRxiv preprint doi: https://doi.org/10.1101/2021.03.20.21253956; this version posted March 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

64 infrastructure and the need of well-trained staff are other limitations commonly faced by these
65 molecular protocols, resulting in surveillance blackouts in many countries. To illustrate the large
66 discrepancies in genomic surveillance data observed during the Covid-19 pandemic, whilst 6.5%
67 (270,762/4.1 x 10^6) of the UK confirmed cases had their genomes sequenced, only 0.03%
68 (3,430/10.5 x 10^6) of the Brazilian confirmed cases were sequenced by early March (11).
69 Therefore, the establishment and standardization of as many molecular protocols as possible that
70 help to scale up the SARS-CoV-2 VOCs screening is highly desirable. Here we propose a rapid
71 and accessible protocol based on Sanger sequencing that is able to identify and discriminate
72 SARS-CoV-2 VOCs, according to each characteristic mutational profile at the Spike-RBD
73 region.

74 In order to access whether the amplicon used in this study is able to cover key SARS-
75 CoV-2 mutations, we accessed Twelve COVID-19 positive samples (RT-PCR - Ct values below
76 25) derived from symptomatic patients of both Pernambuco (Northeast Brazil) and Amazonas
77 (North Brazil) states that had been previously genomic sequenced (8). The study was approved
78 by the local Ethical Committee (CAAE32333120.4.0000.5190). RNA extractions were
79 performed in a BSL-3 facility laboratory with a robotic platform using the Maxwell® 16 Viral
80 Total Nucleic Acid Purification Kit (Promega, Wisconsin-USA), following the manufacturer's
81 instructions. The molecular diagnosis of SARS-CoV-2 was performed using the Kit Molecular
82 BioManguinhos SARS-CoV-2 (E/RP).

83 High Capacity cDNA Reverse-Transcription kit (Applied Biosystems) was used for
84 reverse transcription, following the manufacturer's instruction. Next, cDNA was subjected to
85 PCR with Platinum Taq-polymerase (Invitrogen) and primers flanking the regions between the
86 nucleotide positions 22797 and 23522 of the Wuhan (Wu-1) reference genome, covering key

medRxiv preprint doi: https://doi.org/10.1101/2021.03.20.21253956; this version posted March 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

87 amino acid replacements commonly found in VOCs RBD domain of the Spike protein (76 Left:
88 5'-AGGGCAAACTGGAAAGATTGCT-3' and 77 Right: 5'-
89 CAGCCCCTATTAAACAGCCTGC-3' designed by https://www.protocols.io/view/ncov-2019-
90 sequencing-protocol-bbmuik6w). PCR conditions were: 98 °C for 5 minutes s, 98°C for 30
91 seconds, 59°C for 30 seconds and 72°C for 45 seconds during 35 cycles and final extension of 5
92 min at 72°C. Primer and magnesium chloride concentrations in the PCR were 0.2 μM and 1 mM,
93 respectively. Amplified PCR products were verified in a 1.5% Agarose gel stained with Sybr
94 Safe (Sigma-Aldrich), quantified in a NanoDrop OneC Microvolume UV-Vis Spectrophotometer
95 (Thermo-Fischer, USA) and diluted to 30 ng/uL. Sequencing reactions were performed with
96 BigDye Terminator v3.1 (Applied Biosystems) and ran in capillary electrophoresis (ABI 3500,
97 Applied Biosystems). Contigs from forward and reverse strands were built and analyzed using
98 the CodonCode aligner v3.7.1 software and figures were built using the Biorender platform.
99 Samples were assigned to a lineage according to the mutational profile (Table 1).

100 According to the WGS, from the twelve COVID-19 samples evaluated, seven were from
101 the B.1.1 lineage (non-VOC), three were P.2 and two were P.1. Remarkably, in a blind
102 comparison to WGS (gold standard), all results from the Sanger sequencing method matched
103 those from WGS method. The K417, E484 and N501Y mutations were identified in the P.1 cases
104 and the E484K (in absence of the others) in the P.2 cases (Table 1).

105 Within the sequencing of a single 725 base pairs PCR fragment (Figure 1), this approach
106 could successfully detect VOC-associated mutations and correctly classify samples according to
107 the WGS data. Moreover, the flanked region also covers other relevant circulating RBD
108 mutations (Figure 1) and potentially, new mutations that have not been identified yet. Together,
109 these features overcome some of the limitations of allelic-specific PCR methods, such as the

medRxiv preprint doi: https://doi.org/10.1101/2021.03.20.21253956; this version posted March 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

110 need of one specific probe or primer for each mutation to be evaluated and previous knowledge
111 of the circulating mutations (10). Furthermore, high-quality electropherograms were obtained
112 without a PCR purification step, reducing costs and time of sample processing, which is
113 particularly useful for large-scale application of the method. Another advantage of this approach
114 is that primers can be easily adjusted without major protocol modifications, in case newly
115 described mutations need to be detected. On the other hand, it is important to highlight that
116 Sanger sequencing is normally more time consuming than allelic-specific RT-PCR and hence
117 with a comparative reduced scaling capacity, but it brings some advantages such as more genetic
118 data that helps to tease apart different VOCs and the possibility of detecting new emerging RBD
119 mutations.

120 It is important to highlight that this approach does not substitute WGS and other PCR-
121 based assays and could be used in combination to further validate the VOCs results mainly with
122 WGS to uncover other important mutation at the SAR-CoV-2 genome, but it will allows a much
123 broader network of laboratories to perform molecular surveillance of SARS-CoV-2 VOCs,
124 reporting results within a shorter time frame and in larger amounts, which is of utmost
125 importance in the context of rapid public health decisions in a fast evolving worldwide
126 pandemic.

127

128 ACKNOWLEDGMENTS

129 We would like to thank the COVID-IAM and LACEN-PE teams for providing the
130 samples to sequence the SARS-CoV-2 genomes, the Technological Platform Core and the

medRxiv preprint doi: https://doi.org/10.1101/2021.03.20.21253956; this version posted March 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

131 Bioinformatic Core of the Aggeu Magalhaes Institute for the support with their research
132 facilities.

133

134 **DATA AVAIABILITY**

135 All genomes generated in this study are deposited on GISAID under the accessions:
136 EPI_ISL_500460, EPI_ISL_500461, EPI_ISL_500865, EPI_ISL_500868, EPI_ISL_500872,
137 EPI_ISL_500477, EPI_ISL_500482, EPI_ISL_1239012, EPI_ISL_1239013, EPI_ISL_1239014,
138 EPI_ISL_1239015, EPI_ISL_1239016.

139

140 **FUNDING**

141 Gabriel Luz Wallau was supported by the National Council for Scientific and
142 Technological Development by the productivity research fellowship level 2 (303902/2019-1).

143

144 **DISCLOSURE OF CONFLICTS OF INTEREST**

145 The authors have no competing financial interests to declare.

146

147 **AUTHOR CONTRIBUTIONS**

medRxiv preprint doi: https://doi.org/10.1101/2021.03.20.21253956; this version posted March 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

148 M.F.B conceived the study, performed experiments, collected/analyzed data and drafted the
149 manuscript. L.C.M, V.C.V.C and C.D performed experiments. S.P.B.F and C.F.J.A obtained
150 patient samples, updated the clinical data and corrected the manuscript. M.H.S.P and G.L.W
151 conceived and designed the study, analyzed data and gave the final approval of the version to be
152 submitted.

153

154 **REFERENCES**

155 1. Leung K, Shum MH, Leung GM, Lam TT, Wu JT. Early transmissibility assessment of
156 the N501Y mutant strains of SARS-CoV-2 in the United Kingdom, October to November
157 2020. Euro Surveill. 2021; 26(1): 2002106.
158 2. Makoni M. South Africa responds to the new SARS-CoV-2 variant. The Lancet. 2021;
159 397(10271): 261.
160 3. Faria NR, Mellan TA, Whittaker C, Claro IM, Candido DDS, Mishra S, Crispim MAE.
161 Genomics and epidemiology of a novel SARS-CoV-2 lineage in Manaus, Brazil.
162 medRxiv [Preprint]. 2021; 3:2021.02.26.21252554.
163 4. Annavajhala MK, Mohri H, Zucker JE, et al. A Novel SARS-CoV-2 Variant of Concern,
164 B.1.526, Identified in New York. Preprint. medRxiv [Preprint]. 2021;
165 2021.02.23.21252259.
166 5. Wang P, Wang M, Yu J, et al. Increased Resistance of SARS-CoV-2 Variant P.1 to
167 Antibody Neutralization. BioRxiv [Preprint]. 2021; 2021;2021.03.01.433466.
168 6. Nelson G, Buzko O, Spilman P, Niazi K, Rabizadeh S, Soon-Shiong P. Molecular
169 dynamic simulation reveals E484K mutation enhances spike RBD-ACE2 affinity and the

medRxiv preprint doi: https://doi.org/10.1101/2021.03.20.21253956; this version posted March 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

170 combination of E484K, K417N and N501Y mutations (501Y.V2 variant) induces
171 conformational change greater than N501Y mutant alone, potentially resulting in an
172 escape mutant. G bioRxiv [Preprint]. 2021; 2021.01.13.426558.
173 7. Fontanet A, Autran B, Lina B, Kieny M, Karim S, Sridhar D. SARS-CoV-2 variants and
174 ending the COVID-19 pandemic.. The Lancet. 2021; 397(10278): 952-954.
175 8. Paiva MHS, Guedes DRD, Docena C, Bezerra MF, Dezordi FZ, Machado LC,
176 Krokovsky L, Helvecio E. Multiple Introductions Followed by Ongoing Community
177 Spread of SARS-CoV-2 at One of the Largest Metropolitan Areas of Northeast Brazil.
178 Viruses. 2020; 12(12):1414.
179 9. Chantal B.F. Vogels, Mallery Breban, Tara Alpert1 , Mary E. Petrone, Anne E. Watkins,
180 Emma B. Hodcroft, Christopher E. Mason. PCR assay to enhance global surveillance for
181 SARS-CoV-2 variants of concern. MedRxiv [Preprint]. 2021;
182 https://doi.org/10.1101/2021.01.28.21250486.
183 10. Naveca F, Nascimento V, Souza V., et al. COVID-19 epidemic in the Brazilian state of
184 Amazonas was driven by long-term persistence of endemic SARS-CoV-2 lineages and
185 the recent emergence of the new Variant of Concern P.1. Research Square [Preprint].
186 2021; https://doi.org/10.21203/rs.3.rs-275494/v1.
187 11. GISAID initiative, accessed on 2021 March 02; Available from: http://www.gisaid.org/

188

189 **FIGURE LEGENDS**

190 **Figure 1. Identification of Sars CoV-2 Spike-RBD mutations using Sanger sequencing.**
191 Commonly found RBD mutations flanked by the primer set (nucleotide positions from 22797 to

medRxiv preprint doi: https://doi.org/10.1101/2021.03.20.21253956; this version posted March 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

192 23522 at the Wu-1 genome) used for sequencing, including key mutations to enable identifying
193 variants of concern and interest **(A)**. 725 bp PCR fragments amplified from Sars Cov-2 cDNA
194 **(B)**. Sections from the eletropherograms obtained by Sanger sequencing showing the E484K and
195 N501Y VOC-associated mutations **(C)**.

196

Table 1. Sars cov-2 lineages according to the mutational profile in Sanger sequencing.

		Mutation						
Lineage	**First report**	**K417N**	**K417T**	**L452R**	**S477N**	**E484K**	**N501Y**	**A570**
P.1	Brazil (Amazon)	-	present	-	-	present	present	-
P.2	Brazil	-	-	-	-	present	-	-
B1.1.7	U.K	-	-	-	-	-	present	present
B.1.3.51	South Africa	present	-	-	-	present	present	-
CAL.20C	U.S.A (California)	-	-	present	-	-	-	-
B.1.526	U.S.A (New York)	-	-	-	present	present	-	-

medRxiv preprint doi: https://doi.org/10.1101/2021.03.20.21253956; this version posted March 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

A

S1

S2

333

527

NTD

RBD

FP

HR1

HR2

TA

IT

-K417T/N

-N439K

-L452R

-Y453F

-S477N

-T478K

-E484K

K I A D Y N Y K L P D D F T G C V I A W N S N N L D S K V G G N Y N Y L Y R L F R K S N L K P F E R D I S T E I Y Q A G S T P C N G V E G F N C Y F P L Q S Y

-N501Y/T/S

-A570D

G F Q P T N G V G Y Q P Y R V V V L S F E L L H A P A T V C G P K K S T N L V K N K C V N F N F N G L T G T G V L T E S N K K F L P F Q Q F G R D I A D T T D (...)

B