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

Matheus Filgueira Bezerra¹, Lais Ceschini Machado², Viviane do Carmo Vasconcelos de Carvalho³, Cássia Docena³, Sinval Pinto Brandão-Filho⁴, Constância Flávia Junqueira Ayres², Marcelo Henrique Santos Paiva^{2,5*}, Gabriel Luz Wallau^{2*}.

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

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

³Núcleo de Plataforma Tecnológica – Instituto Aggeu Magalhães; Fiocruz. Recife, PE, Brazil;

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

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

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

* Corresponding authors: Marcelo Paiva & Gabriel Wallau

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

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

ABSTRACT

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

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

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

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

infrastructure and the need of well-trained staff are other limitations commonly faced by these molecular protocols, resulting in surveillance blackouts in many countries. To illustrate the large discrepancies in genomic surveillance data observed during the Covid-19 pandemic, whilst 6.5% (270,762/4.1 x 10⁶) of the UK confirmed cases had their genomes sequenced, only 0.03% (3,430/10.5 x 10⁶) of the Brazilian confirmed cases were sequenced by early March (11). Therefore, the establishment and standardization of as many molecular protocols as possible that help to scale up the SARS-CoV-2 VOCs screening is highly desirable. Here we propose a rapid and accessible protocol based on Sanger sequencing that is able to identify and discriminate SARS-CoV-2 VOCs, according to each characteristic mutational profile at the Spike-RBD region.

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

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

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

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

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

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

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

ACKNOWLEDGMENTS

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

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

133

134 **DATA AVAILABILITY**

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**

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

REFERENCES

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

combination of E484K, K417N and N501Y mutations (501Y.V2 variant) induces conformational change greater than N501Y mutant alone, potentially resulting in an escape mutant. G bioRxiv [Preprint]. 2021; 2021.01.13.426558.

7. Fontanet A, Autran B, Lina B, Kieny M, Karim S, Sridhar D. SARS-CoV-2 variants and ending the COVID-19 pandemic.. The Lancet. 2021; 397(10278): 952-954.

8. Paiva MHS, Guedes DRD, Docena C, Bezerra MF, Dezordi FZ, Machado LC, Krokovsky L, Helvecio E. Multiple Introductions Followed by Ongoing Community Spread of SARS-CoV-2 at One of the Largest Metropolitan Areas of Northeast Brazil. Viruses. 2020; 12(12):1414.

9. Chantal B.F. Vogels, Mallery Breban, Tara Alpert1 , Mary E. Petrone, Anne E. Watkins, Emma B. Hodcroft, Christopher E. Mason. PCR assay to enhance global surveillance for SARS-CoV-2 variants of concern. MedRxiv [Preprint]. 2021; <https://doi.org/10.1101/2021.01.28.21250486>.

10. Naveca F, Nascimento V, Souza V., et al. COVID-19 epidemic in the Brazilian state of Amazonas was driven by long-term persistence of endemic SARS-CoV-2 lineages and the recent emergence of the new Variant of Concern P.1. Research Square [Preprint]. 2021; <https://doi.org/10.21203/rs.3.rs-275494/v1>.

11. GISAID initiative, accessed on 2021 March 02; Available from: <http://www.gisaid.org/>

FIGURE LEGENDS

Figure 1. Identification of Sars CoV-2 Spike-RBD mutations using Sanger sequencing.

Commonly found RBD mutations flanked by the primer set (nucleotide positions from 22797 to

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.

Lineage	First report	Mutation						
		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	-	-