



Check for updates

BRIEF REPORT

A Sanger sequencing-based method for a rapid and economic generation of SARS-CoV-2 epidemiological data: A proof of concept study to assess the prevalence of the A23403G SNP (D614G) mutation in Quito, Ecuador.

[version 1; peer review: 1 approved with reservations, 1 not approved]

Paula Leoro-Garzón¹, Galo Leoro-Monroy¹, David Ortega-Paredes^{2,3},
 Marco Larrea-Álvarez⁴, Gabriel Molina-Cuasapaz^{id}⁵, Pablo González-Andrade⁶,
 Gabriela Sevillano⁶, Camilo Zurita-Salinas⁶, Jeannete Zurita^{id}^{6,7}

¹Departamento de Biología Molecular., Laboratorio Clínico e Inmunológico Inmunolab, Quito, Pichincha, 170403, Ecuador²Unidad de Investigación en Enfermedades Transmitidas por Alimentos y Resistencia a los Antimicrobianos (UNIETAR), Facultad de Veterinaria, Universidad Central del Ecuador, Quito, Pichincha, 170129, Ecuador³Facultad de Ciencias Médicas Enrique Ortega Moreira, Carrera de Medicina, Universidad Espíritu Santo, Samborondón, Guayas, 0901952, Ecuador⁴School of Biological Science and Engineering, Yachay-Tech University, Urcuquí, Imbabura, 100650, Ecuador⁵Facultad de Ciencias Agropecuarias y Recursos Naturales, Carrera de Medicina Veterinaria, Universidad Técnica de Cotopaxi, Latacunga, Cotopaxi, 050101, Ecuador⁶Unidad de Investigaciones en Biomedicina, Zurita & Zurita Laboratorios, Quito, Pichincha, 170104, Ecuador⁷Facultad de Medicina, Pontificia Universidad Católica del Ecuador, Quito, Pichincha, 170143, Ecuador

V1 First published: 31 Mar 2022, 11:383
<https://doi.org/10.12688/f1000research.108799.1>

Latest published: 31 Mar 2022, 11:383
<https://doi.org/10.12688/f1000research.108799.1>

Abstract

The Single Nucleotide Polymorphism (SNP) A23403G associated with the D to G change in position 614 of the SARS-CoV-2 spike protein has recently become dominant. The most utilized and robust approach is the study of whole genome sequences, generally available at public databases. However, this technology is not suited for massive testing as it requires expensive reagents, equipment, and infrastructure. Consequently, developing rapid and accessible protocols will be fundamental for producing epidemiological data linked to this SNP, especially in countries with limited resources. This report has evaluated an easy cost-effective approach, based on sanger sequencing, for detection of the A23403G (D614G) mutation. This strategy was tested in SARS-CoV-2 positive samples collected in Quito during March and October of 2020. In March, a total of 264 out of 1319 samples yielded positive results (20%), while 777 out of 5032 (15%) did so in October. From these cases, almost all samples were associated with the G23403 (G614) variant (>98%). This technique proved to be reliable, reproducible, and might be expandable to study

Open Peer Review

Approval Status ? X

	1	2
version 1	?	X
31 Mar 2022	view	view

1. **Fernanda de-Paris** ^{id}, Hospital de Clínicas de Porto Alegre, Porto Alegre, Brazil

Grazielle Motta Rodrigues, Hospital de Clínicas de Porto Alegre, Porto Alegre, Brazil

2. **Anissa Chouikha**, Institut Pasteur de Tunis, Université Tunis El Manar, Tunis, Tunisia

Any reports and responses or comments on the article can be found at the end of the article.

other mutations without major protocol amendments. The application of this method allowed the production of epidemiological data regarding the A23403G (D614G) mutation in Quito, where no previous reports were available. This approach will be crucial for producing relevant information for public health management, especially during the ongoing pandemic.

Keywords

SARS-CoV-2 infection, D614G, epidemiology, SNP, Sanger sequencing, low-cost method, Quito-Ecuador

Corresponding author: Jeannete Zurita (jjurita@zuritalaboratorios.com)

Author roles: **Leoro-Garzón P:** Funding Acquisition, Methodology, Writing – Original Draft Preparation; **Leoro-Monroy G:** Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Ortega-Paredes D:** Conceptualization, Methodology, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing; **Larrea-Álvarez M:** Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Molina-Cuasapaz G:** Methodology, Writing – Original Draft Preparation, Writing – Review & Editing; **González-Andrade P:** Methodology; **Sevillano G:** Methodology, Writing – Original Draft Preparation; **Zurita-Salinas C:** Funding Acquisition, Methodology, Writing – Review & Editing; **Zurita J:** Funding Acquisition, Methodology, Supervision, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This study was funded by Laboratorio Clínico Inmunolab (research common fund) and Zurita & Zurita Laboratorios (grant number MIC-019) and VetNAAT Laboratory. Funding sources had no influence over study design, data collection, analysis, interpretation or manuscript preparation and submission. This research did not receive any specific grant from funding agencies in the public or not-for-profit sectors.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2022 Leoro-Garzón P *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Leoro-Garzón P, Leoro-Monroy G, Ortega-Paredes D *et al.* **A Sanger sequencing-based method for a rapid and economic generation of SARS-CoV-2 epidemiological data: A proof of concept study to assess the prevalence of the A23403G SNP (D614G) mutation in Quito, Ecuador. [version 1; peer review: 1 approved with reservations, 1 not approved]** F1000Research 2022, 11:383 <https://doi.org/10.12688/f1000research.108799.1>

First published: 31 Mar 2022, 11:383 <https://doi.org/10.12688/f1000research.108799.1>

Introduction

SARS-CoV-2 is the agent responsible for the COVID-19 pandemic that has caused almost 4 million deaths worldwide (World Health Organization (WHO) 2021). The genome of this virus is approximately 30-kb long (Bar-On *et al.* 2020). Two Open Reading Frames (ORFs) (1a and 1b), that cover most of its genome, encode non-structural proteins needed for viral replication as well as an RNA-dependent RNA polymerase (RdRp). Structural proteins, such as the Spike (S) protein, nucleocapsid (N) protein, membrane (M) protein, and envelope (E) protein are encoded by short sub-genomic RNAs (sgRNAs) (Kim *et al.* 2020). The S protein recognizes and binds to the angiotensin-converting enzyme 2 (ACE2) receptor to gain entry to hosts cells.

Evidently, mutations in the gene encoding the S protein alter the affinity to the ACE2 receptor enhances infectivity (Watanabe *et al.* 2020), and increased infection capacity. The aspartate (D) to glycine (G) substitution related to the non-synonymous A23403G SNP (single nucleotide polymorphism) has been associated with the rapid spread of the virus (Yuan *et al.* 2020).

The A23403G (D614G) mutation was first described in early 2020, and rapidly detected in various regions around the globe (Isabel *et al.* 2020; Raghav *et al.* 2020; Dao *et al.* 2021; Elizondo *et al.* 2021; Molina-Mora *et al.* 2021; Pandey *et al.* 2021). Studies related to this mutation have used, in the majority, whole genome sequences (WGS). However, this approach is too expensive to be used in massive testing campaigns, especially in countries with limited resources. Moreover, the number of cases studied by WGS may not necessarily be representative of the number of cases detected by quantitative Polymerase Chain Reaction (qPCR) at these locations.

As a result, several methodologies have been developed as affordable alternatives that provide rapid detection of not only relevant SNPs in the S protein, but also SNPs in variants of concern (Hashemi *et al.* 2020; Bezerra *et al.* 2021; Chakraborty *et al.* 2021; Vogels *et al.* 2021). Low-cost rapid technologies have been put forward to investigate the A23403G (D614G) mutation. These technologies rely on: Probe-based real-time reverse transcriptase PCR (rRT-PCR) (Chan *et al.* 2022), an amplification refractory mutation system (ARMS) (Islam *et al.* 2021), and restriction fragment length polymorphism (RFLP) (Niranji & Al-Jaf 2021). Arguably, it is urgent to establish and standardize as many molecular tools as possible in order to expand the screening capacity of SARS-CoV-2 mutations. Consequently, and as an effort to scale up molecular diagnosis, we have developed and tested a cost-effective method based on Sanger sequencing (Jørgensen *et al.* 2021; Park *et al.* 2021) to detect the A23403G (D614G) mutation.

This investigation was carried out in Quito, the capital city of Ecuador, during March and October of 2020. The first SARS-CoV-2 case detected in the country was registered in March 2020, and from that point the virus has been reported in all 24 provinces. Until the 7th of July 2021, the number of positive cases in Ecuador was 467,878 (Johns Hopkins University & Medicine 2021). However, to this date, there are no studies reporting the prevalence of this mutation in the city or in the country, despite its predominance in various regions (Loureiro *et al.* 2021; Pandey *et al.* 2021; Viedma *et al.* 2021).

Methods

In the present study, we evaluated 1319 samples collected during March and 5032 gathered during October of 2020 (all nasopharyngeal swabs) by Zurita & Zurita Laboratorios and Laboratorio Clínico Inmunolab from both symptomatic and asymptomatic patients (see Underlying data) (Sevillano 2022). The two laboratories carried out detection independently. These clinical laboratories do not perform quantification in routine samples; results were reported as positive or negative. Samples from the first time point (March 2020) were extracted using the High Pure Viral Nucleic Acid kit (Product No. 11858874001, Roche, Germany), and analyzed using the LightMix[®] Modular SARS and Wuhan CoV E-gene kits (Product No. 09155368001, Roche, Germany). Both genes were analyzed using a LightCycler[®] 480 Real-Time PCR. Samples collected in October 2020 were extracted using CommaPrep[®] RNA Extraction Columns (Product No. RP20, Biocomma, China), following the provided guidelines, and analyzed using STANDARD M nCoV Real-Time Detection kit (Product No. M-NCOV-01, SB Biosensor, Korea) on a Stratagene MX3005P Real-Time PCR System or Allplex[™] SARS-CoV-2 Assay (Product No. RV10248X, Seegene Inc, Korea) on CFX96 Touch Real-Time PCR Detection System. Interpretation of results (positive/negative) were carried out based on the guidelines provided by the manufacturers. Both kits are currently approved by the ARCSA (National Agency for Health Regulation, Control and Surveillance of Ecuador) for their use in routine diagnostics.

To estimate the prevalence of the A23403G (D614G) mutation, the sample size was first calculated based on the prevalence of SARS-CoV-2 positive cases, 0.2 for March and October (see Underlying data) (Sevillano 2022). The following equation, with a 10% margin of error and a confidence level of 95%, was used for calculating such value:

$$n = \frac{z^2(P)(1-P)}{d^2} \quad (1)$$

According to [equation \(1\)](#), where z is the statistic corresponding to the level of confidence, P is the expected prevalence, and d is the precision ([Pourhoseingholi et al. 2013](#)). The estimated number was 61, although we analyzed a total of 96 samples per month. Selection of samples was carried out using a random number generator.

Initially, the sequence encoding for the SARS-CoV-2 S protein (Gene Bank accession number: MT252819.1) was utilized for designing primers using the Geneious Prime software Version 2019.2.3 (<https://www.geneious.com/prime/>). Primers flanked the region between nucleotide positions 23,301 and 23,511 of the WH-Human 1 coronavirus (MN908947) reference genome. Primers were aligned to genomes reported in Latin American countries in order to identify potential catastrophic mutations within the annealing sites of the primers. Forward sequence (D614G_F2): 5'-GATGCTGTCCGTGATCCACA-3'; reverse sequence (D614G_R2): 5'-AAACAGCCTGCACGTGTTTG-3'. The resulting amplicon was expected to be 230bp long. In the A23403 (D614) variant, adenine is the primary nucleotide that encodes the amino acid aspartate at position 614 of the polypeptide. Glycine has been observed to replace this aspartate if guanine is present instead of adenine, this variant is known as G23403 (G614).

At the next step, PCR analyses were performed on the positive samples with the following conditions: 45 °C for 15 min, 95 °C for 2 min with 45 cycles of 95 °C for 10 sec and 60 °C for 50 sec, followed by the reading of dissociation curves at temperatures from 55 °C to 95 °C. The reaction used 10 µL GoTaq G2 Hot Start Colorless Master Mix (Product No. M7422, Promega, USA), 50 U of RocketScript RTase H Minus (Product No. BQ-042-101-04, Bioneer, Korea), 1 µL of EvaGreen 20x (Product No. 31000-31000-T, Biotium, USA), 0.75 µL of primers D614G_F2 and D614G_R2 10 mM, 10ng RNase Inhibitor (Product No. RB0478, BioBasic, Canada) and template up to a final volume of 15 µL. Samples were analyzed on a Stratagene MX3005P Real-Time PCR system, using EvaGreen dye, and melting curves (dissociation curves) were analyzed using the MxPro application. This analysis permits the measurement of the temperature at which DNA strands separate into single strands, which provides a reference of the melting temperature at which 50% of DNA is denatured.

Finally, all samples that produced an amplicon with a melting temperature around 81 °C were sent to be sequenced commercially using Sanger technology (Macrogen, Korea). The resulting sequences were aligned to the reference WH-Human 1 coronavirus (MN908947) genome using the Geneious Prime software Version 2019.2.3 (<https://www.geneious.com/prime/>).

Results

At the first sampling point, 20% of the samples (264 out of 1319) tested positive for SARS-CoV-2. Likewise, at a later stage, 15% of the samples (777 out of 5032) were infected by the virus ([Table 1](#)).

Per month, 96 samples were evaluated and, after sequencing, were assigned to their corresponding type ([Table 2](#)). From these samples 99% possessed a guanine at position 23403 (G614), while 1% contained an adenine at this position (D614). In March and October, the G23403 (G614) variant represented more than 98% of the samples (see [Underlying data](#)) ([Sevillano 2022](#)). However, the A23403 (D614) variant were registered only in the samples collected in March.

The sequence of 230 bp ([Figure 1A](#)) was sufficient to effectively identify the A23403G (D614G) mutation ([Figure 1B](#)). Moreover, high resolution melting curve analysis (HRM) was helpful to confirm the amplicon identity before Sanger sequencing, as the T_m value of the expected amplicon ranged from 78 °C to 82 °C. However, this approach was not suitable for genotype differentiation since no marked differences were observed between the wildtype version and the A23403G (D614G) mutation ([Figure 1C](#)).

Table 1. Percentages of positive cases in March and October 2020 among samples collected in Quito, Ecuador.

	Periods of study			
	March		October	
	Counts	%	Counts	%
Positive	264	20	777	15
Negative	1055	80	4255	85
Total	1319	100	5032	100

Table 2. Prevalence of the A23403G (D614G) mutation in the S protein at two different time points.

Variants	Periods of study			
	March 2020		October 2020	
	Counts	%	Counts	%
A23403 (D614)	2	2	0	0
G23403 (G614)	93	97.89	96	100
Total	95	100	96	100

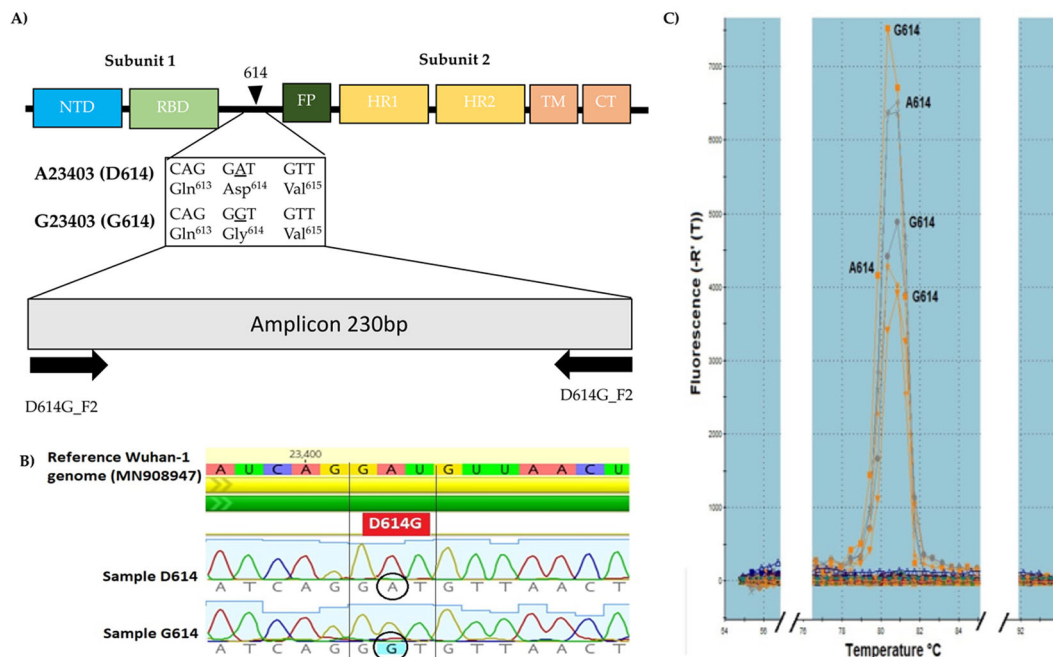


Figure 1. Sanger sequencing for detection of the SARS CoV-2 A23403G (D614G) mutation. A) Adenine to guanine modification and its consequences for the S protein. Primers flanking the area of interest between nucleotide positions 23,301 and 23,511 of the WH-Human 1 coronavirus genome (MN908947), used as reference in the sequencing analysis. B) Segments of the electropherograms by Sanger sequencing indicating the A23403G mutation. C) Melting temperature analysis of selected samples for the missense mutation. All samples had a similar T_m of around 81 °C indicating the correct amplicon size, although not sufficient to discriminate between the G and D types.

Discussion

Since its appearance in late 2019, SARS-CoV-2 has overwhelmed the entire global population with its geographical distribution and mutational rate. The virus has been constantly evolving into new variants with high genetic diversity. To study the G23403 (G614) variant, researchers commonly analyze whole-genome sequences, available in public databases e.g., GISAID (<https://www.gisaid.org/>), as well as sequences of their own.

WGS is the most reliable and utilized method, notwithstanding its cost and inadequacy to work with a large number of samples. WGS is high-priced to be employed as a method for massive testing, predominantly in low-income countries. Consequently, several techniques have been put forward as alternatives for an economic and rapid detection of not only spike mutations, but also variants of concern (Rhee *et al.* 2019; Hashemi *et al.* 2020; Bezerra *et al.* 2021; Chakraborty *et al.* 2021; Islam *et al.* 2021; Park *et al.* 2021; Vogels *et al.* 2021; Chan *et al.* 2022).

In this investigation, the efficacy of a low-cost method, based on sanger sequencing, to detect the A23403G (D614G) mutation was assessed. This approach proved to be affordable, user-friendly and consistent, which simplifies its implementation in common laboratory facilities worldwide; especially in regions with limited health systems such as Sub-Saharan Africa and Latin America (Kirby 2020; Okoi & Bwawa 2020).

The primers employed to produce the amplicon were straightforwardly designed and can be easily adjusted to detect novel variants without major protocol amendments. Arguably, other mutations of relevance could be studied in the same way. For instance, the evaluated technique might prove useful for detecting the N439K mutation, associated with an increased affinity to ACE2 and reduced sensitivity to SARS-CoV-2 antibodies (Thomson *et al.* 2021).

Other approaches have been proposed to study the A23403G (D614G) mutation. For example, one study successfully tested a combined point-of-care nucleic acid and antibody assay (Milcochova *et al.* 2020). Similarly, another method effectively detected the mutation, although based on an engineered Cas12a guide RNA (Meng *et al.* 2021). Despite encouraging results, these techniques appear particularly laborious and expensive for laboratories with limited resources. Rapid and economic approaches have also been developed, which are based on (i) probe-based real-time reverse transcriptase PCR (qRT-PCR) (Chan *et al.* 2022), (ii) an amplification refractory mutation system (ARMS) RT (Islam *et al.* 2021) and (iii) restriction fragment length polymorphism (RFLP) (Niranji & Al-Jaf 2021), the latter method has proved particularly intricate to optimize as original results (Hashemi *et al.* 2020) had to be revised (Niranji & Al-Jaf 2021).

In the present research, a simple straightforward technique consisting of a simple PCR reaction and further sequencing was tested, thus reducing the steps required for sample processing. This study aimed to use melting curve analysis to identify the screened mutation. This approach has been utilized successfully to detect different viruses (Liu *et al.* 2018) and recently applied to detect SARS-CoV-2 mutations (Barua *et al.* 2021; Sarkar *et al.* 2021).

Unfortunately, the melting curves obtained during amplification were not suitable to discriminate between variants. However, they did provide key information about the size of the amplicon, which is fundamental for sequencing the right samples. Clearly, this step must be optimized. Another limitation is that Sanger sequencing could be regarded as time consuming, although it brings several benefits as it is useful to detect developing mutations and even to differentiate variants of concern (Bezerra *et al.* 2021).

Data from the Emergency Operations Center of Pichincha (COE-Pichincha) showed that around 579 positive cases were registered in Quito by March 2020. In October 2020, the official number reported was 15,890 (CEO Provincial Pichincha 2021). However, no reports have been published with regards to the prevalence of the A23403G (D614G) mutation in Quito. In this investigation determined that 190 out of 192 samples (99%) belong to the G23403 (G614) variant, which was dominant at both time points. Other studies have similarly reported a predominance of this SNP in different regions (Loureiro *et al.* 2021; Pandey *et al.* 2021; Viedma *et al.* 2021); these outcomes show the rapid spread of this variant which might relate to a higher transmission rate, as suggested by certain authors (Arora *et al.* 2021; Ozono *et al.* 2021).

As aforementioned, in Ecuador, there are no studies reporting the prevalence of this missense mutation. Hence, to estimate its prevalence in the months studied, this research referred to GISAID (<https://www.gisaid.org/>) to retrieve WGS and metadata of SARS-CoV-2 from the period of study. Our *in-silico* study revealed that from the total of sequences registered in March, five belonged to the G23403 (G614) variant and three to A23403 (D614) variant. Likewise, in October, G23403 (G614) was predominant with four sequences out of five being this type (Table 3). Evidently, a major limitation of studying datasets is that they represent a small fraction of the cases detected by RT-PCR, especially in the periods of study.

The results of this study corroborate what can be inferred from WGS analysis, the G23403 (G614) proved to be the most common in samples from Quito during March and October of 2020. Arguably therefore, using this affordable method permits us to expand the information drawn from databases.

Table 3. Prevalence of the A23403G (D614G) mutation using WGS obtained from GISAID in March and October of 2020.

Variant	Sampling period			
	March 2020		October 2020	
	Counts	%	Counts	%
A23403 (D614)	3	38	1	20
G23403 (G614)	5	62	4	80
Total	8	100	5	100

This scheme was not intended as a replacement of WGS or other PCR-based approaches. Instead, it was devised as a supplementary technique that could be used, in combination with the aforementioned methods, to rapidly generate ample data on the prevalence and epidemiology of relevant SNPs around the globe, mainly in regions with limited facilities and equipment for detection.

The importance of the present approach lies in the speed and easiness with which results can be obtained. This simple technique might help a large group of laboratories to report results quicker and more abundantly, which appears fundamental in the context of public health management during the ongoing pandemic.

The present outcomes demonstrate that the evaluated methodology produced reliable and reproducible information with regards to the A23403G (D614G) mutation. Using this approach, we showed that the G23403 (G614) variant was the most common, during the months of March and October of 2020, among samples from Quito, the capital city of Ecuador. This information contributes to expanding the results obtained by WGS, which will undoubtedly benefit public health decisions during novel outbreaks. Future work must place emphasis on testing a larger number of samples to accurately estimate the prevalence of this mutation in the area, and to assess its relationship with the transmissibility and severity of the disease.

Data availability statement

Underlying data

Repository: A Sanger sequencing-based method for a rapid and economic generation of SARS-CoV-2 epidemiological data: A proof of concept study to assess the prevalence of the A23403G SNP (D614G) mutation in Quito, Ecuador.

DOI: [10.17605/OSF.IO/PMVGJ](https://doi.org/10.17605/OSF.IO/PMVGJ) (Sevillano 2022).

This project contains the following underlying data:

Data file 1. March 2020 Positive SARS-CoV-2 Tests

Data file 2. October 2020 Positive SARS-CoV-2 Tests

Data file 3. Samples distribution by period

Data file 4. March SARS-CoV-2 Sanger sequencing results of samples

Data file 5. October SARS-CoV-2 Sanger sequencing results of samples

Data are available under the terms of the [Creative Commons Zero “No rights reserved” data waiver](#) (CC0 1.0 Public domain dedication).

Ethical approval

The ethical Committee of the General Coordination for Strategic Development in Health, as stipulated by the Ecuadorian Ministry of Public Health, approved the research described herein (Reference number: 074-2020). Data were anonymized for the purpose of protecting private information.

Author contributions

P.L-G.: methodology, editing, funding acquisition, methodology first draft; G.L-M.: conceptualization, editing, first draft; D.O-P.: conceptualization, methodology, supervision, final draft preparation, writing, editing; M.L-A.: conceptualization, final draft preparation, writing, editing; G.M-C.: methodology, editing, first draft, P.G-A.: methodology; G.S.: methodology, first draft; J.Z. and C.Z-S.: methodology, editing, funding acquisition.

References

Arora P, Pöhlmann S, Hoffmann M: **Mutation D614G increases SARS-CoV-2 transmission**. *Signal Transduct Target Ther*. 2021; 6: 1–2.
[Publisher Full Text](#)

Bar-On YM, Flamholz A, Phillips R, et al.: **SARS-CoV-2 (COVID-19) by the numbers**. *Elife*. 2020; 9: e57309.
[PubMed Abstract](#) | [Publisher Full Text](#)

- Barua S, Hoque M, Kelly PJ, *et al.*: **High-resolution melting curve FRET-PCR rapidly identifies SARS-CoV-2 mutations.** *J Med Virol.* 2021; **93**: 5588–5593.
[Publisher Full Text](#)
- Bezerra MF, Machado LC, Carvalho V, *et al.*: **A Sanger-based approach for scaling up screening of SARS-CoV-2 variants of interest and concern.** *Infect Genet Evol.* 2021; **92**: 101348–104910.
[Publisher Full Text](#)
- CEO Provincial Pichincha: **Situación cantonal por COVID-19.** [WWW Document]. 2021.
[Reference Source](#)
- Chakraborty D, Agrawal A, Maiti S: **Rapid identification and tracking of SARS-CoV-2 variants of concern.** *Lancet (London, England).* 2021; **397**: 1346–1347.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Chan CTM, Leung JSL, Lee LK, *et al.*: **A low-cost TaqMan minor groove binder probe-based one-step RT-qPCR assay for rapid identification of N501Y variants of SARS-CoV-2.** *J Virol Methods.* 2022; **299**: 114333.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Dao MH, Phan LT, Cao TM, *et al.*: **Genome-wide analysis of SARS-CoV-2 strains circulating in Vietnam: Understanding the nature of the epidemic and role of the D614G mutation.** *J Med Virol.* 2021; **93**: 5660–5665.
[Publisher Full Text](#)
- Elizondo V, Harkins GW, Mabvukure B, *et al.*: **SARS-CoV-2 genomic characterization and clinical manifestation of the COVID-19 outbreak in Uruguay.** *Emerg Microbes Infect.* 2021; **10**: 51–65.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Hashemi SA, Khoshi A, Ghasemzadeh-moghaddam H, *et al.*: **Development of a PCR-RFLP method for detection of D614G mutation in SARS-CoV-2.** *Infect Genet Evol.* 2020; **86**: 104625.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Isabel S, Graña-Miraglia L, Gutierrez JM, *et al.*: **Evolutionary and structural analyses of SARS-CoV-2 D614G spike protein mutation now documented worldwide.** *Sci Rep.* 2020; **10**: 14031.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Islam MT, Alam ARU, Sakib N, *et al.*: **A rapid and cost-effective multiplex ARMS-PCR method for the simultaneous genotyping of the circulating SARS-CoV-2 phylogenetic clades.** *J Med Virol.* 2021; **93**: 2962–2970.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Johns Hopkins University & Medicine: **COVID-19 Map** [WWW Document]. 2021.
[Reference Source](#)
- Jørgensen TS, Blin K, Kuntke F, *et al.*: **A rapid, cost efficient and simple method to identify current SARS-CoV-2 variants of concern by Sanger sequencing part of the spike protein gene.** *medRxiv.* 2021.2021.03.27.21252266.
- Kim D, Lee J-Y, Yang J-S, *et al.*: **The Architecture of SARS-CoV-2 Transcriptome.** *Cell.* 2020; **181**: 914–921.e10.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Kirby T: **South America prepares for the impact of COVID-19.** *Lancet Respir Med.* 2020; **8**: 551–552.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Liu P, Lu L, Xu M, *et al.*: **A novel multiplex PCR for virus detection by melting curve analysis.** *J Virol Methods.* 2018; **262**: 56–60.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Loureiro CL, Jaspe RC, D'Angelo P, *et al.*: **SARS-CoV-2 genetic diversity in Venezuela: Predominance of D614G variants and analysis of one outbreak.** *PLoS One.* 2021; **16**: e0247196.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Meng Q, Wang X, Wang Y, *et al.*: **Detection of the SARS-CoV-2 D614G mutation using engineered Cas12a guide RNA.** *Biotechnol J.* 2021; **16**: 2100040.
[Publisher Full Text](#)
- Micochova P, Collier D, Ritchie A, *et al.*: **Combined Point-of-Care Nucleic Acid and Antibody Testing for SARS-CoV-2 following Emergence of D614G Spike Variant.** *Cell Reports Med.* 2020; **1**: 100099.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Molina-Mora JA, Cordero-Laurent E, Godínez A, *et al.*: **SARS-CoV-2 genomic surveillance in Costa Rica: Evidence of a divergent population and an increased detection of a spike T11171 mutation.** *Infect Genet Evol.* 2021; **92**: 104872.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Niranji SS, Al-Jaf SMA: **Comments on 'Development of a PCR-RFLP method for detection of D614G mutation in SARS-CoV-2'.** *Infect Genet Evol.* 2021; **87**: 104661.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Okoi O, Bwawa T: **How health inequality affect responses to the COVID-19 pandemic in Sub-Saharan Africa.** *World Dev.* 2020; **135**: 105067.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Ozono S, Zhang Y, Ode H, *et al.*: **SARS-CoV-2 D614G spike mutation increases entry efficiency with enhanced ACE2-binding affinity.** *Nat Commun.* 2021; **12**(12): 1–9.
[Publisher Full Text](#)
- Pandey U, Yee R, Shen L, *et al.*: **High Prevalence of SARS-CoV-2 Genetic Variation and D614G Mutation in Pediatric Patients With COVID-19.** *Open Forum Infect Dis.* 2021; **8**: 1–9.
[Publisher Full Text](#)
- Park SY, Faraci G, Ward PM, *et al.*: **High-precision and cost-efficient sequencing for real-time COVID-19 surveillance.** *Sci Rep.* 2021; **11**: 1–10.
[Publisher Full Text](#)
- Pourhoseingholi MA, Vahedi M, Rahimzadeh M: **Sample size calculation in medical studies.** *Gastroenterol Hepatol From Bed to Bench.* 2013; **6**: 14.
- Raghav S, Ghosh A, Turuk J, *et al.*: **Analysis of Indian SARS-CoV-2 Genomes Reveals Prevalence of D614G Mutation in Spike Protein Predicting an Increase in Interaction With TMPRSS2 and Virus Infectivity.** *Front Microbiol.* 2020; **11**: 2847.
[Publisher Full Text](#)
- Rhee C, Jones T, Hamad Y, *et al.*: **Prevalence, Underlying Causes, and Preventability of Sepsis-Associated Mortality in US Acute Care Hospitals.** *JAMA Netw open.* 2019; **2**: e187571.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Sarkar SL, Alam ASMRU, Das PK, *et al.*: **Development and validation of cost-effective one-step multiplex RT-PCR assay for detecting the SARS-CoV-2 infection using SYBR Green melting curve analysis.** *medRxiv.* 2021; **165**, 1–13.
- Sevillano G: **A Sanger sequencing-based method for a rapid and economic generation of SARS-CoV-2 epidemiological data: A proof of concept study to assess the prevalence of the A23403G SNP (D614G) mutation in Quito, Ecuador.** 2022.
[Reference Source](#)
- Thomson EC, Rosen LE, Shepherd JG, *et al.*: **Circulating SARS-CoV-2 spike N439K variants maintain fitness while evading antibody-mediated immunity.** *Cell.* 2021; **184**: 1171–1187.e20.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Viedma E, Dahdouh E, González-Alba JM, *et al.*: **Genomic epidemiology of SARS-CoV-2 in Madrid, Spain, during the first wave of the pandemic: Fast spread and early dominance by D614G variants.** *Microorganisms.* 2021; **9**: 1–11.
[Publisher Full Text](#)
- Vogels CBF, Breban MI, Ott IM, *et al.*: **Multiplex qPCR discriminates variants of concern to enhance global surveillance of SARS-CoV-2.** *PLOS Biol.* 2021; **19**: e3001236.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Watanabe Y, Berndsen ZT, Raghwanji J, *et al.*: **Vulnerabilities in coronavirus glycan shields despite extensive glycosylation.** *Nat Commun.* 2020; **11**: 1–10.
[Publisher Full Text](#)
- World Health Organization (WHO): **Coronavirus Disease (COVID-19)** [WWW Document]. 2021.
[Reference Source](#)
- Yuan F, Wang L, Fang Y, *et al.*: **Global SNP analysis of 11,183 SARS-CoV-2 strains reveals high genetic diversity.** *Transbound Emerg Dis.* 2020; **68**: 3288–3304.

Open Peer Review

Current Peer Review Status: ? ✖

Version 1

Reviewer Report 06 May 2024

<https://doi.org/10.5256/f1000research.120227.r264014>

© 2024 Chouikha A. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Anissa Chouikha

Institut Pasteur de Tunis, Université Tunis El Manar, Tunis, Tunisia

Reviewer's report for F1000Research

Manuscript title:

A Sanger sequencing-based method for a rapid and economic generation of SARS-CoV-2 epidemiological data: A proof of concept study to assess the prevalence of the A23403G SNP (D614G) mutation in Quito, Ecuador.

Is the work clearly and accurately presented and does it cite the current literature? NO

The authors of the study detail a Sanger sequencing approach to identify the D619G mutation of the SARS-CoV-2 virus. This method aims to aid in the detection of this mutation, particularly in regions lacking Next Generation sequencing capabilities. This mutation was initially identified during the early stages of the COVID-19 pandemic and exhibited a significant impact on the virus's transmissibility. Subsequently, this mutation became prevalent in all circulating SARS-CoV-2 viruses, alongside other mutations of concern within the spike protein. These mutations gave rise to variants of concern such as alpha, delta, and omicron, as well as other variants under surveillance.

Previous studies have already documented a partial Sanger sequencing approach targeting regions within the spike protein. However, in the current context, this study may be considered outdated. The references provided may not reflect the most current literature, as the authors still discuss the D614G mutation as if it were a recent development. As of mid-2024, many real-time PCR assays have been developed for the detection of mutations of concern. Additionally, numerous laboratories have been trained in and have adopted Next Generation sequencing technologies.

Is the study design appropriate and is the work technically sound? Partially

While it's true that Sanger technology offers simplicity and the rapid detection of variants is crucial, the current study's sequencing region and primary objective appear outdated. A similar approach could be applied to detect new variants within the current circulating Omicron variants.

Is the work clearly and accurately presented and does it cite the current literature?

No

Is the study design appropriate and is the work technically sound?

Partly

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Not applicable

Are all the source data underlying the results available to ensure full reproducibility?

No source data required

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Virology, Molecular Biology, Sanger Sequencing, Next generation sequencing, PCR, Loop-Lamp amplification, Measles/Rubella Viruses, Polioviruses, Rotaviruses, SARS-CoV-2, Hepatitis viruses (HAV, HEV, HBV and (HCV)

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Reviewer Report 16 August 2023

<https://doi.org/10.5256/f1000research.120227.r178950>

© 2023 de-Paris F et al. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Fernanda de-Paris

Hospital de Clínicas de Porto Alegre, Porto Alegre, Brazil

Grazielle Motta Rodrigues

Hospital de Clínicas de Porto Alegre, Porto Alegre, Rio Grande do Sul, Brazil

The manuscript offers an alternative method to genomic surveillance based on melting curves and Sanger sequencing methods. The rapid spread of SARS-CoV-2 has shown us the importance of having methodologies capable of rapidly generating epidemiological and genomic data. The subject of the paper could contribute to the surveillance of SARS-CoV-2 variants but needs some

changes to clarify certain issues and make this approach more useful.

Issues pointed on the review form (comments for questions considered partially or unanswered):

- **Is the work clearly and accurately presented and does it cite the current literature?**

Partly. The authors discuss the genomic organization of SARS-CoV-2 and the A23403G mutation but do not fully elaborate on the significance of this mutation in the virus's phylogenetic evolution. The emergence of the A23403G mutation facilitated the rapid global spread and is found in all SARS-CoV-2 lineage B, including the variants of concern and the omicron sublineages. It is expected that all circulating variants at the moment would possess this mutation. It is crucial to include this information to emphasize the importance of genomic surveillance targeting this specific mutation. A reference must be inserted in the final text about this.

- **Are sufficient details of the methods and analysis provided to allow replication by others?**

Partly. The authors discuss the RT-qPCR method utilizing the melting curve approach as a screening method to select positive samples for Sanger sequencing. However, they do not provide a description of the Sanger sequencing method itself, including the fragment amplification and purification procedures. This information is crucial for reproducibility, as the amplification reaction and PCR cycling conditions are specific to each primer pair used.

The study mentions a large number of samples included in order to generate epidemiological surveillance data. However, in the results section, the authors mention that not all samples tested positive. This should be clarified, or alternatively, only the number of samples that were successfully sequenced should be indicated.

Furthermore, the study primarily focuses on samples that may not fully represent the current scenario of SARS-CoV-2. It is important to include samples from other periods of the pandemic to provide a more comprehensive analysis and data. It is crucial to include these samples in the study since the dynamics of SARS-CoV-2 demonstrate that the virus is constantly mutating and adapting. The current subvariants exhibit genomic profiles that significantly differ from the previous variants circulating at the time of the study.

- **Are all the source data underlying the results available ensure full reproducibility?**

The authors do not describe reagents and procedures to obtain cDNA viral. SARS-CoV-2 is a RNA genome virus, so it is crucial a transcriptase reverse reaction to study viral genome by PCR and Sanger Sequencing.

- **Are the conclusions drawn adequately supported by the results?**

The conclusion should be based on the findings of Sanger sequencing. It was Sanger sequencing that correctly discriminated between mutated and non-mutated samples for the targeted viral SNP. The authors need to rewrite the conclusion to emphasize Sanger's results.

Others points to observe:

Point 1: The genomic position should be verified in the methods section. Based on the forward primer and the reference sequence MN908947, the initial position of the amplicon is 23,282. Please ensure to check this information for accuracy.

Point 2: Considering the melting curve approach used as a screening method and the low variability of the analyzed region (when compared to regions such as the RBD region), it may be

more appropriate to focus on evaluating samples that exhibit a melting temperature different from 81 °C. This approach would allow for the sequencing of samples that have a nucleotide sequence distinct from those containing the D614G mutation. Techniques such as High-Resolution Melting (HRM) have the potential to differentiate single nucleotide differences by a melting curve. This is important since the D614G mutation is what facilitated the rapid global spread and is present in all SARS-CoV-2 lineage B, including variants of concern and omicron sublineages. Therefore, it is expected that all samples would contain this mutation.

Point 3: The discussion section should address why the utilization of the Sanger sequencing methodology in this case serves as an alternative approach, highlighting the benefits it offers compared to whole-genome sequencing (WGS) technologies. These benefits of a targeted approach to analyze a single nucleotide polymorphism (SNP) using Sanger sequencing may include factors such as process time, cost-effectiveness, simplified analysis, and the ability to specifically identify and monitor the A23403G SNP, and this could be highlighted in the discussion section.

Point 4: The prevalence equation in the results section is unnecessary and can be removed.

Point 5: The English needs to be revised. Some words like 'gathered' or 'arguably' do not appear to be commonly used in scientific writing.

Point 6: In the methods section, the authors state that 'both genes were analyzed using a LightCycler® 480 Real-Time PCR', but they only cite the E gene. Please verify this information."

Point 7: The Table 1 in the results section presents the percentages of positive cases in both periods included in the study. This data appears to be unnecessary in table format and could be described in the text.

Point 8: The Table 3 in the results section presents the prevalence of the A23403G mutation using WGS data obtained from GISAID during the same period as the study. This information should be described more comprehensively. The data highlights the limited number of Ecuador sample sequences deposited in GISAID, underscoring the urgent need for genomic surveillance, which can be facilitated through the methodology proposed in this study. This information should be reinforced in the authors' discussion to support the importance of searching for the A23403G mutation.

Is the work clearly and accurately presented and does it cite the current literature?

Partly

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

Yes

Are all the source data underlying the results available to ensure full reproducibility?

Partly

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Molecular laboratory diagnosis

We confirm that we have read this submission and believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however we have significant reservations, as outlined above.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research