JMID/
Journal of Microbiology and Infectious Diseases

2022; 12 (3):78-88
doi: 10.5799/jmid.1175386

RESEARCH ARTICLE

Sequencing of S and N genes of SARS-CoV-2 strains circulating in Cuba during March- September 2020

Lissette Pérez[1], Yahisel Tejero[1], Mirtha Aguado[2], Odalys Valdes[1], Mayling Álvarez[1], Guelsys Gonzalez[1], Vivian Kourí[1], María G Guzmán[1]

[1]*Pedro Kourí Tropical Medicine Institute, Department of Virology, Havana, Cuba*
[2]*University of Havana, Faculty of Biology, Center for Proteins Study, Havana, Cuba*

ABSTRACT

Objectives: The first confirmed cases of COVID-19 in Cuba were reported on March 11, 2020, followed by multiple introductions of infected travelers from Europe, America, and Asia. This work aimed to characterize the SARS-CoV-2 strains circulating in Cuba from March to September 2020 by partial nucleotide sequencing of the S and N genes.

Methods: Between March and September 2020, 38 nasopharyngeal exudates from 38 SARS-CoV-2 patients were received at the National Reference Laboratory for Influenza and Respiratory Viruses at the Institute of Tropical Medicine "Pedro Kourí" (IPK). The Sanger sequencing method was used to amplify and sequence a 2539 bp fragment of the spike gene (from position 22020 to 24550) and a 370 bp of the nucleoprotein gene (from position 28340 to 28710). The GISAID database was used to identify the mutation profile of both fragments, and phylogenetic analysis was used to confirm the clades. In addition, clinical and epidemiological data from patients were gathered.

Results: There were 34 and 25 sequences from S and N genes, respectively. In 21 of them, both genes (S and N) were available, whereas, in the remaining 13 and 4, only S or N sequences could be obtained. Based on the presence of the D614G mutation, 32 samples (84.2%) were classified as clade G of SARS CoV-2, and two were classified as Wuhan. No classification was possible in the remaining four (where only the N sequence was available). In one sample each, five different mutations were detected in clade G samples: L517F, L517X, N603T, A846V, and E281V. The 26 N sequences obtained were 100.0% identical to those circulated in most countries.

The G30R mutation was detected in an infected patient in Cuba. Fourteen of the 38 patients studied were imported cases. The first three cases detected with COVID-19 in Cuba were clade G and originated in Italy. Ten individuals were asymptomatic, four presented severe forms of the disease (two fatal), and the remaining presented mild symptoms. No relationship was observed among the clades or the mutational profile with the clinical features, country of origin, and Cuban provinces.

Conclusion: The early establishment of SARS-CoV-2 genetic surveillance in Cuba was helpful for tracking the epidemic. It demonstrated that the SARS-CoV-2 clade G was introduced initially and was the variant that circulated in the country during 2020, although the Wuhan strain was also detected. *J Microbiol Infect Dis 2022; 12(3):77-88.*

Keywords: *sequencing, S gene, N gene, SARS-CoV-2, Cuba*

INTRODUCTION

In December 2019, the Chinese government reported several cases of pneumonia in the Chinese city of Wuhan. The causative agent

Correspondence: Dr. Vivian Kourí, Pedro Kourí Tropical Medicine Institute, Department of Virology, Havana, Cuba
E-mail: vkouri@ipk.sld.cu
Received: 28 January 2022 Accepted: 17 July 2022

was a novel coronavirus called 2019-nCoV after whole genome sequencing [1]. Later, the International Committee on Taxonomy of Viruses (ICTV) officially designated the virus SARS-CoV-2. On March 11, 2020, the WHO declared the COVID-19 outbreak a pandemic, [1] which affected the health care systems worldwide as never before. At the time of the present report (October 2020), around 35.6 million people had been diagnosed with the infection, and over one million deaths were recorded. The countries most severely affected were the USA, Brazil, India, Russia, Mexico, and most European countries [2]. Cuba was no exception; on March 11, 2020, the country's first three imported cases of COVID-19 were diagnosed [3]. Increased cases were subsequently observed; however, the epidemic showed reasonable spreading during 2020. Between March and December 2020, the cumulative incidence of cases was 107.8 per 100 000 population, and 146 individuals died (fatality rate of 1.22 %) [4,5].

The main symptoms of COVID-19 include fever, cough, dyspnea, and muscle pain, but additional symptoms such as diarrhea and vomiting can also occur [6]. SARS-CoV-2, like the rest of the Coronaviridae family, has an unsegmented 30kb single-stranded RNA genome. This genome encodes six reading frames (ORF); four structural proteins: spike (S), nucleocapsid (N), envelope (E), and membrane (M), as well as a large region at the fifth end occupied by ORF1ab, which encodes 16 nonstructural proteins including viral polymerase. The S and N gene products are the significant determinants of antibody response in humans; the S protein also participates in the cellular receptor binding [6].

The genetic characterization of viral pathogens is the basis for developing diagnostic protocols, vaccines, and antiviral drugs. This strategy is also a valuable public health tool to monitor outbreaks, register new lineages, and control diseases through molecular epidemiological studies [7]. Indeed, genomic surveillance allowed the identification of SARS-CoV-2 as a new viral species a few weeks after the first outbreak in China, reinforcing the importance of adopting this tool in pathogen surveillance [8].

Furthermore, the occurrence of mutations is a natural phenomenon in the evolution of viruses, and some of them define the genetic groups that circulate globally, and SARS-CoV-2 is no exception. Since the original Wuhan virus was detected, it has evolved into several clades and variants. The GISAID database recognizes eight clades with 95.0% sequence agreement. Each clade is assigned a letter, and the number of clades has increased throughout the evolution of the pandemic: clades S and L separated initially, then L evolved into V, and G. Clade G gave rise to GH, GR, and GV, with GR subsequently changing into GRY. All clade G (GH, GR, GV, and GRY) had the D614G mutation in the virus spike glycoprotein, which was also present in the concern variants reported later [9].

Research published by WHO in 2020 showed how the SARS-CoV-2 genome has evolved as it has spread worldwide. Within the six clades and 14 subclades identified in this study, the D614G variant was the most frequent SARS-CoV-2 clade [10,11].

At the beginning of the pandemic in January 2020, the clades L, S, and O were the majority. These all declined over the next year as clade G and its descendants increased in number. As of March 2021, the clades L, S, and O were almost extinct, with the new GRY clade accounting for the majority of the G clades [12]. The replacement of glycine (D614G) for aspartic acid 614 of the Spike protein resulted in an increase in the infectivity and transmissibility of the virus. As a result, the other original clades (L, S, O) were rapidly replaced by clade G [10].

In Cuba, the first individuals diagnosed with SARS-CoV-2 were three travelers from Italy. The virus was subsequently detected in other imported cases, mainly from European and American countries, followed by autochthonous transmission, as expected [3].

In this context, this work aimed to characterize the SARS-CoV-2 strains that have been circulating in Cuba since the beginning of the epidemics from March to September 2020 by partial S and N gene sequencing.

METHODS

Study design and samples

A cross-sectional study was conducted for the genomic analysis of 38 clinical specimens (nasopharyngeal swabs) from 38 individuals diagnosed with SARS-CoV-2. The research was conducted in the National Reference

Laboratory for Influenza and Respiratory Viruses at the Institute of Tropical Medicine "Pedro Kourí" (IPK) in the period from March to September 2020.

The subjects and samples were included as part of the genomic surveillance for SARS-CoV-2 established in the country since the beginning of the epidemic. For the selection of samples, we initially focused on travelers from different countries reporting circulation of SARS-CoV-2 However, when autochthonous circulation was suspected; we sequenced samples from different geographic locations in Cuba reporting an increase in the incidence of cases from a different period, and of different severity of clinical cases [13].

The Sanger sequencing method was used to sequence the partial S and N genes of SARS CoV-2. Unfortunately, not all the samples in the study yielded quality results during the amplification or sequencing of S and N genes. Therefore not all samples have data for both genes.

Methodology

The extraction of SARS-CoV-2 RNA from the clinical samples was performed using automatic extraction with QIAcube equipment and the QIAamp Viral RNA Mini-Kit (Qiagen, Germany), following the manufacturer's instructions.

The commercial kit One-Step RT-PCR (QIAGEN, Germany) was used to synthesize cDNA and amplify a fragment of 2649 bp of the S gene (from position 21976 to 24625) and 370 bp of the N gene (from position 28340 to 28710).

Reverse transcription required 45 minutes at 48 °C, followed by 15 minutes of initial denaturation at 95^0C and 45 cycles of denaturation at 95^0C for 1 minute, hybridization at 54^0C for 1 minute, and elongation at 68^0C for 2 minutes. Finally, the last elongation phase was performed at 68^0C for 5 minutes and stored at 4^0C.

Three cycles of RT-PCR for different virus fragments were used to amplify a segment of the S gene. The first cycle amplifies a fragment from position 21976 to 22993, the second from position 22847 to 23812, and the third from position 23681 to 24625. (Table 1). [14] Only one RT-PCR reaction per sample was required for the N gene. The primers used had been previously described by Protocols for SARS-CoV-2 sequencing (Table 1) [14].

Analysis and purification of PCR products

The PCR results were analyzed on a 1.0% agarose gel stained with Gel Red Gel. 5 µL of nucleic material was used, along with 2 µL of Gel Pilot Loading Dye run indicator (5X) (QIAGEN, Germany). By comparing the migration pattern of the molecular weight marker 1kb Gel Pilot Ladder (QIAGEN, Germany), the right size of the amplification was determined.

Sanger nucleic acid sequencing

5 to 6 primers were used on each S gene fragment. A single consensus sequence was derived from two (1736 bp) or three (2539 bp) of the fragments in those viruses where it was possible. Four primers were used to establish a single consensus sequence for the N gene (Table 2).

The sequence reaction mix included 8 µL of DTCS Quick Star Master Mix (Dye Terminator Cycle Sequencing (DTCS) Quick Start Kit (Beckman Coulter, USA), 6 µL of H2O, 1 µL of each primer, and 5 µL of purified DNA, for a total reaction volume of 20 µL. Fifty cycles at 96^0C for 30 seconds, at 50^0C for 20 seconds, and at 60^0C for 4 minutes were performed in this reaction. As described in the commercial DTCS Quick Star Master Mix kit (Beckman Coulter, EU), sequenced products were purified and analyzed on a Beckman Coulter automatic sequencer model CEQTM8800 using the raw data analysis process for PCR products. The primers described in Protocols for SARS-CoV-2 sequencing by the Atlanta CDC for S and N genes were used (Table 2). [14].

Sequence editing and mutation analysis

The sequences obtained were assembled and edited using the Sequencher TM Version 4.10.1 computer tool (Genes Codes Corporation, USA). The complete Wuhan-Hu-1 sequence (NC_045512.2, severe acute respiratory syndrome coronavirus two isolates, complete genome) was used as the reference nucleotide sequence. Mutations were identified using the CoVsurver interpretation algorithm: Mutation Analysis of hCoV-19, from the GISAID database (https://www.gisaid.org/epiflu-applications/covsurver-mutations-app/).

Phylogenetic Analysis

The sequences were aligned using Sequencher TM Version 4.10.1 (Genes Codes Corporation, USA) and minimally edited with Mega 6. The evolutionary history was inferred using the NJ method, based on the Tamura-3 model parameters of the Mega-6 program [15]. The evolutionary history of the analyzed sequences was represented by a consensus tree derived from 1000 copies. The Tamura-3 parameter method was used to calculate evolutionary distances, and the gamma distribution was used to estimate the percentage of variation between the sites. Thirty-three reference sequences were compared with the Cuban sequences obtained in this study.

Statistical analysis

Descriptive statistics and graphics were obtained using Microsoft Office Excel 2010.

Ethics Statement

IPK Ethics Committee in Havana, Cuba, approved the study protocol. The epidemiological and clinical information was obtained from the registry of cases of the surveillance system at the Cuban Ministry of Public Health.

RESULTS

Overall, 34 sequences from the S gene out of 38 samples were obtained. For the N gene, 25 sequences were available. In 21 of the samples, both genes (S and N) were retrieved. However, only S and N sequences could be completed in the remaining 13 and 4 samples, respectively (Table 4).

Thirty-two samples (84.2%) were classified as clade G of SARS-CoV-2, based on the presence of D614G mutation, and two were classified as Wuhan, based on the phylogenetic analysis of the sequence obtained from the S gene and GISAID database (https://www.gisaid.org) (Figure 1). However, it could not be defined if they were G, GH, or GR, since the rest of the mutations were found outside the sequenced region [16]. No classification was possible in the remaining four (where only the N sequence was available). In clade G samples, five different mutations were detected: L517F, L517X, N603T, A846V, and E281V in one sample each [12,17]. The 25 N sequences obtained were 100.0% identical to those circulated in most countries. Only the G30R mutation was detected in an infected patient in Cuba (Table 4, Figure 2).

Table 4 presents the epidemiological and clinical data of the participants in the research. Twenty-one patients (55.3 %) were male, with a mean age of 46 (between 3 and 82 years). Ten (26.3 %) were asymptomatic at the time of the sample, four (10.5 %) acquired serious illnesses, two of which had fatal outcomes (5.3 %), and 24 (63.2%) had a moderate condition. Fourteen (36.9%) of the 38 cases were imported, with ten (71.4%) arriving from Europe. The current investigation includes the first three COVID-19 cases diagnosed in Cuba (hCoV-19/Cuba/949/2020, hCoV-19/Cuba/952/2020, and hCoV-19/Cuba/958/2020). Except for one individual from Spain infected with the original strain Wuhan, all imported patients were infected with clade G of SARS CoV-2.

The present study analyzed 24 samples from autochthonous cases from seven provinces in Cuba, including 11 from a local outbreak that occurred in Ciego de Avila province on September 20, with a high incidence of cases. Clade G was identified in all but one of the cases (Table 4).

DISCUSSION

The samples from the first three patients diagnosed with SARS-CoV-2 in Cuba (imported cases from Italy) and several other imported cases detected by surveillance were included in this investigation [3]. Most patients were infected with the clade G group of SARS-CoV-2, regardless of the country of origin. Consequently, the viral variant disseminated in Cuba was responsible for the first wave of Cuban epidemics, including the local outbreak in Ciego de Avila province in September. This finding agrees with the clades described in several SARS-CoV-2 outbreaks in different countries, which claim that group G accounts for more than 85.0% of the sequences published worldwide [8,17].

Table 1. Primers for the amplification (RT-PCR) of the S and N gene fragments.

S Gen	Position	Primers
Fragment 1		
28 F	21976	CCATTTTTGGGTGTTTATTACC
28 R	22993	TGCTACCGGCCTGATAGATTTC
Fragment 2		
29 F	22847	TTACAGGCTGCGTTATAGCTTGG
29 R	23812	TGCTGCATTCAGTTGAATCACC
Fragment 3		
30 F	23681	ACTCTAATAACTCTATTGCCATACCCAC
30 R	24625	CAGAAGCTCTGATTTCTGCAGC
	N Gen	
36F	28694	CACCAAAAGATCACATTGGCAC
37R	29873	TTTTGTCATTCTCCTAAGAAGC

Table 2. Primers of S and N genes are used during sequencing.

	Position	Primers
	S Gen	
Fragment 1		
28 F	21976	CCATTTTTGGGTGTTTATTACC
28 F	21996	CCACAAAAACAACAAAAGTTGG
29 F	22847	TTACAGGCTGCGTTATAGCTTGG
29 F	22864	GCTTGGAATTCTAACAATCTTG
Fragment 2		
29 F	22864	GCTTGGAATTCTAACAATCTTG
28 R	22993	TGCTACCGGCCTGATAGATTTC
66 L	23182	TTCAACTTCAATGGTTTAACAGGCAC
29 R	23795	TCACCACAAATGTACATTGTAC
Fragment 3		
30 F	23681	ACTCTAATAACTCTATTGCCATACCCAC
67 R	24002	TTGCTTGGTTTTGATGGATCTGG
30 R	24610	CTGCAGCTCTAATTAATTGTTG
30 R	24625	CAGAAGCTCTGATTTCTGCAGC
	N Gen	
36F	28716	CCGCAATCCTGCTAACAATGC
35th	28855	TGAACTGTTGCGACTACGTGATG
36R	29724	TGTGGTGGCTCTTTCAAGTCC
36F	28716	CCGCAATCCTGCTAACAATGC

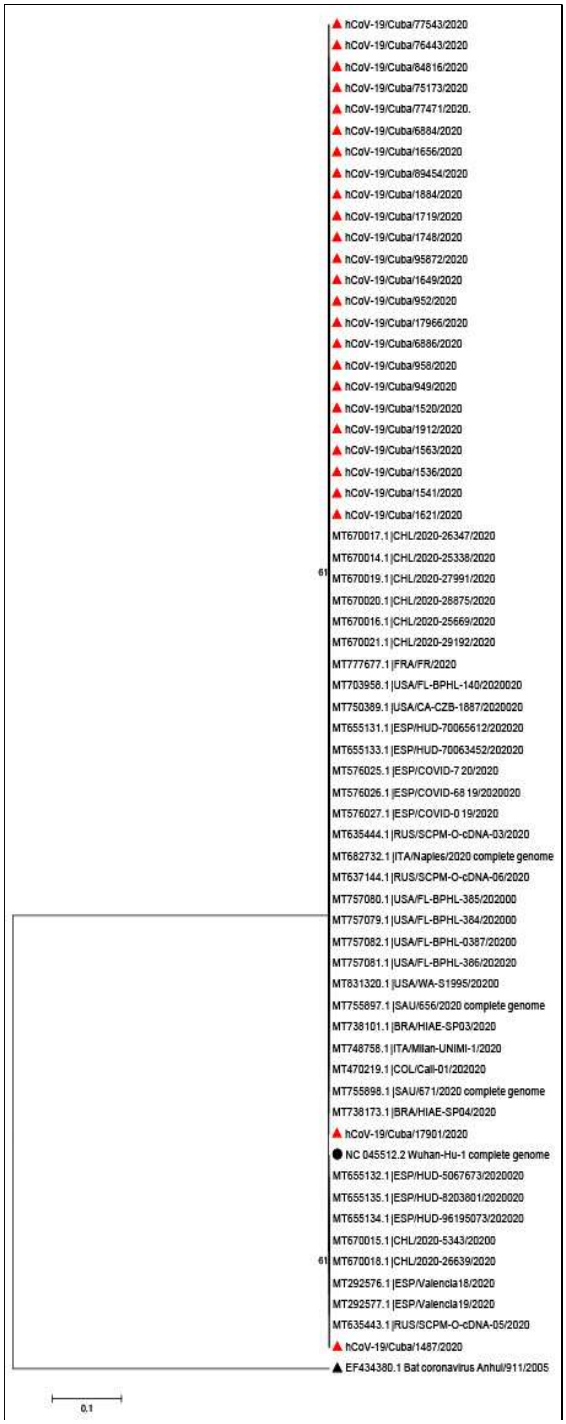

Figure 1. The analysis included 37 reference nucleotide sequences of the S gene and 26 Cuban sequences from the present study (617 bp). The evolutionary history was inferred using the NJ method, based on the Tamura-3 parameter model of the Mega-6 program (5). The consensus tree, inferred from 1000 replicates, was used to represent the evolutionary history of the analyzed sequences. The samples of the present study are marked with a red triangle, the rest are reference sequences, and the bat sequence was used as the outside group, marked with a black triangle. A circle in black shows the Wuhan-Hu-1 sequence.

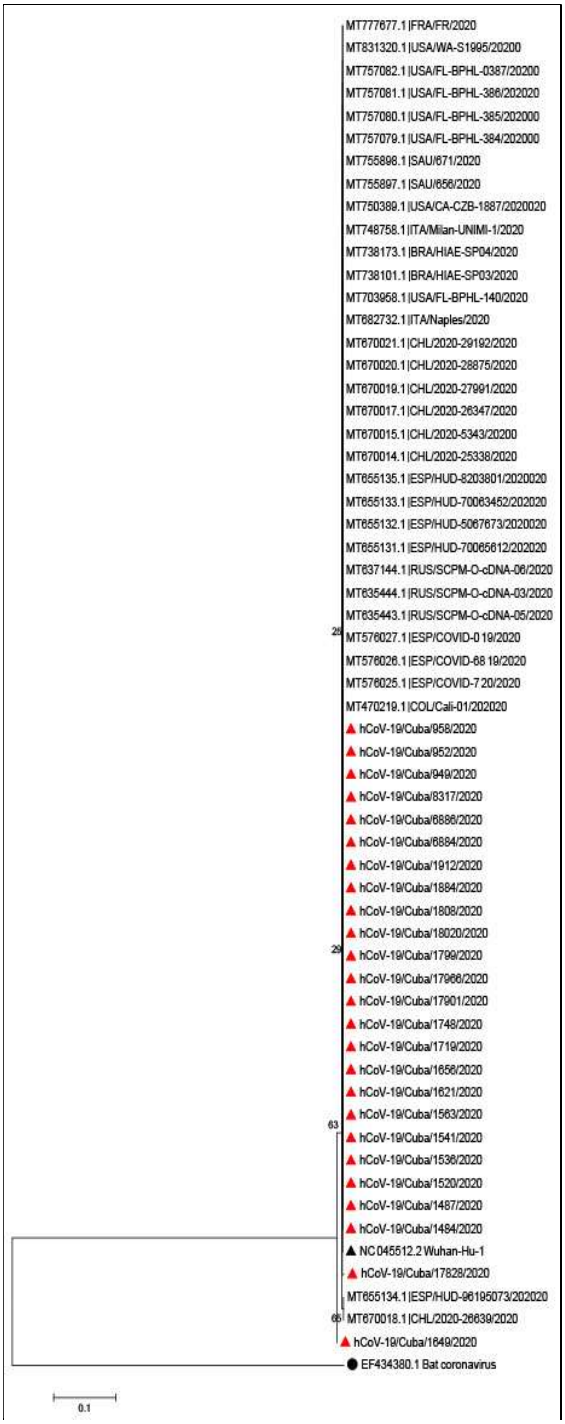

Figure 2. The analysis included 33 reference nucleotide sequences of the N gene and 25 Cuban sequences from the present study (353 bp). The evolutionary history was inferred using the NJ method, based on the Tamura-3 parameter model of the Mega-6 program (5). The consensus tree, inferred from 1000 replicates, was used to represent the evolutionary history of the analyzed sequences. The samples of the present study are marked with a red triangle, the rest are reference sequences, and the bat sequence was used as the outside group, marked with a black triangle. A circle in black shows the Wuhan-Hu-1 sequence.

Table 4. Epidemiological, clinical, and genomic information of SARS-CoV-2 S and N genes from the studied individuals.

Sample ID	Date of sample collection	Country of origin	Country of Residence	Province of Residence/Place of Diagnosis	Sex	Age	Clinical evolution	Mutations (S gene)	Mutations (N gene)	Clade
hCoV-19/Cuba/949/2020	03/11/2020	Italy	Italy	Havana	F	60	Asymptomatic	D614G	No mutations	G
hCoV-19/Cuba/952/2020	03/11/2020	Italy	Italy	Havana	F	57	Asymptomatic	D614G, L517F	No mutations	G
hCoV-19/Cuba/958/2020	03/11/2020	Italy	Italy	Havana	M	65	Severe Respiratory Distress, Pneumonia. Deceased	D614G, L517X	No mutations	G
hCoV-19/Cuba/1487/2020	03/19/2020	Spain	Spain	Havana	M	34	Upper Respiratory Infection, fever, cough	No mutations	No mutations	Wuhan
hCoV-19/Cuba/1520/2020	03/19/2020	Belgium	Belgium	Pinar del Rio	F	25	Flu-like disease, fever, rash, diarrhea	D614G, A846V	No mutations	G
hCoV-19/Cuba/1536/2020	03/19/2020	Cuba	Cuba	Havana	M	31	Flu-like disease, diarrhea, cough, nasal congestion	D614G	No mutations	G
hCoV-19/Cuba/1541/2020	03/19/2020	Guyana	Cuba	Matanzas	F	36	Upper Respiratory Infection, fever, cough, rhinorrhea	D614G	No mutations	G
hCoV-19/Cuba/1484/2020	03/19/2020	Mexico	Mexico	Havana	M	82	Upper Respiratory Infection, fever, cough, pharyngitis	Not	No mutations	NOT
hCoV-19/Cuba/1563/2020	03/20/2020	Cuba	Cuba	Las Tunas	F	38	Cough	D614G	No mutations	G
hCoV-19/Cuba/1621/2020	03/20/2020	Cuba	Cuba	Matanzas	M	23	Fever, cough, rhinorrhea	D614G, N603T	No mutations	G
hCoV-19/Cuba/1649/2020	03/20/2020	Spain	Spain	Havana	M	3	Fever, cough, myalgia, asthenia	D614G	No mutations	G
hCoV-19/Cuba/1656/2020	03/20/2020	Mexico	Cuba	Pinar del Rio	M	39	Fever, asthenia, rhinorrhea	D614G	No mutations	G
hCoV-19/Cuba/1719/2020	03/21/2020	Canada	Canada	Havana	F	55	Asymptomatic	D614G	No mutations	G
hCoV-19/Cuba/1748/2020	03/22/2020	Russia	France	Havana	M	45	Severe Respiratory Distress, myalgia, fever, asthenia, cough	D614G	No mutations	G
hCoV-19/Cuba/1799/2020	03/22/2020	Spain	Cuba	Havana	F	57	Upper Respiratory Infection, fever, cough, rhinorrhea	D614G	No mutations	G
hCoV-19/Cuba/1808/2020	03/22/2020	Spain	Cuba	Havana	F	64	Upper Respiratory Infection, cough, pharyngitis	D614G	No mutations	G
hCoV-19/Cuba/1884/2020	03/23/2020	Cuba	Cuba	Matanzas	F	21	Upper Respiratory Infection, fever, asthenia, cough, pharyngitis	D614G	No mutations	G

hCoV-19/Cuba/1912/2020	03/24/2020	Russia	France	Havana	F	48	Asymptomatic	E281V, D614G	No mutations	G
hCoV-19/Cuba/6884/2020	04/01/2020	Cuba	Cuba	Artemisa	M	64	Fever, asthenia	D614G	No mutations	G
hCoV-19/Cuba/6886/2020	04/01/2020	Cuba	Cuba	Artemisa	F	30	Cough	D614G	No mutations	G
hCoV-19/Cuba/17901/2020	04/01/2020	Cuba	Cuba	Holguin	F	56	Unknown	D614G	No mutations	G
hCoV-19/Cuba/18020/2020	04/01/2020	Cuba	Cuba	Havana	M	47	Unknown	Not	No mutations	Not
hCoV-19/Cuba/8317/2020	04/01/2020	Cuba	Cuba	Havana	F	31	Unknown	Not	No mutations	NOT
hCoV-19/Cuba/17966/2020	04/18/2020	Cuba	Cuba	Camaguey	M	63	Unknown	D614G	No mutations	G
hCoV-19/Cuba/17828/2020	04/18/2020	Cuba	Cuba	Holguin	F	53	Unknown	NOT	G30R	Not
hCoV-19/Cuba/95872/2020	06/07/2020	Cuba	Cuba	Havana	M	53	Unknown	D614G	Not	G
hCoV-19/Cuba/89454/2020	09/07/2020	Cuba	Cuba	Havana	M	68	Pneumonia, Severe Respiratory Distress. Deceased	D614G	Not	G
hCoV-19/Cuba/84816/2020	09/07/2020	Cuba	Cuba	Ciego de Avila	F	42	Anosmia	D614G	Not	G
hCoV-19/Cuba/76443/2020	09/07/2020	Cuba	Cuba	Ciego de Avila	F	42	Asymptomatic	D614G	Not	G
hCoV-19/Cuba/76496/2020	09/07/2020	Cuba	Cuba	Ciego de Avila	M	50	Asymptomatic	D614G	Not	G
hCoV-19/Cuba/73293/2020	09/24/2020	Cuba	Cuba	Ciego de Avila	F	48	Fever, Severe Respiratory Distress.	No mutations	Not	Wuhan
hCoV-19/Cuba/75173/2020	09/25/2020	Cuba	Cuba	Ciego de Avila	M	79	Asymptomatic	D614G	Not	G
hCoV-19/Cuba/77490/2020	09/28/2020	Cuba	Cuba	Ciego de Avila	M	47	Asymptomatic	D614G	Not	G
hCoV-19/Cuba/77543/2020	09/28/2020	Cuba	Cuba	Ciego de Avila	M	42	Rhinorrhea, headache	D614G	Not	G
hCoV-19/Cuba/77550/2020	09/28/2020	Cuba	Cuba	Ciego de Avila	M	30	Anosmia, dysgeusia	D614G	Not	G
hCoV-19/Cuba/77551/2020	09/28/2020	Cuba	Cuba	Ciego de Avila	M	40	Fever	D614G	Not	G
hCoV-19/Cuba/77471/2020	09/28/2020	Cuba	Cuba	Ciego de Avila	M	41	Asymptomatic	D614G	Not	G
hCoV-19/Cuba/77488/2020	09/28/2020	Cuba	Cuba	Ciego de Avila	M	39	Asymptomatic	D614G	Not	G

Early in the pandemic, SARS-CoV-2 variants containing the D614G mutation in the S protein that increases receptor-binding avidity rapidly became dominant in many geographic regions [18,19]. This variant was observed at a low frequency in March 2020 (26.0%) but rapidly increased by April (65.0%) and May (70.0%), indicating a transmission advantage over viruses with D614 [12].

The D614G mutation is a non-synonymous mutation that substitutes aspartic acid for glycine at amino acid position 614 of the viral spike protein. This mutation is involved in the ligand binding to the cellular receptor ACE-2. Structural analysis suggests that D614G alters the receptor binding conformation, making ACE2 binding and fusion more likely, [20] which has been associated with increased virulence, transmissibility, and viral survival capacity. Experimental work using pseudo-typed lentiviruses indicates that D614G increases infectivity in vitro [21]. More recent experimental work compared the spike 614 variants in animal models and human cell cultures using infectious cDNA clones of circulating SARS-CoV-2 strains. Enhanced replication in the upper respiratory tract and enhanced transmission [19] of the D614G variant has been demonstrated in animal models of SARS-CoV-2 infection [22]. Therefore, the D614G mutation seems to confer an increase in viral infectivity [10].

At the time of this study (October 2020), the D614G mutation was detected in 113 countries. Currently, the G (D614G) genotype is present in 170 countries (96.50% of all samples with Spike sequence). The most recent amino acid change occurred in strain hCoV-19/USA/WV-WVU-WV055489/2021, collected in April 2021 (https://www.gisaid.org).

From May to September 2020, the GISAID database revealed that clade D614G was the predominant variant in Latin America (https://covid19dashboard.regeneron.com), and no reports of any variant of concern or interest had been documented [12,23]. On September 2020, the variant Epsilon was detected in the USA, followed by Mexico in November. Gamma and Lambda variants also appeared in Brazil and Peru around that time [24].

Because the fragment size was sequenced, it was impossible to classify the clade G into GH or GR since the rest of the mutations were found outside the sequenced region. [25] It might represent a limitation for the whole analysis and genetic classification; however, we could identify the early circulation of clade G in Cuba, which is the most relevant result. This limitation did not affect the epidemiological or clinical actions taken.

Until September 2020, the L517F mutation, involved in recognition of antibodies, was detected in six sequences and three countries (United States, Spain, and Switzerland). In a deep mutational scanning experiment expressing RBD in a yeast display platform, L517M/K/R slightly increases ACE2 binding, L517G/M increases RBD surface expression. E281V mutation has been detected in 76 sequences (0.05% of all S gene sequences) and six countries: Holland (1), France (7), Taiwan (1), Taipei (1), Canada (1), and Suriname (65). This mutation has been associated with ligand binding [12,26].

The A846V mutation was detected in 79 sequences (0.06% of all S sequences) and 15 countries (United Kingdom, Norway, France, Switzerland, Holland, Belgium, Sweden, Portugal, Austria, New Zealand, Australia, United States, South Africa, Congo, China). This mutation is involved in viral oligomerization interfaces [12,26].

G30R mutation has been detected in 88 samples (0.06% of all N gene sequences) and seven countries: the UK (6), EU (4), Canada (73), Portugal (1), Belgium (2), Norway (2); however, the implications of this mutation have not been defined [12,26].

As observed in Table 4, the proportion of males and females was the same, consistent with data from the Cuban epidemics [3]. The presence of 26.31 percent of asymptomatic persons was expected, given that epidemiological monitoring in Cuba at the time included contact tracing and isolation of all positive cases, resulting in a rate of 30 to 60.0% asymptomatic individuals among identified patients [27].

Unfortunately, two cases (both older than 65 years old) had a fatal outcomes. Since the beginning of the epidemic, the impact of age and comorbidities on the severity of COVID-19 has been documented [28]. From March to May 2020, Cuba exhibited the highest fatality rate, ranging from 2.83%, 4.38%, and 3.79%,

respectively, owing to the low number of cases diagnosed and the lack of clinical management experience of this novel disease. However, reasonable control of the epidemic was achieved by the end of 2020, with a cumulative fatality rate of 0.77 and no service collapse, in contrast to the situation faced in many countries [28].

The present study confronts some critical limitations. Although the sampling selection was made according to PAHO's recommendations, [14] the number of samples tested was not statistically representative of the SARS-CoV-2 infection cases recorded between March and September 2020. Furthermore, not all of the samples included in the study yielded good quality results during amplification or sequencing; as a result, the number of samples sequenced during the study period was not homogeneous.

Whole Genome Sequencing was not performed due to the lack of resources. However, the sequences were obtained to cover a fragment of the S gene from positions 21976 to 23812.

CONCLUSIONS

The early establishment of SARS-CoV-2 genetic surveillance in Cuba was a valuable tool for tracking the epidemic. The present study showed the molecular characterization of SARS-CoV-2 in Cuba during 2020, including the first three patients diagnosed with COVID-19 in March 2020 and other imported and autochthonous cases. It demonstrated that the clade G of SARS-CoV-2 was introduced at the beginning of the pandemic, and it was the circulating variant in the country during this period. However, the Wuhan virus was also detected. The sustained genomic surveillance has allowed managing the epidemic by providing a deeper understanding of SARS-CoV-2, its evolution, and circulation.

ACKNOWLEDGMENTS

The authors want to thank Dr. Eva Harris (Berkeley University, USA), Dr. Kevin Ariën (Institute of Tropical Medicine, Antwerp, Belgium), and Drs. Jairo Mendez and Leticia Franco from PAHO for their collaboration and assistance.

Conflicts of interest: The authors declare no potential conflict of interest relating to regarding article's research, authorship, and publication.

Financial disclosure: None

REFERENCES

1. Zhu N, Zhang D, Wang W, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. The New England journal of medicine 2020; 382:727-733.

2. WHO. Coronavirus Disease (COVID-19) Situation Reports 2020. Available from: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports

3. Beldarraín E, Alfonso I, Morales I, Durán F. Primer acercamiento histórico-epidemiológico a la COVID-19 en Cuba. Ciencias Biomedicas 2020; 10(2).

4. Portal Miranda J. Hitos y alcance de la respuesta cubana a la COVID-19. Boletín de la Cooperación Técnica Cuba 2021; 25(1):4-11.

5. Portal Miranda J. Intervención sanitaria con candidatos vacunales, como estrategia temporal de enfrentamiento a la COVID-19. Revista Cubana de Salud Pública 2021; 48(1): e3513.

6. Hu B, Guo H. Characteristics of SARS-CoV-2 and COVID-19. Nature Reviews Microbiology 2021; 19:141-154.

7. McNamara RP, Caro-Vegas C, Landis JT, et al. High-Density Amplicon Sequencing Identifies Community Spread and Ongoing Evolution of SARS-CoV-2 in the Southern United States. Cell Rep 2020; 33(5):108352.

8. Koyama T, Platt D, Parida L. Variant analysis of SARS-CoV-2 genomes. Bull World Health Organ 2020; 98:495-504.

9. Rambaut A, Holmes EC, O'TooleÁ, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Natur Microbiol 2020; 5(11): 1403–1407.

10. Korber B, Fischer WM, Gnanakaran S, et al. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. Cell 2020; 182: 812-827 e819.

11. Hamed SM, Elkhatib W, Khairalla A, Noreddin A. Global dynamics of SARS-CoV-2 clades and their relation to COVID-19 epidemiology. Scientific Report 2021; 11:8435

12. GISAID: CoVsurver: Mutation Analysis of hCoV-19. 2020. [cited 2020 November 20]. Available from: https://www.gisaid.org/epiflu-applications/covsurver-mutations-app/.

13. PAHO: Guidance for SARS-CoV-2 samples selection for genomic characterization and surveillance 2021. Available from:

https://www.paho.org/en/documents/guidance-sars-cov-2-samples-selection-genomic-characterization-and-surveillance

14. Centers for Diseases Control and Prevention. Protocols for SARS-CoV-2 sequencing. 2020. Available from: https://github.com/CDCgov/SARS-CoV-2_Sequencing.

15. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Molecular Biology and Evolution 2013; 30:2725-2729.

16. Rahimi A, Mirzazadeh A, Tavakolpour S. Genetics and genomics of SARS-CoV-2: A review of the literature with the special focus on genetic diversity and SARS-CoV-2 genome detection. Genomics 2021; 113(1): 1221-1232.

17. Long SW, Olsen RJ, Christensen PA, et al. Molecular Architecture of Early Dissemination and Massive Second Wave of the SARS-CoV-2 Virus in a Major Metropolitan Area. mBio 2020; 11(6).

18. OPS/OMS: Caracterización genómica del SARS-CoV-2 y variantes circulantes en la Región de las Américas. Organización Panamericana de la Salud 2020. Available from: https://www.geuvih.org/wp-content/uploads/2020/10/Caracterizaci%C3%B3n-gen%C3%A9tica-del-SARS-CoV-2-en-las-Am%C3%A9ricas.pdf

19. Chen Z, Chong KC, Wong MCS, et al. A global analysis of replacement of genetic variants of SARS-CoV-2 in association with containment capacity and changes in disease severity. Clinical microbiology and infection: the official publication of the European Society of Clinical Microbiology and Infectious Diseases 2021; 27:750-757.

20. Laha S, Chakraborty J, Das S, Manna SK, Biswas S, Chatterjee R. Characterizations of SARS-CoV-2 mutational profile, spike protein stability and viral transmission. Infect Genet Evol 2020; 85:104445.

21. Zhang L, Jackson CB, Mou H, et al. The D614G mutation in the SARS-CoV-2 spike protein reduces S1 shedding and increases infectivity. Nat Comun 2020; 11:6013.

22. Renn M, Bartok E, Zillinger T, Hartmann G, Behrendte R. Animal models of SARS-CoV-2 and COVID-19 for the development of prophylactic and therapeutic interventions. Pharmacol Ther 2021; 228:107931.

23. WHO: WHO announces simple, easy-to-say labels for SARS-CoV-2 Variants of Interest and Concern, 2021. Available from: https://www.who.int/news/item/31-05-2021

24. Ritchie H, Mathieu E, Rodés-Guirao L, et al. Coronavirus Pandemic (COVID-19) 2020; [cited 2020 December 28]. Available from: OurWorldInData.org.

25. Sengupta A, Hassan S, Choudhuryc PP. Clade GR and clade GH isolates of SARS-CoV-2 in Asia show highest amount of SNPs. Infect Genet Evol 2021; 89:104724.

26. Chen J, Gao K, Wang R, Wei G: Prediction and mitigation of mutation threats to COVID-19 vaccines and antibody therapies. Chem Sci 2021; 12: 6929-6948.

27. CubaData [Internet]. 2020 [cited 2020 November 29]. Available from: https://covid19cubadata.github.io/#cuba.

28. PAHO: Cuba frente a la Covid-19. Boletin de la OPS/OMS en Cuba 2020; 2 (204).