



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Emergence and evolution of SARS-CoV-2 genetic variants during the Cuban epidemic



María G Guzmán^{a,1,*}, Lissette Pérez^{a,1}, Yahisel Tejero^{a,1}, Dailyn Mederos^{a,1}, Mirtha E. Aguado^b, Yanet Pintos^a, Odalys Valdes^a, Mayling Álvarez^a, Liena de Regla Ponce de León^b, Waldemar Baldoquín^a, Yenisleidys Martínez^a, Claudia Figueredo^a, Celine Naranjo^a, Lirialys Nuñez^a, Yanaris López Almaguer^a, José R de Armas Fernández^c, José A Portal Miranda^c, Vivian Kourí^a

^a Department of Virology, Pedro Kourí Tropical Medicine Institute, Havana, Cuba

^b Center of Protein Studies, Faculty of Biology, University of Havana, Havana, Cuba

^c Ministry of Public Health, Havana, Cuba

ARTICLE INFO

Keywords:
Sequencing
S gene
Sars cov-2
Cuba

ABSTRACT

The SARS CoV-2 D614G variant circulated in Cuba in 2020. New viral variants were detected after the opening of the border in November 2020. We show the results of the genomic surveillance in Cuba from December 28, 2020, to September 28, 2021 and their relationship to the epidemiological situation in the country. A total of 1,406 nasopharyngeal exudates from COVID-19 patients were processed for RNA extraction and the 1836 bp fragment of the spike gene was amplified and sequenced. The mutations present were determined using the GISAID database. Prevalence ratios were estimated by fitting Poisson univariate and multivariate regression models to investigate associations between SARS-CoV-2 variant group (VOC, non-VOC) and disease outcome. Seventeen genetic variants were detected including VOC Alpha, Beta, Gamma and Delta, one variant of interest (VOI) (Lambda) and two previous VOI (A.2.5.1 and Zeta/P.2). Beta (34.77%), Delta (24.89%) and D614G (19%) variants were the most frequently detected. By June, Delta increased in frequency, displacing Beta. Disease severity increased significantly with age and VOC (PR =1.98, IC 95%: 1.33–3.05, $p < 0.05$). Genomic surveillance allowed us to identify the upsurge of novel variants. Coinciding with the higher epidemic period, multiple variants were co-circulating. Although we cannot rule out that failure in the transmission containment measures occurred, the increase in the number of cases associated with the circulation of several variants, particularly the Beta and Delta variants is highly suggestive. A greater association of Beta variant with clinical severity and Delta variant with a greater transmissibility was observed.

Background

All viruses evolve over time. The wide and rapid global spread of SARS CoV-2 has favored the emergence of genetic viral variants. As early as February 2020, the D614G mutation within the receptor binding domain (RBD) of the spike (S) protein arose in Europe, and variants carrying this mutation rapidly became the dominants across the world. Since then, several major variants with additional infectious and clinical implications have been identified worldwide. Some of them represent a challenge for global public health because they are associated with an increase in transmissibility, virulence, and disease severity, and a de-

crease in the effectiveness of vaccines, treatments, diagnostic tools or other public health and social measures [1].

On May 31, 2021, the World Health Organization (WHO) established a new classification including variants of concern (VOC), variants of interest (VOI), and variants under monitoring (VUM). Because of the continuous evolution of SARS-CoV-2 and the constant advances in understanding the impacts of variants, these definitions may be periodically adjusted (WHO, <https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/>). Since April 2021, a global exponential increase in VOC Delta was observed. By July, Delta variant was present in almost 90% of samples from around the world, including the Americas. This trend was maintained until the end of November, when the Omicron variant was declared a new VOC, while at the same time the prevalence of other VOC and VOI declined [2]. The rapid evolution of SARS CoV-2

* Corresponding author.

E-mail address: lupe@ipk.sld.cu (M.G. Guzmán).

¹ These authors contributed equally to this work.

makes it necessary to maintain close genomic surveillance to identify the emergence or increase of VOC and VOI and to determine their impact on the epidemiological situation of each country, thus allowing the implementation of control measures.

The Republic of Cuba, located in the Caribbean, has not been exempt from the COVID-19 pandemic. In February 2020, the country established its COVID-19 Prevention and Control Plan that included extensive clinical, epidemiological and laboratory surveillance. Main actions included active search of cases and contact tracing, hospitalization of all patients and isolation of contacts, broad community and intersectoral participation, and strict surveillance in ports and airports, among other measures. The first imported and autochthonous cases were reported in March 2020 [3]. The transmission of SARS CoV-2 gradually spread throughout the country. In December 2020, Cuba reported 12,056 confirmed cases and 146 deaths. In this period, no fatalities were reported in pregnant women and children. The implemented measures allowed the country to maintain a low incidence of the disease (10.01×100,000 inhabitants) and a lethality of 1.21% [4,5]. However, in 2021 the epidemiological situation gradually became more complex, due to the opening of the borders and the arrival of travelers from countries with high transmission rates of SARS CoV-2. The cumulative incidence rate from January to May 2021 was 1117.5 per 100,000 inhabitants [4].

In February 2020, Cuba established laboratory surveillance for SARS CoV-2, including genomic surveillance. The first diagnoses were made in March at the National Reference Laboratory for Influenza and other Respiratory Viruses of the Pedro Kourí Tropical Medicine Institute (IPK). Subsequently, a network with 27 other molecular diagnostic laboratories was established. On a daily basis, IPK receives samples from the laboratory network for quality control and diagnosis confirmation. Genomic surveillance in Cuba made it possible to determine the presence of the D614G variant in the first imported and autochthonous cases. This viral variant continued circulating in 2020 [6].

Objectives: Here, we show the results of the genomic surveillance in Cuba from December 28, 2020, to September 28, 2021, and their relationship to the epidemiological situation in the country. We also aim to examine the association of identified SARS-CoV-2 variants (VOC vs non-VOC) with infection outcomes including severity and deaths.

Study design

Design, setting, and studied samples

This is a cross-sectional study, with viral genome analysis of clinical specimens obtained from patients infected by SARS-CoV-2 from December 28, 2020 to September 28, 2021. The genomic characteristics were described and the relation of the identified variants with clinical outcomes was explored. During the study period, 1406 samples of nasopharyngeal swab samples collected from individuals with a confirmed SARS CoV-2 infection were studied by nucleotide sequencing. These samples included 164 collected from travelers, 117 from fatal cases, 93 from severe and very severe cases and 1032 from non-severe clinical cases including asymptomatic individuals (surveillance samples). Samples from all provinces were represented in the dataset. All samples were re-tested for SARS CoV-2 diagnostic at IPK.

RNA extraction, PCR and nucleotide sequencing

Total RNA was extracted from nasopharyngeal swabs by using automatic extraction with QIAcube equipment and the QIAamp Viral RNA Mini Kit (Qiagen, Germany), following the manufacturer's instructions. The cDNA synthesis and amplification of a 1836 bp fragment of the S gene (positions 21,976 to 23,812) was carried out using the commercial kit One Step RT-PCR (Qiagen, Germany), following the manufacturer's recommendations. RT-PCR was performed using the primers described in protocols for SARS CoV-2 sequencing by the Centers for Disease Control and Prevention (CDC, USA) [7]. The commercial sequencing kit Dye

Terminator Cycle Sequencing (DTCS) Quick Start Kit (Beckman Coulter, USA) was used for the sequence reaction with four primers for S gene sequencing (positions 21,976 to 23,812). Sequenced products were purified following the protocol described in the commercial DTCS Quick Start Master Mix Kit (Beckman Coulter, EU) and analyzed in a Beckman Coulter automatic sequencer model CEQTM8800 using the raw data analysis procedure for PCR products.

Sequence editing and mutation analysis

The obtained sequences 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) was used as the reference nucleotide sequence. Mutations were identified using the CoVsurver interpretation algorithm: Mutation Analysis of hCoV-19, from the GISAID database [8]. Variants designation of Cuban sequences (VOC, VOI, VUM and others) were assigned according to the mutational profile previously described by the GISAID database, for each known variant [8]. Furthermore, those samples with a mutation pattern detected in five or more patients but that had not been previously described by GISAID, were named as mutational patterns (pattern 1 to 4).

Statistical analysis

Univariable analysis

Categorical variables were described using frequency and percentage; continuous variables were described using median and range. Demographic and clinical characteristics of patients in different groups (VOC, non-VOC) were compared using the Wilcoxon-Mann-Whitney rank-sum test for continuous variables and Fisher exact or Pearson χ^2 tests for categorical variables. All tests were 2-tailed, and significance was set at $p < 0.05$.

Multivariable analysis

Un-adjusted (crude) prevalence ratios (PRs) and adjusted prevalence ratios were estimated by fitting Poisson univariate and multivariate regression models with robust estimates to investigate associations between SARS-CoV-2 variant group (VOC, non-VOC) and the outcome of severe disease or death, adjusting for potential confounders (age, sex). Wald tests were used to assess associations between the outcome and variant as predictor, and also between the outcome and interaction terms between variant and age, and sex. Statistical analysis was done using R version 4.0.1 (The R Foundation for Statistical Computing, 2020–06–06).

Epidemiological data

Early in the COVID-19 epidemic, epidemiological data were reported daily and made available to the public at <https://covid19cubadata.github.io/#cuba>. This site was a resource during the data analysis [4].

Ethics

The study was approved by the Institutional Ethics Committee of the Institute of Tropical Medicine Pedro Kourí.

Results

A total of 1406 samples of nasopharyngeal swabs from the same number of patients were selected for genome analysis in the study period. Patients included 671 women (49%), 690 men (51%), median age was 44 years (IQR 25–59), with 1242 cases presumed local infection

Table 1
Demographic and clinical characteristics of studied patients by VOC and non-VOC.

Characteristic	Total N = 1406 ^a	non-VOC, N = 538 ^a	VOC, N = 868 ^a	p-value ^b
Age				0.087
≤44	542 (50%)	246 (51%)	296 (50%)	
45–54	177 (16%)	92 (19%)	85 (14%)	
55–69	197 (18%)	81 (17%)	116 (20%)	
≥70	159 (15%)	63 (13%)	96 (16%)	
Unknown	331	56	275	
Sex				0.14
Female	671 (49%)	249 (47%)	422 (51%)	
Male	690 (51%)	283 (53%)	407 (49%)	
Unknown	45	6	39	
Illness severity				<0.001
Non-severe	1196 (85%)	497 (92%)	699 (81%)	
Severe	210 (15%)	41 (7.6%)	169 (19%)	
Province				<0.001
Artemisa	51 (3.6%)	15 (2.8%)	36 (4.1%)	
Ciego de Ávila	51 (3.6%)	10 (1.9%)	41 (4.7%)	
Cienfuegos	32 (2.3%)	6 (1.1%)	26 (3.0%)	
Camagüey	63 (4.5%)	17 (3.2%)	46 (5.3%)	
Granma	62 (4.4%)	44 (8.2%)	18 (2.1%)	
Guantánamo	55 (3.9%)	44 (8.2%)	11 (1.3%)	
Holguín	37 (2.6%)	8 (1.5%)	29 (3.3%)	
Isla de la Juventud	17 (1.2%)	5 (0.9%)	12 (1.4%)	
La Habana	601 (43%)	218 (41%)	383 (44%)	
Las Tunas	47 (3.3%)	21 (3.9%)	26 (3.0%)	
Matanzas	106 (7.5%)	21 (3.9%)	85 (9.8%)	
Mayabeque	68 (4.8%)	31 (5.8%)	37 (4.3%)	
Pinar del Río	74 (5.3%)	42 (7.8%)	32 (3.7%)	
Santiago de Cuba	62 (4.4%)	22 (4.1%)	40 (4.6%)	
Sancti Spíritus	44 (3.1%)	19 (3.5%)	25 (2.9%)	
Villa Clara	36 (2.6%)	15 (2.8%)	21 (2.4%)	
Origin				0.024
Local	1242 (88%)	462 (86%)	780 (90%)	
Travel	164 (12%)	76 (14%)	88 (10%)	

^a n (%), Median (IQR).

^b Pearson's Chi-squared test; Wilcoxon rank sum test.

(88%) and 164 with a history of recent travel (12%) (Table 1). Six hundred and one (43%) patients were diagnosed in Havana, of whom 138 were classified as severe/fatal (22.9%).

In the first five months of 2021, sixteen genetic variants were detected including the four VOC at that time (Alpha, Beta, Gamma and Delta), one VOI (Lambda), two previous VOI (A.2.5 and Zeta/P.2) and four mutational patterns (local variants). The original Wuhan virus was detected in only one sample collected in Havana in January 2021 from a traveler from the USA. It is striking that most of these variants, including both VOC and VOI, were not established and were declining over time. Overall, 489 (34.8%) samples were classified as VOC Beta (B.1.351), 350 (24.9%) as Delta (B.1.617.2), 27 (1.92%) as Alpha (B.1.1.7), 1 (0.07%) as Wuhan and 2 (0.14%) as Gamma/P1 (B.1.1.28.1). Additionally, 83 (5.9%) were classified as variant A.2.5, 41 (2.91%) as B.1.623, 12 (0.85%) as B.1.575, 11 (0.78%) as Zeta/P2 (B.1.1.28.2), 2 (0.14%) as Lambda and 1 (0.07%) as B.1.519. Four mutational patterns were identified: pattern 1 (D614G+T470K) (38, 2.70%), pattern 2 (S477N+D614G) (5, 0.35%), pattern 3 (D614G+T732A) (43, 3.06%) and pattern 4 (D614G+Q677H) (16, 1.13%). Detected variants and mutational patterns are shown in Table 2.

The distribution of the variants changed over time according to the emergence and expansion of those with the greatest evolutionary advantage. At the end of December 2020, four variants, D614G, Alpha, Beta and A.2.5, were detected in the country, with D614G predominant (Table 2, Fig 1). In January 2021, six more variants were detected including the Wuhan original strain, B.1.575.1, Zeta/P2 as well as the mutational patterns 1, 2 and 3. D614G increased from 72% (18/25) in December to 82.3% (98/119) in January. Two new variants (pattern 4 and B.1.623) were detected in February, accompanied by an increase in Beta variant detection (22/147, 14.9%). Although D614G variant was still widely circulating, its detection diminished to 44.9% in that period

(66/147). In March, B.1.1.519 variant was detected and at this time, Beta and D614G circulated with a similar frequency (30.81%). From April to May, an increase in Beta circulation was observed. Additionally, the variants Gamma/P1, B.1.1.523 and Delta (B.1.617.2) were detected for the first time in April. Delta rapidly increased from 1.44% in May to 27.14% in June, 70.09% in July and 94.07% in August. By September, 100% of the studied samples were classified as Delta. In parallel, the Beta variant diminished from 60.58% in May, to 51.26% in June, 29.06% in July and 5.93% in August. Fig. 1 shows the rate of distribution per month of the major variants circulating in Cuba during the period of study.

As shown in Fig. 2, multiple variants were reported in most of the provinces; however, Beta, initially detected in Havana at the end of December in a South African traveler, progressively extended in March to Mayabeque, Ciego de Ávila, Matanzas and Guantánamo provinces and later to the rest of the country (Fig. 2). On the other hand, Delta, first detected in Matanzas province from a traveler coming from India, in two months rapidly extended from Matanzas to the rest of the provinces. The introduction of both Beta and Delta variants in a province was accompanied by a rapid increase in the incidence of cases. Alpha VOC was first detected in December in a traveler from the UK (Table 2). In February it was detected again in Havana and one month later in Palma Soriano, Santiago de Cuba province and it is not until April that cases with this variant were reported in the provinces of Villa Clara, Matanzas, Holguín, Cienfuegos, Granma and Mayabeque. However, this variant was not associated to an outbreak in any of these provinces. Variant A.2.5, described originally in California, USA [9], was identified in Cuba at the end of December 2020 in a traveler from Panama. This variant ranked fourth in detection preceded by Delta, Beta and D614G, and extended throughout the country until May. The rest of the variants were detected with less frequency (Figs. 1 and 2).

Table 2
Detected SARS CoV-2 variants and mutational patterns from December 28, 2020 to September 28, 2021.

Variant	Month and province of first detection during the study period	December 28-February N = 291 (100%)	March-June N = 781 (100%)	July-September N = 334 (100%)	Total N = 1406 (100%)
D614G ^a	December, Havana	182 (63.2)	84 (10.7)	0 (0.0)	266 (19.06)
Alpha B.1.1.7	December, Havana	3 (1.03)	24 (3.07)	0 (0.0)	27 (1.92)
Beta	December, Havana	23 (7.90)	424 (54.28)	42 (12.57)	489 (34.77)
B.1.351					
A.2.5	December, Ciego de Avila	11 (3.78)	71 (9.09)	1 (0.29)	83 (5.90)
Wuhan ^a	January, Havana	1 (0.34)	0 (0.0)	0 (0.0)	1 (0.07)
Zeta/P2	January, Cienfuegos	3 (1.03)	8 (1.02)	0 (0.0)	11 (0.78)
B.1.575.1	January, Havana	8 (2.74)	4 (0.51)	0 (0.0)	12 (0.85)
Pattern 1 ^b	January, Pinar del Rio	22 (7.56)	16 (2.04)	0 (0.0)	38 (2.70)
Pattern 2 ^c	January, Pinar del Rio	2 (0.68)	3 (0.38)	0 (0.0)	5 (0.35)
Pattern 3 ^d	January, Havana	29 (9.96)	14 (1.79)	0 (0.0)	43 (3.05)
Pattern 4 ^e	February, Camagüey	5 (1.71)	11 (1.40)	0 (0.0)	16 (1.13)
B.1.1.519	March, Havana	0 (0.0)	1 (0.12)	0 (0.0)	1 (0.07)
B.1.623	February, Mayabeque	2 (0.68)	39 (4.99)	0 (0.0)	41 (2.91)
Delta	April, Matanzas	0 (0.0)	59 (7.55)	291 (87.12)	350 (24.89)
B.1.617.2					
Gamma/P1	May, Havana	0 (0.0)	2 (0.25)	0 (0.0)	2 (0.14)
Lambda	May, Havana	0 (0.0)	2 (0.25)	0 (0.0)	2 (0.14)
C37					
B.1.523	May, Camagüey	0 (0.0)	19 (2.43)	0 (0.0)	19 (1.35)

^a Wuhan and D614G SARS CoV-2 variants had already been detected in 2020 in Cuba. In the table we refer only to samples collected for the present study.

^b Pattern 1 (D614G+T470K).

^c Pattern 2 (S477N+D614G).

^d Pattern 3 (D614G+T732A).

^e Pattern 4 (D614G+Q677H). GISAID describes two sub lineages of B.1.1 (lineages B.1.1.222 and B.1.1.322) which have the D614G+T732A mutations in the S gene fragment. In the present study, it was not possible to reach this classification because it would be necessary to sequence another gene. For this reason, the assignment of pattern 3 was maintained for all the samples with this combination of mutations.

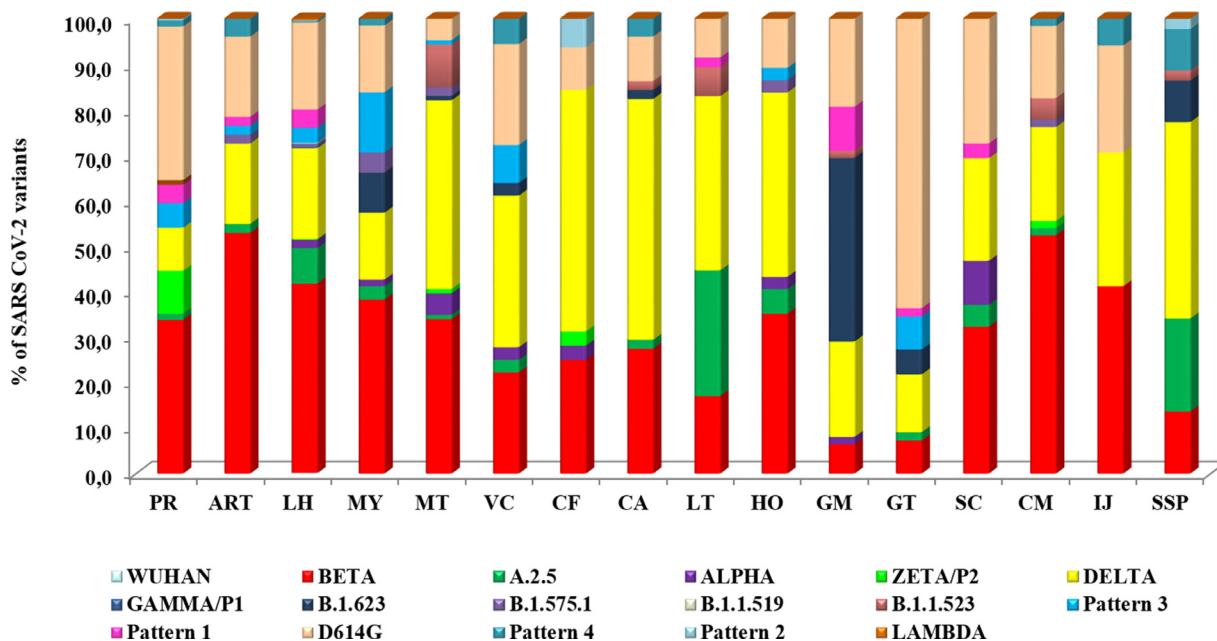


Fig. 1. Percentage of SARS CoV-2 genetic variants detection by month, Cuba, December 28, 2020 to September 28, 2021. At the top, the total number of confirmed COVID-19 cases in the country. Pattern 1: D614G+T470K, Pattern 2: S477N+D614G, Pattern 3: D614G+T732A and Pattern 4: D614G+Q677H.

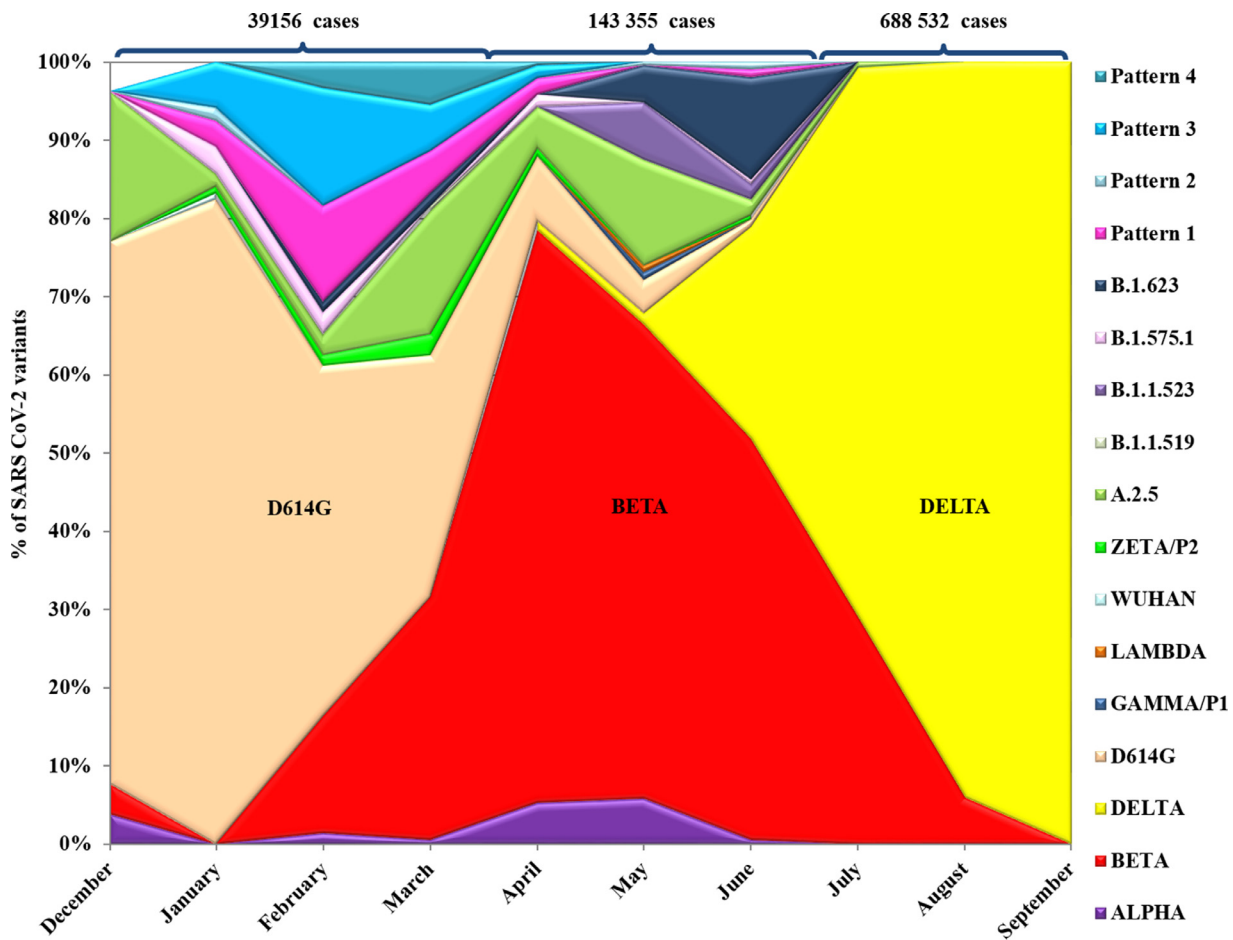


Fig. 2. Percentage of SARS CoV-2 genetic variants distributed by province, December 28, 2020 to September 28, 2021
 Pattern 1: D614G+T470K, Pattern 2: S477N+D614G, Pattern 3: D614G+T732A and Pattern 4: D614G+Q677H. Abbreviations of Cuban Provinces: PR, Pinar del Río; ART, Artemisa; LH, La Habana; MT, Matanzas; VC, Villa Clara; CF, Cienfuegos; CA, Ciego de Avila; LT, Las Tunas; HO, Holguín; GM, Granma; GT, Guantánamo; SC, Santiago de Cuba; CM, Camagüey; IJ, Municipality Isla de la Juventud; SSP, Sancti Spiritus. Western provinces include PR, ART, LH, MY, MT, Central provinces include VC, CF, SSP, CA, CM, LT, Eastern provinces include GM, HG, SC, GT.

Genomic surveillance showed the presence of other local variants not reported internationally at the moment of their detection that we nominated mutational patterns (Table 2). While pattern 2 was detected in only five samples, patterns 1, 3 and 4 were detected in a larger number of cases. Pattern 1, detected in 38 cases, was first described in a patient from Pinar del Río province in early January, but also was detected in Guantánamo, Havana, Granma, Santiago de Cuba, Las Tunas and Artemisa provinces from January to June. Pattern 3, initially recovered in Havana from imported cases from the USA and Colombia, circulated from January to April in the provinces of Havana, Guantánamo, Holguín, Mayabeque, Pinar del Río, Artemisa, Villa Clara and Matanzas, for a total of 43 cases. Pattern 4 was detected in 16 autochthonous cases in February, March and April in the Western and Central provinces, including the Municipality of Isla de la Juventud. The stratification of cases by age group, clinical severity and VOC versus non-VOC demonstrates that disease severity increased significantly with age and VOC were associated with severity (PR = 1.96, IC 95%: 1.34–2.93, $p < 0.001$) even after accounting for age as a confounder (Table 3).

The epidemiological situation of Havana (the capital of Cuba) and Matanzas provinces in 2021 deserve careful attention and illustrate in detail what it was observed in the rest of the country (Fig. 3). On December 28, 2020, four variants were detected in Havana province, including D614G (the predominant), A.2.5, Alpha and Beta. In January, D614G was detected in 43/51 (84.31%) of studied samples and four additional variants were also detected. Although in February, D614G was

still predominant (31/94, 32.97%), Beta variant was detected in 23.40% (22/94) of samples. Other variants were also detected at that time. In March, Beta increased to 45.36% (44/97) and in April was detected in 86.59% (84/97) of the samples studied. Delta variant was identified in Havana in April after Matanzas's first detection of Delta. In June, Beta variant was still predominant in Havana (74/104) and Gamma/P1 was detected for the first time in May (2/104). Further variants were also detected at this time. In June and July, Beta continued circulating but Delta rapidly increased in July (56.25%, 18/32 samples), coinciding with the peak of the epidemics in Havana city, while Beta decreased (43.75%, 14/32 samples). In August, the Delta variant was detected in 91.66% (44/48) and in September in 100% of the 53 studied samples.

The increase in SARS CoV-2 transmission in 2021 was also accompanied by an increase in severity of cases and mortality. Most of the severe, very severe and fatal cases studied were from Havana, where an association of VOC with the severity of the infection was also observed (OR; 2.88; IC 1.8–4.5; $p = 0.0000$) (Table 4), particularly with VOC Beta where 88 out of 250 cases with this variant were classified as severe or fatal, compared to 50 of 351 cases without Beta sequenced samples (OR=3.270; IC 2.20–4.85, $p = 0.0000$).

In Matanzas, D614G variant predominated during the initial months of 2021, although the A.2.5 variant was also detected. In March and April, Beta (78.78%, 26/33) and Alpha (9.09%, 3/33) variants were detected. In April, Delta variant was first identified in the province and by June, Delta was identified in 84.84% (28/33) of the studied samples.

Table 3
Association of SARS-CoV-2 VOC variants and age with disease severity, Cuba, December 28, 2020 – September 28, 2021.

Characteristic	N	Severity			Crude prevalence ratio			Adjusted prevalence ratio		
		Total N = 1406 ^a	Non-severe N = 1196 ^a	Severe N = 210 ^a	PR ^b	95% CI ^b	p-value	PR ^b	95% CI ^b	p-value
VOC		868 (62%)	699 (58%)	169 (80%)	2.55	1.84, 3.64	<0.001	1.96	1.34, 2.93	<0.001
Age	1075									
≤44		542 (50%)	525 (56%)	17 (13%)	—	—	—	—	—	—
45–54		177 (16%)	166 (18%)	11 (8.5%)	1.98	0.90, 4.18	0.077	2.08	0.95, 4.40	0.058
55–69		197 (18%)	155 (16%)	42 (32%)	6.80	3.94, 12.3	<0.001	6.52	3.78, 11.8	<0.001
≥70		159 (15%)	99 (10%)	60 (46%)	12.0	7.19, 21.3	<0.001	11.5	6.84, 20.3	<0.001
Unknown		331	251	80						
Sex	1361									
Female		671 (49%)	565 (49%)	106 (51%)	—	—	—	—	—	—
Male		690 (51%)	587 (51%)	103 (49%)	0.94	0.72, 1.24	0.7	1.14	0.81, 1.62	0.4
Unknown		45	44	1						

^a n (%).

^b PR = Prevalence Ratio, CI = Confidence Interval.

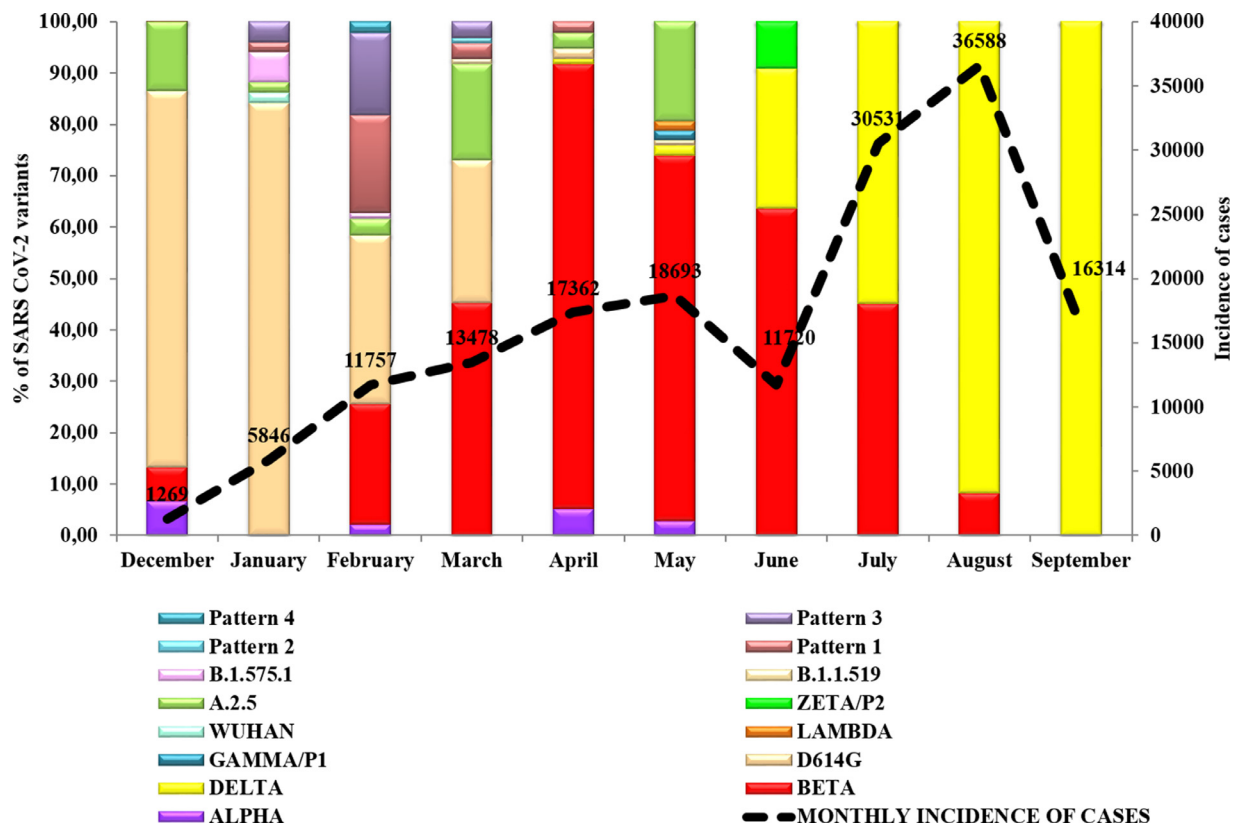


Fig. 3. Percentage of SARS CoV-2 genetic variants related to the incidence of cases per month in Havana province. Dashed lines represents all confirmed cases in Havana (incidence).

Pattern 1: D614G+T470K, Pattern 2: S477N+D614G, Pattern 3: D614G+T732A and Pattern 4: D614G+Q677H.

In July and August, only the Delta variant was detected in the studied samples (Fig. 4).

Discussion

Since the first description of SARS CoV-2, through the end of 2021, more than 5.5 million complete genomic sequences of the virus have been shared worldwide [8,10]. This has allowed monitoring of the evolution of the virus and identifying its association with changes in epidemiological, transmissibility and virulence patterns.

Globally, five VOC have been identified: Alpha, Beta, Gamma, Delta, and Omicron; additionally, there are currently two variants of interest (VOI): Lambda and Mu. In Cuba, despite the strict containment measures

established with the notification of the first COVID-19 autochthonous cases during 2020, a gradual increase in the number of confirmed cases was observed in 2021, largely associated with the introduction into the country of several variants of SARS CoV-2 after the opening of the borders in November 2020. A heterogeneous distribution of SARS CoV-2 variants was observed at national and provincial levels. Interestingly, despite the restrictions in international travel in the country, 17 genetic variants were detected, including the four VOC at that time (Alpha, Beta, Gamma and Delta), one VOI (Lambda), two previous VOI (A.2.5 and Zeta/P.2) and four mutational patterns (local Cuban variants). At the time of the present study, no Mu VOI had been detected, in contrast to other countries in South America [2].

Table 4
VOC and non VOC according to COVID-19 clinical severity in Havana city.

	Severe/very severe and fatal cases	Non severe infections	Total
VOC	111	272	383
Non VOC	27	191	218
Total	138	463	601

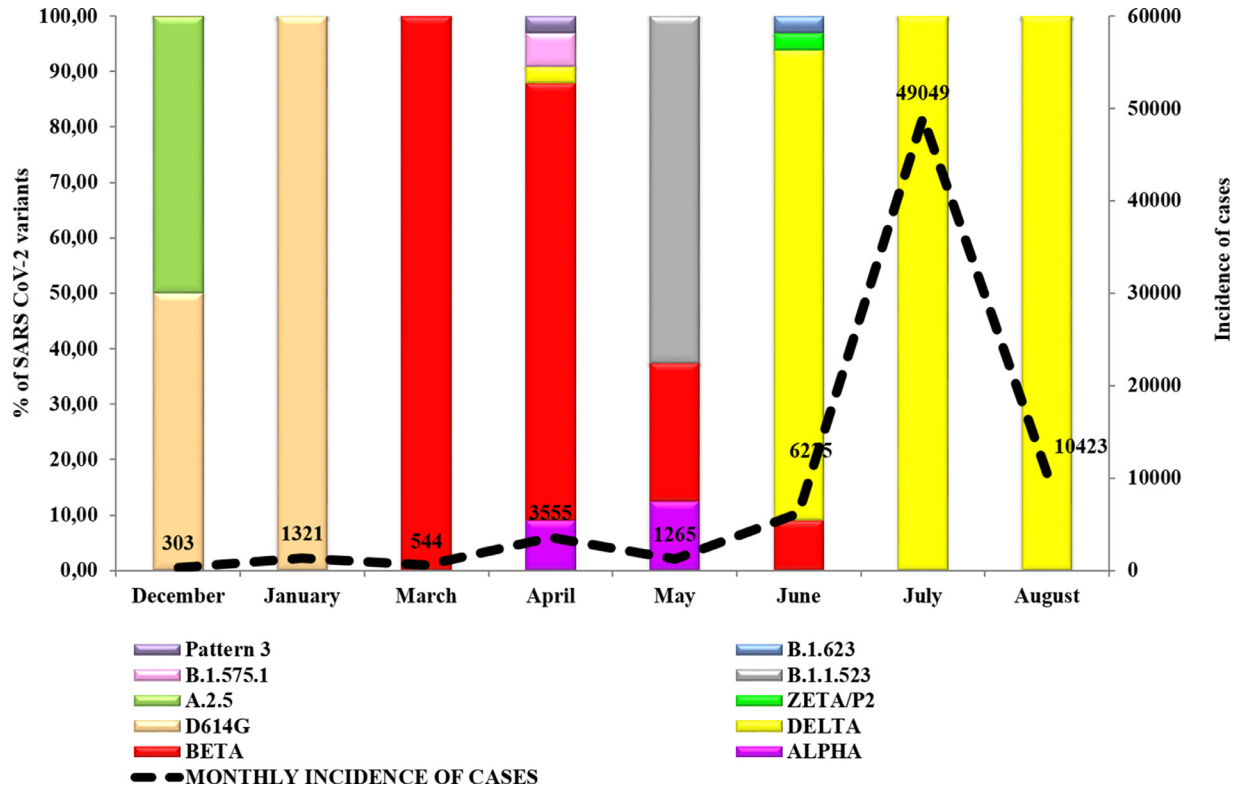


Fig. 4. Percentage of SARS CoV-2 genetic variants related to the incidence of cases per month in Matanzas province. Dashed line represents all confirmed cases in Matanzas (incidence).

Pattern 3: D614G+T732A and Pattern 4: D614G+Q677H.

We can define three epidemic periods in Cuba associated with the circulation of three major variants. The first period included the beginning of the epidemic in March 2020 until November 2020; this period was characterized by quite good control of transmission mainly due to strict epidemiological containment measures, including the lockdown of most social activities and international travel. During that period the D614G variant was detected in more than 90% of the samples sequenced [11]. The second period comprised December 2020 until June 2021, and it was characterized by a sustained increase in the incidence of cases, hospitalizations and mortality. This period can be divided into two waves, with the first wave from December 28, 2020 to February 2021, when circulation of D614G prevailed but the introduction of eleven variants of SARS CoV-2, including the VOC Beta and Alpha, occurred. At that time, 39,156 confirmed cases of COVID-19 were reported (350 cases x 100,000 inhabitants), representing an increase of 4.8 times compared to the total of reported cases from March to November 2020 (8610 cases). The second wave of this period took place between March and June 2021, with a continued increase in the incidence of cases associated with the expansion of the circulation of the Beta variant. In this period, there were 143,355 confirmed cases of COVID-19 (1281 cases x 100,000 inhabitants), an 11.89-fold increase compared to the total reported cases in 2020 and 3.41 times the total reported from December 2020 to February 2021. The third transmission period occurred between July and September, coinciding with the dissemina-

tion of the Delta variant. In this period, 688,532 cases were confirmed (6155 cases x 100,000 inhabitants), an increase of 57.11 times compared to the total reported in 2020, 16.04 times the total reported cases between December 2020 and February 2021, and 4.80 times the total reported cases between March and June 2021 [4]. The expansion of Delta was accompanied by an increase in SARS CoV-2 incidence, initially in Matanzas that gradually extended to the rest of the provinces that was associated with an increase in hospitalizations.

Havana was the province with the highest incidence rates between February and the end of May 2021 (874,3 x 100,000 inhabitants) and a lethality of 1.04%, in a context where the Beta variant was predominant (Fig. 3) [4]. It has been recognized that this variant of concern may lead to a reduction in the neutralizing antibody response to vaccines, in addition to presenting a possible increased risk of hospitalization and mortality. Cases continued to increase and reached the epidemic peak in August with an incidence of 36,588 cases x 100,000 inhabitants, corresponding with the almost exclusive circulation of the Delta strain, considered the most transmissible variant globally at that time [12]. However, when vaccination coverage with the completed three-dose schedule with the Cuban vaccines Abdala and Soberana reached 63.2%, a systematic reduction in incidence was observed [13].

Matanzas, unlike Havana, was characterized by almost complete control of the epidemic until the end of March 2021, when a first increase in the number of seriously ill and deceased patients was observed,

coinciding with the introduction of the Beta variant. This was followed by a much larger second epidemic wave that coincided with the predominance of the Delta variant in the province (Fig. 4). When vaccination coverage was between 4.0 and 35.0%, mortality increased from 1.2 to 22.7 per 100,000 inhabitants. After reaching 36.0% vaccination coverage, mortality and incidence decreased abruptly [13].

The Alpha variant (B.1.1.7), first identified in the county of Kent, England, in September 2020, was dominant in most European countries and other states such as Israel for several months in 2021. This variant was characterized by a significantly higher transmissibility. According to WHO, the Alpha variant has been reported by up to 195 countries [14]. Interestingly, although this variant was first detected in Cuba in December 2020 it did not prevail as a predominant variant, unlike reports from other countries. On the other hand, VOC Beta (B.1.351) that has been detected in 141 countries, mostly in southern African countries with a small extension to countries on all continents including Europe, expanded throughout Cuba, was the most prevalent variant for a period of four months in the Cuban epidemics, and was associated with disease severity [15, 16]. The greater expansion of Beta over Alpha VOC in Cuba could be because there were several introductions of the Beta variant, or that Beta could have gained an evolutionary advantage circulating within the Cuban population with a possible superior genetic susceptibility due to its ethnic origin [17]. Although we cannot rule out that the low number of processed samples could have resulted in lower detection rates of the Alpha variant, it is noteworthy that in provinces that had greater transmission and where genomic surveillance was increased, the presence of the Alpha variant was minimal.

The Delta variant (B.1.617.2), detected in October 2020 in India, quickly became dominant in many countries [18]. This variant is more transmissible with estimates greater than 60% compare to others previously described. According to GISAID, 90% of the global sequences from June to September 2021 were Delta [12,19].

Although we cannot rule out that failure in the local transmission containment measures occurred in this period, the increase in the number of cases associated with the circulation of several variants, in particular the Beta variant initially followed by predominance with the Delta variant, is highly suggestive of the high transmission capacity of these variants. Although Beta and Delta VOC have been associated with an increase in viral transmission and disease severity [12,20], our findings suggest a greater association of the Beta variant with clinical severity and the Delta variant with greater transmissibility.

Natural evolution of viruses highlights the relevance of genomic surveillance. It allowed us to rapidly identify the upsurge of novel variants. The present study reports the molecular epidemiology of SARS-CoV-2 viruses circulating in Cuba during nine months of 2021, coinciding with the peak epidemic period before high vaccine coverage rates were reached. Multiple variants were co-circulating, but a different predominant variant was observed in each epidemic wave, suggesting acquired genetic advantages over other previous circulating variants. At the time of writing this report, transmission in the country is low, with a concomitant reduction in the number of severe, very severe and fatal cases, due to epidemiological measures and Cuba's vaccination strategy with the Soberana and Abdala vaccines [4]. Although the introduction of Omicron in December 2021 was accompanied by an increase in transmission, the high vaccination coverage and the rapid implementation of booster doses have had a positive effect in avoiding a greater number of cases. By February 2022, 9.8 million Cubans (89.2%) were fully vaccinated and 5.9 million have received the booster dose.

Limitations of the study

Whole genome sequencing was not available for this study. The information extracted covers the S gene from positions 21,976 to 23,812. This might have produced difficulties and ambiguous classification in a few local variants (patterns 1 to 5) that could not be identified as new variants by GISAID and for samples with the same mutations in the

S gene that need to be differentiated by the mutations in other parts of the genome. This was the case with B.1.1 (lineages B.1.1.222 and B.1.1.322), which has the D614G+T732A mutations in the S gene fragment. In the present study, it was not possible to reach this classification because it would be necessary to sequence another gene. For this reason, the assignment of pattern 3 was maintained for all the samples with this combination of mutations. In the statistical analysis, other potential confounders such as ethnicity and comorbidity were not available in the datasets of the central laboratory at IPK, and could not be used in the regression models.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors thank Drs. Eva Harris, University of California, Berkeley, USA, Kevin Arien, Institute of Tropical Medicine, Belgium, Leticia Franco and Jairo Mendez, PAHO-WHO, for their collaboration in the development of this study and Linda Lloyd, USA, for review of the manuscript. The authors also thank the PAHO-WHO Representation in Cuba, National Center of Medical Genetics and the Network of Cuban molecular diagnostic laboratories for their support in the collection of the samples.

References

- [1] World Health Organization. (2021). COVID-19 weekly epidemiological update, edition 49, 20 July 2021. World Health Organization. <https://apps.who.int/iris/handle/10665/343120>.
- [2] Salud. Organización Panamericana de la Actualización epidemiológica: Enfermedad por Coronavirus (COVID-19). 30 de octubre de 2021, OPS/OMS, Washington, D.C, 2021 <http://www.paho.org>.
- [3] Organización Panamericana de la Salud. Cuba, a un año de la pandemia. Boletín de la OPS-OMS en Cuba. Hitos y alcance de la respuesta cubana a la COVID-19. Entrevista con el Ministro de Salud Pública, Dr. José Angel Portal. 25 (1) (2021), 4-11.
- [4] Y. Almeida, S. Reyes, EA. Guerra, I. Alonso, Evolución por días de la Tasa de Incidencia (por 100 mil), Covid19 CubaData (2021) <https://covid19cubadata.github.io/#cuba>.
- [5] S. Borroto, Situación de la COVID-19 en Cuba a un año de la pandemia, Bol. de la OPS-OMS en Cuba 25 (1) (2021) 12–18.
- [6] S. Noda, Variantes genéticas aumentan la severidad de la COVID-19, Portal Web del Minist. de Salud Publ. Repub. de Cuba (2021). <https://salud.msp.gob.cu/>.
- [7] Pathogen Discovery Team, Protocols for SARS-CoV-2 sequencing, Centers for Diseases, Control and Prevention (2020).
- [8] S. Elbe and G. Buckland-Merrett, Data, disease and diplomacy: GISAID's innovative contribution to global health. Global Challenges, 1 (2017) 33-46. doi:10.1002/gch2.1018. PMID: 31565258
- [9] X. Deng, MA. Garcia-Knight, MM. Khalid, V. Servellita, C. Wang, MK. Morris, et al., Transmission, infectivity, and antibody neutralization of an emerging SARS-CoV-1 2 variant in California carrying a L452R spike protein mutation, medRxiv (2021), doi:10.1101/2021.03.07.21252647.
- [10] F. González-Candelas, M.A. Shaw, T. Phan, U. Kulkarni-Kale, D. Paraskevis, F. Luciani, et al., One year into the pandemic: short-term evolution of SARS-CoV-2 and emergence of new lineages, Infect. Genet. Evol. 92 (2021).
- [11] R. Alonso, O. Figueredo, L. Fariñas, K. Rodríguez, A. Terrero, R. Carmenate, Ante nuevas variantes de SARS CoV-2, cual es la situación en Cuba y la potencial efectividad de las vacunas, cubanas,CubaDebate (2021) <https://www.cubadebate.cu>.
- [12] S. Shayan, A. Nazanin, F. Michael, Venketaraman vishwanath analysis of the delta variant B.1.617.2 COVID-19, Clin. Pract. 11 (4) (2021) 778–784.
- [13] J. Portal Miranda, Intervención sanitaria con candidatos vacunales, como estrategia temporal de enfrentamiento a la COVID-19, Rev. Cubana Salud Públ. 48 (1) (2021) e3513.
- [14] A. Rambaut, N. Loman, O. Pybus, W. Barclay, J. Barret, A. Carabelli, T. Connor, et al., Preliminary genomic characterization of an emergent SARS-CoV-2 lineage in UK defined by a novel set of spike mutations, Retrieved February 1 (2020).
- [15] H. Tegally, E. Wilkinson, M. Giovanetti, A. Iranzadeh, V. Fonseca, J. Giandhari, et al., Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa, medRxiv (2020).
- [16] S. Alizon, M.T. Sofonea, SARS-CoV-2 virulence evolution: avirulence theory, immunity and trade-offs, J. Evol. Biol. (2021).

- [17] B. Sierra, P. Triska, P. Soares, G. Garcia, A.B. Perez, E. Aguirre, et al., OSBPL10, RXRA and lipid metabolism confer African-ancestry protection against dengue haemorrhagic fever in admixed Cubans, *PLoS Pathog.* 13 (2) (2017).
- [18] J. Wise, et al., Covid-19: UK cases of variant from India rise by 160% in a week, *BMJ* 373 (2021) 1315, doi:10.1136/bmj.n1315.
- [19] V. Tchesnokova, H. Kulasekara, L. Larson, V. Bowers, E. Rechkina, D. Kisiela, et al., Acquisition of the L452R mutation in the ACE2-binding interface of spike protein triggers recent massive expansion of SARS-CoV-2 variants, *J. Clin. Microbiol.* 59 (11) (2021).
- [20] L.J. Abu-Raddad, H. Chemaitelly, H.H. Ayoub, H.M. Yassine, F.M. Benslimane, H.A. Al Khatib, et al., Severity, criticality, and fatality of the SARS-CoV-2 Beta variant, *medRxiv* (2021).