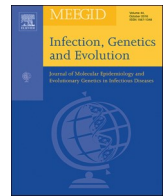




Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

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.



## Research paper

## Substitutions in Spike and Nucleocapsid proteins of SARS-CoV-2 circulating in South America



Carlos Franco-Muñoz<sup>a,b,\*</sup>, Diego A. Álvarez-Díaz<sup>a</sup>, Katherine Laiton-Donato<sup>a</sup>, Magdalena Wiesner<sup>c</sup>, Patricia Escandón<sup>c</sup>, José A. Usme-Ciro<sup>a,d</sup>, Nicolás D. Franco-Sierra<sup>e</sup>, Astrid C. Flórez-Sánchez<sup>f</sup>, Sergio Gómez-Rangel<sup>h</sup>, Luz D. Rodríguez-Calderon<sup>h</sup>, Juliana Barbosa-Ramirez<sup>h</sup>, Erika Ospitia-Baez<sup>h</sup>, Diana M. Walteros<sup>g</sup>, Martha L. Ospina-Martinez<sup>i</sup>, Marcela Mercado-Reyes<sup>a,j</sup>

<sup>a</sup> Unidad de Secuenciación y Genómica, Grupo de Investigación Básica y Aplicada en Enfermedades Emergentes, Dirección de Investigación en Salud Pública, Instituto Nacional de Salud, Bogotá 111321, Colombia

<sup>b</sup> Grupo de Parasitología, Dirección de Investigación en Salud Pública, Instituto Nacional de Salud, Bogotá 111321, Colombia

<sup>c</sup> Grupo de Microbiología, Dirección de Investigación en Salud Pública, Instituto Nacional de Salud, Bogotá, Colombia

<sup>d</sup> Centro de Investigación en Salud para el Trópico-CIST, Universidad Cooperativa de Colombia, Santa Marta, 470003, Colombia

<sup>e</sup> Programa Ciencias de la Biodiversidad, Instituto de Investigación de Recursos Biológicos Alexander von Humboldt, Bogotá 111311, Colombia

<sup>f</sup> Dirección de Redes en Salud Pública, Instituto Nacional de Salud, Bogotá 111321, Colombia

<sup>g</sup> Dirección de Vigilancia en Salud Pública, Instituto Nacional de Salud, Bogotá 111321, Colombia

<sup>h</sup> Grupo de Virología, Dirección de Redes en Salud Pública, Instituto Nacional de Salud, Bogotá 111321, Colombia

<sup>i</sup> Dirección General, Instituto Nacional de Salud, Bogotá 111321, Colombia

<sup>j</sup> Dirección de Investigación en Salud Pública, Instituto Nacional de Salud, Bogotá 111321, Colombia

## ARTICLE INFO

## Keywords:

SARS-CoV-2

Spike

Nucleocapsid

South America

Non-synonymous substitutions

## ABSTRACT

SARS-CoV-2 is a new member of the genus *Betacoronavirus*, responsible for the COVID-19 pandemic. The virus crossed the species barrier and established in the human population taking advantage of the spike protein high affinity for the ACE receptor to infect the lower respiratory tract. The Nucleocapsid (N) and Spike (S) are highly immunogenic structural proteins and most commercial COVID-19 diagnostic assays target these proteins. In an unpredictable epidemic, it is essential to know about their genetic variability. The objective of this study was to describe the substitution frequency of the S and N proteins of SARS-CoV-2 in South America. A total of 504 amino acid and nucleotide sequences of the S and N proteins of SARS-CoV-2 from seven South American countries (Argentina, Brazil, Chile, Ecuador, Peru, Uruguay, and Colombia), reported as of June 3, and corresponding to samples collected between March and April 2020, were compared through substitution matrices using the Muscle algorithm. Forty-three sequences from 13 Colombian departments were obtained in this study using the Oxford Nanopore and Illumina MiSeq technologies, following the amplicon-based ARTIC network protocol. The substitutions D614G in S and R203K/G204R in N were the most frequent in South America, observed in 83% and 34% of the sequences respectively. Strikingly, genomes with the conserved position D614 were almost completely replaced by genomes with the G614 substitution between March to April 2020. A similar replacement pattern was observed with R203K/G204R although more marked in Chile, Argentina and Brazil, suggesting similar introduction history and/or control strategies of SARS-CoV-2 in these countries.

It is necessary to continue with the genomic surveillance of S and N proteins during the SARS-CoV-2 pandemic as this information can be useful for developing vaccines, therapeutics and diagnostic tests.

## 1. Introduction

The recently emerged SARS-CoV-2 responsible for the coronavirus

disease 2019 (COVID-19) pandemic, has increased significantly in the number of cases and deaths, about 70,000 new cases are reported globally.

\* Corresponding author at: Unidad de Secuenciación y Análisis Genómico, Grupo de Investigación Básica y Aplicada en Enfermedades Emergentes, Dirección de Investigación en Salud Pública, Instituto Nacional de Salud, Bogotá 111321, Colombia.

E-mail address: [cefrancom@unal.edu.co](mailto:cefrancom@unal.edu.co) (C. Franco-Muñoz).

<https://doi.org/10.1016/j.meegid.2020.104557>

Received 15 June 2020; Received in revised form 31 August 2020; Accepted 11 September 2020

Available online 17 September 2020

1567-1348/ © 2020 Elsevier B.V. All rights reserved.

(WHO, 2020a, 2020c). The first case of COVID-19 in South America was reported in Brazil on February 26, in a 61 years old man traveling from Italy (gob.br, 2020). In Colombia, the first case of COVID-19 was announced on March 6, in a traveler from Italy, after which the number of patients has exceeded 43,000 and over 1500 deaths (INS, 2020).

The SARS-CoV-2 genome consist of a single, positive-stranded RNA (ssRNA[+]), with 29,903 nucleotides long. The virus has shown to be highly infectious and easily transmitted among human populations (He et al., 2020), even infecting other vertebrate species under laboratory conditions (Shi et al., 2020). The SARS-CoV-2 genome has nine open reading frames (ORFs); the first one, subdivided in ORF1a and ORF1b by ribosomal frameshifting, encodes the polyproteins pp1a and pp1ab which are processed into non-structural proteins involved in subgenomic/genome length RNA synthesis and virus replication. Structural proteins, Spike (S), Envelope (E), Membrane (M), and Nucleocapsid (N) are encoded in subgenomic mRNA transcripts within ORFs 2, 4, 5, and 9, respectively (SIB, 2020; Yount et al., 2005).

Spike protein, a type I membrane glycoprotein, is the most exposed viral protein recognized by the cellular receptor angiotensin-2-converting enzyme (ACE2) during the infection of the lower respiratory tract and considered the main inducer of neutralizing antibodies. The N protein is associated with the RNA genome to form the ribonucleocapsid and is abundantly expressed during infection. Both N and S proteins are highly immunogenic and most commercial COVID-19 diagnostic tests (molecular and immunologic) target these proteins (Álvarez-Díaz et al., 2020; Lee et al., 2020).

Furthermore, non-synonymous mutations in the S and N proteins have been reported, their implications in the potential emergence of antigenically distinct and/or more virulent strains remain to be studied, although it was reported that mutations in the receptor-binding domain (RBD) at the S protein of SARS-CoV related viruses disrupt the antigenic structure and binding activity of RBD to ACE2 (Du et al., 2009). Similarly, non-synonymous mutations could impact the antibody response and the specificity and sensitivity of serological tests for COVID-19 diagnosis is unknown. Thus, identifying variable sites at these proteins can provide a valuable resource for choosing the target antigens for the development of SARS-CoV-2 vaccines, therapeutics, and diagnostic tests (Du et al., 2009; Jacofsky et al., 2020). The objective of this study is to describe the frequency of substitutions in S and N proteins of SARS-CoV-2 in South America.

## 2. Materials and methods

### 2.1. Ethics

This work was developed according to the national law 9/1979, decrees 786/1990 and 2323/2006, which establishes that the Instituto Nacional de Salud (INS) from Colombia is the reference lab and health authority of the national network of laboratories and in cases of public health emergency or those in which scientific research for public health purposes as required. The INS is authorized to use the biological material for research purposes, without informed consent, which includes the anonymous disclosure of results. This study was performed following the ethical standards of the Declaration of Helsinki 1964 and its later amendments. The information used for this study comes from secondary sources of data that were previously anonymized and ensure protect patient data.

### 2.2. Patients and samples

Nasopharyngeal swab samples from patients with suspected SARS-CoV-2 infection were processed for RNA extraction using the automated MagNA Pure LC nucleic acid extraction system (Roche Diagnostics GmbH, Mannheim, Germany) and viral RNA detection was performed by real-time RT-PCR using the SuperScript III Platinum One-Step Quantitative RT-qPCR kit (Thermo Fisher Scientific, Waltham, MA, USA), following the Charité-Berlin protocol (Corman et al., 2020) for the amplification of the SARS-CoV-2 E (betacoronavirus screening assay) and RdRp (SARS-CoV-2 confirmatory assay) genes.

### 2.3. Complete genome sequencing of SARS-CoV-2 through NGS

NGS of SARS-CoV-2 from 43 patients was performed using the amplicon-based Illumina and Nanopore sequencing approaches, ARTIC network protocol (Quick, 2020). Following cDNA synthesis with SuperScript IV reverse transcriptase (Thermo Fisher Scientific, Waltham, MA, USA) and random hexamers (Thermo Fisher Scientific, Waltham, MA, USA), a set of 400-bp tiling amplicons across the whole genome of SARS-CoV-2 were generated using the primer schemes nCoV-2019/V3 (Quick, 2020).

### 2.4. Sequence analysis

Reads were mapped to the Wuhan-Hu-1 reference genome (NC\_045512.2) using BWA (Li et al., 2020) and BBmap (brian-jgi, 2020); then, assembled consensus sequences were submitted to GISAID. Substitution matrices of nucleotides and amino acids of S and N proteins were generated from a multiple sequence alignment with the reference genome against the 43 assembled Colombian SARS-CoV-2 genomes (Table 1) using the Muscle algorithm (Edgar, 2004) in MEGA X (Kumar et al., 2016). Subsequently, 461 SARS-CoV-2 sequences from South American countries, including Argentina, Brazil, Ecuador, Peru, Uruguay and other sequences from Colombia available on the GISAID, NCBI, and GSA databases were analyzed (Supplementary Table S1, and Supplementary Table S2).

## 3. Results

### 3.1. Non-synonymous substitutions in the Spike and Nucleocapsid proteins in Colombia

Several non-synonymous substitutions were observed in the S and N proteins of the Colombian SARS-CoV-2 sequences generated in this study. Three amino acid substitutions were observed in the S protein, D614G was present in 81% (35/43) of the sequences. Furthermore, substitutions G181V and D936Y were found at low frequencies of 2.3% (1/43) and 2.3% (1/43) respectively (Table 1). In the N protein, five amino acid substitutions were found; the most frequent being R203K and G204R in 13.95% (6/43) of the sequences. Amino acid substitutions, R191C, R209I and G238C were found in 4.65% (2/43), 4.65% (2/43) and 6.97% (3/43) of the Colombian sequences, respectively (Table 1). Some nucleotide substitutions were synonymous.

**Table 1**

Nucleotide substitutions and amino acid changes in the Spike and Nucleocapsid proteins of 43 SARS-CoV-2 genomes from Colombia generated in this study. Sequences from other countries of South America were retrieved from GISAID database. Multiple alignment was made using the Muscle algorithm in MEGA X software. Location of the substitutions was estimated based on the reference genome NC\_045512.2. Left panel: nucleotides; right panel: amino acids. The uncovered nucleotide and amino acids during complete genome sequencing are represented by an interrogate mark (?). Nucleotide and amino acid positions are indicated in the number under each gene.

GISAID Accession ID	Sequence ID	Nucleotides												Aminoacids								
		S						N						S			N					
		Spike						Nucleocapsid						Spike			Nucleocapsid					
		2	2	2	2	2	2	2	2	2	2	2	2	0	0	0	0	0	0	0	0	0
	NC_045512_SARS-COV-2_Wuhan-Hu-1	G	A	T	C	G	C	G	G	G	G	G	C	G	D	D	R	R	G	R	G	G
EPI_ISL_456146	hCoV-19/Colombia/Bello/82457/2020	G	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.
EPI_ISL_456118	hCoV-19/Colombia/Medellin/80370/2020	G	.	.	.	.	?	.	.	.	.	.	.	G	.	.	?	.	.	.	.	.
EPI_ISL_456145	hCoV-19/Colombia/Barranquilla/79743/2020	G	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.
EPI_ISL_456120	hCoV-19/Colombia/Bogota/79190/2020	G	.	.	.	.	.	.	.	.	.	.	T	G	.	.	.	.	.	.	.	.
EPI_ISL_456116	hCoV-19/Colombia/Bogota/84861/2020	G	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.
EPI_ISL_456126	hCoV-19/Colombia/Norte_de_Santander/80656/2020	G	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.
EPI_ISL_456155	hCoV-19/Colombia/Pereira/80245/2020	G	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.
EPI_ISL_456130	hCoV-19/Colombia/Cucuta/79504/2020	G	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.
EPI_ISL_456131	hCoV-19/Colombia/Bucaramanga/79752/2020	G	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.
EPI_ISL_456132	hCoV-19/Colombia/Ibague/80202/2020	G	C	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.
EPI_ISL_456134	hCoV-19/Colombia/Cali/79773/2020	G	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.
EPI_ISL_456135	hCoV-19/Colombia/Cali/79779/2020	G	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.
EPI_ISL_456138	hCoV-19/Colombia/Cali/79943/2020	G	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.
EPI_ISL_456150	hCoV-19/Colombia/Cali/79965/2020	G	.	.	T	.	.	.	.	.	.	.	.	G	Y	.	.	.	.	.	.	.
EPI_ISL_456141	hCoV-19/Colombia/Palmira/81271/2020	G	.	?	?	.	.	.	.	.	.	.	.	G	?	.	.	.	.	.	.	.
EPI_ISL_456136	hCoV-19/Colombia/Cali/79929/2020	G	.	.	.	T	.	.	.	.	.	.	.	G	C	.	.	.	.	.	.	.
EPI_ISL_456137	hCoV-19/Colombia/Cali/79930/2020	G	.	.	.	T	.	.	.	.	.	.	.	G	C	.	.	.	.	.	.	.
EPI_ISL_418262	hCoV-19/Colombia/Bogota/78390/2020	G	.	.	.	.	A	A	C	.	.	.	.	G	.	.	K	R	.	.	.	.
EPI_ISL_456149	hCoV-19/Colombia/Bucaramanga/80123/2020	G	.	.	.	.	A	A	C	.	.	.	.	G	.	.	K	R	.	.	.	.
EPI_ISL_456154	hCoV-19/Colombia/Tumaco/103303/2020	T	G	.	.	.	A	A	C	.	.	.	.	V	G	.	K	R	.	.	.	.
EPI_ISL_456156	hCoV-19/Colombia/Barrancabermeja/94514/2020	G	.	.	.	.	A	A	C	.	.	.	.	G	.	.	K	R	.	.	.	.
EPI_ISL_456139	hCoV-19/Colombia/Cali/81251/2020	G	.	.	.	.	A	A	C	.	.	.	.	G	.	.	K	R	.	.	.	.
EPI_ISL_456142	hCoV-19/Colombia/Cali/81279/2020	G	.	.	?	.	A	A	C	.	.	.	.	G	?	.	K	R	.	.	.	.
EPI_ISL_456144	hCoV-19/Colombia/Leticia/103039/2020	G	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.
EPI_ISL_456121	hCoV-19/Colombia/Cartagena/79738/2020	G	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.
EPI_ISL_456148	hCoV-19/Colombia/Cartagena/79890/2020	G	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.
EPI_ISL_456124	hCoV-19/Colombia/Popayan/80705/2020	G	.	.	?	.	.	.	.	.	.	.	.	G	?	.	.	.	.	.	.	.
EPI_ISL_456153	hCoV-19/Colombia/Cienaga/101935/2020	G	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.
EPI_ISL_456125	hCoV-19/Colombia/Santa_Marta/81035/2020	G	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.
EPI_ISL_456147	hCoV-19/Colombia/Bogota/96183/2020	G	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.
EPI_ISL_456122	hCoV-19/Colombia/Cartagena/80191/2020	G	.	.	.	.	.	.	.	.	.	?	.	G	.	.	.	.	.	.	.	?
EPI_ISL_456129	hCoV-19/Colombia/Pereira/82099/2020	G	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.
EPI_ISL_456140	hCoV-19/Colombia/Cartago/81264/2020	G	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.
EPI_ISL_456143	hCoV-19/Colombia/Cali/82307/2020	G	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.
EPI_ISL_456151	hCoV-19/Colombia/Cali/82320/2020	G	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.
EPI_ISL_456117	hCoV-19/Colombia/Medellin/79253/2020	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	C
EPI_ISL_417924	hCoV-19/Colombia/Antioquia/79256/2020	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	C
EPI_ISL_456119	hCoV-19/Colombia/Medellin/80825/2020	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	C
EPI_ISL_456123	hCoV-19/Colombia/Manizales/79489/2020	.	.	.	.	.	?	.	.	.	.	.	?	.	.	.	?	.	.	.	.	?
EPI_ISL_456133	hCoV-19/Colombia/Cali/79449/2020	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
EPI_ISL_456152	hCoV-19/Colombia/Cali/96176/2020	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
EPI_ISL_456128	hCoV-19/Colombia/Armenia/80663/2020	.	.	T	.	.	.	.	T	.	.	.	.	.	.	.	.	.	I	.	.	.
EPI_ISL_456127	hCoV-19/Colombia/Armenia/79911/2020	.	.	T	.	.	.	.	T	.	.	.	.	.	.	.	.	.	I	.	.	.

### 3.2. Non-synonymous substitutions in the Spike and Nucleocapsid proteins in South America

Genomic resource databases, NCBI, GISAID and GSA were consulted to determine the substitutions in S and N proteins of SARS-CoV-2 from South America. A total of 504 genomes reported as of June 3Th 2020, were analyzed, 126 from Colombia (including the 43 genomes reported in this study), 29 from Argentina, 145 from Brazil, 153 from Chile, 4 from Ecuador, 2 from Peru and 45 from Uruguay. Fifty sequences of S and 27 of N were excluded from the analysis because the presence of undetermined bases that did not allow the proper identification of the S and N ORFs in the amino acid substitution matrices.

Twenty-eight and twenty-two non-synonymous substitutions were identified in the sequence of S and N proteins respectively, in genomes of South America (Table S1 and S2). The most frequent in S were D614G (83%) V1176F (2.2%) and P1263L (1.5%), while the most frequent in N were R203K (34.5%), G204R (34.3%), I292T (15.8%) and S197L (3.3%). The remaining substitutions in both, S and N occurred in less than 1% of the sequences. These included G181V and D936Y in S, and R191C and G238C in N, as observed in the Colombian genomes (Fig. 1).

### 3.3. Spatiotemporal distribution of substitutions in Spike and Nucleocapsid

The analysis of substitution frequencies by country shows that D614G substitution in the S protein was frequent in Argentina, Brazil, Chile, Colombia and Peru, with 80–100% of the reported sequences (Fig. 2A). In Ecuador and Uruguay D614 position was predominant by March, however by April the G614 substitution reached 80% in Uruguay. In general, the percentage of genomes in South America with this substitution augmented nearly to 100% from March to April (Fig. 2B).

Non-synonymous substitutions R203K and G204R, which are the hallmarks of the B.1.1 lineage, were the most frequent in the N protein of South American sequences. Both substitutions were frequent in Argentina and Brazil with 55% and 74% of the reported sequences respectively (Fig. 3A). In Ecuador and Chile the occurrence of these substitutions was about 20%, while in Uruguay the frequency was similar to Colombia. Furthermore, the proportion of genomes with this double substitution augmented in Chile, Argentina and Brazil from March to April. In contrast, this proportion increased slightly in Colombia and Uruguay, and remained below 20% (Fig. 3B).

The substitution I292T in the N protein was rare in Argentina (10.7%), Chile (4.6%) and Uruguay (2.2%); and absent in Colombia, Peru and Ecuador. In contrast, this substitution was very frequent in Brazil (56.3%) (Fig. 3C). The spatiotemporal distribution pattern of this substitution was similar to that of R203K and G204R, increasing from March to April in Chile, Argentina and Brazil in contrast to Colombia and Uruguay where this substitution was almost absent in genomes registered on April (Fig. 3D).

## 4. Discussion

The first COVID-19 case in Colombia was confirmed on March 6, 2020, from a traveler who entered the country from Italy on February 26, 2020 (EPI\_ISL\_418262). By June 11, 2020, a total 43,810 confirmed

cases and 1505 deaths have been reported (INS, 2020). This study evidenced the presence of the D614G substitution in the S protein in 89.6% (112/125) of Colombian SARS-CoV-2 sequences, by April 27, 2020, while the first introduced cases presented the conserved position reported in the Wuhan-Hu-1 reference genome (Bhattacharyya et al., 2020).

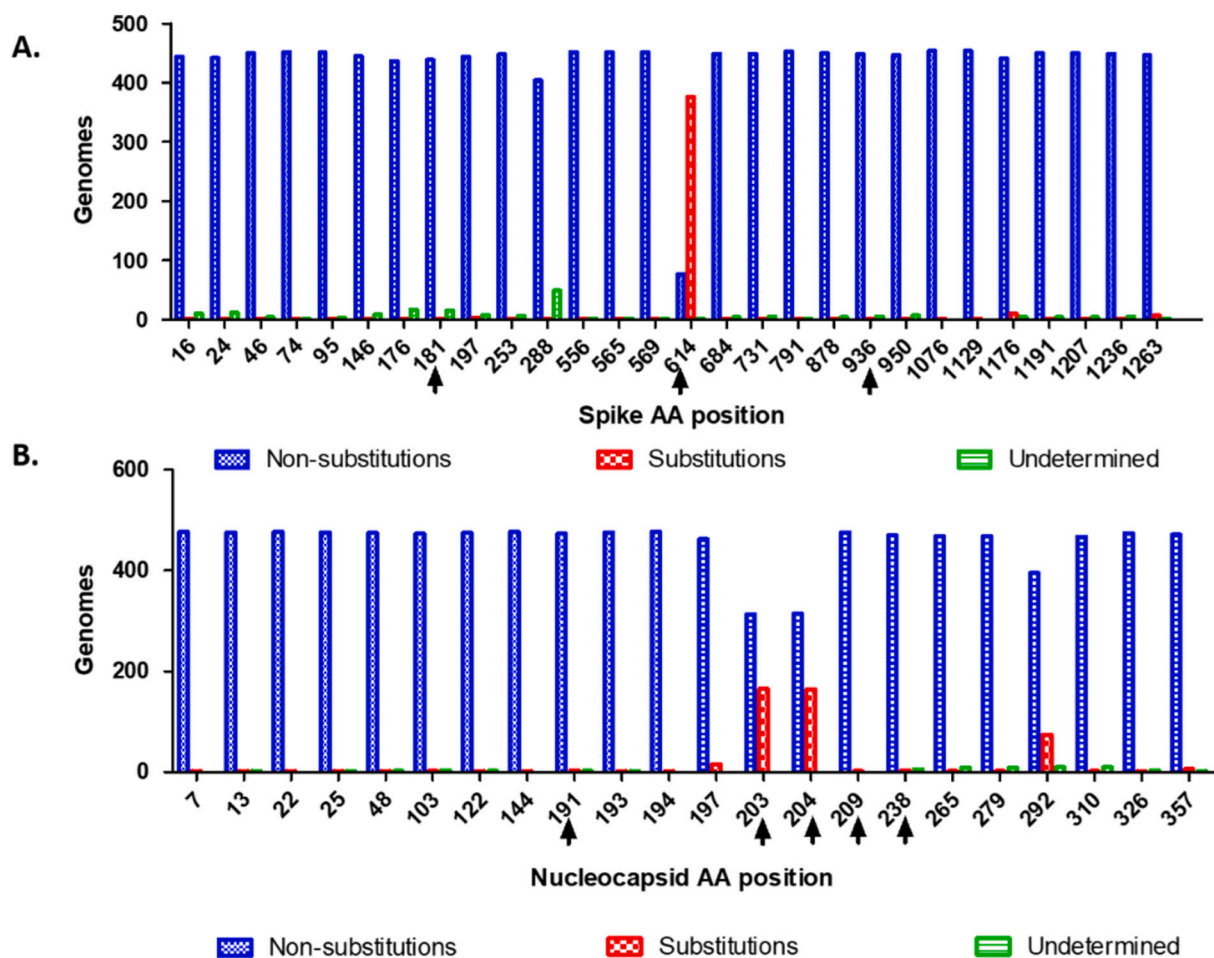
D614G was detected in 85% of sequences being present in most of the South American countries with available genomic information. Several studies have suggested a potential role of the D614G substitution in increase the virus infectivity (Becerra-Flores and Cardozo, 2020; Korber et al., 2020; Nakashima, 2020), transmissibility (Bhattacharyya et al., 2020; Brufsky, 2020), mortality rate and immune system evasion (Kim et al., 2020); However, the information regarding D614G substitution is still inconclusive and it there is not unquestionable evidence of the relation of this mutation with SARS-CoV-2 infectivity and transmission, also it was not possible to rule out the association with founder effects.

On the other hand, the co-occurrence of R203K and G204R substitutions in the N protein, was identified in 34% of South American sequences. The B.1.1 lineage is defined by the three-nucleotide mutation in 2 adjacent codons leading to the two consecutive amino acid changes in the N protein (Bartolini et al., 2020; Rambaut et al., 2020), while most of amino acid changes evidenced in the S and N proteins cannot be directly related to a specific lineage (Bhattacharyya et al., 2020; Korber et al., 2020).

This lineage has been reported in samples from travelers with connection to Italy (Gupta and Mandal, 2020), also observed in the first confirmed case of SARS-CoV-2 in Colombia (EPI\_ISL\_418262) and another patient with travel connection to Spain (EPI\_ISL\_456149) (Table 1). Furthermore, multiple countries outside Italy have reported this lineage among their samples including, Belgium, Switzerland, Vietnam, India, Nigeria and Mexico, demonstrating a wide distribution worldwide (Gupta and Mandal, 2020).

RNA viruses are known to possess high substitution rates compared to DNA viruses, leading to high genetic variability and the rapid action of evolutionary mechanisms of natural selection and genetic drift (Li et al., 2020; Tang et al., 2020). However, SARS-CoV-2 and others coronaviruses have proteins with exonuclease activity, as nsp14, with error correcting capacity (Romano et al., 2020; Subissi et al., 2014). Despite some evolutionary changes may be in fact adaptive, it is important to be careful with conclusions in the absence of an experimental model to evaluate the impact of every mutation in the virus phenotype and virus-host interaction (Villabona-Arenas et al., 2020).

The S and N proteins are the most widely used for serological assays, there are 138 FDA-approved serological tests of which only 24% specifically report the screened antigen of these, 39% use the S and 42% use N, and 18% use both (Supplementary Table S3). Recombinant proteins or synthetic peptides of SARS-CoV-2 are generally explored as alternatives to be used in serological tests and therapeutics against SARS-CoV-2 and related Betacoronaviruses (Du et al., 2009; Jacofsky et al., 2020), considering that S and N proteins are the major immunogenic proteins of SARS and MERS coronavirus and the first choice for producing recombinant antigens (Yan et al., 2020).



**Fig. 1.** Non-synonymous substitutions in the Spike and Nucleocapsid proteins of SARS-CoV-2 in South America. A. Non-synonymous substitution sites in the S protein. Blue bars represent the genomes with conserved positions, red bars represent the genomes with substitution at the indicated site (at least one genome per site). B. Non-synonymous substitutions in the N. Blue bars represent the genome with conserved positions, red bars represent the genomes with substitution at the indicated site (at least one genome). Green bars represent the number of genomes with undetermined amino acids at the indicated site. Black arrows show substitutions found in the 43 Colombian genomes generated in this study. The location of the substitutions was estimated based on the reference genome NC\_045512.2. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

## 5. Conclusion

Amino acid changes were found in the S and N proteins of SARS-CoV-2 circulating in South America, the most frequent being D614G in S and R203K-G204R and I292T in N. It is necessary to continue with genomic surveillance of changes in these proteins during the SARS-CoV-2 pandemic, even more considering that these proteins are the most commonly used in serological and molecular tests.

The identification of nucleotide substitutions, amino acid changes and their frequencies in circulating viruses, can be useful for public health decision-making, including vaccine design efforts, design of SARS-CoV-2 diagnostic tests, and therapeutic compounds.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.meegid.2020.104557>.

## Funding

This study was funded by the National Institute of Health, Bogota, Colombia.

## Declaration of Competing Interest

The authors declare no competing interest.

## Acknowledgements

The authors thank the National Laboratory Network for routine virologic surveillance of SARS-CoV-2 in Colombia. The authors thank all the Colombian and foreign researchers who deposited genomes in GISAID's EpiFlu (TM) Database contributing to genomic diversity and phylogenetic relationship of SARS-CoV-2.

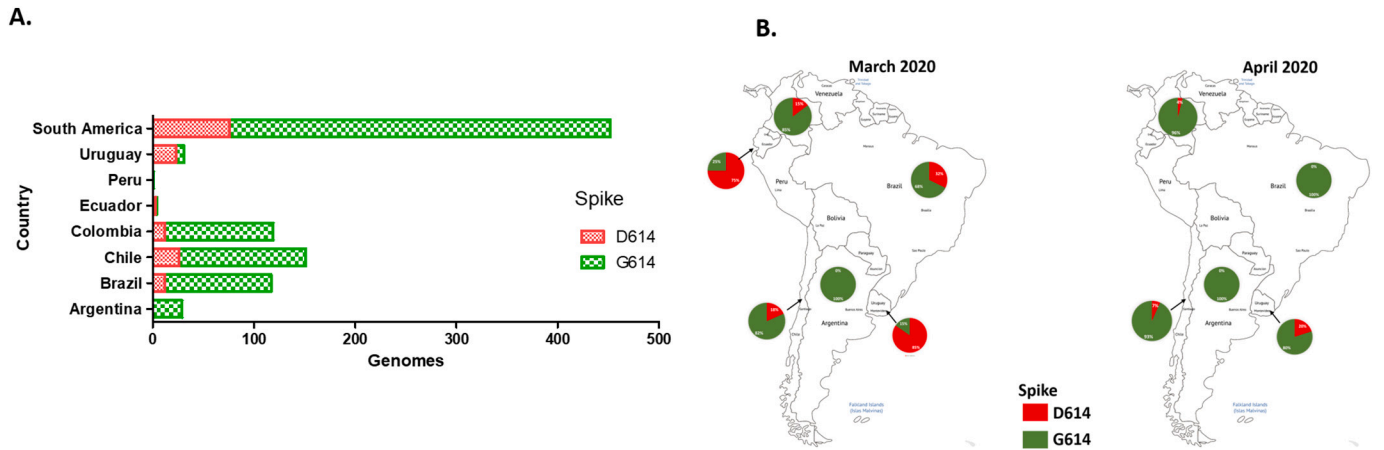


Fig. 2. Spatiotemporal distribution of the D614G substitution in the Spike protein of SARS-CoV-2 in South America. A. Distribution of D614/G614 substitutions in SARS-CoV-2 genomes reported as of Jun 3th 2020 in South American countries B. Spatiotemporal distribution of D614/G614 substitutions in SARS-CoV-2 genomes sequenced between March and April, 2020 in South America, according with the sample collection date.

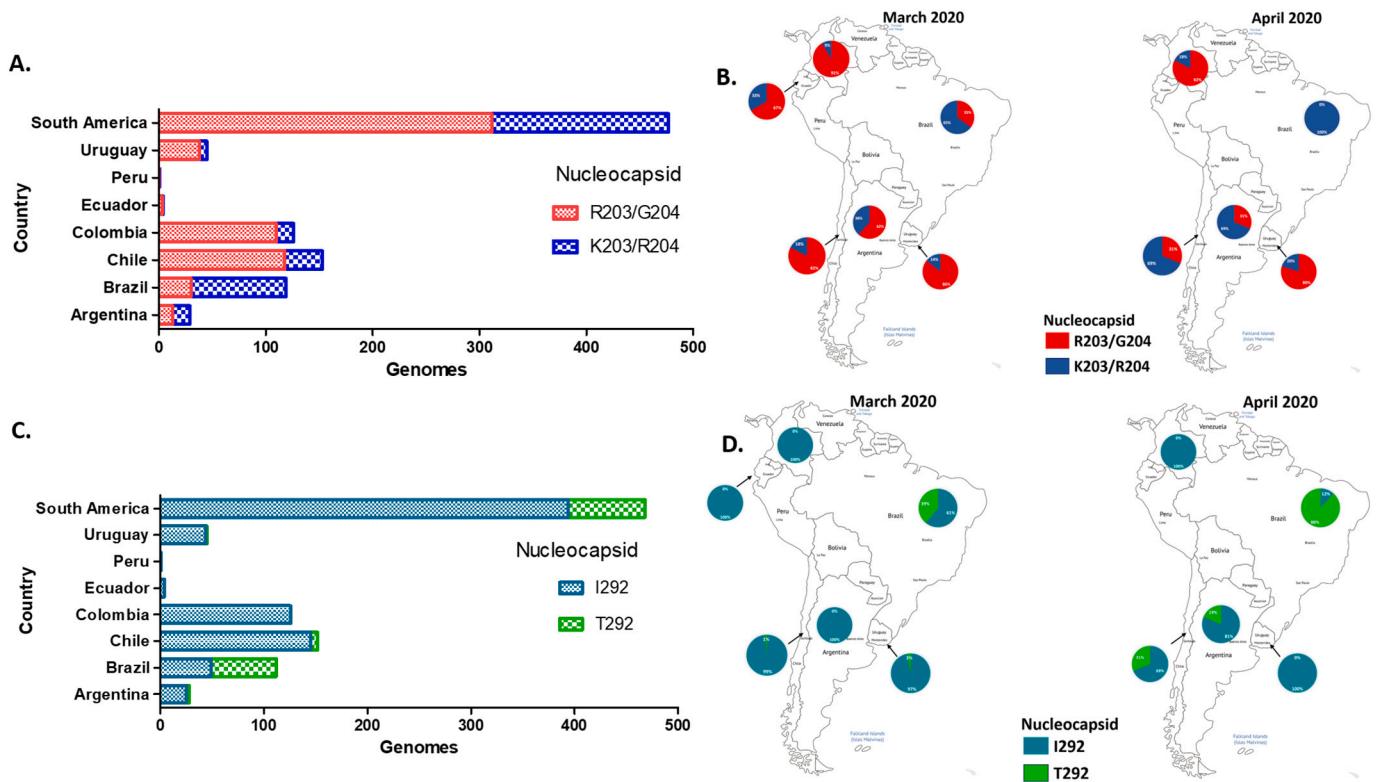


Fig. 3. Spatiotemporal distribution of the most frequent substitutions in the Nucleocapsid protein of SARS-CoV-2 in South America. A. Distribution of R203/G204 and K203/R204 substitutions in SARS-CoV-2 genomes reported as of Jun 3th, 2020 in South American countries. B. Spatiotemporal of R203/G204 and K203/R204 substitutions in SARS-CoV-2 genomes sequenced between March and April, 2020 in South America, according with the sample collection date. C. Distribution of I292/T292 substitutions of nucleocapsid protein in SARS-CoV-2 genomes reported as of Jun 3th, 2020 in South American countries. D. Distribution of I292 and T292 substitutions in SARS-CoV-2 genomes sequenced between March and April, 2020 in South America, according with the sample collection date.

## References

- Álvarez-Díaz, D.A., Franco-Muñoz, C., Laiton-Donato, K., Usme-Ciro, J.A., Franco-Sierra, N.D., Flórez-Sánchez, A.C., Gómez-Rangel, S., Rodríguez-Calderon, L.D., Barbosa-Ramirez, J., Ospitia-Baez, E., Walteros, D.M., Ospina-Martínez, M.L., Mercado-Reyes, M., 2020. Molecular analysis of several in-house rRT-PCR protocols for SARS-CoV-2 detection in the context of genetic variability of the virus in Colombia. *Infect. Genet. Evol.* 84, 104390.
- Bartolini, B., Rueca, M., Gruber, C.E.M., Messina, F., Carletti, F., Giombini, E., Lalle, E., Bordini, L., Matusali, G., Colavita, F., Castilletti, C., Vairo, F., Ippolito, G., Capobianchi, M.R., Di Caro, A., 2020. SARS-CoV-2 phylogenetic analysis, Lazio region, Italy, February–March 2020. *Emerg. Infect. Dis.* 26.
- Becerra-Flores, M., Cardozo, T., 2020. SARS-CoV-2 viral spike G614 mutation exhibits higher case fatality rate. *Int. J. Clin. Pract.* 74, e13525.
- Bhattacharyya, C., Das, C., Ghosh, A., Singh, A.K., Mukherjee, S., Majumder, P.P., Basu, A., Biswas, N.K., 2020. Global Spread of SARS-CoV-2 Subtype with Spike Protein Mutation D614G is Shaped by Human Genomic Variations that Regulate Expression of TMPRSS2 and MX1 Genes. *bioRxiv*.
- brian-jgi, 2020. BMAP Short Read Aligner, and Other Bioinformatic Tools.
- Brufsky, A., 2020. Distinct viral clades of SARS-CoV-2: implications for modeling of viral spread. *J. Med. Virol.* 92, 1386–1390.
- Corman, V.M., Landt, O., Kaiser, M., Molenkamp, R., Meijer, A., Chu, D.K., Bleicker, T., Brünink, S., Schneider, J., Schmidt, M.L., Mulders, D.G., Haagmans, B.L., van der Veer, B., van den Brink, S., Wijsman, L., Goderski, G., Romette, J.-L., Ellis, J., Zambon, M., Peiris, M., Goossens, H., Reusken, C., Koopmans, M.P., Drosten, C., 2020. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro Surv.* 25, 2000045.
- Du, L., He, Y., Zhou, Y., Liu, S., Zheng, B.-J., Jiang, S., 2009. The spike protein of SARS-CoV—a target for vaccine and therapeutic development. *Nat. Rev. Microbiol.* 7, 226–236.
- Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797.
- gob.br, 2020. Coronavirus COVID-19.
- Gupta, A.M., Mandal, S., 2020. Loss of Epitopes from SARS-Cov-2 Proteins for Non-synonymous Mutations: A Potential Global Threat. *OSF Preprints*.
- He, X., Lau, E.H., Wu, P., Deng, X., Wang, J., Hao, X., Lau, Y.C., Wong, J.Y., Guan, Y., Tan, X., 2020. Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat. Med.* 26, 672–675.
- INS, 2020. Coronavirus (COVID - 2019) en Colombia. Instituto Nacional de Salud.
- Jacofsky, D., Jacofsky, E.M., Jacofsky, M., 2020. Understanding antibody testing for COVID-19. *J. Arthroplast.* 35, 574–581.
- Kim, S.-J., Nguyen, V.-G., Park, Y.-H., Park, B.-K., Chung, H.-C., 2020. A novel synonymous mutation of SARS-CoV-2: is this possible to affect their antigenicity and immunogenicity? *Vaccines* 8, 220.
- Korber, B., Fischer, W., Gnanakaran, S.G., Yoon, H., Theiler, J., Abfalterer, W., Foley, B., Giorgi, E.E., Bhattacharya, T., Parker, M.D., 2020. Spike mutation pipeline reveals the emergence of a more transmissible form of SARS-CoV-2. *bioRxiv*.
- Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874.
- Lee, C.Y.-P., Lin, R.T., Renia, L., Ng, L.F., 2020. Serological approaches for COVID-19: epidemiologic perspective on surveillance and control. *Front. Immunol.* 11, 879.
- Li, J., Li, Z., Cui, X., Wu, C., 2020. Bayesian phylodynamic inference on the temporal evolution and global transmission of SARS-CoV-2. *J. Inf. Secur.* 81 (2), 318–356.
- Nakashima, A., 2020. The Global Emergences of Multiple SARS-CoV-2 Sub-Strains: Digital Annotations for Human Behaviors May Assist Automated Retracing of Symptomatic Features and Origins.
- Quick, J., 2020. nCoV-2019 Sequencing Protocol. *protocols.io*.
- Rambaut, A., Holmes, E.C., Hill, V., O’Toole, Á., McCrone, J., Ruis, C., du Plessis, L., Pybus, O.G., 2020. hCoV-2019/Lineages.
- Romano, M., Ruggiero, A., Squeglia, F., Maga, G., Berisio, R., 2020. A structural view of SARS-CoV-2 RNA replication machinery: RNA synthesis, proofreading and final capping. *Cells* 9, 1267.
- Shi, J., Wen, Z., Zhong, G., Yang, H., Wang, C., Huang, B., Liu, R., He, X., Shuai, L., Sun, Z., 2020. Susceptibility of ferrets, cats, dogs, and other domesticated animals to SARS-coronavirus 2. *Science* 368, 1016–1020.
- SIB, 2020. Betacoronavirus. Swiss Institute of Bioinformatics.
- Subissi, L., Posthuma, C.C., Collet, A., Zevenhoven-Dobbe, J.C., Gorbalenya, A.E., Decroly, E., Snijder, E.J., Canard, B., Imbert, I., 2014. One severe acute respiratory syndrome coronavirus protein complex integrates processive RNA polymerase and exonuclease activities. *Proc. Natl. Acad. Sci.* 111, E3900–E3909.
- Tang, X., Wu, C., Li, X., Song, Y., Yao, X., Wu, X., Duan, Y., Zhang, H., Wang, Y., Qian, Z., Cui, J., Lu, J., 2020. On the origin and continuing evolution of SARS-CoV-2. *Natl. Sci. Rev.* 7 (6), 1012–1023.
- Villabona-Arenas, C.J., Hanage, W.P., Tully, D.C., 2020. Phylogenetic interpretation during outbreaks requires caution. *Nat. Microbiol.* 5, 876–877.
- WHO, 2020a. WHO Director-General’s Opening Remarks at the Media Briefing on COVID-19 - 11 March 2020. World Health Organization.
- WHO, 2020c. Novel Coronavirus (2019-nCoV) Technical Guidance: Laboratory Testing for 2019-nCoV in Humans. World Health Organization.
- Yan, Y., Chang, L., Wang, L., 2020. Laboratory testing of SARS-CoV, MERS-CoV, and SARS-CoV-2 (2019-nCoV): current status, challenges, and countermeasures. *Rev. Med. Virol.* e2106.
- Yount, B., Roberts, R.S., Sims, A.C., Deming, D., Frieman, M.B., Sparks, J., Denison, M.R., Davis, N., Baric, R.S., 2005. Severe acute respiratory syndrome coronavirus group-specific open reading frames encode nonessential functions for replication in cell cultures and mice. *J. Virol.* 79, 14909–14922.