Substitutions in Spike and Nucleocapsid proteins of SARS-CoV-2 circulating in Colombia

Short title: Substitutions in SARS-CoV-2 Spike and nucleocapsid proteins

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

a. Unidad de Secuenciación y Genómica. Dirección de Investigación en Salud Pública, Instituto Nacional de Salud. Bogotá - 111321, Colombia

b. Grupo de Salud Materna y Perinatal. Dirección de Investigación en Salud Pública, Instituto Nacional de Salud. Bogotá - 111321, Colombia

c. Grupo de Parasitología. Dirección de Investigación en Salud Pública, Instituto Nacional de Salud. Bogotá - 111321, Colombia

d. Grupo de Microbiología, Dirección de Investigación en Salud Pública, Instituto Nacional de Salud, Bogotá, Colombia

e. Centro de Investigación en Salud para el Trópico–CIST, Universidad Cooperativa de Colombia, Santa Marta. Santa Martha - 470003, Colombia

f. Programa Ciencias de la Biodiversidad, Instituto de Investigación de Recursos Biológicos Alexander von Humboldt, Bogotá - 111311, Colombia.

g. Dirección de Redes en Salud Pública, Instituto Nacional de Salud. Bogotá - 111321, Colombia

h. Dirección de Vigilancia en Salud Pública, Instituto Nacional de Salud. Bogotá - 111321, Colombia

i. Grupo de Virología, Dirección de Redes en Salud Pública, Instituto Nacional de Salud. Bogotá - 111321, Colombia

j. Dirección General, Instituto Nacional de Salud. Bogotá - 111321, Colombia

k. Dirección de Investigación en Salud Pública, Instituto Nacional de Salud. Bogotá - 111321, Colombia

Symbol

*Corresponding author: Grupo de Parasitología. Dirección de Investigación en Salud Pública, Instituto Nacional de Salud. Bogotá - 111321, Colombia. E-mail: cefrancom@unal.edu.co

Abstract

Introduction: SARS-CoV-2 is a new member of the genus *Betacoronavirus*, responsible for the COVID-19 pandemics. The virus crossed the species barrier and established in the human population due to its ability to exploit the ACE receptor for virus entry, which is present and abundant in several tissues, including the lung and respiratory tract, gastrointestinal tract and hearth. Virus interaction with the cellular receptor is mediated by the surface protein, known as Spike. Another structural protein of major importance in the Nucleocapsid, directly interacting with the viral RNA to form the ribonucleocapsid, considered a multifunctional protein, and being the target of the most molecular diagnostics assays. Objective: To describe the frequency of substitutions in spike and nucleocapsid proteins of SARS-CoV-2 circulating in Colombia and evaluate the frequency of these substitutions in SARS-CoV-2 sequences from other countries of South America. Materials and methods: Samples of 43 patients were included for

viral RNA detection by real-time RT-PCR using the Charité-Berlin protocol for the amplification of the SARS-CoV-2 E and RdRp genes. Genome sequences were obtained through the Oxford Nanopore and Illumina MiSeq technologies, following the artic.network "nCoV-2019 sequencing protocol". Available genomes were consulted from GISAID, GenBank, and Genome sequence archive (GSA) and a total of 371 genomes sequences from South America were included. The genome sequences were aligned with the Muscle tool using the MEGA X software. Substitution matrices of the Colombian sequences respect to the reference genome (NC_045512) at the nucleotide and amino acid levels were generated for the spike and nucleocapsid gene.
Results: substitution D614G in the amino acid sequence of spike protein was found in 86.7% of the Colombian sequences; substitutions G181V and D936Y in 2.3%, respectively. Five substitutions were found in the nucleocapsid protein, with substitution R203K and G204R being the most frequent (13.95 %) in Colombia. The substitutions D614G in Spike and R203K-G204R in nucleocapsid have a frequency of 83% and 28% respectively in sequences from South America.
Conclusion: Non-synonymous substitutions were found in the spike and nucleocapsid proteins in Colombian genomes, the most frequent being D614G in Spike and R203K-G204R in nucleocapsid. These substitutions are frequent in the genomes reported for other South American countries. It is necessary to continue with genomic surveillance of the changes in Spike and Nucleocapsid proteins during the SARS-CoV-2 pandemic in Colombia and South America, even more considering that these proteins are the most commonly used antigen in serological tests.

Highlights

- The spike and nucleocapsid proteins of SARS-CoV-2 circulating in Colombia and South-American countries have similar patterns of non-synonymous substitutions

- Substitutions D614G in Spike and R203K-G204R in Nucleocapsid are the most frequent in Colombia and South-American countries

Key words

SARS-CoV-2, spike, nucleocapsid, Colombia, Non-synonymous substitutions, Outbreak lineages.

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, so that daily, about 70,000 new cases are reported globally (WHO, 2020a, 2020c). In Colombia, the first case of COVID-19 was announced by the Nation's Ministry of Health on March 6, in a traveler from Italy, after which the number of patients has exceeded 10,000 and over 500 deaths (INS, 2020). The SARS-CoV-2 genome consist of a single, positive-stranded RNA (ssRNA[+]), with 29,903 nucleotides long. This virus has shown an extraordinary infectiousness and transmission capacity in the human population (He et al., 2020), even exploring other vertebrate species (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 assays (molecular and immunologic) target these proteins (Alvarez-Diaz et al., 2020; Lee, Lin, Renia, & Ng, 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, how 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, Jacofsky, & Jacofsky, 2020). The objective of this study is to describe the frequency of substitutions in Spike and Nucleocapsid proteins of SARS-CoV-2 circulating in Colombia and evaluate the frequency of these substitutions in SARS-CoV-2 sequences from other countries of 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 do 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, 2020a). 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, 2020a).

SARS-CoV-2 specific oligonucleotides were used for the generation of amplicons by means of a high-fidelity DNA polymerase (Q5® High-Fidelity DNA Polymerase - (New England Biolabs Inc., UK, EB), in order to avoid the introduction of artificial substitutions.

2.4 Sequence Analysis

Databases as GISAID, NCBI, and GSA were consulted and a total of 371 sequences were included. The South American sequences were aligned with the Muscle tool (Edgar, 2004) using the MEGA X software (Kumar, Stecher, & Tamura, 2016). Substitutions matrices of SARS-CoV-2 sequences deposited in GISAID (Table 1), compared to the reference genome (NC_045512) at the nucleotide and amino acid levels were generated for the Spike and Nucleocapsid genes and proteins. The Colombian sequences were analyzed to determine the frequency of amino acid changes in Spike and Nucleocapsid proteins. Based on Phylogenetic Assignment of Named Global Outbreak LINeages (PANGOLIN) https://github.com/hCoV- 2019/pangolin different lineages were defined in Colombian genomes.

3. Results

3.1 Non-synonymous substitutions in the Spike and Nucleocapsid proteins

Three amino acid changes were observed in the spike protein of Colombian SARS-CoV-2 sequences compared to the reference sequence (NC_045512.2). Spike protein substitution D614G is present in 81% (35/43) of the sequences reported in this study. The following substitutions were also found in lower frequency: G181V

2.3% (1/43) and D936Y 2.3% (1/43) (Table 1). In nucleocapsid protein, five amino acid changes were found when Colombian genomes reported in this study were compared to the reference sequence (NC_045512.2). The most frequent changes were R203K and G204R in 13.95% (6/43) of the sequences. Amino acid changes 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.

Nonsynonymous substitution D614G in the spike protein was the most frequent observed in Colombian sequences included in this study. Genomic resource databases such as NCBI, GISAID and GSA were consulted to determine if the predominant substitutions in genomes reported in Colombia were present in other countries in South America. A total of 86 genomes from Colombia, 29 from Argentina, 92 from Brazil, 143 from Chile, 4 from Ecuador, 2 from Peru and 11 from Uruguay were reported as of May 19th, 2020. G614 of SARS-CoV-2 was frequent in Argentina, Brazil, Chile, Colombia and Peru, with 80-100% of the reported sequences (Figure 1A). However, in Ecuador and Uruguay, D614G was observed in less percentages, possibly related to the low number of genomes reported or to the introduction history of SARS-CoV-2 in these countries. Finally, some reported genomes had indeterminate nucleotides at the site of the substitution and could not be included in the analysis.

3.2 Spatiotemporal distribution of substitutions in Spike and Nucleocapsid

A temporal analysis of the distribution of D614G in the 131 Colombian reported genomes, including the genomes reported in this study, was performed using the available collection sample data in the associated metadata (Figure 1B). Since the first case report date in Colombia and during March 2020, 14% of Colombian sequences exhibited aspartate at position 614 in the spike protein (D614), while in April 2020, the frequency reduced to 4%.

Nonsynonymous substitutions R203K and G204R (13.95%), which are the hallmarks of the B.1.1 lineage, were the most frequent substitutions in Nucleocapsid protein of Colombian sequences included in this study. Both substitutions are frequent in Argentina and Brazil, present in the 55% of the reported sequences (Fig. 1C). In Peru, Ecuador and Chile the frequency were similar to reported for Colombia, while in Uruguay the substitution was not present. Temporal analysis of the distribution of the R203/G204 and K203/R204 subtypes of the Nucleocapsid for the 131 Colombian genomes shows that the frequency of the K203/R204 subtype was 9% in March 2020, and became 19% in April 2020 (Fig. 1D), suggesting that the frequency of this substitution could increase over time and reach values similar to those reported for Brazil and Chile.

Regarding the other nonsynonymous substitutions that were observed in less frequency, the G181V substitution of the spike protein, was observed in one Colombian sequence and in two sequences from Brazil, while the D936Y substitution of spike protein was not observed in other genomes from South America. The R191C substitution in the nucleocapsid protein was observed in two

Colombian sequences and in another Argentina sequence, R209I was observed in nine sequences from Chile and finally, G238C was exclusive to three Colombian sequences.

3.3 Nonsynonymous substitutions and circulating lineages in Colombia

In the analysis of the nonsynonymous substitutions in the Spike protein of Colombian sequences, the G181V substitution was observed in a Colombian sequence classified as lineage B.1. The D614G substitution was observed in sequences classified as lineage B.1, in sublineages B.1.1, B.1.3, B.1.5, B.1.8, and B.1.11 and lineage B.2. The D936Y substitution was observed in sequences classified as of lineage B.1.5. As for Nucleocapsid protein, the R191C substitution was observed in two Colombian sequences of lineage B.1.5. The R203K and G204R substitutions were observed in sequences classified as lineage B.1. Finally the R209I substitution in sequences classified as lineage B.2.5 and G238C substitution in sequences classified as the lineage A.5.

4. Discussion

The first COVID-19 case in Colombia was confirmed on March 6, 2020, from a person who entered the country from Italy on February 26, 2020. In June 1th, a total de 30.493 cases have been reported with 969 deaths (INS, 2020). This study evidenced the presence of the D614G substitution in spike protein in 89.6% (112/125) of sequences since March 10, 2020, while strains circulating in previous dates reveal the original substitutions reported in Wuhan (Bhattacharyya et al., 2020), including the first Colombian strain sequenced from a symptomatic patient

sampled on March 11, 2020, linked to the second generation of a transmission chain starting with an imported case from Spain.

D614G was detected in 86% of sequences from March, 2020 and in 96% of sequences from April, 2020, being present in most of the geographical departments of Colombia with genomic information available. D614G has been found in 13 departments located from the Caribbean coast to the Amazon region in only 3 months since the first case was reported in Colombia. While several studies have emphasized the potential role of the D614G substitution in favoring the virus transmissibility (Becerra-Flores & Cardozo, 2020; Bhattacharyya et al., 2020; Brufsky, 2020), our data only suggest a founder effect determined by the multiple introduction of G614-bearing B.1 sublineages.

In-silico analysis of D614G substitution in Spike suggests a more stable union with the ligand ACE2 (Becerra-Flores & Cardozo, 2020; Brufsky, 2020; Nakashima, 2020). Preliminary results suggest that the presence of D614G statistically correlate with the mean number of deaths occurred in several European countries and Brazil (Becerra-Flores & Cardozo, 2020). Korber et al proposes an association with infectivity, rapid dispersion and high viral load in infected patients (Korber et al., 2020). Kim et al suggested that this substitution could affect the ability to activate long-term immune response against SARS-CoV-2. Antigenicity analysis of amino acids 613 to 621 from Spike protein related decreases in the epitope recognition in B cells, in charge of stimulating adaptive immune response by producing antibodies. Therefore, this reduced immunogenicity leads to recurrent

infections by this lineage and must be a point to consider in vaccine development (Kim, Nguyen, Park, Park, & Chung, 2020).

On the other hand, the double substitutions R203K and G204R in the Nucleocapsid protein was also identified in South American sequences, in 28% of the isolates with mutations in three consecutives nucleotides. Gupta et al reported this alteration in isolates from Belgium, Brazil, Peru, Mexico, Nigeria, Switzerland, Vietnam and India (Gupta & Mandal). Ayub observed an association between the presences of this substitution with case fatality (Ayub, 2020). While the most amino acid changes evidenced in the Spike and Nucleocapsid proteins cannot be directly related to a specific lineage, as has been reported in previous studies (Bhattacharyya et al., 2020; Korber et al., 2020), the B.1.1 lineage is defined by the three-nucleotide mutation leading to the two consecutives amino acid changes.

RNA viruses are known to possess very high substitution rates (Li, Li, Cui, & Wu, 2020; Tang et al., 2020), leading to high genetic variability and to the action of the evolutionary mechanisms of natural selection and genetic drift. 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.

Recombinant proteins or synthetic peptides from SARS-CoV-2 derived from the reported genomic information is a well-explored alternative for the production of antigens to be used in serological tests, considering that the production of total

viral antigen implies biosafety challenges and the need for a BSL3 laboratory. Taking into account that Spike and Nucleocapsid are the major immunogenic proteins of SARS and MERS coronavirus and the first choice for producing recombinant antigens (Yan, Chang, & Wang, 2020), it is essential to continue with the evolutionary monitoring of genes included in the original designs of serological tests, because amino acid changes can affect the conformational structure and physical-chemical properties of the proteins. Considering that reduced immunogenicity in Spike protein is associated with the D614G amino acid change, it could also be having impact in serosurveillance studies in countries where the subtype G614 is prevalent, such as Spain, where a seroprevalence of 5% was registered (MSCBS, 2020).

5. Conclusion

Amino acid changes were found in the Spike and Nucleocapsid proteins in Colombian genomes of SARS-CoV-2, the most frequent being D614G in Spike and R203K-G204R in Nucleocapsid. These substitutions are frequent in the genomes reported for other South American countries. It is necessary to continue with genomic surveillance of changes in spike and nucleocapsid proteins during the SARS-CoV-2 pandemic in Colombia and South America, even more considering that these proteins are the most commonly antigen/target used in serological and molecular tests. Identification of nucleotide substitutions, amino acid changes and their frequency in the circulating virus in the region is an important information for decision-making of public health, such as if a vaccine candidate is appropriate for a

specific population. Monitoring genetic variability is important for the refinement of SARS-CoV-2 serological and molecular tests.

6. 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.

Conflict of interests

The authors declare no competing interest.

Funding

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

References

Alvarez-Diaz, D. A., Franco-Munoz, C., Laiton-Donato, K., Usme-Ciro, J. A., Franco-Sierra, N. D., Florez, A. C., . . . 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. *medRxiv*, 2020.2005.2022.20107292. doi:10.1101/2020.05.22.20107292

Ayub, M. I. (2020). Reporting Two SARS-CoV-2 Strains Based on A Unique Trinucleotide-Bloc Mutation and Their Potential Pathogenic Difference. *Published online April, 19*.

Becerra□Flores, M., & Cardozo, T. (2020). SARS□CoV□2 viral spike G614 mutation exhibits higher case fatality rate. *International Journal of Clinical Practice*.

Bhattacharyya, C., Das, C., Ghosh, A., Singh, A. K., Mukherjee, S., Majumder, P. P., . . . 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*.

Brufsky, A. (2020). Distinct Viral Clades of SARS□CoV□2: Implications for Modeling of Viral Spread. *Journal of Medical Virology*.

Corman, V. M., Landt, O., Kaiser, M., Molenkamp, R., Meijer, A., Chu, D. K., . . . Drosten, C. (2020). Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin, 25*(3), 2000045. doi:10.2807/1560-7917.ES.2020.25.3.2000045

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. *Nature Reviews Microbiology, 7*(3), 226-236.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res, 32*(5), 1792-1797. doi:10.1093/nar/gkh340

Gupta, A. M., & Mandal, S. Loss of Epitopes from SARS-Cov-2 Proteins for Non-synonymous Mutations: A Potential Global Threat.

He, X., Lau, E. H., Wu, P., Deng, X., Wang, J., Hao, X., . . . Tan, X. (2020). Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nature medicine, 26*(5), 672-675.

INS. (2020). Coronavirus (COVID - 2019) en Colombia. Retrieved from https://www.ins.gov.co/Noticias/Paginas/Coronavirus.aspx

Jacofsky, D., Jacofsky, E. M., & Jacofsky, M. (2020). Understanding Antibody Testing for COVID-19. *The Journal of Arthroplasty*.

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*(2), 220.

Korber, B., Fischer, W., Gnanakaran, S. G., Yoon, H., Theiler, J., Abfalterer, W., . . . 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. *Molecular biology and evolution, 33*(7), 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. *Frontiers in Immunology, 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 Infect*. doi:10.1016/j.jinf.2020.04.016

MSCBS. (2020). *ESTUDIO ENE-COVID19: PRIMERA RONDA ESTUDIO NACIONAL DE SERO-EPIDEMIOLOGÍA DE LA INFECCIÓN POR SARS-COV-2 EN ESPAÑA*. Spain: Ministerio de Sanidad - España Retrieved from https://www.mscbs.gob.es/ciudadanos/ene-covid/home.htm.

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. (2020a). nCoV-2019 sequencing protocol. Retrieved from https://www.protocols.io/view/ncov-2019-sequencing-protocol-bbmuik6w

Shi, J., Wen, Z., Zhong, G., Yang, H., Wang, C., Huang, B., . . . Sun, Z. (2020). Susceptibility of ferrets, cats, dogs, and other domesticated animals to SARS–coronavirus 2. *Science*, *368*(6494), 1016-1020.

SIB. (2020). Betacoronavirus. Retrieved from https://viralzone.expasy.org/764?outline=all_by_species

Tang, X., Wu, C., Li, X., Song, Y., Yao, X., Wu, X., . . . Lu, J. (2020). On the origin and continuing evolution of SARS-CoV-2. *National Science Review*. doi:10.1093/nsr/nwaa036

WHO. (2020a). WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. Retrieved from https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020

WHO. (2020c). Novel Coronavirus (2019-nCoV) technical guidance: Laboratory testing for 2019-nCoV in humans. Retrieved from https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/laboratory-guidance

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. *Reviews in Medical Virology*, e2106.

Yount, B., Roberts, R. S., Sims, A. C., Deming, D., Frieman, M. B., Sparks, J., . . . 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*(23), 14909-14922. doi:10.1128/JVI.79.23.14909-14922.2005

Figures and Tables

Table 1. Nucleotide substitutions and amino acid changes in the Spike and Nucleocapsid proteins of 43 genomes of SARS-CoV-2 from Colombia reported in this study. Location of the substitutions was estimated base on the reference genomic sequence NC_045512.2. Left panel: nucleotides; right panel: amino acids. [?]: Uncovered nucleotide and amino acids during complete genome sequencing.

GISAID Accession ID	Sequence ID	Nucleotides												Aminoacids					
		S					N							S		N			
		Spike					Nucleocapsid							Spike		Nucleocapsid			
		22104	23403	23443	23707	24368	28884	28881	28882	28883	28899	28895	29269	00181	00614	00193	00203	00204	00209	00238
	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
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	.

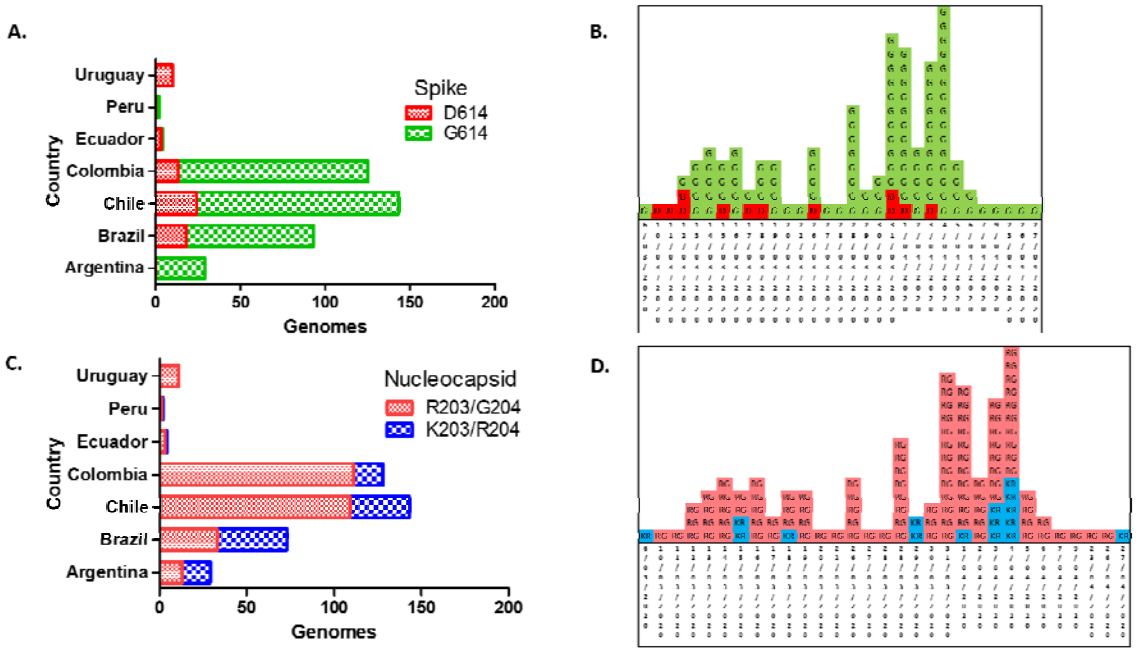
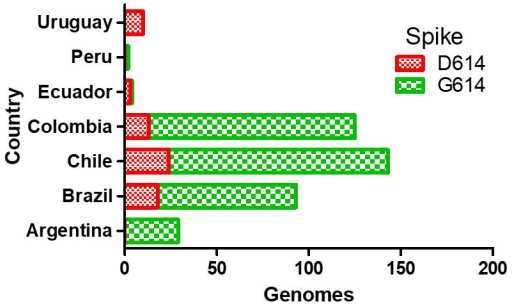
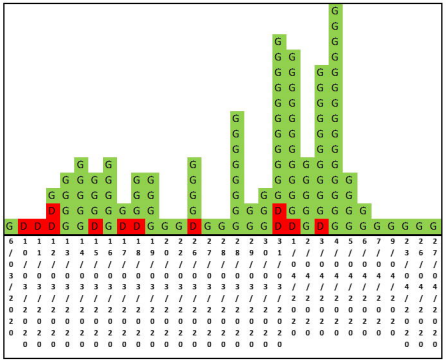


Figure 1. Spatiotemporal distribution of the most frequently substitutions in Spike and Nucleocapsid of SARS-CoV-2. A. Distribution of subtypes D614 and G614 of spike protein in SARS-CoV-2 genomes reported as of May 19th, 2020 in Colombia and other South American countries; in red, variant D614 and in green, variant G614. **B.** Distribution of subtypes D614 and G614 of spike protein in SARS-CoV-2 genomes reported between March and April, 2020 in Colombia, according with the sample collection date, variant D614 and in green, variant G614. **C.** Distribution of subtypes R203/G204 and K203/R204 of nucleocapsid protein in SARS-CoV-2 genomes reported as of May 19th, 2020 in Colombia and other South American countries, in red, variant R203/G204 and in blue, variant K203/R204. **D.** Distribution of subtypes R203/G204 and K203/R204 of nucleocapsid protein in SARS-CoV-2 genomes reported between March and April, 2020 in Colombia, according with the sample collection date in Colombia, in red, variant R203/G204 and in blue, variant K203/R204.

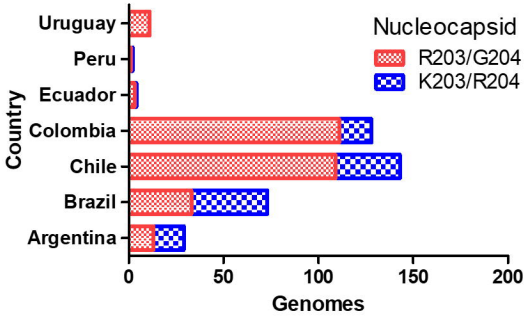
A.


Uruguay
Peru
Ecuador
Colombia
Chile
Brazil
Argentina
Country
Spike
D614
G614
0
50
100
150
200
Genomes

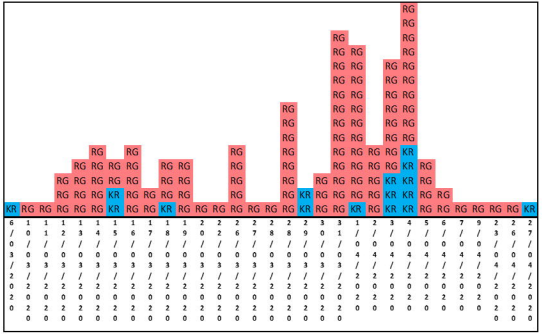
B.



C.


Uruguay
Peru
Ecuador
Colombia
Chile
Brazil
Argentina
Country
Nucleocapsid
R203/G204
K203/R204
0
50
100
150
200
Genomes

D.



GISAID Accession ID	Sequence ID	Nucleotides												Aminoacids						
		S					N							S			N			
		Spike					Nucleocapsid							Spike			Nucleocapsid			
		22104	23403	23407	23408	23604	28841	28842	28843	28844	28845	28846	29669	21614	21619	21646	12013	20049	20049	20238
	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
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	.	.