

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

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

4  
5 Carlos Franco-Muñoz<sup>a, c, \*</sup>, Diego A. Álvarez-Díaz<sup>a, b</sup>, Katherine Laiton-Donato<sup>a</sup>,  
6 Magdalena Wiesner<sup>d</sup>, Patricia Escandón<sup>d</sup>, José A. Usme-Ciro<sup>a, e</sup>, Nicolás D.  
7 Franco-Sierra<sup>f</sup>, Astrid C. Flórez-Sánchez<sup>g</sup>, Sergio Gómez-Rangel<sup>i</sup>, Luz D.  
8 Rodríguez-Calderon<sup>i</sup>, Juliana Barbosa-Ramirez<sup>i</sup>, Erika Ospitia-Baez<sup>i</sup>, Diana M.  
9 Walteros<sup>h</sup>, Martha L. Ospina-Martinez<sup>j</sup>, Marcela Mercado-Reyes.<sup>a, b, k</sup>

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

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

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

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

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

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

22 g. Dirección de Redes en Salud Pública, Instituto Nacional de Salud. Bogotá -  
23 111321, Colombia

24 h. Dirección de Vigilancia en Salud Pública, Instituto Nacional de Salud. Bogotá -  
25 111321, Colombia

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

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

29 k. Dirección de Investigación en Salud Pública, Instituto Nacional de Salud. Bogotá  
30 - 111321, Colombia

31

## 32 **Symbol**

33 \*Corresponding author: Grupo de Parasitología. Dirección de Investigación en  
34 Salud Pública, Instituto Nacional de Salud. Bogotá - 111321, Colombia. E-mail:

35 [cefrancom@unal.edu.co](mailto:cefrancom@unal.edu.co)

36

## 37 **Abstract**

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

51 viral RNA detection by real-time RT-PCR using the Charité-Berlin protocol for the  
52 amplification of the SARS-CoV-2 E and RdRp genes. Genome sequences were  
53 obtained through the Oxford Nanopore and Illumina MiSeq technologies, following  
54 the artic.network "nCoV-2019 sequencing protocol". Available genomes were  
55 consulted from GISAID, GenBank, and Genome sequence archive (GSA) and a  
56 total of 371 genomes sequences from South America were included. The genome  
57 sequences were aligned with the Muscle tool using the MEGA X software.

58 Substitution matrices of the Colombian sequences respect to the reference  
59 genome (NC\_045512) at the nucleotide and amino acid levels were generated for  
60 the spike and nucleocapsid gene.

61 Results: substitution D614G in the amino acid sequence of spike protein was found  
62 in 86.7% of the Colombian sequences; substitutions G181V and D936Y in 2.3%,  
63 respectively. Five substitutions were found in the nucleocapsid protein, with  
64 substitution R203K and G204R being the most frequent (13.95 %) in Colombia.

65 The substitutions D614G in Spike and R203K-G204R in nucleocapsid have a  
66 frequency of 83% and 28% respectively in sequences from South America.

67 Conclusion: Non-synonymous substitutions were found in the spike and  
68 nucleocapsid proteins in Colombian genomes, the most frequent being D614G in  
69 Spike and R203K-G204R in nucleocapsid. These substitutions are frequent in the  
70 genomes reported for other South American countries. It is necessary to continue  
71 with genomic surveillance of the changes in Spike and Nucleocapsid proteins  
72 during the SARS-CoV-2 pandemic in Colombia and South America, even more  
73 considering that these proteins are the most commonly used antigen in serological  
74 tests.

75

## 76 **Highlights**

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

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

82

### 83 **Key words**

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

86

### 87 **1. Introduction**

88 The recently emerged SARS-CoV-2 responsible for the coronavirus disease 2019  
89 (COVID-19) pandemic, has increased significantly in the number of cases and  
90 deaths, so that daily, about 70,000 new cases are reported globally  
91 (WHO, 2020a, 2020c). In Colombia, the first case of COVID-19 was announced by  
92 the Nation's Ministry of Health on March 6, in a traveler from Italy, after which the  
93 number of patients has exceeded 10,000 and over 500 deaths (INS, 2020). The  
94 SARS-CoV-2 genome consist of a single, positive-stranded RNA (ssRNA[+]), with  
95 29,903 nucleotides long. This virus has shown an extraordinary infectiousness and  
96 transmission capacity in the human population (He et al., 2020), even exploring  
97 other vertebrate species (Shi et al., 2020). The SARS-CoV-2 genome has nine  
98 open reading frames (ORFs); the first one, subdivided in ORF1a and ORF1b by  
99 ribosomal frameshifting, encodes the polyproteins pp1a and pp1ab which are  
100 processed into non-structural proteins involved in subgenomic/genome length RNA  
101 synthesis and virus replication. Structural proteins, Spike (S), Envelope (E),  
102 Membrane (M), and Nucleocapsid (N) are encoded in subgenomic mRNA  
103 transcripts within ORFs 2, 4, 5, and 9, respectively (SIB, 2020; Yount et al., 2005)  
104  
105 Spike protein, a type I membrane glycoprotein, is the most exposed viral protein  
106 recognized by the cellular receptor angiotensin-2-converting enzyme (ACE2)  
107 during the infection of the lower respiratory tract and considered the main inducer  
108 of neutralizing antibodies. The N protein is associated with the RNA genome to

109 form the ribonucleocapsid and is abundantly expressed during infection. Both N  
110 and S proteins are highly immunogenic and most commercial COVID-19 diagnostic  
111 assays (molecular and immunologic) target these proteins (Alvarez-Diaz et al.,  
112 2020; Lee, Lin, Renia, & Ng, 2020).

113

114 Furthermore, non-synonymous mutations in the S and N proteins have been  
115 reported, their implications in the potential emergence of antigenically distinct  
116 and/or more virulent strains remain to be studied, although it was reported that  
117 mutations in the receptor-binding domain (RBD) at the S protein of SARS-CoV-  
118 related viruses disrupt the antigenic structure and binding activity of RBD to ACE2  
119 (Du et al., 2009) Similarly, how non-synonymous mutations could impact the  
120 antibody response and the specificity and sensitivity of serological tests for COVID-  
121 19 diagnosis is unknown. Thus, identifying variable sites at these proteins can  
122 provide a valuable resource for choosing the target antigens for the development of  
123 SARS-CoV-2 vaccines, therapeutics and diagnostic tests (Du et al., 2009;  
124 Jacofsky, Jacofsky, & Jacofsky, 2020). The objective of this study is to describe the  
125 frequency of substitutions in Spike and Nucleocapsid proteins of SARS-CoV-2  
126 circulating in Colombia and evaluate the frequency of these substitutions in SARS-  
127 CoV-2 sequences from other countries of South America.

128

## 129 **2. Materials and methods**

### 130 **2.1 Ethics**

131 This work was developed according to the national law 9/1979, decrees 786/1990  
132 and 2323/2006, which establishes that the Instituto Nacional de Salud (INS) from  
133 Colombia is the reference lab and health authority of the national network of  
134 laboratories and in cases of public health emergency or those in which scientific  
135 research for public health purposes as required. The INS is authorized to use the  
136 biological material for research purposes, without informed consent, which includes

137 the anonymous disclosure of results. This study was performed following the  
138 ethical standards of the Declaration of Helsinki 1964 and its later amendments.  
139 The information used for this study comes from secondary sources of data that  
140 were previously anonymized and do protect patient data.

## 141 **2.2 Patients and samples**

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

## 150 **2.3 Complete genome sequencing of SARS-CoV-2 through NGS**

151 NGS of SARS-CoV-2 from 43 patients was performed using the amplicon-based  
152 Illumina and Nanopore sequencing approaches, ARTIC network protocol (Quick,  
153 2020a). Following cDNA synthesis with SuperScript IV reverse transcriptase  
154 (Thermo Fisher Scientific, Waltham, MA, USA) and random hexamers (Thermo  
155 Fisher Scientific, Waltham, MA, USA), a set of 400-bp tiling amplicons across the  
156 whole genome of SARS-CoV-2 were generated using the primer schemes nCoV-  
157 2019/V3 (Quick, 2020a).

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

## 162 **2.4 Sequence Analysis**

163 Databases as GISAID, NCBI, and GSA were consulted and a total of 371  
164 sequences were included. The South American sequences were aligned with the  
165 Muscle tool (Edgar, 2004) using the MEGA X software (Kumar, Stecher, & Tamura,  
166 2016). Substitutions matrices of SARS-CoV-2 sequences deposited in GISAID  
167 (Table 1), compared to the reference genome (NC\_045512) at the nucleotide and  
168 amino acid levels were generated for the Spike and Nucleocapsid genes and  
169 proteins. The Colombian sequences were analyzed to determine the frequency of  
170 amino acid changes in Spike and Nucleocapsid proteins. Based on Phylogenetic  
171 Assignment of Named Global Outbreak LINEages (PANGOLIN)  
172 [https://github.com/hCoV-](https://github.com/hCoV-2019/pangolin) 2019/pangolin different lineages were defined in  
173 Colombian genomes.

## 174 **3. Results**

### 175 **3.1 Non-synonymous substitutions in the Spike and Nucleocapsid proteins** 176

177 Three amino acid changes were observed in the spike protein of Colombian SARS-  
178 CoV-2 sequences compared to the reference sequence (NC\_045512.2). Spike  
179 protein substitution D614G is present in 81% (35/43) of the sequences reported in  
180 this study. The following substitutions were also found in lower frequency: G181V

181 2.3% (1/43) and D936Y 2.3% (1/43) (Table 1). In nucleocapsid protein, five amino  
182 acid changes were found when Colombian genomes reported in this study were  
183 compared to the reference sequence (NC\_045512.2). The most frequent changes  
184 were R203K and G204R in 13.95% (6/43) of the sequences. Amino acid changes  
185 R191C, R209I and G238C were found in 4.65% (2/43), 4.65% (2/43) and 6.97%  
186 (3/43) of the Colombian sequences, respectively (Table 1). Some nucleotide  
187 substitutions were synonymous.

188

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

202

203 **3.2 Spatiotemporal distribution of substitutions in Spike and Nucleocapsid**



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

210

211 Nonsynonymous substitutions R203K and G204R (13.95%), which are the  
212 hallmarks of the B.1.1 lineage, were the most frequent substitutions in  
213 Nucleocapsid protein of Colombian sequences included in this study. Both  
214 substitutions are frequent in Argentina and Brazil, present in the 55% of the  
215 reported sequences (Fig. 1C). In Peru, Ecuador and Chile the frequency were  
216 similar to reported for Colombia, while in Uruguay the substitution was not present.  
217 Temporal analysis of the distribution of the R203/G204 and K203/R204 subtypes of  
218 the Nucleocapsid for the 131 Colombian genomes shows that the frequency of the  
219 K203/R204 subtype was 9% in March 2020, and became 19% in April 2020 (Fig.  
220 1D), suggesting that the frequency of this substitution could increase over time and  
221 reach values similar to those reported for Brazil and Chile.

222

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

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

231

### 232 **3.3 Nonsynonymous substitutions and circulating lineages in Colombia**

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

## 243 **4. Discussion**

244 The first COVID-19 case in Colombia was confirmed on March 6, 2020, from a  
245 person who entered the country from Italy on February 26, 2020. In June 1<sup>th</sup>, a total  
246 de 30.493 cases have been reported with 969 deaths (INS, 2020). This study  
247 evidenced the presence of the D614G substitution in spike protein in 89.6%  
248 (112/125) of sequences since March 10, 2020, while strains circulating in previous  
249 dates reveal the original substitutions reported in Wuhan (Bhattacharyya et al.,  
250 2020), including the first Colombian strain sequenced from a symptomatic patient

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

253

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

263

264 *In-silico* analysis of D614G substitution in Spike suggests a more stable union with  
265 the ligand ACE2 (Becerra-Flores & Cardozo, 2020; Brufsky, 2020; Nakashima,  
266 2020). Preliminary results suggest that the presence of D614G statistically  
267 correlate with the mean number of deaths occurred in several European countries  
268 and Brazil (Becerra-Flores & Cardozo, 2020). Korber et al proposes an association  
269 with infectivity, rapid dispersion and high viral load in infected patients (Korber et  
270 al., 2020). Kim et al suggested that this substitution could affect the ability to  
271 activate long-term immune response against SARS-CoV-2. Antigenicity analysis of  
272 amino acids 613 to 621 from Spike protein related decreases in the epitope  
273 recognition in B cells, in charge of stimulating adaptive immune response by  
274 producing antibodies. Therefore, this reduced immunogenicity leads to recurrent

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

277

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

288

289 RNA viruses are known to possess very high substitution rates (Li, Li, Cui, & Wu,  
290 2020; Tang et al., 2020), leading to high genetic variability and to the action of the  
291 evolutionary mechanisms of natural selection and genetic drift. Despite some  
292 evolutionary changes may be in fact adaptive, it is important to be careful with  
293 conclusions in the absence of an experimental model to evaluate the impact of  
294 every mutation in the virus phenotype and virus-host interaction.

295

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

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

310

## 311 5. Conclusion

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

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

324

## 325 **6. Acknowledgements**

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

330

## 331 **Conflict of interests**

332 The authors declare no competing interest.

333

## 334 **Funding**

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

336

## 337 **References**

- 338 Alvarez-Diaz, D. A., Franco-Munoz, C., Laiton-Donato, K., Usme-Ciro, J. A., Franco-  
339 Sierra, N. D., Florez, A. C., . . . Mercado-Reyes, M. (2020). Molecular analysis of  
340 several in-house rRT-PCR protocols for SARS-CoV-2 detection in the context of  
341 genetic variability of the virus in Colombia. *medRxiv*, 2020.2005.2022.20107292.  
342 doi:10.1101/2020.05.22.20107292
- 343 Ayub, M. I. (2020). Reporting Two SARS-CoV-2 Strains Based on A Unique  
344 Trinucleotide-Bloc Mutation and Their Potential Pathogenic Difference. *Published*  
345 *online April, 19*.
- 346 Becerra-Flores, M., & Cardozo, T. (2020). SARS-CoV-2 viral spike G614 mutation  
347 exhibits higher case fatality rate. *International Journal of Clinical Practice*.
- 348 Bhattacharyya, C., Das, C., Ghosh, A., Singh, A. K., Mukherjee, S., Majumder, P. P., . . .  
349 Biswas, N. K. (2020). Global Spread of SARS-CoV-2 Subtype with Spike Protein  
350 Mutation D614G is Shaped by Human Genomic Variations that Regulate  
351 Expression of Tmprss2 and Mx1 Genes. *bioRxiv*.

- 352 Brufsky, A. (2020). Distinct Viral Clades of SARS-CoV-2: Implications for Modeling of  
353 Viral Spread. *Journal of Medical Virology*.
- 354 Corman, V. M., Landt, O., Kaiser, M., Molenkamp, R., Meijer, A., Chu, D. K., . . . Drosten,  
355 C. (2020). Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR.  
356 *Euro surveillance : bulletin Europeen sur les maladies transmissibles = European*  
357 *communicable disease bulletin*, 25(3), 2000045. doi:10.2807/1560-  
358 7917.ES.2020.25.3.2000045
- 359 Du, L., He, Y., Zhou, Y., Liu, S., Zheng, B.-J., & Jiang, S. (2009). The spike protein of  
360 SARS-CoV—a target for vaccine and therapeutic development. *Nature Reviews*  
361 *Microbiology*, 7(3), 226-236.
- 362 Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high  
363 throughput. *Nucleic Acids Res*, 32(5), 1792-1797. doi:10.1093/nar/gkh340
- 364 Gupta, A. M., & Mandal, S. Loss of Epitopes from SARS-Cov-2 Proteins for Non-  
365 synonymous Mutations: A Potential Global Threat.
- 366 He, X., Lau, E. H., Wu, P., Deng, X., Wang, J., Hao, X., . . . Tan, X. (2020). Temporal  
367 dynamics in viral shedding and transmissibility of COVID-19. *Nature medicine*,  
368 26(5), 672-675.
- 369 INS. (2020). Coronavirus (COVID - 2019) en Colombia. Retrieved from  
370 <https://www.ins.gov.co/Noticias/Paginas/Coronavirus.aspx>
- 371 Jacofsky, D., Jacofsky, E. M., & Jacofsky, M. (2020). Understanding Antibody Testing for  
372 COVID-19. *The Journal of Arthroplasty*.
- 373 Kim, S.-J., Nguyen, V.-G., Park, Y.-H., Park, B.-K., & Chung, H.-C. (2020). A Novel  
374 Synonymous Mutation of SARS-CoV-2: Is This Possible to Affect Their  
375 Antigenicity and Immunogenicity? *Vaccines*, 8(2), 220.
- 376 Korber, B., Fischer, W., Gnanakaran, S. G., Yoon, H., Theiler, J., Abfalterer, W., . . .  
377 Parker, M. D. (2020). Spike mutation pipeline reveals the emergence of a more  
378 transmissible form of SARS-CoV-2. *bioRxiv*.
- 379 Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: molecular evolutionary genetics  
380 analysis version 7.0 for bigger datasets. *Molecular biology and evolution*, 33(7),  
381 1870-1874.
- 382 Lee, C. Y.-P., Lin, R. T., Renia, L., & Ng, L. F. (2020). Serological Approaches for  
383 COVID-19: Epidemiologic Perspective on Surveillance and Control. *Frontiers in*  
384 *Immunology*, 11, 879.
- 385 Li, J., Li, Z., Cui, X., & Wu, C. (2020). Bayesian phylodynamic inference on the temporal  
386 evolution and global transmission of SARS-CoV-2. *J Infect*.  
387 doi:10.1016/j.jinf.2020.04.016
- 388 MSCBS. (2020). *ESTUDIO ENE-COVID19: PRIMERA RONDA ESTUDIO NACIONAL*  
389 *DE SERO-EPIDEMIOLOGÍA DE LA INFECCIÓN POR SARS-COV-2 EN*  
390 *ESPAÑA*. Spain: Ministerio de Sanidad - España Retrieved from  
391 <https://www.mscbs.gob.es/ciudadanos/ene-covid/home.htm>.
- 392 Nakashima, A. (2020). The global emergences of multiple SARS-CoV-2 sub-strains:  
393 Digital annotations for human behaviors may assist automated retracing of  
394 symptomatic features and origins.
- 395 Quick, J. (2020a). nCoV-2019 sequencing protocol. Retrieved from  
396 <https://www.protocols.io/view/ncov-2019-sequencing-protocol-bbmuik6w>



- 397 Shi, J., Wen, Z., Zhong, G., Yang, H., Wang, C., Huang, B., . . . Sun, Z. (2020).  
398 Susceptibility of ferrets, cats, dogs, and other domesticated animals to SARS–  
399 coronavirus 2. *Science*, 368(6494), 1016-1020.
- 400 SIB. (2020). Betacoronavirus. Retrieved from  
401 [https://viralzone.expasy.org/764?outline=all\\_by\\_species](https://viralzone.expasy.org/764?outline=all_by_species)
- 402 Tang, X., Wu, C., Li, X., Song, Y., Yao, X., Wu, X., . . . Lu, J. (2020). On the origin and  
403 continuing evolution of SARS-CoV-2. *National Science Review*.  
404 doi:10.1093/nsr/nwaa036
- 405 WHO. (2020a). WHO Director-General's opening remarks at the media briefing on  
406 COVID-19 - 11 March 2020. Retrieved from  
407 [https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-](https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020)  
408 [the-media-briefing-on-covid-19---11-march-2020](https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020)
- 409 WHO. (2020c). Novel Coronavirus (2019-nCoV) technical guidance: Laboratory testing for  
410 2019-nCoV in humans. Retrieved from  
411 [https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-](https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/laboratory-guidance)  
412 [guidance/laboratory-guidance](https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/laboratory-guidance)
- 413 Yan, Y., Chang, L., & Wang, L. (2020). Laboratory testing of SARS-CoV, MERS-CoV,  
414 and SARS-CoV-2 (2019-nCoV): Current status, challenges, and  
415 countermeasures. *Reviews in Medical Virology*, e2106.
- 416 Yount, B., Roberts, R. S., Sims, A. C., Deming, D., Frieman, M. B., Sparks, J., . . . Baric,  
417 R. S. (2005). Severe acute respiratory syndrome coronavirus group-specific open  
418 reading frames encode nonessential functions for replication in cell cultures and  
419 mice. *J Virol*, 79(23), 14909-14922. doi:10.1128/JVI.79.23.14909-14922.2005  
420
- 421
- 422
- 423
- 424
- 425
- 426
- 427
- 428
- 429
- 430
- 431

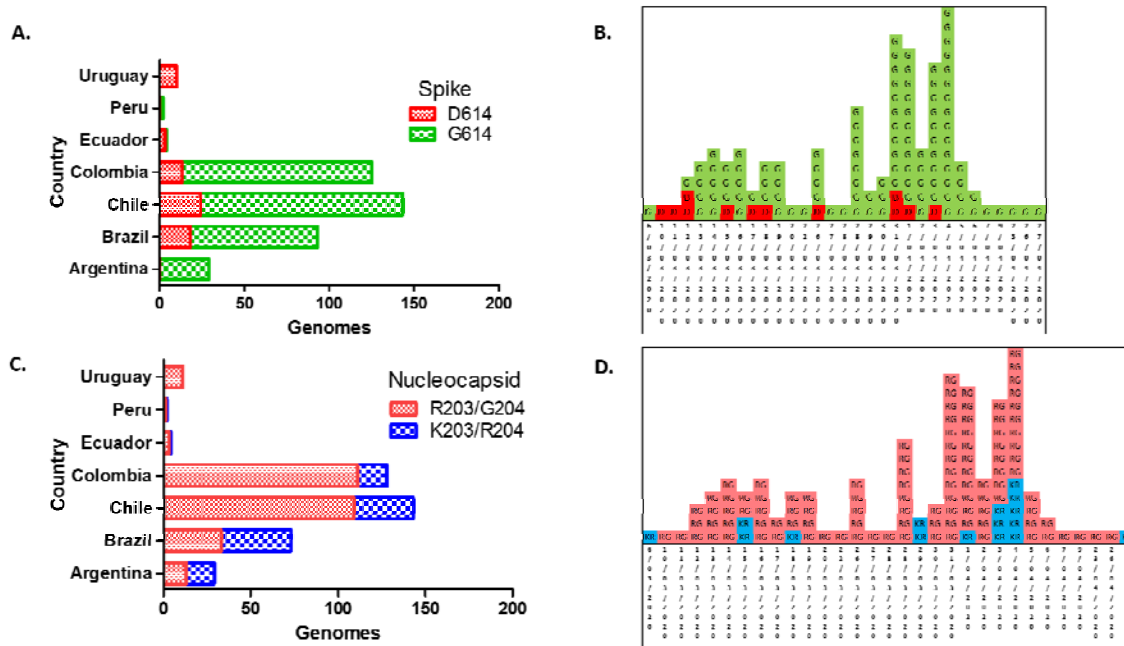


432 **Figures and Tables**

433 **Table 1.** Nucleotide substitutions and amino acid changes in the Spike and Nucleocapsid  
 434 proteins of 43 genomes of SARS-CoV-2 from Colombia reported in this study. Location of  
 435 the substitutions was estimated base on the reference genomic sequence NC\_045512.2.  
 436 Left panel: nucleotides; right panel: amino acids. ? : Uncovered nucleotide and amino acids  
 437 during complete genome sequencing.

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	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	G	R	R	G	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	

439



440

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

454

455

456

457

458

459

460

461

462

463

464

465

466



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	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	C	G	D	R	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	.	.	.	.	.	.	.	.	.	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	.	.	.	.	.	.	.	.	.	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	.	.	.	.	.	.	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	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
EPI_ISL_417924	hCoV-19/Colombia/Antioquia/79256/2020	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	.	.	.	C
EPI_ISL_456119	hCoV-19/Colombia/Medellin/80825/2020	.	.	.	.	.	.	.	.	.	.	.	.	.	T	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	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	I
EPI_ISL_456127	hCoV-19/Colombia/Armenia/79911/2020	.	.	.	.	.	T	.	.	.	.	.	.	.	T	.	.	.	.	I