

1 **Molecular analysis of several in-house rRT-PCR protocols for SARS-CoV-2 detection in the context of**
2 **genetic variability of the virus in Colombia.**

3 Diego A. Álvarez-Díaz^{1,2}, Carlos Franco-Muñoz^{1,3}, Katherine Laiton-Donato¹, José A. Usme-Ciro^{1,4}, Nicolás
4 D. Franco-Sierra⁵, Astrid C. Flórez⁶, Sergio Gómez Rangel⁸, Luz Dary Rodríguez⁸, Juliana Barbosa⁸, Erika
5 Ospitia⁸, Diana Marcela Walteros⁷, Martha Lucia Ospina Martínez⁹, Marcela Mercado-Reyes.^{1,2,10}

6 1. Unidad de Secuenciación y Genómica. Dirección de Investigación en Salud Pública, Instituto Nacional
7 de Salud.

8 2. Grupo de Salud Materna y Perinatal. Dirección de Investigación en Salud Pública, Instituto Nacional de
9 Salud.

10 3. Grupo de Parasitología. Dirección de Investigación en Salud Pública, Instituto Nacional de Salud.

11 4. Centro de Investigación en Salud para el Trópico-CIST, Universidad Cooperativa de Colombia, Santa
12 Marta, Colombia.

13 5. Instituto de Investigación de Recursos Biológicos Alexander von Humboldt.

14 6. Dirección de Redes en Salud Pública, Instituto Nacional de Salud.

15 7. Dirección de Vigilancia en Salud Pública, Instituto Nacional de Salud.

16 8. Grupo de Virología, Dirección de Redes en Salud Pública, Instituto Nacional de Salud.

17 9. Dirección General, Instituto Nacional de Salud.

18 10. Dirección de Investigación en Salud Pública, Instituto Nacional de Salud.

19

20 **SUMMARY**

21 The COVID-19 pandemic caused by SARS-CoV-2 is a public health problem unprecedented in the recent
22 history of humanity. Different in-house real-time RT-PCR (rRT-PCR) methods for SARS-CoV-2 diagnosis
23 and the appearance of genomes with mutations in primer regions have been reported. Hence, whole-
24 genome data from locally-circulating SARS-CoV-2 strains contribute to the knowledge of its global
25 variability and the development and fine tuning of diagnostic protocols. To describe the genetic
26 variability of Colombian SARS-CoV-2 genomes in hybridization regions of oligonucleotides of the main in-
27 house methods for SARS-CoV-2 detection, RNA samples with confirmed SARS-CoV-2 molecular diagnosis
28 were processed through next-generation sequencing. Primers/probes sequences from 13 target regions
29 for SARS-CoV-2 detection suggested by 7 institutions and consolidated by WHO during the early stage of
30 the pandemic were aligned with Muscle tool to assess the genetic variability potentially affecting their
31 performance. Finally, the corresponding codon positions at the 3' end of each primer, the open reading
32 frame inspection was identified for each gene/protein product. Complete SARS-CoV-2 genomes were
33 obtained from 30 COVID-19 cases, representative of the current epidemiology in the country.
34 Mismatches between at least one Colombian sequence and five oligonucleotides targeting the RdRP and
35 N genes were observed. The 3' end of 4 primers aligned to the third codon position, showed high risk of
36 nucleotide substitution and potential mismatches at this critical position. Genetic variability was
37 detected in Colombian SARS-CoV-2 sequences in some of the primer/probe regions for in-house rRT-PCR

38 diagnostic tests available at WHO COVID-19 technical guidelines; its impact on the performance and
39 rates of false-negative results should be experimentally evaluated. The genomic surveillance of SARS-
40 CoV-2 is highly recommended for the early identification of mutations in critical regions and to issue
41 recommendations on specific diagnostic tests to ensure the coverage of locally-circulating genetic
42 variants.

43 **KEYWORDS**

44 SARS-CoV-2, COVID-19 diagnostic testing, Next-Generation Sequencing, RT-PCR, Genetic diversity

45 **HIGHLIGHTS**

- 46 • Colombian SARS-CoV-2 sequences displayed genetic variability in some target regions used for
47 COVID-19 diagnosis.
- 48 • Mismatches in critical primer regions could impact their performance and the rate of false-
49 negative results.

50

51 **INTRODUCTION**

52 In late December 2019 in Wuhan city (China), a new coronavirus called SARS-CoV-2 (initially nCoV-2019)
53 caused the outbreak of a respiratory disease known as the infectious disease due to the new
54 coronavirus (COVID-19) (Zhou et al., 2020; Zhu et al., 2020). Soon after learning about the potential for
55 transmission of this virus in the context of a globalized world, countries took swift and extreme
56 measures such as border closings, rigorous follow-up of contacts, and mandatory preventive isolation. In
57 Colombia, the first case of COVID-19 was identified on March 6, 2020, shortly before the World Health
58 Organization (WHO) declared COVID-19 as a pandemic after it had spread in 114 countries in all
59 continents and having claimed the lives of 4,291 people (WHO, 2020a).

60 The first case of COVID-19 in Colombia was imported from Italy, a country that had the most alarming
61 epidemic peak in Europe at that time. Shortly afterwards, cases of COVID-19 were diagnosed in travelers
62 from other origins, as well as in multiple of their contacts. On April 20, 2020, community transmission
63 cases already exceeded 10% of the total cases registered in the country, which is why the transition to
64 the mitigation phase was declared. Until May 18, 2020, 4,629,000 cases and 297,380 deaths have been
65 reported globally, and in Colombia, 16,295 cases and 592 deaths have been reported (Dong, Du, &
66 Gardner, 2020).

67 SARS-CoV-2 is a betacoronavirus with a positive polarity single-stranded RNA genome of approximately
68 30 kb. This new coronavirus shares a global identity of 96.2% with the bat coronavirus RaTG13 (Zhou et
69 al., 2020), 91.2% with the Malay Pangolin coronavirus isolate, Pangolin-CoV (Zhang, Wu, & Zhang, 2020),
70 and even 97.5% with RmYN02 derived from bat when the ORF1ab gene is exclusively analyzed (Wang,
71 Pipes, & Nielsen, 2020). Its recent origin is enigmatic since there is a high similarity between the amino
72 acid sequence in the receptor-binding domain (RBD domain) of subunit 1 (S1) of the Spike protein of
73 SARS-CoV-2 and that of Pangolin-CoV , but the latter lacks the polybasic furin processing site, exclusive
74 to SARS-CoV-2 (Andersen, Rambaut, Lipkin, Holmes, & Garry, 2020). Accumulated evidence suggests a
75 zoonotic origin of the virus as a result of recombination with a yet unidentified coronavirus or
76 convergent evolution driven by natural selection to optimize interaction with the human ACE2 cell
77 receptor (Wang et al., 2020; WHO, 2020b). After the publication of the complete SARS-CoV-2 genome in

78 genomic data repositories such as NCBI (MN908947.3) and GISAID in mid-January 2020, health agencies
79 and researchers from different countries quickly developed SARS-CoV-2 screening tests based on real-
80 time RT-PCR (rRT-PCR) that amplify different SARS-CoV-2 gene regions. Hundreds of commercial kits are
81 under development and many of them have been licensed for emergency use
82 (<https://www.finddx.org/covid-19/pipeline/>). Various in-house protocols were also developed and
83 published on the website of the World Health Organization (WHO) for informational purposes without
84 implying endorsement, preference or validation by this entity (WHO, 2020c). However, most of these
85 protocols were published during January 2020 when only 230 virus sequences were available that
86 circulated exclusively in Asia and Europe, except for a few cases in the United States and Canada
87 (Holshue et al., 2020).

88 Since then, refinements of these protocols are not known in the context of nearly 30,000 sequences
89 reported on May 18, 2020, worldwide, including Latin America, which provide a more complete
90 perspective of the accumulated genetic variability and sequence particularities of viruses circulating in
91 specific regions that could affect the efficiency and sensitivity of the rRT-PCR protocols currently shared
92 by WHO. The mutation rate of SARS-CoV-2 as a virus with an RNA genome is higher than that of viruses
93 with a DNA genome (Tang et al., 2020), with an estimated mean evolutionary rate of 2.24×10^{-3}
94 substitutions/site/year (Li, Li, Cui, & Wu, 2020); therefore, changes in the sequence could occur over
95 time that compromise the operational performance of diagnostic tests (PAHO, 2020). The objective of
96 this study was to describe the genetic variability of Colombian SARS-CoV-2 genomes in hybridization
97 regions of oligonucleotides of the main in-house methods for SARS-CoV-2 detection.

98

99 **MATERIALS AND METHODS**

100 **Patients and samples**

101 Nasopharyngeal swab samples from patients with suspected SARS-CoV-2 infection were received at the
102 Instituto Nacional de Salud (INS) as part of the virological surveillance of COVID-19 from 11 Colombian
103 departments and the capital district. According to the national law 9/1979, decrees 786/1990 and
104 2323/2006, the INS is the reference lab and health authority of the national network of laboratories and
105 in cases of public health emergency or those in which scientific research for public health purposes as
106 required, the INS may use the biological material for research purposes, without informed consent,
107 which includes the anonymous disclosure of results. This study was performed in accordance with the
108 ethical standards noted in the 1964 Declaration of Helsinki and its later amendments. The information
109 used for this study comes from secondary sources of data that were previously anonymized and do not
110 represent a risk to the community.

111 **RNA extraction and real-time rRT-PCR**

112 Viral RNA was obtained using the automated MagNA Pure LC nucleic acid extraction system (Roche
113 Diagnostics GmbH, Mannheim, Germany) and viral RNA detection was performed by rRT-PCR using the
114 SuperScript III Platinum One-Step Quantitative RT- kit. PCR (Thermo Fisher Scientific, Waltham, MA,
115 USA), following the Charité-Berlin protocol (Victor M. Corman et al., 2020) for the amplification of the
116 SARS-CoV-2 E and RdRp genes.

117 **Next generation sequencing**

118 The complete SARS-CoV-2 genome sequence of 30 patients was obtained through NGS, ten genomes
119 with Oxford Nanopore (Oxford Nanopore Technologies, Oxford, UK) and 20 genomes with Illumina
120 MiSeq (Illumina, San Diego, CA, USA) technologies, following the artic.network "nCoV-2019 sequencing
121 protocol" (Quick, 2020). In both strategies, SARS-CoV-2 specific oligonucleotides were used for the
122 generation of amplicons by means of a Q5[®] high fidelity DNA polymerase (New England Biolabs Inc.,
123 UK), in order to avoid the introduction of artificial mutations. The genomes were assembled by mapping
124 to the reference genome (NC_045512.2) using the BWA (Li et al., 2020) and BBmap (brian-jgi, 2020)
125 software to generate a consensus genome by the two assembly tools.

126 **Genetic diversity analysis**

127 The Colombian genomes and oligonucleotides from of the in-house protocols were aligned with the
128 Muscle tool (Edgar, 2004) using the MEGA X software (Kumar, Stecher, Li, Knyaz, & Tamura, 2018).
129 Substitutions matrices of the Colombian genomes respect to the reference genome (NC_045512) at the
130 nucleotide and amino acid levels were generated for the 13 rRT-PCR protocols published at the WHO
131 website (WHO, 2020c), which several countries have established as their preferred diagnostic protocol
132 for SARS-CoV-2.

133 **Oligonucleotides analysis**

134 Thermodynamic features (priming T_m and ΔG at the variable sites, mispriming and ΔG , hairpin ΔG , and
135 primer dimer ΔG) and the codon position at the 3' end for oligonucleotides with conflicting sites and
136 optimized oligonucleotides were evaluated into the PrimerSelect module of the LaserGene v8.1 suite
137 (DNASTAR Inc. Madison, WI, USA).

138

139 **RESULTS**

140 **Several target regions for SARS-CoV-2 molecular detection using in-house protocols.**

141 A total of 39 primer and probe sequences from the main in-house rRT-PCR protocols for SARS-CoV-2
142 detection published at WHO website were aligned to the reference sequence derived from the first
143 confirmed case at Wuhan, Hubei province, China, and named Wuhan-1 strain (GenBank Accession
144 Number: NC_045512.2). The protocols targeted 13 different genome regions inside the Orf1ab (Nsp9,
145 Nsp10, Nsp11, RdRp, ExoN), E (Envelope) and N (Nucleocapsid) genes with 61,5% (8/13) of the assays
146 targeting the N gene (Figure 1).

147

148 **Point mutations in Colombian SARS-CoV-2 genomes at the target hybridization sequences of some in- 149 house protocols for SARS-CoV-2 detection.**

150 A total 30 whole genome SARS-CoV-2 sequences from Colombia were included in the previously
151 obtained alignment for genetic variability analysis. These sequences derived from cases with
152 confirmation dates from March 6th-24th. 2020 were remitted from 11 departments and the capital
153 district (Antioquia, Bogotá D.C., Bolivar, Caldas, Cauca, Magdalena, Norte de Santander, Quindío,
154 Risaralda, Santander, Tolima and Valle del Cauca) through the National Public Health Laboratories
155 Network to the Colombian National Institute of Health for diagnostic confirmation.

156 From the alignment of the 39 oligonucleotides to the reference and Colombian SARS-CoV-2 sequences, 5
157 showed mismatches with at least one Colombian sequence (Table 1). The conflicting sites in the
158 primer/probe sequences were due to 1) a mismatch between the oligonucleotide and the reference and
159 Colombian sequences or 2) a mismatch between the oligonucleotide and one or more Colombian
160 sequences.

161 The Hong Kong(2020) HKU-NP probe (Chu et al., 2020) (Figure 2) was extremely different to all the
162 reference and Colombian SARS-CoV-2 by four nucleotide sites. This probe also presents the formation of
163 a highly stable hairpin and self-dimer structures (Table S1). The Corman Berlin(2020) RdRP SARSr-F2
164 (Victor M. Corman et al., 2020) (Figure 3) region was variable in one Colombian SARS-CoV-2 sequence
165 from the department of Valle del Cauca (ID: 79943), involving a critical site at the 3' pentamer. This site
166 is supposed to prevent the correct hybridization of the 3' end of the primer, leading to inefficient or
167 unsuccessful extension by the DNA polymerase. The RdRP SARSr-R1 primer (Victor M. Corman et al.,
168 2020) showed a degenerate site which does not comprise the nucleotide found in the reference and
169 Colombian SARS-CoV-2 sequences. However, this mismatch was located at an internal site of the primer,
170 only partially affecting the thermodynamic profile of the primer-target hybridization. The Zhu 2020 CDC-
171 China Set I Probe(ORF1ab) (Zhu et al., 2020) (Figure 4) was found to be almost completely
172 complementary with the Colombian SARS-CoV-2 sequences, except for a viral sequence obtained from a
173 human case in the department of Antioquia (ID: 79253), bearing a single substitution at the seventh
174 probe position, without considerable effect on the thermodynamic features for probe-target
175 hybridization (Table S1). The Zhu 2020 CDC-China Set II Fw(N) primer (Figure 5) hybridization was found
176 to be critically affected by the accumulated genetic diversity of the Colombian SARS-CoV-2 strains. At
177 the 5' end of the primer a triple-nucleotide substitution GGG→AAC in three sequences from Bogotá and
178 Valle del Cauca affected the T_m and ΔG . At the 3' region, two sequences from Quindío displayed a
179 substitution affecting the 3' pentamer stability.

180 The primers and probes sets proposed by Pasteur 2020 (France)nCoV IP2 (Fig. S1), Hong Kong (2020)
181 HKU-ORF1b-nsp14 (Fig. S2), Corman Berlin(2020) E Sarbeco (Fig. S3), CDC 2020(USA) 2019-nCoV N1 (Fig.
182 S4), DMC-MH 2020(Thailand)WH-NIC N (Fig. S4), CDC 2020(USA) 2019-nCoV N3 (Fig. S5), Corman
183 Berlin(2020) N Sarbeco (Fig. S5), Nao 2020(Japan)NIID 2019-nCoV N (Suppl. Fig. 6) and CDC 2020(USA)
184 2019-nCoV N2 (Fig. S6) showed correspondence with the reference sequence and with all the
185 accumulated genetic variability in available sequences of Colombian strains of SARS-CoV-2.

186

187 **The third codon position was found to align with the 3' end of some primers with intended use for** 188 **molecular detection of SARS-CoV-2**

189 Codon positions in coding regions are differentially susceptible to nucleotide substitution, being the
190 third codon position associated with a higher substitution rate. On the other hand, the perfect matching
191 of the last nucleotide at the 3' end of every forward and reverse primer is critical for DNA polymerase-
192 based extension during PCR amplification (Staheli, Ryan, Bruce, Boyce, & Rose, 2009). Therefore, the
193 rational design of primer to be used with rapidly evolving RNA viruses should have the requisite of
194 avoiding third and sometimes first codon positions.

195 Codon position of the last nucleotide at the 3' end of every analyzed primer was identified according to
196 the corresponding open-reading frame. The 3' end of 11 primers corresponded to the first codon

197 position, another 11 primers had 3' ends located at the second codon position and the 3' end of 4
198 primers aligned with the third codon position (Table 1). One of the primers (Nao 2020(Japan)NIID 2019-
199 nCoV N F2) had the 3' end aligned with the first codon position of the codon CUA which encodes for
200 alanine. This codon allows a substitution at the first codon position (C to U) to be synonymous and
201 therefore may escape to the selection pressure at the protein level.

202 The 3' end of the primers CDC 2020(USA) 2019-nCoV N1-F, CDC 2020(USA) 2019-nCoV N3-F, CDC
203 2020(USA) 2019-nCoV N3-R and Hong Kong(2020) HKU-NR aligned with the third codon position of their
204 corresponding ORFs. Two of these primers were the forward and reverse primers proposed to amplify a
205 specific target (CDC 2020(USA) 2019-nCoV N3) at the N gene making this protocol very susceptible to
206 false negative results as the viral genetic variability increases.

207

208 **Discussion**

209 Next Generation Sequencing (NGS) technologies are a valuable tool in determining the whole genome of
210 microorganisms impacting on public health, they have accelerated our ability to understand important
211 factors in the biology of infectious diseases such as the identification of determinants of virulence, drug-
212 resistance associated substitutions, vaccine design and genomic epidemiology, a powerful approach to
213 integrate epidemiologic and genetic information in reconstructing transmission patterns and infectious
214 disease dynamics (Gwinn, MacCannell, & Armstrong, 2019). NGS has also enabled pathogen discovery in
215 a timely manner compared to traditional methods, being essential to the virus taxonomic classification
216 during the beginnings of COVID-19 pandemics (Zhou et al., 2020; Zhu et al., 2020). This robust molecular
217 information is also the raw material for the development and refinement of molecular and serological
218 diagnostic methods (Wang et al., 2020) and all sequence-based designs are susceptible to improvement
219 as the virus disseminates and its genetic variability increases.

220 Mutation is the fundamental source of genetic variation and RNA viruses are particularly susceptible to
221 have high mutation rates during the genome replication (Sanjuan, Nebot, Chirico, Mansky, & Belshaw,
222 2010). Therefore, RNA viruses display high substitution rates when analyzed through time. Although
223 sequence identity between the analyzed genomes and the Wuhan reference strain (NC_045512) was
224 around 99.9%, the estimate substitution rate for SARS-CoV-2 is in the range of $1.67-4.67 \times 10^{-3}$ /site/year
225 (Tang et al., 2020). Therefore, it is evident that the genomic data of this study allow the refinement or
226 even the design of more precise and efficient protocols for the molecular detection of the genetic
227 variants of SARS-CoV-2 circulating in Colombia.

228 An enormous effort of every official and private laboratory around the world has led to the availability of
229 hundreds of commercial kits and in-house protocols for the molecular detection and COVID-19
230 diagnostic confirmation. While this effort is welcome and some recommendations have been provided
231 by regional agencies, the availability of all these methods imply the need for an evidence-based criterion
232 for decision-making about the best option at country level. In this study, based on NGS data of complete
233 SARS-CoV-2 genomes from Colombian cases, timely evidence is provided on the genetic variability of
234 representative strains of the circulating viruses in the country that should be considered for an
235 evidence-based improvement of the routine diagnostic tests.

236 The majority of the evaluated in-house molecular assays displayed genetic stability at the
237 primers/probes target regions. However, some these oligonucleotides displayed mismatches that were
238 considered of minor or major importance for the test performance based on the *in silico* analysis. Some
239 of the analyzed primers/probes displayed mismatches when compared to target sequences of some
240 Colombian SARS-CoV-2 strains. Polymorphic sites in the complete SARS-CoV-2 genomes obtained in this
241 study, were supported by sequencing depths between 123 and 390X sufficient to identify these
242 sequence variations unequivocally. The analyzed genome sequences corresponded to the early
243 introduction and dispersion of the virus in the country. It is expected that some of the identified
244 nucleotide substitutions remain stable in subsequent transmission chains of the corresponding clusters
245 of cases, as demonstrated in the present study as character states shared by two or more sequences.

246 Prior to the arrival of SARS-CoV-2 in Colombia, the Charité-Berlin protocol described by Corman and
247 collaborators was established as a routine in detection (Victor M. Corman et al., 2020) following the
248 recommendation of the Pan American Health Organization (PAHO). Their three molecular targets (RdRp,
249 E and N genes) were analyzed and RdRp displayed some conflicting nucleotide sites. The antisense
250 primer RdRp SARSr R1 presented a degenerate site at the nucleotide position 15,519 equivalent to
251 nucleotides G or C (S) while the reference and all Colombian sequences displayed T at that position.
252 However, this mismatch was expected not to affect seriously the overall stability of its hybridization.
253 There was an important finding in the sense primer RdRp_SARSr_F1 which displayed a mismatch at
254 position 15,451 when compared to the hCoV/Colombia/Valle_del_Cauca/79943/2020 genome. This
255 mismatch located at the second base of the 3' end of the primer is expected to severely affect the
256 primer hybridization and subsequent DNA polymerase-mediated extension. In the present study, we
257 propose to include a degenerate site (R) as follows (RdRP_SARSr-F2 Mod 5'-
258 GTGAAATGGTCATGTGTGGCRG-3'). Thus, covering the genetic variability and improving the
259 thermodynamic stability with this genetic variant to avoid false-negative results (Table S2). The
260 nucleotide identity of the Colombian genomes with respect to the Corman Berlin (2020) E Sarbeco and N
261 Sarbeco designs was 100%. These results supported the PAHO recommendation of diagnostic
262 confirmation based on the single E gene in the context of sustained transmission at high levels in this
263 region where other coronavirus species of the subgenus *Sarbecovirus* are expected to be absent (PAHO,
264 2020).

265 Another critical characteristic of the primers/probes of the current in-house protocols listed by WHO
266 and assessed in the present study was the codon position at the 3' end of the sense and antisense
267 primers in coding regions. The major and minor susceptibility to nucleotide substitutions for the third
268 and second codon positions, respectively, is widely known. As any change at the 3' end can affect the
269 primer hybridization, it is highly recommended that this position aligns with the second (preferable) or
270 first codon position. Some primers and indeed primer sets were found to be coincident with the third
271 codon position. Despite not finding substitutions at these sites in the Colombian strains, that position is
272 expected to be unstable through time.

273 Genomic data available from this study allowed the *in silico* evaluation/refinement of the protocols for
274 molecular detection of SARS-CoV-2 circulating in Colombia. It is highly recommended to establish
275 routine molecular surveillance of the virus in order to determine the real impact of every mutation in
276 the diagnostic test's performance (PAHO, 2020).

277 Finally, the implementation of molecular tests at country-level should be supported by the estimation of
278 the analytical sensitivity (Limit of detection [LOD] in copies/reaction), specificity (Victor M. Corman et
279 al., 2020), and the accumulated genetic variability should be tested during the implementation of a
280 molecular detection protocol including the clinical sensitivity in different biofluids (V. M. Corman et al.,
281 2016), reproducibility, repeatability, inter-operator, inter-instrument, inter-site and inter-batch assays
282 (Hu et al., 2019).

283

284 CONCLUSIONS

285 Detection of SARS-CoV-2 viral RNA using nucleic acid amplification techniques such as rRT-PCR continues
286 to be the gold standard for the diagnosis of COVID-19 (WHO, 2020d). However, all sequence-based
287 methods are susceptible to nucleotide substitution affecting the oligonucleotide hybridization efficiency
288 and resulting in false negatives. Some of the in-house protocols analyzed in the present study require an
289 experimental evaluation of their performance in the context of virus genetic variability. The genomic
290 data of this study allow the refinement or even the design of more precise and efficient protocols for the
291 molecular detection of the genetic variants of SARS-CoV-2 circulating in Colombia. However, more NGS
292 data from Colombian SARS-CoV-2 will be determinant to a better comprehension of the impact of
293 genetic variability on specific molecular assays of routine use as the virus evolves.

294 ACKNOWLEDGMENTS

295 For the joint work with the professionals involved in facing the SARS-CoV-2 pandemic from different
296 aspects and to the Directorates of Public Health Research, Public Health Networks and Surveillance of
297 the National Institute of Health.

298 CONFLICTS OF INTERESTS

299 The authors declare that there is no conflict of interest in the manuscript.

300 FUNDING

301 This study was funded by the National Institute of Health, in Bogotá D.C., Colombia.

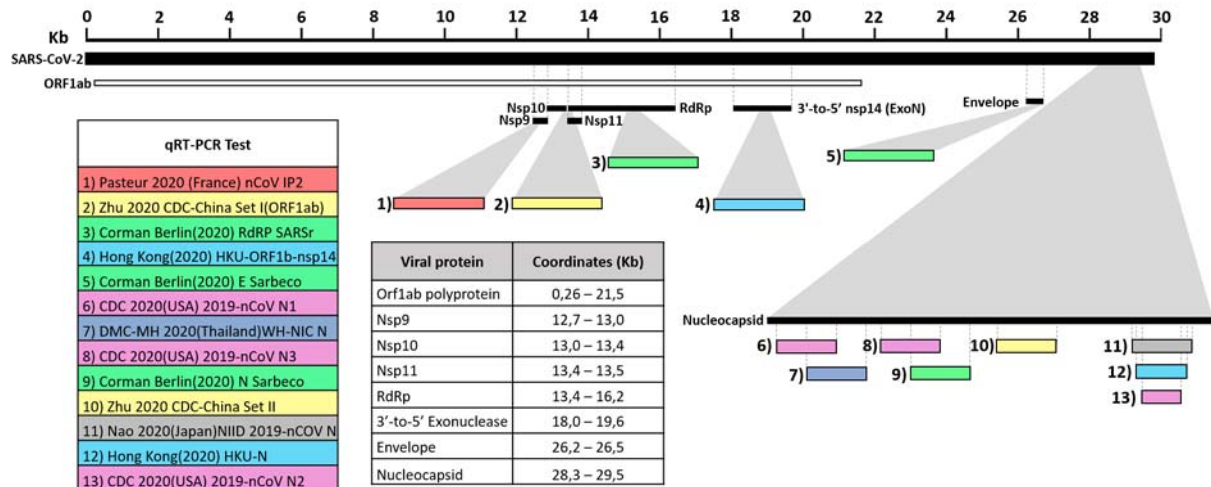
302 REFERENCES

- 303 Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C., & Garry, R. F. (2020). The proximal origin of
304 SARS-CoV-2. *Nat Med*, 26(4), 450-452. doi:10.1038/s41591-020-0820-9
- 305 brian-jgi. (2020). BBMap short read aligner, and other bioinformatic tools. Retrieved from
306 <https://sourceforge.net/projects/bbmap/>
- 307 Corman, V. M., Landt, O., Kaiser, M., Molenkamp, R., Meijer, A., Chu, D. K., . . . Drosten, C. (2020).
308 Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro surveillance :
309 bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin*,
310 25(3), 2000045. doi:10.2807/1560-7917.ES.2020.25.3.2000045
- 311 Corman, V. M., Rasche, A., Baronti, C., Aldabbagh, S., Cadar, D., Reusken, C. B., . . . Drexler, J. F. (2016).
312 Assay optimization for molecular detection of Zika virus. *Bull World Health Organ*, 94(12), 880-
313 892. doi:10.2471/BLT.16.175950

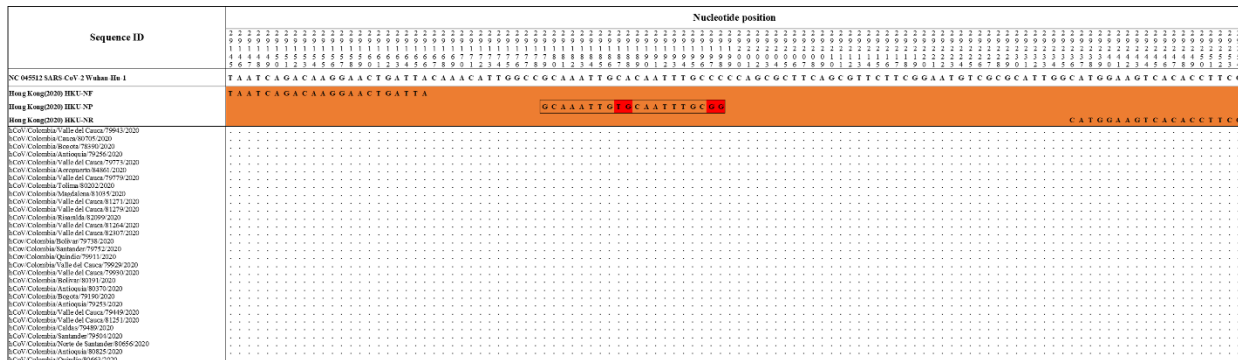
- 314 Chu, D. K. W., Pan, Y., Cheng, S. M. S., Hui, K. P. Y., Krishnan, P., Liu, Y., . . . Poon, L. L. M. (2020).
315 Molecular Diagnosis of a Novel Coronavirus (2019-nCoV) Causing an Outbreak of Pneumonia.
316 *Clin Chem*, 66(4), 549-555. doi:10.1093/clinchem/hvaa029
- 317 Dong, E., Du, H., & Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real
318 time. *Lancet Infect Dis*, 20(5), 533-534. doi:10.1016/S1473-3099(20)30120-1
- 319 Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput.
320 *Nucleic Acids Res*, 32(5), 1792-1797. doi:10.1093/nar/gkh340
- 321 Gwinn, M., MacCannell, D., & Armstrong, G. L. (2019). Next-Generation Sequencing of Infectious
322 Pathogens. *JAMA*, 321(9), 893-894. doi:10.1001/jama.2018.21669
- 323 Holshue, M. L., DeBolt, C., Lindquist, S., Lofy, K. H., Wiesman, J., Bruce, H., . . . Washington State -nCoV, V.
324 C. I. T. (2020). First Case of 2019 Novel Coronavirus in the United States. *N Engl J Med*, 382(10),
325 929-936. doi:10.1056/NEJMoa2001191
- 326 Hu, X. M., Xu, J. X., Jiang, L. X., Deng, L. R., Gu, Z. M., Xie, X. Y., . . . Zhong, T. Y. (2019). Design and
327 Evaluation of a Novel Multiplex Real-Time PCR Melting Curve Assay for the Simultaneous
328 Detection of Nine Sexually Transmitted Disease Pathogens in Genitourinary Secretions. *Front*
329 *Cell Infect Microbiol*, 9, 382. doi:10.3389/fcimb.2019.00382
- 330 Kumar, S., Stecher, G., Li, M., Nknyaz, C., & Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics
331 Analysis across Computing Platforms. *Molecular biology and evolution*, 35(6), 1547-1549.
332 doi:10.1093/molbev/msy096
- 333 Li, J., Li, Z., Cui, X., & Wu, C. (2020). Bayesian phylodynamic inference on the temporal evolution and
334 global transmission of SARS-CoV-2. *J Infect*. doi:10.1016/j.jinf.2020.04.016
- 335 PAHO. (2020). Directrices de Laboratorio para la Detección y el Diagnóstico de la Infección con el Virus
336 COVID-19: Pan American Health Organization.
- 337 Quick, J. (2020). nCoV-2019 sequencing protocol. Retrieved from <https://www.protocols.io/view/ncov-2019-sequencing-protocol-bbmuik6w>
- 338
- 339 Sanjuan, R., Nebot, M. R., Chirico, N., Mansky, L. M., & Belshaw, R. (2010). Viral mutation rates. *J Virol*,
340 84(19), 9733-9748. doi:10.1128/JVI.00694-10
- 341 Staheli, J. P., Ryan, J. T., Bruce, A. G., Boyce, R., & Rose, T. M. (2009). Consensus-degenerate hybrid
342 oligonucleotide primers (CODEHOPs) for the detection of novel viruses in non-human primates.
343 *Methods*, 49(1), 32-41. doi:10.1016/j.ymeth.2009.05.011
- 344 Tang, X., Wu, C., Li, X., Song, Y., Yao, X., Wu, X., . . . Lu, J. (2020). On the origin and continuing evolution
345 of SARS-CoV-2. *National Science Review*. doi:10.1093/nsr/nwaa036
- 346 Wang, H., Pipes, L., & Nielsen, R. (2020). Synonymous mutations and the molecular evolution of SARS-
347 Cov-2 origins. *bioRxiv*, 2020.2004.2020.052019. doi:10.1101/2020.04.20.052019
- 348 WHO. (2020a). WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March
349 2020. Retrieved from [https://www.who.int/dg/speeches/detail/who-director-general-s-
350 opening-remarks-at-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)
- 351 WHO. (2020b). *Coronavirus disease 2019 (COVID-19)*. Retrieved from
352 [https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200423-sitrep-94-
353 covid-19.pdf?sfvrsn=b8304bf0_4](https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200423-sitrep-94-covid-19.pdf?sfvrsn=b8304bf0_4)
- 354 WHO. (2020c). Novel Coronavirus (2019-nCoV) technical guidance: Laboratory testing for 2019-nCoV in
355 humans. Retrieved from [https://www.who.int/emergencies/diseases/novel-coronavirus-
356 2019/technical-guidance/laboratory-guidance](https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/laboratory-guidance)
- 357 WHO. (2020d). *Coronavirus disease (COVID-19)* (Situation Report– 111). Retrieved from
358 [https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200510covid-19-
359 sitrep-111.pdf?sfvrsn=1896976f_2](https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200510covid-19-sitrep-111.pdf?sfvrsn=1896976f_2)
- 360 Zhang, T., Wu, Q., & Zhang, Z. (2020). Probable Pangolin Origin of SARS-CoV-2 Associated with the
361 COVID-19 Outbreak. *Curr Biol*, 30(7), 1346-1351 e1342. doi:10.1016/j.cub.2020.03.022

362 Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., . . . Shi, Z.-L. (2020). Discovery of a novel
363 coronavirus associated with the recent pneumonia outbreak in humans and its potential bat
364 origin. *bioRxiv*, 2020.2001.2022.914952. doi:10.1101/2020.01.22.914952
365 Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., . . . Research, T. (2020). A Novel Coronavirus from
366 Patients with Pneumonia in China, 2019. *N Engl J Med*. doi:10.1056/NEJMoa2001017
367

368 **Figure Legends**



369
 370 **Figure 1. Target hybridization regions of the primers/probes employed by the principal in-house**
 371 **protocols for molecular detection of SARS-CoV-2.** The protocols targeted 13 different genome regions
 372 inside the Orf1ab (Nsp9, Nsp10, Nsp11, RdRp, ExoN), E (Envelope) y N (Nucleocapsid) genes. Target
 373 hybridization regions of the primers / probes employed by the principal in-house protocols for molecular
 374 detection of SARS-CoV-2. The different genes and protein products, as well as the coordinates in
 375 kilobases (Kb) of the genes and protein products to which they are directed were estimated according to
 376 the SARS-CoV-2 reference genome available at GenBank (NC_045512.2).



377
 378 **Figure 2. Genetic diversity at the target region of the Hong_Kong (2020)_HKU-NP assay.** Hong_Kong
 379 (2020) _HKU-NP probe displayed mismatches at positions 29187-88 and 29197-98 (highlighted in red),
 380 with all the Colombian sequences and the RefSeq displaying CA and CC, respectively. Genomic positions
 381 were estimated according to the SARS-CoV-2 reference genome available at GenBank (NC_045512.2).

382

383

384

Sequence ID	Nucleotide position		
	1	2	3
NC_045512 SARS-CoV-2 Wuhan-Hu-1	GTGAAATGGTCAATGTTGGCGG...	CCAGGTGGAACTCATCAGGAGATGC...	TATGCTAAATAGTGTITTTAAACATTG
Corman_Berlin(2020)_RdRP_SARs-F2	GTGARATGGTCATGTTGGCGG	CAGGTGGAACTCATCAGGAGATGC	TATGCTAAATAGTGTITTTAAACATTG
Corman_Berlin(2020)_RdRP_SARs-P2		CAGGTGGAACTCATCAGGAGATGC	TATGCTAAATAGTGTITTTAAACATTG
Corman_Berlin(2020)_RdRP_SARs-R1			TATGCTAAATAGTGTITTTAAACATTG
hCoV/Colombia/Valle del Cauca/79943/2020			
hCoV/Colombia/Cauca/80705/2020			
hCoV/Colombia/Bogota/78399/2020			
hCoV/Colombia/Antioquia/79256/2020			
hCoV/Colombia/Valle del Cauca/79773/2020			
hCoV/Colombia/Aeropuerto/84861/2020			
hCoV/Colombia/Valle del Cauca/79779/2020			
hCoV/Colombia/Tolima/80202/2020			
hCoV/Colombia/Magdalena/81035/2020			
hCoV/Colombia/Valle del Cauca/81271/2020			
hCoV/Colombia/Valle del Cauca/81279/2020			
hCoV/Colombia/Risaralda/82099/2020			
hCoV/Colombia/Valle del Cauca/81264/2020			
hCoV/Colombia/Valle del Cauca/82307/2020			
hCov/Colombia/Bolivar/79738/2020			
hCov/Colombia/Santander/79752/2020			
hCov/Colombia/Quindio/79911/2020			
hCov/Colombia/Valle del Cauca/79929/2020			
hCov/Colombia/Valle del Cauca/79929/2020			
hCov/Colombia/Bolivar/80191/2020			
hCov/Colombia/Antioquia/80370/2020			
hCov/Colombia/Bogota/79190/2020			
hCov/Colombia/Antioquia/79253/2020			
hCov/Colombia/Valle del Cauca/79449/2020			
hCov/Colombia/Valle del Cauca/81125/2020			
hCov/Colombia/Caldas/79489/2020			
hCov/Colombia/Santander/79504/2020			
hCov/Colombia/Norte de Santander/80656/2020			
hCov/Colombia/Antioquia/80825/2020			
hCov/Colombia/Quindio/80663/2020			

385

386 **Figure 3. Genetic diversity at the target region of the Corman_Berlin (2020) _RdRP assay.** The
 387 Colombian genome hCoV/Colombia/Valle_del_Cauca/79943/2020 displayed a substitution (G to A) at
 388 position 15,451 (highlighted in red), affecting the primer Corman_Berlin (2020) _RdRP_SARs-F2
 389 hybridization region. The substitution is located in the penultimate position at the 3' end of the primer,
 390 generating a mismatch that could interfere with the 3' pentamer stability. A degenerate base (S = C or G)
 391 in the primer Corman_Berlin (2020) _RdRP_SARs-R1 at position 15,519 (highlighted in red) produces a
 392 mismatch with all the Colombian genomes and the RefSeq as they have T in the sense sequence, so the
 393 degenerate base should be one that includes an A among degenerate alternatives. Genomic positions
 394 were estimated according to the SARS-CoV-2 reference genome available at GenBank (NC_045512.2).

Sequence ID	Nucleotide position		
	1	2	3
NC_045512 SARS-CoV-2 Wuhan-Hu-1	CCCTGTGGGTTTTACACTTAA...	CCGCTGCGGTTATGTGGAAAGGTTATGG...	TCAGCTGATGCAACAATCGI
Zhu_2020_CDC-China Set I(ORF1ab) fw	CCCTGTGGGTTTTACACTTAA	CAGCTGCGGTTATGTGGAAAGGTTATGG	TCAGCTGATGCAACAATCGI
Zhu_2020_CDC-Set I Probe(ORF1ab)		CAGCTGCGGTTATGTGGAAAGGTTATGG	TCAGCTGATGCAACAATCGI
Zhu_2020_CDC-China Set I(ORF1ab) Rv			TCAGCTGATGCAACAATCGI
hCoV/Colombia/Valle del Cauca/79943/2020			
hCoV/Colombia/Cauca/80705/2020			
hCoV/Colombia/Bogota/78399/2020			
hCoV/Colombia/Antioquia/79256/2020			
hCoV/Colombia/Valle del Cauca/79773/2020			
hCoV/Colombia/Aeropuerto/84861/2020			
hCoV/Colombia/Valle del Cauca/79779/2020			
hCoV/Colombia/Tolima/80202/2020			
hCoV/Colombia/Magdalena/81035/2020			
hCoV/Colombia/Valle del Cauca/81271/2020			
hCoV/Colombia/Valle del Cauca/81279/2020			
hCoV/Colombia/Risaralda/82099/2020			
hCoV/Colombia/Valle del Cauca/81264/2020			
hCoV/Colombia/Valle del Cauca/82307/2020			
hCov/Colombia/Bolivar/79738/2020			
hCov/Colombia/Santander/79752/2020			
hCov/Colombia/Quindio/79911/2020			
hCov/Colombia/Valle del Cauca/79929/2020			
hCov/Colombia/Valle del Cauca/79929/2020			
hCov/Colombia/Bolivar/80191/2020			
hCov/Colombia/Antioquia/80370/2020			
hCov/Colombia/Bogota/79190/2020			
hCov/Colombia/Antioquia/79253/2020			
hCov/Colombia/Valle del Cauca/79449/2020			
hCov/Colombia/Valle del Cauca/81125/2020			
hCov/Colombia/Caldas/79489/2020			
hCov/Colombia/Santander/79504/2020			
hCov/Colombia/Norte de Santander/80656/2020			
hCov/Colombia/Antioquia/80825/2020			
hCov/Colombia/Quindio/80663/2020			

395

396 **Figure 4. Genetic diversity at the target region of the Zhu_2020_CDC-Set I assay.** The Colombian
 397 genome hCoV/Colombia/Antioquia/79253/2020 displayed a substitution (C to T) at
 398 position 13,384 (highlighted in red) where the probe Zhu_2020_CDC-Set I Probe (ORF1ab) hybridizes. Genomic positions
 399 were estimated according to the SARS-CoV-2 reference genome available at GenBank (NC_045512.2).

400

401

	Corman Berlin(2020) RdRP SARSr-P2	14470-15494	NA	None
	Corman Berlin(2020) RdRP SARSr-R1	15530-15505	1	All
	Hong Kong(2020) HKU-ORF1b-nsp14F	18778-18797	2	None
	Hong Kong(2020) HKU-ORF1b-nsp141P	18849-18872	NA	None
	Hong Kong(2020) HKU-ORF1b-nsp14R	18909-18889	1	None
	Corman Berlin(2020) E Sarbeco F1	26269-26294	2	None
Gen E	Corman Berlin(2020) E Sarbeco P1	26362-26357	NA	None
	Corman Berlin(2020) E Sarbeco R1	26381-26360	2	None
	CDC 2020(USA) 2019-nCoV N1-F	28287-28306	3	None
	CDC 2020(USA) 2019-nCoV N1-P	28309-28332	NA	None
	CDC 2020(USA) 2019-nCoV N1-R	28356-28335	2	None
	DMC-MH 2020(Thailand)WH-NIC N-F	28320-28338	2	None
	DMC-MH 2020(Thailand)WH-NIC N-P	28341-28356	NA	None
	DMC-MH 2020(Thailand)WH-NIC N-R	28376-28358	1	None
Gen N	CDC 2020(USA) 2019-nCoV N3-F	28681-28702	3	None
	CDC 2020(USA) 2019-nCoV N3-P	28704-28727	NA	None
	CDC 2020(USA) 2019-nCoV N3-R	28752-28732	3	None
	Corman Berlin(2020) N Sarbeco F1	28706-28724	1	None
	Corman Berlin(2020) N Sarbeco P1	28753-28777	NA	None
	Corman Berlin(2020) N Sarbeco R1	28833-28814	1	None
	Zhu 2020 CDC-China Set II Fw(N)	28881-		hCoV/Colombia/Bogota/78390/2020

	28902		hCoV/Colombia/Valle del Cauca/81279/2020
		2	hCoV/Colombia/Valle del Cauca/81251/2020
			hCov/Colombia/Quindio/79911/2020
			hCoV/Colombia/Quindio/80663/2020
Zhu 2020 CDC-China Set II Probe(N)	28934-28953	NA	Ninguno
Zhu 2020 CDC-China Set II Rv(N)	28979-28958	1	Ninguno
Nao 2020(Japan)NIID 2019-nCoV N F2	29125-29144	1	Ninguno
Nao 2020(Japan)NIID 2019-nCoV N P2	29222-29241	NA	Ninguno
Nao 2020(Japan)NIID 2019-nCoV N R2	29299-29280	2	Ninguno
Hong Kong(2020) HKU-NF	29145-29166	2	Ninguno
Hong Kong(2020) HKU-NP	29179-29198	NA	Ninguno
Hong Kong(2020) HKU-NR	29254-29236	3	Ninguno
CDC 2020(USA) 2019-nCoV N2-F	29164-29183	1	Ninguno
CDC 2020(USA) 2019-nCoV N2-P	29188-29210	NA	All
CDC 2020(USA) 2019-nCoV N2-R	29230-29213	1	Ninguno

414

415

416

417

418

419

420

421

422

Supplementary material

Figure S1. Alignment of Pasteur 2020 (France)nCoV IP2 primer sets with Colombian SARS-CoV-2 genomes. . Genomic positions were estimated according to the SARS-CoV-2 reference genome available at GenBank (NC_045512.2).

Figure S2. Alignment of Hong Kong (2020) HKU-ORF1b-nsp1 primer sets with Colombian SARS-CoV-2 genomes. Genomic positions were estimated according to the SARS-CoV-2 reference genome available at GenBank (NC_045512.2).

Figure S3. Alignment of Corman Berlin(2020) E Sarbeco primer sets with Colombian SARS-CoV-2 genomes. Genomic positions were estimated according to the SARS-CoV-2 reference genome available at GenBank (NC_045512.2).

Figure S4. Alignment of CDC 2020(USA) 2019-nCoV N1 (Fig. S4) and DMC-MH 2020(Thailand)WH-NIC N primer sets with Colombian SARS-CoV-2 genomes. Genomic positions were estimated according to the SARS-CoV-2 reference genome available at GenBank (NC_045512.2).

Figure S5. Alignment of CDC 2020(USA) 2019-nCoV N3 and Corman Berlin(2020) N Sarbeco primer sets with Colombian SARS-CoV-2 genomes. Genomic positions were estimated according to the SARS-CoV-2 reference genome available at GenBank (NC_045512.2).

Figure S6. Alignment of Nao 2020(Japan)NIID 2019-nCoV N and CDC 2020(USA) 2019-nCoV N2 primer sets with Colombian SARS-CoV-2 genomes. Genomic positions were estimated according to the SARS-CoV-2 reference genome available at GenBank (NC_045512.2).

Table S1. Thermodynamic characteristics of the primers/probes showing mismatches with the reference or Colombian SARS-COV-2 sequences.

Table S2. Thermodynamic characteristics of RdRP_SARSr-F2 Mod.