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

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

Infection, Genetics and Evolution 84 (2020) 104390

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

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

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

[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] *Centro de Investigación en Salud para el Trópico–CIST, Universidad Cooperativa de Colombia, Santa Marta 470003, Colombia*
[e] *Programa Ciencias de la Biodiversidad, Instituto de Investigación de Recursos Biológicos Alexander von Humboldt, Bogotá 111311, Colombia*
[f] *Dirección de Redes en Salud Pública, Instituto Nacional de Salud, Bogotá 111321, Colombia*
[g] *Dirección de Vigilancia en Salud Pública, Instituto Nacional de Salud, Bogotá 111321, Colombia*
[h] *Grupo de Virología, Dirección de Redes en Salud Pública, Instituto Nacional de Salud, Bogotá 111321, Colombia*
[i] *Dirección General, Instituto Nacional de Salud, Bogotá 111321, Colombia*
[j] *Dirección de Investigación en Salud Pública, Instituto Nacional de Salud, Bogotá 111321, Colombia*

ARTICLE INFO

Keywords:
SARS-CoV-2
COVID-19 diagnostic testing
Next-generation sequencing
RT-PCR
Genetic diversity

SUMMARY

The COVID-19 pandemic caused by SARS-CoV-2 is a public health problem unprecedented in the recent history of humanity. Different in-house real-time RT-PCR (rRT-PCR) methods for SARS-CoV-2 diagnosis and the appearance of genomes with mutations in primer regions have been reported. Hence, whole-genome data from locally-circulating SARS-CoV-2 strains contribute to the knowledge of its global variability and the development and fine tuning of diagnostic protocols. To describe the genetic variability of Colombian SARS-CoV-2 genomes in hybridization regions of oligonucleotides of the main in-house methods for SARS-CoV-2 detection, RNA samples with confirmed SARS-CoV-2 molecular diagnosis were processed through next-generation sequencing. Primers/probes sequences from 13 target regions for SARS-CoV-2 detection suggested by 7 institutions and consolidated by WHO during the early stage of the pandemic were aligned with Muscle tool to assess the genetic variability potentially affecting their performance. Finally, the corresponding codon positions at the 3′ end of each primer, the open reading frame inspection was identified for each gene/protein product. Complete SARS-CoV-2 genomes were obtained from 30 COVID-19 cases, representative of the current epidemiology in the country. Mismatches between at least one Colombian sequence and five oligonucleotides targeting the RdRP and N genes were observed. The 3′ end of 4 primers aligned to the third codon position, showed high risk of nucleotide substitution and potential mismatches at this critical position. Genetic variability was detected in Colombian SARS-CoV-2 sequences in some of the primer/probe regions for in-house rRT-PCR diagnostic tests available at WHO COVID-19 technical guidelines; its impact on the performance and rates of false-negative results should be experimentally evaluated. The genomic surveillance of SARS-CoV-2 is highly recommended for the early identification of mutations in critical regions and to issue recommendations on specific diagnostic tests to ensure the coverage of locally-circulating genetic variants.

* Corresponding author at: Unidad de Secuenciación y Genómica, Dirección de Investigación en Salud Pública, Instituto Nacional de Salud, Bogotá 111321, Colombia.

E-mail addresses: daalvarezd@unal.edu.co (D.A. Álvarez-Díaz), cefrancom@unal.edu.co (C. Franco-Muñoz), kdlaitond@unal.edu.co (K. Laiton-Donato), jose.usmec@ucc.edu.co (J.A. Usme-Ciro), nfranco@humboldt.org.co (N.D. Franco-Sierra), aflorez@ins.gov.co (A.C. Flórez-Sánchez), sgomezr@ins.gov.co (S. Gómez-Rangel), lrodriguez@ins.gov.co (L.D. Rodríguez-Calderon), jbarbosa@ins.gov.co (J. Barbosa-Ramirez), eospitia@ins.gov.co (E. Ospitia-Baez), dwalteros@ins.gov.co (D.M. Walteros), mospina@ins.gov.co (M.L. Ospina-Martinez), mmercado@ins.gov.co (M. Mercado-Reyes).

https://doi.org/10.1016/j.meegid.2020.104390
Received 27 May 2020; Accepted 31 May 2020
Available online 04 June 2020
1567-1348/ © 2020 Elsevier B.V. All rights reserved.

1. Introduction

In late December 2019 in Wuhan city (China), a new coronavirus called SARS-CoV-2 (initially nCoV-2019) caused the outbreak of a respiratory disease known as the infectious disease due to the new coronavirus (COVID-19) (Zhou et al., 2020; Zhu et al., 2020). Soon after learning about the potential for transmission of this virus in the context of a globalized world, countries took swift and extreme measures such as border closings, rigorous follow-up of contacts, and mandatory preventive isolation. In Colombia, the first case of COVID-19 was identified on March 6, 2020, shortly before the World Health Organization (WHO) declared COVID-19 as a pandemic after it had spread in 114 countries in all continents and having claimed the lives of 4291 people (WHO, 2020a).

The first case of COVID-19 in Colombia was imported from Italy, a country that had the most alarming epidemic peak in Europe at that time. Shortly afterwards, cases of COVID-19 were diagnosed in travelers from other origins, as well as in multiple of their contacts. On April 20, 2020, community transmission cases already exceeded 10% of the total cases registered in the country, which is why the transition to the mitigation phase was declared. Until May 24, 2020, around 5,361,000 cases and 343,000 deaths have been reported globally, and in Colombia, 20,177 cases and 705 deaths have been reported (Dong et al., 2020).

SARS-CoV-2 is a betacoronavirus with a positive polarity single-stranded RNA genome of approximately 30 kb. This new coronavirus shares a global identity of 96.2% with the bat coronavirus RaTG13 (Zhou et al., 2020), 91.2% with the Malay Pangolin coronavirus isolate, Pangolin-CoV (Zhang et al., 2020), and even 97.5% with RmYN02 derived from bat when the ORF1ab gene is exclusively analyzed (Wang et al., 2020). Its recent origin is enigmatic since there is a high similarity between the amino acid sequence in the receptor-binding domain (RBD domain) of subunit 1 (S1) of the Spike protein of SARS-CoV-2 and that of Pangolin-CoV, but the latter lacks the polybasic furin processing site, exclusive to SARS-CoV-2 (Andersen et al., 2020). Accumulated evidence suggests a zoonotic origin of the virus as a result of recombination with a yet unidentified coronavirus or convergent evolution driven by natural selection to optimize interaction with the human ACE2 cell receptor (Wang et al., 2020; WHO, 2020b). After the publication of the complete SARS-CoV-2 genome in genomic data repositories such as NCBI (MN908947.3) and the Global Initiative on Sharing All Influenza Data (GISAID) in mid-January 2020, health agencies and researchers from different countries quickly developed SARS-CoV-2 screening tests based on real-time RT-PCR (rRT-PCR) that amplify different SARS-CoV-2 gene regions. Hundreds of commercial kits are under development and many of them have been licensed for emergency use (https://www.finddx.org/covid-19/pipeline/). Various in-house protocols were also developed and published on the website of the World Health Organization (WHO) for informational purposes without implying endorsement, preference or validation by this entity (WHO, 2020c). However, most of these protocols were published during January 2020 when only 230 virus sequences were available that circulated exclusively in Asia and Europe, except for a few cases in the United States and Canada (Holshue et al., 2020).

Since then, refinements of these protocols are not known in the context of more than 31,000 sequences reported on May 24, 2020, worldwide, including Latin America, which provide a more complete perspective of the accumulated genetic variability and sequence particularities of viruses circulating in specific regions that could affect the efficiency and sensitivity of the rRT-PCR protocols currently shared by WHO. The mutation rate of SARS-CoV-2 as a virus with an RNA genome is higher than that of viruses with a DNA genome (Tang et al., 2020), with an estimated mean evolutionary rate of 2.24×10^{-3} substitutions/site/year (Li et al., 2020); therefore, changes in the sequence could occur over time that compromise the operational performance of diagnostic tests (PAHO, 2020). The objective of this study was to describe the genetic variability of Colombian SARS-CoV-2 genomes in hybridization regions of oligonucleotides of the main in-house methods for SARS-CoV-2 detection.

2. Materials and methods

2.1. Ethics

According to the national law 9/1979, decrees 786/1990 and 2323/2006, the INS 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 may 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 not represent a risk to the community.

2.2. Patients and samples

Nasopharyngeal swab samples from 30 patients with suspected SARS-CoV-2 infection were received at the Instituto Nacional de Salud (INS) between March 6th–24th 2020, as part of the virological surveillance of COVID-19 from 11 Colombian departments and the capital district (Antioquia, Bogotá D.C., Bolivar, Caldas, Cauca, Magdalena, Norte de Santander, Quindío, Risaralda, Santander, Tolima and Valle del Cauca), through the National Public Health Laboratories Network to the INS for diagnostic confirmation.

2.3. RNA extraction and real-time rRT-PCR

Viral RNA was obtained using the automated MagNA Pure LC nucleic acid extraction system (Roche Diagnostics GmbH, Mannheim, Germany) and viral RNA detection was performed by rRT-PCR using the SuperScript III Platinum One-Step Quantitative RT- kit. PCR (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.

2.4. Next generation sequencing

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

2.5. Genetic diversity analysis

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

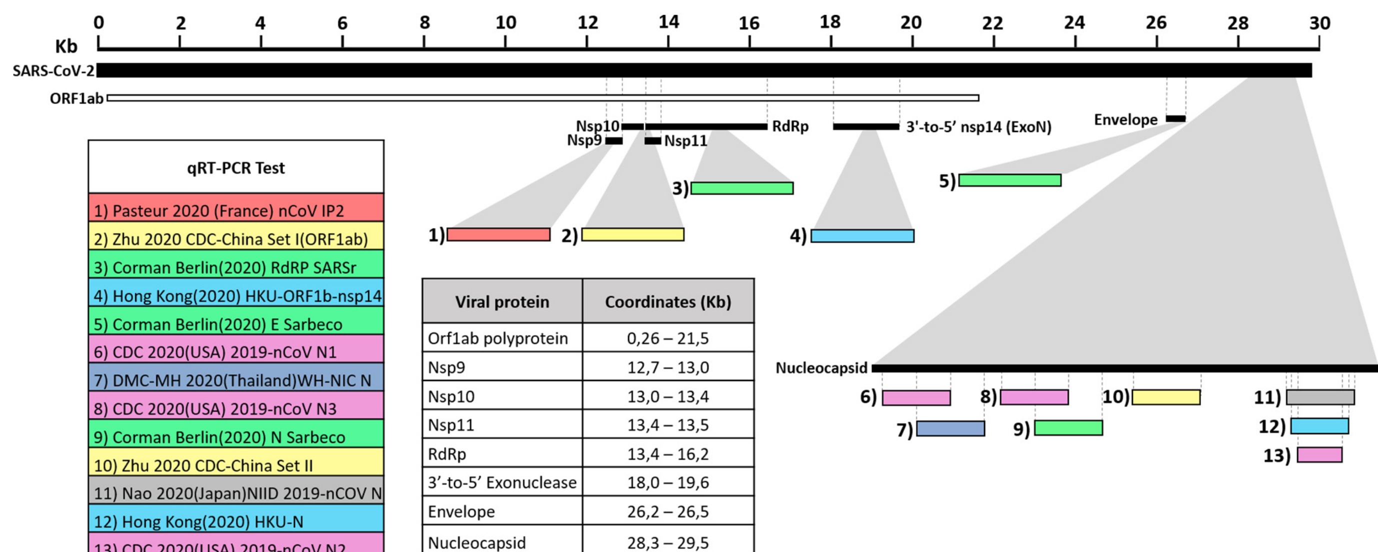

Fig. 1. Target hybridization regions of the primers/probes employed by the principal in-house protocols for molecular detection of SARS-CoV-2. The protocols targeted 13 different genome regions inside the Orf1ab (Nsp9, Nsp10, Nsp11, RdRp, ExoN), E (Envelope) y N (Nucleocapsid) genes. Target hybridization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

2.6. Oligonucleotides analysis

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

3. Results

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

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

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

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

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

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

3.3. The third codon position was found to align with the 3′ end of some primers with intended use for molecular detection of SARS-CoV-2

Codon positions in coding regions are differentially susceptible to nucleotide substitution, being the third codon position associated with a higher substitution rate. On the other hand, the perfect matching of the last nucleotide at the 3′ end of every forward and reverse primer is critical for DNA polymerase-based extension during PCR amplification (Staheli et al., 2009). Therefore, the rational design of primer to be used

Table 1
Target gene/regions and design conflicts of the analyzed primers and probes according to the genetic diversity of the Colombian SARS-CoV-2 strains.

Target gene	Primer/probe name[1]	Coordinates[2]	Codon position at the 3′ end	Genomes presenting mismatches
ORF1ab	Pasteur 2020(France)nCoV IP2-12669Fw	12,690–12,707	1	None
	Pasteur 2020(France)nCoV IP2-12696bProbe	12,717–12,737	NA	None
	Pasteur 2020(France)nCoV IP2-12759Rv	12,797–12,780	2	None
	Zhu 2020 CDC-China Set I(ORF1ab) fw	13,342–13,362	2	None
	Zhu 2020 CDC-Set I Probe(ORF1ab)	13,377–13,404	NA	hCoV/Colombia/Antioquia/79253/2020
	Zhu 2020 CDC-China Set I(ORF1ab) Rv	13,460–13,442	1	None
	Corman Berlin(2020) RdRP SARSr-F2	15,431–15,452	2	hCoV/Colombia/Valle del Cauca/79943/2020
	Corman Berlin(2020) RdRP SARSr-P2	14,470–15,494	NA	None
	Corman Berlin(2020) RdRP SARSr-R1	15,530–15,505	1	All
	Hong Kong(2020) HKU-ORF1b-nsp14F	18,778–18,797	2	None
	Hong Kong(2020) HKU-ORF1b-nsp141P	18,849–18,872	NA	None
	Hong Kong(2020) HKU-ORF1b-nsp14R	18,909–18,889	1	None
Gen E	Corman Berlin(2020) E Sarbeco F1	26,269–26,294	2	None
	Corman Berlin(2020) E Sarbeco P1	26,362–26,357	NA	None
	Corman Berlin(2020) E Sarbeco R1	26,381–26,360	2	None
Gen N	CDC 2020(USA) 2019-nCoV N1-F	28,287–28,306	3	None
	CDC 2020(USA) 2019-nCoV N1-P	28,309–28,332	NA	None
	CDC 2020(USA) 2019-nCoV N1-R	28,356–28,335	2	None
	DMC-MH 2020(Thailand)WH-NIC N-F	28,320–28,338	2	None
	DMC-MH 2020(Thailand)WH-NIC N-P	28,341–28,356	NA	None
	DMC-MH 2020(Thailand)WH-NIC N-R	28,376–28,358	1	None
	CDC 2020(USA) 2019-nCoV N3-F	28,681–28,702	3	None
	CDC 2020(USA) 2019-nCoV N3-P	28,704–28,727	NA	None
	CDC 2020(USA) 2019-nCoV N3-R	28,752–28,732	3	None
	Corman Berlin(2020) N Sarbeco F1	28,706–28,724	1	None
	Corman Berlin(2020) N Sarbeco P1	28,753–28,777	NA	None
	Corman Berlin(2020) N Sarbeco R1	28,833–28,814	1	None
	Zhu 2020 CDC-China Set II Fw(N)	28,881–28,902	2	hCoV/Colombia/Bogota/78390/2020 hCoV/Colombia/Valle del Cauca/81279/2020 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)	28,934–28,953	NA	None
	Zhu 2020 CDC-China Set II Rv(N)	28,979–28,958	1	None
	Nao 2020(Japan)NIID 2019-nCOV N F2	29,125–29,144	1	None
	Nao 2020(Japan)NIID 2019-nCOV N P2	29,222–29,241	NA	None
	Nao 2020(Japan)NIID 2019-nCOV N R2	29,299–29,280	2	None
	Hong Kong(2020) HKU-NF	29,145–29,166	2	None
	Hong Kong(2020) HKU-NP	29,179–29,198	NA	None
	Hong Kong(2020) HKU-NR	29,254–29,236	3	None
	CDC 2020(USA) 2019-nCoV N2-F	29,164–29,183	1	None
	CDC 2020(USA) 2019-nCoV N2-P	29,188–29,210	NA	All
	CDC 2020(USA) 2019-nCoV N2-R	29,230–29,213	1	None

[1] Names were assigned as posted online on the WHO website.
[2] Coordinates estimated from the alignment to the reference sequence NC_045512.2

with rapidly evolving RNA viruses should have the requisite of avoiding third and sometimes first codon positions.

Codon position of the last nucleotide at the 3′ end of every analyzed primer was identified according to the corresponding open-reading frame. The 3′ end of 11 primers corresponded to the first codon position, another 11 primers had 3′ ends located at the second codon

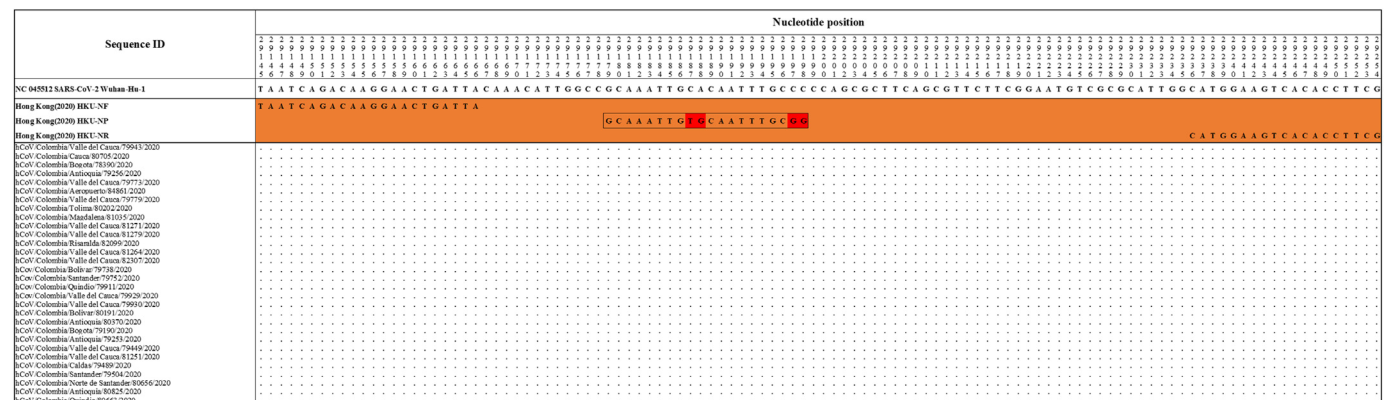

Fig. 2. Genetic diversity at the target region of the Hong_Kong (2020)_HKU-NP assay. Hong_Kong (2020) _HKU-NP probe displayed mismatches at positions 29,187–88 and 29,197–98 (highlighted in red), with all the Colombian sequences and the RefSeq displaying CA and CC, respectively. Genomic positions were estimated according to the SARS-CoV-2 reference genome available at GenBank (NC_045512.2). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Sequence ID	Nucleotide position 15431–15452	Nucleotide position 15469–15494	Nucleotide position 15505–15530
NC 045512 SARS-CoV-2 Wuhan-Hu-1	GTGAAATGGTCATGTGTGGCGG	CCAGGTGGAACCTCATCAGGAGATGC	TATGCTAATAGTGTTTTTAACATTTG
Corman Berlin(2020) RdRP SARSr-F2	GTGARATGGTCATGTGTGGCGG		
Corman Berlin(2020) RdRP SARSr-P2		CAGGTGGAACCTCATCAGGAGATGC	
Corman Berlin(2020) RdRP SARSr-R1			TATGCTAATAGTGTSTTTAACATYTG
hCoV/Colombia/Valle del Cauca/79943/2020	....................A.	....	....
hCoV/Colombia/Cauca/80705/2020	....	....	....
hCoV/Colombia/Bogota/78390/2020	....	....	....
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/79930/2020	....	....	....
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/81251/2020	....	....	....
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	....	....	....

Fig. 3. Genetic diversity at the target region of the Corman_Berlin (2020) _RdRP assay. The Colombian genome hCoV/Colombia/Valle_del_Cauca/79943/2020 displayed a substitution (G to A) at position 15,451 (highlighted in red), affecting the primer Corman_Berlin (2020) _RdRP_SARSr-F2 hybridization region. The substitution is located in the penultimate position at the 3 'end of the promer, generating a mismatch that could interfere with the 3′ pentamer stability. A degenerate base (S = C or G) in the primer Corman_Berlin (2020) _RdRP_SARSr-R1 at position 15,519 (highlighted in red) produces a mismatch with all the Colombian genomes and the RefSeq as they have T in the sense sequence, so the degenerate base should be one that includes an A among degenerate alternatives. Genomic positions were estimated according to the SARS-CoV-2 reference genome available at GenBank (NC_045512.2). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

position and the 3′ end of 4 primers aligned with the third codon position (Table 1). One of the primers (Nao 2020(Japan)NIID 2019-nCOV N F2) had the 3′ end aligned with the first codon position of the codon CUA which encodes for alanine. This codon allows a substitution at the first codon position (C to U) to be synonymous and therefore may escape to the selection pressure at the protein level.

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

4. Discussion

Next Generation Sequencing (NGS) technologies are a valuable tool in determining the whole genome of microorganisms impacting on public health, they have accelerated our ability to understand important factors in the biology of infectious diseases such as the identification of determinants of virulence, drug-resistance associated substitutions, vaccine design and genomic epidemiology, a powerful approach to integrate epidemiologic and genetic information in reconstructing transmission patterns and infectious disease dynamics (Gwinn et al., 2019). NGS has also enabled pathogen discovery in a timely manner compared to traditional methods, being essential to the virus taxonomic classification during the beginnings of COVID-19 pandemics (Zhou et al., 2020; Zhu et al., 2020). This robust molecular

Sequence ID	Nucleotide position 13342–13362	Nucleotide position 13377–13404	Nucleotide position 13432–13450
NC 045512 SARS-CoV-2 Wuhan-Hu-1	CCCTGTGGGTTTTACACTTAA	CCGTCTGCGGTATGTGGAAAGGTTATGG	TCAGCTGATGCACAATCGT
Zhu 2020 CDC-China Set I(ORF1ab) fw	CCCTGTGGGTTTTACACTTAA		
Zhu 2020 CDC-Set I Probe(ORF1ab)		CCGTCTGCGGTATGTGGAAAGGTTATGG	
Zhu 2020 CDC-China Set I(ORF1ab) Rv			TCAGCTGATGCACAATCGT
hCoV/Colombia/Valle del Cauca/79943/2020	....	....	....
hCoV/Colombia/Cauca/80705/2020	....	....	....
hCoV/Colombia/Bogota/78390/2020	....	....	....
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/79930/2020	....	....	....
hCoV/Colombia/Bolivar/80191/2020	....	....	....
hCoV/Colombia/Antioquia/80370/2020	....	....	....
hCoV/Colombia/Bogota/79190/2020	....	....	....
hCoV/Colombia/Antioquia/79253/2020	....	.......T....................	....
hCoV/Colombia/Valle del Cauca/79449/2020	....	....	....
hCoV/Colombia/Valle del Cauca/81251/2020	....	....	....
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	....	....	....

Fig. 4. Genetic diversity at the target region of the Zhu_2020_CDC-Set I assay. The Colombian genome hCoV/Colombia/Antioquia/79253/2020 displayed a substitution (C to T) at position 13,384 (highlighted in red) where the probe Zhu_2020_CDC-Set I Probe (ORF1ab) hybridizes. Genomic positions were estimated according to the SARS-CoV-2 reference genome available at GenBank (NC_045512.2). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Sequence ID	Nucleotide position 28881–28902		Nucleotide position 28934–28979
NC 045512 SARS-CoV-2 Wuhan-Hu-1	GGGGAACTTCTCCTGCTAGAAT	...	TTGCTGCTGCTTGACAGATTGAACCAGCTTGAGAGCAAAATGTCTG
Zhu 2020 CDC-China Set II Fw(N)	GGGGAACTTCTCCTGCTAGAAT		
Zhu 2020 CDC-China Set II Probe(N)			TTGCTGCTGCTTGACAGATT
Zhu 2020 CDC-China Set II Rv(N)			CAGCTTGAGAGCAAAATGTCTG
hCoV/Colombia/Valle del Cauca/79943/2020	....		....
hCoV/Colombia/Cauca/80705/2020	....		....
hCoV/Colombia/Bogota/78390/2020	AAC...................		....
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	AAC...................		....
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	..................T...		....
hCov/Colombia/Valle del Cauca/79929/2020	....		....
hCoV/Colombia/Valle del Cauca/79930/2020	....		....
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/81251/2020	AAC...................		....
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	..................T...		....

Fig. 5. Genetic diversity at the target region of the Zhu_2020_CDC-China_Set II assay. Three Colombian genomes, hCoV/Colombia/Bogota/78390/2020, hCoV/Colombia/ Valle_del_Cauca/81279/2020 and hCoV/Colombia/Valle_del_Cauca/81251/2020 have the same pattern of substitutions (GGG to AAC) in primer Zhu_2020_CDC-China_Set II Fw (N) in the first three nucleotides at the 5′ end (highlighted in red). Furthermore, the hCoV/Colombia/Quindio/79911/2020 and hCoV/Colombia/Quindio/80663/2020 genomes displayed a substitution (G to T) at position 28,899 where this oligonucleotide hybridizes (highlighted in red). Genomic positions were estimated according to the SARS-CoV-2 reference genome available at GenBank (NC_045512.2). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

information is also the raw material for the development and refinement of molecular and serological diagnostic methods (Wang et al., 2020) and all sequence-based designs are susceptible to improvement as the virus disseminates and its genetic variability increases.

Mutation is the fundamental source of genetic variation and RNA viruses are particularly susceptible to have high mutation rates during the genome replication (Sanjuan et al., 2010). Therefore, RNA viruses display high substitution rates when analyzed through time. Although sequence identity between the analyzed genomes and the Wuhan reference strain (NC_045512) was 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 (Tang et al., 2020).

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

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

By the time of this report, more than 31.000 sequences have been made available on GISAID by researchers across the globe; in line with the "*Common primer check for high-quality genomes 2020-05-22*" analysis from this database (BII/GIS, 2020), the Zhu 2020 CDC-China Set II (N) showed the higher percent of genomes with mutation in both, the entire -and the 3′ end primer region, which can affect the sensitivity of this protocol given the genetic variability present in the country. Furthermore, our analysis with the Hong Kong(2020) HKU-NP, showed excessive base mismatch against all tested genomes, so it is highly recommended to improve the design of this probe.

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

Another critical characteristic of the primers/probes of the current

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

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

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

5. Conclusions

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

Supplementary data to this article can be found online at https://doi.org/10.1016/j.meegid.2020.104390.

Funding

This study was funded by the Instituto Nacional de Salud, in Bogotá D.C., Colombia.

Declaration of Competing Interest

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

Acknowledgments

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

References

Andersen, K.G., Rambaut, A., Lipkin, W.I., Holmes, E.C., Garry, R.F., 2020. The proximal origin of SARS-CoV-2. Nat. Med. 26, 450–452.
BII/GIS, 2020. Analysis Update. (GISAID).
brian-jgi, 2020. BBMap Short Read Aligner, and Other Bioinformatic Tools.
Chu, D.K.W., Pan, Y., Cheng, S.M.S., Hui, K.P.Y., Krishnan, P., Liu, Y., Ng, D.Y.M., Wan, C.K.C., Yang, P., Wang, Q., Peiris, M., Poon, L.L.M., 2020. Molecular diagnosis of a Novel Coronavirus (2019-nCoV) causing an outbreak of pneumonia. Clin. Chem. 66, 549–555.
Corman, V.M., Rasche, A., Baronti, C., Aldabbagh, S., Cadar, D., Reusken, C.B., Pas, S.D., Goorhuis, A., Schinkel, J., Molenkamp, R., Kummerer, B.M., Bleicker, T., Brunink, S., Eschbach-Bludau, M., Eis-Hubinger, A.M., Koopmans, M.P., Schmidt-Chanasit, J., Grobusch, M.P., de Lamballerie, X., Drosten, C., Drexler, J.F., 2016. Assay optimization for molecular detection of Zika virus. Bull. World Health Organ. 94, 880–892.
Corman, V.M., Landt, O., Kaiser, M., Molenkamp, R., Meijer, A., Chu, D.K., Bleicker, T., Brünink, S., Schneider, J., Schmidt, M.L., Mulders, D.G., Haagmans, B.L., van der Veer, B., van den Brink, S., Wijsman, L., Goderski, G., Romette, J.-L., Ellis, J., Zambon, M., Peiris, M., Goossens, H., Reusken, C., Koopmans, M.P., Drosten, C., 2020. Detection of 2019 Novel Coronavirus (2019-nCoV) by Real-Time RT-PCR. Euro Surveillance : Bulletin Europeen Sur les Maladies Transmissibles = European Communicable Disease Bulletin. 25 (2000045).
Dong, E., Du, H., Gardner, L., 2020. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect. Dis. 20, 533–534.
Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.
Gwinn, M., MacCannell, D., Armstrong, G.L., 2019. Next-generation sequencing of infectious pathogens. JAMA 321, 893–894.
Holshue, M.L., DeBolt, C., Lindquist, S., Lofy, K.H., Wiesman, J., Bruce, H., Spitters, C., Ericson, K., Wilkerson, S., Tural, A., Diaz, G., Cohn, A., Fox, L., Patel, A., Gerber, S.I., Kim, L., Tong, S., Lu, X., Lindstrom, S., Pallansch, M.A., Weldon, W.C., Biggs, H.M., Uyeki, T.M., Pillai, S.K., Washington State -nCo, V.C.I.T., 2020. First case of 2019 Novel Coronavirus in the United States. N. Engl. J. Med. 382, 929–936.
Hu, X.M., Xu, J.X., Jiang, L.X., Deng, L.R., Gu, Z.M., Xie, X.Y., Ji, H.C., Wang, W.H., Li, L.M., Tian, C.N., Song, F.L., Huang, S., Zheng, L., Zhong, T.Y., 2019. Design and evaluation of a Novel multiplex real-time PCR melting curve assay for the simultaneous detection of nine sexually transmitted disease pathogens in genitourinary secretions. Front. Cell. Infect. Microbiol. 9, 382.
Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. 35, 1547–1549.
Li, J., Li, Z., Cui, X., Wu, C., 2020. Bayesian phylodynamic inference on the temporal evolution and global transmission of SARS-CoV-2. J. Inf. Secur.
PAHO, J., 2020. Directrices de Laboratorio para la Detección y el Diagnóstico de la Infección con el Virus COVID-19. Pan American Health Organization.
Quick, J., 2020. nCoV-2019 Sequencing Protocol. Protocols.io.
Sanjuan, R., Nebot, M.R., Chirico, N., Mansky, L.M., Belshaw, R., 2010. Viral mutation rates. J. Virol. 84, 9733–9748.
Staheli, J.P., Ryan, J.T., Bruce, A.G., Boyce, R., Rose, T.M., 2009. Consensus-degenerate hybrid oligonucleotide primers (CODEHOPs) for the detection of novel viruses in non-human primates. Methods 49, 32–41.
Tang, X., Wu, C., Li, X., Song, Y., Yao, X., Wu, X., Duan, Y., Zhang, H., Wang, Y., Qian, Z., Cui, J., Lu, J., 2020. On the origin and continuing evolution of SARS-CoV-2. Natl. Sci. Rev.
Wang, H., Pipes, L., Nielsen, R., 2020. Synonymous mutations and the molecular evolution of SARS-Cov-2 origins. bioRxiv (2020.2004.2020.052019).
WHO, 2020a. WHO Director-General's Opening Remarks at the Media Briefing on COVID-19 - 11 March 2020. World Health Organization.
WHO, 2020b. Coronavirus Disease 2019 (COVID-19). World Health Organization.
WHO, 2020c. Novel Coronavirus (2019-nCoV) Technical Guidance: Laboratory Testing for 2019-nCoV in Humans. World Health Organization.
WHO, 2020d. Coronavirus Disease (COVID-19).
Zhang, T., Wu, Q., Zhang, Z., 2020. Probable pangolin origin of SARS-CoV-2 associated with the COVID-19 outbreak. Curr. Biol. 30, 1346–1351 e1342.
Zhou, P., Yang, X.L., Wang, X.G., Hu, B., Zhang, L., Zhang, W., Si, H.R., Zhu, Y., Li, B., Huang, C.L., Chen, H.D., Chen, J., Luo, Y., Guo, H., Jiang, R.D., Liu, M.Q., Chen, Y., Shen, X.R., Wang, X., Zheng, X.S., Zhao, K., Chen, Q.J., Deng, F., Liu, L.L., Yan, B., Zhan, F.X., Wang, Y.Y., Xiao, G.F., Shi, Z.L., 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579, 270–273.
Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., Niu, P., Zhan, F., Ma, X., Wang, D., Xu, W., Wu, G., Gao, G.F., Tan, W., China Novel Coronavirus, I., Research, T., 2020. A Novel Coronavirus from patients with pneumonia in China, 2019. N. Engl. J. Med. 382, 727–733.