















Nucleocapsid (N) Gene Mutations of SARS-CoV-2 Can Affect Real-Time RT-PCR Diagnostic and Impact False-Negative Results

Jéssika Cristina Chagas Lesbon ¹, Mirele Daiana Poleti ¹, Elisângela Chicaroni de Mattos Oliveira ¹, José Salvatore Leister Patané ², Luan Gaspar Clemente ³, Vincent Louis Viala ², Gabriela Ribeiro ², Marta Giovanetti ^{4,5}, Luiz Carlos Junior de Alcantara ⁴, Loyze Paola Oliveira de Lima ², Antonio Jorge Martins ², Claudia Renata dos Santos Barros ², Elaine Cristina Marqueze ², Jardelina de Souza Todão Bernardino ², Debora Botequiu Moretti ², Ricardo Augusto Brassaloti ³, Raquel de Lello Rocha Campos Cassano ³, Pilar Drummond Sampaio Correa Mariani ⁶, Svetoslav Nanev Slavov ⁷, Rafael Bezerra dos Santos ⁷, Evandra Strazza Rodrigues ⁷, Elaine Vieira Santos ⁷, Josiane Serrano Borges ⁷, Debora Glenda Lima de La Roque ⁷, Joao Paulo Kitajima ⁸, Bibiana Santos ⁸, Patricia Akemi Assato ⁹, Felipe Allan da Silva da Costa ⁹, Cecilia Artico Banho ¹⁰, Livia Sacchetto ¹⁰, Marilia Mazzi Moraes ¹⁰, Melissa Palmieri ¹¹, Fabiana Erica Vilanova da Silva ¹², Rejane Maria Tommasini Grotto ⁹, Jayme A. Souza-Neto ⁹, Mauricio Lacerda Nogueira ¹⁰, Luiz Lehman Coutinho ³, Rodrigo Tocantins Calado ⁷, Raul Machado Neto ², Dimas Tadeu Covas ², Simone Kashima ⁷, Maria Carolina Elias ², Sandra Coccuzzo Sampaio ² and Heidge Fukumasu ^{1,*}



Citation: Lesbon, J.C.C.; Poleti, M.D.; de Mattos Oliveira, E.C.; Patané, J.S.L.; Clemente, L.G.; Viala, V.L.; Ribeiro, G.; Giovanetti, M.; de Alcantara, L.C.J.; de Lima, L.P.O.; et al. Nucleocapsid (N) Gene Mutations of SARS-CoV-2 Can Affect Real-Time RT-PCR Diagnostic and Impact False-Negative Results. *Viruses* **2021**, *13*, 2474. <https://doi.org/10.3390/v13122474>

Academic Editors: Fabrício S. Campos, Luciana Barros de Arruda and Maite FS. Vaslin

Received: 27 October 2021
Accepted: 1 December 2021
Published: 10 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

- ¹ Laboratory of Comparative and Translational Oncology, Department of Veterinary Medicine, School of Animal Science and Food Engineering, University of Sao Paulo, Pirassununga 13635-900, São Paulo, Brazil; jessika.chagas@usp.br (J.C.C.L.); mirelep@usp.br (M.D.P.); limattos@usp.br (E.C.d.M.O.)
 - ² Butantan Institute, São Paulo 05503-000, São Paulo, Brazil; jose.patane@butantan.gov.br (J.S.L.P.); vincent.viala@butantan.gov.br (V.L.V.); gabriela.ribeiro.br@gmail.com (G.R.); loyze.lima@butantan.gov.br (L.P.O.d.L.); antonio.martins@butantan.gov.br (A.J.M.); claudia.barros@butantan.gov.br (C.R.d.S.B.); elaine.marqueze@butantan.gov.br (E.C.M.); jardelina.bernardino@butantan.gov.br (J.d.S.T.B.); debora.moretti@butantan.gov.br (D.B.M.); raul.machado@butantan.gov.br (R.M.N.); dimas.covas@butantan.gov.br (D.T.C.); carolina.eliassabbaga@butantan.gov.br (M.C.E.); sandra.coccuzzo@butantan.gov.br (S.C.S.)
 - ³ Functional Genomic Center, Department of Animal Science, Luiz de Queiroz School of Agriculture, University of Sao Paulo, Piracicaba 13418-900, São Paulo, Brazil; luan.clemente@usp.br (L.G.C.); ricardo.brassaloti@usp.br (R.A.B.); raquel_ello@yahoo.com.br (R.d.L.R.C.C.); llcoutinho@usp.br (L.L.C.)
 - ⁴ Fundação Oswaldo Cruz, FIOCRUZ, Manguinhos 21040-900, Rio de Janeiro, Brazil; marta.giovanetti@ioc.fiocruz.br (M.G.); alcantaraluiz42@gmail.com (L.C.J.d.A.)
 - ⁵ Institute of Biological Sciences, Federal University of Minas Gerais, Belo Horizonte 31270-901, Minas Gerais, Brazil
 - ⁶ NGS Soluções Genômicas, Piracicaba 13416-030, São Paulo, Brazil; pilarmariani@hotmail.com
 - ⁷ Blood Center of Ribeirão Preto, Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto 14051-060, São Paulo, Brazil; svetoslav.slavov@hemocentro.fmrp.usp.br (S.N.S.); rafaelbezerra@usp.br (R.B.d.S.); evandra@hemocentro.fmrp.usp.br (E.S.R.); elainevs@alumni.usp.br (E.V.S.); josiane.boges@hemocentro.fmrp.usp.br (J.S.B.); debora.laroque@gmail.com (D.G.L.d.L.R.); rtcalado@usp.br (R.T.C.); skashima@hemocentro.fmrp.usp.br (S.K.)
 - ⁸ Mendelics Genomic Analysis, São Paulo 02511-000, São Paulo, Brazil; joao.kitajima@mendelics.com.br (J.P.K.); bibiana.santos@mendelics.com.br (B.S.)
 - ⁹ School of Agricultural Sciences, São Paulo State University, Botucatu 18618-970, São Paulo, Brazil; p.assato@unesp.br (P.A.A.); felipe.allan@unesp.br (F.A.d.S.d.C.); rejane.grotto@unesp.br (R.M.T.G.); jayme.souza-neto@unesp.br (J.A.S.-N.)
 - ¹⁰ Laboratório de Pesquisas em Virologia, Departamento de Doenças Dermatológicas, Infeciosas e Parasitárias, Faculdade de Medicina de São José do Rio Preto, São José do Rio Preto 15090-000, São Paulo, Brazil; ceci.abanho@gmail.com (C.A.B.); liviasacchetto@gmail.com (L.S.); mariliamazzi@hotmail.com (M.M.M.); mauricio.nogueira@edu.famerp.br (M.L.N.)
 - ¹¹ Coordenação de Vigilância em Saúde—Secretaria Municipal da Saúde, São Paulo, São Paulo 01223-906, Brazil; melissapalmieri@prefeitura.sp.gov.br
 - ¹² Laboratory Assistance, Coordination of Primary Care, Municipal Health Department, São Paulo, São Paulo 01223-906, Brazil; fevilanova@prefeitura.sp.gov.br
- * Correspondence: fukumasu@usp.br; Tel.: +55-19-35656864

Abstract: The current COVID-19 pandemic demands massive testing by Real-time RT-PCR (Reverse Transcription Polymerase Chain Reaction), which is considered the gold standard diagnostic test for the detection of the SARS-CoV-2 virus. However, the virus continues to evolve with mutations that lead to phenotypic alterations as higher transmissibility, pathogenicity or vaccine evasion. Another big issue are mutations in the annealing sites of primers and probes of RT-PCR diagnostic kits leading to false-negative results. Therefore, here we identify mutations in the *N* (Nucleocapsid) gene that affects the use of the GeneFinder COVID-19 Plus RealAmp Kit. We sequenced SARS-CoV-2 genomes from 17 positive samples with no *N* gene detection but with *RdRP* (RNA-dependent RNA polymerase) and *E* (Envelope) genes detection, and observed a set of three different mutations affecting the *N* detection: a deletion of 18 nucleotides (Del28877-28894), a substitution of GGG to AAC (28881-28883) and a frameshift mutation caused by deletion (Del28877-28878). The last one cause a deletion of six AAs (amino acids) located in the central intrinsic disorder region at protein level. We also found this mutation in 99 of the 14,346 sequenced samples by the Sao Paulo state Network for Pandemic Alert of Emerging SARS-CoV-2 variants, demonstrating the circulation of the mutation in Sao Paulo, Brazil. Continuous monitoring and characterization of mutations affecting the annealing sites of primers and probes by genomic surveillance programs are necessary to maintain the effectiveness of the diagnosis of COVID-19.

Keywords: COVID-19; RT-PCR; *N* gene; GeneFinder; mutation

1. Introduction

The SARS-CoV-2 virus is the causative agent of the COVID-19 pandemic and real-time RT-PCR (Reverse Transcription Polymerase Chain Reaction) is the gold standard of diagnostic tests. However, there are several commercial and non-commercial tests available with different sensitivities and specificities, the majority of which were designed at the beginning of the pandemic with the available genome sequences at that moment. Twenty months after the first diagnosis, the virus evolved worldwide with different strains emerging on a daily basis. Therefore, mutations conferring phenotypic alterations as higher transmissibility, pathogenicity, or vaccine evasion happened as the ones described in the P.1 (Brazilian, Gamma), B.1.617 (Indian, Delta), and others. Mutations with unknown phenotypic consequences can occur on primers and probes annealing sites from the RT-PCR diagnostic kits. These mutations can lead to false-negative results, increasing the chance of false-negative COVID-19 patients spreading the virus.

To overcome this problem, most of the real-time RT-PCR COVID-19 diagnostic tests detect two or three SARS-CoV-2 gene fragments from *N* (Nucleocapsid), *E* (Envelope), *RdRP* (RNA-dependent RNA polymerase), *S* (spike), *ORF1* (Nonstructural protein), and others. However, there are protocols using only one SARS-CoV-2 gene fragment as is the case of recently developed tests for COVID-19 and Flu detection at the same time by multiplex real-time RT-PCR, increasing the possibility of false-negative SARS-CoV-2 virus detection.

At the beginning of March 2020, we set the Laboratory of Translational and Comparative Oncology of the School of Animal Science and Food Engineering of the University of Sao Paulo at Pirassununga, São Paulo State, Brazil, to fully work on COVID-19 diagnosis by real-time RT-PCR. Since then, we performed more than 130,000 tests for more than 40 cities from southeast of Sao Paulo State so far using three different protocols: the CDC (Centers for Disease Control and Prevention) (*N1*, *N2*, *RNP* Ribonucleoprotein), the GeneFinder COVID-19 Plus RealAmp Kit (*N*, *E*, *RdRP* and human control), and AllPlex 2019-nCoV Assay (*N*, *E*, *RdRP* and human control). Specifically, the GeneFinder kit was used on more than 90,000 samples from the Sao Paulo State COVID-19 Diagnostic Network which our laboratory also integrates. However, in early February 2021, we detected samples with the GeneFinder kit that had high viral load detected by *E* and *RdRP* probes but lacking *N* probe detection. This is a real problem since the *N* probe of the GeneFinder kit is the

most sensitive for SARS-CoV-2 detection in our experience and samples with low viral load (Ct (Cycle threshold) > 33, i.e.,) are detected mainly by *N* probe detection.

Therefore, the objective of this work was to identify the possible mutation(s) in the SARS-CoV-2 *N* gene responsible for this specific problem and check if the problem was extended to other labs from the Sao Paulo State COVID-19 Diagnostic Network. Further characterization of the most frequent mutation in *N* gene affecting the diagnostics is also presented and discussed.

2. Materials and Methods

2.1. Diagnostic Service

The Laboratory of Translational and Comparative Oncology (LOCT-USP) from the School of Animal Science and Food Engineering (FZEA-USP, Pirassununga, SP, Brazil) is BSL-2 certified by CTNBio (Brazilian Technical Committee of Biosafety, Brasilia, DF, Brazil), certified by Instituto Adolfo Lutz for COVID-19 diagnostics and member of the Sao Paulo State COVID-19 Diagnostic Network and the Sao Paulo state Network for Pandemic Alert of Emerging SARS-CoV-2 variants.

2.2. Samples and Extraction of Total Nucleic Acids

Samples from nasopharyngeal and oropharyngeal swabs were collected according to the standard protocol and placed inside transport tubes containing 3 mL sterile saline solution. All samples were transported to the laboratory at a cold temperature (2–8 °C) within 12 h post collection and processed on the day. RNA was isolated from clinical samples (naso and oropharyngeal swabs) using the kit extract–RNA e DNA Viral (Loccus, Cotia, SP, Brazil). Purification by magnetic beads facilitates the isolation process and results in a high yield and purity level of the isolated nucleic acid.

2.3. Real-Time RT-PCR Using GeneFinder COVID-19 Plus RealAmp Kit

The majority of the real-time RT-PCR COVID-19 diagnostic tests were performed with the GeneFinder™ COVID-19 Plus RealAmp RT-PCR master mix (Osang Healthcare Co., Anyang-si, Korea). The purified nucleic acid is reverse transcribed into cDNA and can detect new coronavirus using a real-time RT-PCR probe, through specific primer using three different primers sets for viral detection (*N*, *E*, and *RdRP* gene fragments) and fluorescent probe reactions.

2.4. Genome Sequencing

To find out the *N* gene mutations, we selected 17 representative samples positive for COVID-19 by real-time RT-PCR, which did not present Ct values for the *N* gene. The libraries were constructed using Illumina COVIDSeq™ Test (Illumina Inc, San Diego, CA, USA), according to manufacturer's instructions. The cDNA was carried out on RNA samples isolated and synthesized by reverse transcriptase with random hexamers. The virus genome was amplified using two pool primers in separate PCR reactions. The PCR amplified product was processed for tagmentation and adapter ligation using IDT for Illumina Nextera UD Indexes Set A, B, C, D (384 indexes, 384 samples). The enrichment and cleanup steps were carried out according to the manufacturer's protocol. All samples were processed as batches in a 96-well plate; these 96 libraries were pooled together in a tube. Pooled samples were quantified using Qubit dsDNA High Sensitivity assay kit on a Qubit fluorometer (Invitrogen Inc, Carlsbad, CA, USA), and the fragment sizes were analyzed in Agilent Fragment analyzer 5200 (Agilent Inc, Santa Clara, CA, USA). The pooled library was normalized to 4 nM concentration and denatured with 5 µL of 0.2 N of NaOH. The 1.2 pM library was spiked with 1% PhiX control (PhiX Control v3, Illumina Inc, San Diego, CA, USA) and sequenced on an Illumina MiniSeq platform (Illumina), using a MiniSeq System Mid-Output Kit (300 cycles). The viral isolate sequences were aligned with the reference sequence for SARS-CoV-2 using the Illumina DRAGEN COVIDSeq Test pipeline. Viral strains were classified and mutations analyzed using the software

tools Pangolin (<http://pangolin.cog-uk.io/>, accessed on 5 August 2021), and nextclade (<https://clades.nextstrain.org/>, accessed on 5 August 2021).

2.5. In Silico Evaluation of N Gene Mutations

To evaluate the possible structural impacts of the 203–208 deletion, and considering the fact that the structure of the full length protein is not known, we submitted the N protein sequence to the Robetta server (<https://robetta.bakerlab.org>, accessed on 10 August 2021) for structure prediction and analysis for both full length and mutated proteins. The default parameters were used to produce predicted models using the simultaneous processing of sequence, distance, and coordinate information by the three-track architecture implemented in the RoseTTAfold method [1].

3. Results

3.1. Frequency of Non-N Detection by the USP-Pirassununga COVID-19 Task Force

The USP-Pirassununga COVID-19 Task Force performed over 130,000 real-time RT-PCR tests for COVID-19 detection for more than 40 different cities from the southeast part of Sao Paulo State so far (31 July 2021). Most of these tests (90,045) were performed with the GeneFinder COVID-19 Plus RealAmp Kit (Osang Healthcare Co., Korea) which uses three different primers sets for viral detection (*N*, *E*, and *RDRP* gene fragments). The Ct values for the three viral genes present similarly in positive samples with high and medium viral load (Cts ranging from 12 to 30). Low viral load samples (Cts > 35) were detected with two (generally *E* and *N*) or only one probe (generally *N* probe).

We noted a specific pattern of N^- , E^+ , $RDRP^+$ in the samples from February 2021 in our diagnostic service, being the first sample detected on 26 February 2021. So far, 69 samples were positive with this pattern from a total of 86,393 tests performed with GeneFinder since April 2020 (until 15 July 2021) by our service. This calls our attention for further investigation because the *N* primer set is the most sensitive probe in our conditions and losing this signal could infer false-negative results especially in low viral load samples.

3.2. Genome Sequencing and Identification of N Gene Mutations

Therefore, we sequenced the SARS-CoV-2 genome from 17 representative positive samples, and all were classified as P.1 (or P.1.1) variant. Three different mutations in gene *N* were observed: Del28877-28894 (14/17) causing a deletion of six AA, a substitution of GGG to AAC in 28881-28883 (2/17) changing two AAs and a frameshift mutation caused by a deletion of 28877-28878 (1/17) in Table 1.

Table 1. Mutations in gene N detected by sequencing.

Mutation	Alteration	Frequency
DEL 28877-28894	Deletion N:S202- N:R203- N:G204- N:T205- N:S206- N:P207-	82% (14/17)
G28881A G28882A G28883C	Substitution N:R203K N:G204R	12% (2/17)
DEL 28877-28878	Deletion Frameshift	6% (1/17)

The mutations were further characterized in silico for potential effects on protein function. The mutations causing the six AA deletion and two AA substitutions occurred in the Linker region of the nucleocapsid protein.

3.3. Data from Other Laboratories from the Sao Paulo State COVID-19 Diagnostic Network

For further validation of these findings, we gathered data from two other centers for COVID-19 diagnostic belonging to the Sao Paulo State Network coordinated by Butantan Institute both using the Osang's GeneFinder kit. The service from the School of Medicine in Ribeirão Preto from the University of Sao Paulo processed 76,516 samples from January 1st to 30 June 2021, with 31,899 positive cases. From these, 24 samples fitted the criteria of N^- , E^+ , $RDRP^+$ and four of them were sequenced confirming the presence of DEL 28877–28894. Interestingly, the first sample from this lab was first detected on 17 February 2021, almost simultaneously with our first detection. Data from the major laboratory from the Sao Paulo State COVID-19 Diagnostic Network situated in the Butantan Institute processed 389,273 samples with 201,987 positive diagnostics from January 1st to 31 May 2021. From these, 893 samples (0.23% from all cases) presented the pattern of N^- , E^+ , $RDRP^+$ detection by RT-PCR.

3.4. Presence and Frequency of N Gene Mutations in the Databank of SARS-CoV2 Sequenced Samples from the Sao Paulo COVID-19 Variant Alert Network

The Sao Paulo COVID-19 Variant Alert Network sequenced 14,316 SARS-CoV-2 samples (including 1046 sequenced in our lab) and a total of 111 samples (including our 17 already mentioned samples) (Table S1) have mutations in the same region of the N gene. The most frequent was DEL 28877–28894 (99 samples) but two other mutations were also found: DEL 28877–28886 (11 samples) and DEL 28847–28886 (1 sample). The mutated SARS-CoV-2 viruses were detected in different cities through the Sao Paulo state (Figure 1).

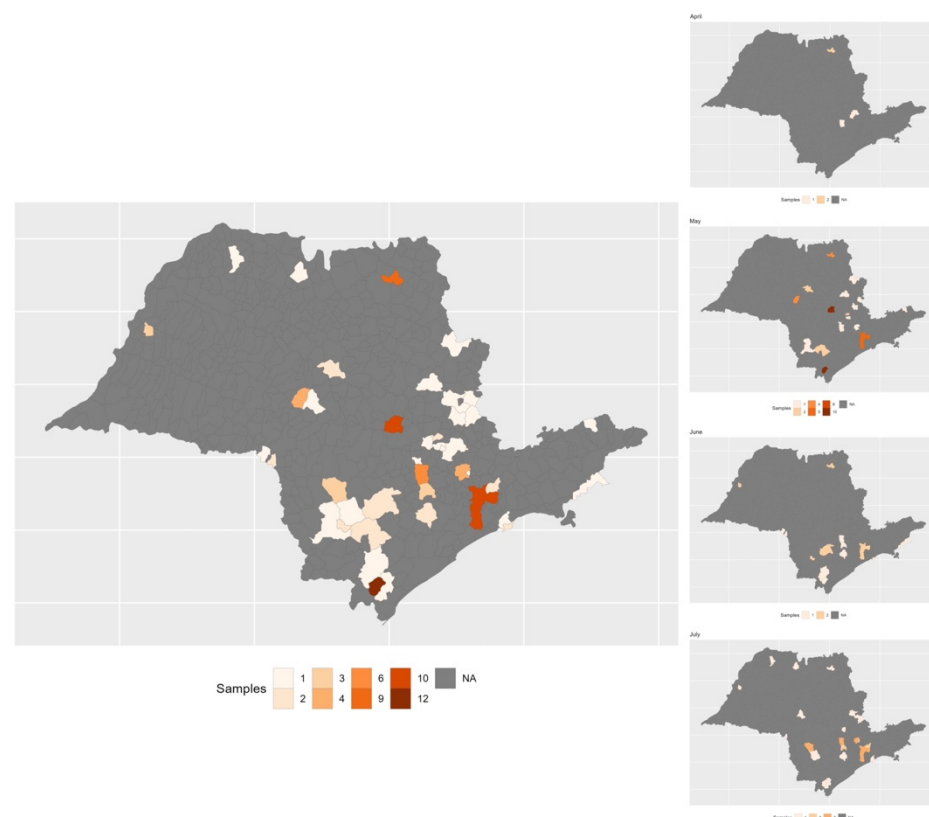


Figure 1. Geolocation of the 111 samples with mutations in the same region of N gene and monthly based frequency from April to July 2021.

3.5. Characterization of DEL 28877–28894 Mutation in SARS-CoV2

The DEL 28877–28894 at protein level produced a deletion of six AAs located in the central intrinsic disorder region (IDR) (182–247) that links the NTD (N-terminal domains) to CTD (C-terminal domains). This is the same region where the highest frequency of

N protein mutations is reported [2–4], including substitutions at amino acids 203 and 204 (also found in this work), which are part of a serine and arginine-rich region comprising residues 184–204 [2–5]. (Figure 2)

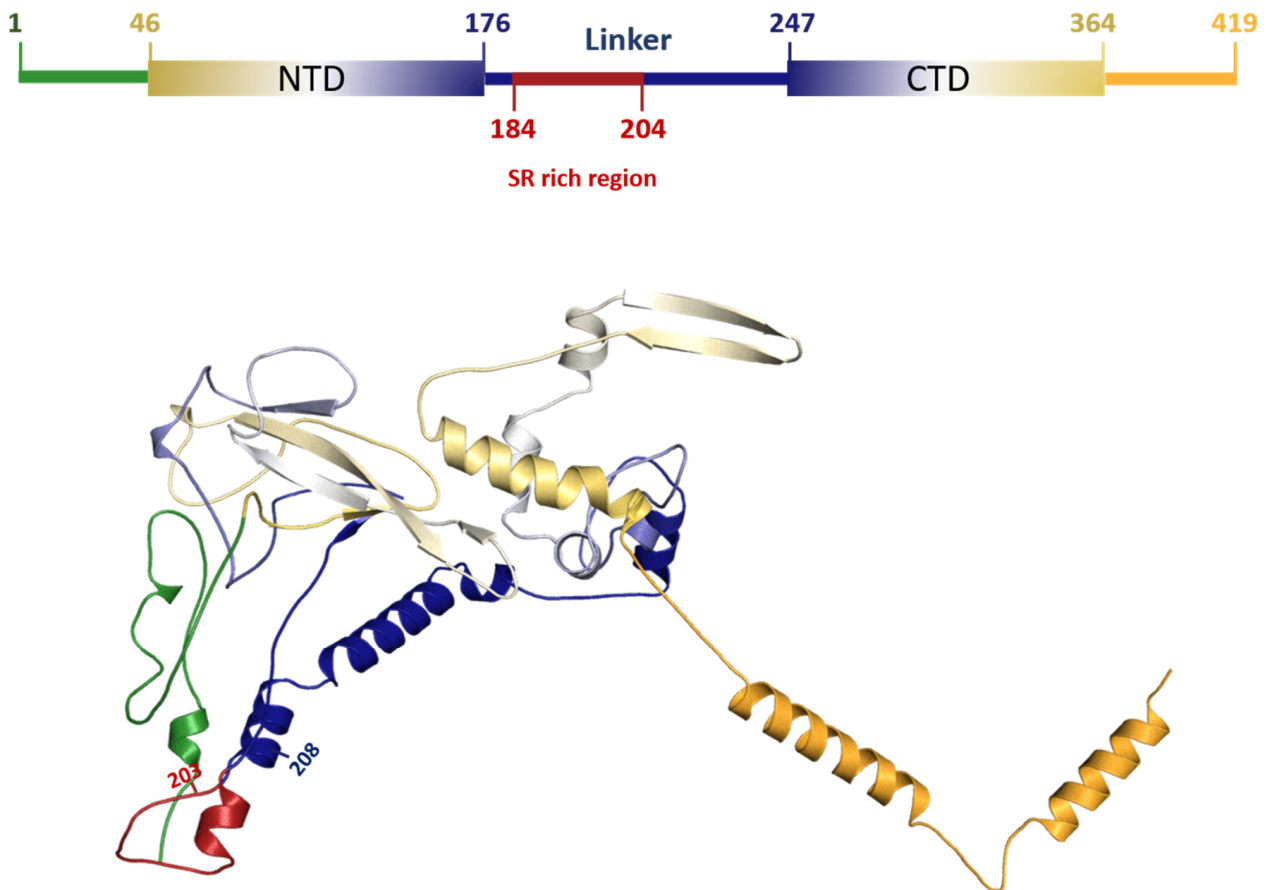


Figure 2. Representative model of protein N indicating the region of the 203–208 deletion, generated by the Robetta server and using the software PyMOL for image generation. The generated model is in compliance with previous work suggesting that the NTD and CTD minimally interact to each other, if at all. They are separated by the linker that is partially extended and with disordered regions have transient α -helices [6–9]. Deletion of the 203–208 fragment will alter physico-chemical properties that might impact on protein mobility between N- and C-terminal domains, as well as in the RNA-mediated phase separation process.

4. Discussion

Commercial SARS-CoV-2 real-time RT-PCR kits generally do not disclose the position of primers and probes for the detection sites but here we describe a set of mutations in *N* gene that affect the detection of the *N* gene by the GeneFinder COVID-19 Plus RealAmp Kit used for the diagnostic in Sao Paulo State since the beginning of the COVID-19 pandemics. Although the failure to detect *N* gene, the GeneFinder assay positively call samples by *E* and/or *RdRP* targets, emphasizing the importance of more than one target in a diagnostic RT-PCR kit. The frequency of these *N* mutations affecting GeneFinder *N* gene detection in the epidemiological data from the Sao Paulo state Network for Pandemic Alert of Emerging SARS-CoV-2 variants is low (~0.78%, 111 samples in 14,316) but these are spread all over the state as showed.

Considering that the *N* gene is the most sensitive probe in our conditions, it is plausible to ponder that false-negative results especially in samples with low viral load can happen due to these *N* mutations. In fact, a similar problem was recently demonstrated by Hasan and colleagues using another FDA (Food and Drug Administration) approved SARS-CoV-2 test (Cepheid Xpert Xpress SARS-CoV-2) that uses two viral targets, *N* and *E* genes [10].

In their work, at the end of October 2020, a mutation in the SARS-CoV-2 *N* gene was suspected when Xpert failed to amplify the *N* gene target in a specimen, despite giving a strong positive result (Ct = 19.8) for the *E* gene. They also detected three more samples in the next two months and after sequencing they found a point mutation (C29200A) in these samples. An analogous issue was showed by Artesi and colleagues that found a recurrent mutation at position 26,340 associated with failure of the *E* gene detection by the cobas SARS-CoV-2 test (Roche) [11]. Here, we consistently showed five different *N* gene mutations that affect the detection of *N* gene by the GeneFinder Kit in more than 100 sequenced samples being the DEL 28877–28894 in Gamma variant the most frequent. Our results agree with the data showing the *N* gene as one of the most non-conservative genes in the SARS-CoV-2 genome [12].

The main roles of the multifunctional nucleocapsid (N) protein in SARS-CoV-2 include viral genome packaging and virion assembly, viral transcription, regulation of transcription in infected cells and in suppression of the host innate immune response [2,3,13,14]. The N protein has 419 amino acids, divided into two main domains (N-terminal (NTD) and C-terminal (CTD)), with well-known structures [2,4,5] (Figure 2). The NTD RNA-binding and CTD dimerization domains range from amino acids 46–176 and 247–364, respectively. These regions are interspersed with three other domains that are intrinsically disordered [2,4,5,15]. The 203–208 deletion found in protein N was previously identified in Australia and Malaysia [2], suggesting that it may confer some adaptive advantage to SARS-CoV-2. This advantage can be a consequence of alterations that occur at both protein and RNA levels. The linker region was experimentally predicted as fundamental for RNA-mediated phase separation [16]. In fact, this IDR may be directly involved in protein–protein interactions to promote phase separation with RNA. Additionally, the region rich in serine and arginine at this IDR was predicted to model the physical properties of the resulting condensate [3,17]. At the genomic RNA level, it was reported that RNA-sequence distinct regions of the viral RNA genome can promote either phase separation or solubilization [18]. Interestingly, the N protein-coding region is predicted to be a phase separation promoter [18]. In addition, at RNA level, the 203–208 deletion can also directly impact the mRNA structure by changing its functional half-life and affecting the regulation of protein expression [19]. Altogether, it is feasible to suggest that 203–208 deletion may direct impact on the process of phase separation, leading to possible optimization in the process of packaging and replication.

5. Conclusions

Here we demonstrated the existence of mutations in the *N* gene that might affect the use of SARS-CoV2 real-time RT-PCR diagnostic kits impacting the false-negative results. These results provide further evidence that existing variants of SARS-CoV2 might escape molecular detection based on nucleic acid amplification tests, especially those ones using a single target of the virus.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/v13122474/s1>; Table S1: Mutations in gene N detected by sequencing.

Author Contributions: Conceptualization and designed the experiments: H.F.; Performed the analysis: J.C.C.L. and M.D.P.; Analyzed the data: J.C.C.L., H.F., E.C.d.M.O., G.R. and L.G.C.; Writing—original draft preparation: J.C.C.L.; writing—review and edition: J.C.C.L., M.D.P., M.G., S.K. and H.F. Molecular screening and produced SARS-CoV-2 genomic data: J.C.C.L., M.D.P., E.C.d.M.O., J.S.L.P., L.G.C., V.L.V., M.G., L.C.J.d.A., L.P.O.d.L., A.J.M., C.R.d.S.B., E.C.M., J.d.S.T.B., D.B.M., R.A.B., R.d.L.R.C.C., P.D.S.C.M., S.N.S., R.B.d.S., E.S.R., E.V.S., J.S.B., D.G.L.d.L.R., J.P.K., B.S., P.A.A., F.A.d.S.d.C., C.A.B., L.S., M.M.M., M.P., F.E.V.d.S., R.M.T.G., J.A.S.-N., M.L.N., L.L.C., R.T.C., R.M.N., D.T.C., S.C.S., M.C.E., S.K. and H.F. All authors have read and agreed to the published version of the manuscript.

Funding: This study was financed by Butantan Institute on behalf of the Rede de Alerta das Variantes da COVID-19, São Paulo Research Foundation (FAPESP) (Grants Number: 2020/10127-1; 2020/05367-3; and 2013/08135-2), the Central Public Health Laboratories, Blood Center of Ribeirao Preto and supported by the Brazilian Ministry of Health and the Pan American Health Organization PAHO/WHO (APO21-00010098), and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (Grant Number: 401119/2020-3). J.C.C.L. was supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)-Finance Code 001.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Patient consent was waived by the Human Ethics Committee of FZEA-USP due to the urgency and number of tests for COVID-19 diagnosis.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: We thank all the authors who have kindly deposited and shared genome data on GISAID (Table S1). The authors acknowledge the technical support of Luciana de Araujo Pimenta, Luiz Aurelio de Campos Crispin, Gabriela Mauric Frossard Ribeiro, Glaucia Maria Rodrigues Borges, Mariane Evaristo e Josiane Serrano Borges from the National Network for Pandemic Alert of SARS-CoV-2 and the contribution of all employees of General Coordination of Public Health Laboratories and professionals of Public Health Laboratories of Brazil and Network for Pandemic Alert of Emerging SARS-CoV-2 Variants for their contribution toward the sequencing effort and for their commitment and work during the fight of the COVID-19 pandemic.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Baek, M.; Di Maio, F.; Anishchenko, I.; Dauparas, J.; Ovchinnikov, S.; Lee, G.R.; Wang, J.; Cong, Q.; Kinch, L.N.; Schaeffer, R.D.; et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **2021**, *373*, 871–876. [[CrossRef](#)] [[PubMed](#)]
2. Rahman, M.S.; Islam, M.R.; Alam, A.S.M.R.U.; Islam, I.; Hoque, M.N.; Akter, S.; Sultana, M.M.R.M.; Hossain, M.A. Evolutionary dynamics of SARS-CoV-2 nucleocapsid protein and its consequences. *J. Med. Virol.* **2021**, *93*, 2177–2195. [[CrossRef](#)] [[PubMed](#)]
3. Zhao, M.; Yu, Y.; Sun, L.-M.; Xing, J.-Q.; Li, T.; Zhu, Y.; Wang, M.; Yu, Y.; Xue, W.; Xia, T.; et al. GCG inhibits SARS-CoV-2 replication by disrupting the liquid phase condensation of its nucleocapsid protein. *Nat. Commun.* **2021**, *12*, 1–14. [[CrossRef](#)]
4. Ye, Q.; West, A.M.V.; Silletti, S.; Corbett, K.D. Architecture and self-assembly of the SARS-CoV-2 nucleocapsid protein. *Protein Sci.* **2020**, *29*, 1890–1901. [[CrossRef](#)] [[PubMed](#)]
5. Peng, Y.; Du, N.; Lei, Y.; Dorje, S.; Qi, J.; Luo, T.; Gao, G.F.; Song, H. Structures of the SARS-CoV-2 nucleocapsid and their perspectives for drug design. *EMBO J.* **2020**, *39*, e105938. [[CrossRef](#)]
6. Cubuk, J.; Alston, J.J.; Incicco, J.J.; Singh, S.; Stuchell-Brereton, M.D.; Ward, M.D.; Zimmerman, M.I.; Vithani, N.; Griffith, D.; Wagoner, J.A.; et al. The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA. *Nat. Commun.* **2021**, *12*, 1–17. [[CrossRef](#)]
7. Savastano, A.; Opakua, A.I.; de Rankovic, M.; Zweckstetter, M. Nucleocapsid protein of SARS-CoV-2 phase separates into RNA-rich polymerase-containing condensates. *Nat. Commun.* **2020**, *11*, 1–10. [[CrossRef](#)]
8. Zeng, W.; Liu, G.; Ma, H.; Zhao, D.; Yang, Y.; Liu, M.; Mohammed, A.; Zhao, C.; Yang, Y.; Xie, J.; et al. Multiple nucleic acid binding sites and intrinsic disorder of severe acute respiratory syndrome coronavirus nucleocapsid protein: Implications for ribonucleocapsid protein packaging. *J. Virol.* **2009**, *83*, 2255–2264. [[CrossRef](#)]
9. Zeng, W.; Liu, G.; Ma, H.; Zhao, D.; Yang, Y.; Liu, M.; Mohammed, A.; Zhao, C.; Yang, Y.; Xie, J.; et al. Biochemical characterization of SARS-CoV-2 nucleocapsid protein. *Biochem. Biophys. Res. Commun.* **2020**, *527*, 618–623. [[CrossRef](#)] [[PubMed](#)]
10. Hasan, M.R.; Sundararaju, S.; Manickam, C.; Mirza, F.; Al-Hail, H.; Lorenz, S.; Tang, P. A Novel Point Mutation in the N Gene of SARS-CoV-2 May Affect the Detection of the Virus by Reverse Transcription-Quantitative PCR. *J. Clin. Microbiol.* **2021**, *59*, 1–3. [[CrossRef](#)] [[PubMed](#)]
11. Artesi, M.; Bontems, S.; Göbbels, P.; Franckh, M.; Maes, P.; Boreux, R.; Meex, C.; Melin, P.; Hayette, M.-P.; Bours, V.; et al. A Recurrent Mutation at Position 26340 of SARS-CoV-2 Is Associated with Failure of the E Gene Quantitative Reverse Transcription-PCR Utilized in a Commercial Dual-Target Diagnostic Assay. *J. Clin. Microbiol.* **2020**, *58*, 1–8. [[CrossRef](#)]
12. Wang, R.; Hozumi, Y.; Yin, C.; Wei, G.-W. Mutations on COVID-19 diagnostic targets. *Genomics* **2020**, *112*, 5204–5213. [[CrossRef](#)]
13. Nabeel-Shah, S.; Lee, H.; Ahmed, N.; Marcon, E.; Farhangmehr, S.; Pu, S.; Burke, G.L.; Ashraf, K.; Wei, H.; Zhong, G.; et al. SARS-CoV-2 Nucleocapsid protein attenuates stress granule formation and alters gene expression via direct interaction with host mRNAs. *BioRxiv* **2020**. [[CrossRef](#)]
14. Cascarina, S.M.; Ross, E.D. A proposed role for the SARS-CoV-2 nucleocapsid protein in the formation and regulation of biomolecular condensates. *FASEB J.* **2020**, *34*, 9832–9842. [[CrossRef](#)] [[PubMed](#)]

15. Neuman, B.W.; Adair, B.D.; Yoshioka, C.; Quispe, J.D.; Orca, G.; Kuhn, P.; Milligan, R.A.; Yeager, M.; Buchmeier, M.J. Supramolecular architecture of severe acute respiratory syndrome coronavirus revealed by electron cryomicroscopy. *J. Virol.* **2006**, *80*, 7918–7928. [[CrossRef](#)] [[PubMed](#)]
16. Lu, S.; Ye, Q.; Singh, D.; Cao, Y.; Diedrich, J.K.; Yates, J.R.; Villa, E.; Cleveland, D.W.; Corbett, K.D. The SARS-CoV-2 nucleocapsid phosphoprotein forms mutually exclusive condensates with RNA and the membrane-associated M protein. *Nat. Commun.* **2021**, *12*, 502. [[CrossRef](#)] [[PubMed](#)]
17. Carlson, C.R.; Asfaha, J.B.; Ghent, C.M.; Howard, C.J.; Hartooni, N.; Safari, M.; Frankel, A.D.; Morgan, D.O. Phosphoregulation of Phase Separation by the SARS-CoV-2 N Protein Suggests a Biophysical Basis for its Dual Functions. *Mol. Cell* **2020**, *80*, 1092–1103.e4. [[CrossRef](#)] [[PubMed](#)]
18. Iserman, C.; Roden, C.A.; Boerneke, M.A.; Sealfon, R.S.G.; McLaughlin, G.A.; Jungreis, I.; Fritch, E.J.; Hou, Y.J.; Ekena, J.; Weidmann, C.A.; et al. Genomic RNA Elements Drive Phase Separation of the SARS-CoV-2 Nucleocapsid. *Mol. Cell* **2020**, *80*, 1078–1091.e6. [[CrossRef](#)] [[PubMed](#)]
19. Mauger, D.M.; Joseph Cabral, B.; Presnyak, V.; Su, S.V.; Reid, D.W.; Goodman, B.; Link, K.; Khatwani, N.; Reynders, J.; Moore, M.J.; et al. mRNA structure regulates protein expression through changes in functional half-life. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 24075–24083. [[CrossRef](#)] [[PubMed](#)]