

## RESEARCH ARTICLE

# SARS-CoV-2 genetic diversity in Venezuela: Predominance of D614G variants and analysis of one outbreak

Carmen L. Loureiro<sup>1</sup>, Rossana C. Jaspe<sup>1</sup>, Pierina D'Angelo<sup>2</sup>, José L. Zambrano<sup>3</sup>, Lieska Rodriguez<sup>2</sup>, Víctor Alarcon<sup>2</sup>, Mariangel Delgado<sup>4</sup>, Marwan Aguilar<sup>2</sup>, Domingo Garzaro<sup>1</sup>, Héctor R. Rangel<sup>1</sup>, Flor H. Pujol<sup>1\*</sup>

**1** Laboratorio de Virología Molecular, Centro de Microbiología y Biología Celular, Instituto Venezolano de Investigaciones Científicas, Caracas, Miranda, Venezuela, **2** Instituto Nacional de Higiene "Rafael Rangel", Caracas, Miranda, Venezuela, **3** Laboratorio de Biología de Virus, Centro de Microbiología y Biología Celular, Instituto Venezolano de Investigaciones Científicas, Caracas, Miranda, Venezuela, **4** Unidad Unidad de Microscopia Electrónica y Confocal, Centro de Microbiología y Biología Celular, Instituto Venezolano de Investigaciones Científicas, Caracas, Miranda, Venezuela

\* [fhpujol@gmail.com](mailto:fhpujol@gmail.com)



## OPEN ACCESS

**Citation:** Loureiro CL, Jaspe RC, D'Angelo P, Zambrano JL, Rodriguez L, Alarcon V, et al. (2021) SARS-CoV-2 genetic diversity in Venezuela: Predominance of D614G variants and analysis of one outbreak. PLoS ONE 16(2): e0247196. <https://doi.org/10.1371/journal.pone.0247196>

**Editor:** Ahmed S. Abdel-Moneim, Taif University, SAUDI ARABIA

**Received:** October 10, 2020

**Accepted:** February 2, 2021

**Published:** February 19, 2021

**Peer Review History:** PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0247196>

**Copyright:** © 2021 Loureiro et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** Nucleotide sequence data have been deposited into the GenBank database under the accession numbers

## Abstract

SARS-CoV-2 is the new coronavirus responsible for COVID-19 disease. The first two cases of COVID-19 were detected in Venezuela on March 13, 2020. The aim of this study was the genetic characterization of Venezuelan SARS-CoV-2 isolates. A total of 7 full SARS-CoV-2 genome sequences were obtained by Sanger sequencing, from patients of different regions of Venezuela, mainly from the beginning of the epidemic. Ten out of 11 isolates (6 complete genomes and 4 partial spike genomic regions) belonged to lineage B, bearing the D614G mutation in the Spike protein. Isolates from the first outbreak that occurred in the Margarita Island harbored an in-frame deletion in its sequence, without amino acids 83–85 of the NSP1 of the ORF1. The search for deletions in 48,635 sequences showed that the NSP1 gene exhibit the highest frequency of deletions along the whole genome. Structural analysis suggests a change in the N-terminal domain with the presence of this deletion. In contrast, isolates circulating later in this island lacked the deletion, suggesting new introductions to the island after this first outbreak. In conclusion, a high diversity of SARS-CoV-2 isolates were found circulating in Venezuela, with predominance of the D614G mutation. The first small outbreak in Margarita Island seemed to be associated with a strain carrying a small deletion in the NSP1 protein, but these isolates do not seem to be responsible for the larger outbreak which started in July.

## Introduction

On March 11, the World Health Organization declared a pandemic COVID-19, the disease caused by a new coronavirus which emerged in Wuhan at the end of 2019. Since then, at the end of January 2021, more than 100 million persons worldwide have contracted the disease, with more than 2,000,000 deaths [1]. Once originated in China, the epidemic center evolved

MT907515-MT907521, MW015946-MW015954 and MW040500-MW040503.

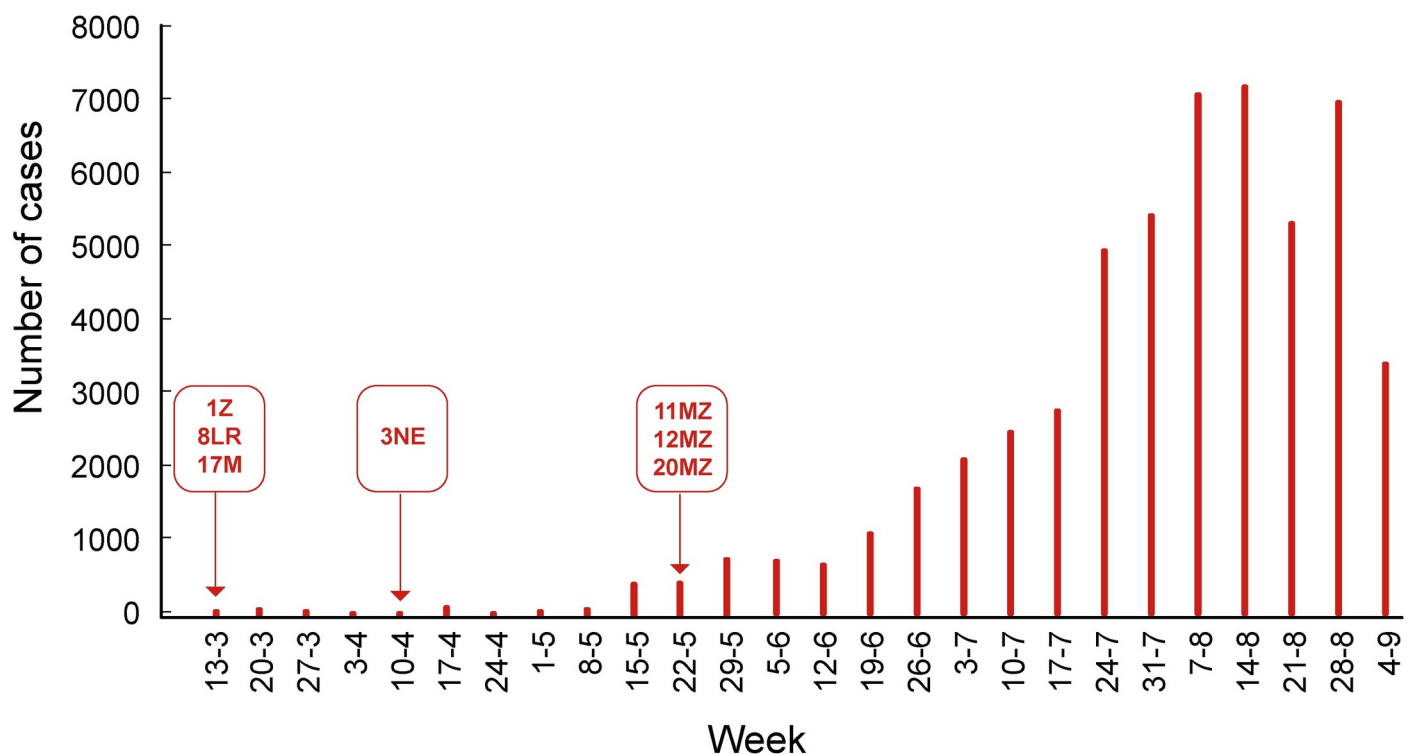
**Funding:** This study was supported by Ministerio del Poder Popular de Ciencia, Tecnología e Innovación of Venezuela.

**Competing interests:** The authors have declared that no competing interests exist.

first in Asia, then in Europe, and in August 2020, the epicenter was located in the Americas, including Latin America.

SARS-CoV-2, the new coronavirus responsible for COVID-19 disease, belongs to the family *Coronaviridae*, genus *Betacoronavirus*, and shares with its predecessor SARS-CoV the subgenera *Sarbecovirus*. After interaction of the Receptor Binding Domain (RBD) of the viral Spike protein with the viral receptor ACE2, SARS-CoV-2 enters the cell through a fusion mechanism. The viral fusion protein is exposed by priming with a proteolytic enzyme [2]. The viruses of this family possess the largest known continuous RNA genome (around 30,000 nt). The RNA-dependent RNA polymerase complex (RdRP) possess proofreading capacity, implying that the mutation rate of these viruses is lower than the one observed for other RNA viruses [3, 4]. On the other hand, these viruses display also a high frequency of recombination, horizontal gene transfer, gene duplication, and alternative open reading frames, which allow them to jump from one species to another [5].

The first case of COVID-19 was detected in Venezuela on March 13, 2020 (Fig 1). The first viral introductions were mainly through air flight passengers, mainly from Europe, during the month of March. Immediately after the two first cases, quarantine and the use of face mask was mandatory. The second wave of viral introductions started in April and includes migrants returning from neighboring countries, mainly Colombia and Brazil. The aim of this study was the genetic characterization of Venezuelan SARS-CoV-2 isolates and the analysis of the ongoing epidemic in two settings: the Margarita Island and the second most populous city, Maracaibo.



**Fig 1. COVID-19 cases in Venezuela, according to official data.** The cases are referred to as the date of report and downloaded from the web page <http://www.ucv.ve/cendes>. The arrows indicate the date of collection of the samples for which complete genome sequences were obtained.

<https://doi.org/10.1371/journal.pone.0247196.g001>

## Materials and methods

### Amplification and sequencing

Clinical samples were processed at the Instituto Nacional de Higiene “Rafael Rangel”, as part of the national diagnostic response to the COVID-19 pandemic, and sent as anonymous samples to Laboratorio de Virología Molecular, Instituto Venezolano de Investigaciones Científicas (IVIC), for sequencing. The samples were chosen according to the following criteria: from the beginning of the epidemic (for complete genome sequencing), from different locations of the country, and exhibiting low Ct values for efficient PCR amplification. This study was approved by the Human Bioethical Committee of IVIC.

Total RNA was extracted from nasopharyngeal or nasal swabs samples using the QIAamp Viral RNA Mini Kit (Hilden, Germany). Nested RT-PCR was carried out using the SuperScript III One-Step RT-PCR Platinum Taq HiFi System (Invitrogen, Thermo Fisher Scientific, USA) with specific Artic primers [6] to produce 3 cDNA/tube in six reactions. In some cases cDNA was generated using SuperScript IV VILO Master Mix with ezDNAase (Invitrogen, Thermo Fisher Scientific, Lithuania). The first and second rounds of PCR were performed using high-fidelity *Taq* platinum DNA polymerase (Invitrogen, Thermo Fisher Scientific, USA) to generate overlapping amplicons, of approximately 2000 nt and 1000 nt respectively. These amplicons cover the entire genome. PCR purified fragments were sent to Macrogen Sequencing Service (Macrogen, Korea) for sequencing. Both strands of DNA were sequenced.

Nucleotide sequence data have been deposited into the GenBank database under the accession numbers MT907515-MT907521, MW015946-MW015954 and MW040500-MW040503.

### Phylogenetic analysis

Sequence alignment was performed by the MUSCLE algorithm and the phylogenetic analysis by the Maximum Likelihood method (1000 bootstrap replicas) with MEGA7 [7]. The best-fitting nucleotide substitution model was chosen according to the Akaike Information Criterion (AIC). The phylogenetic inference was also performed with the Bayesian criterion (TN93). The results were similar, but the tree generated with the GTR+G model exhibited better bootstrap values. The evolutionary history was inferred by using the Maximum Likelihood method and the General Time Reversible model. The tree with the highest log likelihood (-42092.99) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (3 categories (+G, parameter = 0.05000)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This analysis involved 42 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated (complete deletion option). There were a total of 29723 positions in the final dataset. Reference sequences from different lineages [8] were included in the phylogenetic analysis and lineages assigned with the Pangolin COVID-19 Lineage Assigner (<https://pangolin.cog-uk.io/>).

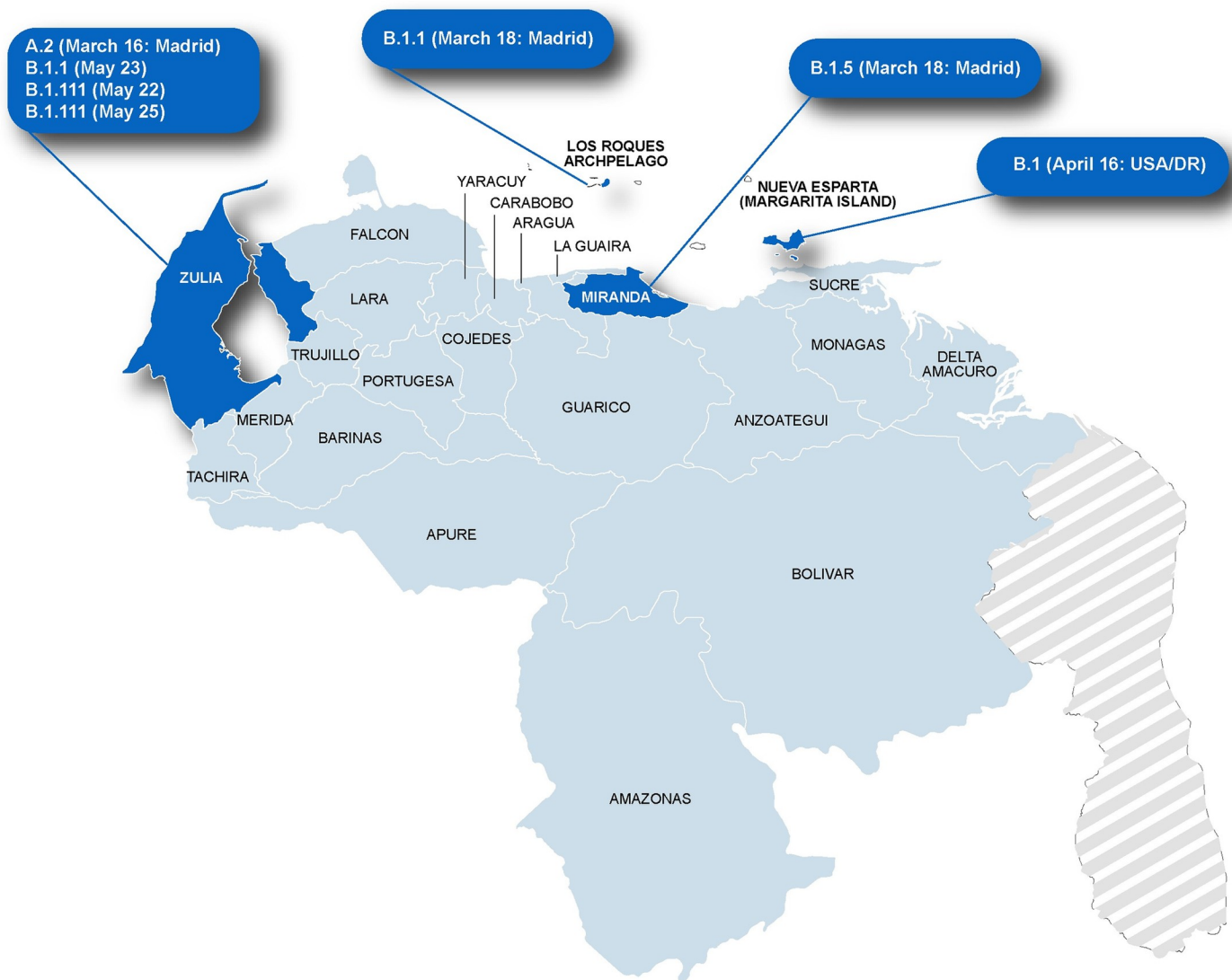
### Protein structure modelling

NSP1 protein sequences from NCO45512\_Wuhan1, MT344955\_2 aa del, and 3NE\_3 aa del were modelled at SWISS-MODEL servers. NSP1 structures were modeled based on the 7K3N Crystal Structure of NSP1 from SARS-CoV-2 from PDB database as a template used to

generate homology structures. Homology structures were analyzed and superimposed by using UCSF Chimera (*ver.* 1.4).

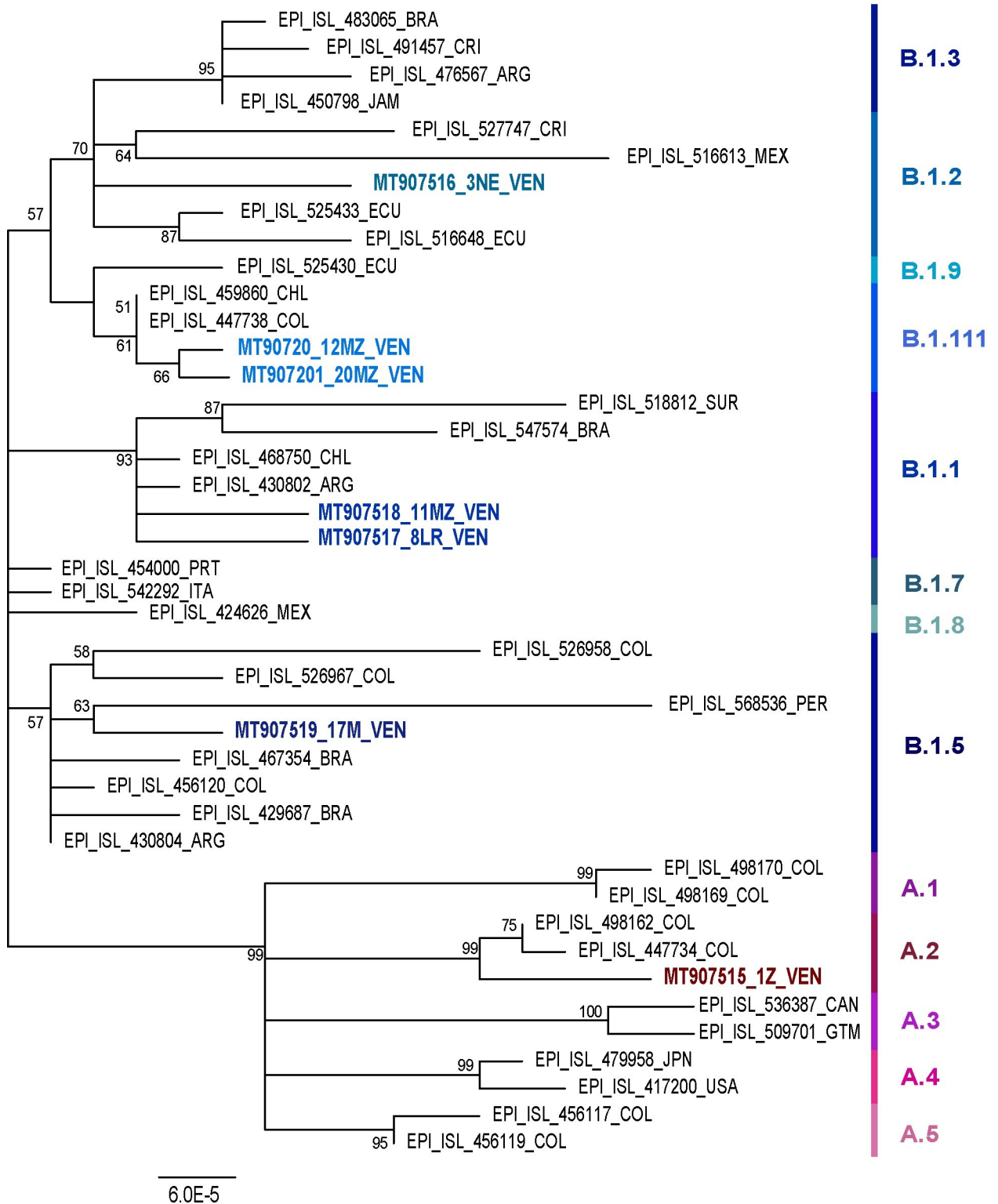
## Results

A total of 7 full genome sequences were obtained from different regions of Venezuela, mainly from the beginning of the epidemic (Figs 1 and 2). One isolate (the oldest studied one, collected on March 16, 2020) belonged to lineage A.2 (bearing a D in position 614 in the spike). The other 6 isolates belonged to lineage B: two B.1, two B.1.1, one B.1.117 and one B.1.5, all with the mutation D614G (Figs 2 and 3). Other 4 isolates from Maracaibo city (Zulia State), for which a 1000 nt sequence surrounding position 23403 was available, also bear the mutation



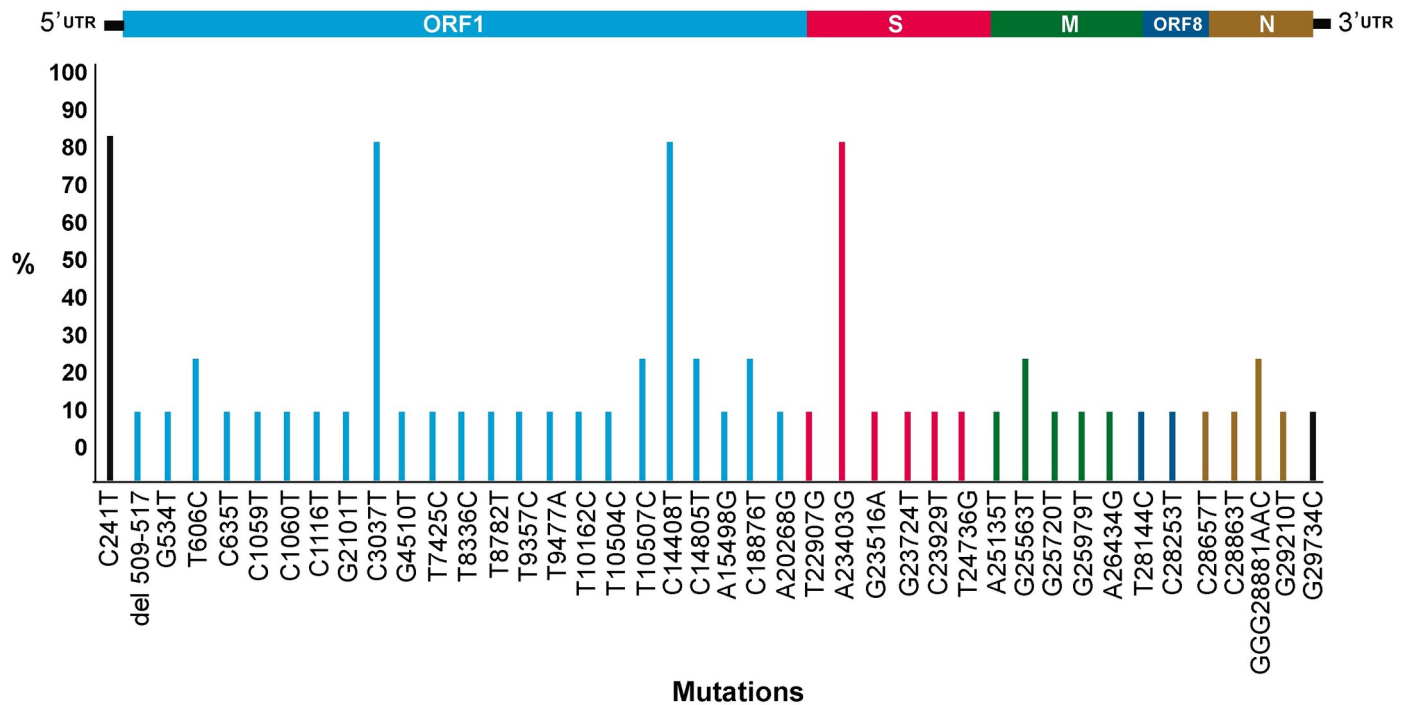
**Fig 2. SARS-CoV-2 Venezuelan isolate lineages according to the place of collection.** The cases are referred to as the date of collection. The country of provenance is indicated for the cases with a recent history of travel and who got probably infected in this country. In the case of Margarita Island, the cases were due to contact with a person coming from the USA and who spent some days in the Dominican Republic (DR). Map of Venezuela was edited and modified from <https://freevectormaps.com>.

<https://doi.org/10.1371/journal.pone.0247196.g002>



**Fig 3. Phylogenetic tree of Venezuelan complete genome isolates.** The evolutionary history was inferred by using the Maximum Likelihood method (500 bootstrap replicas) and the General Time Reversible model. Sequences are shown by their GenBank accession number or GISAID Initiative (<https://www.gisaid.org>) identifier, and country of origin. Venezuelan samples are shown in colors with their isolate name. Lineages are shown in different colors. CRI: Costa Rica, MEX: Mexico, ECU: Ecuador, VEN: Venezuela, ARG: Argentina, BRA: Brazil, JAM: Jamaica, CHL: Chile, COL: Colombia, PER: Peru, CAN: Canada, GTM: Guatemala, JPN: Japan, USA: United States of America, ITA: Italy, SUR: Surinam, PRT: Portugal.

<https://doi.org/10.1371/journal.pone.0247196.g003>



**Fig 4. Frequency of mutations found in Venezuelan SARS-CoV-2 isolates compared to Wuhan MN908947 original isolate.** The position of the mutation in the viral genome is shown.

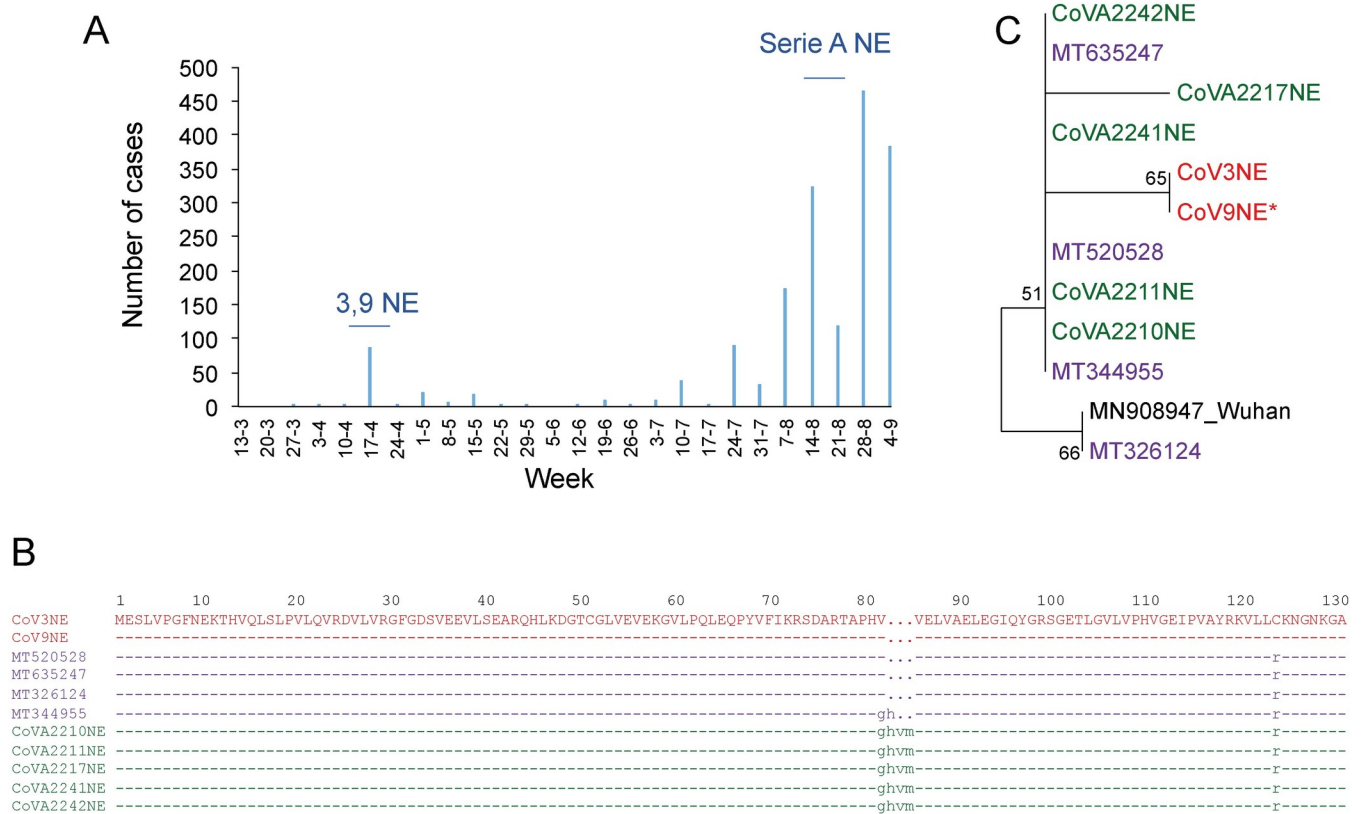
<https://doi.org/10.1371/journal.pone.0247196.g004>

D614G (Accession numbers MW040500-MW040503), for a total of 10/11 D614G isolates. The samples obtained at the beginning of the epidemic (March 2020) were from individuals returning from Spain, except the 3NE sample collected in Margarita Island (Nueva Esparta State). In contrast, the other samples from Maracaibo were collected in May 2020 and were even community cases or from migrants from Colombia (Fig 2). These samples were indeed related to isolates from Colombia (Fig 3).

Each sample harbored a distinct mutation pattern along the whole genome and exhibited a different sequence (Figs 3 and 4). In addition to the D614G mutation in the spike protein characteristic of lineage B (A23403G), all the Venezuelan sequences belonging to lineage B bore the mutation C241T in the UTR, C14408T in the RdRP, and the synonymous mutation C3037T (Fig 4).

Isolate CoV3NE harbors an in-frame deletion in its sequence, lacking amino acids 83–85 of the NSP1 of the ORF1. Another sample from the same outbreak also exhibited the deletion in its sequence. Similar deletions in this region have been reported in sequences from USA (Fig 5B). The March outbreak of COVID-19 in Margarita Island (Fig 5A) had its origin among a group of young baseball players from Venezuelan 's from League A, who had contact with an American baseball scout who came from Philadelphia and traveled to the Dominican Republic before arriving at the Venezuelan island.

Mercatelli *et al.* [9] described the mutations present in 48,635 sequences available in GISAID Initiative database. We analyzed the number of isolates harboring deletions in these sequences, and found that 1,148 sequences out of these 48,635 exhibited at least one deletion. A total of 5,177 deletions were present in these 1,148 sequences, implying an average of 4.5 deletions in these sequences. More than half of these deletions (2,834) were present in the NSP1



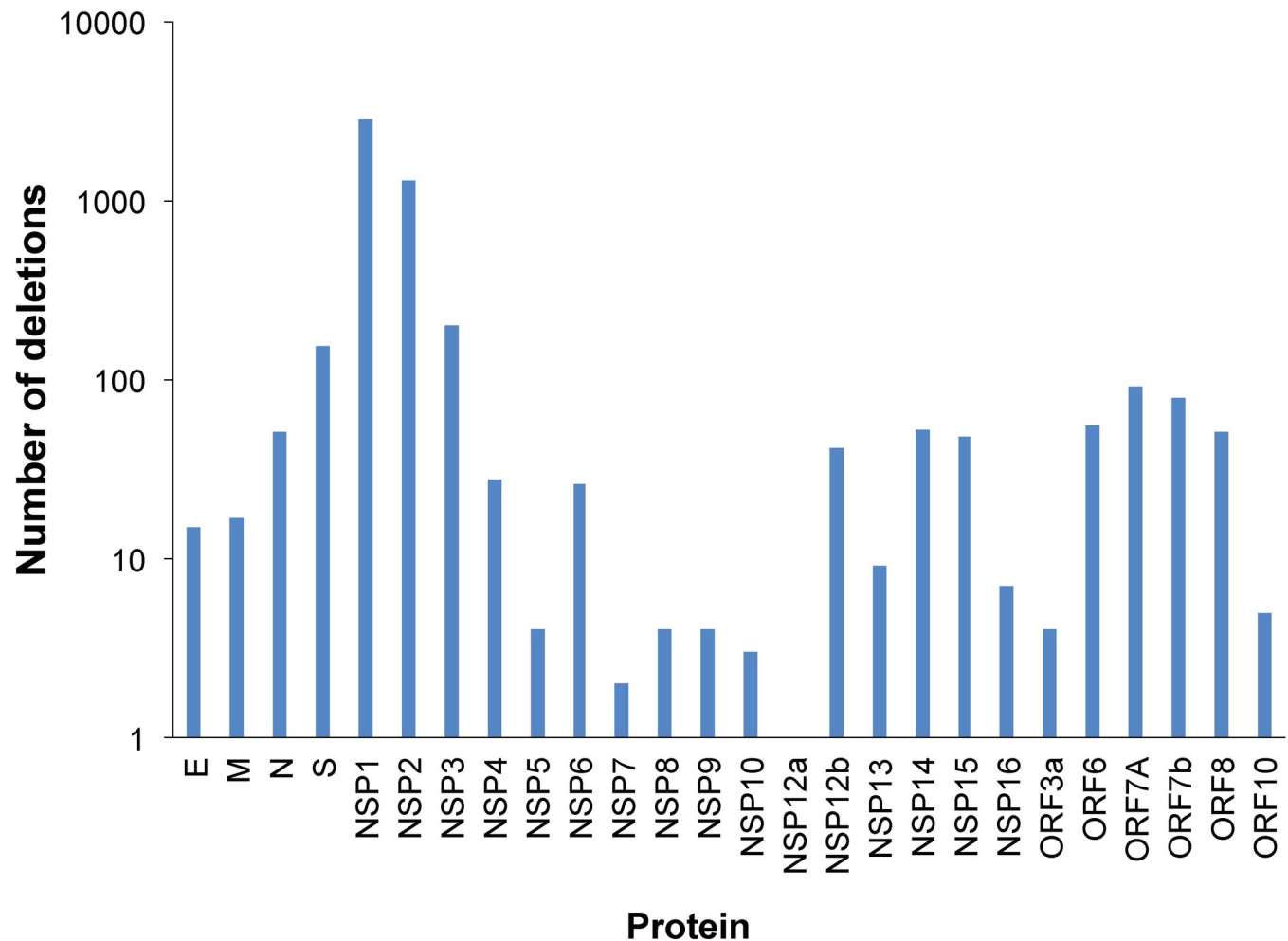
**Fig 5. Analysis of Margarita Island's outbreak, Nueva Esparta State, Venezuela. A:** COVID-19 cases in Nueva Esparta, according to official data. The cases are referred to as the date of the report and were downloaded from the web page <http://www.ucv.ve/cendes>. The horizontal lines indicate the period of time when the samples analyzed were collected. **B:** Phylogenetic tree of the N-terminal region of SARS-CoV-2 genome (nt 40–1946). Genetic distance was estimated by observed divergence, and the phylogenetic tree was constructed with the neighbor-joining method. Samples from Margarita Island (NE) are in red (isolates with deletion) and pink (isolates without deletion), as in C. Samples in blue are from other regions of Venezuela. \*Only 812 nt were available for this sequence. **C:** Amino acid alignment of the N-terminal region of the SARS-CoV-2 NSP1 protein, showing the deletion in NE Venezuelan isolates. Other isolates are named according to their GenBank accession number.

<https://doi.org/10.1371/journal.pone.0247196.g005>

protein (Fig 6). Deletions in the region comprising amino acids 82–85 of NSP1, similar to the one found in this study (Fig 5C), were the most frequent among these sequences.

As noted in Fig 5A, COVID-19 epidemics increased again in Margarita Island since July 2020. In order to evaluate if the strain with the deletion in NSP1 was still the responsible for the new cases, the first 1,900 nt were sequenced in isolates from August. The sequence of 1,900 nt from the 5' end of the genome of 5 isolates was obtained. None of them carry the deletion, and their sequences were different to the one found in the isolates originally circulating in the Island (Fig 5B and 5C). The 5 sequences exhibit at least one nt difference in these 1900 nt, suggesting multiple viral introductions to the Island after July. Other sequences exhibiting a deletion of 2 or 3 amino-acids at the same position were available in the GenBank. These sequences were not related however, to the one found in the isolates from Margarita Island (Fig 5B and 5C), suggesting independent deletion events in these isolates.

The structure of the SARS-CoV-2 NSP1 protein with the 83–85 aa deletion (3NE\_3 aa del) was modeled by structure homology in SWISS-MODEL. Structure comparison between 3NE\_3 aa del with the NSP1 protein of the first reported isolate of SARS-CoV-2 (NCO45512\_Wuhan1) and the NSP1 protein structure with 84–85 aa deletion (MT344955\_2 aa del) showed a structural change (Fig 7A). However, the superimposed NSP1 protein structures (Fig 7B)



**Fig 6. Analysis of deletions in 48,635 sequences of SARS-CoV-2 isolates.** Number of sequences (out of the 48,635 analyzed by Mercatelli et al. [9]) harboring deletions in a viral protein.

<https://doi.org/10.1371/journal.pone.0247196.g006>

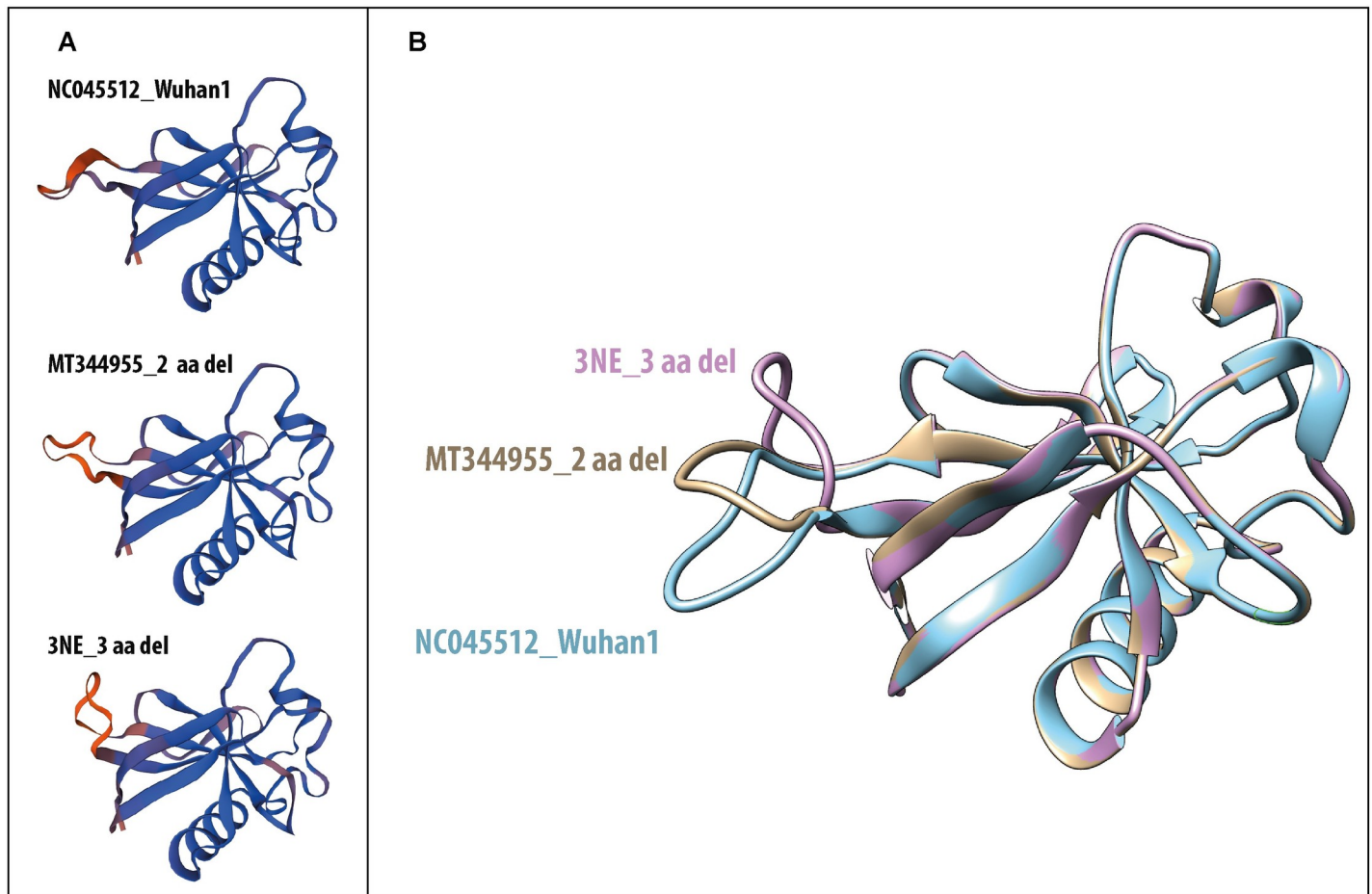
show that these mutations do not affect the core of the protein (elements of the secondary helical structures) and may indicate that this deletion is located in a flexible region.

## Discussion

High diversity was found among the Venezuelan viral isolates sequenced from the early period of the Venezuelan epidemic. The sequences were similar to the one found in other countries of Latin America, with the predominance of lineage B isolates [11–14].

All the lineage B Venezuelan sequences bore the mutation C14408T, which produces the amino acid mutation P314L in the Nsp12 protein, the polymerase in the RdRP. The genomes bearing this mutation were initially associated with significantly higher number of mutations compared to the non-mutated ones [15]. The T14408 is now the most common nucleotide found in the majority of SARS-CoV-2 isolates around the world [16]. As previously described [9], C/T was the most frequent substitution in these complete genome sequences (Table 1). The most probable explanation for this relative abundance is that these genomes are subjected to APOBEC and/or ADAR deaminase activity [17, 18].





**Fig 7. Modelling of SARS-CoV-2 NSP1 protein with and without deletions.** **A:** Comparison of complete NSP1 protein structures between the NSP1 protein structure from the first SARS-CoV-2 isolate reported (NC045512\_Wuhan1), and NSP1 protein structure with two amino acid residues deleted (MT344955\_2 aa del), and NSP1 protein structure reported in this study with three amino acids deleted (3NE\_3 aa del). 7K3N PDB Crystal Structure of NSP1 protein from SARS-CoV-2 was used as a template. **B:** Superimposed structures of the NSP1 proteins from NC045512\_Wuhan1 (Cyan), MT344955\_2 aa del (Yellow), and 3NE\_3 aa del (Pink). Swiss-Model structure homology-modeling servers was used to generate NSP1 SARS-CoV-2 protein structures [10].

<https://doi.org/10.1371/journal.pone.0247196.g007>

Another non-synonymous mutation A23403G, yielding the mutation D614G in the Spike protein, has been widely described. This mutation is not located in the RBD of the Spike protein, the specific region which interacts with the ACE2 receptor. However, *in vitro* studies with pseudotyped viruses expressing the SARS-CoV-2 spike showed that this mutation facilitated the entry of the virus into the cell, probably through favoring the fusion process. It has been suggested that this variant is more easily transmitted, which might account for its present predominance in the new epidemic foci [11, 19]. This variant has also displaced the original one in new infections in many parts of the world [11]. The higher transmissibility nature of the isolates carrying this mutation is not completely demonstrated, however, and there is no evidence of higher severity associated with this variant [20].

Several deletions have been described in SARS-CoV-2, probably due to the recombinogenic capacity of this viral family [9, 21–24]. As shown in this study, deletions were significantly more frequent in the NSP1 protein, and particularly in the region deleted in the Venezuelan isolates from Margarita Island. NSP1 is the first protein coded at the 5' UTR region of the SARS-CoV-2. This protein is a virulence factor, which plays a role in suppressing host innate

immune functions by inhibiting the expression of type I IFN, inhibiting host protein synthesis, and promoting the degradation of host mRNAs [25, 26]. Some deletions in the NSP1 protein of SARS-CoV have been shown to attenuate this virus. Particularly, a deletion of 10 amino acids in SARS-CoV, comprising the 3 amino acids deleted in the CoV3NE isolate, exhibited a small reduction in viral titer when infecting delayed brain tumor (DBT) mouse cells expressing the murine ACE2 receptor. Nevertheless, attenuation of the viral strain was not observed with this deletion, but with others nearest the C-terminal region of the protein [27]. No fatalities were observed in this first outbreak in Margarita Island. However, it is important to consider that many of the infected persons were young. Structural analysis in NSP1 protein showed a change in the N-terminal domain with the presence of 2 or 3 amino-acid deletions. Both the 2 and 3 amino-acid deletions seemed to induce a structural change in a flexible region of the NSP1 protein. This region has been shown not to play an important role of cellular protein translation [28], but has been suggested to play a role in the regulation of cellular mRNA stability or in suppressing host innate immune functions. More studies are needed to evaluate if the deletion found in the CoV3NE isolate affects the replication capacity of this virus. Another deletion in NSP1 protein has been found circulating in some countries in Europe, the USA, and Brazil [24]. This deletion is located in the C-terminal region of the NSP1 protein, and may affect the activity of this protein in the host's gene expression regulation.

The largest deletion described so far in the SARS-CoV-2 genome is located in the ORF8. Since a strong immune response has been observed against this protein, this deletion was originally associated with immune evasion [29]. The function of ORF8 is not completely clear. This protein may be associated with immune evasion according to preliminary evidence. In Singapore, although patients infected with viruses carrying this deletion developed clinically significant illness, infections tended to be milder compared with those caused by the wild-type virus, with less pronounced cytokine release during the acute phase of infection [30].

As expected, high diversity was observed between the different isolates from Maracaibo city, since a significant number of introductions shaped the ongoing outbreak in this city. The same seems to occur since July in Margarita Island, after a probable single introduction of the deleted genomic strain in the first wave of the epidemic. This first strain was introduced by air travel to the island, but since then, air connections have been reduced drastically. The apparent diversity of SARS-CoV-2 isolates circulating in this second wave of the epidemic, might have been introduced by contacts with continental Venezuelan inhabitants, traveling by boat to the island.

In conclusion, a high diversity of SARS-CoV-2 isolates was found circulating in Venezuela, with the predominance of the D614G isolates. The first small outbreak in Margarita Island seemed to be associated with a strain carrying a small deletion in the NSP1 protein, but these isolates do not seem to be responsible for the larger outbreak which started in July.

## Author Contributions

**Conceptualization:** Carmen L. Loureiro, Rossana C. Jaspe, Domingo Garzaro, Héctor R. Rangel, Flor H. Pujol.

**Data curation:** Carmen L. Loureiro, Rossana C. Jaspe, José L. Zambrano, Domingo Garzaro, Héctor R. Rangel, Flor H. Pujol.

**Formal analysis:** Carmen L. Loureiro, Rossana C. Jaspe, José L. Zambrano, Héctor R. Rangel, Flor H. Pujol.

**Funding acquisition:** Héctor R. Rangel, Flor H. Pujol.

**Investigation:** Carmen L. Loureiro, Rossana C. Jaspe, Pierina D'Angelo, José L. Zambrano, Lieska Rodriguez, Víctor Alarcon, Mariangel Delgado, Marwan Aguilar, Domingo Garzaro, Flor H. Pujol.

**Methodology:** Carmen L. Loureiro, Rossana C. Jaspe, Pierina D'Angelo, José L. Zambrano, Lieska Rodriguez, Domingo Garzaro, Héctor R. Rangel, Flor H. Pujol.

**Project administration:** Héctor R. Rangel, Flor H. Pujol.

**Resources:** Carmen L. Loureiro, Lieska Rodriguez, Víctor Alarcon, Héctor R. Rangel, Flor H. Pujol.

**Software:** Carmen L. Loureiro, Rossana C. Jaspe, José L. Zambrano, Héctor R. Rangel, Flor H. Pujol.

**Supervision:** Héctor R. Rangel, Flor H. Pujol.

**Validation:** Carmen L. Loureiro, Rossana C. Jaspe, Héctor R. Rangel, Flor H. Pujol.

**Visualization:** Carmen L. Loureiro, Rossana C. Jaspe, Héctor R. Rangel, Flor H. Pujol.

**Writing – original draft:** Flor H. Pujol.

**Writing – review & editing:** Carmen L. Loureiro, Rossana C. Jaspe, Pierina D'Angelo, José L. Zambrano, Lieska Rodriguez, Víctor Alarcon, Mariangel Delgado, Marwan Aguilar, Domingo Garzaro, Héctor R. Rangel, Flor H. Pujol.

## References

1. Coronavirus disease (COVID-19). [cited 4 Aug 2020]. Available: [https://www.who.int/emergencies/diseases/novel-coronavirus-2019?gclid=CjwKCAjwjLD4BRaiEiwAg5NBFiyd2Q5YJU3kGbD6UQkKmA582H5bl\\_pF1qKtxGD8xmWVhVNFvSE6SRoC6ecQAvD\\_BwE](https://www.who.int/emergencies/diseases/novel-coronavirus-2019?gclid=CjwKCAjwjLD4BRaiEiwAg5NBFiyd2Q5YJU3kGbD6UQkKmA582H5bl_pF1qKtxGD8xmWVhVNFvSE6SRoC6ecQAvD_BwE)
2. Ortega JT, Zambrano JL, Jastrzebska B, Liprandi F, Rangel HR, Pujol FH. Understanding Severe Acute Respiratory Syndrome Coronavirus 2 Replication to Design Efficient Drug Combination Therapies. *Intervirology*. 2020; 1–8. <https://doi.org/10.1159/000512141> PMID: 33099545
3. Denison MR, Graham RL, Donaldson EF, Eckerle LD, Baric RS. Coronaviruses: an RNA proofreading machine regulates replication fidelity and diversity. *RNA Biol*. 2011; 8: 270–279. <https://doi.org/10.4161/rna.8.2.15013> PMID: 21593585
4. Menachery VD, Graham RL, Baric RS. Jumping species—a mechanism for coronavirus persistence and survival. *Curr Opin Virol*. 2017; 23: 1–7. <https://doi.org/10.1016/j.coviro.2017.01.002> PMID: 28214731
5. Peck KM, Burch CL, Heise MT, Baric RS. Coronavirus Host Range Expansion and Middle East Respiratory Syndrome Coronavirus Emergence: Biochemical Mechanisms and Evolutionary Perspectives. *Annu Rev Virol*. 2015; 2: 95–117. <https://doi.org/10.1146/annurev-virology-100114-055029> PMID: 26958908
6. joshquick/artic-ncov2019. In: GitHub [Internet]. [cited 4 Aug 2020]. Available: <https://github.com/joshquick/artic-ncov2019>
7. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol*. 2018; 35: 1547–1549. <https://doi.org/10.1093/molbev/msy096> PMID: 29722887
8. Rambaut A, Holmes EC, O'Toole Á, Hill V, McCrone JT, Ruis C, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol*. 2020. <https://doi.org/10.1038/s41564-020-0770-5> PMID: 32669681
9. Mercatelli D, Giorgi FM. Geographic and Genomic Distribution of SARS-CoV-2 Mutations. *Front Microbiol*. 2020; 11. <https://doi.org/10.3389/fmicb.2020.01800> PMID: 32793182
10. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res*. 2018; 46: W296–W303. <https://doi.org/10.1093/nar/gky427> PMID: 29788355
11. Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, et al. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell*. 2020; 182: 812–827.e19. <https://doi.org/10.1016/j.cell.2020.06.043> PMID: 32697968

12. Xavier J, Giovanetti M, Adelino T, Fonseca V, Barbosa da Costa AV, Ribeiro AA, et al. The ongoing COVID-19 epidemic in Minas Gerais, Brazil: insights from epidemiological data and SARS-CoV-2 whole genome sequencing. *Emerg Microbes Infect.* 2020; 9: 1824–1834. <https://doi.org/10.1080/22221751.2020.1803146> PMID: 32726185
13. Castillo AE, Parra B, Tapia P, Lagos J, Arata L, Acevedo A, et al. Geographical Distribution of Genetic Variants and Lineages of SARS-CoV-2 in Chile. *Front Public Health.* 2020; 8. <https://doi.org/10.3389/fpubh.2020.562615> PMID: 33072699
14. Franco-Muñoz C, Álvarez-Díaz DA, Laiton-Donato K, Wiesner M, Escandón P, Usme-Ciro JA, et al. Substitutions in Spike and Nucleocapsid proteins of SARS-CoV-2 circulating in South America. *Infect Genet Evol.* 2020; 104557. <https://doi.org/10.1016/j.meegid.2020.104557> PMID: 32950697
15. Pachetti M, Marini B, Benedetti F, Giudici F, Mauro E, Storici P, et al. Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. *J Transl Med.* 2020; 18: 179. <https://doi.org/10.1186/s12967-020-02344-6> PMID: 32321524
16. UGUREL OM, ATA O, TURGUT-BALIK D. An updated analysis of variations in SARS-CoV-2 genome. *Turk J Biol.* 2020; 44: 157–167. <https://doi.org/10.3906/biy-2005-111> PMID: 32595352
17. Di Giorgio S, Martignano F, Torcia MG, Mattiuz G, Conticello SG. Evidence for host-dependent RNA editing in the transcriptome of SARS-CoV-2. *Sci Adv.* 2020; 6: eabb5813. <https://doi.org/10.1126/sciadv.abb5813> PMID: 32596474
18. Simmonds P. Rampant C→U Hypermutation in the Genomes of SARS-CoV-2 and Other Coronaviruses: Causes and Consequences for Their Short- and Long-Term Evolutionary Trajectories. Schwemmler M, editor. *mSphere.* 2020; 5: e00408–20, /msphere/5/3/mSphere408-20.atom. <https://doi.org/10.1128/mSphere.00408-20> PMID: 32581081
19. Johnson MC, Lyddon TD, Suarez R, Salcedo B, LePique M, Graham M, et al. Optimized pseudotyping conditions for the SARS-CoV-2 Spike glycoprotein. *J Virol.* 2020. <https://doi.org/10.1128/JVI.01062-20> PMID: 32788194
20. Grubaugh ND, Hanage WP, Rasmussen AL. Making Sense of Mutation: What D614G Means for the COVID-19 Pandemic Remains Unclear. *Cell.* 2020; 182: 794–795. <https://doi.org/10.1016/j.cell.2020.06.040> PMID: 32697970
21. Koyama T, Platt D, Parida L. Variant analysis of SARS-CoV-2 genomes. *Bull World Health Organ.* 2020; 98: 495–504. <https://doi.org/10.2471/BLT.20.253591> PMID: 32742035
22. Islam MR, Hoque MN, Rahman MS, Alam ASMRU, Akther M, Puspo JA, et al. Genome-wide analysis of SARS-CoV-2 virus strains circulating worldwide implicates heterogeneity. *Sci Rep.* 2020; 10: 14004. <https://doi.org/10.1038/s41598-020-70812-6> PMID: 32814791
23. Phan T. Genetic diversity and evolution of SARS-CoV-2. *Infect Genet Evol.* 2020; 81: 104260. <https://doi.org/10.1016/j.meegid.2020.104260> PMID: 32092483
24. Benedetti F, Snyder GA, Giovanetti M, Angeletti S, Gallo RC, Ciccozzi M, et al. Emerging of a SARS-CoV-2 viral strain with a deletion in nsp1. *Journal of Translational Medicine.* 2020; 18: 329. <https://doi.org/10.1186/s12967-020-02507-5> PMID: 32867854
25. Lokugamage KG, Narayanan K, Huang C, Makino S. Severe acute respiratory syndrome coronavirus protein nsp1 is a novel eukaryotic translation inhibitor that represses multiple steps of translation initiation. *J Virol.* 2012; 86: 13598–13608. <https://doi.org/10.1128/JVI.01958-12> PMID: 23035226
26. Thoms M, Buschauer R, Ameisemeier M, Koepke L, Denk T, Hirschenberger M, et al. Structural basis for translational shutdown and immune evasion by the Nsp1 protein of SARS-CoV-2. *Science.* 2020; 369: 1249–1255. <https://doi.org/10.1126/science.abc8665> PMID: 32680882
27. Jimenez-Guardeño JM, Regla-Nava JA, Nieto-Torres JL, DeDiego ML, Castaño-Rodríguez C, Fernandez-Delgado R, et al. Identification of the Mechanisms Causing Reversion to Virulence in an Attenuated SARS-CoV for the Design of a Genetically Stable Vaccine. Drosten C, editor. *PLoS Pathog.* 2015; 11: e1005215. <https://doi.org/10.1371/journal.ppat.1005215> PMID: 26513244
28. Schubert K, Karousis ED, Jomaa A, Scaiola A, Echeverria B, Gurzeler L-A, et al. SARS-CoV-2 Nsp1 binds the ribosomal mRNA channel to inhibit translation. *Nat Struct Mol Biol.* 2020; 27: 959–966. <https://doi.org/10.1038/s41594-020-0511-8> PMID: 32908316
29. Su YCF, Anderson DE, Young BE, Linster M, Zhu F, Jayakumar J, et al. Discovery and Genomic Characterization of a 382-Nucleotide Deletion in ORF7b and ORF8 during the Early Evolution of SARS-CoV-2. *mBio.* 2020; 11. <https://doi.org/10.1128/mBio.01610-20> PMID: 32694143
30. Young BE, Fong S-W, Chan Y-H, Mak T-M, Ang LW, Anderson DE, et al. Effects of a major deletion in the SARS-CoV-2 genome on the severity of infection and the inflammatory response: an observational cohort study. *The Lancet.* 2020; 396: 603–611. [https://doi.org/10.1016/S0140-6736\(20\)31757-8](https://doi.org/10.1016/S0140-6736(20)31757-8) PMID: 32822564