J. Health Biol Sci. 2024;12(1):1-8 doi: 10.12662/2317-3076jhbs.v12i1.5426.p1-8.2024

ORIGINAL ARTICLE

Spike(s) protein gene microevolution of SARS CoV-2 virus in Bolivian Population: correlation between phylogeny and contagion waves

Microevolução do gene da proteína Spike(S) do vírus SARS CoV-2 na população boliviana: correlação entre filogenia e ondas de contágio

Aneliz Ninahuanca Terán[1,2], Emma Torres Tola[1], Daniela Andrea Arteaga Voigt[1], Ruddy Luna Barrón[1]

1. Instituto de Investigaciones Técnico Científicas de la Universidad Policial, Universidad Policial "Mcal. Antonio José de Sucre"(UNIPOL), Miraflores, Bolivia. 2.GENÓMICA B-29 S.C.S., Sopocachi, Bolivia

Abstract

Objective: analyze the population gene structure and phylogeny of the S gene of the SARS CoV-2 virus of COVID-19 (+) patients from the Plurinational State of Bolivia and then correlate its phylogeny with the different waves of contagion. **Methods**: three SARS-CoV-2 samples obtained by nasopharyngeal swabs from positive COVID-19 patients were sequenced by Sanger sequencing. 488 sequences of Bolivian SARS-CoV-2 were downloaded from GISAID until September 25, 2023. The genetic structure and phylogeny were analyzed to correlate the presence of the different variants with the waves of contagion. **Results**: three (3) sequences of 623pb, 1557pb, and 3060pb of the S gene were obtained, and the last one was analyzed with the sequences downloaded from GISAID, obtaining a phylogenetic tree and a network of haplogroups that revealed the formation of clades and nodes that gave rise to variants of the six waves that Bolivian inhabitants faced. The Y508S mutation was identified as new since it was not identified in the CoVariants® and CoV-Glue® databases. **Conclusions**: the phylogeny, haplogroup network, and Person's correlation coefficient revealed the existence of a positive correlation between the microevolution of the analyzed fragment and the appearance of the different waves of contagion.

Keywords: S gene; SARS CoV-2; phylogeny; waves of contagion; Bolivia.

Resumo

Objetivo: analisar a estrutura genética populacional e a filogenia do gene S do vírus SARS CoV-2 de pacientes com COVID-19 (+) do Estado Plurinacional da Bolívia e, em seguida, correlacionar sua filogenia com as diferentes ondas de contágio. **Métodos**: três amostras de SARS-CoV-2 obtidas por swabs nasofaríngeos de pacientes positivos para COVID-19 foram sequenciadas por sequenciamento de Sanger. 488 sequências do SARS-CoV-2 boliviano foram baixadas do GISAID até 25 de setembro de 2023. A estrutura genética e a filogenia foram analisadas para correlacionar a presença das diferentes variantes com as ondas de contágio. **Resultados**: foram obtidas três (3) sequências de 623pb, 1557pb e 3060pb do gene S, sendo a última analisada com as sequências baixadas do GISAID, obtendo uma árvore filogenética e uma rede de haplogrupos que revelaram a formação de clados e nós que deram origem a variantes das seis ondas que os habitantes bolivianos enfrentaram. A mutação Y508S foi identificada como nova, pois não foi identificada nas bases de dados CoVariants® e CoV-Glue®. **Conclusões**: a filogenia, a rede de haplogrupos e o coeficiente de correlação de Person revelaram a existência de uma correlação positiva entre a microevolução do fragmento analisado e o aparecimento das diferentes ondas de contágio.

Palavras-Chave: gene S; SARS CoV-2; filogenia; ondas de contágio; Bolívia.

INTRODUCTION

The SARS-CoV-2 virus responsible for the COVID-19 pandemic originated in Wuhan, China. Since then, a global spread has been experienced, affecting millions of people in various waves of contagion of continuously evolving variants[1]. Its mutation capacity is particularly high, as is common in RNA viruses[2], but has been even more accentuated due to the magnitude of the pandemic, which has led to the emergence of multiple lineages[3].

Various investigations have revealed that the SARS-CoV-2 virus harbors in its different informative regions (proteins E, M, N, S, ORF), e.g. Tang et al. (2020)[4] analyzing 103 genomes revealed the existence of two prevalent types of evolution, type L and type S, the first being derived from the second and more virulent and aggressive. Mousavizadeh & Ghasemi (2021)[5] expose the genotypic and phenotypic differences and similarities of SARS CoV-2 compared to other Betacoronaviruses. Likewise, Zang et al. (2020)[6] report how the detection of SARS CoV-2 in anal swabs was more positive than in nasal swabs.

The genetic material of SARS CoV-2 is a positive single-stranded RNA, 30,000 bases long, making it the largest RNA genome identified to date[7]. The classification of SARS CoV-2 within the Coronaviridae family and the Betacoronavirus genus is based on its genomic structure and the characteristics of its Spike (S) protein that it uses to enter host cells.

In the mature viral particle, the protein S, a type I homotrimeric fusion glycoprotein, is located on the surface of the viral particle and plays a critical role in binding to cellular receptors. In the case of humans, the angiotensin converting molecule II (ACE-

Correspondence: Ruddy Luna Barrón. Av. Abel Iturralde, N° 1122, Zona Miraflores. Bolivia. E-mail: rlunab@gmail.com
Conflict of interesse: There is no conflict of interest on the part of any of the authors.
Received: 5 Ago 2024; Revised: 15 Sep 2024; Accepted: 25 Sep 2024

2) has been identified as the primary receptor for SARS CoV-2. Different investigations have managed to resolve the complete structure of the S protein of SARS CoV-2, whether in its form bound or not to the ACE-2 receptor[8]. This protein consists of approximately 1,273 amino acids, and its domains are clearly defined. Given their importance in virus attachment and entry into host cells, particular emphasis has been placed on mutations affecting the receptor binding domain (RDB)[9]. On the other hand, Resende et al. (2021)[10] describe mutations in other domains, such as the amino (N)-terminal domain (NTD), which also have the potential to induce conformational changes in the structure of the S protein, which may influence its functionality.

Amicone et al. (2021)[11], based on the accumulated mutations and excluding the genes with selection signals in a trial carried out, estimated a spontaneous mutation rate of 1.25×10-6 nt-1 per infection cycle, demonstrating that the Mutation accumulation is heterogeneous throughout the genome and showed that the S gene accumulates mutations at a rate five times greater than the genomic average.

In addition to RNA replication errors, host-mediated genome editing by innate cellular defense mechanisms may introduce a substantial number of targeted mutations into the SARS CoV-2 genome and may, therefore, influence its rate of evolution[12].

The WHO has stopped collecting epidemiological information since August 1, 2023. Until that date, a total of 1,206,420 cases of infection and 22,399 deaths were recorded in the Plurinational State of Bolivia[13]. Worldwide, 6,957,216 deaths were reported, of which 2,959,349 occurred in America[14]. According to the National Institute of Health Laboratories "Dr. Néstor Morales Villazón" (INLASA) from Bolivia, the country faced six (6) waves of contagion in July 2020, January 2021, June 2021, January 2022, July 2022 and December 2022, in which the predominant variants were 20B (B.1.1), 20B (B.1.1) and 20I (Alpha, V1, B1.1.7), 20J (Gamma, V3, P.1), 21K (Omicron, BA.1) and 21A (Delta, B.1.617.2), 22B (Omicron, BA.5), and 22E (Omicron, BQ.1), respectively.[15]

The objective of this study was to analyze the population gene structure and phylogeny of the S gene of the SARS CoV-2 virus of COVID-19 (+) patients from the Plurinational State of Bolivia and then correlate its phylogeny with the different waves of contagion. This is to contribute to the generation of knowledge that is useful for the diagnosis, epidemiological monitoring, and treatment of the patient.

METHODS

RNA extraction and Sanger sequencing

Using the Viral RNA Extraction Ezgene System kit (Biomiga®) and following the manufacturer's technical specifications, RNA was extracted from 3 nasopharyngeal swab samples obtained from patients previously diagnosed with COVID-19 using Real Time PCR. The collection was carried out by personnel from the laboratory of the Genetic Research Center (CINGEN) of Bolivian Police under informed consent.

The quality of the extracts was examined by electrophoretic running on agarose gel (0.8%) with RNA EZ-Vision Loading Buffer and 0.5X TBE buffer. As well as it was quantified by fluorometry (Qubit 2.0., Invitrogen®). The primers used were: 5'TTTTATCTCTTCTTAGTAAAGGTAGAC, 3'CCAGGAGTCAAATAACTTCTATG, 5'CATAGAAGTTATTTGACTCCTGG

3'CCATTGAAGTTGAAATTGACACA, 5' TGTGTCAATTTCAACTTCAATGG

3'CCATTAAACCTATAAGCCATTTGCAT, 5'ATGCAAATGGCTTATAGGTTTAATGG

3'TTGTGAAGATTCTCATAAACAAATCC[16]. The commercial kit SuperScript™IV First-Strand Synthesis System (Invitrogen®) was used for retrotranscription and amplification of the products. Amplicons were visualized on 1.5% agarose gels and compared to the molecular weight control 50 bp DNA Step Ladder (Promega). The amplicons were subjected to alcoholic purification, and then the unbalanced PCR was performed with the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems®) according to the manufacturer's recommendations.

Once the PCR was completed, a second purification was carried out with the BigDye XTerminator purification kit following the manufacturer's recommendations. The products were sequenced on an ABI 3500 Genetic Analyzer (Applied Biosystems®). The obtained sequences were analyzed with Sequencing Analysis v.6.2 software (Applied Biosystems®). The forward and reverse sequences were aligned and edited with Bioedit[17].

Obtaining GISAID sequences

Until September 25, 2023, 488 Bolivian sequences were found in the GISAID database (https://www.gisaid.org/), all downloaded for analysis. Likewise, the Wuhan/WIV04/2019 sequence (EPI_ISL_402124) was downloaded as a reference[10].

Analysis and comparison of sequences

Using the MAFFT v.718 software, the 488 Bolivian sequences extracted from the GISAID database were aligned, plus the three samples for which Sanger sequencing was performed. To avoid errors, those sequences downloaded from GISAID that contained more than four "NNNN..." in a row in the amplified fragments of the three (3) processed samples were eliminated, so only 349 Bolivian sequences remained to compare.

Descriptive statistical analysis

The descriptive statistical analysis of polymorphism, nucleotide diversity, haplotypes number, haplotypic diversity, polymorphic sites, monomorphic sites, and genetic distance was carried out with the DNAsp v.6[19] and Mega 7.0[20] programs. Mega 7.0[20] was used to identify the intrinsic characteristics of the fragment under study.

Inferential statistical analysis

For the phylogenetic analysis, the Mega 7.0[20] program was used with the maximum likelihood (ML) method, Tamura & Nei (1993)[21] substitution model, with a 5-parameter gamma distribution and 1000 bootstraps. Finally, to obtain haplogroup networks, the PopART[22] Program was used using the "median union" or "MJ" inference method. The sequences obtained by Sanger sequencing were uploaded to the GISAID database, which issued the following accession codes: EPI_ISL_18918411, EPI_ISL_18981787 and EPI_ISL_18981788.

For the correlation analysis between the microevolution of the SARS-CoV-2 virus and the chronological emergence of variants due to waves of contagion, the Pearson linear correlation coefficient was used between the genetic distances of the 3060 bp sequences from the S gene, obtained with the method maximum likelihood (ML), Tamura-Nei substitution model, 5-parameter gamma distribution and 1000 bootstraps with the dates of appearance of each variant. (Fig. 5)

RESULTS

Sanger sequenced samples

Of the three samples analyzed by Sanger sequencing, only one was amplified with more than 80% (3060 bp) of the total size of the S gene. This was called CINGEN-IITCUP-LP2/2021 and was annexed to the GISAID database under the Accession ID EPI_ISL_18918411. The other two samples - CINGEN-IITCUP-LP1/2021 and CINGEN-IITCUP-LP3/2024 - were attached to GISAID under the accession IDs EPI_ISL_18981787, EPI_ISL_18981788, whose informative lengths were 1557 bp and 623 bp, respectively.

Comparison between Bolivian sequences

Of the 488 sequences downloaded from GISAID, only 348 complete sequences were obtained according to the amplified fragment of the IITCUP-LP2/2021 sample (3060pb). Thus, 150 Bolivian haplotypes were obtained (Table 1).

Table 1. Number of haplotypes and haplotypic diversity by department of Bolivia according to the S gene fragment analyzed.

N°	Departament	Number of sequences analyzed	Number of haplotypes	Haplotypic diversity
1	La Paz	6	4	0,80000
2	Oruro	1	1*	0
3	Potosí	1	1*	0
4	Cochabamba	14	8	0,86813
5	Chuquisaca	1	1*	0
6	Tarija	1	1*	0
7	Pando	1	3	0,83333
8	Beni	1	1*	0
9	Santa Cruz	319	135	0,95883

* They correspond to haplotypes that are already mentioned in the Santa Cruz groups.

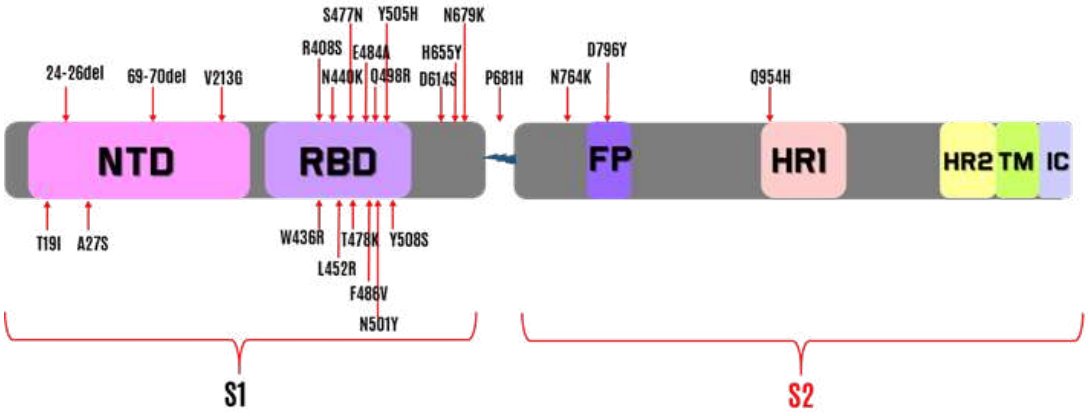
According to the Wuhan standard sequence (EPI_ISL_402124), the CINGEN-IITCUP-LP2/2021 sequence presented 27 amino acid mutations: T19I, L24del, P25del, P26del, A27S, H69del, V70del, V213G, R408S, W436R, N440K, L452R, S477N, T478K, E484A, F486V, Q498R, N501Y, Y505H, Y508S, D614G, H655Y, N679K, P681H, N764K, D796Y and Q954H (Fig. 1). The CINGEN-IITCUP-LP1/2021 sequence presented 7 mutations: T19R, G142D, E156G, F157del, R158del, L452R and T478K. The CINGEN-IITCUP-LP3/2024 sequence presented 10 mutations: T19I, L24del, P25del, P26del, A27S, V83A, G142D, H146Q, Y144del and Q183E.

Of the mutations identified in CINGEN-IITCUP-LP2/2021, 73.53% correspond to mutations present in the Omicron variant, subvariant BA.4 and BA.5 according to the CoVariants® database. Of these, 25 amino acid mutations were identified in the sequence obtained, 5 were not identified due to non-amplification of their segments, and 4 did not mutate. Mutations W436R and Y508S were identified as new mutations for these variants.

In CINGEN-IITCUP-LP3/2024, 10 mutations were identified, among which H146D was not found in the CoVariants® database but was found in the CoV-Glue® database. The other nine mutations correspond to the omicron XBB variant.

In CINGEN-IITCUP-LP1/2021, 7 mutations were identified, of which E156G and R158del were not found in the CoVariants® database but were found in the publication by Tarun, et al. (2022)[23]. The other five (5) mutations correspond to the Delta variant.

Figure. 1 Location of the mutations identified in the S gene of the CINGEN-IITCUP-LP2/2021 sample



Descriptive statistics

Given that in the CINGEN-IITCUP-LP1/2021 and CINGEN-IITCUP-LP3/2024 studies, only sequences less than 50% of the total length of the S gene were amplified, only the descriptive statistics of the Bolivian sequences obtained from GISAID will be provided and the CINGEN-IITCUP-LP2/2021 sequence. Thus, the haplotypic diversity was equal to 0.9615, which represents 96.15%, indicating that of 100 SARS CoV-2 viruses from Bolivia analyzed, 96 are different from each other for the S fragment analyzed (3060 bp). The nucleotide diversity obtained was 0.00486 with a standard deviation equal to 0.00019.

One hundred and seventy (170) polymorphic sites and 2812 monomorphic sites were found, with 78 sites containing gaps and/or deletions. The number of haplotypes and haplotypic diversity by department are shown in Table 1. The average genetic distance was equal to 0.0057. Within the intrinsic characteristics analyzed, a percentage of adenine equal to 30.02%, thymine 32.97%, cytosine 18.82%, and guanine 18.19% was obtained. There is a Transitions/Transversions ratio equal to 1.19.

Inferential statistics

The phylogenetic tree (Fig. 2) obtained for the Bolivian and CINGEN-IITCUP-LP2/2021 sequences reveals the presence of 6 distinct clades, each associated with a year of appearance and a specific variant. The haplotype network obtained (Fig. 4) reveals the presence of 6 haplogroups. The phylogenetic tree (Fig. 3) obtained for the Bolivian sequences and the three (3) sequenced samples with an analyzable size of 623 bp reveals the presence of 6 clades and a clade of two variants separated from the clade to which they belong, that is, to the clade of the 2nd or 1st wave.

The Pearson linear correlation between microevolution and the contagion waves gave a value equal to 0.9359. (Fig. 5)

Figure 2. Phylogenetic tree of the S gene fragment analyzed in Bolivian sequences (3060pb), identifying clades linked to waves and variants.

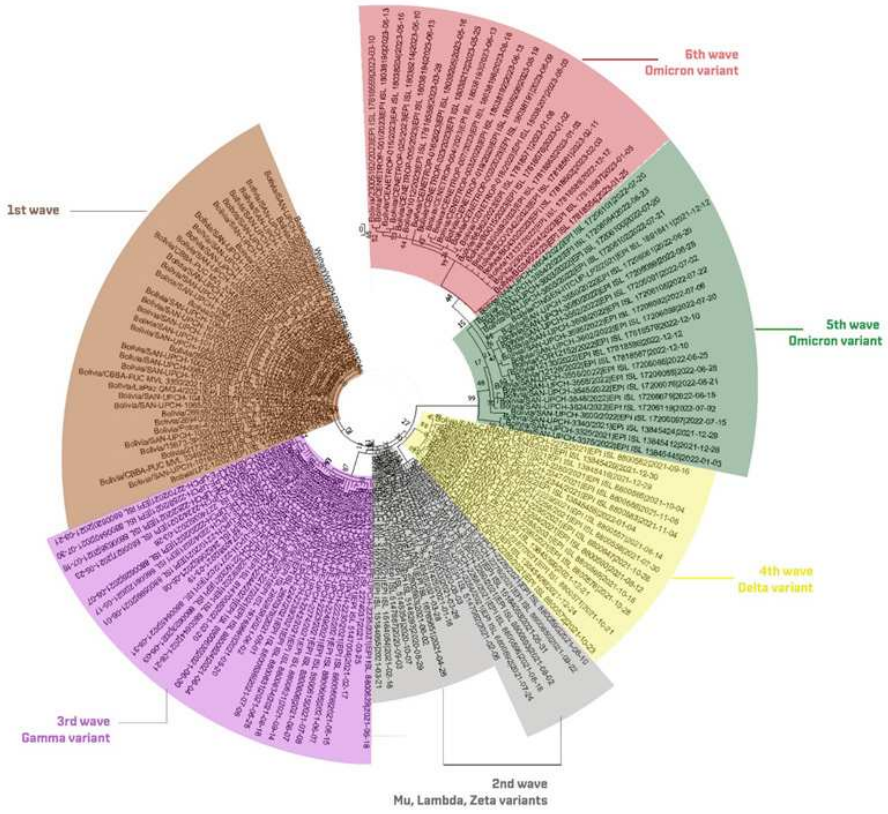


Figure. 3. Phylogenetic tree of the S gene fragment analyzed in Bolivian sequences (623 bp), identifying clades linked to variants.

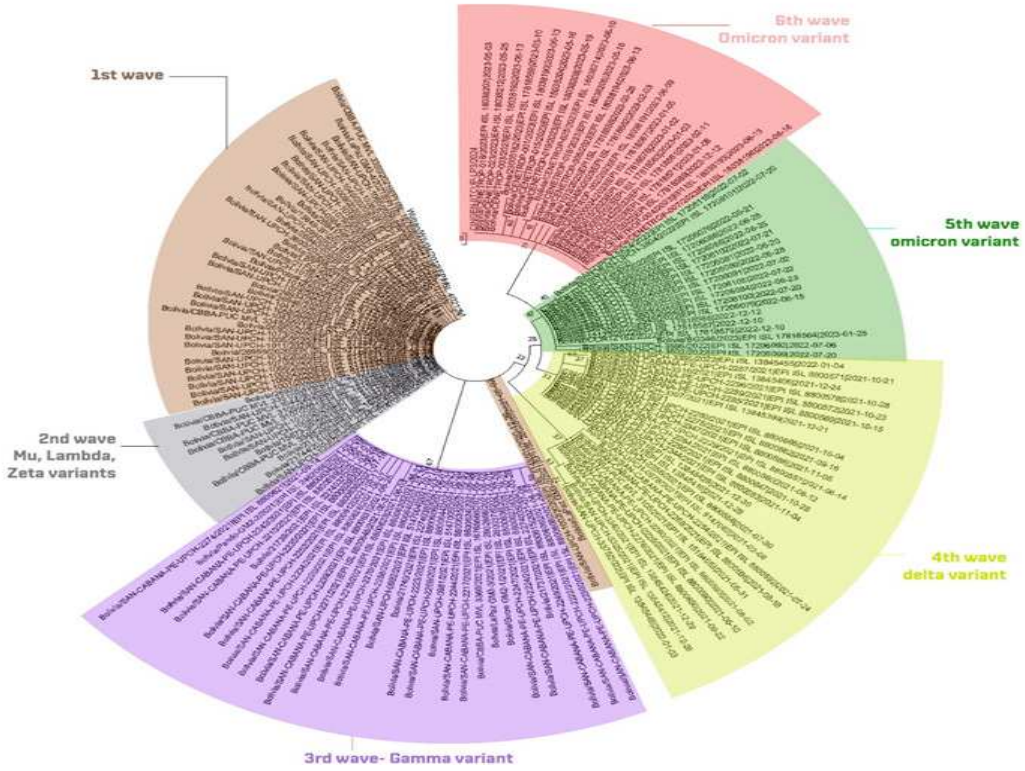


Figure. 4. Network of haplogroups of the S gene (3060 bp) of SARS COV-2 from Bolivia that correspond to five identified

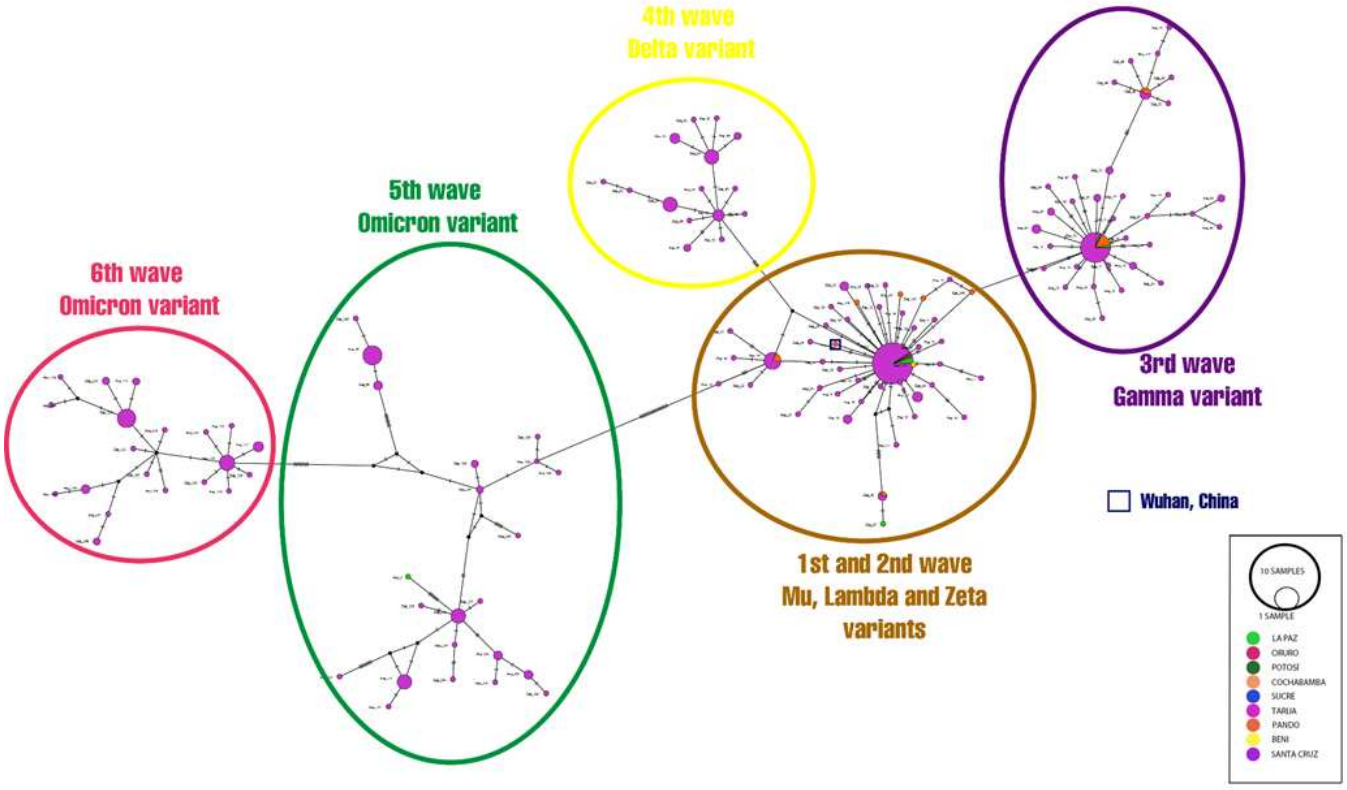
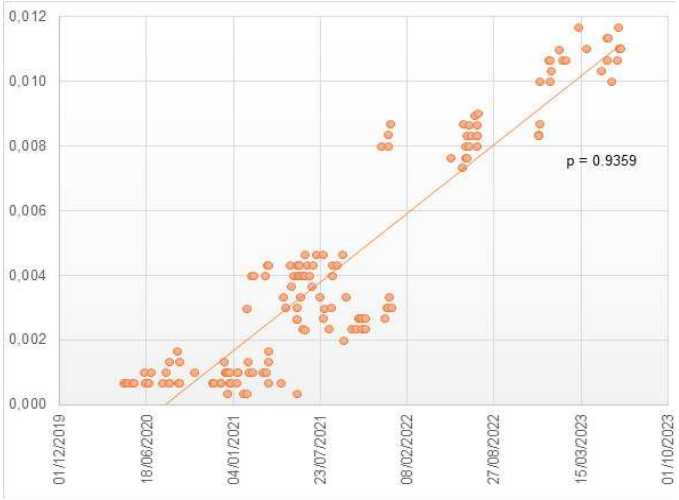


Figure 5. Pearson linear correlation coefficient of the entire analyzed sequence of the S gene (3060 bp)



DISCUSSION

The origin of SARS CoV-2 and its evolutionary relationship remains ambiguous. Several studies attempted to resolve this question using genome-based phylogenetic analyses, but limited progress was made, mainly due to the inability of these methods to reasonably integrate the phylogenetic information of all SARS CoV-2 genes[24].

The values obtained for haplotypic diversity, nucleotide diversity, polymorphic and monomorphic sites, and sites with gaps and/or deletions reveal that the analyzed fragment (3060 bp) is hypervariable since the more sequences are analyzed, the more haplotypes are identified. It may be due to the presence of the NTD and RBD subunits, which are the regions responsible for the recognition and binding of the virus to ACE2[25].

As mentioned by Singh & Yi (2021)[26], the S proteins of coronaviruses are known to undergo frequent sequence changes in nature, including deletions, mutations, and recombinations; in particular, the RBD shows more divergence than other regions, suggesting some alteration in the binding affinity to human ACE2.

The identification of the Y508S mutation in the CINGEN-IITCUP-LP2/2021 sequence represents a significant finding in the context of the evolution of the SARS-CoV-2 virus, particularly in the genetic landscape of the variants circulating in Bolivia. This mutation, as it is not registered in widely used databases such as CoVariants® and CoV-Glue®, suggests the possibility that it is a new and specific variant of this geographic region. The W436R mutation has already been studied for the production of vaccines[27].

The E156G and R158del mutations of the CINGEN-IITCUP-LP1/2021 sequence, although not found in the CoVariants® database, studies such as that of Tarun et al. (2022)[23] report that the combined presence of E156G/Δ157-158 is related to greater infectivity and lower sensitivity to antibodies induced by vaccines; thus, not only changes in RBD can determine the

infectivity and immune escape of the virus. It is why, at the time, the Delta variant was one of the most contagious and lethal[28].

The H146D mutation identified in the CINGEN-IITCUP-LP3/2024 sequence is not new since it is registered on the CoV-Glue® page. Although it was possible to identify the type of variant in the samples sequenced by Sanger, for a definitive classification, it is necessary to have the complete sequence of the S gene[29]. Although the phylogenetic tree (Fig. 3) of the analyzed 623 bp fragment shows six (6) clades defined by related haplotypes based on dates and type of variant, there are two variants that must have been found within the clades of the 1st or 2nd wave but they are separated between the 3rd and 4th wave; however, continue being more old that the variants from 3rd and 4th wave. It suggests that it is necessary to analyze the complete fragment of the S gene to obtain correlation data according to waves of contagion, as observed in the phylogenetic tree of the S gene analyzed based on 3060 bp (Fig. 2).

The correlation is also observed in the haplogroup network in Fig. 4, where the sequence obtained (3060 pb) is found in the haplogroup identified as from the 5th wave when the predominant variants were Omicron BA 4 and 5. Key nodes reflecting genetic diversity between different epidemic waves were found in the haplogroup network (Fig. 4). The first two showed greater genetic variability. However, in later waves, a significant distance between haplotypes is observed, suggesting microevolution of the virus with up to 18 mutational jumps between haplotypes. Some haplotypes, such as CINGEN-IITCUP-LP2/2021, did not have detectable offspring.

On the other hand, it is important to mention that, despite the use of a 623 bp fragment to build a phylogenetic tree (Fig. 3), the arrangement of the samples sequenced by Sanger shows a location in the clade corresponding to its variant, such is the case of the CINGEN-IITCUP-LP1/2021 sample that are found with variants of the 4th wave, that is, Delta variants.

Finally, the Pearson correlation coefficient with a value equal to 0.9359 (Fig. 5) shows a significant positive correlation ($p>0.9$) between the evolution of the S gene of the SARS-CoV-2 variants that were presented in the Bolivian population over time. Well, as time passed, their variability in relation to the Wuhan base variant increased, giving rise to a positive microevolution.

These results underscore the importance of continued genomic surveillance in Bolivia to track the evolution of the virus and detect the emergence of variants that may have implications in terms of transmissibility, severity, and vaccine efficacy. Furthermore, the possible presence of variants with new mutations highlights the need to investigate the immune response and effectiveness of vaccines in the local population.

It is crucial to highlight the fundamental role of Sanger sequencing in this study since it was evident that numerous NGS sequences loaded in GISAID do not present a complete amplification of the S gene, which is why there is no complete information about it in Bolivia. As mentioned by Daniels et al (2021)[16] and other authors, NGS has limitations when analyzing and identifying point mutations of clinical relevance.

At present, the process of generating and acquiring knowledge about SARS-CoV-2 remains a constant challenge. As we understand more about this virus, we are in a stronger position to develop effective tools to combat it directly and to confront other genetically similar viruses. Although next generation sequencing (NGS) has been essential to revealing the genetic information of each variant of the SARS-CoV-2 virus, it is crucial to highlight that accurate confirmation of these variants requires the use of Sanger sequencing. This technique is the gold standard for sequencing and identification of point mutations. In addition, its use should be considered to increase the accuracy of predicting the reliability threshold of the detected variants.[30]

CONCLUSIONS

The population genetic structure and phylogeny of the viral genomes of the SARS-CoV-2 S gene from Bolivian sequences available in the GISAID database from 2020 to September 25, 2023, were analyzed, making the existence of a correlation evident. Between the waves of contagion that the Plurinational State of Bolivia went through and the variants that emerged in each of them, since the formation of six clades concordant with the six waves that occurred was evident, with a significant positive correlation being evident ($p>0.9$) between the genetic distances of the Bolivian variants (with the Wuhan base sequence) and their chronological emergence of variants (microevolution) by waves in the Bolivian population.

Sequences of the S gene of the SARS CoV-2 virus were comparatively studied during the pandemic in Bolivia, contributing significantly to the understanding of its hypervariability, microevolution, and phylogeny throughout the different waves of contagion, which showed a haplotypic diversity of 96.15%, an average genetic distance of 0.0057 and each clade of the phylogenetic tree obtained was shown to harbor variants that arose in the same wave.

In addition, the Y508S mutation was identified as unique in the Bolivian population since it was not found in any other sequence downloaded from GISAID. The information obtained from the amplified fragment of the S gene of the SARS-CoV-2 virus is useful to determine the type of variant in the Bolivian population.

ACKNOWLEDGEMENTS

We thank Laboratorios IFA S.A. for their contribution to the purchase of consumables and reagents for the development of the study. To the Police University of Bolivia for authorizing the work to be carried out in its facilities.

We thank all the authors who uploaded the sequences obtained from SARS CoV-2 from Bolivia to the GISAID platform.

REFERENCES

1. Alisoltani A, Jaroszewski L, Iyer M, Iranzadeh A, Godzik A. Increased Frequency of Indels in Hypervariable Regions of SARS-CoV-2 Proteins-A Possible Signature of Adaptive Selection. Front Genet. 2022 Jun; 13: 875406. doi:10.3389/fgene.2022.875406.

2. Duffy S. Why are RNA virus mutation rates so damn high? PloS Biol. 2018 Aug; 16(8): e3000003. doi:10.1371/journal.pbio.3000003.

3. Attwood SW, Hill SC, Aanensen DM, Connor TR, Pybus OG. Phylogenetic and phylodynamic approaches to understanding and combating the early SARS-CoV-2 pandemic. Nat Rev Genet. 2022 Sep; 23(9): 547–562. doi: 10.1038/s41576-022-00483-8.

4. Tang X, Wu C, Li X, Song Y, Yao X, Wu X, . On the origin and continuing evolution of SARS-CoV-2 . Natl Sci Rev. 2020 Jun; 7(6): 1012-1023. doi: 10.1093/nsr/nwaa036.

5. Mousavizadeh L, Ghasemi S. Genotype and phenotype of COVID-19: Their roles in pathogenesis. J Microbiol Immunol Infec. 2021 Apr; 54(2): 159-163. doi:10.1016/j.jmii.2020.03.022.

6. Zang W, Du R-H, Li B, Zheng X-S, Yang X-L, Hu B, et al. Molecular and serological investigation of 2019-nCoV infected patients: implication of multiple shedding routes. Emerg Microbes Infect. 2020 Feb; 9(1): 386-389. doi: 10.1080/22221751.2020.1729071.

7. Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020 Mar; 579(7798): 270-273. doi:10.1038/s41586-020-2012-7.

8. Wrapp D, Nianshuang W, Corbett KS, Goldsmith JA, Hsieh C-L, Abiona O, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science. 2020 Mar; 367(6483): 1260 – 1263. doi:10.1126/science.abb2507.

9. Salles TS, Cavalcanti AC, Costa FB, Dias VZ, Souza LM, Meneses MD, et al. Genomic surveillance of SARS-CoV-2 Spike gene by sanger sequencing. PLoS One. 2022 Jan; 17(1): e0262170. doi:10.1371/journal.pone.0262170.

10. Resende P, Delatorre E, Gräf T, Mir D, Motta FC, Appolinario LR, et al. Evolutionary Dynamics and Dissemination Pattern of the SARS-CoV-2 Lineage B.1.1.33 During the Early Pandemic Phase in Brazil. Front Microbiol. 2021 Feb; 11: 615280. doi:10.3389/fmicb.2020.615280.

11. Borges V, Alves MJ, Amicome M, Isidro J, Zé-Zé L, Suarte S, et al. Mutation rate of SARS-CoV-2 and emergence of mutators during experimental evolution. BioRxiv. 2021 May: 1-26. doi: 10.1101/2021.05.19.444774.

12. Markov PV, Ghafari M, Beer M, Lythgoe K, Simmonds P, Stilianakis N I, et al. The evolution of SARS-CoV-2. Nat Rev Microbiol. 2023 Jun; 21(6): 361-379. doi:10.1038/s41579-023-00878-2.

13. Salud, O. M. (16 de 09 de 2023). COVID-19 Americas' Regional Dashboard. Obtenido de https://who.maps.arcgis.com/apps/dashboards/c147788564c148b6950ac7ecf54689a0.

14. World Health Organization. Number of COVID-19 cases reported to WHO [Internet]. Geneve: WHO; 2023 [16 de 09 de 2023]. Disponível em: https://covid19.who.int/.

15. INLASA. (16 de 09 de 2023). Dashboard de Vigilancia Genómica. Obtenido de https://vigilancia.inlasa.gob.bo/dashboard-vigilancia/.

16. Daniels R S, Harvey R, Ermetal B, Xiang Z, Galiano M, Adams L, et al. A Sanger sequencing protocol for SARS-CoV-2 S-gene. Influenza Other Respi Viruses. 2021 Nov; 15(6): 707-710. doi:10.1111/irv.12892.

17. Hall T, Alzohairy A M. BioEdit: An important software for molecular biology[Internet]. GERF Bulletin of Biosciences. 2011 Jan [30 de 09 de 2023]; 2(1): 60-61. Disponível em: https://www.researchgate.net/publication/258565830.

18. Katoh K, Misawa K, Kuma K-i, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002 Jul; 30(14): 3059-3066. doi:10.1093/nar/gkf436.

19. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, et al. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. Mol Biol Evol. 2017 Dec; 34(12): 3299-3302. doi:10.1093/molbev/msx248.

20. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Dataset. Mol. Biol. Evol., 33(7): 1870-1874. doi:10.1093/molbev/msw054.

21. Tamura K, Nei M. Estimation of the Number of Nucleotide Substitutions in the Control Region of Mitochondrial DNA in Humans and Chimpanzees. Mol. Biol. Evol. 1993 May; 10(3): 512-526. doi: 10.1093/oxfordjournals.molbev.a040023.

22. Leigh J W, Bryant D. POPART: full-feature software for haplotype network construction. Methods in Ecology and Evolution. 6(9): 1110-1116. doi:10.1111/2041-210X.12410.

23. Mishra T, Dalavi R, Joshi G, Kumar A, Pandey P, Shukla S, et al. SARS-CoV-2 spike E156G/Δ157-158 mutations contribute to increased infectivity and immune escape. Life Sci Alliance. 2022 Mar; 5(7): e202201415. doi: 10.26508/lsa.202201415.

24. Tingting L, Dongxia L, Yang Y, Guo J, Feng Y, Zhang X, et al. Phylogenetic supertree reveals detailed evolution of SARS CoV 2. Sci Rep. 2020 Dec; 10(1): 22366. doi:10.1038/s41598-020-79484-8.

25. Magazine N, Zhang T, Wu Y, McGee M C, Veggiani G, Huang W. Mutations and Evolution of the SARS-CoV-2 Spike Protein. Viruses. 2022 Mar; 14(3): 640. doi:10.3390/v14030640.

26. Singh D, Yi S V. On the origin and evolution of SARS-CoV-2. Exp Mol Med. 2021 Apr; 53(4): 537–547. doi: 10.1038/s12276-021-00604-z.

27. Bergasa-Cáceres F, Rabitz H A. The Promise of Mutation Resistant Drugs for SARS-CoV-2 That Interdict in the Folding of the Spike Protein Receptor Binding Domain. COVID. 2021; 1(1), 288-302. doi:10.3390/covid1010023.

28. Mishra S. ¿Por qué la variante delta es más contagiosa y letal? Una nueva investigación tiene la respuesta [Internet]. National Geographic. 2021 Ago [28 de 11 de 2023]. Disponível em: https://www.nationalgeographicla.com/ciencia/2021/08/por-que-la-variante-delta-es-mas-contagiosa-y-letal-una-nueva-investigacion-tiene-la-respuesta.

29. Murillo E, Palacio-Rua K, Afanador-Ayala C, García-Correa J F, Zuluaga A F. Validación de una nueva estrategia para la identificación de variantes de SARS-CoV-2 mediante secuenciación del gen espiga por Sanger. Enferm Infecc Microbiol Clin. 2023 May; 41(5): 284–289. doi:10.1016/j.eimc.2022.04.014.

30. Mu, W., Lu, H.-M., Chen, J., Li, S., & Elliot, A. M. (2016). Sanger Confirmation Is Required to Achieve Optimal Sensitivity and Specificity in Next-Generation Sequencing Panel Testing. J Mol Diagn. 2016 Nov; 18(6): 923-932. doi: 10.1016/j.jmoldx.2016.07.006.

How to cite this article/ Como citar este artigo:
Terán AN, Tola ET, Voigt DAA, Barrón RL Spike (S) Protein Gene Microevolution of SARS CoV-2 Virus In Bolivian Population: Correlation Between Phylogeny And Contagion WavesJ Health Biol Sci. 2024; 12(1):1-8.