Evolution of SARS-CoV-2 during the first year of the COVID-19 pandemic in Northwestern Argentina

Romina Zambrana Montaño a,b, Andrés Carlos Alberto Culasso a,b, Franco Fernández c, Nathalie Marquez c, Humberto Debat c, Mariana Salmerón d, Ana María Zamora d, Gustavo Ruiz de Huidobro d, Dardo Costas d, Graciela Alabarse d, Miguel Alejandro Charre e, Ariel David Fridman e, Claudia Mamani e, Fabiana Vaca e, Claudia Maza Díaz e, Viviana Raskovsky f, Esteban Lavaque f, Veronica Lesser f, Pamela Cajal f, Fernanda Agüero f, Cintia Calvente f, Carolina Torres a,b,**, Mariana Viegas a,**

** Corresponding author at: Universidad de Buenos Aires, Facultad de Farmacia y Bioquímica, Instituto de Investigaciones en Bacteriología y Virología Molecular (IbaViM), Junín 954, 4to piso, C1113AAD, CABA, Argentina.
* Corresponding author.
E-mail addresses: cторreo@ffyb.uba.ar (C. Torres), viegasmariana@conicet.gov.ar (M. Viegas).

A B S T R A C T

Studies about the evolution of SARS-CoV-2 lineages in different backgrounds such as naive populations are still scarce, especially from South America. This work aimed to study the introduction and diversification pattern of SARS-CoV-2 during the first year of the COVID-19 pandemic in the Northwestern Argentina (NWA) region and to analyze the evolutionary dynamics of the main lineages found. In this study, we analyzed a total of 260 SARS-CoV-2 whole-genome sequences from Argentina, belonging to the Provinces of Jujuy, Salta, and Tucumán, from March 31st, 2020, to May 22nd, 2021, which covered the full first wave and the early second wave of the COVID-19 pandemic in Argentina. In the first wave, eight lineages were identified: B.1.499 (76.9%), followed by N.5 (10.2%), B.1.1.274 (3.7%), B.1.1.348 (3.7%), B.1 (2.8%), B.1.600 (0.9%), B.1.1.33 (0.9%) and N.3 (0.9%). During the early second wave, the first-wave lineages were displaced by the introduction of variants of concern (VOC) (Alpha, Gamma), or variants of interest (VOI) (Lambda, Zeta, Epsilon) and other lineages with more limited distribution. Phylogenetic analyses of the B.1.499 and N.5, the two most prevalent lineages in the NWA, revealed that the rate of evolution of lineage N.5 (7.9 \times 10^{-4} substitutions per site per year, s/s/y) was a ~40% faster than that of lineage B.1.499 (5.6 \times 10^{-4} s/s/y), although both are in the same order of magnitude than other non-VOC lineages. No mutations associated with a biological characteristic of importance were observed as signatures markers of the phylogenetic groups established in Northwestern Argentina, however, single sequences in non-VOC lineages did present mutations of biological importance or associated with VOCs as sporadic events, showing that many of these mutations could emerge from circulation in the general population. This study contributed to the knowledge about the evolution of SARS-CoV-2 in a pre-vaccination and without post-exposure immunization period.
1. Introduction

The COVID-19 pandemic has deeply impacted all populations worldwide, with more than 500 million reported cases so far (Ritchie, 2020). South America has reported approximately 9.2% of those cases and presented epidemiological differences in terms of the lineages and variants that were introduced, emerged, and circulated in the region, especially during the first and the second wave of the COVID-19 pandemics (Faria et al., 2021; Hodcroft, 2021; Torres et al., 2021).

Molecular surveillance became an invaluable tool to monitor circulating viruses and implement public health policies to prevent the introduction, dispersion, and establishment of viruses with concerning characteristics. However, although there are more than 12 million SARS-CoV-2 genomes submitted to international public databases, only 2.0% belong to South American countries (GISAID, https://www.gisaid.org/), limiting the knowledge about the evolutionary characteristics of the virus in the region.

Diversity and evolution in SARS-CoV-2 are reflected in the description of lineages, variants, and clades. On the one hand, the Pango nomenclature system defines “Pango lineages” that are groups of SARS-CoV-2 genome sequences or clusters of infections with shared ancestry (Rambaut et al., 2020). Currently, more than 1700 Pango lineages have been defined (https://www.pango.network/), limiting the knowledge about the evolutionary characteristics of the virus in the region.

Proyecto PAIS is the inter-institutional federal consortium of SARS-CoV-2 genomics in Argentina, created to monitor SARS-CoV-2 diversity and evolution in the country, including surveillance of SARS-CoV-2 variants of public health interest (Ministry of Science Technology and Innovation, 2020). The country reported the first cases of COVID-19 in early March 2020 and experienced a long initial wave from May 2020 up to mid-February 2021, with 2 million cases reported (Ministry of Health of the Nation, 2020a), in a context of border closures and severe restrictions on internal movement in most of that period (Ministry of Health of the Nation, 2020b). Subsequently, from mid-February to the end of September 2021, the country suffered an aggressive second wave, reaching more than 5 million reported cases and 115,000 deaths (Ministry of Health of the Nation, 2021). The National Vaccination Campaign against COVID-19 in Argentina began at the end of December 2020, for the prioritized population (health professionals and strategic personnel), and since March 2021, it has been gradually extended to the general population (starting with those at high risk for severe disease) (Rearte et al., 2022).

Even though millions of sequences have been shared in public databases, studies about the evolution of SARS-CoV-2 lineages in different backgrounds such as naive populations, vaccinees, or individuals with post-infection immunity are still scarce, especially from South America. These studies would allow a better understanding of the different scenarios where SARS-CoV-2 can evolve.

This work aimed to study the introduction and diversification pattern of SARS-CoV-2 during the first year of the COVID-19 pandemic in the Northwestern Argentina region and to analyze the evolutionary dynamics of the main lineages found.

2. Material and methods

2.1. Sample collection and sequencing of SARS-CoV-2

In this study, we analyzed a total of 260 SARS-CoV-2 whole-genome sequences from Northwestern Argentina. Among these, 144 sequences of the total dataset correspond to nasopharyngeal swabs samples from patients confirmed for COVID-19 by real-time RT-PCR that showed a
cycle threshold \((C_T)\) lower than 30. These samples were randomly selected among the total number of positive cases without an epidemiological link or international travel record, from the Provinces of Jujuy \((n = 56)\), Salta \((n = 57)\), and Tucumán \((n = 31)\).

The period of sampling was from March 31st, 2020, to May 22nd, 2021 (epidemiological week (EW) 14/2020 to EW 20/2021), which covered the full first wave and the early second wave of the COVID-19 pandemic in Argentina (Fig. 1).

Nucleic acids were submitted to whole-genome sequencing using the ARTIC protocol with the Oxford Nanopore MinION platform, within the PAIS project.

In addition, sequences from clinical samples were analyzed along with a set of 116 SARS-CoV-2 sequences from samples collected between February and May 2021 in the NWA region, available in the GISAID database (https://www.gisaid.org/) on September 1, 2021, by only selecting complete genomes and excluding those with low coverage: Jujuy \((n = 37)\), Salta \((n = 32)\) and Tucumán \((n = 47)\) (Table S1).

2.2. Ethical statement

The study was revised and approved by the Medical Ethics and Research Committees of “Ricardo Gutierrez” Children’s Hospital, Buenos Aires, Argentina (DI-2020-165-CCABA-HGNRG). Informed consent was not obtained, as patient information was anonymized and de-identified before analysis.

2.3. Viral classification and compilation of datasets by lineage

SARS-CoV-2 lineages of the sequences obtained in this work were preliminarily assigned according to the Phylogenetic Assignment of Named Global Outbreak LiNeages (PANGOLIN) web application (http://pangolin.cog-uk.io/) (Rambaut et al., 2020; O’Toole et al., 2021). According to the PANGO System (Rambaut et al., 2020), seven dominant lineages were determined according to their frequencies in the NWA region: B.1, B.1.1, B.1.499, C.37 (Lambda variant), B.1.1.33, P.1 (Gamma variant), A.2.5.

To separately analyze the introduction and diversification pattern within the main lineages found, seven datasets were built (one for each lineage) that included: the sequences from Argentina from the lineage under analysis, their most similar sequences (the best ten hits for each sequence) from a BLAST analysis conducted against the GISAID database (http://www.gisaid.org, on June 2, 2021) and reference sequences for several lineages and sublineages obtained from the PANGO designation.

2.4. Phylogenetic inference

Sequences for each dataset were aligned using the online version of the MAFFT 7.487 using default parameters (Katoh et al., 2018) (https://mafft.cbrc.jp/alignment/server/), and manually edited in Bioedit version 1.26 (Hall, 1999), trimming the first 54 nucleotides of the 5′ end and the last 67 nts of the 3′ end of the alignment, with respect to the reference hCoV-19/Wuhan/WH04/2020 (EPI_ISL_406801). Final alignments ranged between 29,785 and 29,818 nucleotides in length after the ends were trimmed (details for each dataset can be found in captions of Figs. 3–9).

Maximum likelihood (ML) phylogenetic trees were built by using IQ-TREE v2.1 (Nguyen et al., 2015; Minh et al., 2020), employing the best-fit model of nucleotide substitution as selected by ModelFinder (Kalyaanamoorthy et al., 2017). To evaluate the reliability of the groups and branches obtained in trees, the SH-like approximate likelihood ratio test (1000 replicates) (Guindon et al., 2010) and the ultrafast bootstrap approximation (10,000 replicates) (Hoang et al., 2018) methods were used.

For each analysis, clades have been defined as supported monophyletic groups that included at least two sequences from the NWA region, whereas single sequences were considered sporadic cases.

2.5. Phylodynamics

Phylodynamic analyses were performed to estimate the time to the most recent common ancestors (ancestral dates of divergence), rates of evolution, and demographic reconstructions of the two main lineages found in the NWA region, B.1.499 and N.5. Two datasets including only sequences from the NWA region were compiled.

The temporal signal for both datasets was confirmed using Root-to-tip regression analysis in TempEst v.1.5.3 (Rambaut et al., 2016), showing a positive trend of the root-to-tip genetic distances against sampling dates.
The analyses were carried out using an appropriate substitution model for each dataset according to the Bayesian Information Criterion estimated with ModelFinder in IQ-TREE v2.1 (Kalyaanamoorthy et al., 2017; Minh et al., 2020). The uncorrelated lognormal (UCLN) molecular clock model (Drummond et al., 2006) and the GMRF Bayesian Skyride coalescent model (Minin et al., 2008) implemented in the BEAST v1.10.4 software package, in the CIPRES Science Gateway server (Miller et al., 2010) were used. Analyses were run up to convergence, assessed by effective sample size (ESS) values higher than 200 and 10% of the sampling was discarded as burn-in. Uncertainty in parameter estimates was evaluated in the 95% highest posterior density (HPD95%) interval.

2.6. Mutation analyses

To identify which (if any) of the observed mutations was associated with phenotypic changes (such as an alteration of host-cell receptor binding, or changes in viral fitness or antigenicity), all sequences from NWA introduced in this work and from GISAID were analyzed using the CoVsurver tool available at the GISAID EpiCoV platform and the Outbreak.info tool (https://outbreak.info/). Three analyses were performed: 1. Genome sequences obtained from the NWA were compared with the Wuhan-Hu-4 reference sequence (GISAID: EPI_ISL_402124), searching for mutations in the Spike gene in the lineages with greater predominance in this study; 2. Mutations reported as of biological importance (mostly characteristic for VOCs) but present in non-VOCs lineages were recorded; 3. Signature mutations for the main monophyletic groups containing Argentinean sequences in different lineages were analyzed.

All mutations detected in these analyses were searched in the aforementioned websites and literature, and references were added.
when information associated any mutation with a phenotypic change or a biologically significant role.

### 3. Results

#### 3.1. Lineage distribution

A total of 16 Pango lineages from the 260 whole-genome sequences were found in Northwestern Argentina: a region covered by the Provinces of Jujuy (n = 93), Salta (n = 89), and Tucumán (n = 78).

In the first wave (from March 2020 to February 2021), eight lineages were identified from the 108 SARS-CoV-2 genomes analyzed: the most predominant one was B.1.499 (76.9%) in the three NWA provinces, followed by N.5 (10.2%), B.1.1.274 (3.7%), B.1 (2.8%), B.1.1.13 (0.9%), B.1.1.138 (3.7%), B.1.1.348 (3.7%), B.1.1.247 (5.9%), B.1.1.174 (4.6%), B.1.499 (2.6%), B.1.1.274 (2.0%), B.1.1.151 (0.7%), and N.5 (0.9%).

Instead, in the early second wave (March-May 2021), an increase in the lineage diversity was observed but also a different predominant lineage in each province. We found 12 lineages for the 152 sequences:

- N.5 (27.6%), P.1 (25.7%), C.37 (11.8%), B.1.1.348 (9.9%), A.2.5 (5.9%), B.1.427 (5.9%), B.1.1.7 (4.6%), B.1.499 (2.6%), P.2 (2.6%), B.1.1.274 (2.0%), B.1.1.151 (0.7%), and C.26 (0.7%).

Most of the sequences from the Provinces of Jujuy and Salta were assigned to the lineage N.5 (n = 22/53 and n = 15/53, respectively). Lineage P.1 (VOC Gamma, n = 28/46) showed a marked predominance in Tucumán. In addition, the second most frequent lineage was also different for each province: B.1.1.348 (n = 14/53) in Jujuy; C.37 (VOC Lambda, n = 10/53) in Salta, and B.1.427 (VOC Epsilon, n = 9/46) in Tucumán (Fig. 2A).

#### 3.2. Spread of the Argentinean lineages B.1.499 and N.5 during the first wave and the early second wave

Argentinean sequences were analyzed along with the most similar sequences in the GISAID database to improve the description of monophyletic groups in the NWA, which allows inferring common transmission chains and the establishment of circulating viruses.

Clades containing sequences from the NWA region were described for each lineage and indicated in the Figs. 3–9. Lineage B.1.499 was initially detected in April 2020 and was the most frequently detected lineage of the first wave in the NWA region.
This lineage was introduced at least eight times in this geographic region, originating six clades and two sporadic cases (Fig. 3). Three introductions corresponded to monophyletic groups with sequences from the Provinces of Jujuy and Salta, which suggested common transmission chains between these provinces during the first wave (Fig. 3, clades A–C). Furthermore, another group (Fig. 3, clade E), composed of 12 sequences, corresponded to an introduction in the province of Salta that spread to different cities within the province between June and September 2020.

In contrast, sequences from the Province of Tucumán did not show a phylogenetic association with sequences from the other provinces of the NWA region (Jujuy and Salta), instead, they were more related to sequences from the capital city (Autonomous City of Buenos Aires, CABA) and other provinces located in the Central region of the country, such as Buenos Aires and Córdoba. Four introductions to Tucumán were observed: two corresponded to sporadic cases and the other two involved local groups with viral spread and diversification to different cities of Tucumán during April and May 2020 (Fig. 3, clades D and F).

Lineage N.5 is an Argentinean lineage (derived from the Brazilian lineage B.1.1.33) that replaced lineage B.1.499 at the end of the first wave (early 2021) (Fig. 2A). During the first wave, at least six introductions in the NWA region were observed, three of them occurred in the Province of Jujuy (January 2021), and only one introduction was observed for Salta and Tucumán (February 2021) (Fig. 4, clades H–J).

This lineage predominated during the beginning of the second wave in NWA, in March 2021 (Fig. 2). However, several introductions of lineages P.1 (VOC Gamma) and C.37 (VOI Lambda) were observed since April 2021, increasing their frequencies in the following month.

Phylogenetic analysis of lineage N.5 showed an increase in the number of introductions in the NWA region as well as in the internal clustering, especially in the Provinces of Jujuy and Salta, which suggested transmission chains between these two provinces. Most of the sequences (n = 16) from Jujuy were grouped in a large clade (Fig. 4, clade A) with sequences from Salta (n = 3) and Tucumán (n = 3). In this clade, an inner group was observed (Fig. 4, subclade B) with sequences from the three NWA provinces, showing the internal diversification and dispersion of the lineage in the region between March and May 2021.

Few sporadic cases unrelated to other sequences in the NWA region were found in the three provinces: two introductions in the Province of Salta, one in Jujuy, and another in the province of Tucumán (Fig. 4).

3.3. VOC and VOI associated lineages during the early second wave

VOC Gamma (lineage P.1). At the beginning of the second wave, 15 introductions were observed: seven in Salta, seven in Tucumán, and one in Jujuy. In general, the introductions of the Gamma variant corresponded to sporadic cases except for a large clade (Fig. 5, clade A) containing 18 genomes from the Province of Tucumán obtained between April and May 2021. This large clade also contained two sequences from Salta and Jujuy, and a cluster from Chile, evidencing the establishment
of the lineage in the NWA region, mainly in the province of Tucumán and its association with circulation in the neighboring country of Chile. Additionally, two sequences collected in May 2021 from Tucumán were associated with the P.1.2 sublineage (Fig. 5, clade D).

**VOC Alpha (lineage B.1.1.7).** Four introductions were observed for this variant, but only one introduction corresponded to a monophyletic group of sequences from Salta (Fig. 6, clade A). The remaining samples corresponded to sporadic cases: one from Salta that was associated with sequences from Europe, and two other samples from Salta and Jujuy grouped within a cluster that mostly contained sequences from Argentina. This result showed a brief local transmission period of the Alpha variant during the early second wave, but without establishment in the NWA region.

**VOI Lambda (lineage C.37).** 18 sequences were assigned to the C.37 lineage: ten sequences from the province of Salta, six from Jujuy, and two other samples from Salta and Jujuy grouped within a cluster that mostly contained sequences from Argentina. This result showed a brief local transmission period of the Alpha variant during the early second wave, but without establishment in the NWA region.

**VOI Zeta (lineage P.2).** Four genomes belonging to this lineage were detected during the early second wave (March-May 2021). Two of them belonged to the Province of Salta and formed a cluster along with sequences from Buenos Aires (Fig. 5, clade D). The other sequences corresponded to sporadic introductions without a phylogenetic relationship with other sequences from the NWA region.

**VOI Epsilon (lineage B.1.427).** Nine sequences from the Province of Tucumán were assigned to this lineage, which was originally detected in California, the United States. The genomes were grouped into two clusters (Fig. 8, clades A and B), indicating that there were local transmission chains between March and May 2021.

**Lineage A.2.5.** Nine genomes from NWA (Salta and Tucumán) were assigned to this lineage and were grouped into a monophyletic group (Fig. 9, clade A) with sequences from other Argentinean provinces (Córdoba and Catamarca) and Chile. Internal diversification with three subclades can also be observed.

### 3.4. Phylodynamics of Argentinean lineages in the NWA region

Phylogenetic analyses were performed on lineages B.1.499 and N.5, the two main Argentinean lineages that circulated in the first wave and the early second wave in the NWA region.

**Lineage B.1.499** showed a rate of evolution of $5.6 \times 10^{-4}$ substitutions per site per year ($s/s/y$) (HPD95% = $4.5 \times 10^{-4}$–$6.9 \times 10^{-4}$) and the date of its most recent common ancestor was estimated to be February 12, 2020 (HPD95% = January 8–March 9). In addition, the demographic reconstruction of the lineage showed an increase in the effective number of infections in EW22/2020 (end of May) that persisted...
until a maximum in EW45/2020 (beginning of November) and decreased since then (Fig. 10).

Lineage N.5 showed a rate of evolution of $7.9 \times 10^{-4}$ s/s/y (HPD95% = $5.6 \times 10^{-4} - 1.0 \times 10^{-3}$) and a most recent common ancestor dated on August 23, 2020 (HPD95% = June 15–October 17), while its demographic reconstruction showed no substantial changes, only a trend of a slight increase over time (Fig. 10).

3.5. Mutation analyses

Genomes sequenced obtained from NWA were compared with the Wuhan-Hu-4 reference sequence (GISAID: EPI_ISL_402124) to look for mutations in the Spike gene in the lineages with greater predominance in this study (Fig. 11). In addition, characteristic mutations of VOCs were specially analyzed and were mostly found in single sequences but not in monophyletic groups as signature markers (Table 1).

In addition, we analyzed signature mutations in 19 phylogenetic groups containing Argentinean sequences in different lineages. In total, 42 group-defining mutations were observed: 22 were non-synonymous mutations, 19 were synonymous mutations, and one of them corresponded to a substitution in the non-coding 5’ region of the viral genome (Table 2).

Non-synonymous mutations were found particularly in the ORF1a ($n = 12$) and S ($n = 7$) genes and, to a lesser extent, in the ORF1b, N, and ORF8 genes with one mutation recorded in each of them. In the case of the synonymous mutations, they were mainly detected in the ORF1a ($n = 6$) and ORF1b ($n = 6$) genes and were also observed in the S ($n = 2$), ORF7a ($n = 2$), M ($n = 1$) and ORF3a ($n = 1$) genes.

In the Spike, the non-synonymous mutations were: H49Y (C21707), P681H (C23604A), S373P (T22679C), T51I (C21714T), T95A (A21845G), A701V (C23664T), G1085E (G24816A). Some of these mutations are of particular interest.

The H49Y mutation is one of the changes that characterize the clade A of VOC Alpha from the NWA (Fig. 6, clade A) that also appeared in the lineage B.1.499 (GISAID: EPI_ISL_1396094). The mutation P681H, characteristic of the VOC Alpha, VOC Omicron, and VOI Mu, was found in clade B (lineage B.1.1.274) (Fig. 6, clade B). Finally, the A701V mutation, observed in the VOC Beta and the VOI Iota, was also associated with clade I of lineage N.5 (a clade with nine defining-group mutations) (Fig. 4, clade I).

4. Discussion

In this work, we analyzed the evolution of SARS-CoV-2 during the first year of the COVID-19 pandemic in Northwestern Argentina, covering the first and the early second wave in the region.

The first wave was dominated by the lineage B.1.499, an almost exclusive Argentinean lineage. Its common ancestor was estimated to be dated in mid-February 2020, although the first sample in the country was detected at the beginning of April 2020, in a sample from NWA. The
demographic reconstruction of lineage B.1.499 showed an increase in the effective number of infections that matched the reported cases in the National Surveillance System during the first wave in the NWA, evidencing that this was the lineage that mainly impulses this wave in the region.

The second most prevalent lineage in the first wave in the NWA was the lineage N.5, derived from B.1.1.33, a Brazilian lineage that dispersed mainly into South American countries in the first months of the pandemic (Resende et al., 2021). In this case, the phylodynamic analysis showed a common ancestor diversifying in late-August 2020, much earlier than the first detection in the NWA region (beginning of January 2021), suggesting that the viruses that reached the NWA originated from different transmission chains previously established in other regions of the country.

On the other hand, it was suggested that the emergence of the VOCs Alpha to Delta was driven by an acceleration of the rate of evolution in their branch stems, with estimations rounding an acceleration of 4-fold, with a mean “background” (non-VOC) rate of evolution of $5.8 \times 10^{-4}$ s/s/y and a rate of $2.5 \times 10^{-3}$ s/s/y for the VOC stems (Tay et al., 2022). These results encourage the monitoring of the molecular evolution of SARS-CoV-2 as a tool to understand the context in which a lineage of epidemiological interest can emerge. In this framework, the analysis of the evolutionary dynamics of the two most prevalent lineages in NWA also revealed that the rate of evolution of lineage N.5 ($7.9 \times 10^{-4}$ s/s/y)
was a ~40% faster than that of lineage B.1.499 \((5.9 \times 10^{-4} \text{ s/s/y})\), although both are in the same order of magnitude than other non-VOC lineages (Tay et al., 2022). These results showed the importance of analyzing rates of evolution separately for different lineages, whenever possible, rather than assuming a fixed rate for all, to improve the understanding of the evolutionary dynamics of SARS-CoV-2 lineages.

It should be noted that for the phylodynamic analyses included in this work, a selection of sequences belonging to the lineages B.1.499 and N.5 from the NWA was made to focus on the evolutionary dynamics of viruses from this region. However, to deepen the study of the behavior of the lineages, sequences from the entire country and a longer period would be necessary.

Some few mutations associated with a biological characteristic of importance were observed as signatures markers of the phylogenetic groups established in Northwestern Argentina, and even though positive selection cannot be ruled out, their defining-groups mutations more probably reflect neutral evolution, a major characteristic of the first wave all around the world (Dearlove et al., 2020). Like other first-wave lineages, except for the D614G in Spike (Korber et al., 2020), the P323L in the viral RNA-dependent RNA polymerase protein (Garvin et al., 2020) and other sporadic mutations in Spike, they did not present mutations associated with adaptation to the host as lineages markers (as VOC did) (Martin et al., 2021; Peacock et al., 2021). However, single sequences in non-VOC lineages did present mutations associated with VOCs, showing that many of these mutations could emerge from circulation in the general population.

During the early second wave, the first-wave lineages were displaced by the introduction of VOCs (Alpha, Gamma), VOIs (Lambda, Zeta, Epsilon), and other lineages with more limited distribution.

It is worth noting that VOC Alpha did not become established in the NWA, as it was observed in most other regions outside of South America (Hodcroft, 2021). On the contrary, VOC Gamma (lineage P.1 and derivatives) and VOI Lambda (lineage C.37) were of importance in the second wave of the COVID-19 pandemic in South America, particularly in Argentina (Faria et al., 2021; Romero et al., 2021; Torres et al., 2021).

Several introductions were observed for these variants in the NWA, with sequences associated with samples from other regions of the country or neighboring countries (Brazil and Chile). However, the limited information on molecular surveillance of SARS-CoV-2 in Bolivia and Paraguay does not allow for an evaluation of the complete landscape of the region in the studied period.

Within the NWA region, and despite being composed of three provinces that share borders, it was shown that different routes of viral dispersion impact the epidemiology of the region. For instance, Salta and Jujuy showed evidence of common transmission chains in different lineages, whereas sequences from Tucumán associated with sequences from other provinces in the country, suggesting that internal movement...
due to commercial and touristic routes impact differentially on the provinces.

Finally, this study contributed to the knowledge about the evolution of SARS-CoV-2 in a pre-vaccination and without post-exposure immunization period, gaining relevance for the understanding of the evolutionary behavior of SARS-CoV-2 in a naïve environment. These analyses can establish a background level for new analyses on these or other

Table 1
Significant mutations and their properties in the Spike protein of Northwestern Argentinean sequences from predominant lineages during the first and the early second wave.

| Mutation | GISAID ID | Lineage | VOC/VOI | Biological properties |
|----------|-----------|---------|---------|-----------------------|
| E484K + Δ144 | EPI_ISL_2940196 | B.1.499 | E484K: | Associated with an increment of the immunological resistance to neutralizing elicited by monoclonal and human serum antibodies (Zhou and Wang, 2021). |
| Δ144: | Alpha, Omicron and Eta. | | | |

Table 2
A detailed description of mutations harboured by the monophyletic groups in lineages of the first and the early second wave in the NWA.

| Lineage | Name | Nucleotide mutation | Genes/region | Effect | Amino Acid change |
|---------|------|---------------------|--------------|--------|------------------|
| A.2.5 | A | G443A | ORF1a/msp1 | Non-synonymous | V60I |
| B.1.427 | A | C27143T | ORF1b/msp12 | Synonymous | N207- |
| T15741C | ORF1b/msp12 | Synonymous | N767- |
| B.1.1.7 | A | T2169G | ORF1a/msp2 | Non-synonymous | F455C |
| G14316A | ORF1b/msp12 | Synonymous | Q292- |
| G16221A | ORF1b | Synonymous | P927- |
| G19684T | ORF1b/msp15 | Non-synonymous | H49Y- |
| C21707T | Spike | Non-synonymous | V22L |
| B.1.1.274 | B | C23604A | Spike | Non-synonymous | P681H |
| A25998T | ORF3a | Synonymous | V202- |
| C29375T | N | Synonymous | P683S |
| B.1.1.348 | C | G5572T | ORF1a/msp3 | Non-synonymous | M951I |
| D | T22679C | Spike | Non-synonymous | M951V |
| B.1.499 | A | C7279T | ORF1a/msp3 | Non-synonymous | F1520- |
| C7528T | ORF1a | Synonymous | V1603- |
| B | A23989T | Spike | Synonymous | P809- |
| C11750T | ORF1a/msp6 | Non-synonymous | H47- |
| D | T27534C | ORF7a | Non-synonymous | L260F |
| F | C2334T | ORF1a/msp2 | Non-synonymous | A510V |
| C28054A | ORF8 | Non-synonymous | S54X |
| G | A21845G | Spike | Non-synonymous | T95A |
| A.2.5 | B | C7765T | ORF1a/msp3 | Synonymous | L31- |
| C7765T | ORF1a | Synonymous | —— |
| C3130T | ORF1a/msp3 | Synonymous | Y137- |
| N.5 | A | C17241T | Spike | Non-synonymous | T51I |
| C2344T | ORF1a | Non-synonymous | A510V |
| C28054A | ORF8 | Non-synonymous | S54X |
| G | A21845G | Spike | Non-synonymous | T95A |
| L.18F | Beta and Gamma | | | |
| B | C12049T | ORF1a/msp7 | Synonymous | N69- |
| D | G22801T | Spike | Synonymous | G413- |
| C12049T | ORF1a/msp7 | Synonymous | G413- |
| A454G | ORF2 | Synonymous | Q63- |
| C12049T | ORF1a/msp7 | Synonymous | G413- |
| C12049T | ORF1a/msp7 | Synonymous | G413- |
| C12049T | ORF1a/msp7 | Synonymous | G413- |
| C12049T | ORF1a/msp7 | Synonymous | G413- |
| C12049T | ORF1a/msp7 | Synonymous | G413- |
| C12049T | ORF1a/msp7 | Synonymous | G413- |
| C12049T | ORF1a/msp7 | Synonymous | G413- |
| C12049T | ORF1a/msp7 | Synonymous | G413- |
| C12049T | ORF1a/msp7 | Synonymous | G413- |
| C12049T | ORF1a/msp7 | Synonymous | G413- |
| C12049T | ORF1a/msp7 | Synonymous | G413- |
| C12049T | ORF1a/msp7 | Synonymous | G413- |
| C12049T | ORF1a/msp7 | Synonymous | G413- |
| C12049T | ORF1a/msp7 | Synonymous | G413- |
| C12049T | ORF1a/msp7 | Synonymous | G413- |

(continued on next page)
Table 2 (continued)

| Lineage | Clade name | Nucleotide mutation | Gene/region | Effect | Amino Acid change |
|---------|------------|---------------------|-------------|--------|-------------------|
| P.1     | A          | G24816A             | Spike       | Non-synonymous | G108SE          |
|         |            |                     |             |        |                   |

* The names of the clades correspond to those mentioned in the figures of each lineage.

† Dash (-) represents amino acid maintenance (synonymous change).

‡ H49Y is related to greater stabilizing interaction of the ACE2-Spike complex (Aljindan et al., 2021).

§ P681H and A701V do not play a role in viral infectivity and neutralization potential. However, when in conjunction with the RBD–N501Y mutation, viral infectivity is enhanced (Kuzmina et al., 2022).

SARS-CoV-2 lineages in later periods when other factors that may have an impact on the viral evolution and emergence of variants appeared, such as post-exposure immunity, vaccination, waning immunity, among others.

CRediT authorship contribution statement

**Romina Zambrana Montano:** Formal analysis, Data curation, Investigation, Visualization, Writing – original draft, Writing – review & editing.
**Andrés Carlos Alberto Culasso:** Methodology, Formal analysis, Investigation.
**Franco Fernández:** Methodology, Data curation.
**Nathalie Marquez:** Methodology, Data curation.
**Humberto Debat:** Methodology, Data curation.
**Ana Maria Zamora:** Methodology, Data curation.
**Gustavo Ruiz de Huidobro:** Methodology, Data curation.
**Miguel Alejandro Charre:** Methodology, Data curation.
**Ariel David Fridman:** Methodology, Data curation.
**Claudia Maza Díaz:** Methodology, Data curation.
**Viviana Rasovsky:** Methodology, Data curation.
**Esteban Lavaque:** Methodology, Data curation.
**Veronica Lesser:** Methodology, Data curation.
**Pamela Cajal:** Methodology, Data curation.
**Fernanda Agüero:** Methodology, Data curation.
**Cintia Calvente:** Methodology, Data curation.
**Carolina Torres:** Visualization, Formal analysis, Conceptualization, Funding acquisition, Investigation, Supervision, Writing – original draft, Writing – review & editing.
**Mariana Viegas:** Project administration, Conceptualization, Funding acquisition, Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Data Availability

Data are available at https://doi.org/10.55876/gis8.220621vp and https://doi.org/10.55876/gis8.220615tq.

Acknowledgment

This work was supported by Proyecto IP COVID-19 N° 08 (Ministerio de Ciencia, Tecnología e Innovación, Argentina) and Focem COF 03/11 Covid-19 (Fondo para la Convergencia Estructural del MERCOSUR).

for publication.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.viress.2022.198936.

References

Aljindan, R.Y., Al-Sbahie, A.M., Al-Ohtal, A.J., et al., 2021. Investigation of non-synonymous mutations in the spike protein of SARS-CoV-2 and its interaction with the ACE2 receptor by molecular docking and MM/GBSA approach. Comput. Biol. Med. 135, 104654.

Bersani, A., D’urso, P., Campisi, G., et al., 2021. Role of Q675H mutation in improving SARS-CoV-2 spike interaction with the furin binding pocket. Viruses 13, 1–17.

Dearlove B., Lewitus E., Bai H. et al. A SARS-CoV-2 vaccine candidate would likely match all currently circulating variants. 2020;117:23652–62.

Drummond, A.J., Ho, S.Y.W., Phillips, M.J., et al., 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4, 699–710.

Faría, N.R., Mellan, T.A., Whittaker, C., et al., 2021. Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Malaysia. Brazil. Science 372, 815–821 (80–).

Ferreira, I.A.T.M., Kemp, S.A., Dutri, R., et al., 2021. SARS-CoV-2 B.1.1617 Mutations I455R and E484Q Are Not Synergistic for Antibody Evasion. J. Infect. Dis. 224, 989–994.

Garvin, M.R., Prates, E.T., Pavicic, M., et al., 2020. Potentially adaptive SARS-CoV-2 mutations discovered with novel spatiotemporal and explainable AI models. Genome Biol. 21 https://doi.org/10.1186/s13059-020-02191-0.

Ganey, A.J., Los, A.N., Gentles, L.E., et al., 2021. Antibodies elicited by mRNA-1273 vaccination bind more broadly to the receptor binding domain than do those from SARS-CoV-2 infection. Sci. Transl. Med. 13, 1–13.

Guindon, S., Dufayard, J.F., Lefort, V., et al., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: the performance of PhyML 3.0. Syst. Biol. 59, 370–372.

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids Symp. Ser. 41, 95–98.

Hoang, D.T., Chermomor, O., Van Haeseler, A., et al., 2018. UPPBoot2: improving the outgroup bootstrap approximation. Mol. Biol. Evol. 35, 518–522.

Hodcroft, E.B., 2021. SARS-CoV-2 Mutations and Variants of Interest. CoVariants. DOI. https://covariants.org/.

Kalyanamorthy, S., Minh, B.Q., Wong, T.K.F., et al., 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods 14, 587–589.

Katoh, K., Rozewicki, J., Yamada, K.D., 2018. MAFFT online service: multiple sequence alignment. interactive sequence choice and visualization. Brief. Bioinform. 20, 1160–1166.

Korber, B., Fischer, W.M., Gnanaskaran, S., et al., 2020. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 Virus. Cell 182, 812–827.e19.

Kuzmina, A., Wattad, S., Engel, S., et al., 2022. Functional analysis of spike from SARS-CoV-2 variants reveals the role of distinct mutations in neutralization potential and viral infectivity. Viruses 14. https://doi.org/10.3390/v14040803.

Li, Q., Wu, J., Nie, J., et al., 2020. The impact of mutations in SARS-CoV-2 Spike on viral infectivity and antigenicity. Cell 182, 1284–1294.e9.

Martin, D.P., Weaver, S., Tegally, H., et al., 2021. Convergent evolution of the SARS-CoV-2 N501Y lineages. Cell. https://doi.org/10.1016/j.cell.2021.09.003.

McCallum, M., De Marco, A., Lempp, F.A., et al., 2021. N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2. Cell 184, 2332–2347.e16.

Miller, M.A., Pfiffer, W., Schwartz, T., 2010. Creating the CIPRES science gateway for inference of large phylogenetic trees. In: Proceedings of the Gateway Computing Environments Workshop (GCE), pp. 1–5.

Minh, B.Q., Schmidt, H.A., Chernomor, O., et al., 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. 37, 1530–1534.

Minin, V.N., Bloomquist, E.W., Suchard, M.A., 2008. Smooth skytree through a rough skyline: Bayesian coalescent-based inference of population dynamics. Mol. Biol. Evol. 25, 1459–1471.

Ministry of Science Technology and Innovation, 2020. Argentine Interinstitutional SARS-CoV-2 Genomics Project. Ministry of Science Technology and Innovation. http://intis.gbi.fien.uba.ar/.

Ministry of Health of the Nation, 2020a. COVID-19. Cases registered in the Argentine Republic. Ministry of Health of the Nation. http://datos.salud.gob.ar/dataset/covid-19-casos-registrados-en-la-republica-argentina.

Ministry of Health of the Nation, 2020b. What measures is the government taking? Ministry of Health of the Nation. https://www.argentina.gob.ar/coronavirus/medidas-gobierno.

Ministry of Health of the Nation, 2021. Online Coronavirus Situation Room. Ministry of Health of the Nation. https://www.argentina.gob.ar/sites/default/files/2021/09/sala-situacion-covid-19,30-09-2021.pdf.

Nguyen, L.T., Schmidt, H.A., Van Haeseler, A., et al., 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32, 268–274.

O’Toole, A., Scher, E., Underwood, A., et al., 2021. Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. Virus Evol. 7, 1–9.
Peacock, T.P., Penrice-Randal, R., Hiscox, J.A., et al., 2021. SARS-CoV-2: one year on: evidence for ongoing viral adaptation. J. Gen. Virol. 102 https://doi.org/10.1099/jgv.0.001584.

Rambaut, A., Holmes, E.C., O’Toole, Á., et al., 2020. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat. Microbiol. 5, 1403-1407.

Rambaut, A., Lam, T.T., Carvalho, I.M., et al., 2016. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evol. 2, 1-7.

Rearte, A., Castelli, J.M., Rearte, R., et al., 2022. Effectiveness of rAd26-rAd5, ChAdOx1 nCoV-19, and BBIBP-CorV vaccines for risk of infection with SARS-CoV-2 and death due to COVID-19 in people older than 60 years in Argentina: a test-negative, case-control, and retrospective longitudinal study. Lancet 399, 1254-1264.

Resende, P.C., Delatorre, E., Graf, T., et al., 2021. Evolutionary Dynamics and dissemination pattern of the SARS-CoV-2 Lineage B.1.1.33 during the early pandemic phase in Brazil. Front. Microbiol. 11 https://doi.org/10.3389/fmicb.2020.615280.

Ritchie, H., et al., 2020. Coronavirus Pandemic (COVID-19). Our World Data. https://ourworldindata.org/coronavirus.

Romero, P.E., Dávila-Barclay, A., Salvatierra, G., et al., 2021. The emergence of SARS-CoV-2 Variant Lambda (C.37) in South America. Microbiol. Spectr. 9 https://doi.org/10.1128/spectrum.00789-21.

Tay, J.H., Porter, A.F., Wirth, W., et al., 2022. The emergence of SARS-CoV-2 variants of concern is driven by acceleration of the substitution rate. Mol. Biol. Evol. 39 https://doi.org/10.1093/molbev/msac013.

Torres, C., Mojstiejczuk, L., Acuna, D., et al., 2021. Cost-effective method to perform SARS-CoV-2 variant surveillance: detection of alpha, gamma, lambda, delta, epsilon, and zeta in Argentina. Front. Med. 8 https://doi.org/10.3389/fmed.2021.755463.

Wang, P., Nair, M.S., Liu, L., et al., 2021. Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. Nature 593, 130-135.

Zeng, C., Evans, J.P., Farazone, J.N., et al., 2021. Neutralization of SARS-CoV-2 variants of concern harboring Q677H. mBio 12, 1-7.

Zhou, W., Wang, W., 2021. Fast-spreading SARS-CoV-2 variants: challenges to and new design strategies of COVID-19 vaccines. Signal Transduct. Target. Ther. 6, 1–6.