

Reporte nº17 – Provincia de Córdoba – 19 de marzo de 2021

Consortio interinstitucional para la Secuenciación del genoma y estudios genómicos de SARS-CoV-2
(Proyecto PAIS)

Caracterización genómica de muestras de la provincia de Córdoba entre abril de 2020 y enero de 2021.

RESUMEN

Se realizó la secuenciación del **genoma completo** de SARS-CoV-2 a partir de **93** muestras de la **provincia de Córdoba** correspondientes al periodo comprendido entre **abril de 2020** y **enero de 2021**. **Las secuencias obtenidas fueron asignadas a 6 linajes: B.1.499 (48,4%), N.3 (39,8%), B.1 (3,2%), B.1.1.33 (4,3%), B.1.1.222 (3,2%) y B.1.2 (1,1%)**. Si bien los linajes observados coinciden con los reportados previamente en Argentina, en Córdoba se registró un cambio en la proporción de los linajes en circulación reciente respecto a los previamente detectados en esta región.

Como parte de la evolución natural del virus, se encontraron cambios aminoacídicos en distintas regiones del genoma viral de algunas secuencias, entre ellas en la región que codifica para la proteína Spike. Sin embargo, las implicancias de estos cambios se encuentran bajo análisis y no revisten por el momento consecuencias en la biología viral.

En este informe analiza la circulación de SARS-CoV-2 en la provincia de Córdoba desde el 9/4/2020 al 6/1/2021 (SE 15/2020 a SE 1/2021).

El principal **objetivo** es describir la introducción, circulación y establecimiento de los linajes de SARS-CoV-2 en la provincia de Córdoba y profundizar el estudio del origen y evolución de los brotes relevantes en la provincia.

Este informe complementa los resultados generados en los **Reportes Nº5 y Nº14** cuyos objetivos centrales fueron identificar los linajes virales que ingresaron a Córdoba y monitorear la circulación local del virus a través del muestreo de pacientes con antecedente de viaje a zonas afectadas y del estudio de infecciones adquiridas en la comunidad. En este reporte se realiza además el análisis integrado junto a secuencias previas de SARS-CoV-2 provenientes de Córdoba y de otras provincias de Argentina, obtenidas desde marzo del 2020 por el Consorcio PAIS.

Situación previa de los linajes se SARS-Cov-2 circulantes en la provincia de Córdoba.

En el mes de agosto 2020 se reportó la circulación de los linajes SARS-CoV-2 obtenidos a partir de 22 muestras correspondientes al período marzo (inicio de la pandemia) hasta el 15 de abril 2020, más 3 muestras del mes de julio de relevancia epidemiológica (Reporte PAIS N° 5, <http://pais.qb.fcen.uba.ar/files/reportes/pais-reporte5.pdf>).

Mediante análisis filogenético se determinó que los 22 genomas obtenidos de muestras de la provincia de Córdoba pertenecían al linaje B.1 o sus (sub)linajes derivados, los cuales presentan amplia distribución mundial y habían sido descriptos para las secuencias de SARS-CoV-2 de otras regiones de Argentina.

Situación en el territorio nacional durante el período de análisis.

Respecto de la situación epidemiológica de la nación, al inicio del período de análisis de este informe, se encontraba en vigencia el Aislamiento Social, Preventivo y Obligatorio (ASPO) estricto en todo el territorio argentino, la circulación entre provincias se encontraba restringida y fronteras internacionales cerradas. En el mes de junio la mayor parte del territorio pasó a fase de distanciamiento. Solo cuatro áreas urbanas mantuvieron el aislamiento: el Área Metropolitana Buenos Aires (AMBA), el Departamento San Fernando de la provincia del Chaco, el Departamento Rawson de la provincia de Chubut y el Área Metropolitana Córdoba. El 1° de julio toda la provincia de Córdoba entra a fase de distanciamiento social (DISPO). A partir de diciembre 2020 quedó habilitado el turismo a casas de veraneo en Córdoba para quienes no residían en la provincia y el 1 de enero comenzó la temporada de turismo para el resto del país en el territorio provincial.

Situación epidemiológica de la provincia de Córdoba.

Desde el inicio de la pandemia hasta enero de 2021, en la provincia de Córdoba se realizaron tests con RT-PCR a 594.818 personas. Esto resultó en una tasa de 158.177 personas estudiadas por RT-PCR por cada 1 millón de habitantes. La Figura 1 muestra las determinaciones realizadas mediante RT-PCR por mes.

Hasta el momento del cierre del muestreo se reportaron 130.480 casos confirmados de COVID-19. La Figura 2 muestra los casos confirmados por RT-PCR por mes durante el 2020. El total de decesos por COVID19 ascendió a 2.529, mientras que la proporción de altas en relación con el total de casos confirmados en la provincia de Córdoba fue del 93%.

Finalmente, la Figura 3 muestra la situación ante la pandemia de COVID19 en la provincia de Córdoba, con relación a la evolución del número de casos confirmados, recuperados, fallecidos y activos.

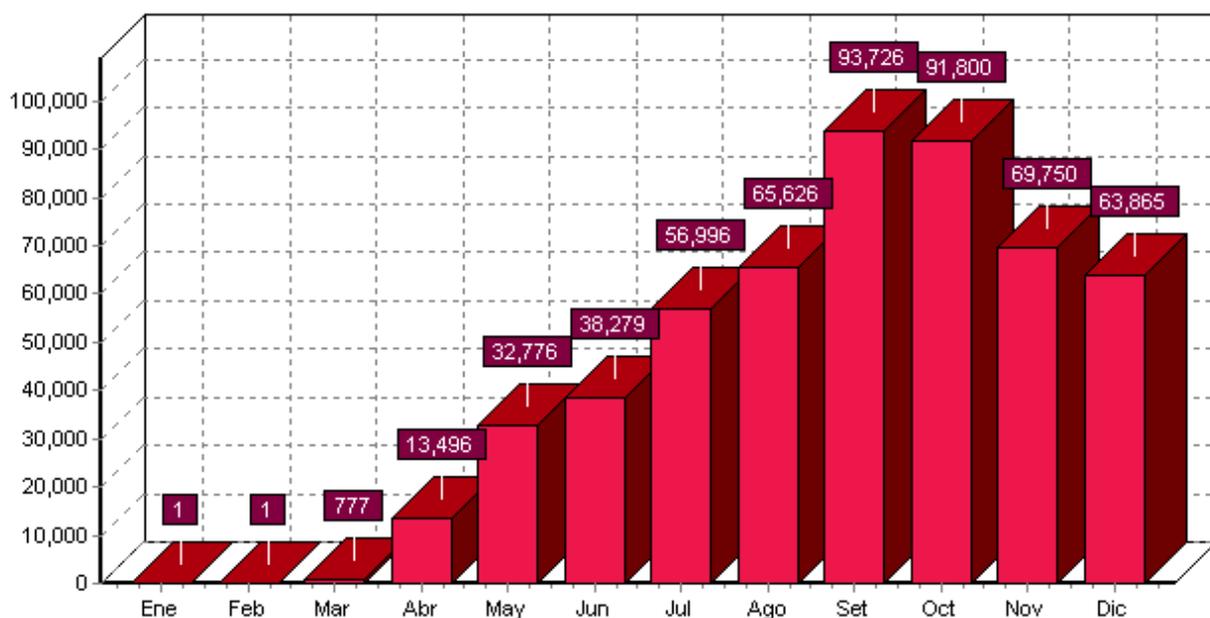


Figura 1. Determinaciones realizadas mediante RT-PCR para detección de SARS-CoV-2 por mes durante el año 2020 en la provincia de Córdoba.

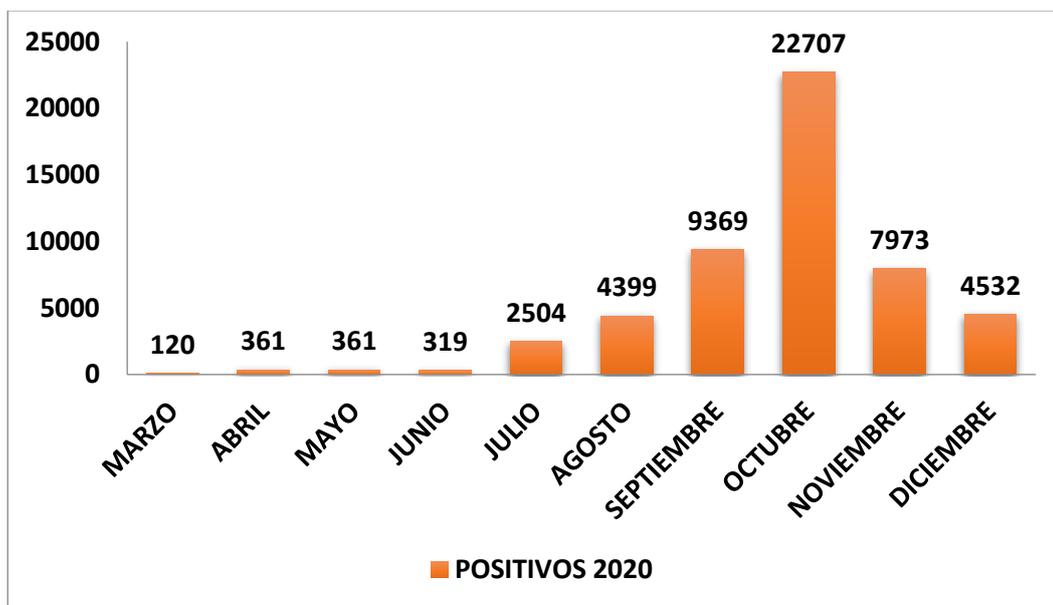


Figura 2. Casos positivos para SARS-CoV-2 por RT-PCR, por mes durante el año 2020 en la provincia de Córdoba.

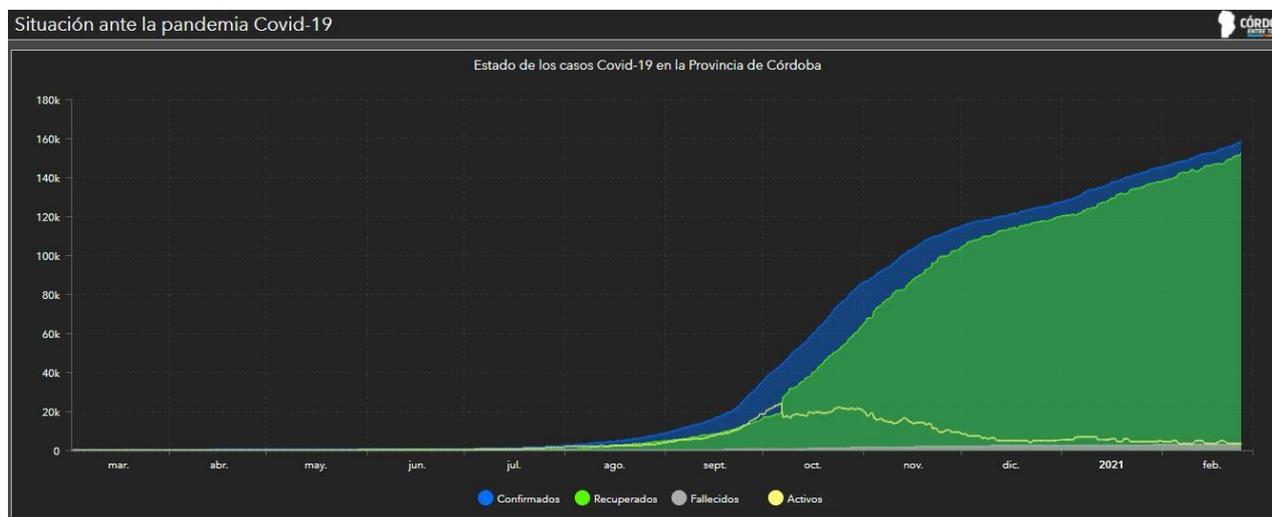


Figura 3. Estado de los casos positivos para la COVID-19 en la provincia de Córdoba. Casos confirmados (azul), recuperados (verde), fallecidos (gris) y activos (amarillo).

Caracterización genómica de SARS-CoV-2 de la provincia de Córdoba

MUESTREO

El criterio de selección de muestras de la provincia para secuenciar consistió en:

- El análisis de diferentes brotes que se presentaron en dicho período:
 - Brote del geriátrico de Saldán (abril-mayo 2020).
 - Brote de Villa Dolores (junio-julio 2020).
- Residentes de localidades con transmisión comunitaria: se seleccionaron muestras de las principales ciudades de Córdoba (aproximadamente 20 muestras de cada mes).

La distribución de muestras seleccionadas según localidad fue la siguiente:

Córdoba Capital (**36**), Río Cuarto (**13**), Villa María (**6**), San Francisco (**5**), Río Tercero (**4**), Oncativo (**4**), Villa Dolores (**4**), Unquillo (**2**), Marcos Juárez (**2**), Monte Cristo (**1**), Oliva (**1**), General Deheza (**1**), Saldán (**1**), Santa Rosa de Calamuchita (**1**), La Falda (**1**), Alta Gracia (**1**), Las Varillas (**1**), Río Segundo (**1**), Monte Buey (**1**), Bell Ville (**1**), Río Ceballos (**1**), Arroyito (**1**), Jesús María (**1**), Monte Maíz (**1**), Corral de Bustos (**1**) y Colonia Caroya (**1**).

Todas las muestras que se analizaron corresponden a extractos de ARN obtenidos de hisopados orofaríngeos y/o saliva confirmados para la COVID-19 por la técnica de referencia de RT-PCR en tiempo real (kit DisCOVERY) y que cumplían con el criterio de una carga viral medida en valores de Ct por debajo de 30.

Estos criterios de selección pueden implicar una subestimación de la diversidad de linajes, sobre todo de los casos que son independientes entre sí y no tienen ningún nexo con *clusters* epidemiológicos identificados.

Estrategia de secuenciación empleada:

La secuenciación de genomas completos fue realizada siguiendo el protocolo de amplificación y secuenciación de ARTIC para MinION modificado (<https://dx.doi.org/10.17504/protocols.io.bbmuik6w>).

Se realizaron los controles necesarios para asegurar calidad en cada una de las corridas siguiendo el SOP “Validación de método de secuenciación de muestras de SARS-CoV-2” desarrollado por la Unidad de Genómica del IABIMO-INTA Castelar del Consorcio PAIS.

RESULTADOS OBTENIDOS

Secuencias obtenidas

Se obtuvieron un total **93 secuencias genómicas de SARS-CoV-2** (profundidad promedio ~700X) provenientes de muestras clínicas de pacientes con la COVID-19.

Estas secuencias serán subidas a la base de datos internacional GISAID una vez concluidos los análisis pertinentes relativos a esta etapa, mientras tanto estarán disponibles en un repositorio local al cual se puede acceder con autorización del Consorcio PAIS.

Distribución de linajes encontrados

Las secuencias del genoma de SARS-CoV-2 obtenidas a partir de individuos infectados de distintos países han sido clasificadas según una taxonomía dinámica de linajes propuesta por el grupo del Dr. Rambaut, cuya última actualización corresponde al 21/02/2021 (Rambaut y col., 2020). Así, de acuerdo con cambios nucleotídicos y su agrupamiento filogenético, el SARS-CoV-2 se clasifica en varios linajes principales y numerosos linajes derivados, denominados con letras y números.

Para el análisis de las secuencias de Córdoba, se realizó la asignación rápida de linajes mediante el programa Pangolin COVID-19 Lineage Assigner (<https://pangolin.cog-uk.io/>) Esta herramienta permite determinar el linaje a través de la identificación de patrones complejos mediante *machine learning*.

Se realizó la confirmación de linajes asignados mediante análisis filogenético, como se describe en el Anexo 1. El análisis e interpretación de los resultados fue realizado por los integrantes del Nodo de Evolución y los grupos de trabajo de la provincia de Córdoba.

Mediante análisis filogenético se determinó que todas las secuencias de la provincia de Córdoba pertenecieron al linaje B.1 o sus (sub)linajes derivados, que posee amplia distribución tanto a nivel mundial como en Argentina.

En mayor detalle, se encontraron representantes de los linajes B.1.499 (45 secuencias: 48,4 %), B.1.1.33 (4 secuencias: 4,3 %) y de su linaje derivado N.3 (37 secuencias: 39,8 %), B.1 (3 secuencias: 3,2 %), B.1.1.222 (3 secuencias: 3,2 %) y B.1.2 (1 secuencia: 1,1 %).

En el gráfico de la Figura 4 se muestra la distribución de linajes hallados en la Provincia de Córdoba hasta el momento, incluyendo las secuencias del Reporte PAIS N°5 y del presente.

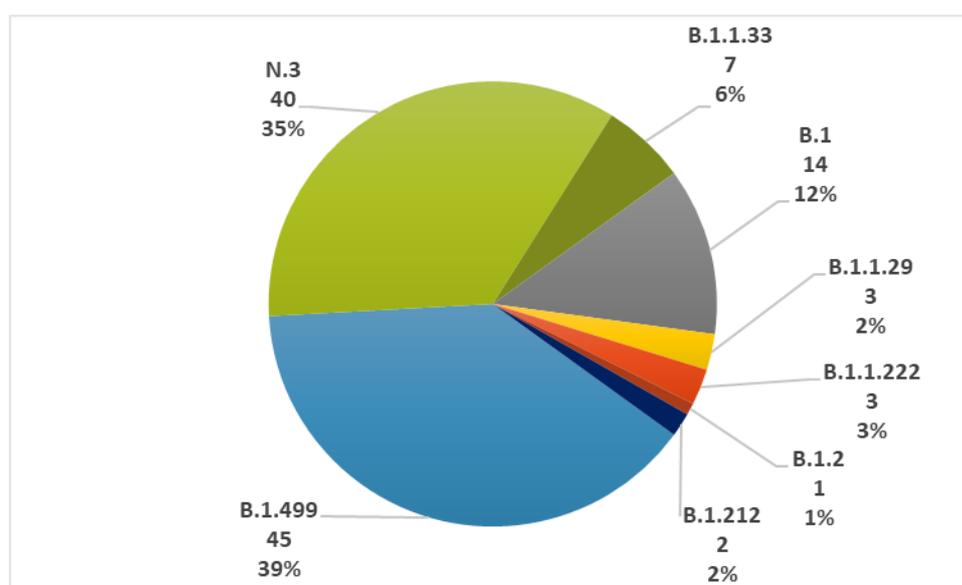


Figura 4. Distribución de Linajes de SARS-CoV-2 hallados en la Provincia de Córdoba (período abril 2020 a enero 2021).

En la Figura 5 se observa la distribución de linajes de SARS-CoV-2 por mes en el período abril 2020 a enero 2021 en la provincia de Córdoba, donde puede observarse un cambio en los linajes detectados. Así, durante los meses de marzo-abril predomina el linaje B.1, y se relaciona con introducciones a partir de personas con antecedentes de viaje y a los dos primeros brotes en Córdoba (en Saldán y en un hospital de la ciudad de Córdoba). En los meses de junio-julio se empiezan a detectar los linajes N.3 y B.1.499 en alta frecuencia. Esto sugiere que ambos linajes se establecieron en la provincia de Córdoba a partir de múltiples ingresos posiblemente desde distintas locaciones (inferido desde su relación cercana con secuencias de distintas regiones, ver sección “Análisis Filogenético”) y que han permanecido circulando en la comunidad.

Finalmente, en los meses de noviembre-enero se detectaron por primera vez los linajes B.1.2 y B.1.1.222, así como se detectó nuevamente el linaje B.1.1.33.

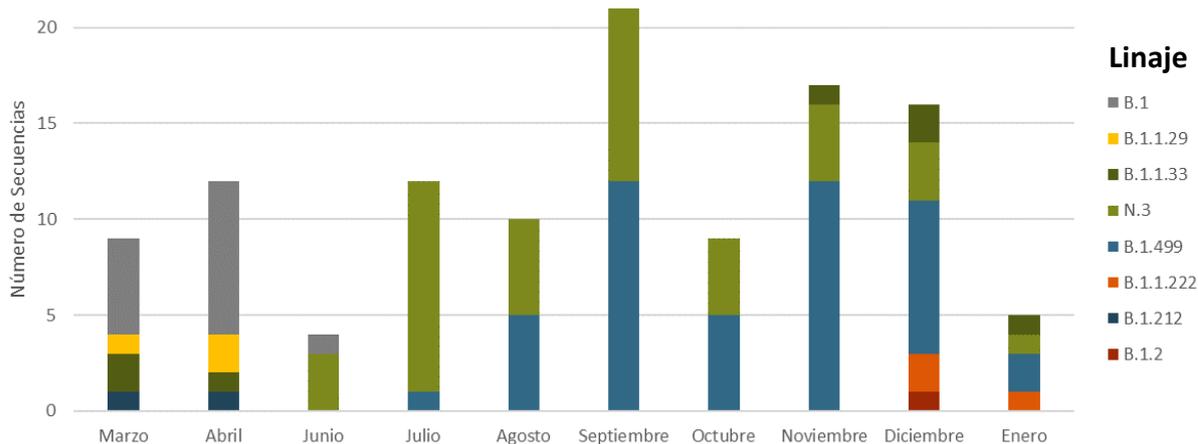


Figura 5. Distribución de linajes de SARS-CoV-2 por mes en el período abril 2020 a enero 2021 en la provincia de Córdoba

Análisis filogenético de linajes individuales

Linaje B.1

Las cuatro secuencias identificadas como B.1 se agruparon junto con otras cinco secuencias de Córdoba reportadas previamente (Figura 6, panel b). Estas cinco secuencias habían sido descritas en el primer Reporte de la provincia (previamente clasificadas como linaje B.1.5, pero re-clasificadas a B.1 según la taxonomía actualizada), corresponde al primer brote de la COVID-19 en Córdoba ocurrido en mes de abril de 2020, relacionado a una institución de cuidados de adultos mayores de la Ciudad de Saldán, departamento Colón. Entre las nuevas muestras analizadas se encuentra la del posible caso índice (PAIS-G0070, 9/4/2020): un profesional de la salud que trabajaba en esta institución, con antecedente de viaje a Chile. Otra de las muestras de este grupo corresponde a un individuo con detección prolongada de RNA viral que fue internado en un hospital de la ciudad de Córdoba y falleció (29/4/2020). Ambas muestras agruparon dentro del mismo clado con las secuencias reportadas previamente de este brote y secuencias del brote ocurrido posteriormente en un nosocomio de la ciudad de Córdoba, confirmando su origen común. La tercera muestra fue obtenida del brote ocurrido en el barrio Villa El Libertador en junio 2020. En conjunto, estas secuencias forman un *cluster* monofilético de Córdoba que sugiere un origen común y que no muestra relación cercana con otras secuencias de Argentina o de otros países reportadas hasta el momento en bases de datos. Esto sugiere que, desde una única y temprana introducción en la provincia en abril de 2020, el virus de este grupo ha permanecido en circulación comunitaria en la ciudad de Córdoba hasta al menos el mes de junio de 2020.

Linaje B.1.2

El linaje B.1.2 ha sido reportado en 51 países, pero el 97% de las secuencias corresponden a los Estados Unidos, donde se lo detectó por primera vez en enero de 2020. Una sola secuencia de la provincia de Córdoba perteneció a este linaje, proveniente de un individuo de la ciudad de Córdoba sin antecedente de viaje diagnosticado en diciembre. Esta secuencia se relacionó a un grupo de secuencias de América del Norte (Canadá y Estados Unidos), también correspondientes al mes de diciembre de 2020 (Figura 6, panel a).

Linaje B.1.1.33

El linaje B.1.1.33 ha sido descrito como un linaje con origen en Brasil, detectado por primera vez en marzo de 2020 en ese país. Si bien hasta el momento ha sido reportado en 27 países, el 73% de las secuencias corresponden a Brasil y otra gran proporción a países sudamericanos, incluyendo Argentina, Chile y Uruguay.

Las secuencias de este reporte no se relacionaron con las secuencias B.1.1.33 reportadas previamente en Córdoba, por lo que se tratarían de distintas introducciones a la provincia. En particular, las cuatro secuencias de este reporte fueron obtenidas de la ciudad de San Francisco, entre los meses de noviembre y enero, y se entremezclaron con secuencias contemporáneas provenientes de la ciudad de Santa Fe, sugiriendo una cadena de transmisión común del virus en esa región (Figura 7, panel c).

Linaje N.3

Del linaje B.1.1.33 surge, al establecerse en la Argentina, el linaje derivado N.3 (alias B.1.1.33.3) detectado por primera vez en abril de 2020. Es reconocido como un linaje argentino, ya que casi la totalidad de las secuencias reportadas en base de datos provienen de nuestro país (n=84), con excepción de un reporte de Chile y uno de Hong-Kong.

Las secuencias de Córdoba del linaje N.3 se distribuyeron en dos grandes grupos monofiléticos, ambos con soporte estadístico, que se encuentran cercanos a secuencias de CABA y provincia de Buenos Aires (Figura 7, paneles a y b).

El grupo I incluye 27 muestras de Córdoba obtenidas entre julio y enero; además, este grupo incluyó una secuencia de Merlo, provincia de Buenos Aires. Diecisiete de las 27 secuencias corresponden a la Ciudad de Córdoba, entre las cuales se entremezclan secuencias de otras localidades sin mostrar un patrón geográfico definido. La amplitud temporal y geográfica de las muestras sugiere el establecimiento y circulación comunitaria de este linaje en la provincia.

El grupo II está conformado únicamente por secuencias de la provincia de Córdoba. Tres secuencias (obtenidas en el mes de julio e informadas previamente) correspondieron a un brote detectado en la Ciudad de Villa Dolores que comenzó en junio (Reporte Nº5). Las nuevas secuencias provienen de muestras tomadas en los meses de junio-julio de 2020 que fueron seleccionadas por su relación con dicho brote. Dos fueron obtenidas de personas que asistieron a un mismo establecimiento bancario al inicio del brote: una

perteneciente a un individuo proveniente de CABA que arribó a Villa Dolores con fines laborales (16/6/2020), quien habría arribado a la localidad cordobesa infectado (posible caso índice); la otra muestra proviene de un individuo que viajó La Falda-Villa Dolores-La Falda (18/6/2020). Las dos muestras restantes, correspondieron a distintos momentos del brote (30/6/2020 y 11/7/2020). Además, este grupo monofilético incluyó secuencias provenientes de localidades ubicadas en el corredor de la ruta 9 (dos de Oncativo, una de Oliva, dos de Villa María y una de la Ciudad de Córdoba), todas obtenidas en el mes de julio. Los resultados mostraron que todas las secuencias agruparon junto a las secuencias de Córdoba obtenidas previamente formando un *cluster* monofilético derivado de los virus detectados en el AMBA. Esto confirmaría la introducción de este virus a Villa Dolores desde el AMBA y su posterior diseminación a otras localidades, como La Falda.

Linaje B.1.1.222

Este linaje circula principalmente en Norteamérica (Estados Unidos y México). En Sudamérica ha sido reportado en Brasil y Chile, en casos aislados. Este es el primer reporte del linaje en Argentina.

Las tres secuencias de este linaje provienen de la ciudad de Córdoba de los meses de diciembre 2020 y enero 2021, dos de ellas provenían de individuos que concurren a un mismo evento social. Todas comparten un ancestro común cercano y el grupo presenta alto soporte estadístico, lo que sugiere que se trataría de una única introducción a la provincia. La secuencia filogenéticamente más cercana corresponde al Reino Unido, agrupando también con secuencias de Estados Unidos y México (Figura 7, panel d).

Linaje B.1.499

Este linaje, plenamente argentino y detectado inicialmente en marzo del 2020, ha sido reportado en diferentes provincias del país, principalmente en la región del AMBA. El mismo fue el mayoritario en las secuencias analizadas en este reporte (48,4%), se detectó en varias regiones de la provincia y en diferentes momentos de la epidemia, sugiriendo su establecimiento y circulación comunitaria en la provincia de Córdoba. Llamativamente, fue el único linaje encontrado en la localidad de Río Cuarto (Figura 8). Las secuencias de Córdoba se distribuyeron en al menos cinco grupos distintos y dos secuencias aisladas, lo que sugiere múltiples introducciones a la provincia, posiblemente desde distintas regiones del país.

El grupo I está conformado por cuatro secuencias de Córdoba y una de la CABA.

El grupo II incluye dos secuencias que comparten un ancestro en común con secuencias de la provincia de Santa Fe.

El grupo III incluye tres secuencias de Córdoba, relacionadas con secuencias de Neuquén y Río Negro, el sur de la provincia de Buenos Aires (Patagones) y una de la CABA.

El grupo IV está conformado enteramente por secuencias de Córdoba obtenidas entre agosto 2020 y enero 2021 e incluye principalmente secuencias de ciudad de Córdoba (n=6), Unquillo (n=2), Villa Dolores (n=1) y San Francisco (n=1).

Proyecto Argentino Interinstitucional de genómica
de SARS-CoV-2

El grupo V es el más numeroso e incluyó a 21 secuencias de Córdoba, dos grupos internos de secuencias de La Pampa y una secuencia de Santa Fe.

Finalmente, se observaron dos secuencias aisladas. Una de ellas proviene de Alta Gracia y se incluyó en un grupo de secuencias del NOA (Figura 8, panel a), mientras que la otra corresponde a la ciudad de Marcos Juárez y agrupó con secuencias de la ciudad de Rosario (Figura 8, panel c).

Proyecto Argentino Interinstitucional de genómica de SARS-CoV-2

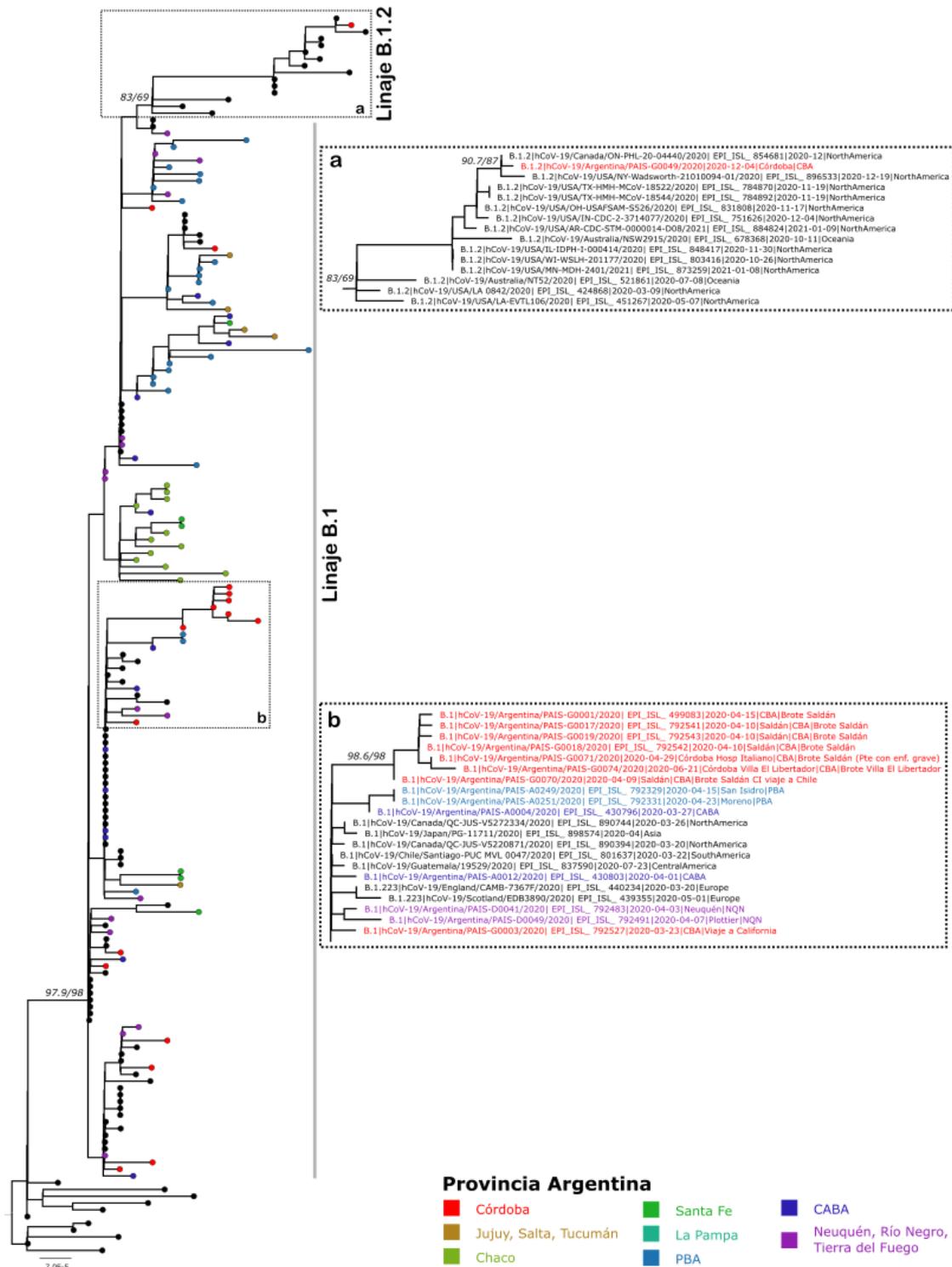


Figura 6. Árbol filogenético de los linajes B.1 y B.1.2 de SARS-CoV-2.

Proyecto Argentino Interinstitucional de genómica de SARS-CoV-2



Figura 8. Árbol filogenético del linaje B.1.499 de SARS-CoV-2. Los triángulos azules en el árbol filogenético del panel de la izquierda corresponden a grupos de secuencias del AMBA que fueron colapsados para mejorar la visualización del resto de los grupos.

Análisis de mutaciones que podrían afectar la detección de los métodos diagnósticos moleculares disponibles en el mercado.

Dado que existe gran diversidad de métodos moleculares disponibles en el país para el diagnóstico de infección por SARS-CoV-2, cada uno con sondas y cebadores específicos localizados en diferentes regiones del genoma, se analizó la presencia de mutaciones en dichas regiones en los genomas de SARS-CoV-2 de la provincia de Córdoba reportadas en este informe. Este análisis se pudo realizar sólo para aquellos ensayos que tienen la información disponible. Cabe aclarar que la presencia de estas mutaciones no implica necesariamente una disminución en la sensibilidad de dichos ensayos.

Se analizaron los cebadores y sondas de los siguientes ensayos (ver referencias en Anexo 1):

- Protocolo de Charité, Berlín. Genes reconocidos: E, N y RdRp.
- Protocolo de The University of Hong Kong. Genes reconocidos: N y RdRp.
- Protocolo de CDC Estados Unidos. Gen reconocido: N (regiones N1 y N2).
- Protocolo de CDC China. Genes reconocidos: ORF1ab y N
- Protocolo de Ministry of Public Health, Tailandia. Gen reconocido: N
- Protocolo de National Institute of Infectious Diseases, Tokio, Japón. Gen reconocido: N
- Protocolo de Institut Pasteur, Paris. Genes reconocidos: E y RdRp.
- Protocolo de KiCqStart de Sigma-Aldrich. Gen reconocido: S

Se reconocieron las secuencias nucleotídicas de los cebadores y sondas utilizando el algoritmo BLASTn a través de la aplicación de un *script* desarrollado en el nodo bioinformático del Consorcio. Se registraron las mutaciones encontradas en los genomas de Córdoba, cuya ubicación es reportada en este informe con respecto a la secuencia de referencia hCoV-19/Wuhan/WIV04/2019:

- Gen RdRp:
 - Se detectaron tres genomas con la mutación t12722c, donde se ubica la región de detección de la sonda del gen RdRp IP2 del protocolo Pasteur.
 - Se detectaron diez genomas con la mutación c14120t, donde se ubica la región de detección de la sonda del gen RdRp IP4 del protocolo del Instituto Pasteur.
 - Se detectó un genoma con la mutación c13458t, donde se ubica la región de detección del cebador antisentido del gen ORF1ab del protocolo del CDC de China.
 - Se detectó un genoma con la mutación c18796t, donde se ubica la región de detección del cebador sentido del gen RdRp del protocolo de la Universidad de Hong Kong.
- Gen S:
 - Se detectó un genoma con la mutación g23593t, donde se ubica la región de detección de la sonda del gen S del protocolo Sigma.

o Se detectaron dos genomas con la mutación c23604a, donde se ubica la región de detección de la sonda del gen S del protocolo Sigma.

- Gen E:

- o Se detectó un genoma con la mutación c26270t, donde se ubica la región de detección del cebador sentido del gen E del protocolo del Instituto Pasteur y de Charité, Berlín.

- Gen N:

- o Se detectó un genoma con la mutación a28365t, donde se ubica la región de detección del cebador antisentido del gen N del protocolo de Tailandia.

- o Se detectó un genoma con una delección 28898-28909, donde se ubica la región de detección del cebador sentido del gen N del protocolo del CDC de China.

- o Se detectó un genoma con la mutación g28936a, donde se ubica la región de detección de la sonda del gen N del protocolo del CDC de China.

- o Se detectaron cuatro genomas con la mutación g29134a, donde se ubica la región de detección del cebador sentido del gen N del protocolo de Japón.

- o Se detectaron cuatro genomas con la mutación c29272t, donde se ubica la región de detección del cebador antisentido del gen N del protocolo de Japón.

- o Se detectó un genoma con la mutación c29280t, donde se ubica la región de detección del cebador antisentido del gen N del protocolo de Japón.

- o Se detectaron tres genomas con la mutación c29218t, donde se ubica la región de detección de la sonda de N2 del protocolo del CDC de Estados Unidos.

- o Se detectó un genoma con la mutación c28311t, donde se ubica la región de detección de la sonda de N1 del protocolo del CDC de Estados Unidos.

- o Se detectaron tres genomas con la mutación g28321t, donde se ubican la región de detección de la sonda de N1 del protocolo del CDC de Estados Unidos, y la región de detección del cebador sentido del gen N del protocolo de Tailandia.

- o Se detectaron 44 genomas con la mutación ggg28881aac, donde se ubica la región de detección del cebador sentido del gen N del protocolo del CDC de China.

- o Se detectaron 42 genomas con la mutación t29148c, donde se ubica la región de detección del cebador sentido del gen N del protocolo de la Universidad de Hong Kong.

Detalle de las sustituciones no sinónimas (cambios aminoacídicos) en las proteínas codificadas en los genomas obtenidos a partir de las muestras de Córdoba.

Para determinar las sustituciones no sinónimas en comparación con la secuencia de referencia hCoV-19/Wuhan/WIV04/2019 se utilizó la herramienta CoVsurver y se corroboraron mediante inspección visual.

Los cambios fueron comparados con los reportados en secuencias de SARS-CoV-2 a nivel mundial utilizando la herramienta COV-GLUE de GISAID (Singer et al. 2020) y la herramienta outbreak.info.

En concordancia con la presencia exclusiva del linaje B.1 y sus derivados en las secuencias virales estudiadas, **todos** los genomas de SARS-CoV-2 poseen la mutación **Spike: D614G**, asociada a mayor transmisibilidad tanto en modelos in vitro como in vivo (Korber et al, 2020; Plante, et al. 2020). Asimismo, **todas** las secuencias poseen la mutación **NSP12 (RdRp)/P323L** -previamente asociada con un potencial poder mutagénico mayor del SARS-CoV-2 por estudios bioinformáticos (Eskier, D. et al, 2020.; Chand, G.B., et. al. 2020).

Los análisis se focalizaron en el gen de la proteína de superficie Spike (S) del SARS-CoV-2 por su importancia biológica en la infección, transmisión, tratamiento e impacto en la vacunación.

Linaje B.1

Las tres secuencias de Saldán y Ciudad de Córdoba identificadas como B.1 poseen cambios descritos previamente en secuencias de SARS-CoV-2 a nivel mundial, incluyendo el cambio en la región Spike_D614G. Dos de ellas (G0071/2020, G0074/2020) comparten además un cambio en NSP3_T181I.

Linaje B.1.2

En la única secuencia de Córdoba identificada como B.1.2 se observaron cambios reportados previamente en secuencias de SARS-CoV-2, incluyendo los cambios en Spike D614G y G1219V. Es la primera vez que se observan esta combinación de mutaciones juntas en la misma secuencia y no ha sido reportada previamente a nivel mundial.

Linaje B.1.1.222

En las tres secuencias de Córdoba identificadas como B.1.1.222 se observaron cambios descritos previamente en secuencias de SARS-CoV-2, incluyendo tres cambios en la región Spike: D614G, D80Y, T732A.

Linaje B.1.1.33

En las cuatro secuencias de San Francisco identificadas como B.1.1.33 se observaron cambios descritos previamente en secuencias de SARS-CoV-2 a nivel mundial; siete de ellos presentes en la región Spike: D614G (4); Q414R (4); T778I (4); G1219C (4); I402V (2); V3G (1); P1162L (1).

En la secuencia que corresponde al inicio del brote se observaron seis cambios en Spike: V3G, I402V, D614G, Q414R, T778I y G1219C. Luego, se observa en secuencias de fecha posterior que algunos cambios se mantuvieron y otros se perdieron. La combinación de los cuatro cambios en Spike D614G, Q414R, T778I y G1219C presentes en las cuatro secuencias, al momento ha sido reportada sólo en Argentina y en los Estados Unidos, a partir del 6 de diciembre de 2020. El resto de las combinaciones de las mutaciones observadas en Spike no ha sido reportado aún a nivel mundial (Tabla 1).

Tabla 1. Combinación de las mutaciones observadas en las secuencias del linaje B.1.1.33.

Muestra	Fecha	V3G	I402V	Q414R	D614G	T778I	P1162L	G1219C	Reportada
G0041/2020	17/11/2020								No
G0060/2020	22/12/2020								No
G0063/2020	29/12/2020								No
G0068/2021	6/1/2021								Desde 6/12/2020 dos veces en Argentina y dos en USA

La mutación Spike_I402V es un cambio poco frecuente, y según un estudio realizado se propone que este cambio podría estar asociado a un leve incremento de la unión al receptor ACE2 (Starr et al. 2021).

Linaje N.3

En las 37 secuencias de este grupo identificadas como pertenecientes al linaje N.3 se observaron cambios reportados previamente en secuencias de SARS-CoV-2 a nivel mundial, incluyendo en la región Spike los cambios: L18F (1), M153T (3), L176F (1), D614G (37), V1228L (4), y otros cambios poco reportados en Spike: E96V (1), G142D (1), A260V (1), R246G (1). En 27 de estas secuencias se observa un cambio en la región NSP3_K589R, que sería marcadora del grupo.

En una secuencia de Villa Dolores del 11 de julio de 2020 (G0077/2020) se observó la co-ocurrencia de las mutaciones Spike_D614G y Spike_R246G; esta combinación ha sido reportada a partir del 3 de enero de 2021 sólo en Irlanda (4) y en Reino Unido (4). En una secuencia de Villa María del 16 de septiembre de 2020 (G0103/2020) se observó la co-ocurrencia de las mutaciones en Spike D614G, V1228L, L18F y A260V. No se hallaron reportes de esta combinación de cambios a nivel mundial. En una secuencia de Arroyito del 16 de septiembre de 2020 (G0110/2020) se observó la co-ocurrencia de las mutaciones Spike_D614G y Spike_E96V; esta combinación ha sido reportada a partir del 19 de enero de 2021 sólo en Turquía (1) y España (1).

Linaje B.1.499

En las secuencias identificadas como B.1.499 provenientes de varias localidades de Córdoba se observan cambios reportados previamente en secuencias de SARS-CoV-2 a nivel mundial, siete de ellos compartidos por las 45 secuencias de este linaje provenientes de la provincia de Córdoba, incluyendo Spike_D614G. En particular para la región Spike se observaron los cambios: T22I (1), D253G (4), V308L (1), T323I (1), Q613H (1), A626S (13), E654Q (24), Q677H (1), P681H (2), L938F (1), T1027I (1), G1219V (1), V1228L (1), y los cambios poco reportados Spike_F65Y (1), D178N (2), G181A (4), E281Q (1), T572S (1), N641Y (1), A783S (1). En particular, el cambio Spike_F65Y que se observa en la secuencia de Rio Cuarto (G0033/2020) es la primera vez que se reporta en el mundo.

Proyecto Argentino Interinstitucional de genómica
de SARS-CoV-2

- En la secuencia de Río Cuarto del 4 de noviembre de 2020 se observó la co-ocurrencia de las mutaciones Spike_D614G y Spike_F65Y; no se hallaron reportes de esta combinación de cambios a nivel mundial.
- En las cuatro secuencias del Grupo I, de las localidades de Corral de Bustos, Marcos Juárez, Villa María y Monte Maíz, entre 22/9/2020 y 18/10/2020, se observó la co-ocurrencia de las mutaciones en Spike: D614G, G181A y D253G. Esta combinación solo se ha reportado a nivel mundial en una secuencia de Argentina de una fecha posterior, del 24/11/2020.
- En una secuencia del Grupo IV, de Córdoba del 4 de enero de 2021 (G0067) se observó la co-ocurrencia de las mutaciones en Spike: D614G, N641Y y A783S; no se hallaron reportes de esta combinación de cambios a nivel mundial.
- En las 24 secuencias (3 del Grupo IV y 21 del Grupo V) se observó la co-ocurrencia de las mutaciones en Spike: D614G, E654Q y en N: T362I, además de un cambio en la región NSP3_M1856I; no se hallaron reportes de esta combinación de cambios a nivel mundial.
- En 13 secuencias (tres del Grupo III y diez del Grupo IV) se observó la co-ocurrencia de las mutaciones en Spike D614G y A626S, ya reportadas a nivel mundial. Sin embargo, en las 10 secuencias del Grupo IV, detectadas desde el 22/9/2020, se observó la co-ocurrencia de las mutaciones en Spike: D614G, A626S junto a N_M210I; esta combinación de mutaciones solo se ha reportado en Islandia a partir del 17 octubre de 2020.
- En una secuencia de Río Cuarto del Grupo V (G0088/2020) se observó co-ocurrencia de las mutaciones Spike_E654Q y Spike_T1027I, reportada anteriormente en Argentina en secuencias de la provincia de La Pampa.
- Las dos secuencias del Grupo II (G0040/2020, G0058/2020) poseen un cambio en NS7b_C41S.

Otros cambios de interés son las inserciones o deleciones en regiones que codifican proteínas virales, las cuales podrían afectar su función. En dos de las secuencias del linaje N.3, se observan deleciones. En una secuencia, una deleción de 12 nucleótidos en la región que codifica para la proteína N que resulta en la pérdida de cuatro aminoácidos (N_del209-212) y en otra secuencia una deleción de tres nucleótidos en la región que codifica para la proteína NSP1 que resulta en la pérdida de un aminoácido (NSP1_M85del). En una de las secuencias del linaje B.1.499 también se observa la deleción de tres nucleótidos en la región que codifica para la proteína NSP1 (NSP1_M85del). Esto demuestra que existen cambios comunes en ciertas posiciones que pueden producirse en secuencias de distintos linajes.

En la Figura 9 se muestran los cambios observados en la región Spike para los distintos linajes identificados en Córdoba, incluyendo aquellos poco frecuentes (amarillo) y los que están relacionados con la unión al receptor celular ACE2 o alguna función de impacto en la biología viral (naranja). Entre paréntesis se incluye el número de secuencias en los que se observó cada cambio representado, excepto para aquellos cambios compartidos por todas las secuencias del linaje.

Proyecto Argentino Interinstitucional de genómica
de SARS-CoV-2

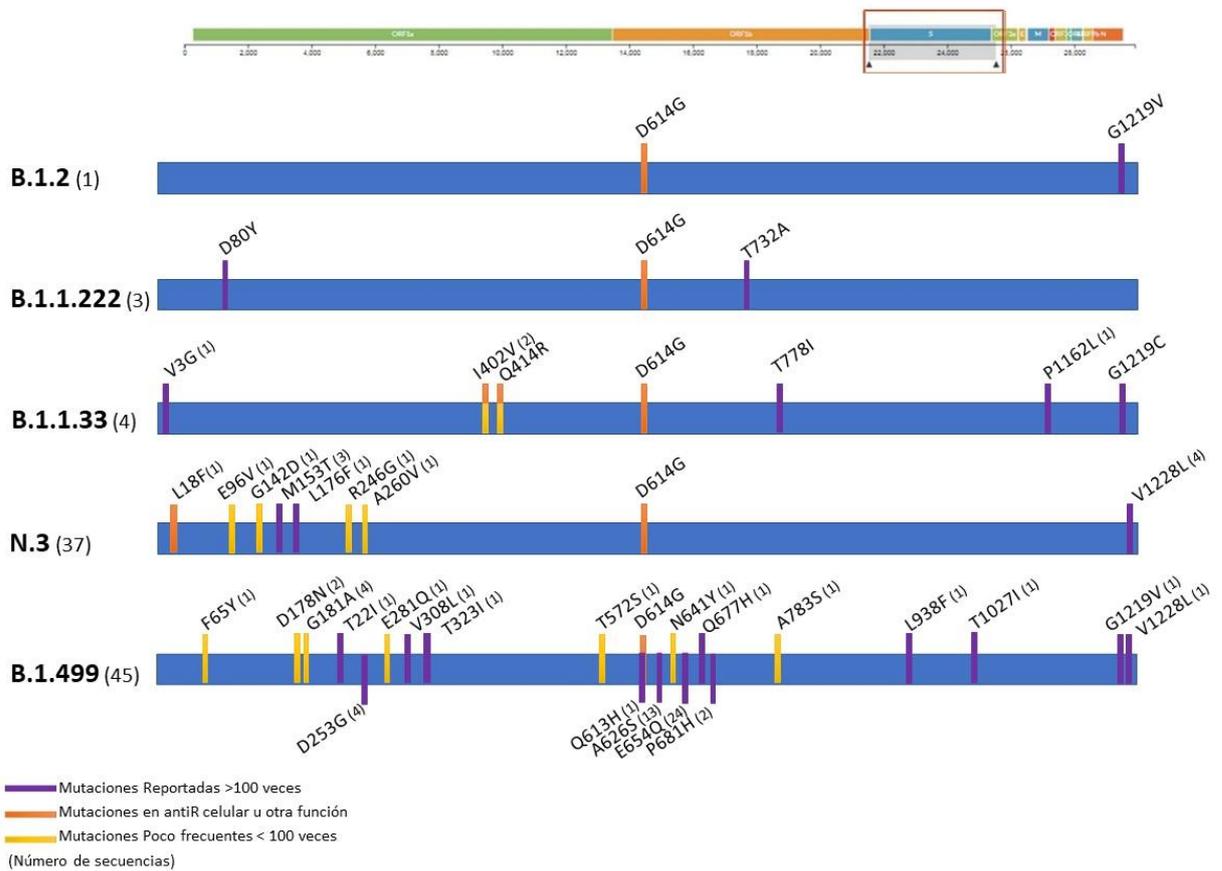


Figura 9. Cambios observados en la región Spike para los distintos linajes identificados en Córdoba.

CONCLUSIONES GENERALES

En este informe se reporta el análisis detallado de **93 secuencias** de SARS-CoV-2 provenientes de muestras clínicas de pacientes de la provincia de Córdoba diagnosticados con la COVID-19 durante el período comprendido entre el 9/4/2020 y el 6/1/2021.

Todas las secuencias corresponden al linaje B.1 o sus derivados, lo que es compatible con la amplia distribución que presenta a nivel mundial y es congruente con lo observado anteriormente para Argentina y para la mayoría de los países de América del Sur.

Cabe destacarse que el análisis comparativo con la primera etapa de estudio de secuencias de esta región del país permitió establecer que se registró un cambio en los linajes en circulación. El linaje B.1.499 fue el mayoritario, mientras que el linaje B.1.1.33 y su derivado N.3, detectado en los casos inicialmente estudiados asociados a viajeros, continuó circulando en alta frecuencia. Esto sugiere que ambos linajes se establecieron en la provincia de Córdoba a partir de múltiples ingresos posiblemente desde distintas locaciones (inferido desde su relación cercana con secuencias de distintas regiones) y algunos de ellos han permanecido circulando en la comunidad. El linaje B.1, responsable de los dos primeros brotes en Córdoba (en Saldán y en un hospital de la ciudad de Córdoba) sólo fue detectado hasta junio 2020, lo que indicaría que su circulación podría haber sido desplazada por otros linajes.

Se detectaron algunos cambios nucleotídicos en zonas de reconocimiento de algunos reactivos de diagnóstico molecular de los que se dispone información; sin embargo, la presencia de estas mutaciones no implica necesariamente una disminución en la sensibilidad de estos ensayos.

Por último, se encontraron algunos cambios aminoacídicos en distintas regiones del genoma viral de algunas secuencias. Se observaron diversas combinaciones de mutaciones en la región que codifica para la proteína Spike, reportadas por primera vez a nivel mundial, entre las que se destaca la mutación Spike_F65Y detectada en una muestra de Río Cuarto. Asimismo, en muestras correspondientes al linaje B.1.1.33 provenientes de San Francisco se detectaron las mutaciones S_Q414R (cuatro muestras) y S_I402V (dos muestras), localizadas en el dominio de unión al receptor ACE2, cambios que podrían estar asociados a un leve incremento de la unión al mismo.

Participantes en este reporte:

Nodo de toma, procesamiento de muestras clínicas y análisis epidemiológicos:

Laboratorio Central, Ministerio de Salud de la provincia de Córdoba: Gabriela Barbás, Gonzalo Castro, Paola Sicilia y Laura López.

Instituto de Virología “Dr. J. M. Vanella” Facultad de Ciencias Médicas - Universidad Nacional de Córdoba: Viviana Ré, María Belén Pisano.

Nodo de secuenciación de Córdoba: IPAVE-INTA-CIAP: Franco Fernández, Nathalie Marquez y Humberto Debat.

Nodo de secuenciación de Rafaela-Santa Fe: EEA Rafaela-INTA-CONICET: Matías Irazoqui, Ariel Amadio.

Nodo evolución: Carolina Torres, Débora Marccone, Laura Mojsiejczuk, María Dolores Blanco Fernández, Andrés Culasso, Humberto Debat, Mariana Viegas.

Fondos:

Proyecto IP COVID-19 N°08, Focem COF 03/11 Covid-19.

ANEXO 1

Análisis filogenéticos

Las secuencias de Argentina fueron analizadas junto con secuencias de referencia de distintos linajes (<https://github.com/hCoV-2019/pangolin>) y secuencias con mejor score de alineamiento por BLAST (entre 10 hits por secuencia incógnita, contra la base de datos de GISAID al 31/01/2021). Los alineamientos se construyeron con el programa MAFFT (<https://mafft.cbrc.jp/alignment/server/>) con parámetros por *default*. El modelo evolutivo apropiado se seleccionó con ModelFinder (Kalyaanamoorthy et al., 2017, <https://doi.org/10.1038/nmeth.4285>) de acuerdo con criterio de Información Bayesiano y los análisis filogenéticos se realizaron por Máxima Verosimilitud con el programa IQ-TREE v. 2.1.2 COVID-edition (Minh y col., 2020, <https://doi.org/10.1093/molbev/msaa015>). Se utilizaron los métodos de *SH-like approximate likelihood ratio test* (1000 réplicas) (Guindon et al., 2010, <https://doi.org/10.1093/sysbio/syq010>) y *Ultrafast bootstrap Approximation* (10000 réplicas) (Hoang y col., 2018, <https://doi.org/10.1093/molbev/msx281>) como métodos para evaluar la confiabilidad de los grupos obtenidos. En las ramas de los árboles (Figuras 5-7) se indican los valores de SH-like/UFB para los grupos relevantes.

Análisis de mutaciones que podrían afectar la detección de los ensayos diagnósticos moleculares disponibles en el mercado.

Se analizaron los cebadores y sondas de los siguientes test:

- Protocolo de Charité, Berlín. Genes reconocidos: E, N y RdRp (https://www.who.int/docs/default-source/coronaviruse/protocol-v2-1.pdf?sfvrsn=a9ef618c_2)
- Protocolo de The University of Hong Kong. Genes reconocidos: N y RdRp. (https://www.who.int/docs/default-source/coronaviruse/peiris-protocol-16-1-20.pdf?sfvrsn=af1aac73_4)
- Protocolo de CDC Estados Unidos. Gen reconocido: N (regiones N1 y N2) (<https://www.cdc.gov/coronavirus/2019-ncov/lab/rt-pcr-panel-primer-probes.html>)
- Protocolo de CDC China. Genes reconocidos: ORF1ab y N (http://ivdc.chinacdc.cn/kyjz/202001/t20200121_211337.html)
- Protocolo de Ministry of Public Health, Tailandia. Gen reconocido: N (https://www.who.int/docs/default-source/coronaviruse/conventional-rt-pcr-followed-by-sequencing-for-detection-of-ncov-rirl-nat-inst-health-t.pdf?sfvrsn=42271c6d_4)
- Protocolo de National Institute of Infectious Diseases, Tokio, Japón. Gen reconocido: N (https://www.who.int/docs/default-source/coronaviruse/method-niid-20200123-2.pdf?sfvrsn=fbf75320_7)
- Protocolo de Institut Pasteur, Paris. Genes reconocidos: E y RdRp. (https://www.who.int/docs/default-source/coronaviruse/real-time-rt-pcr-assays-for-the-detection-of-sars-cov-2-institut-pasteur-paris.pdf?sfvrsn=3662fcb6_2)
- Protocolo de KiCqStart de Sigma-Aldrich. Gen reconocido: S (<https://www.sigmaaldrich.com/covid-19/main-es.html>)

REFERENCIAS

- Eskier D, Suner A, Oktay Y, Karakülah G. 2020. Mutations of SARS-CoV-2 nsp14 exhibit strong association with increased genome-wide mutation load. *PeerJ* 8:e10181 <https://doi.org/10.7717/peerj.10181>
- Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, Hengartner N, Giorgi EE, Bhattacharya T, Foley B, Hastie KM, Parker MD, Partridge DG, Evans CM, Freeman TM, de Silva TI; Sheffield COVID-19 Genomics Group, McDanal C, Perez LG, Tang H, Moon-Walker A, Whelan SP, LaBranche CC, Sapphire EO, Montefiori DC. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell*. 2020 Aug 20;182(4):812-827.e19. doi: 10.1016/j.cell.2020.06.043. Epub 2020 Jul 3. PMID: 32697968; PMCID: PMC7332439.
- Plante JA, Liu Y, Liu J, Xia H, Johnson BA, Lokugamage KG, Zhang X, Muruato AE, Zou J, Fontes-Garfias CR, Mirchandani D, Scharton D, Bilello JP, Ku Z, An Z, Kalveram B, Freiberg AN, Menachery VD, Xie X, Plante KS, Weaver SC, Shi PY. Spike mutation D614G alters SARS-CoV-2 fitness. *Nature*. 2020 Oct 26. doi: 10.1038/s41586-020-2895-3. Epub ahead of print. PMID: 33106671.
- Rambaut y col. 2020. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol* 5, 1403–1407. <https://doi.org/10.1038/s41564-020-0770-5>.
- Singer y col. 2020. CoV-GLUE: A Web Application for Tracking SARS-CoV-2 Genomic Variation Preprints 2020, 2020060225 <https://doi.org/10.20944/preprints202006.0225.v1>.
- Starr y col. 2021. Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. *Cell*. 182(5): 1295–1310.e20. DOI: 10.1016/j.cell.2020.08.012.