

Reporte N°3 – Ciudad Autónoma de Buenos Aires – 28 de julio de 2020

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

En esta etapa del trabajo se buscó identificar el establecimiento de clusters de transmisión locales para el caso de la CABA y su distribución con respecto a las distintas regiones del mismo. El periodo en estudio comprende **desde la semana epidemiológica N°17 hasta la semana epidemiológica N°23** donde la circulación viral autóctona ya se encontraba en curso según el Boletín integrado de vigilancia N°491 (Semana Epidemiológica N°12) del Ministerio de Salud de la Nación Argentina.

En cuanto a la situación epidemiológica de la Nación sobre los Decretos Nacionales relativos al aislamiento social, preventivo y obligatorio (ASPO):

- 12 de marzo de 2020. Argentina (DNU 260_2020) declara la Emergencia Sanitaria
- 16 de marzo de 2020. DNU 274_2020 Prohibición ingreso extranjeros
- 20 de marzo de 2020. DNU 297_2020 Aislamiento Social Obligatorio
- 31 de marzo de 2020. DNU 325_2020 Prorroga DNU 297 hasta 12 abril
- 1 de abril de 2020. DNU 331_2020 Prorroga DNU 274 Ingreso extranjeros
- 11 de abril de 2020. DNU 355_2020 Prorroga ASPO hasta 26 abril.
- 11 de abril de 2020. DNU 365_2020 Prorroga DNU 331
- 26 de abril de 2020: DNU 408_2020 Prorroga ASPO hasta 10 de mayo
- 5 de mayo de 2020 se lanzó un nuevo Dispositivo Estratégico de Testeo para Coronavirus en Territorio Argentino (DETeCTAR)
- 10 de mayo de 2020: DNU 459_2020 Prorroga ASPO hasta 24 de mayo
- 24 de mayo de 2020: DNU 493_2020 Prorroga ASPO hasta 7 de junio

Situación epidemiológica de las regiones analizadas.

Con el objetivo de caracterizar epidemiológicamente el periodo analizado en este trabajo se obtuvo información sobre la cantidad de casos confirmados de COVID-19 y las medidas sanitarias aplicadas en la CABA publicadas en los Boletines Epidemiológicos del Ministerio de Salud de la Ciudad de Buenos Aires.

Desde la fecha de notificación del primer caso confirmado de COVID-19 en Argentina se reportaron se reportaron 9867 casos confirmados de COVID-19 hasta el 4 de junio de 2020 inclusive. Durante el periodo en estudio por este reporte (21 de abril hasta 4 de junio) se confirmaron un total de 9194 los casos de COVID-19 en la CABA.

Las medidas sanitarias implementadas en este periodo incluyeron la ampliación en la definición de casos sospechosos y la implementación del dispositivo DETectAR. Dicho dispositivo tuvo como objetivo inicial en CABA, identificar a los contactos estrechos convivientes de las personas recientemente confirmadas con COVID-19 en los barrios vulnerables (Barrio Padre Mugica (ex Villa 31) y Barrio Padre Rodolfo Ricciardelli (ex Villa 1-11-14)). Para luego extenderse a otros barrios de la Ciudad, empezando por Balvanera, Flores, La Boca, Constitución y Barracas.

Estas medidas sanitarias evidenciaron un marcado aumento en los casos confirmados de COVID-19 que se pueden observar gráficamente en la siguiente figura. En [celeste](#) se muestra el periodo de tiempo que comprende este trabajo.

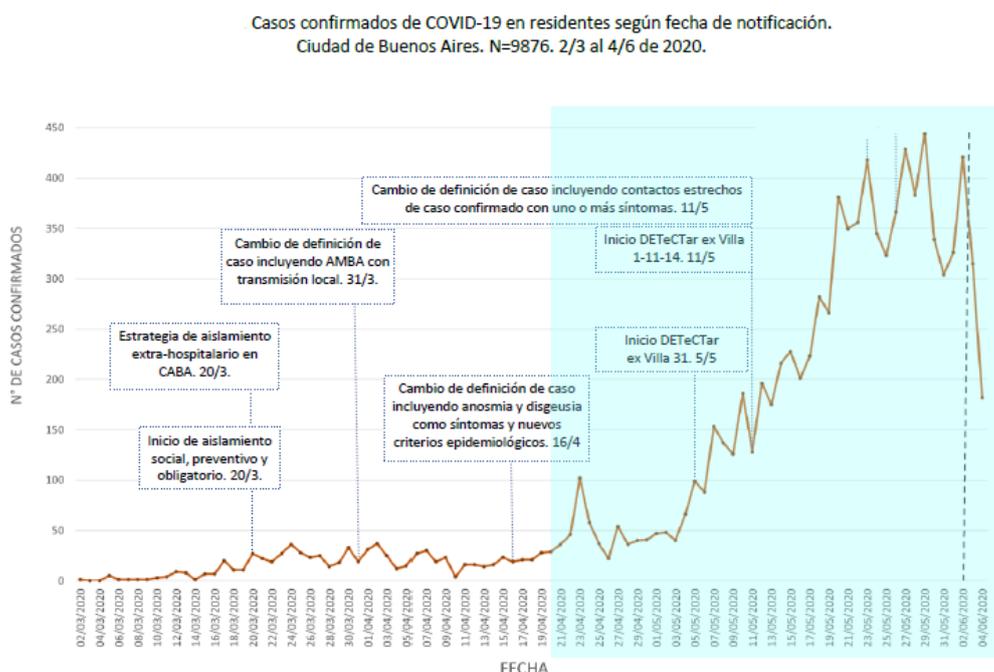


Figura 1. Casos confirmados de COVID-19 en residentes según la fecha de notificación). En [celeste](#) se muestra el periodo de tiempo que comprende este trabajo. Fuente del gráfico base: Boletín Epidemiológico del Ministerio de Salud de la CABA

En el siguiente mapa se pueden ver los barrios de la CABA más afectados al momento del estudio.

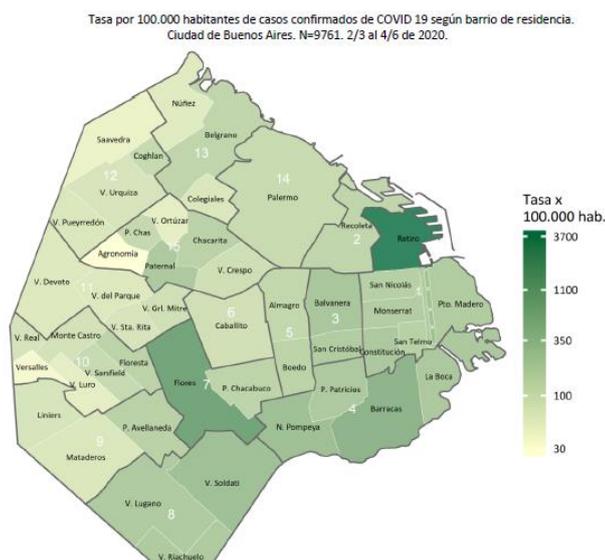


Figura 2. Fuente: Boletín Epidemiológico del Ministerio de Salud de la CABA.

Descripción del muestreo

Los laboratorios que aportaron las muestras y los datos clínico-epidemiológicos de los pacientes fueron:

- Laboratorio de Virología del Hospital de Niños Dr. Ricardo Gutiérrez conformado por: Alicia Mistchenko, Erica Grandis, María Cristina Alvarez López, María Elina Acevedo, Oscar Jacquez, Sofía Alexay, Mariángeles Barreda Frank, María Emilia Villegas, Raquel Barquez, Estela Chascón, Jorgelina Caruso, Mariana Campal, Karina Zacarías, Cristian Díaz, Oscar Luna, Cristian Turchiaro, Julián Cipelli, Guillermo Thomas, Carla Medina, Natalia Labarta.
- Hospital Alemán (CABA) conformado por: Eugenia Ibañez, María Paula Della Latta, Natalia García Allende.
- Unidad de Virología, Centro de Educación Médica en Investigaciones Clínicas "CEMIC" (CABA): Marcela Echavarría

Todas las muestras que se analizaron corresponden a extractos de ARN obtenidos de hisopados nasofaríngeos confirmados para la COVID-19 por la técnica de referencia de RT-PCR en tiempo real y que cumplían con una carga viral medida en valores de Ct que debía estar por debajo de 30.

Dado que desde la semana epidemiológica N°17 hasta la semana epidemiológica N°23 se evidenció el alto impacto de COVID-19 en los barrios vulnerables de la ciudad se decidió en este trabajo focalizar el estudio en los dos barrios en los que se había implementado el dispositivo DETectAR. Además se seleccionaron muestras provenientes de otros barrios de la CABA con el objetivo de obtener datos preliminares sobre la circulación comunitaria del virus allí y su posible relación con las zonas aledañas.

Teniendo en cuenta las zonas seleccionadas para estudiar y para poder analizar posteriormente los resultados obtenidos se dividió a la CABA en 4 zonas:

- CABA NORTE: integrada por los barrios de Nuñez, Belgrano, Saavedra, Coghlan, V. Urquiza, V. Pueyrredón, P. Chas, V. Ortúzar, Chacarita, V. Crespo, Palermo, Recoleta, Retiro, V. Devoto, V. del Parque, V. Gral. Mitre, Caballito, Almagro, Balvanera, San Nicolás, Monserrat, Pto. Madero, V. Sta. Rita.
- CABA SUR: Constitución, San Telmo, La Boca, Barracas, San Cristóbal, Boedo, P. Chacabuco, Flores, Floresta, V. Sarsfield, Monte Castro, V. Real, V. Luro, Versalles, Liniers, Mataderos, P. Avellaneda, N. Pompeya, V. Soldati, V. Lugano, V. Riachuelo.
- BARRIO PADRE MUGICA (ex Villa 31)
- BARRIO PADRE RODOLFO RICCIARDELLI (ex Villa 1-11-14)

En el siguiente mapa se observan las 4 zonas graficadas:

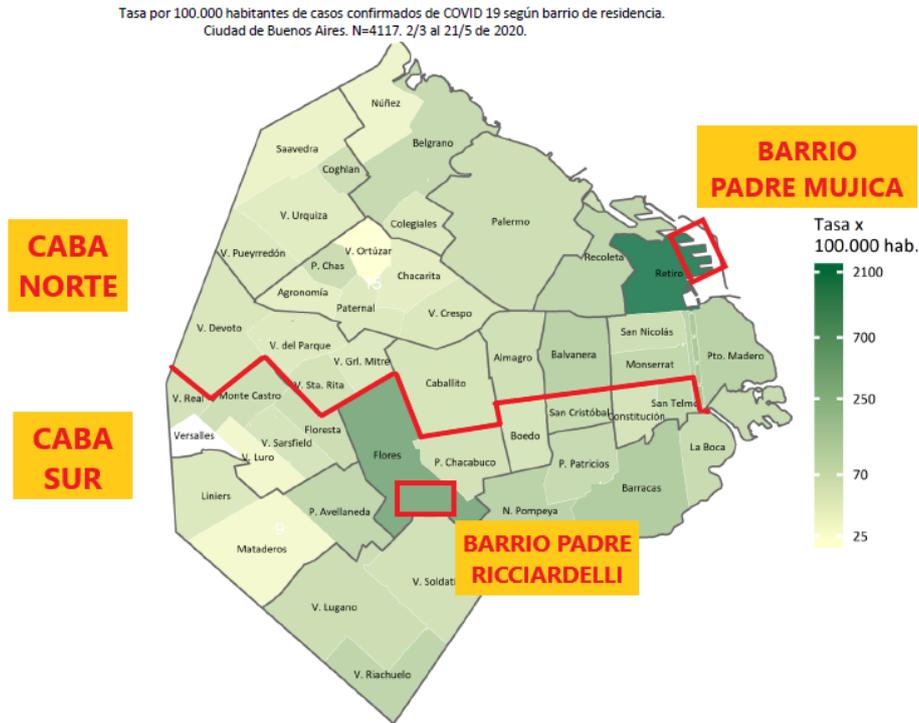


Figura 3. Sectorización de la Ciudad Autónoma de Buenos Aires. Fuente del gráfico base: Boletín Epidemiológico del Ministerio de Salud de la CABA

Muestreo Ciudad Autónoma de Buenos Aires (CABA NORTE y CABA SUR)

Se planificó seleccionar aproximadamente 35 muestras provenientes de la CABA con el objetivo de realizar un análisis preliminar con la proyección de extenderlo en un futuro mediante la realización de un muestreo que sea representativo de todos los barrios de la CABA.

Los criterios que se tuvieron en cuenta para la selección de muestras fueron:

- Seleccionar al menos una muestra de todos barrios de la CABA de los cuales se habían remitido muestras al Laboratorio de Virología del HNRG.
- En cada barrio, seleccionar muestras que estuvieran distribuidas temporalmente.
- En los barrios donde fue posible elegir más de una muestra, priorizar aquellas distanciadas geográficamente.
- Teniendo en cuenta estos criterios se seleccionaron **un total de 37 muestras de la Ciudad Autónoma de Buenos Aires** para la secuenciación del genoma completo de SARS-CoV-2.

En la siguiente tabla se encuentra graficada la cantidad de muestras que se seleccionaron en función del barrio y de la fecha notificación.

Muestreo Barrio Padre Mugica (ex Villa 31)

El Laboratorio de Virología del HNRG diagnosticó un total de 1000 muestras positivas para SARS-CoV-2 durante el periodo en estudio, de las cuales 590 muestras se categorizaron como aptas para la secuenciación del genoma completo del virus.

El proceso de selección de las muestras a secuenciar conllevó distintos criterios aplicados en forma escalonada:

-Se planificó, dada la disponibilidad de los insumos y el número de muestras totales de esta región, seleccionar aproximadamente 120 muestras del Barrio Padre Mugica.

-Se dividió el muestreo en dos grupos, antes y después de la implementación del dispositivo DETectAR (5 de mayo de 2020) con el objetivo de que no haya una sub-representación de los casos reportados en el mes de abril.

- Se planificó seleccionar la misma cantidad de muestras en ambos grupos (pre y post implementación del dispositivo DETecAR)

- Dentro de cada uno de esos grupos se organizaron las muestras en función de la ubicación geográfica delimitada por las manzanas dentro del barrio.

- Se seleccionó al menos una muestra por manzana (siempre y cuando se hubieran reportado casos allí) priorizando que las muestras no provengan de una misma residencia.

- En cada manzana en la que fuera posible, se seleccionaron muestras distribuidas temporalmente.

Teniendo en cuenta estos criterios, se seleccionaron **un total de 117 muestras del Barrio Padre Mugica** para la secuenciación del genoma completo de SARS-CoV-2. La distribución de las mismas se encuentra representada en las siguientes imágenes.

En la siguiente tabla se encuentra graficada la cantidad de muestras que se seleccionaron en función de la manzana y de la fecha de toma de muestra.

PROYECTO ARGENTINO INTERINSTITUCIONAL DE GENÓMICA
DE SARS-COV-2

MZN/ DÍA	ABRIL													MAYO																				
	1 8	1 9	2 0	2 1	2 2	2 3	2 4	2 5	2 6	2 7	2 8	2 9	3 0	1	2	3	4	5	6	7	8	9	1 0	1 1	1 2	1 3	1 4	1 5	1 6	1 7	1 8	1 9	2 0	2 1
1												1				1							1											1
1 B																1									1									
2																	1										1							
3																						1										1		
4												1																						
4 B																				1	1													
5																							1										1	
6																											1							
7							1										2						1											
7 B																											1							
8																										1							1	
9																										1				1				
9 B																									1									
10								1						1		1	1				2										1			
11																																		
11 B																																	1	
12																																		1
12 B																																		
14																																		
14 B																																		
16																																		
17																																		
18																																		
21																																		
22																																		
23																																		
24																																		
25																																		
26																																		
27																																		
29																																		
32																																		
34																																		
35																																		
36																																		
99																																		
100																																		
100 B																																		
101																																		
102																																		
103																																		
104																																		
105																																		
106																																		
108																																		
109																																		
111																																		
112																																		
115																																		
116																																		
117																																		
118																																		
s/mzn																																		

Figura 6. Cantidad de muestras seleccionadas por manzana en función de la fecha (Barrio P. Mugica)

Muestreo del Barrio Padre Rodolfo Ricciardelli (ex Villa 1-11-14)

El Laboratorio de Virología del HNRG diagnosticó un total de 177 muestras positivas para SARS-CoV-2 durante el periodo en estudio, de las cuales 93 muestras se categorizaron como aptas para la secuenciación del genoma completo del virus. El proceso de selección de las muestras a secuenciar conllevó distintos criterios aplicados en forma escalonada.

-Se planificó, dada la disponibilidad de los insumos y el número de muestras totales de esta región, seleccionar aproximadamente 35 muestras del Barrio Padre Rodolfo Ricciardelli.

-Se organizaron las muestras en función de la ubicación geográfica delimitada por las manzanas dentro del barrio y las zonas urbanizadas que lo rodean.

-Se seleccionó al menos una muestra por manzana (siempre y cuando se hubieran reportado casos allí) priorizando que las muestras no provengan de una misma residencia.

-En cada manzana en la que fuera posible, se seleccionaron muestras distribuidas temporalmente.

Teniendo en cuenta estos criterios se seleccionaron **un total de 30 muestras del Barrio Padre Rodolfo Ricciardelli** para la secuenciación del genoma completo de SARS-CoV-2. La distribución de las mismas se encuentra representada en las siguientes imágenes.

En la siguiente tabla se encuentra graficada la cantidad de muestras que se seleccionaron en función de la manzana y de la fecha de toma de muestra.

MZN/DÍA	MAYO																	
	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1																	1	
2											1							
3										1								
4																	1	
9b													1					
10																	1	
11											1							
14											1		1					1
18													1					
19																1		
20										1								
21										1								1
22																	1	
23								1										
24		1								1							1	
25	1																	
26									1								1	
28													1					
29											1							
31													1				1	
BA RIVADAVIA													1					
BA ILLIA																		1
SIN MZN													1					

Figura 8. Cantidad de muestras seleccionadas por manzana en función de la fecha (Bº P. Ricciardelli)

En el siguiente mapa se encuentran representadas la cantidad de muestras seleccionadas en cada manzana.

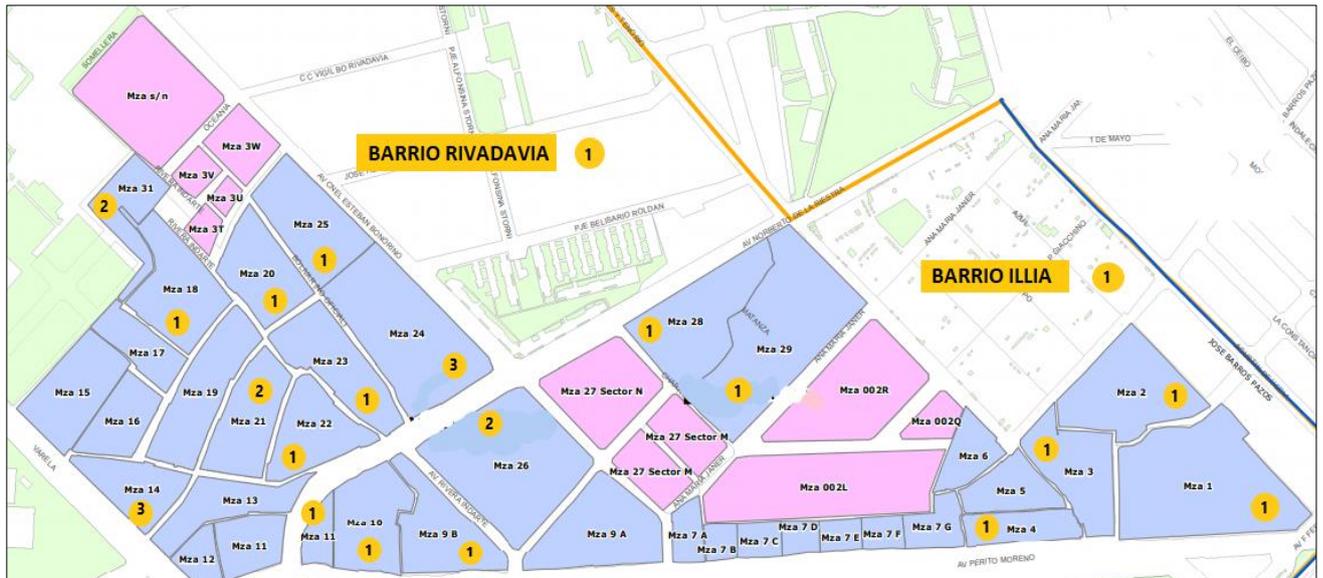


Figura 9. Cantidad de muestras seleccionadas por manzana (Bº Padre Ricciardelli). Fuente del mapa base: Departamento Cartografía. Dirección General de Estadística y Censos (Ministerio de Hacienda, GCBA)

Nodos de secuenciación implicados

Laboratorio de Virología HNRG (CABA), Equipo MinIon, Equipo NextSeq (Unidad de Investigación Traslacional, HNRG). Equipo de trabajo: Laura Valinotto; Mariana Viegas; Mercedes Nabaes; Stephanie Goya; Mónica Natale; Silvina Lusso

Biocódices (CABA), Equipo MiSeq. Equipo de trabajo: Hernán Dopazo, Jeremías Zubrzycki; Juan Manuel Berros

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 para el Consorcio por la Unidad de Genómica del IABIMO-INTA Castelar.

Resultados obtenidos

Secuencias

Se obtuvieron un total de 183 secuencias de SARS-CoV-2 (profundidad promedio ~600X) provenientes de muestras clínicas de pacientes con la COVID-19 distribuidos las zonas descritas en muestreo.

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.

Linajes

Las secuencias del genoma de SARS-CoV-2 obtenidas a partir de individuos infectados de distintos países han sido clasificadas, de acuerdo con cambios nucleotídicos y su agrupamiento filogenético, en dos linajes principales (denominados con letras A y B) y varios linajes internos (A.1-A.9 y sus subgrupos, o B.1-B.22 y sus subgrupos), según la última actualización de la clasificación (Lineages versión 2020-07-20) (Rambaut y col., 2020; <https://doi.org/10.1101/2020.04.17.046086>).

Para la asignación rápida de linajes se utilizaron los programas Pangolin COVID-19 Lineage Assigner y Covidex. Ambas herramientas permiten determinar el linaje, la primera a través una asignación filogenética online usando secuencias de referencia (<https://pangolin.cog-uk.io/>), y la segunda, a través de la identificación de patrones complejos mediante machine learning (<https://sourceforge.net/projects/covidex/>), desarrollada por el Dr. Marco Cacciabue, integrante del Nodo Bioinformático del Consorcio.

Posteriormente el Nodo de Evolución del Consorcio y en particular el grupo de Evolución Viral de la Cátedra de Virología de la FFyB-UBA, realizaron los análisis filogenéticos necesarios para poder confirmar los linajes inicialmente asignados a las secuencias (detalles metodológicos en Anexo). En general, hubo correlación entre los resultados de los dos programas de asignación de linajes rápidos y la obtenida por métodos filogenéticos.

Distribución de linajes encontrados en la Ciudad Autónoma de Buenos Aires

Mediante análisis filogenético se determinó que la mayoría de las secuencias de la CABA pertenecieron al linaje B.1 (o sus (sub)linajes derivados), que ya fue descrito para las secuencias de Argentina reportadas previamente y que posee amplia distribución mundial, y en sólo dos casos, se observó la presencia de linaje B.2 (o su (sub)linaje derivado, B.2.1), asociado inicialmente con países asiáticos pero luego con amplia distribución en Europa, América del Norte y Australia (El árbol filogenético se describe en el Anexo y se muestra en la figura suplementaria 1).

Las secuencias de SARS-CoV-2 de las muestras de CABA (n= 183) correspondieron al linaje B.1 (1,1 %) y sus sublinajes B.1.1 (0,5 %), B.1.1.1 (0,5 %), B.1.1.33 (19,1 %), B.1.3 (76,0 %) y B.1.5 (1,6 %), y al linaje B.2 (0,5 %) y su sublinaje B.2.1 (0,5 %).

Al analizar por separado los linajes encontrados en los barrios vulnerables respecto del resto de CABA, se obtiene:

- **CABA NORTE** (n= 24): B.1.3 (63 %), B.1.1.33 (21 %), B.1 (8 %), B.2 (4 %), B.2.1 (4 %).
- **CABA SUR** (n= 12): B.1.3 (50 %), B.1.5 (25 %), B.1.1.33 (25 %).
- **BARRIO PADRE MUGICA (EX VILLA 31)** (n=117): B.1.3 (100 %).
- **Barrio Padre Rodolfo Ricciardelli (ex Villa 1-11-14)** (n=30): B.1.1.33 (90,0 %), B.1.3 (4 %), B.1.1 (3,3 %), B.1.1.1 (3,3 %).

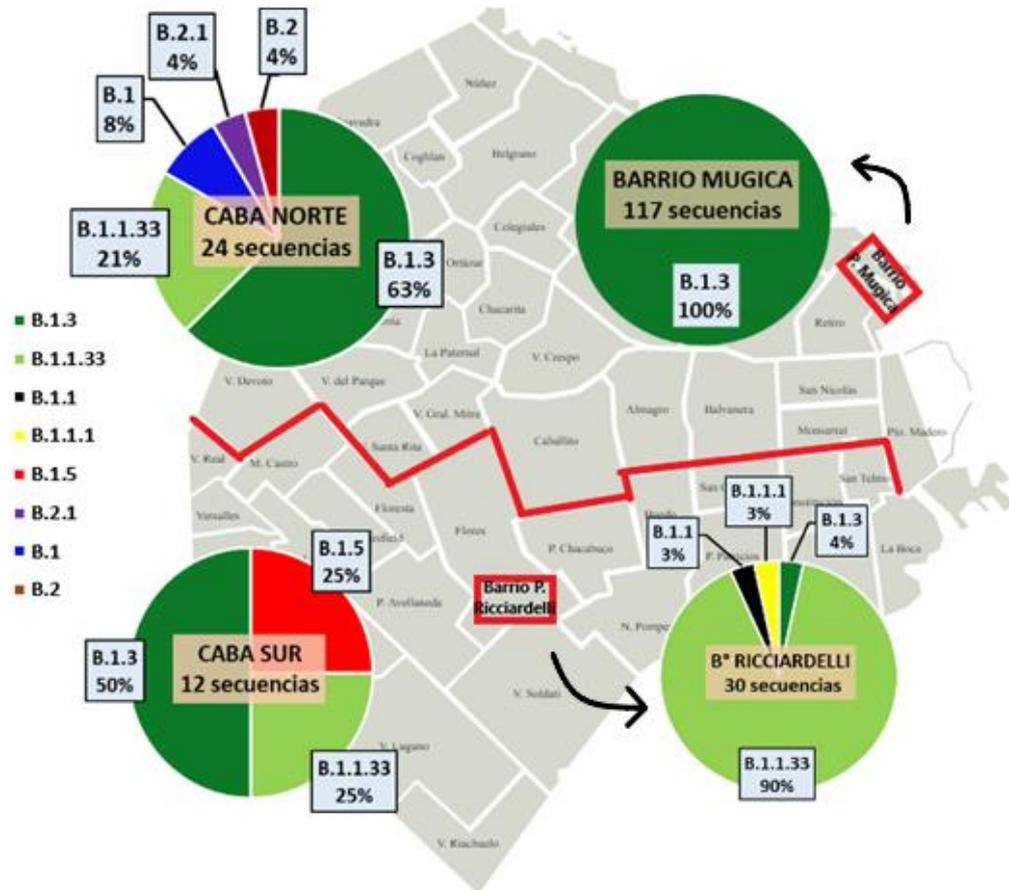


Figura 10. Distribución de los linajes en la CABA.

En los siguientes gráficos se representan la cantidad de secuencias por linaje acumuladas en función del tiempo.

Dada la mayor cantidad de secuencias pertenecientes a los Barrios Mugica y Ricciardelli comparado con el resto de la CABA, se adiciona un gráfico de CABA sin B° MUGICA/B° RICCIARDELLI con el objetivo de que la diversidad en el resto de la CABA se pueda apreciar gráficamente.

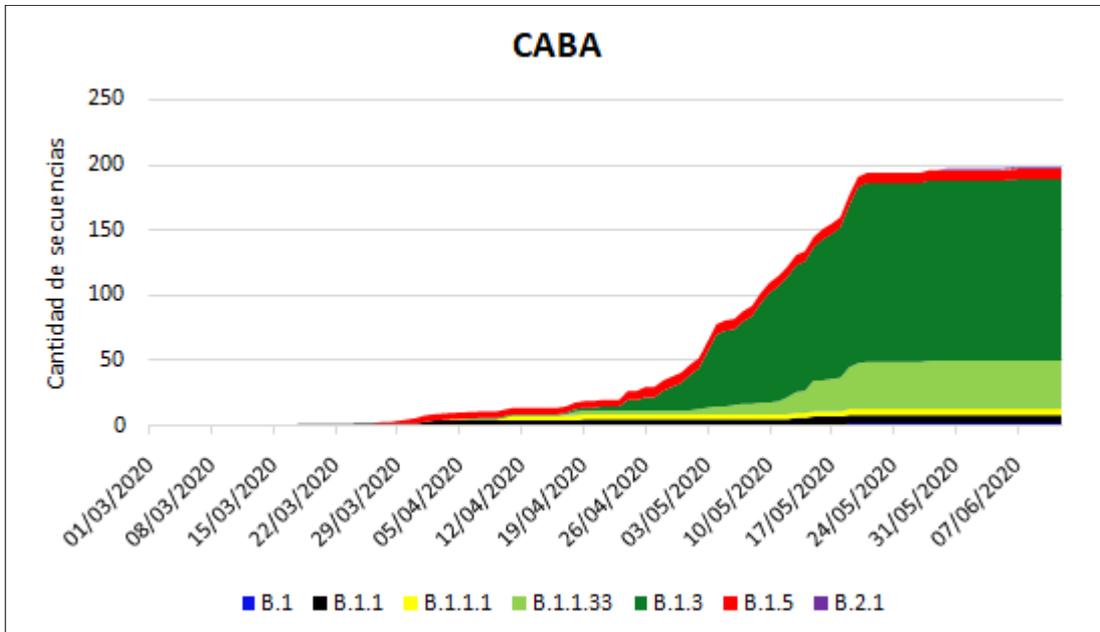


Figura 11. Cantidad de secuencias por linaje acumuladas en función del tiempo (CABA)

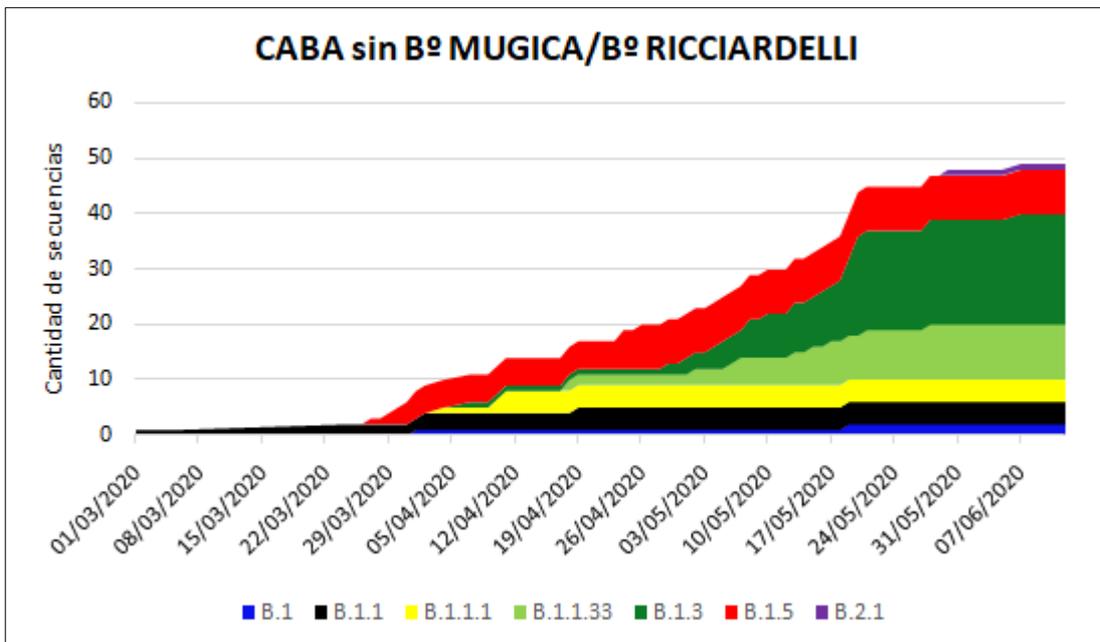


Figura 12. Cantidad de secuencias por linaje acumuladas en función del tiempo (CABA sin BºMugica/BºRicciardelli)

Análisis filogenéticos

Luego de realizar los análisis filogenéticos para cada linaje se procedió a analizar las relaciones filogenéticas dentro de cada linaje de las distintas secuencias de la CABA y su relación con el GBA.

Linaje B.1

Sólo dos secuencias de CABA se asociaron con este linaje. Una secuencia se ubicó entre secuencias de distintos países, mientras que la otra forma un cluster (de moderado soporte) con secuencias de la zona Sur del Gran Buenos Aires (10 secuencias), pero con evidencias de diversificación interna, lo que sugiere que se trata de un cluster de transmisión local (Figura 13.a).

Linaje B.1.1

Sólo una secuencia de CABA pertenece a este linaje y correspondió a una muestra proveniente del Barrio Padre Rodolfo Ricciardelli (ex Villa 1-11-14), sin mostrar agrupamientos de alto soporte significativa asociada a una secuencia de la CABA previamente reportada (Figura 14.c). Cabe destacarse que esta secuencia evidencia una introducción independiente al Barrio Padre Ricciardelli (ex Villa 1-11-14) que la originada por el linaje B.1.1.1 y el B.1.1.33 (ver abajo).

El sistema de nomenclatura utilizado en este reporte indica una subclasificación de este linaje en sublinajes como el B.1.1.1 y B.1.1.33, que se describen a continuación.

Linaje B.1.1.1

Sólo una secuencia de CABA pertenece a este linaje y correspondió a una muestra proveniente del Barrio Padre Ricciardelli (ex Villa 1-11-14) que se asoció con otra secuencia de CABA reportada previamente (Figura 15.b). Cabe destacarse que esta secuencia evidencia una introducción independiente al Barrio Padre Ricciardelli (ex Villa 1-11-14) que la originada por el linaje B.1.1 y el B.1.1.33 (ver abajo).

Linaje B.1.1.33

Las secuencias de CABA de este linaje corresponden al Barrio Padre Ricciardelli (ex Villa 1-11-14) (77,1 % de las B.1.1.33 detectadas) y al resto de CABA (22,9 % de las secuencias B.1.1.33 detectadas).

Estas secuencias se distribuyeron en un gran grupo monofilético con diversificación interna (de moderado soporte), que incluyó todas las secuencias del Barrio Padre Ricciardelli (ex Villa 1-11-14), secuencias de CABA (barrios del Norte y del Sur de la Ciudad), y tres secuencias de la provincia de Buenos Aires (dos de Lomas de Zamora y una de La Plata). Este grupo podría representar una única introducción a CABA, aunque mostraría evidencias de circulación no circunscripta a la ciudad (Figura 14.a).

Linaje B.1.3

Este linaje es el más representado en las secuencias de CABA, constituye el linaje de todas las secuencias del Barrio Padre Mugica (ex Villa 31) y es el mayoritario en las secuencias del resto de CABA (58,3 %), especialmente presente en los barrios del Norte de la Ciudad (14/20 secuencias).

Las secuencias de CABA de este linaje forman un gran grupo (de moderado soporte) con secuencias de la provincia de Buenos Aires, a excepción de una secuencia proveniente de Tierra del Fuego y una de los Estados Unidos. Dentro de este grupo, se observaron algunos clusters de secuencias de CABA, de PBA o de CABA y PBA conjunto, indicativo de que las muestras de esos grupos pertenecen a la misma cadena de transmisión (Figura 16.a).

Si bien algunas secuencias del Barrio Padre Mugica (ex Villa 31) agruparon formando pequeños clusters, en la mayoría de los casos, las secuencias se interrelacionaron entre sí sin un patrón específico de manzana o barrio interno, y se asociaron con secuencias del resto de CABA y de PBA. Estos resultados sugieren que no habría evidencias de circulación viral restringida entre sectores del Barrio 31.

Finalmente, se observó una secuencia de la CABA que no presentó asociación con ninguna secuencia en forma significativa y se asoció con secuencias de distintos países, especialmente los Estados Unidos (Figura 16.c).

Linaje B.1.5

Las tres secuencias de CABA de este linaje corresponden a muestras de la zona Sur de CABA y se distribuyeron en un grupo monofilético con alto soporte formado además por dos secuencias de PBA (Morón y Esteban Echeverría). Este grupo es indicativo de un cluster de circulación entre PBA-CABA (Figura 17).

Linajes B.2 y B.2.1

Una muestra de CABA perteneció al linaje B.2 y otra, a su sublinaje, el B.2.1. La secuencia del linaje B.2 se asocia con secuencias de los Estados Unidos, Europa, Israel y Australia, mientras que la secuencia del linaje B.2.1 se asoció con una de Inglaterra, y éstas con otras secuencias de Inglaterra, Estados Unidos y Francia, principalmente (Figura 18).

Estas dos secuencias representan entonces dos introducciones independientes del virus a nuestro país.

Análisis de estructuración geográfica

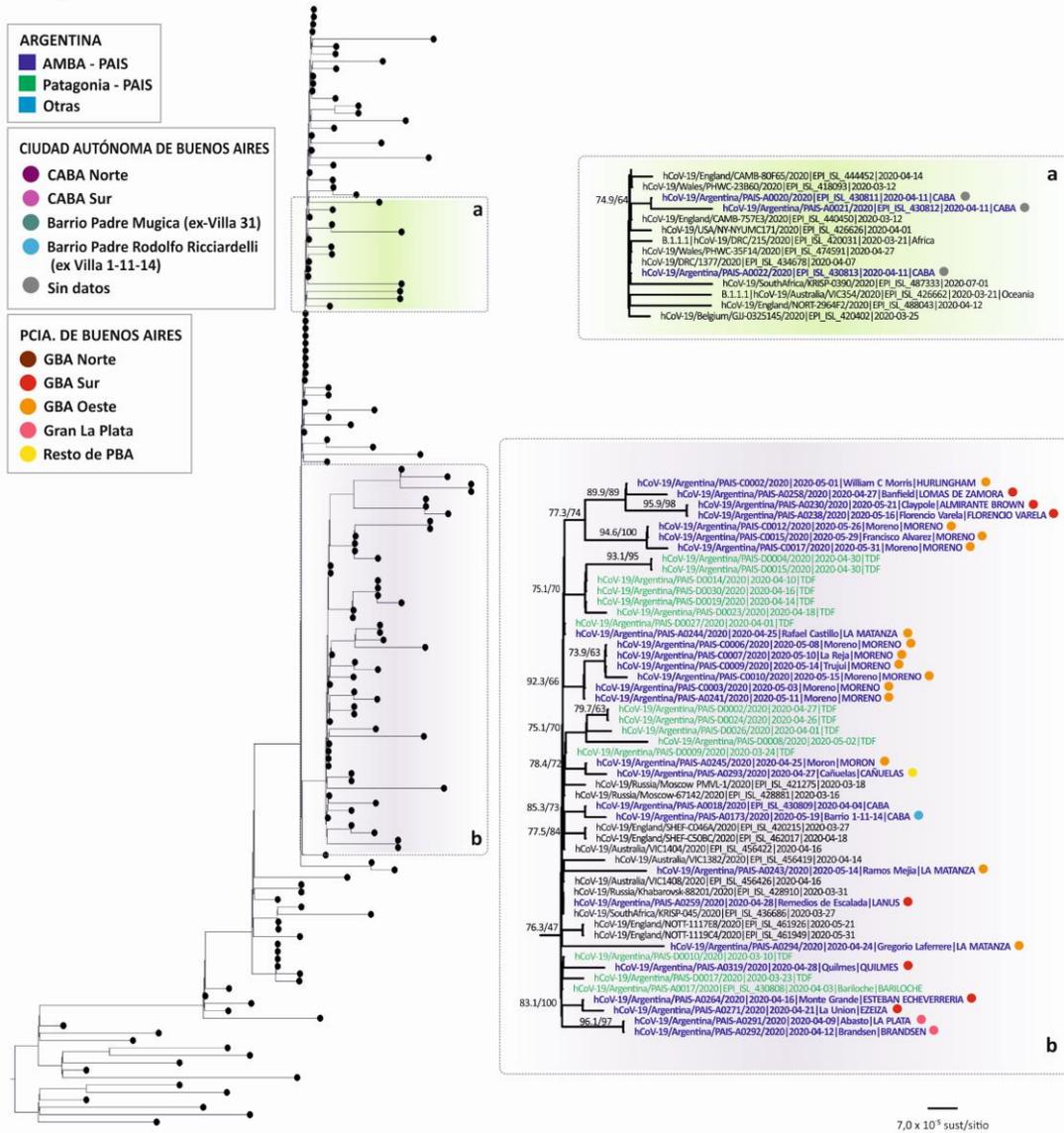
Así como los linajes no se distribuyeron de igual forma entre las regiones geográficas analizadas, los clusters dentro de cada linaje presentaron también un patrón de regionalización. Para evaluar formalmente la probable estructuración geográfica observada, se llevaron a cabo pruebas estadísticas específicas aplicadas a las filogenias moleculares sobre algunos linajes (ver Anexo, para detalles metodológicos y resultados por linaje).

Las regiones consideradas fueron: Barrio P. Mugica (ex Villa 31), Barrio Padre Ricciardelli (ex Villa 1-11-14), CABA Norte, CABA Sur. Mediante estas pruebas se determinó que las secuencias provenientes del Barrio P. Mugica, Barrio Padre Ricciardelli y CABA Sur mostraron estructuración geográfica significativa, es decir, que las secuencias de una misma región no se relacionaron con secuencias en forma aleatoria en los árboles filogenéticos sino que presentaron un **patrón de regionalización, lo que es consistente con transmisión viral en áreas circunscritas, especialmente notable para el Barrio P. Mugica y el Barrio Padre Ricciardelli.**

Por lo tanto, si bien un mismo linaje viral puede encontrarse simultáneamente en distintas regiones geográficas, el patrón general observado es la introducción viral y diversificación zonal restringida, más que la circulación irrestricta de los linajes entre las distintas zonas.

Figura 15. Árbol filogenético del linaje B.1.1.1 de SARS-CoV-2

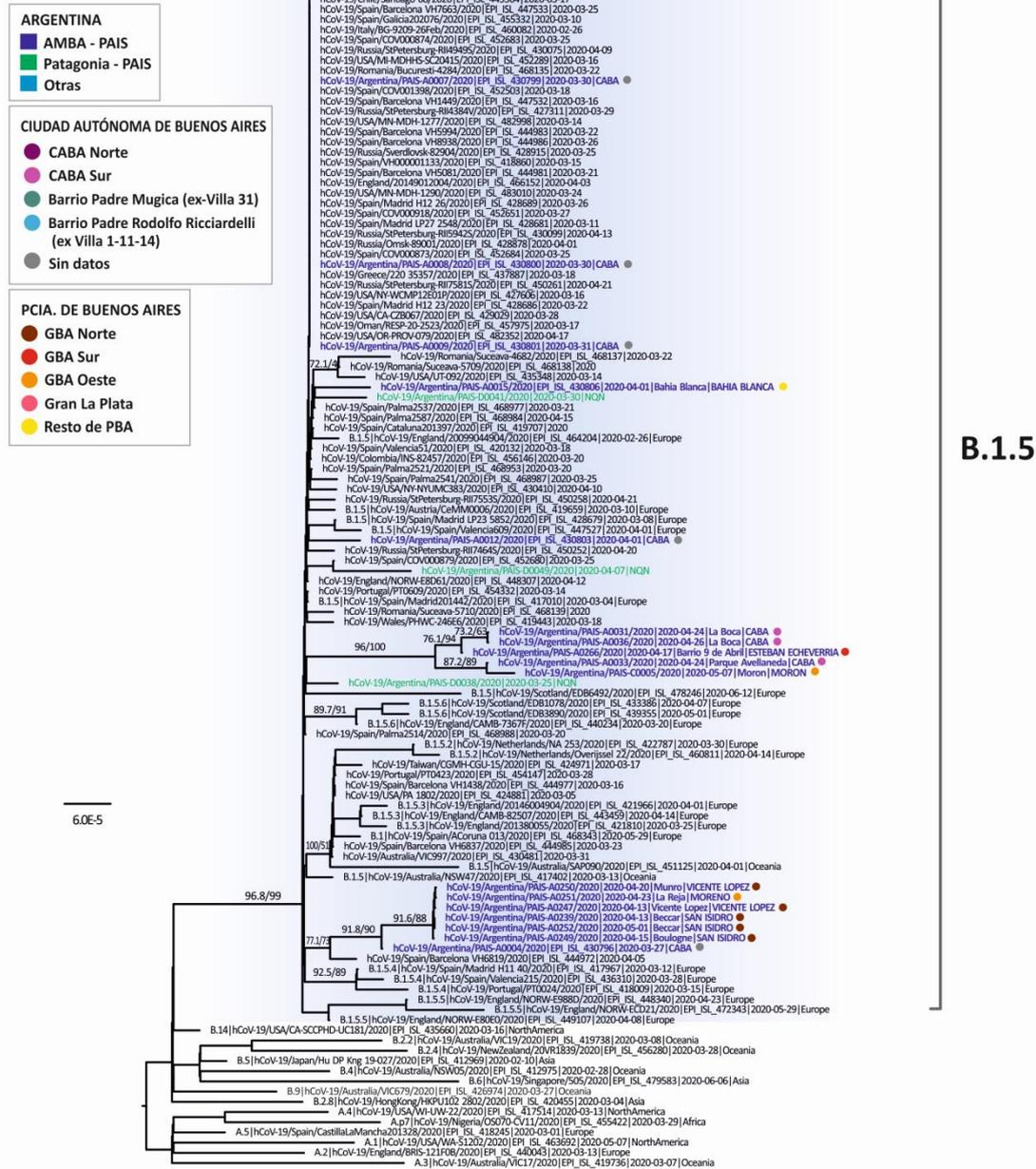
Linaje B.1.1.1



PROYECTO ARGENTINO INTERINSTITUCIONAL DE GENÓMICA DE SARS-CoV-2

Figura 17. Árbol filogenético del linaje B.1.5 de SARS-CoV-2

Linaje B.1.5



PROYECTO ARGENTINO INTERINSTITUCIONAL DE GENÓMICA
DE SARS-CoV-2

Figura 18. Árbol filogenético del linaje B.2 de SARS-CoV-2

Linaje B.2



Análisis de mutaciones en los genomas virales de SARS-CoV-2 reportadas en este informe que podrían afectar la detección por 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. Este análisis se pudo realizar sólo para aquellos métodos que tienen la información disponible (análisis realizado por la Lic. Stephanie Goya, parte del Nodo Central del Consorcio).

Se analizaron los cebadores y sondas de los siguientes test:

- Protocolo de Charité, Berlín. Genes reconocidos: E 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-rihl-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>)

Se reconocieron las secuencias nucleotídicas de los cebadores y sondas utilizando el algoritmo BLASTn a través de la aplicación de un script desarrollado en general por el nodo bioinformático del Consorcio y en particular la herramienta desarrollada por el Ing. Ezequiel Sosa. Se registraron las mutaciones encontradas en los genomas argentinos reportados en este informe con respecto a la cepa de referencia hCoV-19/Wuhan/WIV04/2019:

- Mutaciones en los cebadores y sondas del protocolo de Charité: En 1 genoma se detectó la mutación a15436c localizada en la región de detección del genoma directo del gen RdRp.
- Mutaciones en los cebadores y sondas del protocolo de The University of Hong Kong: En 52 genomas se detectó la mutación t29148c localizada en la región de detección del cebador directo del gen N. En 4 genomas se detectó la mutación c29247t localizada en la región de detección del cebador reverso del gen N. En 1 genoma se detectó la mutación g29254t localizada en la región de detección del cebador reverso del gen N.

**PROYECTO ARGENTINO INTERINSTITUCIONAL DE GENÓMICA
DE SARS-CoV-2**

- Mutaciones en los cebadores y sondas del protocolo CDC China: En 82 genomas se detectó la mutación ggg28881-28883acc localizada en la región de detección del cebador directo gen N. En 1 genoma se detectó la mutación g28883c localizada en la región de detección del cebador directo gen N. En 1 genoma se detectó la mutación g13348a localizada en la región de detección del cebador directo del gen ORF1ab.
- Mutaciones en los cebadores y sondas del protocolo de National Institute of Infectious Diseases, Tokio: En 9 genomas se detectó la mutación g29134a localizada en la región de detección del cebador directo del gen N. En 1 genoma se detectó la mutación c29272t localizada en la región de detección del cebador reverso del gen N.
- Mutaciones en los cebadores y sondas del protocolo de Institut Pasteur: En 4 genomas se detectó la mutación c14120t localizada en la región de detección de la sonda del gen RdRp.
- Mutaciones en los cebadores y sondas del protocolo de KiCqStart: En 1 genoma se detectó la mutación g23653t localizada en la región de detección del cebador reverso del gen S.

Cabe aclarar que la presencia de estas mutaciones no implica necesariamente una baja en la sensibilidad de dichos métodos.

Detalle de las sustituciones no sinónimas (cambios aminoacídicos) en las proteínas codificadas en los genomas obtenidos a partir de las secuencias de la CABA

Para determinar las sustituciones no sinónimas en comparación con la cepa de referencia hCoV-19/Wuhan/WIV04/2019 se utilizó un script desarrollado en conjunto por el Nodo Bioinformático y en particular la herramienta desarrollada por José Matías Irazoqui. Excepto por las dos secuencias pertenecientes al linaje B.2, todos los genomas de SARS-CoV-2 poseen el cambio Spike/D614G, presente en todas las secuencias de linaje B.1 a nivel mundial -previamente asociada a una mayor infectividad in vitro- (Korber et al., *BioRxiv* April 30, 2020. doi: <https://doi.org/10.1101/2020.05.20.104885>) y NSP12 (RdRp)/P323L -con un potencial poder mutagénico mayor del SARS-CoV-2 por estudios bioinformáticos- (Eskier D et al, May 20, 2020. doi: <https://doi.org/10.1101/2020.05.20.104885>). **Sin embargo, no se ha demostrado aún su relevancia in vivo.** Estas mutaciones son características del linaje B y circulan con una muy alta prevalencia a nivel mundial, por lo que su presencia en Argentina confirma la persistencia de las mutaciones en el linaje en su introducción al país.

La proteína de superficie Spike (S) del SARS-CoV-2 es considerada el blanco principal en las inmunoterapias para el tratamiento de la COVID-19, y para el diseño de vacunas (Salvatori, G. et al, *J TranslMed* 18, 222 (2020). <https://doi.org/10.1186/s12967-020-02392-y>). La presencia de mutaciones que afecten su estructura, conformación o glicosilación son de especial interés, por su rol en el plegamiento de las proteínas y la evasión inmune (Watanabe, Y. et al, *NatCommun* 11, 2688 (2020). <https://doi.org/10.1038/s41467-020-16567-0>).

Al analizar las secuencias de AMBA, se observó pérdida de un sitio de N-glicosilación (282) en una de las secuencias obtenidas del Barrio P. Mugica, tratándose del primer reporte de esta mutación a nivel mundial. Si bien esto significa un hallazgo de importancia, será de interés determinar si la mutación fue capaz de ser transmitida en la población.

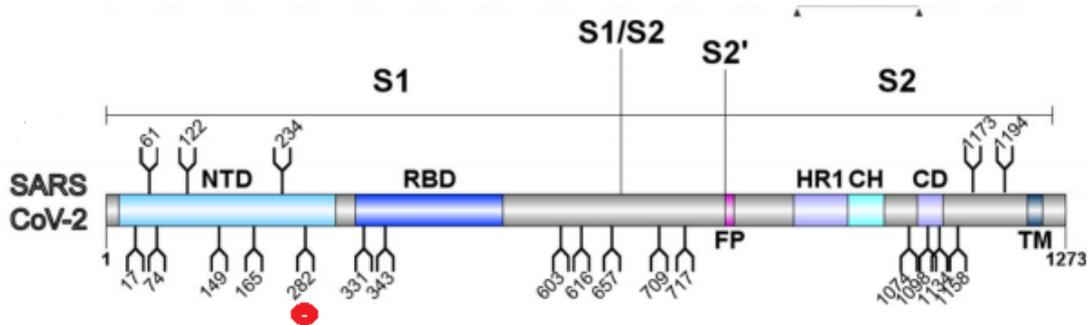


Figura 19: Sitios de N-glicosilación de la proteína S. El círculo rojo marca el sitio donde se observó pérdida de un sitio de glicosilación en una secuencia muestreada en Barrio 31.

Otros cambios de interés son las deleciones e inserciones de aminoácidos en diferentes regiones del genoma del SARS-CoV-2 (Figura 20). En las secuencias de SARS-CoV-2 obtenidas de CABA se observaron 3 secuencias con deleciones o inserciones, 2 de ellas muestreadas en Barrio P. Mugica y una en el barrio de Boedo. **Su relevancia biológica queda por confirmarse.**

Cambio	Gen / Posición	Región de la muestra
DELECIÓN	NSP3/ 199-205	BARRIO P. MUGICA
INSERCIÓN	NSP6/ 35	BARRIO P. MUGICA
DELECIÓN	NSP1/ 85-86	BOEDO

Figura 20. Tabla deleciones o inserciones.

Conclusiones generales sobre los linajes presentes en cada región de la CABA

CABA NORTE y SUR:

- Se observó la presencia mayoritaria de los linajes B.1.3 y B.1.1.33 en toda la CABA.
- B.1.3 es el linaje el mayoritario en las secuencias en las regiones de CABA NORTE y CABA SUR. Estas secuencias se asocian con las presentes en el Barrio P. Mugica, evidenciando cadenas de transmisión conectadas.
- B.1.1.33 es el segundo linaje mayoritario en la CABA y en conjunto con las secuencias de este linaje presentes en el Barrio P. Ricciardelli podrían representar una única introducción a la CABA del mismo, aunque mostraría evidencias de circulación no circunscripta a la ciudad ya que dos secuencias de la Provincia de Buenos Aires forman parte de este grupo.
- En los linajes B.1 y B.1.3 se observa que, las secuencias de CABA están agrupadas con secuencias presentes en la Provincia de Buenos Aires lo que evidencia una cadena de transmisión que conecta ambas regiones del AMBA.
- En CABA SUR se encontraron las únicas 3 secuencias pertenecientes al linaje B.1.5 que están relacionadas con alto sustento con secuencias presentes en la Provincia de Buenos Aires.
- En CABA NORTE están presentes dos secuencias pertenecientes a los linajes B.2 y B.2.1, ambos asociados con secuencias internacionales por lo que estas dos secuencias representarían entonces dos introducciones independientes del virus a nuestro país.

BARRIO PADRE MUGICA:

- En el Barrio P. Mugica se encontró que todas las secuencias pertenecen a un único linaje (B.1.3). Como ya se mencionó, este linaje es el mayoritario en el resto de la CABA y presenta asociaciones con secuencias presentes en la Provincia de Buenos Aires. Esto podría representar que en el Barrio P. Mugica hubo una única introducción o pocas introducciones pero todas ellas provenientes de virus estrechamente relacionados.
- Es importante destacar que sólo en pocos casos se encontró pequeños agrupamientos de alto soporte con secuencias del barrio. Esto evidenciaría que la transmisión del virus no estuvo circunscripta en cada sector particular dentro del barrio. Estos resultados representarían una circulación viral con poca o ninguna restricción intra-barrio que se condice con la situación epidemiológica que presentó el Barrio P. Mugica. En este barrio debido a las condiciones socio-económicas y edilicias, muchas personas no tenían los recursos para realizar el ASPO dentro de sus casas por lo que la implementación del aislamiento social apuntó a limitar la salida de las personas fuera del barrio.

BARRIO PADRE R. RICCIARDELLI:

- Se observó la presencia mayoritaria del linaje B.1.1.33 (90% de las secuencias). Las secuencias de este linaje están relacionadas con secuencias presentes en toda la CABA y también con 3 secuencias del sur de la Provincia de Buenos Aires.
- Al igual que en el Barrio P. Mugica, no se encontró una asociación entre las secuencias de acuerdo con el sector geográfico del barrio, por lo que esto representaría una circulación viral poco restringida dentro del barrio, concordante con que la situación epidemiológica descrita para el Barrio P. Mugica.
- Sólo una secuencia pertenece al linaje B.1.1 sin mostrar agrupamientos de alto soporte significativos. Cabe destacarse que esta secuencia evidencia una introducción independiente al Barrio Padre Ricciardelli (ex Villa 1-11-14) a las originadas por los linajes B.1.1.1 y el B.1.1.33.
- Sólo una secuencia pertenece al linaje B.1.1.1 y se asoció con una secuencia de CABA descrita en el Reporte N° 1. Cabe destacarse que esta secuencia evidencia otra introducción independiente al Barrio 1-11-14 que las arriba descritas.

Cabe destacar que las dos secuencias del Barrio Padre Mugica y del Barrio Ricciardelli incluidas en un análisis anterior del Consorcio PAIS agruparon filogenéticamente con las secuencias de esos mismos sitios incluidas en este análisis. Esto indicaría que el linaje establecido como el prevalente y circulante en cada barrio estuvo presente, al menos, desde la detección de los primeros casos de la COVID-19 en cada uno.

Conclusiones generales

En la CABA se observa presencia de al menos 6 linajes distintos de SARS-CoV-2, con predominancia de dos de ellos (B.1.3 y B.1.1.33). El muestreo intensivo en los Barrios P. Mujica y Padre Ricciardelli mostró una gran homogeneidad en las variantes virales, compatible con una alta transmisibilidad en zonas de conglomerado. Sin embargo, las variantes responsables de los brotes en cada uno de los barrios fueron diferentes (B.1.3 en el Barrio P. Mujica y B.1.1.33 en Barrio Padre Ricciardelli).

Algunos de los genomas de SARS-CoV-2 estudiados presentaron cambios que podrían impactar en sus características biológicas o su diagnóstico molecular, incluyendo la capacidad de ser reconocidos por anticuerpos. Sin embargo, por el momento solo revisten como hallazgos de vigilancia epidemiológica molecular y serán analizados en los siguientes períodos para estudiar su posible relevancia en nuestra población. Estos hallazgos y su relevancia merecen ser confirmados y estudiados en más detalle en estudios posteriores.

Respecto de los análisis de estructuración geográfica, si bien un mismo linaje viral puede encontrarse simultáneamente en distintas regiones geográficas, el patrón general observado es la

introducción viral y diversificación zonal restringida, más que la circulación irrestricta de los linajes entre las distintas zonas.

En el próximo análisis de la CABA, podrán contrastarse estos resultados con los que se obtengan al modificarse las implementaciones del ASPO ocurridas durante los meses de junio y julio, donde comenzó a producirse un movimiento de personas menos restringido entre zonas geográficas del AMBA, acompañado a un aumento progresivo del número de casos de la COVID-19.

Finalmente, cabe destacar la potencialidad de los estudios conjuntos filogenéticos y epidemiológicos para el análisis de casos particulares de cadenas de transmisión viral, así como la importancia de mantener los estudios en el tiempo para dar continuidad al análisis del surgimiento, mantenimiento y desaparición de “clusters” y linajes.

Alcances del análisis

Este estudio preliminar será confirmado utilizando herramientas de análisis más exhaustivas, pero también más demandantes de tiempo de procesamiento de datos. Sin embargo, los resultados aquí expuestos tienen la importancia de aportar más evidencia a las hipótesis de trabajo propuestas, en un tiempo corto, valor fundamental cuando la epidemia está aún en desarrollo.

ANEXO

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 (contra la base de datos de GISAID al 21-07-2020). 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](https://doi.org/10.1038/nmeth.4285)) y los análisis filogenéticos se realizaron por Máxima Verosimilitud con el programa IQ-TREE v.2.0.6 (Minh y col., 2020, <https://doi.org/10.1093/molbev/msaa015>). Se utilizaron los métodos de Ultrafast bootstrap Approximation (10000 réplicas) (Hoang y col., 2018, <https://doi.org/10.1093/molbev/msx281>) y SH-like approximate likelihood ratio test (1000 réplicas) (Guindon et al., 2010, <https://doi.org/10.1093/sysbio/syq010>) como métodos para evaluar la confiabilidad.

Análisis de estructuración geográfica

Para evaluar formalmente estructuración geográfica observada, se llevaron a cabo pruebas estadísticas de asociación aplicadas a las filogenias moleculares sobre algunos linajes: B.1.1 (con la inclusión del B.1.1.33), B.1.1.1, B.1.3, B.1.5.

Estas pruebas permiten determinar si hay evidencias de que las secuencias no se han distribuido al azar en la filogenia y en cambio, presentan asociación bajo alguna condición. Para este análisis se utilizó como condición la localización geográfica de origen, discretizada en las siguientes áreas geográficas:

- **CABA Norte** (según descripción en texto principal, sin Barrio Padre Mugica (ex-Villa 31)).
- **CABA Sur** (según descripción en texto principal, sin Barrio Padre Rodolfo Ricciardelli (ex Villa 1-11-14)).
- **Barrio Padre Mugica (ex-Villa 31).**
- **Barrio Padre Rodolfo Ricciardelli (ex Villa 1-11-14).**

Los análisis fueron llevados a cabo sobre los mismos datasets utilizados para las filogenias presentadas pero sin las secuencias externas al linaje en análisis. Para todos los análisis se estimó la distribución posterior de árboles obtenidos por Métodos bayesianos (MrBayes v3.2.7) y se computó una distribución nula (1000 réplicas) con el programa BaTS v.0.1 (Parker et al. 2008, <https://doi.org/10.1016/j.meegid.2007.08.001>). Para cada carácter discreto (región geográfica), se

estimó el estadístico MC (*maximum monophyletic clade size*) que es una medida del tamaño máximo de los clados monofiléticos.

A continuación, se indican los linajes y las regiones geográficas para los que se rechazó la hipótesis de distribución al azar, lo que permite concluir una estructuración geográfica en los linajes:

- **Linaje B.1.1 y B.1.1.33**

Barrio Padre Rodolfo Ricciardelli (ex Villa 1-11-14) ($p < 0.001$)

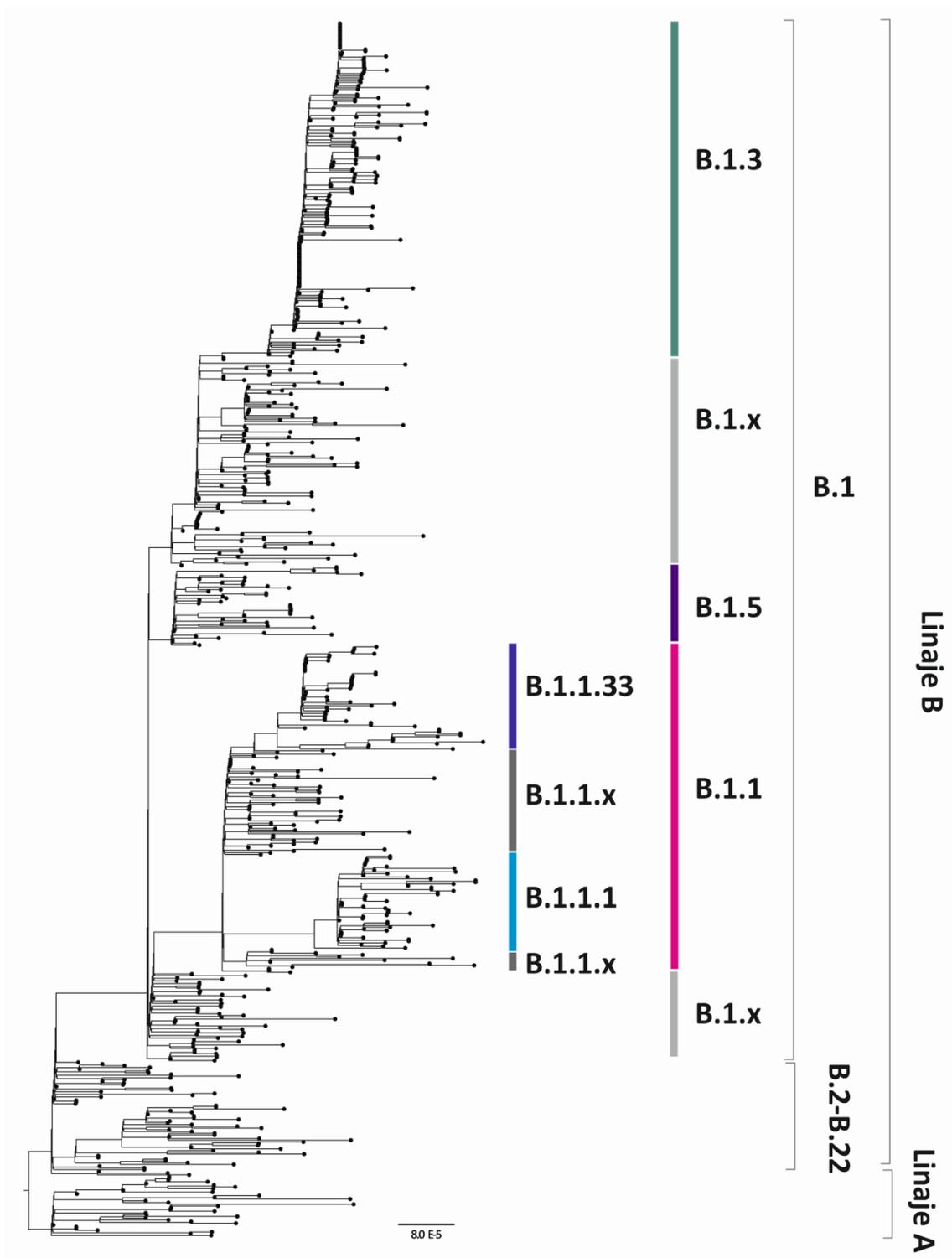
- **Linaje B.1.3**

Barrio Padre Mugica (ex-Villa 31) ($p < 0.001$)

- **Linaje B.1.5**

CABA Sur ($p < 0.002$)

Figura suplementaria 1. Árbol filogenético de los linajes de SARS-CoV-2



Nota

Este informe intenta reflejar la forma de trabajo interdisciplinaria propuesta en el proyecto. Por lo que los datos reportados resultan de la interacción de distintos nodos como los de secuenciación, el bioinformático, de evolución y el invaluable aporte de los epidemiólogos de cada lugar. Intenta demostrar que las diferentes áreas se pueden enriquecer mutuamente cuando se trabaja en equipo. Asimismo, muestra la posibilidad de tener centros de secuenciación genómica descentralizados en distintas regiones del país que están e irán generando habilidades a través de la interacción mutua a medida que este proyecto avance.

En cada análisis particular se informaron los autores participantes, sin embargo, es importante remarcar que si bien hay personas individuales que tienen una autoría especial en varios pasos, cada análisis o secuenciación ha tenido consensos generales discutidos en distintas reuniones por lo que en definitiva todos los integrantes de este Consorcio son autores. A continuación, se listan todos los integrantes del mismo.

Consorcio del Proyecto PAIS (Proyecto Argentino Interinstitucional de genómica de SARS-CoV2)

Nodo Central: grupo de investigación del Laboratorio de Virología Hospital de Niños Ricardo Gutiérrez (HNRG), dirigido por la Dra. Mariana Viegas, Investigador Responsable del proyecto.

Equipo de trabajo: Mariana Viegas; Laura Valinotto; Mercedes Nabaes; Stephanie Goya; Mónica Natale; Silvina Lusso.

Nodos de secuenciación:

1. Laboratorio de Virología HNRG (CABA), Equipo Minlon, Equipo NextSeq (Unidad de Investigación Traslacional, HNRG)
Equipo de trabajo: Laura Valinotto; Mariana Viegas; Mercedes Nabaes; Stephanie Goya; Mónica Natale; Silvina Lusso.
2. Biocódices (CABA), Equipo MiSeq
Equipo de trabajo: Hernán Dopazo, Jeremías Zubrzycki; Juan Manuel Berros
3. IABIMO-INTA Castelar (Provincia de Buenos Aires), Equipo MiSeq, Equipo Minlon
Equipo de trabajo: Andrea Fabiana Puebla, Andrea Verónica Peralta, Marisa Diana Farber, Ana Julia Distéfano, Marianne Graziel Muñoz Hidalgo, Norma Paniego, Mónica Fass, Viviana Cecilia Pedroarias, Sebastián Asurmendi
4. HRU, UNTDF y CADIC-CONICET (Provincia de Tierra del Fuego), Equipo MiSeq
Equipo de trabajo: Santiago Ceballos, Ivan Gramundi, Cristina Nardi, Fernando Gallego,
5. EEA Rafaela del INTA-CONICET (Provincia de Santa Fe), Equipo Minlon
Equipo de trabajo: María Florencia Eberhardt, Cecilia Camussone, Matías Irazoqui, Ariel Amadío.
6. Laboratorio Central de Neuquén (Ministerio de Salud) (Provincia de Neuquén), Equipo MiSeq
Equipo de trabajo: Melina Leonor Mazzeo, Luis Alfredo Pianciola
7. IPAVE-CIAP-INTA (Provincia de Córdoba), Equipo Minlon
Equipo de trabajo: Franco Fernández, Humberto Debat

Nodos bioinformáticos:

1. IQIBICEN - CONICET y Facultad de Ciencias Exactas y Naturales (UBA) - Grupo Marcelo Martí
Equipo de trabajo: Marcelo Martí, Agustín Pardo, Claudio Schuster.
2. IQIBICEN e IC - CONICET y Facultad de Ciencias Exactas y Naturales (UBA) - Grupo Adrián Turjanski. Equipo de trabajo: Adrián Turjanski, Darío Fernández Do Porto, Florencia Castello, Federico Serral, Ezequiel Sosa, Jonathan Zaiat.
3. IABIMO-CICVyA-INTA-CONICET (Provincia de Buenos Aires). Equipo de trabajo: Maximo Rivarola, Sergio Gonzalez, Paula del Carmen Fernández, Sofia Bengoa Luoni, Diego Zavallo, Marco Cacciabue, Laura Lozano Calderón.
4. Biocódices (CABA). Equipo de trabajo: Hernán Dopazo, Jeremias Zubrzycki, Juan Manuel Berros
5. Laboratorio de Virología HNRG (CABA). Equipo de trabajo: Stephanie Goya.
6. EEA Rafaela-INTA-CONICET (Provincia de Santa Fe). Equipo de trabajo: Ariel Amadio, Matías Irazoqui.
7. Cátedra de Virología de Facultad de Farmacia y Bioquímica-UBA (CABA). Equipo de trabajo: Andrés Culasso.
8. Laboratorio Central de Neuquén (Ministerio de Salud) (Provincia de Neuquén). Equipo de trabajo: Carolina Pintos.
9. HRU, UNTDF y CADIC-CONICET (Provincia de Tierra del Fuego). Equipo de trabajo: Santiago Ceballos, Ivan Gramundi.

Nodos de evolución:

1. Laboratorio de Virología, Hospital de Niños Ricardo Gutiérrez (CABA): Mariana Viegas, Stephanie Goya, Mercedes Nabaes
2. Cátedra de Virología de Facultad de Farmacia y Bioquímica-UBA (CABA): Carolina Torres, Andrés Culasso, Laura Mojsiejczuk, María Dolores Blanco Fernandez, Débora Marcone, Viviana Mbayed, Rodolfo Campos, Lucia Vicenta Cavallaro
3. Laboratorio de Biología Celular y Retrovirus, UVEM. Hospital Garrahan (CABA): Paula Aulicino
4. INTA-IABIMO-Evolución (CABA): Guido König, Laura Lozano Calderón
5. FCEN-UBA / IQIBICEN CONICET (CABA): Darío Do Porto, Agustín Pardo, Claudio Schuster
6. HRU, UNTDF y CADIC-CONICET (Provincia de Tierra del Fuego): Santiago Ceballos, Iván Gramundi
7. Laboratorio Central de Neuquén (Ministerio de Salud) (Provincia de Neuquén): Carolina Pintos
8. IPAVE-CIAP-INTA (Provincia de Córdoba): Humberto Debat