

Reporte nº1 - 05 de mayo de 2020

Análisis de secuencias de SARS-CoV-2 obtenidas de pacientes con COVID-19 de Argentina

- Se obtuvieron 26 secuencias de genoma completo de SARS-CoV-2 de Argentina. Entre estas secuencias se encuentran aquellas obtenidas de individuos con historial de viaje registrado, individuos que tuvieron contacto con individuos con COVID-19 e individuos que adquirieron la infección en la comunidad.
- Con respecto a la secuenciación, se observó la presencia de regiones en donde no se pudo determinar la totalidad del genoma del virus (<4%) en alguna de las muestras. Aunque este efecto probablemente se produjo por el alto Ct de las mismas, se analizará en una próxima reunión posibles soluciones entre las cuales se encuentran: saldar esos *drop out* por Sanger mediante un set de primers que abarquen genoma completo (amplicones de 1000 pb solapados), mejorar el *script* de análisis de los *reads*, aumentar la profundidad de cobertura, entre otras.
- Con respecto a los análisis evolutivos, las secuencias del genoma de SARS-CoV-2 obtenidas a partir de individuos infectados de distintos países han sido clasificadas en dos linajes principales (denominado con letras A y B) y varios linajes internos (A1-A5 y sus subgrupos, o B1-B8 y sus subgrupos), de acuerdo con cambios nucleotídicos y su agrupamiento filogenético (Rambaut y col., 2020). Ambos linajes principales presentan amplia distribución mundial, aunque el linaje B reúne a la mayor parte de las secuencias de circulación actual (Figura 1). [Ver Anexo para comparación con la nomenclatura propuesta en [Nextstrain.org](https://nextstrain.org)].
- Mediante análisis filogenético se determinó que las secuencias obtenidas pertenecieron al linaje B.1, que es uno de los principales causantes de los brotes en Europa y América del Norte. Estas secuencias presentaron de 5 a 14 cambios nucleotídicos con respecto a las cepas de referencia “Wuhan-Hu-1/2019” o “Wuhan/WIV04/2019”, de linaje B, y de 7 a 16 cambios nucleotídicos con respecto a la cepa “Wuhan/WH04/2020”, de linaje A, similar a lo hallado para otras secuencias distribuidas a nivel mundial. Estos cambios representan una identidad en el rango de 99,953-99,983 % con respecto a las cepas de referencia de linaje B y de 99,946-99,977 % con respecto a la cepa de linaje A.
- En particular, las secuencias argentinas se asociaron, al menos, con cinco linajes internos: muchas correspondieron al linaje B.1 (n=10), y en los otros casos se pudieron identificar linajes más específicos, como B.1.1 (n=5), B.1.3 (n=3), B.1.5 (n=7) y B.1.27 (n=1) (Figura 2). Cabe destacar que las tres secuencias previas de Argentina también correspondieron al linaje B.1 (n=3).
- Las secuencias argentinas del linaje B.1 están interrelacionadas con secuencias de distintos países, y en algunos casos, forman grupos que podrían representar *clusters* de circulación local dentro del país (como el grupo formado por las cepas A023, A024, A026). A su vez, las secuencias de los dos casos con nexo epidemiológico conocido formaron un grupo monofilético con alto soporte (A005 y A014).
- Las secuencias argentinas del linaje B.1.1 agruparon principalmente con secuencias de Europa del Norte y Rusia. En este grupo se observan secuencias de individuos con historial de viaje registrado (n=3; A020, A021, A022), un individuo que adquirió la infección en la comunidad (A018) y un caso sin información detallada (personal de Policía de Seguridad Aeroportuaria, A017).

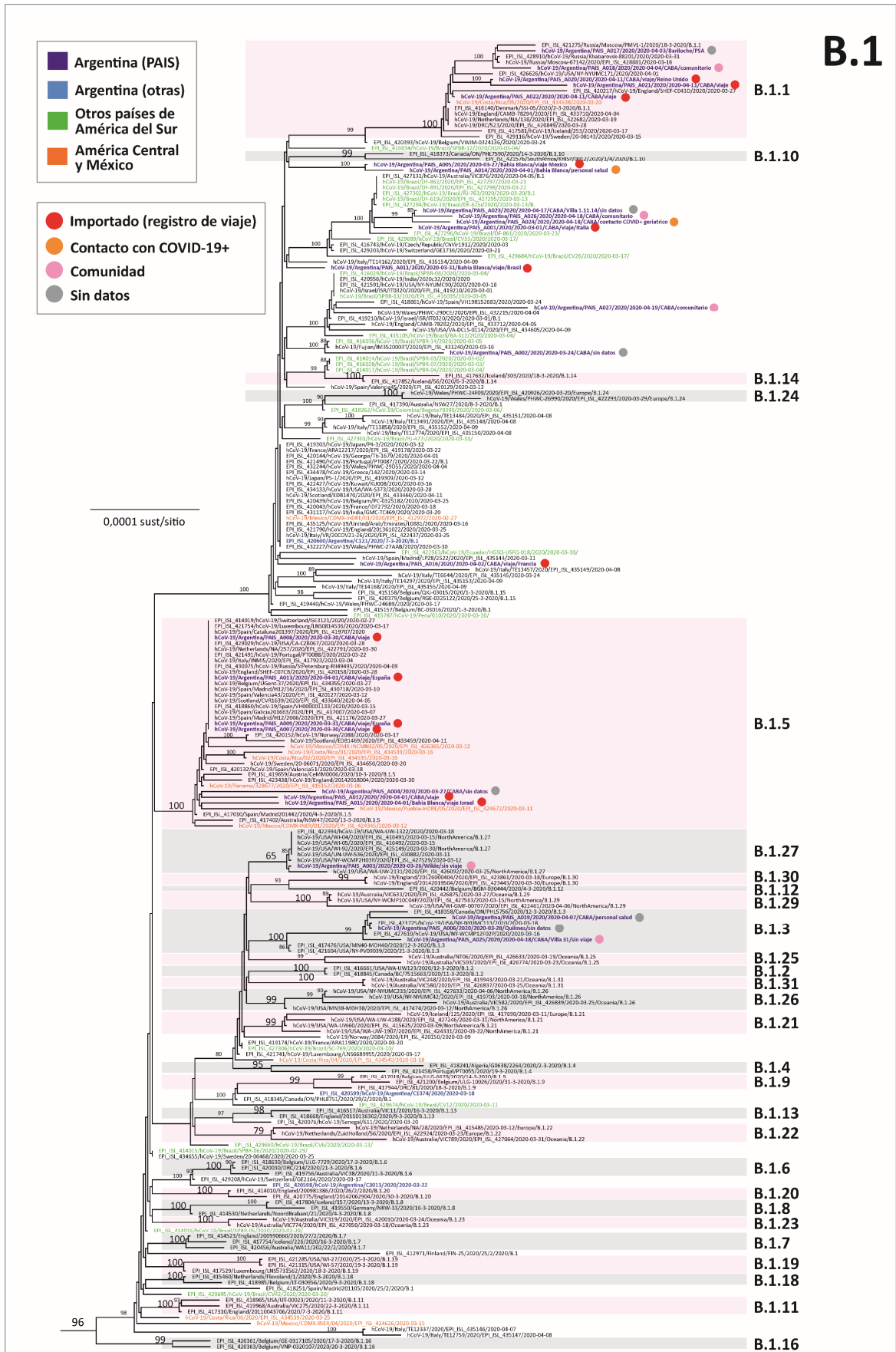
- Las secuencias argentinas del linaje B.1.3 agruparon con virus de América del Norte (Estados Unidos y Canadá). Entre ellas, se encuentra un caso de transmisión comunitaria (residente de Villa 31 en CABA, A025) y otros dos casos de CABA sin información sobre el tipo de contagio (A006 y A019 -personal de salud-).
- Las secuencias argentinas del linaje B.1.5 se asociaron principalmente con virus provenientes de diversos países de Europa. La mayoría de los casos correspondieron a individuos con registro de viaje (n=6; A007, A008, A009, A012, A013, A015), mientras que un caso no presentó información acerca de su posible fuente de contagio (A004).
- La secuencia argentina que forma parte del linaje B.1.27 corresponde a un caso de transmisión comunitaria (A003, residente de Wilde, Pcia. de Buenos Aires). A excepción de la secuencia de Argentina, este linaje está formado exclusivamente por virus circulantes de distintas ciudades de los Estados Unidos.
- Por último, las secuencias de Argentina no presentaron cambios nucleotídicos asociados con posibles fallas en el diagnóstico molecular y del mismo modo, tampoco a nivel global como local, se han encontrado evidencias de que los cambios nucleotídicos o aminoacídicos observados en las secuencias se asocien con cambios en la evolución clínica de la infección.

En conclusión, se observaron múltiples introducciones al país y la posible formación de clusters de transmisión local. El SARS-CoV-2 tanto de secuencias locales como mundiales presenta una velocidad de evolución relativamente lenta (acumulación de cambios genéticos en el tiempo). Sin embargo, dado el alto número de eventos de transmisión a nivel mundial, y la lenta pero progresiva transmisión local, es posible que, en corto tiempo, mediante el aumento en el número de secuencias locales y globales, pueda estudiarse más detalladamente el origen, la diversificación y el comportamiento evolutivo del virus. Cabe destacarse que los resultados de los análisis filogenéticos dependen tanto de las secuencias locales como de las secuencias incorporadas en las bases de datos, por lo que es necesaria una revisión periódica y el reanálisis continuo para aumentar la precisión en las estimaciones e inferencias.

Figura 1. Árbol filogenético de SARS-CoV-2 (el subárbol del linaje B.1 se encuentra ampliado en la Figura 2).



Figura 2. Árbol filogenético de SARS-CoV-2 (subárbol de linaje B.1 de la Figura 1).



ANEXO

Nomenclatura propuesta para la taxonomía de SARS-CoV-2

Rambaut y col. (https://doi.org/10.1101/2020.04.17.046086)	Nextstrain.org (https://nextstrain.org/)
A, A.2, A.3, A.5	B
A, A.1 (A.1.1, A.1.2)	B1
A	B2
A, A4	B4
B.1	A2
B.1, (B.1.1-B.1.31)	A2a
B, B.2	A1a
B.4	A3
B	A6, A7
B.3, B.5, B.6, B.7, B.8	Sin asignar

Asignación de linajes

Para la asignación preliminar de linajes se utilizó la herramienta CoV-GLUE (<http://cov-glue.cvr.gla.ac.uk/#/home>), basada en la nomenclatura propuesta por Rambaut y col. 2020.

Análisis filogenéticos

Las secuencias de Argentina fueron analizadas junto con secuencias de referencia de distintos linajes (<https://github.com/hCoV-2019/pangolin>), secuencias con mejor score de alineamiento por BLAST (contra la base de datos de GISAID al 03-05-2020) y secuencias de América del Sur y América Central (más México). Se intensificó la representación de secuencias de distintos países en los linajes donde se ubicaron las secuencias de Argentina. 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)). Los análisis filogenéticos se realizaron por Máxima Verosimilitud con el programa IQ-TREE v.1.6.12 (Nguyen y col., 2015, <https://doi.org/10.1093/molbev/msu300>) y Ultrafast bootstrap Approximation como método para evaluar la confiabilidad (1000 réplicas) (Hoang y col., 2018, <https://doi.org/10.1093/molbev/msx281>).