

Mutaciones en la Proteína E del SARS-CoV-2 en Argentina.

Autores: C. J. Cimmino¹, A. S. Lerman¹, F. M. Santos¹, R. Campos², O. C. Uez¹, A. Culasso², D. N. Marcone².
1. Instituto Nacional de Epidemiología Dr. J. H. Jara, 2. Facultad de Farmacia y Bioquímica Universidad Nacional de Buenos Aires, 2. CONICET.

Introducción:

Las mutaciones en la proteína **E** (*Envelope*) del SARS-CoV-2 están presentes en las cepas de virus circulantes y pueden influenciar la función viral y la patogénesis (1). Los cambios aminoacídicos en esta proteína tienen implicaciones en el ciclo de vida del virus, la señalización en el hospedador, en la progresión de la enfermedad y el futuro desarrollo de vacunas y agentes terapéuticos (2). Los cambios observados en una región del genoma pueden denominarse mutaciones o polimorfismos dependiendo de la frecuencia en la que se los detecten. Cuando la frecuencia de los cambios es baja (1% o menos) las denominamos “mutaciones”, ya que es probable que haya surgido de manera reciente; cuando las frecuencias son mayores, las denominamos “polimorfismos” (3). La vigilancia Genómica es esencial para el desarrollo de estrategias de control y tratamiento del COVID19. El presente estudio describe la aparición de cambios aminoacídicos respecto de la cepa original detectada en Wuhan en diciembre de 2019, que se encontraron en las variantes de interés (VOI) y variantes de preocupación (VOC) que circularon en la Argentina y las mutaciones que persistieron en las sucesivas variantes de nuestro país.

Objetivo:

Caracterizar las mutaciones presentes en la proteína E en secuencias virales del SARS-CoV-2 reportadas en Argentina desde el inicio de la pandemia.

Materiales y Métodos:

Se descargaron desde el sitio de GISAID (4) las secuencias de genomas completos de SARS-CoV-2 reportadas en Argentina desde marzo de 2020 hasta abril de 2025, filtrando los resultados por calidad de secuencia (solo las confiables) y por porcentaje de cobertura (mayor al 90%). Al mismo tiempo, se descargaron los metadatos asociados en formato tabular. Con esta información, usando el campo de linaje PANGO para cada secuencia, se las clasificó en variantes según el mapeo de variantes publicado por el ECDC (5). Luego, se utilizó un *script* de BASH para generar distintos *datasets* para cada variante de interés. Cada *dataset* fue analizado con el programa *NextClade* (6) para confirmar su linaje / variante. La información de linaje / variante se agrupó por semana epidemiológica (SE) para describir la progresión de la composición de variantes a través del tiempo. Mediante el empleo de un *script* de *Python* (7) los alineamientos en aminoácidos generados por *NextClade* fueron analizados para obtener estadísticas sobre los cambios aminoacídicos en relación con la secuencia de referencia hCoV-19/Wuhan/WIV04/2019 (EPI_ISL_402124). Se compararon los patrones de variación de AA de cada linaje para el gen E entre el período marzo 2020 hasta abril 2025.

Resultados y discusión:

Secuencias: Se descargaron un total de 22.126 secuencias de SARS-CoV-2 que fueron muestreadas entre marzo de 2020 y abril de 2025. En la tabla 1 se puede ver el detalle de la cantidad de secuencias asignadas a cada variante de interés y las fechas de circulación de las mismas.

Patrones de sustitución por variante: Los patrones de sustituciones de AA detectados para cada linaje temprano (B.1.1.33, N.3, N.5 y B.1.499) así como los de las distintas variantes se muestran en la Figura 1. Tanto los linajes tempranos como las variantes **Alfa**, **Delta**, **Gamma**, **Lambda** poseen una mayor diversidad de sustituciones en comparación con los linajes de **Ómicron**. Sin embargo, a pesar de involucrar más posiciones del gen E, las sustituciones se encuentran presentes en menos del 1% de las secuencias, con excepción de E:V58F para el linaje N.3 y para la variante **Lambda**. El patrón difiere con el de **Ómicron** en donde dos sustituciones E:T9I y E:T11A se encuentran en proporciones mayores al 95% para algunas de sus subvariantes.

Sustituciones Comunes/Consistentes (o ausentes) en Clados de Ómicron: E:T9I se observó en muchos de los clados de Ómicron (ej. Ómicron 1, 3, 5, 6, 7, 8, 10, 11, 12), pero no en todos (ej. Ómicron 2, 4 y 9). La sustitución E:T9V se observó solo en Ómicron 4 y 2. Aparentemente, sería una mutación menos común o específica de ciertos sublinajes de Ómicron. La sustitución E:E7K fue observada solo en Gamma, lo que sugiere que este cambio no se ha establecido en otro linaje o se trataría de una sustitución neutral.

Sustituciones Específicas de Variantes/Clados: Se identificaron cambios “marcadores” de variante. Por ejemplo: E:L73P, E:L73S, E:I74N, E:I74F, E:I74M, E:V75F, E:V75L, E:S76N fueron observadas principal o exclusivamente en los clados de Ómicron, y su distribución parece concentrarse en un subconjunto de esos clados (Ómicron 3, 4, 8, 9, 10, 11 y 12). Esto sugiere una evolución divergente dentro de Ómicron. Por otro lado, E:S76N solo aparece en Ómicron 8.

Es importante destacar que el verdadero impacto de cambios deberá evaluarse en el contexto del resto de los genes con los que pueden presentar posibles efectos sinérgicos o antagónicos.

Variante	Cantidad	Fecha Primer Secuencia	Fecha última secuencia
No-VOI	2680	01/03/2020	06/01/2022
Gamma	3110	22/07/2020	22/12/2021
Lambda	1366	08/11/2020	15/11/2022
Delta	4135	09/11/2020	02/02/2022
Alpha	393	21/12/2020	07/12/2021
Epsilon	43	04/01/2021	24/06/2021
Omicrom_02	3043	06/01/2021	07/11/2022
Beta	1	24/04/2021	24/04/2021
Mu	45	30/04/2021	19/12/2021
Eta	1	04/05/2021	04/05/2021
Omicrom_01	89	21/12/2021	12/09/2022
Omicrom_06	177	02/01/2022	06/03/2023
Otras-No Clasificadas	28	07/01/2022	24/02/2025
Omicrom_03	1041	18/01/2022	24/12/2022
Omicrom_05	886	09/05/2022	23/06/2023
Omicrom_04	941	21/05/2022	04/03/2023
Omicrom_07	2086	22/05/2022	23/06/2023
Omicrom_08	225	12/10/2022	20/11/2023
Omicrom_09	946	06/12/2022	22/03/2024
Omicrom_10	573	16/11/2023	27/03/2025
Omicrom_11	210	19/07/2024	26/02/2025
Omicrom_12	107	13/09/2024	28/03/2025
Omicrom_13	0		
Theta	0		
Total	22126	01/03/2020	28/03/2025

Tabla 1. Número de secuencias de Argentina depositadas en GISAID según su clasificación en variantes de interés (VOI) y de preocupación (VOC).

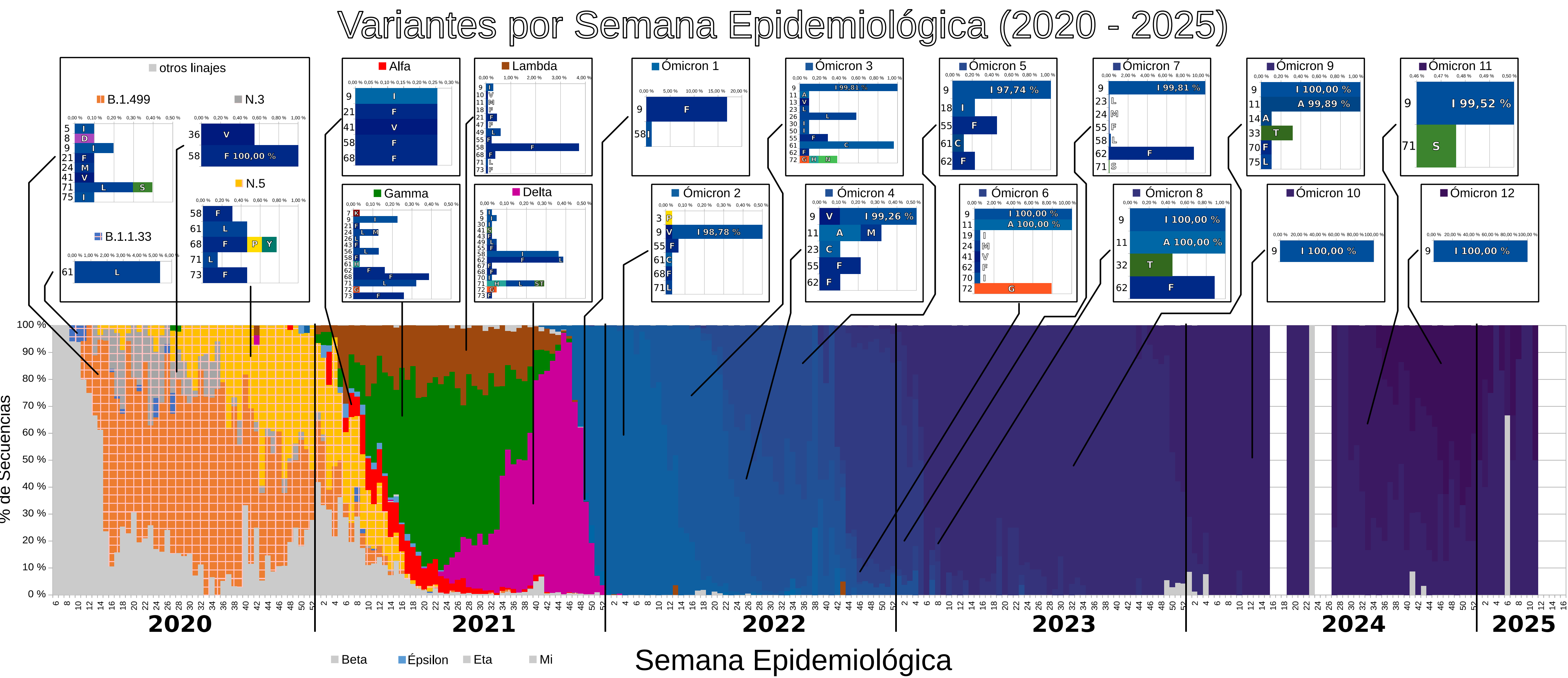


Figura 1. Cambio en la composición de linajes y variantes de SARS-CoV-2 por semana epidemiológica durante el período estudiado (marzo 2020-abril 2025). Las variantes corresponden a las definidas por el Centro para el Control y Prevención de Enfermedades de Europa (ECDC). En el panel superior se muestra para cada variante con más de 50 secuencias, las posiciones y la proporción en las que se observaron sustituciones en el gen E.

Bibliografía:

- Zhou, S., Lv, P., Li, M., Chen, Z., Xin, H., Reilly, S., & Zhang, X. (2023). SARS-CoV-2 E protein: Pathogenesis and potential therapeutic development. *Biomedicine & Pharmacotherapy*, 159, 114242 - 114242. <https://doi.org/10.1016/j.biopha.2023.114242>.
- Liu, B., Yao, Q., Cruz-Cosme, R., Yarbrough, C., Draper, K., Suslovic, W., Muhammad, I., Contes, K., Hillyard, D., Teng, S., & Tang, Q. (2025). Genetic Conservation and Diversity of SARS-CoV-2 Envelope Gene Across Variants of Concern.. *Journal of medical virology*, 97 1, e70136. <https://doi.org/10.1002/jmv.70136>.
- Lacadena J.R. 1999. Capítulo 1 - Conceptos Generales. En *Genética General* (ISBN: 8477386455)
- GISAID - Global Initiative on Sharing All Influenza Data - www.gisaid.org
- ECDC - European Centre for Disease Prevention and Control - <https://www.ecdc.europa.eu/en/covid-19/variants-concern>
- NextClade: Viral genome clade assignment, mutation calling, and sequence quality checks. <https://github.com/nextstrain/nextclade>
- Python Software Foundation. (2024). Python (3.12.3). [Software]. Obtenido de <https://www.python.org>

Conclusiones:

La proteína E ha ido acumulando mutaciones a lo largo del tiempo y en diferentes linajes virales y hoy pueden observarse como polimorfismos. La variación de las frecuencias en la que se observan dichos polimorfismos puede darse por la fijación de nuevas mutaciones o por procesos migratorios que introducen variantes o linajes con proporciones diferentes de los mismos. En la Argentina el proceso de recambio de variantes permite explicar las modificaciones en las frecuencias observadas para estos sitios bajo estudio. En diferentes variantes se pudieron observar la aparición de las mismas mutaciones de forma independiente (convergencia) mientras que otros linajes acumularon un conjunto único de mutaciones (divergencia). Para explicar este patrón se deben estudiar mejor las propiedades del gen E para establecer si se trata de posiciones con plasticidad genética (admiten variaciones sin cambiar el *fitness*) o si se trata de sustituciones que resultan favorables para el desarrollo del virus (aumentan el *fitness*) y por lo tanto hay convergencia. La vigilancia genómica y el desarrollo de herramientas de análisis Bioinformático son fundamentales para rastrear estas mutaciones y comprender su impacto clínico. Sin embargo, la falta de registros completos limita la interpretación de los resultados. Futuros estudios que integren datos genómicos y clínicos son esenciales para dilucidar la relación entre las mutaciones en la proteína E y la patogénesis del SARS-CoV-2.