

CARTA AL EDITOR

ESCASA INFORMACIÓN GENÓMICA EN BASES DE DATOS PÚBLICAS PARA INVESTIGAR EL SARS-COV-2 EN LATINOAMÉRICA

LIMITED GENOMIC INFORMATION IN PUBLIC DATABASES TO INVESTIGATE SARS-COV-2 IN LATIN AMERICA

Pedro Eduardo Romero ^{1,a}

¹ Facultad de Ciencias y Filosofía. Universidad Peruana Cayetano Heredia, Lima, Perú.

^a Licenciado en Biología, doctor en Ciencias Naturales.

Sr. Editor: La caracterización bioinformática de los genomas del Coronavirus-2019 (SARS-CoV-2) es de vital importancia en la pandemia. Por ejemplo, el análisis de la estructura de la proteína viral *spike*, que media la endocitosis por las células respiratorias humanas, permitirá describir epítomos adecuados para la obtención de vacunas ⁽¹⁾. La información genómica también es utilizada en vigilancia y epidemiología molecular para encontrar el origen del virus ⁽²⁾, describir los linajes circulantes y seguir su dispersión a escala global y local ⁽³⁾, como en el caso del proyecto Nexstrain ⁽⁴⁾ que permite visualizar periódicamente las relaciones evolutivas de los linajes virales. A partir de ello, podemos identificar el número de linajes diferentes presentes en cierta región, información que puede utilizarse para generar vacunas adaptadas a genotipos locales. También podemos fechar los eventos de divergencia genética ocurridos desde diciembre de 2019. Esta información, junto con los datos epidemiológicos, nos revela la ruta detallada de dispersión global y dentro de los países afectados.

Los genomas virales vienen siendo depositados en bases de datos internacionales y de acceso abierto, como la Iniciativa global para compartir todos los datos de influenza (GISAID, por sus siglas en inglés) o la base de datos de secuencias genéticas GenBank. A la fecha, existen 3734 secuencias de SARS-CoV-2 en GISAID y 446 en GenBank. El Perú está representado por dos secuencias en GISAID: EPI_ISL_415787, aislada de una mujer de 61 años, y EPI_ISL_416025, aislada de un hombre de

26 años; y solo un genoma se encuentra depositado en GenBank (MT263074). Estas tres secuencias fueron generadas por el Instituto Nacional de Salud.

En Latinoamérica, los datos son escasos. De los 49 genomas reportados en GISAID, Brasil tiene 36 genomas virales; Chile, 7; Colombia, 2; Ecuador, 1; y Panamá, 1. En GenBank, además de Perú, encontramos un registro para Brasil (MT126808) y un registro para Colombia (MT256924). Añadamos a la falta de datos genómicos, el problema de la calidad. Muchos de los genomas de muestras latinoamericanas en GISAID, considerando también la muestra peruana (EPI_ISL_416025), tienen una proporción considerable de regiones ambiguas o que no representan el genoma viral completo, probablemente por errores durante el protocolo de secuenciación. Estos datos incompletos o erróneos no permitirán realizar eficientemente los análisis descritos.

Necesitamos con urgencia más información de los genomas de SARS-CoV-2 que circulan en Latinoamérica para contribuir desde la genómica, la bioinformática y la epidemiología molecular en la lucha contra esta pandemia. Por lo tanto, la difusión de más genomas, con buena calidad de secuenciamiento y en bases de datos públicas, es imprescindible para que investigadores locales y extranjeros conozcan mejor este virus.

Fuente de financiamiento: Autofinanciado.

Conflictos de interés: El autor declara no tener conflictos de interés.

REFERENCIAS BIBLIOGRÁFICAS

1. Bhattacharya M, Sharma AR, Patra P, Ghosh P, Sharma G, Patra BC, *et al.* Development of epitope-based peptide vaccine against novel coronavirus 2019 (SARS-COV-2): Immunoinformatics approach. *J Med Virol.* 2020; 1-14. doi: 10.1002/jmv.25736.
2. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. The proximal origin of SARS-CoV-2. *Nat Med.* 2020;1-3. doi: 10.1038/s41591-020-0820-9.
3. Rife BD, Mavian C, Chen X, Ciccozzi M, Salemi M, Min J, *et al.* Phylodynamic applications in 21 st century global infectious disease research. *Glob Health Res Policy.* 2017;2(1):13. doi: 10.1186/s41256-017-0034-y.
4. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, *et al.* Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics.* 2018;34(23):4121-3. doi: 10.1093/bioinformatics/bty407.

Citar como: Romero PE. Escasa información genómica en bases de datos públicas para investigar el SARS-CoV-2 en Latinoamérica. *Rev Peru Med Exp Salud Publica.* 2020;37(2):374. doi: <https://doi.org/10.17843/rpmesp.2020.372.5396>

Correspondencia: Pedro Eduardo Romero Condori; pedro.romero@upch.pe

Recibido: 26/03/2020 Aprobado: 03/04/2020 En línea: 06/04/2020