

Status: Preprint has been published in a journal as an article  
DOI of the published article: <https://doi.org/10.17843/rpmesp.2021.382.6417>

# Use of complete SARS-CoV-2 genomes to estimate the Rt in Peru during March – April 2020

Pedro Romero, Milagros Sánchez-Yupari, Stephanie Montero, Pablo Tsukayama

<https://doi.org/10.1590/SciELOPreprints.1230>

This preprint was submitted under the following conditions:

- The authors declare that they are aware that they are solely responsible for the content of the preprint and that the deposit in SciELO Preprints does not mean any commitment on the part of SciELO, except its preservation and dissemination.
- The authors declare that the research that originated the manuscript followed good ethical practices and that the necessary approvals from research ethics committees are described in the manuscript, when applicable.
- The authors declare that the necessary Terms of Free and Informed Consent of participants or patients in the research were obtained and are described in the manuscript, when applicable.
- The authors declare that the preparation of the manuscript followed the ethical norms of scientific communication.
- The authors declare that the manuscript was not deposited and/or previously made available on another preprint server or published by a journal.
- The submitting author declares that all authors responsible for preparing the manuscript agree with this deposit.
- The submitting author declares that all authors' contributions are included on the manuscript.
- The authors declare that if the manuscript is posted on the SciELO Preprints server, it will be available under a [Creative Commons CC-BY](#) license.
- The deposited manuscript is in PDF format.
- If the manuscript is being reviewed or being prepared for publishing but not yet published by a journal, the authors declare that they have received authorization from the journal to make this deposit.

Submitted on (YYYY-MM-DD): 2020-09-18

Posted on (YYYY-MM-DD): 2021-03-31

## 1. Título

Uso de genomas completos de SARS-CoV-2 para la estimación del Rt en Perú durante marzo y abril, 2020

### Title

Use of complete SARS-CoV-2 genomes to estimate the Rt in Peru during March – April 2020

## 2. Título corto

Rt usando genomas SARS-CoV-2 peruanos

## 3. Identificación de los autores / Author information

Pedro E. Romero 1,a

<https://orcid.org/0000-0001-9947-3868>

Milagros Sánchez-Yupari 1,b

<https://orcid.org/0000-0003-2113-0253>

Stephanie Montero 2,c

<https://orcid.org/0000-0002-8147-8318>

Pablo Tsukayama 1,d

<https://orcid.org/0000-0002-1669-2553>

## 4. Filiación / Affiliation

1. Facultad de Ciencias y Filosofía. Universidad Peruana Cayetano Heredia, Lima, Perú.
  2. Unidad de Investigación en Enfermedades Emergentes y Cambio Climático. Universidad Peruana Cayetano Heredia, Lima, Perú.
- 
- a. Biólogo con mención en Genética y Biología Celular. Doctor en Ciencias Naturales.
  - b. Bachiller en Ciencias.
  - c. Bióloga Microbióloga Parasitóloga.
  - d. Licenciado en Biología. Doctor en Filosofía en Biología y Ciencias Biomédicas.

## 5. Contribuciones de los autores / Author contribution

PER participó en la concepción y diseño del artículo, análisis e interpretación de datos, redacción del artículo y revisión crítica de la versión final. MSY participó en el análisis e interpretación de datos y redacción del artículo. SM participó en la interpretación de datos, redacción y revisión crítica del artículo. PT participó en la redacción y revisión crítica del artículo. Todos los autores aprueban la versión final.

## **6. Financiamiento / Funding**

PER recibió financiamiento del Fondo Nacional de Desarrollo Científico y Tecnológico y de Innovación Tecnológica (Fondecyt - Perú) - "Proyecto de Mejoramiento y Ampliación de los Servicios del Sistema Nacional de Ciencia, Tecnología e Innovación Tecnológica": Número de contrato 34-2019-FONDECYT-BM-INC. INV.

## **7. Conflictos de interés / Conflict of interest statement**

Los autores declaran no tener conflictos de interés.

## **8. Correspondencia**

Pedro E. Romero  
Facultad de Ciencias y Filosofía - UPCH  
Av. Honorio Delgado 430, 15102 Lima  
pedro.romero@upch.pe

## **Abstract**

Understanding the current SARS-CoV-2 pandemics dynamics using a multidisciplinary approach is essential to improve evidence-based policies. In our study, we estimated the SARS-CoV-2 effective reproductive number ( $R_t$ ) in Peru using information from 69 complete genomes sequenced by INS-Peru and available in the GISAID public database. The tendency of  $R_t$  within March and April 2020 was similar to independent epidemiological reports. The  $R_t$  value decreased ( $R_t < 1$ ) during the first weeks of the pandemics in Peru but increases ( $R_t > 1$ ) before the beginning of April. More genomic information from Peruvian samples, available in public databases, that also represent the broad geographic diversity of Peru will allow better estimations based on genetic data and will complement epidemiological information.

## 1. Título

Uso de genomas completos de SARS-CoV-2 para la estimación del  $R_t$  en Perú durante marzo y abril, 2020

## 2. Resumen

La comprensión de la dinámica de la pandemia más reciente causada por el virus SARS-CoV-2 utilizando un enfoque multidisciplinario es esencial para mejorar la toma de decisiones basadas en evidencia. En nuestro estudio estimamos el número reproductivo efectivo ( $R_t$ ) del SARS-CoV-2 en Perú a partir de información pública de 69 genomas completos generados por el INS-Perú y disponibles en GISAID. La tendencia mostrada por el  $R_t$  durante marzo y abril 2020 fue similar a la del  $R_t$  calculado por métodos convencionales, disminuyendo durante las primeras semanas de marzo ( $R_t < 1$ ) y aumentando moderadamente antes del inicio de abril ( $R_t > 1$ ). La obtención de más información genómica, que pueda ser compartida en bases de datos abiertas, y que tenga una mayor representación geográfica permitirá obtener estimaciones locales de  $R_t$  a partir de datos genéticos y complementar la información epidemiológica.

### Abstract

Understanding the current SARS-CoV-2 pandemics dynamics using a multidisciplinary approach is essential to improve evidence-based policies. In our study, we estimated the SARS-CoV-2 effective reproductive number ( $R_t$ ) in Peru using information from 69 complete genomes sequenced by INS-Peru and available in the GISAID public database. The tendency of  $R_t$  within March and April 2020 was similar to independent epidemiological reports. The  $R_t$  value decreased ( $R_t < 1$ ) during the first weeks of the pandemics in Peru but increases ( $R_t > 1$ ) before the beginning of April. More genomic information from Peruvian samples, available in public databases, that also represent the broad geographic diversity of Peru will allow better estimations based on genetic data and will complement epidemiological information.

## 3. Palabras clave (Keywords) (Fuente: DeCS BIREME)

SARS-CoV-2; evolución molecular; epidemiología; número reproductivo efectivo  
SARS-CoV-2; molecular evolution; epidemiology; effective reproductive number

## 4. Contenido del artículo

### Introducción

La pandemia actual de la enfermedad COVID-19 causada por el virus SARS-CoV-2 fue reportada por primera vez en diciembre 2019 en la ciudad de Wuhan, China y a la fecha, se ha convertido en un problema global de salud pública habiendo afectado a más de 28 millones de personas y producido más de 920 mil decesos (1). El Perú reportó su primer caso de coronavirus el día 6 de marzo 2020. El día 16 de marzo el gobierno peruano decretó el estado nacional de emergencia que incluyó medidas de cuarentena y toque de queda (2). A pesar de estas medidas, la transmisión se generalizó

dentro del país y a la fecha de envío de este trabajo se ha detectado por diagnóstico serológico o molecular más de 720 mil casos y más de 30 mil muertes (3). El uso de herramientas epidemiológicas y estadísticas durante una pandemia es de mucha utilidad en la toma de decisiones por que pueden proveer un panorama inicial de la dinámica de propagación de la enfermedad, y apoyar en el aislamiento precoz de los positivos, el seguimiento de contactos, y selección de medidas de mitigación comunitarias e individuales adecuadas (4). Una de las estimaciones importantes para conocer la velocidad de transmisión potencial durante la fase inicial de la pandemia es el número reproductivo básico ( $R_0$ ) (5). Este número ha sido estimado en 2,3 (Intervalo de confianza 95% [IC]: 2,0, 2,5) tanto para Lima (2) como para Perú (6) durante marzo. Otro valor importante para seguir la transmisión es el número reproductivo efectivo, abreviado como  $R_t$  o  $R_e$ , que mide el número esperado de nuevas infecciones causadas por un individuo infeccioso y se calcula durante el transcurso de la pandemia cuando ya existen individuos no susceptibles en la población (7) Si el  $R_t$  es mayor que 1, los casos se incrementan exponencialmente, si es menor que 1, el número de infectados será cada vez menor (8). El  $R_t$  para Perú fue estimado en 2,36 (IC: 2,11-2,63) durante marzo (7).

La disponibilidad de datos libres en repositorios públicos promueve la ciencia abierta y la rápida toma de decisiones en salud pública (9). Por ejemplo, el primer genoma de SARS-CoV-2 se publicó en enero de 2020 y permitió el desarrollo temprano de pruebas moleculares RT-PCR (10) que se utilizan mundialmente para confirmar la presencia del virus. Asimismo, la información genómica es utilizada en análisis filodinámicos para obtener estimaciones sobre la transmisión de enfermedades emergentes, analizando los procesos de evolución molecular en tiempo real y colaborando con la vigilancia epidemiológica (11). El cálculo del  $R_t$  es posible con modelos similares a los modelos epidemiológicos SIR (susceptibles, infectados, recuperados), utilizando la información filogenética para inferir las tasas de transmisión y recuperación (12). El primer genoma de un caso peruano fue secuenciado por el Instituto Nacional de Salud del Perú (INS) en marzo 2020 (13). Al 31 de agosto de 2020, existían 72 genomas depositados en la base de datos pública *Global Initiative on Sharing All Influenza Data* (GISAID) correspondientes a casos de marzo y abril 2020. Por ello, el objetivo de nuestro estudio fue utilizar la información genómica disponible para Perú y estimar el  $R_t$  durante los dos primeros meses, y de esta manera, evaluar la capacidad de los abordajes moleculares y bioinformáticos como complemento a las estimaciones epidemiológicas tradicionales.

## **Materiales y Métodos**

Las muestras utilizadas en el estudio fueron procesadas por el Laboratorio de Referencia Nacional de Virus Respiratorios y generadas por el Laboratorio de Referencia Nacional de Biotecnología y Biología Molecular, ambos pertenecientes al Centro Nacional de Salud Pública del INS. Los genomas se encuentran depositados en la base de datos pública GISAID y fueron descargados el 01 setiembre 2020. De los 72 genomas disponibles, se utilizaron 69 genomas que cumplían con las siguientes características: genoma completo (>29 000 pares de bases) y un porcentaje de nucleótidos ambiguos menor a 1. Los números de acceso se encuentran en el Material

Suplementario 1. 33 muestras correspondieron a Lima y 36 a otras regiones: (Ancash, 7; Arequipa, 3; Cajamarca, 2; Cusco, 2; Huánuco, 1; Ica, 4; Junín, 2; La Libertad, 3; Lambayeque, 7; Loreto, 5). Las secuencias fueron alineadas con el programa MAFFT (14) junto con el genoma de referencia del SARS-CoV-2 (código de acceso de GenBank: NC\_045512.2). El alineamiento fue editado manualmente en el programa Geneious v.2020.2.3, para eliminar los flancos correspondientes a las regiones no traducidas (UTR). El tamaño final del alineamiento resultó en 29 409 sitios. Este alineamiento puede ser obtenido en el Material Suplementario 2.

Los parámetros del modelo evolutivo fueron estimados en BEAST2 (15) que utiliza una aproximación bayesiana mediante la generación y muestreo de cadenas de Markov Monte Carlo (MCMC). La tasa de sustitución nucleotídica, 23,574 sustituciones por año, fue obtenida a partir del portal de vigilancia genómica NextStrain (16). La tasa dividida entre el tamaño del genoma de referencia NC\_045512.2 (29 903 nucleótidos) resulta en 0,0008 sustituciones/sitio/año. Esta tasa ha sido utilizada en otros estudios de introducciones y transmisión temprana de SARS-CoV-2 (17, 18). La MCMC fue corrida por 10 millones de generaciones, realizando muestreos cada 10 000 generaciones y descartando el primer 25% de resultados (burnin). Se corrieron 10 cadenas MCMC independientes que luego se combinaron en un solo archivo de salida disponible en el Material Suplementario 3. En este archivo se verificó la convergencia de la distribución posterior de los parámetros observando un valor mayor a 200 en el índice ESS (tamaño de muestra estimado). El archivo de entrada puede ser obtenido desde el Material Suplementario 4.

La estimación del  $R_t$  a partir de datos genéticos es posible utilizando el modelo *birth-death skyline* (BDSKY) (19). Los eventos de infección (*birth*) y de “recuperación”, resultantes por muerte, tratamiento o cambios en el comportamiento de los individuos (*death*), son modelados a partir de filogenias generadas por análisis bayesiano. Para ello, es necesario utilizar parámetros a priori que permitan la estimación de una distribución posterior (Tabla 1). Por ejemplo, la tasa de volverse no infeccioso (*becomeUninfectiousrate*) está expresada en unidades por año, y resulta de dividir los días del año entre el promedio del tiempo (días) en el que un individuo es infeccioso ( $365/15 = 25$  unidades, aprox.). El origen de la pandemia fue estimado entre noviembre - diciembre 2019. Finalmente, asignamos al número efectivo reproductivo una media de 1,0 con un valor inicial de 2,0.

## Resultados

La probabilidad posterior, así como la tasa de convertirse en no infeccioso y el número reproductivo efectivo superaron el límite ESS > 200 indicando que existió una convergencia de valores en la distribución posterior (Tabla 2). La tasa de convertirse en no infeccioso (individuo recuperado) se calculó en 33,7; 95% HPD: 23,09 – 43,54 (HPD: intervalo de mayor densidad de probabilidad), correspondiendo a un promedio de 10 días. El  $R_t$  promedio estimado durante la primera quincena de marzo fue 2,13; 95% HPD: 1,09 – 3,64 (Tabla 2, Material Suplementario 5). El valor estimado de  $R_t$  era mayor que 1 días antes del primer reporte de SARS-CoV-2 en Perú (Figura 1, Material Suplementario 5). El valor disminuyó durante la segunda semana de marzo y se mantuvo menor a 1

elevándose nuevamente en la última semana de marzo. Finalmente, el  $R_t$  se mantuvo por encima de 1 hasta la tercera semana de abril.

## **Discusión**

Nuestra investigación muestra concordancia con las estimaciones de  $R_t$  a partir de estudios epidemiológicos (4, 7) y reportes que no han sido revisados por pares (8), proveyendo evidencia independiente para el cálculo de este valor mediante datos genómicos. La disminución del  $R_t$  pocos días antes del inicio de la cuarentena podría explicarse debido a las medidas tomadas por el gobierno la semana previa, suspendiendo clases, y vuelos desde Europa y Asia, así como prohibiendo las reuniones de más de 300 personas (2). El posterior aumento del  $R_t$  podría haberse producido por aglomeraciones recurrentes en lugares públicos como mercados o bancos. El confinamiento es una medida de mitigación estricta que ha funcionado en países con mercados laborales formales, sin embargo, en Perú, la informalidad laboral llegaría a más de 72% de la población económicamente activa. Por ello, el confinamiento suprimió actividades económicas informales significativas y se convirtió en una medida insostenible en el contexto de nuestra dinámica social (20).

Como limitaciones de este estudio tenemos el número limitado de genomas secuenciados durante los meses de marzo y abril. Los genomas incluidos no representan proporcionalmente a las infecciones por SARS-CoV-2 en regiones diferentes de Lima que, determinantes estructurales (gestión de la provisión de salud, acceso a centros de salud, acceso al diagnóstico, capacidad de atención a personas enfermas) y sociales (estatus socioeconómico, comorbilidades, salud mental y factores comportamentales), mostrarían una dinámica de transmisión diferencial. Aunque a la fecha, Perú tiene uno de los números más altos de casos reportados mundialmente, sólo un 24% ha sido confirmado mediante prueba molecular (RT-PCR) (3). Incluso en este escenario, el número de genomas reportados es pequeño por lo que son necesarias más iniciativas de secuenciamiento genómico para obtener un conjunto mayor de datos, idealmente cientos a miles de genomas que permitan un cálculo más certero, así como para conocer la diversidad genética, la evolución y la transmisión del SARS-CoV-2 en Perú. Además de aumentar el número de genomas secuenciados, también es necesario incrementar la representación geográfica de las muestras para poder estimar el  $R_t$  de cada región y seguir de manera más detallada la evolución de la epidemia en Perú.

Aún con las limitaciones descritas, nuestro estudio muestra la relevancia de los datos genómicos para complementar estudios epidemiológicos y la necesaria integración de esta información en un futuro sistema nacional de vigilancia epidemiológica, el cual permita la toma de decisiones sustentadas en evidencia generada mediante ciencia abierta y multidisciplinaria. Finalmente, creemos también necesario promover la formación de recursos humanos capacitados que puedan comprender e integrar métodos epidemiológicos con herramientas bioinformáticas.

## **5. Referencias bibliográficas**

1. WHO Coronavirus Disease (COVID-19) Dashboard [Internet]. Suiza: WHO; 2020 [citado el 02 setiembre 2020]. Disponible en: <https://covid19.who.int>
2. Munayco CV, Tariq A, Rothenberg R, Soto-Cabezas GG, Reyes MF, Valle A, et al. Early transmission dynamics of COVID-19 in a southern hemisphere setting: Lima-Peru: February 29(th)-March 30(th), 2020. *Infect Dis Model.* 2020;5:338-45. doi:10.1016/j.idm.2020.05.001
3. Sala situacional COVID-19 [Internet]. Perú: Ministerio de Salud; 2020 [citado el 02 setiembre 2020]. Disponible en: [https://covid19.minsa.gob.pe/sala\\_situacional.asp](https://covid19.minsa.gob.pe/sala_situacional.asp)
4. Gonzales-Castillo JR, Varona-Castillo L, Dominguez-Morante MG, Ocaña-Gutierrez VR. Pandemia de la COVID-19 y las Políticas de Salud Pública en el Perú: marzo-mayo 2020. *Revista de Salud Pública.* 2020;22(2):1-9. doi:10.15446/rsap.v22n2.87373
5. Ridenhour B, Kowalik JM, Shay DK. Unraveling R0: considerations for public health applications. *Am J Public Health.* 2014;104(2):e32-41. doi:10.2105/AJPH.2013.301704
6. Torres-Roman JS, Kobiak IC, Valcarcel B, Diaz-Velez C, La Vecchia C. The reproductive number R0 of COVID-19 in Peru: An opportunity for effective changes. *Travel Med Infect Dis.* 2020:101689. doi:10.1016/j.tmaid.2020.101689
7. Caicedo-Ochoa Y, Rebellon-Sanchez DE, Penaloza-Rallon M, Cortes-Motta HF, Mendez-Fandino YR. Effective Reproductive Number estimation for initial stage of COVID-19 pandemic in Latin American Countries. *Int J Infect Dis.* 2020;95:316-8. doi:10.1016/j.ijid.2020.04.069
8. Rt de COVID-19 por departamento en Perú [Internet]. Lima: AMIGOCLOUD; 2020 [citado el 05 setiembre 2020]. Disponible en: <https://huaynodata.com/>
9. Moorthy V, Henao Restrepo AM, Preziosi MP, Swaminathan S. Data sharing for novel coronavirus (COVID-19). *Bull World Health Organ.* 2020;98(3):150. doi:10.2471/BLT.20.251561
10. Corman VM, Landt O, Kaiser M, Molenkamp R, Meijer A, Chu DK, et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro Surveill.* 2020;25(3). doi:10.2807/1560-7917.ES.2020.25.3.2000045
11. 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;2:13. doi:10.1186/s41256-017-0034-y
12. Vaughan TG, Leventhal GE, Rasmussen DA, Drummond AJ, Welch D, Stadler T. Estimating Epidemic Incidence and Prevalence from Genomic Data. *Mol Biol Evol.* 2019;36(8):1804-16. doi:10.1093/molbev/msz106
13. Padilla-Rojas C, Lope-Pari P, Vega-Chozo K, Balbuena-Torres J, Caceres-Rey O, Bailon-Calderon H, et al. Near-Complete Genome Sequence of a 2019 Novel Coronavirus (SARS-CoV-2) Strain Causing a COVID-19 Case in Peru. *Microbiol Resour Announc.* 2020;9(19). doi:10.1128/MRA.00303-20
14. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform.* 2019;20(4):1160-6. doi:10.1093/bib/bbx108

15. Bouckaert R, Vaughan TG, Barido-Sottani J, Duchene S, Fourment M, Gavryushkina A, et al. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLoS Comput Biol. 2019;15(4):e1006650. doi:10.1371/journal.pcbi.1006650
16. 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
17. Gambaro F, Behillil S, Baidaliuk A, Donati F, Albert M, Alexandru A, et al. Introductions and early spread of SARS-CoV-2 in France, 24 January to 23 March 2020. Euro Surveill. 2020;25(26). doi:10.2807/1560-7917.ES.2020.25.26.2001200
18. Lai A, Bergna A, Acciarri C, Galli M, Zehender G. Early phylogenetic estimate of the effective reproduction number of SARS-CoV-2. J Med Virol. 2020;92(6):675-9. doi:10.1002/jmv.25723
19. Stadler T, Kuhnert D, Bonhoeffer S, Drummond AJ. Birth-death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). Proc Natl Acad Sci U S A. 2013;110(1):228-33. doi:10.1073/pnas.1207965110
20. Organización Mundial del Trabajo. ¿Respuesta rápida a la COVID-19 en un contexto de alta informalidad? El caso del Perú [Internet]. Ginebra: Organización Mundial del Trabajo; 2020 [citado el 13 de setiembre de 2020]. [http://www.ilo.org/wcmsp5/groups/public/---ed\\_emp/documents/publication/wcms\\_747776.pdf](http://www.ilo.org/wcmsp5/groups/public/---ed_emp/documents/publication/wcms_747776.pdf)

## Tablas

**Tabla 1.** Parámetros *a priori* del modelo BDSKY en BEAST2.

Parámetro	Media	Desviación estándar	Distribución
<i>becomeUninfectiousrate</i>	25	6	Normal
Origen	0,4 (Nov-Dic 2019)	0,03	Normal
Número efectivo reproductivo	1,00	1,00	LogNormal

**Tabla 2.** Parámetros estimados utilizando el modelo BDSKY.

Parámetro	Media	Intervalo 95% HPD	ESS
Probabilidad posterior	-40696,59	-40720,53 – -40674,89	2235
<i>becomeUninfectiousrate</i>	33,77	23,09 – 43,54	3371
Número efectivo reproductivo (primera quincena de marzo)	2,13	1,09 – 3,64	4070

ESS: Tamaño de muestra estimado. HPD: Intervalo de mayor densidad posterior.

**Figura 1.** Valores es

timados de  $R_t$  por el modelo BDSKY. La incertidumbre es representada mediante una distribución 95% HPD. El valor de  $R_t$  disminuye durante las dos primeras semanas de marzo y aumenta moderadamente la última semana de marzo 2020.

