

Primeras seis secuencias del genoma completo de SARS-CoV-2 por NGS en El Salvador

DOI: 10.5377/alerta.v4i1.10682

Noé Rigoberto Rivera¹

Orcid: 0000-0001-9553-0554

Carlos Alexander Ortega Pérez^{1*}

Orcid: 0000-0003-1612-8386

Xochitl Sandoval López²

Orcid: 0000-0002-0988-1313

Carlos Enrique Hernández Ávila³

Orcid: 0000-0002-1230-0102

¹Facultad de Medicina, Universidad de El Salvador, El Salvador; ²Instituto Nacional de Salud, El Salvador; ³Departamento de Gobernanza y gestión del conocimiento, Instituto Nacional de Salud El Salvador.

Recibido: 28 de diciembre de 2020

Aceptado: 15 de enero de 2021

Publicado: 22 de enero de 2021

*Correspondencia: carlos.ortega@ues.edu.sv

Citación recomendada: Rivera NR, Ortega Pérez CA, Sandoval López X, Hernández Ávila CE. Primeras seis secuencias del genoma completo de SARS-CoV-2 por NGS en El Salvador. *Alerta*. 2021;4(1):61-66. DOI 10.5377/alerta.v4i1.10682

Contribución de autoría: NRR1, CAOP2, XSL3, CEHA4. del manuscrito. NRR1, CAOP2, procesaron las muestras y analizaron los datos en las plataformas bioinformáticas.

Resumen

Introducción. En este trabajo se describen las primeras secuencias completas del genoma de SARS-CoV-2 a partir de muestras de pacientes salvadoreños.

Objetivo. Reportar las primeras secuencias completas del genoma del SARS-CoV-2 de casos procedentes de El Salvador.

Metodología. Se realizó una secuenciación masiva en la plataforma MiniSeq Illumina a partir de muestras de secreción nasofaríngea.

Resultados. El análisis filogenético determinó que estas muestras pertenecen al clado 20C secundario de 20A que tiene en común la variante de la mutación D614G de la glicoproteína espícula. La mutación S: D614G fue encontrada en las seis secuencias de SARS-CoV-2. En la plataforma GISAID, las secuencias mostraron pertenecer al clado GH linaje pangolín B.1.2 y B.1.370; ambos linajes están presentes en Estados Unidos.

Conclusión. El análisis filogenético evidenció que estas seis muestras pertenecen al clado 20C, clado secundario de 20A, que tiene en común la variante de la mutación D614G de la glicoproteína espícula.

Palabras clave: SARS-CoV-2, D614G, NGS, 2019-nCoV, COVID-19.

Abstract

Introduction. This work describes the first complete SARS-CoV-2 genome sequences from samples of Salvadorean patients.

Objective. To report the first six complete SARS-CoV-2 genome sequences of six indigenous cases of COVID-19 detected in El Salvador, from nasopharyngeal discharge samples.

Methodology. An NGS next generation sequencing was performed from indigenous samples in the Illumina platform MiniSeq.

Results. The phylogenetic analysis showed that these isolated belong to the clade 20C secondary of 20A, which have in common the S: D614G variant; the D614G mutation of the spike glycoprotein was found in the six sequences of the SARS-CoV-2 genome. Using the GISAID platform, the sequences were found to belong to the clade GH lineage pangolin B.1.2 y B.1.370, both present in the United States.

Conclusion. The phylogenetic analysis showed that these six samples belong to clade 20C, a secondary clade of 20A, which has in common the variant of the spike glycoprotein D614G mutation.

Keywords: SARS -CoV-2, D614G, NGS, 2019-nCoV, COVID-19.

Introducción

El SARS-CoV-2 fue descrito como causante de COVID-19 por primera vez en Wuhan, Hubei, China, el 2 de enero del 2020. En los Estados Unidos de América (EUA), se reportó el primer caso el 20 de enero de 2020, en el estado de Washington D. C., confirmado por los Centros para el Control y la Prevención de Enfermedades de Atlanta (CDC, por sus siglas en inglés)^{1,2}.

El 18 de marzo se reporta el primer caso confirmado de infección por SARS-CoV-2 en El Salvador. A nivel mundial, las tasas de nuevos casos y muertes por COVID-19 continuaron aumentando, con casi 4 millones de nuevos casos y 60 000 nuevas muertes registradas. Al 15 de noviembre de 2020, se había notificado a la OMS 53,7 millones de casos confirmados y 1,3 millones de muertes³.

El 3 de febrero se publicó la secuenciación del ARN de una muestra de líquido de lavado bronco alveolar de un paciente que trabajaba en un mercado de Wuhan y que ingresó en el Hospital Central de Wuhan el 26 de diciembre de 2019, mientras experimentaba un síndrome respiratorio severo que incluía fiebre, mareos y tos. Así, se identificó una nueva cepa de virus ARN de la familia Coronaviridae, que se designó en ese momento como co-

ronavirus «WH-Human 1» (y también se le conoció como «2019-nCoV»). El análisis filogenético del genoma viral completo (29 903 nucleótidos) reveló que el virus estaba estrechamente relacionado en un 89,1% con un grupo de coronavirus similares al SARS del género Betacoronavirus, subgénero Sarbecovirus, encontrado previamente en murciélagos en China⁴.

A partir de ese momento varios países comenzaron a publicar las secuencias obtenidas de pacientes locales y al momento de redacción del presente reporte, la última actualización del 27 de noviembre de 2020 publicada en la plataforma Global Initiative on Sharing All Influenza Data (GISAID, por sus siglas en inglés), registraba 3407 genomas, muestreados entre diciembre de 2019 y noviembre de 2020.

Como país es importante poder obtener las secuencias de SARS-CoV-2 a partir de muestras respiratorias de pacientes locales con COVID-19, pues esta información ayudará a comprender en parte el variado espectro clínico y curso de la enfermedad. Además, desde el punto de vista de la epidemiología molecular, permite comparar las secuencias entre diferentes poblaciones del mundo con las secuencias del país y analizar la evolución del virus a medida avanza la pandemia. Esto será de mucha utilidad para predecir qué podría pasar en los próximos años en el país y cómo responder ante esa eventualidad. Al principio de la pandemia las secuencias estudiadas han permitido determinar linajes, cadenas de transmisión y rutas de infección. Estas secuencias completas también nos permiten evaluar la variabilidad genética del virus a nivel local, su evolución y el poder predecir algunos factores pronósticos de los pacientes. De igual forma, aunque algunos estudios estadísticos proponen que las mutaciones descritas en la mayoría de secuencias reportadas hasta el momento no demuestran mayor relación con la virulencia, patogenicidad e incluso transmisibilidad e infectividad viral⁵, estudios recientes de prestigiosos grupos investigadores han demostrado el aumento del potencial de transmisión de las variantes con la mutación D614G, que predomina a nivel mundial⁶⁻⁹. Los cambios en la secuencia son importantes; sin embargo, las

secuencias también permiten evaluar el grado de conservación de las regiones del genoma viral, las cuales son incidentes para el estudio de métodos diagnósticos, blancos de fármacos antivirales, así como en la elaboración y seguimiento de eficacia de las vacunas.

Se reportan las primeras seis secuencias completas del genoma del SARS-CoV-2 de seis casos de COVID-19 detectados en El Salvador, a partir de muestras de secreción nasofaríngea.

Con la descripción de estas primeras secuencias se inicia una serie de publicaciones sobre la vigilancia genómica del virus SARS-CoV-2 que serán útiles para conocer la evolución del virus, desarrollo de terapias y vacunas.

Metodología

Se analizaron 50 eluidos de ARN de secreción nasofaríngea obtenida de pacientes positivos a la prueba de qRT-PCR para SARS-CoV-2, diagnosticados en el Laboratorio Nacional de Salud Pública del Instituto Nacional de Salud (INS) y la Sección de Virología y Microbiología Molecular, Departamento de Microbiología de la Facultad de Medicina, Universidad de El Salvador. La extracción se realizó con el kit PureLink Viral RNA/DNA Mini Kit Invitrogen de ThermoFisher Scientific. Todos los eluidos fueron conservados a $-80\text{ }^{\circ}\text{C}$ posterior a la extracción.

Se realizó la confirmación diagnóstica por qRT-PCR de los genes E y RdRp con dos protocolos: Charité, Berli/EVAg¹⁰ con la Ag- Path-ID One-Step RT-PCR de ThermoFisher Scientific y con el PowerChek 2019-nCoV Real-time PCR kit (<https://www.fda.gov/media/140069/download>). Las amplificaciones se realizaron en el QuantStudio 5 Real-Time PCR instrument de Applied Biosystems.

De 50 eluidos se escogieron 6 eluidos con CT inferior a 25 con una concentración $\geq 50\text{ ng}/\mu\text{L}$ de ARN determinado por fluorometría con Quantus fluorometer, Promega.

Se utilizó el método de secuenciación con enfoque de amplicones aplicando una variante de reacciones de PCR multiplex de dos grupos con el kit de Paragon CleanPlex® SARS-CoV-2 by SOPHiA, que ha presentado buenos resultados en estudios comparativos de secuenciación masiva (NGS Next- Generation Sequencing)^{11,12}, siguiendo el protocolo del fabricante. Para la generación de los amplicones se utilizó el termociclador eppendorf Flexlid Mastercycler Nexus gradient.

Secuenciación masiva (NGS)

Se utilizó la plataforma MiniSeq Illumina con el MiniSeq Rapid High Output Reagent Kit; como control de calidad interno para la generación, secuenciación y alineación de conglomerados se utilizó PhiX control v3 de Illumina.

Para el almacenamiento, análisis y gestión de datos en formato FASQ producto de la secuenciación, se utilizó la plataforma informática Base Space Sequence de Illumina. Para el alineamiento preliminar, determinación de las variantes por mutaciones y asignación preliminar de clados, se utilizó la plataforma bioinformática SOPHiA-DDM- V5.7.10. Se realizó alineación de las secuencias obtenidas con secuencias consenso NC_045512 en la plataforma en línea FASTA (ebi.ac.uk/tools/sss/fasta/), para generar archivos compatibles para la generación de propuestas filogenéticas como MEGA-X v10.2. Las seis secuencias obtenidas se cargaron para su análisis de variantes por clado a Nexclade beta v0.8.1; finalmente los resultados de las secuencias fueron enviados a GISAID¹³ para la determinación de clado y linaje filogenético, ya que esto permitió clasificar y validar los virus secuenciados en una plataforma ya ampliamente reconocida por la comunidad científica global¹⁴.

Consideraciones éticas

El protocolo de investigación fue sometido y aprobado por el comité local de ética del INS; los investigadores no tuvieron acceso a

la identidad de los pacientes u otros datos sensibles. Las muestras fueron manejadas por un número de referencia asignado por el Laboratorio Nacional de Salud Pública. Los datos de la secuenciación están disponibles en la plataforma GISAID bajo los siguientes registros: EPI_ISL_671974, EPI_ISL_671978, EPI_ISL_672012, EPI_ISL_672570, EPI_ISL_672572, EPI_ISL_672573.

Resultados

En las seis secuencias mediante el alineamiento con el genoma de referencia NC_045512 se observó una concordancia de casi el 100 %, confirmándose que las mismas eran positivas a SARS-CoV-2. De igual forma, las seis secuencias dieron positivo a SARS-CoV-2 en el análisis preliminar en la plataforma bioinformática SOPHiA-DDM-V5.7.10. Además, se analizaron los genes E, M, N, ORF10, ORF1a, ORF1ab, ORF3a, ORF6, ORF7a, ORF7b, ORF8, S, dentro de los cuales se identificó un total de 27 polimorfismos de un solo nucleótido (SNP), desde 10 y hasta 15 SNP en cada una de las seis secuencias. Algunos de los SNPs comunes entre las secuencias, además la mayoría de los SNP identificados, fueron sin sentido (missense), implicando un cambio de aminoácido. Las secuencias basadas en los SNP identificados se agruparon en un solo patrón variante, impulsado principalmente por SNP ubicados en la región S, NSP2, NS3 y gen S, presentando las mutaciones A23403G (S^{D614G}), C1059T (NSP2^{T265I}) y G25563T (NS3^{Q57H}). En el alineamiento con patrón de mutaciones las seis secuencias fueron clasificadas en el clado Nextstrain 20C (Figura 1-A).

Además, se utilizó GISAID para la filogenia de máxima probabilidad de secuencias muestreadas globalmente de SARS-CoV-2^{14,15}. El resultado de la comparación del patrón de mutaciones comparada con la base de secuencias y de manera particular con la secuencia de referencia hCoV19/Wuhan/WIV04/2019, las seis secuencias fueron clasificadas como clado GISAID GH, linaje pangolín B.1.2 y B.1.370, ambos linajes de EUA (Figura 1-B) coincidente con el 20C de Nextstrain.

Tabla 1. Polimorfismos de un solo nucleótido

Polimorfismo de un solo nucleótido		
N_P199L	NSP12_K59N	NSP2_T85I
N_P67S	NSP12_P323L	NSP2_S430L
N_R209K	NSP13_L256F	NSP3_H727Y
NS3_G172V	NSP13_N124S	NSP3_K384N
NS3_P258S	NSP14_N129D	NSP5_L89F
NS3_Q57H	NSP14_P203L	NSP9_S13F
NS7a_T39I	NSP15_V320L	Spike_A623S
NS8_S24L	NSP16_R216C	Spike_D614G
NS8_V62L	NSP2_T265I	Spike_S704L

Al rastrear geográficamente el virus por Nextstrain, las seis secuencias mostraron tener mayor similitud con el clado GH (20C) presente en la región norteamericana (Figura 2).

Figura 1. (A) Filogenia clados 19A, 19B, 20A, 20C según clades. *nextstrain.org/tree*; filogenia según clados GISAID GR, G, GH, O, S, L, V. Las seis secuencias en verde alineadas en clado GH (B); las seis secuencias por alineamiento por perfil de mutaciones fueron clado 20C (Nextstrain), clado GH Linage B.1.2 y B.1.370 (GISAID)

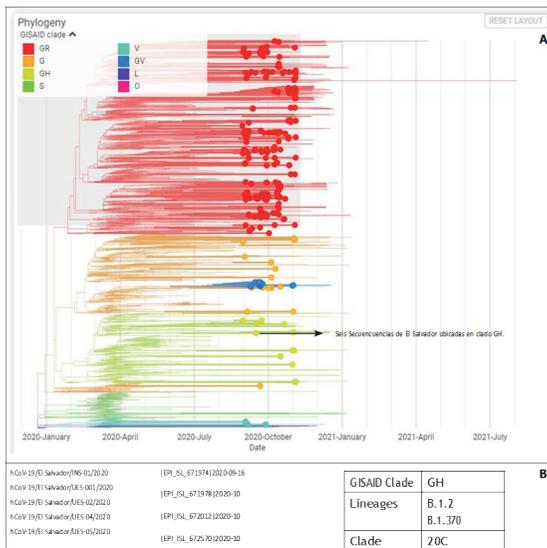
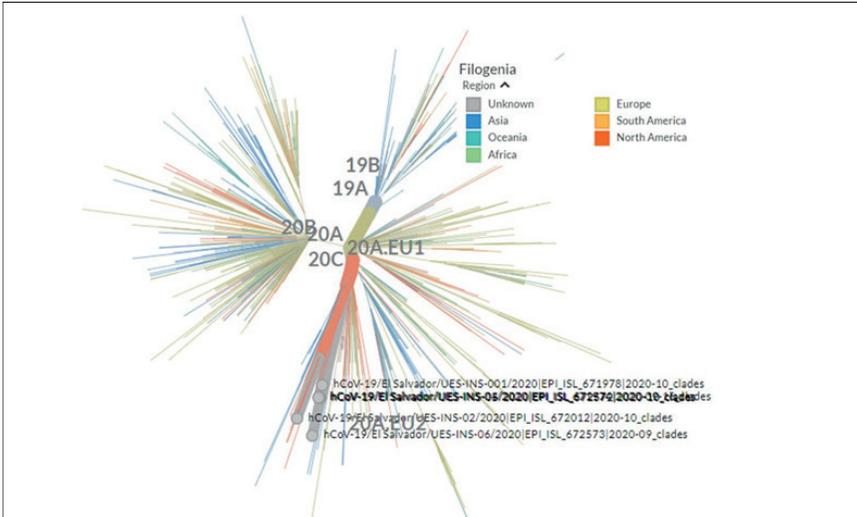


Figura 2. Filogenia por región geográfica para los clados 19A, 19B, 20A, 20C según clades.nextstrain.org/tree. Las seis secuencias en gris alineadas en clado 20C, obsérvese el gran número de secuencias norteamericanas en rojo



Discusión

Las seis secuencias fueron aceptadas y dadas de alta en GISAID. Esta información de SARS-CoV-2 apoya la calidad del proceso de secuenciación, robustece la vigilancia epidemiológica y permitirá seleccionar para el país la vacuna más representativa.

Con el análisis en la plataforma SOPHiA, ninguna de las mutaciones se asoció a mayor virulencia o patogenicidad; sin embargo, sí se encontró la mutación Spike_D614G que actualmente está siendo ampliamente estudiada por su posible relación con mayor infectividad de la proteína espícula, lo cual aumentaría su transmisibilidad⁶.

En el alineamiento con patrón de mutaciones, las seis secuencias fueron clasificadas en el clado Nextstrain 20C, caracterizado por la mutación A23403G, C1059T y G25563T. Esta alcanzó una frecuen-

cia global en torno al 20 % en abril, principalmente por secuencias de EUA^{1,19}; 20C es un clado que se derivó del clado 20A muy frecuente en EUA y que sembró grandes brotes en Europa a principios de 2020, dominando en los posteriores brotes en las Américas. Además, una de las características de este clado es la presencia de la mutación D614G, la cual se ha relacionado con aumento de la infectividad viral^{9,7 16-18}.

Los resultados demostraron que las seis secuencias fueron clasificadas como clado GISAID GH, linaje pangolín B.1.2 y B.1.370, ambos linajes con una alta prevalencia en EUA¹⁶⁻¹⁸. Es probable que la migración influya en este patrón, para lo que se requiere obtener más secuencias de todo el territorio nacional, caracterizar epidemiológicamente para conocer su probable origen, documentando la vigilancia molecular del virus y comprender mejor su evolución en el territorio salvadoreño.

Las mutaciones encontradas serán estudiadas a profundidad y se informarán en una próxima entrega. Los SNPs son candidatos fenotípica o epidemiológicamente interesantes para futuras investigaciones, en combinación con pruebas experimentales la verificación de cualquier fenotipo predicho, como es el caso de la mutación D614G.

Conclusiones

Este es el primer reporte de secuencias completas del genoma de SARS-CoV-2 en El Salvador e identifica una variante predominante en Centro y Norteamérica, específicamente el clado GH, linaje pangolín B.1.2 y B.1.370, ambos linajes con una alta prevalencia en EUA.

El análisis filogenético evidenció que estas seis muestras pertenecen al clado 20C, clado secundario de 20A, que tiene en común la variante de la mutación D614G de la glicoproteína espícula.

Los autores declaran no tener conflictos de interés en el estudio ni haber recibido patrocinio de instituciones privadas; la fuente de

financiamiento utilizada fueron recursos estatales designados para investigación.

Se reconoce el apoyo de las máximas autoridades de la Universidad de El Salvador, Instituto Nacional de Salud y Ministerio de Salud; agradecimientos especiales a la Sección de Virología del Departamento de Microbiología, Facultad de Medicina, Universidad de El Salvador y a la Sección de Virología del Laboratorio Nacional de Salud Pública de El Salvador.

Referencias bibliográficas

1. Harcourt J, Tamin A, Lu X, Kamili S, Sakthivel SK, Murray J, et al. Severe acute respiratory syndrome coronavirus 2 from patient with coronavirus disease, United States. *Emerging Infectious Diseases*. 2020;26(6):1266-1273. DOI:10.3201/EID2606.200516
2. Holshue ML, DeBolt C, Lindquist S, Lofv K.H, Wiesman J, Bruce H, et al. First Case of 2019 Novel Coronavirus in the United States. *New England Journal of Medicine*. 2020;382(10):929-936. DOI:10.1056/nejmoa2001191
3. World Health Organization. Weekly Epidemiological Update on COVID-19. World Health Organization. 2020. Disponible en: <https://www.who.int/publications/m/item/weekly-epidemiological-update---3-november-2020>
4. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, et al. A new coronavirus associated with human respiratory disease in China. *Nature*. 2020;579(7798):265-269. DOI:10.1038/s41586-020-2008-3
5. Van Dorp L, Richard D, Tan CC, Shaw L, Acman M, Balloux F. No evidence for increased transmissibility from recurrent mutations in SARS-CoV-2. *Nature Communications*. 2020. DOI:10.1101/2020.05.21.108506

6. Jackson CB, Zhang L, Farzan M, Choe H. Functional importance of the D614G mutation in the SARS-CoV-2 spike protein. *Biochemical and Biophysical Research Communications*. 2020. DOI:10.1016/j.bbrc.2020.11.026
7. Korber B, Fischer WM, Gnanakaran S, Better K, Yoon H, Theiler J, et al. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell*. 2020;182(4):812-827.e19. DOI:10.1016/j.cell.2020.06.043
8. Zhang L, Jackson CB, Mou H, Ojha A, Peng H, Quilan BD, et al. SARS-CoV-2 spike-protein D614G mutation increases virion spike density and infectivity. *Nature Communications*. 2020;1-9. DOI:10.1038/s41467-020-19808-4
9. Zhang L, Jackson C, Mou H, Ojha M, Rangarajan E.S, IZARD T, et al. The D614G mutation in the SARS-CoV-2 spike protein reduces S1 shedding and increases infectivity. *bioRxiv*. 2020. DOI:10.1101/2020.06.12.148726
10. Corman VM, Landt O, Kaiser M, Molenkamp R, Meijer A, Chu D.K.W, et al. Detection of 2019 -nCoV by RT-PCR. *Euro Surveill*. 2020;25(3):1-8. Disponible en: https://www.who.int/docs/default-source/coronaviruse/wuhan-virus-assay_v1991527e-5122341d99287a1b17c111902.pdf?sfvrsn=d381fc88_21
11. Klempt P, Brož P, Kašný M, Novotný A, Kvapilová K, Kvapil P. Performance of targeted library preparation solutions for SARS-CoV-2 whole genome analysis. *Diagnostics*. 2020;10(10):1-12. DOI:10.3390/diagnostics10100769
12. Charre C, Ginevra C, Sabatier M, Regue H, Destras G, Brun S, et al. Evaluation of NGS- based approaches for SARS-CoV-2 whole genome characterisation. *Virus Evolution*. 2020. DOI:10.1093/ve/veaa075

13. GISAID - phylodynamics of pandemic coronavirus in Latin America and the caribbean. GISAID. 2020. Fecha de consulta: 22 de diciembre de 2020. Disponible en: isaaid.org/epiflu-applications/phylodynamics/
14. Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Global Challenges*. 2017;1(1):33-46. DOI:10.1002/gch2.1018
15. Rambaut A, Holmes EC, Hill V, O'Toole Á, McCrone JT, Ruis C, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nature Microbiology*. 2020;5(11):1403-1407. DOI:10.1038/s41564-020-0770-5
16. Toyoshima Y, Nemoto K, Matsumoto S, Nakamura Y, Kiyotani K. SARS-CoV-2 genomic variations associated with mortality rate of COVID-19. *Journal of Human Genetics*. 2020;65(12):1075-1082. DOI:10.1038/s10038-020-0808-9
17. Hodcroft EB, Zuber M, Nadeau S, Candelas F, Standler T, Nehler RA, et al. Emergence and spread of a SARS-CoV-2 variant through Europe in the summer of 2020. *medRxiv.2020*. DOI: 10.1101/2020.10.25.20219063
18. Pfefferle S, Günther T, Kobbe R, Czech-Sioli M, Indenbirken D, Robitaille A, et al. SARS Coronavirus-2 variant tracing within the first Coronavirus Disease 19 clusters in northern Germany. *Clinical Microbiology and Infection*. 2020. DOI:10.1016/j.cmi.2020.09.034
19. Ministerio de Ciencia y Economía; Instituto de Salud Carlos III. Informe del Grupo de Análisis Científico de Coronavirus del ISCIII (GACC-ISCIII). Madrid. Instituto Carlos III. 2020. 8 p. Disponible en: https://fundadeps.org/wp-content/uploads/2020/04/TRANSMISIÓN-DEL-VIRUS_1.pdf