

Resumen: Análisis genómico de secuencias del virus SARS-CoV-2 de casos costarricenses, marzo 2020 – enero 2021.

Fecha: 08 de febrero de 2021

Introducción

A finales de abril del 2020, por primera vez en Costa Rica, el Inciensa culminó la secuenciación de los primeros seis genomas completos del nuevo coronavirus SARS-CoV-2, causante de la enfermedad COVID-19.

Desde entonces, se ha continuado con la secuenciación de genomas virales de pacientes de todo el país. A continuación, se presenta un resumen del análisis epidemiológico y genómico de los 233 genomas de SARS-CoV-2 obtenidos hasta el momento, de casos presentados en el país. Doscientos veintiocho de estos genomas fueron depositados en la plataforma de la “Global Initiative for Sharing al Influenza Data” (GISAID), para ser compartidos con la comunidad científica nacional e internacional.

La secuenciación de los genomas de SARS-CoV-2 permite: a) brindar información de la dinámica y la diversidad de la población viral que circula en Costa Rica, así como de la relación de las secuencias genómicas y posibles patrones en las rutas de transmisión; b) conocer los grupos genéticos (clados y linajes) circulantes en el territorio nacional; c) identificar marcadores genéticos relevantes y favorecer el seguimiento de posibles cambios estructurales del virus; d) identificar mutaciones en las regiones genómicas utilizadas para la detección molecular del virus; e) contextualizar los resultados obtenidos a nivel nacional dentro del desarrollo y evolución de la pandemia a nivel global; y f) caracterizar los virus causantes de reinfecciones.

Materiales y métodos

Los casos estudiados corresponden a pacientes que presentaron la enfermedad COVID-19 entre la primera semana de marzo de 2020 y la tercer semana de enero 2021. Las muestras fueron recolectadas, entre la semana epidemiológica (SE) 10 a la SE 15 y de la SE 23 a la SE 53 del año 2020 y las SE 1-3 (año 2021). La selección de las muestras incluyó pacientes con un ámbito de edad que comprendía desde los cuatro años hasta los 92 años. Respecto al sexo, el grupo analizado incluye 139 hombres y 92 mujeres (en dos casos, la boleta que acompaña la muestra no especifica este dato). Se procuró contar con muestras provenientes de cada provincia del país y además se incluyeron 11 muestras de defunciones asociadas a COVID-19.

El procesamiento de las muestras en el laboratorio fue realizado por dos grupos de trabajo: el Inciensa secuenció 167 muestras (CRC-001 a CRC-006 y CRC-025 a CRC-234) y la UCR-Facultad de Microbiología, junto con el Instituto Charité (Berlín) secuenciaron 18 muestras (muestras CRC-007 a CRC-024).

Se realizó un análisis descriptivo de las variables epidemiológicas asociadas a los genomas de SARS-Cov-2, que incluyó la información clínico-epidemiológica de las boletas de solicitud de

análisis que acompañaban a las muestras. Además, se desarrolló una estrategia de análisis bioinformático, en cooperación con la UCR-CIET, y se realizaron análisis bioinformáticos adicionales en Inciensa. Se ejecutó un estricto control de calidad para la reconstrucción de los genomas (ensamblaje), se llevó a cabo la identificación de variantes (mutaciones) y los genomas obtenidos fueron comparados filogenéticamente.

Resultados e interpretación

Se identificaron 20 linajes de SARS-CoV-2 (grupos genómicos relacionados) siendo los más frecuentes B.1.1.63 (33,9 %), B.1. (23,2 %), B.1.291 (18,0 %), B.1.1.119 (9,4 %) y B.1.203 (5,1 %). Estos cinco linajes representaron más del 89 % de los genomas estudiados. Durante las primeras cinco semanas se identificaron linajes asociados a pacientes que tenían antecedentes de viaje a Europa (Francia y España), Norte América (EEUU y México) y Sur América (Colombia). Posterior a la SE 24 se identificó el linaje B.1.1, el cual circuló activamente en zona norte del país (Peñas Blancas, el cantón de La Cruz y en Tablillas, en Los Chiles). De Junio 2020 en adelante, este linaje fue identificado en seis de las siete provincias del territorio nacional. Al analizar los linajes por grupo etario, se observó una distribución homogénea de estos en toda la muestra. Al estudiar los linajes por sexo, se observó una distribución similar entre hombres y mujeres, predominando en ambos el linaje B.1.1.63. La distribución de frecuencias de los genomas de SARS-CoV-2 según la clasificación por clado, utilizada en GISAID, fue: 48,9 % (GR), 27,5 % (GH), 21 % (G) y 2,6 % (S) y se muestra en la figura 1.

En cuanto a las mutaciones y la filogenia del virus, se detectó que cada genoma poseía al menos entre 3 y 33 mutaciones, al compararlo con el genoma de referencia Wuhan-Hu-1. Siendo esto consistente con el tiempo transcurrido desde la introducción del virus en la población humana. La variante D614G de la proteína S (espícula viral) predominó en los genomas estudiados (99,1 % de los casos). La variante L84S, utilizada para identificar dos linajes L y S en el ORF8, fue encontrada en seis genomas (2,6 %). Cabe resaltar que los cuatro nuevos genomas pertenecientes al clado S se detectaron entre las SE 51 – 53 (2020) y SE 1-2 (2021), proviniendo dos de ellos de la zona sur del país.

Respecto a las variantes que internacionalmente se consideran de interés en salud pública, no se ha detectado hasta el momento la variante VOC202012/01 (linaje B.1.1.7), ni la variante 501Y.V2 (B.1.351), ni la variante E848K (B.1.1.28) descritas originalmente en el Reino Unido, Sudáfrica y Brasil respectivamente. Sin embargo, dos genomas fueron similares a la variante GH/452R.V1 (B.1.429) y fueron notificadas al Centro Nacional de Enlace del Reglamento Sanitario Internacional (CNE/RSI) del Ministerio de Salud.

Además, se detectó la mutación T1117I en la proteína de la espícula (S) en 84 de los genomas virales (36 % de los casos); los cuales formaron un clúster monofilético claramente distinguible (i.e. origen común) en el árbol filogenético. Esta mutación se acompaña de otras dos, la H145Y y L291L en la nucleocápside viral. La mutación T1117I ha sido escasamente reportada en otras latitudes. Se ha descrito por ahora en al menos 19 países, pero de los 320 genomas disponibles con dicha característica en la base de datos GISAID, más del 26 % (84) provienen de Costa Rica. Modelos estructurales de la proteína S de la variante T1117I predicen que: debido a la

ubicación de la mutación, no se esperan cambios en la interacción de la proteína S con el receptor celular. Y tampoco se predice que acarree cambios en la región inmunodominante de la proteína S que media la respuesta inmune por células B. Se requieren más estudios para determinar las posibles implicaciones de esta variante en la biología del SARS-CoV-2 y la propagación en Costa Rica.

El árbol filogenético evidenció una clara separación de muestras de acuerdo al grupo genético al que pertenecen. Los casos de defunciones se distribuyeron indistintamente en los diversos clados identificados y no se les puede asociar con una menor o mayor mortalidad (Figura 1). Estos datos son consistentes con el hecho de que en la literatura no existe evidencia contundente que relacione la presentación clínica de la COVID-19, con las variantes moleculares de los diferentes genomas descritos a nivel global. También se identificaron dos genomas de SARS-CoV-2 similares a la variante GH/452R.V1 y resultaron genéticamente muy relacionados entre sí formando un clúster.

Ninguna de las variantes identificadas en el estudio se localizó en las regiones de unión de “primers” y sondas de los genes E y RdRP, que se usan en el ensayo de RT-PCR para la detección viral realizada de acuerdo con el protocolo del Instituto de virología Charité (Berlín), a partir de muestras clínicas. También se determinó al menos dos de las muestras analizadas poseían variantes estructurales en el dominio de unión al receptor celular (RBD) de la proteína de la espícula viral (L452R), lo cual indica que se presentaron cambios en la porción de dicha proteína que media la interacción virus-receptor al momento de la infección y deben ser motivo de monitoreo. A la fecha el impacto de esta mutación no ha sido determinado y las investigaciones aún están en desarrollo.

Conclusiones

Los genomas de los virus SARS-CoV-2 que circularon en Costa Rica entre marzo 2020 y enero del 2021 presentaron un patrón similar a lo reportado alrededor del mundo, incluyendo la distribución de casos con las variantes D614G de la proteína S (espícula) y la L84S del ORF8. No se identificaron mutaciones en las regiones de los genes E y RdRP utilizados para el diagnóstico viral, según el protocolo del Instituto de virología Charité (Berlín). Los linajes más comunes en Costa Rica, durante el período de estudio, fueron B.1.1.63, B.1 y B.1.291. Los casos relacionados con defunciones no presentaron un patrón particular de asociación por clado. Se describió el aumento acumulado en la variante T1117I en la espícula viral (36 % de los casos) cuyo significado biológico y epidemiológico aún requiere más estudios. Se identificaron además dos genomas similares a la variante GH/452R.V1 los cuales fueron notificados al CNE/RSI del Min. Salud.

Dada la importancia de estos hallazgos, el Inciensa continuará con la vigilancia genómica del SARS-CoV-2 en nuestro país, para identificar nuevas variantes genéticas o estructurales de importancia epidemiológica para el diagnóstico, prevención y control de la enfermedad.

Elaborado por:
Dr. Francisco Duarte-Martínez, Inciensa
Dr. José Arturo Molina-Mora, Facultad de Microbiología-UCR
Dra. Estela Cordero-Laurent, Inciensa
Red Nacional de Laboratorios

Tree scale: 0.001

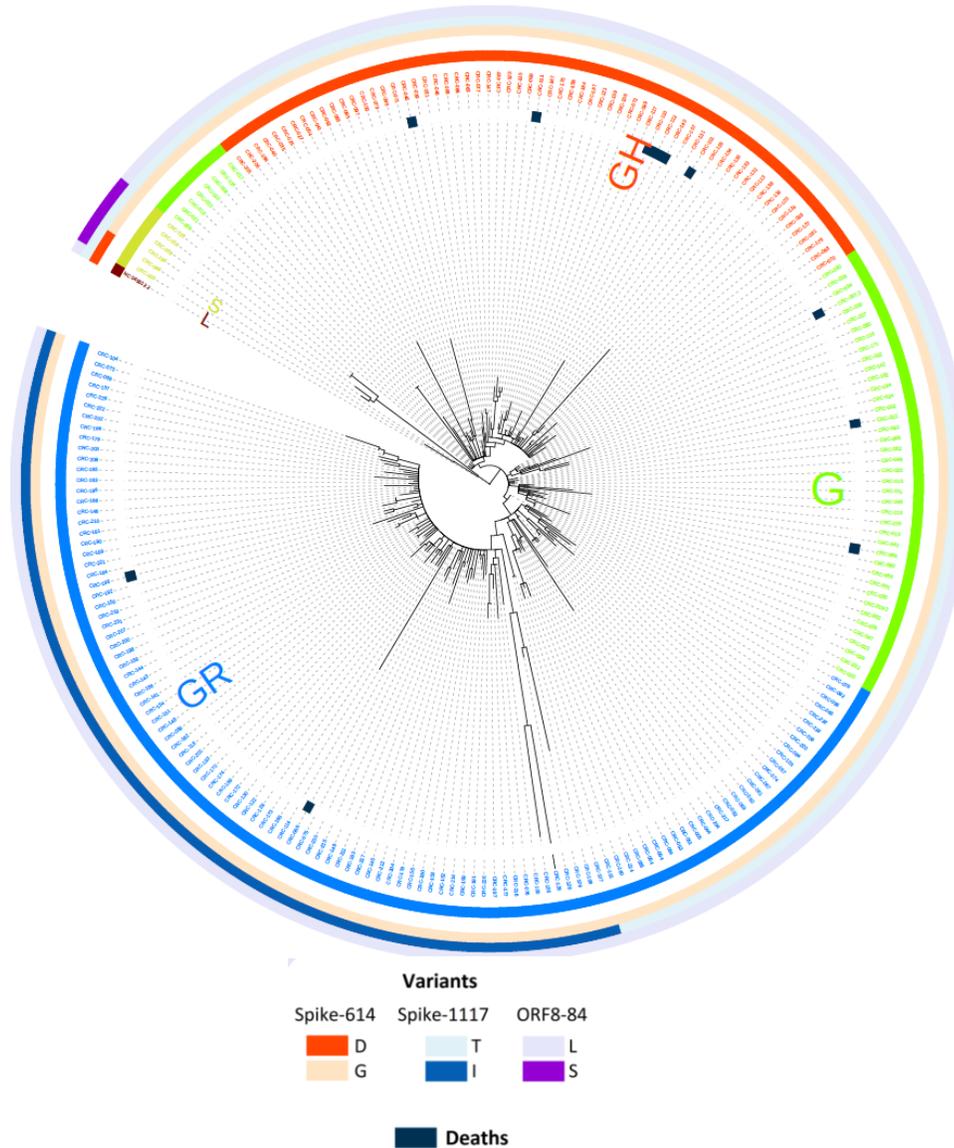


Figura 1. Relaciones filogenéticas de 233 secuencias de SARS-CoV-2 en casos de Costa Rica. Se identifican los clados dados por GISAID, diversas variantes identificadas y las defunciones. Información adicional puede ser consultada en <https://microreact.org/project/97WYgMLAtj9L3TFFhhbxhq>