Análisis genómico de secuencias del virus SARS-CoV-2 de casos costarricenses, marzo - julio 2020.

Fecha: 27 de agosto de 2019

A finales de abril del 2020 por primera vez en Costa Rica, y a pocas semanas de estar circulando en el país, el área de genómica y biología molecular del Inciensa culminó la secuenciación de los primeros seis genomas completos del nuevo coronavirus SARS-CoV-2 causante de la enfermedad COVID-19. Los resultados fueron depositados en la plataforma "Global Initiative on Sharing All Influenza Data" (GISAD), utilizada para compartir este tipo de información de manera global, logrando con esto poner a disposición de la comunidad científica mundial los primeros datos de Costa Rica. En vista de que el Inciensa es el responsable de la vigilancia epidemiológica basada en laboratorio, se continuó con la secuenciación de otros genomas del virus de pacientes de todo el país. Así mismo, la Universidad de Costa Rica (UCR) contribuyó proporcionando 18 genomas obtenidos de muestras facilitadas por varios hospitales nacionales. Este trabajo presenta el análisis epidemiológico y genómico de 70 genomas virales en Costa Rica. Los casos estudiados correspondían a pacientes que presentaron la enfermedad COVID-19 entre los meses de marzo y julio. Se realizó un análisis descriptivo de las variables epidemiológicas asociadas a los genomas y se desarrolló una estrategia de análisis bioinformático en cooperación con la UCR. Se realizó un estricto control de calidad para la reconstrucción de los genomas (ensamblaje), se ejecutó la identificación de variantes (mutaciones) y se comparó filogenéticamente los genomas obtenidos.

Los linajes virales más frecuentes fueron: B.1, B.1.5 y B.1.1 respectivamente y los mismos se distribuyeron de manera homogénea en cuanto al sexo y edad. Los linajes B.1 y B.1.1 presentaron amplia distribución en el territorio nacional y B.1.5 se presentó principalmente en las provincias de San José y Alajuela. Las secuencias genómicas obtenidas durante las semanas epidemiológicas 23-27 formaron un agrupamiento (clúster) evidenciando la transmisión activa del clado B.1.1 en la zona norte del país. Los casos relacionados con defunciones por COVID-19 no generaron un patrón de asociación particular por clado. Se detectó que cada genoma poseía entre 4 y 11 mutaciones comparando con el genoma de referencia Wuhan-Hu-1. La variante D614G de la proteína S predominó en los genomas estudiados. No se identificaron mutaciones en las regiones de los genes E y RdRP utilizadas para el diagnóstico viral según el protocolo de Corman et al. Tampoco se identificaron mutaciones en el dominio de unión al receptor de la espícula viral. Lo anterior resalta la necesidad de que Inciensa continúe con la vigilancia basada en laboratorio de las secuencias genómicas de SARS-CoV-2 que circulan en nuestro país.

Cita sugerida: Duarte-Martínez Francisco[1]* y Molina-Mora José Arturo[2]**, Cordero-Laurent Estela[1], Godínez-Rojas Adriana[1], Calderón-Osorno Melany[1], Corrales-Aguilar Eugenia[2], Brenes-Porras Hebleen[1], Soto-Garita Claudio[1], et al. 2020. **Análisis genómico de secuencias del virus SARS-CoV-2 de casos costarricenses.** Instituto Costarricense de Investigación y Enseñanza en Nutrición y Salud (Inciensa) y Universidad de Costa Rica (UCR), Costa Rica.

[1] Instituto Costarricense de Investigación y Enseñanza en Nutrición y Salud (Inciensa), Tres Ríos.
[2] Centro de Investigación en Enfermedades Tropicales (CIET) y Facultad de Microbiología, Universidad de Costa Rica, San José.
* fduarte@inciensa.sa.cr** jose.molinamora@ucr.ac.cr,

Introducción

COVID-19 (COronaVIrus Disease 2019) es una enfermedad infecciosa causada por el virus SARS-CoV-2, fue descrita por primera vez a finales de diciembre de 2019, en un brote de neumonía atípica en la ciudad de Wuhan, provincia de Hubei, China (Yin, 2020). Posee una presentación clínica poco específica asociada al tracto respiratorio, fiebre, diarrea y casos más graves con dificultad respiratoria, sepsis y shock séptico (Huang et al., 2020). Para finales de agosto 2020, se han identificado más de 23.5 millones de casos alrededor del mundo y 81 000 muertes (OMS, 25 de agosto de 2020, 15:49). Para Costa Rica, más de 36 000 casos han sido reportados (Ministerio de Salud, 27 de agosto 2020), incluyendo más de 380 muertes (el primer caso se reportó el 6 marzo 2020).

El primer genoma completo de SARS-CoV-2 fue publicado el 10 de enero 2020 (Wu et al., 2020). A finales de abril del 2020 por primera vez en Costa Rica, y a pocas semanas de estar circulando en el país, el área de genómica y biología molecular del Inciensa culminó la secuenciación de los primeros seis genomas del nuevo coronavirus SARS-CoV-2. Los resultados fueron depositados en la plataforma "Global Initiative on Sharing All Influenza Data" (GISAD), poniendo a disposición de la comunidad científica mundial los primeros datos de Costa Rica (Inciensa & Ministerio de Salud Pública, 2020). El genoma del SARS-CoV-2 está constituido por una sola cadena de ácido ribonucleico (ARN) de 29 903 bases de longitud. Diversos esquemas de clasificación utilizando marcadores genéticos específicos han sido propuestos para darle seguimiento a la evolución del virus en la población humana. El esquema de clasificación por clados de GISAID y el esquema de clasificación por linajes (Rambaut et al., 2020) han sido ampliamente utilizados para el monitoreo de la actual pandemia.

Por las particularidades inherentes a las tecnologías de secuenciación de ácidos nucleicos existe una complejidad de procesamiento de datos que requiere de un abordaje específico mediado por análisis bioinformáticos. Esta situación imposibilita la aplicación de un análisis universal y estandarizado. Así, el procesamiento de datos de secuencias de ADN es dependiente de algoritmos y modelos matemáticos de alto rendimiento contemplados en la bioinformática (Robasky, Lewis, & Church, 2014; Zhao et al., 2016). En este contexto, en colaboración UCR-Inciensa: se desarrolló un "Protocolo bioinformático y de inteligencia artificial para el apoyo de la vigilancia epidemiológica basada en laboratorio del virus SARS-COV-2 mediante la identificación de patrones genómicos y clínico-demográficos en Costa Rica".

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 la relación de las secuencias genómicas y posibles patrones en las rutas de transmisión; b) conocer los clados y linajes circulantes en el territorio nacional; c) identificar marcadores genéticos relevantes y favorecer el seguimiento de posibles cambios estructurales del virus; d) brindar la capacidad de 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.

Metodología

El presente informe contempla el análisis epidemiológico y genómico de 70 secuencias del virus SARS-CoV-2 de casos de COVID-19 obtenidas entre marzo y julio del año 2020. Las muestras fueron recolectadas en dos momentos epidemiológicos diferentes, entre la semana epidemiológica (SE) 10 a la SE 15 y de la SE 23 a la SE 27. Se hizo un análisis descriptivo de los resultados de la vigilancia de laboratorio de SARS-CoV-2, incluyendo la información clínico-epidemiológica de los pacientes obtenida de las boletas de solicitud de análisis que acompañaban las muestras. La selección de las muestras provenía de pacientes con un ámbito de edad que comprendía desde los cuatro años hasta los 89 años. Cuarenta y seis pacientes masculinos y 24 de pacientes femeninos. Se procuró contar con al menos una muestra proveniente de cada provincia del territorio nacional y se incluyeron cuatro muestras de defunciones asociadas a COVID-19.

El procesamiento de las muestras en el laboratorio fue realizado por dos grupos de trabajo:

- Inciensa: 52 muestras (CRC-001 a CRC-006 y CRC-025 a CRC-070).
- UCR - Instituto Charité (Berlín): 18 muestras (muestras CRC-007 a CRC-024).

En el Inciensa, la secuenciación del genoma completo (WGS, whole genome sequencing) se realizó utilizando la tecnología de secuenciación por síntesis en la plataforma MiSeq de la marca Illumina. Las muestras seleccionadas debían poseer un CT no mayor a 25 para el gen E de SARS-COV-2 en la prueba de diagnóstico de RT-PCR. Cincuenta y dos muestras referidas por los laboratorios de la Caja Costarricense del Seguro Social (CCSS), hospitales privados y por la Morgue Judicial, fueron analizadas utilizando el protocolo recomendado por el Instituto Nacional de Salud Pública de Chile (Castillo et al., 2020). Dieciocho genomas adicionales que fueron secuenciados en el Instituto de Virología del Charité de Berlín en asociación con el Centro de Investigaciones en Enfermedades Tropicales (CIET-UCR) y fueron proporcionados para complementar la vigilancia.

Un protocolo bioinformático y de inteligencia artificial para el apoyo de la vigilancia epidemiológica basada en laboratorio fue estandarizado y aplicado en colaboración con la Universidad de Costa Rica. La metodología general implementada a nivel bioinformático se resume en la Figura 1. Utilizando los datos de secuenciación, para cada caso se procedió a la reconstrucción del genoma (ensamblaje). Se usó la estrategia de evaluación con el criterio 3C (contigüidad, completitud y correctitud) para evaluar la calidad de los genomas ensamblados (Molina-Mora, Campos-Sánchez, Rodríguez, Shi, & García, 2020). Posteriormente se realizó un análisis comparativo que incluyó el estudio de las regiones ensambladas por variabilidad, el análisis de llamado de variantes (para identificar mutaciones) y el establecimiento de relaciones por similitud de secuencia (filogenia). Para esta última se utilizó el criterio de información bayesiano (BIC) y el modelo: HKY+F+I. Todos los protocolos usados a nivel bioinformático fueron estandarizados con los procedimientos y datos de trabajo similares para SARS-CoV-2.

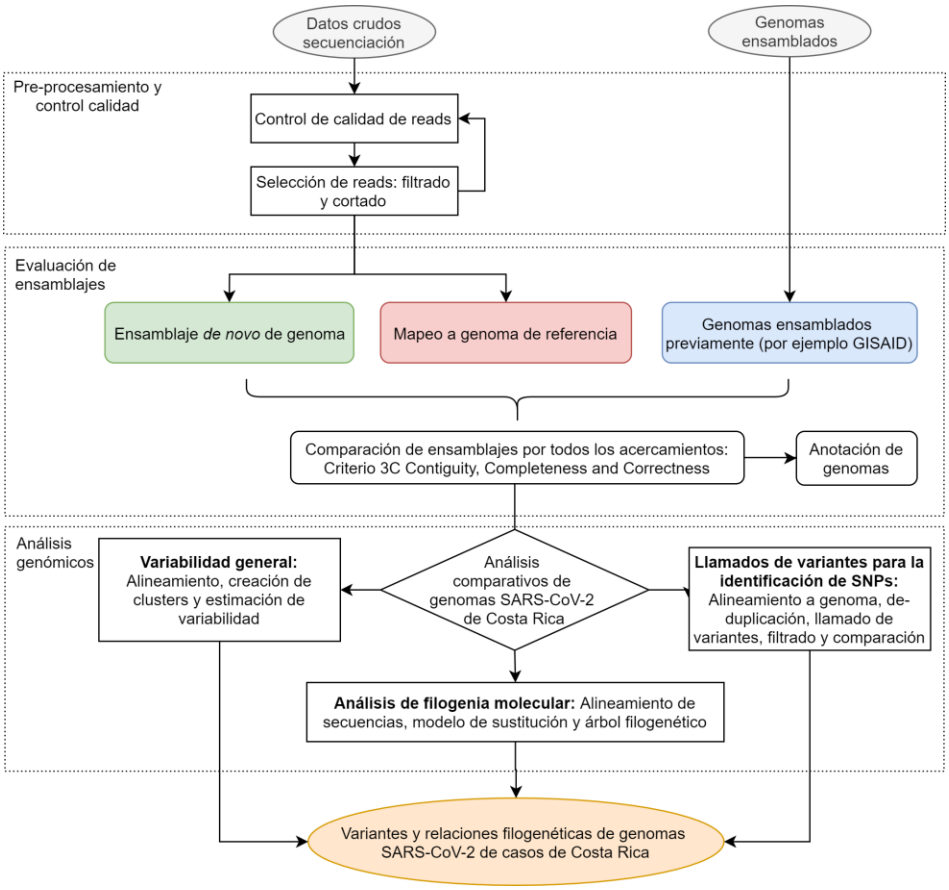


Figura 1. Protocolo bioinformático para el análisis de secuencias del virus SARS-CoV-2.

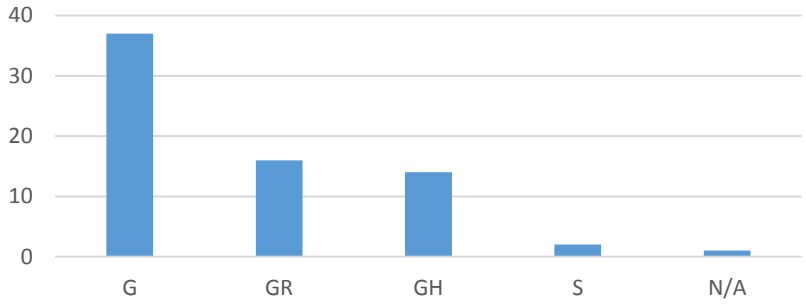
Análisis bioinformáticos adicionales fueron realizados en Inciensa utilizando el software BioNumerics v7.6 de Applied Maths con el plugin SARSCoV2 v0.14, para la detección de SNPs y la extracción de los productos de PCR para los genes E y RdRP de SARS-CoV-2. Todas las secuencias genómicas fueron depositadas en la plataforma de la iniciativa GISAID (www.gisaid.org), donde se catalogaron en clados y linajes (grupos con elevada similitud genética) según los esquemas de clasificación de GISAID y Pangolin (Rambaut el al., 2020). Además, en este repositorio las secuencias genómicas quedaron a disposición de la comunidad científica nacional e internacional.

Resultados y discusión

Descripción epidemiológica

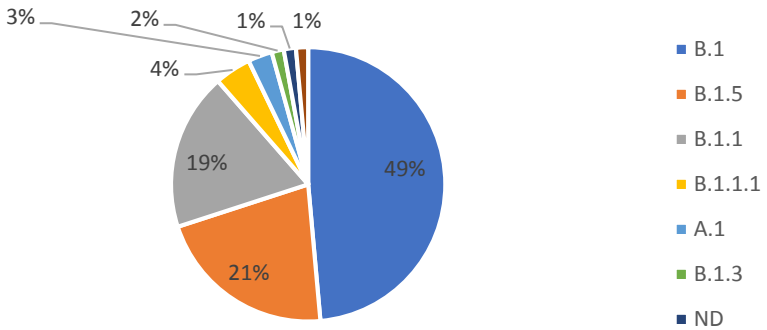
Para el período del estudio (de la SE 10 a 15 y de la SE 23 a 27) se identificaron cuatro clados (grupos genómicos) principales utilizando el esquema de clasificación GISAD. En orden de frecuencia los clados fueron G (37/70), GH (16/70), GR (14/70) y S (2/70). Para uno de los genomas no fue posible la asignación del clado correspondiente (Figura 2). Desde fechas tempranas de la pandemia (SE 10 a SE 12) fue posible documentar la introducción de los cuatro clados antes mencionados.

Figura 2. Distribución de frecuencia de los cados GISAID identificados para los genomas de SARS-CoV-2 de Costa Rica (n =70)



Es importante resaltar que el esquema de clasificación por linaje (clasificación Pangolin) suele utilizarse con mayor frecuencia ya que presenta un mayor poder de resolución que los clados de GISAID. Por lo anterior se describirá el comportamiento epidemiológico del SARS-CoV-2 utilizando la clasificación por linaje. A continuación, se presenta la distribución de frecuencias por linajes para el periodo de estudio (Figura 3).

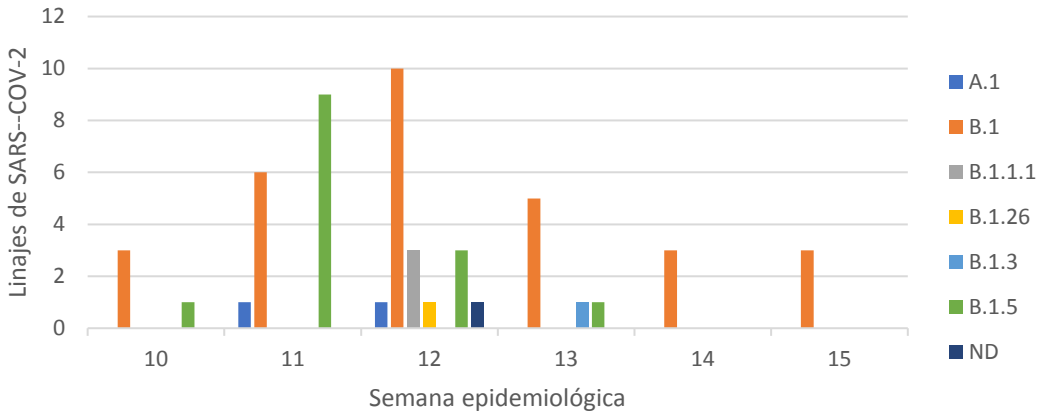
Figura 3. Linajes de SARS-CoV-2 identificados en Costa Rica, marzo - julio 2020
N: 70



Se identificaron 7 linajes diferentes y a uno de los genomas estudiados no fue posible asignarle el linaje correspondiente.

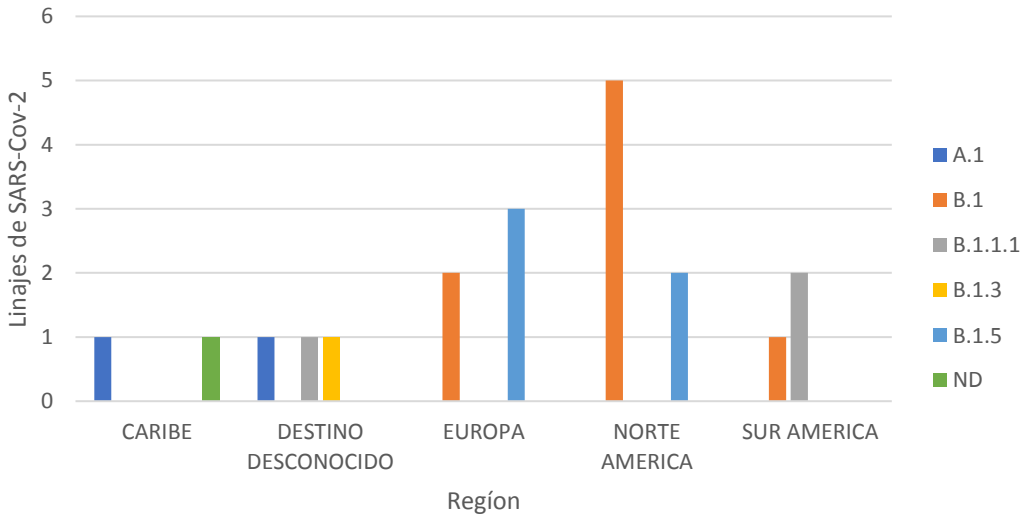
Entre la SE 10 y la SE 15 se identificaron seis linajes, siendo B.1 y B. 1.5. los más frecuentes (Figura 4).

Figura 4. Linajes de SARS-CoV-2 circulantes de la semana epidemiológica 10 a 15, Costa Rica, 2020.
N: 52



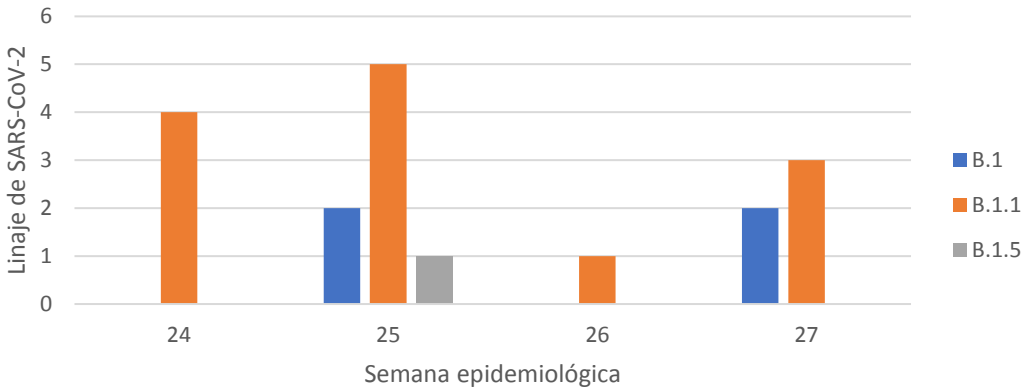
La diversidad antes expuesta podría explicarse principalmente debido a la introducción de los mismos por personas con antecedente de viaje a diversas regiones del mundo. En la Figura 5 se pueden observar los linajes asociados a casos de COVID-19 con antecedente de viaje. Los países visitados por los pacientes incluían E.E.U.U, México, Puerto Rico, Argentina, Perú Colombia, España y Francia.

Figura 5. Linajes de SARS-CoV-2 identificados en pacientes con antecedente de viaje al exterior



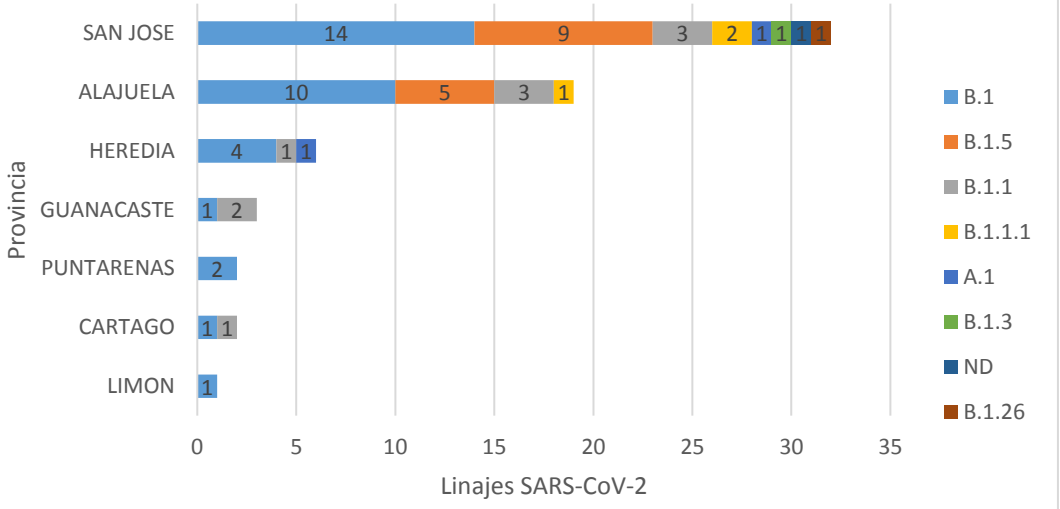
Entre la SE 24 y la SE 27 la diversidad de linajes fue más reducida y solo se detectaron B.1, B.1.1 y B.1.5 (Figura 6), sugiriendo que estos fueron los que lograron diseminarse entre la población para ese periodo de tiempo.

Figura 6. Linajes de SARS-CoV-2 circulantes de la semana epidemiológica 24 a 27, Costa Rica 2020
N: 18



Para las muestras con información de provincia de residencia disponible, la distribución geográfica de los linajes de SARS-CoV-2 puede observase en la figura 7.

Figura 7. Distribución de linajes de SARS-CoV-2 por provincia
N: 65

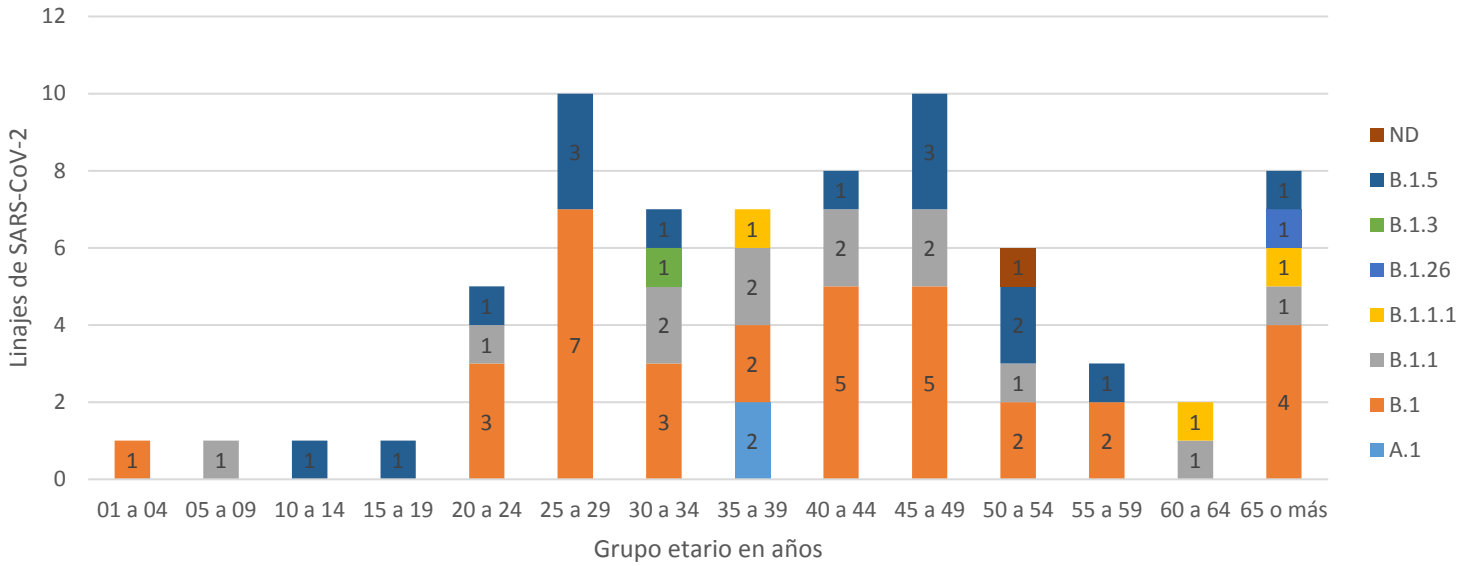


Los linajes B.1 y B.1.1 se documentaron a lo largo territorio nacional. Los mismos fueron identificados en siete y cinco provincias respectivamente. El linaje B.1.5 se encontró en muestras provenientes de San José y Alajuela principalmente. La provincia de San José, al contar con mayor representación en la muestra mostró la diversidad de linajes más elevada.

En el caso particular de la zona norte del país, se analizaron siete muestras (CRC-053, 54, 55, 56, 61, 65, 66) referidas por las A.S. de La Fortuna, Los Chiles y La Cruz, derivadas del proceso de tamizaje comunitario en Peñas Blancas y en Tablillas durante las SE 24 a SE 26. Todos los genomas analizados pertenecieron al linaje B.1.1 indicando que son genéticamente similares y que probablemente se encuentran relacionadas al mismo evento de transmisión (Figura 13).

En el estudio de los linajes por grupo etario se observó una distribución homogénea en toda la muestra. Sin embargo, cabe recordar que se trata de una muestra dirigida que no necesariamente representa a la población general. Se debe recalcar que para esta ocasión se cuenta con pocas muestras de pacientes menores de edad y este aspecto debe profundizarse en un futuro (Figura 8)

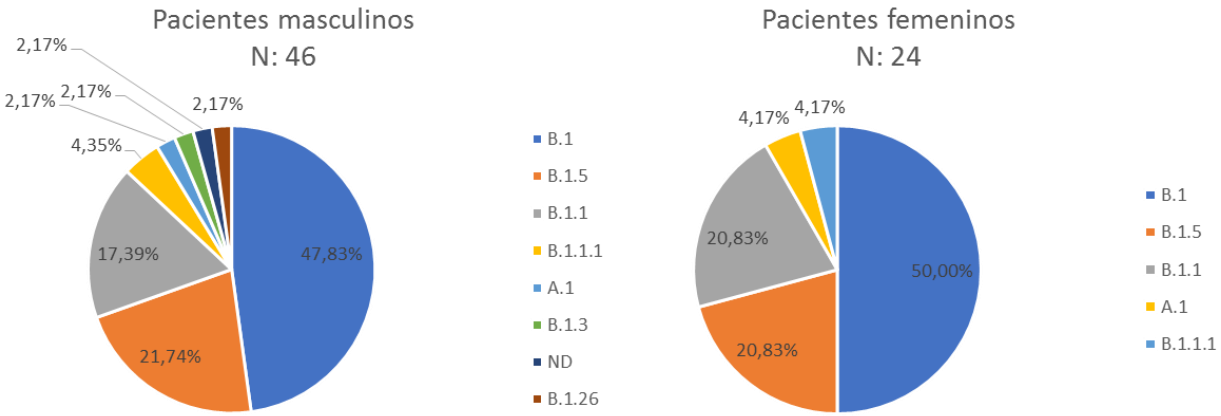
Figura 8. Distribución de linajes de SARS-CoV-2 por grupo etario en Costa Rica, marzo-junio 2020
N: 70



Al analizar los linajes de SARS-CoV-2 por sexo se observó una distribución similar entre hombres y mujeres, prevaleciendo para ambos grupos los linajes B.1, B.1.5 y B.1.1 (Figura 9).

Con respecto a las cuatro defunciones asociadas a COVID-19 incluidas en este estudio se determinó que dos de ellas pertenecía al clado G y las otras dos a los clados GH y GR o a los linajes B.1 y B.1.1 (Figura 13). Un análisis más detallado de las características moleculares (mutaciones y filogenia) de los genomas caracterizados pueden encontrarse en el apartado de "Análisis bioinformático y genómico de las secuencias de SARS-CoV-2."

Figura 9. Distribución de linajes de SARS-CoV-2 por sexo



Análisis bioinformático y genómico de las secuencias de SARS-CoV-2

Tres grandes apartados fueron abordados para el análisis bioinformático y genómico de las secuencias de SARS-CoV-2: control de calidad, análisis de variantes y relaciones filogenéticas de los genomas. Primero, la estrategia incluyó un estricto control de calidad de datos durante la reconstrucción y evaluación de la secuencia del genoma (ensamblaje), evidenciando un ensamblaje completo para la gran mayoría de casos (métrica llamada completitud, Figura 10).

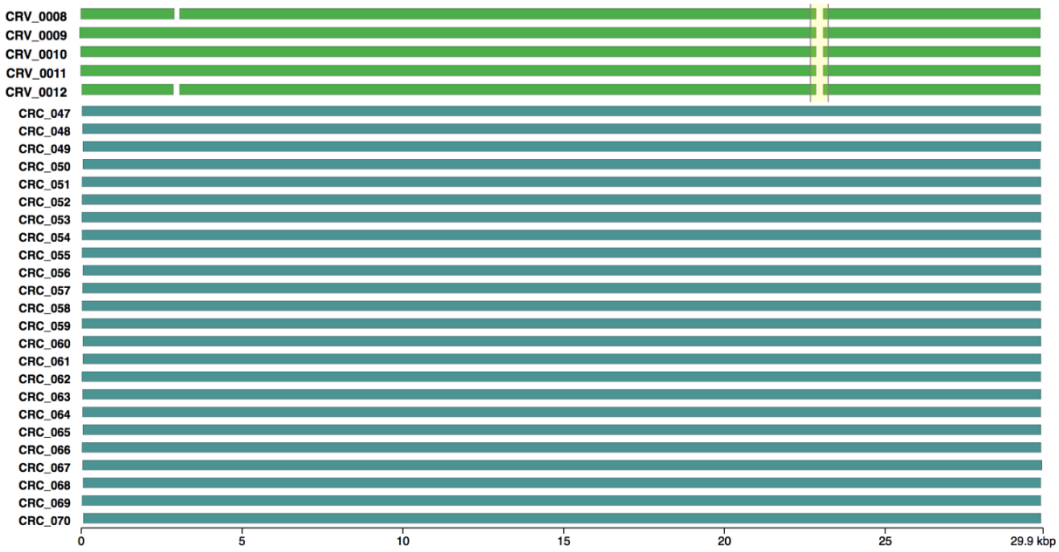


Figura 10. Evaluación de la completitud de genomas de SARS-CoV-2 de Costa Rica.

En algunos casos se presentaron problemas de ensamblaje en ciertas regiones genómicas específicas. En la Figura 2 los espacios en blanco en las muestras de CRV-0008 a CRV-0012 representaron regiones del genoma que no se pudieron obtener. Sin embargo, esos vacíos no lograron comprometer los resultados. Para estos casos se pudo verificar que las regiones bien ensambladas lograron una resolución comparable a los genomas completamente construidos. En general se logró reconstruir la arquitectura completa del virus para la mayoría de casos (Figura 11).

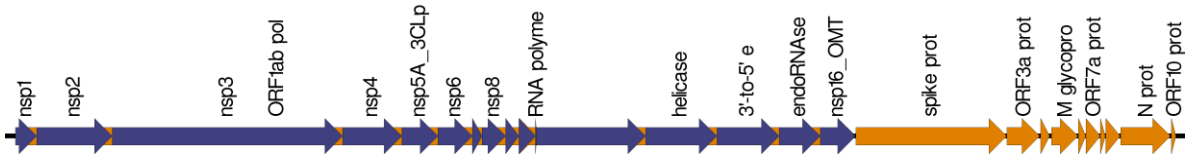


Figura 11. Arquitectura genómica del SARS-CoV-2

En segunda instancia se llevó a cabo el análisis de variantes. Se detectó que cada genoma estudiado poseía entre 4 y 11 mutaciones al ser comparado con el genoma de referencia Wuhan-

hu-1, evidenciando una elevada similitud entre ellos. Estos resultados son consistentes con el corto periodo de tiempo transcurrido desde la introducción del virus en la población humana. El ejemplo para el caso CRC-007 se muestra en la Tabla 1, donde se visualizan las 7 variantes identificadas y su ubicación el genoma.

Tabla 1. Ejemplo del análisis de variantes para el caso CRC-007

CHROM	POS	TYPE	REF	ALT	EVIDENCE	FTYPE	STRAND	NT_POS	AA_POS	EFFECT	LOCUS_TAG	GENE	PRODUCT
NC_045512	241	snp	C	T	T:18 C:0	5'UTR	+			intergenic_region n.241C>T			
NC_045512	3037	snp	C	T	T:20 C:0	mat_peptide	+	2772/21290	924/7095	synonymous_variant c.2772C>T p.Phe924Phe	GU280_gp01	ORF1ab	nsp3
NC_045512	8365	snp	T	C	C:20 T:0	mat_peptide	+	8100/21290	2700/7095	synonymous_variant c.8100T>C p.Asn2700Asn	GU280_gp01	ORF1ab	nsp3
NC_045512	8492	snp	A	G	G:20 A:0	mat_peptide	+	8227/21290	2743/7095	missense_variant c.8227A>G p.Thr2743Ala	GU280_gp01	ORF1ab	nsp3
NC_045512	14408	snp	C	T	T:20 C:0	mat_peptide	+	14143/21290	4715/7095	synonymous_variant c.14143C>T p.Leu4715Leu	GU280_gp01	ORF1ab	RNA-dependent RNA polymerase
NC_045512	23403	snp	A	G	G:20 A:0	CDS	+	1841/3822	614/1273	missense_variant c.1841A>G p.Asp614Gly	GU280_gp02	S	surface glycoprotein
NC_045512	26143	snp	G	A	A:20 G:0	CDS	+	751/828	251/275	missense_variant c.751G>A p.Gly251Ser	GU280_gp03	ORF3a	ORF3a protein

Variantes genómicas relevantes reportadas en la literatura como la mutación Asp614Gly (D614G) de la espícula viral (Korber et al., 2020) son predominantes en los casos de Costa Rica (97% de los genomas), siendo este un patrón similar al resto del mundo (ver Tabla 1, Figura 12 y Figura 13). La variante en el ORF8 Leu84Ser (L84S) sugerida como marcador para identificar dos linajes L y S (Tang et al., 2020) fue encontrada solamente en dos genomas (CRC-003 y CRC-049), lo cual es consistente con la distribución reportada en otras latitudes.

Ninguna de las variantes identificadas en este 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 de acuerdo al protocolo del Hospital Universitario Charité (Corman et al., 2020), a partir de muestras clínicas (dato no mostrado).

Por otra parte, se pudo evidenciar que para las secuencias costarricenses depositadas en GISAID no se documentaron las variantes estructurales (Spike_N439K, T481I, V483A, E484E ni G476S) en el dominio de unión al receptor (RBD) de la proteína de la espícula viral. Lo anterior sugiere que no se han presentado cambios en la porción de la espícula viral que media la interacción virus-receptor al momento de la interacción patógeno – célula blanco.

Las variantes identificadas también se utilizaron para comparar filogenéticamente todos los genomas obtenidos. Esta comparación evidenció una clara separación de acuerdo al clado o linaje (grupo) al que pertenecen (Figura 13). Como se mencionó con anterioridad los clados o linajes presentaron amplia distribución geográfica a lo largo del territorio nacional. Algunos grupos parecen estar enriquecidos por tiempo (semana epidemiológica): las secuencias de las SE 23-27 forman un grupo que se diferencia de las SE 10-15. Ese grupo claramente diferenciado incluye las muestras derivadas del tamizaje comunitario de la Zona Norte (Peñas Blancas y Tablillas) y los datos filogenéticos confirman la transmisión activa del linaje B.1.1 en la zona. Este es un patrón que requiere ser vigilado constantemente para conocer la evolución de los linajes y determinar la

introducción o aparición de un nuevo clado o linaje, o la dispersión y predominancia de alguno de estos en la población.

Como se observa en la Figura 13 solamente dos muestras CRC-003 y CRC-049 se separaron completamente del grupo de genomas, teniendo un perfil específico en el clado S, con versiones de S-614 y ORF8-84 que contrastan con el resto de muestras. Además, se observó que los casos de pacientes fallecidos estaban distribuidos indistintamente en los clados identificados, es decir, que no existe en la muestra estudiada un clado al que se le pueda asociar con una menor o mayor mortalidad. Lo anterior es consistente con que a la fecha no existe evidencia contundente que asocie la presentación clínica de la COVID-19 con las diferencias moleculares de los distintos clados descritos a nivel global (Mercatelli D & Giorgi FM; 2020).

La comparación con genomas de otras latitudes también fue capaz de separar los grupos de acuerdo al clado (Figura 14). Debido a que los análisis filogenéticos dependen de las secuencias usadas y disponibles en las bases de datos, y teniendo en cuenta que el virus presenta una tasa de mutación baja (Mercatelli D & Giorgi FM; 2020), no debe interpretarse que la aparición de un caso de Costa Rica al lado de un genoma de otro país explique su origen. Por ejemplo, si un caso de Costa Rica y un caso de España están contiguos en el árbol filogenético, no podemos asegurar que el caso de Costa Rica fue por un contacto de España. Si se agregan más genomas al árbol su forma puede cambiar y podría ser que por ejemplo ese mismo caso de Costa Rica se agrupe en el nuevo árbol con un genoma de Asia. Es debido a lo anterior que la interpretación de la filogenia siempre debe llevarse a cabo a la luz de la información epidemiológica, datos como el nexo epidemiológico (contactos), antecedente de viajes, fecha de inicio de síntomas, etc., son fundamentales para la contextualización de los resultados.

Para la interpretación de los resultados es importante tener en consideración que no todos los países están realizando de manera sistemática la secuenciación y carga de genomas de SARS-CoV-2 a las bases de datos internacionales, por lo que las mismas presentan un sesgo inherente en cuanto a su representatividad. Por otra parte, cabe recordar que los 70 genomas aquí analizados responden a una selección dirigida de muestras (ya sea por la disponibilidad del material biológico para analizar o por importancia epidemiológica de la muestra) y por consiguiente incluye también un sesgo muestral.

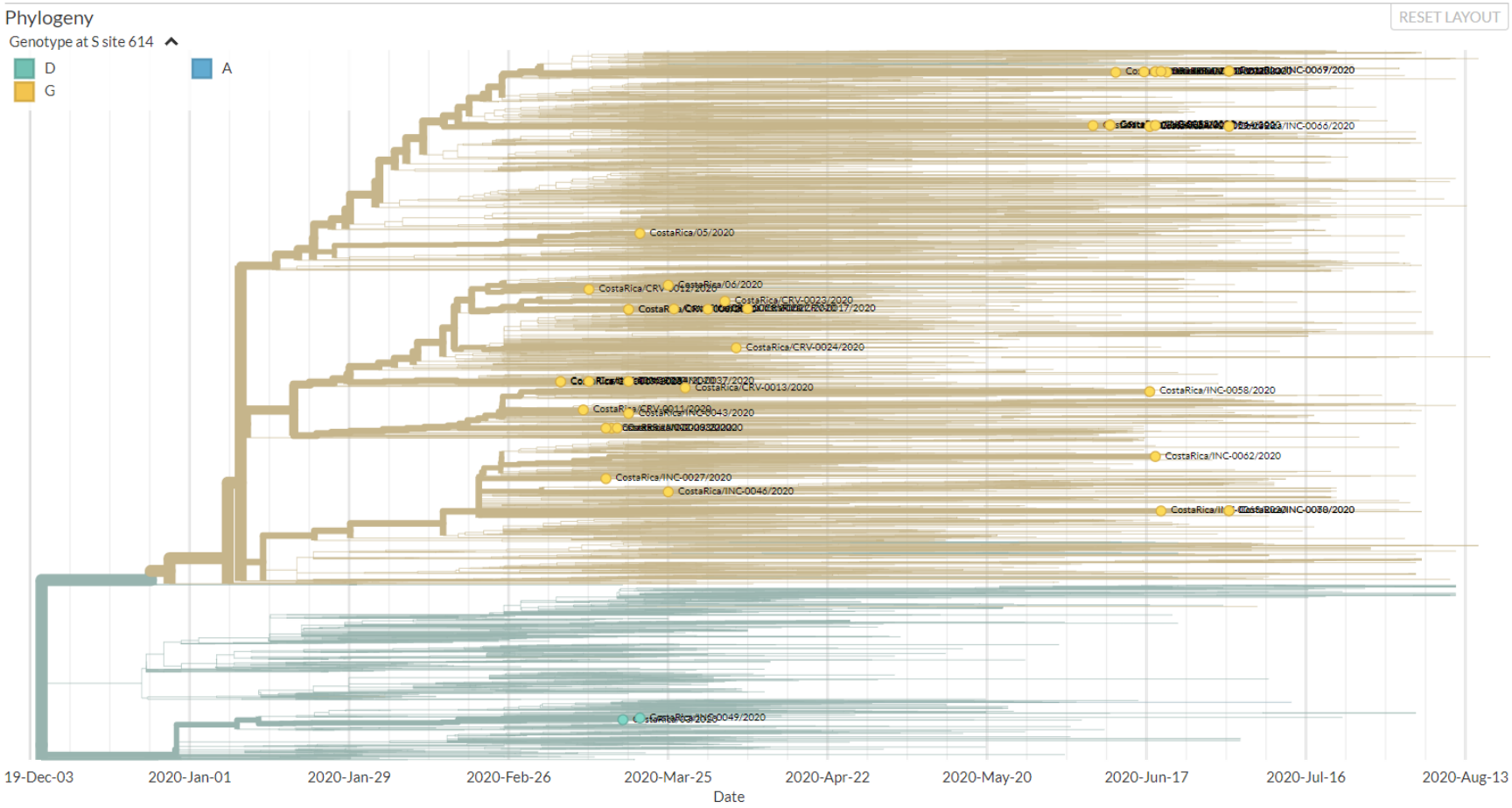


Figura 12. Distribución de secuencias de SARS-CoV-2 de Costa Rica dentro del contexto global. El color del grupo se define por el aminoácido en la posición 614 de la secuencia de aminoácidos de la espícula o proteína S: verde para ácido aspártico (Asp o D) y amarillo para glicina (Gly o G). Fuente: base de datos GISAID (https://www.nextstain.org/).

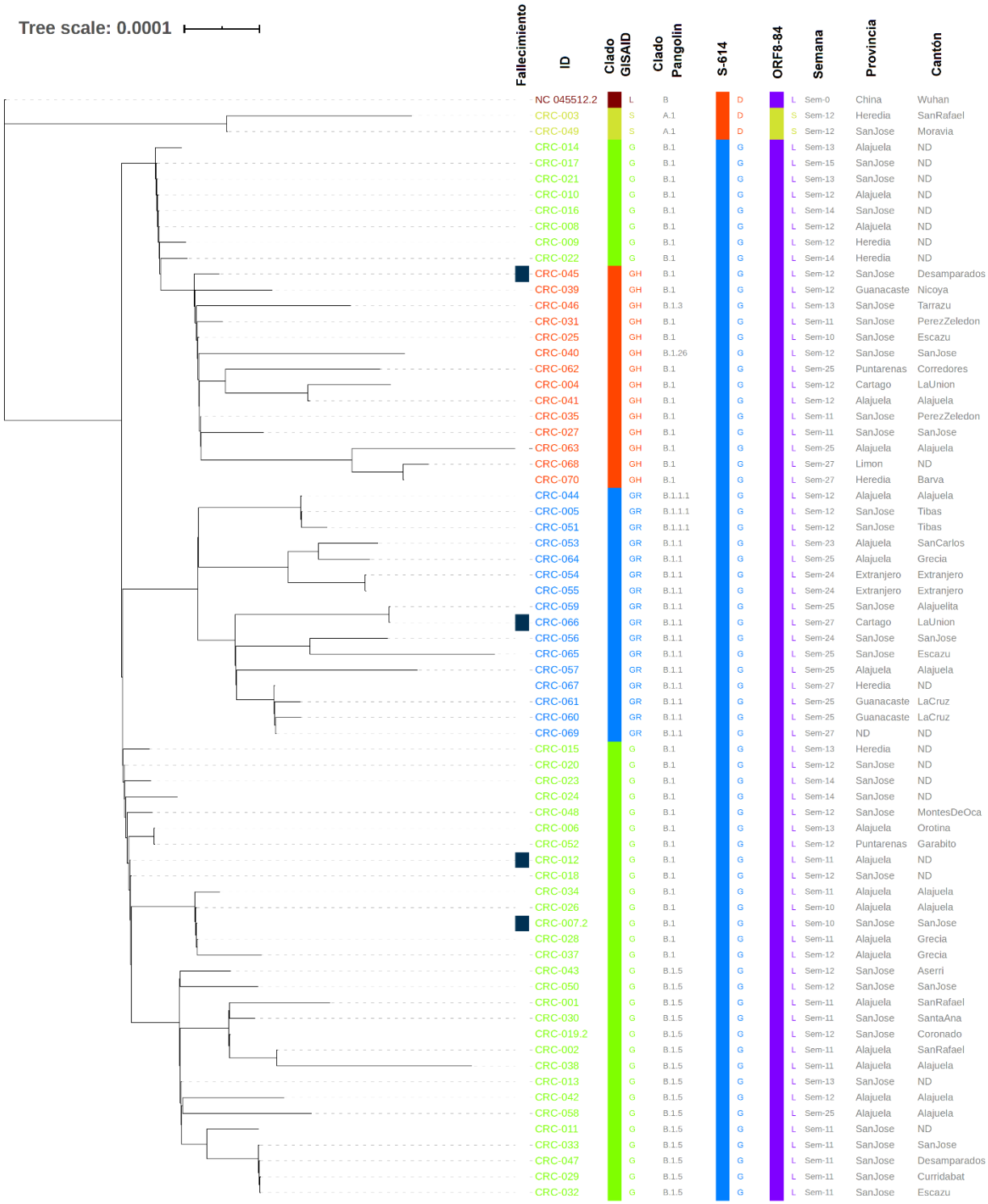


Figura 13. Relaciones filogenéticas de 70 secuencias de SARS-CoV-2 en casos de Costa Rica. Se identifican los clados dados por GISAID, el linaje Pangolin, la variante en la posición 614 de la proteína S, variante en la posición 84 de ORF8, semana epidemiológica, y localidad por provincia y cantón. Las defunciones se resaltaron con un cuadro negro.

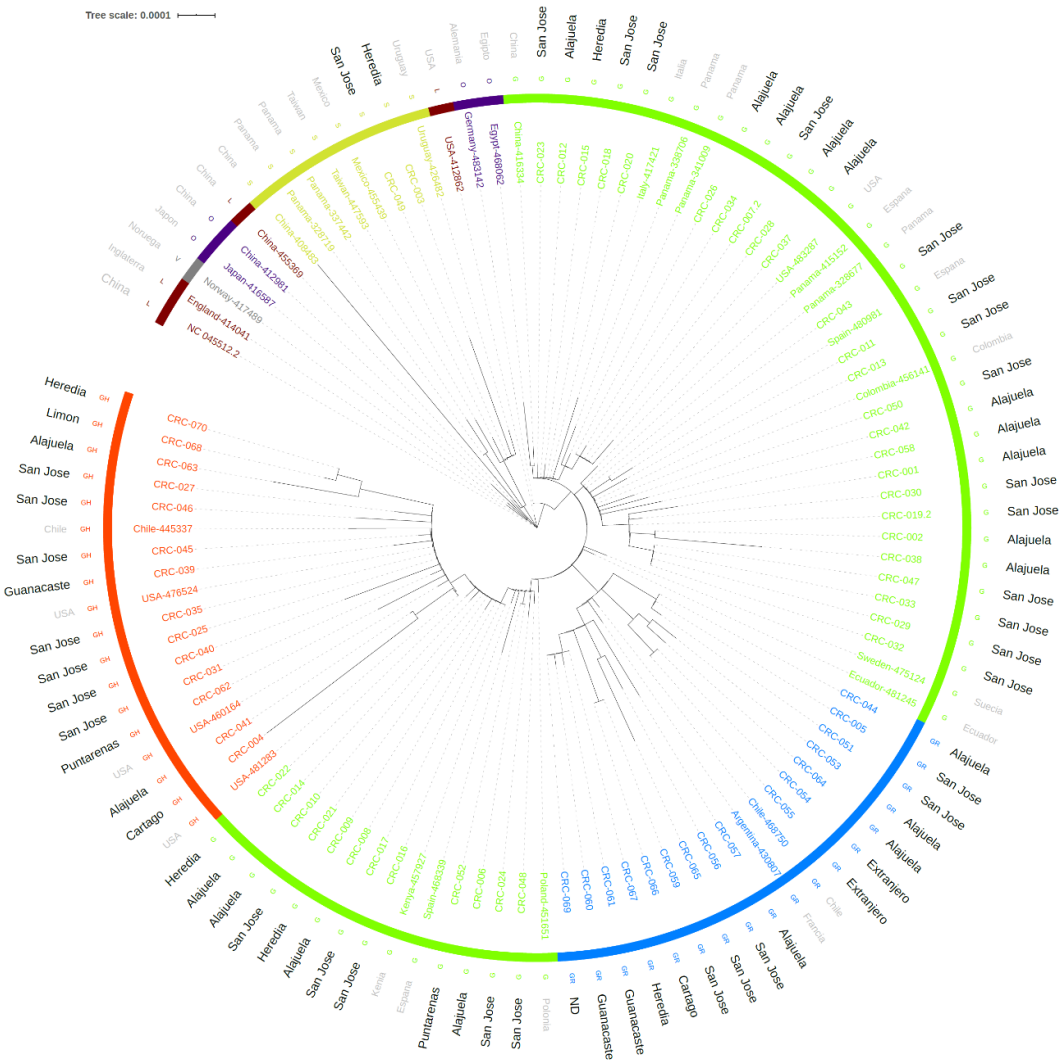


Figura 14. Relaciones filogenéticas de 70 secuencias de SARS-CoV-2 en casos de Costa Rica y otras 34 secuencias del mundo. Se identifican los clados o agrupaciones dadas por la clasificación de GISAID y la zona geográfica.

Conclusiones

El análisis de 70 genomas de SARS-CoV-2 en Costa Rica entre marzo y julio 2020 mostró que los linajes más frecuentemente identificados fueron: B.1, B.1.5 y B.1.1 respectivamente. Los mismos se distribuyeron de manera homogénea en cuanto al sexo y edad de los casos estudiados. Además, los linajes B.1 y B.1.1 presentaron amplia distribución en el territorio nacional y B.1.5 se presentó principalmente en las provincias de San José y Alajuela. De forma interesante, las secuencias de muestras en las semanas epidemiológicas 23-27 forman un agrupamiento (clúster), creando una posible asociación temporal en la cual se evidenció la transmisión activa del clado B.1.1 en la zona norte del país. Por otra parte, los casos asociados a defunciones no generaron un patrón particular por clado. Al estudiar las variantes genómicas (mutaciones) identificadas se reveló un patrón similar a lo reportado alrededor del mundo, incluyendo la distribución de casos con variantes en la posición D614G de la proteína S 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 de Corman et al., y tampoco se identificaron mutaciones en el dominio de unión al receptor de la espícula viral. Lo anterior resalta la necesidad de que Inciensa continúe con la vigilancia basada en laboratorio de las secuencias genómicas que circulan en nuestro país, para darle seguimiento a sus relaciones evolutivas.

Consideraciones importantes y limitaciones del estudio

Los análisis filogenéticos dependen de las secuencias utilizadas y disponibles en las bases de datos nacionales. No todos los países están realizando de manera sistemática la secuenciación y carga de genomas de SARS-CoV-2 a las bases de datos internacionales, por lo que las mismas presentan un sesgo inherente en cuanto a su representatividad.

La interpretación de la filogenia siempre debe llevarse a cabo a la luz de la información epidemiológica, datos como lo es el nexo epidemiológico (contactos), antecedente de viajes, fecha de inicio de síntomas, etc., son fundamentales para la contextualización de los resultados.

Los 70 genomas aquí analizados responden a una selección dirigida de muestras (ya sea por la disponibilidad del material biológico para analizar o por importancia epidemiológica de la muestra) y por consiguiente incluye también un sesgo muestral.

Se evidenciaron deficiencias en el llenado y calidad de la información epidemiológica recolectada en las boletas de solicitud de análisis que acompañan las muestras referidas al Inciensa. Se debe hacer hincapié en mejorar este proceso ya que resulta primordial para garantizar la calidad de la información necesaria para el proceso de vigilancia epidemiológica.

Referencias

Castillo, A. et al. (2020) Phylogenetic analysis of the first four SARS-CoV-2 cases in Chile. J. Med. Virol.(doi:10.1002/jmv.25797)

Corman, V. M., Landt, O., Kaiser, M., Molenkamp, R., Meijer, A., Chu, D. K., Bleicker, T., Brünink, S.,

Schneider, J., Schmidt, M. L., Mulders, D. G., Haagmans, B. L., van der Veer, B., van den Brink, S., Wijsman, L., Goderski, G., Romette, J. L., Ellis, J., Zambon, M., Peiris, M., … Drosten, C. (2020). Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin, 25(3), 2000045. https://doi.org/10.2807/1560-7917.ES.2020.25.3.2000045

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., … Cao, B. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The Lancet, 395(10223), 497–506. https://doi.org/10.1016/S0140-6736(20)30183-5

INCIENSA, & Ministerio de Salud Pública. (2020). Inciensa logra secuenciar el genoma completo del nuevo coronavirus SARS-CoV-2 (COVID-19). Retrieved May 21, 2020, from https://www.ministeriodesalud.go.cr/index.php/centro-de-prensa/noticias/741-noticias-2020/1642-inciensa-logra-secuenciar-el-genoma-completo-del-nuevo-coronavirus-sars-cov-2-covid-19

Korber B, Fischer W, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, Foley B, Giorgi E, Bhattacharya T, Parker M, Partridge D, Evans C, Freeman T, de Silva T, LaBranche C, Montefiori D, on behalf of the Sheffield COVID-19 Genomics Group. Spike mutation pipeline reveals the emergence of a more transmissible form of SARS-CoV-2. doi:10.1101/2020.04.29.069054. PPR:PPR157416.Mercatelli D and Giorgi FM (2020) Geographic and Genomic Distribution of SARS-CoV-2 Mutations. Front. Microbiol. 11:1800. doi: 10.3389/fmicb.2020.01800

Molina-Mora, J.-A., Campos-Sánchez, R., Rodríguez, C., Shi, L., & García, F. (2020). High quality 3C de novo assembly and annotation of a multidrug resistant ST-111 Pseudomonas aeruginosa genome: Benchmark of hybrid and non-hybrid assemblers. Scientific Reports, 10(1), 1392. https://doi.org/10.1038/s41598-020-58319-6

Rambaut, A., Holmes, E.C., O'Toole, Á. et al. (2020). A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat Microbiol https://doi.org/10.1038/s41564-020-0770-5

Robasky, K., Lewis, N. E., & Church, G. M. (2014). The role of replicates for error mitigation in next-generation sequencing. Nature Reviews. Genetics, 15(1), 56–62. https://doi.org/10.1038/nrg3655

Tang, X., Wu, C., Li, X., Song, Y., Yao, X., Wu, X., Duan, Y., Zhang, H., Wang, Y., Qian, Z., Cui, J., & Lu, J. (2020). On the origin and continuing evolution of SARS-CoV-2. National Science Review, nwaa036. https://doi.org/10.1093/nsr/nwaa036

Wu F, Zhao S, Yu B, et al. (2020). A new coronavirus associated with human respiratory disease in China. Nature. 2020;579(7798):265-269. doi:10.1038/s41586-020-2008-3

Yin, C. (2020). Genotyping coronavirus SARS-CoV-2: Methods and implications. Genomics, 112(5), 3588-3596. doi:10.1016/j.ygeno.2020.04.016

Zhao, W., Chen, J. J., Perkins, R., Wang, Y., Liu, Z., Hong, H., … Strain, E. (2016). A novel procedure on next generation sequencing data analysis using text mining algorithm. BMC Bioinformatics, 17(1), 213. https://doi.org/10.1186/s12859-016-1075-9

Agradecimientos:

Inciensa reconoce la contribución de los microbiólogos de los laboratorios que se indican a continuación, quienes suministraron las muestras (CRC-001 a 006 y CRC-025 a CRC-070) y la información clínico-epidemiológico, incluida en este informe.

A.S. Alajuela Central
A.S. Alajuela Norte - Clínica Dr. Marcial Rodríguez
A.S. Alajuela Sur
A.S. Aserrí
A.S. Corredores
A.S. Desamparados 1 - Clínica Dr. Marcial Fallas
A.S. Escazú (COOPESANA)
A.S. Fortuna
A.S. La Cruz
A.S. Los Chiles
A.S. Los Santos
A.S. Mata Redonda - Clínica Dr. Moreno Cañas
A.S. Orotina-San Mateo
A.S. Tibas (COOPESAIN) - Clínica Integrada Rodrigo Fournier
A.S. Tibas-uruca-merced - Clínica Dr. Clorito Picado
Centro Nacional de Rehabilitacion (CENARE)
Clínica Bíblica
EBAIS Concepción Norte
H. de Las Mujeres Dr. Adolfo Carit
H. Nacional de Niños Dr. Carlos Saenz Herrera
H. Dr. Fernando EscaLante PradilLa
H. Dr. Rafael A. Calderón Guardia
H. La Anexión
H. México
H. San Rafael de ALajueLa
H. San Vicente de Paul
Laboratorio clínico San josé
Laboratorio Labin
O.I.J. Morgue Judicial

Se reconoce además la contribución de la Universidad de Costa Rica y de sus colaboradores nacionales e internacionales quienes suministraron y secuenciaron de la muestras CRC-007 a la CRC-024, incluida en este informe. UCR (CIET) y colaboradores: Dra. Eugenia Corrales-Aguilar, Dr. Andres Moreira-Soto, Dr. Ignacio Postigo-Hidalgo, Dr. Ignacio Soto Pacheco, Dr. Jan Felix Drexler, Dr. Hugo Núñez-Navas, Dra. Teresita Somogyi, Dra. Karla Sofía Gutiérrez-Gutiérrez, Dr. Cristian Pérez-Corrales, Dr. Róger Soto-Palma y Dr. Andrei Montero-Bonilla.