Identificación del Cambio de Marco Ribosómico Programado en la pérdida de regiones en el genoma de variantes genéticas de SARS-CoV-2 del departamento de Santander, Colombia

María Alejandra Navarro Corredor

Trabajo de Grado para Optar al Título de Bióloga

Director

Francisco José Martínez Pérez

Doctor en Ciencias Biológicas

Codirector

Lina María Vera Cala

Doctora en Epidemiología

Universidad Industrial de Santander

Facultad de Ciencias

Escuela de Biología

Programa de Biología

Bucaramanga

2023

Dedicatoria

Quiero dedicar este logro a mi padre Héctor José Navarro, a mi tía Mercedes Navarro y a mi tío José Pablo Cala que sin su amor, apoyo incondicional y consejos no habría sido posible. Este logro también es de ustedes, quienes desde el cielo me acompañan y comparten esta gran felicidad conmigo.

Agradecimientos

Agradezco el financiamiento de esta pasantía al proyecto titulado: “Contribución al protocolo de diagnóstico del Coronavirus COVID-19 y del Virus de influenza A H1N1 por RT-PCR en tiempo real de un paso autorizado por la Organización Mundial de la Salud con la inclusión de genes marcadores requeridos para la infección viral y su validación por secuenciación genómica de nueva generación con la tecnología de concentraciones de nucleótidos”, con código registro de la Vicerrectoría de Investigación y Extensión de la Universidad Industrial de Santander No. 76900. Que se aprobó para su ejecución en la MinCienciatón con la Convocatoria: 1015- Invitación a presentar propuestas de proyectos relacionados con la pandemia de COVID- 19, con el contrato No. 369-2020, código 1102101576900.

Al Dr. Francisco José Martínez Pérez por permitirme hacer parte de su equipo, por su paciencia, guía y dedicación.

A Cristian Cadena por sus horas dedicadas a este proyecto, su apoyo y consejos.

A mi madre por apoyarme, por celebrar mis logros y contenerme en los momentos más difíciles, le agradezco por guiarme y enseñarme a afrontar los problemas. Sin su esfuerzo y dedicación no sería la persona que soy hoy.

A Brayan Castillo por ser mi compañero y soporte en este proceso. Le agradezco su paciencia, compresión y amor.

A mi familia, en especial a mi tío Augusto Corredor, quien en sus últimos momentos seguía regalándome esas hermosas palabras llenas de amor y sabiduría.

A Vanessa Blanco por nunca dejarme sola y brindarme su amistad más sincera.

Tabla de Contenido

Pág.

Lista de Tablas

Pág.

Lista de Figuras

Pág.

Lista de Apéndices

Nota: Los apéndices están adjuntos y puede visualizarlos en la base de datos de la biblioteca UIS

Resumen

Título: Identificación del Cambio de Marco Ribosómico Programado en la pérdida de regiones en el genoma de variantes genéticas de SARS-CoV-2 del departamento de Santander, Colombia[*]

Autor: María Alejandra Navarro Corredor[**]

Palabras Clave: SARS-CoV-2, ADNc, Cambio de Marco Ribosómico Programado, identificación de variantes virales, genoma viral.

Descripción: Durante la pandemia de COVID-19, generada por el *Betacoronavirus* SARS-CoV-2, para mejorar el diagnóstico por RT-PCR y la secuenciación de los genomas virales de pacientes, se realizó la inclusión de un reactivo diseñado en función del genoma de SARS-CoV-2, que contiene una solución coadyuvante para desenredar el ARN y permitir la amplificación de la región seleccionada de SARS-CoV-2. Los resultados mostraron una mejora en el sistema de diagnóstico y en la secuenciación genómica. Su aplicación permitió la caracterización de 14 genomas de pacientes diagnosticados con COVID-19 en Santander. Asimismo, de otros genomas con pérdida de codones que presentaron cambios en los Marcos Abiertos de Lectura (ORF), exhibiendo, en algunos codones de paro similitud con el modelo de Cambio de Marco Ribosómico Programado (PRF). Debido a la posibilidad de que los codones de paro pudieran haberse generado durante el ensamble computacional y no tuvieran relevancia biológica, se realizó una validación *in silico* de estos genomas, identificando una posible explicación a la existencia de genomas con estas características, haciendo una identificación del modelo PRF en algunas de las regiones con pérdidas, así como el desarrollo de una base de datos con secuencias de algunas variantes reportadas en la base de datos de GISAID para identificar patrones de similitud entre las secuencias y los genomas aquí obtenidos. La comparación de los genomas secuenciados con respecto a los genomas obtenidos de las variantes reportadas en GISAID evidenció un patrón de pérdida de nucleótidos, regulado por el modelo de PRF.

---

[*] Trabajo de Grado

[**] Facultad de Ciencias. Escuela de Biología. Director: Francisco José Martínez Pérez. Doctor en Ciencias Biológicas. Codirector: Lina María Vera Cala. Doctora en Epidemiología.

Abstract

Title: Identification of Programmed Ribosomal Frameshift in the loss of regions in the genome of genetic variants of SARS-CoV-2 from the department of Santander, Colombia [*]

Author(s): María Alejandra Navarro Corredor[**]

Key Words: Sars-CoV-2, cDNA, Programmed Ribosomal Frameshift, identification of viral variants, viral genome.

Description: During the COVID-19 pandemic, generated by the SARS-CoV-2 *Betacoronavirus*, to improve RT-PCR diagnosis and sequencing of patient viral genomes, the inclusion of a reagent designed based on the SARS-CoV-2 genome, containing a coadjuvant solution to unravel the RNA and allow amplification of the targeted region of SARS-CoV-2, was performed. The results showed an improvement in the diagnostic system and genomic sequencing. Its application allowed the characterization of 14 genomes of patients diagnosed with COVID-19 in Santander. Likewise, other genomes with codon loss presented changes in the Open Reading Frames (ORF), exhibiting, in some stop codons, similarity with the Programmed Ribosomal Frameshift (PRF) model. Due to the possibility that the stop codons could have been generated during computational assembly and had no biological relevance, an *in silico* validation of these genomes was performed, identifying a possible explanation for the existence of genomes with these characteristics, determining the PRF model in some of the regions with losses, as well as developing a database with sequences of some variants reported in the GISAID database to identify patterns of similarity between the sequences and the genomes obtained here. Comparing the sequenced genomes with the genomes obtained from the variants reported in GISAID showed a nucleotide loss pattern regulated by the PRF model.

---

[*] Bachelor thesis
[**]Science Faculty. School of Biology
Adviser: PhD. Francisco José Martinez Pérez. Adviser: PhD. Lina María Vera Cala

Introducción

El Coronavirus del Síndrome Respiratorio Agudo Severo 2 (SARS-CoV-2), perteneciente al género *Betacoronavirus*, causó una nueva enfermedad extendida a nivel mundial conocida como COVID-19, reportada por primera vez en Wuhan, China en diciembre de 2019 y declarada pandemia por la Organización Mundial de la Salud (OMS) en marzo 11 de 2020 (Rahimi et al., 2021; WHO, 2020), siendo este el séptimo coronavirus conocido en infectar a los humanos (Kesheh et al., 2022) y el tercero documentado por contagio zoonótico en causar brotes mundiales en las últimas dos décadas, después del coronavirus asociado al Síndrome Respiratorio Agudo Grave (SARS-CoV) en 2003 y el coronavirus del Síndrome Respiratorio de Oriente Medio (MERS-CoV) en 2012 (Gorbalenya et al., 2020; Kim et al., 2020; Ugurel et al., 2020).

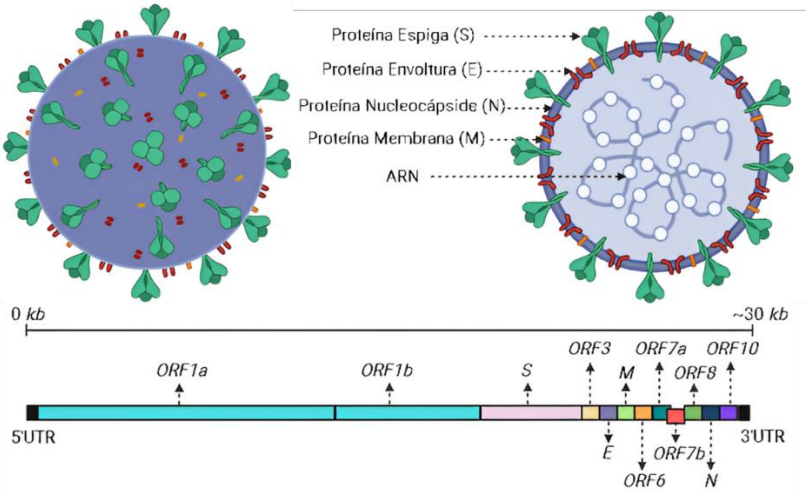
El estado de emergencia sanitaria causado por SARS-CoV-2 desencadenó un esfuerzo global en el desarrollo de estrategias para mitigar su propagación (WHO, 2021). Sin embargo, desde su inicio la respuesta mundial no fue la misma, ya que se desconocían aspectos como el origen del virus, su diagnóstico y tratamiento (Acuti Martellucci et al., 2020); por lo que, la tasa de infección fue mayor a lo esperado (Ashraf et al., 2021). Por consiguiente, la comprensión del genoma de SARS-CoV-2 se convirtió en el punto fundamental e invariable para el desarrollo de una continua actualización de los sistemas de vacunación y diagnóstico por RT-PCR para tener una vigilancia constante de salud pública (Rahimi et al., 2021). De modo que, uno de los principales logros de ese esfuerzo fue la caracterización de su genoma (Kim et al., 2020; Zhu et al., 2020), la cual, se realizó con la secuenciación tipo Illumina, Oxford Nanopore y la verificación de los espacios vacíos determinados por el ensamblaje de las lecturas con la secuenciación Sanger

de los amplicones de PCR 5’ o 3’-RACE (*Rapid Amplification of cDNA Ends*) (Lu et al., 2020; Zhu et al., 2020).

Por lo anterior, se observó que SARS-CoV-2 tiene un genoma de ARN monocatenario de sentido positivo de aproximadamente 30kb con 12 ORF que codifican 27 proteínas y una cola poli(A) en el extremo 3’ (Kim et al., 2020; Rahimi et al., 2021). Los dos primeros ORFs, *ORF1a* y *ORF1ab*, comprenden más de dos tercios del genoma y codifican dos poliproteínas grandes, pp1a y pp1ab, que se escinden proteolíticamente en 16 Proteínas No Estructurales (NSPS), necesarias para la replicación y transcripción del genoma viral (Chiara et al., 2021; de Wit et al., 2016; Kelly et al., 2021; Khailany et al., 2020; Zhang et al., 2020). Seguidamente, se encuentran los genes más cortos que codifican dos tipos de proteínas: 1) estructurales, con el gen *S* (*Spike*) que codifica la proteína espiga, el gen *E* la proteína de envoltura, el gen *M* la proteína de membrana y el gen *N* la proteína de nucleocápside; y 2) accesorias, con los genes *ORF3a, ORF6, ORF7a, ORF7b, ORF8 y ORF10* (Zhu et al., 2020) (Figura 1).

Figura 1

Genoma completo de SARS-CoV-2



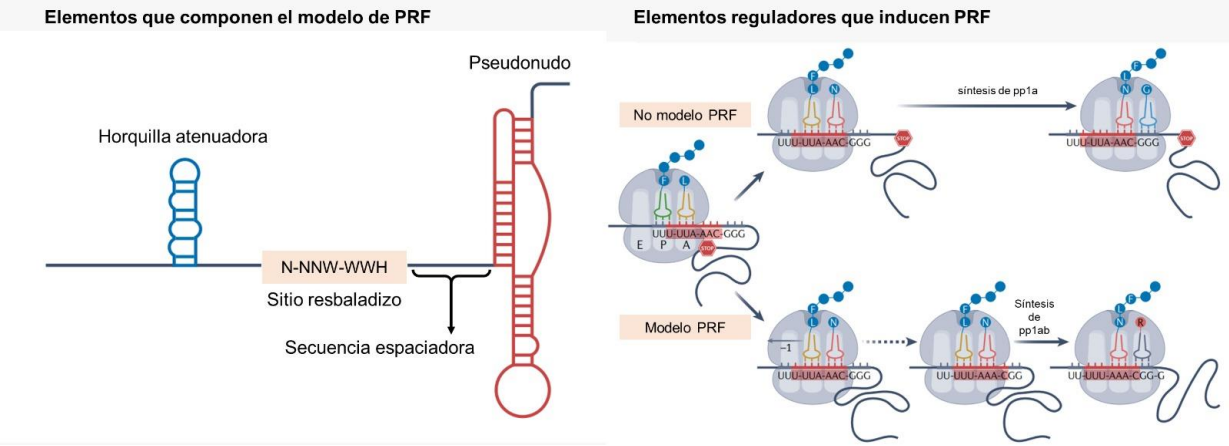
Nota. Se muestra la estructura completa del genoma junto a las proteínas estructurales y accesorias; así como, las partes que conforman la partícula viral de SARS-CoV-2.

Las proteínas no estructurales producidas por las poliproteínas pp1a y pp1ab deben estar presentes en una proporción adecuada para una óptima replicación y con ello propagación del virus, esto es facilitado por un mecanismo regulador denominado Cambio de Marco Ribosómico Programado (PRF) (Roman et al., 2021; Zhang et al., 2020). Este mecanismo es estimulado por un motivo de ARN mensajero (ARNm) estructurado y bien conservado de aproximadamente 80 nucleótidos en el extremo 3’ de *ORF1a* denominado elemento de Estimulación de Desplazamiento de Marco (FSE). Este elemento dirige a los ribosomas para que cambien sus ORFs en una base en la dirección 5’, lo que permite la lectura más allá del codón de paro del gen *ORF1a* (Dinman, 2012; Roman et al., 2021; Zhang et al., 2020).

Este motivo está compuesto por tres elementos importantes: un "sitio resbaladizo" conformado por siete nucleótidos N NNW WWH, donde realmente tiene lugar el cambio traduccional en el marco de lectura pasando a NNN WWW H, siendo NNN = tres bases idénticas, WWW = tres A o tres U, y H ≠ G (Kelly et al., 2021); una secuencia espaciadora corta de generalmente 12 nucleótidos; y una estructura estimulante, que en el caso de SARS-CoV-2 es un pseudonudo de ARNm de tres tallos (Kelly et al., 2021; Roman et al., 2021; Zhang et al., 2020). Así, antes del codón de paro TAA del gen *ORF1a* que generaría la primera poliproteína (pp1a) ocurre un deslizamiento que cambia la traducción formando el codón GTA y se incorpora la Valina, la cual, corresponde a *ORF1ab* codificando una poliproteína diferente (pp1ab). Esta nueva poliproteína al asociarse con el producto de *ORF1a* formará la ARN polimerasa esencial para la infección viral (Kelly et al., 2020). En este sentido, se pueden describir dos elementos que conforman el PRF, aquellos que componen el deslizamiento y los que regulan la traducción (Figura 2).

Figura 2

Modelo de Cambio de Marco Ribosómico Programado (PRF).



Nota. Se observan los elementos que componen y regulan el PRF. Adaptado de Dinman (2012).

Con todo lo anterior, sumado al patrón mutacional acelerado y a la acción evolutiva continua en SARS-CoV-2, se han generado diversas variantes del virus que han ocasionado nuevos picos de infección en la pandemia (Tao et al., 2021; Y. Zhang et al., 2022), la identificación y caracterización de dichas variantes a nivel mundial fue un reto, por lo que, fue necesario implementar una nomenclatura dinámica para clasificar los linajes genéticos de SARS-CoV-2, denominada nomenclatura PANGO (Rambaut et al., 2020). Esta ha sido implementada en varios softwares, siendo los más usados: *Global Initiative on Sharing All Influenza Data* (GISAID) (Shu y McCauley, 2017). Nextclade (Aksamentov et al., 2021) y *Phylogenetic Assignment of Named Global Outbreak Lineages* (PANGOLIN) (O'Toole et al., 2021). Asimismo, para la fácil interpretación del público no científico, la OMS implementó una nueva denominación para las Variantes de Interés (VOI) y Variantes de Preocupación (VOC) empleando letras del alfabeto griego (WHO, 2020). Entre las variantes más destacadas informadas a la fecha, se encuentran: Alpha, Beta, Delta, Gamma, Lambda, Mu y Ómicron. Es de resaltar, que la variante Mu fue

considerada como VOI por la OMS y fue caracterizada en Colombia en enero del 2021 (Halfmann et al., 2022; Hernandez-Ortiz et al., 2022). De igual forma, el estudio de la evolución viral de SARS-CoV-2 requiere una ardua labor de vigilancia genómica, la secuenciación rápida y directamente de la muestra tomada de los pacientes, evitando los cultivos celulares para prevenir los sesgos al identificar variantes no naturales del virus (Chiara et al., 2021).

La secuenciación del genoma facilita la caracterización rápida y precisa de los factores de virulencia del patógeno, infiere la dinámica epidemiológica, permite la detección y el mapeo de los brotes ocasionados por el virus (Gilchrist et al., 2015; Grubaugh et al., 2019). En Colombia, se incentivó la investigación científica en torno a la pandemia como estrategia para buscar soluciones que permitieran afrontarla y fortalecer los sistemas de alerta temprana ante emergencias de salud pública (Álvarez et al., 2020; Rojas-Pineda, 2020). Sin embargo, debido a la complejidad de la secuenciación completa del genoma de SARS-CoV-2, se tuvo que plantear una solución que mejorara la calidad de la secuenciación, denominada Tecnología Genómica UIS (Torres-Jiménez, 2021; Mejía-Ospino et al., 2021), la cual, tiene la capacidad de generar mezclas químicas para la amplificación y secuenciación de polímeros de ácidos nucleicos de genomas en proporciones extremas de nucleótidos, como es el caso de SARS-CoV-2 (Cadena-Caballero et al., 2022).

Por lo cual, en este trabajo se analizan los genomas obtenidos mediante la secuenciación Ion Torrent aplicando la Tecnología Genómica UIS en el marco del proyecto titulado: “Contribución al protocolo de diagnóstico del Coronavirus COVID-19 y del virus de Influenza A (H1N1) por RT-PCR en tiempo real de un paso autorizado por la OMS con inclusión de genes marcadores requeridos para la infección viral y su validación por secuenciación genómica de nueva generación”, para validar las pérdidas de regiones al emplear el modelo PRF e identificando patrones de similitud con secuencias reportadas en otros estudios.

1. Objetivos

1.1 Objetivo General

Analizar los genomas de SARS-CoV-2 secuenciados con Tecnología Genómica UIS para identificar si la pérdida de regiones en los genomas obtenidos son resultado de la polimerización de la tecnología Ion Torrent o son producto del ensamble computacional.

1.2 Objetivos Específicos

Identificar las regiones con pérdida de codones de genomas de SARS-CoV-2 secuenciados con la Tecnología Genómica UIS.

Caracterizar los Marcos Abiertos de Lectura en los genes de los genomas de SARS-CoV-2 obtenidos con la Tecnología Genómica UIS.

Establecer *in silico* la presencia de codones que pudieran favorecer la traducción por medio del modelo de Cambio de Marco Ribosómico Programado de los genomas SARS-CoV-2 caracterizados con la Tecnología Genómica UIS.

Validar *in silico* la relevancia de la pérdida de codones de las variantes genéticas de SARS-CoV-2 identificadas en pacientes diagnosticados con COVID-19 en el departamento de Santander.

2. Competencias

2.1 Competencias Cognitivas

Establece regiones genómicas de SARS-CoV-2 producto de la Tecnología Genómica UIS para validar la traducción del ARN mensajero.

Analiza Marcos Abiertos de Lectura de genomas de SARS-CoV-2 para la identificación de regiones con pérdida de codones que mantienen el Marco Abierto de Lectura.

Determina codones para el modelo de Cambio de Marco Ribosómico Programado de genes de SARS-CoV-2.

Valida *in silico* la relevancia de la pérdida de codones para ensambles producidos por secuenciación genómica.

2.2 Competencias Actitudinales

Aprende a escribir textos científicos para elaboración de informes de resultados de ensambles genómicos.

Aplica principios éticos y bioéticos para procesos de investigación científica.

Formula soluciones viables, rápidas y eficientes para solucionar problemas relacionados a la pérdida de fragmentos de genomas de SARS-CoV-2.

Genera ambiente amigable para la integración en equipos de trabajo.

3. Metodología

3.1 Ensamble de genomas de SARS-CoV-2

Para comprobar la calidad de los datos arrojados en la secuenciación de alto rendimiento con Ion Torrent se utilizó el software FastQC (Andrews, 2010) que, si bien no es una herramienta de análisis para los productos de secuenciación con Ion Torrent, permitió dar un vistazo rápido y sencillo del estado de estos.

El ensamble inicial de las lecturas de cada secuenciación se realizó con el software *Iterative Refinement Meta-Assembler* (IRMA) (Shepard et al., 2016) por el centro de secuenciación GenCore de la Universidad de los Andes en Bogotá-Colombia, con los parámetros por defecto. Sin embargo, para verificar el resultado anterior se comprobó la calidad de los ensambles, con un mapeo de las lecturas de cada ensayo respecto al genoma de referencia (**NC_045512.2)** de la base de datos del GenBank del *National Center for Biotechnology Information* (NCBI) (Sayers et al., 2021), mediante el script MapReads (https://github.com/GenomicUIS/covid/blob/main/src/MapReads.sh) que alineó las lecturas con el genoma de referencia usando el software Bowtie 2 versión 2.4.1 (Langmead y Salzberg, 2012). Los datos resultantes se comprimieron con el script Samscript (https://github.com/GenomicUIS/covid/blob/main/src/samscript.sh). Ambos scripts fueron generados en esta pasantía. Los datos se visualizaron con el software Geneious versión 2022.2.2 (https://www.geneious.com) para generar la secuencia consenso correspondiente a cada genoma.

Para la identificación de cada genoma ensamblado con su respectivo procedimiento, se generó un código de 10 caracteres, en donde, los dos primeros números corresponden al orden consecutivo de las muestras que ingresaron al estudio. Las dos siguientes indican el procedimiento

y/o tratamiento empleado para su secuenciación, a saber: el correspondiente del fabricante “sT”, Síntesis de ADNc con el Primer RACE 3’ “cD”; la inclusión de reactivo dNTPs “dN” y el reactivo desnaturalizante “dR”. Los últimos 6 números indican el mes, día y año de la toma de muestra. Para más información sobre la Tecnología Genómica UIS y el uso de las soluciones “cD, dN y dR” consultar (Cadena-Caballero et al., 2022; Torres-Jiménez, 2021; Mejía-Ospino et al., 2021).

3.2 Caracterización de las secuencias genómicas de SARS-CoV-2

Las secuencias de cada genoma ensamblado se compararon con la base de datos del NCBI con BLAST (Boratyn et al., 2013). Posteriormente, cada genoma se clasificó de acuerdo con la nomenclatura PANGO versión 4.1.3 (Rambaut et al., 2020) con los softwares PANGOLIN versión 1.16 (O’Toole et al., 2021), NEXTCLADE versión 2.9,1 (Aksamentov et al., 2021) y GISAID (Shu y McCauley, 2017).

Además, se realizó una selección de los más plausibles en función de su número de pares de bases para determinar los más cercanos respecto al genoma de referencia de SARS-CoV-2 (**NC_045512.2)** del GenBank (Sayers et al., 2021); el mismo que fue empleado para establecer las regiones de ausencia de nucleótidos en los genomas ensamblados de menor tamaño mediante el alineamiento con el software MAFFT versión 7.458 (Katoh y Standley, 2013) con los parámetros establecidos por el Modelo Evolutivo de Pérdida de ADN (DNA-LM) (Martínez-Pérez et al., 2007), haciendo uso de un script en Biopython (Cock et al., 2009) (https://github.com/druedaplata/bio) generado por el Grupo de Investigación en Cómputo Avanzado y a Gran Escala (CAGE) de la Universidad Industrial de Santander y ejecutado en el Computador de Alto Rendimiento (HPC) del Centro de Supercomputación y Cálculo Científico de la Universidad Industrial de Santander (SC3UIS), GUANE-1 (http://wiki.sc3.uis.edu.co/index.php/Cluster_Guane).

Por último, se realizó la respectiva anotación usando el programa Geneious Prime 2022.2.2. (https://www.geneious.com) para identificar las Secuencias Codificantes (CDS) respecto al genoma de referencia SARS-CoV-2 (**NC_045512.2**) alojado en el GenBank.

3.3 Identificación de los Marcos Abiertos de Lectura (ORFs)

Del alineamiento múltiple de los genomas generados, se identificaron los codones de inicio y terminación de la traducción de cada gen respecto al genoma de referencia empleado previamente. Se determinó el codón ATG (Metionina) como inicio de la traducción y los codones TAA, TAG y TGA, como codones de paro. La verificación de la secuencia de aminoácidos se realizó mediante la traducción conceptual con el Sistema Experto De Análisis De Proteínas (ExPASy) del portal de recursos del Instituto Bioinformático de Suiza (Duvaud et al., 2021). Los aminoácidos de las regiones que diferían respecto a los genes de referencia fueron ubicados en su respectivo codón manualmente. Finalmente, a partir del alineamiento con las respectivas anotaciones se identificaron y señalaron las regiones con variaciones en sus ORFs.

3.4 Identificación de los elementos del modelo PRF de SARS-CoV-2

Para la identificación de la horquilla atenuadora, el sitio resbaladizo y los tres tallos que conforman al pseudonudo de ARNm del modelo de PRF en los sitios con pérdida de codones y/o que pudiesen corresponder al modelo: 1) Se identificó la estructura secundaria de los sitios con pérdida de codones o cambio en el ORF de acuerdo con los parámetros del programa Mfold (Zuker, 2003). 2) Se realizó el modelado de las estructuras secundarias con el editor de estructuras de RNAstructure versión 1.0 (Reuter y Mathews, 2010), empleando como estructura de referencia la propuesta por Kelly et al., (2020). 3) Para determinar el tallo burbuja correspondiente al posible tallo atenuador se seleccionaron aproximadamente 40 nucleótidos rio arriba del codón de paro, para el sitio resbaladizo se tomaron 7 nucleótidos que se aproximara al patrón establecido para el

modelo PRF (NNNWWWH) y para el pseudonudo de ARNm de tres tallos del modelo de PRF se emplearon de 100-120 nucleótidos rio abajo. Es de resaltar, que los apareamientos canónicos o no canónicos que forman el tercer tallo del pseudonudo se determinaron en función del bucle correspondiente al primer tallo. Y 3) se modeló la estructura terciaria con el y con el Software RNAComposer versión 1.0 (Antczak et al., 2017; Popenda et al., 2012)

3.5 Validación *in silico* de pérdida de codones de las variantes de SARS-CoV-2

3.5.1 Construcción de base de genomas de SARS-CoV-2

Se descargaron secuencias genómicas de las variantes reportadas por GISAID desde enero de 2020 hasta enero de 2022, por grupos de un máximo de 10 mil secuencias con los parámetros de genomas completos y de alta cobertura. Posteriormente, se efectuaron dos procesos de curación: en el primero, se eliminaron las secuencias con dos o más nucleótidos indeterminados seguidos, representados con la letra “N” con un script de Biopython (https://github.com/druedaplata/bio); y en el segundo se eliminaron aquellas secuencias irregulares que irrumpían abruptamente el alineamiento. Finalmente, se recuperaron las secuencias que contenían uno o más nucleótidos indeterminados con el script dirty (https://github.com/GenomicUIS/covid/blob/main/src/dirty.py), diseñado en este estudio, para analizar e identificar patrones entre las regiones con nucleótidos determinados de las variantes descargadas con las regiones con pérdidas de los genomas ensamblados.

3.5.2 Alineamiento de secuencias y filogenia

Se generaron dos consensos para cada variante con umbral de frecuencia del 20% y 100% con el programa BioEdit versión 7.0.5.3 (Hall, 1999). Para identificar la similitud nucleotídica de los genomas secuenciados con los genomas reportados en GISAID, estos se alinearon con las secuencias consenso de las variantes con el programa MAFFT versión 7.458 (Katoh y Standley,

2013) empleando los parámetros de DNA-LM (Martínez-Pérez et al., 2007). El alineamiento resultante se editó a un tamaño de 120 caracteres por línea con el software GeneDoc versión 2.7. (Nicholas y Nicholas, 1997). Para la identificación de los consensos de cada variante de SARS-CoV-2 se indica el nombre y el número de genomas empleados para el consenso. A partir del alineamiento se identificaron los ORFs, siguiendo la misma metodología empleada en el inciso 3.3, y se realizó el análisis filogenético usando el método de estimación Maximum Likelihood, a través del software IQ-TREE multicore versión 1.6.12 (Nguyen et al., 2015). La visualización de los árboles se realizó con el software FigTree versión 1.4.4 (Rambaut, 2012).

Por otra parte, para identificar patrones de similitud entre las secuencias descargadas eliminadas y los genomas ensamblados, se recuperaron las secuencias con nucleótidos indeterminados de los genomas descargados de GISAID y se realizaron alineamientos múltiples en MAFFT versión 7.458 (Katoh y Standley, 2013), para las variantes Ómicron, Mu, Lambda y Gamma con el genoma de referencia y los genomas ensamblados. Además, de un alineamiento adicional con los consensos obtenidos de las secuencias con nucleótidos indeterminados de cada variante con un umbral de frecuencia del 20%, junto con el genoma de referencia y los genomas ensamblados.

4. Resultados

4.1 Ensamble de genomas de SARS-CoV-2

De las 14 muestras secuenciadas con el protocolo de Ion Torrent, 9 tuvieron en promedio 1,13 millones de lecturas, otras 2 tuvieron 500000 lecturas y las 3 muestras restantes 60650 lecturas. Además, 12 de estas tuvieron una cobertura cercana al 100%, pero en las muestras

05sT111220 y **06sT110920** fue menor al 30% y 60%, respectivamente. Los ensambles para estas muestras y las empleadas con los tratamientos mostraron con el software IRMA una menor cobertura respecto a Bowtie 2 (Tabla 1, Apéndice A y B). La secuenciación del ADNc con el procedimiento comercial "cD" generó rangos desde 1000 a 150000 secuencias, pero la cobertura en las muestras **07cD120320** y **36cD010821** fue mayor al 90%, con el software Bowtie 2 (Langmead y Salzberg, 2012). La inclusión de la solución dR en la síntesis de ADNc generó una cantidad de secuencias totales menor a un millón con una baja cobertura, pero la calidad (Q) de las secuencias por base (*per base sequence quality*) fue alta en todas las muestras. Además, el contenido de la secuencia por base (*per base sequence content*) cambió en los primeros 60 pb para estabilizarse antes de los 300 pb.

Este patrón se determinó en cuatro de las cinco muestras de ADNc que se sintetizó con la solución dN, mostrando un total de lecturas menor a 151000 con una baja cobertura, contrario a la muestra **07dN120320** que generó un total de 130000 secuencias con alta cobertura (Tabla 1, Apéndice A). Curiosamente, este mismo genoma, tuvo una cobertura del 99.3% respecto al genoma de referencia. Mientras que, el genoma **27sT122620**, obtenido con el método estándar tuvo una cobertura cercana al 97%. La muestra **36dR010821** y las muestras **03102720**, **10120420** con los tratamientos de cD, dR y dN no generaron resultados significativos y por lo tanto no fueron analizadas (Tabla 1, Apéndice A).

Tabla 1

Genomas y ensamblajes de SARS-CoV-2

No.	Código del genoma	ng/µl	Total de lecturas	pb IRMA	% Cobertura	pb Bowtie2	% Cobertura
1	02sT102520	31,80	2240211	29832	99,8%	29860	99,9%
2	03sT102720	45,90	1400386	29827	99,7%	29868	99,9%
3	03cD102720	40,80	150793	1958	6,5%	2622	8,8%
4	03dR102720	42,05	75745	ND	ND	1094	3,7%
5	03dN102720	52,00	150198	ND	ND	552	1,8%
6	05sT111220	55,00	65325	7875	26.3%	9204	30.8%
7	06sT110920	45,90	46657	15830	52.9%	20803	69.6%
8	07sT120320	72,00	1102797	29816	99,7%	29874	99,9%
9	07cD120320	97,00	59622	18363	61,4%	26898	90,0%
10	07dR120320	80,10	46584	12373	41,4%	15108	50,5%
11	07dN120320	112,90	132542	28720	96,0%	29684	99,3%
12	10sT120420	48,60	1498374	29814	99,7%	29885	99,9%
13	10cD120420	83,30	1174	ND	ND	1949	6,5%
14	10dR120420	78,70	12219	ND	ND	1894	6,3%
15	10dN120420	107,10	5606	13758	46%	17237	57,6%
16	12sT120620	41,40	2261112	29823	99,7%	29863	99,9%
17	14sT121420	44,90	4669120	29833	99,8%	29868	99,9%
18	25sT122420	39,80	2156585	29826	99,7%	29874	99,9%
19	27sT122620	55,30	69970	26083	87,2%	28965	96,9%
20	32sT122620	59,20	271442	29820	99,7%	29862	99,9%
21	36sT010821	60,40	2932212	29835	99,8%	29862	99,9%
22	36cD010821	71,50	8368	27004	90,3%	29767	99,5%
23	36dR010821	42,40	650	ND	ND	8113	27,1%
24	36dN010821	69,90	323	ND	ND	19923	66,6%
25	41sT011921	48,60	1900075	29786	99,6%	29865	99,9%
26	44sT012421	49,00	798340	29833	99,8%	29843	99,8%
27	44dR012421	ND	12246	ND	ND	563	1,9%
28	44dN012421	ND	267	ND	ND	215	0,7%

Nota. Esta tabla muestra los 28 genomas ensamblados con los diferentes tratamientos, con su respectiva concentración en nanogramos/microlitros, el número de lecturas obtenidas en la secuenciación, los pares de bases de cada resultado de ensamblaje y la cobertura correspondiente con respecto al genoma de referencia SARS-CoV-2 (NC_045512.2) para los softwares IRMA y Bowtie 2. Además, de los pares de bases obtenidos y aquellas muestras no determinadas (ND).

4.2 Caracterización de las secuencias genómicas de SARS-CoV-2

Los resultados de la secuenciación con la tecnología de Ion Torrent y ensamblados con IRMA mostraron que, de las 14 muestras de ARN seleccionadas en distintas fechas del proyecto,

a través de las lecturas se pudo determinar el genoma completo de 11 secuencias del virus, con un tamaño promedio de 29822 pb, 1 contuvo más de 26000 pb y los 2 restantes tuvieron un tamaño menor a 1600 pb. De igual forma, de las 5 muestras de ADNc enviadas con los tres tratamientos (cD, dR, dN) se obtuvieron 6 resultados positivos, de los cuales: las muestras 03, 07 y 36, con el método comercial, exhibieron un tamaño de 1958 pb, 18363 pb y 27004 pb, respectivamente; las muestras 07 y 10, cuya síntesis del ADNc incluyó el reactivo dN, generaron genomas de 28720 pb y 13758 pb, respectivamente; y la muestra **07dR120320** generó un genoma de 12373 pb. El mismo patrón se observó con los genomas ensamblados con Bowtie 2 (Tabla 1, Apéndice B y E).

El alineamiento tipo BLAST confirmó que, independientemente de su tamaño, la mayoría de los genomas presentan identidad cercana al 100% respecto a otros genomas de SARS-CoV-2 reportados en el GenBank (Sayers et al., 2021) (Apéndice C). Así mismo, la clasificación de acuerdo con la nomenclatura PANGO mostró que aproximadamente el 50% corresponde a los linajes genéticos B.1.1 en las muestras obtenidas del 25 de octubre al 4 de diciembre de 2020. Dos días después, se obtuvo el linaje B.1.1.348 hasta el 19 de enero de 2021 y a partir del 24 de ese mes, se determinó el linaje B.1.111. (Tabla 2, Apéndice D).

Es de resaltar que, la nomenclatura es dinámica y cambiante respecto al software Nextclade en la que 9 genomas ensamblados con IRMA correspondieron al linaje B.1.1.348, pero las muestras **10sT120420** y **32sT122620** con el ensamble Bowtie 2, correspondieron al linaje B.1.1. Mientras que, las muestras **02sT102520** y **44sT012421** correspondieron a los linajes B.1.1.203 y B.1.111. (Tabla 2, Apéndice D).

Tabla 2

Linajes de los genomas de SARS-CoV-2

Código del genoma	Fecha de colección	IRMA			Bowtie 2		
		Pangolin	Nextclade	GISAID	Pangolin	Nextclade	GISAID
02sT102520	25 nov 2020	B.1.1	B.1.1.203	B.1.1.348	B.1.1	B.1.1.203	ND
03sT102720	27 nov 2020	B.1.1	B.1.1.348	B.1.1.348	B.1.1	B.1.1.348	ND
07sT120320	03 dic 2020	B.1.1	B.1.1.348	B.1.1.348	B.1.1	B.1.1.348	ND
07cD120320	03 dic 2020	ND	ND	ND	B.1.1	ND	ND
07dR120320	03 dic 2020	ND	ND	ND	B.1.1	ND	ND
07dN120320	03 dic 2020	ND	ND	ND	B.1.1	B.1.1.348	B.1.1.348
10sT120420	04 dic 2020	B.1.1	B.1.1.371	B.1.1.371	B.1.1	B.1.1	ND
10dN120420	04 dic 2020	ND	ND	ND	B.1.1.348	ND	ND
12sT120620	06 dic 2020	B.1.1.348	B.1.1.348	B.1.1.348	B.1.1.348	B.1.1.348	ND
14sT121420	14 dic 2020	B.1.1	B.1.1.348	B.1.1.348	B.1.1	B.1.1.348	ND
25sT122420	24 dic 2020	B.1.1.348	B.1.1.348	B.1.1.348	B.1.1.348	B.1.1	ND
27sT122620	26 dic 2020	B.1.1.348	B.1.1.348	ND	B.1.1.348	B.1.1.348	B.1.1.348
32sT122620	26 dic 2020	B.1.1.348	B.1.1.348	B.1.1.348	B.1.1.348	B.1.1	ND
36sT010821	08 ene 2021	B.1.1.348	B.1.1.348	B.1.1.348	B.1.1.348	B.1.1.348	ND
36cD010821	08 ene 2021	ND	ND	ND	B.1.1	B.1.1.348	B.1.1.348
36dR010821	08 ene 2021	ND	ND	ND	B.1.1	ND	ND
36dN010821	08 ene 2021	ND	ND	ND	B.1.1	ND	ND
41sT011921	19 ene 2021	B.1.1.348	B.1.1.348	ND	B.1.1.348	B.1.1.348	B.1.1.348
44sT012421	24 ene 2021	B.1.111	B.1.111	B.1	B.1.111	B.1.111	ND

Nota. Esta tabla muestra los linajes de acuerdo con la clasificación PANGO de los genomas ensamblados con los softwares IRMA y Bowtie 2 con Pangolín, Nexclade y GISAID. Además de las muestras no determinadas (ND).

4.3 Identificación de los ORF de SARS-CoV-2

4.3.1 Caracterización de las regiones con pérdidas

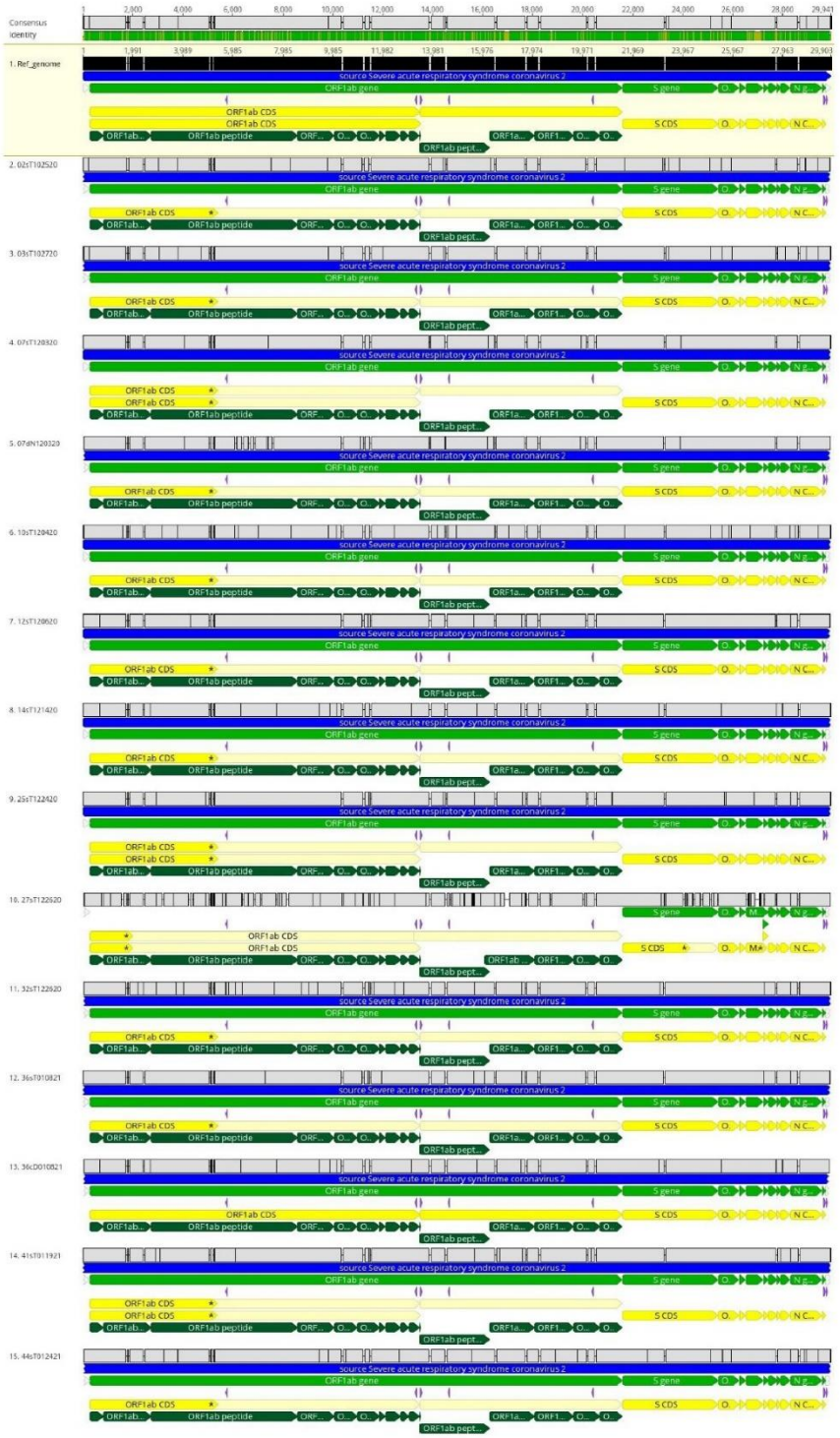
De la traducción conceptual de cada ORF se identificó pérdida de regiones nucleotídicas en los genomas correspondientes a las muestras **07dN120320** y **27sT122620** generando, en algunos casos, cambios en el mismo, especialmente en los genes *ORF1ab* y *S*. El genoma

07dN120320 mostró siete pérdidas nucleotídicas. Las seis primeras en la región del gen *ORF1a* y la restante en *ORF1ab*. Mientras que, el genoma **27sT122620** presentó 22 pérdidas con dos patrones. El primero fue similar para el gen *ORF1a*, del genoma anterior con ocho pérdidas, mientras que en *ORF1ab* fueron cuatro (Tabla 3). En el segundo patrón se determinaron seis pérdidas de regiones en el gen *S*, tres en el gen *M* y una en el gen *ORF6a* (Tabla 3, Apéndice E y G).

La pérdida de estas regiones en ambos genomas se verificó con la secuencia nucleotídica de las lecturas empleados para el ensamble genómico, con lo que se confirmó que la disminución de nucleótidos correspondió a la unión de dos lecturas y en otros casos a una sola (Figura 3, Apéndices B y E). Un aspecto sobresaliente de pérdida de regiones del genoma **07dN120320** es que en solo una de estas no se generó cambio en el ORF. En estas regiones de pérdida, tres de los nuevos ORFs tuvieron el codón de paro TAA y este fue mayor respecto al codón TGA. De igual forma, en las regiones de pérdida del genoma **27sT122620** para los nuevos ORFs el codón de paro mayoritario fue TGA respecto a los codones TAA y TAG que tuvieron la misma cantidad de regiones. En las otras siete no se rompen los ORFs y la traducción continua (Tabla 3, Tabla 4 y Figura 4).

Figura 3

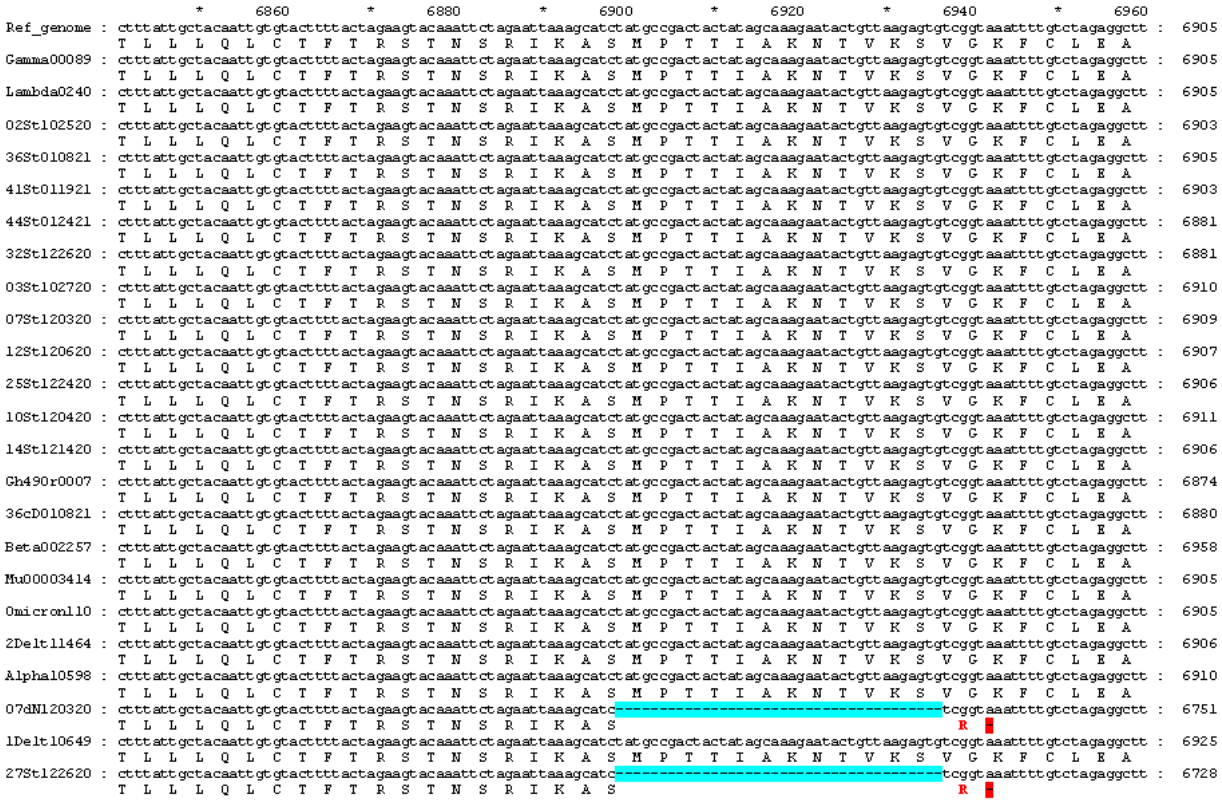
Mapa genómico con los ORFs para los 14 genomas de SARS-CoV-2



Nota. Esta imagen muestra el mapa genómico con los ORFs para los 14 genomas reportados en GISAID de SARS-CoV-2. En la parte superior de la imagen se observa la longitud en pb, el consenso y la identidad de las muestras para todo el alineamiento. La barra negra representa la secuencia de nucleótidos del genoma de referencia (NC_045512.2) y las grises a los genomas ensamblados. En líneas verde claro se representan el tamaño de los genes, en amarillo las Regiones Codificantes (CDS) y en verde oscuro la longitud de los péptidos. Por otro lado, los triángulos morados muestran los sitios de PRF y los asteriscos los sitios con codón de paro. Figura graficada con el software Geneious versión 2022.2.2 (https://www.geneious.com).

Figura 4

ORFs con regiones con pérdida de codones y posible PRF



Nota. Esta figura muestra parte del alineamiento *in silico* de la presencia de codones para los genomas obtenidos y las variantes de SARS-CoV-2 que favorecen la traducción por medio del modelo de PRF. En azul se muestra la región con pérdida de codones y en rojo se señalan los aminoácidos que no siguen el ORF y, el codón de paro (Apéndice G).

4.3.2 Sustituciones de aminoácidos y mutaciones de nucleótidos respecto al genoma de SARS-CoV-2

Se determinó que las sustituciones de aminoácidos ocurrieron principalmente en las proteínas no estructurales y en espiga, en diferente frecuencia. Las sustituciones comunes para los 14 genomas se presentaron en NSP12: P323L y en S: D614G. La misma cantidad de cambios estuvieron en N: G204R y R203K, pero se comparten en 13 genomas. Además, los genomas **12sT120620**, **25sT122420**, **27sT122620**, **32sT122620**, **36sT010821** y **41sT011921** compartieron dos sustituciones una en N: S2Y y la otra en NSP6: V149F. Los genomas **14St121420** y **36cD010821** comparten cinco sustituciones una en S: S477N y las otras en: NSP7a: L102F, NSP4: Y441H, NSP5: V36I, y NSP10: A32V. Los únicos genomas que compartieron el cambio NSP3: Q1125H fueron **32sT122620** y **41sT011921**. Las demás sustituciones presentes en los 14 genomas fueron específicas por posición. Por otro lado, a nivel nucleotídico para los 14 genomas se encontraron 258 mutaciones únicas. Es de resaltar que, según la nomenclatura GISAID el genoma **44sT012421** fue el único que perteneció al clado GH respecto a los demás que se agruparon en el clado GR. Además, de no presentar las sustituciones NSP3: G174D y L357F, que los demás genomas si comparten a excepción del genoma **10sT120420**, y si mostrar una sustitución en S: G1167A y otra en N: T366I (Tabla 4).

Tabla 3

Regiones y/o codones ausentes de los genomas 07dN120320 y 27sT122620

No.	Genoma	Gen	Posición en el genoma	ORF	Aminoácidos ORF	Posición en la proteína	Nuevo dominio ORF de aminoácidos	Nombre de la región
1	27sT122620	ORF1a	1584-1607	Nuevo codón de paro TGA	VVG**EGSEGLN**DNL	438-450	VVGTTFLKYSKKRKSTSILLVTL NLMKRSPLFWHLFLLPQVLLWK L-	BetaCoV_Nsp2_SARS-like 182..818
2			2428-2452	Nuevo codón de paro TGA	KCV**KSREETGL**LMP	719-732	KCVTHASKSPKRNYLLRGRNTS HRSVNRGSCLENW-	Proteína no estructural de betacoronavirus 2 (Nsp2) 182..818
3			3197-3218	Nuevo codón de paro TGA	QEE**DWLDDD**SQQT	975-986	QEEATNCWSTRRQ-	Proteína de función desconocida (DUF3655) 880..1050
4			4248-4273	ORF continua	ITT**YPGQGLNG**YTV	1326-1339	ITTYTV	Superfamilia de macrodominios; cl00019 1233..1358
5			5120-5121	Nuevo codón de paro TAA	KTFYVLP	1616-1622	KTF*YVLP	betaCoV_PLPro. Proteasa similar a la papaína de Betacoronavirus. 1566..1868
6	07dN120320		6080-6120	Pérdida de codones. Continúa ORF un nuevo codón Ala.	KFA**DDLNQLTGYKKPAS**RE L	1935-1955	KFAAREL	SARS-CoV-like_Nsp3_NA. Dominio de unión de ácido nucleico de proteína no estructural 3 BCOV_NSP3E_NAB, PS51945; Perfil de dominio de unión a ácido nucleico (NAB) de Betacoronavirus Nsp3e (MATRIX) 1913..2019.
7	27sT122620		6339-6380	ORF continúa dos nuevos aminoácidos X y Val	DVL**KSEDAQGMDN**LAC	2026-2041	DVLXVLAC	Ninguna región
8	07dN120320		6341-6380	Nuevo codón de paro TAA	NSF**DVLKSEDAQGMDNL**A CE	2023-2042	NSFACLRRSKTSL-EEV	Ninguna región
9			6622-6637	Pérdida de codones continúa ORF	LGL**KTLAT**HGL	2117-2127	LGLHGL	SARS-CoV-like_Nsp3_betaSM marcador específico de Betacoronavirus de proteína no estructural 3 (Nsp3). 2044..2159
10	07dN120320 27sT122620		6843-6882	Nuevo codón de paro TAA	KASMPTTIAKNTVKSVGKF	2191-2209	KASR-	Ninguna región
11	27sT122620		7101-7135	Nuevo codón de paro TAA	TIA**TYCTGSIPCSV**CLS	2276-2293	TIACLSGLDSFRHLSFFRNYTNYH FIF-	Ninguna región
12	07dN120320		7556-7582	Nuevo codón de paro TAA	NGV**RRSFYVYAN**GGK	2428-2442	NGVRR-	Ninguna región
13	27sT122620		8172-8203	ORF continúa	ISAARQGFVDSDVETK	2633-2649	ISAETK	Ninguna región
14	07dN120320	ORF1ab	14485-14502	Nuevos codones y codón de paro TGA	YHF**RELGVVVHN**	4738-4748	YHFSQLQVR-	SARS-CoV-like_RdRp ARN dependiente de ARN polimerasa. 4397..5324

No.	Genoma	Gen	Posición en el genoma	ORF	Aminoácidos ORF	Posición en la proteína	Nuevo dominio ORF de aminoácidos	Nombre de la región
15	27sT22620	ORF1ab	16838-16735	Nuevo codón de paro TAG	TFE**KGDYGDAVVYRGTTTY KLNVGDYFVLTSHTVMPLS APTLVPQEHYVRITGLYPT LNISDEFSSNV**ANY	5523-5593	TFEQIIKRLVCKSILHSRDHLVLV RVILLLA-	BetaCoV_Nsp13-helicase. Dominio helicasa 5575..5914
16			17726-17743	Nuevo codón de paro TAG	VVR**EFLTRN**PAW	5818-5830	VVRTLLGEKLSLFHLIIHRML-	betaCoV_Nsp13-helicase. Dominio helicasa 5575..5914
17			18680-18719	Nuevo codón de paro TGA	RRA**TCFSTASDTYACW**HHS	6137-6155	RRAGIILXDLITSIIRL-	BetaCoV_Nsp14 Proteína no estructural 14 (Nsp14) 5930..6450
18			19971-20008	Nuevo codón de paro TAG	TETICAPLTVFFDGRVDGQV D	6564-6584	TETICAK-	M_alpha_beta_cv_Nsp15-like dominio medio de la proteína no estructural 15 (Nsp15). 6517..6648
19		S	23991-24017	Nuevo codón de paro TGA	DPS**KPSKRSFIEDL**	808-821	DPSRRSTFQQSDTCRCWLHQTIW -	SARS-CoV-like_Spike_SD1-2_S1-S2_S2 Péptido de fusión interno 543..1208
20			24122-24144	Nuevo codón de paro TGA	CAQ**KFNGLTVLPPL**	851-864	CAQSLTALLXATFAHR-	SARS-CoV-like_Spike_SD1-2_S1-S2_S2 543..1208
21			24372-24395	ORF continúa	QDS**LSSTASAL**GKL	935-948	QDSGKL	Repetición de heptada 1" 918..983
22			24682-24731	Nuevo codón de paro TGA	KRV**DFCGKGYHLMSFP**QS A	1038-1056	KRVXFXWKTVSTSWCSLLACDL CPCTRKELHNCSCHLS-	Ninguna región
23			25110-25124	Nuevo codón de paro TAA	ISG**INASVVNIQKE**IDRL	1169-1186	ISGLTASMRLPRI-	Repetición de heptada 1162..1203
24			25330-25347	Nuevo codón de paro TGA	DED**DSEPVLKGV**	1257-1268	DEDAKESNYITHKRTYGFVYENL HNWNCNFEAR-	Ninguna región
25		M	26542-26573	ORF continúa	NGT**ITVEELKK**LLEQ	5-19	NGTLEQ	Ninguna región
26			26624-26628	ORF continúa	ICLLQF	32-37	ICLQF	Ninguna región
27			26856-27066	Nuevo codón de paro TGA	MWS**FNPETNILLNVPLHGT ILTRPLLESELVIGAVILRG HLRIAGHHLGRCDIKDLPK EITVATSRTLSYYKL**GAS	108-182	MWSSPTAWELRSV-	Ninguna región
28		ORF6a	27199-27217	Primera pérdida codón ATG, siquiente ATG	MFHLV	1-19	MRTFK	Ninguna región

Nota. Esta tabla muestra las regiones con pérdidas de codones en los genomas **07dN120320** y **27sT122620**. Se observan: No.: Número de identificación, ORF: las observaciones identificadas en el ORF, Aminoácidos ORF: traducción del ORF en la región señalada del genoma de referencia (NC_045512.2), Nuevo dominio ORF de aminoácidos: la nueva secuencia de aminoácidos en la región con pérdida de codones, Nombre de la región: nombre de la región en donde se presentó la pérdida. En negrita y subrayado se destacan los aminoácidos de la región con pérdida.

Tabla 4

Mutaciones y sustituciones de aminoácidos de los genomas de SARS-CoV-2

No.	Código	Nombre del virus	ID de acceso	Clado	Sustituciones de AA	Mutaciones de nucleótidos
1	02sT102520	hCoV-19/Colombia/SAN-576900-002/2020	EPI_ISL_1576835	GR	Spike A27V, Spike D614G, Spike G1167A, N G204R, N R203K, N T366I, NS3 G174D, NSP3 L357F, NSP12 P323L	C241T, C3037T, A3790T, C7528T, C8299T, G9802T, C14408T, C16293T, C21642T, C23185T, A23403G, C23854T, G25062C, C25521T, G25913A, G28881A, G28882A, G28883C, C29370T
2	03sT102720	hCoV-19/Colombia/SAN-576900-003-RNA/2020	EPI_ISL_1577026	GR	Spike D614G, Spike G1167A, N G204R, N R203K, N T366I, NS3 G174D, NS7a V93L, NS8 P56S, NSP3 A664V, NSP3 L357F, NSP12 P323L	T25A, C241T, C3037T, A3790T, C4710T, G9802T, T11974C, C14408T, A23403G, G25062C, C25521T, G25913A, G27670C, C28059T, G28881A, G28882A, G28883C, C29370T
3	07sT120320	hCoV-19/Colombia/SAN-576900-007-RNA/2020	EPI_ISL_1577182	GR	Spike D614G, Spike G769V, Spike G1167A, N G204R, N R203K, N T366I, NS3 G174D, NSP3 L357F, NSP12 P323L, NSP12 S913L, NSP12 Y149H, NSP15 D91Y	T25A, C241T, C3037T, A3790T, C4048T, C7417T, G9802T, T13885C, C14408T, C16178T, T16419A, G19891T, A23403G, G23868T, G25062C, C25521T, G25913A, G28881A, G28882A, G28883C, C29370T, T29840C
4	07dN120320	hCoV-19/Colombia/SAN-576900-007-2-1/2020	EPI_ISL_16399056	GR	Spike D614G, Spike G769V, Spike G1167A, N G204R, N R203K, N T366I, NS3 G174D, NSP3 L357F, NSP3 Y801N, NSP12 E350Q, NSP12 G352Q, NSP12 P323L, NSP12 R349S, NSP12 S913L, NSP12 V354I, NSP12 Y149H, NSP15 D91Y	C241T, C3037T, A3790T, C4048T, ins5116T, T5120A, G7360K, C7417T, G9802T, T11104C, A11111W, T13885C, C14408T, A14487T, G14488C, A14493G, G14494C, G14495A, T14496A, ins14499CGGTG, G14500A, C16178T, T16419A, G19891T, A23403G, G23868T, G25062C, C25521T, G25913A, G28881A, G28882A, G28883C, C29370T
5	10sT120420	hCoV-19/Colombia/SAN-576900-010-RNA/2020	EPI_ISL_1582589	GR	Spike D614G, Spike L18F, Spike R21I, M R44K, N G204R, N P67S, N R203K, NS3 S171L, NS8 F120S, NSP2 G262S, NSP3 T147I, NSP3 T350I, NSP8 T123I, NSP12 P323L, NSP13 S74P, NSP13 S259L	C241T, G1589A, C3037T, C3159T, C3768T, T6067C, C6196T, A7015G, C8326T, C10138T, C12459T, C14184T, C14408T, C14599T, C14925T, T16456C, C17012T, C21614T, G21624T, A23403G, C25791T, C25904T, G26653A, T28252C, C28472T, G28881A, G28882A, G28883C
6	12sT120620	hCoV-19/Colombia/SAN-576900-012-RNA/2020	EPI_ISL_1628474	GR	Spike D614G, Spike G1167A, N G204R, N R203K, N S2Y, N T366I, NS3 G174D, NS7a Q90stop, NSP3 L357F, NSP6 V149F, NSP12 P323L	C241T, C683T, C3037T, A3790T, C4327T, G9802T, G11417T, C14408T, T15642C, C17550T, A23403G, G25062C, C25521T, G25913A, C27661T, C28278A, G28881A, G28882A, G28883C, C29370T
7	14sT121420	hCoV-19/Colombia/SAN-576900-014-RNA/2020	EPI_ISL_1629710	GR	Spike D614G, Spike G1167A, Spike S477N, N G204R, N R203K, N T366I, NS3 G174D, NS7a L102F, NSP3 L357F, NSP4 Y441H, NSP5 V36I, NSP10 A32V, NSP12 P323L	T25C, C241T, C683T, C2449T, C2710T, C3037T, A3790T, C6310T, C7732T, C9430T, G9802T, T9875C, G10160A, C13119T, C14408T, C15720T, C17502T, G22992A, A23403G, G25062C, G25494T, C25521T, G25913A, C27697T, T27935C, G28881A, G28882A, G28883C, C29370T
8	25sT122420	hCoV-19/Colombia/SAN-576900-025-RNA/2020	EPI_ISL_1582590	GR	Spike D614G, Spike G1167A, N G204R, N R203K, N S2Y, N T366I, NS3 G174D, NS3 L85V, NS3 L95F, NSP3 I733V, NSP3 L357F, NSP3 P74S, NSP6 V149F, NSP12 P323L, NSP16 K160R	T25A, C241T, C2939T, C3037T, A3790T, A4916G, G9802T, G11417T, T14382C, C14408T, T15642C, C17550T, A21137G, A23403G, G25062C, C25521T, T25645G, G25677T, G25913A, G26828T, C28278A, G28881A, G28882A, G28883C, C29370T
9	27sT122620	hCoV-19/Colombia/SAN-576900-027/2020	EPI_ISL_16399398	GR	Spike D614G, Spike G857R, Spike G1044W, Spike G1167A, Spike L858P, Spike N856stop, Spike T859Y, Spike V860C, M K180C, M Y178T, M Y179D, N G204R, N R203K, N S2Y, N T366I, NS3 G174D, NSP3 C1473V, NSP3 D729G, NSP3 D1478I, NSP3 G1300C, NSP3 G1476V, NSP3 I733V, NSP3 L72F, NSP3 L357F, NSP3 L1477stop, NSP3 N1220V, NSP3 P1750S, NSP3 S1475V, NSP3 S1479L, NSP3 S1717P, NSP4 A316V, NSP5 G278A, NSP6 A79S, NSP6 V78C, NSP6 V149F, NSP6 Y80L, NSP8 I132V, NSP12 C730T, NSP12 D92Y, NSP12 D736S, NSP12 D738E, NSP12 D740E, NSP12 E729N, NSP12 E744Q, NSP12 F741M, NSP12 H725del, NSP12 H816P,	T62C, C241T, T734W, A808W, G1125R, C2933T, C3037T, T3616Y, A3790T, del4379_4379, A4905G, A4916G, G5272K, T5577W, T5707C, G6352A, A6377G, A6378T, T6379C, G6617T, A6898G, del6899_6899, del7102_7102, A7103N, ins7156T, del7742_7742, G7743A, T7868C, C7967T, A9204R, C9501T, G9802T, ins10348A, T10443K, G10887C, T11180K, ins11200A, A11201N, G11417T, ins11473T, C11653T, G11873K, A12485G, G13714T, C14408T, del14684_14685, A14754C, C15080Y, C15212T, C15240T, C15536Y, A15578W, A15592C, G15594C, T15595C, A15596G, T15597C, C15601T,

No.	Código	Nombre del virus	ID de acceso	Clado	Sustituciones de AA	Mutaciones de nucleótidos
					NSP12 K438N, NSP12 K718H, NSP12 L723del, NSP12 L727del, NSP12 L731D, NSP12 N734S, NSP12 N743T, NSP12 P323L, NSP12 Q724del, NSP12 R721S, NSP12 R726del, NSP12 R733M, NSP12 R735V, NSP12 T591I, NSP12 V737I, NSP12 V742L, NSP12 Y719R, NSP12 Y732F, NSP13 L83I, NSP13 P82S, NSP13 S80stop, NSP15 A137P, NSP15 A160L, NSP15 A171P, NSP15 D132T, NSP15 E145K, NSP15 E170K, NSP15 F134L, NSP15 G140V, NSP15 G146V, NSP15 G150V, NSP15 G156V, NSP15 G164E, NSP15 G169E, NSP15 I143L, NSP15 I168L, NSP15 K158N, NSP15 L133Y, NSP15 L151Y, NSP15 L167stop, NSP15 N136M, NSP15 N139M, NSP15 N163M, NSP15 P153H, NSP15 Q130K, NSP15 Q152N, NSP15 Q159K, NSP15 R135E, NSP15 R138V, NSP15 S147V, NSP15 S154L, NSP15 S161V, NSP15 T144Q, NSP15 T166H, NSP15 V131stop, NSP15 V141F, NSP15 V148L, NSP15 V155stop, NSP15 V165S, NSP15 V172stop	G15602C, C15603G, T15606C, del15607_15607, T15642C, A15887C, ins16470A, A16471N, C16675T, C17550T, T18732K, del19971_19971, C19972N, ins20136A, C23123Y, C23260Y, C23298Y, A23312W, A23403G, A24123G, ins24124N, del24143_24143, T24144K, A24359M, A24684R, G24692T, A24694G, G24697A, G24698A, G25062C, A25216M, C25521T, G25913A, T26706Y, T27053C, T27054A, A27055C, T27056C, T27057G, A27060T, A27061G, A27062C, T27063C, T27093Y, T27740Y, C28278A, C28657T, G28881A, G28882A, G28883C, C29370T
10	32sT122620	hCoV-19/Colombia/SAN-576900-032-RNA/2020	EPI_ISL_1628499	GR	Spike D614G, Spike G1167A, N G204R, N R203K, N S2Y, N T366I, NS3 G174D, NSP3 A1215T, NSP3 L72F, NSP3 L357F, NSP3 P340L, NSP3 Q1125H, NSP6 V149F, NSP12 P323L	C241T, C2197T, C2933T, C3037T, C3738T, A3790T, T5707C, G6094C, G6362A, G9802T, G11417T, C11653T, C12473T, C12823T, C14408T, T15642C, C16041T, C17550T, A23403G, G25062C, C25521T, G25913A, C28278A, C28657T, G28881A, G28882A, G28883C, C29370T, A29852T, G29853A
11	36sT010821	hCoV-19/Colombia/SAN-576900-036-RNA/2021	EPI_ISL_1577390	GR	Spike D614G, Spike G1167A, N G204R, N R203K, N S2Y, N T366I, NS3 G174D, NSP3 L357F, NSP6 V149F, NSP12 D879Y, NSP12 P323L	T25C, C241T, C3037T, A3790T, C7279T, G9802T, G11417T, C11941T, C14408T, T15642C, G16075T, C17550T, A23403G, G25062C, C25521T, G25913A, C27213T, C28278A, G28881A, G28882A, G28883C, C29370T
12	36cD010821	hCoV-19/Colombia/SAN-576900-036-COM/2021	EPI_ISL_1630836	GR	Spike D614G, Spike G1167A, Spike S477N, N G204R, N R203K, N T366I, NS3 G174D, NS7a ins101FFY, NS7a L102F, NSP3 L357F, NSP4 Y441H, NSP5 V36I, NSP10 A32V, NSP12 P323L	C241T, C683T, C2449T, C2710T, C3037T, A3790T, C6310T, C7732T, C9430T, G9802T, T9875C, G10160A, C13119T, C14408T, C15720T, C17502T, G22992A, A23403G, G25062C, G25494T, C25521T, G25913A, ins27696TTTTTTTAT, C27697T, T27935C, G28881A, G28882A, G28883C, C29370T
13	41sT011921	hCoV-19/Colombia/SAN-576900-041-RNA/2021	EPI_ISL_1630319 2	GR	Spike D614G, Spike G1167A, N G204R, N R203K, N S2Y, N T366I, NS3 G174D, NS7b E33K, NSP3 L357F, NSP3 P125S, NSP3 Q1125H, NSP3 T64I, NSP3 Y801N, NSP6 V149F, NSP12 P323L, NSP15 E223G	A4T, A6C, G7T, T9C, T10G, A12C, A14T, C15G, C16G, C19S, C20A, A22T, G24A, C241T, C2910T, C3037T, C3092T, A3790T, ins5116T, T5120A, G6094C, G9802T, G11417T, C14408T, T15642C, G16647T, C17550T, A20288G, A23403G, G25062C, C25521T, G25913A, G27852A, C28278A, G28881A, G28882A, G28883C, C29370T, G29861T, G29862B, A29863H, G29864V, A29865T, A29866G
14	44sT012421	hCoV-19/Colombia/SAN-576900-044-RNA/2021	EPI_ISL_1629711	GH	Spike D614G, Spike E324Q, N S187L, NS3 Q57H, NSP3 A358V, NSP8 L184F, NSP12 P323L, NSP15 T48I	C241T, C3037T, C3241T, C3792T, C9430T, C10507T, C12641T, C14408T, T15537G, C18877T, C19763T, G22532C, A23403G, G25563T, C28253T, C28833T

Nota. Esta tabla muestra las sustituciones de aminoácidos y mutaciones de nucleótidos en los 14 genomas reportados en GISAID. No.: Número de identificación, ID: Identificador de acceso a las bases de datos GISAID, AA: Sustituciones de aminoácidos.

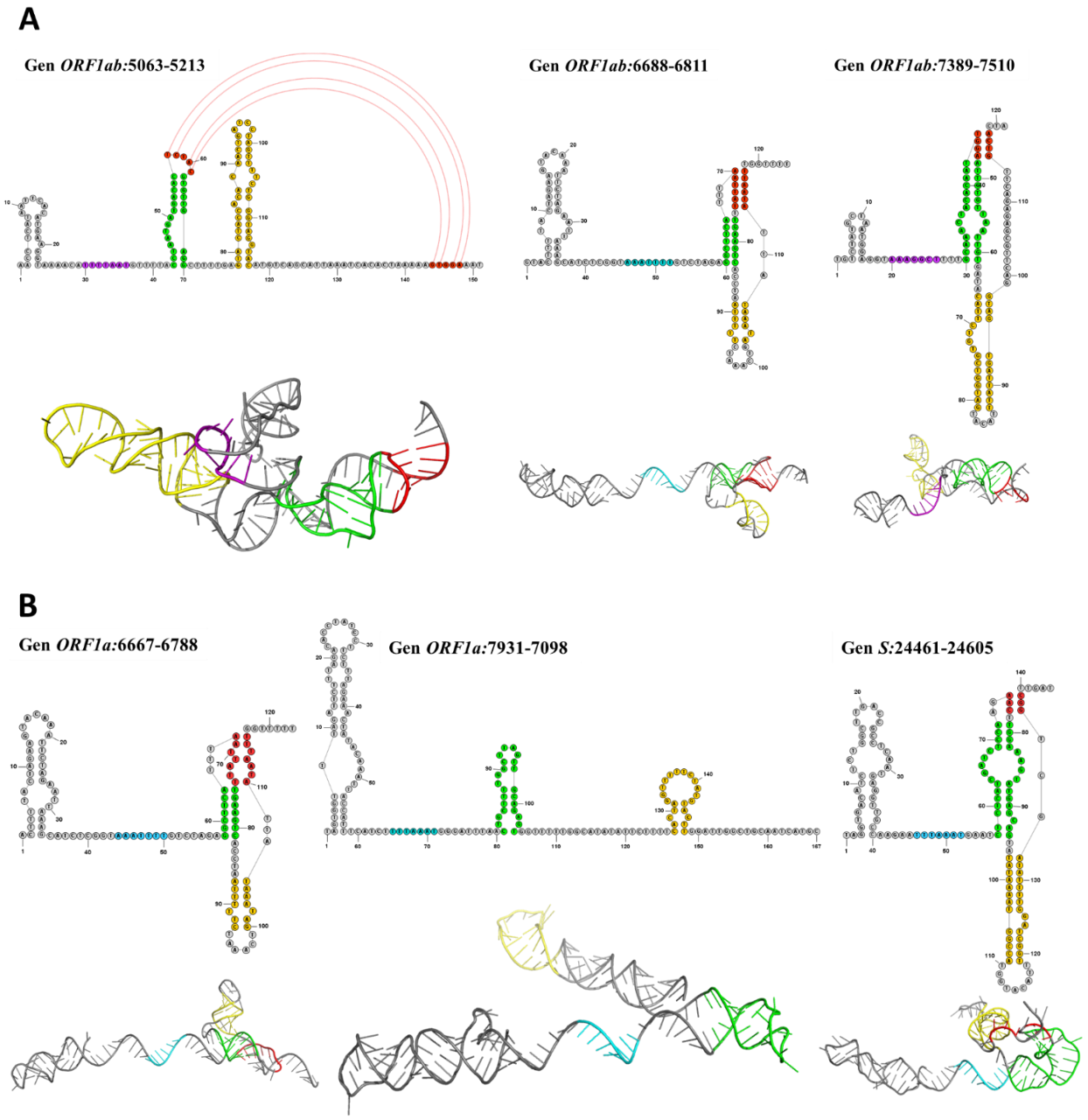
4.4 Identificación del modelo PRF en SARS-CoV-2

La identificación del PRF se llevó a cabo en los genomas **07dN120320** y **27sT122620** quienes para los ORFs con el codón de paro TAA estuvieron inmersos en una región nucleotídica de 133 pb en promedio, cuyos apareamiento canónicos y no canónicos generaron estructuras secundarias similares al modelo PRF. Los primeros cuarenta nucleótidos en promedio río arriba del sitio resbaladizo de las seis regiones determinadas de ambos genomas, la primera parte de la estructura secundaria corresponde a una horquilla atenuadora y rio abajo se produce un pseudonudo similar al pseudonudo de ARNm de tres tallos del modelo antes mencionado cuya estructura cambia entre ambos genomas (Figura 5).

Para el genoma **07dN120320**, la región 6688-6811 del gen *ORF1ab* genera el plegamiento canónico del modelo, en tanto que en las regiones 5063-5213 y 7389-7510 del mismo gen, la distancia que tienen los nucleótidos para el apareamiento con el bucle del primer tallo genera una estructura menos estable respecto a la primera descrita. Por otro lado, en el genoma **27sT122620** las regiones 6667-6788 del gen *ORF1a* y 24461-24605 del gen *S* cumplen con todos los elementos del pseudonudo. La región del 7931-7098 gen *ORF1a* generó dos tallos, pero la distancia entre ellos y la posible región de apareamiento para generar el tercer tallo no cumplen con la estructura canónica de la región del modelo (Figura 5, Apéndice F).

Figura 5

Estructuras de los genomas 07dN120320 y 27sT122620 para las regiones con PRF



Nota. A. **07dN120320** y B. **27sT122620**. Para ambas figuras, de izquierda a derecha se describen: el gen, la posición en nucleótidos, las estructuras secundarias 2D y 3D. De igual forma, se muestran: en azul el sitio resbaladizo canónico respecto a Kelly et al (2020), en gris la horquilla atenuadora, en verde, rojo y amarillo los tres tallos del pseudonudo del PRF, y en morado se muestra el posible sitio resbaladizo no canónico.

4.5 Base de datos y alineamientos

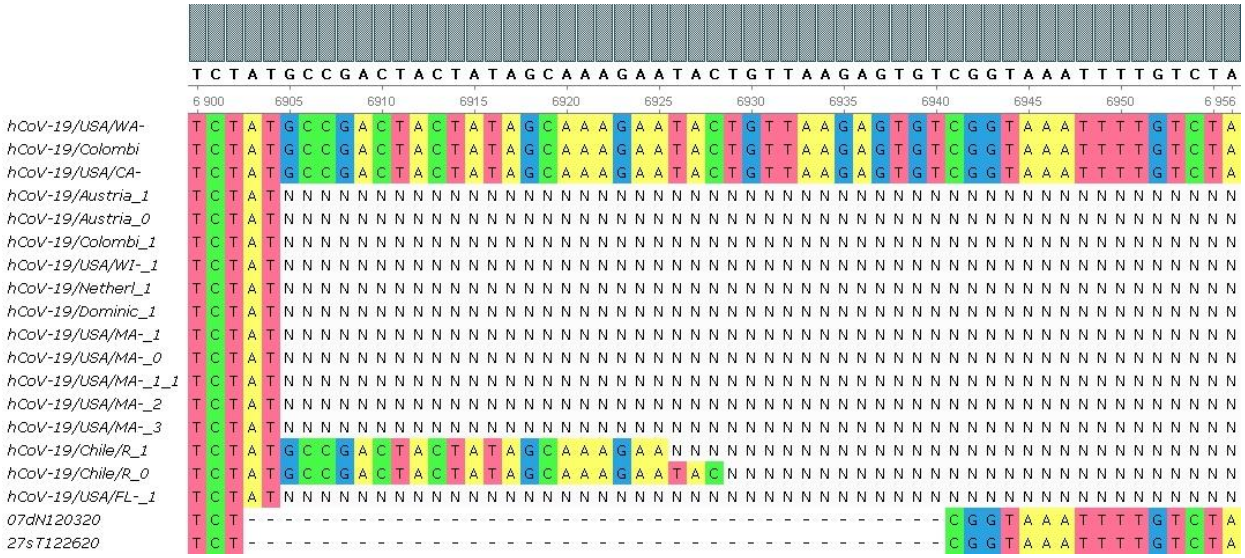
La comparación de la pérdida de regiones y mutaciones de los genomas **07dN120320** y **27sT122620** respecto a la base de datos generada con 80192 genomas de SARS-CoV-2 descargados de GISAID, se realizó de la siguiente manera: 1) De la base de datos anterior, se generó una segunda base de datos con 42357 genomas que no tuvieran 2 o más nucleótidos indeterminados seguidos en sus genomas y posteriormente, se eliminaron genomas que diferían considerablemente con respecto a los demás genomas por variante; con lo que se obtuvo una base de datos final de 38828 genomas. 2) Una tercera base de datos con las secuencias que contenían uno o más nucleótidos indeterminados que contuvo 40627 genomas (Apéndice E y G).

De esta forma se obtuvieron dos bloques de alineación, uno con nucleótidos determinados y otro con nucleótidos indeterminados. Para ambos casos, las secuencias se agruparon por variantes y se alinearon generando 18 secuencias consenso para las variantes Alpha, Beta, Gamma, Delta (1Delta: Norte, Centro y Sur América, Europa y Oceanía; y 2Delta: Caribe, Asia y África), Lambda, Mu, Omicron y GH/490R cada una de estas respecto al genoma de referencia (Apéndice G). Es de resaltar que, en los alineamientos con nucleótidos determinados no se encontraron regiones con pérdida, mientras que en los alineamientos con nucleótidos indeterminados estos se encontraron en regiones hipervariables de SARS-CoV-2 (Figura 6, Apéndice G).

Se encontró patrones de similitud entre las regiones con nucleótidos indeterminados de las variantes Lambda, Mu y Gamma en algunas regiones con pérdida de los genomas **07dN120320** y **27sT122620** (Figura 6, Apéndice G).

Figura 6

Patrón de pérdida de los genomas 07dN120320 y 27sT122620 con VOI y VOC de SARS-CoV-2



Nota. Esta figura muestra parte del alineamiento de las secuencias con nucleótidos indeterminados de la variante Lambda con los 14 genomas ensamblados. Se muestra evidencia de las regiones con pérdida correlacionadas respecto a los genomas **07dN120320** y **27sT122620** (Apéndice G).

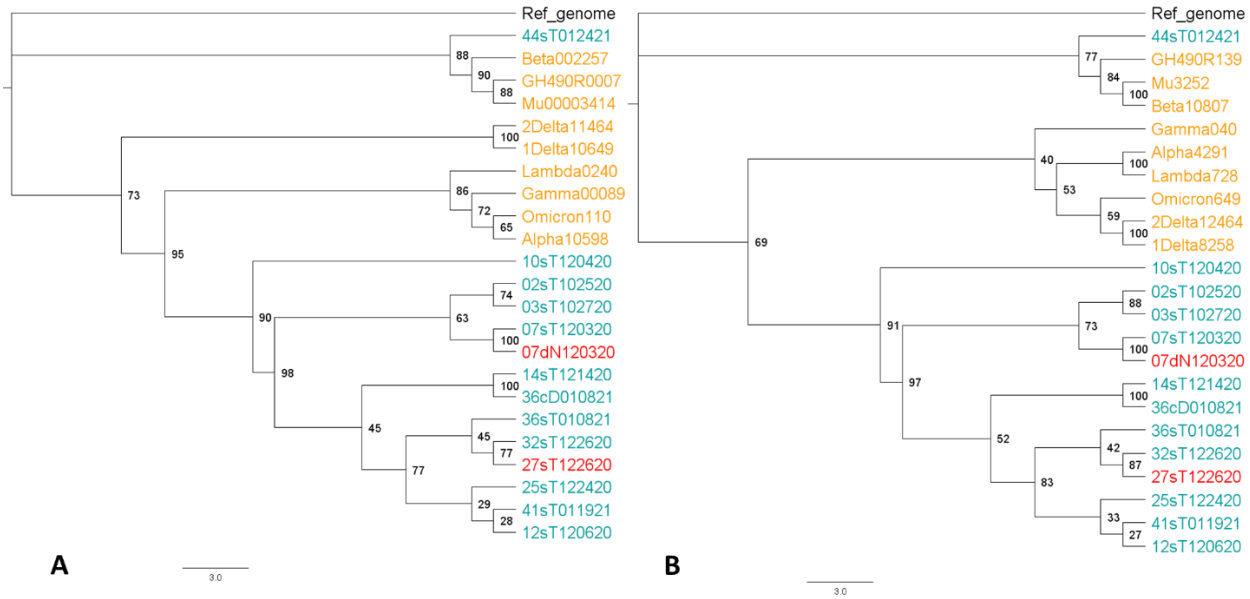
4.6 Relaciones filogenéticas entre los genomas y variantes de SARS-CoV-2

La filogenia de los 14 genomas obtenidos respecto a las variantes de SARS-CoV-2 generadas en la base de datos, mostró dos árboles para las secuencias consenso generadas a partir de genomas sin nucleótidos indeterminados y con nucleótidos indeterminadas (Figura 7 A y B, Apéndice H). Ambas filogenias incluyen los datos mostrados en la Tabla 4, para los 14 genomas secuenciados. En la Figura 7A, 13 genomas se agruparon en un nodo común con soporte de Bootstrap de 90%-91%, respectivamente. En donde, se destacan los genomas **07dN120320** y **27sT122620**, los cuales, presentaron el modelo de PRF y comparten el mismo nodo con la secuenciación del método estándar. A su vez, únicamente el genoma **44sT012421**, colectado el 24 de enero del 2021, comparte nodo para ambos árboles con las variantes Beta, GH/490R y Mu.

En tanto que, se observa una única diferencia entre las filogenias respecto al consenso de la variante Delta, ya que, para el árbol generado con genomas determinados (Figura 7A), estas dos secuencias se agruparon en nodos diferentes respecto a la filogenia con regiones indeterminadas (Figura 7B). Finalmente, el genoma de referencia en ambos casos presentó politomía respecto a la secuencia **44sT012421** y las variantes Beta, GH/490R y Mu (Figura 7, A y B).

Figura 7

Relaciones filogenéticas de los genomas obtenidos respecto a VOI y VOC de SARS-CoV-2



Nota. (A) Se muestran las relaciones filogenéticas sin nucleótidos indeterminados. (B) Se muestran las relaciones filogenéticas con nucleótidos indeterminados. Para ambos casos, las secuencias en naranja se representan los consensos de las variantes: Alpha, Beta, Gamma, Delta, Lambda, Mu, Omicron y GH/490R. En color azul, se representan las secuencias genómicas empleadas en este estudio reportadas en la base de datos GISAID. En rojo las secuencias **07dN120320** y **27sT122620** que presentan el modelo de PRF. El árbol fue realizado con máxima verosimilitud con el software IQ-TREE versión 1.6.12, se muestran los valores de Bootstrap para cada uno de los nodos y el genoma de referencia (NC_045512.2) se empleó como grupo externo. Las secuencias de la variante Delta se dividieron de la siguiente manera: 1Delta (América, Europa y Oceanía); y 2Delta (Caribe, Asia y África) debido al alto número de genomas.

5. Discusión

Uno de los principales logros para mitigar los efectos de la pandemia de SARS-CoV-2 (WHO, 2020) fue la caracterización de su genoma. Lo cual, permitió el desarrollo de los elementos necesarios para su diagnóstico por RT-PCR en tiempo real de un paso mediante el diseño de cebadores y sonda con soluciones coadyuvantes diseñados en función del genoma obtenido por Secuenciación de Nueva Generación (NGS) (Alyafei et al., 2022; Cadena-Caballero et al., 2022; Zhu et al., 2020).

La eficiencia del método aquí descrito quedó corroborada con el aumento en la concentración final de la síntesis del ADNc, en especial con el uso de la solución con concentraciones de nucleótidos (Tabla 1). Esto debido al cambio conformacional de la polimerasa, como se ha reportado en la síntesis del ADNc del genoma de otros virus, por ejemplo, Lentivirus (Coggins et al., 2019).

La correcta identificación y caracterización de las nuevas variantes genómicas de SARS-CoV-2, al igual que sus elementos reguladores como el modelo PRF implica el uso de HPC (Oujja et al., 2021) con algoritmos de ensamble genómico fiables. A raíz de la pandemia se ha diseñado diversas alternativas que optimizan el análisis de los datos de NGS, a partir de métodos de secuenciación basados en amplicones (Dezordi et al., 2022; Fritz et al., 2021; Luo et al., 2022; Rueca et al., 2022). Aquí empleamos IRMA y Bowtie 2 (Langmead y Salzberg, 2012; Shepard et al., 2016) con el fin de corroborar y contrastar los datos genómicos generados con la tecnología Ion Torrent, ya que, se ha reportado que se deben implementar dos o más enfoques de ensamblaje durante estudios de NGS viral, especialmente en entornos clínicos o en métodos experimentales de innovación (Gupta y Kumar, 2022).

Por otro lado, el modelo de PRF no solo se ha reportado para SARS-CoV-2 también se ha descrito para SARS-CoV, MERS, el Virus de la Inmunodeficiencia Humana (VIH) y el Virus del sarcoma de Rous (Ahn et al., 2021; Dinman, 2010; Jacks et al., 1988; Jacks y Varmus, 1985) Nuestros resultados concuerdan con los descrito por Dinman (2010), para SARS-CoV quien demuestra que la longitud y la composición de bases varían en la secuencia espaciadora y son parámetros cruciales para determinar el alcance del PRF de una manera específica en *Betacoronavirus*.

El ensamble de los genomas mostró que, aunque la mayoría de las muestras en las que se obtuvo mayor cantidad de lecturas presentaron una mayor cobertura, no hubo correlación entre el número de lecturas obtenidas en cada secuenciación con la cobertura de cada genoma. Lo anterior, está en concordancia con los resultados obtenidos por Plitnick et al (2021), donde, comparando las tecnologías *Ion Torrent AmpliSeq e Illumina MiSeq ARTIC Protocol*, se observaron niveles de cobertura similares (>98%) para 81/83 muestras primarias que se secuenciaron con ambos métodos.

En la síntesis de ADNc con la inclusión de la solución dR, aunque tuvo unas lecturas con baja cobertura, estas mantuvieron una calidad (Q) por base alta, tratándose posiblemente de ARN subgenómico (ARNsg) (Apéndice E). Así mismo, se encontró un patrón similar en las muestras con la adición de la solución dN, a excepción de la muestra **07dN120320** que, aunque mostró menos lecturas obtuvo una alta cobertura, dándole peso a la hipótesis de que se traten de subgenomas o ARNsg, tal como ha sido reportado en otros estudios (Kim et al., 2020; Long, 2021; Nomburg et al., 2020; Sawicki et al., 2007). Por lo que, aquellos genomas con menos del 90% de cobertura con respecto al genoma de referencia podrían considerarse como ARNsg y son actualmente validados por el grupo de investigación CAGE (Tabla 1, Apéndice E).

El alineamiento tipo BLAST confirmó que, independientemente de su tamaño, la mayoría de los genomas presentan identidad cercana al 100% respecto a otros genomas de SARS-CoV-2 reportados en el GenBank. Con respecto a la clasificación de linajes de acuerdo a la nomenclatura PANGO, se encontró que los linajes obtenidos en nuestras muestras se distribuyeron geográficamente de la siguiente manera: B.1.1.348 Sur América y Estados Unidos; B.1.1.203 Inglaterra y Ecuador; B.1.1.371 en Arabia Saudita, Rumania y Europa occidental, además, contiene B.1.1.161 en la base (Inglaterra); B.1.111 linaje de Sur y Centro América (principalmente Trinidad y Colombia) y un pequeño grupo en Norwich (Inglaterra); B.1.1 linaje europeo con 3 Polimorfismos de Nucleótido Único (SNP) claros “28881GA”, “28882GA”, “28883GC”; y B.1 un gran linaje europeo cuyo origen corresponde al brote del norte de Italia a principios de 2020 (O’Toole, Hill, et al., 2021).

Los genes que presentaron mayor pérdida de regiones para los genomas **07dN120320** y **27sT122620** fueron *ORF1ab* y *S*. Para dar continuidad a la traducción de la segunda parte del gen *ORF1ab* es necesario el PRF, sin embargo, se ha probado que, este suceso tiene una eficiencia entre el 15% y 75%, de acuerdo al sistema de ensayo empleado (Bhatt et al., 2021; Kelly et al., 2020, 2021); considerado por Kelly et al., (2021), una ventaja en tiempo para el virus, ya que al acumular altas concentraciones de las NSPS codificadas por *ORF1a*, ayuda a incapacitar eficientemente la respuesta inmune innata de la célula huésped, permitiendo, al momento de alcanzar una concentración adecuada, una síntesis explosiva de ARN en el momento idóneo. Ahora bien, con respecto al gen *S*: el ARN genómico de SARS-CoV-2 de sentido negativo es un intermediario de la replicación y transcripción, y sirve como molde para la síntesis de ARN genómico de sentido positivo y ARNsg, que codifican las proteínas estructurales y son generados a partir de la transcripción discontinua del ARN genómico de sentido negativo (Long, 2021).

Sin embargo, se han observado en varios estudios empalmes de lecturas no esperadas formando los llamados ARNsg no canónicos (Kim et al., 2020; Long, 2021; Nomburg et al., 2020); que si bien, se han encontrado en diversidad y abundancia, su papel en la evolución y en el ciclo viral sigue en debate (Kim et al., 2020). El mismo evento puede explicar las pérdidas encontradas en los genes *M* y *ORF6a*, así como, las grandes pérdidas en más del 50% del genoma total del resto de genomas analizados en este trabajo.

Entre los dos genomas **07dN120320** y **27sT122620** que presentaron mayor pérdida de los 14 seleccionados, solo en 4 de las 6 estructuras secundarias obtenidas se encontró el sitio resbaladizo con la conformación NNNWWWH (Kelly et al., 2020; Rodnina et al., 2020), dos con AAATTTT, una en el genoma **07dN120320** y dos con TTTAAAT, estas últimas solo en el genoma **27sT122620**. Sin embargo, de acuerdo con los valores de la energía libre Gibbs (ΔG), 5 estructuras secundarias poseen una estructura estable, a excepción de la segunda estructura del genoma **27sT122620,** la cual posee una energía libre muy por encima del 0, lo que indicaría de forma *in silico* que es una estructura poco probable (Mathews y Turner, 2006).

Por otra parte, el clado "G" es la variante con la sustitución aminoacídica en S: D614G que le confiere una mayor infectividad y eficiencia de transmisión al virus. Los clados GH y GR son los descendientes más comunes de este clado. El clado GR, es portador de la combinación de sustituciones N: RG203KR, NSP3: F106F y S: D614G. Y el clado GH lleva las sustituciones ORF3a: Q57H, NSP12b: P314L y S: D614G. (Mercatelli y Giorgi, 2020; Sengupta et al., 2021). De esta forma, aunque el genoma **44sT012421** fue el único que perteneció al clado GH, los 14 genomas seleccionados presentaron la sustitución NSP12b: P314L propio del clado GH, sin embargo, las sustituciones N: R203K y G204R, que son específicas del clado GR es compartida por los 13 genomas a excepción de **44sT012421,** lo que explica la clasificación de este genoma en

el clado GH por GISAID. En consecuencia, que los genomas pertenezcan a los clados descendientes del clado G (GR o GH) no es sorprendente, ya que, a mediados del 2020, fecha en la que fueron tomadas las muestras, los clados más comunes entre los genomas secuenciados de SARS-CoV-2 fueron el clado G y su descendencia, los clados GH y GR, correspondiendo al 74% de todas las secuencias mundiales (Mercatelli y Giorgi, 2020; Hamed et al., 2021).

En relación con los aspectos evolutivos de los genomas descritos y reportados en la base de datos mundial de GISAID, nuestros resultados se correlacionan con otros estudios para la región y para Sudamérica en relación con la prevalencia de las variantes B.1.1 y B.1.1.348 (Castañeda et al., 2021; Ortiz-Pineda y Sierra-Torres, 2022; Ramírez et al., 2021; Ribeiro Dias et al., 2023). Sin embargo, la muestra **44sT012421** colectada el 24 de enero del 2021 en el departamento de Santander, se asocia a la variante Mu (Halfmann et al., 2022; Pascarella et al., 2022; F. Rahimi et al., 2022; Uriu et al., 2022) descrita para Colombia a finales del mismo mes y designada como VOI por la OMS el 31 de agosto del 2021 (WHO, 2022). Esto indica la eficiencia del método aquí descrito, el cual, permite no solo identificar genomas con el modelo de PRF, sino también, caracterizar VOI y VOC de SARS-CoV-2 por NSG empleando Ion Torrent.

6. Conclusiones

La caracterización de los genomas de SARS-CoV-2 con la tecnología Ion Torrent permitió la identificación de las regiones con pérdida en los codones de los genomas secuenciados con la Tecnología Genómica UIS, a partir de muestras de hisopado nasofaríngeo obtenidas de pacientes santandereanos clínicamente diagnosticados.

La identificación de los ORFs en los genes de los genomas de SARS-CoV-2 obtenidos con la Tecnología Genómica UIS permitió corroborar los cambios nucleotídicos específicos para cada uno de los linajes y variantes del virus.

El determinar *in silico* los codones que pudieran favorecer la traducción por medio del modelo PRF en los genomas y subgenomas del virus, debido a las pérdidas y presencia de codones de paro que truncarían los ORFs, haciéndolos inviables, da la posibilidad de descubrir el papel de los ARNsg en el ciclo viral y la evolución de SARS-CoV-2.

La comparación de los genomas secuenciados de este estudio con respecto a los genomas con nucleótidos indeterminados obtenidos de las variantes reportadas en GISAID evidenció un patrón de pérdida de nucleótidos, posiblemente regulado por el modelo de PRF. Así, la existencia de este patrón demuestra la viabilidad de SARS-CoV-2 con un genoma reducido, y con ello el camino evolutivo que puede estar siguiendo el virus.

Se logró determinar que la pérdida de regiones en los genomas obtenidos con la tecnología Genómica UIS son resultado de la polimerización de la tecnología Ion Torrent y no producto del ensamble computacional. Soportado con la identificación del modelo PRF en algunas de estas pérdidas, además del hallazgo de genomas con un patrón similar de pérdida en otros estudios.

7. Recomendaciones

Emplear métodos de NGS que permitan obtener genomas completos de SARS-CoV-2.

Utilizar los agentes coadyuvantes y dNTPs para mejorar la secuenciación en diferentes plataformas de NGS, como Ilumina y Oxford Nanopore.

Corroborar por RT-PCR en tiempo real de un paso la calidad de la muestra que se va a secuenciar para aumentar la probabilidad de caracterizar nuevos linajes y/o variantes de SARS-CoV-2.

Generar nuevos métodos bioinformáticos, desarrollar nuevas alternativas en software y utilizar HPC para optimizar los análisis evolutivos de SARS-CoV-2.

Vigilancia genómica constante de SARS-CoV-2 por medio de HPC y NGS empleando la Tecnología Genómica UIS, para identificar y diseñar estrategias que contrarresten los efectos de nuevas variantes de SARS-CoV-2 y evitar posibles pandemias.

Referencias Bibliográficas

Acuti Martellucci, C., Flacco, M. E., Cappadona, R., Bravi, F., Mantovani, L., y Manzoli, L. (2020). SARS-CoV-2 pandemic: An overview. *Advances in Biological Regulation*, *77*, 100736. https://doi.org/10.1016/j.jbior.2020.100736

Ahn, D. G., Yoon, G. Y., Lee, S., Ku, K. B., Kim, C., Kim, K. D., Kwon, Y. C., Kim, G. W., Kim, B. T., y Kim, S. J. (2021). A novel frameshifting inhibitor having antiviral activity against zoonotic Coronaviruses. *Viruses*, *13*(8), 1639. https://doi.org/10.3390/v13081639

Aksamentov, I., Roemer, C., Hodcroft, E. B., y Neher, R. A. (2021). Nextclade: clade assignment, mutation calling and quality control for viral genomes. *Journal of Open Source Software*, *6*(67), 3773. https://doi.org/10.21105/joss.03773

Álvarez-Díaz, D. A., Laiton-Donato, K., Franco-Muñoz, C., y Mercado-Reyes, M. (2020). Secuenciación del SARS-CoV-2: la iniciativa tecnológica para fortalecer los sistemas de alerta temprana ante emergencias de salud pública en Latinoamérica y el Caribe. *Biomédica*, *40*(2), 188–197. https://doi.org/10.7705/biomedica.5841

Alyafei, K., Ahmed, R., Abir, F. F., Chowdhury, M. E. H., y Naji, K. K. (2022). A comprehensive review of COVID-19 detection techniques: From laboratory systems to wearable devices. *Computers in Biology and Medicine*, *149*, 106070. https://doi.org/10.1016/j.compbiomed.2022.106070

Andrews, S. (2010). FASTQC. A quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Antczak, M., Popenda, M., Zok, T., Sarzynska, J., Ratajczak, T., Tomczyk, K., Adamiak, R. W., y Szachniuk, M. (2017). New functionality of RNAComposer: application to shape the axis

of miR160 precursor structure. *Acta Biochimica Polonica*, *63*(4). https://doi.org/10.18388/abp.2016_1329

Ashraf, O., Virani, A., y Cheema, T. (2021). COVID-19: An update on the epidemiological, clinical, preventive, and therapeutic management of 2019 novel coronavirus disease. *Critical Care Nursing Quarterly*, *44*(1), 128. https://doi.org/10.1097/cnq.0000000000000346

Bhatt, P. R., Scaiola, A., Loughran, G., Leibundgut, M., Kratzel, A., Meurs, R., Dreos, R., O'Connor, K. M., McMillan, A., Bode, J. W., Thiel, V., Gatfield, D., Atkins, J. F., y Ban, N. (2021). Structural basis of ribosomal frameshifting during translation of the SARS-CoV-2 RNA genome. *Science*, *372*(6548), 1306–1313. https://doi.org/10.1126/science.abf3546

Boratyn, G. M., Camacho, C., Cooper, P. S., Coulouris, G., Fong, A., Ma, N., Madden, T. L., Matten, W. T., McGinnis, S. D., Merezhuk, Y., Raytselis, Y., Sayers, E. W., Tao, T., Ye, J., y Zaretskaya, I. (2013). BLAST: a more efficient report with usability improvements. *Nucleic Acids Research*, *41*(W1), W29–W33. https://doi.org/10.1093/nar/gkt282

Cadena-Caballero, C. E., Vera-Cala, L. M., Barrios-Hernandez, C., Rueda-Plata, D., Forero-Buitrago, L. J., Torres-Jimenez, C. S., Lizarazo-Gutierrez, E., Agudelo-Rodriguez, M., Martinez-Perez, F., y Beggs, A. D. (2022). Denaturing and dNTPs reagents improve SARS-CoV-2 detection via single and multiplex RT-qPCR. *F1000Research*, *11*(331), 331. https://doi.org/10.12688/f1000research.109673.1

Castañeda, S., Patiño, L. H., Muñoz, M., Ballesteros, N., Guerrero-Araya, E., Paredes-Sabja, D., Flórez, C., Gomez, S., Ramírez-Santana, C., Salguero, G., Gallo, J. E., Paniz-Mondolfi, A. E., y Ramírez, J. D. (2021). Evolution and epidemic spread of SARS-CoV-2 in Colombia: A year into the pandemic. *Vaccines*, *9*(8), 837. https://doi.org/10.3390/vaccines9080837

Chiara, M., D'Erchia, A. M., Gissi, C., Manzari, C., Parisi, A., Resta, N., Zambelli, F., Picardi, E.,

Pavesi, G., Horner, D. S., y Pesole, G. (2021). Next generation sequencing of SARS-CoV-2 genomes: challenges, applications and opportunities. *Briefings in Bioinformatics*, *22*(2), 616–630. https://doi.org/10.1093/bib/bbaa297

Cock, P. J. A., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., y De Hoon, M. J. L. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, *25*(11), 1422–1423. https://doi.org/10.1093/bioinformatics/btp163

Coggins, S. A., Holler, J. M., Kimata, J. T., Kim, D.-H., Schinazi, R. F., y Kim, B. (2019). Efficient pre-catalytic conformational change of reverse transcriptases from SAMHD1 non-counteracting primate lentiviruses during dNTP incorporation. *Virology*, *537*, 36–44. https://doi.org/10.1016/j.virol.2019.08.010

de Wit, E., van Doremalen, N., Falzarano, D., y Munster, V. J. (2016). SARS and MERS: recent insights into emerging Coronaviruses. *Nature Reviews Microbiology*, *14*(8), 523–534. https://doi.org/10.1038/nrmicro.2016.81

Dezordi, F. Z., Neto, A. M. da S., Campos, T. de L., Jeronimo, P. M. C., Aksenen, C. F., Almeida, S. P., y Wallau, G. L. (2022). ViralFlow: A versatile automated workflow for SARS-CoV-2 genome assembly, lineage assignment, mutations and intrahost variant detection. *Viruses*, *14*(2), 217. https://doi.org/10.3390/v14020217

Dinman, J. D. (2010). Programmed –1 Ribosomal Frameshifting in SARS Coronavirus. *Molecular Biology of the SARS-Coronavirus*, 63–72. https://doi.org/10.1007/978-3-642-03683-5_5

Dinman, J. D. (2012). Mechanisms and implications of programmed translational frameshifting. *Wiley Interdisciplinary Reviews: RNA*, *3*(5), 661–673. https://doi.org/10.1002/wrna.1126

Duvaud, S., Gabella, C., Lisacek, F., Stockinger, H., Ioannidis, V., y Durinx, C. (2021). Expasy,

the Swiss bioinformatics resource portal, as designed by its users. *Nucleic Acids Research*, *49*(W1), W216–W227. https://doi.org/10.1093/nar/gkab225

Fritz, A., Bremges, A., Deng, Z.-L., Lesker, T. R., Götting, J., Ganzenmueller, T., Sczyrba, A., Dilthey, A., Klawonn, F., y McHardy, A. C. (2021). Haploflow: strain-resolved *de novo* assembly of viral genomes. *Genome Biology*, *22*(1), 212. https://doi.org/10.1186/s13059-021-02426-8

Gilchrist, C. A., Turner, S. D., Riley, M. F., Petri, W. A., y Hewlett, E. L. (2015). Whole-Genome sequencing in outbreak analysis. *Clinical Microbiology Reviews*, *28*(3), 541–563. https://doi.org/10.1128/cmr.00075-13

Gorbalenya, A. E., Baker, S. C., Baric, R. S., de Groot, R. J., Drosten, C., Gulyaeva, A. A., Haagmans, B. L., Lauber, C., Leontovich, A. M., Neuman, B. W., Penzar, D., Perlman, S., Poon, L. L. M., Samborskiy, D. V., Sidorov, I. A., Sola, I., y Ziebuhr, J. (2020). The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nature Microbiology*, *5*(4), 536–544. https://doi.org/10.1038/s41564-020-0695-z

Grubaugh, N. D., Ladner, J. T., Lemey, P., Pybus, O. G., Rambaut, A., Holmes, E. C., y Andersen, K. G. (2019). Tracking virus outbreaks in the twenty-first century. *Nature microbiology*, *4*(1), 10–19. https://doi.org/10.1038/s41564-018-0296-2

Gupta, A. K., y Kumar, M. (2022). Benchmarking and assessment of eight *de novo* genome assemblers on viral Next-Generation Sequencing data, including the SARS-CoV-2. OMICS: *A Journal of Integrative Biology*, *2*6(7), 372–381. https://doi.org/10.1089/omi.2022.0042

Halfmann, P. J., Kuroda, M., Armbrust, T., Theiler, J., Balaram, A., Moreno, G. K., Accola, M. A., Iwatsuki-Horimoto, K., Valdez, R., Stoneman, E., Braun, K., Yamayoshi, S., Somsen, E.,

Baczenas, J. J., Mitamura, K., Hagihara, M., Adachi, E., Koga, M., McLaughlin, M., … Kawaoka, Y. (2022). Characterization of the SARS-CoV-2 B.1.621 (Mu) variant. *Science Translational Medicine*, *14*(657). https://doi.org/10.1126/scitranslmed.abm4908

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, *41*, 95–98. https://doi.org/10.14601/phytopathol_mediterr-14998u1.29

Hamed, S. M., Elkhatib, W. F., Khairalla, A. S., y Noreddin, A. M. (2021). Global dynamics of SARS-CoV-2 clades and their relation to COVID-19 epidemiology. *Scientific Reports*, 11(1), 8435. https://doi.org/10.1038/s41598-021-87713-x

Hernandez-Ortiz, J., Cardona, A., Ciuoderis, K., Averhoff, F., Maya, M.-A., Cloherty, G., y Osorio, J. E. (2022). Assessment of SARS-CoV-2 Mu variant emergence and spread in Colombia. *JAMA Network Open*, *5*(3), e224754. https://doi.org/10.1001/jamanetworkopen.2022.4754

Jacks, T., Power, M. D., Masiarz, F. R., Luciw, P. A., Barr, P. J., y Varmus, H. E. (1988). Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. *Nature*, *331*(6153), 280–283. https://doi.org/10.1038/331280a0

Jacks, T., y Varmus, H. E. (1985). Expression of the Rous Sarcoma Virus pol Gene by Ribosomal Frameshifting. *Science*, *230*(4731), 1237–1242. https://doi.org/10.1126/science.2416054

Katoh, K., y Standley, D. M. (2013). MAFFT multiple sequence alignment software Version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

Kelly, J. A., Olson, A. N., Neupane, K., Munshi, S., Emeterio, J. S., Pollack, L., Woodside, M. T., y Dinman, J. D. (2020). Structural and functional conservation of the programmed −1

ribosomal frameshift signal of SARS coronavirus 2 (SARS-CoV-2). *The Journal of Biological Chemistry*, *295*(31), 10741. https://doi.org/10.1074/jbc.ac120.013449

Kelly, J. A., Woodside, M. T., y Dinman, J. D. (2021). Programmed −1 Ribosomal Frameshifting in Coronaviruses: A therapeutic target. *Virology*, *554*, 75–82. https://doi.org/10.1016/j.virol.2020.12.010

Kesheh, M. M., Hosseini, P., Soltani, S., y Zandi, M. (2022). An overview on the seven pathogenic human Coronaviruses. *Reviews in Medical Virology*, *32*(2), e2282. https://doi.org/10.1002/rmv.2282

Khailany, R. A., Safdar, M., y Ozaslan, M. (2020). Genomic characterization of a novel SARS-CoV-2. *Gene Reports*, *19*, 100682. https://doi.org/10.1016/j.genrep.2020.100682

Kim, D., Lee, J. Y., Yang, J. S., Kim, J. W., Kim, V. N., y Chang, H. (2020). The Architecture of SARS-CoV-2 Transcriptome. *Cell*, *181*(4), 914-921. https://doi.org/10.1016/j.cell.2020.04.011

Langmead, B., y Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Long, S. (2021). SARS-CoV-2 Subgenomic RNAs: characterization, utility, and perspectives. *Viruses*, *13*(10), 1923. https://doi.org/10.3390/v13101923

Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., Wang, W., Song, H., Huang, B., Zhu, N., Bi, Y., Ma, X., Zhan, F., Wang, L., Hu, T., Zhou, H., Hu, Z., Zhou, W., Zhao, L., … Tan, W. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The lancet*, *395*(10224), 565-674. https://doi.org/10.1016/s0140-6736(20)30251-8

Luo, X., Kang, X., y Schönhuth, A. (2022). Strainline: full-length de novo viral haplotype

reconstruction from noisy long reads. *Genome Biology*, *23*(1), 29. https://doi.org/10.1186/s13059-021-02587-6

Martínez-Pérez, F., Durán-Gutiérrez, D., Delaye, L., Becerra, A., Aguilar, G., y Zinker, S. (2007). Loss of DNA: A plausible molecular level explanation for crustacean neuropeptide gene evolution. *Peptides*, *28*(1), 76–82. https://doi.org/10.1016/j.peptides.2006.09.021

Mathews, D. H., y Turner, D. H. (2006). Prediction of RNA secondary structure by free energy minimization. *Current Opinion in Structural Biology*, *16*(3), 270–278. https://doi.org/10.1016/j.sbi.2006.05.010

Mejía-Ospino, E., Barrios-Hernández, C. J., Vera-Cala, L. M., Ribón-Gómez, W. A., Martínez-Pérez, F. J., Bautista-Rozo, L. X., Pedraza-Ferreira, G. R., Munive-Argüelles, N. M., Ramírez-Ardila, S. D., Tobe, S., Rodríguez-Vázquez, R., González-Barrios, J. A., y Martínez-Fong, D. (2021). Mezcla de nucelotidos para la amplificacion y secuenciacion de polimeros de acidos nucleicos (Patente de Colombia. No. NC2017/0004500). Universidad Industrial de Santander. Superintendencia de Industria y Comercio.

Mercatelli, D., y Giorgi, F. M. (2020). Geographic and genomic distribution of SARS-CoV-2 mutations. *Frontiers in Microbiology*, *11*. https://doi.org/10.3389/fmicb.2020.01800

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., y Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating Maximum-Likelihood phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274. https://doi.org/10.1093/molbev/msu300

Nicholas, K., y Nicholas, H. (1997). GeneDoc: a tool for editing and annotating multiple sequence alignments.

Nomburg, J., Meyerson, M., y DeCaprio, J. A. (2020). Pervasive generation of non-canonical subgenomic RNAs by SARS-CoV-2. *Genome Medicine*, *12*(1), 108.

https://doi.org/10.1186/s13073-020-00802-w

O'Toole, Á., Hill, V., Pybus, O. G., Watts, A., Bogoch, I. I., Khan, K., Messina, J. P., Tegally, H., Lessells, R. R., Giandhari, J., Pillay, S., Tumedi, K. A., Nyepetsi, G., Kebabonye, M., Matsheka, M., Mine, M., Tokajian, S., Hassan, H., Salloum, T., … Kraemer, M. U. G. (2021). Tracking the international spread of SARS-CoV-2 lineages B.1.1.7 and B.1.351/501Y-V2. *Wellcome Open Research*, *6*, 121. https://doi.org/10.12688/wellcomeopenres.16661.1

O'Toole, Á., Scher, E., Underwood, A., Jackson, B., Hill, V., McCrone, J. T., Colquhoun, R., Ruis, C., Abu-Dahab, K., Taylor, B., Yeats, C., du Plessis, L., Maloney, D., Medd, N., Attwood, S. W., Aanensen, D. M., Holmes, E. C., Pybus, O. G., y Rambaut, A. (2021). Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. *Virus Evolution*, *7*(2). https://doi.org/10.1093/ve/veab064

Ortiz-Pineda, P. A., y Sierra-Torres, C. H. (2022). Evolutionary traits and genomic surveillance of SARS-CoV-2 in South America. *Global Health*, *2022*, 1–9. https://doi.org/10.1155/2022/8551576

Pascarella, S., Bianchi, M., Giovanetti, M., Narzi, D., Cauda, R., Cassone, A., y Ciccozzi, M. (2022). The SARS-CoV-2 Mu variant should not be left aside: It warrants attention for its immuno-escaping ability. *Journal of Medical Virology*, *94*(6), 2479–2486. https://doi.org/10.1002/jmv.27663

Plitnick, J., Griesemer, S., Lasek-Nesselquist, E., Singh, N., Lamson, D. M., y St. George, K. (2021). Whole-genome sequencing of SARS-CoV-2: Assessment of the Ion Torrent AmpliSeq panel and comparison with the Illumina MiSeq ARTIC protocol. *Journal of Clinical Microbiology*, *59*(12). https://doi.org/10.1128/jcm.00649-21

Popenda, M., Szachniuk, M., Antczak, M., Purzycka, K. J., Lukasiak, P., Bartol, N., Blazewicz,

J., & Adamiak, R. W. (2012). Automated 3D structure composition for large RNAs. *Nucleic Acids Research, 40*(14). https://doi.org/10.1093/nar/gks339

Rahimi, A., Mirzazadeh, A., y Tavakolpour, S. (2021). Genetics and genomics of SARS-CoV-2: A review of the literature with the special focus on genetic diversity and SARS-CoV-2 genome detection. *Genomics*, *113*(1), 1221. https://doi.org/10.1016/j.ygeno.2020.09.059

Rahimi, F., Kamali, N., y Bezmin Abadi, A. T. (2022). The Mu strain: the last but not least circulating 'variant of interest' potentially affecting the COVID-19 pandemic. *Future Virology*, *17*(1), 5–8. https://doi.org/10.2217/fvl-2021-0269

Rambaut, A. (2012). FigTree v1. 4. Molecular evolution, phylogenetics and epidemiology. *Edinburgh: University of Edinburgh, Institute of Evolutionary Biology*.

Rambaut, A., Holmes, E. C., O'Toole, Á., Hill, V., McCrone, J. T., Ruis, C., du Plessis, L., y Pybus, O. G. (2020). A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nature Microbiology*, *5*(11), 1403–1407. https://doi.org/10.1038/s41564-020-0770-5

Ramírez, J. D., Florez, C., Muñoz, M., Hernández, C., Castillo, A., Gomez, S., Rico, A., Pardo, L., Barros, E. C., Castañeda, S., Ballesteros, N., Martínez, D., Vega, L., Jaimes, J. E., Cruz-Saavedra, L., Herrera, G., Patiño, L. H., Teherán, A. A., Gonzalez-Reiche, A. S., … y Paniz-Mondolfi, A. (2021). The arrival and spread of SARS-CoV-2 in Colombia. *Journal of Medical Virology*, *93*(2), 1158–1163. https://doi.org/10.1002/jmv.26393

Reuter, J. S., y Mathews, D. H. (2010). RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics*, *11*(1), 129. https://doi.org/10.1186/1471-2105-11-129

Ribeiro Dias, M. F., Andriolo, B. V., Silvestre, D. H., Cascabulho, P. L., y Leal da Silva, M.

(2023). Genomic surveillance and sequencing of SARS-CoV-2 across South America. *Revista Panamericana de Salud Pública*, *47*(1), 2. https://doi.org/10.26633/rpsp.2023.21

Rodnina, M. V, Korniy, N., Klimova, M., Karki, P., Peng, B.-Z., Senyushkina, T., Belardinelli, R., Maracci, C., Wohlgemuth, I., Samatova, E., y Peske, F. (2020). Translational recoding: canonical translation mechanisms reinterpreted. *Nucleic Acids Research*, *48*(3), 1056–1067. https://doi.org/10.1093/nar/gkz783

Rojas-Pineda, E. (2020). Invitación a presentar proyectos que contribuyan a la solución de problemáticas actuales de salud relacionadas con la pandemia de covid-19. MinCiencias. https://minciencias.gov.co/convocatorias/invitacion-para-presentacion-propuestas/invitacion-presentar-proyectos-que-contribuyan

Roman, C., Lewicka, A., Koirala, D., Li, N.-S., y Piccirilli, J. A. (2021). The SARS-CoV-2 programmed −1 ribosomal frameshifting element crystal structure solved to 2.09 Å using Chaperone-Assisted RNA Crystallography. *ACS Chemical Biology*, *16*(8), 1469-1481. https://doi.org/10.1021/acschembio.1c00324

Rueca, M., Giombini, E., Messina, F., Bartolini, B., Di Caro, A., Capobianchi, M. R., y Gruber, C. E. (2022). The easy-to-use SARS-CoV-2 assembler for genome sequencing: Development study. *JMIR Bioinformatics and Biotechnology*, *3*(1), e31536. https://doi.org/10.2196/31536

Sawicki, S. G., Sawicki, D. L., y Siddell, S. G. (2007). A contemporary view of Coronavirus transcription. *Journal of Virology*, *81*(1), 20–29. https://doi.org/10.1128/jvi.01358-06

Sayers, E. W., Cavanaugh, M., Clark, K., Pruitt, K. D., Schoch, C. L., Sherry, S. T., y Karsch-Mizrachi, I. (2021). GenBank. *Nucleic Acids Research*, *49*(D1), D92–D96. https://doi.org/10.1093/nar/gkaa1023

Sengupta, A., Hassan, S. S., y Choudhury, P. P. (2021). Clade GR and clade GH isolates of SARS-

CoV-2 in Asia show highest amount of SNPs. *Infection, Genetics and Evolution*, *89*, 104724. https://doi.org/10.1016/j.meegid.2021.104724

Shepard, S. S., Meno, S., Bahl, J., Wilson, M. M., Barnes, J., y Neuhaus, E. (2016). Viral deep sequencing needs an adaptive approach: IRMA, the iterative refinement meta-assembler. *BMC Genomics*, *17*(1), 1–18. https://doi.org/10.1186/s12864-016-3030-6/figures/7

Shu, Y., y McCauley, J. (2017). GISAID: Global initiative on sharing all influenza data – from vision to reality. *Eurosurveillance*, *22*(13), 30494. https://doi.org/10.2807/1560-7917.es.2017.22.13.30494/cite/plaintext

Tao, K., Tzou, P. L., Nouhin, J., Gupta, R. K., de Oliveira, T., Kosakovsky Pond, S. L., Fera, D., y Shafer, R. W. (2021). The biological and clinical significance of emerging SARS-CoV-2 variants. *Nature Reviews Genetics*, *22*(12), 757–773. https://doi.org/10.1038/s41576-021-00408-x

Torres-Jiménez, C. S. (2021). Contribución al protocolo de secuenciación del genoma de SARS-CoV-2 utilizando la Tecnología Genómica UIS (Tesis de pregrado). Universidad Industrial de Santander, Bucaramanga, Santander, Colombia.

Ugurel, O. M., Ata, O., y Turgut-Balik, D. (2020). An updated analysis of variations in SARS-CoV-2 genome. *Turkish journal of biology*, *44*(3), 157–167. https://doi.org/10.3906/biy-2005-111

Uriu, K., Cárdenas, P., Muñoz, E., Barragan, V., Kosugi, Y., Shirakawa, K., Takaori-Kondo, A., Ito, J., Yamasoba, D., Kimura, I., Suganami, M., Oide, A., Yokoyama, M., Chiba, M., Nakagawa, S., Wu, J., Takahashi, M., Kazuma, Y., Nomura, R., … Sato, K. (2022). Characterization of the immune resistance of severe acute respiratory syndrome Coronavirus 2 Mu variant and the robust immunity induced by Mu infection. *The Journal of Infectious*

Diseases, *226*(7), 1200–1203. https://doi.org/10.1093/infdis/jiac053

World Health Organization. (2020). *WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020*. Recuperado de https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020

World Health Organization. (2021). *WHO Coronavirus (COVID-19) Dashboard | WHO Coronavirus (COVID-19) Dashboard With Vaccination Data*. Recuperado de https://covid19.who.int/

World Health Organization. (2022). *Tracking SARS-CoV-2 variants. World Health Organization*. Recuperado de https://www.who.int/activities/tracking-SARS-CoV-2-variants

Zhang, K., Zheludev, I. N., Hagey, R. J., Wu, M. T.-P., Haslecker, R., Hou, Y. J., Kretsch, R., Pintilie, G. D., Rangan, R., Kladwang, W., Li, S., Pham, E. A., Bernardin-Souibgui, C., Baric, R. S., Sheahan, T. P., D′Souza, V., Glenn, J. S., Chiu, W., y Das, R. (2020). Cryo-electron microscopy and exploratory antisense targeting of the 28-kDa frameshift stimulation element from the SARS-CoV-2 RNA genome. *BioRxiv*. https://doi.org/https://doi.org/10.1101/2020.07.18.209270

Zhang, Y., Zhang, H., y Zhang, W. (2022). SARS-CoV-2 variants, immune escape, and countermeasures. *Frontiers of Medicine*, *16*(2), 196–207. https://doi.org/10.1007/s11684-021-0906-x

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., Niu, P., Zhan, F., Ma, X., Wang, D., Xu, W., Wu, G., Gao, G. F., y Tan, W. (2020). A novel Coronavirus from patients with pneumonia in China, 2019. *New England Journal of Medicine*, *382*(8), 727–733. https://doi.org/10.1056/nejmoa2001017

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research, 31*(13), 3406–3415. https://doi.org/10.1093/nar/gkg595