

Contribución al protocolo de secuenciación del genoma de SARS-CoV-2
utilizando la tecnología genómica UIS

Carolina Sofía Torres Jiménez

Trabajo de Grado para Optar el Título de Bióloga

Director

Francisco José Martínez Pérez

Doctor en Ciencias

Codirector

Lina María Vera Cala

Doctor en Ciencias

Universidad Industrial de Santander

Facultad de Ciencias

Escuela de Biología

Bucaramanga

2021

Dedicatoria

A Carmelo Jiménez, cada día recuerdo tus palabras de orgullo hacia a mí, tu sonrisa al verme y tus lágrimas al despedirte. Este logro, del que de vez en cuando hablábamos, te lo dedico solamente a ti, esperanzada que desde el cielo lo compartas 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 (OMS) 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”, que se aprobó con número de registro 76900, código 1102-101-576900 y contrato RC N° 369-2020 para su ejecución en la convocatoria 1015 de la Mincienciatón. Esta, fue la invitación nacional a presentar propuestas de proyectos relacionados a detener la pandemia de COVID-19.

Al Dr. Francisco José Martínez Pérez por depositar su confianza en mí al permitirme ser parte de este gran proyecto, por su apoyo, dedicación y paciencia en este proceso de aprendizaje y por su gran entrega a la ciencia.

A Cristian Cadena por sus consejos como amigo y colega, y por sus horas dedicadas a este proyecto.

A mis padres y hermanos por su constante acompañamiento, comprensión y apoyo en cada decisión tomada.

A Nicolas Camargo, por ser mi apoyo y compañía en todo momento de la carrera.

Tabla de Contenido

Introducción.....	11
1. Objetivos	18
1.1 Objetivo de la pasantía	18
1.2 Objetivo general	18
1.3 Objetivos específicos	18
2. Competencias adquiridas.....	19
2.1 Competencias cognitivas.....	19
2.2 Competencias actitudinales	19
3. Metodología.....	20
3.1 Identificación de la composición nucleotídica, sitios de poliadenilación y traducción de genes de SARS-CoV-2.....	20
3.2 Diseño de cebadores	21
3.3 Formulación de la síntesis de ADN complementario (ADNc)	22
3.4 Amplificación de librería y protocolo de secuenciación genómica de SARS-CoV-2	23
4. Resultados	24
4.1 Composición nucleotídica y sitios de poliadenilación del genoma de SARSCoV-2.....	24
4.2 Identificación y traducción de los genes de la secuencia consenso del genoma de SARSCoV-2	25
4.3 Implementación de protocolo de síntesis de ADNc y cebadores	29

4.3.1 Cebadores	29
4.3.2 Protocolo de síntesis de ADNc.....	32
4.3.3 Implementación de protocolo de síntesis de ADNc y cebadores	34
4.4 Protocolo de secuenciación genómica de SARS-CoV-2.....	37
5. Discusión	38
6. Conclusiones	43
Referencias bibliográficas	44

Lista de Tablas

	Pág.
Tabla 1. Composición nucleotídica de los genomas consenso mensuales de SARS-CoV-2.....	24
Tabla 2. Posición nucleotídica de los genes de SARS-CoV-2.....	26
Tabla 3. Comparación de número de codones degenerados en enero y octubre de los genes de SARS-CoV-2	29
Tabla 4. Cebadores diseñados para la amplificación del genoma de SARS-CoV-2	30
Tabla 5. Concentraciones de ADNc antes y después de la purificación con perlas magnéticas. .	35

Lista de Figuras

	Pág.
Figura 1. Composición y organización del ARN genómico del virus SARS-CoV-2.....	12
Figura 2. Mapa de restricción de los polilinker añadidos a los cebadores	31
Figura 3. Flujo de trabajo del protocolo de síntesis de ADNc del genoma del virus SARS-CoV-2	33
Figura 4. Gel de agarosa no desnaturizante al 1% con las concentraciones de ADNc de cadena sencilla purificado.	36

Lista de Apéndices

Apéndice A. Alineamiento de secuencias consenso para diseño de cebadores para síntesis de ADNc para secuenciación genómica de SARS-CoV-2.

Apéndice B. Traducción conceptual de genes de SARS-CoV-2 de enero y octubre de 2020.

Apéndice C. Protocolo de secuenciación genómica del genoma del virus SARS-CoV-2.

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

Resumen

Título: Contribución al protocolo de secuenciación del genoma de SARS-CoV-2 utilizando la tecnología genómica UIS*

Autor: Carolina Sofía Torres Jiménez**

Palabras Clave: Secuenciación genómica, SARS-CoV-2, genómica UIS, COVID-19, ADNc.

Descripción: Entre los mecanismos de control y prevención de la pandemia ocasionada por la enfermedad infecciosa denominada COVID-19, causada por el Betacoronavirus SARS-CoV-2, se encuentran el estudio del origen del virus, el diagnóstico y el tratamiento de la enfermedad. Todos estos estudios se basan principalmente en el análisis del genoma viral, por lo que a nivel mundial se incentivó el desarrollo de tecnologías capaces de secuenciar el genoma completo de SARS-CoV-2, con el fin de mejorar los sistemas de diagnóstico y entender la naturaleza del virus. Por este motivo, en este trabajo se analizaron las características nucleotídicas de SARS-CoV-2, a partir de consensos mensuales generados de las secuencias del virus reportadas en la plataforma GISAID entre enero y octubre del año 2020. En función de estas, se diseñaron cebadores fundamentados en los criterios de la tecnología genómica UIS, que contribuyeron eficazmente en las reacciones de síntesis de ADNc. Además, se propone un protocolo con las concentraciones y condiciones de reacción para la síntesis de ADNc y un protocolo de secuenciación para el genoma completo de SARS-CoV-2, ambos con la inclusión de las mezclas COL-UIS y COL-UIS-dNTPs. Estas, junto con los cebadores diseñados, potenciaron la obtención del ADNc del virus SARS-CoV-2 demostrando su capacidad de aplicación en genomas virales ricos en AT. Finalmente, se concluye que la secuenciación del genoma completo del virus SARS-CoV-2 es posible al aplicar los principios de la tecnología genómica UIS, por lo tanto, al incluirla para el diseño de los cebadores, la síntesis de ADNc y en el protocolo de secuenciación propuesto, se favorece el proceso de secuenciación de genomas ricos en AT con la tecnología Ion Torrent.

* Trabajo de Grado

** Facultad de Ciencias. Escuela de Biología. Director: Francisco José Martínez Pérez. Doctor en Ciencias. Codirectora: Lina María Vera Cala. Doctora en Ciencias.

Abstract

Title: Contribution to the SARS-CoV-2 genome sequencing protocol using UIS genomic technology *

Author: Carolina Sofía Torres Jiménez **

Key Words: Genomic sequencing, SARS-CoV-2, UIS genomics, COVID-19, cDNA.

Description: Among the control and prevention mechanisms of the pandemic caused by the infectious disease termed COVID-19, caused by Betacoronavirus SARS-CoV-2, are the study of the origin of the virus, the diagnosis and treatment of the disease. All these studies are mainly based on the analysis of the viral genome, for this reason the development of technologies capable of sequencing the complete SARS-CoV-2 genome was encouraged worldwide, in order to improve diagnostic systems and understand the nature of the virus. For this reason, in this work the nucleotide characteristics of SARS-CoV-2 were analyzed, based on monthly consensus generated from the virus sequences reported on the GISAID platform between January and October 2020. Based on these, primers were designed according to UIS genomic technology, which effectively contributed to cDNA synthesis reactions. In addition, a protocol is proposed with the concentrations and reaction conditions for the synthesis of cDNA and a sequencing protocol for the complete SARS-CoV-2 genome, both with the inclusion of the COL-UIS and COL-UIS-dNTPs mixtures. These, along with the designed primers, enhanced the obtaining of the SARS-CoV-2 virus cDNA, demonstrating its applicability in AT-rich viral genomes. Finally, it is concluded that the sequencing of the complete genome of the SARS-CoV-2 virus is possible by applying the principles of UIS genomic technology, therefore, by including it for the design of the primers, the synthesis of cDNA and in the proposed sequencing protocol, the process of sequencing AT-rich genomes with Ion Torrent technology is favored.

* Bachelor thesis

** Science Faculty. School of Biology.

Adviser: PhD.Francisco José Martínez Pérez. Adviser: PhD. Lina María Vera Cala.

Introducción

Mundialmente se le denomina COVID-19 a la enfermedad infecciosa ocasionada por el coronavirus de tipo 2 causante del Síndrome Respiratorio Agudo Severo, también llamado SARS-CoV-2 por sus siglas en inglés; este es el séptimo coronavirus conocido en infectar a los humanos y junto con SARS-CoV y MERS-CoV, el tercero en causar brotes de enfermedades respiratorias (Moore & June, 2020; Weissleder et al., 2020). Como consecuencia de estos brotes, desde el 2002 los coronavirus emergentes se empezaron a considerar como preocupación de salud pública (Chan et al., 2020; Cui et al., 2019). El primer caso de COVID-19 fue reportado en diciembre de 2019 en Wuhan, China y rápidamente se esparció por toda la región y otros países (Zhu et al., 2020). Para marzo de 2020, ya se reportaban casos de pacientes infectados en 114 países y la Organización Mundial de la Salud (WHO) alertó que se trataba de una pandemia (WHO, 2020). La respuesta de los gobiernos y de las autoridades de la salud a nivel mundial para afrontar esta pandemia no pudo ser la misma que con los brotes causados por SARS-CoV en 2002, ya que en un principio aspectos como el origen de SARS-CoV-2 y el tratamiento a la enfermedad eran totalmente desconocidos, además, la tasa de contagio de SARS-CoV-2 es mucho mayor (Fani et al., 2020; Weissleder et al., 2020).

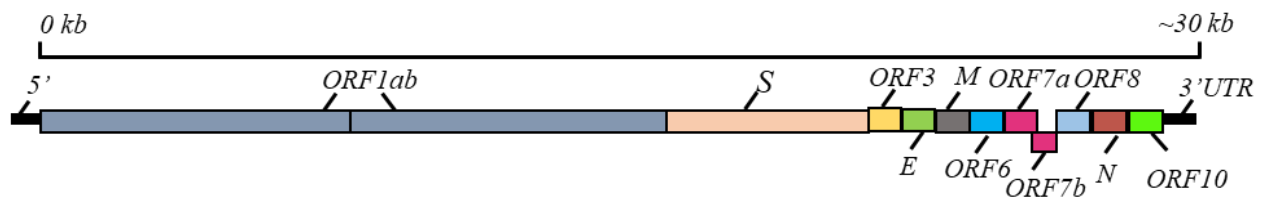
Los primeros estudios en torno a la pandemia se basaron en la identificación del virus, para esto se realizó la primera secuenciación genómica de SARS-CoV-2 mediante una secuenciación meta-transcriptómica complementada por PCR y secuenciación de nueva generación con la tecnología Illumina (F. Wu et al., 2020). Con este estudio, identificaron que SARS-CoV-2 pertenece a la familia Coronaviridae, específicamente al grupo de los coronavirus tipo SARS pertenecientes al género *Betacoronavirus*. Además, presenta un 79% de identidad con el

Betacoronavirus SARS-CoV y una identidad del 86.7 al 89% con un tipo de SARS originado en murciélagos de la especie *Rhinolophus affinis* (F. Wu et al., 2020; Zhou et al., 2020).

El genoma de aproximadamente 30 kb cuenta con la siguiente estructura genómica (Figura 1): desde el extremo 5' el primer gen *ORF1ab* que codifica para proteínas no estructurales, luego el gen *S* que codifica para la proteína espiga (Spike) en la que se encuentra el Dominio de Unión al Receptor (RDB, por sus siglas en inglés) que le confiere al virus la alta afinidad a las células humanas, seguidamente se encuentra los genes *ORF3*, *E* y *M*. Estos dos últimos codifican para las proteínas estructurales de envoltura y de membrana; a continuación los genes *ORF6*, *ORF7a*, *ORF7b*, *ORF8* y el gen *N* que codifica para la nucleocápside y por último el gen *ORF10*. No obstante los detalles exactos de los genes que componen el genoma de SARS-CoV-2 aún no están claros debido a la superposición algunos de estos (Nakagawa & Miyazawa, 2020).

Figura 1

Composición y organización del ARN genómico del virus SARS-CoV-2



Por otro lado, los primeros estudios epidemiológicos rastrearon el virus hasta el mercado de mariscos de Wuhan, China, pero la hipótesis que planteaba que este lugar fue la primera fuente de transmisión se descartó, ya que las muestras recolectadas en el mercado mostraron a un virus totalmente adaptado a células humanas (Wacharapluesadee et al., 2021). Una segunda hipótesis plantea que el virus tiene un origen animal y que el RDB es resultado de la selección natural, esta hipótesis se basa en el porcentaje de identidad que tiene el virus con un grupo de virus encontrados en murciélagos, como se mencionó anteriormente (F. Wu et al., 2020; Zhou et al., 2020). Sin

embargo, ninguno de estos virus es suficientemente similar para ser un predecesor directo, por lo que se sugiere que hay un hospedero intermedio entre murciélagos y humanos (Andersen et al., 2020). Este hospedero intermedio sería, según esta hipótesis, el pangolín *Manis javanica*, ya que algunos coronavirus encontrados en este muestran similitud con el RBD de SARS-CoV-2 (Zhang et al., 2020).

Por otro lado, Sirotkin y Sirotkin en 2020, manifiestan que no se puede descartar la idea que el virus se originara en un laboratorio de virología que hiciera experimentos con coronavirus, esta hipótesis se basa en que la inserción del RBD del pangolín pudo ser resultado de experimentos de mutagénesis hechos en laboratorios y que los investigadores se pudieron infectar accidentalmente (Sirotkin & Sirotkin, 2020). Adicionalmente, explican que, con la tecnología actual, la manipulación del virus pudo haberse hecho desde cualquier laboratorio del mundo que tenga las condiciones (Segreto & Deigin, 2021; Sirotkin & Sirotkin, 2020).

Durante el año 2020, los gobiernos tomaron medidas de prevención para mitigar la propagación de la pandemia, siendo las estrategias más exitosas: el aislamiento preventivo de las personas sanas, el diagnóstico y posterior aislamiento de las personas con síntomas (Bachelet, 2020).

Las principales pruebas diagnósticas de COVID-19 se pueden dividir entre:

- ✓ Prueba serológica, que evalúa el estado inmunológico del paciente a través de las proteínas de respuesta IgG/IgM, distinguen las infecciones pasadas en el paciente debidas a SARS-CoV-2 y determina la exposición viral de la población (Umakanthan et al., 2021; Weissleder et al., 2020).
- ✓ Prueba de antígeno, que sondea las proteínas de la nucleocápside y la proteína espiga del virus mediante pruebas de flujo lateral o Ensayo Inmunoabsorbente

Ligado a Enzimas (ELISA), a través de muestras nasofaríngeas tomadas con hisopos (Umakanthan et al., 2021; Weissleder et al., 2020).

- ✓ Prueba de ácidos nucleicos o prueba diagnóstica de RT-PCR en tiempo real de uno o dos pasos, en la que a partir de muestras nasofaríngeas, orofaríngeas o nasales tomadas por medio de un frotis con hisopo, se transcribe el ARN viral a ADN complementario (ADNc) y posteriormente se amplifica por la RT-PCR y con ellos se logra hacer una detección cualitativa de los ácidos nucleicos del virus (CDC, 2020; Weissleder et al., 2020).

Todas estas pruebas deben ser precisas, ya que, si se da un diagnóstico negativo a un paciente infectado, este podría generar una nueva ola de infección. La precisión de una prueba diagnóstica se determina por su sensibilidad, especificidad y valor predictivo, es decir, evita resultados falsos negativos y falsos positivos (Weissleder et al., 2020). En cuanto a las pruebas serológicas no hay suficiente evidencia de su precisión que respalde su uso (Bastos et al., 2020). Por otro lado, la sensibilidad de las pruebas de anticuerpos depende del momento en que se tome la muestra, ya que esta será baja los primeros días de infección, sin embargo, son útiles para identificar infecciones previas y sirven como complemento a las pruebas de RT-PCR en tiempo real, las cuales son las más utilizadas a nivel mundial y han demostrado ser las más precisas, siempre y cuando se utilice el esputo como tipo de muestra para realizar la prueba (Böger et al., 2021; Deeks et al., 2020; Weissleder et al., 2020). A pesar de esto, también se ha demostrado que las pruebas de RT-PCR en tiempo real presentan un rango entre 2% y 29% de resultados falsos negativos y se considera que los posibles factores causantes de estos resultados son la calidad de la muestra, el tiempo que lleva el paciente presentando los síntomas o las mutaciones del virus que

afectan la prueba (Arevalo-Rodriguez et al., 2020; Gilchrist et al., 2015; Watson et al., 2020; Woloshin et al., 2020)

Considerando que, en los centros de salud autorizados para realizar los protocolos, se siguen correctamente las indicaciones de los procedimientos oficiales para la realización de los diagnósticos en pacientes y el origen biológico de la muestra es el autorizado, la opción más probable que puede afectar los resultados de la prueba es la alta tasa mutacional que presenta el virus, ya que los cebadores que se utilizan en esta fueron desarrollados con base en las primeras cepas reportadas en China y los cambios que ha presentado el genoma del virus podrían afectar la especificidad de estos (Shen et al., 2020). Además, estas mutaciones no solo afectan el diagnóstico sino que pueden variar la virulencia y transmisibilidad del virus (Shen et al., 2020). Por esto, para la identificación de las mutaciones y el estudio de la evolución viral es fundamental que periódicamente se realicen secuenciaciones genómicas del virus, especialmente en los pacientes que presentan los síntomas de COVID-19 pero que la prueba de diagnóstico por RT-PCR presentara un resultado negativo (Phan, 2020).

El estudio de la evolución viral necesita el estado genómico actual del virus, por consiguiente, la secuenciación debe ser rápida y directamente de la muestra tomada de los pacientes, más no de cultivos de poblaciones virales aisladas de líneas celulares infectadas, para evitar sesgos al identificar variantes no naturales del virus ocasionadas por las condiciones controladas del laboratorio (Chiara et al., 2020). Además, la secuenciación del genoma facilita la identificación 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).

Conforme a lo anterior, alrededor del mundo se empezó a tomar la iniciativa para secuenciar genomas de SARS-CoV-2 con el fin de caracterizar su componente genético (Alvarez et al., 2020). Particularmente en Colombia, se incentivó la investigación científica en torno a la pandemia como estrategia para buscar soluciones que permitieran afrontarlo y fortalecer los sistemas de alerta temprana ante emergencias de salud pública (Alvarez et al., 2020; Eduardo Rojas Pineda, 2020). Además, surgió la iniciativa global de compartir los datos obtenidos en bases de datos como GISAID (*Global Initiative on Sharing All Influenza Data*) (Shu & McCauley, 2017) y GenBank del NCBI (*National Center for Biotechnology Information*) (Sayers et al., 2019), de esta manera, se empezaron a caracterizar las diferentes cepas y se evidenciaron los cambios en el genoma del virus con tecnologías desarrolladas a partir de la secuenciación de nueva generación (NGS) (Alvarez et al., 2020). Lo anterior, ha permitido el incremento del conocimiento genómico del virus en aspectos importantes como su tasa de mutación estimada, su proporción nucleotídica y la asignación de linajes a cada cepa mediante las mutaciones presentes en el genoma de SARS-CoV-2 (Candido et al., 2020; Hou, 2020).

Sin embargo, la complejidad de la secuenciación completa del genoma de SARS-CoV-2 se ve reflejada en las secuencias genómicas reportadas en las bases de datos, ya que muchas de estas muestran regiones incompletas o grandes cantidades de gaps en los marcos abiertos de lectura (ORF) de los genes que hacen inconclusa la traducción del ARN mensajero (ARNm). Por consiguiente, se plantea como solución para el mejoramiento de la calidad de la secuenciación considerar aspectos fisicoquímicos y organización del genoma del virus implementando la Tecnología Genómica UIS (Mejía-Ospino et al., 2021), la cual tiene capacidad de generar mezclas químicas para la amplificación y secuenciación de polímeros de ácidos nucleicos de genomas en proporciones extremas de nucleótidos.

En esta pasantía, se contribuye al proceso de secuenciación genómica del virus SARS-CoV-2 mediante el diseño de los protocolos de síntesis de ADN complementario (ADNc) y de secuenciación genómica fundamentados en la tecnología genómica UIS (Mejía-Ospino et al., 2021), que se implementaron para establecer las secuencias genómicas de SARS-CoV-2 de pacientes de Santander, Colombia en el 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 inclusión de genes marcadores requeridos para la infección viral y su validación por secuenciación genómica de nueva generación” presentado por el Grupo de Investigación de Computo Avanzado y a Gran Escala (CAGE) de la Universidad Industrial de Santander y financiado por el Ministerio de Ciencia, Tecnología e Innovación mediante la Mincienciatón (Eduardo Rojas Pineda, 2020).

1. Objetivos

1.1 Objetivo de la pasantía

Contribuir al diseño del protocolo de secuenciación del genoma del nuevo coronavirus SARS-CoV-2 con la tecnología genómica UIS.

1.2 Objetivo general

Elaborar un protocolo para la secuenciación del genoma completo del coronavirus SARS-CoV-2 implementando la tecnología genómica UIS.

1.3 Objetivos específicos

Diseñar oligonucleótidos para la amplificación del genoma completo del virus SARS-CoV-2, en función de la tecnología genómica UIS.

Determinar las condiciones para la síntesis de ADN complementario del genoma completo del virus SARS-CoV-2, en función de la tecnología genómica UIS.

Establecer las condiciones para la secuenciación del genoma completo del virus SARS-CoV-2, con la secuenciación Ion Torrent, en función de la tecnología genómica UIS.

2. Competencias adquiridas

2.1 Competencias cognitivas

Analiza secuencias consenso de genomas virales para diseñar oligonucleótidos de RT-PCR.

Identifica secuencias de genomas virales para el diseño de oligonucleótidos.

Formula condiciones de reacción para la síntesis de ADN complementario.

Elabora procedimientos y condiciones de secuenciación genómica Ion Torrent para obtener genomas en función de la tecnología UIS.

2.2 Competencias actitudinales

Participa en proyectos de prioridad nacional para contribuir a la solución de problemas de principal interés nacional y relevancia a nivel mundial.

Colabora en un grupo de investigación para desarrollar propuestas de innovación.

Desarrolla actitudes éticas y bioéticas para la generación de resultados científicos.

3. Metodología

3.1 Identificación de la composición nucleotídica, sitios de poliadenilación y traducción de genes de SARS-CoV-2

A partir de las secuencias del genoma de SARS-CoV-2 reportadas en las bases de datos GenBank del NCBI (Sayers et al., 2019) y GISAID (Shu & McCauley, 2017), el grupo de investigación generó una secuencia consenso por mes de cada base de datos a partir de enero de 2020. Entre los parámetros para la selección de secuencias se encontraba el de descartar secuencias incompletas o con nucleótidos no determinados para evitar sesgos en los consensos, esto se realizó con la aplicación web prototipo Sequence Manager, que continúa en desarrollo por el grupo de investigación del proyecto. Además, las posiciones con dos o más opciones nucleotídicas en cada secuencia consenso fueron indicadas con el código IUPAC (Dixon et al., 1986) en el programa BioEdit (versión 7.2.5) (Hall, 1999) con un umbral de frecuencia de 100 y con la inclusión de gaps. Finalmente, se utilizaron los recursos de HPC de SC3UIS para alinear las secuencias, esto se realizó mes a mes empleando el software MAFFT (<https://mafft.cbrc.jp/alignment/software/>) con los parámetros correspondientes al Modelo de Pérdida de ADN (Martínez-Pérez et al., 2007).

Para este estudio se escogió el alineamiento de la base de datos GISAID, ya que presentó mayor número de genomas reportados a nivel mundial. Posteriormente, a partir de las secuencias consenso del mes de enero hasta el mes de octubre, se analizó la composición nucleotídica del genoma y se identificaron los posibles sitios de poliadenilación de este, utilizando las secuencias TGTA, reportada como sitio canónico de poliadenilación para algas (Ji et al., 2015), como secuencia control y la secuencia AATAAA correspondiente a sitio canónico de poliadenilación

para mamíferos (Sun et al., 2020); estos análisis se realizaron con la aplicación web prototipo Sequence Manager.

A partir de cada secuencia consenso se realizó la identificación de la posición del codón de inicio de la traducción en el extremo 5' y el codón de paro en el extremo 3' de los genes: *S*, *ORF3a*, *E*, *M*, *ORF6*, *ORF8*, *ORF7b* y *N*, de acuerdo con lo reportado en el genoma de referencia con número de acceso MT844089 de la base de datos del GenBank. Además, para los genes *ORF1ab*, *ORF7a* y *ORF10* se realizó con la Herramienta Básica de Búsqueda de Alineación Local (*BLAST*) (Boratyn et al., 2013) y se identificaron las posiciones nucleotídicas de acuerdo con el genoma de referencia de la base de datos GenBank con número de acceso MW292665 para el gen *ORF1ab* y MW349178 para los genes *ORF7ab* y *ORF10*.

Finalmente, para analizar los cambios aminoacídicos generados por las mutaciones en una posición de un codón con respecto a los genes del primer genoma consenso, se compararon los genes de enero de 2020, con respecto a los de octubre del mismo año. Para esto, se eliminaron los gaps presentes en cada gen y se realizó una comparación de la traducción conceptual de cada gen del consenso generado a partir de la base de datos GISAID con el programa Degenerate Codons (version 1.0), desarrollado previamente por el grupo de investigación.

Para acceder a la plataforma Sequence Manager o al programa Degenerate Codons se debe solicitar la autorización correspondiente al director del grupo CAGE de la Universidad Industrial de Santander, Colombia.

3.2 Diseño de cebadores

A partir de las secuencias consenso mensuales del genoma de SARS-CoV-2 de enero a mayo de 2020, se identificaron las regiones variables, es decir, las que presentaban mayores cambios nucleotídicos y las regiones más conservadas. De igual forma, se establecieron aquellos

sitios con menor número de mutaciones en cada codón, los cuales fueron consideradas como regiones minoritarias. A partir de esta delimitación y siguiendo los principios propuestos por Roche, (2006), para el diseño de los cebadores se escogieron secuencias sentido y anti-sentido de un tamaño de 20 a 30 nucleótidos. Para su selección, estos debían presentar las siguientes características: contenido de G/C variante entre el 40% y el 60%, tener una distribución homogénea de dominios A/T y G/C y una temperatura de fusión (T_m) entre 55° y 65° C, no ser complementarios entre sí en sus extremos 3', además, no ser autocomplementarios en otras regiones y permitir la formación de estructuras secundarias que pudieran favorecer el diseño de sondas tipo escorpión.

Las características fisicoquímicas de los cebadores se determinaron con la herramienta OligoAnalyzer Tool (Owczarzy et al., 2008), la concentración de Cloruro de Magnesio ($MgCl_2$) se determinó de acuerdo con la concentración de la ADN polimerasas que se emplearon en las reacciones de PCR y RT-PCR de todo el proceso. La estructura secundaria de cada cebador para la formación de estructura de tipo escorpión se determinó con la plataforma RNAstructure (Bellaousov et al., 2013). En caso de que la región seleccionada tuviese una temperatura media de fusión menor a 60°C, se diseñó un polilinker rico en G-C que sirviese como adaptador para obtener los amplicones deseados. Los mapas de restricción de los polilinker se generaron con el programa Rebase de New England BioLabs (Roberts et al., 2010) y sus características fisicoquímicas como cebador se indicaron como se mencionó anteriormente.

3.3 Formulación de la síntesis de ADN complementario (ADNc)

La formulación de la síntesis de ADN complementario (ADNc) de SARS-CoV-2, se realizó con base en las características bioinformáticas de su genoma y se sustentó en la enzima transcriptasa reversa SuperScript IV de Invitrogen (Pub. No. MAN0013442), además de las

enzimas que incluyen las mezclas diseñadas con base en la tecnología genómica UIS (Barrios Hernandez et al., 2018). Las concentraciones de los cebadores, agentes desnaturalizantes y de nucleótidos se diseñaron para obtener el ADNc de aproximadamente 30 kb que correspondiese al genoma completo de SARS-CoV-2 y otras para la producción de amplicones de 10kb y 20kb. Las mezclas incluyeron un cóctel de RNasas para la eliminación del RNA mensajero (ARNm) molde del genoma del SARS-CoV-2 y la eliminación del ARN total de la muestra. Finalmente, para la purificación del ADNc, el protocolo incluyó los pasos a seguir del protocolo de purificación de ADN/ARN a partir de perlas magnéticas propuesto por PureDirex (Cat.No. PDM04-0100), así mismo se agregó la reacción de RT-PCR para la amplificación del ADNc y su reamplificación por medio de los adaptadores con las mezclas de ADN polimerasas de la tecnología genómica UIS.

La validación experimental de este protocolo fue realizada por el biólogo contratado como profesional técnico por el proyecto, capacitado para manipular muestras del virus y autorizado a ingresar al Laboratorio Central de Investigaciones de la Facultad de Salud de la Universidad Industrial de Santander (LCI-FS-UIS). Además de implementar el protocolo, verificó la síntesis de ADNc con la determinación de la concentración por espectrofotometría y la integridad del ADNc por electroforesis en gel no desnaturalizante. Los ADNc se enviaron a ser secuenciados por la Unidad de Secuenciación GENECORE de la Universidad de los Andes en Bogotá.

3.4 Amplificación de librería y protocolo de secuenciación genómica de SARS-CoV-2

El protocolo de secuenciación genómica de SARS-CoV-2 se realizó con base en el protocolo de generación de bibliotecas genómicas de la tecnología Ion Torrent (Damerla et al., 2014), al incluir las mezclas COL-UIS y COL-UIS-dNTPs propuestas por la tecnología genómica UIS en los ítems donde se incluye la polimerización de ácidos nucleicos que realiza el equipo

automatizado de los protocolos “Ion AmpliSeq™ Library Preparation on the Ion Chef™ System” y “Ion 520™ & Ion 530™ Kit – OT2”.

4. Resultados

4.1 Composición nucleotídica y sitios de poliadenilación del genoma de SARSCoV-2

El análisis de las secuencias consenso mensuales del genoma de SARS-CoV-2 mostró un genoma mayoritariamente rico en AT (Tabla 1), con una proporción promedio aproximada de 70:30. También, se evidencia que la proporción de dinucleótidos GC en octubre presentó el menor valor, el descenso de esta proporción no es constante y se ve reflejado principalmente en C.

Tabla 1

Composición nucleotídica de los genomas consenso mensuales de SARS-CoV-2

	A	T	C	G	AT	GC
Enero	29.3%	33.0%	17.8%	19.5%	63,2%	37.3%
Febrero	29.3%	33.4%	17.3%	19.4%	62,7%	36.8%
Marzo	25.7%	44.3%	10.9%	18.8%	69,9%	29.8%
Abril	25,3%	44.9%	10.7%	18.8%	70.2%	29.5%
Mayo	26,5%	41.6%	12.4%	18.6%	68.1%	30.9%
Junio	25,7%	43.9%	11.3%	18.5%	69.7%	29.8%
Julio	26,7%	41.8%	12.0%	18.5%	68.5%	30.5%
Agosto	27,1%	41.9%	12.1%	18.0%	68.9%	30.0%
Septiembre	28,0%	39.3%	13.5%	18.1%	67.3%	31.7%
Octubre	25,5%	44.4%	10.6%	18.4%	69.9%	29.1%

Nota: los valores de las columnas AT y GC son la suma de las columnas A + T y G + C, respectivamente, más el porcentaje de aparición respectivo de cada letra en las posiciones con dos o más opciones nucleotídicas.

En cuanto a los posibles sitios de poliadenilación con la secuencia AATAAA correspondiente a sitios canónicos en mamíferos (Sun et al., 2020), se encontraron 14 sitios en el

genoma: 11 de estos dentro del gen *ORF1ab*, 1 dentro del gen *ORF3a*, 1 dentro del gen *M* y 1 dentro del gen *ORF6*. Por otro lado, la secuencia canónica control, TGTAAT reportada para algas (Ji et al., 2015), se encontró en 39 sitios del genoma: 23 de estos dentro del gen *ORF1ab*, 5 dentro del gen *S*, 4 dentro del gen *ORF3a*, 1 dentro del gen *E*, 1 dentro del gen *N* y 5 dentro en las regiones no codificantes ubicadas antes del primer codón de inicio del primer gen y posterior al codón de paro del último gen. Además, se resalta que en las posiciones de la 2441 a la 2450 y de la 19461 a la 19470 hay un solapamiento en la última base de la secuencia canónica de algas y la primera base de la secuencia canónica para mamíferos, y desde la posición 18064 se encuentran dos sitios de poliadenilación canónicos de algas seguidos (Apéndice A).

4.2 Identificación y traducción de los genes de la secuencia consenso del genoma de SARS-CoV-2

Los genes identificados en los consensos se dividen en 11 (Tabla 2), en todos los consensos alineados la posición del primer nucleótido y la posición del último nucleótido de cada gen coincidieron, no obstante, en estos alineamientos se incluyeron gaps (Apéndice A). El gen de mayor longitud identificado fue *ORF1ab* (21323 nucleótidos) y el de menor longitud fue *ORF10* (122 nucleótidos). La ubicación de los genes en el alineamiento de los genomas consenso se encuentran en el apéndice A.

Tabla 2*Posición nucleotídica de los genes de SARS-CoV-2*

GEN	INICIO	FIN
ORF1ab	396	21717
S	21725	25560
ORF3a	25569	26405
E	26430	26657
M	26709	27377
ORF6	27388	27574
ORF7a	27581	27949
ORF7b	27946	28077
ORF8	28084	28452
N	28467	29732
ORF10	29757	29877

Nota: Los valores en la tabla se refieren a la posición del primer nucleótido del codón de inicio y del último nucleótido del codón de paro de los 11 genes de SARS-CoV-2 identificados a partir de consensos mensuales del genoma.

La comparación de las traducciones conceptuales de los genes reportados en enero contra los genes reportados en octubre se presenta en el apéndice B, su análisis encontró que la longitud neta de los genes *ORF1ab*, *S*, *ORF3a* y *N* varía, entendiéndose como longitud neta el número de bases nitrogenadas sin gaps entre ellas. Lo anterior se refleja en el total de codones codificantes de estos genes (Tabla 3), de esta manera, el gen *ORF1ab* presenta 5 codones codificantes más en el mes de octubre, además, el gen pasó de presentar 155 codones degenerados en enero a 1296 en octubre. Estos codones se encuentran a lo largo de todo el gen y su degeneración se ve reflejada en la traducción conceptual del gen, en la cual la mayoría de los codones degenerados tienen de 2 a 4 opciones aminoacídicas para codificar y solo 8 codones en todo el gen tienen 6. Además, dentro del gen hay 6 codones que podrían codificar una señal de paro cortando la traducción del gen.

Los genes *S*, *ORF3a* y *N* presentan un codón de más en el mes de octubre. Para el caso del gen *S* hay un aumento de 212 codones degenerados, los cuales se encuentran a lo largo del gen y tienen de 2 a 4 posibilidades aminoacídicas, con la excepción de 4 codones que presentan 6 posibilidades. Por su parte, el gen *ORF3a*, presenta codones degenerados a lo largo de todo el gen, aumentando la cantidad de estos de enero a octubre con una diferencia de 116 codones degenerados de más. A pesar de ser un gen corto en comparación con los anteriores, los codones degenerados presentan de 2 a 6 posibilidades de aminoácidos para codificar. Finalmente, el gen *N* pasó de tener 11 codones degenerados en enero, todos presentando posibilidad de codificar solo 2 aminoácidos, a presentar 152 en octubre, la mayoría presentando posibilidad de codificar de 4 a 6 aminoácidos; dentro de estos codones a lo largo del gen, 5 presentaron posibilidad de codificar para una señal de paro antes del codón de paro teórico del gen, no obstante, el codón de paro del gen también se encontró degenerado con dos posibilidades de traducción en las cuales ambas codifican una señal de paro.

Los demás genes conservaron la misma longitud en enero y octubre, sin embargo, la cantidad de codones degenerados en los dos meses cambió. El gen *ORF6* presentó 62 codones en los dos meses, de los cuales solo se encontró un codón degenerado en el mes de enero, específicamente el octavo codón. Por otro lado, en el mes de octubre 14 codones se encontraron degenerados, las opciones aminoacídicas de estos van de 2 a 6. Finalmente, entre los codones degenerados en octubre, no se encontró el octavo codón que estaba degenerado en el mes enero.

El gen *M* de 223 codones pasó de tener 7 codones degenerados en enero a tener 39 en octubre, la mayoría de estos con posibilidad de codificar para dos aminoácidos. Por otra parte, el gen *E* presentó 76 codones, de los cuales 3 se encontraban degenerados en enero con la posibilidad

de codificar para dos aminoácidos y en octubre los codones degenerados aumentaron a 14, pero solo uno de estos presentó la posibilidad de codificar para más de 2 aminoácidos.

En relación con los genes *ORF7a* y *b*, se presentaron por separado ya que fueron los únicos que se encontraron superpuestos. El análisis separado de la traducción conceptual de cada gen mostró que el gen *ORF7a* tiene una diferencia de 50 codones degenerados desde enero a octubre, además, dentro del gen se encontraron 3 codones con la posibilidad de codificar para un codón de paro. Sin embargo, el codón de paro teórico del gen también se encontró degenerado, dejando la posibilidad de no codificar para un codón de paro sino para una Leucina, lo que significaría que el marco de lectura de este gen se podría correr incluyendo en su transcripción a los codones pertenecientes al gen *ORF7b*. Por otra parte, el gen *ORF7b* se encuentra superpuesto 4 codones sobre el final del marco de lectura del gen *ORF7a*. Este gen presentó 44 codones de los cuales 18 estaban degenerados en octubre, dejando 5 posibilidades de que codifiquen para codón de paro antes del codón de paro teórico del gen.

El gen *ORF8* presentó una longitud de 366 bases, es decir 122 codones, de los cuales en enero 6 de estos se encontraban degenerados mientras que en octubre se presentaron 52, de los cuales 9 presentaban posibilidad de codificar para codón de paro. Finalmente, el gen *ORF 10*, con una longitud de 117 bases, traducidas en 39 codones, presentó 3 de estos mutados para el mes de enero mientras que en octubre ya presentó 9, entre estos el codón de inicio presenta una mutación en la cual produce Isoleucina en vez de Metionina.

Tabla 3

Comparación de número de codones degenerados en enero y octubre de los genes de SARS-CoV-2

Nombre del gen	Número de codones totales del gen	Número de codones degenerados	
		Enero	Octubre
ORF1ab	7.094 / 7.099 *	155	1296
S	1.274 / 1.275 *	25	237
ORF3a	276 / 277 *	13	129
E	76	3	14
M	233	7	39
ORF6	105	1	24
ORF7a	122	5	55
ORF7b	44	0	18
ORF 8	122	6	52
N	420 / 421 *	11	152
ORF10	39	3	9

Nota: Número de codones totales y número de codones degenerados de cada gen identificado en las secuencias consenso de enero y octubre del genoma de SARS-CoV-2.

* Número de codones totales en enero/octubre

4.3 Implementación de protocolo de síntesis de ADNc y cebadores

4.3.1 Cebadores

En total 8 cebadores fueron diseñados para la amplificación del genoma de SARS-CoV-2 (Tabla 4), 5 de ellos corresponden a cebadores antisentido ubicados a 10 kb, 20 kb y 3 a 30 kb río arriba del extremo 3' UTR. Los 3 restantes corresponden a cebadores en dirección sentido, ubicados posiciones antes del inicio del primer gen ORF1ab (Apéndice A). Cada cebador se diseñó considerando una concentración de 1mM de MgCl₂ como cofactor de la mezcla de ADN

polimerasas termoestables utilizada y la temperatura media de punto de fusión de cada uno se encuentra en un rango de 73.4° a 78.4°C, para evitar la generación de amplificaciones inespecíficas.

Tabla 4

Cebadores diseñados para la amplificación del genoma de SARS-CoV-2

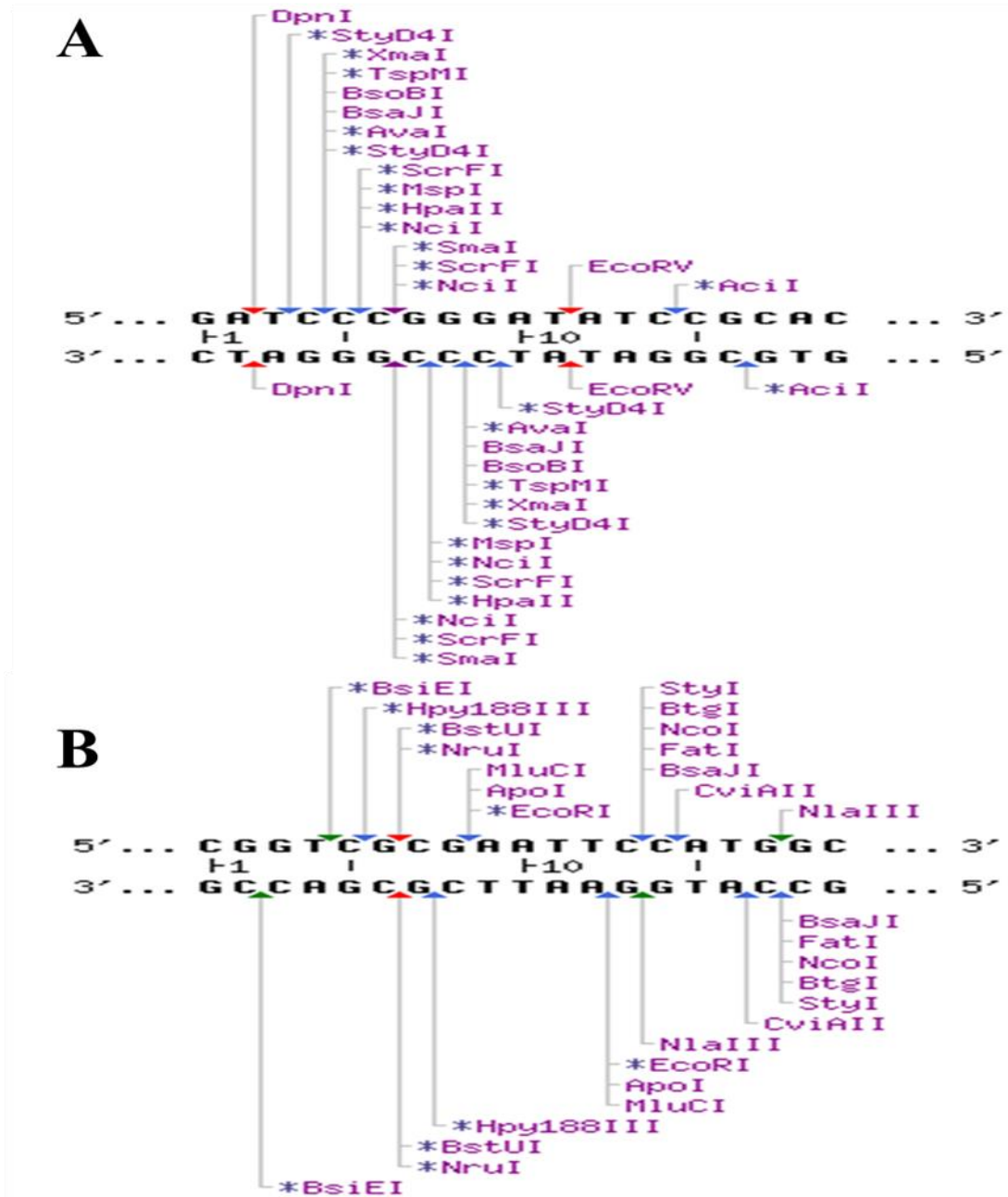
Nombre	Tamaño	Secuencia de cebador más polilinker*	Tm**
FW01COVID19	41	5' GGATCCCGGGATATCCGCACAT GTCTGACAGGACACGAG 3'	75.6 °C
FW02COVID19	41	5' GGATCCCGGGATATCCGCACAC GATCATCAGYACATCTAGG 3'	73.4°C
FW03COVID19	41	5' GGATCCCGGGATATCCGCACAT GTGACCGAAAGGTAAGATG 3'	73.4 °C
RV10kCOVID19	41	3'GTGGTGTGGAGATAGTGGGCC GGTACCTTAAGCGCTGGC 5'	76.4 °C
RV20kCOVID19	43	3'CCATCTCAACTACCAAGTTCATCGCC GGTACCTTAAGCGCTGGC 5'	75.7 °C
RV30kBCOVID19	41	3'CTCGGTGGTGTAAGAGTGGCGCC GGTACCTTAAGCGCTGGC 5'	78.4 °C
RV30kACOVID19	41	3'GTCACACATTGTAATCCCTCCGCC GGTACCTTAAGCGCTGGC 5'	76.2 °C
RV30kAbCOVID19	41	3'CGACGGATATACCTTCTCGGGCC GGTACCTTAAGCGCTGGC 5'	76.2 °C

Nota: Cebadores en dirección sentido: Fw= Forward; 01, 02 o 03= Posición río abajo en región 5' no traducida; COVID19= nombre de SARS-CoV-2. Cebadores en dirección antisentido: Rv= Reverse, 10k, 20k, 30Ak o 30Bk = Pares de bases de ADNc; A, B o Ab= Posición río arriba en extremo 3' UTR. COVID19= nombre de SARS-CoV-2.

*La secuencia polilinker se encuentra resaltada en negrita.

** Temperatura media de fusión.

Por las características nucleotídicas encontradas en el análisis del genoma, a cada cebador se le agregó un polilinker rico en G-C con secuencias ricas en sitios de restricción utilizadas en la clonación con plásmidos o bacteriófagos como *EcoRI* o *XmaI* (Figura 2). De esta manera, en las secuencias sentido de los cebadores se incluyó en el extremo 5' la secuencia GATCCC rica en sitios de restricción y en los cebadores antisentido los sitios de restricción se encuentran repartidos en toda la secuencia (Figura 2). Los polilinker agregados sirven para generar una clonación tipo RACE 5' o 3' de los amplicones, además, si el usuario lo requiere también sirven para realizar una segunda PCR o PCR anidada. Todos ellos, tienen una Tm entre 60° y 65°C.

Figura 2*Mapa de restricción de los polilinker añadidos a los cebadores*

Nota: A. polilinker añadidos a los cebadores en dirección sentido y B. polilinker añadidos a los cebadores en dirección antisentido, diseñados para amplificar el genoma de SARS-CoV-2.

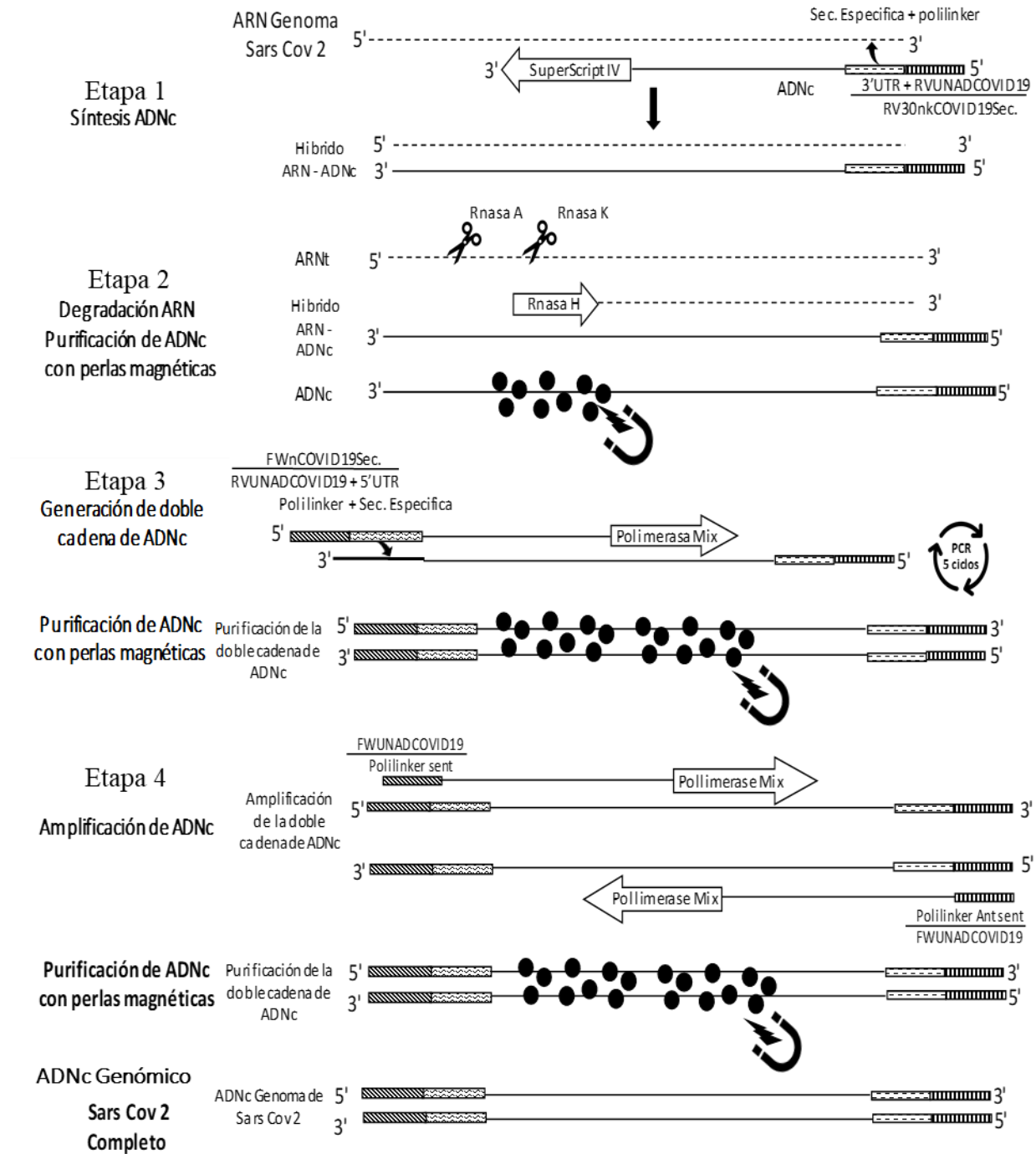
4.3.2 Protocolo de síntesis de ADNc

Los procedimientos explicados en el protocolo para la síntesis de ADNc (Apéndice C) se destacan por la adición de sustancias químicas y oligonucleótidos diseñados con base en la tecnología genómica UIS y características nucleotídicas del genoma de SARS-CoV-2. El objetivo principal es la generación de ADNc de SARS-CoV-2, a partir de ARN total obtenido de muestras de pacientes clínicamente diagnosticados con COVID-19 y su amplificación por RT-PCR.

El protocolo consta de 4 etapas (Figura 3): (1) síntesis de ADN, (2) degradación de ARN y purificación de ADNc con perlas magnéticas, (3) generación de doble cadena de ADNc y purificación de ADNc con perlas magnéticas y (4) amplificación de ADNc.

Figura 3

Flujo de trabajo del protocolo de síntesis de ADNc del genoma del virus SARS-CoV-2



En las dos primeras etapas se desarrollan los pasos para la síntesis y purificación de ADNc del genoma de SARS-CoV-2, en las primeras mezclas se agregan los cebadores diseñados previamente, la enzima SuperScript IV, más cualquiera de las sustancias denominadas COL-UIS-dNTPs o COL-UIS, según el criterio del usuario, que favorecen la síntesis del ADNc. Posteriormente, luego de la formación del híbrido ARN-ADNc, comienza la segunda etapa donde se realiza un cóctel de RNasas para la degradación del ARN total y el ARN del híbrido de ARN-ADNc. Además, se realiza la purificación del ADNc siguiendo los pasos del protocolo de purificación de ARN/ADN con perlas magnéticas propuesto por PureDirex. Finalmente, en la tercera y cuarta etapa, se genera la doble cadena de ADNc y se amplifica, para esto se realizan dos PCR con el ADNc generado en las fases uno y dos, más una mezcla con amortiguadores de alta fidelidad, $MgCl_2$ cebadores diseñados anteriormente y amortiguador COL-UIS o COL-UIS-dNTPs. Finalmente, para la purificación con perlas magnéticas se siguen los pasos del protocolo de purificación de purificación de ARN/ADN con perlas magnéticas propuesto por PureDirex.

Debido a que el protocolo actualmente se encuentra protegido como secreto industrial por la dirección de propiedad intelectual de la vicerrectoría de investigación y extensión de la Universidad Industrial de Santander, se debe realizar la respectiva solicitud para tener acceso a todos los detalles para su ejecución.

4.3.3 Implementación de protocolo de síntesis de ADNc y cebadores

Los resultados de la implementación del protocolo de ADNc utilizando el cebador FW01COVID19, en dos muestras de pacientes positivos para COVID-19 proporcionados el técnico de laboratorio del proyecto se observan en la tabla 5, en la cual se comparan las concentraciones promedio de ácidos nucleicos de dos muestras antes y después de ser purificados con perlas magnéticas luego de cada tratamiento aplicado. El protocolo comercial hace referencia

a una mezcla con la misma proporción de nucleótidos, por otra parte, COL-UIS y dNTPs hacen referencia a las mezclas para tratamiento nucleotídico COL-UIS y COL-UIS-dNTPs, respectivamente, ambas basadas en la tecnología genómica UIS.

Es de resaltar que, en la muestra CO-010 la concentración de ADNc antes y después de la purificación fue mayor con el tratamiento de dNTPs. Por otro lado, con la muestra CO-036 la concentración de ADNc antes de la purificación fue mayor con dNTPs, sin embargo, luego de la purificación el protocolo comercial presentó mayor concentración (Tabla 5). Por otro lado, la implementación de los cebadores RV30kAbCOVID19 y RV30kBCOVID19 no generó resultados.

Tabla 5

Concentraciones de ADNc antes y después de la purificación con perlas magnéticas

Muestra	Tratamiento	Concentración de ácidos nucleicos	
		Antes de la purificación (ng/μl)	Después de la purificación (ng/μl)
CO-010	Comercial	800.2	83.3
	COL-UIS	552.5	78.7
	dNTPs	1529.2	107.1
CO-036	Comercial	503	71.53
	COL-UIS	390.9	42.4
	dNTPs	918.8	69.9

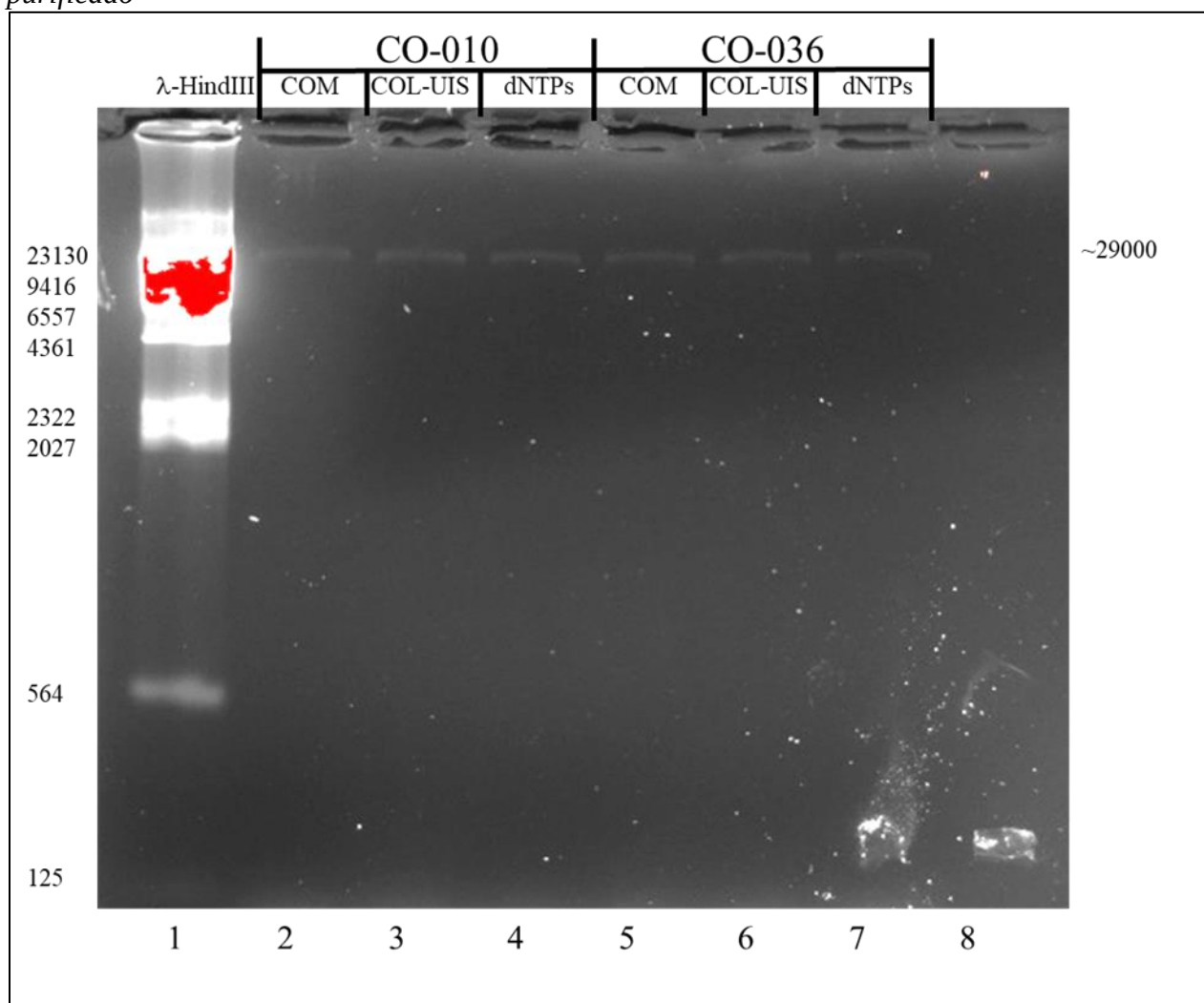
Nota: Los valores de concentración presentados son el resultado del promedio de 3 ensayos por cada tratamiento en cada muestra. CO: Colombia, 010 y 036: No. de la muestra.

Además de la validación de las concentraciones de ADNc de las dos muestras, también fue proporcionado el resultado de una electroforesis en gel de agarosa al 1% de las concentraciones de ADNc de cadena sencilla obtenido en las dos primeras etapas del protocolo de síntesis de ADNc

aplicando los tratamientos: comercial, dNTPs y COL-UIS aplicados a cada muestra (Figura 4). Las bandas se encuentran entre 23,1 kb y 29 kb. En la muestra CO-010 el patrón de migración de las bandas fue mayor para el tratamiento comercial, luego para COL-UIS y por último para dNTPs. En cuanto a la muestra CO-036 la banda con mayor migración fue la de dNTPs, luego la de COL-UIS y por último la comercial.

Figura 4

Gel de agarosa no desnaturizante al 1% con las concentraciones de ADNc de cadena sencilla purificado



Nota: λ -HINDIII= marcador utilizado, COM= tratamiento comercial, COL-UIS= tratamiento con la mezcla COL-UIS, dNTPs= tratamiento con la mezcla de dNTPs, CO= Colombia, 036 y 010= No. De la muestra.

4.4 Protocolo de secuenciación genómica de SARS-CoV-2

El protocolo con las modificaciones de los manuales empleados se encuentra adjunto en el apéndice C. De esta manera, en el manual “Ion AmpliSeq™ Library Preparation on the Ion Chef™ System” se modificó el ítem de dilución de las muestras de ADN ubicado en la página 17 del manual, al agregar las sustancias diseñadas con base en la tecnología genómica UIS. Además, en el manual “Ion 520™ & Ion 530™ Kit – OT2”, se modificó el ítem 2 de la preparación de la solución amplificadora localizada en la página 36, en donde se restó el volumen de los componentes correspondientes a sustancias químicas diseñadas a partir de la tecnología genómica UIS a la cantidad final de agua libre de nucleasas y así, mantener el volumen final solicitado por el proceso. De igual manera, en la mezcla de enzimas del ítem 4 se incluyeron mezclas diseñadas con la tecnología genómica UIS.

La secuenciación de la secuenciación genómica de SARS-CoV-2 a partir de ADNc extraído utilizando el protocolo expuesto en este proyecto se encuentra en proceso de validación por la unidad de secuenciación GENECORE de la Universidad de los Andes de Bogotá, sin embargo, por normatividad de la compañía contratada y la normatividad de la compañía Ion Torrent fabricante de los equipos necesarios para la implementación, el protocolo de secuenciación aquí expuesto no se pudo implementar. Los resultados de los genomas que proporcione GENECORE se mantendrán confidenciales hasta que sean autorizados por la dirección del proyecto y la Vicerrectoría de Investigación y Extensión de la UIS para su publicación.

5. Discusión

El análisis del genoma de SARS-CoV-2 mostró, en su composición nucleotídica, un genoma rico en AT en todas las secuencias consenso estudiadas, solo los porcentajes de GC encontrados en las secuencias consenso de enero y febrero están dentro del rango de 32% a 43% que es considerado normal para los coronavirus (Cui et al., 2019), de marzo a octubre los porcentajes de GC están por debajo de este rango, en especial octubre, el último mes dentro del estudio, de la misma manera las mutaciones en los codones aumentaron en octubre (Apéndice A) y el aumento de AT (Tabla 1), es un reflejo de las variaciones en las posiciones mutadas que tienen opción de presentar más de dos nucleótidos indicando una preferencia de mutación hacia A o T. Estos datos coinciden con otros estudios del genoma que sugieren que los nucleótidos G y C ubicados en cualquiera de las 3 posiciones de un codón mutan a A o T (Hosseini Rad SM & McLellan, 2020; Y. Wang et al., 2020).

El análisis de los sitios de poliadenilación, mostró que el sitio de control negativo TGTA, reportado para algas (Ji et al., 2015), se encuentra 36 veces en el genoma, mientras que el sitio AATAA, reportado para mamíferos (Sun et al., 2020) se encuentra 14 veces. Teniendo en cuenta que el virus infecta en un mayor riesgo a mamíferos (Martínez-Hernández et al., 2020), pocas son las explicaciones que se pueden atribuir para este resultado, además, actualmente no existen estudios que asocien sitios de poliadenilación de algas y coronavirus. Sin embargo, la información sobre los sitios de poliadenilación de los coronavirus que se tiene actualmente, ha demostrado que, tanto en cultivos celulares como en animales, la longitud de la cola poli (A) de los coronavirus es regulada a lo largo del ciclo de vida de estos, de esta manera en la etapa temprana de infección la cola poli (A) aumenta hasta alcanzar su longitud máxima y luego disminuye en la última etapa de infección (Shien et al., 2014; H.-Y. Wu et al., 2013). La eficacia de este alargamiento durante la

infección depende la longitud inicial de la cola poli (A) al momento de la infección, además, esta regulación también favorece a la eficiencia de la traducción viral (Peng et al., 2016; H.-Y. Wu et al., 2013), no obstante, el mecanismo por el cual se regula la cola de poli (A) aún es incierto.

Por otro lado, se lograron identificar los 11 ORF correspondientes a los genes *ORF1ab*, *S*, *ORF3a*, *E*, *M*, *ORF6*, *ORF7a*, *ORF7b*, *ORF8*, *N* y *ORF10* ya descritos para SARS-CoV-2 (Naqvi et al., 2020), siendo *S*, *E*, *M* y *N* genes codificantes para las proteínas estructurales de espiga, de envoltura, de membrana y de la nucleocápside, respectivamente (Fehr & Perlman, 2015), en otros estudios se mencionan 12 marcos de lectura teniendo en cuenta que el gen *ORF1ab* puede considerarse como uno o dos genes dependiendo como se traduzca el marco abierto de lectura, sin embargo siguiendo el genoma que se tomó como referencia para identificar los nucleótidos del gen en este estudio se presentó como uno.

Los gaps encontrados en los alineamientos de los consensos de los genomas mensuales se presumen ser por cambios mutacionales de deleciones o inserciones en alguna de las secuencias teniendo en cuenta que los parámetros de los consensos permitían un umbral de frecuencia de 100, es decir, cualquier cambio en alguna de las secuencias mensuales utilizadas para los consensos se vio reflejado en el resultado de la secuencia consenso final. El cambio en la longitud en algunos genes en los consensos de enero y octubre también puede considerarse como resultado de pérdida de regiones en el genoma viral como parte de su evolución (Phan, 2020), concordando con lo planteado en el Modelo de Pérdida de ADN (DNA-LM) (Martínez-Pérez et al., 2007), esto se puede comprobar con la cepa encontrada en Reino Unido denominada SARS-CoV-2 VUI 202012/01, donde entra las mutaciones encontradas se reporta una supresión en el genoma (Loconsole et al., 2021).

La traducción conceptual de los genes muestra un cambio en el número de codones degenerados desde las primeras secuencias reportadas en enero de 2020 hasta las secuencias reportadas en octubre del mismo año, en todos los genes el número de codones afectados por mutaciones aumentaron, así como las opciones aminoacídicas para codificar de los mismos dando cabida a posibles cambios en la traducción por cada opción aminoacídica. Sin embargo, cabe resaltar que se presentaron casos en los que codones degenerados en enero no se encontraban degenerados en octubre, como fue el caso del gen *ORF6* donde el octavo codón se encontraba degenerado en enero. No obstante, en octubre no lo estaba, sino que era el séptimo codón, por lo tanto hay que considerar entonces errores en el proceso de secuenciación de algunas de las secuencias utilizadas para generar los consensos (Hourdel et al., 2020).

Las consecuencias de los cambios generados en la traducción se ven reflejadas en las opciones que presentan los codones para codificar diferentes aminoácidos (Apéndice B). De las más frecuentes entre los codones son las posibilidades de codificar para una señal de paro dentro del gen, lo que traería como consecuencia la interrupción de la traducción de este en el caso de que el genoma presente las mutaciones que produzcan estas señales, otros casos son los genes que presentan codones degenerados con la posibilidad de codificar para metioninas dentro del gen, esto enviaría señales de inicio de traducción dentro del mismo gen. El caso del gen *ORF7ab*, el cual puede codificarse como un solo gen o dos distintos dependiendo de las mutaciones que se encuentren los codones que comparten los dos genes, es un claro ejemplo de cómo las mutaciones afectan la expresión genética del virus y podrían ser una posible respuesta a cómo varían las diferentes cepas a nivel mundial (S. Wu et al., 2021).

Otro resultado interesante del análisis del genoma de SARS-CoV-2 son las mutaciones que pareciera que asegurasen que ciertas regiones presenten una señal debida, por ejemplo, el caso del

gen *N* en donde aun cuando el último codón del gen se encuentra degenerado las dos opciones de traducción que tiene son señales de paro, de igual manera la región final del gen *ORF7b* la cual presenta varias mutaciones antes del codón de paro teórico, las opciones aminoacídicas de los codones degenerados presentan señales de paro, esto pareciera que asegurase que la traducción del gen sea *ORF7ab* u *ORF7b* termine en esa región para que así el ORF no se corra y no intervenga con la señal de codón de inicio del siguiente gen.

Los análisis previos que se le hicieron al genoma, entre estos la identificación de su composición nucleotídica fueron necesarios, fundamentales e imprescindibles para entender los cambios que se van presentando en el virus e identificar las regiones que podrían denominarse como conservadas, producto de la selección natural, con el fin de utilizarse para el diseño de los cebadores. Además, estos análisis evolutivos funcionaron para evitar el uso de las regiones con alto número de mutaciones para el diseño molecular (C. Wang et al., 2020). No obstante, no todos los cebadores diseñados funcionaron, varias de estas regiones que se creían conservadas presentaron cambios mutacionales en los meses siguientes a los empleados para el diseño (Apéndice A), esto reafirma la necesidad de un seguimiento de las regiones con bajo o alto cambio mutacional en el genoma de SARS-CoV-2 para el estudio biológico del virus y para mejorar el proceso de diagnóstico y monitoreo clínico (C. Wang et al., 2020).

La eficiencia de los demás cebadores más los polilinker demuestra la calidad del diseño con relación a los diseños de ingeniería genética aplicados, lo cual se corroboró en la implementación del protocolo de síntesis de ADNc donde se utilizó uno de ellos. Además, la mezcla COL-UIS-dNTPs sustentada en la tecnología genómica UIS demostró aportar una mayor versatilidad al proceso de síntesis, gracias a que los dNTPs son capaz de aportar las bases para la construcción de la cadena de ADNc (Mejía-Ospino et al., 2021). Por otro lado, la implementación

del protocolo con la mezcla COL-UIS no fue tan eficiente como se esperaba, posiblemente se necesite aumentar las concentraciones a utilizar de acuerdo con las concentraciones de ARN viral en las muestras, el reto de lo anterior es que la cantidad de ARN viral de la muestra depende del nivel de infección del paciente lo cual aún es difícil de determinar.

En cuanto a la inclusión de las mezclas COL-UIS-dNTPs y COL-UIS en los protocolos de secuenciación del genoma de SARS-CoV-2, se contrató a una empresa nacional que tenía la tecnología necesaria para la implementación del protocolo, sin embargo, se presentó la problemática que aun cuando la empresa Ion Torrent, fabricante de los equipos necesarios para el desarrollo del protocolo, permitieron hacer la inclusión de las mezclas en ciertas partes de los protocolos ya establecidos por ellas (Apéndice C), se recibió una negativa en el uso de los equipos para la inclusión de la tecnología genómica UIS en el proceso de secuenciación, ya que por normatividad de la empresa se perdía la garantía. No obstante, como una excepción, se permitió la inclusión del ADNc generado por la tecnología genómica UIS previo a la construcción de la librería, actualmente se está en la espera de este resultado.

Los resultados de la secuenciación serán confidenciales hasta que el Instituto Nacional de Salud, el Ministerio de Ciencia, Tecnología e Innovación de Colombia y la Dirección de Propiedad Intelectual de la Universidad Industrial de Santander, permitan su divulgación y publicación.

6. Conclusiones

El diseño de los cebadores fundamentados en los criterios de la tecnología genómica UIS contribuyen de manera eficaz en los procesos de síntesis ADNc para la amplificación del genoma de SARS-CoV-2 rico en AT.

Las condiciones determinadas en los protocolos expuestos permitieron potenciar la síntesis de ADNc para la secuenciación genómica de SARS-CoV-2, mediante la adición de las mezclas COL-UIS y COL-UIS-dNTPs propuestas por la tecnología genómica UIS.

Las condiciones establecidas con base a la tecnología genómica UIS en el protocolo de secuenciación contribuyen a secuenciar el genoma completo de SARS-CoV-2 y se ajustan para ser aplicadas con la tecnología Ion Torrent de NGS.

La secuenciación del genoma completo del virus SARS-CoV-2 es posible al aplicar los principios de la tecnología genómica UIS basada en el estudio de las condiciones fisicoquímicas del genoma, por lo tanto, al incluirla para el diseño de los cebadores, la síntesis de ADNc y la determinación de las concentraciones de las mezclas COL-UIS y COL-UIS-dNTPs en los protocolos propuestos, se favorece el proceso de secuenciación de genomas ricos en AT con la tecnología Ion Torrent.

Referencias Bibliográficas

- Alvarez, D., Laiton-Donato, K., Franco, C., & Mercado, 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. *Revista Biomédica*, 40(2), 188–197. <https://doi.org/10.7705/biomedica.5841>
- Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C., & Garry, R. F. (2020). The proximal origin of SARS-CoV-2. *Nature Medicine*, 26(4), 450–452. <https://doi.org/10.1038/s41591-020-0820-9>
- Arevalo-Rodriguez, I., Buitrago-Garcia, D., Simancas-Racines, D., Zambrano, P., Campo, R., Ciapponi, A., Sued, O., García, L., Rutjes, A., Low, N., Bossuyt, P., Perez-Molina, J., & Zamora, J. (2020). False-negative results of initial RT-PCR assays for COVID-19: A systematic review. *PLoS ONE*, 15(12), e0242958. <https://doi.org/10.1371/journal.pone.0242958>
- Bachelet, V. (2020). Do we know the diagnostic properties of the tests used in COVID-19? A rapid review of recently published literature. *Medwave*, 20(3), e7891–e7891. <https://doi.org/10.5867/medwave.2020.03.7891>
- Bastos, M., Tavaziva, G., Abidi, S., Campbell, J., Haraoui, L.-P., Johnston, J., Lan, Z., Law, S., MacLean, E., Trajman, A., Menzies, D., Benedetti, A., & Ahmad Khan, F. (2020). Diagnostic accuracy of serological tests for covid-19: Systematic review and meta-analysis. *BMJ*, 370(m2516), 1–13. <https://doi.org/10.1136/bmj.m2516>
- Bellaousov, S., Reuter, J. S., Seetin, M. G., & Mathews, D. H. (2013). RNAstructure: Web servers for RNA secondary structure prediction and analysis. *Nucleic Acids Research*, 41(Web Server issue), 471–474. <https://doi.org/10.1093/nar/gkt290>

- Böger, B., Fachi, M. M., Vilhena, R. O., Cobre, A. F., Tonin, F. S., & Pontarolo, R. (2021). Systematic review with meta-analysis of the accuracy of diagnostic tests for COVID-19. *American Journal of Infection Control*, 49(1), 21–29. <https://doi.org/10.1016/j.ajic.2020.07.011>
- Boratyn, G. M., Camacho, C., Cooper, P. S., Coulouris, G., Fong, A., Ma, N., Madden, T. L., Matten, W. T., McGinnis, S. D., Merezuk, Y., Raytselis, Y., Sayers, E. W., Tao, T., Ye, J., & 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>
- Candido, D. S., Claro, I. M., de Jesus, J. G., Souza, W. M., Moreira, F. R. R., Dellicour, S., Mellan, T. A., du Plessis, L., Pereira, R. H. M., Sales, F. C. S., Manuli, E. R., Thézé, J., Almeida, L., Menezes, M. T., Voloch, C. M., Fumagalli, M. J., Coletti, T. M., da Silva, C. A. M., Ramundo, M. S., ... Faria, N. R. (2020). Evolution and epidemic spread of SARS-CoV-2 in Brazil. *Science*, 369(6508), 1255–1260. <https://doi.org/10.1126/science.abd2161>
- CDC. (2020). Real-Time RT-PCR diagnostic panel. In *Centers for Disease Control and Prevention: Vol. CDC-006-00*. <https://www.fda.gov/media/134922/download>
- Chan, J., Yuan, S., Kok, K.-H., To, K., Chu, H., Yang, J., Xing, F., Liu, J., Yip, C., Poon, R., Tsoi, H.-W., Lo, S., Chan, K.-H., Poon, V., Chan, W.-M., Ip, J., Cai, J.-P., Cheng, V., Chen, H., & Yuen, K.-Y. (2020). A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *The Lancet*, 395(10223), 514–523. [https://doi.org/10.1016/S0140-6736\(20\)30154-9](https://doi.org/10.1016/S0140-6736(20)30154-9)
- Chiara, M., D’Erchia, A. M., Gissi, C., Manzari, C., Parisi, A., Resta, N., Zambelli, F., Picardi, E., Pavesi, G., Horner, D. S., & Pesole, G. (2020). Next generation sequencing of SARS-CoV-2 genomes: challenges, applications and opportunities. *Briefings in Bioinformatics*, 00(00), 1–

15. <https://doi.org/10.1093/bib/bbaa297>
- Cui, J., Li, F., & Shi, Z.-L. (2019). Origin and evolution of pathogenic coronaviruses. *Nature Reviews Microbiology*, 17(3), 181–192. <https://doi.org/10.1038/s41579-018-0118-9>
- Damerla, R. R., Chatterjee, B., Li, Y., Francis, R. J. B., Fatakia, S. N., & Lo, C. W. (2014). Ion Torrent sequencing for conducting genome-wide scans for mutation mapping analysis. *Mammalian Genome : Official Journal of the International Mammalian Genome Society*, 25(3–4), 120–128. <https://doi.org/10.1007/s00335-013-9494-7>
- Deeks, J., Dinnes, J., Takwoingi, Y., Davenport, C., Spijker, R., Taylor-Phillips, S., Adriano, A., Beese, S., Dretzke, J., Ferrante di Ruffano, L., Harris, I., Price, M., Dittrich, S., Emperador, D., Hooft, L., Leeftang, M., Bruel, A., & Group, C. (2020). Antibody tests for identification of current and past infection with SARS-CoV-2. *Cochrane Database of Systematic Reviews*, 6(6), CD013652. <https://doi.org/10.1002/14651858.cd013652>
- Dixon, H. B. F., Bielka, H., & Cantor, C. R. (1986). Nomenclature committee for the International Union of Biochemistry (NC-IUB). Nomenclature for incompletely specified bases in nucleic acid sequences. Recommendations 1984. *Proceedings of the National Academy of Sciences of the United States of America*, 83(1), 4–8. <https://doi.org/10.1073/pnas.83.1.4>
- Eduardo Rojas Pineda. (2020). *Invitación a presentar proyectos que contribuyan a la solución de problemáticas actuales de salud relacionadas con la pandemia de COVID-19*.
- Fani, M., Teimoori, A., & Ghafari, S. (2020). Comparison of the COVID-2019 (SARS-CoV-2) pathogenesis with SARS-CoV and MERS-CoV infections. *Future Virology*, 15(5), 317–323. <https://doi.org/10.2217/fvl-2020-0050>
- Fehr, A. R., & Perlman, S. (2015). Coronaviruses. In H. J. Maier, E. Bickerton, & P. Britton (Eds.), *Methods in Molecular Biology* (Vol. 1282, Issue 1, pp. 1–23). Springer New York.

<https://doi.org/10.1007/978-1-4939-2438-7>

Gilchrist, C. A., Turner, S. D., Riley, M. F., Petri, W. A. J., & 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>

Grubaugh, N. D., Ladner, J. T., Lemey, P., Pybus, O. G., Rambaut, A., Holmes, E. C., & 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>

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

Hosseini Rad SM, A., & McLellan, A. D. (2020). Implications of SARS-CoV-2 Mutations for Genomic RNA Structure and Host microRNA Targeting. *International Journal of Molecular Sciences*, 21(13), 4807. <https://doi.org/10.3390/ijms21134807>

Hou, W. (2020). Characterization of codon usage pattern in SARS-CoV-2. *Virology Journal*, 17(1), 138. <https://doi.org/10.1186/s12985-020-01395-x>

Hourdel, V., Kwasiborski, A., Balière, C., Matheus, S., Batéjat, C. F., Manuguerra, J.-C., Vanhomwegen, J., & Caro, V. (2020). Rapid Genomic Characterization of SARS-CoV-2 by Direct Amplicon-Based Sequencing Through Comparison of MinION and Illumina iSeq100(TM) System. *Frontiers in Microbiology*, 11, 571328. <https://doi.org/10.3389/fmicb.2020.571328>

Ji, G., Wu, X., Li, L., Li, Q. Q., Wu, X., Fu, J., & Chen, G. (2015). PASPA: A web server for mRNA poly(A) site predictions in plants and algae. *Bioinformatics*, 31(10), 1671–1673. <https://doi.org/10.1093/bioinformatics/btv004>

- Loconsole, D., Sallustio, A., Accogli, M., Centrone, F., Capozzi, L., Del Sambro, L., Parisi, A., & Chironna, M. (2021). Genome Sequence of a SARS-CoV-2 VUI 202012/01 Strain Identified from a Patient Returning from London, England, to the Apulia Region of Italy. *Microbiology Resource Announcements*, 10(4), e01487-20. <https://doi.org/10.1128/MRA.01487-20>
- Martínez-Hernández, F., Isaak-Delgado, A. B., Alfonso-Toledo, J. A., Muñoz-García, C. I., Villalobos, G., Aréchiga-Ceballos, N., & Rendón-Franco, E. (2020). Assessing the SARS-CoV-2 threat to wildlife: Potential risk to a broad range of mammals. *Perspectives in Ecology and Conservation*, 18(4), 223–234. <https://doi.org/10.1016/j.pecon.2020.09.008>
- Martínez-Pérez, F., Durán-Gutiérrez, D., Delaye, L., Becerra, A., Aguilar, G., & 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>
- 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., & Martínez-Fong, D. (2021). *Mezcla de nucleótidos para la amplificación y secuenciación de polímeros de ácidos nucleicos* (Patente de Colombia. No. NC2017/0004500). Universidad Industrial de Santander. Superintendencia de Industria y Comercio.
- Moore, J. B., & June, C. H. (2020). Cytokine release syndrome in severe COVID-19. *Science*, 368(6490), 473–474. <https://doi.org/10.1126/science.abb8925>
- Nakagawa, S., & Miyazawa, T. (2020). Genome evolution of SARS-CoV-2 and its virological characteristics. *Inflammation and Regeneration*, 40(17), 1–7. <https://doi.org/10.1186/s41232-020-00126-7>
- Naqvi, A. A. T., Fatima, K., Mohammad, T., Fatima, U., Singh, I. K., Singh, A., Atif, S. M.,

- Hariprasad, G., Hasan, G. M., & Hassan, M. I. (2020). Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: Structural genomics approach. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1866(10), 165878. <https://doi.org/10.1016/j.bbadis.2020.165878>
- Owczarzy, R., Tataurov, A. V., Wu, Y., Manthey, J. A., McQuisten, K. A., Almabrazi, H. G., Pedersen, K. F., Lin, Y., Garretson, J., McEntaggart, N. O., Sailor, C. A., Dawson, R. B., & Peek, A. S. (2008). IDT SciTools: a suite for analysis and design of nucleic acid oligomers. *Nucleic Acids Research*, 36(Web Server issue), W163–W169. <https://doi.org/10.1093/nar/gkn198>
- Peng, Y.-H., Lin, C.-H., Lin, C.-N., Lo, C.-Y., Tsai, T.-L., & Wu, H.-Y. (2016). Characterization of the Role of Hexamer AGUAAA and Poly(A) Tail in Coronavirus Polyadenylation. *PLoS One*, 11(10), e0165077–e0165077. <https://doi.org/10.1371/journal.pone.0165077>
- Phan, T. (2020). Genetic diversity and evolution of SARS-CoV-2. *Infection, Genetics and Evolution*, 81, 104260. <https://doi.org/10.1016/j.meegid.2020.104260>
- Roberts, R. J., Vincze, T., Posfai, J., & Macelis, D. (2010). REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Research*, 38(1), D234–D236. <https://doi.org/10.1093/nar/gkp874>
- Roche. (2006). Primer Design and Template Preparation. In H.-J. Degen, A. Deufel, D. Eisel, Grünewald-Janho, & J. Stefanie Keesey (Eds.), *PCR Applications Manual* (3rd ed., pp. 41–60).
- Sayers, E. W., Cavanaugh, M., Clark, K., Ostell, J., Pruitt, K. D., & Karsch-Mizrachi, I. (2019). GenBank. *Nucleic Acids Research*, 47(D1), D94–D99. <https://doi.org/10.1093/nar/gky989>
- Segreto, R., & Deigin, Y. (2021). The genetic structure of SARS-CoV-2 does not rule out a

- laboratory origin. *BioEssays*, 43(3), 2000240. <https://doi.org/10.1002/bies.202000240>
- Shen, Z., Xiao, Y., Kang, L., Ma, W., Shi, L., Zhang, L., Zhou, Z., Yang, J., Zhong, J., Yang, D., Guo, L., Zhang, G., Li, H., Xu, Y., Chen, M., Gao, Z., Wang, J., Ren, L., & Li, M. (2020). Genomic Diversity of Severe Acute Respiratory Syndrome-Coronavirus 2 in Patients With Coronavirus Disease 2019. *Clinical Infectious Diseases : An Official Publication of the Infectious Diseases Society of America*, 71(15), 713–720. <https://doi.org/10.1093/cid/ciaa203>
- Shien, J.-H., Su, Y.-D., & Wu, H.-Y. (2014). Regulation of coronaviral poly(A) tail length during infection is not coronavirus species- or host cell-specific. *Virus Genes*, 49(3), 383–392. <https://doi.org/10.1007/s11262-014-1103-7>
- Shu, Y., & McCauley, J. (2017). GISAID: Global initiative on sharing all influenza data - from vision to reality. *Euro Surveillance : Bulletin Europeen Sur Les Maladies Transmissibles = European Communicable Disease Bulletin*, 22(13), 30494. <https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494>
- Sirotkin, K., & Sirotkin, D. (2020). Might SARS-CoV-2 Have Arisen via Serial Passage through an Animal Host or Cell Culture?: A potential explanation for much of the novel coronavirus' distinctive genome. *BioEssays*, 42, 2000091. <https://doi.org/10.1002/bies.202000091>
- Sun, Y., Hamilton, K., & Tong, L. (2020). Recent molecular insights into canonical pre-mRNA 3'-end processing. *Transcription*, 11(2), 83–96. <https://doi.org/10.1080/21541264.2020.1777047>
- Umakanthan, S., Chattu, V. K., Ranade, A. V., Das, D., Basavarajegowda, A., & Bukelo, M. (2021). A rapid review of recent advances in diagnosis, treatment and vaccination for COVID-19. *AIMS Public Health*, 8(1), 137–153. <https://doi.org/10.3934/publichealth.2021011>

- Wacharapluesadee, S., Tan, C. W., Maneeorn, P., Duengkae, P., Zhu, F., Joyjinda, Y., Kaewpom, T., Chia, W. N., Ampoot, W., Lim, B. L., Worachotsueptrakun, K., Chen, V. C.-W., Sirichan, N., Ruchisrisarod, C., Rodpan, A., Noradechanon, K., Phaichana, T., Jantararat, N., Thongnumchaima, B., ... Wang, L.-F. (2021). Evidence for SARS-CoV-2 related coronaviruses circulating in bats and pangolins in Southeast Asia. *Nature Communications*, 12(1), 972. <https://doi.org/10.1038/s41467-021-21240-1>
- Wang, C., Liu, Z., Chen, Z., Huang, X., Xu, M., He, T., & Zhang, Z. (2020). The establishment of reference sequence for SARS-CoV-2 and variation analysis. *Journal of Medical Virology*, 92(6), 667–674. <https://doi.org/10.1002/jmv.25762>
- Wang, Y., Mao, J.-M., Wang, G.-D., Luo, Z.-P., Yang, L., Yao, Q., & Chen, K.-P. (2020). Human SARS-CoV-2 has evolved to reduce CG dinucleotide in its open reading frames. *Scientific Reports*, 10(1), 12331. <https://doi.org/10.1038/s41598-020-69342-y>
- Watson, J., Whiting, P. F., & Brush, J. E. (2020). Interpreting a covid-19 test result. *BMJ*, 369, m1808. <https://doi.org/10.1136/bmj.m1808>
- Weissleder, R., Lee, H., Ko, J., & Pittet, M. J. (2020). COVID-19 diagnostics in context. *Science Translational Medicine*, 12(546), 1–5. <https://doi.org/10.1126/scitranslmed.abc1931>
- WHO. (2020). WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. <https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020>
- Woloshin, S., Patel, N., & Kesselheim, A. S. (2020). False Negative Tests for SARS-CoV-2 Infection - Challenges and Implications. *The New England Journal of Medicine*, 383(6), e38. <https://doi.org/10.1056/NEJMp2015897>
- Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., Hu, Y., Tao, Z.-W., Tian, J.-H.,

- Pei, Y.-Y., Yuan, M.-L., Zhang, Y.-L., Dai, F.-H., Liu, Y., Wang, Q.-M., Zheng, J.-J., Xu, L., Holmes, E. C., & Zhang, Y.-Z. (2020). A new coronavirus associated with human respiratory disease in China. *Nature*, 579(7798), 265–269. <https://doi.org/10.1038/s41586-020-2008-3>
- Wu, H.-Y., Ke, T.-Y., Liao, W.-Y., & Chang, N.-Y. (2013). Regulation of coronaviral poly(A) tail length during infection. *PloS One*, 8(7), e70548. <https://doi.org/10.1371/journal.pone.0070548>
- Wu, S., Tian, C., Liu, P., Guo, D., & Zheng, W. (2021). Effects of SARS-CoV-2 mutations on protein structures and intraviral protein-protein interactions. *Journal of Medical Virology*, 93, 0–2. <https://doi.org/10.1002/jmv.26597>
- Zhang, T., Wu, Q., & Zhang, Z. (2020). Probable Pangolin Origin of SARS-CoV-2 Associated with the COVID-19 Outbreak. *Current Biology: CB*, 30(7), 1346-1351.e2. <https://doi.org/10.1016/j.cub.2020.03.022>
- Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., Huang, C.-L., Chen, H.-D., Chen, J., Luo, Y., Guo, H., Jiang, R.-D., Liu, M.-Q., Chen, Y., Shen, X.-R., Wang, X., ... Shi, Z.-L. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, 579(7798), 270–273. <https://doi.org/10.1038/s41586-020-2012-7>
- 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., & 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>