Análisis bioinformático de los cebadores para los genes de Hemaglutinina, Proteína Nuclear y Proteínas de Matriz del virus de Influenza A H1N1 empleados en el diagnóstico a pacientes por RT-PCR en tiempo real, de 2019 a agosto de 2020

Lizeth Johana Forero Buitrago

Trabajo de Grado para Optar al Título de Bióloga

Director

Francisco José Martínez Pérez

Doctor en Ciencias

Codirector

Carlos Jaime Barrios Hernández

Doctor en Ciencias

Universidad Industrial de Santander

Facultad de Ciencias

Escuela de Biología

Bucaramanga

2021

Dedicatoria

A mis padres Elide Buitrago Martínez y José Rafael Forero por ser los seres que amo en esta vida, y que son quienes siempre han creído en mí.

A mi compañero de vida Enem Pinzón Montañez por apoyarme y darme ánimos en todo momento.

A mi profesor de Biología Celular, Dr. Francisco José Martínez Pérez por ayudarme en el proceso de formación.

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 número de contrato 369-2020 y código No. 1102101576900, 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. Además, a la Vicerrectoría de Investigación y Extensión de la Universidad Industrial de Santander por el financiamiento con el número de proyecto 76900.

Al Dr. Francisco José Martínez Pérez y al Dr. Carlos Jaime Barrios Hernández por permitirme hacer parte de su grupo de trabajo.

A la Universidad Industrial de Santander.

A la Facultad de Ciencias de la Universidad Industrial de Santander.

A la Escuela de Biología de la Universidad Industrial de Santander.

Al Laboratorio de Genómica de Celomados (GCL-UIS).

Al Centro de Supercomputación Cálculo Científico SC3-UIS.

Al Grupo de Investigación Computo Avanzado y a Gran Escala (CAGE).

A mis familiares, colegas y amigos por todo su apoyo.

A mi colega Biólogo Cristian Enrique Cadena Caballero por su colaboración y orientación durante el proceso de ejecución de este trabajo. A todos ustedes lectores.

Tabla de contenido

Pág.

Introducción 12

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 20

3. Metodología 21

3.1 Identificación de los genes de HA, NP y M1-M2 del virus de la Influenza A H1N1 en la base de datos IRD 21

3.2 Alineamiento múltiple y depuración de las secuencias de los genes de HA, NP y M1-M2 del virus de la Influenza A H1N1 21

3.3 Generación de secuencias nucleotídicas consenso de los genes de HA, NP y M1-M2 del virus de la Influenza A H1N1 22

3.4 Validación *in silico* de los cebadores creados para los genes de HA, NP y M1-M2 de la Influenza A H1N1 en el GenBank 23

4. Resultados 25

4.1 Identificación de las secuencias genéticas de los genes HA, NP y M1-M2 25

4.2 Alineamiento múltiple y depuración de los genes de HA, NP y M1-M2 27

4.3 Generación de secuencias nucleotídicas consenso de los genes de HA, NP y M1-M2 del virus de la Influenza A H1N1 27

Validación de los cebadores propuestos para los genes de HA, NP y M1-M2 30

5. Discusión 40

6. Conclusiones 44

Referencias bibliográficas 46

Lista de tablas

Tabla 1. Porcentajes de purinas, pirimidinas y nucleótidos degenerados del gen HA 29

Tabla 2. Porcentaje de purinas y pirimidinas para el gen de la HA, NP y M1-M2 30

Tabla 3. Secuencias de cebadores y oligonucleótidos sustrato.....31

Tabla 4. Sondas Taqman para RT-PCR en Tiempo Real de un Paso.....32

Tabla 5. Secuencia de cebadores para secuenciación genómica.....32

Tabla 6. Evaluación de los cebadores con la supercomputadora GUANE-1.....35

Tabla 7. Temperaturas de los cebadores, sondas y oligonucleótidos.....376

Lista de figuras

Figura 1. Secuencias de personas infectadas con el virus de la Influenza A 26

Figura 2. Estructuras secundarias de las sondas y cebadores para RT-PCR 36

Figura 3. Estructuras secundarias para oligonucleótidos de RT-PCR.....37

Figura 4. RT-PCR de los genes del virus de Influenza A H1N1 399

Lista de apéndices

Ver apéndices adjuntos y pueden ser consultados en la base de datos de la Biblioteca UIS

Apéndice A. Búsqueda en IRD mediante el uso de palabras clave (captura de pantalla).

Apéndice B. Secuencias en formato FASTA de personas con el virus de Influenza A H1N1.

Apéndice C. Secuencias, alineamientos y consensos en formato FASTA del 2019-2020.

Apéndice D. Archivo FASTA con los genes HA, NP y M1-M2 (captura de pantalla).

Apéndice E. Posición del codón de inicio y paro de los genes en estudio (captura de pantalla).

Apéndice F. Alineamiento en formato FASTA de los genes HA, NP y M1-M2.

Apéndice G. Hoja de cálculo con porcentajes mensuales de nucleótidos al 20% y 100%.

Apéndice H. Archivos de la validación de los cebadores y sonda diseñados en BLAST.

Apéndice I. Documento de los consensos al 20% y 100% con los cebadores y sonda diseñados.

Apéndice J. Validación de las secuencias consenso con los cebadores y sonda (captura de pantalla).

Glosario

Cebador: es una secuencia corta de ADN de cadena simple que se utiliza en la reacción en cadena de la polimerasa (PCR) & en la síntesis de ADN complementario (ADNc).

Fw: forward, cebador correspondiente a la cadena de ADN en dirección sentido.

HA. hemaglutinina.

IRD: es la base de datos de la investigación de la Influenza (Influenza Research Database).

M1-M2: proteína matriz 1 y proteína de matriz 2.

NA: neuraminidasa.

NCBI: plataforma de las secuencias genómicos (National Center for Biotechnology Information).

NP: nucleoproteína.

OMS: sigla de la Organización Mundial de la Salud.

RT-PCR: la reacción en cadena de la polimerasa con transcriptasa inversa o reversa.

Rv: reverse, cebador correspondiente a la cadena de ADN en dirección anti-sentido.

S4-S5-S7: segmento del gen.

TqM: sonda que genera la señal luminosa en una reacción de polimerización en cadena de RT-PCR en tiempo real (Quince Name Primer 5'-3').

Tm Al: temperatura media de alineamiento en la reacción de PCR o RT-PCR.

Tm: temperatura media de fusión del ADN.

Resumen

Título: Análisis bioinformático de los cebadores para los genes de Hemaglutinina, Proteína Nuclear y Proteínas de Matriz del virus de Influenza A H1N1 empleados en el diagnóstico a pacientes por RT-PCR en tiempo real, de 2019 a agosto de 2020 *****

Autor: Lizeth Johana Forero Buitrago**

Palabras Clave: Bioinformática, Influenza A H1N1, Cebadores de RT-PCR, Diagnóstico molecular.

Descripción:

El virus de la Influenza A H1N1 es un virus zoonótico de genoma monocatenario de ARN en sentido negativo que infecta humanos, cerdos, aves y otros animales. Los diagnósticos epidemiológicos indican que el virus continúa mutando, lo cual, origina nuevas epidemias y pandemias. Las pruebas de RT-PCR con los genes HA, NP y M1-M2 autorizados por la OMS, son la forma más rápida para detectar personas infectadas, pero con la alta tasa mutacional del virus se ha evidenciado que los cebadores y las sondas comerciales no amplifican y generan falsos negativos, aunque las vacunas para la Influenza son actualizadas año a año, las pruebas para diagnóstico no lo son. Los análisis bioinformáticos realizados por el grupo CAGE desde 1991-2019, han demostrado que existen nuevos patrones mutacionales que generan diagnósticos incorrectos o que reconocen otros virus de Influenza. Para descartar falsos negativos, se crearon nuevos cebadores, sondas y oligonucleótidos de los genes HA, NP y M1-M2 del virus de la Influenza A H1N1 y se evaluó su eficiencia *in silico* para implementarlos en los diagnósticos de RT-PCR. Se obtuvieron 7362 secuencias de IRD y se validaron con el software Sequence Manager y un código de selección en la supercomputadora GUANE-1, obteniendo 5872 secuencias para el año 2019 y 1490 para el año 2020. Se construyó una base de datos que permitió validar la eficiencia de los cebadores y sondas, al conocer los nucleótidos mayoritarios A-T y los degenerados R-Y en las secuencias consenso. Se corroboraron los alineamientos en BLAST con los cebadores y sondas de los nucleótidos mayoritarios, los cuales reconocieron sus respectivas secuencias para los años 2019-2020. Se observó con algoritmos-Zuker las secuencias diseñadas fueron eficientes en RT-PCR con el sustrato sintético. En conclusión, el sistema de diagnóstico puede ser eficiente para su aplicación en pacientes.

---

*Trabajo de Grado

**Facultad de Ciencias, Escuela de Biología. Director Francisco José Martínez Pérez. Codirector Carlos Jaime Barrios Hernández.

Abstract

Title: Bioinformatic analysis of the primers for the Hemagglutinin, Nuclear Protein and Matrix Protein genes of the Influenza A H1N1 virus used in the diagnosis of patients by real-time RT-PCR, from 2019 to August 2020*

Author: Lizeth Johana Forero Buitrago**

Key Words: Bioinformatics, Influenza A H1N1, RT-PCR primers, Molecular diagnosis.

Description:

Influenza A H1N1 virus is a single-stranded RNA negative-sense zoonotic virus that infects humans, pigs, birds, and other animals. Epidemiological diagnoses indicate that virus continues generate mutations, which causes new epidemics and pandemics. RT-PCR tests with the HA, NP and M1-M2 genes authorized by WHO are the fastest way to detect infected people, but with the high mutational rate of the virus it has been shown that commercial primers and probes do not they amplify and generate false negatives, although vaccines for influenza are updated from year to year, diagnostic tests are not. Bioinformatic analyzes carried out by the CAGE group from 1991-2019 have shown that there are new mutational patterns that generate incorrect diagnoses or that recognize other influenza viruses. To eliminate the false negatives results, new primers, probes and oligonucleotides of the HA, NP and M1-M2 genes of the Influenza A H1N1 virus were created and their *in silico* efficiency was evaluated to implement them in RT-PCR diagnostics. 7362 IRD sequences were obtained and validated with the Sequence Manager software and a selection code in the GUANE-1 supercomputer, obtaining 5872 sequences for the year 2019 and 1490 for the year 2020. A database was created to allow the validation of the efficiency of the primers and probes, by knowing the majority A-T nucleotides and the degenerate R-Y nucleotides in the consensus sequences. The BLAST alignments were corroborated with the primers and probes of the majority nucleotides, which recognized their respective sequences for the years 2019-2020. It was observed with Zuker algorithms the designed sequences were efficient in RT-PCR with the synthetic substrate. In conclusion, the diagnostic system can be efficient for its application in patients.

---

*Bachelor thesis

**Science Faculty. School of Biology. Director Francisco José Martínez Pérez. Codirector Carlos Jaime Barrios Hernández.

Introducción

La dinámica y el diagnóstico de las enfermedades ocasionadas por virus están relacionadas en parte, por el grado de interacción que tienen con los seres humanos respecto a los aspectos sociales que los involucran (Saccaro, Galfo, Ferranti, Russo, & Menichetti, 2020). Es por ello, que los brotes y cambios mutacionales del virus de la gripa Influenza A H1N1 siguen siendo de interés a nivel mundial (Kim, Vivas Barber, & Lee, 2020), debido a que son uno de los tipos virales más comunes presentes en la humanidad (Jiang et al., 2020). Este virus tuvo impacto en la salud humana y en el desarrollo social durante los años 1918 y 2009 por las pandemias que generaron, por lo que siguen siendo un motivo de estudio debido a los cambios mutacionales desconocidos que ocurren en los nuevos brotes virales (Sullivan, Jacobson, Dowdle, & Poland, 2010).

El éxito del virus depende de la capacidad de transmisibilidad y el potencial infectivo para conservar la viabilidad de replicarse eficientemente en un nuevo huésped (Singanayagam et al., 2020). Por esta razón, se considera que los virus de la Influenza son impredecibles ya que en cualquier momento pueden surgir nuevas cepas contra las cuales existe poca o ninguna inmunidad preexistente en la población humana (Deng et al., 2020). Los virus de la influenza pertenecen a la familia de los *Orthomyxoviridae* (Taubenberger & Kash, 2010). Los tipos del virus de Influenza que causan mayor afectación a la población humana son A y B, mientras que el tipo C causa enfermedades esporádicas principalmente en la aves (Te Velthuis & Fodor, 2016). Los del diferentes subtipos del virus de Influenza A H1N1 se determinan en función de las glicoproteínas que se encuentran en la superficie de la cápside del virus, Hemaglutinina (HA) y Neuraminidasa (NA) (Fujii, Goto, Watanabe, Yoshida, & Kawaoka, 2003; Li et al., 2020). En la actualidad se han confirmado 18 subtipos para la Hemaglutinina, estos a su vez se

dividen en dos pequeños grupos según la deriva antigénica de la proteína; el primer grupo consta de H1, H2, H6, H8, H9, H11, H12, H13, H16, H17 y H18; en el segundo grupo están las variedades H3, H4, H7, H10, H14 Y H15 (Webster & Govorkova, 2015). La Neuraminidasa tiene 9 variantes que son las que le permiten al virus hacer la digestión de los nuevos viriones en forma de secreciones fuera de las células del huésped (Labella & Merel, 2013).

Actualmente, además del subtipo H1N1, los subtipos H5N1, H7N7 y H9N2, también se consideran amenazas ambientales, ya que prevalecen en las aves y pueden infectar a los humanos directamente (Paterson & Fodor, 2012). El genoma de los virus de la Influenza A es alrededor de 13 kilobases de ARN monocatenario de sentido negativo segmentado, estos virus han desarrollado, diferentes mecanismos moleculares que proporcionan la expresión de múltiples proteínas a partir de un solo segmento de gen (Vasin et al., 2014). En este sentido, el genoma de los virus de la Influenza A y B está en 8 segmentos y en siete para los virus de la Influenza C (Paterson & Fodor, 2012; Te Velthuis & Fodor, 2016). Además, el genoma viral es replicado por la ARN polimerasa dependiente del ARN viral (ARNAP), la cual comienza en el extremo 3’ de cada gen para iniciar la síntesis del nuevo ARN viral (Robb et al., 2016), posteriormente, en el núcleo de una célula infectada se importan los ocho segmentos de ARN en un complejo de ribonucleoproteína (RNP) (Gog et al., 2007; Hutchinson & Fodor, 2012) cuya traducción genera las subunidades de la polimerasa (PA, PB1 y PB2) y la nucleoproteína NP las cuales son las responsables de la expresión del genoma. Estas proteínas se asocian a cada segmento de ARN viral para constituir las ribonucleoproteínas virales (RNPv) (Rodriguez-Frandsen et al., 2016). Además se codifican los genes para las proteínas estructurales y funcionales, HA, la NA y las proteínas Matriz (M1- M2) que se encuentran recubriendo toda la cápside del complejo viral (Charyasriwong, Haruyama, & Kobayashi, 2016; Krammer, 2015; Oropeza Fernández et al., 2011; Wunderlich et al., 2009). La HA es una glucoproteína y se encarga de infectar a las células del huésped, mientras que la proteína NA

funciona principalmente como un liberador de viriones al separar los carbohidratos unidos al virus de la superficie celular (Heldt, Kupke, Dorl, Reichl, & Frensing, 2015; Ueda et al., 2013).

El ciclo replicativo del virus de la Influenza se centra en los RNPv, que a su vez dependen de los procesos celulares del huésped para realizar las funciones necesarias para el éxito y finalización de su ciclo replicativo (Eisfeld, Neumann, & Kawaoka, 2015; Gavazzi et al., 2013). Por lo tanto, las personas se contagiarán por medio de contacto o por partículas que quedan suspendidas en el ambiente por estornudos del paciente entre otros (Shah et al., 2019). Es por ello, que los virus de Influenza están en una constante evolución (Sautto & Ross, 2019). La primera solución para prevenir la dispersión de los virus de Influenza es el diagnóstico realizado por el médico, cuando el paciente muestra los síntomas de la infección por virus de Influenza A H1N1, sin embargo, es difícil saber que cepa es la que genera el cuadro clínico considerando únicamente los síntomas del paciente, por ello, que se han desarrollado sistemas de diagnósticos rápidos y precisos para comenzar inmediatamente la terapia viral que será aplicada a la persona infectada (Seki, Oda, & Sugaya, 2020).

Dentro de los diagnósticos rápidos para determinar el tipo de virus de Influenza que genera los cuadros clínicos de gripa se encuentran la toma de radiografías (Schoen, Horvat, Guerreiro, Castro, & Giassi, 2019). También se realizan diagnósticos con el antígeno nucleoproteico viral cuyo resultado se ha estimado en una especificidad del 90 al 95% pero una sensibilidad del 50 al 70% (Datkhaeva, Has, Fitzgerald, & Hughes, 2019; Green & George, 2018). Además se ha implementado el diagnostico con la Reacción en Cadena de la Polimerasa con Transcriptasa Reversa (RT-PCR) en tiempo real, que es más exacto respecto a los diagnósticos anteriormente mencionados ya que por sus características de hibridación entre ácidos nucleicos del virus, se puede diferenciar entre las distintas cepas de Influenza que generan el cuadro clínico (Datkhaeva et al., 2019; Green & George, 2018).

La exactitud del diagnóstico por RT-PCR es debido a que es una prueba exponencial en la cual permite estimar la cantidad de copias del virus en la muestra al medir la cantidad de fluorescencia que se produce durante la polimerización del ARN viral en forma de ADN complementario, a lo que se ha determinado RT-PCR en Tiempo Real (Cobb, Vaks, Do, & Vilchez, 2011). Para lograr lo antes indicado, la reacción utiliza: el segmento de ARN del virus de Influenza A H1N1, el amortiguador, el cofactor de las enzimas, los nucleótidos, los cebadores (oligonucleótidos) complementarios a la región del virus que se amplificará y la molécula que generará la luminosidad, denominada sonda (Shapira et al., 2009; Wu et al., 2010). El proceso es realizado por dos enzimas: la transcriptasa reversa la cual crea una copia de ácido desoxirribonucleico complementario, que se denomina ADN complementario (ADNc) a partir de una molécula de ARN viral, de los 8 segmentos que codifican para las proteínas virales (Boivin, Cusack, Ruigrok, & Hart, 2010). Para terminar con el proceso de amplificación de secuencias virales es necesario utilizar los cebadores que contienen la información de los ácidos nucleicos del virus de la Influenza, los cuales se unen a la cadena complementaria para permitir que la enzima ADN polimerasa termoestable polimerice el fragmento de ADNc seleccionado (Chidlow et al., 2010).

Debido a que la prueba de RT-PCR es exponencial, cada vez que se produce una reacción de polimerización se duplican los segmentos virales y a su vez las moléculas con luminiscencia, permitiendo tener cientos de estas en una prueba, lo cual favorece la detección y el conteo de moléculas virales (Whiley et al., 2009). Otra característica de la RT-PCR es la autenticidad al generar solo copias del virus A H1N1 y no del genoma o ARN del paciente, porque los cebadores de ácidos nucleicos creados para las regiones del virus de Influenza solo se unirán a sus secuencias complementarias del genoma viral por medio de puentes de hidrogeno, además el virus de la Influenza A H1N1 no comparte información genómica con otros seres vivos (Deyde, Sampath, & Gubareva, 2011). Por lo tanto, si la persona se encuentra

infectada con algún virus de Influenza la prueba de RT-PCR en tiempo real mostrará copias virales con luminosidad, por otra parte, la ausencia de luminosidad en la prueba indicará que la persona no tiene el virus y que puede estar enferma debido a otro agente patógeno (Auladell et al., 2019). Cuando los cebadores no se actualizan con regularidad pueden generar diagnósticos falsos negativos, debido a que sus secuencias nucleotídicas pueden tener dos o más nucleótidos diferentes con respecto a los nuevos genomas virales y por tanto no podrán hibridar con los nucleótidos de su segmento complementario (Arevalo-Rodriguez et al., 2020; Cobb et al., 2011; Engelmann, Alidjinou, Lazrek, Dewilde, & Hober, 2017). Esta condición fisicoquímica impedirá la reacción de polimerización, evitando que la molécula luminosa indique la presencia del ARN viral. Estos cambios en el genoma del virus pueden ser producto de mutaciones o de nuevas adaptaciones del virus al huésped (Stellrecht, 2018a, 2018b).

Actualmente se han reportado diversos Kits y publicaciones que muestran diversas sondas y cebadores para el diagnóstico de Influenza A H1N1, sin embargo dentro de los pocos autorizados para el diagnóstico es el de la organización Mundial de la salud que utiliza los genes de HA, NP y M1-M2 para el diagnóstico en pacientes por medio de la RT-PCR cuantitativa de un paso en donde el resultado positivo indica que el virus de Influenza A H1N1 es el que genera la gripa pero cuando el paciente esta clínicamente diagnosticado para gripa y no se genera el diagnóstico se considera un resultado falso negativo (Stellrecht, 2018a). Posterior a la Pandemia de Influenza A H1N1 del año 2009 se demostró que los diagnósticos falsos negativos en pacientes por el sistema antes mencionado se generaron porque los cebadores no podían hibridar con su región complementaria, por lo tanto, perdían su capacidad de reconocer con las nuevas cepas virales (Hoang Vu Mai et al., 2019; Stellrecht, 2018a). Además, los análisis bioinformáticos por el grupo CAGE de 1991 a principio del año 2019 han demostrado que existen nuevos patrones de mutación en los genes de Influenza A H1N1 que generaran diagnósticos incorrectos o que reconocen otros virus de Influenza, dependiendo del

tipo de cebadores y sondas que sean utilizados. Además, se desconoce el nuevo patrón de mutación en los genes de Ha, NP y M1-M2 del año 2019 a Agosto del 2020 y su relación con cebadores y sondas empleadas en el diagnóstico por RT-PCR en tiempo real de un paso. Para dar solución a este problema, se usará la metodología computacional desarrollada por el grupo CAGE que se aplicará en el proyecto autorizado en el Mincienciatón en el área relacionada al diagnóstico de Influenza A H1N1. Con la actualización de los análisis bioinformáticos de los cebadores empleados para el diagnóstico por RT-PCR en tiempo real de un paso se contribuirá a la disminución de los diagnósticos falsos negativos generados por esta metodología.

1. Objetivos

1.1 Objetivo de la pasantía

Actualizar por métodos bioinformáticos la eficiencia de los cebadores diseñados por el proyecto COVID-19 UIS de los genes de Hemaglutinina, Proteína Nuclear, Proteínas de Matriz del virus de la Influenza A H1N1 para el diagnóstico a pacientes por RT-PCR en tiempo real para el año 2019 a agosto del 2020.

1.2 Objetivo general

Determinar la efectividad de los cebadores para los genes de Hemaglutinina, proteína nuclear, proteínas de matriz del virus de la Influenza A H1N1 creados por el proyecto COVID-19 UIS, respecto a las secuencias caracterizadas a partir de pacientes infectados con el virus reportadas en la base de datos pública Influenza Research Database del año 2019 agosto del 2020.

1.3 Objetivos específicos

Diseñar una base de datos con los genes que codifican para Hemaglutinina, proteína nuclear, proteínas de matriz del Virus de la Influenza A H1N1 reportados del año 2019 hasta agosto del 2020 con la base de datos Influenza Research Database.

Establecer *in silico* la eficiencia de los cebadores diseñados para el proyecto COVID-19 UIS respecto a las secuencias consenso de los genes de Hemaglutinina, proteína nuclear,

proteínas de matriz del Virus de la Influenza A H1N1 reportados en la base de datos Influenza Research Database del año 2019 hasta agosto del 2020.

2. Competencias adquiridas

2.1 Competencias cognitivas

Adquiere habilidad de búsqueda bioinformática para una mejor y rápida interpretación de datos nucleotídicos virales.

Analiza secuencias nucleotídicas reportadas para el virus de la Influenza A H1N1 que infectan a seres humanos.

Organiza secuencias en formato tipo FASTA para los genes de hemaglutinina, proteína nuclear, proteínas de matriz del virus de Influenza A H1N1 usados en programas computacionales.

Construye bases de datos con secuencias nucleotídicas para uso en programas de alineamiento múltiple.

Compara secuencias nucleotídicas para validar la efectividad de los cebadores diseñados para proyectos de investigación científica.

Argumenta resultados de análisis bioinformáticos para la selección de la información obtenida de bases de datos.

2.2 Competencias actitudinales

Emprende transformaciones y reformas, para alcanzar logros concretos y a corto plazo relacionados a proyectos científicos.

Adquiere capacidades para desarrollar un aprendizaje autónomo y creativo.

Desarrolla actitud de adaptación para nuevas situaciones con espíritu innovador. Adquiere habilidad de razonamiento crítico y autocrítico para entender la realidad y las ideas que generan los resultados.

Desarrolla aptitud para aplicar el método científico a la resolución de problemas.

3. Metodología

3.1 Identificación de los genes de HA, NP y M1-M2 del virus de la Influenza A H1N1 en la base de datos IRD

La búsqueda avanzada de las secuencias nucleotídicas en la base de datos, IRD se realizó de forma mensual a partir del año 2019 hasta agosto del 2020, mediante palabras claves que fueron consideradas como criterios de selección, siendo estas: tipo de datos, genoma completo, hospedero, tipo de virus: A, segmento del genoma: HA, NP y M1-M2 subtipo: H1N1y la fecha indicada por mes y año. Las secuencias de los genes en estudio se descargaron de la plataforma en formato tipo FASTA mediante la ruta indicada en la base de datos utilizada.

3.2 Alineamiento múltiple y depuración de las secuencias de los genes de HA, NP y M1-M2 del virus de la Influenza A H1N1

La obtención de los alineamientos mensuales de cada segmento se realizó con los recursos computacionales de HPC de SC3UIS con el software MAFFT, al cual, se suministraron los parámetros correspondientes al evolutivo de origen de genes denominado Modelo de Pérdida de ADN: a una penalidad de rompimiento de 9, de extensión de 0,2 y de terminales de 0,4 (Martínez-Pérez et al., 2007; Nakamura, Yamada, Tomii, & Katoh, 2018). El alineamiento resultante se verificó con el software KALING versión 2.0 (Lassmann, Frings, & Sonnhammer, 2009), en la página del Instituto Europeo de Bioinformática (EMBL-EBI) (Madeira et al., 2019), con los parámetros del DNA-LM (Martínez-Pérez et al., 2007). Los resultados finales se almacenaron en archivos tipo FASTA teniendo en cuenta el segmento de

cada gen (S4, S5 y S7) y los criterios de selección registrados en el editor de texto Sublime Text.

Para la depuración de las secuencias con regiones incompletas o nucleótidos no determinados se realizó, con la aplicación prototipo Sequence Manager, que continua en desarrollo por el grupo de investigación del proyecto. Dicha aplicación de forma automática identifica las secuencias con los parámetros antes indicados. Los resultados finales se clasificaron en carpetas y se almacenaron en archivos tipo FASTA para cada segmento de los genes de HA, NP y M1-M2; indicando el mes de cada año y también la cantidad de secuencias encontradas por mes. Los parámetros de captura y uso de la aplicación deben solicitarse por correo electrónico al director del grupo CAGE por continuar como secreto industrial.

3.3 Generación de secuencias nucleotídicas consenso de los genes de HA, NP y M1-M2 del virus de la Influenza A H1N1

Las secuencias precursoras consenso, se obtuvieron con el software BioEdit (Hall, Biosciences, Carlsbad, & Ca, 2013). Se utilizaron los alineamientos mensuales ya establecidos en formato FASTA usando frecuencias de umbral al 20% y 100% para la obtención de los respectivos consensos. En este caso el grupo de investigación CAGE se sustenta en los modelos propuestos por (Day & McMorris, 1992; Kuhner & Yamato, 2017), en donde la frecuencia de umbral minoritaria mostrará las variaciones de más de dos opciones nucleotídicas en una posición,, mientras que el valor máximo corresponde al nucleótido mayoritario de una posición. Posteriormente, se realizó la traducción conceptual de los nucleótidos en la página ExPASY: portal de recursos bioinformáticos de la SIB (Artimo et al., 2012). La información recopilada se almacenó en carpetas indicando cada segmento, mes y año.

3.4 Validación *in silico* de los cebadores creados para los genes de HA, NP y M1-M2 de la Influenza A H1N1 en el GenBank

Los cebadores analizados en esta pasantía fueron diseñados por el grupo de investigación CAGE para el proyecto COVID-19. Es de resaltar que se recibieron dos tipos de diseño para cada gen empleado en el estudio, el primero estuvo conformado por nucleótidos mayoritarios y no presentó alto número de variaciones en cada posición. El segundo correspondió a posiciones degeneradas de nucleótidos y por lo tanto eran hipervariables.

Por lo tanto para su validación *in silico* se empleó el juego de oligonucleótidos mayoritario mediante un alineamiento con la base de datos del National Center for Biotechnology Information (NCBI) (Quiñones et al., 2020) con la aplicación NUCLEOTIDE BLAST (Gotea, Veeramachaneni, & Makaøowski, 2003). El proceso consistió en colocar la secuencia de los cebadores creados para los genes de HA, NP y M1-M2, para ser buscados en la base de datos del NCBI para que generase un resultado de 1000 secuencias nucleotídicas reportadas en esta base de datos. Se crearon archivos tipo FASTA para los resultados indicando el segmento y el sentido de los cebadores, seguidamente del año y del nombre del virus en estudio (H1N1).

La validación de las secuencias consenso obtenidas del año 2019 a agosto del 2020 se realizó con la aplicación prototipo Sequence Manager, del grupo de investigación del proyecto implementada en la supercomputadora GUANE-1.

Para ejecutar el software para la validación de cebadores y sondas, primero se realiza el registro de la secuencia consenso. Inmediatamente se indican las fechas de inicio con el cual se obtuvieron las secuencias para generar la secuencia consenso y la final cuando se generó. Posteriormente, se introduce, una referencia que permita identificar a la secuencia consenso

respecto a otras previamente registradas. Una vez terminado el registro, la aplicación es redirigida a la lista de secuencias genómicas o genes consensos registradas.

Después se genera el registro de cebadores y sondas, en este sentido, en la aplicación prototipo se denominan indistintamente con la palabra "Primer" la cual está precedida por la palabra que indica la dirección del cebador o de la que genera la señal luminosa. Por lo tanto, se tienen tres entradas: (1a) La secuencia nucleotídica del cebador en dirección sentido (5'-3'). (1b) La secuencia nucleotídica de la sonda que genera la señal luminosa en dirección sentido (5'-3'), Quench Primer el cual es opcional y (1c) la secuencia nucleotídica del cebador en dirección antisentido (3'-5'). Es de resaltar que la aplicación genera la posición sentido (5'-3') de forma automática. Adicionalmente, se solicita una referencia bibliográfica correspondiente a la entrada que permita distinguir al cebador que se indica en la entrada y un enlace URL donde se encuentre registrado. Lo anterior puede ser un enlace a un artículo científico o una clave de acceso de la secuencia nucleotídica.

Posteriormente se selecciona una o más secuencias genómicas o genes consensos para que sean relacionados con el cebador que se registró y con ello se logre el análisis Cebador-Consensos. Para facilitar la visualización del resultado, cada cebador es identificado por un color, el cebador sentido en verde, la sonda en amarillo y el cebador antisentido en rojo. Además, se genera una tabla con todas las secuencias consensos relacionados al cebador, se localiza la región nucleotídica con mayor homología, y se indican las mutaciones en cada posición. Además, la aplicación muestra el número total de opciones de secuencias nucleotídicas que se tendrán en función del número de degeneraciones nucleotídicas de los cebadores y sonda. Los resultados se registraron en Word-office versión 2016 señalando en color rojo los códigos de letras que representan a más de un nucleótido.

La determinación de las estructuras secundarias de las sondas, cebadores y oligonucleótidos sustrato se realizó con la aplicación UNAFold Web Server que utiliza los métodos propuestos por (Zuker, 2003).

La validación experimental de los resultados *in silico* de las sondas, cebadores y oligonucleótidos entregados para la realización de esta pasantía se validaron en el Laboratorio Central de Investigaciones de la Facultad de Salud de la Universidad Industrial de Santander, con los protocolos, equipos e insumos adquiridos en el proyecto (Cadena-caballero et al., 2021). Este proceso fue realizado por el Biólogo Cristian Enrique Cadena Caballero.

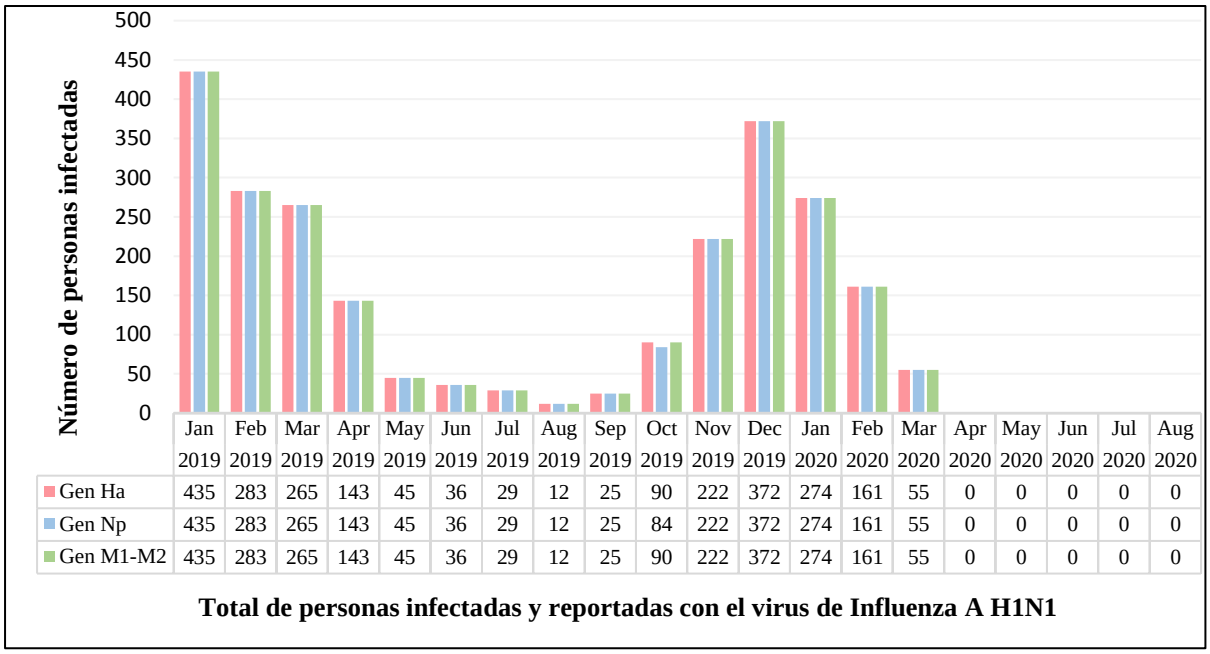
4. Resultados

4.1 Identificación de las secuencias genéticas de los genes HA, NP y M1-M2

La búsqueda en la plataforma IRD con los criterios de selección identificó un total de 7362 secuencias (Apéndice A y B), consideradas por esta base de datos como los reportes de personas infectadas por el virus de Influenza A H1N1 en el año 2019 a marzo del 2020, sin embargo, a partir de abril del 2020 hasta agosto del mismo año no se encontraron resultados de los genes del virus de Influenza A H1N1 publicados en esta base de datos (Apéndice A y B). En el año 2019 se obtuvieron 5872 secuencias, mientras que para el año 2020 se registraron 1490 secuencias. Se encontró que los meses con mayor número de secuencias durante el año 2019 fueron enero, febrero, marzo, noviembre y diciembre, mientras que para el 2020 se reportaron más secuencias en enero y febrero (Fig. 1).

Figura 1

Reporte de personas infectadas con el virus de la Influenza A H1N1



	Jan 2019	Feb 2019	Mar 2019	Apr 2019	May 2019	Jun 2019	Jul 2019	Aug 2019	Sep 2019	Oct 2019	Nov 2019	Dec 2019	Jan 2020	Feb 2020	Mar 2020	Apr 2020	May 2020	Jun 2020	Jul 2020	Aug 2020
Gen Ha	435	283	265	143	45	36	29	12	25	90	222	372	274	161	55	0	0	0	0	0
Gen Np	435	283	265	143	45	36	29	12	25	84	222	372	274	161	55	0	0	0	0	0
Gen M1-M2	435	283	265	143	45	36	29	12	25	90	222	372	274	161	55	0	0	0	0	0

Nota. Cantidad mensual de secuencias reportadas para personas infectadas con el virus de la Influenza A H1N1 en la Plataforma IRD durante el año 2019 a agosto del 2020 para los genes de HA, NP y M1-M2.

La información de las secuencias descargadas de la Plataforma IRD se organizó en carpetas, la primera se nombró Influenza Research Database, dentro de esta se crearon tres carpetas independientes correspondientes al segmento en donde se encuentra cada gen, con los nombres de 4_segment, 5_segment y 7_segment. A su vez, en cada una de estas, se organizaron dos carpetas con los años 2019 y 2020 correspondientes a los registros de secuencias de personas infectadas. Dentro de estas dos carpetas se encuentran tres carpetas más con el nombre de secuencias, alineamientos y consensos, en las mismas los datos en formato FASTA enumerados de 01 a 12, esto para hacer referencia a cada mes del año en estudio (Apéndice C).

4.2 Alineamiento múltiple y depuración de los genes de HA, NP y M1-M2

De las 7362 secuencias reportadas para los genes de esta pasantía, se obtuvieron un total de 7335 secuencias depuradas. Las secuencias de los genes de HA tuvieron en promedio 1725 nucleótidos, seguidamente las secuencias del gen de NP 1518 nucleótidos en cada secuencia, mientras que para el gen de M1-M2, fue el menor reporte de nucleótidos con 996 en cada secuencia (Apéndice D).

Los alineamientos de los genes de HA, NP y M1-M2, mostraron el mismo codón de inicio de la traducción. A saber, ATG que codifica para la Metionina, en cambio difieren en el codón de paro. La HA y M1-M2 presentaron el mismo codón de terminación TAA, por el contrario, el gen de la NP mostró el codón de paro TGA (Apéndice E y F).

4.3 Generación de secuencias nucleotídicas consenso de los genes de HA, NP y M1-M2 del virus de la Influenza A H1N1

Las secuencias consenso que se obtuvieron para cada mes con un umbral del 20% no mostraron cambios significativos en las posiciones de los nucleótidos, ya que se conservó la presencia de las purinas y pirimidinas sin que varias combinaciones de estas arrojaran la traducción del mismo aminoácido, mientras que los resultados que se obtuvieron con un umbral del 100% mostraron cambios en varias de las posiciones comprendidas entre el codón de inicio ATG y el codón de paro (Apéndice G). De manera general los genes de HA, NP y M1-M2 presentaron cambios a nivel de nucleótidos con umbrales del 100%, sin embargo, se encontró que el gen de la HA con un umbral del 20% tuvo variación en la formación de codones donde intervienen dos o tres formas nucleotídicas para un mismo aminoácido. Los cambios en

porcentajes para el gen de la HA que se presentaron durante el mes de Julio variaron en las purinas y pirimidinas. La variación más alta fue de 2,80% en los codones que tienen Adenina y Guanina y 2,60% en los codones que tienen Timina y Citosina. Además, se encontraron otras variaciones porcentuales para el mes de julio del 2019, menores al 0,70% correspondientes a las letras del código genético degenerado M, W, V, D para los codones que codifican para la proteína estructural del virus de la Influenza A H1N1 (Apéndice G y Tabla 1).

El uso de umbrales de incertidumbre del 20% para los genes de HA, NP y M1-M2 arrojó mayor presencia de Adenina y Timina con respecto a los nucleótidos Guanina y Citosina. No obstante, el gen de la HA tuvo menor presencia de Purinas y Pirimidinas durante el mes de Julio del 2019, con un porcentaje de 55,7% y 37,30% de Adenina con Timina y de Guanina con Citosina respectivamente (Tabla 1). Los resultados de las secuencias consenso que se obtuvieron con un umbral de incertidumbre del 100% mostraron aproximadamente una presencia en los codones de 47% al 59% de Adenina y Timina, mientras que para la Guanina y Citosina de un 27% al 40% para el gen que codifica para la Ha. Las secuencias consenso de la NP variaron del 43% al 53% de Adenina y Timina, con respecto a la presencia de un 34% a 43% de pares de bases nucleotídicas Guanina y Citosina. Por otra parte, para M1-M2 se encontró un porcentaje de secuencias que van de 44% al 51% para la adenina y timina, y un porcentaje que varía para las guaninas y citosinas del 37% al 43% para las secuencias en estudio de los pacientes registrados con el virus de la Influenza A H1N1 durante el año 2019 a marzo del 2020 (Tabla 2).

Tabla 1

Porcentajes de purinas, pirimidinas y nucleótidos degenerados del gen HA del virus de la Influenza A H1N1 con un umbral de incertidumbre del 20%

Año	Mes	AT	GC	A	T	C	G	N=A,C,G,T	B=C,G,T	D=A,G,T	H=A,C,T	V=A,C,G	K=G,T	M=A,C	R=A,G	S=C,G	W=A,T	Y=C,T	-	Total
2019	**Enero**	59,10%	40,90%	35,40%	23,70%	18,80%	22,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	99,90%
2019	**Febrero**	59,00%	41,00%	35,40%	23,60%	18,90%	22,10%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%
2019	**Marzo**	59,00%	41,00%	35,40%	23,60%	18,90%	22,10%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%
2019	**Abril**	59,10%	40,90%	35,40%	23,70%	18,90%	22,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%
2019	**Mayo**	59,10%	40,90%	35,40%	23,70%	18,90%	22,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%
2019	**Junio**	59,10%	40,90%	35,40%	23,70%	18,80%	22,10%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%
2019	**Julio**	55,70%	37,30%	33,60%	22,10%	17,20%	20,20%	0,00%	0,00%	0,10%	0,00%	0,10%	0,10%	0,70%	2,80%	0,00%	0,50%	2,60%	0,00%	100,00%
2019	**Agosto**	58,90%	41,10%	35,40%	23,50%	19,00%	22,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	99,90%
2019	**Septiembre**	59,10%	40,90%	35,40%	23,70%	18,80%	22,10%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%
2019	**Octubre**	59,10%	40,90%	35,40%	23,70%	18,80%	22,10%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%
2019	**Noviembre**	59,10%	40,90%	35,40%	23,70%	18,80%	22,10%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%
2019	**Diciembre**	59,10%	40,90%	35,30%	23,80%	18,80%	22,20%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,10%
2020	**Enero**	59,10%	40,90%	35,30%	23,80%	18,80%	22,20%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,10%
2020	**Febrero**	59,00%	41,00%	35,30%	23,70%	18,90%	22,20%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,10%
2020	**Marzo**	59,00%	41,00%	35,30%	23,70%	18,90%	22,20%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,10%

Nota. La abreviación de los nombres de los nucleótidos que forman los aminoácidos (A=Adenina, T=Timina, G=Guanina, C=Citosina). Las demás letras (N, B, D, H, V, K, M, R, S, W, Y) Corresponde a las letras del código degenerado, el cual hace referencia al posible cambio mutacional en la formación de codones para formar el mismo aminoácido.

Tabla 2

Porcentaje de purinas y pirimidinas con un umbral del 100 de incertidumbre para el gen de la HA, NP y M1-M2

Año	Mes	HA		NP		M1-M2	
		AT	**GC**	**AT**	**GC**	**AT**	**GC**
2019	**Enero**	46,10%	27,40%	43,80%	34,70%	44.3%	37.8%
2019	**Febrero**	46,90%	28,70%	44,70%	35,90%	45.2%	39.3%
2019	**Marzo**	47,20%	29,30%	46,20%	36,20%	44.7%	40.8%
2019	**Abril**	51,30%	32,20%	48,80%	39,60%	47.3%	43.1%
2019	**Mayo**	57,10%	38,40%	52,50%	43,10%	50.5%	46.0%
2019	**Junio**	56,60%	38,40%	53,30%	43,40%	51.0%	46.1%
2019	**Julio**	55,70%	37,30%	52,80%	42,80%	50.7%	45.5%
2019	**Agosto**	57,30%	38,90%	53,40%	43,50%	50.9%	46.6%
2019	**Septiembre**	59,10%	40,90%	53,20%	42,70%	50.1%	46.0%
2019	**Octubre**	55,40%	36,20%	52,50%	41,40%	49.4%	45.3%
2019	**Noviembre**	54,20%	34,40%	50,80%	40,90%	48.6%	44.1%
2019	**Diciembre**	50,80%	33,10%	51,00%	39,70%	48.5%	43.0%
2020	**Enero**	50,60%	31,90%	48,50%	39,10%	46.6%	41.6%
2020	**Febrero**	51,60%	32,80%	50,30%	39,40%	48.1%	42.7%
2020	**Marzo**	55,60%	37,20%	53,80%	42,50%	50.9%	44.9%

Nota. AT= (Adenina, Timina), GC= (Guanina, Citosina). HA, NP y M1M2 hace referencia al nombre de los tres genes utilizados para el estudio de la Influenza A H1N1.

Teniendo en cuenta el código genético degenerado se encontró mayor número de cambios en las purinas “R” y en las pirimidinas “Y”. Por lo anterior, dentro de los consensos se encontró que la mayor variación fue para el año 2019 durante el mes de enero en las posiciones para los genes: HA, R=11.90-Y=9.20%, NP R=10.20-Y=7.00% y M1-M2 R=9.40-Y=6.10% (Apéndice G).

Validación de los cebadores propuestos para los genes de HA, NP y M1-M2

Con las secuencias nucleotídicas de los cebadores, sondas y oligonucleótidos diseñados por el grupo de investigación del proyecto, se determinó que para la RT-PCR en Tiempo Real de un Paso hay tres tipos de cebadores y sonda para cada gen, el primero contiene nucleótidos

degenerados para cada codón, el segundo se diseñó con nucleótidos mayoritarios y el tercer grupo correspondió a cebadores para amplificar el gen completo para su secuenciación (Tablas 3, 4 y 5). Es de resaltar que dentro de las moléculas para la RT-PCR de la región del gen empleada para el diagnóstico, el grupo de investigación incluyó, un oligonucleótido que corresponde al sustrato de la reacción con nucleótidos degenerados, el objetivo de este fue fungir como control de la reacción y parámetro de concentración molar en la reacción (Tabla 3). Las sondas tuvieron en el extremo 5’ el fluoróforo 5-carboxifluoreseina para generar la señal luminosa requerida en la cuantificación de la polimerización y en el extremo 3’ un extintor de oscuridad denominado en inglés *Quencher* tipo cereza negra BBQ para absorber la señal luminosa del fluoróforo (Tabla 4). En tanto, que la secuencia de los oligonucleótidos para la amplificación del genoma completo, mostraron únicamente un diseño de nucleótidos degenerados por la alta tasa mutacional presente en la región utilizada del virus (Tabla 5).

Tabla 3

Secuencias de cebadores y oligonucleótidos sustrato

No.	Nombre	pb	Secuencia
0	FwBS420H1N1	20	5'GYAACATTGCTGGCTGGATC 3'
1	FwDS420H1N1	20	5'GYRMYRYYGCTGGMTGGATY 3'
2	RvBS420H1N1	20	5' TCTCCTGGGTAACAYGTTCC 3'
3	RvDS420H1N1	20	5' TCYCCWGGRTAACAHGYYCC 3'
4	TrGtS4H1N1	118	5'GYRMYRYYGCTGGMTGGATYCTGGGAAATCCAGAGTGTGAATCACTCTCYAMAGYAAGRTCATGGYCYTACATTGTDGARAACRTCTAATTCAGACAAGGRRCDTGTTAYCCWGGRGA 3'
5	FwBS520H1N1	20	5'CAATGGCGAAGATGCAACAG 3'
6	FwDS520H1N1	20	5'YRAYGGHGMAGAKDCAACWG 3'
7	RvBS520H1N1	21	5'CATTAGAGAGCACATTCTGGG 3'
8	RvDS520H1N1	21	5'CATTARAGARCACATYCTRGG 3'
9	TrGtS5H1N1	130	5'YRAYGGHGMAGAKDCAACWGCAGGTCTTACTCATATCATGATTTGGCATTCMAAYYTRAATGATGCCACRTAYCARAGRACAAGAGCACTTGTTCGCACTGGAATGGATCCYAGRATGTGYTCTYTAATG 3'
10	FwBS7120H1N1	22	5"CAGTGCTGGTCTRAGAGATGAC 3'
11	FwDS7120H1N1	22	5'CAGTRCHGGTYTRARRGATGAY 3'
12	RvBS7120H1N1	20	5'GGTGCAAGATCCCAATGATG 3'
13	TrGtS71H1N1	126	5'CAGTRCHGGTYTRARRGATGAYCTTCTTGAAAATTTGCAGGCCTACCAGAAGCRRATGGGAGTGCARRTGCRRCGDTTCAARTRAYMYYMTHRYCDYYGCRGCAAAYATCATTGGGATYYTGCAC 3'

14	FwBS7220H1N1	20	5'CTAACCGAGGTYGAAACGCH 3'
15	TrGtS72H1N1	103	5'CTAACCGAGGTYGAAACGCHTACCAGAAGCRRATGGGAGTGCARRTGCRRCGDTTCAARTRAYMYYMTHRYCDYYGCRGCAAAYATCATTGGRATHYTGCACY 3'

Tabla 4

Sondas Taqman para RT-PCR en Tiempo Real de un Paso

No.	Nombre	pb	Secuencia
1	TqMBS420H1N1	34	5" FAM-CTCCACAGCAAGATCATGGTCCTACATTGTGGAA-BBQ 3'
2	TqMDS420H1N1	34	5' FAM-MTCYAMMGMARGRTCATGGYCHTACATTGTDGAA-BBQ 3'
3	TqMBS520H1N1	31	5' FAM-TCCAATCTGAATGATGCCACATATCAGAGGA-BBQ 3'
4	TqMDS520H1N1	31	5 'FAM-TCMAAYYTRAATGATGCCACRTAYCARAGRA-BBQ 3'
5	TqMBS7120H1N1	39	5" FAM-TGCAGATGCAGCGGTTCAARTGATCCTCTCGTCATTGCA-BBQ 3'
6	TqMDS7120H1N1	39	5' FAM-TGCARRTGCRRCGDTTCAARTRAYMYYMTHRYCDYYGCR-BBQ 3'

Tabla 1

Secuencia de cebadores para secuenciación genómica

No.	Nombre	pb	Secuencia
1	FWS4GNO H1N1	26	5'CGGGATGAADRYAATRCTRGTAGTYC 3'
2	RVS4GNO H1N1	27	5'AGGCCTAATGYYARAWRCAWAYYCTAC 3'
3	FWS5GNO H1N1	23	5'ATGGYRTHYCAAGGYACCAAACG 3'
4	RVS5GNO H1N1	24	5'TCATTSTMRTACTCCTCWGCAYTG 3'
5	FWS7GNO H1N1	21	5'GATGAGYCTTCTAACCGAGGTY 3'
6	RVS7GNO H1N1	27	5'TTTYTAYTYTAGCTCTATGTTGACAAA 3'

Los alineamientos tipo BLAST con los cebadores y sondas diseñados con los nucleótidos mayoritarios mostraron que todos estos reconocen sus respectivas secuencias, preferentemente para los años 2019 y 2020 de cepas aisladas de pacientes. Sin embargo, las moléculas para el gen HA también mostraron homología con los virus de Influenza H3N2 que infecta a humanos, con Influenza A H1N1 y H1N2 exclusivo para cerdos. Mientras que los cebadores y sonda para el gen NP se encontró similitud con las cepas de H1N2 y H3N2 que infectan a cerdos. En tanto, que las moléculas diseñadas para el gen M1 mostraron similitud a los virus de Influenza A H1N1 y H1N2 de cerdos (Apéndice H).

Posteriormente, los patrones de similitud fueron congruentes al identificar los cebadores y sonda de los genes HA y NP con la aplicación prototipo Sequence Manager del grupo de investigación instalada en la supercomputadora GUANE-1, cuyo resultado mostró que las sondas y cebadores de cada gen fueron semejantes a su respectiva secuencia en los consensos con nucleótidos mayoritarios, es decir, consensos con un umbral de frecuencia de 20%, pero se observó que el reconocimiento en las secuencias consenso con un umbral de 100% presentó menor similitud y estuvo en función de la secuencia consenso mensual como se observa en la Tabla 6 y Apéndice I. El resultado fue similar para el gen M1-M2. En este sentido las mayores variaciones se determinaron en los alineamientos con las secuencias consenso con un umbral de frecuencia de 100% se determinó en los meses del año 2019 respecto a 2020 que fue minoritaria. En tanto, que los consensos mensuales para el virus de Influenza A H1N1 con los umbrales de frecuencia de 20% la similitud de las secuencias para los tres genes fue del 100% (Apéndice J).

Las predicciones de las estructuras secundarias con base a los algoritmos propuestos por Zuker de cada una de las moléculas indicadas en las Tablas 3, 4 y 5 mostraron que todas formaron una estructura de tallo burbuja la cual correspondió a oligonucleótidos tipo escorpión sea bien en el extremo 5’ o 3’ (Fig. 2) (Zuker, 2003). Los oligonucleótidos que se emplearon para la amplificación del Marco Abierto de Lectura (ORF) de los tres genes también mostraron los respectivos tallos de los cuales los del gen M1-M2, fueron de un menor tamaño respecto a los genes Ha y Np (Fig. 3). Finalmente, las características termodinámicas de todos los cebadores que estuvieron en función de la ADN polimerasa empleada en los protocolos de RT-PCR de la Organización Mundial de la Salud para la identificación del virus de Influenza A H1N1, mostró que a una concentración de 1 mM Mg las sondas con nucleótidos mayoritarios tienen en promedio una Tm de 60°C mientras que los equivalentes con nucleótidos degenerados la Tm tuvo en promedio de 55°C. Este patrón se determinó también para las sondas donde la

Tm de las moléculas tuvo rangos de 67 a 74°C mientras que las sondas con nucleótidos degenerados el rango osciló de 64 a 67°C (Tabla 7).

Tabla 2

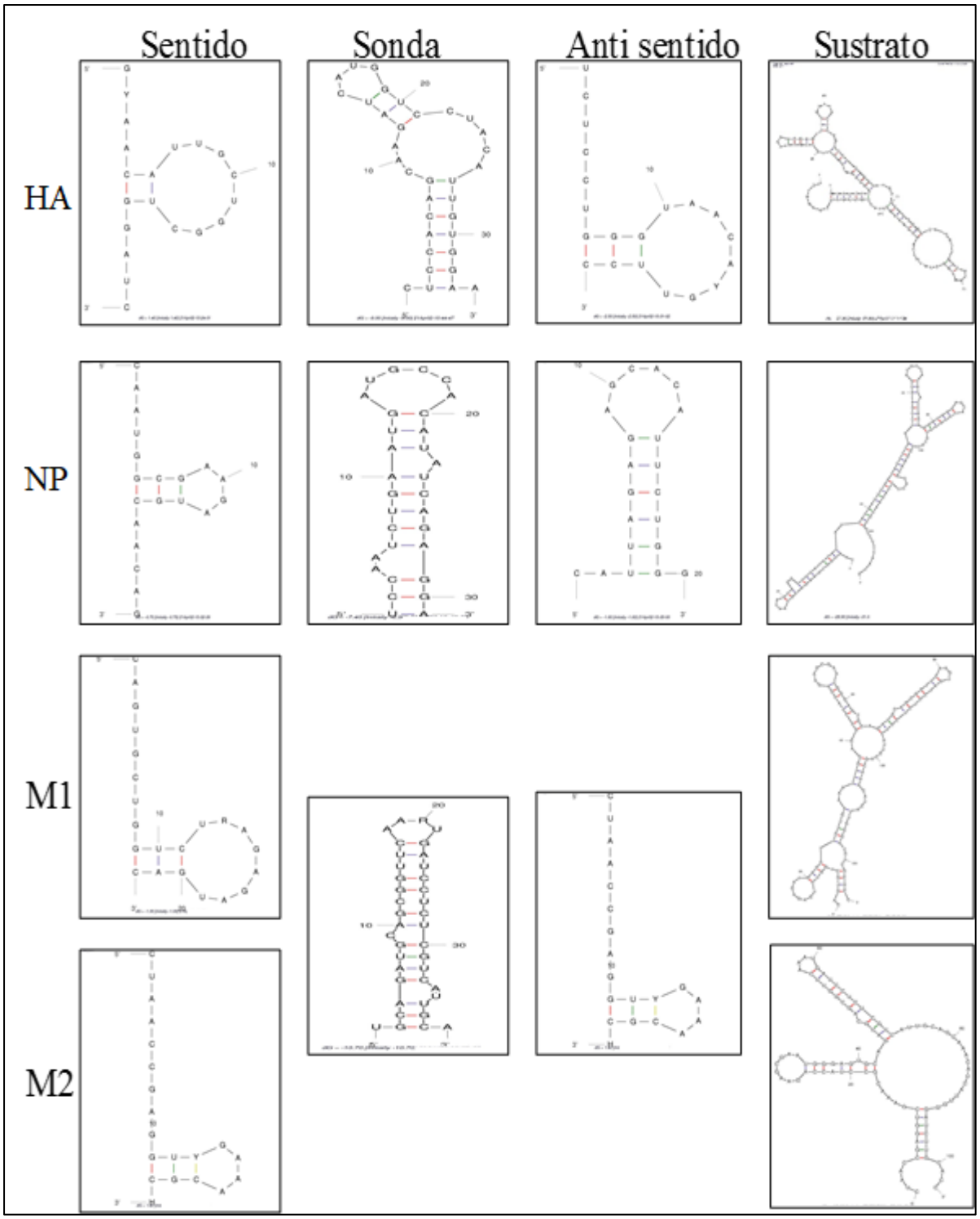
Evaluación de los cebadores con la aplicación prototipo Sequence Manager en la supercomputadora GUANE-1

Frequency threshold 20%		Forward Primer	Probe	Reverse Primer	RT-PCR
Mar/2020.55	S4	GTAACATTGCTGGCTGGATC	TCCACAGCAAGATCATGGTCCTACATTGTGGAA	GGAACGTGTTACCCAGGAGA	(1*1*1) = 1
Feb/2020.161	S4	GTAACATTGCTGGCTGGATC	TCCACAGCAAGATCATGGTCCTACATTGTGGAA	GGAACGTGTTACCCAGGAGA	(1*1*1) = 1
Jan/2020.274	S4	GTAACATTGCTGGCTGGATC	TCCACAGCAAGATCATGGTCCTACATTGTGGAA	GGAACGTGTTACCCAGGAGA	(1*1*1) = 1
Dec/2019.372	S4	GTAACATTGCTGGCTGGATC	TCCACAGCAAGATCATGGTCCTACATTGTGGAA	GGAACGTGTTACCCAGGAGA	(1*1*1) = 1
Nov/2019.222	S4	GTAACATTGCTGGCTGGATC	TCCACAGCAAGATCATGGTCCTACATTGTGGAA	GGAACGTGTTACCCAGGAGA	(1*1*1) = 1
Oct/2019.90	S4	GTAACATTGCTGGCTGGATC	TCCACAGCAAGATCATGGTCCTACATTGTGGAA	GGAACGTGTTACCCAGGAGA	(1*1*1) = 1
Sep/2019.25	S4	GTAACATTGCTGGCTGGATC	TCCACAGCAAGATCATGGTCCTACATTGTGGAA	GGAACGTGTTACCCAGGAGA	(1*1*1) = 1
Aug/2019.12	S4	GTAACATTGCTGGCTGGATC	TCCACAGCAAGATCATGGTCCTACATTGTGGAA	GGAACGTGTTACCCAGGAGA	(1*1*1) = 1
Jul/2019.29	S4	GYAACATTGCTGGCTGGATC	TCCACAGCAAGATCATGGYCCTACATTGTGGAA	GGAACDTGTTACCCAGGAGA	(2*2*3) = 12
Jun/2019.36	S4	GTAACATTGCTGGCTGGATC	TCCACAGCAAGATCATGGTCCTACATTGTGGAA	GGAACGTGTTACCCAGGAGA	(1*1*1) = 1
May/2019.45	S4	GTAACATTGCTGGCTGGATC	TCCACAGCAAGATCATGGTCCTACATTGTGGAA	GGAACGTGTTACCCAGGAGA	(1*1*1) = 1
Apr/2019.143	S4	GTAACATTGCTGGCTGGATC	TCCACAGCAAGATCATGGTCCTACATTGTGGAA	GGAACGTGTTACCCAGGAGA	(1*1*1) = 1
Mar/2019.265	S4	GTAACATTGCTGGCTGGATC	TCCACAGCAAGATCATGGTCCTACATTGTGGAA	GGAACGTGTTACCCAGGAGA	(1*1*1) = 1
Feb/2019.283	S4	GTAACATTGCTGGCTGGATC	TCCACAGCAAGATCATGGTCCTACATTGTGGAA	GGAACGTGTTACCCAGGAGA	(1*1*1) = 1
Jan/2019.35	S4	GTAACATTGCTGGCTGGATC	TCCACAGCAAGATCATGGTCCTACATTGTGGAA	GGAACGTGTTACCCAGGAGA	(1*1*1) = 1
Consensus		**GYAACATTGCTGGCTGGATC**	**TCCACAGCAAGATCATGGYCCTACATTGTGGAA**	**GGAACDTGTTACCCAGGAGA**	
Frequency threshold 100%					
Mar/2020.55	S4	GTAACATTGCMGGCTGGATY	TCCACAGCAARRTCATGGTCMTACATTGTGGAR	GGMACGTGTTACCCAGGAGA	(2*16*2) = 64
Feb/2020.161	S4	GYAACATTGCTGGCTGGAYC	TCCACAGCAAGDTCATGGTCYTACATWGTGGAA	GGAACRTGTTACCCAGGARA	(4*12*4) = 192
Jan/2020.274	S4	GYAACATTGCTGGCTGGATC	TCCRCAGMAARRTCATGGYCYTACATTGTKGAA	GGAACRTGYTACCCAGGAGA	(4*128*4) = 2.048
Dec/2019.372	S4	GYAAYATTGCTGGCTGGATY	TCCACAGMAAGRTCATGGTCCTACATTGTGGAA	GGAACRTGTTACCCAGGRGA	(4*4*4) = 64
Nov/2019.222	S4	GYAACAYTGCTGGMTGGATY	TCCACAGMAAGRTCATGGTCCTACATTGTRGAA	GGAACRTGTTACCCAGGRGA	(16*8*4) = 512
Oct/2019.90	S4	GYAACATTGCTGGCTGGATC	TCCACAGHAAGRTCATGGTCCTACATTGTRGAA	GGAACRTGTTACCCAGGAGA	(4*12*2) = 96
Sep/2019.25	S4	GTAACATTGCTGGCTGGATC	TCCACAGCAAGATCATGGTCCTACATTGTGGAA	GGAACGTGTTACCCAGGAGA	(1*1*1) = 1
Aug/2019.12	S4	GYAACATTGCTGGCTGGATC	TCCACAGCRAGATCATGGTCCTACATTGTGGAA	GGAACRTGTTACCCAGGAGA	(2*2*2) = 8
Jul/2019.29	S4	GYAACATTGCTGGCTGGATC	TCCACAGCAAGATCATGGYCCTACATTGTGGAA	GGAACDTGTTACCCAGGAGA	(2*2*3) = 12
Jun/2019.36	S4	GTAACATTGCTGGCTGGATC	TCCACAGCAAGATCATGGTCMTACATTGTGGAA	GGAACGTGTTACCCAGGAGA	(1*2*1) = 2
May/2019.45	S4	GTAACATYGCTGGCTGGATY	TCCACAGCAAGATCATGGTCCTACATTGTGGAA	GGAACGTGTTACCCAGGAGA	(4*1*1) = 4
Apr/2019.143	S4	GYAACRYYGCTGGMTGGATH	TCCRCAGCAARATCATGGYCYTACATTGTRGAA	GGAACRTGTTACCCAGGAGA	(96*32*2) = 6.144
Mar/2019.265	S4	GYRACAYYGCWGGCTGGATC	TCCMMAGCAARRTCATGGYCYTACATTGTDGAR	GGRRCDTGTTAMCCWGGRGA	(32*384*96) = 1.179.648
Feb/2019.283	S4	GYRAYRYYGCTGGMTGGATY	TCYACAGYAAGRTCATGGYCYTACATTGTDGAA	GGRACRTGTTACCCAGGAGA	(256*96*4) = 98.304
Jan/2019.35	S4	GYRMCATTGCTGGCTGGATC	TCYAMAGCAAGATCATGGYCYTACATTGTDGAA	GGRACRTGTTAMCCWGGAGA	(8*48*16) = 6.144
Consensus		**GYRMYRYYGCTGGMTGGATY**	**TCYAMMGMARGRTCATGGYCHTACATTGTDGAA**	**GGRRCDTGTTAYCCWGGRGA**	

Nota. Se representa a manera de ejemplo, el umbral de frecuencia del 20% y 100% para las secuencias consenso del gen de la HA utilizadas en este estudio desde Enero del 2019 hasta Marzo del 2020. Para cada una de las secuencias se utilizó la fecha mensual de caracterización seguida del año, la cantidad de secuencias utilizadas para realizar el consenso mensual individual y el tipo de segmento empleado (S4). Se muestra en verde los cebadores Forward, en amarillo las sondas y en azul los cebadores Reverse. En cajas rojas y con el código degenerado se representan los cambios mutacionales detectados mensualmente para cada uno de los cebadores y sonda. El consenso mayoritario y degenerado se localiza en negritas al final de cada alineamiento. Por último, se realiza una sumatoria de los cambios mutacionales observados y de la variación de estos para realizar la RT-PCR. Para ver más detalles Apéndice J.

Figura 1

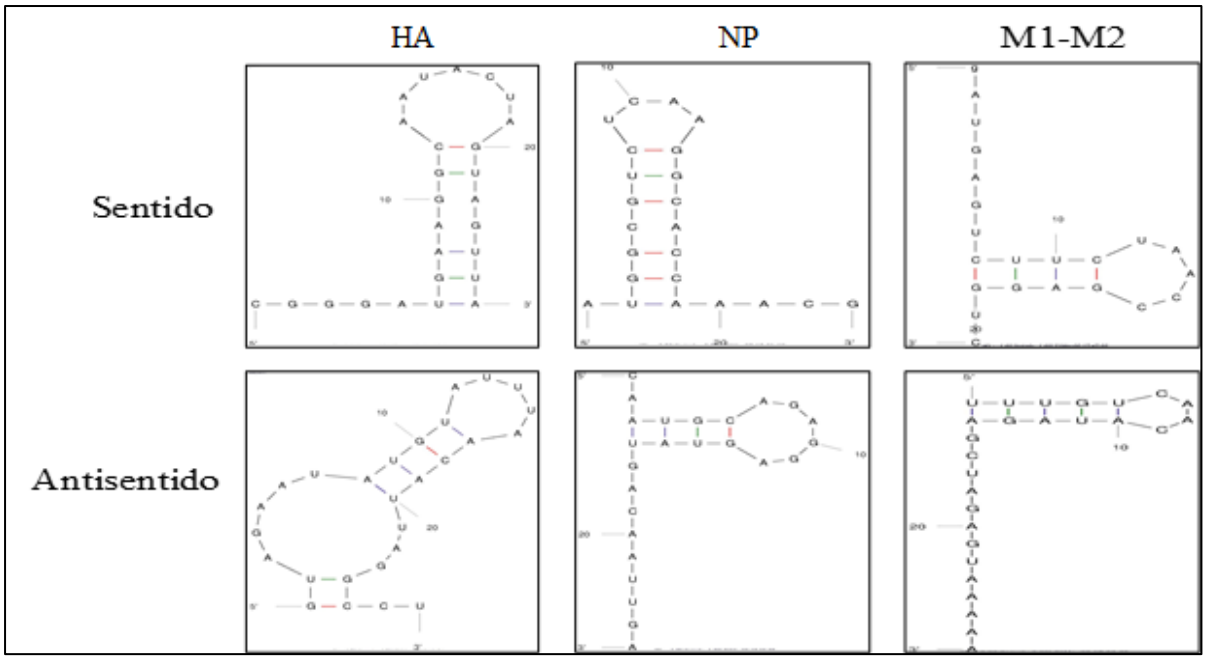
Estructuras secundarias de las sondas y cebadores para RT-PCR



Nota. Todas las moléculas forman tallos burbuja lo que sugiere que la hibridación con su región complementaria no será al 100%.

Figura 2

Estructuras secundarias para oligonucleótidos de RT-PCR



Nota. La estructura secundaria de todas las moléculas formó de manera natural un tallo burbuja para los tres genes.

Tabla 3

Temperaturas de los cebadores, sondas y oligonucleótidos

Gen	Nombre	Nano/Mol RT-PCR	Pb	Tm °C 1 mM/Mg	Tm °C Alin
HA-S4	FwBS420H1N1	400	20	60,2	58
	RvBS420H1N1	400	20	60,3	58
	TqBMS420H1N1	200	34	70,5	58
	FwDS420H1N1	400	20	55	56
	RvDS420H1N1	400	20	58	56
	TqDMS420H1N1	200	34	64,6	56
Np-S5	FwBS520H1N1	400	20	61	59
	RvBS520H1N1	400	21	59,5	59
	TqMBS520H1N1	200	31	67,9	59
	FwDS520H1N1	400	20	54	59
	RvDS520H1N1	400	21	54	59
	TqMDS520H1N1	200	31	63	59
M1-S7	FwBS7120H1N1	400	22	61	59
	RvBS7120H1N1	400	20	60	59
	TqMBS7120H1N1	200	39	74,6	59
	FwDS7120H1N1	400	22	56,1	55
	RvDS7120H1N1	400	20	54,6	55
	TqMBS7120H1N1	200	39	67	55

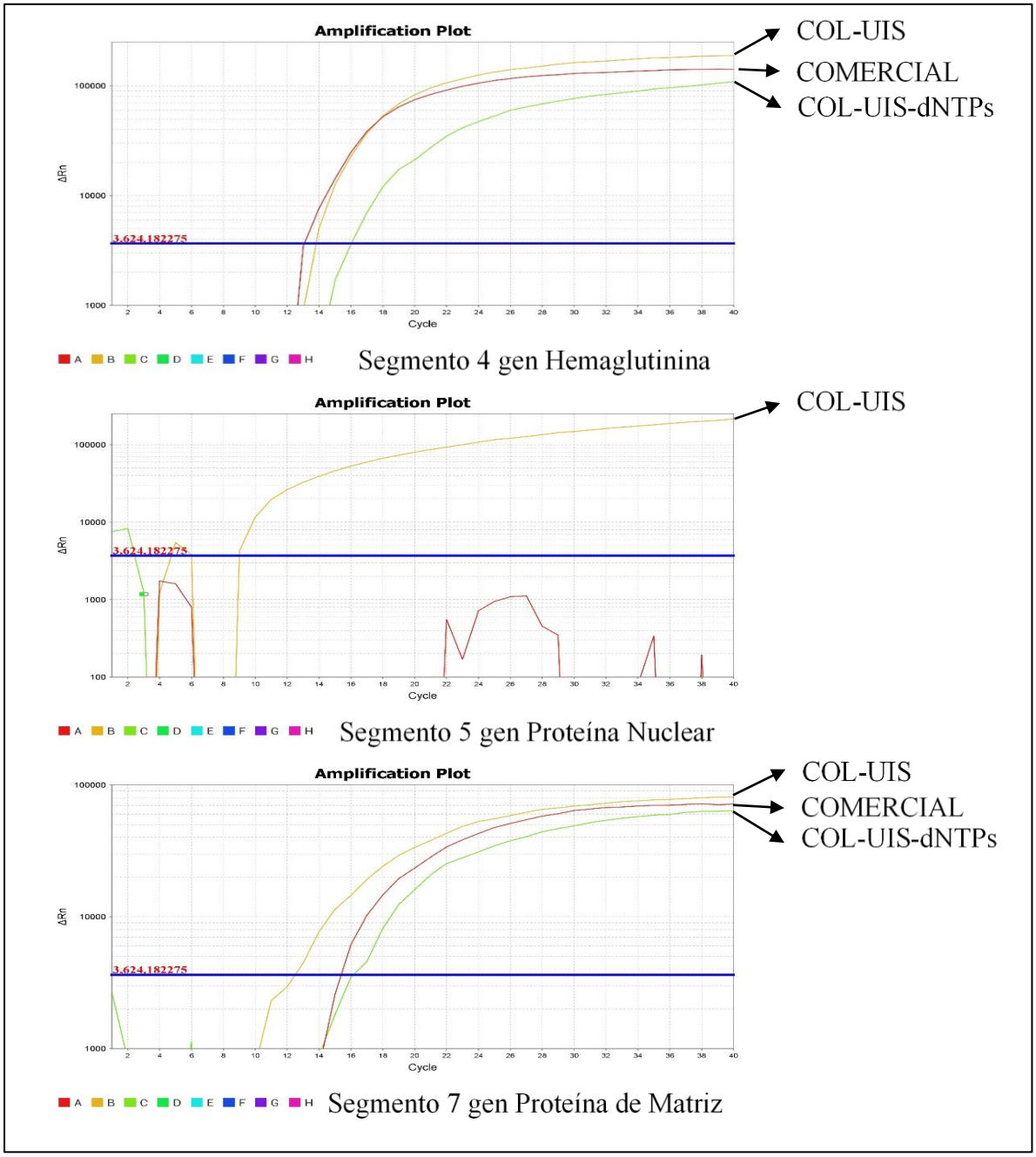
	FwBS7220H1N1	400	20	61,5	59
M2-S7	RvDS7120H1N1	400	20	54,6	55
	TqMBS7120H1N1	200	39	67	55

La validación experimental de estos diseños con el oligonucleótido sustrato, cebadores y sondas en una reacción de RT-PCR en Tiempo Real de un paso se realizó por el profesional Biol. Cristian Enrique Cadena Caballero responsable técnico del proyecto en el Laboratorio Central de Investigaciones de la Facultad de Salud de la Universidad Industrial de Santander. En la reacción se incluyeron las soluciones químicas diseñadas por los investigadores del proyecto con el apoyo del Ministerio de Ciencia Tecnología e Innovación llamadas Colombia-UIS (Col-UIS) y Colombia-UIS-dNTPs (Col-UIS-dNTPs) y se compararon con el kit comercial oficial empleado para el diagnóstico de Influenza A H1N1 por la OMS. Estas soluciones se emplearon para desenrollar el ARN viral y para mejorar la capacidad de polimerización en la reacción, fueron validadas también con el virus de SARS-CoV-2 (Cadena-caballero et al., 2021).

Por lo anterior, en la figura 4 se observan tres reacciones de RT-PCR en Tiempo Real de un Paso para los cebadores, sonda y oligonucleótidos sustrato de los genes HA, NP y M1-M2, en las cuales se destaca la amplificación potencializada de Col-UIS en las tres reacciones respecto a los demás tratamientos. En tanto, los tratamientos comercial y Col-UIS-dNTPs sólo presentaron amplificación para los genes HA y M1-M2. Es de resaltar que para el segmento 5 solo se presentó amplificación con el tratamiento Col-UIS, validando la tecnología desarrollada en el proyecto para genes y genomas virales ricos en regiones Adenina-Timina y al cual hace parte esta pasantía.

Figura 3

RT-PCR de los genes del virus de Influenza A H1N1 con las soluciones COL-UIS (desenrolladora) y COL-UIS-dNTPs (nucleótidos)



Nota. De la parte superior a la parte inferior se encuentran el segmento 4 para el gen HA, el segmento 5 para el gen NP y el segmento 7 para los genes M1 y M2. En amarillo se muestra el resultado con la solución COL-UIS, en rojo el tratamiento comercial y en verde COL-UIS-

dNTPs para cada gráfica. Es de resalta que en el segmento 5 solo se obtuvo amplificación con COL-UIS validando la tecnología para genomas y genes virales ricos en AT.

5. Discusión

El reordenamiento genético del virus de la Influenza A en las poblaciones de aves y de cerdos puede producir nuevos linajes del virus y por ende nuevas pandemias (Pabbaraju et al., 2009). Estudios revelan que la acumulación de mutaciones de aminoácidos en las glicoproteínas del virus de Influenza A H1N1 continúa disminuyendo la efectividad de los diagnósticos moleculares (Opanda, Bulimo, Gachara, Ekuttan, & Amukoye, 2020). El uso de RT-PCR y pruebas químicas en los diagnósticos del virus de la Influenza A, han sido constantes para vigilar la resistencia del virus a los antivirales (Bragstad et al., 2019). No obstante, la postulación de nuevos kits de diagnóstico no es con tanta frecuencia respecto a las nuevas vacunas debido a que se requiere la secuenciación genómica o del gen del virus de la influenza A H1N1y su publicación en base de datos públicas para diseñar y validar las eficiencias de kits de diagnóstico.

Por lo anterior, en esta pasantía se contribuyó a la elaboración de nuevos cebadores, sonda y oligonucleótidos para validar su eficiencia por métodos bioinformáticos y experimentales, que podrán ser aplicados en los diagnósticos de la Medicina actual. En este sentido, este estudio contribuyó al proyecto liderado por Minciencias en el Marco de la Mincienciatón, en donde se ha utilizado la información disponible en la plataforma de IRD de secuencias de personas infectadas con el virus de Influenza A en todo el mundo, con la finalidad de diseñar cebadores para RT-PCR con fines de reducir los falsos negativos y desde la

bioinformática con el apoyo de la supercomputadora GUANE-1 se pudo establecer resultados positivos en el diagnóstico.

El virus de la Influenza A H1N1 continúa presentando cambios genéticos de resistencia a la inmunidad adquirida, por lo tanto puede causar epidemias anuales por lo que continua siendo de interés público para las poblaciones susceptibles a contagios (Gatherer, 2009). Una de las más fuertes en los Estados Unidos fue la soportada durante los años de 2015-2016 donde se empezó a reportar picos altos a partir de octubre del 2015 y a finales de mayo del 2016, lo cual está soportado geográfica y temporalmente. Durante este tiempo el subtipo de Influenza A que más se reportó fue H1N1 seguidamente del subtipo H3N2.

Estos cambios evolutivos en la organización nucleotídica de las cepas de Influenza A HINI también se determinan en los tratamientos farmacológicos empleados en los pacientes ya que se ha demostrado que los antivirales oseltamivir, peramivir y zanamivir, han disminuido su capacidad farmacológica lo cual sugiere que las nuevas modificaciones genéticas los hace más resistentes a la acción del fármaco (Davlin et al., 2016; Petrie et al., 2019; Quiñones et al., 2020). Es por ello que se hace necesario tener herramientas actualizadas para el diagnóstico, ya que los costos de producción de kits de diagnósticos serán afectados por los cambios virales. Por ejemplo, en Japón, se utilizan aproximadamente entre 20 y 40 millones de kits RIDT cada temporada, lo que implica un costo aproximado entre 200 y 400 millones de dólares estadounidenses por año (Green & George, 2018; Seki et al., 2020).

Esta imperiosa necesidad de actualización de sistemas de diagnóstico de virus Influenza A H1N1 también es reflejada en la reacción de RT-PCR en tiempo Real de un Paso la cual ha sido considerada la prueba dorada para el diagnóstico, sin embargo debido a la constante evolución viral las regiones empleadas para el diseño de sondas y cebadores pierden su eficiencia ante las nuevas variantes genómicas y son verificadas periódicamente (Al-Husayni

& Hassoun, 2014; Cissé et al., 2021; Stellrecht, 2018b). Por ejemplo, se sabe que el gen HA tienen una alta tasa de mutabilidad lo que conlleva a que se desajuste el cebador/sonda (Lee et al., 2010).

Es por ello que el Grupos CAGE generó una serie de juegos de sondas y cebadores para el diagnóstico del virus de influenza A H1N1 como parte de los objetivos de la implementación de kits actualizados de diagnóstico con base a los genes Ha, Np y M1- M2. Puesto que las moléculas diseñadas previo a su implementación se diseñaron con los estudios evolutivos del grupo era fundamental si validación *in silico* antes de ser empleadas en muestras de pacientes.

Una de las ventajas de los estudios *in silico* es que permite determinar la eficiencia de una molécula con la información disponible en las bases de datos públicas que en esta pasantía fue la plataforma de IRD de secuencias de personas infectadas con el virus de Influenza A H1N1 del periodo de enero de 2019 donde se encontraron reportes para todos los meses, pero para el año 2020 se obtuvieron reportes hasta marzo de ese año. Los resultados obtenidos fueron similares a los de otros reportes en relación a la publicación de genomas en bases de datos públicas, los cuales indican que el virus de Influenza A H1N1 disminuyó considerablemente a nivel mundial por la aparición del virus SARS-CoV-2 (Olsen et al., 2020).

Aun con la falta de reportes de genomas de Influenza A H1N1 se logró obtener las secuencias representativas de todo el mundo para los periodos establecidos tal como lo establecido en la minería de datos se generó un sistema sencillo y amigable que permitió la organización mensual para el fácil manejo de la información. Esta parte es fundamental en la validación de cebadores y sondas ya que el resultado estará en función de la realidad biológica que represente la base de datos. Este principio se observó al generar las secuencias consenso con los umbrales de frecuencia. Donde claramente se determinó que umbrales de 20% mostraran que no ha existido cambios en los genes de influenza y por lo tanto se puede

considerar que los cebadores y sondas son efectivos. Sin embargo, esta conclusión no es una realidad biológica ya que el umbral de frecuencia de 100% mostró que el patrón de mutación del virus es mayor respecto al umbral anterior y por lo tanto la eficiencia de la RT-PCR estará en función del diseño de cada molécula empleada en la reacción. Lo anterior es validado con los artículos relacionados al diseño de cebadores y sondas para el diagnóstico del virus de influenza A H1N1 propuesto por grupos de investigación en donde utilizan un número determinado de secuencias y los umbrales de frecuencia son generados con los parámetros establecidos por el diseñador del software y por lo tanto los cebadores solo serán efectivos para algunas cepas del virus de Influenza A H1N1 lo cual está en concordancia con algunos conceptos propuestos por (Stellrecht, 2018a).

Lo anterior fue confirmado con los alineamientos de las secuencias consenso mensual en donde se estableció que las regiones empleadas para el diagnóstico viral son más constantes en los alineamientos con umbrales de frecuencia del 20% mientras que los del 100% mostraran la eficiencia de los cebadores y sondas los cuales estuvieron en función del mes empleado para hacer la comparación. No obstante, los cebadores y sondas del grupo CAGE mostraron que reconocerán preferentemente su secuencia complementaria al realizar el alineamiento tipo BLAST en el que se confirma que se tendrán resultados preferentemente de cepas del año 2019 y 2020 pero también tienen la capacidad de reconocer genes de virus de influenza A H1N1 de cerdos y de este organismo también reconocerá genes de las cepas AH1N2 y H2N3 las cuales también infectan a humanos y cerdos como ocurre con los kits de diagnóstico empleados para estos organismos (Haach, Gava, Cantão, & Schaefer, 2020).

6. Conclusiones

Es crucial la actualización continua y la vigilancia genómica por métodos bioinformáticos de los cebadores diseñados en el proyecto COVID-19-UIS para los genes HA, NP y M1-M2 del virus de la Influenza A H1N1 para el año 2019 a agosto del 2020 y futuros, ya que permiten garantizar y brindar eficientemente mejores alternativas de diagnóstico por RT-PCR en tiempo real de un paso contribuyendo a reducir los falsos negativos y facilitar los procesos de tratamiento de los pacientes que lleguen a estar contagiados. Además, un correcto diagnóstico contribuirá a determinar la eficiencia de los programas de vacunación al establecer la ausencia de nuevas variantes virales. Debido a la eficiencia del inmunógeno para generar la respuesta inmune en el paciente.

Los cebadores para los genes de HA, NP y M1-M2 del virus de la Influenza A H1N1 creados por el proyecto COVID-19-UIS demostraron una alta efectividad *in silico* mediante los alineamiento tipo BLAST respecto a las secuencias caracterizadas a partir de pacientes infectados con el virus reportadas en la base de datos pública IRD en el año 2019 agosto del 2020 y experimentalmente mediante el uso de pruebas de diagnóstico como la RT-PCR en tiempo real de un paso con los oligonucleótidos diseñados en el proyecto.

La PCR de transcripción inversa (RT-PCR) sigue siendo una de las tecnologías con mayor ventaja gracias a su universalidad y adaptabilidad para la caracterización de patógenos virales como Influenza A H1N1, al mostrar una especificidad del 100% de los cebadores y sondas diseñado para el proyecto COVID-19-UIS.

Diseñar una base de datos pública con los genes que codifican para HA, NP y M1-M2 del virus de la Influenza A H1N1 de secuencias reportadas no solo es fundamental para el

correcto diagnóstico por RT-PCR en tiempo real de un paso, sino que permite generar un monitoreo constante de los cambios evolutivos que se han registrado antes y después de las diferentes epidemias y pandemias, lo cual, facilita observar los cambios a nivel nucleotídico para entender el comportamiento del virus.

Los estudios genómicos y pruebas de RT-PCR en tiempo real que se le realizan a pacientes con Influenza A H1N1, es información útil para ser registrada en plataformas como IRD. Esto, con el fin de tener un historial genético de los cambios mutacionales que presenta el virus a lo largo del tiempo, por tanto, estos datos facilitan los análisis bioinformáticos, que son necesarios para la actualización de cebadores y sondas en el diagnóstico de pacientes con Influenza A H1N1 y las demás cepas.

Referencias bibliográficas

Al-Husayni, N., & Hassoun, A. (2014). Making the case for RT-PCR for the diagnosis of H1N1. *Journal of Clinical Virology*, *60*(3), 321.

Arevalo-Rodriguez, I., Buitrago-Garcia, D., Simancas-Racines, D., Zambrano-Achig, P., Campo, R. Del, Ciapponi, A., … Zamora, J. (2020). False-negative results of initial RT-PCR assays for COVID-19: a systematic review. *PLoS ONE*, *15*(1), 1–19.

Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Flegel, V., Fortier, A., … Stockinger, H. (2012). ExPASy : SIB bioinformatics resource portal. *Nucleic Acids Research*, *40*(May), 597–603.

Auladell, M., Jia, X., Hensen, L., Chua, B., Fox, A., Nguyen, T. H. O., … Kedzierska, K. (2019). Recalling the future: immunological memory toward unpredictable influenza viruses. *Frontiers in Immunology*, *10*(1), 1–18.

Boivin, S., Cusack, S., Ruigrok, R. W. H., & Hart, D. J. (2010). Influenza A virus polymerase: structural insights into replication and host adaptation mechanisms. *Journal of Biological Chemistry*, *285*(37), 28411–28417.

Bragstad, K., Hungnes, O., Litleskare, I., Nyrerød, H. C., Dorenberg, D. H., & Hauge, S. H. (2019). Community spread and late season increased incidence of oseltamivir-resistant influenza A(H1N1) viruses in Norway 2016. *Influenza and Other Respiratory Viruses*, *13*(4), 372–381.

Cadena-caballero, C. E., Vera-cala, L. M., Barrios-hernandez, C., Rueda-plata, D., Forero-buitrago, L. J., Torres-jimenez, C. S., … Martinez-perez, F. (2021). Denaturing and dNTPs reagents improve SARS-CoV-2 detection via single and multiplex RT-qPCR.

Virus Research, *en evaluac*.

Charyasriwong, S., Haruyama, T., & Kobayashi, N. (2016). In vitro evaluation of the antiviral activity of methylglyoxal against influenza B virus infection. *Drug Discoveries & Therapeutics*, *10*(4), 201–210.

Chidlow, G., Harnett, G., Williams, S., Levy, A., Speers, D., & Smith, D. W. (2010). Duplex real-time reverse transcriptase PCR assays for rapid detection and identification of pandemic (H1N1) 2009 and seasonal influenza A/H1, A/H3, and B viruses. *Journal of Clinical Microbiology*, *48*(3), 862–866.

Cissé, A., Milucky, J., Ilboudo, A. K., Waller, J. L., Bicaba, B., Medah, I., … Tarnagda, Z. (2021). Comparison of performance between Fast Track Diagnostics Respiratory Kit and the CDC global reference laboratory for influenza rRT-PCR panel for detection of influenza A and influenza B. *Influenza and Other Respiratory Viruses*, (August 2020), 381–388.

Cobb, B. R., Vaks, J. E., Do, T., & Vilchez, R. A. (2011). Evolution in the sensitivity of quantitative HIV-1 viral load tests. *Journal of Clinical Virology*, *52*(1), 77–82.

Datkhaeva, I., Has, P., Fitzgerald, K., & Hughes, B. L. (2019). Outcomes of a negative rapid in fluenza diagnostic test in pregnant women. *American Journal of Perinatology*, *1*(212), 90–96.

Davlin, S. L., Blanton, L., Kniss, K., Mustaquim, D., Smith, S., Kramer, N., … Brammer, L. (2016). Influenza Activity — United States, 2015–16 Season and Composition of the 2016–17 Influenza Vaccine. *MMWR. Morbidity and Mortality Weekly Report*, *65*(22), 567–575.

Day, W. H. E., & McMorris, F. R. (1992). Threshold consensus methods for molecular

sequences. *Journal of Theoretical Biology*, *159*(4), 481–489.

Deng, X., Yang, J., Wang, W., Wang, X., Zhou, J., Chen, Z., … Yu, H. (2020). Case fatality risk of novel coronavirus diseases 2019 in China. *MedRxiv Preprint*, *2507*(1), 1–9.

Deyde, V. M., Sampath, R., & Gubareva, L. V. (2011). RT-PCR/electrospray ionization mass spectrometry approach in detection and characterization of influenza viruses. *Expert Review of Molecular Diagnostics*, *11*(1), 41–52.

Eisfeld, A. J., Neumann, G., & Kawaoka, Y. (2015). At the centre: Influenza A virus ribonucleoproteins. *Nature Reviews Microbiology*, *13*(1), 28–41.

Engelmann, I., Alidjinou, E. K., Lazrek, M., Dewilde, A., & Hober, D. (2017). Necessity to critically review the automatic results of the xpert flu assay. *Diagnostic Microbiology and Infectious Disease*, *88*(1), 26–30.

Fujii, Y., Goto, H., Watanabe, T., Yoshida, T., & Kawaoka, Y. (2003). Selective incorporation of influenza virus RNA segments into virions. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(4), 2002–2007.

Gatherer, D. (2009). The 2009 H1N1 influenza outbreak in its historical context. *Journal of Clinical Virology*, *45*(3), 174–178.

Gavazzi, C., Isel, C., Fournier, E., Moules, V., Cavalier, A., Thomas, D., … Marquet, R. (2013). An in vitro network of intermolecular interactions between viral RNA segments of an avian H5N2 influenza A virus: Comparison with a human H3N2 virus. *Nucleic Acids Research*, *41*(2), 1241–1254.

Gog, J. R., Dos Santos Afonso, E., Dalton, R. M., Leclercq, I., Tiley, L., Elton, D., … Digard, P. (2007). Codon conservation in the influenza A virus genome defines RNA packaging signals. *Nucleic Acids Research*, *35*(6), 1897–1907.

Gotea, V., Veeramachaneni, V., & Makaøowski, W. (2003). Mastering seeds for genomic size nucleotide BLAST searches. *Nucleic Acids Research*, *31*(23), 6935–6941.

Green, D. A., & George, K. S. (2018). Rapid antigen tests for Influenza : rationale and significance of the FDA reclassification. *Journal of Clinical Microbiology*, *56*(10), 1–10.

Haach, V., Gava, D., Cantão, M. E., & Schaefer, R. (2020). Evaluation of two multiplex RT-PCR assays for detection and subtype differentiation of Brazilian swine influenza viruses. *Brazilian Journal of Microbiology*, *51*(3), 1447–1451.

Hall, T., Biosciences, I., Carlsbad, & Ca. (2013). Non-motor symptoms and parkinsonism. *Canadian Journal of Neurological Sciences*, *40*(1), 1–2.

Heldt, F. S., Kupke, S. Y., Dorl, S., Reichl, U., & Frensing, T. (2015). Single-cell analysis and stochastic modelling unveil large cell-to-cell variability in influenza A virus infection. *Nature Communications*, *6*(1), 1–12.

Hoang Vu Mai, P., Ung Thi Hong, T., Nguyen Le Khanh, H., Nguyen Thanh, T., Le Thi, T., Nguyen Vu, S., … Le Quynh, M. (2019). Missed detections of influenza A(H1)pdm09 by real-time RT-PCR assay due to haemagglutinin sequence mutation, December 2017 to March 2018, northern Viet Nam. *Western Pacific Surveillance and Response Journal : WPSAR*, *10*(1), 32–38.

Hutchinson, E. C., & Fodor, E. (2012). Nuclear import of the influenza A virus transcriptional machinery. *Vaccine*, *30*(51), 7353–7358.

Jiang, D., Wang, Q., Bai, Z., Qi, H., Ma, J., Liu, W., … Li, J. (2020). Could Environment Affect the Mutation of H1N1 Influenza Virus ? *International Journal of Environmental Research and Public Health Letter*, *17*(3092), 1–10.

Kim, Y., Vivas Barber, A., & Lee, S. (2020). Modeling influenza transmission dynamics with media coverage data of the 2009 H1N1 outbreak in Korea. *PLOS ONE*, *15*(6), 1–21.

Krammer, F. (2015). Emerging influenza viruses and the prospect of a universal influenza virus vaccine. *Biotechnology Journal*, *10*(5), 690–701.

Kuhner, M. K., & Yamato, J. (2017). A Consensus Method for Ancestral Recombination Graphs. *Journal of Molecular Evolution*, *84*(2–3), 129–138.

Labella, A. M., & Merel, S. E. (2013). Influenza. *Medical.Theclinics.Com*, *97*(1), 621–645.

Lassmann, T., Frings, O., & Sonnhammer, E. L. L. (2009). Kalign2: High-performance multiple alignment of protein and nucleotide sequences allowing external features. *Nucleic Acids Research*, *37*(3), 858–865.

Lee, H. K., Lee, C. K., Loh, T. P., Tang, J. W. T., Chiu, L., Tambyah, P. A., … Koay, E. S. C. (2010). Diagnostic testing for pandemic influenza in Singapore: A novel dual-gene quantitative real-time RT-PCR for the detection of influenza A/H1N1/2009. *Journal of Molecular Diagnostics*, *12*(5), 636–643.

Li, L., Chang, D., Han, L., Zhang, X., Zaia, J., & Wan, X. (2020). Multi-task learning sparse group lasso : a method for quantifying antigenicity of influenza A (H1N1) virus using mutations and variations in glycosylation of Hemagglutinin. *BMC Bioinformatics*, *21*(1), 1–23.

Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., … Lopez, R. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research*, *47*(1), 636–641.

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.

Nakamura, T., Yamada, K. D., Tomii, K., & Katoh, K. (2018). Parallelization of MAFFT for large-scale multiple sequence alignments. *Bioinformatics*, *34*(14), 2490–2492.

Olsen, S. J., Azziz-Baumgartner, E., Budd, A. P., Brammer, L., Sullivan, S., Pineda, R. F., … Fry, A. M. (2020). Decreased influenza activity during the COVID-19 pandemic—United States, Australia, Chile, and South Africa, 2020. *American Journal of Transplantation*, *20*(12), 3681–3685.

Opanda, S., Bulimo, W., Gachara, G., Ekuttan, C., & Amukoye, E. (2020). Assessing antigenic drift and phylogeny of influenza A (H1N1) pdm09 virus in Kenya using HA1 sub-unit of the hemagglutinin gene. *PLoS ONE*, *15*(2), 1–14.

Oropeza Fernández, S., Acosta Herrera, B., Piñón Ramos, A., Valdés Ramírez, O., Savón Valdés, C., Arencibia García, A., … Llop Hernández, A. (2011). Diagnóstico molecular del virus influenza A (H1N1) 2009 y otros virus respiratorios, durante la primera ola pandémica en Cuba. *Revista Cubana de Medicina Tropical*, *63*(2), 147–154.

Pabbaraju, K., Wong, S., Wong, A. A., Appleyard, G. D., Chui, L., Pang, X.-L., … Preiksaitis, J. K. (2009). Design and validation of Real-Time Reverse Transcription-PCR assays for detection of pandemic (H1N1) 2009 virus. *Journal of Clinical Microbiology*, *47*(11), 3454–3460.

Paterson, D., & Fodor, E. (2012). Emerging roles for the Influenza A virus nuclear export protein (NEP). *PLoS Pathogens*, *8*(12), 1–9.

Petrie, J. G., Martin, E. T., Truscon, R., Johnson, E., Cheng, C. K., McSpadden, E. J., … Monto, A. S. (2019). Evaluation of correlates of protection against influenza A(H3N2) and A(H1N1)pdm09 infection: Applications to the hospitalized patient population.

Vaccine, *37*(10), 1284–1292.

Quiñones, M., Liou, D. T., Shyu, C., Kim, W., Vujkovic-cvijin, I., Belkaid, Y., & Hurt, D. E. (2020). " METAGENOTE : a simplified web platform for metadata annotation of genomic samples and streamlined submission to NCBI's sequence read archive ." *BMC Bioinformatics*, *21*(378), 1–13.

Robb, N. C., Te Velthuis, A. J. W., Wieneke, R., Tampe, R., Cordes, T., Fodor, E., & Kapanidis, A. N. (2016). Single-molecule FRET reveals the pre-initiation and initiation conformations of influenza virus promoter RNA. *Nucleic Acids Research*, *44*(21), 10304–10315.

Rodriguez-Frandsen, A., De Lucas, S., Pérez-González, A., Pérez-Cidoncha, M., Roldan-Gomendio, A., Pazo, A., … Nieto, A. (2016). HCLE/C14orf166, a cellular protein required for viral replication, is incorporated into influenza virus particles. *Scientific Reports*, *6*(1), 1–12.

Saccaro, L. F., Galfo, V., Ferranti, S., Russo, A., & Menichetti, F. (2020). Severe respiratory failure in an immunocompetent host with invasive pulmonary aspergillosis and H1N1 influenza. *Le Infezioni in Medicina*, *28*(2), 263–267.

Sautto, G. A., & Ross, T. M. (2019). Hemagglutinin consensus-based prophylactic approaches to overcome influenza virus diversity. *Veterinaria Italiana*, *55*(3), 195–201.

Schoen, K., Horvat, N., Guerreiro, N. F. C., Castro, I. De, & Giassi, K. S. De. (2019). Spectrum of clinical and radiographic findings in patients with diagnosis of H1N1 and correlation with clinical severity. *BMC Infectious Diseases*, *19*(964), 1–8.

Seki, Y., Oda, Y., & Sugaya, N. (2020). Very high sensitivity of a rapid influenza diagnostic test in adults and elderly individuals within 48 hours of the onset of illness. *PLoS ONE*,

15(5), 1–11.

Shah, N. K., Upadhyay, K. J., Prajapati, B., Manwar, V., Gamit, D., & Rathwa, K. (2019). A study of clinical profile of Influenza A (H1N1) pdm09 patients at a tertiary care hospital. *International Archives of Integrated Medicine*, *6*(11), 14–18.

Shapira, S. D., Gat-Viks, I., Shum, B. O. V., Dricot, A., de Grace, M. M., Wu, L., … Hacohen, N. (2009). A physical and regulatory map of host-Influenza interactions reveals pathways in H1N1 infection. *Cell*, *139*(7), 1255–1267.

Singanayagam, A., Zhou, J., Elderfield, R. A., Rebecca, F., Ashcroft, J., Galiano, M., … Barclay, W. S. (2020). Characterising viable virus from air exhaled by H1N1 influenza-infected ferrets reveals the importance of haemagglutinin stability for airborne infectivity. *PLoS Pathogens*, *16*(2), 1–22.

Stellrecht, K. A. (2018a). Incidence of matrix genes mutations affecting PCR tests among influenza H3N2 clades circulating during the 2014/15 season. *Diagnostic Microbiology and Infectious Disease*, *91*(3), 239–244.

Stellrecht, K. A. (2018b). The drift in molecular testing for influenza: mutations affecting assay performance. *Journal of Clinical Microbiology*, *56*(3), 1–21.

Sullivan, S. J., Jacobson, R. M., Dowdle, W. R., & Poland, G. A. (2010). 2009 H1N1 Influenza. *Mayo Clinic Proceedings*, *85*(1), 64–76.

Taubenberger, J. K., & Kash, J. C. (2010). Influenza virus evolution, host adaptation, and pandemic formation. *Cell Host and Microbe*, *7*(6), 440–451.

Te Velthuis, A. J. W., & Fodor, E. (2016). Influenza virus RNA polymerase: insights into the mechanisms of viral RNA synthesis. *Nature Reviews Microbiology*, *14*(8), 479–493.

Ueda, R., Sugiura, T., Kume, S., Ichikawa, A., Larsen, S., Miyoshi, H., … Honda, A. (2013). A novel single virus infection system reveals that influenza virus preferentially infects cells in G1 phase. *PLoS ONE*, *8*(7), 1–10.

Vasin, A. V., Temkina, O. A., Egorov, V. V., Klotchenko, S. A., Plotnikova, M. A., & Kiselev, O. I. (2014). Molecular mechanisms enhancing the proteome of influenza A viruses: an overview of recently discovered proteins. *Virus Research*, *185*(2014), 53–63.

Webster, R. G., & Govorkova, E. A. (2015). Continuing challenges in influenza. *Annals of the New York Academy of Sciences*, *1323*(1), 115–139.

Whiley, D. M., Bialasiewicz, S., Bletchly, C., Faux, C. E., Harrower, B., Gould, A. R., … Sloots, T. P. (2009). Detection of novel influenza A(H1N1) virus by real-time RT-PCR. *Journal of Clinical Virology*, *45*(3), 203–204.

Wu, S., Patel, K. B., Booth, L. J., Metcalf, J. P., Lin, H. K., & Wu, W. (2010). Protective essential oil attenuates influenza virus infection: An in vitro study in MDCK cells. *BMC Complementary and Alternative Medicine*, *10*(1), 69.

Wunderlich, K., Mayer, D., Ranadheera, C., Holler, A. S., Mänz, B., Martin, A., … Schwemmle, M. (2009). Identification of a PA-binding peptide with inhibitory activity against influenza A and B virus replication. *PLoS ONE*, *4*(10), 1–12.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*, *31*(13), 3406–3415.