

ANÁLISIS COMPUTACIONAL DE CEBADORES PARA INFLUENZA

Implementación de un procedimiento computacional para establecer la eficiencia de cebadores y sondas para RT-PCR en los genes: Hemaglutinina (*HA*), Proteína Nuclear (*NP*) y Gen de Matriz (*M1-M2*) de Influenza A H1N1

Iver Leandro Daza Sandoval

Trabajo de Grado para Optar el Título de Biólogo

Director

Francisco José Martínez Pérez

Doctor en Ciencias

Codirector

Carlos Jaime Barrios Hernández

Doctor en Informática y Ciencias Computacionales

Universidad Industrial de Santander

Facultad de Ciencias

Escuela de Biología

Bucaramanga

2025

Dedicatoria

A mis padres y hermanos, por el apoyo incondicional y extenso.

A Juliana y a Brandon, confidentes y pilares de toda una carrera.

Agradecimientos

Gracias a la Universidad Industrial de Santander por la formación y apoyo dado durante toda la carrera de Biología.

Gracias al Laboratorio de Genómica Celular Aplicada del Grupo de Investigación en Cómputo Avanzado y a Gran Escala y al Centro de Supercomputación y Cálculo Científico de la Universidad Industrial de Santander por la financiación y apoyo fundamental para la realización de este proyecto.

Gracias al profesor Francisco José Martínez Pérez y al profesor Carlos Jaime Barrios Hernández por la confianza, la formación profesional y la oportunidad de participar en este proyecto.

Gracias a Yordy Cangrejo Useda, Cristian Enrique Cadena Caballero y a Cristina Isabel Carvajal Acuña por la asistencia y consejos para la redacción de este trabajo.

Tabla de Contenido

| | Pág. |
|----------------------------------|-------------|
| INTRODUCCIÓN..... | 13 |
| 1. OBJETIVOS | 20 |
| 1.1 Objetivo General..... | 20 |
| 1.2 Objetivos Específicos | 20 |
| 2. Competencias adquiridas..... | 21 |
| 3. Metodología | 22 |
| 4. Resultados | 27 |
| 5. Discusión | 39 |
| 6. Conclusiones | 43 |
| Referencias Bibliográficas | 43 |

Lista de Tablas

Tabla 1. *Ejemplo de la cantidad de secuencias de los genes HA, NP y M1-M2 del virus de Influenza A H1N1 del año 2018..... 29*

Lista de Figuras

| | Pág. |
|---|-------------|
| Figura 1. Diagrama de flujo de ejecución de algoritmo “Protocolo para Filtrar, Alinear y Reportar Cebadores versión 1.0” (PROT.FAR.Ceb v-1.0)..... | 22 |
| Figura 2. Diagrama de flujo del Módulo 1 para el uso de la aplicación MGFS-v1.0. a partir de la base de datos “NCBI Virus” | 24 |
| Figura 3. Concepción de los fundamentos de las Reglas de Corte | 25 |
| Figura 4. Secuencias descargadas de la base de datos NCBI Virus del Virus de influenza A H1N1 | 27 |
| Figura 5. Ejemplo de utilización de archivos de texto plano generados por el módulo MASN-v1.0 | 29 |
| Figura 6. Validación de la primera Regla Corte con los archivos que contienen las secuencias descargadas se utilizan con el programa MAFFT | 30 |
| Figura 7. Validación de la Segunda Regla Corte con el archivo con las secuencias descargadas se utiliza con el programa MAFFT | 31 |
| Figura 8. Comparación de resultados usando conjuntos de cebadores con sonda y conjuntos sin sonda | 32 |
| Figura 9. Comparación de resultados usando umbrales de consenso diferentes para segmento 4, usando solo conjuntos con sonda y usando todos los cebadores | 34 |

Lista de Apéndices

Apéndice A. Manual de usuario para el uso del algoritmo implementado.

Apéndice B. Código en lenguaje python usado para los módulos.

Apéndice C. Base de datos implementada para el virus de influenza A H1N1, incluye secuencias nucleotídicas del virus y cebadores y sondas obtenidas búsqueda de literatura.

Apéndice D. Alineamientos que muestran la aplicación de la primera y segunda regla de corte para los genes *HA*, *NP*, *M1-M2* del virus de Influenza A H1N1.

Apéndice E. Artículos Reportados en el PubMed con secuencias de cebadores y sondas para identificar los genes *HA*, *NP*, *M1-M2* del virus de Influenza A H1N1 con RT-PCR y/o RT-qPCR.,

Apéndice F. Secuencias consenso de los genes *HA*, *NP*, *M1-M2* del virus de Influenza A H1N1 generados con un umbral de frecuencia del 70%

Apéndice G. Secuencias consenso de los genes *HA*, *NP*, *M1-M2* del virus de Influenza A H1N1 generados con un umbral de frecuencia del 97% .

Apéndice H. Reporte de comparación del consenso elegido con la lista de secuencias de cebadores en formato PDF mostrando porcentaje de identidad respecto a la secuencia consenso y el conjunto de cebadores óptimo

Nota: Los apéndices están adjuntos y puede visualizarlos en la base de datos de la biblioteca UIS y en el siguiente enlace: <https://github.com/GenomicUIS/Influenza-A-H1N1-python-script.git>

Glosario

- Cebador: Secuencia corta de nucleótidos, ya sean de Ácido Desoxirribonucleico (ADN) o Ácido Ribonucleico (ARN), que inicia la replicación del material genético. En PCR, los cebadores marcan los extremos de la región a amplificar. Son diseñados para hibridar a su secuencia complementaria.
- HA: Hemaglutinina, proteína de la superficie del virus de influenza A H1N1. Facilita la unión del virus a células humanas por su receptor. Es el principal antígeno para la generación de vacunas contra el virus.
- HPC: *Computación de Alto Rendimiento (High-Performance Computing)* Uso de supercomputadoras o clúster de computadoras para resolver problemas complejos que requieren gran capacidad de cálculo. Aplicado en simulaciones físicas, modelado climático, bioinformática y análisis de grandes volúmenes de datos.
- IUPAC: *Unión Internacional de Química Pura y Aplicada (International Union of Pure and Applied Chemistry*, por sus siglas en inglés), organización que establece estándares globales en química, al incluir nomenclatura de compuestos, símbolos de elementos y códigos para secuencias genéticas y proteicas.
- MAFFT: *Herramienta de Alineamiento Rápido basado en Transformada de Fourier*, programa informático que alinea múltiples secuencias de ADN, ARN o proteínas. Destaca por su precisión y velocidad, mediante algoritmos matemáticos avanzados. Esencial en estudios evolutivos y comparación de secuencias de ácidos nucleicos y/o genomas.
- ORF (*Open Reading Frame*) *Marco Abierto de lectura*: región del ARN mensajero que contiene los codones que codifican para la respectiva proteína. Se caracteriza para células eucariontes en iniciar con el codón AUG que codifica para metionina y finalizar con tres opciones de codones de paro, a saber: UAG, UGA o UAA.
- MIT: *Licencia del Instituto Tecnológico de Massachusetts*, es una licencia de programa de código abierto (*open source*) permisiva. Permite uso, modificación y redistribución comercial sin restricciones, únicamente se exige incluir el o los derechos de autor (*copyright*) original.
- NP: Nucleoproteína, proteína que se une al material genético en forma de ARN del virus de influenza A H1N1, al empaquetarlo y estabilizarlo. Es crucial para la replicación viral.
- *Código Abierto (open source)*: modelo de desarrollo de programa donde el código fuente es público, gratuito y modificable. Fomenta la colaboración, transparencia y adaptabilidad de cualquier código.
- RT-PCR: *Reacción en Cadena de la Polimerasa con Transcriptasa Inversa*, es una técnica de laboratorio que detecta y amplifica el material genético en forma de ARN. Primero convierte ARN en ADN complementario (ADNc) con una transcriptasa inversa, posteriormente se amplifica la región de interés o todo el ADNc con PCR de punto final.

ANÁLISIS COMPUTACIONAL DE CEBADORES PARA INFLUENZA

- RT-qPCR: *Reacción en Cadena de la Polimerasa con Transcriptasa Inversa en Tiempo Real*. Técnica de laboratorio que amplifica, detecta y cuantifica material genético en forma de ARN por medio de la generación de una señal luminosa generada por una sonda o por un reactivo fluorescente en cada ciclo de amplificación de la región nucleotídica seleccionada en la PCR de punto final.
- RT-qPCR de dos pasos: proceso de RT-qPCR primero se sintetiza y purifica el ADNc para después amplificar y cuantificar la región blanco por PCR de punto final.
- RT-qPCR de un paso: proceso de RT-qPCR en donde no se requiere la purificación del ADNc ya que después de su síntesis, inicia de inmediato la PCR de punto final.
- Sonda: oligonucleótido para reacciones de polimerización en tiempo real de ácidos nucleicos que contiene un fluoróforo unido covalentemente a uno de sus nucleótidos, con el fin de generar la luminiscencia que será cuantificada en cada ciclo de polimerización.
- UGENE: *Herramienta de Análisis Genómico*, es un programa de Código Abierto para analizar secuencias de ADN, ARN o proteínas. Incluye herramientas para alineamiento, anotación genómica, filogenia y visualización de datos.
- Pb: abreviatura para indicar pares de bases en ADN o ARN.
- NCBI: abreviatura para National Center for Biotechnology Information.

Resumen

Título: Implementación de un procedimiento computacional para establecer la eficiencia de cebadores y sondas para RT-PCR en los genes: Hemaglutinina (*HA*), Proteína Nuclear (*NP*) y Gen de Matriz (*M1-M2*) de Influenza A H1N1*

Autor: Iver Leandro Daza Sandoval**

Palabras Clave: PCR, HPC, Python, Bioinformática, Cebador

Descripción: Esta pasantía desarrolló un protocolo computacional en Python para evaluar la eficiencia de cebadores y sondas de RT-PCR en los genes *HA*, *NP* y *M1-M2* del virus Influenza A H1N1 mediante el uso de secuencias consenso. Mediante herramientas bioinformáticas (MAFFT, UGENE) y Computación de Alto Rendimiento (HPC), se realiza el análisis de polimorfismos genómicos que afectan la sensibilidad diagnóstica, identificando regiones críticas para el diseño de oligonucleótidos. El flujo de trabajo se organizó en 3 módulos dentro del programa: Filtrado y alineamiento de secuencias (1-Filtración.py), generación de consenso (2-Alineamiento.py) y evaluación de cebadores con visualización en PDF (3-Reporte.py). El resultado de este flujo es una herramienta open source para análisis rápido de grandes volúmenes de datos para la detección de discrepancias en cebadores reportados, proponiendo ajustes para reducir falsos negativos. Este trabajo aporta una solución reproducible para mejorar diagnósticos del virus de influenza A H1N1, alineándose con estrategias globales de la Organización Mundial de la salud (OMS). Los resultados se publicaron en un repositorio GitHub bajo licencia MIT, lo que facilita que sea adoptado por la comunidad científica.

* Trabajo de Grado

** Facultad de Ciencias. Escuela de Biología. Director: Francisco José Martínez Pérez. Doctor en Ciencias. Codirector: Carlos Jaime Barrios Hernández. Doctor en Informática.

*

Abstract

Title: Implementation of a computational procedure to establish the efficiency of primers and probes for RT-PCR in the genes: Hemagglutinin (*HA*), Nuclear Protein (*NP*), and Matrix Gene (*M1-M2*) of Influenza A H1N1**

Author: Iver Leandro Daza Sandoval⁵

Key Words: PCR, HPC, Python, Bioinformatic, Primer

Description: This internship developed a computational protocol in Python to evaluate the efficiency of RT-PCR primers and probes in the *HA*, *NP*, and *M1-M2* genes of the Influenza A H1N1 virus using consensus sequences. Using bioinformatics tools (MAFFT, UGENE) and High Performance Computing (HPC), genomic polymorphisms affecting diagnostic sensitivity are analyzed, identifying critical regions for oligonucleotide design. The workflow was organized into three modules within the program: sequence filtering and alignment (1-Filtration.py), consensus generation (2-Alignment.py), and primer evaluation with PDF visualization (3-Report.py). The result of this workflow is an open-source tool for rapid analysis of large volumes of data to detect discrepancies in reported primers, proposing adjustments to reduce false negatives. This work provides a reproducible solution to improve influenza A H1N1 virus diagnostics, aligning with global strategies of the World Health Organization (WHO). The results were published in a GitHub repository under an MIT license, facilitating its adoption by the scientific community.

* Degree Work

⁵Faculty of Sciences. School of Biology. Director: Francisco José Martínez Pérez. Doctor of Science. Co-director: Carlos Jaime Barrios Hernández. Doctor of Computer Science.

Introducción

La pandemia ocurrida en el año 2009 por el virus de influenza A H1N1 causó la muerte de una parte de la población humana y por esto se ha estudiado, para prevenir nuevos brotes y mejorar las herramientas de su diagnóstico para combatir posibles sus nuevas variantes (Neumann, Noda, y Kawaoka 2009). Antes de la mencionada pandemia, han surgido otras, como la pandemia española en 1918, y aun así, el virus de influenza A H1N1 continúa su circulación en la población humana a pesar de que la Organización Mundial de la Salud (OMS) declaró el fin de la pandemia en el año 2010 (Monamele et al. 2019). Sin embargo, el virus de influenza A H1N1 persiste y ha generado nuevos brotes sin llegar a producir una pandemia.

El virus de influenza A H1N1 es un virus zoonótico, es decir, posee la facilidad de propagación entre poblaciones humanas y animales debido a la modificación de sus constituyentes antigénicos generados por una continua deriva genética, lo que le permite evadir la respuesta inmune del hospedero (Ravina et al. 2022, Wahlgren 2011). Además, pertenece a la familia Orthomyxoviridae y su genoma está compuesto por 8 hebras individuales de ARN, que codifican para 10 proteínas. Entre ellas se encuentran la nucleoproteína (NP), gen hemaglutinina (HA) y genes de matriz 1 (M1) y 2 (M2). En la cubierta viral HA y NP son mayoritarias respecto a M1 y M2 (Blanco-Lobo et al. 2019; Kumar, Broor, y Rajala 2016).

La proteína HA es fundamental en el proceso invasivo al sistema inmune puesto que el gen que la codifica es más susceptible a mutación (Chen et al. 2024; Wang et al. 2024). Dado que estos cambios son pequeños, se acumulan a lo largo de los años y pueden llegar a ser impedimentos significativos para tener un diagnóstico asertivo en el paciente y, por lo tanto, ha sido necesario actualizar el esquema de vacunación acorde a estos cambios empleando en esta proteína para nuevas vacunas (Sharma et al. 2019; Zhao et al 2025.).

ANÁLISIS COMPUTACIONAL DE CEBADORES PARA INFLUENZA

Se ha propuesto que el origen de las pandemias generadas por el virus de influenza AH1N1 se deben a cambios novedosos en el virus, por ejemplo, la pandemia del año 2009 que se ha postulado que fue causada por una variación del virus de Influenza aviar; mientras que la pandemia española pudo ser generada por una variación en la proteína NP originada en dos variantes del virus proveniente de cerdos (Mänz et al. 2013). De las proteínas del virus de influenza AH1N1 se destacan la HA y NP, que están envueltas en una proteína de matriz envolviendo la membrana viral (Blanco-Lobo et al. 2019). A pesar de los procesos antigénicos de HA antes mencionados, se ha utilizado también a la proteína M2 del virus, ya que posee un ectodominio conservado que puede ser usado para una vacuna de enfoque más universal (Neumann, Noda, y Kawaoka 2009).

Como se indicó anteriormente, más allá de su abundancia, las proteínas HA y NP pasan por presiones de mutagénesis, selección y adaptaciones con cada brote viral (Jones et al. 2019). En este sentido, la liberación de la NP ocurre con la ayuda de la proteína M2, que funciona como canal iónico, después de la fusión con la membrana endosomal (García y Aris-Brosou 2014). Estas características se utilizan para la clasificación de los virus de influenza A, en donde la H hace referencia a la hemaglutinina con 19 subtipos y la N corresponde a la neuraminidasa que participa en la liberación del virus en las células infectadas con 11 subtipos (Rahman et al. 2017; Wahlgren 2011; Er 2025).

Para la detección de virus en la pandemia de 2009 se tuvo una alta demanda de pruebas moleculares, lo que causó una escasez tanto en reactivos y enzimas para procesos de RT-qPCR de un paso, además de personal disponible para dichas tareas (Shu et al. 2021). En este sentido, la OMS ha solicitado a los científicos mejorar estos protocolos para prevención y contención de futuras posibles pandemias en su marco de trabajo del plan 2019 – 2030, respecto al mejoramiento de herramientas de diagnóstico y tratamiento como lo son vacunas, antivirales y de detección como la RT-qPCR en tiempo real de un paso (World Health Organization 2019). Con relación a lo

ANÁLISIS COMPUTACIONAL DE CEBADORES PARA INFLUENZA

anterior, la OMS e instituciones de salud de sus países aliados y con la colaboración de laboratorios académicos y comerciales, se han generado protocolos para la estandarización de técnicas de diagnóstico en el control de enfermedades entre las que se encuentran los virus de Influenza (Tahan et al. 2021; World Health Organization 2019).

En este sentido, existen kits comerciales para una detección más rápida del virus, basados en el reconocimiento antigénico de sus proteínas, como lo podría ser el kit EZ Flu o TRU FLU, pero su potencial de identificación no es óptimo, lo que en ocasiones lleva a imprecisiones en el diagnóstico (Apte y Daniel 2009). Es decir, debido a las nuevas mutaciones en la región de identificación viral se pueden provocar resultados falsos negativos (Hasan et al. 2021), además de crear resistencia a fármacos en uso de los pacientes (García y Aris-Brosou 2014). Lo anterior ha sido verificado con kits comerciales para el diagnóstico del virus de SARS-CoV-2 (Lee et al. 2021) o ciencia básica (Sajal et al. 2024) e incluso la carga viral de la muestra, transporte y procesamiento generan resultados falsos negativos por RT-qPCR (Landry 2011).

Entre los otros métodos para la detección del virus de Influenza A H1N1 se destaca la RT-PCR y/o RT-qPCR de uno o dos pasos, que puede usar muestras tanto nasofaríngeales como orales, además de ser un método de detección más preciso respecto a otros (Majeed y Lee 2021). Cabe resaltar que la RT-qPCR posee una eficiente capacidad de detección, economiza tiempo en la obtención del resultado, incluso en muestras con baja carga viral puede identificar al agente patógeno ya que la señal de fluorescencia podrá ser determinada por el termociclador que se acopla al área objetivo (National Institute of Hygiene and Epidemiology, Hanoi, Viet Nam et al. 2019). Otra ventaja es que en una RT-qPCR se puede emplear más de un gen viral por lo tanto su capacidad de detección es de alto espectro, con un bajo costo de implementación y practicidad (Kudo et al. 2020). Es por ello que la RT-qPCR ha sido considerada esencial en el control de

ANÁLISIS COMPUTACIONAL DE CEBADORES PARA INFLUENZA

enfermedades (Harada et al. 2012; National Institute of Hygiene and Epidemiology, Hanoi, Viet Nam et al. 2019; Tahan et al. 2021; Zhu et al. 2020).

Uno de los elementos fundamentales para tener un óptimo proceso de PCR, RT-PCR o RT-qPCR de uno o dos pasos, son los diseños de cebadores y/o sondas, ya que el correcto diseño de ambas moléculas incrementa la sensibilidad, especificidad y precisión en los sitios de hibridación de los cebadores y/o sonda (Nolan 2013). Es de resaltar que el uso de una sonda con un fluoróforo en el ensayo de RT-qPCR permitirá observar y cuantificar su fluorescencia con el equipo pertinente, la cual es consecuencia de la actividad de la polimerización de la enzima ADN polimerasa termoestable y, por lo tanto, se determina de forma indirecta la hibridación de la sonda y cebadores con su secuencia complementaria y es por ello que es vital que el diseño de la sonda y cebadores sean homólogos a la secuencia blanco para evitar resultados falsos negativos (Nolan 2013, Meuer et al. 2004).

En ese sentido, se ha establecido que el tamaño de los cebadores para PCR o RT-PCR en cualquiera de sus modalidades deben tener un rango de 18 a 30 pb (Meuer et al. 2004). Si los cebadores tienen una mayor cantidad de nucleótidos se aumenta su especificidad y por lo tanto su eficiencia de hibridación es mayor, lo que generará que el producto de amplificación tenga una concentración similar a la de su región blanco. Sin embargo, si el cebador tiene una menor cantidad de nucleótidos, se puede presentar hibridación con otras regiones aun cuando se tenga una mayor eficiencia, por lo que se considera que el cebador presenta una especificidad reducida. Lo anterior conlleva a amplificaciones en áreas que no son de interés lo que causa amplicones adicionales y posiblemente diferentes al esperado (Meuer et al. 2004, Harada et al. 2012).

Diversos autores, como Meuer et al. 2004 y Nolan 2013, indican que en el diseño de cebadores y sondas debe ser altamente riguroso, en especial en lo referente a la generación de

ANÁLISIS COMPUTACIONAL DE CEBADORES PARA INFLUENZA

respuestas rápidas en pandemias, donde la detección exacta del virus u otro agente patógeno contribuirá a las políticas de salud pública y tratamientos virales. Es por ello que para garantizar resultados óptimos en un análisis de PCR, RT-PCR y/o RT-qPCR, los cebadores deben cumplir con las siguientes características clave acordes a los autores antes citados:

- Especificidad estricta. Secuencias de 18-30 pb con contenido GC del 40-60% con una distribución equilibrada de bases, es decir, evitar que el mismo nucleótido esté de manera continua 4 veces o más. Además de impedir la complementariedad en extremos 3' con un $\Delta G \geq -2.0$ kcal para prevenir dímeros inespecíficos.
- Para la amplificación de las secuencias de ARN mensajero eucarionte maduro, los cebadores deben diseñarse para reconocer las uniones exón-exón con el fin de evitar la amplificación de variantes de procesamiento alternativo ARN, según sea el caso.
- La Temperatura de alineamiento de los cebadores y/o sonda deben tener un rango de 55-65°C, con diferencia $\leq 2^\circ\text{C}$ entre ellos, para tener hibridación precisa entre las tres moléculas, lo que maximizará la sensibilidad del proceso en muestras con baja concentración de la región blanco.
- Se debe tener una correcta selección del fluoróforo de la sonda, la cual estará en función de la capacidad de medición del termociclador seleccionado.

La validación bioinformática de las condiciones termodinámicas de los cebadores y sonda puede realizarse con programas comerciales como son Primer3, OligoArchitect u otros del dominio público e incluso por inteligencia artificial; pero no es recomendable ya que por regla, se debe de considerar literatura científica especializada, que muestre la validación del modelo matemático utilizado para el diseño del programa con resultados experimentales para la determinación de

ANÁLISIS COMPUTACIONAL DE CEBADORES PARA INFLUENZA

dímeros entre cebadores y/o sonda, sus estructuras secundarias y homología con otras secuencias nucleotídicas reportadas en bases de datos internacionales.

Debido a la gran cantidad de datos que genera el uso de la reacción de RT-PCR a nivel mundial, los posibles costos incurridos en el uso de cebadores y la necesidad de resultados más precisos, se requiere implementar HPC para complementar y/o mejorar las herramientas de diagnóstico como lo es RT-qPCR para la detección de virus de influenza A H1N1. Lo anterior presentará beneficios significativos en los diseños de identificación viral al reducir el tiempo necesario para la implementación experimental con el fin de obtener resultados, aumentará la cantidad de datos que se pueden usar en los diseños moleculares y por ende la presión del ensayo con cualquier tipo de muestra.

Para contribuir a la generación de soluciones a la problemática antes indicada, en esta pasantía se desarrolló con el Grupo de Investigación en Cómputo Avanzado y a Gran Escala de la Universidad Industrial de Santander una alternativa bioinformática de acceso abierto, que facilitó la implementación de las herramientas bioinformáticas para el análisis evolutivo de secuencias de los genes del virus de influenza A H1N1, con el fin de generar sus secuencias consenso obtenidas con dos umbrales de frecuencia para que, en ellas, se identifiquen las regiones homólogas con cebadores y sondas para el análisis *in silico* de su patrón de hibridación para RT-qPCR.

Lo anterior se realizó con el desarrollo de nuevo código Python que ejecutó algoritmos de alineamiento y consenso de los genes de *HA*, *NP* y *M1-M2* del virus de influenza A H1N1, con el fin de establecer la precisión de cebadores y/o sondas reportados en la literatura respecto a su región consenso complementaria. El resultado permite al usuario evidenciar cebadores óptimos, que amplifican por RT-qPCR regiones de los genes de nuevas cepas virales empleadas para el diagnóstico, ya que esta nueva herramienta bioinformática muestra la probabilidad de obtener

ANÁLISIS COMPUTACIONAL DE CEBADORES PARA INFLUENZA

resultados falsos negativos y de esta forma se contribuirá a economizar los costos y tiempo de la aplicación de la RT-qPCR en laboratorios de ciencia básica y clínicos para la identificación del virus en muestras de pacientes o animales contagiado.

Objetivos

1.1 Objetivo General

- Crear un código con un lenguaje de programación que determine y muestre las diferencias entre los cebadores y sondas de RT-PCR reportadas en la literatura respecto al polimorfismo de su región blanco en los genes *HA*, *NP* y *M1-M2* del virus de influenza A H1N1.

1.2 Objetivos Específicos

- Diseñar un nuevo prototipo computacional que permita la visualización y análisis de resultados RT-PCR para identificar la homología de los genes *HA*, *NP* y *M1-M2* del virus de influenza A H1N1 respecto a sus cebadores y sondas para RT-PCR reportados en la literatura.
- Proponer una solución computacional open source para Influenza A H1N1 que permita el análisis y visualización de problemas asociados en la RT-PCR, soportado en computación de alto rendimiento.
- Establecer lineamientos generales para el desarrollo de herramientas bioinformáticas aplicadas al análisis de resultados de RT-PCR para los genes *HA*, *NP* y *M1-M2* del virus de influenza A H1N1.

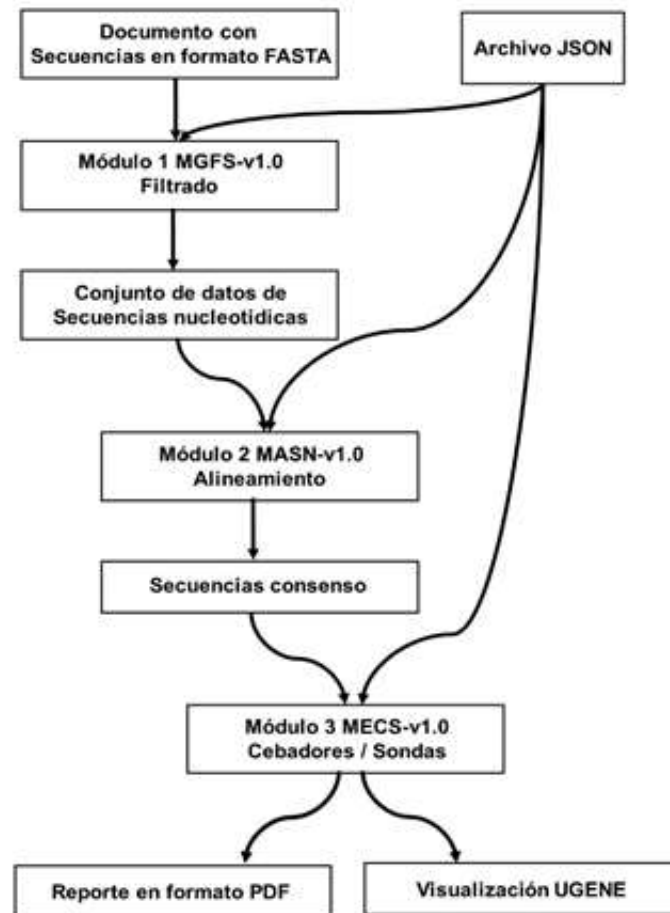
Competencias adquiridas

- Comprende conceptos esenciales del lenguaje de programación seleccionado para el análisis de ácidos nucleicos del virus de influenza.
- Diseña código de programación, con documentación robusta, mediante pautas óptimas para el lenguaje de programación para su uso en el análisis de los genes del virus de influenza A H1N1.
- Estructura y ejecuta el análisis de datos con HPC para hacer búsquedas con los parámetros de programación relacionados al genoma del virus de influenza.
- Expone los puntos clave de los resultados del análisis de PCR a partir del código escrito proponiendo optimizaciones para futuros desarrollos del virus AH1N1 u otros organismos.

3. Metodología

Para la comparación de las sondas y/o cebador de RT-PCR o RT-qPCR, respecto a una secuencia consenso de los genes *HA*, *NP* y *M1-M2* del virus de influenza A H1N1 generada en un intervalo de tiempo y determinar la probabilidad de resultados exitosos o falsos negativos, se generó un programa denominado “Protocolo para Filtrar, Alinear y Reportar Cebadores” versión 1.0 (PROT.FAR.Ceb v-1.0), el cual está diseñado en tres módulos, cuya ejecución sigue un patrón de lógica matemática explicado por el diagrama de flujo que se muestra en la figura 1.

Figura 1. Diagrama de flujo de ejecución del programa “Protocolo para Filtrar, Alinear y Reportar Cebadores” versión 1.0 (PROT.FAR.Ceb v-1.0).



Nota: El diseño del sistema cumple con estándares FAIR (Findable, Accessible, Interoperable, Reusable) mediante el archivo JSON que documenta todos los parámetros analíticos.

Los detalles de los algoritmos computacionales desarrollados para su funcionamiento, se encuentran publicados para el uso de la comunidad nacional e internacional en el repositorio

GitHub del Laboratorio de Genómica Celular Aplicada del Grupo de Investigación en Cálculo Científico y a Gran Escala de la UIS, en la dirección electrónica <https://github.com/GenomicUIS/Influenza-A-H1N1-python-script>, además de un manual que explica los procesos metodológicos de cada módulo (Apéndice A).

3.1. Módulo 1: Generación de Formatos de Secuencias (MGFS-v1.0) para los genes *HA*, *NP* y *M1-M2* del virus de Influenza A H1N1.

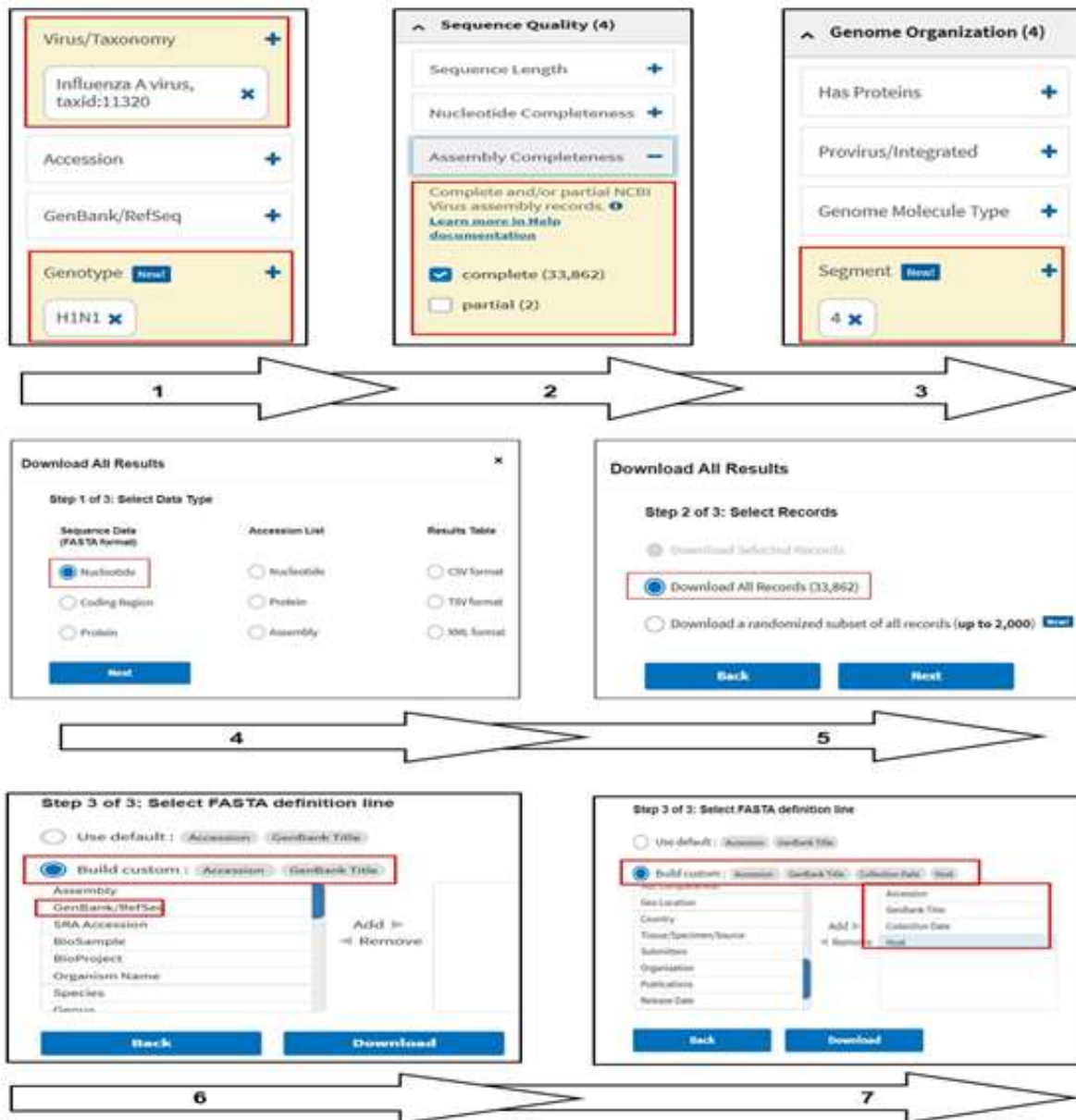
Se identificaron las secuencias de los genes *HA*, *NP* y *M1-M2* del virus de Influenza A H1N1 en la base de datos del GenBank “NCBI Virus” del NCBI (Sayers et al., 2025) por medio de los filtros indicados de la página de la base de datos, estos filtros son: Virus/Taxonomy, Genotype, Segment y Assembly Completeness, tipo de hospedero, fecha de colección (Figura 2). Posteriormente se descargaron en formato FASTA con los siguientes filtros: nucleótidos, descarga de todos los registros, y selección de los parámetros FASTA; entre los parámetros se seleccionó Número de Acceso, título en el GenBank y fecha de colección de la muestra) (Figura 1, Apéndice A). Las secuencias descargadas fueron filtradas automáticamente por el módulo de alineamiento y se generaron dos archivos FASTA: uno con el total de secuencias y el segundo, con las secuencias sin regiones indeterminadas, el cual fue empleado en los siguientes análisis del algoritmo.

3.2. Módulo 2: Análisis de Secuencias Nucleotídicas (MASN-v1.0)

Las secuencias fueron alineadas a partir del programa MAFFT V.7. (Katoh y Standley 2013), implementado en el módulo MASN-v1.0 del programa, con los parámetros por defecto. Debido a que la mayoría de conjuntos de cebadores/sondas están diseñadas para reconocer regiones expresadas de los genes del virus de influenza A H1N1, se realizaron recortes en los alineamientos bajo los parámetros de las “Reglas de corte” (Figura 3), el cual reconoce el codón eucarionte de inicio de la transcripción “AUG” y los codones de parada UAG, UGA y UAA (Apéndice D).

ANÁLISIS COMPUTACIONAL DE CEBADORES PARA INFLUENZA

Figura 2. Diagrama de flujo del Módulo 1 para el uso de la aplicación MGFS-v1.0. a partir de la base de datos “NCBI Virus”



Nota: En el paso 1 se indica la taxonomía y tipo del virus. En el paso dos, se seleccionan genomas completos; y en el 3, el segmento genómico deseado. Para esta representación se utilizó el segmento 4 del Virus de Influenza A H1N1 que corresponde a HA.

Una vez realizada la edición de las secuencias, se generó la secuencia consenso con el umbral del 70% de identidad, con los parámetros de UGENE y Levitsky (Rose et al. 2019) (Apéndice F). Los alineamientos ya editados son procesados por el módulo “MGFS-v1.0.”.

ANÁLISIS COMPUTACIONAL DE CEBADORES PARA INFLUENZA

Figura 3. Concepción de los fundamentos de las Reglas de Corte.



Nota: En la primera Regla de Corte el programa cortará en el segundo codón con mejor promedio como se indica con las tijeras. Pero en la segunda Regla de Corte el Usuario realiza el corte donde lo considere. La regla permite seleccionar codones de paro en extremos 3' UTR de alto número de nucleótidos. En letras Rojas se muestran los posibles codones de inicio y paro de la traducción presentes en un alineamiento hipotético

3.3. Módulo 3: Evaluación de Cebadores y Sondas (MECS-v1.0)

Para validar la efectividad de los cebadores y sondas, se empleó una lista generada por el Laboratorio de Genómica Celular Aplicada, del Grupo de Investigación Cómputo Avanzado y a Gran Escala, que reportó los conjuntos hasta el año 2018. Esta lista se actualizó hasta el año 2024, mediante la búsqueda de literatura en la base de datos el Centro Nacional para la Información Biotecnológica NCBI (Sayers et al., 2025), a partir de las palabras clave “Influenza A H1N1 RT-PCR Human Diagnosis” y “Influenza A H1N1 RT-qPCR Human Diagnosis”. Se utilizaron cebadores y sondas con nucleótidos canónicos, es decir: A, T, G y C, y con nucleótidos degenerados, es decir, con más de dos opciones de nucleótido en la misma posición; y se almacenaron en un archivo de texto plano en formato FASTA, con dirección 5’ – 3’ (Apéndice C).

ANÁLISIS COMPUTACIONAL DE CEBADORES PARA INFLUENZA

A partir de las secuencias consenso generadas en el módulo anterior, se realizó la comparación de los conjuntos identificados para cada uno de los genes y se seleccionaron aquellos que presentaron mayor homología respecto a las secuencias consenso, lo cual se corroboró mediante el valor porcentual obtenido por el análisis del programa (Apéndice H). El cebador sentido, la sonda y el cebador antisentido, se identificaron dentro del alineamiento por los colores rojo, azul y verde, respectivamente. Los análisis de cada gen, se almacenaron en archivos PDF, con los valores obtenidos para cada conjunto. Además, la eficiencia del programa se comparó con los programas Primer 3, PrimerQuest y Primer BLAST al utilizar la secuencia consenso y los cebadores y/o sondas resultantes del programa.

4. Resultados

4.1. Secuencias obtenidas con el Módulo Generación de Formatos de Secuencias (MGFS-v1.0) para los genes *HA*, *NP* y *MI-M2* del virus de Influenza A H1N1.

Para la validación de la aplicación del Módulo MGFS-v1.0 se utilizaron las secuencias de los genes *HA*, *NP* y *MI-M2* del Virus de Influenza A H1N1, reportados en la base de datos “NCBI Virus” del año 2009 al 2024. Hasta el 24 de agosto de 2025, se reportaron un total de 23,949 secuencias para el gen *HA*, mientras que para los genes *NP* y *MI-M2* el total de secuencias fue de 23,946. Por lo tanto, el Módulo MGFS-v1.0 descargó un total 71,841 secuencias en tres documentos independientes en formato FASTA para cada gen (Figura 4, Apéndice C).

Figura 4. Secuencias descargadas de la base de datos NCBI Virus del Virus de influenza A H1N1

| Segment | Count | Selected |
|---------|----------|-------------------------------------|
| 1 | (23,945) | <input type="checkbox"/> |
| 2 | (23,946) | <input type="checkbox"/> |
| 3 | (23,946) | <input type="checkbox"/> |
| 4 | (23,949) | <input checked="" type="checkbox"/> |
| 5 | (23,946) | <input checked="" type="checkbox"/> |
| 6 | (23,946) | <input type="checkbox"/> |
| 7 | (23,946) | <input checked="" type="checkbox"/> |
| 8 | (23,946) | <input type="checkbox"/> |

Nota: Los genes *HA*, *NP* y *MI-M2* del virus de Influenza A H1N1 en la base de datos se indican con numeración, es decir 4, 5 y 7 respectivamente. En la página se indican en la casilla de verificación que al ser seleccionada se colorea en azul. La selección se indica con rectángulos rojos.

ANÁLISIS COMPUTACIONAL DE CEBADORES PARA INFLUENZA

La validación del filtro con los archivos antes indicados, se evidenció con las secuencias reportadas en el año 2018 para los tres genes, de las cuales se recuperó el 7% del total de secuencias (Tabla 1). Además, el archivo con las secuencias de genes sin nucleótidos y/o regiones indeterminadas, reportó un mayor número respecto a las que sí presentaron indeterminaciones. Es de resaltar que la depuración por defecto de los 12 meses del año seleccionado que realizó el Módulo, presentó el mismo patrón de resultados (Tabla 1).

4.2 Alineamientos de los genes *HA*, *NP* y *MI-M2* del virus de Influenza A H1N1 con el Módulo

2. Análisis de Secuencias Nucleotídicas (MASN-v1.0).

El Módulo MASN-v1.0 generó correctamente los alineamientos esperados con el programa MAFFT V.7 a partir de cada archivo de los genes *HA*, *NP* y *MI-M2* del virus de Influenza A H1N1 del año 2018 obtenidos en el módulo MGFS-v1.0. No se determinaron inconsistencias o regiones indeterminadas en los alineamientos de cada uno de los genes. Los alineamientos se visualizaron con otro programa público de edición y análisis de alineamientos de ácidos nucleicos: Bioedit (Figura 5).

A partir de estos alineamientos, se aplicó la primera regla de corte y se evidenció que todas las secuencias de los 3 genes reportados para el año 2018, presentaron uno o dos codones de inicio de la traducción AUG; así como un corte en el codón de paro identificado de acuerdo con la moda establecida para el total de secuencias, independientemente de si correspondía o no al codón de paro canónico. El resultado visualizado con el programa UGENE, corroboró la eliminación de las regiones no deseadas de los extremos 5' y 3' de las secuencias de los tres alineamientos, de acuerdo con la primera regla de corte (Figura 6, Apéndice D).

Además, los alineamientos generados con el módulo MASN-v1.0 se emplearon para validar la Segunda Regla de Corte, al eliminar regiones específicas seleccionadas por el usuario, la cual se validó mediante la visualización en UGENE (Figura 7).

ANÁLISIS COMPUTACIONAL DE CEBADORES PARA INFLUENZA

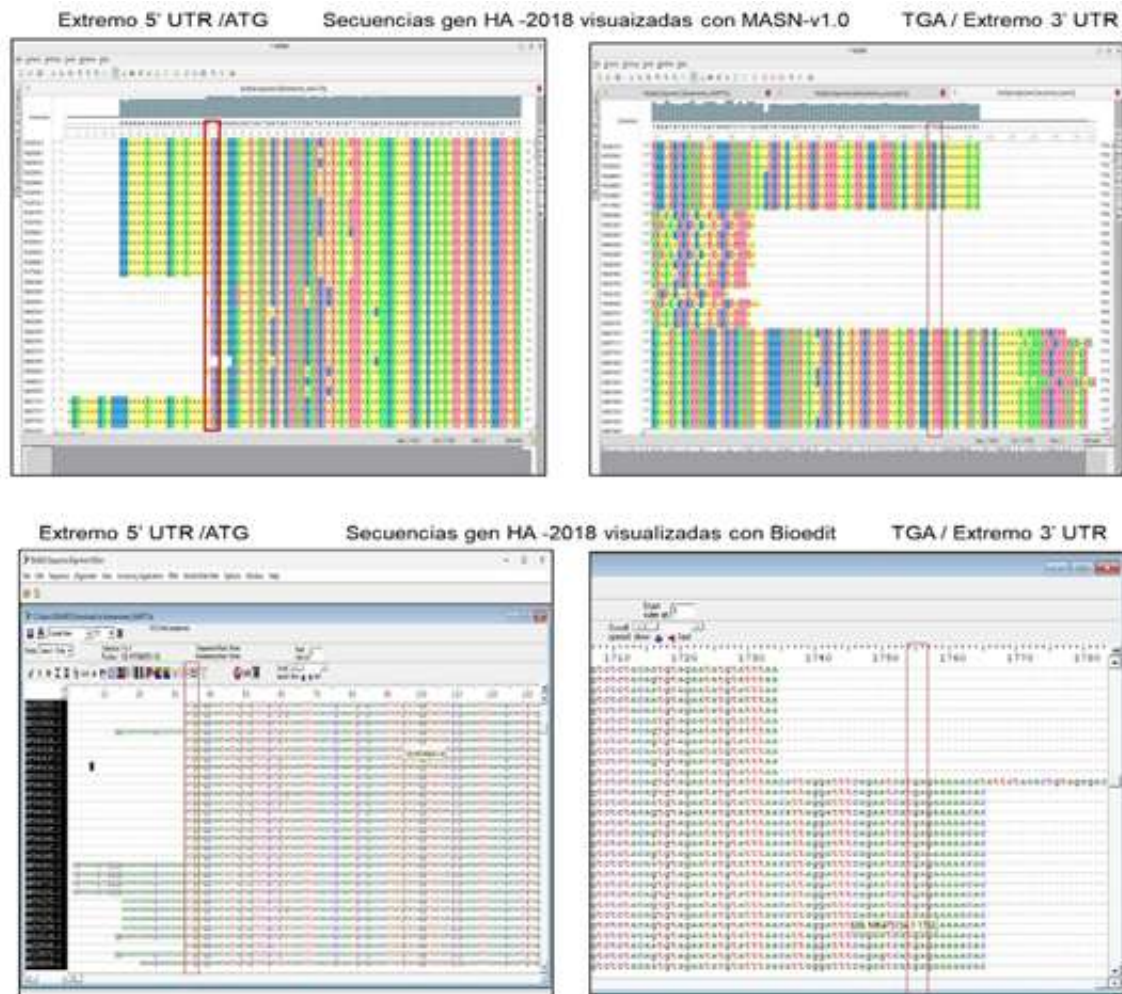
Tabla 1

Ejemplo de la cantidad de secuencias de los genes HA, NP y M1-M2 del virus de Influenza A H1N1 del año 2018.

| Gen de Influenza A H1N1 | Secuencias reportadas de 2009 a 2024 | Secuencias reportadas en el año 2018 con nucleótidos y regiones determinadas | Secuencias reportadas en el año 2018 con nucleótidos y regiones indeterminadas | Total de Secuencias del año 2018 |
|---------------------------------|--------------------------------------|--|--|----------------------------------|
| <i>Hemaglutinina</i> | 22,722 | 1,674 | 6 | 1680 |
| <i>Nucleoproteína</i> | 23,946 | 1674 | 3 | 1677 |
| <i>Proteína de Matriz M1-M2</i> | 23,946 | 1675 | 2 | 1677 |

Figura 5

Ejemplo de utilización de archivos de texto plano generados por el módulo MASN-v1.0



Nota: En la parte superior de la imagen se muestra la correcta ejecución del programa al colocar todas las secuencias seleccionadas de la base de datos NCBI Virus. En la parte inferior, el resultado es el mismo al utilizar el programa publico Bioedit. En cajas rojas se muestran los codones de inicio y paro de la traducción del ARNm del segmento 4 del virus de influenza reportadas para el año 2018.

ANÁLISIS COMPUTACIONAL DE CEBADORES PARA INFLUENZA

Figura 6

Validación de la primera Regla Corte con los archivos que contienen las secuencias obtenidas con el Módulo MASN-v1.0.

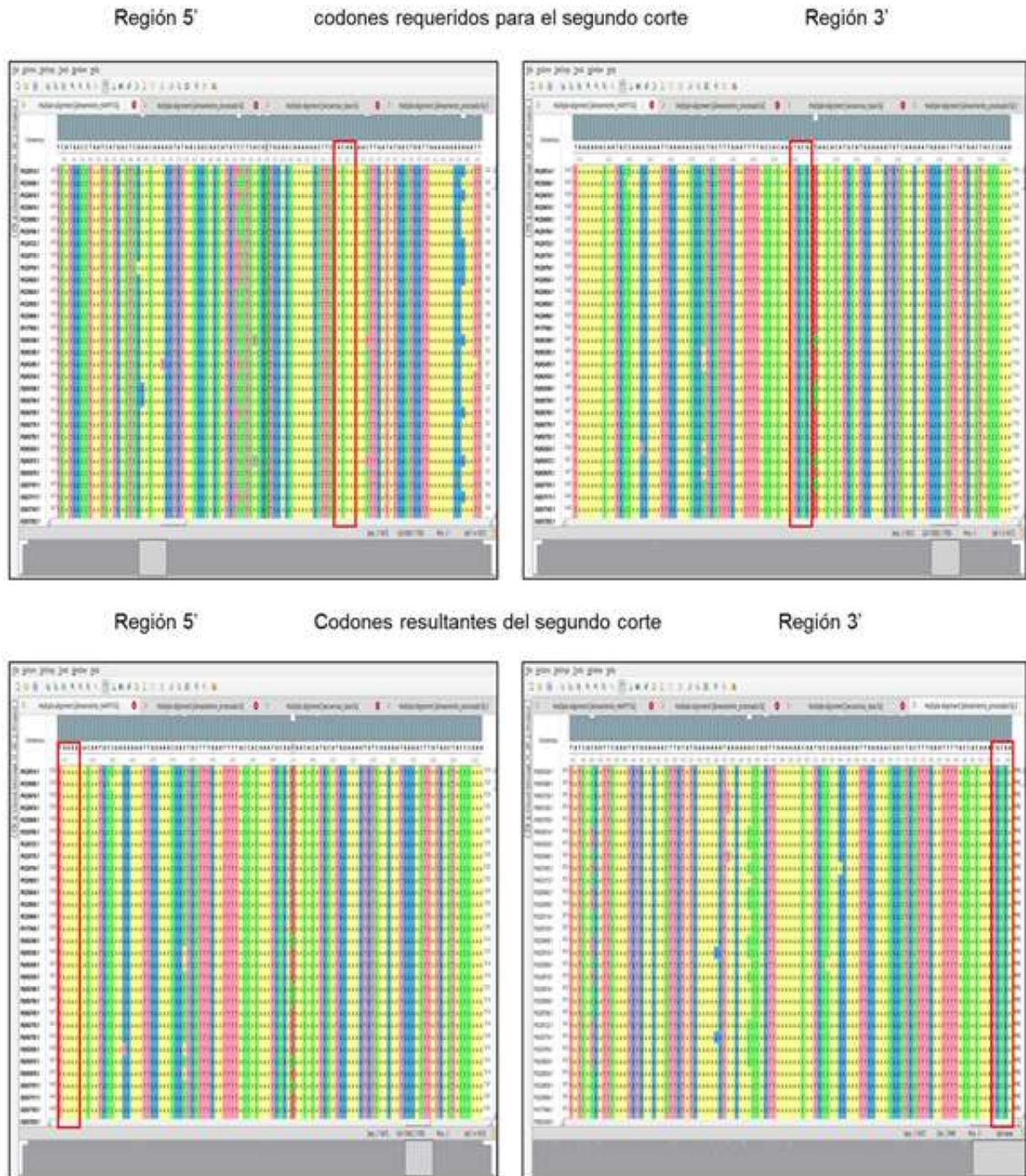


Nota: Al aplicar la primera regla de corte el modulo elimina los nucleótidos rio arriba y rio abajo de los codones de inicio y paro de la traducción (indicados en rectángulos rojos), lo que produce el documento deseado. El corte de la región 3' no necesariamente puede corresponder al reconocido por el ribosoma para detener la síntesis de la proteína. Las imágenes corresponden a visualización de una parte de los alineamientos del gen HA del virus de Influenza A H1N1 con el programa UGENE.

ANÁLISIS COMPUTACIONAL DE CEBADORES PARA INFLUENZA

Figura 7

Validación de la Segunda Regla Corte con el archivo con las secuencias alineadas con el Módulo MASN-v1.0.



Nota: Al aplicar la segunda regla de corte, el modulo elimina los nucleótidos rio arriba y rio abajo de las posiciones seleccionadas por el usuario, lo que produce el documento deseado. Las imágenes corresponden a la visualización de una parte de los alineamientos del gen HA del virus de Influenza A H1N1 con el programa UGENE.

4.3. Módulo 3. Evaluación de Cebadores y Sondas (MECS-v1.0)

La base de datos de cebadores y sondas generada por el grupo de investigación CAGE, reportó 23 juegos de cebadores y/o sondas para RT-PCR o RT-qPCR hasta el año 2018. Mientras que la búsqueda en la base de datos PubMed del año 2019 al 2024 con los criterios de búsqueda, generó un total de 94 publicaciones; de estas, 2 mostraron las secuencias de cebadores y sondas validados de forma experimental y los demás utilizaron kits comerciales u otros previamente reportados. No obstante, se incluyó un tercer juego para RT-PCR del gen *HA* de un artículo utilizado por el grupo CAGE del año 2020 (Apéndice E).

La información anterior se incluyó en el documento de texto plano requerido por el módulo MECS-v1.0 para identificar al mejor juego de cebadores y sondas, respecto a las secuencias consenso generadas con con el módulo MASN-v1.0, para los umbrales de frecuencia del 70 y 90% de los tres genes, desde el año 2009 al 2024. La aplicación de esta fase del protocolo para el gen *HA* para el año 2018, indicó el juego de cebadores reportados por Sharma et al., (2018), como los mejores de acuerdo con los consensos con el 70 y 97% de umbral, con un puntaje de identidad del 0.99 y 0.97 respectivamente (Figura 8).

ANÁLISIS COMPUTACIONAL DE CEBADORES PARA INFLUENZA

Los resultados anteriores fueron similares para las secuencias consenso de los genes *NP* y *M1-M2* del año 2018, con un umbral del 70%: el módulo MECS-v1.0 indicó que los mejores cebadores para el gen *NP* son los propuestos por González- Barrios et al., (2017) para RT-PCR, con un puntaje de identidad de 0.96; mientras que los indicados para el gen *M1-M2*, fueron los propuestos por Shin et al., (2011), con un puntaje de 0.78.

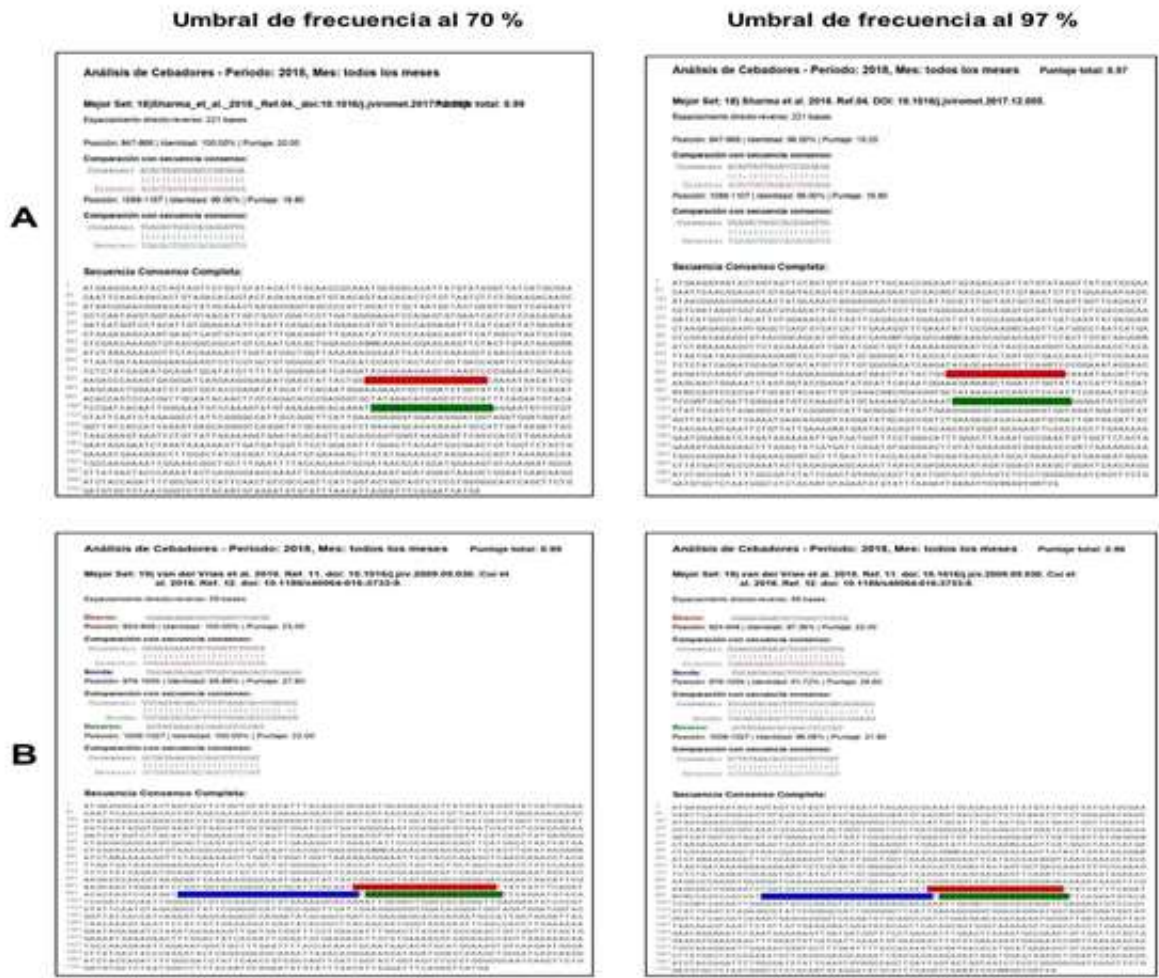
Sin embargo, a diferencia del anterior resultado, la secuencia consenso generada con el umbral de frecuencia del 70 y 97%, indicó que los cebadores y sondas ideales para el diseño de RT-qPCR para los genes *NP* y *M1-M2*, correspondieron a los propuestos por Gonzalez-Barrios et al en 2017, pero con un puntaje de 0.77 para ambos. Lo anterior se debió a la escasa homología entre el cebador sentido respecto a la secuencia consenso (Apéndice F y G). Los resultados de los años 2009 al 2024 para los genes *HA*, *NP* y *M1-M2* se encuentran en el Apéndice H.

El formato de salida de los resultados presentó la siguiente estructura: se indicó el juego de cebadores y/o sonda seleccionado, el autor y el DOI de la publicación donde se reporta la RT-PCR o RT-qPCR; el número de pares de bases entre el último nucleótido del cebador sentido respecto al primer nucleótido del cebador antisentido; así como la secuencia del cebador, la posición de la molécula en la secuencia consenso, el porcentaje de identidad y el Puntaje de similitud. La siguiente información que se mostró, fue la comparación de nucleótidos respecto a la secuencia consenso, en donde, las posiciones idénticas se indicaron con una línea vertical; las posiciones con dos opciones del código IUPAC, se indicaron con dos puntos; y las posiciones donde el carácter del código indica tres opciones, se mostró con un punto en la parte superior del nucleótido de la sonda. Este formato fue el mismo para el cebador sentido, la sonda y el antisentido, los cuales se representaron con letras rojas, azules y verdes, respectivamente. En la parte inferior se mostró la secuencia consenso y la posición de cada molécula en una caja con los colores antes mencionados (Figura 8).

ANÁLISIS COMPUTACIONAL DE CEBADORES PARA INFLUENZA

Figura 9

Comparación de resultados usando umbrales de consenso diferentes para segmento 4, usando solo conjuntos con sonda y usando todos los cebadores.



Nota: Los resultados en (A), corresponden a la evaluación realizada por el módulo MECS-v1.0 para el gen HA del virus de influenza A H1N1 con la secuencia consenso anual de 2018 con los umbrales de frecuencia del 70% y 97% y correlacionando la base de datos de los cebadores de RT-PCR y RT-qPCR. Los resultados en (B), corresponden a los parámetros de las secuencias consenso indicados en (A), pero la base de datos correspondió a cebadores y sondas para RT-qPCR.

Además, al utilizar este resultado, pero con los programas Primer 3, PrimerQuest y Primer BLAST no se generaron resultados en ninguno de los casos.

5. Discusión

El virus de la influenza A H1N1 representa una problemática de salud a nivel mundial (Neumann, Noda, y Kawaoka 2009), por lo cual se requiere la mejora e implementación de nuevas tecnologías para su identificación (World Health Organization 2019). Para esto, se han implementado diversas técnicas entre las que destacan la RT-PCR y la RT-qPCR. Sin embargo, la alta tasa de mutación del virus ha dificultado su identificación debido a los constantes cambios en los genes de referencia (Mänz et al. 2013), al igual que en otros virus como la variante Delta de SARS-CoV-2 (Rajib et al., 2022) o el virus de la rabia (Gigante et al., 2025). Para contrarrestar esto, en la comunidad científica se han reportado diversos conjuntos de cebadores y sondas para RT-PCR y RT-qPCR, los cuales hibridan sobre diferentes regiones del genoma del virus (Shu et al. 2021). A pesar de la variedad de conjuntos que se han reportado, no se ha investigado sobre la eficiencia de estos para identificar el virus en diferentes períodos de tiempo y de acuerdo con las mutaciones que se presentan.

La implementación del protocolo diseñado en esta pasantía permite realizar los análisis requeridos para determinar cebadores y sondas óptimos para la identificación del virus de influenza A H1N1. Mediante el primer módulo, es decir MGFS-v1.0, se pudieron filtrar las secuencias correspondientes al segmento 4 para el año 2018, así como en los intervalos de los años 2000 a 2008 y abril de 2009 a 2024. Anudado a lo anterior, se obtuvieron secuencias con regiones indeterminadas y secuencias sin estas regiones, lo cual es importante para la comparación de los cebadores y sondas, ya que esto asegura la integridad de las secuencias a emplear y optimiza la comparación de los conjuntos. Así mismo, la caracterización en un período de tiempo permite identificar variantes específicas, lo cual tiene aplicación no sólo en el virus de Influenza A H1N1 sino en otros de importancia médica como el virus del dengue o el SARS-CoV-2, y de igual forma en análisis de genética de poblaciones.

ANÁLISIS COMPUTACIONAL DE CEBADORES PARA INFLUENZA

Dada la naturaleza del genoma del virus de la Influenza A H1N1, se debe caracterizar correctamente la región a identificar y las variaciones que allí se encuentren. Mediante la implementación del módulo 2, MASN-v1.0, las secuencias alineadas con los parámetros de MAFFT V.7 son recortadas para reconocer los sitios de inicio y parada de la traducción, a partir del reconocimiento de los codones universales de inicio y parada. El poder realizar esta delimitación permite comparar los juegos de cebadores y sondas únicamente con las regiones codificantes del genoma viral. Sin embargo, los codones de inicio y parada pueden variar entre las secuencias, ya sea por efectos de la secuenciación o por las características mismas del virus, es decir, posibles variantes (Mänz et al. 2013). Este tipo de análisis se puede realizar a partir del alineamiento obtenido en el módulo 2, lo cual permite al usuario delimitar de manera manual las regiones a comparar. Este procedimiento optimiza la búsqueda de cebadores y sondas específicos ya que se pueden descartar regiones con mucha variación nucleotídica, lo cual genera inconvenientes en el diseño de las moléculas.

Es de resaltar que el protocolo realiza el recorte de secuencias de manera automática, esto con el fin de evitar las regiones 5' UTR y 3' UTR, ya que pueden generar ruido en el análisis y aumentar los requerimientos de cómputo. Aunque este procedimiento también se puede hacer manualmente, es importante tener un conocimiento previo de las moléculas a estudiar para evitar un recorte arbitrario que pueda generar resultados biológicamente incongruentes.

El diseño de cebadores y sondas para la identificación de ácidos nucleicos, ya sea por RT-PCR, RT-qPCR o cualquiera de las variaciones de la cadena en reacción de la polimerasa, requiere de un procedimiento sumamente cuidadoso para asegurar el éxito de la reacción, lo cual puede representar disminución en los costos económicos y de tiempo (Nolan 2013; Chen et al., 2025). Es por esto por lo que se debe asegurar que las secuencias de referencia sean de la mayor calidad posible, para que el diseño de cebadores y sondas esté acorde a lo requerido por el investigador. En este sentido, la fase final de la implementación del módulo permitió identificar correctamente

ANÁLISIS COMPUTACIONAL DE CEBADORES PARA INFLUENZA

el mejor conjunto de cebadores y sonda para la identificación del segmento 4 del virus de la Influenza A H1N1, los cuales correspondieron a los reportados por Poddar et al., (2002), desde el año 2000 hasta marzo de 2009.

Sin embargo, a partir de abril del año 2009 y hasta el 2024, el mejor conjunto fue el reportado por Sharma et al., (2018). Estos resultados sugieren que el genoma del virus pudo cambiar entre marzo y abril de 2009 como ha sido reportado para el patrón evolutivo de los genes *HA*, *NP* y *M1-M2* del virus de Influenza A H1N1 (González-Barrios et al., 2022), y del gen de la proteína no estructural (*NS*) del virus de la Influenza A en humanos (Nielsen et al., 2024); e incluso con el patrón mutacional del gen *HA* del virus de influenza A H1N1 que afectó la estructura cuaternaria de la proteína del año 2009 respecto a la del año 2018 (Pushan et al 2025). Por lo tanto, el nuevo juego de cebadores reportados por Sharma et al., (2018) que seleccionó el módulo es el óptimo para la identificación viral. Este resultado demuestra la correcta ejecución de los módulos y la efectividad del programa, al distinguir cambios en las secuencias en un período determinado de tiempo, lo cual representa una ventaja en la identificación del virus y su escalabilidad en otros tipos de análisis genéticos y moleculares.

Algunas limitaciones en cuanto al uso del sistema operativo son aparentes con las herramientas MAFFT V7 y UGENE. El primero sólo puede ejecutar sus funciones siempre y cuando su versión correspondiente, disponible en línea y diseñada para Sistema Operativo Windows 10/11, se encuentre junto al código a ejecutar. En distribuciones Linux no es tan restrictivo, dado que existe una versión especialmente diseñada para instalarse en este tipo de distribuciones. Sin embargo, en estas distribuciones es necesario compilar e instalar el paquete de UGENE para ejecutar mediante línea de comandos de algoritmo requerido (explicación detallada del proceso en el Apéndice A). Estas funciones, aunque limitadas, pueden ser ejecutadas desde equipos de baja capacidad y es posible escalar esta solución computacional para análisis mayores

ANÁLISIS COMPUTACIONAL DE CEBADORES PARA INFLUENZA

en HPC. Si este fuese el objetivo, en los parámetros se pueden configurar el número de núcleos a usar para el análisis.

Este código modular puede ser ejecutado con secuencias de otros organismos. Dependiendo de su tamaño, se puede usar un algoritmo específico a través de su edición en los parámetros. Por defecto, selecciona el algoritmo más adecuado de acuerdo a la cantidad de secuencias a alinear. Por otro lado, la lista de cebadores es determinante para el análisis por lo que se recomienda seguir el formato de cebadores implementado en este análisis antes de ejecutar las funciones del módulo. Es de resaltar que este protocolo, a diferencia de otros programas como Primer3 (Kõressaar et al., 2018) o PrimerQuest (IDT, <https://www.idtdna.com/pages/tools/primerquest>) que sugieren posibles cebadores y sondas, identifica el mejor conjunto para amplificar una molécula, por lo cual se puede implementar como una herramienta de apoyo durante la generación de nuevos cebadores y sondas, no sólo para el virus de la influenza, sino que también en otras moléculas.

La aplicación de cada módulo de manera independiente y los archivos obtenidos en cada paso, permiten que la implementación de este protocolo no se limite únicamente a la búsqueda de cebadores y sondas, sino que también permita al usuario la manipulación de secuencias para estudios de sistemática o genética de poblaciones. De igual forma, la flexibilidad al utilizar secuencias consenso mayoritarias o con nucleótidos degenerados, permite la identificación de secuencias específicas o, su uso en el diseño de oligonucleótidos degenerados con aplicación en metagenómica.

6. Conclusiones

El porcentaje de identidad proporcionado en este código muestra una idea más clara de los cebadores adecuados a usar al determinar si existe una alta probabilidad de obtener resultados falsos negativos. Dado que el código es modular, se pueden reestructurar de acuerdo a las necesidades o incluso reemplazarlos por procesos más óptimos. El diseño es flexible y operativamente es funcional en varios sistemas, por lo que escalarlo para análisis de muchas más secuencias, e incluso en paralelo, podría ser determinante para reducir tiempos de diagnóstico y costos prediciendo cual puede ser la mejor opción para diagnósticos RT-PCR.

Debido a su diseño en módulos, se podría reemplazar o ignorar el uso de uno cualquiera de los aquí presentados, mientras se conserve el formato de resultado para los módulos restantes. Su flexibilidad para operar en distintos sistemas operativos facilita que esta herramienta pueda ser usada por cualquier usuario, incluso si no se usa a gran escala en HPC, pudiendo aprovechar el uso de varios núcleos para el análisis. Puesto que las comparaciones de los cebadores con la secuencia consenso usan el código IUPAC, esta herramienta también podría operar con secuencias nucleotídicas degeneradas o de mayoría, acorde a las necesidades del usuario. En este trabajo se uso como base secuencias de influenza A H1N1 y cebadores para su amplificación reportados en la literatura, sin embargo estos algoritmos usados podrían ayudar no solo a reducir el número de falsos negativos que ocurren, sino a la elaboración de nuevos cebadores que eviten dichos falsos negativos tanto en el virus de influenza A y sus variantes como lo sería para otros organismos así como para facilitar la elaboración de nuevos cebadores para enfrentar nuevos desafíos para las técnicas de diagnóstico actuales.

Referencias Bibliográficas

- Apte, Arun, y Saurabha Daniel. 2009. «PCR Primer Design». *Cold Spring Harbor Protocols* 2009(3): pdb.ip65. doi:10.1101/pdb.ip65.
- Blanco-Lobo, Pilar, Aitor Nogales, Laura Rodríguez, y Luis Martínez-Sobrido. 2019. «Novel Approaches for The Development of Live Attenuated Influenza Vaccines». *Viruses* 11(2): 190. doi:10.3390/v11020190.
- Chen, Nan, Renxi Wang, Wanlu Zhu, Xiangjun Hao, Jing Wang, Guojiang Chen, Chunxia Qiao, et al. 2024. «Development and Characterization of an Antibody That Recognizes Influenza Virus N1 Neuraminidases» ed. Victor C. Huber. *PLOS ONE* 19(5): e0302865. doi:10.1371/journal.pone.0302865.
- Chen, Z., Yan, L., Liu, J., Zuo, W., Xu, Q., Qiao, S., Liu, S., Zheng, Y., Lin, H., Yang, L., Wang, B., Song, L., Li, T., Zhang, D., He, S., Ye, H., Zhang, J., Ge, S., Zhang, S., & Xia, N. (2025). An ultra-sensitive, multiplexed, and cost-effective POCT system for the detection of co-infecting respiratory viruses, including SARS-CoV-2, Flu A, Flu B, and RSV, within 30 min. *Journal of pharmaceutical and biomedical analysis*, 260, 116765. <https://doi.org/10.1016/j.jpba.2025.116765>.
- Er, J. C. (2025). Improving Influenza Nomenclature Based on Transmission Dynamics. *Viruses*, 17(5), 633. <https://doi.org/10.3390/v17050633>
- Garcia, Vanessa, y Stéphane Aris-Brosou. 2014. «Comparative Dynamics and Distribution of Influenza Drug Resistance Acquisition to Protein M2 and Neuraminidase Inhibitors». *Molecular Biology and Evolution* 31(2): 355-63. doi:10.1093/molbev/mst204.
- Gigante, C. M., Wicker, V., Condori, R. E., Wilkins, K., & Li, Y. (2025). Postmortem Diagnosis of Rabies in Animals by the Updated, Multiplexed LN34 Real-Time Reverse Transcription-Polymerase Chain Reaction Assay. *Journal of visualized experiments : JoVE*, (219), 10.3791/66761. <https://doi.org/10.3791/66761>.
- González-Barrios, J. A., Thompson-Bonilla, M. R., Martínez-Pérez, F., Barrios-Hernández, C. J., Bautista-Rozo, L. X., Rodríguez-Vázquez, R., & Martínez-Fong, D. (2022). Oligonucleótidos y proceso para detectar el virus de influenza A H1N1 (Patente No. MX/E/2016/029524). Instituto Mexicano de la Propiedad Intelectual.
- Harada, Daisuke, Ritsuo Nishiuchi, Yuka Iwasaki, Hirokazu Watanabe, Chiho Tokorodani, Akane Kanazawa, Hisako Kiguchi, et al. 2012. «Reliability of a Rapid Test for the Clinical Diagnosis of Influenza A/H1N1 2009». *Scandinavian Journal of Infectious Diseases* 44(10): 776-81. doi:10.3109/00365548.2012.686670.
- Hasan, Mohammad Rubayet, Sathyavathi Sundararaju, Chidambaram Manickam, Faheem Mirza, Hamad Al-Hail, Stephan Lorenz, y Patrick Tang. 2021. «A Novel Point Mutation in the N Gene of SARS-CoV-2 May Affect the Detection of the Virus by Reverse Transcription-Quantitative PCR» ed. Alexander J. McAdam. *Journal of Clinical Microbiology* 59(4): e03278-20. doi:10.1128/JCM.03278-20.

- Jones, Sara, Shijulal Nelson-Sathi, Yejun Wang, Raji Prasad, Sabrina Rayen, Vibhuti Nandel, Yueming Hu, et al. 2019. «Evolutionary, Genetic, Structural Characterization and Its Functional Implications for the Influenza A (H1N1) Infection Outbreak in India from 2009 to 2017». *Scientific Reports* 9(1): 14690. doi:10.1038/s41598-019-51097-w.
- Kudo, Eriko, Benjamin Israelow, Chantal B. F. Vogels, Peiwen Lu, Anne L. Wyllie, Maria Tokuyama, Arvind Venkataraman, et al. 2020. «Detection of SARS-CoV-2 RNA by Multiplex RT-qPCR» ed. Bill Sugden. *PLOS Biology* 18(10): e3000867. doi:10.1371/journal.pbio.3000867.
- Kumar, Deepshikha, Shobha Broor, y Maitreyi S. Rajala. 2016. «Interaction of Host Nucleolin with Influenza A Virus Nucleoprotein in the Early Phase of Infection Limits the Late Viral Gene Expression» ed. Kevin Harrod. *PLOS ONE* 11(10): e0164146. doi:10.1371/journal.pone.0164146.
- Köressaar, T., Lepamets, M., Kaplinski, L., Raime, K., Andreson, R., & Remm, M. (2018). Primer3_masker: integrating masking of template sequence with primer design software. *Bioinformatics* (Oxford, England), 34(11), 1937–1938. <https://doi.org/10.1093/bioinformatics/bty036>.
- Landry M. L. (2011). Diagnostic tests for influenza infection. *Current opinion in pediatrics*, 23(1), 91–97. <https://doi.org/10.1097/MOP.0b013e328341ebd9>
- Lee, S., Won, D., Kim, C. K., Ahn, J., Lee, Y., Na, H., Kim, Y. T., Lee, M. K., Choi, J. R., Lim, H. S., & Lee, K. R. (2021). Novel indel mutation in the N gene of SARS-CoV-2 clinical samples that were diagnosed positive in a commercial RT-PCR assay. *Virus research*, 297, 198398. <https://doi.org/10.1016/j.virusres.2021.198398>
- Majeed, Abdul, y Sungchang Lee. 2021. «Applications of Machine Learning and High-Performance Computing in the Era of COVID-19». *Applied System Innovation* 4(3): 40. doi:10.3390/asi4030040.
- Mänz, Benjamin, Dominik Dornfeld, Veronika Götz, Roland Zell, Petra Zimmermann, Otto Haller, Georg Kochs, y Martin Schwemmler. 2013. «Pandemic Influenza A Viruses Escape from Restriction by Human MxA through Adaptive Mutations in the Nucleoprotein» ed. Juan Ortin. *PLoS Pathogens* 9(3): e1003279. doi:10.1371/journal.ppat.1003279.
- Meuer, S., Wittwer, C., & Nakagawara, K. (Eds.). (2004). RT-PCR application manual (3rd ed.). Roche Applied Science. <https://gene-quantification.de/ras-pcr-application-manual-3rd-ed.pdf>
- Monamele, Chavely Gwladys, Hermann Landry Munshili Njifon, Marie-Astrid Vernet, Mohamadou Ripa Njankouo, Sebastien Kenmoe, Ali Ahmed Yahaya, Louis Deweerdt, et al. 2019. «Molecular Characterization of Influenza A(H1N1)Pdm09 in Cameroon during the 2014-2016 Influenza Seasons» ed. Mathilde Richard. *PLOS ONE* 14(1): e0210119. doi:10.1371/journal.pone.0210119.
- National Institute of Hygiene and Epidemiology, Hanoi, Viet Nam, Phuong Hoang, Hang Nguyen, Huong Tran, Thuy Nguyen, Anh Nguyen, Thanh Le, et al. 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* 10(1): 32-38. doi:10.5365/wpsar.2018.9.3.003.
- Neumann, Gabriele, Takeshi Noda, y Yoshihiro Kawaoka. 2009. «Emergence and Pandemic Potential of Swine-Origin H1N1 Influenza Virus». *Nature* 459(7249): 931-39. doi:10.1038/nature08157.
- Nielsen, B. F., Berrig, C., Grenfell, B. T., & Andreasen, V. (2024). One hundred years of influenza A evolution. *Theoretical population biology*, 159, 25–34. <https://doi.org/10.1016/j.tpb.2024.07.005>.
- Nolan, T. (Ed.). (2013). PCR technologies: A technical guide. Sigma-Aldrich. <https://www.sigmaaldrich.com/deepweb/assets/sigmaaldrich/marketing/global/documents/197/153/pcr-technologies-guide.pdf>.
- PrimerQuest PCR & QPCR Primer Design Tool | IDT. (n.d.). Integrated DNA Technologies. <https://www.idtdna.com/pages/tools/primerquest>.
- Pushan, S. S., Samantaray, M., Rajagopalan, M., & Ramaswamy, A. (2025). Structural dynamics of influenza A (H1N1) hemagglutinin protein: a comparative study of Indian (2018) isolate with its evolutionary neighbor, Californian (2009) strain. *Journal of biomolecular structure & dynamics*, 43(12), 6064–6077. <https://doi.org/10.1080/07391102.2024.2317985>.
- Rahman, Mozibur, Sheikh Ariful Hoque, Mohammad Ariful Islam, y Sabita Rezwana Rahman. 2017. «Molecular Analysis of Amantadine-Resistant Influenza A (H1N1 Pdm09) Virus Isolated from Slum Dwellers of Dhaka, Bangladesh». *Virus Genes* 53(3): 377-85. doi:10.1007/s11262-017-1447-x.
- Rajib, S. A., Ogi, Y., Hossain, M. B., Ikeda, T., Tanaka, E., Kawaguchi, T., & Satou, Y. (2022). A SARS-CoV-2 Delta variant containing mutation in the probe binding region used for RT-qPCR test in Japan exhibited atypical PCR amplification and might induce false negative result. *Journal of infection and chemotherapy : official journal of the Japan Society of Chemotherapy*, 28(5), 669–677. <https://doi.org/10.1016/j.jiac.2022.01.019>.
- Ravina, Paramjeet Singh Gill, Ashok Kumar, Jagriti Narang, Minakshi Prasad, y Hari Mohan. 2022. «Molecular Detection of H1N1 Virus by Conventional Reverse Transcription PCR Coupled with Nested PCR». *Sensors International* 3: 100178. doi:10.1016/j.sintl.2022.100178.
- Sajal, S. S. A., Islam, D. Z., Khandker, S. S., Solórzano-Ortiz, E., Fardoun, M., Ahmed, M. F., Jamiruddin, M. R., Azmuda, N., Mehta, M., Kumar, S., Haque, M., & Adnan, N. (2024). Strategies to Overcome Erroneous Outcomes in Reverse Transcription-Polymerase Chain Reaction (RT-PCR) Testing: Insights From the COVID-19 Pandemic. *Cureus*, 16(11), e72954. <https://doi.org/10.7759/cureus.72954>
- Sharma, Vikrant, Manisha Sharma, Divya Dhull, Sulochana Kaushik, y Samander Kaushik. 2019. «Phylogenetic Analysis of the Hemagglutinin Gene of Influenza A(H1N1)Pdm09 and A(H3N2) Virus Isolates from Haryana, India». *VirusDisease* 30(3): 336-43. doi:10.1007/s13337-019-00532-7.

- Shu, Bo, Marie K. Kirby, William G. Davis, Christine Warnes, Jimma Liddell, Ji Liu, Kai-Hui Wu, et al. 2021. «Multiplex Real-Time Reverse Transcription PCR for Influenza A Virus, Influenza B Virus, and Severe Acute Respiratory Syndrome Coronavirus 2». *Emerging Infectious Diseases* 27(7): 1821-30. doi:10.3201/eid2707.210462.
- Tahan, Stephen, Bijal A. Parikh, Lindsay Droit, Meghan A. Wallace, Carey-Ann D. Burnham, y David Wang. 2021. «SARS-CoV-2 E Gene Variant Alters Analytical Sensitivity Characteristics of Viral Detection Using a Commercial Reverse Transcription-PCR Assay» ed. Yi-Wei Tang. *Journal of Clinical Microbiology* 59(7): e00075-21. doi:10.1128/JCM.00075-21.
- Wahlgren J. Influenza A viruses: an ecology review. *Infect Ecol Epidemiol.* 2011;1. doi: 10.3402/iee.v1i0.6004. Epub 2011 Feb 11. PMID: 22957113; PMCID: PMC3426330.
- Wang, Ningning, Wei Lu, Li Yan, Mengru Liu, Feihu Che, Yue Wang, Chunli Yang, et al. 2024. «Epidemiological and Genetic Characterization of the Influenza A (H1N1) Virus in Hangzhou City in 2023». *Frontiers in Public Health* 12: 1464435. doi:10.3389/fpubh.2024.1464435.
- World Health Organization. 2019. *Global influenza strategy 2019-2030*. . técnico. <https://iris.who.int/handle/10665/311184>.
- Zhao, Y., Guo, S., Liu, J., Wang, Y., Wang, B., Peng, C., & Du, E. (2025). Adjuvant-free, self-assembling ferritin nanoparticle vaccine coupled with influenza virus hemagglutinin protein carrying M1 and PADRE epitopes elicits cross-protective immune responses. *Frontiers in immunology*, 16, 1519866. <https://doi.org/10.3389/fimmu.2025.1519866>
- Zhu, Na, Dingyu Zhang, Wenling Wang, Xingwang Li, Bo Yang, Jingdong Song, Xiang Zhao, et al. 2020. «A Novel Coronavirus from Patients with Pneumonia in China, 2019». *New England Journal of Medicine* 382(8): 727-33. doi:10.1056/NEJMoa2001017.