



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
MAESTRÍA Y DOCTORADO EN CIENCIAS DE LA PRODUCCIÓN Y SALUD ANIMAL

DESARROLLO DE ESTRATEGIAS DE SECUENCIACIÓN Y ANÁLISIS DE GENOMAS
COMPLETOS DE VIRUS ANIMALES

TESIS

QUE PARA OPTAR POR EL GRADO DE:

Maestro en Ciencias

PRESENTA:

RONALD RANDY MUÑOZ MEJÍA

DR. ROGELIO ALEJANDRO ALONSO MORALES

Facultad de Medicina Veterinaria y Zootecnia

DR. JOSÉ IVÁN SÁNCHEZ BETANCOURT

Facultad de Medicina Veterinaria y Zootecnia

DRA. CLAUDIA RANGEL ESCAREÑO

Instituto Nacional de Medicina Genómica



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas

Tesis Digitales

Restricciones de uso

DERECHOS RESERVADOS ©

PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

CONTENIDO

1. INTRODUCCIÓN.....	8
1.1. Secuenciación del ácido desoxirribonucleico	8
1.1.1. Ácido desoxirribonucleico	8
1.1.2. Tecnologías de secuenciación	9
1.1.3. Métodos de secuenciación de primera generación.....	12
1.1.3.1. Secuenciación de Sanger.....	12
1.1.3.2. Secuenciación de Maxam y Gilbert	12
1.1.3.3. Secuenciación fluorescente.....	12
1.1.4. Métodos de secuenciación masiva (NGS).....	13
1.1.4.1. Pirosecuenciación	13
1.1.4.2. Secuenciación por hibridación.....	13
1.1.4.3. Secuenciación por ligación.....	14
1.1.4.4. Secuenciación por semiconductor	14
1.2. Métodos de ensamblado	15
1.2.1. Ensamblado con secuencia de referencia	15
1.2.2. Estrategias para el ensamblado <i>de novo</i>	15
1.2.2.1. Overlap Layout Consensus (OLC).....	15
1.2.2.2. Gráficas de Bruijn (DBG).....	16
1.2.2.3. Greedy Assmebly Algorithms	16
1.3. Virus de la influenza	16
1.3.1. Taxonomía y clasificación	16
1.3.2. Características morfológicas del virus	17
1.3.3. Organización del genoma del virus de influenza tipo A.....	17
1.3.4. Ecología del virus de influenza tipo A	18
1.3.5. Evolución molecular de los virus de influenza tipo A	19
1.3.6. Variaciones antigénicas de los virus de la influenza tipo A	20
1.3.6.1. Deriva antigénica	20
1.3.6.2. Cambio antigénico	21
1.3.7. Epidemiología de los virus de la influenza tipo A.....	22
1.4. Influenza aviar.....	22

1.4.1.	Virus de influenza aviar de alta patogenicidad (IAAP)	23
1.4.2.	Virus de influenza aviar de baja patogenicidad (IABP)	23
1.4.3.	Importancia en salud pública.....	23
1.4.4.	Influenza aviar en México.....	24
1.5.	Influenza porcina.....	24
1.5.1.	Importancia en salud pública.....	24
1.5.2.	Subtipos de influenza tipo A circulantes en cerdos.....	25
1.6.	Virus de Newcastle	28
1.6.1.	Taxonomía y clasificación	28
1.6.2.	Características y genoma.....	28
2.	JUSTIFICACIÓN.....	29
3.	HIPÓTESIS.....	30
4.	OBJETIVO GENERAL	30
5.	OBJETIVOS ESPECÍFICOS.....	30
6.	ESTRATEGIA EXPERIMENTAL.....	31
7.	METODOLOGÍA DEL ENSAMBLADO DE GENOMAS	32
8.	MATERIALES Y MÉTODOS.....	33
8.1.	Muestras.....	33
8.2.	Preparación de las bibliotecas para secuenciación	35
8.3.	Evaluación de la calidad de las lecturas	36
8.4.	Ensamblado del genoma.....	37
8.5.	Análisis bioinformático de las secuencias del virus de influenza aviar A/H5N2.....	38
8.6.	Alineamientos	43
8.7.	Análisis filogenético del virus de influenza aviar A/H5N2.....	44
8.8.	Análisis de recombinación genómica del virus de influenza aviar A/H5N2.....	45
9.	RESULTADOS.....	46
9.1.	Secuenciación.....	46
9.2.	Calidad de las lecturas	49
9.3.	Ensamblado del genoma.....	52
9.4.	Alineamiento de las secuencias del virus de influenza aviar A/H5N2	55
9.5.	Análisis filogenético de las secuencias del virus de influenza aviar A/H5N2.....	60

9.6.	Análisis de recombinación genómica	70
10.	DISCUSIÓN.....	78
11.	CONCLUSIONES	82
12.	REFERENCIAS	84

ÍNDICE DE FIGURAS

Figura 1	20
Figura 2.....	36
Figura 3.....	50
Figura 4.....	51
Figura 5.....	52
Figura 6.....	53
Figura 7.....	53
Figura 8.....	57
Figura 9.....	57
Figura 10.....	59
Figura 11.....	61
Figura 12.....	62
Figura 13.....	63
Figura 14.....	64
Figura 15.....	65
Figura 16.....	66
Figura 17.....	67
Figura 18.....	68
Figura 19.....	69
Figura 20.....	70
Figura 21.....	71
Figura 22.....	72
Figura 23.....	73
Figura 24.....	74

Figura 25.....	75
Figura 26.....	76
Figura 27.....	77

ÍNDICE DE CUADROS

Cuadro 1.....	10
Cuadro 2.....	11
Cuadro 3.....	18
Cuadro 4.....	26
Cuadro 5.....	34
Cuadro 6.....	35
Cuadro 7.....	38
Cuadro 8.....	39
Cuadro 9.....	42
Cuadro 10.....	42
Cuadro 11.....	44
Cuadro 12.....	46
Cuadro 13.....	47
Cuadro 14.....	47
Cuadro 15.....	48
Cuadro 16.....	49
Cuadro 17.....	55
Cuadro 18.....	56
Cuadro 19.....	58

1. INTRODUCCIÓN

1.1. Secuenciación del ácido desoxirribonucleico

1.1.1. Ácido desoxirribonucleico

El ácido desoxirribonucleico (ADN) fue aislado por primera vez por el médico Friedrich Miescher en 1869, mientras realizaba trabajos para determinar la composición química de material purulento en vendas quirúrgicas, al cual llamó nucleína ya que había sido extraído de núcleos celulares¹. Fue hasta 1919 que Phoebus Levene identificó los componentes del ADN, al describir que se encontraba formado por una base nitrogenada, un azúcar y un fosfato². Posteriormente en 1930 el mismo Levene en conjunto con Albrecht Kossel demostraron que la nucleína en un inicio descrita por Miescher era un ácido desoxirribonucleico formado por 4 bases nitrogenadas, un azúcar desoxirribosa y un grupo fosfato, el cual pensaban era una cadena corta donde las bases se repetían secuencialmente. En 1928 Frederick Griffith propuso el principio transformante al notar una transformación entre cepas bacterianas de *Streptococcus pneumoniae* por medio de transferencia de una sustancia activa. Fue hasta 1944 que Oswald Avery, Colin MacLeod y Maclyn McCarty determinaron que el responsable del principio transformante era el ADN³.

En 1940 todavía se tenían reservas en aceptar al ADN como el material genético, principalmente debido a su simplicidad química, ya que era descrito como un polímero largo compuesto por 4 sub unidades distintas, las cuales se encontraban unidas químicamente. Finalmente en 1952 se logró elucidar el papel del ADN en la herencia genética, esto debido al trabajo de Alfred Hershey y Martha Chase con el fago T2, en el que comprobaron que el fago transmitía su información genética en su ADN y no en sus proteínas⁴.

En 1953 James Watson y Francis Crick, apoyados en los datos de difracción de rayos X proporcionados por Rosalind Franklin, proponen el modelo de la doble hélice del ADN⁵ por el cual reciben el premio Nobel en 1962 junto con Maurice Wilkins⁶.

1.1.2. Tecnologías de secuenciación

La secuenciación del ADN tiene como finalidad determinar el orden de la secuencia de nucleótidos de una molécula de ADN, ya sea de una región o de la molécula completa. Al determinar el orden de nucleótidos de la molécula completa se conoce el genoma de un organismo, el cual determina su información genética que es heredada de generación en generación²³.

La secuenciación de ácidos nucleicos inicia en 1976 cuando Walter Fiers logra secuenciar el primer genoma completo de un organismo, el del bacteriófago MS2, el cual consiste de aproximadamente 3500 nucleótidos (nt)⁷, pero es hasta el siguiente año (1977), cuando surgen dos técnicas modernas de secuenciación de ADN; la secuenciación química de Maxam y Gilbert⁸ y la secuenciación enzimática de Frederick Sanger⁹, quien obtiene la secuencia completa del genoma del bacteriófago **Φ x174** representado por 5,375 nt.

Algunos de los avances científicos que permitieron un desarrollo en los métodos automatizados de secuenciación de ácidos nucleicos fue el desarrollo de la técnica de la reacción en cadena de la polimerasa en 1985 por Kary Mullis¹⁰ y el aislamiento de la polimerasa termoestable de la bacteria *Thermus aquaticus* en 1988 por Innis¹¹. En 1986, Smith reporta la primera técnica de secuenciación automatizada, la cual permitía obtener información de hasta 200 nt¹².

Durante la última década se han desarrollado nuevas tecnologías de secuenciación automática, denominadas NGS (de sus siglas en inglés *Next Generation Sequencing*). A diferencia de la técnica de secuenciación por Sanger, las NGS permiten obtener millones de secuencias de ADN, lo que ha

permitido disminuir los tiempos y costos de secuenciación¹³, no obstante, los distintos equipos ofrecen limitantes particulares de cada tecnología, lo que ha resultado en un reto para el análisis bioinformático de los datos (Cuadro 1).

Cuadro 1. Rendimiento de las principales tecnologías de secuenciación masiva en el mercado. A manera de resumen se resalta las limitantes más importantes de cada una de las tecnologías. Una de las principales diferencias entre los equipos es el tamaño de las lecturas generadas¹⁴.

MÉTODO	LONGITUD DE LECTURAS	LIMITANTES
Sanger	1000	Bajo rendimiento, de baja calidad en los dos extremos
Roche 454®	200-700	Errores en homo-polímeros
Illumina®	150	Baja calidad en el extremo 3 '
SOLID®	75	Difícil de trabajar y lecturas cortas
IonTorrent®	100-400	Alta tasa de error, error en homo-polímeros
PacBio®	>9000	Muy alta tasa de error, puede iterar

Dentro de las principales tecnologías NGS desarrolladas, se encuentra la secuenciación por hibridación de Illumina®¹⁵, cuya principal ventaja es la cantidad de información generada y como principal limitante el tamaño de las lecturas, así como el Ion Personal Genome Machine (PGM) de Life Technologies™¹⁶, el cual ofrece velocidad y bajo costo de la reacción, sin embargo tiene poca

precisión y baja cobertura. En el cuadro 2 se muestran las ventajas y limitaciones que ofrecen las principales tecnologías de secuenciación.

Cuadro 2. Ventajas y limitaciones de las principales tecnologías de secuenciación masiva. Se detalla características de los 4 equipos más utilizados en procesos de secuenciación.

TECNOLOGÍA	VENTAJAS	LIMITACIONES
454 Roche® Pirosecuenciación	<ul style="list-style-type: none"> • Tiempo de secuenciación (~10 horas por corrida) • Alto rendimiento; 400-600 mega bases/experimento • Alta certeza 	<ul style="list-style-type: none"> • Lecturas cortas (250-500 pb) • Dificultades secuenciando homopolímeros
Illumina® Secuenciación por síntesis	<ul style="list-style-type: none"> • Bajo costo • Alto rendimiento; hasta 200 gigabases/ experimento • Alta certeza • No hay problemas con homopolímeros. 	<ul style="list-style-type: none"> • Lecturas cortas (~100 pb en 2010) • Tiempo de secuenciación (2-11 días)
Solid® Secuenciación por ligación	<ul style="list-style-type: none"> • Bajo costo • Alto rendimiento; hasta 300 Giga bases/experimento • Alta certeza • No hay problemas con homopolímeros. 	<ul style="list-style-type: none"> • Lecturas cortas (~75 pb) • Tiempo de secuenciación (4-8 días)
Ion torrent® Semiconductor	<ul style="list-style-type: none"> • Tiempo de secuenciación (~2 días por corrida) • Bajo costo y Alta certeza 	<ul style="list-style-type: none"> • Corto rendimiento y Baja calidad en lecturas • Dificultades secuenciando homopolímeros

1.1.3. Métodos de secuenciación de primera generación

1.1.3.1. Secuenciación de Sanger

El método de secuenciación de Sanger (secuenciación enzimática) fue desarrollado por Frederick Sanger en 1975 y es la técnica más utilizada actualmente. Este método consiste en la síntesis de una cadena complementaria de ADN a partir de una cadena sencilla de ADN que sirve como molde, utilizando nucleótidos modificados químicamente. Estos nucleótidos tienen reemplazado el grupo hidroxilo en el extremo 3' por un átomo de hidrógeno, esto evita la elongación de la cadena de ADN ya que el grupo hidroxilo es necesario para la formación de los puentes fosfodiéster entre los nucleótidos. Al final se obtienen fragmentos de ADN de distintos tamaños que terminan con un dideoxinucleótido, los cuales pueden ser visualizados en una película de rayos X, utilizando moléculas dATP marcadas con fósforo 32 ¹⁷.

1.1.3.2. Secuenciación de Maxam y Gilbert

Este método también es conocido como secuenciación química y fue publicado en 1977, el cual requiere marcar radioactivamente el extremo 5' de la molécula de ADN, la cadena de doble hebra de ADN es tratada químicamente para generar secuencias de uno o dos nucleótidos (G, A+G, C, C+T) para obtener fragmentos marcados que pueden ser separados por tamaño mediante electroforesis para posteriormente ser visualizados en una película de rayos X¹⁸.

1.1.3.3. Secuenciación fluorescente

El método de secuenciación fluorescente fue publicado en 1986 por Leroy Hood. En este método la molécula de ADN es marcada con un *primer* etiquetado fluorescentemente, utilizando un fluoróforo distinto para cada uno de los cuatro nucleótidos. A diferencia de la secuenciación enzimática y la secuenciación química que utilizan exposición a rayos X, este método detecta los nucleótidos mediante un láser que excita el fluoróforo con el que es marcado el nucleótido¹⁹.

1.1.4. Métodos de secuenciación masiva (NGS)

1.1.4.1. Pirosecuenciación

Este método fue desarrollado en 1996 por Mostafa Ronaghi y Pal Nyrén y fue la primera tecnología de secuenciación masiva en comercializarse. La pirosecuenciación se basa en detectar el pirofosfato (PPi) liberado con la incorporación de nucleótidos durante la replicación del ADN, la enzima sulfurilasa convierte el PPi en Adenosín Tri Fosfato (ATP), el cual es utilizado por la enzima luciferasa para emitir fotones²⁰, el sistema 454 de Roche© asocia el tiempo de la emisión de luz con la incorporación de los nucleótidos para inferir cual base es la que está siendo incorporada²¹.

1.1.4.2. Secuenciación por hibridación

El sistema Illumina™ utiliza la tecnología de secuenciación de Solexa. Debido a que las lecturas obtenidas con Illumina™ son más cortas que las obtenidas con el equipo 454 de Roche©, no puede resolver repeticiones cortas, sin embargo el costo por base es más bajo y el rendimiento es mayor. La secuenciación por hibridación consta de dos partes, en primer lugar se realiza un fraccionamiento del ADN, seguido de la amplificación clonal de fragmentos de ADN sobre la superficie sólida de una celda de flujo. De esta forma se crean racimos de fragmentos clonados. El siguiente paso consiste en la secuenciación mediante síntesis. El sistema se basa en el empleo de los cuatro nucleótidos (A, C, T, G) marcados con fluorescencia de distintos colores para la secuenciación de los millones de racimos presentes en la superficie de la celda, estos nucleótidos modificados contienen terminaciones reversibles, lo que permite que cada ciclo del proceso de secuenciación ocurra en presencia de los cuatro nucleótidos simultáneamente²².

1.1.4.3. Secuenciación por ligación

La secuenciación SOLiD (*Sequencing by Oligonucleotide Ligation Detection*), también conocida como método de codificación de dos bases, ya que utiliza un código de color de dos bases para diferenciar entre polimorfismos (SNP) y errores en la secuenciación, fue implementada por Applied Biosystems® en 2007. Las lecturas obtenidas con este equipo son más cortas que las obtenidas con Illumina™, sin embargo el rendimiento es mayor²³. El método SOLiD consiste en una amplificación por emulsión de PCR, seguido por la creación de la biblioteca de secuencias de ADN para ligarlas con el ADN amplificado por medio de una ADN ligasa. En la secuenciación por ligación cada fluoróforo es asociado a secuencias de dos nucleótidos (AT, CG, etc.)²⁴.

1.1.4.4. Secuenciación por semiconductor

El Ion Personal Genome Machine (PGM) de la empresa Life Technologies™ es un secuenciador a base de chips de iones. Este equipo utiliza la tecnología de secuenciación a base del uso de los ISFET (*Ion Sensible Field Electrical Transistor*: transistor de campo eléctrico sensible a iones) dentro de un circuito integrado del tipo CMOS (*Complementary metal-oxide-semiconductor*: Semiconductor Complementario Metal-Óxido). El proceso general comprende cargar la muestra de ADN y anclarla a unas pequeñas esferas de acrilamida en el chip, posteriormente se añaden los di nucleótidos trifosfatados (dNTP). Cada vez que uno de los nucleótidos encuentre su complementario en la hebra molde se formará el enlace y empezará a sintetizarse la nueva cadena, por cada enlace la reacción libera un ion de Hidrógeno (H⁺), el cual cambiará ligeramente el pH y es detectado por los sensores del semiconductor para producir una señal eléctrica en función a la diferencia de potencial generado (ΔV). Los componentes electrónicos transformarán esta señal eléctrica en una digital y así se procede a leer las millones de secuencias generadas. Todo el proceso toma aproximadamente 4 segundos y la precisión de secuenciación es superior al de los métodos tradicionales basados en luminiscencia¹⁶.

1.2. Métodos de ensamblado

Existen dos enfoques para el ensamblado de un genoma, uno es el ensamblado con secuencia de referencia y el segundo es el ensamblado *de novo*.

1.2.1. Ensamblado con secuencia de referencia

En la estrategia de ensamblado con referencia se utiliza el genoma del organismo más cercanamente relacionado como guía para comparar las lecturas obtenidas de la secuenciación, este método requiere menos capacidad de cómputo y es más rápido comparado con el enfoque de ensamblado *de novo*, sin embargo, puede introducir sesgos y resulta apropiado solamente cuando se dispone con un genoma de referencia²⁵.

1.2.2. Estrategias para el ensamblado *de novo*

El ensamblado *de novo* está dirigido a la reconstrucción de genomas en los que no se cuenta con una secuencia de referencia, para este enfoque existen tres estrategias: *Overlap-Layout-Consensus* (OLC), gráficas de Brujin (DBG) y el método *Greedy Assmebly Algorithms* que puede utilizar OLC o DBG.

1.2.2.1. Overlap Layout Consensus (OLC)

Este algoritmo consiste de 3 pasos principales: sobrelapamiento, diseño y consenso. En el sobrelapamiento se construye un gráfico básico en el que todas las lecturas son conectadas con todas, indicando donde existen sobrelapamientos. En la etapa de diseño se comprime el gráfico para generar contiguos, en base a la ruta más óptima y con mejor calificación, y finalmente la etapa de consenso determina el ensamblado basado en el gráfico generado en los dos pasos anteriores^{52 53}.

1.2.2.2. Gráficas de Brujin (DBG)

Es el método más nuevo y se basa en la teoría de grafos de Euler. A diferencia del OLC que utiliza caminos de Hamilton, de los cuales no se han podido desarrollar algoritmos eficientes, existen algoritmos eficientes para determinar caminos eulorianos⁵². El algoritmo inicia calculando palabras de tamaño k para poder construir el gráfico, cada palabra es almacenada en memoria una vez, evitando la redundancia de los otros métodos, y haciendo uso eficiente de esta, a través de tablas de hash⁵³.

1.2.2.3. Greedy Assmebly Algorithms

Éste fue el primer algoritmo para abordar la estrategia de ensamblado *de novo*. El primer paso del algoritmo consiste en el cálculo del alineamiento pareado para todas las lecturas, estos alineamientos son ponderados basados en la longitud del solapamiento y el porcentaje de bases que alinean²⁶. Las lecturas con la puntuación más alta son unidas en una sola para generar un contiguo, este paso se repite hasta que no se puedan generar más contiguos²⁷.

1.3. Virus de la influenza

1.3.1. Taxonomía y clasificación

El virus de la influenza pertenece a la familia *Orthomixoviridae*, la cual consta de 6 géneros: *Influenzavirus A*, *Influenzavirus B*, *Influenzavirus C*, *Isavirus*, *Quaranjavirus* y *Thogotovirus*²⁸. Se clasifica en 3 tipos: A, B y C. Únicamente los virus tipo A se clasifican en subtipos, basados en las propiedades antigénicas de las glicoproteínas de superficie hemaglutinina (HA) y neuraminidasa (NA)²⁹.

1.3.2. Características morfológicas del virus

Los viriones presentan forma esférica o filamentosa con prolongaciones superficiales que contienen las glicoproteínas de superficie hemaglutinina (HA) y neuraminidasa (NA), su tamaño varía entre 100 y 300 nanómetros (nm) y su morfología es de simetría helicoidal²⁹.

1.3.3. Organización del genoma del virus de influenza tipo A

Los virus de la influenza tienen un genoma ARN segmentado de cadena sencilla de polaridad negativa (-ssARN). Los virus de la influenza tipo A constan de 8 segmentos que codifican para 11 proteínas³⁰, los segmentos 1, 3, 4, 5 y 6 codifican cada uno para una proteína diferente, la proteína polimerasa básica 2 (PB2), la proteína polimerasa ácida (PA), la proteína HA, la proteína nucleoproteína (NP) y la proteína NA respectivamente. Mientras que los segmentos 2, 7 y 8 codifican cada uno para 2 proteínas, la polimerasa básica 1 (PB1) y la proteína de fusión (PB1-F2), la proteína de matriz 1 (M1) y la proteína de matriz 2 (M2) y la proteína no estructural 1 (NS1) y la proteína no estructural 2 (NS2) respectivamente^{31,32} (Cuadro 3).

Cuadro 3. Organización del genoma y proteínas del virus de influenza tipo A. Tamaño y producto de los 8 segmentos del genoma del virus de influenza A. Los segmentos 7 y 8 codifican para 2 proteínas, el resto de los segmentos codifica para 1 proteína.

SEGMENTO	TAMAÑO	GEN	PRODUCTO
1	2341	PB2	Transcriptasa (unión a cap)
2	2341	PB1	Transcriptasa (elongación)
	237	PB1-F2	Promueve apoptosis
3	2233	PA	Transcriptasa (actividad proteasa)
4	1778	HA	Hemaglutinina
5	1565	NP	Nucleoproteína
6	1467	NA	Neuraminidasa
7	1027	M1	Proteína de matriz
		M2	Proteína integral de membrana
8	890	NS1	No estructural 1
		NS2	No estructural 2

1.3.4. Ecología del virus de influenza tipo A

En base a su historia natural, el virus de la influenza tipo A tiene como hospedero natural a las aves acuáticas silvestres, sin embargo, también puede infectar humanos, cerdos, caballos, mamíferos marinos, perros, gatos²⁹ y murciélagos³³. El receptor celular que es utilizado por el virus para su replicación difiere entre aves y mamíferos, los virus que infectan aves se unen a receptores de ácido **siálico con enlaces α 2,3-Galactosa** mientras que los virus que infectan mamíferos se unen a **receptores de ácido siálico con enlaces α 2,6-Galactosa**, sin embargo ambos receptores pueden hallarse en el tracto respiratorio del cerdo y del ser humano. A diferencia del tracto respiratorio

humano que presenta receptores $\alpha 2,3$ -Galactosa en el tracto respiratorio bajo y $\alpha 2,6$ -Galactosa en el tracto respiratorio alto, el cerdo presenta ambos receptores en el tracto respiratorio alto³⁴, haciéndolo susceptible a ser infectado tanto por cepas virales humanas como por aviares, pudiendo actuar como vehículos de mezcla y facilitar la transmisión interespecie³⁵ (Figura 1).

Previo a la identificación de los subtipos H17N10 y H18N11 del virus de la influenza en murciélagos se habían identificado 16 HA y 9 NA circulando en reservorios aviares³⁶. Los virus de la influenza B y C únicamente causan enfermedades leves en humanos³⁷, aunque existen reportes donde se ha identificado a la foca como susceptible de ser infectado por el virus de la influenza B³⁸.

1.3.5. Evolución molecular de los virus de influenza tipo A

Un proceso importante en la dinámica evolutiva de estos virus es el evento de reordenamiento genético, que es el resultado de un intercambio de segmentos genómicos virales cuando una célula es infectada por cepas de virus humanos y animales, el virus emplea este mecanismo como un proceso de adaptación para infectar un nuevo huésped (Figura 1)³⁹.

La evolución de los virus de influenza también es ocasionada por la alta tasa de mutación en las proteínas de superficie HA y NA, principalmente ocasionada por la falta de actividad de corrección de errores de la RNA polimerasa viral³², la presión de selección ocasionada por la evasión de la respuesta inmune del hospedador y la ocasionada por la resistencia a antivirales³².

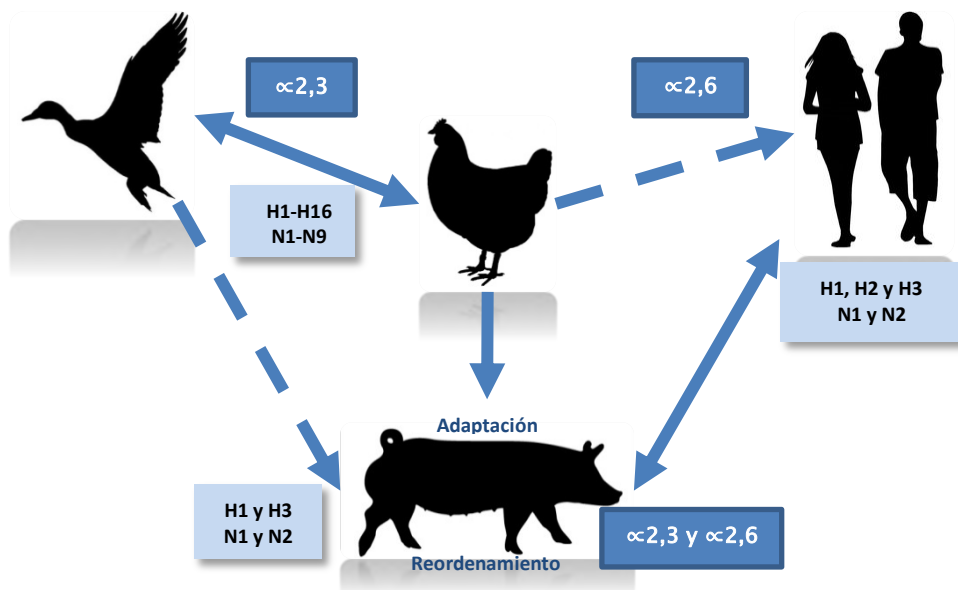


Figura 1. Transmisión interespecie del virus de influenza tipo A. Las flechas marcan el flujo de infección entre diferentes especies. Las líneas punteadas indican la vía con menor ocurrencia.

1.3.6. Variaciones antigénicas de los virus de la influenza tipo A

Debido a la estructura segmentada del genoma del virus de la influenza y la variación antigénica de las proteínas virales de superficie, se propician cambios continuos dados por dos mecanismos, deriva antigénica (*antigenic drift*) y cambio antigénico (*antigenic shift*), lo que resulta en virus capaces de evadir la respuesta inmune adaptativa en el hospedador⁴⁰.

1.3.6.1. Deriva antigénica

El cambio gradual que ocurre mediante mutaciones puntuales en los 2 genes que contienen el material genético que codifica para las 2 proteínas de superficie HA y NA se denomina deriva antigénica, ésta ocurre únicamente en virus de la influenza tipo A y B. Estas mutaciones son el

resultado de la evolución gradual de cepas virales como respuesta a la presión selectiva para evadir el sistema inmune del hospedador⁴¹.

Las mutaciones ocurren con cada ciclo de replicación del virus y se dan con mayor frecuencia en los sitios de unión a anticuerpos de las proteínas de superficie, aproximadamente una vez cada diez mil nucleótidos (1×10^4), lo que representa casi la totalidad de la longitud del genoma del virus, la ARN polimerasa viral comete un error en la inserción de un nucleótido (debido a la ausencia de enzimas de prueba de lectura de ARN) lo que ocasiona que cada virión porte al menos una mutación, sin embargo la mayoría de este tipo de mutaciones son neutrales⁴⁰.

1.3.6.2. Cambio antigénico

El cambio antigénico ocurre únicamente en virus de la influenza tipo A y es el resultado de la sustitución de los segmentos genómicos completos entre distintas cepas virales, principalmente la proteína HA y con menos frecuencia la proteína NA, lo que a su vez resulta en nuevos subtipos virales, lo anterior puede ocurrir mediante dos vías: contagio directo entre animales (aves y cerdos) y humanos, o mediante una mezcla de genes de virus de origen humano y animal para crear un nuevo subtipo de virus humano mediante un proceso denominado recombinación genética. Se estima que el cambio antigénico ocurre aproximadamente 3 veces cada 100 años (pandemias de 1918, 1957 y 1968)⁴¹.

Un factor importante que contribuye para que ocurran eventos de reordenamiento genético es la co-circulación de diferentes subtipos, esto debido a que una célula puede ser infectada con más de un subtipo viral al mismo tiempo, lo que permite el reordenamiento de los segmentos genómicos del virus en la replicación³⁵.

1.3.7. Epidemiología de los virus de la influenza tipo A

La influenza es una enfermedad respiratoria aguda que causa epidemias anuales y ocasionalmente pandemias. La tasa de incidencia anual de la influenza estacionaria a nivel mundial en humanos es del 5 al 10% en adultos, y del 20 al 30% en niños. La enfermedad es causa de hospitalización y muerte, sobre todo en los grupos de alto riesgo (niños menores de 6 años, ancianos y enfermos crónicos). Estas epidemias causan en todo el mundo alrededor de 3 a 5 millones de casos de enfermedad grave y entre 250 000 y 500 000 muertes anuales⁴².

Durante el siglo XX los virus de la influenza tipo A han sido la causa de pandemias en humanos. En 1918 la influenza española ocasiono la muerte de ~50 millones de personas en todo el mundo causada por el subtipo H1N1, en 1957 con la influenza Asiática murieron ~4 millones de personas por el subtipo H2N2 y en 1968 el virus H3N2 ocasiono la muerte de ~2 millones de personas en la influenza de Hong Kong⁴³.

En junio 11 del 2009 la Organización Mundial de la Salud (OMS) declaró la primera pandemia en 41 años ocasionada por el subtipo H1N1pdm (pandémico) 2009 que consistía de un triple reordenamiento estrechamente relacionado a los virus de la influenza tipo A que circulaban en Norteamérica, Asia y Europa en aves, cerdos y humanos respectivamente⁴⁴.

1.4. Influenza aviar

La mayoría de las especies aviares domésticas y silvestres son susceptibles a la infección por el virus de la influenza aviar tipo A (AIV), los cuales se encuentran agrupados en dos patotipos, virus de influenza aviar de baja patogenicidad (IABP) y virus de influenza aviar de alta patogenicidad (IAAP), esta última perteneciente a la lista A de epizootias (OIE 2016), sin embargo, si un subtipo es patogénico para una especie puede no serlo para otra⁴⁵.

1.4.1. Virus de influenza aviar de alta patogenicidad (IAAP)

Aunque se han descrito 16 HA en aves, únicamente H5 y H7 están asociadas con virus IAAP. En marzo de 1997 se reportó un brote en Hong Kong del subtipo H5N1 en aves de corral, ese mismo año se identificaron personas infectadas con virus idénticos a los aviares⁴⁶, ésta fue la primera vez que se reportaban virus en humanos que no habían pasado por un proceso de adaptación previo en mamíferos^{47 48}. En 2001 se identificaron seis genotipos, debido a los eventos de reordenamiento del virus IAAP H5N1 entre cepas de virus de aves silvestres y domésticas⁴⁹.

1.4.2. Virus de influenza aviar de baja patogenicidad (IABP)

Debido a la prevalencia de virus IABP, estas han adoptado mecanismos para mutar a IAAP y ocasionalmente causar muerte por coinfección con patógenos secundarios⁵⁰.

Desde 1993 variantes de baja patogenicidad del subtipo H5N2 se encuentran circulando en México, las cuales se han establecido de forma endémica adaptándose a propagarse en aves vacunadas^{50 51}.

Entre los años 1994 y 2004 se detectó IABP en aves de corral y silvestres en Europa, África, Asia y América del norte, donde particularmente se tuvo que sacrificar a cuatro millones de pavos⁵².

1.4.3. Importancia en salud pública

Hasta 2014, la Organización Mundial de la Salud (OMS) reportó más de seiscientos casos de infecciones humanas del subtipo de alta patogenicidad H5N1 de origen aviar. El primer caso de infección en humanos en el continente americano fue reportado en enero de 2014 en Canadá⁵³. Estos incidentes muestran que la especificidad del receptor no es un factor determinante en la infección a otro hospedero⁵⁴. Así mismo, los virus aviares no siempre necesitan un hospedador intermediario, ni eventos de reordenamiento con subtipos circulantes para poder infectar a un hospedero humano⁵⁵.

1.4.4. Influenza aviar en México

En Mayo de 1994 se identificó una cepa de IABP H5N2 en aves de corral, sin embargo desde finales de 1993 se habían detectado los primeros casos sospechosos de IA en México. Los virus de IABP se esparcieron por el territorio nacional y eventualmente mutaron a una cepa IAAP, principalmente identificada en los estados de Puebla y Querétaro⁵⁶. El gobierno implementó un programa de vacunación en 1995 para el control de las cepas de IABP y de IAAP, sin embargo pese a que no se han reportado brotes del virus de IAAP H5N2 desde 1996, virus de IABP continúan circulando y esparciéndose a otros países de Centroamérica. En 2000 se reportaron virus genéticamente relacionados con las cepas mexicanas de IABP H5N2 en Guatemala y El Salvador⁵⁷.

1.5. Influenza porcina

El virus de influenza porcina (VIP) participa activamente en el complejo respiratorio porcino, se presenta con cuadro respiratorio (tos, estornudos y congestión ocular) y fiebre, que se extiende al 100% de los animales. La recuperación se produce de 5 a 7 días, pero el impacto en la condición corporal es notorio⁵⁸.

La importancia de su estudio y control radica principalmente desde el punto de vista epidemiológico, ya que el cerdo es considerado como riesgo potencial de diseminación del virus entre especies. Lo anterior supone mayor susceptibilidad de los cerdos a los virus de influenza aviar y humano, sumado a que la recombinación genética de dichos virus frecuentemente emerge de los cerdos⁵⁹.

1.5.1. Importancia en salud pública

Entre 1958 y 2005 se han reportado 37 casos de influenza tipo A de origen porcino en humanos⁶⁰.

La mayoría de las infecciones por influenza porcina son clínicamente muy similares a las provocadas por virus de influenza humana, a pesar de esto han existido casos fatales en humanos infectados con

virus de influenza porcina, principalmente del subtipo H1N1. En 1976, cuando ocurrió el llamado **“incidente New Jersey” en los Estados Unidos, cerca de 500 humanos se infectaron con un virus H1N1 idéntico al aislado de cerdos en ese momento.** No se pudo probar que los cerdos fueran la fuente del virus, además la cepa inicialmente causaba una enfermedad leve o asintomática en humanos y en todo caso no era pandémica³⁵.

1.5.2. Subtipos de influenza tipo A circulantes en cerdos

Actualmente los subtipos circulantes a nivel mundial de influenza en cerdos son: H3N2, H1N1 y H1N2 (Cuadro 4). Los VIP H1N1 que predominan en Europa son completamente de origen aviar y fueron introducidos por patos silvestres. En Estados Unidos circulan dos cepas de VIP H1N1, los virus **“clásicos” H1N1, que permanecen desde inicios del siglo XX y también virus producto de reordenamiento con glicoproteínas de superficie del virus clásico y proteínas internas del virus H3N2 o del VIP H1N2⁴⁴.** En el cuadro 2 se puede observar que pese a que son 3 cepas las circulantes en los cerdos a nivel mundial, existen una gran cantidad de variantes genéticas que son el producto de co-circulación de cepas genéticamente diferentes que dan origen a eventos de reordenamiento en estos virus.

Cuadro 4. Variantes genéticas de subtipos del virus de influenza porcina a nivel mundial. Asia representa un punto geográfico importante para posibles recombinaciones genéticas del virus, al presentar tanto cepas de origen Europeo como Norte Americano.

SUBTIPO	VARIANTE	REGIÓN	ORIGEN
H1N1	Clásica	Norte América	pdmH1N1 (1918)
	Clásica	Asia	pdmH1N1 (1918)
	Europea tipo aviar	Europa	Aviar
	Europea tipo aviar	Asia	Europeo aviar
H3N2	Human-like	Norte América	Humano
	Human-like	Asia	Humano
	Human-like	Europa	Humano
	Doble reordenamiento	Norte América	HA, NA y PB1 (H3N2 Humano) NS, NP, M, PB2 y PA (csH1N1)
	Triple reordenamiento	Norte América	NS, NP y M (csH1N1) HA, NA y PB1 (H3N2 Humano) PB2 y PA (H1N1 Aviar)

	Doble reordenamiento		Europa	HA y NA (PC-like) 6 segmentos (EA-like)
	Sydney 1997		Asia	H3N2 Humana
	Doble reordenamiento		Asia	HA y NA (PC-like) 6 segmentos (EA-like)
H1N2	Triple reordenamiento		Norte América	NA, PB1 (Humano) HA, M, NP y NS (Cerdo) PA y PB2 (Aviar)
	Doble reordenamiento		Norte América	NA y PB1 (trH3N2) HA, NP, M, NS, PA y PB2 (csh1N1)

1.6. Virus de Newcastle

1.6.1. Taxonomía y clasificación

El virus de Newcastle pertenece al orden de los *Mononegavirales*, el cual comprende las familias *Bornaviridae*, *Filoviridae*, *Nyamiviridae*, *Rhabdoviridae* y *Paramyxoviridae*, esta última se divide en dos subfamilias, la *Pneumovirinae* y la *Paramyxovirinae* que consta de siete géneros: *Aquaparamyxovirus*, *Avulavirus*, *Ferlavirus*, *Henipavirus*, *Morbillivirus*, *Respirovirus* y *Rubulavirus*. El género *Avulavirus* se comprende de nueve especies, una de ellas el virus de *Newcastle* y ocho especies del virus *Avian paramyxovirus*⁶¹.

De acuerdo a su patogenicidad las cepas del virus de *Newcastle* se clasifican en: asintomáticas, lentogénicas, mesogénicas, neurotrópicas velogénicas y viscerotrópicas velogénicas⁶¹.

1.6.2. Características y genoma

El virus presenta formas pleomorfas con prolongaciones superficiales, generalmente esféricas entre 100 y 500nm, aunque también filamentosas de hasta 100nm, se encuentran provistos de envoltura con simetría helicoidal. Tiene un genoma ARN de cadena sencilla de polaridad negativa (-ssRNA), de 15,186 Kb de tamaño y codifica para 6 proteínas, la Nucleoproteína (N), la Fosfoproteína (P), la proteína Matriz (M), la proteína de Fusión (F), la proteína Hemaglutinina-Neuraminidasa (HN) y la proteína Polimerasa larga (L)⁶².

2. JUSTIFICACIÓN

Los genomas de los virus animales son relativamente pequeños, pero con un alto grado de variabilidad, ocasionado principalmente por la baja fidelidad de la polimerasa viral, conocer a detalle estas variaciones y el impacto de las mismas resalta la importancia del estudio de su genoma completo.

Una estrategia disponible para la secuenciación del genoma completo es el uso de las tecnologías de la secuenciación masiva, que presentan la ventaja de generar grandes cantidades de información en un solo ensayo. Adicionalmente mediante el uso de herramientas bioinformáticas tanto de dominio público como comercial se puede realizar el análisis de los datos generados por los equipos de secuenciación.

Actualmente una de las limitantes para abordar estudios de secuenciación de genomas completos en virus es la falta de información que dirija a los investigadores en la elección de la estrategia que facilite el análisis de la información generada a partir de estas tecnologías.

Particularmente resulta interesante tomar como modelo el virus de influenza aviar H5N2 y estudiar la evolución de su genoma completo en México, ya que hasta ahora se han realizado estudios de la secuencia del gen de la hemaglutinina que han mostrado que las cepas circulantes han estado cambiando continuamente. Identificar cómo ha cambiado el genoma completo del virus y los linajes que se encuentran circulando actualmente en nuestro país, permitirá identificar las variantes genéticas circulantes en México y determinar la evolución del virus a través del tiempo.

3. HIPÓTESIS

El uso de las plataformas de secuenciación masiva en la secuenciación del genoma completo del virus de Newcastle y el virus de influenza aviar H5N2 permitirá identificar la estrategia más eficiente para la secuenciación, ensamblado y análisis de virus animales.

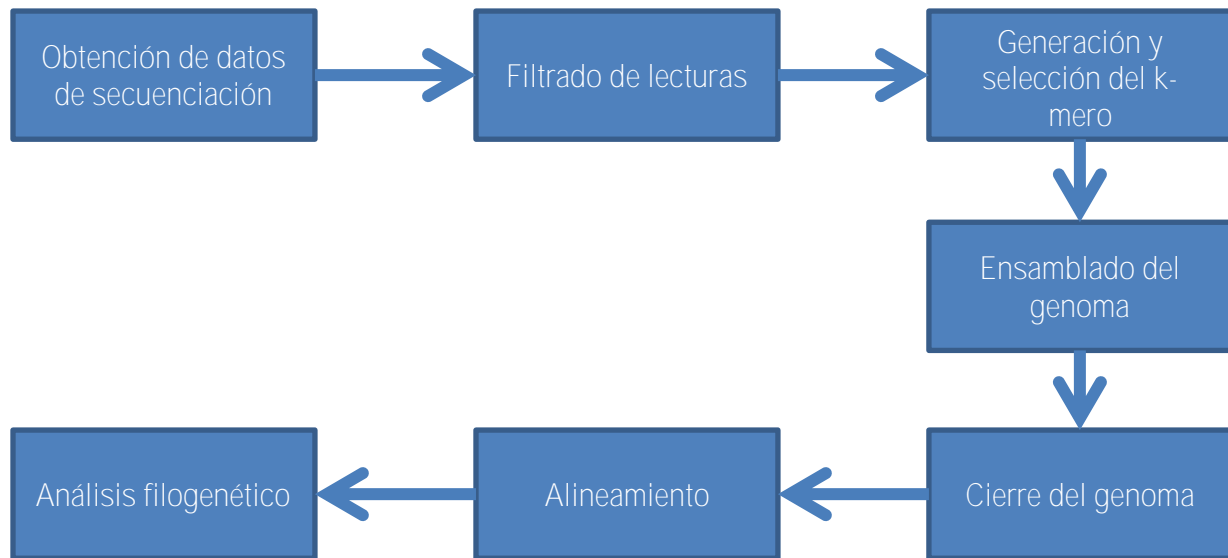
4. OBJETIVO GENERAL

Evaluar diferentes plataformas de secuenciación masiva y herramientas bioinformáticas para la secuenciación y ensamblado de genomas virales tomando como modelos el virus de Newcastle y el virus de influenza aviar H5N2.

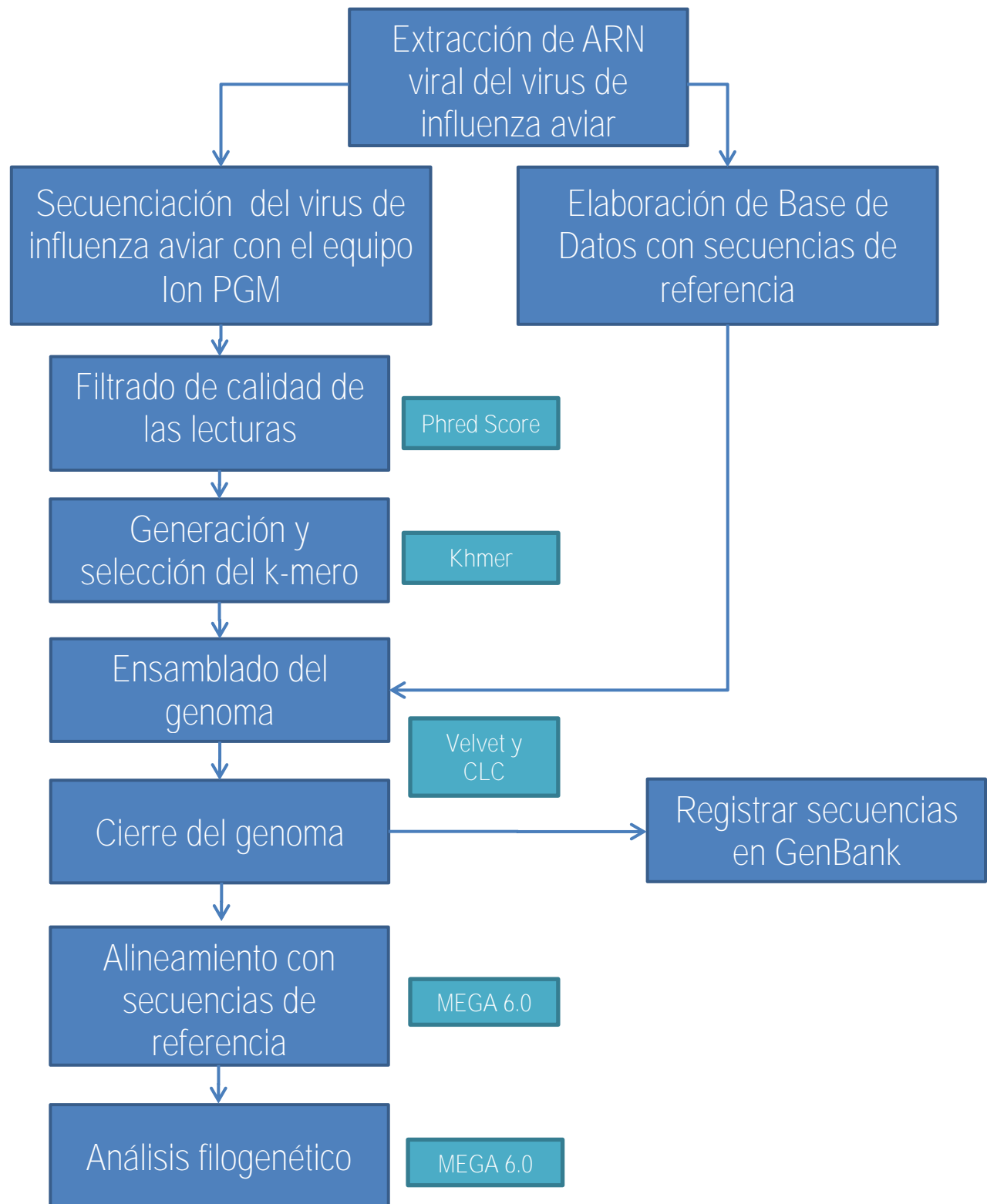
5. OBJETIVOS ESPECÍFICOS

- I. Evaluar las plataformas de secuenciación masiva Ion PGM y Genome Analyzer IIX para determinar la mejor estrategia de secuenciación del genoma del virus de Newcastle y el virus de influenza aviar.
- II. Utilizar software comercial y de dominio público para el ensamblado de genomas del virus de Newcastle y del virus de influenza aviar a partir de datos obtenidos por secuenciación masiva.
- III. Obtener el ensamblado de los genomas del virus de Newcastle y del virus de influenza aviar.
- IV. Determinar relación evolutiva del genoma del virus de influenza aviar mediante un análisis filogenético, utilizando secuencias de referencia del virus.

6. ESTRATEGIA EXPERIMENTAL



7. METODOLOGÍA DEL ENSAMBLADO DE GENOMAS



8. MATERIALES Y MÉTODOS

Se utilizaron dos plataformas de secuenciación masiva para determinar la secuencia del genoma completo de muestras del virus de Influenza Aviar y del virus de Newcastle. Las muestras del virus de Newcastle secuenciadas en este trabajo se utilizaron para la estandarización del proceso de ensamblado a partir de datos generados por secuenciación masiva.

La secuenciación de las muestras del virus de influenza aviar se realizó en la Facultad de Medicina Veterinaria y Zootecnia (FMVZ), en el Departamento de Medicina y Zootecnia de Cerdos de la Universidad Nacional Autónoma de México (UNAM), utilizando la tecnología de secuenciación por semiconductor de Thermo Fisher Scientific Inc.®, la secuenciación de las muestras del virus de Newcastle se llevó a cabo en el Instituto Nacional de Medicina Genómica (INMEGEN), utilizando la tecnología de secuenciación por síntesis de Illumina® (Illumina, San Diego, CA).

8.1. Muestras

Se secuenciaron cinco muestras de influenza aviar (Cuadro 5) y doce muestras del virus de Newcastle (Cuadro 6). Las muestras fueron proporcionadas por un laboratorio de diagnóstico privado.

Cuadro 5. Muestras del virus de influenza aviar procesadas en el equipo **Ion PGM™**. Año y origen de los Virus de influenza aviar secuenciados.

MUESTRA	AÑO	PAÍS	ESTADO
JX-1	2014	México	-
JX-2	2014	México	-
JX-3	2014	México	-
JX-4	2014	México	-
M11	2005	México	Querétaro

Cuadro 6. Muestras del virus de Newcastle procesadas en el equipo Genome Analyzer IIX. Información de los virus de Newcastle secuenciados. Se desconoce el año de aislamiento de la muestra.

MUESTRA	PAÍS
New 1	México
New 2	México
New 3	México
New 4	México
New 5	México
New 5 a	México
New N5	México
New N5 a	México
New LS	México
New M	México
JV-01	México
JV-02	México

8.2. Preparación de las bibliotecas para secuenciación

Los ocho segmentos genómicos de las cinco muestras de influenza aviar fueron secuenciadas en el equipo de secuenciación masiva **Ion PGM™**, en un chip **Ion 314™ Chip Kit v2**⁶³.

Para la generación de las bibliotecas se utilizó el kit comercial **PathAmp™ FluA Reagents**⁶⁴ (Figura 2). Para la elaboración de las bibliotecas se siguieron las especificaciones del fabricante. Se ligaron los adaptadores para la síntesis del DNA complementario (cDNA), se realizó el enriquecimiento por PCR,

posteriormente se llevó a cabo la purificación del producto de PCR. Para la construcción de las bibliotecas se fragmentó aleatoriamente el cDNA y se ligaron los adaptadores con las etiquetas.

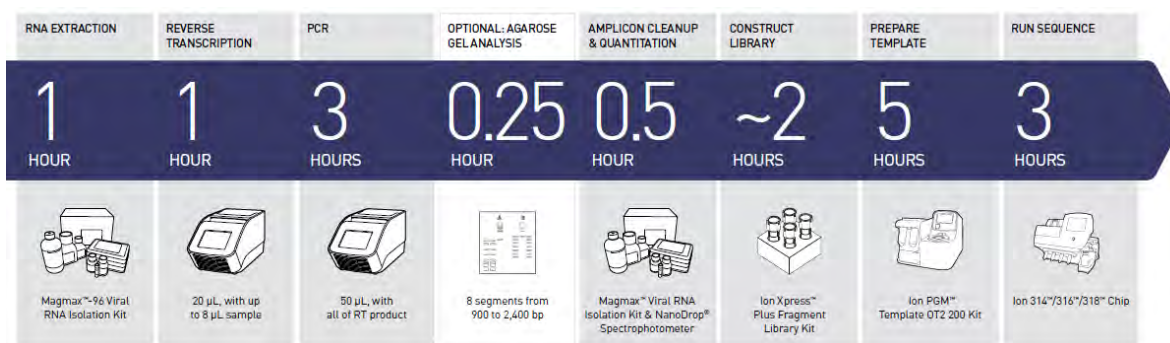


Figura 2. Metodología de secuenciación del virus de influenza tipo A en el equipo Ion PGM. Proceso de secuenciación del virus de influenza empleando el kit comercial *PathAmp™ FluA Reagents*.

Las muestras del virus de Newcastle se secuenciaron en el equipo de Illumina Genome Analyzer Iix. Para la elaboración de las bibliotecas se siguieron las especificaciones del fabricante, se fragmentó el ADN y se añadieron las secuencias de adaptadores a los extremos, se realizó la hibridación con las sondas de captura y finalmente se realizó la amplificación mediante PCR de las regiones de alta especificidad⁶⁵.

8.3. Evaluación de la calidad de las lecturas

La calidad de las lecturas obtenidas de la secuenciación fue evaluada utilizando el programa FastQC® (*High Throughput Sequence QC Report*)⁶⁶. FastQC es una herramienta que genera

estadísticos de las secuencias y parámetros para la evaluación de la calidad de datos generados por secuenciación masiva.

Para evaluar la calidad de las lecturas se evaluó el contenido de GC, el valor Q (*Phred score*)⁶⁷, el cual estima la probabilidad de que cada uno de los nucleótidos secuenciados sea erróneo; donde un valor Q de 30 equivale a un error por cada mil bases, así como la longitud de las lecturas.

8.4. Ensamblado del genoma

Para realizar el ensamblado del genoma a partir de datos de secuenciación masiva se siguió la estrategia de ensamblado con genoma de referencia, utilizando el programa *CLC Genomics Workbench 7®* (QIAGEN Company). Las secuencias genómicas utilizadas como referencia para el mapeo de las lecturas fueron obtenidas de la base de datos GenBank® (NCBI).

Para el ensamblado de las muestras del virus de Newcastle se utilizó como referencia el genoma del virus de La Sota (Número de acceso de GenBank® KC844235).

Para el ensamblado del genoma de las muestras del virus de influenza aviar se utilizaron 3 genomas del subtipo H5N2 para las muestras M11, JX-1, JX2 Y JX-3, mientras que para la muestra JX-4 se utilizaron 2 genomas del subtipo H7N3 (Cuadro 7).

Cuadro 7. Secuencias utilizadas para el ensamblado del genoma de las muestras del virus de influenza aviar. Las secuencias utilizadas como referencia fueron descargadas de la base de datos GenBank®.

SUBTIPO	AÑO	PAÍS	ESTADO	NÚMERO DE ACCESO
H5N2	1994	México	-	GU186573:GU186580
H5N2	2006	México	Veracruz	KM455878:KM455885
H5N2	2010	México	Aguascalientes	KM657856:KM657863
H7N3	2006	México	-	KC669383:KC669390
H7N3	2012	México	Jalisco	JX317626,JX397993 JX465631:JX465636

8.5. Análisis bioinformático de las secuencias del virus de influenza aviar A/H5N2

Las secuencias ensambladas fueron exportadas al formato FASTA para poder realizar los análisis de las mismas. Las secuencias utilizadas para los análisis filogenéticos del en este trabajo fueron descargadas de la base de datos GenBank® (Cuadros 8 y 9).

Para el análisis de las secuencias de las muestras de influenza aviar se utilizaron un total de 47 secuencias de genomas completos de influenza aviar A H5N2 de México, Guatemala y El Salvador (Cuadro 8) y 2 secuencias de genomas completos de influenza porcina A H5N2 de Corea como outgroup (Cuadro 9) para realizar los alineamientos y las construcciones filogenéticas. Para el gen HA se realizó un análisis adicional con secuencias parciales, para el cual se utilizaron 119 secuencias.

Cuadro 8. Secuencias utilizadas para los análisis filogenéticos de las secuencias del virus de influenza aviar. Secuencias de genomas completos de influenza aviar H5N2 utilizadas en el análisis filogenético.

IDENTIFICADOR	AÑO	PAÍS	ESTADO	NÚMERO DE ACCESO
28159-232	1994	México	Hidalgo	CY005832:CY005838, CY006040
14589-660	1994	México	Jalisco	FJ610108:FJ610115
31381-3	1994	México	-	GU186565:GU186572
31381-7	1994	México	-	GU186573:GU186580
31381-991	1994	México	-	GU052659:GU052666, U37165
14585-622	1994	México	Puebla	GU186539:GU186544, GU186547, U37179
14586-654	1994	México	Puebla	GU052667:GU052674
14587-644	1994	México	Puebla	FJ610094:FJ610101
28159-488	1995	México	Chiapas	FJ610082:FJ610087, AY497163, AY497068
28159-232	1995	México	Hidalgo	GU052682:GU052689
28159-541	1995	México	-	GU052690:GU052697
28159-538	1995	México	Morelos	GU052698:GU052705
28159-474	1995	México	Puebla	GU052675:GU052681, AY497117
7653-20	1995	México	Querétaro	FJ610076:FJ610081, U85390, U79448
14588-19	1995	México	Querétaro	CY015097:CY015103,

				AB558474
28159-398	1995	México	Veracruz	GU186588:GU186593, AY497160, AY497065
37821-771	1996	México	-	GU052588:GU052595
22019-853	1996	México	Querétaro	GU052763:GU052770
15406	1997	México	Chiapas	GU052596:GU052603
15408	1997	México	Chiapas	GU052604:GU052611
15224	1997	México	Chiapas	CY005839:CY005844, CY014717, CY006041
15407	1997	México	-	GU186545:GU186550, AY497172, AY497077
124-3705	1998	México	Aguascalientes	AY497081, AY497099, AY497107, AY497115, AY497123, AY497131, AY497139, AY497176
228-4471	1998	México	Jalisco	GU052732:GU052739
230-4956	1998	México	Jalisco	GU052706:GU052713
22184	1998	México	-	GU052612:GU052619
227-4353	1998	México	Morelos	GU052714:GU052719, AY497082, AY497177
231-5284	1998	México	Puebla	AY497084, AY497100, AY497108, AY497116, AY497124, AY497132, AY497140, AY497179
232-6169	1998	México	Veracruz	GU052720:GU052725, AY497180, AY497085
45511-1	2000	Guatemala	-	GU052628:GU052635
102711-1	2001	El	-	AY497093, AY497102,

		Salvador		AY497110, AY497118, AY497126, AY497134, AY497142, AY497188
102711-2	2001	El Salvador	-	GU052636:GU052643
194573	2002	Guatemala	-	GU186557:GU186564
1433-4	2005	México	Durango	KM455886:KM455893
1433-8	2005	México	Durango	KM455910:KM455917
1433-6	2005	México	Hidalgo	KM455902:KM455909
1433-5	2005	México	-	KM455894:KM455901
1433-15	2006	México	Coahuila	KM455942:KM455949
1433-9	2006	México	Durango	KM455918:KM455925
1433-10	2006	México	Durango	KM455926:KM455933
1433-13	2006	México	-	KM455934:KM455941
1433-1	2006	México	Veracruz	KM455878:KM455885
IA05	2009	México	Coahuila	KM657864:KM657871
IA13	2010	México	Aguascalientes	KM657856:KM657863

Cuadro 9. Secuencias de influenza porcina utilizadas como outgroup en el análisis filogenético. Se decidió utilizar como *outgroup*, secuencias de influenza A/H5N2 presente en cerdos de Corea por su origen aviar y su distancia filogenética.

IDENTIFICADOR	AÑO	PAÍS	ESTADO	NÚMERO DE ACCESO
C12	2008	Corea	-	FJ461592:FJ461599
C13	2008	Corea	-	FJ461600:FJ461607

Para el análisis de las secuencias de las muestras de Newcastle se utilizaron secuencias del gen HN (hemaglutinina-neuraminidasa), así como los genomas completos de la cepa chimalhuacan, NDV-P05, 01/10 y La Sota (Cuadro 10).

Cuadro 10. Secuencias del virus de Newcastle utilizados para el ensamblado de las secuencias. Genomas completos y gen HN utilizados en los alineamientos.

IDENTIFICADOR	AÑO	PAÍS	NÚMERO DE ACCESO
Chimalhuacan	2014	México	KJ577136
NDV-P05	2012	México	HM117720
01/10	2010	México	JX974435
HN/M	2008	México	KF910962
La Sota	2013	China	KC844235

8.6. Alineamientos

Los alineamientos de las secuencias de ambos virus se realizaron utilizando el programa MEGA 6⁶⁸ con el algoritmo de ClustalW⁶⁹, con los parámetros por default. Para la edición de las secuencias del virus de Newcastle se utilizó el programa Clone Manager 9[©] (*Scientific & Educational Software*).

La evaluación de los alineamientos se realizó con el cálculo de la distancia media global (OMD: overall mean distance), para que un alineamiento sea confiable el valor OMD debe ser menor a 0.33⁷⁰.

Para las secuencias del virus de Newcastle se alinearon las secuencias de los genes F y HN (402 - 3554) así como el genoma completo.

Las secuencias del virus de influenza aviar fueron alineadas tomando las regiones codificantes de cada uno de los ocho segmentos genómicos del virus por separado (Cuadro 11). Para el gen HA se realizó un alineamiento adicional de la secuencia parcial del gen (472-1050).

Cuadro 11. Alineamiento de los segmentos genómicos del virus de influenza aviar. Se alinearon las regiones codificantes de los ocho segmentos genómicos.

Segmento	Región
PB2	14-2293
PB1	13-2286
PA	25-2175
HA	14-1708
NP	19-1515
NA	1-1410
M	14-995
NEP	15-852

8.7. Análisis filogenético del virus de influenza aviar A/H5N2

Las construcciones de los árboles filogenéticos se realizaron utilizando el programa MEGA 6^{©58}, por el método de Máxima verosimilitud (*Maximum likelihood*)⁷¹ con un *bootstrap* de 1000 réplicas, siendo el propuesto por la herramienta en base a las características de las secuencias.

Se realizó el análisis filogenético para los ocho segmentos del virus de influenza aviar por separado, en relación a secuencias reportadas de genomas completos de México, Guatemala y El Salvador (Cuadro 8).

8.8. Análisis de recombinación genómica del virus de influenza aviar A/H5N2

Los *heatmaps* son una herramienta gráfica comúnmente utilizada para el análisis de niveles de expresión de un grupo de genes⁷², en este trabajo evaluamos su uso para elucidar posibles recombinaciones en el genoma del virus.

Con el objetivo de evaluar posibles recombinaciones en los segmentos genómicos del virus de influenza aviar se construyeron *heatmaps* para cada uno de los ocho segmentos, los cuales fueron creados utilizando el programa estadístico R^{©73} con las secuencias utilizadas para los alineamientos (Cuadro 8).

Para la elaboración de los *heatmaps* como herramienta que permita determinar el flujo de los segmentos genómicos entre las cepas mexicanas de influenza aviar de los años 1994 a 2014 se utilizaron secuencias completas del genoma del virus de influenza aviar.

9. RESULTADOS

Para poder investigar la evolución del genoma del virus de influenza aviar se secuenció y se analizó la secuencia completa de cada uno de sus segmentos.

9.1. Secuenciación

De las cinco muestras procesadas del virus de influenza aviar **en el equipo Ion PGM™ se logró** obtener lecturas suficientes para realizar el ensamblado del genoma completo de cuatro (JX-1, JX-2, JX-3 y JX-4). Tanto la profundidad como la cobertura obtenida para cada uno de los segmentos fue muy variable en todas las muestras (Cuadros 12, 13, 14 y 15).

Cuadro 12. Lecturas obtenidas para la muestra JX-1 secuenciada en el equipo Ion PGM™. Los segmentos con menos cobertura fueron los tres segmentos de las polimerasas.

JX-1		
SEGMENTO	LECTURAS	PROFUNDIDAD
PB2	286	24.70
PB1	613	52.94
PA	337	30.18
HA	756	87.20
NP	810	106.44
NA	998	145.27
MP	1,450	289.42
NEP	1,362	314.91

Cuadro 13. Lecturas obtenidas para la muestra JX-2 secuenciada en el equipo Ion PGM™. El segmento con menos cobertura fue el segmento de la nucleoproteína.

JX-2		
SEGMENTO	LECTURAS	PROFUNDIDAD
PB2	555	47.92
PB1	466	40.24
PA	621	55.62
HA	1042	113.99
NP	340	41.13
NA	401	55.02
MP	725	141.72
NEP	604	136.1

Cuadro 14. Lecturas obtenidas para la muestra JX-3 secuenciada en el equipo Ion PGM™. Los segmentos con menos cobertura fueron los tres segmentos de las polimerasas.

JX-3		
SEGMENTO	LECTURAS	PROFUNDIDAD
PB2	500	41.55
PB1	486	40.22
PA	390	33.15
HA	626	67.3
NP	745	91.14
NA	382	52.11
MP	900	178.43
NEP	626	140.45

Cuadro 15. Lecturas obtenidas para la muestra JX-4 secuenciada en el equipo Ion PGM™. El segmento con menos cobertura fue el segmento tres que codifica para la polimerasa acida.

JX-4		
SEGMENTO	LECTURAS	PROFUNDIDAD
PB2	682	48.98
PB1	282	21.97
PA	102	8.09
HA	696	74.76
NP	1125	130.92
NA	955	131.83
MP	2199	418.31
NS	2187	491.88

De las doce muestras procesadas del virus de Newcastle por secuenciación masiva, se obtuvo lecturas suficientes para el ensamblado del genoma completo de siete. Se obtuvo entre 1.8 y 10.9 millones de lecturas por cada muestra con una cobertura de ~95% del genoma viral y una profundidad de entre 12000x y 70,000x (Cuadro 16).

Cuadro 16. Lecturas obtenidas por muestra en el equipo Genome Analyzer IIx. Todas las lecturas tuvieron una longitud de 105pb.

ID	LECTURAS OBTENIDAS	PROFUNDIDAD
New M	9165127	60146
New 3	3973006	26072
New 2	4800937	31506
New 5	4806581	31543
New 5 a	1850573	12144
New N5	2442186	16026
New LS	10995772	72159

9.2. Calidad de las lecturas

Mientras que con Ion PGM™ se obtuvieron lecturas de tamaños variables (Figura 3) con la tecnología de Illumina® todas tuvieron el mismo tamaño (105nt), lo que facilitó el proceso de ensamblado de las secuencias.

La calidad de las lecturas obtenidas con el equipo Genome Analyzer IIx tuvieron un valor Q promedio de 34 (Figura 4) donde hasta la posición 50 de la lectura mantuvieron una Q 40, lo que equivale a un error por cada 10000nt que es cercano al tamaño del genoma del virus, con el **equipo Ion PGM™** las lecturas presentaron un valor Q promedio de 28 (Figura 5).

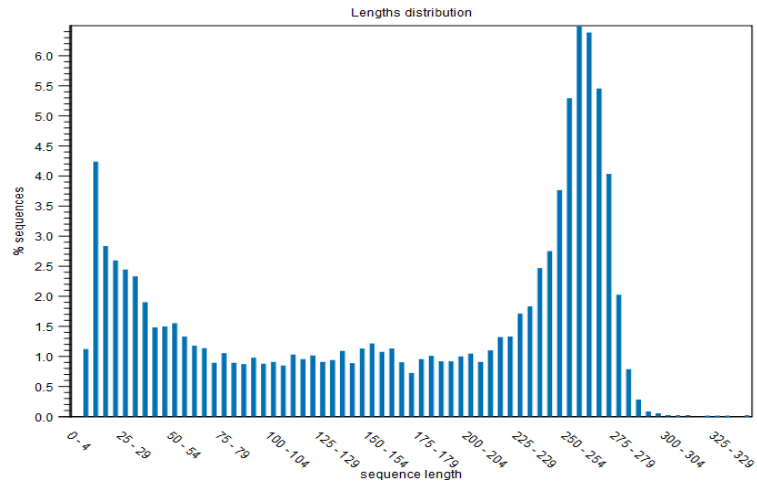


Figura 3. Tamaño de las lecturas obtenidas con el **equipo Ion PGM™**. Las lecturas presentaron un tamaño muy variable, debido a las características técnicas del equipo.

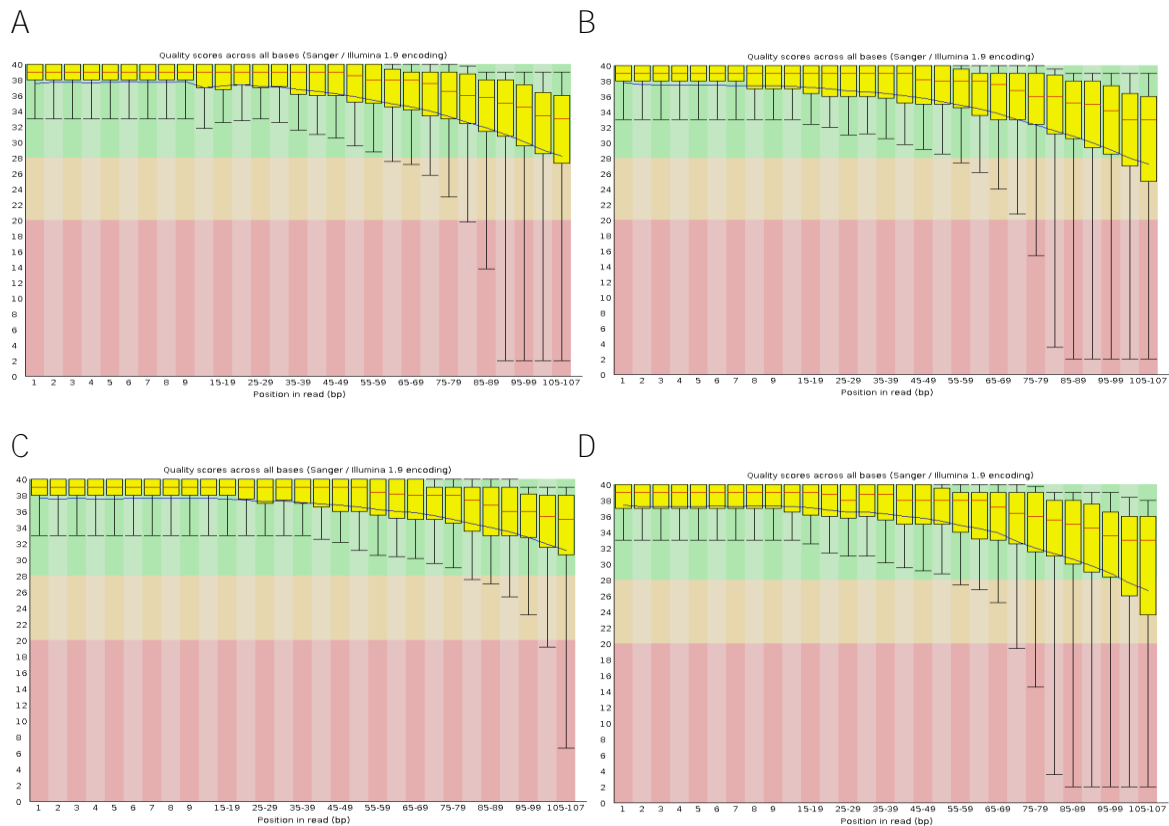


Figura 4. Calidad de las lecturas obtenidas en el equipo Genome Analyzer Ix para las muestras del virus de Newcastle. Se muestra la distribución de la calidad de las lecturas para las muestras New M, New 3, New 2 y New LS en 4A, 4B, 4C y 4D respectivamente. En la barra amarilla se muestra el rango intercuartil del valor Q por posición de la lectura, el valor promedio fue de Q 34, en la línea central en color naranja se muestra el valor medio.

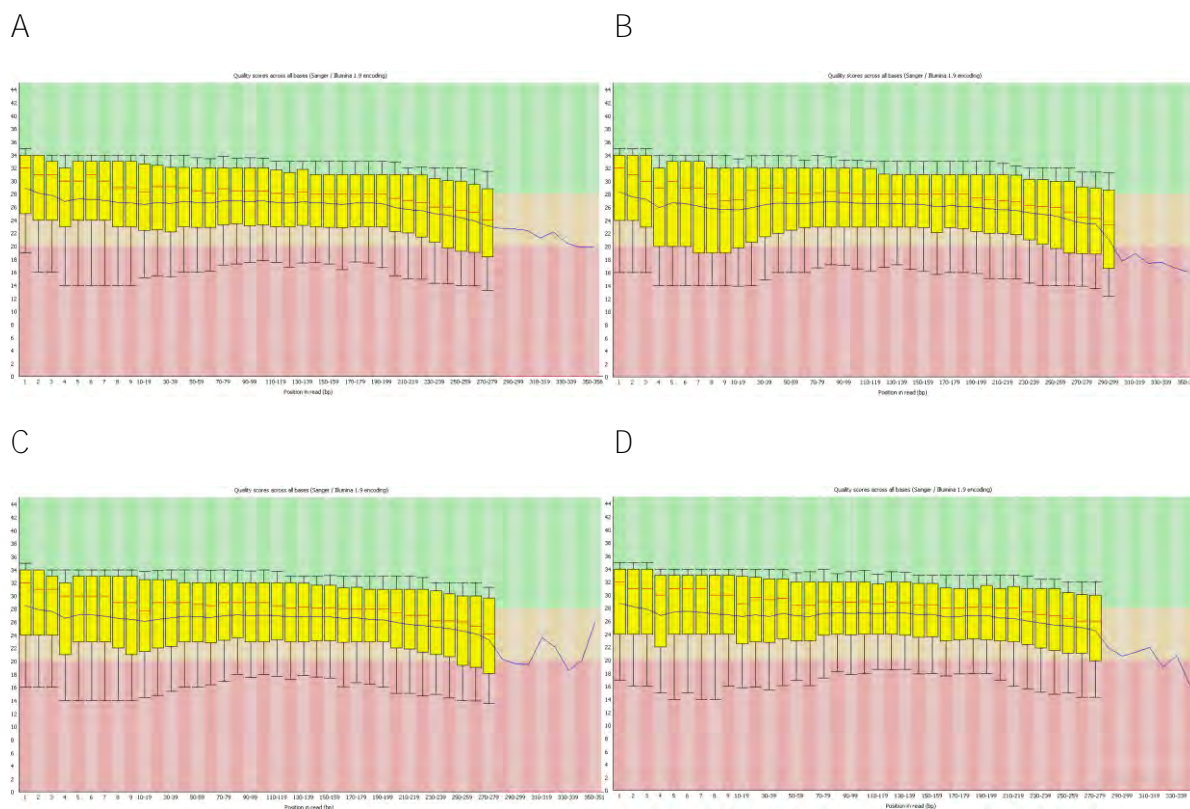


Figura 5. Calidad de las lecturas obtenidas en el equipo **Ion PGM™** para las muestras del virus de influenza aviar A/H5N2. En 5A, 5B, 5C y 5D se puede ver la distribución de la calidad de las lecturas para las muestras JX-1, JX-2, JX-3 y JX-4 respectivamente. En la barra amarilla se representa el rango intercuartil del valor Q por posición en la lectura, el valor promedio fue de Q 28, en la línea central en color naranja se muestra el valor medio.

9.3. Ensamblado del genoma

Una vez evaluada la calidad de las lecturas se les eliminaron las etiquetas que se utilizaron como identificadores al momento de realizar la secuenciación. Como filtro de calidad, se descartaron tanto las lecturas de mala calidad ($Q < 20$) como las inespecíficas para las secuencias del virus de influenza aviar (Figura 6) y el virus de Newcastle (Figura 7).

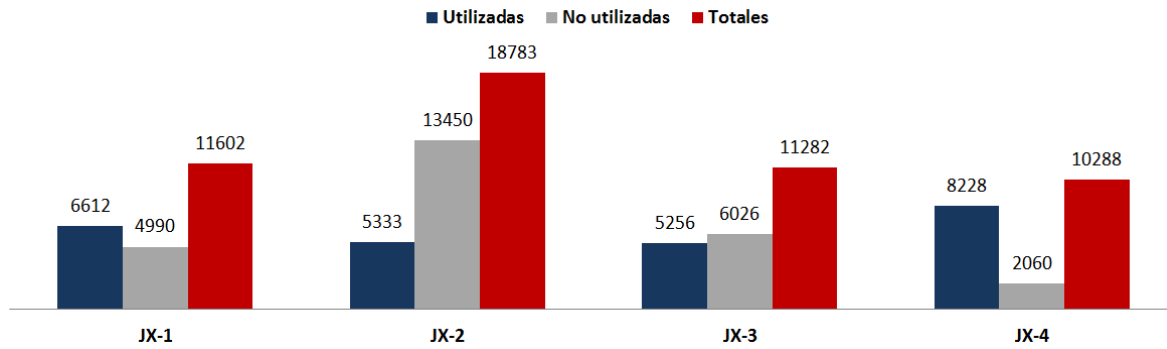


Figura 6. Lecturas obtenidas para cada muestra del virus de influenza aviar A/H5N2 en el equipo **Ion PGM™**. Lecturas utilizadas para el ensamblado (azul), lecturas eliminadas por el filtro de calidad o que no alinearon (gris) y lecturas crudas obtenidas del equipo de secuenciación (rojo) para las muestras JX-1, JX-2, JX-3 y JX-4 respectivamente.

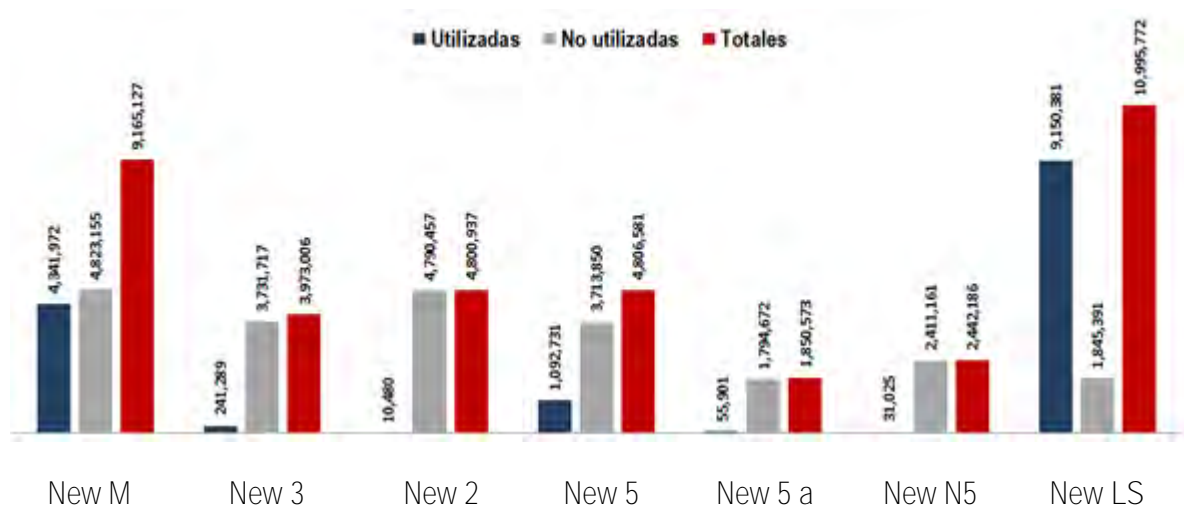


Figura 7. Lecturas obtenidas para cada muestra del virus de Newcastle en el equipo Genome Analyzer IIx. Lecturas utilizadas para el ensamblado (azul), lecturas eliminadas por el filtro de calidad o que no alinearon (gris) y lecturas crudas obtenidas del equipo de secuenciación (rojo) para las muestras New M, New 3, New 2, New 5, New 5 a, New N5 y New LS respectivamente.

Posteriormente al filtrado de calidad, las lecturas se mapearon con un genoma de referencia (Cuadro 7) para poder extraer la secuencia consenso para cada una de las muestras.

Debido a que los 3 segmentos de las polimerasas para cada una de las muestras del virus de influenza aviar no tuvieron una cobertura del 100% (extremos sin lecturas), posiblemente por la cantidad baja de lecturas y a que las secuencias de estos segmentos genómicos son las de longitud más larga, se decidió rellenar los huecos con la secuencia de referencia. Ya que los extremos de las secuencias son altamente conservados entre los genomas del virus de influenza A no se afectó la calidad de los ensamblados.

Para la muestra New 2 del virus de Newcastle no se logró obtener una cobertura del 100% pese a obtener una profundidad mayor a 30000X, siendo imposible obtener una secuencia consenso de buena calidad. Para la muestra New M inicialmente se obtuvo una secuencia con bases ambiguas, sin embargo se editó revisando puntualmente el ensamblado, logrando corregir la secuencia. Las muestras New 3, New N5, New LS, New 5 y New 5 a no presentaron problemas para el ensamblado.

Las secuencias consenso del virus de influenza aviar fueron validadas con la herramienta *Influenza Virus Resource package annotation tool*[®] (NCBI). Las secuencias del virus de Newcastle se validaron tomando en cuenta el marco de lectura correcto para el genoma de estos virus.

9.4. Alineamiento de las secuencias del virus de influenza aviar A/H5N2

La confiabilidad de los alineamientos utilizados en las construcciones filogenéticas del virus de influenza fue evaluada utilizando el valor OMD (Cuadro 17) calculado con MEGA 6[©]. El valor OMD para todos los alineamientos fue menor a 0.33 lo que confirma la certeza de los mismos.

Cuadro 17. Distancia media global de los alineamientos de cada segmento. Todos los alineamientos presentaron un valor OMD < 0.33.

SEGMENTO	OMD
PB2	0.059
PB1	0.052
PA	0.051
HA	0.083
NP	0.049
NA	0.066
M	0.047
NEP	0.05

Adicionalmente se realizó un análisis BLASTN[®] (NCBI) de las secuencias de las proteínas de superficie HA y NA de cada muestra, se obtuvo un porcentaje de identidad arriba del 95%, con una cobertura del 100% y un valor E de cero (Cuadro 18). Las muestras JX-2 y JX-3 presentaron la mayor identidad con una cepa de 2010 de influenza aviar H5N2 (Figura 8), mientras que la muestra JX-1 con una cepa de 2006 y una cepa de 2009 de influenza aviar H5N2 para los genes HA y NA

respectivamente. La muestra JX-4 tuvo 98% de identidad con una cepa de influenza humana H7N3 de 2012 de origen aviar (Figura 9).

Cuadro 18. Alineamiento local de los genes HA y NA del virus de influenza aviar A/H5N2.

Porcentaje de identidad y secuencia del alineamiento local de las proteínas de superficie.

IDENTIFICADOR	GEN	% IDENTIDAD	IDENTIDAD OBTENIDA	NÚMERO DE ACCESO DE GENBANK
JX-1	HA	95%	Hidalgo/2006/H5N2	KM368291
JX-1	NA	96%	Coahuila/2009/H5N2	KM657869
JX-2	HA	97%	Aguascalientes/2010/H5N2	KM657859
JX-2	NA	96%	Aguascalientes/2010/H5N2	KM657861
JX-3	HA	96%	Aguascalientes/2010/H5N2	KM657859
JX-3	NA	95%	Aguascalientes/2010/H5N2	KM657861
JX-4	HA	98%	México/2012/H7N3	CY125728
JX-4	NA	98%	México/2012/H7N3	CY125730

A



B

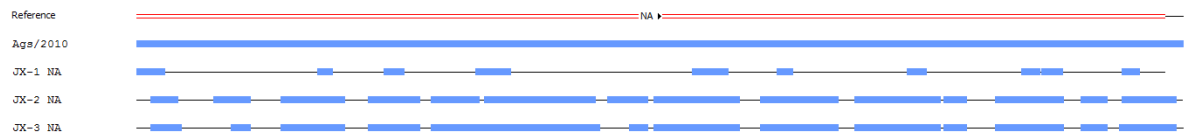


Figura 8. Identidad en los genes de superficie HA y NA del virus de influenza aviar A/H5N2. En 8A y 8B se muestra el porcentaje de identidad de las muestras JX-1, JX-2 y JX3 con la cepa de 2010 (KM657859, KM657861) para los genes HA y NA respectivamente. Las regiones en azul son las que presentan 100 % de identidad con la referencia.

A



B

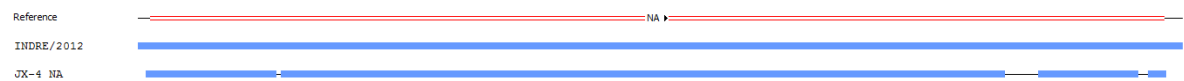


Figura 9. Identidad en los genes de superficie HA y NA del virus de influenza aviar A/H7N3. En 9A y 9B se muestra el porcentaje de identidad de la muestra JX-4 alineado con la secuencia de (JX397993, JX317626) para los genes HA y NA respectivamente, las regiones en azul son las que presentan 100 % de identidad.

Se identificaron las muestras del virus de Newcastle con BLASTN®, se determinó que la cepa New M pertenece al genotipo V (Cuadro 20), el cual es prevalente en México. Las muestras New 3, New N5 y New LS presentaron mayor porcentaje de identidad con la cepa La Sota (Cuadro 19). Para las muestras New 5 a y New 5 se identificó un genoma híbrido, con genoma de la cepa La Sota, de genotipo II, y un segmento del genotipo V (Figura 10).

Cuadro 19. Identidad del genoma completo del virus de Newcastle. Se muestra la identidad del genoma completo obtenido con la herramienta BLASTN®.

IDENTIFICACION	CALIDAD DE LA SECUENCIA	IDENTIDAD OBTENIDA
New 2	Baja calidad	Newcastle
New 3	Buena calidad	LaSota
New 5	Buena calidad	Hibrido LaSota / Genotipo V en region F HN
New N5	Buena calidad	La sota
New 5 a	Buena calidad	Hibrido LaSota / Genotipo V en región F HN
New LS	Buena calidad	LaSota
New M	Buena calidad	Newcastle genotipo V

El sitio de inserción del genoma V (4945 – 8097) para las muestras New 5 a y New 5 contiene los genes F y HN, con un tamaño de 3602pb.

Las muestras New 5 a y New 5 tienen regiones con alto porcentaje de identidad con el genoma de la cepa Chimalhuacan (Figura 10).

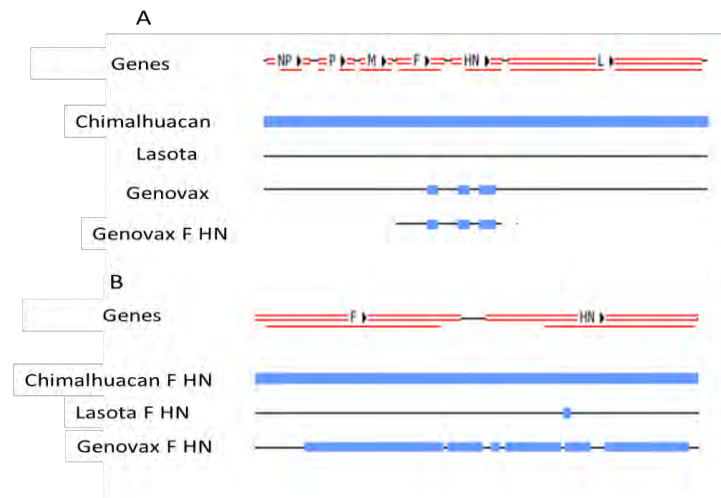


Figura 10. Alineamiento muestras New 5 a y New 5. En A se muestra esquema de identidad entre los genomas de las cepas Chimalhuacan, La Sota, y un segmento de Genovax que contiene los genes F y HN. Las regiones azules son 100% idénticas. En B se presenta la identidad para el segmento que contiene los genes F y HN.

9.5. Análisis filogenético de las secuencias del virus de influenza aviar A/H5N2

El análisis filogenético se realizó con secuencias completas y parciales del genoma del virus de influenza aviar.

Para las reconstrucciones filogenéticas de las secuencias de genomas completos del virus de influenza aviar se emplearon 47 secuencias de genomas completos subtipo H5N2 de México, Guatemala y El Salvador, obtenidas de GenBank®. Para la construcción filogenética de las secuencias parciales del gen HA se utilizaron 119 secuencias del subtipo H5N2 aisladas entre los años 1994 y 2010, todas ellas mexicanas, igualmente obtenidas de GenBank®, para ambos análisis se utilizaron 2 secuencias subtipo H5N2 de cerdos de Corea como *outgroup*, así como las 3 muestras de 2014 secuenciadas para este trabajo.

En la figura 11 se observa la reconstrucción filogenética para las secuencias parciales del gen HA, estas se agrupan en ocho linajes, cuatro de los cuales fueron descritos en 2004 por Chang-Won Lee y cols. (Puebla, Jalisco, A y B), y cuatro nuevos (C, D, E y F), dentro de éstos últimos se agrupan las muestras secuenciadas en este trabajo. Las muestras JX-2 y JX-3 se agrupan en el linaje F mientras que la muestra JX-1 se agrupa en el linaje E, lo que supone distintos linajes circulando al mismo tiempo, resaltado en rojo en el linaje Jalisco se muestra la cepa vacunal (28159-232).

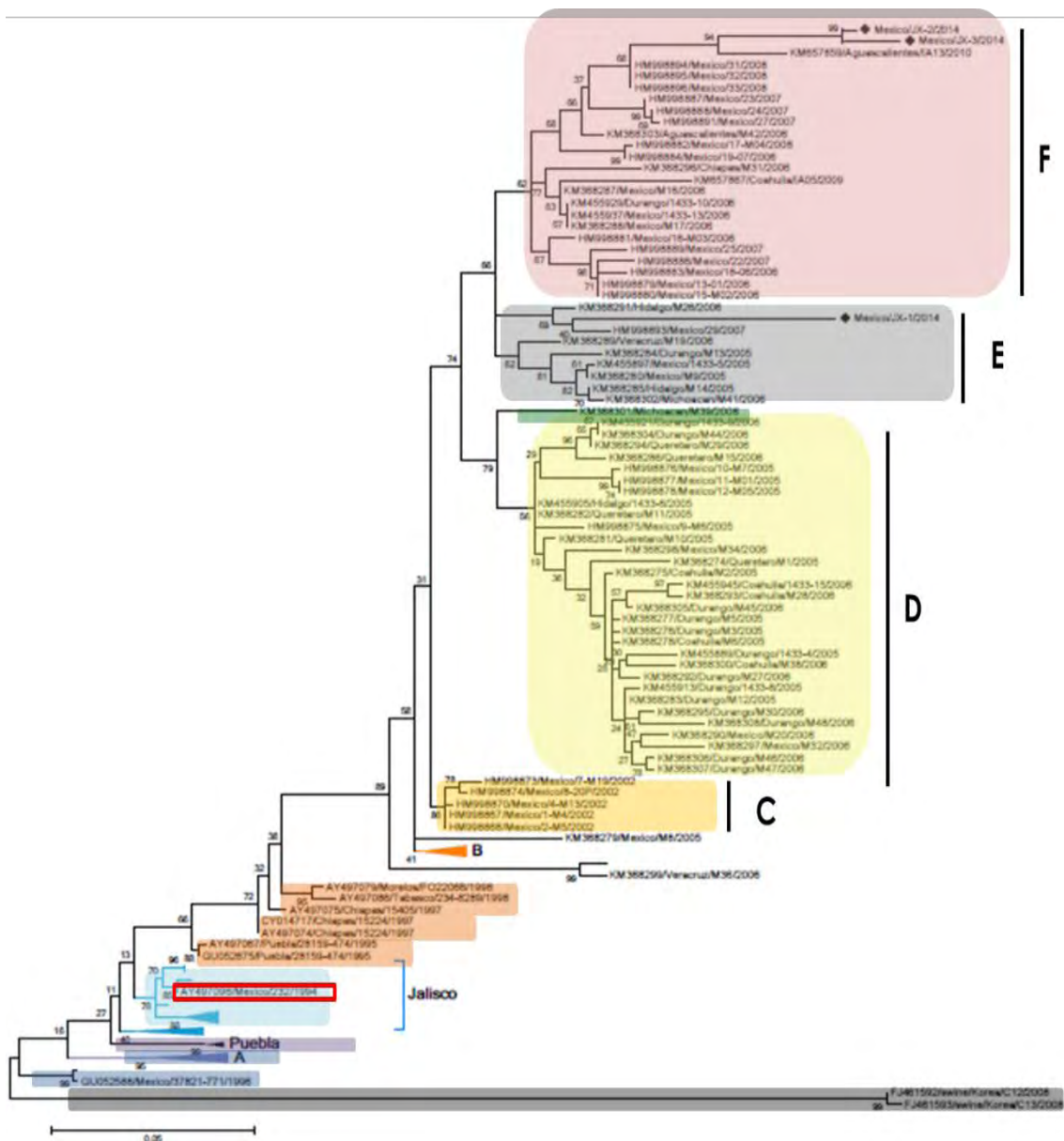


Figura 11. Reconstrucción filogenética de la de la secuencia parcial (472 - 1050) del gen HA del virus de influenza aviar A/H5N2 (n=119), muestras de 1994-2014, por el método de máxima verosimilitud (ML). Se indican los linajes en colores. Las secciones filogenéticas son identificadas por color: en morado linaje Puebla, en azul linaje A, en azul claro linaje Jalisco, en naranja linaje B, en gris claro linaje E y en rosa linaje F. Las muestras **indicadas con ♦ corresponden a las secuenciadas** en este trabajo, agrupadas en el linaje F.

Para el análisis filogenético del genoma completo del virus de influenza aviar se generaron ocho reconstrucciones filogenéticas para cada uno de los segmentos, se identificaron los 4 linajes descritos en 2004 por Chang-Won Lee y colaboradores (Puebla, Jalisco, A y B en color morado, azul claro, azul oscuro y naranja respectivamente) y dos nuevos linajes (E y F en color gris y rosa). En la reconstrucción filogenética del gen HA, JX-1, JX-2 y JX-3 se agrupan en el linaje F (Figura 12).

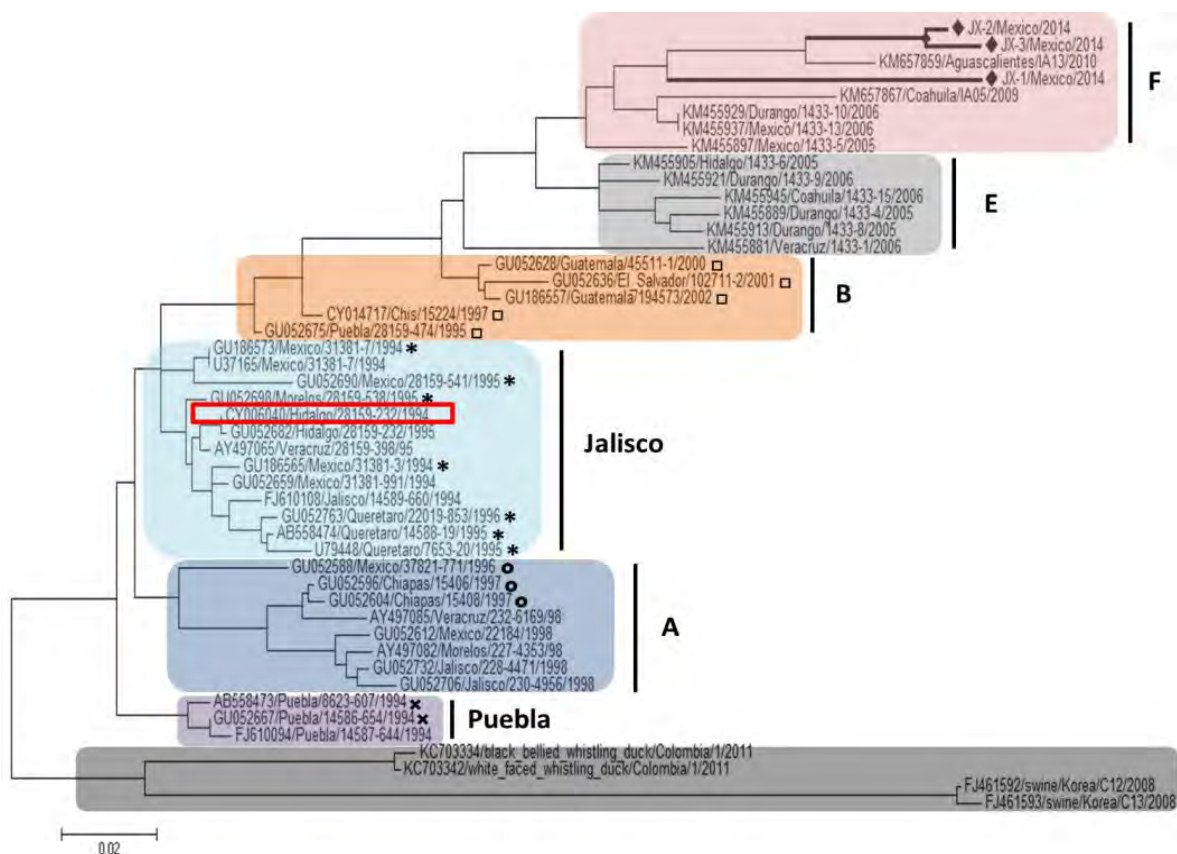


Figura 12. Reconstrucción filogenética del gen HA del virus de influenza aviar A/H5N2 (n=47), muestras de 1994-2014, por el método de máxima verosimilitud (ML). El esquema de colores corresponde al utilizado en la Figura 11. Las marcas \square , $*$, \circ , \times se utilizan para identificar las secuencias de los linajes B, Jalisco, A y Puebla respectivamente, utilizadas en el estudio de Chang-Won Lee. La cepa vacunal se señala en cuadro rojo, agrupada con el linaje Jalisco. Las muestras indicadas con \blacklozenge corresponden a las secuenciadas en este trabajo, agrupadas en el linaje F.

En la reconstrucción filogenética del gen NA la muestra JX-1 se agrupa en el linaje F, las muestras JX-2 y JX-3 se agrupan en el linaje E. Cepas pertenecientes al linaje E se agruparon en el linaje F. El linaje Puebla se encuentra agrupado con cepas pertenecientes a los linajes Jalisco y B, la cepa vacunal se agrupó en el linaje Jalisco (Figura 13).

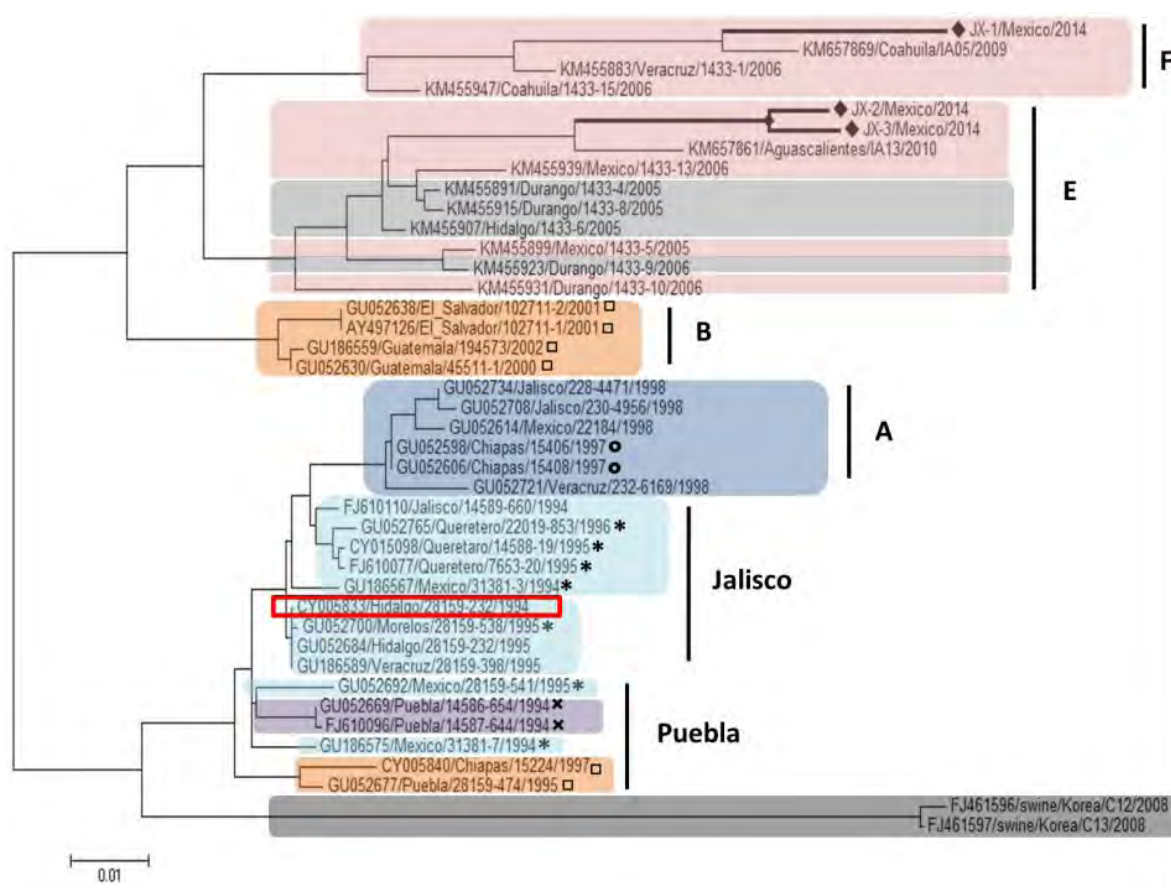


Figura 13. Reconstrucción filogenética del gen NA del virus de influenza aviar A/H5N2 (n=47), muestras de 1994-2014, por el método de máxima verosimilitud (ML). El esquema de colores y la identificación de las muestras corresponden al utilizado en la Figura 12. En el linaje Puebla se agrupan cepas del linaje B y Jalisco. El linaje E se agrupa con cepas del linaje F. La cepa vacunal se indica en rojo en el linaje Jalisco.

En la reconstrucción filogenética del gen PB2 la muestra JX-1 se agrupa en el linaje F, es de resaltar que las muestras JX-2 y JX-3 se agrupan con una de las cepas de virus de cerdos (C12), utilizadas como *outgroup*. Entre los linajes E y F se observó un reagrupamiento de las cepas entre ambos linajes. El linaje Puebla se encuentra agrupado con una cepa perteneciente al linaje Jalisco, mismo que se agrupó con cepas del linaje B. La cepa vacunal se agrupó en el linaje Jalisco (Figura 14).

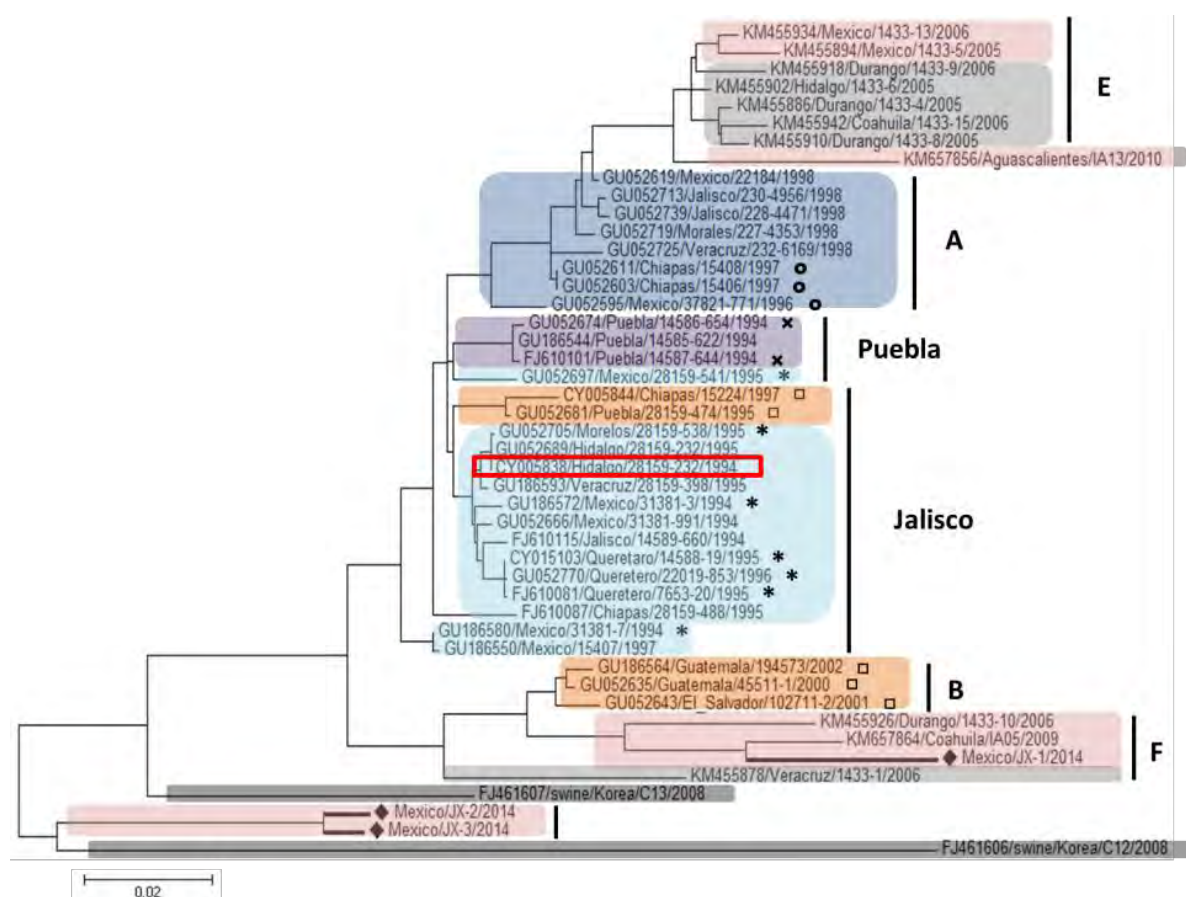


Figura 14. Reconstrucción filogenética del gen PB2 del virus de influenza aviar A/H5N2 (n=47), muestras de 1994-2014, por el método de máxima verosimilitud (ML). El esquema de colores y la identificación de las muestras corresponden al utilizado en la Figura 12. La muestra JX-1 se agrupa en el linaje F, las muestras JX-2 y JX-3 se encuentran lejanas a los linajes identificados. El linaje Jalisco se agrupa con cepas del linaje B.

En la reconstrucción filogenética del gen PB1 las muestras JX-1, JX-2 y JX-3 no se agruparon con los linajes aviarios mexicanos. JX-2 y JX-3 se encuentran más cercanas con las cepas pertenecientes a virus de cerdos utilizadas como *outgroup*, posiblemente por la falta de secuencias actuales; esta observación se apoya con los resultados observados en la construcción filogenética del gen PB2. El linaje Puebla se encuentra agrupado con cepas de los linajes A, B y Jalisco. Cepas de los linajes A, E y F se encuentran cercanas al linaje B. La cepa vacunal se agrupó en el linaje Jalisco (Figura 15).

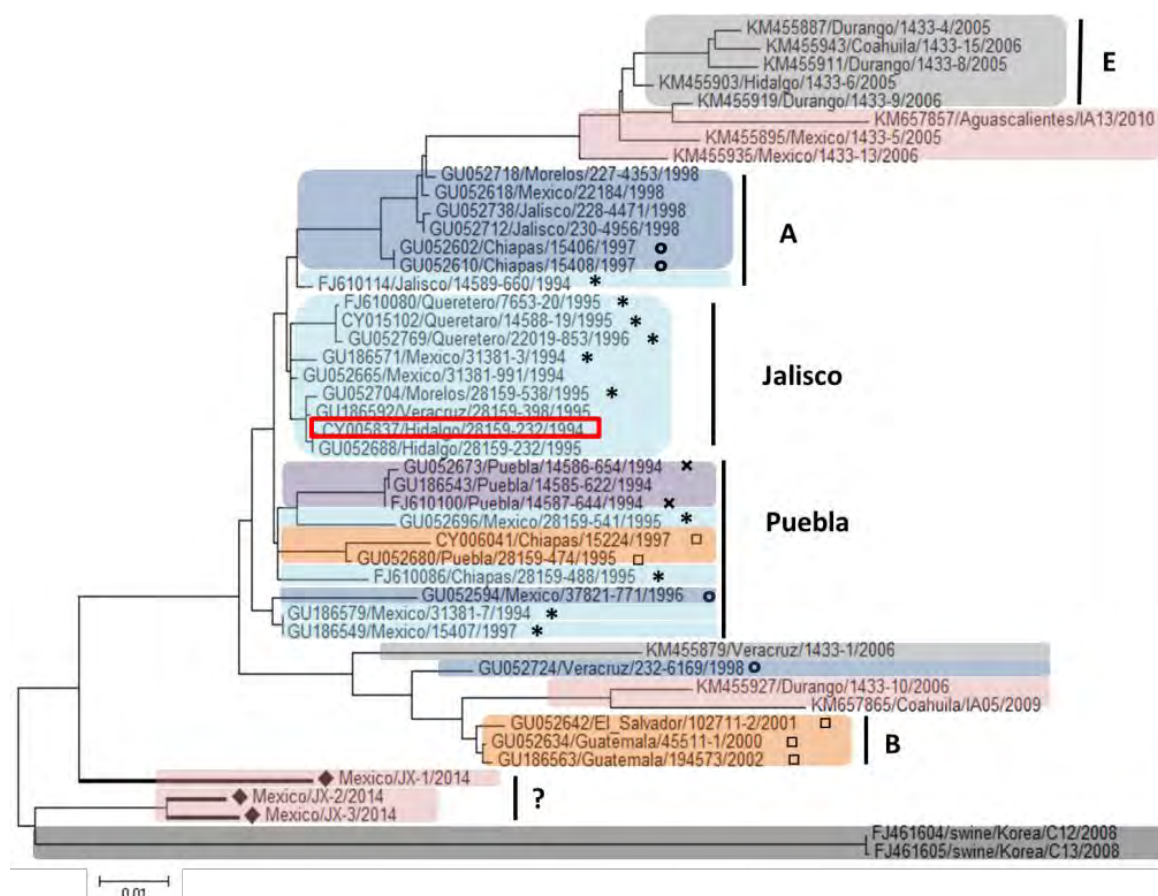


Figura 15. Reconstrucción filogenética del gen PB1 del virus de influenza aviar A/H5N2 (n=47), muestras de 1994-2014, por el método de máxima verosimilitud (ML). El esquema de colores y la identificación de las muestras corresponden al utilizado en la Figura 12. Cepas de linaje A, E y F cercanas al linaje B. La cepa vacunal se indica en rojo en el linaje Jalisco.

En la reconstrucción filogenética del gen PA las muestras JX-2 y JX-3 se agruparon en el linaje F, la muestra JX-1 se agrupó con cepas de virus de cerdos utilizadas como *outgroup*, lo que puede deberse a la falta de secuencias disponibles circulando actualmente; esta observación se apoya con los resultados observados en las reconstrucciones filogenéticas de los genes PB1 y PB2. El linaje Puebla se encuentra agrupado con cepas de los linajes A, B y Jalisco. Cepas de los linajes E y F se encuentran cercanas al linaje B. La cepa vacunal se agrupó en el linaje Jalisco (Figura 16).

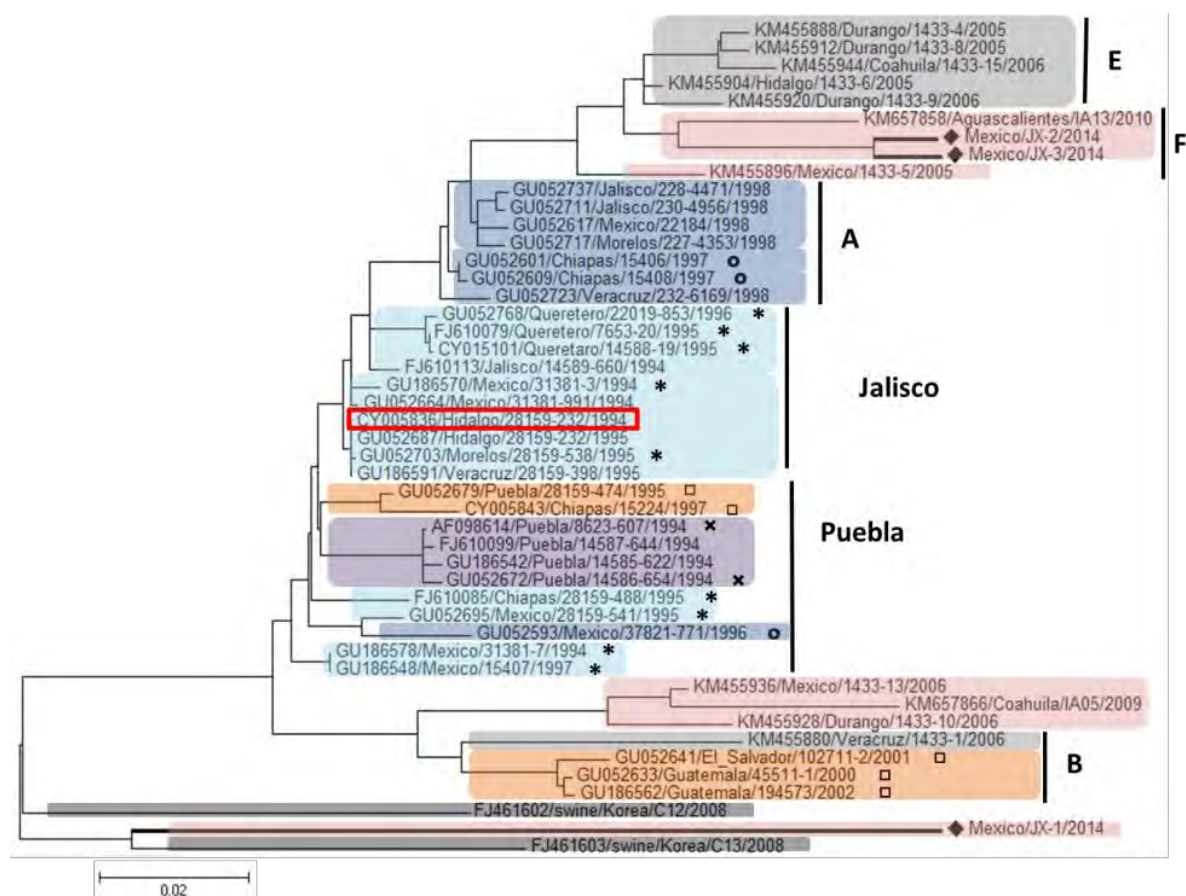


Figura 16. Reconstrucción filogenética del gen PA del virus de influenza aviar A/H5N2 (n=47), muestras de 1994-2014, por el método de máxima verosimilitud (ML). El esquema de colores y la identificación de las muestras corresponden al utilizado en la Figura 12. El linaje Puebla se agrupa con secuencias de los linajes A, B y Jalisco. La muestra JX-1 no se agrupa con los linajes.

La topología de la reconstrucción filogenética del gen NP fue muy similar a la del gen HA, las muestras JX-1, JX-2 y JX-3 se agruparon en el linaje F. Cepas del linaje Jalisco se encuentran lejanas al mismo y a los otros linajes, formando un clado separado de los linajes descritos. La cepa vacunal se agrupó en el linaje Jalisco (Figura 17).

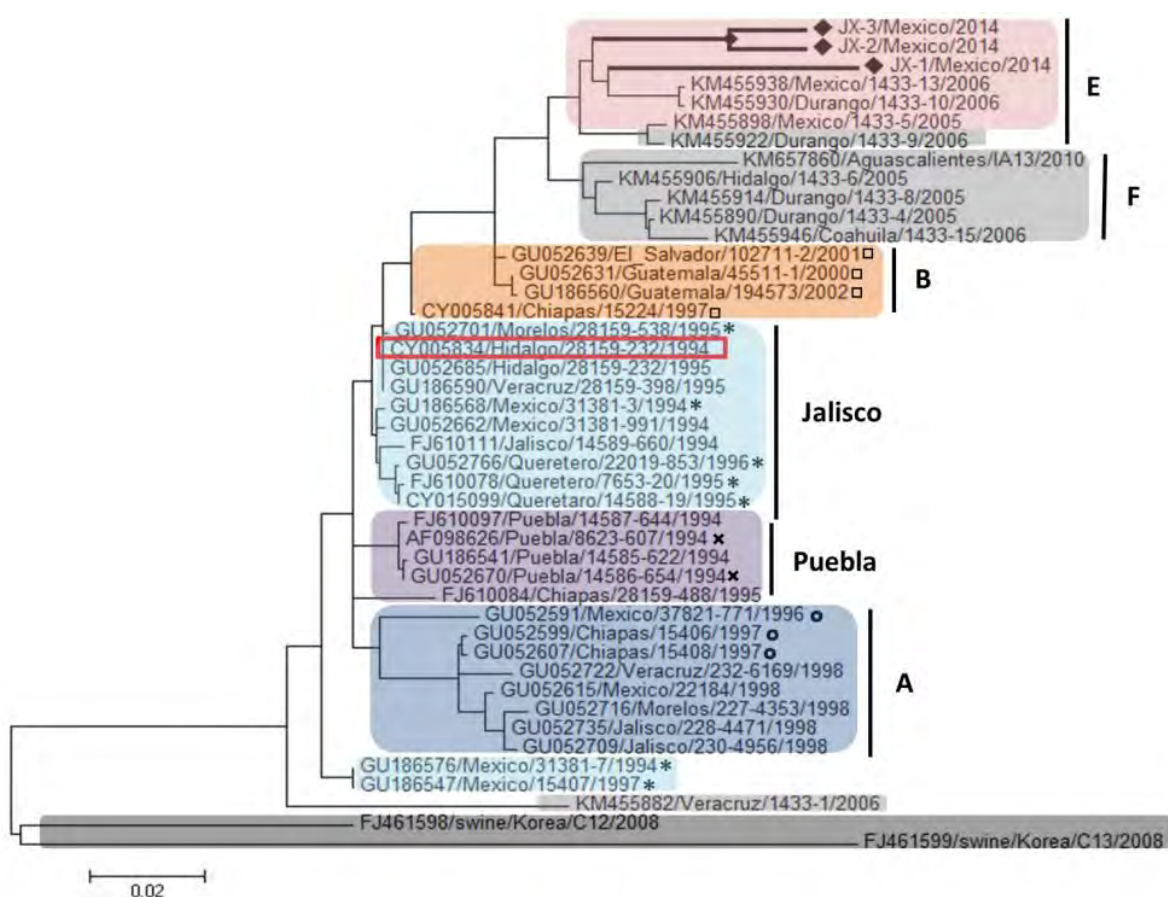


Figura 17. Reconstrucción filogenética del gen NP del virus de influenza aviar A/H5N2 (n=47), muestras de 1994-2014, por el método de máxima verosimilitud (ML). El esquema de colores y la identificación de las muestras corresponden al utilizado en la Figura 12. Cepas pertenecientes al linaje Jalisco se agrupan separadas en un clado diferente.

En la reconstrucción filogenética del gen M las muestras JX-1, JX-2 y JX-3 se agruparon en el linaje F. El caso particular de una cepa del linaje E, la cual se encuentra cercana a los linajes B, E y F, formando un clado separado representado por una sola secuencia, posiblemente por la falta de cepas secuenciadas pertenecientes a este linaje. El linaje A se encuentra agrupado con cepas del linaje Jalisco. La cepa vacunal se agrupó en el linaje Jalisco (Figura 18).

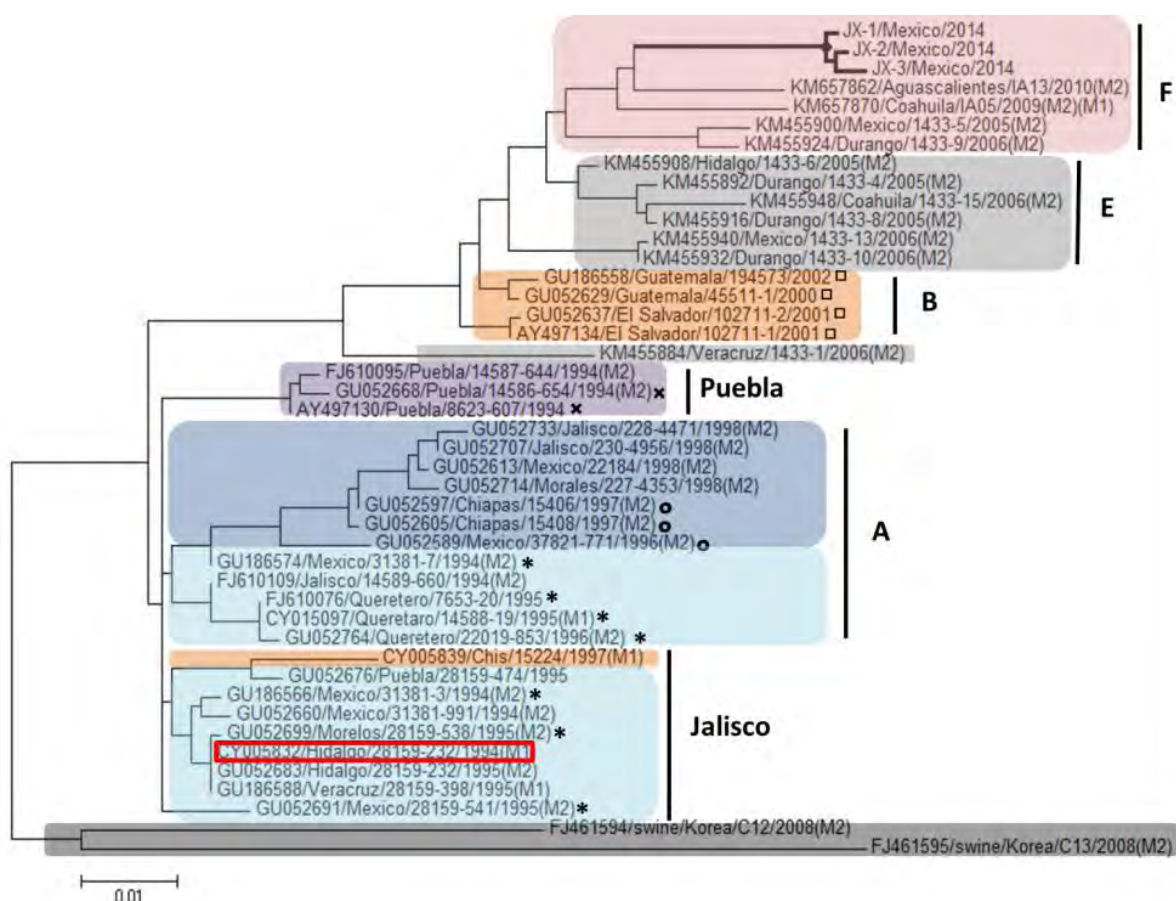


Figura 18. Reconstrucción filogenética del gen M del virus de influenza aviar A/H5N2 (n=47), muestras de 1994-2014, por el método de máxima verosimilitud (ML). El esquema de colores y la identificación de las muestras corresponden al utilizado en la Figura 12. El linaje A se encuentra agrupado con cepas del linaje Jalisco.

En la reconstrucción filogenética del gen NEP las muestras JX-1, JX-2 y JX-3 se agruparon en el linaje F. Cepas del linaje Puebla se agrupan con cepas de los linajes A y Jalisco. Para el linaje Jalisco las cepas forman dos clados separados sin embargo agrupadas con cepas del mismo linaje. La cepa vacunal se agrupó en el linaje Jalisco (Figura 19).

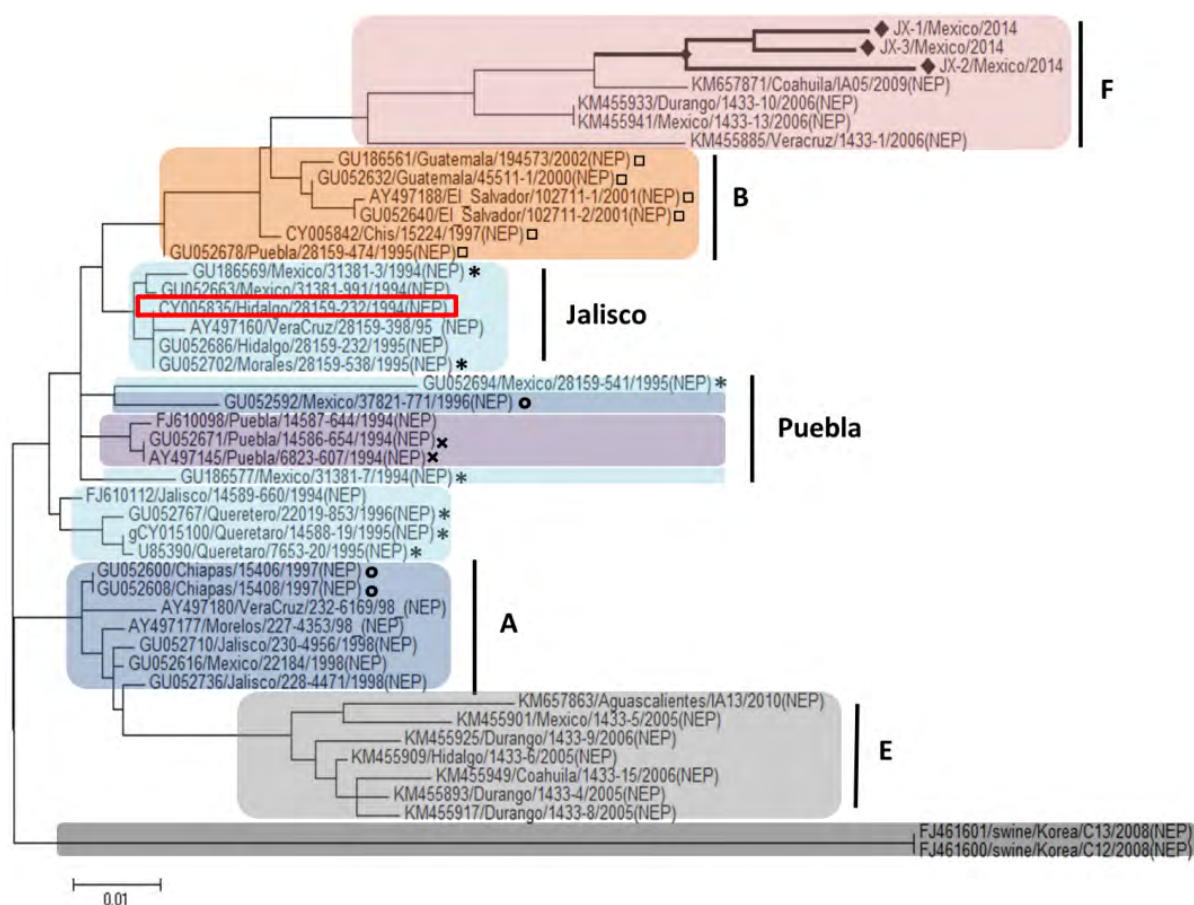


Figura 19. Reconstrucción filogenética de la proteína NEP del virus de influenza aviar A/H5N2 (n=47), muestras de 1994-2014, por el método de máxima verosimilitud (ML). El esquema de colores y la identificación de las muestras corresponden al utilizado en la Figura 12. El linaje Puebla se encuentra conformado con cepas de los linajes A y Jalisco. La cepa vacunal se indica en rojo en el linaje Jalisco.

9.6. Análisis de recombinación genómica

La muestra JX-1 presentó una identidad del 93% con las cepas 1433-5/2005, 1433-10/2006 y 1433-13/2006 mientras que las muestras JX-2 y JX-3 presentaron una identidad del 96% con la cepa IA13/2010 de Aguascalientes (Figura 20). La cepa de Guatemala 194573/2002 tiene una identidad del 96% con las cepa de Puebla 28159-474/1995. La cepa vacunal 28159-232/1994 tiene un porcentaje de identidad del 98% con cepas del linaje Jalisco (31381-7/1994, 31381-991/1994 y 31381-3/1994).

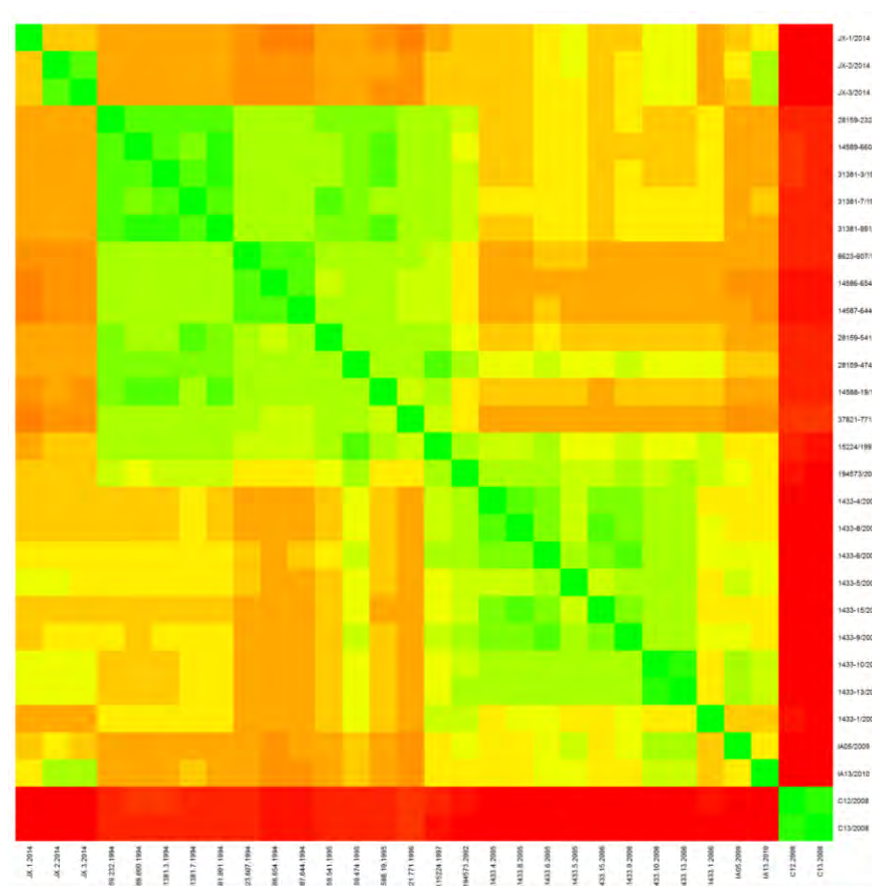


Figura 20. Heatmap del gen HA de cepas Mexicanas de influenza aviar A/H5N2 (n=47), muestras de 1994-2014. Las cepas son listadas en orden cronológico (a excepción de las muestras de este trabajo y las utilizadas como *outgroup*, que son las primeras tres y las últimas dos respectivamente) e identificadas con la clave de la identificación de la cepa y el año. La intensidad de color (verde = alto; rojo = bajo) refleja el nivel de identidad entre las cepas.

En la Figura 21 se muestra el heatmap del gen NA generado con los porcentajes de identidad de las cepas mexicanas H5N2, la muestra JX-1 tuvo un 96% de identidad con la cepa de Coahuila IA05/2009, mientras que las muestras JX-2 y JX-3 tuvieron un 95% de identidad con la cepa de Aguascalientes IA13/2010. La cepa vacunal (28159-232/1994) presentó una identidad del 100% con las cepas 28159-398/1995 y 28159-232/1995 de Veracruz e Hidalgo respectivamente, también pertenecientes al linaje Jalisco.

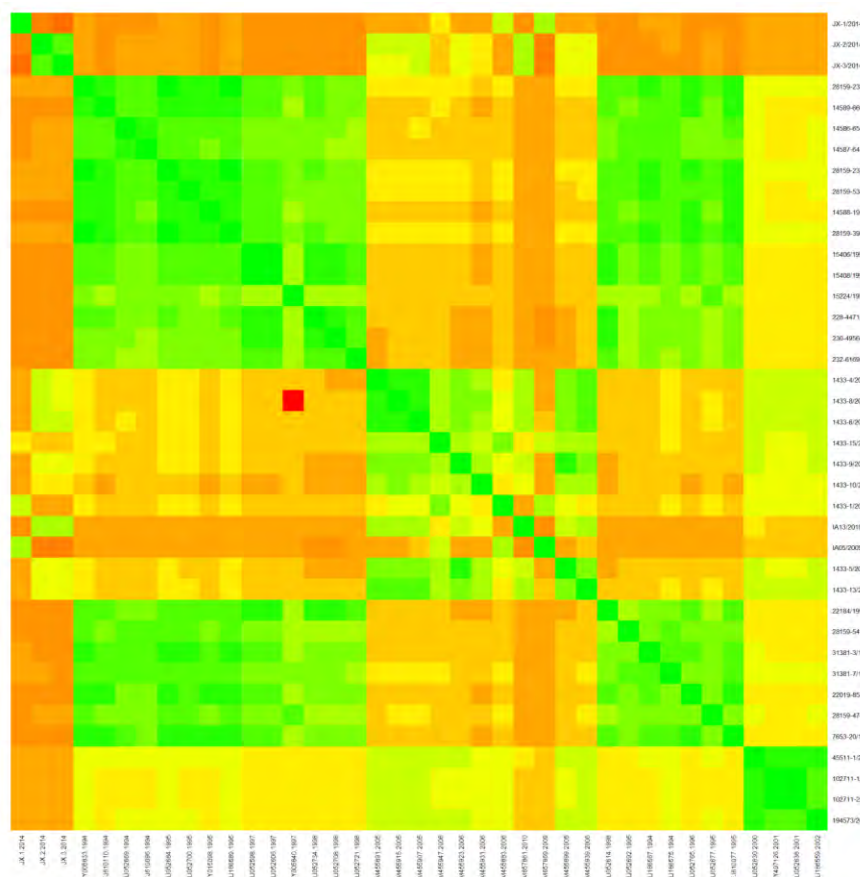


Figura 21. Heatmap del gen NA de cepas Mexicanas de influenza aviar A/H5N2 (n=47), muestras de 1994-2014. Las cepas son listadas en orden cronológico (a excepción de las muestras de este trabajo y las utilizadas como *outgroup*, que son las primeras tres y las últimas dos respectivamente) e identificadas con la clave de la identificación de la cepa y el año. La intensidad de color (verde = alto; rojo = bajo) refleja el nivel de identidad entre las cepas.

Las muestras JX-1, JX-2 y JX-3 presentaron un 96% de identidad con las cepas de Aguascalientes y Coahuila (IA13/2010, IA05/2009). La cepa de Veracruz 232-6169/1998 presentó el 71% de identidad. Las cepas de Guatemala (194573/2002, 45511-1/2000, 45511-1/2000) y El Salvador (102711-1/2001, 102711-2/2001) tuvieron un 98% de identidad con la cepa de Hidalgo 1433-6/2005. Al igual que con el gen NA, la cepa vacunal (28159-232/1994) presentó una identidad del 100% con las cepas 28159-398/1995 28159-232/1995 de Veracruz e Hidalgo respectivamente (Figura 22).

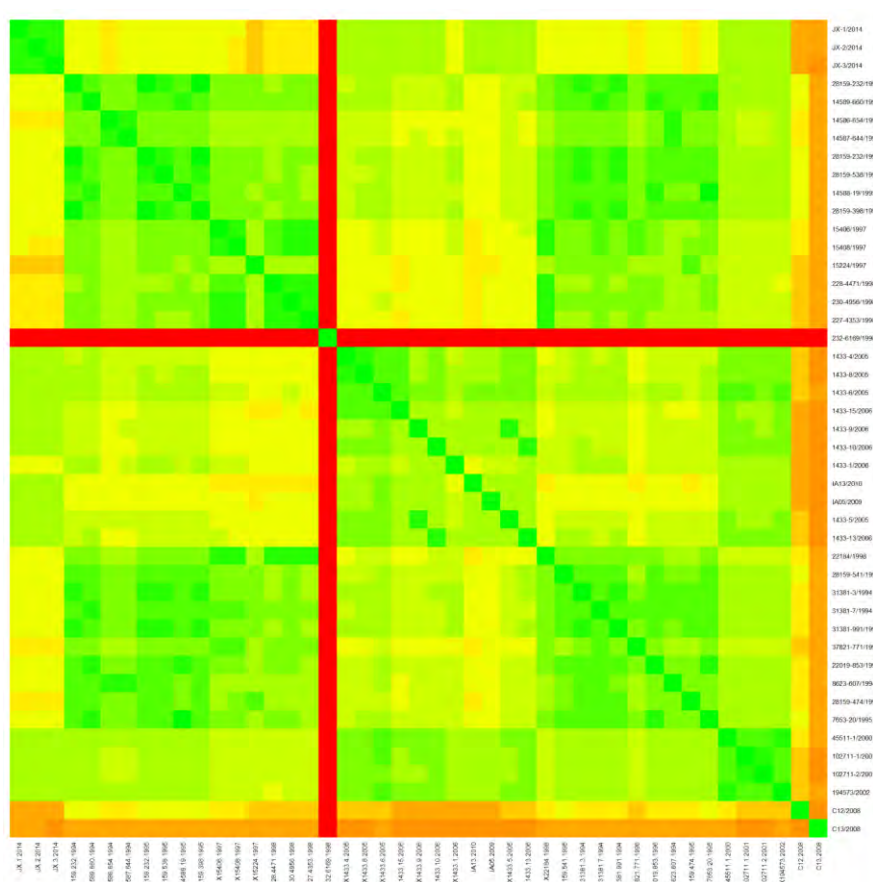


Figura 22. Heatmap del gen M de cepas Mexicanas de influenza aviar A/H5N2 (n=47), muestras de 1994-2014. Las cepas son listadas en orden cronológico (a excepción de las muestras de este trabajo y las utilizadas como *outgroup*, que son las primeras tres y las últimas dos respectivamente) e identificadas con la clave de la identificación de la cepa y el año. La intensidad de color (verde = alto; rojo = bajo) refleja el nivel de identidad entre las cepas.

En la Figura 24 se muestra el porcentaje de identidad entre las cepas para el gen NP. Para las muestras JX-1, JX-2 y JX-3 se tuvo un 95% de identidad con la muestra de Durango 1433-10/2006. Las cepas de Guatemala (194573/2002, 45511-1/2000, 45511-1/2000) y El Salvador (102711-1/2001, 102711-2/2001) tuvieron un 98% de identidad con la cepa de Chiapas 15224/1997. Al igual que con los segmentos NA, M y NEP, la cepa vacunal 28159-232/1994 presentó 100% de identidad con las cepas de Veracruz e Hidalgo.

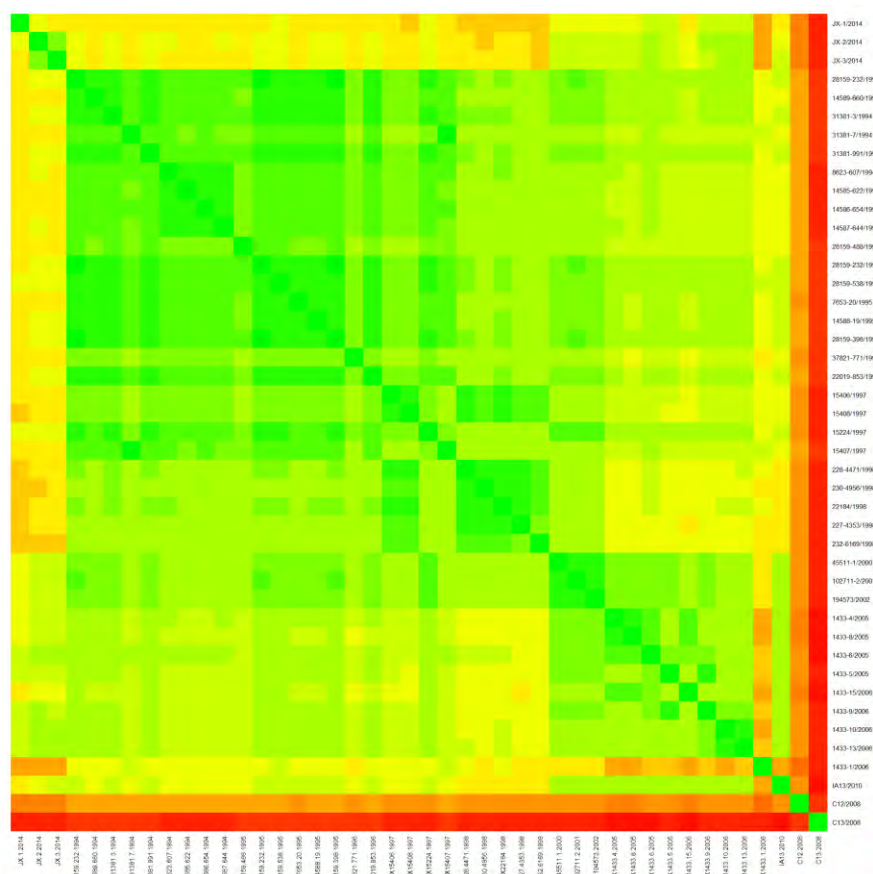


Figura 24. Heatmap del gen NP de cepas Mexicanas de influenza aviar A/H5N2 (n=47), muestras de 1994-2014. Las cepas son listadas en orden cronológico (a excepción de las muestras de este trabajo y las utilizadas como *outgroup*, que son las primeras tres y las últimas dos respectivamente) e identificadas con la clave de la identificación de la cepa y el año. La intensidad de color (verde = alto; rojo = bajo) refleja el nivel de identidad entre las cepas.

En la Figura 25 se muestra la identidad obtenida de las cepas mexicanas H5N2 para el gen PA. La muestra JX-1 presentó el 87% de identidad con cepas de los años 1994-1996, mientras que las muestras JX-2 y JX-3 presentaron el 95% de identidad con la cepa de Durango 1433-9/2006. Las cepas de Guatemala y El Salvador tuvieron un 95% de identidad con la cepa vacunal, que al igual que con los segmentos NA, M y NEP y NP, presentó 100% de identidad con las cepas de Veracruz e Hidalgo.

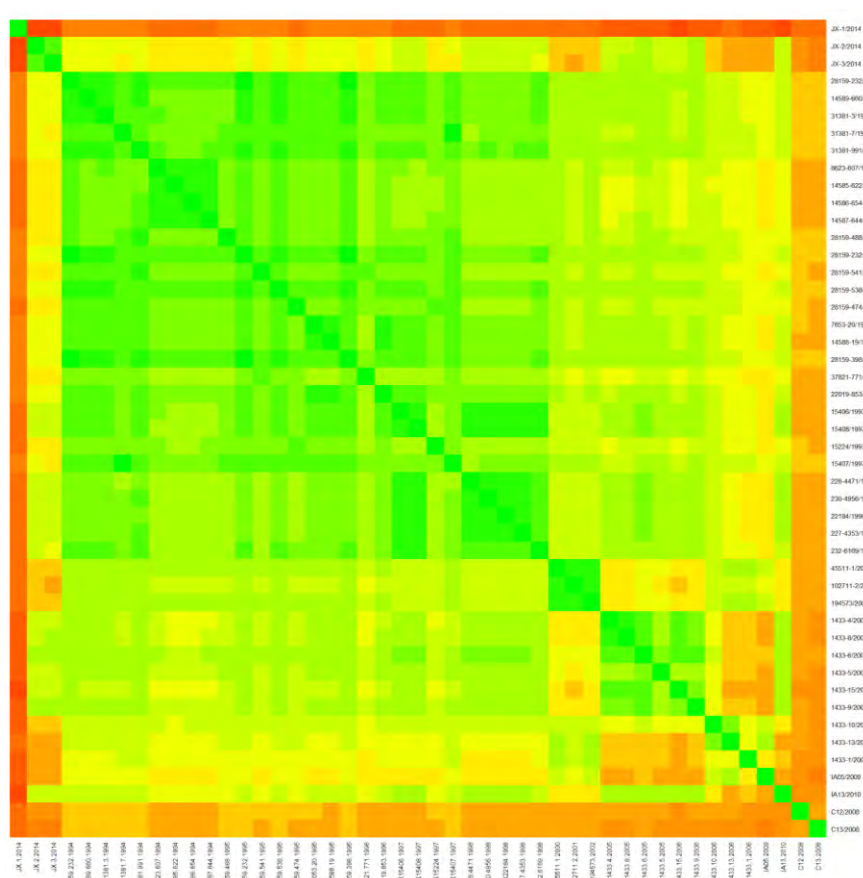


Figura 25. Heatmap del gen PA de cepas Mexicanas de influenza aviar A/H5N2 (n=47), muestras de 1994-2014. Las cepas son listadas en orden cronológico (a excepción de las muestras de este trabajo y las utilizadas como *outgroup*, que son las primeras tres y las últimas dos respectivamente) e identificadas con la clave de la identificación de la cepa y el año. La intensidad de color (verde = alto; rojo = bajo) refleja el nivel de identidad entre las cepas.

En la Figura 26 se observa el porcentaje de identidad de la secuencia nucleotídica entre las cepas mexicanas para el gen PB1. Las muestras JX-1, JX-2 y JX-3 tuvieron el 94% de identidad con cepas de 1994-1996. Las cepas de Guatemala y El Salvador presentaron el 98% de identidad con la cepa de Veracruz 232-6169/1998. La cepa vacunal presentó 99% de identidad con las cepas de Veracruz e Hidalgo. La cepa de Veracruz 1433-1/2006 tuvo el 94% de identidad con cepas de 1994-1998.

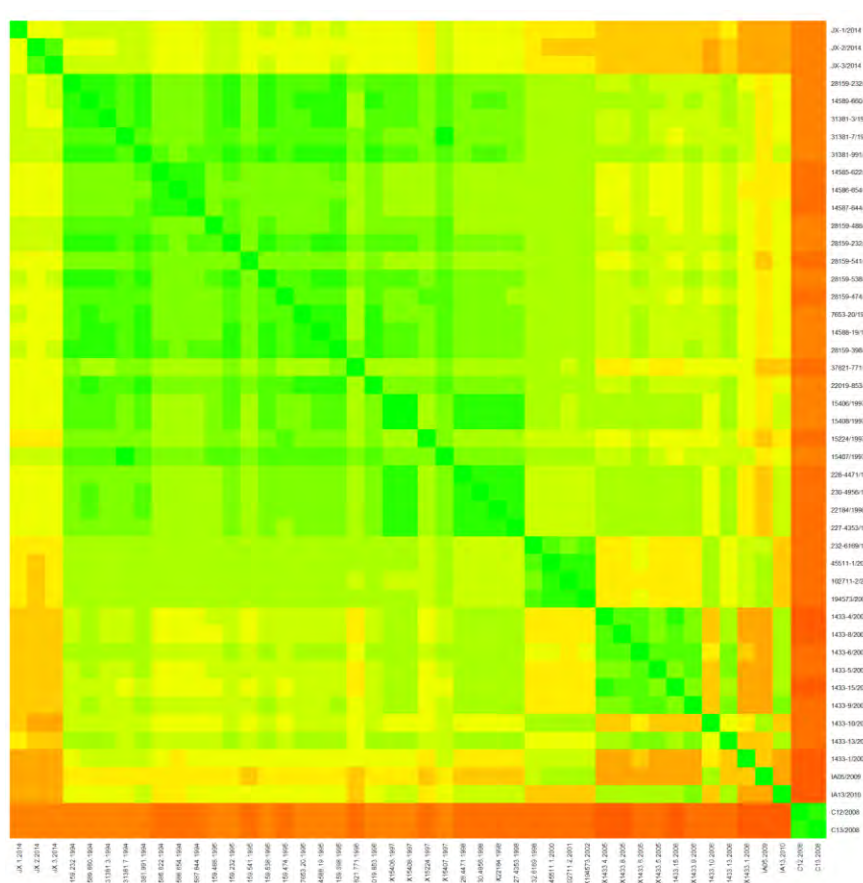


Figura 26. Heatmap del gen PB1 de cepas Mexicanas de influenza aviar A/H5N2 (n=47), muestras de 1994-2014. Las cepas son listadas en orden cronológico (a excepción de las muestras de este trabajo y las utilizadas como *outgroup*, que son las primeras tres y las últimas dos respectivamente) e identificadas con la clave de la identificación de la cepa y el año. La intensidad de color (verde = alto; rojo = bajo) refleja el nivel de identidad entre las cepas.

En el caso articular del heatmap para el gen PB2, las muestras JX-2 y JX-3 presentaron el 89% de identidad con muestras de 1994 a 1996, dentro de las cuales se encuentra la cepa vacunal (28159-232/1994), la cual presentó el 100% y el 99% de identidad con las cepas 28159-232/1995 y 28159-398/1995 respectivamente. La muestra JX-1 presentó el 95% de identidad con la muestra de Coahuila IA05/2009, mientras que las cepas de Guatemala y El Salvador tuvieron el 95% de identidad con la cepa 31381-7/1994 (Figura 27).

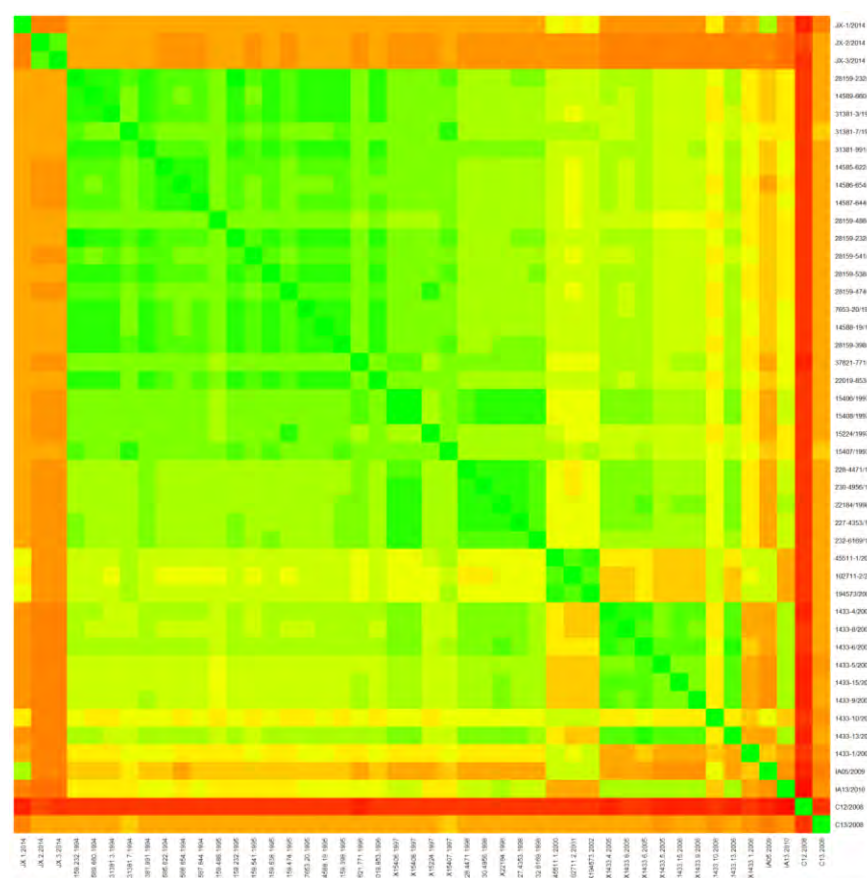


Figura 27. Heatmap del gen PB2 de cepas Mexicanas de influenza aviar A/H5N2 (n=47), muestras de 1994-2014. Las cepas son listadas en orden cronológico (a excepción de las muestras de este trabajo y las utilizadas como *outgroup*, que son las primeras tres y las últimas dos respectivamente) e identificadas con la clave de la identificación de la cepa y el año. La intensidad de color (verde = alto; rojo = bajo) refleja el nivel de identidad entre las cepas.

10. DISCUSIÓN

El uso de las tecnologías de secuenciación de siguiente generación (NGS) en la genómica, particularmente en estudios del genoma completo es una técnica de reciente aplicación en la virología⁷⁴. Se ha reflejado su impacto principalmente en dos ámbitos, reducción de costos y mayor rendimiento en la información generada, lo que ha llevado a diferentes grupos de trabajo a enfocar sus estudios bajo el uso de las mismas⁷⁵.

El empleo de estas herramientas ha presentado un impacto positivo en el área, contribuyendo en estudios tanto de evolución genómica⁷⁶ como en trabajos enfocados en identificar la diversidad genética en virus⁷⁷. A propósito de la contribución en el área de la virología con la incorporación de estas tecnologías enfocadas en el estudio del virus de la influenza, Suxiang Tong y colaboradores (cols.) identificaron en 2011 al murciélago como una nueva especie hospedera del virus de la influenza tipo A (H17N10, H18N11)³³, así mismo se han identificado un gran número de SNP (polimorfismo de un sólo nucleótido) que conllevan a sustituciones de aminoácidos en sitios antigénicos de la proteína HA, así como confirmar la presencia de patógenos previamente reportados por su papel en la infección en conjunto con el virus que resulta en la severidad de la enfermedad, como *Streptococcus pneumoniae*⁷⁸, sin embargo, existen grupos de trabajo que encuentran reservas en su utilización, principalmente debido al bajo número de lecturas obtenidas que corresponden al organismo de estudio⁷⁹.

Como principal objetivo de este trabajo nos planteamos utilizar estas tecnologías con la intención de identificar la estrategia óptima para la secuenciación del genoma completo de virus de interés veterinario. El genoma completo de 11 muestras fue sujeto a secuenciación masiva para posteriormente realizar el análisis de las secuencias ensambladas en conjunto con información de la

Base de Datos del virus de Influenza del NCBI®. El impacto y la importancia de la técnica se refleja en el impacto que tiene tanto en salud animal⁸⁰ y humana⁸¹, propiamente recae en el alcance de la misma para el estudio de la evolución y la identificación de nuevas especies de microorganismos aislados de muestras clínicas.

La cobertura obtenida de la secuenciación de las muestras fue suficiente para observar variación entre los linajes del virus de influenza aviar y poder realizar un análisis filogenético de cada uno de los segmentos genómicos a fin de determinar la evolución del genoma completo del virus a través del tiempo. Para este propósito el equipo de Illumina® presentó una ventaja, al generar lecturas del mismo tamaño, de 105 nt, lo que facilitó de manera importante la reconstrucción de las secuencias al realizar el análisis bioinformático de las mismas⁸². La herramienta de Illumina® tuvo mayor rendimiento en cobertura y profundidad (cuadro 16) que la herramienta de Thermo fisher® concordando con los resultados obtenidos en otros estudios⁸³.

Cabe destacar que en trabajos donde el alcance del proyecto este enfocado en determinar mutaciones claves en el objeto de estudio resulta conveniente obtener una profundidad mayor a 20 X⁸², así como una calidad arriba de 30 Q en las lecturas generadas que de confiabilidad al estudio, la plataforma de illumina® presentó esta ventaja (figura 4), ya que se obtuvo una tasa de error menor que con el ION PGM contrastando con los resultados obtenidos por Quail y cols. en 2012⁸³.

En 1995 Horimoto y cols. realizaron un análisis filogenético de la secuencia parcial de la proteína HA y concluyen que los virus mexicanos tienen como ancestro común a una cepa Norteamericana de 1991 circulando en aves silvestres acuáticas⁸⁴, lo que podría suponer el origen de este segmento más no de los restantes. Partiendo de estas premisas, este trabajo resalta la importancia y la necesidad de realizar estudios de evolución molecular del genoma completo del virus de influenza aviar

periódicamente con el objetivo de generar información actualizada de las cepas circulando actualmente en nuestro país.

Bajo el enfoque del estudio del genoma completo del virus de influenza se ha estudiado el proceso de adaptación del virus de aves silvestres a domésticas⁸⁵, eventos de reordenamiento del virus en el tiempo que han tenido impacto en salud pública a nivel global⁸⁶, se ha logrado la identificación del origen de cada uno de los segmentos de su genoma⁸⁷, sin embargo los estudios que se han realizado previamente en México, particularmente de las cepas aviares mexicanas H5N2 se han enfocado exclusivamente en el análisis de la secuencia nucleotídica de una región de sus proteínas, principalmente las proteínas de superficie HA y NA. Actualmente en GenBank® se tienen 577 secuencias completas de los diferentes segmentos del virus de cepas Mexicanas A/H5N2, sin embargo solamente se tienen 47 genomas completos disponibles, siendo la más reciente la cepa Aguascalientes/IA13 del año 2010⁸⁸. En el presente estudio se utilizaron estas tecnologías NGS con el fin de obtener la secuencia completa del genoma del virus de influenza aviar H5N2 de cepas circulando actualmente en nuestro país, logrando de esta manera contribuir con información genómica de cepas mexicanas actuales.

Como objetivo adicional del trabajo se realizó un análisis filogenético para estudiar la diversidad genética del virus y la evolución molecular que ha presentado su genoma en el tiempo desde el brote de 1994 en nuestro país.

No obstante que el análisis filogenético de los 8 segmentos del virus de influenza aviar exhibieron patrones de evolución similar, encontramos sólida evidencia filogenética de eventos de posible reordenamiento afectando la mayoría de los segmentos del genoma del virus (Figuras 13, 14, 15, 16, 18 y 19), esto principalmente derivado de la co-circulación de distintos linajes de cepas mexicanas H5N2, lo anterior es apoyado por el estudio de Holmes y cols. en virus humanos H3N2, que remarcan

el papel de la co-circulación de distintos linajes en el tiempo lo cual conlleva a frecuentes eventos de reordenamiento dentro del mismo subtipo⁸⁹.

Adicionalmente a los linajes previamente descritos Puebla, Jalisco, A y B⁵⁰ identificamos 4 nuevos linajes C, D, E y F (Figura 11) en el análisis de la secuencia parcial del segmento 4 del virus (HA) y únicamente 2 de estos (E y F) en el análisis de la secuencia completa de cada uno de los segmentos genómicos, probablemente como resultado de eventos de deriva génica a la que el virus se ha visto expuesto en nuestro país, uno de los factores principales que posiblemente ha contribuido a estos eventos es el programa de vacunación implementado en aves de corral desde el brote de 1994 con el fin de controlar y erradicar los virus LPAI H5N2, lo que ha permitido al virus evadir la respuesta inmune del huésped como ocurre con los virus humanos donde frecuentemente el virus se enfrenta a esta presión selectiva²⁹.

También se realizaron gráficas *heatmap* de cada uno de los segmentos genómicos del virus para proveer una medida cuantitativa del flujo génico entre cepas de diferentes años derivado de eventos de reordenamiento de los segmentos genómicos. Para la construcción de los gráficos se partió de los porcentajes de identidad de los alineamientos generados para las construcciones filogenéticas de cada uno de los segmentos genómicos del virus. Interesantemente para el segmento HA y NA de linajes actuales del virus se encuentra mayor identidad con cepas que prevalecen en el tiempo de años de 1996 y 1998 (Figura 20 y 21).

Nuestros análisis sugieren que eventos de reordenamiento genético dentro de cepas del mismo subtipo A/H5N2 están ocurriendo en México, dentro de los ocho segmentos genómicos del virus, los cuales han presentado un papel importante en la historia evolutiva del virus A/H5N2 en México.

11. CONCLUSIONES

Con los resultados obtenidos podemos concluir que con las 2 plataformas utilizadas en el estudio se logró secuenciar los genomas del virus de Newcastle y el virus de influenza aviar H5N2, sin embargo el equipo de illumina® presentó las ventajas de generar mayor profundidad por muestra, mayor cobertura del genoma, mayor calidad en las lecturas y lecturas del mismo tamaño.

Una de las variables más significativa entre los dos equipos de secuenciación utilizados en este trabajo fue la calidad en las lecturas, Q 34 para illumina® por Q 28 obtenida con Ion PGM®, sin embargo concluimos que este factor puede verse mitigado con una profundidad mayor a 20 X para genomas virales.

Al no obtener la cobertura del genoma completo, se puede adoptar la estrategia de ensamblado con genoma de referencia, para este propósito la calidad y el tamaño de las lecturas es relevante, ya que requirió de menor esfuerzo el tratar la información homologada en tamaños que el reto que representó el ensamblado a partir de lecturas de tamaño variable como las generadas por el equipo Ion PGM®.

En cuanto al software podemos concluir que la herramienta comercial *CLC Genomics Workbench 7®* presenta la ventaja de la interface de usuario frente al software de uso comercial como la herramienta Velvet, adicional a que con ambas herramientas se obtuvieron resultados similares en la generación de los ensamblados de las lecturas.

Si bien el análisis filogenético del virus de influenza aviar A/H5N2 indica la presencia de dos nuevos linajes que se conservan en las reconstrucciones filogenéticas de los ocho segmentos del genoma del virus, consideramos que es necesario contar con un banco de secuencias del virus por región y por periodo de tiempo para poder confirmarlo.

Es importante generar información actualizada a fin de contar con una base de datos de virus A/H5N2 circulando en México.

Las muestras secuenciadas de influenza aviar se agrupan en linajes separados al linaje vacunal, esto puede indicar que eventos de deriva antigénica multilinaje están ocurriendo en el linaje mexicano H5N2. Posiblemente la persistencia de los virus de influenza aviar H5N2 en el campo está viéndose favorecida por su diferencia con la cepa vacunal.

Debido al patrón en los heatmaps podemos concluir que los virus actuales presentan una tasa de mutación diferente para cada segmento genómico del virus, así mismo nuestros resultados sugieren la conservación de segmentos genómicos de diferentes años en linajes actuales del virus lo cual puede deberse al favorecimiento en la presión selectiva para algunos segmentos del genoma del virus, sin embargo para confirmarlo se requiere mayor estudio en el tema.

12. REFERENCIAS

- ¹ Dahm, R. (2005). Friedrich Miescher and the discovery of DNA. *Developmental Biology*, 278(2), 274-288.
- ² Levene, P. A. (1919). The structure of yeast nucleic acid. *Studies from the Rockefeller Institute for Medical Research*, 30, 221.
- ³ Avery, O. T., MacLeod, C. M., & McCarty, M. (1944). Studies on the chemical nature of the substance inducing transformation of pneumococcal types induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. *The Journal of experimental medicine*, 79(2), 137-158.
- ⁴ Hershey, A. D., & Chase, M. (1952). Independent functions of viral protein and nucleic acid in growth of bacteriophage. *The Journal of general physiology*, 36(1), 39-56.
- ⁵ Watson, J. D., & Crick, F. H. (1953). Molecular structure of nucleic acids. *Nature*, 171(4356), 737-738.
- ⁶ Wilkins, M. H. F., Stokes, A. R., & Wilson, H. R. (1953). Molecular structure of nucleic acids: molecular structure of deoxypentose nucleic acids. *Nature*, 171(4356), 738-740.
- ⁷ Fiers, W., Contreras, R., Duerinck, F., Haegeman, G., Iserentant, D., Merregaert, J., ... & Ysebaert, M. (1976). Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene.
- ⁸ Maxam, A. M., & Gilbert, W. (1977). A new method for sequencing DNA. *Proceedings of the National Academy of Sciences*, 74(2), 560-564.
- ⁹ Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, 74(12), 5463-5467.

-
- ¹⁰ Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., ... & Erlich, H. A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, 239(4839), 487-491.
- ¹¹ Innis, M. A., Myambo, K. B., Gelfand, D. H., & Brow, M. A. (1988). DNA sequencing with *Thermus aquaticus* DNA polymerase and direct sequencing of polymerase chain reaction-amplified DNA. *Proceedings of the National Academy of Sciences*, 85(24), 9436-9440.
- ¹² Smith, L. M., Sanders, J. Z., Kaiser, R. J., Hughes, P., Dodd, C., Connell, C. R., ... & Hood, L. E. (1986). Fluorescence detection in automated DNA sequence analysis.
- ¹³ Barzon, L., Lavezzo, E., Militello, V., Toppo, S. and Pal`u, G. (2011). Applications of next-generation sequencing technologies to diagnostic virology. *International journal of molecular sciences*, 12(11), pp.7861--7884.
- ¹⁴ Lam, H. Y., Clark, M. J., Chen, R., Chen, R., Natsoulis, G., O'hualachain, M., ... & Butte, A. J. (2012). Performance comparison of whole-genome sequencing platforms. *Nature biotechnology*, 30(1), 78-82.
- ¹⁵ Quail, M., Swerdlow, H. and Turner, D. (2009). Improved protocols for the illumina genome analyzer sequencing system. *Current protocols in human genetics*, pp.18--2.
- ¹⁶ Hinz, W., et al. (2011). An integrated semiconductor device enabling non-optical genome sequencing. *Nature*. doi:10.1038/nature10242.
- ¹⁷ Sanger, F., & Coulson, A. R. (1975). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of molecular biology*, 94(3), 441-448.

-
- ¹⁸ Maxam, A. M., & Gilbert, W. (1977). A new method for sequencing DNA. *Proceedings of the National Academy of Sciences*, 74(2), 560-564.
- ¹⁹ Smith, L. M., Sanders, J. Z., Kaiser, R. J., Hughes, P., Dodd, C., Connell, C. R., ... & Hood, L. E. (1986). Fluorescence detection in automated DNA sequence analysis.
- ²⁰ Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlén, M., & Nyrén, P. (1996). Real-time DNA sequencing using detection of pyrophosphate release. *Analytical biochemistry*, 242(1), 84-89.
- ²¹ Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., ... & Volkmer, G. A. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 437(7057), 376-380.
- ²² Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., ... & Anastasi, C. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *nature*, 456(7218), 53-59.
- ²³ Mardis, E. R. (2008). The impact of next-generation sequencing technology on genetics. *Trends in genetics*, 24(3), 133-141.
- ²⁴ Valouev, A., Ichikawa, J., Tonthat, T., Stuart, J., Ranade, S., Peckham, H., ... & Johnson, S. M. (2008). A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning. *Genome research*, 18(7), 1051-1063.
- ²⁵ Pop, M. (2009). Genome assembly reborn: recent computational challenges. *Briefings in bioinformatics*, bbp026.
- ²⁶ Pop, M. (2009). Genome assembly reborn: recent computational challenges. *Briefings in bioinformatics*, bbp026.

-
- ²⁷ Miller, J. R., Koren, S., & Sutton, G. (2010). Assembly algorithms for next-generation sequencing data. *Genomics*, 95(6), 315-327.
- ²⁸ ICTV International Committee on Taxonomy of Viruses [Online] // International Committee on Taxonomy of Viruses. - ICTV, 2014. - 2014. - <http://ictvonline.org/virusTaxonomy.asp>.
- ²⁹ Bouvier, N. M., & Palese, P. (2008). The biology of influenza viruses. *Vaccine*, 26, D49-D53.
- ³⁰ Chen, W., Calvo, P. A., Malide, D., Gibbs, J., Schubert, U., Bacik, I., ... & Henklein, P. (2001). A novel influenza A virus mitochondrial protein that induces cell death. *Nature medicine*, 7(12), 1306-1312.
- ³¹ Suzuki, Y. (2006). Natural selection on the influenza virus genome. *Molecular biology and evolution*, 23(10), 1902-1911.
- ³² **Szewczyk, B., Bieńkowska**-Szewczyk, K., & Król, E. (2014). Introduction to molecular biology of influenza A viruses. *Acta Biochim Pol*, 61(3), 397-401.
- ³³ Tong, S., Li, Y., Rivailler, P., Conrardy, C., Castillo, D. A. A., Chen, L. M., ... & Turmelle, A. S. (2012). A distinct lineage of influenza A virus from bats. *Proceedings of the National Academy of Sciences*, 109(11), 4269-4274.
- ³⁴ Walther, T., Karamanska, R., Chan, R. W., Chan, M. C., Jia, N., Air, G., ... & Haslam, S. M. (2013). Glycomic analysis of human respiratory tract tissues and correlation with influenza virus infection. *PLoS Pathog*, 9(3), e1003223.
- ³⁵ Van Reeth, K. (2007). Avian and swine influenza viruses: our current understanding of the zoonotic risk. *Veterinary research*, 38(2), 243-260.

-
- ³⁶ Mehle, A. (2014). Unusual influenza A viruses in bats. *Viruses*, 6(9), 3438-3449.
- ³⁷ Matsuzaki, Y., Sugawara, K., Mizuta, K., Tsuchiya, E., Muraki, Y., Hongo, S., ... & Nakamura, K. (2002). Antigenic and genetic characterization of influenza C viruses which caused two outbreaks in Yamagata City, Japan, in 1996 and 1998. *Journal of clinical microbiology*, 40(2), 422-429.
- ³⁸ Osterhaus, A. D. M. E., Rimmelzwaan, G. F., Martina, B. E. E., Bestebroer, T. M., & Fouchier, R. A. M. (2000). Influenza B virus in seals. *Science*, 288(5468), 1051-1053.
- ³⁹ Smith, G. J., Vijaykrishna, D., Bahl, J., Lycett, S. J., Worobey, M., Pybus, O. G., ... & Rambaut, A. (2009). Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. *Nature*, 459(7250), 1122-1125.
- ⁴⁰ Boni, M. F. (2008). Vaccination and antigenic drift in influenza. *Vaccine*, 26, C8-C14.
- ⁴¹ Carrat, F., & Flahault, A. (2007). Influenza vaccine: the challenge of antigenic drift. *Vaccine*, 25(39), 6852-6862.
- ⁴² OMS Organización Mundial de la Salud [Online] Nota descriptiva N°211 Marzo de 2014 - <http://www.who.int/mediacentre/factsheets/fs211/es/>
- ⁴³ Salomon, R., & Webster, R. G. (2009). The influenza virus enigma. *Cell*, 136(3), 402-410.
- ⁴⁴ Smith, G. J., et al (2009). Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. *Nature*. doi:10.1038/nature08182
- ⁴⁵ Alexander, D. J. (2000). A review of avian influenza in different bird species. *Veterinary microbiology*, 74(1), 3-13.
- ⁴⁶ Centers for Disease Control and Prevention (CDC. (1998). Update: isolation of avian influenza A (H5N1) viruses from humans--Hong Kong, 1997-1998. *MMWR. Morbidity and mortality weekly report*, 46(52-53), 1245.

-
- ⁴⁷ Claas, E. C., Osterhaus, A. D., Van Beek, R., De Jong, J. C., Rimmelzwaan, G. F., Senne, D. A., ... & Webster, R. G. (1998). Human influenza A H5N1 virus related to a highly pathogenic avian influenza virus. *The Lancet*, 351(9101), 472-477.
- ⁴⁸ Bender, C., Hall, H., Huang, J., Klimov, A., Cox, N., Hay, A., ... & Subbarao, K. (1999). Characterization of the surface proteins of influenza A (H5N1) viruses isolated from humans in 1997–1998. *Virology*, 254(1), 115-123.
- ⁴⁹ Li, K. S., Guan, Y., Wang, J., Smith, G. J. D., Xu, K. M., Duan, L., ... & Peiris, J. S. M. (2004). Genesis of a highly pathogenic and potentially pandemic H5N1 influenza virus in eastern Asia. *Nature*, 430(6996), 209-213.
- ⁵⁰ Lee, C. W., Senne, D. A., & Suarez, D. L. (2004). Effect of vaccine use in the evolution of Mexican lineage H5N2 avian influenza virus. *Journal of virology*, 78(15), 8372-8381.
- ⁵¹ Escorcia, M., Vázquez, L., Méndez, S. T., Rodríguez-Ropón, A., Lucio, E., & Nava, G. M. (2008). Avian influenza: genetic evolution under vaccination pressure. *Virol J*, 5(1), 15.
- ⁵² Alexander, D. J. (2007). Summary of avian influenza activity in Europe, Asia, Africa, and Australasia, 2002-2006. *Avian diseases*, 51(s1), 161-166.
- ⁵³ CDC (2014). Highly pathogenic avian influenza a (H5N1) in people.
- ⁵⁴ Zhou, N. N., Shortridge, K. F., Claas, E. C., Krauss, S. L., & Webster, R. G. (1999). Rapid evolution of H5N1 influenza viruses in chickens in Hong Kong. *Journal of virology*, 73(4), 3366-3374.
- ⁵⁵ Matrosovich, M., Zhou, N., Kawaoka, Y., & Webster, R. (1999). The surface glycoproteins of H5 influenza viruses isolated from humans, chickens, and wild aquatic birds have distinguishable properties. *Journal of virology*, 73(2), 1146-1155.

-
- ⁵⁶ Villareal, C. L., & Flores, A. O. (2003). The Mexican avian influenza (H5N2) outbreak. *Avian Diseases*, 47, 18-22.
- ⁵⁷ Suarez, D. L., Spackman, E., & Senne, D. A. (2003). Update on molecular epidemiology of H1, H5, and H7 influenza virus infections in poultry in North America. *Avian Diseases*, 47(s3), 888-897.
- ⁵⁸ Munster, V., et al. (2009). Pathogenesis and transmission of swine-origin 2009 A (H1N1) influenza virus in ferrets. *Science*, 325(5939), pp.481--483.
- ⁵⁹ Van Reeth, K., De Vleeschauwer, A., Kyriakis, K. and Pensaert, M. (2006). Influenza in birds, pigs and humans: old theories versus current viewpoints. *Proceedings of the 19th IPVS Congress*, Copenhagen, Denmark.
- ⁶⁰ Myers, K., Olsen, C. and Gray, G. (2007). Cases of swine influenza in humans: a review of the literature. *Clinical infectious diseases*, 44(8), pp.1084--1088.
- ⁶¹ Alexander, D. J. (2000). Newcastle disease and other avian paramyxoviruses. *Revue Scientifique et Technique-Office International des Epizooties*, 19(2), 443-455.
- ⁶² Seal, B. S., King, D. J., & Sellers, H. S. (2000). The avian response to Newcastle disease virus. *Developmental & Comparative Immunology*, 24(2), 257-268.
- ⁶³ Thermo Fisher Scientific Inc. © (2015). **Ion PGM™ Hi-Q™** Sequencing Kit User Guide.
- ⁶⁴ Thermo Fisher Scientific Inc. © (2014). **PathAmp™ FluA Reagents** User Guide.
- ⁶⁵ Minoche, A. E., Dohm, J. C., & Himmelbauer, H. (2011). Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and genome analyzer systems. *Genome biology*, 12(11), R112.
- ⁶⁶ FASTQC High Throughput Sequence QC Report© (2011). Simon Andrews, Pierre Lindebaum, Brian Howard & Phil Ewels. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

-
- ⁶⁷ Ewing, B., Hillier, L., Wendl, M. C., & Green, P. (1998). Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. *Genome research*, 8(3), 175-185.
- ⁶⁸ Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular biology and evolution*, 30(12), 2725-2729.
- ⁶⁹ Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic acids research*, 22(22), 4673-4680.
- ⁷⁰ Kumar, S., & Filipski, A. (2007). Multiple sequence alignment: in pursuit of homologous DNA positions. *Genome research*, 17(2), 127-135.
- ⁷¹ Felsenstein, J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of molecular evolution*, 17(6), 368-376.
- ⁷² Eisen, M. B., Spellman, P. T., Brown, P. O., & Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences*, 95(25), 14863-14868.
- ⁷³ Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., ... & Hornik, K. (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome biology*, 5(10), R80.
- ⁷⁴ Aherfi, S., Boughalmi, M., Pagnier, I., Fournous, G., La Scola, B., Raoult, D., & Colson, P. (2014). Complete genome sequence of Tunisvirus, a new member of the proposed family Marseilleviridae. *Archives of virology*, 159(9), 2349-2358.

⁷⁵ Capobianchi, M. R., Giombini, E., & Rozera, G. (2013). Next generation sequencing technology in clinical virology. *Clinical Microbiology and Infection*, 19(1), 15-22.

⁷⁶ Croville, G., Soubies, S. M., Barbieri, J., Klopp, C., Mariette, J., Bouchez, O., ... & Guérin, J. L. (2012). Field monitoring of avian influenza viruses: whole-genome sequencing and tracking of neuraminidase evolution using 454 pyrosequencing. *Journal of clinical microbiology*, 50(9), 2881-2887.

⁷⁷ Rosseel, T., Lambrecht, B., Vandenbussche, F., van den Berg, T., & Van Borm, S. (2011). Identification and complete genome sequencing of paramyxoviruses in mallard ducks (*Anas platyrhynchos*) using random access amplification and next generation sequencing technologies. *Virology journal*, 8(1), 1.

⁷⁸ Kuroda, M., Katano, H., Nakajima, N., Tobiume, M., Aina, A., Sekizuka, T., ... & Hata, S. (2010). Characterization of quasispecies of pandemic 2009 influenza A virus (A/H1N1/2009) by de novo sequencing using a next-generation DNA sequencer. *PLoS One*, 5(4), e10256.

⁷⁹ Kuroda, M., Katano, H., Nakajima, N., Tobiume, M., Aina, A., Sekizuka, T., ... & Hata, S. (2010). Characterization of quasispecies of pandemic 2009 influenza A virus (A/H1N1/2009) by de novo sequencing using a next-generation DNA sequencer. *PLoS One*, 5(4), e10256.

⁸⁰ Ludwig, G. V., Calle, P. P., Mangiafico, J. A., Raphael, B. L., Danner, D. K., Hile, J. A., ... & McNamara, T. (2002). An outbreak of West Nile virus in a New York City captive wildlife population. *The American journal of tropical medicine and hygiene*, 67(1), 67-75.

⁸¹ Grard, G., Fair, J. N., Lee, D., Slikas, E., Steffen, I., Muyembe, J. J., ... & Makuwa, M. (2012). A novel rhabdovirus associated with acute hemorrhagic fever in central Africa. *PLoS Pathog*, 8(9), e1002924.

⁸² Seto, D. (2010). Viral genomics and bioinformatics. *Viruses*, 2(12), 2587-2593.

⁸³ Quail, M. A., Smith, M., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., ... & Gu, Y. (2012). A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC genomics*, 13(1), 1.

⁸⁴ Horimoto, T., Rivera, E., Pearson, J., Senne, D., Krauss, S., Kawaoka, Y., & Webster, R. G. (1995). Origin and molecular changes associated with emergence of a highly pathogenic H5N2 influenza virus in Mexico. *Virology*, 213(1), 223-230.

⁸⁵ Croville, G., Soubies, S. M., Barbieri, J., Klopp, C., Mariette, J., Bouchez, O., ... & Guérin, J. L. (2012). Field monitoring of avian influenza viruses: whole-genome sequencing and tracking of neuraminidase evolution using 454 pyrosequencing. *Journal of clinical microbiology*, 50(9), 2881-2887.

⁸⁶ Nelson, M. I., Viboud, C., Simonsen, L., Bennett, R. T., Griesemer, S. B., George, K. S., ... & Taubenberger, J. K. (2008). Multiple reassortment events in the evolutionary history of H1N1 influenza A virus since 1918. *PLoS Pathog*, 4(2), e1000012.

⁸⁷ Pasick, J., Berhane, Y., Joseph, T., Bowes, V., Hisanaga, T., Handel, K., & Alexandersen, S. (2015). Reassortant highly pathogenic influenza A H5N2 virus containing gene segments related to Eurasian H5N8 in British Columbia, Canada, 2014. *Scientific reports*, 5.

⁸⁸ Bao, Y., Bolotov, P., Dernovoy, D., Kiryutin, B., Zaslavsky, L., Tatusova, T., ... & Lipman, D. (2008). The influenza virus resource at the National Center for Biotechnology Information. *Journal of virology*, 82(2), 596-601.

⁸⁹ Holmes, E. C., Ghedin, E., Miller, N., Taylor, J., Bao, Y., St George, K., ... & Taubenberger, J. K. (2005). Whole-genome analysis of human influenza A virus reveals multiple persistent lineages and reassortment among recent H3N2 viruses. *PLoS Biol*, 3(9), e300.