

**VIROMA DE MUESTRAS DE AIRE DE LA
ESTACIÓN DEL METRO Y EL HOSPITAL 20 DE
NOVIEMBRE EN LA CIUDAD DE MÉXICO**

TESIS

QUE PARA OPTAR POR EL GRADO DE:

MAESTRO EN CIENCIAS GENÓMICAS

PRESENTA:

MARTÍNEZ AMADOR ULICES FABRICIO

DIRECTOR:

DRA. CLAUDIA SELENE ZARATE GUERRA

SISTEMA BIBLIOTECARIO DE INFORMACIÓN Y DOCUMENTACIÓN



UNIVERSIDAD AUTÓNOMA DE LA CIUDAD DE MÉXICO COORDINACIÓN ACADÉMICA

RESTRICCIONES DE USO PARA LAS TESIS DIGITALES

DERECHOS RESERVADOS[©]

La presente obra y cada uno de sus elementos está protegido por la Ley Federal del Derecho de Autor; por la Ley de la Universidad Autónoma de la Ciudad de México, así como lo dispuesto por el Estatuto General Orgánico de la Universidad Autónoma de la Ciudad de México; del mismo modo por lo establecido en el Acuerdo por el cual se aprueba la Norma mediante la que se Modifican, Adicionan y Derogan Diversas Disposiciones del Estatuto Orgánico de la Universidad de la Ciudad de México, aprobado por el Consejo de Gobierno el 29 de enero de 2002, con el objeto de definir las atribuciones de las diferentes unidades que forman la estructura de la Universidad Autónoma de la Ciudad de México como organismo público autónomo y lo establecido en el Reglamento de Titulación de la Universidad Autónoma de la Ciudad de México.

Por lo que el uso de su contenido, así como cada una de las partes que lo integran y que están bajo la tutela de la Ley Federal de Derecho de Autor, obliga a quien haga uso de la presente obra a considerar que solo lo realizará si es para fines educativos, académicos, de investigación o informativos y se compromete a citar esta fuente, así como a su autor ó autores. Por lo tanto, queda prohibida su reproducción total o parcial y cualquier uso diferente a los ya mencionados, los cuales serán reclamados por el titular de los derechos y sancionados conforme a la legislación aplicable.

INTEGRACIÓN DEL JURADO:

Presidente: Dra. María Elizbeth Álvarez Sánchez

Secretario: Dra. Claudia Selene Zarate Guerra

Vocal: Dra. Sofia Lizeth Alcaraz Estrada

Plantel de Adscripción:

PLANTEL DEL VALLE, UACM

DIRECTOR

DRA. CLAUDIA SELENE ZARATE GUERRA

UACM, PLANTEL DEL VALLE

Agradecimientos

A la **Universidad Autónoma de la Ciudad de México** por participar en la formación de nuevos estudiantes con excelencia académica.

A la **Dra. Selene Zarate** quién formó parte integral en mi formación, por la dirección de este proyecto de investigación y por el excelente trato en el laboratorio.

A la **Dra. Sofia Alcaraz** por todas las facilidades logísticas y experimentales otorgadas, así como sus comentarios y para la realización de esta tesis.

A la **Dra. Martha Yocupicio** por permitirme realizar algunas actividades en su laboratorio y por sus señalamientos para la conclusión de esta tesis.

A la **Dra. Elizbeth Álvarez** por sus acertados comentarios en los seminarios y para el desarrollo de este escrito.

A la **Dra. Mariana Peimbert** por participar en el diseño del proyecto, así como en sus aportaciones materiales y académicas para la conclusión de este proyecto.

A todos los profesores que integran el posgrado e invitados por los cursos impartidos.

A **Hugo** por compartir un poco de su conocimiento siempre con la mejor disposición.

A mis compañeros de generación.

Finalmente, agradezco el apoyo a los proyectos UACM PI-CCyT-2019-05 y SECITI/102/2017, por el financiamiento y apoyo a la realización de este proyecto. Asimismo, agradezco al CONACyt por la beca número 493777 otorgada durante el período de agosto de 2017 a julio 2019 y a la beca otorgada a través del proyecto SECITI para la conclusión de este trabajo.

Dedicatorias

A **mis padres**, quienes siempre han estado conmigo y me aman profundamente.

A **mis hermanos**, con quienes siempre compartiré un lazo inquebrantable.

A la mujer más maravillosa, quién forma parte importante de mi vida e hizo posible el que me convirtiera en padre. Te amo **Ana Laura**.

A mi pequeña hija **Violeta Erika** quién es y será mi mayor fuente de inspiración, dedicación y esfuerzo.

A mis compañeros del laboratorio de bioinformática, **Carla, Jose y Enrique**, les deso mucho éxito. A **Carolina y Apolinar** de la UAM Cuajimalpa por la ayuda en los muestreos y por sus valiosas aportaciones.

A todos los que directa e indirectamente han aportado cosas valiosas en mi vida y compartido tiempo conmigo.

Índice general

Resumen.....	1
Abstract.....	2
1. Introducción.....	3
1.1. Virus.....	3
1.2. Propiedades de los virus.....	3
1.3. Clasificación de los virus.....	4
2. Metagenómica.....	9
2.1. Metagenómica viral.....	10
3. Antecedentes directos.....	16
3.1. Dinámica de bioaerosoles en interiores.....	16
3.2. Bioaerosoles: importancia en la transmisión de enfermedades.....	17
3.3. Transmisión de virus en ambientes interiores.....	18
3.4. Transmisión de virus en hospitales.....	19
3.5. Estudios de metagenoma en ambientes urbanos.....	21
3.6. Clasificación de secuencias metagenómicas: métodos y retos.....	24
4. Hipótesis.....	26
5. Justificación.....	26
6. Planteamiento del problema.....	27
7. Objetivos.....	28
7.1. General.....	28
7.2. Particulares.....	28
8. Estrategia experimental.....	29
9. Materiales y métodos.....	30
9.1. Muestras.....	30
9.2. Extracción de ácidos nucleicos (RNA).....	31
9.3. Análisis de calidad de las secuencias.....	32
9.4. Reclutamiento de fragmentos.....	33
9.5. Análisis de diversidad.....	34
9.6. Análisis de variantes (SNVs) en los virus compartidos en todas las muestras.....	38

10. Resultados.....	40
10.1.Extracción de ácidos nucleicos.....	40
10.2.Análisis de calidad de los datos de secuenciación.....	41
10.3.Asignación taxonómica.....	44
10.4.Análisis de variantes virales.....	54
10.5.Determinación de SNVs.....	57
10.6.Comparación entre muestras.....	59
11. Discusión.....	62
12. Conclusiones.....	67
13. Perspectivas.....	69
14. Referencias.....	70
15. Trabajos presentados en congresos.....	79

Índice de figuras

1. Ciclo de vida de un virus de RNA
2. Clasificación de virus de acuerdo a la organización de su material genético
3. Tasas de mutación y sustitución en virus de DNA y RNA
4. Secuencias alojadas en INDSC del año 2000 a 2014
5. Flujo de los principales pasos en un estudio de datos de secuenciación de nueva generación
6. Dinámica y transmisión de partículas aerosolizadas a través del aire y su deposición en el tracto respiratorio humano
7. Diagrama del plan general de trabajo
8. RNA total extraído para cada muestra
9. Número total de *reads* entregados por el IBT
10. Calidad de un archivo .fastq visualizado con FASTQC
11. Cantidad de reads sin duplicados ni adaptadores
12. Diversidad viral a nivel de reads
13. Microfotografía electrónica del adenovirus humano del tipo C
14. Microfotografía electrónica del virus de poliedrosis nuclear de *Spodoptera littoralis* (izquierda) y de un fago de *Staphylococcus sp* (derecha)
15. Diversidad viral a nivel de *contigs*
16. Tablet mostrando *reads* y su alineamiento en una secuencia de referencia
17. Regiones compartidas en las tres especies virales en todas las muestras
18. Correlación de variantes presentes en los tres virus circulantes en todas las muestras

Índice de tablas y cajas

Cajas

1. Principales características de los virus

Tablas

1. Número de secuencias virales alojadas en RefSeq y totales de NCBI
2. Virus de interés en enfermedades respiratorias encontrados en diferentes áreas de un hospital
3. Promedios de temperatura y humedad
4. Resultados de BLASTx de la muestra de aire de taquilla del día 12 de noviembre a partir de los *contigs* obtenidos
5. Resultados de BLASTx de la muestra de aire del vagón del día 12 de noviembre a partir de los *contigs* obtenidos
6. Resultados de BLASTx de la muestra de aire de la sala de espera del hospital del día 12 de noviembre a partir de los *contigs* obtenidos
7. Resultados de BLASTx de la muestra de aire de taquilla del día 26 de noviembre a partir de los *contigs* obtenidos
8. Resultados de BLASTx de la muestra de aire del vagón del día 26 de noviembre a partir de los *contigs* obtenidos
9. Resultados de BLASTx de la muestra de aire la sala de espera del hospital del día 26 de noviembre a partir de los *contigs* obtenidos
10. Frecuencia de variantes por posición en el genoma encontradas

Abreviaturas

μl: Microlitro

μm: Micrometros

16S rRNA: ARN ribosomal 16S o componente de la subunidad menor (30S) de los ribosomas procariotas que se une a la secuencia Shine-Delgarno

BAM: Binary Aligned Map o mapa binario de alineamientos

Base calling: Llamado de bases, es el proceso de incorporación de nucleótidos en una plataforma de secuenciación

Binning: Agrupamiento de *reads* o *contigs* en un grupo taxonómico particular

BLAST: Basic Local Alignment Search Tool por sus siglas en inglés, o Herramienta Básica de Búsqueda de Alineamientos Locales en español

CDMX: Ciudad de México

Contigs: Segmentos contiguos de DNA que representan una región consenso

DNA: Ácido desoxirribonucleico

dNTPs: Deoxiribonucleótidos trifosfatados

dsDNA: Genoma viral de DNA de doble cadena o bicatenario

dsRNA: Genoma viral de RNA de doble cadena o bicatenario

E-value: Valor esperado, disminuye exponencialmente al aumentar el score del alineamiento.

FASTQ: formato de fichero informático basado en texto con score de calidad

Gap: Espacio vacío en un alineamiento

GC: Porcentaje de citosina y guanina

GenBank: Base de datos de secuencias genéticas

hMPV: Metapneumovirus humano

IAH: Infecciones adquiridas en hospitales

IBT: Instituto de Biotecnología de la UNAM

ICTV: Comité Internacional de Taxonomía Viral

ID: Identificador

IMG/VR: Genomas de Microbios Integrados/Virales

INSDC: International Nucleotide Sequence Database Collaboration. En español es la Base de Datos de Secuencias de Nucleótidos de Colaboración Internacional.

k -mero: Es una secuencia de k nucleótidos

Kpb: Kilopares de bases

MetaSUB: Metagenomics and Metadesign of Subways and Urban Biomes. En español significa Metagenómica y Metadiseño de Biomas Urbanos y Subterráneos

mOsm: Osmolaridad

mRNAs: RNA mensajero

NCBI: National Center for Biotechnology Information por sus siglas en inglés. En español es el Centro Nacional para Información Biotecnológica

Output: Archivo de salida

PCR: Reacción en cadena de polimerasa

PEAT: Paired End Adapter Trimming

PH: Potencial de hidrógeno, medida para determinar el grado de alcalinidad o acidez de una disolución

Primer: Segmento de ácidos nucleicos que sirve como punto de partida para una PCR

Reads: Fragmentos de DNA obtenidos de una plataforma de secuenciación

RefSeq: Base de datos de secuencias de referencia anotadas y curadas, alojadas en NCBI

RIN: Número de integridad de RNA

RNA: Ácido ribonucleico

rpm: Revoluciones por minuto

RSV: Virus Sincitial Respiratorio

SAM: Sequence Aligned Map o mapa de alineamientos de secuencias

SARS: Severe Acute Respiratory Syndrome, en español Síndrome Respiratorio Agudo Severo

Script: Archivo de ordenes basado en texto simple

Secuenciación Shotgun:

SNVs: Variante de un solo nucleótido

ssDNA: Genoma viral de DNA de cadena sencilla o monocatenario

ssRNA: Genoma viral de RNA de cadena sencilla o monocatenario

UUSMB: Unidad Universitaria de Secuenciación Masiva y Bioinformática

Resumen

Los virus son las entidades biológicas más abundantes en la Tierra. Tienen la habilidad de infectar bacterias, arqueas y eucariotas. Son microorganismos que necesitan apoderarse de la maquinaria de una célula hospedera para poder replicarse. Los virus más estudiados son aquellos patógenos para los humanos, de los cuales existen aproximadamente unas 200 especies. Su estudio es difícil, ya que las técnicas tradicionales como cultivos o serotipificación están orientadas a identificar virus específicos, dejando sin describir una gran parte de ellos. En los últimos años, el estudio de metagenomas y el desarrollo de tecnologías de secuenciación de nueva generación han permitido conocer aquella porción que no se podía identificar por métodos tradicionales. Por otro lado, las personas pasan una gran parte de su tiempo en ambientes con las condiciones adecuadas para el establecimiento y desarrollo de agentes infecciosos como virus, estos ambientes tienen la particularidad de ser cerrados, donde el recambio de aire y la poca ventilación permiten la transmisión de enfermedades. En este estudio, a través de secuenciación *shotgun*, se analizó la composición viral y las variantes genómicas presentes en el aire de la taquilla y en el interior de un vagón de la estación “Hospital 20 de noviembre” del Sistema de Transporte Colectivo del Metro de la Ciudad de México y de una sala de espera de un hospital cercano, los días 12 y 26 de noviembre de 2018. Los resultados muestran que las familias virales más frecuentes son *Myoviridae*, *Adenoviridae* y *Baculoviridae* en todos los sitios de muestreo en ambos ambientes. Particularmente, el fago de *Staphylococcus sp*, el adenovirus humano del tipo C y el virus de polihedrosis nuclear de *Spodoptera littoralis*, para cada familia, respectivamente, lo que es interesante ya que bacterias, humanos e invertebrados contribuyen al viroma de estos ambientes. Además, el análisis de variantes presentes en regiones genómicas indica que los virus circulantes en el aire de algunos sitios son similares a las que poseen los virus que circulan en otros sitios ya sean del 12 o del 26 de noviembre.

Abstract

Viruses are the most abundant biological entities on Earth. They have the ability to infect bacteria, archaea and eukaryotes. They are microorganisms that need to seize the machinery of a host cell in order to replicate. The most studied viruses are those pathogens for humans, of which there are approximately 200 species. Their study is difficult, because traditional techniques such as cultures or serotyping are aimed at identifying specific viruses, leaving a large part of them undescribed. In recent years, the study of metagenomes and the development of next generation sequencing technologies have allowed us to know that portion that could not be identified by traditional methods. On the other hand, people spend a large part of their time in environments with the appropriate conditions for the establishment and development of infectious agents such as viruses, these environments have the particularity of being closed, where the exchange of air and poor ventilation allow the diseases transmission. In this study, through shotgun sequencing, the viral composition and genomic variants present were analyzed in samples obtained from the area of the the ticket office of the “Hospital November 20” station and inside a train car of the corresponding subway line of the México’s City Collective Transportation System “Metro”. Additionally, a waiting room of a nearby hospital was also sampled. The samples were taken , on November 12 and 26, 2018. The results showed that the most frequently sampled viral families were *Myoviridae*, *Adenoviridae* and *Baculoviridae* at all sampling sites in both environments. Particularly, the presence of the phage of *Staphylococcus* sp, the human adenovirus type C and the nuclear polyhedrosis virus of *Spodoptera littoralis* were detected. This observation is interesting because it shows that bacteria, humans and invertebrates contribute to the virome of these environments. In addition, the analysis of variants present in some genomic regions of these viruses showed that the viruses circulating in the air of some sites are similar to those possessed by viruses circulating in other sites either from November 12 or 26, indicating that these areas are not isolated and therefore virus are able to move freely between them.

1. Introducción

1.1 Virus

Los virus son las entidades biológicas más abundantes de la biosfera y tienen la habilidad de infectar a bacterias, arqueas y eucariotas. Se estima que los virus son al menos 10 veces más abundantes que las bacterias, y se calcula que existen alrededor de 10^7 virus por mililitro y 10^9 partículas virales por gramo en aguas marinas y 10^8 por gramo de sedimentos y suelos (Paez-Espino *et al.*, 2018; Roux *et al.*, 2014; Allen L. Z. *et al.*, 2017). Se ha sugerido que el número de virus en la Tierra excede considerablemente el número de estrellas en el universo, más o menos 10^{23} partículas virales en total (100 sextillones) (Dimmock N. J. *et al.*, 2016).

1.2 Propiedades de los virus

Asumiendo que algunas de las características de todos los virus pueden ser similares entre ellos, es posible comparar las propiedades que son exclusivas de virus con respecto a las de sus hospederos. Una diferencia entre la célula hospedera y los virus es que las células hospederas tienen ambos tipos de ácidos nucleicos que son DNA y RNA, los virus solo contienen uno de los dos y al igual que las células de su hospedero, su información genética esta codificada por su material genético. Otra diferencia entre ambos es que los virus son replicados solamente a partir de su material genético, nunca de un virus preexistente, mientras que las células hospederas se replican a partir de una célula preexistente de forma integral con todos sus componentes (Dimmock N. J. *et al.*, 2016). En la Caja 1 se enlistan algunas propiedades comunes a todos los virus.

- Los virus tienen genoma de ácidos nucleicos de DNA o de RNA, nunca de ambos.
- Comparado con el genoma de una célula, los genomas virales son pequeños, sin embargo, existen virus con genomas de 3000 hasta 600,000 nucleótidos.
- Los genomas virales son asociados con proteínas que forman así la partícula viral, pero en algunos virus esta nucleoproteína esta rodeada de otras proteínas o de una bicapa lipídica.

- Las proteínas externas de la partícula viral les permiten reconocer a la célula hospedera correcta y entrar en ella.
- Solo pueden replicarse en células vivas, son entidades biológicas obligadas.

Caja 1. Principales características de los virus

Los virus solo pueden ser detectados de forma indirecta y son pequeños simbioses obligados que no pueden replicarse por sí solos, ya que requieren de la maquinaria de una célula hospedera para hacerlo. Aunque muchos de ellos son conocidos por causar enfermedades, otros coexisten pacíficamente con sus hospederos (Dimmock N. J. *et al.*, 2016). Una vez que infectan a una célula, los virus son capaces de apoderarse de la maquinaria celular para producir más partículas virales.

1.3 Clasificación de los virus

La clasificación viral puede ser de acuerdo a las enfermedades que ocasionan, al organismo al que infectan, a la morfología de la partícula, al tipo de genoma refiriéndose a que los genomas virales pueden ser de DNA o RNA , de cadena simple o doble, lineales o circulares, segmentados o no segmentados y pueden ser de polaridad negativa o positiva. Estas características de los genomas virales son la base de la clasificación de Baltimore la cual considera los mecanismos de replicación, de acuerdo al el tipo de ácido nucleico que permiten la producción de mRNAs y proteínas virales (Wylie *et al.*, 2013). (Figura 2). La partícula viral infecciosa es llamada virión, y consiste de los ácidos nucleicos y una cápside externa compuesta en su mayoría por proteínas y en ocasiones estar rodeada por una bicapa lipídica.

Los mecanismos utilizados por los virus para su multiplicación se llevan a cabo en función de la naturaleza de su material genético y debido a que no poseen las estructuras necesarias para traducir a los mRNAs, la célula hospedera es quien los aporta, consiguiendo así que sus mRNAs sean reconocidos por el ribosoma del hospedero y traducidos a proteínas. Por convención todos los mRNA son designados como de polaridad positiva (+). Las hebras de DNA o RNA viral que son complementarias al mRNA son designadas como de polaridad negativa (-). Usando

esta terminología, así como la información sobre el proceso de replicación se han definido siete grupos virales:

Grupo I: Virus de DNA de doble hebra (dsDNA), no importa la designación de polaridad negativa o positiva dado que mRNA proviene de cualquier hebra. La transcripción ocurre de manera similar a la de la célula hospedera.

Grupo II: Virus de DNA de cadena sencilla (ssDNA) de polaridad positiva. El DNA puede ser convertido a DNA de doble hebra antes de la síntesis de mRNA.

Grupo III: Contiene todos los virus de RNA de cadena doble (dsRNA), su genoma es segmentado y el mRNA es sintetizado de un molde de cada segmento.

Grupo IV: Son virus de RNA de cadena sencilla (ssRNA) de polaridad positiva como el mRNA y pueden ser traducidos directamente a proteínas, sin embargo, la síntesis de una hebra de RNA complementaria al genoma precede la síntesis de mRNA.

Grupo V: Engloba a virus de RNA de cadena sencilla (ssRNA) de polaridad negativa que es complementario a la secuencia de mRNA, cuya síntesis requiere de la generación de nuevos genomas virales y de un dsRNA intermediario el cual es usado como molde para la replicación

Grupo VI: Contiene virus denominados retrovirus con genoma de RNA de cadena sencilla (ssRNA) y que generan un intermediario de dsDNA como preludio a la replicación.

Grupo VII: Son los retrovirus que pueden ser convertidos del grupo I al grupo VII. Esto es basado en la replicación de su genoma de dsDNA a través de un intermediario de ssRNA(+)

Con respecto a su ciclo de vida, cada virus tiene sus particularidades, dependiendo si su genoma es de DNA o RNA, si es envuelto o no, sin embargo, podemos resumirlo en las siguientes etapas (Dimmock N. J. *et al.*, 2016) (Figura 1):

Ciclo de vida de un virus de RNA

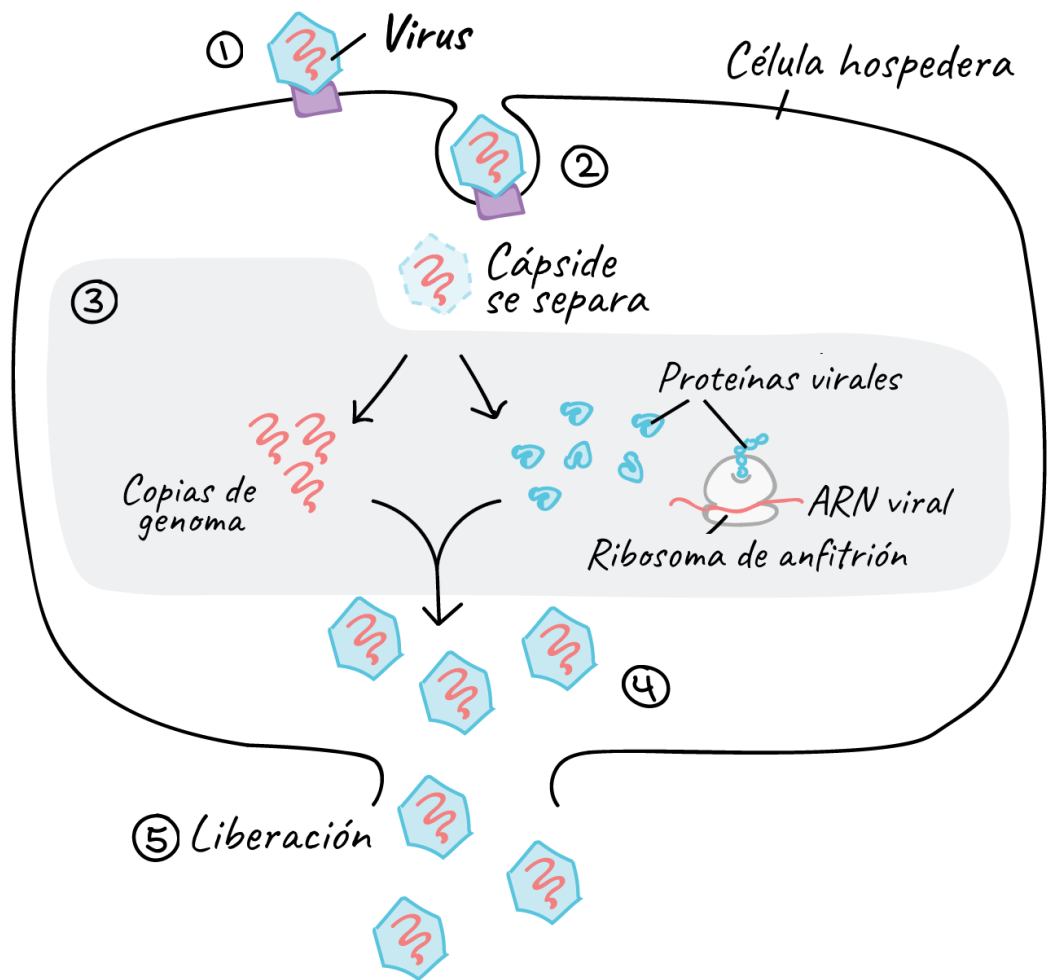


Figura 1. Diagrama general del ciclo de vida de un virus con genoma de RNA. Se pueden observar 5 etapas principales, las cuales son: {entrada, desnudamiento, replicación, encapsidación y salida. Tomado de [khanacademy.org](https://www.khanacademy.org)

Las etapas principales de un ciclo de vida se muestran en la Figura 1 y se explican a continuación:

1. Reconocimiento y anclaje: las proteínas virales reconocen receptores específicos de la célula hospedera.
2. Fusión de membranas: la membrana lipídica se fusiona con la envoltura lipídica de los virus. Los virus no envueltos
3. Liberación de genoma e integración: la partícula viral desnuda su genoma, dependiendo del tipo de virus, este se puede integrar al genoma del hospedero,

- sirviendo así de molde para la generación de más copias en el núcleo o bien, replicarse en citoplasma y ser traducido para la síntesis de proteínas virales.
- Replicación de genoma y síntesis de proteínas: se producen más copias del genoma viral y los mRNA sirven de molde para la síntesis de proteínas estructurales y no estructurales.
 - Encapsidación y liberación: las copias del genoma viral son encapsidadas por proteínas estructurales, formando la partícula viral y promoviendo la lisis de la célula hospedera.

Clasificación de virus de acuerdo a la organización de su material genético

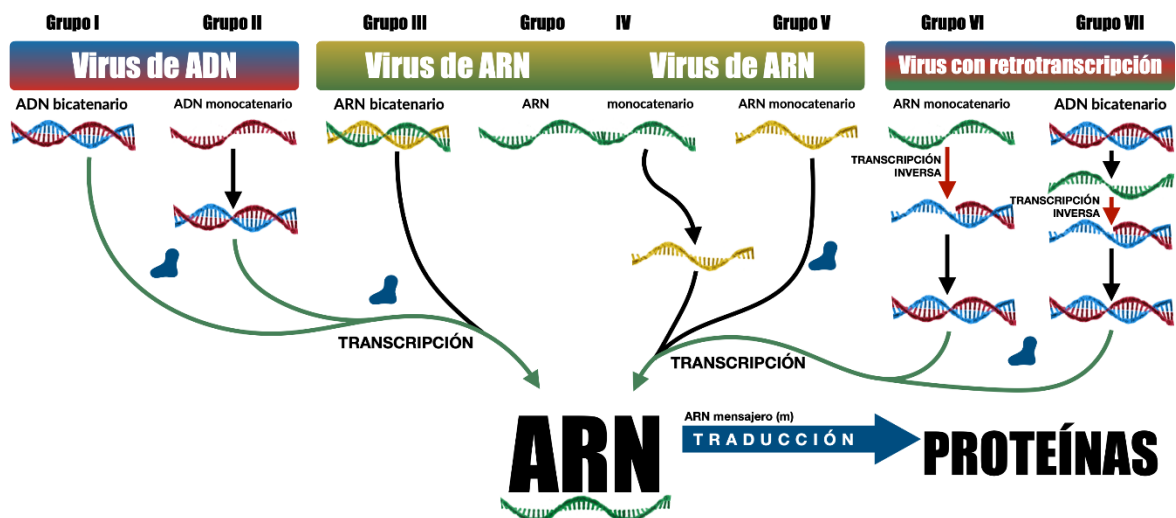


Figura 2. Se muestra la clasificación de Baltimore para virus, la cual está basada en la naturaleza de su genoma, incluidos los virus de DNA, los de RNA y los retrovirus, así mismo la vía que siguen para transcribir a RNA y su posterior traducción a proteínas.

Los virus más estudiados son aquellos que resultan patógenos para los humanos o para especies de interés alimenticio, de los cuales se han descrito aproximadamente 200 especies virales, además de que se descubren nuevos virus patógenos en una tasa de dos por año (Bibby, K. 2013). El gran impacto en la salud humana y en especies animales y vegetales de importancia económica ha hecho de los virus un importante objeto de estudio durante décadas (Orton R.J. *et al.*, 2016).

Una de las características que dificulta el control y prevención de algunas enfermedades virales es el hecho de que algunos virus evolucionan rápidamente debido al gran tamaño de sus poblaciones y a sus elevadas tasas de replicación. Particularmente los virus de RNA presentan altas tasas de mutación debido a la pobre fidelidad de sus RNA polimerasas, a la generación de grandes poblaciones virales que poseen una gran variabilidad, lo que les permite adaptarse rápidamente a nuevos ambientes y a presiones de selección como tratamientos con fármacos (Orton R.J. *et al.*, 2016). Además, se ha observado que las tasas de mutación y sustitución de algunos virus de DNA de cadena sencilla es equiparable con la de los virus de RNA (Figura 3) (Duffy S. *et al.*, 2008).

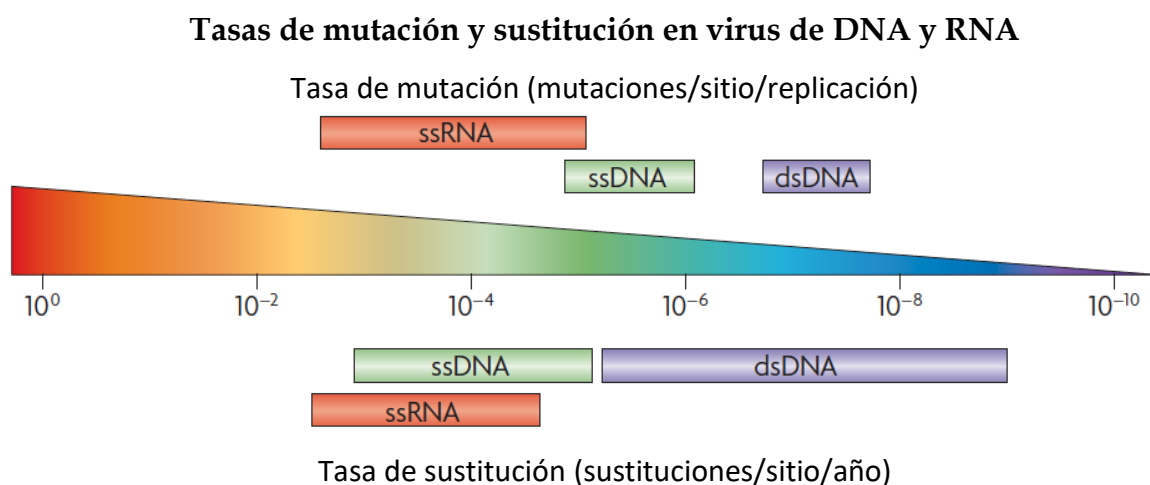


Figura 3. Tasa de mutación en los genomas virales expresada como mutaciones por sitio por ronda de replicación para virus con diferente arquitectura genómica es resumidos en la parte superior de la figura, mientras que la tasa de sustitución en los genomas virales está dada por sustituciones por sitio por año, la cual se muestra en la parte inferior de la figura. Modificado de Duffy S. *et al.*, 2008.

2. Metagenómica

El término “metagenómica” fue acuñado por el ecologista de microbios de suelo Jo Handelsman en 1998. La definición general de metagenómica o genómica de comunidades se refiere al análisis de la información genética o de genomas colectivos de microorganismos dentro de una muestra obtenida de un ambiente dado independiente de cultivos (Williamson S. J 2011). Los estudios de metagenómica producen como resultado una colección de información genética a través de secuenciación de DNA o RNA de una comunidad. El análisis subsecuente puede enfocarse a identificar un marcador conservado para la identificación de las especies presentes en dicha comunidad o bien un análisis del tipo de genes presentes y las funciones bioquímicas correspondientes (Izard J., 2014).

Los datos de secuencias metagenómicas son típicamente usados para responder las dos siguientes preguntas fundamentales: ¿qué hay ahí? y ¿qué función tiene ahí? Los datos funcionales y taxonómicos de estudios metagenómicos han revolucionado nuestro entendimiento de la diversidad de microbios y su papel en las comunidades (Williamson S., 2011). Durante los últimos años, la metagenómica ha permitido identificar la biodiversidad presente en diferentes ambientes de manera mucho más profunda que los datos que se obtenían con los métodos convencionales. Actualmente las tecnologías de secuenciación no están restringidas a bacterias, por lo tanto, cualquier muestra biológica que consiste en una mezcla compleja de organismos puede ser analizada. Diversos tipos de muestras dentro de la biosfera están siendo exploradas usando estas técnicas, tales como el perfilado de organismos presentes en aires interiores de ambientes urbanos, en océanos, en la determinación del microbioma en el tracto digestivo en humanos o el descubrimiento de patógenos no identificados previamente (Behzad H. *et al.*, 2015).

2.1 Metagenómica viral

Los virus son considerados como el mayor reservorio de diversidad genética. Se ha reportado que los virus pueden afectar los ciclos biogeoquímicos y la dinámica de los ecosistemas (Jover L. F. *et al.*, 2014).

Como tal, un viroma puede ser definido como el conjunto de virus que se encuentran en un ambiente particular. Desde el contexto del cuerpo humano, por ejemplo, varios viromas deben ser considerados puesto que existen diferentes compartimentos con comunidades virales distintas (Zarate S., *et al.*, 2018). La exploración como tal de un viroma, es decir el conjunto de virus presentes en un ambiente, comenzó en 2002 con la descripción de las poblaciones virales en océanos, que contenían una considerable complejidad genética (Paez- Espino *et al.*, 2016) y ha continuado con la descripción de más viromas oceánicos (Angly F. E. *et al.*, 2006; Allen L. Z. *et al.*, 2017), en suelo (Reavy B. *et al.*, 2014; Williamson K. E. *et al.*, 2017), en animales de granjas (Shan T. *et al.*, 2011) y casi en cualquier ambiente donde existe dinámica entre los microorganismos que lo habitan.

Sin duda, el estudio del cuerpo humano bajo las condiciones fisiológicas de salud y enfermedad es donde se han realizado la mayor cantidad de estudios donde se analizan las comunidades virales (Carding S. R. *et al.*, 2017; Cadwell Ken, 2015).

Con la llegada de las tecnologías de secuenciación de nueva generación o secuenciación de alto rendimiento se ha revolucionado el campo de la microbiología ecológica y se han llevado a cabo los estudios clásicos de ambientes a otro nivel. Este tipo de tecnología ha conducido al establecimiento de la “metagenómica” (Oulas *et al.*, 2015).

En bacterias el gen 16S rRNA está conservado en todas las especies, sin embargo, su secuencia tiene la variabilidad suficiente para resolver un metagenoma bacteriano presente en una muestra incluso por debajo del nivel taxonómico de género. En contraste, la detección de virus usando metagenómica es complicada debido a que los virus no comparten un elemento genético en común (Lundberg *et al.*, 2013). Para superar esto, avances en tecnologías de secuenciación de nueva generación y en

técnicas de ensamble de las secuencias, han conducido a la mejora en la obtención de fragmentos virales más largos. De hecho, los *contigs* que se generan por el solapamiento de *reads* representan casi por completo los genomas virales, claro, que también depende de otros factores como la naturaleza de la muestra, la concentración de ácidos nucleicos y un correcto proceso de preparación de bibliotecas, así como de la secuenciación misma (Roux *et al.*, 2014). Existen dos principales estrategias de secuenciación de material genético, con diferencias metodológicas y ventajas, así como desventajas. Una de ellas es la secuenciación de amplicones, la cual consiste en restringir con el uso de *primers* dirigidos una región genómica de interés. La limitación de este método es que no se puede proveer de un panorama general de diversidad, sin embargo, la profundidad obtenida de las secuencias es perfecta para llevar a cabo un análisis de variantes de los virus presentes en una muestra. En contraste, la secuenciación *shotgun* permite obtener un panorama general de la diversidad viral encontrada en una muestra determinada, comparando los datos resultantes con las secuencias contenidas en las bases de datos públicas usando algoritmos de búsqueda comparativa como BLAST (Ranjan *et al.*, 2015). Por otro lado, el uso de la metagenómica permite el examen y análisis de la composición taxonómica y potencial metabólico de la microbiota encontrada en diferentes ambientes (Thomas T. *et al.*, 2015).

Hasta 2015, el número de genomas y de segmentos de genomas incluidos en la base de datos RefSeq de NCBI era limitado (Tabla 1), (Figura 4) (Brister *et al.*, 2014).

Número de secuencias virales alojadas en RefSeq y totales de NCBI

Tipo de genoma viral	Segmentos de genoma en RefSeq	Segmentos de genoma totales
dsDNA	1755	3023
dsRNA	919	17929
ssDNA	669	6692
ssRNA(-)	187	4384
ssRNA(+)	917	14441

Tabla 1. Número de segmentos virales validados y de RefSeq por tipo de genoma viral. La tabla no incluye secuencias de virus de influenza, debido a que están almacenadas en una base de datos especializada.

Secuencias alojadas en INDSC del año 2000 a 2014

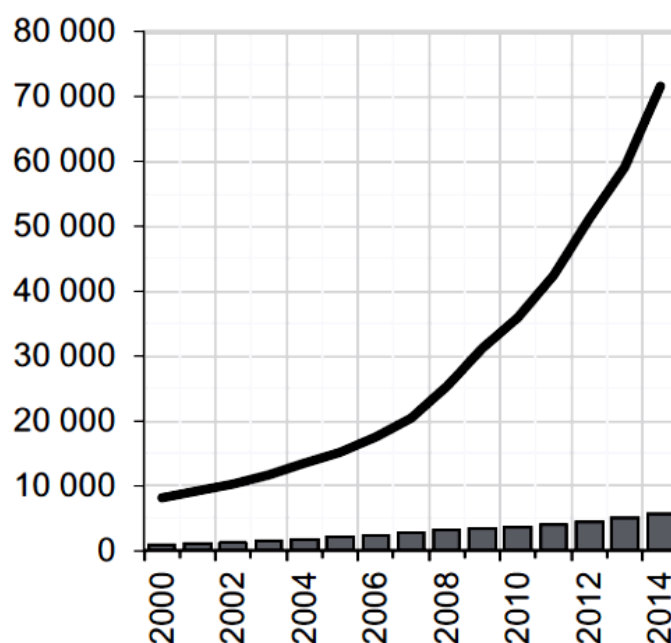


Figura 4. Número de segmentos de virus validados. En el eje de Y de la gráfica se indica el número de segmentos validados y disponibles en INSDC que están representados por la línea negra, y en el eje X se ubican los segmentos de virus de RefSeq que se muestran por las columnas grises. Secuencias de influenza no son incluidas.

Los brotes de nuevos y mortales virus resaltan la vulnerabilidad global a las enfermedades emergentes, muchos de los cuales tienen un impacto sanitario y económico masivo (Carroll D. *et al.*, 2018).

Con respecto al estudio de virus, la metagenómica supera las principales limitaciones de las herramientas clásicas para detección viral. De hecho, a diferencia de las técnicas tradicionales para identificación viral y microbiana, la metagenómica no requiere ni cultivos ni aislamiento para la identificación de las especies presentes en la muestra, por lo cual es particularmente adecuada para proveer un panorama global de la diversidad funcional y potencialmente identificar nuevas especies. En principio, esta estrategia permite la identificación de cualquier organismo, incluyendo aquellos no

detectados comúnmente debido a la dificultad de aislamiento y crecimiento bajo condiciones de laboratorio (Fancello *et al.*, 2012).

Seguido del proceso de secuenciación, viene el proceso de armar contigs a partir de reads, el cual puede tomar dos rutas diferentes: 1) ensamble basado en referencia o 2) ensamble *de novo*. La elección de cuál es el mejor depende de la necesidad del análisis, por ejemplo, el ensamble *de novo* puede ser usado, en teoría, aun cuando existe un genoma de referencia, si el poder computacional lo permite (Oulas *et al.*, 2015).

El ensamble que se realiza contra genomas de referencia implica el uso de uno o más genomas como un mapa para crear contigs que pueden representar genomas o partes de genomas pertenecientes a un género o especie específica (Oulas *et al.*, 2015). (Figura 5)

Flujo de los principales pasos en un estudio de datos de secuenciación de nueva generación

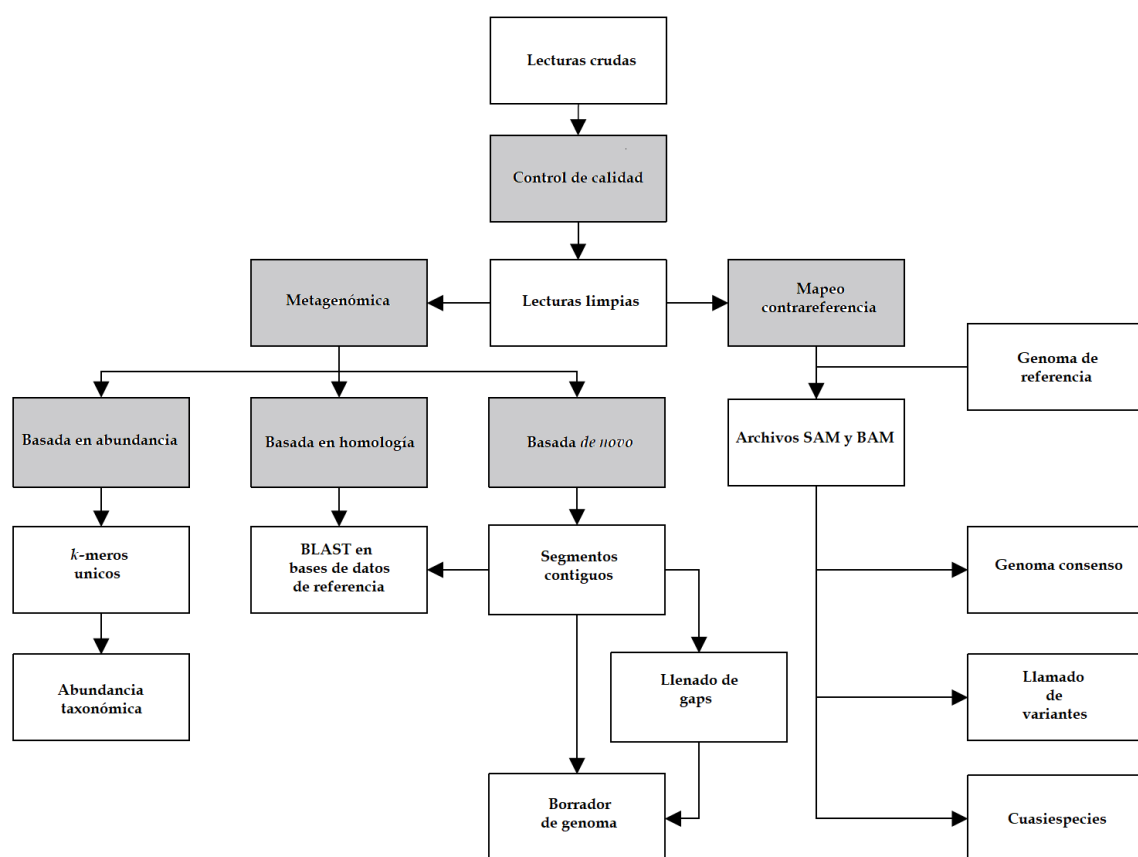


Figura 5. Principales estrategias de un análisis de datos de secuenciación shotgun. Se ilustran los pasos principales involucrados en la generación de un genoma consenso, ensamble *de novo* y metagenómica.

A pesar de las muchas ventajas que pueden ofrecer los estudios metagenómicos para caracterizar la variabilidad genética presente en una muestra, están limitados por la cantidad de información que se encuentra en las bases de datos que se usan para clasificar las secuencias obtenidas. Es muy común que muchas secuencias obtenidas de un metagenoma no puedan ser clasificadas debido a que no presentan ninguna similitud con las bases de datos de referencia (Edwards y Rohwer, 2005).

Otro de los retos de los estudios de metagenómica viral es que la cantidad de material genético de origen viral en una mezcla compleja es muy limitada. Esto se debe a que el tamaño de los genomas virales es en general varios órdenes de magnitud menor que el de bacterias y eucariontes, además, la ausencia de un marcador genético universal para virus, como la subunidad ribosomal 16s en bacterias, obliga a dirigir un estudio de esta naturaleza a través de secuenciación shotgun, ya que no se necesita el uso de primers que amplifiquen una región específica, puesto que en esencia no se sabe que virus son los que están presentes en una muestra (Thomas *et al.*, 2012).

La aleatoriedad de la secuenciación masiva asegura que algunas especies de microorganismos que pueden ser no detectados por métodos tradicionales van a ser representados por al menos algunos segmentos pequeños de secuencias. Esta tecnología puede ser aplicada al campo médico investigando nuevas cepas o especies de microbios, brotes y enfermedades complejas (Alavandi y Poomina, 2012).

Tomando como punto de referencia a la metagenómica, y al estudio de microorganismos sin la necesidad de cultivos previos, se acuñó el término viroma. El viroma se refiere a la colección de ácidos nucleicos, ya sean de DNA o RNA, contenidos en una comunidad viral asociados a un ecosistema o ambiente particular. La palabra, es, desde luego, derivada de virus y genoma, y fue usada por primera vez por Forest Rohwer para describir los metagenomas virales determinados a través de secuenciación shotgun. Todos los macroorganismos tienen un viroma particular, que

incluye fagos de bacterias y virus de células eucariontes. Los viromas son importantes en ciclos de energía y nutrientes, desarrollo de inmunidad y son una fuente importante de genes a través de su conversión lisogénica. En ambientes marinos, los virus han recibido considerable atención en las últimas dos décadas ya que la lisis de células hospedadas por virus marinos libera materia orgánica disuelta y particulada, las cuales regresan al ambiente donde pueden ser asimiladas por microorganismos o exportadas desde la superficie hasta la profundidad oceánica. De hecho, los virus son los mayores directores de mortalidad de fitoplancton y bacterias heterotróficas (Jover *et al.*, 2014).

Al respecto de su importancia en el desarrollo del sistema inmune, un ejemplo claro es la infección con el virus del Zika, donde dicha infección estimula la producción de interferón tipo I (α , β), tipo II (γ) y tipo III (λ) y numerosos genes estimulados por interferón que limitan la infección (Pierson y Graham, 2016).

A la fecha, gracias a los avances en tecnologías para el estudio de virus, y a estudios que demuestran que los virus son más abundantes que otros microorganismos, el estudio de metagenomas en otro tipo de ambientes como hospitales, sistemas de transporte de pasajeros, habitaciones, entre otros, ha tomado mayor relevancia debido a la posibilidad de transmisión de patógenos, lo que representa un problema de salud pública.

3. Antecedentes directos

3.1 Dinámica de bioaerosoles en interiores

Esta bien establecido que los humanos pasamos mucho de nuestro tiempo en interiores como el hogar, el trabajo, hospitales, escuelas y sistemas de transporte. Además, es bien sabido que los ambientes interiores ocupados por humanos contienen abundante material de origen microbiano, sin embargo, se sabe poco acerca de la dinámica de los microorganismos que están presentes (Figura 6) (Nazaroff W. W. 2014).

Dinámica y transmisión de partículas aerosolizadas a través del aire y su deposición en el tracto respiratorio humano

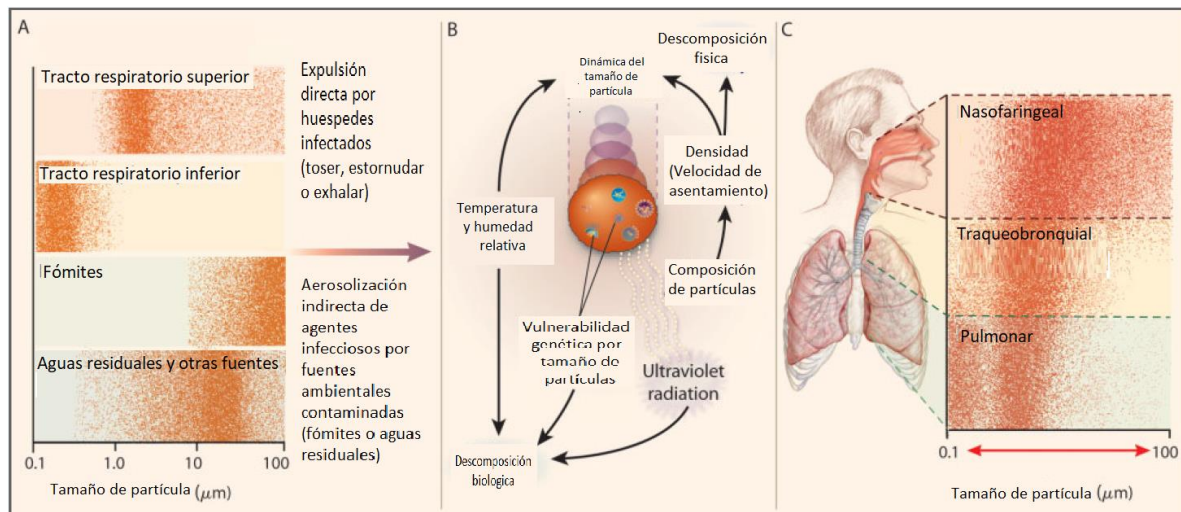


Figura 6. Vía aerobiológica para la transmisión de enfermedades respiratorias comunicables. Ya sea un humano infectado o una superficie contaminada, cada fuente genera partículas virales con diferente tamaño (A). El tiempo que permanece en el aire depende de su tamaño, composición y factores ambientales (B). La porción del tracto respiratorio del hospedero susceptible que inhala las partículas y son depositadas depende de su tamaño. Modificado de Roy C. J. *et al.*, 2004.

Existen algunas razones para estar interesados en la microbiología de ambientes urbanos, o de ambientes que fueron construidos por los seres humanos, las cuales derivan de tres conceptos primarios. Primero, la exposición a bioaerosoles puede

causar o contribuir al desarrollo de enfermedades importantes. Segundo, síntomas adversos de problemas respiratorios correlacionan bien con indicadores de humedad en interiores. Tercero, la microbiología de los espacios que ocupamos puede influir al microbioma humano y así conferir salud o enfermedad, aunado a factores como la dieta o el estilo de vida en general (Mubareka S. *et al.*, 2019).

3.2 Bioaerosoles: importancia en la transmisión de enfermedades

Bioaerosol es una contracción de “biological aerosol” por su nombre en inglés y aerosol se refiere a la suspensión de partículas en un gas, en este sentido, todos los microorganismos, incluidos virus, bacterias, hongos y químicos o moléculas asociadas a los microbios como endotoxinas y micotoxinas también son parte de estos (Verreault D. *et al.*, 2008). Son una suspensión de sólidos o partículas líquidas, cuyos tamaños oscilan de 0.001 a 100 μm .

El incremento de la globalización, urbanización e industrialización tienen un gran impacto en el ambiente y estilo de vida humanos. Aunque la mayor atención de los estudios ha sido dirigida hacia las consecuencias económicas y ecológicas, es claro que los cambios sociales son responsables del incremento en la aparición de nuevos virus, que pueden ser patógenos para humanos u otras especies (Lam *et al.*, 2016).

Desde ese enfoque, una densidad poblacional alta, coloca a la Ciudad de México como un sitio donde pueden ocurrir epidemias emergentes, particularmente lugares o ambientes donde hay intercambio de microbiota debido a la necesidad de estar o trasladarse de un sitio a otro. Como ejemplo, durante la pandemia por el virus de influenza H1N1, el número de personas infectadas fue de aproximadamente 70 mil y alrededor de mil muertes (López-Cervantes M. *et al.*, 2010).

Para entender mejor la dinámica de los microbios de las ciudades en relación con la salud humana y crear ambientes saludables en sitios que las personas visitan frecuentemente, algunos investigadores han comenzado a estudiar el microbioma presente en sistemas de transporte de pasajeros. Un proyecto internacional llamado MetaSUB fue iniciado para buscar el microbioma de las principales redes de transporte

en más de 50 ciudades, entre ellas, Singapur, Boston, Shanghái, París, Estocolmo y San Paulo, con el objetivo de mejorar la salud pública (Coleman K. K. *et al.*, 2018).

Los estudios de bioaerosoles han demostrado que las concentraciones de bacterias en sistemas de transporte subterráneo tienden a fluctuar temporalmente y están ligadas a la presencia y actividades humanas. Dicha información demuestra el posible papel del transporte público en la transmisión de enfermedades, sin embargo, pocos estudios se han enfocado en la existencia de virus transportados por aire en sistemas de transporte público (Browne A. *et al.*, 2016).

Distintos tipos de dispositivos para captar el aire han sido usados en los últimos años, dependiendo del diámetro aerodinámico de las partículas, las propiedades de adhesión, movimiento Browniano, así como la inercia de las mismas. Como tal, los métodos de recolección de muestras van desde los basados en gravedad, donde se colocan recipientes con medios de cultivo y atrapan solo las partículas que caen, otros son los recolectores con filtros, estos están dirigidos a la captación de virus que son transportados en el aire. Un tercer método de recolección es el hisopado de las superficies de filtros purificadores de aire. En todos los casos, se requiere de suficiente concentración de virus para que estos puedan ser detectados (Verreault D. *et al.*, 2008).

3.3 Transmisión de virus en ambientes interiores

Algunos estudios han propuesto que los virus pueden transmitirse de un hospedero a otro por aire ya que ellos lo usan para transportarse. Recientemente, un estudio retrospectivo de cohorte conducido después de una epidemia por el virus de SARS (Severe Acute Respiratory Syndrome) en Hong Kong en 2003, sugiere que en efecto, la transmisión por aire juega un papel importante en el esparcimiento de la enfermedad (Yu. I. T. *et al.*, 2005), sin embargo, los bioaerosoles son también responsables de la transmisión de otras enfermedades virales (Johnson N. R. *et al.*, 20146).

Diferentes sitios funcionan como reservorios o vectores desde los cuales puede ocurrir la transmisión de virus y otros agentes infecciosos (Huttenhower *et al.*, 2016). Las personas de las sociedades modernas pasan hasta el 90% de su vida en lugares cerrados

por lo cual están constantemente expuestos al contenido presente en el aire y superficies de dichos hábitats interiores. Ese aire consta de un numeroso contenido de partículas de origen biológico (Taewon *et al.*, 2018) que han mostrado su capacidad para causar reacciones adversas a la salud como infecciones, hipersensibilidad y reacciones tóxicas debido a la naturaleza y combinación de las partículas presentes y a la exposición por inhalación (Yu *et al.*, 2004). Cabe enfatizar que la relación bioaerosoles-humano es estrictamente la que ocurre comúnmente en espacios ocupados por la mayoría de las personas, como casas, hospitales, oficinas, escuelas, es decir, los sitios donde el tiempo de ocupación es mayor o donde la densidad de personas es elevada. No se incluyen ambientes industriales donde la exposición es ocupacional ya que, al tratarse de lugares con una naturaleza y procesos diferentes, la esencia del aire que ahí se encuentra es, desde luego, más tóxica (Eames I. *et al.*, 2009).

Los sistemas de transporte de pasajeros proveen el ambiente ideal para la transmisión de microorganismos, ya que acarrean un gran número de personas junto con sus microbiomas, diariamente. Particularmente, el Sistema de Transporte Colectivo Metro de la Ciudad de México, transporta más de 4 millones de usuarios diariamente. Tan solo en el año 2018, un total de 1,647,475,013 usuarios fueron transportados, posicionándolo como el segundo sistema de transporte más ocupado de América, después del subterráneo de Nueva York, y el noveno en el mundo. Con respecto a la estación Hospital 20 de noviembre de la línea 12 del Sistema de Transporte Colectivo del Metro, de octubre a diciembre de 2018, el número de usuarios transportados ascendió a 1,063,834 (STC, 2019).

El movimiento masivo de personas que respiran el mismo aire y tocan las mismas superficies, promueven un intercambio a gran escala de microbiota humana y ambiental. Entre los factores físicos que influyen directamente en el microbioma están la temperatura y la humedad, aunque existen otros como las diferencias en los sistemas de ventilación, la limpieza de la infraestructura y la contaminación. En el mundo, existen pocos reportes donde se describa el microbioma de los sistemas de transporte

de pasajeros, sin embargo, pueden ser tomados como referencia para llevar a cabo nuevos proyectos en nuestro país. (Hernández A., *et al.*, 2019).

3.4 Transmisión de virus en hospitales

En los hospitales existe la presencia de agentes patógenos que provienen de pacientes que acuden por enfermedades y cirugías, sin embargo, existen también infecciones nosocomiales las cuales representan un serio problema de salud pública (Beggs C. B. 2003).

Partiendo de lo anterior, se conoce que los sistemas de transporte de pasajeros no son los únicos ambientes urbanos donde ocurre el desarrollo y transmisión de organismos patógenos. Los hospitales son otro sitio donde los individuos son frecuentemente expuestos ocupacional o transitoriamente a diversos organismos como bacterias o virus a través de varias rutas como lesiones, cuidado directo de pacientes y al ambiente contaminado, ya sea en aire o superficies (Ribeiro E., 2017). (Tabla 2)

El estudio de los bioaerosoles resulta de particular interés debido a que son sitios donde la erradicación y control de las infecciones resulta difícil, debido a que son ambientes con las condiciones adecuadas para que los microorganismos desarrollen resistencia a medicamentos como antibióticos y antivirales (Clark R. P. *et al.*, 2009).

Virus de interés en enfermedades respiratorias encontrados en diferentes áreas de un hospital

Sitio y muestras analizadas	Periodo de estudio	Virus encontrados
Hospital pediátrico, aire y superficies	3 meses	Norovirus, adenovirus, RSV, rotavirus, hMPV
Unidad de cuidados intensivos	6 meses	Rotavirus
Sala de espera	15 meses	Adenovirus, norovirus
Unidad pediátrica primaria de inmunodeficiencia	12 meses	Norovirus, astrovirus, rotavirus
Unidad pediátrica general	6 meses	Norovirus, astrovirus, rotavirus

Tabla 2. Contaminación por virus en muestras de interiores de un hospital. RSV= Virus sincitial respiratorio, hMPV= Metapneumovirus humano.

La prevalencia de virus patógenos potencialmente transmitidos por aire y microgotas representan un problema significativo para trabajadores, pacientes o personas que transitan en hospitales y sistemas de transporte de pasajeros, además de la habilidad de los virus de permanecer en superficies o en aire (Tringe *et al.*, 2007).

La transmisión de enfermedades a través del aire es una vía por la cual se generan infecciones como tuberculosis y aspergilosis, así mismo, otros microorganismos como *Acinetobacter baumannii* y *Pseudomonas aeruginosa* (Beggs C. B. 2003).

3.5 Estudios de metagenoma en ambientes urbanos

Los estudios metagenómicos son aplicados comúnmente para investigar los genomas específicos, conocidos o desconocidos, de organismos cultivables o no, que están presentes dentro de una comunidad ambiental bajo estudio, sin embargo, las secuencias completas de genes que codifican proteínas pueden ofrecer conocimiento funcional invaluable acerca de las comunidades. Por estas razones, un ensamble de *reads* (fragmento de DNA obtenido de una plataforma de secuenciación) cortos en *contigs* (segmentos contiguos de DNA que representan una región consenso) genómicos y su orientación a *scaffolds* (porción de secuencia de un genoma compuesto de *contigs* y *gaps*) es a menudo realizado para otorgar una vista más concisa y compacta de la comunidad bajo investigación (Oulas *et al.*, 2015).

En los últimos años, nuevos trabajos han caracterizado la microbiota y el metagenoma de ambientes urbanos como hospitales y sistemas de transporte de pasajeros aportando datos que son de ayuda para entender sitios de contacto entre humanos y el mundo microbiano. De hecho, el cómo los humanos interactúan con nuevas especies de bacterias o virus adquiridas depende del ambiente que transitan, los tipos de superficies que tocan y la dinámica física del ambiente en la ciudad (The MetaSUB International Consortium, 2016).

Estudios basados en secuenciación de DNA de microbios revelan que las comunidades microbianas de ambientes urbanos son fuertemente influenciadas por sus ocupantes humanos (Hsu., *et al*, 2017). En un estudio realizado por Robertson *et al*, 2013, en el metro de Nueva York, se examinó el contenido del aire de varias estaciones mediante secuenciación masiva de amplicones para la subunidad pequeña ribosomal (16s rRNA). Los resultados mostraron que los *phylum* que predominan son actinobacterias, proteobacteria, firmicutes y en algunos casos bacteroidetes, los cuales son propios de la microbiota humana.

Por otro lado, en la red de transporte subterráneo de Hong Kong, Leung y cols, analizaron los bioaerosoles colectados en varios sitios y a través de la secuenciación del gen 16S rRNA hicieron la descripción de las comunidades microbianas. Los resultados arrojaron que los *phylum* más abundantes y que están presentes en todas las muestras son proteobacterias, actinobacterias, firmicutes y *Deinococcus-Thermus*, esto es muy similar a lo reportado por Robertson et al. 2013.

En Barcelona, se examinó la composición y abundancia de bioaerosoles recolectados en el transporte subterráneo de pasajeros durante un periodo de invierno. Los resultados mostraron que las bacterias existen en un orden de 10^4 bacterias/m³, que la composición es homogénea dentro de los trenes, plataformas y andenes, además es dominada por el taxa *Methylobacterium*. Con respecto a virus, se hallaron secuencias que pertenecen al virus de influenza del tipo A y B, lo que posiciona a este tipo de ambientes urbanos como sitios de transmisión de patógenos (Triadó-Margarit, *et al.*, 2016).

Dado que los sistemas de transporte de pasajeros no son los únicos ambientes urbanos donde las personas pasan gran parte de su tiempo, ni tampoco donde existen las condiciones y factores adecuados para el establecimiento, desarrollo y transmisión de microbios. Históricamente, la microbiota de los hospitales se ha estudiado apoyándose en métodos basados en cultivos, sin embargo, esto sesga el análisis a patógenos específicos en lugar de analizar el microbioma completo. Es claro que los hospitales son sitios donde existen microorganismos como bacterias, hongos y virus que han

desarrollado resistencia a fármacos y sustancias desinfectantes, lo cual tiene una profunda implicación en las infecciones adquiridas en hospitales (IAH) o nosocomiales, conduciendo a muertes de pacientes o infecciones de personas durante visitas, motivo por el cual también es importante su estudio (Tong *et al.*, 2017).

Dicha situación es atribuida a dos categorías de microorganismos transmitidos a través del aire: patógenos y microbios oportunistas ambientales, estos últimos afectan a pacientes comprometidos inmunológicamente. Las principales rutas de transmisión de IAH son: contacto directo, transporte por microgotas, por vectores y a través del aire, en esta última los microbios pueden permanecer por largos períodos de tiempo y viajar grandes distancias (Lai *et al.*, 2015).

En ese sentido, se han llevado a cabo algunos estudios donde se describe la diversidad de microbios presentes en hospitales, con un enfoque metagenómico aprovechando las herramientas de secuenciación de nueva generación. Uno de ellos consistió en un análisis a través de secuenciación masiva donde se examinaron muestras de los tractos respiratorios inferior y superior en niños que presentaban infección respiratoria y que habían resultado negativos para 6 bacterias y 15 virus por PCR. Los datos de las lecturas determinaron que eran portadores de RSV-A en mayor frecuencia, coronavirus y rinovirus, entre otros. Adicionalmente, pacientes ambulatorios y hospitalizados resultaron positivos para algunos virus respiratorios (Taboada *et al.*, 2014).

Los estudios que abordan el análisis de aires o superficies en hospitales son pocos, sin embargo, existen algunos. Uno de ellos es el realizado en un hospital afiliado al sistema de salud de San Diego, CA en donde se colectaron 63 muestras de aire de 8 sitios que incluyeron salas de espera, recepción y afuera del hospital a través de seis meses y se analizaron por secuenciación shotgun. Se identificaron segmentos genómicos relacionados a patógenos oportunistas como *Aspergillus*, *Penicillium* y *Stenotrophomonas*. Además, se determinó que algunos reads pertenecientes a *Stenotrophomonas* indican homología con una cepa clínica multirresistente a antibióticos, así como la presencia de secuencias que codifican a genes de resistencia (King P. *et al.*, 2016).

Los análisis de viomas han permitido un mejor entendimiento de la dinámica de las comunidades virales incluyendo diversidad taxonómica y genotípica, capacidad funcional, biogeografía y evolución. La metagenómica viral es también una poderosa herramienta que nos permite revelar la presencia de nuevos virus de DNA o RNA en una gran variedad de ambientes (Williamson S. J. 2011)

Es evidente que la metagenómica de bioaerosoles de ambientes urbanos puede ser una poderosa herramienta que drásticamente incrementará el número de secuencias virales reportadas a la fecha.

Es sorprendente que, a pesar de los avances recientes en técnicas de secuenciación de nueva generación, aún hay pocos estudios en la literatura que describan el viroma de ambientes urbanos (Dávila-Ramos S. *et al.*, 2019).

El impacto de los microbios en el aire y su contribución al ecosistema global es desproporcionalmente estudiado, aun cuando dichos microorganismos son ubicuos en el aire, por lo cual la metagenómica de virus en ambientes urbanos representa un reto, el cual puede ser superado con avances en las plataformas de secuenciación, con más y mejores bases de datos, así como avances en herramientas de bioinformática (Behzad H., *et al.*, 2015).

3.6 Clasificación de secuencias metagenómicas: métodos y retos.

Caracterizar la diversidad taxonómica de comunidades microbianas es uno de los objetivos primarios de los estudios metagenómicos. Una gran cantidad de microbios residen en diversos ambientes, por lo cual no pueden ser cultivados. Estas comunidades microbianas muestran variaciones espacio-temporales en su composición taxonómica. Un primer paso en un análisis metagenómico es descifrar la estructura de la comunidad, categorizando los microorganismos que residen en un ambiente y cuantificando la diversidad en términos de riqueza o abundancia. En el contexto de comunidades microbianas, el termino especie se refiere a un rango fundamental en la jerarquía taxonómica.

A pesar de la disponibilidad de herramientas para el *binning* (agrupamiento de *reads* o *contigs* a un grupo taxonómico particular) de las secuencias metagenómicas, hay dificultades que deben ser superadas para un correcto análisis, las cuales se discuten a continuación.

Existen limitaciones en la fase inicial de un proyecto metagenómico como la adecuada recolección de las muestras, su preparación, el aislamiento del material genético y la mejor elección de la plataforma de secuenciación. Desde luego, que en ocasiones se debe ajustar el protocolo a lo disponible en el centro de trabajo, no obstante, hay que cuidar lo mejor posible cada una de estas etapas para obtener los mejores resultados (Mande S. S. *et al.*, 2012).

4. Hipótesis

Los viromas presentes en cada ambiente urbano serán diferentes, sin embargo, será posible encontrar virus compartidos en algunos sitios muestreados de ambos ambientes debido a que son fácilmente transportados por aire y por las personas que los transitan.

5. Justificación

Es importante estudiar y analizar la diversidad de virus presentes en ambientes urbanos, considerando que son entidades abundantes en toda la biosfera y que algunos de ellos suelen ser patógenos para diferentes especies, incluido el ser humano. Considerando también que ambientes urbanos como los sistemas de transporte de pasajeros y hospitales son sitios con las condiciones adecuadas para el establecimiento, desarrollo y evolución de comunidades de microorganismos, entre ellos los virus, será interesante conocer las especies virales que están circulando y que pueden ser potencialmente dañinas para las personas que concurren en dichos sitios. Por lo que en este trabajo analizaremos los viromas presentes en muestras de aire de la sala de un hospital y en el aire de una taquilla y el interior de una vagón del Metro de la CDMX cercana, esto en los días 12 y 26 de noviembre de 2018, para establecer si existen similitudes entre los virus circulantes en ambos ambientes urbanos.

6. Planteamiento del problema

En México hasta el momento no existen estudios enfocados en el análisis de la diversidad y variabilidad de virus en ambientes urbanos como hospitales y sistemas de transporte de pasajeros.

Este estudio puede ser abordado con el uso de la tecnología de secuenciación de segunda generación, la cual permite identificar transferencias genómicas en los virus circulantes.

7. Objetivos

7.1 General:

- Describir el viroma presente en muestras de aire colectadas en la sala de espera del hospital 20 de noviembre, en taquillas y el interior de un vagón del metro de la Ciudad de México.

7.2 Particulares:

- Determinar las secuencias de los virus de RNA presentes en el aire de una sala de espera del hospital 20 de noviembre, en taquillas y el interior de un vagón del metro de la Ciudad de México
- Determinar la diversidad taxonómica viral presente en las muestras de aire.
- Comparar la diversidad viral de ambos ambientes urbanos mediante el análisis de variantes.

8. Estrategia experimental

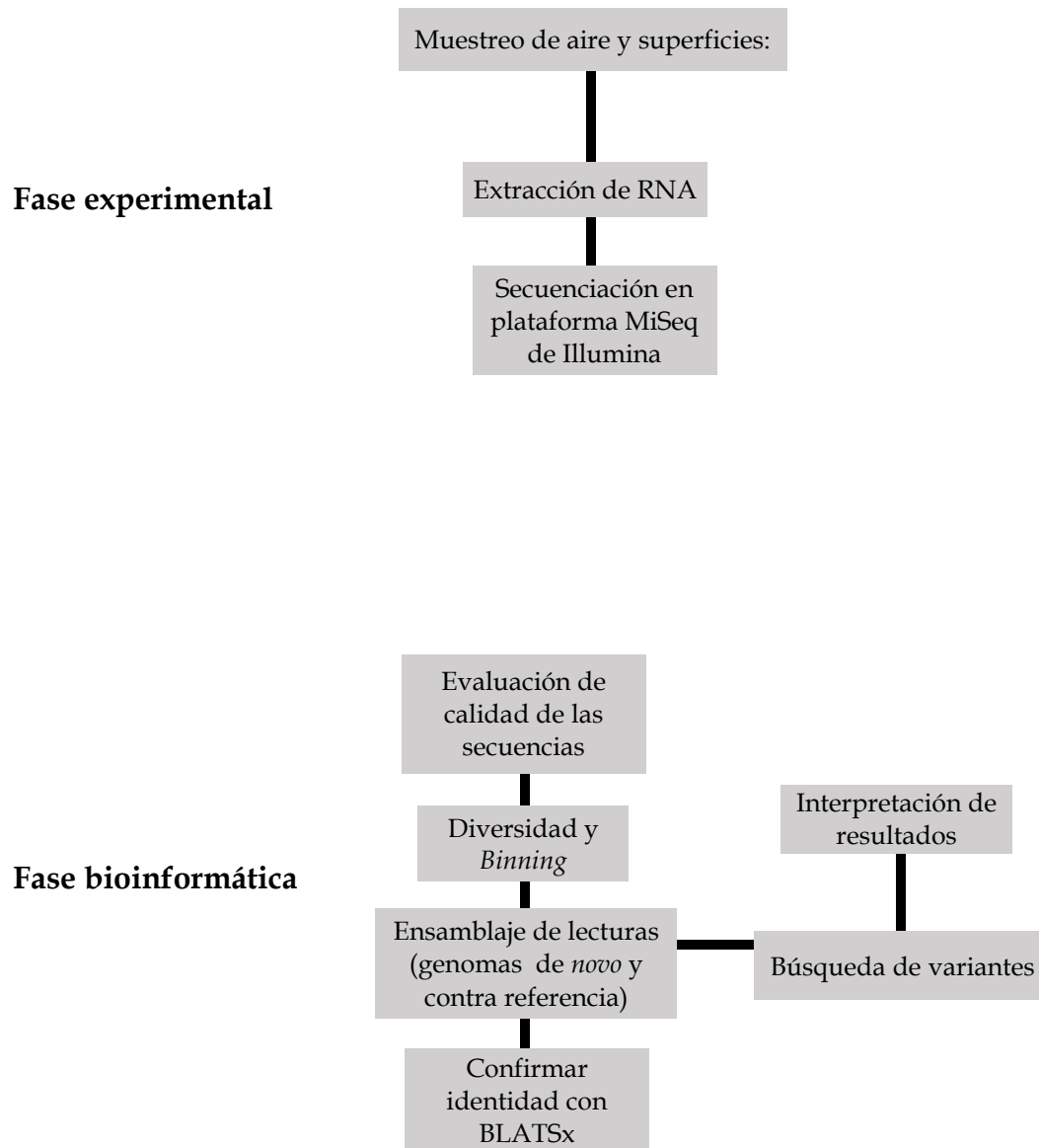


Figura 7. Diagrama del plan general de trabajo

9. Materiales y métodos

9.1 Muestras

Debido a la naturaleza del estudio, por tratarse de lugares públicos y que pertenecen a instituciones gubernamentales, se tramitaron los permisos correspondientes.

Los sitios elegidos para tomar las muestras de aire fueron los siguientes: el área de taquilla y el interior de un vagón de la estación hospital 20 de noviembre de la CDMX, para el hospital se eligió una sala de espera. La toma de muestras se realizó los días 12 y 26 de noviembre, considerando que están dentro de una temporada con un alto índice de incidencias de enfermedades respiratorias, dando como resultado un total de seis muestras.

Las muestras se colectaron con un dispositivo de captación de aire llamado Coriolis Air® (Bertin Instruments), el cual conlleva un vial que contenía 10 ml de solución salina de Hank's con las siguientes características: tiene un pH balanceado entre 7.2 y 8.0 y su osmolaridad es de 320 mOsm/kg, lo que es bueno para preservar las partículas virales en un ambiente osmóticamente equilibrado y sin nutrientes para el crecimiento de otros microorganismos como bacterias. El dispositivo se programó para funcionar a una velocidad de 200 litros/segundo durante tres ciclos de 10 minutos. Al inicio de cada ciclo se registraron las condiciones de temperatura y humedad.

El vial que contenía los 10 ml de solución salina de Hank's con las partículas recuperadas fue centrifugado a 14,000 rpm durante 5 minutos para bajar las partículas grandes como bacterias y dejar en el sobrenadante las partículas más pequeñas como virus, ambas fases se fraccionaron en volúmenes de 1000 µl y se almacenaron a -70°C hasta su uso. Para el proyecto solo se usaron los sobrenadantes puesto que se considera que por su tamaño la mayoría de los virus quedarían en suspensión y se eliminarían las bacterias completas, lo que permitiría enriquecer el RNA viral en la muestra.

9.2 Extracción de ácidos nucleicos (RNA)

Para aislar el RNA de los sobrenadantes, se utilizó el kit QIAamp Viral RNA Mini Kit (QIAGEN, CA) y con la finalidad de concentrar una buena cantidad de RNA y siguiendo las instrucciones del fabricante se procedió de la siguiente manera:

Se concentraron 5 alícuotas de 140 μ l de cada una de los sobrenadantes de las seis muestras, se les añadió solución de lisis con acarreador de RNA, se hicieron los lavados correspondientes para eliminar el exceso de solución de lisis y finalmente se eluyeron en 30 μ l de agua inyectable. La cantidad de RNA total se midió en un equipo NanoDrop 2000 (Thermo Scientific).

Con el objetivo de verificar si el RNA estaba integro, se corrió 1 μ l de la elución de RNA en gel de agarosa teñido con bromuro de etidio, esto para asegurar que se enviaba en buen estado a la Unidad de Síntesis y Secuenciación de DNA del Instituto de Biotecnología de la UNAM, donde fueron procesadas de acuerdo a los protocolos de la Unidad, los cuales consisten en lo siguiente.

Se separaron las moléculas de RNA mediante electroforesis capilar en un Agilent 2100 Bioanalyzer para determinar el RIN (RNA Integrity Number), que indica el grado de degradación de las muestras de RNA.

Debido a que las plataformas de secuenciación de nueva generación como Illumina requieren de DNA para ser procesado, el RNA que se aisló de las muestras es retrotranscrito a cDNA el cual ya es utilizado para la plataforma de secuenciación.

Posteriormente, se llevó a cabo la construcción de bibliotecas genómicas lo cual consistió en fragmentar el DNA y añadir índices y adaptadores en los extremos 5' y 3', esto se realiza ya que los adaptadores sirven como cebadores durante la secuenciación y los índices ayudan a identificar las muestras de cada corrida.

La secuenciación se efectuó en la plataforma MiSeq de Illumina *paired end*, la cual se basa en la incorporación de deoxiribonucleótidos trifosfatados (dNTPs) que tienen la

capacidad de ser terminadores reversibles de la polimerización, por ello, se conoce como secuenciación reversible por síntesis. Esta plataforma integra los pasos de generación de grupos de segmentos de DNA, amplificación de los mismos, secuenciación y análisis de datos en el mismo instrumento. Las secuencias *paired end* facilitan el ensamblado de los genomas. Los reads reportados por la Unidad del Instituto de Biotecnología tuvieron una longitud de 2 X 75 nucleótidos.

9.3 Análisis de calidad de las secuencias

Los datos reportados por la Unidad de Síntesis y Secuenciación de DNA consisten en archivos en formato FASTQ, los cuales son archivos de texto que contienen secuencias de nucleótidos y su correspondiente score de calidad (PHRED score).

Todas las secuencias FASTQ fueron sometidas a un análisis de calidad utilizando en software FastQC, el cual muestra un reporte gráfico para las secuencias de cada read, incluyendo principalmente el promedio del score de calidad, la posición de los nucleótidos, el % de GC, secuencias sobrerrepresentadas y si aún contienen adaptadores. Por fortuna, en ninguno de los archivos fastq se encontraron nucleótidos de baja calidad, por lo que no fue necesario eliminarlos.

Continuando con el procesamiento de los datos, el siguiente paso fue la remoción de adaptadores presentes, para esto se utilizó la herramienta PEAT, la cual funciona específicamente para secuencias *paired end*. La línea de comandos es la siguiente:

```
./PEAT_linux paired -1 PATH/archivo_R1.fastq -2 PATH/archivo_R2.fastq -o  
PATH/archivo_R1_sin_adaptadores.fastq PATH/archivo_R2_sin_adaptadores.fastq --  
adapter_contexts report.txt
```

Con el objetivo de agilizar el mapeo y aprovechar mejor los recursos del equipo de cómputo se eliminaron secuencias repetidas al 100% de identidad, para esto se empleó la herramienta CD-HIT, en la cual los comandos utilizados son los siguientes:

```
./cd-hit-dup -i PATH/archivo_R1_sin_adaptadores.fastq -i2  
PATH/archivo_R2_sin_adaptadores.fastq -o PATH/archivo_R1_sindups.fastq -o2  
PATH/archivo_R2_sindups.fastq
```

A partir de estos archivos limpios de adaptadores y sin secuencias repetidas, se trabajó en el análisis metagenómico.

9.4 Reclutamiento de fragmentos

Al tratarse de un estudio metagenómico, particularmente enfocado a la detección de virus, se siguió la estrategia a continuación descrita:

Se descargó una base de datos de secuencias de nucleótidos virales, para esto se eligió la base de datos de NCBI. Para facilitar la búsqueda y el reclutamiento de *reads* con señales virales, se construyó un índice a partir de dicha base de datos. Para esto se utilizó la herramienta SMALT con la siguiente línea de comandos:

Para crear el índice:

```
PATH/smalt index -k 13 -s 4 virusdb_index PATH/virusdb.fasta
```

Para reclutar los *reads* a partir del índice:

```
PATH/smalt map -f samsoft -o archivo.sam -y 0.6 -i 320 -j 80 -l pe  
PATH/virusdb_index PATH/archivo_R1_sindups.fastq PATH/archivo_R2_sindups.fastq
```

Como resultado del mapeo o reclutamiento de *reads* se obtiene un archivo .sam, a partir del cual se pueden extraer los *reads* que tuvieron correspondencia con la base de datos de virus. Con ayuda de la herramienta SAMtools es posible manipular el archivo .sam, las líneas de comandos ejecutadas fueron las siguientes:

Para convertir a un archivo más compacto (.bam)

```
PATH/samtools view -b -S -o PATH/archivo.bam PATH/archivo.sam
```

Extraer y escribir en otro archivo .sam los *reads* que tuvieron correspondencia con la base de datos de virus

```
PATH/samtools view -f 2 PATH/archivo.bam | tee archivo_hits.sam
```

Extraer y escribir en otro archivo los *reads* que no tuvieron correspondencia con la base de datos de virus

```
PATH/samtools view -F 2 PATH/archivo.bam | tee archivo_no_hits.sam
```

Contar el número de *reads* que correspondieron a la base de datos de virus

```
PATH/samtools view -c -f 2 archivo.bam
```

Finalmente, los *reads* que correspondieron a la base de datos de virus se escribieron en un archivo FASTQ con el uso de los siguientes comandos

Reads forward:

```
PATH/cut -f 1,2,10,11 archivo.sam | awk '$2==83||$2==99' | cut -f 1,3,4 | awk  
'{print $1"\n"$2"\n"+" "\n"$3}' | sed "s/NS500/@NS500/g" > archivo_virus_F.fastq
```

Reads reverse:

```
PATH/cut -f 1,2,10,11 archivo.sam | awk '$2==147||$2==163' | cut -f 1,3,4 | awk  
'{print $1"\n"$2"\n"+" "\n"$3}' | sed "s/NS500/@NS500/g" > archivo_virus_R.fastq
```

Cabe mencionar que los archivos FASTQ para cada sitio de muestreo son dos, que corresponden a un *forward* y un *reverse*, para los cuales se hizo exactamente el mismo procedimiento arriba descrito y a partir de los archivos FASTQ que hicieron match con la base de datos de virus, se continuó con las siguientes etapas del análisis.

9.5 Análisis de diversidad

Para obtener un panorama general de la diversidad viral presente en los sitios de muestreo, se utilizó la herramienta Kaiju para asignar taxonómicamente, esta herramienta utiliza una base de datos de proteínas de NCBI, toma los *reads* que están en nucleótidos y los traduce virtualmente a proteínas, asignando cada *read* a un grupo taxonómico y agrupándolos en un taxa particular. En este caso, se indicó a Kaiju que la asignación fuera a nivel de familia, sin embargo, al existir secuencias virales que corresponden a virus sin este nivel taxonómico, los agrupo a nivel de orden. La línea de comandos empleada para Kaiju es la siguiente:

Para crear una base de datos de proteínas virales

```
PATH/./kaiju-makedb -s viruses
```

Al finalizar la ejecución de este comando, nos quedan tres archivos importantes, “kaiju_db_viruses.fmi”, “nodes.dmp” y names.dmp”.

Para llevar a cabo la asignación taxonómica:

```
PATH/./kaiju -t nodes.dmp -f kaiju_db_viruses.fmi -i archivo.fasta -o  
PATH/archivo.txt -a mem -v
```

Para añadir nombres a los *reads* clasificados:

```
PATH/./kaiju-addTaxonNames -t nodes.dmp -n names.dmp -i PATH/archivo.txt -o  
PATH/archivo_nombres.txt
```

El resultado de esto es un archivo de texto con 8 columnas, las cuales se explican en la figura de abajo:

1. Indica si el *read* fue clasificado o no clasificado (C o U)
2. ID del *read*
3. ID del taxón en NCBI al *read* asignado
4. Longitud o score del mejor alineamiento usado para la clasificación
5. El ID del taxón de todas las secuencias en la base de datos con el mejor alineamiento
6. El número de acceso de todas las secuencias en la base de datos con el mejor alineamiento
7. Fragmento de la secuencia que tuvo correspondencia
8. Nombre del nivel taxonómico o incluso del género viral al que pertenece el *read*

Con la intención de construir diagramas de *pie*, se utilizó un *script* dentro de Kaiju para obtener un *output* (archivo de salida) para Krona. El comando fue el siguiente:

```
PATH/./kaiju2krona -t nodes.dmp -n names.dmp -i PATH/archivo.txt -o  
PATH/archivo.krona
```

Finalmente, para obtener el diagrama de *pie* a partir del archivo para Krona, se utilizó la herramienta Kronatools, la cual contiene un *script* para visualizar en el navegador de internet el gráfico. Se empleó la línea de comandos:

```
PATH/./ImportText.pl -o PATH/archivo.html PATH/archivo.krona
```

Dado que en un estudio metagenómico los *reads* son informativos, pero no reflejan con certeza la diversidad encontrada, el siguiente paso fue ensamblar genomas *de novo* a partir de los archivos FASTQ que hicieron *match* con la base de datos de virus. Para esto se utilizó la herramienta SPAdes con la variante meta, especializada en realizar ensambles a partir de datos complejos, como metagenomas. Para esto, la línea de comando utilizada fue:

```
PATH./spades.py -o PATH/archivo_meta --meta -1 PATH/archivo_virus_F.fastq -2  
PATH/archivo_virus_R.fastq
```

SPAdes es una herramienta que utiliza *reads paired end* en formato FASTQ y FASTA, emplea diferente tamaño de *k-meros* para conseguir los mejores ensamblados, como resultado nos otorga un archivo .FASTA que contiene el ID por nodo, la longitud del contig ensamblado y la secuencia en nucleótidos.

Con la finalidad de confirmar la diversidad viral presente en las muestras y corroborar la identidad de los ensambles, se sometieron a asignación taxonómica con Kaiju nuevamente y a una búsqueda a través de alineamientos múltiples con la herramienta BLAST+ instalada localmente a través de la variante BLASTx, con la cual se utilizó una base de datos de proteínas virales descargada de NCBI. El uso de *contigs* para el análisis posterior tiene la ventaja adicional de optimizar los recursos de cómputo y el delimitar la búsqueda a una base de datos de proteínas virales para evitar sesgos en los resultados. Los comandos utilizados para realizar la búsqueda en BLAST+ fueron los siguientes:

Para añadir la base de datos a BLAST+:

```
PATH/makeblastdb -in proteinas_virales.fasta -dbtype prot -out  
proteinas_virales_db
```

Para realizar la búsqueda de los *contigs* ensamblados, los cuales están dentro de un archivo de texto simple (.txt):

```
PATH/blastx -db proteinas_virales_db -query contigs.fasta -out contigs.xml -outfmt  
6 -num_threads 4 -max_target_seqs 5
```

Cabe mencionar que elegir la extensión .xml y el formato de salida 6 nos da un archivo de salida en formato tabular que facilita la identificación de los resultados. Por otra parte, el ajuste en “num_threads 4” optimiza el uso de los núcleos lógicos del procesador del equipo de cómputo y “max_target_seqs 5” limita la búsqueda a los primeros cinco resultados con que se alinean los *contigs*. Explicando las secciones de las que consta el archivo son las siguientes:

1. ID de la secuencia
2. ID del genoma de referencia
3. Porcentaje de alineamientos idénticos
4. Longitud del alineamiento
5. Número de alineamientos incorrectos
6. Número de espacios vacíos en la secuencia
7. Inicio del alineamiento en la secuencia problema
8. Fin del alineamiento en la secuencia problema
9. Inicio del alineamiento en la secuencia de referencia
10. Fin del alineamiento en la secuencia de referencia
11. Valor E
12. Bit score

9.6 Análisis de variantes (SNVs) en los virus compartidos en todas las muestras

De acuerdo a los resultados arrojados por el reclutamiento de *reads*, y a la asignación taxonómica a nivel de *reads* y de *contigs* mediante Kaiju, existen tres familias virales que son las más frecuentes y compartidas en todas las muestras. Teniendo esto en cuenta, se realizaron tres mapeos para cada muestra, cada uno con una base de datos de cada familia frecuente. Como resultado de esto y con ayuda de la herramienta Tablet, se pueden visualizar de manera gráfica las secuencias de referencia con las que hay un mayor número de *reads* reclutados, así mismo la región donde los *reads* alinean mejor. Esto nos ayuda, además, a elegir la región con respecto al genoma de referencia, que se tomó para analizar las variantes en las regiones de los virus en común en todas las muestras.

Otro criterio para delimitar la región que fue sometida al análisis de variantes es, desde luego, la profundidad o cobertura de la secuenciación, la cual debe ser mayor o igual a 50 ($\geq 50x$), lo cual indica el número de veces que fue secuenciado ese genoma o región.

Para conocer la cobertura, la herramienta que se utilizó fue QualiMap. Entre otras cosas, nos describe el tamaño de la secuencia de referencia, el número de reads totales, número de reads mapeados y no mapeados, porcentaje de GC y la cobertura a través del genoma de referencia, lo cual es fundamental para poder llevar a cabo el análisis de variantes.

La herramienta VirVarSeq es la que se empleó para dicho análisis. Esta herramienta permite la detección de variantes de baja y de muy baja frecuencia a nivel de codón dentro de una población viral en datos de secuenciación de plataformas de nueva generación como Illumina. VirVarSeq incluye un *script* en el cual se indican varios parámetros que delimitan la región a analizar, los cuales son:

1. *startpos*: posición en la secuencia de referencia a nivel de nucleótido donde inicia la secuencia consenso.

2. *endpos*: posición en la secuencia de referencia a nivel de nucleótido donde termina la secuencia consenso.
3. *region_start*: posición de la secuencia de referencia a nivel de nucleótido donde empieza el análisis de variantes. Puede ser igual o mayor a *startpos*.
4. *region_len*: número de posiciones en codones que necesitan ser cubiertos en la tabla de variantes.

Como resultado de ejecutar VirVarSeq, obtenemos una carpeta que incluye, entre otras cosas, una tabla donde se indica el nombre de la muestra, la posición por codón, el codón de la secuencia de referencia, el codón encontrado en el análisis con las variantes, ya sea en la primera, segunda o tercera posición, el aminoácido que es codificado por el codón de referencia, el aminoácido codificado por las variantes encontradas y la frecuencia con que se encuentran las variantes en una posición. Se considero una frecuencia de 0.01, lo que representa que las variantes están en el 10% del total en una posición.

Finalmente, para determinar si las variantes presentes en las regiones ensambladas de los virus que están presentes en todas las muestras son o no las mismas, se realizó un análisis de correlación de Pearson, donde se reporta si hay relación entre las variantes encontradas por posición en el genoma en una especie viral que circula en todos los sitios de muestreo.

10. Resultados

Promedios de temperatura y humedad

Sitio	Temperatura (°C)	Humedad (%)
Aire taquilla 12 nov	21.66	47.33
Aire vagón 12 nov	23.33	41
Sala de espera 12 nov	22.33	41
Aire taquilla 26 nov	21.00	48.33
Aire vagón 26 nov	20	40.33
Sala de espera 26 nov	19.6	53

Tabla 3. Se muestran los promedios de las tres lecturas de las condiciones de temperatura y humedad. Se observa que las variaciones son mínimas, excepto en la sala de espera del hospital el día 26 de noviembre, en donde la humedad es mayor.

Los promedios de las lecturas de temperatura y humedad se mantuvieron estables, sin embargo, el aumento de temperatura es mayor el día 12 de noviembre, mientras que disminuye para el día 26, lo que puede deberse a la cercanía a la temporada de invierno. Respecto a la humedad, esta es menor el día 12 de noviembre, probablemente debido a que la cantidad de personas era mayor en el metro y en el hospital.

10.1 Extracción de ácidos nucleicos

A partir de cada muestra se realizó la extracción de RNA como se describió en Materiales y métodos. La cantidad de RNA total aislada de cada muestra fue cuantificada con la ayuda de un NanoDrop 2000, y los resultados se muestran en la Figura 8. No hay relación entre la cantidad de RNA obtenido, ya que a mayor humedad y menor temperatura, la cantidad de RNA aislado fue mayor, ocurriendo lo contrario en condiciones de menor humedad y mayor temperatura, siendo mayor la cantidad de RNA en el aire del vagón en ambas fechas de muestreo.

Con el objeto de obtener las secuencias, se enviaron 30 µl de RNA de cada muestra a la Unidad de Síntesis y Secuenciación de DNA del IBT de la UNAM, donde se verificó la

concentración de RNA, se construyeron las bibliotecas y se obtuvieron secuencias *pair end* utilizando a plataforma MiSeq de Illumina .

RNA total extraído para cada muestra

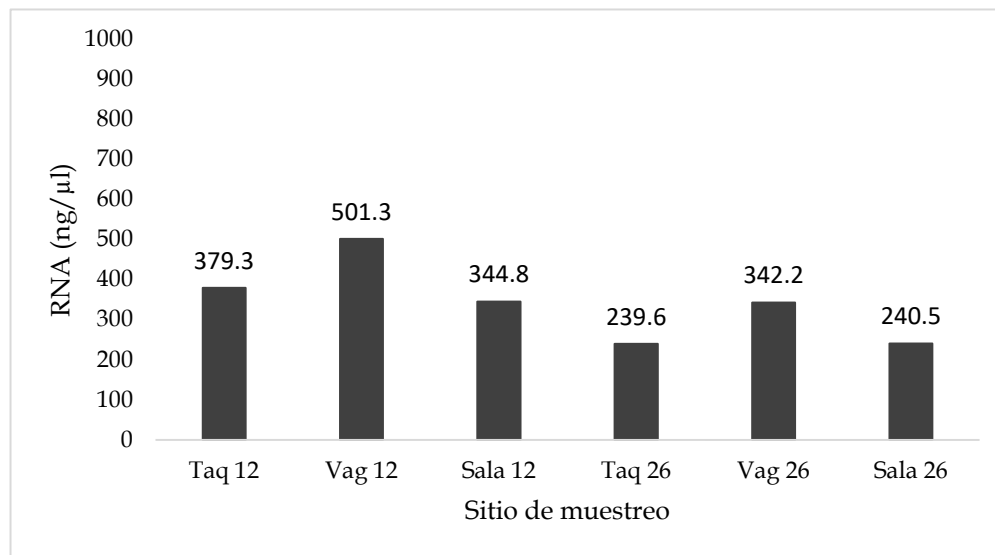


Figura 8. Cantidad de RNA total aislado de cada muestra. En general, en las muestras que corresponden al día 12 de noviembre se obtuvo la mayor cantidad de RNA, en contraste con las muestras colectadas el día 26 de noviembre.

10.2 Análisis de calidad de los datos de secuenciación

El número de reads para cada muestra que fueron entregados por la Unidad Universitaria de Secuenciación Masiva y Bioinformática de la UNAM se reportan en la Figura 9. La muestra con el mayor número de reads corresponde al aire colectado en la sala de espera del hospital 20 de noviembre el día 26 de noviembre con 5,724,105x2 reads, mientras que la muestra que tiene el menor número de reads es la que corresponde al aire de la sala de espera del hospital 20 de noviembre del día 12 de noviembre con 2,159,224x2. Cabe mencionar que no se observó correlación en el número de *reads* obtenidos con respecto a la cantidad de RNA aislado.

Número total de *reads* entregados por el IBT

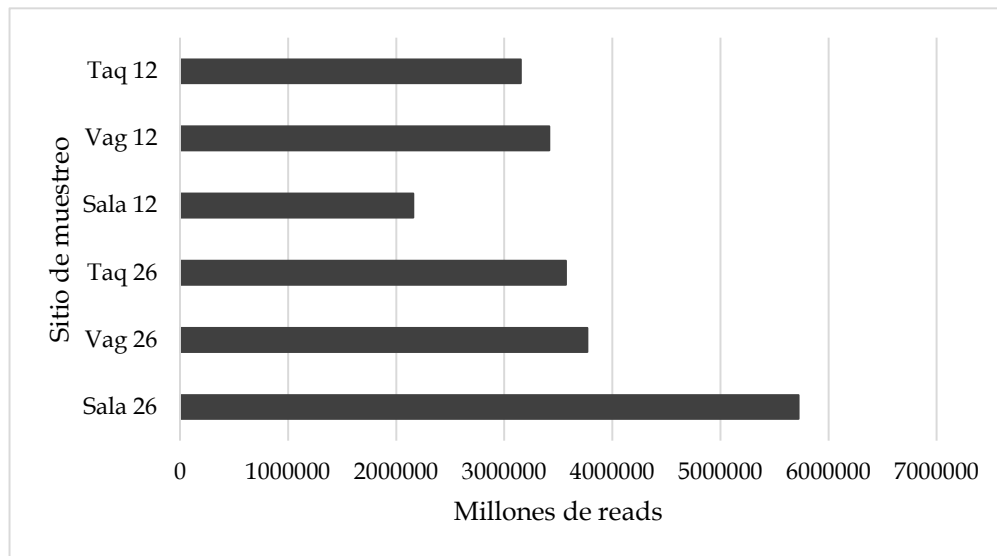


Figura 9. Número de *reads* entregados por el IBT. La muestra con mayor número de *reads* entregados corresponde a la del aire de la sala de espera del hospital del día 26 de noviembre, mientras que la tuvo menos *reads* es la misma, pero del día 12 de noviembre,

Como parte del proceso de verificación de calidad y filtrado de secuencias, se utilizó la herramienta FASTQC para obtener el score de calidad. De acuerdo a lo observado en FASTQC, todas las muestras tienen un score de PHRED mayor a 30 (Figura 10), sin embargo, debido al tipo de secuenciación en la plataforma MiSeq de Illumina, es común que la calidad disminuya a medida que aumenta el llamado de bases o *base calling*. El porcentaje de GC para todas las muestras presenta múltiples picos, lo que indica que la naturaleza de la muestra es heterogénea y diversa.

Calidad de un archivo .fastq visualizado con FASTQC

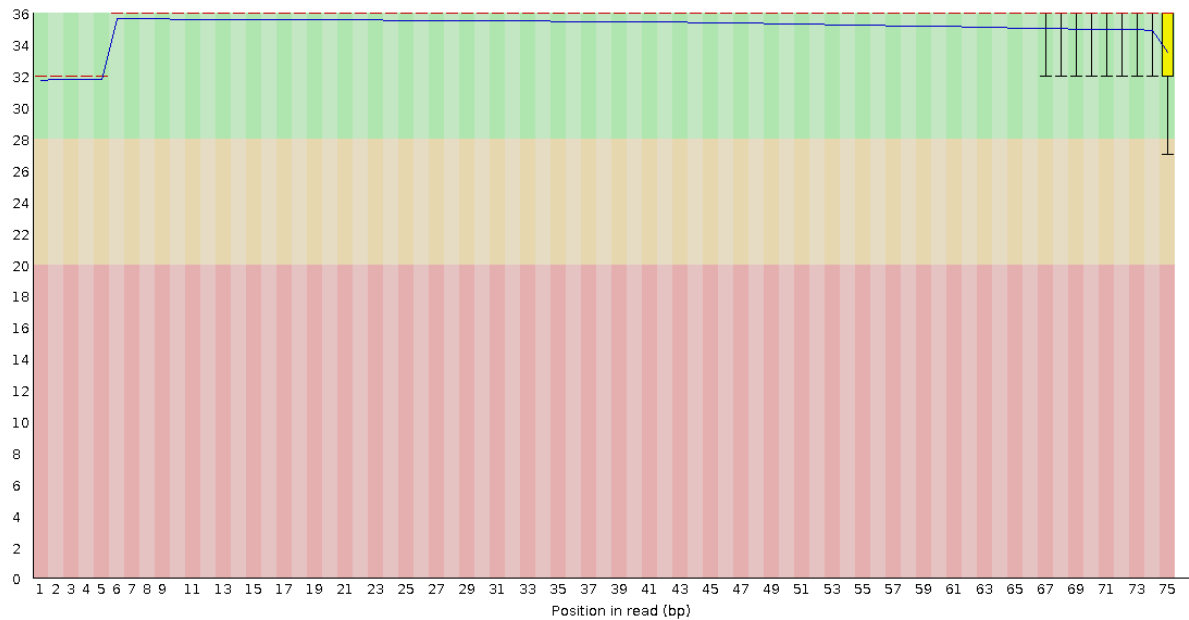


Figura 10. Imagen de un archivo .fastq visualizado con FASTQC. Se muestra la longitud de los *reads* en el eje X y el score de PHRED en el eje Y, así mismo, que la calidad de los nucleótidos disminuye en los últimos nucleótidos.

Debido a que FASTQC indicó que todas las secuencias carecían de nucleótidos de baja calidad y de adaptadores, se procedió a eliminar únicamente las secuencias repetidas. Una vez hecho esto, el número de reads disminuyó considerablemente, lo cual se muestra en la figura 11, puesto que alrededor del 20% al 50% de las secuencias originales fue eliminado.

Las secuencias resultantes de esta etapa, se consideraron como aptas para continuar con los demás análisis.

Cantidad de reads sin duplicados ni adaptadores

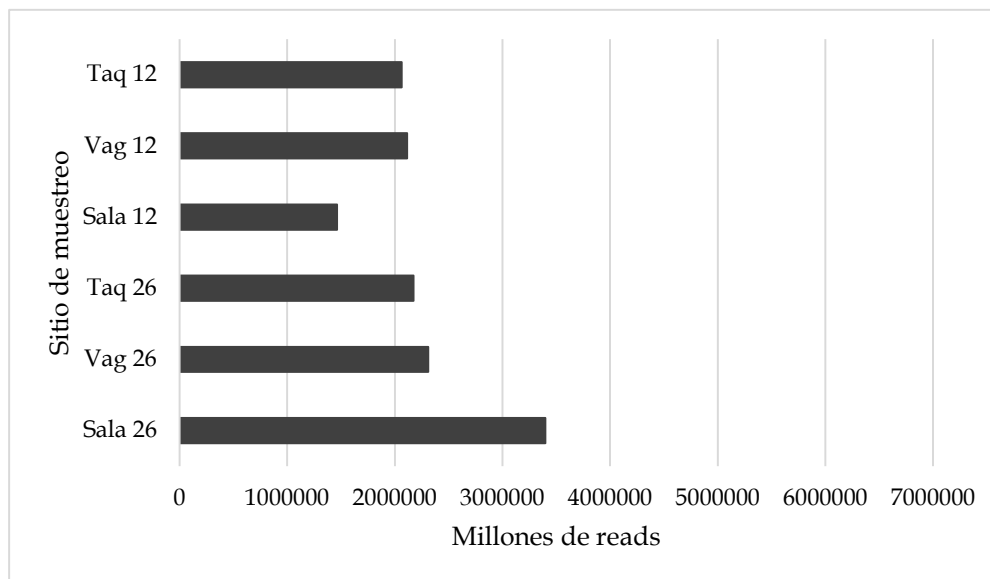


Figura 11. Número de reads sin adaptadores ni secuencias repetidas (x2). Se eliminaron del 20% al 50% del total de secuencias entregadas.

10.3 Asignación taxonómica

Con el fin de obtener un panorama general de la diversidad viral que se encuentra circulando en el aire de ambos ambientes, se hizo una asignación taxonómica con la herramienta Kaiju. Los resultados se muestran en la Figura 12, donde se observa que el mayor número de *reads* virales para cada muestra corresponden a las familias *Adenoviridae*, seguido de *Baculoviridae* y *Myoviridae*. Cabe mencionar que se lograron clasificar algunos reads para la familia *Orthomyxoviridae*, específicamente para el virus de influenza A, en las muestras de aire del vagón del 12 de noviembre, de la sala de espera del hospital y del aire del vagón del día 26 de noviembre, sin embargo, fueron muy pocos.

Diversidad viral a nivel de reads

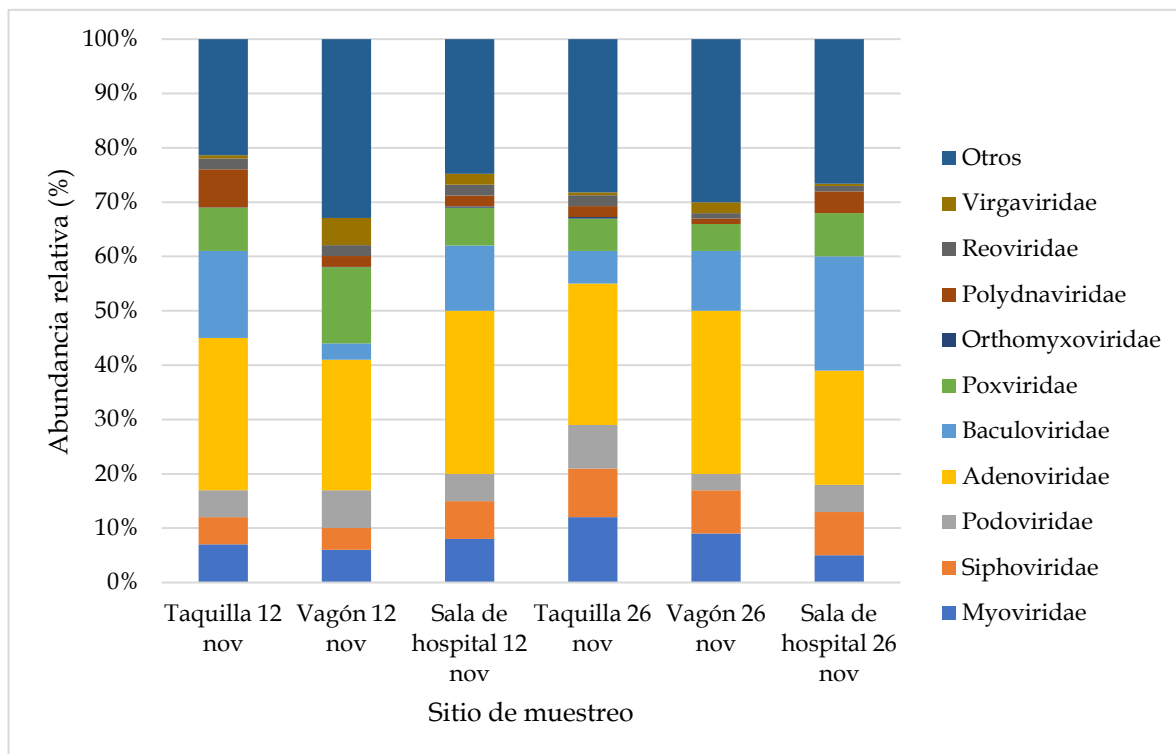


Figura 12. Diversidad viral de *reads* a nivel de familia. Las familias virales más frecuentes son *Myoviridae*, *Siphoviridae* y *Podoviridae*, estas pertenecen al orden *Caudovirales*, así mismo las familias *Adenoviridae*, *Baculoviridae* y *Poxviridae*.

Es interesante que en ambos ambientes y en todos los sitios de muestreo exista la presencia de secuencias que pertenecen a la familia *Adenoviridae*, particularmente, al hacer un screening con mayor resolución, determinamos que los *reads* pertenecen específicamente a adenovirus humano tipo C, el cual es un virus patógeno que ocasiona diferentes enfermedades en los seres humanos (Almand E. *et al.*, 2017).

Al respecto, el adenovirus humano del tipo C (Figura 13).es un virus con genoma de DNA de cadena doble, lineal no segmentado que va de los 30 a 38 kpb. Son virus no envueltos y ocasionan conjuntivitis, gastroenteritis y enfermedades de las vías respiratorias.

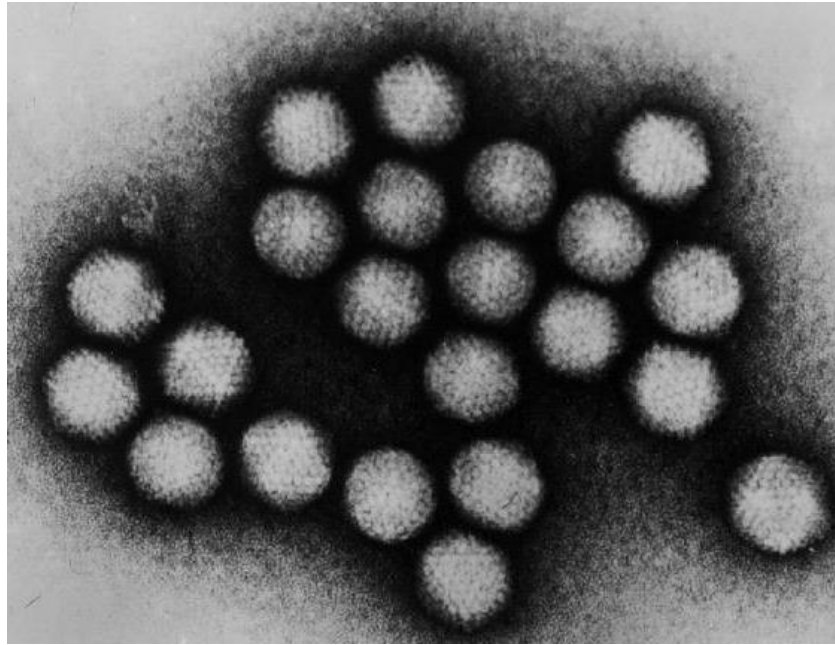


Figura 13. Microfotografía electrónica del adenovirus humano del tipo C.

Otros virus identificados son el virus de la polihedrosis nuclear de *Spodoptera littoralis* (polilla común) y un fago de *Staphylococcus sp* (Figura 14)(bacteria), lo cual es interesante ya que el humano, la polilla y bacterias están contribuyendo material al viroma de muestras de aire.

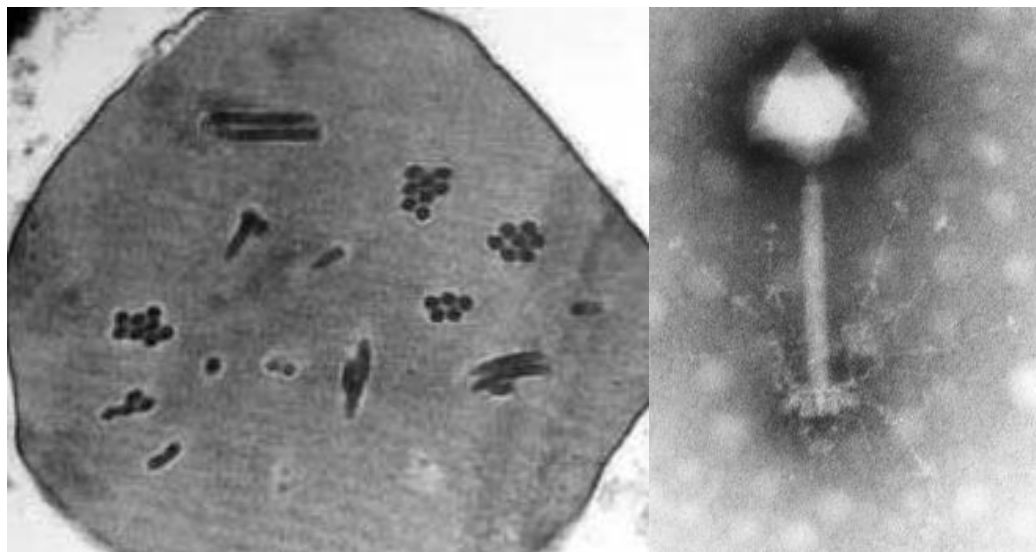


Figura 14. Microfotografía electrónica del virus de poliedrosis nuclear de *Spodoptera littoralis* (izquierda) y de un fago de *Staphylococcus sp* (derecha).

Una vez conseguido el panorama general de la diversidad viral presente, se continuó con el ensamble de *contigs*, con la intención de confirmar la diversidad además de identificar regiones genómicas para realizar el análisis de variantes y optimizar la capacidad del equipo de cómputo, debido a que la verificación de la asignación taxonómica tradicionalmente se hace mediante un blastx que es computacionalmente costoso.

Como se mencionó en materiales y métodos los *contigs* de cada muestra se construyeron con la herramienta SPAdes, y se obtuvieron en promedio 25 *contigs* por cada muestra que oscilan entre 80 y 1000 nucleótidos de longitud. Una vez obtenidos los *contigs* fueron sometidos a asignación taxonómica con Kaiju como se muestra en la Figura 15. Además, la identidad de los *contigs* fue verificada mediante la herramienta BLAST+ en la modalidad BLASTX. Los resultados se muestran en las tablas 1 a 6, que muestran solo los primeros 10 *contigs* de cada muestra.

Diversidad viral a nivel de *contigs*

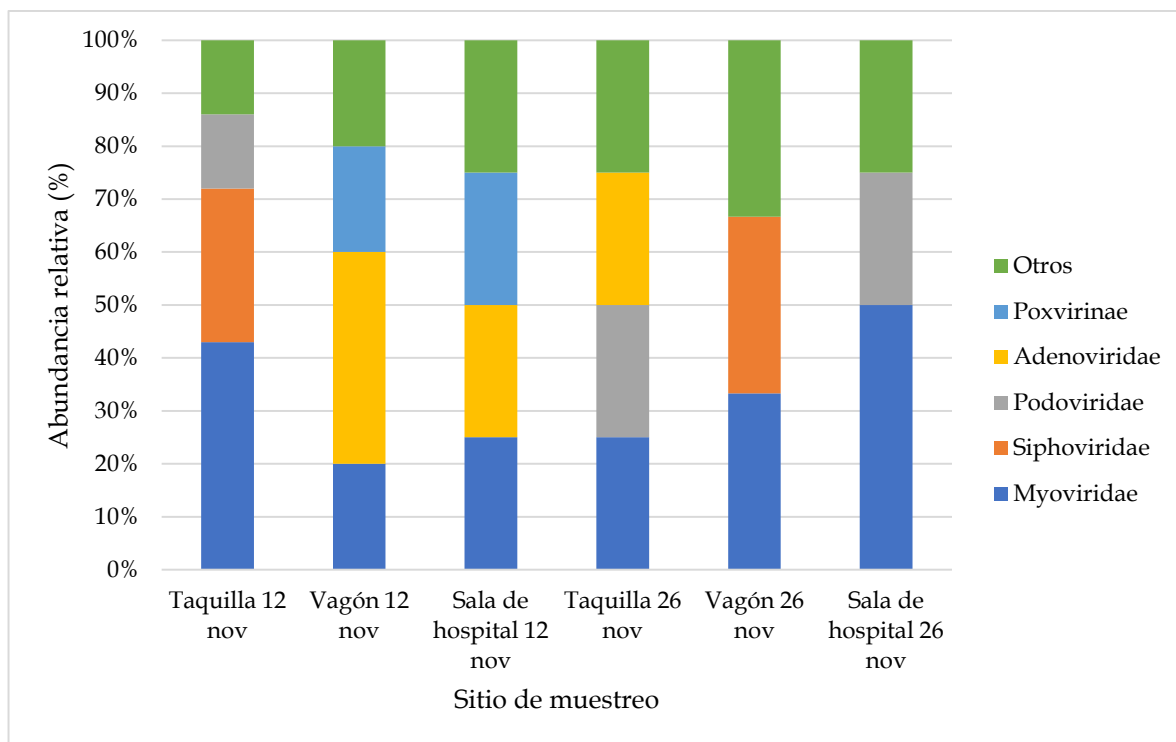


Figura 15. Diversidad viral de *contigs* a nivel de familia. Se observa que en todas las muestras siguen siendo frecuentes las familias virales *Myoviridae*, mientras que las familias *Siphoviridae*, *Podoviridae*, *Adenoviridae* y *Poxviridae* están presentes de manera heterogénea.

Es importante señalar que cuando se realiza el análisis con Kaiju a nivel de *reads* se identifica una mayor cantidad de especies virales que con *contigs*, esto se debe a que existen miles de *reads* que pertenecen a una familia viral en particular, mientras que en algunas familias virales los *reads* obtenidos no son contiguos por lo que no es posible ensamblar *contigs*.

Contig	Longitud	Número de acceso	% Identidad	E-value	Nombre	Familia
1	508	YP_195211.1	87.574	4.56e-93	photosystem II D1 protein [Synechococcus phage S-PM2]	Myoviridae
2	485	YP_007518291.1	93.548	1.74e-104	photosystem II D1 protein [Synechococcus phage S-RIM8 A.HR1]	Myoviridae
3	322	YP_009293522.1	98.980	1.84e-65	transposase [Salmonella phage SJ46]	Myoviridae
4	308	YP_004301069.1	31.373	0.13	hypothetical protein 65p232 [Aeromonas virus 65]	Myoviridae
5	289	YP_009624869.1	67.750	4.41e-35	Beta-lactamase SHV-2 [Escherichia phage RCS47]	Myoviridae
6	281	NP_037691.1	100	3.43e-63	IS903 transposase [Escherichia virus HK022]	Myoviridae
7	245	YP_009301715.1	87.654	7.59e-37	photosystem II protein D1 [Cyanophage S-RIM32]	Myoviridae
8	236	YP_009617199.1	86.111	3.15e-12	lactose operon repressor [Enterobacteria phage DE3]	Myoviridae
9	226	YP_009336789.1	81.081	4.61e-39	hypothetical protein 1 [Hubei tombus-like virus 8]	Riboviria
10	209	YP_006607828.1	40.625	2.8	hypothetical protein [Apocheima cinerarium nucleopolyhedrovirus]	Baculoviridae

Tabla 4. Resultados de BLASTx de la muestra de aire de taquilla del día 12 de noviembre a partir de los *contigs* obtenidos

Contig	Longitud	Número de acceso	% Identidad	E-value	Nombre	Familia
1	1001	YP_009301715.1	86.787	0.0	photosystem II protein D1 [Cyanophage S-RIM32]	Myoviridae
2	497	YP_009624869.1	66.917	8.58e-52	Beta-lactamase SHV-2 [Escherichia phage RCS47]	Myoviridae
3	363	YP_001551773.1	100	7.59e-54	control protein E4orf4 [Human mastadenovirus C]	Adenoviridae
4	342	NP_040533.1	100	2.18e-11	fiber [Human mastadenovirus C]	Adenoviridae
5	332	YP_007237719.1	34.286	3.2	putative virion structural protein [Erwinia phage phiEaH2]	Myoviridae
6	325	YP_009118491.1	39.024	2.3	immediate early protein 2 [Condylorrhiza vestigialis MNPV]	Baculoviridae
7	304	YP_089649.1	96.970	1.03e-40	hypothetical protein VT2-Sap9a [Enterobacteria phage VT2-Sakai]	Myoviridae
8	238	YP_009329619.1	90.476	4.77e-06	Kelch repeat and BTB domain-containing protein A55 [BeAn 58058 virus]	Poxviridae
9	230	YP_009337198.1	69.565	1.21e-16	hypothetical protein 2 [Changjiang tombus-like virus 8]	Riboviria
10	229	YP_009288106.1	25	2.2	terminase [Mycobacterium phage Tonenili]	Myoviridae

Tabla 5. Resultados de BLASTx de la muestra de aire del vagón del día 12 de noviembre a partir de los *contigs* obtenidos.

Contig	Longitud	Número de acceso	% Identidad	E-value	Nombre	Familia
1	915	YP_009301715.1	88.197	0.0	photosystem II protein D1 [Cyanophage S-RIM32]	Myoviridae
2	292	YP_009336668.1	36.585	0.92	hypothetical protein 2 [Wenling picorna-like virus 7]	Riboviria
3	287	NP_040519.1	100.00	0.11	encapsidation protein 52K [Human mastadenovirus C]	Adenoviridae
4	263	YP_009480767.1	36.364	3.2	F-box domain containing protein [Pandoravirus macleodensis]	No clasificado
5	240	YP_009252368.1	29.825	0.43	replication associated protein [Faeces associated gemycircularvirus 22]	Genomoviridae
6	236	YP_009624869.1	65.385	7.5e-23	Beta-lactamase SHV-2 [Escherichia phage RCS47]	Myoviridae
7	229	YP_009165706.1	36.364	4.7	hypothetical protein [Perigonia lusca single nucleopolyhedrovirus]	Baculoviridae
8	208	YP_009337198.1	55.882	1.6e-17	hypothetical protein 2 [Changjiang tombus-like virus 8]	Riboviria
9	172	YP_003284237.1	40	4.1	ORF3 protein [Sugarcane bacilliform Guadeloupe D virus]	Caulimoviridae
10	167	YP_009118491.1	39.024	0.51	immediate early protein 2 [Condylorrhiza vestigialis MNPV]	Baculoviridae

Tabla 6. Resultados de BLASTx de la muestra de aire de la sala de espera del hospital del día 12 de noviembre a partir de los *contigs* obtenidos.

Contig	Longitud	Número de acceso	% Identidad	E-value	Nombre	Familia
1	995	YP_009301715.1	87.009	0.0	photosystem II protein D1 [Cyanophage S-RIM32]	Myoviridae
2	324	YP_009337198.1	56.410	3.09e-21	hypothetical protein 2 [Changjiang tombus-like virus 8]	Riboviria
3	294	YP_009118491.1	39.024	0.57	immediate early protein 2 [Condylorrhiza vestigialis MNPV]	Baculoviridae
4	255	NP_040519.1	100.000	0.050	encapsidation protein 52K [Human mastadenovirus C]	Adenoviridae
5	231	YP_007517762.1	30.000	1.1	hypothetical protein [Pelagibacter phage HTVC010P]	Myoviridae
6	219	YP_009008326.1	87.500	2.07e-39	photosystem II D2 protein [Synechococcus phage ACG-2014h]	Myoviridae
7	218	YP_009292464.1	40.000	2.7	HNH endonuclease [Gordonia phage Eyre]	Myoviridae
8	217	YP_009624869.1	66.667	9.47e-28	Beta-lactamase SHV-2 [Escherichia phage RCS47]	Myoviridae
9	216	YP_009481323.1	48.148	0.59	F-box domain containing protein [Pandoravirus macleodensis]	No clasificado
10	211	YP_009217185.1	41.935	0.59	hypothetical protein [Pseudomonas phage PhiPA3]	Myoviridae

Tabla 7. Resultados de BLASTx de la muestra de aire de taquilla del día 26 de noviembre a partir de los *contigs* obtenidos.

Contig	Longitud	Número de acceso	% Identidad	E-value	Nombre	Familia
1	975	YP_009301715.1	87.926	0.0	photosystem II protein D1 [Cyanophage S-RIM32]	Myoviridae
2	263	YP_009336791.1	84.058	1.12e-32	hypothetical protein 3 [Hubei tombus-like virus 8]	Riboviria
3	260	YP_009617199.1	83.871	1.06e-08	lactose operon repressor [Enterobacteria phage DE3]	Myoviridae
4	239	YP_007517762.1	30.000	1.2	hypothetical protein [Pelagibacter phage HTVC010P]	Myoviridae
5	224	YP_009551924.1	42.424	0.25	putative P1a protein [Apple-associated luteovirus]	Riboviria
6	171	YP_009118491.1	39.024	0.32	immediate early protein 2 [Condylorrhiza vestigialis MNPV]	Baculoviridae
7	170	YP_005087338.1	43.478	6.5	predicted protein [Cyanophage P-SSP2]	Myoviridae
8	168	YP_008770036.1	36.364	1.8	tail fiber [Bacillus phage CampHawk]	Myoviridae
9	157	YP_009195056.1	41.463	0.43	RNA ligase [Citrobacter phage Margaery]	Myoviridae
10	153	YP_009165018.1	45.833	0.38	hypothetical protein ml_52 [Mollivirus sibericum]	No clasificado

Tabla 8. Resultados de BLASTx de la muestra de aire del vagón del día 26 de noviembre a partir de los *contigs* obtenidos.

Contig	Longitud	Número de acceso	% Identidad	E-value	Nombre	Familia
1	609	YP_009301715.1	88.596	2.40e-153	photosystem II protein D1 [Cyanophage S-RIM32]	Myoviridae
2	328	YP_009118491.1	39.024	2.3	immediate early protein 2 [Condylorrhiza vestigialis MNPV]	Baculoviridae
3	301	YP_655224.1	30.667	0.16	gp28 [Mycobacterium phage PBI1]	Myoviridae
4	279	YP_008318989.1	34.146	6.4	hypothetical protein pdul_cds_319 [Pandoravirus dulcis]	No clasificado
5	279	YP_007877859.1	85.870	1.57e-54	photosystem II D1 protein [Cyanophage Syn30]	Myoviridae
6	276	YP_009292464.1	42.857	0.72	HNH endonuclease [Gordonia phage Eyre]	Myoviridae
7	244	YP_005087436.1	87.500	3.52e-33	photosystem II reaction center protein PsbA/D1 [Cyanophage 9515-10a]	Myoviridae
8	229	YP_009337198.1	69.565	1.13e-16	hypothetical protein 2 [Changjiang tombus-like virus 8]	Riboviria
9	153	YP_002922522.1	62.500	1.2	RnlB RNA ligase 2 [Enterobacteria phage JS10]	Myoviridae
10	147	YP_002332339.1	51.724	1.3	hypothetical protein PPMP38_gp27 [Pseudomonas virus MP38]	Myoviridae

Tabla 9. Resultados de BLASTx de la muestra de aire de la sala de espera del hospital del día 26 de noviembre a partir de los *contigs* obtenidos.

Como se puede observar en las Tablas 1-6, los virus de las familias *Baculoviridae* y *Myoviridae* son comunes a todas las muestras mientras que los que pertenecen a la familia *Adenoviridae* solo están presentes en las muestras de aire de vagón y sala de espera del hospital del día 12 de noviembre y de aire de taquilla del día 26 de noviembre, además, se resaltan en colores aquellos *contigs* que BLASTx identifico compartidos en las diferentes muestras.

10.4 Análisis de variantes virales

Con el objetivo de determinar si existen variantes polimórficas en los virus circulantes en los distintos sitios de muestreo de ambos ambientes, se realizó un análisis de variantes de un solo nucleótido (SNVs) presentes en los virus que fueron identificados en todos los sitios de muestreo. Para este análisis se consideró que las regiones conseguidas en el ensamble estuvieran compartidas en todas las muestras y que tuviera una profundidad de secuenciación de por lo menos 50x.

En la Figura 16 se muestra un ejemplo de una región del adenovirus humano tipo C que fue identificada en todos los sitios de muestreo. Se visualiza con la herramienta Tablet, la cual nos permite delimitar e identificar claramente la presencia de sitios polimórficos (cuadros rojos).

Tablet mostrando *reads* y su alineamiento en una secuencia de referencia

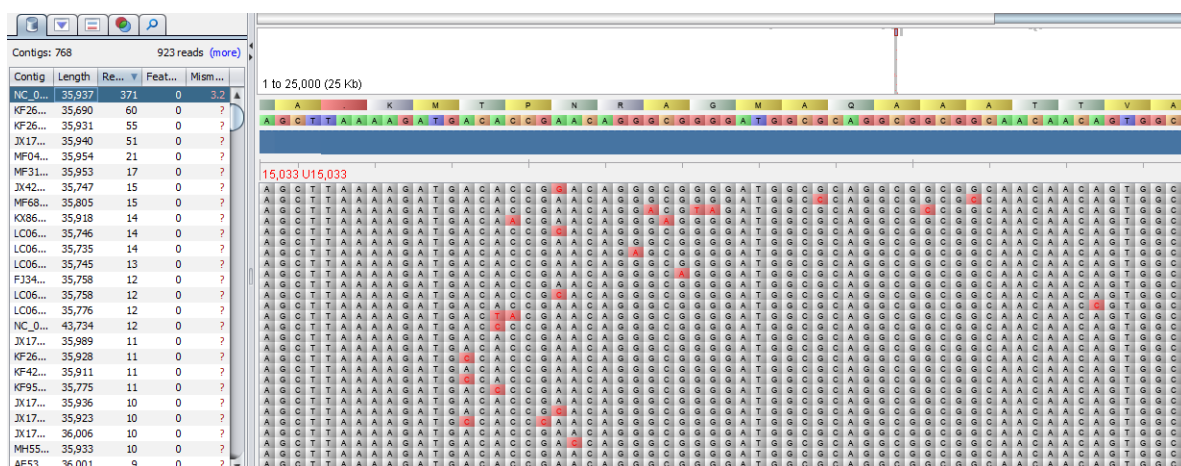


Figura 16. Imagen de un archivo de secuenciación visualizado en Tablet y su correspondencia con un genoma de referencia. Se muestra una secuencia de referencia

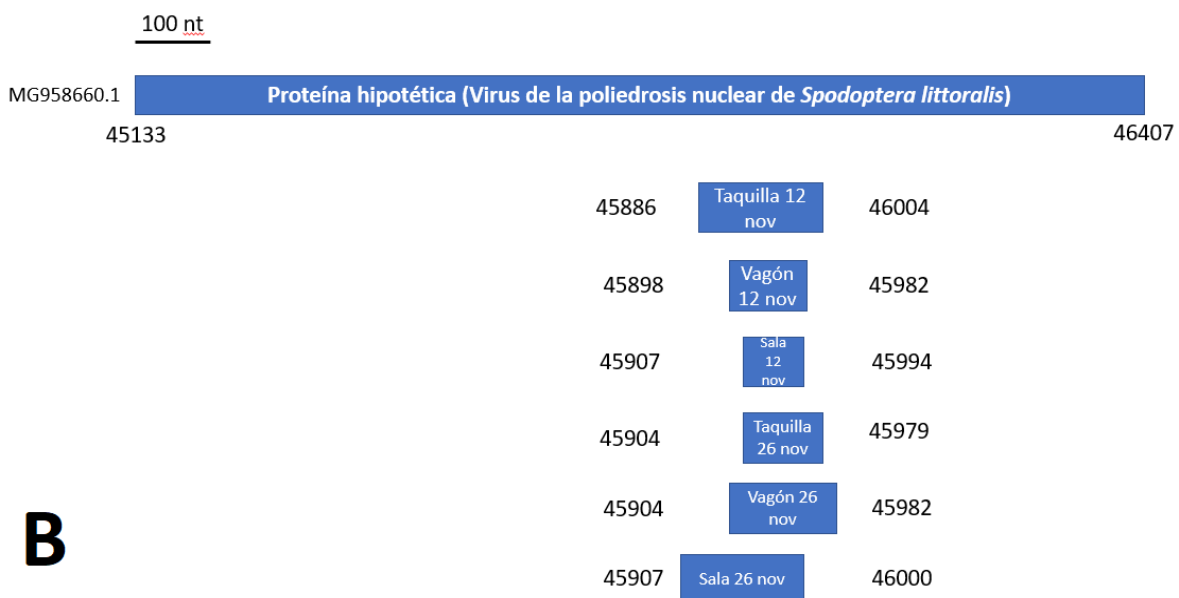
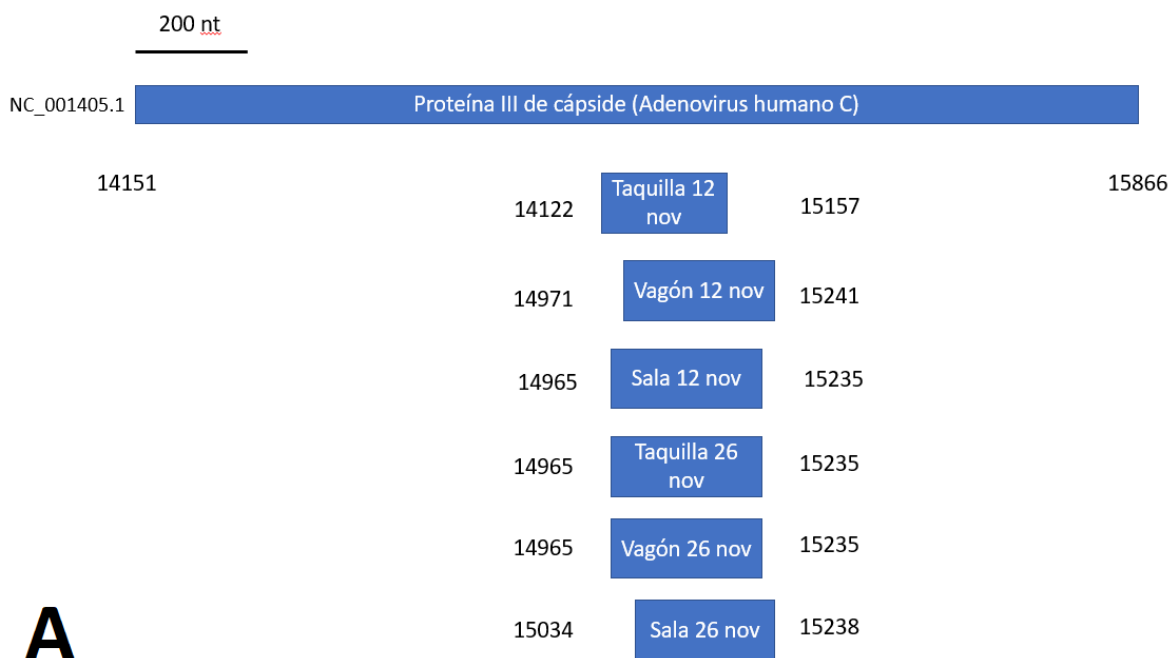
y el número de *reads* que alinean a esta, los números de acceso, la posición en el genoma, así como las variantes presentes.

A partir de las secuencias obtenidas se lograron ensamblar algunos *contigs* que estaban presentes en todas las muestras y que pertenecen a tres especies virales. En primer lugar, se obtuvo un *contig* que corresponde al adenovirus humano del tipo C, cuya longitud de la secuencia compartida en las seis muestras es de aproximadamente 200 nucleótidos y corresponde a una proteína estructural de cápside III denominada “penton base” o base pentona (Figura X) que junto con las fibras de adherencia vírica de cápside tiene la función de dirigir la entrada del virus a la célula. Tanto la base pentona como las fibras son inmunógenos y contienen antígenos específicos (Zubieta C., *et al.*, 2005).

En segundo lugar, se consiguió el ensamble de una región de alrededor de 150 nucleótidos que pertenece a al gen que codifica para una proteína hipotética (Figura 17) en el virus de la poliedrosis nuclear de la polilla común (*Spodoptera littoralis*). De esta proteína se desconoce su función

Finalmente se obtuvieron *contigs* de aproximadamente 170 nucleótidos pertenecen a una región de una proteína hipotética de un bacteriófago de *Staphylococcus sp* (Figura X).

Regiones compartidas en las tres especies virales en todas las muestras



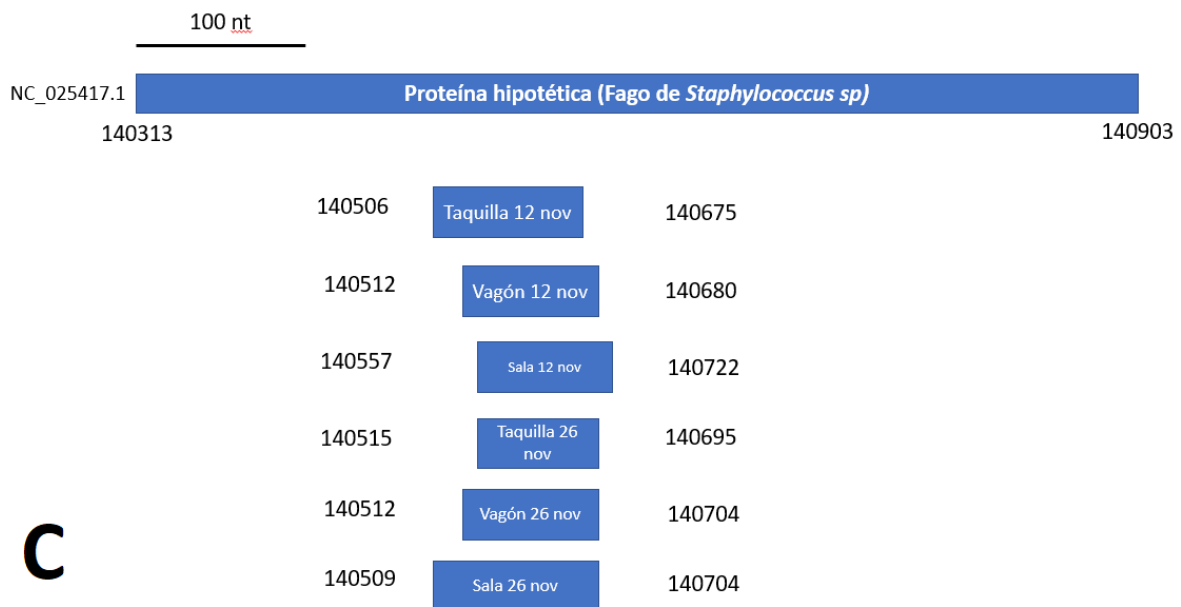


Figura 17. Esquemas de las regiones que se lograron ensamblar de los tres virus presentes en todas las muestras de ambos ambientes. A) Proteína de cápside III de adenovirus humano C, B) Proteína hipotética del virus de poliedrosis nuclear de *Spodoptera littoralis*, C) Proteína hipotética de un fago de *Staphylococcus sp*. Se indica la región de inicio y de termino con respecto al tamaño del gen, así como la muestra a la que corresponden.

10.5 Determinación de SNVs

Una vez obtenidos los *contigs*, se delimitó una región que abarcará la misma posición en todos los virus para todas las muestras presentes, esto con la finalidad de analizar las variantes presentes.

La herramienta VirVarSeq nos dio como archivos de salida unas tablas donde detalla, entre otras cosas, la posición en la región del genoma, el codón presente en la secuencia de referencia, el codón que resulta del cambio de nucleótido en la región donde hay reclutamiento de reads, el aminoácido de referencia y el aminoácido alternativo, así como la frecuencia con que están presentes en un sitio. Se consideraron como SNVs aquellos en los que las variantes estuvieran en una frecuencia mayor que 0.01, lo que corresponde a un porcentaje mayor del 10%. Para el virus de poliedrosis nuclear de

Spodoptera littoralis se encontraron 127, 145, 102, 112, 105 y 93 variantes de un solo nucleótido en las muestras de aire de taquilla, aire de vagón y aire de la sala de espera de los días 12 y de aire de taquilla, aire de vagón y aire de la sala de espera del 26 de noviembre, respectivamente. En la Tabla 10 se muestra un ejemplo de esos datos.

POS EN GENOMA	CODON_REF	CODON_ALT	AA_REF	AA_ALT	S/NS	FRECUENCIA
4989	GGA	CGA	G	R	NS	0.016393443
4989	GGA	TGA	G	.	NS	0.016393443
4990	TCA	GCA	S	A	NS	0.040983607
4990	TCA	ACA	S	T	NS	0.016393443
4990	TCA	TAA	S	.	NS	0.016393443
4990	TCA	TCC	S	S	S	0.016393443
4991	CCT	GGC	P	G	NS	0.016393443
4992	ACG	AAG	T	K	NS	0.040983607
4992	ACG	ACT	T	T	S	0.016393443
4993	ATG	ATT	M	I	NS	0.024590164
4993	ATG	ACC	M	T	NS	0.016393443
4995	TGG	TCT	W	S	NS	0.016393443
4995	TGG	TTC	W	F	NS	0.016393443
4996	AGG	TGG	R	W	NS	0.024590164
4996	AGG	TGC	R	C	NS	0.016393443
4997	GTG	GGG	V	G	NS	0.06557377
4997	GTG	AAA	V	K	NS	0.032786885
4997	GTG	GTT	V	V	S	0.024590164
4998	GTA	GGA	V	G	NS	0.016393443
4999	ACA	CAT	T	H	NS	0.016393443
5000	TTC	CTC	F	L	NS	0.016393443
5000	TTC	TCG	F	S	NS	0.016393443

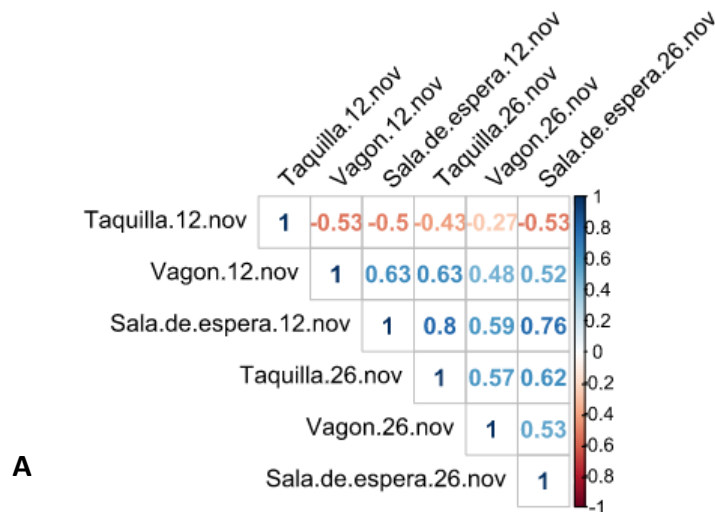
Tabla 10. Variantes del virus de poliedrosis nuclear correspondiente a la muestra de aire de taquilla del 12 de noviembre. En la tabla se muestran las posiciones 4989 a la 5000, donde hay diferentes sustituciones en cada posición.

Describiendo la tabla mostrada arriba, se incluye el codón de referencia, el alternativo con la sustitución, el aminoácido de referencia y el aminoácido que resulta de la sustitución, así mismo si es sinónima o no y la frecuencia con que ocurre a través de la profundidad de la secuenciación.

10.6 Comparación entre muestras

Del análisis de dichas variantes, se realizó una prueba de correlación de Pearson para demostrar si existe correlación entre las variantes que están presentes en las tres especies virales en todas las muestras. Se diseñó un *script* en lenguaje de programación R, donde se contrasta si los virus circulantes presentan variantes similares en los diferentes ambientes. En la Figura 18 se muestran las matrices de correlación obtenidas. En todos los casos, el valor de p es <0.05 .

Correlación de variantes presentes en los tres virus circulantes en todas las muestras



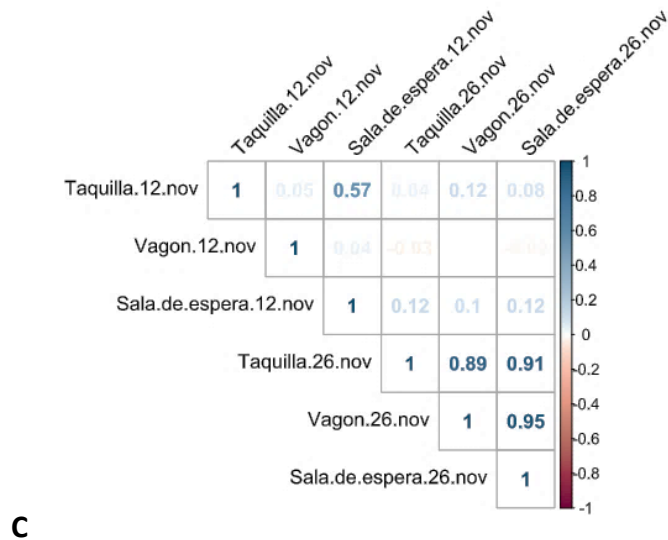
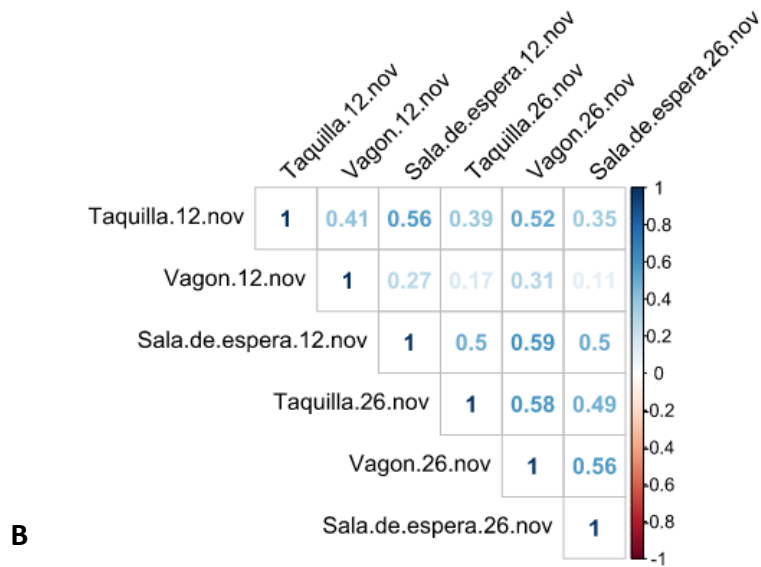


Figura 18. Matrices de correlación de Pearson. El color azul indica si hay relación entre los virus circulantes en los diferentes sitios de muestreo en los días 12 y 26 de noviembre, mientras que el color rojo indica que no hay relación, así mismo se muestra el coeficiente de correlación de lado derecho. **A)** Adenovirus humano **C**, **B)** virus de poliedrosis nuclear de *Spodoptera littoralis* y **C)** fago de *Staphylococcus sp.*

Se puede observar que para el caso de adenovirus humano **C**, las variantes que están presentes en la muestra de aire de la taquilla del día 12 de noviembre no tiene nada que ver con las variantes encontradas en las demás muestras, sin embargo las variantes que

posee el virus circulante de la sala de espera del hospital del día 12 de noviembre tienen una alta correlación con las encontradas en la muestra de taquilla del día 26 de noviembre y de la misma sala de espera del día 26 de noviembre.

Respecto al virus de poliedrosis nuclear de *Spodoptera littoralis*, las variantes son similares en todos los sitios ya que tienen una correlación positiva, no obstante, en proporción las variantes presentes en el aire de la taquilla del día 12 de noviembre son las mismas que las encontradas en el aire de la taquilla del día 26 de noviembre, de la sala de espera del 12 de noviembre y en el aire del vagón del día 26 de noviembre, ya que su coeficiente de correlación es mayor a 0.5.

En el caso de fago de *Staphylococcus sp* presente, el resultado indica que las variantes que se encuentran en dicho virus son similares en las muestras correspondientes al aire de la taquilla del 12 de noviembre con la del aire de la sala de espera del día 12 de noviembre, además las variantes del aire de la taquilla del día 26 de noviembre son similares que las del aire del vagón y sala de espera del día 26 de noviembre. En estos casos el coeficiente de correlación es mayor a 0.9, lo que indica que pueden ser las mismas variantes con mayor certeza.

11. Discusión

La metagenómica es una herramienta poderosa para estudiar la composición de viromas procedentes de diversos ambientes, que van desde los encontrados en ecosistemas, dentro de organismos como los seres humanos, hasta de ambientes urbanos como sistemas de transporte de pasajeros y hospitales. El estudio de virus dentro de un ambiente particular tiene diferentes aplicaciones en otros campos, por ejemplo, las implicaciones epidemiológicas que la presencia de virus patógenos podría tener; sin embargo, debido a la dificultad de aislar partículas virales, existen pocos estudios donde se caractericen las comunidades virales de ambientes urbanos. De hecho, la metagenómica ha permitido identificar asociaciones entre comunidades microbianas y enfermedades humanas, lo cual ha permitido el descubrimiento de nuevos virus, como por ejemplo el virus de SARS (síndrome respiratorio agudo severo), y el virus de la influenza aviar (Ryu, Wang-Shick 2017). Sin embargo, mientras los microbiomas asociados con el cuerpo humano y con ambientes naturales como suelo y océanos han sido extensamente investigados, solo unos cuantos trabajos han caracterizado las comunidades microbianas asociadas a ambientes meramente urbanos.

El campo de la genómica viral y bioinformática está experimentando una fuerte expansión gracias a las tecnologías de secuenciación de nueva generación, que provee una secuenciación rápida y menos costosa y el subsecuente ensamble en un gran número de genomas virales. Además, permite el análisis de la diversidad genética viral intra-hospedera y dinámica de “cuasispecies” la cual es relevante para preguntas biológicas en la transmisión viral, resistencia a vacunas y salto de hospedero. La secuenciación de nueva generación permite también la rápida identificación virus conocidos o potencialmente nuevos.

La calidad del aire en sitios públicos es un problema importante de salud pública. Consecuentemente con el incremento en el conocimiento de problemas de salud asociados a la contaminación en el aire, el interés en el estudio de los organismos presentes en el aire, o bioaerosoles ha incrementado también. Debido a su tamaño

microscópico, los bioaerosoles pueden permanecer en el aire y ser transportados por largos periodos de tiempo y viajar grandes distancias. Estos pueden llegar fácilmente al tracto respiratorio y dependiendo de su composición, una fracción de los organismos patógenos como bacterias o virus pueden ser transportados y ocasionar enfermedades infecciosas o reacciones alérgicas.

En el presente estudio, debido a la difícil naturaleza de la muestra, es decir, la relativamente poca cantidad de biomasa presente en el aire, de la cuál la fracción minoritaria corresponde a los virus, las condiciones ambientales que promueven la degradación de las partículas virales y de sus ácidos nucleicos, así como otros factores presentes en el aire como contaminantes, limpiadores, desinfectantes y la misma competencia entre microorganismos como bacterias, dificultó y limitó la obtención de secuencias de origen viral, lo cual está acorde a lo reportado en otros análisis de viomas (Roux S., *et al.*, 2011). Aunque no existen kits optimizados para el procesamiento de muestras de bioaerosoles, como ocurre con muestras clínicas o de suelo, se pudo obtener la cantidad de ácidos nucleicos suficiente para la secuenciación en la plataforma MiSeq, misma concentración que no pudo ser cuantificada de manera consistente, ya que la cuantificación realizada por el servicio de secuenciación difirió en gran medida de la realizada por nosotros en el laboratorio. Esto pudo ser debido a que los métodos usados por la UUSMB para estimar la calidad del RNA se basan en la detección de RNA ribosomal, el cual no es de importancia para este trabajo.

Como era de esperarse, al hacer el mapeo o reclutamiento de reads contra una base de datos de virus, el porcentaje de secuencias que pertenecen a virus es muy bajo, esto se debe a varios factores. El primero es que en las bases de datos aún son pocas las secuencias que pertenecen a virus, en comparación con lo que hay para bacterias, plásmidos, humano u otros organismos. Segundo, el tamaño de los genomas virales es mucho más pequeño comparado con el de otros organismos, por lo que es de esperarse que la mayoría de los reads pertenezcan a bacterias y eucariontes. Tercero, hacen falta optimizar los métodos de purificación de partículas virales de una muestra de aire, además de probar diferentes protocolos de extracción de ácidos nucleicos virales. En

este último sentido la purificación de partículas virales no es sencilla cuando se desea preservar la diversidad viral. Por ejemplo, la filtración tradicional (0.22 μm) elimina virus de gran tamaño como los Pandoravirus que han sido descritos recientemente (Philippe N., *et al.*, 2013) mientras que otras estrategias como la ultracentrifugación sólo han sido optimizadas para virus de interés específicos.

En la parte que corresponde a la diversidad viral encontrada en este estudio, si bien es cierto que son pocas las secuencias virales, fueron suficientes para otorgarnos un panorama general de lo que circula diariamente en la estación del metro, dentro de un vagón y en una sala de espera de un hospital del sector público.

Los resultados indican que la mayor parte de secuencias virales corresponden al orden *Caudovirales*, enfáticamente a la familia *Myoviridae*, lo cual concuerda con otros estudios. Dentro de esta familia se ubican, en su mayoría, virus que infectan a bacterias, conocidos como bacteriófagos. Si partimos del hecho que las bacterias son el segundo grupo taxonómico más abundante en la Tierra, no es de sorprender que también lo sean los organismos que las infectan y que les sirven como hospedero para replicar. La segunda familia viral más frecuente, es *Adenoviridae*. En este grupo taxonómico, la mayor parte de los reads pertenecen a un tipo de virus de especial atención ya que es infectivo para humanos y otros mamíferos. Son virus de DNA de cadena doble y son no envueltos. Se han descrito más de 60 serotipos agrupados en 7 especies (A-G). Típicamente causan enfermedades respiratorias como resfriados, conjuntivitis, crop, bronquitis y neumonía. En los niños generalmente causan infecciones en los tractos respiratorio e intestinal. Finalmente, el tercer grupo viral más frecuente corresponde a la familia *Baculoviridae*. Dentro de este grupo taxonómico se ubican virus que infectan invertebrados, como insectos, artrópodos, etc. Particularmente los hallados en este estudio corresponden a virus que infectan a polillas, denominados virus de la poliedrosis nuclear que infecta a la polilla común *Spodoptera littoralis*. Esto es bastante lógico debido a que, en las ciudades, existen una enorme cantidad de insectos, entre ellos las polillas, las cuales en ocasiones están en interiores, en este caso, el metro de la

Ciudad de México y un hospital. Cabe mencionar que estos virus no son patógenos para seres humanos.

Otros estudios donde describen la estructura de los viromas presentes en ambientes urbanos como hospitales y sistemas de transporte de pasajero, arrojan que los principales agentes patógenos son algunos virus de interés clínico como el virus de influenza A y B (Bllachere F. M. *et al.*, 2013), norovirus (Alsved M. *et al.*, 2019).

Una de las observaciones más interesantes del presente estudio es el hecho de que se encontró una alta correlación entre las variantes virales encontradas en los sitios de muestreo. Cabe destacar que en algunos de los virus se encontró una asociación temporal entre las variantes circulantes, es decir las variantes cambian con el día de muestreo. Sin embargo, en ningún caso se observó una fuerte asociación geográfica, lo que sugiere que los virus son intercambiados libremente entre estos ambientes, lo que puede ser importante desde el punto de vista epidemiológico.

Debido a la naturaleza de los virus, a su abundancia en la Tierra y a sus elevadas tasas de mutación, su estudio resulta todo un reto, sin embargo, los avances técnicos y metodológicos han mejorado en los últimos 50 años, permitiendo mayor precisión y caracterización de estas entidades. El estudio de viromas ha permitido mostrar un panorama general de todo lo encontrado en una muestra obtenida de un ambiente determinado, lo cual ayuda en campos como la medicina o epidemiología, ya que el objetivo es encontrar principalmente virus que son infecciosos para seres humanos cuyas rutas de transmisión no necesite de vectores.

El estudio de los metagenomas resulta de particular dificultad debido a la poca disponibilidad de secuencias en las bases de datos, principalmente en el estudio de los virus, las secuencias existentes en las bases de datos son aún muy reducidas, dejando a un lado en el análisis más del 90% del total de secuencias. Además, la dificultad para tomar las muestras de la mejor forma, es decir, hay diferencias entre utilizar dispositivos de captación de aire y tomar una muestra directamente de una superficie, en este sentido, la concentración de virus si bien es mayor, gran parte de las partículas

virales se encuentran degradadas debido a la presencia de RNAasas. Otro obstáculo es el aislamiento del material genético, ya que la cantidad, tamaño de partícula viral y su genoma son mucho más pequeños que el de otros microorganismos como bacterias, hongos y restos celulares humanos, lo que representa una competencia al momento de hacer la extracción. Finalmente, la purificación de los ácidos nucleicos virales, representan en conjunto, grandes retos en el campo de la metagenómica viral, sin embargo, los esfuerzos continúan para obtener cada vez mejores resultados.

12. Conclusiones

El estudio de un viroma por métodos tradicionales como cultivos e identificación por anticuerpos es de gran ayuda, sin embargo, deja de lado un enorme espectro ya que la mayoría de virus no pueden ser identificables mediante estos ensayos. El uso de tecnología de secuenciación masiva permite el análisis de todos los virus presentes en una muestra compleja, no obstante, también es limitado debido a que hasta ahora, la búsqueda de secuencias virales es conducida esencialmente en la base de datos GenBank o RefSeq, ambas de NCBI, de acuerdo a los criterios de clasificación de secuencias virales. Existen otras plataformas como IMG/VR (Integrated Microbial Genome/Virus) que integra alrededor de 760,000 genomas virales y fragmentos de genomas provenientes de bases de datos públicas, sin embargo, una de las limitantes para poder utilizar los datos alojados en IMG/VR es que se requiere de una capacidad de computo mayor, con la cual no fue posible contar. En este trabajo y como un estudio exploratorio, se utilizó la base de datos de virus de RefSeq, la cual excluye algunas categorías de datos que no pueden ser realmente procesados. Por otro lado, no todas las secuencias tienen clasificación taxonómica de acuerdo al Comité Internacional de Taxonomía de Virus (ICTV por sus siglas en ingles). Al respecto, el número de secuencias virales existentes en GenBank es de casi dos millones, de las cuales, solo 3,279 están registradas como genomas en RefSeq y de estas, solo 1,800 están clasificadas a nivel de género en el ICTV.

A pesar de lo difícil que es el estudio de un metagenoma, particularmente de virus, se logro la identificación de secuencias virales que pertenecen principalmente a las familias *Myoviridae*, *Baculoviridae* y *Adenoviridae*. Específicamente, son secuencias de virus que infectan a bacterias, invertebrados y mamíferos. Existe relación con los virus circulantes en ambos ambientes, es decir, algunos virus presentan las mismas variantes genéticas, como el caso del adenovirus humano del tipo C, que está presente en el aire del de la sala de espera del hospital del día 12 y 26 de noviembre, así como en el aire de la taquilla del día 12 y 26 de noviembre. Con respecto a los virus de la poliedrosis

nuclear de *Spodoptera littoralis*, se determinó que las variantes genéticas son similares en todos los sitios muestreados.

El de fago de *Staphylococcus sp* encontrado, indica que tiene variantes similares en los que corresponden al aire de la taquilla de los días 12 y 26 de noviembre, así mismo con las encontradas en el aire de la sala de espera del día 12 de noviembre. vagón y sala de espera del día 26 de noviembre.

Es claro que muchos organismos, como bacterias o virus, pueden ser fácilmente transportados a través del aire, además pueden permanecer por grandes períodos de tiempo en sitios donde las personas pasan gran parte de su tiempo, ya sea de manera transitoria u ocupacional, como sistemas de transporte de pasajeros y hospitales.

A pesar de que hace falta mejorar los métodos y herramientas para el estudio de metagenomas, fue posible obtener un panorama general de lo que circula en ambos ambientes urbanos. Cabe mencionar que se encontraron secuencias que pertenecen a agentes patógenos respiratorios, como virus respiratorios de influenza A, pero no de otros tipos como virus sincitial respiratorio o de parainfluenza humana.

Por último, es difícil caracterizar la diversidad viral presente en muestras de aire debido a la poca biomasa que se encuentra, además, la poca optimización de los métodos de extracción de material genético viral es otro obstáculo, las bases de datos de virus son todavía pequeñas, lo que deja un gran sesgo sin identificar, aunado a esto, las herramientas bioinformáticas enfocadas a virus no están del todo perfeccionadas.

13. Perspectivas

- Incluir también a los virus de DNA, extrayendo el material genético con un kit para DNA
- Realizar muestreos en superficies y comparar la diversidad viral presente en el aire y en fómites
- Utilizar otras bases de datos virales para mejorar el reclutamiento de reads con señales virales
- Probar otras herramientas para el *sorteo* de *reads* o *contigs*, así como para mejorar la construcción de ensamblajes

14. Referencias

1. Access, O. (2016) The Metagenomics and Metadesign of the Subways and Urban Biomes (MetaSUB) International Consortium inaugural meeting report. *Microbiome*. 4, 24.
2. Alavandi, S. V., & Poornima, M. (2012). Viral metagenomics: A tool for virus discovery and diversity in aquaculture. *Indian Journal of Virology*, 23(2), 88–98. <https://doi.org/10.1007/s13337-012-0075-2>
3. Allen, R., Iii, W., Brown, J., Colby, S., Overall, C. C., Lee, J., Zucker, J., Glaesemann, K. R., Jansson, C., & Janet, K. (2017). ATLAS (Automatic Tool for Local Assembly Structures) - a comprehensive infrastructure for assembly, annotation, and genomic binning of metagenomic and metatranscriptomic data. *PeerJ*, 5, 1–11. <https://doi.org/10.7287/peerj.preprints.2843v1>
4. Almand, E. A., Moore, M. D., & Jaykus, L. A. (2017). Virus-bacteria interactions: An emerging topic in human infection. *Viruses*, 9(3), 1–10. <https://doi.org/10.3390/v9030058>
5. Alves, L. D. F., Westmann, C. A., Lovate, G. L., De Siqueira, G. M. V., Borelli, T. C., & Guazzaroni, M. E. (2018). Metagenomic Approaches for Understanding New Concepts in Microbial Science. *International Journal of Genomics*, 2018. <https://doi.org/10.1155/2018/2312987>
6. Anderson, B. D., Lednicky, J. A., Torremorell, M., & Gray, G. C. (2017). The use of bioaerosol sampling for airborne virus surveillance in swine production facilities: A mini review. *Frontiers in Veterinary Science*, 4(JUL), 1–8. <https://doi.org/10.3389/fvets.2017.00121>
7. Bahcall, O. G. (2015). Metagenomics: Urban microbiome. *Nature Reviews Genetics*, 16(4), 194. <https://doi.org/10.1038/nrg3921>

8. Behzad, H., Gojobori, T., & Mineta, K. (2015). Challenges and opportunities of airborne metagenomics. *Genome Biology and Evolution*, 7(5), 1216–1226. <https://doi.org/10.1093/gbe/evv064>
9. Bibby, K. (2013). Metagenomic identification of viral pathogens. *Trends in Biotechnology*, 31(5), 275–279. <https://doi.org/10.1016/j.tibtech.2013.01.016>
10. Brister, J. R., Ako-Adjei, D., Bao, Y., & Blinkova, O. (2015). NCBI viral Genomes resource. *Nucleic Acids Research*, 43(D1), D571–D577. <https://doi.org/10.1093/nar/gku1207>
11. Carding, S. R., Davis, N., & Hoyles, L. (2017). Review article: the human intestinal virome in health and disease. *Alimentary Pharmacology and Therapeutics*, 46(9), 800–815. <https://doi.org/10.1111/apt.14280>
12. Carroll, D., Daszak, P., Wolfe, N. D., Gao, G. F., Morel, C. M., Morzaria, S., Pablos-Méndez, A., Tomori, O., & Mazet, J. A. K. (2018). The Global Virome Project. *Science*, 359(6378), 872–874. <https://doi.org/10.1126/science.aap7463>
13. Charles, T. C., Liles, M. R., & Sessitsch, A. (2017). Functional metagenomics: Tools and applications. In *Functional Metagenomics: Tools and Applications*. <https://doi.org/10.1007/978-3-319-61510-3>
14. Choi, J. Y., Zemke, J., Philo, S. E., *et al* (2018) Aerosol Sampling in a Hospital Emergency Room Setting: A Complementary Surveillance Method for the Detection of Respiratory Viruses. *Front. Public Heal.* 6, 174
15. Coleman, K. K., Nguyen, T. T., Yadana, S., Hansen-Estruch, C., Lindsley, W. G., & Gray, G. C. (2018). Bioaerosol Sampling for Respiratory Viruses in Singapore’s Mass Rapid Transit Network. *Scientific Reports*, 8(1), 1–7. <https://doi.org/10.1038/s41598-018-35896-1>
16. Dávila-Ramos, S., Castelán-Sánchez, H. G., Martínez-ávila, L., Sánchez-Carbente, M. D. R., Peralta, R., Hernández-Mendoza, A., Dobson, A. D. W., Gonzalez, R. A.,

- Pastor, N., & Batista-García, R. A. (2019). A review on viral metagenomics in extreme environments. *Frontiers in Microbiology*, 10(OCT), 1–19. <https://doi.org/10.3389/fmicb.2019.02403>
17. Dowd, K. A., Ko, S. Y., Morabito, K. M., Yang, E. S., Pelc, R. S., DeMaso, C. R., Castilho, L. R., Abbink, P., Boyd, M., Nityanandam, R., Gordon, D. N., Gallagher, J. R., Chen, X., Todd, J. P., Tsybovsky, Y., Harris, A., Huang, Y. J. S., Higgs, S., Vanlandingham, D. L., ... Graham, B. S. (2016). Rapid development of a DNA vaccine for Zika virus. *Science*, 354(6309), 237–240. <https://doi.org/10.1126/science.aai9137>
 18. Duffy, S., Shackelton, L. A., & Holmes, E. C. (2008). Rates of evolutionary change in viruses: Patterns and determinants. *Nature Reviews Genetics*, 9(4), 267–276. <https://doi.org/10.1038/nrg2323>
 19. Fancello, L., Raoult, D., & Desnues, C. (2012). Computational tools for viral metagenomics and their application in clinical research. *Virology*, 434(2), 162–174. <https://doi.org/10.1016/j.virol.2012.09.025>
 20. Fernandes-Matano, L., Monroy-Muñoz, I. E., Angeles-Martínez, J., et al (2017) Prevalence of non-influenza respiratory viruses in acute respiratory infection cases in Mexico. *PLoS One*. 10.1371/journal.pone.0176298
 21. Forouzan, E., Shariati, P., Mousavi Maleki, M. S., Karkhane, A. A., & Yakhchali, B. (2018). Practical evaluation of 11 de novo assemblers in metagenome assembly. *Journal of Microbiological Methods*, 151(February), 99–105. <https://doi.org/10.1016/j.mimet.2018.06.007>
 22. Harris, Z. N., Dhungel, E., Mosior, M., & Ahn, T. H. (2019). Massive metagenomic data analysis using abundance-based machine learning. *Biology Direct*, 14(1), 12. <https://doi.org/10.1186/s13062-019-0242-0>

23. Hernández, A. M., Vargas-Robles, D., Alcaraz, L. D., & Peimbert, M. (2019). Station and train surface microbiomes of Mexico City's metro (subway/underground). *BioRxiv*, 735027. <https://doi.org/10.1101/735027>
24. Hsu, T., Joice, R., Vallarino, J., Abu-Ali, G., Hartmann, E. M., Shafquat, A., DuLong, C., Baranowski, C., Gevers, D., Green, J. L., Morgan, X. C., Spengler, J. D., & Huttenhower, C. (2016). Urban Transit System Microbial Communities Differ by Surface Type and Interaction with Humans and the Environment. *MSystems*, 1(3), 1–18. <https://doi.org/10.1128/msystems.00018-16>
25. Huang, T. (2018). *Computational Systems Biology: Methods and Protocols*. <https://doi.org/10.1007/9781493977178>
26. Ilyinskii, P. O., Roy, C. J., O'Neil, C. P., Browning, E. A., Pittet, L. A., Altreuter, D. H., Alexis, F., Tonti, E., Shi, J., Basto, P. A., Iannaccone, M., Radovic-Moreno, A. F., Langer, R. S., Farokhzad, O. C., von Andrian, U. H., Johnston, L. P. M., & Kishimoto, T. K. (2014). Adjuvant-carrying synthetic vaccine particles augment the immune response to encapsulated antigen and exhibit strong local immune activation without inducing systemic cytokine release. *Vaccine*, 32(24), 2882–2895. <https://doi.org/10.1016/j.vaccine.2014.02.027>
27. Jover, L. F., Effler, T. C., Buchan, A., Wilhelm, S. W., & Weitz, J. S. (2014). The elemental composition of virus particles: Implications for marine biogeochemical cycles. *Nature Reviews Microbiology*, 12(7), 519–528. <https://doi.org/10.1038/nrmicro3289>
28. King, P., Pham, L. K., Waltz, S., Sphar, D., Yamamoto, R. T., Conrad, D., Taplitz, R., Torriani, F., & Forsyth, R. A. (2016). Longitudinal metagenomic analysis of hospital air identifies clinically relevant microbes. *PLoS ONE*, 11(8), 1–14. <https://doi.org/10.1371/journal.pone.0160124>

29. Kumar, V., Maitra, S. S., & Shukla, R. N. (2015). Environmental Metagenomics: The Data Assembly and Data Analysis Perspectives. *Journal of The Institution of Engineers (India): Series A*, 96(1), 71–83. <https://doi.org/10.1007/s40030-014-0102-y>
30. Lai, Y. Y., Li, Y., Lang, J., Tong, X., Zhang, L., Fang, J., Xing, J., Cai, M., Xu, H., Deng, Y., Xiao, F., & Tian, G. (2015). Metagenomic human respiratory air in a hospital environment. *PLoS ONE*, 10(10), 1–15. <https://doi.org/10.1371/journal.pone.0139044>
31. Lam, T. T.-Y., Zhu, H., Guan, Y., and Holmes, E. C. (2016) Genomic Analysis of the Emergence, Evolution, and Spread of Human Respiratory RNA Viruses. *Annu. Rev. Genomics Hum. Genet.* **17**, 193–218
32. Leung, M. H. Y., Wilkins, D., Li, E. K. T., Kong, F. K. F., & Lee, P. K. H. (2014). Indoor-air microbiome in an urban subway network: Diversity and dynamics. *Applied and Environmental Microbiology*, 80(21), 6760–6770. <https://doi.org/10.1128/AEM.02244-14>
33. Mubareka, S., Groulx, N., Savory, E., Cutts, T., Theriault, S., Scott, J. A., Roy, C. J., Turgeon, N., Bryce, E., Astrakianakis, G., Kirychuk, S., Girard, M., Kobinger, G., Zhang, C., & Duchaine, C. (2019). Bioaerosols and transmission, a diverse and growing community of practice. *Frontiers in Public Health*, 7(FEB). <https://doi.org/10.3389/fpubh.2019.00023>
34. Nazaroff, W. W. (2016). Indoor bioaerosol dynamics. *Indoor Air*, 26(1), 61–78. <https://doi.org/10.1111/ina.12174>
35. Nelson, K. E. (2015). Genes, Genomes and Metagenomes: Basics, Methods, Databases and Tools. In *Encyclopedia of Metagenomics* (Vol. 1). <https://doi.org/10.1007/978-1-4899-7478-5>
36. Núñez, A., Amo de Paz, G., Rastrojo, A., García, A. M., Alcamí, A., Gutiérrez-Bustillo, A. M., & Moreno, D. A. (2016). Monitoring of airborne biological particles

- in outdoor atmosphere. Part 1: Importance, variability and ratios. *International Microbiology*, 19(1), 1-13. <https://doi.org/10.2436/20.1501.01.258>
37. Oliver, J. (2013). 済無No Title No Title. In *Journal of Chemical Information and Modeling* (Vol. 53, Issue 9). <https://doi.org/10.1017/CBO9781107415324.004>
 38. Orton, R. J., Gu, Q., Hughes, J., Maabar, M., Modha, S., Vattipally, S. B., Wilkie, G. S., & Davison, A. J. (2016). Bioinformatics tools for analysing viral genomic data. *OIE Revue Scientifique et Technique*, 35(1), 271-285. <https://doi.org/10.20506/rst.35.1.2432>
 39. Oulas, A., Pavloudi, C., Polymenakou, P., Pavlopoulos, G. A., Papanikolaou, N., Kotoulas, G., Arvanitidis, C., & Iliopoulos, I. (2015). Metagenomics: Tools and insights for analyzing next-generation sequencing data derived from biodiversity studies. *Bioinformatics and Biology Insights*, 9, 75-88. <https://doi.org/10.4137/BBI.S12462>
 40. Paez-Espino, D., Eloë-Fadrosh, E. A., Pavlopoulos, G. A., Thomas, A. D., Huntemann, M., Mikhailova, N., Rubin, E., Ivanova, N. N., & Kyrpides, N. C. (2016). Uncovering Earth's virome. *Nature*, 536(7617), 425-430. <https://doi.org/10.1038/nature19094>
 41. Prussin, A. J., & Marr, L. C. (2015). Sources of airborne microorganisms in the built environment. *Microbiome*, 3, 78. <https://doi.org/10.1186/s40168-015-0144-z>
 42. Rascovan, N., Duraisamy, R., & Desnues, C. (2016). Metagenomics and the Human Virome in Asymptomatic Individuals. *Annual Review of Microbiology*, 70(1), 125-141. <https://doi.org/10.1146/annurev-micro-102215-095431>
 43. Robertson, C. E., Baumgartner, L. K., Harris, J. K., Peterson, K. L., Stevens, M. J., Frank, D. N., & Pace, N. R. (2013). Culture-independent analysis of aerosol microbiology in a metropolitan subway system. *Applied and Environmental Microbiology*, 79(11), 3485-3493. <https://doi.org/10.1128/AEM.00331-13>

44. Rosario, K., Fierer, N., Miller, S., Luongo, J., & Breitbart, M. (2018). Diversity of DNA and RNA Viruses in Indoor Air As Assessed via Metagenomic Sequencing. *Environmental Science and Technology*, 52(3), 1014–1027. <https://doi.org/10.1021/acs.est.7b04203>
45. Roux, S., Tournayre, J., Mahul, A., Debroas, D., & Enault, F. (2014). Metavir 2: New tools for viral metagenome comparison and assembled virome analysis. *BMC Bioinformatics*, 15(1). <https://doi.org/10.1186/1471-2105-15-76>
46. Sanjuan, R., Nebot, M. R., Chirico, N., Mansky, L. M., and Belshaw, R. (2010) Viral Mutation Rates. *J. Virol.* **84**, 9733–9748
47. Schoenfeld, T., Patterson, M., Richardson, P. M., Wommack, K. E., Young, M., & Mead, D. (2008). Assembly of viral metagenomes from Yellowstone hot springs. *Applied and Environmental Microbiology*, 74(13), 4164–4174. <https://doi.org/10.1128/AEM.02598-07>
48. Storey, D. G., Cote, M. J., Dimock, K., & Kang, C. Y. (1987). Nucleotide sequence of the coding and flanking regions of the human parainfluenza virus 3 hemagglutinin - Neuraminidase gene: Comparison with other paramyxoviruses. *Intervirology*, 27(2), 69–80. <https://doi.org/10.1159/000149722>
49. Taboada, B., Espinoza, M. A., Isa, P., *et al* (2014) Is there still room for novel viral pathogens in pediatric respiratory tract infections? *PLoS One*. 10.1371/journal.pone.0113570
50. Thomas, T., Gilbert, J., & Meyer, F. (2014). Metagenomics: A guide from sampling to data analysis. *The Role of Bioinformatics in Agriculture, Figure 1*, 357–383. <https://doi.org/10.1201/b16568>
51. Thumbi, D. K., Béliveau, C., Cusson, M., Lapointe, R., & Lucarotti, C. J. (2013). Comparative Genome Sequence Analysis of *Choristoneura occidentalis* Freeman

- and *C. rosaceana* Harris (Lepidoptera: Tortricidae) Alphabaculoviruses. *PLoS ONE*, 8(7). <https://doi.org/10.1371/journal.pone.0068968>
52. Tong, X., Xu, H., Zou, L., Cai, M., Xu, X., Zhao, Z., Xiao, F., & Li, Y. (2017). High diversity of airborne fungi in the hospital environment as revealed by meta-sequencing-based microbiome analysis. *Scientific Reports*, 7(January), 1–8. <https://doi.org/10.1038/srep39606>
 53. Triadó-Margarit, X., Veillette, M., Duchaine, C., Talbot, M., Amato, F., Minguillón, M. C., Martins, V., de Miguel, E., Casamayor, E. O., & Moreno, T. (2017). Bioaerosols in the Barcelona subway system. *Indoor Air*, 27(3), 564–575. <https://doi.org/10.1111/ina.12343>
 54. Tringe, S. G., Zhang, T., Liu, X., Yu, Y., Lee, W. H., Yap, J., Yao, F., Suan, S. T., Ing, S. K., Haynes, M., Rohwer, F., Wei, C. L., Tan, P., Bristow, J., Rubin, E. M., & Ruan, Y. (2008). The airborne metagenome in an indoor urban environment. *PLoS ONE*, 3(4). <https://doi.org/10.1371/journal.pone.0001862>
 55. Trujillo, M. E., Beltrán, R., Garcia, G., Chávez, F., Blanco, M. Á., Sarmiento, R. E., & Bolaños, D. (2017). Virus y el sector pecuario. *La Virología En México. Situación Actual, Retos y Oportunidades.*, 216. http://www.ibt.unam.mx/computo/pdfs/La_Virologia_en_Mexico_Situacion_Actual_4.pdf
 56. Viegas, C., Viegas, S., Gomes, A. Q., Täubel, M., & Sabino, R. (2017). Exposure to microbiological agents in indoor and occupational environments. *Exposure to Microbiological Agents in Indoor and Occupational Environments*, 1–415. <https://doi.org/10.1007/978-3-319-61688-9>
 57. Wadapurkar, R. M., & Vyas, R. (2018). Computational analysis of next generation sequencing data and its applications in clinical oncology. *Informatics in Medicine Unlocked*, 11(May), 75–82. <https://doi.org/10.1016/j.imu.2018.05.003>

58. Wong-Chew, R. M., Espinoza, M. A., Taboada, B., *et al* (2015) Prevalence of respiratory virus in symptomatic children in private physician office settings in five communities of the state of Veracruz, Mexico. *BMC Res. Notes*. **8**, 1–8
59. Yooseph, S., Andrews-Pfannkoch, C., Tenney, A., McQuaid, J., Williamson, S., Thiagarajan, M., Bami, D., Zeigler-Allen, L., Hoffman, J., Goll, J. B., Fadrosh, D., Glass, J., Adams, M. D., Friedman, R., & Venter, J. C. (2013). A metagenomic framework for the study of airborne microbial communities. *PLoS ONE*, *8*(12). <https://doi.org/10.1371/journal.pone.0081862>
60. Zárate, S., Hernández-Perez, F., Taboada, B., Martínez, N. E., Alcaráz-Estrada, S. L., del Moral, O., & Yocupicio-Monroy, M. (2019). Complete genome of DENV2 isolated from mosquitoes in Mexico. *Infection, Genetics and Evolution*, *71*, 98–107. <https://doi.org/10.1016/j.meegid.2019.03.018>
61. Zheng, T., Li, J., Ni, Y., Kang, K., Misiakou, M. A., Imamovic, L., Chow, B. K. C., Rode, A. A., Bytzer, P., Sommer, M., & Panagiotou, G. (2019). Mining, analyzing, and integrating viral signals from metagenomic data. *Microbiome*, *7*(1), 1–15. <https://doi.org/10.1186/s40168-019-0657-y>

15. Trabajos presentados en congresos



La Sociedad Mexicana de Virología y la Asociación Mexicana de Bioseguridad
otorgan la presente

CONSTANCIA

a MARTÍNEZ AMADOR ULICES FABRICIO, Zárate Selene, Peimbert Torres Mariana, Alcaraz Estrada Sofia Lizeth.

por su valiosa exposición del trabajo libre en **PRESENTACIÓN ORAL intitulado**

Viroma de muestras de aire de la estación del metro y el hospital 20 de Noviembre en la Ciudad de México

presentado en el **XI CONGRESO NACIONAL DE VIROLOGÍA**

que tuvo lugar del 4 al 7 de septiembre de 2019 en los Espacios Magnos de la Universidad de Guanajuato.
Ciudad de Guanajuato, Guanajuato, México, a 7 de septiembre de 2019.


Dr. Ramón González García Conde
Presidente
Sociedad Mexicana de Virología AC
Folio: XICNV-CONF22




QFB Lissete Valenzuela Fabris
Presidenta
Asociación Mexicana de Bioseguridad AC