



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

POSGRADO EN CIENCIAS BIOLÓGICAS

FACULTAD DE MEDICINA

BIOMEDICINA

VIROMA RESPIRATORIO DE INDIVIDUOS MEXICANOS INFECTADOS CON

VIRUS DE INMUNODEFICIENCIA HUMANA

TESIS

QUE PARA OPTAR POR EL GRADO DE:

MAESTRA EN CIENCIAS BIOLÓGICAS

PRESENTA:

Q.F.B. RINCÓN RUBIO ALMA CAROLINA

TUTOR PRINCIPAL DE TESIS: DR. SANTIAGO ÁVILA RÍOS

INSTITUTO NACIONAL DE ENFERMEDADES RESPIRATORIAS, INER

COMITÉ TUTOR: DR. ENRIQUE HERNÁNDEZ LEMUS

INSTITUTO NACIONAL DE MEDICINA GENÓMICA, INMEGEN

DR. JOSÉ IVÁN SÁNCHEZ BETANCOURT

FACULTAD DE MEDICINA VETERINARIA Y ZOOTECNIA, UNAM

MÉXICO, CD. MX., OCTUBRE, 2018



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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



OFICIO CPCB/869/2018

Asunto: Oficio de Jurado para Examen de Grado.

Lic. Ivonne Ramírez Wence
Directora General de Administración Escolar, UNAM
Presente

Me permito informar a usted que el Subcomité de Biología Experimental y Biomedicina del Posgrado en Ciencias Biológicas, en su sesión ordinaria del día 06 de agosto de 2018, aprobó el siguiente jurado para la presentación del examen para obtener el grado de **MAESTRA EN CIENCIAS BIOLÓGICAS** de la alumna **RINCÓN RUBIO ALMA CAROLINA** con número de cuenta **516021230**, con la tesis titulada "**VIROMA RESPIRATORIO DE INDIVIDUOS MEXICANOS INFECTADOS CON VIRUS DE INMUNODEFICIENCIA HUMANA**", realizada bajo la dirección de del **DR. SANTIAGO ÁVILA RÍOS**:

Presidente: DR. CARLOS FEDERICO ARIAS ORTIZ
Vocal: DR. ALEJANDRO MANUEL GARCÍA CARRANCÁ
Secretario: DR. JOSÉ IVÁN SÁNCHEZ BETANCOURT
Suplente: DR. CARLOS CABELLO GUTIÉRREZ
Suplente: DR. ENRIQUE HERNÁNDEZ LEMUS

Sin otro particular, me es grato enviarle un cordial saludo.

ATENTAMENTE
"POR MI RAZA HABLARÁ EL ESPÍRITU"
Cd. Universitaria, Cd, Mx., a 21 de septiembre de 2018


DR. ADOLFO GERARDO NAVARRO SIGÜENZA
COORDINADOR DEL PROGRAMA



AGRADECIMIENTOS

Agradezco al Posgrado en Ciencias Biológicas de la UNAM por la calidad de su programa, sus excelentes cátedras y, sobre todo, por permitirme ser parte de la comunidad UNAM.

Agradezco también el apoyo recibido por CONACYT con la beca otorgada, durante 24 meses, en el marco de la "Convocatoria de Becas Nacionales 2016 Primer Periodo".

Por último, agradezco enormemente al Dr. Santiago Ávila Ríos quien fungió como mi Tutor Principal y a los Doctores Enrique Hernández Lemus y José Iván Sánchez Betancourt, miembros de mi Comité Tutor. Su tiempo, guía, consejos y apoyo fueron fundamentales para la realización de este proyecto y para mi formación académica.

AGRADECIMIENTOS PERSONALES

Agradezco a mi madre y a mi padre, Gabriela y Eduardo, por su amor y apoyo especialmente en los momentos difíciles, los quiero mucho.

Agradezco infinitamente a la Doctora Celia Boukadida por su paciencia, apoyo, consejos y, sobre todo, su gran disposición para explicar; su guía y enseñanzas fueron indispensables para la realización de este proyecto. Además, su ejemplo de disciplina y constancia me ha motivado a dar siempre lo mejor de mí.

Por supuesto le doy las gracias al Doctor Santiago Ávila Ríos por su constante apoyo, confianza y su genuino interés en mi formación profesional; sus comentarios me han hecho crecer académica y personalmente. Además, con su ejemplo me ha demostrado que se puede ser un gran investigador y una excelente persona.

Agradezco especial y efusivamente a Ana Paola Carranco por su fundamental apoyo en la instalación del programa bioinformático y por sus consejos en el análisis de datos; sus aportaciones fueron primordiales para este trabajo.

Asimismo, agradezco a todos los pacientes que voluntariamente aceptaron participar en el protocolo al que pertenece este proyecto, sin su colaboración nada de esto hubiera sido posible. Al mismo tiempo, expreso mi gratitud a las médicas Sandra Astorga y Melissa Lomelí, y a Fernando Díaz por su gran contribución al invitar a los pacientes a participar; además le doy las gracias a la Doctora Isabel León por la revisión de los expedientes clínicos.

Agradezco también a las Doctoras Olivia Sánchez, Dina Martínez y Teresa Aguirre, médicas adscritas al Instituto Nacional de Enfermedades Respiratorias (INER), por la realización de los procedimientos médicos para la obtención de las muestras respiratorias. Su participación y la de los médicos residentes a su cargo, fue imprescindible para que este trabajo pudiera llevarse a cabo.

Además, le doy las gracias a todo el personal de recepción y toma de muestra del Departamento de Investigación en Enfermedades Infecciosas (CIENI), INER, por la obtención de muestras sanguíneas de los individuos que conformaron nuestra cohorte, su colaboración fue fundamental. También reconozco y agradezco la participación del personal del Laboratorio de Diagnóstico Viroológico (LDV) del CIENI, INER; su contribución fue de gran ayuda.

Igualmente, expreso mi agradecimiento a Ángel León, Monserrat Chávez y a las Doctoras Perla del Río y Sandra Pinto por su gran colaboración en el procesamiento y almacenamiento de las muestras utilizadas en este proyecto; el trabajo en conjunto fue primordial para que esto pudiera llevarse a cabo.

Por supuesto estoy agradecida con el grupo de Investigación Básica del CIENI, INER, conformado por Claudia García, Daniela Tapia, César Cortés, Margarita Matías, Maribel Soto, Marissa Pérez, Dulce López, Humberto Valenzuela, Olivia Briceño, Daniela Garrido, Elvira Piten, Loana Álvarez y Silvia del Arenal, quienes de alguna u otra forma contribuyeron a la realización de este proyecto.

Asimismo, agradezco al personal administrativo, de informática y de mantenimiento del CIENI, INER por el apoyo brindado durante todo este tiempo.

Finalmente agradezco al Doctor Gustavo Reyes Terán, director del CIENI, INER, por darme la oportunidad de integrarme a su equipo de investigación, además de brindarme su constante apoyo y confianza.

ÍNDICE

LISTA DE FIGURAS

LISTA DE TABLAS

RESUMEN	1
ABSTRACT	2
1. INTRODUCCIÓN	3
2. OBJETIVOS	8
2.1 Objetivo general.....	8
2.2 Objetivos específicos	8
3. HIPÓTESIS	8
4. ANTECEDENTES	9
5. MÉTODOLÓGÍA	11
5.1 Selección de la cohorte	11
5.2 Obtención de las muestras.....	11
5.3 Purificación y concentración de las partículas virales.....	12
5.4 Extracción de ácidos nucleicos	12
5.5 Transcripción inversa y amplificación SISPA (<i>Sequence-Independent, Single-Primer Amplification</i>)	13
5.6 Purificación y cuantificación de los productos de amplificación.....	13
5.7 Secuenciación de las muestras con tecnología de siguiente generación.....	14
5.8 Análisis bioinformático.....	15
5.9 Determinación de la carga viral de VIH y CMV por qPCR.....	16
6. RESULTADOS	18
6.1 Características generales de la cohorte.....	18
6.2 Validación de la metodología de caracterización del viroma respiratorio.....	20
6.2.1 Identificación de virus humanos en las muestras respiratorias.....	20
6.2.2 Confirmación de la detección de VIH y CMV	22
6.3 Caracterización del viroma respiratorio	24
6.3.1 Clasificación de secuencias totales	24
6.3.2 Clasificación de secuencias virales de acuerdo al hospedero.....	25
6.3.3 Identificación de secuencias de virus humanos mediante NGS	27
6.3.4 Prevalencia de virus humanos.....	30
6.3.5 Abundancia de virus humanos	32
6.3.5.1 Efecto de la inmunodeficiencia causada por la infección de VIH.....	32

6.3.5.2	Efecto de variables clínicas y epidemiológicas.....	34
6.3.6	Coberturas de los genomas virales.....	35
6.3.7	Análisis filogenético de las secuencias de VPH	36
7.	DISCUSIÓN	41
8.	CONCLUSIONES	47
9.	LITERATURA CITADA	49

LISTA DE FIGURAS

FIGURA 1. Correlación del número de copias de CMV (A) y de VIH (B) identificadas por PCR en tiempo real y el número de secuencias normalizadas encontradas por NGS.	23
FIGURA 2. Clasificación de secuencias totales en lavados orales (A) y LBA (B).....	25
FIGURA 3. Clasificación de secuencias virales	26
FIGURA 4. Heatmaps de secuencias virales identificadas en muestras de lavado oral (A) y LBA (B).....	28
FIGURA 5. Comparación del número de secuencias virales obtenidas en los lavados orales y los LBAs.	29
FIGURA 6. Prevalencia de virus humanos, citomegalovirus, papillomavirus, anellovirus y herpesvirus en lavados orales y LBAs.....	31
FIGURA 7. Comparación del número de secuencias virales obtenidas entre grupos en los lavados orales (A) y los LBAs (B)	32
FIGURA 8. Comparación del número de secuencias obtenidas entre grupos en los lavados orales y los LBAs.....	34
FIGURA 9. Mapas de cobertura del genoma viral.....	35
FIGURA 10. Árbol filogenético de papillomavirus de los géneros beta (A) y gamma (B)	38

LISTA DE TABLAS

TABLA 1. Características de la cohorte.....	19
TABLA 2 Características de los virus humanos identificados en las muestras respiratorias de los individuos de nuestra cohorte.....	21

RESUMEN

INTRODUCCIÓN. La infección por VIH se caracteriza por la pérdida progresiva de linfocitos T CD4+. Como resultado, las personas en etapas avanzadas de esta infección son altamente susceptibles a una gran variedad de enfermedades; entre las cuales está la neumonía, que aún representa una causa significativa de morbilidad y mortalidad en personas con VIH, especialmente en países como México que se caracterizan por un diagnóstico tardío de esta infección. Sin embargo, el papel de los virus respiratorios en la adquisición y desarrollo de neumonía en personas con inmunodeficiencia asociada a VIH aún no ha sido dilucidada. Por ello, el objetivo de este estudio fue caracterizar las poblaciones virales (viroma) presentes en el tracto respiratorio de personas VIH positivas y con sospecha de neumonía.

MÉTODOS. Se obtuvieron lavados orales (LO) y broncoalveolares (LBA) de 31 adultos con neumonía, 5 de ellos eran VIH negativos y los 26 restantes estaban infectados por VIH. El grupo de personas VIH negativas fue dividido en 4 subgrupos basados en el grado de inmunodeficiencia (conteo de células T CD4+ ≥ 200 ; $50 \leq CD4 < 200$; $20 \leq CD4 < 50$ y $CD4 < 20$). Se utilizó secuenciación de siguiente generación para la caracterización del viroma de las muestras orales y broncoalveolares. Las partículas virales fueron concentradas, se extrajeron los ácidos nucleicos y se realizó la retrotranscripción del RNA. El cDNA resultante junto con el DNA inicial, fueron amplificados de manera aleatoria y secuenciados usando la plataforma de Illumina con el equipo NextSeq. Las secuencias virales fueron identificadas usando el programa computacional SURPI.

RESULTADOS. Se obtuvieron 11.2 [IQR, 10.2 - 12.1] y 9.0 [IQR, 7.4 - 10.1] millones de secuencias de alta calidad en los LOs y LBAs, respectivamente. Se identificaron secuencias de virus humanos en todos los LOs y en 87% de los LBAs. Se detectaron secuencias de VIH en 15.4% y 57.7% de los LOs y LBAs de los individuos VIH positivos, respectivamente. Los virus más prevalentes en los LOs fueron los anellovirus [83.9%], virus de Epstein-Barr y papillomavirus [80.6%, en ambos casos], virus de herpes humano-7 (VHH-7) [67.7%], VHH-6 [54.8%] y citomegalovirus (CMV) [48.4%]. Por otro lado, los virus más frecuentemente detectados en los LBAs fueron anellovirus [64.5%], CMV [41.9%] y papillomavirus [29%]. Además, se identificaron patógenos respiratorios en varias muestras: rinovirus fue encontrado en 4 LOs y 2 LBAs [12.9% y 6.5%, respectivamente], adenovirus fue detectado en 1 LO [3.2%], metapneumovirus en 1 LO y 1 LBA, mientras que enterovirus, virus de la influenza y coronavirus fueron encontrados en 1 LBA. También se detectó pegivirus humano en 4 LOs y 4 LBAs [12.9%], VHH-1 en 3 LOs [9.7%], VHH-8 y poliomavirus humano en 2 OL [6.5%] y finalmente, se detectaron secuencias de parvovirus humano en 1 LO y 1 LBA [3.2%]. De manera importante, en varias muestras se encontró una gran cantidad de secuencias virales y se obtuvieron genomas casi completos de algunos virus (papillomavirus, virus de la influenza, parvovirus y poliomavirus humano), lo cual refleja altas cargas virales de estos patógenos. Además, se identificó un nuevo tipo de *Gammapapillomavirus*. Finalmente, y de manera interesante, las muestras de los pacientes VIH positivos severamente inmunocomprometidos mostraron una tendencia al incremento en el número y diversidad de secuencias virales.

CONCLUSIÓN. La presencia de una gran variedad de virus patógenos y comensales en el tracto respiratorio de personas infectadas con VIH, sugiere que dicha infección podría estar relacionada con una expansión del viroma respiratorio. Es necesario analizar más muestras para poder describir el efecto de la inmunodeficiencia asociada a este virus en la composición del viroma respiratorio y su contribución al desarrollo de enfermedades respiratorias.

ABSTRACT

BACKGROUND. HIV infection is marked by the progressive depletion of CD4+ T lymphocytes. As a result, people with advanced HIV disease are highly susceptible to a wide variety of infections. Among these, pneumonia is a significant cause of HIV-associated morbidity and mortality, especially in countries like Mexico that are characterized by late HIV diagnosis. However, the role of respiratory viruses in the acquisition and development of pneumonia in immunocompromised HIV-infected persons remains to be elucidated. The objective of this study was to characterize the viral populations (virome) present in the respiratory tract of HIV-infected subjects with pneumonia.

METHODS. Oral lavages (OL) and bronchoalveolar lavages (BAL) were obtained from 31 adults with pneumonia, including 5 HIV-negative and 26 HIV-infected individuals. The HIV-positive group was divided into 4 subgroups based on the degree of immunodeficiency (CD4 T cell count ≥ 200 ; $50 \leq \text{CD4} < 200$; $20 \leq \text{CD4} < 50$ and $\text{CD4} < 20$). Metagenomic next-generation sequencing was used to characterize the virome in OL and BAL samples. Viral particles were concentrated, nucleic acids were extracted, RNA was reverse-transcribed and cDNA and DNA were randomly amplified and sequenced using the NextSeq Illumina platform. Viral sequences were identified using the SURPI computational pipeline.

RESULTS. We obtained 11.2 [IQR, 10.2 - 12.1] and 9.0 [IQR, 7.4 - 10.1] million high-quality sequences per OL and BAL, respectively. Human virus sequences were detected in all OL and in 87% of BAL. HIV sequences were detected in 15.4% and 57.7% of OL and BAL from HIV-infected individuals, respectively. The most prevalent viruses in OL were anelloviruses [83.9%], Epstein-Barr virus and papillomaviruses [80.6%], human herpes virus 7 (HHV-7) [67.7%], HHV-6 [54.8%] and cytomegalovirus (CMV) [48.4%]. Viruses frequently detected in BAL included anelloviruses [64.5%], CMV [41.9%] and papillomaviruses [29%]. Respiratory pathogens were identified in several samples: rhinovirus was found in 4 OL and 2 BAL [12.9% and 6.5%, respectively], adenovirus was detected in 1 OL [3.2%], metapneumovirus were identified in 1 OL and 1 BAL, while enterovirus, influenza virus and coronavirus were found in 1 BAL. In addition, human pegivirus was detected in 4 OL and 4 BAL [12.9%], HHV-1 was identified in 3 OL [9.7%], HHV-8 and human polyomavirus were found in 2 OL [6.5%] and human parvovirus were detected in only 1 OL and BAL [3.2%]. Importantly, some samples contained large numbers of viral sequences and several near full-length (papillomavirus, influenza virus, human parvovirus and human polyomavirus) genomes were obtained, reflecting high viral loads. In addition, a new *Gammapapillomavirus* type was identified. Interestingly, respiratory samples from highly immunocompromised HIV-infected individuals showed a trend towards an increase in the number and diversity of viral sequences.

CONCLUSION. The presence of a variety of human viruses including pathogens and commensals in the respiratory tract of HIV-infected persons, suggests that HIV infection may be associated with an expansion of the respiratory virome. Additional samples will be analyzed in order to further characterize the effect of HIV-associated immunodeficiency on the composition of the respiratory virome and its potential contribution to the development of respiratory diseases.

1. INTRODUCCIÓN

El término "microbioma" fue propuesto por J. Lederberg y A. McCray en 2001 para hacer referencia a la comunidad de microorganismos comensales, simbióticos y patógenos distribuidos en alguna parte del cuerpo (Lederberg & McCray, 2001). Sin embargo, hoy en día este término es empleado en un sentido más amplio y puede ser definido como la suma total de todos los microorganismos y sus genomas asociados con algún órgano o tejido. Mientras que, para designar únicamente a la comunidad microbiana constituida por bacterias, arqueas, hongos, virus y protozoos, se utiliza el vocablo "microbiota" (Shukla *et al.*, 2017). A su vez, se usan las palabras "bacterioma", "micoma" y "viroma" para referirse, respectivamente, a los componentes bacterianos, fúngicos y víricos del microbioma. Para este trabajo resulta de especial interés el viroma humano, el cual está compuesto no solamente por virus capaces de infectar células del hospedero, sino también por elementos génicos virales integrados en nuestros cromosomas y virus que infectan otros organismos del microbioma (Virgin, 2014).

El estudio del viroma humano es relevante por varias razones, una de las primordiales y la que ha motivado la mayor parte de las investigaciones realizadas hasta ahora, es la existencia de varios patógenos virales que tienen un impacto importante en la salud (Virgin, 2014). Sin embargo, en los últimos años se ha propuesto que el viroma tiene muchas más implicaciones en el proceso salud-enfermedad del hospedero debido, entre otras cosas, a que algunos de ellos tienen la capacidad de infectar de manera crónica y sistémica a los individuos. Se estima que una persona sana alberga más de diez infecciones virales persistentes, no obstante, este número puede ser mucho mayor, ya que el viroma humano no ha sido completamente caracterizado (Virgin, 2014). Con base en lo anterior puede decirse que, al igual que las bacterias, los virus logran comportarse como comensales o patógenos oportunistas (Cadwell, 2015). Ejemplo de ello son algunos miembros de la familia de los herpesvirus, poliomavirus, papillomavirus, adenovirus, circovirus y anellovirus (Virgin, 2014; Cadwell, 2015).

Además de lo anterior, se ha propuesto que los virus tienen efectos inmunomoduladores que generan consecuencias para el hospedero más allá de la defensa antiviral. Esta inmunomodulación puede provocar un incremento (inmunopotenciación) o un decremento (inmunosupresión) de la inmunidad o modificar la naturaleza de las reacciones inmunológicas; además puede ser local o sistémica, cambiar a través del tiempo y diferir entre los distintos compartimientos anatómicos (Cadwell, 2015). Asimismo, los virus pueden alterar las características fisiológicas del hospedero tanto por la interacción directa con éste, como por su posible influencia en otros componentes del microbioma (Zárate *et al.*, 2017). Así, algunos microorganismos que forman parte del microbioma, pueden facilitar o disminuir la infectividad de ciertos virus de manera local e incluso en sitios distales. De manera similar algunos elementos del viroma, sobre todo los bacteriófagos, tienen un papel importante en la regulación de la

estructura de las poblaciones bacterianas (Zárate *et al.*, 2017). Por todo lo anterior es posible que el viroma altere la susceptibilidad a enfermedades complejas o infecciones secundarias en los individuos (Cadwell, 2015). Asimismo, una vez que se haya determinado la composición del viroma en personas sanas, la alteración de éste podría ser utilizada como un marcador de inflamación (Cadwell, 2015) o inmunosupresión.

Por otra parte, la detección, caracterización y clasificación de los virus puede contribuir a prevenir y dar tratamiento a las infecciones causadas por ellos, lo cual ha motivado el desarrollo de técnicas y métodos para su identificación (Datta *et al.*, 2015). Algunas de las técnicas más utilizadas durante el siglo pasado fueron el cultivo celular, la microscopía electrónica y el uso de sueros de referencia de pacientes previamente infectados o vacunados (Mokili *et al.*, 2012). Sin embargo, estas técnicas son costosas, tardadas, poco sensibles o específicas y, en el caso de la propagación viral, se requiere encontrar las condiciones que se parezcan al ecosistema que comparten el virus y su hospedero (Mokili *et al.*, 2012). Posteriormente surgieron técnicas como la Reacción en Cadena de la Polimerasa (PCR), la PCR cuantitativa (qPCR) y los microarreglos, las cuales son muchos más rápidas que las técnicas clásicas y han permitido el descubrimiento de nuevos genotipos virales; no obstante, sólo permiten la identificación de un virus o grupo de ellos de manera simultánea (Datta *et al.*, 2015; Mokili *et al.*, 2012). Con la finalidad de superar esta última limitación surgió la "metagenómica", que consiste en el estudio de la totalidad del material genético presente en una muestra (Datta *et al.*, 2015). Esto, aunado al desarrollo de técnicas de amplificación de ácidos nucleicos de manera aleatoria y el uso de secuenciación de siguiente generación (NGS, por sus siglas en inglés), ha revolucionado el campo de la microbiología. De manera particular, estas técnicas novedosas han beneficiado enormemente a la virología ya que permiten detectar cualquier virus que esté presente en una muestra determinada, independientemente de si se puede propagar o no o de si es conocido o nuevo (Mokili *et al.*, 2012). Más aún, la detección de los virus puede realizarse a partir de cantidades relativamente pequeñas de material genético y, en algunos casos, puede secuenciarse casi la totalidad del genoma viral (Datta *et al.*, 2015), lo que ha conducido a un aumento significativo en el descubrimiento de virus nuevos en años recientes.

Gracias al desarrollo de estas nuevas tecnologías y a un creciente interés en el estudio del microbioma humano se ha podido descubrir que varios sitios anatómicos que antes eran considerados estériles, en realidad poseen comunidades microbianas que pueden verse alteradas durante el curso de una enfermedad. Este es el caso del microbioma del tracto respiratorio inferior, cuyo estudio ha sido de gran interés en los últimos años. Los resultados de estas investigaciones han mostrado asociaciones entre la constitución del microbioma respiratorio y el desarrollo o exacerbación de enfermedades como el asma, la fibrosis quística, la enfermedad pulmonar obstructiva crónica y la fibrosis pulmonar idiopática; sin embargo, la

mayor parte de los estudios se han enfocado en el componente bacteriano del microbioma (Dickson *et al.*, 2016).

Con base en lo anterior, resulta fácil pensar que algunas condiciones clínicas que modifiquen la constitución anatómica o fisiológica del hospedero también pueden ocasionar cambios en el ambiente que deriven en una alteración del microbioma (Dickson *et al.*, 2016). Un claro ejemplo de ello, es la infección por el Virus de Inmunodeficiencia Humana (VIH), cuyos efectos deletéreos en el sistema inmunológico del hospedero tiene repercusiones sistémicas.

El VIH infecta y destruye principalmente a los linfocitos T CD4+, los cuales tienen un papel fundamental en la respuesta inmunológica. Debido a ello, cuando no hay intervención clínica, el avance de la infección por VIH lleva a una inmunosupresión severa que condiciona al individuo a una alta susceptibilidad a infecciones causadas por microorganismos patógenos y oportunistas. De manera general y en ausencia de atención farmacológica, la infección por este virus puede dividirse en tres etapas que se describen a continuación. La primera es la fase aguda, que incluye el evento de trasmisión, la infección de células de la mucosa y la diseminación a tejidos linfoides (Deeks *et al.*, 2015). Esta etapa se caracteriza por una intensa replicación viral, una alta viremia (de hasta 10^7 copias de RNA/mL de sangre) y una rápida y pronunciada depleción de células T CD4+. Se considera que esta fase concluye cuando, debido a la aparición de linfocitos T CD8+ con actividad citotóxica frente al VIH, se logra un control parcial de la infección con la consiguiente disminución de la carga viral y una moderada recuperación de células T CD4+ en sangre periférica; esto sucede hasta alcanzar un punto de equilibrio (*set point*) donde, tanto la pérdida de células como el incremento de la carga viral, son paulatinas (Moir *et al.*, 2011). Una vez ocurrido esto, se considera que se ha llegado a la etapa crónica de la infección, la cual puede durar varios años y se caracteriza por la constante replicación viral y la activación inmune sostenida. A pesar de que en esta fase aumenta la susceptibilidad del individuo a contraer otras infecciones, la mayoría de ellos permanecen asintomáticos (Moir *et al.*, 2011). Finalmente, cuando el conteo de células T CD4+ es menor a 200 células por μL de sangre, se considera que la persona ha llegado a la etapa de SIDA (Síndrome de Inmunodeficiencia Adquirida), la cual está aunada a la aparición de infecciones oportunistas (Selik *et al.*, 2014).

El uso del tratamiento antirretroviral (TAR) y antibióticos profilácticos ha disminuido considerablemente la aparición de infecciones oportunistas; sin embargo, las neumonías aún representan una causa significativa de morbilidad y mortalidad en personas con VIH (Skalski & Limper, 2016), en particular en países de ingresos medios/bajos como México, en el que aún prevalece la detección tardía de la infección por VIH (Crabtree *et al.*, 2011). Además, se ha observado que la inmunosupresión asociada a este patógeno puede provocar cambios en la composición del viroma entérico y plasmático (Li *et al.*, 2013; Monaco *et al.*, 2016). En 2013, Li y colaboradores compararon el viroma plasmático de personas VIH positivas con bajos (<20

células/ μL) y altos (>700 células/ μL) conteos de células T CD4+. Las personas altamente inmunocomprometidas (<20 células T CD4+/ μL) tuvieron un mayor número de secuencias de anellovirus (Li *et al.*, 2013). Esta familia viral se caracteriza por tener una prevalencia extremadamente alta ($\sim 90\%$), una distribución cosmopolita, una gran diversidad genómica y un aparente pan-tropismo a nivel del hospedero; además, aunque ha tratado de asociarse con múltiples enfermedades, no existe evidencia de su relación con ninguna de ellas (Spandole *et al.*, 2015). Asimismo, los resultados de otra investigación mostraron que las heces de personas en etapa de SIDA poseían una cantidad de secuencias de adenovirus significativamente mayor a la encontrada en personas VIH negativas o VIH positivas con más de 200 células CD4+/ μL . Los autores proponen que esta alteración podría contribuir a la enteropatía asociada al SIDA y a la progresión de la enfermedad (Monaco *et al.*, 2016).

El efecto del funcionamiento deficiente del sistema inmunológico en la composición del viroma humano también ha sido estudiado en personas con terapia de inmunosupresión después de haber recibido algún trasplante de órganos (De Vlaminck *et al.*, 2013; Young *et al.*, 2015). En estas investigaciones se encontró una correlación inversa entre la carga viral de anellovirus y el grado de inmunodeficiencia de los pacientes. Estos hallazgos fueron descritos tanto en el tracto respiratorio (Young *et al.*, 2015) como en el plasma (De Vlaminck *et al.*, 2013) de personas trasplantadas. Además, en este último caso, al ser un estudio longitudinal, se pudo observar una expansión general de los virus cuando el efecto del tratamiento inmunosupresor era mayor, seguida de un decremento de la carga viral total al disminuir las dosis de dicha terapia (De Vlaminck *et al.*, 2013).

Asimismo, en estudios posteriores, se han buscado asociaciones entre el viroma de ciertos sitios anatómicos con la probabilidad de desarrollar complicaciones subsecuentes a algún trasplante de órganos (Legoff *et al.*, 2017) (Abbas *et al.*, 2017). Legoff y colaboradores, por ejemplo, analizaron de manera longitudinal los cambios ocurridos en el viroma entérico de pacientes antes y después de recibir un trasplante de células hematopoyéticas pluripotenciales; esto con la finalidad de encontrar posibles correlaciones entre las modificaciones en la composición y diversidad de las poblaciones virales intestinales, con el desarrollo de la enfermedad de injerto contra huésped manifestada en este mismo órgano. Los resultados de este estudio mostraron un incremento en la proporción de secuencias virales identificadas en las heces de los individuos trasplantados, durante las seis semanas siguientes al procedimiento. Más aún, en el caso particular de las personas que presentaron esta enfermedad, la expansión del viroma fue especialmente importante en virus de DNA persistentes (anellovirus, herpesvirus, papillomavirus y poliomavirus). Finalmente, y de manera interesante, se encontró asociación entre la infección por picobirnavirus con el subsecuente desarrollo de esta complicación post-trasplante (Legoff *et al.*, 2017), lo cual podría tener implicaciones clínicas importantes.

Aunado a lo anterior, otro grupo de investigadores estudiaron la posible relación del viroma pulmonar con la disfunción primaria del injerto (DPI), tras el trasplante de dichos órganos (Abbas *et al.*, 2017). Para ello, se caracterizó la totalidad de las comunidades virales presentes en los pulmones de los donadores, justo antes del procedimiento, y de los individuos receptores, una vez terminada la re-perfusión de estos mismos órganos. Además, con la finalidad de identificar probables diferencias en el viroma peri-operatorio de personas que presentaron DPI con aquellas que no tuvieron complicaciones, se analizaron de manera conjunta muestras de lavado broncoalveolar de ambos tipos de pacientes y, al mismo tiempo, se obtuvieron muestras de voluntarios sanos. Uno de los hallazgos más relevantes fue que los niveles de DNA de torquetenovirus, pertenecientes a la familia de los anellovirus, fueron 100 veces mayores en los donadores en comparación a lo encontrado en individuos sanos. Lo anterior puede deberse a la inmunodeficiencia inducida en estas personas, previo al trasplante. Otro dato interesante fue que, a pesar de que las cantidades absolutas de este tipo de anellovirus no fueron diferentes por estatus de DPI, los receptores con esta complicación presentaron menores incrementos de torquetenovirus después del procediendo quirúrgico. Una posible explicación de ello es que la DPI está asociada con activación inmunológica y con una disminución de la viabilidad del tejido lo que, en conjunto, podría limitar la replicación viral (Abbas *et al.*, 2017).

2. OBJETIVOS

2.1 Objetivo general

- Determinar el efecto de la inmunodeficiencia asociada a la infección por VIH sobre la composición del viroma del tracto respiratorio de individuos mexicanos.

2.2 Objetivos específicos

- Estandarizar un método que permita identificar los virus presentes en el tracto respiratorio de individuos con síntomas de infección respiratoria, a partir de muestras de lavados broncoalveolares (LBA) y lavados orales, mediante el uso de secuenciación de siguiente generación.
- Caracterizar el viroma respiratorio de individuos mexicanos con infección respiratoria, con y sin VIH.

3. HIPÓTESIS

La infección por VIH provocará cambios en el viroma respiratorio, asociados a la inmunodeficiencia característica de la enfermedad, con respecto a individuos no infectados por este virus. El patrón de composición del viroma respiratorio de los individuos infectados por VIH cambiará de acuerdo con la etapa de la infección y, en particular, al grado de inmunodeficiencia en la que se encuentre el paciente.

4. ANTECEDENTES

Hasta el momento, es poco lo que se sabe acerca de las comunidades virales del tracto respiratorio; no obstante, estudios recientes han sentado las bases para su caracterización. En el caso particular del viroma oral se ha encontrado que, a pesar de que existen diferencias individuales en su composición (Pride *et al.*, 2012; Abeles *et al.*, 2014), los bacteriófagos de las familias *Siphoviridae* y *Myoviridae*, además de fagos de *Streptococcus* y los herpes virus, están ampliamente distribuidos en adultos jóvenes sanos (Pérez-Brocá & Moya, 2018). Aunado a ello, se ha propuesto que, cuando no existen alteraciones ambientales drásticas, la composición del viroma de la cavidad oral es estable a lo largo del tiempo (Pride *et al.*, 2012). Sin embargo, la mayoría de estas investigaciones se han realizado en cohortes pequeñas y se han enfocado en virus de DNA.

En cuanto al viroma de las partes distales del tracto respiratorio, la información es aún más limitada y los pocos estudios publicados hasta ahora se han realizado en personas donantes o receptoras de un trasplante pulmonar (Young *et al.*, 2015; Abbas *et al.*, 2017). Estas investigaciones encontraron que la carga viral de anellovirus en lavados broncoalveolares (LBAs) es mayor en las personas bajo tratamiento inmunosupresor, tanto donantes (Abbas *et al.*, 2017) como receptoras (Young *et al.*, 2015), en comparación con individuos sanos; por esta razón, se ha propuesto que la carga viral de estos virus podría ser un marcador de inmunosupresión. Además, en estos estudios también se identificaron algunas secuencias de virus de herpes humano 4 y 7 (VHH-4 y VHH-7) y virus del papiloma humano (VPH), aunque en una frecuencia y proporción muy baja (Young *et al.*, 2015).

Por otra parte, el viroma respiratorio de personas con síntomas de neumonía o infección del tracto respiratorio superior, ha sido examinado en muestras de lavado nasal o nasofaríngeo de pacientes pediátricos. En este estudio se identificó la presencia de virus conocidos por ser patógenos respiratorios, en ambos grupos de pacientes. Estos virus incluyeron al virus sincicial respiratorio, rinovirus, coronavirus, metapneumovirus, virus de la influenza y parainfluenza. Asimismo, se encontró una baja prevalencia de anellovirus, rotavirus, astrovirus, VPH y herpesvirus, así como virus de animales y plantas (Taboada *et al.*, 2017). De manera interesante, este estudio es uno de los pocos realizados hasta ahora en los que se identificaron, de manera conjunta, virus de DNA y de RNA.

Como puede verse, el componente viral de microbioma respiratorio ha sido poco estudiado y caracterizado; no obstante, estudios del viroma de otros sitios anatómicos han identificado modificaciones en su composición, asociadas a la inmunosupresión inducida por el VIH (Li *et al.*, 2013; Monaco *et al.*, 2016). Además, se ha observado una relación entre la infección por picobirnavirus con el desarrollo de complicaciones post-trasplante (Legoff *et al.*, 2017). En conjunto, estos hallazgos sugieren que ciertos elementos del viroma podrían ser utilizados como

marcadores de inmunosupresión o ayudar a predecir el desarrollo o exacerbación de enfermedades.

Con base en los datos que se tienen hasta el momento, resulta evidente que el estudio del viroma respiratorio de personas infectadas por VIH podría contribuir al entendimiento, prevención, diagnóstico y tratamiento de enfermedades respiratorias asociadas a la infección por este patógeno. Sin embargo, la constitución del microbioma pulmonar, particularmente del viroma, aún no ha sido descrita en personas que viven con VIH. Por ello, en el presente estudio se propone caracterizar el viroma respiratorio de individuos mexicanos VIH positivos con diferentes grados de inmunodeficiencia y con síntomas de neumonía, a partir de muestras de lavado broncoalveolar (LBA) y lavado oral, y compararlo con el de personas VIH negativas con síntomas de infección respiratoria. Asimismo, dado a que es más probable encontrar nuevos virus humanos en muestras provenientes de personas inmunocomprometidas, al estudiar el viroma de pacientes en etapa de SIDA habrá mayor probabilidad de identificar nuevas cepas virales patógenas u oportunistas o con tropismo previamente no identificado (Delwart, 2013).

5. METODOLOGÍA

5.1 Selección de la cohorte

En este estudio se incluyeron 31 pacientes, de entre 18 y 50 años de edad, que tenían síntomas de infección respiratoria, requerían una muestra de LBA para diagnóstico clínico y no habían recibido tratamiento inmunosupresor o quimioterapia. De los 31 pacientes incluidos 26 eran individuos infectados por VIH que fueron divididos en cuatro subgrupos, de acuerdo con el grado de inmunodeficiencia, de la siguiente manera: $CD4 \geq 200$, $50 \leq CD4 < 200$, $20 \leq CD4 \leq 50$ y $CD4 < 20$. Para el grupo de personas menos inmunocomprometidas se reunieron 5 muestras, mientras que los 3 grupos restantes estuvieron conformados por 7 individuos cada uno. Además, se incluyeron 5 muestras de personas VIH negativas.

Para cada individuo participante en el estudio se tomaron muestras de lavado oral y LBA, las cuales fueron recolectadas y almacenadas en el Departamento de Investigación en Enfermedades Infecciosas (CIENI) del Instituto Nacional de Enfermedades Respiratorias (INER), una vez que los participantes dieron consentimiento escrito para el uso de las muestras en investigación básica. Este proyecto fue aprobado por el Comité de Ética en Investigación del INER con el código B08-16.

5.2 Obtención de las muestras

Para este estudio se utilizaron muestras de lavados orales y LBAs. Las muestras orales fueron obtenidas mediante un enjuague vigoroso, por toda la cavidad oral, con 15 mL de solución de cloruro de sodio (NaCl) al 0.9%, durante un minuto. Posteriormente, se realizó un enjuague similar con un colutorio de clorhexidina al 0.12%, para disminuir la posible contaminación de las muestras del tracto respiratorio inferior con microorganismos de la boca.

Por otra parte, los LBAs consistieron en la instilación, a través del tracto respiratorio y por medio del broncofibroscopio, de 3 jeringas con 60 mL de solución de NaCl al 0.9% (volumen total de 180 mL); cada instilación fue seguida inmediatamente de la aspiración del líquido, con una tasa de recuperación de entre 40 y 70% del volumen instilado. Cada vez que dicha solución es instilada y recuperada ésta penetra cada vez más en las vías aéreas distales, por lo que la primera jeringa permite un mejor muestreo a nivel bronquial y las siguientes a nivel alveolar. Dado que nuestro interés era tener un muestreo que incluyera las partes más distales del tracto respiratorio inferior, se utilizaron las jeringas 2 y 3 para la caracterización del viroma. Para ello, el contenido de ambas jeringas fue mezclado antes del almacenamiento y procesamiento de las muestras. Los LBAs utilizados en este estudio no poseían restos visibles de sangre.

Las muestras de lavado oral y de LBA fueron centrifugadas a 2,500 rpm (revoluciones por minuto) durante 10 minutos y los sobrenadantes fueron criopreservados a -80°C en tubos falcon de 15 mL hasta su utilización.

Es importante mencionar que además de estas muestras, se incluyeron dos tipos de controles negativos. Uno de ellos consistió en solución salina que fue instilada y recuperada por el broncofibroscopio, y otro fue este mismo tipo de solución colocada en un recipiente estéril de polipropileno (usados también para la recuperación de los lavados orales). Ambos fueron sometidos a todo el proceso descrito en los siguientes apartados. Esto se hizo con el propósito de identificar posibles secuencias contaminantes.

5.3 Purificación y concentración de las partículas virales

La purificación de las partículas virales tuvo la finalidad de eliminar las células humanas, bacterianas y fúngicas, además de disminuir la cantidad de ácidos nucleicos libres y concentrar los viriones. Para ello, se descongelaron a temperatura ambiente, alícuotas de 8 a 13 mL de lavados orales, LBAs y de ambos tipos de controles negativos, y se centrifugaron durante 10 min a 4,000 g y 4°C para sedimentar las células presentes en las muestras. Posteriormente, fueron filtradas en membranas de acetato de celulosa con poros de 0.45 µm (Tube Top Vacuum Filter System, Corning) para retener fragmentos celulares que no hubieran podido ser eliminados con la centrifugación. Luego de ello, con la finalidad de concentrar las partículas virales y de reducir el volumen a 200 µL, las muestras fueron pasadas por filtros que tenían una membrana de 100 kDa (Amicon Ultra-15 Centrifugal Unit, Millipore). De esta manera, los viriones y las macromoléculas fueron retenidas en el filtro. Finalmente, las muestras concentradas fueron tratadas con DNasa y RNasa (Turbo DNase, Ambion y RNase I, Invitrogen) para degradar los ácidos nucleicos libres.

5.4 Extracción de ácidos nucleicos

La extracción de RNA y DNA viral a partir de lavados orales y LBAs se realizó con el estuche de reactivos QIAmp Viral RNA Mini Kit (QIAGEN, Valencia, CA), el cual se basa en el principio de la cromatografía de adsorción mediante el uso de columnas de sílica. Para ello, primero fue necesario agregar una solución de lisis que, además de lisar las partículas virales para liberar los ácidos nucleicos, inactivó las RNasas y DNasas que hubieran podido degradarlos y aseguró un pH óptimo para la unión de los ácidos nucleicos a la columna de sílica. A continuación, se hicieron dos lavados con la finalidad de eliminar contaminantes celulares y proteínas y, finalmente, el RNA y el DNA viral fueron eluidos con un buffer de baja fuerza iónica que contenía azida de sodio para inhibir el crecimiento microbiano. Es importante mencionar que la extracción se realizó, contrario a las recomendaciones del fabricante, sin utilizar RNA acarreador para evitar la interferencia de éste durante la secuenciación.

5.5 Transcripción inversa y amplificación SISPA (*Sequence-Independent, Single-Primer Amplification*)

Los lavados orales y los LBAs fueron sometidos a secuenciación tipo *shotgun*, es decir, secuenciación de todos los ácidos nucleicos extraídos después de la purificación y concentración de las partículas virales. Esta técnica se empleó debido a que permite la amplificación de todos los virus presentes en las muestras. Para ello fue necesario estandarizar una técnica conocida como SISPA (*Sequence-Independent, Single-Primer Amplification*), la cual utiliza cebadores aleatorios (*random primers*). Brevemente podemos decir que las muestras fueron sometidas a transcripción inversa para la obtención de cDNA (DNA complementario) de los virus de RNA y, posteriormente, se realizó la amplificación del cDNA y el DNA viral. Como se mencionó anteriormente, se usaron cebadores aleatorios, los cuales consisten en una secuencia fija y en una secuencia variable de nueve nucleótidos. Esta última se hibridó con las regiones del RNA que tuvieran una secuencia complementaria y sirvió como cebador para la polimerización del cDNA y la síntesis de la segunda cadena. Finalmente, para la amplificación sólo se agregó la secuencia fija que sirvió como cebador, al reconocer la secuencia conocida que fue integrada al cDNA y el DNA viral.

5.6 Purificación y cuantificación de los productos de amplificación

Después de que el cDNA y el DNA fueron amplificados, se purificaron y se seleccionaron los fragmentos mayores a 150 pares de bases (pb) mediante perlas magnéticas AMPure XP (Beckman Coulter). Este sistema está basado en la inmovilización reversible de fase sólida y consiste, brevemente, en que estas perlas con propiedades paramagnéticas se unen selectivamente a fragmentos de DNA mayores a 100 pb (aunque la eficiencia de purificación y el porcentaje de recuperación pueden optimizarse al variar la razón perlas: producto de amplificación). Posteriormente, se colocan en una barra magnética que permite que dichas perlas, junto con los productos deseados, queden retenidos mientras se realizan lavados con etanol para eliminar posibles contaminantes como cebadores, nucleótidos, sales y enzimas. Finalmente, se eluye en un buffer que facilita la recuperación del DNA de interés (Beckman Coulter, 2013).

A continuación, los productos purificados fueron cuantificados con el kit Qubit® dsDNA HS (Life Technologies), el cual consiste en marcadores fluorescentes que se unen selectivamente al DNA de doble cadena. Posteriormente, la concentración es calculada al comparar la fluorescencia emitida por cada muestra con una curva estándar que es construida usando estándares de DNA incluidos en el estuche de reactivos.

5.7 Secuenciación de las muestras con tecnología de siguiente generación

Una vez que se han amplificado los ácidos nucleicos, es necesario que cumplan con ciertas características para que puedan ser secuenciados mediante las tecnologías de siguiente generación. Este proceso es conocido como preparación de “bibliotecas genómicas”. En este caso, se usó el estuche de reactivos Nextera XT DNA Library Preparation (Illumina) para preparar las bibliotecas a partir de 1ng de DNA y cDNA viral. De manera general, la preparación de bibliotecas con esta química consiste en la fragmentación del DNA y la posterior adición de índices y adaptadores en los extremos 5' y 3' de los fragmentos resultantes; esto se realiza en un solo paso por la acción de transposomas especialmente diseñados para ello. Tanto los adaptadores como los índices son secuencias cortas de nucleótidos; los primeros sirven como cebadores durante la secuenciación y los segundos son utilizados para identificar las distintas muestras de una corrida. Para incorporar estos oligonucleótidos a los fragmentos de DNA de interés, fue necesario realizar una PCR (Reacción en Cadena de la Polimerasa) con pocos ciclos para evitar al máximo sesgos de amplificación (Nextera® XT DNA Sample Preparation Guide, 2012). Una vez hecho esto, se seleccionaron los fragmentos mayores a 300 pb mediante perlas magnéticas AMPure XP (Beckman Coulter), se cuantificaron con el kit Qubit® dsDNA HS (Life Technologies), se determinó el tamaño promedio de cada biblioteca y se validó su calidad con un bioanalizador (Agilent 2100). Este último utiliza los principios de la electroforesis en gel tradicional, con la diferencia de que se realiza en un chip; así, los fragmentos de DNA fueron separados por tamaño al ser colocados en una matriz de polímero y ser sometidas a un gradiente de voltaje (Agilent 2100 Bioanalyzer 2100, Expert User's Guide). Luego de ello, con base en la concentración (en ng/ μ L) y en el tamaño medio de las bibliotecas, se calculó la concentración molar de cada una de ellas y posteriormente, el total de las bibliotecas genómicas incluidas en cada corrida fueron mezcladas de manera equimolar para obtener una concentración final de 4 nM. Finalmente, la mezcla de bibliotecas fue secuenciada mediante el equipo NextSeq (Illumina) con ciclos de 2 x 150 y en la versión 2 del equipo de reactivos NextSeq 500/550 Mid Output.

Por otro lado, las plataformas de secuenciación de Illumina se basan en la secuenciación en fase sólida con deoxirribonucleótidos trifosfato (dNTPs) que tienen la capacidad de ser terminadores reversibles de la polimerización, por ello, se conoce como secuenciación por síntesis. Los templados de DNA son generados usando la amplificación por puente de fragmentos de este ácido nucleico unidos a una celda, lo que da como resultado múltiples copias de la misma molécula de DNA, dichos agrupamientos de moléculas idénticas son conocidos como *clusters* (Pickrell *et al.*, 2012). La amplificación por puente requiere que los adaptadores que son incorporados a los fragmentos de DNA de interés durante la preparación de las bibliotecas, sean complementarios a unas secuencias de nucleótidos que están unidas a la celda y sirven como cebadores sentido y antisentido. Debido a ello, es necesario desnaturalizar el DNA de doble

cadena para generar DNA monocatenario. Una vez ocurrido esto, el DNA se hibrida con los oligonucleótidos de la superficie de la celda que funcionan como cebadores y se genera una primera copia del DNA de interés, posteriormente las moléculas de DNA iniciales son removidas y la copia, que queda unida directamente a la celda, adquiere una estructura similar a un puente para que el extremo 3' que quedó libre se hibride con el cebador antisentido, se polimerice y finalmente se desnaturalice para que el ciclo vuelva a empezar. De esta forma se obtienen millones de copias del DNA inicial (Buermans & den Dunnen, 2014).

Posteriormente, los *clusters* de DNA generados quedan inmovilizados en la superficie sólida; una vez que ocurrió esto, se agregan los cuatro dNTPs terminadores reversibles marcados con fluorescencia, en ciclos de secuenciación. La incorporación de dichos dNTPs interrumpe la síntesis de DNA, se detecta el color de la fluorescencia, se remueve dicho marcaje y un nuevo ciclo comienza (Pickrell *et al.*, 2012). Una de sus principales ventajas es que puede generar entre 25 y 6 mil millones de secuencias en una sola corrida, dependiendo del equipo y el tipo del estuche de reactivos utilizados; en el caso de NextSeq 500/550 Mid-Output Kit v2 usado en este proyecto, se pueden obtener hasta 260 millones de secuencias (Illumina). Además, es muy fácil de usar, lo que ha favorecido que esta plataforma sea de las más utilizadas para la NGS. Sin embargo, hay que tener presente que la calidad de las secuencias decae durante los últimos ciclos y que, debido a que los distintos dNTPs son incorporados a la reacción de manera simultánea, su principal fuente de error es la sustitución de bases (Pickrell *et al.*, 2012).

5.8 Análisis bioinformático

Para la realización de estos análisis se contó con el apoyo de personas expertas en bioinformática del grupo de investigación del CIENI. El primer paso del análisis consistió en remover las secuencias duplicadas que son introducidas principalmente durante la PCR de amplificación. Para ello se utilizó la herramienta FastUniq, la cual está optimizada para identificar y remover, de forma rápida, secuencias duplicadas entre millones de pares de secuencias cortas y complementarias obtenidas por NGS (Xu *et al.*, 2012). Una vez hecho esto, se removieron los adaptadores de las secuencias únicas mediante el programa cutadapt (Martin, 2011). Los criterios para ello fueron los siguientes: se tomaron en cuenta sólo aquellas secuencias con una longitud mayor a 50 nucleótidos; un mismo adaptador podía ser identificado y removido si se encontraba hasta 8 veces en una sola secuencia; se consideró una tolerancia máxima de error del 10% y finalmente, un adaptador fue reconocido como tal si se identificaba, por lo menos, el 70% de su secuencia.

Después de estos primeros pasos, se utilizó el pipeline SURPI (*Sequence-based Ultrarapid Pathogen Identification*) que permite la identificación de patógenos presentes en muestras clínicas a partir

de datos obtenidos mediante NGS (Naccache *et al.*, 2014). De manera general, este programa realiza el análisis de la siguiente manera:

- **Pre-procesamiento de datos.** Se remueven las secuencias de baja calidad y baja complejidad y posteriormente, se eliminan las secuencias que se consideren contaminación de genoma humano.
- **Alineamiento de secuencias.** Las secuencias de buena calidad son alineadas con todas las secuencias de nucleótidos presentes en la biblioteca del NCBI (*National Center for Biotechnology Information*), lo cual permite la identificación de bacterias, hongos, parásitos y virus.
- **Ensamblaje *de novo*.** Para las secuencias que no pudieron ser identificadas se realiza un ensamblaje *de novo* y luego de ello, son alineadas con una base de datos de proteínas virales y con todas las secuencias de proteínas del NCBI. Finalmente, SURPI reporta una lista de todas las secuencias clasificadas taxonómicamente y una tabla con el número de secuencias encontradas para cada organismo identificado.

Una vez que las secuencias fueron analizadas por este programa, se realizó una revisión manual para eliminar aquellas que hubieran sido mal clasificadas o estuvieran presentes en el control negativo. Finalmente, para que los números de secuencias fueran comparables entre las muestras y grupos, se realizó una normalización considerando 10 millones de secuencias de alta calidad para cada muestra.

5.9 Determinación de la carga viral de VIH y CMV por qPCR

La determinación de la carga viral de CMV y VIH en los LBAs fue realizada por el equipo de trabajo del Laboratorio de Diagnóstico Viroológico (LDV) del CIENI. Para ello, se usaron los estuches de reactivos LightCycler® CMV Quantitative Kit y Abbott RealTime HIV-1, cuya sensibilidad es de 10 copias de CMV/200µL y < 40 copias de VIH/600µL, respectivamente.

De manera general, la qPCR consiste en la cuantificación del DNA o RNA presente en una muestra mediante la evaluación de curvas de amplificación en las que se representa la fluorescencia detectada contra el número de ciclos de PCR. Así, el número de copias de un ácido nucleico puede determinarse a partir del número de ciclos necesarios para que la señal alcance un determinado nivel de fluorescencia llamado umbral, el cual debe interceptar la curva en la fase exponencial y es determinado de manera arbitraria (Rodríguez & Rodríguez, 2006). El número de ciclos necesarios para interceptar el valor umbral se llama "ciclo umbral" o "*threshold cycle*" (Ct) y es inversamente proporcional al número de copias iniciales de DNA o RNA (Rodríguez & Rodríguez, 2006).

En el caso particular de la qPCR para la identificación de VIH utilizada en este trabajo, las muestras cuya carga viral es menor al límite de detección pueden ser reportadas como indetectables (cuando el equipo no detecta ninguna señal) o detectables con < 40 copias/600 μ L (cuando existe señal, pero no puede determinarse el número de copias [de 1 a 39]). Por ello, para facilitar los análisis, las muestras que fueron reportadas como detectables con < 40 copias/600 μ L se consideraron con 20 copias de VIH/mL.

6. RESULTADOS

6.1 Características generales de la cohorte

Se analizaron las muestras de lavado oral y LBA de 31 individuos con síntomas de infección respiratoria que requerían una muestra de LBA para diagnóstico clínico. Cinco de estos pacientes eran VIH negativos y los 26 restantes estaban infectados por VIH y tenían diferentes grados de inmunodeficiencia, por lo que fueron divididos en cuatro subgrupos de acuerdo con el conteo de células CD4+. Las muestras fueron recolectadas durante un año (22 de junio de 2016 - 16 junio de 2017). Las características clínicas más importantes se presentan en la Tabla 1.

Como puede observarse en la Tabla 1, la mayor parte de los individuos infectados por VIH son hombres jóvenes. Esto concuerda con el predominio característico de pacientes del sexo masculino y de personas jóvenes en la población infectada por VIH en nuestro país (CENSIDA, 2017). Por otro lado, cabe destacar que se encontraron diferencias significativas entre grupos en el número de células T CD4+ ($p < 0.0001$) y en la razón de células CD4/CD8 ($p = 0.0008$), lo cual es esperado debido a la selección y estratificación de los individuos incluidos en cada grupo. Asimismo, no se observaron diferencias estadísticamente relevantes en características como la edad, sexo, tratamiento con antibióticos o antirretrovirales, tabaquismo, agente etiológico de la neumonía y el número de secuencias de alta calidad obtenidas en cada grupo. Lo anterior es importante puesto que estas particularidades podrían influir en los resultados de este proyecto.

Por otro lado, la carga viral de VIH en plasma no fue significativamente diferente en los grupos de pacientes infectados por este virus, a pesar de que la mediana de este parámetro clínico fue muy distinta en cada subgrupo. Estas diferencias se deben a que en nuestra cohorte se incluyeron individuos VIH positivos con y sin tratamiento antirretroviral, lo que generó una gran diversidad en las cargas virales de VIH.

	VIH - n=5	VIH CD4 ≥ 200 n=5	VIH 50 ≤ CD4 < 200 n=7	VIH 20 ≤ CD4 < 50 n=7	VIH CD4 < 20 n=7	p-value	
Edad (años) [mediana (IQR)]	39 (29 – 43)	35 (20 – 45)	31 (22 – 34)	36 (28 – 44)	28 (23 – 35)	p = 0.415	
Sexo (hombres) [n (%)]	2 (40%)	4 (80%)	6 (85.7%)	6 (85.7%)	6 (85.7%)	p = 0.303	
Conteo de células T CD4+ (células/mm ³) [mediana (IQR)]	521 (412 – 869)	232 (222 – 448)	61 (56 – 170)	35 (27 – 40)	12 (8- 19)	p < 0.0001	
Razón CD4/CD8 [mediana (IQR)]	1.79 (1.55 – 2.8)	0.5 (0.23 – 0.56)	0.14 (0.03 – 0.22)	0.05 (0.03 – 0.22)	0.03 (0.03 – 0.13)	p = 0.0008	
Carga viral de VIH (copias/mL) [mediana (IQR)]	NA	108 (<40 – 249,195)	199,955 (84 – 933,417)	344,539 (16,239- 552,071)	34,516 (<40 -503,105)	p = 0.3834	
Tratamiento antirretroviral [n (%)]	NA	4 (80%)	2 (28.6%)	1 (14.3%)	2 (28.6%)	NS	
Tratamiento con antibióticos [n(%)]	5 (100%)	3 (60%)	6 (85.7%)	6 (85.7%)	6 (85.7%)	NS	
Tabaquismo [n (%)]	2 (40%)	4 (80%)	3 (42.9%)	3 (42.9%)	2 (28.6%)	NS	
Agente etiológico de la neumonía [n (%)]							
Identificado	3 (60%)	4 (80%)	6 (85.7%)	6 (85.7%)	7 (100%)		
<i>Pneumocystis jirovecii</i>	0	1 (20%)	4 (57.1%)	5 (71.4%)	4 (57.1%)		
Otros hongos (<i>Mucorales</i> , <i>H. capsulatum</i> , <i>Cryptococcus sp.</i> , <i>A. flavus</i>)	1 (20%)	1 (20%)	0	0	2 (28.6%)		
<i>Mycobacterium tuberculosis</i> o <i>M. bovis</i>	2 (40%)	0	1 (14.3%)	0	1 (14.3%)	NS	
Otras bacterias (<i>Chlamydophila pneumoniae</i> , <i>K. pneumonia</i> , <i>S. hominis</i> , <i>M. avium</i> , <i>M. intracellulare</i>)	0	1 (20%)	0	0	3 (42.9%)		
Virus (Influenza, metapneumovirus)	0	1 (20%)	0	1 (14.3%)	0		
Otro (no microbiológico)	0	0	1 (14.3%)	1 (14.3%)	0		
No identificado	2 (40%)	1 (20%)	1 (14.3%)	1 (14.3%)	0		
Número de secuencias pre-procesadas de alta calidad obtenidas con NGS (millón de secuencias) [mediana (IQR)]	LBA	9.69 (7.15 – 10.85)	8.62 (7.07 – 9.64)	9.45 (8.53 – 10.44)	9.34 (7.47 – 9.99)	9.00 (6.97 – 9.28)	p = 0.773
	ORAL	11.55 (10.66 – 11.65)	10.18 (9.24 – 12.57)	11.54 (10.88 – 12.31)	12.00 (10.14 – 12.35)	10.91 (8.80 – 12.10)	p = 0.798

Tabla 1. Características de la cohorte. Abreviaturas: IQR: Rango intercuartil, NA: No aplica; NS: No significativo

Por otra parte, es importante resaltar que en la mayoría de los pacientes (83.87%) pudo identificarse el agente etiológico de la afección respiratoria; los patógenos hallados fueron virus, bacterias y hongos. El agente infeccioso más frecuentemente encontrado en personas infectadas por VIH fue *Pneumocystis jirovecii* (53.84%), el cual es un hongo ubicuo que comúnmente causa infecciones oportunistas en individuos con conteos de células CD4+ menores a 200 células/mL (CDC, 2017). Además, en 12.9% del total de individuos de la cohorte, se identificaron micobacterias causantes de tuberculosis, hongos distintos a *P. jirovecii* y otras bacterias. Resulta interesante que en 2 personas (6.5%) se hallaron virus respiratorios, los cuales fueron Influenza A H3N2 y metapneumovirus. Asimismo, en los dos grupos que incluían a los individuos con mayor grado de inmunodeficiencia, se encontraron 4 casos en los que 2 agentes infecciosos fueron reconocidos. Las combinaciones halladas fueron las siguientes: PCP (*Pneumocystis pneumonia*) e Influenza A H3N2; PCP y *Cryptococcus sp.*; PCP y *Mycobacterium avium*; PCP y *Chlamydomphila pneumoniae*.

6.2 Validación de la metodología de caracterización del viroma respiratorio

6.2.1 Identificación de virus humanos en las muestras respiratorias

Una de las maneras más sencillas de validar la metodología propuesta y estandarizada para este estudio, es analizar las características de los virus identificados. Con esta finalidad, en la Tabla 2 se presentan los distintos virus humanos encontrados en las muestras respiratorias de nuestra cohorte. Los virus humanos identificados pertenecen a 12 familias virales y tienen características muy diversas. Como puede verse en la Tabla 2, se identificaron partículas virales envueltas y sin envoltura, de 30 a más de 200 nm de diámetro, con genomas de DNA y RNA, de cadena sencilla y doble y de 2.8 a casi 300 kb (kilobases). Con base en esto, podemos asegurar que la metodología empleada permitió la identificación de una gran variedad de virus humanos con características muy distintas.

Familia viral	Especie	Genoma	Tamaño del virión	Otras características
<i>Adenoviridae</i>	HAdVs	dsDNA lineal ~36 kb	~90 nm	- Partícula no envuelta - Cápside icosaédrica
<i>Anelloviridae</i>	Anellovirus	ssDNA (-) circular 2.86 – 3.9 kb	30 - 50 nm	- Partícula no envuelta - Cápside icosaédrica
<i>Coronaviridae</i>	CoV	ssRNA (+) lineal 26 – 32 kb	118 -136 nm	- Partícula envuelta - Esférica con moderado grado de pleomorfismo - Genoma más grande de todos los virus de RNA
<i>Flaviviridae</i>	HPgV	ssRNA (+) lineal 10 - 11 kb	~50 nm	- Partícula envuelta - Esférica
<i>Herpesviridae</i>	VHH-1	dsDNA lineal 124 – 295 kbp	120 – 260 nm	- Partícula envuelta - Esférica
	(EBV) VHH-4			
	(CMV) VHH-5			
	VHH-6			
	VHH-7			
VHH-8				
<i>Orthomyxoviridae</i>	Influenza virus	ssRNA (-) lineal segmentado, 8 segmentos Tamaño de los segmentos: de 890 a 2,341 nt. Total 13.5 kb	~100 nm	- Partícula envuelta - Pleomórficas: esférica pero también puede ser filamentosa
<i>Papillomaviridae</i>	VPH	dsDNA circular ~8 kbp	~60 nm	- Partícula no envuelta - Cápside icosaédrica
<i>Paramyxoviridae</i>	MPVh	ssRNA (-) lineal ~13.2 kb	150 - 600 nm	- Partícula envuelta - Pleomórficas: esférica, pero también puede ser filamentosa
<i>Parvoviridae</i>	hPV	ssDNA lineal ~5 kb	~25 nm	- Partícula no envuelta - Cápside icosaédrica
<i>Picornaviridae</i>	RVH	ssRNA (+) lineal 7- 8.8 kb	~30 nm	- Partícula no envuelta - Cápside icosaédrica
	EVH			
<i>Polyomaviridae</i>	HPyV	dsDNA circular ~5 kbp	45 – 50 nm	- Partícula no envuelta - Cápside icosaédrica
<i>Retroviridae</i>	VIH	ssRNA (+) lineal dimérico ~9.7 kb	~100 nm	- Partícula envuelta - Esférica

Tabla 2. Características de los virus humanos identificados en las muestras respiratorias de los individuos de nuestra cohorte. Abreviaturas. dsDNA: DNA de doble cadena, ssDNA: DNA de cadena sencilla, ssRNA: RNA de cadena sencilla, kb: kilobases, kbp: kilopares de bases, nt: nucleótidos, HAdVs: Adenovirus humano, HPgV: Pegivirus humano, VHH: Virus de Herpes humano, EBV: Virus de Epstein Barr, CMV: Citomegalovirus, VPH: Virus del Papilloma Humano, hPV: Parvovirus Humano, RVH: Rinovirus Humano, HPyV: Poliomasvirus Humano, CoV: Coronavirus, MPVh: Metapneumovirus humano, EVH: Enterovirus humano (Knipe *et al.*, 2013 y Spandole *et al.*, 2015).

6.2.2 Confirmación de la detección de VIH y CMV

Con la finalidad de validar la metodología empleada para la caracterización del viroma respiratorio, se consideró útil comparar los resultados obtenidos por NGS con lo hallado con una técnica sensible y específica como la PCR cuantitativa (qPCR). En esta comparación únicamente se incluyeron VIH y CMV debido a la alta prevalencia de estos virus en la cohorte estudiada (ver en la sección 6.3.3 Identificación de secuencias de virus humanos mediante NGS). Además, al ser virus de RNA y DNA, respectivamente, permite validar que la metodología propuesta sea adecuada para la identificación de ambos tipos de virus.

Se realizó la determinación de la carga viral de CMV y VIH en el Laboratorio de Diagnóstico Viroológico (LDV) del CIENI, en los 26 LBAs de pacientes infectados por VIH. En la Figura 1 se muestran las correlaciones obtenidas entre el número de copias identificadas por qPCR y el número de secuencias normalizadas encontradas por NGS. En el caso del CMV, los resultados de PCR en tiempo real confirmaron la ausencia del virus en 10 de las 13 muestras negativas por NGS y la presencia del virus en 12 de las 13 muestras positivas por NGS (resultados idénticos en 22 de 26 muestras, que representan el 85%). En 3 muestras se detectó CMV por qPCR y no por NGS y en un LBA se detectó por NGS, pero no se identificó por qPCR.

En el caso del VIH, los resultados de PCR en tiempo real confirmaron la ausencia del virus en 10 de las 11 muestras negativas por NGS y la presencia del virus en las 15 muestras positivas por NGS (resultados idénticos en 25 de 26 muestras, que representan el 96%). Únicamente en una de las muestras se identificó VIH por qPCR, pero no por NGS.

Además, en el caso de ambos virus, se observa una correlación positiva entre el número normalizado de secuencias obtenido por NGS y la carga viral obtenida por qPCR. Dichas correlaciones resultan relevantes debido a que la qPCR es una técnica cuantitativa que determina el número de genomas virales presentes en una muestra, mientras que la técnica de NGS utilizada en este estudio es semicuantitativa y permite la detección de ácidos nucleicos virales presentes en partículas virales intactas.

Estos resultados, aunados a la identificación de partículas virales con características muy diversas, confirman la adecuada obtención, procesamiento y almacenamiento de las muestras, y validan la metodología experimental (purificación y concentración de las partículas virales, extracción de ácidos nucleicos y uso de cebadores aleatorios), la plataforma de secuenciación (NextSeq, Illumina) y la estrategia de análisis (*pipeline* SURPI y revisión manual de las secuencias).

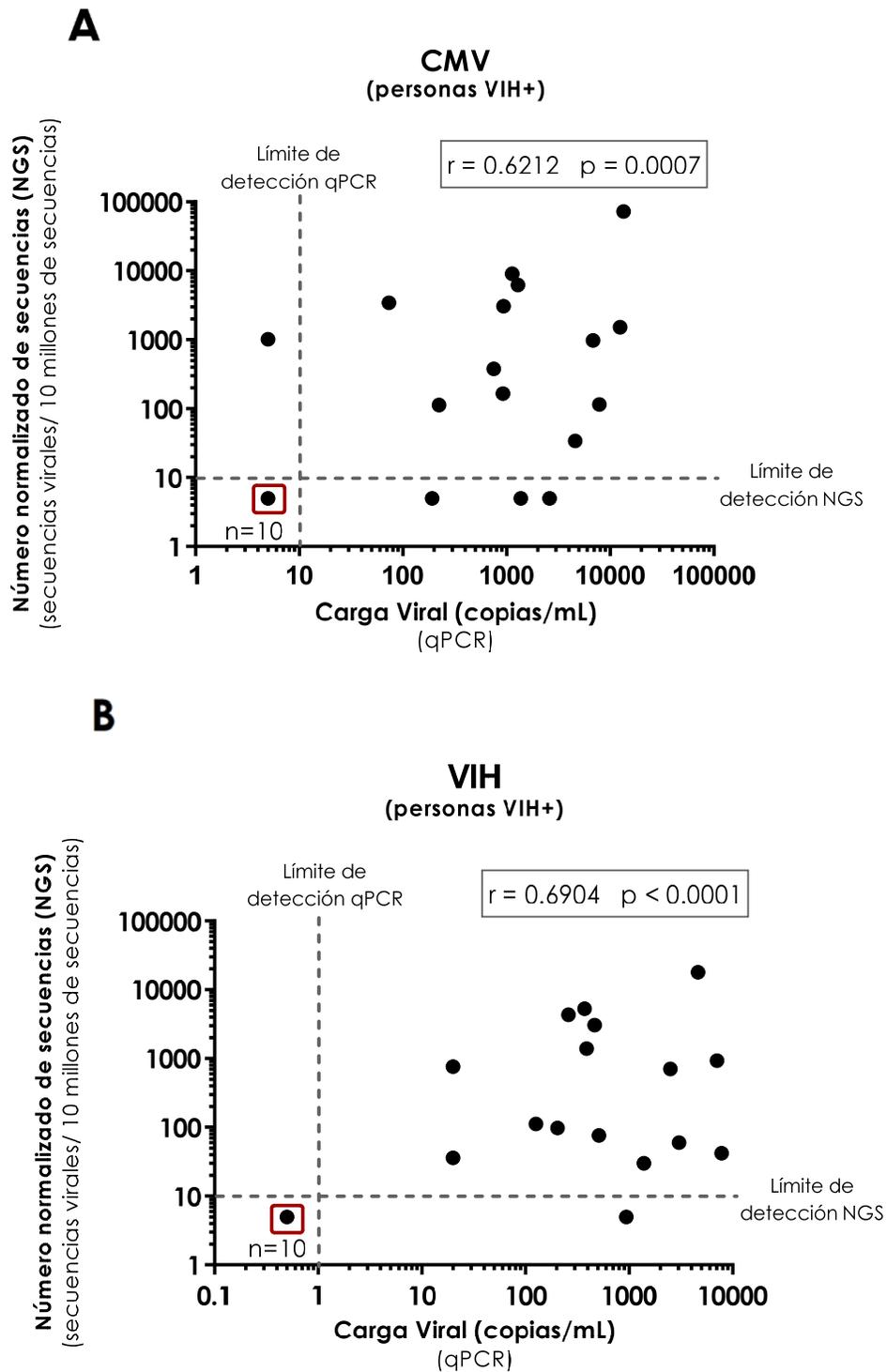


Figura 1. Correlación del número de copias de CMV (A) y de VIH (B) identificadas por PCR en tiempo real y el número de secuencias normalizadas encontradas por NGS. Los números de secuencias virales fueron normalizados considerando 10 millones de secuencias de alta calidad. Los límites de detección de cada técnica son indicados en líneas punteadas. Ambos ejes son representados en escala logarítmica. El cálculo de la correlación se realizó con la prueba Spearman.

6.3 Caracterización del viroma respiratorio

6.3.1 Clasificación de secuencias totales

La mediana del total de secuencias pre-procesadas (después de remover las secuencias duplicadas, los fragmentos correspondientes a los adaptadoras y las secuencias de baja calidad o complejidad) obtenidas en las muestras de lavado oral fue de 11,221,103 secuencias (IQR, 10,187,979 - 12,107,077), mientras que para los LBAs fue de 8,996,501 secuencias (IQR, 7,397,371 - 10,121,460). Para poder comparar los números de secuencias entre pacientes y grupos, éstas fueron normalizadas considerando 10 millones de secuencias de alta calidad.

Asimismo, la totalidad de secuencias obtenidas en cada muestra fueron clasificadas, de acuerdo a lo identificado en SURPI, en las siguientes categorías de secuencias: virales, bacterianas, humanas, pertenecientes a otros organismos y no identificadas. El porcentaje de cada tipo de secuencias en las diferentes muestras es representado en la Figura 2. En ella puede observarse que la mayor parte de las secuencias encontradas en los lavados orales no fueron asociadas con organismos conocidos (mediana, 72.99%; IQR, 42.86% - 91.01%); mientras que, en los LBAs, la mayoría de las secuencias eran de origen humano (mediana, 82.31%; IQR, 74.30% - 87.18%). En cuanto a las secuencias virales, la mediana del porcentaje de éstas halladas en la cavidad oral, fue de 1.230% (IQR, 0.519% - 17.024%); mientras que, en las muestras de tracto respiratorio inferior, la mediana de lo correspondiente a secuencias virales fue de 0.061% (IQR, 0.018% - 0.385%). Asimismo, llama la atención que, en los lavados orales, existe una variación interindividual importante; lo cual resulta particularmente evidente en las secuencias virales, puesto que en 14 muestras de la cavidad oral (45.16%) el porcentaje de secuencias virales fue menor al 1%, en 3 muestras (9.68%) cerca de la mitad del total de secuencias correspondía a virus y en 1 de los lavados orales casi el 90% de lo identificado era de origen viral. Finalmente, es importante mencionar que en ningún caso se encontraron diferencias significativas en la clasificación de las secuencias entre los distintos grupos de pacientes ($p > 0.05$).

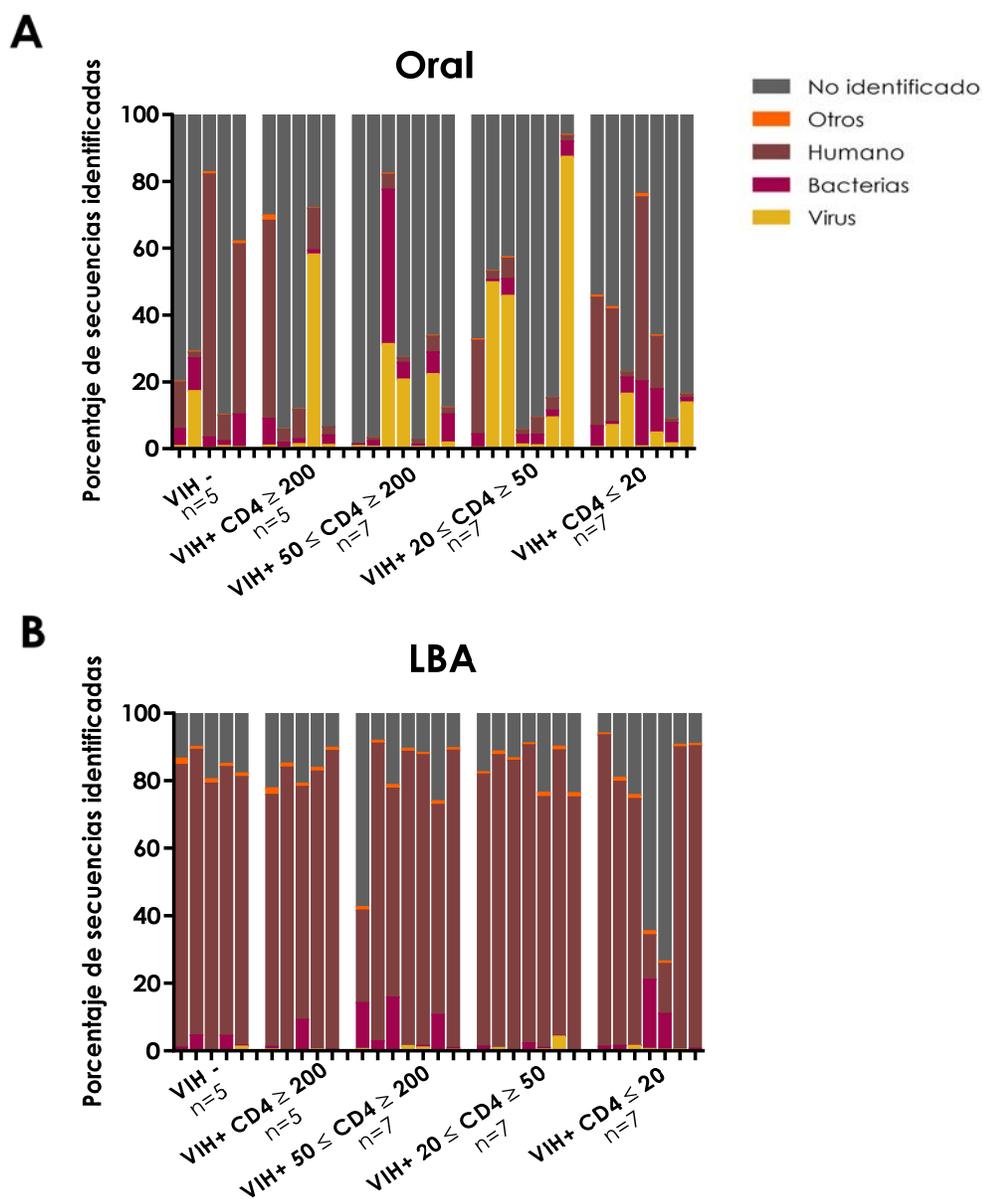


Figura 2. Clasificación de secuencias totales en lavados orales (A) y LBA (B). Se muestra la proporción de secuencias clasificadas por SURPI. Cada barra corresponde a un paciente. En cada grupo, las muestras fueron ordenadas de acuerdo al conteo de células T CD4+ de manera decreciente (de izquierda a derecha).

6.3.2 Clasificación

6.3.3 de secuencias virales de acuerdo al hospedero

De manera importante y como puede observarse en las Figuras 3C y 3D, encontramos grandes diferencias en el número de secuencias virales entre los individuos (oral: mediana, 72,156; IQR, 24,531-1.675x10⁶; LBA: mediana, 4,085; IQR, 510 – 17,587). Asimismo, las secuencias virales fueron clasificadas de acuerdo al

hospedero que dichos virus son capaces de infectar. Los virus de vertebrados fueron los más abundantes en la mayoría de las muestras tanto orales (mediana, 80.98%; IQR, 44.43% - 98.82%) como broncoalveolares (mediana, 98.57; IQR, 86.78% - 99.71%). Además, en 8 de 31 lavados de la cavidad oral (25.8%), más de la mitad de secuencias virales correspondían a fagos (mediana, 13.86%; IQR, 0.979% - 54.54%), mientras que únicamente en 3 LBAs (9.67%) esta clase de virus fueron los más abundantes (mediana, 0.9780%; IQR, 0.019% - 11.56%). Asimismo, en ambos tipos de lavados fueron identificados virus de plantas, protistas, arqueas, invertebrados y algunos cuyo hospedero no pudo ser identificado con las referencias disponibles en la base de datos del NCBI de agosto de 2015; no obstante, estos virus se encontraron en baja proporción y frecuencia (Figura 3).

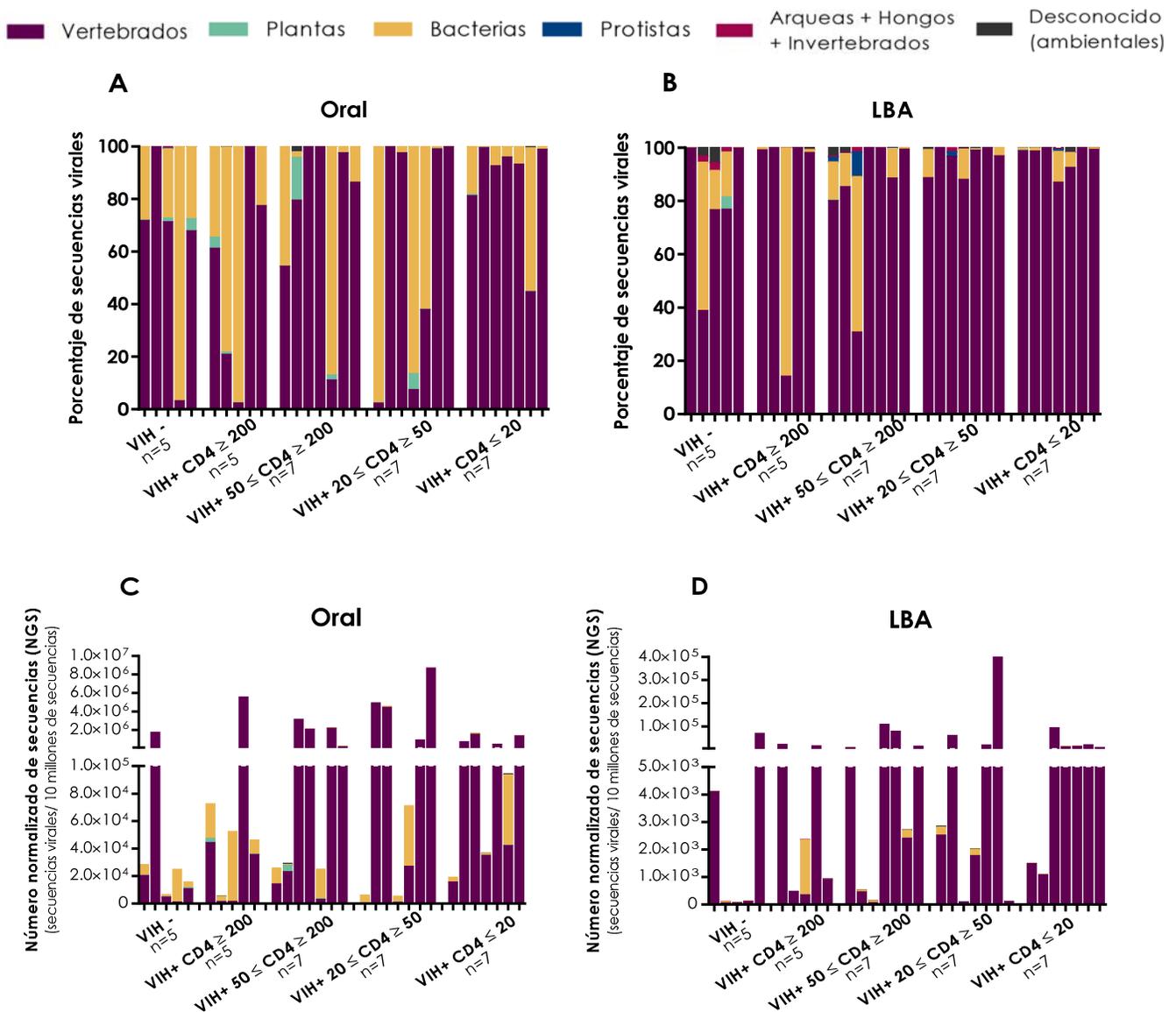


Figura 3. Clasificación de secuencias virales. Se muestra la proporción de secuencias virales de acuerdo al hospedero en lavados orales (A) y LBA (B). En las secciones inferiores se presentan el número normalizado de secuencias virales en las muestras orales (C) y broncoalveolares (D). Cada barra corresponde a un paciente. En cada grupo, las muestras fueron ordenadas de acuerdo al conteo de células T CD4+ de manera decreciente (de izquierda a derecha).

6.3.4 Identificación de secuencias de virus humanos mediante NGS

Una vez que las secuencias virales fueron clasificadas, decidimos enfocarnos en el análisis de los virus humanos debido a su importancia clínica y a que representan la mayoría de los virus identificados.

Para considerar que cierto virus estaba presente en una muestra se tomaron en cuenta solo aquellos que tuvieran más de diez secuencias, que no se hubieran identificado en los controles negativos y que se encontraran en más de una región del genoma viral. El número de secuencias de virus humanos, en escala logarítmica, es presentado a manera de heatmap en la Figura 4.

La representación en forma de heatmap del número de secuencias encontradas en cada muestra permite comparar fácilmente los resultados obtenidos. Es notorio, por ejemplo, que se reconocieron distintos virus pertenecientes a diferentes familias virales. Los virus con mayor prevalencia en el tracto respiratorio superior fueron: anellovirus con 83.9%, virus de Epstein Barr (EBV) y papillomavirus con 80.6%, en ambos casos; seguidos de los herpesvirus tipo 7 y 6 con una prevalencia del 67.7% y 54.8%, respectivamente. Asimismo, en poco menos de la mitad de los pacientes (48.4%) se encontraron secuencias de citomegalovirus (CMV); mientras que pegivirus humano y rinovirus fueron hallados 12.2% de ellos. Además, otros virus fueron identificados en algunas de las muestras orales de nuestra cohorte, este fue el caso de los herpesvirus tipo 1 y 8 que se encontraron en 9.7% y 6.5% de los lavados orales, respectivamente; poliomavirus con una prevalencia del 6.5% y finalmente, adenovirus, metapneumovirus y parvovirus que se encontraron en una sola muestra (3.2%). En cuanto al VIH, se lograron identificar secuencias de este virus en 4 de los 26 pacientes infectados (15.4%) y de manera relevante, ninguno de éstos 4 individuos había recibido TAR al momento de la toma de muestra; esto corresponde a casi la cuarta parte (23.5%) de las personas VIH positivas sin TAR.

Por otro lado, a nivel broncoalveolar, los virus ubicuos mayormente encontrados fueron anellovirus (64.5%) y CMV (41.9%). Además, en casi una tercera parte de este tipo de muestras se identificaron papillomavirus (29%); mientras que pegivirus y herpesvirus tipo 6 y 7, tuvieron una prevalencia más baja (12.9%, 6.5% y 3.2%, respectivamente). Por otra parte, resulta interesante la identificación de distintitos virus respiratorios como rinovirus, que estuvo presente en 2 de estas muestras (6.5%), además de coronavirus, virus de la influenza, metapneumovirus y enterovirus, cada uno hallado en una muestra distinta del tracto respiratorio inferior (3.2%). De manera importante, en los pacientes en los que se identificaron secuencias de metapneumovirus e influenza A H3N2, dichos virus también fueron reportados como agentes etiológicos de la neumonía en las pruebas realizadas para diagnóstico clínico. Finalmente, es importante resaltar que en más de la mitad de las personas VIH positivas (57.7%) se encontraron secuencias de este virus, lo cual corresponde a 15 de las 17 muestras (88.2%) de pacientes que no habían recibido TAR antes de la toma de muestra. La alta prevalencia de este patógeno en los lavados broncoalveolares se debe a la replicación del VIH en el tracto respiratorio inferior, en particular en los macrófagos alveolares (Koziel *et al.*, 1999).

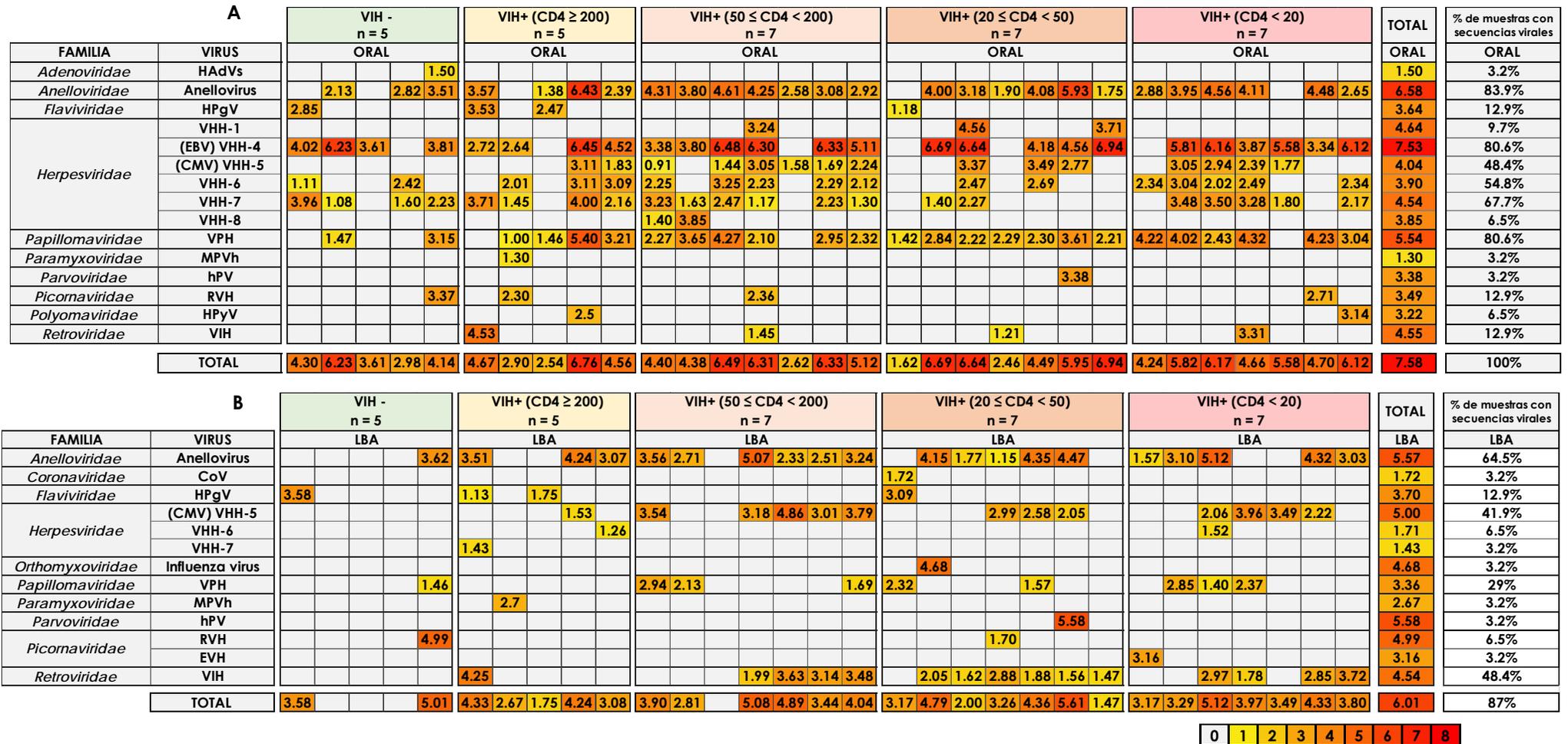


Figura 4. Heatmaps de secuencias virales identificadas en las muestras de lavado oral (A) y LBA (B) de cada individuo. Las secuencias fueron normalizadas considerando 10 millones de secuencias de alta calidad. El número de secuencias se representa en escala logarítmica. El código de colores se muestra en la parte inferior derecha de la figura. En cada grupo, las muestras fueron ordenadas de acuerdo al conteo de células T CD4+ de manera decreciente (de izquierda a derecha).

Abreviaturas. n: número de individuos, HAdVs: Adenovirus humano, HPgV: Pegivirus humano, VHH: Virus de Herpes humano, EBV: Virus de Epstein Barr, CMV: Citomegalovirus, VPH: Virus del Papilloma Humano, hPV: Parvovirus Humano, RVH: Rinovirus Humano, HPyV: Poliomasvirus Humano, CoV: Coronavirus, MPVh: Metapneumovirus humano, EVH: Enterovirus humano.

Por otra parte, estos heatmaps sugieren que se hallaron una mayor cantidad de virus humanos en los lavados orales que en los LBAs. Dichos resultados fueron corroborados al comparar el número de secuencias de virus humanos totales en cada tipo de muestra ($p < 0.0001$). Más aún, la cantidad de secuencias de herpesvirus y papillomavirus también fue significativamente mayor en las muestras orales, en comparación con los LBAs ($p < 0.0001$, en ambos casos). Lo anterior es esperado ya que la cavidad oral está directamente expuesta al ambiente. En contraste, el número de secuencias de VIH tendió a ser mayor en el tracto respiratorio inferior ($p = 0.0554$) (Figura 5).

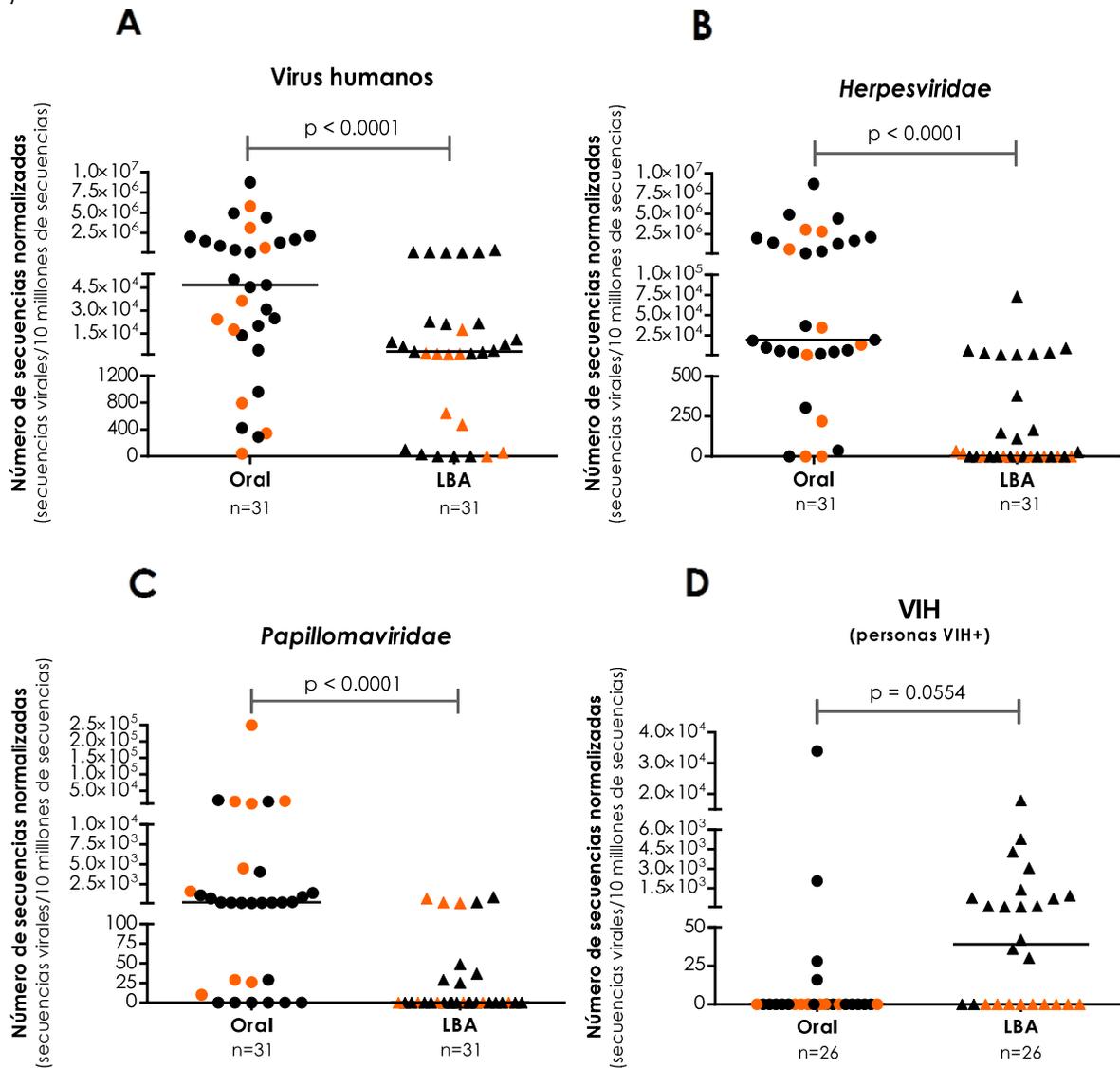
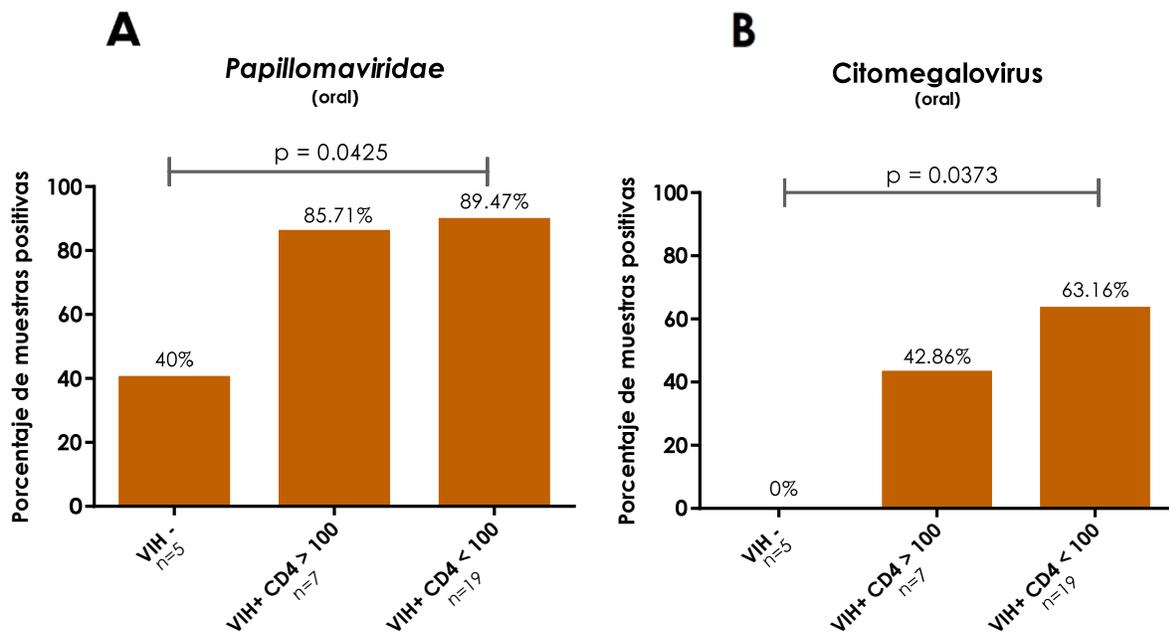


Figura 5. Comparación del número de secuencias virales obtenidas en los lavados orales y los LBAs. Las secuencias fueron normalizadas considerando el número de secuencias virales por cada 10 millones de secuencias de alta calidad. Las líneas horizontales indican la mediana de cada grupo. Los íconos anaranjados representan a los pacientes que recibían tratamiento antirretroviral (TAR) al momento de la toma de muestra. La comparación entre grupos se realizó con la prueba no paramétrica de Wilcoxon.

6.3.5 Prevalencia de virus humanos

El análisis de los posibles cambios en la composición del viroma respiratorio puede realizarse de dos maneras distintas. La primera de ellas es identificar y contrastar la presencia o ausencia de cada virus o familia viral en los diferentes grupos de pacientes; de esta manera se identifican posibles diferencias en la prevalencia. La segunda forma consiste en comparar el número de secuencias virales entre los grupos. Los resultados de esta comparación reflejan probables modificaciones en la abundancia viral.

Cuando se comparó la prevalencia de los virus o familias virales más comunes en nuestra cohorte se encontró que los papillomavirus fueron más frecuentemente identificados en los lavados orales de personas VIH positivas con conteos de CD4 menores a 100 células/ μ L de sangre, en comparación con este mismo tipo de muestra de individuos sin VIH ($p=0.0425$). Asimismo, al comparar estos dos grupos también se reconoció que el CMV fue más prevalente en los individuos con mayor grado de inmunosupresión, tanto en los lavados orales como en los LBAs ($p=0.0373$ y $p=0.0411$, respectivamente). Además, en el caso particular de las muestras del tracto respiratorio inferior, la identificación de virus humanos totales fue más frecuente en personas infectadas por VIH en comparación con los individuos VIH negativos ($p= 0.0117$). Más aún, en los LBAs, tanto los anellovirus como los herpesvirus fueron más comúnmente encontrados en el grupo con mayor grado de inmunodeficiencia ($p=0.0474$ y $p=0.0411$, respectivamente) (Figura 6).



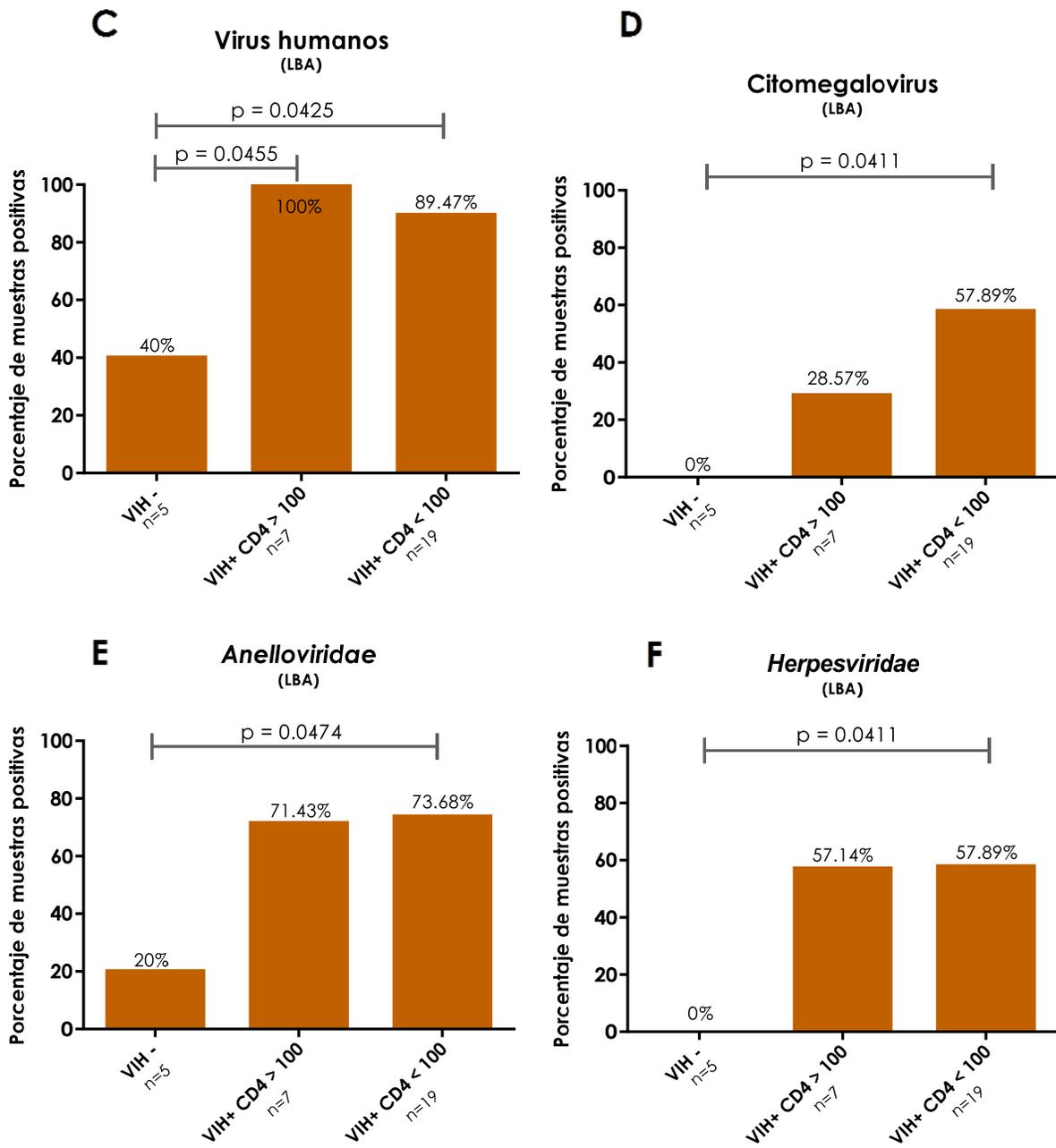
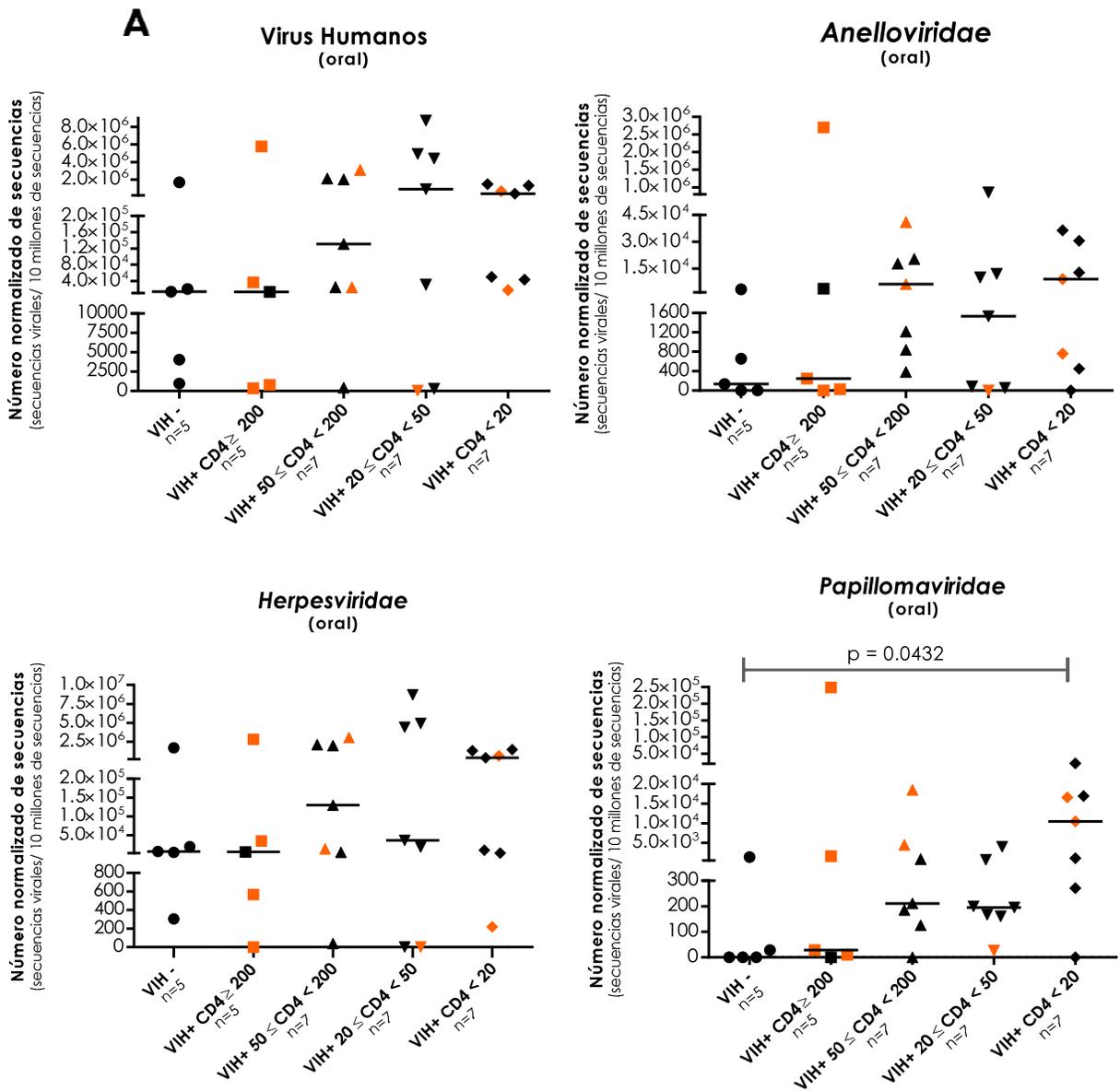


Figura 6. Prevalencia de virus humanos, citomegalovirus, papillomavirus, anellovirus y herpesvirus en lavados orales y LBAs. Se muestran los virus o familias virales cuya prevalencia, en lavados orales o LBAs, fue significativamente diferente entre los individuos sin VIH, VIH+ con más 100 células CD4+/μL de sangre y VIH+ altamente inmunocomprometidos (<100 células CD4+/μL). La comparación entre grupos se realizó con la prueba de Fisher.

6.3.6 Abundancia de virus humanos

6.3.5.1 Efecto de la inmunodeficiencia causada por la infección de VIH

Con la finalidad de identificar posibles diferencias en el número de secuencias virales entre los 5 grupos de estudio, se compararon la cantidad de secuencias de los virus humanos totales (sin considerar al VIH) y de las familias virales más prevalentes. Los resultados mostraron que existen diferencias significativas en el número de secuencias de papillomavirus en las muestras orales de pacientes con VIH y con menos de 20 células CD4+/ μ L de sangre, en comparación con los individuos VIH negativos ($p=0.0432$) (Figura 7).



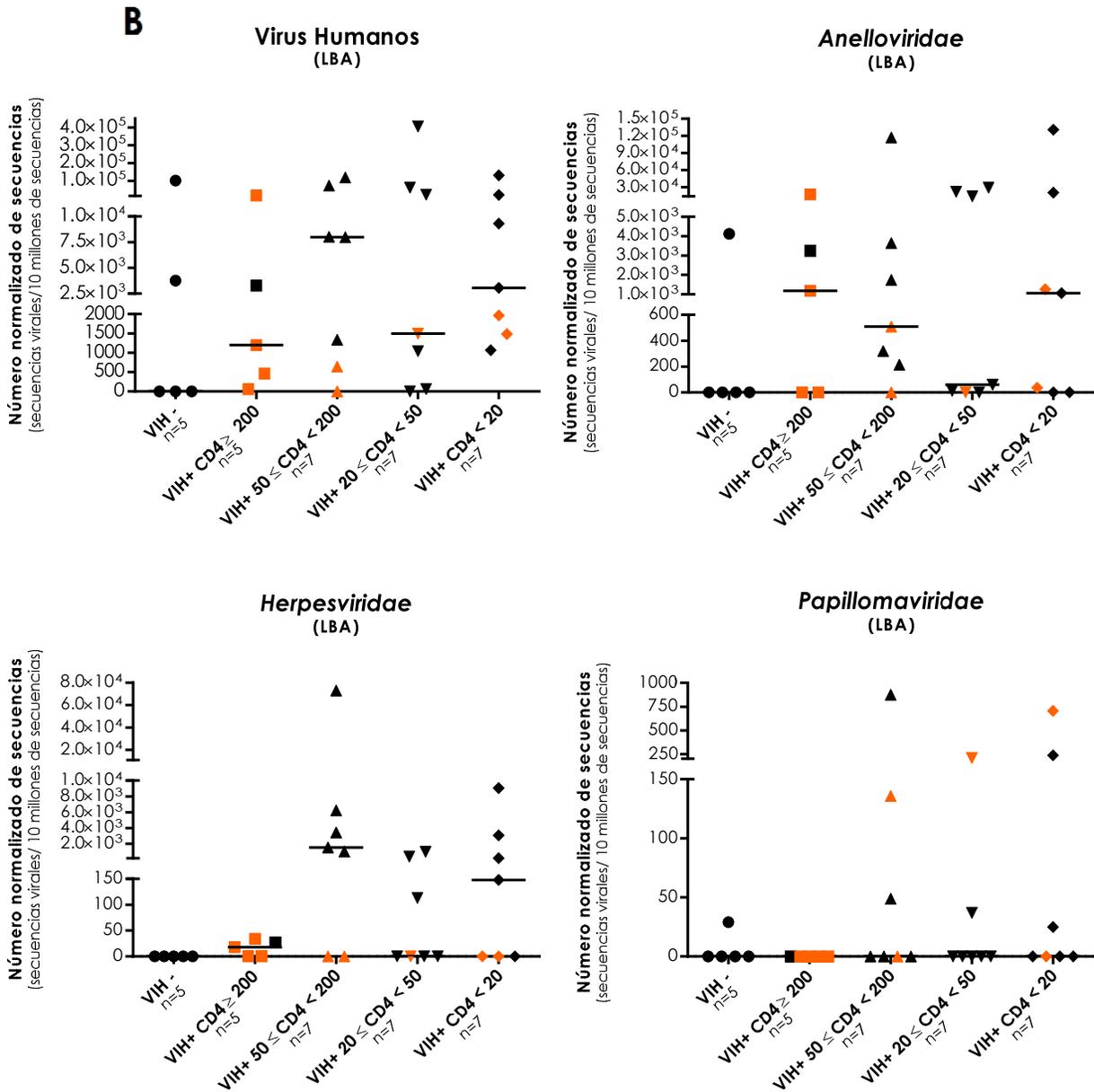


Figura 7. Comparación del número de secuencias virales obtenidas entre grupos en los lavados orales (A) y los LBAs (B). Las secuencias fueron normalizadas considerando el número de secuencias virales por cada 10 millones de secuencias de alta calidad. Las líneas horizontales indican la mediana de cada grupo. Los íconos anaranjados representan a los pacientes que recibían tratamiento antirretroviral (TAR) al momento de la toma de muestra. La comparación entre grupos se realizó con la prueba no paramétrica de Kruskal-Wallis.

6.3.5.2 Efecto de variables clínicas y epidemiológicas

Debido a que existen varios factores que pudieran influir en la composición del viroma respiratorio, se realizó un análisis multivariante en el que se compararon las abundancias relativas de las secuencias encontradas en las muestras de pacientes, con distintas características clínicas y epidemiológicas. Para ello, se consideró $q=0.05$ como valor umbral de significancia y se utilizó el programa MaAsLin (*Multivariate Association with Linear Models*) disponible en línea (Morgan *et al.*, 2012). Las variables analizadas fueron: sexo, edad, tabaquismo, estatus de VIH, tratamiento antirretroviral, con antibióticos, antimicóticos o antivirales, diagnóstico de la afección respiratoria, carga viral de VIH, conteo de células CD4+ y CD8+, razón de células CD4/CD8 y estación del año.

En el caso de los lavados orales se observó que las personas que fuman tienen una menor abundancia relativa de secuencias bacterianas ($p=0.0007$; $q=0.0437$), en comparación con las personas que no fuman, independientemente de las otras variables. Esto sugiere que el hábito tabáquico provoca alteraciones en el microbioma oral. Por otra parte, en los LBAs se encontró asociación entre la administración de tratamiento antirretroviral y la ausencia de secuencias de VIH en el tracto respiratorio inferior ($p=0.0001$, $q=0.0265$) (Figura 8). Lo cual es esperado debido a que el tratamiento antirretroviral provoca una disminución significativa en la carga viral de VIH en los pulmones (Twigg *et al.*, 2008).

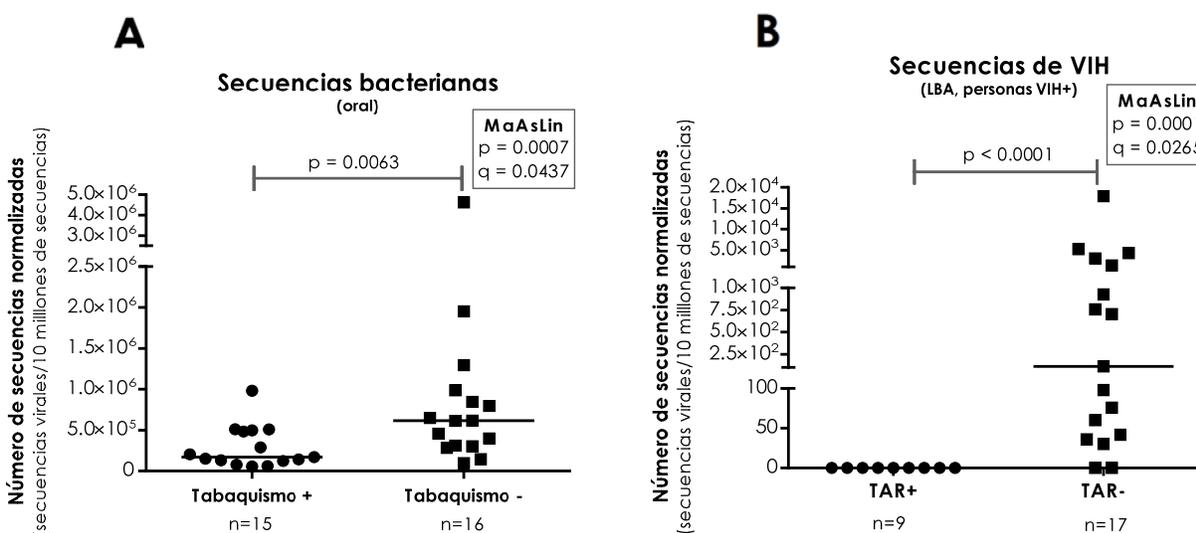
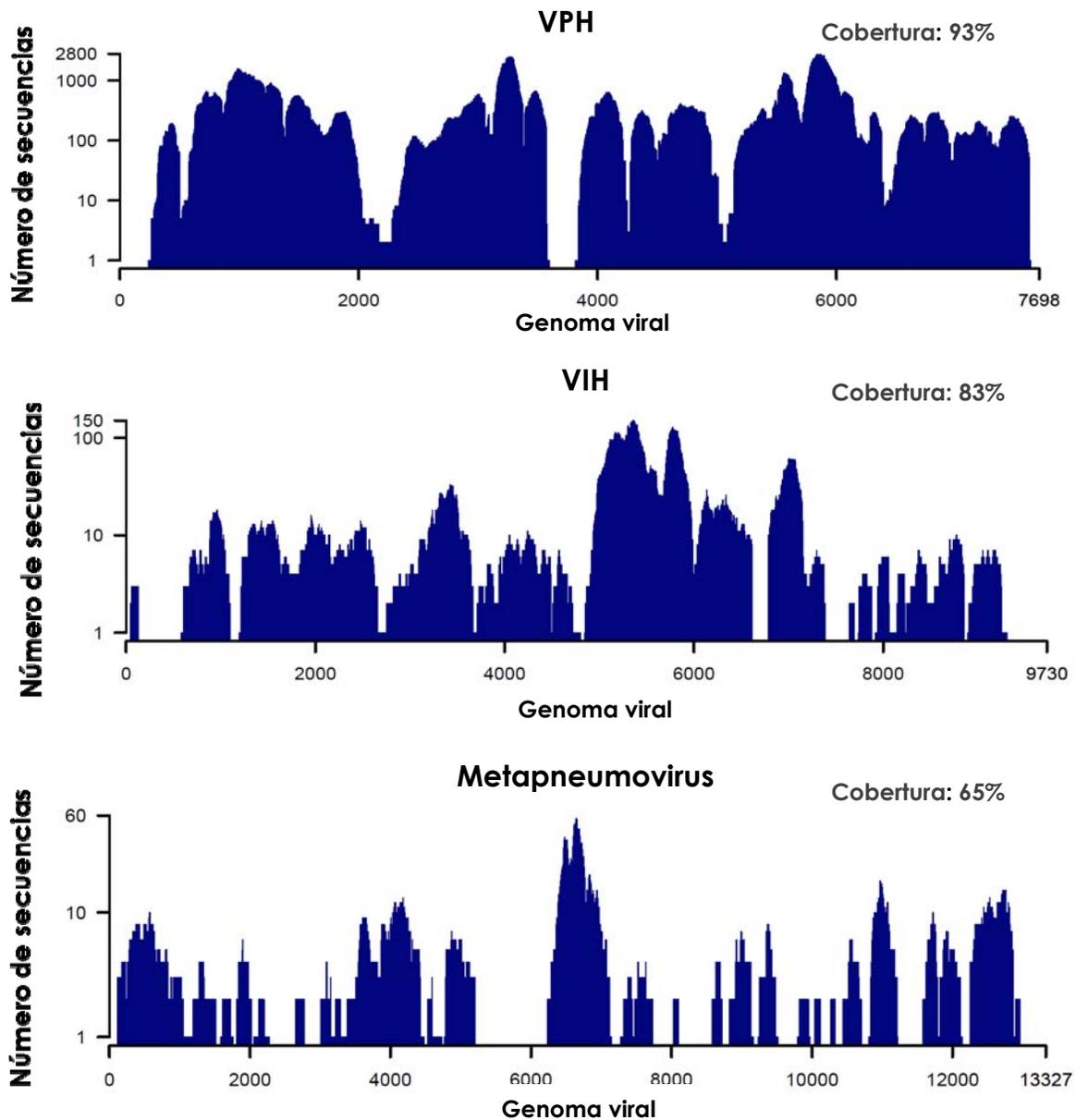


Figura 8. Comparación del número de secuencias obtenidas entre grupos en los lavados orales y los LBAs. Se muestra la comparación de secuencias bacterianas en los lavados orales de personas que fuman y que no fuman (A); además se presenta la diferencia en el número de secuencias de VIH presentes en los LBAs de pacientes con y sin TAR (B). Las líneas horizontales indican la mediana de cada grupo. La comparación entre grupos se realizó con la prueba no paramétrica de Mann-Whitney. El análisis multivariante se hizo con el programa MaAsLin.

6.3.7 Coberturas de los genomas virales

Como se mencionó en apartados anteriores, una de las condiciones para asegurar la presencia de cierto virus en una muestra fue que se encontraran secuencias en más de una región del genoma viral, por lo que todos los virus identificados en este estudio tuvieron esta característica. Asimismo, en algunos casos se obtuvieron los genomas completos o casi completos de algunos virus, este fue el caso de VPH, influenza, parvovirus y poliomavirus.

La cobertura de los genomas virales depende tanto del número de secuencias encontradas como del tamaño del genoma viral. Por ello, en algunos casos, a pesar de identificarse una gran cantidad de secuencias virales, la cobertura fue muy baja debido al mayor tamaño del genoma de referencia. Algunos ejemplos de mapas de cobertura del genoma viral se presentan en la Figura 9.



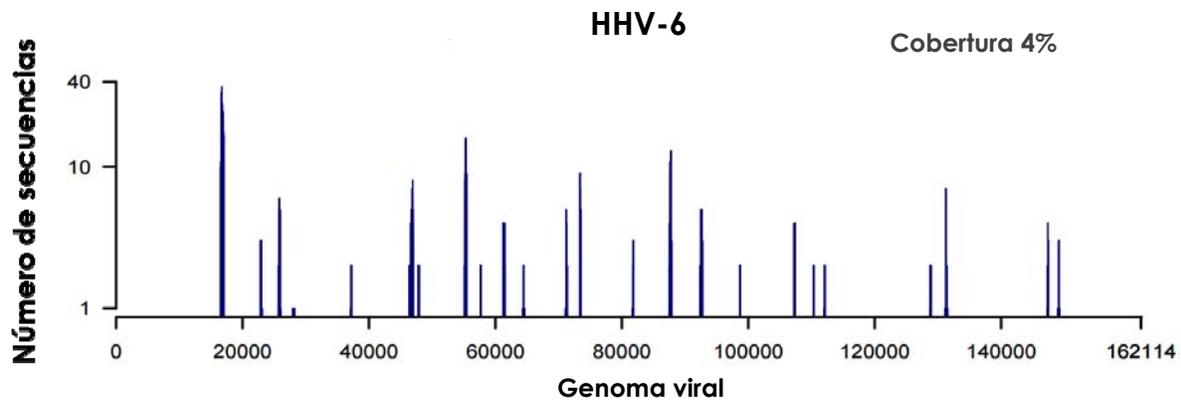


Figura 9. Mapas de cobertura del genoma viral. Se muestran mapas con distintos porcentajes de cobertura. En el eje x se representa la posición en el genoma viral y en el eje y se presentan el número secuencias que alinearon con el genoma de referencia. En la parte superior derecha de cada representación se muestra el porcentaje de cobertura. La representación gráfica se realizó con el programa RStudio.

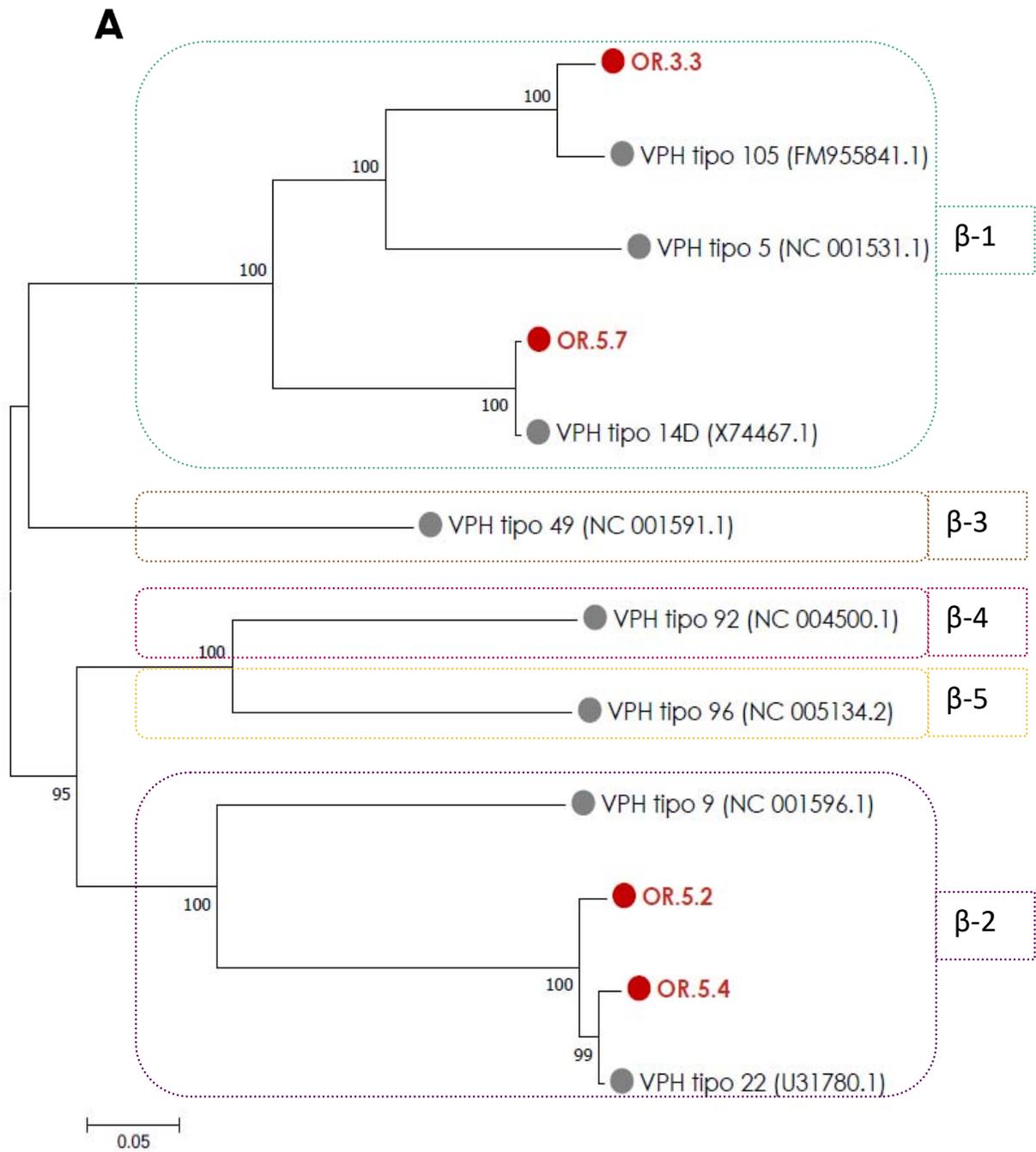
6.3.8 Análisis filogenético de las secuencias de VPH

Debido a que la infección por VPH puede tener importantes implicaciones clínicas y a la alta prevalencia de este grupo de virus en la cohorte en estudio, se consideró necesario determinar la diversidad taxonómica de las secuencias encontradas e identificar los tipos de los papillomavirus presentes en las muestras. Los papillomavirus identificados pertenecen a los géneros *Betapapillomavirus* y *Gammapapillomavirus*, por lo que se construyeron dos árboles filogenéticos, uno para cada uno de estos géneros de VPH. Para ello, se incluyeron algunas secuencias de cada muestra que, en conjunto, tuvieran una cobertura igual o mayor al 89% del genoma viral más cercano reportado en las bases de datos del NCBI. Asimismo, se consideraron secuencias de referencia representativas de la diversidad genómica de cada uno de los géneros encontrados; además se incluyeron los genomas virales que, de acuerdo a lo reportado por SURPI, eran los más estrechamente relacionados con los identificados en las muestras de interés. Las secuencias usadas como referencia corresponden a genomas virales completos. Todas las secuencias incluidas en cada árbol filogenético fueron alineadas de manera global con el programa CLC Genomics Workbench (QIAGEN). Los árboles filogenéticos fueron construidos usando el método de máxima verosimilitud (MV) y el Modelo General de Tiempo Reversible; su significancia se probó mediante 1,000 réplicas de *bootstrap* (Figura 10).

Resulta interesante que en 10 lavados orales el número de secuencias fue suficiente para cubrir un alto porcentaje del genoma viral y en consecuencia, ser incluidos en este análisis. Por otro lado, los géneros de VPH identificados fueron beta (Figura 10 A) y gamma (Figura 10 B). En el caso del género *Betapapillomavirus*, todas las distintas especies fueron enmarcadas en recuadros punteados; mientras que para el género *Gammapapillomavirus* sólo se especificaron las especies

de los tipos virales más cercanos a los encontrados en nuestra cohorte. Asimismo, vale la pena mencionar que, en la mayoría de los casos, las variantes virales secuenciadas tuvieron un porcentaje de identidad mayor al 95% con la secuencia de referencia más estrechamente relacionada; a excepción de la muestra OR.2.4, cuya secuencia fue más disímil a la usada como referencia (81.9% de identidad). Lo cual es relevante debido a que podría tratarse de un nuevo tipo de VPH. Para determinar si se trata de un nuevo tipo de *Gammapapillomavirus*, se hizo el alineamiento de la secuencia identificada en este estudio, con la referencia más cercana (VPH tipo 201[KP692115.1]) en la región del gen L1 de este virus, ya que la clasificación filogenética de VPH se realiza con base en el porcentaje de identidad de esta región del genoma (Knipe *et al.*, 2013). Los resultados de este alineamiento indican que ambas secuencias tienen una identidad del 81.2% en el sitio del gen L1. Debido a que dentro de una especie de papillomavirus, los tipos de VPH comparten entre el 71% y el 89% de su secuencia (Knipe *et al.*, 2013), este resultado corrobora el hallazgo de un nuevo tipo de *Gammapapillomavirus*.

Por otro lado, llama la atención la diversidad de secuencias de VPH identificadas en las muestras orales. Dichas secuencias pertenecen a 2 géneros distintos y a 7 especies diferentes. Los géneros de VPH a los que pertenecen las secuencias de papillomavirus identificadas en este estudio (beta y gamma), incluyen a muchos tipos virales considerados como comensales (Knipe *et al.*, 2013). Asimismo, es importante mencionar que no se identificaron secuencias de VPH del género alpha. Este género es el de mayor importancia clínica debido a que agrupa distintos tipos de VPH, de alto y bajo riesgo, asociados al desarrollo de cáncer genital y otras mucosas (Knipe *et al.*, 2013).



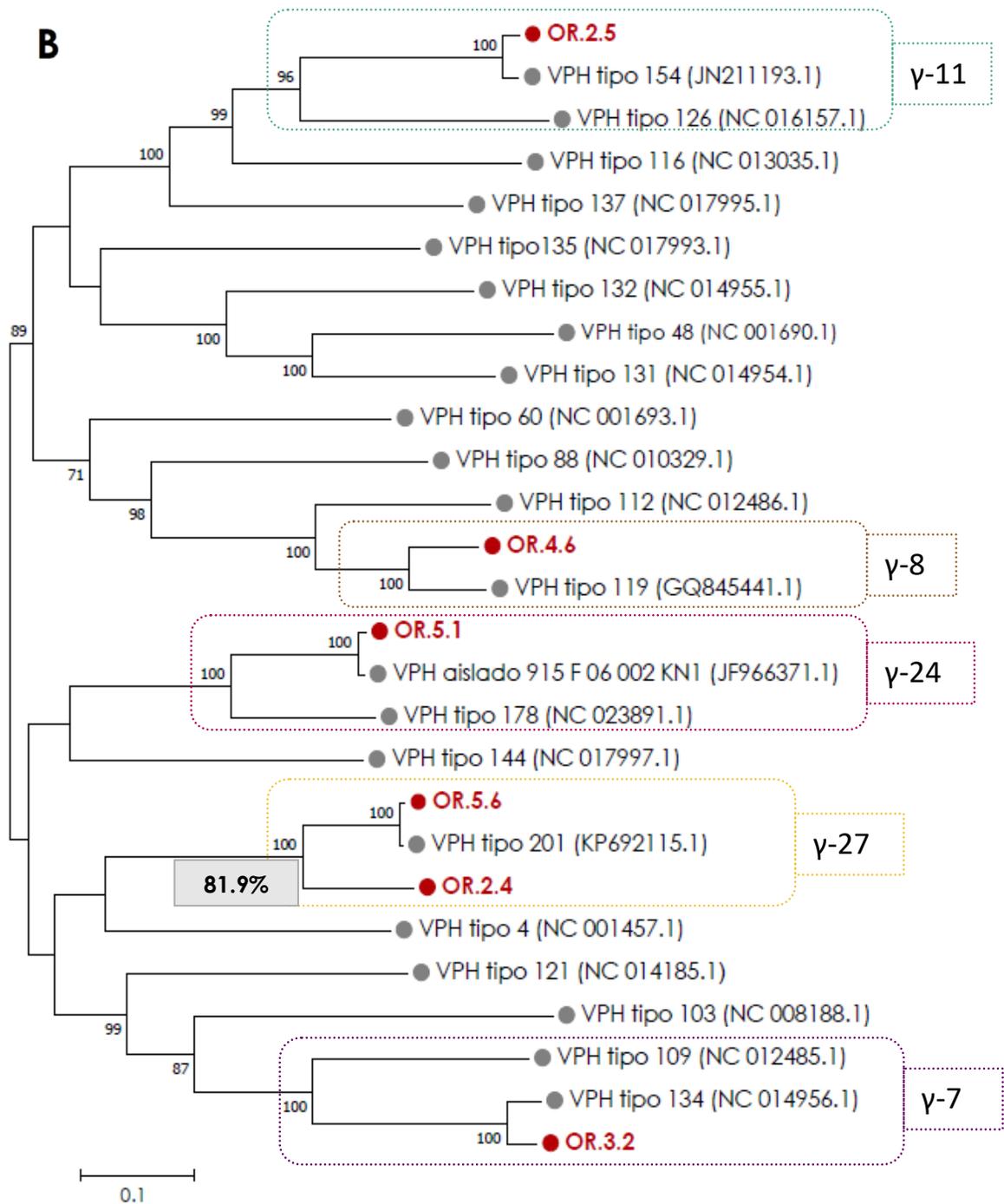


Figura 10. Árboles filogenéticos de papillomavirus de los géneros beta (A) y gamma (B). Se incluyeron las secuencias de cada muestra que tuvieran una cobertura igual o mayor al 89% del genoma viral más cercano. Éstas se representan en color rojo y las claves con las que son nombradas corresponden a lo siguiente: OR, es el tipo de muestra (oral); el primer número corresponde al grupo de pacientes de acuerdo al grado de inmunodeficiencia (1 para los VIH negativos y 5 para VIH positivos con menos de 20 células CD4+/ μ L) y el segundo número hace referencia a la posición dentro del grupo, cuando las muestras fueron ordenadas de acuerdo al conteo de células T CD4+ de manera decreciente. En color gris se presentan los genomas completos de diferentes especies de VPH representativos de la diversidad genómica de los géneros beta (A) y gamma (B), además de los genomas virales más estrechamente relacionados con los

identificados en las muestras. El porcentaje de identidad de la secuencia encontrada en la muestra OR.2.4 y la usada como referencia, se muestran en el recuadro gris al lado izquierdo del nodo interno. Las secuencias de referencia corresponden a genomas virales completos. Todas las secuencias incluidas en cada árbol filogenético fueron alineadas de manera global con el programa CLC Genomics Workbench (QIAGEN). La filogenia fue inferida utilizando el método de Máxima Verosimilitud en el software MEGA 7. Se muestra el árbol consenso inferido a partir de 1,000 réplicas de *bootstrap*. Algunas especies virales se muestran enmarcadas en recuadros punteados de distintos colores.

7. DISCUSIÓN

En el presente estudio se caracterizó el viroma del tracto respiratorio superior e inferior de individuos mexicanos VIH positivos y con síntomas de neumonía, con la finalidad de aportar datos que permitan determinar el efecto de la inmunodeficiencia asociada a este patógeno sobre la composición del componente viral del microbioma del tracto respiratorio. Para ello, se seleccionó una cohorte de 26 personas infectadas por VIH y con diferentes grados de inmunodeficiencia; además se incluyeron, como controles, a 5 pacientes VIH negativos con sospecha de neumonía. La particularidad de nuestro grupo de estudio es el enriquecimiento en personas en etapa de SIDA, lo cual está relacionado con el diagnóstico tardío característico de países con ingresos bajos o medios (Crabtree *et al.*, 2011). Además, dicha característica también está vinculada a que la cohorte está conformada por individuos con síntomas respiratorios y, debido a la alta prevalencia (70% - 80%) de PCP (*Pneumocystis pneumonia*) en personas con SIDA (CDC, 2017), era de esperarse que gran parte de los individuos VIH positivos incluidos en este estudio estuvieran en etapas avanzadas de la infección.

De manera relevante, el presente trabajo es uno de los primeros estudios realizados para la descripción del viroma respiratorio en personas infectadas por VIH. Además, la metodología estandarizada para este proyecto permitió reconocer virus de DNA y RNA; lo cual representa una ventaja en comparación con otros estudios del viroma humano que se han enfocado en la identificación de virus de DNA (Pérez-Broca & Moya, 2018; Pride *et al.*, 2012; De Vlaminc *et al.*, 2013; Young *et al.*, 2015). La detección y el análisis de virus de DNA y RNA es importante debido a que el viroma humano está conformado por ambos tipos de virus, por ello, a pesar de que la identificación de los virus de RNA sea metodológicamente más compleja, es necesaria para caracterizar completamente el viroma del tracto respiratorio. Más aún, muchos virus respiratorios tienen un genoma de RNA, ejemplo de ello son: influenza, parainfluenza, virus sincicial respiratorio (VSR), metapneumovirus, rinovirus, coronavirus y enterovirus. Además, varios de estos virus fueron identificados en nuestra cohorte, lo que hace aún más evidente la importancia de la identificación de virus de RNA para la descripción del viroma respiratorio superior e inferior.

Por otro lado, existe un gran número de factores que pueden intervenir en la constitución del microbioma respiratorio, algunos de ellos son la edad, el sexo, el tratamiento con antibióticos, antivirales o antifúngicos, el hábito tabáquico, la estación del año, el estado general de salud, las infecciones respiratorias y los padecimientos crónicos, los hábitos de higiene y la alimentación. Además, en nuestra cohorte en particular, debemos añadir a estos factores, la inmunodeficiencia asociada a la infección por VIH y el tratamiento con antirretrovirales. No obstante, el impacto de cada uno de estos elementos en la composición del viroma no ha sido completamente determinado. Por esta razón, algunas de estas variables fueron recabadas y comparadas entre los distintos grupos de nuestra cohorte. A excepción del número de células T CD4+ y la razón de

células CD4/CD8, ninguna diferencia significativa fue identificada entre los subgrupos (Tabla 1). Los efectos de todas estas variables de confusión podrían enmascarar otras posibles disparidades entre los subgrupos de la cohorte, asociados a la progresión de la infección por VIH. Sin embargo, muchas de estas variables no pueden ser eliminadas ni controladas, sobre todo en una cohorte conformada por individuos que, por su estado de salud e historia clínica, requerían la obtención de una muestra de LBA para el diagnóstico clínico mediante un procedimiento altamente invasivo.

Posiblemente debido a todos los factores comentados previamente, se observó una gran variación interindividual en la proporción de secuencias identificadas y clasificadas en los lavados orales. Ejemplo de ello es que, aunque en la mayoría de las muestras orales un alto porcentaje del total de las secuencias no fueron asociadas con organismos conocidos, en 4 de ellas las secuencias virales fueron las más abundantes (Figura 2A). Por el contrario, en los LBAs las variaciones entre individuos fueron menos notorias (Figura 2B). Esto podría deberse a que la cavidad oral está más expuesta al ambiente y, por lo tanto, los factores ambientales y los hábitos de higiene podrían tener una mayor influencia en la composición del microbioma de este sitio anatómico. Asimismo, la gran cantidad de secuencias obtenidas de las muestras orales cuyo origen no pudo ser reconocido, podrían corresponder a secuencias virales, sobre todo bacteriófagos, que aún no han sido identificados. Estudios previos han estimado que la concentración de partículas virales en la saliva es de 10^8 partículas/mL y éstas pertenecen mayoritariamente a bacteriófagos (Pride *et al.*, 2012). Además, en la cuarta parte de los lavados orales incluidos en nuestro estudio, más de la mitad de las secuencias virales correspondían a fagos (Figura 3A). Aunado a esto, las bases de datos de secuencias virales no incluyen la totalidad de la diversidad viral ya que muchos virus, en particular fagos, aún no han sido identificados (Paez-Espino *et al.*, 2016). Por estas razones, es posible que una parte de las secuencias de origen desconocido correspondan a virus no identificados, principalmente a fagos.

En cuanto a los LBAs, en la mayoría de las muestras más del 70% del total de secuencias eran humanas (Figura 2B). Estos resultados hacen evidente que la proporción de secuencias del hospedero y, por lo tanto, de lo correspondiente al microbioma, depende en gran medida del tipo de muestra y del compartimiento anatómico del que ésta proceda. Los LBAs contienen una gran cantidad de células humanas, el número aproximado es de 110 células/ μ L y, en las personas sanas no fumadoras, la proporción de células encontradas es la siguiente: más del 85% corresponde a macrófagos alveolares, entre 10% y 15% a linfocitos, \leq 5% a células epiteliales, \leq 3% a neutrófilos y \leq 1% a eosinófilos (Meyer *et al.*, 2012). Además, es necesario tomar en cuenta que el genoma de una célula humana diploide (6.4×10^9 pb) es aproximadamente 1.5 millones de veces más grande que el genoma de un virus promedio (10^4 pb); por ello, la presencia de algunos genomas celulares residuales en las muestras, después del enriquecimiento en partículas virales, tiene importantes repercusiones en la cantidad de secuencias humanas identificadas. Asimismo,

la alta variación en la proporción de secuencias humanas obtenidas en distintos tipos de muestra, ha sido previamente reportado en estudios metagenómicos (Naccache *et al.*, 2014). Resultados similares fueron obtenidos por el grupo de investigación del CIENI, al secuenciar muestras de heces (n> 100) y tejido. En las muestras fecales menos del 1% correspondieron a secuencias humanas, mientras que en las de origen tisular más del 90% de las secuencias provinieron del hospedero.

En lo que respecta a los virus humanos, observamos una mayor prevalencia de papillomavirus (VPH) y citomegalovirus (CMV) en las muestras orales de personas más inmunocomprometidas en comparación con aquellas VIH negativas (Figuras 6A y 6B). Además, en los LBAS se encontró que CMV, anellovirus y herpesvirus fueron más prevalentes en el grupo de individuos con mayor grado de inmunodeficiencia (Figuras 6C, 6E y 6F). Asimismo, la identificación de virus humanos (sin considerar al VIH) fue más frecuente en las muestras broncoalveolares de personas VIH positivas en comparación con aquellas no infectadas por este virus (Figura 6D). Sin embargo, a excepción de los papillomavirus en los lavados orales, no detectamos diferencias significativas en el número de secuencias virales entre los distintos grupos, lo cual refleja la ausencia de diferencias en la abundancia viral. Con base en estos resultados, consideramos que una de las posibles razones por las que no se observó un incremento significativo en las secuencias virales en las personas con mayor grado de inmunodeficiencia, es que el número de individuos en cada subgrupo no es suficiente para contrarrestar el efecto de la variación interindividual y de las variables de confusión. Estudios previos encontraron diferencias en la composición del viroma de varios compartimientos anatómicos, entre personas con diferentes grados de inmunodeficiencia, cuando los grupos estaban conformados por al menos 12 individuos (Young *et al.*, 2015; Monaco *et al.*, 2016) o cuando la comparación se realizó entre grupos con características extremas (<20 células CD4+/ μ L y >700 células CD4+/ μ L) (Li *et al.*, 2013). Debido a ello, una de las principales limitantes de este estudio es el pequeño número de muestras de cada subgrupo, por lo que se planea continuar el reclutamiento de pacientes, en este protocolo vigente, hasta contar con 25 individuos en cada subgrupo.

A pesar de esta limitación, el hallazgo de una mayor prevalencia de virus pertenecientes a las familias *Papillomaviridae*, *Anelloviridae* y *Herpesviridae* en el grupo de personas más inmunocomprometidas, en comparación con aquellas VIH negativas, sugiere que la inmunodeficiencia asociada al VIH podría estar relacionada con la expansión de algunos virus en el tracto respiratorio. Esto podría deberse a la reactivación de virus en estado de latencia, al incremento en la susceptibilidad a nuevas infecciones, a la reducción del control inmunológico de la replicación viral y a una duración más prolongada de las infecciones virales.

Por otra parte, encontramos que los virus más frecuentemente identificados en ambos tipos de muestras respiratorias fueron los anellovirus (OR: 83.9%, LBA: 64.5%) (Figura 4). Los cuales pertenecen a la familia *Anelloviridae*, recientemente descubierta (Feer *et al.*, 2018). Esta familia viral posee 4 géneros distintos: torque teno virus (*Alphatoquevirus*), torque teno midi virus

(*Betatorquevirus*), torque teno mini virus (*Gammatorquevirus*) y *Gyrovirus* y se caracteriza por tener una diversidad genética extremadamente alta, ser adquiridos poco después del nacimiento y establecer infecciones persistentes (Spandole *et al.*, 2015; Feer *et al.*, 2018). Además, llama la atención su elevada prevalencia (~90%) y su aparente pan-tropismo, pues se ha identificado en muestras de saliva, orina, sudor, lágrimas, secreciones nasales, heces, hisopados faríngeos, hígado, bilis, tejido gástrico, secreción cervical, semen, pelo, piel, médula ósea, linfa, ganglios, músculos, sangre del cordón umbilical, tiroides, pulmones, bazo, páncreas, riñones, líquido cefalorraquídeo, tejido nervioso y células mononucleares de sangre periférica (PMBCs) (Spandole *et al.*, 2015); sin embargo, su posible efecto patogénico no ha sido determinado (Spandole *et al.*, 2015; Feer *et al.*, 2018). Así, la alta prevalencia de este tipo de virus en nuestra cohorte, concuerda con la previamente reportada. Además, en muchas de las muestras, se identificaron varios géneros de anellovirus (datos no mostrados). Esto coincide con el hallazgo de diversos tipos de anellovirus en el tracto respiratorio de individuos que habían recibido trasplante de pulmón (Young *et al.*, 2015). En el trabajo presentado aquí, observamos que los anellovirus fueron más comúnmente encontrados en las muestras del tracto respiratorio inferior de personas VIH positivas con conteos de CD4 menores a 100 células/ μ L de sangre, en comparación con individuos sin VIH (Figura 6E). Este resultado es consistente con los de otros grupos de investigación en los que se identificó un incremento significativo en la abundancia de anellovirus en muestras respiratorias y plasmáticas de pacientes con mayor grado de inmunosupresión (Li *et al.*, 2013; De Vlaminc *et al.*, 2013; Young *et al.*, 2015). En conjunto, estos resultados apoyan la idea de que la carga viral de anellovirus podría ser usada como un marcador del estado general de inmunosupresión (De Vlaminc *et al.*, 2013).

Por otro lado, los virus pertenecientes a la familia *Papillomaviridae*, también fueron altamente prevalentes en nuestra cohorte (OR: 80.6%, LBA: 29%) (Figura 4). Esta familia viral comprende virus de DNA no envueltos que infectan células epiteliales (Knipe *et al.*, 2013) incluyendo las de la base de la lengua, la garganta y las amígdalas (Raymond *et al.*, 2018). Estos virus pueden inducir lesiones benignas en la piel y en las mucosas (Knipe *et al.*, 2013), aunque algunos tipos de VPH de alto riesgo, particularmente el VPH-16, han sido fuertemente asociados con la patogénesis del carcinoma de células escamosas orofaríngeas (Panwar *et al.*, 2014). La prevalencia de VPH oral encontrada en nuestra cohorte, es considerablemente más alta a la reportada en varias cohortes de personas sanas. Por ejemplo, cuando Kreimer y colaboradores realizaron una revisión sistemática en la que incluyeron 18 estudios publicados entre 1997 y 2009, encontraron que la prevalencia de VPH oral era de 4.5% en 4,070 individuos (Kreimer *et al.*, 2010). Una prevalencia similar fue reportada en 5,579 hombres y mujeres estadounidenses de entre 14 y 69 años. La prevalencia general de VPH en lavados orales fue de 6.9%, aunque en los hombres fue significativamente mayor (10.1%) que en las mujeres (3.6%) (Gillison *et al.*, 2012). Por otro lado, se identificó VPH en el 23.2% de una cohorte brasileña (Tristão *et al.*, 2012). Estas diferencias sugieren que la prevalencia de este virus podría ser diferente en distintas poblaciones y, en el caso de nuestra cohorte, es necesario considerar la influencia de otras variables. Una de las más importantes es la infección por VIH, ya que estudios previos han mostrado que la prevalencia de

VPH oral es mayor en personas VIH positivas. Por ejemplo, Fatahzadeh y su grupo de trabajo reportaron una prevalencia del 87% en una cohorte de individuos de entre 50 y 69 años de edad infectados con VIH (Fatahzadeh *et al.*, 2013). Este resultado es muy similar a lo encontrado en los lavados orales de las personas VIH positivas incluidas en este trabajo (88.5%). Asimismo, en otro estudio se identificó que la prevalencia de VPH de alto riesgo (tipos 16, 18 y 33) es mayor en personas VIH positivas (13.7%) en comparación con individuos VIH negativos (4.5%); además, los conteos de células T CD4+ menores a 200 células/mL están asociados con la identificación de VPH en biopsias de amígdalas (Kreimer *et al.*, 2004). Este hallazgo coincide con los resultados de nuestro estudio, ya que los virus pertenecientes a la familia *Papillomaviridae* fueron más prevalentes en personas VIH positivas con menos de 100 células CD4+/ μ L (89.47%), que en el grupo de individuos sin VIH (40.0%) (Figura 6A). Más aún, esta familia viral fue la única que tuvo una abundancia significativamente mayor en los lavados orales de personas más inmunocomprometidas, en comparación con aquéllas no infectadas por VIH (Figura 7A). El aumento en la abundancia de VPH sugiere que existe una mayor replicación viral debido a la reducción del control inmunológico provocada por la inmunosupresión asociada al VIH. Por otra parte, los análisis filogenéticos de este tipo de virus indicaron que las secuencias con una alta cobertura ($\geq 89\%$) del genoma viral, pertenecen a los géneros *Betapapillomavirus* y *Gammapapillomavirus*, los cuales incluyen muchos de los papillomavirus considerados comensales (Knipe *et al.*, 2013) y no se ha encontrado una asociación entre su identificación en la cavidad oral y las prácticas sexuales (Fatahzadeh *et al.*, 2013; Nunes *et al.*, 2016), lo que sugiere otras formas de transmisión. Por último, resulta de suma importancia la identificación de un nuevo tipo de *Gammapapillomavirus* debido a las posibles nuevas asociaciones entre algunos tipos de VPH y el desarrollo y progresión de ciertas enfermedades, sobre todo en personas con inmunosupresión.

Finalmente, debido a la gran cantidad de factores de confusión mencionados en párrafos anteriores, consideramos relevante identificar mediante un análisis multivariante, el posible efecto de algunas variables clínicas y epidemiológicas en la composición del microbioma respiratorio de los individuos de nuestra cohorte. En el caso de los lavados orales, se observó que el tabaquismo está relacionado con una menor cantidad de secuencias bacterianas (Figura 8). Estos resultados concuerdan con que el hábito tabáquico modifica el microambiente oral y favorece la alteración de varios componentes del microbioma de este compartimiento anatómico. Las diferencias en el bacterioma de personas que fuman y aquéllas que no lo hacen, han sido reportadas previamente (Wu *et al.*, 2016). De manera interesante, se ha encontrado que el tabaquismo genera un ambiente favorable para las bacterias anaerobias estrictas y facultativas, y limita la propagación de bacterias aerobias, debido a que provoca una disminución en la disponibilidad de oxígeno (Wu *et al.*, 2016). Por otro lado, estas diferencias no fueron encontradas en los LBAs. Este resultado coincide con lo encontrado en un estudio previo en el que se comparó el bacterioma de muestras orales y LBAs de 64 individuos fumadores (n=19) y no fumadores (n=45) (Morris *et al.*, 2013). Los datos obtenidos indicaron que no existen diferencias significativas en el bacterioma del tracto respiratorio inferior asociadas al hábito tabáquico, por lo que la influencia

de este factor podría ser menos importante en las partes distales del tracto respiratorio (Morris *et al.*, 2013).

8. CONCLUSIONES

La metodología estandarizada para este proyecto permitió la identificación de virus con características muy distintas, mediante el uso de Secuenciación de Siguiete Generación, en muestras del tracto respiratorio superior e inferior. Para ello, se reunió una cohorte de 31 individuos, de entre 18 y 50 años, con síntomas de infección respiratoria que incluyó 5 personas VIH negativas y 26 infectadas por VIH y con diferentes grados de inmunodeficiencia. Nuestra cohorte tuvo un predominio de hombres jóvenes y la mayoría de los individuos VIH positivos, habían llegado a la etapa de SIDA. Estas características son propias de la población infectada por VIH en nuestro país y además, está relacionada con los criterios de inclusión establecidos para la participación en este estudio. Por otro lado, no se encontraron diferencias significativas entre los distintos subgrupos de la cohorte, en las características recabadas, que pudieran influir en la composición de viroma respiratorio.

Una de las principales ventajas de la metodología empleada en este trabajo es que, a diferencia de otros estudios del viroma, fue planeada y optimizada para la identificación de virus de DNA y RNA, lo cual es necesario para la caracterización completa de las comunidades virales del tracto respiratorio. Además, dicha metodología mostró resultados concordantes con los obtenidos mediante qPCR; con ello se aseguró la validación de los métodos experimentales y de la estrategia de análisis de datos.

Por otra parte, la proporción de secuencias clasificadas como de origen viral fue muy baja en la mayoría de las muestras tanto orales como broncoalveolares. Esto ha sido previamente reportado en otros estudios y está relacionado con las diferencias en el tamaño del genoma del hospedero y otros componentes del microbioma, en comparación con los genomas virales. Además, es necesario considerar que las bases de datos de secuencias virales no incluyen la totalidad de la diversidad viral debido a que muchos virus aún no han sido identificados.

En cuanto a los virus humanos, se identificaron anellovirus, distintos tipos de herpesvirus, papillomavirus, pegivirus, poliomavirus, parvovirus, VIH y virus que son ampliamente reconocidos por ser patógenos respiratorios como rinovirus, metapneumovirus, mastadenovirus, coronavirus, enterovirus, adenovirus y virus de la influenza. Además, se encontró una mayor diversidad y cantidad de virus en las muestras orales en comparación con los LBAs; estas diferencias fueron significativas para los virus humanos totales y los pertenecientes a las familias virales *Herpesviridae* y *Papillomaviridae*.

Asimismo, a pesar de que sólo se encontraron diferencias significativas en la abundancia de papillomavirus en los lavados orales de personas VIH positivas con mayor grado de inmunodeficiencia, comparado con el grupo de pacientes no infectados por VIH; las diferencias en la prevalencia de varios virus o familias virales entre distintos subgrupos de la cohorte, sugiere

que la infección por VIH podría estar asociada con la expansión del viroma respiratorio superior e inferior. Además, de manera interesante, se obtuvieron varios genomas completos o casi completos de algunos virus como VPH, virus de la influenza, parvovirus y poliomavirus humano, lo cual podría indicar altos niveles de replicación y, por lo tanto, una alta carga viral de estos patógenos en el tracto respiratorio de personas VIH positivas. No obstante, es necesario analizar más muestras de personas VIH positivas e individuos sin VIH para poder describir el efecto de la inmunodeficiencia asociada a este virus en la composición del viroma respiratorio y su contribución al desarrollo de enfermedades respiratorias.

Por otro lado, resulta de suma relevancia la identificación y secuenciación de casi la totalidad del genoma de un nuevo tipo de *Gammapapillomavirus*.

Finalmente, a pesar de que este estudio tiene limitaciones, es uno de los primeros cuyo objetivo es la caracterización del componente viral del microbioma respiratorio y el primero en enfocarse en personas VIH positivas, por lo que sienta las bases para estudios posteriores y aporta información valiosa al conocimiento del viroma humano.

9. LITERATURA CITADA

1. Abbas, A.A, Diamond, J.M., Chehoud, C., Chang, B., Kotzin, J.J., Young, J.C., ... Collman R.G. (2017). The Perioperative Lung Transplant Virome: Torque Teno Viruses Are Elevated in Donor Lungs and Show Divergent Dynamics in Primary Graft Dysfunction. *American Journal of Transplantation*, 17(5), 1313-1324. <http://doi.org/10.1111/ajt.14076>
2. Abeles, S. R., Robles-Sikisaka, R., Ly, M., Lum, A. G., Salzman, J., Boehm, T. K., & Pride, D. T. (2014). Human oral viruses are personal, persistent and gender-consistent. *The ISME Journal*, 8(9): 1753–1767. <http://doi.org/10.1038/ismej.2014.31>
3. Agencourt AMPure XP. Information for Use Guide PCR Purification. August 2013. Beckman Coulter, Inc.
4. Agilent 2100 Bioanalyzer. 2100 Expert User's Guide. May 2005. Agilent Technologies, Inc. 2000-2005.
5. Buermans, H.P.J., & den Dunnen, J.T. (2014). Next generation sequencing technology: Advances and applications. *Biochimica et Biophysica Acta, (BBA) - Molecular Basis of Disease*. 1842, 1932–1941. <https://doi.org/10.1016/j.bbadis.2014.06.015>
6. Cadwell, K. (2015). The virome in host health and disease. *Immunity*, 42(5), 805–813. <http://doi.org/10.1016/j.immuni.2015.05.003>
7. Centro Nacional para la Prevención y Control del VIH y el SIDA. (2017). Vigilancia Epidemiológica de casos de VIH/SIDA en México Registro Nacional de Casos de SIDA Actualización al Cierre del 2016.
8. Crabtree-Ramírez, B., Caro-Vega, Y., Shepherd, B. E., Wehbe, F., Cesar, C., Cortés, C., ... on behalf of the CCASAnet Team. (2011). Cross-Sectional Analysis of Late HAART Initiation in Latin America and the Caribbean: Late Testers and Late Presenters. *PLoS ONE*, 6(5), e20272. <http://doi.org/10.1371/journal.pone.0020272>
9. Datta, S., Budhauriya, R., Das, B., Chatterjee, S., Vanlalhmuaaka, & Veer, V. (2015). Next-generation sequencing in clinical virology: Discovery of new viruses. *World Journal of Virology*, 4(3), 265–276. <http://doi.org/10.5501/wjv.v4.i3.265>
10. De Vlamincq, I., Khush, K. K., Strehl, C., Kohli, B., Neff, N. F., Okamoto, J., ... Quake, S. R. (2013). Temporal Response of the Human Virome to Immunosuppression and Antiviral Therapy. *Cell*, 155(5), 1178–1187. <http://doi.org/10.1016/j.cell.2013.10.034>
11. Deeks, S. G., Overbaugh, J., Phillips, A., & Buchbinder, S. (2015). HIV infection. *Nature Reviews Disease Primers*, 1: 15035. <http://doi.org/10.1038/nrdp.2015.35>
12. Delwart, E. (2013). A Roadmap to the Human Virome. *PLoS Pathogens*, 9(2), e1003146. <http://doi.org/10.1371/journal.ppat.1003146>
13. Dickson, R. P., Erb-Downward, J. R., Martinez, F. J., & Huffnagle, G. B. (2016). The Microbiome and the Respiratory Tract. *Annual Review of Physiology*, 78, 481–504. <http://doi.org/10.1146/annurev-physiol-021115-105238>
14. Fatahzadeh, M., Schlecht, N. F., Chen, Z., Bottalico, D., McKinney, S., Ostolozza, J., ... Burk, R. D. (2013). Oral Human Papillomavirus Detection in Older Adults Who Have Human Immunodeficiency Virus Infection. *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*, 115(4), 505–514. <http://doi.org/10.1016/j.oooo.2012.11.004>
15. Freer, G., Maggi, F., Pifferi, M., Di Cicco, M. E., Peroni, D. G., & Pistello, M. (2018). The Virome and Its Major Component, Anellovirus, a Convolutated System Molding Human Immune Defenses and Possibly Affecting the Development of Asthma and Respiratory Diseases in Childhood. *Frontiers in Microbiology*, 9, 686. <http://doi.org/10.3389/fmicb.2018.00686>
16. Gillison, M. L., Broutian, T., Pickard, R. K. L., Tong, Z., Xiao, W., Kahle, L., ... Chaturvedi, A. K. (2012). Prevalence of Oral HPV Infection in the United States, 2009–2010. *JAMA*, 307(7), 693–703. <http://doi.org/10.1001/jama.2012.101>
17. Knipe, D.M., & Howley, P. (Ed.). (2013). *Fields Virology*. Philadelphia, USA: Wolters Kluwer Health/Lippincott Williams & Wilkins

18. Koziel, H., Kim, S., Reardon, C., Li, X., Garland, R., Pinkston, P., & Kornfeld H. (1999). Enhanced *in vivo* human immunodeficiency virus-1 replication in the lungs of human immunodeficiency virus-infected persons with *Pneumocystis carinii* pneumonia. *American Journal of Respiratory and Critical Care Medicine*, 160 (6): 2048–55. <https://doi.org/10.1164/ajrccm.160.6.9902099>
19. Kreimer, A.R., Alberg, A.J., Daniel, R., Gravitt, P.E., Viscidi, R., Garrett, E.S., ... Gillison, M.L. (2004). Oral Human Papillomavirus Infection in Adults Is Associated with Sexual Behavior and HIV Serostatus. *The Journal of Infectious Diseases*, 189(4):686-98. <https://doi.org/10.1086/381504>
20. Kreimer, A.R., Bhatia, R.K., Messegue, A.L., González, P., Herrero, R., & Giuliano, A.R. (2010). Oral Human Papillomavirus in Healthy Individuals: A Systematic Review of the Literature. *Sexually Transmitted Diseases*, 37(6):386-91. <http://doi.org/10.1097/OLQ.0b013e3181c94a3b>
21. Lederberg, J., & McCray, A.T. (2001). 'Ome Sweet 'Omics - A Genealogical Treasury of Words Genealogical Treasury of Words. *Scientist*;15(7):8.
22. Legoff, J., Resche-Rigon, M., Bouquet, J., Robin, M, Naccache, S.N., Mercier-Delarue, S., ..., Chiu C.Y. (2017). The eukaryotic gut virome in hematopoietic stem cell transplantation: new clues in enteric graft-versus-host disease. *Nature Medicine*, 23(9):1080-1085. <https://doi.org/10.1038/nm.4380>
23. Li, L., Deng, X., Linsuwanon, P., Bangsberg, D., Bwana, M. B., Hunt, P., ... Delwart, E. (2013). AIDS Alters the Commensal Plasma Virome. *Journal of Virology*, 87(19), 10912–10915. <http://doi.org/10.1128/JVI.01839-13>
24. Martin M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal*, 17(1):10–12.
25. Meyer, K.C., Raghu, G., Baughman, R.P., Brown, K.K., Costabel, U., du Bois R.M., ... Wood, B. (2012). An official American Thoracic Society clinical practice guideline: the clinical utility of bronchoalveolar lavage cellular analysis in interstitial lung disease, *American Journal of Respiratory and Critical Care Medicine*, 185(9):1004-14. <http://doi.org/10.1164/rccm.201202-0320ST>
26. Moir, S., Chun, T.W., & Fauci, A.S., (2011). Pathogenic mechanisms of HIV disease. *Annual Review of Pathology: Mechanisms of Disease*, 6:223-48. <https://doi.org/10.1146/annurev-pathol-011110-130254>
27. Mokili, J. L., Rohwer, F., & Dutilh, B. E. (2012). Metagenomics and future perspectives in virus discovery. *Current Opinion in Virology*, 2 (1), 63-77. <https://doi.org/10.1016/j.coviro.2011.12.004>.
28. Monaco, C. L., Gootenberg, D. B., Zhao, G., Handley, S. A., Ghebremichael, M. S., Lim, E. S., ... Virgin, H. W. (2016). Altered Virome and Bacterial Microbiome in Human Immunodeficiency Virus-Associated Acquired Immunodeficiency Syndrome. *Cell Host & Microbe*, 19(3), 311–322. <http://doi.org/10.1016/j.chom.2016.02.011>
29. Morgan, X.C., Tickle, T.L., Sokol, H., Gevers, D., Devaney, K.L., Ward, D.V., ... Huttenhower, C. (2012). Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol.* Apr 16;13(9): R79.
30. Naccache, S. N., Federman, S., Veeraraghavan, N., Zaharia, M., Lee, D., Samayoa, E., ... Chiu, C. Y. (2014). A cloud-compatible bioinformatics pipeline for ultrarapid pathogen identification from next-generation sequencing of clinical samples. *Genome Research*, 24(7), 1180–1192. <http://doi.org/10.1101/gr.171934.113>
31. Nextera® XT DNA Sample Preparation Guide, 2012
32. Nunes, E. M., Sudenga, S. L., Gheit, T., Tommasino, M., Baggio, M. L., Ferreira, S., ... Slichero, L. (2016). Diversity of Beta-papillomavirus at anogenital and oral anatomic sites of men: The HIM Study. *Virology*, 495, 33–41. <http://doi.org/10.1016/j.virol.2016.04.031>
33. Paez-Espino, D., Eloie-Fadros, E.A., Pavlopoulos, G.A., Thomas, A.D., Huntemann, M., Mikhailova, N. ... Kyrpides, N.C. (2016). Uncovering Earth's viroma. *Nature*, 536(7617):425-30. <https://doi.org/10.1038/nature19094>
34. Panel on Opportunistic Infections in HIV-Infected Adults and Adolescents. (2017). Guidelines for the prevention and treatment of opportunistic infections in HIV-infected adults and adolescents: recommendations from the Centers for Disease Control and Prevention, the National Institutes of Health, and the HIV Medicine Association of the Infectious Diseases Society of America. http://aidsinfo.nih.gov/contentfiles/lvguidelines/adult_oi.pdf.

35. Panwar, A., Batra, R., Lydiatt, W.M., Ganti, A.K. (2014). Human papilloma virus positive oropharyngeal squamous cell carcinoma: A growing epidemic. *Cancer Treatment Reviews*, 40(2):215-9. <https://doi.org/10.1016/j.ctrv.2013.09.006>
36. Pérez-Brocal, V. & Moya, A., (2018). The analysis of the oral DNA virome reveals which viruses are widespread and rare among healthy young adults in Valencia (Spain). *PLoS ONE*, 13(2): e019186. <http://doi.org/10.1371/journal.pone.0191867>
37. Pickrell, W.O., Rees, M.I., & Seo-Kyung Chung. (2012). Next Generation Sequencing Methodologies- An overview. Elsevier, 89, 1-26. <https://doi.org/10.1016/B978-0-12-394287-6.00001-X>
38. Pride, D. T., Salzman, J., Haynes, M., Rohwer, F., Davis-Long, C., White, R. A., ... Relman, D. A. (2012). Evidence of a robust resident bacteriophage population revealed through analysis of the human salivary virome. *The ISME Journal*, 6(5): 915–926. <http://doi.org/10.1038/ismej.2011.169>
39. Raymond, S., Liu, B. & Taioli, E. (2018). Racial differences in the prevalence of oncogenic oral human papillomavirus infection types in the United States adult population. *Head & Neck*, 00:1–9. <https://doi.org/10.1002/hed.25317>
40. Rodríguez M & Rodríguez W. (2006). PCR en tiempo real. Cuernavaca: IBT-UNAM
41. Selik, R.M., Mokotoff, E.D, Branson, B., Owen, S. M., Whitmore, S., & Hall, H.I., (2014). Revised Surveillance Case Definition for HIV Infection — United States, 2014. *Centers for Disease Control & Prevention (CDC)*, 63(3): 1-10. <http://www.jstor.org/stable/24832587>
42. Shukla S. D., Budden K. F., Neal R. & Hansbro P. M. (2017). Microbiome effects on immunity, health and disease in the lung. *Clinical & Translational Immunology*, 6, e133; doi:10.1038/cti.2017.6; published online 10 March 2017
43. Skalski, J. H., & Limper, H. A. (2016). Fungal, Viral, and Parasitic Pneumonias Associated with Human Immunodeficiency Virus. *Seminars in Respiratory and Critical Care Medicine*, 37(02), 257-266. DOI: 10.1055/s-0036-1578802
44. Spandole, S., Cimponeriu, D., Berca, L.M., & Mihăescu G. (2015). Human anelloviruses: an update of molecular, epidemiological and clinical aspects. *Archives of Virology*. 160(4): 893–908. <https://doi.org/10.1007/s00705-015-2363-9>
45. Taboada, B., Espinoza, M. A., Isa, P., Aponte, F. E., Arias-Ortiz, M. A., Monge-Martínez, J., ... Arias, C. F. (2014). Is There Still Room for Novel Viral Pathogens in Pediatric Respiratory Tract Infections? *PLoS ONE*, 9(11), e113570. <http://doi.org/10.1371/journal.pone.0113570>
46. Tristão, W., Ribeiro, R.M., Oliveira, C.A., Betiol, J.C., & Bettini, J de S. (2012). Epidemiological study of HPV in oral mucosa through PCR. *Brazilian Journal of Otorhinolaryngology*, 78(4):66-70.
47. Twigg, H.L., Weiden, M., Valentine, F., Schnizlein-Bick, C. T., Bassett, R., Zheng, L., ... Bucy, R.P. (2008). Effect of Highly Active Antiretroviral Therapy on Viral Burden in the Lungs of HIV-Infected Subjects. *The Journal of Infectious Diseases*, 197(1):109–116. <https://doi.org/10.1086/523766>
48. Virgin, H. W. (2014). The virome in mammalian physiology and disease. *Cell*, 157(1), 142–150. <http://doi.org/10.1016/j.cell.2014.02.032>
49. Xu, H., Luo, X., Qian, J., Pang, X., Song, J., Qian, G., ... Chen, S. (2012). FastUniq: A Fast De Novo Duplicates Removal Tool for Paired Short Reads. *PLoS ONE*, 7(12), e52249. <http://doi.org/10.1371/journal.pone.0052249>
50. Young, J. C., Chehoud, C., Bittinger, K., Bailey, A., Diamond, J. M., Cantu, E., ... Collman, R. G. (2015). Viral metagenomics reveal blooms of anelloviruses in the respiratory tract of lung transplant recipients. *American Journal of Transplantation: Official Journal of the American Society of Transplantation and the American Society of Transplant Surgeons*, 15(1): 200–209. <http://doi.org/10.1111/ajt.13031>
51. Zárata, S., Taboada, B., Yocupicio-Monroy, M., & Arias, C.F. (2017). Human virome. *Archives of Medical Research*, 48(8):701-716. <https://doi.org/10.1016/j.arcmed.2018.01.005>