



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**  
**PROGRAMA DE MAESTRÍA Y DOCTORADO EN CIENCIAS BIOQUÍMICAS**  
**INSTITUTO DE BIOTECNOLOGÍA**

**METAGENÓMICA VIRAL Y BACTERIANA DE**  
**ENFERMEDADES RESPIRATORIAS GRAVES**

# **T E S I S**

QUE PARA OPTAR POR EL GRADO DE:

**MAESTRO EN CIENCIAS**

P R E S E N T A

L.C.G. Ana Georgina Cobián Güemes

DIRECTOR DE TESIS:

Dr. Carlos Federico Arias Ortiz  
Instituto de Biotecnología, UNAM

INTEGRANTES DE COMITÉ TUTOR:

Dr. Esperanza Martínez Romero  
Centro de Ciencias Genómicas, UNAM

Dr. Enrique Merino Pérez  
Instituto de Biotecnología, UNAM

**CUERNAVACA, MORELOS**

**Marzo 2013**

## Integrantes de jurado de examen

Presidente

Dr. Katy Juárez López                      Instituto de Biotecnología, UNAM

Secretario

Dr. Rosa Gutiérrez Ríos                      Instituto de Biotecnología, UNAM

Vocal

Dr. Julio Collado Vides                      Centro de Ciencias Genómicas, UNAM

Vocal

Dr. Guillermo Gosset Lagarda                      Instituto de Biotecnología, UNAM

Vocal

Dr. Ramón González García-Conde                      Facultad de Ciencias, UAEM

**Este trabajo está dedicado**

**A mis papás**

**A Adrián**

**A mis abuelitos, primos y tíos.**

**Al Dr. Carlos Arias**

**Por confiar en mí y apoyarme siempre.**

## **AGRADECIMIENTOS**

Este trabajo fue realizado en el laboratorio de los doctores Carlos Arias y Susana López ubicado en el Instituto de Biotecnología, campus Morelos de la Universidad Nacional Autónoma de México con recursos aportados por el Consejo Nacional de Ciencia y Tecnología (CONACyT) y por el Instituto de Ciencia y Tecnología del Distrito Federal (ICyTDF)

Mi sincero agradecimiento al Dr. Carlos Arias por dirigir este proyecto con entusiasmo y darme la oportunidad de conocer y trabajar en el fascinante mundo de la virología. Por brindarme enseñanzas y consejos tanto académicos como de vida. Por apoyarme al conseguir todas las herramientas necesarias para la realización de este proyecto, desde muestras clínicas hasta equipo de cómputo. Carlos, gracias por tu amistad y por confiar en mí y mis ideas.

Agradezco al Dr. Pavel Isa por compartir estos años de trabajo y hacerme crecer al plantearme nuevos retos en el análisis de los datos. Por tomarme en cuenta en diversos trabajos sobre influenza. Por procesar las muestras de este trabajo. Por su amistad.

A la Dra. Susana López por estar al pendiente de mí y acogerme como miembro del Grupo Arias-López. Susana, gracias por tus consejos que fueron siempre amables y sencillos.

A los integrantes de mi comité tutor, el Dr. Enrique Merino y la Dra. Esperanza Martínez, agradezco sus valiosos comentarios y sugerencias sobre el rumbo del proyecto. Enrique, gracias por guiarme en el camino bioinformático cuando me sentía perdida. Esperanza gracias por mostrarme ideas nuevas en la forma de ver el proyecto.

A los doctores Katy Juárez, Rosy Gutiérrez, Julio Collado, Guillermo Gosset y Ramón González les agradezco su tiempo y dedicación al revisar mi tesis. Sus comentarios y correcciones enriquecieron mucho este trabajo.

A todos los miembros del Grupo Arias-López, gracias por compartir estos años de trabajo y amistad. Al Dr. Ernesto Méndez, en paz descanse, por sus comentarios sobre quasiespecies y taxonomía viral. Especialmente agradezco a MariLo Soto del Río y Marina Escalera porque formamos un equipo de trabajo maravilloso; a mis compañeros de mesa Lili, Diego, Luis Paulin y Enrique por su compañerismo; y a Blanca Taboada por sus sugerencias.



Grupo Arias-López, 2010. IBT. Cuernavaca Morelos

Agradezco a las siguientes personas que amablemente me apoyaron en este proyecto:

A Adrián Cantú por escucharme y ayudarme cuando los problemas de programación se complicaban. A Patrick Louchouart por asesorarme en la terminología médica y fisiología del aparato respiratorio humano. A Leonardo Collado por acompañarme en el aprendizaje del manejo de datos de secuenciación masiva.

A la Unidad Universitaria de Secuenciación Masiva, Verónica Jiménez y Ricardo Grande. A Jérôme Verleyen por ayudarme a conocer y trabajar en el cluster, siempre amable, servicial y amigable. A la Unidad Universitaria de Apoyo Bioinformático: Alejandro y Karel. A los miembros de la unidad de docencia, al Dr. Enrique Rudiño, Jalil, Toño y Gloria. A Arely y Cruz. A la comunidad del IBT por hacer de este lugar de trabajo un ambiente agradable y de compañerismo. A mis compañeros de generación de maestría.

Al personal del Instituto Nacional de Enfermedades Respiratorias por recolectar y proporcionarnos las muestras clínicas por medio del Dr. Francisco Quiñonez.



# ÍNDICE

Integrantes de jurado de examen.....	1
<b>AGRADECIMIENTOS</b> .....	2
<b>ÍNDICE</b> .....	4
Índice de Figuras.....	6
Índice de Tablas.....	8
Índice de Material Suplementario.....	8
<b>RESUMEN</b> .....	9
<b>INTRODUCCIÓN</b> .....	10
<b>Metagenómica</b> .....	10
<i>Perfiles 16S de rRNA</i> .....	10
<i>Técnicas de secuenciación masiva de DNA</i> .....	12
<b>Metagenómica viral</b> .....	14
<i>Retos en la exploración de la diversidad viral en enfermedades humanas</i> .....	14
<i>Aplicaciones de la secuenciación masiva en virología clínica</i> .....	16
<i>Descubrimiento de virus en animales mediante NGS</i> .....	17
<i>Descubrimiento de virus en enfermedades humanas mediante NGS</i> .....	18
<b>El microbioma pulmonar humano</b> .....	19
<i>Características anatómicas e inmunológicas de los pulmones</i> .....	19
<i>Métodos de toma de muestras pulmonares</i> .....	20
<i>Enfermedades del tracto respiratorio bajo y su impacto en el microbioma pulmonar</i> .....	20
<b>Influenza</b> .....	21
<i>Panorama general de la pandemia 2009-2010</i> .....	21
<i>Transmisibilidad</i> .....	22
<i>Orígenes de la pandemia</i> .....	23
<i>Situación en México de mayor mortalidad que en el resto del mundo</i> .....	26
<i>Casos hospitalizados, obtención de muestras</i> .....	26
<i>Neumonía bacteriana como complicación de la enfermedad</i> .....	27
<i>Métodos de diagnóstico de influenza</i> .....	27
<b>HIPÓTESIS</b> .....	28
<b>OBJETIVO</b> .....	28
<b>MATERIALES Y MÉTODOS</b> .....	29
<b>Pacientes</b> .....	29
<b>Diagnóstico molecular de influenza A</b> .....	30
<i>RT-PCR cuantitativo</i> .....	31
<i>Seeplex</i> .....	32
<i>PCR</i> .....	32
<i>Amplificación del genoma completo y secuenciación</i> .....	33
<b>Secuenciación masiva de DNA en la plataforma Illumina GAIIx</b> .....	34
<i>Extracción de ácidos nucleicos</i> .....	34

<i>Preparación de la biblioteca para secuenciación</i> .....	34
<i>Generación de clusters</i> .....	35
<i>Secuenciación mediante síntesis</i> .....	37
<i>Análisis de imágenes</i> .....	37
<b>Procesamiento de las secuencias</b> .....	39
<b>RESULTADOS</b> .....	42
<b>Protocolo Bioinformático A</b> .....	43
<b>Protocolo Bioinformático B</b> .....	47
<b>Protocolo Bioinformático C</b> .....	47
<b>PhiX como línea control</b> .....	51
<b>Identificación del virus de influenza A</b> .....	52
<b>Secuenciación del genoma completo de H1N1pdm2009</b> .....	55
<b>Análisis de las secuencias de os virus de influenza A</b> .....	57
<b>Resumen de diagnóstico molecular de influenza A</b> .....	58
<b>Rinovirus</b> .....	59
<b>Bacterias</b> .....	62
Abundancia de especies bacterianas.....	62
Diversidad bacteriana .....	68
Muestra 8883, bacteria predominante <i>Haemophilus influenzae</i> .....	70
Muestra 8836, bacteria predominante <i>Pseudomonas aeruginosa</i> .....	70
Bacterias patógenas.....	72
<b>DISCUSIÓN</b> .....	74
<b>Sobre la metodología</b> .....	74
<b>Sobre la obtención y preparación de muestras</b> .....	75
<b>Sobre los protocolos bioinformáticos</b> .....	76
<b>PhiX</b> .....	78
<b>Influenza</b> .....	79
<b>Rinovirus</b> .....	83
<b>Diversidad viral</b> .....	84
<b>Diversidad bacteriana</b> .....	88
<b>Bacterias patógenas</b> .....	94
<b>PERSPECTIVAS</b> .....	95
<b>Identificación de virus desconocidos</b> .....	95
<b>Aplicación de este método en diagnóstico</b> .....	95
<b>BIBLIOGRAFÍA</b> .....	96

## **Índice de Figuras**

Figura 1 Regiones hipervariables de la subunidad ribosomal 16S

Figura 2: Casos reportados de influenza en 2009 en México

Figura 3: Valores de número de reproducción estimado para las pandemias de influenza.

Figura 4: Origen del virus influenza A H1N1 2009

Figura 5: Procedimiento de secuenciación en GAIIx

Figura 6: Preparación de biblioteca para secuenciación

Figura 7: Generación de clusters

Figura 8: Secuenciación mediante síntesis

Figura 9: Proceso de análisis de imágenes.

Figura 10: Calidad Phred

Figura 11: Diagrama de flujo de los protocolos bioinformáticos

Figura 12: Filtrado de secuencias

Figura 13: Proceso de paralización de la búsqueda vs NT

Figura 14: Visualización de virus en la muestra 9140 en MEGAN

Figura 15. Visualización en MEGAN del resultado del protocolo A en la muestra PhiX

Figura 16. Línea control PhiX, MEGAN 100 secuencias

Figura 17: Distribución longitud de alineamiento de lecturas identificadas como influenza

Figura 18: Cobertura errónea del genoma de influenza para la muestra 3974

Figura 19: Cobertura del genoma de influenza en las muestras 6985, 8836 y 7192

Figura 20: Estructura de HA.

Figura 21: Cobertura del genoma de rinovirus humano 73

Figura 22. Contigs resultado del ensamblado de novo mediante Velveth

Figura 23: Filogenia de rinovirus.

Figura 24: Número de secuencias por especie bacteriana para la muestra 9041.

Figura 25: Especies bacterianas con 10 o más secuencias en la muestra 9041

Figura 26: Especies bacterianas identificados en los pacientes fallecidos

Figura 27: Especies bacterianas identificados en los pacientes vivos

Figura 28: Mapa de calor de especies bacterianas

Figura 29: Phyla en todas las muestras

Figura 30: Mapa de calor con datos a nivel de familia

Figura 31: Resultados del pipeline A en la muestra 8883

Figura 32: Bacterias patógenas

Figura 33: Familias bacterianas

## **Índice de Tablas**

Tabla 1: Tecnologías de secuenciación de DNA

Tabla 2: Segmentos del virus de influenza A y proteínas que codifican.

Tabla 3: Datos clínicos de los 14 pacientes participantes en este estudio

Tabla 4: Resultados de diagnóstico molecular de influenza A

Tabla 5: Resultados de la secuenciación

Tabla 6: Secuencias virales, protocolo bioinformático A

Tabla 7: Secuencias virales, protocolo bioinformático B y C

Tabla 8. Resumen de la información sobre el diagnóstico de influenza A en las 14 muestras

Tabla 9: Especies bacterianas predominantes en cada muestra.

Tabla 10: Bacterias patógenas en las muestras

Tabla 11: Proporciones adenovirus vs genoma humano

Tabla 12: Comparación de estudios metagenómicos de vías respiratorias

## **Índice de Material Suplementario**

S1: Medio REMEL

S2: Protocolo diagnóstico influenza A pandémica 2009 CDC

S3: Protocolo extracción de ácidos nucleicos

S4: Protocolo preparación de muestras para secuenciar DNA genómico

S5: Gráficas abundancia de organismos 14 muestras

S6: Visualización en MEGAN de cada muestra

## **RESUMEN**

La metagenómica es la disciplina que estudia a los microorganismos que componen un ambiente particular, en un estudio metagenómico no es necesario cultivar cada organismo para poder estudiarlo. La metagenómica se basa en el estudio del material genético total contenido en un ambiente particular, brindando así la ventaja de poder estudiar la totalidad de organismos presentes sin la necesidad de dirigir la búsqueda hacia un grupo particular.

Las enfermedades respiratorias son una de las principales causas de mortalidad a nivel mundial, por lo que elucidar qué tipo de organismos las causan y cómo interactúa entre ellos y con el hospedero es un asunto de interés tanto científico como práctico.

Durante la epidemia del virus de influenza A H1N1 que ocurrió en el año 2009, en México se reportó una mayor mortalidad que en el resto del mundo, también fueron identificados varios casos graves que culminaron en decesos en los que no fue detectado el virus de influenza A H1N1 2009. Este hecho llevó al planteamiento de varias interrogantes sobre la virulencia del virus circulante en México y la existencia de agentes agravantes del padecimiento. En este trabajo presentamos un enfoque metagenómico para elucidar estas interrogantes.

El enfoque metagenómico de este trabajo consiste en analizar, mediante secuenciación masiva de DNA, el material genético extraído del tracto respiratorio de 14 pacientes que padecieron de síndrome respiratorio severo durante la epidemia de influenza en el año 2009 en México. Mediante este estudio logramos identificar varios elementos tanto virales como bacterianos que pudiesen ser los agentes causales del síndrome respiratorio severo que sufrieron estos pacientes o simplemente ser parte del microbioma pulmonar humano.

Las aportaciones de este trabajo son valiosas tanto en el aspecto bioinformática como en el biológico. Las principales contribuciones científicas generadas por este trabajo son: el desarrollo e implementación de protocolos bioinformáticos para el análisis de datos provenientes de microbioma humano y los resultados biológicos derivados de estos protocolos.

# INTRODUCCIÓN

## Metagenómica

El término metagenómica fue acuñado en 1998 y se refiere al análisis tanto funcional como basado en secuencia de DNA<sup>1</sup> del conjunto de genomas microbianos contenidos en una muestra ambiental<sup>2</sup>.

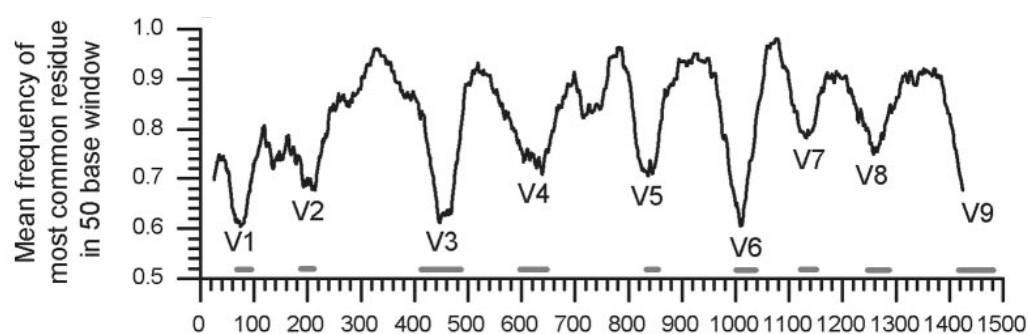
La metagenómica es la disciplina que permite el estudio genómico de comunidades de organismos sin la necesidad de cultivarlos, esto mediante el análisis de las secuencias totales de DNA extraído directamente del ambiente a estudiar<sup>3</sup>. De esta manera no existen perturbaciones en la adaptación de los organismos a condiciones de cultivo en un laboratorio. Sin embargo, los datos obtenidos, al provenir de comunidades microbianas heterogéneas y debido a las técnicas de secuenciación masiva per se, son complejos y parciales.

Previo al surgimiento de la metagenómica, los estudios sobre diversidad microbiana estaban basados en la información genómica y morfológica obtenida de cultivos clonales de cada organismo. Se estima que mas del 99% de los microorganismo no son cultivables<sup>4</sup>. Los primeros análisis metagenómicos revelaron la existencia de muchos más microorganismos de los que pueden ser cultivados<sup>5</sup>, abriendo así una ventana a la exploración de la diversidad microbiana mediante métodos independientes de cultivo y redefiniendo la noción del número de especies existentes, sobre todo, bacterianas.

### *Perfiles de 16S rRNA*

Los primeros trabajos en metagenómica bacteriana se basaron en el análisis de las regiones hipervariables que se encuentran en la subunidad ribosomal 16S<sup>6</sup>. El primer trabajo en el que fueron tomadas muestras directamente del ambiente y sus regiones hipervariables 16S fueron estudiadas es el realizado en el año 1991 sobre una comunidad marina de picoplancton<sup>7</sup>. En esta comunidad descubrieron dos grupos bacterianos filogenéticamente distintos a los que habían sido descritos previamente así como varias proteobacterias previamente descritas.

Los ribosomas están compuestos por tres macromoléculas de RNA ribosomal (rRNA) acopladas a varias proteínas, todo esto organizado en dos subunidades<sup>8</sup>. Los ribosoma son una de las moléculas biológicas que poseen mayor grado de conservación evolutiva ya que su función es esencial para toda célula, esta característica ha permitido utilizarlos como marcadores en la construcción de filogenias que pretenden incluir a todos los organismos vivos<sup>9</sup>. En bacterias, la subunidad 16S posee 9 regiones hipervariables en las que existe una divergencia tal en secuencia que permite, al realizar análisis filogenéticos de estas regiones, identificar su origen a nivel de especie, la variabilidad de la subunidad 16S en varias bacterias se ilustra en la Figura 1. Por lo anterior el análisis de las regiones hipervariables de los ribosomas ha sido la técnica clásica utilizada en metagenómica bacteriana.



**Figura 1:** Regiones hipervariables de la subunidad ribosomal 16S<sup>10</sup>, el eje x muestra la longitud de la subunidad 16S, el eje y muestra la media de las frecuencias de los residuos más comunes en ventanas de 50 nucleótidos. El análisis fue realizado en 4,383 secuencias de cepas tipo de *Escherichia coli*. Las 9 regiones hipervariables se denominan V1,...,V9.

Los ribosomas de arqueas poseen un tamaño y composición similar a los ribosomas bacterianos, poseen también una subunidad 16S y es posible realizar análisis metagenómicos con esta subunidad<sup>11</sup>.

En eucariotes, la subunidad 18S es utilizada para la identificación a nivel de especie; en particular en el caso de hongos, la región transcrita interna espaciadora (ITS) del rRNA permite la discriminación entre especies<sup>12</sup>.



### *Técnicas de secuenciación masiva de DNA*

Una estrategia distinta utilizada en metagenómica es la secuenciación masiva de todo el DNA presente en una muestra. Mediante esta estrategia es posible conocer no sólo qué microorganismos están presentes, sino también qué genes conforman al metagenoma.

Las secuencias que se obtienen mediante secuenciación masiva son cortas, entre 35 y 700 nucleótidos de longitud, e identificar su origen representa un problema bioinformático complejo. A pesar de lo anterior, las técnicas actuales de secuenciación masiva de DNA han permitido realizar estudios metagenómicos exitosos en diversos ecosistemas, incluyendo diferentes tipos de muestras biológicas humanas, tales como sangre, tracto intestinal<sup>13</sup> y tracto respiratorio<sup>14</sup>.

La secuenciación masiva de DNA (NGS, por sus siglas en inglés Next Generation Sequencing) es el procedimiento mediante el cual es posible leer la sucesión de nucleótidos en varios fragmentos diferentes de DNA de manera simultánea. Contrario a la secuenciación mediante Sanger<sup>15</sup> en la que sólo es posible leer un fragmento de DNA a la vez. Las diferencias entre las tecnologías NGS y la secuenciación tipo Sanger son que, en NGS el análisis es masivo y en paralelo, brinda un alto rendimiento y costos reducidos. Aunque mediante NGS la secuenciación de genomas se ha simplificado, los análisis posteriores de los datos y el significado biológico de los mismos son el cuello de botella en la comprensión de los genomas.

Las plataformas de secuenciación de DNA se dividen en primera, segunda y tercera generación; siendo la secuenciación por Sanger la primera generación, la secuenciación basada en fluoróforos la segunda generación y la secuenciación de molécula única sin previa amplificación mediante PCR y en tiempo real es la tercera generación.

Existen varias plataformas para la secuenciación de DNA, los costos y la cantidad de datos obtenidos son variables, en la Tabla 1 se presenta un comparativo. Los costos que se presentan son sólo una referencia ya que existen variaciones dependiendo de la versión y aplicación específica de cada plataforma.

G <sup>n</sup>	Compañía	Instrumento	Principio de secuenciación	Longitud máxima	Número de bases <sup>a</sup>	Costo <sup>b</sup>	Ventajas	Desventajas
1°	Applied Biosystems	Sanger 3730xl	Terminación de cadena por dideoxy	400~900pb	1.9 a 84 Kb	\$2400	Alta calidad Longitud	Costo elevado Bajo rendimiento
2°	Roche	454 GS FLX	Pirosecuenciación	700pb	0.7Gb	\$10	Longitud	Error en homopolímeros > 6nt Costo elevado
2°	Illumina	HiSeq 2000	Secuenciación mediante síntesis	50SE, 50PE, 101PE	600Gb	\$0.07	Alto rendimiento	Ensamblado de lecturas tan cortas
2°	Applied Biosystems	SOLiDv4	Ligación y decodificación de dos bases	50+35pb ó 50+50pb	120Gb	\$0.13	Exactitud	Ensamblado de lecturas
2°	Complete Genomics	Polonator <sup>16</sup>	Ligación y decodificación de nueve bases	26pb (13+13)↑	3.9Gb	\$0.65↓	Código abierto Proyección de menores costos	Poca información
2°	Ion Torrent	Ion PGM	Semiconductores con una química de secuenciación simple	~200pb↑ en promedio	0.1Gb↑	\$0.1	Desarrollo a la par de los microchips Rapidez Bajo costo	Taza de error elevada Variación en el tamaño de lecturas
3°	Pacific Biosciences	SMRT	Liberación de fluoróforo mediante desplazamiento por síntesis de DNA, la señal es captada en tiempo real.	~1300pb en promedio <sup>17</sup>	¿?	¿?	No utiliza amplificación por PCR Longitud de lecturas más grande hasta el momento Rapidez	Aún en desarrollo, no comercializado. Costo elevado Bajo número de lecturas por corrida
3°	Oxford Nanopore	¿?	DNA de cadena sencilla se hace pasar por un bioporo de diámetro nanométrico embebido en una bicapa lipídica que facilita el intercambio de iones. Se hace pasar un flujo iónico constante y se miden las variaciones en el mismo.	Potencialmente >5 000pb a una velocidad de 1pb/ns <sup>18</sup>	¿?	¿?	Longitud de lecturas Rapidez	Aún en desarrollo, no comercializado

**Tabla1:** Tecnologías de secuenciación de DNA<sup>19</sup>. <sup>n</sup>Generación, <sup>a</sup>Por corrida completa <sup>b</sup>Por un millón de nucleótidos en dólares estadounidenses, ↑ con tendencia a aumentar, ↓ con tendencia a disminuir, ¿? no existe información al respecto. SE:single end, PE:pair end

## **Metagenómica viral**

El surgimiento del campo de estudio conocido como metagenómica viral lo marcó la publicación de dos metagenomas marinos en el año 2002. Estos trabajos demostraron que, mediante secuenciación masiva de DNA, es posible obtener información acerca de los organismos virales que habitan un ecosistema sin necesidad de cultivarlos<sup>20</sup>. Los virus son las entidades biológicas más abundantes en los océanos, en estos trabajos se identificaron como predominantes a las familias de fagos de DNA, así como algunos virus de algas, estimando que una comunidad viral marina contiene entre 300 y 7,000 subtipos virales. Posterior a este trabajo se han descubierto una gran cantidad de fagos marinos que fungen como promotores de la diversidad bacteriana<sup>21</sup>.

La metagenómica viral ha sido una herramienta poderosa en el descubrimiento de virus asociados a enfermedades de etiología desconocida que comparten poca identidad con virus conocidos o que no habían sido asociados anteriormente a una enfermedad.

### *Retos en la exploración de la diversidad viral en enfermedades humanas*

Los virus son las entidades biológicas más abundantes en el planeta<sup>22</sup>. En un ambiente marino los virus constituyen el 94% de todas las partículas que contienen ácidos nucleicos, con un estimado de  $10^{30}$  bacteriófagos presentes en la biosfera<sup>23</sup>. A la fecha sólo contamos con 3703 genomas virales de referencia<sup>24</sup> y en varios estudios de metagenomas virales el 90% de las secuencias no tienen similitud con alguna secuencia reportada en el banco de datos de secuencias mundial (GenBank), lo que nos habla de una baja representación de secuencias virales en las bases de datos<sup>3</sup> y por lo tanto de una diversidad viral aún no explorada.

Los virus no poseen un marcador filogenético conservado, como el gen 16S rRNA en procariotes o el 18S rRNA en eucariotes, que pueda utilizarse para determinar su historia evolutiva y diversidad, siendo este hecho un reto en su estudio y monitoreo. Sin embargo, con estrategias similares a los perfiles de DNA ribosomal es posible clasificar mediante secuencia a ciertos grupos virales usando como marcador filogenético a las proteínas de la cápside del virus o sus polimerasas de DNA o RNA, aunque no todo el universo viral podría ser representado en una sola filogenia.

Debido a la falta de un marcador filogenético, a la dificultad de aislarlos individualmente y a la diversidad aún no explorada que inferimos existe, un enfoque metagenómico en el estudio de ecosistemas virales nos dará un panorama más amplio y completo de la diversidad e interacciones virales existentes que el estudio de cada virus individualmente.

Mediante las técnicas actuales de secuenciación masiva de DNA es posible obtener millones de fragmentos de secuencias de una muestra heterogénea sin previa clonación o amplificación de los ácidos nucleicos presentes en la misma. Esta herramienta ha sido de gran utilidad en estudios metagenómicos basados en el análisis de material genético recolectado directamente de muestras ambientales. Este nuevo método permite el estudio de organismos que no son fácilmente cultivables en el laboratorio y de organismos en su ambiente natural<sup>25</sup>.

Al presentarse una enfermedad de etiología desconocida en una población, especialmente humana, es de suma importancia identificar el agente causal de la misma. Los virus de RNA son los principales agentes de enfermedades emergentes y re-emergentes<sup>26</sup>, también son capaces de traspasar la barrera de la especie que infectan y establecerse en un hospedero nuevo. En particular, gran parte de las enfermedades respiratorias son causadas por virus, siendo las neumonías la primera causa de mortalidad en niños menores de 5 años<sup>27</sup>. El diagnóstico del patógeno involucrado muchas veces es incierto dada la similitud de síntomas en diferentes infecciones.

### *Aplicaciones de la secuenciación masiva en virología clínica*

En el campo de salud humana la secuenciación masiva se ha aplicado al estudio detallado de varias vacunas virales, incluyendo la trivalente oral de poliovirus, en la que se detectaron mutaciones y variantes importantes<sup>28</sup>. Rubeola, fiebre amarilla, varicela y parotiditis, en las que se descubrieron virus contaminantes como virus de leucocis aviar y retrovirus de simio. En el caso de la vacuna oral de rotavirus, rotarix, se encontró que ésta estaba fuertemente contaminada con circovirus porcino-1, virus altamente prevalente en cerdos pero no patogéno para humanos<sup>29</sup>.

Los laboratorios de vigilancia epidemiológica o de diagnóstico ocasionalmente encuentran muestras provenientes de enfermedades infecciosas cuyos agentes etiológicos no pueden ser detectados por los métodos usuales, como cultivo celular, pruebas serológicas o amplificación de ácidos nucleicos. En estos casos la metagenómica es una excelente herramienta para identificar a los agentes infecciosos responsables del padecimiento. Un ejemplo es la identificación de virus de RNA conocidos y desconocidos en muestras de tejidos de animales enfermos en las que no se había podido detectar previamente ningún agente infeccioso asociado<sup>30</sup>.

La detección de virus en muestras clínicas se puede realizar mediante la amplificación de DNA con sondas específicas para un determinado tipo de virus. Al ser una técnica altamente específica no es capaz de detectar virus divergentes, requiere de un previo conocimiento de la secuencia viral, y las tasas de fracaso en la identificación son variadas. Por ejemplo, 30-85% en encefalitis, 12% en parálisis flácida, y 18-62% en hepatitis de virus no conocidos<sup>29</sup>. Una opción más es la detección de varias familias virales simultáneamente a través de microarreglos<sup>31</sup>, para lo cual es necesario que el virus a identificar posea una alta identidad con los oligonucleótidos impresos en el vidrio y que hibride correctamente para dar una buena señal, perdiendo así la detección de virus divergentes.

Una alternativa para identificar virus en aislados clínicos es el análisis de los ácidos nucleicos totales presentes en las muestras mediante secuenciación masiva de DNA. Esta técnica se ha utilizado eficientemente para identificar, entre otros, patógenos virales como

influenza, norovirus, hepatitis, arenavirus y poliomavirus, en especímenes clínicos en los que previamente no se habían descrito<sup>14</sup>, también fue posible detectar patógenos asociados a fibrosis quística no identificados anteriormente<sup>32</sup>

#### *Descubrimiento de virus en animales mediante NGS*

Las técnicas de NGS han sido un agente catalizador en el desarrollo de métodos que tienen como objetivo identificar a los agentes causales de enfermedades de etiología desconocida que muchas veces son virus. A continuación serán presentados algunos ejemplos exitosos del uso de NGS y metagenómica en el descubrimiento de nuevos virus causantes de enfermedades en animales.

El primer ejemplo es el estudio realizado en tortugas marinas con fibropapilomatosis, en el que se descubrió un nuevo virus con un genoma circular de DNA de cadena sencilla de aproximadamente 1,800 nucleótidos, “Sea turtle tornovirus 1”, demostrándose así el potencial de la metagenómica para encontrar nuevos virus directamente en tejidos de animales afectados por un cierto padecimiento<sup>33</sup>. Existen varios ejemplos más, como la identificación de un picornavirus en pavos con enfermedades entéricas de causa desconocida y de gran impacto económico en granjas productoras<sup>34</sup>. También existe otro trabajo en el que se identifica un virus re-emergente de la familia *Dicistroviridae* en abejas<sup>35</sup>, presuntamente causante de parálisis severa y devastación de la colonia, aunque estos resultados han sido altamente cuestionados, ya que la presencia del virus no necesariamente causa la devastación de la colonia<sup>36</sup>. También se descubrió un astrovirus que causa daño en el sistema nervioso, esto en una población de visones<sup>37</sup>.

El ejemplo más reciente es el descubrimiento y caracterización genómica de un virus en boas<sup>38</sup> causante de una enfermedad letal cuyos síntomas incluyen parálisis en las boas y se caracteriza por la aparición de cuerpos de inclusión en las células. El virus descubierto en este trabajo es un arenavirus altamente divergente a los previamente descritos.

### *Descubrimiento de virus en enfermedades humanas mediante NGS*

Existen varios casos en los que ha sido posible identificar nuevos virus en enfermedades humanas de etiología previamente desconocida.

En el 75% de los casos de encefalitis no se detecta algún patógeno asociado, un nuevo astrovirus denominado HAstV-PS<sup>39</sup>, fue descubierto en tejido cerebral, obtenido postmortem, de un muchacho de 15 años que sufrió encefalitis. Los astrovirus son virus no envueltos de cadena sencilla de RNA que típicamente se transmiten por vía fecal-oral y causan enfermedades gastrointestinales. La identificación de este nuevo astrovirus es un caso interesante ya que es la primera vez que se encuentra a un astrovirus como agente causal de enfermedades en el sistema nervioso. Posteriormente también se ha identificado a HAstV-PS en muestras fecales obtenidas de un brote de gastroenteritis en niños<sup>40</sup>, lo que sugiere que este virus está circulando en poblaciones humanas. En las enfermedades gastrointestinales, aproximadamente en un 30-50% de los casos no es posible detectar al agente causal. En un estudio realizado en muestras diarréicas pediátricas descubrieron un nuevo virus perteneciente al género picornavirus, lo nombraron klassevirus-1<sup>41</sup>. El descubrimiento se logró mediante la secuenciación masiva, en la plataforma Illumina, de un pool de 141 muestras. El virus más cercano a klassevirus-1 es aichivirus, con quien comparte una identidad de ~50%, por lo que se determinó que klassevirus es un nuevo género dentro de picornavirus. Sin embargo no está aún determinado si klassevirus causa alguna enfermedad o si es un agente causal de gastroenteritis.

En 2008 en Sudáfrica surgió un brote hemorrágico en humanos con una fatalidad del 80%, la causa era desconocida. Como parte del programa de vigilancia epidemiológica del centro de prevención de enfermedades (CDC) de Estados Unidos se tomaron muestras sanguíneas y de hígados de 5 casos, se extrajo RNA total y fue pirosecuenciado. Como resultado descubrieron 3 contigs pertenecientes a un nuevo arenavirus, diseñaron oligonucleótidos para amplificar las regiones faltantes y completaron el genoma del nuevo virus denominado Lujo virus (LUJV), las relaciones filogenéticas de este nuevo virus son muy interesantes. LUJV constituye una rama única del nodo ancestral de los arenavirus del viejo mundo,

además es el primer arenavirus de fiebre hemorrágica que se ha descubierto en las últimas tres décadas<sup>42</sup>.

En los tres casos presentados, el uso de las tecnologías de secuenciación masiva es indispensable para el descubrimiento de virus divergentes de los ya conocidos. Algunos otros ejemplos en enfermedades humanas, son el descubrimiento de Henen Fever Virus (HNF) un bunyavirus hemorrágico descubierto en China<sup>43</sup>, un subtipo nuevo de ebola descubierto en Uganda<sup>44</sup> y varios nuevos picobirnavirus en muestras fecales<sup>45</sup>.

### **El microbioma pulmonar humano**

*Características anatómicas e inmunológicas de los pulmones*<sup>46</sup>.

El tracto respiratorio bajo, compuesto de la tráquea y los pulmones, posee una estructura y función diferentes a las del tracto respiratorio alto que está altamente colonizado por microorganismos.

Los pulmones están divididos en 3 secciones de acuerdo a su estructura y función: bronquios, bronquiolos y alveolos. Los bronquios y bronquiolos son los conductos aéreos primarios, están rodeados de cartílago que permite el flujo aéreo hacia la parénquima, o tejido alveolar de los pulmones, en donde ocurre el intercambio gaseoso. Los conductos aéreos están cubiertos de epitelio ciliado intercalado con diferentes tipos de células que secretan mucinas, proteínas inmunomoduladoras, tensoactivos y proteasas. En conjunto las barreras químicas y físicas protegen de daño físico o biológico al establecer un flujo de transporte mucosal hacia la parte superior de los pulmones.

Además de las barreras fisicoquímicas que aportan la cilia y la mucosa, el epitelio está también protegido por varias células inmunes como células dendríticas, células de Langerhans, linfocitos T y mastocitos que responden contra los diferentes antígenos inhalados estableciendo así una barrera inmunológica robusta en contra de posibles patógenos, ayudando así a mantener los alveolos estériles.



Se piensa que, en conjunto con la barrera física que aporta la nariz y la mucosa del tracto respiratorio alto, las defensas son suficientes para mantener al tracto respiratorio bajo estéril.

#### *Métodos de toma de muestras pulmonares.*

Existen varias técnicas para tomar muestras del nicho ecológico pulmonar. En orden de la menos a la más invasiva son: inducción de esputo, lavado broncoalveolar, cepillado bronquial, biopsia endobronquial y biopsia transbronquial.

La inducción de esputo y el lavado broncoalveolar muestrean a las células y patógenos que pueden ser lavados del lumen pulmonar. El cepillado bronquial brinda acceso a las células y microbios que están adheridos a la superficie del lumen de las vías respiratorias bajas. La biopsia bronquial toma una muestra de tejido de las vías respiratorias grandes brindando acceso a microbios invasivos dentro del tejido pulmonar. La biopsia transbronquial permite muestrear las vías aéreas más pequeñas así como los alveolos.

El lavado broncoalveolar consiste en introducir un broncoscopio a través de la boca o la nariz hasta los pulmones, dejar fluir una solución salina en un segmento del pulmón y recolectarla para posteriormente ser examinada<sup>47</sup>. Es la técnica más común para muestrear en el tracto respiratorio bajo: colonizaciones o infecciones bacterianas, componentes de los fluidos que secreta el epitelio y para determinar la composición proteica de las vías aéreas. Frecuentemente es utilizada para monitorear el estado inmunológico del pulmón del paciente mediante la medición de los niveles de células y patógenos. Es una técnica invasiva y por lo tanto solo se realiza en pacientes enfermos.

#### *Enfermedades del tracto respiratorio bajo y su impacto en el microbioma pulmonar.*

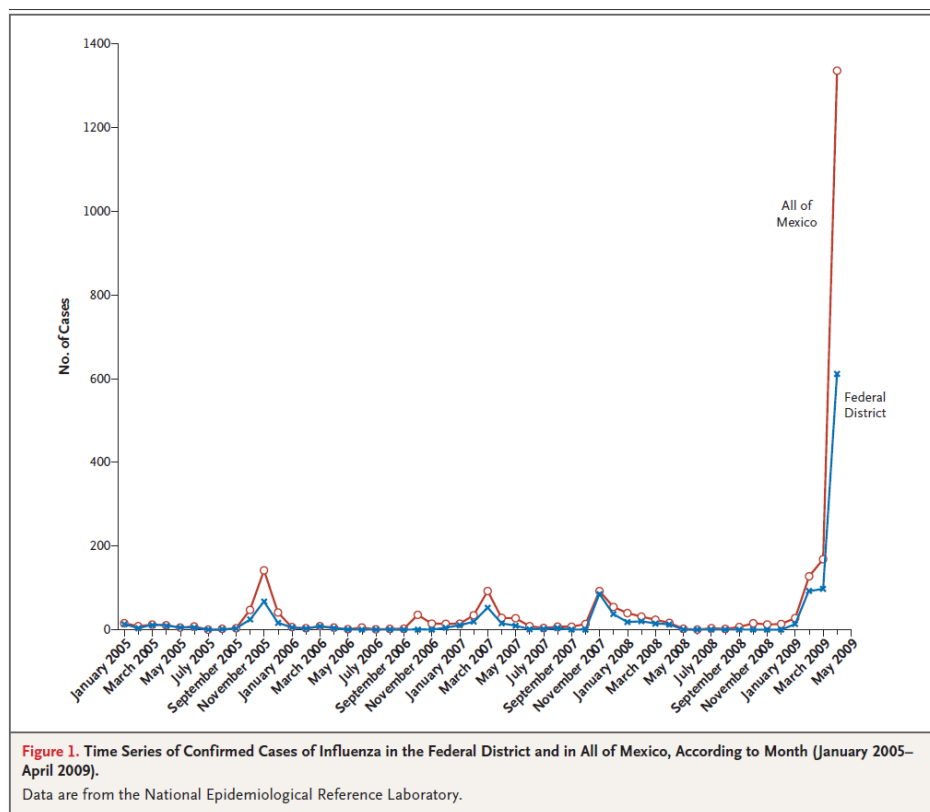
Las enfermedades respiratorias crónicas son asma, fibrosis quística (CF), bronquiectasias y enfermedad crónica de obstrucción pulmonar (COPD). Estas enfermedades frecuentemente llevan a una disfunción del flujo mucosal y/o el sistema inmunológico pulmonar. En los casos de deficiencia inmune, como la causada por el virus de inmunodeficiencia humana (VIH), la homeostasis inmune de los pulmones se ve afectada.

# ANTECEDENTES

## Influenza

### *Panorama general de la pandemia de 2009-2010*

A principios del mes de marzo de 2009 se reportó en México una mayor cantidad de enfermedades de vías respiratorias respecto al mismo mes en años anteriores, las cuales se confirmaron posteriormente como un brote del virus de influenza A subtipo H1N1<sup>48</sup>.



**Figura 2:** Casos reportados de influenza en 2009 en México

En México los virus de influenza A y B circulan a lo largo de todo el año, con predominio en los meses de invierno. Los subtipos antigénicos de influenza A más frecuentes en poblaciones humanas son H1, H3, N1 y N2; de 1994 a 2005 el 91% de las enfermedades por influenza fueron del subtipo H3N2, 8% por H1N1 y 1% por H1N2<sup>49</sup>.

El Sistema Nacional de Vigilancia Epidemiológica, con base en la información recabada en todo el país de la temporada 2008/2009 detectó tres aspectos particulares de esta epidemia<sup>50</sup>:

- Aumento en el número de casos de influenza
- Desplazamiento en el pico de la temporada que habitualmente aparece en Diciembre y en esta ocasión se presentó en Marzo de 2009
- Incremento en la proporción de infecciones debidas a virus de influenza tipo B
- Aparición de casos de neumonía grave de rápida evolución con letalidad mayor al 1%

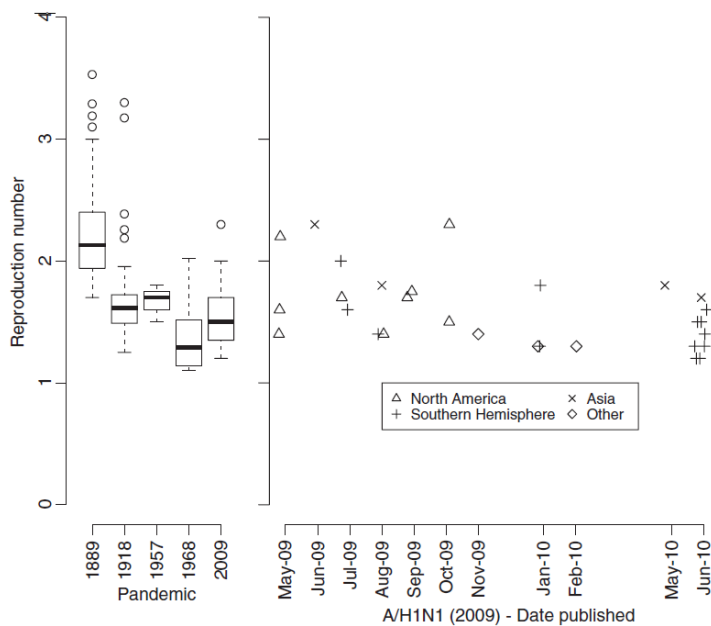
La mortalidad asociada al virus de influenza causante de la pandemia A H1N1 2009 (H1N1pdm09) en los casos iniciales ocurridos en México se determinó como 0.45%. Este índice es ligeramente alto comparado con la letalidad de 0.1% de influenza estacional, pero bajo comparado con 35% a 65% del virus H5N1 y entre 2.5% y 5% del virus de influenza de la pandemia de 1918<sup>51</sup>. Estos índices son una referencia pero no son directamente comparables ya que la letalidad no solo depende del virus per se, sino también depende de las condiciones y el tratamiento que cada paciente recibe.

#### *Transmisibilidad.*

Los parámetros que la Organización Mundial de la Salud (OMS) recomienda medir cuando surge una epidemia son: periodo de incubación, es decir el tiempo entre la infección y los síntomas; intervalo serial, el tiempo entre los síntomas de un primer caso y un segundo caso; y la tasa o número de reproducción ( $R$ ), esto es el número promedio de casos secundarios por caso primario, un valor de  $R$  mayor a 1 se considera como un brote epidémico de la enfermedad. Estos parámetros son esenciales para estimar la eficacia de las medidas estratégicas para contrarrestar una pandemia. También son importantes al comparar varios brotes epidémicos a través de la historia.

El parámetro para el cual existe más información en las últimas cuatro pandemias de influenza (1889, 1918, 1957 y 1986) es el número de reproducción. Los estimados van de 1 a 6 dependiendo del lugar, tiempo, ola y métodos de estimación.

Al utilizar toda la información publicada relacionada con A/H1N1 (2009) hasta Julio de 2010, se muestra que la media de intervalo serial (SI) fue  $< 3$  días y el número de reproducción ( $R$ ) entre 1:2 y 1:3 con una media de 1:1.5, como se muestra en la Figura 3<sup>52</sup>, este  $R$  es muy cercano al estimado en pandemias anteriores. Por lo anterior, la pandemia de 2009 no fue mucho más letal ni con una mayor transmisibilidad que las anteriores.



**Figura 3:** Valores de número de reproducción estimado para las pandemias de influenza.

### *Orígenes de la pandemia*

En el caso de la pandemia de influenza de 2009-2010, el origen del virus fue porcino, producto de un virus triple rearrreglante<sup>53</sup>. El virus de influenza A posee un genoma de RNA de cadena sencilla de polaridad negativa, el cual está organizado en 8 segmentos que codifican 10 y algunas veces 11 proteínas cuya función se describe en la Tabla 2.

Segmento	Nombre	Proteínas	Principales funciones
1	PB2	PB2	Componente de RNA polimerasa, reconocimiento de cap
2	PB1	PB1, PB1-F2	Componente de RNA polimerasa, actividad endonucleasa y elongación
3	PA	PA	Componente de RNA polimerasa, proteasa
4	HA	HA	Glicoproteína de superficie, unión al receptor celular, actividad de fusión, antígeno mayor
5	NP	NP	Unión a RNA, síntesis de RNA, importación de RNA al núcleo
6	NA	NA	Glicoproteína de superficie, actividad neuraminidasa
7	M	M1, M2	Proteína de matriz, interactúa con vRNPs y las glicoproteínas de superficie, exporte nuclear, canal iónico, ensamblaje
8	NS	NS1, NEP/NS2	Proteína multifuncional, antagonista de interferón. Exporte nuclear de vRNPs

**Tabla 2:** Segmentos del virus de influenza A y proteínas que codifican.

Al ser un genoma segmentado es posible que ocurran eventos de rearrreglos genómicos al co-infectar, dos o más virus de subtipos diferentes a un mismo hospedero. Éste fue el caso del origen del virus de influenza H1N1pdm2009. La envoltura del virus está compuesta por una membrana lipídica que alberga a las proteínas hemaglutinina (HA) y neuraminidasa (NA). HA y NA al estar expuestas confieren las características antigénicas de cada virus<sup>54</sup>. Las aves acuáticas son un hospedero en el que se han encontrado los 16 tipos de HA y los 9 de NA existentes, por lo que se cree que en este ecosistema es en donde surge la mayor proporción de rearrreglos entre diferentes subtipos<sup>55</sup>.

La combinación de segmentos del virus de influenza H1N1pdm2009 no se había encontrado anteriormente ni en cerdos ni en humanos. El origen de cada segmento se discute a continuación y se ilustra en la Figura 4.

NA y M provienen del linaje porcino Euroasiático; se piensa que estos genes fueron introducidos de aves a cerdos aproximadamente en 1979 y desde entonces no había sido reportado fuera de Europa y Asia. HA, NP y NS pertenecen al linaje porcino clásico; fueron introducidos en la población porcina cerca de 1918; estos segmentos posteriormente formaron parte del linaje conocido como triple rearreglante. PB2 y PA son del linaje conocido como triple rearreglante porcino; estos segmentos, originalmente de origen aviar, entraron a la población de cerdos norteamericanos cerca de 1998. PB1 fue introducido de humanos a cerdos y pasó a formar parte del linaje triple rearreglante en Norteamérica. Fue introducido de aves a humanos aproximadamente en 1968<sup>52</sup>. Las fechas en las que se cree fue introducido cada segmento, son las fechas en las que se detectaron por vez primera, sin embargo pudieron haberse introducido mucho antes y ser detectados tiempo después.

Dado que todos los segmentos del virus H1N1pdm2009 han circulado en algún momento en poblaciones porcinas, es altamente probable que el rearreglo ocurriera en cerdos. El rearreglo final pudo haber surgido en cualquier parte del mundo y probablemente circuló en poblaciones porcinas sin ser detectado hasta que se transmitió a humanos<sup>56</sup>.

Las secuencias con las que los segmentos del virus H1N1pdm2009 tienen mayor identidad fueron aisladas hace 10 o más años, lo que sugiere que este virus estuvo circulando en poblaciones de cerdos sin ser detectado previamente.



**Figura 4:** Origen del virus influenza A H1N1 2009

### *Situación en México de mayor mortalidad que en el resto del mundo*

En los primeros meses de la epidemia de influenza A, se reportó en México una mayor mortalidad que en el resto del mundo. Así, hasta el 29 de mayo de 2009 se habían reportado 85 muertes (1.73%) de 4910 casos confirmados con infección de influenza H1N1pdm09. México fue el país que reportó más casos severos que culminaron en decesos, mientras que en otros países predominaron los casos leves de la infección<sup>57</sup>. Las explicaciones para esta observación en su momento, entre otras, fueron las siguientes:

- El virus circulante en México es más virulento que el que circula en otras partes del mundo.
- Siendo la región donde primero se detectó el virus, el sistema médico-hospitalario fue sorprendido y no se dio la atención adecuada.
- Existe algún otro agente patogénico, que cocirculó con el virus H1N1pdm09, agravando la situación de los pacientes.

Los virus que circularon en México se detectaron también en otros lugares del mundo, en varios países se reportaron algunas secuencias 100% idénticas a las aisladas en México.

### *Casos hospitalizados, obtención de muestra.*

Durante los meses de la pandemia, el Instituto Nacional de Enfermedades Respiratorias (INER) de la Secretaría de Salud recibió un número mayor al usual para esa época del año de casos de neumonía grave con síntomas parecidos a los de influenza A, como son fiebre, tos, disnea, niveles altos de lactato deshidrogenasa en sangre y neumonía bilateral. Del 24 de marzo al 24 de abril de 2009 se presentaron un total de 214 consultas en sala de urgencias a causa de neumonía o enfermedades parecidas a influenza. 98 de estos pacientes fueron hospitalizados. De algunos de ellos se tomaron muestras nasofaríngeas al ser admitidos al hospital y de otros se obtuvieron muestras de lavados bronco-alveolares al complicarse el curso de la enfermedad y ser intubados.

### *Neumonía bacteriana como complicación de la enfermedad*

Una neumonía bacteriana secundaria a la causada por una infección con el virus de influenza A puede presentarse días después del inicio de la enfermedad y a diferencia de la neumonía primaria, que continúa con el curso clínico de influenza, la bacteriana ocasiona una reaparición de la fiebre, expectoración purulenta y nuevas opacidades radiográficas. Los gérmenes secundarios comúnmente son *Streptococcus pneumoniae*, *Haemophilus influenzae* y *Moraxella catarrhalis*. En las neumonías bacterianas secundarias asociadas a influenza hay una sobrerrepresentación de *Staphylococcus aureus*, comparado con la neumonía bacteriana sin influenza y se han reportado casos en los que *S. aureus* es resistente a metilciclina<sup>58,59</sup>. Esta bacteria puede generar una alta mortalidad en estos casos. En 25% de los casos en los que se aplica ventilación mecánica existe complicación bacteriana nosocomial asociada al ventilador<sup>59</sup>

### *Métodos de diagnóstico de influenza*

En muchos de los pacientes hospitalizados en el INER se detectó el nuevo virus H1N1pdm2009 usando el protocolo de PCR del Centro para Prevención y Control de Enfermedades (CDC) de los EUA. Sin embargo, en otros pacientes con síntomas similares a influenza, no se detectó el virus. En el periodo del 24 de Marzo al 29 de Abril de 2009, la Secretaría de Salud en México, recolectó 8817 especímenes nasofaríngeos de pacientes con síndrome respiratorio severo, de los cuales 2582 fueron positivos al virus H1N1pdm2009, representando esto un 29% de casos positivos<sup>60</sup>.

El propósito de este proyecto fue analizar mediante metagenómica la diversidad viral y bacteriana presente en las muestras de varios pacientes del INER en las cuáles se detectó el virus de influenza y otro grupo en las que no se detectó el virus.



## **HIPÓTESIS**

Es posible detectar virus conocidos e identificar virus desconocidos mediante el uso de secuenciación masiva de DNA en muestras clínicas de pacientes con síndrome respiratorio severo.

## **OBJETIVO**

Identificar los patógenos virales y bacterianos presentes en muestras obtenidas de pacientes con síndrome respiratorio severo positivos y negativos a influenza.

### **Objetivos particulares**

1. Secuenciar con tecnologías de segunda generación muestras de aspirados broncoalveolares o lavados nasofaríngeos de pacientes con enfermedad respiratoria severa en las cuáles se haya detectado el virus H1N1pdm2009, o que hubieran cursado con una enfermedad similar a influenza pero sin detección del virus.
2. Diseñar el pipeline bioinformático necesario para el análisis metagenómico de los datos de secuenciación masiva.
3. Determinar el origen de las secuencias obtenidas, ya sea DNA humano, o de patógenos bacterianos o virales.

## **MATERIALES Y MÉTODOS**

### **Pacientes**

Seleccionamos a 14 pacientes para este trabajo; todos ellos acudieron al INER durante la pandemia de influenza entre mayo y septiembre de 2009 y fueron diagnosticados con “enfermedad respiratoria severa”. Seis de estos pacientes, fallecieron, los cuales fueron atendidos en la Unidad de Cuidados Intensivos (UIC). Los otros 8 pacientes sobrevivieron y se atendieron como consulta externa, o fueron internados en la Sala General o en la UIC. Los días de enfermedad fueron variables entre 3 y 15. El rango de edades fue entre 28 y 44 años, con una media de 36 años para los fallecidos, y un rango de edad entre 8 y 70 para los que se dieron de alta. Todos los pacientes fueron de sexo masculino. Estos datos se muestran en la Tabla 3.

Se obtuvieron muestras de lavados broncoalveolares para los pacientes hospitalizados que fueron intubados en la UIC y se tomaron muestras nasofaríngeas al ingresar al hospital para los demás. Todos los pacientes otorgaron consentimiento informado para la toma de muestras.

Los lavados broncoalveolares se realizaron con una solución salina normal (Cloruro de Sodio 0.9%) y se colocaron en medio de transporte Remel, se almacenaron a  $-70^{\circ}\text{C}$ . Las muestras nasofaríngeas fueron tomadas de la parte profunda de la faringe utilizando un hisopo estéril de dacrón, posteriormente se colocaron en medio de transporte Remel (ver material suplementario S1) y se almacenaron a  $-70^{\circ}\text{C}$ . Todos los pacientes fueron tratados con el antiviral oseltamivir. Los pacientes hospitalizados fueron tratados, además de con el antiviral, con los antibióticos ceftriaxona y claritromicina.

Muestra	Edad	Ubicación	Diagnóstico	SRIS	Enfermedad crónica	Complicación	Días de enfermedad	Tipo muestra	Muerte	Fecha
6985	43	UCI	SP	Si	HTAS	No	8	Broncoalveolar	si	09/08/2009
7192	28	UCI	SP	Si	No	No	14	Broncoalveolar	si	16/08/2009
7425	44	UCI	SP	Si	Obeso	NAVM	6	Broncoalveolar	si	22/08/2009
8261	28	UCI	SP	Si	No	No	13	Broncoalveolar	si	15/09/2009
8836	37	UCI	SP	Si	No	No	4	Broncoalveolar	si	30/09/2009
9041	41	UCI	SP	Si	No	IRA	8	Broncoalveolar	si	06/10/2009
3974	70	UCI	P	Si	No	No	NA	Broncoalveolar	no	NA
8410	31	EXT	ILI	No	No	No	NA	Nasofaríngeo	no	NA
8883	8	ISG	P	No	No	No	15	Nasofaríngeo	no	NA
9140	63	UCI	SP	Si	DM	NAVM	5	Broncoalveolar	no	NA
9232	33	ISG	P	No	No	No	4	Nasofaríngeo	no	NA
9275	42	ISG	P	No	No	No	8	Nasofaríngeo	no	NA
9362	33	EXT	ILI	No	No	No	3	Nasofaríngeo	no	NA
9434	51	EXT	ILI	No	No	No	7	Nasofaríngeo	no	NA

**Tabla 3:** Datos clínicos de los 14 pacientes participantes en este estudio

UCI=unidad de cuidado intensivo EXT=externo ISG=internado en sala general SP=neumonía severa P=neumonía ILI=enfermedad parecida a influenza SRIS=respuesta inflamatoria sistémica HTAS=hipertensión arterial sistémica DM=diabetes mellitus NAVM=neumonía asociada a ventilación mecánica

### Diagnóstico molecular de influenza A

Las muestras obtenidas de los pacientes se dividieron en varias alícuotas con el objetivo de probar distintos métodos de diagnóstico molecular del virus de influenza A, todos estos métodos se basan en la técnica de reacción en cadena de la polimerasa (PCR) y varían en los subtipos de virus de influenza A que son capaces de detectar.

La Reacción en Cadena de la Polimerasa (PCR)<sup>61</sup> es una técnica mediante la cual es posible obtener muchas copias (amplificar) un fragmento de DNA específico. Esto se logra gracias a las propiedades intrínsecas de desnaturalización y apareamiento del DNA a distintas temperaturas. Para amplificar una región es necesario conocer la secuencia de los extremos que la flanquean. La especificidad en la amplificación de un fragmento está determinada por el diseño de sondas (primers) con una secuencia complementaria a cada extremo de la región que se desea amplificar.

El diagnóstico molecular mediante PCR se basa en el diseño de pares de primers que sólo amplifiquen un fragmento de DNA específico del patógeno, tal fragmento tendrá un tamaño esperado que será determinado mediante un gel de agarosa. Después de que la muestra es amplificada, el producto de PCR es sometido a una electroforesis en gel de agarosa junto a un marcador de peso molecular y así, si al término de la electroforesis, el fragmento de DNA de tamaño esperado está presente, es conclusivo que el DNA del patógeno está presente en la muestra.

#### *RT-PCR cuantitativo (qRT-PCR)*

Una variación a la técnica de PCR es el PCR cuantitativo, en el que en cada ciclo de amplificación una señal fluorescente es medida permitiendo así conocer la cantidad de moléculas en cada ciclo.

Se realizó RT-PCR cuantitativo en el laboratorio de diagnóstico del INER para detectar influenza H1N1pdm09 de acuerdo al protocolo recomendado por la Organización Mundial de la Salud (OMS) y el Centro para la Prevención de Enfermedades de Estados Unidos (CDC) (ver material suplementario S2). El protocolo consiste en el uso, por separado, de 3 pares de oligonucleótidos, el primero identifica influenza A (InfA) y está dirigido al segmento M; el segundo identifica influenza de origen porcino (swInfA) dirigido al segmento NP; y el tercero al virus H1N1pdm09 (swH1) dirigido al segmento HA. Para que una muestra sea diagnosticada como positiva al virus pandémico es necesario que, en no más de 40 ciclos, sea positiva para InfA y alguno de los subtipos swInfA o swH1<sup>62</sup>, si sólo es positiva para InfA es considerada como una muestra que contiene virus de influenza no pandémico.

Cinco de los 14 pacientes resultaron positivos. Los nueve pacientes en los que no se detectó el virus H1N1pdm09 presentaron síntomas similares a los causados por el virus de influenza. Una segunda alícuota se utilizó para realizar los siguientes ensayos en el Instituto de Biotecnología:

## *Seeplex*<sup>63</sup>

Se utilizó el kit comercial “Seeplex RV15 ACE Detection” ,producido por la compañía Seegene, para la identificación de 15 virus respiratorios en las muestras, tales virus son:

- Adenovirus /A/B/C/D/E
- Coronavirus 229E/NL63
- Parainfluenza 1
- Parainfluenza 2
- Parainfluenza 3
- Coronavirus OC43
- Rinovirus A/B/C
- Influenza A
- Sincicial Respiratorio A
- Sincicial Respiratorio B
- Bocavirus 1/2/3/4
- Influenza B
- Metapneumovirus
- Parainfluenza 4
- Enterovirus

El diagnóstico mediante Seeplex se basa en pares de oligonucleótidos que se usan en 3 reacciones de PCR separadas, seguidas del análisis de los productos de cada una en un gel de agarosa. Al comparar las bandas presentes en cada gel, se determina la presencia de cada patógeno.

El límite de detección de influenza A por este método es de 100 copias del genoma viral y las sondas están dirigidas al segmento 7 que codifica para el gen M, por lo que no distinguen entre subtipos de influenza A.

Mediante este método, cinco muestras fueron positivas a influenza A, una positiva a rinovirus y ocho negativas, los resultados se muestran en la Tabla 4.

## *PCR*

Los 8 segmentos del virus influenza A poseen una región conservada de 12 nucleótidos en ambos extremos que se utilizan para amplificar específicamente los 8 segmentos en una sola reacción de PCR<sup>64</sup>. Este método se utilizó como una segunda opción para detectar la

presencia de virus de influenza en las muestras. De este producto se llevó a cabo un PCR con sondas específicas para diferentes genes de influenza: HA, NA, PB1 y M

*Amplificación del genoma completo y secuenciación*

En tres de las muestras utilizamos el producto de amplificación del genoma completo descrito anteriormente para secuenciarlas en la plataforma Illumina, obteniendo millones de lecturas de 36 nucleótidos de longitud. Estas secuencias se alinearon al genoma de referencia A/Netherlands/2009, ya que fué uno de los primero genomas publicados del virus pandémico, utilizando el programa Mapping ans Assembling with Qualities<sup>65</sup> (MAQ), y así recuperamos el genoma casi completo de una muestra y el genoma parcial de dos de ellas.

En la Tabla 4 se resumen los resultados del diagnóstico de influenza A realizado en cada una de las muestras.

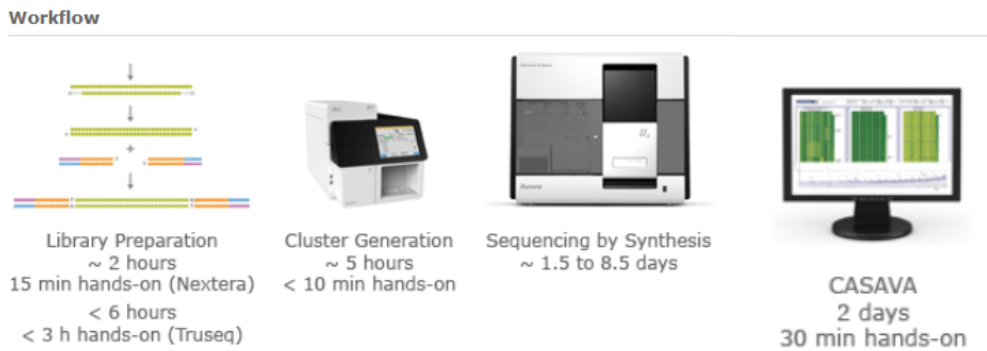
Muestra	Muerte	Seplex	RT-PCR CDC <sup>a</sup>	PCR gene M	Genoma <sup>b</sup>
6985	Si	Influenza A	Positivo	Positivo	95%
7192	Si	Influenza A	Positivo	Positivo	76%
7425	Si	Negativo	Positivo	Positivo	NA
8261	Si	Influenza A	Positivo	Positivo	61%
8836	Si	Influenza A	Positivo	Positivo	NA
9041	Si	Influenza A	Positivo	Positivo	NA
3974	No	Negativo	Negativo	Negativo	NA
8410	No	Negativo	Negativo	Negativo	NA
8883	No	Negativo	Negativo	Negativo	NA
9140	No	Rinovirus	Negativo	Negativo	NA
9232	No	Negativo	Negativo	Negativo	NA
9275	No	Negativo	Negativo	Negativo	NA
9362	No	Negativo	Negativo	Negativo	NA
9434	No	Negativo	Negativo	Negativo	NA

**Tabla 4.** Resultados de diagnóstico molecular de influenza A en las 14 muestras

NA=no analizado, a=realizado en el INER, b=obtenido mediante secuenciación masiva

## Secuenciación masiva de DNA en la plataforma Illumina GAIIx

Los ácidos nucleicos totales de cada muestra fueron extraídos para ser secuenciados en la plataforma Illumina GAIIx, la secuenciación consta de cuatro pasos (Figura 5): preparación de la biblioteca; generación de clusters; secuenciación mediante síntesis; y análisis de imágenes.



**Figura 5:** Procedimiento de secuenciación en GAIIx

### *0.-Extracción de ácidos nucleicos.*

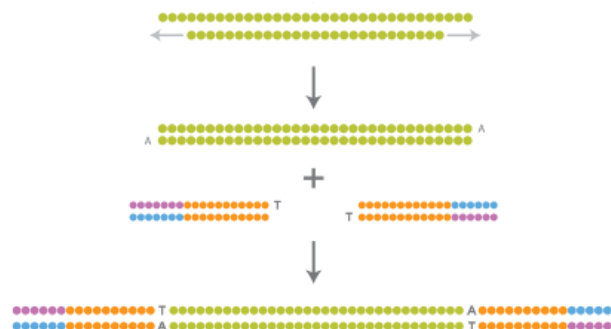
La extracción de ácidos nucleicos totales fue realizada con el kit comercial “PureLink® Viral RNA/DNA Kit” de la compañía invitrogen (ver material suplementario S3). Se basa en el principio de digerir las proteínas de la muestra mediante un buffer de lisis capaz de abrir las partículas virales y adicionalmente, el uso de la serin proteasa, proteinasa K, que digiere rápidamente a las nucleasas evitando que los ácidos nucleicos sean degradados.

Una modificación es que en vez de utilizar tRNA de levadura como acarreador, se utilizan 5 microgramos por muestra de acrilamida linear. Posteriormente la muestra se pasa por una matriz de silica en la que los ácidos nucleicos quedan inmovilizados y después son eluidos quedando así los ácidos nucleicos totales libres de restos protéicos.

### *1.-Preparación de la biblioteca para secuenciación*

Una vez extraídos los ácidos nucleicos totales de cada muestra, se utilizó el kit para secuenciar DNA genómico de Illumina siguiendo las instrucciones sugeridas (ver material suplementario S4), el proceso se describe a continuación.

El primer paso del proceso es fragmentar el DNA de la muestra en moléculas de 800 pares de bases o menos mediante nebulización. Este proceso genera fragmentos de DNA de doble cadena con extremos 5' o 3' sobresalientes. Los extremos tienen que ser reparados para obtener fragmentos con extremos romos. Este paso involucra la digestión de los 3' protuberantes mediante la actividad exonucleasa de la polimerasa de T4 y la extensión de los extremos 3' recesivos por el fragmento Klenow de la DNA polimerasa de *E. coli*. Finalmente la polinucleotidil cinasa de T4 es utilizada para fosforilar los extremos 5'. El segundo paso es agregar una adenina a los extremos 3' de cada fragmento de DNA, usando la actividad polimerasa del fragmento Klenow (3' a 5' exo minus). Este paso es necesario para ligar a cada fragmento los adaptadores que poseen una timina protuberante en su extremo 3'. El tercer paso es una ligación entre los fragmentos y los adaptadores, se lleva a cabo por una DNA ligasa, el proceso se ilustra en la Figura 6. El último paso es la separación, selección y purificación de los fragmentos de aproximadamente 300 pares de bases (adaptador más fragmento) mediante electroforesis en gel de agarosa al 2% y el producto se lleva a una concentración de 18 picomolar.



**Figura 6:** Generación de biblioteca para secuenciación

## 2.- Generación de clusters

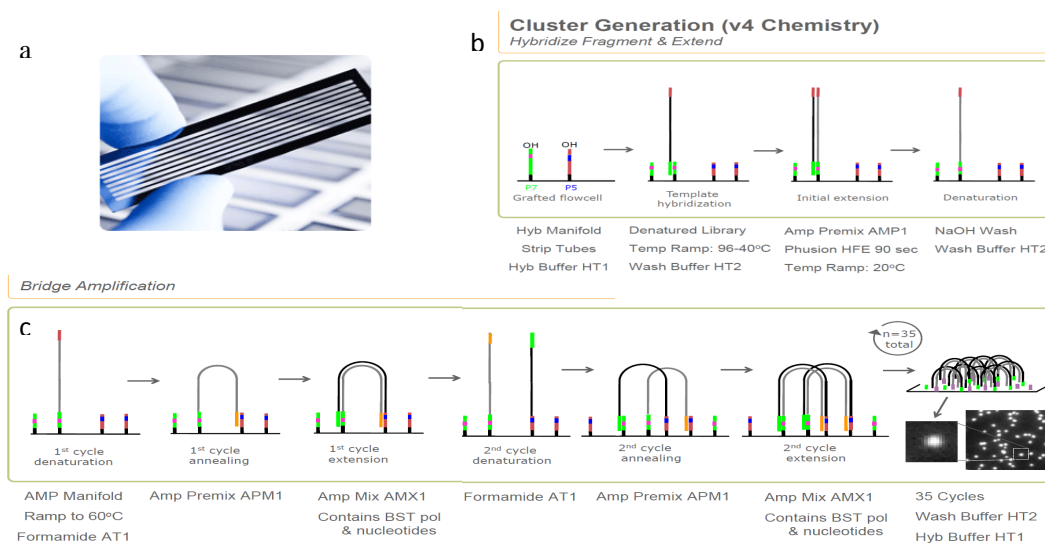
Este paso se lleva a cabo en la unidad generadora de clusters, un equipo separado del secuenciador, en el que se coloca un portaobjetos (flowcell, Figura 7a) con 8 hendiduras longitudinales (carriles), y se inyecta cada muestra en un carril mediante flujo capilar.



Cada fragmento de DNA en la muestra posee un adaptador complementario a un oligonucleótido ya impreso en el vidrio, con el que hibridará quedando fijo. La librería que se inyecta en el flowcell está desnaturalizada (cadena sencilla de DNA) y mediante un gradiente de temperatura de 96-40°C las secuencias hibridan con su complementario impreso en el vidrio, se agregan polimerasa y nucleótidos y se hace un gradiente de temperatura hasta 20°C, propiciando la síntesis de la cadena complementaria que quedará unida al oligo impreso en el vidrio, en condiciones desnaturalizantes de NaOH se eluye la cadena original (Figura 7b).

Una vez unida cada secuencia al vidrio, siguen 35 ciclos de amplificación para generar los clusters o polonies que son grupos de secuencias provenientes de un solo templado. Los pasos en cada ciclo de la generación de clusters son: hibridación, amplificación, linearización, bloqueo e hibridación (Figura 7c).

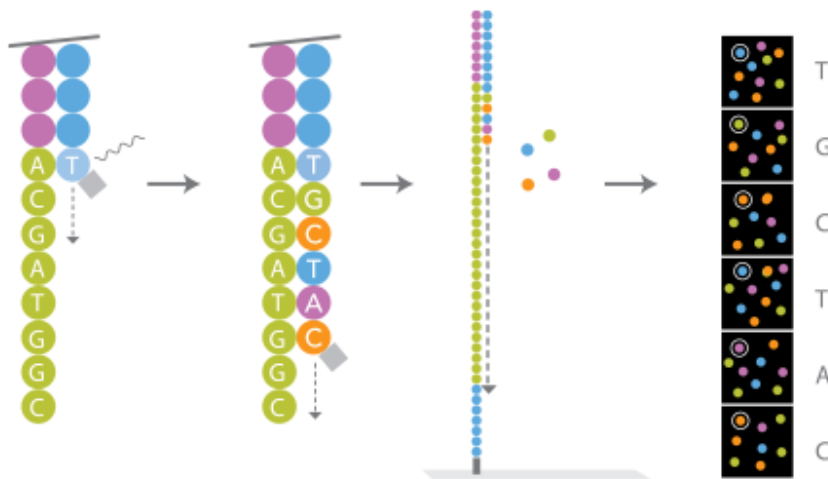
Para nuestras muestras utilizamos 2 flowcells y colocamos una muestra en cada carril



**Figura 7:** Generación de clusters: a) Flowcell b) hibridación del fragmento c) amplificación en puente

### 3.-Secuenciación mediante síntesis

Una vez generados los clusters, el flowcell se coloca en el secuenciador en donde se dejan fluir cíclicamente por todos los carriles las 4 bases nitrogenadas marcadas cada una con un fluoróforo distinto, junto con una DNA polimerasa de alta fidelidad. Posteriormente, mediante un lector laser y un lente se toman fotografías de todo el flowcell en cada ciclo, dependiendo del número de ciclos que se realicen, será la longitud de las secuencias. Un ciclo consiste de los siguientes pasos: agregar NTP's y polimerasa, incorporación del nucleótido y toma de imágenes, corte del terminador y la marca fluorescente. En las 14 muestras se realizaron 75 ciclos de secuenciación.



**Figura 8:** Secuenciación mediante síntesis.

### 4.-Análisis de imágenes

Las imágenes obtenidas del flowcell son procesadas en tiempo real. Primero, a partir de las imágenes TIF se identifican las coordenadas, intensidades y se estima el ruido de cada cluster; posteriormente se utilizan las intensidades y ruido para obtener la secuencia de cada





**Figura 9:** Proceso de análisis de imágenes.

Como control del proceso de secuenciación masiva, en un carril de cada flowcell se coloca una muestra control, esta muestra contiene al bacteriófago PhiX. Este carril se utiliza para medir el número de clusters generados y la calidad de las secuencias, entre otros parámetros. En los análisis posteriores también utilizaremos este carril como control y lo denominaremos PhiX.

### Procesamiento de las secuencias

Los archivos de secuencias en formato fastq fueron transferidos al cluster del Instituto de Biotecnología en donde fueron almacenados y procesados utilizando varios programas escritos por nosotros en diferentes lenguajes de programación, como son: bash, perl, python. También utilizamos en el análisis programas como BLAST, BLAT, MAQ, SMALT, VELVET y MEGAN, entre otros. Su uso será explicado a continuación.

El algoritmo BLAST<sup>67</sup> (Basic Local Alignmen Search Tool) fué diseñado para comparar secuencias de nucleótidos u aminoácidos. Los elementos necesarios para realiza la búsqueda son: una secuencia problema o query, y una base de datos sobre la cual realizar la búsqueda. La búsqueda es realizada mediante alineamientos entre la secuencia problema y cada una de las secuencias en la base de datos. Ciertos parámetros son calculados para cada alineamiento, como identidad y similitud entre ambas secuencias, e-valor y p-valor, entre otros.

BLAST es un algoritmo heurístico que no garantiza encontrar el mejor alineamiento, sin embargo es mucho más rápido que el algoritmo Smith-Waterman<sup>68</sup> que sí lo garantiza. La velocidad de BLAST sobre Smith-Waterman, aunada a las medidas de confianza para cada alineamiento, han hecho de BLAST el más popular de los métodos de comparación de secuencias. Al definir los puntos de corte es posible identificar con confianza el mejor o mejores hits de cada secuencia problema en la base de datos.

La complejidad de BLAST es  $m \times n$ , siendo  $m$  el número de secuencias problema y  $n$  el tamaño de la base de datos. Esto significa que el tiempo que tarde una búsqueda es directamente proporcional al tamaño de la base de datos con la cual se compare.

BLAT<sup>69</sup> (BLAST-Like Alignment Tool) es un algoritmo con el mismo objetivo, pero más eficiente que BLAST. La eficiencia se debe a una estrategia distinta en la forma en la que los alineamientos son calculados, BLAT crea un índice de todos los  $k$ -meros que no sobrelapen de la base de datos. Este índice es cargado en la memoria RAM sólo una vez para realizar todas las búsquedas, así los alineamientos se basan en comparaciones de  $k$ -meros reduciendo el espacio de búsqueda al ir recorriendo el índice. Los parámetros que se miden para cada alineamiento son equivalentes a los usados en BLAST.

MAQ (Mapping and Assembling with Qualities) es un algoritmo diseñado para ensamblar genomas. Las tecnologías de secuenciación masiva arrojan fragmentos pequeños de secuencias que es necesario ensamblar para realizar análisis posteriores. MAQ requiere de un genoma de referencia que se utiliza como molde sobre el cuál se acomodan o mapean cada uno de los pequeños fragmentos y posteriormente se recupera el genoma que fue secuenciado, tal genoma se denomina concenso. MAQ permite identificar variantes dentro de una misma posición en el genoma así como la calidad de las mismas.

SMALT<sup>70</sup> es un algoritmo que alinea secuencias cortas a un genoma de referencia, es más eficiente que MAQ y los parámetros que se pueden modificar son más variados que en MAQ. La eficiencia radica en la forma de realizar el mapeo, SMALT construye un índice de  $k$ -meros sobre el cuál ensambla cada una de las secuencias cortas. La creación del índice se puede realizar con un tamaño de  $k$ -mero y sobrelape entre  $k$ -meros definido por el usuario, las variaciones en este parámetro afectan tanto la sensibilidad como la eficiencia de

la búsqueda, entre más pequeño sea el k-mero la búsqueda es más amplia pero consume más recursos computacionales. SMALT no posee funciones que permitan llamar un consenso ni identificar variantes, sin embargo existe un grupo de herramientas denominado samtools<sup>71</sup> que permite post-procesar los datos del alineamiento generado por SMALT y recuperar el genoma consenso, variantes y SNPs.

VELVET<sup>72</sup> es un ensamblador de novo, esto significa que es capaz de ensamblar lecturas pequeñas provenientes de secuenciación masiva, sin requerir un genoma de referencia. Los fragmentos de secuencias resultantes de un ensamblado de novo se denomina contigs. La estrategia que velvet utiliza es indexar cada read en un tamaño de k-mero determinado y denominar a cada uno de estos nodos, después el problema se simplifica en encontrar el camino que visite más nodos basado en el solapamiento de los k-meros y por lo tanto de las secuencias. La calidad del ensamblado final al utilizar velvet depende de varios parámetros, como son: que tan representado está el genoma en los reads, la cobertura de cada contig, el k-mero que se utilice, etc. Para utilizar velvet adecuadamente es necesario hacer una exploración de parámetros en cada grupo de datos ya que la optimización es distinta dependiendo de las características de cada grupo de secuencias a ensamblar.

NT es la base de datos que contiene todos los genomas secuenciados que han sido depositados en NCBI. Para este proyecto utilizamos la base de datos NT publicada en 2011.

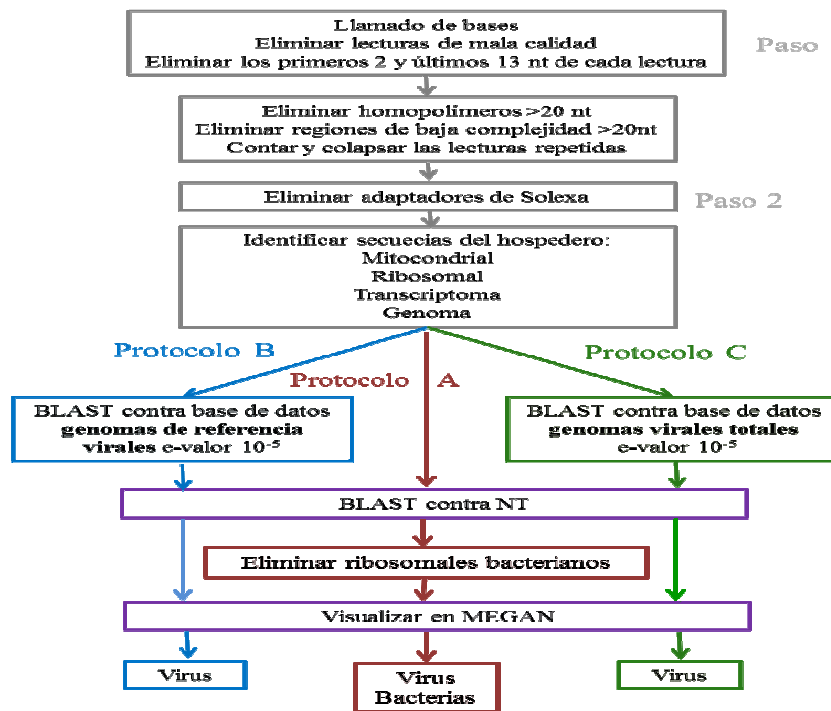
MEGAN<sup>73</sup> es un grupo de programas con un ambiente gráfico que permite visualizar los resultados de BLAST contra la base de datos NT o un subgrupo de la misma utilizando la taxonomía de NCBI. MEGAN utiliza un algoritmo de LCA (lowest common ancestor) que asigna cada lectura a un taxón dado que el nivel taxonómico al que es asignado refleje el nivel de conservación de la secuencia.

Además de visualizar los datos, en MEGAN es posible realizar análisis como medidas de diversidad, análisis de vías metabólicas representadas en el set de datos, entre otros. Es posible comparar varios sets de datos a la vez y aplicar análisis estadísticos sobre los mismos.

## RESULTADOS

Como resultado de la secuenciación obtuvimos entre 13 y 29 millones de lecturas de 75 nucleótidos de longitud para cada una de las muestras, Tabla 5 columna “lecturas totales”. El análisis de estos datos es el tema central de este proyecto.

Durante el desarrollo del proyecto se diseñaron y evaluaron 3 protocolos bioinformáticos para el análisis de los datos: el protocolo A, analiza las secuencias mediante BLAST contra NT (toda la base de nucleótidos de Genebank); el protocolo B, compara las secuencias con la base de datos de genomas de referencia virales; el protocolo C, compara las secuencias con todos los genomas y secuencias virales disponibles. Los protocolos se ilustran en la Figura 11.



**Figura 11:** Diagrama de flujo de los protocolos bioinformáticos

### **Protocolo Bioinformático A**

#### **Paso 1**

A) Filtrar las secuencias por calidad, utilizando sólo aquellas cuya calidad phred sea mayor o igual a 30, es decir que tienen una probabilidad del 99.9% de ser correctas. La calidad de las secuencias en general decae en el ciclo 60, así que decidimos cortar los últimos 13 nucleótidos y los primeros 2 en todas las lecturas, quedándonos así con secuencias de 60 nucleótidos.

B) Eliminar las secuencias con homopolímeros mayores o iguales a 20 nucleótidos, al igual que las secuencias con regiones de baja complejidad mayores o iguales a 20 nucleótidos. Estas secuencias fueron eliminadas ya que se sabe que existen en varios organismos y no podrían ser asignadas a un taxón específico, el corte de 20 nucleótidos o más es porque representan 1/3 de la lectura, con 40 nucleótidos que no sean homopolímero o de baja complejidad sí es posible asignar un taxón, pero con menos la confiabilidad en la asignación decrece.

C) Contar y eliminar las secuencias repetidas, manteniendo a un representante de ellas. El número de secuencias restantes después de aplicar estos filtros para cada carril se muestra en la Tabla 5, en la columna “Después de filtrado 1”.

#### **Paso 2**

Mediante alineamientos usando BLAST identificamos las lecturas que corresponden a los adaptadores de Solexa usados en la secuenciación, así como las secuencias del hospedero (transcriptoma, DNA mitocondrial, RNA ribosomal y genómico). Este primer filtro no identificó a todas las secuencias del hospedero, por lo que adicionalmente utilizamos un filtrado con MAQ y SMALT. Para este análisis utilizamos la versión GRCh37.p2 del genoma humano completo, publicada en Junio de 2011. Ésta contiene 22 cromosomas, genoma mitocondrial, cromosoma X, cromosoma Y, regiones que aún no se ha determinado el cromosoma al que pertenecen y regiones que han sido acomodadas en

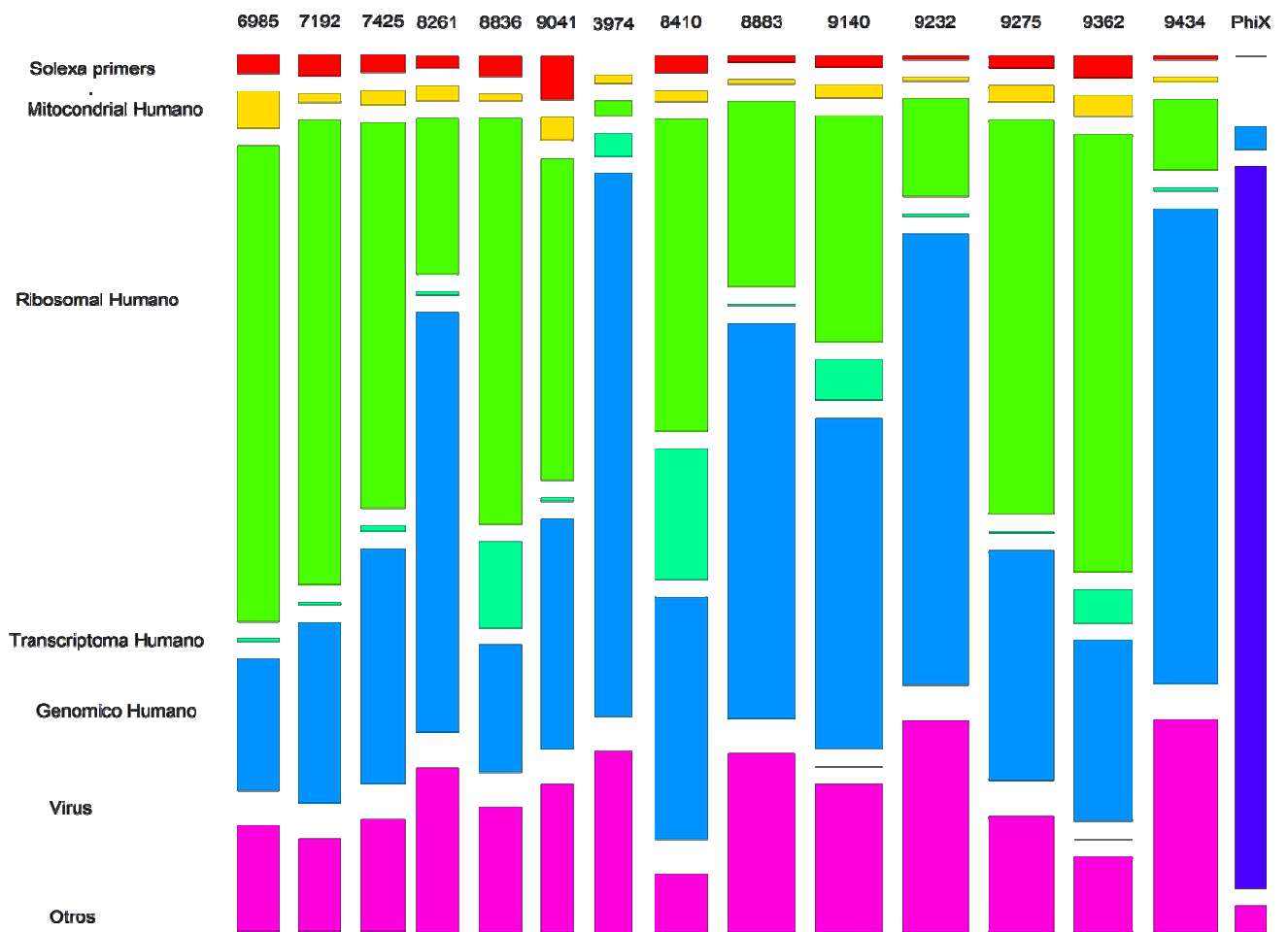


cromosomas equivocados. Las secuencias identificadas se separaron del análisis, su proporción en cada muestra se ilustra en la Figura 12.

Mediante tres rondas de mapeos con MAQ y una ronda posterior de mapeos con SMALT logramos identificar la mayor cantidad de secuencias del hospedero, reduciendo así el total de secuencias que será posteriormente alineado mediante BLAST contra toda la base de datos NT, siendo éste el paso limitante en el tiempo de análisis. El número de lecturas restantes después de estos filtrados se muestra en la Tabla 5 en la columna “Después de filtrado 2”.

Muestra	Lecturas totales	Después de filtrado 1	% del total <sup>a</sup>	Después de filtrado 2
6985	18,369,422	7,208,627	39	490,187
7192	17,990,508	6,356,492	35	432,241
7425	18,455,959	7,889,970	43	536,518
8261	18,061,802	9,629,224	53	770,338
8836	17,934,978	6,329,366	35	NA
9041	13,825,428	7,795,762	56	530,112
3974	15,671,919	10,562,580	67	739,381
8410	22,600,338	5,009,324	22	340,634
8883	29,047,159	19,751,128	68	1,343,077
9140	28,598,135	16,151,859	56	1,098,326
9232	28,068,333	22,646,202	81	1,539,942
9275	28,112,969	12,474,024	44	848,234
9362	25,114,571	7,249,397	29	492,959
9434	27,755,105	22,513,782	81	1,530,937
PhiX	12,842,818	1,306,969	10	72,227
<b>Total</b>	<b>322,449,444</b>	<b>162,874,706</b>	<b>51</b>	<b>10,765,113</b>

**Tabla 5** Resultados de la secuenciación masiva para las 14 muestras y el control PhiX. <sup>a</sup> % del total, es la proporción de lecturas únicas respecto al total de lecturas.

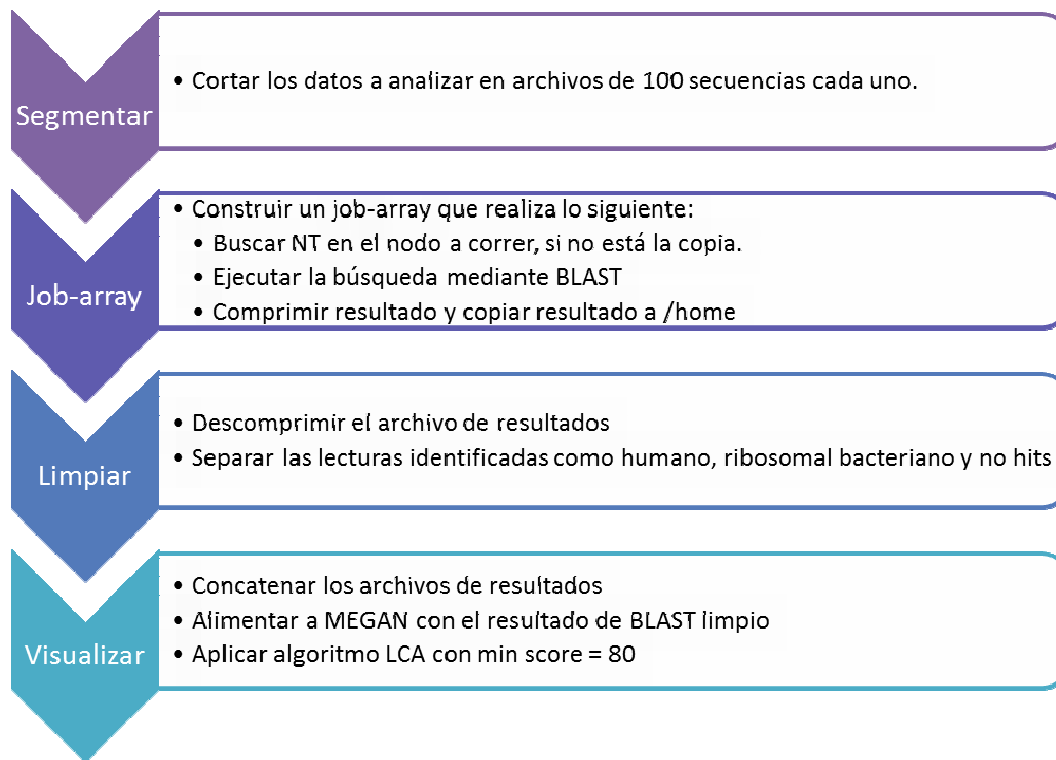


Composición de las muestras

**Figura 12** Filtrado de secuencias en las 14 muestras y la línea control PhiX. Las proporciones se muestran en número absoluto de lecturas.

### Paso 3

Identificar el origen de las lecturas restantes mediante BLAST contra la base de datos NT que contiene todas las secuencias de organismos conocidos disponibles hasta Febrero de 2011; el corte de e-valor fue  $10^{-5}$ . Con el objetivo de optimizar la búsqueda de cada secuencia en la base de datos NT, diseñamos una estrategia paralelizada, aprovechando así la capacidad del cluster del IBT. El diagrama de flujo se muestra en la Figura 13.



**Figura 13:** Proceso de paralización de la búsqueda de cada secuencia en la base de datos NT, utilizando BLAST

Con esta paralización, utilizando 60% de la capacidad del cluster del IBT que son 240 cores, el procesamiento de cada muestra tomó ~3 días.

El formato de resultados de BLAST que utilizamos es el extendido. De estos resultados omitimos aquellas secuencias que tuvieran hits con genoma humano o con ribosomas bacterianos. Separamos del análisis las secuencias que no obtuvieron hits con la base de datos NT.

Visualizamos los resultados de BLAST vs NT en MEGAN4, programa que aplica un algoritmo de LCA (last common ancestor) para asignar a cada lectura, basándose en sus hits de BLAST, a un nivel y grupo taxonómico.

Los virus identificados mediante este protocolo se muestran en la Tabla 6

## **Protocolo Bioinformático B**

### **Los protocolos B y C comparten los Pasos 1 y 2 del protocolo A.**

Dado que el tiempo que tomaría analizar todas las muestras con el diagrama de flujo mostrado en la Figura 8 sería demasiado largo, diseñamos un análisis distinto para los datos con el fin de identificar las secuencias virales presentes en las 14 muestras. Después de los pasos 1 y 2 del protocolo A, comparamos todas las secuencias contra la base de datos de genomas de referencia virales de NCBI (3703 genomas). Las secuencias que dieron hits, las comparamos con toda la base de datos NT para asegurar que los hits que dieron en la primera búsqueda fueran realmente los mejores hits contra todas las secuencias disponibles, así sólo las secuencias que dieran el mismo organismo en ambas búsquedas fueron tomadas en cuenta como verdaderos (Tabla 7).

La probabilidad de que una de estas secuencias sea producto del azar es  $= \frac{1}{4^{60}} = 7.5e-37$ , dado que en cada posición pueden existir cuatro nucleótidos y las secuencias son de longitud 60, las posibles secuencias que se pueden formar con cuatro nucleótidos y de longitud sesenta son  $4^{60}$ , la probabilidad de que una de ellas sea por azar es  $1/4^{60}$ . Por lo que en un universo de ~900 millones de nucleótidos en un carril, la probabilidad de que una secuencia sea idéntica a la de algún organismo de manera azarosa, es muy baja. Finalmente, para visualizar los resultados de BLAST utilizamos MEGAN (Figura 14).

## **Protocolo Bioinformático C**

A partir de la base de datos NT de NCBI construimos una base de datos con todas las secuencias con la palabra “virus” o “phage” en el título, con el objetivo de extender el espacio de búsqueda de secuencias virales. El resultado fue una base de datos con 228,890 secuencias, en la que existe redundancia y un sesgo hacia secuencias de virus patógenos humanos como VIH. Esta base de datos fue utilizada para comparar mediante BLAST cada muestra. Los resultados fueron los mismos que con el protocolo B.

Virus	PhiX	6985	7192	7425	8261	8836	9041	3974	8410	8883	9140	9232	9275	9362	9434
Influenza A H1N1		4	27		6	1	2								
Influenza A H3N2									18		1	1		5	6
Human adenovirus C	2	18	11	22	20	9	48	46	13	2	4	2	3	17	5
Human adenovirus 2	0	1	1	1	5	0	9	5	6	0	1	0	0	1	0
Human adenovirus 6	0	2	0	8	7	1	6	3	5	1	1	1	0	5	0
Aeromonas phage											1				1
Bacillus phage 250			1												
Burkholderia phage	0	0	3	2	0	0	1	1	17	0	1	2	0	23	0
Dengue virus															1
Enterobacteria phage	1	1						2		1					
Enterobacteria phage PhiX	1218001														
Hepatitis B virus						7									
Hepatitis C virus	12														
Human endogenous retrovirus K		16	15	7	10	1	9	27	6	45	45	45	5	11	45
Human herpesvirus 1				33											
Human herpesvirus 4				2			2			3				1	1
Human klassevirus 1				1											
Human papillomavirus								92							10
Human rhinovirus A									6	1	24801				1
Lactobacillus johnsonii prophage					2										
Lactococcus phage															1
Porcine encephalomyocarditis virus								3							
Propionibacterium phage	0	1		1			12		3			9	1	350	6
Pseudomonas phage						5									
Semliki forest virus															1
Simian sapelovirus 2											1				
Staphylococcus phage		71									1			5	
Streptococcus phage									1	1					
Tobacco etch virus															1
Torque teno virus								4							
UR2 sarcoma virus						5									

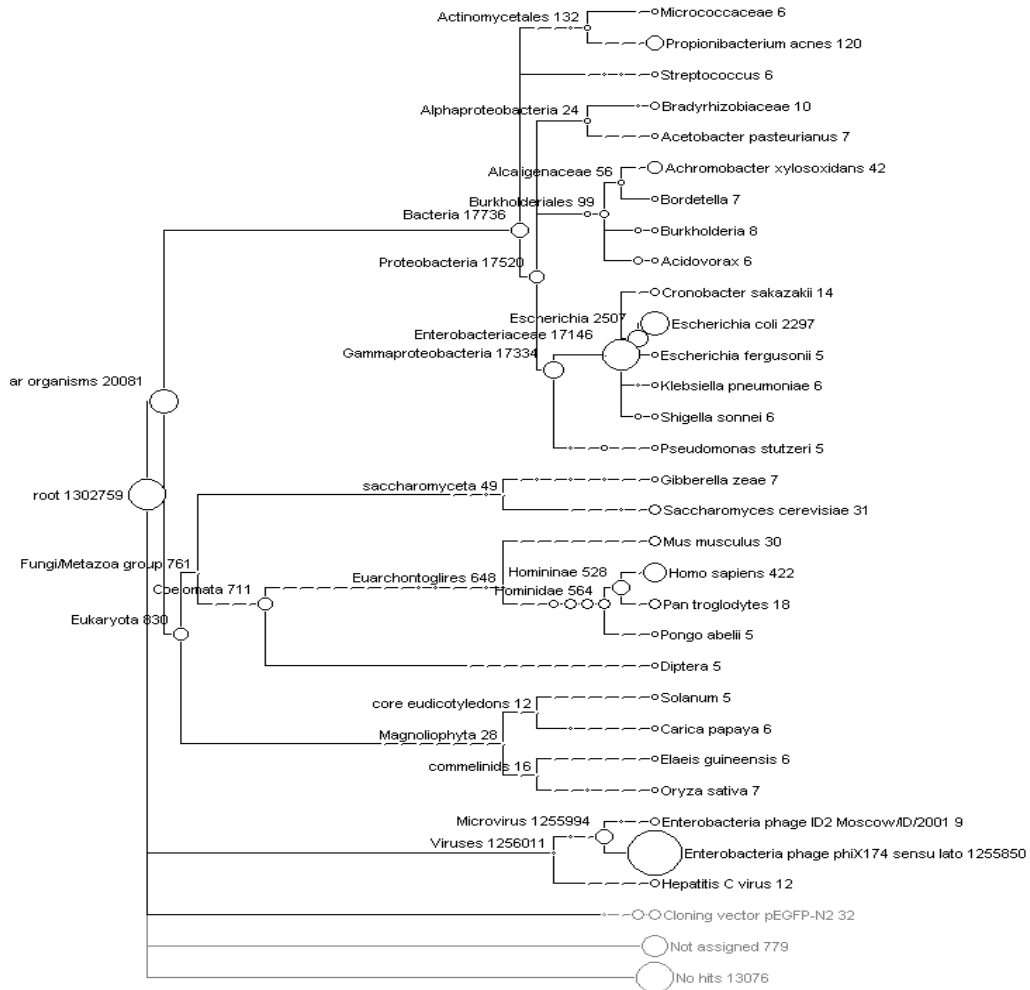
**Tabla 6:** Número de secuencias virales en las 14 muestras y el control PhiX identificadas mediante el protocolo bioinformático A

Virus	PhiX	6985	7192	7425	8261	8836	9041	3974	8410	8883	9140	9232	9275	9362	9434
Influenza A H1N1 2009		2	27	1	12	3	3				1				
Influenza A H3N2 2009-2010									19			1	4	3	6
Human adenovirus C		20	12	45	44	23	68	61	17	4	4	2	5	16	3
Human adenovirus 6		6		6	1			1	3	1				1	
Human adenovirus 2				1	1	4	6	3							
Human adenovirus C		20	12	45	44	23	68	61	17	4	4	2	5	16	3
Human adenovirus 6		6		6	1			1	3	1				1	
Human adenovirus 2				1	1	4	6	3							
Aeromonas phage phiO18P											1			1	
Anelloviridae								2							
Burkholderia phage phiE202				3			1	3	23		5	2	2	26	
Cloning vectors								3							
Ennerobacteria phage T1								2							
Enterobacteria phage f1				1											
Enterobacteria phage M13					1										
Enterobacteria phage phiX174	48704						1								
Enterobacteria phage RB43										1					
Enterobacteria phage T1		1													
Human enterovirus 76											31				
Human herpesvirus 1				39											
Human herpesvirus 4				5			2			3			1	1	
Human herpesvirus 4															1
Human herpesvirus 6					1										
Human herpesvirus 7 strain RK							1								
Human klassevirus 1				1											
Human papillomavirus															1
Human papillomavirus type 112															1
Human papillomavirus type 21								10							
Human papillomavirus type 4								19							
Human papillomavirus type 66															1
Human Rhinovirus A									1		2822				
Human Rhinovirus C											1				
Lactobacillus johnsonii prophage					2										
Lactobacillus phage Lb-338-1													1		
Lactobacillus phage Lrm 1					1										
Molluscum contagiosum virus subtype 1															1
Moloney murine leukemia virus					1										
Propionibacterium phage PA6		6		3	2		42		6			28	4	845	18
Pseudomonas phage DMS3						10									
Pseudomonas phage MP38						2									
Ralstonia phage phiRSA1		1						3	7		3			3	
Roseolovirus						1									
Simian sapelovirus 2											2				
Staphylococcus phage CNPH82		51													7
Staphylococcus phage PH15												1	2	4	1
Staphylococcus phage phi2958PVL											1		1	1	



## PhiX como línea control

Aplicamos el protocolo bioinformático A, a la línea control que contiene al bacteriófago PhiX. PhiX se utiliza como control interno del proceso de secuenciación, con el objetivo de normalizar las intensidades de los clusters, verificar el número de los mismos y las calidades de cada lectura.

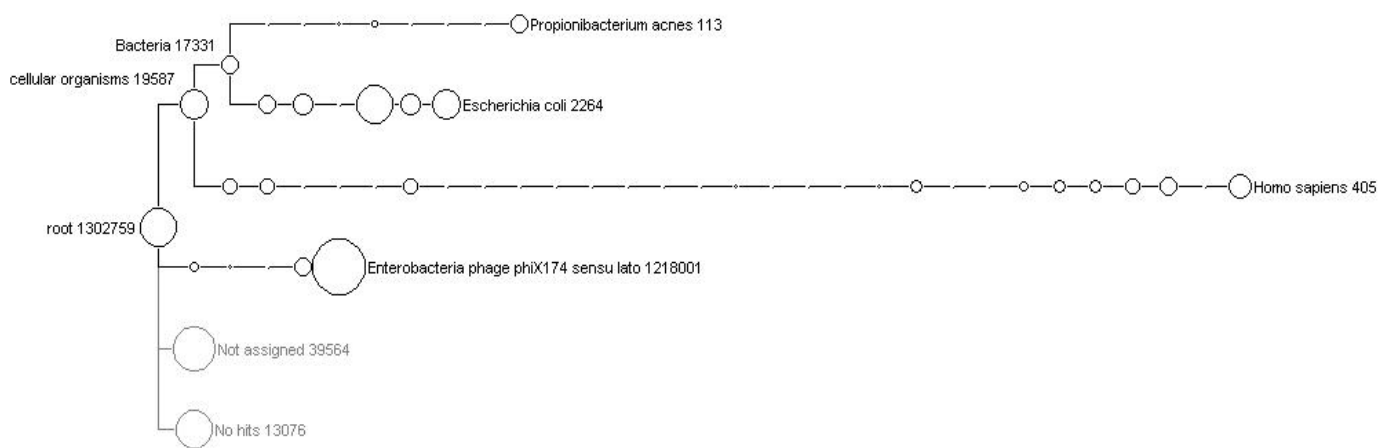


**Figura 15.** Visualización en MEGAN del resultado del protocolo A en la muestra PhiX. Sólo se muestran los organismos que tienen más de 5 lecturas asignadas.



Los resultados del pipeline completo en esta línea control, en la que sólo deberíamos encontrar lecturas de PhiX, se muestran en la Figura 15. Encontramos que el 96.4% de las lecturas únicas pertenecen a PhiX, aunque también detectamos varias secuencias de diferentes organismos:

Si mostramos sólo los organismos que tienen asignadas más de 100 lecturas observamos en la Figura 16 que, además del bacteriófago PhiX sólo encontramos a *Propionibacterium acnes*, *Escherichia coli* y *Homo sapiens* como contaminantes.



**Figura 16.** Línea control PhiX, sólo se muestran los organismos que tienen asignadas más de 100 lecturas.

### Identificación del virus de influenza.

En trabajos anteriores en el laboratorio, en los cuales hemos utilizado el método de secuenciación masiva de Illumina, hemos podido determinar la secuencia completa del genoma de virus de influenza crecidos en embrión de pollo o en cultivo celular<sup>74</sup>. Esto lo hemos logrado ensamblando el genoma viral usando como molde el genoma de un virus de

referencia y el programa MAQ. En el caso de nuestras muestras clínicas no fue posible ensamblar el genoma mediante esta estrategia ya que el número de lecturas de influenza A es muy bajo, aún en las muestras que sabemos fueron positivas por qRT-PCR, por lo que en este caso utilizamos una estrategia diferente en los mismos datos, con la que tampoco recuperamos el genoma de influenza A de estas muestras. Presentamos esta estrategia porque ilustra un problema de mapeo que es importante tomar en cuenta en dato metagenómicos.

La estrategia que probamos consistió en: del set de secuencias únicas ya filtradas realizamos BLASTN utilizando como base de datos el genoma A/Netherlands/2009 con un corte de e-valor  $e^{-05}$ . De esta manera obtuvimos hits para todas las muestras con una extensión de palabra, o alineamiento, de 18-25nt (Figura 17). Con los resultados de BLAST mapeamos cada segmento al genoma de referencia de influenza y pudimos erróneamente ensamblar 22% del genoma (Figura 18). Sin embargo, al hacer BLAST contra todo NT de estos hits, la mayoría eran falsos positivos, ya que alineaban mejor con secuencias humanas o de otros organismo que con influenza, sólo una pequeña porción que eran las lecturas que alinearon 50 nucleótidos o más son las verdaderas positivas. Presentamos estos resultados como evidencia que un alineamiento de 20 de 60 nucleótidos a un genoma de referencia no es suficiente ni adecuado para obtener el genoma de influenza A, sin embargo, como se muestra en la Figura 15, los alineamientos de más de 50 nucleótidos si son confiables.

A continuación construimos una base de datos con todos los genomas de influenza disponibles (~15,000) y, por segmento, colapsamos las secuencias que tuviesen 99% de identidad entre ellas utilizando el programa CD-HIT<sup>75</sup>. Con esta base de datos comparamos todas las lecturas de cada muestra. Por este método obtuvimos los mismos resultados que al hacer la búsqueda utilizando sólo los genomas de referencia de influenza. Así obtuvimos una pequeña cantidad de secuencias específicas para influenza en cada muestra (entre 1 y 20 secuencias por muestra)

Distribución de longitud del alineamiento H1N1 lane4

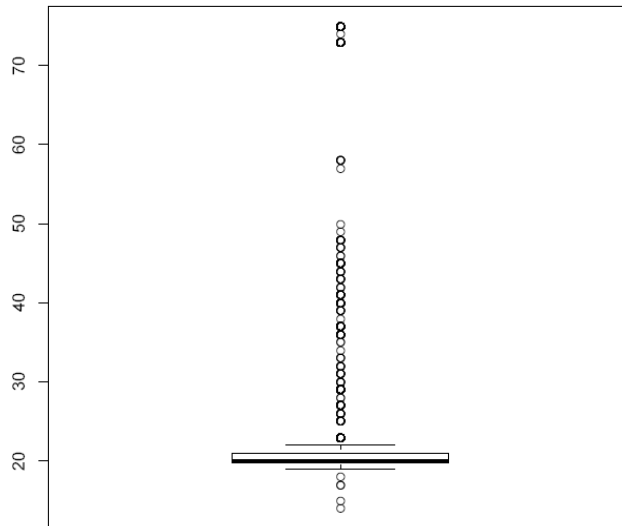


Figura 17 Distribución de longitud de alineamiento de las lecturas identificadas como influenza.

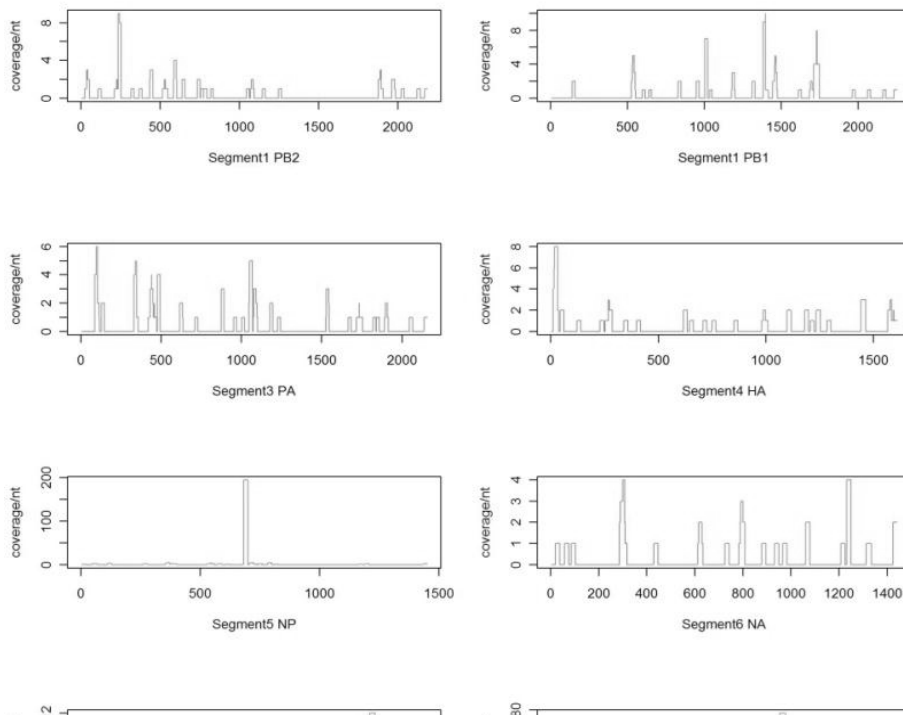
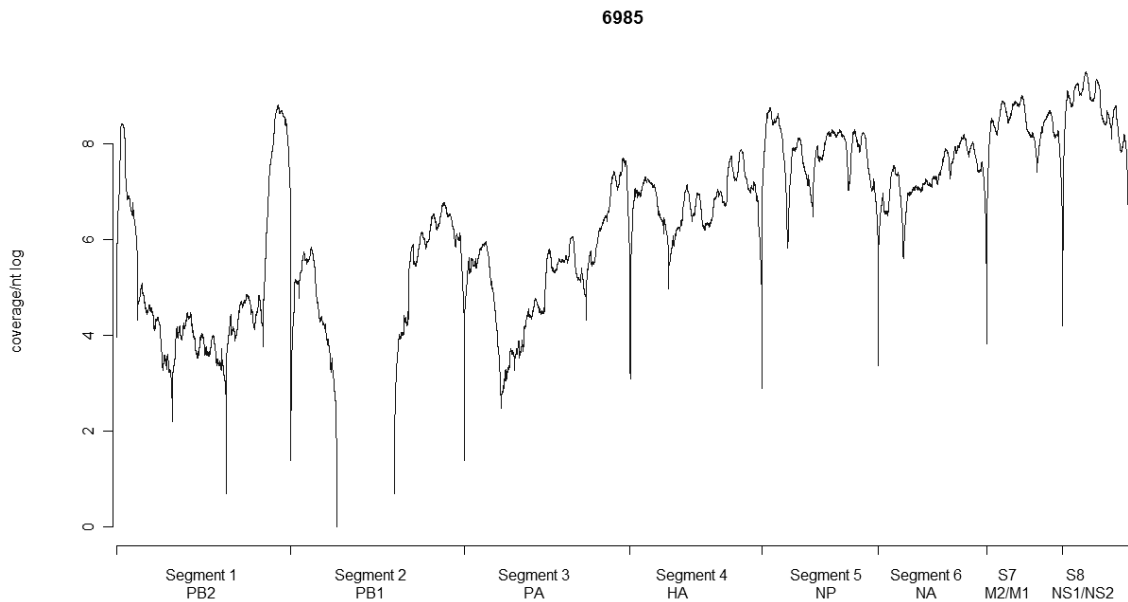
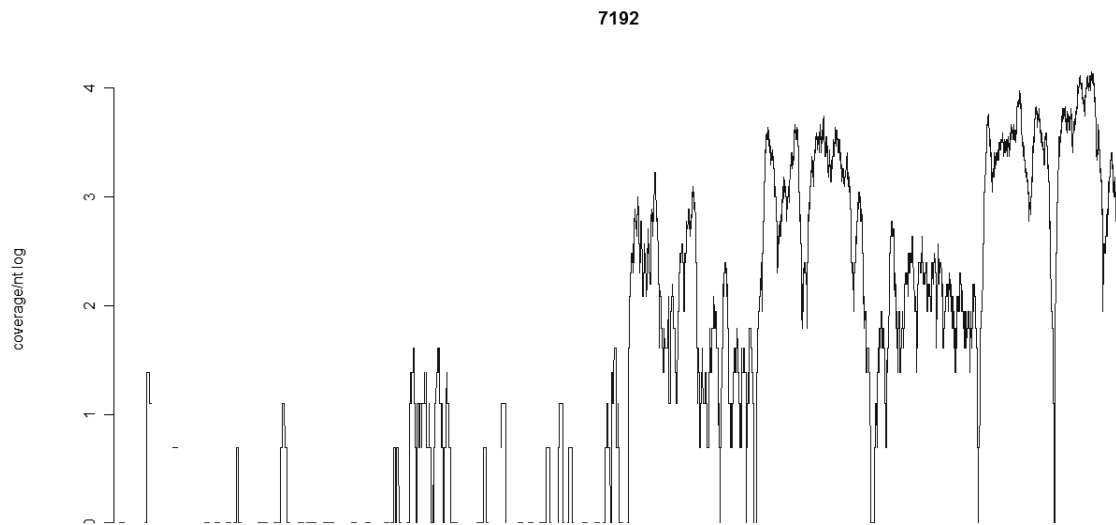
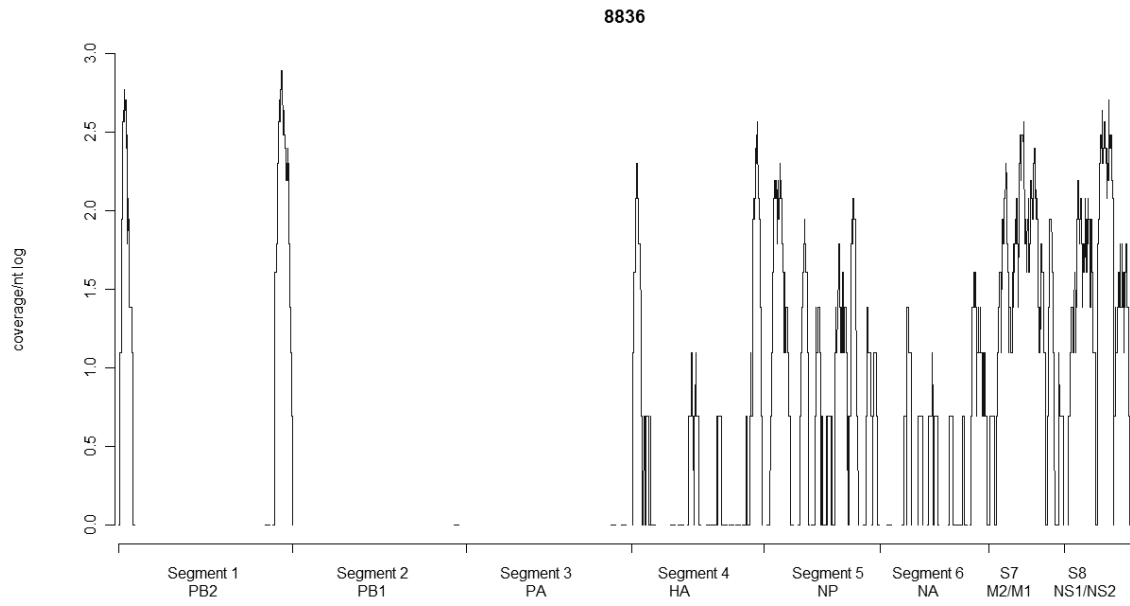


Figura 18 Cobertura errónea del genoma de influenza para la muestra 3974 al ensamblar con fragmentos de 20 nucleótidos.

## Secuenciación del genoma completo de H1N1pdm2009

Con el objetivo de recuperar el genoma completo del virus H1N1pdm2009 presente en nuestras muestras, realizamos una nueva secuenciación masiva en 3 muestras con el siguiente método: Amplificar el genoma completo de influenza A (descrito en la parte de diagnóstico), generar las bibliotecas para secuenciación tipo Illumina, colocar un código de barras de 5 nucleótidos a cada muestra, colocar 5 muestras por carril (multiplexing) y secuenciar con una longitud de 35 ciclos. Ensamblamos cada genoma utilizando MAQ. Como resultado recuperamos un 95% del genoma de influenza A en la muestra 6985, 68% en la muestra 8836 y 44% en la muestra 7192, confirmando así la presencia del virus de influenza A en estas 3 muestras y obteniendo el genoma del virus pandémico aislado de nuestras muestras. En la Figura 19 se ilustra la cobertura obtenida para cada genoma.





**Figura 19:** Coberturas de los 3 genomas de influenza A ensamblados parcialmente provenientes de las muestras 6985, 8836 y 7192.

## Análisis de la secuencia de los virus de influenza

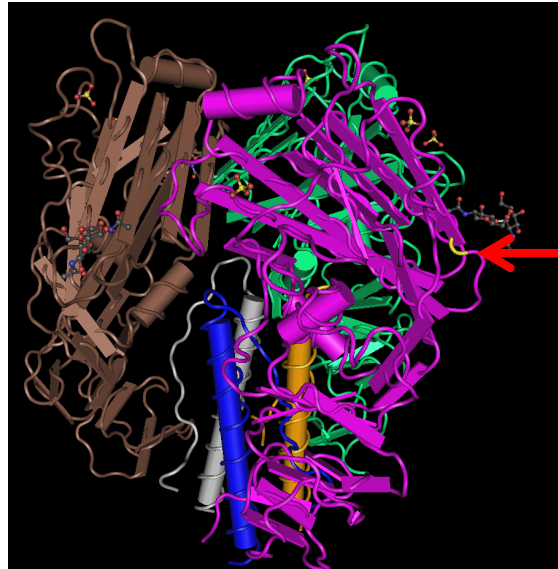
Con el objetivo de analizar a detalle el genoma de la secuencia que obtuvimos del virus de influenza A de la muestra 6985, lo comparamos con las secuencias AH1N1pdm09 disponibles en NCBI aisladas en México. Con estas secuencias realizamos alineamientos e identificamos 5 cambios a nivel de aminoácidos presentes en la muestra 6985 respecto a las demás secuencias, estos cambios son:

HA Asn179Ser

NA Asn248Asp

NS1 Lys66Glu

NS2 Gln2Arg, Leu120Phe



**Figura 20** Estructura de HA. En amarillo se muestra la posición 179, que está implicada en el reconocimiento de ácido siálico.

En las secuencias provenientes de las muestras 7192 y 8836 la mutación Asn179Ser en la hemaglutinina no está presente, ambas secuencias, al igual que el resto de las aisladas en México que se utilizaron para esta comparación poseen asparagina en la posición 179. En la neuraminidasa, la mutación Asn248Asp está presente en las secuencias 6985 y 7192; la secuencia de la muestra 8836 en la posición 248 posee asparagina al igual que las demás secuencias aisladas en México.

En NS1, la mutación Lys66Glu sólo está presente en la secuencia 6985; las secuencias 7192 y 8836, así como las demás aisladas en México poseen ácido glutámico en la posición 66.

## Resumen de diagnóstico molecular de influenza A

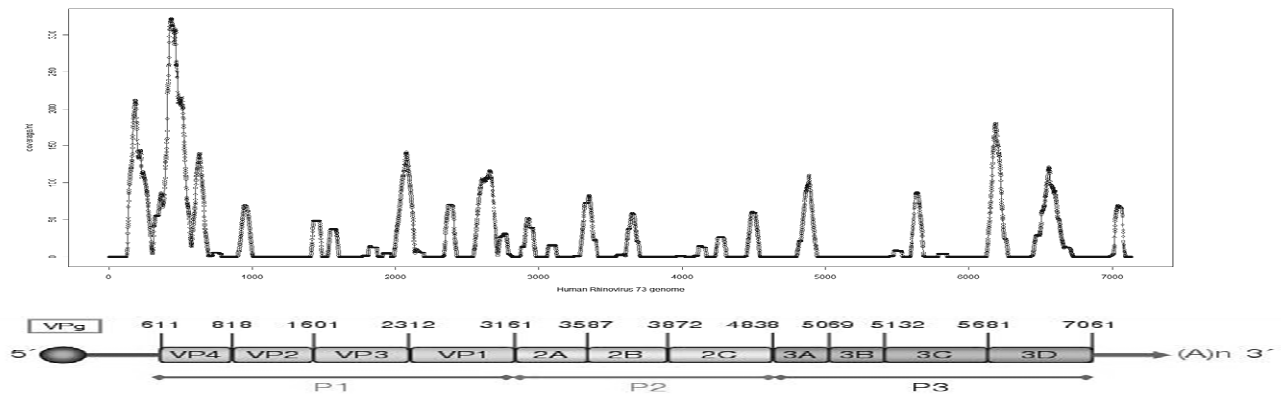
Los diferentes métodos de diagnóstico molecular de influenza A tienen resultados similares entre ellos para 5 muestras, como se ilustra en la Tabla 8. En la muestra 7425 el kit comercial Seeplex no fue capaz de detectar el virus. El número de lecturas que identificamos como influenza A mediante los protocolos B y C es constante, aunque el protocolo A identifica menos lecturas de influenza respecto a los protocolos B y C, excepto en la muestra 6985. Mediante los protocolos de secuenciación masiva pudimos identificar secuencias de virus H3N2 que no fueron detectadas por ninguno de los métodos convencionales. En las muestras 3974 y 8883 no detectamos ningún tipo de virus de influenza.

Muestra	Seeplex	PCR	PCR	Genoma	PipelineB		Pipeline	Pipeline
		INER	Gene M		H1N1	H3N2	A	C
6985	influenza A	Positiva	Positiva	Completo	2	0	4	2
7192	Influenza A	Positiva	Positiva	Parcial	27	0	27	27
7425	Negativa	Positiva	Positiva	NA	1	0	0	1
8261	Influenza A	Positiva	Positiva	Parcial	12	0	6	12
8836	Influenza A	Positiva	Positiva	NA	3	0	1	3
9041	Influenza A	Positiva	Positiva	NA	3	0	2	3
3974	Negativa	Negativa	Negativa	NA	0	0	0	0
8410	Negativa	Negativa	Negativa	NA	0	19	18	19
8883	Negativa	Negativa	Negativa	NA	0	0	0	0
9140	Rinovirus	Negativa	Negativa	NA	1	0	1	1
9232	Negativa	Negativa	Negativa	NA	0	1	1	1
9275	Negativa	Negativa	Negativa	NA	0	4	0	4
9362	Negativa	Negativa	Negativa	NA.	0	3	5	3
9434	Negativa	Negativa	Negativa	NA	0	6	6	6

**Tabla 8.** Resumen de la información sobre el diagnóstico de influenza A en las 14 muestras de este estudio. Los números representan cuántas lecturas únicas se identificaron mediante cada método

## Rinovirus

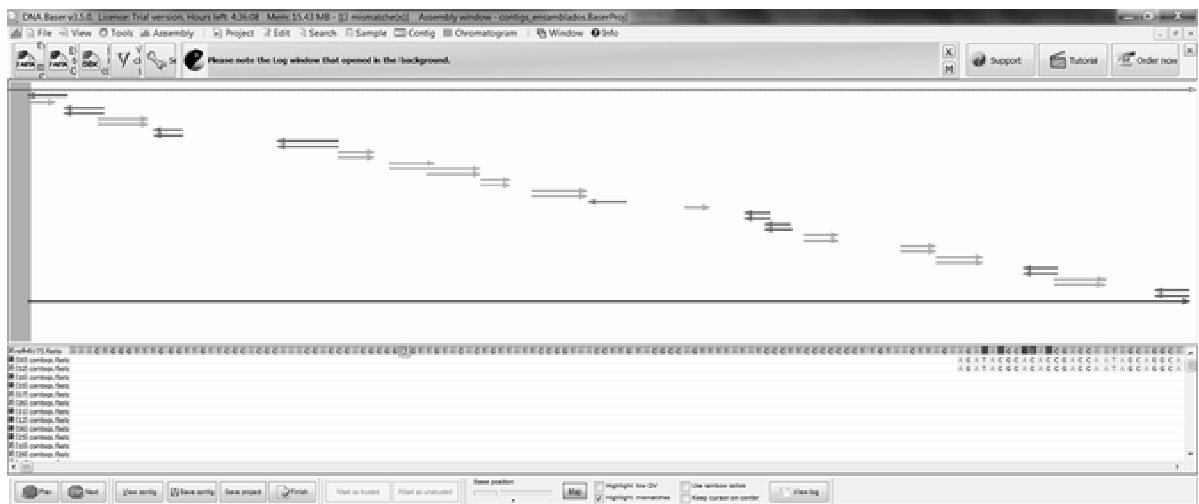
En la búsqueda de secuencias virales con el pipeline B, identificamos en la muestra 9140 ~1600 lecturas únicas asignadas a rinovirus humano. Con el objetivo de recuperar la secuencia del genoma de este virus construimos una base de datos que contenía 122 secuencias de genoma completo de todos los tipos de rinovirus caracterizados hasta Enero de 2012, y la utilizamos como genoma de referencia. Mapeamos todas las lecturas utilizando MAQ, obteniendo 3,202 lecturas únicas, equivalentes a 65,855 lecturas absolutas (2% del total de lecturas del carril), pudiendo así identificar a un rinovirus de humano (HRV) 73 como el virus presente en la muestra. El ensamblado cubrió un 48.8% del genoma con una cobertura de 50X (cada nucleótido fue leído en promedio 50 veces) (Figura 21). También utilizamos el programa BWA<sup>76</sup> para ensamblar el genoma de rinovirus, utilizando la misma base de datos, y obtuvimos 4,195 lecturas únicas equivalentes a 67,014 lecturas absolutas, de las cuales 3679 lecturas pertenecen a HRV 73 y 162 a HRV 13. Adicionalmente, utilizamos SMALT para mapear las lecturas contra la misma base de datos de genomas de rinovirus, mapeando 10,370 lecturas únicas, equivalentes a 252,745 lecturas absolutas.



**Figura21** Cobertura del genoma de rinovirus humano 73, 48.8% del genoma tiene una cobertura de 50X

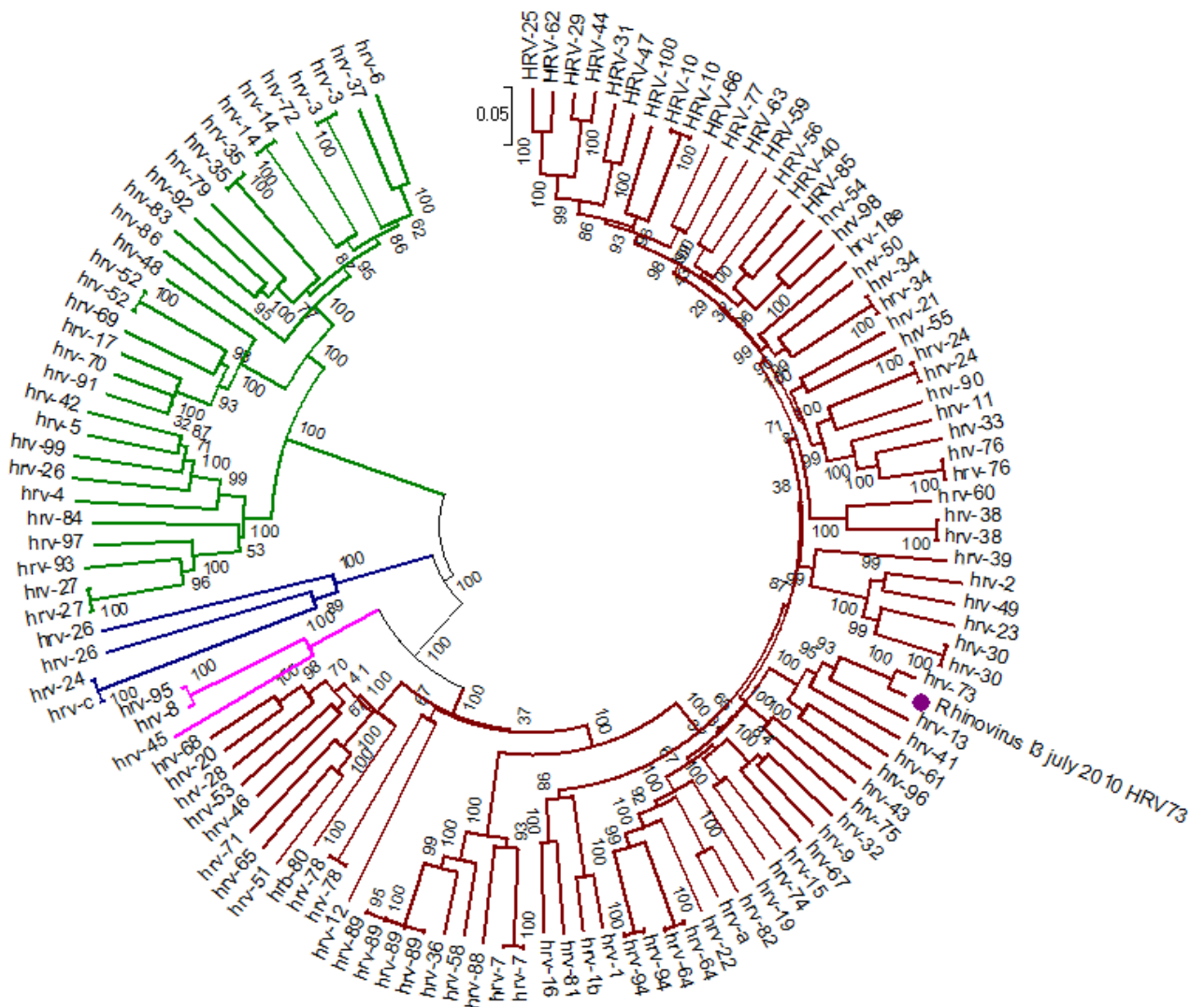


Con el objetivo de ensamblar de novo el genoma de rinovirus, utilizamos Velveth sobre las lecturas preciamente identificadas mediante SMALT, como resultado recuperamos varios contigs que en total suman 4,762 nucleótidos, constituyendo así un 65.7% del genoma total de rinovirus (Figura 22).



**Figura 22.** Contigs resultado del ensamblado de novo mediante Velveth

El genoma de rinovirus 73 está compuesto por 7,140 nucleótidos. Mediante este análisis recuperamos, de la muestra 9140, 65.7% de la secuencia viral con una cobertura de 50X. Con este nuevo genoma y el grupo de 122 secuencias de los subtipos de rinovirus humanos realizamos un análisis filogenético, se muestra en la Figura 23.

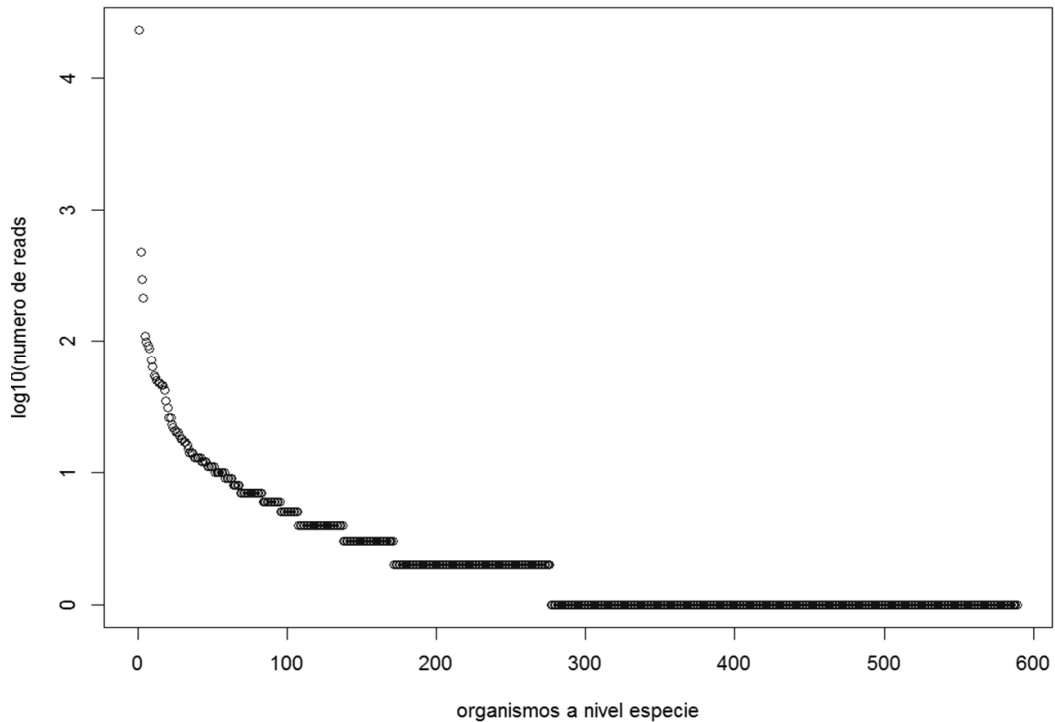


**Figura 23.** Filogenia de rinovirus. Filogenia de los diferentes subgrupos de rinovirus humanos, grupo A en rojo, clado D en rosa, grupo C en azul y grupo B en verde. La secuencia recuperada de la muestra 9140 se muestra con un círculo morado, esta secuencia claramente pertenece al grupo A, el más prevalente en poblaciones humanas.

## Bacterias

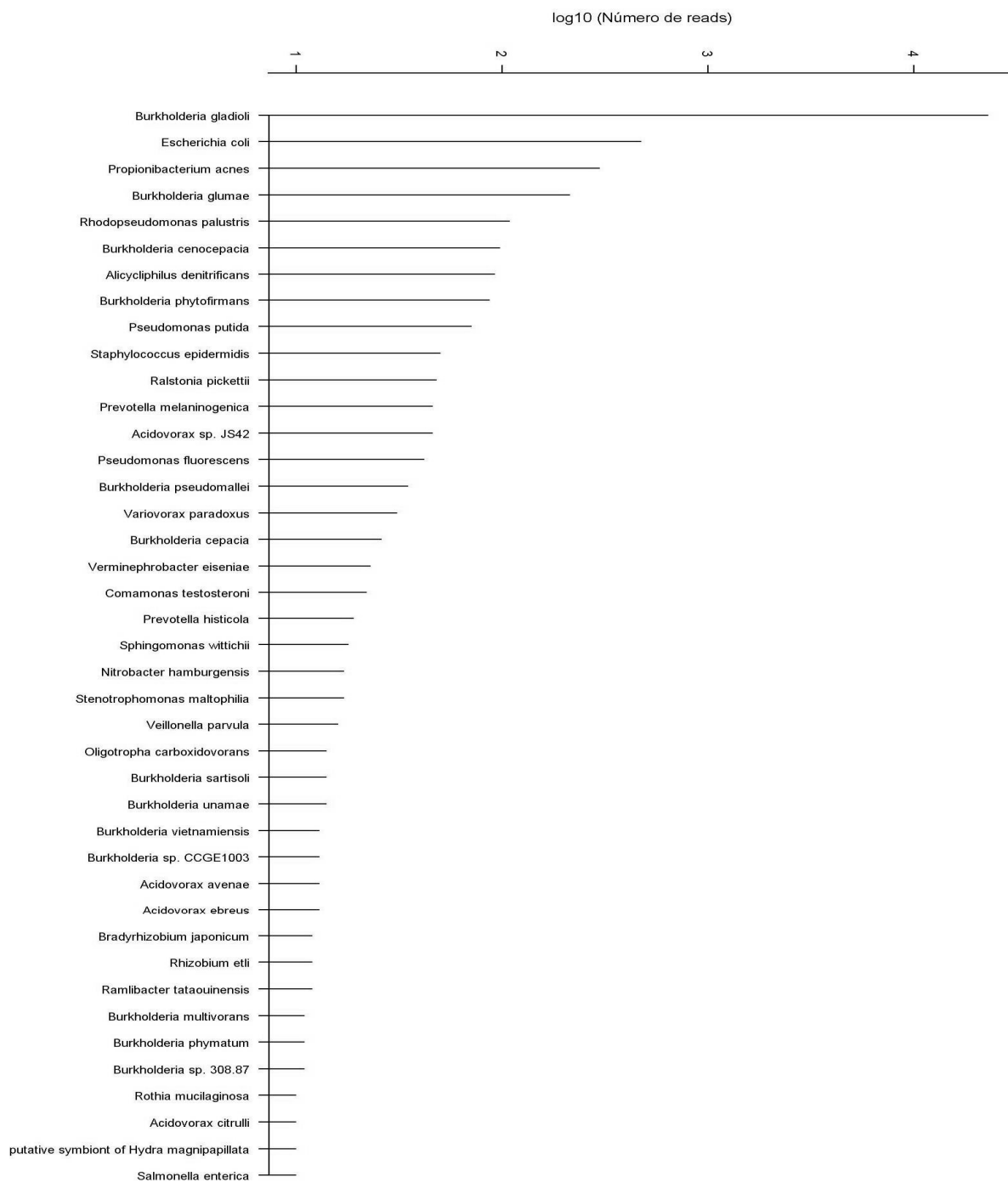
### Abundancia de especies bacterianas.

El pipeline A fue aplicado a todas las muestras y los resultados analizados en MEGAN. A partir de estos datos extrajimos el número de secuencias encontradas para cada especie bacteriana y los graficamos utilizando el programa de análisis estadístico R<sup>77</sup>. La Figura 24 presenta, como ejemplo, la gráfica para la muestra 9041, en la que observamos un comportamiento de pocas especies con más de 100 lecturas y muchas especies con menos de 10 lecturas, lo cual refleja el predominio de algunas bacterias en el ecosistema.



**Figura 24.** Número de secuencias por especie bacteriana para la muestra 9041.

Al analizar este tipo de gráficas para todas las muestras, decidimos sólo considerar para análisis posteriores los organismos que tuviesen 10 secuencias o más. Con este corte generamos nuevamente gráficas de abundancia de especies bacterianas para cada muestra. En la Figura 25 se muestra el caso de la muestra 9041



**Figura25** Especies bacterianas con 10 o más secuencias en la muestra 9041

Las gráficas de cada muestra se proporcionan en el material suplementario (S5).

El número de especies bacterianas identificadas con más de 10 lecturas va desde 19 hasta 176 en cada muestra (ver Tabla 9), la bacteria más abundante en la mayoría de las muestras es *Burkholderia gladioli*. En la muestra 8836 es *Pseudomonas aeruginosa* y en la muestra 8883 *Haemophilus influenzae*, en la Tabla 9 se muestran el primer, segundo y tercer lugar en abundancia bacteriana.

Muestra	# especies	1°	2°	3°
6985	74	<i>Burkholderia gladioli</i>	<i>Staphylococcus epidermis</i>	<i>Prevotella melaninogenica</i>
7192	97	<i>Burkholderia gladioli</i>	<i>Prevotella melaninogenica</i>	<i>Veillonella parvula</i>
7425	44	<i>Burkholderia gladioli</i>	<i>Prevotella melaninogenica</i>	<i>Prevotella denitricola</i>
8261	30	<i>Burkholderia gladioli</i>	<i>Rothia mucilaga</i>	<i>Escherichia coli</i>
8836	107	<b><i>Pseudomonas aeruginosa</i></b>	<i>Burkholderia glumae</i>	<i>Pseudomonas anguilliseptica</i>
9041	41	<i>Burkholderia gladioli</i>	<i>Escherichia coli</i>	<i>Propionibacterium acnes</i>
3974	139	<i>Burkholderia gladioli</i>	<i>Burkholderia glumae</i>	<i>Burkholderia cenocepacia</i>
8410	148	<i>Burkholderia gladioli</i>	<i>Burkholderia glumae</i>	<i>Burkholderia cenocepacia</i>
8883	61	<b><i>Haemophilus influenzae</i></b>	<i>Burkholderia gladioli</i>	<i>Nicotella semolina</i>
9140	118	<i>Burkholderia gladioli</i>	<i>Burkholderia glumae</i>	<i>Moraxella nonliquefaciens</i>
9232	28	<i>Burkholderia gladioli</i>	<b><i>Moraxella catarrhalis</i></b>	<i>Moraxella nonliquefaciens</i>
9275	19	<i>Burkholderia gladioli</i>	<i>Propionibacterium acnes</i>	<i>Burkholderia glumae</i>
9362	176	<i>Burkholderia gladioli</i>	<i>Propionibacterium acnes</i>	<b><i>Moraxella catarrhalis</i></b>
9434	75	<i>Burkholderia gladioli</i>	<i>Veillonella parvula</i>	<b><i>Moraxella catarrhalis</i></b>

**Tabla9:** Especies bacterianas predominantes en cada muestra.

A nivel de especie, en la Figura 26 y en la Figura 27 se muestran gráficas que, dependiendo del tamaño de la letra con la que está escrita cada especie representa la abundancia de la misma, en una escala logarítmica. Los colores corresponden a la clasificación a nivel de clase de cada bacteria.

En la Figura 26 se muestra la información de abundancia a nivel de especie de los pacientes que fallecieron, en la Figura 27 para los vivos.

Especies bacterianas > 10 lecturas en fallecidos



Especies bacterianas > 10 lecturas en vivos



**Figura 26:** Especies bacterianas identificadas en los pacientes fallecidos. Bacterias con >10 lecturas asignadas, las proporciones se muestran en escala logarítmica.

**Figura 27:** Especies bacterianas identificadas en los pacientes vivos. Bacterias con >10 lecturas asignadas, las proporciones se muestran en escala logarítmica.

Generamos un mapa de calor en el que, dependiendo de la intensidad de cada cuadro, es el número de lecturas correspondiente a cada especie bacteriana en una escala logarítmica. Se muestra en la Figura 28.

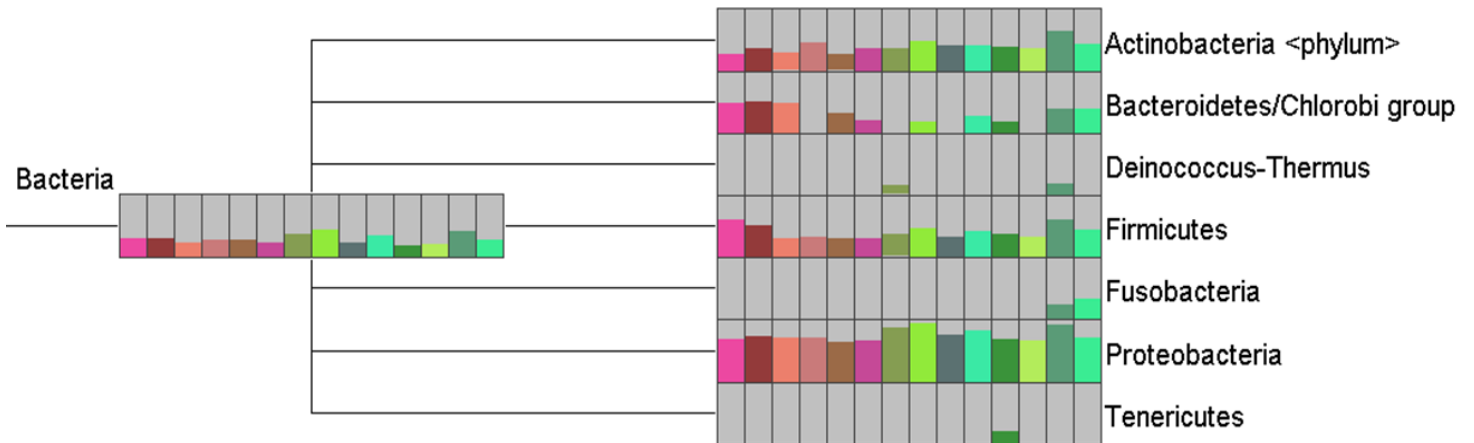
**Figura 28:** Mapa de calor de especies bacterianas identificadas en todas las muestras, la intensidad del color es proporcional a una escala logarítmica del número de lecturas asignadas a cada especie.





## Diversidad bacteriana.

El pipeline A fue aplicado exitosamente en todas las muestras. Como resultado obtuvimos archivos de BLAST en formato extendido y con ellos alimentamos a MEGAN4, que aplica un algoritmo de LCA<sup>78</sup> mediante el cual asigna cada lectura, ya sea a una especie o al nivel taxonómico que logre identificar. La taxonomía de NCBI tiene 35 phyla bacterianos; en nuestras muestras identificamos 4 phyla (Figura 29) comunes a todas las muestras: Actinobacteria, Bacteroidetes, Firmicutes y Proteobacteria.



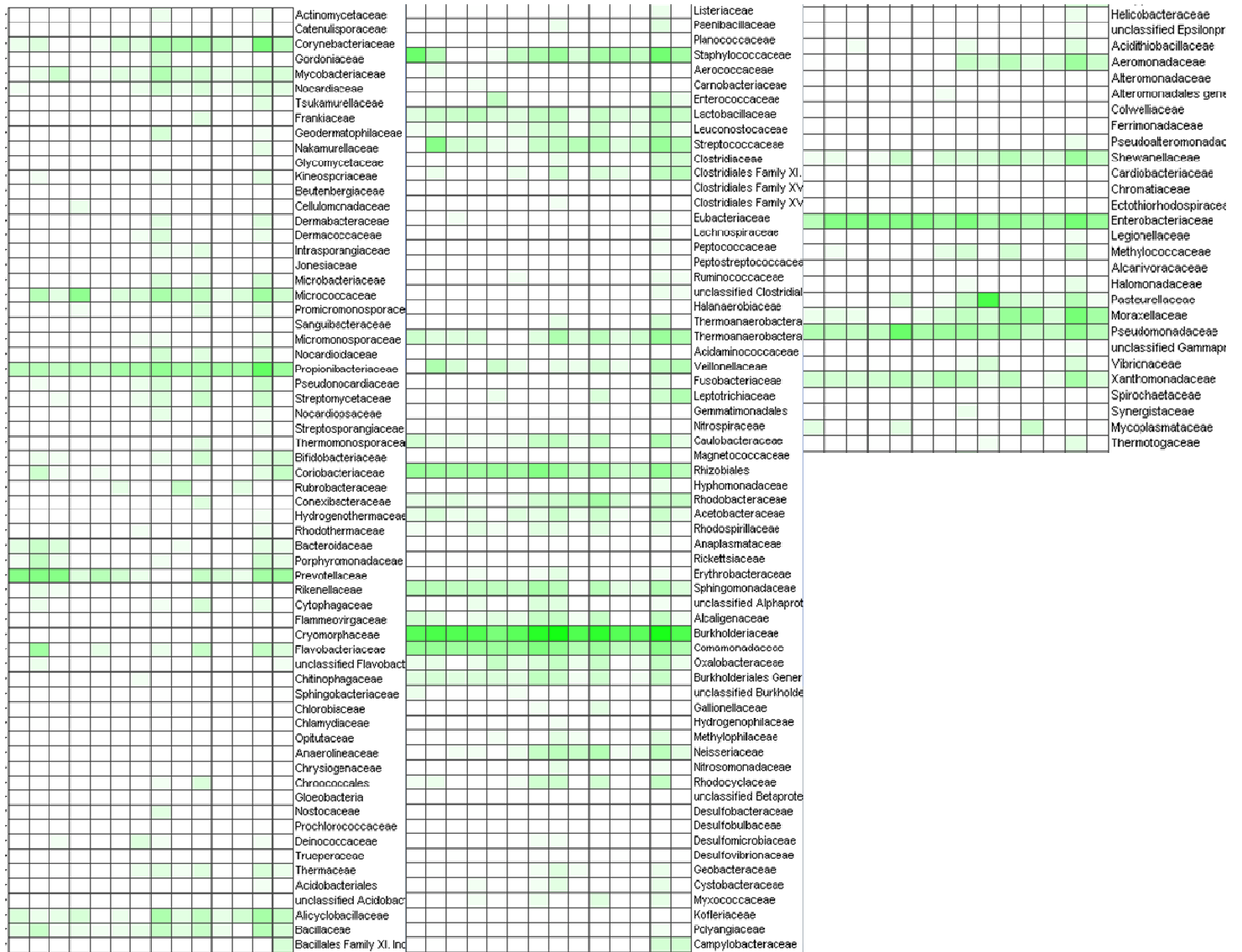
**Figura 29.** Visualización en MEGAN4 a nivel de phylum del número de secuencias asignadas por muestra. Muestras de pacientes que fallecieron en tonos magenta, muestras de pacientes vivos en tonos verdes. Aplicando LCA con mínimo 10 lecturas y un score de 80.

Para el grupo total de muestras, generamos “mapas de calor” a nivel de familia (Figura 30).

Las familias bacterianas, y phylum al que pertenecen, presentes en todas las muestras fueron, en orden de mayor a menor:

- Burkholderiaceae Proteobacteria
- Enterobacteriaceae Proteobacteria
- Propionibacteriaceae Actinobacteria

- Pseudomonaceae Proteobacteria
- Comamonadaceae Proteobacteria
- Streptococcaceae Firmicutes



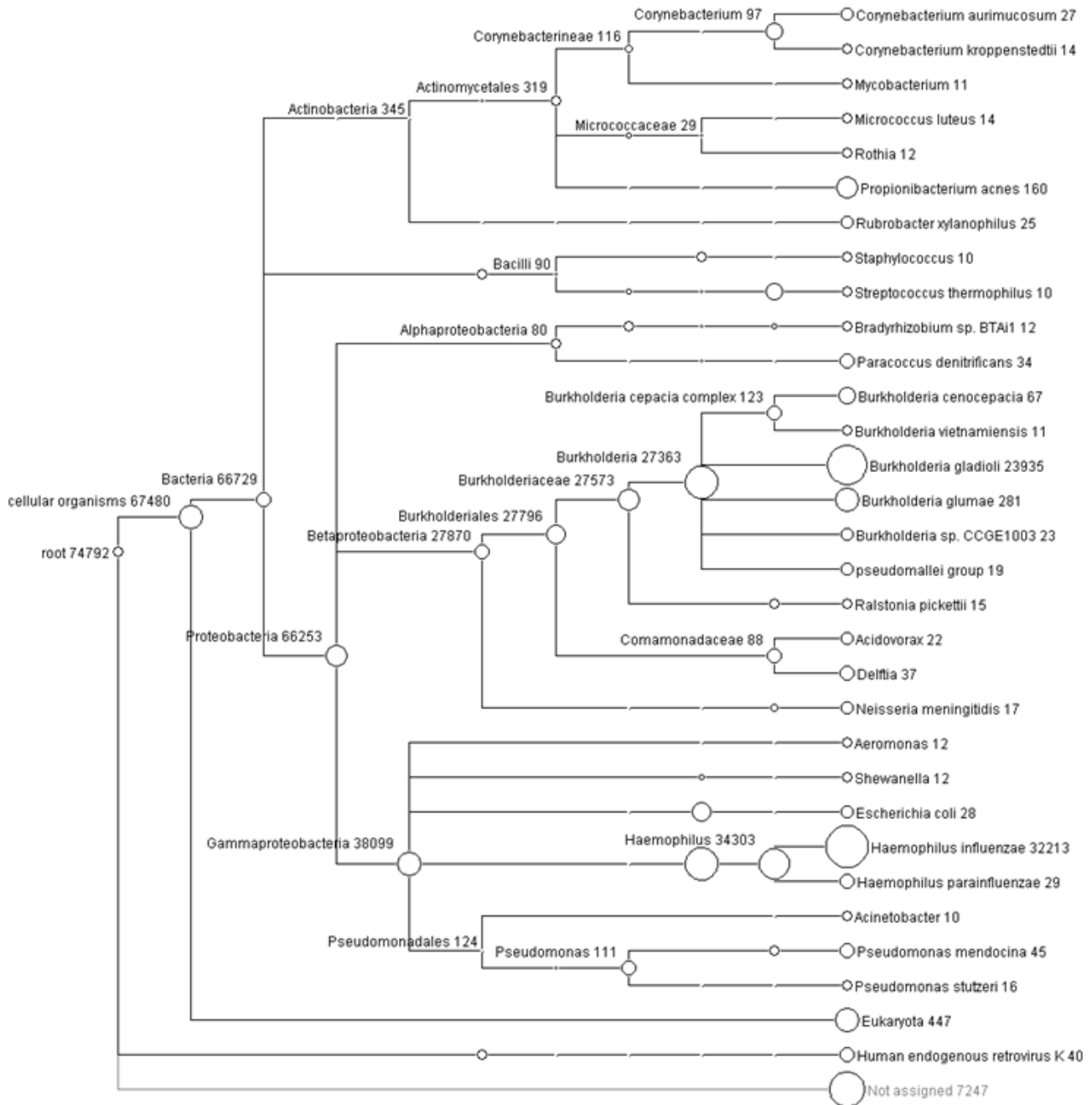
**Figura 30.** Mapa de calor con datos a nivel de familia de las muestras de todos los pacientes, las primeras 6 columnas son los que fallecieron, las 8 siguientes son los vivos.

**Muestra 8883, bacteria predominante *Haemophilus influenzae***

Como resultado del análisis completo de esta muestra, utilizando el protocolo A, visualizamos los resultados de BLAST contra NT en MEGAN (Figura 31). Esta muestra es interesante ya que no encontramos lecturas de influenza A, pero el paciente estuvo internado en la sala general, al aplicar el protocolo A encontramos que la bacteria *Haemophilus influenzae* es la predominante en esta muestra; ésta tiene asignadas 32,213 lecturas únicas, que representan un 48% del total de lecturas asignadas a bacterias.

**Muestra 8836, bacteria predominante *Pseudomonas aeruginosa***

Al aplicar el protocolo bioinformático A en esta muestra, identificamos 5,739 lecturas únicas como *Pseudomonas aeruginosa*, estas lecturas representan el 46% del total de lecturas identificadas como bacterianas.

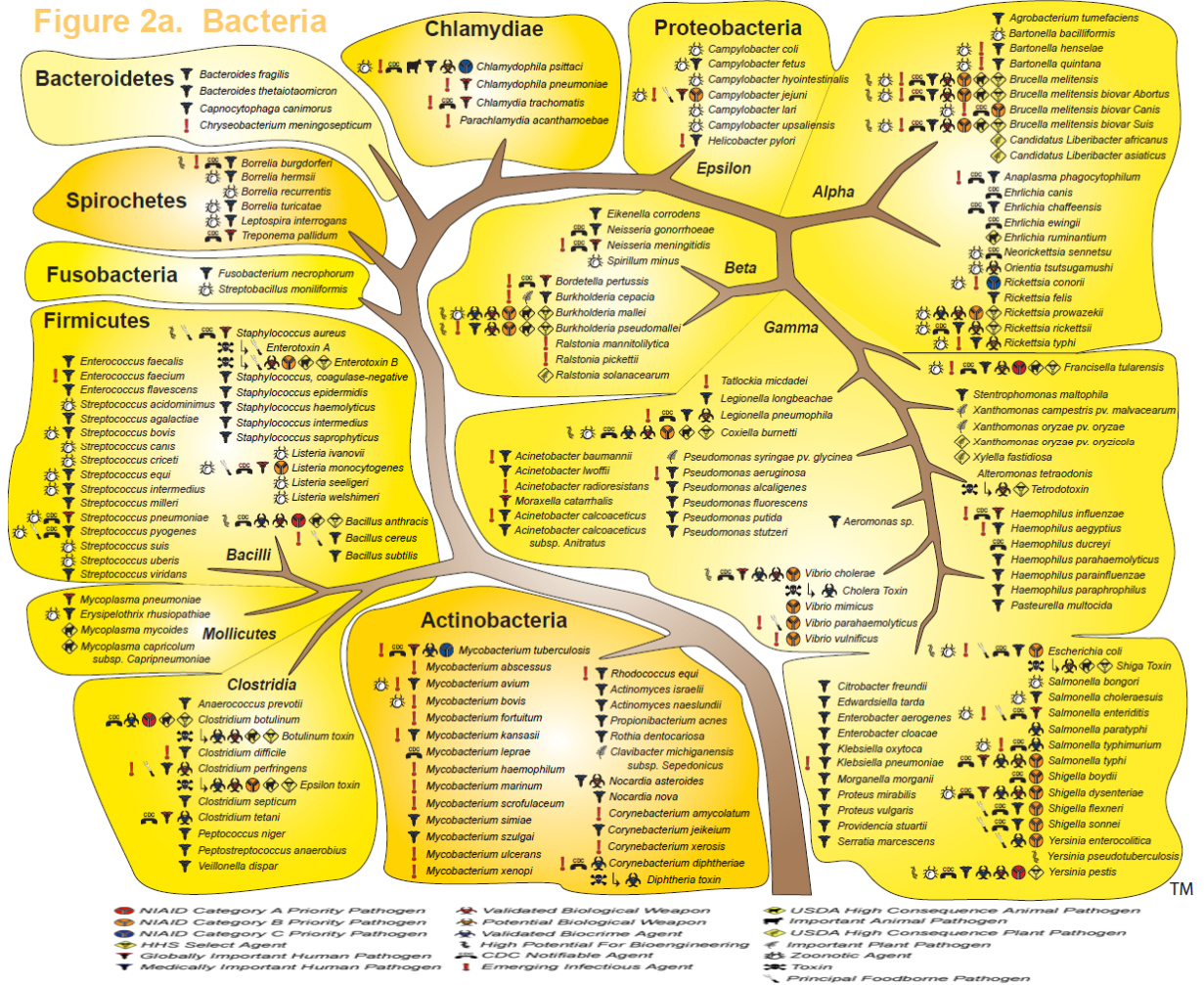


**Figura 31.** Resultados del pipeline A en la muestra 8883, visualizados en MEGAN. Sólo se muestran los organismos con más de 10 lecturas asignadas.

Las figuras de la visualización en MEGAN de los resultados de BLAST contra NT para cada muestra se muestran en el Material Suplementario S6

## Bacterias patógenas

Es de especial interés conocer qué bacterias patógenas están presentes en cada una de las muestras, por lo tanto basados en la literatura<sup>79</sup> definimos un grupo de bacterias patógenas a buscar en nuestros datos. Nos basamos en las definidas como patógenas humanas, que se muestran en la siguiente Figura 32.



**Figura 32:** Bacterias patógenas. Tomado de <http://www.microbialrosetastone.com>

Las bacterias patógenas presentes en todas nuestras muestras se muestran en la Tabla 10. Sólo tomamos en cuenta aquellas que tienen 10 o más lecturas por especie.

Phylum	Especie	6985	7192	7425	8261	8836	9041	3974	8410	8883	9140	9232	9275	9362	9434
Actinobacteria	<i>Corynebacterium jeikeium</i>	0	0	0	0	0	0	0	0	0	0	0	0	48	0
Actinobacteria	<i>Propionibacterium acnes</i>	39	61	33	71	37	153	156	504	160	196	201	168	8621	309
Firmicutes	<b><i>Staphylococcus aureus</i></b>	57	0	0	0	0	0	0	0	0	22	139	0	76	0
Firmicutes	<i>Staphylococcus epidermis</i>	4322	51	0	0	0	26	87	169	0	26	0	0	2130	72
Firmicutes	<i>Staphylococcus haemolyticus</i>	0	0	0	0	0	0	0	0	0	47	0	0	359	40
Firmicutes	<i>Staphylococcus saprophyticus</i>	0	0	0	0	0	0	0	19	0	0	0	0	24	0
Firmicutes	<b><i>Streptococcus pneumoniae</i></b>	0	108	0	0	0	0	0	0	0	0	0	0	31	0
Firmicutes	<i>Anaerococcus prevotii</i>	0	0	0	0	0	0	0	0	0	0	0	0	18	24
Firmicutes	<b><i>Mycoplasma pneumoniae</i></b>	0	0	0	0	0	0	0	0	0	0	20	0	0	0
Proteobacteria	<b><i>Burkholderia cepacia</i></b>	91	179	131	115	310	73	1314	3749	123	1920	193	154	5097	306
Proteobacteria	<b><i>Burkholderia pseudomallei</i></b>	0	10	0	0	0	0	54	187	0	150	0	0	150	0
Proteobacteria	<b><i>Neisseria meningitidis</i></b>	0	0	0	0	0	0	13	21	17	23	0	0	11	0
Proteobacteria	<i>Ralstonia pickettii</i>	39	31	14	19	25	33	146	110	15	43	9	11	137	9
Proteobacteria	<i>Acinetobacter baumannii</i>	0	0	0	0	0	0	0	12	0	61	0	0	168	19
Proteobacteria	<i>Acinetobacter calcoaceticus</i>	0	0	0	0	0	0	0	0	0	14	0	0	27	0
Proteobacteria	<i>Aeromonas sp.</i>	0	0	0	0	0	0	0	25	12	52	0	16	268	20
Proteobacteria	<i>Enterobacter cloacae</i>	0	0	0	0	0	0	0	13	0	22	0	0	99	11
Proteobacteria	<i>Escherichia coli</i>	0	295	311	252	267	266	27	446	28	98	19	29	517	66
Proteobacteria	<b><i>Haemophilus influenzae</i></b>	0	0	0	0	0	0	0	27	32213	11	0	0	17	0
Proteobacteria	<b><i>Haemophilus parainfluenzae</i></b>	0	0	0	0	0	0	0	0	29	0	0	0	31	0
Proteobacteria	<b><i>Klebsiella pneumoniae</i></b>	0	0	0	0	0	0	0	0	0	30	17	11	208	31
Proteobacteria	<b><i>Moraxella catarrhalis</i></b>	0	0	0	0	0	0	0	0	0	192	299	0	2251	132
Proteobacteria	<b><i>Pseudomonas aeruginosa</i></b>	0	0	0	11	5773	0	22	63	55	70	10	10	205	12
Proteobacteria	<b><i>Pseudomonas fluorescens</i></b>	33	18	17	12	13	21	148	25	0	28	0	0	50	10
Proteobacteria	<i>Pseudomonas putida</i>	14	12	0	0	14	18	23	31	0	31	0	0	58	0
Proteobacteria	<i>Pseudomonas stutzeri</i>	0	0	0	0	0	0	9	25	16	20	0	0	110	0

**Tabla 10** Bacterias patógenas presentes en las muestras. Se indica el número de secuencias encontrado para cada especie. Las especies marcadas en negritas se asocian típicamente a enfermedades respiratorias.

# DISCUSIÓN

## **Sobre la metodología.**

Este estudio nos ha permitido explorar la diversidad tanto viral como bacteriana en muestras aisladas de personas con enfermedad respiratoria de diferente intensidad. La ventaja al ser un estudio metagenómico es que nos permite explorar más allá de los patógenos asociados tradicionalmente a enfermedades respiratorias; también nos permitió conocer un poco sobre el microbioma que reside en el tracto respiratorio humano.

Al secuenciar los ácidos nucleicos totales con el objetivo de conocer la composición microbiana de cada muestra logramos identificar bacterias y virus simultáneamente en un solo ensayo, esto sin la necesidad de tener información a priori de cuáles de estos organismos estarían presentes en cada muestra. El diagnóstico molecular clásico se basa en el diseño de sondas específicas para amplificar el material genético de un organismo o grupo de organismos, esta aproximación refleja un sesgo en el tipo de organismos que se detectan ya que sólo los organismos hacia los que está dirigida la búsqueda son detectados. En este aspecto, nuestra estrategia metagenómica brinda información sobre el ecosistema representado en cada muestra y no sólo sobre los virus o bacterias patógenos que se busquen.

El estado del arte de la metagenómica mediante secuenciación masiva aún es un enfoque sesgado, ya que sólo las secuencias que presentan una homología con las ya conocidas son detectadas, este enfoque depende completamente de la representación de organismos de los cuales ya se tiene secuencia. Mediante este enfoque en el análisis de los datos no es posible detectar a los organismos que aún no han sido secuenciados o aquellos que aún no han sido descubiertos, sin embargo si sus secuencias están presentes en la muestra, posteriormente podrían desarrollarse nuevas estrategias no basadas en búsqueda por homología para elucidar su origen.

Al comparar la metodología y tipo de tecnología de secuenciación utilizada en este trabajo con la tecnología tradicionalmente utilizada para estudios metagenómicos, que es la pirosecuenciación, concluimos que son tecnologías complementarias ya que cada una tiene distintas bondades. Mientras la tecnología Illumina brinda mayor profundidad la pirosecuenciación aporta fragmentos de DNA más largos e informativos.

La pirosecuenciación en la plataforma 454 fue el método pionero en estudios metagenómicos ya que al obtener tamaños de lecturas de hasta 500 pares de bases es más sencillo ensamblar cada lectura en contigs más grandes y realizar búsqueda de genes. En estudios metagenómicos cuyo objetivo es elucidar las vías metabólicas de un ecosistema la plataforma 454 es la indicada para ello. Actualmente el instrumento GS FLX Titanium XL+ es capaz de secuenciar hasta un millón de lecturas con una moda de 700 pares de bases. Mediante esta plataforma se ha logrado la identificación de virus previamente desconocidos como el caso del virus "lujo"<sup>42</sup> en muestras clínicas de pacientes con fiebre hemorrágica. Un caso más de éxito es el descubrimiento de un arenavirus<sup>80</sup> en pacientes que recibieron trasplantes de diversos órganos.

En el descubrimiento de virus una de las ventajas que aporta la plataforma 454 es la longitud de las lecturas que permite rescatar la secuencia del genoma viral con mayor facilidad. Entre las desventajas, comparado con la plataforma Illumina, se encuentran el costo y la poca profundidad que se obtiene.

### **Sobre la obtención y preparación de muestras**

La obtención y preparación de las muestras es uno de los primeros pasos en los que el método se puede mejorar. Al analizar los datos de este proyecto, el DNA humano presente en cada muestra representa cerca de un 80%, por lo que sólo 30% de las lecturas obtenidas mediante secuenciación masiva son de utilidad en la identificación de microorganismo presentes en el tracto respiratorio. Al tratar las muestras con DNAsa I, previo a la extracción de ácidos nucleicos, es posible eliminar la mayoría del DNA humano, aunque en este tipo de tratamientos también se perdería el DNA bacteriano. Si el enfoque es completamente viral, sería conveniente además de tratar con DNAsas utilizar filtrado de la muestra por tamaño, para sólo dejar pasar a las partículas virales. Otra opción es eliminar el RNA ribosomal al pasar la muestra por columnas que tiene sondas complementarias a regiones ribosomales, así estas secuencias quedarán unidas a la columna y la demás muestra quedará limpia de las mismas.

Entre otras consideraciones, es importante tener datos clínicos certeros y detallados de cada uno de los pacientes, tales datos deberían incluir placas torácicas, factores de riesgo,



factores demográficos, tratamiento administrado, días de enfermedad y progresión. Al contar con estos datos, los resultados del tipo de virus y bacterias muestreados podrían correlacionarse con los datos clínico, brindando la posibilidad de encontrar algún factor que propicie la colonización por cierto tipo de microorganismos. Es importante conocer, en el caso de las muestras de lavados broncoalveolares, el lugar específico en los pulmones de donde la muestra fue tomada, ya que en estudios previos se ha observado que existe una flora bacteriana distinta en cada localidad pulmonar<sup>81</sup>. Al no contar con estos datos, la correlación entre el tipo y abundancia bacteriana encontrados no se puede contrastar entre muestras al nivel de detalle que lo hacen otros estudios.

Es también necesario, contar con controles en cada uno de los pasos, desde la obtención de muestras, es importante contar con una muestra de sólo las soluciones que se utilizaron ya sea en el lavado broncoalveolar o en el lavado nasofaríngeo, tal muestra tratada al igual que las obtenidas de los pacientes nos daría información sobre posibles contaminantes en el proceso de obtención de las muestras.

También es deseable contar con un control de una persona saludable, la muestra debía ser tomada en la misma fecha y lugar que el de los pacientes enfermos, podría ser algún familiar que acompañase al paciente. Este tipo de muestras, sobre todo en el caso de los lavados broncoalveolares, es difícil de obtener ya que la toma de muestra es un procedimiento invasivo.

### **Sobre los protocolos bioinformáticos.**

El diseño detrás de los tres protocolos bioinformáticos presentados en este trabajo se basa en la idea de identificar, mediante homología, el origen de cada una de las secuencias obtenidas mediante secuenciación masiva. Este objetivo es complejo, ya que los distintos mecanismos evolutivos a los que están sujetos todos los seres vivos han resultado en grupos de genes compartidos por varios organismos, en genes adquiridos por transferencia horizontal, en regiones repetitivas y de baja complejidad. Intentamos tomar en cuenta estos mecanismos al determinar el origen de cada secuencia, sin embargo aún es necesario explorar esta área.

El primer paso es asegurar que las secuencias son confiables, es decir que la calidad de cada nucleótido es suficiente para utilizarlo, en este estudio sólo trabajamos con secuencias en las que el estimado de confianza fuese mayor al 99.9% por esto eliminamos las secuencias de baja calidad así como los primeros dos y últimos 13 nucleótidos de cada lectura en los que la calidad decrecía drásticamente.

Las secuencias de homopolímeros están presentes en muchos organismos y es imposible identificar a cuál de ellos pertenecen, por lo que por la nula información que podrían brindar a una asignación taxonómica fueron eliminadas, el mismo caso se aplica para las secuencias de baja complejidad. Elegimos un corte de si 1/3 o más nucleótidos de cada lectura era homopolímero o de baja complejidad esta fuera eliminada del análisis. Este corte lo determinamos mediante pruebas empíricas en las que observamos que con 40 o más nucleótidos es posible asignar un nivel taxonómico a la lectura, pero con menos de 40 nucleótidos era problemático.

En las primeras exploraciones de una muestra encontramos que existía un número importante de lecturas idénticas. Tales lecturas pueden tener dos orígenes: el primero es ampliaciones de un fragmento mediante PCR (duplicados de PCR); el segundo es que en la muestra realmente existiesen más de una secuencia idénticas. Es indistinguible determinar cuál de estas posibilidades dio origen a los duplicados, para hacerlo sería necesario acompañar cada experimento de un PCR cuantitativo para cada secuencia. Decidimos colapsar las lecturas idénticas para reducir el tamaño del grupo de datos y mantuvimos el conteo de las lecturas colapsadas. Observamos que las proporciones de organismos predominantes son independientes del número de duplicados de PCR.

Las proporciones de adaptadores de Solexa que se concatenan entre ellos y constituyen una lectura fueron bajas en todas las muestras, eliminarlos redujo un poco el set de datos.

Las secuencias del hospedero en cada muestra son más del 80%, esta proporción, para futuros análisis debe disminuir al pre-procesar las muestras como ya se discutió al inicio de esta sección. Es necesario reducir las secuencias del hospedero ya que estas secuencias reducen la proporción de lecturas microbianas.

Una de las consideraciones importantes en el diseño de los protocolos bioinformáticos es la noción de garantizar, en la medida de lo posible, que la asignación de cada lectura es la mejor respecto al universo de organismos secuenciados. Un ejemplo es una lectura en la que al compararla sólo con la base de datos de genomas virales resulta en un alineamiento de 40 de 60 nucleótidos al genoma de influenza, pero al compararla con todo NT, alinean 60 de 60 nucleótidos en el genoma humano. En este caso si sólo limitamos la búsqueda a una base de datos viral corremos el riesgo de obtener varios falsos positivos. En el protocolo A garantizamos los mejores hits de cada lectura al hacer la búsqueda en NT, en los protocolos B y C, sólo son asignadas como verdaderos positivos aquellas lecturas que tanto en la base de datos de genomas virales, como en NT su mejor hit es el mismo. Si una lectura alinea con un virus, pero al buscarla contra NT alinea mejor con otro organismo, es asignada al segundo organismo.

Este tipo de consideraciones es esencial es un acercamiento metagenómico, porque debido al tamaño y diversidad de cada set de datos es riesgoso comparar contra sólo un genoma de referencia que puede resultar en un ensamblado ficticio.

El corte por e-valor que utilizamos de e-5 en todas las búsquedas lo determinamos mediante la experiencia y revisión de la literatura, ya que este corte ha sido utilizado en otros reportes<sup>32</sup> e incluso se han utilizado e-valores menos estrictos como e-3.

El uso de MEGAN en las asignaciones taxonómicas nos permitió determinar el origen de cada lectura al nivel taxonómico que fuese confiable en cada caso.

## **PhiX**

El bacteriófago PhiX que se utiliza como control interno de la secuenciación masiva es una muestra pura proporcionada por la compañía Illumina: generalmente en cada corrida se coloca PhiX en un carril con el objetivo de extraer información sobre intensidades, calidad de los clusters formados, concentración del DNA utilizado, etc. En nuestro caso sólo se colocó PhiX en una de las dos corridas y lo utilizamos también como control de nuestros protocolos bioinformáticos. En el protocolo B obtuvimos ~40,000 secuencias únicas de PhiX, en el protocolo C ~70,000 secuencias únicas y con el protocolo A ~1.2 millones de

secuencias únicas que representan un 96.4% de las lecturas totales de ese carril. De estos datos podemos comprobar que tanto el protocolo B como el C están subestimando la cantidad de secuencias virales, el protocolo A fue capaz de recuperar todas las secuencias de PhiX contenidas en la muestra, por lo tanto como método de identificación de secuencias virales, el protocolo A recupera una mayor cantidad de información.

Además de encontrar secuencias de PhiX como era de esperarse, usando el protocolo A, pipeline ~13,000 lecturas no dieron hits contra la base de datos NT, 779 lecturas no fueron asignadas a ningún organismo ya que su identidad fue muy baja, 568 fueron humanas y 45,418 fueron sorprendentemente asignadas a otros organismos. Un total de 27 organismos fueron identificados con más de 5 lecturas, entre los que destacan *Propionibacterium acnes* (133 lecturas), *E. coli* (17,520 lecturas), *Alcaligenaceae* (56 lecturas), vectores de clonación (105 lecturas), así como de particular curiosidad, 12 lecturas de Rhizobiales y 12 lecturas del virus hepatitis C. Se ha aislado *Propionibacterium acnes* de la superficie de la piel tanto sana como con acné<sup>82</sup>, *E. coli* creemos que viene de contaminación de otras secuenciaciones, ya que el subtipo es K12, al igual que *rhizobiales* también parecen ser contaminación de secuenciaciones previas llevadas a cabo en la Unidad Universitaria de Secuenciación Masiva de DNA.

¿Qué sentido tienen estos hallazgos? Nos dejan notar que la secuenciación masiva es altamente sensible, por lo que es propensa a contaminaciones que podrían mal interpretarse como organismos presentes en la muestra original. Como en todo experimento de laboratorio existen errores, pero mientras más conocimiento tengamos sobre las posibles fuentes de contaminación nos permitirá tomarlas en cuenta e interpretar con precaución los resultados.

## **Influenza**

Dadas las evidencias experimentales y los resultados del análisis bioinformático, concluimos que en las muestras que está presente el virus de influenza H1N1pdm2009 son: 6985, 7192, 7425, 8261, 8836, 9041 y 9140, todas ellas provenientes de pacientes hospitalizados en la unidad de cuidados intensivos, 6 de ellos fallecieron. Los resultados de los protocolos bioinformáticos A, B y C son congruentes entre ellos, si bien son muy pocas

las lecturas que logramos identificar como virus de influenza A, los resultados obtenidos con el PCR recomendado por la OMS y el kit diagnóstico comercial Seeplex confirmaron la presencia del virus en 6 de 7 muestras. En la muestra 9041 es posible que el número de copias presente del genoma del virus fuera menor al límite de detección del PCR-CDC y del PCR múltiple que usa Seeplex y sólo detectamos la presencia de H1N1pdm2009 mediante secuenciación masiva.

En las muestras 8410, 9232, 9275, 9362 y 9434 detectamos lecturas correspondientes a virus de influenza A H3N2. Ninguno de estos pacientes murió ni estuvo en sala de cuidados intensivos, por lo que el virus estacional H3N2 circuló simultáneamente con el virus H1N1pdm09 pero causando aparentemente menor severidad en la enfermedad. Un estudio realizado en hurones infectados con distintos tipos de virus de influenza A demuestra que la severidad de los síntomas y el tiempo de recuperación son mayores al infectar con H1N1pdm09 que con H3N2 estacional<sup>83</sup>. Los autores describen que H1N1pdm09 causa inflamación severa en los pulmones, que puede llevar a un paro respiratorio y muerte, así como un periodo prolongado de fiebre, posiblemente provocado por la secreción prolongada de citosinas proinflamatorias. Mientras que H3N2 estacional sólo causó un periodo de 1 a 2 días de fiebre con recuperación pronta. Si extrapolamos los hallazgos de este trabajo a nuestras muestras humanas, podría explicar por qué en los pacientes que se detectó H3N2 sobrevivieron mientras que los infectados con H1N1pdm09 fallecieron.

Las 3 pruebas de diagnóstico molecular que fueron realizadas en estas muestras están dirigidas al gene M, por lo que debería haberse detectado la presencia de H3N2, lo cual no sucedió.

Este fenómeno puede explicarse si el virus H3N2 tuviera un gene M divergente del detectado por los métodos de diagnóstico molecular, impidiendo así su detección. Sin embargo gracias a la estrategia metagenómica mediante secuenciación masiva fue posible detectarlo en las muestras previamente negativas a H1N1pdm09, identificando así indicios de la circulación de virus H3N2 estacionales durante la pandemia de influenza en México en 2009.

En las muestras 3974 y 8883 no detectamos la presencia de influenza A, ni por PCR ni por secuenciación masiva, por lo que concluimos que estos pacientes no estuvieron infectados con influenza, pero sí presentaron síntomas similares a los causados por el virus de influenza A, de acuerdo a los datos proporcionados por el INER. En conclusión, en 7 muestras detectamos influenza A H1N1, en 5 muestras influenza A H3N2 y en 2 muestra no detectamos influenza.

La secuenciación masiva de DNA de muestras clínicas directas, utilizando la plataforma Illumina, detectó muy pocas secuencias del virus de influenza A por razones desconocidas hasta ahora. Una posibilidad es que el genoma del virus, siendo de 13,000 nucleótidos, se diluye en la abundancia de secuencias del hospedero, cuyo genoma consta de  $3.2 \times 10^9$  nucleótidos, dificultando su detección. Hay que recordar que de la muestra se extrajeron ácidos nucleicos totales sin ningún paso de enriquecimiento previo para partículas virales (ver Materiales y Métodos). Es importante también considerar el estadio de la infección en el que se tomó la muestra. Estos pacientes estaban hospitalizados y muy probablemente las muestras fueron tomadas hacia el final de la infección, cuando necesitaron ventilación mecánica y fue posible tomar las muestras. Al final de la infección es cuando hay menor carga viral, el pico de secreción viral se encuentra en las primeras 48 horas, con  $10^3$  a  $10^7$  TCID<sub>50</sub>/mL y declina paulatinamente hasta que después de 6 días es casi indetectable<sup>56</sup>. En contraste con los resultados de influenza, la presencia de rinovirus se detectó con claridad en la muestra 9140, probablemente porque la muestra se tomó durante la fase inicial de la infección.

En la literatura existen trabajos con resultados similares a nuestros hallazgos, en los cuales se reporta que en muestras de lavados nasofaríngeos de pacientes positivos a influenza por RT-PCR, el virus no se ha podido encontrar de manera eficiente a través de secuenciación masiva mediante pirosecuenciación, obteniéndose sólo de 20 a 240 lecturas<sup>14</sup>. En otro trabajo similar a este proyecto, utilizando muestras aisladas de México, Canadá y Estados Unidos, se encontró que en las muestras de México que no fueron tratadas con DNAsas se obtuvieron muy pocas lecturas de influenza y coberturas muy pobres (utilizando el método de secuenciación de Illumina), argumentando que la preparación de la muestra era la causa<sup>84</sup>.

En las muestras 6985, 7192 y 8836 logramos recuperar parcialmente el genoma del virus H1N1pdm2009. Los ácidos nucleicos de los que se partió para amplificar estos genomas vienen de las mismas muestras en las que con los protocolos A, B y C se detectaron muy pocas lecturas de influenza A, sin embargo, al amplificar específicamente el genoma de influenza A obtuvimos 96% del genoma viral.

En los 3 genomas parciales que recuperamos del virus H1N1pdm2009 buscamos marcadores de virulencia o mutaciones ya reportadas asociadas a una mayor virulencia. No encontramos ninguna; sólo identificamos 5 mutaciones particulares del virus presente en la muestra 6985 respecto a los circulantes en México en la misma época. De estas 5 mutaciones una, en la hemaglutinina (Asn179Ser) está en el sitio de reconocimiento de ácido sialico, que podría alterar la unión del virus a su receptor celular.

En su conjunto, los datos anteriores nos sugieren que la mayor mortalidad por el virus H1N1pdm09 observada en México en relación a otros países no se debe a que los virus circulantes en nuestro país fueran más virulentos que los circulantes en el resto del mundo en la misma época ya que existen secuencias aisladas en diversas partes del mundo que son 100% idénticas a las aisladas en México; es decir que si la virulencia estuviese asociada a algún cambio en aminoácidos, éste no se encuentra exclusivamente en las variantes aisladas en México.

La pandemia se presentó desfasada al brote de influenza estacional, por lo que los servicios médicos probablemente no estaban preparados para contender con un nuevo virus de influenza. Lo anterior, aunado al momento durante el curso de la infección en que los pacientes acudieron en las primeras fases de la epidemia al hospital, ya que generalmente sólo acudían los que llevaban varios días enfermos (Quiñones, comunicación personal), siendo esta una causal de fracaso en el tratamiento con antivirales.

Otro factor que puede haber contribuido a un índice de mayor mortalidad asociada al virus H1N1pdm09 al inicio de la pandemia en México, es que sólo se reportaban los casos muy graves, por lo que las estimaciones de tasas de letalidad del virus fueron elevadas. Al transcurrir los meses y existir el reporte de pacientes no graves, estas tasas descendieron. El Instituto Nacional de Referencia Epidemiológica reportó que durante el mes de Abril de

2009 la letalidad del virus de influenza A circulante en México era del 1%<sup>49</sup>; posteriormente, al confirmar más casos, se estableció que la letalidad del virus H1N1pHm09 en México es 0.4% con una alta transmisibilidad<sup>85</sup>. En Tailandia la letalidad fue de 0.58%<sup>86</sup> y en los Estados Unidos 0.15%<sup>87</sup>, aunque estos resultados son variables dependiendo de la corte de pacientes que se utilice. En Inglaterra estimaron una letalidad de 0.024%<sup>88</sup> al tomar en cuenta tanto a los casos hospitalizados como a los muy leves.

La exploración a nivel molecular del virus causante de esta pandemia fue importante para descartar la posibilidad de circulación de un virus con marcadores de virulencia o mutaciones asociadas a una mayor fatalidad en el curso de la enfermedad, afortunadamente no fue el caso. Más adelante discutiremos la presencia de bacterias asociadas a la epidemia de influenza A.

## **Rinovirus**

En la muestra 9140 encontramos mediante los protocolos A, B y C la presencia de rinovirus, también confirmada por Seplex. Con el protocolo A identificamos ~24,000 lecturas, en el B y C ~2,800 lecturas, y al ensamblar el genoma obtuvimos ~65,000 lecturas. El serotipo encontrado para este virus de acuerdo al análisis filogenético realizado fue rinovirus humano 73, uno de los virus circulantes en poblaciones humanas; existen alrededor de 100 serotipos y los más prevalentes cambian cada año<sup>89</sup>.

Las enfermedades respiratorias causadas por rinovirus son de las más comunes, los síntomas inician como un resfriado común, con estornudos, obstrucción nasal, inflamación de garganta, dolor de cabeza, tos y malestar general; la fiebre generalmente no se presenta en infecciones por rinovirus. Aproximadamente 1/3 de las infecciones por rinovirus son asintomáticas, la duración de la infección va de 1 ó 2 días hasta un mes en las más graves, las complicaciones incluyen infección en el tracto respiratorio bajo que predispone a la presencia de agentes bacterianos<sup>53</sup>.

En el análisis filogenético este virus claramente pertenece al grupo A, uno de los más abundantes en las poblaciones humanas. En la filogenia incluimos secuencias de los grupos



A, B, C y el recientemente descrito clado D con el objetivo de determinar a cuál de estos pertenecía nuestra secuencia. En México hay una mayor prevalencia del grupo A y B, aunque también se han encontrado infecciones por virus del grupo C<sup>90</sup>.

El paciente en el que encontramos rinovirus A 73 estuvo internado en la unidad de cuidados intensivos con diagnóstico de neumonía severa, inflamación sistémica severa y neumonía asociada a ventilación mecánica. Este paciente también presenta diabetes mellitus y su edad en el momento de la infección era de 63 años. No falleció.

En 2002, durante la temporada de infecciones respiratorias en la que predomina el virus de influenza A, se detectó en un grupo de 100 adultos mayores hospitalizados y graves la presencia de rinovirus en 4 de ellos y no de influenza, todos ellos se recuperaron<sup>91</sup>. Ese trabajo sentó un precedente sobre la asociación de rinovirus no sólo a enfermedades respiratorias leves, sino también a algunos casos severos en los que no se identifica otro patógeno viral.

Un brote de rinovirus en un asilo de ancianos en 2003 culminó en el deceso de 13 personas, de las cuales en 7 se identificó rinovirus serotipo 82 como único agente infeccioso<sup>92</sup>, siendo éste uno de los primeros reportes en el que rinovirus causa enfermedades respiratorias graves que culminan en decesos. Las implicaciones futuras de la identificación de rinovirus 73 en un caso de neumonía severa, como el caso del paciente 9140, son importantes ya que un patógeno típicamente asociado a enfermedades leves, como es rinovirus, debe vigilarse ya que puede causar también complicaciones graves en adultos mayores.

### **Diversidad viral**

La metodología que utilizamos en la preparación de las muestras no implicó ningún paso para enriquecer secuencias virales, por lo que el número de secuencias virales que observamos es muy bajo comparado con el número de lecturas del hospedero. Esto refleja el panorama del ambiente pulmonar en el que aunque el número de partículas virales por célula es alto, la mucosa tanto pulmonar como nasal contiene células del sistema inmune que fueron extraídas en los lavados y las secuencias del genoma viral se diluyen.

El virus que identificamos en todas las muestras fue adenovirus humano C, el número de lecturas va desde 2 hasta 68, estos número a simple vistan parecen muy bajos, sin embargo si ponderamos el tamaño del genoma y número de células humanas comparado con el número de partículas virales y tamaño de su genoma, estos números son proporcionales. En un estudio sobre la prevalencia de adenovirus C en muestras obtenidas de amigdalectomías, cuantificaron que por cada  $10^7$  células humanas existen hasta  $10^6$  partículas virales<sup>93</sup>. Si tomamos en cuenta que el tamaño del genoma humano haploide es, ~3,000 millones de nucleótidos y el genoma de adenovirus C, ~35,000 nucleótidos, como se muestra en la Tabla 11, por cada nucleótido de adenovirus existen ~3,000,000 de nucleótidos humanos, esto en las condiciones del reporte. En nuestro caso, en la muestra 6985 encontramos 18 lecturas únicas de adenovirus y ~ 6,000,000 de genoma humano, esto equivale a que por cada nucleótido de adenovirus hay ~300,000 nucleótidos humanos. Así concluimos que 18 lecturas de adenovirus en una muestra son representativas y proporcionales de una infección.

	<b>Adenovirus C humano</b>	<b>Homo sapiens</b>
<b>tamaño genoma (nt)</b>	35,000	2 * 3,000,000,000
<b># células o partículas virales</b>	10,000,000	100,000,000
<b># nucleótidos totales</b>	3.5E+11	6E+17
<b>adenovirus/humano</b>	<b>5.8E-07</b>	
<b>reads 6985</b>	18	6,487,764
<b>adenovirus/humano</b>	<b>2.7E-6</b>	

**Tabla 11:** Proporciones de adenovirus vs genoma humano

En todas las muestras encontramos adenovirus C, un patógeno que se ha descrito como causal del 5-10% de los casos de neumonías y enfermedades respiratorias severas respectivamente. La prevalencia de diferentes tipos de adenovirus se ha reportado de hasta el 70% en niños menores de 1 año, dependiendo de la zona geográfica. La prevalencia en adultos del grupo C no se ha analizado con detalle (con técnicas sensibles que no dependan de anticuerpos) y de serotipos del grupo B se está apenas empezando a determinar. En todo caso, son comúnmente asociados a infecciones del tracto respiratorio inferior.

En el estudio realizado por Rohwer en 2009 sobre metagenómica de muestras de pacientes con y sin fibrosis quística identificaron como virus predominantes a adenovirus y herpesvirus tanto en muestras de pacientes enfermos como de los sanos. En el trabajo de Lysholm en 2012<sup>94</sup> también detectaron adenovirus C en un pool de muestras de lavados nasofaríngeos de personas con enfermedades respiratorias, es interesante que también detectaron tanto influenza H1N1 como H3N2.

El segundo virus que identificamos en todas las muestras es retrovirus endógeno humano K (HERV-K), el 8% del genoma humano está compuesto de elementos retrovirales que se cree están inactivos, a excepción retrovirus K que se ha propuesto sigue activo en el genoma humano desde la divergencia entre humanos y chimpancés<sup>95</sup>, la presencia de estas secuencias en nuestro grupo de datos va desde 1 hasta 45 lecturas, es interesante notar que sólo el protocolo bioinformático A detecta este tipo de virus, ya que en las bases de datos de genomas de referencia virales no están contenidos estos elementos. En el estudio de Rohwer de 2009 detectaron retrovirus, es interesante ya que su metodología incluye el tratamiento con DNAsa I y la filtración por tamaño de las partículas virales. Sugieren que la presencia de retrovirus es por intermediarios de DNA de infecciones persistentes en los pacientes con fibrosis quística y no por similitud con los elementos insertados en el genoma humano. En nuestro caso creemos que son elementos presentes en el genoma humano, ya que se encuentran en todas las muestras y no utilizamos un tratamiento para deshacernos del DNA humano. La variación en el número de lecturas que cada muestra presenta podrían reflejar el número de copias de estos elementos que existe en el genoma de cada paciente, sin embargo es una hipótesis que no puede probarse o descartarse con los datos que tenemos.

En cinco muestras identificamos herpesvirus, la muestra con más número de lecturas fue la 7425 con 44 lecturas identificadas como herpesvirus humano 4 que también recibe el nombre de virus Epstein-Barr, este paciente falleció. En el reporte de Rohwer 2009 identifican lecturas de virus de Epstein-Barr sólo en los pacientes con fibrosis quística, la presencia de este virus se ha asociado a un padecimiento pulmonar exacerbado<sup>96</sup>.

En dos muestras identificamos virus de papiloma, una de ellas, fue la muestra 3974 en la que no identificamos influenza. Se ha reportado la presencia de virus de papiloma en tumores pulmonares<sup>97</sup> así como en las vías respiratorias altas de niños<sup>98</sup>. En el trabajo de Rohwer 2009 también identifican lecturas de virus de papiloma tanto en los pacientes con CF como en los sanos.

En una muestra identificamos hepatitis B y en la muestra control PhiX identificamos hepatitis C. En una muestra identificamos una lectura de virus de dengue. En la muestra 7425 identificamos una lectura de klassevirus, un virus recientemente descrito<sup>99</sup> a partir de un pool de muestras gastroentéricas pediátricas. Los siguientes virus fueron identificados en muestras separadas: virus de sarcoma UR2, sapelovirus de simio 2, virus de encefalomiocarditis porcina, virus del bosque de semliki, virus de mosaico del tabaco, torqueteno virus y anelovirus. Estos dos últimos también fueron identificados en el trabajo de Lysholm 2012 en el pool de muestras de personas con enfermedades respiratorias. Aunque desconocemos las implicaciones de la presencia de estos virus en el padecimiento de cada paciente, es el tipo de información interesante que se puede obtener de un estudio metagenómico.

Los fagos que identificamos en las muestras, reflejan la diversidad bacteriana de las mismas. En todos los casos en los que detectamos el fago, detectamos también la bacteria hospedera del mismo. Por lo tanto la metagenómica viral se puede utilizar como monitoreo de algunas bacterias. De interés, en las muestras encontramos fagos de Enterobacteria, Burkolderia, Lactobacillus, Pseudomonas, Staphylococcus y Streptococcus.

Las secuencias virales identificadas en el análisis realizado están sin duda presentes en estas muestras, aunque si fueron o no la causa del síndrome respiratorio severo no lo podemos saber. Sin embargo, sí nos dan una idea de la diversidad de virus eucariotes y bacterianos que existe en el tracto respiratorio humano, así como una prueba de principio sobre la posibilidad de identificar virus en muestras directamente extraídas de pacientes sin previa amplificación ni cultivo de los virus.

## Diversidad bacteriana

El comportamiento de los datos respecto a la abundancia de especies bacterianas refleja pocos organismos con alta cobertura y muchos con baja, esto nos habla de organismos predominantes en este tipo de muestras.

Para todos los análisis decidimos sólo tomar en cuenta a las especies bacterianas con 10 o más lecturas asignadas, este corte es arbitrario y tiene el objetivo de poder comparar las muestras entre sí con mayor facilidad. La reducción del número de bacterias asignadas con este corte fue aproximadamente de un 90%; para la muestra 9041 en la que al tomar en cuenta las bacterias identificadas por sólo una lectura o más tenemos ~600 bacterias, al cortar a 10 o más lecturas nos quedamos con sólo 41 bacterias. Al reducir los datos a analizar pudimos identificar bacterias comunes entre varios metagenomas.

*Burkholderia gladioli* es la bacteria más abundante en 12 de las 14 muestras analizadas. Cada muestra tiene entre 600 y 800,000 lecturas asignadas a este organismo, el cual se ha reportado como patógeno oportunista nosocomial. También se ha descrito en metagenomas de personas con fibrosis quística y como una infección reincidente en pacientes que han recibido trasplantes de pulmón<sup>100</sup>, por lo que no resulta extraño encontrarla en este metagenoma, sin embargo, es necesario investigar cuál podría ser su rol, si acaso alguno, en la progresión de la enfermedad de los pacientes. En la muestra 8836 *Pseudomonas aeruginosa* es la bacteria más abundante con 68,392 lecturas y en la muestra 8883 *Haemophilus influenzae* es la bacteria predominante con 42,527 lecturas, ambas son bacterias patógenas asociadas a enfermedades respiratorias, por lo que creemos que fueron las causantes de los padecimientos de los pacientes.

La segunda bacteria más abundante en 4 muestras es *Burkholderia glumae*, en dos muestras *Prevotella melaninogenica*, en 2 muestras *Propionibacterium acnes* y sólo en una muestra *Staphylococcus epidermis*, *Rothia mucilaga*, *Escherichia coli*, *Veillonella parvula* y *Moraxella catarrhalis*, esta última reportada como patógena en tracto respiratorio

El establecimiento de un punto de corte en 10 secuencias por especie nos permitió observar con claridad la distribución de organismos presentes en las muestras sin el ruido de

organismos muy poco abundantes. El número de bacterias identificado en cada muestra es variable, con un mínimo de 19, máximo de 176 y media de 67; no hubo una diferencia significativa entre el número de bacterias de los pacientes vivos y aquellos que fallecieron.

En la taxonomía de NCBI existen 35 phyla, 4 de los cuales estuvieron presentes en todas las muestras: Actinobacterias, Bacteroidetes, Firmicutes y Proteobacterias. No encontramos ninguna diferencia clara entre los dos grupos de datos (vivos y muertos) a nivel de phylum.

A nivel de especie existen bacterias específicas para el grupo de fallecidos: *Rhosopseudomonas palustri*, *Stenotrophomonas maltophila*, *Variovorax paradoxus*, *Pseudomonas fluorescens*, *Burkholderia phytofirmans*, *Burkholderia pseudomallei*, *Ralstonia picketti*, *Acidovorax citruli*, *Plevotella melaninogenica*, *Lactobacillus casei*, *Sphingomonas wittchii*. No podemos correlacionar la presencia de estas bacterias con la gravedad de los pacientes, ya que la diferencia observada podría deberse al sitio del tracto respiratorio del que se tomó la muestra. En el caso de los fallecidos todas las muestras fueron tomadas de tracto respiratorio bajo, mientras que de los vivos es de tracto respiratorio alto. Las especies bacterianas presentes sólo en los pacientes vivos son: *Staphylococcus epidermis*, *Thermobacterium thermosa*, *Paracoccus denitrificans*, *Burkholderia vietnamensis*, *Burkholderia thailandensis*, *Comamonas testosteroni*, *Uncultured beta proteobacterium*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa*, *Pseudomonas mendocina*, *Pseudomonas stutzeri* estas especies podrían representar a bacterias que crecen en tracto respiratorio alto.

Es interesante comparar los resultados emanados de este estudio con otros trabajos similares que han sido publicados. Para comprender los resultados de cada trabajo, a continuación presento un resumen del tipo de muestras y métodos que cada uno de ellos utiliza; posteriormente se presentará la comparación de los resultados con este estudio. Los artículos publicados con los que compararé este estudio son:

Lysholm, 2012.

Caracterización del microbioma viral en pacientes con infecciones severas del tracto respiratorio bajo, utilizando secuenciación metagenómica.

210 muestras de aspirados nasofaríngeos fueron recolectadas de Marzo 2004 a Mayo de 2005 en Estocolmo, para realizar diagnóstico de infecciones del tracto respiratorio.

Las muestras fueron mezcladas en un pool y tratadas con DNAsas, después extrajeron los ácidos nucleicos totales y los amplificaron mediante hexámeros al azar; posteriormente fueron secuenciadas en la plataforma 454. El análisis fue mediante comparaciones con la base de datos NT con un corte de e-valor de  $10^{-5}$ , la asignación se realizó sólo con el mejor hit para cada secuencia. 12% de las secuencias no tuvieron homología con NT.

Sze, 2012.

El microbioma del tejido pulmonar en enfermedad crónica obstructiva de pulmón (COPD)

Muestras de tejido pulmonar de 8 personas no fumadoras, 8 de fumadores sin COPD, 8 de pacientes con COPD muy severa, y como control positivo 8 de pacientes con fibrosis quística (CF), también usaron agua estéril como control negativo.

El análisis consistió en amplificación de la región hipervariable V2 de la subunidad ribosomal 16S y cuantificación mediante qPCR. Amplificación de fragmentos de 550 nucleótidos de las regiones hipervariables V1 a V3 y posterior pirosecuenciación. Identificaron de 20 a 252 células bacterianas por cada 1,000 células humanas.

Erb-Downward, 2011.

Análisis del microbioma pulmonar en fumadores “sanos” y en COPD

Este estudio analizó muestras de lavados broncoalveolares de 14 voluntarios y 8 especímenes de tejido pulmonar, estos fueron obtenidos de personas que recibieron un trasplante de pulmón.

Las muestras fueron tratadas con proteinasa K, y la región hipervariable V1 a V3 fue amplificada y posteriormente secuenciada en la plataforma 454. También realizaron PCR cuantitativo de los 16S.

Greninger, 2010.

Un análisis metagenómico de la infección por el virus Influenza A pandémico (2009 H1N1) en pacientes de Norte América.

Muestras de 17 lavados nasofaríngeos tomadas de pacientes con síntomas de influenza A, las muestras fueron tomadas en las etapas tempranas de la pandemia de 2009, son aislados de México, Canadá y Estados Unidos.

Las muestras fueron hibridadas en un microarreglo de detección viral y paralelamente secuenciadas en la plataforma Illumina con una secuenciación tipo pair-end, utilizaron 3 carriles y barcode en cada muestra. Las muestras de Mexico no fueron tratadas con DNAsa I y las demás si.

Sapkota, 2010.

Patógenos humanos que son abundantes en el metagenoma bacteriano de cigarrillos.

Analizan cuatro marcas de cigarrillos, de cajetillas nuevas seleccionan cuatro cigarrillos y de los cuales extraen ácidos nucleicos mediante un kit comercial para extracción de DNA de suelos. Amplificaron algunas regiones de la subunidad 16s rRNA que posteriormente clonaron y seleccionaron algunas clonas para secuenciar mediante Sanger. Las muestras paralelamente fueron hibridadas en un microarreglo diseñado para detectar bacterias a diferentes niveles taxonómicos.

En la Tabla 12 se encuentra la comparación entre los primeros cuatro trabajos y este trabajo a nivel de familia. Sólo uno de los trabajos fue realizado en la plataforma Illumina GAIIx, pero a una profundidad menor de la que nosotros utilizamos. Las comparaciones inician a nivel de phylum identificados por cada estudio, los phyla identificados por los cinco estudios y que podríamos decir que son parte del microbioma core del tracto respiratorio humano son: Proteobacteria, Firmicutes, Actinobacteria y Bacteroidetes.

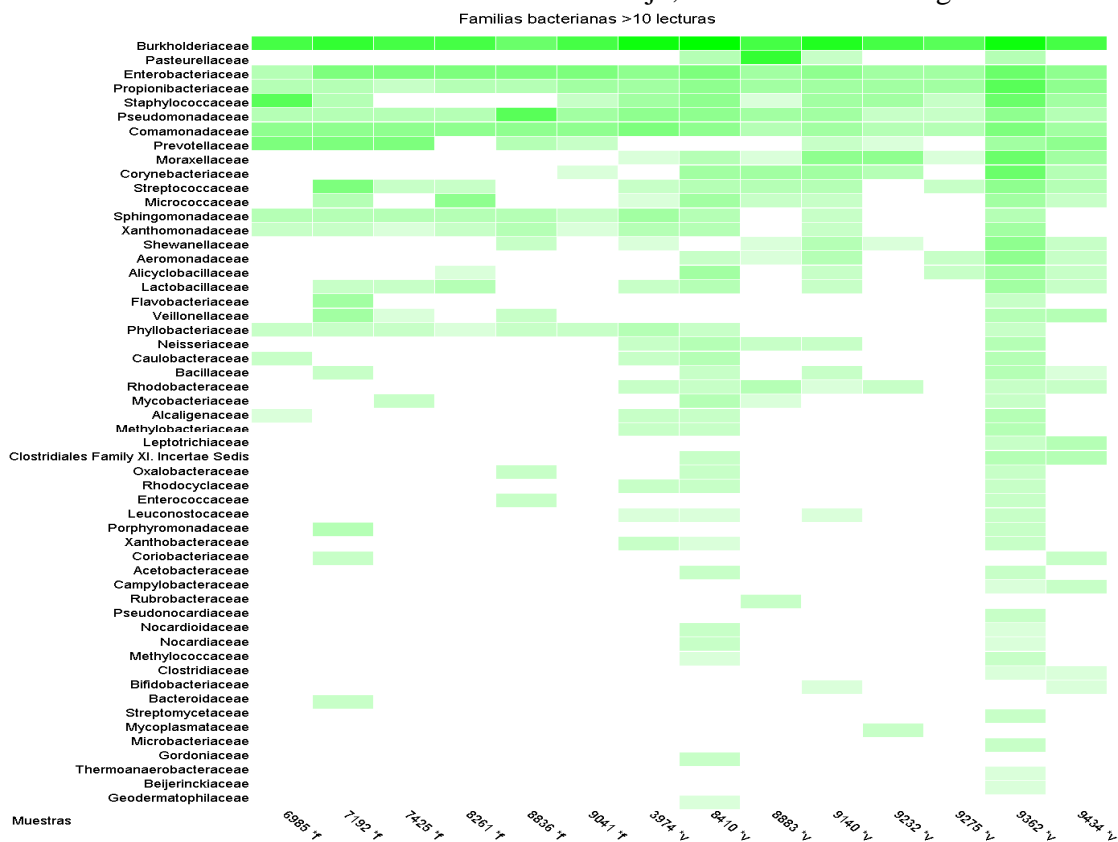


Tecnología	Illumina GAIIX	454 GS-FLX	454 GS-FLX	454 FLX-titanium	Illumina GAIIX
Región amplificada	Todo	Todo	V1 a V3 16S	V1 a V3 16S	Todo
Tratamiento DNAsa I	No	Si	No	No	Algunas
<i>Actinobacteria</i>	Si	Si	Si	Si	Si
<i>Bacteroidetes</i>	Si	Si	Si	Si	Si
<i>Firmicutes</i>	Si	Si	Si	Si	Si
<i>Proteobacteria</i>	Si	Si	Si	Si	Si
<i>Tenericutes</i>	1 muestra	No	1 muestra	No	Si
<i>Cyanobacteria</i>	No	No	3 muestras	No	No
<i>Fusobacteria</i>	2 muestras	No	No	Si	No
<i>Spirochaetes</i>	No	No	No	Si	No
<i>Deinococcus</i>	2 muestras	No	No	No	No
<i>Burkholderiaceae</i>	1°	No	2° COPD	11°	5°
<i>Pasteurellaceae</i>	2°	4°	No	6°	5°
<i>Enterobacteriaceae</i>	3°	9°	No	10°	1°
<i>Propionibacteriaceae</i>	4°	No	No	No	5°
<i>Staphylococaceae</i>	5°	6°	No	No	5°
<i>Pseudomonadaceae</i>	6°	5°	1° CF	1°	4°
<i>Comamonadaceae</i>	7°	1°	1° nonsmoker	No	4°
<i>Prevotellaceae</i>	8°	No	No	3°	No
<i>Moraxellaceae</i>	9°	11°	No	11°	2°
<i>Corynebacteriaceae</i>	10°	No	No	No	No
<i>Streptococaceae</i>	11°	7°	No	2°	3°
<i>Mycrococcineae</i>	12°	No	No	8°	No
<i>Sphingobacteriaceae</i>	13°	10°	No	No	No
<i>Xanthomonadaceae</i>	14°	No	No	No	No
<i>Lactobacillaceae</i>	15°	2°	1° COPD	No	4°
<i>Shewanellaceae</i>	15°	No	No	No	5°
<i>Aeromonadaceae</i>	15°	No	No	No	No
<i>Alicyclobacillaceae</i>	15°	No	No	No	No
<i>Flovabacteriaceae</i>	15°	No	No	No	No
<i>Veillonellaceae</i>	16°	No	No	4°	No
<i>Neisseriaceae</i>	16°	No	No	8°	5°
<i>Phylobacteriaceae</i>	16°	No	No	No	No
<i>Mycoplasmaceae</i>	17°	No	No	No	5°
<i>Bacillaceae</i>	17°	No	No	8°	No
<i>Fusobacteriaceae</i>	No	No	No	4°	No
<i>Alcaligenaceae</i>	No	3°	2° CF	No	5°
<i>Carnobacteriaceae</i>	No	10°	No	8°	4°
<i>Oxalobacteriaceae</i>	No	No	No	No	5°
<i>Clostridiaceae</i>	No	No	No	No	5°
<i>Porphyramonadaceae</i>	No	No	No	5°	5°
<i>Methylophilaceae</i>	No	8°	No	No	No
<i>Chitinophagaceae</i>	No	9°	No	No	No
<i>Hydrogonophilaceae</i>	No	11°	No	No	No
<i>Cytophagaceae</i>	No	11°	No	No	No
<i>Leptotrichiaceae</i>	No	No	No	7°	No
<i>Lacnospiraceae</i>	No	No	No	8°	No

<i>Actinomycetae</i>	No	No	No	8°	No
<i>Campilobacteriaceae</i>	No	No	No	9°	No
<i>Cortobacteriaceae</i>	No	No	No	9°	No
<i>Erysipelotrichaceae</i>	No	No	No	9°	No
<i>Lachnospiraceae</i>	No	No	No	9°	No
<i>Elostridiales</i>	No	No	No	9°	No
<i>Bacteroidaceae</i>	No	No	No	9°	No

**Tabla 12:** Comparación de estudios metagenómicos de vías respiratorias. Los campos en color azul son información sobre el método. Los campos en color morado son clasificación a nivel phylum. Los campos en negro son clasificación a nivel familia.

Las familias bacterianas identificadas en este trabajo, se muestran en la Figura 33



**Figura 33:** Familias bacterianas identificadas en cada muestra con más de 10 lecturas. La escala de color representa el número de lecturas en escala logarítmica.

No identificamos a un grupo de bacterias específico para cada grupo de pacientes, tampoco lo logran otros estudios.

### **Bacterias patógenas**

Las bacterias que identificamos como patógenas humanas y están en las muestras de todos los pacientes incluidos en este estudio son: *Propionibacterium acnes*, *Burkholderia cepacia*, *Burkholderia pseudomallei*, *Ralstonia pickettii*, *Aeromonas sp* y *Escherichia coli*.

El número de bacterias patógenas identificadas en los pacientes con infecciones no graves fue mayor a las encontradas en los que fallecieron; la media en este caso fue de 16, comparado con 9 en los pacientes que fallecieron. En el conjunto de datos completo, el menor número de bacterias patógenas fue 8 y el mayor 31. Creemos que la presencia de un menor número de bacterias patógenas en los pacientes graves y que fallecieron se debe al tratamiento antibiótico que recibieron que consistió en ceftriaxona y claritromicina. Los pacientes que no fueron hospitalizados sólo recibieron tratamiento antiviral. La bacteria patógena *Burkholderia cepacia* está presente en todas las muestras.

Cabe resaltar que la muestra 9362 presenta un gran número de bacterias patógenas, 31 de 38 incluidas en la base de datos identificadas en todo el set de datos, aunque no tenemos los datos clínicos completos de esta persona podemos inferir que sufrió un cuadro clínico grave.

En las muestras 9232 y 9275 no hay ningún tipo de *Pseudomonas* mientras que en todas las demás muestras sí. En la muestra 8836 *Pseudomonas aeruginosa* es el organismo predominante con 68,000 lecturas, mientras que en la muestra 8883 *Haemophilus influenzae* es el predominante con 42,000 lecturas

En un estudio<sup>27</sup> en el que de pacientes con influenza buscan bacterias asociadas mediante secuenciación masiva con 454 entre las bacterias, que representaron el ~1.5% del total de lecturas, las más abundantes fueron *Streptococcus pneumoniae*, *Moraxella bovis*,

*Moraxella bovoculi*, *Haemophilus influenzae* y *Escherichia coli*. En otro trabajo<sup>29</sup> con el objetivo de explorar la diversidad bacteriana y viral en el tracto respiratorio utilizaron secuenciación mediante 454 y encontraron que el 4.5% de las secuencias fueron bacterianas y 1.9% virales. Las bacterias sólo las caracterizaron a nivel de género, siendo los más abundantes *Moraxellaceae* y *Streptococcus*. En la parte viral el 55% de las lecturas fueron de la familia *Orthomyxoviridae* (familia a la cual pertenece el virus de influenza), seguido de fagos, algunos virus de plantas, 16 lecturas del virus de la leucemia murina y una lectura con virus de ebola

## **PERSPECTIVAS**

### **Identificación de virus desconocidos**

La cantidad de secuencias de virus conocidos que obtuvimos en este estudio, a excepción de rinovirus A, son muy bajas, por lo que es predecible que si existiera un nuevo virus en estas muestras no podríamos muestrearlo lo suficiente como para recuperar su genoma completo a menos que la muestra fuera tomada en la fase aguda de la enfermedad. Sin embargo, lo que se podría hacer al respecto es ensamblar *de novo* las lecturas que no pudieron ser asignadas a ninguna de las categorías anteriores, para obtener contigs más largos y utilizarlos para hacer búsquedas de patrones de genes mediante Modelos Ocultos de Markov entrenados con sets de datos de cápsides virales, polimerasas u otras proteínas virales (MetaGeneMarkS, Glimmer, MetaGene) y predicción de tRNAs, entre otros.

### **Aplicación de este método en diagnóstico**

Esta metodología es útil en el diagnóstico de aquellas muestras que por diagnóstico molecular clásico no es posible identificar a un agente causal del padecimiento. En un futuro podría utilizarse como una herramienta cotidiana en el diagnóstico y descubrimiento de virus, a medida que las tecnologías de secuenciación masiva decrezcan sus precios y el tiempo de análisis sea más rápido, esta metodología podrá usarse de forma eficiente.

## BIBLIOGRAFÍA

---

- <sup>1</sup> Watson, J. D., & Crick, F. H. C. (1953). Molecular structure of nucleic acids. *Nature*, (4356), 737–738.
- <sup>2</sup> Handelsman, J., Rondon, M., Brady, S., Clardy, J., & Goodman, R. (1998). Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chemistry & Biology*, 5(10), 245–249.
- <sup>3</sup> Wooley, J. C., Godzik, A., & Friedberg, I. (2010). A primer on metagenomics. *PLoS computational biology*, 6(2), e1000667. doi:10.1371/journal.pcbi.1000667
- <sup>4</sup> Amann, R. I., Ludwig, W., & Schleifer, K. H. (1995). Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiological reviews*, 59(1), 143–69.
- <sup>5</sup> Hugenholtz, P., Goebel, B. M., & Pace, N. R. (1998). Impact of Culture-Independent Studies on the Emerging Phylogenetic View of Bacterial Diversity, *180*(18), 4765–4774.
- <sup>6</sup> Lane, D. J., Pace, B., Olsen, G. J., Stahl, D. a, Sogin, M. L., & Pace, N. R. (1985). Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proceedings of the National Academy of Sciences of the United States of America*, 82(20), 6955–9.
- <sup>7</sup> Schmidt, T. M., DeLong, E. F., & Pace, N. R. (1991). Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. *Journal of bacteriology*, 173(14), 4371–8.
- <sup>8</sup> Nelson, D. Cox, M.,(2008) Lehninger Principles of Biochemistry, Fifth edition, Freeman, USA.
- <sup>9</sup> Doolittle, W. F. (1999). Phylogenetic Classification and the Universal Tree. *Science*, 284(5423), 2124–2128. doi:10.1126/science.284.5423.2124
- <sup>10</sup> Ashelford, K. E., Chuzhanova, N. A., Fry, J. C., Jones, A. J., & Weightman, A. J. (2005). At Least 1 in 20 16S rRNA Sequence Records Currently Held in Public Repositories Is Estimated To Contain Substantial Anomalies. *Applied and environmental microbiology*, 71(12), 7724–7736. doi:10.1128/AEM.71.12.7724
- <sup>11</sup> Bintrim, S., & Oodman, R. (1997). Molecular phylogeny of Archaea from soil. *Proceedings of the National Academy of Sciences of the United States of America*, 94(January), 277–282.
- <sup>12</sup> Makimura K, et.al.(2001) Fungal flora on board the Mir-Space Station, identification by morphological features and ribosomal DNA sequences. *Microbiology and Immunology*, 45(5):357-63.
- <sup>13</sup> Reyes, A., Haynes, M., Hanson, N., Angly, F. E., Andrew, C., Rohwer, F., & Gordon, J. I. (2011). Viruses in the fecal microbiota of monozygotic twins and their mothers. *Viruses* 466(7304), 334-338. doi:10.1038/nature09199.

- 
- <sup>14</sup> Nakamura, S., Yang, C.-S., Sakon, N., Ueda, M., Tougan, T., Yamashita, A., Goto, N., et al. (2009). Direct metagenomic detection of viral pathogens in nasal and fecal specimens using an unbiased high-throughput sequencing approach. *PloS one*, 4(1), e4219. doi:10.1371/journal.pone.0004219
- <sup>15</sup> Sanger, F., & Nicklen, S. (1977). DNA sequencing with chain-terminating. *Proceedings of the National Academy of Sciences of the United States of America*, 74(12), 5463–5467.
- <sup>16</sup> <http://www.polonator.org/index.htm>
- <sup>17</sup> Chin, C.-S., Sorenson, J., Harris, J. B., Robins, W. P., Charles, R. C., Jean-Charles, R. R., Bullard, J., et al. (2011). The origin of the Haitian cholera outbreak strain. *The New England journal of medicine*, 364(1), 33–42. doi:10.1056/NEJMoa1012928
- <sup>18</sup> Timp, W., Mirsaidov, U. M., Wang, D., Comer, J., Aksimentiev, A., & Timp, G. (2010). Nanopore Sequencing: Electrical Measurements of the Code of Life. *IEEE transactions on nanotechnology*, 9(3), 281–294. doi:10.1109/TNANO.2010.2044418
- <sup>19</sup> Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., et al. (2012). Comparison of next-generation sequencing systems. *Journal of biomedicine & biotechnology*, 2012, 251364. doi:10.1155/2012/251364
- <sup>20</sup> Breitbart, M., Salamon, P., Andresen, B., Mahaffy, J. M., Segall, A. M., Mead, D., Azam, F., et al. (2002). Genomic analysis of uncultured marine viral communities. *PNAS*.
- <sup>21</sup> Mann, N. H., Cook, A., Bailey, S., Clokie, M., Amanullah, A., Azam, N., Balliet, A., et al. (2003). Bacterial photosynthesis genes in a virus. *Nature*, 424(August), 741-742.
- <sup>22</sup> Edwards, R. a, & Rohwer, F. (2005). Viral metagenomics. *Nature reviews. Microbiology*, 3(6), 504-10. doi:10.1038/nrmicro1163
- <sup>23</sup> Brüßow, H., & Hendrix, R. W. (2002). Phage genomics: small is beautiful. *Cell*, 108(1), 13-6. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11792317>
- <sup>24</sup> National Center for Biotechnology Information <http://www.ncbi.nlm.nih.gov/genome/>
- <sup>25</sup> Marco, D, et.al. (2010). *Metagenomics: Theory, Methods and Applications*. Caister Academic Press, UK.
- <sup>26</sup> Holmes, E. C. (2009). The comparative genomics of viral emergence. *PNAS*, 1-5.
- <sup>27</sup> World Health Organization,, World Health Statistics 2010: Estimated deaths by cause among children aged <5 years, WHO, 2010, Geneva
- <sup>28</sup> Neverov, A., & Chumakov, K. (2010). Massively parallel sequencing for monitoring genetic consistency and quality control of live viral vaccines. *Proceedings of the National Academy of Sciences of the United States of America*, 107(46), 20063–8. doi:10.1073/pnas.1012537107

- 
- <sup>29</sup> Victoria, J. G., Wang, C., Jones, M. S., Jaing, C., McLoughlin, K., Gardner, S., & Delwart, E. L. (2010). Viral nucleic acids in live-attenuated vaccines: detection of minority variants and an adventitious virus. *Journal of virology*, *84*(12), 6033-40. doi:10.1128/JVI.02690-09
- <sup>30</sup> Victoria, J. G., Kapoor, A., Dupuis, K., Schnurr, D. P., & Delwart, E. L. (2008). Rapid identification of known and new RNA viruses from animal tissues. *PLoS pathogens*, *4*(9), e1000163. doi:10.1371/journal.ppat.1000163
- <sup>31</sup> Wang, D., Coscoy, L., Zylberberg, M., Avila, P. C., Boushey, H. a, Ganem, D., & DeRisi, J. L. (2002). Microarray-based detection and genotyping of viral pathogens. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(24), 15687-92. doi:10.1073/pnas.242579699
- <sup>32</sup> Willner, D., Furlan, M., Haynes, M., Schmieder, R., Angly, F. E., Silva, J., Tammadoni, S., et al. (2009). Metagenomic analysis of respiratory tract DNA viral communities in cystic fibrosis and non-cystic fibrosis individuals. *PLoS one*, *4*(10), e7370. doi:10.1371/journal.pone.0007370
- <sup>33</sup> Ng, T. F. F., Manire, C., Borrowman, K., Langer, T., Ehrhart, L., & Breitbart, M. (2009). Discovery of a novel single-stranded DNA virus from a sea turtle fibropapilloma by using viral metagenomics. *Journal of virology*, *83*(6), 2500-9. doi:10.1128/JVI.01946-08
- <sup>34</sup> Day, J. M., Ballard, L. L., Duke, M. V., Scheffler, B. E., & Zsak, L. (2010). Metagenomic analysis of the turkey gut RNA virus community. *Virology journal*, *7*(1), 313. BioMed Central Ltd. doi:10.1186/1743-422X-7-313
- <sup>35</sup> Genersch, E., & Aubert, M. (2010). Emerging and re-emerging viruses of the honey bee (*Apis mellifera* L.). *Veterinary research*, *41*(6), 54. doi:10.1051/vetres/2010027
- <sup>36</sup> Tokarz, R., Firth, C., Street, C., Cox-foster, D. L., & Lipkin, W. I. (2011). Lack of Evidence for an Association between Iridovirus and Colony Collapse Disorder. *PLoS ONE*, *6*(6), 2009–2011. doi:10.1371/journal.pone.0021844
- <sup>37</sup> Blomstro, A., & Wide, F. (2010). Detection of a Novel Astrovirus in Brain Tissue of Mink Suffering from Shaking Mink Syndrome by Use of Viral Metagenomics. *Journal of clinical microbiology*, *48*(12), 4392–4396. doi:10.1128/JCM.01040-10
- <sup>38</sup> Stenglein, M. D., Sanders, C., Kistler, A. L., Ruby, J. G., Franco, J. Y., Reavill, D. R., & Dunker, F. (2012). Identification , Characterization , and In Vitro Culture of Highly Divergent Arenaviruses from Boa Constrictors and Annulated Tree Boas : Candidate Etiological Agents for Snake Inclusion Body Disease. *mBio*, *3*(4). doi:10.1128/mBio.00180-12
- <sup>39</sup> Quan, P.-lan, Wagner, T. A., Briese, T., Torgerson, T. R., Hornig, M., Tashmukhamedova, A., et al. (2010). Astrovirus Encephalitis in Boy with Agammaglobulinemia, X.-linked. *Emerging Infectious Diseases*, *16*(6). doi:10.3201/eid1606.091536
- <sup>40</sup> Finkbeiner, S. R., Li, Y., Ruone, S., Conrardy, C., Gregoricus, N., Toney, D., Virgin, H. W., et al. (2009). Identification of a novel astrovirus (astrovirus VA1) associated with an outbreak of acute gastroenteritis. *Journal of virology*, *83*(20), 10836–9. doi:10.1128/JVI.00998-09

- 
- <sup>41</sup> Greninger, A. L., Runckel, C., Chiu, C. Y., Haggerty, T., Parsonnet, J., Ganem, D., & DeRisi, J. L. (2009). The complete genome of klassevirus - a novel picornavirus in pediatric stool. *Virology journal*, 6, 82. doi:10.1186/1743-422X-6-82
- <sup>42</sup> Briese, T., Paweska, J. T., McMullan, L. K., Hutchison, S. K., Street, C., Palacios, G., Khristova, M. L., et al. (2009). Genetic detection and characterization of Lujo virus, a new hemorrhagic fever-associated arenavirus from southern Africa. *PLoS pathogens*, 5(5), e1000455. doi:10.1371/journal.ppat.1000455
- <sup>43</sup> Xu, B., Liu, L., Huang, X., Ma, H., Zhang, Y., Du, Y., Wang, P., et al. (2011). Metagenomic analysis of fever, thrombocytopenia and leukopenia syndrome (FTLS) in Henan Province, China: discovery of a new bunyavirus. *PLoS pathogens*, 7(11), e1002369. doi:10.1371/journal.ppat.1002369
- <sup>44</sup> Towner, J. S., Sealy, T. K., Khristova, M. L., Albariño, C. G., Conlan, S., Reeder, S. a, Quan, P.-L., et al. (2008). Newly discovered ebola virus associated with hemorrhagic fever outbreak in Uganda. *PLoS pathogens*, 4(11), e1000212. doi:10.1371/journal.ppat.1000212
- <sup>45</sup> Van Leeuwen, M., Williams, M. M. W., Koraka, P., Simon, J. H., Smits, S. L., & Osterhaus, A. D. M. E. (2010). Human picobirnaviruses identified by molecular screening of diarrhea samples. *Journal of clinical microbiology*, 48(5), 1787–94. doi:10.1128/JCM.02452-09
- <sup>46</sup> Frasser RS (2005) Histology and gross anatomy of the respiratory tract. In: Hamid Q. (eds) *Physiology bases of respiratory disease*. BK Decker, Hamilton.
- <sup>47</sup> Henderson AJ (1994) Bronchoalveolar lavage, *Arch Dis Child* 70:167-169
- <sup>48</sup> Chowell, G., Bertozzi, S. M., Colchero, M. A., Lopez-Gatell, H., Alpuche-Aranda, C., Hernandez, M., & Miller, M. a. (2009). Severe respiratory disease concurrent with the circulation of H1N1 influenza. *The New England journal of medicine*, 361(7), 674-9. doi:10.1056/NEJMoa0904023
- <sup>49</sup> Bright, R. a, Medina, M.-jo, Xu, X., Perez-Oronoz, G., Wallis, T. R., Davis, X. M., Povinelli, L., et al. (2005). Incidence of adamantane resistance among influenza A (H3N2) viruses isolated worldwide from 1994 to 2005: a cause for concern. *Lancet*, 366(9492), 1175-81. doi:10.1016/S0140-6736(05)67338-2
- <sup>50</sup> Cordova-Villalobos et.al., La epidemia de Influenza A/H1N1 en México, editorial medica Panamericana, México, 2010.
- <sup>51</sup> Maines, T. R., Jayaraman, A., Belser, J. a, Wadford, D. a, Pappas, C., Zeng, H., Gustin, K. M., et al. (2009). Transmission and pathogenesis of swine-origin 2009 A(H1N1) influenza viruses in ferrets and mice. *Science (New York, N.Y.)*, 325(5939), 484–7. doi:10.1126/science.1177238
- <sup>52</sup> Boëlle, P.-Y., Ansart, S., Cori, A., & Valleron, A.-J. (2011). Transmission parameters of the A/H1N1 (2009) influenza virus pandemic: a review. *Influenza and other respiratory viruses*, 5(5), 306–16. doi:10.1111/j.1750-2659.2011.00234.x



- 
- <sup>53</sup> Garten, R. J., Davis, C. T., Russell, C. a, Shu, B., Lindstrom, S., Balish, A., Sessions, W. M., et al. (2009). Antigenic and genetic characteristics of swine-origin 2009 A(H1N1) influenza viruses circulating in humans. *Science (New York, N.Y.)*, 325(5937), 197-201. doi:10.1126/science.1176225
- <sup>54</sup> Knipe D, Howley P. et.al (2007). *Fields Virology*, Fifth edition, *Lippincott Williams & Wilkins*, USA
- <sup>55</sup> Webster, R. G., Bean, W. J., Gorman, O. T., Chambers, T. M., & Kawaoka, Y. (1992). Evolution and ecology of influenza A viruses. *Microbiological reviews*, 56(1), 152–79.
- <sup>56</sup> Arias, C. F., Escalera-Zamudio, M., Soto-Del Río, M. D. L. D., Cobián-Güemes, A. G., Isa, P., & López, S. (2009). Molecular anatomy of 2009 influenza virus A (H1N1). *Archives of medical research*, 40(8), 643-54. doi:10.1016/j.arcmed.2009.10.007
- <sup>57</sup> Perez-Padilla, R., de la Rosa-Zamboni, D., Ponce de Leon, S., Hernandez, M., Quiñones-Falconi, F., Bautista, E., Ramirez-Venegas, A., et al. (2009). Pneumonia and respiratory failure from swine-origin influenza A (H1N1) in Mexico. *The New England journal of medicine*, 361(7), 680-9. doi:10.1056/NEJMoa0904252
- <sup>58</sup> Cheng, V. C. C., Lau, Y.-K., Lee, K.-L., Yiu, K.-H., Chan, K.-H., Ho, P.-L., & Yuen, K.-Y. (2009). Fatal co-infection with swine origin influenza virus A/H1N1 and community-acquired methicillin-resistant *Staphylococcus aureus*. *The Journal of infection*, 59(5), 366–70.
- <sup>59</sup> Murray, R. J., Robinson, J. O., White, J. N., Hughes, F., Coombs, G. W., Pearson, J. C., Tan, H.-L., et al. (2010). Community-acquired pneumonia due to pandemic A(H1N1)2009 influenza virus and methicillin resistant *Staphylococcus aureus* co-infection. *PloS one*, 5(1), e8705.
- <sup>60</sup> Chowell, G., Bertozzi, S. M., Colchero, M. A., Lopez-Gatell, H., Alpuche-Aranda, C., Hernandez, M., & Miller, M. a. (2009). Severe respiratory disease concurrent with the circulation of H1N1 influenza. *The New England journal of medicine*, 361(7), 674–9. doi:10.1056/NEJMoa0904023
- <sup>61</sup> US Patent: Process for Amplifying Nucleic Acids, US 4,683,202 - July 28, 1987, K. Mullis
- <sup>62</sup> CDC protocol of realtime RTPCR for influenza A(H1N1), 28 April 2009
- <sup>63</sup> Kim, S. R., Ki, C.-S., & Lee, N. Y. (2009). Rapid detection and identification of 12 respiratory viruses using a dual priming oligonucleotide system-based multiplex PCR assay. *Journal of virological methods*, 156(1-2), 111-6. Elsevier B.V. doi:10.1016/j.jviromet.2008.11.007
- <sup>64</sup> Zhou, B., Donnelly, M. E., Scholes, D. T., St George, K., Hatta, M., Kawaoka, Y., & Wentworth, D. E. (2009). Single-reaction genomic amplification accelerates sequencing and vaccine production for classical and Swine origin human influenza a viruses. *Journal of virology*, 83(19), 10309-13. doi:10.1128/JVI.01109-09

- 
- <sup>65</sup> Li, H., Ruan, J., & Durbin, R. (2008). Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome research*, 18(11), 1851–8. doi:10.1101/gr.078212.108
- <sup>66</sup> Cock, P. J. A., Fields, C. J., Goto, N., Heuer, M. L., & Rice, P. M. (2010). The Sanger FASTQ file format for sequences with quality scores, and the Solexa / Illumina FASTQ variants, 38(6), 1767–1771. doi:10.1093/nar/gkp1137
- <sup>67</sup> Altschup, S. F., Gish, W., Pennsylvania, T., & Park, U. (1990). Basic Local Alignment Search Tool 2Department of Computer Science, 403–410.
- <sup>68</sup> Smith, T. F., & Waterman, M. S. (1981). Identification of common molecular subsequences. *Journal of molecular biology*, 147(1), 195–7. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/7265238>
- <sup>69</sup> Kent, W. J. (2002). BLAT---The BLAST-Like Alignment Tool. *Genome Research*, 12(4), 656–664. doi:10.1101/gr.229202
- <sup>70</sup> <http://www.sanger.ac.uk/resources/software/smalt/>
- <sup>71</sup> <http://samtools.sourceforge.net>
- <sup>72</sup> Zerbino, D. R., & Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome research*, 18(5), 821–9. doi:10.1101/gr.074492.107
- <sup>73</sup> Huson, D. H., Auch, A. F., Qi, J., & Schuster, S. C. (2007). MEGAN analysis of metagenomic data, 377–386. doi:10.1101/gr.5969107.
- <sup>74</sup> Escalera-Zamudio, M., Cobián-Güemes, G., de Los Dolores Soto-Del Río, M., Isa, P., Sánchez-Betancourt, I., Parissi-Crivelli, A., Teresa Martínez-Cázares, M., et al. (2012). Characterization of an influenza A virus in Mexican swine that is related to the A/H1N1/2009 pandemic clade. *Virology*, 433, 176–182. doi:10.1016/j.virol.2012.08.003
- <sup>75</sup> Ā, W. L., & Godzik, A. (2006). Cd-hit : a fast program for clustering and comparing large sets of protein or nucleotide sequences, 22(13), 1658–1659. doi:10.1093/bioinformatics/btl158
- <sup>76</sup> Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* (Oxford, England), 25(14), 1754–60. doi:10.1093/bioinformatics/btp324
- <sup>77</sup> <http://www.r-project.org/>
- <sup>78</sup> Huson D., S. Mitra, H. Ruscheweyth et.al, (2011) Integrative analysis of environmental sequences using MEGAN4. *Genome Research*.
- <sup>79</sup> Ecker, D. J., Sampath, R., Willett, P., Wyatt, J. R., Samant, V., Massire, C., Hall, T. a, et al. (2005). The Microbial Rosetta Stone Database: a compilation of global and emerging infectious microorganisms and bioterrorist threat agents. *BMC microbiology*, 5(Cdc), 19. doi:10.1186/1471-2180-5-19

- 
- <sup>80</sup> Palacios, G., Ph, D., Druce, J., Du, L., Tran, T., Birch, C., Briese, T., et al. (2008). A New Arenavirus in a Cluster of Fatal Transplant-Associated Diseases. *The New England journal of medicine*, 991–998.
- <sup>81</sup> Erb-Downward, J. R., Thompson, D. L., Han, M. K., Freeman, C. M., McCloskey, L., Schmidt, L. a, Young, V. B., et al. (2011). Analysis of the lung microbiome in the “healthy” smoker and in COPD. *PloS one*, 6(2), e16384. doi:10.1371/journal.pone.0016384
- <sup>82</sup> Richard A. Bojar, Keith T. Holland (2004). Acne and propionibacterium acnes, *Clinics in Dermatology*, Volume 22, Issue 5, September-October 2004, Pages 375-379, ISSN 0738-081X,
- <sup>83</sup> Huang, S. S. H., Banner, D., Fang, Y., Ng, D. C. K., Kanagasabai, T., Kelvin, D. J., & Kelvin, A. a. (2011). Comparative analyses of pandemic H1N1 and seasonal H1N1, H3N2, and influenza B infections depict distinct clinical pictures in ferrets. *PloS one*, 6(11), e27512. doi:10.1371/journal.pone.0027512
- <sup>84</sup> Greninger, A. L., Chen, E. C., Sittler, T., Scheinerman, A., Roubinian, N., Yu, G., Kim, E., et al. (2010). A Metagenomic Analysis of Pandemic Influenza A (2009 H1N1) Infection in Patients from North America. (R. Tripp, Ed.) *PLoS ONE*, 5(10), e13381. doi:10.1371/journal.pone.0013381
- <sup>85</sup> Fraser, C., Donnelly, C. a, Cauchemez, S., Hanage, W. P., Van Kerkhove, M. D., Hollingsworth, T. D., Griffin, J., et al. (2009). Pandemic potential of a strain of influenza A (H1N1): early findings. *Science (New York, N.Y.)*, 324(5934), 1557–61. doi:10.1126/science.1176062
- <sup>86</sup> de Silva, U. C., Warachit, J., Waicharoen, S., & Chittaganpitch, M. (2009). A preliminary analysis of the epidemiology of influenza A(H1N1)v virus infection in Thailand from early outbreak data, June-July 2009. *Euro surveillance : bulletin européen sur les maladies transmissibles = European communicable disease bulletin*, 14(31), 6–8. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/19660247>
- <sup>87</sup> Louie, J. K., Acosta, M., Winter, K., Jean, C., Gavali, S., Schechter, R., Vugia, D., et al. (2009). Factors associated with death or hospitalization due to pandemic 2009 influenza A(H1N1) infection in California. *JAMA : the journal of the American Medical Association*, 302(17), 1896–902. doi:10.1001/jama.2009.1583
- <sup>88</sup> Donaldson, L. J., Rutter, P. D., Ellis, B. M., Greaves, F. E. C., Mytton, O. T., Pebody, R. G., & Yardley, I. E. (2009). Mortality from pandemic A/H1N12009 influenza in England: public health surveillance study. *BMJ*. doi:10.1136/bmj.b5213
- <sup>89</sup> Palmenberg Ann C., Rathe, Jennifer, L. S. (2011). Analysis of Complete Genome Sequences of Human Rhinovirus. *Journal of Allergy*, 125(6), 1190-1201. doi:10.1016/j.jaci.2010.04.010.Analysis
- <sup>90</sup> Landa-cardena, A., Morales-romero, J., García-roman, R., Cobián-Güemes, A. G., Pitalúa-cortés, F., Mora, S. I., & Montero, H. (2012). Clinical Characteristics and Genetic Variability of Human Rhinovirus in Mexico. *Viruses*, 4, 200–210. doi:10.3390/v4020200

- 
- <sup>91</sup> Falsey, A. R., Walsh, E. E., & Hayden, F. G. (2002). Rhinovirus and coronavirus infection-associated hospitalizations among older adults. *The Journal of infectious diseases*, 185(9), 1338–41. doi:10.1086/339881
- <sup>92</sup> Louie, J. K., Yagi, S., Nelson, F. a, Kiang, D., Glaser, C. a, Rosenberg, J., Cahill, C. K., et al. (2005). Rhinovirus outbreak in a long term care facility for elderly persons associated with unusually high mortality. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 41(2), 262–5. doi:10.1086/430915
- <sup>93</sup> Garnett, C. T., Erdman, D., Xu, W., & Gooding, L. R. (2002). Prevalence and Quantitation of Species C Adenovirus DNA in Human Mucosal Lymphocytes, 76(21), 10608–10616. doi:10.1128/JV1.76.21.10608
- <sup>94</sup> Lysholm, F., Wetterbom, A., Lindau, C., Darban, H., Bjerkner, A., Fahlander, K., Lindberg, a M., et al. (2012). Characterization of the viral microbiome in patients with severe lower respiratory tract infections, using metagenomic sequencing. *PloS one*, 7(2), e30875. doi:10.1371/journal.pone.0030875
- <sup>95</sup> Belshaw, R., Dawson, A. L. A., Woolven-allen, J., Redding, J., Burt, A., & Tristem, M. (2005). Genomewide Screening Reveals High Levels of Insertional Polymorphism in the Human Endogenous Retrovirus Family HERV-K ( HML2 ): Implications for Present-Day Activity †, 79(19), 12507–12514. doi:10.1128/JVI.79.19.12507
- <sup>96</sup> Winnie GB, Cowan RG.(1992). Association of Epstein-Barr virus infection and pulmonary exacerbations in patients with cystic fibrosis. *Pediatr Infect Dis J*. 1992 Sep;11(9):722-6.PMID:1333067
- <sup>97</sup> Klein F, Kotb WFMA, Petersen I (2008) Incidence of human papilloma virus in lung cancer. *Lung Cancer*. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19019488>
- <sup>98</sup> Zawadzka-Głós L, Jakubowska A, Chmielik M, Bielicka A, Brzewski M (2003) Lower airway papillomatosis in children. *Int J Pediatr Otorhinolaryngol* 67: 1117–1121.
- <sup>99</sup> Greninger, A. L., Runckel, C., Chiu, C. Y., Haggerty, T., Parsonnet, J., Ganem, D., & DeRisi, J. L. (2009). The complete genome of klassevirus - a novel picornavirus in pediatric stool. *Virology journal*, 6, 82. doi:10.1186/1743-422X-6-82
- <sup>100</sup> Murray, S., Charbeneau, J., Marshall, B. C., & LiPuma, J. J. (2008). Impact of burkholderia infection on lung transplantation in cystic fibrosis. *American journal of respiratory and critical care medicine*, 178(4), 363-71. doi:10.1164/rccm.200712-1834OC

  
**remel**

## MicroTest™ M4RT®

### Medio multimicrobiano

(Español)

#### USO PREVISTO

MicroTest™ M4RT® de Remel es un medio líquido cuyo uso se recomienda en los procedimientos cualitativos, para el transporte de muestras clínicas hacia el laboratorio, en los procedimientos microbiológicos de virus y clamidias.

#### RESUMEN Y EXPLICACIÓN

Los virus y las clamidias son sensibles a las condiciones ambientales adversas y requieren un medio de transporte específico para mantener su viabilidad (e infectividad) durante su traslado al laboratorio. Con tal fin, son ideales las formulaciones que contienen proteínas para la estabilización, antibióticos para minimizar la contaminación bacteriana y micótica, y un tampón para mantener un pH neutro.

#### PRINCIPIO

El medio M4RT® consiste en una solución salina equilibrada de Hank modificada, complementada con albúmina sérica bovina, gelatina, sacarosa y ácido glutámico. El pH se equilibra con tampón HEPES. Como indicador de pH se usa rojo fenol. Para inhibir el crecimiento de bacterias y levaduras competidoras se añaden al medio gentamicina y anfotericina B. El medio es isotónico y no tóxico para las células huésped de mamífero, lo que permite su uso en procedimientos de centrifugación y cultivo celular para el diagnóstico precoz (*shell vial assay*). Se añaden crioprotectores para garantizar la viabilidad del organismo durante el proceso de congelación y descongelación. No se ha incluido suero bovino completo en esta formulación porque puede inhibir la adhesión y crecimiento de mixovirus y paramixovirus.

#### REACTIVOS (FÓRMULA CLÁSICA)\*

Ingredientes de los reactivos:

Salino de Hank equilibrado	Tampón HEPES
Albúmina sérica bovina	Rojo fenol
Gelatina	Gentamicina
Sacarosa	Anfotericina B
Ácido L-glutámico	

pH 7,3 +/- 0,2 a 25 °C

\*Ajustado según necesidades para cumplir los estándares de funcionamiento.

#### PRECAUCIONES

Este producto sólo es para uso diagnóstico *in vitro* y debe ser utilizado por personal con la formación adecuada. Se tomarán precauciones frente a los riesgos microbiológicos, esterilizando correctamente las muestras, envases y medios después de su uso. Se deben leer y seguir atentamente las instrucciones.

#### ALMACENAMIENTO

Este producto se presenta listo para su uso y no requiere más preparación. Se debe almacenar en su envase original a una temperatura de 2 a 30 °C hasta el momento de su uso. No calentar en exceso. No incubar antes de su uso. El almacenamiento inadecuado provocará la pérdida de actividad antimicrobiana.

#### DETERIORO DEL PRODUCTO

Este producto no se debe usar si: 1) hay signos de contaminación; 2) hay signos de fugas; 3) el color ha cambiado y no es rosa claro; 4) se ha sobrepasado la fecha de caducidad, o 5) hay otros signos de deterioro.

#### OBTENCIÓN, ALMACENAMIENTO Y TRANSPORTE DE LAS MUESTRAS

Las muestras se deben obtener y manipular siguiendo las normas recomendadas.<sup>2,4</sup> Para mantener la viabilidad óptima, transportar la muestra al laboratorio en cuanto sea posible. Aunque el medio M4RT® puede mantener incluso los microorganismos frágiles durante periodos

de tiempo prolongados a temperatura ambiente, se recomienda refrigerar las muestras a una temperatura de 2 a 8°C o mantenerlas en hielo después de su obtención y durante el transporte. Si se va a retrasar mucho el procesado, las muestras se deben enfriar a -70°C o menos y el transporte se hará en hielo seco. Las muestras se deben enviar según las normas federales, estatales y hospitalarias y se procesarán en cuanto sea posible tras su llegada al laboratorio.

#### MATERIALES NECESARIOS PERO NO SUMINISTRADOS

(1) Jeringuillas, agujas; (2) pinzas estériles; (3) medio suplementario; (4) asa de inoculación, dispositivo de esterilización en asa; (5) incubadoras, sistemas ambientales alternativos; (6) microorganismos para el control de calidad, y (7) los materiales necesarios para la identificación del microorganismo. Consultar el equipo necesario para la obtención de la muestra en las referencias apropiadas.<sup>2,4</sup>

#### PROCEDIMIENTO

La obtención adecuada de la muestra del paciente es fundamental para el éxito del aislamiento y la identificación de los microorganismos infecciosos. Las muestras se deben obtener lo antes posible después del inicio de la enfermedad.

1. Obtener la muestra en cuanto sea posible tras el inicio de los síntomas.
2. Extraer el tapón del vial en condiciones asépticas.
3. Insertar la torunda en el medio.
4. Romper limpiamente el mango de la torunda por la línea ranurada.
5. Volver a poner el tapón en el vial y cerrarlo bien.
6. Etiquetar con la información adecuada del paciente.
7. Enviar al laboratorio para su procesamiento con el menor retraso posible.

#### INTERPRETACIÓN DE LOS RESULTADOS

Este medio de transporte sirve como vehículo para mantener la viabilidad del microorganismo durante el transporte de la muestra al laboratorio.

#### CONTROL DE CALIDAD

En todos los lotes de MicroTest™ M4RT® se ha estudiado la contaminación microbiológica, la toxicidad de las células huésped y la capacidad para mantener la viabilidad de los microorganismos deseados (ver la lista a continuación), y el resultado obtenido ha sido aceptable. El estudio de los microorganismos de control se debe realizar de acuerdo con los procedimientos de control de calidad establecidos en el laboratorio. Si se observan resultados anómalos en el control de calidad, no se informará de los resultados de ese paciente.

#### CONTROL

Herpes simplex tipo 2 ATCC® VR-734  
Virus sincitial respiratorio ATCC® VR-1302  
Citomegalovirus ATCC® VR-977  
*Chlamydia trachomatis* ATCC® VR-880

#### LIMITACIONES

1. Manejar las muestras en condiciones asépticas.
2. La situación, el momento y el volumen de las muestras obtenidas para cultivo son variables significativas para obtener resultados fiables del cultivo. Seguir las normas recomendadas para la obtención de las muestras.<sup>2,4</sup>
3. M4RT® está indicado únicamente como medio de obtención y transporte de virus y clamidias. Este medio puede servir como crioprotector de virus clínicos, incluidos el citomegalovirus y el virus varicela-zoster.
4. Como las torundas de alginato cálcico son tóxicas para muchos virus encapsulados y pueden interferir con las pruebas de anticuerpos fluorescentes, no se deben usar para obtención de la muestra. Las torundas con mango de madera pueden contener toxinas y formaldehídos y no se deben usar. Las torundas con punta de algodón o políester son adecuadas cuando procede obtener la muestra con torunda.

#### VALORES ESPERADOS

Los resultados dependerán en gran medida de la obtención correcta y adecuada de la muestra, así como de los oportunos transporte y procesado en el laboratorio.

### CARACTERÍSTICAS DE FUNCIONAMIENTO

El medio de transporte M4RT® de Remel se comparó con el medio de transporte comercial y estándar usado habitualmente para el transporte y mantenimiento de virus y clamidias. Los resultados fueron equivalentes o superiores a los obtenidos con el medio de comparación. El porcentaje de recuperación de cada agente a una temperatura de 2 a 8 °C fue el siguiente:

MICROORGANISMO	8 HORAS	24 HORAS	48 HORAS
Virus sincitial respiratorio	79%	65%	52%
Virus del herpes simple 1 McInyre		84%	65%
Citomegalovirus (AD 169)		65%	38%
Influenzae A H3N2 Shanghai 87		89%	72%
<i>Chlamydia trachomatis</i>	80%	48%	35%

### BIBLIOGRAFÍA

- Koneman, E.W., S.D. Allen, W.M. Janda, P.C. Schreckenberger, and W.C. Winn, Jr. 1992. Color Atlas and Textbook of Diagnostic Microbiology. 4<sup>th</sup> ed. J.B. Lippincott Co., Philadelphia, PA.
- Murray, P.R., E.J. Baron, J.H. Jorgensen, M.A. Pfaller, and R.H. Tenover. 2003. Manual of Clinical Microbiology. 8<sup>th</sup> ed., Vol. 2. ASM, Washington, D.C.
- Moffett, M.B., J.L. Young, and R.D. Stuart. 1948. Br. Med. J. 2:421-424.
- Mahoney, J.B. and M.A. Chernesky. J. Clin. Microbiol. 22:865-867.
- Gleaves, C.A., R.L. Hodinka, S.L.G. Johnston, and E.M. Swierkosz. 1994. Cumitech 15A. Laboratory Diagnosis of Viral Infections. ASM, Washington, D.C.
- Forbes, B.A., D.F. Sahm, and A.S. Weissfeld. 2002. Bailey and Scott's Diagnostic Microbiology. 11<sup>th</sup> ed. Mosby, St. Louis, MO.

### Símbolos

<b>REF</b>	Número de catálogo
<b>IVD</b>	Dispositivo médico para diagnóstico <i>in vitro</i>
<b>LAB</b>	Para el uso del laboratorio
	Consulte las instrucciones de uso
	Límite de temperatura (Temp. de almacenamiento)
<b>LOT</b>	Código de lote (Número de lote)
	Fecha de caducidad
<b>EC REP</b>	Representante autorizado en Europa

M4RT® es una marca registrada de Remel Inc.  
 MicroTest™ es una marca comercial de Remel Inc.  
 ATCC® es una marca registrada de American Type Culture Collection.



IFU 12505, Revisado el 2005-08-24

Impreso en los EE.UU.

### PRESENTACIONES DE ENVASADO

REF.	Tubo de centrifuga cónico de 15 ml 3 ml llenado	Bolas de vidrio	Torundas con punta de polyester		Tarjeta de ID de la muestra	* Bolsa con cierre de cremallera	Tamaño del envase
			Sobre con 2 torundas con mango de plástico (mujeres)	Sobre con 1 torunda con el mango de plástico y 1 torunda con el cuerpo de acero inoxidable (varones y mujeres)			
12505	√	√					72
12506	√	√					12
12552	√	√	√			√	100
12553	√	√	√		√	√	100
12576	√	√	√	√		√	100
12577	√	√	√	√	√	√	100
12578	√	√		√		√	100
12579	√	√		√	√	√	100

\* Bolsa con cierre de cremallera se piensa para el empaquetado del producto y no el transporte del espécimen.

12076 Santa Fe Drive, Lenexa, KS 66215, EE.UU.

Información General: (800) 255-6730 Servicio Técnico: (800) 447-3641 Pedidos: (800) 447-3635

Teléfono nacional/internacional: (913) 888-0939 Fax internacional: (913) 895-4128

Internet: www.remel.com Correo electrónico: remel@remel.com



## **CDC protocol of realtime RTPCR for influenza A(H1N1)**

**28 April 2009**

*revision 1 (30 April 2009)*

*revision 2 (6 October 2009)*

The WHO Collaborating Centre for influenza at CDC Atlanta, United States of America, has made available the protocol, attached, of realtime RTPCR for influenza A(H1N1).

# ***CDC Realtime RTPCR (rRTPCR) Protocol for Detection and Characterization of Swine Influenza (version 2009)***

## ***General Comments***

**Disclaimer:** Names of vendors or manufacturers are provided as examples of suitable product sources. Inclusion does not imply endorsement by the Centers for Disease Control and Prevention.

**Assumptions:** This procedure assumes a basic familiarity with rRT-PCR assays.

**Assay principle:** The *CDC Realtime RTPCR (rRTPCR) Protocol for Detection and Characterization of Swine Influenza* includes a panel of oligonucleotide primers and dual-labeled hydrolysis (Taqman<sup>®</sup>) probes to be used in real-time RT-PCR assays for the *in vitro* qualitative detection and characterization of swine influenza viruses in respiratory specimens and viral cultures. The InfA primer and probe set is designed for universal detection of type A influenza viruses. The swInfA primer and probe set is designed to specifically detect all swine influenza A viruses. The swH1 primer and probe set is designed to specifically detect swine H1 influenza. This assay is utilized for testing influenza A positive respiratory specimens (unsubtypable) taken from suspect swine influenza A infected patients.

**Protocol Use Limitations:** These protocols were optimized using quantitative one-step probe RT-PCR (Invitrogen SuperScript<sup>™</sup>III Platinum<sup>®</sup> One-Step Quantitative Kit) that have been shown to produce comparable results on 96-well format thermocycler systems such as Applied Biosystems<sup>™</sup> real-time PCR systems (7000, 7300, 7500, etc.), BioRad real-time PCR detection systems (iQ<sup>™</sup> or iQ5<sup>™</sup>) or Stratagene QPCR instruments (MX4000<sup>®</sup>, MX3000<sup>®</sup> or MX3005<sup>®</sup>).

**Safety Information:** Specimen processing should be performed in accordance with pertaining national biological safety regulations.

**Acceptable specimens:** Respiratory specimens including: bronchoalveolar lavage, tracheal aspirates, sputum, nasopharyngeal or oropharyngeal aspirates or washes, and nasopharyngeal or oropharyngeal swabs. Swab specimens should be collected only on swabs with a synthetic tip (such as polyester or Dacron<sup>®</sup>) and an aluminum or plastic shaft. Swabs with cotton tips and wooden shafts are not recommended. Specimens collected with swabs made of calcium alginate are not acceptable.

Rejection criteria:

- Specimens not kept at 2-4°C (≤4 days) or frozen at -70°C or below.
- Inappropriate specimens not listed above.

**Nucleic acid extraction:** Performance of RT-PCR amplification based assays depends on the amount and quality of sample template RNA. RNA extraction procedures should be qualified and validated for recovery and purity before testing specimens. Commercially available



extraction procedures including QIAamp® Viral RNA Mini Kit, or RNeasy® Mini Kit (QIAGEN), Roche MagNA Pure Compact RNA Isolation Kit, MagNA Pure LC RNA Isolation Kit II, and Roche MagNA Pure Total Nucleic Acid Kit have been shown to generate highly purified RNA when following manufacturer's recommended procedures for sample extraction.

## *Materials*

### **Reagents:**

1. One-step quantitative RT-PCR probe hydrolysis (e.g., Taqman®) kit
  - Invitrogen SuperScript™III Platinum® One-Step Quantitative Kit (cat# 11732-020 or 11745-100).
2. Molecular grade sterile distilled water (RNase and DNase free)
3. Forward and reverse primers (40µM)
4. Dual-labeled probes (10µM)
5. Positive control materials

### **Supplies:**

1. Laboratory marking pen
2. Cooler racks for 1.5 microcentrifuge tubes and 96-well 0.2ml PCR reaction tubes
3. 20µl and 200µl adjustable pipettes and aerosol barrier tips
4. 0.2ml PCR reaction tube strips or plates
5. Optical strip caps
6. Sterile, nuclease free 1.5 ml microcentrifuge tubes
7. Disposable powder-free gloves

### **Equipment:**

1. Microcentrifuge
2. Vortex
3. Real-time PCR detection system with a 96-well format thermocycler reaction block.

## *Procedure*

### **Preparation:**

#### **1. Avoiding sample contamination**

Because of the sensitivity of fluorogenic 5' nuclease assays, special precautions must be taken to avoid false positive amplifications. The following precautionary steps are recommended:

- (a) Maintain separate areas for assay setup and handling of nucleic acids.
- (b) Maintain separate, dedicated equipment (e.g., pipettes, microcentrifuges) and supplies (e.g., microcentrifuge tubes, pipette tips) for assay setup and handling of extracted nucleic acids.
- (c) Wear a clean lab coat and powder-free disposable gloves (not previously worn) when setting up assays.
- (d) Change gloves between samples and whenever you suspect they may be contaminated.
- (e) Keep reagent and reaction tubes capped or covered as much as possible.

#### **2. Equipment preparation**

Work surfaces, pipettes, and centrifuges should be cleaned and decontaminated with cleaning products such as 5% bleach, “DNAzap™” or “RNase AWAY®” to minimize risk of nucleic acid contamination.

### 3. Reagent preparation

**NOTE:** Keep all reagents on cold rack during assay set up.

#### (a) Primers and probes

- Thaw frozen aliquots of primer and probes (Thawed aliquots of probes may be stored in the dark up to 3 months at 2-8°C. Do not re-freeze probes).
- Vortex all primers and probes.
- Briefly centrifuge all primers and probes and then place in cold rack.

#### (b) Realtime RTPCR reagents

- Place Master Mix and enzyme in cold rack
  - Thaw the 2X Reaction Mix vial.
  - Mix the 2X Reaction Mix by inversion.
- 
- Briefly centrifuge 2x Reaction Mix and enzyme then place in cold rack

### Tests for each RT-PCR run

1. Each sample RNA extract is tested by separate primer/probe sets: InfA, Universal swine (swFluA), Swine H1 (swH1) and RNaseP (RP). The RNaseP primer and probe set targets the human RNase P gene and thus serves as an internal positive control for human nucleic acid.
2. No template controls (NTC) and positive template controls (PTC) for all primer/probe sets should be included in each run.
3. Human Specimen Control (HSC) provides a secondary negative control that validates the nucleic extraction procedure and reagent integrity.

### Reaction setup

Reaction assay mixtures are made as a cocktail and dispensed into the 96-well reaction plate. Water and extracted nucleic acid or positive template controls are then added to the appropriate test reactions and controls.

1. Label one 1.5 ml microcentrifuge tube for each primer/probe set.
2. Determine the number of reactions (N) to set up per assay. It is necessary to make excess reaction cocktail to allow for the NTC, PTC, HSC reactions and pipetting error. See below:
  - If number of samples (n) including controls = 1 to 14, then  $N = n + 1$
  - If number of samples (n) including controls > 15, then  $N = n + 2$
3. Master Mix: calculate the amount of each reagent to be added for each primer/probe set reaction master mix. The calculations are as follows:

Reagent	Volume of Reagent Added per Reaction
Nuclease-free Water	N x 5.5 µl
Forward Primer	N x 0.5 µl
Reverse Primer	N x 0.5 µl
Probe	N x 0.5 µl
SuperScript™ III RT/Platinum® Taq Mix	N x 0.5 µl
2X PCR Master Mix	N x 12.5 µl
<b>Total Volume</b>	<b>N x 20.0 µl</b>

After addition of the water, mix reaction mixtures by pipetting up and down. Do not vortex.

5. Centrifuge for 5 sec to collect contents at bottom of the tube, and then place the tube in cold rack.

6. Set up reaction strip tubes or plates in 96-well cooler rack.

7. Dispense 20µl of each master mix into each well going across the row as shown below:

#### Example Test Setup

	1	2	3	4	5	6	7	8	9	10	11	12
A	InfA	InfA	InfA	InfA	InfA	InfA	InfA	InfA	InfA	InfA	InfA	InfA
B	sw InfA	sw InfA	sw InfA	sw InfA	sw InfA	sw InfA	sw InfA	sw InfA	sw InfA	sw InfA	sw InfA	sw InfA
C	swH1	swH1	swH1	swH1	swH1	swH1	swH1	swH1	swH1	swH1	swH1	swH1
D	RP	RP	RP	RP	RP	RP	RP	RP	RP	RP	RP	RP
E	InfA	InfA	InfA	InfA	InfA	InfA	InfA	InfA	InfA	InfA	InfA	InfA
F	sw InfA	sw InfA	sw InfA	sw InfA	sw InfA	sw InfA	sw InfA	sw InfA	sw InfA	sw InfA	sw InfA	sw InfA
G	swH1	swH1	swH1	swH1	swH1	swH1	swH1	swH1	swH1	swH1	swH1	swH1
H	RP	RP	RP	RP	RP	RP	RP	RP	RP	RP	RP	RP

#### Example Sample Setup

	1	2	3	4	5	6	7	8	9	10	11	12
A	NTC	S1	S3	S5	S7	S9	S11	S13	S15	S17	S19	PTC
B	NTC	S1	S3	S5	S7	S9	S11	S13	S15	S17	S19	PTC
C	NTC	S1	S3	S5	S7	S9	S11	S13	S15	S17	S19	PTC
D	NTC	S1	S3	S5	S7	S9	S11	S13	S15	S17	S19	PTC
E		S2	S4	S6	S8	S10	S12	S14	S16	S18	HSC	
F		S2	S4	S6	S8	S10	S12	S14	S16	S18	HSC	
G		S2	S4	S6	S8	S10	S12	S14	S16	S18	HSC	
H		S2	S4	S6	S8	S10	S12	S14	S16	S18	HSC	

**Note:** Negative template controls (NTC) should be added first (column 1) before any of the samples are added to check for contamination in the master mix. HSC should be added after the samples have been added (column 11) to check for cross-contamination during sample preparation or addition. Positive template controls (PTC) should be added last after all samples and NTCs are sealed.

8. Before moving the plate to the nucleic acid handling area, set up the NTC reactions for column 1 in the assay set-up area. As shown above, samples can be added by column.

9. Pipette 5 µl of nuclease free water into the NTC wells. Cap NTC wells.
10. Cover the reaction plate and move the reaction plate to the nucleic acid handling area.
11. Vortex the tubes containing the samples for 5 sec. Centrifuge tubes for 5 sec.
12. Set up the extracted nucleic acid samples in the cold rack.
13. As shown above, samples can be added by column. Pipette 5 µl of the first sample into all the wells labeled for that sample (for example, Sample “S1” as shown above). Change tips after each addition
14. Cap the column to which the sample has been added. This will help to prevent sample cross contamination and enable you to keep track of where you are on the plate.
15. Change gloves when necessary to avoid contamination.
16. Repeat steps 13. through 15. for the remaining samples.
17. Add 5 µl of HSC extracted sample to the HSC wells (column 11). Cap HSC wells.
18. Finally, pipette 5 µl of positive template control RNA into all PTC wells. Cap PTC wells.
19. If using 8-tube strips, label the TAB of each strip to indicate sample position (DO NOT LABEL THE TOPS OF THE REACTION TUBES!). Briefly centrifuge tube strips for 10-15 seconds. Return strip tubes to cold rack.  
If using plates, centrifuge at 500 x g for 30 seconds at 4°C. Return to cold rack.

### RT-PCR amplification conditions

The reaction volume is 25µl. Program the thermocycler as follows:

Reverse Transcription	50°C for 30 min
Taq inhibitor activation	95°C for 2 min
PCR amplification (45 cycles)	95°C for 15 sec 55°C for 30 sec*

\* Fluorescence data (FAM) should be collected during the 55°C incubation step.

### Interpretation/examination:

1. The NTC reactions for probe/primer sets should not exhibit fluorescence growth curves that cross the threshold line. If a false positive occurs with one or more of the primer and probe NTC reactions, sample contamination may have occurred. Invalidate the run and repeat the assay with stricter adherence to the procedure guidelines.
2. All clinical samples should exhibit RP reaction curves that cross the threshold line at or before 37 cycles, thus indicating the presence of sufficient RNA from human RNase P gene indicating the specimen is of acceptable quality. However, it is possible that some samples may fail to give positive reactions due to low cell numbers in the original clinical sample. Also, samples taken from animal/avian species or cell culture typically exhibit either no RP reaction, or a weak RP reaction. Failure to detect RNase P in any of the clinical samples may indicate:
  - (a) Improper extraction of nucleic acid from clinical materials resulting in loss of RNA or carry-over of RT-PCR inhibitors from clinical specimens
  - (b) Absence of sufficient human cellular material in sample to enable detection
  - (c) Improper assay set up and execution
  - (d) Reagent or equipment malfunction
3. The HSC should NOT exhibit fluorescence growth curves for primer/probe sets InfA, swFluA, or swH1 that cross the threshold line within 40 cycles. If any influenza specific primer/probes exhibit a growth curve that crosses the threshold line, interpret as follows:

- (a) Contamination of RNA extraction reagents may have occurred. Invalidate the run and confirm reagent integrity of RNA extraction reagents prior to further testing.
  - (b) Cross contamination of samples occurred during RNA extraction procedures or assay setup. Invalidate the run and repeat the assay with stricter adherence to procedure guidelines.
- 4.** PTC reactions should produce a positive result with the InfA, swInfA, swH1, and RP reactions before 40 cycles. If expected positive reactivity is not achieved, invalidate the run and repeat the assay with stricter adherence to procedure guidelines. Determine the cause of failed PTC reactivity, implement corrective actions, and document results of the investigation and corrective actions. Do not use PTC reagents that do not generate expected result.
- 5.** When all controls meet stated requirements, a specimen is considered presumptive positive for influenza A virus if the InfA reaction growth curves cross the threshold line within 40 cycles. If the reaction for influenza A is positive, it may also be positive for Univ SW and/or SW H1. A specimen is considered presumptive positive for swine influenza A/H1 if BOTH the InfA and the respective subtype (swInfA or swH1) reaction growth curves cross the threshold line within 40 cycles. If a specimen is positive for InfA and only one of the subtype reactions or positive for InfA only, contact CDC for guidance.
- 6.** When all controls meet the stated requirements, a specimen is considered negative for influenza virus if growth curves for neither InfA cross the threshold within 40 cycles.

**Limitations:**

1. Analysts should be trained and familiar with testing procedures and interpretation of results prior to performing the assay.
2. A false negative result may occur if inadequate numbers of organisms are present in the specimen due to improper collection, transport or handling.
3. A false negative result may occur if an excess of DNA/RNA template is present in the in the reaction. If inhibition of the RP control reaction is noted for a particular sample, extracted RNA can be tested at 2 or more dilutions (e.g., 1:10 and 1:100) to verify the result.

## **PRIMER AND PROBE SETS**

*Note: primer/probe sets may undergo periodic modification as information about current circulating viruses is known.*

<b>Primers and Probes</b>	<b>Sequence (5'&gt;3')</b>	<b>Working Concentration</b>
InfA Forward	GAC CRA TCC TGT CAC CTC TGA C	40 μM
InfA Reverse	AGG GCA TTY TGG ACA AAK CGT CTA	40 μM
InfA Probe <sup>1</sup>	TGC AGT CCT CGC TCA CTG GGC ACG	10 μM
SW InfA Forward	GCA CGG TCA GCA CTT ATY CTR AG	40 μM
SW InfA Reverse	GTG RGC TGG GTT TTC ATT TGG TC	40 μM
SW InfA Probe <sup>2</sup>	CYA CTG CAA GCC CA" T" ACA CAC AAG CAG GCA	10 μM
SW H1 Forward	GTG CTA TAA ACA CCA GCC TYC CA	40 μM
SW H1 Reverse	CGG GAT ATT CCT TAA TCC TGT RGC	40 μM
SW H1 Probe <sup>2</sup>	CA GAA TAT ACA "T"CC RGT CAC AAT TGG ARA A	10 μM
RnaseP Forward	AGA TTT GGA CCT GCG AGC G	40 μM
RnaseP Reverse	GAG CGG CTG TCT CCA CAA GT	40 μM
RnaseP Probe <sup>1</sup>	TTC TGA CCT GAA GGC TCT GCG CG	10 μM

<sup>1</sup> TaqMan® probes are labeled at the 5'-end with the reporter molecule 6-carboxyfluorescein (FAM) and with the quencher, Blackhole Quencher 1 (BHQ1) (Biosearch Technologies, Inc., Novato, CA) at the 3'-end.

<sup>2</sup> Taqman® probes are labeled at the 5'-end with the reporter molecule 6-carboxyfluorescein (FAM) and quenched internally at a modified "T" residue with BHQ1, with a modified 3'-end to prevent probe extension by Taq polymerase.



## PureLink<sup>®</sup> Viral RNA/DNA Kits

For rapid, efficient purification of viral nucleic acids from cell-free samples

Catalog Numbers 12280-050, 12280-096

Document Part Number 25-0916

Publication Number MAN0000562

Revision 2.0

For Research Use Only. Not for use in diagnostic procedures.



2

## Contents

Experienced Users' Mini Kit Procedure .....	4
Experienced Users' 96 Kit Procedure.....	5
Kit Contents and Storage .....	7
<b>Introduction.....</b>	<b>8</b>
Overview .....	8
<b>Methods .....</b>	<b>11</b>
Before Starting .....	11
Purification Procedure for Mini Kit .....	15
Example of Expected Results for Mini Kit .....	20
Purification Procedure for 96 Kit.....	21
Example of Expected Results for 96 Kit.....	28
Troubleshooting .....	29
<b>Appendix .....</b>	<b>31</b>
Accessory Products .....	31
Technical Support .....	32
Purchaser Notification.....	33

3

## Experienced Users' Mini Kit Procedure

**Introduction** This quick reference sheet is included for experienced users of the PureLink<sup>®</sup> Viral RNA/DNA Mini Kit. If you are a first time user, follow the detailed protocol in this manual.

Step	Action
Preparing Lysates	<p>The lysate preparation protocol is described below for 200 <math>\mu</math>L starting material. If you wish to process &gt;200 <math>\mu</math>L (<math>\leq</math>500 <math>\mu</math>L) sample volume, scale-up the reagent volumes accordingly.</p> <ol style="list-style-type: none"> <li>1. Add 25 <math>\mu</math>L Proteinase K into a sterile microcentrifuge tube.</li> <li>2. Add 200 <math>\mu</math>L of cell-free sample into the microcentrifuge tube. <b>Note:</b> If you are processing &lt;200 <math>\mu</math>L sample, adjust final volume of the sample to 200 <math>\mu</math>L using PBS (phosphate buffered saline) or 0.9% NaCl.</li> <li>3. Add 200 <math>\mu</math>L Lysis Buffer (containing 5.6 <math>\mu</math>g Carrier RNA). Close the tube lid and mix by vortexing for 15 seconds.</li> <li>4. Incubate at 56°C for 15 minutes.</li> <li>5. Add 250 <math>\mu</math>L 96–100% ethanol to the tube, close the lid, and mix by vortexing for 15 seconds.</li> <li>6. Incubate the lysate for 5 minutes at room temperature.</li> </ol>
Purification Procedure	<ol style="list-style-type: none"> <li>1. <b>Add</b> above lysate to the Viral Spin Column in a collection tube.</li> <li>2. Centrifuge the column at 6800 <math>\times</math> g for 1 minute. Discard the collection tube. Place the spin column in a new Wash Tube.</li> <li>3. <b>Wash</b> the column with 500 <math>\mu</math>L Wash Buffer (W5) with ethanol. Centrifuge at 6800 <math>\times</math> g for 1 minute. Discard the flow through.</li> <li>4. <b>Repeat</b> wash Step 3 with 500 <math>\mu</math>L Wash Buffer (W5) once.</li> <li>5. Discard the collection tube and place the spin column in another, clean Wash Tube.</li> <li>6. Centrifuge the spin column at maximum speed for 1 minute to remove any residual Wash Buffer (W5).</li> <li>7. Place the spin column in a clean 1.7-mL Recovery Tube.</li> <li>8. <b>Elute</b> with 10–50 <math>\mu</math>L sterile RNase-free water (E3) supplied with the kit (add water to the center of the cartridge).</li> <li>9. Incubate at room temperature for 1 minute. Centrifuge the spin column at maximum speed for 1 minute to elute nucleic acids. <i>The Recovery Tube contains purified viral nucleic acids. Discard the spin column.</i></li> <li>10. Store purified viral RNA/DNA at –80°C or use RNA/DNA for the desired downstream application.</li> </ol>

4

## Experienced Users' 96 Kit Procedure

### Introduction

This quick reference sheet is included for experienced users of the PureLink® 96 Viral RNA/DNA Kit. If you are a first time user, follow the detailed protocol in this manual.

Step	Action
Preparing Lysates	<p>The lysate preparation protocol is described below for <b>200 µL</b> starting material.</p> <ol style="list-style-type: none"> <li>Add 25 µL Proteinase K to each well.</li> <li>Add 200 µL of cell-free sample to the well.</li> </ol> <p><b>Note:</b> If you are processing &lt;200 µL sample, adjust final volume of the sample to 200 µL using PBS (phosphate buffered saline) or 0.9% NaCl.</p> <ol style="list-style-type: none"> <li>Add 200 µL Lysis Buffer (containing 5.6 µg Carrier RNA). Mix by pipetting up and down 3–5 times.</li> <li>Incubate at 56°C for 15 minutes. Briefly centrifuge.</li> <li>Add 250 µL 96–100% ethanol to each well and mix by pipetting up and down 3–5 times.</li> <li>Incubate the lysate for 5 minutes at room temperature.</li> </ol>
Purification Procedure with Centrifugation	<p>Perform all centrifuge steps at 25°C.</p> <ol style="list-style-type: none"> <li>Transfer each lysate to the 96-well Filter Plate. Cover unused wells with foil tape.</li> <li>Place the 96-well Filter Plate onto a 96-well, Deep-Well Plate and centrifuge at <math>\geq 2250 \times g</math> for 5 minutes.</li> <li>Remove the flow-through and place the Filter Plate back on the Deep-Well Plate.</li> <li>Add 500 µL Wash Buffer (W5) with ethanol.</li> <li>Centrifuge 5 minutes at <math>\geq 2250 \times g</math> until liquid passes through the Filter Plate.</li> <li>Remove the flow-through and place the Filter Plate back on the Deep-Well Plate.</li> <li>Repeat Steps 4–6.</li> <li>Centrifuge the Filter Plate at <math>\geq 2250 \times g</math> for 10 minutes to dry.</li> <li>Add 100 µL RNase-free water and incubate the plate for 1 minute at room temperature.</li> <li>Place the Filter Plate onto a Receiver Plate. Centrifuge at <math>\geq 2250 \times g</math> for 5 minutes.</li> <li>Use the RNA/DNA for the desired downstream application. To store the purified RNA/DNA, cover the wells with foil tape, and store at –80°C.</li> </ol>

5

## Experienced Users 96 Kit Procedure, Continued

Step	Action
Purification Procedure with Vacuum	<ol style="list-style-type: none"> <li>Place a 96-well Filter Plate in top half of a vacuum manifold.</li> <li>Transfer lysates from Preparing Lysates, Step 6 to the Filter Plate. Cover unused wells with foil tape.</li> <li>Apply vacuum for 2 minutes at room temperature. Release.</li> <li>Add 500 µL Wash Buffer (W5) with ethanol.</li> <li>Apply vacuum 2 minutes at room temperature. Release.</li> <li>Repeat Steps 4 and 5.</li> <li><b>Tap the Filter Plate on a stack of paper towels and pat dry.</b></li> <li>Return the Filter Plate to the vacuum manifold and apply vacuum for 10 minutes at room temperature. Release.</li> <li>Replace the waste collection tray with the Receiver Plate and place the Filter Plate on top of the Receiver Plate.</li> <li>Add 150 µL sterile, RNase-free water (E3) and incubate the plate for 1 minute at room temperature.</li> <li>Apply vacuum for 2 minutes at room temperature. Release.</li> <li>The RNA/DNA is eluted into the Receiver Plate in a volume of 120–130 µL. Use the RNA/DNA for the desired downstream application. To store the purified RNA/DNA, cover the wells with foil tape, and store at –80°C.</li> </ol>

6

## Kit Contents and Storage

### Shipping and Storage

All components of the PureLink® Viral RNA/DNA Kits are shipped at room temperature.

Upon receipt, store all kit components at room temperature, **except store Carrier RNA at –20°C.**

**Note:** The Proteinase K solution is stable for 1 year when stored at room temperature. For long-term storage (>1 year) or if room temperature is >25°C, store the Proteinase K solution at 4°C.

### Types of Kits

This manual supports the following kits.

Kit	Reactions	Catalog No.
PureLink® Viral RNA/DNA Mini Kit	50	12280-050
PureLink® 96 Viral RNA/DNA Kit	384	12280-096

### Contents

The components and amounts included in the PureLink® Viral RNA/DNA Kits are listed in the following table.

**Note:** Since the Mini Kit is designed for purifying samples using a starting volume of  $\leq 500$  µL, some reagents in the kit maybe provided in excess in the amount needed. The 96 Kit sample volume is limited to 200 µL or less.

Component	12280-050	12280-096
Viral Lysis Buffer (L22)	32 mL	100 mL
Wash Buffer (W5) (5X)	15 mL	2 × 87.5 mL
Proteinase K (20 mg/mL) in storage buffer (proprietary)	2 × 1.6 mL	10 mL
Carrier RNA (lyophilized)	310 µg	2.2 mg
Sterile, RNase-free Water (E3)	15.5 mL	75 mL
Viral Spin Columns with Collection Tubes	50 each	—
Wash Tubes (2.0-mL)	2 × 50	—
Recovery Tubes (1.5-mL)	50 each	—
PureLink® Filter Plate	—	4 each
Receiver Plate	—	4 each
Deep-Well Plate	—	6 each
Foil Tape	—	3 × 4/bag

7

## Introduction

### Overview

#### Introduction

The PureLink® Viral RNA/DNA Kits provide a rapid and efficient method to simultaneously purify viral RNA/DNA from fresh or frozen cell-free biological fluids (plasma, serum, cerebrospinal fluid) and cell culture supernatants.

The PureLink® Viral Mini Kit is specifically designed to isolate high-quality viral nucleic acids from a variety of RNA and DNA viruses within 45 minutes using low elution volumes that allow sensitive downstream analysis.

The PureLink® 96 Viral Kit is specifically designed for obtaining the highest yield with samples of low titer. The kit allows high throughput isolation of viral nucleic acid using Tecan or Beckman instruments.

The purified viral RNA/DNA is devoid of proteins and nucleases, and is suitable for use in downstream applications that allow viral detection and genotyping.

#### System Overview

Both kits allow efficient lysis of viral particles at elevated temperatures using Proteinase K and selective binding of viral nucleic acids to the silica matrix under highly denaturing conditions.

The viral particles in the cell-free samples are lysed using Proteinase K and Lysis Buffer (L22) containing Carrier RNA at 56°C. The Lysis Buffer (L22) is specifically formulated to allow lysis of different types of viral particles.

Ethanol is added to the lysate to a final concentration of 37% and the sample is loaded onto a silica spin column (Mini Kit) or silica Filter Plate (96 Kit). The viral RNA/DNA molecules bind to the silica-based media and impurities such as proteins and nucleases are removed by thorough washing with Wash Buffer. The RNA/DNA is then eluted in sterile, RNase free water.

8



## Before Starting, Continued

### Preparing Carrier RNA

The recommended purification protocol uses 5.6 µg Carrier RNA per sample (for ≤500 µL sample). If you wish to use less Carrier RNA per sample, you need to validate the amount of Carrier RNA needed for each sample type and downstream application.

To prepare Carrier RNA (5.6 µg/sample):

1. Add 310 µL RNase-free Water (included with the kit) to 310 µg lyophilized Carrier RNA supplied in a tube with the kit to obtain 1 µg/µL Carrier RNA stock solution.
2. Mix thoroughly and aliquot the solution into smaller aliquots. Store the aliquots at -20°C. Avoid repeated freezing and thawing.
3. Calculate the volume of Lysis Buffer/Carrier RNA mix required to process the desired number of samples simultaneously using the following formula:  

$$N \times 0.21 \text{ mL (volume of Lysis Buffer/reaction)} = A \text{ mL}$$

$$A \text{ mL} \times 28 \text{ µL/mL} = B \text{ µL}$$
 where  
 N = number of samples  
 A = calculated volume of Lysis Buffer (L22)  
 B = calculated volume of 1 µg/µL Carrier RNA stock solution to add to Lysis Buffer (L22)
4. Thaw the required amount of 1 µg/µL Carrier RNA stock solution.
5. In a sterile tube, add the volume of Carrier RNA stock solution (B, calculated as above) to the volume of Lysis Buffer (A, calculated as above). Mix gently by pipetting up and down. Avoid vortexing as it generates foam.
6. Store at 4°C until use. **Use the buffer within 1 hour.**

#### Example:

The example below shows the amount of Lysis Buffer (L22) and Carrier RNA stock solution required to process 10 samples using the above formula:

$$10 \times 0.21 \text{ mL (volume of Lysis Buffer/reaction)} = 2.1 \text{ mL}$$

$$2.1 \text{ mL} \times 28 \text{ µL/mL} = 58.8 \text{ µL}$$

To prepare Lysis Buffer containing Carrier RNA for processing 10 samples, mix 58.8 µL Carrier RNA stock solution with 2.1 mL Lysis Buffer (L22).

13

## Before Starting, Continued

### Instrument Compatibility for 96-Well Plates

The PureLink® Filter Plates are compatible with the following instruments:

- **Vacuum Manifold:** The manifold must accommodate the PureLink® Plates and be capable of collecting the filtrate (e.g., UniVac 3 Vacuum Manifold System from Whatman)
- **Centrifuge:** Must be capable of centrifuging 96-well plates obtaining  $\geq 2250 \times g$ , and accommodate a 5.0-cm microtiter plate stack.
- **Automated Liquid Handling Workstation:** The workstation must be equipped with a vacuum manifold and a vacuum source. The PureLink® Filter Plate has been tested and qualified for use on the Biomek® FX and the Tecan Freedom EVO™ Workstations.

For the Biomek® FX, you may use the Small Vacuum Adapter Collar (Whatman Cat. no. 7705-0120) with PureLink® Plates.

The PureLink® Filter Plates are compatible with the QIAvac Vacuum Manifold, but **not** with the BioRobot® Workstations.

### Calibrating Vacuum for Use with 96-Well Plates

We recommend using a vacuum pressure of -12 to -15 in. Hg to obtain the best results.

Using higher vacuum pressure than the recommended pressure may cause sample splattering or inefficient DNA binding, while using lower vacuum pressure will affect the elution resulting in lower recovery.

To check the vacuum pressure:

1. Place an unused PureLink® Filter Plate on top of the vacuum manifold.
2. Apply vacuum and check the vacuum pressure on the vacuum regulator (usually attached to the manifold or a vacuum pump).
3. Adjust the vacuum pressure on the regulator to obtain the recommended pressure of -12 to -15 in. Hg.

**Note:** During purification the vacuum pressure may exceed the recommended value.

14

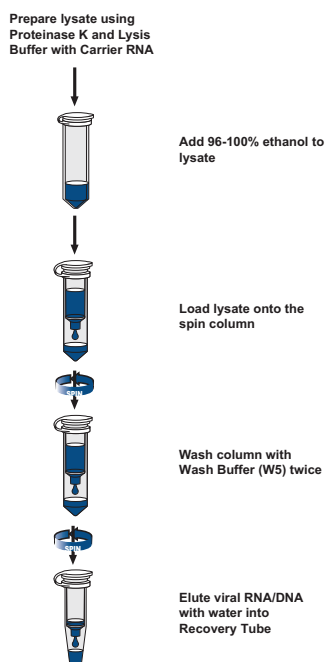
## Purification Procedure for Mini Kit

### Introduction

The viral nucleic acid purification procedure is described below using spin column based centrifugation in a total time of ~45 minutes.

### Experimental Overview

The flow chart for purifying viral RNA/DNA using the PureLink® Viral RNA/DNA Mini Kit is shown below.



15

## Purification Procedure for Mini Kit, Continued



Follow the recommendations below to obtain the best results:

- Perform all centrifugation steps at room temperature
- Be sure to add ethanol to Wash Buffer (W5) before use (page 12)
- Perform the recommended wash steps to obtain high-quality RNA
- Always pipet water in the center of the Viral Spin Column and perform a 1 minute incubation before elution

### Important

- Handle all viruses in compliance with established institutional guidelines. Since safety requirements for use and handling of viruses may vary at individual institutions, we recommend consulting the health and safety guidelines and/or officers at your institution.
- Be sure to take the appropriate precautions (wear a laboratory coat, disposable gloves, and eye protection) when handling viral samples.
- The eluates collected during wash steps contain biohazardous waste. Dispose the eluate and collection tubes appropriately as biohazardous waste.

### Elution Volume

The PureLink® Viral RNA/DNA Mini Kit utilizes low, recommended elution volume of 10–50 µL to elute viral nucleic acid resulting in highly concentrated viral nucleic acids that is required for sensitive downstream applications. You may elute the viral nucleic acids in an elution volume ranging from 10–150 µL depending on your downstream applications.

**Note:** Using larger elution volume decreases the viral nucleic acid concentration in the eluate.

16

## Overview, Continued

### Advantages

The PureLink® Viral RNA/DNA Mini Kit provides the following advantages:

- Rapid and efficient purification of high-quality viral nucleic acid using spin column-based centrifugation with no sample cross-contamination
- Specifically designed to purify viral RNA and DNA from ≤500 µL cell-free samples within 45 minutes
- Ability to elute viral nucleic acids in low elution volumes of 10–50 µL to allow sensitive downstream analysis

In contrast, the PureLink® 96 Viral RNA/DNA Kit provides:

- Highest sensitivity for samples with very low viral titer
- High throughput processing of samples
- Automation using Tecan or Beckman robotic systems

Lastly, **both** kits provide:

- Purified nucleic acid free of contaminants such as proteins and nucleases
- Reliable performance of the purified viral nucleic acids in downstream applications

### Carrier RNA

The Carrier RNA included with the PureLink® Viral RNA/DNA Kits is yeast tRNA (page 31). The presence of an excess amount of Carrier RNA as compared to viral nucleic acids during lysate preparation and purification:

- Increases the binding of viral nucleic acids to the silica matrix
- Reduces any viral nucleic acid degradation from nucleases present in the sample

The purification protocol recommends using 5.6 µg Carrier RNA for 200–500 µL of sample. Most of the Carrier RNA is removed during the purification process because it is <200 bp and any remaining Carrier RNA does not interfere with downstream applications. However, depending on your application, you may validate the assay using less Carrier RNA.

9

## Methods

### Before Starting

#### Introduction

Review the information in this section before starting. Guidelines are included for the recommended amount of starting material for use and to obtain high-quality RNA.



#### **WARNING! GENERAL CHEMICAL HANDLING.**

For every chemical, read the Safety Data Sheets (SDSs) and follow the handling instructions. Wear appropriate protective eyewear, clothing, and gloves. Safety Data Sheets (SDSs) are available from [www.lifetechnologies.com/support](http://www.lifetechnologies.com/support).

The PureLink® Plant Total DNA Purification Kit buffers contain guanidine hydrochloride. Contact with acids or bleach liberates toxic gases. DO NOT ADD acids or bleach to any liquid wastes containing this product.



#### **Important**

If there is any precipitate present in the buffers, warm the buffer up to 25°C to 37°C to dissolve the precipitate before use.

#### General Handling of RNA

Observe the following guidelines to prevent RNase contamination:

- Use disposable, individually wrapped, sterile plastic ware
- Use only sterile, new pipette tips (aerosol-barrier pipet tips recommended) and microcentrifuge tubes
- Wear latex gloves while handling reagents and RNA samples to prevent RNase contamination from the surface of the skin
- Always use proper microbiological aseptic techniques when working with RNA
- Use RNase AWAY® Reagent (page 31) to remove RNase contamination from surfaces

11

## Overview, Continued

### Proteinase K

The Proteinase K is used for efficient lysis of viral particles. Proteinase K is active in the highly denaturing conditions of the lysis step.

### Downstream Applications

The purified viral RNA and DNA is suitable for use in RT-PCR, qRT-PCR, and qPCR, and can be used for:

- Viral load monitoring
- Viral detection
- Viral genotyping

### Mini Kit Specifications

Starting Material:	≤500 µL cell-free sample
Binding Capacity:	~5 µg nucleic acid
Column Reservoir Capacity:	700 µL
Wash Tube Capacity:	2.0 mL
Recovery Tube Capacity:	1.5 mL
Centrifuge Compatibility:	Capable of centrifuging at >10,000 × g
Elution Volume:	10–50 µL

### 96 Kit Specifications

Dimensions:	Standard SBS (Society for Biomolecules Screening) footprint
Starting Material:	≤200 µL cell-free sample
Binding Capacity:	At least 40 µg nucleic acid
Filter Reservoir Capacity:	1.0 mL
Receiver Plate Capacity:	0.33 mL (0.1–0.2 mL)
Deep-Well Plate Capacity:	1.09 mL
Centrifuge Compatibility:	Capable of centrifuging at ≥2250 × g Bucket depth 5 cm
Elution Volume:	100–150 µL

10

## Before Starting, Continued

### Starting Material

The PureLink® Viral RNA/DNA Kits are designed to isolate viral nucleic acid from cell-free biological fluids such as plasma, serum, and CSF (cerebrospinal fluid) as well as cell culture supernatant using fresh or frozen samples.

To obtain high yield of viral nucleic acids and minimize any degradation, follow these guidelines:

- Collect the sample (such as plasma or serum) and proceed immediately to the purification protocol (page 15 if using the Mini Kit; page 21 if using the 96 Kit). If desired, you can store the sample at 4°C for short-term storage (≤4 hours) or freeze the sample at –20°C or –80°C for long-term storage.
- Do not freeze-thaw the plasma or serum sample more than once.
- Remove any visible cryoprecipitates from samples by centrifugation at ~7000 × g for 2–3 minutes. Use the clear supernatant immediately for purification.
- If you need to concentrate the cell culture supernatant use appropriate centrifugal concentrators.

### Sample Volume

The PureLink® Viral RNA/DNA Mini Kit can process sample volumes of ≤500 µL without preparation of any additional buffers or carry over of any impurities. After preparing the lysate, if the final lysate volume is >600 µL, you need to perform multiple loadings of the final lysate onto the spin column.

The 96 Kit processes sample volumes of ≤200 µL.

### Preparing Wash Buffer

#### For the Mini Kit:

Add 60 mL 96–100% ethanol to 15 mL Wash Buffer (W5) included with the kit.

#### For the 96 Kit:

Add 350 mL of 96–100% ethanol to 87.5 mL Wash Buffer (W5) included in the kit.

Place a check in the box on the Wash Buffer label to indicate the ethanol is added. Store the Wash Buffer (W5) with ethanol at room temperature.

12

## Purification Procedure for Mini Kit, Continued

### Materials Needed

#### Components supplied by the user

- Cell-free samples (such as plasma or serum samples)
- Appropriate amount of Lysis Buffer (L22) containing Carrier RNA, page 13
- Heat block set to 56°C
- 96–100% ethanol
- Sterile 1.5-mL or 2-mL microcentrifuge tubes
- Microcentrifuge capable of centrifuging  $>10,000 \times g$
- *Optional:* Sterile, RNase-free PBS or 0.9% NaCl

#### Components supplied with the Kit

- Wash Buffer (W5)
- Sterile, RNase-free Water (E3)
- Viral Spin Column in Collection Tubes
- Wash Tubes and Recovery Tubes

### Preparing Lysate

The lysate preparation protocol is described below for 200  $\mu\text{L}$  starting material. If you wish to process  $>200 \mu\text{L}$  ( $\leq 500 \mu\text{L}$ ) sample volume, scale-up the reagent volumes accordingly.

**Note:** There is no need to scale-up the amount of Carrier RNA. Use  $\leq 5.6 \mu\text{g}$  Carrier RNA per sample volume of  $\leq 500 \mu\text{L}$ .

1. Add 25  $\mu\text{L}$  Proteinase K (included with the kit) into a sterile microcentrifuge tube.
2. Add 200  $\mu\text{L}$  of cell-free sample (equilibrated to room temperature) into the microcentrifuge tube.  
**Note:** If you are processing  $<200 \mu\text{L}$  sample, adjust final volume of the sample to 200  $\mu\text{L}$  using PBS or 0.9% NaCl.
3. Add 200  $\mu\text{L}$  Lysis Buffer (containing 5.6  $\mu\text{g}$  Carrier RNA). Close the tube lid and mix by vortexing for 15 seconds.
4. Incubate at 56°C for 15 minutes.
5. Briefly centrifuge the tube to remove any drops from the inside of the lid.
6. Proceed immediately to **Binding and Washing Step**, page 18.

17

## Purification Procedure for Mini Kit, Continued

### Binding and Washing Step

1. Add 250  $\mu\text{L}$  96–100% ethanol to the lysate tube to obtain a final ethanol concentration of 37%, close the lid, and mix by vortexing for 15 seconds.  
**Note:** If you are processing up to 10 samples, you may add ethanol to all tubes and then vortex each tube.
2. Incubate the lysate with ethanol for 5 minutes at room temperature.
3. Briefly centrifuge the tube to remove any drops from the inside of the lid.
4. Transfer the above lysate with ethanol ( $\sim 675 \mu\text{L}$ ) onto the Viral Spin Column.
5. Close the lid and centrifuge the column at  $\sim 6800 \times g$  for 1 minute. Discard the collection tube with the flow-through.  
**Note:** If you are processing  $>200 \mu\text{L}$  starting material, you need to perform multiple loadings of the lysate by transferring any remaining lysate to the same Viral Spin Column and centrifuge at  $6800 \times g$  for 1 minute.
6. Place the spin column in a clean Wash Tube (2 mL) included with the kit and add 500  $\mu\text{L}$  Wash Buffer (W5) with ethanol to the spin column.
7. Close the lid and centrifuge the column at  $\sim 6800 \times g$  for 1 minute. Discard the flow-through and place the spin column back into the Wash Tube.  
**Note:** Additional Wash Tubes are available separately (page 31), if you do not wish to reuse the Wash Tube.
8. Add 500  $\mu\text{L}$  Wash Buffer (W5) with ethanol into the spin column.
9. Close the lid, centrifuge at  $\sim 6800 \times g$  for 1 minute. Discard the Wash Tube containing the flow-through.
10. Place the spin column in another clean, Wash Tube (2 mL) included with the kit.
11. Centrifuge the column at maximum speed in a microcentrifuge for 1 minute to dry the membrane completely. Discard the Wash Tube with the flow-through.
12. Proceed to the **Elution Step**, page 19.

18

## Purification Procedure for Mini Kit, Continued

### Elution Step

1. Place the Viral Spin Column in a clean 1.5-mL Recovery Tube supplied with the kit.
2. Add 10–50  $\mu\text{L}$  of Sterile, RNase-free water (E3) to the center of the column. Close the lid.  
**Note:** You may use an elution volume of  $\leq 150 \mu\text{L}$  for elution (page 16).
3. Incubate at room temperature for 1 minute.
4. Centrifuge the column at maximum speed for 1 minute. The Recovery Tube contains purified viral nucleic acids. Remove and discard the spin column.
5. Store the purified RNA/DNA at  $-80^\circ\text{C}$  or use the RNA/DNA for the desired downstream application.

### Analyzing Viral RNA/DNA

Since the amount of viral RNA/DNA present in cell-free body fluids is low and there is considerable amount of Carrier RNA in the purified viral RNA/DNA sample, we do not recommend using UV absorbance at 260 nm or Quant-iT<sup>TM</sup> Assay Kits to determine the viral nucleic acid yields.

To determine viral nucleic acid yield, use qRT-PCR or RT-PCR for RNA virus, and qPCR and PCR for DNA virus using appropriate viral-specific probes.

To analyze viral nucleic acid size, use agarose gel electrophoresis followed by hybridization using viral specific labeled probes and autoradiography.

19

## Example of Expected Results for Mini Kit

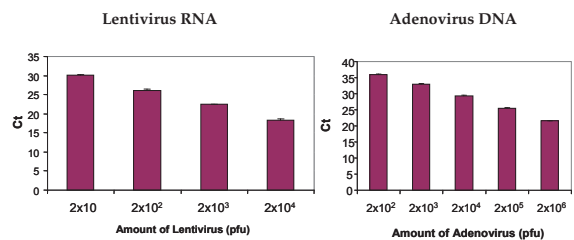
### Results

Examples of results obtained after purification of viral RNA/DNA using the PureLink<sup>®</sup> Viral RNA/DNA Mini Kit are shown in the following figure.

Serum samples (200  $\mu\text{L}$ ) were spiked with lentivirus RNA (left panel) or adenovirus DNA (right panel) at the indicated pfu. Viral RNA/DNA was purified using the PureLink<sup>®</sup> Viral RNA/DNA Mini Kit as described in this manual. Elution was performed with 50  $\mu\text{L}$  RNase-free water.

For lentivirus RNA sample, 5  $\mu\text{L}$  of the purified RNA was used to perform qRT-PCR using the SuperScript<sup>™</sup> III Platinum<sup>®</sup> One-Step qRT-PCR Kit w/ROX with TaqMan<sup>®</sup> primers in an ABI 7700 instrument. For adenovirus DNA sample, 5  $\mu\text{L}$  of purified DNA was used to perform qPCR using the Platinum<sup>®</sup> Quantitative PCR SuperMix-UDG w/ROX kit (page 31) with LUX<sup>™</sup> primers in an ABI 7700 instrument.

**Results:** Consistently lower Ct values indicate an increase in sensitivity of detection for viral RNA and DNA using the PureLink<sup>®</sup> Viral RNA/DNA Mini Kit.

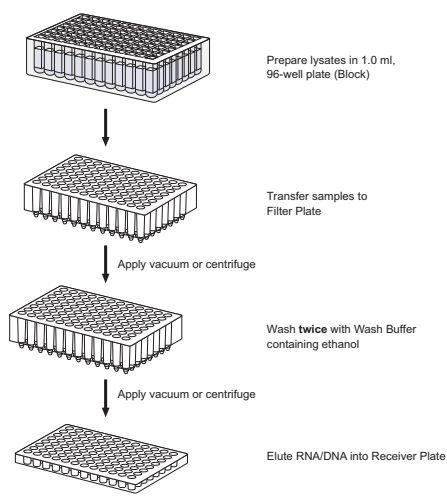


20

## Purification Procedure for 96 Kit

**Introduction** The viral nucleic acid purification procedure described here uses plate-based centrifugation or vacuum for high-throughput processing of samples.

**Experimental Overview** The flow chart for purifying viral RNA/DNA using the PureLink® 96 Viral RNA/DNA Kit is shown in the following diagram.



21

## Purification Procedure for 96 Kit, Continued



Follow the recommendations below to obtain the best results:

- Perform all centrifugation steps at room temperature
- Be sure to add ethanol to Wash Buffer (W5) before use (page 12)
- Perform the recommended wash steps to obtain high-quality RNA
- Always pipet water in the center of the silica matrix and perform a 1 minute incubation before elution

### Important

- Handle all viruses in compliance with established institutional guidelines. Since safety requirements for use and handling of viruses may vary at individual institutions, we recommend consulting the health and safety guidelines and/or officers at your institution.
- Be sure to take the appropriate precautions (wear a laboratory coat, disposable gloves, and eye protection) when handling viral samples.
- The eluates collected during wash steps contain biohazardous waste. Dispose the eluate and collection tubes appropriately as biohazardous waste.

### Elution Volume

The PureLink® 96 Viral RNA/DNA Kit elutes viral nucleic acid in 100–150  $\mu\text{L}$  for downstream applications.

**Note:** Using larger elution volume decreases the viral nucleic acid concentration in the eluate.

22

## Purification Procedure for 96 Kit, Continued

**Automation Scripts** For automation scripts, contact Technical Support page 32.

### Materials Needed

- Cell-free samples (such as plasma or serum samples)
- Appropriate amount of Lysis Buffer (L22) containing Carrier RNA, page 13
- Heat block for 96-well plate, water bath, or incubator set to 56°C
- 96–100% ethanol
- Sterile, RNase-free PBS or 0.9% NaCl
- Vacuum manifold and a vacuum pump for 96-well plates (producing pressure of –12 to –15 in. Hg or –800 to –900 mbar) or automated liquid handling workstations
- or
- Centrifuge with buckets to hold 96-well plates and are 5 cm (2 inches) deep. Centrifuge speed must be  $\geq 2250 \times g$  and temperature should be 25°C

*Components supplied with the kit*

- Wash Buffer (W5)
- Sterile, RNase-free Water (E3)
- 96-well PureLink® Filter Plates
- 96-well Receiver Plates
- 96-well Deep-Well Plates
- Foil tape

23

## Purification Procedure for 96 Kit, Continued

### Preparing Lysate

The lysate preparation protocol is described below for 200  $\mu\text{L}$  starting material.

1. Take one Deep-Well Plate and add 25  $\mu\text{L}$  Proteinase K per well for each sample.
2. Add 200  $\mu\text{L}$  of cell-free sample (equilibrated to room temperature) to each well.  
**Note:** If sample is less than 200  $\mu\text{L}$ , adjust the final volume of the sample to 200  $\mu\text{L}$  using sterile PBS or 0.9% NaCl.
3. Add 200  $\mu\text{L}$  Lysis Buffer (containing 5.6  $\mu\text{g}$  Carrier RNA). Mix by pipetting up and down 3–5 times.
4. Seal the plate with the foil tape and incubate at 56°C for 15 minutes.
5. Centrifuge the plate briefly to collect the liquid in the bottom of the well.
6. Add 250  $\mu\text{L}$  96–100% ethanol to each well to obtain a final ethanol concentration of 37% and mix by pipetting up and down 3–5 times.  
**Note:** If you are processing multiple samples, you may add ethanol to all wells and then mix.
7. Incubate the lysate with ethanol for 5 minutes at room temperature.
8. If you are using centrifugation for purification, follow the procedure on page 25. If you are using a vacuum manifold for purification, proceed to page 26.

24

## Purification Procedure for 96 Kit, Continued

### Purification Procedure Using Centrifugation

1. Transfer each lysate sample to a well of the 96-well Filter Plate. Cover any unused wells with foil tape.
2. Place the 96-well Filter Plate onto a new or used 96-well Deep-Well Plate and centrifuge at  $\geq 2250 \times g$  for 5 minutes.
3. Remove the flow-through and place the Filter Plate back on the Deep-Well Plate.
4. Add 500  $\mu\text{L}$  Wash Buffer (W5) with ethanol into each well of the 96-well Filter Plate.
5. Centrifuge at maximum speed until all liquid passes through the Filter Plate, about 5 minutes.
6. Remove the flow-through and place the Filter Plate back on the Deep-Well Plate.
7. Repeat Steps 4–6.
8. Centrifuge the Filter Plate at  $\geq 2250 \times g$  for 5 minutes to dry the plate completely.
9. Add 100  $\mu\text{L}$  RNase-free water (E3) to the center of the membrane in each well and incubate the plate for 1 minute at room temperature.
10. Place the Filter Plate onto a Receiver Plate (supplied in the kit). Centrifuge at  $\geq 2250 \times g$  for 5 minutes.
11. Use the RNA/DNA for the desired downstream application. To store the purified RNA/DNA, cover the wells with foil tape, and store at  $-80^\circ\text{C}$ .

25

## Purification Procedure for 96 Kit, Continued

### Purification Procedure Using a Vacuum Manifold

Assemble the vacuum manifold as per the manufacturer's instructions.

1. Place a 96-well Filter Plate into the vacuum manifold designed to hold a 96-well plate.
2. Transfer all of the lysates from each well of the Deep-Well Plate to a fresh well in the Filter Plate. Cover unused wells with foil tape.
3. Apply vacuum for 2 minutes at room temperature and release vacuum.
4. Add 500  $\mu\text{L}$  Wash Buffer (W5) with ethanol into each well of the 96-well Filter Plate.
5. Apply vacuum for 2 minutes at room temperature. Release vacuum.
6. Repeat Steps 4 and 5.
7. After releasing the vacuum, tap the Filter Plate with the filter side down on a stack of paper towels and pat dry the plate. Alternatively, you can centrifuge the Filter Plate at  $\geq 2250 \times g$  for 5 minutes to dry the plate completely.
8. Place the Filter Plate on the vacuum manifold and apply vacuum for 10 minutes at room temperature. Release vacuum.
9. Proceed to elution using vacuum manifold (Steps 10–12, below) or using centrifugation (Steps 9–10, page 25).
10. Place the Receiver Plate included in the kit in the vacuum manifold in place of the waste collection tray and place the Filter Plate on top of the Receiver Plate.
11. Add 150  $\mu\text{L}$  of Sterile, RNase-free water (E3) to the center of the membrane in each well of the Filter Plate and incubate the plate for 1 minute at room temperature.
12. Apply vacuum for 2 minutes at room temperature. Release vacuum. The RNA/DNA is eluted into the Receiver Plate in a volume of 120–130  $\mu\text{L}$ .
13. Use the RNA/DNA for the desired downstream application. To store the purified RNA/DNA, cover the wells with foil tape, and store at  $-80^\circ\text{C}$ .

26

## Purification Procedure for 96 Kit, Continued

### Analyzing Viral RNA/DNA

Since the amount of viral RNA/DNA present in cell-free body fluids is low we recommend that you do not use UV absorbance to determine yield. Use qRT-PCR or RT-PCR for RNA virus, and qPCR and PCR for DNA virus using appropriate viral-specific probes to determine yield or presence of viral nucleic acid.

To analyze viral nucleic acid size, use agarose gel electrophoresis followed by hybridization using viral specific labeled probes and autoradiography.

27

## Example of Expected Results for 96 Kit

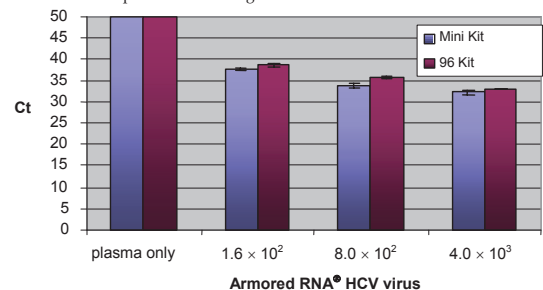
### Results

Examples of results obtained after purification of viral RNA/DNA using the PureLink® 96 Viral RNA/DNA Kit are shown in the following figure.

Serum or plasmid samples (200  $\mu\text{L}$ ) were spiked with Armored RNA® Hepatitis C Virus (HCV) virus at the indicated pfu. Viral RNA was purified using the PureLink® 96 Viral RNA/DNA Kit or the PureLink® RNA/DNA Mini Kit as described in this manual.

qRT-PCR was performed on the resulting samples using the SuperScript™ III One-Step qRT-PCR Kit w/Platinum® Taq with TaqMan® primers in an ABI 7700 instrument.

**Results:** Consistently lower Ct values indicate an increase in sensitivity of detection for viral RNA and comparable performance using either PureLink® kit.



28

## Troubleshooting

Problem	Cause	Solution
Low nucleic acid yield	Incomplete lysis or the column or matrix is clogged	Be sure to use the appropriate volumes of reagents during lysate preparation. If you are processing >200 µL sample for the Mini Kit, adjust the reagent volumes accordingly to obtain complete lysis.
		Do not use more than 200 µL sample when using the 96 Kit.
		If cryoprecipitates are visible in frozen viral samples, remove the cryoprecipitates by centrifugation (page 12) to avoid clogging of the column or matrix.
	Poor quality of sample material	Precipitates in buffers may affect lysis. Dissolve precipitates by warming the buffers at 25°C to 37°C.
		Avoid repeated freezing and thawing of samples. Use fresh samples and process immediately after collection or use samples thawed only once for best results.
		Check the quality of the RNA in the original samples using qRT-PCR or RT-PCR.
Lysis Buffer without Carrier RNA used or Carrier RNA inactivated	To prepare lysates from cell-free samples, use Lysis Buffer (L22) with Carrier RNA (page 13). Once the Carrier RNA is reconstituted in water, aliquot the Carrier RNA and store at -20°C. Do not perform multiple freeze-thaw cycles.	
Incorrect binding conditions	For efficient binding of viral nucleic acids, always <b>add</b> ethanol to the lysate to a final concentration of 37% prior to loading the lysate onto the column or matrix.	
Ethanol not added to Wash Buffer (W5)	Be sure to add 96–100% ethanol to Wash Buffer (W5) as described on page 12. Do not use denatured 95% ethanol.	
Incorrect elution conditions	Add water to the center of the silica matrix and perform incubation for 1 minute with water before eluting.	

29

## Appendix

### Accessory Products

#### Additional Products

The following products are also available from Life Technologies. For details on these products, visit [www.lifetechnologies.com](http://www.lifetechnologies.com) or contact Technical Support (page 32).

Product	Quantity	Catalog No.
Wash Tubes (2.0 mL)	100	12282-100
RNase AWAY®	250 mL	10328-011
PureLink® Foil Tape	50 pieces	12261-012
Viral Lysis Buffer (L22)	500 mL	12282-500
UltraPure™ DEPC-treated Water	1 L	750023
UltraPure™ DNase/RNase-Free Distilled Water	500 mL	10977-015
Yeast tRNA	25 mg	15401-011
Phosphate Buffered Saline (PBS), 1X	500 mL	10010-023
SuperScript™ III One-Step RT-PCR System with Platinum® <i>Taq</i> DNA Polymerase	100 reactions	12574-026
RNA UltraSense™ One-Step Quantitative RT-PCR System	100 reactions	11732-927
Platinum® Quantitative PCR SuperMix-UDG w/ROX	100 reactions	11743-100

31

## Troubleshooting, Continued

Problem	Cause	Solution
Low nucleic acid yield, continued	RNA quantitation performed using UV absorbance	Since viral nucleic acids are present in low amounts in cell-free samples, <b>do not</b> use UV absorbance for quantitation. Analyze viral nucleic acids using qRT-PCR, RT-PCR, qPCR, or PCR.
	RNA degraded	Follow the guidelines on page 11 to prevent RNase contamination.
Poor performance of nucleic acids in downstream enzymatic reactions	Poor quality of samples	Always use fresh samples or samples frozen at -80°C. For lysis, process the sample quickly to avoid degradation.
	Presence of ethanol or use of denatured 95% ethanol in purified nucleic acids	Traces of ethanol from the Wash Buffer (W5) can inhibit downstream enzymatic reactions. To remove Wash Buffer (W5), discard Wash Buffer (W5) flow through. Always use a new Wash Tube or Receiver Plate and completely dry the column or membrane. Use only 96–100% ethanol. Do not use denatured 95% ethanol.
	Assay may be sensitive to Carrier RNA concentration	You may need to optimize the amount of Carrier RNA that is required for optimal purification and is suitable for your downstream applications.
	Reagents for enzymatic reactions inactive	Ensure that the enzymes and reagents used for performing downstream applications have not expired or inactivated. Repeat the reaction with fresh enzyme and reagents.
	Viral nucleic acid eluate too dilute	Optimize the amount of viral nucleic acid eluate required for your specific application and perform elution using the desired kit and elution volume (10–150 µL).
Carrier RNA not enough to process samples	Incorrect Carrier RNA amount used per sample We recommend using a maximum of 5.6 µg Carrier RNA per sample when processing ≤500 µL sample volume. <b>Do not</b> increase the Carrier RNA amount when processing 500 µL sample volume using the Mini Kit.	

30

## Technical Support

#### Obtaining Support

For the latest services and support information for all locations, go to [www.lifetechnologies.com](http://www.lifetechnologies.com).

At the website, you can:

- Access worldwide telephone and fax numbers to contact Technical Support and Sales facilities
- Search through frequently asked questions (FAQs)
- Submit a question directly to Technical Support ([techsupport@lifetech.com](mailto:techsupport@lifetech.com))
- Search for user documents, SDSs, vector maps and sequences, application notes, formulations, handbooks, certificates of analysis, citations, and other product support documents
- Obtain information about customer training
- Download software updates and patches

#### Safety Data Sheets (SDS)

Safety Data Sheets (SDSs) are available at [www.lifetechnologies.com/support](http://www.lifetechnologies.com/support).

#### Certificate of Analysis

The Certificate of Analysis provides detailed quality control and product qualification information for each product. Certificates of Analysis are available on our website. Go to [www.lifetechnologies.com/support](http://www.lifetechnologies.com/support) and search for the Certificate of Analysis by product lot number, which is printed on the box.

#### Limited Product Warranty

Life Technologies Corporation and/or its affiliate(s) warrant their products as set forth in the Life Technologies' General Terms and Conditions of Sale found on Life Technologies' website at [www.lifetechnologies.com/termsandconditions](http://www.lifetechnologies.com/termsandconditions). If you have any questions, please contact Life Technologies at [www.lifetechnologies.com/support](http://www.lifetechnologies.com/support).

32



## Purchaser Notification

### Robotic Scripts Disclaimer

Any scripts or methods for robotic systems have been developed with minimal involvement from the manufacturer and are provided only as a convenience to customers to use at their own risk. Life Technologies assumes no responsibility or liability for any special, incidental, direct, or consequential loss or damages whatsoever. Life Technologies provides no warranties of any kinds, express or implied, including without limitation implied warranties or merchantability or fitness for a particular purpose.

### Disclaimer

LIFE TECHNOLOGIES CORPORATION AND/OR ITS AFFILIATE(S) DISCLAIM ALL WARRANTIES WITH RESPECT TO THIS DOCUMENT, EXPRESSED OR IMPLIED, INCLUDING BUT NOT LIMITED TO THOSE OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, OR NON-INFRINGEMENT. TO THE EXTENT ALLOWED BY LAW, IN NO EVENT SHALL LIFE TECHNOLOGIES AND/OR ITS AFFILIATE(S) BE LIABLE, WHETHER IN CONTRACT, TORT, WARRANTY, OR UNDER ANY STATUTE OR ON ANY OTHER BASIS FOR SPECIAL, INCIDENTAL, INDIRECT, PUNITIVE, MULTIPLE OR CONSEQUENTIAL DAMAGES IN CONNECTION WITH OR ARISING FROM THIS DOCUMENT, INCLUDING BUT NOT LIMITED TO THE USE THEREOF.

© 2012 Life Technologies Corporation. All rights reserved. The trademarks mentioned herein are the property of Life Technologies Corporation or their respective owners. RNase *AWAY* is a registered trademark of Molecular Bio-Products, Inc. TaqMan is a registered trademark of Roche Molecular Systems, Inc. BioRobot is a registered trademark of QIAGEN Corp. Freedom EVO is a trademark of Tecan. Biomek is a registered trademark of Beckman Coulter, Inc.

## Notes

# Preparing Samples for Sequencing Genomic DNA

FOR RESEARCH ONLY

## Topics

- 3 Introduction
- 5 Kit Contents and Equipment Checklist
- 7 Fragment the Genomic DNA
- 11 Perform End Repair
- 12 Add 'A' Bases to the 3' End of the DNA Fragments
- 13 Ligate Adapters to DNA Fragments
- 14 Purify Ligation Products
- 16 Enrich the Adapter-Modified DNA Fragments by PCR
- 17 Validate the Library



This publication and its contents are proprietary to Illumina, Inc., and are intended solely for the contractual use of its customers and for no other purpose than to operate the system described herein. This publication and its contents shall not be used or distributed for any other purpose and/or otherwise communicated, disclosed, or reproduced in any way whatsoever without the prior written consent of Illumina, Inc.

For the proper operation of this system and/or all parts thereof, the instructions in this guide must be strictly and explicitly followed by experienced personnel. All of the contents of this guide must be fully read and understood prior to operating the system or any of the parts thereof.

FAILURE TO COMPLETELY READ AND FULLY UNDERSTAND AND FOLLOW ALL OF THE CONTENTS OF THIS GUIDE PRIOR TO OPERATING THIS SYSTEM, OR PARTS THEREOF, MAY RESULT IN DAMAGE TO THE EQUIPMENT, OR PARTS THEREOF, AND INJURY TO ANY PERSONS OPERATING THE SAME.

Illumina, Inc. does not assume any liability arising out of the application or use of any products, component parts, or software described herein. Illumina, Inc. further does not convey any license under its patent, trademark, copyright, or common-law rights nor the similar rights of others. Illumina, Inc. further reserves the right to make any changes in any processes, products, or parts thereof, described herein without notice. While every effort has been made to make this guide as complete and accurate as possible as of the publication date, no warranty or fitness is implied, nor does Illumina accept any liability for damages resulting from the information contained in this guide.

© 2008 Illumina, Inc. All rights reserved. **Illumina, Solexa, Making Sense Out of Life, Oligator, Sentrix, GoldenGate, DASL, BeadArray, Array of Arrays, Infinium, BeadXpress, VeraCode, IntelliHyb, iSelect, and CPro** are registered trademarks or trademarks of Illumina. All other brands and names contained herein are the property of their respective owners.

## Workflow

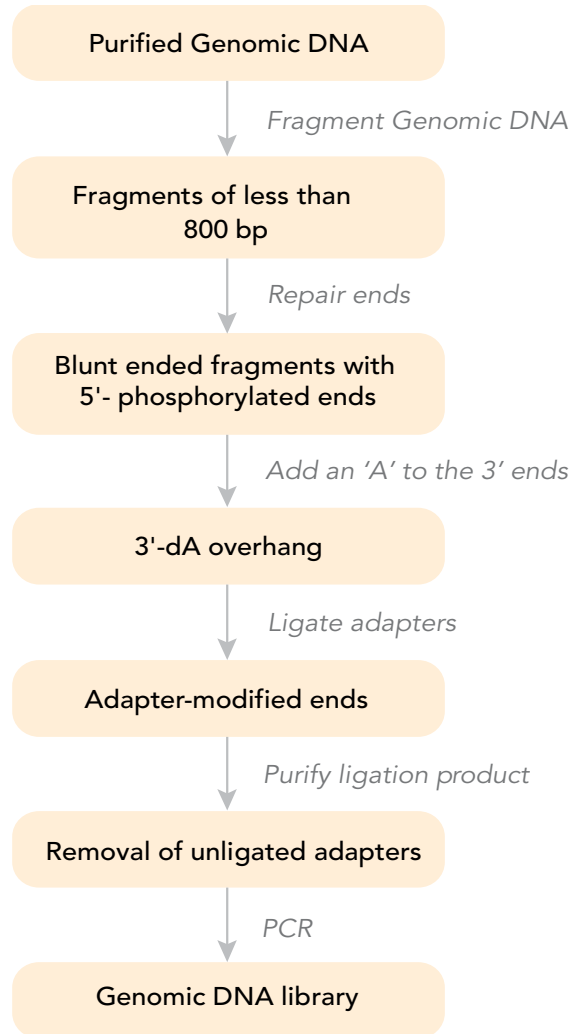


Figure 2 Sample Preparation Workflow

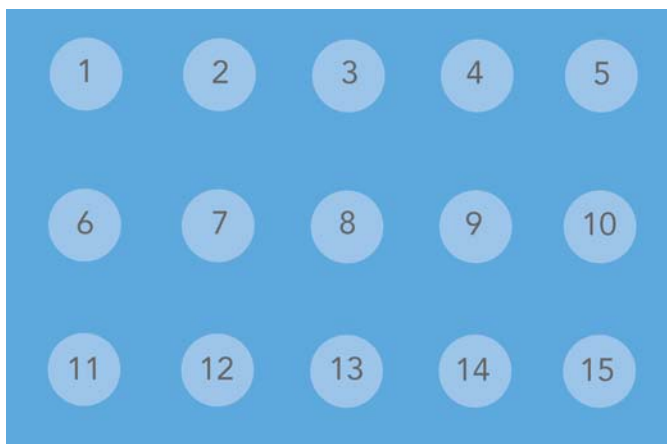
## Kit Contents and Equipment Checklist

Check to ensure that you have all of the reagents identified in this section before proceeding to sample preparation.

### Genomic DNA Sample Prep Kit, Box 1

Store at -20°C

This box is shipped at -80°C. As soon as you receive it, store the following components at -20°C.

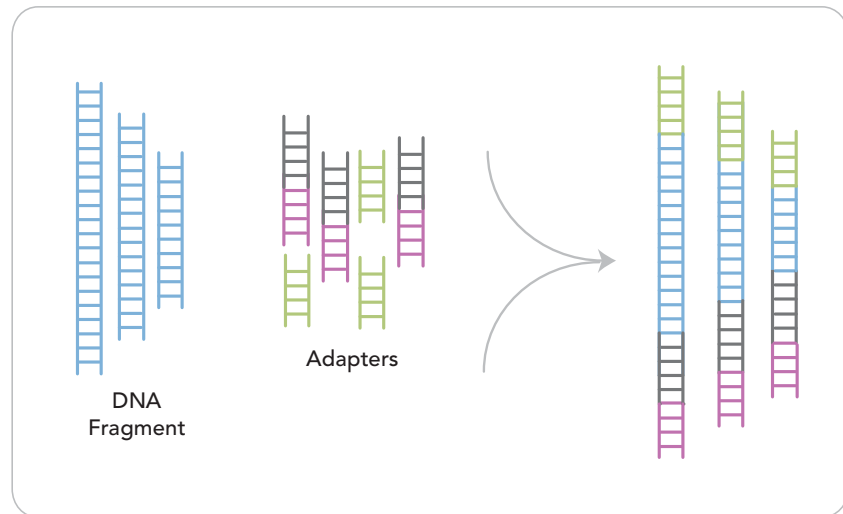


**Figure 3** Genomic DNA Sample Prep Kit, Box 1

1. T4 DNA Ligase Buffer with 10 mM ATP, part # 1000534
2. Klenow DNA Polymerase, part # 1000515
3. Klenow Buffer, part # 1000535
4. 2X DNA Ligase Buffer, part # 1000523
5. Phusion DNA Polymerase (Finnzymes Oy), part # 1000524
6. 10 mM dNTPs Mix, part # 1001932
7. T4 PNK, part # 1000519
8. 1 mM dATP, part # 1000520
9. Adapter Oligo Mix, part # 1000521
10. PCR Primer 1.1, part # 1000537
11. T4 DNA Polymerase, part # 1000514
12. Empty
13. Klenow Fragment (3' to 5' exo minus), part # 1000536
14. DNA Ligase, part # 1000522
15. PCR Primer 2.1, part # 1000538

# Introduction

This protocol explains how to prepare libraries of genomic DNA for analysis on the Illumina Cluster Station and Genome Analyzer. You will add adapter sequences onto the ends of DNA fragments to generate the following template format:



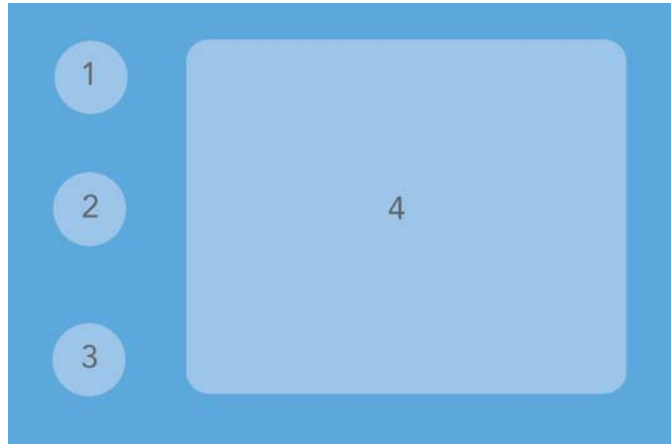
*Figure 1* Fragments after Sample Preparation

The adapters contain sequences that correspond to the two surface-bound amplification primers on the flow cells used in the Cluster Station.

## Genomic DNA Sample Prep Kit, Box 2

### Store at Room Temperature

This box is shipped at room temperature. Store the following components at room temperature.



*Figure 4* Genomic DNA Sample Prep Kit, Box 2

1. Nebulization Buffer, part # 1000466
2. TE Buffer, part # 1000465
3. Ultra Pure Water, part # 1000467
4. Nebulizer Kit, part # 1000541

## Equipment Checklist

Check to ensure that you have all of the necessary user-supplied equipment before proceeding to sample preparation.

- ▶ Benchtop microcentrifuge
- ▶ Benchtop centrifuge with swing-out rotor
- ▶ Dark Reader transilluminator or UV transilluminator
- ▶ Disposable scalpels
- ▶ Electrophoresis unit
- ▶ Gel trays and tank
- ▶ Thermal cycler or heat block

## Fragment the Genomic DNA

This protocol fragments the genomic DNA using a nebulization technique, which fragments DNA to less than 800 bp in minutes using a disposable device. Nebulization generates double-stranded DNA fragments comprised of 3' or 5' overhangs.

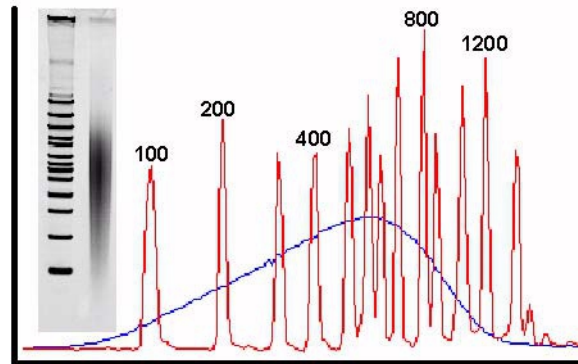


Figure 5 Fragment Genomic DNA

### Consumables Illumina-Supplied

- ▶ Nebulizers (box of 10 nebulizers and vinyl accessory tubes)
- ▶ Nebulization buffer (7 ml)
- ▶ TE Buffer

### User-Supplied

- ▶ QIAquick PCR Purification Kit (QIAGEN, #28104)
- ▶ Purified DNA (1–5  $\mu$ g, 5  $\mu$ g recommended)  
DNA should be as intact as possible, with an OD<sub>260/280</sub> ratio of 1.8–2
- ▶ Compressed air of at least 32 psi
- ▶ Clamp (1 per nebulizer)
- ▶ PVC tubing
  - Fisher Scientific, catalog # 14-176-102
  - Nalgene Labware, catalog # 8007-0060

Table 1 PVC Tubing Dimensions

ID	OD	Wall	Length
1/4 in.	3/8 in.	1/16 in.	1 meter



**CAUTION**

If you intend to nebulize DNA that could possibly contain any pathogenic sequences such as pathogenic viral DNA, perform the nebulization process under containment conditions (e.g., a biosafety cabinet) to prevent exposure to aerosols.

## Procedure

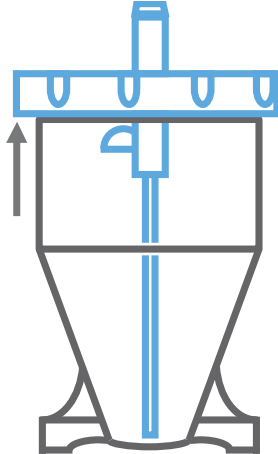
The DNA sample to be processed should be highly pure, having an OD<sub>260</sub>/280 ratio of between 1.8 and 2, and should be as intact as possible.



### NOTE

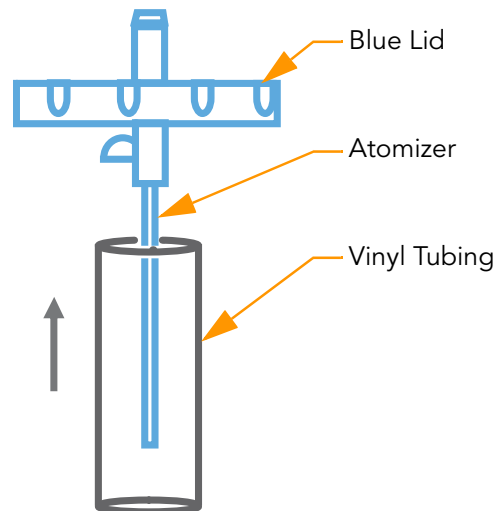
If you are not familiar with this shearing method, Illumina recommends that you test this procedure on test samples before proceeding with your sample DNA.

1. Remove a nebulizer from its plastic packaging and unscrew the blue lid.



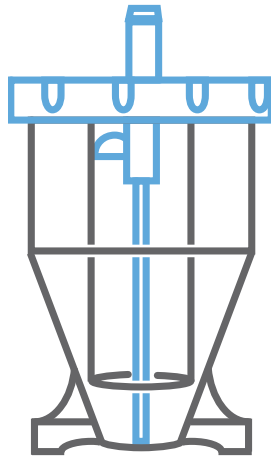
*Figure 6* Remove the Nebulizer Lid

2. Using gloves, remove a piece of vinyl tubing from its packaging and slip it over the central atomizer tube. Push it all the way to the inner surface of the blue lid.



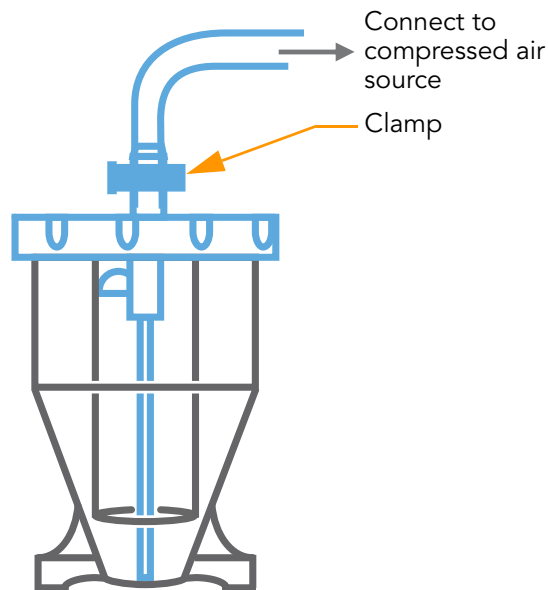
*Figure 7* Assemble the Nebulizer

3. Add 1–5  $\mu\text{g}$  of purified DNA in a total volume of 50  $\mu\text{l}$  of TE buffer to the nebulizer.
4. Add 700  $\mu\text{l}$  nebulization buffer to the DNA and mix well.
5. Screw the lid back on (finger-tight).



**Figure 8** Replace the Nebulizer Lid

6. Chill the nebulizer containing the DNA solution on ice.
7. Connect the compressed air source to the inlet port on the top of the nebulizer with the PVC tubing, ensuring a tight fit. Secure with the small clamp.



**Figure 9** Connect Compressed Air

8. Bury the nebulizer in an ice bucket and place it in a fume hood.
9. Use the regulator on the compressed air source to ensure the air is delivered at 32–35 psi.
10. Nebulize for 6 minutes. You may notice vapor rising from the nebulizer; this is normal.
11. Centrifuge the nebulizer at 450 xg for 2 minutes to collect the droplets from the side of the nebulizer. If necessary, use an old nebulizer as a counter-balance.



12. If a centrifuge is not available, then use 2 ml of the binding buffer (PB or PBI buffer) from the QIAquick PCR Purification Kit to rinse the sides of the nebulizer and collect the DNA solution at the base of the nebulizer.
13. Measure the recovered volume. Typically, you should recover 400–600  $\mu$ l.
14. Follow the instructions in the QIAquick PCR Purification Kit to purify the sample solution and concentrate it on one QIAquick column, eluting in 30  $\mu$ l of EB.

## Perform End Repair

This protocol converts the overhangs resulting from fragmentation into blunt ends, using T4 DNA polymerase and *E. coli* DNA polymerase I Klenow fragment. The 3' to 5' exonuclease activity of these enzymes removes 3' overhangs and the polymerase activity fills in the 5' overhangs.

### Consumables **Illumina-Supplied**

- ▶ T4 DNA ligase buffer with 10mM ATP
- ▶ dNTPs mix
- ▶ T4 DNA polymerase
- ▶ Klenow DNA polymerase
- ▶ T4 PNK
- ▶ Water

### User-Supplied

- ▶ QIAquick PCR Purification Kit (QIAGEN, part # 28104)

### Procedure

1. Prepare the following reaction mix:
  - DNA sample (30  $\mu$ l)
  - Water (45  $\mu$ l)
  - T4 DNA ligase buffer with 10mM ATP (10  $\mu$ l)
  - dNTPs mix (4  $\mu$ l)
  - T4 DNA polymerase (5  $\mu$ l)
  - Klenow DNA polymerase (1  $\mu$ l)
  - T4 PNK (5  $\mu$ l)The total volume should be 100  $\mu$ l.
2. Incubate in the thermal cycler for 30 minutes at 20°C.
3. Follow the instructions in the QIAquick PCR Purification Kit to purify on one QIAquick column, eluting in 32  $\mu$ l of EB.

## Add 'A' Bases to the 3' End of the DNA Fragments

This protocol adds an 'A' base to the 3' end of the blunt phosphorylated DNA fragments, using the polymerase activity of Klenow fragment (3' to 5' exo minus). This prepares the DNA fragments for ligation to the adapters, which have a single 'T' base overhang at their 3' end.

### Consumables **Illumina-Supplied**

- ▶ Klenow buffer
- ▶ dATP
- ▶ Klenow exo (3' to 5' exo minus)

### User-Supplied

- ▶ MinElute PCR Purification Kit (QIAGEN, part # 28004)



This protocol requires a QIAquick MinElute column rather than a normal QIAquick column.

### Procedure

1. Prepare the following reaction mix:
  - DNA sample (32  $\mu$ l)
  - Klenow buffer (5  $\mu$ l)
  - dATP (10  $\mu$ l)
  - Klenow exo (3' to 5' exo minus) (3  $\mu$ l)The total volume should be 50  $\mu$ l.
2. Incubate for 30 minutes at 37°C.
3. Follow the instructions in the MinElute PCR Purification Kit to purify on one QIAquick MinElute column, eluting in 10  $\mu$ l of EB.

# Ligate Adapters to DNA Fragments

This protocol ligates adapters to the ends of the DNA fragments, preparing them to be hybridized to a flow cell.

## Consumables

### Illumina-Supplied

- ▶ DNA ligase buffer
- ▶ Illumina adapter oligo mix
- ▶ DNA ligase

### User-Supplied

- ▶ QIAquick PCR Purification Kit (QIAGEN, part # 28104)

## Procedure

This procedure uses a 10:1 molar ratio of adapter to genomic DNA insert, based on a starting quantity of 5 µg of DNA before fragmentation. If you started with less than 5 µg, titrate the volume of adapter reagent accordingly to maintain the 10:1 ratio of DNA.

1. Prepare the following reaction mix:

- DNA sample (10 µl)
- DNA ligase buffer (25 µl)
- Adapter oligo mix (10 µl)
- DNA ligase (5 µl)

The total volume should be 50 µl.

2. Incubate in a thermal cycler for 15 minutes at 20°C.

3. Follow the instructions in the QIAquick PCR Purification Kit to purify on one QIAquick column, eluting in 30 µl of EB.

## Purify Ligation Products

This protocol purifies the products of the ligation reaction on a gel to remove all unligated adapters, remove any adapters that may have ligated to one another, and select a size-range of templates to go on the cluster generation platform.

### Consumables User-Supplied

- ▶ Certified low-range Ultra Agarose (BIO-RAD, part # 161-3106)
- ▶ 50x TAE buffer
- ▶ Distilled water
- ▶ Ethidium bromide
- ▶ Loading buffer (50 mM Tris pH 8.0, 40 mM EDTA, 40% (w/v) sucrose)
- ▶ Low molecular weight DNA ladder (NEB, part # N3233L)
- ▶ QIAquick or MinElute Gel Extraction Kit (QIAGEN, part # 28704 or part # 28604)

### Procedure

**CAUTION**

Illumina does not recommend purifying multiple samples on a single gel due to the risk of cross-contamination between libraries.

**CAUTION**

It is important to perform this procedure exactly as described, to ensure reproducibility.

1. Prepare a 50 ml, 2% agarose gel with distilled water and TAE. Final concentration of TAE should be 1X at 50 ml.
2. Add ethidium bromide (EtBr) after the TAE-agarose has cooled. Final concentration of EtBr should be 400 ng/ml (i.e., add 20 µg EtBr to 50 ml of 1X TAE).
3. Add 3 µl of loading buffer to 8 µl of the ladder.
4. Add 10 µl of loading buffer to 30 µl of the DNA from the purified ligation reaction.
5. Load all of the ladder solution to one lane of the gel.
6. Load the entire sample in another lane of the gel, leaving a gap of at least one empty lane between ladder and sample.
7. Run the gel at 120 V for 60 minutes.
8. View the gel on a Dark Reader transilluminator or a UV transilluminator.

9. Excise a region of gel with a clean scalpel. The gel slice should contain the material in the 150–200 bp range.

**NOTE**

You can select more than one size-range of adapter-ligated DNA by excising slices from different parts of the gel. A relatively short insert template is 150–200 bp, while 300–650 bp is a long insert template.

10. Using a Gel Extraction Kit, do one of the following:
  - If the gel slice is less than 400 mg, use one column from a QIAquick Gel Extraction Kit and elute in 30  $\mu$ l EB.
  - If the gel slice is more than 400 mg, use two MinElute columns, elute each one in 15  $\mu$ l EB, and pool.

## Enrich the Adapter-Modified DNA Fragments by PCR

This protocol uses PCR to selectively enrich those DNA fragments that have adapter molecules on both ends, and to amplify the amount of DNA in the library. The PCR is performed with two primers that anneal to the ends of the adapters. The number of PCR cycles is minimized to avoid skewing the representation of the library.

### Consumables **Illumina-Supplied**

- ▶ Phusion DNA polymerase (Finnzymes Oy)
- ▶ PCR primer 1.1
- ▶ PCR primer 2.1
- ▶ Ultra pure water

### User-Supplied

- ▶ QIAquick PCR Purification Kit (QIAGEN, part # 28104)

### Procedure

This protocol assumes 5 µg of DNA input into library prep. If you use 0.5 µg, adjust the protocol as described in the following table.

Input of DNA to Library Prep	Volume of Purified Library into PCR	Volume of Water	Number of PCR Cycles
5 µg	1 µl	22 µl	10
0.5 µg	2 µl	21 µl	12

1. Prepare the following PCR reaction mix:
  - DNA (1 µl)
  - Phusion DNA polymerase (Finnzymes Oy) (25 µl)
  - PCR primer 1.1 (1 µl)
  - PCR primer 2.1 (1 µl)
  - Water (22 µl)

The total volume should be 50 µl.

2. Amplify using the following PCR protocol:
  - a. 30 seconds at 98°C
  - b. 10 cycles of:
    - 10 seconds at 98°C
    - 30 seconds at 65°C
    - 30 seconds at 72°C
  - c. 5 minutes at 72°C
  - d. Hold at 4°C
3. Follow the instructions in the QIAquick PCR Purification Kit to purify on one QIAquick column, eluting in 30 µl of EB.

## Validate the Library

Illumina recommends performing the following quality control steps on your DNA library.

1. Determine the concentration of the library by measuring its absorbance at 260 nm. The yield from the protocol should be between 500 and 1000 ng of DNA.
2. Measure the 260/280 ratio. It should be approximately 1.8.
3. Load 10% of the volume of the library on a gel and check that the size range is as expected. It should be similar in size to the size-range excised during the gel purification step.
4. To determine the molar concentration of the library, examine the gel image and estimate the median size of the library smear. This is generally about 450 bp for a long insert library and about 170 bp for a short insert library.
  - a. Multiply this size by 650 (the molecular mass of a base-pair) to get the molecular weight of the library.
  - b. Use this number to calculate the molar concentration of the library.
5. Clone 4% of the volume of the library into a sequencing vector.
  - a. Sequence individual clones by conventional Sanger sequencing.



### NOTE

The 5' ends of the library molecules are not phosphorylated and therefore require a phosphorylated vector for cloning.

- b. Verify that the insert sequences are from the genomic source DNA.

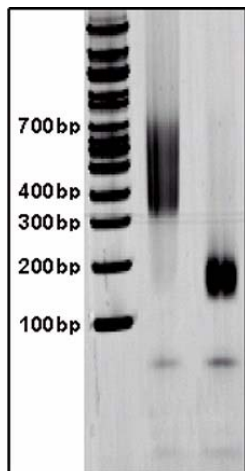


Figure 10 Sequencing Gel

This example shows a library run on a 4–20% TBE polyacrylamide gel, stained with Vistra Green (GE Healthcare # RPN5786) and visualized on a fluorescence scanner. The smear on the middle lane shows a long insert library, and the smear on the right shows a short insert library. The left



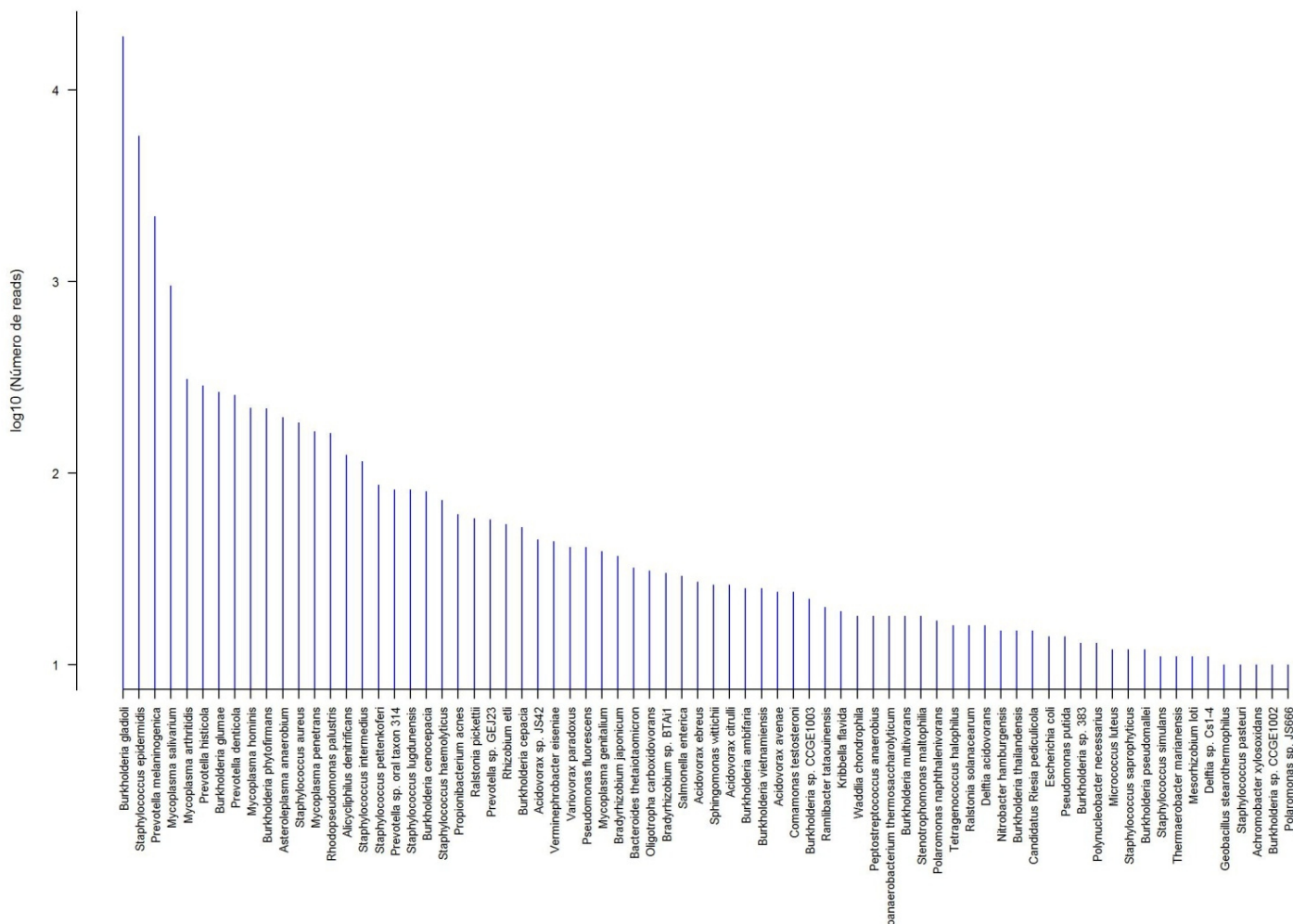
lane shows a marker ladder. The two bands less than 100 bp in size in the middle and right lanes are primers from the enrichment PCR step and have no effect on the subsequent formation of clusters.



Illumina, Inc.  
9885 Towne Centre Drive  
San Diego, CA 92121-1975  
+1.800.809.ILMN (4566)  
+1.858.202.4566 (outside North America)  
techsupport@illumina.com  
[www.illumina.com](http://www.illumina.com)

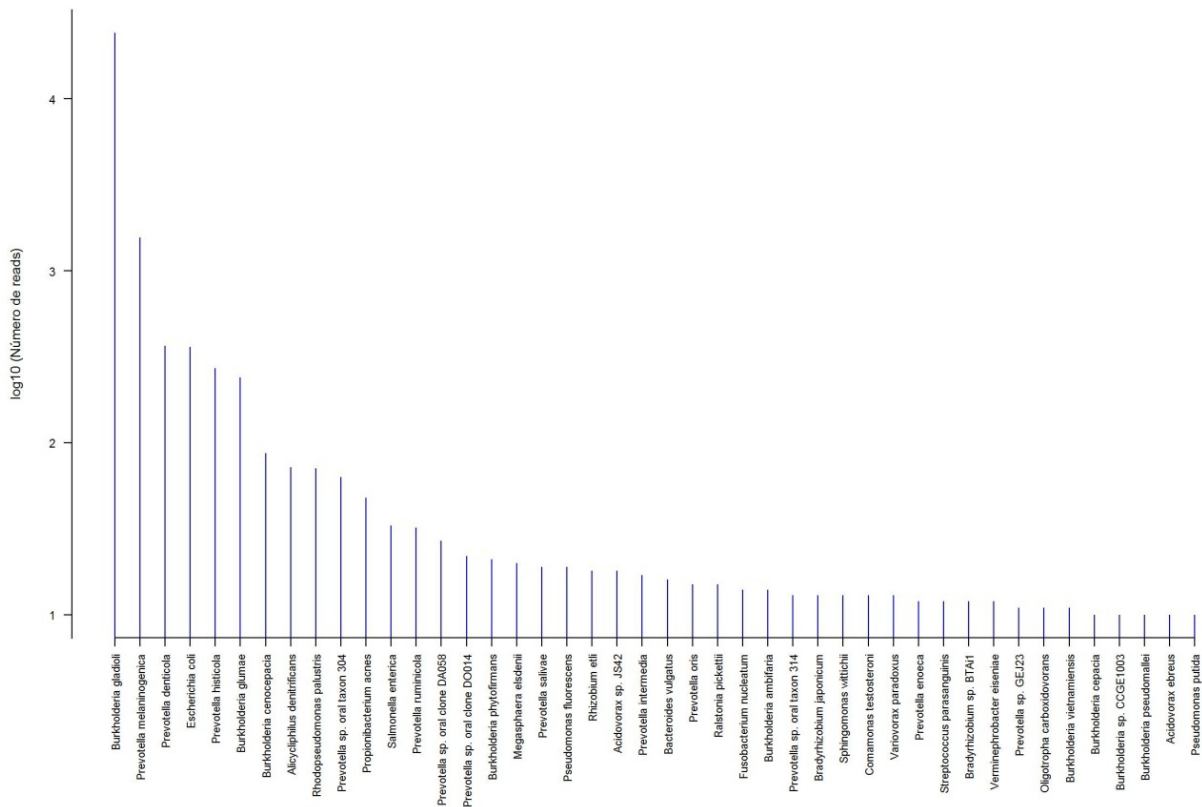


**Material suplementario S5.** Gráficas de abundancia de especies bacterianas con 10 o más reads.

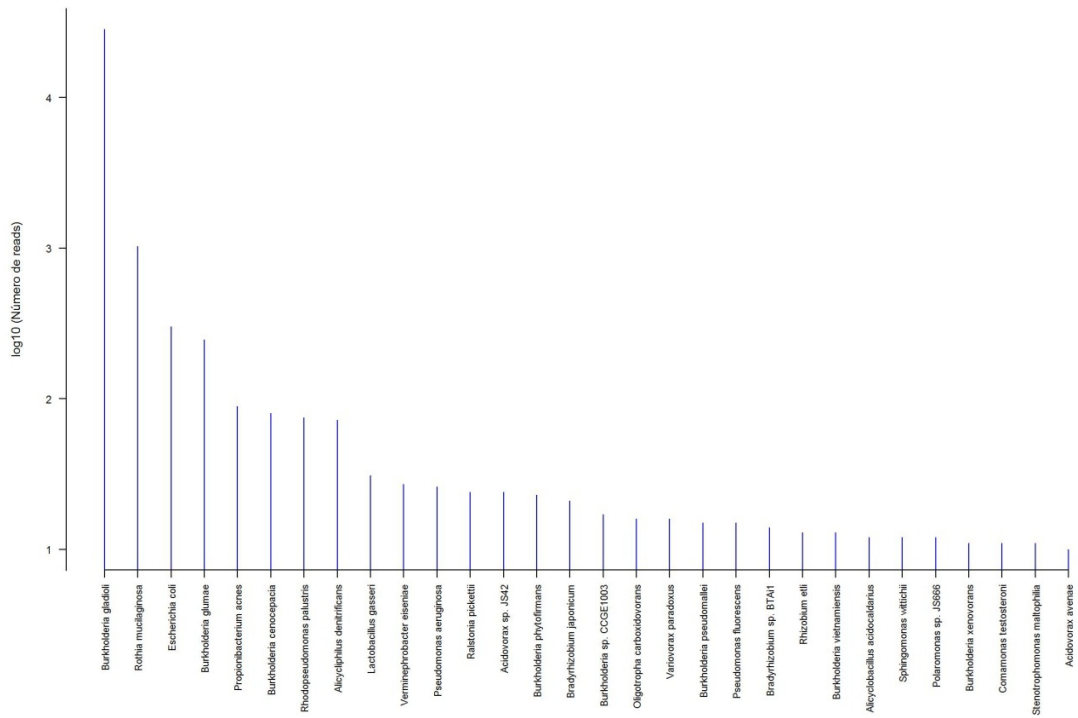


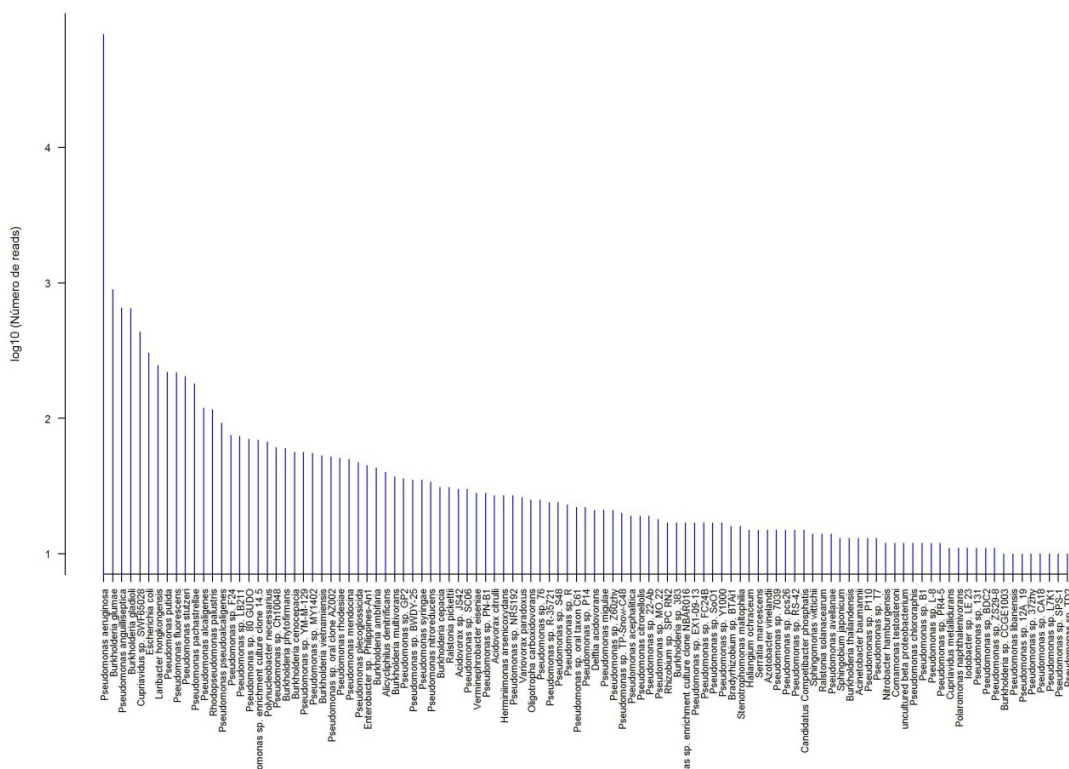
Muestra 6985

Muestra 7425

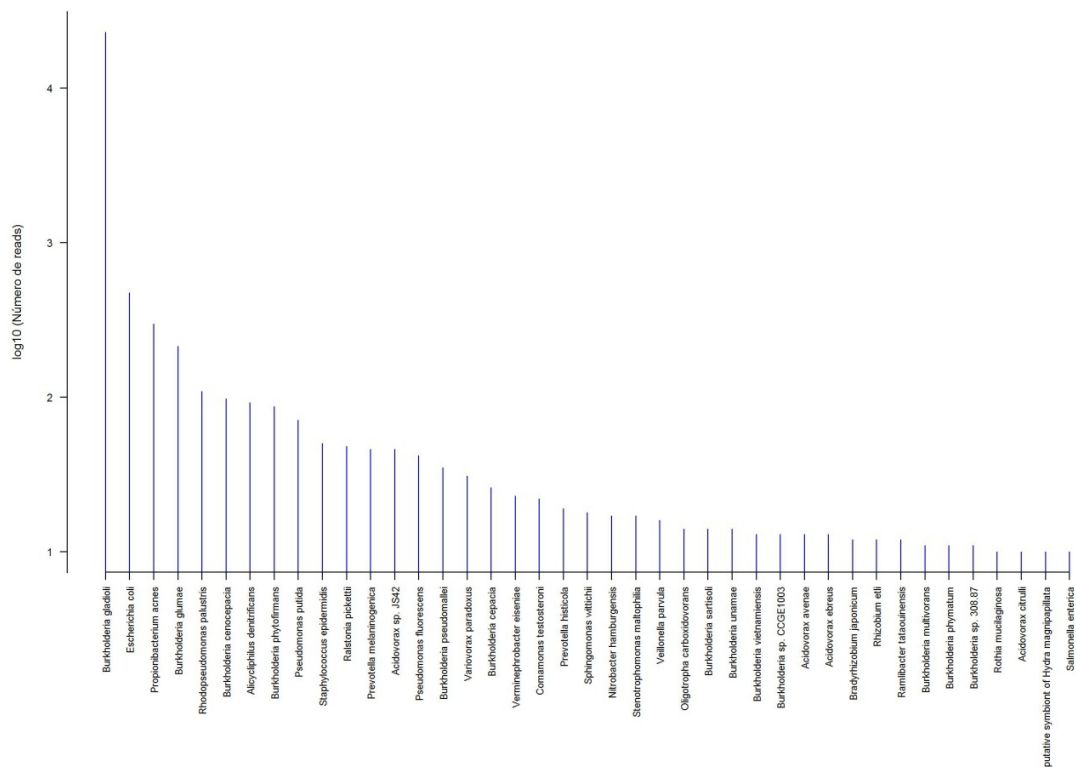


Muestra 8261



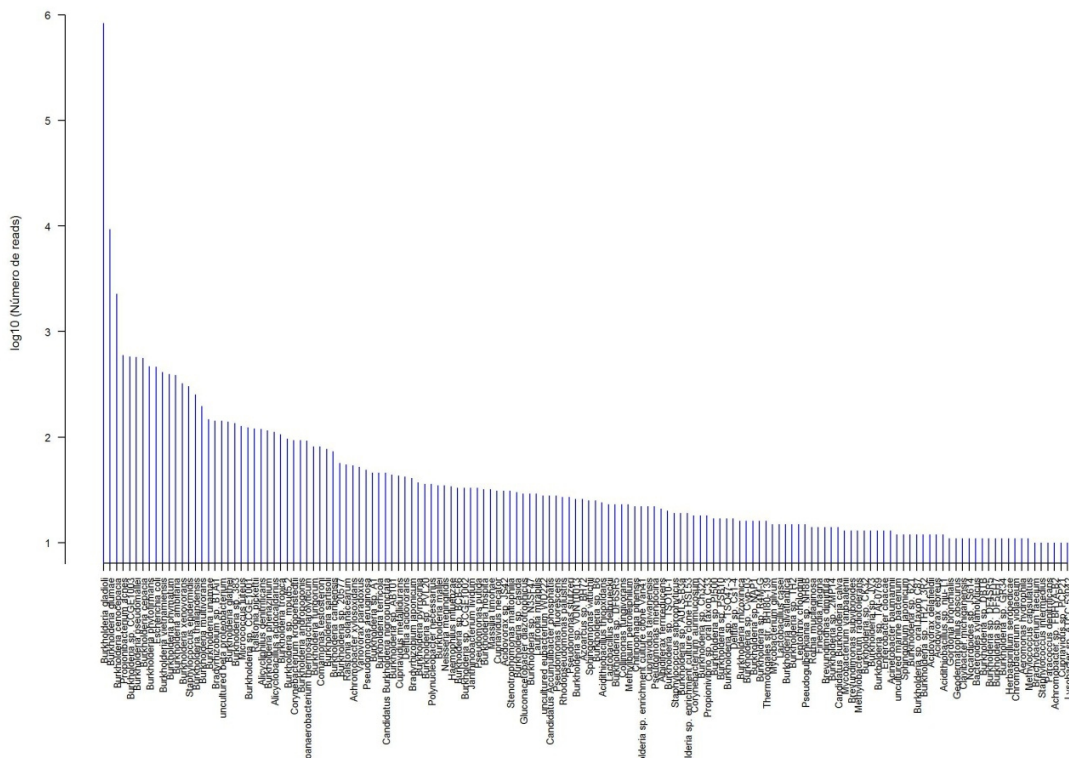


Muestra 8836

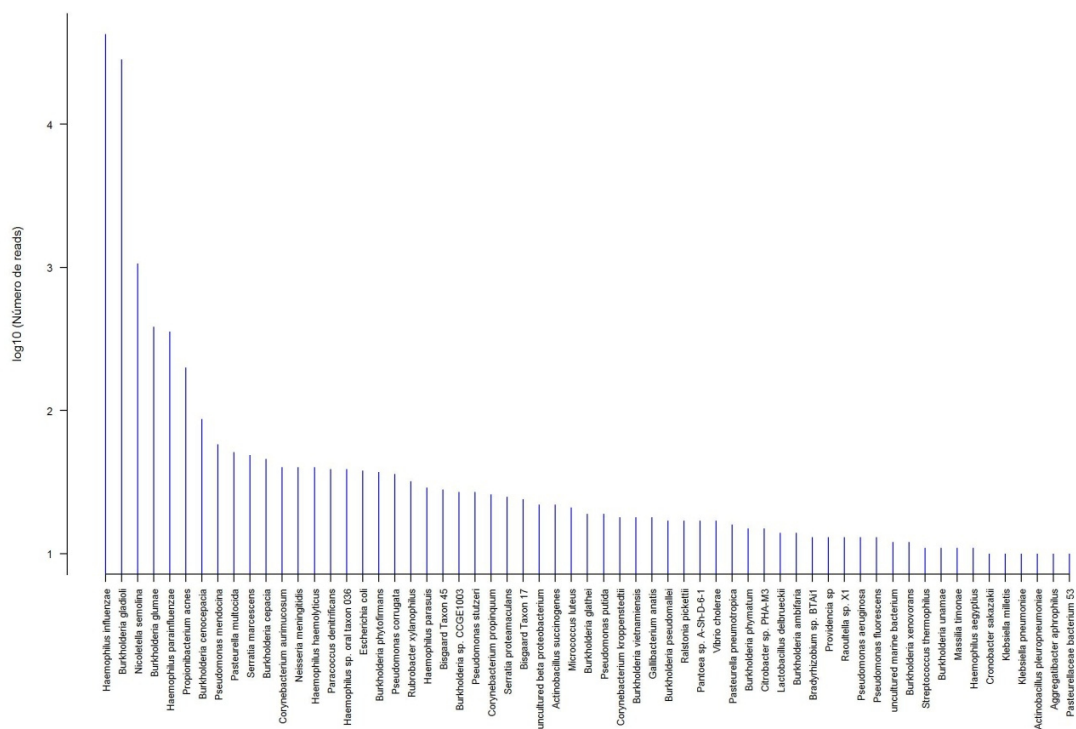


Muestra 9041

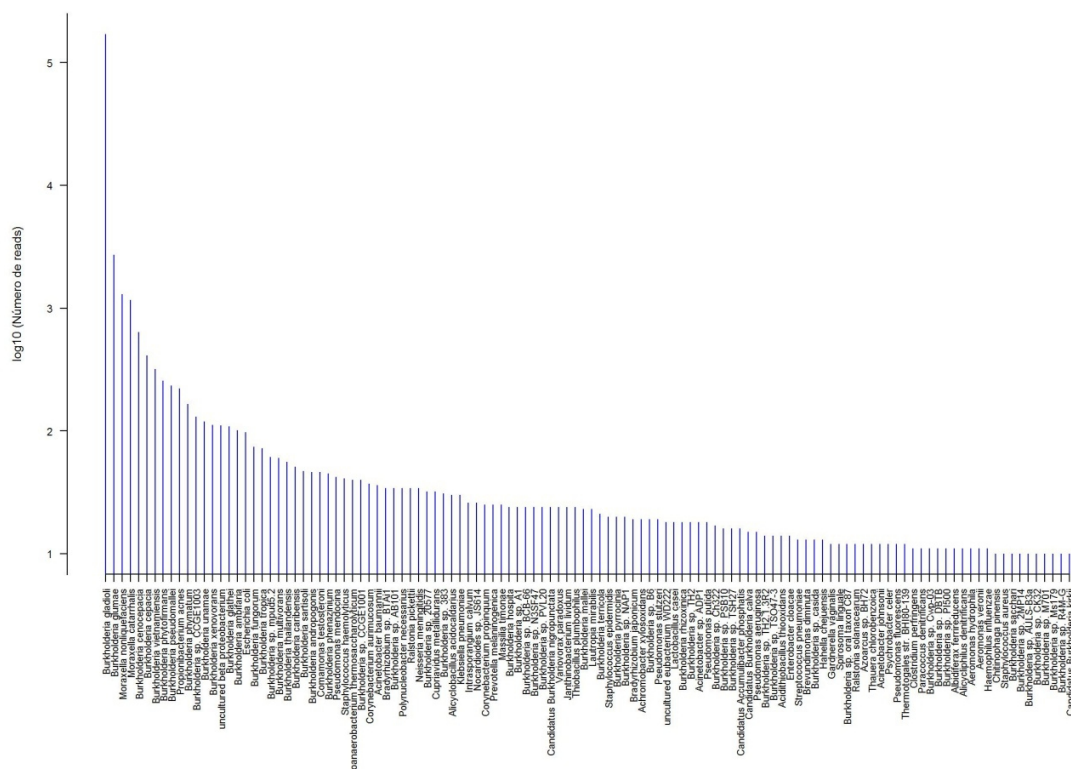




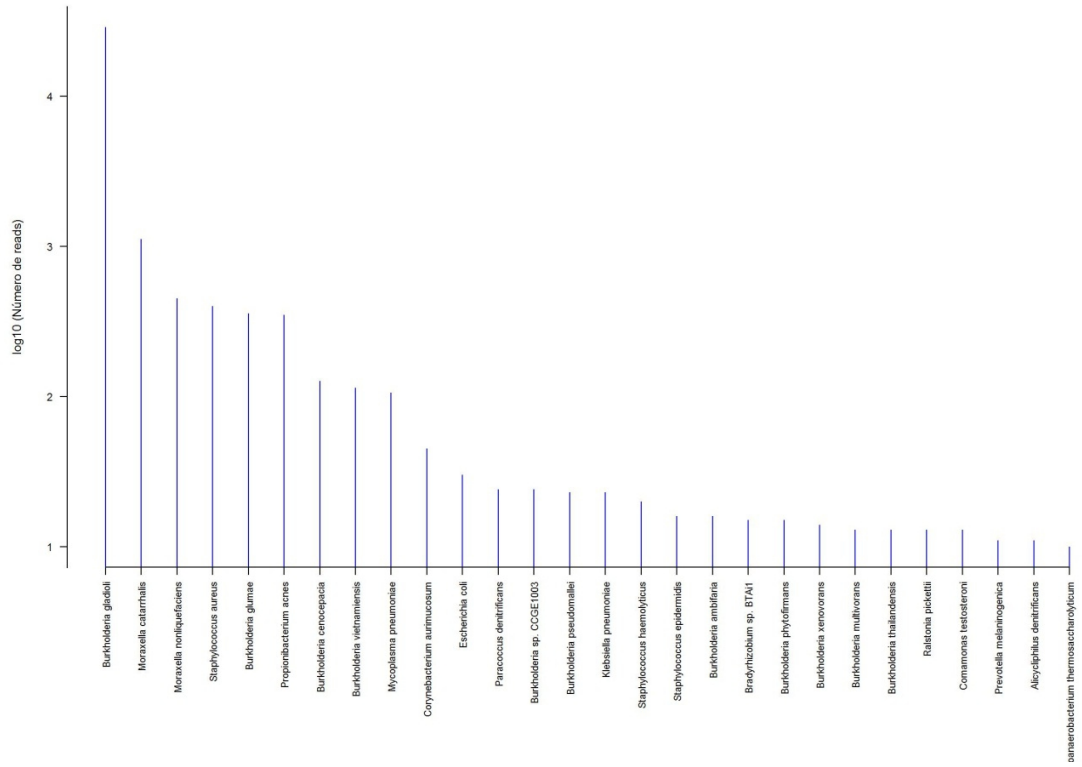
Muestra 8410



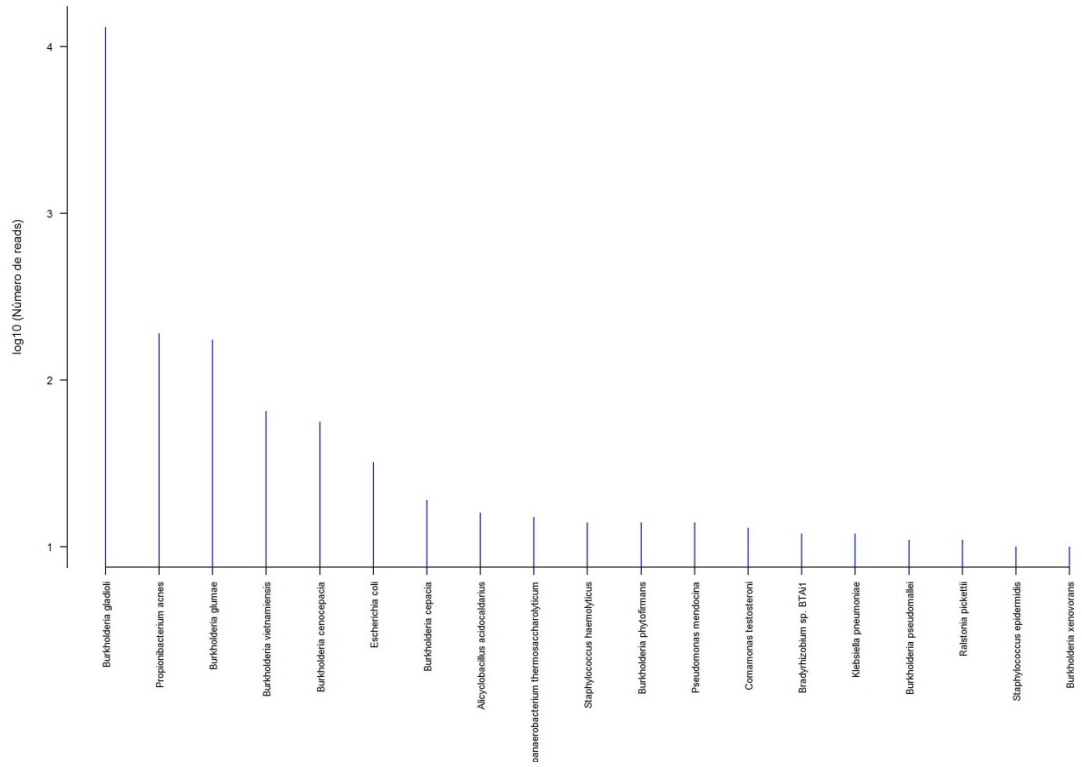
Muestra 8883



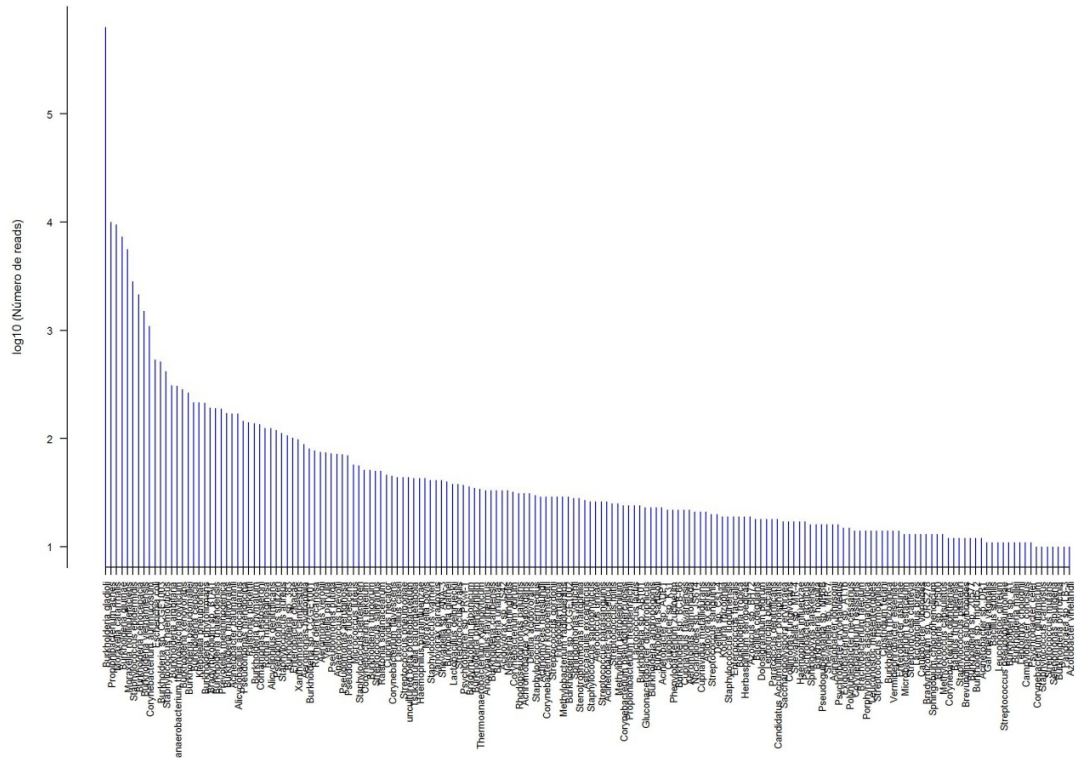
Muestra 9140



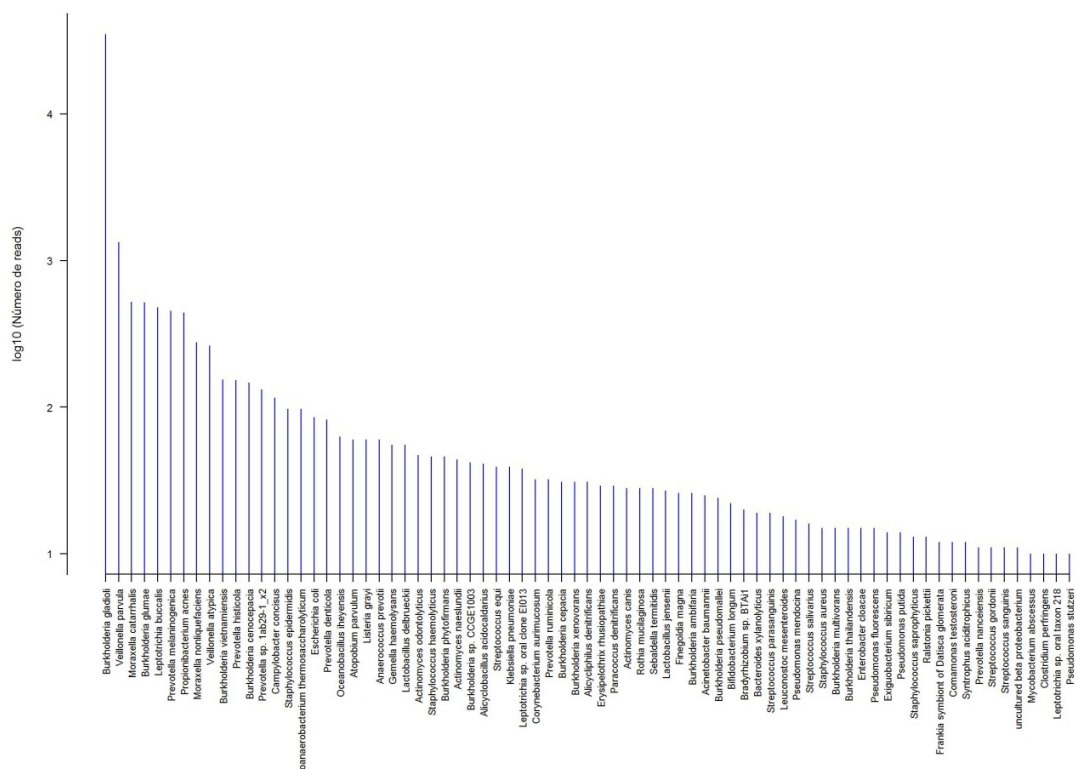
Muestra 9232



Muestra 9275



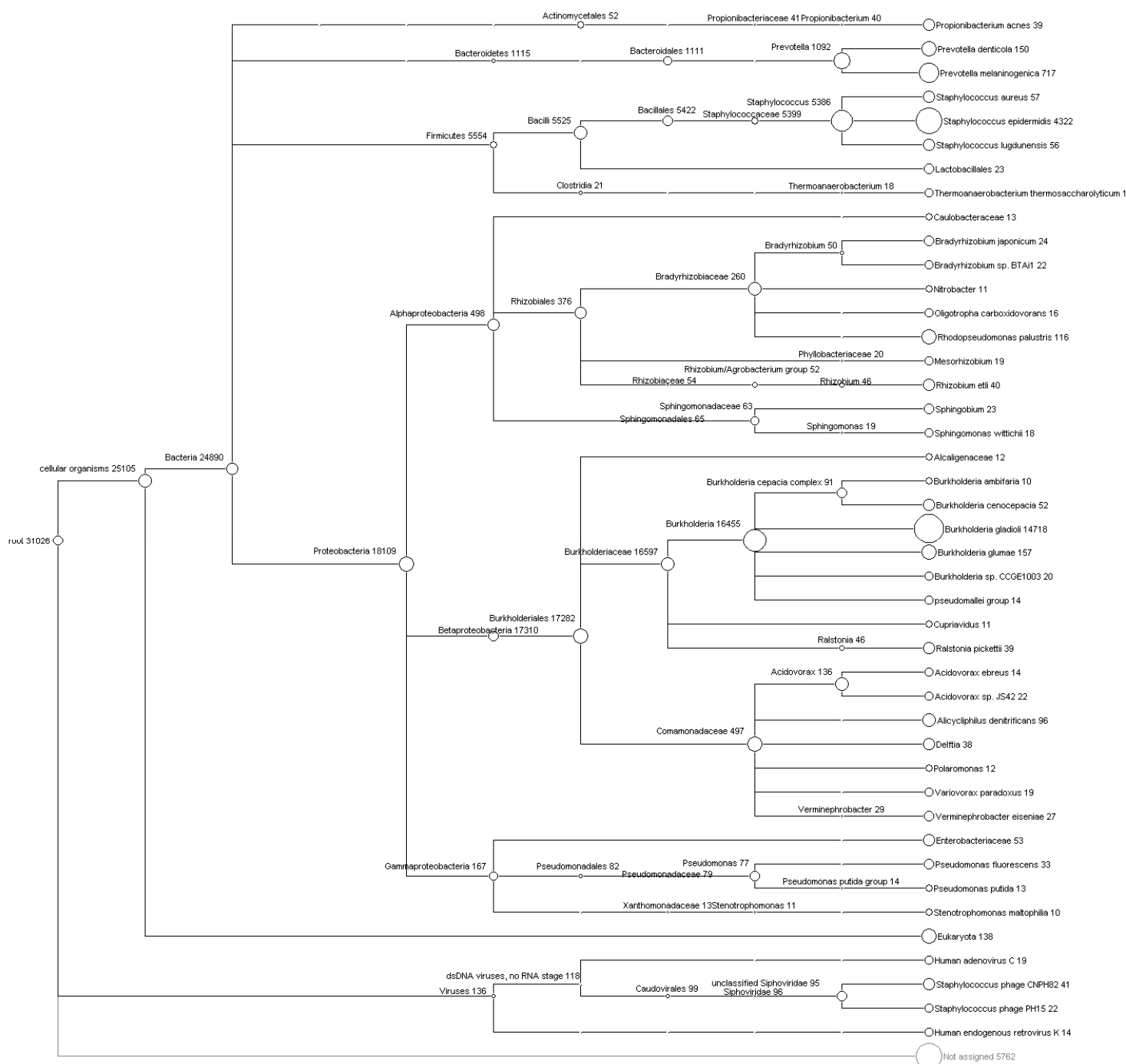
Muestra 9362



Muestra 9434

**Material suplementario S6:** Visualización en MEGAN de cada muestra, procesada mediante el protocolo bioinformático A, sólo se muestran los organismo identificados con más de 10 lecturas

**Muestra 6985**

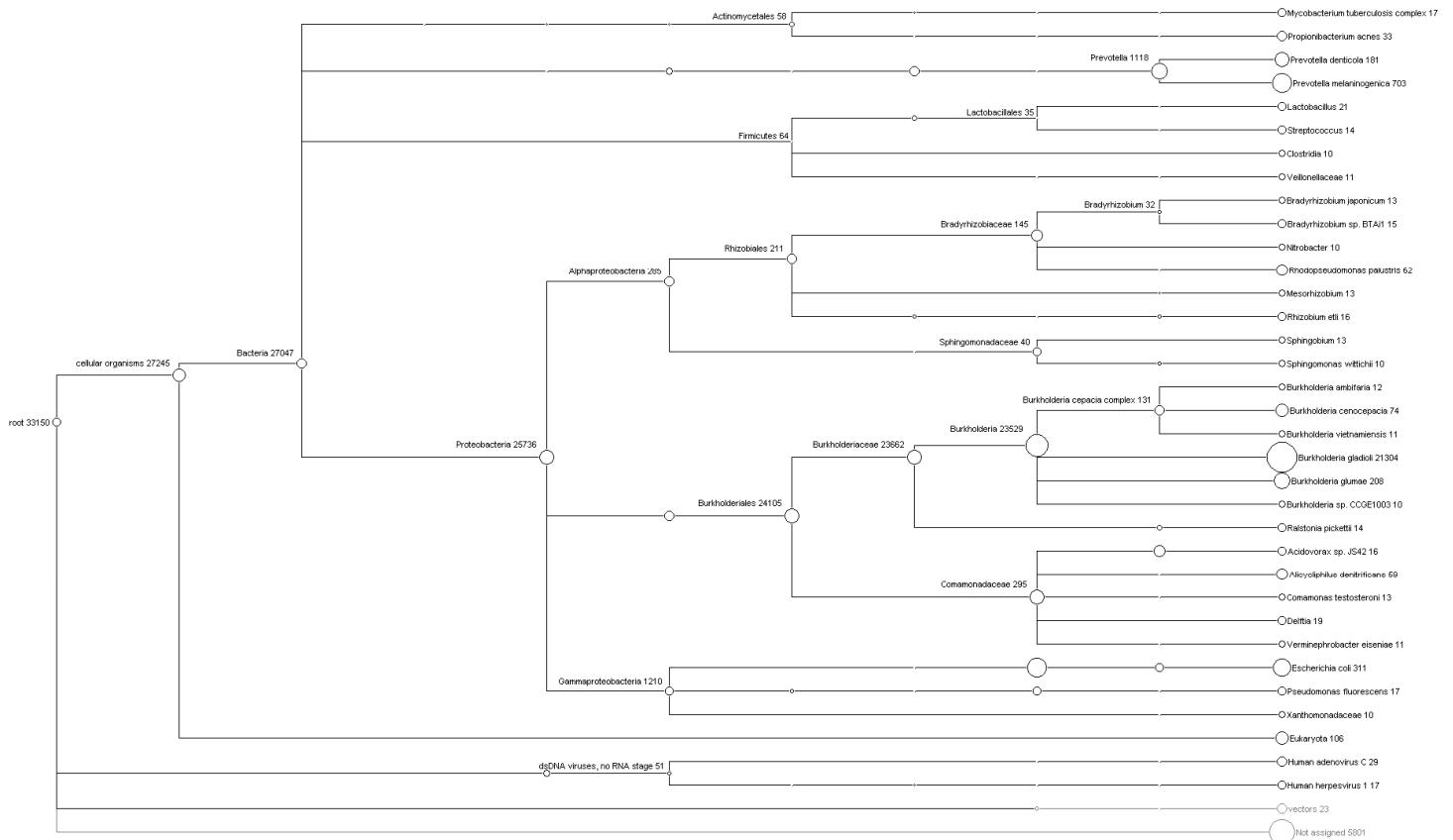


Metagenómica viral y bacteriana de enfermedades respiratorias graves  
L.C.G. Ana Georgina Cobián Güemes



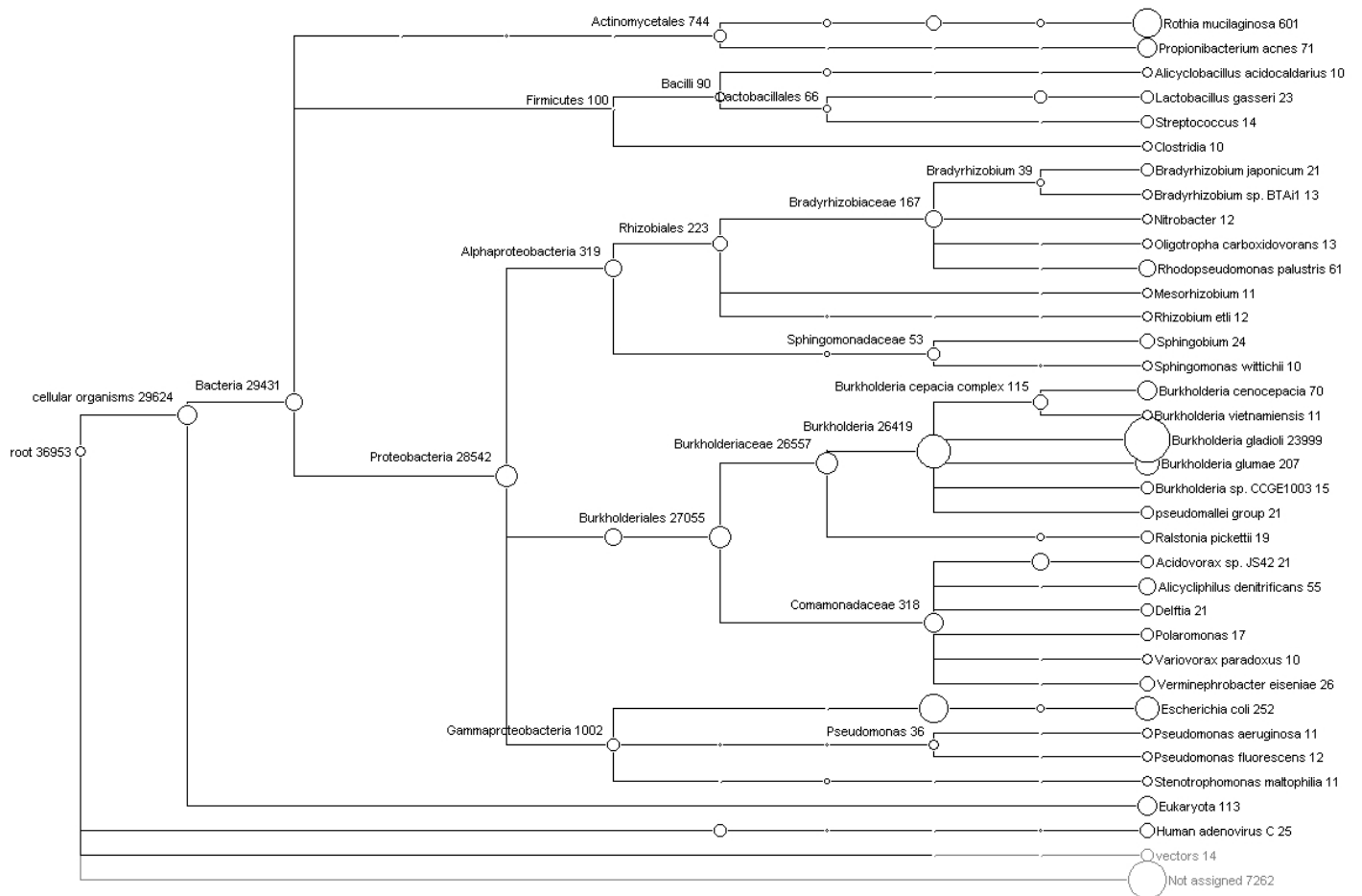


Muestra 7425



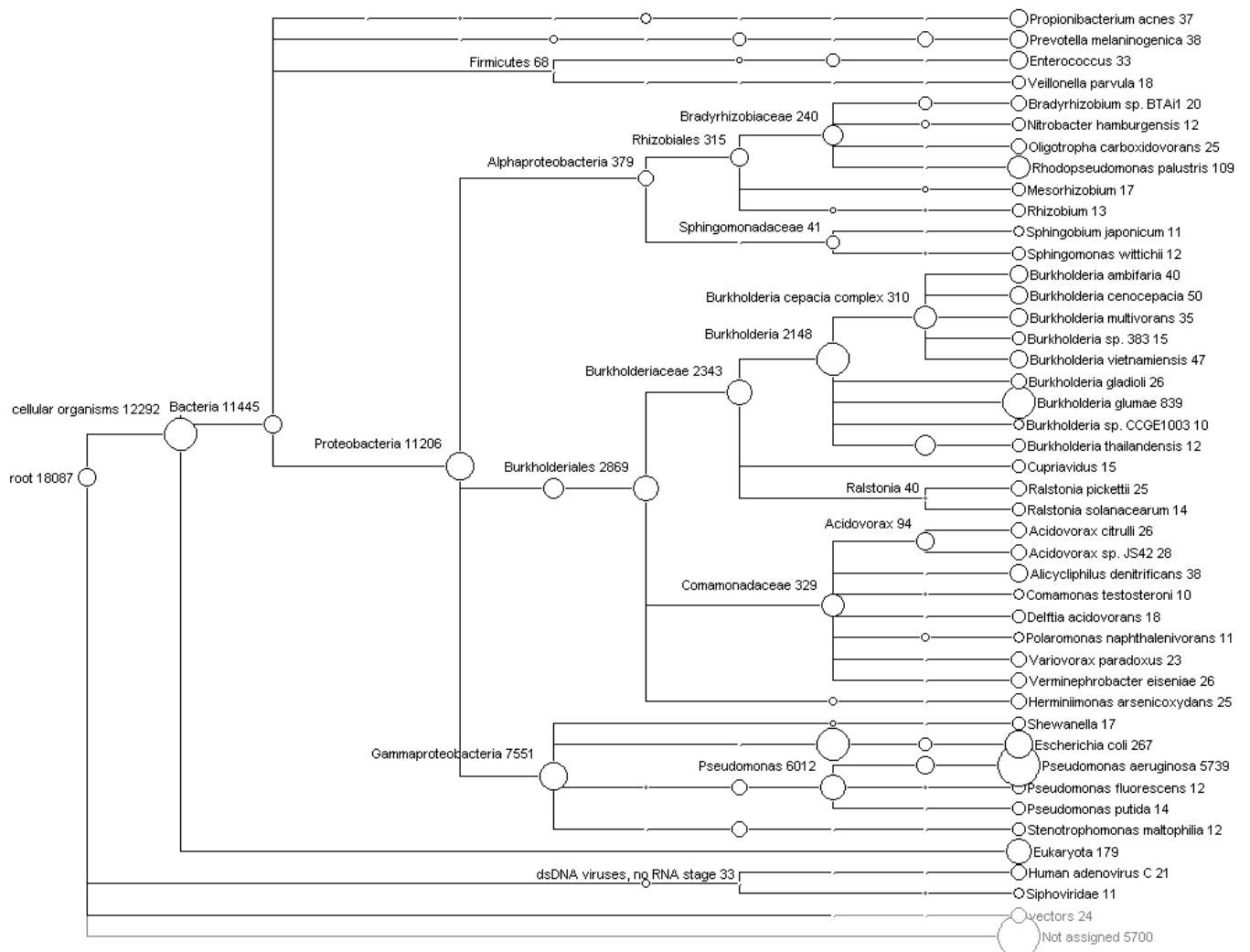
Metagenómica viral y bacteriana de enfermedades respiratorias graves  
 L.C.G. Ana Georgina Cobián Güemes

**Muestra 8261**



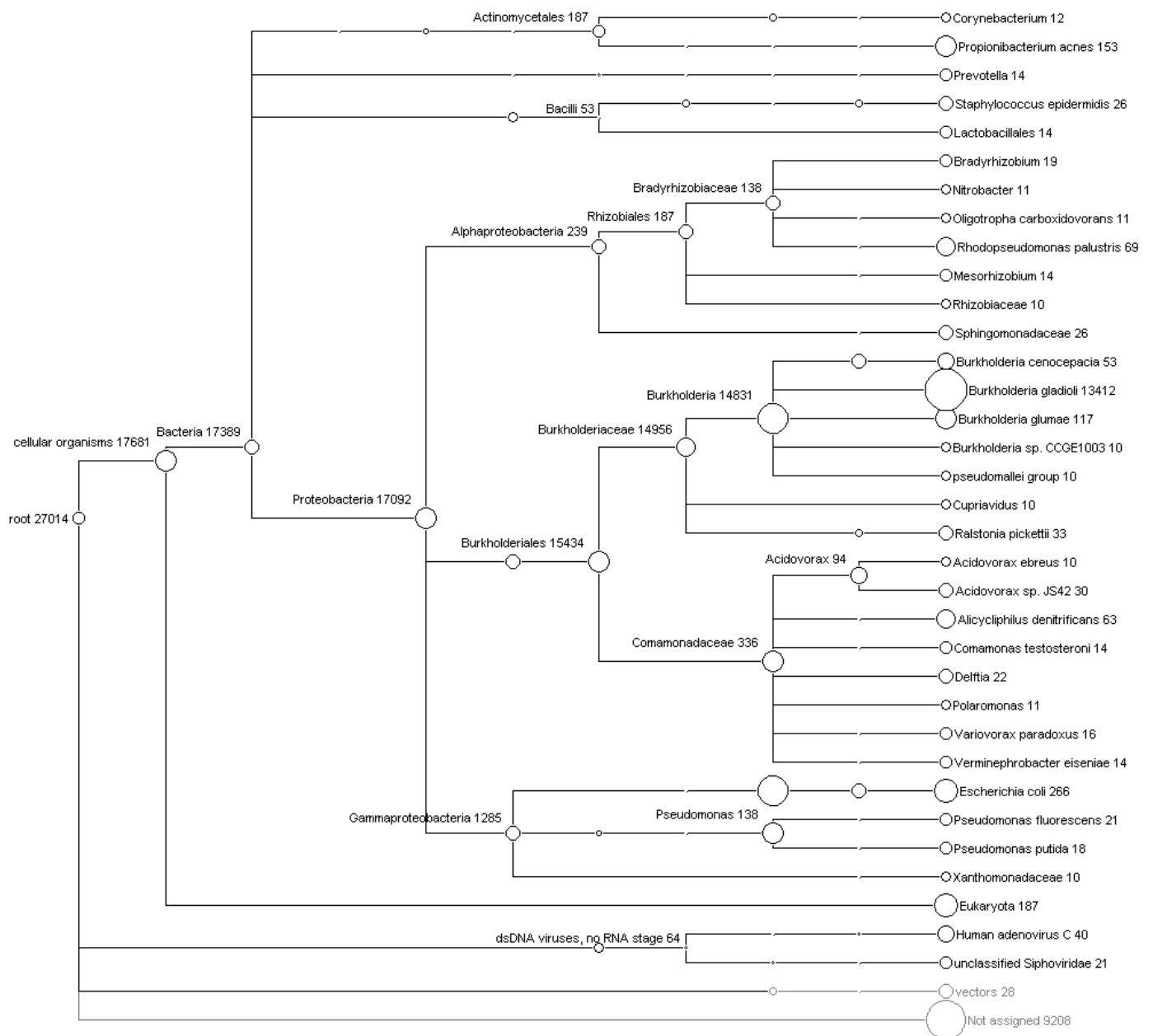
Metagenómica viral y bacteriana de enfermedades respiratorias graves  
 L.C.G. Ana Georgina Cobián Güemes

**Muestra 8836**



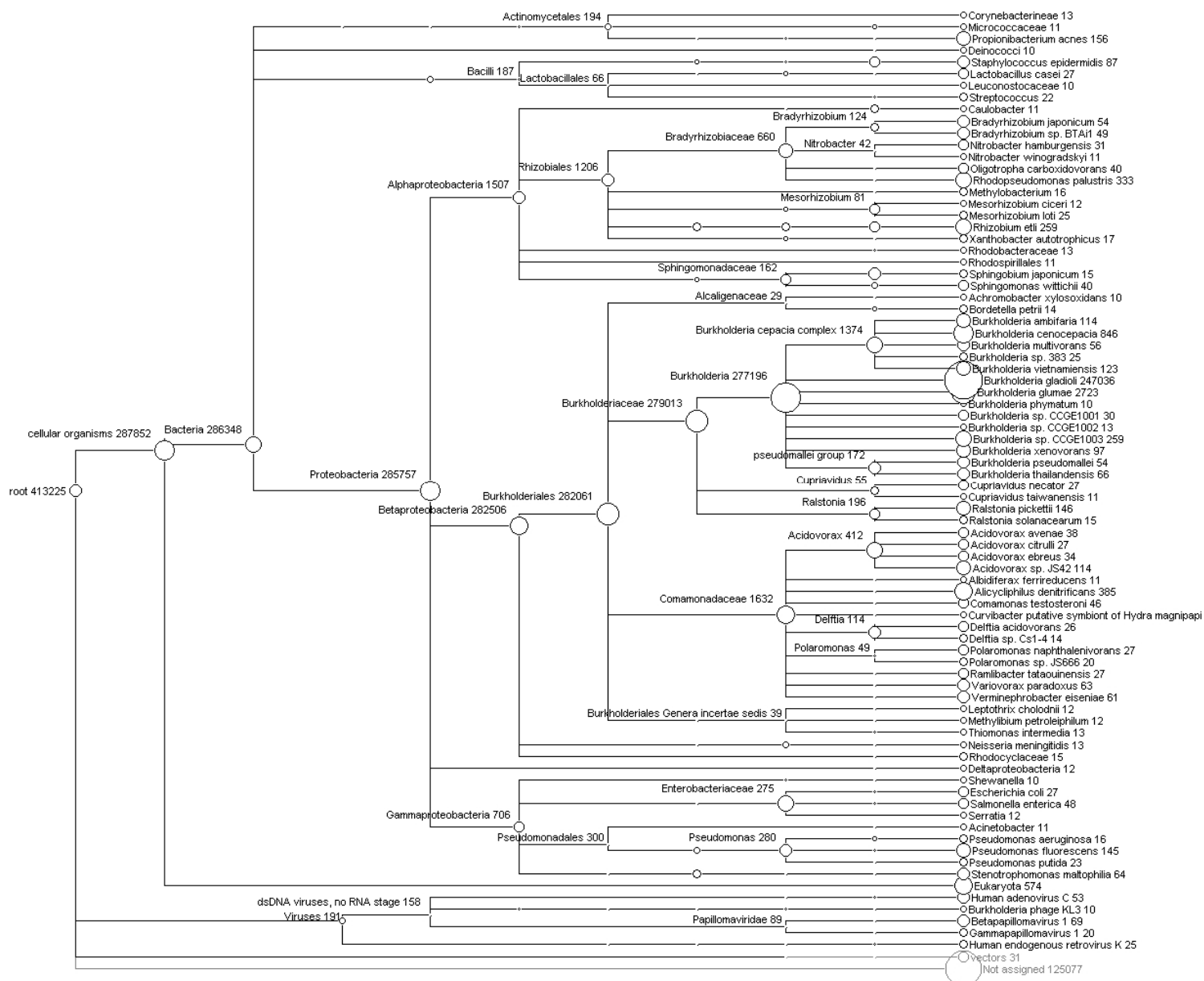
Metagenómica viral y bacteriana de enfermedades respiratorias graves  
 L.C.G. Ana Georgina Cobián Güemes

**Muestra 9041**



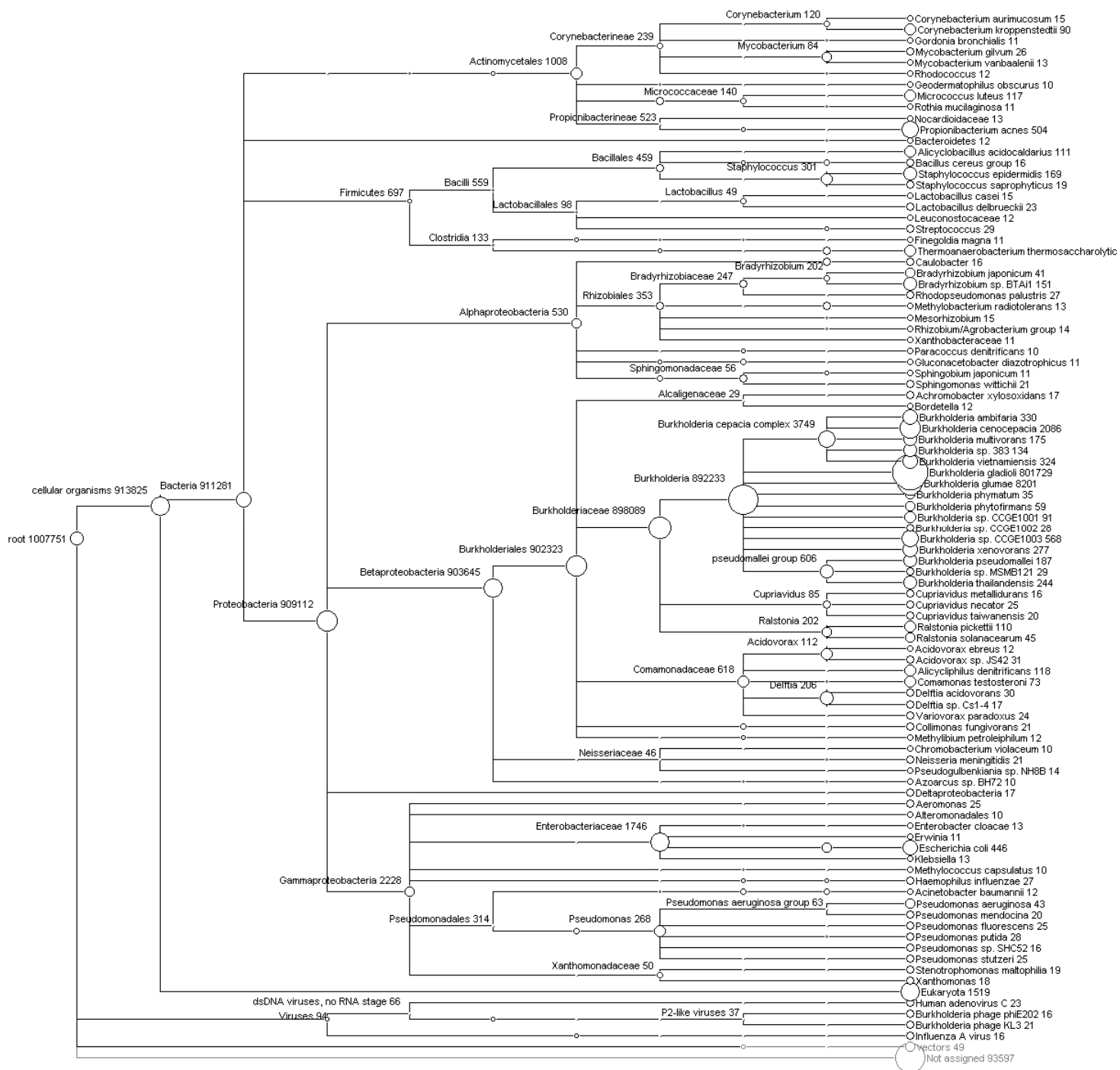
Metagenómica viral y bacteriana de enfermedades respiratorias graves  
L.C.G. Ana Georgina Cobián Güemes

Muestra 3974



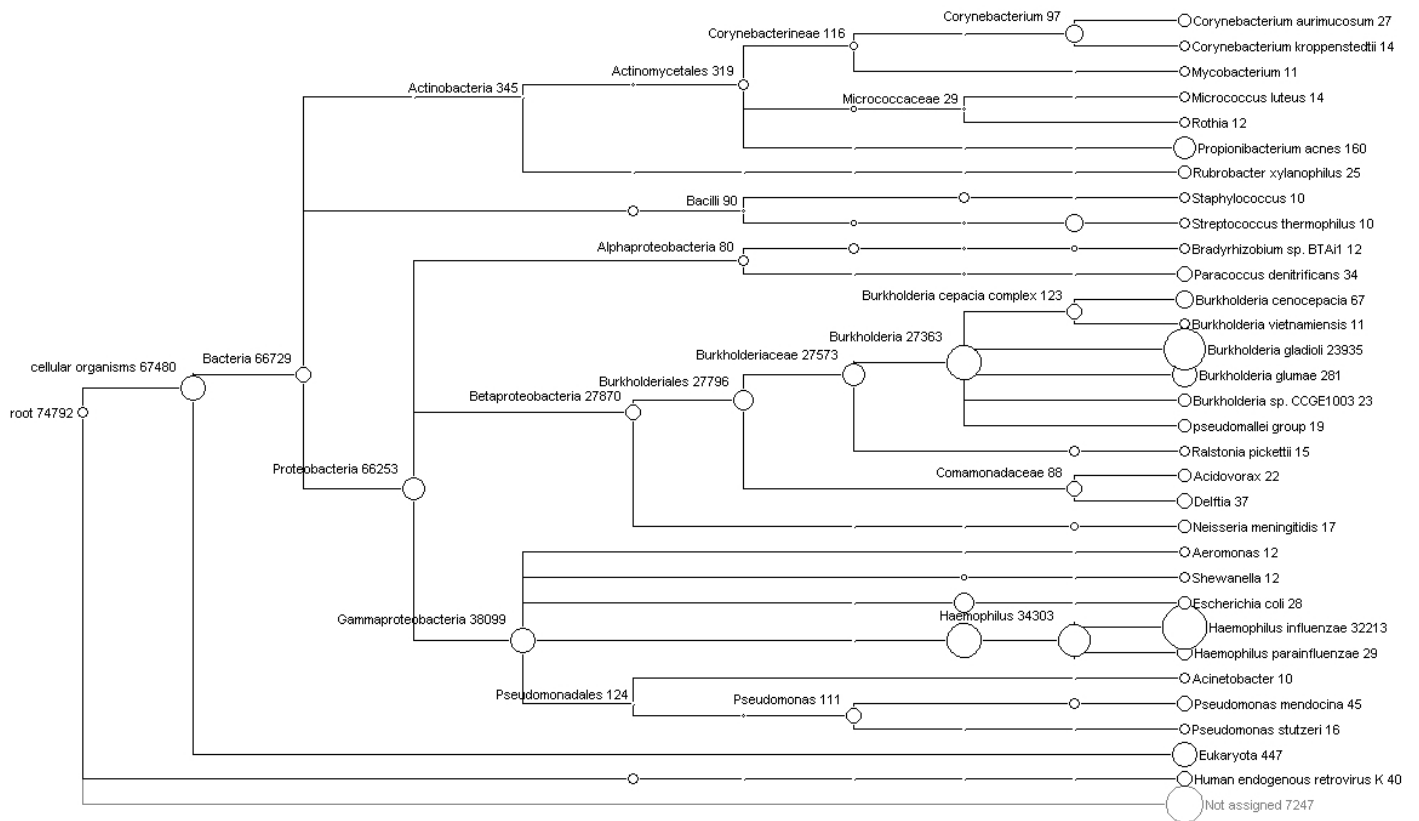
Metagenómica viral y bacteriana de enfermedades respiratorias graves  
L.C.G. Ana Georgina Cobián Güemes

Muestra 8410



Metagenómica viral y bacteriana de enfermedades respiratorias graves  
L.C.G. Ana Georgina Cobián Güemes

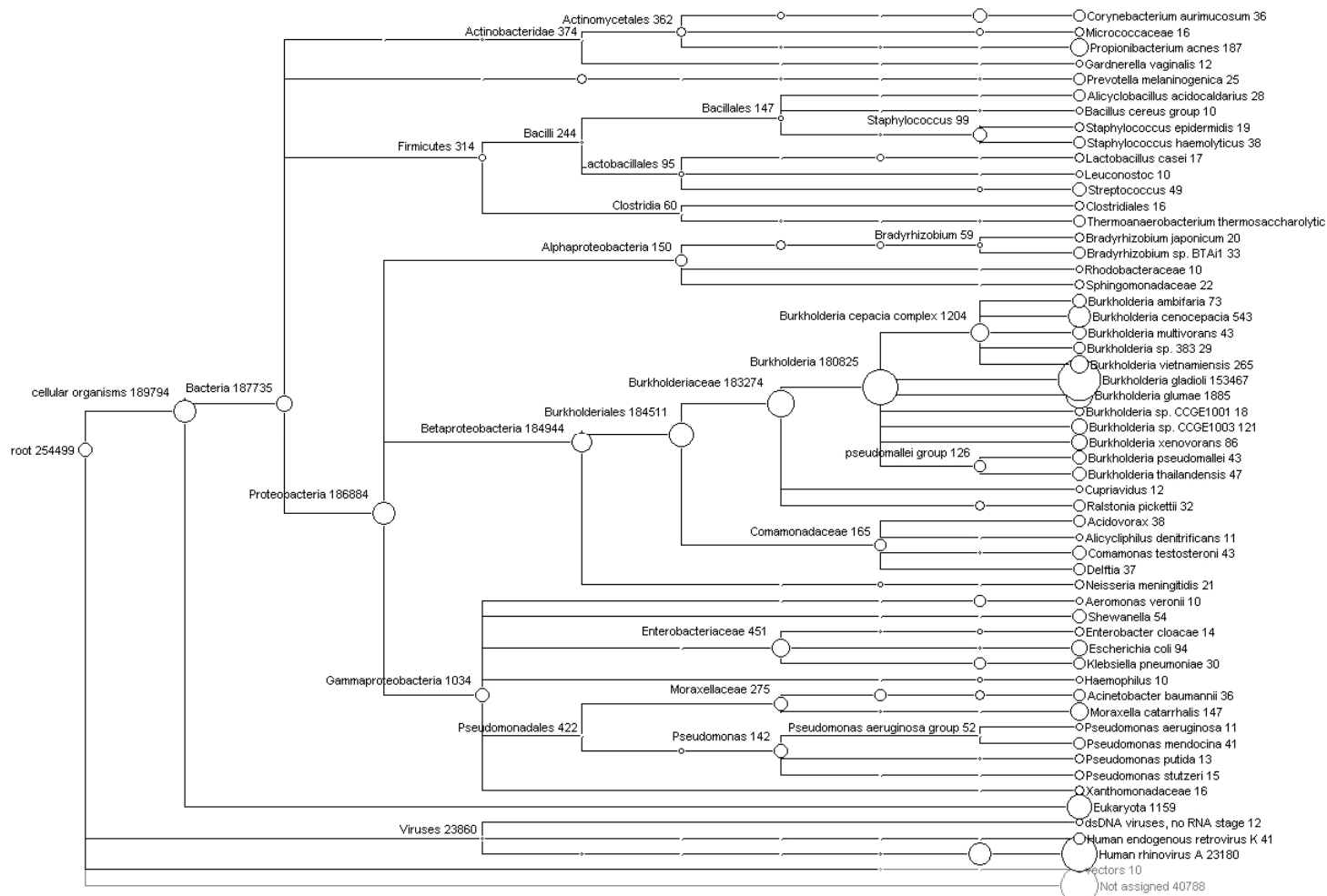
**Muestra 8883**



Metagenómica viral y bacteriana de enfermedades respiratorias graves  
 L.C.G. Ana Georgina Cobián Güemes

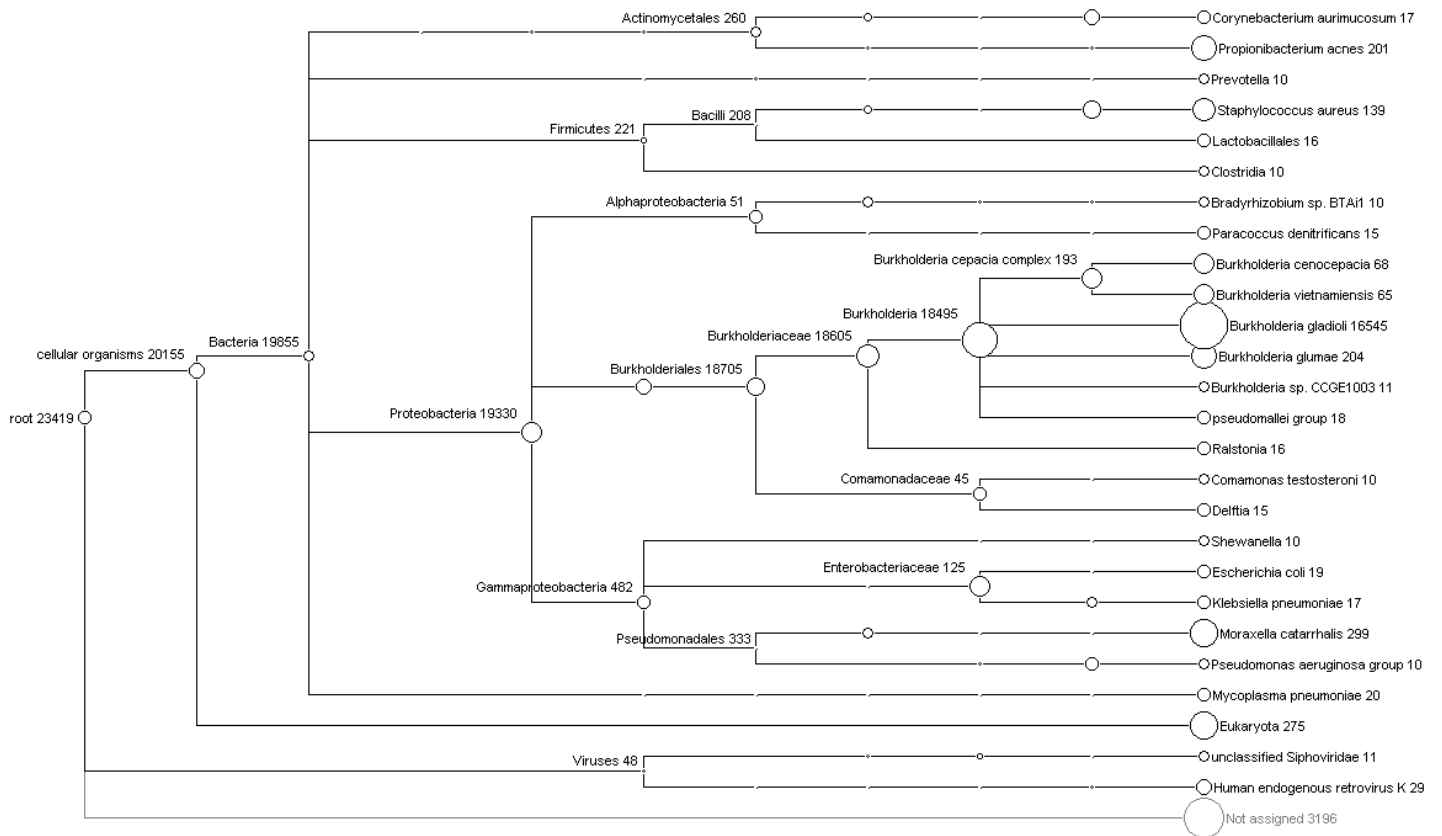


Muestra 9140



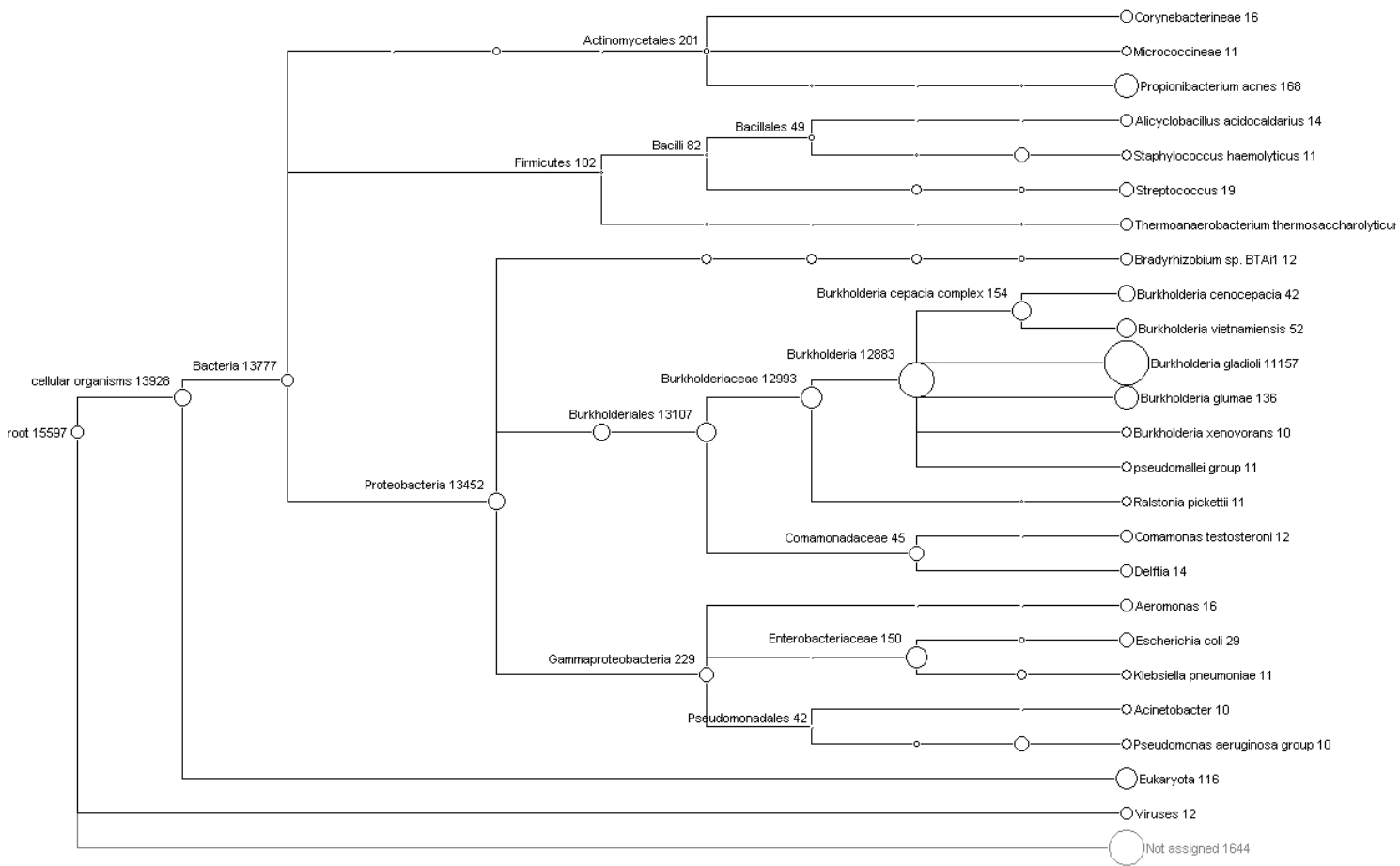
Metagenómica viral y bacteriana de enfermedades respiratorias graves  
L.C.G. Ana Georgina Cobián Güemes

**Muestra 9232**

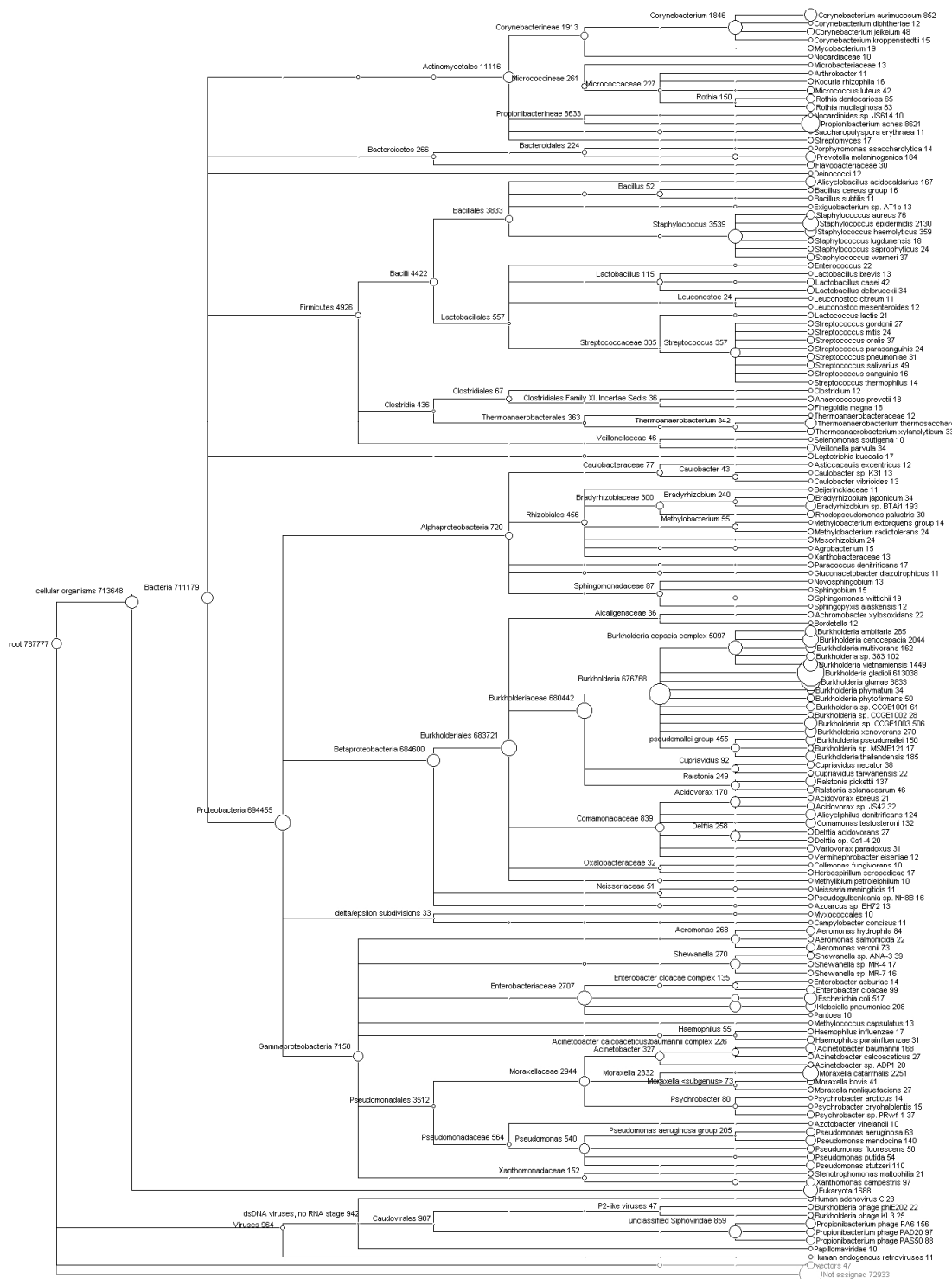


Metagenómica viral y bacteriana de enfermedades respiratorias graves  
L.C.G. Ana Georgina Cobián Güemes

**Muestra 9275**

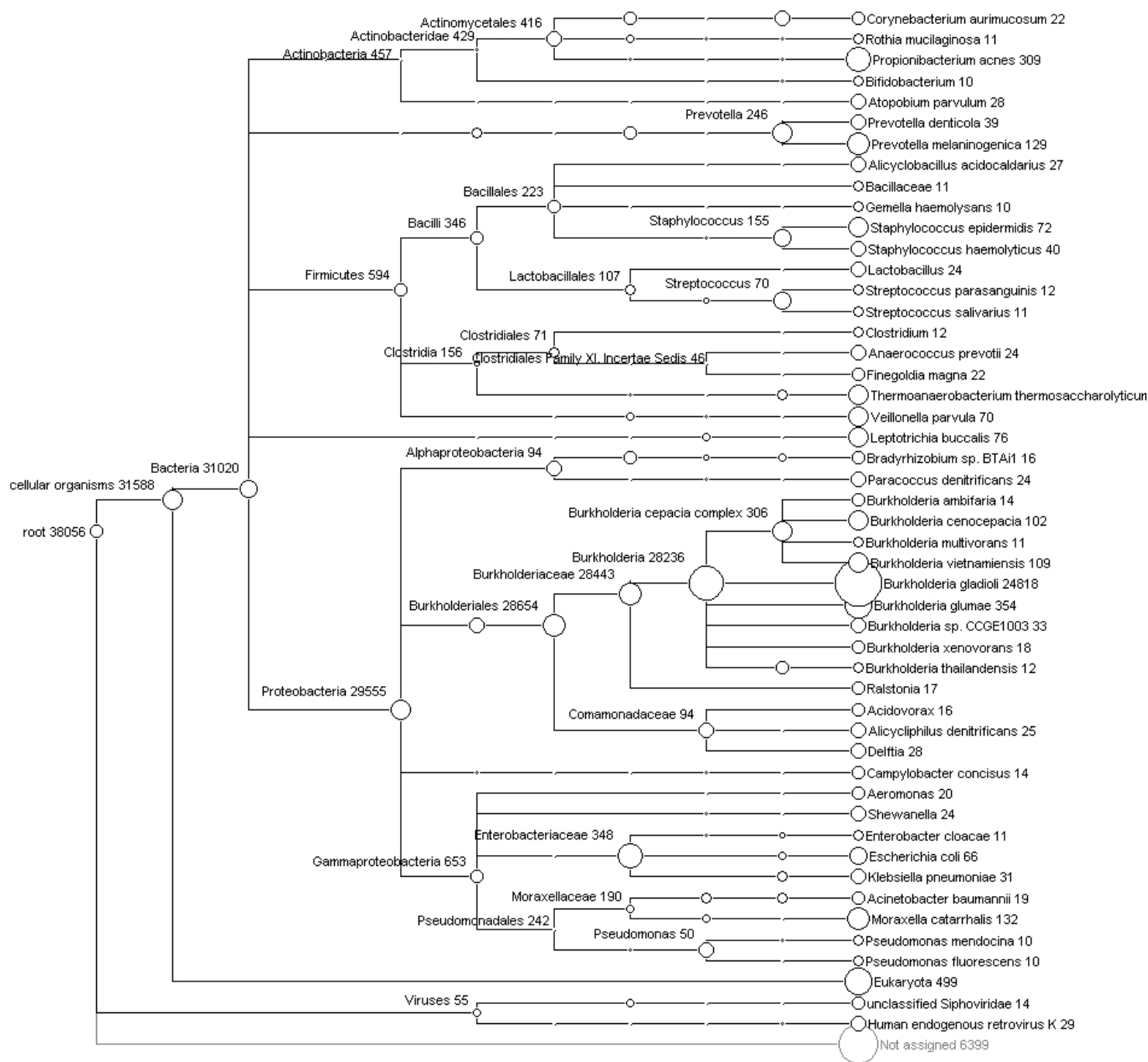


### Muestra 9362



Metagenómica viral y bacteriana de enfermedades respiratorias graves  
L.C.G. Ana Georgina Cobián Güemes

**Muestra 9434**



Metagenómica viral y bacteriana de enfermedades respiratorias graves  
L.C.G. Ana Georgina Cobián Güemes

REVIEW ARTICLE

## Molecular Anatomy of 2009 Influenza Virus A (H1N1)

Carlos F. Arias, Marina Escalera-Zamudio, María de los Dolores Soto-Del Río,  
Ana Georgina Cobián-Güemes, Pavel Isa, and Susana López

*Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, Mexico*

Received for publication September 22, 2009; accepted October 8, 2009 (ARCMED-D-09-00444).

---

Influenza A viruses are a major cause of morbidity and mortality worldwide and affect large segments of the population every year. The nature of their genome, formed by eight segments of single-stranded RNA, favors the constant evolution of the virus by two main mechanisms: the accumulation of single nucleotide mutations in the viral genes introduced by an error-prone viral RNA polymerase and the reassortment of genes between two strains of different origin. The viral genome encodes 11 proteins. Most have been shown to play a role in shaping the virulence scenario of influenza A viruses, including the adaptation of infection and transmission into new host species, the ability to modulate the host immune response, and the capacity to replicate efficiently at low temperature. On the surface of the virus particles there are two principal polypeptides, the hemagglutinin (HA) and the neuraminidase (NA), which are the target for the neutralizing antibodies immune response. There are 16 HA and 9 NA different subtypes in the influenza A virus that circulate in humans and animals. When a virus strain with a new HA or NA subtype appears in the human population by genetic reassortment, it usually causes a pandemic because there is no preexisting immunity against the new virus. This was the case for the three pandemics that occurred during the last century (1918, 1957, and 1968) and also for the first pandemic of the 21<sup>st</sup> century, caused by the currently circulating A (H1N1) 2009 virus, which was generated by gene reassortment between a virus present in pigs of North America and a virus that circulates in the swine population of Euroasia. © 2009 IMSS. Published by Elsevier Inc.

---

*Key Words:* Influenza, Biology, Pathogenesis, Virulence, Pandemics.

### Influenza Viruses

Influenza viruses of the family Orthomyxoviridae are classified into three different types: A, B, and C. Influenza A viruses infect a great variety of avian and mammalian species including humans. In contrast, there is only one known subtype of influenza B virus that primarily infects humans, thus the pandemic potential of influenza B viruses is very low, although they can cause severe respiratory illness. Influenza C viruses cause mild respiratory illness in humans and some animals, and it has been the least studied type (1,2).

It has been estimated that every year about 500 million people are infected with this virus, causing about 250–500 thousand deaths globally. Most of these cases occur during

the coldest months of the year, and they are known as seasonal influenza. Since 1977 two subtypes of influenza A H1N1 and H3N2, and influenza B, circulate seasonally in the human population. Given the epidemic and pandemic potential of the influenza A viruses, attention of the national and international health organizations is focused on them. Based on the antigenicity of their hemagglutinin (HA) and neuraminidase (NA) proteins, influenza A viruses are classified into different subtypes; there are 16 subtypes of HA (H1–H16) and 9 subtypes of NA (N1–N9) (3,4). Wild aquatic birds like ducks, gulls, and seabirds are the most important reservoir of these viruses because in this population circulate all the described H and N subtypes. It is thought that they are the source of viruses that infect other avian and mammalian species. Many subtypes of influenza A infect asymptotically or cause mild symptoms in birds; however, some strains of subtypes H5 and H7 are highly pathogenic and cause severe illness and death in some wild as well as

---

Address reprint requests to: Carlos F. Arias, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Av. Universidad 2001, Colonia Chamilpa, Cuernavaca, Morelos 61210, Mexico; Phone: (+52) (777) 311-4701; E-mail: [arias@ibt.unam.mx](mailto:arias@ibt.unam.mx)

domestic avian species. Similar to avian infections, some influenza strains can asymptotically infect pigs, whereas other strains can cause flu-like symptoms in these animals like fever, cough, and nasal secretion. In contrast to birds, pigs are infected seasonally primarily with influenza subtypes H1N1, H3N2, and H1N2 (2).

### Virus Structure

Influenza A viruses have a pleomorphic appearance when observed under the electron microscope. Some particles appear spherical with an average diameter of 100 nm, whereas others show elongated forms of about 300 nm in length. The viral particle has a lipid envelope where three viral proteins are inserted, the glycoproteins HA and NA, and small amounts of the transmembrane protein M2. The lipid bilayer containing HA, NA, and M2 surround an inner protein layer formed by the matrix protein M1 that encloses the viral genome associated with the nucleoprotein NP and with small amounts of PB1, PB2, and PA that form the RNA polymerase complex. Inside the viral particle it is also the nuclear export protein NEP, also known as nonstructural protein NS2 (5,6).

The genome of influenza A viruses is formed by eight segments of single-stranded RNA of negative polarity that vary in size from 2350 nucleotides (nt) to 890 nt. In total, the genome contains about 13,600 nt and encodes 11 proteins. All segments code for one protein except the PB1 gene, which also encodes PB1–F2 in some viral strains, the matrix gene that codes for M1 and M2, and the smallest gene that codes for NS1 and NS2 (NEP). All viral genes have 5' and 3' non-translated regions (NTR) of variable length that contain the replication and encapsidation signals. The last 13 and 12 nt of the 5' and 3' ends, respectively, are highly conserved among all genes of the virus (1,6).

### Virus Replication Cycle

To initiate infection, the HA protein on the virus surface interacts with sialic acid (SA)-containing cell receptors on the surface of the target cell (see below). After cell attachment, the virus particle is internalized through classical receptor-mediated endocytosis in clathrin-coated vesicles. Acidification of endosomes is critical for the next step of the replication cycle of the virus in two manners: a) the low pH provokes a conformational change in the hemagglutinin that favors the fusion of the viral and cellular membranes, allowing the virus to enter the cell cytoplasm, and b) the M2 protein, which functions as a proton channel, allows the acidification of the interior of the viral particle, promoting the uncoating of the virus and allowing the ribonucleoprotein complexes to dissociate from the viral particle, such that they are released into the cytoplasm and transported to the cell nucleus. The replication

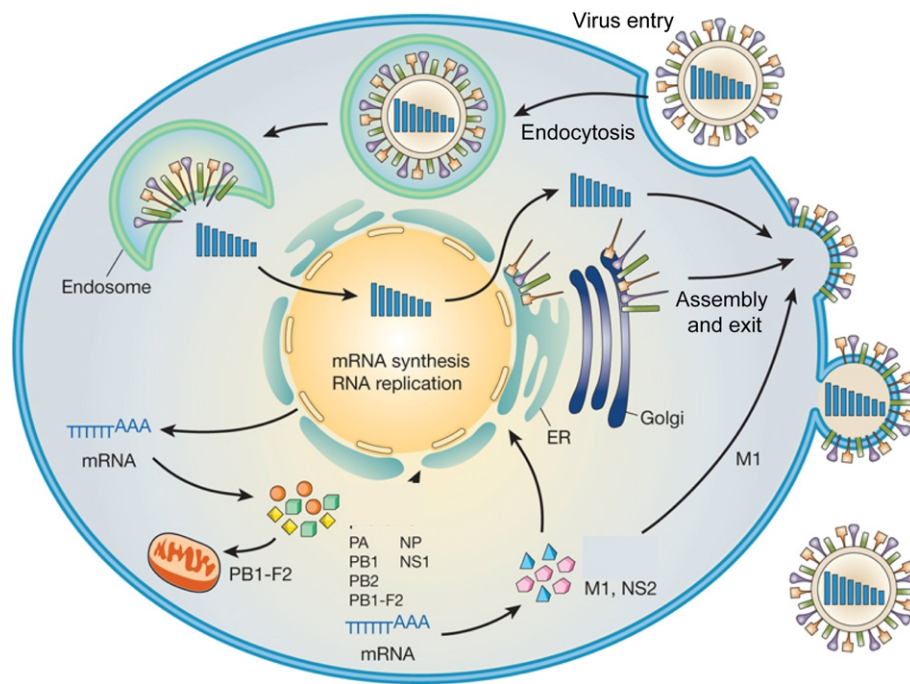
and transcription of the viral RNAs are carried out by the RNA polymerase complex in the nucleus of the infected cell and are exported to the cell cytoplasm by the proteins NEP and M1, where viral mRNAs are translated. The progeny virions assemble and bud at the plasma membrane that has been previously modified with the viral membrane proteins HA, NA, and M2 (Figure 1) (6).

### Antigenic Variation and Evolution of Influenza A Viruses

Influenza A viruses evolve constantly using different mechanisms. The most important is the lack of proofreading activity of the viral RNA polymerase during replication of the influenza genomic RNA segments, which results in a high level of point mutations (antigenic drift) (7). The rate of mutation during replication of the influenza genome is about 1 nucleotide change for every copied genome ( $7.3 \times 10^{-5}$  mutations/replicated nucleotide) (8). The evolutionary rate differs among the viral segments, probably reflecting differences in selective pressure of the infected host; thus, the highest evolutionary rates are observed in proteins that are target of neutralizing antibodies—hemagglutinin and neuraminidase—and, for unknown reasons, the M2 protein. The mutation frequency in the H and N amino acid sequences during a 1-year period is estimated to be <1%, and it takes 2–5 years to achieve differences which can be detected in serological tests such as the hemagglutination inhibition assay. The mean evolutionary rate of the influenza A (H1N1) 2009 epidemic virus is  $3.66 \times 10^{-3}$  substitutions/site/year (9). The permanent antigenic drift in the principal targets of the immunological response, the proteins HA and NA, is the cause of the constant need to review the viruses that are included in the yearly prepared vaccines. The antigenic drift that has occurred in influenza A viruses for thousands of years has led to the current diversity of HA and NA subtypes (10) (Figure 2).

The second mechanism of influenza virus evolution results from the segmented nature of its genome, which permits the formation of new progeny viruses having novel combination of segments when two or more different virus subtypes infect a single cell (antigenic shift). This process is capable of introducing new proteins in circulating viral populations that can drastically change the biological properties of the virus. Antigenic shift is commonly associated with appearance of pandemic influenza viruses; for instance, the acquisition of new PB1, H and/or N proteins to which the human population was not immune, played a major role in the 1957 and 1968 pandemics (11). Co-circulating seasonal virus strains can also reassort frequently, with so far unknown impact on antigenic evolution (12).

A third mechanism associated with evolution of influenza A viruses is recombination by template switching. This type



**Figure 1.** Replication cycle of influenza A viruses. After receptor-mediated endocytosis, the viral ribonucleoprotein (vRNP) complexes are released into the cytoplasm and subsequently transported to the nucleus where replication and transcription take place. Messenger RNAs are exported to the cytoplasm for translation. Early viral proteins, that is, those required for replication and transcription, are transported back to the nucleus. Late in the infection cycle, the M1 and NS2 proteins facilitate the nuclear export of newly synthesized vRNPs. PB1-F2 associates with mitochondria. The assembly and budding of progeny virions occurs at the plasma membrane. (Modified with permission from Neumann, Noda & Kawaoka, *Nature* 459:931, 2009). Color version of this figure available online at [www.arcmedres.com](http://www.arcmedres.com).

of recombination may involve genetic material from more than one origin (for instance, ribosomal RNA and viral RNA) or two different viral RNA segments, and these nonhomologous recombination events could be associated with changes in viral pathogenicity (13,14). On the other hand, occurrence of homologous recombination, involving template switching between homologous RNA segments from two different influenza virus strains, has been more controversial. Whereas this process was found to be involved in evolution of the avian influenza isolates, it seems to be a rare event in human influenza A viruses (15,16). Nonetheless, the increased identification of recombinant viruses in recent years may indicate that this sort of recombination occurs more often than previously believed.

### Viral Tropism and Receptor Specificity

Influenza A viruses infect a wide variety of animals. The initial step of influenza virus cell infection is the recognition by the viral HA of *N*-acetylneuraminic acid (SA) molecules on the surface of host cells. SAs are nine-carbon monosaccharides commonly found at the terminal end of many glycoconjugates, usually bound to galactose through  $\alpha$ 2,3 or  $\alpha$ 2,6 linkages (Figure 3). SA is a major determinant of the tropism of influenza viruses given the high specificity of viral strains for the type of SA linkage. Thus, viruses that infect humans

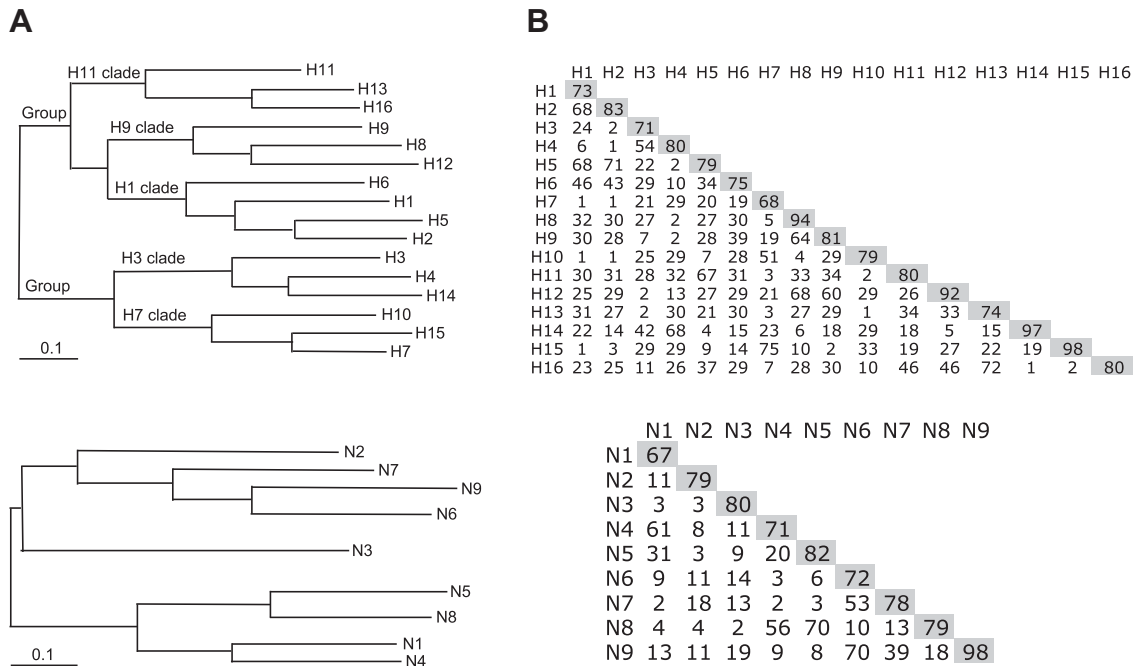
bind preferentially to SA with  $\alpha$ 2,6 linkages, whereas avian viruses bind mostly to SA in an  $\alpha$ 2,3 configuration. The human tracheal epithelial tissue preferentially express SA with  $\alpha$ 2,6 linkages, whereas avian epithelial gastrointestinal tissue (where influenza viruses replicate in avian hosts) contain mainly SA in an  $\alpha$ 2,3 configuration. Of relevance, pig tracheal tissue has both kinds of SA linkages, and swine can thus be infected with viruses that recognize both types or receptors, hence, these animals have often been proposed as a mixing vessel for the generation of new strains from co-infections of avian and mammalian viruses (1,6,17).

### Molecular Determinants of Virulence

Most of the viral proteins have been shown to play a role in shaping the virulence scenario of influenza A viruses, including the adaptation of infection and transmission into new host species, the ability to modulate the host immune response, and the capacity to replicate efficiently at low temperature, among others (Table 1). The proteins that have been most characterized with regard to their pathogenic potential are HA, PB1, PB2, and NS1. On the other hand, two proteins are associated with the development of resistance to antiviral drugs, NA and M2.

It has been shown that virulence in influenza A viruses can be directly influenced by exchange of complete

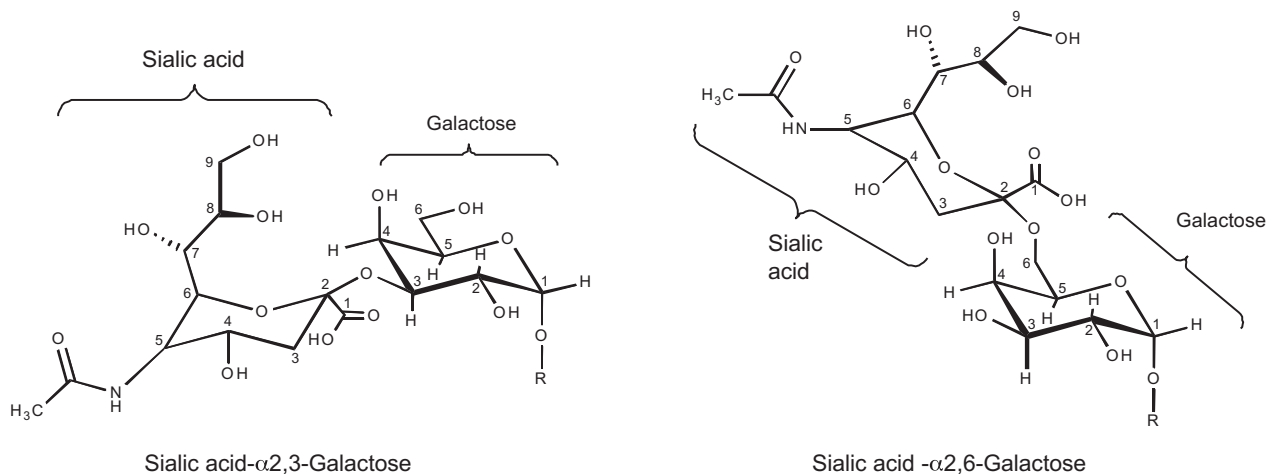




**Figure 2.** (A) Phylogenies of 16 influenza virus HA subtypes and of 9 NA subtypes. Maximum-likelihood trees were generated by comparing the nucleotide sequences. (B) Nucleotide identity (%) among the consensus sequences of each subtype of HA and NA. All the available sequences at the NCBI database were downloaded for each subtype of HA and NA. Those sequences within each subtype with more than 90% identity were removed to eliminate redundancy, and the consensus sequence was obtained for each HA and NA subtype. The generated consensus sequences were pairwise compared, and are shown in the figure. Shaded values refer to the lowest percent identity of the non redundant sequences within each subtype of HA and NA.

genomic segments or by single point mutations. For example, it has been shown that the full gene segments PB1, HA, and NA of the pandemic strain influenza A (H1N1) 1918 are critical for the virulence of this virus strain in mice (18). On the other hand, a variety of point mutations have been described in different viral proteins, which have been described to be involved virulence (see below and Table 1). It is important, however, to interpret with caution the presence of the reported mutations as

definitive determinants of virulence in humans because mutations in some viral strains have been associated with virulence in particular animal models but not in others, or have not been found to be relevant epidemiologically. For instance, some strains of influenza H5N1 have been classified as highly pathogenic in domestic and wild birds from epidemiological data, yet when experimental infections have been conducted in mammals such as mice, these strains only cause mild symptoms (19). Likewise, the



**Figure 3.** Linkages of sialic acid to galactose recognized by influenza viruses.

current pandemic strain influenza A (H1N1) 2009 has been reported to replicate more efficiently in mice and ferrets than a seasonal H1N1 virus and to cause more severe pathological lesions in the lungs of infected mice, ferrets and non-human primates than a currently circulating human H1N1 virus (20,21); however, epidemiologically the pandemic influenza A (H1N1) 2009 virus has been found not to be more virulent in people than seasonal viruses.

Furthermore, highly virulent strains not always contain all identified virulence markers, whereas some low virulence strains might present some of them; thus, there is not always a correlation between proposed virulence markers and virulence. In addition, it is important to keep in mind that the virulence of influenza viruses in mammals is thought to be polygenic; thus, given the diversity of viruses and the polygenic nature of this trait, general conclusions about virulence cannot be drawn with confidence from individual studies (19). These observations make it difficult to reach a consensus regarding which mutations are relevant virulence factors in mammals, particularly in human infections. Nonetheless, available data suggest that a multibasic sequence at the HA proteolytic cleavage site and the presence of a lysine at position 627 of PB2 (see below) seem to be universal determinants of viral pathogenicity.

### Hemagglutinin and Its Role in Viral Tropism and Virulence

Hemagglutinin plays an important role during the cell entry of influenza viruses. This protein is key during the initial steps of infection because it is responsible for attachment of the virus to SA, the cellular receptor. As mentioned above, this interaction explains, at least in part, the host range and tissue tropism of influenza viruses (2). The amino acids in the active site of HA that play a critical role in receptor binding and specificity have been well characterized. For instance, the presence of glutamine and glycine at amino acid residues 226 and 228 in HA of the H2 and H3 subtypes allows recognition of avian SA receptors ( $\alpha$ 2,3), whereas leucine and serine residues at the same positions confer binding to human SA receptors ( $\alpha$ 2,6) (22–25). On the other hand, for HA of the H1 subtype, glutamic acid and glycine residues at positions 190 and 225 are responsible for the interaction with avian SA receptors, whereas the presence of aspartic acid at the same positions provide binding specificity to human SA receptors (26–28). The influenza A (H1N1) 2009 strains that have been characterized so far have aspartic acid at amino acid residues 190 and 225, explaining, at least partially, its fit capability to infect human hosts (29,30).

In addition to being responsible for virus cell attachment, hemagglutinin is critical for virus cell entry because it mediates the fusion of viral and cellular membranes. In

addition, it is the main target of neutralizing antibodies (2). HA is a trimeric rod-shaped protein that possesses a transmembranal domain inserted into the viral membrane and a globular head comprised of three identical domains that form a spike projecting away from the viral surface. The globular head has at least five predicted antigenic sites that surround the receptor-binding site in the A (H1N1) 1918 strain (31). The HA protein is synthesized as a precursor HA<sub>0</sub> molecule that is posttranslationally cleaved into domains HA<sub>1</sub> and HA<sub>2</sub>. This proteolytic cleavage is essential for viral replication because, under the low pH conditions present in the endocytic vesicles, it allows a conformational change of HA that exposes a hydrophobic fusion peptide located at the amino terminus of the HA<sub>2</sub> domain. This fusion peptide is essential to promote the fusion of the viral and cellular membranes that permit the exit of the viral ribonucleoprotein complex into the cytoplasm (32). The HA protein is a critical determinant of virulence and pathogenicity of influenza viruses with a clear link between cleavage and virulence. Highly virulent strains of avian influenza (H5 and H7) have an HA protein with multiple basic amino acid residues at its cleavage site that are recognized by ubiquitous proteases; thus, infection with these strains can result in systemic spread of the virus. On the other hand, low virulence strains have HA proteins with only one arginine residue at their cleavage site and are only cleaved by tissue-specific proteases such as trypsin, resulting in infections that are usually localized and mild (33). The influenza A (H1N1) 2009 strains possess only one arginine at the HA cleavage site (29,30).

### PB1-F2: An Apoptosis-inducing Factor

PB1-F2 is a small protein (87–89 amino acids) encoded in an alternative reading frame of the PB1 polymerase gene. Although this protein is not expressed in all influenza strains, it has been reported that it contributes to the pathogenicity of the viral strains in a mouse model by inducing apoptosis. This pro-apoptotic factor interacts with two proteins of the mitochondrial membrane, promoting permeabilization of the membrane and release of cytochrome c (34). It has also been shown that PB1-F2 can exacerbate the pro-inflammatory response of the host during a primary viral infection in mice, which increases the frequency and severity of secondary bacterial infections. The mechanism through which this occurs is not known (35). It has been shown that a single amino acid change (asparagine to serine at position 66) correlates with the virulence of H5N1 strains. This amino acid substitution is also present in the PB1-F2 protein of the influenza A (H1N1) 1918 pandemic virus (36). The PB1 gene of the influenza A (H1N1) 2009 viruses encodes a truncated, and thus most probably non-functional, PB1-F2 protein of 11 amino acids (29,30). Furthermore, mutations

**Table 1.** Amino acid mutations associated with pathogenesis and virulence of influenza A viruses, as well as with resistance to antiviral drugs

Gene	Mutation	Effect	Strain	Model	A/H1N1/2009 <sup>a</sup>	References
PB2	E627K	-627K associates with increased transmission. -Important determinant of host range. Avian strains have 627E, human strains 627K. -627K associates with advantageous replication at 33°C and confers air transmission of avian virus. -627K is linked to increased virulence in mice.	H5N1 H1N1/1918	Guinea pigs Squirrel monkey Mouse Human	99.6% 627E 0.4% 627K	38-41, 69,70
PB2	D701N	Modulates polymerase activity in avian and mammalian cells, determining host range.	H7N7	Mouse	100% 701D	17
PB2	S714R	Enhances polymerase activity and increases virulence in mice.	H7N7	Mouse	100% 714S	37
PB2	K318R K335R/Q	-Correlate with high pathogenicity in mice, in the presence of additional mutations. -Suggested as mammalian virulence determinants by multivariate analysis.	H5N1	Computational, <sup>b</sup> mouse	100% 318R 100% 335G	19,72
PB1	L13P S678N	L13P confers enhanced activity of viral polymerase, changes affinity for PA binding and is associated with increased virulence.	H7N7	Mouse	100% 13P 99.3% 678S 0.7% 678N	37
PB1	I317M/V	Correlates with high pathogenicity in mice, in the presence of additional mutations.	H5N1	Computational, <sup>b</sup> mouse	99.7% 317M 0.3% 317V	72,73
PB1-F2	N66S	Increases pathogenicity, w/weight loss, decreased survival, and increase in virus replication.	H5N1 H1N1/1918	Mouse	Truncated protein	36
PA	K615N	Enhances activity of viral RNA polymerase and stimulates viral replication and pathogenicity.	H7N7	Mouse	100% 615N	71
PA	I127V	Associates with increased virulence.	H5N1	Computational <sup>b</sup>	99.7% 127V 0.3% 127I	19,74
HA	D190E D225G	Cause a switch in receptor binding preference from human to avian receptor. This results in an incapability of transmission via respiratory droplets, though pathogenicity in mammals is maintained.	H1N1/1918 H5N1	Ferrets	90.7% 190D 7.4% 190N 1.8% 190V 95.4% 225D 1.8% 225E 1.8% 225G 1.0% 225N	26-28
HA	A200T	Contributes to resistance to zanamivir.	H1N1 (human)	<i>In vitro</i>	98.6% 200A 1.4% 200T	51
HA	Q192R N182K	Involved in the binding of H5N1 influenza A viruses to human-type receptors.	H5N1 H1N1 human isolates	<i>In vitro</i>	70% 192Q 30% 192K 70% 182I 30% 182V	76
HA	Q226L G228S	In H2 and H3 viruses 226Q and 228G correlate with binding to avian receptors; 226L and 228S correlate with the shift from avian to human receptor binding.	Human isolates	<i>In vitro</i>	95% 226Q 5% 226R 100% 228G	22-24
HA	S227I	Reduces the ability to support virulence. Amino acid located in the receptor-binding pocket of HA.	H5N1	Mice	99.6% 227E 0.4% 227K	77

NP	N319K	Enhances binding to mammalian importin- $\alpha$ proteins, increases polymerase activity. Related to host range specificity.	H7N7	<i>In vitro</i>	99.4% 319K 0.6% 319N	42
NA	E119G/A/V/D R292K H274Y	-E119G/A/V/D and R292K contribute to resistance to the antiviral drug zanamivir. -Oseltamivir resistance is associated to H274Y (N1), R292K (N2), and E119 (N2) mutations. -Resistance to new drugs: Peramivir resistance is associated to R292K; A-315675 resistance is associated to changes in 119.	H1N1/1933 H1N9 H1N1/1947 H1N1 (human)	<i>In vitro</i>	100% 119E 100% 292R 85% 274H 15% 274Y	52,79
NA	294N, 116V, 117I, 136Q, 150K, 222I	Changes in these amino acids may contribute to generate resistance to NA inhibitors.	H5N1	<i>In vitro</i>	100% 294N 100% 116V 100% 117I 100% 136Q 100% 150K 100% 222I	80
NA	N146K, S219T, A272V Deletion 245-248	Oseltamivir resistance conferred by 3 aa substitutions and a 4-aa deletion.	H3N2	Medical case	100% 146N 100% 219W 100% 272A	50
M2	S31N	Confers cross-resistance to adamantane drugs.	Human isolates	<i>In vitro</i>	100% 31N	48
M2	L26F, V27A, A30T, G34E	Confer cross-resistance to adamantane drugs.	Swine isolates H1N1 H1N2 H3N2	<i>In vitro</i>	100% 26L 100% 27V 100% 30A 100% 34G	49
NS1	D92E	Confers cytokine resistance (particularly to INF- $\beta$ and TNF- $\alpha$ ). Increases virulence in infected animals.	H5N1 A/swine/Texas/ 4199-2/98	Swine	100% 92D	81,82
NS1	A149V	Attenuates the NS1 antagonism to interferon.	H5N1	<i>In vitro</i>	100% 149A	46
NS1	N127T/D/V/R/A	The 127 site mediates both inhibition of PKR activation and temporal regulation of viral RNA synthesis.	H5N1	<i>In vitro</i>	99.4% 127N 0.6% 127T	83
NS1	PDZ domain ESEV EPEV KSEV RSKV RSEV	-Increased virulence in animal models. The increment in virulence is not related to unbalanced cytokine production. -ESEV and EPEV of avian strains correlate with high virulence. -KSEV of H1N1/1918 correlates with virulence. -RSKV and RSEV, common in human strains, correlate with low virulence.	H1N1/1918 H5N1 Human isolates	Mouse	NS1 is truncated. The 11 carboxy-terminal amino acids are deleted. It lacks the PDZ domain	44
NS1	P42S	42S is critical for virulence of H5N1 strains and to antagonize host cell interferon induction.	H5N1	Mouse	100% 42S	45
NS1	N189D/G T/Y195S	Associate with increased virulence.	H5N1	Computational <sup>b</sup>	99% 189G 1% 189D 100% 195S	19,74
NS2	I31M Y56H/L	Associate with increased virulence.	H5N1	Computational <sup>b</sup>	99.8% 31M 0.2% 31I 100% 56H	19,74

Mutations identified as adaptation markers of avian viruses able to replicate in mice: L13P and S678N in PB1, D701N and S714R in PB2, K615N in PA, and N319K in NP.

<sup>a</sup>The information on the H1N1/2009 epidemic virus was obtained from (29,30), or it is based on our analysis of virulence-related mutations of complete protein sequences available at NCBI Influenza Virus Resource on September 4, 2009.

<sup>b</sup>Theoretical data. Not tested experimentally.

in the full-length polymerase subunit PB1 protein such as L13P and S678 N (Table 1) have been found to correlate with an increased virus replication in mice and a consequent enhanced virulence (37).

### **PB2 and Viral Replication at Low Temperatures**

Various mutations in PB2 have been associated with diverse virulence traits including transmission and host range restriction. The substitution of glutamic acid at amino acid residue 627 (present in most avian viruses) for lysine (the amino acid present in most human influenza strains) converts an H5N1 virus from nonpathogenic to pathogenic in mice (38). Viruses with lysine at position 627 also replicate better at 33°C (the temperature found in the human upper respiratory tract—as compared to 41°C, the body temperature of birds) than do viruses with glutamic acid at that position, whereas both variants replicate efficiently at 37°C (30,40). Efficient replication not only in the lower but also in the upper respiratory tract of mammals may facilitate transmission. In fact, a Lys627Glu mutation reduces the transmissibility of human influenza viruses in guinea pigs (38). Furthermore, the PB2 protein of the 1918 pandemic strain was found to be required for airborne transmission and efficient replication of the virus at 33°C (41). Lysine at position 627 is now recognized as a determinant of viral pathogenicity in several mammalian species (30).

Other adaptive mutations in PB2 such as Asp701Asn enhance binding of this protein to importin  $\alpha$ 1 (a cellular nuclear import factor) in mouse cells and correlate with a better replication in mammalian cells (42). These data suggest that efficient transmission among humans is favored in strains that contain the molecular markers 627Lys and 701Asn in PB2. The current pandemic virus influenza A (H1N1) 2009 has glutamic and aspartic acids at positions 627 and 701, respectively, in the great majority of the virus sequences reported (29,30).

### **NS1 and Control of the Host Immune Response**

NS1, as a nonstructural protein of influenza A viruses, is not a component of the viral particles but is expressed at very high levels in infected cells. This protein is dimeric, localizes to the nucleus, and has the capacity to bind double-stranded RNA (dsRNA). It antagonizes the cellular interferon response by binding dsRNA, thus preventing the activation of PKR (a dsRNA-dependent protein kinase) as well as of transcription factors and interferon- $\beta$  stimulated gene products (2,6). NS1 is a non-essential virulence factor that has accessory functions during viral infection and it may differ among viruses in its ability to counteract the cellular interferon response and thereby modulate viral pathogenicity. The exact mechanism by which this protein acts is not yet understood. Several mutations described in

NS1 have been implicated in virulence and pathogenicity because they confer differential ability to modulate the host immune response. For instance, the change of aspartic acid for glutamic acid at position 92 correlates with high pathogenicity in avian strains (H5N1) (43); likewise, the NS1 protein of highly pathogenic H5N1 strain and the 1918 H1N1 pandemic virus possess a four-residue motif (either ESEV or EPEV) at the carboxy-terminal end of the protein that forms a potential PDZ domain (a protein-protein recognition module present in many cellular proteins that initiate diverse cell-signalling cascades) associated with high virulence. Low virulence strains lack the PDZ binding domain or have a modified version of it (RSKV or RSEV) (44). Interestingly, the influenza A (H1N1) 2009 strains sequenced so far have a deletion of the 11 carboxy-terminal amino acids and thus lack the PDZ domain and have aspartic acid at position 92 (29,30). In addition, it has been shown that NS1 proteins from H5N1 strains are able to exacerbate pro-inflammatory cytokine production (TNF- $\alpha$  and IFN- $\beta$ ), which generates an unbalanced immune response that often results in clinical complications (45,46).

### **Antivirals**

#### *M2 Channel Inhibitors*

M2 is a tetrameric integral membrane protein that possesses a short ectodomain, a transmembrane domain, and an endodomain that is acylated and phosphorylated. This protein is the main target for antiviral drugs of the adamantanes class, whose action consists in blocking the ion channel activity of M2. The adamantanes, amantadine and rimantadine, have a tricyclic structure and have antiviral properties against all subtypes of influenza A virus but not against influenza B or C viruses (47). Adamantanes have been used as first-choice antiviral drugs against outbreaks of influenza A viruses for many years. However, the viruses resistant to these drugs have been increasing globally; nowadays most seasonal H1N1 and H3N2 strains of human and swine origin show resistance to adamantanes including the influenza A (H1N1) 2009 viruses (29). Mutations in M2 associated with resistance to these drugs have been identified as single amino acid substitutions at positions 26, 27, 30, 31, and 34, located in the transmembrane region of the protein (48,49).

#### *Neuraminidase Inhibitors*

The NA of influenza viruses is a homotetrameric glycoprotein anchored by a fibrous stalk in the viral membrane. The protein possesses a globular head comprised of four monomers that constitute the active site composed of nine conserved residues. Its primary role in the infectious cycle is to liberate the viral progeny from infected cells. Its enzymatic activity catalyzes SA removal from its linkage to galactose, thereby destroying the receptor and allowing the virus to disseminate and infect other cells (2,6). Furthermore, this



protein is also the main target of the antiviral drugs zanamivir and oseltamivir. These drugs closely resemble the structure of the natural substrate of the neuraminidase and thus prevent the removal of the SA residue from the glycopeptide receptor by the viral neuraminidase. These inhibitors are active against both influenza A and B viruses.

Subtype-specific NA mutations conferring resistance to NA inhibitors have been reported during *in vitro* passages of the virus and in isolates from clinical cases, albeit at adamantane-much lower frequency than adamantane-resistant variants (50–53). However, a growing number of influenza A strains seems to be developing resistance to NA inhibitors; thus, seasonal H1N1 strains show resistance to oseltamivir attributable to a mutation at amino acid position 274, where a histidine amino acid residue is substituted by a tyrosine (54). Of particular interest, it has been demonstrated that ~95% of the circulating strains in the U.S. during the 2008–2009 winter season have proven to be resistant to oseltamivir (55,56). Of concern, although most strains of the characterized current pandemic virus have been shown to be sensitive to this drug, is that there are already some reports showing the emergence of oseltamivir-resistant influenza A (H1N1) 2009 viruses in various countries including Japan, Denmark, Holland, and the U.S.

#### Alternative Antivirals

The finding of resistant viruses to the antiviral drugs in use is not unexpected given the plasticity of the influenza viral genome; thus, the development of new strategies for the treatment of influenza infections is a must. Fortunately, there is a list of several experimental antivirals that are now at different stages of development. One of these new drugs is T-705, a nucleoside analog that inhibits the polymerase activity of influenza viruses A, B, and C (57). Phase II clinical trials with T-705 have been completed in Japan, and phase III trials are scheduled (40). Peramivir and CS-8958 (the pro-drug of R-125489) (58,59) are two NA inhibitors that are structurally different from SA that have been proven to effectively block the enzymatic activity of the NA of the highly pathogenic influenza H5N1 virus strains. These two NA-inhibitors are in phase II clinical trials (30). A recent study reports the isolation of a high affinity neutralizing antibody against the hemagglutinin ectodomain that binds to an epitope that is recalcitrant to neutralization escape, and represents a promising strategy for broad-spectrum protection against seasonal and pandemic influenza viruses (60).

#### Origin of the Genomic Segments of Pandemic Influenza Virus A (H1N1) 2009

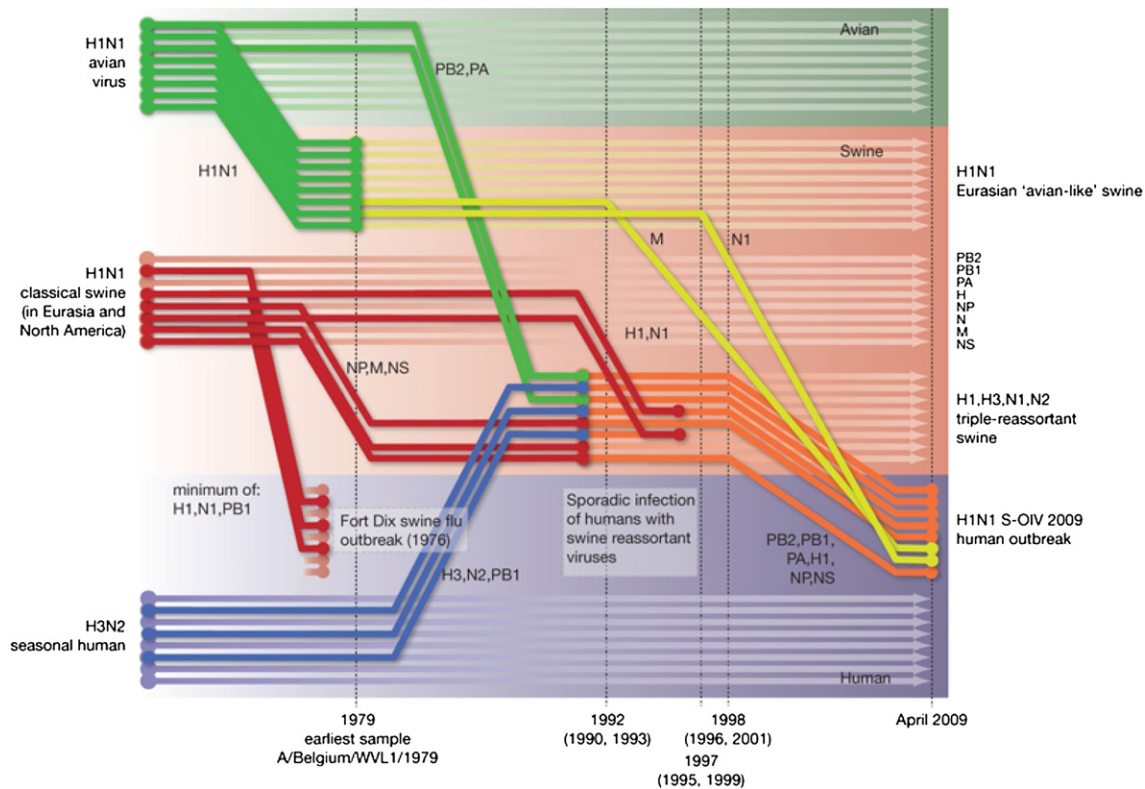
The lineages of influenza A viruses that circulate in swine in different regions of the world where they have been char-

acterized are very well defined. Classical swine H1N1 viruses have circulated in pigs in North America and other parts of the world since at least 1930 when they were detected for the first time (61). In 1998, a triple H3N2 reassortant virus was reported in the U.S. The genome of this virus was composed of genomic segments derived from three different origins: avian strains that circulate in wild aquatic birds of the U.S., classical swine H1N1 viruses, and seasonal H3N2 human viruses (62). This triple reassortant became established in the swine population of North America and has circulated widely since then. The triple reassortant viruses experienced subsequent reassortment of genes with the classical swine H1N1 to generate triple reassortant viruses belonging to subtypes H1N1 and H1N2 (63), all of which circulate currently in North America. In Mexico there is scarce information about the strains that circulate in pigs, and epidemiological surveys are urgently needed.

In Europe, an avian H1N1 virus was introduced directly into the pig population, and it became established in swine. This virus, known as avian-like swine H1N1, was first detected in 1979 and gradually replaced the classical swine H1N1 virus (64). The swine virus from avian origin exchanged genomic segments with the human H3N2 seasonal virus to generate H3N2 reassortants (65). In Asia, all the viruses described in North America and Europe circulate in the pig population: the classical swine H1N1 virus, the European avian-like swine H1N1, the European H3N2 reassortants, and the triple reassortant viruses H3N2, H1N1, and H1N2 described in North America (9,29,66). Of interest, the Euroasiatic avian-like swine H1N1 virus has not been described so far in the pig population of North America (9,29).

Phylogenetic studies comprising dozens of genomic sequences of strains of the virus A (H1N1) 2009 isolated from different parts of the world indicate that this virus is of swine origin because its genes have been present in viral lineages that have circulated in swine for at least the last 10 years (9,29,67). The immediate precursor of the current epidemic virus is one of the triple reassortant H1Nx viruses described above that acquired the NA and the M protein genes from an Euroasiatic avian-like swine H1N1 virus (9,29,67). The origin of each of the genes of this new influenza virus A (H1N1) 2009 virus is shown in Figure 4. The geographical location and time when the reassortment event that gave origin to the new epidemic virus is not known, although it is thought that the influenza A (H1N1) 2009 virus could have been circulating in pigs for years without being detected and before having been passed on to humans (9). Phylogenetic studies suggest that the transmission of the virus from pigs to humans occurred sometime during the second semester of 2008, several months before the alert of the outbreak was given (9,68).

The influenza A (H1N1) 2009 virus rapidly dispersed worldwide in a few months, and a second wave of virus



**Figure 4.** Reconstruction of the sequence of reassortment events leading up to the emergence of swine-origin influenza virus (S-OIV). Shaded boxes represent host species; avian (green), swine (red) and human (grey). Colored lines represent interspecies-transmission pathways of influenza genes. The eight genomic segments are represented as parallel lines in descending order of size. Dates marked with dashed vertical lines on ‘elbows’ indicate the mean time of divergence of the S-OIV genes from corresponding virus lineages. Reassortment events not involved with the emergence of human disease are omitted. Fort Dix refers to the last major outbreak of S-OIV in humans. The first triple-reassortant swine viruses were detected in 1998, but to improve clarity the origin of this lineage is placed earlier. Taken with permission from Smith et al. *Nature* 459, 1122–1125, 2009. Color version of this figure available online at [www.arcmedres.com](http://www.arcmedres.com).

infections is expected to occur in the cold months of the 2009–2010 season in the northern hemisphere. It is important to keep a close epidemiological surveillance of the influenza A viruses that circulate in the human population to detect virus strains that might become resistant to the available antiviral drugs, as well as to detect viral strains with increase virulence either by antigenic shift or antigenic drift. In this regard, the transmission of the influenza A (H1N1) 2009 virus from humans to swine (Argentina, Australia, Canada, Ireland, and Norway) and turkeys (Chile) that has been reported in several countries is of concern because these viruses could eventually return to humans after having experienced reassortment events in pigs with avian or swine viruses. Thus, it is also very important that close epidemiological surveillance be also carried out in pig and bird populations.

#### Acknowledgments

This work was partially supported by grant #CB-2006 #54492 from CONACyT-Mexico. MEZ and MDSR were supported by scholarships from CONACyT-Mexico.

#### References

1. Strauss J, Strauss E. Minus-strand RNA viruses. In: *Viruses and Human Disease*. 2<sup>nd</sup> ed. Oxford, UK: Elsevier; 2008. pp. 137–175.
2. Wright P, Neumann G, Kawaoka Y. Orthomyxoviruses. In: Knipe D, Howley P, eds. *Fields Virology*. 5<sup>th</sup> ed. Philadelphia: Lippincott Williams & Wilkins; 2006. pp. 1691–1740.
3. Fouchier RA, Munster V, Wallensten A, et al. Characterization of a novel influenza A virus hemagglutinin subtype (H16) obtained from black-headed gulls. *J Virol* 2005;79:2814–2822.
4. Webster RG, Bean WJ, Gorman OT, et al. Evolution and ecology of influenza A viruses. *Microbiol Rev* 1992;56:152–179.
5. Bouvier NM, Palese P. The biology of influenza viruses. *Vaccine* 2008;26(Suppl 4):D49–D53.
6. Palese P, Shaw M. Orthomyxoviridae: the viruses and their replication. In: Knipe D, Howley P, eds. *Fields Virology*. 5<sup>th</sup> ed. Philadelphia: Lippincott Williams & Wilkins; 2006. pp. 1647–1689.
7. Both GW, Siegman LJ, Bellamy AR, et al. Coding assignment and nucleotide sequence of simian rotavirus SA11 gene segment 10: location of glycosylation sites suggests that the signal peptide is not cleaved. *J Virol* 1983;48:335–339.
8. Drake JW. Rates of spontaneous mutation among RNA viruses. *Proc Natl Acad Sci USA* 1993;90:4171–4175.
9. Smith GJ, Bahl J, Vijaykrishna D, et al. Dating the emergence of pandemic influenza viruses. *Proc Natl Acad Sci USA* 2009;106:11709–11712.

10. Suzuki Y, Nei M. Simulation study of the reliability and robustness of the statistical methods for detecting positive selection at single amino acid sites. *Mol Biol Evol* 2002;19:1865–1869.
11. Scholtissek C, Rohde W, Von Hoyningen V, et al. On the origin of the human influenza virus subtypes H2N2 and H3N2. *Virology* 1978;87:13–20.
12. Holmes EC, Ghedin E, Miller N, et al. Whole-genome analysis of human influenza A virus reveals multiple persistent lineages and reassortment among recent H3N2 viruses. *PLoS Biol* 2005;3: e300.
13. Orlich M, Gottwald H, Rott R. Nonhomologous recombination between the hemagglutinin gene and the nucleoprotein gene of an influenza virus. *Virology* 1994;204:462–465.
14. Suarez DL, Senne DA, Banks J, et al. Recombination resulting in virulence shift in avian influenza outbreak, Chile. *Emerg Infect Dis* 2004;10:693–699.
15. Boni MF, Zhou Y, Taubenberger JK, et al. Homologous recombination is very rare or absent in human influenza A virus. *J Virol* 2008;82:4807–4811.
16. He CQ, Xie ZX, Han GZ, et al. Homologous recombination as an evolutionary force in the avian influenza A virus. *Mol Biol Evol* 2009;26:177–187.
17. Ma W, Lager KM, Vincent AL, et al. The Role of Swine in the Generation of Novel Influenza Viruses. *Zoonoses Public Health*; 2009.
18. Pappas C, Aguilar PV, Basler CF, et al. Single gene reassortants identify a critical role for PB1, HA, and NA in the high virulence of the 1918 pandemic influenza virus. *Proc Natl Acad Sci USA* 2008;105:3064–3069.
19. Lycett SJ, Ward MJ, Lewis FI, et al. Detection of mammalian virulence determinants in highly pathogenic avian influenza H5N1 viruses: multivariate analysis of published data. *J Virol* 2009;83:9901–9910.
20. Itoh Y, Shinya K, Kiso M, et al. In vitro and in vivo characterization of new swine-origin H1N1 influenza viruses. *Nature* 2009;460:1021–1025.
21. Munster VJ, de Wit E, van den Brand JM, et al. Pathogenesis and transmission of swine-origin 2009 A (H1N1) influenza virus in ferrets. *Science* 2009;325:481–483.
22. Barnett JM, Cadman A, Burrell FM, et al. In vitro selection and characterisation of influenza B/Beijing/1/87 isolates with altered susceptibility to zanamivir. *Virology* 1999;265:286–295.
23. Connor RJ, Kawaoka Y, Webster RG, et al. Receptor specificity in human, avian, and equine H2 and H3 influenza virus isolates. *Virology* 1994;205:17–23.
24. Naeve CW, Hinshaw VS, Webster RG. Mutations in the hemagglutinin receptor-binding site can change the biological properties of an influenza virus. *J Virol* 1984;51:567–569.
25. Rogers GN, Paulson JC. Receptor determinants of human and animal influenza virus isolates: differences in receptor specificity of the H3 hemagglutinin based on species of origin. *Virology* 1983;127:361–373.
26. Glaser L, Stevens J, Zamarin D, et al. A single amino acid substitution in 1918 influenza virus hemagglutinin changes receptor binding specificity. *J Virol* 2005;79:11533–11536.
27. Stevens J, Blixt O, Glaser L, et al. Glycan microarray analysis of the hemagglutinins from modern and pandemic influenza viruses reveals different receptor specificities. *J Mol Biol* 2006;355:1143–1155.
28. Tumpey TM, Maines TR, Van Hoeven N, et al. A two-amino acid change in the hemagglutinin of the 1918 influenza virus abolishes transmission. *Science* 2007;315:655–659.
29. Garten RJ, Davis CT, Russell CA, et al. Antigenic and genetic characteristics of swine-origin 2009 A(H1N1) influenza viruses circulating in humans. *Science* 2009;325:197–201.
30. Neumann G, Noda T, Kawaoka Y. Emergence and pandemic potential of swine-origin H1N1 influenza virus. *Nature* 2009;459:931–939.
31. Stevens J, Corper AL, Basler CF, et al. Structure of the uncleaved human H1 hemagglutinin from the extinct 1918 influenza virus. *Science* 2004;303:1866–1870.
32. Stegmann T, White JM, Helenius A. Intermediates in influenza induced membrane fusion. *EMBO J* 1990;9:4231–4241.
33. Horimoto T, Nakayama K, Smeekens SP, et al. Proprotein-processing endoproteases PC6 and furin both activate hemagglutinin of virulent avian influenza viruses. *J Virol* 1994;68:6074–6078.
34. Zamarin D, Garcia-Sastre A, Xiao X, et al. Influenza virus PB1–F2 protein induces cell death through mitochondrial ANT3 and VDAC1. *PLoS Pathog* 2005;1: e4.
35. McAuley JL, Hornung F, Boyd KL, et al. Expression of the 1918 influenza A virus PB1–F2 enhances the pathogenesis of viral and secondary bacterial pneumonia. *Cell Host Microbe* 2007;2:240–249.
36. Conenello GM, Zamarin D, Perrone LA, et al. A single mutation in the PB1–F2 of H5N1 (HK/97) and 1918 influenza A viruses contributes to increased virulence. *PLoS Pathog* 2007;3:1414–1421.
37. Gabriel G, Dauber B, Wolff T, et al. The viral polymerase mediates adaptation of an avian influenza virus to a mammalian host. *Proc Natl Acad Sci USA* 2005;102:18590–18595.
38. Steel J, Lowen AC, Mubareka S, et al. Transmission of influenza virus in a mammalian host is increased by PB2 amino acids 627K or 627E/701N. *PLoS Pathog* 2009;5:e1000252.
39. Hatta M, Hatta Y, Kim JH, et al. Growth of H5N1 influenza A viruses in the upper respiratory tracts of mice. *PLoS Pathog* 2007;3:1374–1379.
40. Massin P, van der Werf S, Naffakh N. Residue 627 of PB2 is a determinant of cold sensitivity in RNA replication of avian influenza viruses. *J Virol* 2001;75:5398–5404.
41. Van Hoeven N, Pappas C, Belser JA, et al. Human HA and polymerase subunit PB2 proteins confer transmission of an avian influenza virus through the air. *Proc Natl Acad Sci USA* 2009;106:3366–3371.
42. Gabriel G, Herwig A, Klenk HD. Interaction of polymerase subunit PB2 and NP with importin alpha1 is a determinant of host range of influenza A virus. *PLoS Pathog* 2008;4:e11.
43. Seo SH, Hoffmann E, Webster RG. Lethal H5N1 influenza viruses escape host anti-viral cytokine responses. *Nat Med* 2002;8:950–954.
44. Jackson D, Hossain MJ, Hickman D, et al. A new influenza virus virulence determinant: the NS1 protein four C-terminal residues modulate pathogenicity. *Proc Natl Acad Sci USA* 2008;105:4381–4386.
45. Jiao P, Tian G, Li Y, et al. A single-amino-acid substitution in the NS1 protein changes the pathogenicity of H5N1 avian influenza viruses in mice. *J Virol* 2008;82:1146–1154.
46. Li Z, Jiang Y, Jiao P, et al. The NS1 gene contributes to the virulence of H5N1 avian influenza viruses. *J Virol* 2006;80:11115–11123.
47. Varghese JN, Laver WG, Colman PM. Structure of the influenza virus glycoprotein antigen neuraminidase at 2.9 Å resolution. *Nature* 1983;303:35–40.
48. Deyde VM, Xu X, Bright RA, et al. Surveillance of resistance to adamantanes among influenza A(H3N2) and A(H1N1) viruses isolated worldwide. *J Infect Dis* 2007;196:249–257.
49. Schmidtke M, Zell R, Bauer K, et al. Amantadine resistance among porcine H1N1, H1N2, and H3N2 influenza A viruses isolated in Germany between 1981 and 2001. *Intervirology* 2006;49:286–293.
50. Abed Y, Nehme B, Baz M, et al. Activity of the neuraminidase inhibitor A-315675 against oseltamivir-resistant influenza neuraminidases of N1 and N2 subtypes. *Antiviral Res* 2008;77:163–166.
51. Baz M, Abed Y, Boivin G. Characterization of drug-resistant recombinant influenza A/H1N1 viruses selected in vitro with peramivir and zanamivir. *Antiviral Res* 2007;74:159–162.
52. Gubareva LV, Bethell R, Hart GJ, et al. Characterization of mutants of influenza A virus selected with the neuraminidase inhibitor 4-guanidino-Neu5Ac2en. *J Virol* 1996;70:1818–1827.
53. Gubareva LV, Kaiser L, Hayden FG. Influenza virus neuraminidase inhibitors. *Lancet* 2000;355:827–835.
54. Colman PM. New antivirals and drug resistance. *Annu Rev Biochem* 2009;78:95–118.
55. Dharan NJ, Gubareva LV, Meyer JJ, et al. Infections with oseltamivir-resistant influenza A(H1N1) virus in the United States. *JAMA* 2009;301:1034–1041.



56. Poland GA. Valuing influenza vaccine: medical, economic, and social benefits. *Clin Infect Dis* 2009;48:299–301.
57. Furuta Y, Takahashi K, Fukuda Y, et al. In vitro and in vivo activities of anti-influenza virus compound T-705. *Antimicrob Agents Chemother* 2002;46:977–981.
58. Babu YS, Chand P, Bantia S, et al. BCX-1812 (RWJ-270201): discovery of a novel, highly potent, orally active, and selective influenza neuraminidase inhibitor through structure-based drug design. *J Med Chem* 2000;43:3482–3486.
59. Yamashita M, Tomozawa T, Kakuta M, et al. CS-8958, a prodrug of the new neuraminidase inhibitor R-125489, shows long-acting anti-influenza virus activity. *Antimicrob Agents Chemother* 2009;53:186–192.
60. Sui J, Hwang WC, Perez S, et al. Structural and functional bases for broad-spectrum neutralization of avian and human influenza A viruses. *Nat Struct Mol Biol* 2009;16:265–273.
61. Shope R. The etiology of swine influenza. *Science* 1931;73: 215–215.
62. Webby RJ, Swenson SL, Krauss SL, et al. Evolution of swine H3N2 influenza viruses in the United States. *J Virol* 2000;74:8243–8251.
63. Vincent AL, Lager KM, Ma W, et al. Evaluation of hemagglutinin subtype 1 swine influenza viruses from the United States. *Vet Microbiol* 2006;118:212–222.
64. Pensaert M, Ottis K, Vandeputte J, et al. Evidence for the natural transmission of influenza A virus from wild ducts to swine and its potential importance for man. *Bull World Health Organ* 1981;59:75–78.
65. Brown IH. The epidemiology and evolution of influenza viruses in pigs. *Vet Microbiol* 2000;74:29–46.
66. Jung K, Song DS. Evidence of the co-circulation of influenza H1N1, H1N2 and H3N2 viruses in the pig population of Korea. *Vet Rec* 2007; 161:104–105.
67. Trifonov V, Khiabani H, Rabadan R. Geographic dependence, surveillance, and origins of the 2009 influenza A (H1N1) virus. *N Engl J Med* 2009;361:115–119.
68. Fraser C, Donnelly CA, Cauchemez S, et al. Pandemic potential of a strain of influenza A (H1N1): early findings. *Science* 2009;324: 1557–1561.
69. Kawaoka Y, Webster RG. Sequence requirements for cleavage activation of influenza virus hemagglutinin expressed in mammalian cells. *Proc Natl Acad Sci USA* 1988;85:324–328.
70. Subbarao EK, Kawaoka Y, Murphy BR. Rescue of an influenza A virus wild-type PB2 gene and a mutant derivative bearing a site-specific temperature-sensitive and attenuating mutation. *J Virol* 1993;67:7223–7228.
71. Gabriel G, Abram M, Keiner B, et al. Differential polymerase activity in avian and mammalian cells determines host range of influenza virus. *J Virol* 2007;81:9601–9604.
72. Chen W, Calvo PA, Malide D, et al. A novel influenza A virus mitochondrial protein that induces cell death. *Nat Med* 2001;7: 1306–1312.
73. Katz JM, Lu X, Tumpey TM, et al. Molecular correlates of influenza A H5N1 virus pathogenesis in mice. *J Virol* 2000;74:10807–10810.
74. Subbarao K, Shaw MW. Molecular aspects of avian influenza (H5N1) viruses isolated from humans. *Rev Med Virol* 2000;10:337–348.
75. Gubareva LV, Robinson MJ, Bethell RC, et al. Catalytic and framework mutations in the neuraminidase active site of influenza viruses that are resistant to 4-guanidino-Neu5Ac2en. *J Virol* 1997;71:3385–3390.
76. Yamada S, Suzuki Y, Suzuki T, et al. Haemagglutinin mutations responsible for the binding of H5N1 influenza A viruses to human-type receptors. *Nature* 2006;444:378–382.
77. Hatta M, Gao P, Halfmann P, et al. Molecular basis for high virulence of Hong Kong H5N1 influenza A viruses. *Science* 2001;293:1840–1842.
78. Wu WL, Chen Y, Wang P, et al. Antigenic profile of avian H5N1 viruses in Asia from 2002 to 2007. *J Virol* 2008;82:1798–1807.
79. Gubareva LV. Molecular mechanisms of influenza virus resistance to neuraminidase inhibitors. *Virus Res* 2004;103:199–203.
80. Deyde VM, Nguyen T, Bright RA, et al. Detection of molecular markers of antiviral resistance in influenza A (H5N1) viruses using a pyrosequencing method. *Antimicrob Agents Chemother* 2009;53:1039–1047.
81. Seo SH, Hoffmann E, Webster RG. The NS1 gene of H5N1 influenza viruses circumvents the host anti-viral cytokine responses. *Virus Res* 2004;103:107–113.
82. Solorzano A, Webby RJ, Lager KM, et al. Mutations in the NS1 protein of swine influenza virus impair anti-interferon activity and confer attenuation in pigs. *J Virol* 2005;79:7535–7543.
83. Min JY, Li S, Sen GC, et al. A site on the influenza A virus NS1 protein mediates both inhibition of PKR activation and temporal regulation of viral RNA synthesis. *Virology* 2007;363:236–243.

Article

## Clinical Characteristics and Genetic Variability of Human Rhinovirus in Mexico

Adriana Landa-Cardena <sup>1,†</sup>, Jaime Morales-Romero <sup>1,†</sup>, Rebeca García-Roman <sup>1</sup>, Ana Georgina Cobián-Güemes <sup>2</sup>, Ernesto Méndez <sup>2,‡</sup>, Cristina Ortiz-Leon <sup>1</sup>, Felipe Pitalúa-Cortés <sup>3</sup>, Silvia Ivonne Mora <sup>2</sup> and Hilda Montero <sup>1,\*</sup>

<sup>1</sup> Instituto de Salud Pública, Universidad Veracruzana, Av. Luis Castelazo Ayala s/n, Col. Industrial Ánimas, 91190, Xalapa, Veracruz, México; E-Mails: nita\_18lc@hotmail.com (A.L.-C.); jamorales@uv.mx (J.M.-R.); rebgarcia@uv.mx (R.G.-R.); cortiz@uv.mx (C.O.-L.)

<sup>2</sup> Instituto de Biotecnología, Universidad Nacional Autónoma de México, Av. Universidad 2001, 62210, Cuernavaca, Morelos, México; E-Mails: acobian@ibt.unam.mx (A.G.C.-G.); sivonnemor@gmail.com (S.I.M.)

<sup>3</sup> Hospital Civil ‘Luis F. Nachón’, Secretaría de Salud, Pedro Rendón 1, Esq. Nicolás Bravo, 91000, Xalapa, Veracruz, México; E-Mail: pitalua-felipe1@hotmail.com (F.P.-C.)

† These authors contributed equally to this work.

‡ Deceased.

\* Author to whom correspondence should be addressed; E-Mail: hmontero@uv.mx (H.M.); Tel.: +52-22-88-41-89-00, ext. 13323; Fax: +52-22-88-41-89-35.

Received: 15 November 2011; in revised form: 17 January 2012 / Accepted: 17 January 2012 /

Published: 25 January 2012

---

**Abstract:** Human rhinovirus (HRV) is a leading cause of acute respiratory infection (ARI) in young children and infants worldwide and has a high impact on morbidity and mortality in this population. Initially, HRV was classified into two species: HRV-A and HRV-B. Recently, a species called HRV-C and possibly another species, HRV-D, were identified. In Mexico, there is little information about the role of HRV as a cause of ARI, and the presence and importance of species such as HRV-C are not known. The aim of this study was to determine the clinical characteristics and genetic variability of HRV in Mexican children. Genetic characterization was carried out by phylogenetic analysis of the 5'-nontranslated region (5'-NTR) of the HRV genome. The results show that the newly identified HRV-C is circulating in Mexican children more frequently than HRV-B but not

as frequently as HRV-A, which was the most frequent species. Most of the cases of the three species of HRV were in children under 2 years of age, and all species were associated with very mild and moderate ARI.

**Keywords:** HRV; HRV-C; wheezing; genotyping

---

## 1. Introduction

Infants are one of the populations affected most by acute respiratory infection (ARI). The causes and severity of ARIs have been studied in different populations, and human rhinovirus (HRV) was one of the viruses most frequently associated with ARIs and could be the cause of more than 50% of ARI cases in some pediatric populations [1,2]. In most cases, HRV is associated with self-limiting illness with clinical characteristics similar to those of other viral infections such as fever, rhinorrhea, cough, and wheezing. However, HRV infections could cause complications such as otitis media, bronchiolitis, and pneumonia, which could require hospitalization [1,3,4]. The impact of HRV is not only on ARI; there is evidence that moderate to severe infection caused by HRV is a risk factor for wheezing illnesses and the development of asthma in stages posterior to infection [4]. Therefore, this virus is an important pathogen in public health.

HRV is an enterovirus that belongs to the family *Picornaviridae*. HRV is an enveloped virus and has a single strand of positive-sense RNA genome of approximately 7.2 kb in length. Serology analysis has detected 99 serotypes of HRV, and nucleotide sequencing analysis of the total and partial HRV genome has been used to classify this virus too. On the basis of that analysis, approximately 150 strains have been found and, until recently, were grouped into two species: HRV-A and HRV-B. Recently, HRV-C, a new genotype, was identified, and another possible species, HRV-D, has been proposed [5–8]. After the identification of HRV-C, many epidemiologic studies were initiated in different regions of the world to determine the importance of the new HRV species as a cause of ARI. Despite little published information, some data show that HRV-C types have been found in cases with a more severe disease [9,10]. However, this information is not clear, because some studies show that HRV-C could be a cause of mild or moderate disease in the infant population [11].

The importance of HRV as a causal agent of ARI in Mexican children has been little studied. This study describes the frequency and clinical characteristics of HRV infection in children under 6 years old and, on the basis of 5'-nontranslated region (5'-NTR) sequencing, determines the species circulating in this population.

## 2. Results

### 2.1. HRV Detection

A screening of 124 nasal swab samples was carried out. In all of the samples, the presence of the HRV-RNA genome was examined by reverse transcription-polymerase chain reaction (RT-PCR) by using primer pairs that encompassed the 400-bp fragment of the 5'-NTR in accordance with a previous report [12]. The samples were obtained from children younger than 72 months old (6 years), the

average age was  $18 \pm 15.6$  months, and the 75% of them were not more than 25 months old. Of the total samples, 21 were HRV-positive (21 out of 124, 16.9%). However, most of the HRV-positive cases were in the first or second year of life (18 out of 21, 85.7%); the age range was 2 to 48 months.

## 2.2. Signs and Symptoms

Clinical manifestations of HRV-positive children were compared with those of HRV-negative children, and the analysis is summarized in Table 1. Note that there are missing data; those data were not found in the medical record of the child or the parents did not provide them. In general, statistically significant differences were not found in the symptoms of the HRV-positive and HRV-negative children. Wheezing, a lower respiratory tract symptom, was found in 42.9% of the HRV-positive cases, and the severity of signs and symptoms of children with wheezing and HRV was not different to a statistically significant degree from that of children without HRV.

**Table 1.** Clinical characteristics of children with human rhinovirus infection compared with those without this virus.

	HRV-positive N = 21	HRV-negative N = 103	OR (95% CI)	P value
Age in months, median (Min-Max)	15.0 (2.0 to 48.0)	15.0 (0.12 to 67.9)	-	0.99
Younger than 2 years old, n (%)	17 (81.0)	67 (65.0)	2.3 (0.7 to 8.7)	0.16
Female, n (%)	12 (57.1)	47 (45.6)	1.59 (0.6 to 4.5)	0.34
History of asthma, n (%)	0 (0.0)	0 (0.0)	---	---
Previous allergic diseases, n (%)	0 (0.0)	9 (8.7)	---	0.36
Clinical symptoms, n/N (%) [missing values]				
Rhinorrhea	20/21 (95.2) [0]	90/99 (90.9) [4]	2.0 (0.2 to 44.5)	0.99
Cough	19/21 (90.5) [0]	89/95 (93.7) [8]	0.64 (0.1 to 5.0)	0.64
Tearing	6/21 (28.6) [0]	42/101 (41.6) [2]	0.56 (0.2 to 1.7)	0.13
Sore eyes	5/19 (26.3) [2]	25/98 (25.5) [5]	1.04 (0.3 to 3.5)	0.99
Fever $>37.5$ °C	5/21 (23.8) [0]	33/100 (33.0) [3]	0.63 (0.2 to 2.1)	0.41
Diarrhea	5/21 (23.8) [0]	19/102 (18.6) [1]	1.4 (0.4 to 4.7)	0.56
Wheezing	9/21 (42.9) [0]	55/102 (53.9) [1]	0.6 (0.2 to 1.8)	0.36
Dyspnea	11/21 (52.4) [0]	50/102 (49.0) [1]	1.1 (0.4 to 3.2)	0.78

n/N = subjects with the characteristic of interest/data available. Proportions were compared by using the chi-square or Fisher exact test. Medians were compared by using the Mann-Whitney *U* test. CI, confidence interval; HRV, human rhinovirus; Min-Max, minimum and maximum values; OR, odds ratio.

To analyze the ARI severity caused by HRV, we used a classification that groups the severity of infection into four types: very mild, mild, moderate, and severe [13]. On the basis of the symptoms, 42.85% of the HRV-positive cases were associated with very mild ARI, 52.38% with mild ARI, and 4.76% with moderate ARI. Only one HRV-positive child required hospitalization (1 out of 21, 4.76%).

### 2.3. HRV Genotyping

For the genotyping of the 21 HRV-positive samples, the products of the 5'-NTR and VP4/VP2 regions amplified by RT-PCR, which have been used previously for the genotyping of this virus [12,14], were sequenced. The 400-bp amplified fragment of 5'-NTR was sequenced in all positive samples of HRV. However, only the 549-bp amplified fragment of VP4/VP2 was detected and sequenced in 7 of 21 HRV-positive samples.

The Basic Local Alignment Search Tool (BLAST) was used to conduct a comparative analysis between all of the sequences we obtained and the reported sequences in GenBank. All of the sequences of the 5'-NTR region were used for a phylogenetic analysis (Figure 1). One sample was not included in the construction of a phylogenetic tree, because only 150 nucleotides were identified; however, its sequence was similar to reported sequences, ensuring that the sample corresponded to the HRV-C species. The phylogenetic analysis shown in Figure 1 indicates that the viruses of the three species of HRV—A, B, and C—are circulating in the population of children analyzed. Of the three species of HRV, the A species was the most frequent (61.9%), followed by HRV-C (28.57%) and HRV-B (9.52%). The genotyping of the seven samples obtained by the VP4/VP2 method turned out to be the same as the one obtained by the 5'-NTR method, and so the results were confirmed.

After the species were found, we analyzed a possible relation between severities of the ARI and the HRV species. The results did not show a statistically significant association between severity and HRV species. However, HRV-A was present in very mild ARI in 35.7% of cases, mild ARI in 57.1% of cases, and moderate ARI in 7.1% of cases. HRV-B was associated only with very mild ARI in 100% of cases, and the newly identified HRV-C was found in very mild and mild ARI in 60% and 40% of cases, respectively.

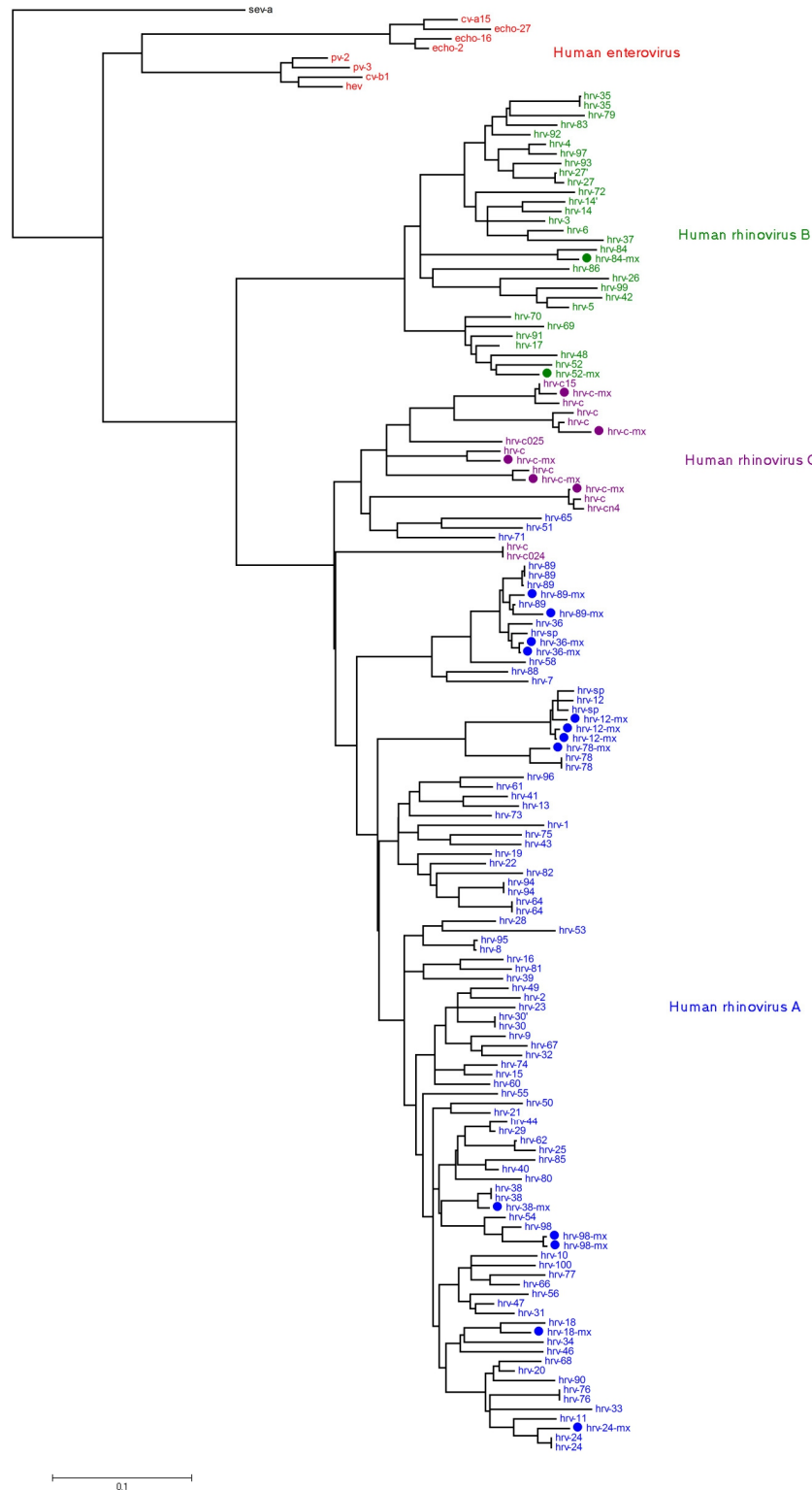
## 3. Experiment

### 3.1. Study Design

A cross-sectional study was carried out in children who were under 6 years old and who had ARI during one of two periods. Children in the first period, from November to December 2008, were treated at the 'Dr. Gastón Melo' medical unit, and children in the second period, from December 2010 to September 2011, were recruited from the 'Dr. Luis F. Nachón' hospital. Both institutions are located in Xalapa, Veracruz, México, which has a latitude of 19°30'43"N and an altitude of 1,427 m. The study was approved by the ethics committees of the health department of Veracruz state (Secretaría de Salud de Veracruz) and the 'Dr. Luis F. Nachón' hospital. The parents of recruited infants in this study provided written informed consent for a sample to be taken and analyzed. The inclusion criteria were age of fewer than 6 years and presentation with an ARI. A structured questionnaire was used to obtain the following information: age, sex, symptoms, asthma and allergy/atopy status, and ARI severity. Severity of illness was divided into four grades—very mild (upper respiratory tract signs or symptoms are present), mild (lower respiratory tract signs or symptoms are present but hospitalization was not required), moderate (lower respiratory tract signs or symptoms are present and hospitalization was required using oxygen saturations in air smaller than 93% on pulse oximetry), and severe (the same

characteristics to moderate but the child required oxygen saturations in air more than 93%)—in accordance with the classification by Bezerra and colleagues [13].

**Figure 1.** Phylogenetic tree of 5'-nontranslated region (5'-NTR) sequences of human rhinovirus (HRV) from nasal swabs. The phylogenetic tree was constructed by means of the maximum likelihood method. Reference strains of each species obtained from GenBank are indicated, and the species assignment in this study is indicated by circles. Accession numbers are JQ000010 to JQ000029.



### 3.2. Clinical Samples

One hundred twenty-four nasal swabs were collected with sterile rayon-tipped applicators (Puritan Medical Products Co. LLC, Guilford, ME, USA) in accordance with World Health Organization recommendations [15]. Some studies have concluded that there is no significant difference in specificity and sensitivity between nasal swab and nasopharyngeal aspirate specimens when PCR was used for respiratory virus detection [16,17]. The swab was placed into a tube that contained Leibovitz viral transport medium with 100 IU/mL penicillin (American Type Culture Collection, Manassas, VA, USA), 0.01 mg/mL streptomycin, and 0.25 µg/mL amphotericin B (American Type Culture Collection). The samples were maintained at 4 °C and transported to the virology laboratory of the Institute of Public Health at Veracruzana University (Instituto de Salud Pública, Universidad Veracruzana), where the samples were centrifuged at 7,000g for 5 minutes and the supernatant was stored at −80 °C until the analysis.

### 3.3. Nucleic Acid Extraction and RT-PCR Assays

Nucleic acids were extracted with a PureLink viral RNA/DNA kit (Invitrogen Corporation, Carlsbad, CA, USA) in accordance with the instructions of the manufacturer. A screening for detecting HRV was carried out by RT-PCR that amplifies the 5'-NTR of HRV genome by means of an AccessQuick™ RT-PCR System (Promega Corporation, Madison, WI, USA) in accordance with the instructions of the manufacturer and with previously described primers and methods [12], and 0.02 µg of HRV-RNA and 0.5 µM of each primer were used. Also, a 549-bp fragment was obtained by amplifying the VP4/VP2 region of the HRV genome in accordance with the previously published primers and protocol [14], except that the alignment was carried out at 42 °C for 1 minute. Finally, amplified PCR products were purified with a High Pure PCR Product Purification kit (Roche, Mannheim, Germany). In all cases, the reference strain HRV 17 (American Type Culture Collection) was used as the positive control.

### 3.4. Nucleotide Sequence Analysis

The purified PCR products were sequenced by the forward primer by means of the Taq FS dye terminator cycle sequencing fluorescence-based sequencing method and a Model 3730 sequencer obtained from PerkinElmer Life and Analytical Sciences, Inc. (Waltham, MA, USA) and Applied Biosystems (Life Technologies, Carlsbad, CA, USA).

### 3.5. Phylogenetic Analysis

The nucleotide sequences were subjected to an online National Center for Biotechnology Information BLAST search, based on this results the species and subtype were defined by a 95% identity or greater (Supplementary Table 1). The best hit was included in the phylogenetic analysis as well as all of the complete rhinovirus genomes available at GenBank. As outgroups, eight human enterovirus sequences were used. This dataset comprises 153 nucleotide sequences. Alignments of 360 nucleotides of the 5'-NTR regions of this dataset were performed with ClustalW, and the evolutionary history was inferred by using the maximum likelihood method based on the Kimura two-parameter model and

1,000 bootstrap replicates. Evolutionary analyses were conducted with MEGA version 5 (Center for Evolutionary Medicine and Informatics, The Biodesign Institute, Tempe, AZ, USA).

### 3.6. Statistical Analysis

Clinical symptoms, including disease severity, were compared by using the chi-square or Fisher exact test. Age was expressed as a median. Medians were compared by using the Mann-Whitney *U* test. Odds ratio and 95% confidence interval were used to identify the association of these variables with HRV infection. Disease severity was compared for each species. A *P* value not more than 0.05 was considered statistically significant. Statistical analysis was performed with SPSS software version 18 (SPSS Inc., Chicago, IL, USA).

## 4. Discussion and Conclusions

This study shows the frequency of HRV infection in a sample population of children who were under 6 years old and who presented with an ARI. This study is the first report of genetic variability of HRV in a population of children from Mexico (in this case, a part of eastern Mexico with a latitude of 19° 30' 43" N and an altitude of 1,427 m). It is important to mention that, during the collection period, HRV was a cause of ARI principally in children under 2 years old, and signs and symptoms were similar to those of ARIs caused by other pathogens, as was reported in another study [18]. HRV was found in only 17% of the ARI cases in our population. However, different frequencies of HRV as a cause of ARI have been reported [11,19,20]. We consider that the time of year of sampling and the population could be factors affecting the frequency of HRV in our study.

Our study included as many outpatient children as hospitalized children. This allowed us to determine the importance of HRV as a cause of morbidity and to measure the frequency with which HRV could cause symptomatic ARI but not necessarily require hospitalization. In addition, some studies suggest that HRV infection at a young age is a factor associated with the development of asthma or wheezing illnesses, not only when HRV is a cause of moderate or severe ARI [3,4] but also in outpatients who present with wheezing [21–23], and the allergic sensitization may play an important role [24]. In the population in this study, HRV was a cause of very mild to moderate infections. It is important to highlight that the severity of signs and symptoms of children with wheezing and HRV was not different to a statistically significant degree from that of children without HRV; this finding is in contrast to that of other studies, in which the clinical characteristics of children with HRV and wheezing were more severe [3]. However, in this study (in accordance with previous studies [21,22]), 42.9% of the children infected with HRV presented with wheezing. These data could suggest that about 45 of every 100 children infected with HRV could develop asthma or wheezing illnesses in posterior stages of infection. This could have a high impact on the quality of life of the infant and their parents and have economic consequences for the family and the health institutions. Given that HRV infection is a cause of morbidity and mortality and that the consequence of HRV infection is a chronic illness, this virus is a great problem in public health. However, the role of HRV as a risk factor to develop chronic illnesses in Mexican children needs to be studied further to establish control strategies to the infection like a rhinovirus vaccine or adequate and opportune treatments. To this end, Lehtinen and colleagues [23] found that the treatment with prednisolone may decrease the probability of



recurrent wheezing in a child infected with rhinovirus; this represents a possible alternative for decreasing the problems associated with rhinovirus infection.

In this study, other viral agents were not detected. However, there are reports that suggest a low probability of co-infections in people with HRV infection and other viral pathogens [25–27]. Likewise, there is evidence that HRV and virus co-infections could be associated with a more severe ARI [28–30]. In this study, our samples correspond only to very mild to moderate infections, suggesting that HRV single or in co-infection with other pathogens (not detected in this study) did not generate severe ARIs. It is very important that more studies be carried out to address this matter.

Few reports show the genetic variations of the principal viruses causing ARIs in Mexico. Thus, it is very important to determine the genetic properties of viruses causing ARIs in this population. Our study was limited to an analysis of sequences of specific regions of the HRV genome which have been used for genotyping of the virus, and our interest was to determine the species circulating in the infant population. With this aim, an analysis was carried out with specific regions of 5'-NTR and VP4/VP2 of HRV genome in HRV-positive samples. Of all of the HRV-positive samples examined by 5'-NTR amplification, only 7 were amplified in the VP4/VP2 region. In this study, the method of amplification of the 5'-NTR region was the most efficient, and this is consistent with previous observations [9]. The genotyping obtained by the VP4/VP2 method turned out to be the same as the one obtained by the 5'-NTR method.

The phylogenetic analysis shows that the three species—HRV-A, HRV-B, and the newly identified HRV-C—are circulating in children in the population we analyzed. HRV-C was found with a frequency less than that of HRV-A but greater than that of HRV-B and this also occurred in other populations studied [9]. Interestingly, previous reports suggest that the HRV-C species is more pathogenic since HRV-C infection could cause symptoms of greater severity than HRV-A and HRV-B infections could [9,10,31]. The relation between severity and HRV species associated with the ARI is not clear in this study. Although statistically significant differences were not found, our results, in contrast to those of other works [9,10,31], suggest that HRV-C is associated only with very mild and mild illnesses.

It is important to mention that the only case of an HRV-positive child who required hospitalization was an HRV-A species. Also, this child had a premature birth, a characteristic described as a possible risk factor of a more severe ARI [32], and this has happened with other respiratory viruses [33]. This suggests that the population characteristics such as the species of the virus causing the infection, plus other possible factors, could have an effect on the severity of the infection.

In conclusion, this study shows that the newly identified HRV-C species is present in the Mexican infant population. The data in this work establish a basis for carrying out new studies in a bigger sample of cases with longer periods of inclusion for a better definition of the molecular epidemiology of this important HRV, and all of its species, in Mexican children.

## Acknowledgments

We would like to express our gratitude to our late colleague Ernesto Méndez for his invaluable contributions to this work and for his friendship. He will be remembered with great fondness. We thank Patricia Meza and Alberto Navarrete for helping to secure the use of the facilities where this

study was carried out and Alfonso Hernandez and Martín Ortiz for collecting samples. This study was supported by grant 109397 from the National Council for Science and Technology-Mexico. Adriana Landa-Cardena and Martin Ortiz were supported by a CONACYT scholarship.

### Conflict of Interest

The authors declare no conflict of interest.

### References and Notes

1. Greenberg, S.B. Respiratory consequences of rhinovirus infection. *Arch. Intern. Med.* **2003**, *163*, 278–284.
2. Tregoning, J.S.; Schwarze, J. Respiratory viral infections in infants: Causes, clinical symptoms, virology, and immunology. *Clin. Microbiol. Rev.* **2010**, *23*, 74–98.
3. Jartti, T.; Korppi, M. Rhinovirus-induced bronchiolitis and asthma development. *Pediatr. Allergy Immunol.* **2011**, *22*, 350–355.
4. Gern, J.E. The ABCs of rhinoviruses, wheezing, and asthma. *J. Virol.* **2010**, *84*, 7418–7426.
5. Lee, W.M.; Kiesner, C.; Pappas, T.; Lee, I.; Grindle, K.; Jartti, T.; Jakiela, B.; Lemanske, R.F., Jr.; Shult, P.A.; Gern, J.E. A diverse group of previously unrecognized human rhinoviruses are common causes of respiratory illnesses in infants. *PLoS One* **2007**, *2*, doi:10.1371/journal.pone.0000966.
6. Palmenberg, A.C.; Spiro, D.; Kuzmickas, R.; Wang, S.; Djikeng, A.; Rathe, J.A.; Fraser-Liggett, C.M.; Liggett, S.B. Sequencing and analyses of all known human rhinovirus genomes reveal structure and evolution. *Science* **2009**, *324*, 55–59.
7. McIntyre, C.L.; McWilliam Leitch, E.C.; Savolainen-Kopra, C.; Hovi, T.; Simmonds, P. Analysis of genetic diversity and sites of recombination in human rhinovirus species C. *J. Virol.* **2010**, *84*, 10297–10310.
8. Simmonds, P.; McIntyre, C.; Savolainen-Kopra, C.; Tapparel, C.; Mackay, I.M.; Hovi, T. Proposals for the classification of human rhinovirus species C into genotypically assigned types. *J. Gen. Virol.* **2010**, *91*, 2409–2419.
9. Piralla, A.; Rovida, F.; Campanini, G.; Rognoni, V.; Marchi, A.; Locatelli, F.; Gerna, G. Clinical severity and molecular typing of human rhinovirus C strains during a fall outbreak affecting hospitalized patients. *J. Clin. Virol.* **2009**, *45*, 311–317.
10. Renwick, N.; Schweiger, B.; Kapoor, V.; Liu, Z.; Villari, J.; Bullmann, R.; Miething, R.; Briese, T.; Lipkin, W.I. A recently identified rhinovirus genotype is associated with severe respiratory-tract infection in children in Germany. *J. Infect. Dis.* **2007**, *196*, 1754–1760.
11. Smuts, H.E.; Workman, L.J.; Zar, H.J. Human rhinovirus infection in young African children with acute wheezing. *BMC Infect. Dis.* **2011**, *11*, doi:10.1186/1471-2334-11-65.
12. Kiang, D.; Kalra, I.; Yagi, S.; Louie, J.K.; Boushey, H.; Boothby, J.; Schnurr, D.P. Assay for 5' noncoding region analysis of all human rhinovirus prototype strains. *J. Clin. Microbiol.* **2008**, *46*, 3736–3745.

13. Bezerra, P.G.; Britto, M.C.; Correia, J.B.; Duarte Mdo, C.; Fonceca, A.M.; Rose, K.; Hopkins, M.J.; Cuevas, L.E.; McNamara, P.S. Viral and atypical bacterial detection in acute respiratory infection in children under five years. *PLoS One* **2011**, *6*, doi:10.1371/journal.pone.0018928.
14. Mizuta, K.; Hirata, A.; Suto, A.; Aoki, Y.; Ahiko, T.; Itagaki, T.; Tsukagoshi, H.; Morita, Y.; Obuchi, M.; Akiyama, M.; *et al.* Phylogenetic and cluster analysis of human rhinovirus species A (HRV-A) isolated from children with acute respiratory infections in Yamagata, Japan. *Virus Res.* **2010**, *147*, 265–274.
15. World Health Organization (WHO). WHO Guidelines for the collection of human specimens for laboratory diagnosis of avian influenza infection. Available online: [http://www.who.int/influenza/human\\_animal\\_interface/virology\\_laboratories\\_and\\_vaccines/guidelines\\_collection\\_h5n1\\_human\\_s/en/index.html#](http://www.who.int/influenza/human_animal_interface/virology_laboratories_and_vaccines/guidelines_collection_h5n1_human_s/en/index.html#) (accessed on 7 September 2009).
16. Heikkinen, T.; Marttila, J.; Salmi, A.A.; Ruuskanen, O. Nasal swab versus nasopharyngeal aspirate for isolation of respiratory viruses. *J. Clin. Microbiol.* **2002**, *40*, 4337–4339.
17. Sung, R.Y.T.; Chan, P.K.S.; Choi, K.C.; Li, A.M.C.; Nelson, E.A.S. A comparative study of nasopharyngeal aspirate and nasal swab specimens for the diagnosis of acute viral respiratory infection. *Hong Kong Med. J.* **2009**, *15*, 24–27.
18. Piotrowska, Z.; Vazquez, M.; Shapiro, E.D.; Weibel, C.; Ferguson, D.; Landry, M.L.; Kahn, J.S. Rhinoviruses are a major cause of wheezing and hospitalization in children less than 2 years of age. *Pediatr. Infect. Dis. J.* **2009**, *28*, 25–29.
19. Seah, S.G.; Lim, E.A.; Kok-Yong, S.; Liaw, J.C.; Lee, V.; Kammerer, P.; Metzgar, D.; Russell, K.L.; Tan, B.H. Viral agents responsible for febrile respiratory illnesses among military recruits training in tropical Singapore. *J. Clin. Virol.* **2010**, *47*, 289–292.
20. Pogka, V.; Kossivakis, A.; Kalliaropoulos, A.; Moutousi, A.; Sgouras, D.; Panagiotopoulos, T.; Chrousos, G.P.; Theodoridou, M.; Syriopoulou, V.P.; Mentis, A.F. Respiratory viruses involved in influenza-like illness in a Greek pediatric population during the winter period of the years 2005–2008. *J. Med. Virol.* **2011**, *83*, 1841–1848.
21. Jackson, D.J.; Gangnon, R.E.; Evans, M.D.; Roberg, K.A.; Anderson, E.L.; Pappas, T.E.; Printz, M.C.; Lee, W.M.; Shult, P.A.; Reisdorf, E.; *et al.* Wheezing rhinovirus illnesses in early life predict asthma development in high-risk children. *Am. J. Respir. Crit. Care Med.* **2008**, *178*, 667–672.
22. Lemanske, R.F., Jr.; Jackson, D.J.; Gangnon, R.E.; Evans, M.D.; Li, Z.; Shult, P.A.; Kirk, C.J.; Reisdorf, E.; Roberg, K.A.; Anderson, E.L.; *et al.* Rhinovirus illnesses during infancy predict subsequent childhood wheezing. *J. Allergy Clin. Immunol.* **2005**, *116*, 571–577.
23. Lehtinen, P.; Ruohola, A.; Vanto, T.; Vuorinen, T.; Ruuskanen, O.; Jartti, T. Prednisolone reduces recurrent wheezing after a first wheezing episode associated with rhinovirus infection or eczema. *J. Allergy Clin. Immunol.* **2007**, *119*, 570–575.
24. Jackson, D.J.; Evans, M.D.; Gangnon, R.E.; Tisler, C.J.; Pappas, T.E.; Lee, W.M.; Gern, J.E.; Lemanske, R.F., Jr. Evidence for a causal relationship between allergic sensitization and rhinovirus wheezing in early life. *Am. J. Respir. Crit. Care Med.* **2011** Sep 29, [Epub ahead of print].

25. Greer, R.M.; McErlean, P.; Arden, K.E.; Faux, C.E.; Nitsche, A.; Lambert, S.B.; Nissen, M.D.; Sloots, T.P.; Mackay, I.M. Do rhinoviruses reduce the probability of viral co-detection during acute respiratory tract infections? *J. Clin. Virol.* **2009**, *45*, 10–15.
26. Wang, Z.; Malanoski, A.P.; Lin, B.; Long, N.C.; Leski, T.A.; Blaney, K.M.; Hansen, C.J.; Brown, J.; Broderick, M.; Stenger, D.A.; *et al.* Broad spectrum respiratory pathogen analysis of throat swabs from military recruits reveals interference between rhinoviruses and adenoviruses. *Microb. Ecol.* **2010**, *59*, 623–634.
27. Casalegno, J.S.; Ottmann, M.; Duchamp, M.B.; Escuret, V.; Billaud, G.; Frobert, E.; Morfin, F.; Lina, B. Rhinoviruses delayed the circulation of the pandemic influenza A (H1N1) 2009 virus in France. *Clin. Microbiol. Infect.* **2010**, *16*, 326–329.
28. Calvo, C.; Garcia-Garcia, M.L.; Blanco, C.; Vazquez, M.C.; Frias, M.E.; Perez-Brena, P.; Casas, I. Multiple simultaneous viral infections in infants with acute respiratory tract infections in Spain. *J. Clin. Virol.* **2008**, *42*, 268–272.
29. Cilla, G.; Onate, E.; Perez-Yarza, E.G.; Montes, M.; Vicente, D.; Perez-Trallero, E. Viruses in community-acquired pneumonia in children aged less than 3 years old: High rate of viral coinfection. *J. Med. Virol.* **2008**, *80*, 1843–1849.
30. Jartti, T.; Lee, W.M.; Pappas, T.; Evans, M.; Lemanske, R.F., Jr.; Gern, J.E. Serial viral infections in infants with recurrent respiratory illnesses. *Eur. Respir. J.* **2008**, *32*, 314–320.
31. Miller, E.K.; Khuri-Bulos, N.; Williams, J.V.; Shehabi, A.A.; Faouri, S.; Al Jundi, I.; Chen, Q.; Heil, L.; Mohamed, Y.; Morin, L.L.; *et al.* Human rhinovirus C associated with wheezing in hospitalised children in the Middle East. *J. Clin. Virol.* **2009**, *46*, 85–89.
32. Van Piggelen, R.O.; van Loon, A.M.; Krediet, T.G.; Verboon-Macielek, M.A. Human rhinovirus causes severe infection in preterm infants. *Pediatr. Infect. Dis. J.* **2010**, *29*, 364–365.
33. Gooch, K.L.; Notario, G.F.; Schulz, G.; Gudkov, K.M.; Buesch, K.; Khong, H.; Campbell, A. Comparison of risk factors between preterm and term infants hospitalized for severe respiratory syncytial virus in the Russian Federation. *Int. J. Womens Health* **2011**, *3*, 133–138.



## Characterization of an influenza A virus in Mexican swine that is related to the A/H1N1/2009 pandemic clade

Marina Escalera-Zamudio<sup>a</sup>, Georgina Cobián-Güemes<sup>a</sup>, María de los Dolores Soto-del Río<sup>a</sup>, Pavel Isa<sup>a</sup>, Iván Sánchez-Betancourt<sup>b</sup>, Aurora Parissi-Crivelli<sup>c</sup>, María Teresa Martínez-Cázares<sup>c</sup>, Pedro Romero<sup>a</sup>, Lauro Velázquez-Salinas<sup>d</sup>, Belem Huerta-Lozano<sup>d</sup>, Martha Nelson<sup>e</sup>, Hilda Montero<sup>f</sup>, Pablo Vinuesa<sup>g</sup>, Susana López<sup>a</sup>, Carlos F. Arias<sup>a,\*</sup>

<sup>a</sup> Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, Mexico

<sup>b</sup> Facultad de Medicina Veterinaria y Zootecnia, Universidad Nacional Autónoma de México, D.F., Mexico

<sup>c</sup> Laboratorio Estatal de Salud Pública, Servicios de Salud de Veracruz, Veracruz, Mexico

<sup>d</sup> Comisión Mexico-Estados Unidos para la Prevención de la Fiebre Aftosa y otras Enfermedades Exóticas de los Animales, D.F., Mexico

<sup>e</sup> Fogarty International Center, National Institutes of Health, Bethesda, Maryland, United States

<sup>f</sup> Instituto de Salud Pública, Universidad Veracruzana, Xalapa, Veracruz, Mexico

<sup>g</sup> Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, Mexico

### ARTICLE INFO

#### Article history:

Received 25 January 2012

Returned to author for revisions

30 July 2012

Accepted 1 August 2012

Available online 24 August 2012

#### Keywords:

Influenza

Pandemics

Swine

Genetic diversity

Evolution

La Gloria

Mexico

### ABSTRACT

In the spring of 2009, swine-origin influenza H1N1pdm09 viruses caused the first influenza pandemic of this century. We characterized the influenza viruses that circulated early during the outbreak in Mexico, including one newly sequenced swine H1N1pdm09 virus and three newly sequenced human H1N1pdm09 viruses that circulated in the outbreak of respiratory disease in La Gloria, Veracruz. Phylogenetic analysis revealed that the swine isolate (A/swine/Mexico/4/2009) collected in April 2009 is positioned in a branch that is basal to the rest of the H1N1pdm09 clade in two (NP and PA) of the eight single-gene trees. In addition, the concatenated HA–NA and the complete whole-genome trees also showed a basal position for A/swine/Mexico/4/2009. Furthermore, this swine virus was found to share molecular traits with non-H1N1pdm09 H1N1 viral lineages. These results suggest that this isolate could potentially be the first one detected from a sister lineage closely related to the H1N1pdm09 viruses.

© 2012 Elsevier Inc. All rights reserved.

### Introduction

Influenza viruses are one of the most important pathogens of humans, causing seasonal epidemics and occasional pandemics. Three major influenza pandemics were recorded in the 20<sup>th</sup> century occurring in 1918, 1957, and 1968. During March and early April 2009 the first cases of an A/H1N1 virus were detected in Mexico and the United States, and the virus rapidly spread worldwide by human-to-human transmission, causing the first influenza pandemic of the 21<sup>st</sup> century (Dawood et al., 2009). Genetic characterization and phylogenetic analyses of these viruses showed that the pandemic strain was from swine origin, and represented a novel reassortant virus containing the PB1, PB2, PA, HA, NP, and NS segments from North American triple

\* Correspondence to: Instituto de Biotecnología, Universidad Nacional Autónoma de México, Av. Universidad 2001, Colonia Chamilpa, Cuernavaca, Morelos 62210, Mexico. Fax: +52 777 3172399.

E-mail address: [arias@ibt.unam.mx](mailto:arias@ibt.unam.mx) (C.F. Arias).

reassortant swine viruses, while the NA and M genes were derived from the Eurasian swine lineage (Garten et al., 2009). Despite the large increase of influenza gene sequencing that has taken place in the last few years, the answer to when, where, and how the pandemic H1N1pdm09 virus arose in swine is still unknown, as the vast majority of publicly available sequences are from human and avian viruses, and not from swine strains.

Mexico, and particularly the town of La Gloria, Veracruz, was identified as the place where the pandemic virus emerged (Fraser et al., 2009; Lopez-Cervantes et al., 2009; MMWR, 2009b). A 5-year old boy from this town was the first confirmed human case of H1N1pdm09 influenza virus, detected during an outbreak of influenza-like respiratory illness that initiated in early March 2009. This case was for some time dubbed by the media as “patient zero”, although a later report confirmed that a person from Mexico City developed symptoms on 11<sup>th</sup> March, and possibly there could still be earlier cases that had occurred during February in San Luis Potosí (Cohen, 2009). Early H1N1pdm09 virus infections were also detected in two children that had

become sick on late March in San Diego and Imperial Counties, in California (MMWR, 2009a).

The lack of systematic surveillance and sampling of influenza viruses in swine populations in North America, including Mexico, as well as worldwide, have impeded efforts to trace the geographical origin of the Eurasian and North American swine viruses that originally reassorted to generate the pandemic H1N1pdm09 strain. Previous analyses estimated 9–17 years (depending on the gene) of unsampled evolution between the H1N1pdm09 viruses and the most closely related influenza viruses in swine, which were isolated in Asia (Smith et al., 2009b). To date, Asia remains the only region where both Eurasian and North American swine lineages are known to co-circulate in swine. However, evolutionary analyses of human H1N1pdm09 viruses in Asia do not support an Asian origin of the 2009 pandemic (Lemey et al., 2009 PLoS Currents). Thus, it was of interest to characterize the influenza viruses that circulated in Mexican swine during the early stages of the pandemic, and compare their genomic sequences with other pandemic viruses isolated worldwide, including the early isolates from California.

Early after its emergence, the H1N1pdm09 virus was differentiated into two clusters (Fereidouni et al., 2009) and seven subclades (Nelson et al., 2009), although differences in the pathogenic or biological properties of these groups remains unknown. Nelson et al. reported that the early viruses from Mexico belonged to subclade 2, while those of California belonged to subclade 1 (Nelson et al., 2009). Viruses belonging to subclade 7, which include most of the H1N1pdm09 virus strains characterized from the spring 2009 wave in New York, have become the dominant strains in most places.

In this work, we report the genetic characterization of the H1N1pdm09 viruses that circulated in humans during the respiratory outbreak of March 2009 in La Gloria, and also of an H1N1pdm09 pandemic isolate from a pig in a farm located in central Mexico early during the outbreak. Phylogenetic analysis of the sequences obtained revealed that the swine isolate represents a basal branch to the 2009 pandemic clade and possesses ancestral molecular traits shared with viruses belonging to the human H1N1 seasonal lineage, but not with other 2009 pandemic viruses. The human H1N1pdm09 isolates from La Gloria shared a high sequence identity with early H1N1pdm09 isolates from California.

## Results

### Sequence analysis of H1N1/2009 viruses isolated from La Gloria and a pig farm

The outbreak of acute respiratory illness in La Gloria, Veracruz, was characterized by a large number of cases (616 or 28.5% of the population) reported during March 5th to April 10th. Initial testing of four respiratory specimens collected during this period identified two patients as positive for seasonal influenza A (H3N2), one for seasonal influenza B, and one patient for novel influenza A/H1N1 virus (MMWR, 2009b). However, recent data based on a serological census of La Gloria showed that the influenza-like outbreak was largely caused by the novel H1N1pdm09 virus, which infected about 24% of the population (unpublished data). Towards the end of the outbreak, in late March and early April, nasal-swab samples from ill patients were collected by the public health personnel of Veracruz. We isolated the H1N1pdm09 virus from three of those samples (La Gloria-3,-4, and -8) and determined their full genomic sequence using the Illumina's sequencing platform. Sequence comparisons showed that two (La Gloria-3 and -4) of the three viruses were identical at

**Table 1**  
Summary of the influenza A sequence reads obtained.

	Total number of reads <sup>a</sup>	Influenza reads <sup>b</sup>	Total bases covered <sup>c</sup>	Genome depth coverage
La Gloria-3	8,576,322	380,093 (4.43%)	13,592 (99.7%)	1006x
La Gloria-4	7,609,171	242,307 (3.18%)	13,556 (99.4%)	643x
La Gloria-8	10,421,316	44,998 (0.43%)	13,543 (99.3%)	120x
Swine-4	10,593,972	101,038 (0.95%)	13,564 (99.5%)	268x

<sup>a</sup> Number of good quality reads

<sup>b</sup> The percentage of influenza reads out of the total number of reads is shown in parenthesis

<sup>c</sup> The percentage of the influenza genome covered, using as reference an H1N1 genome of 1362 nt, is shown in parenthesis

the nucleotide level, and had only two amino acids differences with the third virus (La Gloria-8). These viruses also shared a high identity with the early California isolates (> 99.8%, 8 amino acid changes), and in general with all the so far characterized pandemic viruses.

Of interest, all three isolates from La Gloria presented the same single nucleotide polymorphisms (SNPs), one in the PB2 gene and three in the NP gene, suggesting a single introduction event of H1N1pdm09 virus in La Gloria community (Table 2). Of these SNPs, three were silent, and only one led to an amino acid change (nucleotide position 301 in the NP open reading frame). The two SNPs in NP co-localized so close together that it was possible to analyze their co-segregation. In fact, SNPs at positions 1155 and 1167 gave rise to only two viral populations, as C-1155 always co-segregated with G-1167, while T-1155 co-segregated with A-1167. In all SNPs observed, the predominant nucleotide was the one present in that position in > 98% of pandemic 2009 virus in the Influenza Virus Research database available at GenBank (Table 2).

We also determined the complete nucleotide sequence of the genome of a H1N1pdm09 pandemic influenza virus isolated early during the pandemic from a pig with respiratory illness in a farm in central Mexico (A/swine/Mexico/4/2009) (Table 1). This viral isolate also presented SNPs in the PB2, PB1, HA, and NA genes (Table 2), although at nucleotide positions different from the La Gloria viruses. Out of these, two SNPs were silent, and similar to La Gloria viruses the predominant nucleotide at each nucleotide position was the one present in > 82% of pandemic 2009 viruses (Table 2).

As mentioned, early after their emergence, the H1N1pdm09 virus differentiated into seven different subclades (Nelson et al., 2009). Based on the presence of genetic markers for each subclade and from the phylogenetics trees, La Gloria isolates grouped within subclade 2, while the A/swine/Mexico/4/2009 isolate could not be classified into any of the previously reported subclades (Table 3).

### Phylogenetic analysis of swine-4 and La Gloria viruses

The phylogenetic relationship of the A/swine/Mexico/4/2009 and La Gloria viruses with other H1N1pdm09 viruses was determined using single gene coding-region alignments under the maximum likelihood criteria (ML). Interestingly, we found that for the NP, PA, and concatenated HA-NA gene trees, the A/swine/Mexico/4/2009 isolate represented a basal branch to the rest of the pandemic viruses with significant branch support (aLRT and bootstrap  $\geq$  0.74, 69; Fig. 1, Supplementary Figs. S-3

and S-8). However, in the evolutionary analysis of the remaining viral genes, the phylogenetic relationship between the swine isolate and other pandemic viruses could not be fully resolved (PB2, HA, NA, and NS gene trees; [Supplementary Figs S-1, S-4, S-5, and S-7](#)), or the swine isolate clustered within the pandemic clade (PB1 and M gene trees; [Supplementary Figs. S-2 and S-6](#)).

To improve the resolution, we performed a phylogenetic analysis using the concatenated whole-genome coding regions, as no genetic shift events have been detected among the recently circulating swine and pandemic lineages. This strategy allowed us to integrate the phylogenetic signal of every viral gene in a single tree. The ML tree obtained shows that the A/swine/Mexico/4/2009 virus is positioned on a basal branch to the rest of pandemic sequences, with significant support value (aLRT and bootstrap  $\geq 0.92$ , 70; [Supplementary Fig. S-9](#)). As expected, La Gloria viruses grouped near other early human H1N1pdm09 viruses (as indicated by the collection date, showed on the right side of each virus name), such as A/California/04/2009 virus and the A/California/07/2009 vaccine strain, as it is clearly seen in the concatenated HA–NA gene tree (aLRT  $\geq 70$ ; [Supplementary Fig. S-8](#)).

**Table 2**  
Ambiguities (SNPs) at nucleotide positions observed in viral samples.

Sample	Gene	Nt <sup>a</sup>	Amb <sup>b</sup>	1 <sup>st</sup> choice <sup>c</sup>	2 <sup>nd</sup> choice <sup>c</sup>	H1N1/2009 <sup>d</sup>
La Gloria-3	PB2	99	R	G (461x)	A (338x)	G (98%)
	NP	301	R	G (196x)	A (109x)	G (98%)
	NP	1155	Y	C (928x)	T (383x)	C (98%)
	NP	1167	R	G (1056x)	A (504x)	G (98%)
La Gloria-4	PB2	99	R	G (162x)	A (104x)	G (98%)
	HA	1289	R	A (27x)	G (3x)	A (99%)
	NP	301	R	G (88x)	A (30x)	G (98%)
	NP	1155	Y	C (594x)	T (183x)	C (98%)
	NP	1167	R	G (748x)	A (336x)	G (98%)
La Gloria-8	PB2	99	R	G (44x)	A (24x)	G (98%)
	PB1	641	R	A (17x)	G (1x)	A (99%)
	NP	301	K	G (26x)	T (3x)	G (98%)
	NP	1155	Y	C (172x)	T (32x)	C (98%)
	NP	1167	R	G (204x)	A (54x)	G (98%)
Swine-4	PB2	1030	R	G (65x)	A (55x)	G (97%)
	PB1	1251	R	G (134x)	A (77x)	G (98%)
	HA	106	K	G (85x)	T (52x)	G (98%)
	HA	1161	K	T (158x)	G (17x)	T (99%)
	HA	1171	R	G (124x)	A (12x)	G(82%)
	NA	1326	Y	C (345x)	T (63x)	C (99%)

<sup>a</sup> Nucleotide numbering begins at each open reading frame.

<sup>b</sup> Ambiguity observed, R=A/G, Y=C/T, K=T/G.

<sup>c</sup> 1st choice and 2nd choice nucleotides observed in a given position. In brackets is number of times a given nucleotide was sequenced.

<sup>d</sup> Most common nucleotide in the position observed among H1N1pdm09 pandemic strains.

**Table 3**  
H1N1pdm09subclades defined by amino acid polymorphisms.<sup>a</sup>

	PB2	PB1	PA	HA	NP	NA	M1/2	NS1	NS2
<b>Subclade<sup>b</sup></b>									
<b>1</b>			S224*P	S100P*, A214T*, V338I*					
<b>2</b>			M581L*		T373I				
<b>3</b>									
<b>4</b>	V649I*	I667T		K2E, Q310H*	V100I	V106I			E63K
<b>5</b>					V100I*	V106I*, N248D			
<b>6</b>					V100I	V106I, N248D			
<b>7</b>				S220T	V100I*	V106I, N248D*		I123V	
<b>Swine-4</b>				214T					
<b>La Gloria 3/4, 8</b>			581L		373I				

<sup>a</sup> As defined in Nelson et al. (2009). The asterisk (\*) denotes changes that are present in a majority, but not all of the isolates from a subclade

<sup>b</sup> In the case of the A/swine/Mexico/4/2009 and La Gloria viruses, only the amino acids that correspond to a change associated with a given subclade are shown.

### Polymorphisms in the swine-4 virus

We further analyzed the genomic sequence of A/swine/Mexico/4/2009 to identify all non-synonymous changes in the eight viral genes that could differentiate this isolate from other pandemic viruses. All individual gene sequences available for non-pandemic H1N1 and for pandemic H1N1pdm09 viruses were analyzed separately in SeaView and JalView alignment editors, changes were detected manually, and frequencies were obtained in order to identify polymorphisms or mutations. We considered a mutation as a change present in less than 1% of the population, and a polymorphism as a change occurring in at least 1% of the population. We detected fourteen amino acid changes distributed along six genes in swine-4 that were different from other pandemic viruses: R126H and E636G in PB1; N222K and V407I in PA; P200S, A214T, and T220S in HA; K77R and D53E in NP; S35G, I106V, P126H, and D248N in NA; and V123I in NS1. The amino acid changes in the HA protein (positions 200, 214, and 220) are located in the receptor recognition region, and position 220 corresponds to the Ca antigenic site as in the HA1 numbering scheme (Lin et al., 2009). All other changes identified were neither located in a biologically relevant position in any of the viral proteins nor have been detected as being under positive selection in previous studies. Seven of the changes found in the swine-4 isolate were present in less than 1% of the pandemic and non-pandemic virus population, and were thus considered as mutations. Polymorphic sites HA T220S, NA I106V, and NS1 V123I are present in 13–30% of the pandemic virus population.

Of note, polymorphisms V407I (PA), D53E (NP), P200S (HA), and P126H (NA) in A/swine/Mexico/4/2009 are highly conserved among the non-pandemic A/H1N1 virus population, regardless their animal host, geographical origin or year of isolation (Table 4), occurring in 83 to >99% of the non-pandemic A/H1N1 sequences analyzed. On the other hand, while the four non-synonymous changes became fixed in most of the human seasonal and animal non-pandemic A/H1N1 viruses, these were absent in most of the pandemic viral population (Table 4). Few other pandemic viruses were found to have one or two of these markers; nevertheless, none was found to have all four changes.

The polymorphisms found in A/swine/Mexico/4/2009 do not seem to be an artifact derived from adaptation of the viruses to grow in eggs, since La Gloria viruses characterized in this work were also egg-grown but do not share these changes with the swine isolate. Furthermore, the mutations that lead to the adaptation of influenza viruses to grow in eggs have been well characterized before, and usually occur in HA and NA, altering the receptor-binding properties of the virus (Brookes et al., 2010; Gambaryan et al., 1999; Matrosovich et al., 2004; Mochalova et al., 2003). Furthermore, we determined the A/swine/Mexico/4/2009 HA and NA gene sequences by conventional PCR and Sanger







outbreak of March 2009 indicates that these viruses are very closely related to the strains detected about the same time (early April 2009) in California. Recent evolutionary analysis suggests that the 2009 pandemic viruses emerged after several reassortment events and circulated undetected in mammalian species over a period of years before they were introduced into humans (Smith et al., 2009a). It has been reported that the genetic segments of the H1N1pdm09 viruses had been circulating in swine for 9 to 17 years (depending on the gene) previous to the outbreak (Smith et al., 2009b), while the M and NA genes derived from Eurasian swine lineage had not been detected circulating in the American continent before (Smith et al., 2009b). In addition, it has been estimated that the virus responsible for the 2009 influenza pandemic was probably transmitted to humans during the second half of 2008, several months before its identification.

It is of interest that the D53E (NP) and V407I (PA) ancestral polymorphisms described in the A/swine/Mexico/4/2009 isolate are present in the same two viral genes whose phylogenetic analysis showed the A/swine/Mexico/4/2009 isolate to represent a basal branch to the rest of the pandemic viruses. These findings support the hypothesis that the A/swine/Mexico/4/2009 isolate has ancestral traits not shared with other pandemic viruses. On the other hand, in the evolutionary analysis of the other viral genes the swine isolate clustered with other 2009 pandemic viruses, and ancestral polymorphisms were absent, or in the case of the HA and NA genes (where polymorphisms P200S and P126H are located, respectively) the phylogenetic relationship could not be fully resolved and remained as a polytomy.

Detection of swine influenza viruses in Hong Kong that share 7 of the 8 genes with the A/H1N1pdm09 pandemic virus (Smith et al., 2009b) suggests that the Most-Recent Common Ancestor that originally re-assorted to generate the genetic composition of the H1N1pdm09 strain could have emerged and circulated, as stated, anywhere in the world. Our observations suggest that although the swine isolate is closely related to the H1N1pdm09 viruses, it could belong to a sister lineage (a similar virus to the H1N1pdm09 strain) that co-circulated in pigs early during the pandemic. Alternatively, the A/swine/Mexico/4/2009 isolate could represent a man-to-pig transmission event, as the pandemic virus was widespread in Mexico at the time that A/swine/Mexico/4/2009 was isolated (22nd of April, 2009). Although the whole-genome, HA–NA, PA, and NP trees indicate that A/swine/Mexico/4/2009 could be a sister lineage, the clustering of A/swine/Mexico/4/2009 within the pandemic clade for other gene segments clearly indicates that the ancestral origins of the H1N1pdm09 virus in swine cannot be fully resolved. Additional sequencing of swine influenza viruses in Mexico has not identified other A/swine/Mexico/4/2009-like or H1N1pdm09 viruses, raising further questions about the evolutionary origins of the H1N1pdm09 virus.

## Methods

### *Ethics statement*

Analysis of human clinical samples was approved by the Bioethics Committee of the Instituto de Biotecnología. The samples from patients with acute respiratory illness were collected by personnel of the Ministry of Health of the State of Veracruz, under the Mexican Official Norm NOM-024-SSA2-1994 for Prevention and Control of Acute Respiratory Infections in the Primary Attention of Health. Informed consent was not obtained as the analysis of samples for pathogens is part of the mandate of clinical testing for the public health agency of the State of Veracruz, and samples as well as demographic, clinical, and laboratory data were de-identified prior to analysis.

With regard to the animal sacrificed, approval by a review board was not necessary since sacrifice of the sick animal was based on the Mexican Official Norm NOM-033-ZOO-1995 for humanitarian sacrifice of animals, with the owner consent. In the case of campaigns to screen for diseases, there is a sample size calculated by the General Directorate of Epidemiology, and occasionally, like in this case, the sampling includes the collection of animal organs, in agreement with the animal farm owners.

### *Sample collection, viral isolation, and RNA extraction*

Samples from patients with acute respiratory illness collected on April 3<sup>rd</sup> 2009 during the outbreak of La Gloria that had been found positive for H1N1pdm09 by real-time PCR were obtained from the BSL-3 laboratory of the Ministry of Health of the State of Veracruz. Filtrates of nasal human samples or swine lung and trachea homogenates were inoculated and grown in embryonated chicken eggs in the Veterinary School of the National University of Mexico. The H1N1pdm09 virus could be isolated from three different patients from La Gloria (isolates 3, 4, and 8). The swine respiratory sample was collected on April 22<sup>nd</sup> 2009 in a farm in the State of Queretaro, Mexico. Out of 360 asymptomatic pigs in the farm, two showed high temperature (41 °C), were agitated and had nasal flow. One of the animals was sacrificed, and during the pathology studies a generalized pneumonia with abscesses was observed. Samples were taken from these two animals; nevertheless, influenza virus sequences were only detected in one of the two samples taken. The swine H1N1pdm09 viral sequence was detected by real-time PCR at the Mexico–United States Commission for the prevention of Foot and Mouth Disease and Other Exotic Animal Diseases (CPA), and confirmed at the Instituto de Biotecnología/UNAM, in Cuernavaca, Morelos, in BSL-2 facilities. Viral RNA from the viruses was extracted from allantoic liquid using either a Purelink viral RNA/DNA extraction kit (Invitrogen), with linear acrylamide (Ambion) as carrier (La Gloria viruses), or with Trizol and precipitated with isopropanol (swine virus).

### *Illumina library preparation, sequencing, and analysis*

Whole-genome sequencing was performed at the next-generation sequencing core facility of the National University of Mexico located at the Instituto de Biotecnología in Cuernavaca, Morelos. Two hundred base-pair-sized libraries, prepared using Illumina's mRNA-Seq 8 sample prep kit as suggested by the manufacturer, were loaded in separate lanes of a flow cell from a Genome Analyzer IIx (Illumina, San Diego, CA). Sequencing was performed through 36 cycles of single base pair extensions. Image analysis was carried out using Genome Analyzer Pipeline Version 1.4. Assembly of viral reads was carried out with MAQ (Mapping and Assembly with Qualities) (Li et al., 2008) version 0.7.1, using influenza A isolate A/Netherlands/602/2009 as the assembly reference genome (Genbank ID accession numbers for the eight genes: CY046940, CY046941, CY046942, CY039527, CY046943, CY039528, CY046944, CY046945). Coverage graphics were generated using R2.12 (Team, 2010). Any reads that did not match the reference genome were not considered for further analysis. From each sample we obtained between 7.6 and 10.5 million 36-nt long reads, out of which 0.4 to 4.4% corresponded to influenza H1N1pdm09 sequences (Table 1). In total, more than 99.3% of each viral genome was sequenced, with a 120x to 1006x coverage (Table 1). All samples showed homogeneous gene segment coverage (Supplementary Fig. S-10). All sequences reported in this work were submitted to GenBank, except for the A/Mexico/LaGloria-3/2009, as it was identical to the A/Mexico/LaGloria-4/2009 virus. ID accession numbers for A/Mexico/LaGloria-4/2009:

CY077600, CY077601, CY077602, CY077603, CY077604, CY077605, CY077606, CY077607; A/Mexico/LaGloria-8/2009; CY077608, CY077609, CY077610, CY077611, CY077612, CY077613, CY077614, CY077615; A/swine/Mexico/4/2009: CY053642, CY053643, CY053644, CY053645, CY053646, CY053647, CY053648, CY053649) and are publicly available at the National Center for Biotechnology (NCBI) Influenza Virus Resource (IVR). Single nucleotide polymorphisms (SNPs) were identified in positions that had at least 10x coverage using MAQ and analyzed by Mapview (Bao et al., 2009).

#### *Phylogenetic analysis of H1N1pdm09 viruses*

To better visualize the phylogenetic relationships of pandemic viruses and other swine influenza viruses, a data set of approximately 94 to 115 sequences (number depending on the gene, the lowest being for PA) from swine influenza viruses collected from 2000 to 2011 in Asia and America where selected from the Influenza Virus Resource-NCBI database available as of June 2012. Only full-length sequences representing a balanced selection of the diversity of recently circulating swine H1N1 viruses isolated worldwide were chosen. Triple-reassortant H1N2 viruses described previously as being the most closely related to the pandemic lineage were also included in all datasets, except for the NA alignment (Smith et al., 2009b). Since the PA gene segment of 2009 pandemic viruses is from avian origin, PA gene sequences from avian H1N1 influenza viruses collected from 2000 and later in Asia or America were also included in the PA dataset.

In the case of 2009 pandemic viruses selection, swine viruses with early and late collection dates from all over the world were included. For the A/swine/Mexico/4/2009 phylogenetic analysis, only human strains relevant to this analysis such as the vaccine strain A/California/07/2009, the A/California/04/2009 virus, and other Mexican isolates were included, together with the sequences reported in this work. For the La Gloria virus phylogenetic analysis, we added 32 selected human H1N1pdm09 viral sequences to the HA–NA concatenated gene alignment, originally composed of 115 sequences. Initial analysis using both local and global alignment algorithms (BLAST, MAFFT) showed that the percentage identity between all pandemic viruses, including the strains used in this study is very high in all genes (> 99.8% at the nucleotide level), so 100% identical sequences were eliminated. Distantly related viral lineages to the 2009 pandemic viruses that could reduce tree resolution (e.g. seasonal H1N1 human viruses) were excluded from all the datasets.

Individual coding regions from complete virus genomes were extracted from the database and aligned using MAFFT algorithm (Kato et al., 2009). ModelTest (Posada and Crandall, 1998) was applied to each set of sequences to determine the best-fit substitution model, and phylogenetic analysis was performed using single-gene or concatenated nucleotide coding regions under the Maximum Likelihood (ML) criterion using PhyML (Guindon and Gascuel, 2003). Individual ML trees were inferred under the GTR+G model, using the aLRT (SH-like) and bootstrap analysis implemented in PhyML 3.0 to evaluate branch support. Only values of aLRT (SH-like) and bootstrap near or above .70 or 70 were considered as significant. To evaluate the robustness of our trees, we built topologies under BioNJ alternatively (HKY, 100 bootstraps for branch support), and obtained the same results (data not shown). All trees were mid-point rooted for clarity purposes.

#### *Polymorphism identification and mapping*

Two different data sets were used to detect polymorphisms among different viral populations. The first set (set one) was used

to identify polymorphisms between our isolates and other H1N1pdm09 pandemic viruses, and it comprised amino acid sequence data for all pandemic viral genomes available in the IVR–NCBI database as of September 2010. The second set (set two) was used to detect polymorphisms between our isolates and non-pandemic H1N1 viruses; thus, it contained individual amino acid sequences for all viral genes corresponding to all non-pandemic A/H1N1 viruses, regardless their animal origin, collection place, and date, available in the IVR–NCBI database, also until September 2010. For both data sets, approximately 5700 individual full-length complete protein sequences were aligned by MAFFT as described above. The pandemic viral sequences reported in this work (La Gloria viruses and the A/swine/Mexico/4/2009 isolate) were included in the data sets and predominant nucleotides in SNPs positions were used. Variable sites were manually detected in La Gloria viruses and the swine isolate as compared to other viral pandemic sequences (set one alignment), or other non-pandemic H1N1 viruses (set two alignment) by using SeaView (Gouy et al., 2010) and JalView (Clamp et al., 2004) alignment editors. Polymorphism frequencies were calculated with SequenceLogo (Crooks et al., 2004).

#### **Acknowledgments**

We thank Ricardo Grande and Verónica Jiménez, of the next-generation sequencing core facility of the National University of Mexico for their technical support. We are grateful to all personnel from the molecular biology and virus isolation units of CPA for their work in the diagnosis of the Querétaro swine pandemic influenza virus. This work was supported by Grants 5549 and I0110/184/09 from the National Council for Science and Technology (Conacyt)-Mexico, and Grant PICOSI09–209 from the Instituto de Ciencia y Tecnología del Distrito Federal, Mexico. Marina Escalera-Zamudio, María de los Dolores Soto del Río y Georgina Cobián-Güemes were supported by a scholarship from CONACYT-Mexico.

#### **Appendix A. Supporting information**

Supplementary data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.virol.2012.08.003>.

#### **References**

- Bao, H., Guo, H., Wang, J., Zhou, R., Lu, X., Shi, S., 2009. MapView: visualization of short reads alignment on a desktop computer. *Bioinformatics* 25, 1554–1555.
- Brookes, S.M., Nunez, A., Choudhury, B., Matrosovich, M., Essen, S.C., Clifford, D., Slomka, M.J., Kuntz-Simon, G., Garcon, F., Nash, B., Hanna, A., Heegaard, P.M., Queguiner, S., Chiapponi, C., Bublot, M., Garcia, J.M., Gardner, R., Foni, E., Loeffen, W., Larsen, L., Van Reeth, K., Banks, J., Irvine, R.M., Brown, I.H., 2010. Replication, pathogenesis and transmission of pandemic (H1N1) 2009 virus in non-immune pigs. *PLoS One* 5 (e9068).
- Clamp, M., Cuff, J., Searle, S.M., Barton, G.J., 2004. The Jalview Java alignment editor. *Bioinformatics* 20, 426–427.
- Cohen, J., 2009. A New Patient Zero? *Science Insider*, AAAS.
- Crooks, G.E., Hon, G., Chandonia, J.M., Brenner, S.E., 2004. WebLogo: a sequence logo generator. *Genome Res.* 14, 1188–1190.
- Dawood, F.S., Jain, S., Finelli, L., Shaw, M.W., Lindstrom, S., Garten, R.J., Gubareva, L.V., Xu, X., Bridges, C.B., Uyeki, T.M., 2009. Emergence of a novel swine-origin influenza A (H1N1) virus in humans. *N. Engl. J. Med.* 360, 2605–2615.
- Fereidouni, S.R., Beer, M., Vahlenkamp, T., Starick, E., 2009. Differentiation of two distinct clusters among currently circulating influenza A(H1N1)v viruses. *Euro. Surveill.* 14 (46), pii: 19409. March–September 2009, 14, 2009 Nov 19.
- Fraser, C., Donnelly, C.A., Cauchemez, S., Hanage, W.P., Van Kerkhove, M.D., Hollingsworth, T.D., Griffin, J., Baggaley, R.F., Jenkins, H.E., Lyons, E.J., Jombart, T., Hinsley, W.R., Grassly, N.C., Balloux, F., Ghani, A.C., Ferguson, N.M., Rambaut, A., Pybus, O.G., Lopez-Gatell, H., Alpuche-Aranda, C.M., Chapela, I.B., Zavala, E.P., Guevara, D.M., Checchi, F., Garcia, E., Hugonnet, S., Roth, C., 2009. Pandemic

- potential of a strain of influenza A (H1N1): early findings. *Science* 324, 1557–1561.
- Gambaryan, A.S., Robertson, J.S., Matrosovich, M.N., 1999. Effects of egg-adaptation on the receptor-binding properties of human influenza A and B viruses. *Virology* 258, 232–239.
- Garten, R.J., Davis, C.T., Russell, C.A., Shu, B., Lindstrom, S., Balish, A., Sessions, W.M., Xu, X., Skepner, E., Deyde, V., Okomo-Adhiambo, M., Gubareva, L., Barnes, J., Smith, C.B., Emery, S.L., Hillman, M.J., Rivaller, P., Smagala, J., de Graaf, M., Burke, D.F., Fouchier, R.A., Pappas, C., Alpuche-Aranda, C.M., Lopez-Gatell, H., Olivera, H., Lopez, I., Myers, C.A., Faix, D., Blair, P.J., Yu, C., Keene, K.M., Dotson Jr., P.D., Boxrud, D., Sambol, A.R., Abid, S.H., St George, K., Bannerman, T., Moore, A.L., Stringer, D.J., Blevins, P., Demmler-Harrison, G.J., Ginsberg, M., Kriner, P., Waterman, S., Smole, S., Guevara, H.F., Belongia, E.A., Clark, P.A., Beatrice, S.T., Donis, R., Katz, J., Finelli, L., Bridges, C.B., Shaw, M., Jernigan, D.B., Uyeki, T.M., Smith, D.J., Klimov, A.I., Cox, N.J., 2009. Antigenic and genetic characteristics of swine-origin 2009 A(H1N1) influenza viruses circulating in humans. *Science* 325, 197–201.
- Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52, 696–704.
- Gouy, M., Guindon, S., Gascuel, O., 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* 27, 221–224.
- Katoh, K., Asimenos, G., Toh, H., 2009. Multiple alignment of DNA sequences with MAFFT. *Methods Mol. Biol.* 537, 39–64.
- Lemey, P., Suchard, M., Rambaut, A., 2009. Reconstructing the initial global spread of a human influenza pandemic: A Bayesian spatial–temporal model for the global spread of H1N1pdm. *PLoS Curr.* 1, RRN1031. [revised 2009 Sep 2].
- Li, H., Ruan, J., Durbin, R., 2008. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res.* 18, 1851–1858.
- Lin, T., Wang, G., Li, A., Zhang, Q., Wu, C., Zhang, R., Cai, Q., Song, W., Yuen, K.Y., 2009. The hemagglutinin structure of an avian H1N1 influenza A virus. *Virology* 392, 73–81.
- Lopez-Cervantes, M., Venado, A., Moreno, A., Pacheco-Dominguez, R.L., Ortega-Pierres, G., 2009. On the spread of the novel influenza A (H1N1) virus in Mexico. *J. Infect. Dev. Countries* 3, 327–330.
- Matrosovich, M.N., Matrosovich, T.Y., Gray, T., Roberts, N.A., Klenk, H.D., 2004. Human and avian influenza viruses target different cell types in cultures of human airway epithelium. *Proc. Natl. Acad. Sci. U S A* 101, 4620–4624.
- MMWR Morb Mortal Wkly Rep. 2009a. Swine Influenza A (H1N1) Infection in Two Children—Southern California, March–April 2009. *MMWR Morb Mortal Wkly Rep* 58, pp. 400–402.
- MMWR Morb Mortal Wkly Rep. 2009b. Update: Novel Influenza A (H1N1) Virus Infection—Mexico, March–May, 2009. *MMWR Morb Mortal Wkly Rep* 58, pp. 585–589.
- Mochalova, L., Gambaryan, A., Romanova, J., Tuzikov, A., Chinarev, A., Katinger, D., Katinger, H., Egorov, A., Bovin, N., 2003. Receptor-binding properties of modern human influenza viruses primarily isolated in Vero and MDCK cells and chicken embryonated eggs. *Virology* 313, 473–480.
- Nelson, M., Spiro, D., Wentworth, D., Beck, E., Fan, J., Ghedin, E., Halpin, R., Bera, J., Hine, E., Proudfoot, K., Stockwell, T., Lin, X., Griesemer, S., Kumar, S., Bose, M., Viboud, C., Holmes, E., Henrickson, K., 2009. The early diversification of influenza A/H1N1pdm. *PLoS Curr.* 1, RRN1126.
- Posada, D., Crandall, K.A., 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14, 817–818.
- Smith, G.J., Bahl, J., Vijaykrishna, D., Zhang, J., Poon, L.L., Chen, H., Webster, R.G., Peiris, J.S., Guan, Y., 2009a. Dating the emergence of pandemic influenza viruses. *Proc. Natl. Acad. Sci. U S A* 106, 11709–11712.
- Smith, G.J., Vijaykrishna, D., Bahl, J., Lycett, S.J., Worobey, M., Pybus, O.G., Ma, S.K., Cheung, C.L., Raghwani, J., Bhatt, S., Peiris, J.S., Guan, Y., Rambaut, A., 2009b. Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. *Nature* 459, 1122–1125.
- Team, R.D.C., 2010. R: A language and Environment for Statistical Computing, R Foundation for Statistical Computing Vienna.