



CENTRO DE INVESTIGACIÓN Y DE ESTUDIOS AVANZADOS DEL
INSTITUTO POLITÉCNICO NACIONAL

UNIDAD ZACATENCO

DEPARTAMENTO DE GENÉTICA Y BIOLOGÍA MOLECULAR

**“Genómica comparativa de fagos de *Pseudomonas*: mosaicismo observado
en cinco fagos novedosos”**

TESIS

Que presenta

M. en C. Victor Missael Flores López

Para obtener el grado de

DOCTOR EN CIENCIAS

EN LA ESPECIALIDAD DE GENÉTICA Y BIOLOGÍA MOLECULAR

Director de la tesis: Dr. Gabriel Guarneros Peña

Ciudad de México

Junio, 2017

El presente trabajo fue realizado con el apoyo de la beca otorgada por el Consejo Nacional de Ciencia y Tecnología con número de becario: 206969.

De igual manera, este trabajo formó parte del proyecto “Genómica estructural de las proteínas de los bacteriófagos” financiado por el Consejo Nacional de Ciencia y Tecnología con número de proyecto: 166814

Finalmente, el presente trabajo también contó con el apoyo del Instituto de Ciencia y Tecnología del Distrito Federal como parte del proyecto de Investigación en Ciencia Básica PICSA 11-107

Agradecimientos

El presente trabajo fue realizado gracias al apoyo incondicional de mi familia; gracias a la guía experta de mi director de tesis; gracias a las valiosas aportaciones de mi comité tutorial; gracias a la ayuda proporcionada por el cuerpo académico y administrativo del departamento de Genética y Biología Molecular de CINVESTAV Zacatenco; gracias a la convivencia diaria con mis compañeros y amigos del programa de posgrado; y al último pero no por ello menos importante, gracias al amor, al apoyo, a la guía, a la ayuda y a la amistad de Denisse Flores quien me motiva día a día a seguir queriendo ser el mejor.

El presente trabajo se realizó gracias a todos y cada uno de ustedes.

Contenido

Resumen.....	1
Abstract.....	2
Introducción.....	3
Antecedentes.....	11
Justificación.....	12
Objetivos.....	13
Materiales y métodos.....	14
Resultados.....	21
Discusión.....	33
Conclusiones.....	35
Perspectivas.....	37
Bibliografía.....	38
Anexos.....	43
Firmas del jurado.....	84

Resumen

Los bacteriófagos son virus que son capaces de infectar células bacterianas. Se cree que son las entidades biológicas más diversas y abundantes en el planeta. Consecuentemente, se cree que son uno de los principales reservorios de información genética y además se especula que podrían ser una de las principales fuentes de variación entre organismos relacionados debido a su interacción con ellos.

La diversidad y abundancia de los fagos se ve reflejada en sus genomas y mientras que algunos estudios describen genomas fágicos con elevados porcentajes de similitud a secuencias previamente reportadas, algunos otros muestran fagos altamente novedosos cuyas secuencias genómicas no guardan relaciones detectables con otros organismos.

En el presente trabajo se realizó la descripción genómica de los fagos de *Pseudomonas aeruginosa* PaMx11, PaMx25, PaMx28, PaMx42 y PaMx74 a través de la asignación de funciones a proteínas codificadas en sus genomas empleando criterios de similitud de secuencia, dominios conservados y homología estructural.

Adicionalmente, mediante el uso de diversas estrategias de genómica comparativa se lograron determinar relaciones de similitud existentes entre los genomas de fagos de *Pseudomonas aeruginosa* y otros genomas fágicos. A través de estas estrategias se logró la detección de conexiones entre grupos que se consideraban previamente no relacionados. Siendo dichas conexiones, genomas fágicos con características que los asocian a más de un grupo de forma simultánea.

El mosaicismo inherente de los genomas fágicos es evidente en el genoma de los fagos PaMx25, PaMx28 y PaMx74, sin embargo, las conexiones observadas entre los fagos pertenecientes a distintos grupos únicamente pudieron ser detectadas mediante la inclusión de genomas de reciente descripción.

Teniendo esto en consideración, se cree que el número creciente de genomas fágicos depositados en bases de datos, permitirá la detección de una mayor cantidad de conexiones entre grupos de fagos que se creían previamente no relacionados con la posibilidad de formar un espacio genómico continuo.

Abstract

Bacteriophages are viruses that are able to infect bacterial cells. It is believed that they are the most diverse and abundant biological entities on the planet. Hence, it is also believed that phages represent one of the most important reservoirs of genetic information, and likely, one of the main sources of genetic variation observed among genomes of related bacterial species.

The high abundance and diversity of phages is observed through their genomes, and whereas some studies describe phage genomes displaying high nucleotide identities to known sequences in databases; other studies present novel sequences that appear not to be related to known genomes.

In this study, the genomic characterization of the *Pseudomonas* phages PaMx11, PaMx25, PaMx28, PaMx42 and PaMx74 was performed through functional assignment of the putative proteins encoded in their genomes by means of several computational strategies.

Additionally, by using different approaches of comparative genomics, several relationships were detected among the genomes of phages infecting *Pseudomonas* and phages infecting other bacterial species. At least three groups that previously appeared to be unrelated were connected by means of the inclusion of the PaMx25, PaMx28 and PaMx74 genomes. The inherent mosaicism of phage genomes was evident in such genomes, nevertheless, the connections that bridged the previously unrelated phage groups could only be detected by including recently described genomes.

Taking this into account, it is believed that as more phage genomes are described, the number of connections bridging apparently unrelated phage groups will also increase, and eventually the formation of a genomic continuum could be possible.

1. Introducción

1.1 Los bacteriófagos

1.1.1 Descripción.

Los bacteriófagos, también llamados fagos, son virus que infectan células procarióticas, al igual que otros virus, son estructuras proteicas que envuelven un ácido nucleico, DNA o RNA de cadena doble o sencilla en una topología circular o lineal. La morfología de los fagos es muy diversa, sin embargo se cree que los fagos con cola son los más abundantes [Guttmann, Raya & Kutter, 2005].

1.1.2. La infección fágica.

La infección fágica (Figura 1) comienza cuando una partícula viral reconoce un receptor en la superficie bacteriana mediante estructuras específicas denominadas anti-receptores. Una vez que el fago ha llevado a cabo una interacción estable con la célula, algunas proteínas estructurales hidrolizan componentes de las barreras celulares para que el material genético del virus pueda ingresar a la célula hospedera [Guttmann, Raya & Kutter 2005].

En el proceso de introducción del material genético del fago están involucrados mecanismos bioquímicos y fisicoquímicos tales como motores enzimáticos y la presión hidrostática de la cápside viral [Grayson & Molineux, 2007].

Una vez dentro de la célula, el material genético del fago puede completar un ciclo replicativo o ser degradado por la maquinaria enzimática del hospedero. Los bacteriófagos pueden replicarse por lo menos de dos formas distintas. Una de ellas es la infección lítica en donde se producen partículas virales en el interior de la bacteria y su liberación ocasiona la lisis celular; la otra estrategia de replicación corresponde a la infección lisogénica, en la cual no se lleva a cabo la producción de partículas virales, sin embargo, el material genético de los fagos se integra al genoma de la célula hospedera replicándose a la par del genoma bacteriano [Hyman & Abedon, 2010].

El estado de lisogénia es estable, sin embargo es reversible y puede dar lugar al ciclo lítico bajo las circunstancias ambientales, energéticas y metabólicas adecuadas. El DNA viral integrado en el cromosoma de la bacteria se denomina profago y cada vez que la bacteria se divide, el profago se replica junto con el DNA del hospedero. Este arreglo ofrece al virus un mecanismo alternativo para hacer más copias de su genoma [Casjens 2003].

A través del ciclo lisogénico, es posible que tanto las bacterias como los fagos adquieran material genético uno del otro. Hay casos en donde los fagos permanecen en la bacteria como profagos, y estos pueden fijarse en el cromosoma mediante la acumulación de mutaciones deletéreas y rearrreglos, convirtiéndose así, en profagos crípticos [Casjens 2003]. Los profagos crípticos son secuencias remanentes de uno o varios eventos lisogénicos en los genomas bacterianos. Por otro lado, también existen fagos que encapsulan DNA que proviene del hospedero, y al infectar a una segunda célula introducen material genético adicional que podría conferir al hospedero nuevos genotipos que conllevan a la expresión de nuevos fenotipos, a este proceso se le conoce como transducción.

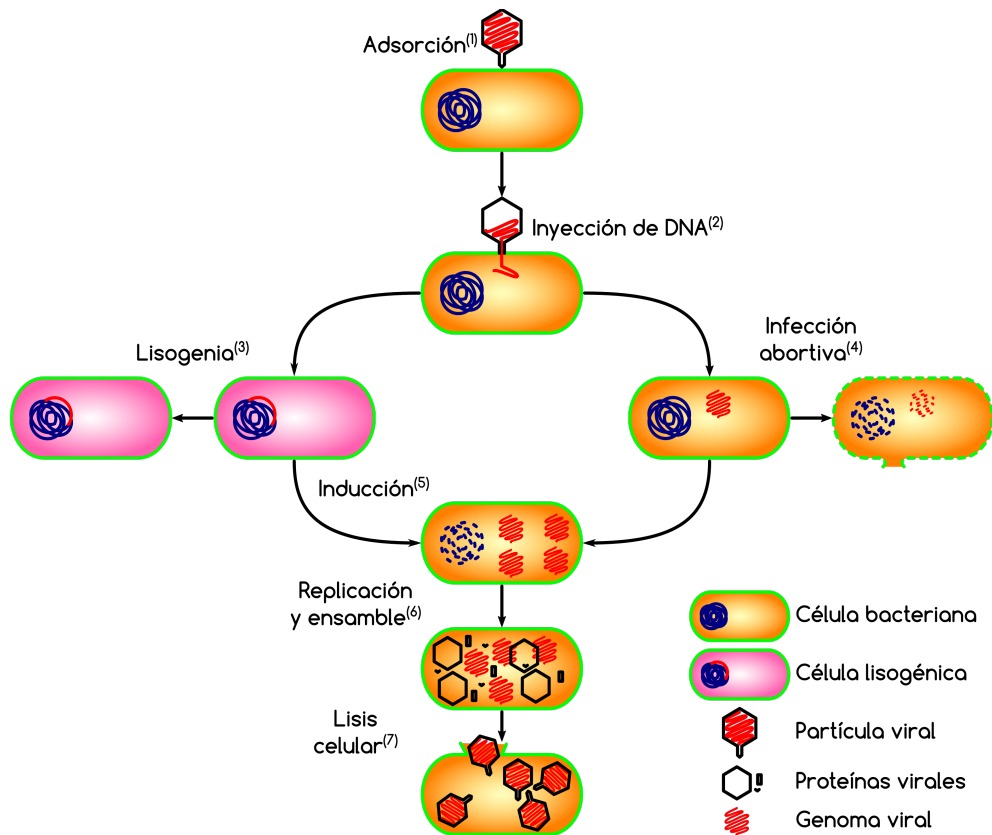


Figura 1: La infección fágica comienza con la adsorción de una partícula viral a la superficie de una célula bacteriana⁽¹⁾. Una vez que la partícula viral reconoce su receptor, el fago introduce su material genético al citoplasma⁽²⁾. El DNA viral una vez dentro de la célula puede integrarse en el cromosoma bacteriano dando lugar al ciclo lisogénico, replicándose sincronizadamente con el cromosoma bacteriano⁽³⁾. Alternativamente, las células bacterianas pueden hacer frente a la infección fágica mediante un proceso llamado infección abortiva, en el que la célula entra en muerte celular programada, terminando prematuramente el ciclo replicativo del fago⁽⁴⁾. El estado lisogénico en las bacterias puede ser muy estable, sin embargo, bajo condiciones de stress, la célula bacteriana activa la respuesta SOS, la cual puede inducir el desarrollo del ciclo lítico⁽⁵⁾. En el ciclo lítico se producen las proteínas necesarias para el ensamble de nuevas partículas virales, de igual manera, el material genético del fago se replica activamente⁽⁶⁾. Una vez producidas las nuevas partículas virales, enzimas líticas y proteínas transmembranales dan lugar a la lisis celular⁽⁷⁾, liberando más partículas virales al medio.

1.2. Bacteriófagos en el ambiente

A través de observaciones microscópicas y análisis metagenómicos, se ha propuesto que los fagos son las entidades biológicas más abundantes y diversas en el planeta [Bergh et al., 1989]. En general, se estima que existen aproximadamente 10^{31} partículas fágicas en la biosfera, superando a los procariotes en por lo menos un orden de magnitud [Brussow & Kutter, 2005].

Como consecuencia de la abundancia de los fagos, se ha estimado que globalmente pueden suscitarse hasta 10^{24} infecciones fágicas por segundo [Pedulla et al., 2003] trayendo como consecuencia que incluso los eventos que podrían considerarse raros ocurran con cierta frecuencia [Chibanni-Chennoufi et al., 2004].

Debido a su abundancia, los fagos tienen un efecto directo en las poblaciones bacterianas, y de manera igualmente significativa, en la movilización de materia orgánica en los ambientes donde se localizan [Proctor & Fuhrman 1990, Fuhrman 1992]. Por otro lado, el ciclo lisogénico, tiene diversas consecuencias, entre las que encontramos: movilización de material genético entre bacterias [Lindell et al., 2004. Beumer & Robinson, 2005], impacto en las poblaciones bacterianas mediante la inducción del ciclo lítico y adquisición de genes mediante la integración del genoma fágico. Debido a lo anterior, se cree que los fagos son uno de los principales motores de la evolución y diversidad bacteriana [Ohnishi et al., 2001; Casjens, 2003; Weinbauer & Rassoulzadegan, 2004].

1.3 Genómica de bacteriófagos

A pesar de que los bacteriófagos son altamente diversos a nivel genético [Breitbart et al., 2002], se ha visto que una fracción significativa de los fagos reportados hasta el momento presentan cierto sesgo morfológico al pertenecer al orden *Caudovirales*, es decir, fagos con cabeza y cola. Por un lado, se sabe que las estructuras icosaédricas son energéticamente favorables en sistemas biológicos [Caspar & Klug, 1962], de modo que las cápsides fágicas pueden presentar una organización similar a pesar de que las proteínas que las conforman no presenten similitud de secuencia. Esto último tiene como consecuencia que diversos fagos presenten una morfología muy conservada a pesar de estar filogenéticamente alejados [Caspar & Klug, 1962]. Estas observaciones se vuelven evidentes tanto en fagos aislados de ambientes naturales como en fagos modelo. Históricamente se han reportado bacteriófagos que a pesar de ser genéticamente distintos, muestran una morfología semejante, tal es el caso de los fagos P22 y T7. Ambos fagos están clasificados como virus pertenecientes a la familia *Podoviridae* (fagos con caudas cortas) sin embargo, no comparten homología significativa en su material genético. Por otro lado, también se ha reportado el escenario opuesto, es decir, fagos genéticamente relacionados pero con morfologías distintas: mientras que los bacteriófagos P22 y lambda son

morfológicamente distintos, su material genético es muy similar, al grado de poderse obtener híbridos funcionales entre ambos fagos [Botstein & Herskowitz, 1974].

La capacidad de los fagos P22 y lambda de formar híbridos viables, sugería que las funciones codificadas en sus genomas eran esencialmente las mismas y podían ser intercambiadas sin mayor complicación. Estos hallazgos sentaron las bases para el establecimiento de la teoría modular de evolución de los genomas fágicos. Esta teoría postula que la recombinación homóloga es la principal responsable del intercambio de material genético, sin embargo, la viabilidad de la progenie producida, esta dada por la compatibilidad de las funciones albergadas en los módulos intercambiados.

La teoría modular de evolución por si misma, no es capaz de explicar la diversidad genética observada en los bacteriófagos reportados hasta el momento ya que la recombinación homóloga no es indispensable para el intercambio de módulos entre genomas relacionados. Trayendo como consecuencia que los módulos intercambiados no necesariamente estén relacionados con funciones similares.

Adicionalmente, se han reportado genomas fágicos que son altamente conservados y no muestran variaciones en sus genomas [Flores *et al.*, 2012 & Cruz-Plancarte *et al.*, 2016], por lo que las funciones codificadas en sus genomas podrían ser no reemplazables, incluso si se presentara recombinación entre dichos fagos y otras entidades biológicas.

Lo anterior trae como consecuencia que se puedan observar distintos tipos de genomas fágicos: genomas altamente conservados a pesar de su separación espacio temporal; y genomas compuestos de regiones con orígenes diversos donde cada segmento puede tener una historia evolutiva distinta a la del resto del genoma. A este fenómeno se le conoce como mosaicismo, ya que los genomas fágicos son esencialmente mosaicos con segmentos de orígenes diversos (Figura 2). A partir de estas observaciones se ha postulado que podría existir una poza genética a la cual los bacteriófagos tienen un acceso no uniforme, mediado esencialmente por su rango de hospedero [Hendrix *et al.*, 1999].

Por lo anterior, se ha llegado a la conclusión que los bacteriófagos difícilmente siguen un modelo tradicional de herencia vertical, por ello se ha propuesto que su evolución pueda ser explicada a través de un modelo reticular [Lima-Mendez *et al.*, 2008]. Esto tiene como consecuencia una dificultad inherente en la clasificación de los bacteriófagos, ya que un mismo genoma podría estar relacionado con más de un grupo de fagos. No obstante, existen trabajos en donde se han propuesto esquemas de clasificación integrales que consideran las características morfológicas, genómicas y biológicas de los bacteriófagos para poder ubicarlos dentro de grupos concertados [Lavigne *et al.*, 2008; Lavigne *et al.*, 2009 & Adriaenssens *et al.*, 2014].

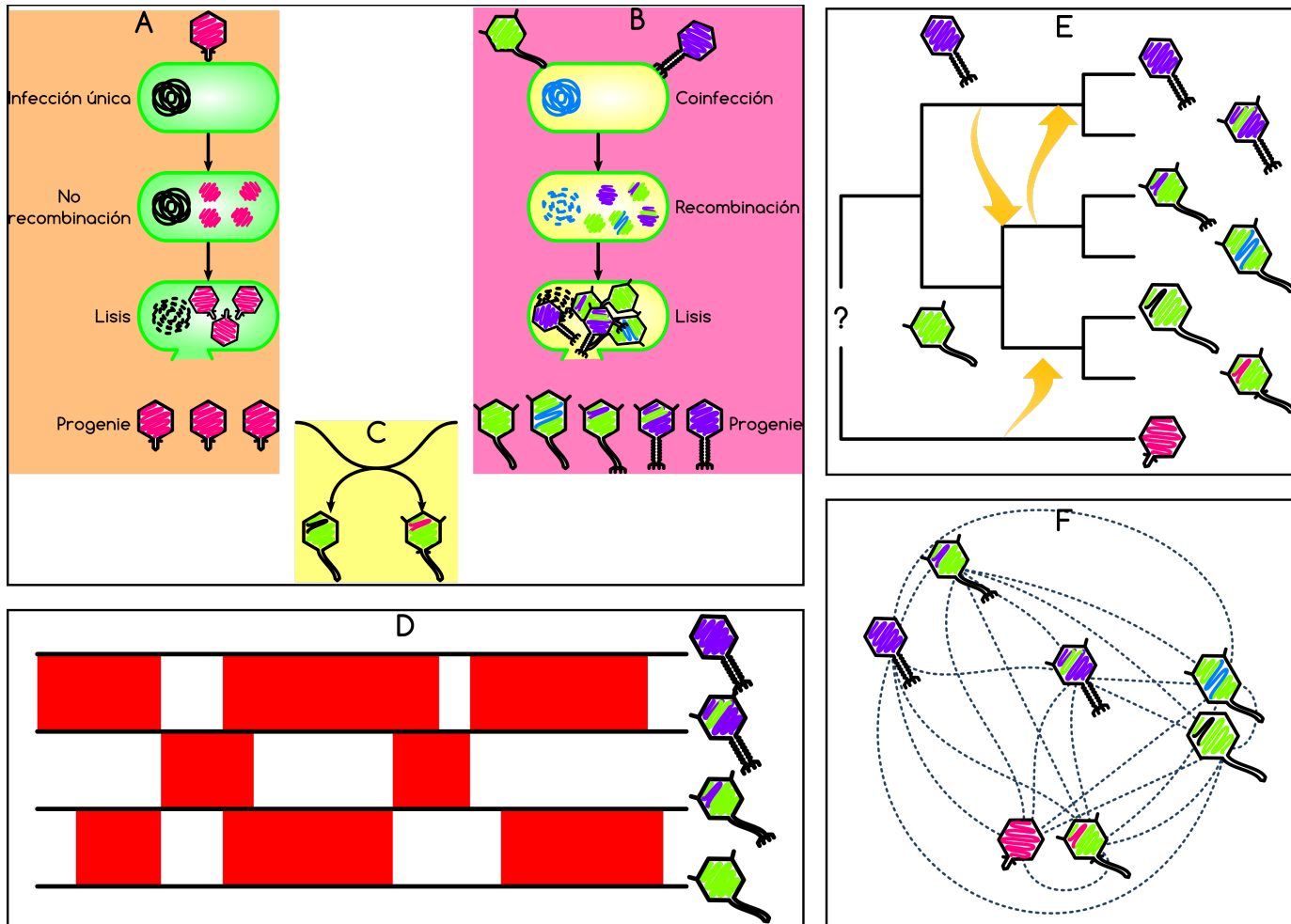


Figura 2: **A.** La falta de recombinación resulta en genomas altamente conservados, por lo que no se observa mosaicismo. **B.** La teoría modular de la evolución fágica señala que la recombinación es el mecanismo por el cual se genera el mosaicismo en los genomas fágicos. **C.** Un adendo a la teoría modular, indica que no es necesaria la recombinación homóloga para el intercambio de material genético, pero que es necesaria una poza genética común a la que los fagos pueden tener acceso dependiendo principalmente de su rango de hospedero. **D.** La comparación de los genomas fágicos nos permite evidenciar el mosaicismo, representado por zonas de alta similitud, intercaladas con regiones que pueden provenir de otro linaje. **E.** Lo anterior tiene como consecuencia que la inferencia filogenética tradicional resulte complicada debido a que la transferencia horizontal de material genético (representado por flechas naranjas) no permite la determinación de relaciones de ancestría. **F.** Considerando las limitaciones de la inferencia filogenética tradicional, se ha optado por representar las relaciones existentes entre genomas fágicos como redes filogenómicas que consideran la transferencia horizontal de material genético y permiten observar relaciones múltiples que pasarían desapercibidas utilizando árboles filogenéticos.



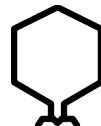
1.4 Los bacteriófagos de *Pseudomonas*

Las bacterias pertenecientes al género *Pseudomonas* pueden ser infectadas por una gran variedad de bacteriófagos que van desde los *Inovirus* con genomas relativamente cortos (~3.6 kbp) hasta los fagos gigantes *Phikzvirus* con genomas de 200 a 300 kbp [Krylov et al., 2007].

Debido a la importancia médica, biotecnológica y agropecuaria de las bacterias que pertenecen al género *Pseudomonas*, sus bacteriófagos han sido ampliamente estudiados y muchos de ellos han sido completamente secuenciados. Los hallazgos realizados en estas secuencias indican que los fagos de *Pseudomonas* representan un gran reservorio de información genética novedosa: en 201 genomas disponibles de dichos fagos se han encontrado 18,866 genes codificantes, de los cuales, únicamente 4,208 proteínas tienen una función asignada.





Al momento de la publicación de este trabajo, existen reportados 209 fagos de *Pseudomonas* en bases de datos, de los cuales, se cuenta con la secuencia genómica de 201. Estos fagos se distribuyen en 22 grupos con base en su similitud de secuencia nucleotídica y aminoacídica, morfología microscópica y propiedades biológicas. Aún así, una fracción significativa de los fagos de *Pseudomonas* no han sido asignados a un grupo y se clasifican únicamente con base en su morfología. A continuación se muestran los diversos géneros de fagos que infectan *Pseudomonas*, así como algunas de sus propiedades más relevantes, tales como su morfología y tipo de ácido nucleico. Los géneros mostrados corresponden con la nomenclatura actualmente aceptada por el *International Comitee for the Taxonomy of Viruses*.

Tabla 1: Géneros de fagos de *Pseudomonas* aceptados por el ICTV

Género	Mat. Genético	Morfología	Integrantes ¹
<i>Elvirus</i>			EL, OBP
<i>Kpp10virus</i>			CHA_P1, KPP10 , P3_CHA, PAK_P3, PAK_P5, vB_PaeM_PAO1_Ab03, vB_PaeM_PAO1_Ab04, vB_PaeM_PAO1_Ab06, vB_PaeM_PAO1_Ab17, vB_PaeM_PS24
<i>Peduovirus</i>			phiCTX
<i>Pakpunavirus</i>			C11, JG004, K5, K8, PA10, PAK_P1 , PAK_P2, PAK_P4, PaoP5, PaP1, phiMK, vB_PaeM_C2-10_Ab02, vB_PaeM_C2-10_Ab08, vB_PaeM_C2-10_Ab1, Zigelbrucke
<i>Pbunavirus</i>			14-1, DL52, DL60, DL68, F8, Gallinipper, JG024, Jollyroger, KPP12, KPP22, KPP22M1, KPP22M2, KPP22M3, Kraken, Kula, LBL3, LMA2, Nemo, Nessie, NH-4, NP3, PA5, PB1 , phiKT28, phiKTN6, Poseidon, S12-1, Smee, SN, SPM-1, Triton, vB_Pae_PS44, vB_PaeM_CEB_DP1, vB_PaeM_PAO1_Ab27
<i>Phikzvirus</i>	dsDNA		201phi2-1, KTN4, phiKZ , PhiPA3
No clasificados			Lu11, PaBG, phi3, phiPsa374, vB_PsyM_KIL1, vB_PsyM_KIL2, vB_PsyM_KIL3, vB_PsyM_KIL3b, vB_PsyM_KIL4, vB_PsyM_KIL5
<i>F116virus</i>			F116 , H66, LKA5
<i>Lit1virus</i>			DL64, LIT1 , Pa2, PA26, PEV2, phi176, RWG, vB_PaeP_C2-10_Ab09, vB_PaeP_MAG4, YH30, YH6
<i>Luz7virus</i>			KPP21, LUZ7
<i>Luz24virus</i>			DL54, LUZ24 , MR299-2, PaP3, PhiCHU, phiIBB-PAA2, TL, vB_PaeP_C1-14_Or, vB_PaeP_C2-10_Ab22, vB_PaeP_p2-10_Or1
<i>Phikmvvirus</i>			DL62, LKA1, LKD16, LUZ19, MPK6, MPK7, phi-2, phikF77, phiKMV , phiNFS, PT2, PT5, vB_Pae-TbilisiM32, vB_PaeP_PAO1_1-15pyo, vB_PaeP_PAO1_Ab05, vB_PaeP_PPA-ABTNL, YMC11/06/C171_PPU_BP
<i>T7virus</i>			gh-1, Pf-10, Phi-S1, phi15, phiIBB-PF7A, phiPsa17, phiPSA2, PPPL-1, UNO-SLW1, UNO-SLW2, UNO-SLW3, UNO-SLW4
No clasificados			119X, AF, PaMx33, PaMx35, PaMx41, PaMx43, PaMx46, PaP2, R18, tf, UFV-P2, Andromeda, Bf7, VSW-3

¹Los representantes de cada género se muestran resaltados en negritas

Tabla 1 (continuación): Géneros de fagos de *Pseudomonas* aceptados por el ICTV

Género	Mat. Genético	Morfología	Integrantes ¹
<i>Ab18virus</i>	dsDNA		PaMx11, vB_PaeS_PAO1_Ab18
<i>D3112virus</i>			D3112 , DMS3, FHA0480, H70, JBD24, JBD26, JBD30, JBD5, JBD69, JBD88a, JBD93, JD024, MP22, MP29, MP38, MP42, MP48, PA1phi, PaMx73, vB_PaeS_PAO1_Ab30, YMC11/11/R1836
<i>D3virus</i>			D3 , PMG1
<i>Np1virus</i>			NP1 , PaMx25
<i>Pamx74virus</i>			PaMx28, PaMx74 , AAT-1
<i>Septima3virus</i>			73 , PaMx42, vB_Pae_PS9N, vB_Pae-Kakheti25, vB_PaeS_SCH_Ab26
<i>Yuavirus</i>			AN14, LKO4, M6, MP1412, PAE1, YuA
No clasificados		B3, F10, JBD44, MD8, PAJU2, phi1, phi2, phi297, phiPSA1, SM1, YMC/01/01/P52_PAE_BP, YMC11/02/R656, YMC11/07/P54_PAE_BP, JBD18, JBD25	
<i>Cystovirus</i>	dsRNA		phi6, phi10, phi12, phi13, phi2954, phi7, phi8, phi9
<i>Inovirus</i>	ssDNA		pf1, pf3
<i>Levivirus</i>	ssRNA		PP7, PRR1

¹Los representantes de cada género se muestran resaltados en negritas

2. Antecedentes

Desde su descubrimiento han sido propuestos distintos esquemas para la clasificación de los bacteriófagos. Actualmente no existe un método lo suficientemente robusto que permita la identificación y clasificación de todos los fagos dentro de grupos definidos [Ackerman, 2005]. Esto es particularmente cierto con los genomas que presentan un alto grado de mosaicismo o cuando se trata de genomas aparentemente novedosos que no muestran relaciones evidentes con grupos definidos de fagos.

Se han implementaron métodos computacionales para la clasificación de fagos pertenecientes a las familias *Podoviridae*, *Myoviridae* y *Siphoviridae* [Lavigne et al., 2008, Lavigne et al., 2009, Adriaenssens et al., 2014]. Estos métodos consideran el número de genes similares entre múltiples genomas, para posteriormente obtener proporciones de genes compartidos y así crear grupos de fagos relacionados ya sea a nivel nucleotídico o aminoacídico. No obstante, las limitaciones de dichos métodos siguen siendo los genomas con alto grado de mosaicismo o los genomas novedosos.

En 2012 nuestro grupo de trabajo, identificó fagos de *Pseudomonas* que posiblemente representaban nuevas especies fágicas, todos ellos pertenecientes a la familia *Siphoviridae*, sin embargo únicamente se emplearon datos fenotípicos para su consideración como especies nuevas, mientras que los métodos computacionales utilizados no revelaban si dichas especies pertenecían a grupos conocidos de fagos previamente descritos [Sepúlveda-Robles et al., 2012].

En 2013 se aisló al fago de *Vibrio* phiJM2012, y mediante una estrategia filogenómica reticular, se detectaron relaciones de similitud entre phiJM2012 y los fagos de *Pseudomonas* del género *Phikzvirus* [Jang et al., 2013]. No obstante, estas estrategias presentan ciertas limitaciones, tales como el tamaño de las bases de datos consultadas, el grado de refinamiento en las anotaciones genómicas y la metodología empleada para evaluar las relaciones entre los fagos estudiados.

Finalmente, en 2015 se llevó a cabo la comparación de más de 600 genomas de mycobacteriófagos, logrando la identificación de un continuo genómico en donde los límites previamente establecidos entre los grupos de fagos construidos, poco a poco parecían desaparecer con la inclusión de nuevos genomas que conectaran grupos adyacentes de fagos [Pope et al., 2015]. La noción del continuo genómico ya había sido propuesta [Lawrence et al., 2002], sin embargo, no fue hasta 2015 que se obtuvo evidencia a favor de dicho modelo.

3. Justificación

En el laboratorio se cuenta con una colección de fagos de *Pseudomonas* así como con las secuencias genómicas de 53 de ellos. Sin embargo, seis de los fagos pertenecientes a dicha colección mostraron secuencias y patrones moleculares que los ubicaban fuera de los grupos propuestos en ese momento. Adicionalmente, se desconocían las funciones albergadas en dichos genomas, así como las relaciones existentes entre ellos y los genomas de otros fagos. Por ello, el presente trabajo consistió en la determinación de las funciones codificadas en cinco de los seis genomas novedosos, así como el diseño e implementación de estrategias computacionales y experimentales que permitieran detectar relaciones genómicas entre los fagos de estudio y otras entidades biológicas.

4. Objetivos

4.1 Objetivo general

Realizar un estudio genómico funcional y comparativo de fagos que infectan *Pseudomonas* sp. centrado en los fagos PaMx11, PaMx25, PaMx28, PaMx42 y PaMx74

4.2 Objetivos particulares

- Ensamblar y refinar las secuencias genómicas de los fagos PaMx11, PaMx25, PaMx28, PaMx42 y PaMx74
- Realizar la anotación de dichos genomas y asignar funciones a los genes encontrados en dichos genomas mediante homología de secuencia y dominios conservados
- Comparar las secuencias nucleotídicas de los genomas de fagos de *Pseudomonas* para establecer grupos de fagos relacionados estrechamente.
- Comparar las secuencias aminoacídica de las proteínas codificadas en los genomas de fagos de *Pseudomonas* para establecer grupos de fagos relacionados distantemente.
- Evaluar si los ORFs homólogos encontrados en las comparaciones múltiples presentan arreglos similares mediante comparaciones pareadas
- Determinar la asignación taxonómica de los fagos caracterizados en este trabajo con base en los lineamientos del comité internacional de taxonomía de virus (ICTV por sus siglas en inglés).
- Determinar cuales podrían ser las relaciones evolutivas de los fagos de estudio con respecto de otros genomas depositados en bases de datos mediante el uso de redes filogenómicas

5. Materiales y métodos

5.1 Bacteriófagos empleados en este estudio

Para el presente estudio se seleccionaron los fagos representantes de nuevas especies de fagos de *Pseudomonas*, descritos en el trabajo “High diversity and novel species of *Pseudomonas aeruginosa* bacteriophages” [Sepúlveda-Robles *et al.*, 2012]. En este trabajo nos referiremos a los fagos PaMx11, PaMx25, PaMx28, PaMx42 y PaMx74 (números de acceso JQ067087.2, JQ067084.3, JQ067089.2, JQ067092.2 y JQ067093.2 respectivamente) como los fagos de estudio.

5.2 Secuenciación de los genomas de los fagos de estudio

La secuenciación del genoma de PaMx11 se llevó a cabo en CINVESTAV unidad Zacatenco, en el Laboratorio de Referencia y Apoyo para la Caracterización de Genomas, Transcriptomas y Microbiomas del departamento de Genética y Biología Molecular, empleando la tecnología Ion Torrent.

La secuenciación de los genomas de los fagos PaMx25 y PaMx28 se efectuó en la unidad de servicios genómicos del Laboratorio Nacional de Genómica para la Biodiversidad (LANGEBIO) a través de la tecnología Roche 454.

La secuenciación de los genomas de los fagos PaMx42 y PaMx74 se realizó empleando el servicio de secuenciación masiva de la compañía Macrogen (Macrogen Inc. Seoul, Korea).

Las lecturas obtenidas en la secuenciación de los genomas fueron ensambladas usando MIRA v3.4.0.1 [Chevreux *et al.*, 1999] y los contigs obtenidos fueron reensamblados usando velvet v1.1.02 [Zerbino & Birney, 2008]. Finalmente, las lecturas obtenidas fueron mapeadas en los ensamblados obtenidos para determinar variaciones puntuales en la secuencia usando SHRiMP v2.2.3 [David *et al.*, 2011].

La secuencia del fago PaMx25 fue completada empleando secuenciación Sanger con los oligonucleótidos CTTACTGCTGAGGTTGAGCCC y GGTCAGTATGAGGACTCCAACGA, los cuales alinean con el genoma en las posiciones 39,551 y 40,046 respectivamente.

La topología de las secuencias obtenidas fue ajustada de modo tal que los fagos pudieran ser comparados gráficamente con sus homólogos más cercanos.

5.3 Identificación de ORFs

Las secuencias genómicas de los fagos empleados fueron analizadas empleando los programas GeneMark [Borodovsky *et al.*, 2003] y rbs_finder.pl [Suzek *et al.*, 2001] para la búsqueda de marcos de lectura abierta y para la identificación de sitios de unión a ribosoma

respectivamente. Los resultados fueron revisados manualmente y fueron editados con base en los resultados obtenidos para fagos de *Pseudomonas* previamente depositados en bases de datos.

5.4 Asignación de funciones a los genes encontrados

Con la finalidad de encontrar las posibles funciones de las proteínas putativas codificadas en los genomas analizados, las secuencias aminoacídicas de los posibles ORFs fueron estudiadas empleando 4 estrategias:

1. Búsqueda de homología a nivel de secuencia aminoacídica

Las secuencias aminoacídicas deducidas de los ORFs predichos fueron sometidas a búsquedas tipo BLASTp [Altschul et al., 1997] empleando la base de datos no redundante del *National Center for Biotechnology and Information* (<http://www.ncbi.nlm.nih.gov>). Las búsquedas se realizaron con un valor de expectancia menor o igual a 1×10^{-4} . Los resultados fueron filtrados, eliminando los hits que mostraran únicamente similitud localizada en regiones de las secuencias menores al 35% de la proteína más pequeña, y cuya región alineada fuera inferior al 50% de la proteína más pequeña.

2. Búsqueda de dominios conservados

Mediante el uso de InterProScan (<http://www.interpro.ebi.ac.uk>) [Jones et al., 2014] se identificó la presencia de dominios conservados en las secuencias aminoacídicas codificadas en los genomas de fagos de *Pseudomonas*. Los dominios conservados fueron de utilidad para la asignación de funciones a las proteínas que mostraran similitud contra secuencias de función desconocida, o que no mostraran similitud de secuencia contra ninguna secuencia en bases de datos.

3. Búsqueda de homología estructural

Las secuencias aminoacídicas de las proteínas putativas codificadas en los genomas analizados, fueron modeladas empleando el servidor I-TASSER [Yang et al., 2015]. Los modelos obtenidos fueron comparados contra estructuras depositadas en el Protein Data Bank (<http://www.rcsb.pdb.org>) y se consideraron como homólogos aquellos pares de estructuras que mostraran una desviación media entre átomos (RMSD por sus siglas en inglés) menores a 2 Å, y que mostraran un TM-score mayor a 0.75.

4. Identificación de las proteínas estructurales del virión mediante espectrometría de masas

Las partículas virales de los fagos de estudio fueron purificadas a partir de gradientes de CsCl [Sepúlveda-Robles et al., 2012]. Posteriormente, las proteínas estructurales fueron resueltas en geles de poliacrilamida al 14% y teñidas con azul de coomassie. Las bandas fueron

escindidas del gel y las proteínas obtenidas fueron procesadas en un espectrómetro de masas LTQ-Orbitrap Velos. Es necesario indicar, que el procedimiento experimental fue realizado amablemente por Omar Sepúlveda Robles como parte del trabajo publicado a partir de los resultados obtenidos en este estudio. Los espectros obtenidos, fueron procesados empleando el programa TransProteomic Pipeline v4.8 [Deutsch et al., 2015]. Las proteínas estructurales fueron identificadas con base en la cobertura y el número de espectros obtenidos a partir de los datos de espectrometría de masas.

Los resultados fueron consolidados empleando el programa Artemis v14 [Carver et al., 2012], las funciones asignadas, así como las secuencias actualizadas fueron depositadas en la base de datos GenBank. (<http://www.ncbi.nlm.nih.gov/genbank>) bajo los números de acceso JQ067087.2 (PaMx11); JQ067084.3 (PaMx25); JQ067089.2 (PaMx28); JQ067092.2 (PaMx42 y JQ067093.2 (PaMx74).

5.5 Genómica comparativa

Los genomas de los fagos de estudio fueron comparados a nivel de secuencia nucleotídica y aminoacídica para investigar las relaciones existentes entre dichos genomas y otras secuencias depositadas en bases de datos. Para llevar a cabo dichas comparaciones, se emplearon los genomas de 201 fagos que infectan *Pseudomonas*, en el anexo 1 se muestran las características de dichos genomas.

Las comparaciones a nivel de nucleótidos se realizaron empleando el programa MUMmer [Delcher et al., 2003] mientras que las comparaciones a nivel de aminoácidos se realizaron mediante BLASTp [Altschul et al., 1997]. A continuación se describirá brevemente la estrategia de cada comparación.

5.5.1 Comparaciones nucleotídicas.

Las secuencias de los genomas fágico en formato fasta fueron comparadas de forma pareada con el programa MUMmer, el cual arroja tablas que indican los intervalos comparados y los porcentajes de identidad entre dichos intervalos, a continuación se muestra un ejemplo de lo que arroja MUMmer

[S1]	[E1]	[S2]	[E2]	[LEN 1]	[LEN 2]	[% IDY]	[TAGS]
10	1183	10	1171	1174	1162	95.91	NC_007807 KU884561
1184	24214	1347	24389	23031	23043	95.45	NC_007807 KU884561
25180	25660	25050	25530	481	481	85.68	NC_007807 KU884561
25895	34691	25774	34569	8797	8796	96.30	NC_007807 KU884561
34914	42651	34795	42562	7738	7768	94.24	NC_007807 KU884561
42713	43336	42624	43237	624	614	88.38	NC_007807 KU884561

Donde: [S1] y [S2] corresponden a las coordenadas del inicio del alineamiento para la secuencia 1 y 2 respectivamente. [E1] y [E2] corresponden a las coordenadas del fin del alineamiento para la secuencia 1 y 2 respectivamente. [Len1] y [Len2] corresponde a la longitud de la región alineada para la secuencia 1 y 2 respectivamente. [%Idy] representa el porcentaje de identidad para el alineamiento especificado. El campo marcado como [Tags] corresponde a los nombres de las secuencias 1 y 2.

El porcentaje de identidad global entre dos genomas, fue calculado tomando en consideración la suma ponderada de los porcentajes de identidad entre las secuencias comparadas.

5.5.2 Comparaciones aminoacídicas.

Las secuencias aminoacídicas de las proteínas codificadas en los genomas fágico fueron empleadas para construir una base de datos tipo BLAST, posteriormente, cada proteína se reto contra la base de datos, usando 0.0001 como umbral en el valor de expectancia. Al concluir las comparaciones, se obtuvo una tabla de hits con la siguiente estructura:

Query	Subject	Qln	Sln	Qst	Sst	Qnd	Snd	Scr	e-val	Len	Pos
NC_004665.51	NC_019923.105	562	599	123	170	500	533	204	4e-17	386	160
NC_004665.59	NC_022096.394	709	1032	294	490	708	928	122	5e-07	464	168
KT887557.36	NC_004665.51	609	562	169	123	533	500	222	2e-19	391	167
JN254801.73	NC_004665.59	545	709	156	326	545	709	300	5e-29	441	173
JN254801.89	NC_004665.51	569	562	40	35	551	535	327	9e-33	532	232
NC_004665.51	NC_028919.85	562	569	74	81	535	551	268	5e-25	495	210
NC_004665.59	NC_028919.67	709	545	392	231	709	545	307	9e-30	369	151
NC_004665.66	NC_028919.57	314	276	81	44	215	203	111	1e-06	169	68
NC_004466.81	NC_004665.59	545	709	156	326	545	709	300	5e-29	441	173
NC_004466.98	NC_004665.51	569	562	40	35	551	535	328	7e-33	532	232

Donde las primeras dos columnas corresponden al número de acceso de las secuencias comparadas (*Query & Subject*). Las columnas 3 y 4 indican las longitudes de dichas secuencias (*Query length* → *Qln*; *Subject length* → *Sln*). Las columnas 5 y 6 muestran las coordenadas del inicio del alineamiento para cada secuencia (*Query start* → *Qst*; *Subject start* → *Sst*). Las columnas 7 y 8 muestran las coordenadas del fin de dicho alineamiento (*Query end* → *Qnd*; *Subject end* → *Snd*). Las columnas 9 a 12 muestran el *score* (*Scr*), e-value, longitud (*Len*) y número de *matches* positivos (*Pos*) de cada alineamiento respectivamente.

Dado que la tabla obtenida, no refleja por si misma las relaciones existentes entre dos genomas fágicos, se obtuvieron dos medidas a partir de dicha tabla: el porcentaje de ORFs homólogos entre dos genomas, y el score acumulado a partir de dichos ORFs homólogos.

En este trabajo, se definió un par de ORFs como homólogos, si cumplía con los siguientes requisitos:

1. Ambas secuencias deben ser de longitudes comparables
2. Las regiones repetitivas entre ambas secuencias no deben aportar mucho al score global
3. La longitud de la región alineada entre las dos secuencias debe ser significativa.

5.5.3 Optimización de parámetros

Tanto MUMmer como BLASTp cuentan con parámetros que al ser alterados pueden modificar severamente los resultados obtenidos. En el caso de MUMmer, el principal valor a considerar es la longitud de quiebre (*break length*) es decir, el número de residuos a considerar en un alineamiento de baja calidad antes de calificar una región como no homóloga, es decir, que tan tolerante es el alineamiento a una serie de mismatches dentro de un alineamiento. Por otro lado, los valores de corte para las comparaciones empleando BLASTp, ya han sido reportados previamente, sin embargo, los parámetros para la definición de un par de ORFs como homólogos no han sido establecidos con claridad. En este trabajo, los parámetros que afectan el establecimiento de un par de ORFs como homólogos, es la diferencia en la longitud de dichos ORFs, la tolerancia a repetidos en los alineamientos, y la fracción que representa la región alineada con respecto de la longitud de los ORFs.

Con la finalidad de determinar cuales eran los parámetros óptimos para cada programa, se realizaron comparaciones múltiples empleando diversos rangos en los valores de cada parámetro. Para ello, se consideraron los genomas de fagos de *Pseudomonas* que tuvieran una asignación taxonómica reconocida y se compararon dichos genomas contra el genoma de referencia de dicho grupo. Cada set de comparaciones se realizó empleando un valor distinto, y los parámetros óptimos fueron determinados con base en curvas de rarefacción (Anexo 2).

5.6 Detección de grupos de fagos relacionados

Una vez que se optimizaron los parámetros, se llevaron a cabo comparaciones globales entre los genomas de los fagos PaMx11, PaMx25, PaMx28, PaMx42 y PaMx74; contra secuencias depositadas en bases de datos. Los resultados obtenidos fueron expresados en forma de matrices de similitud, y a partir de estas matrices, se construyeron grupos de fagos relacionados. Dichos grupos se definieron como aquel conjunto de genomas que mostraran porcentajes detectables de identidad a nivel de secuencia nucleotídica y que mostraran por lo menos 40% de ORFs homólogos [Adriaenssens et al., 2014]. No obstante, para la evaluación de las relaciones distantes en los genomas comparados, se consideró un umbral en el porcentaje de ORFs homólogos mayor o igual a 20%. Este valor fue elegido tomando en consideración la naturaleza quimérica observada en los genomas de los fagos PaMx25, PaMx28 y PaMx74 (ver

sección 6.4.3). Para validar la formación de grupos, se compararon los resultados obtenidos, con la clasificación actual de fagos de *Pseudomonas* (Anexo 3).

5.7 Detección de regiones de sintenia entre genomas

Si bien el porcentaje de ORFs homólogos provee de una medida robusta para evaluar las posibles relaciones entre genomas, esta aproximación no permite inferir si los ORFs homólogos entre dos genomas se ubican dentro de un mismo contexto genómico. Por ello, se decidió seleccionar genomas de referencia para cada grupo, y compararlos contra los genomas de los fagos PaMx11, PaMx25, PaMx28, PaMx42 y PaMx74. Los genomas de referencia fueron seleccionados con base en la calidad de sus anotaciones y la presencia de literatura donde se hayan descrito dichos genomas. En la siguiente tabla se muestran los genomas representantes de cada género.

Tabla 2: Genomas fágicos seleccionados como representantes de los géneros a los que pertenecen los fagos en estudio

Fago	Número de acceso	Familia	Género	Longitud (bp)	Fagos comparados
vB_PaeS_PA01_Ab18 ¹	NC_026594	<i>Siphoviridae</i>	<i>Ab18virus</i>	56,537	PaMx11
NP1	NC_031058	<i>Siphoviridae</i>	<i>Np1virus</i>	58,566	PaMx25
vB_Pae-Kakheti25 ²	NC_017864	<i>Siphoviridae</i>	<i>Septima3virus</i>	42,844	PaMx42
AAT-1 ³	KU204984	<i>Siphoviridae</i>	<i>Pamx74virus</i>	57,599	PaMx28 PaMx74

¹Essoh *et al.*, 2015 ²Karumidze *et al.*, 2012 ³Andrade-Domínguez & Kolter, 2016

Los mapas genómicos de los fagos seleccionados fueron comparados empleando tBLASTx con un umbral de valor de expectancia de 1×10^{-4} y una longitud de alineamiento mayor o igual a 50 residuos.

5.8 Comparaciones genómicas con fagos de otros géneros

Además de las comparaciones realizadas contra genomas de fagos de *Pseudomonas*, las secuencias de los fagos PaMx11, PaMx25, PaMx28, PaMx42 y PaMx74 fueron comparadas contra genomas de fagos pertenecientes a otros géneros y/o que infectaran especies bacterianas diferentes a *Pseudomonas*. Las estrategias de comparación fueron las mismas que se emplearon para las comparaciones contra fagos de *Pseudomonas*.

5.9 Posición filogenómica de los fagos estudiados

Dado que ninguno de los fagos estudiados presentaba un gen en común, y que con base en las comparaciones gráficas, mostraban un alto grado de mosaicismo, se optó por una

estrategia filogenómica para evaluar las posibles relaciones evolutivas entre los fagos comparados. Para cumplir este objetivo, se empleó como medida de asociación, el porcentaje de ORFs homólogos entre genomas y su score acumulado BLASTp. A partir de estas datos se construyó una red asociativa en donde los genomas y sus relaciones fueron representados como nodos y las aristas que los unen, respectivamente. La posición filogenómica de cada fago fue evaluada empleando un algoritmo de disposición ponderando en las aristas el porcentaje de ORFs homólogos, o el score acumulado de BLASTp (*edge-weighted spring-embedded layout algorithm: EWSE*). Dicho algoritmo supone que cada arista une dos nodos con una fuerza atractora determinada por el porcentaje de ORFs homólogos, o su score acumulado BLASTp. En este trabajo se empleó una simulación de 10,000 iteraciones en donde la posición final de los genomas estudiados, está determinada por su grado de asociación con otros genomas, es decir, los genomas que están altamente asociados, forman grupos compactos donde las aristas entre nodos tienen una longitud mínima. Por otro lado, los genomas que están lejanamente relacionados, muestran aristas largas. Finalmente, a través de esta aproximación se puede estimar cual es el grupo más cercano para aquellos genomas que muestren relaciones con más de un grupo.

5.10 Material bibliográfico derivado de este trabajo

Los resultados correspondientes a la caracterización del genoma del fago PaMx25 fueron publicados en la revista científica *Archives of virology* [Flores *et al.*, 2017] como parte de los requisitos para acreditar el grado de doctorado en ciencias.

6. Resultados

6.1 Refinamiento de las secuencias de los fagos estudiados

A partir de los datos de secuenciación obtenidos, se logró determinar las secuencias genómicas completas de los fagos PaMx11, PaMx25 y PaMx42, obteniéndose secuencias de 59,878 bp, 57,899 bp y 43,225 bp respectivamente. Por otro lado, los genomas de los fagos PaMx28 y PaMx74 no pudieron ser completados en su totalidad, en el caso del fago PaMx28 se lograron secuenciar 55,008 bp, mientras que el fago PaMx74 presentó únicamente 58,627 bp.

A pesar de que las secuencias de los fagos PaMx28 y PaMx74 presentaron regiones que no pudieron ser secuenciadas, las comparaciones gráficas (ver sección 6.4.3) indicaron que las regiones no secuenciadas no implicaban una falta considerable de información, asimismo, las conclusiones de este trabajo permanecerían válidas aún después de completar dichos genomas.

6.2 Anotación de genomas

Mediante el uso de búsquedas tipo BLAST y búsquedas de dominios conservados con InterProScan, se logró la asignación de funciones a 180 genes de un total de 360 genes presentes en los 5 genomas analizados. Los ORFs identificados fueron agrupados con base en su función dentro de 5 categorías (Tabla 3). La descripción completa de las funciones identificadas en los genomas estudiados se muestra en el anexo 4

Tabla 3: Resumen de las funciones encontradas en los genomas de los fagos en estudio

Genoma	Empaquetamiento de DNA	Morfogénesis	Lisis celular	Replicación de DNA	Metabolismo de nucleótidos	Función desconocida
PaMx11	3	17 (11)	3	6	10	43
PaMx25	3	18 (15)	4	11	9	30
PaMx28	3	17 (17)	3	5	8	39
PaMx42	3	17 (14)	5	6	2	24
PaMx74	3	17	3	5	7	41

Los números en cada celda corresponden al número de ORFs encontrados en cada genoma para cada categoría. Los números entre paréntesis, indican el número de ORFs verificados por espectrometría de masas

6.3 Funciones específicas:

6.3.1 Empaquetamiento de DNA

En los genomas estudiados se logró la identificación de los ORFs codificantes de las subunidades menor y mayor de la terminasa (genes *terS* y *terL*, respectivamente) así como la proteína portal a través de homología de secuencia, dominios conservados y homología estructural (anexo 4). Adicionalmente, la proteína portal fue identificada experimentalmente a partir de los resultados de espectrometría de masas en los cinco genomas (sección 6.3.2).

Los resultados de homología estructural, indicaron que la proteína portal de los cinco fagos estudiados es homóloga (RMSD de 1.38 a 2.62) a la proteína portal del fago SPP1 que infecta a *Bacillus* (Num. Acc. PDB: 2JES; Anexo 4) [Lebedev et al., 2007].

Por otro lado, las subunidades mayores de la terminasa de los fagos PaMx11, PaMx28, PaMx42 y PaMx74 resultaron ser homólogos estructurales (RMSD de 0.89 a 1.71) de la subunidad mayor de la terminasa del fago Sf6 (Num. Acc. PDB: 4IDH) que infecta a *Shigella* [Zhao et al., 2013]. Del mismo modo, las subunidades menores de la terminasa de los fagos PaMx11, PaMx28, y PaMx74 mostraron homología estructural (RMSD de 1.69 a 2.5) con la subunidad menor de la terminasa del fago Sf6 (Num. Acc. PDB: 3HEF) [Zhao et al., 2010]. No obstante, la subunidad mayor de la terminasa del fago PaMx25 mostró homología estructural (RMSD = 1.71) con la subunidad mayor del fago T4 (Num. Acc. PDB: 3CPE) [Sun et al., 2008].

La maquinaria de empaquetamiento de DNA de los bacteriófagos es un potente motor que involucra la translocación de DNA hacia el interior de las cápsides pre ensambladas en el citoplasma bacteriano [Isidro et al., 2004]. Dicho proceso es inherentemente complejo, y las proteínas involucradas participan en distintas etapas. En ese sentido, se sabe que la proteína portal del fago SPP1 no solamente sirve como punto de anclaje para el complejo de terminasas, sino que también participa en la expansión de la cápside, liberación de la proteasa de andamiaje y finalmente en la translocación de DNA hacia el interior de la cápside [Isidro et al., 2003].

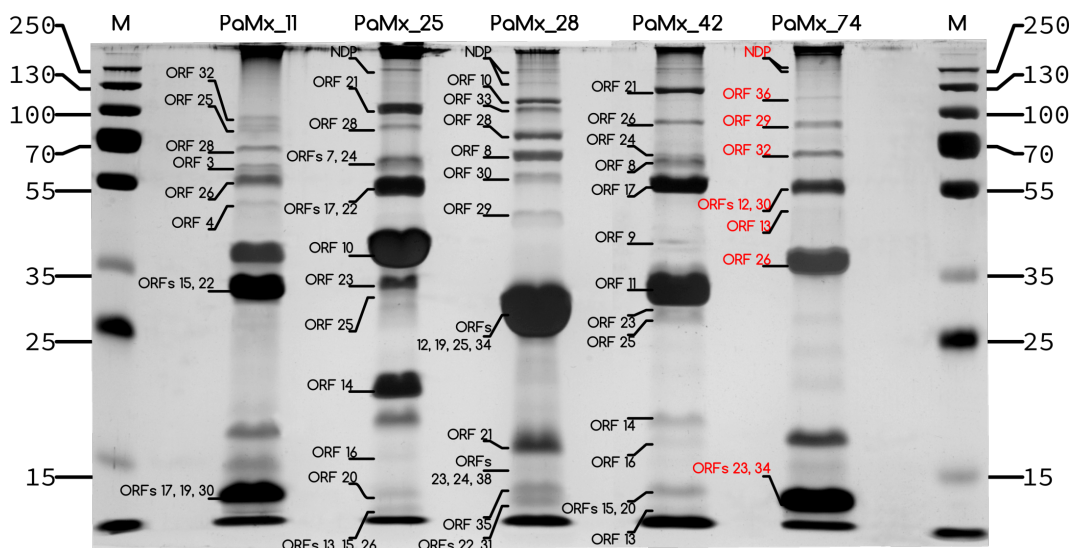
El otro componente principal del motor de empaquetamiento es el complejo de terminasas. En este complejo, la subunidad menor formando un anillo homomultimérico de , dando lugar a un canal a través del cual pasa el DNA fágico. A su vez, la subunidad mayor de la terminasa se asocia en homopentámeros u homohexámeros, la función de la subunidad mayor de la terminasa es translocar el DNA a través de la hidrólisis de ATP [Sun et al., 2008], y de cortar el DNA una vez que la cápside ha recibido suficiente DNA.

Los resultados derivados de la detección de homología estructural, sugieren que el mecanismo de empaquetamiento de DNA de los fagos estudiados, podría ser similar al de los fagos SPP1, Sf6 y T4. Dado que la proteína portal de los cinco fagos mostró similitud con la de

SPP1, podría pensarse que el acoplamiento entre el complejo de terminasas y la procápside es similar al del fago SPP1. Asimismo, el proceso de translocación y empaquetamiento de DNA en los fagos de estudio podría ser de tipo *headfull*, dada la similitud estructural de las subunidades mayores de las terminasas con aquellas de los fagos Sf6 y T4, ambos empaquetadores tipo *headfull* [Casjens et al., 2005]. A pesar de que estas observaciones son consistentes con los modelos de empaquetamiento actuales, será necesario realizar los experimentos adecuados para validar las predicciones realizadas.

6.3.2 Morfogénesis

Con base en homología de secuencia, se logró la asignación de por lo menos 17 proteínas estructurales en cada genoma estudiado, adicionalmente, la presencia de dominios conservados en dichas secuencias, permitió la identificación de la proteína portal, la proteína F de morfogénesis, las proteínas mayores de cápside y cola; así como de la proteína de medida de la cola. Los resultados obtenidos en los ensayos de espectrometría de masas confirmó la mayoría de las predicciones (Figura 3, Anexo 4). En general, se obtuvieron espectros de buena calidad (Anexo 5), sin embargo, las proteínas del fago PaMx74 mostraron baja cobertura, por lo que la identificación de las proteínas presentes en cada banda se logró con base en la similitud de secuencia de dichas proteínas con las del fago PaMx28 (ver sección 6.4.3).



6.3.3 Lisis celular

Comúnmente, los genes asociados con la lisis del hospedero, se encuentran adyacentes en los genomas fágicos y codifican para proteínas de membrana y proteínas que degradan la pared celular [Young, 2014]. A través de la homología de secuencia y de la presencia de dominios conservados, se identificaron los módulos de lisis en los cinco fagos estudiados. El módulo de lisis del fago PaMx42, codificado en los ORFs 1 al 5, es similar al del fago Kakheta25. No obstante, mientras que el genoma del fago Kakheta25 presenta una endolisina en dicho módulo, el genoma de PaMx42 codifica una proteína putativa similar a glicosil hidrolasas del género *Burkholderia* (Anexo 4).

El módulo de lisis del fago PaMx25 es similar al del fago NP1, y a su vez, ambos muestran un arreglo parecido al del fago Kakheta25, a pesar de que la secuencia aminoacídica es distinta. En ese sentido, los genes codificantes de las proteínas putativas de membrana del fago PaMx25 tienen homólogos en los genomas de *Methylovorus*, *Novosphingobium* y *Xylella* (Anexo 4).

Por otro lado, los módulos de lisis de los fagos PaMx11, PaMx28 y PaMx74, mostraron un arreglo similar al de los módulos de lisis de los fagos Ab18 y AAT-1 respectivamente (Figura 4). Sin embargo, la endolisina putativa codificada en el genoma de PaMx11, presentó una secuencia similar a la del fago Kakheta25 (Anexo 4). Mientras que las endolisinas putativas codificadas en los genomas de PaMx28 y PaMx74 fueron similares a las del fago PY100 que infecta a *Yersinia* [Schwudke *et al.*, 2003]. La diversidad observada en los módulos de lisis de los fagos de estudio, tanto en posición como en secuencia aminoacídica, sugieren que dichos módulos pudieron haber sido transferidos horizontalmente, ya que de haber tenido un origen común, la señal de similitud a nivel de secuencia aminoacídica sería detectable como lo fue en el caso de los genes codificantes de las proteínas estructurales (Figura 6).

6.3.4 Replicación y reparación de DNA

Los ORFs asociados con la replicación de DNA en los fagos estudiados mostraron una gran diversidad ya que ninguno de los módulos identificados era idéntico entre ellos, sin embargo, se encontraron similitudes localizadas.

De modo tal que en el genoma de PaMx42 se encontraron seis ORFs, codificantes de DNA polimerasa, *clamp* replicativo. DNA helicasa, exonucleasa RecB, ATPasa RecA y primasa-helicasa (Anexo 4). Cada uno de estos seis ORFs presentaron homólogos en el genoma de PaMx25, no obstante, en dicho genoma se identificaron ORFs adicionales codificantes de una DNA topoisomerasa, DNA ligasa, RNAsa H-like, endonucleasa DenV y una helicasa UvrD (Anexo 4). A diferencia del módulo de replicación y reparación de DNA de PaMx42, el módulo correspondiente en el genoma de PaMx25 presentó varias interrupciones por otras secuencias codificantes, notablemente por un conjunto de ORFs asociados con la síntesis de derivados de 7-

deazaguanina (Sección 6.3.5 y Figura 4). El módulo de replicación y reparación de DNA de PaMx25, si bien mostró 6 ORFs homólogos con el de PaMx42, observamos que presentó un arreglo muy conservado con respecto al del colifago 9g (Figura 6), incluso en las regiones no homólogas con PaMx42.

En el caso del módulo de replicación del fago PaMx11, si bien presenta cuatro genes homólogos con el genoma del fago YuA, también se encontraron dos genes adicionales, que no están presentes en YuA.

Finalmente, en los genomas de PaMx28 y PaMx74 se identificaron cinco ORFs codificantes de proteínas putativas relacionados con el proceso de replicación: DNA polimerasa, DNA helicasa, primasa-polimerasa, DNA ligasa y una proteína de unión a DNA. Tres de estos ORFs mostraron similitud de secuencia aminoacídica con genes del fago RDJL Phi 1 de *Roseobacter* [Huang et al., 2011].

Los módulos de morfogénesis de los fagos de estudio presentaron un arreglo más o menos conservado al realizar comparaciones con otros fagos de *Pseudomonas* (ver sección 6.4.3). Por el contrario, los módulos de replicación y reparación de DNA mostraron una gran diversidad de ORFs que además presentaron similitud con secuencias pertenecientes a fagos que infectan otros géneros bacterianos, tales como *Escherichia* y *Roseobacter* (sección 6.4.3).

6.3.5 Metabolismo de nucleótidos

Los ORFs ubicados en esta categoría, fueron subdivididos en dos grandes grupos, aquellos genes que codifican enzimas putativas que participan en la biosíntesis de nucleótidos; y genes que codifican enzimas hidrolíticas específicas de nucleótidos. En el primer grupo, encontramos ribonucleótido reductasas (RNR) de clases Ia (*nrdA* y *nrdB*) y RNRs de clase II (*nrdJ*); dCMP deaminasas (*cd*), timidilato sintasas tipo *thyA* y *thyX*; y dihidrofolato reductasas (*frd*) (anexo 4).

La presencia de estos genes no es uniforme en los genomas estudiados, lo que sugiere que dichos genes podrían ser no esenciales. Asimismo, estos genes se localizaron de forma dispersa a lo largo de los genomas estudiados, inclusive, dentro de módulos no relacionados con metabolismo de nucleótidos. El genoma de PaMx25, mostró además un conjunto de ORFs continuos relacionados con la biosíntesis de queuosina, un nucleótido hipermodificado derivado de GTP. Tradicionalmente, se había asociado a la queuosina con la modificación de RNAs de transferencia, específicamente tRNAs con el anticodón GU[A,C,G,T][Morris & Elliot, 2001], sin embargo, recientemente se demostró que en bacteriófagos y otros elementos móviles, las enzimas codificadas en estos genes pueden modificar el DNA genómico con derivados de 7-deazaguanina [Thiaville et al., 2016]. En uno de los sistemas en donde se verificó dicha modificación, fue el genoma del fago 9g que infecta *Escherichia coli* [Kulikov et al., 2014].

La modificación del DNA con derivados de 7-deazaguanina, resulta en la resistencia del DNA a la digestión con enzimas de restricción. De forma similar, en el genoma de PaMx25 existen dichos genes, y de igual manera, su DNA es resistente al corte por endonucleasas cuando se sometió a la digestión con las enzimas *EcoRI*, *HinDIII* & *NdeI* [Sepúlveda-Robles et al., 2012]. Lo anterior es altamente sugerente del papel del grupo de genes asociado con la síntesis de derivados de 7-deazaguanina en el genoma de PaMx25.

En el segundo grupo de genes de metabolismo de nucleótidos, se tienen aquellos que codifican hidrolasas y pirofosfohidrolasas putativas específicas de nucleótidos trifosforilados (NTPasas y NPPasas, respectivamente). La diferencia principal entre ambos tipos de enzimas radica en que las NTPasas liberan nucleótidos bi fosforilados mientras que las NPPasas liberan nucleótidos monofosforilados. Hasta el momento no se sabe cual es la función que desempeñan las NPPasas en sistemas fágicos, sin embargo, dado que los ORFs codificantes de dichas enzimas fueron identificados en los cinco genomas estudiados, consideramos que podrían jugar un rol importante. Adicionalmente, los ORFs codificantes de NPPasas no mostraron una posición genómica conservada (Figura 4), lo cual podría sugerir que se trata de material genético transferido horizontalmente. Aunado a lo anterior, en el genoma de PaMx74 se lograron identificar dos ORFs codificantes de NPPasas en posiciones distintas del genoma, reforzando la noción de que dichos ORFs han sido transferidos horizontalmente. Por otro lado, los ORFs codificantes de NTPasas sólo fueron identificados en los genomas de los fagos PaMx11, PaMx25 y PaMx28. A diferencia de los ORFs codificantes de NPPasas, la presencia no uniforme de los ORFs codificantes de NTPasas sugiere que dichas enzimas podrían ser no esenciales.

6.4 Genómica comparativa

6.4.1 Comparaciones genómicas múltiples entre fagos de *Pseudomonas*

Durante el proceso de anotación genómica fue evidente que los fagos estudiados, mostraban similitud con secuencias previamente reportadas en bases de datos (anexo 4). Para averiguar cuales serían las secuencias más relacionadas con nuestros fagos de estudio, se procedió a realizar comparaciones globales múltiples empleando como métricas de similitud, el porcentaje global de identidad a nivel de secuencia nucleotídica, y el porcentaje de ORFs homólogos entre pares de genomas a nivel de secuencia aminoacídica.

Con la finalidad de validar el método empleado, se realizaron comparaciones globales entre fagos de *Pseudomonas* con clasificación taxonómica conocida (Anexo 3). A partir de estas comparaciones, notamos que la mayoría de los géneros propuestos y/o aceptados por el *International Comitee for the Taxonomy of Viruses*, mostraban porcentajes elevados tanto de ORFs homólogos como de identidad de secuencia nucleotídica (Anexo 3).

Considerando lo anterior, se logró la asignación taxonómica de los fagos de estudio. De modo que el fago PaMx42 se ubicó dentro del género *Septima3virus*; PaMx25 dentro del género *Np1virus*; los fagos PaMx28 y PaMx74 forman parte del género *Pamx74virus*; y el fago PaMx11 se ubicó dentro del género *Ab18virus* [Essoh et al., 2015].

A pesar de que las asociaciones internas en los géneros *Septima3virus*, *Np1virus*, *Pamx74virus* y *Ab18virus* son bastante sólidas, estos géneros muestran una proporción significativa de ORFs homólogos entre ellos, y entre el género *Yuavirus* [Ceyssens et al., 2008]. Esta observación nos hace suponer que estos géneros no están del todo aislados. Las comparaciones a nivel de secuencia nucleotídica, muestran que entre géneros no hay una asociación cercana, por lo que de existir una relación de ancestría entre ellos, no debería ser reciente. Adicionalmente, nos enfrentamos al problema de que si bien dos genomas pueden mostrar una proporción significativa de ORFs homólogos, su arreglo puede ser muy diferente, lo cual sugeriría que más allá de una relación evolutiva, ambos genomas pudieron haber experimentado transferencia horizontal de material genético.

6.4.2 Detección de regiones sinténicas

Con la finalidad de verificar el arreglo de los genes homólogos entre genomas, realizamos comparaciones pareadas empleando tBLASTx. Los mapas comparativos, muestran que los fagos en estudio, presentan regiones altamente sinténicas con los genomas previamente identificados en las comparaciones múltiples (Figura 4). De hecho, cada uno de los genomas comparados, mostraba una organización casi idéntica. Estos resultados, indican que los fagos en estudio, en efecto pertenecen a los géneros indicados en las comparaciones múltiples.

6.4.3 Comparaciones genómicas múltiples con fagos de otros géneros

El siguiente paso lógico consistió en la comparación de los fagos en estudio con fagos de otros géneros y/o que infectaran hospederos distintos a *Pseudomonas*. Para lograr este objetivo, se seleccionaron fagos que mostraran un porcentaje de ORFs homólogos superior a 20 con respecto de los genomas en estudio, y se realizaron las comparaciones de la misma forma que con los fagos de *Pseudomonas*.

Inicialmente, se construyó una matriz con los porcentajes de ORFs homólogos e identidad de secuencia nucleotídica entre fagos. Nuevamente, verificamos que los géneros identificados previamente mostraban una buena proporción de ORFs homólogos con otros géneros, esta vez sin embargo, no solamente detectamos ORFs homólogos con respecto del género *Yuavirus*, sino también con fagos de los géneros *Titanvirus* [Bollivar et al., 2016], *Nonagvirus* [Kulikov et al., 2014], *Seuratvirus* [Carstens et al., 2015], *Ssp2virus*, así como con fagos no clasificados (Figura 5).

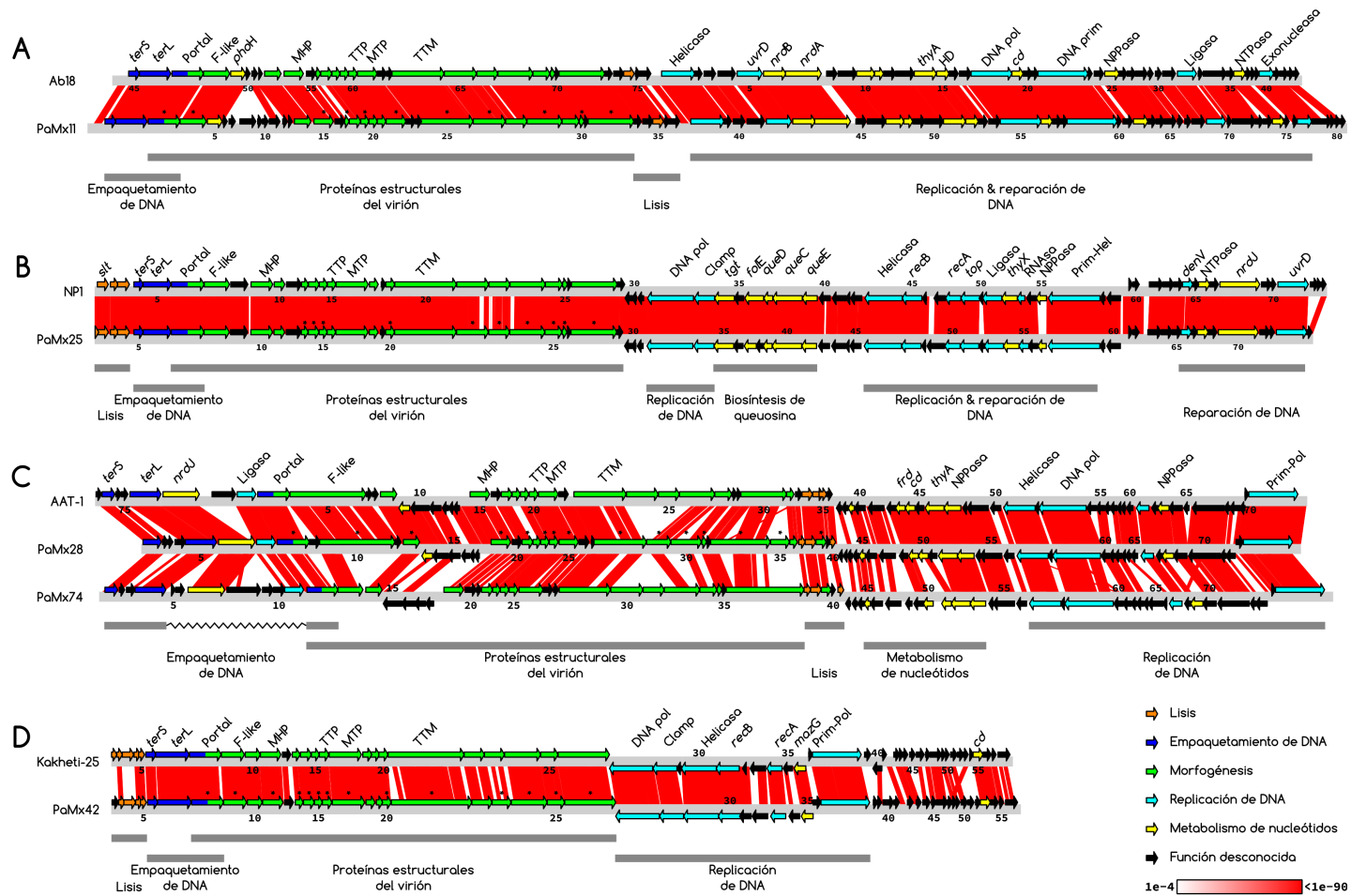


Figura 4: Comparaciones genómicas de los fagos estudiados contra genomas de referencia en cada género. A: *Ab18virus*. B: *Np1virus*. C: *Pamx74virus*. D: *Septima3virus*. Las comparaciones se realizaron empleando tBLASTx, los genomas se muestran como barras grises claras, los ORFs identificados como flechas de colores. La designación funcional se indica en la leyenda. Los bloques que conectan cada genoma representan similitud de secuencia aminoacídica, los colores de cada bloque corresponden a un e-value, la escala se muestra en la leyenda. Los módulos funcionales de cada genoma se muestran como barras oscuras debajo de cada comparación. Los genes mostrados son: *terS* y *terL*, subunidades menor y mayor de la terminasa; *phoH*, ATPasa parecida a PhoH; MHP, proteína mayor de cápside; TTP, proteína terminadora de cauda; MTP, proteína mayor de cauda; TTP, proteína de medida de la cauda; *uvrD*, helicasa UvrD; *nrdA*, *nrdB* & *nrdJ*, ribonucleotido reductasas; *thyA* & *thyX*, timidilato sintasas; *cd*, dCMP deaminasa; *slt*, transglicosilasa lítica; *tgt*, tRNA ribosil transferasa; *foIE*, GTP ciclohidrolase; *queD*, 6-piruvil tetrahidrobiopterin sintasa; *queC*, preQ0 sintasa; *queE*, 7-carboxi-7-deazaguanina sintasa; *recB*, exonucleasa RecB; *recA*, recombinasa A; *top*, DNA topoisomerasa; *denV*, glicosilasa de reparación de DNA; *frdA*, dihidrofolato reductasa; *mazG*, pirofosfohidrolasa parecida a MazG

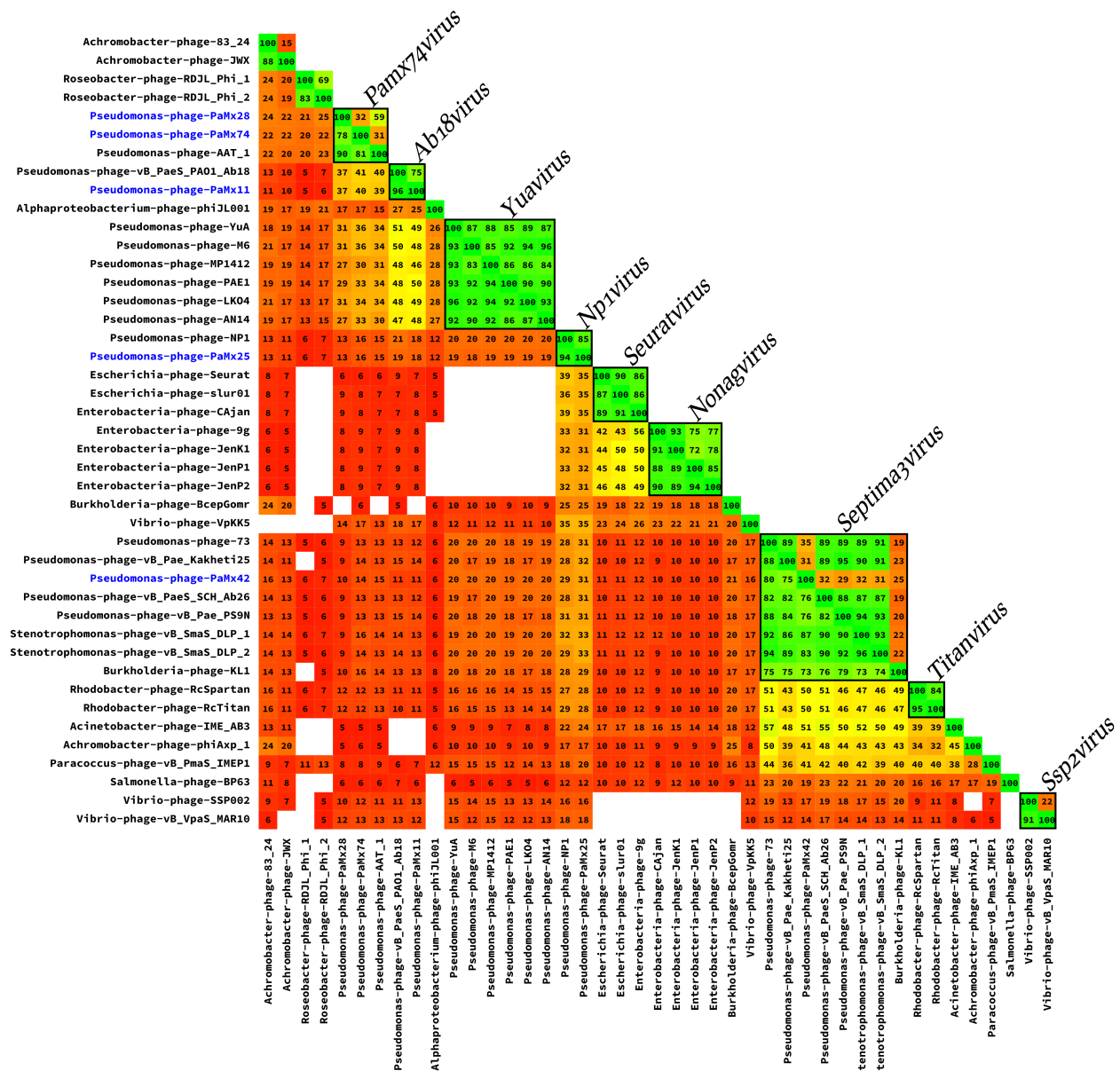


Figura 5: Comparación de los genomas de los fagos estudiados con genomas de fagos de otros géneros. A partir de las comparaciones recíprocas se construyó una diagonal separando la matriz en dos secciones. Por encima de la diagonal se muestran los porcentajes de identidad a nivel de secuencia nucleotídica entre genomas. Por debajo de la diagonal se muestran los porcentajes de ORFs homólogos. Los géneros reportados se muestran encerrados en recuadros negros. Como puede apreciarse, los fagos en estudio (azul), no solamente están relacionados con los géneros descritos anteriormente, sino también con fagos de otros géneros, e incluso con fagos que infectan hospederos distintos de *Pseudomonas*.

De modo tal que decidimos realizar las comparaciones pareadas entre genomas representativos que nos permitieran determinar, si existía una organización similar de sus genomas, que fuera indicativo de una posible relación de ancestría.

Dado que el genoma del fago PaMx74 resultó muy similar al del fago PaMx28, fue descartado de las comparaciones, de forma similar, el genoma del fago PaMx42 fue eliminado al claramente pertenecer al género *Septima3virus*. De modo tal que únicamente fueron comparados los genomas de los fagos de *Pseudomonas* PaMx11, PaMx25, PaMx28 y YuA; el fago 9g de *Escherichia*, y el fago RDJL Phi 1 de *Roseobacter*. (Figura 6).

Esta comparación mostró que el genoma de PaMx11 es muy similar al de YuA, sin embargo, las regiones en donde no es similar a YuA, es similar al genoma de PaMx28 en la región de genes estructurales (Figura 6 panel A). De igual manera, el genoma de PaMx25, es altamente sinténico con el genoma de YuA en la región de genes estructurales, pero no así en los módulos de replicación & reparación de DNA y de metabolismo de nucleótidos. No obstante, tanto el genoma de PaMx25 como el de PaMx28, muestran conexiones con otros genomas fágicos en las regiones de metabolismo de nucleótidos, y de replicación y reparación de DNA. Así, mientras que la primera mitad del genoma de PaMx25 muestra una relación directa con el genoma de YuA, la segunda mitad está claramente asociada con el genoma del fago 9g. De igual manera, el genoma de PaMx28, y por extensión el de PaMx74 (Figura 4), muestra una relación dual con el genoma de PaMx11 y el del fago RDJL Phi 1.

La comparación entre los genomas de YuA, RDJL Phi 1 y 9g (Figura 6 panel B) indica claramente que estos fagos no están relacionados, sin embargo, la inclusión de los genomas de PaMx11, PaMx25 y PaMx28, forman conexiones entre estos grupos. Este hallazgo indica que lo que normalmente concebimos como grupos aislados de fagos, posiblemente sea consecuencia de un muestreo insuficiente, ya que conforme más genomas fágicos son secuenciados, sería posible establecer más conexiones, quizá hasta el punto en el que los límites entre *grupos* sean imperceptibles.

6.5 Posición filogenómica de los fagos en estudio

Si bien no se tiene un mecanismo establecido para explicar el mosaicismo en los genomas fágicos, se ha postulado que la transferencia horizontal mediada por recombinación sea la responsable de dicho fenómeno. Lo anterior tiene como consecuencia que las relaciones evolutivas en los fagos posiblemente no sigan un modelo tradicional de herencia vertical, por lo que la inferencia evolutiva a través de árboles filogenéticos es hasta cierto punto limitada. En vez, se ha postulado que las asociaciones en los fagos podrían ser mejor explicadas a través de representaciones reticulares.

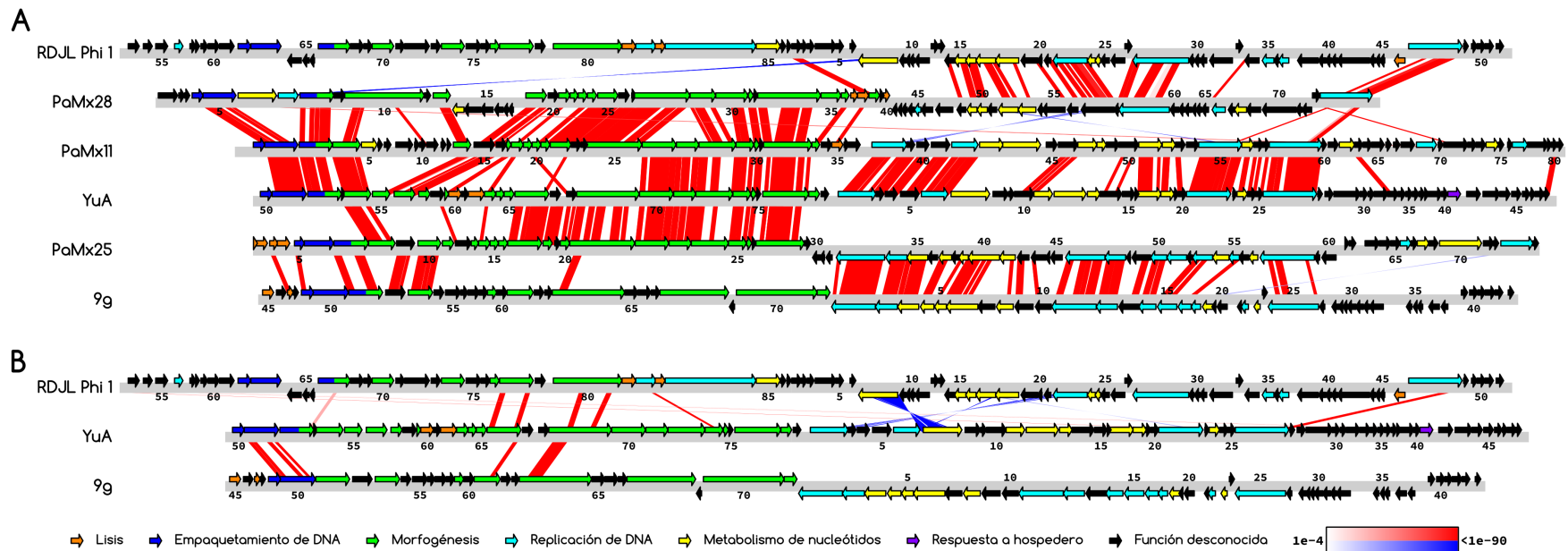


Figura 6: Comparaciones pareadas de los fagos en estudio. Las comparaciones se realizaron del mismo modo que en la figura 4. Los fagos PaMx11, PaMx25 y PaMx28 como conectores entre los géneros *Yuavirus* y *Nonagvirus*, representados por los fagos YuA y 9g respectivamente, así como con el fago RDJL Phi 1 (A). El mosaicismos en estos genomas es evidente, ya que en las regiones donde un par de genomas no son homólogos, lo son con su pares adyacentes. Los genomas de PaMx11, PaMx25 y PaMx28 forman conexiones entre YuA, 9g y RDJL Phi 1, los cuales de otra manera no muestran una relación clara y aparentan estar completamente separados (B)

En este estudio se construyó una red filogenómica con la finalidad de poder determinar cual es la posición de un genoma en un escenario donde podría estar asociado a más de un grupo. De este modo, una vez que se aplicó el algoritmo de disposición, se obtuvo la red mostrada en la figura 7. En ella, vemos que los genomas de los fagos PaMx25 y NP1, si bien están asociados muy cercanamente con los genomas del género *Septima3virus*, también lo están con los genomas de los géneros *Yuavirus*, *Nonagvirus* y *Seuratvirus*. Lo mismo ocurre con los genomas de los fagos PaMx11, PaMx28 y PaMx74, los cuales se ubican como grupos transitorios entre el género *Yuavirus* y los genomas de los fagos RDJL Phi 1 y RDJL Phi 2 de *Roseobacter*.

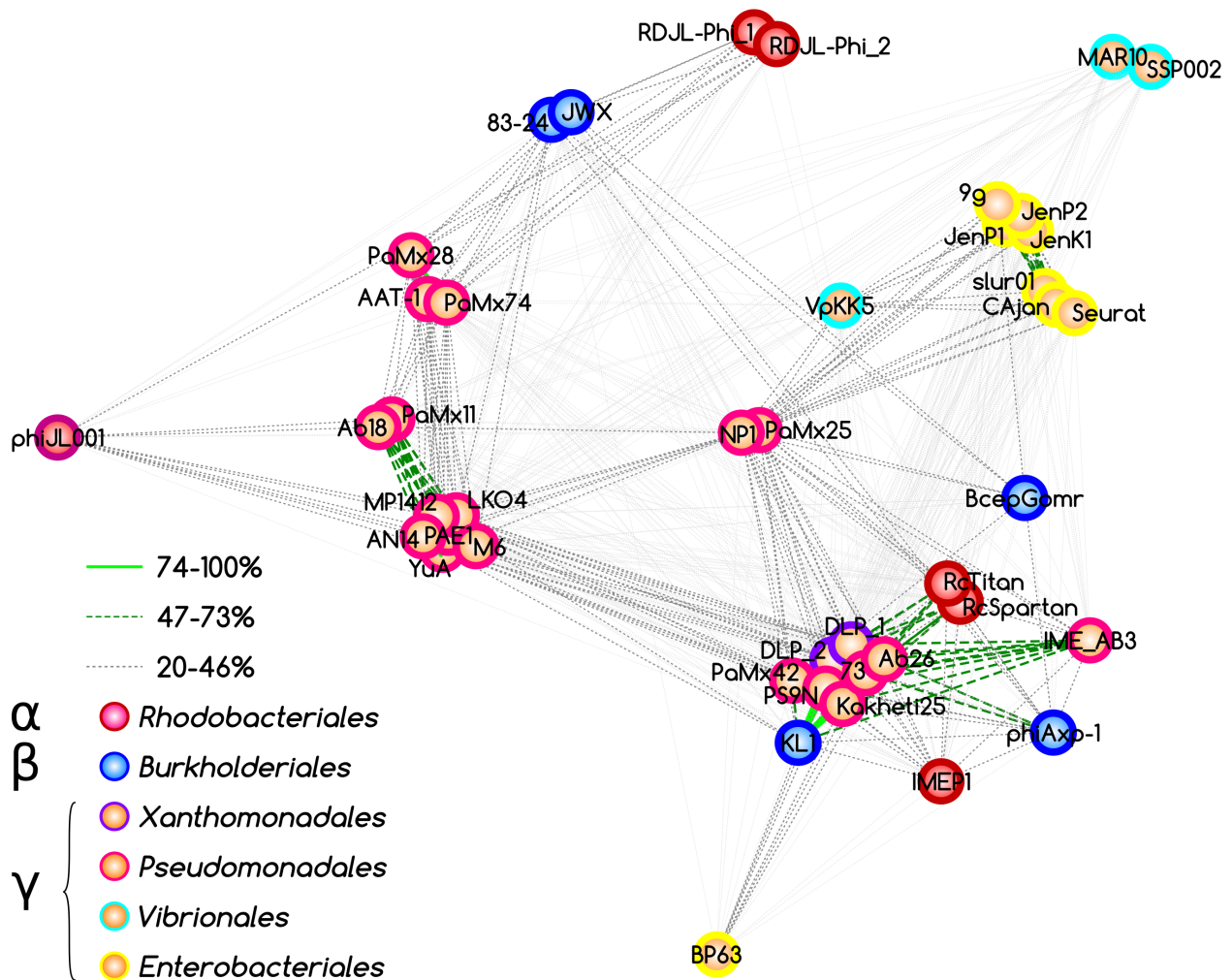


Figura 7: Red de asociación entre los genomas estudiados y genomas depositados en bases de datos. Cada nodo corresponde con un genoma, los nombres de los fagos están indicados junto a cada nodo. La clasificación taxonómica de los hospederos de los fagos fue codificada en color (ver leyenda). Las aristas representan la posible relación entre dos genomas, dada por el porcentaje de ORFs homólogos (el intervalo se muestra en la leyenda). La red fue sometida a un algoritmo de disposición ponderado en las aristas donde el porcentaje de ORFs homólogos hacía las veces de fuerza atractora entre genomas. Con la excepción del genoma de PaMx42, los genomas de los fagos en estudio se ubicaron en puntos intermedios entre géneros fágicos previamente descritos.

7. Discusión

Las comparaciones genómicas han sido ampliamente utilizadas para la descripción de fagos que podrían estar relacionados dado que infectan hospederos filogenéticamente cercanos [Desiere et al., 1999 & Desiere et al., 2001]. Desde las comparaciones empleando métodos experimentales como la hibridación de ácidos nucleicos, hasta la comparación de secuencias nucleotídicas empleando programas de cómputo [Botstein & Herskowitz, 1974; Juhala et al., 2000; Pedulla et al., 2003 & Sepúlveda-Robles et al., 2012]. Ya sea *in silico* o experimentalmente, se ha llegado a la conclusión de que los genomas fágicos son mosaicos que constan de segmentos provenientes de orígenes diversos.

En este trabajo, a partir de la comparación de secuencias, logramos detectar que nuestros fagos de estudio presentan genomas con un elevado grado de mosaicismo ya que aunque los genes estructurales de los cinco fagos presentaron un arreglo muy similar, los genes involucrados en el metabolismo de nucleótidos o de replicación de DNA no guardaban una relación detectable (Figuras 4 y 6). El mosaicismo detectado en los cinco genomas estudiados no solamente involucró a fagos de *Pseudomonas*, también se logró demostrar sintenia con genomas de fagos aparentemente no relacionados como 9g o RDJL Phi 1.

Los eventos que originan el mosaicismo en los genomas fágicos podrían darse por recombinación mediada por enzimas codificadas ya sea en los genomas fágicos o de los hospederos. No obstante, el intercambio de material genético en los fagos podría estar limitado por el rango de hospedero de los fagos involucrados. En ese sentido, los genomas empleados en este estudio, si bien infectan todos ellos a *Pseudomonas aeruginosa*, presentan dos características que apuntan hacia eventos no recientes. En primer lugar tenemos que ninguno de los fagos comparados muestra identidad de secuencia nucleotídica, suponiendo que los posibles eventos de intercambio de material genético entre ellos ocurrieran recientemente, se esperaría que mostraran cierto grado de identidad de secuencia nucleotídica. En segundo lugar, la presencia de regiones de sintenia entre fagos que infectan hospederos filogenéticamente separados, pero todos ellos pertenecientes al *phylum Proteobacteria*, podría sugerir una relación de ancestría paralela a la divergencia del *phylum Proteobacteria*.

Dado que los estudios de genómica comparativa realizados al momento en bacteriófagos únicamente han mostrado sintenia detectable sólo entre fagos cuyos hospederos pertenecen a la misma clase taxonómica [Desiere et al., 1999; Desiere et al., 2001; Mahony et al., 2013], resulta atractivo analizar más a fondo las posibles relaciones de nuestros fagos de estudio ya que los hospederos de dichos fagos abarcan un intervalo filogenético muy amplio al pertenecer a distintas clases taxonómicas.

La importancia de nuestros fagos no solamente radica en el intervalo filogenético cubierto, sino también en la formación de conexiones entre grupos de fagos que previamente se consideraban no relacionados. Esto es, los genomas de los fagos YuA, RDJL Phi 1 y 9g, no muestran identidad de secuencia nucleotídica entre sí, a su vez, presentan un reducido porcentaje de ORFs homólogos (Figura 5) y no muestran regiones de sintenia entre ellos (Figura 6). Sin embargo, el panorama cambia radicalmente cuando se incluyen en las comparaciones los genomas de los fagos PaMx11, PaMx25 y PaMx28, ya que estos genomas forman conexiones que permiten establecer una relación indirecta entre los fagos YuA, 9g y RDJL Phi 1 (Figuras 6 y 7). En otras palabras, la inclusión de los genomas de PaMx11, PaMx25 y PaMx28, desvanece los límites existentes entre grupos de fagos que previamente se consideraban aislados. Esta idea no es novedosa, anteriormente se había propuesto el modelo de un continuo genómico, en donde los grupos de fagos podrían ser únicamente consecuencia de una exploración limitada del espacio genómico que representan los fagos [Lawrence et al., 2002]. Recientemente, el modelo del continuo genómico fue abordado a través de la comparación de más de 600 genomas de *Mycobacterium*, y se observó que en efecto, existen genomas que podrían ser puntos de transición entre dos genomas que se consideraban no relacionados [Pope et al., 2015].

El resultado anterior, si bien marca un precedente en la genómica de bacteriófagos, presenta una limitación importante que radica en que todos los genomas comparados, provienen de fagos que infectan bacterias de una misma especie. De modo tal que el intercambio de material genético es esperable debido a que el rango de hospedero de dichos fagos no supone una barrera para el acceso a una poza común de material genético.

En el presente trabajo, el número de genomas comparados es considerablemente menor, sin embargo, se lograron detectar puntos de transición entre genomas. Lo anterior podría cambiar significativamente al incrementar el número de genomas por comparar. Esto es, es posible que conforme se secuencien más genomas fágicos, se logren desvanecer las barreras entre grupos de fagos que actualmente se consideran no relacionados, formando un verdadero continuo incluso entre fagos que infecten distintos hospederos.

8. Conclusiones

Se completaron las secuencias genómicas de los fagos PaMx11, PaMx25 y PaMx42. A pesar de que las secuencias de los fagos PaMx28 y PaMx74 no lograron completarse las comparaciones genómicas con fagos de referencia indican que las regiones no cubiertas fueron no significativas y las conclusiones de este trabajo permanecerán validas incluso después de que dichos genomas se completen.

La anotación funcional de los genomas estudiados arrojó un total de 160 genes con función asignada, siendo las categorías de genes estructurales y de metabolismo de nucleótidos las más representadas en dichos genomas. La combinación de las metodologías empleadas rindió buenos resultados dado que aunque muchos de los genes mostraron baja similitud contra secuencias en bases de datos, o ausencia de dominios conservados, el uso de modelado tridimensional y espectrometría de masas permitió la asignación de funciones a dichos genes.

Las comparaciones genómicas realizadas a nivel de secuencia nucleotídica revelaron que los genomas de estudio eran altamente similares a las secuencias de fagos dentro de géneros ya reportados. Esto indica que si bien nuestros fagos representan especies novedosas, forman parte de géneros que no existen de forma aislada en la naturaleza.

Las comparaciones genómicas mediante el porcentaje de ORFs homólogos, indicó claramente que los fagos en estudio presentaban relaciones con fagos de otros géneros, e incluso con fagos que infectan bacterias distintas de *Pseudomonas*. Si bien esta estrategia por si misma no permite inferir relaciones evolutivas, resulta útil en la selección de candidatos a ser comparados empleando metodologías más específicas.

Las comparaciones pareadas contra genomas de referencia mostraron que nuestros fagos de estudio presentan regiones altamente sinténicas con otros fagos que a su vez parecerían no estar relacionados. Esto tiene como consecuencia que aunque no se logre establecer *aún* un continuo genómico entre los fagos comparados, poco a poco se van desvaneciendo los límites entre grupos de fagos. Conforme se incluyan nuevas secuencias de más bacteriófagos, en algún momento podría ser factible establecer un modelo de continuidad genómica en donde los grupos de fagos sean indistinguibles.

Al momento no se cuenta con un modelo establecido que explique las relaciones evolutivas de los fagos en estudio. Sin embargo, la construcción de redes de similitud con los genomas en estudio, la determinación de regiones de sintenia y, la comparación de la clasificación taxonómica de los hospederos que infectan dichos fagos mostró dos aspectos relevantes:

- Las relaciones encontradas en nuestros fagos de estudio, abarcan un rango taxonómico muy amplio
 - Previamente sólo se habían logrado inferir relaciones en fagos que infectan hospederos dentro de la misma clase taxonómica, nuestros resultados lograron cubrir hospederos de distintas clases dentro de un mismo *phylum*.
- La ausencia de identidad a nivel de secuencia nucleotídica en los genomas comparados, indica que las posibles relaciones entre ellos, no son recientes. Al ponderar esta hipótesis con la separación filogenética de los hospederos de los fagos estudiados, resulta interesante proponer un modelo de coevolución.
 - En este modelo, podría haber existido un ancestro común a, por lo menos, los fagos comparados en este estudio
 - El ancestro de dicho ancestro común pudo haber sido una proteobacteria de un linaje previo, o muy cercano, a la separación en las clases α , β & γ .

9. Perspectivas

El presente trabajo abordó de forma sistemática las propiedades genómicas de cinco fagos novedosos que formaron conexiones genómicas entre grupos de fagos que previamente se consideraban no relacionados. A pesar del esfuerzo realizado se requiere del abordaje teórico y experimental para tener un panorama más completo de los fagos de estudio.

Dentro de las perspectivas contempladas se tienen las siguientes

- Completar las secuencias genómicas de los fagos PaMx74 y PaMx28
- Determinar experimentalmente el mecanismo de empaquetamiento de los fagos en estudio
- Identificar las proteínas estructurales del fago PaMx74
- Determinar bioquímicamente la actividad de las enzimas codificadas en los genomas de los fagos de estudio
- Determinar si las enzimas codificadas en los genomas analizados son esenciales para el desarrollo de los fagos de estudio
- Realizar comparaciones genómicas con nuevos genomas depositados en bases de datos con la finalidad de detectar nuevas posibles conexiones entre los genomas de los fagos de estudio y otros genomas fágicos
- Determinar si los fagos comparados en este estudio pudieran haber seguido un modelo co-evolutivo con sus hospederos.

10. Bibliografía

1. Ackerman HW, Bacteriophage Classification. In Bacteriophages Biology and Applications. 2005. Kutter E & Sulakvelidze A. (ed.). Boca Ratón, Florida.
2. Adriaenssens EM, Mattheus W, Cornelissen A, Shaburova O, Krylov VN, Kropinski AM, Lavigne R. Complete genome sequence of the giant *Pseudomonas* phage Lu11. J. Virol. 2012. 86(11):6369-70.
3. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997. 25(17):3389-402.
4. Andrade-Domínguez A, Kolter R. Complete Genome Sequence of *Pseudomonas aeruginosa* Phage AAT-1. Genome Announc. 2016. 4(4):e00165-16.
5. Bergh O, Børsheim KY, Bratbak G, Heldal M. High abundance of viruses found in aquatic environments. Nature. 1989. 340(6233):467-8.
6. Beumer A, Robinson JB. A broad-host-range, generalized transducing phage (SN-T) acquires 16S rRNA genes from different genera of bacteria. Appl. Environ. Microbiol. 2005. 71(12):8301-4.
7. Borodovsky M, Mills R, Besemer J, Lomsadze A. Prokaryotic gene prediction using GeneMark and GeneMark.hmm. Curr. Protoc. Bioinformatics. 2003. Chapter 4. Unit4.5.
8. Botstein D. A theory of modular evolution for bacteriophages. Ann. N. Y. Acad. Sci. 1980. 354:484-90.
9. Botstein D, Herskowitz I. Properties of hybrids between *Salmonella* phage P22 and coliphage lambda. Nature. 1974. 251(5476):584-9.
10. Breitbart M, Salamon P, Andresen B, Mahaffy JM, Segall AM, Mead D, Azam F, Rohwer F. Genomic analysis of uncultured marine viral communities. Proc. Natl. Acad. Sci. U.S.A. 2002. 99(22):14250-5.
11. Brüssow H. & Kutter E. Phage Ecology. In Bacteriophages Biology and Applications. 2005. Kutter E & Sulakvelidze A. (ed.). Boca Raton, Florida.
12. Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. Bioinformatics. 2012. 28(4):464-9.
13. Casjens S. Prophages and bacterial genomics: what have we learned so far? Mol. Microbiol. 2003. 49(2):277-300.
14. Casjens SR, Gilcrease EB, Winn-Stapley DA, Schicklmaier P, Schmieger H, Pedulla ML, Ford ME, Houtz JM, Hatfull GF, Hendrix RW. The generalized transducing *Salmonella* bacteriophage ES18: complete genome sequence and DNA packaging strategy. J Bacteriol. 2005. 187(3):1091-104.
15. Caspar DL & Klug A. Physical principles in the construction of regular viruses. Cold Spring Harb Symp Quant Biol. 1962. 27:1-24.
16. Chevreaux B, Wetter T, Suhai S. Genome Sequence Assembly Using Trace Signals and Additional Sequence. Information Computer Science and Biology: Proceedings of the German Conference on Bioinformatics (GCB). 1999. 45-56.
17. Chibani-Chennoufi S, Bruttin A, Dillmann ML, Brüssow H. Phage-host interaction: an ecological perspective. J. Bacteriol. 2004. 186(12):3677-86.

18. Cruz-Plancarte I, Cazares A, Guarneros G. Genomic and Transcriptional Mapping of PaMx41, Archetype of a New Lineage of Bacteriophages Infecting *Pseudomonas aeruginosa*. *Appl Environ Microbiol*. 2016. 82(22):6541-6547.
19. David M, Dzamba M, Lister D, Ilie L, Brudno M. SHRiMP2: sensitive yet practical SHort Read Mapping. *Bioinformatics*. 2011. 27(7):1011-2.
20. Delcher AL, Salzberg SL, Phillippy AM. Using MUMmer to identify similar regions in large sequence sets. *Curr. Protoc. Bioinformatics*. 2003. Chapter 10. Unit 10.3.
21. Demuth J, Neve H, Witzel KP. Direct electron microscopy study on the morphological diversity of bacteriophage populations in Lake Plußsee. *Appl. Environ. Microbiol*. 1993 59(10):3378-84.
22. Desiere F, Lucchini S, Brüßow H. Comparative sequence analysis of the DNA packaging, head, and tail morphogenesis modules in the temperate cos-site *Streptococcus thermophilus* bacteriophage Sfi21. *Virology*. 1999. 260(2):244-53.
23. Desiere F, Mahanivong C, Hillier AJ, Chandry PS, Davidson BE, Brüßow H. Comparative genomics of lactococcal phages: insight from the complete genome sequence of *Lactococcus lactis* phage BK5-T. *Virology*. 2001. 283(2):240-52.
24. Deutsch EW, Mendoza L, Shteynberg D, Slagel J, Sun Z, Moritz RL. Trans-Proteomic Pipeline, a standardized data processing pipeline for large-scale reproducible proteomics informatics. *Proteomics Clin. Appl*. 2015. 9(7-8):745-54.
25. Essoh C, Latino L, Midoux C, Blouin Y, Loukou G, Nguetta SP, Lathro S, Cablanmian A, Kouassi AK, Vergnaud G, Pourcel C. Investigation of a Large Collection of *Pseudomonas aeruginosa* Bacteriophages Collected from a Single Environmental Source in Abidjan, Côte d'Ivoire. *PLoS One*. 2015. 10(6):e0130548.
26. Fiers W, Contreras R, Duerinck F, Haegeman G, Iserentant D, Merregaert J, Min Jou W, Molemans F, Raeymaekers A, Van den Berghe A, Volckaert G, Ysebaert M. Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene. *Nature*. 1976. 260(5551):500-7.
27. Flores V, López-Merino A, Mendoza-Hernandez G, Guarneros G. Comparative genomic analysis of two brucellaphages of distant origins. *Genomics*. 2012. 99(4):233-40.
28. Flores V, Sepúlveda-Robles O, Cazarez A, Kameyama L, Guarneros G. Comparative genomic analysis of *Pseudomonas aeruginosa* phage PaMx25 reveals a novel siphovirus group related to phages infecting hosts of different taxonomic classes. *Arch Virol*. 2017. doi: 10.1007/s00705-017-3366-5.
29. Fuhrman JA. Bacterioplankton roles in cycling of organic matter: The microbial food web. In *Primary Productivity and Biogeochemical Cycles in the Sea*. 1992. Falkowski PG and Woodhead AD (ed.). Plenum, New York.
30. Grayson P, Molineux IJ. Is phage DNA 'injected' into cell? biologists and physicists can agree. *Curr. Opin. Microbiol*. 2007. 10(4):401-9.
31. Guttman B, Raya R & Kutter E. *Basichage Biology*. In *Bacteriophages Biology and Applications*. 2005. Kutter E & Sulakvelidze A. (ed.). Boca Raton, Florida.
32. Hatfull GF. Bacteriophage genomics. *Curr. Opin. Microbiol*. 2008. 11(5):447-53.
33. Hatfull GF, Hendrix RW. Bacteriophages and their genomes. *Curr. Opin. Virol*. 2011. 1(4):298-303.

34. Hendrix RW, Smith MC, Burns RN, Ford ME, Hatfull GF. Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. *Proc. Natl. Acad. Sci. U.S.A.* 1999. 96(5):2192-7.
35. Huang S, Zhang Y, Chen F, Jiao N. Complete genome sequence of a marine roseophage provides evidence into the evolution of gene transfer agents in alphaproteobacteria. *Virology*. 2011. 8:124.
36. Hyman P, Abedon ST. Bacteriophage host range and bacterial resistance. *Adv. Appl. Microbiol.* 2010. 70:217-48.
37. Isidro A, Henriques AO, Tavares P. The portal protein plays essential roles at different steps of the SPP1 DNA packaging process. *Virology*. 2004. 322(2):253-63.
38. Jang HB, Fagutao FF, Nho SW, Park SB, Cha IS, Yu JE, Lee JS, Im SP, Aoki T, Jung TS. Phylogenomic network and comparative genomics reveal a diverged member of the Φ KZ-related group, marine vibrio phage Φ JM-2012. *J. Virol.* 2013. 87(23):12866-78.
39. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R, Hunter S. InterProScan 5: genome-scale protein function classification. *Bioinformatics*. 2014. 30(9):1236-40.
40. Juhala RJ, Ford ME, Duda RL, Youlton A, Hatfull GF, Hendrix RW. Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. *J. Mol. Biol.* 2000. 299(1):27-51.
41. Karumidze N, Thomas JA, Kvatadze N, Goderdzishvili M, Hakala KW, Weintraub ST, Alavidze Z, Hardies SC. Characterization of lytic *Pseudomonas aeruginosa* bacteriophages via biological properties and genomic sequences. *Appl. Microbiol. Biotechnol.* 2012. 94(6):1609-17.
42. Kingsford CL, Ayanbule K, Salzberg SL. Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake. *Genome Biol.* 2007. 8(2):R22.
43. Krylov VN, Dela Cruz DM, Hertveldt K, Ackermann HW. "phiKZ-like viruses", a proposed new genus of myovirus bacteriophages. *Arch. Virol.* 2007. 152(10):1955-9.
44. Kulikov EE, Golomidova AK, Letarova MA, Kostryukova ES, Zelenin AS, Prokhorov NS, Letarov AV. Genomic sequencing and biological characteristics of a novel *Escherichia coli* bacteriophage 9g, a putative representative of a new *Siphoviridae* genus. *Viruses*. 2014. 6(12):5077-92.
45. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. Versatile and open software for comparing large genomes. *Genome Biol.* 2004. 5(2):R12.
46. Kwan T, Liu J, Dubow M, Gros P, Pelletier J. Comparative genomic analysis of 18 *Pseudomonas aeruginosa* bacteriophages. *J. Bacteriol.* 2006. 188(3):1184-7.
47. Latino L, Essoh C, Blouin Y, Vu Thien H, Pourcel C. A novel *Pseudomonas aeruginosa* bacteriophage, Ab31, a chimera formed from temperate phage PAJU2 and *P. putida* lytic phage AF: characteristics and mechanism of bacterial resistance. *PLoS One*. 2014. 9(4):e93777.
48. Lavigne R, Darius P, Summer EJ, Seto D, Mahadevan P, Nilsson AS, Ackermann HW, Kropinski AM. Classification of *Myoviridae* bacteriophages using protein sequence similarity. *BMC Microbiol.* 2009. 9:224.

49. Lavigne R, Seto D, Mahadevan P, Ackermann HW, Kropinski AM. Unifying classical and molecular taxonomic classification: analysis of the *Podoviridae* using BLASTP-based tools. *Res. Microbiol.* 2008. 159(5):406-14.
50. Lawrence JG, Hatfull GF, Hendrix RW. Imbroglios of viral taxonomy: genetic exchange and failings of phenetic approaches. *J. Bacteriol.* 2002. 184(17):4891-905.
51. Lebedev AA, Krause MH, Isidro AL, Vagin AA, Orlova EV, Turner J, Dodson EJ, Tavares P, Antson AA. Structural framework for DNA translocation via the viral portal protein. *EMBO J.* 2007. 26(7):1984-94.
52. Lima-Mendez G, Van Helden J, Toussaint A, Leplae R. Reticulate representation of evolutionary and functional relationships between phage genomes. *Mol. Biol. Evol.* 2008. 25(4):762-77.
53. Lindell D, Sullivan MB, Johnson ZI, Tolonen AC, Rohwer F, Chisholm SW. Transfer of photosynthesis genes to and from *Prochlorococcus* viruses. *Proc. Natl. Acad. Sci. U.S.A.* 2004 Jul. 101(30):11013-8.
54. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 1997. 25(5):955-64.
55. Mahony J, Martel B, Tremblay DM, Neve H, Heller KJ, Moineau S, van Sinderen D. Identification of a new P335 subgroup through molecular analysis of lactococcal phages Q33 and BM13. *Appl. Environ. Microbiol.* 2013. 79(14):4401-4409.
56. Morris RC, Elliott MS. Queuosine modification of tRNA: a case for convergent evolution. *Mol Genet Metab.* 2001. 74(1-2):147-59.
57. Ohnishi M, Kurokawa K, Hayashi T. Diversification of *Escherichia coli* genomes: are bacteriophages the major contributors? *Trends Microbiol.* 2001. 9(10):481-5.
58. Pedulla ML, Ford ME, Houtz JM, Karthikeyan T, Wadsworth C, Lewis JA, Jacobs-Sera D, Falbo J, Gross J, Pannunzio NR, Brucker W, Kumar V, Kandasamy J, Keenan L, Bardarov S, Kriakov J, Lawrence JG, Jacobs WR Jr, Hendrix RW, Hatfull GF. Origins of highly mosaic mycobacteriophage genomes. *Cell.* 2003. 113(2):171-82.
59. Pope WH, Bowman CA, Russell DA, Jacobs-Sera D, Asai DJ, Cresawn SG, Jacobs WR, Hendrix RW, Lawrence JG, Hatfull GF, Science Education Alliance Phage Hunters Advancing Genomics and Evolutionary Science, Phage Hunters Integrating Research and Education, Mycobacterial Genetics Course. Whole genome comparison of a large collection of mycobacteriophages reveals a continuum of phage genetic diversity. *Elife.* 2015. 4.
60. Proctor L & Fuhrman JA. Viral mortality of marine bacteria and cyanobacteria. *Nature.* 1990. 343:60-62.
61. Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, Fiddes CA, Hutchison CA, Slocombe PM, Smith M. Nucleotide sequence of bacteriophage phi X174 DNA. *Nature.* 1977. 265(5596):687-95.
62. Sepúlveda-Robles, Kameyama L, Guarneros G. High diversity and novel species of *Pseudomonas aeruginosa* bacteriophages. *Appl. Environ. Microbiol.* 2012. 78(12):4510-4515.
63. Sun S, Kondabagil K, Draper B, Alam TI, Bowman VD, Zhang Z, Hegde S, Fokine A, Rossmann MG, Rao VB. The structure of the phage T4 DNA packaging motor suggests a mechanism dependent on electrostatic forces. *Cell.* 2008. 135(7):1251-62.
64. Suttle CA. Viruses in the sea. *Nature.* 2005. 437(7057):356-61.

65. Suzek BE, Ermolaeva MD, Schreiber M, Salzberg SL. A probabilistic method for identifying start codons in bacterial genomes. *Bioinformatics*. 2001. 17(12):1123-30.
66. Sváb D, Bálint B, Maróti G, Tóth I. A novel transducible chimeric phage from *Escherichia coli* O157:H7 Sakai strain encoding Stx1 production. *Infect. Genet. Evol.* 2015. 29:42-7.
67. Thiaville JJ, Kellner SM, Yuan Y, Hutinet G, Thiaville PC, Jumpathong W, Mohapatra S, Brochier-Armanet C, Letarov AV, Hillebrand R, Malik CK, Rizzo CJ, Dedon PC, de Crécy-Lagard V. Novel genomic island modifies DNA with 7-deazaguanine derivatives. *Proc. Natl. Acad. Sci. U.S.A.* 2016. 113(11):E1452-9.
68. Torrella F, Morita RY. Evidence by electron micrographs for a high incidence of bacteriophage particles in the waters of Yaquina Bay, Oregon: ecological and taxonomical implications. *Appl. Environ. Microbiol.* 1979. 37(4):774-8.
69. Weinbauer MG, Rassoulzadegan F. Are viruses driving microbial diversification and diversity? *Environ. Microbiol.* 2004. 6(1):1-11.
70. White HE, Sherman MB, Brasilès S, Jacquet E, Seavers P, Tavares P, Orlova EV. Capsid structure and its stability at the late stages of bacteriophage SPP1 assembly. *J Virol.* 2012. 86(12):6768-77.
71. Wommack KE, Hill RT, Kessel M, Russek-Cohen E, Colwell RR. Distribution of viruses in the Chesapeake Bay. *Appl. Environ. Microbiol.* 1992. 58(9):2965-70.
72. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: protein structure and function prediction. *Nat. Methods*. 2015. 12(1):7-8.
73. Young R. Phage lysis: three steps, three choices, one outcome. *J. Microbiol.* 2014. 52(3):243-58.
74. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 2008. 18(5):821-9.
75. Zhao H, Christensen TE, Kamau YN, Tang L. Structures of the phage Sf6 large terminase provide new insights into DNA translocation and cleavage. *Proc. Natl. Acad. Sci. U.S.A.* 2013. 110(20):8075-80.
76. Zhao H, Finch CJ, Sequeira RD, Johnson BA, Johnson JE, Casjens SR, Tang L. Crystal structure of the DNA-recognition component of the bacterial virus Sf6 genome-packaging machine. *Proc. Natl. Acad. Sci. U.S.A.* 2010. 107(5):1971-6.

11. Anexos

Anexo 1. Genomas de fagos de *Pseudomonas* empleados en este trabajo

Fago	Número de acceso	Familia	Género	Longitud del genoma	
EL	NC_007623	<i>Myoviridae</i>	<i>Elvirus</i>	211215	
OBP	NC_016571			284757	
CHA_P1	NC_022974			88255	
KPP10	NC_015272			88322	
P3_CHA	KC862296			88097	
PAK_P3	NC_022970			88097	
PAK_P5	NC_022966			88135	
vB_PaeM_PAO1_Ab03	NC_026587			<i>Kpp10virus</i>	86246
vB_PaeM_PAO1_Ab04	LN610581				86668
vB_PaeM_PAO1_Ab06	LN610582				84759
vB_PaeM_PAO1_Ab17	LN610576	83598			
vB_PaeM_PS24	NC_028882	84583			
phiCTX	NC_003278	<i>Myoviridae</i>	<i>P2virus</i>	35580	
C11	NC_028652	<i>Myoviridae</i>	<i>Pakpunavirus</i>	94109	
JG004	NC_019450			93017	
K5	NC_030910			93754	
K8	NC_028817			93879	
PA10	KY000083			91212	
PAK_P1	NC_015294			93198	
PAK_P2	NC_022967			92495	
PAK_P4	NC_022986			93147	
PaoP5	NC_029083			93464	
PaP1	NC_019913			91715	
phiMK	NC_031110			93129	
vB_PaeM_C2-10_Ab02	LN610572			93848	
vB_PaeM_C2-10_Ab08	LN610575			93503	
vB_PaeM_C2-10_Ab1	NC_019918			92777	
Zigelbrucke	KY073228			92338	
14/01/17	NC_011703	66235			
DL52	KR054028	65867			
DL60	NC_028745	66103			
DL68	NC_028971	66111			
F8	NC_007810	66015			
Gallinipper	KT372690	<i>Myoviridae</i>	<i>Pbunavirus</i>	65917	
JG024	NC_017674			66275	
Jollyroger	KT372691			65795	
KPP12	NC_019935			64144	
KPP22	LC105987			64415	

Anexo 1. Genomas de fagos de *Pseudomonas* empleados en este trabajo

Fago	Número de acceso	Familia	Género	Longitud del genoma
KPP22M1	LC105988			64415
KPP22M2	LC105989			64415
KPP22M3	LC105990			64415
Kraken	KT372692			65762
Kula	KT372693			65762
LBL3	NC_011165			64427
LMA2	NC_011166			66530
Nemo	KT372694			66283
Nessie	KT372695			65762
NH-4	NC_019451			66116
NP3	KU198331			66063
PA5	KY000082			66182
PB1	NC_011810	<i>Myoviridae</i>	<i>Pbunavirus</i>	65764
phiKT28	KP340287			66381
phiKTN6	KP340288			65994
Poseidon	KT372696			65762
S12-1	LC102730			66257
Smee	KT372697			66278
SN	NC_011756			66390
SPM-1	KF981875			65729
Triton	KT372698			65762
vB_Pae_PS44	NC_028939			68871
vB_PaeM_CEB_DP1	KR869157			66158
vB_PaeM_PAO1_Ab27	NC_026586			66299
201phi2-1	NC_010821			316674
KTN4	KU521356			279593
phiKZ	NC_004629	<i>Myoviridae</i>	<i>Phikzvirus</i>	280334
PhiPA3	NC_028999			309208
Lu11	NC_017972	<i>Myoviridae</i>	No asignado	280538
PaBG	NC_022096	<i>Myoviridae</i>	No asignado	258139
phi3	NC_030940	<i>Myoviridae</i>	No asignado	32637
phiPsa374	NC_023601	<i>Myoviridae</i>	No asignado	97906
vB_PsyM_KIL1	NC_030934	<i>Myoviridae</i>	No asignado	90552
vB_PsyM_KIL2	KU130127	<i>Myoviridae</i>	No asignado	92466
vB_PsyM_KIL3	KU130128	<i>Myoviridae</i>	No asignado	92068
vB_PsyM_KIL3b	KU130131	<i>Myoviridae</i>	No asignado	92095
vB_PsyM_KIL4	KU130129	<i>Myoviridae</i>	No asignado	92816
vB_PsyM_KIL5	KU130130	<i>Myoviridae</i>	No asignado	93385

Anexo 1. Genomas de fagos de *Pseudomonas* empleados en este trabajo

Fago	Número de acceso	Familia	Género	Longitud del genoma
F116	NC_006552			65195
H66	KC262634	<i>Podoviridae</i>	<i>F116virus</i>	65270
LKA5	KC900378			64746
DL64	NC_028885			72378
LIT1	NC_013692			72544
Pa2	NC_027345			73008
PA26	JX194238			72321
PEV2	NC_031063			72697
phi176	KM411960	<i>Podoviridae</i>	<i>Lit1virus</i>	73048
RWG	KM411958			72646
vB_PaeP_C2-10_Ab09	NC_024140			72028
vB_PaeP_MAG4	NC_031104			72979
YH30	NC_029101			72192
YH6	NC_027388			73050
DL54	NC_028919			45673
LUZ24	NC_010325			45625
MR299-2	JN254801			44789
PaP3	NC_004466			45503
PhiCHU	NC_028933	<i>Podoviridae</i>	<i>Luz24virus</i>	45626
phiIBB-PAA2	NC_022971			45344
TL	NC_023583			45696
vB_PaeP_C1-14_Or	HE983844			45469
vB_PaeP_C2-10_Ab22	NC_026599			45808
vB_PaeP_p2-10_Or1	NC_019813			44030
KPP21	NC_029017	<i>Podoviridae</i>	<i>Luz7virus</i>	73420
LUZ7	NC_013691			74901
DL62	NC_028836			42508
LKD16	NC_009935			43200
LUZ19	NC_010326			43548
MPK6	NC_022746			42957
MPK7	NC_022091			42874
phikF77	NC_012418			43152
phiKMV	NC_005045	<i>Podoviridae</i>	<i>Phikmvvirus</i>	42519
phiNFS	KU743887			42351
PT2	NC_011107			42961
PT5	EU056923			42954
vB_Pae-TbilisiM32	NC_017865			42966
vB_PaeP_PAO1_1-15pyo	LN610580			42750
vB_PaeP_PAO1_Ab05	NC_026602			43639
vB_PaeP_PPA-ABTNL	NC_027375			43227

Anexo 1. Genomas de fagos de *Pseudomonas* empleados en este trabajo

Fago	Número de acceso	Familia	Género	Longitud del genoma
LKA1	NC_009936			41593
phi-2	NC_013638	<i>Podoviridae</i>	<i>Phikmvvirus</i> no clasificados	43144
YMC11/06/C171_PPU_BP	NC_030923			45417
Andromeda	NC_031014			40008
Bf7	NC_016764	<i>Podoviridae</i>	<i>Phikmvvirus</i> no clasificados	40058
VSW-3	KX066068			40556
gh-1	NC_004665			37359
Pf-10	NC_027292			39167
Phi-S1	NC_021062			40192
phi15	NC_015208			39562
phiIBB-PF7A	NC_015264			40973
phiPsa17	KR091952	<i>Podoviridae</i>	<i>T7virus</i>	40525
phiPSA2	NC_024362			40472
PPPL-1	NC_028661			41149
UNO-SLW1	KX431888			39215
UNO-SLW2	KX449361			39167
UNO-SLW3	KX449362			39092
UNO-SLW4	KX449363			39136
119X	NC_007807	<i>Podoviridae</i>	No asignado	43365
AF	NC_019923	<i>Podoviridae</i>	No asignado	42689
PaMx33	KU884561	<i>Podoviridae</i>	No asignado	43265
PaMx35	KU884562	<i>Podoviridae</i>	No asignado	43733
PaMx41	KU884563	<i>Podoviridae</i>	No asignado	43490
PaMx43	KU884564	<i>Podoviridae</i>	No asignado	43223
PaMx46	KU884565	<i>Podoviridae</i>	No asignado	43266
PaP2	NC_005884	<i>Podoviridae</i>	No asignado	43783
R18	LC102729	<i>Podoviridae</i>	No asignado	63560
tf	NC_017971	<i>Podoviridae</i>	No asignado	46271
UFV-P2	NC_018850	<i>Podoviridae</i>	No asignado	45517
PaMx11	NC_028770			59878
vB_PaeS_PAO1_Ab18	NC_026594	<i>Siphoviridae</i>	<i>Ab18virus</i>	56537
D3112	NC_005178			37611
DMS3	NC_008717			36415
FHA0480	JN808773			37374
H70	NC_027384	<i>Siphoviridae</i>	<i>D3112virus</i>	37359
JBD24	NC_020203			37095
JBD26	JN811560			37840
JBD30	NC_020198			36947
JBD5	NC_020202			37740

Anexo 1. Genomas de fagos de *Pseudomonas* empleados en este trabajo

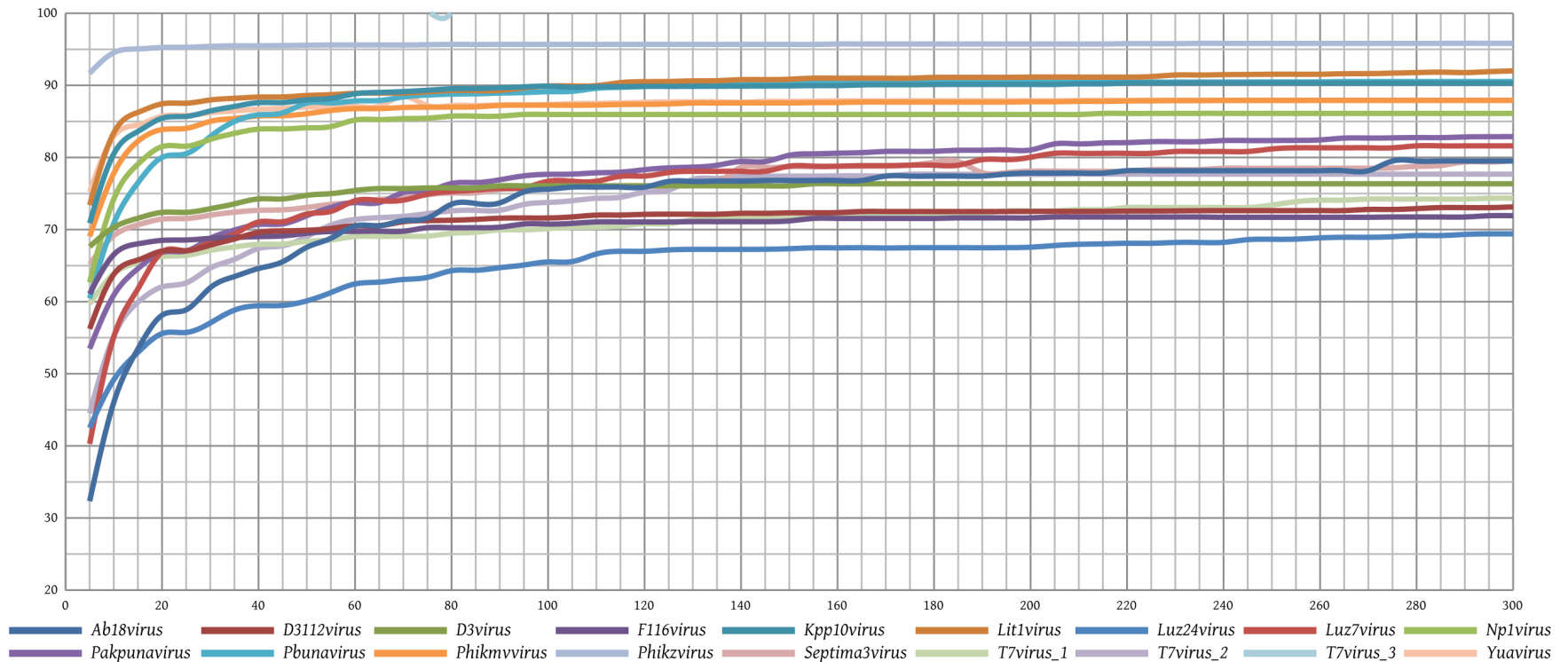
Fago	Número de acceso	Familia	Género	Longitud del genoma
JBD69	NC_030908			36938
JBD88a	NC_020200			36429
JBD93	NC_030918			36629
JD024	NC_024330			37380
MP22	NC_009818			36409
MP29	NC_011613			36632
MP38	NC_011611	<i>Siphoviridae</i>	<i>D3112virus</i>	36885
MP42	NC_018274			36847
MP48	NC_024782			36838
PA1phi	NC_023700			34553
PaMx73	JQ067085			36570
vB_PaeS_PAO1_Ab30	NC_026601			37238
YMC11/11/R1836	KT968832			37714
D3	NC_002484	<i>Siphoviridae</i>	<i>D3virus</i>	56426
PMG1	NC_016765			54024
PaMx28				
PaMx74		<i>Siphoviridae</i>	<i>Pamx74virus</i>	
AAT-1	KU204984			57599
NP1	NC_031058	<i>Siphoviridae</i>	<i>Np1virus</i>	58566
PaMx25	JQ067084			57899
73	NC_007806			42999
PaMx42	NC_028879			43225
vB_Pae_PS9N	KM434185	<i>Siphoviridae</i>	<i>Septima3virus</i>	43047
vB_Pae-Kakheti25	NC_017864			42844
vB_PaeS_SCH_Ab26	NC_024381			43056
AN14	KX198613			60973
LKO4	KC758116			61818
M6	NC_007809	<i>Siphoviridae</i>	<i>Yuavirus</i>	59446
MP1412	NC_018282			61167
PAE1	NC_028980			62181
YuA	NC_010116			58663
B3	NC_006548	<i>Siphoviridae</i>	No asignado	38439
F10	NC_007805	<i>Siphoviridae</i>	No asignado	39199
JBD44	NC_030929	<i>Siphoviridae</i>	No asignado	49033
MD8	NC_031091	<i>Siphoviridae</i>	No asignado	43277
PAJU2	NC_011373	<i>Siphoviridae</i>	No asignado	46872
phi1	KT887557	<i>Siphoviridae</i>	No asignado	57218
phi2	NC_030931	<i>Siphoviridae</i>	No asignado	41871
phi297	NC_016762	<i>Siphoviridae</i>	No asignado	49135

Anexo 1. Genomas de fagos de *Pseudomonas* empleados en este trabajo

Fago	Número de acceso	Familia	Género	Longitud del genoma
phiPSA1	NC_024365	<i>Siphoviridae</i>	No asignado	51090
SM1	KU245542	<i>Siphoviridae</i>	No asignado	93191
YMC/01/01/P52_PAE_BP	JX403939	<i>Siphoviridae</i>	No asignado	49381
YMC11/02/R656	NC_028657	<i>Siphoviridae</i>	No asignado	60919
YMC11/07/P54_PAE_BP	NC_030909	<i>Siphoviridae</i>	No asignado	50183
JBD18	NC_027986	<i>Siphoviridae</i>	No asignado	39014
JBD25	NC_027992	<i>Siphoviridae</i>	No asignado	39552

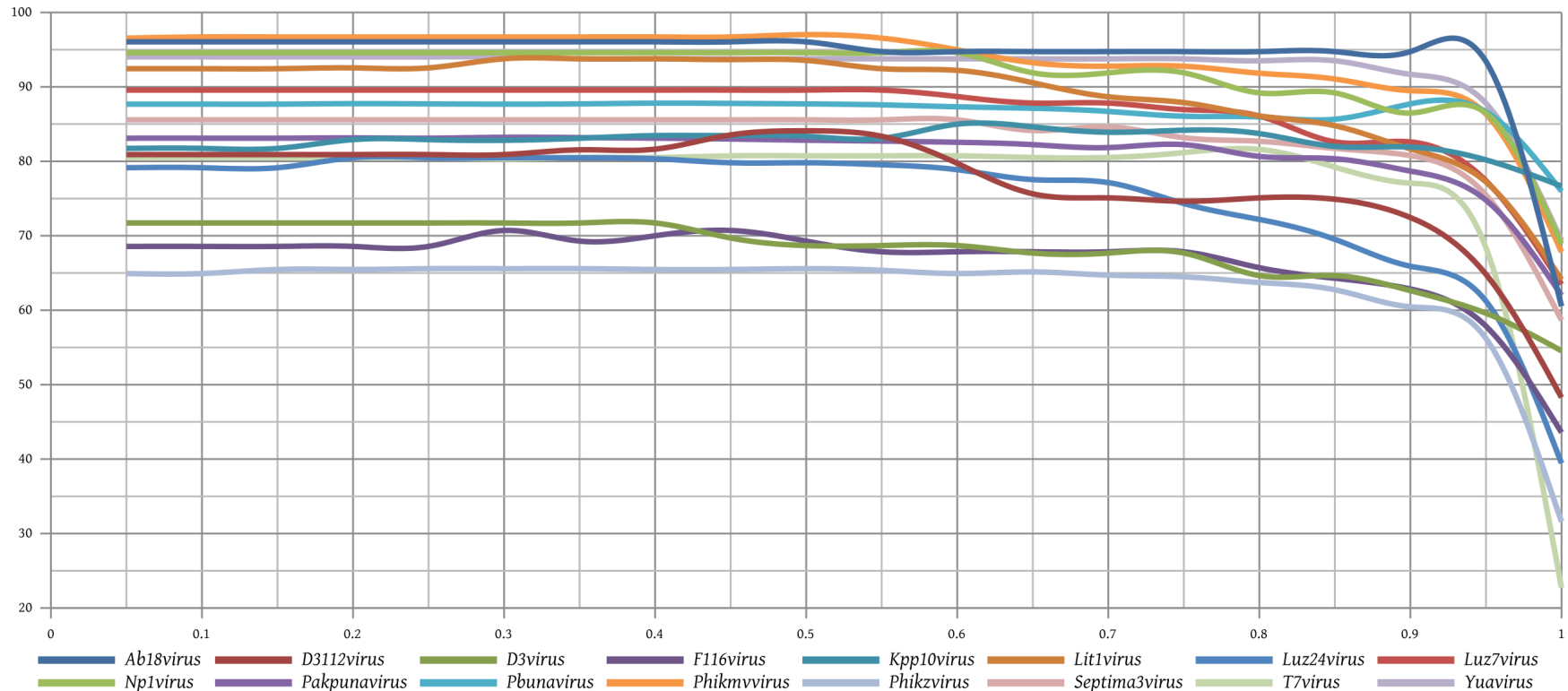
Anexo 2. Optimización de los parámetros empleados en las comparaciones globales

Comparación a nivel de nucleótidos: longitud de corte



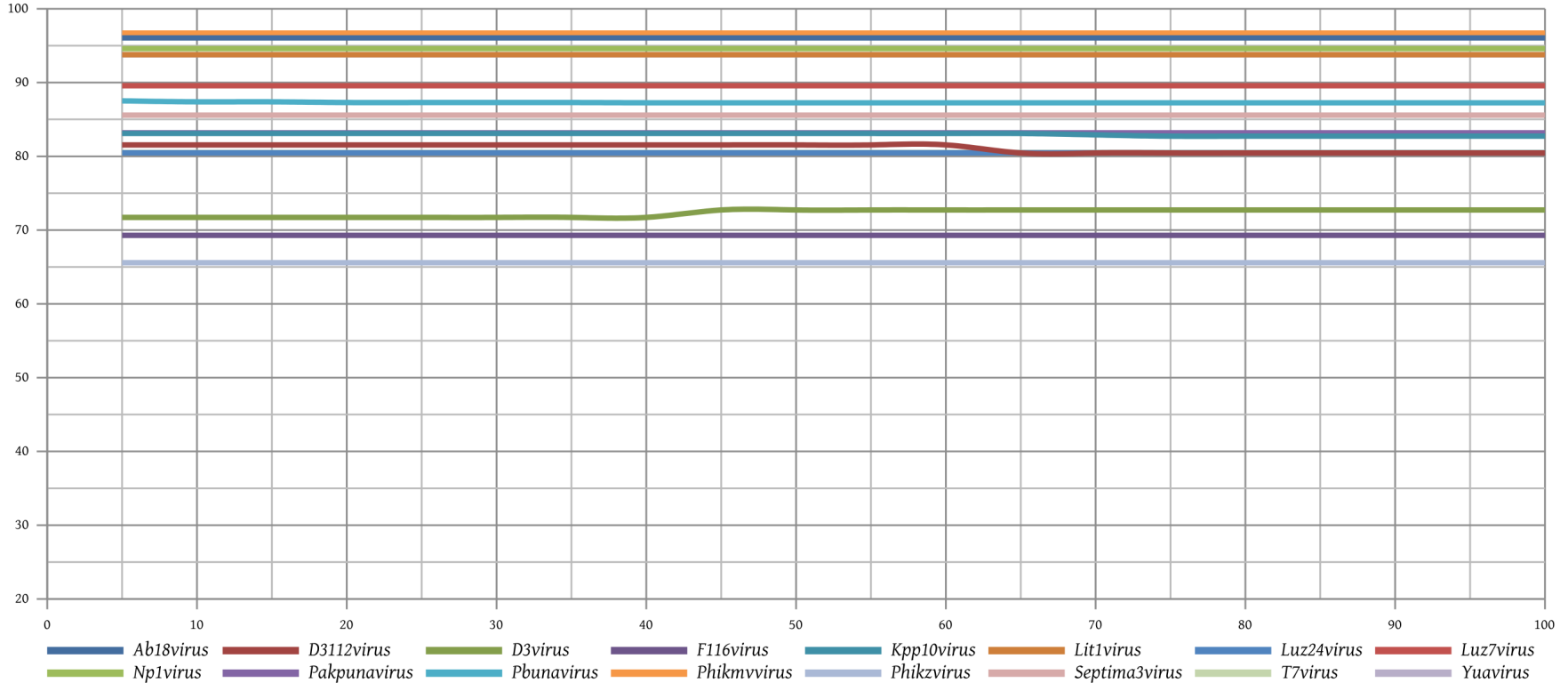
En el eje de las abscisas se indica la longitud de quiebre en las comparaciones múltiples usando MUMmer. Las ordenadas representan el porcentaje de identidad promedio observado entre los integrantes de cada grupo de fagos (anexo 1) comparados con el genoma de referencia de dicho grupo. Se eligió el valor de 120 residuos debido a que a partir de este punto, el incremento en el valor de corte, no representaba una ganancia proporcional en el porcentaje de identidad global.

Anexo 2: Porcentaje de ORFs homólogos: comparación de las longitudes de los posibles ORFs homólogos



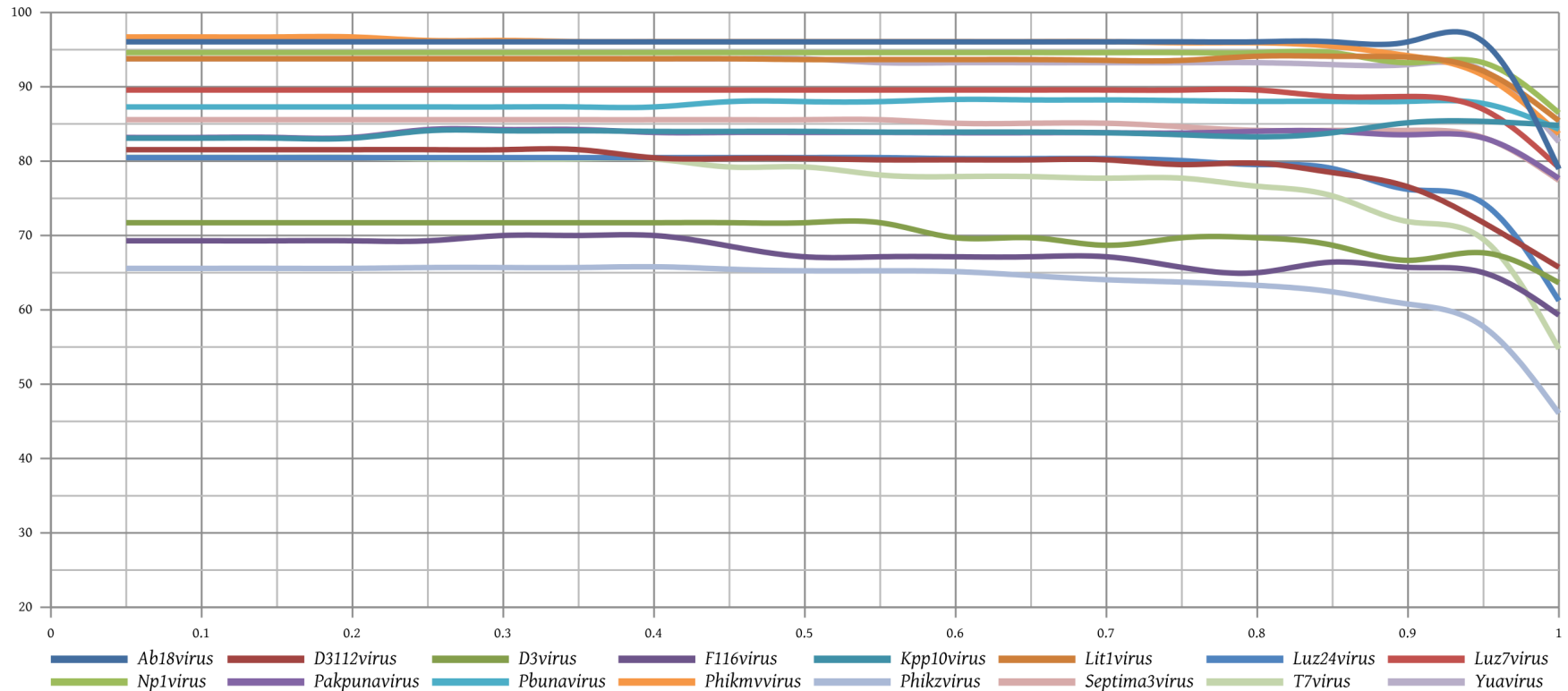
Para considerar un par de ORFs como homólogos, primeramente se optimizó el umbral del cociente de la longitud del ORF más corto entre la longitud del ORF más largo. En el eje de las abscisas se muestra el valor de corte para el cociente mencionado, los pares de ORFs que mostraran un cociente de sus longitudes por debajo del valor de corte, eran descartados y no se consideraban homólogos. Las ordenadas representan el porcentaje de ORFs homólogos promedio observado entre los integrantes de cada grupo de fagos (anexo 1) comparados con el genoma de referencia de dicho grupo. Se eligió el valor de umbral de 0.35 ya que en este punto se mantenía estable el porcentaje de ORFs homólogos y no se comprometía el filtrado por el cociente entre la longitud del alineamiento entre los posibles ORFs homólogos entre la longitud del ORF más corto.

Anexo 2: Porcentaje de ORFs homólogos: Tolerancia a la presencia de repetidos entre los posibles ORFs homólogos



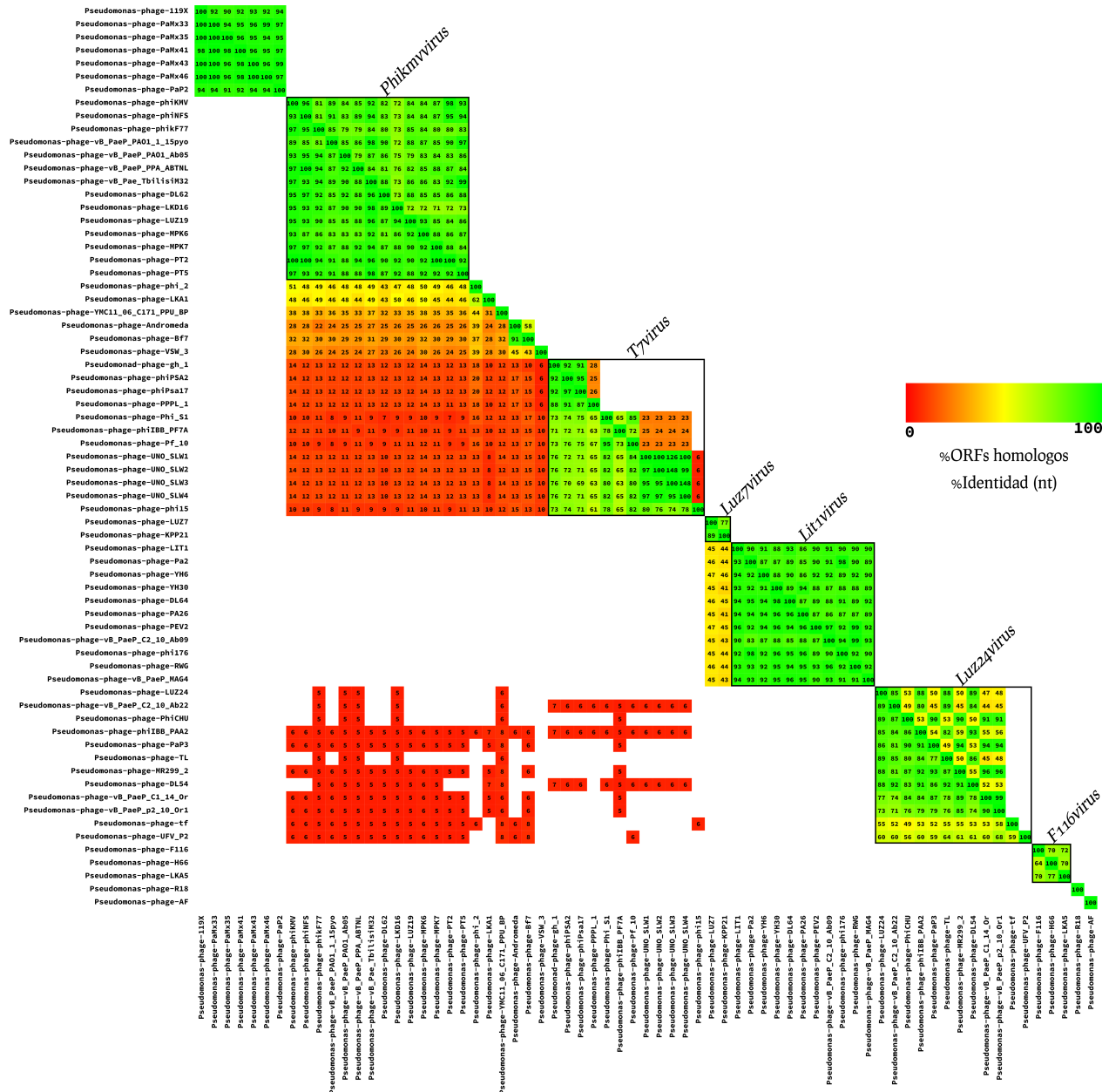
Dado que la presencia de regiones repetitivas podría incrementar los scores obtenidos en las comparaciones aminoacídicas usando BLASTp, se aplicó un filtro para descartar dichas regiones repetitivas en caso de existir. En el eje de las abscisas se muestra el valor de corte para la tolerancia a repetidos entre los posibles ORFs homólogos. Aquellos pares de ORFs que mostraran regiones repetitivas con una longitud superior a la del valor de corte, fueron penalizadas restando dicha longitud repetitiva a la longitud del alineamiento. Las ordenadas representan el porcentaje de ORFs homólogos promedio observado entre los integrantes de cada grupo de fagos (anexo 1) comparados con el genoma de referencia de dicho grupo. Se eligió el valor de umbral de 35 residuos, sin embargo, en la mayoría de los grupos evaluados, la presencia de repetidos tuvo una aportación no significativa.

Anexo 2: Porcentaje de ORFs homólogos: comparación de la longitud del alineamiento de los posibles ORFs homólogos con la longitud de los mismos



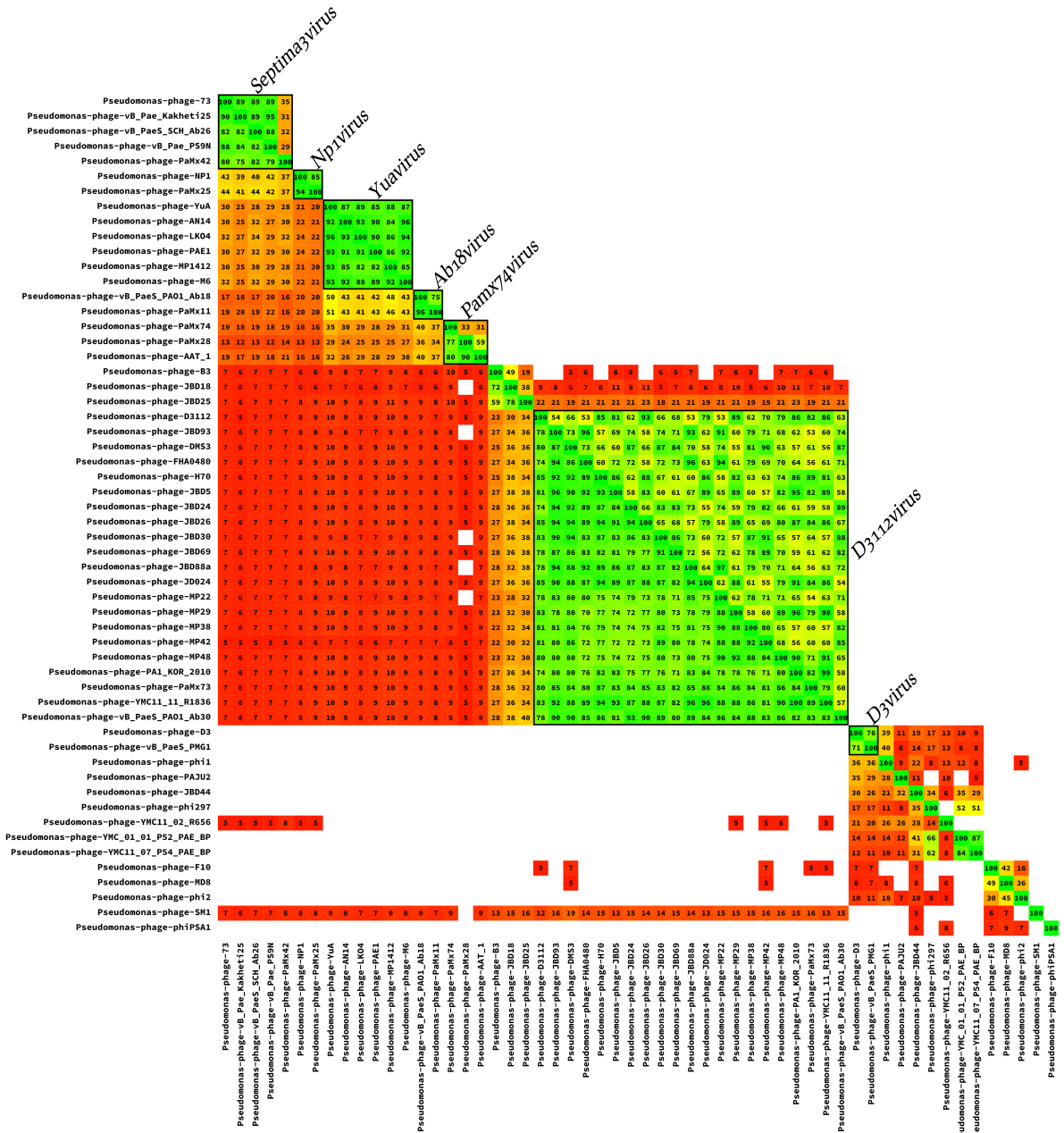
El último paso para considerar un par de ORFs como homólogos, fue la optimización del umbral del cociente de la longitud de la región alineada para dichos ORFs entre la longitud del ORF más largo. En el eje de las abscisas se muestra el valor de corte para el cociente mencionado, los pares de ORFs que mostraran un cociente de la longitud de la región alineada, entre la longitud del ORF más largo por debajo del valor de corte, eran descartados y no se consideraban homólogos. Las ordenadas representan el porcentaje de ORFs homólogos promedio observado entre los integrantes de cada grupo de fagos (anexo 1) comparados con el genoma de referencia de dicho grupo. Se eligió el valor de umbral de 0.5 ya que en este punto se mantenía estable el porcentaje de ORFs homólogos y se eliminaban hits anómalos correspondientes a ORFs muy pequeños comparados con ORFs muy largos que mostraran regiones de alta homología sin que esto significara que fueran homólogos.

Anexo 3: Familia *Podoviridae*



Los fagos empleados se muestran en el anexo 1. A partir de las comparaciones reciprocas se construyó una diagonal separando la matriz en dos secciones. Por encima de la diagonal se muestran los porcentajes de identidad a nivel de secuencia nucleotídica entre genomas. Por debajo de la diagonal se muestran los porcentajes de ORFs homólogos. Los grupos previamente reportados se encuentran encerrados en recuadros negros. Como se puede apreciar, estos grupos muestran porcentajes detectables de identidad nucleotídica y altos porcentajes de ORFs homólogos. Los agrupamientos obtenidos corresponden con los reportados por el ICTV

Anexo 3: Familia *Siphoviridae*



Los fagos empleados se muestran en el anexo 1. A partir de las comparaciones reciprocas se construyó una diagonal separando la matriz en dos secciones. Por encima de la diagonal se muestran los porcentajes de identidad a nivel de secuencia nucleotídica entre genomas. Por debajo de la diagonal se muestran los porcentajes de ORFs homólogos. Los grupos previamente reportados se encuentran encerrados en recuadros negros. Como se puede apreciar, estos grupos muestran porcentajes detectables de identidad nucleotídica y altos porcentajes de ORFs homólogos. Los agrupamientos obtenidos corresponden con los reportados por el ICTV

Anexo 4. Funciones codificadas en los genomas estudiados. PaMx11

ORF	Función asignada	Similitud de secuencia ¹	Dominios conservados ²	Homología estructural ³
01	Subunidad menor de la terminasa	YP_009125148 (2 e-113)	-	3HEFB (2.5)
02	Subunidad mayor de la terminasa	YP_001595874 (0.0)	PF03237 (2.2 e-20)	4IDHa (1.28)
03	Proteína portal*	YP_001595875 (5 e-161)	PF13264 (1.5 e-38)	2JESa (2.62)
04	Proteína de morfogénesis*	YP_001595877 (3 e-133)	PF04233 (6.3 e-10)	2QKGa (2.69)
05	PhoH-like	YP_009125152 (2 e-144)	PF02562 (1.2 e-70)	3B85b (1.66)
06	-	YP_009125153 (2 e-34)	-	3DWKc (2.78)
07	-	YP_001039717 (6 e-08)	-	4B63a (3.56)
08	-	WP_044864770 (9 e-14)	-	2NXFa (3.46)
09	-	-	-	4FWEa (2.63)
10	-	-	-	1CX8b (3.55)
11	-	YP_009125154 (8 e-49)	-	2L82a (3.24)
12	-	YP_001595870 (4 e-16)	-	2YBFa (1.92)
13	-	YP_009125155 (4 e-28)	-	2CYDa (2.27)
14	Proteína estructural de virión	YP_001595878 (2 e-105)	-	3PDYa (2.76)
15	Proteína mayor de cápside*	YP_001595879 (4 e-156)	SSF56563 (5.5 e-22)	3E8Ka (3.13)
16	-	YP_001595881 (4 e-14)	-	1E29a (2.03)
17	Proteína estructural de virión*	YP_009125159 (7 e-34)	-	3RKOn (3.44)
18	Proteína estructural de virión	YP_001595886 (3 e-29)	-	1XN8a (2.93)
19	Proteína estructural de virión*	YP_009125161 (3 e-82)	-	3OSQa (4.28)
20	Proteína estructural de virión	YP_009125162 (7 e-79)	-	2COKa (3.06)
21	Proteína terminadora de cauda	YP_009125163 (6 e-84)	PF13554 (5.1 e-15)	4HUDa (3.69)
22	Proteína mayor de cauda*	YP_009125164 (0.0)	-	1W8Na (2.18)
23	-	YP_009125165 (5 e-75)	PF08748 (4.9 e-5)	2K30a (3.35)
24	-	YP_009125166 (3 e-62)	-	2GSCb (2.87)
25	Proteína de medida de la cauda*	YP_009125167 (0.0)	-	2BKUb (0.89)
26	Proteína estructural de virión*	YP_001595893 (0.0)	G3DSA:2.60.120.260 (2.5 e-4)	2CN3b (2.45)
27	Proteína estructural de virión	YP_001595894 (6 e-178)	-	1MAFh (1.70)
28	Proteína estructural de virión*	YP_001595895 (1 e-77)	-	4ACQa (2.00)
29	Proteína estructural de virión	YP_001595896 (6 e-53)	PF09356 (8.9 e-19) PF09931 (2.2 e-11)	3IAXa (3.33)
30	Proteína estructural de virión*	YP_001595897 (3 e-11)	-	2ZZGb (2.92)
31	-	YP_009125173 (7 e-31)	-	2Y9Ua (1.84)
32	Proteína estructural de virión*	YP_001595899 (2 e-149)	PF13550 (1.2 e-21)	1BY5a (1.03)
33	-	YP_009125175 (1 e-71)	PF10983 (1.1 e-26)	1QWZa (3.74)
34	Proteína de membrana	YP_009125176 (7 e-92)	-	1ILS4a (2.21)

¹Números de acceso de las proteínas similares en bases de datos, los números entre parentesis indican el e-value obtenido con BLASTp

²Números de acceso de los dominios conservados y su e-value (números entre parentesis) obtenido con InterProScan

³Números de acceso de las estructuras similares en bases de datos. Los números entre parentesis indican el RMSD obtenido con TM-Align

*Proteínas identificadas por espectrometría de masas

Anexo 4: PaMx11

ORF	Función asignada	Similitud de secuencia ¹	Dominios conservados ²	Homología estructural ³
35	Endolisina	YP_006299867 (2 e-54)	SSF54001 (1.96 e-06) TIGR02594 (1.5 e-43)	3I86b (1.67)
36	Proteína de membrana	YP_009125178 (2 e-28)	-	3PFDa (2.16)
37	-	YP_009125179 (1 e-65)	-	2Y44a (2.66)
38	Helicasa	YP_001595825 (0.0)	PF00176 (2.5 e-18) PF00271 (4.5 e-7)	3MWYw (1.69)
39	-	YP_001595826 (2 e-39)	-	2GRUa (2.56)
40	-	YP_001595827 (1 e-62)	-	3COLa (2.11)
41	-	YP_001595828 (5 e-103)	-	2QH8a (3.20)
42	Helicasa UvrD	YP_001595829 (1 e-148)	PF13538 (1.2 e-24) PF13604 (3.2 e-21)	3UPUa (2.13)
43	Subunidad menor de la ribonucleotido reductasa	YP_009125109 (0.0)	PF00268 (3 e-11) G3DSA:1.10.620.20 (8.4 e-35)	4D8Fa (1.71)
44	Subunidad mayor de la ribonucleotido reductasa	YP_009125110 (0.0)	PF02867 (2 e-25) PF00317 (2.6 e-5)	3HNDa (1.11)
45	-	YP_009125111 (5 e-91)	-	1LS4a (2.96)
46	-	YP_001595833 (7 e-97)	-	4ACQa (1.65)
47	Oxidoreductasa dependiente de FAD	YP_009125113 (0.0)	PF01266 (2.5 e-11) SSF51905 (8.63 e-17)	1NG4b (1.78)
48	Acetiltransferasa	YP_009125114 (5 e-80)	PF00583 (7.4 e-14) SSF55729 (1.05 e-21)	2CNTD (1.21)
49	-	YP_009125115 (4 e-121)	PF13671 (1.8 e-6) SSF52540 (2.35 e-11)	1EX7a (2.87)
50	-	YP_009125116 (0.0)	-	4B21a (4.06)
51	Timidilato sintasa	YP_001595840 (2 e-140)	PF00303 (3.1 e-45) SSF55831 (3.14 e-50)	3KJSa (1.05)
52	Proteína de dominio HD	YP_001595841 (2 e-42)	SSF109604 (7.22 e-19) G3DSA:1.10.3210.10 (1.4 e-23)	2PAUb (1.86)
53	-	YP_001595843 (7 e-26)	PF11753 (2.1 e-5)	3UQ3a (2.04)
54	-	YP_009125120 (6 e-54)	-	1XWYa (3.93)
55	DNA polimerasa	YP_001595844 (0.0)	PF00476 (7.4 e-43) SSF56672 (2.89 e-68)	1KFSa (0.79)
56	dCMP deaminasa	YP_009125122 (0.0)	PF00383 (1.6 e-21) SSF53927 (5.27 e-28)	2W4La (1.45)
57	-	YP_001595847 (3 e-35)	-	1ZP7b (3.17)
58	-	YP_001595848 (1 e-13)	PF13560 (2.3 e-7) SSF47413 (7.19 e-5)	3F51a (2.08)
59	Primasa	YP_001595849 (0.0)	TIGR01613 (3.5 e-5)	1QBKb (2.14)
60	-	YP_009125126 (2 e-33)	-	3LESa (2.73)
61	-	YP_009125127 (9 e-68)	-	1AEPa (2.94)

¹Números de acceso de las proteínas similares en bases de datos, los números entre parentesis indican el e-value obtenido con BLASTp

²Números de acceso de los dominios conservados y su e-value (números entre parentesis) obtenido con InterProScan

³Números de acceso de las estructuras similares en bases de datos. Los números entre parentesis indican el RMSD obtenido con TM-Align

Anexo 4: PaMx11

ORF	Función asignada	Similitud de secuencia ¹	Dominios conservados ²	Homología estructural ³
62	Pirofosfohidrolasa	YP_009125128 (2 e-134)	PF01503 (3.5 e-12) G3DSA:1.10.3420.10 (7.5 e-26)	3MQUa (3.45)
63	-	YP_009125129 (3 e-79)	-	2GFNb (3.52)
64	-	YP_009125130 (2 e-69)	-	1F7Ba (2.44)
65	-	YP_009125131 (1 e-91)	-	4G0Ba (3.37)
66	-	YP_009125132 (7 e-40)	-	1ITZa (3.40)
67	-	YP_009125133 (2 e-47)	-	3KKDc (3.24)
68	-	YP_009125134 (4 e-110)	-	1WZGb (4.72)
69	DNA ligasa	YP_009125135 (0.0)	PF14743 (2.1 e-15) PF01068 (4.8 e-12)	2Q2Ta (2.07)
70	-	YP_009125136 (1 e-11)	-	2FY0a (2.97)
71	-	YP_009125137 (0.0)	SSF55874 (1.32 e-5)	1X8Kb (1.51)
72	-	YP_009125138 (1 e-58)	-	3IYGh (3.74)
73	-	WP_044027045 (3 e-26)	-	1B6Ch (2.51)
74	Trifosfato hidrolasa	YP_009125139 (6 e-105)	SSF52540 (3.76 e-5)	4GBMa (3.08)
75	-	YP_009125140 (2 e-37)	-	3CHQa (3.63)
76	-	YP_009125142 (1 e-30)	-	3F22c (1.29)
77	Exonucleasa	YP_009125143 (2 e-149)	PF12843 (4.5 e-6) SSF53098 (7.59 e-18)	1ZBHc (2.35)
78	-	-	-	3OV0a (1.42)
79	-	-	-	3KO1a (2.10)
80	-	YP_009125146 (9 e-41)	-	3PESa (1.00)
81	-	YP_009125147 (2 e-48)	PF14216 (3 e-26)	2ILYb (3.56)

¹Números de acceso de las proteínas similares en bases de datos, los números entre parentesis indican el e-value obtenido con BLASTp

²Números de acceso de los dominios conservados y su e-value (números entre parentesis) obtenido con InterProScan

³Números de acceso de las estructuras similares en bases de datos. Los números entre parentesis indican el RMSD obtenido con TM-Align

Anexo 4: PaMx25

ORF	Función asignada	Similitud de secuencia ¹	Dominios conservados ²	Homología estructural ³
01	Proteína de membrana	-	-	3G9Ra (0.51)
02	Transglicosilasa lítica	WP_015830271 (7 e-77)	SSF53955 (1.15 e-6) G3DSA:1.10.530.10 (3.2 e-6)	2Y8Pa (1.18)
03	Proteína de membrana	EA013176 (7 e-22)	-	1GAKa (2.36)
04	Proteína de membrana	WP_015830269 (5 e-18)	-	3L75c (3.9)
05	Subunidad menor de la terminasa	AKA60867 (1 e-14)	PF03592 (1.4 e-5)	2Q01b (3.93)
06	Subunidad mayor de la terminasa	CDM51558 (0.0)	PF03237 (1.8 e-12)	3CPEa (1.71)
07	Proteína portal*	YP_223984 (4 e-166)	PF13264 (1 e-27)	2JESa (1.38)
08	Proteína de morfogénesis	YP_223986 (1 e-128)	TIGR01641 (1.9 e-11)	1SLYa (2.45)
09	-	YP_223990 (2 e-70)	-	1QUUa (2.08)
10	Proteína mayor de cápside*	YP_223991 (2 e-139)	-	2FS3a (2.83)
11	Proteína estructural de virión	-	-	2A9Sa (1.99)
12	-	AJT60492 (2 e-11)	-	1H9Ta (2.46)
13	Proteína estructural de virión*	-	-	-
14	Proteína estructural de virión*	YP_001595886 (4 e-26)	-	1XN8a (2.1)
15	Proteína estructural de virión*	YP_224000 (8 e-25)	-	1UTEa (4.03)
16	Proteína terminadora de cauda*	YP_001595888 (4 e-20)	PF13554 (1.3 e-7)	4HUDa (3.40)
17	Proteína mayor de cauda*	YP_001595889 (0.0)	-	1DABa (1.38)
18	Proteína estructural de virión	YP_001595890 (5 e-47)	-	3PF0a (2.59)
19	-	YP_006299882 (7 e-22)	-	1QJ2b (3.53)
20	Proteína estructural de virión*	YP_224005 (2 e-36)	PF04883 (2.5 e-8)	1EBTa (1.7)
21	Proteína de medida de la cauda*	YP_001595892 (5 e-154)	TIGR02675 (2.2 e-15)	2XWUb (2.27)
22	Proteína estructural de virión*	YP_001595893 (0.0)	SSF49785 (2.6 e-5)	3A0Fa (1.34)
23	Proteína estructural de virión*	YP_001595894 (1 e-143)	-	2ZJ6a (2.45)
24	Proteína estructural de virión*	YP_001595895 (9 e-74)	-	-
25	Proteína estructural de virión*	YP_001595896 (7 e-53)	PF09356 (1.8 e-18)	2OJHa (3.09)
26	Proteína estructural de virión*	YP_009125172 (8 e-37)	-	4ARTa (3.29)
27	-	YP_009125173 (1 e-10)	-	3K1Rb (1.62)
28	Proteína estructural de virión*	YP_001595899 (3 e-148)	PF13550 (4.6 e-24)	3A0Fa (1.54)
29	-	YP_009125175 (5 e-23)	-	2HKEb (3.45)
30	-	-	-	1ZXXa (3.64)
31	-	YP_006299921 (3 e-8)	-	3VA9a (1.69)
32	-	WP_018211163 (4 e-4)	-	3TXYa (2.58)
33	DNA polimerasa	YP_006299891 (2 e-153)	PF00136 (4.9 e-5) SSF56672 (2.78 e-29)	1TGOa (0.99)
34	Clamp de replicación	YP_006299892 (4 e-25)	-	1UNNa (1.62)
35	tRNA ribosiltransferasa	YP_009032326 (1 e-159)	-	1IT7b (1.82)

¹Números de acceso de las proteínas similares en bases de datos, los números entre parentesis indican el e-value obtenido con BLASTp

²Números de acceso de los dominios conservados y su e-value (números entre parentesis) obtenido con InterProScan

³Números de acceso de las estructuras similares en bases de datos. Los números entre parentesis indican el RMSD obtenido con TM-Align

*Proteínas identificadas por espectrometría de masas

Anexo 4: PaMx25

ORF	Función asignada	Similitud de secuencia ¹	Dominios conservados ²	Homología estructural ³
36	-	AKA60899 (5 e-38)	-	1OT5b (4.63)
37	GTP ciclohidrolasa	EPF71056 (2 e-87)	PF01227 (2.8 e-72) SSF55620 (1.11 e-72)	1IS7a (0.79)
38	-	-	-	1K66b (3.8)
39	QueD-like	WP_046118971 (2 e-74)	PF01242 (4.5 e-28) SSF55620 (1.5 e-31)	3QNAa (1.58)
40	QueC-like	YP_009032329 (1 e-153)	PF06508 (4.9 e-50)	1ECFa (2.32)
41	QueE-like	AIW03906 (5 e-79)	PF04055 (2.6 e-11) PF13394 (1 e-16)	3C8Fa (2.08)
42	-	YP_009032332 (8 e-27)	-	2OCYa (2.43)
43	-	YP_009032333 (6 e-17)	-	-
44	-	-	-	3V39a (2.33)
45	-	YP_009032360 (2 e-14)	-	3EZUa (1.79)
46	Helicasa	YP_006299894 (0.0)	PF00271 (3.5 e-12) SSF52540 (1 e-37)	2W00b (1.84)
47	Exonucleasa RecB	YP_009032335 (5 e-113)	-	4IC1a (2.06)
48	-	WP_017980597 (4 e-6)	-	1URUa (1.9)
49	-	YP_009126577 (2 e-42)	PF05037 (1.4 e-24)	4LESa (3.25)
50	ATPasa RecA	YP_009032337 (6 e-88)	PF13479 (5.5 e-30)	2DR3d (2.84)
51	DNA topoisomerasa	WP_012019844 (7 e-104)	SSF56726 (5.4 e-11) G3DSA:3.40.1360.10 (4.3 e-7)	1D3Yb (2.03)
52	-	-	-	3RGFb (3.18)
53	DNA ligasa	WP_042589654 (4 e-87)	PF01068 (2.1 e-18) PF14743 (1.3 e-15)	2Q2Ud (0.89)
54	Timidilato sintasa	WP_017532241 (4 e-49)	PF02511 (6.6 e-45) SSF69796 (8.5 e-47)	1O26b (1.19)
55	Ribonucleasa H-like	YP_009032339 (3 e-26)	PF00075 (1.7 e-10) SSF53098 (8.73 e-17)	3U3Gd (1.71)
56	-	YP_009032340 (3 e-15)	-	3ZCOa (1.84)
57	Pirofosfohidrolasa	EGT92009 (2 e-46)	PF03819 (9.5 e-9) SSF101386 (2.1 e-12)	3CRAb (3.3)
58	Primasa helicasa	AJT60548 (0.0)	PF13148 (3.8 e-11)	3EA5b (1.88)
59	-	AIW03924 (2 e-04)	-	3CS1a (2.77)
60	-	-	-	3RAUa (2.26)
61	-	-	-	3LD2a (2.4)
62	-	YP_006560789 (4 e-42)	-	1GAXa (2.74)
63	-	WP_017979863 (2 e-17)	-	4A04a (1.96)
64	-	YP_006299893 (2 e-18)	-	4FZ4a (1.31)

¹Números de acceso de las proteínas similares en bases de datos, los números entre parentesis indican el e-value obtenido con BLASTp

²Números de acceso de los dominios conservados y su e-value (números entre parentesis) obtenido con InterProScan

³Números de acceso de las estructuras similares en bases de datos. Los números entre parentesis indican el RMSD obtenido con TM-Align

*Proteínas identificadas por espectrometría de masas

Anexo 4: PaMx25

ORF	Función asignada	Similitud de secuencia ¹	Dominios conservados ²	Homología estructural ³
65	-	WP_044027045 (5 e-28)	-	1VJYa (3.45)
66	Glicosilasa DenV de reparación de DNA	AGE55633 (3 e-46)	PF03013 (3.7 e-34) SSF47077 (1.96 e-38)	1ENK (0.93)
67	-	WP_020327578 (6 e-11)	SSF57756 (3.35 e-5) G3DSA:4.10.60.10 (2.7 e-4)	2JPSa (2.68)
68	-	YP_009125139 (5 e-45)	SSF52540 (9.4 e-8)	2ZQ5a (2.51)
69	-	-	-	3GE8d (4.19)
70	Ribonucleotido reductasa	AIW56646 (0.0)	SSF51998 (1.62 e-103) G3DSA:3.90.1390.10 (1.9 e-16)	1XJEa (1.22)
71	-	YP_009032343 (5 e-36)	-	1GEFa (2.93)
72	-	-	-	3I9Ya (2.4)
73	Helicasa UvrD	BAR61886 (1 e-98)	PF00580 (6.2 e-18) PF13361 (5.1 e-13)	2IS6b (1.4)
74	-	-	-	2WWAa (3.18)

¹Números de acceso de las proteínas similares en bases de datos, los números entre parentesis indican el e-value obtenido con BLASTp

²Números de acceso de los dominios conservados y su e-value (números entre parentesis) obtenido con InterProScan

³Números de acceso de las estructuras similares en bases de datos. Los números entre parentesis indican el RMSD obtenido con TM-Align

Anexo 4: PaMx28

ORF	Función asignada	Similitud de secuencia ¹	Dominios conservados ²	Homología estructural ³
01	-	CCA79406 (3 e-25)	-	1UUQa (3.77)
02	-	-	-	1T90b (3.44)
03	-	-	SSF57756 (7.9 e-5)	3CRLb (3.08)
04	Subunidad menor de la terminasa	-	G3DSA:2.60.40.10 (4.1 e-04)	2IG2h (2.54)
05	Subunidad mayor de la terminasa	YP_001595874 (0.0)	PF03237 (6.8 e-19)	4IDHa (1.39)
06	Ribonucleotido reductasa	YP_001595830 (0.0)	PF02867 (2.2 e-111) SSF51998 (7.32 e-129)	1ZYZa (0.97)
07	DNA ligasa	YP_009125135 (2 e-79)	PF01068 (4.7 e-17) SSF56091 (3.4 e-26)	-
08	Proteína portal*	YP_001595875 (1 e-158)	PF13264 (5.4 e-37)	2JESa (1.95)
09	-	ADJ21941 (8 e-16)	-	3LDYa (1.34)
10	Proteína de morfogénesis*	YP_001595877 (2 e-84)	PF04233 (1.7 e-13) SSF56399 (4.52 e-25)	1U6Gc (2.03)
11	-	-	-	3UBPc (3.76)
12	Proteína estructural de virión*	YP_001595878 (5 e-44)	-	4HPQc (2.99)
13	YbiA-like	WP_010348170 (3 e-57)	PF08719 (4.9 e-48) SSF143990 (1.44 e-43)	2B3Wa (1.19)
14	-	-	-	1R88b (2.52)
15	-	YP_009098835 (7 e-39)	-	1ENIa (2.14)
16	-	WP_045111943 (4 e-07)	-	3HF3b (3.19)
17	-	WP_020470063 (9 e-11)	-	3IJ6d (3.13)
18	-	-	-	2YN6a (2.38)
19	Proteína mayor de cápside*	YP_001595879 (2 e-145)	SSF56563 (3.4 e-24)	2FS3a (2.30)
20	-	YP_001595881 (6 e-09)	-	2OUTa (3.14)
21	Proteína estructural de virión*	YP_001595886 (1 e-27)	-	1ZTSa (3.04)
22	Proteína estructural de virión*	WP_025423048 (4 e-12)	-	-
23	Proteína estructural de virión*	YP_009125162 (2 e-40)	PF04883 (1.8 e-04)	1AEPa (1.64)
24	Proteína terminadora de cauda*	YP_009125163 (2 e-21)	PF13554 (8.3 e-19)	4HUDa (3.75)
25	Proteína mayor de cauda*	YP_009125164 (5 e-166)	-	3TKVa (2.16)
26	-	YP_009125165 (6 e-42)	-	1AEPa (1.93)
27	-	YP_009125166 (9 e-19)	-	1W37a (2.86)
28	Proteína de medida de la cauda*	YP_009125167 (0.0)	-	3EA5b (2.23)
29	Proteína estructural de virión*	YP_001595895 (9 e-07)	-	3SCYa (2.37)
30	Proteína estructural de virión*	YP_001595896 (3 e-32)	PF09931 (2 e-10) PF09356 (1.5 e-10)	-
31	Proteína estructural de virión*	YP_001595897 (2 e-14)	-	4ARTa (3.4)
32	-	WP_042089778 (2 e-16)	-	2KJ8a (2.33)

¹Números de acceso de las proteínas similares en bases de datos, los números entre parentesis indican el e-value obtenido con BLASTp

²Números de acceso de los dominios conservados y su e-value (números entre parentesis) obtenido con InterProScan

³Números de acceso de las estructuras similares en bases de datos. Los números entre parentesis indican el RMSD obtenido con TM-Align

*Proteínas identificadas por espectrometría de masas

Anexo 4: PaMx28

ORF	Función asignada	Similitud de secuencia ¹	Dominios conservados ²	Homología estructural ³
33	Proteína estructural de virión*	WP_011634507 (0.0)	PF13550 (2 e-20) SSF53300 (9.2 e-13)	2ZXQa (1.73)
34	Proteína estructural de virión*	WP_004817414 (5 e-11)	-	3C3Xb (2.11)
35	Proteína estructural de virión*	YP_009125175 (8 e-46)	PF10983 (2 e-25)	2YMUa (4.02)
36	Proteína de membrana	BAO53034 (2 e-15)	-	2CAZe (2.74)
37	Endolisina	CAJ28446 (4 e-37)	PF13539 (4.8 e-07) SSF55166 (4.79 e-13)	2VO9a (1.96)
38	Proteína estructural de virión	YP_004421769 (2 e-18)	-	-
39	-	BAO53036 (2 e-07)	-	1VGMb (2.06)
40	Proteína de membrana	BAO53037 (1 e-17)	-	2KUBa (1.41)
41	-	-	-	1Q09a (2.55)
42	-	-	-	3VA9a (1.97)
43	-	YP_009126412 (5 e-05)	-	3AYFa (3.15)
44	-	-	-	1DGJa (3.53)
45	Proteína de dominio HIRAN	WP_012843048 (1 e-13)	PF08797 (5.4 e-12)	3K2Ya (0.55)
46	-	YP_004421779 (4 e-04)	-	3IUVa (2.28)
47	-	WP_031568659 (3 e-18)	-	3LICa (3.02)
48	-	YP_004421782 (2 e-10)	-	1ZP7b (2.72)
49	Dihidrofolato reductasa	YP_004421783 (2 e-15)	SSF53597 (6.81 e-16) G3DSA:3.40.430.10 (1.9 e-13)	3TQ8a (1.88)
50	dCMP deaminasa	YP_004421784 (4 e-24)	PF00383 (1.6 e-15) SSF53927 (6.67 e-27)	2W4La (1.55)
51	-	ADV26973 (6 e-17)	-	3LDYa (1.15)
52	Timidilato sintasa	YP_004421785 (4 e-72)	PF00303 (4.7 e-44) SSF55831 (1.96 e-48)	1CI7b (1.85)
53	Pirofosfohidrolasa	YP_004421786 (3 e-30)	PF01503 (2.9 e-19) G3DSA:1.10.3420.10 (3.9 e-24)	3AYFa (5.5)
54	-	YP_001595850 (9 e-08)	-	3HL9d (2.96)
55	-	YP_004421787 (4 e-46)	-	3MDUa (6.02)
56	-	YP_004421789 (4 e-22)	-	2OTOd (2.79)
57	Helicasa	YP_004421790 (5 e-123)	PF00176 (9.00 e-08) SSF52540 (3.42 e-31)	1Z3Ix (0.85)
58	-	-	-	2FNCA (3.07)
59	DNA polimerasa	YP_004421796 (3 e-123)	PF00476 (1.5 e-10) SSF56672 (1.89 e-59)	1KFSa (0.87)
60	-	-	-	3P8Cb (3.7)
61	-	-	-	1VDKa (3.46)
62	-	-	-	3Q6Ka (3.72)

¹Números de acceso de las proteínas similares en bases de datos, los números entre parentesis indican el e-value obtenido con BLASTp

²Números de acceso de los dominios conservados y su e-value (números entre parentesis) obtenido con InterProScan

³Números de acceso de las estructuras similares en bases de datos. Los números entre parentesis indican el RMSD obtenido con TM-Align

*Proteínas identificadas por espectrometría de masas

Anexo 4: PaMx28

ORF	Función asignada	Similitud de secuencia ¹	Dominios conservados ²	Homología estructural ³
63	-	-	-	3IZ0d (2.87)
64	-	-	-	2XCLa (3.69)
65	-	WP_009367862 (8 e-06)	-	2QIBa (2.99)
66	Proteína de unión a DNA	WP_039072596 (2 e-41)	PF10991 (8.2 e-44) SSF50249 (9.25 e-19)	1JE5a (3.54)
67	-	WP_003667885 (8 e-21)	PF14216 (9 e-25)	2IXDa (3.1)
68	Trifosfato hidrolasa	YP_009125139 (7 e-49)	SSF52540 (2.5 e-08)	1T8Ua (3.37)
69	-	CDO37627 (8 e-04)	-	2WV1b (4.1)
70	-	AJD82885 (8 e-50)	PF10926 (1.5 e-43, 5.8 e-22)	-
71	-	-	-	3RGAA (4.04)
72	-	-	-	2V71b (2.42)
73	-	-	-	3CRJd (2.48)
74	Primasa polimerasa	YP_004421815 (0.0)	PF09250 (4.00 e-22) PF08707 (7.2 e-10)	4A2La (1.71)

¹Números de acceso de las proteínas similares en bases de datos, los números entre parentesis indican el e-value obtenido con BLASTp

²Números de acceso de los dominios conservados y su e-value (números entre parentesis) obtenido con InterProScan

³Números de acceso de las estructuras similares en bases de datos. Los números entre parentesis indican el RMSD obtenido con TM-Align

Anexo 4: PaMx42

ORF	Función asignada	Similitud de secuencia ¹	Dominios conservados ²	Homología estructural ³
01	-	WP_043345172 (5 e-10)	-	2OF5l (3.49)
02	Holina	YP_006560753 (1 e-19)	-	4OR2a (2.48)
03	Endolisina	WP_017330071 (2 e-45)	SSF53955 (3.22 e-25) G3DSA:1.10.530.40 (1.90 e-28)	3HDEa (1.01)
04	Proteína Rz	YP_006299868 (6 e-26)	-	4K12b (1.69)
05	Proteína Rz1	YP_006299869 (9 e-30)	-	2QPQa (3.23)
06	Subunidad menor de la terminasa	YP_006299870 (2 e-82)	-	1UG3b (4.09)
07	Subunidad mayor de la terminasa	YP_006299871 (0.0)	PF03237 (2 e-08)	4IDHa (1.38)
08	Proteína portal*	YP_006299872 (0.0)	PF13264 (7.8 e-43)	2JESm (1.63)
09	Proteína de morfogénesis*	YP_006299873 (0.0)	-	4L6Ya (1.83)
10	Proteína de andamiaje	YP_006299874 (2 e-132)	-	-
11	Proteína mayor de cápside*	YP_006299875 (0.0)	-	4AN5a (2.10)
12	-	YP_006299876 (8 e-56)	-	2YL8a (4.14)
13	Proteína estructural de virión*	YP_006299877 (7 e-31)	-	3B9Wa (3.92)
14	Proteína estructural de virión*	YP_006299878 (6 e-111)	-	1ZTSa (2.67)
15	Proteína estructural de virión*	YP_006299879 (5 e-72)	-	2PP6a (2.33)
16	Proteína terminadora de cauda	YP_006299880 (8 e-92)	-	4HUDa (3.88)
17	Proteína mayor de cauda*	YP_006299881 (0.0)	-	4KH3a (1.29)
18	Proteína estructural de virión	YP_006299882 (5 e-83)	-	2DVMa (3.96)
19	-	YP_006299882 (3 e-50)	-	3IW0a (3.02)
20	Proteína estructural de virión*	YP_006299884 (4 e-78)	-	4KT3a (3.67)
21	Proteína de medida de la cauda*	YP_006299885 (0.0)	TIGR02675 (1 e-20)	4BUJb (2.44)
22	Proteína estructural de virión	YP_006299886 (1 e-166)	-	4IRTa (4.37)
23	Proteína estructural de virión*	YP_006299887 (6 e-101)	-	2BBKh (1.36)
24	Proteína estructural de virión*	YP_006299888 (0.0)	-	3CM9s (1.66)
25	Proteína estructural de virión*	YP_006299889 (6 e-159)	PF09356 (5.8 e-14) PF09931 (7.5 e-07)	3B7Fa (2.84)
26	Proteína estructural de virión*	YP_006299890 (0.0)	PF13550 (3.4 e-26)	4ACQa (2.32)
27	DNA polimerasa	YP_006299891 (0.0)	SSF56672 (1.17 e-32) PF00136 (4 e-09)	3IAYa (1.86)
28	Clamp de replicación	YP_006299892 (0.0)	-	3D1Fv (1.98)
29	Helicasa	YP_006299894 (0.0)	PF00271 (1.3 e-6) SSF52540 (1.49 e-31)	-
30	Exonucleasa RecB	YP_006299895 (0.0)	-	3T33a (1.11)
31	-	WP_032432683 (3 e-36)	PF11672 (1.1 e-36)	1TICb (3.27)
32	-	YP_006299897 (3 e-86)	PF05037 (2.3 e-12)	4BMLa (2.70)
33	ATPasa RecA	YP_006299898 (4 e-160)	PF13479 (4.9 e-22)	3CMTa (4.16)

¹Números de acceso de las proteínas similares en bases de datos, los números entre parentesis indican el e-value obtenido con BLASTp

²Números de acceso de los dominios conservados y su e-value (números entre parentesis) obtenido con InterProScan

³Números de acceso de las estructuras similares en bases de datos. Los números entre parentesis indican el RMSD obtenido con TM-Align

*Proteínas identificadas por espectrometría de masas

Anexo 4: PaMx42

ORF	Función asignada	Similitud de secuencia ¹	Dominios conservados ²	Homología estructural ³
34	-	YP_006299899 (9 e-30)	-	2HSBa (2.21)
35	Pirofosfohidrolasa	YP_006299900 (4 e-96)	SSF101386 (1.77 e-07)	2Q4Pa (2.51)
36	-	-	G3DSA:1.10.8.320 (5.30E-04)	3KZ1a (3.60)
37	Primasa helicasa	YP_006299902 (0.0)	-	1QGKa (1.48)
38	-	-	-	3KNTc (3.20)
39	-	GAL57139 (2 e-09)	-	3HVAb (2.34)
40	-	-	-	2ZZWb (3.26)
41	-	EXB25985 (1 e-19)	-	3KE9b (2.32)
42	-	YP_006299906 (2 e-27)	-	2GPQa (3.20)
43	-	YP_006299908 (1 e-37)	-	3AQGb (2.69)
44	-	YP_006299910 (3 e-18)	-	1CIYa (3.57)
45	-	YP_006299911 (1 e-29)	-	3HWCb (2.52)
46	-	YP_006299912 (1 e-10)	-	3EXGf (4.70)
47	-	YP_006299917 (2 e-38)	-	2WUSr (2.66)
48	-	YP_006299913 (3 e-22)	-	2VFBa (2.61)
49	-	YP_006299914 (2 e-14)	-	2DKEa (2.47)
50	-	YP_006299915 (1 e-41)	-	2GBBa (3.84)
51	-	-	-	1E7Ua (3.59)
52	-	-	-	3P0Ka (2.58)
53	dCMP deaminasa	YP_006299919 (7 e-81)	PF00383 (5 e-21) SSF53927 (3.35 e-25)	2W4Le (1.42)
54	-	YP_006299920 (5 e-46)	-	2BNXb (3.39)
55	-	YP_006299921 (5 e-19)	-	1E9Rb (2.39)
56	-	YP_006299922 (2 e-87)	-	1EK5a (4.73)

¹Números de acceso de las proteínas similares en bases de datos, los números entre parentesis indican el e-value obtenido con BLASTp

²Números de acceso de los dominios conservados y su e-value (números entre parentesis) obtenido con InterProScan

³Números de acceso de las estructuras similares en bases de datos. Los números entre parentesis indican el RMSD obtenido con TM-Align

Anexo 4: PaMx74

ORF	Función asignada	Similitud de secuencia ¹	Dominios conservados ²	Homología estructural ³
01	Subunidad menor de la terminasa	CCA79406 (2 e-26)	-	3HEFb (1.69)
02	-	-	-	4IJJa (1.96)
03	-	-	-	3S9Xa (4.03)
04	Subunidad mayor de la terminasa	YP_001595874 (0.0)	PF03237 (2 e-19)	4IDHa (1.23)
05	-	-	-	2FQ3a (2.01)
06	-	-	-	3PMCa (1.65)
07	Ribonucleotido reductasa	YP_001595830 (0.0)	PF02867 (5.4 e-101) SSF51998 (5.49 e-131)	1XJEa (0.79)
08	-	WP_043764852 (2 e-108)	-	4NEIa (1.92)
09	-	WP_046981916 (6 e-51)	-	3DTDc (2.53)
10	-	-	-	2RAEa (3.59)
11	DNA ligasa	YP_009125135 (3 e-69)	PF14743 (8.5 e-14) SSF56091 (1.21 e-27)	2Q2Ta (0.64)
12	Proteína portal	YP_001595875 (2 e-154)	PF13264 (2.7 e-37)	2JESa (1.69)
13	Proteína de morfogénesis	YP_001595877 (4 e-102))	PF04233 (1.1 e-12) TIGR01641 (6.7 e-16)	3J7Ja (1.34)
14	Proteína estructural de virión	YP_001595878 (8 e-40)	-	-
15	-	-	-	-
16	-	YP_009098835 (5 e-39)	-	-
17	-	WP_021158921 (8 e-06)	PF05259 (1.4 e-04)	-
18	-	-	-	-
19	Major head protein	YP_001595879 (5 e-147)	PF05065 (1.7 e-06) SSF56563 (1.83 e-25)	-
20	-	YP_001595881 (9 e-07)	-	-
21	-	YP_009125159 (2 e-10)	-	-
22	Proteína estructural de virión	YP_001595886 (3 e-25)	-	-
23	Proteína estructural de virión	WP_025423048 (2 e-13)	-	-
24	Proteína estructural de virión	YP_009125162 (5 e-41)	PF04883 (2.2 e-04)	-
25	Tail terminator protein	YP_009125163 (9 e-22)	PF13554 (1.4 e-18)	-
26	Major tail protein	YP_009125164 (2 e-104)	-	-
27	-	YP_009125165 (6 e-45)	-	3M1Lc (3.58)
28	-	YP_009125166 (9 e-38)	-	3PMDa (3.28)
29	Tail tape measure protein	YP_009125167 (0.0)	-	4HNWa (0.97)
30	Proteína estructural de virión	YP_001595893 (0.0)	PF00754 (6.5 e-09) SSF49785 (4.47 e-09)	4LGNa (1.19)
31	Proteína estructural de virión	YP_001595894 (2 e-164)	-	1MAF (1.10)
32	Proteína estructural de virión	YP_001595895 (3 e-76)	-	3S8Va (2.20)
33	Proteína estructural de virión	YP_001595896 (1 e-58)	PF09356 (8.7 e-20) PF09931 (1.5 e-13)	3B7Fa (2.68)

¹Números de acceso de las proteínas similares en bases de datos, los números entre parentesis indican el e-value obtenido con BLASTp

²Números de acceso de los dominios conservados y su e-value (números entre parentesis) obtenido con InterProScan

³Números de acceso de las estructuras similares en bases de datos. Los números entre parentesis indican el RMSD obtenido con TM-Align

Anexo 4: PaMx74

ORF	Función asignada	Similitud de secuencia ¹	Dominios conservados ²	Homología estructural ³
34	Proteína estructural de virión	YP_001595897 (7 e-13)	-	4BHRa (2.11)
35	-	YP_001595898 (2 e-14)	-	3IDWa (1.85)
36	Proteína estructural de virión	YP_009125174 (0.0)	PF13550 (1.7 e-26) PF00754 (1.3 e-70)	4BXSv (2.28)
37	Proteína de membrana	BAO53034 (2 e-17)	-	3L1La (3.41)
38	Endolisina	CAJ28446 (3 e-40)	SSF55166 (5.57 e-14)	2MXZa (2.21)
39	Proteína estructural de virión	YP_004421769 (4 e-19)	-	3EUHb (3.78)
40	-	BAO53036 (9 e-11)	-	1YAKd (2.77)
41	Proteína de membrana	BAO53037 (7 e-18)	-	2R17c (2.47)
42	-	-	-	4WB1a (3.54)
43	-	YP_001595851 (2 e-04)	-	4IGGa (3.13)
44	-	-	-	3KYQa (3.36)
45	Proteína de dominio HIRAN	WP_012843048 (9 e-14)	PF08797 (1.7 e-12)	3K2Ya (0.5)
46	-	WP_032964266 (5 e-10)	-	2YFAa (3.41)
47	-	WP_029672564 (3 e-19)	-	1ZU4a (2.41)
48	-	-	-	3MJ5a (3.72)
49	-	YP_004421782 (3 e-18)	-	1ZP7a (1.55)
50	Dihidrofolato reductasa	YP_004421783 (2 e-15)	SSF53597 (4.45 e-13)	3TQAa (1.32)
51	dCMP deaminasa	YP_004421784 (8 e-29)	PF00383 (1.4 e-21) SSF53927 (1.08 e-31)	2W4LF (0.99)
52	Timidilate sintasa	YP_004421785 (7 e-70)	PF00303 (4.4 e-44) SSF55831 (6.02 e-38)	4DQ1a (1.27)
53	Pirofosfohidrolasa	YP_004421786 (1 e-16)	PF01503 (3 e-20)	3CRAa (4.26)
54	-	YP_001595850 (4 e-07)	-	4P78c (1.90)
55	-	YP_004421787 (4 e-46)	-	4X96d (1.53)
56	-	YP_004421789 (7 e-23)	-	4CKGa (3.32)
57	Helicasa	YP_004421790 (9 e-126)	PF00176 (3.9 e-11) SSF52540 (3.55 e-31)	1Z63b (1.18)
58	-	-	-	3ERRa (3.11)
59	DNA polimerasa	YP_004421796 (6 e-126)	PF00476 (2.2 e-10) SSF56672 (3.22 e-61)	2KFNa (0.93)
60	-	WP_047028429 (3 e-23)	PF11672 (6.7 e-28)	3APGa (3.95)
61	-	-	-	4BUJb (3.59)
62	-	-	-	2A7Va (3.26)
63	-	-	-	3FMEa (4.02)
64	-	-	-	2WPNa (3.17)
65	-	WP_039567398 (3 e-22)	PF10979 (6.4 e-13)	2RAEa (1.95)
66	-	-	-	1DKNa (2.59)
67	Proteína de unión a DNA	WP_039072596 (6 e-41)	PF10991 (1.4 e-41) SSF50249 (5.82 e-17)	1JE5a (3.27)

¹Números de acceso de las proteínas similares en bases de datos, los números entre parentesis indican el e-value obtenido con BLASTp

²Números de acceso de los dominios conservados y su e-value (números entre parentesis) obtenido con InterProScan

³Números de acceso de las estructuras similares en bases de datos. Los números entre parentesis indican el RMSD obtenido con TM-Align

Anexo 4: PaMx74

ORF	Función asignada	Similitud de secuencia ¹	Dominios conservados ²	Homología estructural ³
68	-	WP_037625068 (5 e-22)	PF14216 (3.2 e-25)	3HZXb (3.54)
69	Pirofosfohidrolasa	YP_001210245 (2 e-13)	-	4A01 (5.14)
70	-	CDO37627 (5 e-06)	-	2NX4a (3.14)
71	-	AJD82885 (4 e-45)	PF10926 (3.6 e-42)	3J4Aa (0.95)
72	-	-	-	1TE1a (3.36)
73	-	YP_004421814 (2 e-04)	-	4UOSa (1.99)
74	-	-	-	4P1Ta (2.62)
75	Primasa polimerasa	YP_004421815 (0.0)	PF09250 (1.5 e-23) SSF56747 (5.17 e-15)	1VW1a (0.75)

¹Números de acceso de las proteínas similares en bases de datos, los números entre parentesis indican el e-value obtenido con BLASTp

²Números de acceso de los dominios conservados y su e-value (números entre parentesis) obtenido con InterProScan

³Números de acceso de las estructuras similares en bases de datos. Los números entre parentesis indican el RMSD obtenido con TM-Align

Anexo 5: Identificación de las proteínas estructurales del fago PaMx11

Main Entry Accession	NTT	Probability	# Tot PSMs	% Coverage / Weight	% Spectrum ids / NSP
ALH23677.1 portal protein [Pseudomonas phage PaMx11]	1.0000		530	42.2%	63.30%
VLRPGSVELYR	0.9995		69	1.00	19.93
LTITTDPOGK	0.9995		66	1.00	19.93
GFVHVLIDFFR	0.9995		58	1.00	19.93
FFLLAGSGLSPEDADLVVGNR	0.9995		49	1.00	19.93
AGPGAQTATAR	0.9995		37	1.00	19.93
EFDDDYKER	0.9840		36	1.00	19.93
DAGIQLPK	0.9995		29	1.00	19.93
ILEEETEVTENEAAVRR	0.9995		24	1.00	19.93
DLEALEDM ₁₄ SYGATHLK	0.9995		24	1.00	19.93
AGPDTESASFK	0.9995		20	1.00	19.93
KAWLEELK	0.9993		20	1.00	19.93
AGPDTESASFVK	0.9995		19	1.00	19.93
REFDDDYKER	0.9459		19	1.00	19.93
FYYVEHTGAANAGQK	0.9995		16	1.00	19.93
KAWLEELIKR	0.9965		14	1.00	19.93
ILEEETEVTENEAAVRR	0.9995		14	1.00	19.93
KDPDDKAPENLPPDOPPASAGGSD	0.9995		6	1.00	19.93
IRVLRPGSVELYR	0.3383		6	1.00	19.93
VARFFLLAGSGLSPEDADLVVGNR	0.9204		3	1.00	19.93
LEDEWATGLDAZPFVYYAGTR	0.9709		3	1.00	19.93
AWLEELIKR	0.9969		11	1.00	19.93
AWLEELIKR	0.9868		11	1.00	19.93
Main Entry Accession	NTT	Probability	# Tot PSMs	% Coverage / Weight	% Spectrum ids / NSP
ALH23678.1 morphogenesis protein [Pseudomonas phage PaMx11]	1.0000		730	52.2%	25.11%
ERFPGAFCAGL	0.9997		19	1.00	25.89
STH ₁₄₇ PVLGGEEELVDKAFVTDTR	0.9998		17	1.00	25.89
FFGAFCAGL	0.9998		10	1.00	25.89
WTSTLDGR	0.9998		9	1.00	25.89
LGLTQGENLDTIVRR	0.9998		8	1.00	25.89
VEDEKLEQSR	0.9998		8	1.00	25.89
GGALDDIYDR	0.9998		7	1.00	25.89
I _{LH} ₁₄₇ DDISAVR	0.9998		6	1.00	25.89
STH ₁₄₇ PVLGGEEELVDK	0.9998		6	1.00	25.89
RGGALDDIYDR	0.9998		6	1.00	25.89
HATAMERFDLALRHQIDLR	0.9998		5	1.00	25.89
LLIDWFSSLR	0.9998		4	1.00	25.89
EDFRADAEK	0.9998		3	1.00	25.89
AGNELSTADLR	0.9998		3	1.00	25.89
WSA ₁₄₇ SPAERNR	0.9998		2	1.00	25.89
WSA ₁₄₇ SPAER	0.9998		2	1.00	25.89
LLAETADAVAR	0.9997		2	1.00	25.89
LFDALR	0.9998		2	1.00	25.89
WTSTLDGR ₁₂₆ SSIC ₁₂₆ R	0.9955		1	1.00	25.91
VVGTAKHYSDGVLTLR	0.9962		1	1.00	25.91
STHVPVLGGEEELVDKAFVTDTR	0.997		1	1.00	25.89
RDLWGLADGTPLSFTPDQEK	0.9992		1	1.00	25.89
NYSDGVLTLR	0.9998		1	1.00	25.89
LGLTQGENLDTIVRRVVGTK	0.9993		1	1.00	25.89
LGLTQGENLDTIVR	0.9998		1	1.00	25.89
LFDALRHQIDLR	0.9963		1	1.00	25.91
I _{LH} ₁₄₇ DDISAVREAAWK	0.9990		1	1.00	25.89
GGALDDIYDRAGNELSTADLR	0.9998		1	1.00	25.89
Main Entry Accession	NTT	Probability	# Tot PSMs	% Coverage / Weight	% Spectrum ids / NSP
ALH23689.1 major head protein [Pseudomonas phage PaMx11]	1.0000		1340	77.8%	100.00%
VTSELTIIIGDAEVNLIQATR	0.9990		373	1.00	15.39
LALDDLAVGVZENIVNRR	0.9990		151	1.00	15.39
SFNALLR	0.9990		136	1.00	15.39
GLAALPGVTN	0.9990		120	1.00	15.39
GLGGASIGDVLTPSGAEVPAYR	0.9990		113	1.00	15.39
SSDGNQTAIQVASK	0.9990		99	1.00	15.39
H ₁₄₇ Y ₁₄₇ EL ₁₄₇ ZIPFDGIAGNALAYNR	0.9990		89	1.00	15.39
WYWC ₁₆₀ GLALFSEK	0.9990		81	1.00	15.39
WYWC ₁₆₀ LALFSEK	0.9990		79	1.00	15.39
ENVLGDVVAVGVDATITAK	0.9990		70	1.00	15.39
NDFIPVNIQVGTSTNCL ₁₆₀ TVFAGTFDDGSR	0.9990		27	1.00	15.39
THGIAGLTAEAAAGIQIEDVGASETKDNVITR	0.9990		13	1.00	15.39
LDELMDITDKDGEVDVLTNHAR	0.9990		12	1.00	15.39
DGEVDVLTNHAR	0.9990		10	1.00	15.39
THGIAGLTAEAAAGIQIEDVGASETK	0.9990		5	1.00	15.39
HYEHPFDGIAGNALAYNRENVLGDVVAVGVDATITAK	0.4088		2	1.00	15.39
APATFTKYSELTIIGDAEVNLIQATR	0.9989		1	1.00	15.39
Main Entry Accession	NTT	Probability	# Tot PSMs	% Coverage / Weight	% Spectrum ids / NSP
ALH23691.1 virion structural protein [Pseudomonas phage PaMx11]	1.0000		13	34.7%	37.8%
VTEVFASDAGSLNLE	0.9994		7	1.00	0.91
VAMVPAAGR	0.9928		6	1.00	1.00
Main Entry Accession	NTT	Probability	# Tot PSMs	% Coverage / Weight	% Spectrum ids / NSP
ALH23693.1 virion structural protein [Pseudomonas phage PaMx11]	0.9994		17	9.3%	5.39%
VTPGATQQQLK	0.9994		17	1.00	0.01
Main Entry Accession	NTT	Probability	# Tot PSMs	% Coverage / Weight	% Spectrum ids / NSP
ALH23696.1 major tail structural protein [Pseudomonas phage PaMx11]	0.9058		7	6.1%	0.48%
NDRQZADR	0.8626		6	1.00	0.31
H ₁₄₇ MGRSH ₁₄₇ R	0.3148		1	1.00	0.01

Proteínas estructurales identificadas en el fago PaMx11. En cada bloque se muestran los péptidos identificados, la probabilidad de dicha identificación, número de espectros mapeados, cobertura, y fracción de espectros identificados por péptido con respecto al número de péptidos *hermanos*. Los ORFs identificados fueron los siguientes: ORF_03, ALH23677; ORF_04, ALH23678; ORF_15, ALH23689; ORF_17, ALH23691; ORF_19, ALH23693 & ORF_22, ALH23696

Anexo 5: Identificación de las proteínas estructurales del fago PaMx11

Main Entry Accession	NTT	Probability	# Tot PSMs	% Coverage / Weight	% Spectrum ids / NSP
ALH23699.1 tail tape measure protein [Pseudomonas phage PaMx11]	1.0000		52	69.0%	82.48%
LDKNDADQLAR	0.9999		56	1.00	36.05
YAAQW ₁₄₇ EGLYAAR	0.9999		49	1.00	36.05
LWNLQLEP ₁₅₈ SR	0.9998		39	1.00	36.05
LOGEQRW ₁₅₈ EVSLR	0.9999		39	1.00	36.05
SYAGEQSSAYR	0.9998		35	1.00	36.05
ATFSDDTIEAAGOR	0.9998		35	1.00	36.05
NTPVGLAQATEGFTK	0.9998		26	1.00	36.05
ALAGLEQOR	0.9999		20	1.00	36.05
AQELGFPANLAEM ₁₄₇ AR	0.9998		19	1.00	36.05
SQLFTEEEAZOESYNRR	0.9998		17	1.00	36.05
H ₁₄₇ KQIIAEIK	0.9999		17	1.00	36.05
DVDIQASR	1.0000		16	1.00	36.05
LKIDLGAASKAEGASSR	0.9999		14	1.00	36.05
LKIQDFDEN ₁₄₇ QDAEM ₁₄₇ AR	0.9999		13	1.00	36.05
IESVEALEAAR	1.0000		12	1.00	36.05
EASLTDIM ₁₄₇ R	1.0000		12	1.00	36.05
LKEFGDIK	0.9996		11	1.00	36.05
RKQIIAEIK	1.0000		9	1.00	36.05
LR EEYDRAAEAR	0.9998		8	1.00	36.05
EFENLVASLR	1.0000		7	1.00	36.05
EASLTDIM ₁₄₇ R	0.9999		7	1.00	36.05
ALHRYGNTAAAM ₁₄₇ QIK	0.9999		7	1.00	36.05
TEEEALAEYIK	0.9999		6	1.00	36.05
EKALAGLEQOR	0.9999		6	1.00	36.05
TQALLTSGEQLFGGLAGLAK	0.9998		6	1.00	36.05
SLDASSRAEAEOR	0.9999		6	1.00	36.05
LRAPVATLATAM ₁₄₇ ALNKK	0.9998		6	1.00	36.05
SQLFTEEEAZOESYNR	0.9998		4	1.00	36.05
NTGAGS ₁₄₇ LDIMAR	0.9875		4	1.00	36.81
IESVEALEAARR	1.0000		4	1.00	36.05
DLGGAASKAEGASSR	0.9999		4	1.00	36.05
ALFAVSIK	0.9998		4	1.00	36.05
VAEER ₁₄₇ KALAGLEQOR	1.0000		3	1.00	36.05
DVDIQASRDKAAAR	1.0000		3	1.00	36.05
EFGIKSSSEGOR	0.9865		2	1.00	36.05
EEYDRAAEAR	0.9921		2	1.00	36.73
YAAQW ₁₄₇ EGLYAAR EASLTDIM ₁₄₇ R	0.9862		1	1.00	36.64
RKQIIAEIKSLDASSR	0.7645		1	1.00	36.05
NAAEIEQYLMALGENEFAGAM ₁₄₇ DNR	0.9999		1	1.00	36.05
LVRAAYAVDAIK	0.9992		1	1.00	36.07
DM ₁₄₇ SOM ₁₄₇ IEAVADATTGFEFLKEFGDIK	0.9989		1	1.00	36.09
Main Entry Accession	NTT	Probability	# Tot PSMs	% Coverage / Weight	% Spectrum ids / NSP
ALH23700.1 viiion structural protein [Pseudomonas phage PaMx11]	1.0000		26	37.8%	86.26%
IVSFLTTPALVAANDWEVLR	0.9999		63	1.00	12.96
IIVDIPFSSAVGDKR	0.9999		30	1.00	12.96
HGVYPPGAPSDPTSDNR	0.9999		30	1.00	12.96
DNVAVTNNLFEPSTATHR	0.9999		30	1.00	12.96
YSYANR	0.9999		21	1.00	12.96
APGNSGSESYVGIK	0.9999		21	1.00	12.96
TTTSTYVNSM ₁₄₇ ZR	0.9999		19	1.00	12.96
TVVVFQAYR	0.9999		15	1.00	12.96
DNVAVTNNLFEPSTATHRK	0.9999		15	1.00	12.96
NTVHEFWALSD	0.9999		11	1.00	12.96
INPVSPTPGAHAYWR	0.9999		8	1.00	12.96
WATLYLR	0.9999		6	1.00	12.96
H ₁₄₇ L ₁₄₇ LLTDSGEVANHFGEVZLR	0.9999		6	1.00	12.96
EC ₁₄₇ LGGGYLQPC ₁₄₇ DM ₁₄₇ LQK	0.9995		2	1.00	12.96
Main Entry Accession	NTT	Probability	# Tot PSMs	% Coverage / Weight	% Spectrum ids / NSP
ALH23702.1 viiion structural protein [Pseudomonas phage PaMx11]	1.0000		26	43.2%	75.98%
SIDM ₁₄₇ ZONGSGTISVDIK	0.9999		37	1.00	20.96
YTFNFPDQKPTLR	0.9999		31	1.00	20.96
VWVFNH ₁₄₇ PQTK	0.9999		30	1.00	20.96
TVLVEDLNER	0.9999		30	1.00	20.96
VYIGPVWAE ₁₄₇ TR	0.9999		24	1.00	20.96
AVTSTPVM ₁₄₇ ETP	0.9999		21	1.00	20.96
VTANKDLGSGSLPQYR	0.9999		20	1.00	20.96
TSHVDSAK	0.9999		19	1.00	20.96
SNIDFTIASAEFR	0.9999		15	1.00	20.96
TYSLSTINGAPVZNAK	0.9999		15	1.00	20.96
H ₁₄₇ PSFLDDYIYR	0.9999		10	1.00	20.96
RNSIDFTIASAEFR	0.9999		9	1.00	20.96
QATHEFOLDTYQFSR	0.9999		7	1.00	20.96
STFDLDRQER	0.9999		6	1.00	20.96
IDIPHDYAGR	0.9999		6	1.00	20.96
NGTQVFSR	0.9999		6	1.00	20.96
AZATQWHR	0.9999		6	1.00	20.96
TVLVEDLNERISR	0.9999		3	1.00	20.96
LPLGSSYLNAVAPLRFAR	0.9999		2	1.00	20.96
GKDVLTDRTVLVEDLNER	0.9999		2	1.00	20.96
QRAFVLP ₁₄₇ WNR	0.9999		1	1.00	20.96
AKAIATQWHRVYIGPVWAE ₁₄₇ TR	0.9999		1	1.00	20.96
Main Entry Accession	NTT	Probability	# Tot PSMs	% Coverage / Weight	% Spectrum ids / NSP
ALH23704.1 viiion structural protein [Pseudomonas phage PaMx11]	0.9995		6	13.0%	1.61%
EIPVLFGR	0.9994		6	1.00	0.28
EIPVLFGRK	0.2771		1	1.00	1.00

Proteínas estructurales identificadas en el fago PaMx11. En cada bloque se muestran los péptidos identificados, la probabilidad de dicha identificación, número de espectros mapeados, cobertura, y fracción de espectros identificados por péptido con respecto al número de péptidos hermanos. Los ORFs identificados fueron los siguientes: ORF_25, ALH23699; ORF_26, ALH23700; ORF_28, ALH23702 & ORF_30, ALH23704.

Anexo 5: Identificación de las proteínas estructurales del fago PaMx11

Main Entry Accession	NTT	Probability	# Tot PSMs	% Coverage / Weight	% Spectrum ids / NSP
ALH23706.1 <i>vijion structural protein [Pseudomonas phage PaMx11]</i>		1.0000	86	31.6%	29.5%
NVPSTAGYAVVTGVRPSSDATNAK	■	0.9999	17	1.00	15.67
EGGVSGRVDIDH ₁₄ GGPTQGR	■	0.9999	14	1.00	15.67
AYKPYPPGQLR	■	0.9999	12	1.00	15.67
GVLDTVATTHPLGTR	■	0.9999	8	1.00	15.67
VDIDH ₁₄ GGPTQGR	■	0.9999	6	1.00	15.67
WAFWASR	■	0.9982	5	1.00	15.69
LLPSTGGSTLGINSATQNVPEVGR	■	0.9999	4	1.00	15.67
ALSTPLASATZYANRK	■	0.9999	4	1.00	15.67
NSYLQGR	■	0.9989	3	1.00	15.68
DGSSLSLQHWFEFR	■	0.9999	3	1.00	15.67
VTAVSDTSLTVGR	■	0.9966	2	1.00	15.70
QFPFGTNGTLATOAAAR	■	0.9998	2	1.00	15.67
GDQNZSVSWAHRDR	■	0.9820	2	1.00	15.80
SGYGSYGYSYGGV	■	0.9999	1	1.00	15.67
SEINGDMPAHLIR	■	0.9903	1	1.00	15.75
IQLWVSR	■	0.9999	1	1.00	15.67
INGQOYPSVSR	■	0.9929	1	1.00	15.73

Proteínas estructurales identificadas en el fago PaMx11. En cada bloque se muestran los péptidos identificados, la probabilidad de dicha identificación, número de espectros mapeados, cobertura, y fracción de espectros identificados por péptido con respecto al número de péptidos hermanos. En la imagen se muestran los péptidos correspondientes al ORF_32.

Anexo 5: Identificación de las proteínas estructurales del fago PaMx25

Main Entry Accession	NTT	Probability	# Tot PSMs	% Coverage / Weight	% Spectrum ids / NSP
ALH23823.1 portal protein [Pseudomonas phage PaMx25]	1.0000	100%	114	51.1%	53.3%
ATAYGELELLLR	1.0000	100%	200	1.00	2876
DIIVTPDEPPLGLGR	1.0000	100%	90	1.00	2876
SGDLNNSGQNTESGSALQTR	1.0000	100%	99	1.00	2876
YIGVNSQGLSEQR	1.0000	100%	69	1.00	2876
VTPNLEAFDFEHL47SGK	1.0000	100%	65	1.00	2876
RINEEQLVYGR	1.0000	100%	61	1.00	2876
TPTIELPANH147EALRDK	1.0000	100%	60	1.00	2876
TPTIELPANH147EALR	1.0000	100%	56	1.00	2876
INEEQLVYGR	1.0000	100%	46	1.00	2876
STDFEIVSQTK	1.0000	100%	42	1.00	2876
LALAVYR	1.0000	100%	41	1.00	2876
TGAGSH47EVEGGDAK	1.0000	100%	35	1.00	2876
AKGETYLPPTK	0.9999	100%	35	1.00	2876
NIDGGEADEGEAR	1.0000	100%	35	1.00	2876
IPFVFNITK	1.0000	100%	30	1.00	2876
IC136A0M147GANPPDDVK	1.0000	100%	24	1.00	2876
DLVDFH47TAR	1.0000	100%	20	1.00	2876
AVFDYVK	1.0000	100%	18	1.00	2876
YSEFKEDWVH147R	1.0000	100%	16	1.00	2876
GVTLDCIPFVFNITK	0.9999	100%	15	1.00	2876
TDATEDEGTPLR	1.0000	100%	14	1.00	2876
DKATAYGELELLLR	1.0000	100%	13	1.00	2876
LALAVYRGEADYR	0.9998	100%	11	1.00	2876
GETYLPPTK	1.0000	100%	9	1.00	2876
QNLFM147GGDTLVVGGVK	1.0000	100%	8	1.00	2876
EAYDAVK	0.2320	100%	7	1.00	2953
LNLVVLDESGFRR	0.9982	100%	4	1.00	2884
LNLVVLDESGFR	0.9996	100%	3	1.00	2876
RSTDFEIVSQTK	0.9994	100%	2	1.00	2876
TH147GAPLSK	0.9994	100%	1	1.00	2876
KSIH147LVDR	0.9909	100%	1	1.00	2905
ALH23820.1 major head protein [Pseudomonas phage PaMx25]	1.0000	100%	149	46.2%	99.0%
SSGVFNWLFGR	0.9992	100%	436	1.00	1163
VSDYTR	0.9992	100%	264	1.00	1163
NNLIDFVSDVNGNAVR	0.9992	100%	179	1.00	1163
IPTFLGR	0.9992	100%	176	1.00	1163
IADVVPPELFSPYVQ0M147TQEK	0.9992	100%	150	1.00	1163
IGTATEQVR	0.9992	100%	94	1.00	1163
KPSAGNGGG0DLFNR	0.9992	100%	90	1.00	1163
DLNDNAENVSTDDPATDSTPNK	0.9992	100%	91	1.00	1163
EVIVDDGVPR	0.9992	100%	74	1.00	1163
VPTVDR	0.9992	100%	20	1.00	1163
VPTVDRKPSAGNGGG0DLFNR	0.7370	100%	10	1.00	1163
SGAIVLDAQLNSALAGGLTFNEPSFK	0.9666	100%	1	1.00	1168
H147LQNLIDFVSDVNGNAVR	1.0000	100%	1	1.00	1163
ALH23816.1 vision structural protein [Pseudomonas phage PaMx25]	1.0000	100%	403	52.1%	79.8%
GLVVSASQVIR	1.0000	100%	237	1.00	896
TGAADFADTDIPVK	1.0000	100%	79	1.00	896
TTTQALAWPR	0.9999	100%	70	1.00	896
ALTAELAPDLK	1.0000	100%	73	1.00	896
ATDYIEGR	0.9999	100%	50	1.00	896
AC180C180EYALR	1.0000	100%	25	1.00	896
VDAAGLVVATIQC	1.0000	100%	17	1.00	896
KTTTQALAWPR	1.0000	100%	16	1.00	896
RAC180C180EYALR	1.0000	100%	14	1.00	896
GVTWGTGADAVK	1.0000	100%	14	1.00	896
VDAAGLVVATIK	0.0766	100%	4	1.00	930
ALH23815.1 vision structural protein [Pseudomonas phage PaMx25]	1.0000	100%	30	30.1%	3.0%
LDATAADTSHPWKGGAPAVALAK	0.9964	100%	26	1.00	341
LDATAADTSHPWK	0.9943	100%	26	1.00	332
GPGAPAVALAK	0.9890	100%	26	1.00	332
VLPVFPVPSGSLGR	0.9842	100%	20	1.00	332
DIVKEELLDR	0.4179	100%	2	1.00	330
ALH23814.1 tail terminator protein [Pseudomonas phage PaMx25]	1.0000	100%	24	49.2%	2.0%
TGTVYIQVFPVGGDSK	0.9977	100%	15	1.00	239
AGIDAKQLVDFAFQAR	0.9977	100%	11	1.00	239
H124HEAGTSGAFQQLNVLATFSYDDVR	0.1734	100%	1	1.00	239
DILGFFK	0.9686	100%	1	1.00	231

Proteínas estructurales identificadas en el fago PaMx25. En cada bloque se muestran los péptidos identificados, la probabilidad de dicha identificación, número de espectros mapeados, cobertura, y fracción de espectros identificados por péptido con respecto al número de péptidos hermanos. Los ORFs identificados fueron los siguientes: ORF_07, ALH23823; ORF_10, ALH23820; ORF_13, ALH23817; ORF_14, ALH23816; ORF_15, ALH23815 & ORF_16, ALH23814.

Anexo 5: Identificación de las proteínas estructurales del fago PaMx25

Main Entry Accession	NTT	Probability	# Tot PSMs	% Coverage / Weight	% Spectrum ids / NSP
ALH23813.1 major tail structural protein [Pseudomonas phage PaMx25]	1.0000		79	72.2%	39.8%
LAVEQDQAITLPLENNAEAK	0.9997		125	1.00	19.23
IFFGSLVLIKESDPTLTK	0.9997		80	1.00	19.23
NLALLNDIPLLSLGDGR	0.9997		61	1.00	19.23
SGTRPTLVPGSAFNTSSDFSR	0.9997		57	1.00	19.23
FGHTLLFQSFVYLPDVAGGV	0.9997		52	1.00	19.23
SVASASTAVVINEVITDEASPPAAK	0.9997		43	1.00	19.23
KTTAPH ₁₄ NTAATPTIGVTGASK	0.9997		42	1.00	19.23
IFFGSLVLIK	0.9997		39	1.00	19.23
LN ₁₄ WGFFFADIR	0.9997		36	1.00	19.23
IDSNVTGLR	0.9997		33	1.00	19.23
FAEESLTK	0.9997		32	1.00	19.23
VSVNMSYLFEDK	0.9997		24	1.00	19.23
NPZINPSR	0.9997		19	1.00	19.23
TTAPH ₁₄ NTAATPTIGVTGASK	0.9997		18	1.00	19.23
ASGTDFTTLGLAGEWVYLGDDAVGTQFNRR	0.9997		16	1.00	19.23
TLPSPVWVPLPNSYADFQGGIATVAR	0.9997		11	1.00	19.23
NNSDVTIDV ₁₄ LJK	0.9997		7	1.00	19.23
IFFGSLVLIKESDPTLTKR	0.9997		6	1.00	19.23
VSNTPDAEVTGDK	0.9997		4	1.00	19.23
TYAAAAGLTFLAN ₁₄ W ₁₄ ASFGVGTANGLK	0.9892		3	1.00	19.23
LEAVGYOFASATV ₁₄ W ₁₄ ASFGVGTANGLK	0.2512		1	1.00	19.98
LALASSTDAAPTPLFAFATENSLVNNVSPNCK	0.0668		1	1.00	20.23
ALH23810.1 virion structural protein [Pseudomonas phage PaMx25]	1.0000		103	50.1%	16.02%
DALDGTGAAZAEQGAETAK	0.9995		35	1.00	3.09
SNWVNLGAPADGTR	0.9995		29	1.00	3.09
LDNGISAGAPSGFVEK	0.9995		29	1.00	3.09
ARSNWVNLGAPADGTR	0.7684		6	1.00	3.46
VALAVDSAVVZATPVDGTR	0.6189		1	1.00	3.03
ALH23809.1 tail tape measure protein [Pseudomonas phage PaMx25]	1.0000		180	63.7%	64.20%
SGGITLTOEETAALR	0.9961		236	1.00	52.29
AEEFNISILEGAPILQOVVANHIDGAEQSGIK	0.9961		102	1.00	52.29
NLEDVQNTAEK	0.9961		73	1.00	52.29
TIDATAGLLTGLGIAVR	0.9961		66	1.00	52.29
FHYDDIPLGK	0.9961		57	1.00	52.29
QGFTEIVELYSR	0.9997		56	1.00	52.29
LGWITGV ₁₄ LN ₁₄ W ₁₄ STFK	0.9961		49	1.00	52.29
SLNTIAPVEGAK	0.9961		47	1.00	52.29
AAAITQAI ₁₄ NTYEGATK	0.9961		47	1.00	52.29
ALGALTAFAVDQIR	0.9961		46	1.00	52.29
YADAWASAGQIR	0.9961		46	1.00	52.29
NFFSGLASLQSNIEELAAIGK	0.9961		46	1.00	52.29
SGKELGASQALTK	0.9997		40	1.00	52.29
NFFSGLASLQSNIEK	0.9961		40	1.00	52.29
ALGALTAFAVDQIRK	0.9961		40	1.00	52.29
IATKDIAEAAAVQEQLYK	0.9961		39	1.00	52.29
KYADAWASAGQIR	0.9997		35	1.00	52.29
ALAOGGIVGAHAAVVVAQGNQAVAIR	0.9961		33	1.00	52.29
SAVM ₁₄ DLLLETVKFDR	0.9961		31	1.00	52.29
DAEAAAVQEQLYK	0.9961		31	1.00	52.29
TLAASSTSAQDQAGLH ₁₄ W ₁₄ Q ₁₄ ALGSGQIVR	0.9961		29	1.00	52.29
WVDGVFTR	0.9961		28	1.00	52.29
SAVM ₁₄ DLLLETVK	0.9997		25	1.00	52.29
ASFTISQGFLLISNAFH ₁₄ R	0.9997		25	1.00	52.29
QNDLSEQDYSNAK	0.9997		24	1.00	52.29
ELGASQALTK	0.9997		23	1.00	52.29
AASGV ₁₄ ILIR	0.9997		23	1.00	52.29
QNELVAQDDH ₁₄ LANSVEQR	0.9961		22	1.00	52.29
AVGEDIAVFGDLK	0.9997		16	1.00	52.29
FHYDDIPLGK ₁₄ NR	0.9961		15	1.00	52.29
FFETYGNIAQSIDAGFQDGGFH ₁₄ EK	0.9961		15	1.00	52.29
LEH ₁₄ WCAEETLNK	0.9995		14	1.00	52.29
AFAGVPDAI ₁₄ ELFNR	0.9996		14	1.00	52.29
WIGQTD ₁₄ SALGSH ₁₄ LG ₁₄ W ₁₄ K	0.9997		13	1.00	52.29
VOLAQR ₁₄ ENELK	0.9979		13	1.00	52.29
LLGH ₁₄ SR	0.9997		13	1.00	52.29
GLITTDQAR	0.9997		12	1.00	52.29
AFDANLAGSEETDAK	0.9997		12	1.00	52.29
VOLAQR ₁₄ ENELK ₁₄ R	0.9997		11	1.00	52.29
SAVM ₁₄ DLLLETVKFDRK	0.9973		10	1.00	52.29
NKLAQYTFYDGVGSGFAGVAK	0.9961		10	1.00	52.29
FTEN ₁₄ GK	0.9991		10	1.00	52.29
RALGALTAFAVDQIRK	0.9990		9	1.00	52.29
AFSPD ₁₄ VELEK	0.9990		9	1.00	52.29
QNELVAQDDH ₁₄ LANSVEQR	0.9967		7	1.00	52.29
AASGV ₁₄ DLK	0.9996		7	1.00	52.29
EL ₁₄ EKL ₁₄ NELR	0.7581		6	1.00	52.81
AR ₁₄ OVES ₁₄ YV ₁₄ IEK	0.9997		6	1.00	52.29
NLEDVQNTAEK ₁₄ ASGV ₁₄ DLK ₁₄ R	0.9787		4	1.00	52.36
TYNVAQV ₁₄ YV ₁₄ NI ₁₄ Z ₁₄ YV ₁₄ NR	0.9965		3	1.00	52.30
QVDTKFFETYGNIAQSIDAGFQDGGFH ₁₄ EK	0.9997		2	1.00	52.29
LEMAK ₁₄ AEETLNK ₁₄ AVAK	0.9435		2	1.00	52.46
ALGALTAFAVDQIRCYADAWASAGQIR	0.9997		2	1.00	52.29
RALGALTAFAVDQIR	0.1340		1	1.00	53.15
QVESVH ₁₄ K ₁₄ IEK	0.1230		1	1.00	53.16
FHYDDIPLGK ₁₄ NR ₁₄ ENDEQIR	0.0762		1	1.00	53.29
AFDANLAGSEETDAK ₁₄ FGK	0.9652		1	1.00	53.40

Proteínas estructurales identificadas en el fago PaMx25. En cada bloque se muestran los péptidos identificados, la probabilidad de dicha identificación, número de espectros mapeados, cobertura, y fracción de espectros identificados por péptido con respecto al número de péptidos hermanos. Los ORFs identificados fueron los siguientes: ORF_17, ALH23813; ORF_20, ALH23810 & ORF_21, ALH23809.

Anexo 5: Identificación de las proteínas estructurales del fago PaMx25

Main Entry Accession	NTT	Probability	# Tot PSMs	% Coverage / Weight	% Spectrum ids / NSP
ALH23808.1 <i>virion structural protein [Pseudomonas phage PaMx25]</i>		0.9999	13	7.3%	0.99%
APLSSSHYPOH ₁₄₂ ZR		0.8642	3	1.00%	1.49
DNIAALTNLTEPTNVOYR		0.9982	4	1.00%	1.36
YSYANR		0.4847	11	1.00%	1.86
Main Entry Accession	NTT	Probability	# Tot PSMs	% Coverage / Weight	% Spectrum ids / NSP
ALH23807.1 <i>virion structural protein [Pseudomonas phage PaMx25]</i>		1.0000	849	96.2%	3170%
LFZAIAQGTANR		0.9989	196	1.00%	9.70
AAFQSFQNWAGSGR		0.9998	77	1.00%	9.70
VFPYVAK		0.9998	66	1.00%	9.70
GADYSSNSGYSYGSM ₁₄₇ VR		0.9998	65	1.00%	9.70
FGC ₁₆₀ 5SLDLFNPAPGGGR		0.9989	62	1.00%	9.70
GSAYFNFR		0.9998	27	1.00%	9.70
GIAYLVV		0.9989	24	1.00%	9.70
YGITLNGSDGYAAGNAWDKQGVYVVR		0.9998	14	1.00%	9.70
LFZAIAQGTANRWATVGGSGR		0.9433	11	1.00%	9.91
YM ₁₄₇ DM ₁₄₇ TNVLPDGEFTIGSDTWK		0.2763	2	1.00%	10.42
TLYAEVEITTGSM ₁₄₇ R		0.9386	2	1.00%	9.91
GTTPNEQ ₁₆₀ VLFPLVTGGPPAYHFFAPDSK		0.9998	11	1.00%	9.70
Main Entry Accession	NTT	Probability	# Tot PSMs	% Coverage / Weight	% Spectrum ids / NSP
ALH23806.1 <i>virion structural protein [Pseudomonas phage PaMx25]</i>		1.0000	77	17.4%	5.42%
APSFDDVYVR		1.0000	31	1.00%	7.83
FLLEDAQFSR		1.0000	29	1.00%	7.83
LPLETNPK		1.0000	19	1.00%	7.83
TDIDSFNWEOR		1.0000	12	1.00%	7.83
AGIWAALR		0.9999	8	1.00%	7.83
SAVVSLSER		1.0000	7	1.00%	7.83
DVGQVAATIAIAETPES		1.0000	7	1.00%	7.83
TVEVWSAYFDPK		1.0000	4	1.00%	7.83
HMKTIGFK		0.9996	11	1.00%	7.91
Main Entry Accession	NTT	Probability	# Tot PSMs	% Coverage / Weight	% Spectrum ids / NSP
ALH23805.1 <i>virion structural protein [Pseudomonas phage PaMx25]</i>		1.0000	62	18.0%	6.32%
LGQTFPSPVVR		0.9993	59	1.00%	2.00
TSQWQIEVQC ₁₆₀ ESVFTSISK		0.9993	2	1.00%	2.00
QTTDTFHDISK		0.0722	1	1.00%	3.00
NPYGSSIA		0.9993	1	1.00%	2.00
Main Entry Accession	NTT	Probability	# Tot PSMs	% Coverage / Weight	% Spectrum ids / NSP
ALH23804.1 <i>virion structural protein [Pseudomonas phage PaMx25]</i>		0.9669	7	11.7%	1.22%
ETPVLFGR		0.9669	7	1.00%	1.01
Main Entry Accession	NTT	Probability	# Tot PSMs	% Coverage / Weight	% Spectrum ids / NSP
ALH23802.1 <i>virion structural protein [Pseudomonas phage PaMx25]</i>		1.0000	304	53.0%	23.69%
GVLDSPVVMHALGAR		0.9997	53	1.00%	31.11
GTLAEGSAPVQTVTIQAR		0.9997	46	1.00%	31.11
SVTLEEFIDVILK		0.9997	40	1.00%	31.11
VANVELGALDSNVVK		0.9997	36	1.00%	31.11
HIDGSLFVDR		0.9997	31	1.00%	31.11
QNGIAQWYSK		0.9997	24	1.00%	31.11
QFPGFTNKIATK		0.9996	24	1.00%	31.11
DGQIQLODHWQWFR		0.9997	24	1.00%	31.11
RPSFWGR		0.9990	19	1.00%	31.11
INTQAVPDTVR		0.9997	19	1.00%	31.11
ALSVPLASGTIYANR		0.9997	19	1.00%	31.11
IQLWAVR		0.9983	16	1.00%	31.12
LGALVPGFR		0.9987	14	1.00%	31.11
GVVSAVFR		0.9996	14	1.00%	31.11
YGIQLW ₁₄₇ R		0.9993	13	1.00%	31.11
GDQDITVAIVHR		0.9996	13	1.00%	31.11
AAIGDM ₁₄₇ NPAPHIR		0.9994	13	1.00%	31.11
LGQATLVDTAAASIGPEANTTYTC ₁₆₀ R		0.9997	11	1.00%	31.11
LGALVPGFRGVVSAVFR		0.9982	11	1.00%	31.12
AASLNIGDGVVLSWPR		0.9987	9	1.00%	31.11
QVVLGH ₁₄₇ NPVLYK		0.9986	8	1.00%	31.11
NTIGELINSVTVVWDASTGK		0.9997	8	1.00%	31.11
ALSVPLASGTIYANR		0.9979	7	1.00%	31.12
VFPDQVVFETDVEYATGEVAR		0.9987	6	1.00%	31.11
NNSVTVQDIALAQQQSVVSTTK		0.9997	5	1.00%	31.11
LLPTTAK		0.9633	5	1.00%	31.23
GDQDITVAIVHRDR		0.9794	5	1.00%	31.18
GGDINSLLVLDSSVDK		0.9974	4	1.00%	31.15
IKLPLTTAK		0.9327	3	1.00%	31.26
GGDINSLLVLDSSVDKDFR		0.9997	3	1.00%	31.11
DRLOQTATLVDTAAASIGPEANTTYTC ₁₆₀ R		0.9902	2	1.00%	31.14
INTQAVPDTVRGDQDITVAIVHR		0.8634	1	1.00%	31.48
EGVSGAVDIDM ₁₄₇ GGDAQTPNSYLVSK		0.9997	1	1.00%	31.11

Proteínas estructurales identificadas en el fago PaMx25. En cada bloque se muestran los péptidos identificados, la probabilidad de dicha identificación, número de espectros mapeados, cobertura, y fracción de espectros identificados por péptido con respecto al número de péptidos hermanos. Los ORFs identificados fueron los siguientes: ORF_22, ALH23808; ORF_23, ALH23807; ORF_24, ALH23806; ORF_25, ALH23805; ORF_26, ALH23804 & ORF_28, ALH23802.

Anexo 5: Identificación de las proteínas estructurales del fago PaMx28

Main Entry Accession	NTT	Probability	# Tot PSMs	% Coverage / Weight	% Spectrum ids / NSP
ALH23608.1 portal protein [Pseudomonas phage PaMx28]					
FPLLAASGVPADK	1.0000	100	66.4%	91.96%	21.87
ATLLNM ₁₄₇ TEQTLDTZAGKPR	0.9992	92	1.00	1.00	21.87
VGWDEVVTR	0.9992	71	1.00	1.00	21.87
VTIGPNFLTTEASDGK	0.9992	72	1.00	1.00	21.87
ATVATPSSAYQR	0.9992	67	1.00	1.00	21.87
WNPETGEPVVR	0.9992	62	1.00	1.00	21.87
ALD ₅ ESSYLASTVR	0.9992	59	1.00	1.00	21.87
GLSHLVLEHPTPEER	0.9992	49	1.00	1.00	21.87
DFLPOYENETMK	0.9992	39	1.00	1.00	21.87
WYVYDHTGAZIK	0.9992	31	1.00	1.00	21.87
DAGRDFLPOYENETMK	0.9992	30	1.00	1.00	21.87
KRPGDQATGR	0.9992	29	1.00	1.00	21.87
INADVLDSEADAELDALIK	0.9992	20	1.00	1.00	21.87
WELWAPDEK	0.9992	19	1.00	1.00	21.87
M ₁₄₇ TDVLGGTEAM ₁₄₇ R	0.9992	15	1.00	1.00	21.87
EOVVLGGDVP ₅₀ EELTEDIDM ₁₄₇ IGNNLHAF ₃₄₀ R	0.9992	10	1.00	1.00	21.87
VLEPGRWELWAPDEK	0.9292	9	1.00	1.00	21.96
DFODC ₁₆₀ VELM ₁₄₇ QYTDWLGLEDGGSVR	0.9992	7	1.00	1.00	21.87
KTLLDENK	0.9989	6	1.00	1.00	21.87
SGAEDLASLEDQMATYGAEFMR	0.9936	6	1.00	1.00	21.86
NVLTVSR	0.8930	6	1.00	1.00	22.01
TLLDENK	0.9902	4	1.00	1.00	21.88
IRKETTVER	0.1451	3	1.00	1.00	22.73
KATVATPSSAYQR	0.9036	2	1.00	1.00	21.99
GEEHVEDGTTALGYVPLVTFYAGK	0.1109	1	1.00	1.00	22.76
DFLPOYENETM ₁₄₇ EAR	0.0658	1	1.00	1.00	22.87
ALH23610.1 morphoanthesis protein [Pseudomonas phage PaMx28]					
AKAAGYGVKTKNK	1.0000	20	50.0%	25.00%	35.36
APIPGGVLPEVAGIK	0.9998	14	1.00	1.00	35.29
VGTPPOAGFAIK	0.9998	12	1.00	1.00	35.29
LGLLAGETNDQLVAR	0.9998	11	1.00	1.00	35.29
IEQMR.LGLLAGETNDQLVAR	0.9998	11	1.00	1.00	35.29
TGVNVSTFSK	0.9998	10	1.00	1.00	35.29
SVFGSIQEPGASDLEELAK	0.9998	10	1.00	1.00	35.29
AVVENLEQALPANK	0.9998	10	1.00	1.00	35.29
ADPDGYAKASTLAEKSTVK	0.9836	10	1.00	1.00	35.37
STVAPVLAGEEVQDRPTZDTR	0.9998	8	1.00	1.00	35.29
LGLLAGETNDQLVARQR	0.9879	8	1.00	1.00	35.42
FNFSQRAAGLR	0.9761	8	1.00	1.00	35.42
AVAGSPINQVPLEGRIGK	0.9998	7	1.00	1.00	35.29
LEGTEAGLDAK	0.9998	7	1.00	1.00	35.29
YAGSVEAGKLSMNAQQR	0.9090	6	1.00	1.00	35.62
MPTANEELNATIRHOIK	0.9822	6	1.00	1.00	35.37
GLELLKQGVGNISDVATTLDK	0.9710	6	1.00	1.00	35.42
AAGPYLDAFAAPK	0.9998	6	1.00	1.00	35.29
LKEAADYELSK	0.9998	6	1.00	1.00	35.29
EGLPPPPR	0.9996	6	1.00	1.00	35.29
EAAGDKWGH ₁₄₇ SEAR	0.1192	5	1.00	1.00	36.17
VPTGGLAQAK	0.9998	4	1.00	1.00	35.29
SVLAEFPEAK	0.9997	4	1.00	1.00	35.29
EFODDVLGK	0.9997	4	1.00	1.00	35.29
YQVASKAAGKPYLDAFAAPK	0.9729	3	1.00	1.00	35.41
TRAKREVEFR	0.9138	3	1.00	1.00	35.60
SLQPLTEAEAAAIR	0.9998	3	1.00	1.00	35.29
QGPYSLQDLK	0.9993	3	1.00	1.00	35.29
AELGGKLVVQVPGVGLK	0.9998	3	1.00	1.00	35.29
YASDPHLQAYVDAHQGLAK	0.9998	3	1.00	1.00	35.29
SQLVKDGLLPNK	0.0806	2	1.00	1.00	35.29
QEYDPASTSLASASYKSELNK	0.4305	2	1.00	1.00	35.86
LRA ₁₄₇ LADVRR	0.1766	2	1.00	1.00	36.12
ALLQGGH ₁₄₇ PDQVVLK	0.8433	2	1.00	1.00	35.75
AAVSGPVAPPK	0.9998	2	1.00	1.00	35.29
VYGRSARQINFGQAELOK	0.2538	1	1.00	1.00	36.04
WVATLDGR	0.1309	1	1.00	1.00	36.16
WKGH ₁₄₇ SEARR	0.0803	1	1.00	1.00	36.29
VLNEALR	0.9985	1	1.00	1.00	35.30
SYTGSTYRVLNEALR	0.2802	1	1.00	1.00	36.01
QHDLLLPDDVLAAVKASAK	0.0803	1	1.00	1.00	36.29
QHDLLLPDDVLAAVK	0.9997	1	1.00	1.00	35.29
GYTANPKLEGFZAGLDAK	0.9206	1	1.00	1.00	35.69
EVDFRAEAR.EAAGDK	0.4305	1	1.00	1.00	35.86
EFQDDVLGGCAQLFR	0.3606	1	1.00	1.00	35.90
DKWADENGGAPPTSTNYEQLR	0.3573	1	1.00	1.00	35.94
DKWADENGGAPPTSTNYEK	0.9765	1	1.00	1.00	35.42
DGVLDTTR.EAEN ₁₄₇ EAR	0.3966	1	1.00	1.00	35.90
AGLFRDGH ₁₄₇ PLDKFVDEK	0.1342	1	1.00	1.00	36.16
ALH23612.1 vision structural protein [Pseudomonas phage PaMx28]					
VZK.EGEEYVVR	1.0000	1	25.0%	0.70%	2.66
AFESEAPNSGVIQATQKPAQK	0.7816	3	1.00	1.00	2.67
M ₁₄₇ DFEFLK	0.3739	2	1.00	1.00	3.06
GFOAQLQTK	0.9989	2	1.00	1.00	2.44
YLVTAAVQAZAGHK	0.5110	1	1.00	1.00	2.90
EGEEYVVR	0.0658	1	1.00	1.00	3.44

Proteínas estructurales identificadas en el fago PaMx28. En cada bloque se muestran los péptidos identificados, la probabilidad de dicha identificación, número de espectros mapeados, cobertura, y fracción de espectros identificados por péptido con respecto al número de péptidos hermanos. Los ORFs identificados fueron los siguientes: ORF_08, ALH23608; ORF_10, ALH23610 & ORF_12, ALH23612.

Anexo 5: Identificación de las proteínas estructurales del fago PaMx28

Main Entry Accession	NTT	Probability	# Tot PSMs	% Coverage / Weight	% Spectrum ids / NSP
ALH23619.1 major head protein [Pseudomonas phage PaMx28]		1.0000	259	68.0%	93.61%
AAATFTQVSTLTIIGDAEVNGLIQATR	■	0.9996	620	1.00	13.96
NDWIFZINQITGTSTNTTTFVAGTLDGSR	■	0.9996	507	1.00	13.96
VWYIC ₁₆₀ GLALFNK	■	0.9996	402	1.00	13.96
AYLALLR	■	0.9996	273	1.00	13.96
ENALGDVH ₁₄₇ AGVGSITIAK	■	0.9996	206	1.00	13.96
EVGKEDADESITR	■	0.9996	154	1.00	13.96
GLAALLEGVTN	■	0.9996	131	1.00	13.96
SHGIAGLTAER	■	0.9996	70	1.00	13.96
ALGGTSPGDIVTH ₁₄₇ ASGR	■	0.9996	70	1.00	13.96
HASZINIK	■	0.9996	54	1.00	13.96
AVQIASK	■	0.9996	50	1.00	13.96
RAYLALLR	■	0.9996	41	1.00	13.96
SGDNDQK	■	0.9996	19	1.00	13.96
EDADESITR	■	0.9996	6	1.00	13.96
LAQDELAVGVZENVTVNEHFELLPFDIGNALAYNR	■	0.9996	4	1.00	13.96
AYLALLRALGGTSPGDIVTH ₁₄₇ ASGR	■	0.0872	1	1.00	14.98
ALH23621.1 tail structural protein [Pseudomonas phage PaMx28]		1.0000	722	75.4%	80.61%
GLPPALVSAAC ₁₆₀ H ₁₄₇ LADR	■	0.9999	316	1.00	11.26
VQNEGTIEGANSYVDAAAVR	■	0.9999	142	1.00	11.26
TDEELVSAAVR	■	0.9999	110	1.00	11.26
GGVTSFLR	■	0.9999	106	1.00	11.26
FSDFIQK	■	0.9999	91	1.00	11.26
ATQYLDK	■	0.9999	86	1.00	11.26
LOGDQNR	■	0.9999	86	1.00	11.26
ALTYPLM ₁₄₇ PDPTFDASGR	■	0.9999	69	1.00	11.26
ITGYQLYR	■	0.9999	60	1.00	11.26
LTEVQPIK	■	0.9999	42	1.00	11.26
GVDLSAK	■	0.9998	19	1.00	11.26
IAEKLTVEGPIKIQK	■	0.9927	6	1.00	11.26
ALTYPLM ₁₄₇ PDPTFDASGRIAEK	■	0.3516	1	1.00	11.94
ALH23622.1 tail structural protein [Pseudomonas phage PaMx28]		1.0000	724	61.7%	24.03%
LVYIQAERPR	■	0.9996	61	1.00	4.10
LVDGETIK	■	0.9996	30	1.00	4.10
LVDGETIR	■	0.9996	20	1.00	4.10
GSQAPTVSPVAGIVTSGAAVFPVWGVVDVSEK	■	0.9996	14	1.00	4.10
RGRLVYIQAERPR	■	0.1076	11	1.00	4.10
LEVDZLAIEFGQAC ₁₆₀ ILR	■	0.9996	11	1.00	4.10
ALH23623.1 tail structural protein [Pseudomonas phage PaMx28]		1.0000	303	45.9%	10.31%
LVZIATPVDSGR	■	0.9996	34	1.00	4.30
GNWQTTINSPAAALDR	■	0.9996	30	1.00	4.30
ASVLELFK	■	0.9996	17	1.00	4.30
RASVLELFK	■	0.9999	13	1.00	4.30
IEYDQWR	■	0.9999	9	1.00	4.30
IKHVAQWR	■	0.1994	3	1.00	5.13
IEYDQWRQAPEGVIRK	■	0.1309	1	1.00	5.13
IEYDQWRQAPEGVIR	■	0.0733	1	1.00	5.13
ALH23624.1 tail terminator protein [Pseudomonas phage PaMx28]		1.0000	491	31.7%	59.40%
ADELSDFIK	■	0.9999	273	1.00	5.00
EVDGWYR	■	0.9999	106	1.00	5.00
LAHGGVELVTSC ₁₆₀ GR	■	0.9999	73	1.00	5.00
VSH ₁₄₇ TVTWEAR	■	0.9999	69	1.00	5.00
GREVDGWYR	■	0.9999	47	1.00	5.00
RLAHGGVELVTSC ₁₆₀ GR	■	0.9999	23	1.00	5.00
ALH23625.1 major tail structural protein [Pseudomonas phage PaMx28]		1.0000	163	32.0%	6.30%
SNVSGQVYAFENSTLLDK	■	0.9990	66	1.00	3.12
IVAESAYGTFANPVFDIIR	■	0.9990	60	1.00	3.12
FINETESSIVFELPDGAGNLYTVTLPFR	■	0.9990	34	1.00	3.12
LLEAVLGGTWEVDTFPAVGTDR	■	0.9990	18	1.00	3.12
ESLQSEELRDR	■	0.1236	3	1.00	4.10

Proteínas estructurales identificadas en el fago PaMx28. En cada bloque se muestran los péptidos identificados, la probabilidad de dicha identificación, número de espectros mapeados, cobertura, y fracción de espectros identificados por péptido con respecto al número de péptidos hermanos. Los ORFs identificados fueron los siguientes: ORF_19, ALH23619; ORF_21, ALH23621; ORF_22, ALH23622; ORF_23, ALH23623; ORF_24, ALH23624 & ORF_25, ALH23625.

Anexo 5: Identificación de las proteínas estructurales del fago PaMx28

Main Entry Accession	NTT	Probability	# Tot PSMs	% Coverage / Weight	% Spectrum ids / NSP
ALH23628.1 tail tape measure protein [Pseudomonas phage PaMx28]	1.0000		2044	65.6%	95.52%
ADELGFPANLAEM ₁₄₇ AR	0.9863		140	1.00	47.36
LTGPLTAATSATAALSK	0.9863		124	1.00	47.36
IQTQLGAASOM ₁₄₇ FDGLAGLAK	0.9863		100	1.00	47.36
DLSDM ₁₄₇ VEAVADATTGSEFER	0.9863		107	1.00	47.36
AAVDFLIAAFK	0.9863		90	1.00	47.36
GAEELEGLR	0.9863		89	1.00	47.36
DISQTVDELTAFIGK	0.9863		83	1.00	47.36
AFSVAQAAH ₁₄₇ SISTGLAK	0.9863		75	1.00	47.36
ATSGLTGAFTR	0.9996		72	1.00	47.36
AGVAFEALR	0.9996		68	1.00	47.36
M ₁₄₇ SLLDAALAEK	0.9863		64	1.00	47.36
LWNLGLTPSER	0.9996		61	1.00	47.36
ALADLESOR	0.9863		58	1.00	47.36
QFDVFNAGLVATGSSSEK	0.9863		56	1.00	47.36
TEEESQESYAR	0.9996		54	1.00	47.36
EFAAQTPYSLDOAVGGFTK	0.9863		54	1.00	47.36
ALTSYQNTASAM ₁₄₇ SK	0.9863		43	1.00	47.36
OM ₁₄₇ SLESELVTETER	0.9863		51	1.00	47.36
ATFNDDTVAGVQOR	0.9996		50	1.00	47.36
SEEEAIOASYEK	0.9863		44	1.00	47.36
ALFATSK	0.9996		43	1.00	47.36
VESLQVAEADRR	0.9996		42	1.00	47.36
EALYNGLLTEEH ₁₄₇ LR	0.9863		42	1.00	47.36
LATDALOGLTOM ₁₄₇ ZIASGELEAYLK	0.9863		35	1.00	47.36
LGENDFAGAM ₁₄₇ QOR	0.9863		33	1.00	47.36
ASDDVAAAR	0.9996		29	1.00	47.36
AZIEANTGVESDOR	0.9863		24	1.00	47.36
LKEFGDK	0.9982		23	1.00	47.36
RLNQDFQOEOAM ₁₄₇ EQOR	0.9863		22	1.00	47.36
VSAYAGAFVDGK	0.9863		21	1.00	47.36
LEADRAEQLR	0.9992		21	1.00	47.36
EFOSVVESLR	0.9996		20	1.00	47.36
GLTSGGGAEK	0.9996		18	1.00	47.36
EREFOSVVESLR	0.9996		18	1.00	47.36
LEADRAEQLK	0.9980		16	1.00	47.36
DNFGWEDEGK	0.9996		16	1.00	47.36
AZIEANTGVESDORR	0.9994		16	1.00	47.36
LNQDFQOEOAM ₁₄₇ EQOR	0.9863		14	1.00	47.36
SYAGEQSTAYK	0.9996		11	1.00	47.36
SITVQFDWGR	0.9989		11	1.00	47.36
KAZIEANTGVESDORR	0.9996		11	1.00	47.36
VESLQVAEADR	0.9996		10	1.00	47.36
EYEEQOR	0.9155		10	1.00	47.36
KAZIEANTGVESDOR	0.9996		9	1.00	47.36
VSLTFQGVTTTIGHNAETEEYLTK	0.9996		6	1.00	47.36
ANTAEGSELR	0.9992		6	1.00	47.36
SSDKVDK	0.2304		3	1.00	47.93
SEEEAIOASYEKR	0.9707		3	1.00	47.24
NFPMNR	0.0771		2	1.00	65.15
ALH23629.1 vision structural protein [Pseudomonas phage PaMx28]	1.0000		800	53.1%	69.00%
GDVLTDPNVLVR	0.9999		101	1.00	14.88
VQAVKPVSVGR	0.9998		72	1.00	14.88
AKGAVQLSAAPFQHR	0.9999		67	1.00	14.88
YTOGGAATISVPEVPGI	0.9998		63	1.00	14.88
M ₁₄₇ RQVWPVZPKK	0.9999		50	1.00	14.88
TTLQWHSR	0.9998		49	1.00	14.88
HAFWYLSHNPDLVAR	0.9999		43	1.00	14.88
VVEVLNR	0.9998		38	1.00	14.88
FDADRVEIK	0.9998		31	1.00	14.88
YGAPETQTH ₁₄₇ SHHPLTR	0.9999		29	1.00	14.88
ELDETDM ₁₄₇ TGIIVGVER	0.9998		20	1.00	14.88
WAQAPFVVSNDVRR	0.9999		20	1.00	14.88
GLTESLEWLTQVIR	0.9998		13	1.00	14.88
GLTESLEWLTQVIR _{C100} QAGEK	0.9645		10	1.00	15.09
VKAVGAVQLSAAPFQHR	0.9999		6	1.00	14.88
QTM ₁₄₇ RQVWPVZPKK _{T100} WLDPAEYGAAC	0.9999		6	1.00	14.88
WAQAPFVVSNDVRR	0.0520		1	1.00	15.88
M ₁₄₇ RQVWPVZPKKGLTESLEWLTQVIR	0.1033		1	1.00	15.77
EOHYRELDQNTGIIVGVER	0.0504		1	1.00	15.88

Proteínas estructurales identificadas en el fago PaMx28. En cada bloque se muestran los péptidos identificados, la probabilidad de dicha identificación, número de espectros mapeados, cobertura, y fracción de espectros identificados por péptido con respecto al número de péptidos hermanos. Los ORFs identificados fueron los siguientes: ORF_28, ALH23628 & ORF_29, ALH23629.

Anexo 5: Identificación de las proteínas estructurales del fago PaMx28

Main Entry Accession	NTT	Probability	# Tot PSMs	% Coverage / Weight	% Spectrum ids / NSP
ALH23630.1 <i>virion structural protein [Pseudomonas phage PaMx28]</i>		1.0000	636	68.1%	72.25%
FVEH ₁₄ GFVDPFAR	1.0000	121	1.00	19.97	
NHFLVPSGWR	1.0000	97	1.00	19.97	
NPFAGSSIV	1.0000	89	1.00	19.97	
LTGVRPSGVDIK	1.0000	88	1.00	19.97	
NNNLYFVLDR	1.0000	73	1.00	19.97	
SGSH ₁₄ AEELVVGTR	1.0000	71	1.00	19.97	
GIDATVVEELK	1.0000	62	1.00	19.97	
DGIQSISSSTASGGLYNALLTPDYEDIK	1.0000	61	1.00	19.97	
QSLIDVLDLR	1.0000	36	1.00	19.97	
FMLEENIGGFDKIPTR	1.0000	51	1.00	19.97	
ANVAALTPWEGTAVNYPM ₁₄ GFAR	1.0000	47	1.00	19.97	
AC ₁₀₀ FFTDGVPEPSSADAASLAADM ₁₄ ZEAR	1.0000	43	1.00	19.97	
RGIDATVVEELK	1.0000	19	1.00	19.97	
AVDNDVSYGVAVLFDFTVOLAK	1.0000	15	1.00	19.97	
LYPGC ₁₀₀ DR	0.9999	13	1.00	19.97	
LFTQOOGYTSYGH ₁₄ ZVYGLLVAK	1.0000	12	1.00	19.97	
LWVYTNAPYEVTYAGEVYAPAAVGHSEVSEIK	1.0000	10	1.00	19.97	
ANLQITTFDVIINDAR	1.0000	7	1.00	19.97	
SGSNALVVGTRTM ₁₄ DIK	0.1173	2	1.00	20.86	
HSAAELTVRFVENGVDVFPFNNLYFVLDR	0.9975	2	1.00	20.01	
YORM ₁₄ C ₁₀₀ FHALYGR	0.1119	1	1.00	20.86	
NNNLYFVLDRSGSH ₁₄ AEELVVGTR	0.1962	1	1.00	20.77	
M ₁₄ C ₁₀₀ FHALYGR	0.9693	1	1.00	20.20	
FVIGHAGALTLRPNMESLR	0.9967	1	1.00	20.03	
ALH23631.1 <i>virion structural protein [Pseudomonas phage PaMx28]</i>		1.0000	770	29.9%	32.40%
EIPVVFGR	0.9996	134	1.00	100	
DTGPNVAVGDLR	0.9996	36	1.00	100	
ALH23633.1 <i>virion structural protein [Pseudomonas phage PaMx28]</i>		1.0000	764	40.4%	80.44%
TIQRI ₁₄ ZIQLLK	0.9997	88	1.00	28.40	
VAEVAIGDGR	0.9997	85	1.00	28.40	
DVSGDAIK	0.9997	81	1.00	28.40	
AKFPYPPGDFR	0.9997	78	1.00	28.40	
VTNPEVALVETVFR	0.9997	74	1.00	28.40	
ASIPTVESFK	0.9997	74	1.00	28.40	
LGALVPAFR	0.9997	71	1.00	28.40	
KPDLFGGESR	0.9997	59	1.00	28.40	
QVAFAPYYELVGR	0.9997	54	1.00	28.40	
INGLYFPEGELR	0.9997	52	1.00	28.40	
LLSSGLK	0.9997	51	1.00	28.40	
LELWVSR	0.9997	50	1.00	28.40	
QVYIGLNPYLK	0.9997	42	1.00	28.40	
DAWVGNR	0.9997	37	1.00	28.40	
SWFSALASGEER	0.9997	34	1.00	28.40	
GEVLTWARR	0.9997	33	1.00	28.40	
QGISPQWPTK	0.9997	31	1.00	28.40	
HADAALYDR	0.9997	27	1.00	28.40	
DDVDVSTLLHLEANTIK	0.9997	24	1.00	28.40	
EC ₁₀₀ LQPTM ₁₄ GYAEALDDASFR	0.9997	22	1.00	28.40	
EGGIQSDIDM ₁₄ PGPEVQWVNDYLSK	0.9997	17	1.00	28.40	
AVIZADR	0.9997	17	1.00	28.40	
HADAALYDRK	0.9997	13	1.00	28.40	
LGIC ₁₀₀ LWDR	0.9994	8	1.00	28.40	
QGSTVDTLQAGNPDAGYVGAAVTPGGSAINSR	0.9997	7	1.00	28.40	
KVTGVYK	0.9992	6	1.00	28.40	
YYLGM ₁₄ M ₁₄ VLC ₁₀₀ HGPVQK	0.9997	3	1.00	28.40	
LQOTGGSLGFLDASISPEAGTYYTFR	0.9905	2	1.00	28.52	
LTRDDVDVSTLLHLEANTIK	0.7901	2	1.00	28.94	
IAPVSGAGEVSLDQASALLVQVSGR	0.9997	2	1.00	28.40	
ALH23634.1 <i>virion structural protein [Pseudomonas phage PaMx28]</i>		0.8678	1	8.7%	0.14%
NASTLAANLFC ₁₀₀ SFLYGADGYTDVQK	0.8678	1	1.00	0.01	
ALH23635.1 <i>virion structural protein [Pseudomonas phage PaMx28]</i>		1.0000	712	45.9%	238.01%
AAGWSAGTAT	0.9996	82	1.00	4.40	
VAVWSQADNAWAFVPR	0.9996	22	1.00	4.40	
LGWVAVVEDEEK	0.9996	20	1.00	4.40	
FGVHLVSK	0.9996	17	1.00	4.40	
LGWVAVVEDEEKLTAIK	0.9996	3	1.00	4.40	
LGRFGVHLVSK	0.4039	1	1.00	4.96	
ALH23638.1 <i>virion structural protein [Pseudomonas phage PaMx28]</i>		0.5041	1	37.3%	0.21%
STEGAAAERGGATAPRATPKR	0.5041	1	1.00	0.01	
EDVAREK	0.1263	1	1.00	0.71	
AELSAQVATLRDVAR	0.0865	1	1.00	0.81	
ADALSSAVDRNDQARQEAR	0.0469	1	1.00	0.51	

Proteínas estructurales identificadas en el fago PaMx28. En cada bloque se muestran los péptidos identificados, la probabilidad de dicha identificación, número de espectros mapeados, cobertura, y fracción de espectros identificados por péptido con respecto al número de péptidos hermanos. Los ORFs identificados fueron los siguientes: ORF_30, ALH23630; ORF_31, ALH23631; ORF_33, ALH23633; ORF_34, ALH23634; ORF_35, ALH23635 & ORF_38, ALH23638.

Anexo 5: Identificación de las proteínas estructurales del fago PaMx42

Main Entry Accession	NTT	Probability	# Tot PSMs	% Coverage / Weight	% Spectrum ids / NSP
ALH23593.1 portal protein [Pseudomonas phage PaMx42]	1.0000		1630	47.2%	78.16%
NVSAAFEMALK	0.9999		116	1.00	19.93
LLPH ₁₄ 7YLIR	0.9999		115	1.00	19.93
AGILVDYPTTEGGNGASVADLEAGK	0.9999		115	1.00	19.93
VPALLNP.LV.NATGGGINLTLSK	0.9999		110	1.00	19.93
TATEELLSAASGGSTLSSATK	0.9999		99	1.00	19.93
LDEEGYVMEIWR	0.9999		74	1.00	19.93
YLPH ₁₄ FNASDQSK	0.9999		61	1.00	19.93
LLQAEPTNH ₁₄ 1K	0.9999		61	1.00	19.93
GGIP.LPTGADAK	0.9998		61	1.00	19.93
AVSLNLAYSR	0.9999		46	1.00	19.93
DAIAGEPTVK	0.9998		35	1.00	19.93
AVFYNVAR	0.9999		35	1.00	19.93
GALFTTEH ₁₄ R	0.9999		34	1.00	19.93
YDAYLAR	0.9998		33	1.00	19.93
QIIEEWOK	0.9998		19	1.00	19.93
WIGAGDAGVK	0.9999		15	1.00	19.93
KAVSLNAYSR	0.9999		13	1.00	19.93
FELNTDFDIAR	0.9999		12	1.00	19.93
IKPTLVVYAPTEIINWR	0.9999		10	1.00	19.93
GTVNFGRS	0.9995		4	1.00	19.94
RIIIEWQK	0.9980		3	1.00	19.96
ALH23592.1 morpho genesis protein [Pseudomonas phage PaMx42]	1.0000		303	21.2%	20.69%
LAVYVEGK	0.9990		36	1.00	9.54
EIIEEGAR	0.9995		31	1.00	9.54
LYDISTR	0.9996		28	1.00	9.54
WVQDGLR	0.9995		21	1.00	9.54
DISEVLK	0.9996		19	1.00	9.54
AFH ₁₄ 7AADLEVNR	0.9989		16	1.00	9.54
YEADPLTLDLFR	0.9996		14	1.00	9.54
IYSAYTSQLEOLK	0.9990		14	1.00	9.54
YEADPLTLDLFR	0.9982		11	1.00	9.54
DISEVLK	0.9895		10	1.00	9.50
QPLEVDDILGTGGELAR	0.9990		9	1.00	9.54
ALH23590.1 major head protein [Pseudomonas phage PaMx42]	1.0000		1422	97.7%	100.00%
QQIDLNAATGGAILLOGAAHGDFSDVAFFAK	0.9990		462	1.00	13.96
VYQADNSYNGVIR	0.9990		402	1.00	13.96
NAYSGVAADK	0.9985		148	1.00	13.96
LFTYGTNVIR	0.9982		145	1.00	13.96
LYGNLNNSER	0.9990		144	1.00	13.96
SPTDAALFTSTNWR	0.9990		132	1.00	13.96
LDPGFR	0.9967		74	1.00	13.96
DLAGVIVEH	0.9990		66	1.00	13.96
YDATGNTPDQPTWNNLNNGQAK	0.9990		61	1.00	13.96
HLVDTSVK	0.9990		48	1.00	13.96
YATSEKDLAVIVEH	0.9990		47	1.00	13.96
TGDELIR	0.9983		42	1.00	13.96
VAAGTPPIR	0.9979		35	1.00	13.96
FNAYSGVAADK	0.9990		33	1.00	13.96
FGDSSQZAAHIL ₁₄ 1HSTPH ₁₄ 4HK	0.9990		20	1.00	13.96
ALH23588.1 virion structural protein [Pseudomonas phage PaMx42]	1.0000		630	91.9%	100.00%
LNAAADKPYEVTVH ₁₄ 1NGAANTK	0.9990		252	1.00	5.99
ITGVIGSGTVANLVNPTTK	0.9990		224	1.00	5.99
LAPC ₁₄ DFVAGTVPTZINAK	0.9990		156	1.00	5.99
TVLDPDAPVUNLPNTQAVVTNGQK	0.9990		111	1.00	5.99
VVFFTAGTPTSGELAAZAK	0.9990		82	1.00	5.99
AVVTLSAS	0.9990		37	1.00	5.99
KVIFFTAGTPTSGELAAZAK	0.9990		32	1.00	5.99
ALH23587.1 virion structural protein [Pseudomonas phage PaMx42]	1.0000		465	35.0%	46.81%
TSPTQALAWPR	0.9998		240	1.00	4.13
TGVFLNC ₁₄ DEVPSNVZPK	0.9998		92	1.00	4.13
ATDYLEAQC ₁₄ R	0.9998		81	1.00	4.13
GTELPDQDD.EVAAH ₁₄ 1LIR	0.9998		61	1.00	4.13
RTSPTQALAWPR	0.9998		11	1.00	4.13
YQGRKTSPTQALAWPR	0.1373		1	1.00	4.99
ALH23586.1 virion structural protein [Pseudomonas phage PaMx42]	1.0000		303	80.2%	41.89%
QTH ₁₄ 1ETFTIK	0.9998		119	1.00	6.89
IANDIVLSPNGQK	0.9998		93	1.00	6.89
VLTYVFGA	0.9998		49	1.00	6.89
GSNH ₁₄ 1GLH ₁₄ 1GRVPPFNLIK	0.9998		38	1.00	6.89
DVTIC ₁₄ FLVDR	0.9997		8	1.00	6.89
VZDDAAFPDF.SUPHNPGPATPEDK	0.9998		6	1.00	6.89
DVZRDGVELR	0.9998		6	1.00	6.89
QTMETFTZKGTVEPK	0.9769		1	1.00	6.89

Proteínas estructurales identificadas en el fago PaMx42. En cada bloque se muestran los péptidos identificados, la probabilidad de dicha identificación, número de espectros mapeados, cobertura, y fracción de espectros identificados por péptido con respecto al número de péptidos hermanos. Los ORFs identificados fueron los siguientes: ORF_08, ALH23593; ORF_09, ALH23592; ORF_11, ALH23590; ORF_13, ALH23588; ORF_14, ALH23587 & ORF_15, ALH23586.

Anexo 5: Identificación de las proteínas estructurales del fago PaMx42

Main Entry Accession	NTT	Probability	# Tot PSMs	% Coverage / Weight	% Spectrum ids / NSP
ALH23585.1 tail terminator protein [Pseudomonas phage PaMx42]	1.0000		403	66.2%	35.8%
INELDPPEELYER	0.9998		145	1.00	6.75
SNTQAFALGGTLAK	0.9998		113	1.00	6.75
M142EFDDAK	0.9998		43	1.00	6.75
WGGVYR	0.9997		38	1.00	6.75
FNVTVEFYDELG	0.9998		35	1.00	6.75
YTASGLVYVYVFC160PK	0.9998		18	1.00	6.75
QTVFEQITLSTC160EGKPGQK	0.9998		9	1.00	6.75
NAFRGICTTPGK	0.9416		11	1.00	6.75
Main Entry Accession	NTT	Probability	# Tot PSMs	% Coverage / Weight	% Spectrum ids / NSP
ALH23584.1 major tail structural protein [Pseudomonas phage PaMx42]	1.0000		2441	69.4%	100.00%
TVASSTATTVVNETLIDEAAPPAGVK	0.9998		842	1.00	19.59
NAGTFDZPLLGLGGGR	0.9998		323	1.00	19.59
ALVFDOTFFAAATETGTGK	0.9998		258	1.00	19.59
LLOGFFADAR	0.9991		227	1.00	19.59
IAAPGEDAVNTSSDIYR	0.9998		216	1.00	19.59
FANNVGYR	0.9987		190	1.00	19.00
LNADLTFAC160DNTHR	0.9992		159	1.00	19.59
ELASTAPINGAAVAISGVAASTK	0.9998		130	1.00	19.59
ANADVGLSVISAAK	0.9991		90	1.00	19.59
TYTVASGGAFAAM144LVNATGFANANNGLK	0.9998		86	1.00	19.59
KGITIDLDASGGFNSDFTK	0.9998		80	1.00	19.59
WVFIGDAASNR	0.9990		69	1.00	19.59
GTITDLDASGGFNSDFTK	0.9998		65	1.00	19.59
QLPPTPNVYGLSPNSYSDFGGELSTVAR	0.9998		62	1.00	19.59
IDNITGLLAYAEEEC160LAK	0.9998		61	1.00	19.59
QLSGGLNGVDAEYLEGAVANFTLNIPQADK	0.9989		35	1.00	19.00
APDPSR	0.9844		25	1.00	19.00
SGDAGDTIK	0.9827		21	1.00	19.00
LFVGVVZIK	0.9861		15	1.00	19.00
HNVLDPASSNPALFGYVSEANVSINNVSPNK	0.9182		1	1.00	19.96
LFVGVVZIKENKIPALTIK	0.9997		1	1.00	19.59
Main Entry Accession	NTT	Probability	# Tot PSMs	% Coverage / Weight	% Spectrum ids / NSP
ALH23581.1 viroion structural protein [Pseudomonas phage PaMx42]	1.0000		33	63.1%	3.27%
LNDGYSAGAPAGFVER	0.9978		16	1.00	2.38
TPVDTSQALSNWR	0.9978		12	1.00	2.38
HANDLLSLAVSLERK	0.0406 [D]		3	1.00	3.00
ASAAETINAAKSVLQMK	0.2217		3	1.00	3.00
RLNDDYSAGAPAGFVER	0.9978		1	1.00	2.38
Main Entry Accession	NTT	Probability	# Tot PSMs	% Coverage / Weight	% Spectrum ids / NSP
ALH23580.1 tail tape measure protein [Pseudomonas phage PaMx42]	1.0000		2200	69.7%	96.12%
TAVDTANAAAANDR	0.9998		215	1.00	45.44
LATAQDQTTAATAANAAAQTR	0.9998		190	1.00	45.44
LWAEISLEDGFK	0.9998		111	1.00	45.44
NSAIEFFGELNKK	0.9998		110	1.00	45.44
SAYELVNFER	0.9998		87	1.00	45.44
TTPTZGSLTVLKK	0.9998		86	1.00	45.44
DLDQDLQLLQLLPLK	0.9998		76	1.00	45.44
LLDLQAISSQEQYSR	0.9998		74	1.00	45.44
VYRLICEISANK	0.9998		69	1.00	45.44
ANKATNQLSAAQSK	0.9998		69	1.00	45.44
SGGAM1427GGVVEGLNLR	0.9998		67	1.00	45.44
FDZYNEAZGPQR	0.9998		62	1.00	45.44
AZLLWADN144PK	0.9998		62	1.00	45.44
TGHAALENTAK	0.9998		60	1.00	45.44
ZQAVQDQIVQOQK	0.9998		60	1.00	45.44
GAGANQLGGDQDAK	0.9998		60	1.00	45.44
LONVTESSQVVTLTIK	0.9998		59	1.00	45.44
AATAQVQGTAAQNLAAATTR	0.9998		56	1.00	45.44
M1449GAGQTEFESWIK	0.9998		55	1.00	45.44
ITSEVLNFAQLZQK	0.9998		50	1.00	45.44
ELFDLANR	0.9998		49	1.00	45.44
INTQLDNE144NR	0.9998		45	1.00	45.44
LTLQDANALSDQEAALM144DASVTK	0.9998		44	1.00	45.44
M144FTLQPOR	0.9998		43	1.00	45.44
AGVEETATAFTR	0.9998		42	1.00	45.44
AVSENM1447PH144VLDVAVK	0.9998		39	1.00	45.44
M144DQIEEELIQR	0.9998		36	1.00	45.44
LATAQDQTAATAAQAQAQANLAAQNTSATASQR	0.9998		32	1.00	45.44
GITLMAEELTQLR	0.9998		29	1.00	45.44
EEOQVM1447QVNDLLTIK	0.9998		28	1.00	45.44
AGQASQTAGQDAGYVK	0.9998		28	1.00	45.44
DYNATLEASK	0.9998		27	1.00	45.44
ALIVSGATQEAESALLQLSQGFNAQK	0.9998		27	1.00	45.44
ALNVPIZR	0.9998		23	1.00	45.44
TOVAM1447ANVEAALQR	0.9998		20	1.00	45.44
VQDKVDSSTSIK	0.9998		16	1.00	45.44
IKLDDQNM1447LNAASGFFGQ1447QLQIK	0.9945		15	1.00	45.40
ATGVTAAASK	0.9998		15	1.00	45.44
SSASSGLSR	0.9998		13	1.00	45.44
LOGDEFR	0.9998		13	1.00	45.44
ROQYIDQLK	0.9938		9	1.00	45.47
QDLDLSEQTAATH1447R	0.9996		8	1.00	45.44
LTETZINK	0.9950		6	1.00	45.40
QOYIDQLK	0.0763		4	1.00	46.44
DYNATLEASKK	0.9897		3	1.00	45.40
SOEDSLR	0.9878		1	1.00	45.50
LOSELAR	0.9687		1	1.00	45.50
KLLDLGASQEQYSR	0.9600		1	1.00	45.50

Proteínas estructurales identificadas en el fago PaMx42. En cada bloque se muestran los péptidos identificados, la probabilidad de dicha identificación, número de espectros mapeados, cobertura, y fracción de espectros identificados por péptido con respecto al número de péptidos hermanos. Los ORFs identificados fueron los siguientes: ORF_16, ALH23585; ORF_17, ALH23584; ORF_20, ALH23581 & ORF_21, ALH23580.

Anexo 5: Identificación de las proteínas estructurales del fago PaMx42

Main Entry Accession	NTT	Probability	# Tot PSMs	% Coverage / Weight	% Spectrum ids / NSP
ALH23578.1 <i>virion structural protein [Pseudomonas phage PaMx42]</i>		1.0000	816	68.0% / 68.79%	
NGSSSSVHSGTYGYADK	0.9998	110	1.00	5.59	
IDNLKPGSLVTVYGPDQWK	0.9998	115	1.00	5.59	
QSSGLTIVVPPNWR	0.9998	99	1.00	5.59	
VAQSSGKQVDFM ₁₄ HSGEM ₁₄ DSSGTR	0.9998	85	1.00	5.59	
SLISGEIGTGGHFTASVDSR	0.9998	62	1.00	5.59	
TAIENAC ₁₀ VANGITLNGILSK	0.9998	62	1.00	5.59	
GGNLGTHC ₁₀₀ AGIDNALPQNSATVLPVK	0.9998	31	1.00	5.59	
Main Entry Accession	NTT	Probability	# Tot PSMs	% Coverage / Weight	% Spectrum ids / NSP
ALH23577.1 <i>virion structural protein [Pseudomonas phage PaMx42]</i>		1.0000	476	34.5% / 36.32%	
VFGPIAVWR	1.0000	71	1.00	15.24	
YSVNEVAETTLVC ₁₆₀ K	1.0000	48	1.00	15.23	
ISFLP ₁₄ R	1.0000	48	1.00	15.23	
LVVQPSDGAFTTIEGVEYAEK	1.0000	45	1.00	15.23	
IHVSPQIEDLQTIASVQTR	1.0000	34	1.00	15.23	
LFPIVTHR	1.0000	33	1.00	15.24	
TLFAH ₁₄ R	1.0000	31	1.00	15.23	
VIFDOSTK	1.0000	28	1.00	15.24	
LWIPFQYATR	1.0000	26	1.00	15.23	
VWVPTYR	1.0000	24	1.00	15.24	
LGDDKVTWK	0.9997	23	1.00	15.24	
IQRPEE ₁₄ DYWK	1.0000	22	1.00	15.23	
YSEVEYLPVGDVZGR	1.0000	19	1.00	15.23	
FYSDL ₁₄ AAVGR	0.9998	9	1.00	15.24	
SQLTFRPSTVTVLPTVYLGVPVFDK	0.9998	8	1.00	15.28	
KVWPTFR	0.3170	8	1.00	15.62	
YVEEETASGDR	1.0000	4	1.00	15.23	
Main Entry Accession	NTT	Probability	# Tot PSMs	% Coverage / Weight	% Spectrum ids / NSP
ALH23576.1 <i>virion structural protein [Pseudomonas phage PaMx42]</i>		1.0000	100	34.4% / 7.98%	
DYGFOTPPALDLTYR	0.9956	36	1.00	3.51	
GPVASIINGEDATLR	0.9956	30	1.00	3.51	
DAATFVPIWK	0.9955	27	1.00	3.51	
FGNILOGKIPNVYVQPC ₁₆ QNVLFDER	0.9956	7	1.00	3.51	
FNNIINFGG ₁₆₀ PVYVPGESNNPFTNGIS	0.1925	5	1.00	4.00	
Main Entry Accession	NTT	Probability	# Tot PSMs	% Coverage / Weight	% Spectrum ids / NSP
ALH23575.1 <i>virion structural protein [Pseudomonas phage PaMx42]</i>		1.0000	481	36.9% / 14.68%	
TPAPATGATSTAAWAR	0.9997	70	1.00	20.54	
IDLGLTSGEVK	0.9997	48	1.00	20.54	
EAQALSVFNDADAGLETGTQTR	0.9997	38	1.00	20.54	
VTPNEALSTFDSSGR	0.9997	37	1.00	20.54	
STEIS ₁₆₀ QTTAEVASK	0.9997	36	1.00	20.54	
LGDFQVPR	0.9997	35	1.00	20.54	
LPYQPNLTLDSGR	0.9997	32	1.00	20.54	
GLVDTVPASHAGDR	0.9997	32	1.00	20.54	
TTWSTFM ₁₆₀ CR	0.9997	23	1.00	20.54	
QC ₁₆₀ DGLLYODPATSK	0.9997	21	1.00	20.54	
LLDTTTSKLP ₁₅₅ SAPTVAQAQNGR	0.9997	18	1.00	20.54	
QLSLVPLVK	0.9997	17	1.00	20.54	
GGLGQTFYDGR	0.9998	14	1.00	20.54	
KVSSGLFSK	0.9998	13	1.00	20.54	
YGDPI ₁₄ PLVWGTVR	0.9998	12	1.00	20.54	
DLLEEI ₁₄ R	0.9998	10	1.00	20.54	
QDYDINTLPVLDESIDK	0.9998	8	1.00	20.54	
YGDPIPLVWGTVRQK	0.9489	6	1.00	20.15	
SREAOALSVFNDADAGLETGTQTR	0.9155	4	1.00	20.05	
LTPQDYDINTLPVLDESIDK	0.3690	3	1.00	21.17	
ASTLFPQGVFLN ₁₆₀ QFN ₁₄ W ₁₄ R ₁₄ VSK	0.3678	2	1.00	30.08	
R ₁₅ REAOALSVFNDADAGLETGTQTR	0.3669	1	1.00	30.08	
NFSKTTWSTFM ₁₆₀ AVTFK	0.4484	1	1.00	21.09	
DLLEEI ₁₆₀ DGLLYODPATSK	0.8461	1	1.00	30.99	

Proteínas estructurales identificadas en el fago PaMx42. En cada bloque se muestran los péptidos identificados, la probabilidad de dicha identificación, número de espectros mapeados, cobertura, y fracción de espectros identificados por péptido con respecto al número de péptidos hermanos. Los ORFs identificados fueron los siguientes: ORF_23, ALH23578; ORF_24, ALH23577; ORF_25, ALH23576 & ORF_26, ALH23575.

12. Firmas del jurado

Dr. Gabriel Guarneros
Peña

Dr. Luis Kameyama
Kawabe

Dra. Rosa María
Bermúdez Cruz

Dr. Pablo Vinuesa
Fleischman

Dr. Javier Hernández
Sánchez