



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**

---

---

**POSGRADO EN CIENCIAS BIOLÓGICAS**

**FACULTAD DE CIENCIAS**

**Evolución Molecular de la Interacción entre los  
Genes de Profagos y sus Hospederos  
Procariontes: El Caso de Sulfolobus spp.**

**T E S I S**  
**QUE PARA OBTENER EL GRADO ACADÉMICO**  
**DE MAESTRA EN CIENCIAS BIOLÓGICAS**  
**(ORIENTACIÓN EXPERIMENTAL)**  
**P R E S E N T A**  
**Y E T Z I R O B L E S B U C I O**

**DIRECTOR DE TESIS:**  
**DR. ANTONIO E. LAZCANO ARAUJO REYES**

**MÉXICO, D.F.**

**MAYO, 2008**



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

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

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

## *AGRADECIMIENTOS*

*Al Posgrado en Ciencias Biológicas*

*Al Dr. Antonio E. Lazcano Araujo Reyes por su asesoría y dirección en esta tesis.*

*Al comité tutorial formado por la Dra. Carmen Gómez Eichelmann y el Dr. Pedro Miramontes Vidal .*

*Al apoyo proporcionado por el Consejo Nacional de Ciencia y Tecnología a través de la beca con no. de registro 187234.*

*Este trabajo fue elaborado en el laboratorio de Microbiología de la Facultad de Ciencias en la Universidad Nacional Autónoma de México.*

## *AGRADECIMIENTOS*

- *A la Universidad Nacional Autónoma de México*
- *Al jurado que evaluó este trabajo: Dr. Víctor Valdes, Dra. Carmen Gómez, Dr. Arturo Becerra, Dr. Pedro Miramontes y Dr. Antonio Lazcano.*
- *Al Dr. Lorenzo Segovia por la lectura y comentarios al trabajo.*
- *Al Dr. Arturo Becerra y Dr. Luis Delaye por su apoyo técnico-académico para la realización de esta tesis.*
- *A la Mtra en C. Ana María Velasco Velasco por su guía académica y comentarios acerca del proyecto.*
- *A los miembros del laboratorio Sara Islas, Ricardo, Mario Rivas, Jesús y Ervin por su amistad!*
- *A Elisa Solís, Daniela Sosa, Mónica Naves Fernández y Gabriela Naves Fernández por los buenos ratos que hemos pasado.*
- *A las amistades que han perdurado Citlali E., Jessica A. Guzmán y Miriam B.*

*Al amor incondicional y eterno apoyo de papá y mamá*

*A mi hermano Héctor y anexos perrunos*

*A mi hermana Rosa*

*A Dorjee Sempa*

*A los que partieron para volver a encontrarnos*

*A los maestros de todos los tiempos*

*OM*

# ÍNDICE

## RESUMEN

<b>1. INTRODUCCIÓN</b>	<b>1-7</b>
1.1 Biología de Bacteriófagos	<b>8-15</b>
1.2 Características Generales de los Bacteriófagos	<b>16-30</b>
<b>2. OBJETIVOS</b>	<b>31</b>
<b>3. HIPÓTESIS</b>	<b>32</b>
<b>4. METODOLOGÍA</b>	<b>33</b>
4.1 Base de Datos y Búsqueda en Estructura Primaria	<b>34-37</b>
<b>5. RESULTADOS</b>	<b>38</b>
5.1 Caracterización Funcional de las Proteínas Virales	<b>39-40</b>
5.2 Bacteriófagos y Hospederos	<b>41-43</b>
5.3 Alineamientos	<b>44-57</b>
<b>6. DISCUSIÓN</b>	<b>58-63</b>
<b>7. CONCLUSIONES</b>	<b>64</b>
<b>8. SUPLEMENTOS</b>	<b>65-71</b>
<b>9. REFERENCIAS</b>	<b>72-76</b>

# ÍNDICE DE FIGURAS

## INTRODUCCIÓN

**Figura 1.** Virus de enterobacteria P2 teñido con 5% UA. **17**

La barra representa 100nm. Micrografía realizada por Dr. Maria Schnos, Departamento de Virología Molecular, Universidad de Wisconsin, Madison. USA. Imagen obtenida de la base de datos ICTV.

**Figura 2.** Bacteriófago Cp-1. **18**

La barra representa 50nm y el virus fue tratado con PT. Micrografía realizada por Dr. Hans-W Ackermann, Departamento de Microbiología y Bioquímica, Laval University, Quebec, Canadá. Imagen obtenida de la base de datos ICTV

**Figura 3.** Virus del bacteriófago  $\lambda$  teñido con 5% UA. **20**

La barra representa 100nm. Micrografía realizada por Dr. Maria Schnos, Departamento de Virología Molecular, Universidad de Wisconsin, Madison. USA. Imagen obtenida de la base de datos ICTV.

**Figura 4.** Bacteriófago PM2. **21**

La barra representa 100nm y el virus fue tratado con 2% PT. Micrografía realizada por Dr. Hans-W Ackermann, Departamento de Microbiología y Bioquímica, Laval University, Quebec, Canadá. Imagen obtenida de la base de datos ICTV.

**Figura 5.** Bacteriófago  $\phi$ X174. **22**

La barra representa 100nm y el virus fue tratado con 2% PTA. Micrografía realizada por el Dr. Hans-W Ackermann, Departamento de Microbiología y Bioquímica, Laval University, Quebec, Canadá. Imagen obtenida de la base de datos ICTV.

**Figura 6.** Bacteriófago  $\phi$ 6. **23**

La barra representa 100nm y el virus fue tratado con 2% PTA. Micrografía realizada por el Dr. Hans-W Ackermann. Departamento de Microbiología y Bioquímica, Laval University, Quebec, Canadá. Imagen obtenida de la base de datos ICTV.

**Figura 7.** Bacteriófago MS2. **26**

La barra representa 100nm y el virus fue tratado con 2% PTA. Imagen obtenida de la base de datos ICTV. Micrografía realizada por el Dr. Hans-W Ackermann, Departamento de Microbiología y Bioquímica, Laval University, Quebec, Canadá.

**Figura 8.** Bacteriófago H75. **27**

La barra representa 100nm y el virus fue tratado con 2% PTA. Micrografía realizada por el Dr. Hans-W Ackermann, Departamento de Microbiología y Bioquímica, Laval University, Quebec, Canadá. Imagen obtenida de la base de datos ICTV.

**Figura 9.** Bacteriófago 37-61. **28**

La barra representa 100nm y el fago fue tratado con 2% PTA. Micrografía realizada por el Dr. Hans-W Ackermann, Departamento de Microbiología y Bioquímica, Laval University, Quebec, Canadá. Imagen obtenida de la base de datos ICTV.

**Figura 10.** Bacteriófago filamentoso SIFV. **30**

Partículas de *Sulfolobus islandicus* filamentous virus (SIFV) tratadas con uranilacetato al 2%. Imagen obtenida ICTV.

## **METODOLOGÍA**

**Diagrama de Flujo** **33**

**Figura 11.** Los 275 genomas utilizados en la base de datos de bacteriófagos y sus porcentajes. **35**

**Figura 12.** Genomas de eucariontes y procariontes utilizados para la segunda base de datos y sus porcentajes. **35**

## **RESULTADOS**

**Figura 13.** Relación de los 287 resultados para los bacteriófagos que infectan y no infectan arqueas. **44**

## **SUPLEMENTOS**

**Figura 15.** Los 339 genomas bacterianos clasificados de acuerdo a su taxonomía y sus porcentajes. **65**

**Alineamientos** **67-71**

## ÍNDICE DE TABLAS

### INTRODUCCIÓN

<b>Tabla 1.</b> Factores de virulencia codificados en bacteriófagos.	<b>7</b>
--	----------

### METODOLOGÍA

<b>Tabla 2.</b> Archivos creados por Ancestor 7.0/BLAST.	<b>36</b>
--	-----------

<b>Tabla 3.</b> Tabla correspondiente a un grupo de hospederos bacterianos.	<b>37</b>
---	-----------

### RESULTADOS

<b>Tabla 4.</b> No. de proteínas virales con al menos un 'hit' de acuerdo a la clasificación taxonómica de los bacteriófagos.	<b>38</b>
---	-----------

<b>Tabla 5.</b> Porcentajes de proteínas de acuerdo a su anotación.	<b>39</b>
---	-----------

<b>Tabla 6.</b> Proteínas cuyas anotaciones forman el grupo de las proteínas hipotéticas y su porcentaje.	<b>40</b>
---	-----------

<b>Tabla 7.</b> Hospederos y bacteriófagos encontrados en las bases de datos empleadas en este estudio.	<b>42</b>
---	-----------

<b>Tabla 8.</b> Resumen de las proteínas virales que produjeron alineamientos significativos con proteínas hipotéticas de arqueas.	<b>46</b>
--	-----------

<b>Tabla 9.</b> Resumen de las proteínas hipotéticas virales que produjeron alineamientos significativos con proteínas de eucariontes, bacterias y arqueas.	<b>49</b>
---	-----------

<b>Tabla 10.</b> Resumen de los dominios funcionales identificados para las proteínas hipotéticas virales.	<b>54</b>
--	-----------

### DISCUSIÓN

<b>Tabla 11.</b> Virus que infectan Crenarchaeota utilizados en este estudio.	<b>58</b>
---	-----------

**Tabla 12.** Relación entre las proteínas hipotéticas virales y la presencia de sus homólogos en los tres dominios. **61**

## **SUPLEMENTOS**

**Tabla 1S.** Clasificación taxonómica y número de genomas de arqueas utilizadas para este estudio. **66**

**Tabla 2S.** Clasificación taxonómica y número de genomas de eucariontes utilizados en este estudio. **66**



## 1. INTRODUCCIÓN

Los bacteriófagos pasaron desapercibidos casi 40 años después del inicio de la bacteriología en los 1880s. Los primeros experimentos en este campo estaban enfocados en observaciones de cultivos bacterianos, y eran estudiados como un organismo completo más que como una población dinámica de células individuales. Sin embargo, hubo reportes que sugerían la presencia de agentes que las podían destruir en medios de cultivo. En 1896 Hankin reportó que las aguas de los ríos Jumna y Ganges en India podían matar muchos tipos de bacterias, especialmente *Vibrio cholerae*, y pudo demostrar que la actividad aséptica podía ser filtrada y sensible al calor, con lo que concluyó que la propiedad bactericida era un agente químico (1, 2). En 1901 Emmerich y Löw describieron una sustancia de cultivos autolisados que provocaba lisis en varios cultivos de bacterias, podía curar infecciones experimentales y proporcionar inmunidad a infecciones subsecuentes (3). Los primeros experimentos realizados con bacteriófagos utilizaron cultivos bacterianos en medio sólido y se basaron en la observación de colonias lisadas (placas).

En 1915 Frederick W. Twort publicó un estudio donde reportó un fenómeno que él describió como 'glassy transformation', lo que marca el inicio de la investigación en bacteriófagos como fue reconocido ocho años después. Independientemente al trabajo de Twort, Félix d'Herelle en 1917 describió un "microbio" que era antagonista a bacterias y las lisaba en cultivos líquidos y sólidos (4). Él también observó parches cristalinos de bacterias infectadas en agar sólido, y a estas estructuras las denominó placas.



D'Herelle llamó a estos microbios bacteriófagos y observó que estas partículas se multiplicaban indefinidamente, que necesitaban células vivas para hacerlo, y que la lisis celular era un proceso requerido para producir virus. Él imaginó a un bacteriófago como un microbio organizado, y se percató que las placas eran una manera de contabilizarlo. Así mismo, con el uso de cultivos líquidos d'Herelle demostró que los fagos se multiplican por pasos, y los identificó como ciclos de infección, multiplicación, liberación de partículas y reinfección.

La investigación de d'Herelle tuvo dos vertientes importantes: una se enfocó al uso terapéutico de bacteriófagos en enfermedades infecciosas, y la segunda a la biología de estas partículas. En sus primeros estudios con muestras de pacientes disentéricos d'Herelle observó que los títulos de fagos aumentaban al igual que la recuperación de los pacientes, y pensó que estas partículas podían representar agentes de resistencia a enfermedades infecciosas. A partir de ello d'Herelle abogó para que los bacteriófagos fueran utilizados como agentes terapéuticos en la era previa a los antibióticos. La idea de que los bacteriófagos eran virus que infectan bacterias tardó en ser aceptada. Hasta principios de la década de los cuarenta, los grupos líderes en microbiología pensaban que el fenómeno de lisis en bacterias asociadas a fagos resultaba de una activación autocatalítica de enzimas endógenas. Para responder estas críticas, d'Herelle estudió la biología de los bacteriófagos como parte de su segunda línea de investigación mencionada antes.

Los estudios de d'Herelle sugirieron que los bacteriófagos son parásitos obligados intracelulares y que el rango de especificidad en hospederos y sus propiedades antigénicas eran características de “razas” de bacteriófagos. La visión de d'Herelle para utilizar bacteriófagos como agentes de inmunidad no era bien recibida por los bacteriólogos de la época como Jules Bordet, quien recibió el premio Nobel en 1919 al describir la bacteriolisis mediada por complemento.



Bordet comenzó a trabajar con fagos para estudiar su papel en el fenómeno que había descrito. Aisló bacterias de animales infectados y estas tenían la propiedad de producir bacteriófagos cuando eran cultivados en medio líquido o con extractos obtenidos del peritoneo de estos animales. Bordet denominó a estas bacterias lisogénicas, y pensó que los bacteriófagos eran producidos por las bacterias debido a una perturbación en el estado metabólico bacteriano. Bordet no apoyaba la idea de los bacteriófagos como microbios, sino como enzimas líticas. Las diferencias científicas entre Bordet y d'Herelle generó lo que se denominó la controversia Twort-d'Herelle, que duró 10 años y terminó en 1932 cuando se pudo comprobar que los fenómenos y los extractos estudiados por Twort y d'Herelle eran idénticos.

En los 1920s y 1930s d'Herelle trabajó con la epidemia del cólera en India. A partir de esta época la terapia con fagos fue sometida a gran escrutinio por parte de la comunidad científica y asociaciones médicas. A finales de la década de los treinta el 'Council on Pharmacy and Chemistry' (Estados Unidos) evaluó esta terapia y concluyó que no sólo había reportes positivos y negativos, sino que también la falta de protocolos estándares para preparar las mezclas virales dificultaba la comparación entre los estudios publicados. Con la Segunda Guerra Mundial y el descubrimiento de los antibióticos el interés por estudiar a los bacteriófagos se abatió en los Estados Unidos, aunque en Europa continuaron los estudios pero en un contexto militar, como fue demostrado por la Ex-Unión Soviética cuando utilizó la terapia por fagos en heridos de guerra.

Cabe señalar que al mismo tiempo que se desarrollaban los estudios de terapia viral en la década de los treinta se investigaba también la composición química de los bacteriófagos.



El desarrollo de nuevas técnicas experimentales como la ultra centrifugación y filtración por membranas coloidales permitió distinguir que había bacteriófagos de diferentes tamaños. No fue sino hasta 1940 que con la ayuda del microscopio electrónico se pudieron vislumbrar las diferentes morfologías de los bacteriófagos.

En los años siguientes, los bacteriófagos comenzaron a ser utilizados como herramientas para estudiar mecanismos de mutación. Max Delbrück y Salvador Luria diseñaron un método matemático e indirecto, 'the fluctuation test', para demostrar que bacterias mutantes (resistentes a la infección por fagos) estaban presentes en una población bacteriana antes de ser expuesta a dosis letales de fagos (5). En 1952 Joshua y Esther Lederberg diseñaron un método directo para comprobar que las mutaciones ocurrían al azar y en forma independiente a los procesos de selección. La técnica experimental consistía en transferir las colonias de una caja de cultivo sobre terciopelo, para luego colocarla en otra caja de cultivo y crear réplicas conociendo la ubicación exacta de la colonia estudiada. Ellos utilizaron esta técnica para estudiar la resistencia bacteriana a la infección viral y a la estreptomycin. En su estudio concluyeron que mutantes surgían al azar antes de la aplicación del agente selectivo (fago antibiótico) (6).

Por otro lado, Alfred Hershey y Raquel Rotman elucidaron el comportamiento genético de los bacteriófagos cuando estudiaron la morfología de las placas producidas por diferentes mutantes virales, para demostrar que cruces de fagos pueden obtenerse cuando se infectan células con diferentes fagos (7). La mayor parte de los estudios estaban enfocados en bacteriófagos que causaban lisis, a pesar de que el fenómeno de lisogenia ya estaba descrito por Bordet. En 1950 en Paris, André Lowff y Antoinette Gutmann describieron este fenómeno y designaron a la forma latente del fago como profago. Ellos monitorearon la inducción y liberación de partículas virales en células individuales con microscopio (8).



Con el avance en la genética de los bacteriófagos fue descubierto un fenómeno llamado modificación inducida por hospedero. Salvador Luria y Mary Human lo describieron en el fago T2 (9), y Giuseppe Bertani y Jean Weigle en el fago lambda (10). Ambos grupos encontraron que el rango de hospederos para un fago (de acuerdo a su eficiencia para producir placas) era determinado algunas veces por la cepa donde se infectara. De alguna manera la bacteria modificaba al fago y esto alteraba su habilidad para infectar otras cepas. Estudios posteriores durante la década de los sesentas y principios de los setentas revelaron que estas modificaciones eran hospedero-específicas, e incluían metilaciones o glicosilaciones en el DNA del fago y nucleasas secuencia-específicas.

Los hallazgos científicos derivados del estudio de los bacteriófagos permitieron comprobar el flujo de información propuesto por Francis Crick entre el DNA-DNA y DNA-RNA-proteína. La evidencia experimental que apoyó la idea del RNA como molécula intermediaria fue proporcionada por Volkin et al. Ellos encontraron que en células infectadas por el bacteriófago T2, la composición de bases en las moléculas de RNA era similar a la encontrada en el genoma de DNA del fago, mientras que en células no infectadas esto no sucedía. Se dedujo así que la función de un gen viral requería de la síntesis de moléculas de RNA diferentes a las presentes en células no infectadas, y que la composición de las bases y la información contenida en el RNA estaba determinada por el DNA del fago (11). El descubrimiento de bacteriófagos con genomas de RNA en 1961 (12) confirmó lo postulado por Crick cuando se demostró que estas moléculas funcionan de acuerdo al flujo de información predicho. Gracias al desarrollo de técnicas experimentales en bioquímica, genética y microbiología se han podido describir los mecanismos de replicación, transcripción y ensamblaje de partículas en bacteriófagos.



En los últimos años el estudio genómico de los fagos ha tomado gran importancia debido a dos aspectos importantes: se ha calculado que representan la población de sistemas biológicos más abundantes de la tierra (aprox.  $10^{31}$  individuos) (13, 14, 15) y evidencia apunta a que son, en buena medida, responsables de muchos de los cambios en genomas bacterianos. Por ejemplo, es bien sabido que los fagos son vectores importantes para la transferencia de DNA entre cepas bacterianas, y estudios genómicos han revelado que entre 10-20% de algunos genomas bacterianos están compuestos por genes de fagos (16). Los mecanismos hasta ahora caracterizados para estos cambios son transducción, transformación y conjugación.

Existen diferentes tipos de bacteriófagos que pueden contener como material genético una cadena sencilla o doble de DNA o RNA y pueden seguir los ciclos lítico o lisogénico de infección. Los líticos infectan a sus hospederos y programan la síntesis de viriones causando la muerte de la bacteria, mientras que los lisogénicos establecen una relación no letal con su hospedero mediante la integración de su genoma o parte de este en el cromosoma bacteriano (profago) o, como en los casos de los fagos P1, N15, LE1,  $\phi$ 20 y  $\phi$ BB-1, permanecen como plásmidos lineales o circulares (16). Los bacteriófagos en estado lisogénico pueden inducirse para entrar al ciclo lítico. La mayor parte de los genes virales, incluyendo aquellos involucrados en la lisis y la producción de viriones que se encuentran en un profago, no se expresan. Sin embargo, en ciertas ocasiones, algunos genes del profago se expresan en la etapa lisogénica dentro del hospedero, alterando sus propiedades. Estos se denominan genes de conversión lisogénica (LCG, por sus siglas en inglés). Los efectos son variados, y pueden ir desde proteger a la bacteria de una futura infección por otro fago hasta aumentar la virulencia del hospedero patógeno (17, 18).



Por ejemplo, en *Corynebacterium diphtheriae* se encuentra un profago que contiene un gen que codifica para la toxina de la difteria, y en *V. cholerae* serogrupo O1, la toxina de cólera está codificada en un gen del fago filamentosos CTX $\phi$ . Otros ejemplos de este fenómeno se incluyen en la tabla 1 (18, 19).

Tabla 1. Factores de virulencia codificados en bacteriófagos.

Bacteria	Bacteriofago	Gen	Proteína
<i>Escherichia coli</i> 0157:H7	933, H-19B $\phi$ FC3208	stx hly2	Toxina Shiga Enterohemolisina
<i>Shigella flexneri</i>	Sf6	oac	Acetilasa antígeno O
<i>Salmonella enterica</i>	Gifsy-2	sodC-1	Superóxido dismutasa
<i>Vibrio cholerae</i>	CTX $\phi$	ctxAB	Toxina de cólera
<i>Clostridium botulinum</i>	Phage C1	C1	Neurotoxina
<i>Corynebacterium diphtheriae</i>	$\beta$ -phage	tox	Toxina de la difteria
<i>Streptococcus pyogenes</i>	T12	speA	Toxina eritrogénica
<i>Staphylococcus aureus</i>	NA	see, sel	Enterotoxina
<i>Pseudomonas aeruginosa</i>	$\phi$ 13	entA	Enterotoxina A

*E. coli*, *S. flexneri*, *S. enterica*, *V. cholerae*, *P. aeruginosa* son Gammaproteobacterias; *C. botulinum*, *S. pyogenes*, *S. aureus* son Firmicutes y *C. diphtheriae* es Actinobacteria. Tabla modificada de Ref 19.

En la actualidad ha aumentado el interés por estudiar a los bacteriófagos debido al papel que desempeñan para proveer de material genético a sus hospederos más aún, falta por esclarecer muchos de los mecanismos moleculares que estas partículas biológicas utilizan para permanecer latentes. Las herramientas de la genómica y la información en las bases de datos ofrecen una estrategia para abordar la compleja relación evolutiva que existe entre los bacteriófagos y sus hospederos. Una aplicación de estas herramientas se muestra en el trabajo de Liu et al (20) donde secuenciaron 26 genomas de bacteriófagos que infectan *S. aureus* e identificaron 31 familias de polipéptidos cuya expresión inhibe el crecimiento de esta bacteria patógena, sugiriendo así el desarrollo de antibióticos a partir de su descubrimiento.



## 1.1 Biología de Bacteriófagos

Históricamente los bacteriófagos que han sido más estudiados a nivel molecular son los que infectan *Escherichia coli* como f1, M13, fd y  $\lambda$ . Los fagos filamentosos f1, M13 y fd tienen una morfología común, y su estrategia de infección es lítica, mientras que el bacteriófago  $\lambda$  puede tener dos ciclos biológicos; el lítico y el lisogénico. El colifago  $\lambda$  puede insertar su genoma en el cromosoma bacteriano y generar una fase lisogénica. En la fase lítica grupos de genes son expresados en un orden definido (temprana, media, tardía), mientras que en la fase lisogénica solo un grupo limitado de genes se expresa para mantener el estado de profago. Las células que contienen profagos son lisogénicas y en estas puede inducirse la fase lítica para producir partículas virales. Se estima que una célula de 10,000 por generación expresa de forma espontánea la mayoría de los genes del colifago  $\lambda$  y cuando las células se lisan, liberan aprox. 100 partículas virales. Cabe señalar que aquellos bacteriófagos que lisan a sus hospederos se les denomina virulentos mientras que a los que no siempre hacen esto se les conoce como 'temperate'. Cuando una bacteria posee un profago se establece 'inmunidad', que permite prevenir que otros fagos similares infecten a la misma célula. Generalmente un genoma bacteriano contiene una copia de un profago particular. Abajo se muestran las etapas del ciclo lítico del bacteriófago  $\lambda$ , que ha sido estudiado extensamente (21).

### Absorción e inyección

Las fibras largas de la cola de  $\lambda$  hacen contacto con la membrana del hospedero. La fibra central se une a la proteína de unión a maltosa LamB. El DNA viral es inyectado al interior del hospedero.



## Transcripción

Una RNA polimerasa dependiente de DNA en *E. coli* transcribe los genes de  $\lambda$  a partir de los promotores  $P_L$  y  $P_R$  en ambas direcciones. La transcripción finaliza en los terminadores  $T_{L1}$  y  $T_{R1}$ . Si el producto del gen N (pN) está presente permite continuar la transcripción y expresa genes involucrados en la recombinación (*exo*, *bet* y *gam*), replicación (O y P) y regulación positiva (Q). pN está involucrado en la antiterminación y su forma de actuar es la siguiente: nutR es un sitio de unión a pN que está en el transcrito producido a partir del promotor  $P_R$ . pN se une a nutR y a la RNA polimerasa evitando que esta enzima no reconozca a  $T_{R1}$  u otro terminador de la transcripción. pN actúa con las proteínas del hospedero Nus (algunas involucradas en la terminación de la transcripción en *E. coli*). El producto del gen Q está regulado por pN y es un regulador positivo de la expresión tardía de genes. Estos genes codifican proteínas de la cápside, cola y proteínas implicadas en la lisis. La proteína Q reconoce un sitio de unión (*qut*) en el DNA viral y se une a la RNA polimerasa promoviendo la expresión de todos los genes tardíos sin terminación.

## Replicación de DNA

El DNA de  $\lambda$  es de doble cadena y lineal (50kb). En cada extremo de la doble cadena hay un segmento de cadena sencilla de 12 nt que se llama *cos*. Estas cadenas sencillas son complementarias y se unen para formar un círculo de DNA que tiene dos discontinuidades en las cuales la DNA ligasa actúa para cerrar el círculo. Cuando el DNA está cerrado en sus extremos y los productos de los genes O y P (pO y pP respectivamente) han sido generados, la replicación comienza en el sitio *ori*. Primero pO se une a *ori* en el DNA y después pP se une a pO. Este complejo proteico causa que la helicasa DnaB de *E. coli* se una al DNA cerca de *ori* y separe la doble hélice.



Después de separar las cadenas se producen 'primers' por la primasa DnaG de *E. coli* y esto facilita la replicación en ambas direcciones. En la replicación se forma una molécula de DNA en forma de theta que se separa en dos círculos al termino de cada ronda. Después de varios ciclos replicativos "θ" inicia la replicación unidireccional mejor conocida como replicación de círculo rodante donde la cadena líder se replica de forma continua mientras que la cadena retrasada requiere de 'primers' producidos por la primasa DnaG. Aquí se genera una molécula larga de DNA que se llama "concatenado". La proteína Gam de  $\lambda$  inhibe a la exonucleasa RecBC del hospedero que podría digerir al "concatenado".

### **Formación de partículas virales**

Las cápsides consisten de 12 copias de pB organizadas en un sitio llamado "portal" donde el DNA es introducido a la partícula viral y es aquí donde la cola del fago se une. La cabeza consiste de 415 copias de pE y 415 copias de pD que forman un icosaedro (12 vértices y 20 caras). Para empacar el DNA en las partículas virales, un complejo de pA y pNu1 se une a las cápsides vacías y corta el DNA "concatenado" en los sitios *cos*. Una vez que el DNA está dentro de las cabezas se unen las colas y la partículas virales están completas. Los productos de los genes *V*, *I*, *K*, *L* y *M* forman las colas.

### **Lisis**

El producto del gen R es transportado a través de la membrana celular a la pared celular por el producto del gen S. pR corta uniones en la pared celular del hospedero causando la ruptura de la membrana y la liberación de las partículas virales.



Después de aprox. 10-15 min de que  $\lambda$  ha infectado a su hospedero se establece cual ciclo (lítico o lisogénico) tomará la infección. Los eventos que ocurren durante esta decisión y el establecimiento de la lisogenía se describen abajo.

### **Decisión Lisis-Lisogenía**

Durante esta decisión la proteína reguladora CI (represor  $\lambda$ ) se expresa a partir del promotor  $P_{RE}$ , y su expresión depende del activador CII. De acuerdo a la literatura, los mecanismos que determinan los niveles de este activador no están bien caracterizados pero aparentemente dependen de la fisiología del hospedero. Cuando los niveles de CII son altos, la expresión de CI dirige el establecimiento de la lisogenía. Si CII está ausente o en niveles bajos, la célula sigue el ciclo lítico. CII es inestable y susceptible a degradación por las proteasas bacterianas HflB o FtsH. Así mismo, CIII de  $\lambda$  modula la actividad de FtsH y se cree que actúa como inhibidor competitivo. CII y CIII se expresan en la etapa temprana de la infección y a partir de los promotores  $P_R$  y  $P_L$  respectivamente. El crecimiento con bajos niveles de nutrientes o a baja temperatura favorecen la lisogenía.

### **Lisogenía e Inducción del profago**

CI se expresa continuamente del promotor  $P_{RM}$  y mantiene el estado lisogénico. CI regula la actividad de  $P_{RM}$  de tal forma que mantiene un nivel relativamente constante de sí mismo durante el ciclo de crecimiento.  $P_{RM}$  se encuentra en una región regulatoria llamada  $O_R$ . Dos proteínas regulatorias, CI y Cro, actúan en esta región. Si las células contienen CI y no Cro, la lisogenía se establece y persiste la expresión de CI. Por otro lado, si hay Cro y no CI, se presenta el ciclo lítico y Cro sigue expresándose.



El estado lisogénico es bastante estable, pero puede cambiar a un estado lítico cuando el hospedero presenta daño al DNA o inhibición en la replicación de DNA. En estos casos, la respuesta SOS se activa en las células para reparar el daño al DNA o restaurar la replicación. El fago  $\lambda$  toma ventaja de esta respuesta celular y utiliza este mecanismo para inactivar CI. A este proceso se le denomina inducción del profago.

El sistema de regulación SOS involucra dos proteínas: LexA que reprime aprox. 40 genes (genes SOS) y RecA que cuando se activa promueve la auto-proteolisis de LexA inactivándolo. RecA está involucrado en los procesos de recombinación homóloga y reparación de DNA. Si una célula contiene un profago de  $\lambda$ , RecA activado promueve de forma indirecta la degradación de CI y LexA. La acción de RecA modifica el dominio de unión a DNA de CI y previene que dímeros de este se unan al operador (21).

### **Otros mecanismos para la inducción de profagos**

Los bacteriófagos 186 y CTX $\Phi$  presentan mecanismos de inducción, dependientes del sistema SOS, y diferentes al descrito anteriormente para  $\lambda$ . Ambos profagos son inducidos por el tratamiento de agentes que activan la respuesta SOS. El bacteriófago 186 tiene un gen llamado *tum* que está regulado por LexA del hospedero. Cuando el sistema SOS se activa, la proteína Tum se expresa, y actúa como antirepresor antagonizando la función de CI, también presente en 186. El bacteriófago CTX $\Phi$ , que infecta a *V. cholerae*, puede ser inducido por mitomicina C. Su mecanismo de inducción no es claro aún, pero se piensa que su represor no se degrada (21).



## El caso del profago P1

A diferencia de los profagos  $\lambda$  y P2 que se integran al cromosoma de *E. coli*, el profago P1 permanece en el hospedero como plásmido de DNA de copia única. Las células hijas reciben una copia exacta de P1, y aquellas que pierden al profago mueren por acción de una proteína que es liberada de un complejo inhibidor después de que P1 se ha perdido de la célula (45, 46).

## Transducción especializada

La transducción especializada fue descrita en  $\lambda$ , y es el mecanismo por el cual genes bacterianos localizados en ambos extremos del profago pueden ser incorporados a su genoma. El genoma de  $\lambda$  se integra en el cromosoma de *E. coli* entre los genes *gal*, involucrados en la fermentación de la galactosa, y *bio*, responsables de la biosíntesis de biotina. Este proceso requiere de la proteína de integración Int de  $\lambda$  y de proteínas del hospedero llamadas factores de integración. Un evento de entrecruzamiento (crossover) entre el sitio *attP* en el fago y *attB* en *E. coli* integra el genoma de  $\lambda$  en su cromosoma generando sitios híbridos, *attB/P* y *attP/B*. El genoma viral integrado tiene un nuevo orden de genes. La liberación del profago  $\lambda$  requiere de las proteínas Xis e Int, si un evento de recombinación “ilegítima” ocurre entre el DNA del profago y el cromosoma bacteriano, la molécula resultante de DNA viral puede contener los genes *gal* o *bio*. Si esta molécula de DNA es empaquetada puede transferir estos genes a otra cepa de *E. coli* (45).



## Transducción generalizada

Este es otro mecanismo para transferir genes entre bacterias donde el bacteriófago empaca DNA del hospedero dentro de su cápside o cabeza. Los fagos que empaquetan DNA del hospedero no llevan material genético viral, y la longitud del DNA bacteriano empaquetado es igual a la longitud del DNA viral. E. S. Lennox demostró la transducción generalizada utilizando el colifago P1 cuya cápside puede contener 100kb de DNA de *E. coli* (47).

## Virus de Arqueas

Se han aislado partículas virales de *Halobacterium* spp. y *Methanobacterium* spp. (Euryarchaeota). Se encontró que la mayoría de estos presentan morfologías similares a las familias *Myoviridae* y *Siphoviridae*. Además, se han aislado virus de *Sulfolobus* y *Thermoproteus* (Crenarchaeota) que presentan morfologías únicas y se han clasificado en cuatro familias virales: *Rudiviridae*, *Lipothrixviridae*, *Fuselloviridae* y *Guttaviridae*, para esta última no se ha encontrado una cepa que pueda propagarse establemente en el laboratorio (22, 23, 24). Los genomas de estas familias son de doble cadena de DNA, son lineales para *Lipothrixviridae* y *Rudiviridae*, y circulares para *Fuselloviridae* y *Guttaviridae*. Los genomas de *Fuselloviridae* existen dentro de la célula como plásmidos, y tienen la capacidad de integrarse en el genoma de sus hospederos por la acción de una integrasa codificada en su genoma. La integración ocurre en un gen específico de tRNA como sucede para algunos bacteriófagos, en este caso la integración en su genoma deja intacto al gen de la integrasa mientras que en los *Fuselloviridae* queda dividido (23).



Se ha encontrado que la producción del Fusellovirus SSV1 puede ser inducida por exposición a radiación UV o tratamiento con mitomicina, y esto no mata al hospedero (25). También se ha encontrado que los genomas lineales no codifican integrasas y no hay evidencia que apunte a que se integran en el genoma de *Sulfolobus*. Los miembros de estas familias parecen propagarse de forma lisogénica, hasta el momento el virus TTV1 (*Lipothrixviridae*) exhibe un comportamiento lítico. Esta forma de propagación viral refleja la necesidad de minimizar o evitar el contacto con las condiciones extremas e inestables del ambiente donde habitan sus hospederos.

Se han propuesto otras tres familias virales que infectan Crenarchaeota pero aún no han sido aceptadas por ICTV (International Committee for Taxonomy of Viruses), estas son *Globuloviridae*, *Bicaudaviridae* y *Ampullaviridae* (24). Este comité clasifica a los bacteriófagos de acuerdo a su morfología, y en la actualidad existe una controversia con respecto a la denominación de las partículas virales que infectan arqueas pues no se parecen a los bacteriófagos que fueron aislados en bacterias, y por lo tanto, este término no concuerda con la biología de estas partículas que infectan a otros procariontes. A las partículas virales que infectan arqueas se les denomina virus. La distinción de términos entre bacteriófagos y virus se mantiene a lo largo de este trabajo.



## 1.2 Características Generales de los Bacteriófagos

**Orden:** Caudovirales (26, 27)

**Familias:** Myoviridae, Podoviridae, Siphoviridae

**Morfología:** Consisten de una cabeza, cola, 'spikes' y fibras. La cápside no tiene envoltura. La cabeza tiene una simetría de icosaedro. La cola es helicoidal y puede ser contráctil, rígida o flexible, y presentar forma de bastón o tener estructura de discos aplanados. La cola puede presentar un collar o vaina (sheath) que está separada de la cabeza a través de un cuello y fibras.

**Genoma:** No está segmentado y contiene una molécula lineal de DNA de doble cadena. Puede estar circularizado. En los extremos presenta secuencias redundantes, y proteínas enlazadas covalentemente. Estos pueden ser cohesivos o romos. El DNA de doble cadena puede tener gaps de cadena sencilla.

**Proteínas:** El genoma codifica proteínas estructurales y no estructurales.

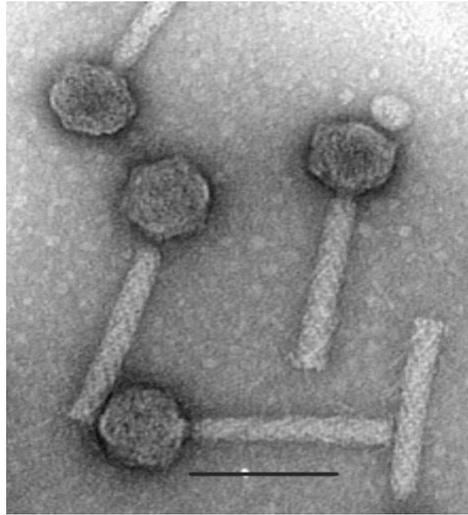


Figura 1. Virus de enterobacteria P2 (26).

**Familia: Myoviridae** (26)

**Género: virus parecidos a T4, P1, P2 (Figura 1), Mu, SPO1 y  $\phi$ H.**

**Morfología:** Consisten de una cabeza y cola, que carecen de envoltura. La cabeza está separada de la cola por un complejo de proteínas que forman un tubo central y una vaina contráctil que presentan fibras largas y cortas. La cola es helicoidal y la cabeza tiene simetría de icosaedro. La cápside está formada por 152 capsómeros, la cola es rígida, larga y contráctil.

**Genoma:** No está segmentado y consiste de una molécula lineal de DNA de doble cadena. El genoma presenta la base 5-hidroxi-metil citosina en lugar de timidina.

**Proteínas:** Codifica proteínas estructurales y no estructurales.

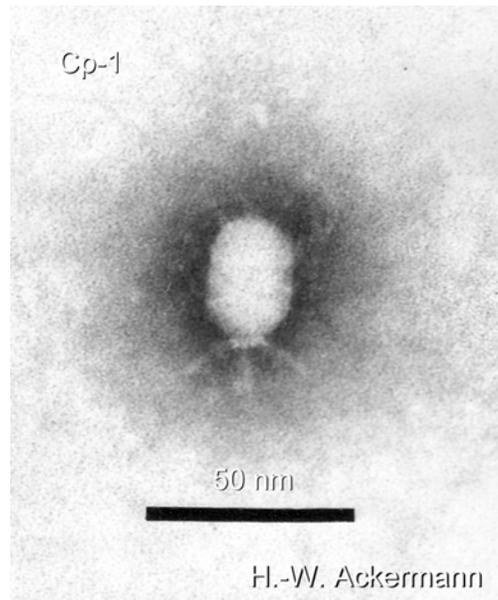


Figura 2. Bacteriófago Cp-1 (26).

**Familia:** Podoviridae (26)

**Género:** T7, P22 y  $\phi$  29

**Morfología:** Los fagos tienen cabeza y cola, no tienen envoltura. La cabeza tiene una simetría de icosaedro y la cola es helicoidal. La cabeza es hexagonal y consiste de 72 capsómeros. La cola es corta, gruesa, en forma de bastón, no contráctil y consiste de discos aplanados (Figura 2).

**Genoma:** No está segmentado y contiene una molécula lineal de DNA de doble cadena. El genoma tiene secuencias terminales redundantes.

**Proteínas:** Codifica proteínas estructurales y no estructurales. Los viriones consisten de al menos 9 proteínas estructurales localizadas en la cabeza y cápside. Presenta genes que codifican a la polimerasa y enzimas como la transferasa.



**Organización Genómica y Replicación:** Los fagos son absorbidos por la cola a través de receptores localizados en la pared celular del hospedero y entran mediante la acción de la lisozima. Las partículas virales son degradadas en el citoplasma. La transcripción viral está regulada temporalmente, hay dos clases de genes: tempranos y tardíos. Los genes con funciones relacionadas están agrupados en el genoma y la replicación es bidireccional. Las proteínas virales se acumulan en el citoplasma, y las proteínas de la cabeza se ensamblan con el genoma cuando este ha sido cortado. Los viriones se forman en el espacio perinuclear y son liberados a través de la lisis celular.

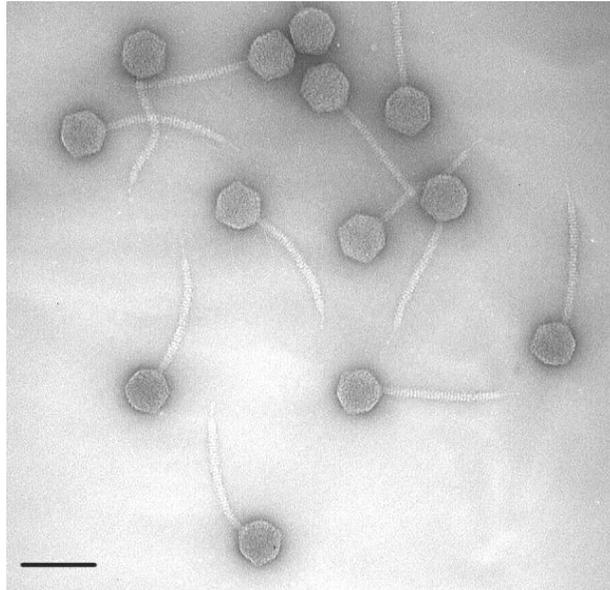


Figura 3. Virus del bacteriófago  $\lambda$  (26).

**Familia:** Siphoviridae (26)

**Géneros:**  $\lambda$  (Figura 3), T1, T5, L5, c2,  $\psi$ M1

**Morfología:** Los fagos consisten de una cabeza y cola, y carecen de envoltura. La cabeza tiene simetría de icosaedro y una cola helicoidal, la cápside es hexagonal y consiste de 72 capsómeros. La cola es filamentosa, no contráctil y tiene fibras cortas terminales.

**Genoma:** No está segmentado y contiene una molécula lineal de DNA de doble cadena. El genoma tiene extremos cohesivos.

**Proteínas:** Codifica proteínas estructurales y no estructurales.

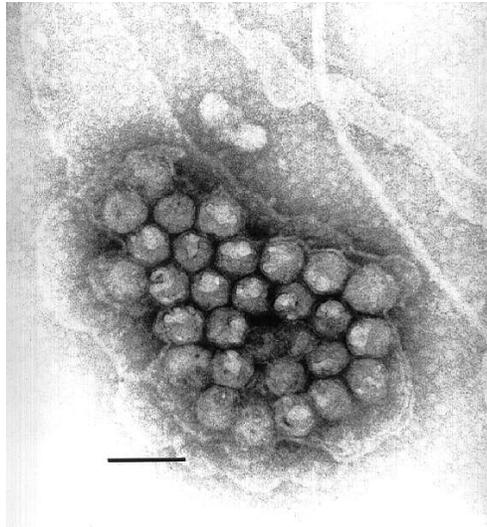


Figura 4. Bacteriófago PM2 (26).

**Familia:** Corticoviridae  
**Género:** Corticovirus (26, 28)

**Morfología:** Los fagos consisten de una cápside, nucleocápside y una membrana lipídica interna localizada entre estas. Ambas cápsides son redondas, presentan simetría de icosaedro, y están formadas por tres capas. La cápside exterior presenta proteínas que se proyectan de la superficie y no tiene envoltura (Figura 4).

**Genoma:** No está segmentado y contiene una molécula circular de DNA superenrollado de doble cadena.

**Proteínas:** Codifica proteínas estructurales y no estructurales.

**Lípidos:** Están presentes en la membrana interna ubicada entre la cápside y nucleocápside. Los lípidos virales forman una bicapa lipídica. La membrana viral contiene fosfatidil glicerol y fosfatidil etanolamina.

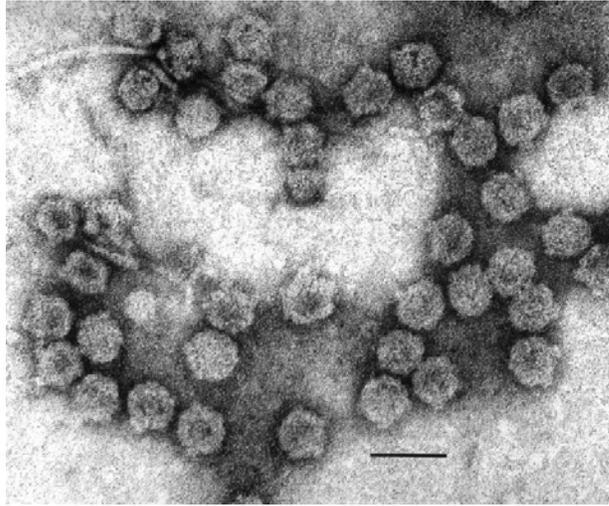


Figura 5. Bacteriófago  $\phi$ X174 (26).

**Familia:** Microviridae (26)

**Género:** Microvirus, Spiromicrovirus, Chlamydiamicrovirus, Bdellomicrovirus.

**Morfología:** Los viriones consisten de una cápside, no tienen envoltura, son redondos y con simetría de icosaedro. La cápside consiste de 12 capsómeros y tiene proteínas que se proyectan de la superficie (Figura 5).

**Genoma:** No está segmentado y contiene una molécula circular de DNA de cadena sencilla +.

**Proteínas:** Codifica proteínas estructurales y no estructurales.

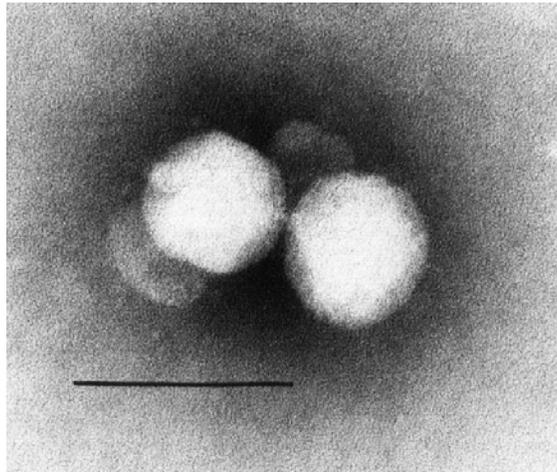


Figura 6. Bacteriófago  $\phi 6$  (26).

**Familia: Cystoviridae**

**Género: Cystovirus (26)**

**Morfología:** Los fagos consisten de una envoltura, cápside y un complejo proteico asociado a la polimerasa. La cápside viral tiene envoltura. Los viriones son esféricos y están cubiertos de proteínas de absorción P3 y proteínas de fusión P6. La cápside y nucleocápside son redondas y presentan simetría de icosaedro, la última contiene el complejo de polimerasa (Figura 6).

**Genoma:** Está compuesto de tres segmentos lineales de RNA de doble cadena. Cada partícula viral tiene los tres segmentos.

**Proteínas:** Codifica proteínas estructurales y no estructurales que están localizadas en la envoltura (P3, P6, P9, P10 y P13), nucleocápside (P8 y P5) y el complejo de polimerasa (P1, P2, P4 y P7).



**Proteínas estructurales:** P10 es una proteína de membrana que tiene actividad lítica y sufre modificaciones postraduccionales que incluyen fosforilación y mistirilación. La proteína P3 está involucrada en la absorción y puede fosforilarse. La proteína P6 es membranal y sirve de anclaje para que P3 se fusione a la membrana. La proteína de envoltura P9 está involucrada en el ensamblaje, P8 se une al genoma, P5 posee actividad lítica e inicia la lisis celular cuando se ensamblan las partículas virales. Las proteínas P1, P2, P4 y P7 forman el complejo de polimerasa y son expresadas en la fase temprana de la infección. Así mismo, P8, P5, P1, P2, P4 y P7 son proteínas de la nucleocápside.

**Lípidos:** Están localizados en la envoltura. Los lípidos virales provienen de la membrana del hospedero.

**Organización del Genoma y Replicación:** Los fagos son absorbidos a través de receptores localizados en el pili que se retraen y acercan al virus hacia la membrana del hospedero. Al ponerse en contacto, ambas membranas se fusionan y este proceso es seguido por la acción de una enzima lítica. En el citoplasma del hospedero las partículas virales se degradan. Cuando se libera el complejo de polimerasa se activa la transcripción de genes de expresión temprana.

**Transcripción:** Está temporalmente regulada y hay genes expresados en las fases temprana y tardía de la infección.

Segmento L: Codifica proteínas de forma unidireccional y el transcrito es policistrónico, codifica cuatro proteínas que constituyen el complejo de polimerasa y se expresa en la fase temprana.



Segmento M: También presenta una organización unidireccional, presenta cinco genes que codifican proteínas de membrana y 'spikes'.

Segmento S: El transcrito producido es policistrónico y codifica tres proteínas estructurales para la cápside. También presenta genes que codifican para una endopeptidasa (implicada en la lisis y entrada de la partícula viral), y proteínas involucradas en el ensamblaje de la cápside.

**Familia: Plasmaviridae**

**Género: Plasmavirus (26)**

**Morfología:** Los viriones consisten de una envoltura, un complejo nucleoproteico y cápside. La cápside tiene envoltura y en las partículas virales no se detecta una estructura regular, es común encontrar estructuras esféricas que carezcan genoma superenrollado.

**Genoma:** Contiene una molécula circular de DNA superenrollado de doble cadena.

**Proteínas:** El genoma codifica proteínas estructurales. Los fagos tienen al menos cuatro proteínas estructurales que se localizan en la envoltura.

**Lípidos:** Se encuentran en la envoltura. Los lípidos virales provienen de la membrana del hospedero y forman una bicapa.

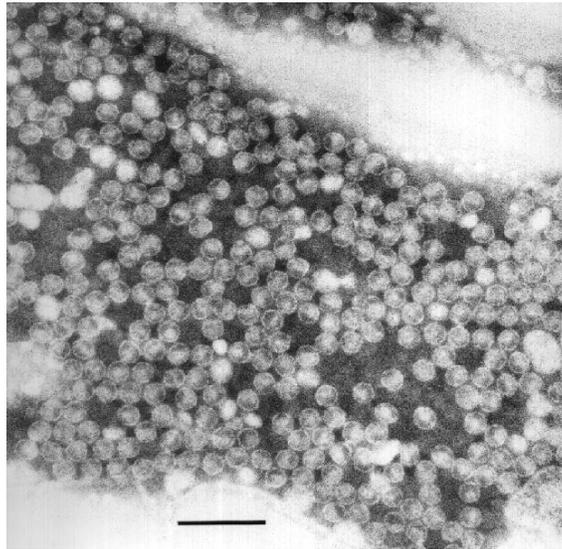


Figura 7. Bacteriófago MS2 (26).

**Familia:** Leviviridae

**Géneros:** Allolevivirus, Levivirus (26)

**Morfología:** Los fagos consisten de una cápside, nucleocápside y no tienen envoltura. Ambas cápsides son redondas y presentan simetría de icosaedro. La cápside consiste de 32 capsómeros (Figura 7).

**Genoma:** Contiene una molécula lineal de RNA de cadena sencilla +.

**Proteínas:** Codifica proteínas estructurales y no estructurales. Los viriones consisten de dos proteínas estructurales localizadas en la cápside.

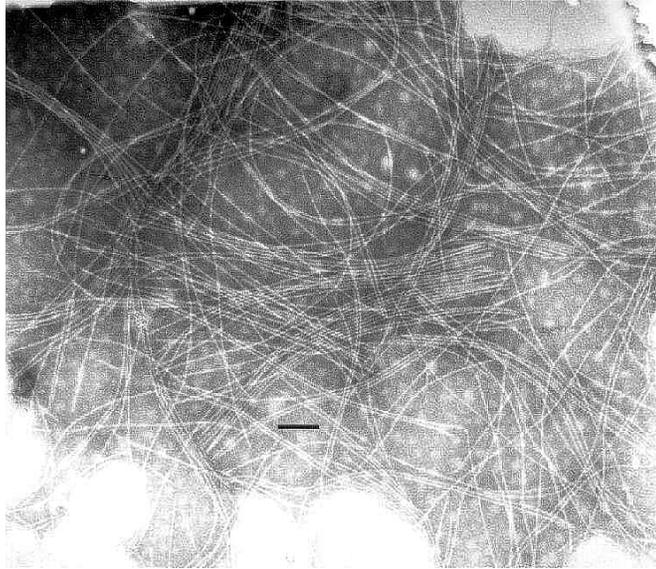


Figura 8. Bacteriófago H75 (26).

**Familia:** Inoviridae

**Género:** Inovirus, Plectovirus (26)

**Morfología:** Los fagos consisten de una cápside, nucleocápside y no tienen envoltura. Ambas cápsides son oblongadas. La cápside tiene forma de bastón, y los viriones son de diferentes longitudes (Figura 8).

**Genoma:** Contiene una molécula circular de DNA de cadena sencilla +.

**Proteínas:** Codifica proteínas estructurales y no estructurales.

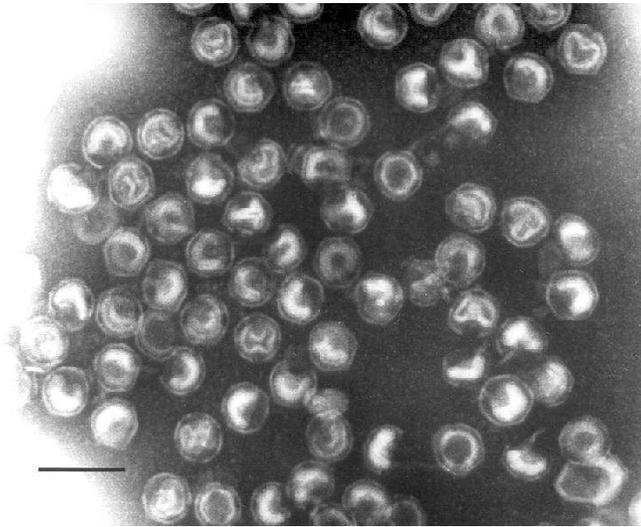


Figura 9. Bacteriófago 37-61 (26).

**Familia: Tectiviridae**

**Género: Tectivirus (26)**

**Morfología:** Los viriones consisten de una cápside/nucleocápside, una membrana lipídica interna y no presentan cola ni envoltura. Ambas cápsides son redondas y presentan simetría de icosaedro. La cápside está compuesta por dos capas, la externa es hexagonal y presenta proteínas que se proyectan de la superficie, mientras que la interna consiste de una vesícula lipoproteíca (Figura 9).

**Genoma:** Contiene una molécula lineal de DNA enrollado de doble cadena.

**Proteínas:** El genoma codifica proteínas estructurales.

**Lípidos:** Están localizados en la vesícula interna y forman una bicapa. Los lípidos virales vienen del hospedero.



**Familia: Rudiviridae**

**Género: Rudivirus (26)**

**Morfología:** Consisten de una cápside/nucleocápside oblongada, no tienen envoltura y exhiben simetría de icosaedro. La cápside tiene forma de bastón y es rígida.

**Genoma:** Contiene una molécula lineal de DNA de cadena sencilla.

**Proteínas:** Codifica proteínas estructurales.

**Familia: Fuselloviridae**

**Género: Fusellovirus (26)**

**Morfología:** Consisten de una nucleocápside y una envoltura. Los viriones tienen forma de limón y son de diferentes tamaños, son flexibles, tienen protuberancias que se extienden a través de la envoltura y en un extremo presentan fibras cortas. Las envolturas virales contienen dos proteínas virales y una proteína asociada a DNA.

**Genoma:** Contiene una molécula circular de DNA superenrollado de doble cadena.

**Proteínas:** Codifica proteínas estructurales.

**Lípidos:** Están localizados en la envoltura y provienen de las membranas de los hospederos.

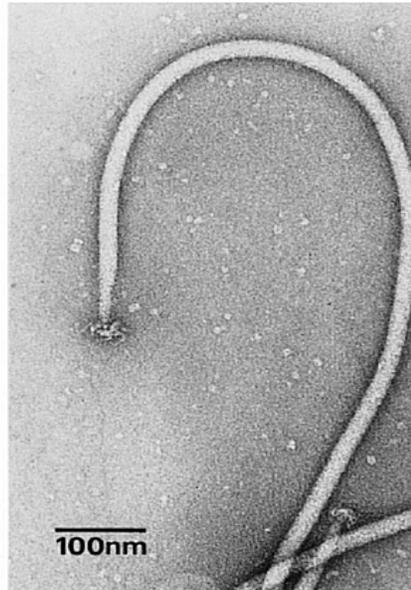


Figura 10. Bacteriófago filamentososo SIFV (26).

**Familia:** *Lipothrixviridae* (26)

**Género:** *Alphalipothrixvirus*, *Betalipothrixvirus*, *Gammalipothrixvirus*.

**Morfología:** Los viriones consisten de una envoltura y nucleocápside, tienen forma de bastón, son rígidos y presentan proteínas que se proyectan desde la nucleocápside y a través de la envoltura. La cápside/nucleocápside es oblongada y presenta simetría helicoidal (Figura 10).

**Genoma:** Contiene una molécula lineal de DNA de doble cadena.

**Proteínas:** El genoma codifica proteínas estructurales y no estructurales.

**Lípidos:** Están localizados en la envoltura y provienen de la membrana de los hospederos. La membrana viral incluye glicolípidos y fosfolípidos.



## 2. OBJETIVOS

Con el propósito de entender la dinámica evolutiva del sistema fago-hospedero, en términos de transferencia de genes y su mantenimiento en los genomas del fago y su hospedero, se realizó un estudio comparativo utilizando como modelo los genomas y las diferentes proteínas de fagos y varios organismos (eucariontes y procariontes) reportados en las bases de datos NCBI (<http://www.ncbi.nlm.nih.gov/genomes/static/phg.html>) y Kegg (<http://www.genome.jp/kegg/kegg2.html>). Se generó una base de datos que contuvo todas las secuencias de proteínas de los bacteriófagos y virus de arqueas reportados en la base de datos NCBI. Se realizaron alineamientos pareados entre las secuencias de proteínas de fagos y sus hospederos usando BLAST (29).

El análisis de los resultados se llevó a cabo en dos etapas, la primera involucró un análisis general de los 7154 alineamientos obtenidos a partir de BLAST, estudiando su distribución en los diferentes hospederos y las anotaciones que presentaban tales proteínas. En una segunda etapa, se seleccionaron las proteínas de los virus que infectan Crenarchaeota, específicamente *Sulfolobus* spp., para analizar detalladamente su distribución y para describir, de ser posible, sus funciones ocupando la base de datos pfam (<http://pfam.sanger.ac.uk>).

Los objetivos para este proyecto se resumen de la siguiente manera: (1) generar una base de datos que contenga todas las secuencias de proteínas de los bacteriófagos y virus de arqueas reportados en la base de datos NCBI, (2) diseñar una metodología para facilitar el análisis de toda la información obtenida, y (3) identificar proteínas cuya posible función pueda ser probada experimentalmente a largo plazo (homólogos celulares que puedan estar adoptando nuevas funciones).



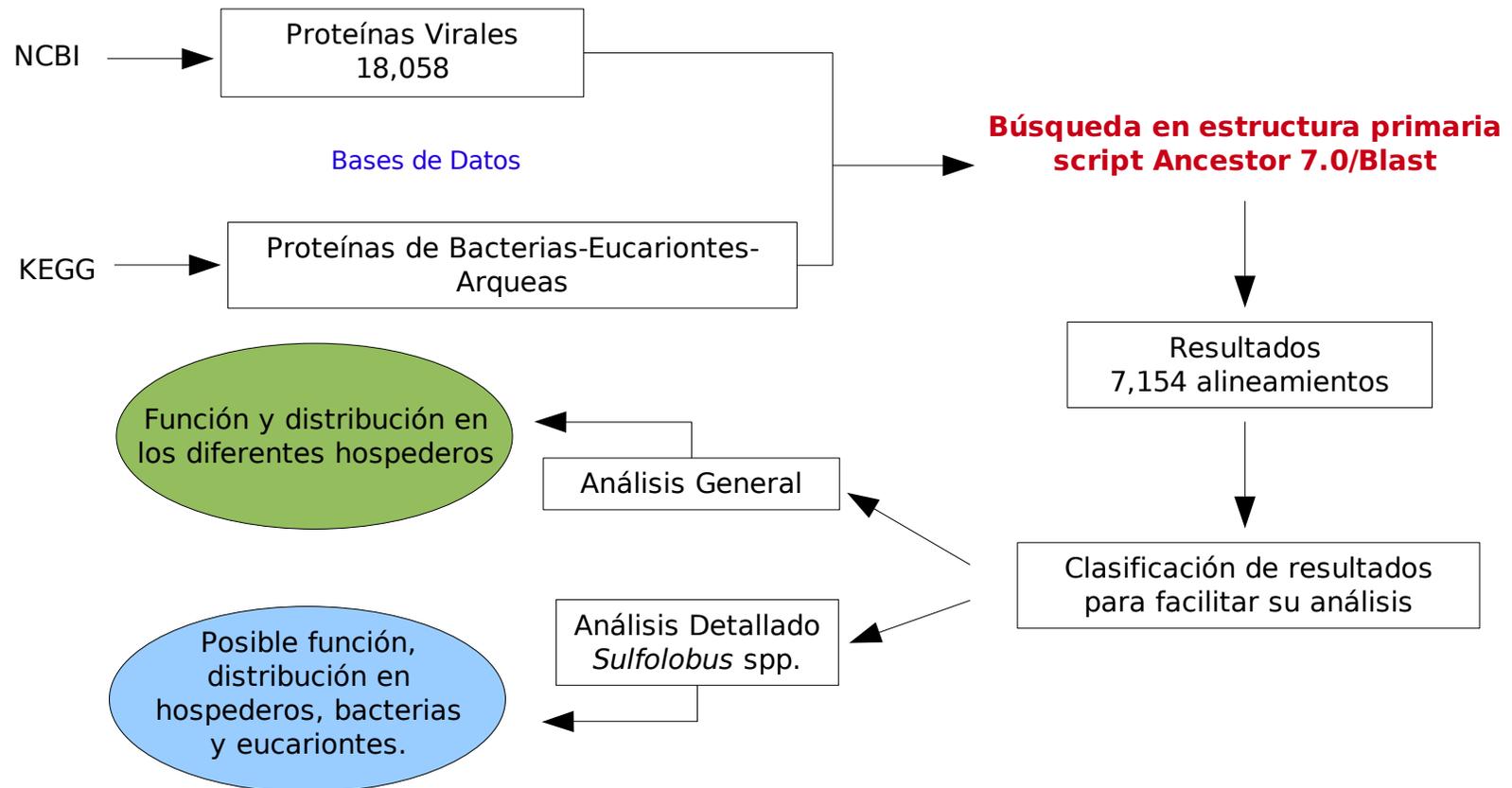
### **3. HIPÓTESIS**

Los genes xenólogos resultado de la interacción entre los bacteriófagos y sus hospederos procariontes deben presentar un sesgo funcional hacia cualquier etapa del ciclo biológico viral que beneficie la producción de partículas virales y su mantenimiento en el genoma viral.



#### 4. METODOLOGÍA

El siguiente diagrama de flujo muestra el plan de trabajo que se siguió:





#### **4.1 Bases de Datos y Búsqueda en Estructura Primaria**

Se crearon dos bases de datos: la primera contiene las secuencias de proteínas virales reportadas para 275 genomas de bacteriófagos obtenidos de NCBI, y la segunda incluye todas las secuencias de proteínas reportadas para 450 genomas de procariontes y eucariontes de la base de datos KEGG. Para la construcción de ambas se utilizaron genomas completamente secuenciados. A continuación se muestra la distribución taxonómica de los diferentes genomas utilizados para la base de datos de bacteriófagos (Figura 11) y para los procariontes y eucariontes (Figura 12). Para ver en detalle la distribución de los genomas de bacterias y eucariontes consultar la Figura 1S y la Tabla 2S en suplementos.

Se realizaron alineamientos pareados entre ambas bases de datos, proteínas virales contra proteínas de bacterias, eucariontes y arqueas. Se utilizó el script Ancestor 7.0 que automatiza el proceso realizado por BLAST (29). Esta búsqueda fue en estructura primaria de proteínas, y el valor de corte utilizado fue de  $e^{-0.000001}$ . Este valor de corte seleccionó aquellas secuencias de proteínas que estuvieran cercanas al criterio de 22% de identidad y 35% de similitud para identificar homología. Por cada resultado arrojado por BLAST (29) se creaban cinco archivos y todos los resultados positivos se guardaban en una tabla. Los diferentes archivos creados se ejemplifican en la tabla 2.

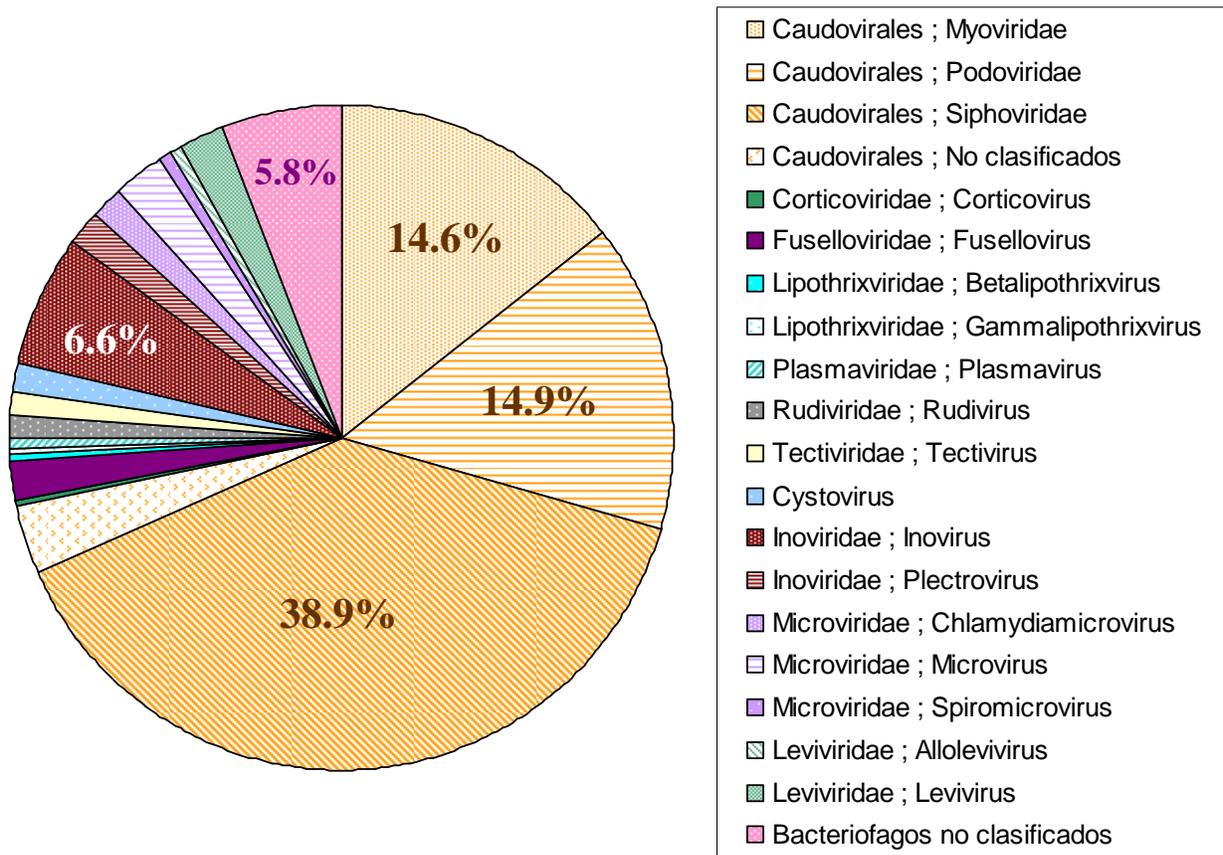


Figura 11. Los 275 genomas utilizados en la base de datos de bacteriofagos y sus porcentajes. La mayoría pertenecen al orden Caudovirales y a las familias Myoviridae (14.6%), Podoviridae (14.9%) y Siphoviridae (38.9%).

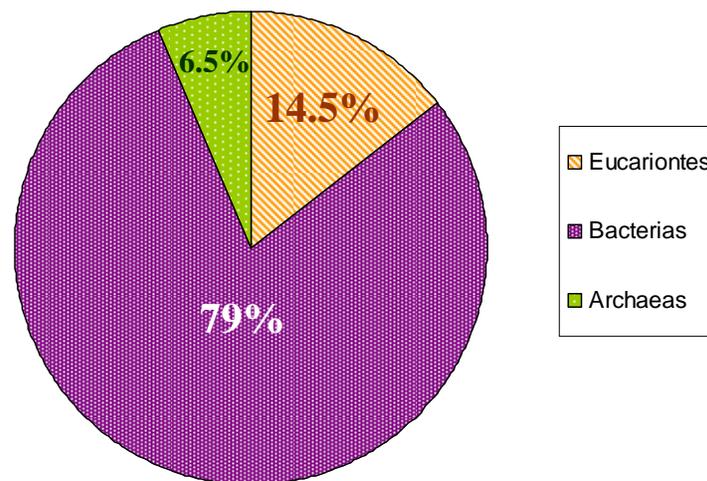


Figura 12. Genomas de eucariontes y procariontes ocupados para la segunda base de datos. El 79% de los genomas pertenecen al dominio bacteria.



Tabla 2. Archivos creados por Ancestor 7.0/BLAST.

Archivo	Información que contiene
Alineación	Alineaciones de proteínas en formato de texto
Dominios	Dominios encontrados en las alineaciones
Genes	Lista de identificadores para las proteínas en cada alineación
secBLAST	Lista de secuencias de proteínas en formato fasta
Lista	Lista de resultados desplegados por genoma
outPSI	Alineación desplegada por BLAST

En total se obtuvieron 7154 alineamientos que se analizaron de forma general de acuerdo a la anotación y función que presentaban tales proteínas. Para realizar un análisis detallado se clasificaron los resultados de acuerdo a los hospederos presentes en la base de datos, bacterias y arqueas. Se obtuvieron los siguientes grupos: Gammaproteobacterias, Betaproteobacterias, Epsilonproteobacterias, Deltaproteobacterias, Alfacaproteobacterias, Firmicutes, Actinobacterias, Clamidiales, Espiroquetas, Cianobacterias, Bacteroidetes, Chlorobi, Deinococcus, Euryarcheota y Crenarchaeota. Abajo se muestra un ejemplo de como están organizados los resultados en cada grupo (Tabla 3). Se eligió a Crenarchaeota, específicamente *Sulfolobus* spp. para analizarlo.



Tabla 3. Tabla correspondiente a un grupo de hospederos bacterianos.

	Chlorobium tepidum	Chlorobium chlorochromatii	Pelodictyon luteolum
>vir:46 001447-Acholeplasma phage L2-integrase	0	1	0
>vir:113 005885-Actinoplanes phage phiAsp2-Pas55	1	1	1
>vir:147 000935-Acyrtosiphon pisum bacteriophage APSE-1-P13	0	1	0
>vir:225 006557-Bacillus clarkii bacteriophage BCJA1c-DEAD box family helicase	0	2	1
>vir:268 006557-Bacillus clarkii bacteriophage BCJA1c-lysin	1	1	1
>vir:290 002649-Bacillus phage GA-1 virion-peptidoglycan hydrolase	1	1	1
>vir:322 001423-Bacillus phage PZA-morphogenesis protein B	0	1	0
>vir:464 005135-Bacteriophage 44RR2.8t-topoisomerase II large subunit	1	1	1
>vir:465 005135-Bacteriophage 44RR2.8t-hypothetical protein	1	1	1
>vir:656 005135-Bacteriophage 44RR2.8t-dCMP deaminase	1	1	1
>vir:692 005135-Bacteriophage 44RR2.8t-hypothetical protein	2	2	3
>vir:699 005135-Bacteriophage 44RR2.8t-topoisomerase II medium subunit	1	1	1
>vir:714 005356-Bacteriophage 77-77ORF015	1	1	1
>vir:724 005356-Bacteriophage 77-77ORF001	1	0	1
>vir:890 003216-Bacteriophage A118-putative integrase	0	0	2
>vir:923 003216-Bacteriophage A118-putative SSB protein	2	2	2

Estas bacterias están clasificadas en el grupo Clorobi. La columna izquierda muestra las anotaciones de las proteínas virales, arriba están los nombres de los organismos de donde provienen las proteínas que alinearon con las proteínas virales. En verde están los resultados con al menos un 'hit'. Por falta de espacio solo se muestra una parte de la tabla.



## 5. RESULTADOS

Las familias virales Myoviridae, Podoviridae y Siphoviridae concentran el 88.9% de los resultados obtenidos por BLAST. Las proteínas reportadas para 37 genomas de bacteriófagos no dieron resultados, esto puede ser por: (a) un efecto del valor de corte utilizado para este estudio, o (b) los hospederos que estos bacteriófagos infectan no estuvieron incluidos en la base de datos. En total 7154 proteínas virales dieron al menos un alineamiento significativo 'hit' (Tabla 4) de un total de 18062 proteínas en la base de datos viral.

Tabla 4. No. de proteínas virales con al menos un 'hit' de acuerdo a la clasificación taxonómica de los bacteriófagos.

<b>Clasificación de Bacteriófagos</b>	<b>No. proteínas</b>
Caudovirales ; Myoviridae	1377
Caudovirales ; Podoviridae	720
Caudovirales ; Siphoviridae	4097
Caudovirales ; No clasificados	167
Corticoviridae ; Corticovirus	4
Fuselloviridae ; Fusellovirus	46
Lipothrixviridae ; Betalipothrixvirus	6
Lipothrixviridae ; Gammalipothrixvirus	1
Plasmaviridae ; Plasmavirus	2
Rudoviridae ; Rudovirus	12
Tectiviridae ; Tectivirus	35
Cystovirus	1
Inoviridae ; Inovirus	65
Inoviridae ; Plectrovirus	2
Microviridae ; Chlamydiamicrovirus	11
Microviridae ; Microvirus	1
Microviridae ; Spiromicrovirus	1
Leviviridae ; Allolevivirus	0
Leviviridae ; Levivirus	0
Bacteriofagos no clasificados	606
<b>TOTAL</b>	<b>7154</b>



## 5.1 Caracterización Funcional de las Proteínas Virales

Se clasificaron las 7154 proteínas virales para conocer cuantas proteínas están anotadas y cuales son sus funciones. Primero se contabilizaron las proteínas anotadas contra las hipotéticas y se encontró lo siguiente (Tabla 5).

Tabla 5. Porcentajes de proteínas de acuerdo a su anotación.

<b>Proteínas</b>	<b>Base de Datos (%)</b>	<b>Resultados (%)</b>
Hipotéticas	61.33	55.51
Anotadas con una función	38.14	43.84
Anotadas con dos funciones	0.53	0.66
<b>TOTAL</b>	<b>100</b>	<b>100</b>

En la base de datos de las proteínas virales más de la mitad están consideradas como ORF o hipotéticas, este comportamiento se refleja también en las proteínas obtenidas después de la búsqueda. En este trabajo el grupo de las proteínas hipotéticas incluía proteínas cuya anotación no especificaba una función clara. En la Tabla 6 se presentan los nombres de las proteínas que se incluyeron en este grupo. En cuanto a las proteínas anotadas se encontró que hay proteínas asignadas a una función y proteínas asignadas con dos funciones.

Las proteínas asignadas con dos funciones están involucradas en la replicación del genoma viral, recombinación, regulación transcripcional, formación de partículas virales, degradación de proteínas y ácidos nucleicos.



Tabla 6. Proteínas cuyas anotaciones forman el grupo de las proteínas hipotéticas y su porcentaje.

	<b>Proteína</b>	<b>Base de Datos (%)</b>	<b>Resultados (%)</b>
	Desconocida	8.5	4.73
	Putativa	0.97	2.19
	Dominio conservado	0.07	0.05
	ORF	28.49	44.72
	Hipotética	52.42	38.55
	Hipotética conservada	2.51	1.84
Proteínas de Fago	Proteínas de fago (varias)	5.85	7.2
	Proteína relacionada al fago	0.01	0.03
	phi like protein	0.07	0
	Proteína predicha de 4.5-86.4Kd	0.7	0.4
	Homólogo al gen 17 del fago phi29	0.01	0
	Proteína similar de fago	0.02	0.03
	Proteína putativa de fago	0.02	0
	Proteína conservada de fago	0.25	0.2
	Proteína conservada de función desconocida	0.04	0
	Proteína hipotética de fago	0.05	0.05
		<b>100</b>	<b>100</b>



## 5.2 Bacteriófagos y Hospederos

Debido a la cantidad importante de resultados obtenidos, el análisis no se realizó tomando en cuenta todo el universo de información. Para hacer un análisis detallado se buscó un grupo de bacteriófagos cuyos hospederos estuviesen incluidos en la base de datos de los organismos. Para ello se consultó la literatura y se encontró que los bacteriófagos que conforman la base de datos a estudiar infectan a cianobacterias, firmicutes, proteobacterias y crenarchaeota (Tabla 7). Debido a que los genomas de *Aeromonas hydrophila* y *Aeromonas salmonicida* no están secuenciados, no se les incluyó en la base de datos de organismos de este estudio. Pese a que hay 25 genomas de actinobacterias y 22 genomas de euryarchaeotas reportados aquí, no se encontraron reportes de fagos que tuvieran a estos procariontes como hospederos. Solo para el bacteriófago L5 no se encontró información debido a que se retiró de la base de datos del NCBI y no se incluyó. Los siguientes bacteriófagos no tienen hospederos bien caracterizados: Para el bacteriófago JL001 se reporta que su hospedero puede ser una bacteria perteneciente a las alpha-proteobacterias y que se encuentra asociada a las esponjas marinas (30), en cuanto al bacteriófago APSE-1 se reporta que su hospedero es el endosimbionte secundario de *Acyrtosiphon pisum* que ha sido identificado como miembro de Enterobacteriaceae y está asociado con *Escherichia coli* (31). Finalmente, para el bacteriófago RB43 no se encontró alguna publicación pero en la ficha técnica reportada en la base de datos del NCBI se registra que su hospedero pertenece también a Enterobacteriaceae.



Tabla 7. Hospederos y bacteriófagos encontrados en las bases de datos empleadas en este estudio.

Clasificación de Bacteriofagos	Cyanobacteria (17)	Deinococcus-Thermus (4)	Firmicutes (83)	Fusobacteria (1)	Planctomycetes (1)	Proteobacteria (177)	Spirochaetes (7)	Thermotogae (1)	Crenarchaeota (5)	Euryarchaeota (22)	Nanoarchaeota (1)
	Caudovirales ; Myoviridae (40)	*		*			*				
Caudovirales ; Podoviridae (41)			*			*					
Caudovirales ; Siphoviridae (107)			*			*					
Caudovirales ; No clasificados (9)											
Corticoviridae ; Corticovirus (1)						*					
Fuselloviridae ; Fusellovirus (5)									*		
Lipothrixviridae ; Betalipothrixvirus (1)									*		
Lipothrixviridae ; Gammalipothrixvirus (1)									*		
Plasmaviridae ; Plasmavirus (1)			*								
Rudiviridae ; Rudivirus (3)									*		
Tectiviridae ; Tectivirus (3)			*								
Cystovirus (4)						*					
Inoviridae ; Inovirus (18)			*			*					
Inoviridae ; Plectrovirus (4)			*			*					
Microviridae ; Chlamydia microvirus (5)			*			*					
Microviridae ; Microvirus (7)			*			*					
Microviridae ; Spiromicrovirus (1)			*			*					
Leviviridae ; Allovivirus (2)						*					
Leviviridae ; Levivirus (6)						*					

Arriba la clasificación taxonómica y el no. de organismos procariontes que conforman la base de datos de proteínas no virales (en verde el dominio arquea), del lado izquierdo la clasificación y el no. de bacteriófagos que forman la base de datos de proteínas virales. Las estrellas indican la presencia de fagos y sus respectivos hospederos en ambas bases de datos.



El grupo seleccionado para este trabajo conforma los virus que infectan arqueas específicamente Crenarcheota (Tabla 1S). Este grupo incluye los genomas de cinco arqueas y 10 virus. Los genomas de arqueas son: *Aeropyrum pernix*, *Sulfolobus solfataricus*, *Sulfolobus tokodaii*, *Sulfolobus acidocaldarius* y *Pyrobaculum aerophilum* y los virus son: *Acidianus filamentous* phage 1, *Sulfolobus islandicus* filamentous virus, *Sulfolobus* spindle-shaped virus 1, *Sulfolobus* spindle-shaped virus 2, *Sulfolobus* spindle shaped virus Kamchatka, *Sulfolobus* spindle-shaped virus Ragged Hills, *Sulfolobus tengchongensis* spindle shaped virus STSV1, *Sulfolobus* turreted icosahedral virus, *Sulfolobus* virus SIRV1 y *Sulfolobus* virus SIRV2. Los criterios que se tomaron en cuenta para seleccionar a este grupo de virus y arqueas fueron: (1) un grupo limitado de hospederos procariontes para facilitar la clasificación de resultados, en la base de datos hay un sesgo con los genomas de bacterias siendo los más secuenciados, y (2) en términos biológicos, las bacterias se hallan en numerosos ecosistemas mientras que muchas de las arqueas secuenciadas viven en lugares con condiciones extremas.

La ventaja de utilizar un pequeño grupo de arqueas es que permite un análisis práctico de los 'hits'. Además, poco se sabe de los virus de arqueas y su relación con estas. La información obtenida en este estudio se puede actualizar cuando halla otros genomas de arqueas secuenciados.



### 5.3 Alineamientos

Se obtuvo al menos un alineamiento significativo para 287 proteínas virales contra las proteínas de las arqueas. El 20.9% de estas proteínas pertenecen a los virus que infectan a Crenarcheota y el 79.1% son proteínas que están en los genomas de bacteriófagos que de acuerdo a la literatura no infectan arqueas. Algunos de estos bacteriófagos incluyen a 44RR2.8t, T5, Mu, P22, Mycobacteriophage Che8, *Staphylococcus* phage K, entre otros. Abajo se muestran las proteínas anotadas e hipotéticas para los 287 resultados obtenidos en relación a los fagos que infectan y no infectan arqueas (Figura 13).

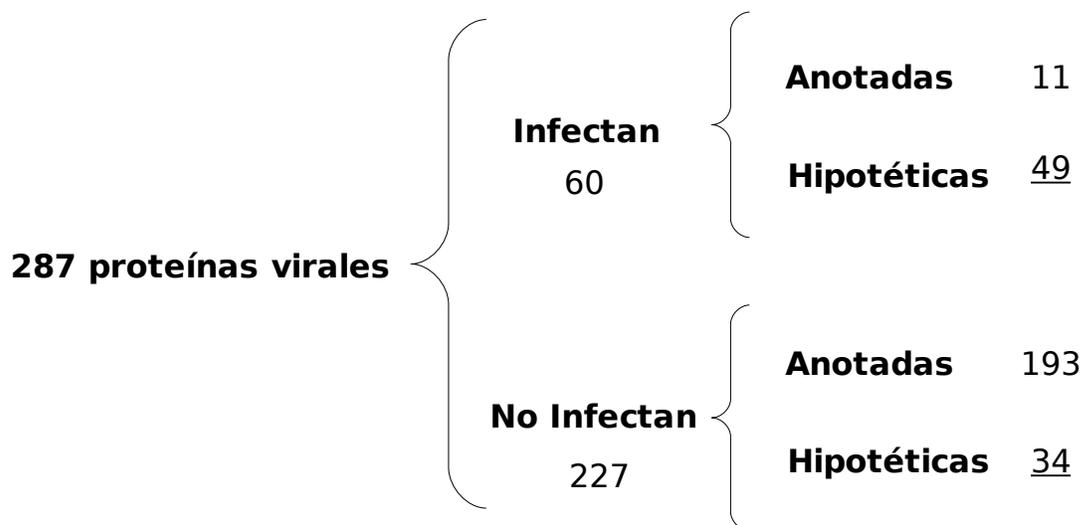


Figura 13. Relación de los 287 resultados para los virus que infectan y no infectan arqueas. Los resultados se clasificaron en proteínas con anotación e hipotéticas y los resultados de interés para este estudio están subrayados.

Dentro de las 60 proteínas pertenecientes a partículas virales que infectan arqueas nuestro grupo de interés comprende las 49 proteínas hipotéticas pues es aquí donde podremos identificar nuevas proteínas que sean producto de la relación virus-arquea.



El criterio para seleccionar tales proteínas estuvo basado en la anotación que tienen en la base de datos NCBI, e incluyen proteínas descritas como ORF, desconocida e hipotética como se menciona en la tabla 6. El grupo de 34 proteínas hipotéticas virales que pertenecen a bacteriófagos que infectan bacterias pero no arqueas nos puede brindar información acerca de posibles eventos de transferencia horizontal.

Los grupos de proteínas hipotéticas mencionados anteriormente fueron analizados detalladamente para estudiar de donde provenían. Se encontró que hubo dos tipos de alineamientos: (a) proteínas hipotéticas que alineaban con proteínas hipotéticas; y (b) proteínas hipotéticas que alineaban con proteínas ya caracterizadas. En la Tabla 8 se muestra un resumen de 38 proteínas hipotéticas virales que produjeron alineamientos significativos con otras proteínas hipotéticas. En la tabla 9 se encuentran los resultados obtenidos de los alineamientos entre las proteínas hipotéticas virales y proteínas celulares caracterizadas. Finalmente, para confirmar la posible presencia de dominios funcionales dentro de las proteínas virales se realizó una segunda búsqueda utilizando la herramienta BLAST en la base de datos pfam. Los resultados de esta búsqueda están en la Tabla 10. Ejemplos de los alineamientos múltiples se encuentran en suplementos.



Tabla 8. Resumen de las proteínas virales que produjeron alineamientos significativos con proteínas hipotéticas de arqueas.

<b>Base de datos (Anotación, ID)</b>	<b>Bacteriófago y Hospedero</b>	<b>Tamaño (aa)</b>	<b>Hits (Anotación)</b>	<b>Organismo</b>
Hipotética, YP_003733.1	Acidianus filamentus virus 1/Acidianus	59	3 Hipotéticas	Arquea (sto, sso, sai)
Hipotética, NP_046894.1	Mycobacteriophage D29/Mycobacterium tuberculosis	240	1 Hipotética	Arquea (pai)
Hipotética, NP_445671.1	Sulfolobus islandicus filamentous virus/Sulfolobus islandicus	232	4 Hipotéticas	Arqueas (sto, sai, sso)
Hipotética, NP_445692.1	Sulfolobus islandicus filamentous virus/Sulfolobus islandicus	88	1 Hipotética	Arquea (sso)
Hipotética, NP_039777.1	Sulfolobus spindle-shaped virus 1/Sulfolobus shibatae	251	2 Hipotéticas	Arquea (ape)
Hipotética, NP_039781.1	Sulfolobus spindle-shaped virus 1/Sulfolobus shibatae	244	1 Hipotética	Arquea (sai)
Hipotética, NP_039794.1	Sulfolobus spindle-shaped virus 1/Sulfolobus shibatae	102	1 Hipotética	Arquea (sto)
Hipotética, NP_039795.1	Sulfolobus spindle-shaped virus 1/Sulfolobus shibatae	129	3 Hipotéticas	Arquea (sso,sto)
Hipotética, NP_039798.1	Sulfolobus spindle-shaped virus 1/Sulfolobus shibatae	792	2 Hipotéticas	Arqueas (sso,sto)
ORF 233, NP_944455.1	Sulfolobus spindle-shaped virus 2/Sulfolobus	233	1 Hipotética	Arquea (ape)
ORF 211, NP_944461.1	Sulfolobus spindle-shaped virus 2/Sulfolobus	211	1 Hipotética	Arquea (sai)
ORF 100, NP_944473.1	Sulfolobus spindle-shaped virus 2/Sulfolobus	100	1 Hipotética	Arquea (sto)
ORF 155, NP_944475.1	Sulfolobus spindle-shaped virus 2/Sulfolobus	155	3 Hipotéticas	Arqueas (sso, sto)
ORF 305, NP_944476.1	Sulfolobus spindle-shaped virus 2/Sulfolobus	305	1 Hipotética	Arquea (sso)
ORF 809, NP_944478.1	Sulfolobus spindle-shaped virus 2/Sulfolobus	809	2 Hipotéticas	Arquea (sso, sto)



Tabla 8. Resumen de las proteínas virales que produjeron alineamientos significativos con proteínas hipotéticas de arqueas.

<b>Base de Datos (Anotación, ID)</b>	<b>Bacteriófago y Hospedero</b>	<b>Tamaño (aa)</b>	<b>Hits (Anotación)</b>	<b>Organismo</b>
ORF A231, NP_963972.1	Sulfolobus spindle-shaped virus Kamchatka-1 / Sulfolobus sp.	231	1 Hipotética	Arquea (ape)
ORF A231, NP_963972.1	Sulfolobus spindle-shaped virus Kamchatka-1 / Sulfolobus sp.	231	1 Hipotética	Arquea (ape)
ORF A460, NP_963977.1	Sulfolobus spindle-shaped virus Kamchatka-1 / Sulfolobus sp.	460	1 Hipotética	Arquea (sai)
ORF B98, NP_963986.1	Sulfolobus spindle-shaped virus Kamchatka-1 / Sulfolobus sp.	98	1 Hipotética	Arquea (sto)
ORF B158, NP_963988.1	Sulfolobus spindle-shaped virus Kamchatka-1 / Sulfolobus sp.	158	3 Hipotéticas	Arquea (sso, sto)
ORF B111, NP_963989.1	Sulfolobus spindle-shaped virus Kamchatka-1 / Sulfolobus sp.	111	1 Hipotética	Arquea (sto)
ORF B252, NP_963990.1	Sulfolobus spindle-shaped virus Kamchatka-1 / Sulfolobus sp.	252	3 Hipotéticas	Arqueas (sso, sto)
ORF C108, NP_963991.1	Sulfolobus spindle-shaped virus Kamchatka-1 / Sulfolobus sp.	108	2 Hipotéticas	Arqueas (sso, sto)
ORF B793, NP_963992.1	Sulfolobus spindle-shaped virus Kamchatka-1 / Sulfolobus sp.	793	2 Hipotéticas	Arqueas (sso, sto)
ORF A247, NP_963931.1	Sulfolobus spindle-shaped virus Ragged Hills / Sulfolobus sp.	247	1 Hipotética	Arquea (ape)
ORF D212, NP_963934.1	Sulfolobus spindle-shaped virus Ragged Hills / Sulfolobus sp.	212	1 Hipotética	Arquea (sai)
ORF A102b, NP_963949.1	Sulfolobus spindle-shaped virus Ragged Hills / Sulfolobus sp.	102	1 Hipotética	Arquea (sto)



Tabla 8. Resumen de las proteínas virales que produjeron alineamientos significativos con proteínas hipotéticas de arqueas.

<b>Base de Datos (Anotación)</b>	<b>Bacteriófago y Hospedero</b>	<b>Tamaño (aa)</b>	<b>Hits (Anotación)</b>	<b>Organismo</b>
ORF C150, NP_963952.1	Sulfolobus spindle-shaped virus Ragged Hills / Sulfolobus sp.	150	3 Hipotéticas	Arquea (sso, sto)
ORF C113, NP_963953.1	Sulfolobus spindle-shaped virus Ragged Hills / Sulfolobus sp.	113	2 Hipotéticas	Arquea (sto)
ORF C247, NP_963954.1	Sulfolobus spindle-shaped virus Ragged Hills / Sulfolobus sp.	247	1 Hipotética	Arquea (sso)
ORF B812, NP_963957.1	Sulfolobus spindle-shaped virus Ragged Hills / Sulfolobus sp.	812	2 Hipotéticas	Arqueas (sso, sto)
Hipotética, YP_077211.1	Sulfolobus tengchongensis spindle-shaped virus STSV1/Sulfolobus tengchongensis	266	3 Hipotéticas	Arqueas (sso, sto, pai)
Hipotética, YP_024996.1	Sulfolobus turreted icosahedral virus/Sulfolobus solfataricus	61	3 Hipotéticas	Arqueas (sto,sso, sai)
Hipotética, NP_666594.1	Sulfolobus virus SIRV-1/Sulfolobus islandicus	306	3 Hipotéticas	Arqueas (sto, sso, pai)
Hipotética, NP_666595.1	Sulfolobus virus SIRV-1/Sulfolobus islandicus	399	1 Hipotética	Arquea (sto)
Hipotética, NP_666602.1	Sulfolobus virus SIRV-1/Sulfolobus islandicus	59	2 Hipotéticas	Arquea (sso, sai)
Hipotética, NP_666547.1	Sulfolobus virus SIRV-2/Sulfolobus islandicus	310	3 Hipotéticas	Arqueas (sto, sso, pai)
Hipotética, NP_666548.1	Sulfolobus virus SIRV-2/Sulfolobus islandicus	399	1 Hipotética	Arquea (sto)
Hipotética, NP_666555.1	Sulfolobus virus SIRV-2/Sulfolobus islandicus	59	2 Hipotéticas	Arquea (sso, sai)

Claves: *P. aerophilum* (pai), *S. tokodaii* (sto), *S. acidocaldarius* (sai), *S. solfataricus* (sso) y *A. pernix* (ape). En gris la única proteína viral que está codificada en el genoma de un fago que no infecta arqueas.



Tabla 9. Resumen de las proteínas hipotéticas virales que produjeron alineamientos significativos con proteínas de eucariontes, bacterias y arqueas.

<b>Proteína Viral Anotación, ID Bacteriófago Hospedero</b>	<b>Tamaño (aa)</b>	<b>No. de Hits, Anotación Organismos donde se encontraron</b>
Hipotética, NP_932363.1 Bacteriophage 44RR2.8t <i>Aeromonas salmonicida</i>	389	554 Glicina/Serina hidroximetiltransferasa y aminotransferasa clase I y II. Eucariontes, Arqueas y Bacterias.
Hipotética, NP_943996.1 Bacteriophage Aeh1 <i>Aeromonas hydrophila</i>	309	40 UV damage endonuclease. Eucariontes, Arqueas y Bacterias.
Desconocida, NP_945015.1 Bacteriophage Felix 01 <i>Salmonella</i>	298	427 Fosforibosil transferasa. Eucariontes, Arqueas y Bacterias.
Desconocida, NP_945016.1 Bacteriophage Felix 01 <i>Salmonella</i>	293	428 Fosforibosilpirofosfato sintetasa. Eucariontes, Arqueas y Bacterias.
Desconocida, NP_570363.1 Cyanophage P60 <i>Synechococcus sp. WH7803</i>	235	78 Timidilato sintasa dependiente de flavina. Eucariontes, Arqueas y Bacterias.



Tabla 9. Resumen de las proteínas hipotéticas virales que produjeron alineamientos significativos con proteínas de eucariontes, bacterias y arqueas.

<b>Proteína Viral Anotación, ID Bacteriófago Hospedero</b>	<b>Tamaño (aa)</b>	<b>No. de Hits, Anotación Organismos donde se encontraron</b>
Hipotética, NP_861740.1 Enterobacteria phage RB69 <i>Escherichia coli</i>	389	7 Proteasa, proteasa de la familia colagenasa. Arqueas (sso, sto, mka, mbu, mja)
Desconocida, NP_046960.1 Methanobacterium phage psiM2 <i>Methanobacterium thermoautotrophicum</i>	127	18 Hipotética, regulador transcripcional de la familia Cro/CI. Bacterias y Arqueas (ape, mmp).
Desconocida, NP_071814.1 Methanothermobacter wolfeii prophage psiM100 <i>Methanothermobacter wolfeii</i>	129	19 Hipotética, regulador transcripcional de la familia Cro/CI. Bacterias y Arqueas (mmp, pai, ape).
Hipotética, NP_438138.1 Streptococcus pyogenes phage 315.4 <i>Streptococcus pyogenes MGAS315</i>	125	19 Hipotética, regulador transcripcional de la familia Cro/CI y proteína ribosomal 50s. Bacterias y Arqueas (mmp, ape).



Tabla 9. Resumen de las proteínas hipotéticas virales que produjeron alineamientos significativos con proteínas de eucariontes, bacterias y arqueas.

<b>Proteína Viral Anotación, ID Bacteriófago Hospedero</b>	<b>Tamaño (aa)</b>	<b>No. de Hits, Anotación Organismos donde se encontraron</b>
ORF 328, NP_944456.1 Sulfolobus spindle-shaped virus 2 <i>Sulfolobus</i>	328	7 Fragmento C-terminal de integrasa e hipotética. Arqueas (sso, tko, pho)
ORF F340, NP_963973.1 Sulfolobus spindle-shaped virus Kamchatka-1 <i>Sulfolobus sp.</i>	340	8 Fragmento homólogo a la integrasa SSV1 e hipotética. Arqueas (sso, pho, tko, sto).
ORF B494, NP_963976.1 Sulfolobus spindle-shaped virus Kamchatka-1 <i>Sulfolobus sp.</i>	494	10 DNA helicase, DNA repair helicase, ATP-dependent helicase. Arqueas (sso, sto, sai, pai, pto, ape, mst, tac) y Eucariontes (ddpo)
ORF B137, NP_963996.1 Sulfolobus spindle-shaped virus Kamchatka-1 <i>Sulfolobus sp.</i>	137	1 Proteína estructural hipotética. Arquea (sto)



Tabla 9. Resumen de las proteínas hipotéticas virales que produjeron alineamientos significativos con proteínas de eucariontes, bacterias y arqueas.

<b>Proteína Viral Anotación, ID Bacteriófago Hospedero</b>	<b>Tamaño (aa)</b>	<b>No. de Hits, Anotación Organismos donde se encontraron</b>
ORF D355, NP_963933.1 Sulfolobus spindle-shaped virus Ragged Hills <i>Sulfolobus sp.</i>	355	10 Fragmento homólogo a la integrasa SSV1 e hipotética. Arqueas (sso, sto, tko, pho).
Hipotética, YP_077212.1 Sulfolobus tengchongensis spindle-shaped virus STSV1 <i>Sulfolobus tengchongensis</i>	187	8 Hipotética y archaeosine/queuine tRNA ribosyltransferasa. Arqueas (sso, sto, pto, tvo, tac, neq, mhu, mba)
Hipotética, YP_077221.1 Sulfolobus tengchongensis spindle shaped virus STSV1 <i>Sulfolobus tengchongensis</i>	468	6 Proteína involucrada en la biosíntesis de polisacáridos e hipotética. Arqueas (sso, sto, sai, tac)
Hipotética, YP_025023.1 Sulfolobus turreted icosahedral virus <i>Sulfolobus solfataricus</i>	557	2 Hipotética y Autotransportador externo de membrana. Arquea (sto) y Betaproteobacteria (bxo).
Hipotética, NP_666569.1 Sulfolobus virus SIRV-2 <i>Sulfolobus islandicus</i>	121	2 Resolvasa Holliday-junction. Arqueas (sso, sai)



Tabla 9. Resumen de las proteínas hipotéticas virales que produjeron alineamientos significativos con proteínas de eucariontes, bacterias y arqueas.

<b>Proteína Viral Anotación, ID Bacteriófago Hospedero</b>	<b>Tamaño (aa)</b>	<b>No. de Hits, Anotación Organismos donde se encontraron</b>
Hipotética, NP_666569.1 Sulfolobus virus SIRV-2 <i>Sulfolobus islandicus</i>	121	2 Resolvasa Holliday-junction. Arqueas (sso, sai)

En amarillo se muestran las proteínas que provienen de arqueas. Claves: *P. aerophilum* (pai), *S. tokodaii* (sto), *S. acidocaldarius* (sai), *S. solfataricus* (sso), *A. pernix* (ape), *Methanopyrus kandleri* (mka), *Methanococcoides burtoni* (mbu), *Methanococcus jannaschii* (mja), *Methanococcus maripaludis* (mmp), *Pyrococcus horikoshi* (pho), *Thermococcus kodakaraensis* (tko), *Picrophilus torridus* (pto), *Methanosphaera stadtman* (mst), *Thermoplasma acidophilum* (tac), *Thermoplasma volcanium* (tvo), *Nanoarchaeum equitans* (neq), *Methanospirillum hungatei* (mhu), *Methanosarcina barkeri* (mba), *Drosophila pseudoobscura* (ddpo) y *Burkholderia xenovorans* (bx).



Tabla 10. Resumen de los dominios funcionales identificados para las proteínas hipotéticas virales.

Proteína viral (ID)	Proteína/Dominio	Vía metabólica/Proceso Biológico
<p>Hipotética (NP_932363.1)</p>	 <p>Serina-hidroximetiltransferase SHMT lleva la conversión de serina y glicina utilizando PLP como cofactor.</p> <p>Reacción: 5,10-metilenotetrahidrofolato + glicina + H<sub>2</sub>O = tetrahidrofolato y L-serina</p>	<p>Metabolismo de glicina, serina y treonina, degradación de lisina, metabolismo de cianoaminoácido, 'one carbon pool by folate' y metabolismo de metano.</p>
<p>Hipotética (NP_943996.1)</p>	 <p>Endonucleasa UV UvdE Endonucleasa que está involucrada en el primer paso dentro de una vía alternativa para reparar DNA dañado por radiación UV.</p>	<p>Reparación de nucleótidos por excisión.</p>
<p>Desconocida (NP_945015.1)</p> <p>Desconocida (NP_945016.1)</p>	 <p>Ribosafosfato difosfoquinasa/Dominio fosforibosil transferasa Este dominio está presente en las enzimas fosforibosiltransferasas (PRTase) que llevan a cabo la reacción de transferencia de un grupo fosfato a partir de PRPP.</p> <p>Reacción: ATP+ D-ribosa 5 fosfato = AMP + 5-fosfo-α-D-ribose 1-difosfato (PRPP)</p>	<p>Síntesis de nucleótidos, vía de salvamento, metabolismo de purinas y vía de las pentosas fosfato.</p>



Tabla 10. Resumen de los dominios funcionales identificados para las proteínas hipotéticas virales.

Proteína viral (ID)	Proteína/Dominio	Vía metabólica/Proceso Biológico
Desconocida (NP_570363.1)	<p style="text-align: center;"><b>Thy1</b></p> <p style="text-align: center;">Timidilato sintasa (FAD)</p> <p>Thy1 es una proteína que complementa los niveles de timidina necesarios durante el crecimiento en los organismos donde se encuentra. No presenta homología con timidilato sintasa.</p> <p>Reacción:  <span style="color: red;">5,10-metilenotetrahidrofolato + dUMP + FADH<sub>2</sub> = dTMP + tetrahidrofolato + FAD</span></p>	Metabolismo de purinas, one carbon pool by folate.
Hipotética (NP_861740.1)	<p style="text-align: center;"><b>Peptidase_U32</b></p> <p>Esta proteína pertenece a la familia de las peptidasas U32, un ejemplo de estas es la colagenasa que degrada colágeno tipo I.</p>	Degradación de lisina y metabolismo de biotina.
Desconocida (NP_046960.1)  Desconocida (NP_071814.1)  Desconocida (NP_438138.1)	<p style="text-align: center;"><b>ASCH</b></p> <p style="text-align: center;">Dominio ASCH</p> <p>Adopta una estructura de barril beta. Se piensa que funciona como un sitio de unión a RNA durante la coactivación, procesamiento de RNA y posiblemente durante la regulación traduccional en procariontes.</p>	Desconocida



Tabla 10. Resumen de los dominios funcionales identificados para las proteínas hipotéticas virales.

Proteína viral (aa)	Proteína/Dominio	Vía metabólica/Proceso Biológico
<p>ORF B494 (NP_963976.1)</p>	 <p>Dominio DEAD 2 Región conservada dentro de las helicasas de unión a DNA parecidas a RAD3. Se encuentran en procariontes, eucariontes y arqueas. RAD3 está involucrada en la reparación de nucleótidos por excisión y forma parte del factor de transcripción TFIIH en levadura.</p>	<p>Funciones incluyen unión a DNA y ATP, actividad de helicasa de DNA dependiente de ATP.</p>
<p>ORF B137 (NP_963996.1)</p>	 <p>Proteína de función desconocida Pertenece a una familia de varias proteínas presentes en <i>Sulfolobus</i> y en virus que infectan a estas arqueas, se desconoce su función.</p>	<p>Desconocida</p>
<p>Hipotética (YP_077201.1)</p>	 <p>Proteína hipotética presente en arqueas. Esta familia comprende proteínas hipotéticas de arqueas, son de aprox. 120 aa, y parecen ser específicos a <i>Sulfolobus</i> spp. Se desconoce su función.</p>	<p>Desconocida</p>
<p>Hipotética (YP_077221.1)</p>	 <p>Proteína involucrada en la biosíntesis de polisacáridos.</p>	<p>Miembros de esta familia son proteínas integrales de membrana, la mayoría involucradas en la producción de polisacáridos.</p>



Tabla 10. Resumen de los dominios funcionales identificados para las proteínas hipotéticas virales.

Proteína viral (aa)	Proteína/Dominio	Vía metabólica/Proceso Biológico
Hipotética (NP_666569.1)	 Resolvasa Holliday junction (hjc)	Pertenece a una familia de resolvasas presentes en arqueas y bacterias. Esta proteína es equivalente a RuvC pero su secuencia no es similar.
Hipotética* (YP_025023.1)	 Dominio Beta de un autotransportador (rojo) y pertactina (verde).	<p>Dominio Beta responsable de la secreción de la proteína que lo presente.</p> <p>Pertactina (P69) es un factor de virulencia de <i>Bordetella pertussis</i>, ha sido identificado como una adhesina.</p>

\* En la secuencia de esta proteína solo se encontró el dominio de pertactina, el dominio beta no está presente.



## 6. DISCUSIÓN

El estudio de virus que infectan arqueas es de gran interés porque sirven como modelos experimentales para estudiar la bioquímica y biología molecular de organismos que habitan en ambientes extremos. Es importante recordar que el grupo analizado en este estudio fueron los virus que infectan Crenarchaeota, y son los que presentan morfologías diferentes a los virus descritos para Euryarchaeota y el resto de los bacteriófagos. Abajo se muestran los virus seleccionados para este trabajo y de qué organismo fueron originalmente aislados (Tabla 11).

Tabla 11. Virus que infectan Crenarchaeota utilizados en este estudio.

<b>Virus</b>	<b>Hospedero</b>
<i>Acidianus filamentous phage 1</i>	<i>Acidianus</i> (32)
<i>Sulfolobus islandicus filamentous virus</i>	<i>Sulfolobus islandicus</i> (33)
<i>Sulfolobus spindle-shaped virus 1</i>	<i>Sulfolobus shibatae</i> (34)
<i>Sulfolobus spindle-shaped virus 2</i>	<i>Sulfolobus</i> (34)
<i>Sulfolobus tengchongensis spindle shaped virus STSV1</i>	<i>Sulfolobus tengchongensis</i> (35)
<i>Sulfolobus turreted icosahedral virus</i>	<i>Sulfolobus solfataricus</i> (36)
SIRV 1	<i>Sulfolobus islandicus</i> (37)
SIRV 2	<i>Sulfolobus islandicus</i> (37)
<i>Sulfolobus spindle-shaped virus Kamchatka</i>	<i>Sulfolobus</i> sp. (38)
<i>Sulfolobus spindle-shaped virus Ragged Hills</i>	<i>Sulfolobus</i> sp. (38)

Como se muestra en la tabla anterior, todos los virus a excepción de *Acidianus filamentous phage 1* fueron aislados de *Sulfolobus* sp. *S. solfataricus* es una de las arqueas mejor caracterizadas. *Sulfolobus* sp. son acidófilos aerobios que crecen a temperaturas de 70°C-87°C y con un pH de 1.5-5.5, y han sido aislados en varios lugares como el parque nacional Yellowstone (USA), Japón, Islandia, Nueva Zelanda, El Salvador, República Dominicana e Italia (22).



Este grupo de virus es atractivo por sus características biológicas ya que su ciclo de propagación aparenta ser lisogénico sin comprometer al hospedero. Además no parecen ser tan ubicuos como los bacteriófagos descritos hasta ahora que infectan bacterias, pues sus hospederos están en lugares que requieren de altas temperaturas para sobrevivir.

Con los resultados obtenidos en este trabajo se identificaron dos grupos de proteínas dentro del grupo de virus que infectan Crenarchaeota: (a) un grupo de proteínas hipotéticas que están en genomas virales y que tienen proteínas homólogas en genomas de arqueas exclusivamente, y (b) un conjunto de proteínas hipotéticas que están codificadas en los genomas virales y que tienen homólogos en bacterias, arqueas y eucariontes. Como se aprecia en la Tabla 8 para el grupo (a) solo hay una proteína viral hipotética, que está codificada en el genoma de Mycobacteriophage D29, y que tuvo un 'hit' que corresponde a una proteína en el genoma de la arquea *P. aerophilum*. Este resultado fue inesperado ya que se puede predecir que una proteína del Mycobacteriophage D29 tenga homólogos en su hospedero, *M. tuberculosis* o en otras bacterias, pero no en una arquea. Habría que analizar con más detalle este caso y rastrear esta proteína en genomas bacterianos usando otro valor de corte. Independiente al caso mencionado antes, con el resto de las proteínas hipotéticas cuya función se desconoce experimentalmente, se puede sugerir que representan un 'pool' de genes exclusivos para arqueas y que pueden ser la materia prima para generar otros genes. Daubin et al (39) sugieren que las interacciones moleculares entre los fagos y sus hospederos componen un mecanismo probable para generar nuevos genes. Aquí no se puede asegurar que estas proteínas provengan de virus sino que estos al insertarse en el cromosoma de sus hospederos tomaron genes que ahora forman parte de su genoma. Sería interesante analizar la posición de estos genes en los genomas de *Sulfolobus* spp. para saber si pertenecen a profagos o si son secuencias que posiblemente funcionen como materia prima.



En el caso del grupo (b) se encontró la siguiente relación de 'hits' (Tabla 16). A cinco proteínas virales hipotéticas se les encontraron homólogos en los tres dominios de la vida, estas proteínas están codificadas en los genomas de bacteriófagos que no infectan arqueas. Cuatro parecen estar involucradas en la síntesis de nucleótidos (hidroximetiltransferasa, fosforibosil transferasa, fosforibosilpirofosfato sintetasa, y timidilato sintasa dependiente de flavina), y una en la reparación de DNA dañado por radiación UV. Estas funciones están conservadas en los tres dominios y son fundamentales para la vida (40).

Esta información nos indica que es probable que estas proteínas están siendo transportadas por virus, y sería interesante evaluar si han adoptado funciones involucradas en la biología de estos como ha ocurrido con las glicosiltransferasas (41), y en el caso de P-SSP7, P-SSM2 y P-SSM4 que infectan cianobacterias se encontró que los tres tienen genes fotosintéticos que se presumen son importantes para mantener activos los centros de reacción de la fotosíntesis y así asegurar el flujo de energía durante la infección (42). Se puede sugerir que los genes identificados en este estudio y que parecen estar involucrados en la síntesis de nucleótidos han sido adquiridos por los bacteriófagos para cumplir con la demanda de nucleótidos dentro del hospedero durante la infección.

Por otro lado, cuatro proteínas virales hipotéticas presentaron homólogos en bacterias y arqueas, y estas últimas se encuentran en genomas de bacteriófagos que infectan bacterias (*Streptococcus*) y arqueas (*Methanobacterium*, *Methanothermobacter* y *Sulfolobus*). En este caso, se encontró que tres son homólogas a reguladores transcripcionales de la familia Cro/C1. No es sorprendente encontrar proteínas relacionadas a reguladores transcripcionales ya que los virus de arqueas parecen tener un ciclo lisogénico, la regulación en la producción de partículas virales debe estar a cargo de represores o activadores como en el caso de los bacteriófagos.



La identificación de estas proteínas ayudará a diseñar experimentos para evaluar su actividad y estudiar la regulación transcripcional en sus hospederos.

Tabla 12. Relación entre las proteínas hipotéticas virales y la presencia de sus homólogos en los tres dominios.

	Arqueas	Eucariontes	Bacterias
Bacteriophage 44RR2.8t/ <i>A. salmonicida</i>	*	*	*
Bacteriophage Aeh1/ <i>A. hydrophila</i>	*	*	*
Bacteriophage Felix 01/ <i>Salmonella</i>	*	*	*
Bacteriophage Felix 01/ <i>Salmonella</i>	*	*	*
Cyanophage P60/ <i>Synechococcus</i> sp. WH7803	*	*	*
Enterobacteria phage RB69/ <i>E. coli</i>	*		
<i>Methanobacterium</i> phage psiM2/ <i>M. thermoautotrophicum</i>	*		*
<i>Methanothermobacter wolfeii</i> prophage psiM100/ <i>M. wolfeii</i>	*		*
<i>Streptococcus pyogenes</i> phage 315.4/ <i>S. pyogenes</i> MGAS315	*		*
<i>Sulfolobus</i> spindle-shaped virus 2/ <i>Sulfolobus</i>	*		
<i>Sulfolobus</i> spindle-shaped virus Kamchatka-1/ <i>Sulfolobus</i> sp.	*		
<i>Sulfolobus</i> spindle-shaped virus Kamchatka-1/ <i>Sulfolobus</i> sp.	*	*	
<i>Sulfolobus</i> spindle-shaped virus Kamchatka-1/ <i>Sulfolobus</i> sp.	*		
<i>Sulfolobus</i> spindle-shaped virus Ragged Hills/ <i>Sulfolobus</i> sp.	*		
<i>Sulfolobus tengchongensis</i> STSV1/ <i>S. tengchongensis</i>	*		
<i>Sulfolobus tengchongensis</i> STSV1/ <i>S. tengchongensis</i>	*		
<i>Sulfolobus</i> turreted icosahedral virus/ <i>S. solfataricus</i>	*		*
<i>Sulfolobus</i> virus SIRV-2/ <i>S. islandicus</i>	*		

Finalmente, se encontraron ocho proteínas virales cuyos homólogos fueron hallados en genomas de arqueas. Todas las proteínas, a excepción de un caso, están en virus que infectan arqueas.



Las funciones que se identificaron incluyen tres integrasas, tRNA ribosiltransferasa, una proteína involucrada en la biosíntesis de polisacáridos, resolvasa Holliday junction, colagenasa, y una proteína estructural. Asimismo, se encontró una proteína viral que presentó homólogos en eucariontes y arqueas, posiblemente se trata de una helicasa.

La distribución de estos resultados refleja los diferentes eventos de transferencia horizontal que han sostenido estos organismos y el transporte de genes a través de los bacteriófagos. Para corroborar estos resultados, se buscó que partes de las secuencias presentan homología. En todas las secuencias se encontraron dominios homólogos en la base de datos Pfam.

Para rastrear la historia evolutiva de cada gen tendríamos que contar con más secuencias de virus para hacer un análisis comparativo. En un trabajo realizado por Koonin, Prangishvili y Garrett se analizaron secuencias de proteínas virales utilizando PSI-BLAST y buscaron dominios conservados. Ellos utilizaron una base de datos viral diferente a la empleada aquí y encontraron 15 proteínas que son compartidas por diferentes virus que infectan Crenarchaeota. Asimismo, delimitaron un grupo de genes que tiene homólogos en arqueas y bacterias. Algunas de las funciones predichas para estas proteínas incluyen: DNA metilasas, proteína de unión a RNA, acetiltransferasa, helicasa, resolvasa, epimerasa, permeasa y nucleasa (43). Comparando los resultados entre el estudio de Koonin et al y este se encontró una proteína en común (NP\_666548.1). Koonin et al sugieren que esta proteína puede ser una permeasa y aquí se encontró que es hipotética y no tiene homólogos. Esto puede ser reflejo de la base de datos con la cual compararon las proteínas virales.



En cuanto a las funciones biológicas predichas en este estudio, la mayoría están involucradas en la síntesis de nucleótidos y conversión de aminoácidos aunque hay casos interesantes como por ejemplo, la proteína involucrada en la biosíntesis de polisacáridos y la colagenasa.

Con respecto a las funciones biológicas representadas en la base de datos, se encontró que la mitad de las proteínas virales pertenecientes a bacteriófagos no están caracterizadas y están asignadas como hipotéticas, ORFs o desconocidas. En un estudio realizado por Liu et al agruparon familias de proteínas de bacteriófagos en lo que ellos denominaron POG (Phage Orthologous Groups) y utilizaron bacteriófagos cuyo genoma es de doble cadena de DNA (44). Las categorías funcionales que asignan a sus grupos incluyen transcripción, replicación, división celular, transporte de nucleótidos, lisis entre otras. Estas funciones no son diferentes a las encontradas en este estudio pese a que se incluyeron todos los tipos de fagos. Esto refleja un sesgo en los genomas secuenciados de bacteriófagos ya que la mayoría son de doble cadena de DNA



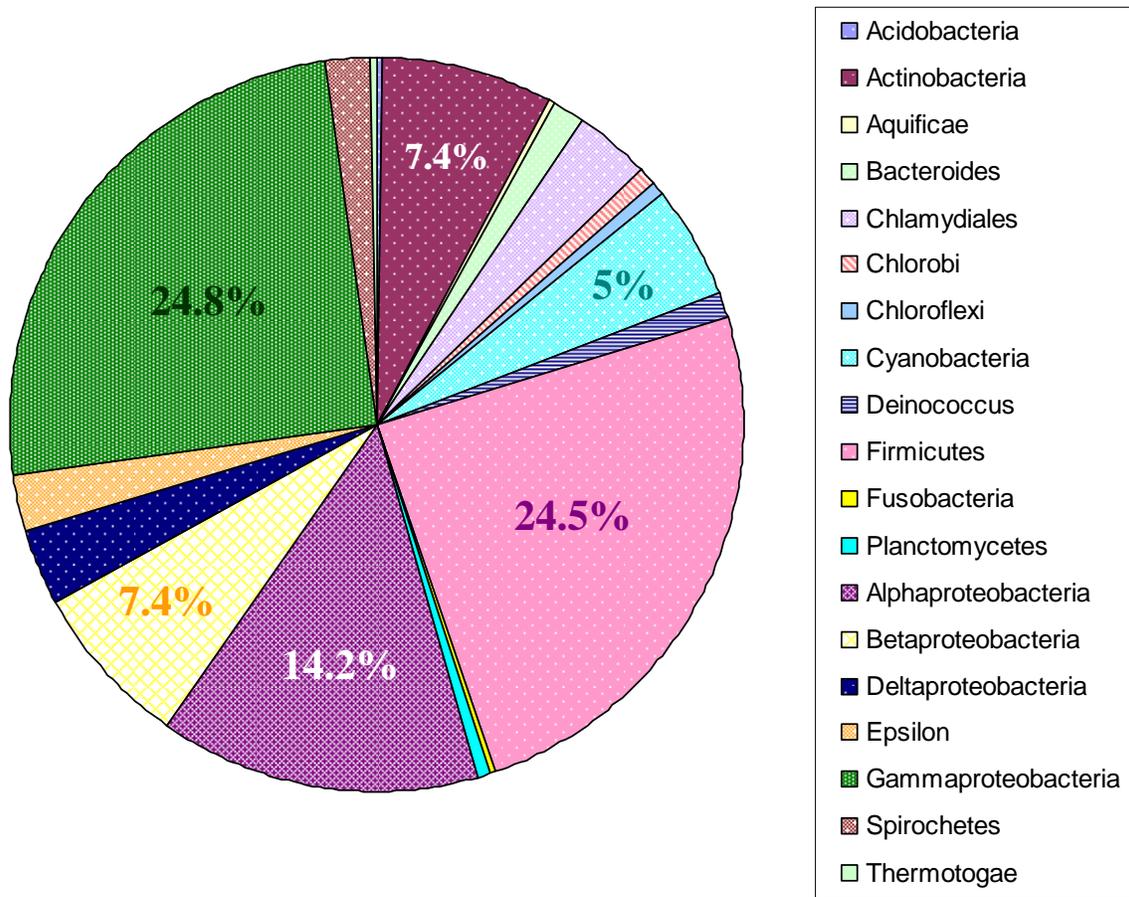
## 7. CONCLUSIONES

La mayoría de los genomas secuenciados de bacteriófagos son aquellos cuyos genomas son de doble cadena de DNA, y tienen como hospederos a firmicutes y proteobacterias. Esto representó un sesgo al momento de crear las bases de datos y analizar los resultados. El análisis de los resultados obtenidos para *Sulfolobus* spp. permitió identificar dos grupos de proteínas y sus posibles dominios funcionales. El primer grupo consistió de proteínas homólogas presentes solo en *Sulfolobus* y los virus que lo infectan. Como no se les encontró algún dominio funcional, se sugiere que dentro de la dinámica evolutiva virus-arquea estas secuencias puedan funcionar como materia prima para generar nuevos genes. Finalmente, el otro grupo de proteínas virales tuvo homólogos en los tres dominios de la vida, y las funciones predichas para estas son fundamentales para el sostenimiento de la misma. Todo esto sugiere que los virus que infectan *Sulfolobus* mantienen genes dentro de sus genomas que codifiquen proteínas cuyas funciones ayuden a mantener al hospedero vivo y no comprometan la producción de partículas virales en ambientes extremos.

Hasta ahora estos virus parecen ser únicos en su morfología y características biológicas, y poco se sabe de su interacción evolutiva con sus hospederos. Una manera de abordar su estudio es a través de la bioinformática y la genómica comparativa. Los genomas hasta ahora secuenciados no representan significativamente su diversidad biológica, sin embargo con la información obtenida aquí, se pueden diseñar modelos experimentales que permitan vislumbrar las funciones que estos posibles dominios tengan dentro del hospedero y del ciclo biológico viral. Este estudio permitirá definir con más detalle la metodología empleada para esclarecer las relaciones evolutivas virus-hospedero.



## 8. SUPLEMENTOS



- **Figura 1S.** Los 339 genomas bacterianos clasificados de acuerdo a su taxonomía y sus porcentajes. Actinobacteria (7.4%), Cyanobacteria (5%), Firmicutes (24.5%), Alphaproteobacteria (14.2%), Betaproteobacteria (7.4%) y Gammaproteobacteria (24.8%).



- **Tabla 1S.** Clasificación taxonómica y número de genomas de arqueas utilizadas para este estudio.

	<b>Clasificación</b>	<b>No. de Genomas</b>
Crenarchaeota; Thermoprotei	Desulfurococcales	1
	Sulfolobales	3
	Thermoproteales	1
Euryarchaeota	Archaeoglobi; Archaeoglobales	1
	Halobacteria; Halobacteriales	4
	Methanobacteria; Methanobacteriales	2
	Methanococci; Methanococcales	2
	Methanomicrobia; Methanomicrobiales	1
	Methanomicrobia; Methanosarcinales	4
	Methanopyri; Methanopyrales	1
	Thermococci; Thermococcales	4
Thermoplasmata; Thermoplasmatales	3	
Nanoarchaeota	Nanoarchaeum	1
<b>TOTAL</b>		<b>28</b>

- **Tabla 2S.** Clasificación taxonómica y número de genomas de eucariontes utilizados en este estudio.

	<b>Clasificación</b>	<b>No. de Genomas</b>
Alveolata; Apicomplexa	Coccidia	2
	Haemosporida	3
	Piroplasmida	2
Entamoebidae	Entamoeba	1
Euglenozoa	Kinetoplastida; Trypanosomatidae	3
Fungi	Ascomycota; Pezizomycotina	5
	Ascomycota; Saccharomycotina	11
	Ascomycota; Schizosaccharomycetes	1
	Basidiomycota; Hymenomycetes	2
	Microsporidia; Unikaryonidae	1
Metazoa	Arthropoda; Hexapoda	5
	Chordata; Craniata	13
	Chordata; Urochordata	1
	Nematoda; Chromadorea	2
Mycetozoa	Dictyosteliida; Dictyostelium	1
Rhodophyta	Bangiophyceae; Cyanidiales	1
Viridiplantae	Streptophyta; Embryophyta	2
<b>TOTAL</b>		<b>56</b>

```

          :*::*::  * : * *::*:*  . : * :  :: *  :: . :  : .  . * . .
ljo_LJ0305 -----MKILLSIKPKYVSSIMNGTKKFEFR-RKIFKR-KDVTIVVVYATKPIGKVVGEFEIKQVISTPKSVWNMTAQA-GLDKLDFYNYFEGLD 88
plu_plu4361 -----MKVLLSIKPEYVDRILDGSKKFEFR-KIAFKN-NQVQSVVIYATMPIGMIVGEFEIKEIISNSPSVWVEMTHKFA-GTTKDFFDYFEGRE 88
vir_8689 -----MKVLLSVKPKYVEKIMSGDKRYEFR-KTIWKK--KINEVYIYSTTPEKKIVASFTYNKVIKEDPQLWFEFYHES-GLTQGEFFDYFRGIK 87
mmp_MMP0784 -----MTILLSIKPKYVKEILNGSKKYEFR-KSIFKKYDKNELVFIYSSYPVKRIVGTFSVGDIIENCPKILWNEFKNVS-GIKESFFKYFKEKD 89
pai_PAE0017 -----MSIKPKFGEEILKGLKKYELRRLVG-PLVEPGDLLYLFTKPAVIGRFTAGIVFLVFPAGSISRLLGELGVDVGIGAEDLEYVKGAK 86
ape_APE1779 MGAVLDEDEILYLMSIRPOYARAIMAGRKKYELRRIHGVPPIEESIIIVYASGNVKSIIGEFOAKRVIQATPEKIWSIASKPG-SGVGEDAWQYIRGAK 99
ruler 1.....10.....20.....30.....40.....50.....60.....70.....80.....90.....100

```



```

          . * :  : .  . * * :  : .
ljo_LJ0305 EAFAIQISKVTPYKKQLDLSLELDGN--LKAPOSFIYVING----- 126
plu_plu4361 KAVAIQIGNVKKYDKPLPLNMLGQG--ITAPQSYRYSLSA----- 125
vir_8689 KGYAIPIKELKPFDPPELS-LN-N--IRPPQSFYRYS--EEQVSHLI----- 129
mmp_MMP0784 KGYAIQINKLNEFKNPVNPKDINPD--FRAPOSFCYISQNFEEELFKI----- 136
pai_PAE0017 YAMLIQVKNPEACTAPVRLSDA----GLRPPPSYRRLKGEG---AEALSRLCVRS--- 134
ape_APE1779 RAIAIEVGFRRLYERPVTLEETIRRIIPGWNPPFSYTTILEPGNPIYELLIKRLRRRQHQL 158
ruler .....110.....120.....130.....140.....150.....

```



### CLAVES

ljo\_LJ0305: *Lactobacillus johnsonii* L50305 Lj965 prophage protein  
 plu\_plu4361: *Photorhabdus luminescens* hypothetical protein  
 vir\_8689: *Methanothermobacter wolfeii* prophage psiM100-unknown protein\_idNP\_071814.1  
 mmp\_MMP0784: *Methanococcus maripaludis* S2 hypothetical protein  
 pai\_PAE0017: *Pyrobaculum aerophilum* hypothetical protein  
 ape\_APE1779: *Aeropyrum pernix* hypothetical protein  
 NOTA: Los siguientes alineamientos siguen el mismo patrón de colores para los aminoácidos.

- M, I, A, L, V, F, C, W
- N, T, S, Q
- K, R
- P
- E, D
- Y, H
- G

```

      *: ** * : :. **. :*:*:***.*.:. :.***:***.* : .** *** :. :.:. *. *: **::* .:* : . .*:*:*:*: : *
sso_SSO1176 -MNRDIGKNAERELVSIILRGEFNAVRIPTSNSSPNPLPDIFATKGNILLSECKSTWENKVKVKEHQVRKLLDFLSMFTMKG-VPLIAIKFKQVHEWRV 98
sai_Saci_1741 -MNRDIGKSAERELVSIILRSEGFNAVRIPTSNSSPNPLPDIFATKGNILLAIECKSTWEHKVKVKDROVIKLFEFLSMFTMEG-RALIAVKFKEIHKWKV 98
vir_13989 MNIRQSGKYVEYKLEILEENGFKALRIPVSGTGKQALPDIIATKDNTIYPIEVKSTSKDVITIKKFQIEKLFEFCEIFDFCNCQPLVAIHKKYKTVIV 100
ruler 1.....10.....20.....30.....40.....50.....60.....70.....80.....90.....100

```



```

      : .:. . :.:.
sso_SSO1176 LVPEKAEDIIVTIDNSIPIEDLF-KILEKRIEK--ILTP- 135
sai_Saci_1741 MEIREIKEVEVTVDNSIFLENYISQFLENYIPIQAGRELSKL 139
vir_13989 YTLPODVRAKEKIKFKYGINS----- 121
ruler .....110.....120.....130.....140.

```



### CLAVES

sso\_SSO1176: *S. solfataricus* holliday junction resolvase  
sai\_Saci\_1741: *Sulfolobus acidocaldarius* hjc; conserved archaeal holliday junction resolvase  
vir\_13989: *Sulfolobus* virus SIRV2 hypothetical protein

- M, I, A, L, V, F, C, W
- N, T, S, Q
- K, R
- P
- E, D
- Y, H
- G

```

      :  :*:*:*:  * : * * :*:*:  .  .  :*:  *  *:*  *  :  * : * .  . :  .  .  :*  . :
vir_12658 -----MNKIEAIIISIKPQFVDEIMKGNKRFEFRKSF-LKSIP--DRCYIYSTKPVGKIVGFFTIIKQVLRDEPEKIWOQTSKKSGITKDFDFEYYSGKNK 91
spm_spyM18_1781 -----MNKIEAIIISIKPQFVDEIMKGNKRFEFRKSF-LKSIP--DRCYIYSTKPVGKIVGFFTIIKQVLRDEPEKIWOQTSKKSGITKDFDFEYYSGKNK 91
spg_SpyM3_1233 -----MNKIEAIIISIKPQFVDEIMKGNKRFEFRKSF-LKSIP--DRCYIYSTKPVGKIVGFFTIIKQVLRDEPEKIWOQTSKKSGITKDFDFEYYSGKNK 91
sph_MGAS10270_Spy1329 -----MNKIEAIIISIKPQFVDEIMKGNKRFEFRKSF-LKSIP--DRCYIYSTKPVGKIVGFFTIIKQVLRDEPEKIWOQTSKKSGITKDFDFEYYSGKNK 91
mmp_MMP0784 -----MTILLSIKPKVVKIILNGSKKYEFRKSI-FKKYDKNELVFIYSSYPVKRIVGTFVSGDIIENCPKILWNEFKNVSGIKESSEFFKYFKEKDK 90
ape_APE1779 MGAVLDEDEILYLMSIRPQYARAIMAGRKKYELRRIHGVPPIEEGSIIIVYASGNVKSIIIGEFQAKRVIQATPEKIWSIASKPGSGVGEDAWQYIRGAKR 100
ruler 1.....10.....20.....30.....40.....50.....60.....70.....80.....90.....100

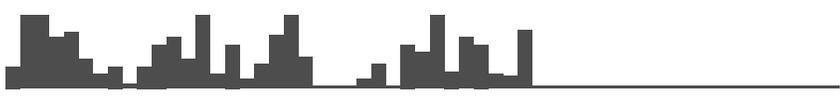
```



```

      .  **:::  :  *  .  . : *
vir_12658 AVAIEIDRLFAYKTPKNYHEID---IKGKPPQTYKTI----- 125
spm_spyM18_1781 AVAIEIDRLFAYKTPKNYHEID---IKGKPPQTYKTI----- 125
spg_SpyM3_1233 AVAIEIDRLFAYKTPKNYHEID---IKGKPPQTYKTI----- 125
sph_MGAS10270_Spy1329 AVAIEIDRLFAYKTPKNYHEID---IKGKPPQTYKTI----- 125
mmp_MMP0784 GYAIQINKLNEFKNPVNPKDIN---PDFRAPQSFYIS-QNFEEELFKI----- 136
ape_APE1779 AIAIEVGFRRLYERPVTLEEIRRIIPGWNPFYSYTIPEGNPIYELLIKRLRRRQHQL 158
ruler .....110.....120.....130.....140.....150.....

```



### CLAVES

- vir\_12658: *Streptococcus pyogenes* phage 315.4 provirus hypothetical protein
- spm\_spyM18\_1781: *S. pyogenes* (serotype M18) hypothetical protein
- spg\_SpyM3\_1233: *S. pyogenes* MGAS315 (serotype M3) hypothetical protein
- sph\_MGAS10270\_Spy1329: *S. pyogenes* MGAS10270 (serotype M3) phage transcriptional regulator
- mmp\_MMP0784: *Methanococcus maripaludis* S2 hypothetical protein
- ape\_APE1779: *A. pernix* hypothetical protein

- M, I, A, L, V, F, C, W
- N, T, S, Q
- K, R
- P
- E, D
- Y, H
- G

```

::* : :.*.* .:* .* : *:*:* ** :
bar_GBAA5592 MHKGCVFIMIMRFGYVSH-AMALWDCSPAKITITFTSFQKLSKQEREDKLYDVTKONLEHTIRILHYNIAHEIP--LYRLSSSIVPLATHPEVEFDYIGAF 97
ban_BA5592 MHKGCVFIMIMRFGYVSH-AMALWDCSPAKITITFTSFQKLSKQEREDKLYDVTKONLEHTIRILHYNIAHEIP--LYRLSSSIVPLATHPEVEFDYIGAF 97
btk_BT9727_5031 -----MIMRFGYVSH-AMALWDCSPAKITITFTSFQKLSKQEREDKLYDVTKONLEHTLRILHYNIAHEIP--LYRLSSSIVPLATHPEVEFDYIGAF 89
bce_BC5347 -----MLIRFGYVSH-AMALWDCSPAKITITFTSFQKLSKQEREDKLYHVIRONLEHTIRILHYNIAHEIP--LYRLSSSIVPLATHPEVEFDYIGVF 89
pca_Pcar_1346 -----MIRLGLCCQ-FRQAP--IHFKRTTAKALKKLDRIEQLNRLDRLCLKNAHSLVTAVEEVHRLGIG--AFRVCSNLLPLYTHPEVGYIEELP 86
vir_1114 -----MIGFACQMGVEVINGEFKQFKDTKFGVTRVCDMDKLNIEIAAKKLLADKVNNDNLOALENLVLRVAKLPQNRRLRVTSDLLPLFTHPKYGLYEDLN 95
ruler 1.....10.....20.....30.....40.....50.....60.....70.....80.....90.....100

```



```

. : : .....*:*:*.*:*.* :*:. : : : : : : .**** * : : : : :
bar_GBAA5592 TP----LWRKIGALIKEHNLRI SFHPNQFTLFTS DKPHITTNAITDMTYHYKVLD AIGIAD--SSYINIHVGGAYGNKEKAIERFHENIKKLPAHIKKQM 191
ban_BA5592 TP----LWRKIGALIKEHNLRI SFHPNQFTLFTS DKPHITTNAITDMTYHYKVLD AIGIAD--SSYINIHVGGAYGNKEKAIERFHENIKKLPAHIKKQM 191
btk_BT9727_5031 TP----LWRKIGALIKEHNLRI SFHPNQFTLFTS DKPHITTNAITDMTYHYKVLD AIGIAD--SSYINIHVGGAYGNKEKAIERFHENIKKLPAHIKKQM 183
bce_BC5347 TP----LWRKIGALIKEHNLRI SFHPNQFTLFTS DKPHITTNAITDMTYHYKILDAIGIAD--SSYINIHVGGAYGNKEKAIERFHENIKKLPHTHIKQM 183
pca_Pcar_1346 NGEETIYTKLTSVKKIKDRYNLRLSFHPDQFTLLSSPRPEVTSASIKDLEYQAMLASLIG---ADAINIHGGGAYGDKRSALARLAEQVSLPKAVRSRL 182
vir_1114 KR--ITYKLALIGDVAQRYGIRLSMHPSQYTVVNSDKPEVVDAAALIDLEMTFILKHLVPMGEGIIINHLNGKTDIIP-----VERMSDHLKSYL 184
ruler .....110.....120.....130.....140.....150.....160.....170.....180.....190.....200

```



```

*:*:*:* * : : : : :*:*:* ** .* : : ** : . .:*:*.. : : : : :
bar_GBAA5592 TLENDDKTCCTAETLSICQEKIPFVFDYHHHMANLCE----EPLLELLPAIFETWSHTNIVPKVHISSPKS--KKEFRAHAEYIDLEFIKPFHLVAKK 284
ban_BA5592 TLENDDKTCCTAETLSICQEKIPFVFDYHHHMANLCE----EPLLELLPAIFETWSHTNIVPKVHISSPKS--KKEFRAHAEYIDLEFIKPFHLVAKK 284
btk_BT9727_5031 TLENDDKTYTTAETLSICQEKIPFVFDYHHHMANLCE----EPLLELLPAIFETWSHTNIVPKVHISSPKS--KKEFRAHAEYIDLEFIKPFHLVAKK 276
bce_BC5347 TLENDDKTYTTAETLSICQKENIPFVFDYHHHMANLCE----EPLLELLPAIFETWSHTNISPKVHISSPRS--EKEFRAHAEYIDLEFIKPFHLIAKK 276
pca_Pcar_1346 TLENDDHTYTVQDLFPICEQLSIPLVYDVHHHRCNPPD----LSLAEATEACLQTWQHKREPLFHISSPKHGWNKKPKPHADYIDIDDF-----TEW 272
vir_1114 TFENDEKIGILTRILEVCEKYGFPLVFDIHHHRCNKT DGGILLRGT DPLAHRILGTWIGRKCRAKMHVSQSRG--TDNFRDLCAHSDMIYEAIEEVLST 282
ruler .....210.....220.....230.....240.....250.....260.....270.....280.....290.....300

```



```

. * :*:* *:*:* : : : :
bar_GBAA5592 INHNFDIMIESKQKDLAMLQFIQELSSIRGIKRISSTLQW 325
ban_BA5592 INHNFDIMIESKQKDLAMLQFIQELSSIRGIKRISSTLQW 325
btk_BT9727_5031 INHNFDIMIESKQKDLAMLQFIQELSSIRGIKRISSTLQW 317
bce_BC5347 HNHNFDMIESKQKDLALFQLIDELSAIRGIKRISGAMLQW 317
pca_Pcar_1346 KQLDATIDVEAKAKELAVVKLQELG-LPPWQAVY----- 306
vir_1114 YIRSWDIMVEAKFKNVAAAKLAGQIYW----- 309
ruler .....310.....320.....330.....340.

```



**CLAVES**

bar\_GBAA5592: *Bacillus anthracis* Ames uvvE; putative UV damage endonuclease  
ban\_BA5592: *B. anthracis* Ames uvvE; putative UV damage endonuclease  
btk\_BT9727 5031: *Bacillus thuringiensis* uvvE; putative UV damage endonuclease  
bce\_BC5347: *Bacillus cereus* ATCC 14579 uvvE; putative UV damage endonuclease  
pca\_Pcar\_1346: *Pelobacter carbinolicus* UV-endonuclease-like  
vir\_1114: Bacteriophage Aeh1 hypothetical protein

```

* . * * : : * . * : : * * : : * : : * : : * : : * : :
vir_6172 MNTKTTAQQSHQHPHSTLNFNPTFNMSKVSLISVTPDAEELIAYLARVS-----NKDAKLGDPADRLIRYLKIKHGHSPPFEMAHMVVEIETT-RAIA 91
pmt_PMT1550 -----MNSVSLISVTPKAEELIAYCARVS-----NPQNQDNPSERLIRYLKIKHRHWSPPFEMAHIIVMEIGTT-RAIA 66
pmn_PMN2A_1609 -----MSSVQLITSTPNAEKSMAYIARVS-----NPKNQDNDDFTKLIYCYKNEHWSVFEQAYMTLQIETT-RGIA 66
pma_Pro0273 -----MNNVHLVLTATPDAEKTMAIYARVS-----NPKNQANEEFGKLIKCYCINNEHWSVFEQAYMTLQIETT-RAIA 66
pmm_PMM0243 -----MSKVELISLTPAEKTMAYIARVS-----NPSNQANDKFAGLLRYCYKKEHWSVFEQSCMTLKIETT-RGIA 66
sso_SSO0321 -----MISVKLVSYTNDGKVIARIAAKMSRSRKGWDVYHEKDMTDDEIETWIRDAILHGYWSVLEHSVYTFSTIETSRVAS 75
ruler 1.....10.....20.....30.....40.....50.....60.....70.....80.....90.....100

```



```

* : : * : : * : : * : : * : : * : : * : : * : : * : :
vir_6172 PQILRHRFSFQEFSSQRYAVADLGEFVPE--ELRRQDTKNRONSIDDLPAEIVAEFQSEIDEVVFETIRVLYQRMLSYGVAKECARGILPLNTPTKLYMSG 189
pmt_PMT1550 AQILRHRFSFQEFSSQRYADVSTMAFAKPP-SLRRQDIKNRONSIDDLQETVSKLSSKANLLFEOSRDLYQEMLAAGVAKECAREVLPLATPTKIYMG 165
pmn_PMN2A_1609 AQILRHRFSFQEFSSQRYADSMQLG-EIPIPELRRQDNKNRONSISDLPEIINTFNKKIKHQFVQNKELYEEMLEAGIAKECARFVLPLATPTRIYMTG 165
pma_Pro0273 AQILRHRFSFQEFSSQRYADSTQLG-EIPIPELRRQDSKNRONSISDLDEDIANDFREKIKLQFKNSLDLYQEMLDKGVAKESARFILPLATPTRIYMTG 165
pmm_PMM0243 AQILRHRFSFQEFSSQRYAETSLLGNEIPIPNLRRQDQKNRONSIDDIPELQIKFSEKISKHFQANKLYEEMLEAGIAKECARFIMPLATPTRIYMTG 166
sso_SSO0321 HOLVRHRIASYTQMSHRFAKP-----IDEYYKPIIPPSAKKRKSKELVEKAYKEAYDNYTLLSGVPEEDARYVLPNGVNTINIVVTM 157
ruler .....110.....120.....130.....140.....150.....160.....170.....180.....190.....200

```



```

. * . : : * * : * * : * : . : . * : .
vir_6172 TIRSWIHYCQLRCNGTQEEHRLIAEECKNLLYANLPNVAKAAFMS----- 235
pmt_PMT1550 SVRSWLHYIDLRSAHGTOKEHQDVAIKCREILTKEMPCVCNAMWP----- 210
pmn_PMN2A_1609 SCRSWIHYINLRTGHGTOKEHMEIAQECKNIFSKYPPVSSALNWLFE----- 212
pma_Pro0273 SCRSWLHYINLRSGHGTOKEHLIIAKECKKIFTLQFPKVSEALGWIES----- 213
pmm_PMM0243 SCRSWIHYIQLRSKEGTQKEHMEIAEDCKRVFIKYFPVSEALNWE----- 212
sso_SSO0321 NARELYNFFSLRLCSRAQWEIRAIWKMLEEYKVVHPRLFKYTGPNCCIHENFIRNENESITLEDIFKDYKLEFISORCIEGVLRDGIKKCIINSRSVLD 257
ruler .....210.....220.....230.....240.....250.....260.....270.....280.....290.....300

```



```

vir_6172 --- 235
pmt_PMT1550 --- 210
pmn_PMN2A_1609 --- 212
pma_Pro0273 --- 213
pmm_PMM0243 --- 212
sso_SSO0321 NIK 260
ruler ...

```

### CLAVES

vir\_6172: Cyanophage P60 unknown protein  
pmt\_PMT1550: *Prochlorococcus marinus* thyX; thymidylate synthase [EC:2.1.1.-]  
pmn\_PMN2A\_1609: *P. marinus* NATL2A thymidylate synthase [EC:2.1.1.148] [KO:K03465]  
pma\_Pro0273: *P. marinus* SS120 THY1; predicted alternative thymidylate synthase [EC:2.1.1.-]  
pmm\_PMM0243: *P. marinus* MED4 possible Thy1 protein homolog [EC:2.1.1.-]  
sso\_SSO0321: *S. solfataricus* thymidylate synthase [EC:2.1.1.148] [KO:K03465]



## 9. REFERENCIAS

1. Waldor, M. K., Friedman, D. I., and Adhya S. L. **Phages: Their Role in Bacterial Pathogenesis and Biotechnology.** 2005. ASM Press. Washington, DC. Pg 3-17.
2. Hankin, E. H. 1896. **L'action bactéricide des eaux de la Jumna et du Gange sur de microbe du choléra.** *Ann. Inst. Pasteur.* 10: 511-523.
3. Emmerich, R., Löw, O. 1901. **Die künstliche Darstellung der immunisierenden Substanzen (Nucleasen-Immunproteïdine) und ihre Verwendung zur Therapie der Infektionskrankheiten und zur Schutzimpfung and Stelle de Heilserums.** *Z. Hyg. Infektkrankh.* 36: 9-36.
4. D'Herelle, F. 1917. **Sur un microbe invisible antagoniste des bacilles dysentériques.** *C. R. Acad. Sci. Paris.* 165: 373-375.
5. Luria, S. E., Delbrück, M. 1943. **Mutations of bacteria from virus sensitivity to virus resistance.** *Genetics.* 28: 491-511.
6. Lederberg, J., Lederberg E. M. 1952. **Replica plating and indirect selection of bacterial mutants.** *J. Bacteriol.* 63: 399-406.
7. Hershey, A. D., Rotman, R. 1949. **Genetic recombination between host-range and plaque-type mutants of bacteriophage in single bacterial cells.** *Genetics.* 34: 44-71.
8. Lwoff, A., Gutmann, A. 1950. **Recherches sur un Bacillus megathérium lysogène.** *Ann. Inst. Pasteur.* 78: 711-739.
9. Luria, S. E., Human, M. L. 1952. **A nonhereditary host-induced variation of bacterial viruses.** *J. Bacteriol.* 64: 557-569.
10. Bertani, G., Weigle, J. J. 1953. **Host-controlled variation in bacterial viruses.** *J. Bacteriol.* 65: 113-121.
11. Volkin, E., Astrachan, L. 1956. **Phosphorus incorporation in *Escherichia coli* ribonucleic acid after infection with bacteriophage T2.** *Virology.* 2: 149-161.



12. Loeb, T., Zinder, N. D. 1961. **A bacteriophage containing RNA.** *Proc. Natl. Acad. Sci. USA.* 47: 282-289.
13. Papke T. R., Doolittle W. F. 2003. **Phage Evolution: New Worlds of Genomic Diversity.** *Curr. Biol.* 13: R606-R607.
14. Hendrix, R. W. 2003. **Bacteriophage genomics.** *Curr. Opin. Microbiol.* 6: 506-511.
15. Wilhelm, S. W., Brigden, S. M., Suttle, C. A. 2002. **A dilution technique for the direct measurement of viral production: A comparison in stratified and tidally mixed coastal waters.** *Microb. Ecol.* 43: 168-173.
16. Casjens, S. 2003. **Prophages and bacterial genomics: what have we learned so far?** *Mol. Microbiol.* 49(2): 277-300.
17. Canchaya, C., Fournous, G., Chibani-Chennoufi, S., Dillmann, M. L., Brüssow, H. 2003. **Phage as agents of lateral gene transfer.** *Curr. Opin. Microbiol.* 6: 417-424.
18. Boyd, E. F., Davis, B. M., Hochhut, B. 2001. **Bacteriophage-bacteriophage interactions in the evolution of pathogenic bacteria.** *TRENDS in Microbiol.* 9(3): 137-144.
19. Brüssow, H., Canchaya, C., Hardt W. D. 2004. **Phages and the Evolution of Bacterial Pathogens: from Genomic Rearrangements to Lysogenic Conversion.** *Microbiol Mol Biol Rev.* 68(3): 560-602.
20. Liu, J., Dehbi, M., Moeck, G. et al. 2004. **Antimicrobial drug discovery through bacteriophage genomics.** *Nature Biotech.* 22(2): 185-191.
21. Waldor, M. K., Friedman, D. I., and Adhya S. L. **Phages: Their Role in Bacterial Pathogenesis and Biotechnology.** 2005. ASM Press. Washington, DC. Pg 37-54.
22. Rice, G., Stedman, K., Snyder, J., Wiedenheft, B., Willits, D., Brumfield, S., McDermott, T., and Young, J. M. 2001. **Viruses from extreme thermal environments.** *PNAS.* 98(23): 13341-13345.
23. Prangishvili, D., and Garrett, R. A. 2004. **Exceptionally diverse morphotypes and genomes of crenarchaeal hyperthermophilic viruses.** *Biochem. Soc. Trans.* 32(part 2): 204-208.



24. Prangishvili, D., and Garrett, R. A. 2005. **Viruses of hyperthermophilic Crenarchaea**. *TRENDS in Microbiol.* 13(11): 535-542.
25. Schleper, C., Kubo, K., Zillig, W. 1992. **The particle SSV1 from the extremely thermophilic archaeon *Sulfolobus* is a virus: Demonstration of infectivity and of transfection with viral DNA**. *Proc. Natl. Acad. Sci.* 89: 7645-7649.
26. ICTVdB - **The Universal Virus Database**, version 4, April 2006. <http://www.ncbi.nlm.nih.gov/ICTVdb/ICTVdb/> Last updated: 25 September 2006 by Cornelia Büchen-Osmond. Copyright © 2002 International Committee on Taxonomy of Viruses. All rights reserved.
27. Yu, M. X., Slater, M. R., Ackermann, H.-W. 2005. **Isolation and characterization of *Thermus* bacteriophages**. *Arch. Virol.* 151: 663-679.
28. Krupovič, M., Vilen, H., Bamford, J. K. H., Kivelä, H. M., Aalto, J.-H., Savilahti, H., Bamford, D. H. 2006. **Genome characterization of lipid-containing marine bacteriophage PM2 by transposon insertion mutagenesis**. *Am. Soc. Microbiol.*
29. Stephen, F. A., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., Lipman, D. J. 1997. **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs**. *Nucleic Acids Res.* 25: 3389-3402.
30. Lohr, E. J., Chen, F., Hill, R. T. 2005. **Genomic analysis of bacteriophage  $\Phi$ JL001: Insights into its interaction with a sponge-associated alpha-proteobacterium**. *Appl. Env. Microbiol.* 71(3): 1598-1609.
31. Van der Wilk, F., Dulleman, A. M., Verbeek, M., Van den Heuvel, J. F. J. M. 1999. **Isolation and Characterization of APSE-1, a Bacteriophage Infecting the Secondary Endosymbiont of *Acyrtosiphon pisum***. *Virology.* 262: 104-113.
32. Bettstetter, M., Peng, X., Garrett, R. A., Prangishvili, D. 2003. **AFV1, a novel virus infecting hyperthermophilic archaea of the genus *Acidianus***. *Virology.* 315: 68-79.
33. Arnold, H. P., Zillig, W., Ziese, U., Holz, I., Crosby, M., Utterback, T., Weidmann, J. F., Kristjanson, J. K., Klenk, H. P., Nelson, K. E., Fraser, C. M. 2000. **A novel Lipothrixvirus, SIFV, of the extremely thermophilic crenarchaeon *Sulfolobus***. *Virology.* 267: 252-266.



34. Stedman, K. M., She, Q., Phan, H., Arnold, H. P., Holz, I., Garrett, R. A., Zillig, W. 2003. **Relationships between fuselloviruses infecting the extremely thermophilic archaeon *Sulfolobus*: SSV1 and SSV2.** *Res. Microbiol.* 154: 295-302.
35. Xiang, Z., Chen, L., Huang, X., Luo, Y., She, Q., Huang, L. 2005. ***Sulfolobus tengchongensis* spindle-shaped virus STSV1: Virus-host interactions and genomic features.** *J. Virol.* 8677-8686.
36. Rice, G., Tang, L., Stedman, K., Roberto, F., Spuhler, J., Gillitzer, E., Johnson, J. E., Douglas, T., Young, M. 2004. **The structure of a thermophilic archaeal virus shows a double-stranded DNA viral capsid type that spans all domains of life.** *PNAS.* 101(20): 7716-7720.
37. Peng, X., Blum., H., She, Q., Mallok, S., Brügger, K., Garrett, R. A., Zillig, W., Prangishvili, D. 2001. **Sequences and replication of genomes of the archaeal Rudiviruses SIRV1 and SIRV2: Relationships to the archaeal Lipothrixvirus SIFV and some eukaryal viruses.** *Virology.* 291: 226-234.
38. Wiedenheft, B., Stedman, K., Roberto, F., Willits, D., Gleske, A-K., Zoeller, L., Snyder, J., Douglas, T., Young, M. 2004. **Comparative genomic analysis of hyperthermophilic archaeal *Fuselloviridae* viruses.** *J. Virol.* 78(4): 1954-1961.
39. Daubin, V., Ochman, H. 2004. **Start-up entities in the origin of new genes.** *Curr. Opin. Gen. Dev.* 14: 616-619.
40. Delaye, L., Becerra, A., Lazcano, A. 2005. **The last common ancestor: What's in a name?** *Orig. Life Evol. Bios.* 35: 537-554.
41. Markine-Goriaynoff, N., Gillet, L., Van Etten, J. L., Korres, H., Verma, N., Vanderplasschen, A. 2004. **Glycosyltransferases encoded by viruses.** *J. Gen. Virol.* 85: 2741-2754.
42. Sullivan, M. B., Coleman, M. L., Weigele, P., Rohwer, F., Chisholm, S. W. 2005. **Three *Prochlorococcus* cyanophage genomes: signature features and ecological interpretation.** *PLoS Biol.* 3(5): e144.
43. Prangishvili, D., Garrett, R. A., Koonin, E. V. 2006. **Evolutionary genomics of archaeal viruses: unique viral genomes in the third domain of life.** *Virus Res.* 117: 52-67.



44. Liu, J., Glazko, G., Mushegian, A. 2006. **Protein repertoire of double-stranded DNA bacteriophages** *Virus Res.* 117: 68-80.
45. Waldor, M. K., Friedman, D. I., and Adhya S. L. **Phages: Their Role in Bacterial Pathogenesis and Biotechnology.** 2005. ASM Press. Washington, DC. Pg 18-36.
46. Lehnerr, H., Maguin, E., Jafri, S., Yarmolinsky, M. B. 1993. **Plasmid addiction genes of bacteriophage P1: *doc*, which causes cell death on curing of prophage, and *phd*, which prevents host death when prophage is retained.** *J. Mol. Biol.* 233: 414-428.
47. Lennox, E. S. 1955. **Transduction of linked genetic characters of the host by bacteriophage P1** *Virology.* 1: 190-206.