



FACULTAD DE CIENCIAS AGRARIAS  
FACULTAD DE CIENCIAS BIOQUÍMICAS Y FARMACÉUTICAS  
UNIVERSIDAD NACIONAL DE ROSARIO

**“Anotación genómica y análisis comparativo entre bacteriófagos de  
*Staphylococcus aureus*”**

Tec. Sup. en Genética Soledad Telma Carrasco

TRABAJO FINAL PARA OPTAR AL TÍTULO DE ESPECIALISTA EN  
BIOINFORMÁTICA

DIRECTOR: Dr. Cristian Suárez

2018

**Anotación genómica y análisis comparativo entre bacteriófagos de  
*Staphylococcus aureus***

Soledad Telma Carrasco

Téc. Sup. en Genética – Inst. Superior Part. Nro. 4080 “TECNOLOGÍA MÉDICA”

Este Trabajo Final es presentado como parte de los requisitos para optar al grado académico de Especialista en Bioinformática, de la Universidad Nacional de Rosario y no ha sido previamente presentada para la obtención de otro título en ésta u otra Universidad. El mismo contiene los resultados obtenidos en investigaciones llevadas a cabo en Laboratorio de Microbiología Molecular de la Facultad de Ciencias Médicas U.N.R., durante el período comprendido entre Octubre de 2016 y Marzo de 2018, bajo la dirección del Dr. Cristian Suárez.

Carrasco Soledad T.

Suárez Cristian A.

Defendida: .....de 2018.

# Agradecimientos

Al Dr. Cristian Suárez por su apoyo y paciencia hasta último momento, pese a mis preguntas repetitivas.

Al Dr. Ricardo Morbidoni por abrirme las puertas de su laboratorio y darme la oportunidad de seguir creciendo profesionalmente.

Al grupo de microbiología por integrarme y acompañarme.

Al Dr. Guillermo Pratta quién me motivó a seguir esta carrera y a avanzar profesionalmente, siempre dispuesto a escuchar mis inquietudes y a dar un buen consejo.

A Martín, mi compañero en este camino.

# Dedicatorias

A Bautista, por ser la luz que le faltaba a este Sol. Porque si faltaba una razón eras vos.

## Publicaciones

Parte de este trabajo formó parte del artículo científico “Broad-range lytic bacteriophages that kill *Staphylococcus aureus* local field strains” publicado el 25 de Julio de 2017 en la revista PLoS ONE 12(7): e0181671. Su autoría le pertenece a Virginia Abatángelo, Natalia Peressutti Bacci, Carina A. Boncompain, Ariel A. Amadio, Soledad Carrasco, Cristian A. Suárez y Héctor R. Morbidoni.

# Abreviaturas y Símbolos

AA: aminoácido/s

ADN: ácido desoxiribonucleico

ARNt: ácido ribonucleico de transferencia

ARNt-Arg: ácido ribonucleico de transferencia de arginina

ARNt-Asp: ácido ribonucleico de transferencia de aspargina

ARNt-His: ácido ribonucleico de transferencia de histidina

ARN-Met: ácido ribonucleico de transferencia de metionina

ARNt-Phe: ácido ribonucleico de transferencia de fenilalanina

ARNt-Trp: ácido ribonucleico de transferencia de triptófano

BLAST: Basic Local Alignment Search Tool

Fago/s: bacteriófago/s

Kpb: kilopares de bases

ORFs: marco abierto de lectura

pb.: pares de bases

RTL: región terminal larga, zona comprendida entre los ORFs *treA* y *bofL*

# Resumen

*Staphylococcus aureus* es un patógeno bacteriano capaz de causar una amplia variedad de enfermedades tanto en humanos como en animales. Con el fin de entender la mecánica de su patogenia y de buscar tratamientos alternativos es que se recurre al uso de bacteriófagos. Los mismos son virus que infectan bacterias de manera altamente específica y tienen la capacidad de destruirlas. En este trabajo se realizó la anotación y el análisis comparativo de cuatro bacteriófagos *provenientes de diferentes orígenes* capaces de infectar *S. aureus*, dos de ellos líticos y dos temperados. Centrándose en sus diferencias se propuso la utilización de diferentes herramientas bioinformáticas y el análisis de diferentes genes para su estudio. Se logró caracterizar a los bacteriófagos y analizar tanto las diferencias como las similitudes genéticas con los pertenecientes a su misma familia. Se analizaron las endolisinas codificadas por estos fagos, de manera interesante un fago (vB\_SauS\_308) solo contiene el dominio CHAP de ésta. Por último, se realizó la búsqueda de genes que codifiquen factores de virulencia y toxinas.

# Abstract

## **Genomic annotation and comparative analysis between *Staphylococcus aureus* bacteriophages**

*Staphylococcus aureus* is a bacterial pathogen capable of causing a variety of diseases both in humans and in animals. We resort to the use of bacteriophages in order to understand the pathogenesis and look for alternative treatments. Phages are viruses that infect bacteria in a specific way and have the ability to lyse them. In this study the annotation and the comparative analysis of four bacteriophages from environmental, animal and human origin capable of infect *S. aureus* was carried out.

Considering the differences between the phages the utilization of different bioinformatics tools and the selection of distinct genes has been proposed. It has been possible to describe the distinctive features and the similarities with those belonging to the same family. Endolysins were analysed, and one phage (vB\_SauS\_308) only shows CHAP domain. To finalize the search for genes encoding virulence factors and toxins was carried out.

# Índice

Agradecimientos.....	3
Dedicatorias.....	4
Publicaciones.....	5
Abreviaturas y Símbolos.....	6
Resumen.....	7
Abstract.....	8
Índice.....	9
Introducción.....	11
Objetivos.....	14
Materiales.....	15
Materiales y Métodos.....	15
Fuente de bacteriófagos.....	15
Anotación de genomas.....	15
Análisis comparativo de Caudovirales.....	17
Herramientas para análisis comparativo de fagos Siphoviridae.....	18
Herramientas para análisis comparativo de fagos Myoviridae.....	19
Resultados y Discusión.....	20
Características generales.....	20
Organización del genoma.....	22
Características generales de los genomas de bacteriófagos estafilocócicos..	27
Análisis comparativo de fagos Siphoviridae.....	29
Módulo de lisogenia.....	32
Módulo de lisis.....	33
Módulo de replicación del ADN.....	33
Análisis comparativo de fagos Myoviridae.....	34
Conclusiones.....	38
Bibliografía.....	40
ANEXOS.....	45
Anexo I – <i>script</i> genomediagram.py.....	45
Anexo II – <i>script</i> heatmap.r.....	46

Anexo III – <i>script</i> circularfilo.r.....	46
Anexo IV – <i>script</i> holinlength.pl.....	46
Anexo V – Anotación de genomas.....	47

# Introducción

Un tema que alerta a la salud pública a nivel mundial es la resistencia antimicrobiana que no solo presume mayores costos económicos sino que aumenta el riesgo de muerte y disminuye la probabilidad de obtener mejores resultados en pacientes con diversas patologías (“WHO,” 2018). La resistencia es un proceso que ocurre en forma natural a través de largos períodos de tiempo, es por esto, que el problema no es la resistencia per se sino la celeridad con la que estos cambios se presentan en las últimas décadas.

En bacterias la resistencia se puede dar por varias vías, entre las cuales podemos nombrar las mutaciones genéticas espontáneas y la transferencia horizontal de genes de una bacteria resistente a una que no lo es. Los antibióticos eliminan las bacterias sensibles a la droga administrada dejando aquellas resistentes, lo que conlleva a la reproducción de éstas últimas. Esto se debe al mal uso y al abuso de los antibióticos, dando como resultado la pérdida de la eficacia de los mismos (Ventola, 2015).

Nos encontramos en la era post-antibióticos en donde estos están perdiendo su eficacia debido a que las bacterias son, no sólo resistentes a una droga, sino a varias a la vez (multi-resistencia) (WHO, 2014). La falta de nuevas opciones terapéuticas que reemplacen los antibióticos motivó replantearse viejas opciones como el uso de bacteriófagos (virus que infectan bacterias) los cuáles fueron dejados de lado con la aparición de los antibióticos. Los fagos han evolucionado conjuntamente con sus hospedadores durante miles de millones de años y, por lo tanto, han adquirido mecanismos para contrarrestar las defensas bacterianas, esto los hace ser específicos y letales para la bacteria blanco y no suponen ningún riesgo para la salud de los humanos (Burrowes, Harper, Anderson, McConville, & Enright, 2011; Paul et al., 2011). Además, los cócteles de bacteriófagos tienen bajo costo y pueden ser modificados mediante la incorporación de nuevos bacteriófagos o la

modificación genética de los existentes. De esta manera uno puede contrarrestar la aparición de bacterias resistentes a los mismos.

Entre los principales patógenos bacterianos que la Organización Mundial de la Salud (OMS) pone en alerta se encuentra *Staphylococcus aureus*. Estas bacterias son capaces de causar una amplia variedad de enfermedades debido a que coloniza piel y mucosas tanto del ser humano como de los animales pudiendo derivar en infecciones de piel (como por ejemplo forúnculos e impétigo), hasta dolencias más graves como bacteriemia, sepsis, endocarditis bacteriana y Síndrome de shock tóxico (Deghorain & Van Melderren, 2012; Paul et al., 2011). Hasta el momento se han reportado cepas de *S. aureus* resistentes a meticilina (SARM) y otra cepas resistentes a vancomicina (SARV) (Gardete & Tomasz, 2014; McGuinness, Malachowa, & DeLeo, 2017; Medina et al., 2013).

Una de las características entre las diferentes cepas de *S. aureus* es la presencia de elementos genéticos móviles, entre ellos los bacteriófagos. Además, estos últimos aumentan la plasticidad genómica de *S. aureus* durante la infección, facilitando la adaptación del patógeno a las condiciones del hospedador. Los bacteriófagos pueden portar factores de virulencia que favorecen la patogenicidad de la bacteria, entre estos podemos encontrar estafiloquinasas (*sak*) responsable de la destrucción de tejidos, leucocidinas (*lukF-PV*, *lukS-PV*, *lukM*, *lukF*) causantes de las infecciones necróticas, enterotoxinas (*sea*, *seg*, *sek*, *sek2*, *sep*, *seq*) que causan, entre otros, el Síndrome de shock tóxico e intoxicación por alimentos, y la toxina exfoliativa A (*eta*) causante de las infecciones severas de piel (Deghorain & Van Melderren, 2012; Goerke et al., 2009).

La motivación por el estudio de bacteriófagos de *S. aureus* podemos separarla en distintas áreas: 1) debido a la participación que tienen en la patogenicidad de *S. aureus* (fagos temperados), 2) ya que tienen la capacidad de destruir específicamente al patógeno y poseen amplio rango de hospedador (fagos líticos) y 3) la potencialidad de las enzimas líticas codificadas por los fagos (endolisinas) las cuales tienen elevado potencial como agentes antibacterianos.

Para la utilización de fagos como agentes de biocontrol se debe realizar un análisis profundo de los fagos en cuestión, en donde los procedimientos convencionales de biología no son suficientes. Es necesaria la anotación de los genomas para realizar la búsqueda e identificación de genes tanto de interés biotecnológico como asociados a patogénesis. Además, la comparación entre genomas es importante para poder clasificar los fagos y relacionarlos con fagos reportados; para todo esto es preciso utilizar herramientas bioinformáticas que faciliten tanto la obtención de datos, como su interpretación y visualización.

# Objetivos

## Objetivo general:

- Realizar la anotación genómica de cuatro bacteriófagos así como el análisis y la comparación de los mismos mediante la utilización de diferentes herramientas bioinformáticas.

## Objetivos específicos:

- Buscar e identificar genes de interés biotecnológico, como endolisinas y enzimas despolimerizantes, y analizar y comparar los mismos.
- Buscar genes o islas de genes asociados a patogénesis, especialmente en los bacteriófagos lisogénicos.
- Realizar la comparación entre genomas y la identificación y el análisis de enzimas claves para la clasificación y la actividad.

# Materiales y Métodos

## Fuente de bacteriófagos

Se realizó la anotación, análisis y comparación de cuatro bacteriófagos pertenecientes al Laboratorio de Microbiología Molecular de la Facultad de Ciencias Médicas de la Universidad Nacional de Rosario, provenientes de muestras de suelo (vB\_Sau\_CG y vB\_Sau\_S24), hisopado nasal humano (MatT) y manos de carniceros (308). Los mismos fueron secuenciados en INDEAR (Instituto de Agro-Biotecnología de Rosario) mediante un equipo Hiseq/Illumina y ensamblados en el mismo lugar utilizando el pipeline A5 (Tritt, Eisen, Facciotti, & Darling, 2012). Luego de analizar el tamaño y la organización de los fagos se siguieron las directrices de nomenclatura propuestas por Kropinski *et al.* (Kropinski, Prangishvili, & Lavigne, 2009) para renombrar a dos fagos como vB\_SauS\_308 y vB\_SauS\_MatT, nomenclatura que será utilizada en este trabajo cuando sea necesaria. Mientras que para los dos restantes se mantuvo el nombre utilizado en el artículo de Abatángelo y col. del mismo laboratorio (Abatángelo et al., 2017).

## Anotación de genomas

La anotación de los bacteriófagos se llevó a cabo mediante el uso del software DNA Master (<http://phagesdb.org/DNAMaster/>), el cuál realiza la predicción de los marcos abiertos de lectura hipotéticos (ORFs) utilizando los resultados programa GeneMark (Besemer, 2001); su input es la secuencia de nucleótidos de los genomas en formato FASTA. Luego la anotación fue revisada manualmente comparando con los datos obtenidos por el servidor RAST (<http://rast.nmpdr.org>) (Aziz et al., 2008), siguiendo las directrices del manual de anotación de DNA Master. Se evaluó la similitud de todas las proteínas con secuencias ya conocidas, tanto de bacterias

como de bacteriófagos, depositadas en la base de datos pública del NCBI (National Center for Biotechnology Information). Para esto se corrió a través de DNAMaster el servidor BLAST (Basic Local Alignment Search Tool), con un cutoff del E-value de  $1e-3$ . Se alinearon las secuencias con BLASTN y se visualizaron con ACT (Artemis Comparison Tool) (Carver et al., 2005) para así verificar si aquellas secuencias que no arrojaron positivos o que dieron como resultado “proteínas hipotéticas” tenían homología con alguno de los fagos secuenciados, también se utilizó Mauve (Darling, Mau, Blattner, & Perna, 2004) para el análisis de regiones conservadas.

Se prosiguió con la búsqueda e identificación de genes e islas de genes asociados a patogénesis mediante BLASTP e IslandViewer (Bertelli et al., 2017; Langille, Hsiao, & Brinkman, 2010).

La predicción de dominios y de motivos se realizó con Pfam e InterProScan (Jones et al., 2014). Los ARNt putativos fueron predichos a través de tRNA Scan-SE (Lowe & Chan, 2016) y ARAGORN (Laslett & Canback, 2004).

Una vez finalizada la anotación se prosiguió al reordenamiento de los ORFs putativos ubicando en primer lugar al que codifica la unidad mayor de la terminasa en fagos Myoviridae y para integrasa en los Siphoviridae.

Los diagramas de los genomas fueron realizados con Python versión 2.7.12 (<http://www.python.org>) con el paquete GenomeDiagram (Pritchard, White, Birch, & Toth, 2006) cuyo *script* (ANEXO I) fue extraído de “Biopython Tutorial and Cookbook” (Chang et al., 2010) y corrido en el sistema operativo Ubuntu 16.04.4 LTS.

## Análisis comparativo de Caudovirales

La mayoría de los bacteriófagos pertenecen al orden de Caudovirales, lo cuales poseen ADN doble hebra, entre los 18 y 550 kpb y poseen cápsides icosaedricas. Este orden se divide en fagos pertenecientes a las siguientes familias: Myoviridae, Siphoviridae y Podoviridae.

Se recolectaron los genomas de 54 fagos pertenecientes al orden Caudovirales de la base de datos del NCBI (EU418428.2, FR852584.1, JX080300.2, JX080301.2, JX080302.2, JX080303.2, JX080304.2, JX080305.2, JX875065.1, KP687431.1, KP687432.1, KR902361.1, KR908644.1, NC\_004678.1, NC\_004679.1, NC\_005356.1, NC\_005356.1, NC\_005880.2, NC\_007021.1, NC\_007045.1, NC\_007046.1, NC\_007047.1, NC\_007048.1, NC\_007049.1, NC\_007050.1, NC\_007051.1, NC\_007052.1, NC\_007053.1, NC\_007055.1, NC\_007056.1, NC\_007057.1, NC\_007058.1, NC\_007059.1, NC\_007060.1, NC\_007061.1, NC\_007062.1, NC\_007063.1, NC\_007064.1, NC\_007065.1, NC\_007066.1, NC\_019448.1, NC\_019726.1, NC\_020877.1, NC\_022090.1, NC\_022918.1, NC\_022920.1, NC\_023009.1, NC\_023573.1, NC\_025416.1, NC\_025417.1, NC\_025426.1, NC\_028765.1, NC\_028962.1, NC\_030933.1). Sus secuencias, junto a la de los fagos de interés, fueron ordenadas arbitrariamente de forma tal que el inicio fuese el ORF del gen de la subunidad mayor de la terminasa. Esta enzima es la encargada del reconocimiento del ADN y el inicio del empaquetamiento en el fago.

Con el fin de llevar a cabo un análisis comparativo con los cuatro bacteriófagos de interés se realizó una comparación de fragmentos del tipo todos contra todos con GENEES (Ågren, Sundström, Håfström, & Segerman, 2012) en su versión 2.2.1 (Parámetros: "Accurate"; Método: "BLASTN"; Reference: phage K - NC\_005880.2) El resultado fue graficado con R software – Bioconductor (RC Gentleman et al., 2004) versión 3.2.3, utilizando la función heatmap (Robert Gentleman, Carey, Huber, Irizarry, & Dudoit, 2005) (ANEXO II). La red filogenética fue representada con

SplitsTree4 (Huson & Bryant, 2006) con el método Neighbor-Net (Bryant & Moulton, 2004).

## Herramientas para análisis comparativo de fagos Siphoviridae

Los genomas de bacteriófagos pertenecientes a la familia Siphoviridae fueron cargados en DNAMaster y luego exportados en formato GenBank para ingresarlas a la base de datos de uso local del programa Phamerator (Cresawn et al., 2011; Russell & Hatfull, 2017). Este programa compara los genomas de los fagos, sus genes y sus productos, a través de BLASTP y ClustalW. La elección de esta herramienta se fundamenta no solo en su interfaz gráfica (que permite una buena visualización de los resultados de las comparaciones) sino en que agrupa las proteínas hipotéticas en “Phamilies” (también llamadas “Phams”) según el puntaje de los alineamientos.

Dichas secuencias extraídas fueron alojadas en una carpeta (una por grupo) para crear archivos multifasta (fasta2multifasta.py - Jonathan Grandaubert); en el caso de ser necesaria su traducción (holinas, polA, dnaC y dnaD) se utilizó el script en python “ConvertFasta.py” (<https://github.com/>) modificando el parámetro “codonTableID = 11”. Las secuencias fueron alineadas con ClustalW y los resultados convertidos a formato Phylip utilizando Biopython (biopython.org); esta conversión de formato fue necesaria para luego realizar los árboles filogenéticos circulares (Fig. 11) con R software (ANEXO III).

En el caso de las holinas, además, fue necesario el cálculo de longitud de las secuencias proteicas, para lo cual se realizó un script en BioPerl, con la ayuda del paquete Bio::SeqIO (<http://bioperl.org/>) (ANEXO IV).

## Herramientas para análisis comparativo de fagos Myoviridae

Se realizaron dos alineamientos de las secuencias nucleotídicas con ClustalW (mediante línea de comandos de CLUSTAL 2.1 Multiple Sequence Alignments instalada por default en BioLinux, uno del fago vB\_Sau\_CG vs. el fago K y el otro de este último con vB\_Sau\_S24. Se realizó un dotplot con NUCmer (Delcher, Phillippy, Carlton, & Salzberg, 2002; Kurtz et al., 2004) eligiendo los parámetros por default. Se finalizó con una comparación mediante BLASTN visualizándose con CGView Comparison Tools, que también fue utilizada para los fagos Siphoviridae, utilizando el script `build_blast_atlas.sh` (Grant & Stothard, 2008).

# Resultados y Discusión

## Características generales

El análisis de los genomas anotados (ANEXO V) ubicó 211 secuencias codificantes en vB\_Sau\_S24, 224 en vB\_Sau\_CG, 69 en vB\_SauS\_308 y 71 vB\_SauS\_MatT. El análisis de la presencia de ARN de transferencia, utilizando ARAGORN (Laslett & Canback, 2004) y tRNA-Scan (Lowe & Chan, 2016) mostró que el primer fago codifica dos ARNt (ARNt-Asp y ARNt-Arg) y el segundo cinco (ARNt-His, ARNt-Met, ARNt-Trp, ARNt-Phe, ARNt-Asp), mientras que los fagos restantes no presentan ARNt, como era de esperarse.

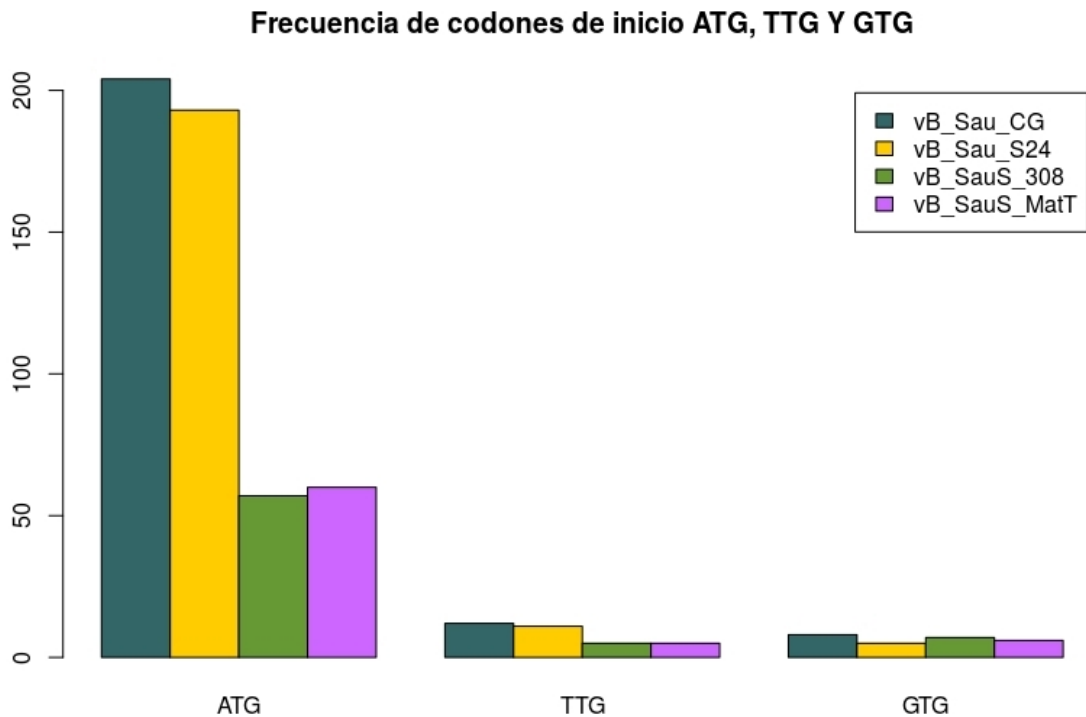
La cantidad de ORFs es proporcional al tamaño del genoma, siendo el genoma de vB\_SauS\_308 el más pequeño, abarca 39.974 pb. y contiene 69 marcos abiertos de lecturas (ORFs), mientras que vB\_SauS\_MatT comprende 71 ORFs en 44.300 pb. El contenido G+C% es de 34,11% y 34,71% respectivamente. Los genomas de los fagos vB\_Sau\_S24 y vB\_Sau\_CG son mayores, teniendo el primero 139.997 pb. y 211 ORFs, y 142.934 pb. y 229 ORFs el segundo (Tabla 1). La superposición de genes es una característica común en virus, que les permite incrementar su número de proteínas sin necesidad de aumentar su tamaño (Chirico, Vianelli, & Belshaw, 2010). Esta característica se encuentra presente en los fagos estudiados, siendo 30 los ORFs que se comportan de esta forma en vB\_Sau\_CG, 23 en vB\_Sau\_S24 y 14 tanto en vB\_SauS\_308 como en vB\_SauS\_MatT.

**Tabla 1:** Características genómicas generales de los bacteriófagos vB\_Sau\_CG, vB\_Sau\_S24, vB\_SauS\_308 y vB\_SauS\_MatT

Bacteriófago	Tamaño genómico (bp)	% G+C	CDS	% codificante	Densidad genica (gen/Kbp)	ARNt	Longitud promedio (bp)
vB_Sau_CG	142934	30.51	224	90.4	1.57	5	577
vB_Sau_S24	139997	30.86	209	89.8	1.49	2	602
vB_SauS_308	39974	34.11	69	84.3	1.56	-	541
vB_SauS_Mat_T	44300	34.71	71	94.4	1.60	-	589

La cadena directa codifica 167, 158, 62 y 64 ORFs y la reversa 62, 53, 7 y 7 ORFs para vB\_Sau\_CG, vB\_Sau\_S24, vB\_SauS\_308 y vB\_SauS\_MatT respectivamente. La comparación con secuencias disponibles en bases de datos públicas mostraron que vB\_Sau\_S24 tiene 10/138 (7,25%) proteínas hipotéticas sin homología en bases de datos y vB\_Sau\_CG 4/158 (2,53%); solo una de estas proteínas es homóloga entre los fagos. Mientras que todas las proteínas hipotéticas de vB\_SauS\_308 y vB\_SauS\_MatT son homólogas con al menos una proteína en base de datos públicas. La densidad codificante, definida como “la fracción de pares de bases que se encuentran en los codones” (Fickett & Guigó, 1993), es mayor al 80% en todos los casos.

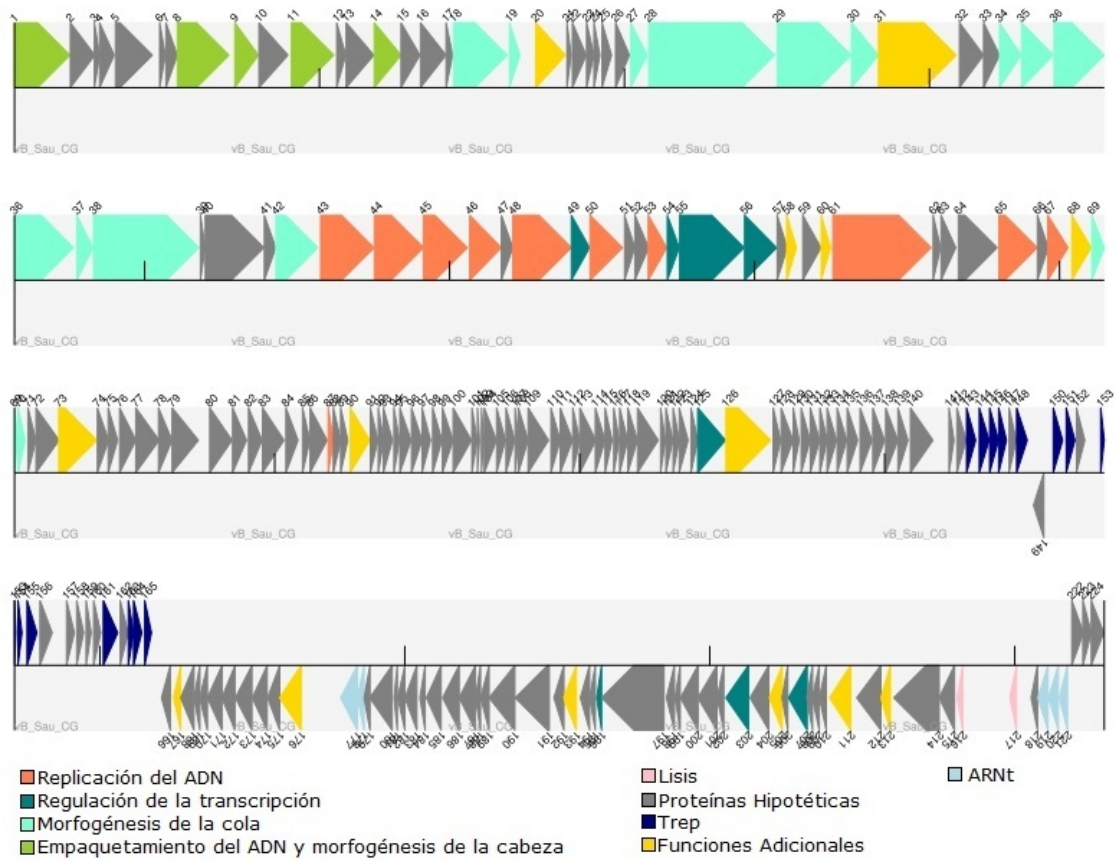
Se observó que el 89,7% de los ORFs inician con un codón AUG, mientras que solo el 10,3% lo hacen con un codón TTG o GTG (5,76% y 4,54% respectivamente) (ver Fig 1.), lo cuál contradice los propuestos que aseveran que los codones AUG y GTG tienen aproximadamente la misma frecuencia (Russell & Hatfull, 2015). El ORF vBSauCG\_102 es el de menor tamaño, con 69 pb. en su hebra directa codificando para una proteína hipotética, mientras que el ORF de mayor tamaño es el vB\_SauS308\_60 que se ubica en la misma dirección, posee 4650 pb. y codifica para la proteína tape-measure encargada de determinar la longitud de la cola del fago.



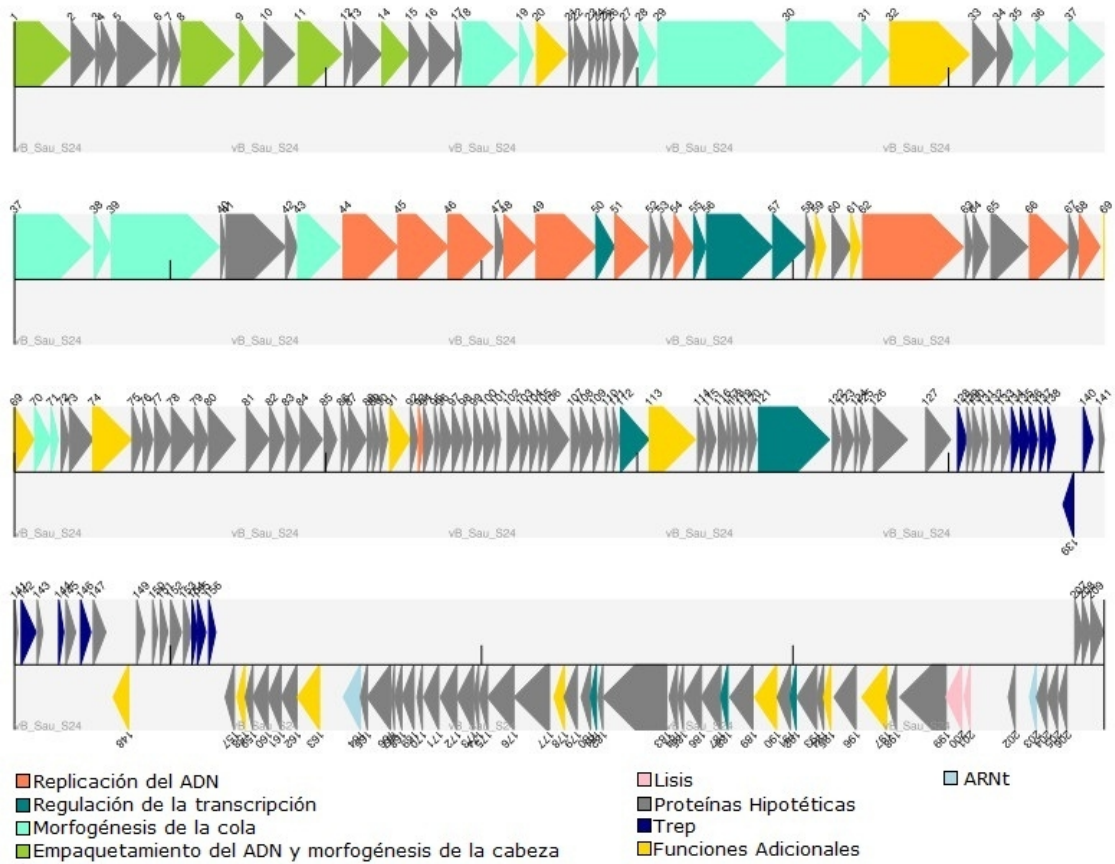
**Figura 1:** Frecuencia de codones de inicio ATG, TTG y GTG en bacteriófagos vB\_Sau\_CG, vB\_Sau\_S24, vB\_SauS\_308 y vB\_SauS\_MatT

## Organización del genoma

En las figuras 2 a 5 se esquematizan la organización de los genomas. Se puede observar que no se respeta la organización modular descrita por Deghorain y Van Maderen (Deghorain & Van Melderren, 2012), en donde para los bacteriófagos Myoviridae (Fig. 2 y 3) debería seguir el orden: empaquetamiento del ADN y morfogénesis de la cápside, morfogénesis de la cola, replicación del ADN, regulación de la transcripción, y lisis.

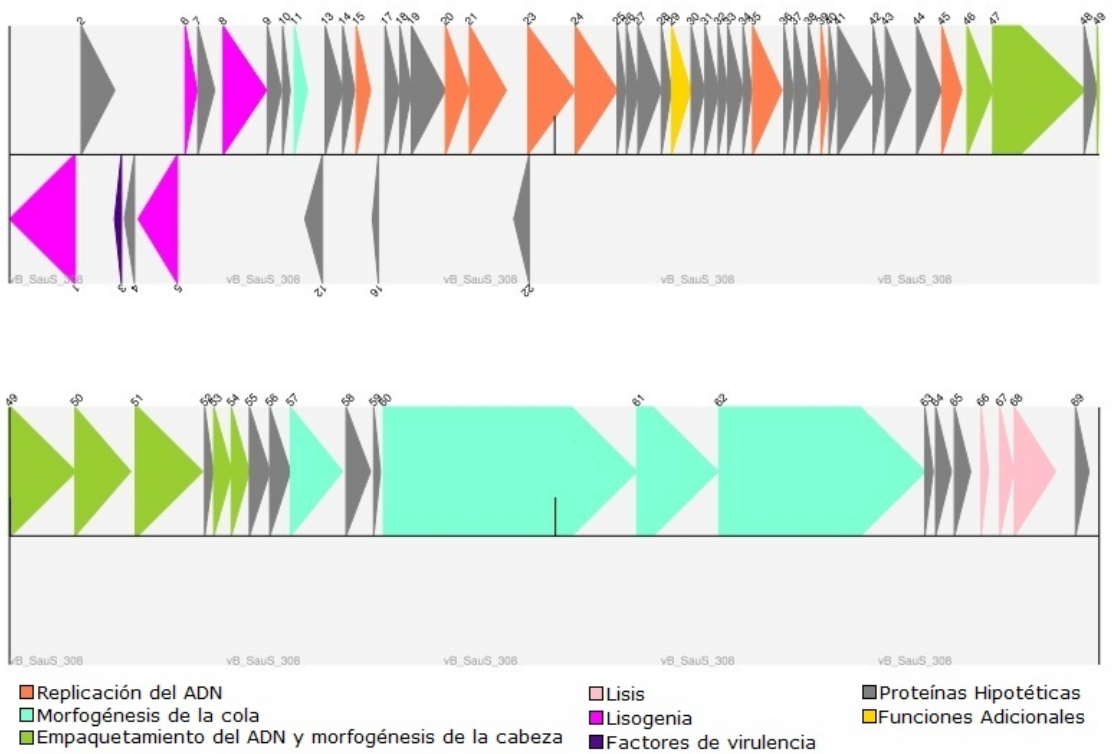


**Figura 2:** Organización del genoma del fago vB\_Sau\_CG. Los ORFs fueron coloreados según su función.

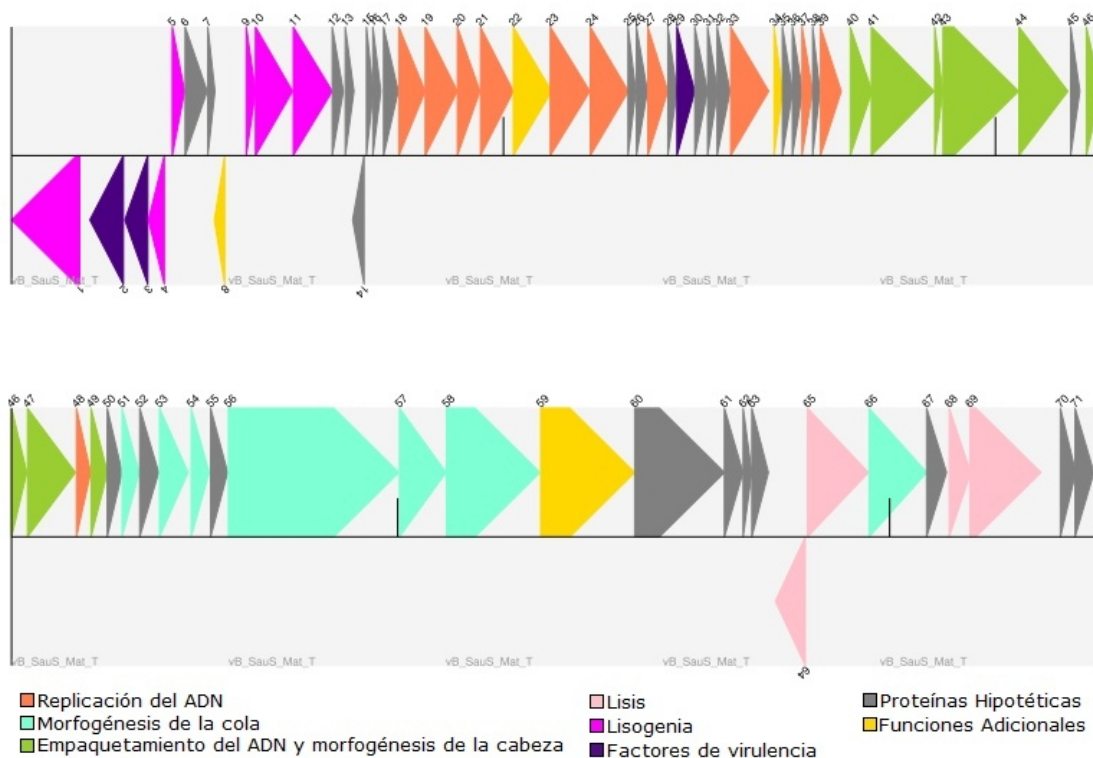


**Figura 3:** Organización del genoma del fago vB\_Sau\_S24. Los ORFs fueron coloreados según su función.

Los fagos Siphoviridae (Fig. 4 y 5) por el contrario, sí respetan un orden, contienen además, como era de esperarse, el módulo de lisogenia. Un análisis global de la organización de los genes revela que la mayoría se transcribe de una sola hebra, formándose clusters que se transcriben de la opuesta. Solo el 22% de los ORFs se ubican en la hebra reversa, siendo estos los que codifican proteínas que participan en la lisogenia y las involucradas en el metabolismo de ARN.



**Figura 4:** Organización del genoma del fago vB\_SauS\_308. Los ORFs fueron coloreados según su función.



**Figura 5:** Organización del genoma del fago vB\_SauS\_MatT. Los ORFs fueron coloreados según su función.

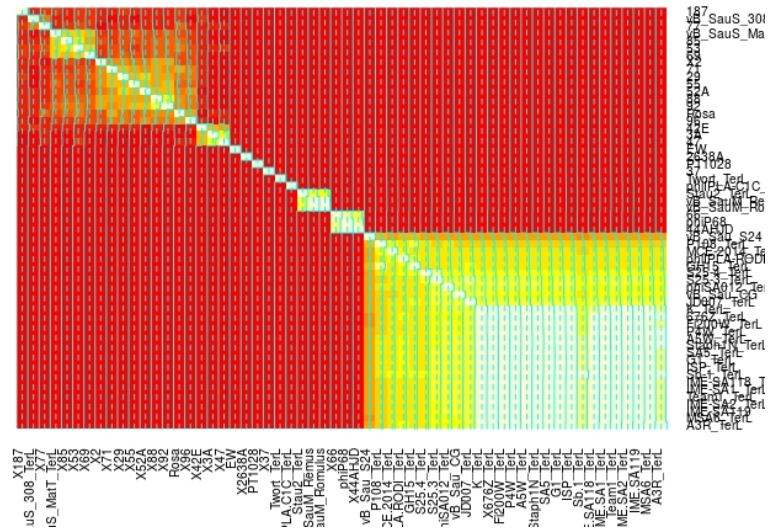
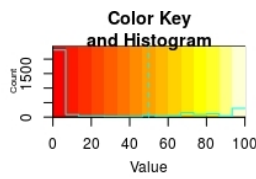
La búsqueda de genes que codifiquen para factores de virulencia demostró que vB\_Sau\_CG y vB\_Sau\_S24 (fagos líticos) no contienen genes que codifiquen proteínas asociadas a la virulencia o proteínas tóxicas. Mientras que, los fagos vB\_SauS\_308 y vB\_SauS\_MatT poseen genes hipotéticos que codifican para estafiloquinasa (*sak*), leucocidina Pantón-Valentine (*PVL*) e isla patogénica SaPI<sub>n2</sub>.

La identificación de endolisinas resultó en la presencia de ellas en todos los fagos, con excepción de vB\_SauS\_308. Éste fago solo posee el dominio CHAP, dominio que generalmente viene acompañado por los dominios N-acetylmuramoyl-L-alanine amidase y Bacterial SH3 domain, que sí se encuentran presente en el resto de los fagos.

## Características generales de los genomas de bacteriófagos estafilocócicos

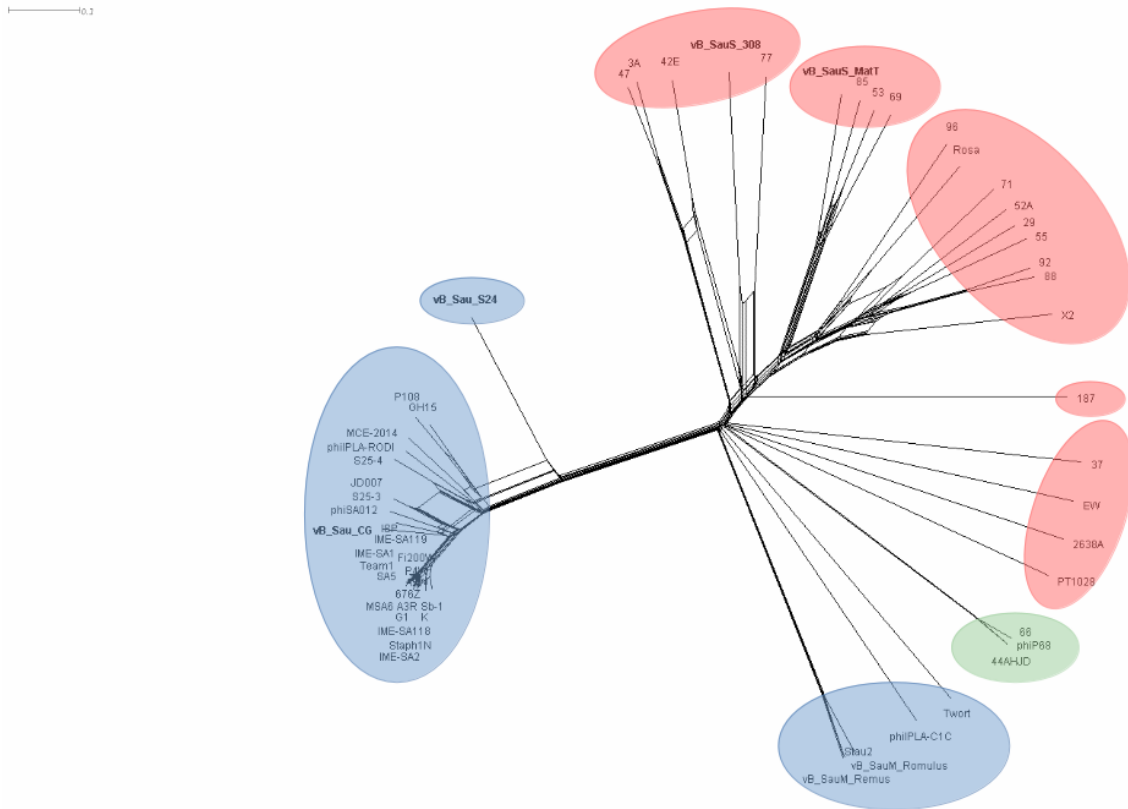
Los bacteriófagos vB\_Sau\_CG, vB\_Sau\_S24, vB\_SauS\_308 y vB\_SauS\_MatT fueron comparados con 54 fagos obtenidos de la base de datos del NCBI que infectan bacterias de la familia *Staphylococcus aureus* con ADN doble cadena lineal. Se realizó la clasificación de ellos según lo propuesto por Pelletier *et al.* (Kwan, Liu, DuBow, Gros, & Pelletier, 2005) en tres clases basándose en el tamaño de sus genomas (Clase I: <20 Kb; Clase II: ≈40 Kb; Clase III: >125 Kb), observándose una correlación entre ésta clasificación y la morfológica, siendo los fagos de la familia Podoviridae aquellos con los menores genomas (Clase I) 66, 44AHJD, phiP68 y PT1028; los Myoviridae con los mayores (Clase III) 676Z, A3R, A5W, Fi200W, G1, GH15, IME-SA1, IME-SA118, IME-SA119, IME-SA2, ISP, JD007, K, MCE-2014, MSA6, P108, P4W, phiPLA-C1C, phiPLA-RODI, phiSA012, S25-3, S25-4, SA5, Sb-1, Staph1N, Stau2, Team1, Twort, vB\_SauM\_Remus, vB\_SauM\_Romulus, vB\_Sau\_CG, y vB\_Sau\_S24; y los Siphoviridae (Clase II) 29, 37, 47, 53, 55, 69, 71, 77, 85, 88, 92, 96, 187, 2638A, 3A, 42e, 52A, EW, ROSA, X2, vB\_SauS\_308 y vB\_SauS\_MatT los intermedios.

Las secuencias nucleotídicas de los 58 fagos fueron comparadas mediante el uso del software Gegenees (Ågren *et al.*, 2012), previo arreglo de las secuencias ubicando en el primer ORF la unidad mayor de la terminasa. Esta comparación (Fig. 6) del tipo todos contra todos demostró que, en general, los fagos pertenecientes a la misma clase (I, II, III) poseen secuencias nucleotídicas relacionadas, no así, aquellos pertenecientes a otras clases.



**Figura 6:** Heat map (producido con Bioconductor - función heatmap()) para la comparación del tipo fragmentada todos vs. todos

La red filogenética (Fig. 7) mostró la formación de diversos clados, ubicándose los fagos de interés en distintos grupos en concordancia al los niveles de identidad que mostraba el análisis con ClustalW: 84% entre vB\_Sau\_S24 y vB\_Sau\_CG y 15% entre vB\_SauS\_308 y vB\_SauS\_MatT. Hacia derecha de la imagen se puede observar cómo agrupan los fagos pertenecientes a la familia Podoviridae y Siphoviridae, mientras que los Myoviridae tienden a localizarse a la izquierda, con la excepción de los fagos Twort, philPLA-C1C, Stau2, vB\_SauM\_Remus y vB\_SauM\_Romulus (Vandersteegen et al., 2013).



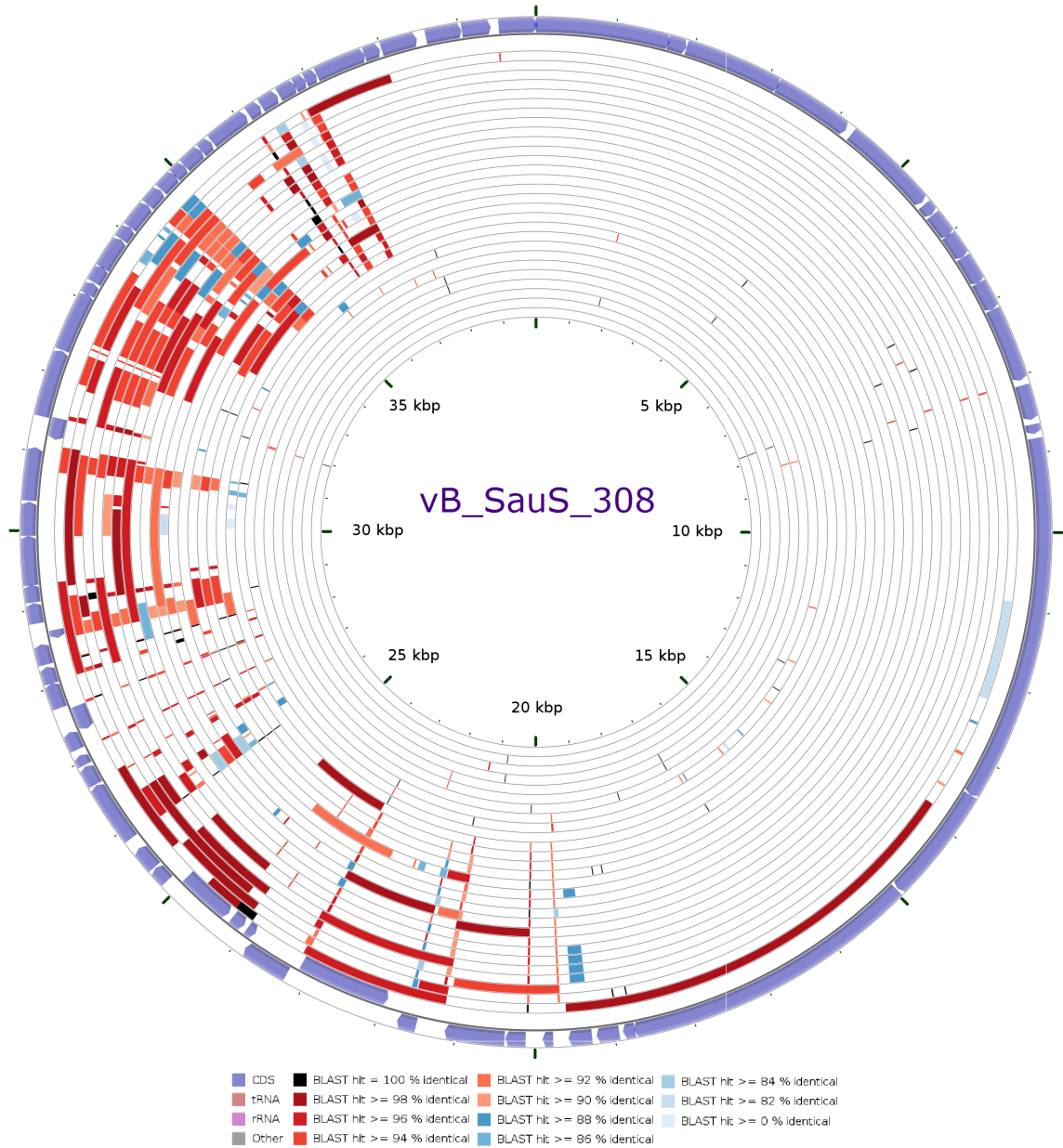
**Figura 7:** Análisis de la red de bacteriófagos Caudovirales con SplitsTree4. Cincuenta y cuatro genomas disponibles en la base de datos fueron alineados con Gegenees y la red armada con SplitsTree4 bajo el método Neighbor-Net. Se observa como los fagos de interés agrupan en distintos clusters (Azul: Myoviridae, Rojo: Siphoviridae, Verde: Podoviridae)

En virtud de lograr una mejor presentación de los datos se procede al análisis de los fagos de interés según pertenezcan a la familia Myoviridae o Siphoviridae.

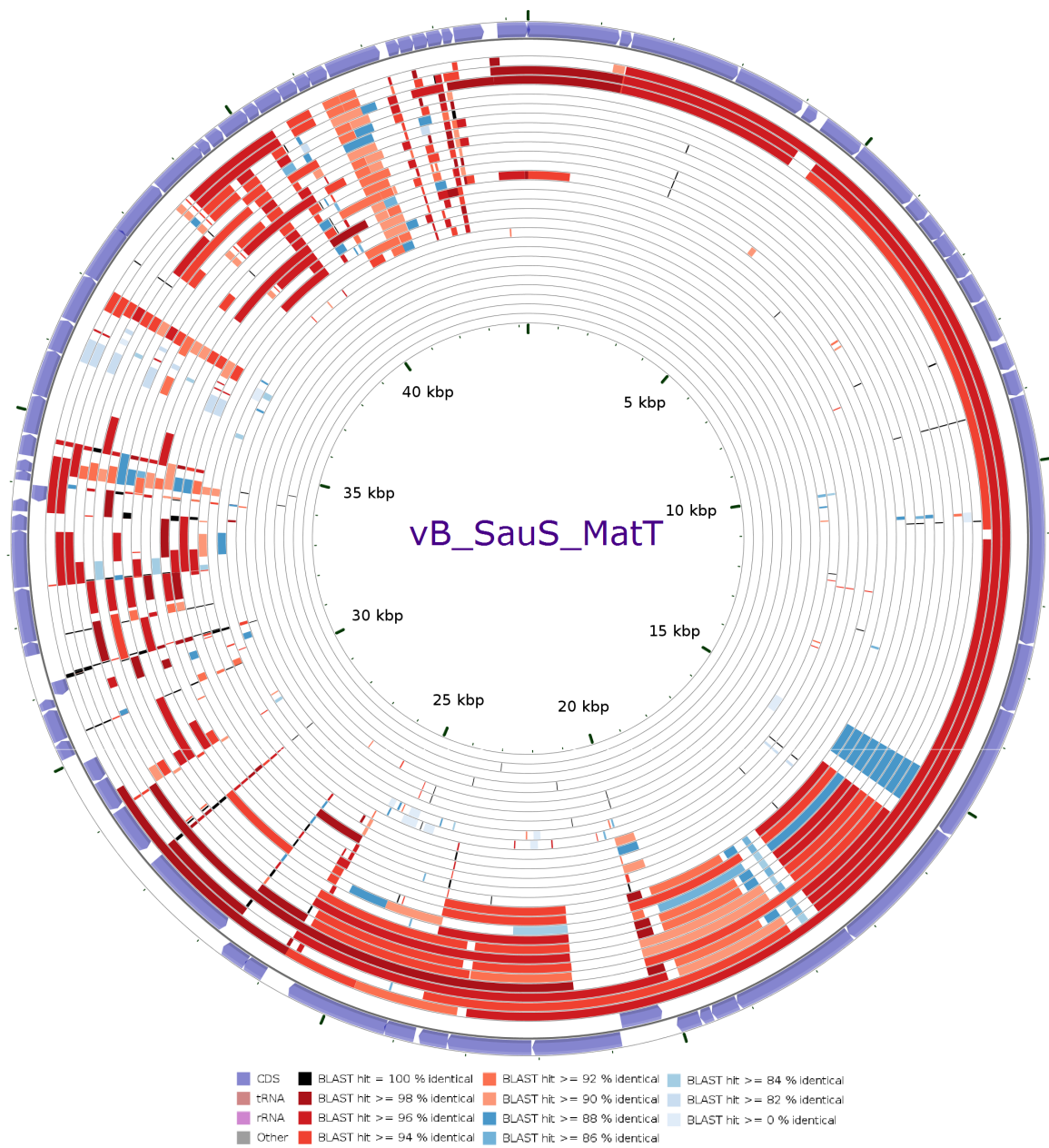
## Análisis comparativo de fagos Siphoviridae

Pese a que los bacteriófagos exhiben, en general, una organización modular similar y poseen genes con funciones conservadas, los mismos presentan una gran variación en su secuencia nucleotídica (Fig. 8 y 9). Es por esto que desde hace tiempo se intenta buscar un nuevo sistema de clasificación que no dependa exclusivamente del alineamiento de los genomas sino en la elección de una (o varias) proteína representativa de al menos dos módulos y el análisis de su

agrupación (Kahánková et al., 2010; Goerke et al., 2009). En el presente trabajo se optó por analizar los módulos de lisogenia, lisis, replicación del ADN y los factores de virulencia, los resultados se resumen en la Tabla 2.



**Figura 8:** Mapa CGView que compara el fago vB\_SauS\_308 con otros fagos Caudovirales. El anillo mas externo corresponde a vB\_SauS\_308, utilizado como genoma de referencia. Los anillos internos corresponden al alineamiento con BLASTN con cada genoma analizado, variando su color según el porcentaje de similitud.



**Figura 9:** Mapa CGView que compara el fago vB\_SauS\_MatT con otros fagos Caudovirales. El anillo mas externo corresponde a vB\_SauS\_MatT, utilizado como genoma de referencia. Los anillos internos corresponden al alineamiento con BLASTN con cada genoma analizado, variando su color según el porcentaje de similitud.

**Tabla 2:** Clasificación *in-silico* de genomas *Siphoviridae* según su organización modular basados en las proteínas integrasa y holina y en los factores de replicación (*polA*, *dnaC* or *dnaD*).

Bacteriófago	Grupos Integrasa	Longitud Holina (pb)	Factores de Replicación	Factores de Virulencia
3A	-	303	PolA	
47	T-Int1	303	PolA	
77	T-Int1	303	dnaC	
ROSA	T-Int1	303	dnaC	
<b>vB_SauS_308</b>	T-Int1	255	dnaC	<i>sak, PVL</i>
52A	T-Int1	438	dnaC	
53	S-Int2	303	dnaC	
85	S-Int2	438	dnaC	<i>Pathogenicity island SaPin2, PVL</i>
<b>vB_SauS_MatT</b>	S-Int2	438	dnaC	
92	S-Int2	438	dnaC	
X2	S-Int2	438	dnaD	
2638A	T-Int3	249	PolA	
42e	T-Int3	255	dnaD	
37	T-Int4	423	dnaC	
EW	T-Int5	435	dnaD	
29	T-Int6	438	dnaD	
69	T-Int6	438	dnaC	
88	T-Int6	438	dnaC	
187	T-Int6	255	dnaC	
55	T-Int7	438	dnaC	
71	T-Int7	414	dnaD	<i>ear</i>
96	T-Int8	303	dnaD	

Los mismos tipos se colorean de igual forma. Se nombran los grupos de Int1 a Int8 anteponiendo "S" o "T" según se trate de Serina-integrasas o Tirosina-integrasas.

## Módulo de lisogenia

Este grupo de genes se encuentra presente en fagos temperados, los cuales tienen la capacidad de integrarse en la bacteria huésped y se mantienen integrados en el genoma de dicha bacteria hasta que son estimulados para ingresar en el ciclo lítico. Las proteínas encargadas de esta función, motivo por el cual fueron elegidas para el análisis, son las integrasas, proteínas bien conservadas que se clasifican en dos grupos mayores Serina-integrasas y Tirosina-integrasas (Fogg, Colloms, Rosser, Stark, & Smith, 2014). Del análisis se excluyó al fago 3A debido a la ausencia de la proteína. Se procedió al alineamiento con ClustalW y a la esquematización del árbol filogenético circular en donde se observa la formación de ocho subgrupos (ver Fig.

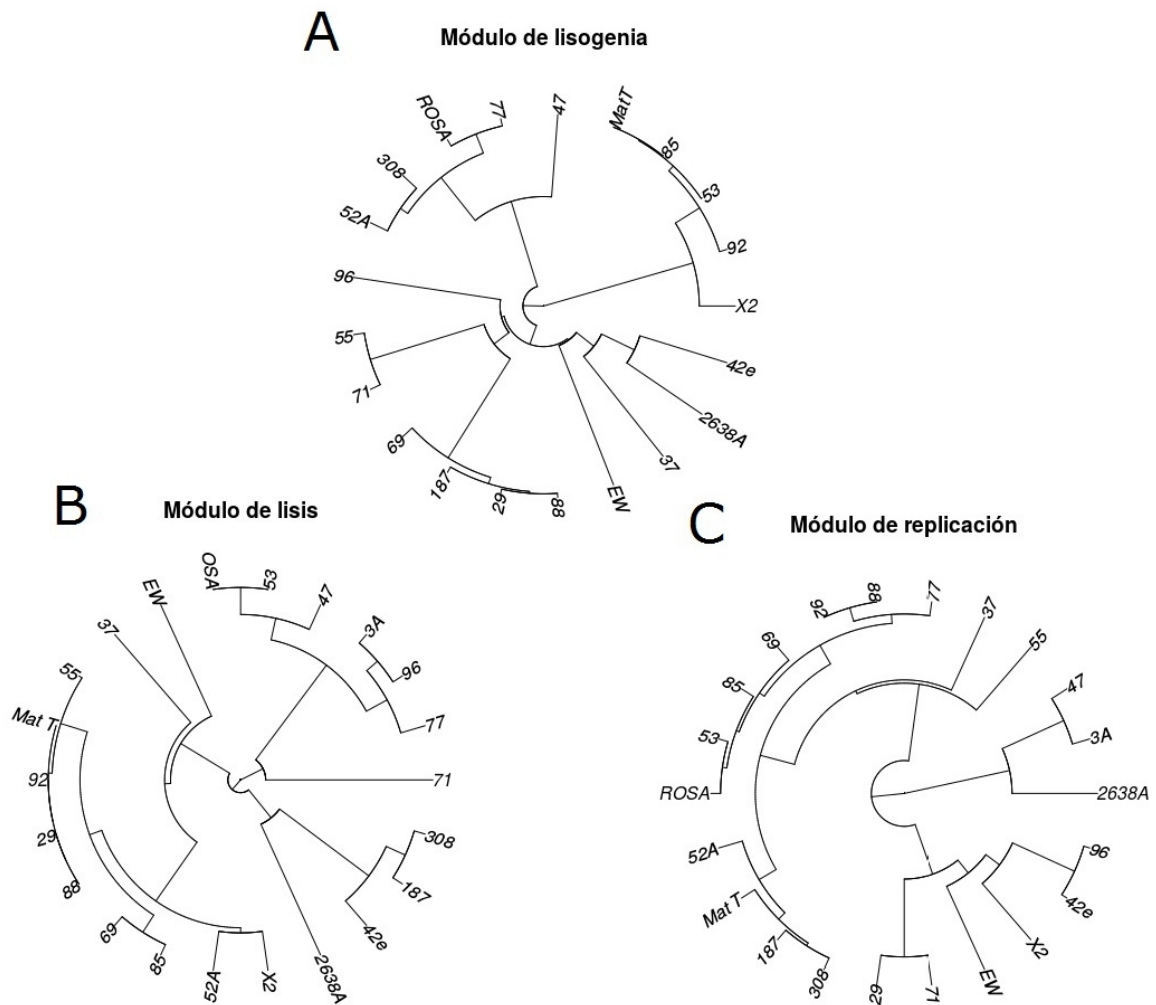
11-A). Los fagos con una identidad mayor al 70% fueron ubicados en el mismo grupo, otorgando una variación de este porcentaje de 76 y el 100% en fagos pertenecientes al mismo grupo. Los grupos pertenecientes a la misma raíz poseen una identidad del 50% aproximadamente, mientras que los que pertenecen a distintas no alcanzan el 24%.

## Módulo de lisis

Este módulo codifica dos proteínas, holina y endolisina, la primera es capaz de degradar la membrana celular del huésped dando fin al ciclo lítico (Mesnage & Foster, 2013). Su clasificación se basa en el polimorfismo de longitud del gen, que varía desde las 255 pb. hasta las 438 pb. Se puede observar que presentan siete longitudes distintas, sin embargo, se las clasifica en cinco grupos debido a que los fagos EW y 37 presentan deleciones que le confieren una homología del 64-66% con el resto del grupo. El análisis de las secuencias proteicas detecta cuatro mutaciones puntuales en las holinas “255” y “303”, y catorce en la holina “438”.

## Módulo de replicación del ADN

Mediante el programa Phamerator (Cresawn et al., 2011) se analizaron los módulos de replicación de los 22 fagos. Se tuvo en cuenta la presencia de los genes: ADN polimerasa A (polA), dnaC y dnaD (Kahánková et al., 2010; Weigel & Seitz, 2006). Se pudo observar la formación de cinco familias, o Phams según el programa, en donde cada gen formaba una, con la excepción de dnaD que formaba tres familias (a: X2, 45e y 96; b: EW; c:27 y71). Se ahondó en la diferencia que llevó a esto, estudiando tanto la longitud de la secuencia proteica como su alineamiento, y viendo que aquellas que conformaban la misma Pham, y por lo tanto compartían mayor porcentaje de identidad, también poseían longitudes similares (a: entre 251 y 254 AA, b: entre 265 y 268 AA y c: 234 AA).

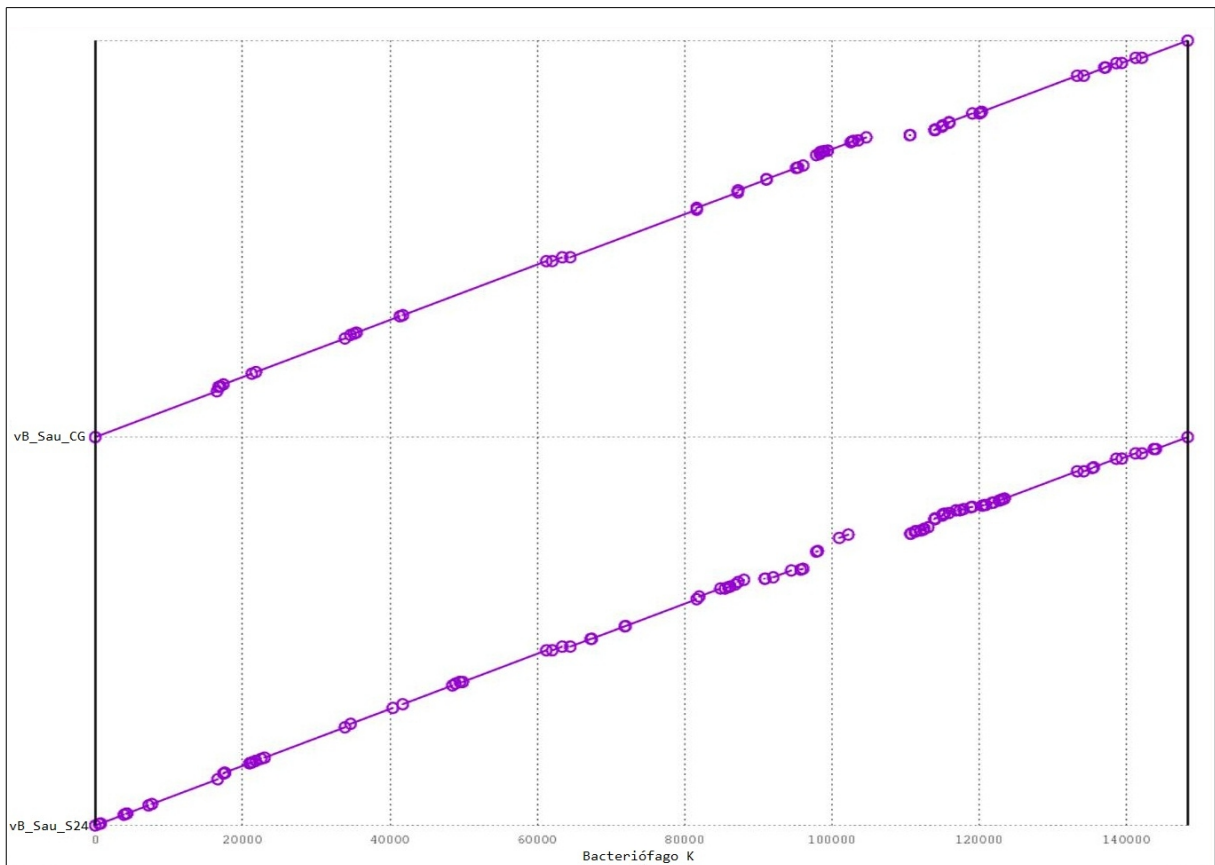


**Figura 10:** Árboles filogenéticos circulares que ilustran la similitud de las secuencias nucleotídicas de tres módulos de fagos: (A) módulo de lisogenia; (B) módulo de lisis; y (C) módulo de replicación del ADN. Los nombres de los fagos se encuentran escritos a continuación de cada rama. Los genomas fueron alineados con ClustalW y los gráficos realizados en R software.

## Análisis comparativo de fagos Myoviridae

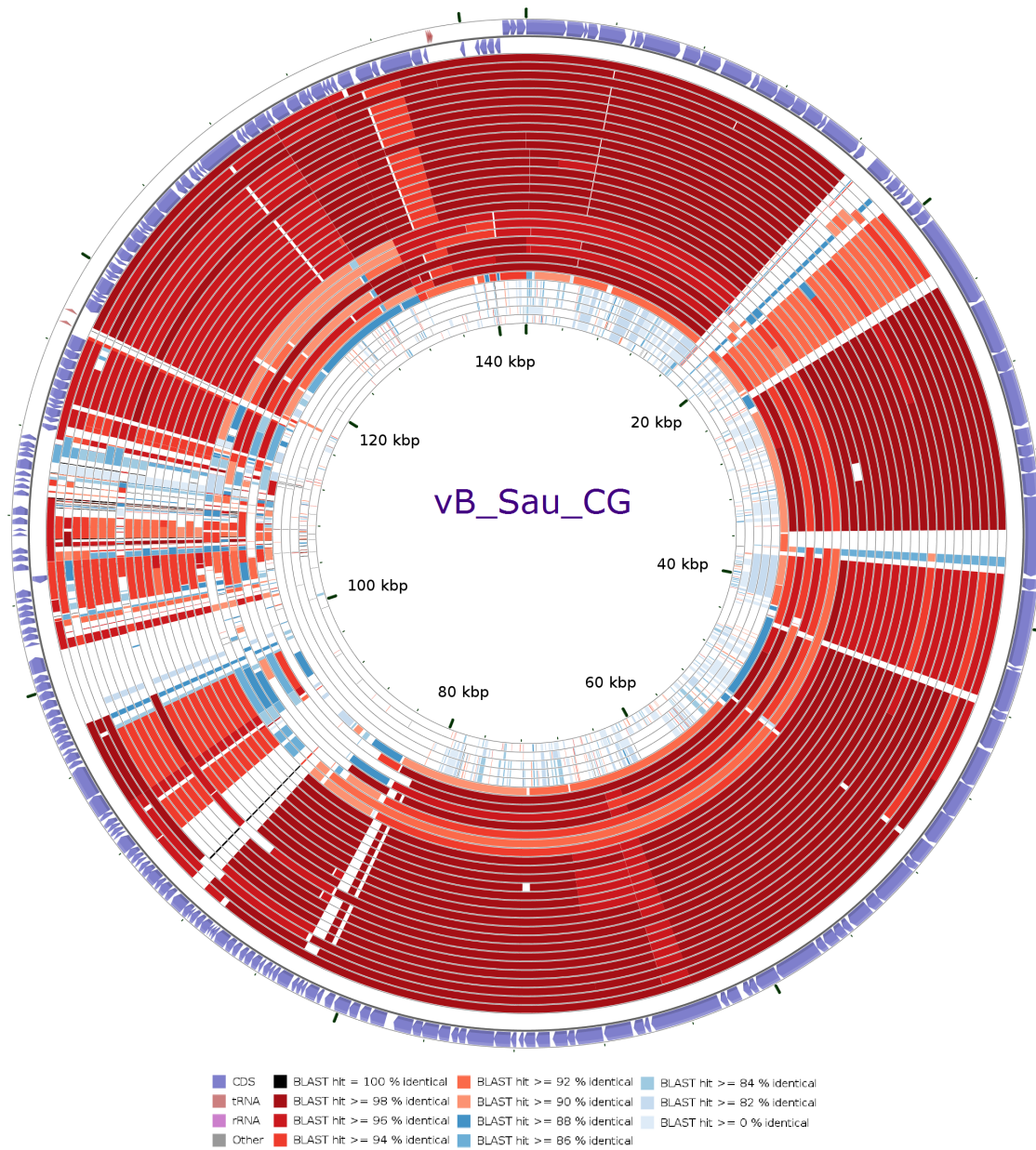
Son los fagos pertenecientes a la familia Myoviridae aquellos con más regiones conservadas, es por esto que para la comparación de vB\_Sau\_S24 y vB\_Sau\_CG se utilizó el fago K, que es uno de los miembros más caracterizados de esta familia (Alves et al., 2014; O'Flaherty et al., 2004; Sanz-Gaitero, Keary, Garcia-Doval, Coffey, & van Raaij, 2014).

Los resultados de los alineamientos de estos fagos arrojaron altos niveles de identidad (91% y 82%, respectivamente). Utilizando el pipeline NUCmer del paquete MUMmer 3.2 (Delcher et al., 2002; Kurtz et al., 2004) se llevó a cabo un alineamiento punto a punto; los resultados mostrados con un dotplot (Fig. 12) revelan una extensa homología en la secuencias, con diferencias en la repetición terminal larga (RTL) representadas por el gap.

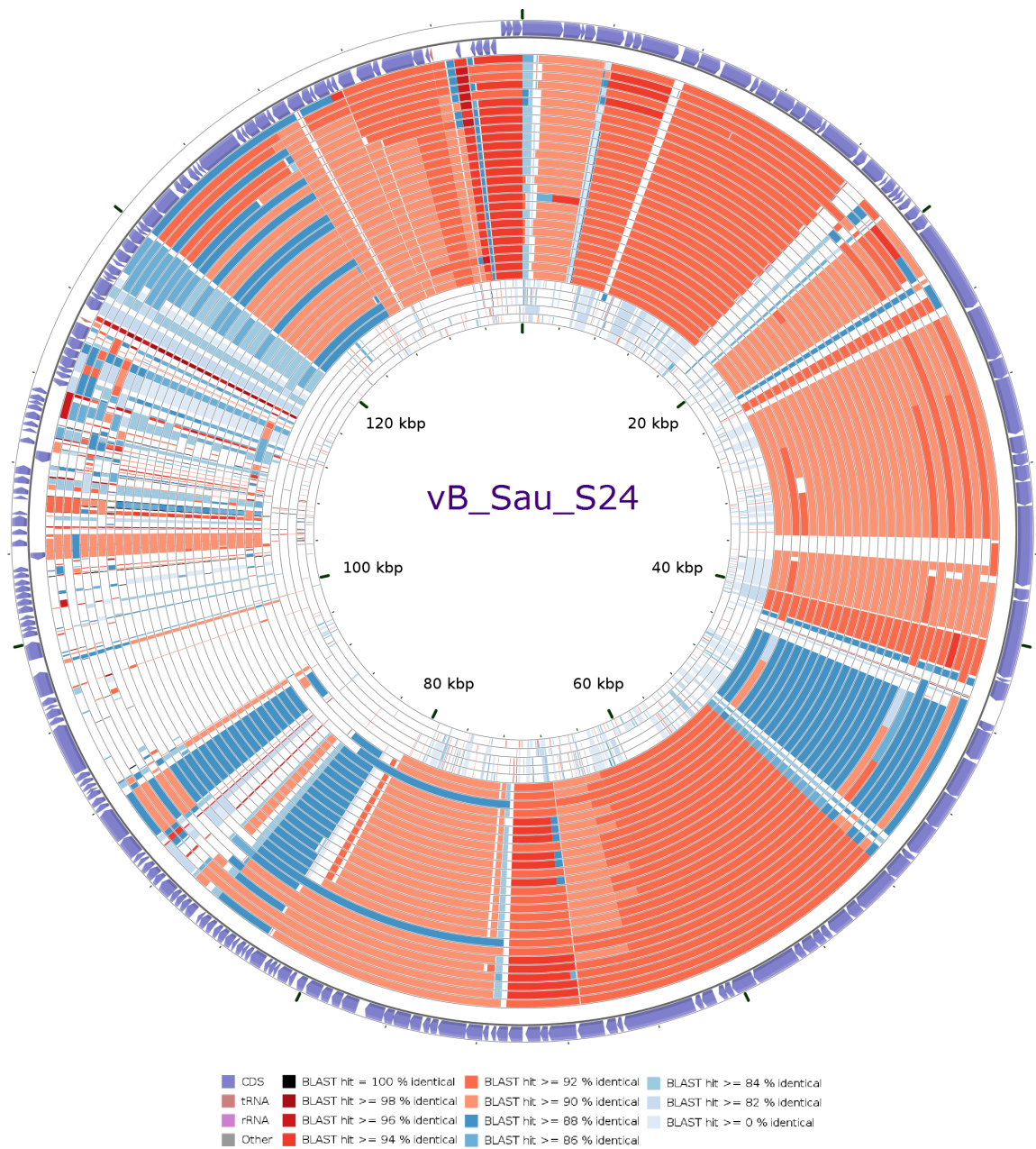


**Figura 11:** Alineamiento dot-plot utilizando NUCmer. En el eje X se representa la secuencia nucleotídica del genoma del bacteriófago K y en el eje Y los genomas de vB\_Sau\_CG y vB\_Sau\_S24.

Para completar el análisis se compararon los dos fagos de interés con todos los pertenecientes a la familia Myoviridae a través de CGViewer Comparison Tool (Fig. 13 y 14) utilizando un BLASTN. Se pudo ver que ambos fagos comparten zonas de gran similitud con la mayoría de los fagos alineados (zonas oscuras) y, en concordancia con el resultado arrojado por NUCmer, menos similitud en la región RTL (ORFs vBSauCG\_143 a vBSauCG\_167 y vBSauS24\_128 a vBSauS24\_158).



**Figura 12:** Mapa CGView que compara el fago vB\_SauS\_CG con otros fagos Caudovirales. El anillo mas externo corresponde a vB\_SauS\_CG, utilizado como genoma de referencia. Los anillos internos corresponden al alineamiento con BLASTN con cada genoma analizado, variando su color según el porcentaje de similitud.



**Figura 13:** Mapa CGView que compara el fago vB\_SauS\_S24 con otros fagos Caudovirales. El anillo mas externo corresponde a vB\_SauS\_S24, utilizado como genoma de referencia. Los anillos internos corresponden al alineamiento con BLASTN con cada genoma analizado, variando su color según el porcentaje de similitud.

# Conclusiones

El estudio de bacteriófagos, tanto con fines comerciales como por la búsqueda de conocimiento, es un tema que motiva a investigadores en todas partes. Frecuentemente los mismos son realizados por institutos que se encuentran fuera del país, lo cual genera poca información de fagos con actividades anti-estafilocócicas. Sin embargo, en nuestro país la información local es escasa. Para ampliar los conocimientos con cepas locales se realizó el estudio de cuatro fagos obtenidos de muestras de distinta procedencia.

Los cuatro bacteriófagos con los que se trabajó pertenecen al orden Caudovirales, clasificados como Myoviridae o Siphoviridae por poseer genomas entre 139-150 Kpb y 39-45 Kpb, respectivamente.

Los fagos fueron clasificados, según su tamaño y la organización de sus genes en Clase III y Clase II. Teniendo en cuenta su relación filogenética y sus características generales se pudo comparar a nuestros fagos pertenecientes a la familia Myoviridae con el fago K, que es el más reportado en la literatura. Sin embargo, observamos que los fagos pertenecientes a la familia Siphoviridae presentan mayores diferencias con otros de su familia, lo cual es atribuible a su naturaleza lisogénica, ya que al integrarse muchas veces puede llevarse consigo genes pertenecientes al hospedador.

Los genomas vB\_Sau\_CG y vB\_Sau\_S24 no contienen genes que codifiquen proteínas asociadas a la virulencia o proteínas tóxicas, lo cual es deseable si en el futuro se los desea tomar en cuenta para aplicaciones biotecnológicas (biocontrol). Sin embargo, en los fagos temperados se encontró la presencia de los factores de virulencia: estafiloquinasa, leucocidina de Pantón-Valentine y una isla patogénica (SaPin2). Por otro lado, llevamos a cabo el análisis de la presencia de endolinas, que son encargadas junto con la holina de romper la pared celular para permitir la liberación de la progenie fágica. Este análisis demostró la presencia de estas

enzimas en todos los fagos. Interesantemente, la endolisina del fago vB\_SauS\_308 sólo contiene uno de los dominios activos CHAP (endopeptidasa) y no se encontró el dominio AMI (amidasa) ni el dominio de unión a la célula SH3, los cuales se encuentran presentes en la mayoría de las endolisinas de fagos de *S. aureus*. Razón por la cual se lo considera de interés para próximos estudios.

La caracterización según la organización modular de los fagos vB\_SauS\_308 y vB\_SauS\_MatT con otros fagos de la misma familia mostró un alto uso de las holinas del grupo “303” y “438” y del factor de replicación dnaC sobre otros. No pudo encontrarse una asociación entre la elección entre integrasa-holina o integrasa-factor de replicación. Pese a lo mencionado por otros autores (Kahánková et al., 2010), no se pudo atribuir ninguna relación entre el módulo de lisogenia y la presencia de factores de virulencia.

Es de destacar que pese a la baja homología entre los fagos Siphoviridae obtenidos de la base de datos y los del laboratorio, sus secuencias si presentan proteínas bien conservadas para las funciones estudiadas (integración, replicación y lisis).

## Bibliografía

(<http://rast.nmpdr.org/>). (n.d.). Retrieved from <http://rast.nmpdr.org/>

Abatángelo, V., Peressutti Bacci, N., Boncompain, C. A., Amadio, A. A., Carrasco, S., Suárez, C. A., & Morbidoni, H. R. (2017). Broad-range lytic bacteriophages that kill *Staphylococcus aureus* local field strains. *PLOS ONE*, *12*(7), e0181671. <https://doi.org/10.1371/journal.pone.0181671>

Ågren, J., Sundström, A., Håfström, T., & Segerman, B. (2012). Gegenees: Fragmented alignment of multiple genomes for determining phylogenomic distances and genetic signatures unique for specified target groups. *PLoS ONE*, *7*(6). <https://doi.org/10.1371/journal.pone.0039107>

Alves, D. R., Gaudion, A., Bean, J. E., Perez Esteban, P., Arnot, T. C., Harper, D. R., ... Jenkins, A. T. A. (2014). Combined use of bacteriophage K and a novel bacteriophage to reduce *Staphylococcus aureus* biofilm formation. *Applied and Environmental Microbiology*, *80*(21), 6694–6703. <https://doi.org/10.1128/AEM.01789-14>

Aziz, R. K., Bartels, D., Best, A., DeJongh, M., Disz, T., Edwards, R. A., ... Zagnitko, O. (2008). The RAST Server: Rapid annotations using subsystems technology. *BMC Genomics*, *9*, 1–15. <https://doi.org/10.1186/1471-2164-9-75>

Bertelli, C., Laird, M. R., Williams, K. P., Lau, B. Y., Hoad, G., Winsor, G. L., & Brinkman, F. S. L. (2017). IslandViewer 4: Expanded prediction of genomic islands for larger-scale datasets. *Nucleic Acids Research*, *45*(W1), W30–W35. <https://doi.org/10.1093/nar/gkx343>

Besemer, J. (2001). GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Research*, *29*(12), 2607–2618. <https://doi.org/10.1093/nar/29.12.2607>

biopython.org. (n.d.). Retrieved from <https://biopython.org>

Bryant, D., & Moulton, V. (2004). Neighbor-Net: An Agglomerative Method for the Construction of Phylogenetic Networks. *Molecular Biology and Evolution*, *21*(2), 255–265. <https://doi.org/10.1093/molbev/msh018>

Burrowes, B., Harper, D. R., Anderson, J., McConville, M., & Enright, M. C. (2011). Bacteriophage therapy: Potential uses in the control of antibiotic-resistant pathogens. *Expert Review of Anti-Infective Therapy*. <https://doi.org/10.1586/eri.11.90>

- Carver, T. J., Rutherford, K. M., Berriman, M., Rajandream, M. A., Barrell, B. G., & Parkhill, J. (2005). ACT: The Artemis comparison tool. *Bioinformatics*, *21*(16), 3422–3423. <https://doi.org/10.1093/bioinformatics/bti553>
- Chang, J., Chapman, B., Friedberg, I., Hamelryck, T., Hoon, M. De, Cock, P., ... Wilczy, B. (2010). Biopython Tutorial and Cookbook. *Update*, *2010*(November), 15–19. <https://doi.org/10.1145/360262.360268>
- Chirico, N., Vianelli, A., & Belshaw, R. (2010). Why genes overlap in viruses. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1701), 3809–3817. <https://doi.org/10.1098/rspb.2010.1052>
- Cresawn, S. G., Bogel, M., Day, N., Jacobs-Sera, D., Hendrix, R. W., & Hatfull, G. F. (2011). Phamerator: A bioinformatic tool for comparative bacteriophage genomics. *BMC Bioinformatics*, *12*(1), 395. <https://doi.org/10.1186/1471-2105-12-395>
- Darling, A. C. E., Mau, B., Blattner, F. R., & Perna, N. T. (2004). Mauve : Multiple Alignment of Conserved Genomic Sequence With Rearrangements Mauve : Multiple Alignment of Conserved Genomic Sequence With Rearrangements, 1394–1403. <https://doi.org/10.1101/gr.2289704>
- Deghorain, M., & Van Melderren, L. (2012). The staphylococci phages family: An overview. *Viruses*, *4*(12), 3316–3335. <https://doi.org/10.3390/v4123316>
- Delcher, A. L., Phillippy, A., Carlton, J., & Salzberg, S. L. (2002). Fast algorithms for large-scale genome alignment and comparison. *Nucleic Acids Research*, *30*(11), 2478–2483. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12034836>
- Fickett, J. W., & Guigó, R. (1993). Estimation of protein coding density in a corpus of DNA sequence data. *Nucleic Acids Research*, *21*(12), 2837–2844. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/8332493>
- Fogg, P. C. M., Colloms, S., Rosser, S., Stark, M., & Smith, M. C. M. (2014). New applications for phage integrases. *Journal of Molecular Biology*, *426*(15), 2703–2716. <https://doi.org/10.1016/j.jmb.2014.05.014>
- Gardete, S., & Tomasz, A. (2014). Mechanisms of vancomycin resistance in *Staphylococcus aureus*. *The Journal of Clinical Investigation*, *124*(7), 2836–2840. <https://doi.org/10.1172/JCI68834.2836>
- Gentleman, R., Carey, V., Bates, D., Bolstad, B., Dettling, M., Dudoit, S., ... Zhang, J. (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology*, *5*(10), R80. <https://doi.org/10.1186/gb-2004-5-10-r80>

- Gentleman, R., Carey, V. J., Huber, W., Irizarry, R. A., & Dudoit, S. (2005). *Bioinformatics and computational biology solutions using R and Bioconductor. Statistics for biology and health*. <https://doi.org/10.1016/B978-0-12-387667-6.00013-0>
- Goerke, C., Pantucek, R., Holtfreter, S., Schulte, B., Zink, M., Grumann, D., ... Wolz, C. (2009). Diversity of prophages in dominant *Staphylococcus aureus* clonal lineages. *Journal of Bacteriology*, *191*(11), 3462–3468. <https://doi.org/10.1128/JB.01804-08>
- Grant, J. R., & Stothard, P. (2008). The CGView Server: a comparative genomics tool for circular genomes. *Nucleic Acids Research*, *36*(Web Server issue), 181–184. <https://doi.org/10.1093/nar/gkn179>
- <http://bioperl.org/>. (n.d.). Retrieved from <http://bioperl.org/>
- <http://phagesdb.org/DNAMaster/>. (n.d.). Retrieved from <http://phagesdb.org/DNAMaster/>
- <https://github.com/>. (n.d.). Retrieved from <https://github.com/>
- Huson, D. H., & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, *23*(2), 254–267. <https://doi.org/10.1093/molbev/msj030>
- Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., ... Hunter, S. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics*, *30*(9), 1236–1240. <https://doi.org/10.1093/bioinformatics/btu031>
- Kahánková, J., Pantůček, R., Goerke, C., Růžicková, V., Holochová, P., & Doškař, J. (2010). Multilocus PCR typing strategy for differentiation of *Staphylococcus aureus* siphoviruses reflecting their modular genome structure. *Environmental Microbiology*, *12*(9), 2527–2538. <https://doi.org/10.1111/j.1462-2920.2010.02226.x>
- Kropinski, A. M., Prangishvili, D., & Lavigne, R. (2009). Position paper: The creation of a rational scheme for the nomenclature of viruses of Bacteria and Archaea. *Environmental Microbiology*, *11*(11), 2775–2777. <https://doi.org/10.1111/j.1462-2920.2009.01970.x>
- Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., & Salzberg, S. L. (2004). Versatile and open software for comparing large genomes. *Genome Biology*, *5*(2), R12. <https://doi.org/10.1186/gb-2004-5-2-r12>

- Langille, M. G. I., Hsiao, W. W. L., & Brinkman, F. S. L. (2010). Detecting genomic islands using bioinformatics approaches. *Nature Reviews Microbiology*, 8(5), 373–382. <https://doi.org/10.1038/nrmicro2350>
- Laslett, D., & Canback, B. (2004). ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Research*, 32(1), 11–16. <https://doi.org/10.1093/nar/gkh152>
- Lowe, T. M., & Chan, P. P. (2016). tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Research*, 44(W1), W54–W57. <https://doi.org/10.1093/nar/gkw413>
- McGuinness, W. A., Malachowa, N., & DeLeo, F. R. (2017). Vancomycin Resistance in *Staphylococcus aureus*. *The Yale Journal of Biology and Medicine*, 90(2), 269–281. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/28656013>  
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5482303>
- Medina, G., Egea, A. L., Otth, C., Otth, L., Fernández, H., Bocco, J. L., ... Sola, C. (2013). Molecular epidemiology of hospital-onset methicillin-resistant *Staphylococcus aureus* infections in Southern Chile. *European Journal of Clinical Microbiology & Infectious Diseases*, 32(12), 1533–1540. <https://doi.org/10.1007/s10096-013-1907-8>
- Mesnage, S., & Foster, S. J. (2013). N-Acetylmuramoyl-L-alanine Amidase. In *Handbook of Proteolytic Enzymes* (pp. 1401–1407). <https://doi.org/10.1016/B978-0-12-382219-2.00315-X>
- National Center for Biotechnology Information. (n.d.). Retrieved from <https://www.ncbi.nlm.nih.gov/>
- O’Flaherty, S., Coffey, A., Edwards, R., Meaney, W., Fitzgerald, G. F., & Ross, R. P. (2004). Genome of Staphylococcal Phage K: A New Lineage of Myoviridae Infecting Gram-Positive Bacteria with a Low G+C Content. *Journal of Bacteriology*, 186(9), 2862–2871. <https://doi.org/10.1128/JB.186.9.2862-2871.2004>
- Paul, V., Sundarajan, S., Rajagopalan, S., Hariharan, S., Kempashanaiah, N., Padmanabhan, S., ... Ramachandran, J. (2011). Lysis-deficient phages as novel therapeutic agents for controlling bacterial infection. *BMC Microbiology*, 11. <https://doi.org/10.1186/1471-2180-11-195>
- Pritchard, L., White, J. A., Birch, P. R. J., & Toth, I. K. (2006). GenomeDiagram: A python package for the visualization of large-scale genomic data. *Bioinformatics*, 22(5), 616–617. <https://doi.org/10.1093/bioinformatics/btk021>

- Russell, D. A., & Hatfull, G. F. (2015). *DNA Master Quick Start Guide*.
- Russell, D. A., & Hatfull, G. F. (2017). PhagesDB: The actinobacteriophage database. *Bioinformatics*, 33(5), 784–786. <https://doi.org/10.1093/bioinformatics/btw711>
- Sanz-Gaitero, M., Keary, R., Garcia-Doval, C., Coffey, A., & van Raaij, M. J. (2014). Crystal structure of the lytic CHAPK domain of the endolysin LysK from *Staphylococcus aureus* bacteriophage K. *Virology Journal*, 11(1), 133. <https://doi.org/10.1186/1743-422X-11-133>
- Tritt, A., Eisen, J. A., Facciotti, M. T., & Darling, A. E. (2012). An Integrated Pipeline for de Novo Assembly of Microbial Genomes. *PLoS ONE*, 7(9). <https://doi.org/10.1371/journal.pone.0042304>
- Vandersteegen, K., Kropinski, A. M., Nash, J. H. E., Noben, J.-P., Hermans, K., & Lavigne, R. (2013). Romulus and Remus, Two Phage Isolates Representing a Distinct Clade within the Twortlikevirus Genus, Display Suitable Properties for Phage Therapy Applications. *Journal of Virology*, 87(6), 3237–3247. <https://doi.org/10.1128/JVI.02763-12>
- Ventola, C. L. (2015). The antibiotic resistance crisis: part 1: causes and threats. *P & T: A Peer-Reviewed Journal for Formulary Management* (2015), 40(4), 277–283. <https://doi.org/Article>
- Weigel, C., & Seitz, H. (2006). Bacteriophage replication modules. *FEMS Microbiology Reviews*, 30(3), 321–381. <https://doi.org/10.1111/j.1574-6976.2006.00015.x>
- WHO. (2014). Antimicrobial resistance. Global report on surveillance. *World Health Organization*, 61(3), 383–394. <https://doi.org/10.1007/s13312-014-0374-3>
- WHO. (2018).

# ANEXOS

## Anexo I – *script* genomediagram.py

```
from Bio import SeqIO
A_rec = SeqIO.read("myphage.gb", "gb")

from reportlab.lib.colors import grey, aquamarine, gold, yellowgreen, coral, teal, navy,
lightblue, pink, magenta, indigo

A_colors = [magenta*1 + [grey]*4 + [indigo]*1 + [grey]*1 + [magenta]*2 + [grey]*1 +
[magenta]*1 + [coral]*2 + [grey]*4 + [grey]*4 + [grey]*3 + [grey]*5 + [gold]*1 +
[teal]*2

from Bio.Graphics import GenomeDiagram

name = "myphagename"
gd_diagram = GenomeDiagram.Diagram(name)
max_len = 0
for record, gene_colors in zip([A_rec], [A_colors]):
    max_len = max(max_len, len(record))
    gd_track_for_features = gd_diagram.new_track(1,
        name=record.name,
        greytrack=True,
        start=0, end=len(record))
    gd_feature_set = gd_track_for_features.new_set()

    i = 0
    for feature in record.features:
        if feature.type != "gene":
            #Exclude this feature
            continue
        gd_feature_set.add_feature(feature, sigil="ARROW",
            color=gene_colors[i], arrowshaft_height=1.0, label=True,
            name = str(i+1),
            label_position="start",
            label_size = 5, label_angle=45)

        i+=1

gd_diagram.draw(format="linear", pagesize='A4', fragments=2,
    start=0, end=max_len)
gd_diagram.write(name + ".pdf", "PDF")
gd_diagram.write(name + ".png", "PNG")
```

## Anexo II – *script heatmap.r*

```
source("http://bioconductor.org/biocLite.R")
biocLite("DESeq")
library("DESeq")
install.packages("gplots")

#leo los datos Gegenees
file <- read.table("/ubicacion/myarchivo.tab", header = TRUE, row.names=1)

#data frame a matrix numerica
data_matrix <- data.matrix(file)

#graficar
heatmap.2(data_matrix, Rowv = FALSE, Colv = FALSE, dendrogram = 'none')
```

## Anexo III – *script circularfilo.r*

```
library (ape)

#leo los datos, los necesito como una matrix
data <- read.dna("/ubicacion/myarchivo.phylip")
mydata <- as.matrix(data)

#calculo de distancias y formacion clusters
mydist <- dist(mydata)
mycluster <- hclust(mydist)

#grafico
plot(as.phylo(mycluster), type="fan", main="Módulo_nombre")
```

## Anexo IV – *script holinlength.pl*

```
#!/usr/bin/perl

use strict; use warnings;
use Bio::SeqIO;
use Bio::Seq;

# gets a sequence from a file
my $seqio = Bio::SeqIO->new( '-format' => 'fasta' , -file => 'myarchivo.fasta');
my $seqobj = $seqio->next_seq();

print "Seq_id: ", $seqobj->id(), " length: ", $seqobj->length(),"\n";
```

## Anexo V – Anotación de genomas

**Tabla 3:** Anotación de genomas de bacteriófagos Myoviridae

	<b>Función putativa</b>	<b>Dirección</b>	<b>vB_Sau_CG</b>	<b>vB_Sau_S24</b>
<b>Empaquetamiento del ADN y morfogénesis de la cabeza</b>	Terminase large subunit	+	1	1
	Portal protein	+	8	8
	Pro-head protease	+	9	9
	Major capsid protein	+	11	11
	Capsid protein	+	14	14
<b>Morfogénesis de la cola</b>	Major tail sheath protein	+	18	18
	Tail tube protein	+	19	19
	Tail morphogenetic protein (TmpB)	+	27	28
	Tail tape measure	+	28	29
	Tail murein hydrolase (TAME)	+	29	30
	Peptidoglycan Hydrolase (TmpE)	+	30	31
	Baseplate wedge subunit (BmpA)	+	34	35
	Baseplate morphogenetic protein (BmpB)	+	35	36
	Tail morphogenetic protein (TmpF)	+	36	37
	Baseplate morphogenetic protein (BmpC)	+	37	38
	Adsorption-associated tail protein (TmpG)	+	38	39
	Tail fiber protein	+	42	43
	Tail morphogenetic protein (TmpH)	+	69	70
	Tail morphogenetic protein (TmpI)	+	70	71
	<b>Replicación del</b>	DNA helicase A	+	43
Rep protein		+	44	45

<b>ADN</b>	DNA helicase B	+	45	46
	Recombination exonuclease A	+	46	48
	Recombination exonuclease B	+	48	49
	DNA primase	+	50	51
	Resolvase	+	53	54
	DNA polymerase A	+	61	62
	Repair recombinase	+	65	66
	RNA polymerase sigma factor	+	67	68
	DNA sliding clump inhibitor	+	87	93
<b>Regulación de la transcripción</b>	Anti-sigma factor	+	49	50
	Ribonucleotide reductase, stimulatory protein	+	54	55
	Ribonucleotide reductase, large subunit	+	55	56
	Ribonucleotide reductase, minor subunit	+	56	57
	Ribose-phosphate pyrophosphokinase	+	125	112
	RNA ligase	+	-	121
	Nucleoside triphosphate pyrophosphohydrolase	-	196	181
	Nucleoside 2-deoxyribosyltransferase	-	203	188
	Ribonuclease	-	207	192
<b>Lisis</b>	Endolysin (N-acetylmuramoyl-L-alanine amidase)	-	216	200
	Holin	-	217	201
<b>Funciones adicionales</b>	Putative intron-encoded nuclease	+	20	20
	Glycerophosphoryl diester phosphodiesterase	+	31	32
	Thioredoxin	+	58	59
	DNA binding protein	+	60	61
	Ig like protein	+	68	69
	Metallophosphoesterase	+	73	74
	RuBisCO small subunit	+	90	91
	Nicotinamide phosphoribosyl transferase (NadV)	+	126	113
HNH homing endonuclease	-	-	148	

	BofL	-	167	158
	Serine/Threonine protein phosphatase	-	176	163
	AAA family ATPase	-	193	178
	PhoH related protein	-	205	190
	Transcriptional regulator	-	211	195
	Transglycosylase	-	213	197
	tRNA-Asp		+	+
	tRNA-Arg		-	+
	tRNA-Phe		+	-
	tRNA-Trp		+	-
	tRNA-Met		+	-
	tRNA-His		+	-

**Tabla 4:** Anotación de genomas de bacteriófagos Siphoviridae

	<b>Función putativa</b>	<b>Dirección</b>	<b>vB_SauS_308</b>	<b>vB_SauS_MatT</b>
<b>Lisogenia</b>	Integrasa	-	1	1
	CI-like repressor	-	5	4
	Cro-like repressor	+	6	-
	Transcriptional regulator	+	-	5 y 9
	Antirepressor	+	8	10 y 11
<b>Replicación del ADN</b>	Mu Gam like protein	+	-	18
	SS-DNA binding protein	+	15 y 20	19 y 20
	HNHc nuclease	+	21	21
	Replication initiation protein	+	23	23
	DNA replication protein dnaC	+	24	24
	DNA helicase	+	-	27
	dUTPase	+	35	33
	Transcriptional activator RinB	+	39	37
	Transcriptional activator RinA	+	-	39
	HNH endonuclease	+	45	-
	Phage transcriptional terminator	+	-	48
<b>Empaquetamiento del ADN y morfogénesis de</b>	Terminase small subunit	+	46	40
	Terminase large subunit	+	47	41
	Portal protein	+	49	42-43
	Prohead serine protease	+	50	-

<b>la cabeza</b>	Capsid protein	+	51	-
	Phage head morphogenesis protein	+	-	44
	Phage capsid and scaffold	+	-	46
	Major capsid protein	+	-	47
	Phage head-tail connector protein (HTCP)	+	53 y 54	49
<b>Morfogénesis de la cola</b>	Pùtative tail component	+	-	51
	Tail tube protein	+	57	53
	Phage tail assembly chaperone protein	+	-	54
	Tail tape measure protein	+	60	56
	Tail protein	+	61	57
	Minor tail protein	+	62	-
	Endopeptidase		11	58
	Tail fiber protein	+	-	66
<b>Lisis</b>	Holin	+	66 y 67	68
	Endolysin (N-acetylmuramoyl-L-alanine amidase)	+	-	69
	CHAP domain	+	68	64
	Glucosaminidase	+	-	65
<b>Funciones adicionales</b>	Extracellular adherence protein (MAP domain)	+	-	8
	AP2 domain	+	-	22
	Coiled coil region	+	29	
	Transmembrane region	+	-	34
	Putative major teichoic acid biosynthesis protein C	+	-	59