



UNIVERSIDAD NACIONAL DE ROSARIO

FACULTAD DE CIENCIAS BIOQUÍMICAS Y FARMACEÚTICAS

“Bioinformática aplicada a la caracterización genómica y clasificación de bacteriófagos de *Staphylococcus aureus*”

Ma. Soledad Rita Moyano

Tesina para optar al título de Licenciada en Biotecnología

Director: Dr. Cristian Alejandro Suárez

Rosario, Argentina 2019

INDICE

1.	Abreviaturas	2
2.	Lenguaje especial	3
3.	Resumen	4
4.	Introducción	5
4.1.	<i>Staphylococcus aureus</i>	5
4.2.	Bacteriófagos	6
4.2.1.	Ciclo lisogénico.....	7
4.2.2.	Ciclo lítico.....	9
4.3.	Características generales de los bacteriófagos.	11
4.3.1.	Bacteriófagos de <i>Staphylococcus aureus</i>	13
4.3.2.	Características globales de los fagos de <i>S. aureus</i>	14
4.3.3.	Clasificación y evolución de la familia de fagos de <i>Staphylococcus</i>	15
4.3.4.	Uso de fagos de <i>Staphylococcus</i> para terapia fágica y otras aplicaciones biotecnológicas. 17	
5.	Objetivos	20
6.	Materiales y métodos	21
6.1.	Bacteriófagos utilizados.....	21
6.2.	Anotación genómica del bacteriófago I74 de SA	22
6.3.	Genómica comparativa	23
7.	Resultados	26
	CAPITULO 1	26
	7.1. Análisis genómico del nuevo bacteriófago de <i>Staphylococcus aureus</i> (I74) de la familia Siphoviridae.....	26
	7.2. Análisis genómico del Siphostaphylofago	27
	CAPITULO 2	33
7.3.	Análisis genómico comparativo del Siphostaphylofago	33
7.3.1.	Phamerator	33
7.3.2.	Análisis de la holina de I74	36
7.3.3.	Análisis CG view	37
7.3.4.	Análisis de la similitud de secuencia.....	39
7.3.5.	Análisis Gegenees	41
7.3.6.	<i>Staphylofagos</i> homólogos con I74	42
7.3.7.	Análisis del %GC de los Siphostaphylofagos	43
7.3.8.	Análisis filogenético de los Siphostaphylofagos	44
8.	Discusión	57
9.	Conclusión	60
10.	Bibliografía	61

1. Abreviaturas

Escisionasa (*xis*)

GC (guanina-citocina)

HIF (factor de integración del hospedador bacteriano)

ICTV (Comité internacional en Taxonomía de virus)

Integrasa (*int*)

kPb (kilopares de bases)

MRSA (*Staphylococcus aureus* resistentes a meticilina)

NCBI (Centro Nacional de Información Biotecnológica)

SA (*Staphylococcus aureus*)

SAPIs (Islas de patogenicidad de *Staphylococcus aureus*)

Sitio attachment del fago (*attP*)

Sitio attachment de la bacteria (*attB*)

Sitios híbridos del profago (*attL* y *attR*, izquierdo y derecho respectivamente)

Staphylococcus aureus (SA)

Staphylococcus coagulasa negativa (CoNS)

UV (ultravioleta)

Tapemeasure protein (TMP)

2. Lenguaje especial

ANI (identidad de nucleótido en promedio)

Core (hace referencia al núcleo, en este caso de una secuencia ubicada en el centro de otra secuencia)

Fis (Factor de estimulación de inversión, proteína bacteriana)

H (proteína de morfogénesis de la cápside)

Helper (hace referencia a un ayudante o auxiliar para una determinada acción)

Hit (hace referencia a la coincidencia de una secuencia de DNA o proteína con otras presentes en la base de datos)

Ma (algoritmo Manhattan)

MHP (proteína de la cápside mayor)

mHP (proteína de la cápside menor)

Gene content dissimilarities (GCD)

Port (proteína portal)

Prot (proteasa)

Script (archivo de órdenes)

Syphofagos (se refiere a los bacteriófagos pertenecientes a la familia de *Siphoviridae*)

Syphostaphylofagos (se refiere a los bacteriófagos de SA de la familia *Siphoviridae*)

TerS (terminasa subunidad pequeña)

TerL (terminasa subunidad grande)

Orphans (huérfanos)

3. Resumen

El incremento de la resistencia bacteriana (como en *Staphylococcus aureus*) por la presión selectiva que representa la utilización de antibióticos a gran escala, sobre todo en nuestros hospitales (también en sectores agrícolas y alimenticios), ha permitido la diseminación de cepas con mecanismos de resistencia que, en muchas ocasiones, nos dejan prácticamente sin alternativas para el tratamiento de las infecciones bacterianas; culminando en consecuencias mortales. Por eso, es necesario el desarrollo de nuevas terapias para el control de infecciones producidas por estos microorganismos.

Los bacteriófagos son virus que infectan bacterias de una manera altamente específica y tienen la capacidad de destruirlas. Por otro lado, los bacteriófagos poseen un rol crucial en patogénesis y virulencia debido a la capacidad que tienen algunos de ellos de favorecer la transferencia de genes. En la actualidad, gracias al aumento constante del número de bacteriófagos aislados y secuenciados, sumado al desarrollo de diversas herramientas para analizar dichos fagos, ha resurgido el interés para utilizar a los fagos con finalidades terapéuticas y/o diagnósticas.

Durante la presente tesina, se ha llevado a cabo la caracterización y análisis bioinformático de un bacteriófago específico de *Staphylococcus aureus* (fago I74, previamente aislado en el mismo laboratorio de trabajo), perteneciente a la familia Siphoviridae. Los análisis realizados comprenden, la anotación genómica y genómica comparativa utilizando fagos presentes en las bases de datos. Esta gran cantidad de información genética de numerosos bacteriófagos de *S. aureus*, permite un enfoque sistemático para estudiar la diversidad y evolución de los mismos y aportan datos sobre la transferencia de genes de virulencia entre las cepas bacterianas.

4. Introducción

4.1. *Staphylococcus aureus*

Staphylococcus aureus es un coco gram positivo que tiene la capacidad de causar infecciones tanto en humanos como animales. Este patógeno oportunista genera un amplio rango de infecciones tanto sistémicas como locales. Actualmente, a nivel mundial existe una creciente preocupación por el incremento en la resistencia a diferentes antimicrobianos de este patógeno. Siendo la resistencia a meticilina una de las más relevantes (SAMR), por lo que se requieren de medidas urgentes para prevenir las infecciones llevadas a cabo por este patógeno.

Una característica importante de *S. aureus* es su capacidad para codificar muchos factores de virulencia, los cuales facilitan la diseminación en los tejidos de los hospedadores y el escape al sistema inmune del mismo. Los mismos pueden estar codificados por elementos genéticos móviles o estar presentes en profagos integrados en el genoma de diferentes cepas de *S. aureus* (**Pantůcek et al. 2004; Goerke et al. 2009**).

La adquisición o la pérdida de bacteriófagos temperados durante la infección, incrementa la plasticidad genómica en cepas huéspedes, así facilita la adaptación del patógeno a varias condiciones (**Moore and Lindsay 2001; Resch et al. 2005; Goerke, Koller, and Wolz 2006**).

Unas de las causas de variación en el contenido de profagos en las cepas de *S. aureus*, se debe aparentemente a que el fago puede integrarse dentro del cromosoma huésped en al menos nueve sitios de inserción distintos. Estos sitios son reconocidos por la enzima denominada integrasa que es específica del profago. Los mismos tienen estructura modular y poseen mosaicismo debido principalmente a recombinación y transferencia horizontal de secuencias.

El estudio de los mecanismos de transferencia de los profagos, es por lo tanto de gran importancia, no solamente para la determinación de la alta variabilidad genómica en *S. aureus*, sino también desde las perspectivas genómicas globales para la elucidación de la evolución y la virulencia de las cepas bacterianas.

4.2. Bacteriófagos

Los bacteriófagos, que también se pueden denominar con el término abreviado de fagos, son virus que infectan bacterias. Estos parásitos intracelulares estrictos fueron descubiertos independientemente por Twort en 1915 y por d'Herelle en 1917 (**Ackermann and DuBow 1987**) y en la actualidad sabemos que son ubicuos en todos los ambientes (incluyendo el agua, el aceite y el aire) en los que sus hospedadores, las bacterias, proliferan (**Zhang et al. 2017**). Una característica destacable de estos agentes biológicos genéticamente diversos, es que poseen una alta especificidad por su hospedador.

Además, los fagos juegan un rol crítico no solamente en la adaptación y ecología de las bacterias para los nuevos ambientes, sino en la patogénesis de las infecciones bacterianas humanas (**Tony Kwan et al. 2005**).

Los bacteriófagos se pueden dividir en dos grandes grupos dependiendo de su ciclo de vida (**Figura 1**):

- Fagos temperados o avirulentos, en los que el DNA viral puede integrarse en el genoma de la bacteria infectada y replicarse junto a esta, transmitiéndose de una generación a otra sin que se produzca la lisis celular. Este fenómeno se conoce como lisogenia; a la cepa así infectada como lisógeno y al DNA viral integrado como profago. En determinadas condiciones (de temperatura, pH, agentes mutagénicos, etc), el profago puede comenzar un ciclo lítico que conducirá a la lisis celular.
- Fagos líticos o virulentos, que provocan la lisis de las células hospedadoras, liberando una nueva progenie fágica en un proceso conocido como ciclo lítico.

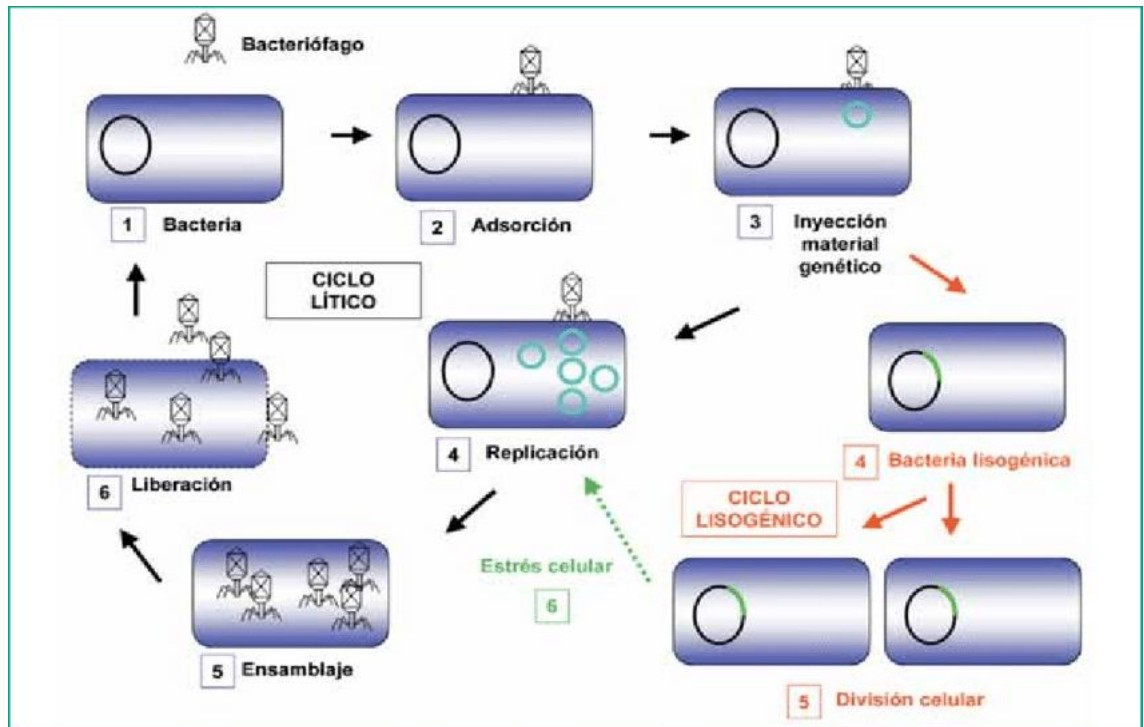


Figura 1: Representación esquemática de los ciclos de vida de los bacteriófagos. Fuente: https://www.researchgate.net/publication/319351357_La_comercializacion_de_organismos_modificados_geneticamente_aspectos_normativos_e_institucionales

4.2.1. Ciclo lisogénico

Tras la adsorción y la penetración del DNA, los fagos temperados son capaces de establecer una relación con el hospedador. El DNA viral se integra como un profago en el genoma bacteriano, pudiendo replicarse junto a este durante numerosas generaciones. Las células lisogenizadas presentan inmunidad frente a la infección, es decir, no pueden volver a ser infectadas por el mismo virus. Además de la inmunidad y de algunos casos concretos de conversión lisogénica, en general exhiben un fenotipo normal (Cluzel, Serio, and Accolas 1987).

En determinadas condiciones, el profago puede comenzar un ciclo lítico, que conduce a la destrucción de la celular hospedadora, liberándose nuevos viriones. Este proceso conocido como inducción puede producirse espontáneamente o provocarse con luz UV o con mutágenos como la mitomicina C. Los genes implicados en el establecimiento y mantenimiento de la lisogenia esta agrupados en el módulo de la lisogenia. Estos módulos tienen una organización genética muy similar entre todos los

fagos de la familia *Siphoviridae* que infectan bacterias gram-positivas de bajo contenido en GC (Lucchini, Desiere, and Brüssow 1999).

-Interruptor genético. Mecanismo que determina que tras la infección de un fago temperado se desencadene el ciclo lítico o ciclo lisogénico. En la mayoría de los casos se basa en dos proteínas codificadas por el genoma viral y que actúan como reguladores de la transcripción. Estas proteínas son el represor, CI y Cro (Ptashne 1986). CI reprime la expresión de los genes líticos provocando el establecimiento del ciclo lisogénico. Además, es responsable de la inmunidad, ya que también inhibe la síntesis de RNA de otros viriones que intenten infectar la misma célula bacteriana. Por el contrario, Cro inhibe la expresión de los genes lisogénicos tempranos, incluido CI y activa la expresión de los genes líticos. Debido a que el represor es capaz de reprimir la transcripción de Cro, existe una competencia entre ambas proteínas, cuyo resultado determinara cuál de los dos ciclos, líticos o lisogénicos, se lleva a cabo.

-Escisión. La integración da lugar a la formación de dos sitios de unión híbridos, en las intersecciones de los genomas fágico y bacteriano. Estos sitios de unión, llamados *attL* y *attR*, comparten el mismo core que *attP* y *attB* y van a permitir que la integración sea un proceso reversible, de tal forma que cuando se produce la inducción, tiene lugar una recombinación sitio-específica entre ellos (Figura 2), escindiéndose el profago del genoma bacteriano e iniciando así el ciclo lítico. Un gen adyacente a *int* (integrasa), el gen *xis*, codifica una escisionasa que unida a la integrasa, es la que cataliza la escisión del profago (Echols, Lu, and Burgers 1983).

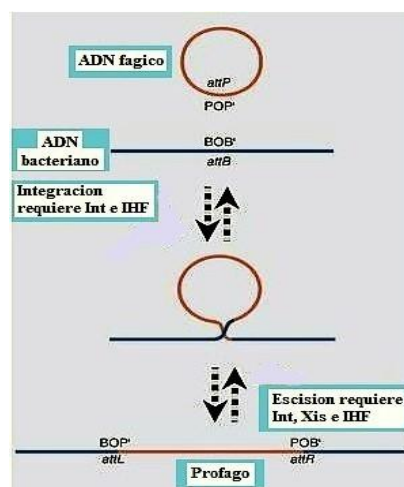


Figura 2. Recombinación sitio específica fago-hospedador (Fuente: <http://fbio.uh.cu/sites/genmol/conf5/conf5/p03.htm>).

Recombinación sitio específica fago-hospedador

La recombinación sitio-especifica consiste en la unión y ruptura de DNA, que es llevado a cabo por enzimas específicas de fagos, llamadas recombinasas sitio-especifica o integrasas. Las recombinasas se dividen en dos familias, dependiendo del residuo aminoácido que interviene en la unión covalente entre el DNA y la proteína; serin- o tirosin-recombinasas. La tirosin-recombinasa une y rompe hebras simples de a pares para formar un intermediario de unión Holiday. Por el contrario la serin-recombinasa corta todas las hebras promoviendo la religación e intercambio de hebras.

Además, la integrasa posee una secuencia de nucleótidos bien conservada entre los grupos de fagos, esto refleja la diversidad de la población de fagos de *S. aureus*. También permite la predicción de la localización cromosomal del profago y por último, está relacionado estrechamente a los genes de virulencia del profago y podría por lo tanto transportar información acerca del potencial patogénico de *S. aureus* (Goerke et al. 2009).

4.2.2. Ciclo lítico

Adsorción. El primer paso de la infección consiste en la adsorción del fago a los receptores presentes en la superficie celular de la bacteria. Este acoplamiento es muy específico, de forma que solo tiene lugar cuando la cola del fago y los receptores tienen configuraciones complementarias.

Inyección del DNA fágico. La inyección es el proceso mediante el cual, el DNA pasa desde la cabeza del fago hasta el interior del citoplasma bacteriano a través de la cola. La partícula fágica vacía, permanecerá en el exterior. La inyección del DNA parece ser dependiente de la producción celular de energía y de la síntesis de proteínas (Watanabe and Takesue 1973; Watanabe, Takesue, and Ishibashi 1979; Watanabe et al. 1991). Poco después de la inyección del DNA viral se detiene la síntesis de ácidos nucleicos y proteínas del hospedador y comienza a degradarse su genoma. La maquinaria metabólica del hospedador se pone al servicio del virus, sintetizando sus ácidos nucleicos y proteínas.

Transcripción. La transcripción de los fagos se clasifica en función del tiempo que transcurre entre la infección y el momento en que se detecta el ARN correspondiente. Según el tipo de bacteriófago, existe transcripción temprana,

intermedia y tardía. Se comprobó que la transcripción temprana es llevada a cabo por la ARN polimerasa del hospedador (**Chandry, Davidson, and Hillier 1994; Parreira et al. 1996**). El inicio de la transcripción intermedia y tardía requiere de la síntesis previa de una proteína activadora codificada por el genoma viral.

Replicación. La replicación del ADN se produce posteriormente a la transcripción temprana y se considera el evento que marca la transición de las funciones virales tempranas a las tardías. Los mARN tardíos se sintetizan tras la replicación, realizándose la transcripción a partir de las copias del genoma sintetizadas *de novo*. Las regiones de replicación caracterizadas parecen estar muy conservadas (**Desiere, Lucchini, and Brüssow 1999**) y se identificaron distintos genes que codifican proteínas implicadas en la replicación como helicasas, primasas, proteínas de unión al DNA simple hebra, topoisomerasas, metilasas o proteínas organizadoras del replisoma. Los orígenes de replicación se caracterizan por ser regiones no codificadoras, con alto contenido en pares de bases AT y que presentan numerosas repeticiones directas e invertidas (**Lu et al. 2010**).

Biosíntesis de las proteínas estructurales. El mARN tardío dirige la síntesis de tres tipos diferentes de proteína: las proteínas estructurales (cabeza y cola); las proteínas que participan en el ensamblaje del virión (proteínas de andamio, participan en el ensamblaje de las cápsides y son eliminadas una vez concluida su construcción) y las proteínas responsables de la lisis celular y la liberación de los fagos. Los genes que codifican las proteínas estructurales están agrupados en el genoma viral constituyendo los denominados módulos morfogenéticos. Esta organización es parecida entre fagos que infectan distintas bacterias, pero apenas existe similitud entre la secuencia de aminoácidos de los productos de distintos genes (**Gutiérrez et al. 2014**).

Empaquetamiento del DNA esta mediado por enzimas denominadas terminasas. Estas proteínas están compuestas, por una subunidad mayor y una menor, son endonucleasas que favorecen la interacción específica entre el ADN y la procabeza viral y dirigen el empaquetamiento del ADN en su interior, en un proceso dependiente de ATP (**Black 1989**). Existen dos mecanismos distintos de empaquetamiento de los concatémeros de DNA en el interior de las cabezas:

-Sitios *pac*. La terminasa inicia el empaquetamiento uniéndose al DNA y cortando el concatémero por una secuencia de reconocimiento específica conocida

como sitio *pac*. El DNA penetra en la procápside hasta que está totalmente llena y es precisamente la capacidad de la cabeza, la que determina el sitio en el cual la terminasa realiza el segundo corte. La capacidad de las cabezas permite empaquetar un DNA un poco más largo que el genoma completo del fago, lo que provoca la existencia de secuencias repetidas en los extremos (redundancia terminal) y que diferentes viriones presenten secuencias distintas en los extremos (permutación circular) (**Black 1989**). La terminasa posee dominio terminasa_3.

-Sitios *cos*. El mecanismo de empaquetamiento es sitio específico. La terminasa reconoce unas secuencias determinadas, denominadas *cos*, que flanquean el genoma completo del fago y determinan los sitios de inicio y terminación del empaquetamiento. Todo DNA comprendido entre ambos extremos cohesivos y que corresponde a un genoma completo, habrá sido empaquetado. La terminasa posee dominio terminasa_1.

Lisis celular y liberación de las partículas fágicas. Ocurre la lisis de las células hospedadoras y la liberación de la progenie viral. Generalmente, la lisis celular requiere la acción de dos enzimas codificadas por el genoma fágico: la lisina, que actúa degradando el peptidoglicano de la pared celular y la holina que va a provocar la desorganización de la membrana celular, generando poros que permiten el acceso de la lisina a la pared celular (**Sheehan et al. 1996**).

El tiempo mínimo necesario que transcurre entre la infección y la liberación de los nuevos viriones, conocido como periodo de latencia, depende del virus y del hospedador, variando en los fagos entre los 10 y 140 minutos. El tamaño de explosión es también heterogéneo, pudiendo variar entre los 10 y 400 viriones liberados por célula infectada. (**Klaenhammer and Fitzgerald 1994**).

4.3. Características generales de los bacteriófagos.

Las partículas infecciosas, consisten en un fragmento de material genético rodeado por una envoltura proteica que lo protege del ambiente y le sirve como vehículo de transmisión de una célula hospedadora a otra (**Figura 5**). Los viriones son mucho más pequeños que las bacterias a las que infectan, siendo imposible su visualización en el microscopio óptico. Los análisis morfológicos se realizan mediante la observación en el microscopio electrónico de preparaciones teñidas negativamente.

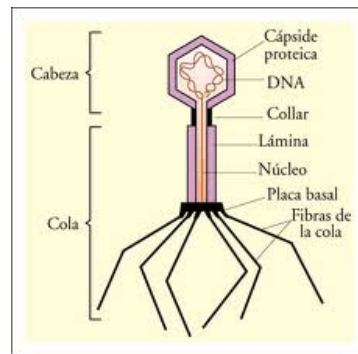


Figura 5: Morfología del virus (Fuente: <https://epidemiologiamolecular.com/bacteriofagos/>)

Tradicionalmente los bacteriófagos se clasificaron según el modelo de Bradley (**Bradley 1967**) que los divide según la cola en tres morfotipos: A) con cola contráctil; B) con cola no contráctil y larga y C) con cola no contráctil y corta. A su vez, se pueden clasificar basándose en la forma de la cabeza: 1) cabezas isométricas; 2) cabezas alargadas y pequeñas y 3) cabezas alargadas y grandes. En la taxonomía moderna, los tres morfotipos principales, A, B y C, se conocen como familias *Myoviridae*, *Siphoviridae* y *Podoviridae*, respectivamente (**Figura 6**).

A pesar de que la clasificación morfológica de los bacteriófagos es la más extendida, su significado evolutivo no está claro. Así por ejemplo, los mapas genéticos del podovirus P22 y el siphovirus λ comparten una estructura tan semejante que su separación taxonómica es cuestionable (**Brüssow and Hendrix 2002**).

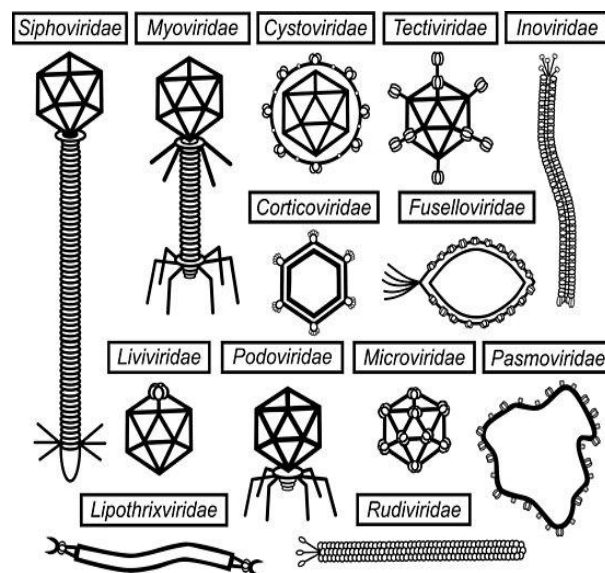


Figura 6: Familias de fagos (Fuente: <https://www.redbubble.com/es/people/bacteriophage/works/23845913-familias-de-bacteri-fagos?p=poster> imagen Luiz Albino, Brasil).

-Composición de los viriones. Los fagos contienen material genético ADN lineal doble o simple hebra o DNA circular o ARN simple o doble hebra. El ADN de los fagos se tiene que empaquetar en la cabeza de los viriones y por lo tanto, el tamaño de ésta, limita su longitud. El tamaño del genoma es específico para cada fago y varía normalmente entre 18 y 134 Kb (**Hatfull 2008**). En los más pequeños, la estructura es tan compacta que son habituales los solapamientos entre genes (**Moineau 1999**).

-Contenido en GC. Este parámetro es específico para cada fago y normalmente refleja el contenido en GC de la célula hospedadora evidenciando la coevolución parasito-hospedador (**Klaenhammer and Fitzgerald 1994**).

-Organización modular. Los genomas de los bacteriófagos están organizados en unidades funcionales multigénicas conocidas como módulos. Estos módulos codifican funciones biológicas concretas, como puede ser la construcción de la cápside o el establecimiento de la lisogenia. La comparación de las cada vez más numerosas secuencias de genomas completas disponibles, está permitiendo confirmar la teoría modular de la evolución de los fagos (**Desiere, Lucchini, and Èssow 1998**) según la cual, el producto y la unidad de la evolución, es una familia de elementos genéticos intercambiables o módulos.

-Proteínas estructurales. Estas proteínas (cabeza y cola) suelen ser inmunogénicas (**Moineau 1999**).

4.3.1. Bacteriófagos de *Staphylococcus aureus*

Debido a su rol crucial en virulencia y patogénesis, los fagos de *Staphylococcus aureus* han sido estudiados ampliamente. La mayoría de ellos, codifican y diseminan factores de virulencia. Además, sus mecanismos contribuyen a la adaptabilidad y versatilidad extraordinaria de estos patógenos prominentes debido a un incremento en la plasticidad genómica.

A nivel morfológico, todos estos fagos caracterizados pertenecen al orden *Caudovirales* y son principalmente *Siphoviridae* (familia) temperados. A nivel molecular, las genómicas comparativas revelaron un mosaicismo extenso, con los genes organizados en módulos funcionales que se intercambian frecuentemente entre fagos

(**Kahánková et al. 2010**). Todos estos aspectos, son de crucial importancia para nuestro entendimiento de la emergencia y evolución de los patógenos entre especies bacterianas tales como *Staphylococcus*.

4.3.2. Características globales de los fagos de *S. aureus*

Como la gran mayoría de los fagos, los de *S. aureus* poseen DNA doble hebra y en su mayoría son temperados. Los cuales están insertos en el cromosoma bacteriano como profagos, algunos de ellos son líticos debido a mutaciones en las funciones lisogénicas. De acuerdo a la clasificación morfológica, propuesto previamente por Ackermann, los *Siphoviridae* están compuestos por una cápside icosaédrica y una cola no contráctil que termina en una estructura, placa basal. Además, según la categoría del tamaño del genoma pertenecen a la clase II (tamaño de genoma 40 Kb, intermedio). El tamaño se correlaciona con la clasificación morfológica (**Deghorain and Van Melderren 2012**).

Las regiones codificantes están estrechamente juntas, con muy pocas y pequeñas regiones intergénicas y una alta densidad génica (1,67 genes/Kb en promedio). Generalmente, su contenido GC es similar al de su hospedador (**T. Kwan et al. 2005**).

Los genomas de *Siphoviridae de S. aureus* muestran la estructura típica de la familia morfológica. Cinco módulos funcionales que se disponen como sigue; lisogenia, metabolismo del ADN, empaquetamiento del ADN y morfogénesis de la cápside, morfogénesis de la cola y la lisis celular del huésped. El módulo del metabolismo del ADN puede ser dividido en funciones de regulación y replicación. Cuando presentan factores de virulencia, generalmente están codificados corriente abajo del módulo de lisis (**Brussow, Canchaya, and Hardt 2004; Canchaya et al. 2003**).

En algunos casos, ellos están insertados entre los módulos de metabolismo del ADN y la lisogenia (**Bae et al. 2006**). Los genes son generalmente transcriptos en la misma hebra, excepto para los clústeres pequeños, tales como los genes involucrados en la integración del genoma del huésped (**T. Kwan et al. 2005; Canchaya et al. 2003**). En ocasiones los módulos no están bien definidos y los genes de la lisis y la cola están solapados (**Deghorain and Van Melderren 2012**).

La morfogénesis de la cápside y el empaquetamiento del ADN muestra un típico patrón de genes de los géneros de fagos como Sfi11 y Sfi21 (**Figura 7**) (**Canchaya et al. 2003; Brüssow and Desiere 2001**).

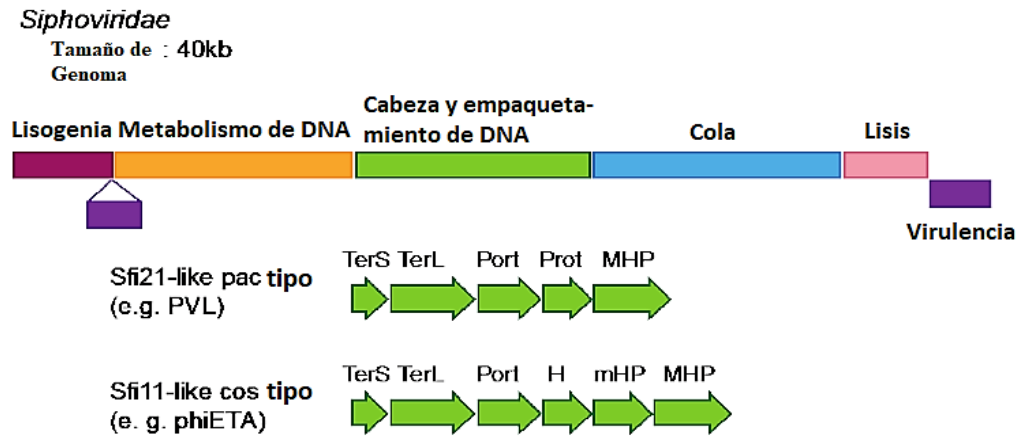


Figura 7: Módulos funcionales en genomas *Siphoviridae* (**Deghorain and Van Melderren 2012**). **TerS:** subunidad pequeña de terminasa; **TerL:** subunidad grande de terminasa; **Port:** proteína portal; **Prot:** proteasa; **MHP:** proteína de la cápside mayor; **mHP:** proteína de la cápside menor; **H:** proteína de morfogénesis de la cápside.

4.3.3. Clasificación y evolución de la familia de fagos de *Staphylococcus*.

Las clasificaciones previamente propuestas para fagos de *S. aureus* se basaron en sus propiedades líticas, serotipos, morfología, en tamaño y número de proteínas de viriones y en organización y tamaños de genomas, las cuales se determinaron por hibridación del DNA o patrones de restricción de la endonucleasa (**Łobocka et al. 2012**). Más recientemente, los progresos en genómica y bioinformática han permitido métodos de clasificación alternativos.

Estudios de genómicas comparativa llevo a la subdivisión de las familias morfológicas en subfamilias y género. Siguiendo una clasificación propuesta por Brussow y Desiere (aunque aun no reconocida por ICTV, Comité internacional en Taxonomía de virus); los fagos de *Siphoviridae* de bacteria gram positiva de bajo contenido GC, incluyendo *Staphylococcus*, son categorizado como fagos tipo Sfi11 o Sfi21, por varios autores (**Brüssow 2001**). Esta distinción está basada, en los patrones de genes de la cápside. Los fagos tipo Sfi21 comparten características en la región de la cápside con *E. coli* HK97 y usa una estrategia basada en sitio *cos* similar, para el empaquetamiento del DNA. En *S. aureus*, el género tipo Sfi21 se subdividió en tres grupos: el primero está representado por los fagos phiPVL, phiPV83, phi13 y

phiSa3mw; el segundo por phiSLT, phiSa2mw y phi12 y el tercero por phiMu50A y phiN315 (**Canchaya et al. 2003**). Estos muestran una morfología de cápside distinta (isométrica o elongada) y codifican patrones de genes típicos de los fagos Sfi21 (proteína mayor de cápside, proteasa y proteína portal). Los fagos tipo *pac*, como Sfi11, están relacionados al fago SPP1 de *B. subtilis* y difiere notablemente de los fagos tipo Sfi21, por la ausencia del gen que codifica la proteasa (**Brüssow 2001; Le Marrec et al. 1997**).

Otro enfoque de clasificación se basó en genes marcadores específicos dentro de los genomas de *Siphoviridae* de *S. aureus*, propuestos. Esta consiste, en la detección por PCR de genes representativos, de diferentes tipos o categorías de fagos. Esto es útil para la detección de profagos de *S. aureus*, el cual es de gran interés para estudios epidemiológicos (**Lindsay and Holden 2004; Kahánková et al. 2010**). Los marcadores abarcan genes codificantes para componentes estructurales; tales como fibras de cola, proteínas de cápside (**Pantůček et al. 2004**) o genes de integrasas (**Goerke et al. 2009**). En el último caso, la clasificación se correlaciona con sitios de integración distintos en el cromosoma huésped. Una buena correlación entre el tipo de integrasa y determinantes de virulencia, han sido reportados (**Goerke et al. 2009**).

Sin embargo, estos métodos no proveen información acerca de la estructura de mosaico, aunque la detección de genes representativos de cada módulo funcional podría proveer algún indicio de mosaicismo (**Kahánková et al. 2010**). La organización génica en mosaico es consistente con la teoría de evolución modular basado en intercambios de módulos, por transferencia horizontal y eventos de recombinación (**Hatfull and Hendrix 2011**). Los fagos de *S. aureus* son altamente mosaicistas (**Nelson et al. 2006; Morgan and Pitts 2008**) indicando que los intercambios de genes son comunes, dentro de este grupo de fagos. Intercambios no solamente de genes individuales, sino también de dominios de proteínas o grupos de genes, tales como los módulos funcionales (**Hatfull and Hendrix 2011**). Los eventos de intercambio están favorecidos entre fagos del mismo rango de tamaño genómico, el cual corresponde a distintas familias morfológicas (**Belcaid, Bergeron, and Poisson 2010**).

Un aspecto interesante del mosaicismo, es que la comparación de un gran número de genomas, permite el establecimiento de relaciones filogenéticas de regiones específicas en lugar de los fagos enteros. Árboles filogenéticos de genes representativos

revelan distintas historias evolutivas para distintos módulos (**Deghorain and Van Melderren 2012**). El mosaicismo abarca genomas de fagos de *Siphoviridae* de *S. aureus* y especies de CoNS, sugiriendo intercambios de genes entre especies, el cual podría ser de crucial importancia para la patogénesis de *Staphylococcus* (**Deghorain and Van Melderren 2012**). La nomenclatura de Siphofagos de *S. aureus* es inconsistente. La secuenciación de genomas bacterianos y fágicos y estudios microarreglos de cepas *S. aureus* revelaron una alta diversidad de profagos que confunden la clasificación. Actualmente, se sugirió la clasificación de los bacteriófagos en familias de acuerdo a estudios de homología de los genes que codifican para la integrasa (**Lindsay and Holden 2004**).

4.3.4. Uso de fagos de *Staphylococcus* para terapia fágica y otras aplicaciones biotecnológicas.

Debido al interés renovado en terapia fágica, un gran número de fagos han sido aislados y caracterizados, para uso potencial contra infecciones de *Staphylococcus*, tanto en humanos como en animales y preservación de comidas. La mayoría de los fagos seleccionados para terapia fágica o preservación de comida son estrictamente líticos (**Abatángelo et al. 2017**) debido a las características no deseadas de los fagos temperados (resistencia de cepas lisógenas a infección de fagos y transducción de genes del hospedador) (**Lu and Koeris 2011; Mann 2008**).

El uso directo de lisinas es una alternativa frecuente para el uso de fagos enteros (**O'Flaherty et al. 2005**). Otra característica importante de los fagos seleccionados para aplicaciones biotecnológicas y terapéuticas que posean un amplio rango de hospedador, por ejemplo el fago K es específico para aislados clínicos de *S. aureus* particulares, tales como cepas CoNS particulares, mientras phiStau2 muestra un rango de hospedador específico para aislados clínicos de *S. aureus*, pero es ineficiente contra las cepas de CoNS probadas (**Hsieh et al. 2011**).

Los cocteles de fagos permiten un tratamiento específico contra patógenos sin afectar la flora comensal (**Merabishvili et al. 2009**). La gran cantidad de fagos existentes en el medioambiente posibilita el desarrollo de la terapia fágica (se estima más de 10^{31} partículas fágicas en la biosfera), abriendo las perspectivas para novedosos enfoques terapéuticos (**Lu and Koeris 2011**).

Una alternativa innovadora para la utilización de fagos, es su utilización como biosensor, ya que son altamente específicos para su hospedador. Esto puede llevarse a cabo mediante la modificación genética de los mismos, a modo de obtener un sistema de detección basado en la interacción específica del fago con su hospedador. Ejemplo: biosensor basado en un fago luminiscente independiente de sustrato para detectar específicamente bacteria entérica como *E. coli* (**Franche, Vinay, and Ansaldi 2017; Wan et al. 2007; Li et al. 2010; Zourob and Ripp 2010; Tolba et al. 2010**).

Para todas las aplicaciones que involucren bacteriófagos es importante conocer su secuencia genómica completa. Teniendo en cuenta todo lo mencionado anteriormente, esta tesina se focaliza en la caracterización funcional y genética completa de un bacteriófago temperado de *Staphylococcus aureus* de la familia *Siphoviridae* nombrado como I74. Este fago fue aislado y secuenciado en el Laboratorio de Microbiología Molecular, Facultad de Ciencias Médicas, UNR. El aislamiento y secuenciación de cada vez más fagos de *S. aureus* demanda el desarrollo de nuevos métodos de clasificación y comparación, lo que nos motivó a la utilización de diferentes herramientas bioinformáticas, muchas de ellas desarrolladas para fagos de otros hospedadores. Este fago se aisló mediante la inducción con mitomicina C en la cepa clínica de *S. aureus* H7 (hospitalaria). Sus placas de lisis son heterogéneas en tamaño y tiene una morfología definida. Por microscopía electrónica de transmisión (MET, **Figura 8**) se determinó que pertenece a la familia *Siphoviridae*, presentando colas largas y flexibles no contráctiles (longitud: 160nm) y cabeza icosaédrica (diámetro: 44nm). Además, su secuencia genómica fue determinada observándose que cuenta con un genoma de 43685 pb. Posee un rango de hospedador bajo (39%). Cuenta con una característica llamativa; se postuló que dicho fago no se integra al genoma del hospedador, produciendo placas de lisis claras (**Tesis Peressutti, 2018**).



Figura 8. Micrografía electrónica de transmisión del bacteriófago I74

(Tesis Peressutti, 2018)

5. Objetivos

Como objetivo general de esta tesina, nos propusimos llevar a cabo en la caracterización del genoma del bacteriófago I74. El mismo es miembro de la familia Siphoviridae y específico de *Staphylococcus aureus*. Además, analizar su relación filogenética con otros bacteriófagos del mismo hospedador y familia, mediante análisis bioinformáticos.

Los objetivos específicos a desarrollar en este proyecto son:

- Realizar la predicción de los marcos abiertos de lectura del bacteriófago I74.
- Realizar la predicción de funciones utilizando diversas herramientas bioinformáticas.
- Analizar las relaciones filogenéticas aplicando diferentes programas a partir de genomas completos y/o genes individuales.
- Comparar los resultados obtenidos a través de nuestros estudios de clustering de fagos de *S. aureus* con otros métodos de clasificación sugeridos en la bibliografía.

6. Materiales y métodos

6.1. Bacteriófagos utilizados

Durante el desarrollo de esta tesina utilizaremos la secuencia genómica del bacteriófago I74 de *Staphylococcus aureus* (SA) inducido de la cepa H7 de origen clínico. La búsqueda de fagos de la familia Siphoviridae con sus secuencias completas y depositadas en la base de datos de NCBI arrojó los resultados que se listan en la Tabla I (21 de junio del 2018).

Tabla I. Bacteriófagos de la base de datos NCBI con su respectivo código de acceso NCBI.

BACTERIOFAGOS	CODIGO NCBI	BACTERIOFAGOS	CODIGO NCBI
3A	NC_007053.1	92	NC_007064.1
3MRA	NC_028917.1	96	NC_007057.1
phiSa2wa_st121mssa	NC_004615.1	187	NC_007047.1
21	MG029518.1	B166	NC_028859.1
29	NC_007061.1	B236	NC_028915.1
23MRA	NC_028775.1	DW2	NC_024391.1
37	NC_007055.1	EW	NC_007056.1
42e	NC_007052.1	vB_SauS_IMEP5	KX156762
47	NC_007054.1	JS01	NC_021773.2
52A	NC_007062.1	LH1	JX174275
53	NC_007049.1	P240	KY056620
55	NC_007060.1	P282	KT809368
55-2	KR709302.1	P630	KT809369
55-3	KR709303.1	P954	NC_013195.1
69	NC_007048.1	P1105	KT878766
71	NC_007059.1	phi2958PVL	NC_011344.1
77	NC_005356.1	vB_SauS_phi2	NC_028862.1
80	NC_030652.1	phiETA2	NC_008798.1
85	NC_007050.1	phiETA3	NC_008799.1
88	NC_007063.1	phiJB	NC_028669.1
phiMR11	AB370268.1	UPMK_2	MG564297
phiMR25	NC_010808.1	X2	NC_007065.1
phiPV83	AB044554.1	SH-St 15644	MG770897.1
phiPVL108	NC_008689.1	IME1346_01	KY653125.1

phiPVL-CN125	NC_012784.1	IME1354_01	KY653126.1
phiSaus-IPLA35	NC_011612.1	IME1361_01	KY653123.1
phiSaus-IPLA88	NC_011614.1	SA97	NC_029010.1
ROSA	NC_007058.1	StauST398-1	NC_021326.1
VB-SauS-SA2	MH356730.1	StauST398-4	NC_023499.1
SA7	KY695153	StauST398-5	NC_023500.1
SA12	NC_021801.1	StauST398-2	NC_021323.1
SA13	NC_021863.1	StauST398-3	NC_021332.1
SAP-26	NC_014460.1	YMC/09/04/R1988	NC_022758.1
SP5	JX274646.1	phiBU01	NC_026016.1
SP6	JX274647.1	SA137ruMSSAST121PVL	MH384261.1
phiSa2wa_st1	MF580410.1	SA1014ruMSSAST7	MH384259.1
phiSa2wa_st5	MG029509	2638A	NC_007051.1
phiSa2wa_st22	MG029510	SMSAP5	NC_019513.1
TEM123	NC_017968.1	11	NC_004615.1
TEM126	HQ127381		

6.2. Anotación genómica del bacteriófago I74 de SA

Para los análisis bioinformáticos se utilizó una computadora con sistema operativo Ubuntu Linux 14.04 LTS donde se instalaron todos los programas mencionados en la siguiente tesina.

Para llevar a cabo la anotación genómica se utilizaron las herramientas bioinformáticas; DNA Master, Rast y Genemark (<https://phagesdb.org/DNAMaster/>; **Besemer and Borodovsky 2005**; <http://exon.gatech.edu/GeneMark/>) para determinar los ORFs hipotéticos (marcos abiertos de lectura) de dicho fago. Luego se realizó una revisión manual de los mismos y se fue modificando aquellos ORFs que no tenían un codón de inicio y terminación definido o se solapaban con otros ORFs. La elección del codón de inicio se realizó teniendo en cuenta el Z-value (menor) priorizando una mayor longitud del ORF. La función de las proteínas se predijo mediante el programa BLASTp. Una vez finalizada la anotación genómica, se extrajo el mapa genómico del fago I74 del programa DNA Master.

Mediante la herramienta bioinformática, Artemis 16.0.11 (Carver et al. 2008), se extrajo información de la secuencia génica del fago I74. Los datos extraídos fueron; densidad génica, longitud en promedio, % codificante y % GC.

6.3. Genómica comparativa.

La asignación de las familias o “phams” de los productos génicos de los fagos se realizó mediante el programa Phamerator (Cresawn et al. 2011). Para ello primero se generó una base de datos MySQL, luego se cargó en ella los genomas de los fagos en formato genbank. Los fagos que no se encontraban ordenados, con el gen de la integrasa en la primera posición, fueron reordenados mediante DNA Master. Se realizó la comparación de los genomas utilizando ClustalW y BLAST a través de scripts incorporados en Phamerator. Se realizó la asignación de familias, utilizando los criterios: E-value menores de 10^{-50} (BLASTn) y una identidad de secuencia mayor a 32,5% (ClustalW). Aquellos ORFs que cumplen con dichos valores se consideran una familia específica. Este programa permite exportar mapas de fagos donde se muestran las áreas que comparten homologías y representaciones circulares de los miembros de determinadas familias y como se relacionan.

Por último, se extrajo el contenido de GC de los fagos de la base de datos NCBI de Phamerator. Con los datos extraídos, se realizó el boxplot y los análisis estadísticos descriptivos para comparar el %GC de los clusters, utilizando la herramienta estadística R 3.5.1. Entre los análisis estadísticos se aplicó el test Fligner-Killeen para determinar si existe homogeneidad de las variancias entre los clústeres, el test Shapiro Wilk para determinar si hay distribución normal en los clústeres, el test Kruskal-Wallis para determinar si hay diferencias de %GC entre los clústeres y el test Tukey para las comparaciones múltiples entre los clústeres y determinar si existe diferencias significativas en %GC entre los mismos.

Se utilizó el paquete CG view Comparison Tool (CCT) para generar gráficos circulares, que representan la comparación de los genomas y proteínas de los fagos utilizando BLASTn y BLASTp, respectivamente (Grant and Stothard 2008). Para construir los gráficos se asignó como genoma de referencia a I74 y el resto de los genomas se compararon con éste. Para ello se utilizó el script “*build_blast_atlas.sh. - p*”.

Para construir la red filogenética se generó una base de datos con los genomas de los fagos en formato genbank, en el programa bioinformático **Gegenees 2.2.1** (Ågren et al. 2012), el cual realiza un alineamiento fragmentado. Luego se exportó dicho alineamiento en un archivo con formato nexus. Por último, con dicho archivo se construyó la red filogenética con el software Splits Tree 4.14.6 (Huson 1998) utilizando el método de algoritmo basado en distancias “Neighbor-Net”.

El análisis de homología general de las secuencias de aminoácidos de las proteínas tapemeasure (TMP) de los fagos, reproducido en forma de gráfico de puntos (dot-plot), se realizó con el programa **GEPARD** (Krumstiek, Arnold, and Rattei 2007). Para ello, se generó un archivo multifasta con las secuencias de las TMP a analizar. Las secuencias fueron ordenadas manualmente hasta dar patrones similares; así se definió claramente los clústeres.

En el software **CAFE** (Lu et al. 2017) (acceso libre en <https://github.com/younglululu/CAFE>), se incorporaron los genomas de los fagos en formato fasta en un orden específico, teniendo en cuenta los parámetros $K=8$ y el algoritmo Manhattan. Así se construyó el heatmap. Luego con los valores de distancias de los pares de genomas fágicos, arrojados en dicho software se realizó un dendograma en R-Studio, utilizando los paquetes Circlize y Dendextend, utilizando UPGMA como método de clustering.

Se utilizó el programa **Mash** (Ondov et al. 2016) que realiza comparaciones libres de alineamiento (<https://github.com/marbl/mash>). El input de este programa son los genomas en formato fasta. Se utilizaron un tamaño de sketch de 50000 y un $k=17$ mediante los comandos:

```
>mash sketch -k 17 -s 50000 -i -o
```

```
>mash dist -t
```

La utilización de estos parámetros se debió a que se ajustaron mejor a los resultados esperados. Los resultados se exportaron en formato tab; en el cual se encuentran las distancias de los pares de genomas fágicos. Luego en R-Studio 1.1.456 se realizó un dendograma, utilizando los paquetes Circlize y Dendextend.

Las secuencias de aminoácidos de las TMPs, de las subunidades grandes de las terminasas (también se analizaron con BLASTp y Phamerator) y las secuencias de nucleótidos de las integrasas se alinearon utilizando ClustalW en el programa **MEGA X** (Kumar et al. 2018) (www.megasoftware.net, acceso libre). Los árboles filogenéticos se construyeron mediante el método basado en distancias “Neighbor-Joining”. El

porcentaje de árboles replicados con el Test Bootstrap (*500 replicates*), la distancia evolutiva se computó con el método de corrección de Poisson y todas las posiciones conteniendo datos ausentes y gaps fueron eliminados (delección completa).

Por otro lado, para determinar el tipo de integrasa en cada bacteriófago, se analizaron las secuencias de aminoácidos con las herramientas BLASTp, Pfam (Finn et al. 2016), Phamerator (Cresawn et al. 2011) y HHPRED (Soding, Biegert, and Lupas 2005). Luego para la construcción del árbol filogenético, se usaron las secuencias de los nucleótidos de las integrasas. La distancia evolutiva se computó con el método de Maximun composite likelihood.

Para realizar el cálculo del parámetro **GCD** (disimilitud en el contenido de genes) (Pope et al. 2017) versus los clústeres, se utilizaron los datos de familias asignadas por Phamerator. Los datos requeridos se extrajeron de las bases de datos de este programa utilizando Libre Office Base. Se utilizó un script en Python; *analyze_pham_data.py*

(https://github.com/tmavrich/mavrich_hatfull_nature_micro_2017), el cual utiliza los datos obtenidos anteriormente. Este script nos proporciona la proporción de familias compartidas (término entre paréntesis en la ecuación). Con estos datos podemos calcular el valor de GCD utilizando la siguiente ecuación:

$$\text{GCD} = 1 - \left(\frac{\frac{\text{Shared phams}}{\text{Total phams in genome A}} + \frac{\text{Shared phams}}{\text{Total phams in genome B}}}{2} \right)$$

Figura 9. Calculo de GCD. (Pope et al. 2017)

GCD es calculado por determinación de la proporción de phams relativos compartidos para cada número total de genomas de phams y después se promedia las proporciones a la mitad (Figura 9). Por último, se generó el boxplot en RStudio utilizando los paquetes gplots y RColorBrewer.

7. Resultados

CAPITULO 1

7.1. Análisis genómico del nuevo bacteriófago de *Staphylococcus aureus* (I74) de la familia Siphoviridae

Como se ha mencionado anteriormente, existen diversas aplicaciones biotecnológicas que involucran el uso de bacteriófagos. Entre ellas podemos mencionar, la terapia fágica, el uso como biosensores y la aplicación de enzimas codificadas por los fagos con actividad antimicrobianas (lisinas). Por otro lado, la capacidad de los fagos temperados de poder integrarse en el genoma del hospedador y llevarse consigo genes de éste cuando comienza el ciclo lítico lo transforma en un excelente vehículo para genes de virulencia como toxinas y resistencia a antimicrobianos. Para el estudio de todas ellas, un requisito esencial es el conocimiento de la secuencia completa de los fagos y las enzimas que ellos codifican. En este trabajo se caracterizó bioinformáticamente el fago I74, miembro de la familia Siphoviridae específico de *S. aureus* el cual fue aislado en el Laboratorio de Microbiología Molecular de la Fac. De Cs. Médicas, UNR. Luego fue secuenciado por el INDEAR (Instituto de Agrobiotecnología de Rosario) y cargado en el programa Rast. Este último, se utilizó para seleccionar el contig apropiado desde el ensamblado crudo. Aquellos contigs que contenían secuencias genéticas provenientes del hospedador y tamaño pequeño se eliminaron. Luego, el contig seleccionado de I74 (contig 1) fue, importado en el programa DNA master.

7.2. Análisis genómico del Siphostaphylofago

Anotación genómica

Para llevar a cabo la anotación del genoma del bacteriófago I74 se utilizó el programa DNAMaster. Este programa junto a otras herramientas fueron desarrollados por el grupo del Dr. Graham Hatfull para el análisis de fagos pertenecientes al filo actinobacterias. Este programa nos permite a través de la utilización de herramientas como Genmark y Glimmer (integrados en DNAMaster) hacer las predicciones de marcos abiertos de lectura (del inglés, Open Reading Frames, ORFs). En simultáneo la secuencia genómica del fago se carga en servidor online de Genmark obteniéndose una gráfica con el potencial codificante para cada gen. Con ambos datos (DNAMaster y Genmark) se procede a la inspección manual de cada ORF predicho y su corrección si fuera necesario.

Una vez identificados los ORFs se realiza la predicción de función (**Tabla II**). La función biológica se predijo por similitud con secuencias conocidas de fagos o bacterias depositadas en la base de datos y los *hits* positivos se identificaron con el BLASTp. También se utilizó la herramienta HHPred, el cual es un servidor para la predicción de estructura y función de proteínas. Además, se ordenó el genoma del fago tomando arbitrariamente como ORF inicial, el gen correspondiente a la integrasa.

Pudimos asignarle función probable al 42,25% de los ORFs en promedio, utilizando BLASTp, mientras que el 57,75% restante no se correlacionó con proteínas de función conocida para bacterias o fagos de la base de datos de GenBank (corresponden a proteínas hipotéticas).

Tabla II. Funciones biológicas predichas de los ORFs del fago I74.

ORF	FUNCION	ORF	FUNCION	ORF	FUNCION
1	integrasa	26	Atpasa familia AAA	51	proteína hipotética
2	escisionasa	27	proteína hipotética	52	proteína hipotética
3	proteína hipotética	28	proteína hipotética	53	proteína de cola
4	exonucleasa	29	proteína hipotética	54	proteína Chaperona
5	regulador transcripcional Familia XRE	30	proteína hipotética	55	proteína hipotética
6	regulador transcripcional Familia XRE	31	proteína hipotética	56	proteína Tapemeasure
7	proteína hipotética	32	proteína hipotética	57	proteína de cola
8	proteína hipotética	33	proteína hipotética	58	peptidasa
9	antirepresor	34	acetiltransferasa	59	proteína hipotética
10	proteína hipotética	35	proteína hipotética	60	proteína hipotética
11	proteína hipotética	36	pirofosfatasa dUTP	61	proteína hipotética
12	proteína hipotética	37	proteína hipotética	62	proteína hipotética
13	oxidoreductasa	38	proteína hipotética	63	proteína hipotética
14	proteína hipotética	39	activador transcripcional RinB	64	proteína con dominio CHAP
15	proteína hipotética	40	proteína hipotética	65	proteína hipotética
16	proteína hipotética	41	terminasa subunidad pequeña	66	Proteína de fibra
17	proteína hipotética	42	terminasa subunidad grande	67	proteína hipotética
18	proteína hipotética	43	proteína portal	68	holina
19	proteína hipotética	44	proteína menor de cápside	69	proteína con dominio CHAP
20	proteína hipotética	45	proteína hipotética	70	proteína hipotética
21	proteína hipotética	46	proteína hipotética	71	proteína hipotética
22	proteína de unión a ADN simple hebra	47	proteína mayor de cápside		
23	proteína hipotética	48	proteína hipotética		
24	proteína contiene dominio AP2	49	proteína adaptador cola- cabeza		
25	hidrolasa α/β	50	proteína adaptador cola- cabeza		

El mapa genómico de I74 fue exportado desde DNA Master y las funciones predichas de los genes fueron indicados, como se observa en la **Figura 10**.

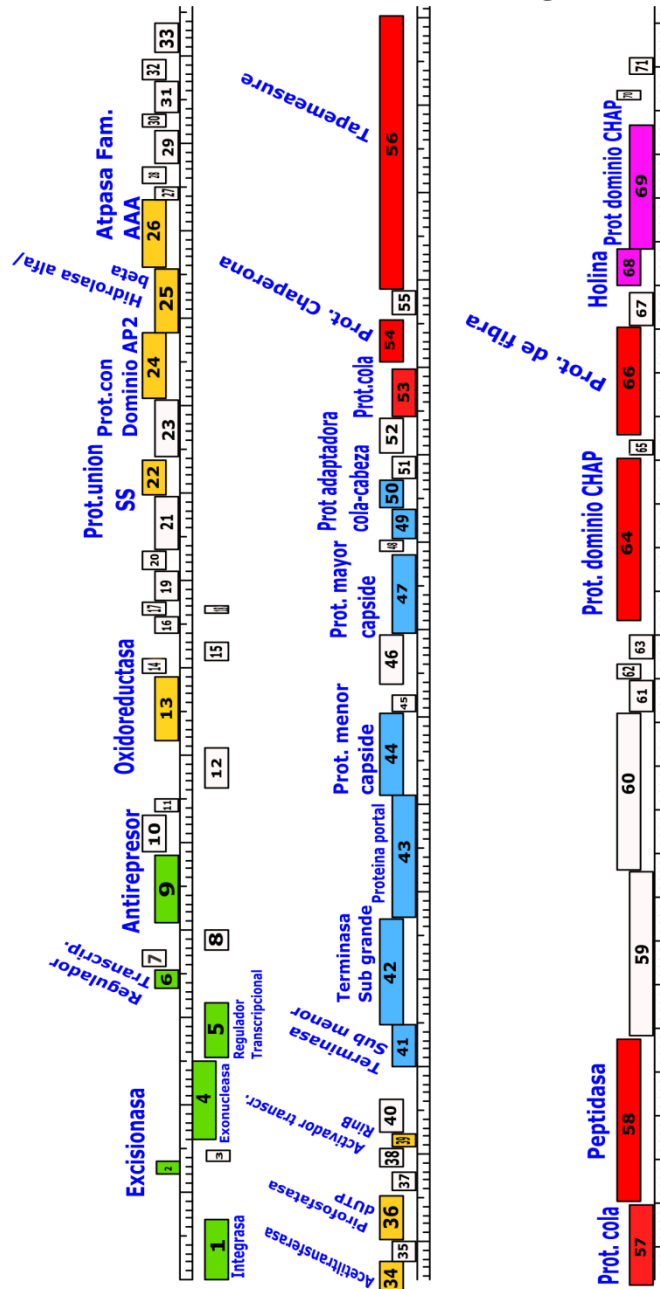


Figura 10. Mapa genómico del bacteriófago I74 extraído del programa DNA Master. Módulo de lisogenia (verde), módulo de metabolismo de DNA (naranja), módulo de morfogénesis de cápside (celeste), módulo de morfogénesis de cola (rojo), módulo de lisis (rosa) y ORFs hipotéticos (blanco).

Al analizar el genoma del fago I74 perteneciente a la familia *Siphoviridae*, se observa que muestra la estructura típica de dicha familia morfológica (Xia et al. 2013). Se disponen en cinco módulos funcionales según un orden característico (Figura 11).

Estos módulos poseen genes conservados y grandes regiones de genes que codifican para funciones desconocidas. En los mismos hay genes codificantes de funciones de morfogénesis de cápside y cola del fago, replicación y transcripción de ADN, lisogenia y lisis celular.

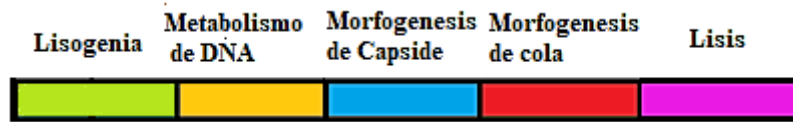


Figura 11. Estructura modular de los fagos de *Staphylococcus* de la familia *Siphoviridae*.

Se observa en el mapa genómico que la mayoría de los genes son transcritos hacia la derecha, salvo los genes que codifican para la integrasa (**ORF 1**), exonucleasa (**ORF 4**), el regulador transcripcional (**ORF 5**) y funciones hipotéticas (**ORF 3, 8, 12, 15 y 18**) que se transcriben hacia la izquierda. Cada uno de los módulos codifica para genes específicos, detallados a continuación:

Lisogenia (ORFs 1, 2, 4, 5, 6 y 9): genes involucrados en la lisogenia y la regulación de la transcripción (integrasa y proteínas regulatorias; CI, Cro respectivamente). El cambio entre el ciclo lítico y lisogénico, es mayormente controlado por un circuito molecular similar al fago λ ; el fago permanecerá en estado lisogénico si predomina CI, pero será transformado al ciclo lítico si predomina Cro. El modo de acción de CI varía entre fagos de distintas familias (**Ganguly et al. 2009**). El **análisis con Pfam** mostró que el ORF5 posee homología con CI, mientras que el ORF6 posee homología de dominio correspondiente con Cro.

El ORF 5 posee un motivo estructural mayor hélice-giro-hélice (HTH3, 55AAs), capaz de unirse al DNA, presentes en reguladores transcripcionales. Clan: CLO123, E-value=1,1x10⁻⁸. Además posee una serin peptidasa (peptidasa S24, 70AAs), este dominio posee un plegamiento complejo de hojas betas plegadas, el cual contiene una estructura barril, tipo SH3. Clan: CLO299, E-value=5,4x10⁻¹⁴.

El ORF6 (55AAs) posee un motivo estructural mayor, hélice-giro-hélice (HTH3), capaz de unirse al DNA. Clan: CLO123. E-value = 1,4x10⁻⁶.

Metabolismo de DNA (ORFs 13, 22, 24, 25, 26, 34, 36 y 39): este módulo está compuesto por genes que codifican para proteínas involucradas en la replicación del DNA (proteína de unión a DNA, proteína de replicación DnaC). Varios genes del

módulo podrían estar involucrados en el cambio del metabolismo del DNA del hospedador para que se dé la replicación del DNA fágico. (**Christie and Dokland 2012; Novick, Christie, and Penadés 2010**).

Se realizó la búsqueda de genes de patogenicidad y resistencia a antibióticos del fago I74, mediante los servidores online VirulenceFinder-2.0 y ResFinder-3.1 respectivamente, el cual no se reportó ningún gen con dichas funciones. También se hizo la búsqueda por inspección manual de los genes de dicho fago sin dar resultados positivos.

Morfogénesis de cápside (ORFs 41, 42, 43, 44, 47, 49 y 50): entre los genes pertenecientes a este módulo se encuentran, los de la subunidad pequeña y grande de la terminasa y genes estructurales de la cápside (proteínas portal y morfogenéticas de cápside). La proteína RinA ejerce una función regulatoria. La actividad de activador de la transcripción tanto de la proteína RinA como RinB fue demostrado, que modula la expresión del gen de la integrasa (**Ye and Lee 1993**). La proteína de la cola y la cabeza mayor son las proteínas más abundantes en los viriones maduros.

Morfogénesis de cola (ORFs 53, 54, 56, 57, 58, 64 y 66): entre éstos genes se hallan los codificantes para proteínas estructurales de la cola (tape measure, proteínas morfogenéticas de la cola, proteína de la fibra de la cola, proteína de la placa basal, proteína de hidrólisis de la pared celular). **Análisis con Pfam** del ORF54 y ORF66. El ORF54 posee homología con una proteína chaperona (111 AAs) que se ensambla a la cola del fago y cubre la proteína tapemeasure para evitar que esta forme un complejo no productivo o un precipitante. Clan: CLO567. E-value= $5,2 \times 10^{-11}$. El ORF66 posee homología con un dominio N-terminal (Bppu N, 148 AAs). E-value: $3,2 \times 10^{-29}$. Además tiene función de cola de fago (PTR, 60 AAs). E-value: $2,3 \times 10^{-7}$.

Lisis (ORFs 68 y 69): está compuesto en *S. aureus* por dos enzimas, que comprende una holina y una endolisina. Los sustratos específicos de varias endolisinas fueron estudiados en detalle, poseen actividad amidasa y endopeptidasa (**Navarre et al. 1999**). **Análisis con Pfam** del ORF69, el cual tiene homología con un dominio CHAP (83 AAs). Clan: CLO125. E-value: $3,6 \times 10^{-11}$; un dominio amidasa N-acetylmuramoyl-L-alanina (126 AAs). E-value: $3,9 \times 10^{-15}$ y un dominio SH3 bacteriano (65 AAs). E-value: $3,9 \times 10^{-24}$.

Las holinas son un grupo funcional de proteínas más diverso y pequeñas. Causan la lisis del hospedador en un tiempo preciso programado genéticamente. Las endolisinas son proteínas con una o más actividades muralítica (capacidad lítica) (**Sass and Bierbaum 2007**).

Las endolisinas presentan dominios específicos (**Figura 12**) que permiten a la enzima anclarse selectivamente sobre una pared bacteriana (CBD, del inglés Cell Binding Domain) concreta, al tiempo que poseen dos dominios catalíticos capaz de hidrolizar el peptidoglicano bacteriano (AMI2 o AMI3 y CHAP). Los ensayos con estas enzimas han demostrado que son efectivos bactericidas con importantes aplicaciones en medicina y biotecnología (**Schmelcher, Donovan, and Loessner 2012**). Actualmente, en el Laboratorio de Microbiología Molecular de la Facultad de Ciencias Médicas (UNR) se están desarrollando estudios específicos en endolisinas.



Figura 12. Análisis con pfam del ORF69 (<https://pfam.xfam.org/search/sequence>).

Artemis, es una herramienta bioinformática de comparación; principalmente es usada para la visualización y anotación de datos de secuencias genómicas, particularmente para genomas microbianos (**Carver et al. 2008**). Se procedió a cargar la secuencia del fago I74 en formato genbank la cual fue generada mediante DNAMaster, y con esta herramienta, se extrajeron los datos genómicos del fago;

Longitud en promedio de los genes: 573 pb (no posee intrones)

Densidad génica: 1,625 genes/Kb (615 bases por gen)

Porcentaje codificante (%): 93,2.

Estos datos muestran una alta densidad génica y un elevado % codificante, muy característico de los bacteriófagos.

CAPITULO 2

7.3. Análisis genómico comparativo del Siphostaphylofago

El incremento reciente en secuencias genómicas de fagos disponibles en la base de datos públicas (828 en el orden de *Caudovirales*, a partir del 20 de junio del 2013; <https://www.ncbi.nlm.nih.gov/>), junto con el desarrollo de las herramientas bioinformáticas, han posibilitado la clasificación de los bacteriófagos en base a sus relaciones genéticas.

Hasta la fecha, un gran número de fagos de *Staphylococcus aureus* han sido descritos, pero una caracterización génica y funcional completa han sido realizado solamente en un número limitado.

Una de las clasificaciones de los *Siphovirus* de *Staphylococcus* los agrupa en tres géneros: “3alikevirus”, “77likevirus” y “Phietalikevirus”. La clasificación se basó en análisis comparativos de secuencias de proteínas y nucleótidos. Los tres grupos tienen dominios conservados, que posee actividad hidrolítica de peptidoglicano, identificado como parte de las proteínas TMPs (“3alikevirus” y “77likevirus”) o como proteínas asociadas a viriones individuales (“Phietalikevirus”). Además los bacteriófagos que corresponden al género “3alikevirus” comparten estrechamente proteínas de empaquetamiento y relacionadas al procesamiento del ADN, mientras que los bacteriófagos del género “Phietalikevirus” codifican proteínas específicas de cola para la interacción con el huésped. Estas propiedades son consideradas distintivas para estos géneros. Mientras que los fagos que no se han podido clasificar en ninguna de estas categorías se los llamó “huérfanos”, los cuales poseen una organización más divergente, pero ellos comparten algunas propiedades con los miembros de estos géneros propuestos (**Gutiérrez et al. 2014**).

Uno de los objetivos de este trabajo de tesina fue realizar una clasificación de fagos temperados mediante la utilización de diferentes herramientas bioinformáticas, actualmente disponibles, y compararlas.

7.3.1. Phamerator

Es un programa desarrollado en lenguaje Python (**Cresawn et al. 2011**), para análisis de genómica comparativa y representaciones comparativas de genomas de bacteriófagos. Se basa en clasificar los genes que codifican proteínas, dentro de familias

o “phams” de secuencias relacionadas, usando ClustalW y BLASTn para realizar las comparaciones. Además, se puede generar, una representación circular de familias de genes facilitando el análisis de sus historias evolutivas, que migran a través de poblaciones de fagos por intercambio génico horizontal (Pope et al. 2017).

Este programa fue desarrollado en un principio para ser utilizado dentro del programa SEAPHAGES, pero en 2016 el grupo de Hope y colaboradores modificó el código de manera de poder generar bases de datos locales (Merrill et al. 2016) .

Para realizar este análisis como primera medida, hay que generar nuestra propia base de datos de fagos. Las secuencias en formato Genbank de 78 fagos (Tabla I, sección Materiales y métodos) fueron obtenidas de la base de datos del NCBI. Posteriormente, los genomas fueron cargados en la base de datos (Mysql) en el programa Phamerator. Una vez que Phamerator realiza la predicción de las familias, las mismas pueden observarse mediante el uso de diferentes colores asignados arbitrariamente en el mapa de los fagos (Figura 13).

Mediante este análisis pudimos observar que la mayoría de los ORFs del fago I74 se le ha asignado una familia y la misma es compartida con otros genes presentes en fagos que componen nuestra base de datos. Ejemplo: ORF 1 pertenece a la familia 176 y la misma esta compuesta por 19 miembros de la base de datos. Un solo ORF no tiene familia asignada, por lo tanto no comparten similaridad con los otros fagos de la base de datos; a estos se los denomina “*Orphans*” (del ingles, huérfanos). En este análisis se identificaron 678 familias en total, de los bacteriofagos de *S. aureus* en estudio.

Además esta herramienta, me permitió predecir la función biológica de algunos ORFs de los bacteriófagos que componen nuestra base de datos (ejemplo: integrasa, terminasa, etc), ya que algunos genes en formato Genbank no poseen una función asignada.

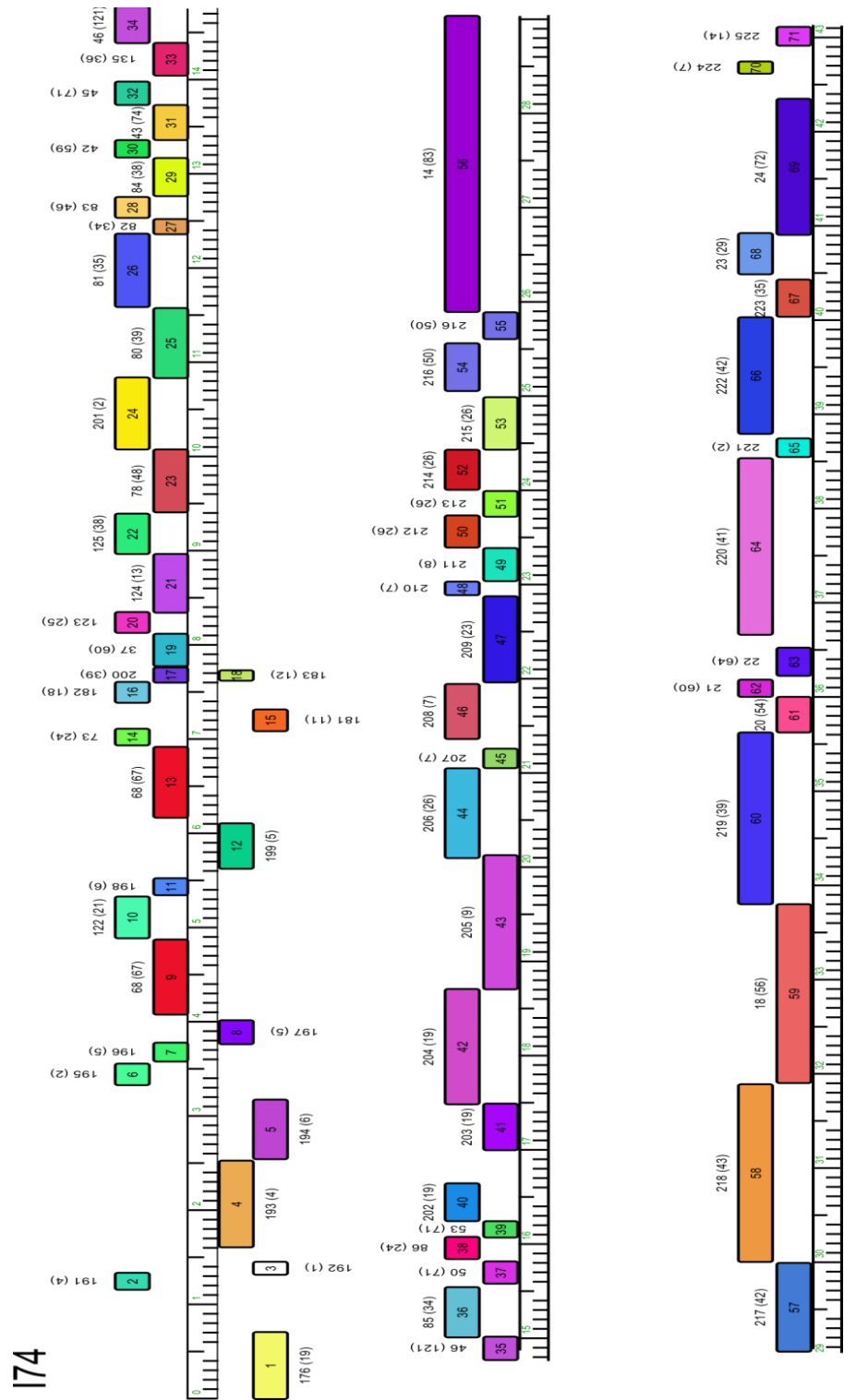


Figura 13. Mapa de los ORFs del genoma del bacteriófago I74 extraído del programa Phamerator. El ORF 3 al no pertenecer a ninguna familia se lo clasifica como orphan (color blanco), mientras que el pham 192 tiene un miembro.

7.3.2. Análisis de la holina de I74

El ORF 68 del fago I74 codifica para una proteína denominada holina; de acuerdo con las familias asignadas por Phamerator, esta enzima pertenece al pham23. Esta familia está compuesta por 29 miembros (SA12, SA13, SA2, SP5, SP6, StauST398-2, StauST398-3, UPMK-2, phi2, 11, 29, 37, 52A, 55, 55-2, 55-3, 69, 80, 85, 88, 92, DW2, EW, I74, IME1354_01, phiMR11, phiMR25, IPLA88 y X2; **Figura 14**).

Se extrajo del Phamerator el mapa circular de las holinas (pham23) de los bacteriófagos del NCBI que tienen alta homología con el gen que codifica para la holina en el fago I74. Aquellos fagos que no se encuentran en dicha familia, se determinó que pertenecen a dos familias diferentes; pham137 con 34 miembros y pham63 con 14 miembros y dos fagos (3MRA y 2638A) que no codifican para una holina en sus genomas (datos no mostrados).

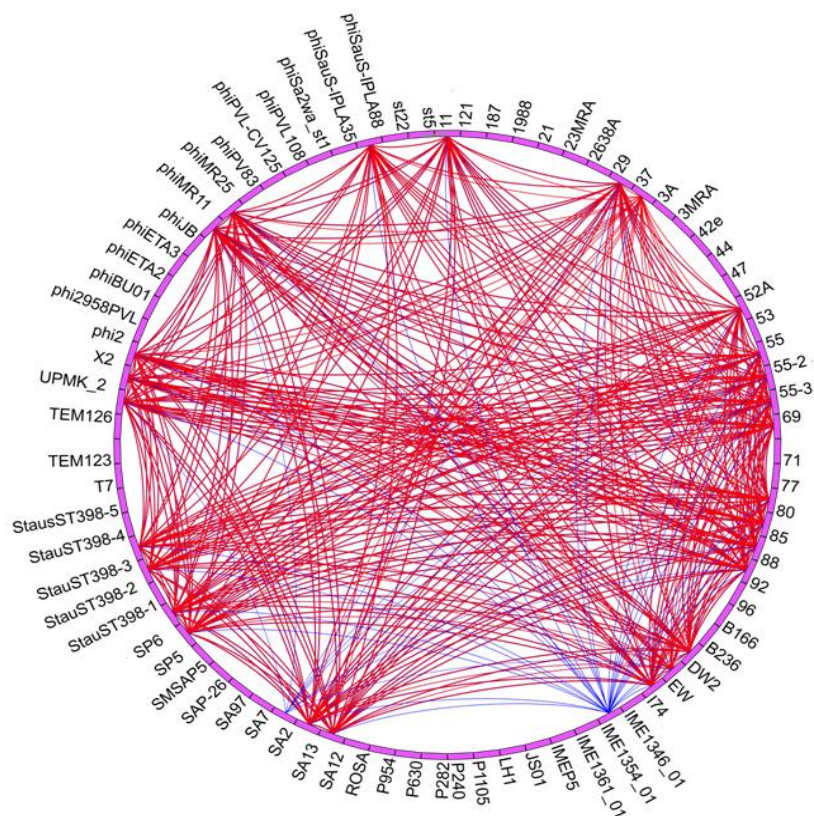
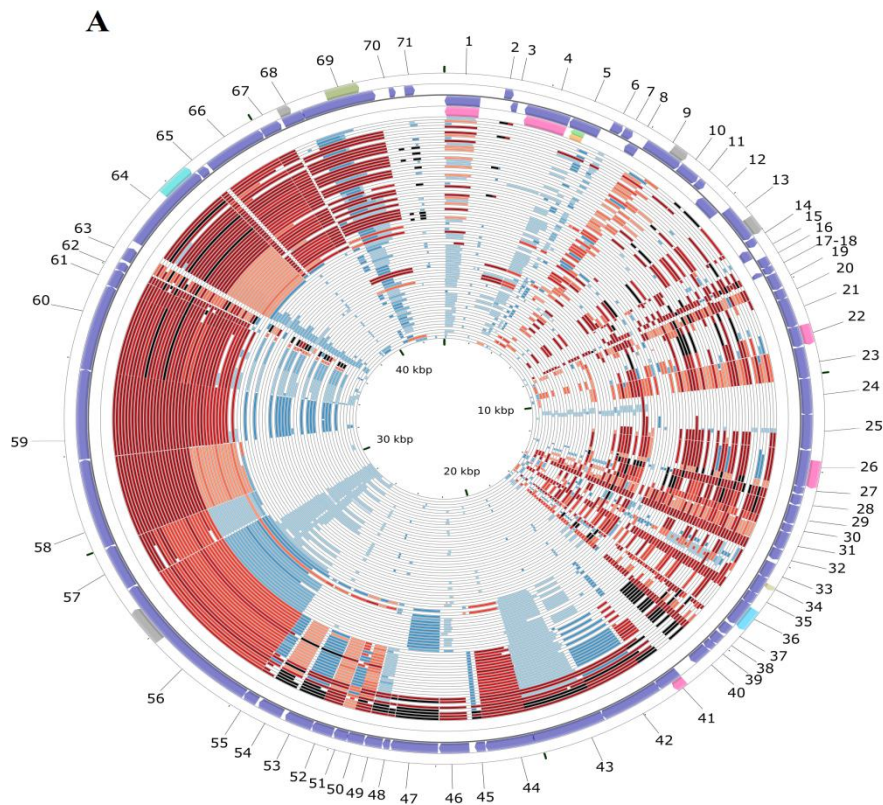


Figura 14. Gráfico circular de la familia de las holinas (pham23). Los productos génicos conectados por líneas rojas son incluidos en este pham debido a que el E-value es $<1e-50$ (BLAST). Los productos génicos conectados por líneas azules son incluidos en el pham debido a que ellos comparten más de 32.5% de identidad (ClustalW).

7.3.3. Análisis CG view

El programa CG View Comparison Tool realiza un análisis comparativo mediante BLAST, y los representa mediante gráficos circulares (Grant, Arantes, and Stothard 2012). En dicho estudio, se compararon los mismos fagos utilizados para generar la base de datos de Phamerator. En la **Figura 15** se observan los resultados obtenidos de la comparación mediante BLASTp del fago I74, utilizado como referencia y representado en el anillo más exterior versus el resto de los fagos. En estos gráficos, se indican los números de ORFs correspondientes. El número de ORF se encuentra marcado para una mejor identificación de los mismos. El porcentaje de homología en las distintas regiones del genoma se representa según la escala de colores de las figuras 5B y 6B (porcentajes de identidad de los gráficos circulares representada por colores).

El orden de los fagos es decreciente respecto a la homología entre el fago I74 y los demás fagos del NCBI (dato no mostrado). Los gráficos de genomas de fagos que poseen mayor homología con I74, son los siguientes: phiJB, phiETA3, 96, 3MRA, 29, B166, DW2, 88, StauST398-3, 80, 55-3, B236, SA13, UPMK-2, 55-2, 52A, 55, ROSA, 92, phiMR11, 71, StauST398-1, X2 y StauST398-5.

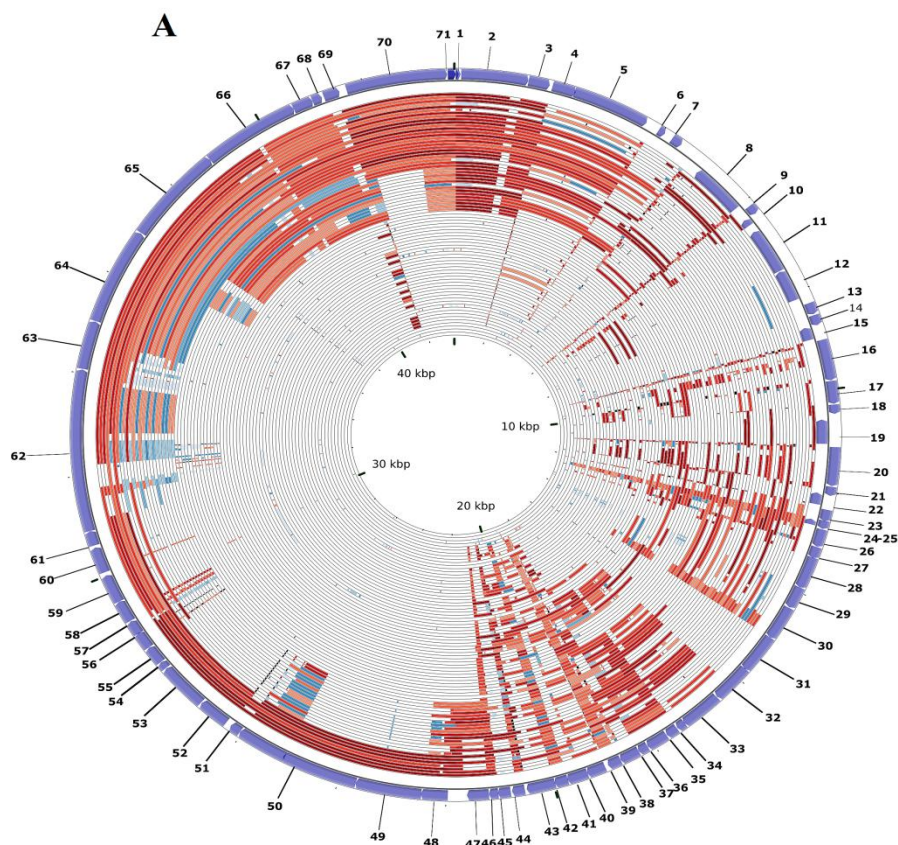


B



Figura 15. A. Gráfico circular de proteínas del fago I74 comparado con 78 fagos de *Staphylococcus* de la base de datos. **B.** Código de colores de acuerdo al % de homología según BLASTp.

Según el gráfico circular (**Figura 15.A**), se observó un alto porcentaje de homología total (50%-90%, **Figura 15.B**) en la región que abarca el ORF 56 al ORF 69, entre los fagos. En la región restante hay variabilidad en el porcentaje de homología total. Las zonas de mayor homología se deben a la presencia de secuencias de genes conservados pertenecientes a los módulos morfológicos de cola y lisis, principalmente.



B



Figura 16. A. Gráfico circular del DNA del fago I74 comparado con 78 *Syphofagos* de *Staphylococcus* de la base de datos. **B.** Código de colores de acuerdo al % de homología según Blastn.

Observando el porcentaje de homología total a nivel de ADN (**Figura 16.B**) en el gráfico circular, existe una alta homología total (90%-98% gama de rojos) entre los fagos, en la región del genoma que abarca del ORF 1 a 5 y del 62 a 71. En la región restante se observa variabilidad de la misma.

El orden de los fagos es decreciente respecto a la homología entre el fago I74 y los demás fagos del NCBI (dato no mostrado). Los gráficos de genomas de fagos que poseen mayor homología con I74, son los siguientes: phiJB, 3MRA, phiETA3, DW2, 96, StauST398-3, B166, 29, 88, ROSA, 55-3, SA13, 80, B236, 55-2, UPMK-2, 55, 52A, 92, phiMR11, 71, X2, StauST398-5 y StauST398-1.

Comparando ambos gráficos circulares (**Figura 15.A y Figura 16.A**), se observa una diferencia significativa de homología total a nivel de nucleótidos, pudiendo concluir el alto porcentaje de homología total entre el bacteriófago I74 y los bacteriófagos de la base de datos a nivel de aminoácidos, teniendo en cuenta los ORFs de los fagos en estudio.

7.3.4. Análisis de la similitud de secuencia

La gran cantidad de datos genómicos que existe para micobacteriofagos (fagos que infectan micobacterias) ha impulsado el desarrollo de muchas herramientas y métodos para la clasificación de estos fagos. Se ha demostrado que el uso de un único gen, en este caso TMP (proteína tapemeasure) era suficiente para la clasificación de los fagos en diferentes clústeres (**Smith et al. 2013**).

Se realizó un análisis mediante dot-plot utilizando el programa GEPARD con las secuencias de aminoácidos de las TMPs de los fagos. Estas secuencias se reordenaron manualmente (formato multifasta), hasta reproducir agrupamiento definidos de fagos de acuerdo al patrón. Las líneas sólidas implican una alta homología entre las secuencias comparadas. De esta manera, se determinó claramente, la formación de 5 grupos (clústeres), representadas como patrones de líneas más intensas (**Figura 17**).

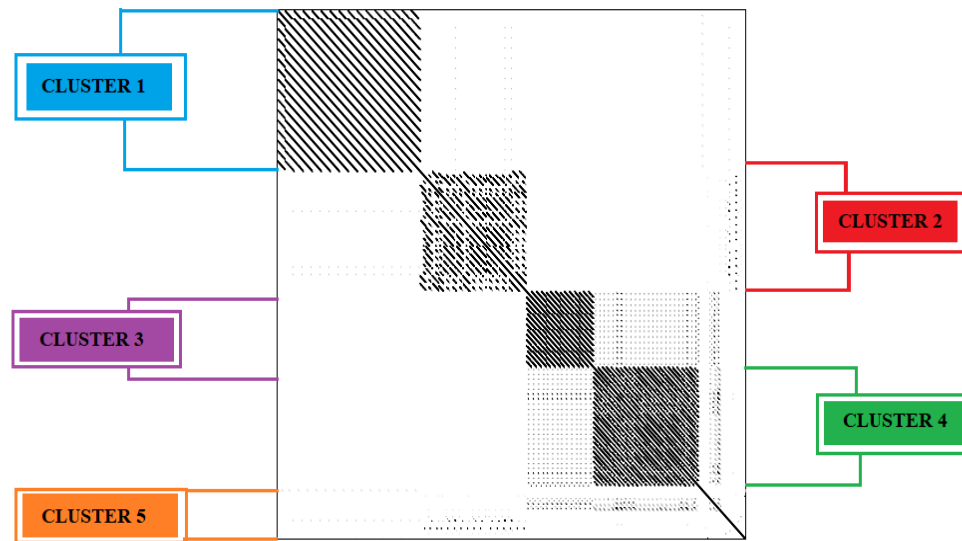


Figura 17. Gráfico de puntos de los 79 bacteriófagos de *Staphylococcus aureus*. En base a la homología de secuencias de aminoácidos de las TMPs, se observan 5 grupos (Referencias, **Tabla III**). El clúster 5 no está bien definido

Tabla III. Referencias de colores de los bacteriófagos de *Staphylococcus aureus* agrupados en clústeres

CLUSTER 1	CLUSTER 2	CLUSTER 3	CLUSTER 4	CLUSTER 5
3A	P954	SA7	29	IMEP5
42e	phiPVL-CN125	53	88	187
st5	phiPVL108	85	phiMR11	SA2
47	77	phiMR25	SA13	37
SMSAP5	P1105	SP5	52A	EW
1988	P630	TEM123	DW2	2638A
44	IME1346_01	SA12	StauST398-3	IME1354_01
phiSauS-IPLA35	P282	SP6	96	
21	st22	TEM126	ROSA	
P240	23MRA	phiSauS-IPLA88	phiJB	
phiSa2wa_st1	StauST398-4	SA97	StauST398-1	
121	IME1361_01	SAP-26	55-3	
LH1	T7	11	55-2	
phi2958PVL	phiBU01	69	55	
StauST398-2	JS01	phiETA2	71	
phi2			B236	
phiPV83			80	

			UPMK_2	
			92	
			X2	
			3MRA	
			I74	
			StauST398-5	
			B166	
			phiETA3	

7.3.5. Análisis Gegenees

Este programa utiliza un algoritmo que aplica alineamiento fragmentado para facilitar los análisis comparativos de cientos de genomas microbianos. Los genomas son fragmentados y comparados todos contra todos, utilizando BLASTn. Además, da un panorama filogenómico de los genomas y los alineamientos pueden entonces estar minados por regiones genómicas, donde existen matchs en grupos de genes con regiones conservadas (Ågren et al. 2012).

En este estudio se determinó la relación filogenética entre el bacteriófago I74 y los *Siphofagos* de *Staphylococcus aureus* de la base de datos. Se utilizaron los genomas de 79 bacteriófagos, alineándose y así se construyó la red filogenética (**Figura 18**). La red filogenética, se construyó con el software SplitsTree 4.14.6 (Huson 1998) utilizando el método basado en distancias “Neighbor-Net”. Se observó, una clara distinción de cinco clústeres al igual que en el análisis de GEPARD. Los círculos mostrando los clústeres se realizaron en función de los datos de GEPARD a modo de comparación. En ambos análisis el bacteriófago I74 pertenece al clúster 4 (**círculo rojo**). El fago 187 perteneciente al clúster 5 en GEPARD, se encuentra distantemente relacionado en el análisis de Gegenees, pudiendo ser clasificado como singleton (unico miembro).

Realizando una comparación de estos resultados con los obtenidos con CG View, los fagos mas homólogos con I74 se ubican en el mismo clúster, lo mismo sucede con los resultados de GEPARD, por lo que existe correspondencia entre los agrupamientos o clustereado.

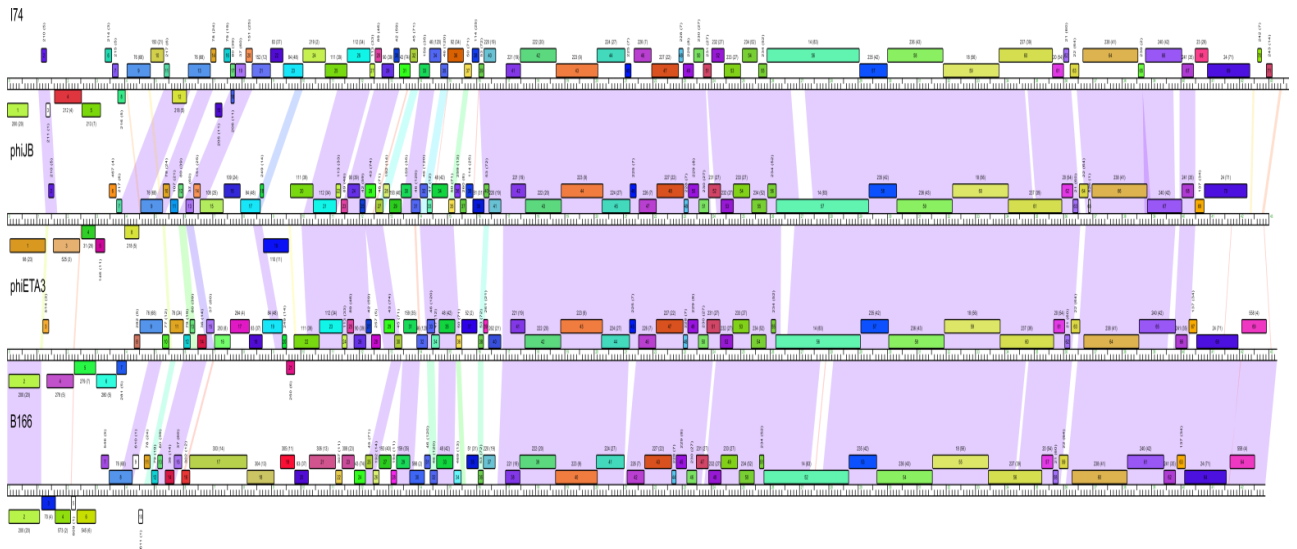


Figura 19. Mapas genómicos de los fagos más homólogos a I74. Sombreado lila: homología alta (E-value 10^{-50}). Sombreado rojo: homología baja (<32,5% identidad). No sombreado/blanco: no hay similitud significativa.

En la **Figura 19** se observa, que hay mayor homología en los módulos de morfogénesis de cápside y de cola; entre los genomas de los bacteriófagos I74, phiJB, phiETA3 y B166. Esto también puede observarse en los mapas circulares de CG View.

7.3.7. Análisis del %GC de los Siphostaphylofagos

Con la herramienta bioinformática Phamerator, se extrajeron los datos del contenido de GC (%GC) de los fagos. Luego se realizó un boxplot con RStudio (**Figura 20**) y se aplicó estadística descriptiva para determinar la mediana y la variancia de cada clúster. En el boxplot, se relacionó el %GC con los clústeres de los bacteriófagos. El fago SA2 se decidió excluirlo del análisis (presenta un %GC de 31,92, inusualmente bajo), ya que dispersa significativamente los valores de %GC de los demás fagos en el clúster 5. El fago I74 se ubica en el clúster 4 y tiene un %GC de 35,25.

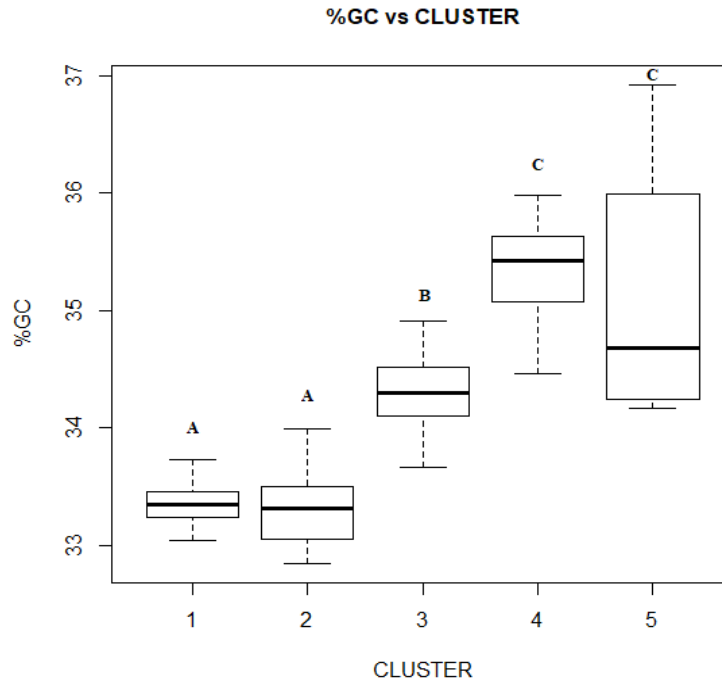


Figura 20. Relación del % GC versus clústeres de los *Siphostaphylofagos*. Los boxplots indicados con diferentes letras (A, B y C) muestran grupos que son estadísticamente diferentes ($p < 0.05$). Esto se demostró mediante la utilización del test de comparaciones múltiples de Tukey.

En la **Figura 20**, se observa que los clústeres 1 y 2 poseen medianas similares (33,35 y 33,31%GC respectivamente) mientras que los clústeres 4 y 5 sus medianas difieren levemente (35,36 y 34,66%GC, respectivamente) pero difieren significativamente con respecto a los clústeres anteriores. El clúster 3 posee una mediana de 34,31%GC; distinto de los demás clústeres. Las variancias de los clústeres 2, 3 y 4 son casi similares, mientras que los clusteres 1 y 5 difieren entre si y con los anteriores. Mediante la utilización de tests de comparaciones multiples se observa que existen diferencias significativas en %GC entre el clúster 3 y los demas clústeres; mientras que entre los clústeres 1 y 2 no hay diferencias significativas al igual que los clústeres 4 y 5. Estos últimos difieren significativamente con los demás clústeres. En el clúster 5 existe una alta dispersión con respecto a los demás clusteres.

7.3.8. Análisis filogenético de los *Siphostaphylofagos*

Análisis filogenético mediante la utilización del programa CAFE

Este programa se basa en una plataforma de análisis libre de alineamiento basado en *k-mer* general para estudiar las relaciones entre genomas y metagenomas.

Permite la visualización de las medidas de disimilitud de pares de genomas; incluyendo dendogramas, heatmap, análisis de las coordenadas principales y muestreo de redes.

La **Figura 21** muestra el dendrograma generado por CAFÉ, lo que permite obtener una visualización clara de las relaciones filogenéticas del bacteriófago I74 con los *Siphofagos* de la base de datos. Los datos obtenidos mediante este análisis se corresponden con los estudios previos (GEPARD), por lo que el bacteriófago I74 se ubica correctamente en el clúster 4. Existe un grupo de fagos que se observa filogenéticamente más distantes del resto, los cuales pertenecen a diferentes clústeres asignados por otros análisis, los cuales los hemos marcado con asterisco.

En este estudio se incorporaron en el programa en un orden específico los genomas de los fagos, teniendo en cuenta como se había realizado el análisis de GEPARD. Utilizamos el algoritmo Manhattan (Ma) con parámetros preestablecidos variando los valores de $K=4$, 6 y 8. Se utilizó el algoritmo Ma, ya que se ajustó con mayor precisión a los resultados esperados. Con los parámetros $K=4$ y 6 (análisis no mostrado), no se observó una clara distinción de clústeres, contrario a lo observado con $K=8$, por lo tanto, se trabajó con este valor de K (**Lu et al. 2017**). Podemos observar un patrón de colores distinto (gama de azules) en cada clúster, esto se debe a la diferencia en el grado de homología entre los pares de fagos. Los clústeres se evidencian por zonas más oscuras (mayor homología) entre los pares de fagos (**Figura 22**).

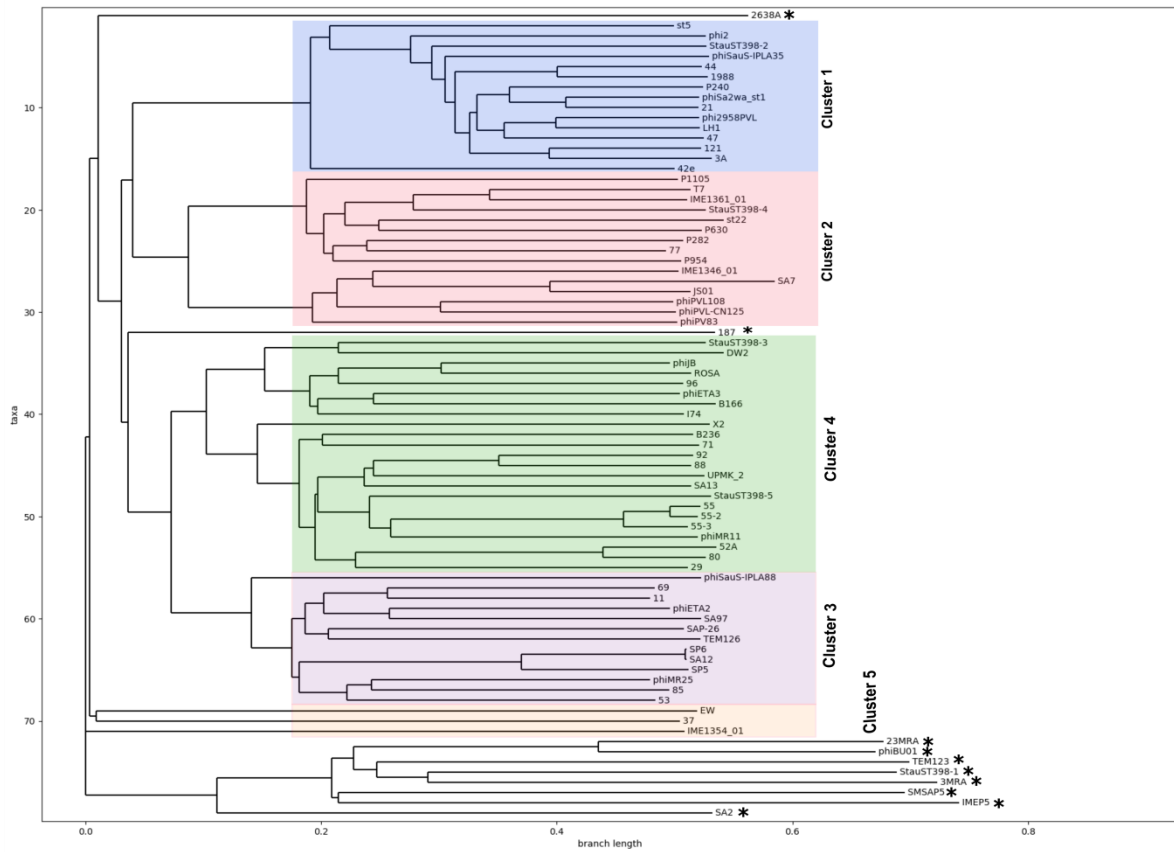


Figura 21. Dendrograma obtenido mediante CAFÉ. Análisis filogenético del bacteriófago I74 con respecto a los demás *Siphofagos* de la base de datos. Colores de los clústeres en base al GEPARD; **clúster 1:** azul, **clúster 2:** rojo; **clúster 3:** violeta; **clúster 4:** verde; **clúster 5:** naranja. (*) Fagos que no comparten homología con ningún clúster.

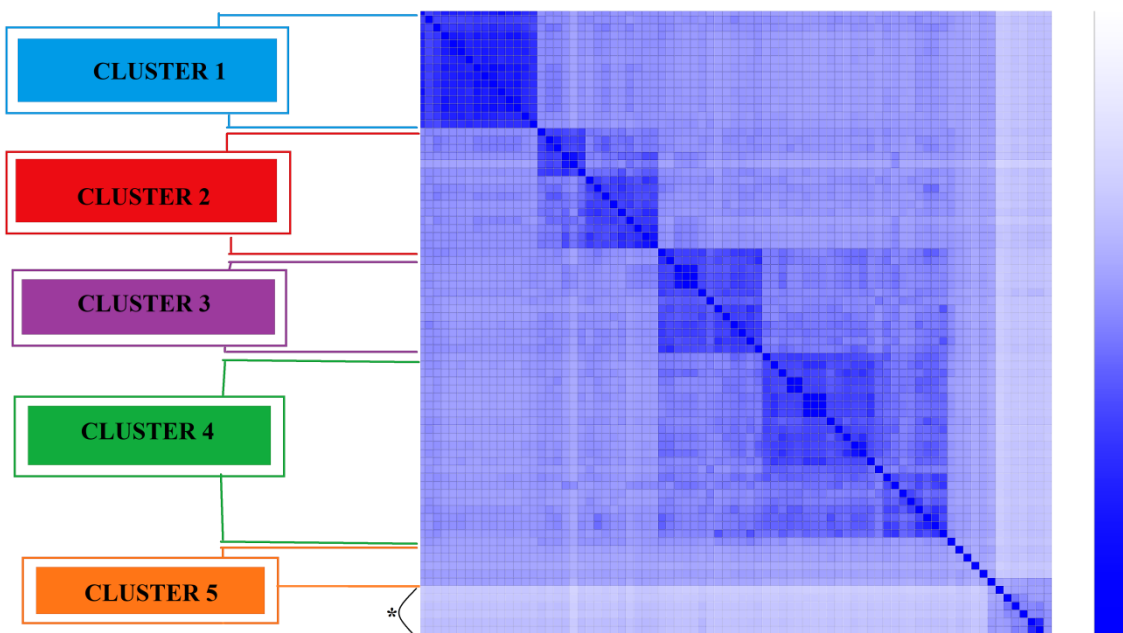


Figura 22. Heatmap (Ma, K=8), gráfica de puntos de los 79 bacteriófagos de *Staphylococcus aureus* en base a la homología general de las secuencias de nucleótidos. Tanto en el eje y como el eje x se encuentran los bacteriófagos de SA. (*) Fagos que no comparten homología con ningún clúster.

Análisis mediante el cálculo de la distancia Mash

Mash, se basa en la técnica de reducción de dimensionalidad MinHash. Reduce secuencias grandes y *sets* de secuencias a pequeñas, *sketches* representativos, del cual las distancias de mutaciones globales pueden ser rápidamente estimadas. La distancia de *Mash* puede ser rápidamente computada por los *sketches* reducido en tamaño solamente, ya que produce un resultado que se correlaciona fuertemente con las medidas basadas en alineamientos, tales como el ANI (identidad de nucleótido en promedio) (Ondov et al. 2016).

Las similitudes de las secuencias de nucleótidos de los genomas enteros fue calculado usando Mash. Para fagos se ha observado que el incremento en el tamaño de *sketch* incrementa la precisión en el cálculo de las distancias, cuando se compararon genomas distantemente relacionados.

Debido a que Mash es un cálculo libre de alineamiento de similitud, no se correlaciona estrictamente con la divergencia evolutiva, y esta característica distintiva del no alineamiento tanto como la composición de bases, pueden influir en los valores de Mash (Mavrich and Hatfull 2017).

En este estudio se determinó la relación filogenética de I74 con los Siphofagos de la base de datos utilizando los genomas de los mismos (Figura 23). Comparando con GEPARD, Gegenees y CAFE algunos bacteriófagos no se encuentran en los mismos clústeres previos. Esto puede deberse a la utilización de diferentes algoritmos y parámetros los cuales pueden requerir mas ajustes. Con el cálculo de las distancia y la construcción del dendrograma, en este caso los clústeres 3 y 4 (donde se encuentra el fago I74) forman un único clúster que nombramos 3 (violeta) y el clúster 5 en el cluster 4 (verde). El fago I74 se lo muestra remarcado mediante un círculo rojo en el clúster 3.

Comparando este análisis con los anteriores, se observó que el fago 187 difiere notablemente del GEPARD, pudiendo reasignarlo al clúster 3, ya que se encuentran cerca (Gegenees) o en la misma rama (Mash y CAFE), así se pudo predecir la alta

relación filogenética de este fago con los demás fagos de este clúster.

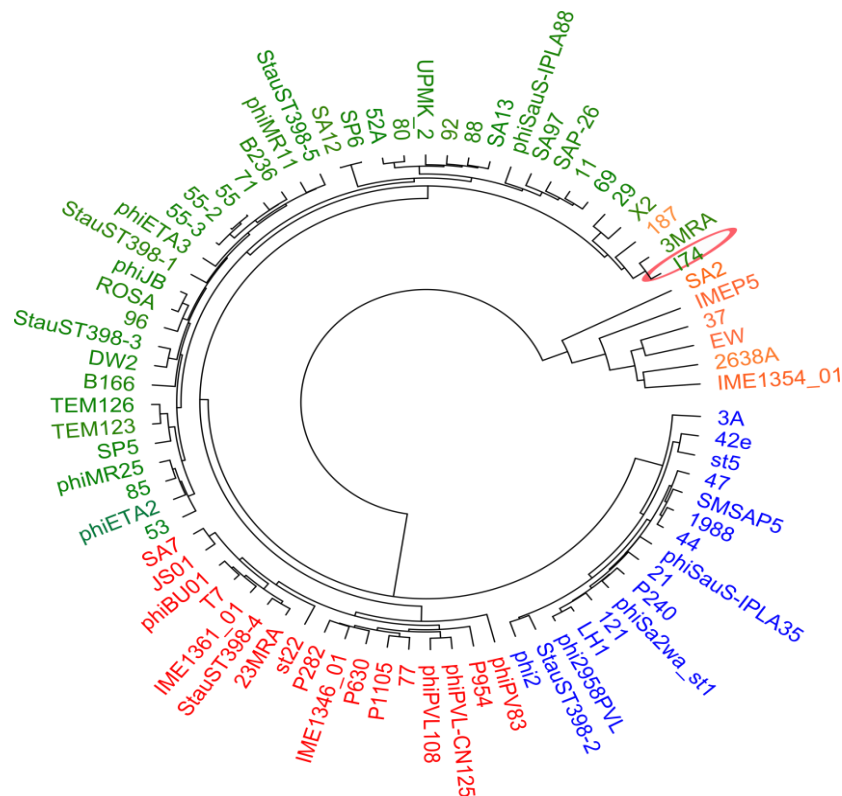


Figura 23. Dendrograma construido con los valores de distancia de Mash. Los colores de los clústeres se definieron según el GEPARD; clúster 1: azul; clúster 2: rojo; clúster 3: verde; clúster 4: naranja. Los parámetros utilizados fueron; *mash sketch -k 17 -s 50000 -i -o* y *mash dist -t*, se ajustaron mejor a los resultados esperados.

Análisis de disimilitudes del contenido de genes para pares de genomas (GCD) de Staphylofagos

GCD, es un parámetro que se utiliza para determinar las relaciones entre fagos. Determina la disimilitud del contenido genético por comparaciones de a pares de genomas entre los 79 fagos de *S. aureus*. Se calcula como la proporción de phams (familias) relativos compartidos para cada número total de genomas de phams, promediando las dos proporciones y sustrayendo desde 1. El rango de valores desde 1, donde los genes no están compartidos (los dos fagos son 100% disimilares), a 0 donde el contenido génico es idéntico.

Se construyó una tabla con los valores de GCD que posee el fago I74 contra todos los fagos *S. aureus* y sus respectivos clústeres. Los valores de ésta tabla fueron cargados en RStudio 1.1.456, así se generó una gráfica de boxplot con los valores de

GCD (mostrándose la mediana y la dispersión de los mismos) en función de los clústeres (**Figura 24**).

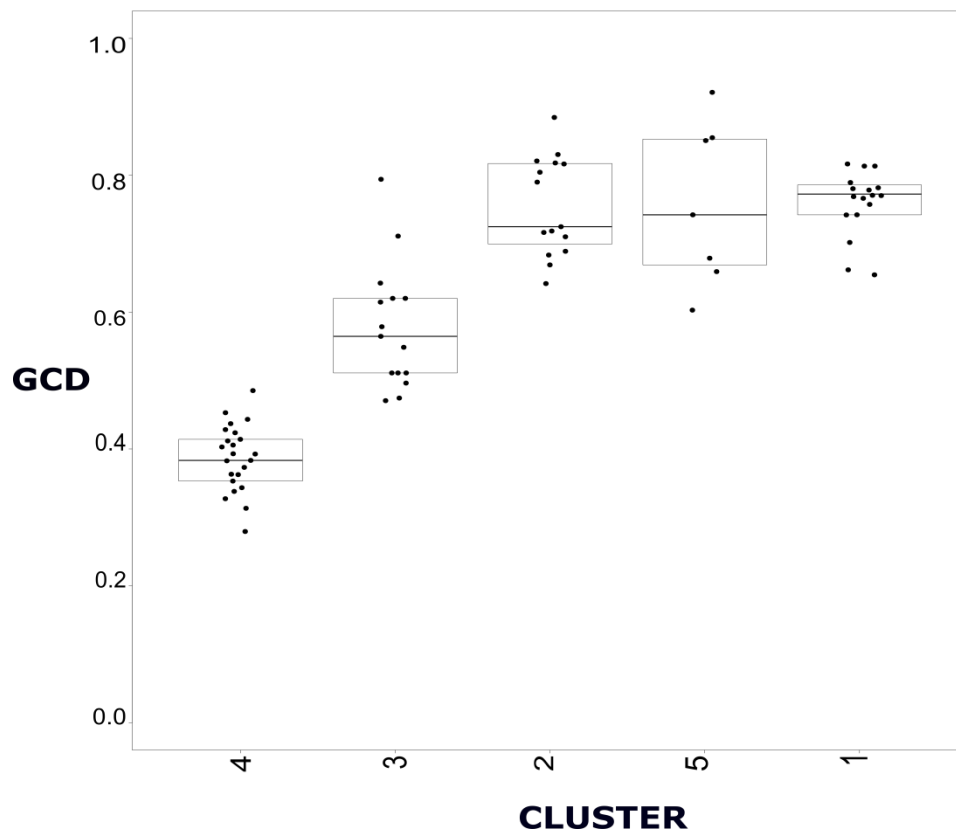


Figura 24. Análisis GCD de *Staphylofagos* de SA en función de los diferentes clústeres.

En este estudio (**Figura 24**), se observó la diversidad de los fagos de *S. aureus*, representado por la distribución de los valores de GCD (genes compartidos entre dos genomas de fagos). En este caso, se acotó al estudio de la disimilitud del contenido genético sólo entre el fago I74 en función de los demás fagos.

Cuando se compara un fago con otros miembros del mismo clúster, los valores de GCD deberían ser bajos, ya que ellos deberían compartir una más alta proporción de sus genes, mientras que si lo comparamos con fagos de diferentes clústeres, los valores de GCD deberían incrementarse (**Pope et al. 2017**).

Cuando se compara I74 contra los miembros del clúster 4 se observa un valor de GCD promedio $<0,4$. Cuando se compara con el resto de los clústeres se observaron valores promedios de GCD mayores, lo que confirma que I74 comparte un número

mayor de genes con los fagos de dicho clúster. Además hay poca dispersión con respecto a la mediana entre I74 y los fagos de este clúster, por lo que se puede verificar que están relacionados.

En el clúster 3, el valor de GCD promedio $<0,6$ similar a lo anterior. Pero hay mayor dispersión con respecto a la mediana, por lo que entre I74 y los fagos de este clúster habría una mayor diferencia entre los genes compartidos. En los clústeres 2, 5 y 1 se observan valores de GCD promedio $>0,7$; por lo que hay baja similitud genética entre el fago I74 y los demás fagos de estos clústeres. Además, en el clúster 1 hay poca dispersión con respecto a la mediana, lo que indicaría una menor variabilidad génica entre los fagos perteneciente a este clúster. Con los clústeres 2 y 5, I74 comparte una baja proporción de sus genes, siendo el clúster 5 el que posee mayor variabilidad de genes. Por lo que, este estudio concuerda con los análisis previos, ya que se observa variabilidad entre los clústeres y además el fago I74 tiene mayor similitud genética con los fagos del clúster 4.

Análisis de las proteínas TMPs

Este estudio se llevó a cabo con el software Mega X (Análisis de genéticas evolucionaria molecular), el cual implementa métodos analíticos y herramientas para filogenómica y filomedicina. Ha sido actualizado para diversos análisis de evolución molecular (**Kumar et al. 2018**).

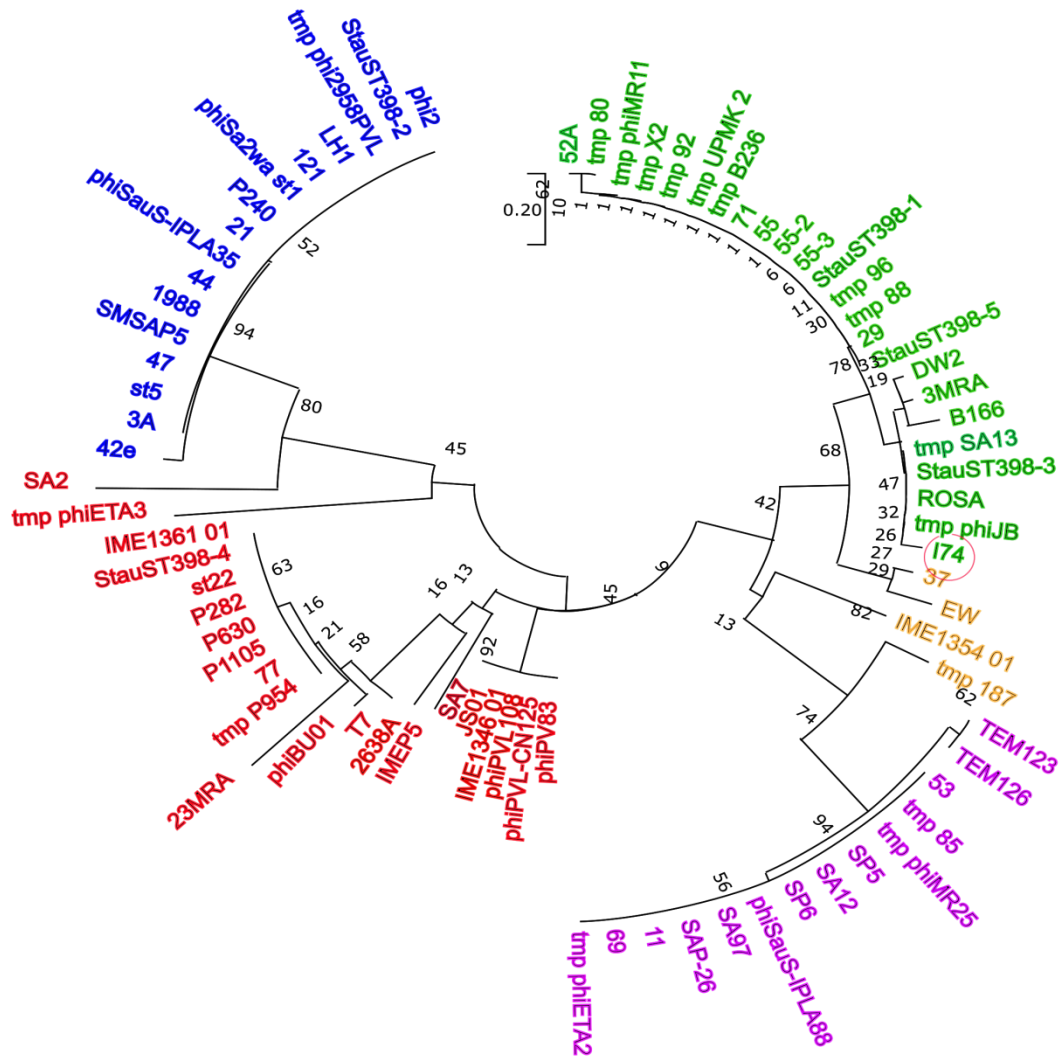


Figura 25. Análisis filogenético de las TMPs de los bacteriófagos de *S. aureus* de la familia *Siphoviridae*. Método Neighbor-Joining. El alineamiento se realizó mediante el algoritmo Muscle.

Se usó la secuencia TMP de fagos, debido a que es una proteína conservada y como se mencionó anteriormente en el análisis GEPARD, se demostró en un estudio que el método de gen único, identifica relaciones de fagos tanto si se utiliza el genoma completo o el gen de TMPs para la comparación. En el árbol filogenético, la TMP del bacteriófago I74 (**Figura 25, círculo rojo**) comparte mayor homología con las TMP de los fagos del clúster 4.

Análisis de la subunidad grande de las terminasas

Los bacteriófagos con cola contienen un cromosoma, DNA doble hebra lineal, los cuales tienen un tamaño de 18 a 500 kpb (determinado de los fagos con cola

conocidos, miembros del orden *Caudovirales*). Estos cromosomas lineales pueden tener un tipo de extremo conocido, que pueden ser; extremos cohesivos (5' o 3' extensiones simple hebra), repeticiones terminales directas permutadas circularmente, repeticiones terminales directas largas o cortas, secuencias de DNA huésped terminales o proteínas terminales unidas covalentemente. Estos tipos diferentes de extremos, reflejan estrategias diferentes de replicación de DNA y especialmente difieren las acciones de las terminasas (enzimas) durante el empaquetamiento del DNA (crean distintos extremos de DNA del fago). Estas enzimas son levemente variables entre los bacteriófagos, pero sigue siendo la más conservada entre los mismos (**Casjens 2003**).

En general la determinación de la secuencia genómica completa de la terminasa, no elucidada la naturaleza de estos extremos, se requieren de otros análisis experimentales.

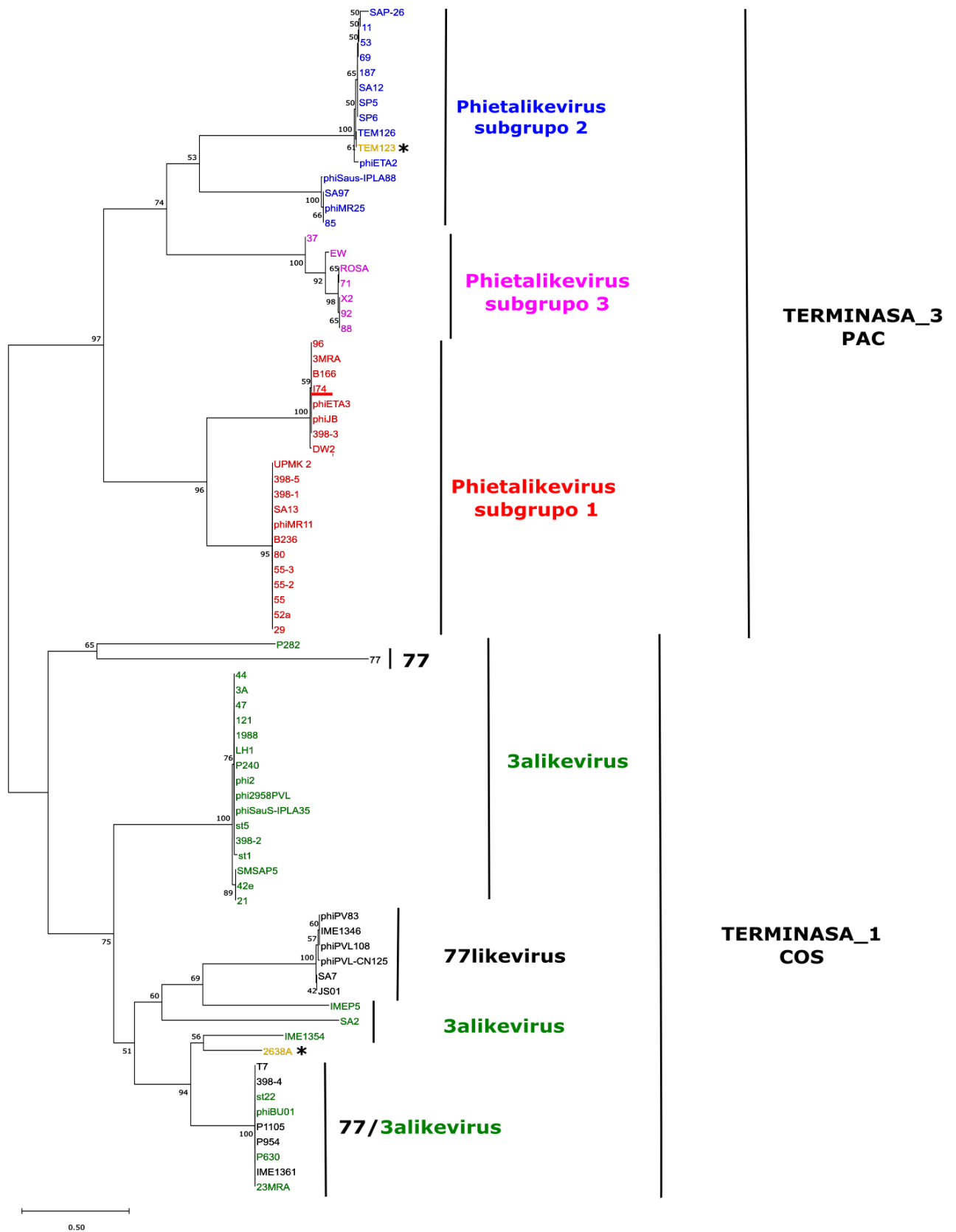


Figura 26. Análisis filogenético de las subunidades grandes de las terminasas de los bacteriófagos de la familia *Siphoviridae* de *Staphylococcus*. (*) Fagos Orfans (comparten algunas propiedades con los

bacteriófagos de estos géneros propuestos en la figura). Método de corrección de Poisson. Alineamiento mediante ClustalW.

Se realizó el análisis filogenético por comparación de las secuencias de aminoácidos de las subunidades grandes de las terminasas. Los resultados del análisis se muestran en la **Figura 26**, donde se observa en el árbol filogenético de estas proteínas, lo que se correlaciona al tipo de extremo de ADN que producen. (Chang et al. 2015).

Se observó que la subunidad grande de la terminasa del bacteriófago I74 se encuentra en el grupo de *phietalikevirus subgrupo 1* por lo que posee una terminasa con dominio *terminasa_3* y por lo tanto genera extremos de ADN tipo *pac* para empaquetar su genoma (**línea roja en la Figura 26**). Esto se verificó, utilizando la herramienta BLASTp donde se observó que la terminasa de I74 contiene el dominio “*terminasa_3*” de la familia PBSX.

Cabe destacar, que los grupos 3a y 77 se definieron teniendo en cuenta el gen que codifica para un activador de la integrasa, RinA o RinB y en caso que dicho gen no estuviese presente, se tuvo en cuenta el gen que codifica para la amidasa o helicasa y ADN polimerasa A. El grupo 3a se distingue por la presencia de fagos que tienen RinA y el grupo 77 por helicasa y ADN polimerasa A o amidasa (Casjens & Gilcrease 2009). Dada esta clasificación, y comparándola con los grupos definidos anteriormente, no hubo una correlación marcada. Igualmente, se observó que el fago I74 se relaciona principalmente con la mayoría de los fagos que pertenecen al clúster 4.

Análisis de las integrasas

Los bacteriófagos temperados llevan a cabo procesos de integración, escisión e inversión de segmentos de DNA definidos que ocurren comúnmente a través de una recombinación sitio-específica, dichos procesos le permiten integrar su genoma al cromosoma bacteriano. Esto es llevado a cabo por enzimas específicas de fagos, llamadas recombinasas sitio-específicas o integrasas. Las recombinasas se dividen en dos familias, dependiendo del residuo aminoácido que interviene en la unión covalente de ADN-Proteína en una reacción intermedia y su homología de secuencia de aminoácidos; serin o tirosin recombinasas (o integrasas). Sus mecanismos de recombinación son totalmente distintos. Además, la integrasa posee la secuencia de nucleótidos bien conservada entre los grupos de fagos, esto refleja la diversidad de la población de fagos de *S. aureus*.

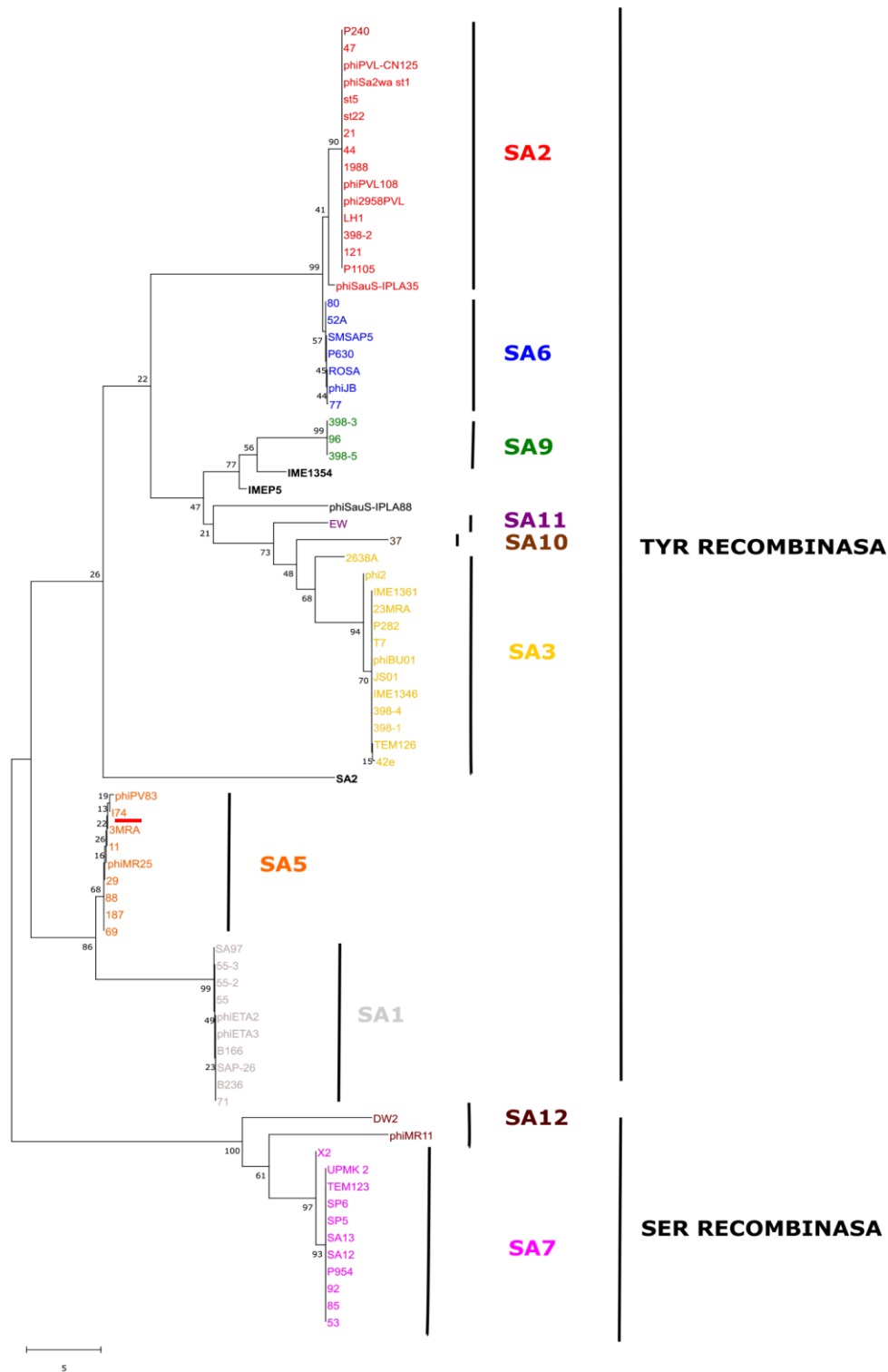


Figura 27. Análisis filogenético de las integrasas. Colores definidos según el tipo de integrasa de cada fago. Método “Neighbor-Joining Tree”. La distancia evolucionaria se computó con el Método de Maximun composite likelihood y el alineamiento mediante ClustalW.

En la **Figura 27**, se observa el análisis filogenético realizado sobre las integrasas de los bacteriófagos. Existen estudios donde se han analizado y clasificado a las integrasas de fagos de *S. aureus* en diferentes grupos (SA1 a 12) (**Goerke et al. 2009**). De la misma forma, se incluyeron las secuencias de las integrasas de los fagos que estamos analizando, y mediante la observación de las relaciones filogenéticas podemos predecir el tipo de integrasa que posee.

Para la identificación del gen de la integrasa en todos los fagos analizados hicimos uso del programa Phamerator. Donde estaba definido que el ORF 1 del fago I74 codifica para la integrasa, se determinó a que familia corresponde y que fagos se hallaban en la misma. Esto me permitió predecir los genes que codifican para integrasa en algunos fagos, que su anotación no estaba completa en la base de datos.

Haciendo el análisis de las integrasas de los bacteriófagos por la herramienta BLASTp y los programas bioinformáticos HHPRED, Phamerator y Pfam, se determinó que el fago I74 posee una integrasa de la familia tirosin-recombinasa (grupo SA5, **línea roja en la Figura 27**). Se realizó el análisis para todos los fagos, y luego para la construcción del árbol filogenético, se utilizó las secuencias de los nucleótidos de los mismos. Además el fago I74 posee el gen escisionasa, poco común en fagos de *Staphylococcus* (**Tesis Peressutti, 2018**). También se observó, que difiere notablemente del GEPARD y el análisis con las terminasas, ya que el I74 no se encuentra agrupado con los mismos bacteriófagos del clúster 4.

8. Discusión

Actualmente, el auge en el desarrollo de las herramientas bioinformáticas acompañada con el creciente aislamiento y caracterización de los bacteriófagos temperados de *Staphylococcus aureus*, han permitido avanzar en el entendimiento y obtener una gran cantidad de información respecto a la genética y las relaciones filogenéticas de estos fagos.

Staphylococcus aureus es un patógeno de relevancia mundial debido a las infecciones que causa tanto en humanos como animales. Posee varios factores de virulencia algunos codificados por bacteriófagos que permiten o favorecen que esta bacteria pueda adaptarse y colonizar diferentes nichos. Entre ellos podemos encontrar, leucocidina Pantón-Valentina (*lukSF*), toxina exfoliativa A (*eta*), proteína SasX anclada a la pared celular y el clúster de evasión inmune compuesto de enterotoxina S (*sea*), estafiloquinasa (*sak*), la proteína inhibitoria de quimiotaxis (*chp*) y el inhibidor del complemento estafilococal (*scn*). (Carroll, Cafferkey, and Coleman 1993; Lee and Iandolo 1986).

El estudio de fagos capaces de infectar a este patógeno es de importancia, desde dos puntos muy distintos, por un lado aporta información para ayudar a entender los mecanismos de transferencias de genes (muchos de estos involucrados en la patogenia), y por otro lado permite evaluar el uso de los mismos o enzimas que ellos codifican (por ejemplo, endolisinas) como agentes de biocontrol.

El desarrollo de este trabajo de tesina proporcionó un avance en el estudio de la filogenética, genómica comparativa y caracterización genómica del bacteriófago temperado llamado I74, el cual tiene la capacidad de infectar específicamente cepas de *S. aureus*.

Como primer punto, hemos realizado la caracterización genómica de este fago, para ello se analizaron los datos del ensamblado del mismo, seleccionando el que poseía el genoma completo. Pudimos observar que posee un tamaño de genoma de 43685 pb, perteneciendo a la clase II (familia *Siphoviridae*) y tiene la capacidad de integrarse ya que se obtuvo mediante inducción con mitomicina C (Tesis Peressutti, 2018). Posee un bajo %GC, codifica para 71 ORFs, el 57,75% de los genes poseen funciones hipotéticas

y no posee intrones. Pudimos determinar la organización modular de estos genes, rasgo característico de estos organismos.

Es imprescindible resaltar que para considerar utilizar un fago como agente de biocontrol (tanto su uso en animales, medioambiente y en humanos), el mismo debe cumplir la condición de ser estrictamente lítico. Esto se debe a que algunos fagos temperados poseen genes codificantes de factores de virulencia, los cuales pueden utilizarlos como vehículo de intercambio genético con la bacteria por transducción (**Pantůček et al. 2004; Goerke et al. 2009**). En este trabajo realizamos una búsqueda exhaustiva de la presencia de genes de virulencia y de resistencia a antibióticos en este fago, obteniendo resultados negativos. Sin embargo, a pesar de carecer de dichos genes aun posee la capacidad de integrarse y puede adquirirlos en el pasaje sobre otras cepas de *S. aureus*, por lo que no puede utilizarse como agente de biocontrol.

El análisis genómico comparativo a nivel de aminoácidos y nucleótidos entre I74 y los distintos *Staphylofagos* incluidos desde la base de datos de NCBI, mostró mayor homología a nivel de aminoácidos. Lo cual es lógico pensando en la conservación de las funciones que llevan a cabo muchas de estas proteínas. Esto se vió claramente en los análisis de CG View. Un menor grado de homología a nivel de ADN, se puede deber a cambios en las frecuencias de uso de codones, las mutaciones y eventos de transferencia horizontal/vertical de genes entre fago-hospedador bacteriano, los cuales dirigen su adaptación a un nicho ecológico específico (**Hatfull and Hendrix 2011**).

El modulo de lisis está compuesto por una lisina (endolisina) y una holina, esta última forma poros en la membrana de la bacteria permitiendo el acceso de la lisina al peptidoglicano. El análisis de las holinas mediante Phamerator, nos permitió clasificar las mismas en tres familias (pham23, pham137 y pham63). La holina codificada por el fago I74 (ORF68) pertenece al pham23, y posee alta homología con 29 miembros.

Posteriormente, nos planteamos la utilización de diferentes métodos bioinformáticos para el agrupamiento o clusterización de los fagos. Se ha demostrado para fagos que infectan micobacterias que la utilización del gen que codifica para la proteína Tapemeasure es suficiente para agrupar los fagos en clústeres. Por ello, analizamos si esto es extrapolable a los fagos de *S. aureus*. En este caso, las secuencias de aminoácidos de las TMPs fueron analizadas en el programa GEPARD, donde se obtuvo un gráfico de dot-plot donde se observa la formación de cinco clústeres de fagos

bien definidos. Estos resultados también concuerdan con el árbol filogenético de las TMPs (MEGA X). Luego, realizamos un análisis por homología de nucleótidos de los genomas totales mediante el programa Gegenees. Este análisis se ajustó a lo observado mediante el uso de GEPARD, formándose cinco clústeres, en este caso se aprecia que los miembros del clúster 5 se encuentran más distantemente relacionados entre sí. Nuestro fago I74 se encuentra agrupado en el clúster 4 en ambos análisis junto con 24 miembros más, cabe señalar que es el clúster más numeroso.

De la red filogenética obtenida por Gegenees, se seleccionaron tres de los fagos más relacionados a I74 y luego se extrajo el mapa de los alineamientos genómicos de los mismos realizado por Phamerator. En el mismo, se observó mayor homología en el módulo morfogenético de cabeza (terminasas) y cola (TMPs) y baja homología en el módulo de lisis. Esto se comprobó con el CG View de proteínas donde existe alta homología (rojo) entre los ORFs de cabeza y cola y baja homología entre los ORFs de holina (celeste) de los fagos que se encuentran en el clúster 4. Estos resultados se corresponden con el análisis de las holinas, donde se determinó que las pertenecientes al pham23 (I74) difieren de aquellos que se encuentran en el clúster 4.

La variabilidad en el contenido en GC de los fagos puede ser indicio de la adaptación a diferentes hospedadores. Para ello, analizamos como varía el contenido en GC de los fagos en función de los clústeres asignados. Este análisis mostró que existe similitud en este parámetro entre algunos clústeres (1 vs 2 y 4 vs 5). También pudimos observar que un miembro del clúster 5 el fago 2638A, posee el %GC (36,92) más alto con respecto a los demás fagos y además en este clúster existe mucha variabilidad de %GC entre los fagos. Este resultado concuerda con los obtenidos en Gegenees donde los miembros de este clúster están más alejados entre sí. También podemos mencionar que en promedio todos los clústeres poseen mayor contenido en GC que la bacteria hospedadora cuyo %GC es de 32.7%.

Los avances en secuenciación y la obtención de cantidades enormes de datos genómicos y de secuencia han impulsado en el último tiempo a desarrollar diferentes herramientas bioinformáticas para su análisis (**Zielezinski et al. 2017**). Los métodos comparación libres de alineamiento son un ejemplo de ello. Para hacer uso de estas herramientas elegimos los programas CAFE y Mash. Los cálculos de la distancia de Mash y su posterior grafica mediante dendograma mostró que los clústeres 3 y 4 (se

ubica I74) forman un único grupo de fagos relacionados. Además, el análisis realizado con CAFE, mostró a diferencia de Mash una separación entre clúster 3 y 4, salvo por la presencia de algunos fagos que mediante este análisis se mostraron distantemente relacionados. Observamos también que el fago 187, es un posible singleton más relacionado al clúster 4. Cabe mencionar, que estos análisis también arrojaron resultados que muestran la existencia de baja homología entre los fagos del clúster 5 (distancia evolutiva grande) y divergen significativamente de la mayoría de los clústeres en estudio.

Debido a la incorporación de cada vez más genomas secuenciados de micobacterias en la base de datos phagesde.org, es cada vez más complicado su agrupamiento en clústeres. Este grupo desarrollo un método de comparación que involucra el cálculo del parámetro GCD (**Pope et al. 2017**). El análisis de este parámetro mostró que nuestro fago I74 comparte en promedio un mayor número de genes con los miembros del clúster 4. Observándose nuevamente que el contenido de genes es más variable cuando se compara con los miembros de clúster 5.

El análisis de la terminasa codificada por el fago I74 nos permite demostrar que el mismo posee una forma de empaquetamiento utilizando sitios *pac* y se clasifica en el grupo Phietalikevirus subgrupo 1 (**Chang et al. 2015**). Por otro lado, el estudio de la integrasa mostró que la misma es de tipo tirosin-recombinasa y se clasifica en el grupo SA5 (**Goerke et al. 2009**).

9. Conclusión

Esta tesina se focalizó en el estudio de un bacteriófago temperado de la familia *Siphoviridae* específico de *S. aureus*, y obtener un amplio conocimiento a nivel genómico de fagos de esta especie. Para esto, nos centramos en la caracterización y análisis filogenéticos del mismo mediante la utilización de herramientas bioinformáticas.

Se trabajó con un bacteriófago inducido por Mitomicina C, de una cepa H7 de *S. aureus* hospitalaria, y pudimos determinar mediante su caracterización la presencia de genes que codifican para funciones líticas y lisogénicas, destacando la escisionasa poco común en estos tipos de fagos.

El desarrollo de este trabajo, nos permitió mediante el uso de numerosas herramientas bioinformáticas la anotación de un fago de *S. aureus* y su comparación genómica con fagos presentes en la base de datos. Estos resultados indicaron que el fago I74 pertenece a un grupo nombrado clúster 4 junto con 24 miembros más, donde se observó que entre los miembros hay una alta homología en los módulos morfogenéticos de cabeza y cola, y además poseen diferencias en los módulos lisogénicos y de lisis. Por lo tanto, hemos obtenido una amplia información a nivel genómico y proteico, como también un enfoque sistemático en la diversidad y evolución de los mismos.

10. Bibliografía

1. Abatángelo, Virginia, Natalia Peressutti Bacci, Carina A. Boncompain, Ariel A. Amadio, Soledad Carrasco, Cristian A. Suárez, and Héctor R. Morbidoni. 2017. “Broad-Range Lytic Bacteriophages That Kill *Staphylococcus Aureus* Local Field Strains” edited by R. K. Aziz. *PLOS ONE* 12(7):e0181671.
2. Ackermann, Hans-Wolfgang and Michael S. DuBow. 1987. *Viruses of Prokaryotes*. CRC Press.
3. Ågren, Joakim, Anders Sundström, Therese Håfström, and Bo Segerman. 2012. “Gegenees: Fragmented Alignment of Multiple Genomes for Determining Phylogenomic Distances and Genetic Signatures Unique for Specified Target Groups” edited by N. Ahmed. *PLoS ONE* 7(6):e39107.
4. Bae, Taeok, Tadashi Baba, Keiichi Hiramatsu, and Olaf Schneewind. 2006. “Prophages of *Staphylococcus Aureus* Newman and Their Contribution to Virulence.” *Molecular Microbiology* 62(4):1035–47.
5. Belcaid, Mahdi, Anne Bergeron, and Guylaine Poisson. 2010. “Mosaic Graphs and Comparative Genomics in Phage Communities.” *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology* 17(9):1315–26.
6. Besemer, John and Mark Borodovsky. 2005. “GeneMark: Web Software for Gene Finding in Prokaryotes, Eukaryotes and Viruses.” *Nucleic Acids Research* 33(Web Server issue):W451-4.
7. Black, L. W. 1989. “DNA Packaging in DsDNA Bacteriophages.” *Annual Review of Microbiology* 43(1):267–92.
8. Bradley, D. E. 1967. “Ultrastructure of Bacteriophage and Bacteriocins.” *Bacteriological Reviews* 31(4):230–314.
9. Brussow, H., C. Canchaya, and W. D. Hardt. 2004. “Phages and the Evolution of Bacterial Pathogens: From Genomic Rearrangements to Lysogenic Conversion.” *Microbiology and Molecular Biology Reviews* 68(3):560–602.
10. Brüßow, H. and F. Desiere. 2001. “Comparative Phage Genomics and the Evolution of Siphoviridae: Insights from Dairy Phages.” *Molecular*

- Microbiology* 39(2):213–22.
11. Brüssow, Harald. 2001. “Phages of Dairy Bacteria.” *Annual Review of Microbiology* 55(1):283–303.
 12. Brüssow, Harald and Roger W. Hendrix. 2002. “Phage Genomics: Small Is Beautiful.” *Cell* 108(1):13–16.
 13. Canchaya, Carlos, Caroline Proux, Ghislain Fournous, Anne Bruttin, and Harald Brüssow. 2003. “Prophage Genomics.” *Microbiology and Molecular Biology Reviews: MMBR* 67(2):238–76, table of contents.
 14. Carroll, J. D., M. T. Cafferkey, and D. C. Coleman. 1993. “Serotype F Double- and Triple-Converting Phage Insertionally Inactivate the *Staphylococcus Aureus* β -Toxin Determinant by a Common Molecular Mechanism.” *FEMS Microbiology Letters* 106(2):147–55.
 15. Carver, Tim, Matthew Berriman, Adrian Tivey, Chinmay Patel, Ulrike Böhme, Barclay G. Barrell, Julian Parkhill, and Marie-Adèle Rajandream. 2008. “Artemis and ACT: Viewing, Annotating and Comparing Sequences Stored in a Relational Database.” *Bioinformatics* 24(23):2672–76.
 16. Casjens, Sherwood. 2003. “Prophages and Bacterial Genomics: What Have We Learned so Far?” *Molecular Microbiology* 49(2):277–300.
 17. Casjens, Sherwood R. and Eddie B. Gilcrease. 2009. “Determining DNA Packaging Strategy by Analysis of the Termini of the Chromosomes in Tailed-Bacteriophage Virions.” *Methods in Molecular Biology (Clifton, N.J.)* 502:91–111.
 18. Chandry, P. S., B. E. Davidson, and A. J. Hillier. 1994. “Temporal Transcription Map of the *Lactococcus Lactis* Bacteriophage Sk1.” *Microbiology* 140(9):2251–61.
 19. Chang, Yoonjee, Hakdong Shin, Ju-Hoon Lee, Chul Park, Soon-Young Paik, and Sangryeol Ryu. 2015. “Isolation and Genome Characterization of the Virulent *Staphylococcus Aureus* Bacteriophage SA97.” *Viruses* 7(10):5225–42.
 20. Christie, Gail E. and Terje Dokland. 2012. “Pirates of the Caudovirales.” *Virology* 434(2):210–21.
 21. Cluzel, P. J., J. Serio, and J. P. Accolas. 1987. “Interactions of *Lactobacillus Bulgaricus* Temperate Bacteriophage 0448 with Host Strains.” *Applied and Environmental Microbiology* 53(8):1850–54.

22. Cresawn, Steven G., Matt Bogel, Nathan Day, Deborah Jacobs-Sera, Roger W. Hendrix, and Graham F. Hatfull. 2011. "Phamerator: A Bioinformatic Tool for Comparative Bacteriophage Genomics." *BMC Bioinformatics* 12(1):395.
23. Deghorain, Marie and Laurence Van Melderen. 2012. "The Staphylococci Phages Family: An Overview." *Viruses* 4(12):3316–35.
24. Desiere, Frank, Sacha Lucchini, and Harald Brüssow. 1999. "Comparative Sequence Analysis of the DNA Packaging, Head, and Tail Morphogenesis Modules in the Temperate Cos-Site Streptococcus Thermophilus Bacteriophage Sfi21." *Virology* 260(2):244–53.
25. Desiere, Frank, Sacha Lucchini, and Harald Bru Èssow. 1998. *Evolution of Streptococcus Thermophilus Bacteriophage Genomes by Modular Exchanges Followed by Point Mutations and Small Deletions and Insertions.*
26. Echols, H., C. Lu, and P. M. Burgers. 1983. "Mutator Strains of Escherichia Coli, MutD and DnaQ, with Defective Exonucleolytic Editing by DNA Polymerase III Holoenzyme." *Proceedings of the National Academy of Sciences of the United States of America* 80(8):2189–92.
27. Finn, Robert D., Penelope Coggill, Ruth Y. Eberhardt, Sean R. Eddy, Jaina Mistry, Alex L. Mitchell, Simon C. Potter, Marco Punta, Matloob Qureshi, Amaia Sangrador-Vegas, Gustavo A. Salazar, John Tate, and Alex Bateman. 2016. "The Pfam Protein Families Database: Towards a More Sustainable Future." *Nucleic Acids Research* 44(D1):D279–85.
28. Franche, Nathalie, Manon Vinay, and Mireille Ansaldi. 2017. "Substrate-Independent Luminescent Phage-Based Biosensor to Specifically Detect Enteric Bacteria Such as E. Coli." *Environmental Science and Pollution Research* 24(1):42–51.
29. Ganguly, Tridib, Malabika Das, Amitava Bandhu, Palas K. Chanda, Biswanath Jana, Rajkrishna Mondal, and Subrata Sau. 2009. "Physicochemical Properties and Distinct DNA Binding Capacity of the Repressor of Temperate Staphylococcus Aureus Phage Φ11." *FEBS Journal* 276(7):1975–85.
30. Goerke, C., J. Koller, and C. Wolz. 2006. "Ciprofloxacin and Trimethoprim Cause Phage Induction and Virulence Modulation in Staphylococcus Aureus." *Antimicrobial Agents and Chemotherapy* 50(1):171–77.

31. Goerke, C., R. Pantucek, S. Holtfreter, B. Schulte, M. Zink, D. Grumann, B. M. Broker, J. Doskar, and C. Wolz. 2009. "Diversity of Prophages in Dominant *Staphylococcus Aureus* Clonal Lineages." *Journal of Bacteriology* 191(11):3462–68.
32. Grant, Jason R., Adriano S. Arantes, and Paul Stothard. 2012. "Comparing Thousands of Circular Genomes Using the CGView Comparison Tool." *BMC Genomics* 13(1):202.
33. Grant, Jason R. and Paul Stothard. 2008. "The CGView Server: A Comparative Genomics Tool for Circular Genomes." *Nucleic Acids Research* 36(Web Server issue):W181-4.
34. Gutiérrez, Diana, Evelien M. Adriaenssens, Beatriz Martínez, Ana Rodríguez, Rob Lavigne, Andrew M. Kropinski, and Pilar García. 2014. "Three Proposed New Bacteriophage Genera of Staphylococcal Phages: '3alikevirus', '77likevirus' and 'Phietalikevirus.'" *Archives of Virology* 159(2):389–98.
35. Hatfull, Graham F. 2008. "Bacteriophage Genomics." *Current Opinion in Microbiology* 11(5):447–53.
36. Hatfull, Graham F. and Roger W. Hendrix. 2011. "Bacteriophages and Their Genomes." *Current Opinion in Virology* 1(4):298–303.
37. Hsieh, Sue-Er, Hsueh-Hsia Lo, Shui-Tu Chen, Mong-Chuan Lee, and Yi-Hsiung Tseng. 2011. "Wide Host Range and Strong Lytic Activity of *Staphylococcus Aureus* Lytic Phage Stau2." *Applied and Environmental Microbiology* 77(3):756–61.
38. Huson, D. H. 1998. "SplitsTree: Analyzing and Visualizing Evolutionary Data." *Bioinformatics* 14(1):68–73.
39. Kahánková, Jana, Roman Pantůček, Christiane Goerke, Vladislava Růžičková, Pavla Holochová, and Jiří Doškař. 2010. "Multilocus PCR Typing Strategy for Differentiation of *Staphylococcus Aureus* Siphoviruses Reflecting Their Modular Genome Structure." *Environmental Microbiology* 12(9):2527–38.
40. Klaenhammer, T. R. and G. F. Fitzgerald. 1994. "Bacteriophages and Bacteriophage Resistance." Pp. 106–68 in *Genetics and Biotechnology of Lactic Acid Bacteria*. Dordrecht: Springer Netherlands.
41. Krumsiek, J., R. Arnold, and T. Rattei. 2007. "Gepard: A Rapid and

- Sensitive Tool for Creating Dotplots on Genome Scale.” *Bioinformatics* 23(8):1026–28.
42. Kumar, Sudhir, Glen Stecher, Michael Li, Christina Knyaz, and Koichiro Tamura. 2018. “MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms” edited by F. U. Battistuzzi. *Molecular Biology and Evolution* 35(6):1547–49.
43. Kwan, T., J. Liu, M. DuBow, P. Gros, and J. Pelletier. 2005. “The Complete Genomes and Proteomes of 27 Staphylococcus Aureus Bacteriophages.” *Proceedings of the National Academy of Sciences* 102(14):5174–79.
44. Kwan, Tony, Jing Liu, Michael DuBow, Philippe Gros, and Jerry Pelletier. 2005. “The Complete Genomes and Proteomes of 27 Staphylococcus Aureus Bacteriophages.” *Proceedings of the National Academy of Sciences of the United States of America* 102(14):5174–79.
45. Lee, C. Y. and J. J. Iandolo. 1986. “Lysogenic Conversion of Staphylococcal Lipase Is Caused by Insertion of the Bacteriophage L54a Genome into the Lipase Structural Gene.” *Journal of Bacteriology* 166(2):385–91.
46. Li, Suiqiong, Yugui Li, Huiqin Chen, Shin Horikawa, Wen Shen, Aleksandr Simonian, and Bryan A. Chin. 2010. “Direct Detection of Salmonella Typhimurium on Fresh Produce Using Phage-Based Magnetoelastic Biosensors.” *Biosensors and Bioelectronics* 26(4):1313–19.
47. Lindsay, Jodi A. and Matthew T. G. Holden. 2004. “Staphylococcus Aureus : Superbug, Super Genome?” *Trends in Microbiology* 12(8):378–85.
48. Łobocka, Małgorzata, Monika S. Hejnowicz, Kamil Dąbrowski, Agnieszka Gozdek, Jarosław Kosakowski, Magdalena Witkowska, Magdalena I. Ulatowska, Beata Weber-Dąbrowska, Magdalena Kwiatek, Sylwia Parasion, Jan Gawor, Helena Kosowska, and Aleksandra Głowacka. 2012. “Genomics of Staphylococcal Twort-like Phages - Potential Therapeutics of the Post-Antibiotic Era.” Pp. 143–216 in *Advances in virus research*. Vol. 83.
49. Lu, Timothy K. and Michael S. Koeris. 2011. “The next Generation of Bacteriophage Therapy.” *Current Opinion in Microbiology* 14(5):524–31.
50. Lu, Yang Young, Kujin Tang, Jie Ren, Jed A. Fuhrman, Michael S. Waterman, and Fengzhu Sun. 2017. “CAFE: ACcelerated Alignment-FrEe Sequence Analysis.” *Nucleic Acids Research* 45(W1):W554–59.
51. Lu, Z., E. Altermann, F. Breidt, and S. Kozyavkin. 2010. “Sequence

- Analysis of Leuconostoc Mesenteroides Bacteriophage Phi1-A4 Isolated from an Industrial Vegetable Fermentation.” *Applied and Environmental Microbiology* 76(6):1955–66.
52. Lucchini, Sacha, Frank Desiere, and Harald Brüssow. 1999. “Similarly Organized Lysogeny Modules in Temperate Siphoviridae from Low GC Content Gram-Positive Bacteria.” *Virology* 263(2):427–35.
53. Mann, Nicholas H. 2008. “The Potential of Phages to Prevent MRSA Infections.” *Research in Microbiology* 159(5):400–405.
54. Le Marrec, C., D. van Sinderen, L. Walsh, E. Stanley, E. Vlegels, S. Moineau, P. Heinze, G. Fitzgerald, and B. Fayard. 1997. “Two Groups of Bacteriophages Infecting Streptococcus Thermophilus Can Be Distinguished on the Basis of Mode of Packaging and Genetic Determinants for Major Structural Proteins.” *Applied and Environmental Microbiology* 63(8):3246–53.
55. Mavrigh, Travis N. and Graham F. Hatfull. 2017. “Bacteriophage Evolution Differs by Host, Lifestyle and Genome.” *Nature Microbiology* 2:17112.
56. Merabishvili, Maya, Jean-Paul Pirnay, Gilbert Verbeken, Nina Chanishvili, Marina Tediashvili, Nino Lashkhi, Thea Glonti, Victor Krylov, Jan Mast, Luc Van Parys, Rob Lavigne, Guido Volckaert, Wesley Mattheus, Gunther Verween, Peter De Corte, Thomas Rose, Serge Jennes, Martin Zizi, Daniel De Vos, and Mario Vaneechoutte. 2009. “Quality-Controlled Small-Scale Production of a Well-Defined Bacteriophage Cocktail for Use in Human Clinical Trials” edited by D. M. Ojcius. *PLoS ONE* 4(3):e4944.
57. Merrill, Bryan D., Andy T. Ward, Julianne H. Grose, and Sandra Hope. 2016. “Software-Based Analysis of Bacteriophage Genomes, Physical Ends, and Packaging Strategies.” *BMC Genomics* 17(1):679.
58. Moineau, Sylvain. 1999. “Applications of Phage Resistance in Lactic Acid Bacteria.” *Antonie van Leeuwenhoek* 76(1/4):377–82.
59. Moore, P. C. L. and J. A. Lindsay. 2001. “Genetic Variation among Hospital Isolates of Methicillin-Sensitive Staphylococcus Aureus: Evidence for Horizontal Transfer of Virulence Genes.” *Journal of Clinical Microbiology* 39(8):2760–67.
60. Morgan, G. J. and W. B. Pitts. 2008. “Evolution without Species: The Case of Mosaic Bacteriophages.” *The British Journal for the Philosophy of*

Science 59(4):745–65.

61. Navarre, W. W., H. Ton-That, K. F. Faull, and O. Schneewind. 1999. “Multiple Enzymatic Activities of the Murein Hydrolase from Staphylococcal Phage Phi11. Identification of a D-Alanyl-Glycine Endopeptidase Activity.” *The Journal of Biological Chemistry* 274(22):15847–56.
62. Nelson, Daniel, Raymond Schuch, Peter Chahales, Shiwei Zhu, and Vincent A. Fischetti. 2006. “PlyC: A Multimeric Bacteriophage Lysin.” *Proceedings of the National Academy of Sciences of the United States of America* 103(28):10765–70.
63. Novick, Richard P., Gail E. Christie, and Jose R. Penadés. 2010. “The Phage-Related Chromosomal Islands of Gram-Positive Bacteria.” *Nature Reviews Microbiology* 8(8):541–51.
64. O’Flaherty, S., A. Coffey, W. J. Meaney, G. F. Fitzgerald, and R. P. Ross. 2005. “Inhibition of Bacteriophage K Proliferation on Staphylococcus Aureus in Raw Bovine Milk.” *Letters in Applied Microbiology* 41(3):274–79.
65. Ondov, Brian D., Todd J. Treangen, Páll Melsted, Adam B. Mallonee, Nicholas H. Bergman, Sergey Koren, and Adam M. Phillippy. 2016. “Mash: Fast Genome and Metagenome Distance Estimation Using MinHash.” *Genome Biology* 17(1):132.
66. Pantůcek, R., J. Doskar, V. Růžicková, P. Kaspárek, E. Oráčová, V. Kvardová, and S. Rosypal. 2004. “Identification of Bacteriophage Types and Their Carriage in Staphylococcus Aureus.” *Archives of Virology* 149(9):1689–1703.
67. Parreira, R., R. Valyasevi, A. L. Lerayer, S. D. Ehrlich, and M. C. Chopin. 1996. “Gene Organization and Transcription of a Late-Expressed Region of a Lactococcus Lactis Phage.” *Journal of Bacteriology* 178(21):6158–65.
68. Pope, Welkin H., Travis N. Mavrigh, Rebecca A. Garlena, Carlos A. Guerrero-Bustamante, Deborah Jacobs-Sera, Matthew T. Montgomery, Daniel A. Russell, Marcie H. Warner, Science Education Alliance-Phage Hunters Advancing Genomics and Evolutionary Science Science Education Alliance-Phage Hunters Advancing Genomics and Evolutionary Science (SEA-PHAGES), and Graham F. Hatfull. 2017. “Bacteriophages of

- Gordonia Spp. Display a Spectrum of Diversity and Genetic Relationships.” *MBio* 8(4):e01069-17.
69. Ptashne, Mark. 1986. *A Genetic Switch: Gene Control and Phage [Lambda]*. Cell Press.
70. Resch, Alexandra, Birgit Fehrenbacher, Klaus Eisele, Martin Schaller, and Friedrich GÃ¶tz. 2005. “Phage Release from Biofilm and Planktonic *Staphylococcus Aureus* Cells.” *FEMS Microbiology Letters* 252(1):89–96.
71. Sass, Peter and Gabriele Bierbaum. 2007. “Lytic Activity of Recombinant Bacteriophage Phi11 and Phi12 Endolysins on Whole Cells and Biofilms of *Staphylococcus Aureus*.” *Applied and Environmental Microbiology* 73(1):347–52.
72. Schmelcher, Mathias, David M. Donovan, and Martin J. Loessner. 2012. “Bacteriophage Endolysins as Novel Antimicrobials.” *Future Microbiology* 7(10):1147–71.
73. Sheehan, Michelle M., JosÃ©L Garcia, Rubens LÃ³pez, and Pedro Garcia. 1996. “Analysis of the Catalytic Domain of the Lysin of the Lactococcal Bacteriophage Tuc2009 by Chimeric Gene Assembling.” *FEMS Microbiology Letters* 140(1):23–28.
74. Smith, Kyle C., Eduardo Castro-Nallar, Joshua NB Fisher, Donald P. Breakwell, Julianne H. Grose, and Sandra H. Burnett. 2013. “Phage Cluster Relationships Identified through Single Gene Analysis.” *BMC Genomics* 14(1):410.
75. Soding, J., A. Biegert, and A. N. Lupas. 2005. “The HHpred Interactive Server for Protein Homology Detection and Structure Prediction.” *Nucleic Acids Research* 33(Web Server):W244–48.
76. Tolba, M., O. Minikh, L. Y. Brovko, S. Evoy, and M. W. Griffiths. 2010. “Oriented Immobilization of Bacteriophages for Biosensor Applications.” *Applied and Environmental Microbiology* 76(2):528–35.
77. Wan, Jiehui, Huihua Shu, Shichu Huang, Ben Fiebor, I. Hsuan Chen, Valery A. Petrenko, and Bryan A. Chin. 2007. “Phage-Based Magnetoelastic Wireless Biosensors for Detecting *Bacillus Anthracis* Spores.” *IEEE Sensors Journal* 7(3):470–77.
78. Watanabe, K. and S. Takesue. 1973. “Energy Requirement for the Formation of Blender-Resistant Complexes in *Lactobacillus* Phage Infection.” *Journal*

- of General Virology* 20(3):319–26.
79. Watanabe, K., S. Takesue, and K. Ishibashi. 1979. “Adenosine Triphosphate Content in *Lactobacillus Casei* and the Blender-Resistant Phage-Cell Complex-Forming Ability of Cells on Infection with PL-1 Phage.” *Journal of General Virology* 42(1):27–36.
80. Watanabe, Kenji, Mitsuhiro Shirabe, Yukihiko Nakashima, and Yukari Kakita. 1991. *The Possible Involvement of Protein Synthesis in the Injection of PL-1 Phage Genome into Its Host, LactobaciZZus Casei*. Vol. 137.
81. Xia, Jufeng, Jianjun Gao, Norihiro Kokudo, Kiyoshi Hasegawa, and Wei Tang. 2013. “Methicillin-Resistant *Staphylococcus Aureus* Antibiotic Resistance and Virulence.” *Bioscience Trends* 7(3):113–21.
82. Ye, Z. H. and C. Y. Lee. 1993. “Cloning, Sequencing, and Genetic Characterization of Regulatory Genes, *RinA* and *RinB*, Required for the Activation of *Staphylococcal* Phage Φ 11 Int Expression.” *Journal of Bacteriology* 175(4):1095–1102.
83. Zhang, Qian, Shaozhen Xing, Qiang Sun, Guangqian Pei, Shi Cheng, Yannan Liu, Xiaoping An, Xianglilan Zhang, Yonggang Qu, and Yigang Tong. 2017. “Characterization and Complete Genome Sequence Analysis of a Novel Virulent Siphoviridae Phage against *Staphylococcus Aureus* Isolated from Bovine Mastitis in Xinjiang, China.” *Virus Genes* 53(3):464–76.
84. Zielezinski, Andrzej, Susana Vinga, Jonas Almeida, and Wojciech M. Karlowski. 2017. “Alignment-Free Sequence Comparison: Benefits, Applications, and Tools.” *Genome Biology* 18(1):186.
85. Zourob, Mohammed and Steven Ripp. 2010. “Bacteriophage-Based Biosensors.” Pp. 415–48 in *Recognition Receptors in Biosensors*. New York, NY: Springer New York.
86. https://www.researchgate.net/publication/319351357_La_comercializacion_de_organismos_modificados_geneticamente_aspectos_normativos_e_institucionales. 08/03/19. 15:03.
87. <http://fbio.uh.cu/sites/genmol/confs/conf5/p03.htm>. 08/03/19. 16:36.
88. <https://epidemiologiamolecular.com/bacteriofagos/> 11/02/19 08:59.
89. <https://www.redbubble.com/es/people/bacteriophage/works/23845913-familias-de-bacteri-fagos?p=poster> 11/02/19 09:07

90. <https://phagesdb.org/DNAMaster/>. 09/03/19 15:30
91. <http://exon.gatech.edu/GeneMark/> 09/03/19 15:40
92. <https://github.com/younglululu/CAFE> 09/03/19 16:06
93. www.megasoftware.net. 09/03/19 16:17
94. <https://pfam.xfam.org/search/sequence>. 09/03/19. 17:26
95. Tesis Peressutti. 2018. Aislamiento y utilización biotecnológica de bacteriofagos del género Staphylococcus: su aplicación para el diagnóstico rápido y usos terapéuticos. FbioyF, UNR.
96. https://github.com/tmavrich/mavrich_hatfull_nature_micro_2017. 26/03/19. 10:29
97. [https://www.ncbi.nlm.nih.gov/genome/?term=Staphylococcus%20aureus\[Organism\]&cmd=DetailsSearch](https://www.ncbi.nlm.nih.gov/genome/?term=Staphylococcus%20aureus[Organism]&cmd=DetailsSearch). 29/03/19. 09:54